Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computers and Electronics in Agriculture
40 (2003) 65–85

Computers and electronics in agriculture

www.elsevier.com/locate/compag

Neural network-based detection of mechanical, sensor and biological faults in deep-trough hydroponics

K.P. Ferentinos [a,*], L.D. Albright [a], B. Selman [b]

[a] *Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA*
[b] *Department of Computer Science, Cornell University, Ithaca, NY, 14853, USA*

Abstract

In this work, two separate fault detection models are developed: one for the detection of faulty operation of a deep-trough hydroponic system which is caused by mechanical, actuator or sensor faults, and one for the detection of a category of biological faults (i.e. specific stressed situations of the plants), namely the "transpiration fault". The neural network methodology was proved to be successful in the task of fault detection, in both applications. The general fault detection model was capable of detecting a faulty situation in a very short time, in most cases within 20 or 40 min. In the case of the biological fault detection model, the "transpiration fault" was generally detected within 2–3 h. The results show that both neural networks have useful generalization capabilities. The influences of the conditions of the plants to the measured root zone variables are also investigated and they show that biological faults in general cannot be detected in this kind of cultivation systems using the measurable variables used in this work. The most probable explanation is the inertia of overwhelming mass of the nutrient solution (compared to the mass of the plants), in a way that it becomes a limitation factor to the influence of plants condition to their root zone microenvironment.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Fault detection; Hydroponics; Feedforward neural networks; Biological faults; Backpropagation algorithms

* Corresponding author.
E-mail addresses: kpf3@cornell.edu (K.P. Ferentinos), lda1@cornell.edu (L.D. Albright), selman@cs.cornell.edu (B. Selman).

0168-1699/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0168-1699(03)00012-7

1. Introduction

The goal of every greenhouse facility is to optimize the quantity as well as the quality of production. Optimization is achieved using automated systems in order to control the environment inside the greenhouse so that optimal conditions for the specific cultivated plant are approximated. In addition, hydroponic cultivation gives the opportunity to control root environment precisely and, thus, to have more extensive plant production system control.

Loss of control over a greenhouse is a phenomenon usually having negative effects on the production and, consequently, the profit of the facility. Certain failures (faults) are easily noticed, while others can be difficult to detect. Of course, a feedback-controlled greenhouse may be able to maintain desired conditions even when some parts of the control mechanism are out of order. However, especially when faults concern the plants, the effects may be disastrous for the entire production. Thus, detection and diagnosis of possible faults in that area become very important. In particular, that which is the most important is rapid detection of developing faults. That is, detection at the earliest possible stage of slowly developing faults, as well as quick identification of the problem. In order to enable fast detection, an on-line identification system is necessary.

Detection of a stressed plant is very important for its final production. The motivation to detect faults in the plants themselves was the assumption that stressed plants that react with their microenvironment in a hydroponic system would affect this microenvironment in a detectable way, via the collected measurements. For example, the rate of change of pH conveys the rate of nitrate uptake, which is a measure of growth rate. EC can be used as an estimate of the concentration of nutrient elements in the solution and so as an estimate of the plants' nutrient uptake. The quantity of dissolved oxygen (DO) in the nutrient solution can give a clear indication of the metabolic rate of the plant roots and, hence, their vitality.

Artificial intelligence (AI) has developed to such a degree that it can lead to an intelligent control system capable of self-examination. The combined information of different sensors, through AI methodologies such as neural networks (NNs), can lead to quality classification of isolated information derived from specific sensors or actuators. In this way, a fault detection and diagnosis system could eventually be developed, capable of detecting and identifying specific failures in parts (sensors or actuators) of the hydroponic system, by simply reading the collected measurements of the system sensors and actuators. As an extension of the "speaking plant" approach (Hashimoto, 1989), the hydroponic system, which consists of mechanical and electronic equipment and also of cultivated plants, could be considered as one hybrid (bio-mechatronic) system. In this extent, biological failures, in the sense of stressed situations of the plants, could possibly be detected by a fault detection system, using simply the readings of the above-mentioned measurements.

Behaviors of normal and malfunctioning systems may differ completely and those differences can be detected by measuring the environmental variables of the system. The main advantage of the use of NNs is that values of environmental variables and the corresponding states of the system are unnecessary. Thus, one need not have an accurate model of the hydroponic system, from which several residuals or fault signatures are to be extracted and used for final fault detection. NNs have been proved capable of identifying faults in several complex biological processes (Parlos et al., 1994; Sorsa et al., 1991; Venkatasubramanian and Chan, 1989; Watanabe et al., 1989) in which, neither analytical models nor intermediate residual calculations were used.

2. Materials and methods

Lettuce plants (*Lactuca sativa*, var. Vivaldi) were cultivated in a continuous production deep-trough hydroponic system, consisted of three small growing ponds (tanks). The ponds of this system were filled with nutrient solution (Sonneveld and Straver, 1994) and the plants were placed in specially placed holes on Styrofoam panels that covered the ponds and floated on the surface. Each stainless steel pond had an area of 0.75 m^2 (1.25 m $\times$ 0.60 m). Seeds were sown in small rockwool cubes with volume around 3.5 cm^3, each of which had a small hole (around 0.2 cm^3) filled with peatlite, wherein each seed was placed. Seedlings were transplanted from a growth chamber into the ponds 12 days after sowing. Foam spacers of 2 cm width were used to provide sufficient room for the plants as they grew larger and were added incrementally as plants were larger. Each pond had a range of plant ages ranging from 12-day-old, which was the transplanting age of the plants from the germination chambers to the ponds, to 27-day-old. Every 2 days, the largest plants of the hydroponic system (of 28 days of age) were harvested, the rest of the plants were moved upwards, 12-day-old plants were transplanted from the growth chamber into the ponds and new seeds were sown in the growth chamber. In this way, a “continuous production” type of system was developed, which resembles real-life hydroponic production systems more closely than other techniques that have been used in neural network modeling applications (e.g. Ferentinos and Albright, 2002) and, in addition, it produces nearly constant conditions.

Desired values of the environmental parameters during the operation of the hydroponic system were an air temperature of 24 °C during the day and 19 °C during the night, relative humidity from 30 to 70% and a light integral of 17 mol m^{-2} per day of photosynthetically active radiation. For the nutrient solution, the pH set point was 5.8, the EC was maintained between 1150 and 1250 µS cm^{-1} and the DO was maintained between 6 and 7.5 mg/l.

2.1. Normal and faulty situations

The procedure of training a fault detection NN requires an accurate definition of “normal operation”, defined in our case as unstressed plants in a system that is in control. NNs do not need the existence of some representation of normal operation by means of specific values of the environmental variables in order to learn the pattern. Thus, we need to know only when the system and, in extension, the plants are in conditions considered to be normal by the producers, and also to know which training data sets correspond to those normal conditions. The values of the measured variables mentioned before were considered to be the normal values for the system and the plants were considered to be normal as long as they appeared healthy. We also need to define the "faulty operation" and to categorize this kind of operation into different types of faulty operations, one for each different kind of fault. In order to take data sets for each kind of fault, we have to impose those faults and take the corresponding measurements of the microenvironment variables.

