Empirical Whole Model Validation Case Study: the PASSYS Reference Wall

Søren Østergaard Jensen
Thermal Insulation Laboratory
Technical University of Denmark

The methodology for validation of Building Energy Simulation Programs developed within the CEC concerted action PASSYS by the Model Validation and Development subgroup was presented at the IBPSA conference "Building Simulation '91" (Jensen and van de Perre 1991, Palomo et al 1991). Part of this validation methodology is a methodology for Empirical Whole Model Validation. Several case studies using this methodology have now been carried out. One of these case studies implies the investigation of a data set obtained in a PASSYS test cell with the PASSYS Reference Wall. The paper describes this case study in order to demonstrate the PASSYS methodology for empirical whole model validation - especially the part of the methodology dealing with comparison of measurements and predictions.

Introduction

The aim of performing empirical whole model validation is to detect if a model is capable of describing the reality correctly. It is a non-trivial task to perform as it requires expertise in experimental design, modelling principles and simulation techniques. In many earlier validation studies comparisons have been performed for relatively few physical parameters (eg only the total heat demand), making it difficult to identify the cause of the observed discrepancies between measured and predicted values. In some cases discrepancies have been reduced by adjusting one or more of the determining parameters without knowing the real cause of the deviations, as the number of accurately measured parameters was insufficient. In such cases it cannot be claimed that the model has been proven valid. The comparison between measurements and predictions has furthermore often been performed only by comparing curves. This method is highly subjective and gives only little information on what may cause deviations between measurements and predictions.

An empirical validation study should be carefully planned before starting. The experiment should be designed so that the necessary data is available in terms of eg scanning interval, location and accuracy of sensors. Guidance in this area can be obtained by modelling the phenomena before doing the experiments. Also, the expected analysis procedure of the data should be planned before the study. Statistical techniques should be applied in the comparison of measurements and predictions in order to increase the quality of and confidence in the validation result and to maximize the information about the capability of the model to describe the reality.

The PASSYS Methodology

There is a need for an empirical whole model validation methodology. Such a methodology has been developed by the Model Validation and Development subgroup; a methodology which ensures that the information on program performance and the cause of discrepancies is maximized. The methodology comprises six stages:

1) definition of scope, type and nature of the physical and numerical experiment,
2) implementation of the physical experiment on site,
3) processing of the measured data,
4) performance of simulations,
5) analysis of the results and assessment of the sensitivity,
6) documentation of the data set and validation work.

The empirical whole model validation methodology is shown graphically in figure 1 - more details can be found in (Jensen and van de Perre 1991, Palomo et al 1991, Jensen 1993a and b). It should be remembered that empirical whole model validation is only one among several approaches in a validation methodology. Empirical whole model validation cannot stand alone.

A validation study should further always be carried out by at least two teams: a leading team performing the exercise and a reviewing team. In this way the performing team can receive input and criticism and the calculations are impartially checked. This should ensure the quality of the validation study.

Thermal Insulation Laboratory, Building 118,
Technical University of Denmark,
DK-2800 Lyngby, Denmark
Phone: + 45 45 93 44 77, Fax: + 45 45 93 17 55,
E-mail: kim @ lfv.unic.dk

