A Confirmation Technique for Thermal Performance Simulation Models

T J Williamson
Department of Architecture
University of Adelaide
AUSTRALIA

ABSTRACT

This paper deals with the problem of empirical validation of thermal performance computer programs. It begins with a brief review of a number of techniques which have been used as a measure of the goodness-of-fit between measured and predicted data in a variety of empirical validation exercises. Several inadequacies inherent in existing techniques are identified as,

a) no attempt is made to take into account the severity of the validation test.
b) none give a single measure of the success (or otherwise) of the test.
c) isolation of sources of error are difficult.
d) tests cannot be used easily for internal validation and/or algoriffim "tuning".

An argument is presented that an objective technique for establishing the accuracy of simulation predictions and which addresses these inadequacies is required. To satisfy this requirement a confirmation factor Cs based on an inequality coefficient is defined. Following from *Cs* it is shown that a degree of confirmation *D* may be evaluated.

Multiple input variables may be taken into account by calculating *Cs* and *D* from the estimates of the Principal Components derived from independent linear fúnctions of the original variables.

The confirmation technique is illustrated with an example comparing the predictions of the computer program TEMPAL with internal temperatures derived from monitoring a house. Possible sources of program error are investigated and the improvements made with a new program *EnCom2* are shown.

INTRODUCTION

Since the mid 1970's energy and thermal performance analysis computer programs have been employed for the quantitative evaluation of existing and proposed buildings. While the descriptive power of these programs continues to be enhanced the problem of confirming their adequacy for particular uses remains a contentious issue. Despite some impressive work in recent years aimed at developing empirical validation methodologies (Judkoff et al., 1983; Bloomfield, 1985; Eppel, Lomas, & Martin, 1993) and intermodal comparison techniques (Judkoff & Neymark, 1993) there continues to be an obvious need to develop consistent, logical and objective techniques for confirming the adequacy of these programs. Inadequacies which can be identified with techniques used todate are,

a) no attempt is made to take into account the severity of the validation test.
b) none give a single measure of the success (or otherwise) of the test.
c) isolation of sources of error are difficult.
d) tests cannot be used easily for internal validation and/or algorithm "tuning".

This paper outlines a technique which overcomes these inadequacies and which can be employed in empirical validation techniques for establishing the accuracy of simulation predictions. Of particular interest is the validation of the predictions of internal thermal conditions when heating or cooling plant are not operating ie. the free-running condition.

BACKGROUND

The purpose in conducting a thermal analysis of a building using a computer program is to predict some aspect of the building's thermal sub-system in order to provide information for decision making. Ideally we would wish any model used for reliable decision making to accurately represent reality. The very nature of models generally means however that many aspects of reality that in fact affect behaviour have been excluded because they have been deemed by the model builder to be unimportant. Empirical validation is seen as being concerned with examining the correspondence between reality (or at least a sub-system of reality) and the model predictions.

A variety of authors have provided definitions of empirical validation as it relates to building thermal analysis models. Some such definitions are,

- *'..comparing program predictions with the corresponding results from actual buildings...' (Clark & Forrest, 1978)*

- *'..the comparison of the predictions of the model with physical reality.' (Bowman & Lomas, 1985)*

- *'..testing the theoretical correctness of a calculation method and the numerical and mathematical procedures used to solve the resulting model.' (Bloomfield, 1985).*

Few authors have discussed directly the philosophical nature of the empirical validation problem applied to thermal analysis models.

VALIDATION VERSUS CONFIRMATION

The logical difficulties of inductive inference inherent in each of the definitions above point to an attitude adopted by most validationists which can be characterised as,

' measurements (m) are said to support a model's predictions (e) whenever m agree with e. Only if e are a counter-instance of m (ie. the measurements are nothing like the predictions), is the model likely to be rejected.'

Popper (1983) points out that adopting this type of uncritical methodology can lead to no other conclusion than statements like 'the model gives good predictions for the building being investigated' or 'predictions have shown good agreement with the monitored values'. Rather than validation Popper suggests a more critical attitude could be stated as;

'one looks for instances of falsification or refutation of a model. If one does not succeed we may speak of the confirmation of the model.'