Because NN was trained off-line (meaning that the data sets were first collected and then used for training), there was no way of mistakenly having unhealthy plants in data sets of "normal operation". The "faulty operation" consisted of three different kinds of faults:

Actuator/mechanical faults. These are failures in some actuator or some mechanical part of the hydroponic system. The actuator fault considered was the malfunction of the pH control pump (fault type 1) and the mechanical fault was the malfunction of the nutrient solution circulation pump (fault type 2).
Sensor faults. These are failures in the sensors of the system. The ones considered were (a) pH sensor failure (fault type 3) and (b) EC sensor failure (fault type 4).
Plant (or biological) faults. These are problems in the cultivated plants themselves and are divided into (a) root area faults and (b) shoot area faults.

For the first two categories of faulty operations, several faults were especially imposed on the actuators and sensors of the system in order to train the NN model and investigate its inherent fault detection capabilities. The failures in the pumps were reproduced by turning off the pumps at specific times for some known periods of time. The sensor faults were reproduced by adding a periodically changing noise (sine wave) to the readings of the pH or EC sensors.

For the third category, however, faults were imposed directly on the plants. These faults can be divided into two categories: root zone faults and shoot zone faults. Four different series of experiments were performed:

- The largest plants (of ages of 21, 23, 25 and 27 days) were removed from the ponds and held with their roots exposed to air for periods of 5 min. In this way, possible problems in the root zone of the plants were imitated.
- Several leaves of the largest plants were removed. This action imposed permanent damage on the plants and imitated the effects of major problems in the shoot zone of the plants.
- The canopy of the largest plants was disturbed for intervals of 5 min and slightly damaged. This imitated effects of similar but less influential shoot problems than the previous series of experiments.
- The largest plants (of ages of 23, 25 and 27 days) were covered with transparent plastic bags. This imposed a temporary fault that imitated major problems in the shoot zone.

2.2. General fault detection neural network

The feedforward methodology of NNs was used (a good theoretical reference is Fine, 1999) for the fault detection task. As possible inputs of the NN model, the environmental parameters (air temperature, relative humidity and light intensity), the measurable variables of the microenvironment of the plant (pH, EC, DO, nutrient solution temperature and transpiration rate of the plants) and the control signals of the pH and the DO control schemes (amounts of acid and oxygen added, respectively) were considered. As explained below, biological faults showed no correlation with any measured variable except for the transpiration rate. On the other hand, transpiration rate was not affected by the first two fault categories (mechanical/actuator and sensor faults). Therefore, two separate NN models were developed: one for the detection of mechanical/actuator and sensor faults, and a second one for the detection of biological (or plant) faults.

The first fault detection neural network (FDNN) model was developed in order to detect general failures in the hydroponic system of either mechanical or electronic (sensor) nature. The network inputs were the pH, the EC, the DO and the temperature of the nutrient solution and also the air temperature, relative humidity and light intensity at the level of the plants, together with the control signals of the pH and the DO. In addition to these inputs, one-step and two-step histories of the pH, EC and DO variables were included. That is, for each of these variables, three inputs existed: one for time t (current time), one for time $t-1$ (previous time step) and one for time $t-2$ (two time steps before). Thus, the total number of inputs was 15. The time step was 20 min. The network had a single binary output that represented either normal (0) or faulty (1) situation. That is, all mechanical and sensor faults were treated as a general “faulty operation”. (A model that, in addition, performs isolation of each of those faults, is presented in Ferentinos and Albright, 2003).

Several different architectures of one-hidden-layer and two-hidden-layer networks were tested, with two different activation functions (logistic and hyperbolic tangents). The training methodology was the Backpropagation Training Algorithm (Rumelhart et al., 1986). Four different multidimensional minimization algorithms (steepest descent, conjugate gradient, quasi-Newton and Levenberg–Marquardt algorithm) were tested. An on-line adjustable learning rate performed better than a constant one. In the steepest descent and the conjugate gradient algorithms, the Hessian was used at every iteration to solve for the “best” learning rate. For the other two algorithms, the “best” learning rate was calculated with an approximate line search using a cubic interpolation. The final NN fault detection system was tested using new data and its generalization capabilities were explored.

2.3. Biological fault detection neural network

Contrary to the assumptions and expectations mentioned in Section 1, the first three series of biological faults showed no detectable effect on the measurable variables. Not even the transpiration rate, a variable strongly linked with plants’ physiological condition, was noticeably affected. In the case of the first fault type, this can be explained by the large buffering capacity of the nutrient solution, which masks the effects of the fault. In the case of the second fault type, probably the transpiration rate was not influenced because of its compensation by the loss of plant sap from the cut surfaces of the plants. Finally, in the case of the third fault type, it seems that the disturbance of the plants' canopy was not sufficient to influence the system's measured variables. The large buffering capacity of the nutrient solution was again the major factor.

The fourth fault type, covering of the plants with transparent plastic bags, showed some effect only in the transpiration rate (for this reason it is referred as "transpiration fault" from now on). The other variables of the system were not affected. Transpiration rate, even though it was highly correlated with the environmental conditions of the greenhouse (temperature, light intensity and relative humidity), gave a clear indication of the existence of stressed plants. In order to show the effect of this fault to the transpiration rate, a series of graphs is presented. In these graphs, the cumulative differences of evapotranspiration rates of the ponds of the hydroponic system are graphed over time. In the cases of fault existence, one pond maintained normal conditions while the other contained the stressed plants. The evapotranspiration rate was evaluated by continuously measuring the weights of the ponds.

Fig. 1a shows the cumulative evapotranspiration rate differences between two ponds, say "pond 1" and "pond 2", that both have normal plants. Normally, the differences should be constant and not increasing. However, because of the air flow above the ponds, if the plotted cumulative rates are the differences in the form "pond 2 minus pond 1", then "pond 2" always has larger evaporation than "pond 1". All the experiments shown in this section were performed under these air flow conditions in the greenhouse. A change of the air flow pattern after data collection proved that the evapotranspiration rate differences were caused by the air flow above the ponds. In all graphs, cumulative values were initialized to zero each time a transplanting took place in the hydroponic system (every 2 days). The main reason, besides the convenience of having smaller values of cumulative rates for easier interpretation and comparison, was that during transplanting, the ponds were manually disturbed and the measured weights during those periods did not represent the real weights of the ponds.

Fig. 1b shows the cumulative evapotranspiration rate differences when the second fault type (removed leaves) was imposed at around time interval No. 2000. At that point, half of the leaves of the largest plants of "pond 1" were removed, while nothing changed in "pond 2". The normally expected result of reduced transpiration in "pond 1", would be an increment of the difference between transpiration rates of "pond 2" and "pond 1", which clearly did not happen. A period of 3 days after the occurrence of the fault is shown in the graph, and nothing changed in the usual differences between the two ponds. All curves in Fig. 1b are very similar to the ones in Fig. 1a and in both plots the maximum cumulative differences do not exceed 1.5 kg at the end of the 2-day periods. As mentioned before, this behavior is explained if the additional evaporation from the cuts of the removed leaves is taken into account.

