EMPIRICAL VALIDATION OF THREE THERMAL SIMULATION PROGRAMS USING DATA FROM A PASSIVE SOLAR BUILDING

Herbert Eppel & Kevin J. Lomas

ECADAP Centre, Institute of Energy and Sustainable Development,
De Montfort University, The Gateway, Leicester LE1 9BH, England
Fax: +44 116 257 7449 e-mail: herb@dmu.ac.uk

ABSTRACT

There is a continuing need to validate detailed thermal simulation programs of buildings. One way of doing this is to compare program predictions with measured building performance data. This is known as empirical validation. Data from the US National Institute of Standards and Technology passive solar test facility in Gaithersburg, MD, were used to assess predictions of ESP, HTB2 and SERI-RES. The results are tested for significance by means of Monte Carlo sensitivity analysis. A way of improving the power of empirical validation using side-by-side comparisons is suggested. The work provides benchmarks for testing other thermal simulation programs.

1. INTRODUCTION

The use of Detailed Thermal Simulation Programs of Buildings (DSPs) is increasing, but still limited in practice for a variety of reasons. On of them is the lack of confidence in their applicability and reliability. Validation can enhance the credibility of the programs, lead to improvements in the software and to increased confidence in their use. The validation of complex computer programs is not an easy task. In fact it has been argued that absolute verification is actually impossible [Channel 4, 1992]. Because of this, validation must be seen as an ongoing activity, with a number of complementary techniques being used to try and assess and improve program accuracy [e.g. Bloomfield at al, 1992].

The empirical validation technique [e.g. Lomas, 1991a] (Figure 1) is employed in this study to evaluate the performance of three DSPs, namely ESP-r [ESRU 1993a&b], HTB2 [Alexander, 1992] and SERI-RES [Palmiter & Wheeling, 1983; Haves & Littler, 1987]. All three are state-of-the art public domain programs. They use the finite difference method for modelling conduction, but a range of algorithms and submodels for the other thermophysical processes.

The program predictions will be compared with measured data collected at the US National Institute of Standards and Technology (NIST). This has been identified as one of only a handful of datasets available in the UK, which are suitable for empirical validation of 'building envelope' programs [Lomas, 1991b]. A way of increasing the power of an empirical validation dataset by using side-by-side comparisons is illustrated.

2. DESCRIPTION OF THE NIST PASSIVE SOLAR TEST FACILITY

The NIST Passive Solar Test Building is located in Gaithersburg, Maryland, USA, at 39 °N latitude, 77.3 °W longitude, and at an elevation of 127 m above sea level, on an open field site with no shading from the surroundings. The building is a rectangular one-story, slab-on-grade, wooden framed structure with the long axis running from east to west. It is divided into four cells which are separated from each other by heavily insulated walls (Figure 2) [Mahajan, 1984a&b; Sheheen, 1983].

Cell 4 is a 'direct-gain' test cell, with south-facing glass doors and a thermal storage wall on the north side. Cell 3 is a control cell with a more conventionally sized south-facing window. The investigation will initially concentrate on cell 4, with data from cell 3 being used later for a side-by-side comparison.

Both cells were modelled as one zone, i.e. they were not sub-divided further. The properties of the materials were generally taken from the NIST handbook [Sheheen, 1983] and were mostly based on the ASHRAE Handbook of Fundamentals [ASHRAE, 1981]. Standard CIBSE values [CIBSE, 1986] were used for surface heat transfer coefficients and air gap resistances where required by the programs.

The floor was modelled as a 1.3 m thick layer of earth (of which 0.3 m had real and 1.0 m had low capacitance) underlying the slab and gravel construction. This provides an adequate thermal resistance between the cell and the fixed deep ground temperature, but reduces the heat capacitance of the earth to prevent excessive model preconditioning times [Lomas, 1988].

