6829

Building and Environment, Vol. 28, No. 2, pp. 175–187, 1993.
Printed in Great Britain.

0360-1323/93 $6.00+0.00
Pergamon Press Ltd.

Empirical Validation of the Model SERI-RES Using Data from Test Rooms

C. J. MARTIN*
D. M. J. WATSON*

Comparisons have been made between the predictions of the thermal simulation model SERI-RES and measurements made in a double glazed test room. A sensitivity analysis technique has been used to determine the maximum deviation between simulated and measured results which can be accounted for in terms of uncertainties in the inputs supplied to the model, and the actual deviations observed have been found to be greater than this, implying that there are significant errors within the model. A novel analysis technique has been used to determine the source of those errors, and has concluded that the bulk of them stem from the operation of the auxiliary heat source in the room.

1. INTRODUCTION

COMPUTER models are increasingly being used to predict the thermal performance of buildings. For a new building they may be used at the design stage to determine aspects of the design, or to ensure compliance with regulations. For existing buildings they may be used to compare rival upgrade options. On a national level they can give guidance as to the energy saving potential of chosen alternative strategies. In the U.K., for example, the Department of Energy Passive Solar Programme makes extensive use of building thermal simulation models to predict the energy savings which can be obtained by incorporating passive solar design features in buildings.

In all such applications it is clearly important that the models used produce reliable predictions, and that they have credibility. A range of exercises can be performed to build confidence in the programs, and, taken together, these are referred to as 'model validation'. The first such exercise is analytical testing, where the model is asked to make predictions for cases which are so simple that the correct solutions can be calculated explicitly. The second type of test is the intermodel comparison, in which the predictions of alternative simulation programs are compared. The third type of test which can be applied to a thermal simulation model is empirical validation, in which the predictions of the simulation model are compared with data measured in real buildings. The whole process of model validation was analysed extensively in a recent U.K. study carried out jointly by the Building Research Establishment and the Science and Engineering Research Council [1], and the work described in this paper attempts to build on the conclusions of that study.

The appeal of empirical validation as a means of increasing the credibility of a building simulation model is immediately apparent. To represent a real building any model inevitably makes a great many simplifications. Some of these simplifications may be made in the interests of, for example, simplifying data entry or increasing computational efficiency. The impact of these approximations can be assessed by comparison with an alternative model which does not make them. Other simplifications however, are made of necessity rather than convenience. The underlying physical processes may be so complex that they cannot be modelled thoroughly, or that the uncertainties associated with attempting to model them are so large as to render the results useless. Empirical validation represents the only way in which the impact of this second type of simplification can be assessed.

Empirical validation may at first appear to be a simple task, but in practice it is extremely difficult to carry out convincingly. The many factors which influence the thermal performance of a building interact in a complex way, and are closely related. When agreement between model predictions and measured data is poor, it can be very difficult to determine the source of the discrepancies observed. More seriously, when agreement between predictions and data is good, it may only be as a result of several errors cancelling out.

To explore further the form that an empirical validation project should take it is necessary to specify more succinctly just what can be expected from such an exercise. It would be unrealistic to suggest that a single test could in some sense 'certify' a simulation model as being completely satisfactory. Rather, the process of confidence building will be gradual, each empirical validation exercise indicating new areas of application in which the model may be used with confidence.

In situations where the model performs well, and this good performance can be shown not to be the result of several errors cancelling out, credibility is immediately enhanced. However there is a second possible outcome to any validation activity: the model may fail to perform satisfactorily. In this case the validator will be called upon to identify the sources of the observed errors. Armed with

*The Energy Monitoring Company Ltd. Unit 3, Chapel Court, Newport Pagnell, MK16 0EW, U.K.

this information, the model developer can return to his or her code, and improve the performance of the model. Once again, credibility is increased, although this time it may be with the associated cost of doubt cast on previous predictions of the model. It is however preferable to have an improved model and to understand that previous simulation results may be uncertain than to continue to use a model containing errors in a state of ignorant bliss. In many respects validation work which finds errors in a model, and then allows them to be identified and rectified, is even more valuable than a test which simply concludes that the model performs acceptably in one particular application.

There are thus two aims in any validation task: being certain that good performance implies a good model, and being able to identify the sources of bad performance. To achieve these aims, great care is required when designing the empirical validation experiment, when collecting the data, and when carrying out the simulations for comparison with that data.

Preliminary validation attempts of any kind will seek to eliminate as many sources of complexity and uncertainty as possible. In this way the number of competing influences on the performance of the test building are minimized, making it simpler to achieve the aims identified above. In the case of empirical validation this simplification will largely be achieved by appropriate choice of test building. Inevitably there will be a compromise in the choice of building—too complex a building may defeat the aims of validation, but a structure which is so artificial as to be unrepresentative will mean that the findings of the project will lack credibility with real building practitioners.

In the past, individual test rooms have proved to be useful vehicles for empirical validation. Uncertainties due to occupancy, complex interactions with other building zones, infiltration and ground floor heat loss can be eliminated. At the same time such rooms can be exposed to real climate, can be of a representative size and can retain realistic construction details. The Energy Monitoring Company (EMC) currently operates a test facility where nine such test rooms are sited, and these have been used to provide the experimental data used in the work described here. The site and the test rooms are described in detail in [2].

This paper describes a model validation exercise carried out at the EMC test facility. In Section 2 some of the problems previously encountered when comparing data not originally intended for validation with the predictions of simulation models are reviewed. From this there follows a discussion of how best to operate test buildings for validation purposes, and of how subsequently to operate the model for the purposes of comparison. In Section 3 a validation experiment using these improved techniques is described. Preliminary comparisons between the measured data and simulation results indicate that there are discrepancies between the two. A differential sensitivity study is used to determine whether these divergences are significant, that is whether they are larger than could be expected due to uncertainties in the inputs to the model alone. When it is found that they are, a cross-correlation/deconvolution technique is used to determine which driving forces (and hence which areas of the model) are responsible for the errors. The paper concludes that test rooms have been shown to be a useful vehicle in which to carry out empirical model validation. Further work which is now underway is briefly described.

2. OPTIMAL OPERATION OF TEST BUILDINGS AND SIMULATION MODELS FOR EMPIRICAL VALIDATION

2.1. *Review of problems inherent in existing data sets*

Previous attempts at empirical model validation using data from the EMC test roms, had revealed that the available data sets were not ideally suited to the task [3]. Two principal problems were identified:

(a) Cross-correlation analysis of the resulting simulation errors failed to identify correctly the source of those errors [3]. The reason for this was found to be correlations between the forces driving the rooms: solar radiation, external temperature and auxiliary energy input.

(b) In the interests of realism the data sets, which had been gathered as part of a previous project, featured thermostatic control of the rooms to a fixed setpoint for part of the day. In other words, a control device was used to vary the power input to the rooms, to achieve and then maintain a given temperature within each room.

Significant problems arose when comparing the simulated and measured results. The comparison was carried out by supplying to the model the heating setpoint schedule that had been implemented in reality. During periods when the heating was off the heat input was, of course, correctly predicted as zero, but there were errors in the predicted temperatures as the building reacts to other influences. When the heating cycle began there were errors in the prediction of the amount of energy required to bring the room to the setpoint, and simultaneously errors in the predicted temperatures as the room warms up. Finally, when the room and simulation model reached the setpoint there were no further errors in the predicted temperature, but there were errors in the prediction of the energy required to maintain that setpoint.

Clearly there is nothing that can be done about correlations between solar radiation and external temperature—these variables are beyond the experimenters' control. However, it may at least be possible to choose a heater operation strategy which will reduce, or perhaps even eliminate, correlations between this driving force and the others, and this possibility was explored before collecting the data sets which will be analysed here.

