MAKALAH PENELITIAN

IDENTIFICATION OF CUMULATIVE FRUIT RESPONSES DURING STORAGE USING NEURAL NETWORKS

(Identifikasi Terhadap Respon Kumulatif Buah Selama Penyimpanan Dengan Metode Neural Network)

Wahyu Purwanto*)

ABSTRACT

Neural networks are useful to identify complex nonlinear relationships between input and output of a system. Cumulative fruit responses such as water losses and ripening during storage are characterized non-linearly. For identification, several patterns of these cumulative responses, as affected by environmental factors, are often conducted by repeating the experiment several times under different environmental conditions. It is not well-known how many response patterns (training data sets) are necessary for an acceptable identification. This research explores an effective way to identify the cumulative responses of tomato during storage using neural networks. Firstly, data for identification were obtained from a mathematical model. Secondly, the relationship between the number of response pattern and the estimation error were investigated. The estimated error becomes smaller when the number of response pattern is three or more. This suggests that three types of response patterns allow cumulative responses to be successfully identified. Besides, an addition of linear data (1, 2, .., N) as input variable significantly improves the identification accuracy of the cumulative response. Finally, the identification of actual data was implemented based on these results and the satisfactory results will be obtained.

Keywords: Storage process, dynamic system, cumulative fruit responses, identification, neural networks

INTRODUCTION

In fruit production processes, the final quality of fruit is dependent on the storage process. Up to now, cold storage is one of the method to control environmental factors. Under this condition, however, the quality of fruit seems to be never improved while its freshness can be maintained. In order to keep the fruit quality, it is important to control the environmental factors in the storage building adequately with respect to the physiological status of fruit.

There are many types of fruit responses during the storage process. Cumulative fruit responses such as water losses and color change, which are characterized by the change in one direction (increase or decrease), are well used for evaluating the fruit quality. These data are important in controlling the fruit quality. Models of these cumulative responses to environmental factors are necessary for realizing the optimal control of the fruit storage process.

System identification is one of the modeling techniques for an unknown dynamic system (Eykhoff, 1974). In the modeling techniques, a model is built based on actual input and output data measured from a real system. Neural networks have a capability to identify complex nonlinear relationships between input and output data sets with their own high learning ability (Hinton, 1992). In recent years, neural networks have been widely applied to the identification of agricultural production systems (Hirafuji, 1991; Honjo *et al.*, 1992; Seginer and McClenden, 1992; Morimoto *et al.*, 1995). Purwanto *et al* (1996) showed that a neural network model was more effective than an ARMA model for identifying such nonlinear complex systems as plant responded to environmental factors. Hsu *et al* (1995) also showed the same results in another research.

The problem in getting the cumulative responses for identification using neural networks is the reason to repeat the same experiment several times under different environmental conditions in order to obtain several types of input and output data sets. It is however not well-known how many response patterns the input and the output data sets are required for acceptable identification. It is also difficult to identify the dynamic relationships between the input of environment and the output of cumulative responses, because the cumulative responses always increase (or decrease) in one direction, regardless the change direction of input.

In this study, the effective way to identify such cumulative responses of tomato and the number of response patterns (training data set) required for acceptable identification will be examined through both simulation and actual experiment using neural networks.

MATERIALS AND METHODS

Data for identification

Data obtained from a mathematical model (Estimated data) : Firstly, the data for identification were obtained from a mathematical model in order to make the identification exactly under no measuring noise. The model is given by a first order differential equation as follows.

$$X = A\,x + B\,u, \quad y = C\,X \qquad 1$$

$$z = \int y\, dt \qquad 2$$

*) *Dept. of Agro-industrial Engineering Faculty of Agricultural Technology, Gadjah Mada University*

where x is a state variable, u is an input variable which corresponds to environmental factor, y is an output variable which corresponds to the rate (or velocity) of fruit response, A, B and C are coefficient factors and here given as -0.9, 0.9 and 1.0, respectively. By integrating y, the cumulative response z can be obtained, as expressed in equation 2. The output y is called *"velocity response (output 1)"* and the output z *"cumulative response (output 2)"*. It is no problem to use a linear model in order to examine the relationship between the number of data pattern and the estimated error using neural networks, since the fruit responses are characterized by nonlinearity.