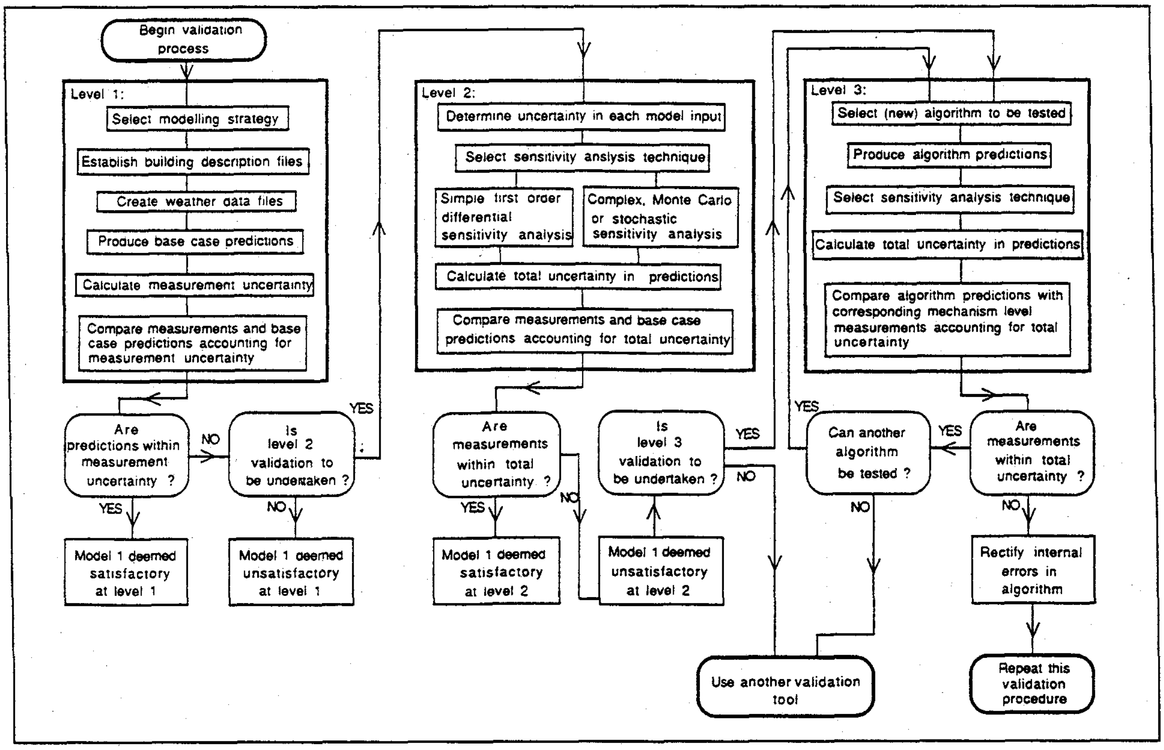

Figure 1: Three-level Empirical Validation Methodology [Lomas, 1991a]

The very close agreement between the measured value of the conduction loss coefficient (67 ± 7 W/K, excluding infiltration) [Sheheen, 1983] and the value calculated by SERI-RES (66.73 W/K) showed that, to within experimental accuracy, the building had been modelled correctly.

The cells and the surrounding climate were extensively monitored during several experiments, with data from one of these (January 24th to February 12th 1984) being used in this study (Figure 3). During the experiments no auxiliary cooling was provided. The heating setpoint temperature was

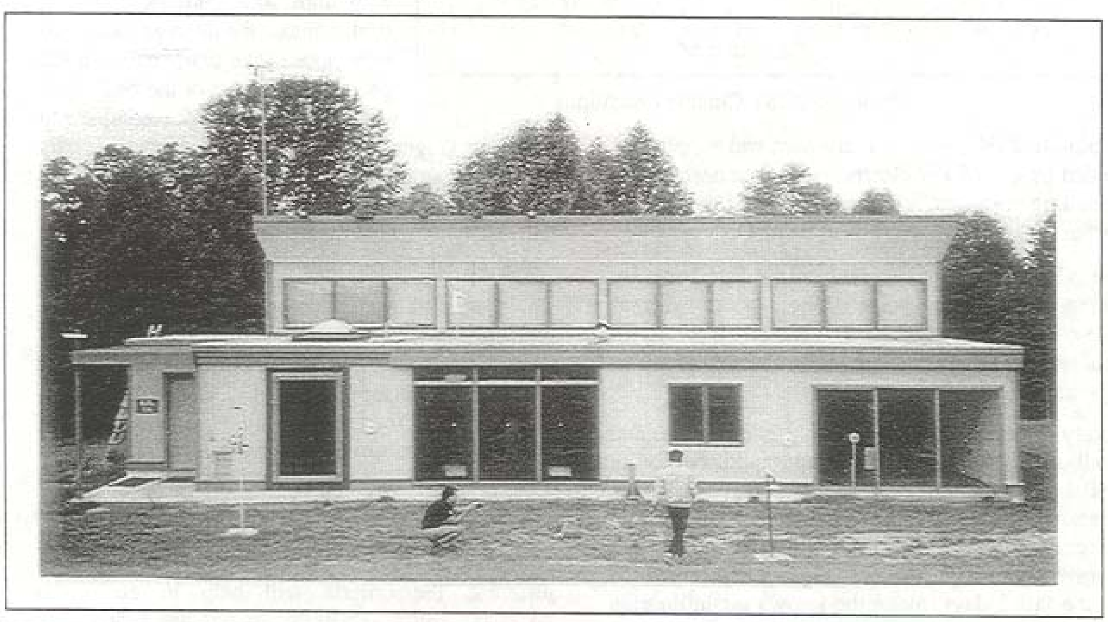

Figure 2: The NIST Passive Solar Test Building (South Elevation)

