Project title:	Improving the accuracy of tomato yield prediction through the use of Intelligent Systems techniques
Project number:	CP 68
Project leader:	Evor Hines, School of Engineering University of Warwick
Report:	Annual report 2011
Previous report:	N/A
Key staff:	Steve Adams
Location of project:	Warwick University
Industry Representative:	Paul Howlett, Wight Salads Group
Date project commenced:	February 2010
Date project completed (or expected completion date):	January 2013

DISCLAIMER:

The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

[Name]
[Position]
[Organisation]

Signature .... Date ....

[Name]
[Position]
[Organisation]

Signature .... Date ....

Report authorised by:

[Name]
[Position]
[Organisation]

Signature .... Date ....

[Name]
[Position]
[Organisation]

Signature .... Date ....

CONTENTS

GROWER SUMMARY

Headline

Intelligent Systems techniques have the potential to predict weekly tomato yields accurately

Background

Supermarkets require reliable supplies of high quality fruit in agreed quantities. At present, however, growers lack the ability to be able to predict their tomato yields with a high degree of accuracy, or to manipulate their pattern of yield to a pre-determined schedule. Tomato yields can vary from week to week, therefore, accurate predictions are vital to enable any shortfall in supply to be sourced in advance, or in the case of a surplus for a promotion to be organised or an alternative market to be found.

While there are a number of tomato models they tend to focus on predicting photosynthesis and total yield, they perform poorly when trying to predict weekly yields. Some commercial software was developed (by Letgrow.com) which aimed to address this issue, although it had limited success and is no longer available.

The aim of this study was to accurately determine greenhouse tomato weekly yields based on studying the environmental variables (temperature, solar radiation, CO_2 concentration and vapour pressure deficit) within the same greenhouse. The work used a prediction tool applied to these variables for process control purposes. The focus was on Intelligent Systems (IS) techniques, in this case an artificial neural network (ANN) which is a group of artificial neurons which can be connected together to perform complex tasks.

Summary

A four-year ‘dataset’ comprising weekly tomato yields and daily internal glasshouse environmental measurements from a commercial nursery was used to design and train a model predicting weekly tomato yields. Another dataset with similar climatic variables and weekly yields from experimental trials at Warwick HRI was used to validate the model.

Current experimental results suggest that the developed Intelligent System can play a role in the accurate prediction of tomato weekly yields. Although there were some differences among tomato cultivars, the IS technique was robust in dealing with imprecise data, and has a learning capability when presented with new scenarios.

A further look at other IS techniques will be done in future, such as Fuzzy logic algorithm, Fuzzy Neural Network, GNMM algorithm, and others. These will be investigated to explore the improvement that could be achieved on the prediction process for tomato yields, and the enhancement of the prediction process accuracy.

Action Points for Growers

None to date.

SCIENCE SECTION

Introduction

A reliable supply of high quality food in agreed quantities is an essential element of the modern food supply chain with impact on growers, supermarkets and ultimately the consumer. This is a particular challenge when the crop is relatively susceptible to harsh conditions as is the case with tomatoes. Growers have increased food quality and yield in many parts of the world through the use of greenhouses, where the environmental conditions can be controlled, and by selecting better cultivars Adams and Valdes (2002). However, weekly yields can fluctuate and this can pose problems of both over-demand and over-production in the absence of accurate yield predictions. In this respect growers and scientists seek ways to forecast tomato yield so as to be able to plan greenhouse operations, marketing and so on so to deliver, *inter alia,* cost reductions and thus increased profits Santos *et al*. (1992).

A large number of prediction models and prototypes have been developed in the past few decades, making use of specialist knowledge of tomato physiology and growing conditions. However, although they tend to deal accurately with total yields they are poor at accommodating weekly yield fluctuations Adams and Valdes (2002).

Two of the most popular physiological process based growth models are TOMGRO Jones *et al*. (1991) and (1999) and TOMSIM Heuvelink (1999), which model growth and yield as a function of climate and physiological parameters. The use of these models is limited by their complexity, especially for practical application by growers and it is difficult to obtain the initial conditions for the parameters required for implementation Heuvelink (1999).

Mathematical model based prediction systems consider parameters such as greenhouse temperatures, radiation levels, CO_2 and vapor pressure deficit (VPD) as the input variables to the system. They then process the available data using the model and generate an estimate of tomato yield as the output. For example, Kano and Bavel (1988) developed a deterministic model based on a photosynthesis equation and a carbon accumulation model. This took the concentration of CO_2, the temperature and the light level as inputs to calculate tomato yield, where their results delivered a weekly yield prediction accuracy of 69%.