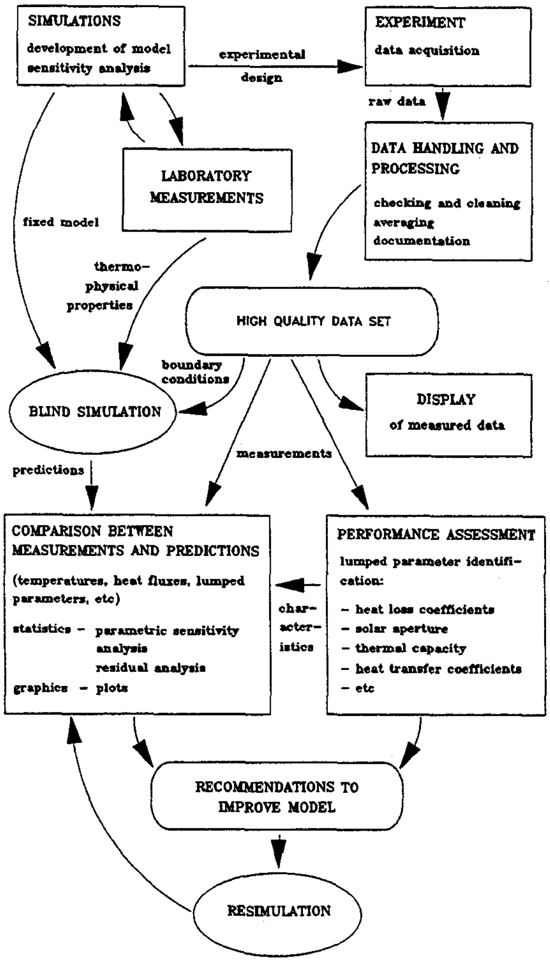

Figure 1. Outline of the PASSYS empirical whole model validation methodology.

The PASSYS Reference Wall

The PASSYS project has now come to an end and several case studies using the above-mentioned empirical whole model validation methodology have been carried out (Jensen 1993b). One of these case studies was performed on a data set obtained in a PASSYS test cell with the PASSYS Reference wall. The Reference wall has been tested at all 11 PASSYS test sites. The investigation of one of the obtained data sets with the Reference wall is in the following used to illustrate the PASSYS empirical whole model validation methodology.

The PASSYS Reference wall is a 2.75 x 2.75 m^2 lightweight wall consisting of 100 mm insulation foam with 13 mm plywood on each side. A double-glazed window with an area of 1.15 m^2 is centrally located in the wall. The wall is made of materials with well known thermo-physical properties which are typical for lightweight insulated walls.

The design of the Reference wall, the instrumentation of the wall and the test cell and the used test sequence were based on recommendations from the Model Validation and Development subgroup. The set-up of the experiment with the Reference wall was guided by simulations using a model of the test cell with the Reference wall. The simulation program applied in the validation exercise was ESP version 6.30a (Clarke 1985). By performing sensitivity studies with the simulation model important information were obtained on which parameters was critical and should be measured with extra care. It was eg shown that even a rather low infiltration rate had an important influence on the performance of the test cell, so it was recommended that the test cell should be sealed very carefully.

The simulation model developed for guiding the set-up of the experiment was after the experiment used for the comparison of measurements and predictions - the validation exercise was, therefore, blind.

The data set investigated in the present paper was obtained at the German PASSYS test site in Stuttgart during the period August 9-October 6, 1991. The validation exercise was performed by the Thermal Insulation Laboratory, Technical University of Denmark and reviewed by Eenheid Zone, Vrije Universiteit Brussel, Belgium.

The data set was chosen for release as a PASSYS one-minute scan high quality data set. The data set has as such been included in a comprehensive cleaning procedure where outliers, smaller gaps, etc have been corrected using advanced statistical tools.

Due to a power failure of 6 hours at the test site, the data set had to be treated as two separate periods - see figs. 2-3. The test strategy used was the common strategy developed within PASSYS: One week start-up period at constant test room temperature, 2 weeks of constant minimum power input to the test room, 2 weeks of high constant power input, 2 weeks of moderate constant power input and 2 weeks with a pseudo random binary power input - a so-called ROLBS-signal (Jensen 1993b).

Comparison of measurements and predictions

In the following the PASSYS methodology for comparing measurements and predictions will be illustrated by describing the major findings from the different steps of the methodology: From graphical comparison of measurements and predictions over assessment of uncertainties to residual analysis comprising graphical display, investigation of auto-correlation functions, density power spectra and cross-correlation functions.

The simulation model of the test cell with the Reference wall from the set-up of the experiment was used in the comparison of measurements and predictions. Smaller modifications were, however, applied to the model in order to adapt local conditions - eg location, site exposure and obstructions.

Using the measured climate data, service room temperature and heat input from the heating systems as input, the predicted test room temperature has been compared with the measured test room temperature. The comparison between the 'blind' model and the measurements is shown in fig. 2. The residuals (measured minus predicted values) are also shown. A fairly large trend (systematical error) between measurements and predictions is observed, while the dynamic response of the test cell seems to be well represented by the model.