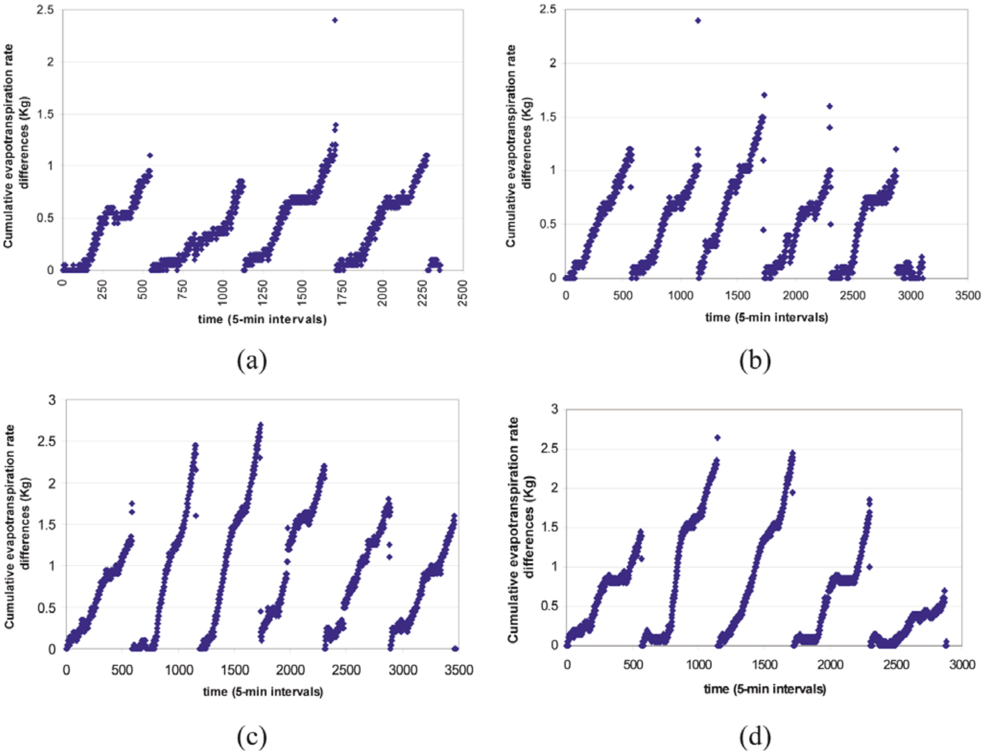

Fig. 1. Cumulative evapotranspiration rate differences between: (a) two ponds in normal operation, (b) a pond in normal operation and a pond with fault type 2 (removed leaves)—fault was imposed around interval No. 2000, (c, d) a pond in normal operation and a pond with fault type 4 (covered plants) in both graphs—faults were imposed around interval No. 1000 (c) and interval No. 900 (d).

Thus, no indication of the fault appeared in this case. Similar results were obtained by the imposed faults of removing the largest plants from the tanks for periods of 5 min or disturbing their leaves (fault types 1 and 3, respectively).

In Fig. 1c and d, two different experiments of imposing the "transpiration fault" are presented. As can be seen from these graphs, the evapotranspiration rate differences between normal plants and stressed plants increase in quite a large degree during the faults, as they reach or even exceed the value of 2.5 kg in most cases. This increment was larger during the first days because during that period more plants were covered with plastic bags. One day after the initiation of the fault, one series of stressed (covered) plants was harvested and after that, every 2 days, one more series was harvested, up to a period of 5 days from the beginning of the fault, when all stressed plants had been harvested and evapotranspiration returned to normal values. In Fig. 1c, the fault was imposed at interval No. 1000, while in Fig. 1d at interval No. 900.

In summary, the investigation of the influence of biological faults in the main measurable variables of the hydroponic system, i.e. pH, EC, DO and transpiration rate, showed that all tested faults had no detectable effect on these variables, except for the fault that consisted of covering the plants with transparent plastic bags ("transpiration fault"), which gave a significant influence in the transpiration rate of the plants. Based on this fact, a biological fault detection neural network (BFDNN) model was developed, in order to make this biological fault detectable during the operation of the hydroponic system.

As in the general fault detection model, the feedforward methodology of NNs was used for this model. All the analysis previously described used comparison of the investigated pond of the hydroponic system with another pond, which was known to be under normal operation. Obviously, if we need to construct a fault detection system for real-life application, the existence of such an "ideal" pond for reference cannot be assumed. Thus, the constructed NN model used only data from one pond and no comparisons with another pond. The inputs to the network were the weight of the pond and the environmental variables of the greenhouse, air temperature, light intensity at the level of the plants and relative humidity. In addition, the temperature of the nutrient solution was included as input. The single output of the network represented either normal (0) or faulty (1) operation.

The biological fault considered before, i.e. the "transpiration fault", is by nature very slowly developing, compared to the other types of faults considered here. In addition, the real effect of the fault is not on the weight itself but rather on the rate of weight change. Thus, a sufficient amount of weight history values is required by the NN for the extraction of the desired results, which is the detection of the fault in a reasonable time. The exact amount of that history, how many past weight values are going to be used, cannot be known a priori. Several sizes were tried. More specifically, networks with 5, 10, 15 and 20 inputs of weight values were tested. Thus, the NNs had a total of 9, 14, 19 and 24 inputs, respectively, including air temperature, light intensity, relative humidity and temperature of nutrient solution.

2.4. Training of the general FDNN

The general fault detection NN model was trained with experimental data collected from all three ponds of the system, for both normal and faulty situations. Approximately 15 (non-continuous) days of data formed the final training set. The time step of these data was 20 min. Thus, the training set was based on 1050 entries for each input of the neural network.

The training process included two basic parts. The first part, the preliminary training process, determined the best combination of network architecture and training algorithm. Several candidate network topologies were trained (both 1-HL and 2-HL networks) with all four training algorithms and the results were compared. The second part of the training process, which was the basic training process, focused on training the best combination of architecture and algorithm. Based on the preliminary training, the network architecture/algorithm combination that gave the best results was the 2-HL NN with nine nodes in the first hidden layer and nine nodes in the second hidden layer, trained with the quasi-Newton algorithm.

The basic training process had a goal to further train the selected NN with the best possible algorithm for this system and architecture, which was proven to be the quasi-Newton algorithm. Many different random initial network parameters were tested in that training. Also, several values of the coefficient of the penalty term for the regularization (λ), varying by the order of 5, were tried. Both logistic and hyperbolic tangent activation functions (functions of hidden nodes) were tested. The results of these tests showed that the value of λ that leads to the minimum sum squared error (SSE) was $\lambda = 0.05$ and that the network with logistic activation functions performed better than the one with hyperbolic tangent activation functions. Thus, the final NN model consisted of a network with 15 inputs, two hidden layers with nine nodes each that had logistic activation functions, and one output. The preferred training algorithm was the backpropagation training algorithm using the quasi-Newton multidimensional minimization algorithm with parameter $\lambda = 0.05$.

