# THE PREDICTION AND VALIDATION OF GREENHOUSE TOMATO YIELD USING MATHEMATICAL MODELS AND EXPERT SYSTEMS

by

**WINSTON C. TANG**

B.A.Sc., The University of British Columbia, 1993

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Bio-Resource Engineering

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1995

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Bio-Resource Engineering

The University of British Columbia
Vancouver, Canada

Date April 24, 1995

DE-6 (2/88)

**Abstract**

Greenhouse tomato yields were predicted using two mathematical models developed in this study - the empirical and deterministic models. Weekly yield predictions for an entire growing season were compared with actual yields and results from an expert system model developed by the Agassiz Research Station.

The deterministic math model involved using first principle equations of photosynthesis and respiration to simulate crop growth. Utilizing a known tomato yield conversion factor, net photosynthesis rates ($P_{net}$) were converted to weekly yield predictions and compared with actual recorded yields. A deterministic model using two week cumulations of $P_{net}$ converted to yield was used successfully to predict actual tomato yields 6 weeks ahead of time with a root-mean-square-error (RMSE) of 0.38 kg/m$^2$.

The empirical math model employed regression techniques to fit historical greenhouse climate data to recorded yields. Correlations between light, temperature, and weekly tomato yields were derived into equations to predict yields for future growing seasons. An empirical model cumulating 3, 6, and 9 weeks of light and temperature data was developed to predict yields 4 weeks ahead of time with a RMSE of 0.45 kg/m$^2$.

When one-week-ahead predictions from the Agassiz expert system model were compared with actual recorded yields a RMSE of 0.401 kg/m$^2$ was calculated. The expert system model utilizing trend recognition techniques was also used as a comparison with the two math models. When compared and ranked for prediction accuracy, application flexibility, and user-friendliness, the expert system was chosen as the overall best model for tomato yield prediction.

# TABLE OF CONTENTS

## TABLE OF CONTENTS



## LIST OF TABLES

## LIST OF FIGURES

**Figure**



### *Appendix Figures*

## LIST OF FIGURES (cont'd)

**Figure**

## Acknowledgements

I would like to express my sincerest thanks and gratitude to my thesis advisor, ***Dr. Anthony K. Lau***, for his thoughtful advice and assistance throughout this project. The kindness, encouragement, and friendship that has been shared will not be forgotten. Thank you Anthony, for all your help throughout my university experience.

A special thank you to ***Dr. Wei C. Lin***, for providing results from the Agassiz Expert System for comparison with the mathematical models developed in this study. Thank you Wei, your assistance, advice, and explanations were always helpful and will always be remembered.

To ***Dr. Sie-Tan Chieng***, thank you for your insightful ideas and suggestions throughout this research project. Your many helpful suggestions and recommendations were always welcome and greatly appreciated. Thank you Sie-Tan!

A special thank you to my wonderful ***Princess*** who helped me endure all the long hours of writing and analysis. You were always there for me when I needed encouragement, compassion, or ranch-dipped pretzels. Thank you so much for all the love, laughter, and the most unforgettable moments of my life! I.L.U.F.B. U.R.E.2.Me.W.B. I.W.2.M.U. W, W.U.M.M.? What do you say?

"If you believe, it does exist. If you are determined, it can be yours.." - *WCT*

W² & Pekkle (TLF)

## Introduction

Greenhouse produced crops have become an important part of the agricultural industry (BCFA, 1994) and are well-known in the marketplace for their high quality and consistency. However, while striving to provide a top quality product, greenhouse producers have had to incur heavy capital costs, and these costs are then inevitably reflected to the consumer. Consequently, any advances that may improve the yield and lower the cost of producing greenhouse crops will benefit both growers and consumers.

Current decision-making for greenhouse climate control and production relies heavily on the past experience of the individual grower and general guidelines published in production manuals. As a result, grower yields tend to vary greatly on an individual and yearly basis. In an attempt to solve these problems, computers have been implemented in climate control to regulate the aerial and root-zone environmental conditions for crops and thereby improve yields. However, computer climate control still remains largely dependent on settings prescribed by individual growers. Although computers have been useful in providing automated control of environmental parameters (Hashimoto et al., 1993), the actual dynamic relationships that exist between climatic conditions and crop yields still remain unclear.

Models of the greenhouse environment and simulations of crop growth are needed to enhance environmental control strategies and to encourage growth of the entire greenhouse industry (Jones et al., 1991).

One approach to improving crop yields and identifying plant response relationships lies in developing mathematical growth models that can predict future yields given a set of climatic conditions. Successful growth models can then help to identify the environmental factors with the greatest influence on crop yield and aid the grower in making essential management decisions to improve productivity. Affected decisions may range from simple adjustments of lighting or temperature setpoints to justifying the cost of investing in new equipment. A valid growth model can also be a useful tool for the greenhouse grower by providing a quick estimate of future yields based on current practises.

Another approach for yield prediction involves expert system modeling which is an advanced information processing method that is able to recognize patterns and compensate for incomplete or ambiguous data (Rehbein, et al., 1992). Provided with historical data, a properly trained expert system will be able to analyse trends in production as they relate to input data (ie. light, temperature, and other greenhouse climatic variables) and predict future yields.

By providing reliable future yield and growth forecasts, both the mathematical modeling and expert system approaches aim to improve the overall productivity of greenhouse crops and enhance their marketability to the consumer. Successful yield prediction is a valuable tool for determining better greenhouse management practises, reducing resource wastage, and identifying the relationships between plant responses, crop yield, and climate.

## 1.0 Theory and Review

As world populations continue their steady growth, greenhouse production systems must become more efficient to meet the increasing demands for food. Much of the recent advances in greenhouse crop production have been linked to the application of computers in climate control. Computer automation has already allowed greenhouse environmental conditions to be properly monitored and maintained to increase crop productivity while reducing energy requirements. However, in order to make further advances in greenhouse productivity a better understanding of plant responses to climatic conditions is required (Tantau, 1980). The emphasis on developing innovative methods for optimizing greenhouse production has intensified globally over the past decade (Jones, 1991). Devising models or methods for predicting future crop yields is an effective way to investigate plant growth and climate relationships.

### 1.1 The Development of Mathematical Yield Models

Developing mathematical models to simulate crop growth is one method that can be used to predict final yield. Math models are generally divided into two main classes - those that employ the deterministic method and those that apply the empirical approach. (***Figure 1***).

The deterministic approach relies on general laws or basic first principles to derive a scientifically-based model. Researchers have already used this approach to produce models to make predictions of canopy photosynthesis (Acock et al, 1978), and plant transpiration (Stanghellini and van Meurs, 1992;Yang et al., 1990).

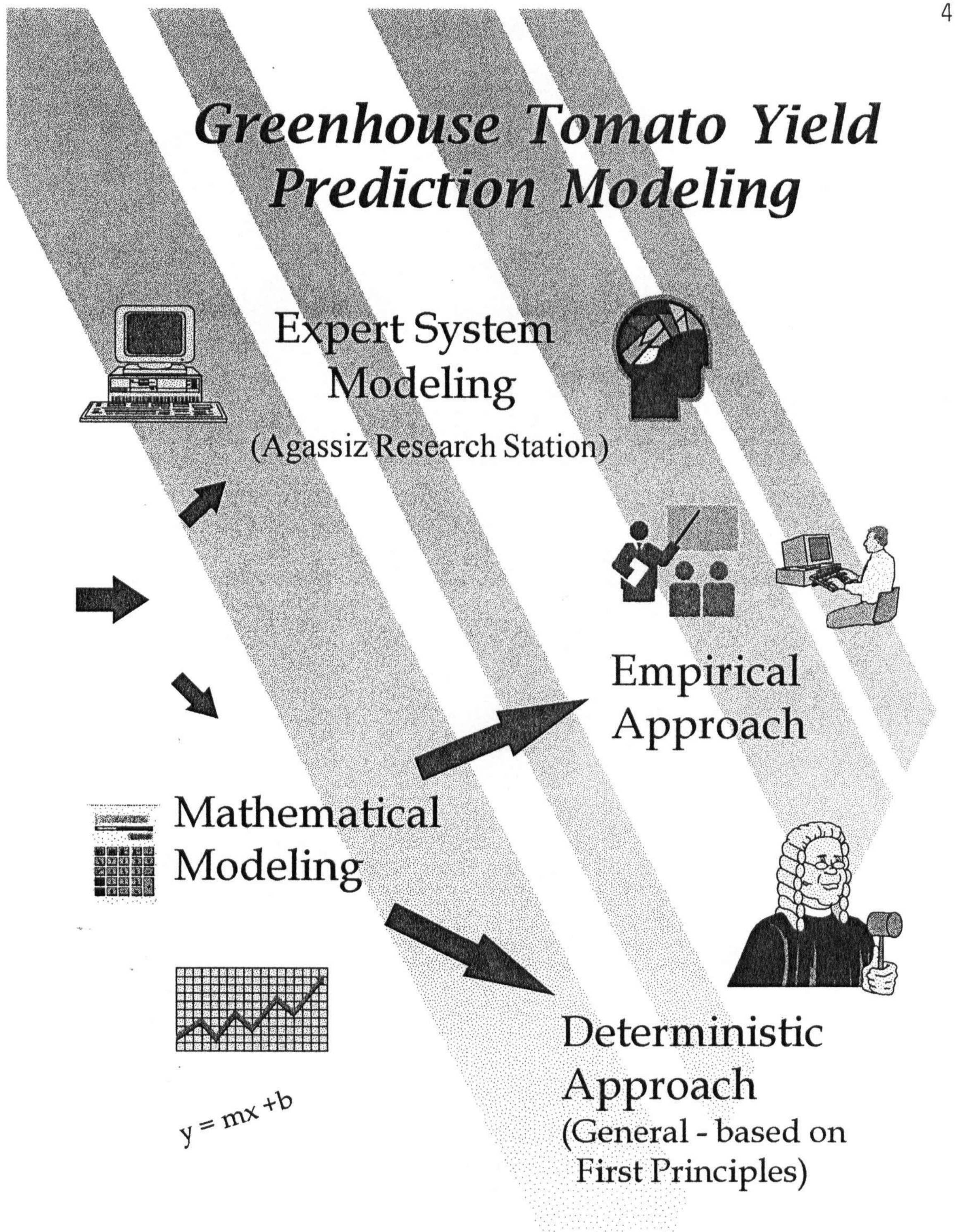



*Figure 1.* Shown above are the approaches that will be used to predict Greenhouse Tomato Crop Yield.

(Illustration by W.C.Tang, 1995)

The deterministic approach could also be useful in predicting yield. By applying the first principles that relate plant photosynthesis, respiration, and plant growth, estimates of yield could be made. Deterministic equations involve systems of differential equations to describe dynamic processes such as plant growth. If plant models are developed using this approach, each reaction between the seedling to harvest stage is described by a differential equation (ie. change in dry matter production). However, as plant growth typically involves long time spans, most studies limit their scope for testing plant models to only a few months (Takakura, 1993). The ultimate goal of the deterministic model is to reach the best possible agreement between the model predictions and the experimental data over a set time period.

Empirical models examine historical data and use mathematical regression techniques to find correlations between independent and dependent variables. In this case, the independent variables would be the environmental parameters such as greenhouse temperature, light, and humidity and the dependent variable would be the actual yield recorded for that period. Empirical models search for basic cause and effect relationships to simplify a complicated process into a functional equation which can be applied to the general case. The problem that exists for equations derived purely from an empirical approach is that they are generally linear-based (straight line equations) while plant growth is an example of a non-linear, dynamic system (Takakura, 1993). As a result, empirical models are restricted to providing good estimates within the range of data they were derived from.

Beyond the range of values found in the experimental case, empirical models can only provide extrapolations as a basis for predictions.

## 1.2 The Application of Expert Systems for Yield Prediction

An innovative approach to modeling involves the application of expert system models in greenhouse crop production (***Figure 1***). An expert system model is simply a method of processing information that involves a self-learning model that adapts to changing data inputs (ie. environmental parameters). Expert systems are based on the most up-to-date '*expert*' information in a field of study and incorporate this knowledge in decision-making processes and weighting factors (Levine et al, 1990). For predicting tomato crop yield, expert information would include up-to-date research findings of plant-climate relationships from greenhouse growers and plant physiologists. This expert knowledge would then be incorporated into a computer program by applying weighting factors to measured environmental parameters such as light, temperature, and humidity (***Figure 2***).

By relating the weighted environmental parameters to recorded actual yields, an expert system model can provide yield predictions. The advantage of using expert systems for yield prediction is their potential ability to provide nonlinear systems control for cases such as plant growth and their ability to forecast yields even in cases where data sets are incomplete or ambiguous. This is accomplished by assigning weights ($W_i$) to all selected data inputs ( $X_i$ ) (ie. light, temperature, humidity values) based on expert (greenhouse growers, plant physiologists) knowledge and then applying these weighted data as input for

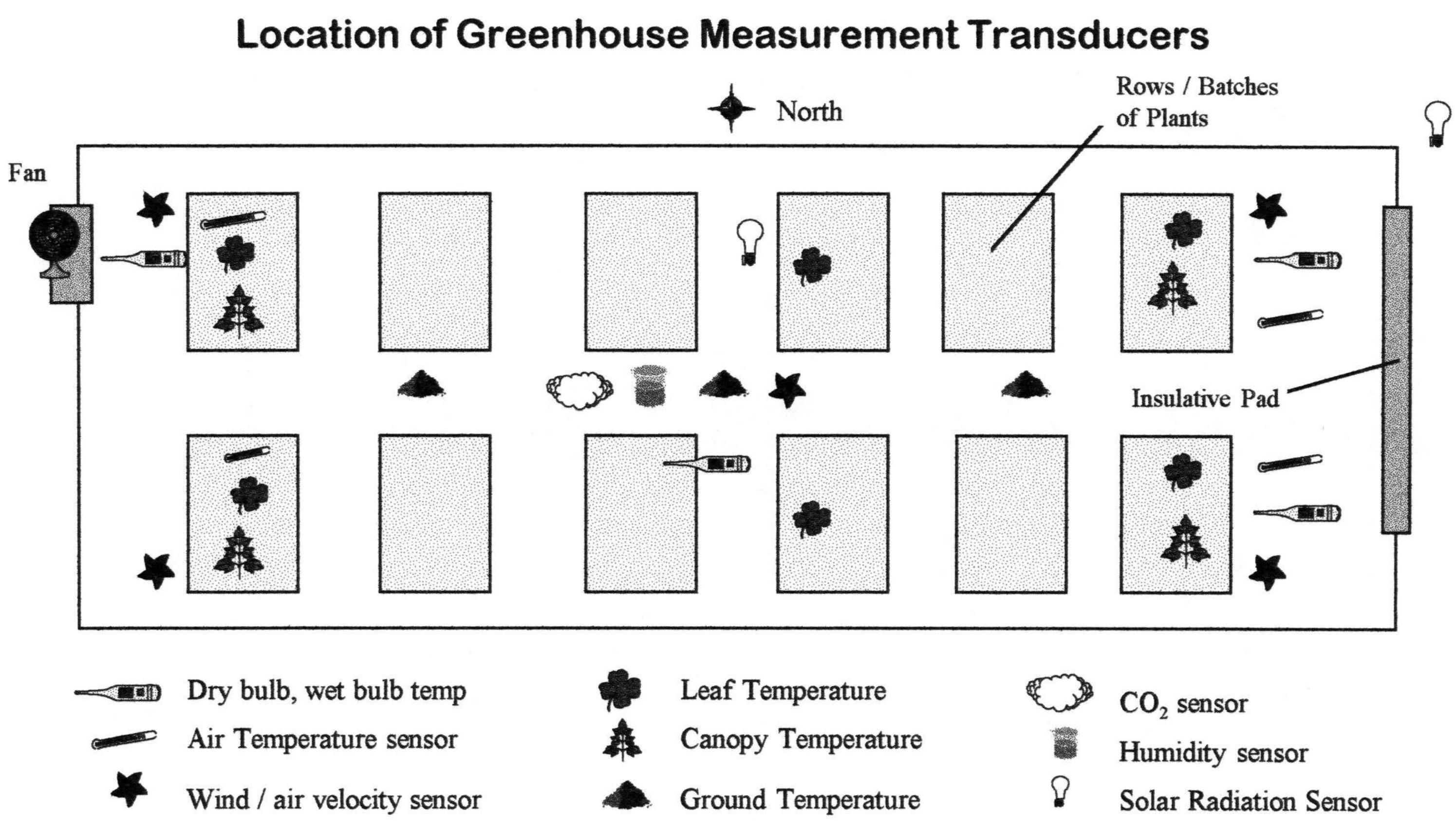



***Figure 2***. Typical Location of Greenhouse Measurement Transducers (Willits, 1993) Illustration by WCTang

a processing element (internal algorithm) which in turn produces one output value. In reaction to the data input, the model will then gradually adjust the weighting of each factor to produce more accurate and reliable predictions and minimize the error for the next iteration. By repeating this process, expert systems learn to recognize trends in the data inputs and are able to make predictions based on historical cases. Ideally, a properly trained expert system would then be interfaced with an on-line greenhouse climate control system to determine the best environmental setpoints to maximize future crop yields and minimize operating costs.