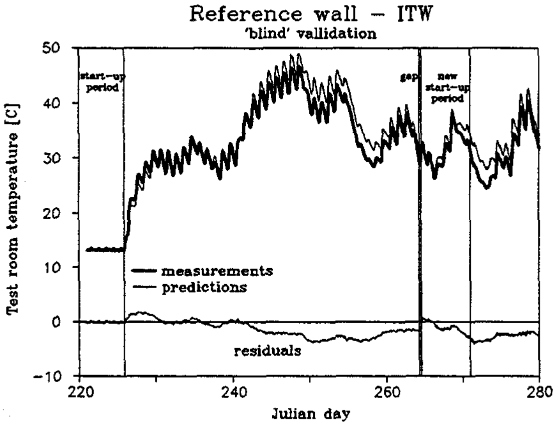

Figure 2. Results from the 'blind' validation.

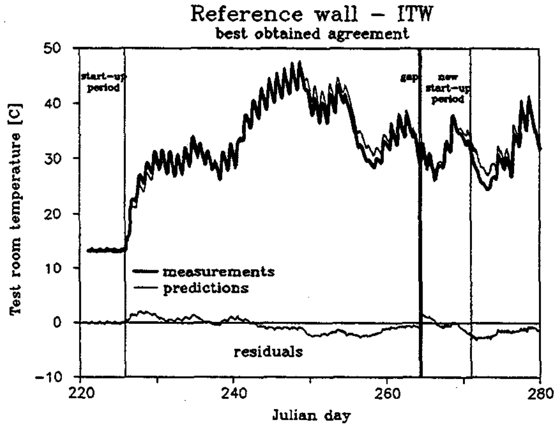

Figure 3. The result from the simulation with increased conductivity of the insulation.

Based on the findings from parameter identifications on data sets from experiments with the so-called Calibration wall (400 mm insulation foam), the capacity of the test cell (not the south wall) was adjusted. The capacity of the edge constructions of the 'blind' model was found to be too high. Further it was found, that the internal convection was higher than calculated by known formulas. It is, therefore, justifiable to decrease the capacity of the test cell and to use measured convection coefficients for all opaque surfaces. This reduced the mean value of the residuals from -1.44 to -1.30°C for the first period and from -3.56 to -2.69°C for the second period. The variance of the residuals was reduced from 5.35 to 2.35°C^2 and from 2.33 to 0.40°C^2 for the two periods.

The variance was thus decreased very much but there is still a trend - although a bit smaller. The trend is partly caused by the absence of temperature dependent conductivities and maybe by an incorrect representation of the edge constructions of the test cell. Figure 3 shows the comparison between measurements and predictions, where the conductivity of the insulation materials of the test cell is increased by 7.4%. The reason for this is, that the mean temperature of the insulation materials during the experiment was approximately 25°C, while the conductivity of the insulation was measured at 10°C. Measurements have shown that the conductivity increases by about 5% for each 10°C increase of the temperature level. Figure 3 shows a much smaller trend. The mean value of the residuals is now -0.49 and -1.81°C and the variance is 1.54 and 0.47°C^2 for the two periods.

Differential Sensitivity Analysis

A differential sensitivity analysis was then carried out. In this technique perturbed runs, where each model input parameter is changed by ± their standard deviation, are performed. Based on the perturbed runs the overall uncertainty band of the simulation was calculated. The analysis was, further, performed in order to give an idea of which parameters had major influence on the simulation and to analyse if the observed discrepancies could be explained by the uncertainty of the input parameters to the simulation.

The most important input parameters to the simulation (causing a change of more than 0.2°C of the internal air temperature) were found to be:

- the ground reflection
- the conductivity - especially the conductivity of the edge constructions
- the ambient temperature
- the wind speed
- the global radiation on horizontal
- the view factor to the surroundings

The overall result from the differential sensitivity analysis is shown in fig. 4. The residuals with the corresponding error bands are shown. The error bands are normalized: High case minus predictions and low case minus predictions. The error bands on the measured temperatures are not included. These error bands are estimated to be ±0.2°C, but may be larger due to stratification of the air in the test room - see later.

