Discrete-time nonlinear FIR models with integrated variables for greenhouse indoor temperature simulation

Manuel R. Arahal, Francisco Rodríguez, Armando Ramírez-Arias and Manuel Berenguel

Abstract— This paper shows how Nonlinear Finite Impulse Response (NFIR) models realized by artificial neural networks can be used for developing simulation models of the inside temperature of greenhouses.

The proposed NFIR models use integrated variables to reduce the number of past values needed as inputs. Several NFIR models have been developed using past data following a systems identification methodology. All data have been obtained from a real greenhouse in Southern Spain dedicated to tomato crop. The NFIR models are later compared with a model based on first principles.

The results obtained in the a posteriori application of the models to new real data show that the performance of the NFIR model with integrated variables compares well with that of a first principles model, although the generalization capabilities of the latter are superior.

I. INTRODUCTION

Crop production in greenhouses has experimented a large expansion recently especially in some regions of Europe. The requirements for climate control has propelled the research in this area and today there is a number of papers dealing with modelling [1], short term climate control [2] and long horizon control devised either to minimize the energy consumption [3] or maximize economic profit [4].

For climate control purposes it is convenient to obtain short term prediction models that allow the controller to compute the best control move. A hierarchical scheme has been proposed in [4] by the authors to control crop growth by means of manipulating climate variables and taking into account optimal production criteria (maximum benefits). In this situation the temporal horizon considered can span a whole season, hence the need for simulation models such as the one proposed here.

Artificial neural networks (ANN) can be used for developing simulation models in many ways (see [5] and [6] for a review of ANN applications in greenhouses). In this paper we have restricted ourselves to feedforward networks whose input vector contain past values of the independent variables. This kind of models is a nonlinear extension of the Finite Impulse Response (FIR) family well known in the signal processing and control literature.

There are three main differences with previous work, namely (i) the use of discrete time black box models for simulation (not just one step ahead prediction), (ii) the use of real data for training and validation spanning a whole season, and (iii) the system identification approach employed for variable selection, order selection and network training and testing.

The models have been developed using past data and compared with new data. All data have been obtained from a real greenhouse in Southern Spain during normal operation producing tomato crop. The neural models are also compared with a model based on first principles whose parameters have been selected to match the same greenhouse that provided the data.

In the next section the data used for developing the models is shown, followed by section III where the neural models structures and the development of the actual models are explained. Similarly, in section IV a model based in first principles derived by the authors elsewhere is presented here and compared with the NFIR models with both past and new data. The paper ends with some conclusions.

II. DATA

Historical data have been obtained from a real greenhouse belonging to the installations of the Estación Experimental Las Palmerillas, located in El Ejido (Almería, Southeast Spain, Lon. 2° 43' W, Lat. 36° 48' N, Alt. 151 m). A typical greenhouse of this zone was used with 876 m^2 of soil surface and 4.4 m high, formed by two symmetric flat slope roof with five North-South oriented naves (7.6 $\times$ 23.2 m). The greenhouse has PE film of 200 microns thick as covering material and lays on a structure made of galvanized steel. The climate control actuators installed in the greenhouse are natural vents (lateral and roof), shade screen and air heaters.

The data adquisition system provides values of control actions and internal and external conditions sampled each minute. Not all data have been disclosed in order to perform a final test comparing different models. Figure 1 shows graphically the disposition of the different data sets. Boxes in gray correspond to days whose data has been used for training or validation. The last boxes (marked with stripes) belong to the end of the season and have not been used except for final comparisons.

A subset of all variables are just considered as candidates to be part of the input vector of the neural networks based on previous experience on the problem and on the fact that the model will be used for simulation (not just prediction). They are presented in table I divided into two groups: control actions (CA) and disturbances (D).
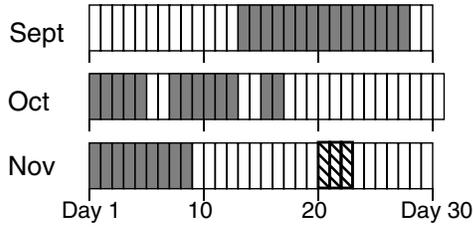
Fig. 1. Marked boxes indicate days when data have been collected. Days in grey have been used for training or validation, the striped boxes are days used just for the final comparison between models.

