Predicting Lettuce Canopy Photosynthesis with Statistical and Neural Network Models

Jay Frick,1 Cyrille Precetti,2 and Cary A. Mitchell3
National Aeronautics and Space Administration Specialized Center of Research and Training in Bioregenerative Life Support, Purdue University, 1165 Horticulture Building, West Lafayette, IN 47907-1165

ADDITIONAL INDEX WORDS. Lactuca sativa, CELSS, chaos, crop optimization, environmental control, hydroponics, NASA

ABSTRACT. An artificial neural network (NN) and a statistical regression model were developed to predict canopy photosynthetic rates (Pn) for 'Waldman's Green'1 leaf lettuce (Lactuca sativa L.). All data used to develop and test the models were collected for crop stands grown hydroponically and under controlled-environment conditions. In the NN and regression models, canopy Pn was predicted as a function of three independent variables: shootzone CO2 concentration (600 to 1500 mmol mol-1), photosynthetic photon flux (PPF) (600 to 1100 μmol m-2 s-1), and canopy age (10 to 20 days after planting). The models were used to determine the combinations of CO2 and PPF setpoints required each day to maintain maximum canopy Pn. The statistical model (a third-order polynomial) predicted Pn more accurately than the simple NN (a three-layer, fully connected net). Over an 11-day validation period, average percent difference between predicted and actual Pn was 12.3% and 24.6% for the statistical and NN models, respectively. Both models lost considerable accuracy when used to determine relatively long-range Pn predictions (>6 days into the future).

The U.S. National Aeronautics and Space Administration (NASA) is considering the use of higher plants to sustain human life during long-term space exploration and colonization missions (Salisbury and Clark, 1996). The proposed controlled environment life-support system (CELSS) will be at least partially bioregenerative and feasible during missions when resupply from Earth is either impossible or impractical. Through the biological processes of photosynthesis and transpiration, higher plants in a CELSS will absorb CO2, release O2, produce fresh food, and provide a continual source of potable water (Hoff et al., 1983).

Space structures typically are designed with major emphasis on weight reduction (Allen and Haisler, 1985). The resulting constraints suggest that a CELSS should be as small as possible, with minimal resource storage buffers. This further implies a highly dynamic system, capable of responding rapidly to changes in state or demand. To maintain stability and manage such a biomechanical system, comprehensive crop-response models are needed. Such models must accurately predict real-time crop responses with respect to changes in the environment and the changing needs of attending space travelers.

Because plants are living organisms, predictable and consistent crop responses often are difficult to obtain. It is well known that crop dynamics are a function of genetic diversity as well as complex interactions with the environment (Milthorpe and Moorby, 1979). However, other chaotic factors also could play a role. For example, it is unknown if subtle environmental variations early in the growth cycle influence long-term crop responses. All such factors should be addressed when developing comprehensive crop-response models for CELSS design and simulation.

In this study, we developed a simple neural network (NN) as an alternative to a statistical regression model to describe photosynthetic response (Pn) during crop development. NNs form a unique class of mathematical algorithms constructed from sets of simple computing elements that have been assembled to perform a specific task (Machado et al., 1992). Information is processed as it passes through each of the elements (or neurons) within the network. NNs were developed by McCulloch and Pitts (1943), who described the computing elements as analogous to biological neurons in the human brain.

The motivation for using NNs lies in their learning ability. A mathematical proof exists that demonstrates that any continuous function can be reproduced (or approximated) with a three-layer NN, provided enough neurons are in the hidden layer. Recently, NNs have been reported to learn input-output relationships without the statistics required for a regression approach (Bailey and Thompson, 1990a). NN models can be developed when there is limited information about the relationship between independent and dependent variables (Varle, 1994), and they can be retrained as additional experimental data become available. In contrast, incorporating new data into an existing statistical model is difficult, unless entire series of experiments are carefully preplanned for that purpose. A drawback of NNs is that they provide no information regarding statistical significance of independent variables or significance of interactions among the variables, nor do they offer any insight into the biological or physical processes underlying a given input-output data set. Additional information detailing the relationship between statistics and NNs can be found in Ripley (1993), Weiss and Kulikowski (1991), and White (1992).

The objective of this study was to develop and evaluate dynamic models to quantify lettuce canopy Pn. Models incorporating multiple interdependent variables and those capable of near real-time Pn predictions were of primary interest. Algorithms that predict crop response and yield are needed for CELSS development and are relevant in other applications of controlled-environment agriculture as well. For example, these efforts may serve as a precursor to phasic environmental control strategies or for active-control scenarios, whereby environmental conditions are optimized in real time, depending on continuous feedback response signals from the crop.