Climate change and environmental factors may greatly affect agricultural production in the world, thus a great deal of research has been devoted to studying the potential impacts of climate on crop production within greenhouses. In most such studies, crop management techniques were employed to investigate the cause of crop and yield fluctuation on a weekly timescale, this includes managing growth factors so that they are available to the plant at the right time in the desired amounts as yield variability may be caused by changes of these factors Heuvelink (1999).

Tomato growers may be contracted to sell agreed quantities of produce to supermarkets. However, tomato yields often vary from week to week, and so the ability to accurately predict future yields would give them a competitive advantage. For example if a grower is able to predict that they will to have insufficient yield in a given week, they could source additional produce from elsewhere. Conversely, if they are predicted to have excess yield they could look for alternative markets or arrange promotions to compensate. As a result there have been considerable research efforts to develop accurate tomato yield prediction systems Mukherjee and Sastri (2004).

The approach here is based on a prediction tool applied to a number of variables for process control purposes. The focus in this work is on Intelligent Systems (IS) techniques, in this case an artificial neural network (ANN) Hashimoto (1997), Morimoto *et al.* (1996), which is a group of artificial neurons which can be connected together to perform complex tasks. These methods are particular useful when the data are composite and multivariate Chen *et al*. (2008).

The aim of this study was to accurately determine greenhouse tomato yield based on studying the environmental variables within the same greenhouse. The system we developed takes into account various growing conditions and factors that may affect plant development. Environmental conditions that are understood to influence the growth and productivity of tomato plants include air temperature (day and night), fruit temperature, radiation, CO_2 concentration, fruit load, plant density, stress and sometimes the soil. For example, a fluctuation of temperature mostly affects the rate of fruit ripening and the rate of fruit growth. The research work of Willits and Peet (1998) also suggests that warmer conditions in the greenhouse at night can improve the quality and quantity of tomatoes. The relationship between temperature and yield is complex Fitz-Rodriguez and Giacomelli (2009). Studies in Fitz-Rodriguez and Giacomelli (2009) and Santos *et al.* (1992) have shown that the sensitivity of fruits to temperature changes over time, as fruits become more sensitive to temperature as they approach maturity. This explains why raising the greenhouse temperature results in a peak in yield followed a few days later by a yield reduction Adams and Valdes (2002). However, temperature fluctuations do not significantly influence the overall tomato yields when compared with temperature controlled growing conditions however they do influence weekly tomato yields in greenhouses Koning (1988) and (1990).

Studies funded by the UK Department for Environment, Food and Rural Affairs (DEFRA) were concerned with trying to understand how the aerial environment affected the pattern of tomato yields. They showed that the primary cause of fluctuations in yield was due to the effect of temperature on fruit ripening Adams *et al.* (2001a) and (2001b). Another important environmental factor is CO_2, which is considered to be the key part of photosynthesis; therefore CO_2 enrichment in greenhouses can significantly increase tomato yields Nilsen *et al.* (1983).

Materials and methods

Depending on the type of the connections between their artificial neurons, ANNs can be divided into two classes, Feed Forward Networks (FFNs) and Feedback Networks (FBNs) Gardner *et al.* 1990). In terms of training an ANN there are three broad categories of learning: unsupervised, supervised and reinforcement. In FFNs, the neurons are connected in a strict hierarchical manner, where the outputs of the neurons in one layer connect to the inputs of the neurons in the next layer and there is no loop back from one layer to another. In addition, there is no interaction between the neurons in the same layer. In FBNs, the connection patterns of the neurons are more flexible and the outputs of neurons in a subsequent layer are allowed to connect (loop back) to neurons in a preceding layer(s) Krose and Smagt, (1996); Russell and Norvig (1995). In this work the focus was on supervised FFN networks.

Since a greenhouse system can be considered as a nonlinear, multivariable, system which is open to the external environment Bennis *et al.* (2008), black box models have been used to tackle this kind of problem, They are of particular utility when the data are composite and multivariate Chen *et al.* (2008), and ANNs are one type of such models. ANNs have been employed in agriculture previously, for example Kalogirou *et al.* 2000 presented their application to a range of issues concerning energy related problems to predict the temperature and CO_2 inside a greenhouse in order reduce power consumption. The ANN was shown to be a robust method, which is very effective at, for example, predicting some of the salient environmental factors inside the greenhouse. ANN models have also been used to optimize greenhouse operation in addition to experimental data fitting Linker *et al.*

1998. The work in Linker *et al.* 1998 indicated that the models used not only fitted the data well but also produced reasonable optimization results in controlling the greenhouse environment; they achieved up to 94% accuracy in predicting the temperature of the greenhouse for 5 and 10 minutes ahead of the sensor signal.