2.2. *Choice of test room operating strategy*

In designing test room experiments to collect data for the purposes of model validation, a number of aspects of room operation are under the control of the experimenter. The most obvious control mechanism is that of auxiliary heater scheduling. In an attempt to reduce the problems outlined in the previous section three possible heater operating strategies were considered:

—operating the test rooms in pairs each with a 12 h ON/12 h OFF schedule, but in antiphase between the two rooms. One test room is heated from 6:00 to 18:00, and the other from 18:00 to 6:00 the following day. Taken individually the data sets will suffer from exactly the same problems as those collected previously, but analysing the differences in performance may allow the effects due to heater operation to be separated from those of climate. This technique has proved useful in previous work [4].

—operating the test rooms to a thermostatic schedule which is not a multiple of 24 h. With this strategy the operation of the heater drifts in and out of phase with the meteorological variables, allowing the different sources of error to be separated. Again, the technique has proved useful in previous test room work [5].

—randomizing the operation of the heater to ensure that its output does not correlate with external variables. This is a well known system identification technique which has been used in the past to identify the response of test buildings [6, 7]. To our knowledge, however, it has not previously been used in the context of empirical model validation.

It seems that any of these operating strategies could provide data sets which will allow the source of simulation errors to be isolated more effectively. However, the first suffers from a number of disadvantages: it requires that the test rooms are configured in identical pairs for each experiment, implicitly assuming that these pairs are perfectly matched. The resulting data then has to be interpreted in terms of the difference in performance between the pair of identical buildings operated with different heating strategies. Any attempt to interpret the data in the way in which the model will normally be used. to make predictions about a single building, causes the problems identified in the earlier data sets to reappear. The second strategy appears to solve these problems, providing data from a single test room which will, over time, allow the influence of the auxiliary heat source to be separated from that of climate. The problem with this operating strategy is that although the chosen heater operation will not be correlated with meteorological variables it still has a statistical structure of its own, in that it consists of the same sequence repeated many times, which in turn makes the extraction of the causes of simulation errors less robust.

The third strategy avoids these problems. One criticism which can be levelled at such randomised operation of the room heat source is that it is unrealistic. However, the purpose of model validation is to establish that the representation of a building within the model is satisfactory over a wide range of operating conditions. In fact, the type of randomized heater operation proposed represents a very stringent test of any model. Consider a heater sequence in which the heater is either on or off, and its state at each time step is decided randomly, for example by tossing a coin. In this sequence information about the past operation of the heater provides no information about the likely future operation. This property will make the sequence a very difficult one for a simulation to follow: it minimizes the chance of errors from different sources cancelling out.

One potential problem with the random sequence described is that it gives relatively little weight to the longer time constants present in buildings. Simple calculation indicates that if such a sequence is produced at hourly timesteps there will be very few periods of uninterrupted heating longer than about six hours. Since inverting a sequence of this type yields another sequence of the same type there are also few uninterrupted cooling periods longer than about six hours.

The EMC test rooms have a relatively short time constant, of between 8 and 12 h. For these rooms the heater sequence described is appropriate. For buildings with longer time constants, however, the sequence is unlikely to stress sufficiently the slower heat transfer processes. One solution to this problem, which has been proposed for use in the PASSYS project, is the Randomly Ordered Logarithmic Binary Sequence (ROLBS) [8], in which the required number of occurrences of each length of heating pulse is specified and the pulses are then arranged in random order. Whilst this sequence retains its lack of correlation with the meteorological variables driving the room, the accentuation of the lower frequencies inevitably causes unwanted auto-correlations within the test sequence. A second solution, currently being investigated at EMC, uses a digital filter to emphasize the low frequencies in a white noise test sequence. Because the filter which was used is known, its inverse can be used to 're-whiten' the resulting experimental data before it is analysed. However, because of the fast response of the EMC test rooms neither of these techniques was considered necessary here, and it was decided that the rooms would be operated with a straightforward randomized heater schedule of the type initially described.

2.3. *Operation of the model*

The notion of 'blind' simulation runs is central to the approach to model validation adopted in this work. The model user who attempts to predict the performance works in ignorance of the measured values which he or she is called upon to predict. The reason for insisting on this mode of operation is simple: it has been demonstrated [3] that it will almost always be possible for a modeller to adjust the input parameters of the test room model to bring its predictions into line with observations. In the case of unconscious adjustment, bias will be introduced into the simulation results which may vary between users and between models. In the case of conscious adjustment, the inputs to the model can almost always be systematically altered (within physically reasonable bounds) to achieve a high level of agreement—empirical validation as a test of the model becomes useless.

For the simulations described here, no modifications were made to the SERI-RES descriptions of the EMC test rooms which had been prepared some years prior to data collection [9]. We can thus be confident that the simulations were carried out 'blind'.

In the previous section, randomized operation of the test room heat source was chosen, on the grounds that it would yield data most likely to allow the errors due to different test room driving forces to be isolated. Perhaps the obvious way in which to carry out a simulation to compare with this data is to apply the same pseudorandom sequence to the model of the test room. The resulting simulations of zone and surface temperatures can then be compared with the measured values to assess the quality of the simulation. We will refer to this mode of model operation as 'power scheduled'.

The heater power-scheduled approach to model operation has a further important benefit when used with the model SERI-RES. The zone temperature predicted by SERI-RES can only be interpreted directly as an air temperature under a fairly restricted set of operating circumstances [10]. When these requirements are not fulfilled, employing the simulated zone temperature as a thermostat temperature results in the simulation being physically incorrect. In a power-scheduled run these discrepancies will still be apparent when predicted zone and measured air temperatures are compared, but predicted and measured surface temperatures and heat fluxes should still be predicted correctly, and can be compared directly.

The principal disadvantage of heater power-scheduled operation of the model is that it provides no information about how good the model is at predicting the energy consumption of the room, the actual consumption having been fed to the model as part of the simulation process. Since the prediction of auxiliary energy consumption is one of the principal applications of the model, its performance on this task is of interest. To obtain predictions of energy consumption, the model can be fed with hourly details of the temperature within a test room. If this value is used as the setpoint, the model will predict the amount of auxiliary energy (which may be positive or negative) that is required to maintain that temperature. This prediction can then be compared with the actual energy consumption measured. The model also produces predictions of the surface temperatures and heat fluxes which can, again, be compared with the measured values. We refer to this mode of operation as 'zone temperature scheduled'.

To summarize, there are two ways of operating the simulation model. The first, heater power scheduling, is to apply the same heat input to the model as was applied to the test rooms, and observe the quality of the predictions of the resulting temperatures and heat fluxes. The second approach, zone temperature scheduling, is to force the model zone temperature to follow the air temperature measured in the test room, and examine the prediction of the energy required to do this, and of the resulting surface temperatures and heat fluxes. The consistent use of one of these strategies throughout a simulation run will eliminate the types of problems previously encountered in interpreting the results of comparisons between the measured data and corresponding simulations. The availability of these twin strategies is in no way dependent on the fact that the room is being operated with a randomized heater schedule rather than under thermostatic control. Either technique can be used with data sets featuring any kind of heater control, including unheated operation.

3. COMPARISON OF MODEL PREDICTIONS WITH MEASURED DATA

Data was collected from three test rooms equipped with single glazing, double glazing, and an insulated opaque infill panel, for a period of 50 days. The data were recorded at five-minute intervals, and subsequently integrated into hourly average values for use with SERI-RES, which was used to produce predictions at hourly intervals. The data set is continuous over the 50 day acquisition period.

A total of six simulations were then carried out, comprising a power scheduled and temperature scheduled run for each of the three rooms. The results presented here are based on the results for the double glazed room only.

3.1. *Preliminary comparisons between simulation results and measured data*

Figure 1 shows the simulated zone temperature and the air temperature measured in the double-glazed test room, over five days of the overall 50 day experimental period. The period shown was chosen because it contained a good range of meteorological driving forces, and thus demonstrates the full range of test room response. This simulation was 'power scheduled' and thus the model has been called upon to predict zone temperature (as shown), surface temperatures, and surface heat fluxes. Apart from an obvious offset, the initial appearance is of quite good agreement, with the model appearing to follow the temperature fluctuations observed in reality quite well. Figure 2 shows the difference between the simulated and measured temperatures, hereafter referred to as the 'simulation error'. Examination of Fig. 2 reveals quite large temperature prediction errors, which were not immediately apparent on Fig. 1, and indicates one of the dangers of presenting empirical validation results in that form alone.