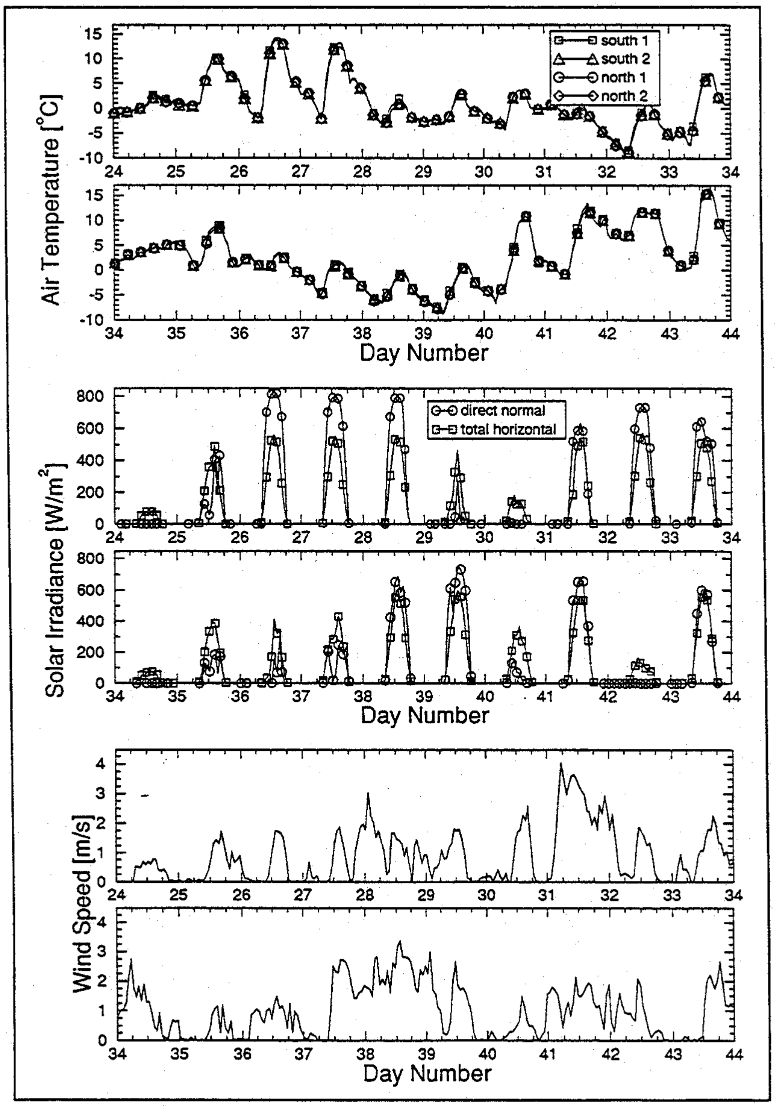

Figure 3: NIST Climate Conditions

nominally 20°C, and auxiliary heat was supplied as needed by a 3.76 kW electric resistance heater. For modelling purposes, it was assumed that the heating system was perfectly controlled to 20°C.

The NIST data were centered on the half hour, i.e. averaged from hour-point to hour-point, whereas ESP-r and the SERI-RES version used expect weather data to be centered on the hour. Appropriate adjustments had to be made.

Hourly measured values of both air infiltration and small internal heat gains (from equipment) were available, although air infiltration was only measured for the first 13 days of the 20-day period being considered. For modelling purposes, the average value from the previous 13 days was used for the last 7 days. Since the known air infiltration rates varied considerably (Figure 4), hourly values were used as input for the three programs. This was easily achieved in ESP-r by creating files containing appropriate values for each time step. In order to fairly assess the validity of HTB2 and SERI-RES, the two programs were modified to enable the hourly values of infiltration and casual gains to be fed in together with the climate data.

Care was taken to produce an optimum strategy for node placement, preconditioning time and the number of time steps. Four days preconditioning were found to be sufficient for the NIST simulations, leaving 16 out of the 20 days for the comparisons.

3. BASE-CASE RESULTS

The performance parameters of primary interest were the air temperature and the heating load in the test cells. The measured heating energy (Figure 4) was adjusted by deducting the energy consumed by the destratification fan, which had not been recorded separately (its energy use was added to the casual gains). The internal air temperature had been measured at numerous points and, although the range of the individual measurements was quite large (in cell 4 typically up to 6 °C for days with high solar radiation) due to stratification, the average value was very close to the temperature at the geometrical centre of the cell. This was therefore used for comparison purposes (Figure 4), because the programs only predict a single value. (SERI actually predicts a so-called enclosure temperature, which is a mix of air and radiant temperatures, rather than pure air temperature).

The measured values, together with the base-case predictions of maximum temperature and total heating energy for cell 4, are shown in Table 1 for the 16-day comparison period. ESP-r and SERI-RES both predict maximum and minimum temperatures within 1.5°C of the measured values. This is Excellent agreement according to the classification adopted in the Applicability Study (AS) work undertaken at DMU [Lomas, 1992]. Although the context of this work was somewhat different, the criteria will help in an initial appraisal, before sensitivity studies are undertaken. The maximum temperature predicted by HTB2 is