Actual data (fruit responses) : Secondly, actual data were used for identification. The input is a storage temperature and the outputs are fruit responses such as the water loss and the color change of tomato (*Lycopersicon esculentum* Mill. cv. Momotaro). The sampling interval was one day. The experiments were carried out using a chamber (Tabai-espec, LHU-112M), where the temperature and the relative humidity are controlled at the accuracy of 0.1°C and 2% RH, respectively. For simpler identification of fruit responses, the data were sampled at the rate of one data in a day. The water loss of the fruit was estimated from the weight using an electric balance (Sartorius, LC-621S). The color (ripening) of tomato was evaluated using a color sensor (Minolta, CR-200b). For the evaluation of the color, which is the range of green to redness, a hue angle (degree) in the Lch method was used.

Neural networks for identification

Neural networks are capable of recognizing the relationships between the input and the output of an ill-defined system with their own learning ability. Figure 1 shows a three layer neural network for identifying the dynamic responses of a single input single output (SISO) system. Moreover, a single input multi output (SIMO) system was also supposed.

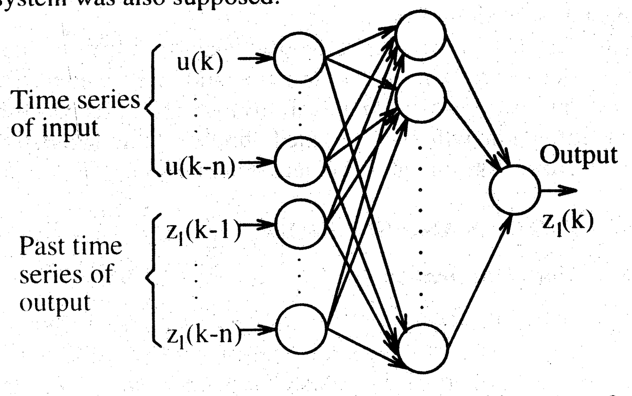


Fig. 1. A three layer neural network architectures for system identification (u(k): input, z_1(k): output, n: number of system parameter).

The input is the temperature {u(k)} and the outputs are two types of fruit responses, the water loss {z_1(k)} and the color {z_2(k)} of fruit, which are both given by cumulative responses. As shown in the figure, for the identification of a SISO system, the current output, z_1(k), is estimated from n*th* historical output data {z_1(k-1), ..., z_1(k-n)} and (n+1)*th* historical input data {u(k), u(k-1), ..., u(k-n)} (k: sampling time point, n: number of system parameter). So, the neuron number in the input layer is (2n+1) and that in the output is 1. In the case of a single input and two output system, on the other hand, the input neuron number is (3n+1). The neuron number in the hidden layer was determined through trial and error. The learning algorithm of neural networks is an error back propagation (Rumelhart *et al.*, 1986; Chen *et al.*, 1990). The learning rate and the momentum factor are 0.02 and 0.8, respectively. The iteration number for learning is 30,000.

IDENTIFICATION OF MODEL DATA

Cumulative and velocity responses obtained from a mathematical model

To begin with, the data for identification were obtained from the model given by equations 1 and 2, and then divided into two groups: the training data set for building a model and the testing data set for evaluating the accuracy of the model. Figures 2 and 3 show the training data set, which has seven types of response patterns, and the testing data set, which has six types of response patterns, respectively.

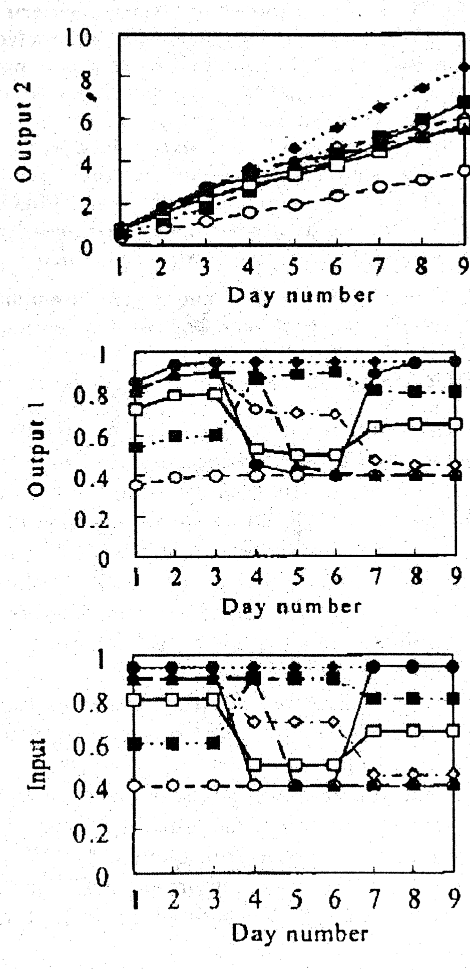


Fig. 2. Training data set consists of seven types of velocity and cumulative response patterns, obtained from a mathematical model.