The next stage in the validation process is to determine whether these discrepancies are significant in the context of an empirical validation trial.

3.2. *Sensitivity studies*

Any detailed model of a building will, by definition, require as input a great deal of information. Generally this will include parameters describing the location of the building, the geometry of the structure, the thermophysical properties of the materials used in the construction of the building, and the heating and/or cooling plant installed in the building. When a simulation is actually performed using the model of the building further inputs are required, in the form of meteorological data and information about how the building was operated. Inevitably, there will be uncertainties in all of these inputs. In this section a definition of whether a given discrepancy between simulation results and data is significant is adopted which is based on whether the uncertainties in inputs to the model can account for the discrepancy observed.

To carry out this test, the maximum amount of discrepancy which can be accounted for in terms of uncertainty in the input parameters supplied to the model must be established. If the discrepancy which has been observed is greater than this, the model can *never* be made to represent the observed reality by adjusting the input parameters within their accepted ranges. In this case the model itself is deficient—it cannot adequately represent reality. If the observed discrepancy between

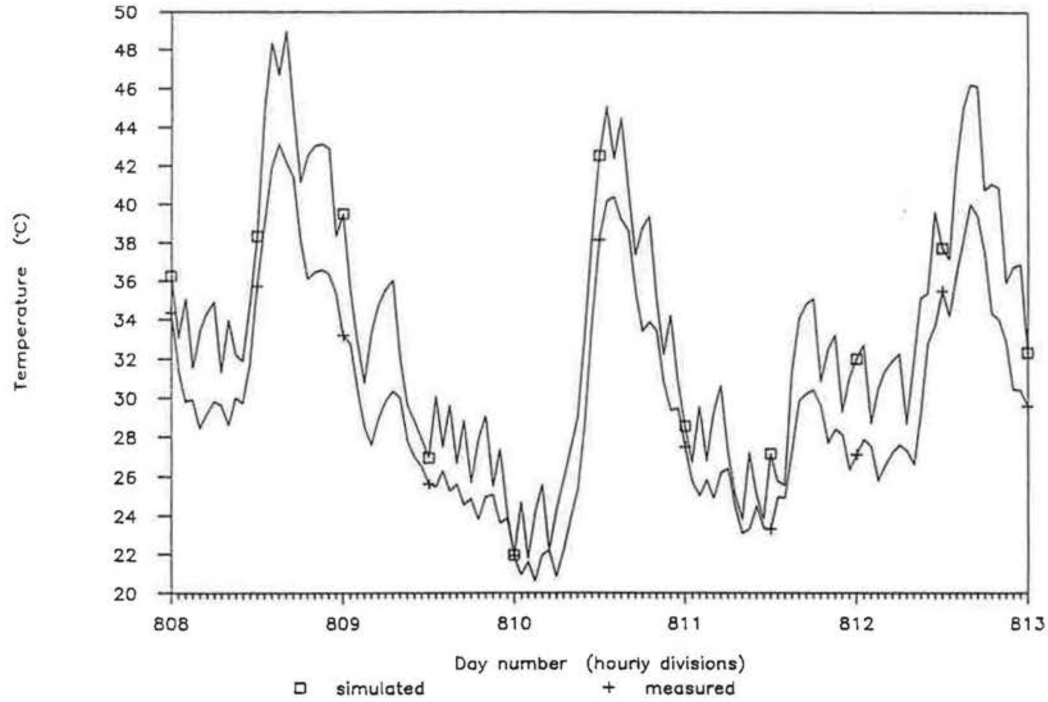

Fig. 1. SERI-RES predicted and measured zone temperature in a double-glazed test room.

model and reality can be entirely accounted for by uncertainties in the input parameters then the body of the model *may* be without fault (or 'valid'). In this case the only way to proceed is to obtain better values for at least some of the input parameters. When this is done the consequent uncertainty in the predicted values will decrease, increasing the power of the validation test. At this stage the input uncertainties may no longer be able to account for the discrepancies observed, and the model is concluded to be flawed, or it may be found to be satisfactory even at this higher level of accuracy, increasing credibility yet further.

The uncertainties in the model parameters may arise from two sources. The first is measurement error when the parameters are determined. For example in measuring the thermal conductivity of a slab of insulation

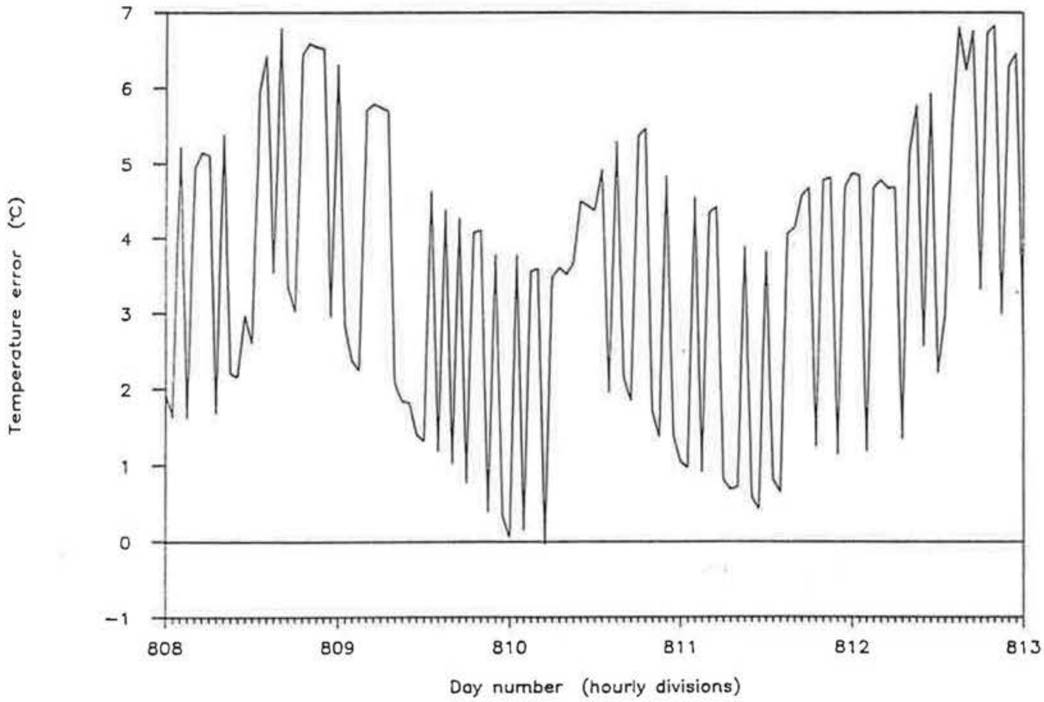

Fig. 2. SERI-RES zone temperature prediction error in a double glazed test room.

there are errors in the measurement of the temperature drop across the slab, and in the measurement of the resulting heat flow. As a consequence there will be uncertainty in the quoted conductivity. The second source of error is due to bad parameterization. In the case of the insulation, for example, the real conductivity may not be the single constant value which is sought. There may be variation with age, with manufacturing batch, or with operating conditions which are not accounted for.

The net effect of both these types of error is the same—incorrect values are used to describe the properties of the building which is to be simulated. What is required is a method of establishing the effect of each of those uncertainties on the output of the model, and of combining those effects to obtain the overall output uncertainty. One technique for performing this analysis was proposed in the SERC/BRE study [1], and has been applied here. For each input parameter, a simulation is carried out with the parameter set to its maximum possible value, and a second is performed with the parameter set to its minimum possible value. The change in simulation output from the base case simulation, in which all the input parameters were set to their nominal values is then calculated for both cases. This process is repeated for all the model inputs (or, more often, for the subset believed to have the most significant impact on the results). The resulting output perturbations are then grouped into those tending to increase the output, and those tending to decrease it. The magnitudes of these increases and decreases are then combined in quadrature to obtain an estimate of the overall increase or decrease in output which could occur as a result of uncertainties in the input parameters supplied to the model. When uncertainty bands are to be established for a succession of predictions, for example hourly temperature values, this procedure is carried out for each point. One consequence of this is that if variations in a parameter cause errors of different signs at different times that parameter will implicitly be assumed to vary over time, always adopting its 'worst case' value. For some parameters this may be a reasonable assumption, for others (for example dimensions) it is clearly not.