Expert system models have recently been tested in determining ideal temperature setpoints in greenhouses (Lacroix, et al., 1993). In this case, an artificial expert system temperature controller was created to predict crop responses to changes in greenhouse setpoint temperatures. By examining and testing various temperature setpoints with this technique, researchers were able to reduce energy consumption by 10% while maintaining productivity.

In another study, expert system methods were combined with standard data analysis techniques to predict apple quality (Boucherau et al., 1992). Measurements of apple sugar content obtained from infra-red spectrometry were used as a basis for determining quality. By employing standard regression techniques, a simple linear model was produced. An expert system was then employed to handle the non-linear data components and this improved the accuracy of apple quality predictions by five per cent.

Expert systems have also helped determine suitable environmental settings for greenhouse lettuce crops (Seginer et al., 1992). Historical data was used to train an expert system to aid in management decisions and determine environmental setpoints (ie. temperature settings).

When compared, mathematical models provide reasonable linear projections of future values but are less flexible than expert system models in terms of overall application.

Mathematical models require many parameters for the submodels that account for various physiological growth and development processes (***Figures 6a & 6b***). These parameters need to be determined by scientific experiments or at the very least approximated.

Uncertainties are therefore bound to be associated with mathematical models. Conversely, expert systems focus on trend and pattern analysis eliminating the need to approximate environmental coefficients to provide predictions.

Expert systems are able to adapt and learn by continually analysing trends to minimize the error with each subsequent prediction.

### 1.3 Research Objectives

The main objective of this study is to adapt (deterministic model) and develop a model (empirical model) that can successfully predict greenhouse tomato yield, of the tomato fruit itself, in terms of quantity (kg/m$^2$).

Steps (***Figure 3***) required to achieve the research objectives include:

1. To collect and analyse historical environmental and crop yield data from greenhouse growers via graphical hourly plots of greenhouse data to identify any obvious trends, patterns, or relationships.

2. To develop and test a deterministic yield prediction model (one based on first principles and known yield conversion factors).

3. To develop an empirical yield prediction model from multiple regression analysis of greenhouse data. The best preliminary model will then be selected for further development to determine the best overall empirical model.

4. To analyse yield predictions from an alternative, expert system model adapted by the Agassiz Research Station (Lin, 1994).

5. To compare and validate the empirical, deterministic, and Agassiz expert system yield prediction models. The best overall prediction model will then be identified.

## 2.0 Materials and Methods

### 2.1 Data Source

All of the data used in this study will be collected from growth records of **greenhouse tomatoes** provided by independent growers and the B.C. Greenhouse Growers Co-operative.

# Research Objectives

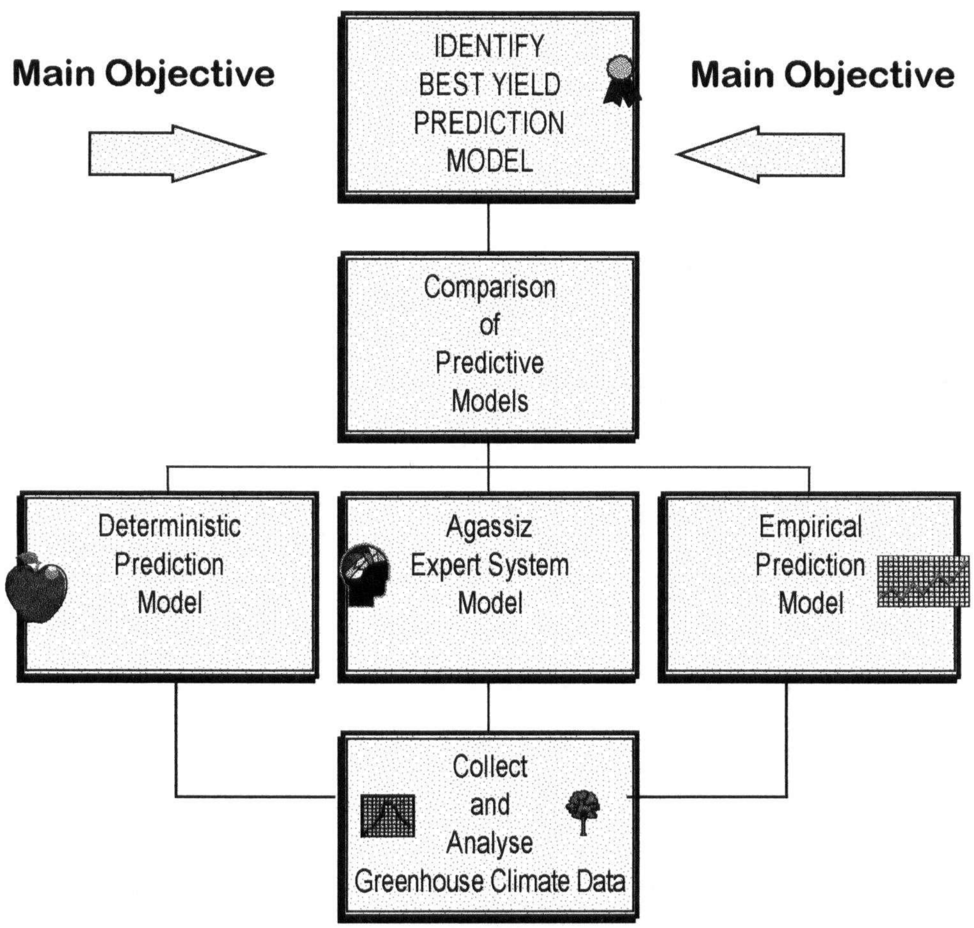



**Figure 3.** Shown above are the research objectives in this study.

## 2.2 Data Collection

Temperature, carbon dioxide, humidity, and solar radiation data used in this study were collected by a greenhouse climate computer with measurement transducers. Typical locations for these transducers within a greenhouse environment are shown on ***Figure 2***.

## 2.3 Data Analysis

The environmental data required for developing the various mathematical and expert system models will be obtained from two independent tomato greenhouse growers. In order to maintain the confidentiality of their information, data shall be identified as either from Greenhouse Grower 'A' or Greenhouse Grower 'B' (***Figure 4***).

**Greenhouse Grower A's data set includes hourly environmental values (values for every hour of the day)** for light intensity, greenhouse temperature, inside relative humidity, carbon dioxide concentrations, and actual recorded yield for a number of growing seasons during 1992. Hourly data is necessary for developing a deterministic mathematical growth model based on first principles (ie. photosynthesis and respiration equations) as the model attempts to simulate actual crop growth in order to provide yield predictions.

# *Data Collection & Analysis*

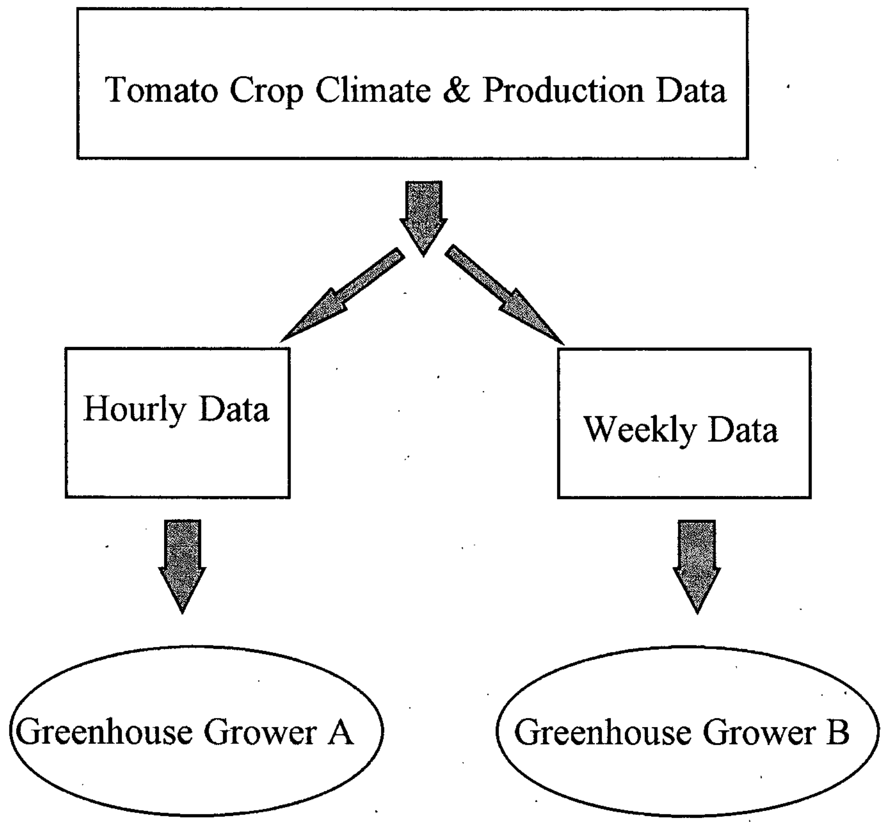



***Figure 4***. Shown above are the data sets used in this study

One complete set of greenhouse tomato hourly climate data will be plotted for a 60-day period (1440 hours) during the 1992 growing season. Graphical plots of: inside temperature, outside temperature, solar intensity (light), relative humidity, and carbon dioxide concentrations, will be analysed for any obvious trends or patterns in greenhouse conditions or management practises.

**Greenhouse Grower B's data set includes weekly values** for: light intensity (weekly cumulative total in $W/m^2$), average temperatures (a weekly average based on every hour temperature readings), day and night temperatures (weekly averages of hourly day-time and night-time temperatures, respectively), day and night humidity (weekly averages of hourly day and night humidity readings, respectively) and weekly recorded actual yield values ($kg/m^2$) for the years between 1987 to 1993 (excluding 1988 due to missing data). Weekly data from grower B will be used in this study to prepare an empirical math model using mathematical regression methods. The same weekly data from grower B was also used in testing the expert system model developed by the Agassiz Research Station (Agassiz, B.C.), a Research Branch of Agriculture Canada. The weekly predictions from the Agassiz expert system will then be used for comparison purposes.

## 2.4 Selecting the Deterministic Math Model

Using the hourly data provided by Greenhouse Grower A, a deterministic crop growth model will be developed to predict yield (***Figure 5***). The first principle equations that will be used for this approach are mathematical relationships between environmental

# Development of the Deterministic Model

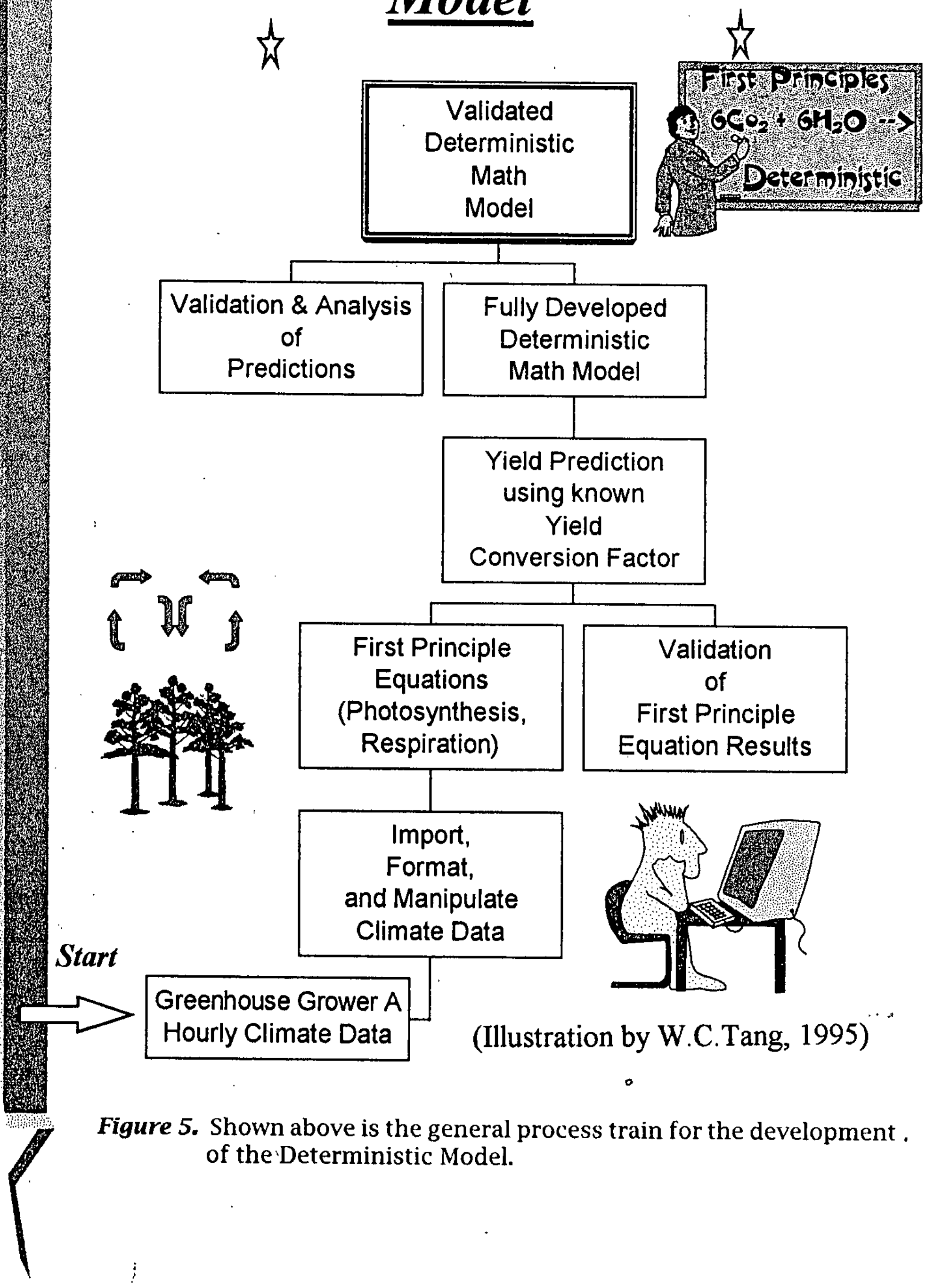



***Figure 5.*** Shown above is the general process train for the development of the Deterministic Model.

conditions, dry weight growth, respiration rate, and photosynthesis as detailed in literature (Jones, 1991) and shown on ***Figures 6a & 6b***. An alternative method to compute crop yield is by applying a multiplication factor that relates net photosynthesis to crop yield. For tomatoes, this was found to be 7 (Enoch, 1978) with the following assumptions: greenhouse light transmission of 80%, plant production based on $CO_2$ uptake at optimal leaf temperature, and that one absorbed $CO_2$ molecule is used to create one molecule of dry matter ($CH_2O$), with 50% of this dry matter being fruit yield, representing between 5% - 7% of the total fruit weight including both wet and dry portions (Ho and Hewitt, 1983) (***Figure 6c***).