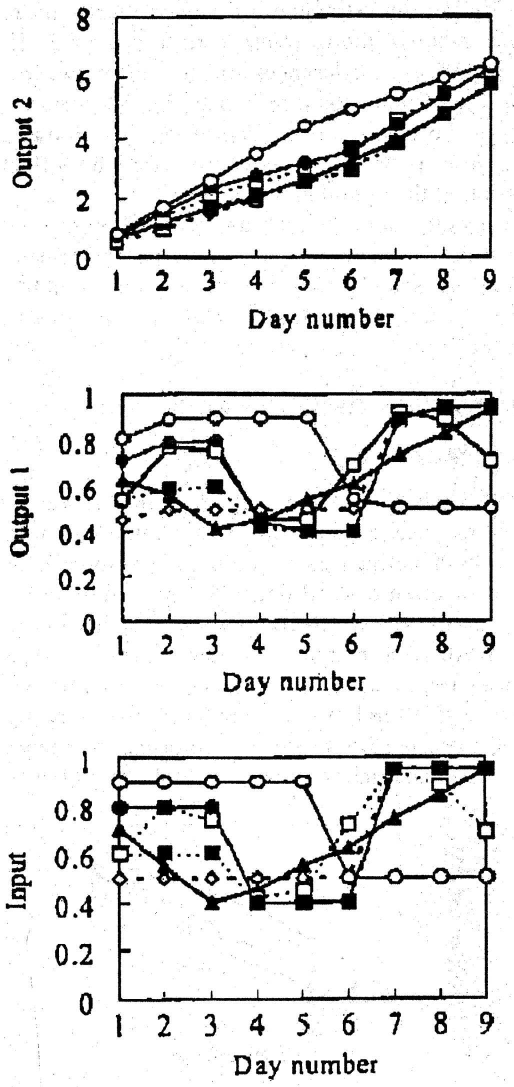


Fig. 3. Testing data set consists of six types of velocity and cumulative response patterns, obtained from a mathematical model

All data were obtained from a mathematical model. In both figures, the bottom figures indicate the input data, the middle ones the velocity responses, and the top ones the cumulative responses which were obtained by integrating the velocity responses. Each response has nine data. Such data number can be seen scant for effective identification. The training data set and the testing data set were obtained from the same model but they were independent with each other.

Comparison of the estimated accuracy using neural networks and least squares

Table 1 shows the comparisons of the estimated error in the identification of the velocity responses and the cumulative responses using the neural network and least squares methods (Auto Regressive Moving Average and Moving Average models).

Table 1. Comparison of the estimated error for the velocity and cumulative responses obtained from a mathematical model using the neural network.

Data	**Estimated Error**	
	Neural Network	Least Squares
Velocity response	*0,0033*	*0,0019*
Cumulative response	*0,1066*	*1,2478*

For identification using the neural network, the system parameter number n and the hidden neuron number in the neural network were 1 and 5, respectively. Here, the minimum value of system parameter number n=1 was used for this identification from a viewpoint of computational time saving. In three methods, the estimated error for the velocity responses are much smaller than that for the cumulative responses. This is caused by the velocity responses being quite similar to the change pattern of the input. From the estimated error for the velocity responses, it was slightly smaller with an ARMA (*Auto Regressive Moving Average*) and a MA (*Moving Average*) than with the neural network while all the errors were enough small for identification. However, the estimated error for the cumulative response was much smaller by the neural network and an ARMA model than by a MA model. Thus, it is found that the neural network is superior to a MA model for the identification of the cumulative responses from several simulation.