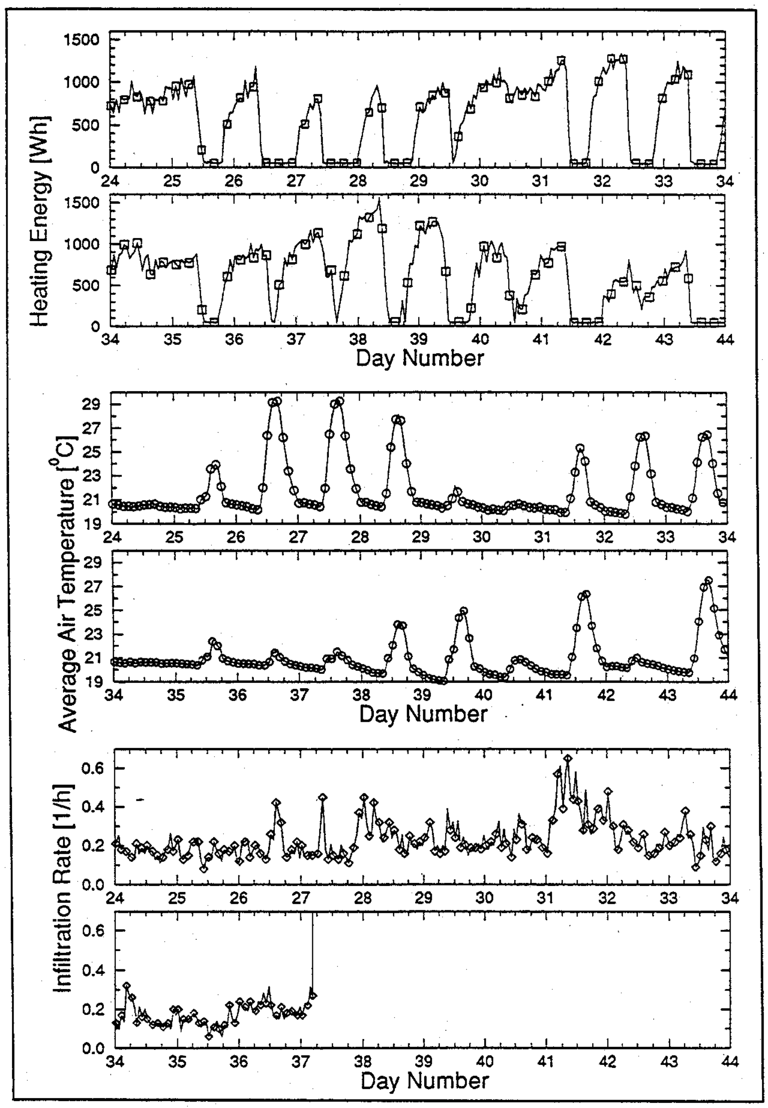

Figure 4: NIST Measured Performance (Cell 4)

2.9°C lower than the measured value. Nevertheless, this could still be classified as Good agreement.

Table 1: Comparison of measured and predicted air temperatures and heating energy demands

		Maximum Temperature	Total Heating Energy
Measured		28.0°C	219.1 kWh
ESP-r	predicted	26.5°C	132.9 kWh
	difference	-1.5°C	-39.3%
HTB2	predicted	25.1°C	250.5 kWh
	difference	-2.9°C	+14.5%
SERI-RES	predicted	27.7°C	209.3 kWh
	difference	-0.3°C	-4.5%

The total heating energy predicted by SERI-RES is below the measurement by just 4.5%. This could be classified as Excellent agreement. According to the AS classification, the heating energy predictions of both HTB2 and ESP-r were Bad, but the HTB2 value (over-prediction of 14.5%) was better than the ESP-r value (under-prediction of 39.3%).

A more detailed picture of the program performance emerges when one compares the predictions with the measurements on an hourly basis (Figures 5 and 6 - ignore the uncertainty bands for the time being). The severe heating energy under-prediction of ESP-r is illustrated in Figure 5. The jagged measured energy line suggests poor control of the heating system in the test cell.

The temperature plots (Figure 6) confirm the Excellent agreement between the temperatures predicted by SERI-RES and the measured values, which was observed when looking at the maximum temperature in isolation. The measured air temperatures, during periods when the heating was on, frequently lie above the setpoint (by almost 1°C on occasions), and sometimes below the setpoint. This could either be due to stratification effects or to poor control of the heating system response. These effects will be taken into account in the sensitivity analysis (Section 4).

A visual impression of the differences between measured and predicted air temperatures and heating energies, on an hour-by-hour basis, is given in Figure 7. An alternative way of showing these differences graphically is to plot frequency curves of temperature and heating energy differences (Figure 8). These plots confirms the earlier observation that SERI-RES performs best for both parameters, and that ESP-r underpredicts the heating energy demand.

The measurement uncertainty was quoted as ± 0.1% for auxiliary energy and ± 0.5°C for temperature [Mahajan, 1984a]. It is apparent (from Table 1 and Figures 5 and 6) that the predictions of all three programs differ by more than this measurement uncertainty. Therefore, all three programs are deemed to be Unsatisfactory at Level 1 of the empirical validation process (Figure 1), and a more detailed assessment at Level 2 of the methodology is required.