2.5. Training of BFDNN

Similar to the training process described before, a preliminary training process of the BFDNN took place in order to determine the best combination of network architecture and training algorithm, by training several candidate network topologies (both 1-HL and 2-HL networks) with all four training algorithms (steepest descent, conjugate gradient, quasi-Newton and Levenberg–Marquardt) and comparing the results. Here, the additional parameter of the number of inputs, as explained before, was considered. The rest of the training process focused on training the best combination of architecture and algorithm. Many different random initial network parameters were tested with weights ranging in the area $[-1/d, 1/d]$, where d is the number of network inputs (Duda et al., 2001). Because of the large amount of data and the complexity that the varying number of inputs incorporated, the computational power consumption factor became a limitation. Thus, only logistic activation functions were used, as they seem to be more successful than hyperbolic tangent in most relevant applications, and also the penalty term for the regularization (λ) was kept constant. These parameters were investigated later, during the final training of the network. The output node had a linear summation function.

The training set consisted of data collected during four repetitions of the experiment of covering the plants with transparent plastic bags, as well as normal data collected during normal operation of the system. The experiments of "faulty situation" took place during the period from February 2001 to the end of April 2001. The total amount of data summed up to around 17.5 days. The time step used was 10 min. Thus, part of the training set with faulty data consisted of 2500 entries (columns in matrix form). 17.5 days of normal data were also used, from the afternoon of March 25, 2001, until early in the morning of April 12. Thus, the entire training data set consisted of 5000 columns. Each column was of the following form:

$$[\mathrm{weight}(t) \quad \mathrm{weight}(t-1) \quad \mathrm{weight}(t-2) \quad \cdots \quad \mathrm{nutr}_{\mathrm{sol}}\mathrm{temp}(t) \quad \mathrm{air}_{\mathrm{temp}}(t) \quad \mathrm{light}(t) \quad \mathrm{rel}_{\mathrm{hum}}(t)]^{\mathrm{T}}.$$

The number of weight values that were included depended on the network that was used, as it was explained before. In addition, the output element of the network was included at the bottom of each column. Thus, the first half of the columns had the value 1 as the last element, while the rest half columns the value 0.

The results of the preliminary training with all four training algorithms, with the exception of the conjugate gradient algorithm in the case of 1-HL networks, showed that the fourth NN ("NN4") that has the largest history of weight values, outperformed the other three in most cases. More specifically, the performance achieved by all four networks was proportional to the size of the weight values history. In most cases of different numbers of nodes in the hidden layer, NN4 outperformed the other networks, while in general, the performance gets worse as weight values history decreases (NN3 → NN2 → NN1). Several 2-HL network architectures were also tried. Like in the case of 1-HL architectures, in general, the performance was proportional to the size of the weight values history; thus NN4 gave the best results in most cases.

The NN that gave the best performance in the preliminary training explorations was a 2-HL network with 24 inputs (NN4), with 14 nodes in the first hidden layer and seven nodes in the second hidden layer, trained with the quasi-Newton algorithm. This network was further trained for 1000 iterations. Again, several initial random weights in the range $[-1/24, 1/24]$ were used, in order to avoid local minima. Several values of the penalty term for the regularization (λ) as well as of the other algorithm parameters were tested. In addition, both logistic and hyperbolic tangent activation functions were used. The best performance was given with $\lambda = 0.1$ and logistic activation functions. The mean-squared error of the final NN was equal to 0.0984.

3. Results

3.1. Testing results of the general FDNN

Testing consisted of presenting new data to the trained NN model and exploring its generalization capabilities. The main goal in a fault detection system is not only detection of the existence of a fault, but also its rapid detection. Especially, when we deal with slowly developing faults, the time factor becomes more important because these faults are more difficult to detect as they begin. Six testing data sets were presented to the NN fault detection system. These data sets are described in Table 1. Each set starts with data of normal operation and some specific fault is imposed at some known time, except for the last data set that contains only normal data. The faults, after their introduction, last up to the end of the testing tests.

The output of the NN was considered to represent a faulty operation if it had a value greater than 0.5, while for values smaller than 0.5 a normal operation was assumed. Normal operation of the pH control pumps consists of temporal instantaneous operations of the pumps every time the pH control scheme of the hydroponic system requires the addition of acid in the nutrient solution. The

Table 1
Description of testing data sets for the general FDNN

Data set	Fault type	Beginning date/time	End date/time	No. of 20 intervals	Interval that fault was introduced
1	1	November 8, 10:00 A.M.	November 10, 11:50 A.M.	150	16
2	2	November 18, 8:30 A.M.	November 18, 6:30 P.M.	32	18
3	2	December 5, 11:10 A.M.	December 5, 5:25 P.M.	19	9
4	3	November 30, 11:00 A.M.	December 2, 6:40 P.M.	167	16
5	4	December 11, 8:00 A.M.	December 12, 7:10 A.M.	69	16
6	No fault	November 14, 5:00 P.M.	November 15, 2:20 P.M.	64	–

amounts of added acid are quite small compared to the mass of the nutrient solution, but the change of pH is performed rather fast, usually in a couple of minutes. On the other hand, the circulation pumps operate continuously, because their duty is to continuously circulate the nutrient solution of the ponds.

Fig. 2 shows the NN output on the first data set. The "pH control pump out of order" fault was introduced at the 16th interval. The fault was detected by the network within only two step intervals (point 18), i.e. a period of 40 min. After that the network gives a steady and strong indication that the operation is not normal, with values very close to 1.

Fig. 3 shows the network output for the second testing set, which contains a "circulation pump out of order" fault. The data again start with normal operation and the fault is introduced in the 18th interval. It is clear here that this kind of fault is detected very fast. Even from the 18th data point, the output of the network is 0.69, which is considered as a weak fault indication. The next data point gives a value of 0.91 that strongly indicates the existence of faulty operation. A disadvantage here can be considered the fact that the output drops below 0.5 (normal operation), 40 min after the occurrence of the fault and stays in that area for an hour. After that period, it returns to the faulty indication. This is not something important, if we consider the fact that the indication of some fault is already given at a very early stage and also that the kinds of faults considered are supposed to be irreversible without the interaction of a human factor with the system. The third testing set contains the same fault type as the previous one that is now introduced in the ninth data point. This time it takes 20 min to the network to indicate a possible fault (value