In this work, four three networks were designed and trained: two to investigate the five environmental variables of temperature, solar radiation, CO_2 concentration and VPD using the WSG and WHRI datasets; the third and fourth to make use of only four variables by excluding VPD and leaving everything else as per the first network. A Multi-Layered Perception (MLP) ANN was used in this work. The performance measure was the Mean Square Error (MSE) between the data points (y_n) and the ANN estimates ($\hat{y}_n$), given for *N* yield data points by

Eq. 1:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2$$

The software package MATLAB version 7.10 Neural Networks Toolbox was used to train, validate and test the ANN prediction models.

In an ANN, each artificial neuron is a simple arithmetic unit Gardner *et al.* (1990), which computes its output by summing the weighted input signals and applying an appropriate mathematical transformation to the sum. Thus, the key components of a conventional artificial neuron are the input signals, weights, bias, activation function and the output signal. Here, an MLP network with five inputs, one hidden layer, and a single output was implemented. The process involved in the development of the prediction system is considered in three main stages: (i) input variable selection, (ii) development of the predictor and (iii) evaluation of the predictor. The general architecture of the MLP consists of one input vector, two hidden layers, and a single output.

The MSE was calculated as the difference between the target output and the network output. Initially the network was trained for 1000 iterations, using the Levenberg-Marquardt learning algorithm Marquardt (1963). After training the network to a satisfactory performance, the independent validation set was then used to evaluate the prediction performance of our ANN giving the results discussed in the next section.

Results

The evaluation of our MLP predictor was performed by training it in the first instance with the datasets collected between 2004 and 2007, whereas for the second training case only independent data from “New site” in 2008 has been used. The prediction result generated by our MLP predictor is illustrated in Figure 1 together with the actual measured yield data.

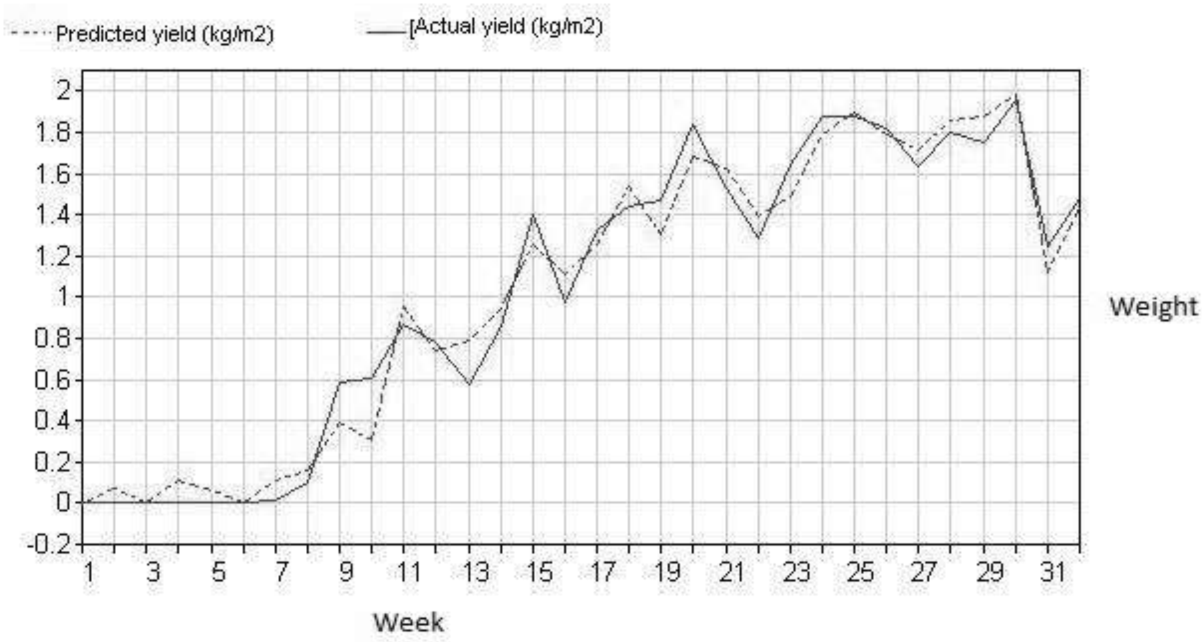

Figure 1: Prediction results using CO_2,

After running the Levenberg-Marquardt algorithm for 1000 iterations, the MSE of the predicted results was 0.005 and the percentage error (PE) in the prediction was 25%. It can be seen that the general pattern of the predicted values roughly follows the desired data and that, as expected, further training improved the performance of the MLP; especially for points with high fluctuation.