Materials and Methods

EQUIPMENT AND APPARATUS. Lettuce canopy net Pn data were generated from experiments conducted using the controlled-environment Minitron II facility at Purdue University (Knight et al.,

Received for publication 2 Oct. 1997. Accepted for publication 18 May 1998. Purdue University Agricultural Experiment Station journal paper no. 15489. Research was supported in part by NASA grant NAGW-2329. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked "advertisement" solely to indicate this fact.  
1Research engineer.  
2Postdoctoral research assistant.  
3Professor, Dept. of Horticulture, Purdue University; to whom reprint requests should be addressed.
refilled with single-strength Hoagland's nutrient solution, EC = 2500 mS cm⁻¹ (Hoagland and Arnon, 1950) and PPF was increased to 600 μmol·m⁻²·s⁻¹. PPF was provided by a combination of high pressure sodium, metal halide, and incandescent flood lamps. The incandescent output was continuously variable, making fine gradations in PPF possible. All other environmental conditions were controlled to the same setpoints as described. Environmental conditions in both Mintron chambers were maintained at the above levels until day 10 after planting.

Treatments to enhance photosynthesis began on day 10 after planting. Day 10 was chosen because canopy Pn was not detectable before this time due to relatively small plant size compared to the total chamber volume. Treatments consisted of altering PPF and CO₂ levels, and then recording the corresponding steady-state canopy Pn response associated with each treatment combination. Three PPF levels were investigated: 600, 850, and 1100 μmol·m⁻²·s⁻¹. The CO₂ concentrations examined were 600, 1050, and 1500 μmol·mol⁻¹. Each day the full factorial of CO₂ and PPF (nine treatment combinations) was run within each Mintron chamber. Each treatment combination was maintained for 50 consecutive minutes after steady-state conditions were achieved. The resulting canopy Pn was measured repeatedly and averaged over these 50-min treatment intervals.

The order of PPF and CO₂ treatment combinations was randomly selected at the beginning of each day. After Pn data had been collected for all nine treatment combinations, setpoints for PPF and CO₂ were reset to 850 μmol·m⁻²·s⁻¹ and 1050 μmol·mol⁻¹, respectively, for the remainder of the day's 20-h photoperiod. This entire procedure was repeated each day, starting on day 10 and ending on day 20. The Pn data derived from this experiment were used to develop the regression and NN models.

In a subsequent experiment, the procedure was duplicated in as exact a manner as possible, except that daily treatments consisted of setting CO₂ and PPF setpoints to levels predicted by the models to result in maximum canopy Pn. Each day from day 10 to 20, the maximum Pn treatment setpoints (via both models) were run within each Mintron chamber, and corresponding canopy Pn was measured. The data from this second experiment were used to evaluate the performance and accuracy of both the statistical and NN models.

The accuracy of each model was evaluated in two ways: First, the absolute difference between predicted and actual Pn (e.g., predicted maximum Pn – measured Pn) was determined for each day during the evaluation period (days 10 to 20 after planting). Therefore, Pn differences that approached 0 indicated a highly accurate prediction by the model. Secondly, percent differences (Holman, 1984) between predicted and actual Pn were calculated for each model, and an overall average percent difference was calculated over the 10-d test period.

**Model development.** The statistical, regression model was developed using SAS software (SAS Institute, Cary, N.C.). Experimental design was treated as a split block, with each Mintron chamber as a separate experimental unit. Main effects and interactions among the dependent variables were completely partitioned into 198 degrees of freedom (Steel and Torrie, 1980). Next, a Bartlet's test was used to check for error homogeneity (Steel and Torrie, 1980). Since error variance increased with the mean, the data were transformed (using a square-root function) in accordance with the Box-Cox analysis of transformations (Box et al., 1978). After generating the ANOVA tables, insignificant terms (P = 0.05) were eliminated. In the final step, best-fit regression coefficients were calculated to complete the Pn model.

Developing the NN model was largely a trial-and-error process. From early simulations we determined that relatively simple NN topology would suffice to represent our Pn data. The raw data were
Table 1. Regression model used to predict lettuce canopy photosynthesis (Pn). Pn equation is valid only when independent variables are within given limits.