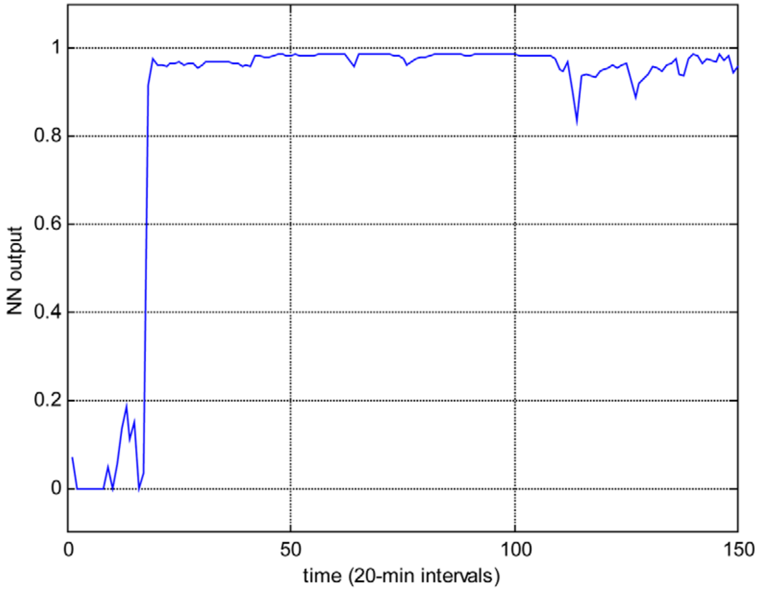

Fig. 2. Output of the FDNN for testing data set 1 (fault type 1); fault introduced at time step No. 16.

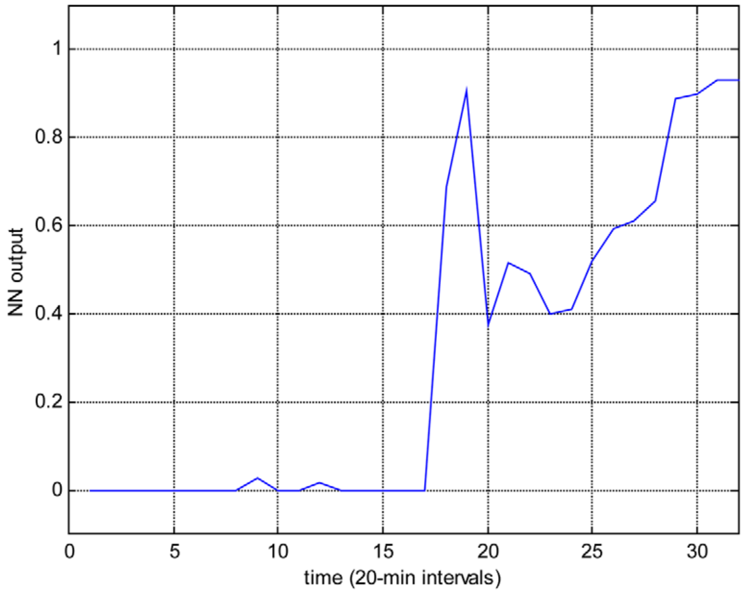

Fig. 3. Output of the FDNN for testing data set 2 (fault type 2); fault introduced at time step No. 18.

0.64) while in a total of 1-h period after the introduction of the fault, the output becomes high enough (0.82) to strongly indicate the existence of faulty operation.

The fourth testing data set has a total of 167 20-min intervals and the “failure in pH sensor” fault was introduced in the 16th point. The output of the NN model is presented in Fig. 4. Similar to the previous case, it takes one step (20 min) for the network to indicate a possible fault with an output of 0.60, while at the next 20-min step the output becomes 0.91 and stays in that area. The rather periodical fluctuations of the output are caused by the nature of the sensor fault. As mentioned in Section 2.1, this kind of fault was reproduced by adding a periodically changing noise to the readings of the pH sensor. The form of noise is a sine function. Thus, in the points where the noise becomes small, the fault confidence of the network decreases. However, this does not cause the network to exit the “faulty situation” area.

The fifth testing data set contains the implementation of the “failure in EC sensor” fault. The set has 69 20-min intervals and the fault was introduced in the 16th interval. As can be seen in Fig. 5, the fault is detected 4 h after its beginning. Moreover, several hours later, when the noise of the sensor failure becomes small, the network indicates normal operation for that period. It seems that this specific fault causes some problems for the detection process, probably because no information about the control signal for the EC is present; EC was controlled manually in the hydroponic system.

Finally, the last testing data set contains only normal operation data and it was used to check the network ability to continuously recognize normal operation. The

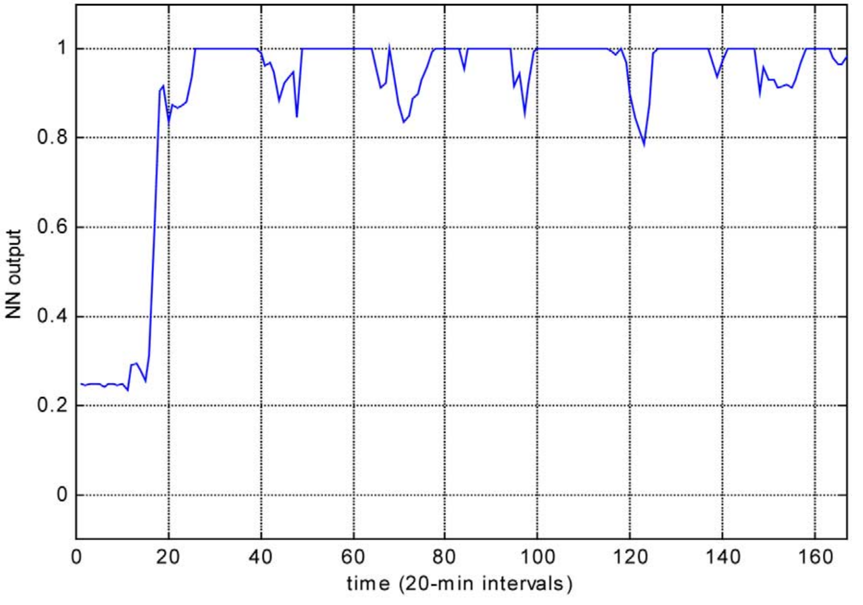

Fig. 4. Output of the FDNN for testing data set 4 (fault type 3); fault introduced at time step No. 16.

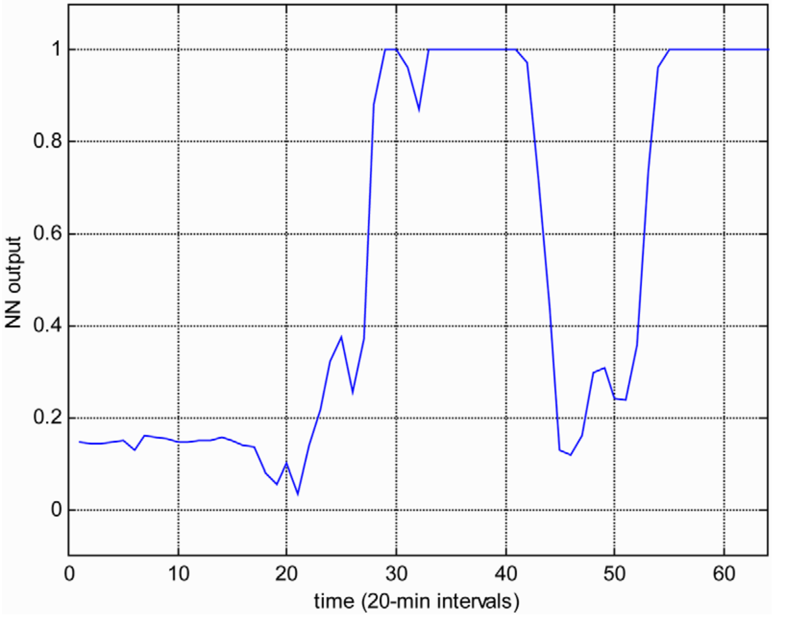

Fig. 5. Output of the FDNN for testing data set 5 (fault type 4); fault introduced at time step No. 16.