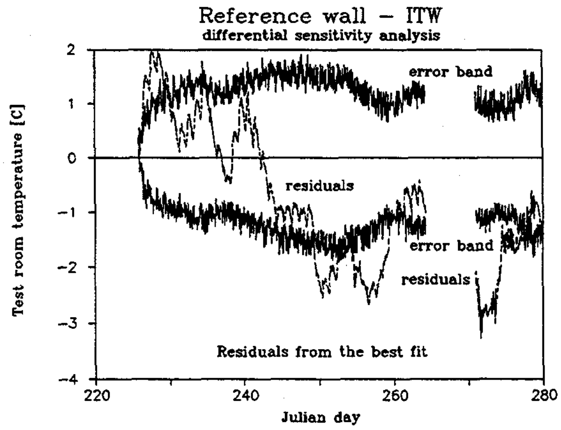

Figure 4. The residuals from fig. 3 and the error bands from the differential sensitivity analysis.

From fig. 4 it is seen that the experiment is rather well determined - the error bands lie within ±1.9°C. One third of the residuals falls outside the error band. The residuals are, although, within two times the error band.

The uncertainty of some of the parameters is, however, rather optimistic - especially for the climate parameters. The uncertainty of the measured solar radiation is the one given by the manufacturer. This uncertainty is often reported to be higher. An increase of the uncertainty is further introduced when using the measured solar radiation for calculating the solar incidence on inclined surfaces. The wind speed is measured at 10 m height from which the local wind speed along the surfaces is calculated - this procedure introduces large uncertainties. Measurements (Jensen 1989) have shown that the ambient temperature at the south side of the test cell may be several degrees higher than the ambient temperature at the north side during day-time. The heat input to the test cell from the heating system was, during the experiment, very small ie the main input came from the fan of the heating system. The uncertainty of the set back of the window in the wall was not included in the above-described differential sensitivity.

A new differential sensitivity analysis was carried out where the uncertainty of the above-mentioned parameters was increased within reasonable limits.

In the first differential sensitivity analysis the error bands lay within ±1.9°C. With the increased uncertainties the error bands lie within ±2.4°C, making more of the residuals from fig. 4 fall within the error bands.

The differential sensitivity analyses show that the main part of the observed trend of the residuals may be explained by the uncertainty of the input parameters to the simulation model and the absence of temperature dependent conductivities. This, however, only tells that ESP by an appropriate set of input parameters is capable of reproducing the behaviour of the test cell, rather than that ESP can predict the response of the test cell. The next step in the PASSYS model validation methodology is, therefore, to use more advanced statistical tools to investigate the residuals.

In this analysis the residuals from the simulation with decreased capacity of the test cell and measured convection coefficients were used, as the residuals in fig. 3 were obtained by increasing the conductivity of the insulation with a fixed value. This is not correct. The conductivity should be different for the different layers and should further vary over time as a function of the temperature of the insulation material.

Residual Analysis

First the residuals were plotted together with the input to the test cell. Some of these plots are shown in figs. 5-6 - the two start-up periods are not shown. A strong correlation between the residuals and the service room temperature and the residuals and the heat input is seen. As, most of the time, the service room temperature equaled the test room temperature, this correlation is really not correct. It hides another correlation, that the heat loss by conduction increases with the temperature level in the test room, which is also supported by the observed correlation between the residuals and the ambient temperature. The residuals further expose daily fluctuations, which seem to be caused by the climate. More dynamic fluctuations are observed in the second period. It is highly probable that this is caused by the high frequency heat input from the heating system - see fig 6.

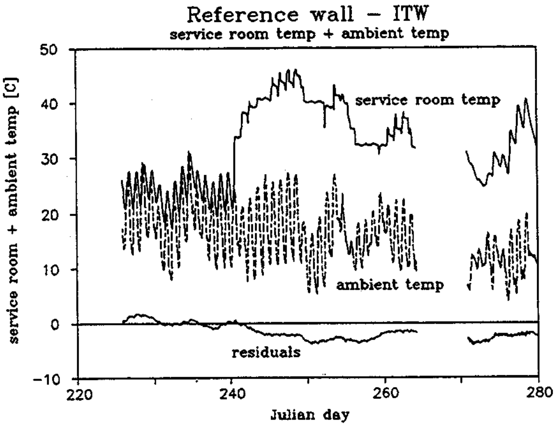

Figure 5. Service room and ambient temperature.