In order to narrow the scope of this study, only hourly data from the 1992 growing season for Grower A will be used. The 1992 environmental data contained the most complete set of readings for greenhouse climate: light intensity, temperature, and carbon dioxide concentrations. (The deterministic math model will not utilize humidity data because transpiration is not being modeled.) Using a computer spreadsheet program, the appropriate column headings for the data will be configured and formula cells based on first equations found in literature (Jones, 1991) will be applied. Predictions for respiration, gross photosynthesis, net photosynthesis, and crop dry weight are then obtained. The results of net photosynthesis can now be converted to crop yield (***Figure 6c***) using the generalized multiplication factor of 7 (Enoch, 1978). Weekly cumulations of net photosynthesis based yield predictions will then be plotted and compared with actual recorded tomato yields for 1992. Once these immediate yield results are validated and corrected for missing data, the model will be used to predict yield 4 and 6 weeks ahead.

## $P_g$ - Gross Photosynthesis Equation

$$P_g = D \frac{\tau * C * p(\theta)}{K} * \ln\left[\frac{\alpha * K * I_o + (1-m) * \tau * C * p(\theta)}{\alpha * K * I_o * \exp(-KL) + (1-m) * \tau * C * p(\theta)}\right]$$

where $P_g$ = gross photosynthesis measured in units of $g\ (CH_2O)\ /\ m^2$-h

$D$ = coefficient to convert photosynthesis calculations from $\mu mol\ (CO_2)\ /\ m^2$-s to $g\ (CH_2O)\ /\ m^2$-h
$\tau$ = leaf conductance to $CO_2$, $\mu mol\ (CO_2)\ /\ m^2$-s
$C$ = $CO_2$ concentration of the air, ($\mu mol\ (CO_2)\ /\ mol\ (air)$) = ppm
$p(\theta)$ = dimensionless function of temperature, the effect of temperature on the maximum rate of photosynthesis for a single leaf, as a quadratic equation:

$$p(\theta) = \{ 1 - \left( (\theta_h - \theta) / (\theta_h - \theta_l) \right)^2 \}$$

where: $\theta_h$ is the temperature at which leaf photosynthesis is maximum, °C
$\theta$ is the measured temperature in the greenhouse, °C
$\theta_l$ is the temperature at which leaf photosynthesis is zero, °C

$\alpha$ = leaf light utilization efficiency, $\mu mol\ (CO_2)\ /\ \mu mol$ (photon)
$K$ = canopy light extinction coefficient
$I_o$ = light flux density at the top of the canopy, $\mu mol$ (photon) / $m^2$-s
$m$ = light transmission coefficient of leaves
and $L$ = canopy leaf area index, $m^2$ (leaf) / $m^2$ (ground)

## $R_m$ - Maintenance Respiration Rate

$$R_m = k_m * \exp(0.0693\ [\ \theta - 25\ ])$$

where $R_m$ = maintenance respiration rate, $g\ (CH_2O)\ /\ g$ tissue - h (maintenance respiration is the loss of $CO_2$ due to breakdown and re-synthesis of existing tissue and depends on temperature)
$k_m$ = respiration rate at 25°C, $g\ (CH_2O)\ /\ g$ tissue - h
$\theta$ = measured greenhouse/inside temperature, °C

***Figure 6a.*** Shown above are the $P_g$ and $R_m$ equations used in developing the deterministic math model (Jones, 1991).

## dW/dt - Crop Dry Weight Growth Rate

$$dW/dt = E\,(P_g - R_m * W)$$

| | | | |
|---|---|---|---|
| where | $dW/dt$ | = | rate of dry weight of the crop, g (tissue) - h |
| | $W$ | = | total plant dry weight, $g / m^2$ |
| | $E$ | = | conversion efficiency of $CH_2O$ to plant tissue, g(tissue) / g($CH_2O$) |
| | $R_m$ | = | maintenance respiration rate, g ($CH_2O$) / g(tissue) - h |
| and | $P_g$ | = | canopy gross photosynthesis rate, g ($CH_2O$) / $m^2$ - h |

## Pnet - Net Photosynthesis

$$P_n = P_g - R_m * W$$

| | | | |
|---|---|---|---|
| where | $P_n$ | = | net photosynthesis rate, g ($CH_2O$) / $m^2$ - h |
| | $P_g$ | = | canopy gross photosynthesis rate, g ($CH_2O$) / $m^2$ - h |
| | $R_m$ | = | maintenance respiration rate, g ($CH_2O$) / g(tissue) - h |
| and | $W$ | = | total plant dry weight, $g / m^2$ |

## Coefficient Values used in these Equations

$\alpha = 0.056$ $\theta_h = 30°C$, $\theta_l = 5°C$ $\tau = 0.0664$ $k_m = 0.0006$ $m = 0.10$ $D = 0.108$

$E = 0.70$ $K = 0.58$ L = varies from 0.6 to 3.31 (de Koning, 1993)

***Figure 6b.*** Shown above are the dW/dt and $P_n$ equations, and coefficient values used in developing the deterministic math model (Jones, 1991).

# Photosynthesis:

$$6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$$

*Simplified Carbon Conversion for determining Yield Conversion Factor:*

$$1\ CO_2 \rightarrow 1\ CH_2O \text{ molecule}$$

$$\downarrow \qquad\qquad \downarrow$$

(MW) = 44 g (MW) = 30 g
(MW) = Molecular Weight

- For 1 molecule of $CO_2$ used, (ie. every 44 g of $CO_2$ used) we get 30 g of $CH_2O$.
- 1/2 of this 30 g of $CH_2O$ is partitioned to become fruit yield (dry portion) = 15 g.
- This 15g is 5% of the total weight of the fruit (wet and dry portions)
- Therefore, the total fruit weight is actually: 15 g * (100 / 5) = 300 grams
- For 44 g of $CO_2$ used (absorbed) we have 300 g of fruit weight.

∴ *1 g of $CO_2$ used leads to ( 300 g / 44g ) = 7 g of fruit weight (wet & dry).*

---

Enoch's Generalized Multiplication Factor for Tomato Yield (Enoch, 1978)

---

$$\frac{30\text{ g}}{44\text{g}} * \frac{1}{2} * \frac{100}{5} \simeq 7 = \text{Yield Multiplication Factor for Tomatoes}$$

- Molecular Weight for $CO_2$ (44 g) and $CH_2O$ (30 g)
- Half the dry matter is yield (1/2).
- This dry matter represents 5% of total fruit weight (dry & wet) (100/5).

---

***Figure 6c.*** Shown above is Enoch's Yield Conversion Factor (1978) for tomatoes that will be used in developing the deterministic math model.

Predicted and actual yield values will then be compared with the standard error of estimate ($S_e$) or root-mean-square-error (RMSE) method (***Figure 9***) for evaluating the accuracy of results using the deterministic modeling approach.

### 2.5 Developing the Empirical Math Model

Weekly light and temperature data from Greenhouse Grower B will be used in this portion of the study to develop an empirical prediction model using multiple regression techniques (***Figure 7***). Regression techniques are useful for deriving linear relationships between input and output variables. Such analysis will be carried out in the advanced math tools available in spreadsheet programs such as Quattro Pro (for simple regression) as well as with more advanced statistical analysis packages such as Systat (for multiple and stepwise regressions).

The first step in developing the empirical model involves identifying the primary environmental condition that is most closely linked to tomato crop production. From literature review McAvoy et al. (1989), have found strong relationships between photosynthetic photo flux (light intensity) and total yield. In this study they performed comparisons and regressions of the total photosynthetic photon flux (PPF) received during a 60-day production cycle. A close linear trend ($r^2 = 0.947$) between recorded tomato yields and total PPF was observed.

Following a similar approach, the weekly data for 1987 to 1993 from Greenhouse Grower B will be analysed for a light/yield linear relationship. Nine week (63 day) cumulative light intensity (PPF) values will be plotted against recorded yield. This method

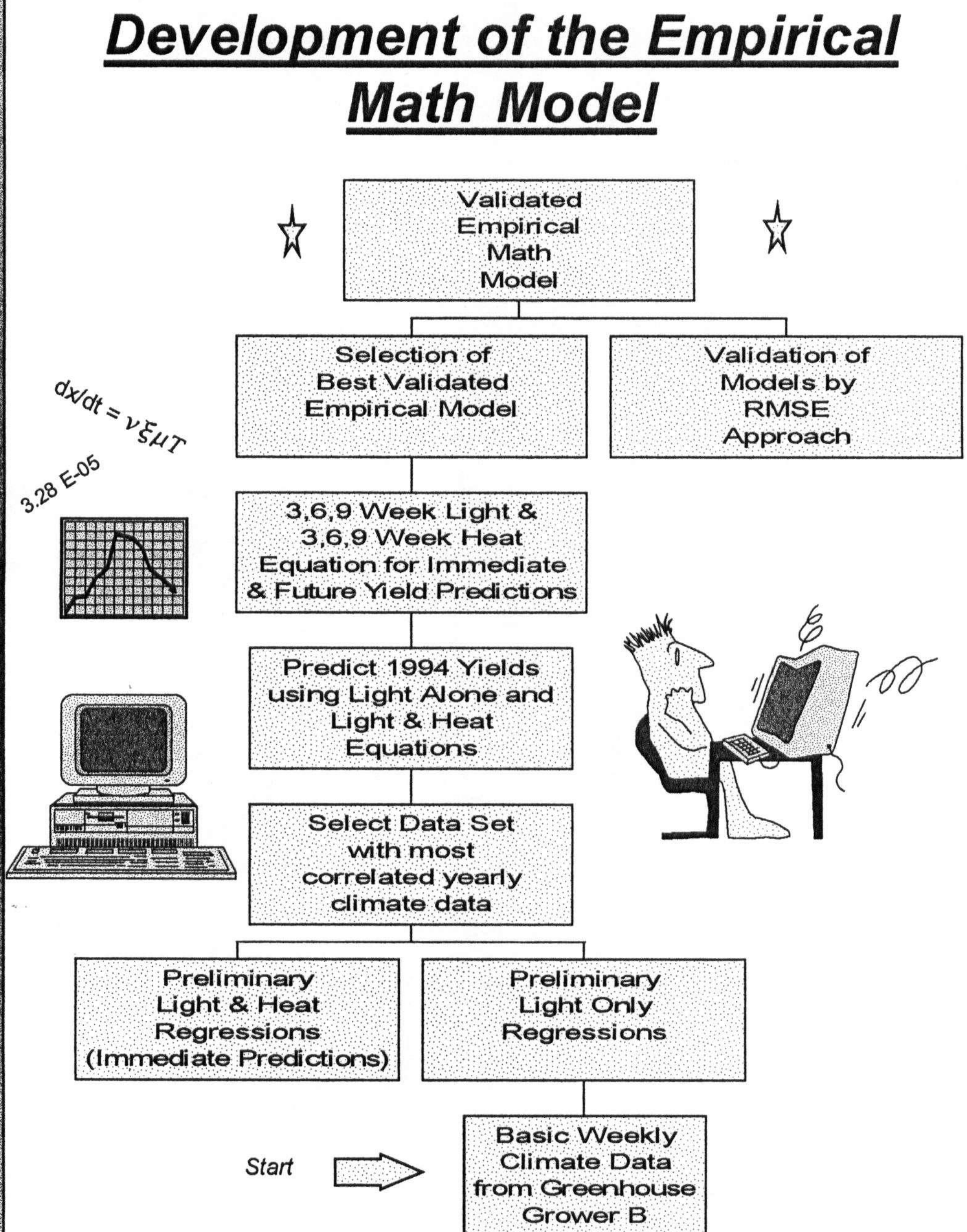



*Figure 7.* Shown above is the general process train for developing the empirical math model.

will identify the yearly data that has the closest relationship between 9-week cumulated PPF and yield (in g/plant), as indicated by the statistical $r^2$ , and derive a set of linear mathematical expressions for each year of data as well as for 6- year cumulative data (1987, 1989, 1990, 1991, 1992, 1993, excluding 1988 data). The $r^2$, or correlation coefficient, measures the linear association or clustering around a line with higher values indicating a closer fit to actual yields (Freedman et al, 1991).

Extending the light and yield relationship further, a second set of regression results will be performed by cumulating light data for 9 weeks ($CL_9$) in units of $W/m^2$ and regressing these totals against recorded yield in units of ($kg/m^2$). These linear equations based on 9-week summations of light will then be used to predict 1994 crop yields based on light readings recorded for 1994. The actual 1994 yields and immediate weekly predicted yields from these equations will then be plotted on the same graph and compared. A standard error of estimate ($S_e$) or root-mean-square-error analysis (RMSE) will also be performed on each set of predictions versus actual yield values (***Figure 9***) to determine the equation that most closely approximates the actual yield.

The next step in developing the empirical model will be to investigate the relationships between yield and other recorded environmental parameters. Tomato plants have been found to be very responsive to changes in temperature and light levels and the status of these conditions strongly affects carbon fixation and dry matter partitioning into fruit (Jones et al, 1991).

Humidity also affects tomato growth but to a lesser degree, and a single trial run will be included development of the empirical model to examine its effectiveness as a factor in yield predictions, research time-permitting.