Symbol	Variable	Type
v^1	Aggregated ventilation actuation	CA
v^2	Heating actuation	CA
v^3	External temperature	D
v^4	Wind speed	D
v^5	Radiation (PAR)	D

TABLE I

SUBSET OF VARIABLES CONSIDERED AS CANDIDATES FOR INPUT

VARIABLES.

It has to be noted that in order to perform a simulation the input signals must be either forecasted (this is quite feasible with temperature and radiation), controlled (as is the case with the actuations) or selected in some way to provide a realistic scenario. This latter case correspond to wind direction and speed that are difficult to forecast but whose effect can be considered statistically.

III. DESCRIPTION OF THE MODELS

There are many model structures available for identification. In this application we are interested in obtaining a discrete-time model for simulation where the real output of the system $(y(k))$ is not known at any time (except perhaps at the beginning of the simulation $k = 0$).

A model that uses past values of y (i.e. an autoregressive model) must then use its own output $\hat{y}$ as an estimation of the true output and use it recursively during the entire simulation time. This can cause a built-up of the simulation error producing errors that are larger for larger simulation horizons. A model that uses just past values of the input signal belongs to the family of Finite Impulse Response (FIR) models. The output of a FIR model is obtained as a linear combination of past values of the system's input. Since the real output of the system is not needed, this kind of models produces simulation errors that are independent of the simulation horizon. Also, any FIR model obtained by identification is stable (in the BIBO sense) since the output of the model is obtained as a combination of past input values.

In a FIR model the convolution of the system's impulse response is truncated to the first m terms. This poses serious problems when the model is intended to represent a system that has poorly damped poles. In fact, a large number of past values of u may be needed to represent a simple transfer function model. Also, model variance error is usually high
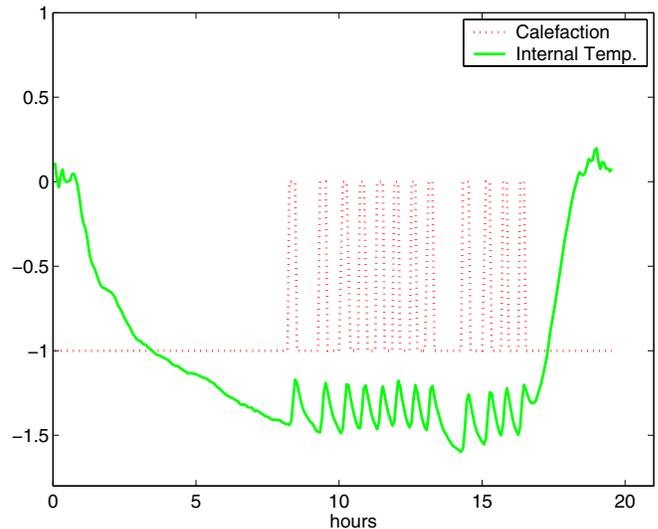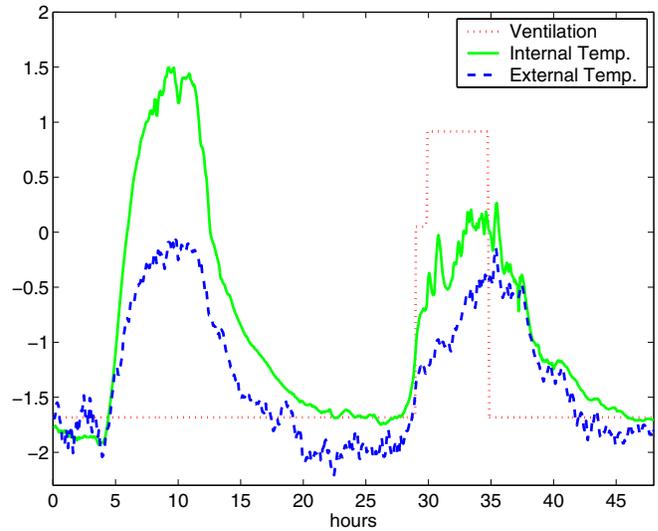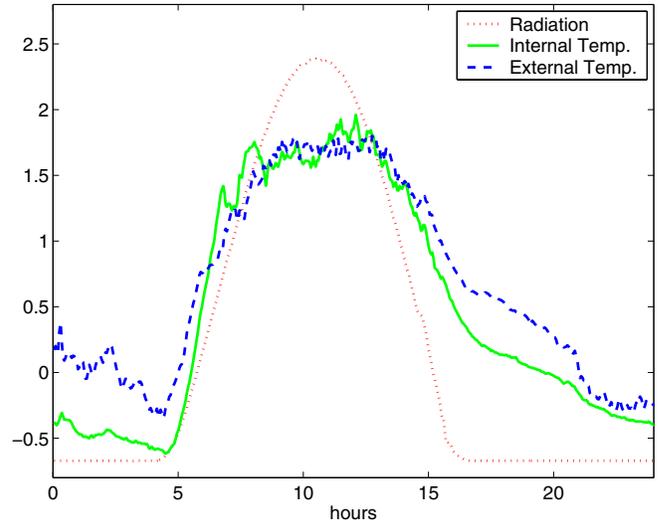