Repeating the training with the same dataset but without considering the VPD as an input variable led to different results as shown in Figure 2 where the MSE of the predicted results is 0.0018 and the PE in this prediction was 18%.

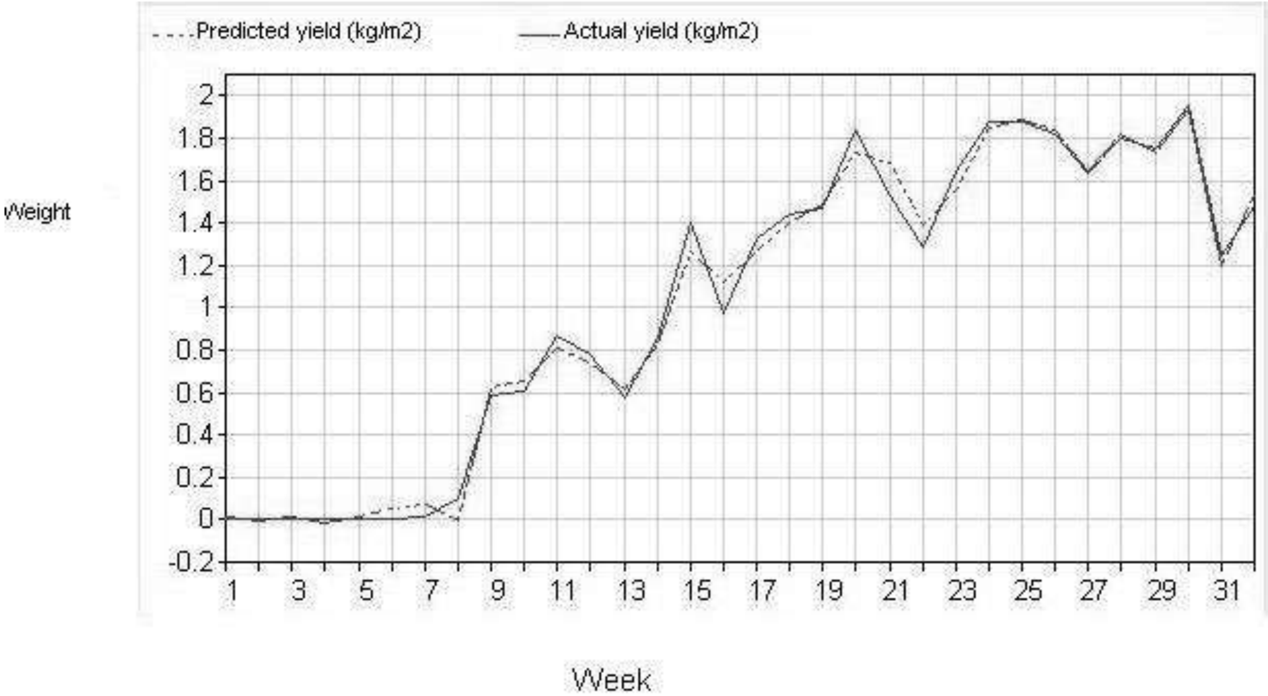

Figure 2. Prediction results without considering VPD

Figure 2 shows the results obtained without considering VPD as an input variable for the WSG dataset. The effect of omitting VPD was to improve the predictive performance pattern and lead to good prediction accuracy. This was a strong indication that VPD does not have a significant impact when predicting tomato yields and that the model without it could be used in predicting the total yield.

The same model, training algorithm and given parameters were used to predict the weekly yield for a different dataset which was provided by Warwick HRI. The data was collected between 1999 and 2000. It contained the weekly tomato yields in a greenhouse and the daily internal environmental measurements of the greenhouse comprising of temperature, solar radiation, CO_2 concentration and VPD. The yield data used to develop the prediction system was recorded on a weekly basis from the middle of March (in calendar week No. 11) when the first fruits were picked, up until the middle of October (in calendar week No. 40).