As a further step in the empirical analysis, the effect of temperature will be entered into the regression model.

The general rule for tomato production is that the 24-hour greenhouse temperature is generally responsible for the rate of growth, with higher average temperatures increasing growth rates. The target setpoint temperature for tomatoes reported in the 1993 Greenhouse Vegetable Production Guide is 19 °C (B.C. Ministry of Agriculture, 1993).

Below 10°C, little tomato growth / photosynthetic activity occurs. Provided with this information, a cumulation of temperature in terms of heat units (Wolf et al., 1986) will be performed. Each heat unit (HU) is defined as the difference between the daily 24-hour average temperature and a baseline temperature of 10 °C. The first trial run using this approach will involve cumulating 9 weeks of recorded temperatures and subtracting the baseline 10°C from each value ( 9 weeks of values x 10°C = 90°C subtracted from 9 week cumulative total of temperature readings). These adjusted values will then be referred to as 9-week cumulative heat units ($CH_9$). By performing a multiple step-wise regression analysis on both the 9 week heat units ($CH_9$) and 9-week cumulative light values ($CL_9$), linear relationships involving heat, light, and yield will result. This approach will provide immediate yield predictions and provide an indication as to which factor (light or temperature) plays a bigger role in tomato yield prediction.

Continuing on with the study, the yearly data set that contains the best linear relationship with light (highest $r^2$ - value from 9-week light (PPF) regression) will be further analysed. Combinations of the environmental factors; 3, 6, and 9 week cumulative heat units ($CH_3$, $CH_6$, $CH_9$) and 3, 6, and 9 week cumulative light values ($CL_3$, $CL_6$, $CL_9$) will be regressed to determine the best prediction equation ($Y_p$) for yield ($Y_p = f\{CH_3, CH_6, CH_9, CL_3, CL_6, CL_9\}$). However, at this point of the analysis process instead of using the model to produce immediate weekly yield values, the model as a true predict-ahead model.

The initial regression results of empirical model identified the most correlated data, and whether light or temperature was the greatest influencing factor on yield. However, an empirical model that uses historical data to predict crop yield a few weeks ahead is necessary if the model is to be more flexible for growers and the Greenhouse Growers Co-operative Association. By using the cumulated values of light and temperature and regressing them with actual yield results recorded four weeks ahead, a four-week ahead predictive model can be obtained.

As a final step of the empirical approach to yield prediction, the 1994 yield will be predicted 4 weeks ahead by an multi-variable linear equation that combines 3,6, and 9 week cumulative light and 3, 6, and 9 week cumulative heat data. This equation will be evaluated by standard error analysis ($S_e$ or RMSE) for usefulness and flexibility of the model to the greenhouse grower to identify the best overall empirical model.

## 2.6 The Agassiz Expert System Model

The results from an expert system, tomato yield prediction model, developed by the Agassiz Research station (Lin, 1994) will be analysed and ***used solely as a comparison model*** for the two mathematical models developed in this study.

---

***The specific methodology used to develop the Agassiz expert system is currently being licensed and must remain confidential as part of a collaborative agreement to protect intellectual rights between the Agassiz Research Station, the University of British Columbia, and the participating members of the greenhouse industry.***

---

However, an overview of the general methodology used to develop the Agassiz expert system is presented on ***Figure 8***. First, the same weekly environmental data (actual yield, light intensity, temperature, and humidity readings from 1987 to 1993) from Greenhouse Grower B used in developing the empirical math model, was imported into the expert system. Then by using up-to-date ***expert*** knowledge from research findings, greenhouse growers, and plant physiologists (Lin, 1994), weighting of environmental variables as they related to yield were applied in training the expert model.

The advantage to the Agassiz expert system is that as the database of expert knowledge grows, the model can be modified to reflect these new findings by adjusting the environmental weighting factors. The expert system is also a self-learning system, in that it recognizes trends between environmental variables and actual yield, and adjusts the weighting factors to reduce the error with each subsequent yield prediction.

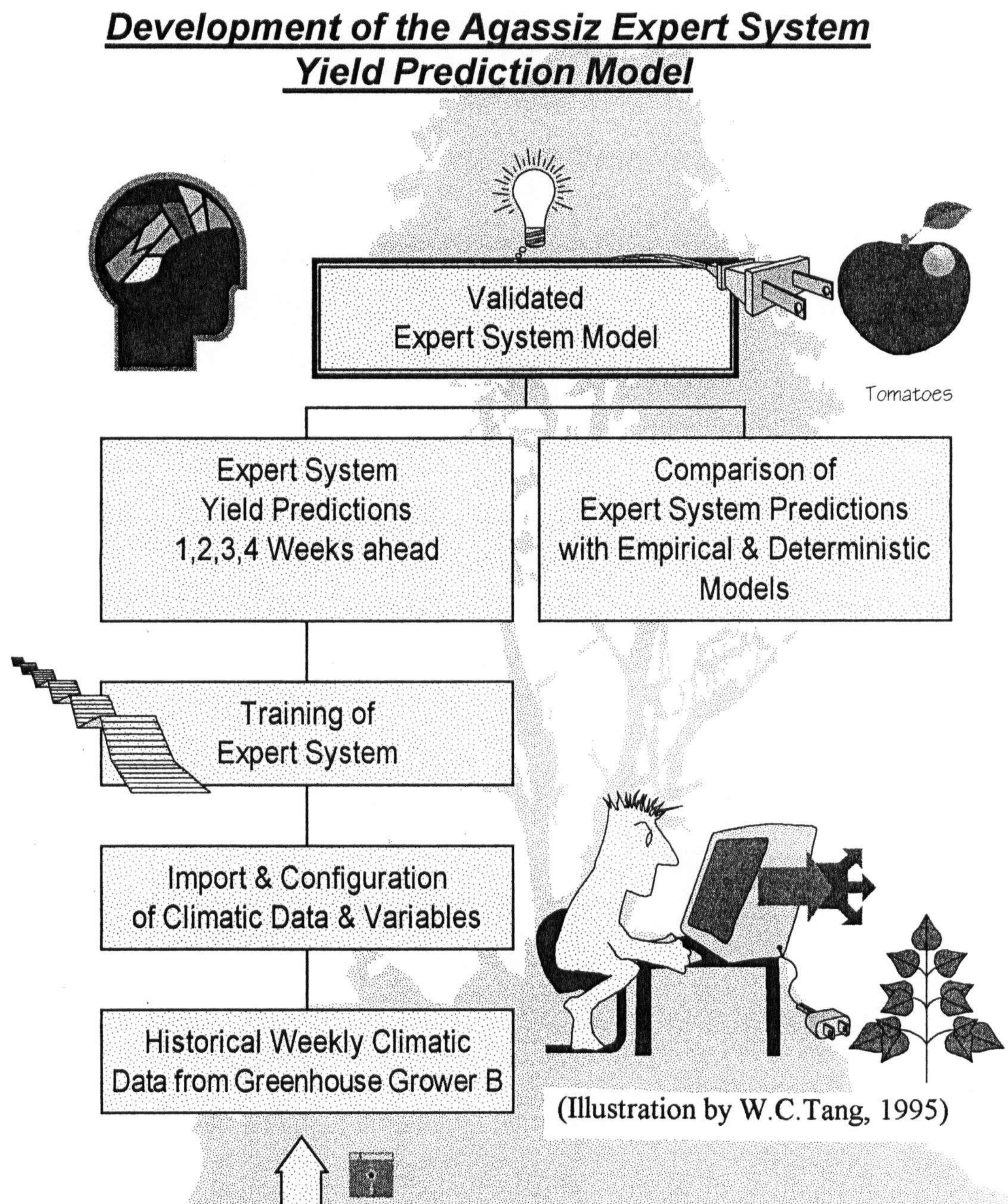



*Figure 8.* Shown above is the general process scheme for developing the Agassiz Expert System yield prediction model.

The researchers at Agassiz have configured their expert system to provide 1 week, 2 week, 3 week, and 4 week ahead predictions of **1994 yield** and these results will then be validated by comparison with actual recorded yields by the RMSE approach. The results from the Agassiz expert system will also be compared with the predictions from the mathematical models developed in this study.

## 2.7 Validation and Comparison of the Yield Predictive Models

In order to be a useful management tool for greenhouse growers, crop growth and yield prediction models must be validated before they are used (de Koning, 1993a). The validation of the deterministic, empirical, and Agassiz expert system model can be performed by analysing the magnitude of the error in their predictions.

Calculation of the standard error of estimate ($S_e$) (Alder and Roessler, 1960) or root-mean-square-error (RMSE) for prediction results measures the overall size of differences between predicted and actual values for yield (Freedman et al, 1991 and Kozak, 1995).

The $S_e$ or simplified RMSE equation used in this study is shown on ***Figure 9***. The magnitude of these error values depends on the range of values being compared and the number of predicted values. With a larger range and greater number of values, generally larger error values are expected. In this study, typical tomato crop yield results range from 0 kg/m$^2$ to 4 kg/m$^2$, with a minimum of 30 predicted yield values provided by each model. To provide a basis for validation and comparison, models providing error values of less than 0.8 (kg/m$^2$) for a year of predictions (approximately 50% error for an average 1.6 kg/m$^2$ yield throughout the year) will be considered valid models in this study.

## Root-Mean-Square-Error Analysis (RMSE)
## or
## Standard Error of Estimate ( $S_e$ )

$$\text{RMSE or } S_e = \sqrt{\frac{\sum_{i=1}^{N}(\text{Actual - Predicted})^2}{N}}$$

- $S_e$ = Standard Error of Estimate, $kg/m^2$
- RMSE = Root Mean Square Error, $kg/m^2$
- Predicted = Predicted Yield Value, $kg/m^2$
- Actual = Actual Yield Value, $kg/m^2$
- N = Total Number of Predictions

***Figure 9***. Shown above is the $S_e$ or RMSE method used to compare the predictive models in this study. Lower values indicate less error and better predictions.

(Freedman et al, 1991. Alder & Roessler, 1960. and Kozak, 1995)

The lowest RMSE or $S_e$ values indicate the least error between predicted and actual values (ideally an error value of 0 kg/m$^2$), and best predictive performance.

The best overall model will be determined by the accuracy of predictions (lowest RMSE or $S_e$ values) for an entire growing season and during periods of peak productivity (RMSE for a specific peak interval), flexibility of the model for application in site-specific and general cases, and overall user-friendliness.

## 3.0 Results and Discussion

### 3.1 Analysis of Graphical Plots of Hourly Greenhouse Climate Data

Shown in Appendix A are graphical plots of inside greenhouse temperature, outside temperature, solar intensity, carbon dioxide concentrations, and relative humidity for a 60-day growth period (between January 1, 1992 to March 5, 1992).

Trends observed during the 60-day growth period included the maintenance of a fairly consistent inside greenhouse temperature between the 18 to 23°C range, with peaks in daily inside temperature corresponding directly to peaks in solar intensity (***Figures A-1, A-2, A-3 in Appendix A***). Solar intensity peaks are recorded for noon periods each day.

Guidelines from the 1993 Greenhouse Vegetable Production Guide indicate average 24 hour temperatures as being responsible for the rate of growth, with higher averages leading to faster growth rates. The target setpoint temperature for tomatoes is currently 19°C (B.C. Ministry of Agriculture, Fisheries, and Foods, 1993). Successful environmental temperature control for this plotted growth period is evidenced by the minimal fluctuations even during periods of outside temperature extremes including sub-zero periods.

Plots of carbon dioxide concentrations for the same period (***Figures A-4, A-5, A-6 in Appendix A***) revealed a range between 200 to 2000 parts per million (ppm), which is typical enrichment for a tomato crop (Gould, 1983). Concentrations frequently reached their peaks during the noon hour of each day, and showed a daily maximum fluctuation of about 1800 ppm. Inside venting for temperature control can be held accountable for part of the variability in carbon dioxide concentrations.

Relative humidity readings plotted for the same growth period showed peaks and valleys approximately every ten hours, with peak humidity during late afternoon and evening hours. Humidity was found to vary greatly between 60 to 95 percent for the early part of the growth period (January 1 to January 24, 1992) with less variation occurring during the following two months (January 25 to March 5, 1992). Low humidities for a few hours a day will not hinder the overall climate balance within the greenhouse but levels should typically be maintained above 50% to prevent extreme moisture losses. Proper maintenance of high humidity levels (80% and above) will encourage stomatal opening and increased carbon dioxide uptake, leading to increased water evaporation, and cooling of the plant canopy (B.C. Ministry of Agriculture, Fisheries, and Foods, 1993).

The patterns discovered from these plots illustrate both the importance of climate control and the possibility of applying trend recognition techniques to predict these parameters. Trends in greenhouse climate can then be related to actual recorded yields for particular growth periods for the development of a yield prediction model.

### 3.2 Results of the Deterministic Math Model

Applying the first principle equations of photosynthesis, respiration, and dry weight growth rate detailed in literature (Jones, 1991) a deterministic math model was developed. Using the yield conversion factor of 7 described in literature (Enoch, 1978) hourly net photosynthesis calculations were converted into hourly yield predictions (ie. hourly $P_{net}$ values multiplied by 7). Monthly variations in leaf area index (LAI) throughout the year (***Appendix B***) affected the photosynthetic values and were corrected for accordingly. Highest LAI values were reported from literature (de Koning, 1993a) for the peak growth period between March to June.

A sample of the spreadsheet used in developing the deterministic math model is shown in ***Appendix B***. For graphical plotting of the results, the hourly yield predictions were summed into weekly cumulative totals before plotting versus week number. Week numbers correspond to an annual chronological scheme where week 1 represents the first week in January and where week 52 represents the last week of December.

The initial plot of immediate (same week) yield predictions compared with actual recorded yield (***Figure 10***) produced an good range of results. By cumulating net photosynthetic ($P_{net}$ ) rates for periods of 1 week (158 hours) and converting these totals to yields (1 Week C. $P_{net}$ Yield) a low range of yield predictions resulted. The magnitude of the predictions were approximately half of the actual recorded yield for the 1 week cumulations, indicating that a week of net photosynthetic activity may not be sufficient for predicting actual yields.

