Computers and Electronics in Agriculture 79 (2011) 82–89

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data

David L. Ehret [a,*], Bernard D. Hill [b], Tom Helmer [a], Diane R. Edwards [c]

[a] Pacific Agri-Food Research Centre, Agriculture and Agri-Food Canada, PO Box 1000, Agassiz, BC, Canada V0M 1A0
[b] Lethbridge Research Centre, Agriculture and Agri-Food Canada, PO Box 3000, Lethbridge, AB Canada T1J 4B1
[c] Department of Biological Sciences, University of Calgary, Calgary, AB, Canada T2N 1N4

ARTICLE INFO

Article history:
Received 7 March 2011
Received in revised form 12 May 2011
Accepted 29 July 2011

Keywords:
Automation
Crop monitoring
Crop models
Greenhouse
Neural networks
Tomato

ABSTRACT

The recent development of tools to automatically monitor important crop attributes *in situ* such as yield, growth and water use offers an opportunity to relate real-time crop status to current environmental conditions. In this study, continuous minute-by-minute measurements of crop yield, growth and water use averaged over weekly, daily, or hourly intervals throughout the growing season were used to determine crop response to changes in the greenhouse environment. The data were obtained from crop monitoring stations established in both commercial and research greenhouses. Crop yield measurements obtained from the monitoring system were generally in very close agreement with yields recorded over a much larger area in the commercial greenhouse. Yield was more closely related ($R^2 = 0.65$) to radiation from the previous week than to radiation in the current week ($R^2 = 0.56$). In addition, a neural network (NN) model of yield which included radiation as an input was better at predicting yield in the following week ($R^2 = 0.70$) than yield in the current week ($R^2 = 0.57$). These results indicate a lag effect of radiation on yield. Similarly, yield was more positively related to growth from the previous week ($R^2 = 0.32$) than to growth from the current week ($R^2 = 0.17$). Neural network models of daily growth at both sites ($R^2 = 0.74, 0.69$) included day of the year, temperature and CO_2 as inputs. A negative relationship between day of the year and daily growth indicates a decline in crop vigor through the measurement period. Neural network models of daily crop water use for the two sites were stronger ($R^2 = 0.91, 0.85$) than those for growth, highlighting the difference in physiological complexity between the two. A model of canopy water status as affected by environmental conditions was generated using hourly measures of tomato canopy mass change. Although the rate of canopy mass gain through the day was often constant, there were days when the plant experienced periods of reduced mass gain mid-day. On those days, the amount of deviation from a constant rate was positively related to radiation, day temperature and water use, suggesting periods of water stress. With subsequent recovery of mass gain rates late afternoon, these deviations did not affect canopy growth for the day. Overall, automated monitoring provides new information on the crop which may readily be incorporated into models of crop performance.

1. Introduction

Climate and substrate sensors are used in the commercial growing of greenhouse tomatoes to provide up-to-the-minute information about the greenhouse environment. These sensors furnish information for record keeping and control purposes. However, they provide only indirect information on the status of the crop. Physiological information could provide important supplementary data, especially if integrated into control systems or computer models. For example, rather than calculating transpiration rate from microclimate date, transpiration could be measured directly. A range of physiological parameters such as growth, photosynthesis, transpiration, and leaf temperature may be automatically monitored with a variety of instruments, either remotely, for example with imaging systems (Morden et al., 1997), or through a number of physical contact methods (Ehret et al., 2001). Recently, sap flow and stem diameter measurements have been intensively studied as tools with which to schedule irrigation (Steppe et al., 2008; De Swaef et al., 2009; De Swaef and Steppe, 2010). Another technique using balances and weighing lysimeters was developed years ago to measure transpiration in a variety of cropping situations (Van Bavel and Myers, 1962; Grimmond et al., 1992; Van Meurs and Stanghellini, 1992), or to calibrate other sensors (Baille et al., 1992). In most cases, the mass of the plant is necessarily a component of the overall mass being recorded. However, some indeterminate vines such as tomatoes and long English cucumbers are trained in such a way that the mass of the canopy is largely

* Corresponding author. Tel.: +1 604 796 1712; fax: +1 604 796 0359.
E-mail address: david.ehret@agr.gc.ca (D.L. Ehret).

0168-1699/$ - see front matter Crown Copyright © 2011 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2011.07.013

supported by an overhead crop wire. De Koning and Bakker (1992) showed that the daily mass gain of a single suspended tomato plant could be accurately recorded with an electronic force gauge suspended from the crop wire. From these principles, the Crop*Assist* monitoring system was developed which could continuously and automatically record yield, growth and water use in greenhouse vine crops such as tomato (Helmer et al., 2005) or herbaceous crops such as pineapple sage (Ehret et al., 2004).

The goal of this study was to demonstrate the potential of the Crop*Assist* automated crop monitoring system to provide data with which to develop models of tomato crop yield, growth and water use in response to changes in greenhouse climate.

2. Materials and methods

2.1. Greenhouse conditions and crop culture

Plants were grown in greenhouses at Gipaanda Greenhouses (GIP) and at the Pacific Agri-Food Research Centre (PARC) in 2002. Both were Venlo style glasshouses, which are gutter-connected greenhouses with high walls and passive roof-top ventilation. The tomato plants (*Solanum lycopersicum* L.) cv. Rapsodie (Rogers/Syngenta Seeds, Boise, ID) were grafted onto cv. Maxifort rootstock (deRuiter Seeds, Bergschenhoek, The Netherlands). In both cases, plants were grown according to commercial production guidelines established for British Columbia, Canada (BCMAFF, 1996). The substrate was yellow cedar sawdust at both sites. Site-specific details are given below.