Diagrammatically the confirmation problem may be presented as shown on Figure 1. The values of v and p which produce the 'real world' results m, if used as input for the model will produce e which should be a sufficiently good approximation of m taking into account the intended use of the model. Criteria for falsification or refutation should be predetermined.

CONFIRMATION TECHNIQUES

A variety of 'subjective' and 'objective' parametric and non-parametric statistical techniques may be used to judge the goodness-of-fit between measured and predicted data. Standard statistical tests are, however, fraught with problems if there is a high degree of auto-correlation in the variables. This is likely to be the case with variables associated with the thermal environment of buildings. Several tests which may be suitable are described below.

Turing's Test

Turing (1956) describes an experiment which, with appropriate modifications, can provide a simple objective model evaluation technique. In this modified test people who are knowledgeable about the operation of thermal performance systems would be presented with two sets of information, one set generated by the simulation model and the other gathered from measurements of the building. The information may consist of outcome data or graphical presentations, frequency plots etc. The observers would be asked if they can discriminate between the system and the model outputs. Failure to discriminate provides confirmation of the model.

Goodness-of-Fit Tests

When the information from a simulation program and the corresponding simuland data is a time series eg hourly internal temperatures, and presented in graphical form, eight techniques have been suggested by Cyert (1966) to check their correspondence.

1. analyse the number of turning points.
2. analyse the timing of the turning points. The model predictions may lag or lead the measured data.
3. analyse the direction of the turning points.
4. analyse the amplitude of fluctuations for corresponding time segments.
5. analyse the average amplitude over the whole time series.
6. analyse the simultaneity of turning points for different variables.
7. analyse the average value of variables.
8. analyse the exact matching of variables.

More objective measures of the goodness-of-fit may be obtained from the following statistics. The error between two time series X and Y at simultaneous time can be represented as:

$$Z_t = X_t - Y_t$$

The mean difference for n time increments in the data series is then given as:

$$Z_m = \frac{\sum Z_t}{n}$$

The absolute value of this difference is:

$$|Z_m| = \frac{\sum Abs(Z_t)}{n}$$

while the root-mean-square of the errors is given by:

$$R = \sqrt{\sum Z_t^2 / n}$$

Another measure commonly used for analysing the correspondence of two data sets is by computing the correlation coefficient.. This technique measures the degree of fit of the linear regression relationship

$$X_t = a + bY_t$$

Perfect correspondence would be represented by a=0 and b=1 for all time.

Another statistic suggested by Theil (1961) is known as the inequality coefficient U where,

$$U = \frac{\sqrt{\frac{1}{n}\sum (X_t - Y_t)^2}}{\sqrt{\frac{1}{n}\sum X_t^2} + \sqrt{\frac{1}{n}\sum Y_t^2}}$$

This coefficient describes the inequality in the time series due to three sources, unequal tendency (mean), unequal variation (variance) and imperfect covariation (covariance). U=0 when X_t = Y_t at all times.

Two non-parametric statistical tests may also be used to test aspects of goodness-of-fit. The Ψ2 test and the Kolmogorov-Smirnov two sample test may be applied to determine whether there are differences between distributions of the predicted and measured data series. Ψ2 tests for differences between predicted and observed frequencies of binned data while the Kolmogorov-Smirnov test compares the cumulative frequency distributions of the continuous data sets.

Time series analysis techniques also provide several methods of testing the correspondence between two time series. The cross-correlation function has been used to examine the relationship between two time series in the time domain, while the squared coherency function measures the linear correlation of the series at different frequencies. The cross-correlation technique was used by Williamson and Coldicutt (1986) to assess the behaviour of a thermal performance program which incorporated evaporative and heat-pump cooling plant modelling. More recently Palomo et.al. (1991) as part of the CEC PASSYS project developed several statistical tools based on time series analysis as part of the empirical validation of thermal performance simulation programs.