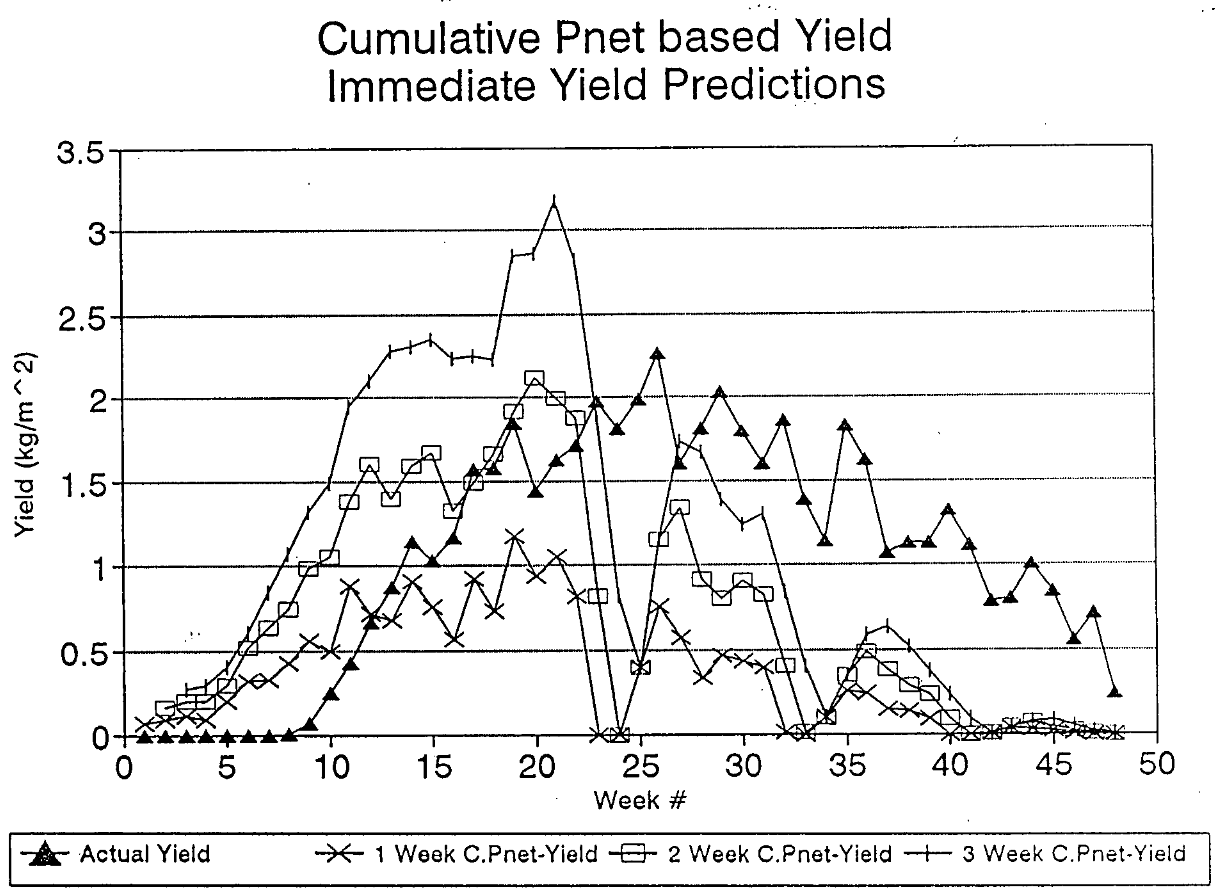



Figure 10 Shown above are the immediate yield predictions using the yield conversion factor applied to Pnet.

The idea of cumulating net photosynthetic activity and converting to yield was carried on further by cumulating 2 and 3 weeks of $P_{net}$ in order to obtain a better fit of the actual yield curve. The plotted results (***Figure 10***) revealed the closest correlation between a two week cumulation of $P_{net}$ and actual yield. For the first half of the year (up to week 23) predictions showed a good range of peaks simulating actual recorded yield. Predicted yield peaks were similar especially during the 10 to 20 week period after a short initial period of over-prediction (weeks 1 to 10). The over-prediction during the initial few weeks was expected as the deterministic model assumes that all net photosynthetic activity is being directly converted to yield. In actual greenhouse practise however, tomato crops are not being harvested for the first 8 to 9 weeks at the start of the growth season.

The general shapes of the curves for each of the weekly cumulation trials were similar and encountered a shared problem. In cases of missing or incomplete data sets, yield predictions cannot be made accurately. Yield predictions for weeks 23 to 25 and 32 to 35 show evidence of the problem that exists when required values for the climatic variables used in predictive equations are missing. The successful prediction and cumulation of $P_{net}$ activity relies on complete data sets. Calculated cumulative yields following periods of missing data lead to inaccurate and incomplete predictions due to zero value predictions being included in the weekly totals. These zero values are then carried forward to the next week of predictions and then the effect is magnified as these zero values dampen the results leading to under-predictions.

In order to compensate for those weeks of missing or ambiguous data, manual extrapolation of the predictions between those weeks with complete data had to be performed. The two intervals of weeks with incomplete or zero data (weeks 23 to 25 and 32 to 35) were corrected by extrapolating prediction lines between the two weeks with reliable data (weeks 22 to 26 and 32 to 36 respectively). The yield prediction corrections for missing data are shown as intervals connected with dotted lines on the future yield prediction plots that followed. In order for the deterministic prediction model to be useful for a grower, it must be able to forecast yields ahead of time.

Predictions of yield were made 4 weeks ahead of time using the same 1, 2, and 3 week cumulations of net photosynthetic converted activity (***Figure 11***). Better results were found for the 4 week ahead set of data than those reported for the immediate, same week predictions. In fact, by predicting 4 weeks ahead, the problem of over-prediction during the first nine weeks of the growing season was alleviated. The 4 week ahead shift in yield prediction also led to closer alignment of the predicted and actual yield peaks and valleys. In particular, the 2 week cumulation of $P_{net}$ showed the closest relationship to actual yield and was calculated to have a RMSE of 0.464 $kg/m^2$ for the full year of predictions. The dotted intervals between weeks 25 to 29 and 36 to 38 show the correction by extrapolation for missing data during these periods (***Figure 11***). However, correction for yield predictions in this manner was still not sufficient to maintain the close prediction trend experienced in the first 25 weeks plotted.

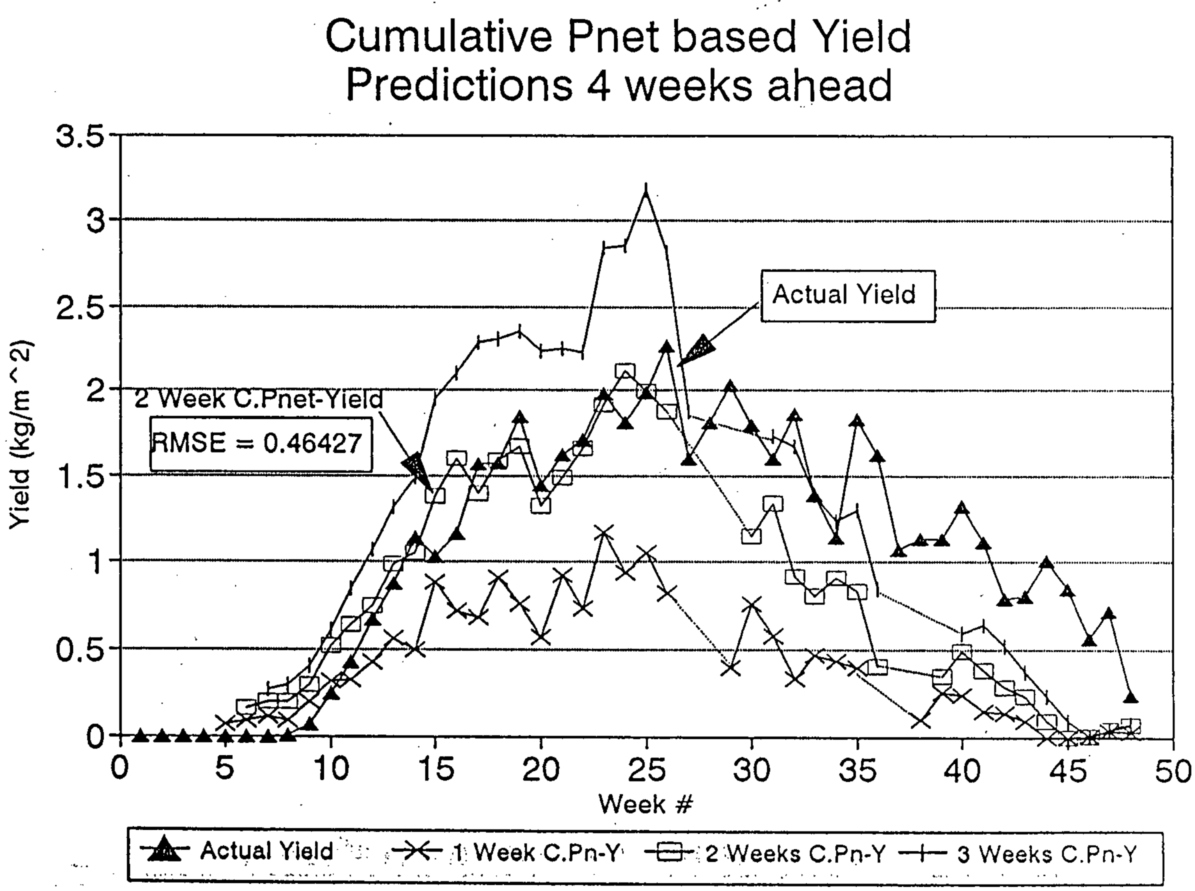



Figure 11 Shown above are the 4 week ahead yield predictions using the yield conversion factor applied to Pnet

The reason for the loss in accuracy and precision after the periods of missing data is that the deterministic model attempted to simulate actual, continuous growth of a tomato crop. Any discontinuities, or periods of missing data would not only affect that particular period, but also the summations leading to yield predictions for subsequent time periods. The effect of the missing data periods was even magnified further for the last quarter of predictions (weeks 35 to 52).

In a further attempt to provide predictions further ahead, yields were predicted 6 weeks ahead of time, again by shifting immediate yields forward and comparing by the RMSE method. Even better results were found for predicting 6 weeks ahead as shown on the plotted graph (***Figure 12***). The major peaks in yield became even more closely aligned with actual yields and the 6 week shifting reduced the prediction error of the extrapolated intervals with missing data. Root-mean-square analysis of the 2 week cumulative yield plot for 6 weeks ahead produced a value of 0.3818 $kg/m^2$, even closer than the 4 week ahead plot.

The use of two week cumulative $P_{net}$ values converted to yield for 6 week ahead predictions actually becomes a 8 week ahead forecast considering 2 weeks of historical data are required to make these predictions. Basis for cumulating two week $P_{net}$ activity as a yield predictor was found from the rate of starch accumulation of a tomato plant found from literature. Daily starch accumulation for tomato crops was found to reach its peak during the 15 to 25 day period after pollination (Ho and Hewitt, 1986).

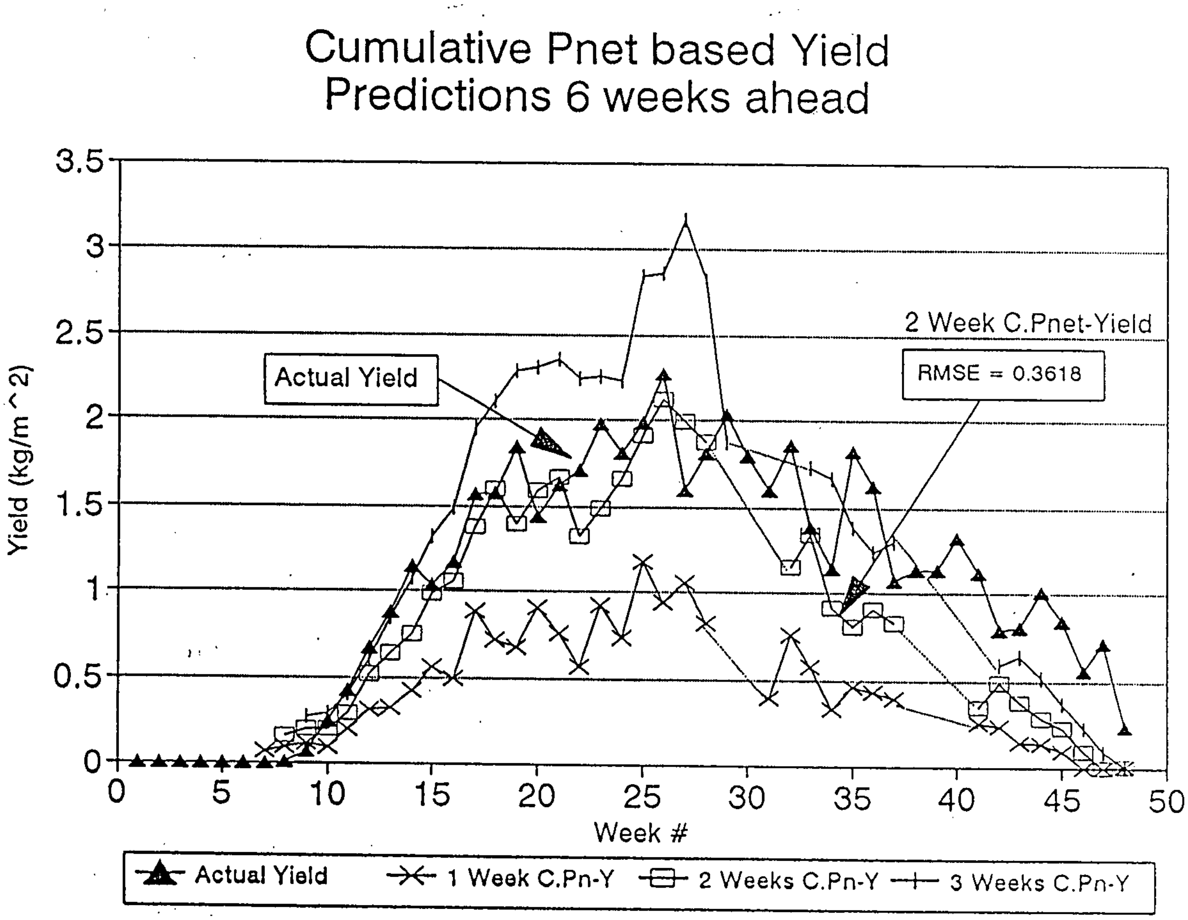



Figure 12 Shown above are the 6 week ahead yield predictions using the yield conversion factor applied to Pnet.

As the rate of starch accumulation may vary slightly for different tomato varieties, it is likely that the peak starch accumulation period for the tomatoes used in this study were near the two week (14 day) interval.

From literature, the peak of starch accumulation accounts for approximately 30% of the daily accumulated dry matter by day 20 after fruit pollination (Ho and Hewitt, 1983). Use of the two week $P_{net}$ results and converting with the assumptions (***Figure 6c***) built into the yield conversion factor (Enoch, 1978) for yield predictions may be justified. As much of the starch and dry weight growth of the tomato may have occurred by the two week rapid growth period, the conversion factor already has an assumption that only 50% of the dry matter is being converted to yield. With the 50% assumption of dry matter to yield conversion the magnitude of the prediction error is already being reduced as this ratio may vary under changing climatic conditions. For example, under conditions of very low light flux densities (below 40 $W/m^2$) no net photosynthesis was recorded for tomatoes of any size (Tanaka et al., 1974), however based on the spreadsheet structure of the deterministic model and first principle equations, a value greater than zero may be calculated.

One possible reason for the discrepancy between results in actual practise and in theory is that emphasis of the effects of light in the generalized photosynthetic equation may not be great enough. The effect of temperature, carbon dioxide concentrations, leaf area index, and light are all incorporated into the photosynthetic equation and each additional parameter may dampen the significant effect of light levels on $P_{net}$ activity. Consequently, when $P_{net}$ values are directly converted into yield predictions using the conversion factor, the deterministic model may over-predict during periods of low light levels and under-predict under high light conditions.