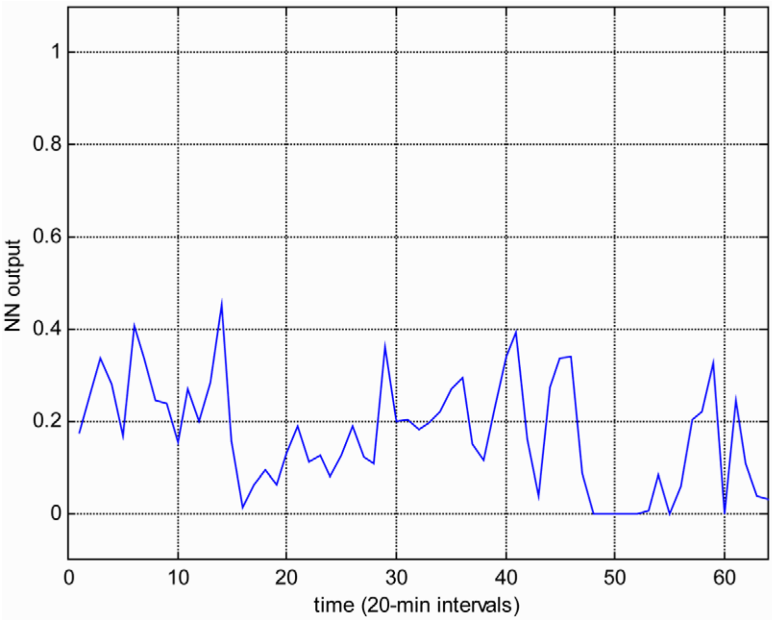

Fig. 6. Output of the FDNN for testing data set 6 (normal operation).

output of the network is shown in Fig. 6. The graph shows a period of almost one whole day and the output is always below 0.4, indicating normal operation during the entire period.

3.2. Testing results of BFDNN

Three testing data sets were constructed for the testing of the developed NN model. The first two sets contained a “transpiration fault” and were used to test the rapidness of the detection of the fault by the model, while the third one contained only normal data and was used to test the performance of the system during normal operation and the probabilities of false alarms. More details about the testing data sets are shown in Table 2.

Table 2
Description of testing data sets

Data set	Fault	Beginning date/time	End date/time	No. of 10 intervals	Interval that fault was introduced
1	Yes	March 6, 7:50 A.M.	March 11, 11:55 P.M.	820	30
2	Yes	May 23, 7:00 A.M.	May 28, 12:45 P.M.	760	30
3	No	April 19, 6:00 A.M.	April 22, 1:00 P.M.	474	–

The NN output was trained to be 0 during normal operation and 1 during faulty operation. It was a continuous output with values from 0 to 1, thus some thresholds had to be implemented in order to decide below which value normal operation is assumed and above which value a fault is detected. These output thresholds were set to 0.4 and 0.6, respectively. When the output was between these two values, then the system could assume that neither normal nor faulty operation exists, i.e. an unknown condition with a possibility of fault. An even more sophisticated technique could use the logic that the upper threshold is used for shifting from normal to faulty operation and the lower limit is used for shifting from faulty to normal operation. That is, the interval [0.4, 0.6] holds the previously known condition of the output decision. This was the approach used to extract the following conclusions.

Fig. 7a shows the NN output during the first testing data set. The time step is 10 min. It can be seen from the graph that even though the fault is detected, there are large intervals that the output returns back to normal (below 0.4). This happened twice during the main period of the fault and it also happened toward the end of the fault (after interval No. 720) where the fault has started to vanish, as most of the covered plants have been harvested and only one series of plants remains covered. Therefore, the operation has already started to become normal again. These can be better seen in Fig. 7b, where the hourly averages of the NN output are shown, for the entire data set. The fact that the system mistakenly gives a normal operation indication at some periods during the occurrence of the fault is not considered to be important, because the indication of the fault has already been given for a satisfactorily long time. The nature of the fault is such that usually it cannot be reversed without human intervention; therefore, after the fault has been detected, it is not important if the system fails to continuously detect it. The important matter in the detection of a fault is the rapidness of detection. The fault started at interval No. 30. It is detected by the NN after 3 h and the output stays above the lower threshold (0.4), giving a strong indication that a fault has occurred.

In Fig. 8a, the NN output during the second testing data set is shown. As in the case of the first testing set, even though the fault is detected, there are quite large intervals that the output returns to normal (below 0.4). This can also be seen in Fig. 9b, where the hourly averages of the network are shown. Some of these mistakes in classification are probably caused by the disturbances of the ponds during the transplanting routine every 2 days. Some others simply reflect the inability of the network to correctly classify the situation in the “faulty” area. As mentioned before, because of the fact that the fault is irreversible on its own, correct detection at the earliest possible stage after its occurrence is the important factor. The fault started at interval No. 30. It is first detected by the NN after 2 h but the confidence of the network is not very strong, as the output oscillates between the two thresholds and even drops below the lower threshold. However, the fault indication becomes strong 1.5 h after its first detection, that is 3.5 h after the fault occurrence (around interval No. 50), when the output goes again above the upper threshold (0.6) and stays there.

Finally, the NN model was tested during normal operation of the system, for three and a half days (third data set). Fig. 9a shows the network’s output in 10 min