As described, the technique also requires a number of assumptions concerning the linearity of the model with respect to its input parameters, and the statistical independence of the uncertainties in those parameters. Whilst many of these assumptions may be valid, some can be refuted directly [3], and thus the method will not be exact. However, recent comparisons between the results of this technique and more sophisticated Monte Carlo methods has indicated that agreement is generally close [11].

Figure 3 shows the error in predicted zone temperature, hour by hour, which can be attributed to uncertainties in the 27 most important inputs required by the SERI-RES model of the double glazed EMC test room. A total of 55 simulations have thus been carried out to generate these bounds. Also shown on the figure is the simulation error which was actually observed. The interpretation of the figure is straightforward. If the observed error always lies within the range which can be attributed to uncertainties in the input parameters then these uncertainties can account for those errors, and the model itself cannot be faulted. If, however, the observed error falls outside the range generated by the sensitivity study then that error cannot be accounted for in terms of input uncertainties, and the model itself must be at fault. This is clearly the case for the data shown on Fig. 3, and it is concluded that the simulation errors which we have observed point to a significant failing in the model SERI-RES when used to model a heated test room.

3.3. *Detailed analysis of simulation error structure*

The analysis of the previous section has demonstrated that the divergences observed between the measured data and the simulations of SERI-RES cannot be accounted for in terms of uncertainty in the model input parameters alone—that is they are *significant* to the empirical validator. This in turn implies that there are one or more errors within the model itself. The next step in the empirical validation process is to locate the source of those errors. If this is achieved, recommendations can be made as to which parts of the model should be refined.

A number of diagnostic techniques were outlined in the SERC/BRE model validation project [1]. The most powerful of these, when it comes to identifying sources of error, centre around cross-correlation analysis. These techniques will be exploited and subsequently extended here.

We begin by examining the cross correlations between the simulation errors and the quantities driving the building. The latter fall into two categories: those which are used in the simulation, modelled driving forces, and those which are not, which we term unmodelled driving forces. In the first category fall external temperature, solar radiation and auxiliary energy input. In the second category are wind data, and the net sky to ground radiation exchange at the site.

3.3.1. *Cross-correlation of errors with driving forces.* The cross correlation function gives a measure of how closely two sequences of data are related, as a function of time delay. In analysing the discrepancies between building models and real buildings the inclusion of time delay is essential, as both systems are dynamic, and thus errors may be delayed from the driving forces which cause them.

The cross correlation function always lies within the range -1 to 1. Its absolute value gives a direct indication of how closely the two sequences are related, a value of 1 implying perfect correlation, and a value of zero implying no dependence at that particular time delay.

Figure 4 shows the cross-correlation of the error in predicted zone temperature in the double glazed room with the modelled driving forces, and Fig. 5 shows the corresponding correlations with the unmodelled driving forces. All of the cross-correlation functions contain elements that are statistically significant, that is they lie outside the 5% confidence intervals shown on the graphs. In previous work, however, it has been demonstrated that examination of cross-correlation functions alone is not a conclusive way of identifying the source of simulation errors, because of the inter-correlations between the quantities driving the rooms [3]. To separate out the influences of the different driving forces, and express

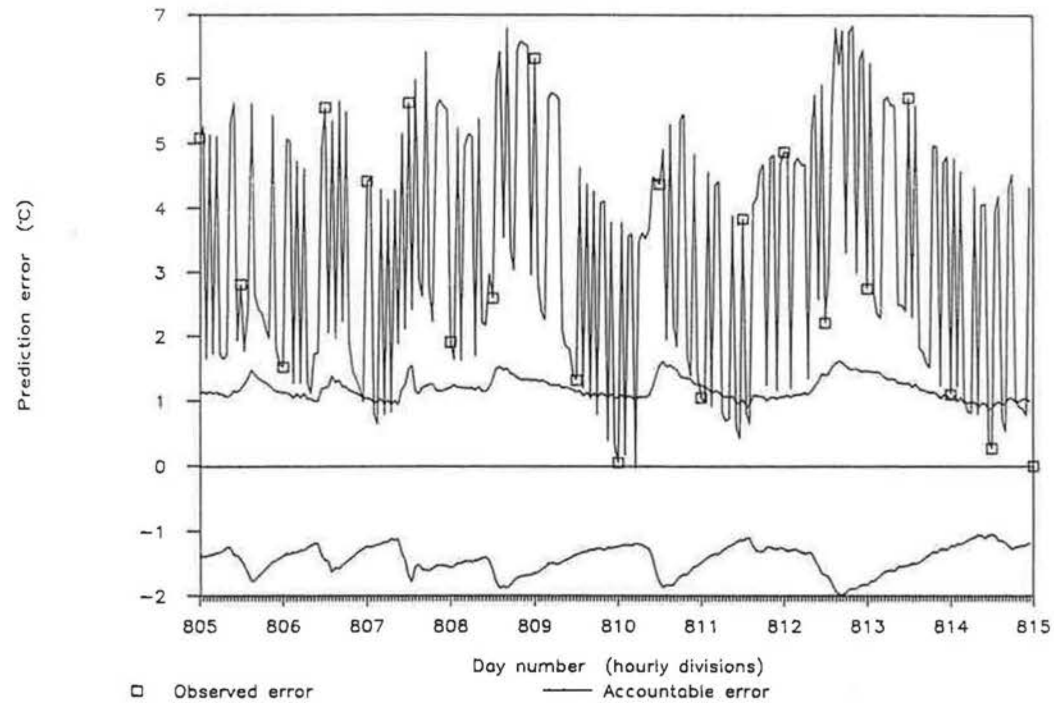

Fig. 3. Simulation error attributable to uncertainties in input parameters and actual error.

them in a form which is independent of the particular weather statistics during the validation experiment a deconvolution technique is employed.

3.3.2. *Deconvolution of errors due to driving forces.* In order to separate out the various sources of error a simple model of the error process is postulated. The proportion of the error in any predicted variable due to a given driving force is assumed to be related to that driving force by a linear, time invariant dynamic system. In reality, these assumptions will always be violated to a greater or lesser extent. For example the dependence of simulation error on solar radiation may vary over time as solar geometry changes. Other sources of error may violate the linearity assumption, for example errors in external surface heat loss may depend on wind velocity, solar