In the process of this work, a severe bug was discovered in a new release of one of the programs

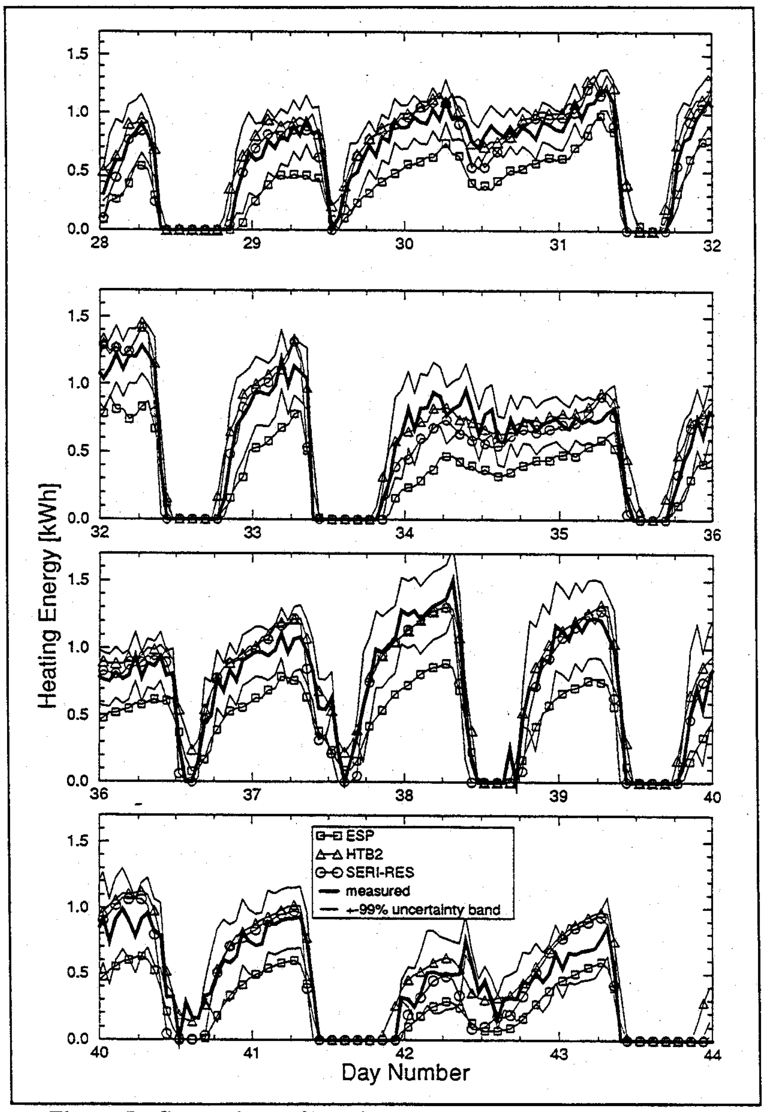

Figure 5: Comparison of Predicted and Measured Heating Energy Demands in Cell 4, and estimated 99-precentile uncertainty band

used. This had remained unnoticed by the program authors because it only appeared with certain types of heating system. This incident illustrates that validation has to be seen as an ongoing activity, undertaken by both program authors and program users. The routine validation of thermal simulation programs will be assisted by the production of computer programs which automatically test new software releases against existing benchmarks, such as those developed in recent IEA and European work [Jensen, 1993; Lomas et al., 1994; Judkoff & Neymark, 1994; Martin et al., 1994]. The datasets generated as part of this study can play a part in this process.

4. SENSITIVITY ANALYSIS

The comparisons so far have not taken account of the combined effect of the uncertainty in the measured performance of the test cell and the data describing it - which are entered into the programs. It is crucial (at Level 2) to account for these in order to assess whether the divergences between a program and the measurements are likely to be due to this uncertainty or to internal error in the program combined, perhaps, with inappropriate modelling assumptions. Much effort was therefore directed towards sensitivity analysis.

In a previous paper, three sensitivity analysis techniques, Differential Sensitivity Analysis (DSA), Monte Carlo Analysis (MCA), and Stochastic Sensitivity Analysis (SSA), were appraised using the programs, ESP-r, HTB2, and SERI-RES [Lomas & Eppel, 1992]. It was found that SSA had severe drawbacks, whereas DSA and MCA were particularly suitable for this type of investigation, and the Monte Carlo method was chosen for this study.

In MCA all the inputs are perturbed simultaneously, so the method fully accounts for any interactions between the inputs and, in particular, any synergistic effects. Furthermore any non-linearities in the input/output relationships are fully accounted for. It has been shown that, irrespective of the number of uncertain program inputs, only marginal changes in the estimated total uncertainties are obtained after 60 to 80 simulations [Lomas & Eppel, 1992]. The obvious disadvantage of MCA is that, because the inputs are varied simultaneously, the sensitivities of the predictions to the individual input parameter changes are not divulged, but for empirical validation this is rather less relevant.

Estimating the uncertainty in each input is a difficult task, although a great deal of progress was made during other recent projects [Lomas & Bowman, 1987; Pinney et al., 1991]. In this work, the upper and lower bound uncertainty values, such that there was only a very small chance (say 1%) that the actual values could lie beyond them, were estimated from published literature. The distribution of the errors was assumed to be roughly normal and the value midway between the upper and lower bounds was chosen as the base-case, or modal value, for each parameter.