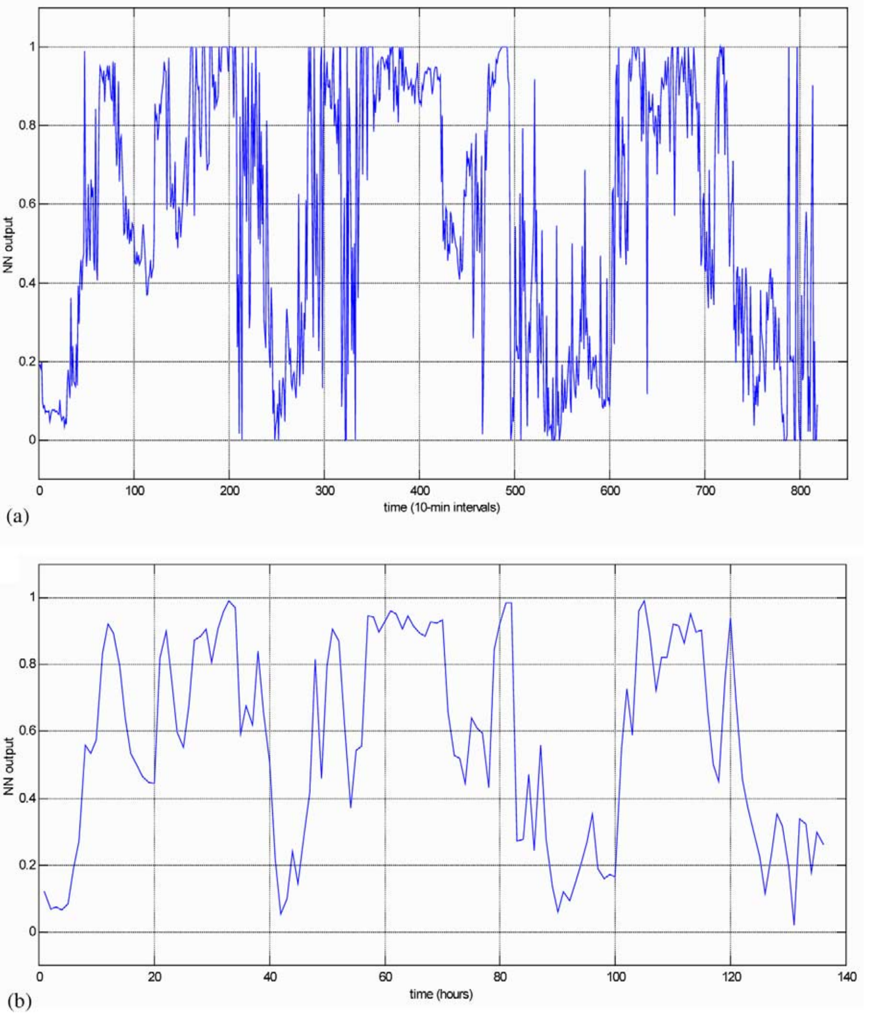

Fig. 7. BFDNN output during the first testing data set (a) and hourly averages of the output (b).

intervals during the entire set and Fig. 9b shows the hourly averages of the output. Generally, the model's output stays in the area of normal operation, below 0.4. There are only four cases of false alarm, which are very short in duration, especially three of them. The percentage of false alarm in the entire data set is 2.11%. As can be seen in the graph of the hourly averages of the output (Fig. 9b), the three of the four events of false alarm are removed with this kind of filtering, dropping the percentage of false alarm to 1.26%. However, averaging the output NN, while decreasing the possibility of false alarm, makes the system slower in detecting a possible fault.

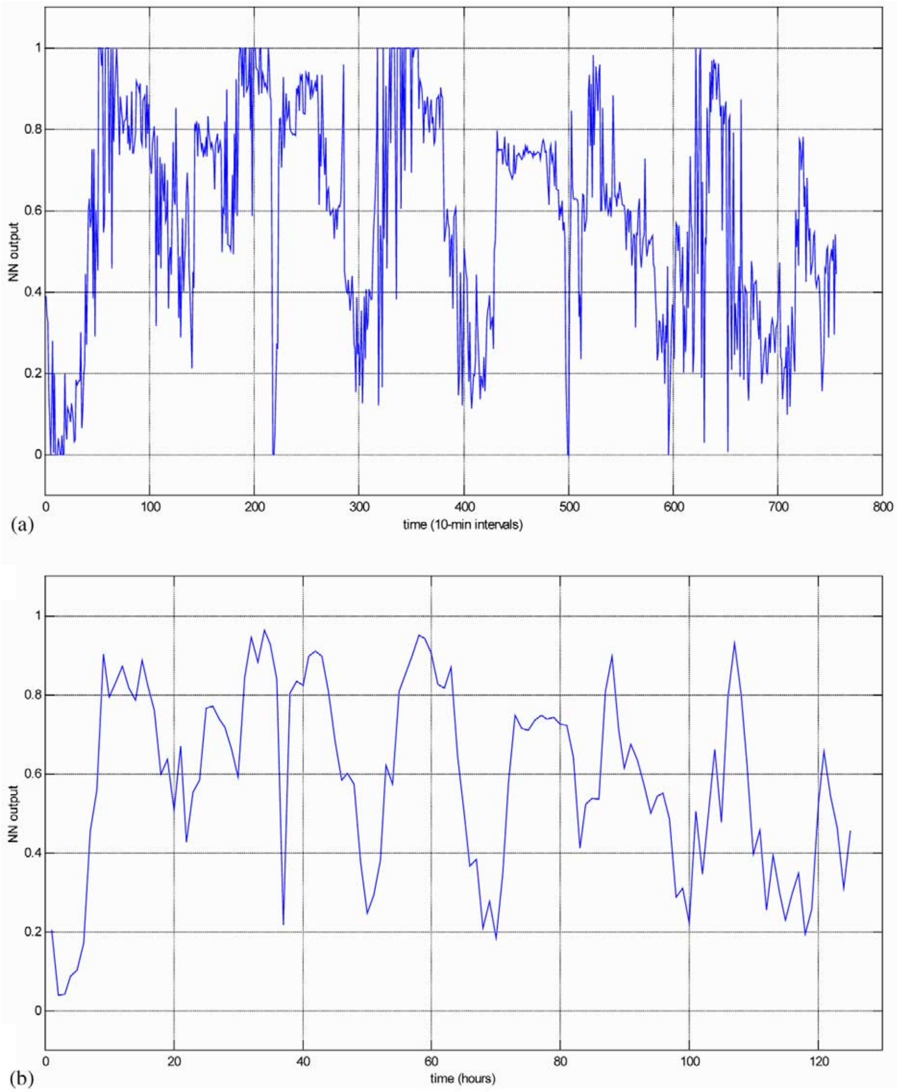

Fig. 8. BFDNN output during the second testing data set (a) and hourly averages of the output (b).

4. Discussion and conclusions

In this work, two separate fault detection models were developed: one for the detection of faulty operation of the hydroponic system which was caused by mechanical, actuator or sensor faults, and one for the detection of a category of biological or plant faults, namely the ''transpiration fault''. The NN methodology proved to be successful in the task of fault detection, in both applications. The