The greenhouses were divided into four compartments, namely B8, B9, B10 and B11; each compartment being subject to different plant regimes and were monitored separately. However, the variation in the environment between the compartments in every year was within 20% of the average.

The results obtained when predicting the weekly yields for the year 2000 performed very well and gave the results shown in Figure 3 shows prediction results while Figure 4 below show the results obtained without considering VPD as an input variable for the WHRI dataset, where the overall results appear in Table 1.

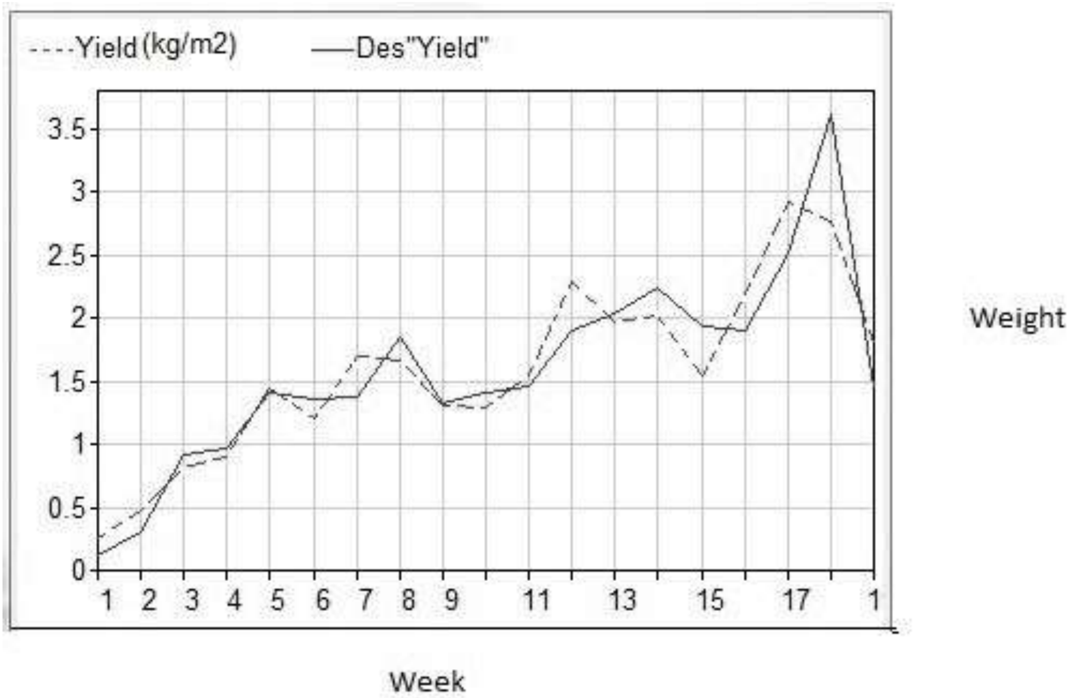

Figure 3. Prediction results for WHRI dataset

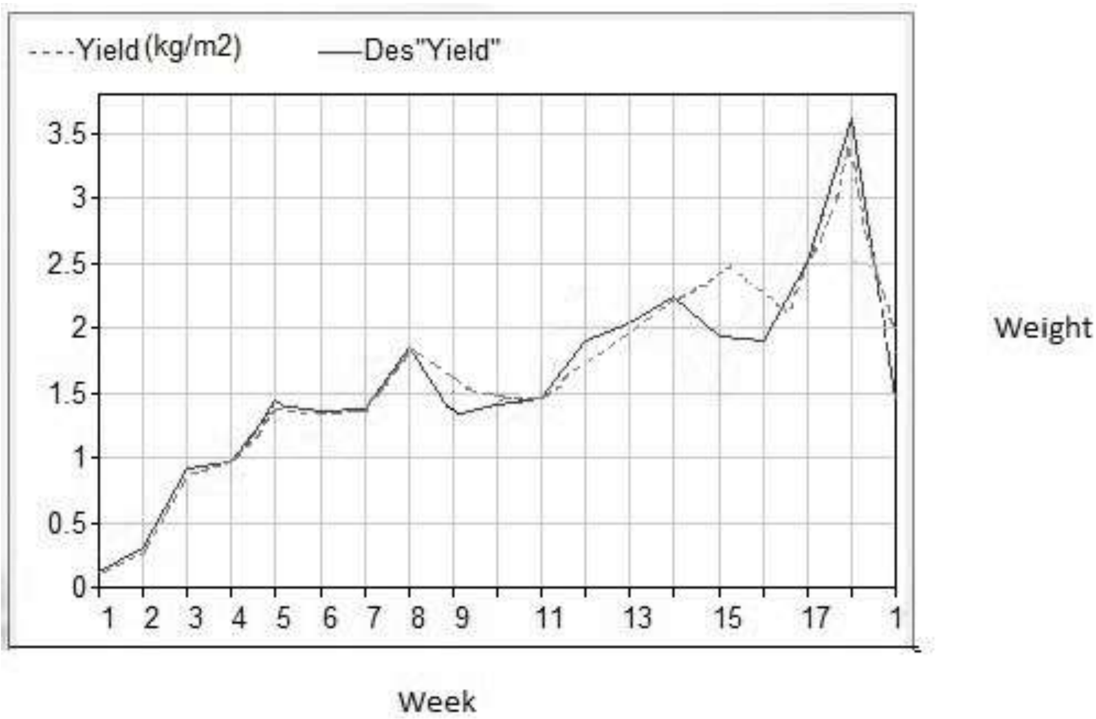

Figure 4. Prediction results for WHRI dataset without VPD

Table 1. Overall results

Dataset (Input variables)	Method	Accuracy
WSG (Temp, VPD, Light, CO_2,Yield)	MLP	87%
WSG (Temp, Light, CO_2,Yield)	MLP	90%
WHRI (Temp, VPD, Light, CO_2,Yield)	MLP	85%
WHRI (Temp, Light, CO_2,Yield)	MLP	89%

The results shown in Table 1 show that the 87% accuracy obtained using the five environmental variables in the WSG dataset was substantially improved to 90% by excluding the VPD as an input variable. The corresponding performances using the WHRI data were 85% and 89%. Hence, data from commercial operations were used to design and train neural networks to predict weekly tomato yields. Although the resulting neural network models are capable of predicting the seasonal and weekly variations of the fruit, they do not provide a mechanistic explanation of the factors influencing these fluctuations. However, knowing this information in advance could be valuable for growers for making decisions on climate and crop management.

Some of the possible advantages of the neural network model implemented in this study include:

(1) many growers currently record the input parameters of the model making it easy to implement

(2) the model can “learn” from datasets with new scenarios (new cultivars, different control strategies, improved climate control, etc.)

(3) less experienced growers could use the system because the decision making process of the expert growers incorporated into the data used in the trained networks and production could thereby become more reliable.

Figure 5 shows the convergence that yield trend line takes in 37 weeks.