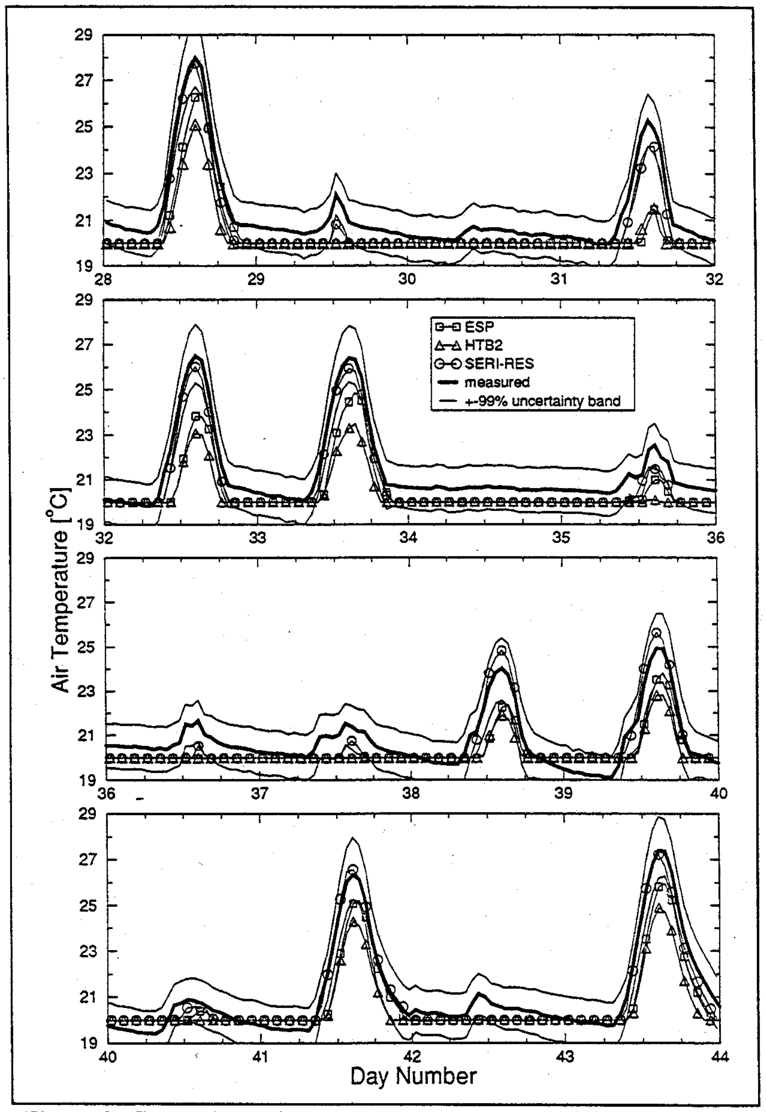

Figure 6: Comparison of Predicted and Measured Air Temperatures in Cell 4, and estimated 99-precentile uncertainty band

It can be seen that, whereas the ESP-r results clearly lie outside the uncertainty bands on more than 1 in 100 occasions (Figures 5 and 6), the heating energy results for HTB2 and SERI-RES are within those bands virtually all the time. Thus HTB2 and SERI-RES are deemed to be Satisfactory at Level 2 for heating energy predictions, whereas ESP-r is not. SERI-RES is also Satisfactory at Level 2 for its 'air temperature' predictions, whereas ESP-r and HTB2 are not.

Ideally, the sensitivity analysis should be undertaken for each program. However, as a first indication, and in order to reduce the time required for the analysis, only one program was used to estimate the total uncertainties. SERI-RES was chosen for this purpose, because the program requires the user to specify some highly uncertain parameters (in particular window U-value, and internal and external surface heat transfer coefficients). The total uncertainty was therefore likely to be greater, rather than smaller, than that which would be appropriate for the other two programs. (It has previously been shown that the three programs produced very similar sensitivities to changes in individual input parameters [Lomas & Eppel, 1992].) The uncertainty due to measurement errors was added to the total uncertainty estimated from 100 Monte Carlo runs.

5. DISCUSSION AND FUTURE WORK

The tendency for ESP-r to underpredict heating energy demands is consistent with the results obtained in the recent IEA Annex 21/ Task 12 empirical validation exercise [Lomas et al, 1994] and the intermodel comparison work (BESTEST) conducted by the same group [Judkoff & Neymark, 1994]. The better performance of SERI-RES is also consistent with the findings of the IEA empirical validation study. Attempts will be made to find the reasons for the discrepancies between the program predictions and the measurements.

In this work, the uncertainty bands were found to be very wide. Efforts to reduce this width would be beneficial for gaining further insight. One way of doing this is to perform a 'side-by-side' or 'matched pair' experiment. In such an experiment, the performance of two buildings, which only differ in one significant feature and are otherwise "as identical as possible" [Fracastoro & Lyberg, 1983], is compared. The uncertainty associated with the **difference** in performance (e.g. energy saving) should, in theory, be smaller than the uncertainty associated with the absolute performance values in two separate experiments (or buildings). Efforts are currently underway to apply this technique using the NIST data, and initial results appear promising.

6. CONCLUSIONS

- The continuing trend towards the use of Detailed Simulation Programs (DSPs) to analyse the thermal performance of buildings is recognised.