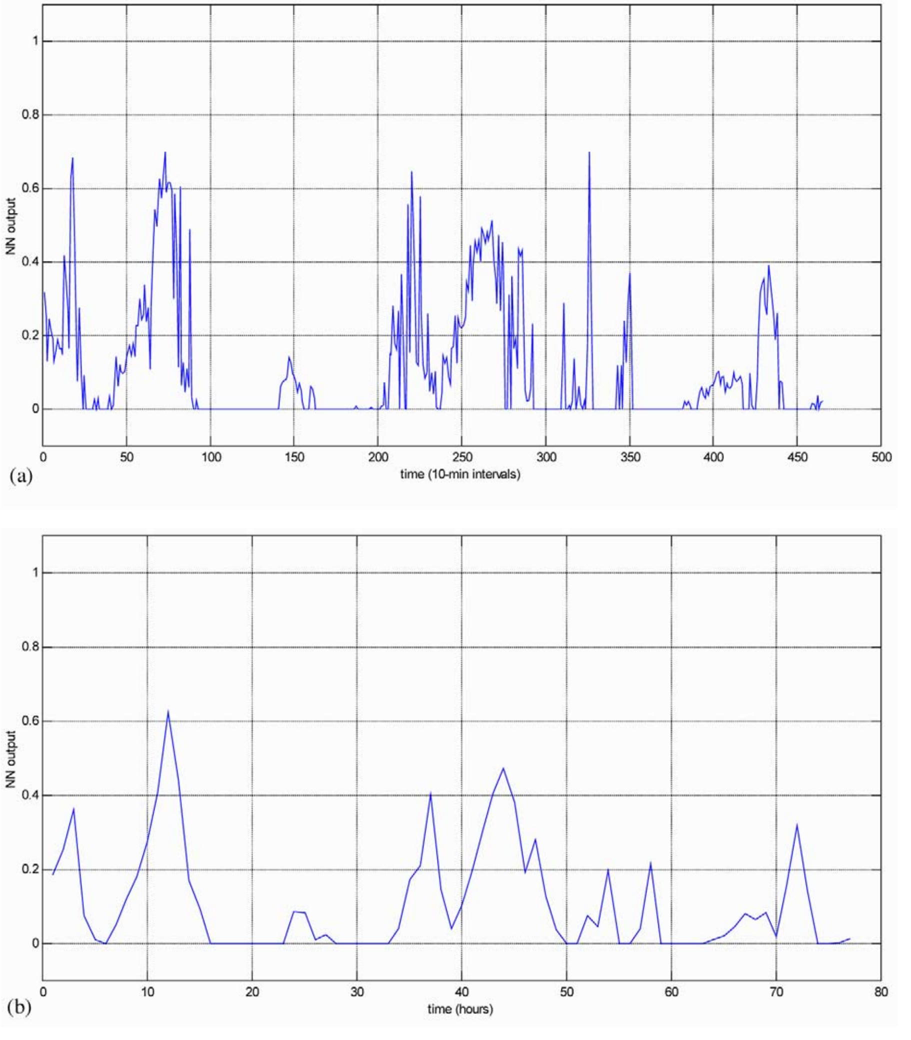

Fig. 9. BFDNN output during the third testing data set (a) and hourly averages of the output (b).

general fault detection model was capable of detecting a faulty situation in a very short time, in most cases within 20 or 40 min. In the case of the biological fault detection model, the ‘‘transpiration fault’’ was generally detected within 2–3 h, which is a reasonable time if one thinks the slow nature of biological interaction between the plants and their environment. The results showed that both NNs had useful generalization capabilities.

The exploration of the influences of biological faults to the hydroponic system variables showed that the major system variables used in the general fault detection model, pH, EC and DO, were not influenced by the imposed biological faults, i.e. by the stressing of the plants in various ways. This was something unexpected, considering the dynamics of the root microenvironment of the plants in a deep-trough hydroponic system and also the known high influence relation between root zone conditions and the condition of the plants. If we assume that the specific variables measured in the hydroponic system are a representation of the root zone microenvironment, it seems that the interaction between the microenvironment and plants is highly biased toward the direction that leads to the plants. The condition of the plants does not seem to influence the measured variables of the microenvironment. The only variable that was influenced to a measurable extent was the evapotranspiration rate. Evapotranspiration measurements could be related to a specific biological fault. However, this variable is a combination of evaporation of the nutrient solution, which is a physical process, and transpiration of the plants, which is a biological process. Moreover, even this variable was not affected by the rest of the faults imposed on the plants. The most probable explanation is that the effect on the transpiration rate was too small for differences to be detected by the measuring system. Another possible reason is that the transpiration rate was measured together with evaporation (evapotranspiration rate), and the latter was covering any small trends or fluctuations of the former.

The conclusion of the explorations of the influences of biological faults on the system measurable variables was that biological faults in general cannot be detected in this kind of cultivation systems using the measurable variables that were used in this work. The interaction between plants and their root zone microenvironment is not equally balanced, as the condition of the plants is highly influenced by the conditions in their root zone microenvironment, while these microenvironment conditions are not influenced in the same degree by the conditions of the plants. The most probable explanation of this property is the inertia that the overwhelming mass of the nutrient solution (compared to the mass of the plants) produces, in a way that it becomes a limitation factor to the influence of plants condition to their root zone microenvironment.

The successful detection of the “transpiration fault”, even though it is understandable that that fault is a severe effect to the plants and represents an extreme, rather unrealistic situation, is motivation for further research in the area of biological fault detection, possibly using additional measured variables of the plant microenvironmental conditions like, for example, nitrate concentration of the nutrient solution. Another approach could be the application of this kind of detection system to different types of hydroponics, such as nutrient film technique (NFT) systems. In NFT, the plants influence the nutrient solution to a greater extent because there is significantly less nutrient solution mass per plant and the buffering factor is less. Finally, an approach that could lead to the desired detection of biological faults could be the combination of the proposed methods with the use of more “plant-oriented” measurements that could more precisely represent the actual conditions of the plants than the measured variables of this work.

Acknowledgements

The helpful advice of Prof. Norman R. Scott of the Department of Biological and Environmental Engineering at Cornell University, is gratefully acknowledged.

References

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, 2nd ed. Wiley, New York.

Ferentinos, K.P., Albright, L.D., 2002. Predictive neural network modeling of pH and electrical conductivity in deep-trough hydroponics. Trans. ASAE 45 (6), 2007–2015.

Ferentinos, K.P., Albright, L.D., 2003. Fault detection and diagnosis in deep-trough hydroponics using intelligent computational tools. Biosystems Eng. 84 (1), 13–30

Fine, T.L., 1999. Feedforward Neural Network Methodology. Springer, New York.

Hashimoto, Y., 1989. Recent strategies of optimal growth regulation by the speaking plant concept. Acta Hort. 260, 115–121.

Parlos, A.G., Muthusami, J., Atiya, A.F., 1994. Incipient fault detection and identification in process systems using accelerated neural network learning. Nucl. Technol. 105, 145–161.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

Sonneveld, C., Straver, N., 1994. Nutrient solutions for vegetables and flowers grown in water or substrates. Proefstation voor de bloemisterij en glasgroente, vol. 8. Naaldwijk, The Netherlands.

Sorsa, T., Koivo, H.N., Koivisto, H., 1991. Neural networks in process fault diagnosis. IEEE Trans. Syst., Man, Cybern. 21 (4), 815–825.

Venkatasubramanian, V., Chan, K., 1989. A neural network methodology for process fault diagnosis. AIChE J. 35, 1993–2002.

Watanabe, K., Matsuura, I., Abe, M., Kubota, M., Himmelblau, D.M., 1989. Incipient fault diagnosis of chemical processes via artificial neural networks. AIChE J. 35, 1803–1812.