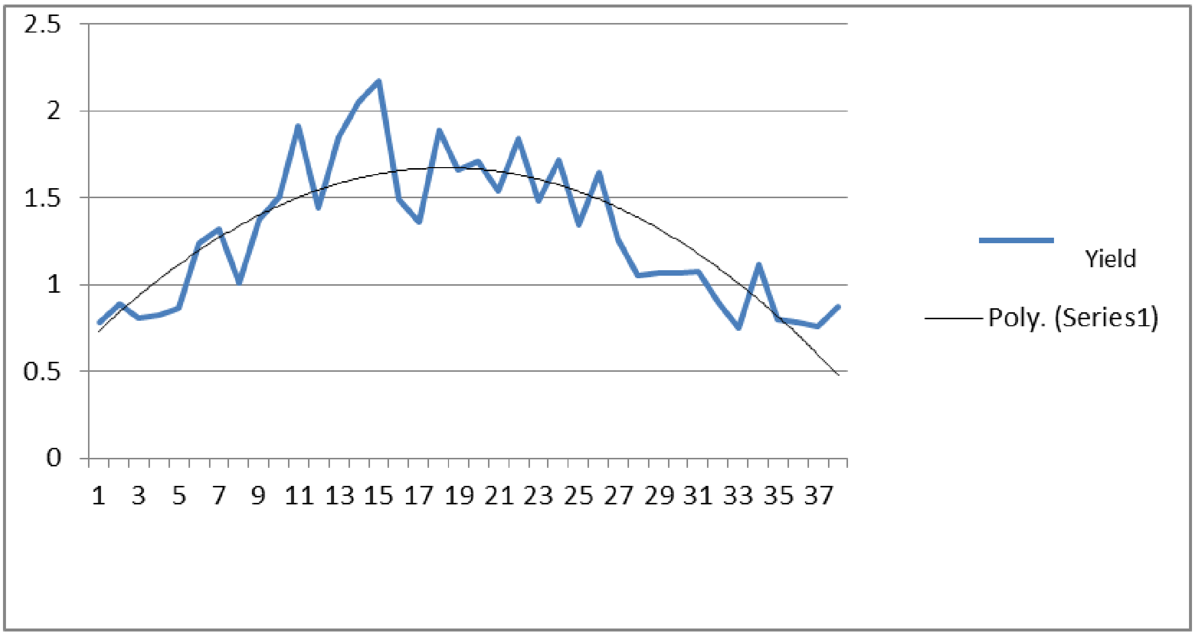

Figure 5. Yield Trend line with polynomial Regression

Discussion

The work presented here used ANNs with a limited number of variables in the experiments and only 1000 generations of MLP iteration. In the context of the number of potential solutions and investigated different tomato cultivars in different greenhouses, this only covers a small part of the prospective solutions. Thus, further work is needed to explore the use of different Intelligent Systems techniques to for example optimize the environmental variables which might affect the greenhouse yield fluctuation and to further explore datasets with different cultivars. The next step will be to develop an integral system that, based on the current growth mode and on the current predicted yield, is able to suggest corrective measures to steer the plant growth mode toward the appropriate direction given by the plant characteristics, the market demand and the seasonal climate expectations.

Conclusions

An Artificial Neural Network (ANN) model has been successfully developed to predict tomato yields in greenhouses. Yield prediction is a complex process and may be affected by a number of factors. Several major steps were taken to predict tomato yield. The first step was to predict the weekly yield for a given greenhouse, using the normal environmental dataset which included temperature, CO_2, radiation and VPD, to predict the weekly yields for the years 2004 to 2007. These variables were used to predict the yield for the next year 2008 of the WSG dataset using an MLP. When employing five environmental variables, the weekly yield prediction accurately was 87%. In a second experiment the WSG dataset VPD parameter was not included and this resulted in an accuracy of 89%. Training and testing an MLP using the WHRI dataset with five environmental variables achieved 85% accuracy in weekly yield prediction, which improved to 89% when only the VPD was excluded.

In overall terms our results indicate that Intelligent Systems techniques such as ANNs can play a role in the accurate prediction of tomato yields and may therefore provide benefits to all those who have an interest in the tomato food chain, ranging from the growers through to the consumers. It is possible to implement intelligent System (IS) techniques, including neural networks, for modeling and predicting one of the used plant parameters for changing the operation of a greenhouse tomato production system. Data from experimental trials and from a commercial operation allowed the modeling of tomato responses in a wide range of growing conditions, increasing the possibilities for application. Although there were some differences among tomato cultivars, the IS techniques were robust in dealing with imprecise data, and they have a learning capability when presented with new scenarios.

Glossary

ANN	Artificial Neural Network
IST	Intelligent System Technique
MLP	Multi Layer Perceptron
NN	Neural Network
GNMM	Genetic Neural Mathematical Method
MSE	Mean Squared Error

Knowledge and Technology transfer:

- HDC Annual Studentship Conference, Lincolnshire, Feb 2010, improving the accuracy of tomato yield prediction through the use of Intelligent Systems techniques. Kifaya Qaddoum, D. Iliescu, E. Hines, A. Steven, F. Zhang, University of Warwick (poster presentation).
- Isle of Wight visit to WSG (Wight Salad Group), June 2010, to discuss the applicability of our methods in the real work environment. Presentations given to show research results and how to employ it as a graphical user interface.

- Meeting at School of Engineering at Warwick University December 2010 with WSG (Wight Salad Group), and Becky Turner (HDC), to discuss the development process of the Graphical User Interface. Presentations given to show research results and how to show the progress.
- Isle of Wight visit to WSG (Wight Salad Group), May 2011, this visit was to demonstrate and discuss in details the designed GUI, also to clarify few points of

References

1. Adams, S. R., Cockshull, K. E., & Cave, C. R. J. (2001a). Effect of Temperature on the Growth and Development of Tomato Fruits. Annals of Botany, 88, 869-877.

2. Adams, S. R., & Valdes, V. M. (2002). The effect of periods of high temperature and manipulating fruit load on the pattern of tomato yields. Journal of Horticultural Science & Biotechnology, 77(4), 461-466.

3. Adams, S. R., Valdes, V. M., Cave, C. R. J., & Fenlon, J. S. (2001b). The impact of changing light levels and fruit load on the pattern of tomato yields. Journal of Horticultural Science & Biotechnology, 76(3), 368-373.