MODEL CONFIRMATION

The goodness-of-fit techniques outlined above tell us nothing of the ways in which the inner workings and processes of a model operates. None are a sufficient test of a model's performance and cannot be used themselves to confirm the adequacy of a model. A further demand in order to properly measure the performance of a model in a given test is to account for that part of measured behaviour *m* which can be accounted for by the input variables *v* at simultaneous time. To illustrate this point, consider a well shaded lightweight structure with a high ventilation rate. In this circumstance an estimate of the internal temperature may simply be the external temperature. If a model is to be considered adequate, *e* should provide a better estimate of *m* than *v*.

The correspondence between *e* and *m* can be calculated from Theil's inequality coefficient and represented by *U(e,m)*. Similarly the correspondence between *m* and *v* can be represented by *U(m,v)*. A measure of the model's confirmation is then given by;

$$C_S = U(m,v) - U(e,m)$$

C_S may be described as a *confirmation factor*. The factor seems intuitively to be a satisfactory indicator of a model's performance in a validation test. The value of C_S represents a single measure combining the difficulty of the validation test and how well the model, given the input parameters, performs. The following results may be derived from C_S,

(1) as *m* and *v* become unrelated at time t, for example by *m* lagging *v*, then *U(m,v)* tends towards 1. Also as *m* approaches *e* at all times t, *U(e,m)* tends towards 0.

(2) if *m* approximates to *e* at times t then C_S is positive. If *m* undermines *e* (so that *v* provides a better estimate of *m* than does the model prediction *e*), then C_S is negative.
Only when *m* is very unequal to *v* and *e* is nearly equal to *m* is confirmation strong.

(3) the maximum value of C_S is 1 when, and only when, *U(m,v)* = 1 and *U(e,m)* = 0.

(4) the maximum value C_S can attain for a particular test is equal to *U(m,v)*. This then represents a measure of the severity of the confirmation test. Following from this the *degree of confirmation* may be evaluated as,

$$D = \frac{U(m,v) - U(e,m)}{U(m,v)}$$

The formulation for C_S assumes a single input variable. Multiple input variables such as dry bulb temperature, humidity, solar radiation and wind speed as well as building occupancy variables are however used for most dynamic thermal performance simulation programs. These may be combined into a number of uncorrelated time series using the technique of principal components described by Kendall (1975; 1983). If we have n observations of p variables of v, then some new variables $\xi_1, \xi_2, \xi_3 \ldots\ldots \xi_p$ may be defined which are linear functions of the v's but are themselves uncorrelated and therefore mutually independent. The linear relationship is described by,

$$\xi_i = \sum_{i=1}^{p} l_{ij} v_j$$

The new functions ξ_i are known as principal components. The value ξ_1 is termed the first principal component and ξ_2 the second etc. The l's correspond to the components of the eigenvectors of the covariance (or correlation) matrix with the principal components being ordered by the eigenvalues (variances) of that matrix. Generally each principal component can be interpreted in terms of some physical 'real' description of the process. Where, however, the variables are measured at different scales principal components analysis based on the correlation matrix is often recommended (Jolliffe, 1986). The principal component coefficients in this case are more difficult to interpret directly.

CONFIRMATION EXAMPLE

In this part of the paper the use of the model confirmation technique outlined above is illustrated. The exercise involves 'validating' the program TEMPAL*. Extensive validation tests have previously been conducted on TEMPAL in a variety of situations (Coldicutt, Coldicutt, & Coldicutt, 1978; Williamson, 1984) and each has reported satisfactory results related to the intended use of the program.

The data used for model confirmation in this exercise were derived from extensive monitoring of the CSIRO experimental Low Energy Consumption House (LECH) in Highett, Melbourne .

* TEMPAL, A thermal performance simulation package written by A.B. & E. Coldicutt, Department of Architecture, University of Melbourne.

Goodness-of-Fit Tests

Figure 2. shows the measured and TEMPAL predicted environmental temperatures in the living area, designated Zone 1 of the house, during a free-running 7 day period. The predicted output results are from the first simulation run. The Cyert examination technique may be applied to this graphical comparison of results. Although differences are obvious one may conclude that, given the temperature measurement was taken at only one part of the zone, there appears to be sufficient correspondence to use the predictions as the basis of design decisions.