Other possible sources of error for the predictions is the effect of respiration upon the magnitude of $P_{net}$. It is known that respiration rates in mature fruits and young fruits vary greatly. Respiration rates in two-week-old tomatoes were reported to change from 0.4-0.6 mg $CO_2$ per gram of fresh weight per hour to 0.05 to 0.07 mg $CO_2$ per gram of fresh weight per hour as they reached maturity (Tanaka et al., 1974). The deterministic model developed here does not account for the variability of respiration rates due to fruit age and this may be a source of error for the predictions.

### 3.3 Results of the Empirical Math Model

Following studies that found a close correlation between photosynthetic photon flux (PPF) received during a 60-day production cycle and tomato yield (McAvoy et al, 1989), an empirical model was developed. Weekly solar radiation data from Greenhouse Grower B in units of moles/$m^2$ were plotted versus crop yield (g/plant) assuming a tomato plant density of 2.5 tomato plants / $m^2$. Results from these plots of total PPF for a 9-week (approximately 60-days) production period versus crop yield were plotted and linear regressions were performed to derive equations for each year of data.

A summary of the regression equations derived from the total PPF plots can be found in the appendix (***Table A-1 in Appendix C***). For each plot, nine weeks of PPF were cumulated and then plotted against yield (g/plant) to find the year with the closest correlation (***Figures A-11, A-12, A-13, A-14, A-15, A-16 and A-17 in Appendix C***). As well, for one plot, data from all 6 years (1987 to 1993 inclusive, excluding 1988 due to missing data) was plotted as a 1987-1993 graph to identify a general trend relationship between light and yield (***Figure 13***). The 1987-1993 plot had a calculated $r^2$ value of 0.668 which indicates a fairly consistent and direct relationship between light and yield for 6 years of historical data.

From $r^2$ analysis, or comparison by the correlation coefficient, the 1992 data set showed the closest fit with the highest $r^2$ value of 0.746. The higher the $r^2$ value (with maximum value being 1) the better the linear fit between the independent variable of light (PPF) and dependent variable (yield).

The yearly scatter plots showed a fairly good linear trend between light and yield with $r^2$ results ranging from 0.61 to 0.746. The $r^2$ correlation results calculated from Grower B's weekly data were generally lower than those reported in literature for tomato crops (McAvoy, 1989) where an $r^2$ value of 0.896 (or r = 0.947) was obtained.

From the results of these preliminary plots, a second set of regression equations was obtained by plotting light and yield data in different units (***Table A-2 in Appendix D***). Light data was cumulated for 9 weeks (assuming a 60-day production period as before) in units of $W/m^2$ and regressed with yield measured in units of $kg/m^2$ to develop an empirical model which would produce more meaningful results (in more useful and common units) for greenhouse growers.

The whole range of regression equations derived from each year of data were used to predict 1994 tomato yields and compared with actual yields reported for the same period (***Figures A-18 & A-19 in Appendix D***).

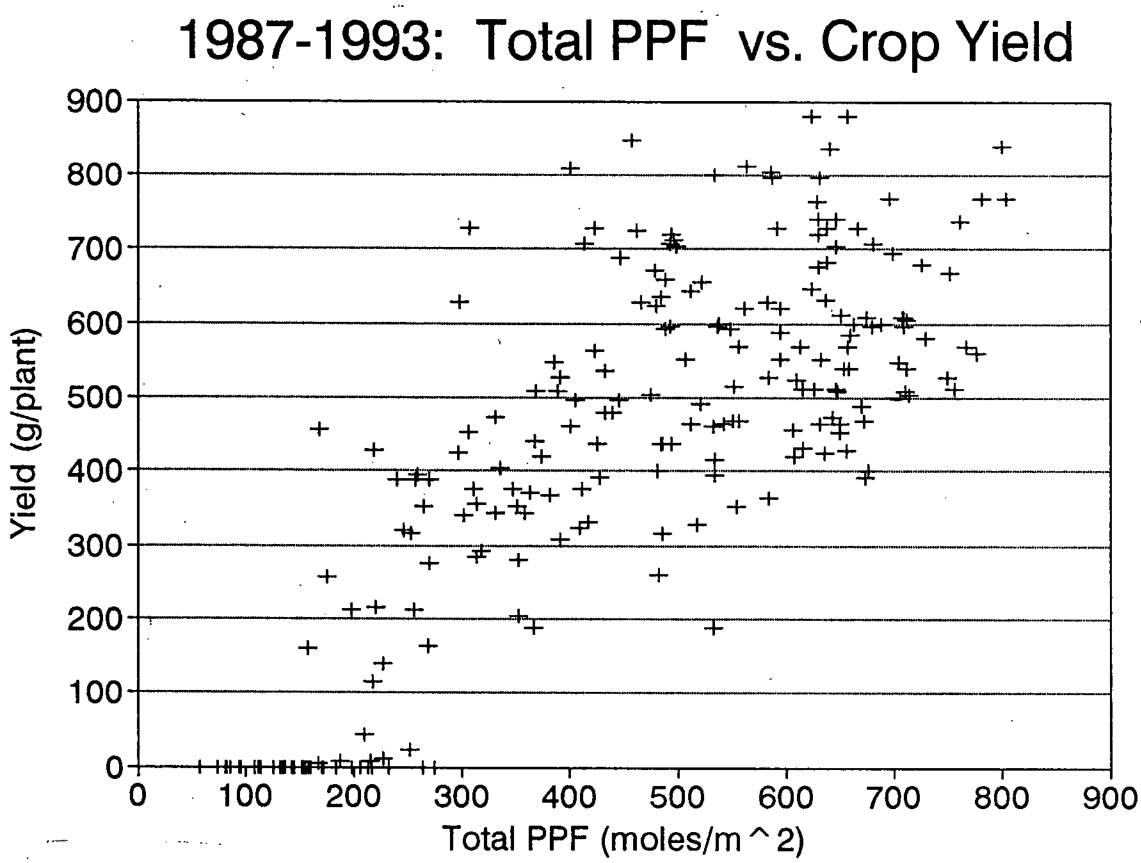



Figure 13 Trend identification of light (PPF) vs. Crop Yield

Analysis by the root-mean-square-error analysis revealed the best results (RMSE = 0.382 kg/m$^2$, $r^2$ = 0.625) when the equation derived from the 1990 set of data based on light alone ($CL_9$) was used to predict yields for 1994.

The coefficients for 1990 set of data were then identified as predictive equation 1.1 (Table 1, and Appendix C - Table A-1). Using a similar approach, the regression equation derived from the 1992 set of data was identified as predictive equation 1.2 to provide an upper-range value for prediction (***Table 1a***). When equation 1.2 (1992 equation) was used to predict 1994 tomato yields an RMSE value of 0.4574 kg/m$^2$ ($r^2$ = 0.465) was calculated. Plots of both sets of same-week yield predictions from equation 1.1 and 1.2 against actual recorded yields is shown on ***Figure 14***. From an examination of the plotted predicted yields, the regression equations were found to produce fairly good results for the first 20 weeks of the growing season when yields rose steadily.

However, for certain intervals of rapid growth (between weeks 20 to 30 and weeks 40 to 46), peaks in actual recorded yields could not be adequately predicted by the regression models. In fact from week 20 onward, the light-only based models ($CL_9$) were unable to predict the yield extremes found in actual greenhouse production practises. The greatest discrepancy in actual recorded yields and predicted yields was found during weeks 25 and 32 where the regression models consistently under-predicted actual yields. Although an empirical model developed solely on cumulative light had its basis, adequate correction for predictions during the peak periods of actual yield could not be obtained using a light-only model.

Table 1a

# EMPIRICAL MATH MODEL
## Results for 1994 Yield Predictions

### 1.0 Light Alone ($CL_9$)

$$Y_{p1} = a_0 + a_1(CL_9) \qquad (1)$$

| | $a_0$ | $a_1$ | $r^2$ | RMSE |
|---|---|---|---|---|
| Predictive equation (1.1) (1990 Data) | -0.07057 | 0.0000165 | 0.625 | 0.382832 |
| Predictive equation (1.2) (1992 Data) | -0.08383 | 0.0000145 | 0.465 | 0.457382 |

The predictive equation (1.1) is based on fitting 1990 data to regression equation (1) and produced a RMSE of 0.34827 and $r^2 = 0.625$ for 1994 yield predictions. Predictive equation (1.2) is based on fitting 1992 data to regression equation (1) and produced a RMSE of 0.457382 and $r^2 = 0.465$ when used to predict 1994 yield.

### 2.0 Light and Heat ($CL_9$ and $CH_9$)

$$Y_{p2} = a_0 + a_1(CL_9) + a_2(CH_9) \qquad (2)$$

| | $a_0$ | $a_1$ | $a_2$ | $r^2$ | RMSE |
|---|---|---|---|---|---|
| Predictive equation (2.1) (1992 Data) | 2.39584 | 0.00001513 | -0.03003 | 0.68 | 0.353099 |
| Predictive equation (2.2) (1990 Data) | 2.38054 | 0.00002444 | -0.033809 | 0.247 | 0.542205 |

The predictive equation (2.1) is based on fitting 1992 data to regression equation (2) and produced a RMSE of 0.353 and $r^2 = 0.68$ for 1994 yield predictions. Predictive equation (2.2) is based on 1990 data and produced a RMSE of 0.542 and $r^2$=0.247 when used to predict 1994 yield.

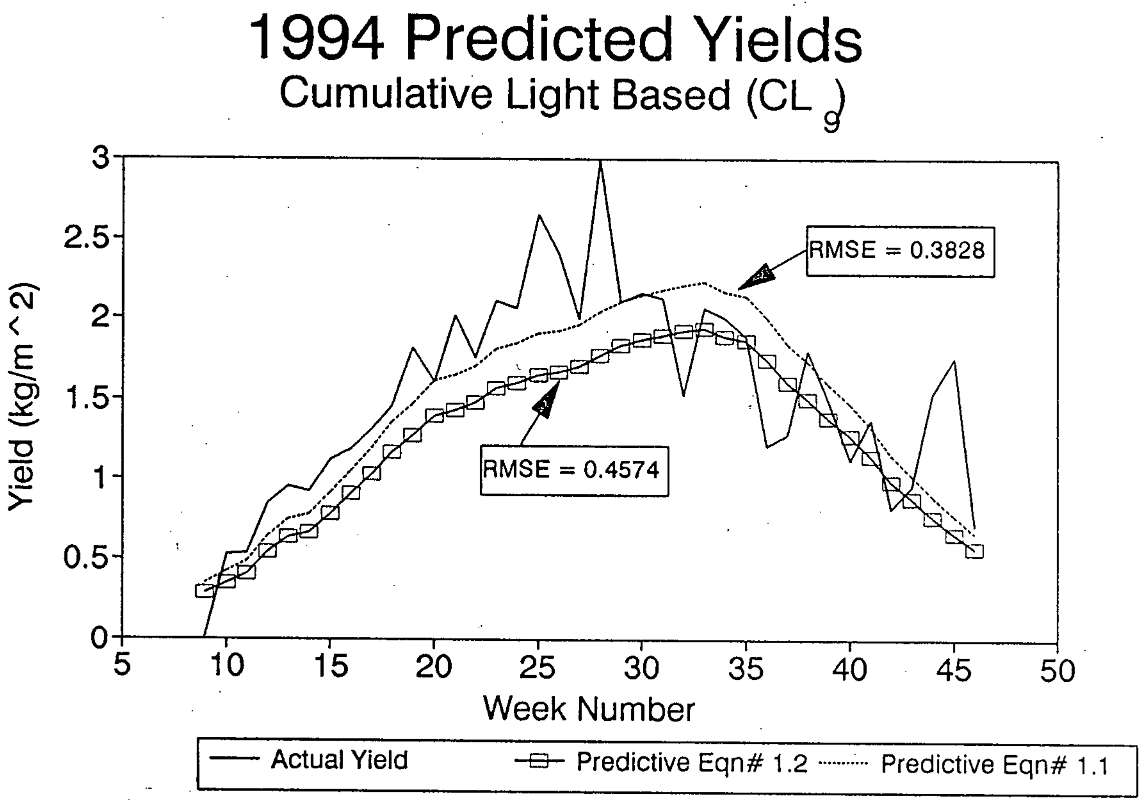



Figure 14 Predicted Yields for 1994 production based on cumulative light data

Greenhouse growers require the best yield forecasts during peak growth periods to ensure that availability of adequate harvesting supplies and workers to capitalize on the crop productivity.

In an attempt to provide closer predictions during these peak yield periods, another climatic variable had to be incorporated into the regression equation. From literature review, temperature was also found to strongly affect the yield of tomato crops (Jones et al, 1991). As well, in another study, the development time between flowering and harvest for tomato crops was found to be mainly dependent on temperature (de Koning, 1993b).

Consequently, the next climatic variable entered into the empirical model for predicting tomato yield was temperature.

Using a cumulative heat unit approach described in literature (Wolf et al., 1986) 9-week cumulations of corrected temperatures ($CH_9$) were regressed with 9-week cumulations of light data ($CL_9$) to create a light and heat based empirical model.

Multiple step-wise regressions were performed for each year of data to provide detailed analysis of the relationships between corrected temperatures ($CH_9$), light ($CL_9$), and yield (***Appendix D***). The resulting regression equations from this approach were then summarized (***Appendix D, Table A-2***) and used to predict 1994 tomato yields. Prediction plots versus actual recorded yields for 1994 were performed (***Appendix D, Figures A-20, A-21***). Analysing the plotted results it was found that the incorporation of the temperature variable provided better predictions during the peak production periods. Where the light-only plots ($CL_9$) were unable to predict for these extreme intervals, the 9 week cumulative temperature ($CH_9$) and light ($CL_9$) based equations were able to correct for the peak periods. Isolating the equations from the same two years (1990 and 1992) of predictions (Table 1), a good range of predictions was obtained. The best predictions were found using the heat and light method with the 1992 set of data (RMSE = 0.353 kg/m$^2$) and was identified as predictive equation 2.1 (***Table 1a***).

The 1990 equation was identified as predictive equation 2.2 (RMSE = 0.542 kg/m$^2$) both for comparison with the previous light-only model ($CL_9$) and also to illustrate the improved predictions during the peak yield periods using this method. From the plot of predicted yields and actual yield (***Figure 15***) a better range of predictions for high and low values were found. Predictive equation 2.1 (1992 data) was found to produce yield predictions very close to the actual yields for the first 25 weeks of the production period. Problems with predicting yields during peak periods still existed using equation 2.1 but predictions were generally higher during these intervals producing an overall improved and lower RMSE of 0.353 kg/m$^2$.

Using predictive equation 2.2, yields were generally over-predicted for the entire production periods (RMSE = 0.542 kg/m$^2$) but the possibility for adequate correction using this method ($CH_9$ & $CL_9$ regressions) was shown. If RMSE analysis was performed for the peak periods alone, equation 2.2 with its over-predictions may actually be proven to provide the most useful results to a greenhouse grower. During peak periods it may be preferable for a grower to be prepared to handle higher yields than those actually experienced to maximize productivity.