<table>
<thead>
<tr>
<th>Term no.</th>
<th>Model term</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept = 31.135</td>
<td>---</td>
</tr>
<tr>
<td>2</td>
<td>Time</td>
<td>-5.335058</td>
</tr>
<tr>
<td>3</td>
<td>Time²</td>
<td>0.352644</td>
</tr>
<tr>
<td>4</td>
<td>Time³</td>
<td>-8.526×E-03</td>
</tr>
<tr>
<td>5</td>
<td>CO₂</td>
<td>-2.7849×E-02</td>
</tr>
<tr>
<td>6</td>
<td>CO₂×time</td>
<td>4.388×E-03</td>
</tr>
<tr>
<td>7</td>
<td>CO₂×time²</td>
<td>-2.61×E-04</td>
</tr>
<tr>
<td>8</td>
<td>CO₂×time³</td>
<td>6.118×E-06</td>
</tr>
<tr>
<td>9</td>
<td>PPF</td>
<td>-6.5942×E-02</td>
</tr>
<tr>
<td>10</td>
<td>PPF²</td>
<td>1.988×E-05</td>
</tr>
<tr>
<td>11</td>
<td>PPF×time</td>
<td>8.532×E-03</td>
</tr>
<tr>
<td>12</td>
<td>PPF×time²</td>
<td>-3.95×E-04</td>
</tr>
<tr>
<td>13</td>
<td>PPF×time³</td>
<td>8.799×E-06</td>
</tr>
<tr>
<td>14</td>
<td>PPF²×time</td>
<td>-1.386×E-06</td>
</tr>
<tr>
<td>15</td>
<td>CO₂×PPF</td>
<td>7.9839×E-05</td>
</tr>
<tr>
<td>16</td>
<td>CO₂×PPF²</td>
<td>-1.6702×E-08</td>
</tr>
<tr>
<td>17</td>
<td>CO₂×PPF×time</td>
<td>-1.2236×E-05</td>
</tr>
<tr>
<td>18</td>
<td>CO₂×PPF×time²</td>
<td>6.96×E-07</td>
</tr>
<tr>
<td>19</td>
<td>CO₂×PPF×time³</td>
<td>-1.5219×E-08</td>
</tr>
<tr>
<td>20</td>
<td>CO₂×PPF²×time</td>
<td>1.06285×E-09</td>
</tr>
</tbody>
</table>

\[ Pn = \left[ \text{Intercept} + \sum \text{Term}_i \times \text{Coefficient}_i \right]^2, \text{ with } i = 2 \text{ to } 20. \]

\[ \text{Independent variables: time, 10 to 20 d after planting; CO}_2 \text{, level, 600 to 1500 } \mu\text{mol-m}^{-2}\text{-s}^{-1}; \text{ and PPF, 600 to 1100 } \mu\text{mol-m}^{-2}\text{-s}^{-1}. \]

\[ Pn = [\text{Intercept} + \sum \text{Term}_{(i)} \times \text{Coefficient}_{(i)}]^2, i = 2 \text{ to } 20. \]

In general, first-, second-, and third-order terms containing time (days after planting) were significant in the statistical model. This result appears reasonable since a third-order polynomial may approximate a sigmoid-type response curve. Only first-order effects of CO₂ level were significant, whereas first- and second-order terms of PPF were included in the model. The statistical model is valid only when CO₂, PPF, and time are all within ranges of the treatment combinations used to collect the original data set.

A NN model for canopy Pn was developed as an alternative to the traditional statistical approach. Final architecture included a simple, fully connected, three-layer NN, with three nodes in the input layer (to represent CO₂, PPF, and time), three nodes in the hidden layer, and one node in the output layer (to represent canopy Pn). The resulting NN model had an \( R^2 \) of 0.96 and a maximum delta of 0.075. Variations of this NN architectures were attempted, but model \( R^2 \) did not change significantly with other designs.

**Using the Models.** Both models were used to predict the maximum and minimum Pn that could be achieved on any given day during a crop growth cycle (Fig. 1). It follows that any desired Pn between the daily maximum and minimum values also is attainable by setting CO₂ and PPF at some appropriate levels. The maximum and minimum Pn predicted for each day depended on our experimental ranges of CO₂ and PPF. For example, certainly Pn < 0 can be obtained if respiration plants are contained in total darkness. However, 0 PPF was out of the boundaries of the original data set used to generate the models.

Both models revealed a continuous sigmoidal-type relationship to describe canopy Pn over time (Fig. 1). However, the curvature of the NN response model was more gradual than that of the statistical model. The models predicted a maximum Pn potential of \( = 0 \) (on

---

**Results**

**Model Specifics.** Canopy Pn data collected from the first experiment were used to develop the statistical and NN models. In both approaches, Pn was determined as a function of three independent, input variables: 1) CO₂ level, 2) PPF, and 3) time after planting. It was not possible to model Pn throughout the entire growth cycle (days 0 to 20) because Pn was not detectable before day 10 due to small plant size in their lag phase of development.

The function used to model canopy Pn took the form of a complex polynomial with 20 terms (Eq. [2]). Table 1 contains specific information regarding each term in the model. The regression model fit the experimental data set with an \( R^2 = 0.994 \).