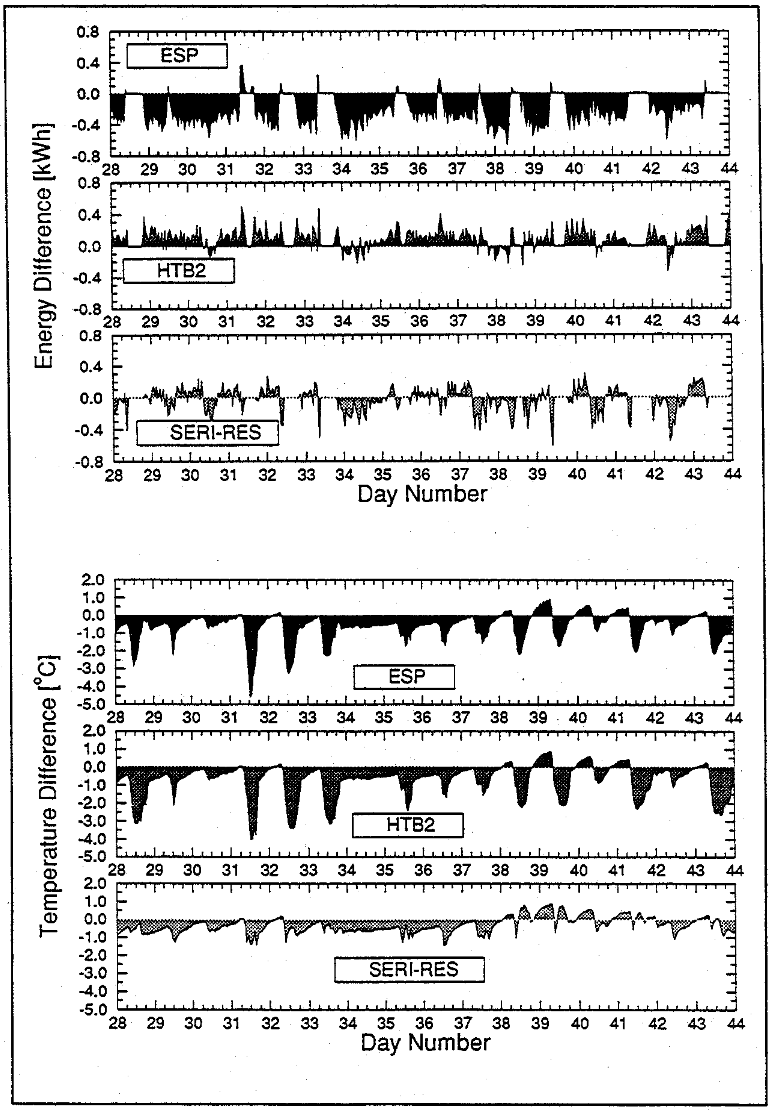

Figure 7: Hourly Differences Between Measured and Predicted Parameters

- The need for further program development and validation work is also recognised.
- A recognized empirical validation methodology was employed to compare the performance of ESP, HTB2 and SERI-RES with data collected at the NIST passive solar test site.
- The Monte Carlo sensitivity analysis technique is a powerful tool for determining the inherent uncertainty to be attributed to program predictions.
- The inherent uncertainty was found to be rather large in this particular case. A potentially generally applicable way of reducing the width of the uncertainty band has been suggested.
- The revelation of a serious bug in a new release of one of the programs confirmed that validation has to be seen as an ongoing process.
- The NIST empirical validation study is one in a series of recent high-quality validation exercises. The combined power of these validation tools can help to significantly enhance the predictive ability of thermal building simulation programs.

REFERENCES

Alexander, D.K., 1992, *HTB2 User Manual*, Release 1.10rev1, UWCC Cardiff, 125pp.

ASHRAE Handbook of Fundamentals, 1981, American Society of Heating, Refrigeration and Air Conditioning Engineers, Atlanta, GA, USA.

Bloomfield, D.P., Lomas, K.J. and Martin, C.J., 1992 *Assessing programs which predict the thermal performance of buildings*, BRE Information Paper IP 7/92, 3pp.

Channel 4, 1992, *Dispatches*, Television documentary, 4.3.92

CIBSE Guide, 1986, *Volume A: Design Data*, The Chartered Institution of Building Services Engineers, London

ESRU, 1993a, *ESP-r User Guide* Version 8 Series, ESRU Manual U93/2, University of Strathclyde, Glasgow.

ESRU, 1993b, *Summary of ESP-r's Product Model*, ESRU Report TR93/6, University of Strathclyde, Glasgow, 72pp.

Fracastoro, G.V. and Lyberg, M.D., 1983, *Guiding principles concerning design of experiments, instrumentation and measuring techniques,* IEA Energy Conservation in Buildings and Community Systems Programme Annex III, Swedish Council for Building Research, D II, ISBN 91-540-3955-X

Haves, P. and Littler, J., 1987, *Refinements to SERI-RES*; The final report from the Research in Building Group, Polytechnic of Central London, for the Energy Technology Support Group, ETSU S-1130.

Jensen, S.O. (Ed.), 1993, *The PASSYS Project, Subgroup Model Validation Development, Final Report*, Parts I and II, 1986-1992 Commission of the European Communities, DGXII.