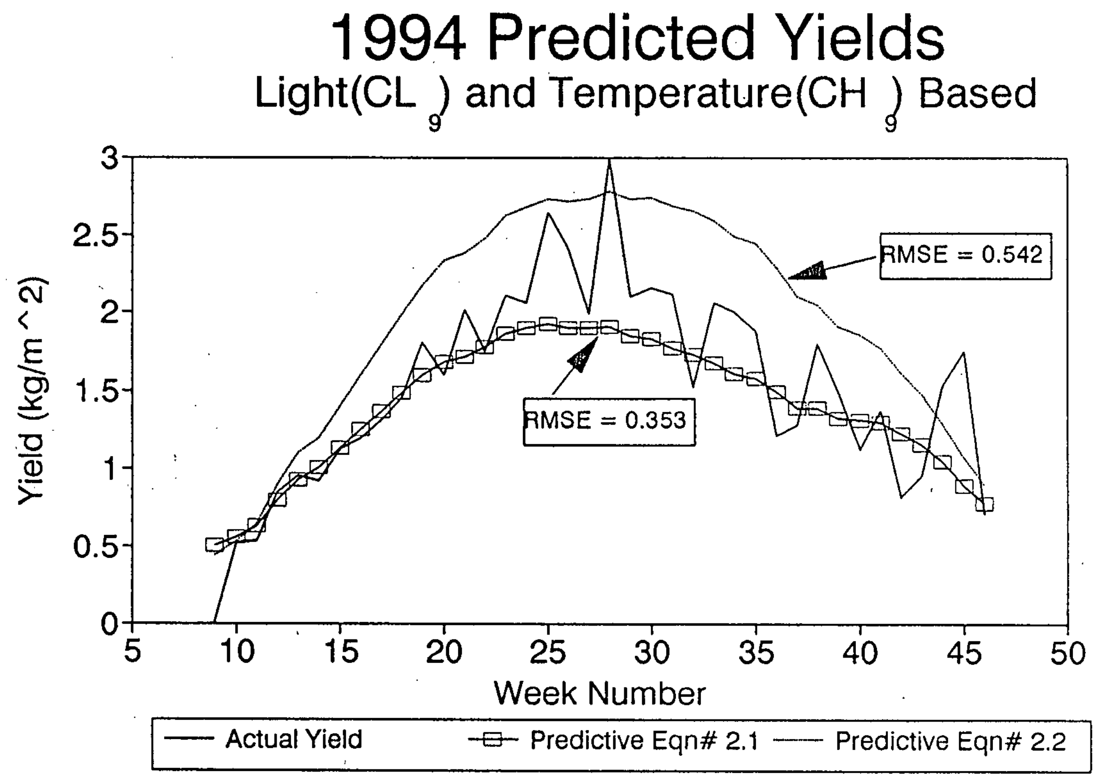



Figure 15 Predicted yields for 1994 production based on cumulative light and heat data.

The encouraging results from these light and heat multiple regressions led to experimentation with other combinations of cumulative light and heat data. In particular, it was postulated that by cumulating heat and light units for different time intervals (using a similar approach as used for the 9-week period) 3 and 6 week cumulations would be made.

For this experimental process, only one yearly characteristic set of data was used for developing the empirical model.

As detailed in the materials and methods section the year of data with the best linear relationship with light was further analysed. The 1992 data set was chosen for further analysis because it had the closest relationship between total PPF and crop yield (g/plant) from the first set of regression results ($r = 0.746$).

As well, the 1992 equation was selected for a better comparison of results between the deterministic and empirical models. A comparison of prediction results from the deterministic and empirical models developed from the same year of data (the former being hourly and the latter being weekly 1992 data) would ideally identify the better model rather than variations in predictions based on climatic differences.

By incorporating the 3, 6 and 9-week cumulative totals for light and heat into an all-inclusive equation, Equation 3.1 (***Table 1b***) (where $Yp = f\{CH_3, CH_6, CH_9, CL_3, CL_6, CL_9\}$) it was hoped that this method would correct for yield peaks and provide an overall improvement in prediction accuracy. Using three light and heat variables in the yield prediction was based on the fact that each interval (3, 6, and 9 weeks) may have varying effects on the development of the tomato crop related to yield.

Table 1b

# EMPIRICAL MATH MODEL

## 3.0 Light and Heat ($CL_{3,6,9}$ and $CH_{3,6,9}$) - Immediate Yield Prediction

$$Y_{p3} = a_0 + a_1(CL_3) + a_2(CL_6) + a_3(CL_9) + a_4(CH_3) + a_5(CH_6) + a_6(CH_9) \qquad (3)$$

| | $a_0$ | $a_1$ | $a_2$ | $a_3$ | |
|---|---|---|---|---|---|
| Equation (3.1) (1992 Data) | 0.34917 | 0.0000355 | 0.00000548 | -0.000667 | |
| | $a_4$ | $a_5$ | $a_6$ | $r^2$ | RMSE |
| | -0.001127 | -0.03108 | -0.02212 | 0.527 | 0.429766 |

The predictive equation (3.1) is based on fitting 1992 data to regression equation (3) with a RMSE of 0.429766 and $r^2$ = 0.527 for 1994 predictions.

## 4.0 Light and Heat ($CL_{3,6,9}$ and $CH_{3,6,9}$) - Future Yield (4 Weeks ahead)

$$Y_{p4} = a_0 + a_1(CL_3) + a_2(CL_6) + a_3(CL_9) + a_4(CH_3) + a_5(CH_6) + a_6(CH_9) \qquad (4)$$

| | $a_0$ | $a_1$ | $a_2$ | $a_3$ | |
|---|---|---|---|---|---|
| Equation (4.1) (1992 Data) | 0.34917 | 0.0000355 | 0.00000548 | -0.000667 | |
| | $a_4$ | $a_5$ | $a_6$ | $r^2$ | RMSE |
| | -0.001127 | -0.03108 | -0.02212 | 0.491 | 0.44579 |

The predictive equation (4.1) is based on fitting 1992 data to regression equation (4) and produced a RMSE of 0.44579 and $r^2$ = 0.491 for 1994 predictions when using 3,6, and 9 week cumulated data (4 weeks ahead of actual yield) to make future yield predictions.

$$Y_{p4} = 0.35 + 0.0000355(CL_3) + 0.0000055(CL_6) - 0.00067(CL_9) - 0.0011(CH_3) - 0.031(CH_6) - 0.022(CH_9)$$

RMSE = 0.4458 kg / m$^2$

By incorporating each cumulative interval into the yield predictive equation an attempt was made to have a model that would predict for peak periods and more closely follow the trends of the actual yields. Regressing (***Appendix D***) and plotting the results of this approach (***Figure 16***) a better pattern of predictions was found even though prediction error increased (RMSE = 0.43 kg/m$^2$). Predicted yields showed more high and low variation and demonstrated a greater degree of association with the pattern of tomato crop productivity. From these regression results it became more evident that light or solar intensity was the primary influencing factor affecting yield with temperature effects being the secondary factor mainly influencing peak periods of productivity. From these stages in developing the empirical model, an attempt was made to predict yield weeks ahead of time to be a useful management tool for growers.

From the results of the three approaches used to develop the empirical model (light-only, light and heat, and 3 variables of light and heat) it was decided that the method using 3 variables of light and heat be pursued for creating a 4-week predict ahead model. Two approaches were used for developing this 4-week predict ahead model. The first approach involved cumulating 3, 6, and 9-week data for light and heat and regressing these results with actual yield values recorded 4 weeks ahead of time. Previously, predictions were only made for the same week by relating 3, 6, and 9-week data leading up to that week. With the 4 week ahead approach, the actual yields from 1992 were related to data cumulated 4 weeks prior to the actual reported yield.

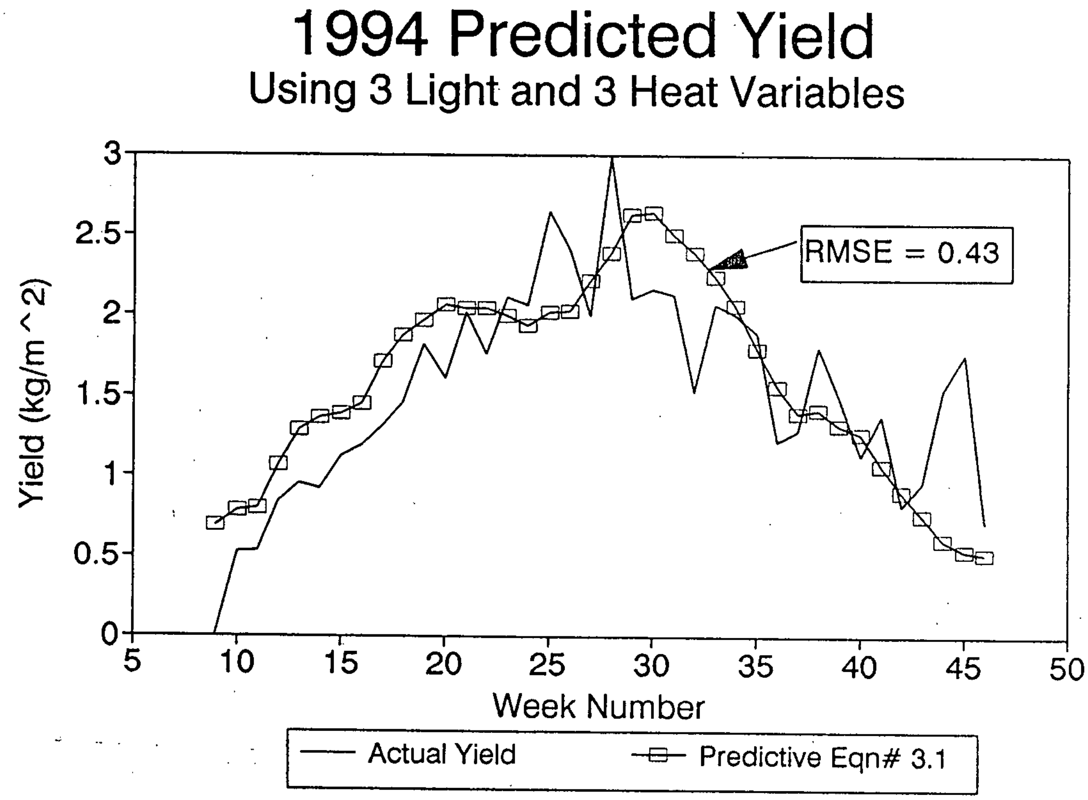



Figure 16 Predicted Immediate yield using 3 variables of light and 3 variables of heat. (CL3, CL6, CL9, and CH3, CH6, CH9)

The regression equation derived using this approach (*Appendix D, Alternative Regression Method*) produced predictions with fairly high error and low correlation with environmental variables (RMSE value of 0.615 $kg/m^2$ and $r^2 = 0.321$)

The second approach that was employed to predict 4 weeks ahead, was to simply use the same prediction equation developed for immediate, same-week predictions for the 3 variables of light and heat (***Equation 4.1, Table 1b***) and use weekly light and temperature values 4 weeks ahead in the equation. Use of this method produced a much lower RMSE value (0.45 $kg/m^2$) and showed better correlation ($r^2 = 0.491$) with environmental variables than predictions made using the regression equation derived from 4 week ahead data and yield (***Appendix D, Alternative Regression Method***). The predictions using the same week equation (4.1) with 4 week ahead data are shown on ***Figure 17.*** Analysing and comparing with the actual yield plot, prediction error for the first 25 weeks is minimal.

For the remaining 25 weeks of the growing season, the highest prediction error interval still remains between week 25 to 30 (June, July) when crop productivity is typically highest. A reason for the greater error during this interval include the strong influence of warmer temperatures on crop development time during this period which cannot be easily accommodated by a regression model that tries to predict weekly yields for an entire growing season. In fact, researchers are still having difficulty estimating the amount of plant assimilates (nutrients from photosynthesis and root absorption) being formed and translated at different temperatures.

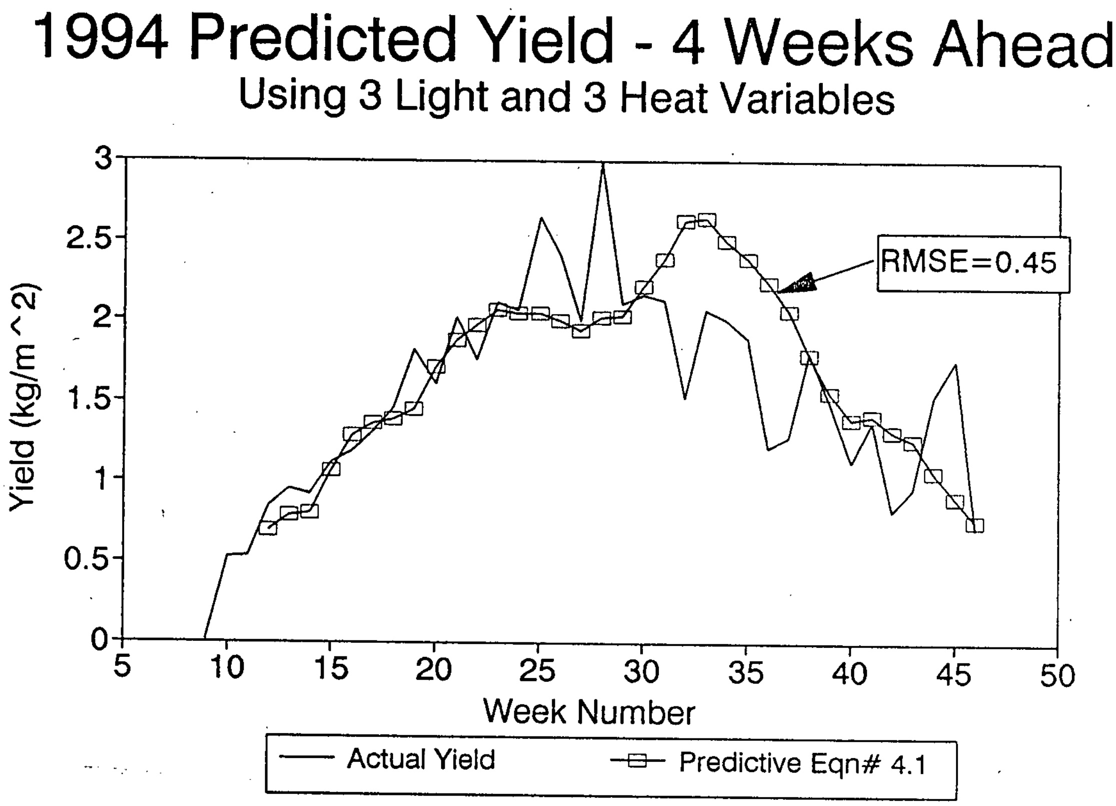



Figure 17 Predicted yield for 4 weeks ahead using 3 light and 3 heat variables. (CL3, CL6, CL9, and CH3, CH6, CH9)

Nutrient translocation rates have been reported to be about four times higher at 27°C than at 15°C (Barendse, 1993) and this may provide a clue to sharp increases in productivity during high temperature intervals.