Improvement of the identification of cumulative responses

From Table 1, the estimated error for the cumulative response was larger than that for the velocity response. A further step was required to improve the performance of the neural network shown in Fig.1 because the cumulative responses are more important than the velocity responses for control. Improvisation was done by adding a linear data $\{d(k)=1,2, ..., N\}$ as one of the input variables because the cumulative responses are characterized by the change in one direction (only increasing or decreasing direction), regardless of the change direction of the input. In this case, the current output $z_l(k)$ is estimated from *nth* historical output data $\{z_l(k-1), ..., z_l(k-n)\}$, linear data $d(k)$ and $(n+1)$*th* historical input data $\{u(k), u(k-1), ..., u(k-n)\}$. The neuron number in the input layer is $(2n+2)$ and that in the output is 1. This improved type is called a "type II neural network", and the basic type is called a "type I neural network".

Figure 4 represents the estimated errors obtained from the identification of the testing data, as a function of the number of system parameter n, using the type I and a type II neural networks. The error by the type I neural network dramatically dropped over n=2. On the other hand, the estimated error by the Type II neural network showed a markedly low value from n=1. Under n=1, therefore, the error by the type II became much lower than that by the type I. This is probably due to appropriate matching of cumulative responses and linear data. This result shows that the type II neural network provides much better performance for identification at the smallest value n=1. This is apparently profitable for the prediction of small number of data. The case of n=2 may be also attractive because of its smaller error.

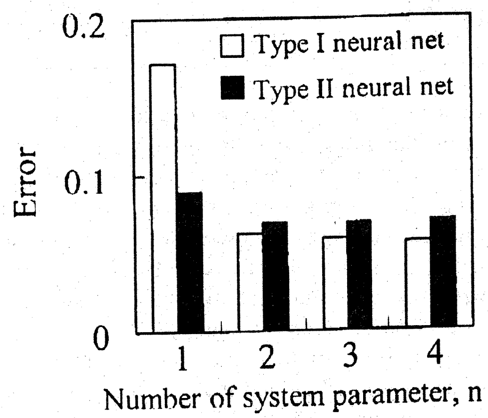


Fig. 4. Estimated errors, as a function of the number of system parameter n, in the identification of cumulative responses.

Comparison of the estimated error under the different number of response pattern

The type II neural network was superior to the type I neural network under n=1. It was also found that the estimated error under n=2 was enough small to identify the cumulative responses. The last exploration of this study is to examine the number of the response pattern required for acceptable identification of the cumulative responses under n=1 or 2 using the type II neural network. Figure 5 shows the relationship between the number of the response pattern and the estimated error under n=1 using the type II neural network. Only the testing data set was used for this examination.

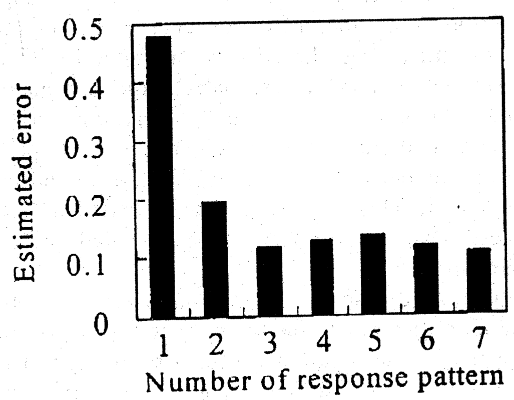


Fig. 5. Estimated error, as a function of the number of the response pattern, obtained from identification of the cumulative responses using the type II neural network (n=1).

From the figure, the estimated error dramatically dropped when the number of the response pattern is 2 or 3. There was no significant differences in the error when the number of the response pattern is over 3. This feature was the same as the case of n=2. Therefore, it is found that three response patterns is allowable for the effective identification of the cumulative responses.

This result suggests that an acceptable model for control can be obtained by repeating the experiment at least three times under different temperature conditions. In this case, it may be necessary to well change the time course of the input.

IDENTIFICATION OF FRUIT RESPONSES

Actual fruit responses

Figure 6 shows the daily changes in the water loss and the color (hue angle in the Lch method) of tomato, which are both defined as a cumulative response, under different temperature conditions. Relative humidity was kept constant (90% RH). Seven types of responses patterns were obtained. Each response has eight data. These types of responses (cumulative responses) are often observed in the area of agricultural production. As shown in the figure, it is found that the response of the color can be seen quite nonlinear while that in the water loss looks more linear.