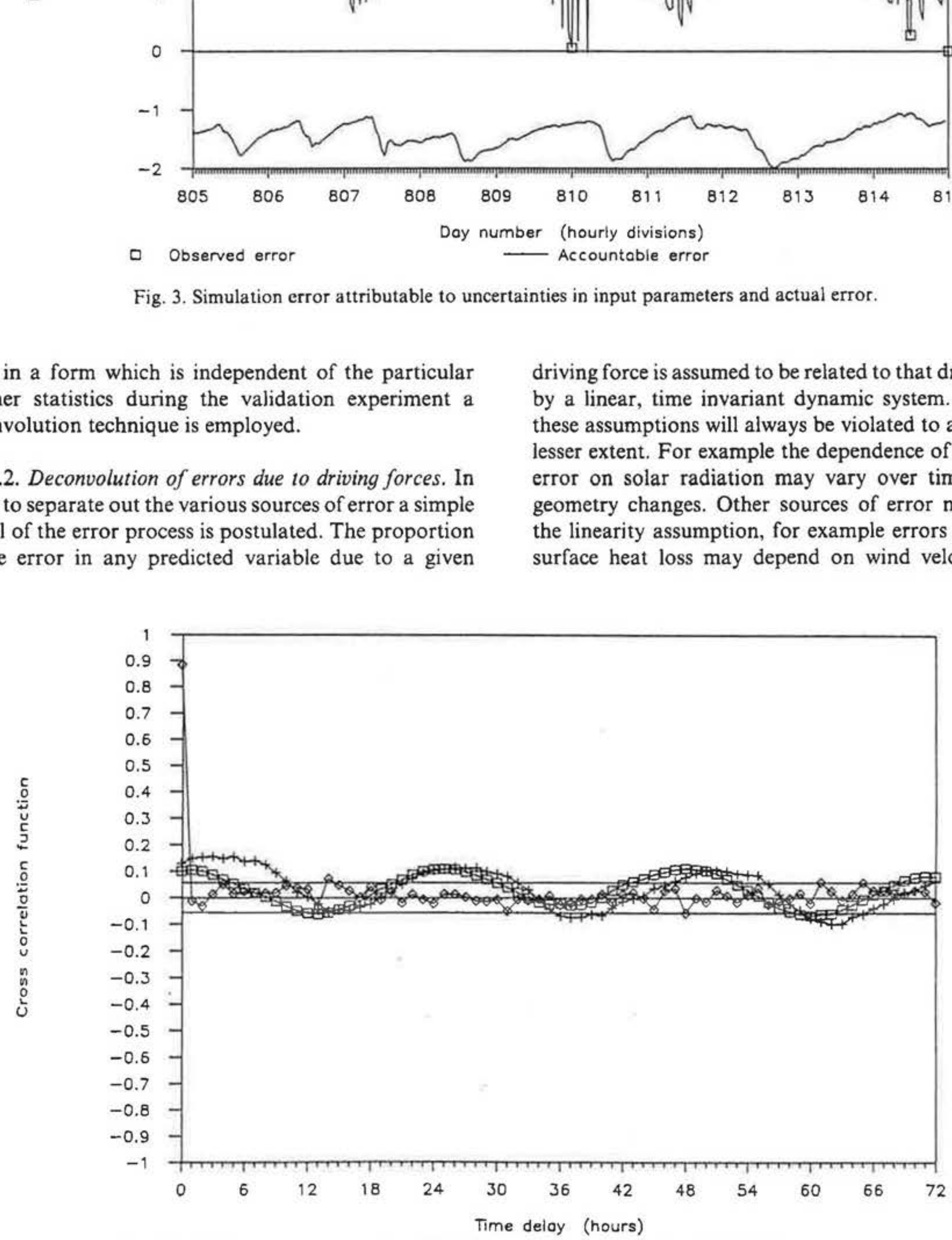

Fig. 4. Cross-correlations between zone temperature prediction error and modelled driving forces.

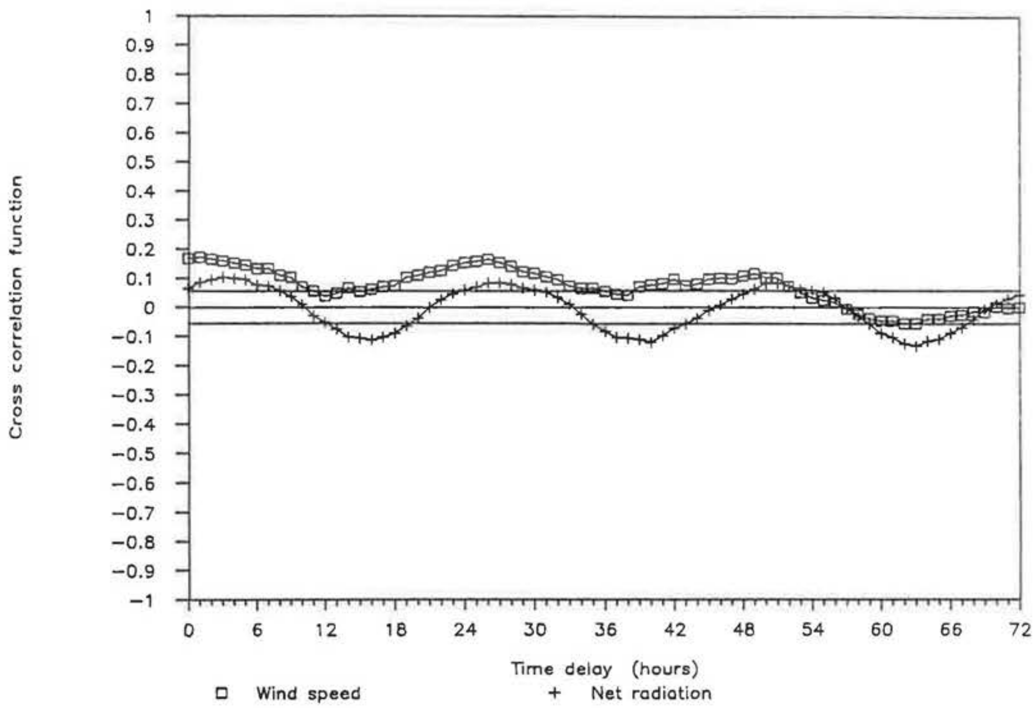

Fig. 5. Cross-correlations between zone temperature prediction error and unmodelled driving forces.

radiation and net radiation exchange in a complex, nonlinear way. We shall see however, that the assumptions of time invariance and linearity allow a large proportion of the simulation errors to be accounted for.

Each of the linear, time-invariant systems which describes the way in which simulation error can be related to a driving force is fully characterized by its impulse response. A technique has been developed which allows these impulse responses to be extracted from simulated and measured data, and measurements of the driving forces. The technique is described in outline in the Appendix to this paper. The results presented in [3] used a simplified form of the same analysis, also described in the Appendix.

Figure 6 shows the result of deconvolving the contributions of the principal driving forces to the error in simulating the test room zone temperature. The system inputs are the driving forces defined earlier: heater power, solar radiation, ambient temperature, wind speed and net radiation exchange in turn. The impulse responses shown relate the difference between the simulated and measured test room zone temperatures dynamically to the driving forces listed. To obtain a representative plot, the impulse responses have been scaled by the mean value of each of the driving forces. This choice of scaling has a straightforward interpretation. If each of the forces driving the test room were simply constant at its mean value over the period of the experiment then the contribution from each source to the simulation error would equilibrate at a constant value given by the area under the impulse response curve multiplied by the mean value of the driving force. In practice, of course, the driving forces are not constant, and their relative contributions vary over time. However, this scaling does serve to give a reasonable indication of the relative importance of each source of error. The figure demonstrates that by far the most significant source of error when predicting the test room air temperature is operation of the auxiliary heat source, the contributions from the other driving forces being insignificant in comparison.

Figure 7 shows the impulse responses contributing to back wall temperature error. These responses are significant in magnitude, and we see that there are significant contributions to the error from heater operation and solar radiation.

Finally, Fig. 8 shows the corresponding disaggregation of the errors in the back wall surface heat flux. Here, once again, the contribution from the auxiliary heat input overshadows that from other sources of error. Interestingly, the error is seen to be positive for the first hour (the simulated heat flux is larger than that measured), and then becomes negative for the following three hours. The overall integrated error is thus close to zero, which indicates that very good agreement will be observed when the long term mean values are examined. This conclusion has been confirmed by reference back to the original simulated and measured results. This demonstrates clearly the dangers inherent in using simpler heater schedules to produce model validation data. Consider an experiment in which the heater is simply operated at a constant power. After the first few hours of the experiment the error in the rediction of back wall heat flux which has been detected here will become self cancelling, leading to the erroneous conclusion that the quantity is being correctly predicted. This problem occurs because of the inappropriate choice of heater schedule, and is avoided when heater operation is randomized, demonstrating further the power of such heater schedules.