4. Bennis, N. J. Duplaix, G. Enea, M. Haloua and Youlal H. (2008). Greenhouse climate modeling and robust control. Comput. Electronics Agric., 61: 96-1 07.

5. Chen, S.H., Jakeman A.J. and Norton, J.P. (2008). Artificial intelligence techniques: An introduction for their use for modeling environmental systems. Mathematics Comput. Simulation, 78: 379-400.

6. Fitz-Rodriguez, E and Giacomelli, G.A. (2009). Yield Prediction and Growth Mode Characterization of Greenhouse Tomatoes with Neural Networks and Fuzzy Logic. Transactions of the ASABE (American Society of Agricultural and Biological Engineers). Vol. 52(6): 2115-2128

7. Heuvelink, E. (1995). Growth, development, and yield of a tomato crop: Periodic destructive measurements in a greenhouse. Scientia Hort. 61(1-2): 77-99.

8. Heuvelink, E. (1996). Tomato growth and yield: quantitative analysis and synthesis. PhD diss. Wageningen, The Netherlands: Wageningen Agricultural University.

9. Heuvelink, E. (1999). Evaluation of a dynamic simulation model for tomato crop growth and development. Ann. Botany 83(4): 413-422.

10. Kano, A., & Bavel, C. H. M. v. (1988). Design and Test of a Simulation Model of Tomato Growth and Yield in a Greenhouse. Engei Gakkai zasshi (Japanese), 56(4), 408-416.

11. Koning, A. N. M. d. (1988). The effect of different day/night temperature regimes on growth, development and yield of glasshouse tomatoes. Journal of Horticultural Science, 63, 465-471.

12. Koning, A. N. M. d. (1990). Long-term temperature integration of tomato growth and development under alternating temperature regimes. Scientia Horticulturae, 45, 117-127.

13. Krose, B., & Smagt, P. v. d. (1996). An introduction to Neural Networks.

14. Hashimoto, Y., 1997. Application of artificial neural networks and genetic algorithms to agricultural systems. Comput. Electronics Agric., 18: 71-72.

15. Kalogirou, S.A., 2000. Applications of artificial neural networks for energy systems. Applied Energy, 67: 17-35.

16. Jones, J. W., E. Dayan, L. H. Allen, H. van Keulen, and H. Challa. 1991. A dynamic tomato growth and yield model (TOMGRO). Trans. ASAE 34(2): 663-672.

17. Jones, J. W., A. Kening, and C. E. Vallejos. 1999. Reduced state variable tomato growth model. Trans. ASAE 42(1): 255265.

18. Gardner J.W., Hines E.L., Wilkinson M. (1990). Application of artificial neural networks to an electronic olfactory system, Measurement Science & Technology, 1(5), pp.446-451, ISSN:0957-0233 0957-0233

19. Linker, R., I. Seginer and P.O. Gutman, (1998). Optimal CO_2 control in a greenhouse modeled with neural networks. Comput. Electronics Agric., 19: 289-31 0.

20. Marquardt, D. (1963) An algorithm for least squares estimation of non-linear parameters. J. Soc. Ind. Appl. Math. 431–441.

21. Mukherjee, J., & Sastri, C. V. S. (2004). Fruit Yield Predicting Model of Tomato using Spectral and Hyperspectral Indices. Journal of the Indian Society of Remote Sensing, 32(2004), 301-306.

22. Morimoto, T., K. Hatou and Hashimoto Y. (1996). Intelligent control for a plant production system. Control Eng. Practice, 4: 773-784.

23. Nilsen, S., Hovland, K., Dons, C., & Sletten, S. P. (1983). Effect of CO_2 enrichment on photosynthesis, growth and yield of tomato. Scientia Horticulturae, 20, 1-Patel, V.C., R.W.

24. McClendon and J.W. Goodrum, (1998). Color computer vision and artificial neural networks for the detection of defects in poultry eggs. Artificial Intel. Rev., 12: 176-193.

25. Russell, S., & Norvig, P. (1995). Artificial intelligence: a modern approach.

26. Santos, J. R. A., Gomez, A. A., & Rosario, T. L. (1992). A model to predict the yield of determinate tomatoes. Scientia Horticulturae, 50(1-2), 89-105.

27. Willits, D. H., & Peet, M. M. (1998). The effect of night temperature on greenhouse grown tomato yields in warm climates. Agricultural and Forest meteorology, 92, 191-202.