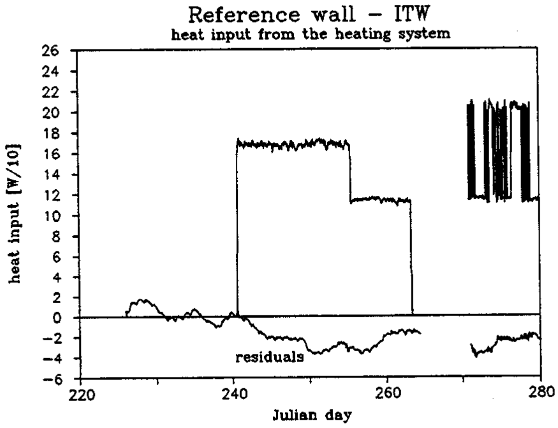

Figure 6. Heat input from the heating system.

Auto-correlation:

The auto-correlation function for both the residuals and the differentiated residuals was then analysed in order to determine if the trend could be removed by differentiation. It is necessary to remove the steady state trend as it will hide important information in the dynamic area. The analysis showed that the residuals only had to be differentiated once in order to remove the steady state trend. In the following, both non-differentiated and differentiated residuals are therefore investigated.

Density Power Spectrum:

Figure 7 shows the density power spectrum for the two periods. A more careful investigation of the data set has revealed (Jensen 1993b), that in order to be in good agreement, the curves for the non-differentiated residuals should lie below the thick lines in fig. 7.

It is seen that for the first period the model very well represents the dynamic response of the test cell while steady state problems occur. However, for the second period there are problems at frequencies between 0.3-0.45 (corresponding to 2.2-3.3 hours). These problems are thought to be caused by the heating system - the high frequency heat input during the second period (see fig. 6); however, as the time constant of the heating system is far smaller than 2-3 hours, the only reason for the problems seems to be, that the test room was heated through textile hoses right under the ceiling (the hoses were not inclined as later decided ie going from ceiling to floor). The heating system, therefore, tends to heat the test room by filling it from the top with warm air. This stratification of the air during heat input may also explain part of the trend observed in figs. 2-3. This means that the observed residuals may partly be caused by a wrongly measured mean temperature of the test room.

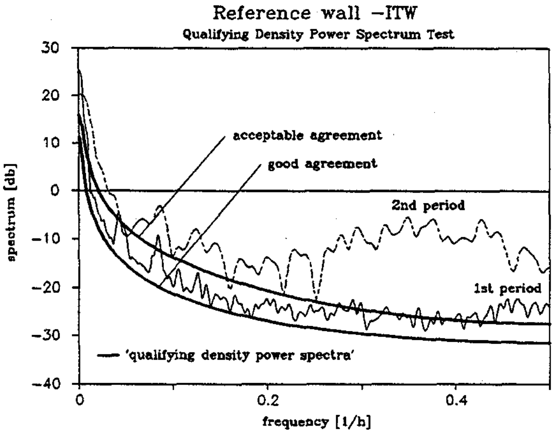

Figure 7. The density power spectrum for the non-differentiated residuals for both periods.

The thick lines shown in fig. 7 is a new way to compare measurements and predictions and is called Qualifying Density Power Spectrum Test (Jensen 1993b). The thick lines in fig. 8 were obtained before the here described analysis of the data set was performed. The Qualifying Density Power Spectrum Test is, therefore, blind.

Cross-correlation:

The next step in the analysis of the residuals is to compute the cross-correlation functions between the residuals and the input parameters. Figures 8-11 show the cross-correlations between the non-differentiated residuals and the input parameters for both periods. The horizontal lines are the ±95% confidence interval for the non-zero correlation hypothesis. The parameters are not cross-correlated, if the curve stays between these two lines.

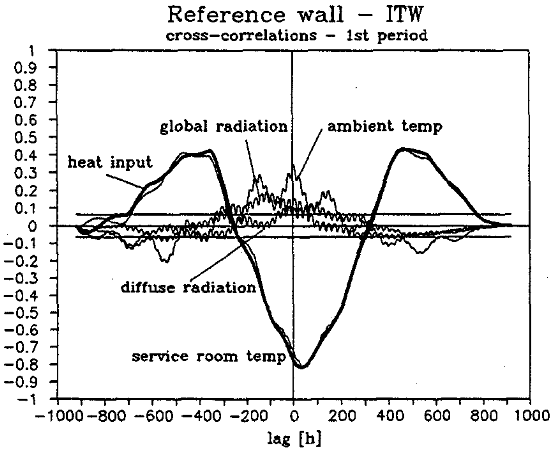

Figure 8. Cross-correlations between the non-differentiated residuals and some input parameters for the first period.