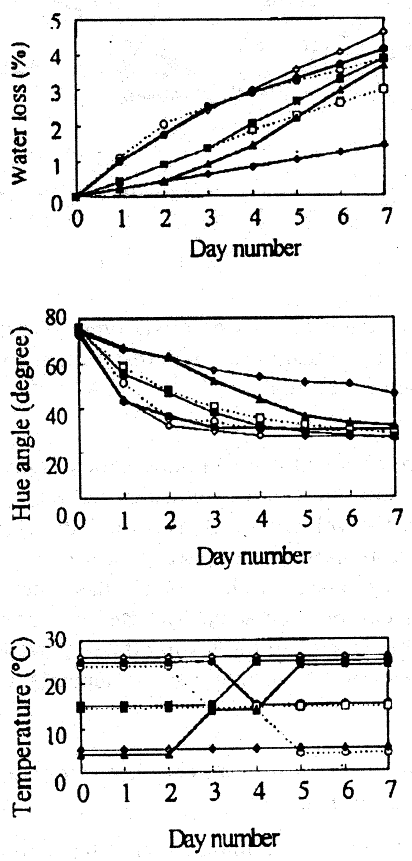
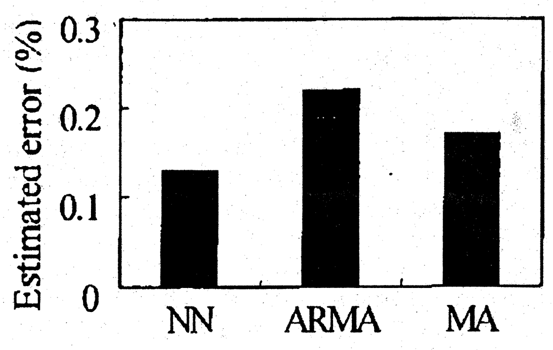


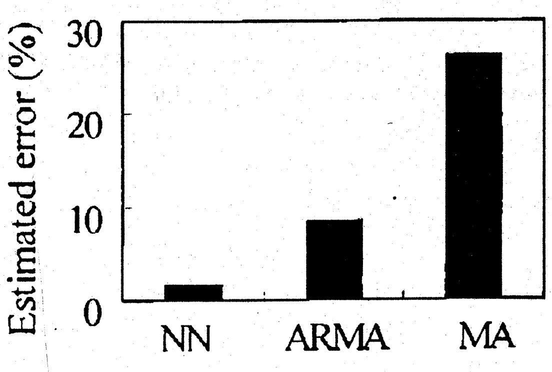
Fig. 6. Daily changes in the water loss and the color (hue angle in the Lch method) of tomato (Actual).

Identification of actual cumulative fruit responses

Next, the actual data shown in Fig.6 are identified. Figure 7 (a) and (b) show the comparisons of the estimated errors in the identifications of the water loss and the color of fruit using the type II neural network (NN), an ARMA model and a MA model.



(a) Water loss of fruit



(b) Fruit color

Fig. 7. Comparisons of the estimated error using the type II neural network, an ARMA model and a MA model.

In the identification of the color, it is found that the estimated error is much smaller by the type II neural network than by an ARMA. This is probably due to the strong nonlinear relationship between the color of fruit and the temperature. This means that the type II neural network is superior to an ARMA model for the identification of nonlinear cumulative responses.

Figure 8 represents the comparisons between the estimated responses, calculated from the type II neural network, and the observed responses of water loss. The system parameter number and the hidden neuron number were 1 and 5, respectively.

It is found that all the estimated responses are closely related to the observed responses.

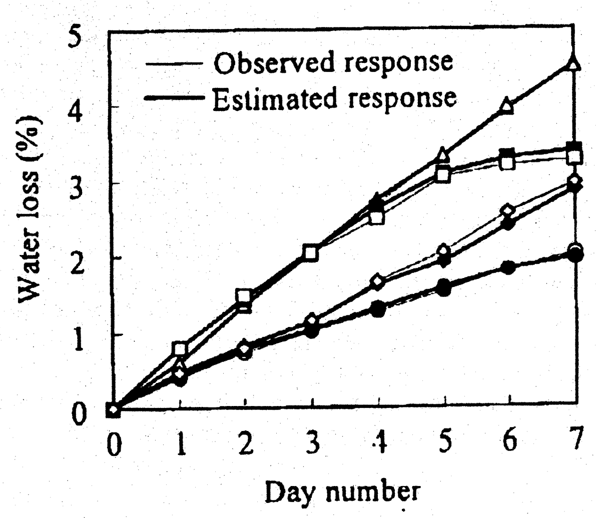


Fig. 8 Comparisons of the estimated responses and the observed responses of the water loss of tomato, as affected by temperature (Actual data).