The various single value statistics comparing the measured and predicted time series are shown in Table 1. The absolute mean error, RMS error and the correlation coefficient and the inequality coefficient provide results which give some confidence that the simulation has produced sufficiently good results. The ψ^2 and K-S tests, however, clearly demonstrate that the two time series distributions are significantly different.

Confirmation Analysis

The covariance matrix was estimated from the four input variables dry bulb temperature ($v1$), direct beam solar radiation ($v2$), diffuse horizontal solar radiation ($v3$), and wind speed ($v4$). The resulting eigenvalue and eigenvector values are shown in Tables 2. The first principal component could be described as the total radiation vector and accounts for more than 93% of the variation. The second component accounts for around 6% of the variation and is dominated by the diffuse radiation component. The last two principal component are associated with the temperature and wind inputs respectively.

Because the input variables are measured at different scales the eigenvalues and eigenvectors were also estimated from the correlation matrix. These are shown in Table 3. The first vector accounts for 57% of the variation in this case and (roughly) gives equally weighting to all four input variables. The second component represents a contrast between the temperature and direct radiation variables and the diffuse radiation and wind speed variables. It is difficult to attribute a physical description to this component and it appears simply to be an artefact of the analysis. The third component is a contrast between temperature with wind and the radiation variables. This is interpreted as describing the processes of ventilation and the external film coefficients associated with the glazing elements. The forth principal component again seems to be an artefact of the analysis expressing a contrast between temperature with diffuse radiation and direct radiation with wind speed.

The confirmation factors and degrees of confirmation are shown in Table 4 calculated for the full 7 day test period. The values C_S and D relate to the principal components calculated from the covariance matrix and the C_S's and D' values relate to those calculated from the correlation matrix.

An examination of Table 4 will help to identify which areas of the computer model are operating incorrectly or which could be improved. First we can calculate the overall severity of the validation exercise as the average of C_S/D' to be 0.625, indicating a reasonable complexity of test. Gross errors found from experience to be suggested by $D'<0.7$ are not present in this example. To improve the computer model our aim is to maximise the D and D' values. We can see from ξ_1 of D' that some general improvements can be made and that in particular, from ξ_3 of D, those algorithms which deal with the external temperature should perhaps be examined first to improve the performance of the computer model. Improving an existing apparently satisfactory program requires an iterative process where small successive improvements are made as errors are eliminated and algorithms improved.

This process was adopted in the development of a new simulation package *EnCom2* based on the same computational technique as TEMPAL but incorporating improved heat flow algorithms and new operating features including the most up-to-date algorithms for calculating external film coefficients, improved shading calculations, the provision of non-isotropic sky conditions and many other features. Figure 3 shows the predicted temperatures for the 7 day period as before with all input parameters the same as the original TEMPAL run. The result appears, at least visually, to indicate a better overall match with the measured conditions. The degree of confirmation factors D', Table 5, show a small but positive increase, indicating an overall improvement in the predictions of the new program. The goodness-of-fit statistics, Table 6, compared with the Table 1, verify this improvement.

DISCUSSION

If it was only necessary to establish the truth or non-truth of a computer model validation would be relatively easy. The problem however is not that simple as at least three questions need to be answered, first, the degree of correspondence to reality, second, the severity of the test and thirdly, whether the result is sufficient to provide confidence that the model can be used in a given situation for decision making. The confirmation factor and the degree of confirmation statistics proposed above are an attempt to provide objective answers to these questions. In addition the statistics may be interpreted to indicate which areas of a program could be improved. The confirmation factor and degree of confirmation tests can be built into software development modules to ensure that changes made to programs are indeed improvements.

A minimum acceptable program performance level can be established based on the degree of confirmation factor. Perhaps a little more experience is required in using this technique, but at this stage, $D'>0.80$ would seem to ensure a program of sufficient accuracy for most design decision making.