Fig. 2. Views of the historical data used for training: evolution of some normalized variables during three different time periods. On each plot the variable and its effect on the inside temperature of the greenhouse is shown.

4159

for a FIR model due to its high order. Bias error is also high due to the necessary truncation of the convolution product. In this application a method is derived to diminish the drawbacks while retaining the benefits of using a FIR-like model structure. The rationale behind the method is that for many physical systems the output does not react to short variations in the input but rather to the accumulated effect of these variations. This is specially true for sampled systems with short sampling periods. In these situations it makes sense to use as model inputs the integrated values of the input variables instead of the variables themselves, reducing the number of values actually used as input variables of the model. In the case presented here, the output of the model is the indoor temperature. It can be deduced from physical considerations (and confirmed by observation) that this variable does not change significantly due to high frequency changes of some inputs such as radiation or heating. In other words, some input variables have to maintain their value for many sampling periods in order to exert changes in the output. This provides a justification for the use of integrated variables.

The NFIR model structure is the nonlinear counterpart of the FIR family. The input variables used are the same, but in the first case the function that combines them is a nonlinear mapping produced in this paper by an artificial neural network. As a result, the input vector for the ANN at sample time k is computed as:

$$\mathbf{z}(k) = \begin{array}{l} [\quad v^1(k - d^1 - 1), ..., v^1(k - d^1 - n^1), \\ ..., v^{nv}(k - d^{nv} - n^{nv})] \end{array} \quad (1)$$

where all variables included (v^1 to v^{nv}) are control actions or disturbances. The values d^j $j = 1, ..., nv$ are the dead time for variable j. Model orders are the number of lagged values n^j of each variable from $j = 1$ to $j = nv$. The dimension of the input vector is thus $\dim \mathbf{z} = \sum_{j=1}^{nv} n^j$

The output variable is assumed to be a nonlinear function of the input vector plus a white noise signal n

$$y(k) = \mathrm{f}(\mathbf{z}(k)) + n(k + 1) \quad (2)$$

In this paper we have first determined the time lag for variables v^3 and v^5 because they show daily cycles that are followed by the output. Then the effect of both variables in the output has been crudely determined by a low order model and subtracted from the output. The residual has been used to determine time lags for other variables using a correlation technique with controlled portions of the data where the input variable has significant variations. The results are: $d^1 = 1$, $d^2 = 6$, $d^3 = 1$, $d^4 = 1$ and $d^5 = 2$.

A. Integrated variables

A variation of NFIR models consists of including integrated values of some variables. Then, instead of using $v(k - d - i)$ for $i = 1, 2..., n$ it is possible to accumulate the effect in just m sums with $m < n$ being the new variables

$$w(k, j) = \sum_{i=s_j}^{i=f_j} v(k - d - i) \quad (3)$$

Variable	model orders	Windows lengths
v^1	$n^1 = 1$	25
	$n^1 = 2$	18, 40
	$n^1 = 3$	16, 41, 90
	$n^1 = 4$	16, 41, 90, 162
	$n^1 = 5$	16, 41, 90, 162, 283
v^2	$n^2 = 1$	19
	$n^2 = 2$	12, 20
	$n^2 = 3$	10, 18, 23
	$n^2 = 4$	10, 18, 23, 35
	$n^2 = 5$	10, 18, 23, 35, 54
v^3	$n^3 = 1$	45
	$n^3 = 2$	32, 130
	$n^3 = 3$	29, 110, 220
	$n^3 = 4$	29, 110, 220, 312
	$n^3 = 5$	29, 110, 220, 312, 534
v^4	$n^4 = 1$	14
	$n^4 = 2$	14, 24,
	$n^4 = 3$	14, 24, 45
	$n^4 = 4$	14, 24, 45, 74
	$n^4 = 5$	14, 24, 45, 74, 114
v^5	$n^5 = 1$	54
	$n^5 = 2$	40, 120
	$n^5 = 3$	32, 91, 215
	$n^5 = 4$	30, 87, 198, 279
	$n^5 = 5$	27, 85, 198, 279, 540