The GIP site was a commercial greenhouse located in Delta, BC (N49°03′45″, W122°06′16″). The number of stems was adjusted during the season by controlling the growth of side shoots. In commercial tomato greenhouses, the number of stems per plant is generally manipulated to be in synch with seasonal changes in light conditions (leaf area per stem is relatively constant). Thus when light levels are high, stem density is increased in order to take advantage of the higher incident light for photosynthesis. Automated crop monitoring began on April 3, 2002 at which time stem density was 3.67 stems m^{-2}. Stem density was reduced to 2.57 stems m^{-2} during the week of August 18, 2002 and maintained at that value until crop monitoring was terminated on November 14, 2002. Daily sums for global solar radiation ($MJ\,m^{-2}\,d^{-1}$) from a pyranometer located on the greenhouse roof, and greenhouse day, night and mean 24 h values for temperature, relative humidity (RH), and CO_2 concentration ($\mu L\,L^{-1}$) were recorded during the crop monitoring period.

The PARC site was a research greenhouse located in Agassiz, BC (N49°14′37″, W121°45′53″). Data were collected from four independent greenhouse compartments with a growing area of 65 m^2 each. Crop stem density was 2.5 stems m^{-2} in each compartment. Two compartments were enriched with CO_2 and the other two were maintained at ambient CO_2, all part of a separate experiment on CO_2 enrichment. Each compartment had independent climate and irrigation control. Climate data as described for GIP were logged at 5 min intervals throughout the experiment and tallied for the day.

2.2. Crop monitoring

Crop*Assist* monitoring stations were established at the GIP and PARC sites according to the methods of Helmer et al. (2005). Basically, Crop*Assist* uses two pairs of load cells. An upper set weighs the crop canopy for continuous measurement of daily growth (midnight to midnight), fruit harvests, and state of canopy hydration through the day. To accomplish this, stems are transferred from the overhead crop wire found in all modern tomato greenhouses to a beam placed parallel to the crop wire but suspended from the load cells. The upper load cells were S-beam tension/compression cells (Revere Transducers, Model 363D3-20T1-50lb, Tustin, CA, USA). Each cell had a capacity of 22 kg with a readability between 1 and 5 g. The combined error due to hysteresis and non-linearity was between 0.02% and 0.03% of full scale, with temperature sensitivities between 0.001% and 0.002% of load per °C. A lower set of load cells measures crop water use (transpiration plus water contributing to growth), irrigation events, percent leaching, and substrate moisture content. This is achieved by placing plants (in containers of substrate) in a trough positioned on top of two single point load cells (HBM, Model SP4-30 kg, Marlboror, MA, USA), each with 30 kg capacity, mounted in aluminum brackets, and positioned approximately 3 m apart. Initially, the upper and lower load cells utilize the same plants. However, greenhouse tomato training practices dictate that as these indeterminant plants grow, the stems must be lowered and incrementally moved along the row, all in the same direction. Since the upper load cells and beam are stationary, stems will eventually be moved off one end of the beam, while new ones will be moved onto the other end, all the while maintaining the same number of stems on the beam.

At GIP, one station was placed approximately midway along a typical 100 m row in a greenhouse zone of 55,000 m^2. Six plants were positioned in the trough (for measurement by the lower load cells) and 12 stems were suspended from the overhead beam (for measurement by the upper load cells). The station was designed to function inconspicuously, requiring minimal attention by the grower and with no special considerations required by the greenhouse workers. At PARC, each of the 4 compartments had an independent Crop*Assist* station located in the middle of the growing area and midway along a 10 m row. Ten plants were placed in the trough and either 9 or 10 stems were suspended from the overhead beam.

2.3. Neural network analysis

Neural network (NN) models were developed with the data provided by the Crop*Assist* stations. Datasets to predict yield, growth and water use consisted of different combinations of time (days or weeks) and greenhouse environmental inputs (temperature, radiation, CO_2 levels, RH, and so on) paired with known outputs (actual yield, growth, and water use) on a case-by-case basis. Preliminary work showed that NN models of daily growth were somewhat improved by using growth values averaged for the current day plus the following day rather than each day separately. This procedure was followed throughout the study. Data were prepared for NN modeling by reducing the input measurements to two decimals or less, then randomizing the order of the cases presented to the NN software.

Feed-forward NN models were constructed using NeuralWorks Predict®, v3.21 software (NeuralWare®, Carnegie, PA) on a desktop PC (2.4 GHz). Briefly, this involved: (1) selecting training and validation subsets, (2) analyzing and transforming data, (3) selecting variables, (4) network construction and training, and (5) model verification. Details on how Predict® conducts NN modeling have been previously published (Hill et al., 2002; Ehret et al., 2008).

NN modeling is an iterative trial and error process initiated using different random seed numbers. The software generates a random number from the seed number to initialize the genetic algorithm used in variable selection, and to initialize the connection weights when constructing the neural networks. For all predictions, 30 different random seed NN models were considered. Usually, no one absolute best NN model was obtained but rather a few different models with similar performance. The 'best' NN models were chosen on the basis of a suitable NN architecture (i.e. a minimum number of input neurons connected to a hidden