REFERENCES

Bloomfield, D. (1985). Appraisal Techniques for Methods of Calculating the Thermal Performance of Buildings. Building Services Engineering Research & Technology(1), 13-20.

Bowman, N. T., & Lomas, K. J. (1985). Empirical Validation of Dynamic Thermal Computer Models of Buildings. Building Services Engineering Research & Technology(4), 153-162.

Clark, J. A., & Forrest, I. (1978). ABACUS Occasional Paper: 61. Validation of the ESP Thermal Simulation Program. Glasgow: University of Strathclyde.

Coldicutt, A. B., Coldicutt, E., & Coldicutt, S. (1978). An Investigation of the Thermal Performance and Life-time Costs of Public Housing in Victoria and Tasmania (AHRC Project No.57). Melbourne: Australian Housing Research Council.

Cyert, R. M. (1966). A Description and Evaluation of Some Firm Simulation. In Computer Models and Games. Proc. IBM Scientific Computing Symposium. White Plains NY: IBM.

Eppel, H., Lomas, K. J., & Martin, C. (1993). Empirical Validation in International Energy Agency Annex 21/Task 12: Final Report. Leicester, UK: De Montfort University.

Jolliffe, I. T. (1986). Principal Component Analysis. New York: Springer-Verlag.

Judkoff, R., & Neymark, J. (1993). Building Energy Simulation Test (BESTTEST) and Diagnostic Method: A Procedure for Testing the Ability of

Whole Building Energy Simulation Programs to Thermally Model the Building Fabric (IEA Technical Report). Golden CO.: NREL.

Judkoff, R., Wortman, D., O'Doherty, & Burch, J. (1983). A Methodology for Validating Building Energy Analysis Simulations (SERI/TR-254-1508 Report ed.). Golden, CO.

Kendall, M. G. (1975). Multivariate Analysis. London: Griffin.

Kendall, M., Stuart, A., & Ord, J. K. (1983). The Advanced Theory of Statistics (4th ed., Vol. 3). London: Charles Griffin & Co.

Palomo, E., Marco, J., & Madsen, H. (1991). Methods to Compare Measurements in Simulations. In J. A. Clarke, J. W. Mitchell, & R. C. Van de Perre (Eds.), Proc. of Conference Building Simulation'91 (pp. 570-576). Nice, France: International Building Performance Simulation Association.

Palomo, E., Tellez, F. M., Marco, J., & Heras, M. R. (1991). Evaluation of Passive Solar Components by Identification Techniques: Tests for Model Validation Purposes. In S. Alvarez, J. L. De Asiain, S. Yannas, & E. De Oliveria Fernandes (Eds.), Proc. of Conference PLEA'91: Architecture and Urban Space (pp. 491-497). Seville, Spain: International PLEA Organization.

Popper, K. R. (1983). Realism and the Aim of Science. New Jersey: Rowman and Littlefield.

Theil, H. (1961). Economic Forecasts and Policy. Amsterdam: North-Holland Publishing.

Turing, A. M. (1956). Can a Machine Think. In J. R. Newman (Ed.), The World of MAthematics (Vol. 4). New York: Simon and Schuster.

Williamson, T. J. (1984). Project Report: 75. An Evaluation of Thermal Performance Computer Programs. Canberra: Australian Housing Research Council.

Williamson, T. J., & Coldicutt, S. (1986). Comparison of evaporative cooling and refrigerative air-conditioning for dwellings in South Australia. Adelaide: Department of Architecture, The University of Adelaide.