TABLE II

WINDOWS LENGTHS CHOSEN TO PRODUCE INTEGRATED VARIABLES.

for $j = 1$ to $j = m$. Obviously the initial and final index for the sums must verify: $s_1 = 1$, $s_{\alpha+1} = f_\alpha + 1$ for all $\alpha = 1, ..., m - 1$ and $f_m = n$.

The initial (s) and final index (f) define a time window in which the integration of variable v takes place to yield the integrated variable w. The lengths ($f - s + 1$) of these time windows do not need to be equal. A simple method to derive windows lengths from data is to compute the correlation between the output variable and the integrated variable for different lengths and choosing the best values. Results are more accurate if just controlled portions of the data are used, for instance, to compute correlations between integrated heating action and inside temperature one has to use data of nights were the heating has been used.

It is worth noting that the proposed method does not imply the training of models at all (unlike forward inclusion that will be used later for model order selection). The results obtained for variables v^1 to v^5 are shown in Table II. It can be seen that windows lengths depends on model order as one should have expected since the relative effect of past inputs in the output decreases with time.

Using the integrated variables, the input vector for a neural NFIR model at sample time k is computed as:

$$\mathbf{z}(k) = \begin{array}{l} [\quad w^1(k, 1), ..., w^1(k, n^1), ... \\ w^5(k, 1), ..., w^5(k, n^5) \ ] \end{array} \quad (4)$$

At this point the only structural parameter that need to be selected is the number of inputs for each integrated variable n^1 to n^5.

It has to be noted that the use of integrated variables produces some loss of high frequency information. This can pose a problem if the model is later used in situations where this frequency is of importance, for instance in model based control. The windows lengths and the sample time determine the cutoff frequency. In this application, the sample time is fixed and cannot be increased. On the other hand, the windows lengths are selected in such a way that the resulting integrated variables allow the construction of an accurate simulation model.

B. Input variable selection

For linear model structures Akaike's criterion and others may be used to choose the model order that achieves a balance between accuracy and complexity of the model. For neural models other procedures must be used. Here we restrict ourselves to forward inclusion, leaving the exploration of other techniques for future work. In forward inclusion lagged variables are included in the model one by one, selected according to how much they help reducing the model error.

Since the expected nature of the relationship between input variables and observed output (or residuals of previous models) is nonlinear it does not make sense to look for linear correlations among them. The procedure to select variables should be done in a different manner. A possible approach is to use a nonlinear measure of correlation such as the one proposed in [7] or orthogonal forward selection and backward elimination [8].

The problem arises on when to stop adding independent variables, since any new variable will introduce new degrees of freedom in the model that will allow to reduce the error in the available data. Validation techniques must be carefully used to perform model selection.

C. Model creation and selection

The adjustable parameters of the models are obtained minimizing a quadratic criterion of the simulation error $e(k) = y(k) - \hat{y}(k)$. In order to assess the goodness of the models the root mean squared error (RMSE) is used as a figure of merit. For a given set S of data the root mean squared error over this set is computed as:

$$J_S = \sqrt{\frac{1}{N_S} \sum_{t=1}^{N_S} e(t)^2} \qquad (5)$$

being N_S the number of observations in the set.

Usually a training set (TS) is used to minimize J, producing a set of model parameters. The usual practice is to produce many models of different orders or types and to compare them using some left aside data named validation set (VS). It is expected that models with a larger number of parameters will produce a lower value of J_{TS}. The comparison on the VS will hopefully provide an estimate of the generalization capabilities of each model. The selection of the final model should be based on this estimation.