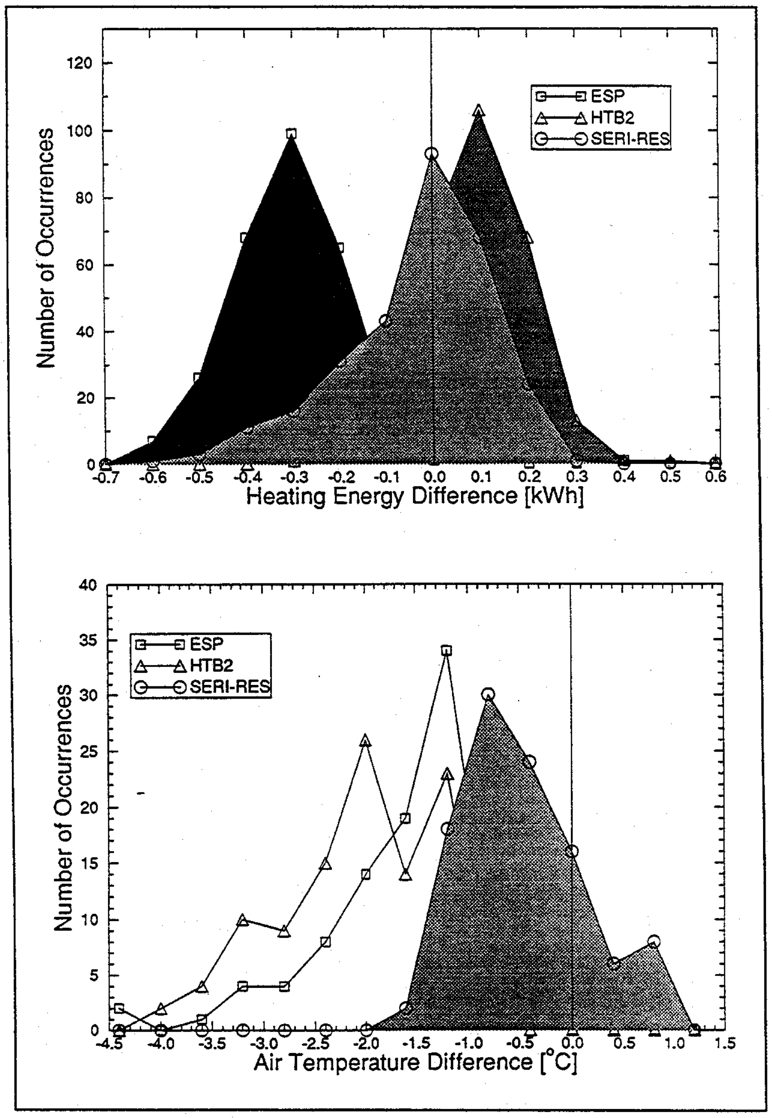

Figure 8: Heating Energy and Air Temperature Difference Frequency Curves

Judkoff, R. and Neymark, J., 1994, *Building Energy Simulation Test (BESTEST) and Diagnostic Method*, International Energy Agency Task Solar Heating and Cooling Task 12B/ Energy Conservation in Buildings and Community Systems Annex 21C, NREL/TP-472-6231, US National Energy Laboratory, Golden, CO

Lomas, K.J. and Bowman, N.T., 1987, *Developing and testing tools for empirical validation*, Ch.14, Vol.4 of SERC/BRE final report on "An investigation into analytical and empirical validation techniques for dynamic thermal models", SERC GR/C/62871, 79pp.

Lomas, K.J., 1988, *Ground modelling*, Applicability Study 1, Research Note 2, for ETSU, 6pp.

Lomas, K.J., 1991a, *Dynamic thermal simulation models of buildings: New method for empirical validation*, BSER&T 12(1) 25-37

Lomas, K.J., 1991b, *Availability of monitored hourly building performance data for validating dynamic thermal models of buildings*, BSER&T 12(2) 71-74

Lomas, K.J. and Eppel, H., 1992, *Sensitivity analysis techniques for building thermal simulation programs*, Energy and Buildings, 19, pp.21-44.

Lomas, K.J., 1992, *Applicability Study 1, Executive Summary*, ETSU S 1213, 75pp.

Lomas. K.J., Eppel, H., Martin, C. and Bloomfield, D., 1994, *Empirical Validation of Thermal Building Simulation Programs using Test Room Data, Volume 1: Final Report*, International Energy Agency Solar Heating and Cooling Task 12B/ Energy Conservation in Buildings and Community Systems Annex 21C, IEA21RR6/94, BRE, Watford, UK

Mahajan, Bal M., 1984a, *National Bureau of Standard passive solar test facility - instrumentation and site handbook*, NBSIR 84-2911

Mahajan, Bal M., 1984b, *Short duration winter-time performances of different passive solar systems*, NBSIR 84-2930

Martin, C., Seale, C. and Eppel, H., 1994, *Analytical testing of detailed thermal simulation programs*, Proc. BEPAC 94 Conf., York, UK, pp.S27-S32.

Palmiter, L. and Wheeling, T., 1983, *SERI-RES manual*, Version 1.0, Solar Energy Research Institute, Golden, CO, USA, 356pp.

Pinney, A., Parand, F., Lomas, K.J. Bland, B.H. and Eppel, H., 1991, *The choice of uncertainty limits on program input parameters within the Applicability Study 1 project*, Applicability Study 1, Research Report 18, for ETSU, 28pp.

Sheheen, Nancy, 1983, *National Bureau of Standards direct-gain test cell site handbook*, LA-9786-MS.