---

Fig. 1. Daily range of attainable canopy photosynthesis (Pn) rates as predicted by the statistical and neural network models and with dependent variables held within specific limits.
day 10) up to 65 μmol·m⁻²·s⁻¹ CO₂ (on day 20). It should be noted that 0 Pn reflected the sensitivity of our environmental control equipment. Hence, the canopy was photosynthetically active, but this activity was not detectable until after day 10.

In addition to predicting the daily range of possible Pn, the models also were used to determine setpoints for CO₂ and PPF required to achieve maximum potential Pn for each day in the growing cycle (Fig. 2). Daily maximum Pn values were available from the data used to generate Fig. 1. Starting with the desired maximum Pn, models were solved iteratively to determine CO₂ and PPF combinations required each day to obtain the predicted maximum Pn (Fig. 2). This day-to-day optimization approach is potentially valuable since energy and resource savings may be realized without compromising crop growth.

Suggested CO₂ and PPF setpoints differed substantially between the statistical and NN models (Fig. 2). According to the statistical model, PPF should be set at 900 μmol·m⁻²·s⁻¹ on day 10, and then increased gradually to 1100 μmol·m⁻²·s⁻¹ by day 18. This demonstrates that a phasic environmental control strategy may be applicable for lettuce stands without compromising potential Pn. The statistical model suggested very slight alterations in daily CO₂ setpoints, starting with 1400 μmol·mol⁻¹ on day 10 and increasing to 1500 μmol·mol⁻¹ by day 14 (Fig. 2). It is unknown if such a subtle change in CO₂ levels realistically affects crop development. The statistical model also revealed that CO₂ and PPF could be decreased near the end of the growth cycle (day 19) without significant reduction of maximum Pn.

According to the NN model, daily setpoints for CO₂ and PPF could not be altered greatly and still achieve the maximum predicted Pn (Fig. 2). For all practical purposes, the NN suggested a constant PPF of 1040 to 1050 μmol·m⁻²·s⁻¹ throughout the cropping cycle. Similarly, a fairly constant CO₂ setpoint of 1500 μmol·mol⁻¹ was determined from the NN model. The NN model also implied no Pn response to CO₂ on day 10. Therefore, there was no suggested CO₂ setpoint for that day.

Validating the Models. In a follow-up experiment, daily CO₂ and PPF levels were set to those conditions determined by the statistical and NN models to maintain maximum Pn (Fig. 2). Canopy Pn was measured under these conditions and compared to the daily, predicted maximum target Pn (Fig. 1). Overall, the statistical model predicted canopy Pn more accurately than the NN model (Fig. 3 and Fig. 4). In the statistical approach, average percent difference between predicted Pn and actual Pn was 12.3%, when results were averaged over an 11-d evaluation period (days 10 to 20 after planting). In comparison, average percent difference for the NN model averaged 24.6% over a 10-d period (days 11 to 20).

For both models, measured canopy Pn typically was less than the models had predicted (Fig. 3). Furthermore, divergence between predicted and actual Pn values became greater as the canopy aged (Fig. 4). For the statistical approach, actual Pn began to diverge from the predicted values by day 16. The situation was more extreme in the NN model, with divergence beginning by day 12.

Discussion

The Pn models reported in this study are relatively simple compared to those that ultimately will be required for a CELSS. Of the 10 or more proposed CELSS candidate species (Hoff et al., 1983; Salisbury and Clark, 1996), all will require models to predict total crop response (e.g., Pn, transpiration, respiration, and edible biomass yield and quality). These models must predict crop response as a function of multiple, environmental variables. For example, models must simultaneously incorporate variables such as CO₂, PPF, temperature, spectral quality (for reproductive crops), photoperiod, and humidity. Rootzone variables such as nutritional regimes and pH also have been shown to be important when quantifying crop responses (Frick and Mitchell, 1993; Frick et al., 1994).

These dynamic models were developed in part to test the possibil-
Diversity between predicted and actual Pn values over time (as seen with both models, Figs. 3 and 4). One explanation for this divergence may involve chaotic-type effects, where subtle environmental differences early in the growth cycle may result in unpredictable, potentially large discrepancies in the expected crop response. In chaos theory this concept commonly referred to as the butterfly effect (Gleick, 1987).

More research is needed before attributing our observed divergence entirely to chaos. However, assuming chaos plays even a minor role in crop responses, then CELSS development will encounter a new level of complexity. Not only will it be difficult to create consistent and accurate crop models, but long-term system stability also might be questionable. For instance, suppose an environmental perturbation occurs in a CELSS (e.g., setpoint changes, equipment malfunction, etc.), then the following questions arise: What magnitude and duration of disturbance causes a significant deviation from model predictions? During which crop growth phase (if any) is it safe for such a disturbance to occur? To answer these questions, models of all types should be explored and refined.

**Literature Cited**