Model denomination	orders $(n^1, ..., n^5)$	J_{TS+VS}
NFIR^{12212}	(1, 2, 2, 1, 2)	0.74°C
NFIR^{12213}	(1, 2, 2, 1, 3)	0.68°C
NFIR^{12214}	(1, 2, 2, 1, 4)	0.63°C
NFIR^{12314}	(1, 2, 3, 1, 4)	0.61°C
NFIR^{12315}	(1, 2, 3, 1, 5)	0.60°C
NFIR^{13315}	(1, 3, 3, 1, 5)	0.59°C

TABLE III

COMPARISON OF SEVERAL NFIR MODELS.

For any particular model structure, a neural network has to be constructed. This typically involves training more that just a network because the number of nodes of the network, the initial value for its weights, the training rule and the number of iterations are all factors that have an influence on the performance of the network. There is a number of techniques for neural model selection [9] or regularization theory [10] can be tested in order to try to improve the generalization capabilities of the models. In this work we will make an intensive use of the VS for neural selection.

In this application the VS consists of 70 % of all data pairs available up to November 9th. The other 30 % constitutes the TS. Data from November 21th, 22th and 23th (see figure 1) will be referred to as comparison set (CS) and will not be used in any way until the final comparison.

Before comparing the neural approach with first principles models it is interesting to show the performance obtained by the different neural models. Table III shows also some figures of merit: the root mean squared error (RMSE) in Celsius degrees for available data set (comprising both TS and VS) for some selected models.

Figure 3 shows the temperature for a day in which the heating was used and the simulation obtained using model NFIR^{12314} realized by a neural network of 20 hidden nodes.

IV. COMPARISON WITH A FIRST PRINCIPLES MODEL

This model is derived in [1] although its physical parameters have been refitted to the particular greenhouse were data has been obtained. This procedure has been straightforward for some magnitudes (mass, length, and so on) but more elaborate in others (e.g. heat transfer coefficients).

Table IV shows the RMSE for some NFIR models and for the first principles model (FPM). It can be seen that the black-box models achieve a smaller simulation error in the data set that has been used for validation. However, the FPM provides a better simulation in the comparison set (CS). This data set has been spared for this comparison (see figure 1). This set was not disclosed until the end and so was not used to guide the model selection.

It is also worth noting that the NFIR model that performs better in the CS is not the same that did better during training and validation. This is due to the fact that model selection based on set aside data is often biased towards higher orders and thus prone to overfitting.

The simulation error is, for all models, less than 1°C RMS. This is more than enough for the purpose of hierarchical
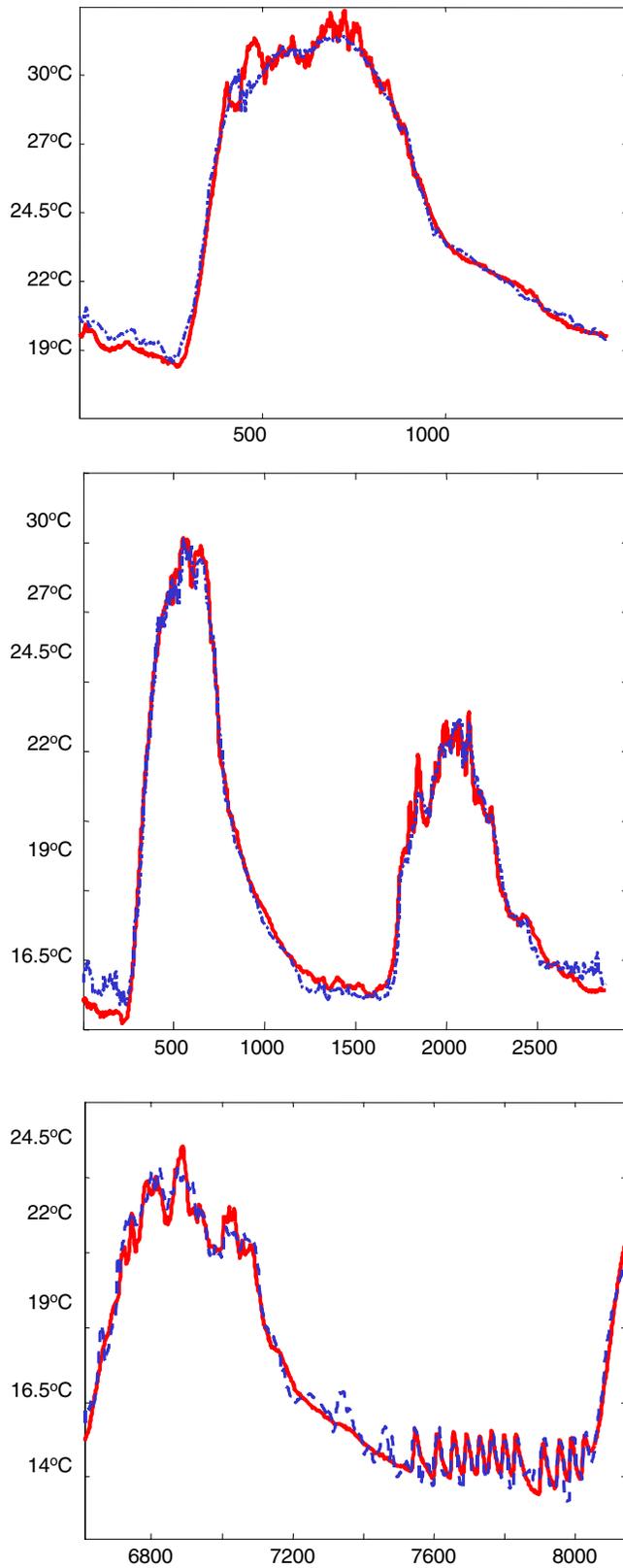
Model	J_{VS}	J_{CS}
NFIR[12214]	0.79°C	1.09°C
NFIR[12314]	0.78°C	1.01°C
NFIR[12315]	0.72°C	1.12°C
NFIR[13315]	0.66°C	1.18°C
FPM	0.88°C	0.93°C