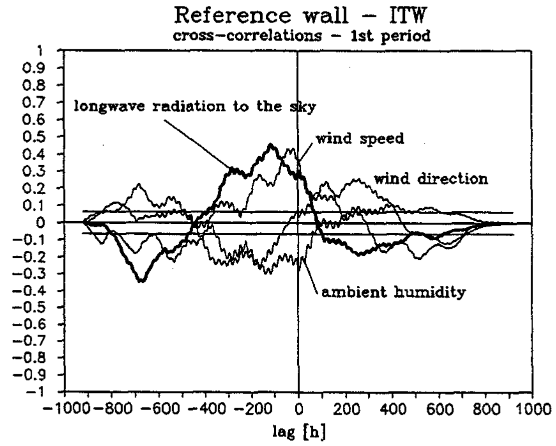

Figure 9. Cross-correlations between the non-differentiated residuals and some input parameters for the first period.

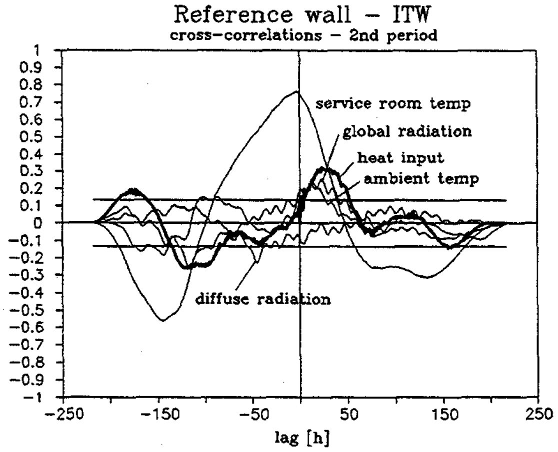

Figure 10. Cross-correlations between the non-differentiated residuals and some input parameters for the second period.

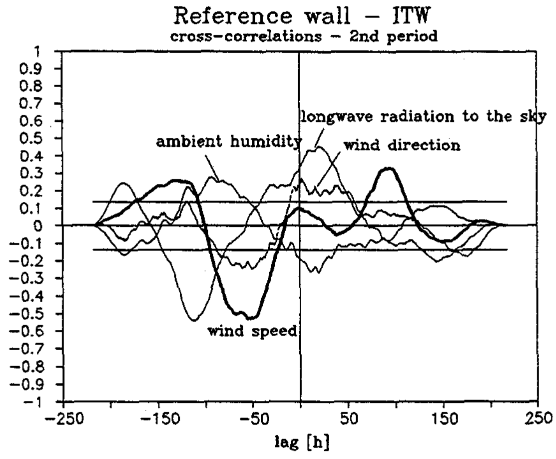

Figure 11. Cross-correlations between the non-differentiated residuals and some input parameters for the second period.

Figure 8 shows a strong cross-correlation between the non-differentiated residuals and the service room temperature and the heat input. However, for the second period (fig. 10) only a minor cross-correlation with the heat input is observed. This supports the hypothesis, that it is not the service room temperature or the heat input which are responsible for the large trend, but the temperature level, which again points to the absence of temperature dependent conductivities, problems with the edge constructions and to a possibly wrongly measured air temperature. The figures show that the residuals are cross-correlated with all the investigated input parameters. The ambient temperature, solar radiation, humidity and wind direction are however of minor importance, while the wind direction and the external longwave radiation to the sky have a stronger cross-correlation. From earlier experience (Jensen, 1989) it is known that the test cell and especially glazed surfaces are sensitive to the external heat transfer - convection and radiation. The calculation of the external convection and the longwave radiation exchange during night-time is currently not determined well enough.