Table 1. Goodness-of-Fit Statistics

Statistic	Value TEMPAL Run
Absolute Mean error	1.07
Root-mean-square error	1.22
Linear correlation coefficient	0.982
Theil's Inequality coefficient	0.027
Ψ^2 probability (DF= 15)	0.002
Kolmogorov-Smirnov probability	0.003

Table 2. EigenValues and Vectors of the Covariance Matrix

EigenValues	**EigenVectors**			
	v1	*v2*	*v3*	*v4*
105752.1	0.0124	0.9788	0.2045	0.0016
7531.0	-0.0109	-0.2043	0.9788	0.0037
31.3	0.9992	-0.0144	0.0080	0.0366
1.2	-0.0366	-0.0003	-0.0043	0.9993

Table 3. EigenValues and Vectors of the Correlation Matrix

EigenValues	EigenVectors			
	v1	*v2*	*v3*	*v4*
2.29	0.4673	0.5686	0.4988	0.4577
0.80	-0.7091	-0.2162	0.4963	0.4518
0.62	0.2260	-0.3429	-0.5127	0.7540
0.29	0.4772	-0.7158	0.4919	-0.1341

Table 4: Confirmation and Degree of Confirmation Factors TEMPAL Results

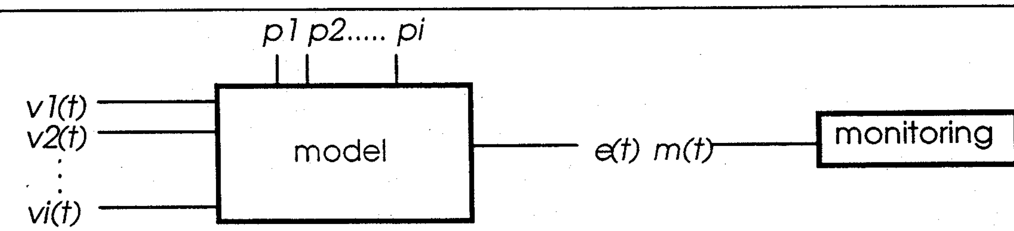
	ξ_1	ξ_2	ξ_3	ξ_4	Average
Confirmation Factor C_ς	0.892	0.679	0.188	0.865	0.656
Degree *D*	0.970	0.962	0.875	0.970	0.944
Confirmation Factor C'_ς	0.379	0.751	0.469	0.509	0.526
Degree *D'*	0.800	0.888	0.832	0.843	0.841

Table 5: Degree of Confrimation *D'* *EnCom2* Results

	$\xi 1$	$\xi 2$	$\xi 3$	$\xi 4$	Average
Degree D'	0.835	0.900	0.867	0.873	0.870

Table 6. Goodness-of-Fit Statistics

Statistic	Value *EnCom2* Run
Absolute Mean error	0.57
Root-mean-square error	0.73
Linear correlation coefficient	0.982
Theil's Inequality coefficient	0.016
Ψ^2 probability (DF= 15)	0.323
Kolmogorov-Smirnov probability	0.627



where *e(t)* model estimates/predictions/output
v1(t),v2(t)......vi(t) - input (exogenous) variables
p1,p2....pi - input parameters
m(t) - real world measured values