TABLE IV

RMS ERRORS OF SOME NFIR MODELS AND A FIRST PRINCIPLES MODEL
(FPM).

control that motivated their development. However, the FPM took vastly more experimentation and work to be developed (see [1]).

V. CONCLUSIONS

The paper has shown that the NFIR model family is suitable for producing low order simulation models of the inside temperature of greenhouses provided that integrated variables are used.

A simple method for determining the time windows to be used to obtain integrated values of input variables has been presented.

The results show that the performance of the neural model compares well with that of a first principles model, although the generalization capabilities of the latter are superior. However the time spent to produce the NFIR model is much smaller. Consequently the work to be done to produce simulation models for a new built greenhouse (with different structure and/or size) is smaller using the black box data-driven approach.

REFERENCES

[1] Rodrí guez, F. (2002). Modelado y control jerárquico de crecimiento de cultivos bajo invernadero. PhD thesis. University of Almerí a. Spain.

[2] Rodrí guez, F., M. Berenguel and M.R. Arahal (2001). Feedforward controllers for greenhouse climate control based on physical models. *Proceedings of the European Control Conference ECC01* pp. 2158–2163.

[3] van Ooteghem, R. J. C., J. D. Stigter, L. G. van Willigenburg and G. van Straten (2003). Optimal control of a solar greenhouse. *Proceedings of the European Control Conference ECC03*.

[4] Rodrí guez, F., M. Berenguel and M.R. Arahal (2003). A hierarchical control system for maximizing profi t in greenhouse crop production. *Proceedings of the European Control Conference ECC03*.

[5] Linker, R. and I. Seginer (2004). Greenhouse temperature modeling: a comparison between sigmoid neural networks and hybrid models. *Mathematics and computers in simulation* pp. 19–29.

[6] Ferreira, P. M., E. A. Faria and A. E. Ruano (2001). Neural network models in greenhouse environmental control. *Neurocomputing* pp. 51–75.

[7] Yuan, J.-L. and T.L. Fine (1998). Neural-network design for small training sets of high dimension. *IEEE Transactions on neural networks* **9**, 266–280.

[8] Mao, K. Z. (2004). Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **34**, 629–634.

[9] Moody, J. and J. Utans (1995). Architecture selection strategies for neural networks: application to corporate bond rating prediction. In: *Neural Networks in the Capital Markets* (A.-P. Refenes, Ed.). Wiley. Chichester. pp. 277–300.

[10] Guo, P., M.R. Lyu and C.L.P. Chen (2002). Regularization parameter estimation for feedforward neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **33**, 35–44.

Fig. 3. Actual interior temperature (solid line) and simulation obtained by model NFIR[12314] (dashed line) in a clear day (upper plot), in a day when ventilation was in use (middle plot) and in a day with heating (lowerplot). In all cases the horizontal scale is given in sampling periods (minutes).