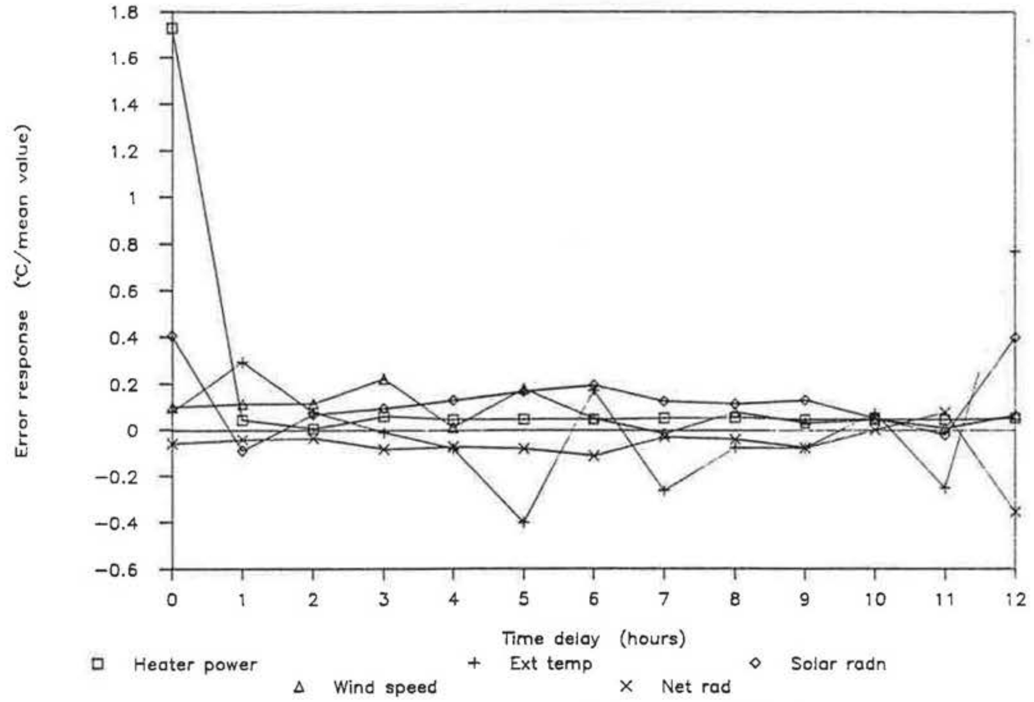

Fig. 6. Response of zone temperature prediction error to driving forces.

3.3.3. *Reconstitution of observed errors.* Once the error impulse responses have been obtained, it is possible to generate the hour by hour contributions of each of the driving forces to the overall discrepancy between the simulated and measured quantities.

Figure 9 shows the total error from the five driving forces considered, together with the actual discrepancy between simulated and measured results over the same period. Back wall temperature has been chosen to demonstrate this analysis because it has been seen previously to contain significant contributions from a number of driving forces. The figure indicates that a large proportion of the error observed has been accounted for by the simplified error model. The remaining, unexplained, portion of the simulation error may be due to unmodelled driving forces which have not been measured but which

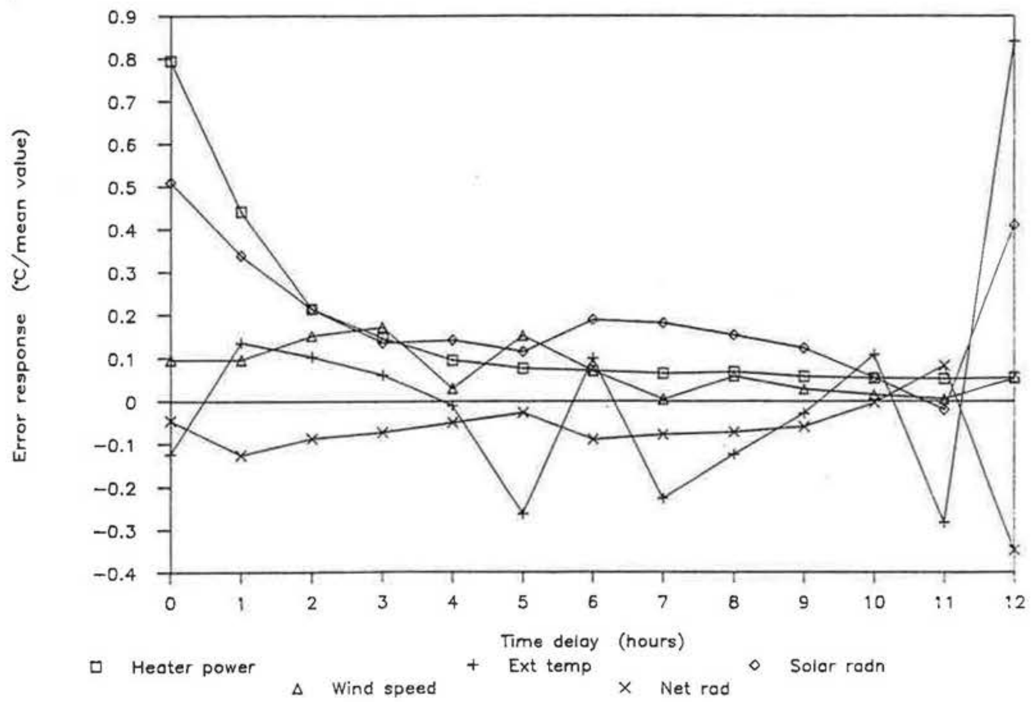

Fig. 7. Response of back wall temperature prediction error to driving forces.

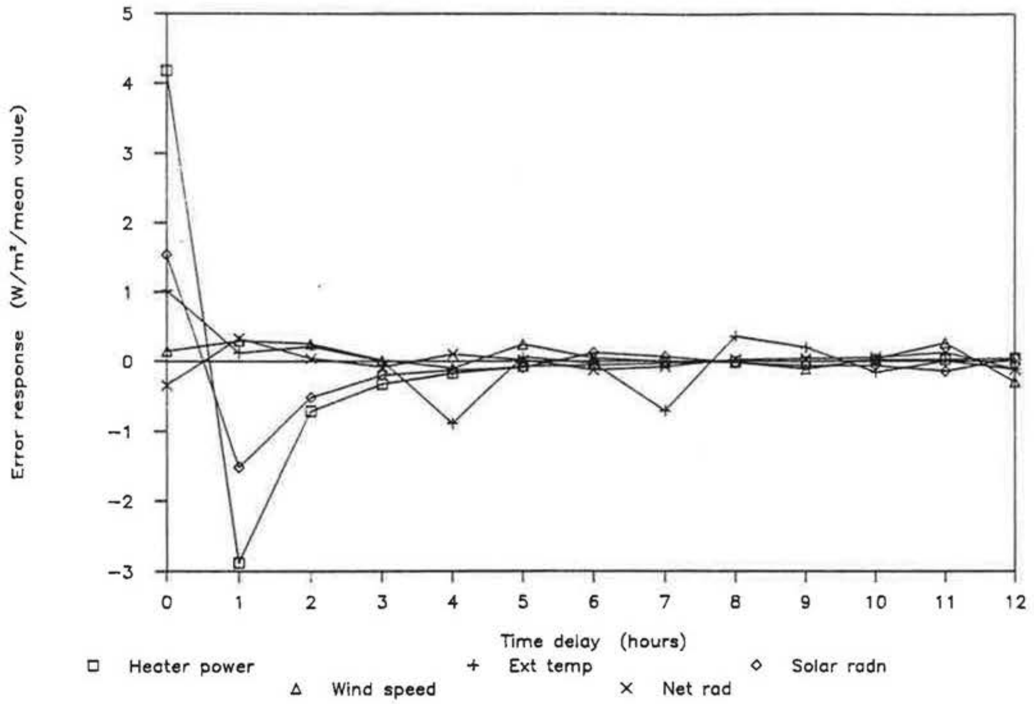

Fig. 8. Response of back wall heat flux prediction error to driving forces.

affect the performance of the building, or to the break down of the time invariance and linearity assumptions upon which this analysis depends. Inevitably, there will be mechanisms present in the test rooms which cause both assumptions to be violated to some extent. However, the close agreement between the linearly reconstructed and observed simulation errors implies that effects from these sources are relatively small.