Figure 9 shows estimated relationships between the temperature and the water loss of tomato under different days after storage, estimated using the type II neural-network model. These relationships were obtained by plotting the values on 3, 4, 5 and 6 days after storage in each cumulative response under different constant temperatures. Closed circle on the figure represent observed values. Relative humidity is assumed to be kept constant (90%RH).

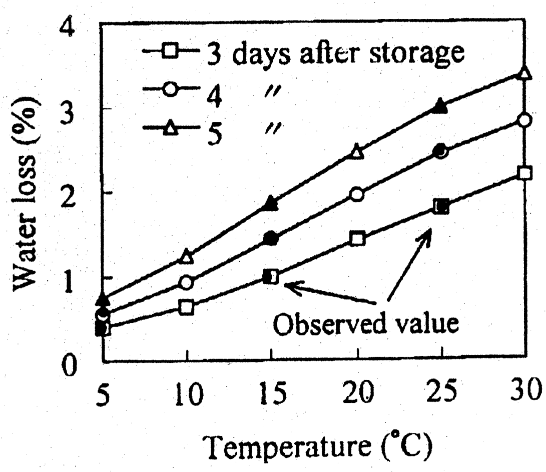


Fig. 9. Estimated relationships between the temperature and the water loss of tomato, which are made using values on 3, 4 and 5 days after storage in their cumulative responses (Simulation)

It can be seen that the estimated values are quite similar to the observed values. Slight non-linearity is observed in the relation between the temperature and the water loss.

CONCLUSION

In this study, a three-layer neural network was applied to the identification of the cumulative responses such as the water loss and redness of tomato, as affected by environmental factors. It is concluded that the neural network approach for the identification of cumulative responses are resulting best performance when the system parameter number n=1, the number of response pattern 3 or more, and adding a linear data to the input variable of the neural network like the Type II neural network. This implies that the type II neural network is effective for the identification of cumulative responses under small data number from the viewpoint of computational time saving. The techniques obtained here can be applicable to a wide variety of identification problems in plant production systems.

REFERENCES

Chen, S., S.A. Billings, and P.M. Grant (1990): Non-linear system identification using neural network. *International Journal of Control,* 51(6), 1191-1214 .

Eykhoff, P. (1974): *System identification.* John Wiley and Sons, London.

Hinton, G.E. (1992): How neural network learn from experience. *Scientific American*, Sept. 1992, 105-109.

Hirafuji, M. (1991): A plant growth model by neural networks and L-system. *Preprints of 9th IFAC/IFORS Symposium Identification System Parameter Estimation,* 605-609.

Honjo, T., T. Okuya and A. Okuya (1992): Estimation of water and nutrient consumption in hydroponic tomato greenhouse. *SHITA Journal,* 3(2), 97-102.

Hsu, K.L., H.V. Gupta and S. Sorooshian (1995): Artificial neural network modeling of the rainfall-runoff process. *Water resources Research,* 31(10), 2517-2530.

Morimoto, T, I. Cho and Y. Hashimoto (1991): Identification of hydroponics in an advanced control system of the greenhouse. *Preprints 9th IFAC/IFORS Symposium Identification System Parameter Estimation,* 610-615.

Morimoto, T., T. Torii and Y. Hashimoto (1995): Optimal control of physiological processes of plants in a green plant factory. *Control Engineering Practice,* 3(4), 505-511.

Purwanto, W., T. Morimoto, Y. Hashimoto. 1996: Simulative Estimation for the Identification of the Cumulative Response of a Plant Using Neural Networks. Journal of SHITA, Vol: 8 (2), pp. 112-118.

Rumelhart, J.E., G.E. Hinton, and R.J. Williams (1986): Learning representations by back-propagation error. *Nature*, 323 (9), 533-536.

Seginer, I. and McClenden (1992): Methods for optimal control of the greenhouse environment. *Trans. of ASAE,* 35(4), 1299-1307.

Thornley, J.H.M. (1976): *Mathematical models in plant physiology.* Academic Press, London.