Figures 12-15 show the most important cross-correlations between the differentiated residuals and the input parameters. For the first period only the ambient temperature, the global solar radiation and the heat input are cross-correlated with the residuals and, therefore, shown here (the diffuse radiation shows a similar pattern as the global solar radiation although less cross-correlation than fig. 13, but is strongly cross-correlated with the global solar radiation and is, therefore, not shown here). The heat input is, however, of minor importance for the first period - only a small peak in lag 0 is observed. For the second period only the heat input is cross-correlated with the differentiated residuals. These figures show, that it is the solar radiation and also the ambient temperature, which are responsible for the daily fluctuations, while it is the heat input which creates the very fast fluctuations in the second period. There is, in the documentation of the data set, no information on how clean the window was. Dirt on the window will only reduce the mean value of the residuals by up to 0.2°C but has a larger influence on the spot residuals during day-time. As the walls of the test cell are heavily insulated the cross-correlation between the differentiated residuals and the ambient temperature is believed mainly to be caused by the heat transfer through the window eg a wrong value of the thermal resistance of the window.

Conclusion

The aim of the here described exercise was to test and show the use of the whole model validation methodology developed within PASSYS.

It is hoped that the present paper has shown how powerful a tool the methodology is when used correctly. The methodology offers more objective tools than normally used to investigate the reasons for the observed discrepancies between measurements and predictions.

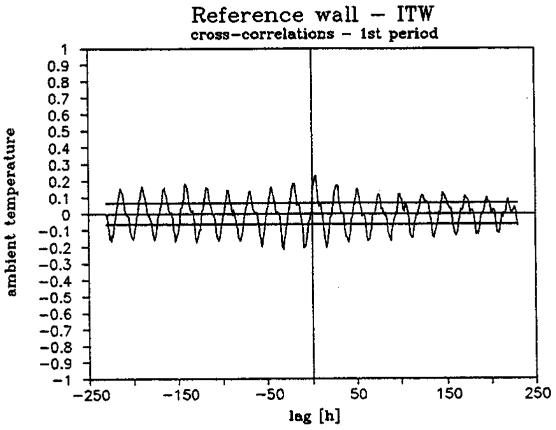

Figure 12. Cross-correlations between the differentiated residuals and the ambient temperature for the first period.

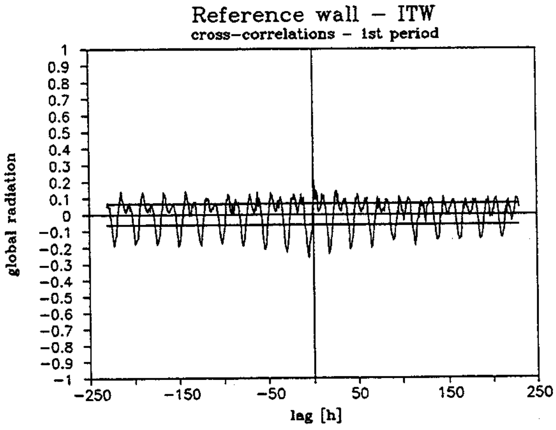

Figure 13. Cross-correlations between the differentiated residuals and the global horizontal radiation for the first period.

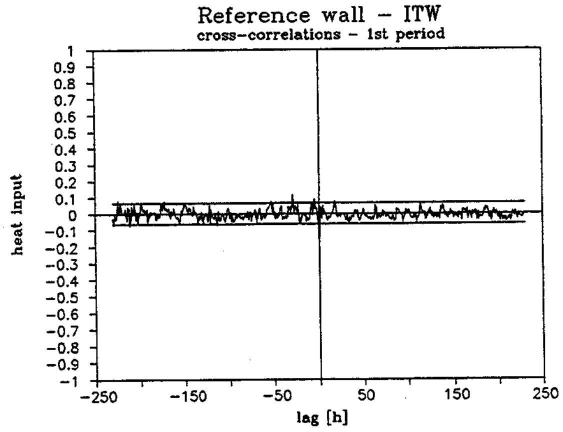

Figure 14. Cross-correlations between the differentiated residuals and the heat input for the first period.

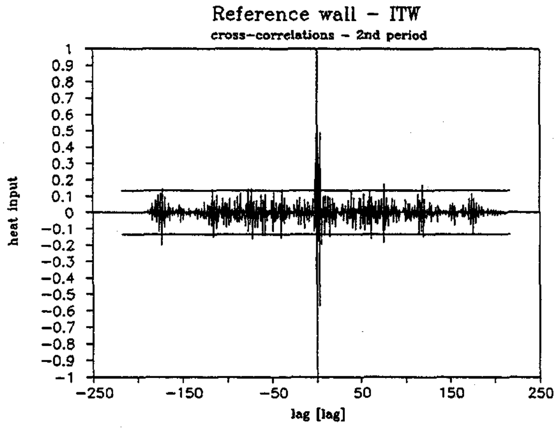

Figure 15. Cross-correlations between the differentiated residuals and the heat input for the second period.