Figure 10 shows the relative contributions (over the whole 50 day simulation period) of each driving force to the variance of the simulation error, providing a picture of how the errors in predicting back wall temperature are built up, and demonstrating that the dominant sources of error is, in this case, solar radiation. This conclusion is not general to the remainder of the quantities predicted and measured in the test room. Back wall temperature

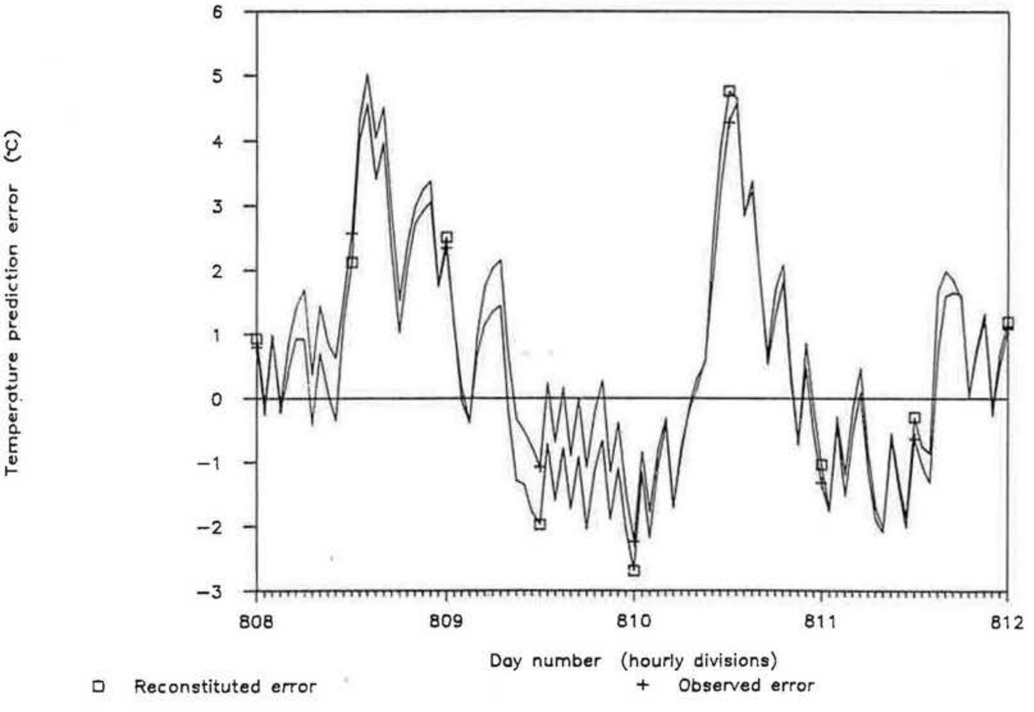

Fig. 9. Reconstituted and observed errors in test room back wall temperature prediction.

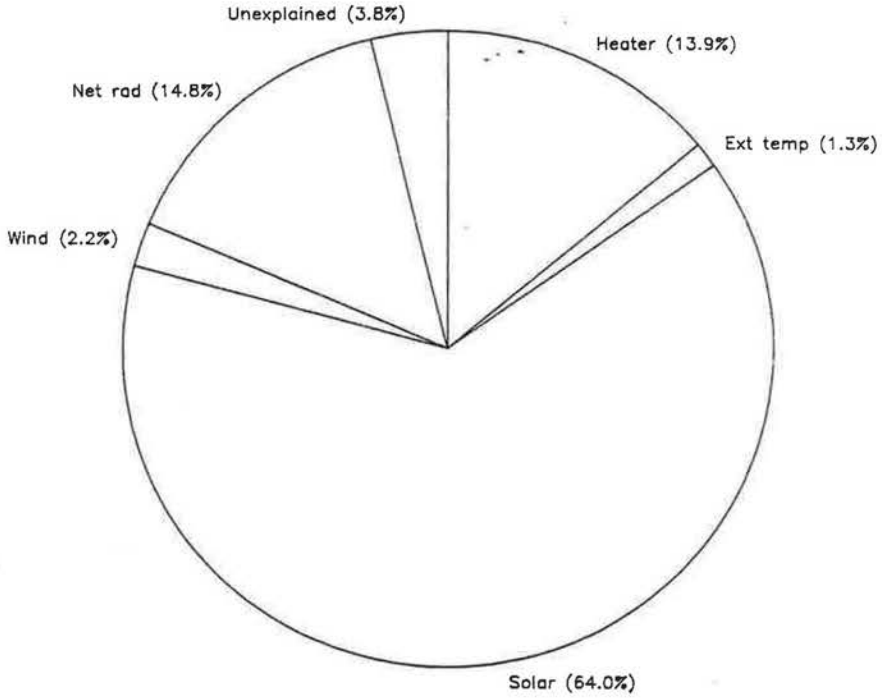

Fig. 10. Relative contributions of driving forces to the reconstituted prediction errors.

was deliberately chosen as the quantity with which to demonstrate the error reconstitution technique because its prediction error contained components from several sources. As the impulse responses presented earlier indicated, the dominant source of error in the simulation is auxiliary energy input. For some of the quantities predicted, for example zone temperature, it is the major contributor to the discrepancies between simulation and reality.

4. CONCLUSIONS

It has been demonstrated that data from test cells can form the starting point for detailed empirical validation work on computer simulation models. It has further been shown that, with the appropriate operation of both test rooms and simulation model, the sources of discrepancies between predictions and measured data can be identified, providing useful information to model users, and in the longer term, to the developers of future models.

The analysis presented implies that, for many of the quantities predicted by SERI-RES, the most significant source of error is operation of the auxiliary heat source. Hence the model should be able to predict the performance of an unheated room well. As a test of this conclusion, and thus of the deconvolution analysis described, further data sets have now been collected in which the test rooms are unheated. Subsequent comparison with the predictions of the model demonstrated a high level of agreement, confirming the conclusion of the analysis described in this paper, and providing a first indication that the technique satisfactorily identifies the source of discrepancies between model and data. Further work is now pursuing three lines of investigation:

—analytical tests which will allow verification of the validation techniques presented above are being developed,
—a combination of further analysis and further experimental work has been used to determine why significant errors appear when SERI-RES is used to handle heated buildings, and finally,
—the impact of the errors detected in test rooms on simulation results for more realistic buildings is being assessed analytically.

Acknowledgements—The work described in this paper was partly funded by the U.K. Department of Energy, and managed by the Energy Technology Support Unit at Harwell. The Energy Monitoring Company would like to thank both organizations for their support.

REFERENCES

1. K. J. Lomas and N. T. Bowman, An investigation into Analytical and Empirical Validation Techniques for Dynamic Thermal Models of Buildings. A SERC/BRE Collaborative Project. Volume 4. (September 1987).
2. Description of the EMC Test Room Facility 1987–88. Energy Monitoring Company Report to ETSU, Harwell Laboratories, OXON. S1197-P4. October 1986.

3. Detailed Model Comparisons: An empirical validation exercise using SERI-RES. Energy Monitoring Company Report to ETSU, Harwell Laboratories, OXON. S1197-I. June 1991.
4. An experimental investigation of the influence of heater type on passive solar building performance. Energy Monitoring Company Report to ETSU, Harwell Laboratories, OXON. S1197-P2. September 1988.
5. C. Martin and D. M. J. Watson, Further experiments in a highly instrumented test room—1990. Energy Monitoring Company report to British Gas plc (February 1990).
6. K. M. Letherman, C. J. Palin and P. M. Park, The measurement of Dynamic Thermal Response in Rooms using Pseudo-random Binary Sequences. *Bldg and Environ.* **17**, 11–16 (1982).
7. C. Martin and D. M. J. Watson, Stochastic heating trial in the British Gas test room—1990. Energy Monitoring Company report to British Gas plc (March 1989).
8. Subgroup Test Methodologies Final Report. 1986–89. Edited by Conphoebus and TNO. 060-89-PASSYS-TME-FP-026. PASSYS Project Phase I. CEC DG XII.
9. K. J. Lomas, Simulations using SERI-RES and HTB2 for comparison with measured data from the EMC Passive Solar Test Rooms. Leicester Polytechnic (1988).
10. P. Haves, Modelling different forms of heat input using SERI-RES. Polytechnic of Central London (April 1985).
11. H. Eppel, F. Parand, A. Pinney and K. Lomas, Developing and Proving Sensitivity Analysis Techniques. ETSU Applicability Study Research Report 13. Leicester Polytechnic report to ETSU, Harwell Laboratories, OXON (February 1990).