Figure 1. The Confirmation Problem

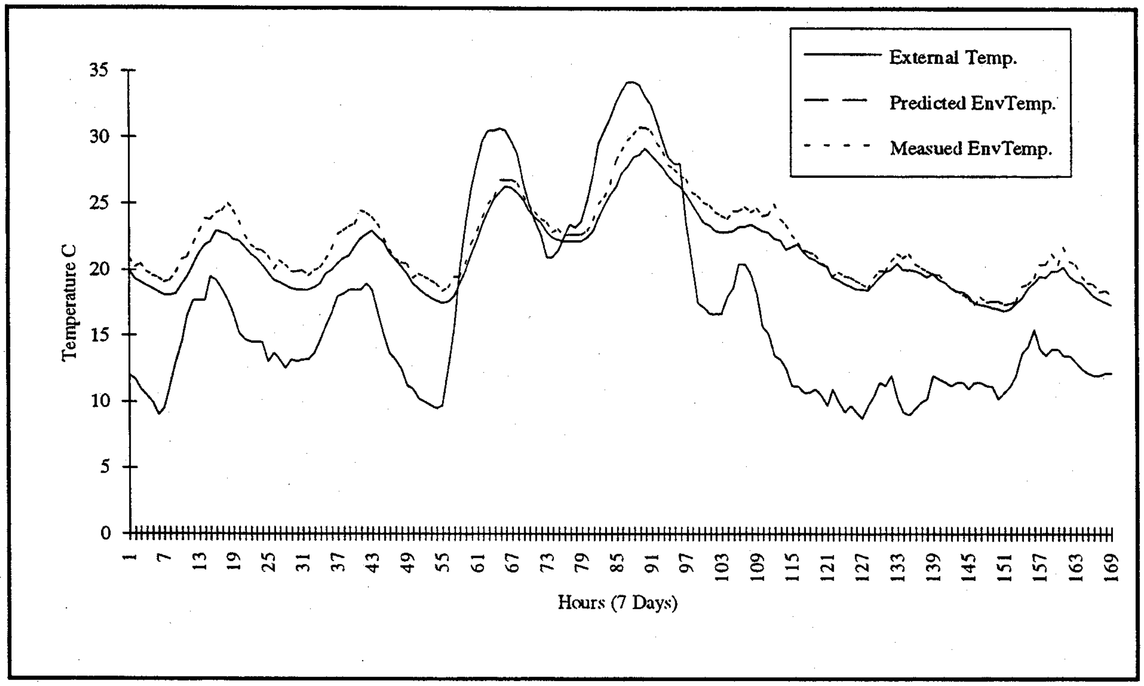


Figure 2: Comparison of Measured and TEMPAL Predicted Zone 1 Environmental Temperatures

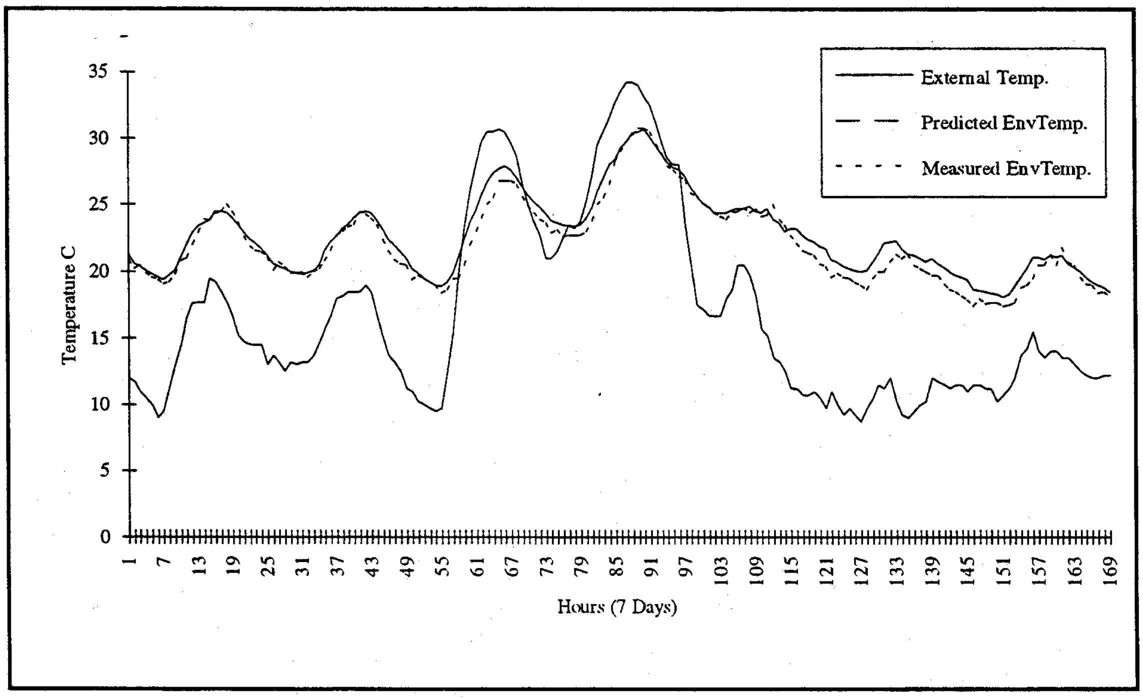


Figure 3: Comparison of Measured and *Encom2* Predicted Zone 1 Environmental Temperatures