In the present paper the reasons for the observed disagreement were found to be: For the steady state problems: The absence of temperature dependent conductivities, an incorrect representation of the edge constructions, a possible error in the way the measured test cell temperature is determined, the external heat exchange to the surroundings - especially the convection and the longwave radiation exchange - and the incoming solar radiation to the room. For the dynamic problems (which are small): The thermal resistance of the window, the incoming solar radiation to the room and a possible error in the measured test cell temperature.

Based on the validation study advice has been given on how to investigate the above-mentioned problems - these are described in (Jensen 1993b). The empirical whole model validation exercise has exposed several problems which, however, only can be solved by applying several of the other validation approaches - especially single process validation techniques. This again highlights that empirical whole model validation is only one of several validation techniques in a validation methodology.

Further validation activities on the described data set will be helped by the fact that the experiment was well planned - the presence of a large sensor set eg 7 air temperature sensors, 20 internal surface temperature sensors, a temperature sensor array in the reference wall, additional climate sensors (eg a pyrgeometer measuring the longwave radiation exchange to the sky), etc. The used test strategy has, furthermore, excited the test cell in such a way that problems have been exposed, which would have remained hidden if the test strategy only contained a period of constant temperature in the test room. However, the data set turned out to be of a poorer quality than expected (however, still of a higher quality than normally seen) - the wind speed was, for part of the time, not measured at the site, the ambient surface temperatures show large disagreement, not enough measurements on local climate conditions, etc. Several of the reasons for the observed residuals may, therefore, remain hidden.

The test cell is a very special construction - low heat fluxes through the envelope (an overall UA-value including the Reference wall of about 12 W/K), a large external surface/volume ratio and nearly no infiltration. It is, therefore, very likely that several of the problems, observed in the validation study, are of minor importance in real, less heavily insulated, buildings - eg the problem of modelling the edge constructions and the absence of temperature dependent conductivities, while other phenomena as user behaviour and infiltration often play a more important role - also in well insulated low energy buildings. The recommendations for modifications to the model or the program should thus also be based on common sense - on which processes are of major importance.

A new method for comparing measurements and predictions, and for comparing the performance of different models and simulation programs - the Qualifying Density Power Spectrum Test - has been introduced. The method offers the opportunity not only to investigate the absolute difference between the measurements and the predictions, but also to pin-point at which frequencies the problems occur - if they are steady state problems or dynamic problems, it will make the search for the causes of the problems much easier. It is further possible to classify simulation programs by the area in which they perform well: In the steady state area, the dynamic area or both. In this way it will be easier for the user to choose the right program for the actual problem.

References

Clarke, J.A. 1985. *Energy Simulation in Building Design*. Adam Hilger Ltd, Bristol and Boston.

Jensen, S.Ø. (ed). 1989. *The PASSYS Project Phase 1 - Subgroup Model Validation and Development, Final Report - 1986-1989*. Commission of the European Communities, DGXII.

Jensen, S.Ø. and R. Van de Perre 1991. "Tools for Whole Model Validation of Building Simulation Programs - Experience from the CEC Concerted Action PASSYS". In *Proceedings of Building Simulation '91* (Sophia-Antipolis, Nice, France, August 20.22). IBPSA, pp. 574-553.

Jensen, S.Ø, (ed) 1993a. *Validation of Building Energy Simulation Programs - a Methodology - Part I. Research report from the PASSYS Subgroup Model Validation and Development*. Commission of the European Communities, DGXII. (will be published in 1993).

Jensen, S.Ø, (ed) 1993b. *Validation of Building Energy Simulation Programs - a Methodology - Part II. Research report from the PASSYS Subgroup Model Validation and Development*. Commission of the European Communities, DGXII. (will be published in 1993).

Palomo, E., Marco, J. and Madsen, H. 1991. Methods to Compare Measurements and Simulations. In *Proceedings of Building Simulation '91* (Sophia-Antipolis, Nice, France, August 20.22). IBPSA, pp. 574-553.