APPENDIX
DESCRIPTION OF THE CROSS-CORRELATION/DECONVOLUTION TECHNIQUE

Any linear, time invariant system can be completely described in terms of its impulse response. This is the output of the system when its input is an impulse, which is unity at time zero and zero at all other times. It is a requirement of any physical system that its impulse response is zero before time zero. If not, the system would begin to respond to inputs before they were actually applied—it would be non-causal.

The ability of the impulse response to characterise a system completely stems from the fact that an arbitrary input sequence can be broken down into a series of scaled and time-shifted impulses. Because the system is postulated to be linear, the output due to a scaled impulse is simply the impulse response suitably scaled. If the system is also time invariant, the response to an input shifted in time is simply the response shifted in time. Again because the system is linear the total output is simply the sum of the outputs due to each part of the decomposed input sequence.

This process of building up the system output in response to an arbitrary input is written formally as the discrete time convolution sum. For the univariate case where the experiment begins at time zero, and all inputs to the system are zero before this, it is:

$$y[t] = \sum_{k=0}^{\infty} h[k]x[t-k] \tag{1}$$

where:

$y[t]$ is the system output.
$x[t]$ is the system input. and
$h[t]$ is the system impulse response.

Equation (1) allows the response of a system with a known impulse response to be generated for any arbitrary input. In approaching the problem of identifying the sources of simulation errors the output of the system (in this case the simulation error) is known, as is the input (the measured driving force). The impulse response connecting the two data sequences is the required quantity, and the convolution equation given above must be inverted to obtain it, a process known as deconvolution. Equation (1) can be inverted directly to obtain a recursive solution for the impulse response $h[t]$:

$$h[0] = \frac{y[0]}{x[0]} \quad \text{followed by} \quad h[t] = \frac{y[t] - \sum_{k=0}^{t-1} h[k]x[t-k]}{x[0]}. \tag{2}$$

In practice, however, equation (2) does not represent a good way of obtaining an impulse response from input and output data sequences. Simple examination of the equations shows that the value of $h[0]$ is based on only one pair of data points. Any errors in $h[0]$, caused for example, by noise in those readings, will then be propagated through to the estimate of $h[1]$. Uncertainties in $h[0]$ and $h[1]$ are then carried forward into the estimate of $h[2]$ and so on. The process thus has very poor performance when faced with any noise in the measured data.

A second problem arises with the state of the system at time zero. Implicit in the formulation of the response equation was the assumption that there are no inputs to the system before $t = 0$, that is the system is settled in quiescent state when the inputs begin. In any real experiment there are likely to be driving forces (for example climate) which cannot be switched off up to the point at which the experiment begins. Thus at the start of the test the system retains some memory of the disturbances which occurred earlier. This effect is most pronounced in the measurements taken at the start of the experiment, the very values which are to be used to initiate the solution process by deducing $h[0]$.

The solution to both of these problems lies with the cross-covariance function. The cross-covariance between two sequences of numerical data, $\{x[t]\}$ and $\{y[t]\}$, is defined by:

$$r_{xy}[k] = E((x[j]-\bar{x})(y[j+k]-\bar{y})) \tag{3}$$

where:

E denotes the expectation (or averaging) operation over the dummy variable j,
$\bar{x}$ denotes the mean of the sequence $\{x[t]\}$, and
$\bar{y}$ denotes the mean of the sequence $\{y[t]\}$, that is:

$$\bar{x} = E(x[j]) \quad \bar{y} = E(y[j]).$$

The auto-covariance function of a sequence is simply the cross-covariance of the sequence with itself.

If equation (3) is substituted into equation (1), a little algebraic manipulation yields:

$$r_{yx}[t] = \sum_{k=0}^{\infty} h[k]r_{xx}[t-k]. \tag{4}$$

This rather remarkable result reveals that the cross-covariance between the input and output of a system is given by the convolution of the input signal auto-covariance with the system impulse response. The relation given is the discrete time form of the Weiner–Hopf equation.

The Weiner–Hopf equation can be inverted as before, to yield the impulse response of a system from the appropriate auto-covariance and cross-covariance functions. In this way, both of the problems outlined earlier are avoided. The terms in the covariance functions are the results of averaging over the whole data set, and thus the effect of noise in the data is greatly reduced. If the experiment is significantly longer than the time constants of the system under investigation the influence of disturbances before the start of the experiment will also be minimal. The combination of cross-correlation and deconvolution thus provides a robust way of obtaining impulse responses from streams of input and output variables.

The problem posed in the main text requires that the relative contributions from a number of inputs, or driving forces, $x_1[t]$, $x_2[t]$, $x_3[t]$ etc. to a single output be determined. Expressed in terms of the impulse responses to each of those inputs, $h_1[t]$, $h_2[t]$, $h_3[t]$ etc., the analysis model is:

$$y[t] = \sum_{k=0}^{\infty} h_1[k]x_1[t-k] + \sum_{k=0}^{\infty} h_2[k]x_2[t-k] + \sum_{k=0}^{\infty} h_3[k]x_3[t-k] \ldots \quad (5)$$

For the purposes of solution, equation (5) can be cast in matrix form. Here, for the sake of brevity, we consider a case with two independent variables, the impulse response of each of which has only three non-zero elements. The extension to larger cases (for example the five independent variables with impulse responses of 13 elements each used in the main text) is straightforward. If the length of the input and output sequences is $n+1$ points, then equation (5) can be written:

$$\mathbf{y} = X\mathbf{h} \quad (6)$$

where:

$$\mathbf{y} = \begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[n] \end{bmatrix}$$

$$X = \begin{bmatrix} x_1[0] & 0 & 0 & x_2[0] & 0 & 0 \\ x_1[1] & x_1[0] & 0 & x_2[1] & x_2[0] & 0 \\ x_1[2] & x_1[1] & x_1[0] & x_2[2] & x_2[1] & x_2[0] \\ x_1[3] & x_1[2] & x_1[1] & x_2[3] & x_2[2] & x_2[1] \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1[n] & x_1[n-1] & x_1[n-2] & x_2[n] & x_2[n-1] & x_2[n-2] \end{bmatrix}$$

and:

$$\mathbf{h} = \begin{bmatrix} h_1[0] \\ h_1[1] \\ h_1[2] \\ h_2[0] \\ h_2[1] \\ h_2[2] \end{bmatrix}.$$

Equation (6) can be solved for $\mathbf{h}$, the required impulse response terms, using the pseudo-inverse of the non-square matrix X:

$$\mathbf{h} = (X^T X)^{-1} X^T \mathbf{y}.$$

The elements of the matrices $X^T X$ and $X^T\mathbf{y}$ correspond to the cross-correlations and auto-correlations used previously in the solution of the univariate case. Thus this method for solving the multivariate case demonstrates the same robustness described for its much simpler counterpart. The quality of the solution obtained can be established by using the measured input sequences and the derived impulse responses in equation (5) to attempt to 'reconstitute' the observed output (in this case simulation error). Close agreement between the reconstituted output and that observed indicates that the postulated model is representing physical reality well.

Clearly the analysis required for the multivariate case is much more complex than that for the univariate case, because of the possibility of cross-correlations between the various input sequences. When these inter-correlations do not exist, or can be ignored, the multivariate solution can be replaced by a univariate solution for each impulse response in turn. In this case the reconstitution of the errors using equation (5) provides a test of how well the model derived using the simplified analysis accounts for the errors observed. This is the simplified analysis used in the earlier work referred to in the main text. The results presented in this paper, however, used the full solution method described above.