Similarly, tomato development times between flowering to harvesting of the tomato fruit have been found to vary from 70 days at 17°C to 44 days at 25°C, demonstrating strong a strong relationship to temperature (de Koning, 1993b). In order to improve the accuracy of predictions during the peak temperature and solar intensity periods (weeks 25 to 30), it may be necessary to develop regression equations for those specific intervals. Actual yield and environmental data for those peak periods may need to be isolated and regressed separately to derive separate equations for peak productivity intervals throughout the year. The grower may then use these 'peak period equations' to provide an upper-end estimate of yields during high productivity periods.

When a trial test run was performed to incorporate the effects of humidity as well as light and heat (***Appendix E**, Humidity Trial Regression Results & Figure A-22*), prediction accuracy was found to be worse than with only light and heat, (RMSE = 0.472 kg/m$^2$).

Overall, the empirical model with 4-week ahead predictions using equation (4.1) was found to be the most flexible and useful for actual application by greenhouse growers.

### 3.4 Results of the Agassiz Expert System

Using weekly data from Greenhouse Grower B, the Agassiz Expert System model was trained to provide tomato yield predictions 1, 2, 3, and 4 weeks ahead of time. The specific weighting and configuration of environmental variables was decided upon by the plant physiologist researchers at Agassiz (Lin, 1994) and the detailed methodology used in the development of the model remained confidential for licensing purposes. However, prediction results using this model were shared for academic comparison with the empirical and deterministic models developed in this study. The tabulated results from the Agassiz Expert System are shown on ***Table 2.***

For comparison with the two mathematical models in the study, RMSE analysis was performed on the four sets of predictions. Weekly predictions were identified as PRE-1, PRE-2, PRE-3, and PRE-4 results, representing predictions made 1, 2, 3, and 4 weeks before actual yields are recorded, respectively. Error analysis by RMSE revealed a range from a high 0.722 $kg/m^2$ for the PRE-2 results (2 weeks before actual yield is recorded) to a low value of 0.401 $kg/m^2$ for the PRE-1 results.

The best predictions in terms of lowest error when compared with actual yield were found for the PRE-1 (1 week before actual yield is recorded) set of results with an RMSE of 0.401 $kg/m^2$. The predictions from the Agassiz model were then compared graphically by plotting the PRE-1 results with actual recorded yields for better analysis of prediction performance throughout the growth season (***Figure 18***).

Comparing the actual and predicted yields (PRE-1) the advantages of the trend recognition techniques employed by the Agassiz expert system become more evident. The expert system is capable of predicting more variability in yield trends throughout the growing season than the mathematical models. During the first 20 weeks, although predictions have less accuracy than those of the mathematical models, the general trend of actual yields is well predicted. Between weeks 20 to 30 there is evidence of the improved ability to predict for peaks and valleys in crop productivity during intervals of high temperature and solar intensity.

## 1994 TOMATO YIELD PREDICTION
### EXPERT SYSTEM RESULTS

| WEEK | PRE-4 | PRE-3<br>KG/M ^ 2 | PRE-2 | PRE-1 | ACTUAL YIELD<br>KG/M ^ 2 | 1994 DATE<br>(Mon --> Sun) |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 6 | | PREDICTED | | | ACTUAL | |
| 7 | | YIELDS | | | YIELD | |
| 8 | | | | | | |
| 9 | NA | NA | NA | NA | 0.01 | FEB28-MAR6 |
| 10 | NA | NA | NA | 0.47 | 0.53 | MAR07-MAR13 |
| 11 | NA | NA | 0.66 | 0.66 | 0.54 | MAR14-MAR20 |
| 12 | NA | 0.77 | 0.77 | 0.47 | 0.85 | MAR21-MAR27 |
| 13 | 0.85 | 0.82 | 0.24 | 0.66 | 0.95 | MAR28-APR03 |
| 14 | 1.16 | 0.54 | 0.39 | 0.95 | 0.92 | APR04-APR10 |
| 15 | 1.12 | 0.77 | 1.11 | 0.95 | 1.12 | APR11-APR17 |
| 16 | 0.36 | 1.12 | 1.14 | 1.26 | 1.19 | APR18-APR24 |
| 17 | 1.28 | 1.58 | 1.23 | 0.82 | 1.31 | APR25-MAY01 |
| 18 | 1.45 | 1.25 | 0.85 | 1.23 | 1.45 | MAY02-MAY08 |
| 19 | 1.63 | 0.96 | 0.87 | 1.56 | 1.81 | MAY09-MAY15 |
| 20 | 1.26 | 0.87 | 1.45 | 1.38 | 1.6 | MAY16-MAY22 |
| 21 | 1.47 | 1.77 | 1.19 | 1.87 | 2.92 | MAY23-MAY29 |
| 22 | 2.1 | 1.33 | 1.84 | 1.92 | 1.75 | MAY30-JUN05 |
| 23 | 1.93 | 1.74 | 1.62 | 1.75 | 2.11 | JUN06-JUN12 |
| 24 | 2.08 | 1.62 | 1.46 | 1.71 | 2.06 | JUN13-JUN19 |
| 25 | 2.31 | 1.42 | 1.39 | 1.99 | 2.65 | JUN20-JUN26 |
| 26 | 2.05 | 1.51 | 1.59 | 2.18 | 2.41 | JUN27-JUL03 |
| 27 | 2.09 | 1.53 | 1.83 | 1.83 | 1.99 | JUL04-JUL10 |
| 28 | 1.88 | 1.97 | 1.54 | 1.74 | 2.99 | JUL11-JUL17 |
| 29 | 2.52 | 1.83 | 3.34 | 2.41 | 2.1 | JUL18-JUL24 |
| 30 | 2.35 | 2.94 | 4.34 | 1.93 | 2.16 | JUL25-JUL31 |
| 31 | 2.63 | 3.63 | 2.12 | 1.67 | 2.12 | AUG01-AUG07 |
| 32 | 3.94 | 1.78 | 1.48 | 1.72 | 1.52 | AUG08-AUG14 |
| 33 | 1.36 | 1.48 | 1.65 | 1.88 | 2.06 | AUG15-AUG21 |
| 34 | 1.48 | 1.59 | 1.92 | 2.14 | 2 | AUG22-AUG28 |
| 35 | 1.49 | 1.58 | 1.96 | 1.94 | 1.88 | AUG29-SEP04 |
| 36 | 1.05 | 1.41 | 1.83 | 1.7 | 1.21 | SEP05-SEP11 |
| 37 | 1.31 | 1.32 | 1.56 | 1.3 | 1.28 | SEP12-SEP18 |
| 38 | 1.32 | 1.24 | 1.37 | 1.31 | 1.8 | SEP19-SEP25 |
| 39 | 1.29 | 1.22 | 1.31 | 1.81 | 1.49 | SEP26-OCT02 |
| 40 | 1.04 | 1.1 | 1.49 | 0.99 | 1.12 | OCT03-OCT09 |
| 41 | 0.93 | 1.1 | 0.78 | 0.78 | 1.37 | OCT10-OCT16 |
| 42 | 1.06 | 1.12 | 0.74 | 0.7 | 0.82 | OCT17-OCT23 |
| 43 | 1.62 | 1.12 | 0.54 | 0.96 | 0.95 | OCT24-OCT30 |
| 44 | 1.27 | 1.23 | 1.06 | 1.48 | 1.53 | OCT31-NOV06 |
| 45 | 1.21 | 0.99 | 1.44 | 1.61 | 1.75 | NOV07-NOV13 |
| 46 | 0.96 | 0.77 | 1.26 | 2 | 1.75 | NOV14-NOV20 |
| 47 | 0.7 | 1.08 | 1.63 | 1.4 | 0.71 | NOV21-NOV27 |
| RMSE | 0.63 | 0.603 | 0.722 | 0.401 | | |

NA = NOT AVAILABLE BECAUSE OF NO PREVIOUS ACTUAL YIELD RECORD

Table 2 Tabulated results for the Agassiz Expert System

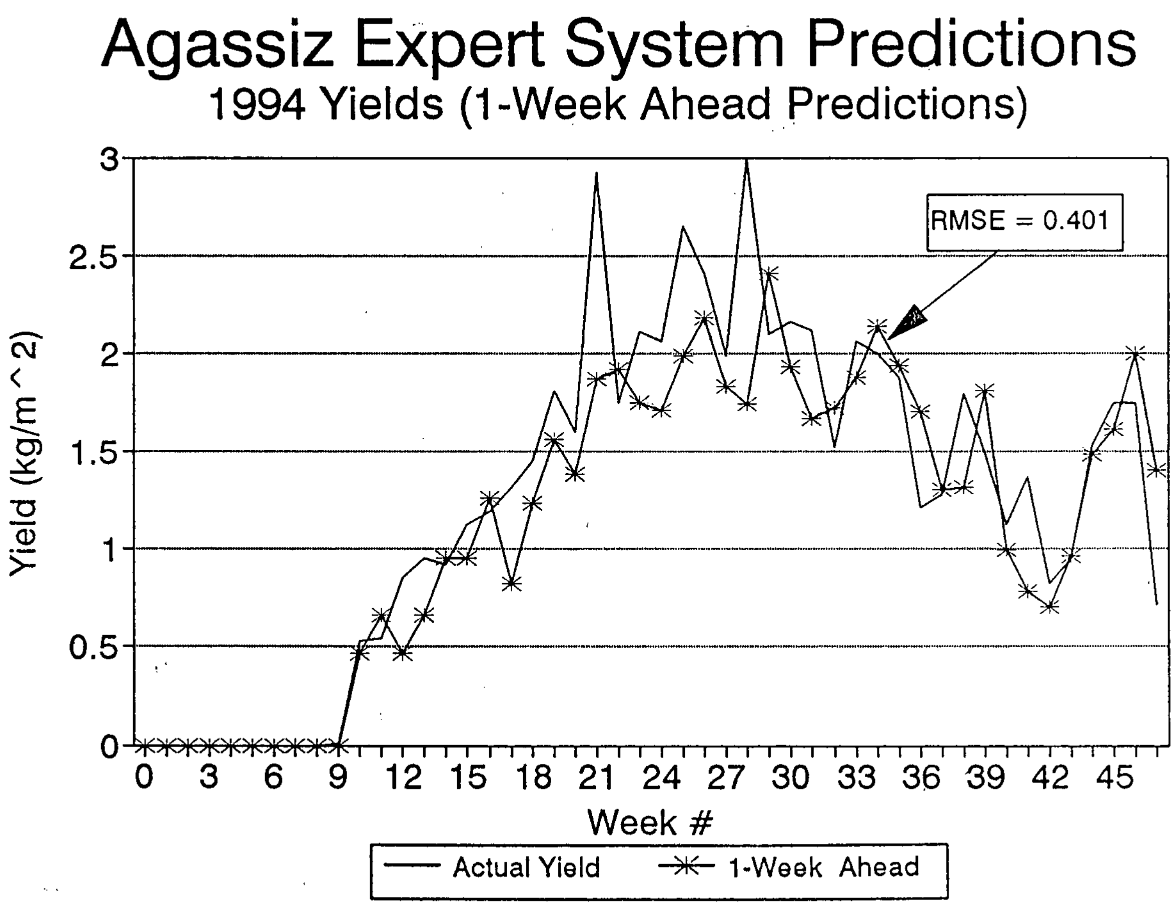



Figure 18 Shown above are the 1-Week Ahead Yield Predictions for 1994 using the Agassiz Expert System.

For the last interval, from weeks 30 to 50, the expert system continues to demonstrate its general ability to provide close predictions for peak productivity periods and an excellent forecast of the yield trend expected.

For the purpose as a greenhouse management tool, yield predictions 1-week ahead (PRE-1) may be sufficient for preparing for weekly harvests depending on the size of the greenhouse operation. Larger operations may require an earlier forecast of yields for decision-making regarding capital cost investments such as the purchase of higher capacity harvesting equipment. However, the advantage of the predictions provided by the Agassiz Expert System is its ease of application, and flexibility.

With the expert system designed to provide a range of predictions (1, 2, 3, and 4 week ahead), the same model could be applied to another set of data from another greenhouse site quite easily as only the weighting factors and input data may change. In fact, in terms of flexibility and user-friendliness, the expert system has the greatest potential and advantage. As the database of actual yield and environmental parameters grows, the predictions by the trend recognition expert system will improve. The expert system will be able to account for fluctuations in yield and better accommodate for yearly trends than the mathematical models, without tedious modifications. As an added advantage, the expert system approach begins as a general model for yield prediction, but then as it is trained with grower and site specific data it learns yield trends. The expert system model becomes particular to trends of a specific grower and gradually becomes a customized model.

The results from the Agassiz expert system demonstrate the dynamic and flexible nature of this approach and encourages further research and experimentation with this method of yield prediction.

### 3.5 Validation and Comparison of Results from the Yield Predictive Models

The guideline for validation of the models in this study was established to be an RMSE error value of less than 0.8 $kg/m^2$ as mentioned earlier (***Section** 2.7*).

For the deterministic model based on the first-principle equations of net photosynthesis and respiration, the best model developed was a 2-week cumulative $P_{net}$ based yield model predicting 6 weeks ahead with an RMSE value of 0.3618 $kg/m^2$. The error value obtained within the set guideline required for validation so the model was considered valid. The advantages of the deterministic modeling method was that it was flexible, and based on well-accepted first-principle equations. The model was considered flexible because the equations applied relied purely on environmental parameters and did not attempt to measure trends or simulate patterns of yield productivity. Consequently, by theory, the model should be applicable to any greenhouse site without bias, and provide the same predictions for any tomato crop under the same environmental conditions. The deterministic model is not season-dependent nor grower-specific because the structure of this approach relies only on the magnitude of environmental parameters such as solar intensity, temperature, and carbon dioxide concentrations.

Disadvantages to the deterministic approach include the strict environmental monitoring required, the use of many growth parameters, assumptions in the yield conversion factor, and the tedious nature of application. In order to provide accurate and useful predictions, the development of the model required a full year of hourly environmental data (8760 hours / year). Missing or ambiguous data for any extended periods of time led would result in 'zero values' for predictions during that interval because the equation would not have any input values. The problem of missing data becomes greater once the effect of cumulating weekly net photosynthetic rates ($P_{net}$) is considered. If cumulative totals of $P_{net}$ are being converted and used to predict yield, low totals due to missing data for normally productive yield periods would lead to inaccurate predictions. In this study, zero yield periods were solved by extrapolation between known prediction points but without the far prediction point (weeks ahead of time) no yield value could be reported. Aside from the problem of missing data, it may not be time or cost-effective to require such a strict environmental monitoring scheme to ensure proper predictions by this method.

The use of a great number of coefficients and growth parameters is also a weakness of the deterministic approach (***Figures 6a & 6b***). Small errors in assumptions or parameter values become magnified with subsequent predictions and finding the source of the error from the many assumptions becomes a difficult task.

Another uncertainty in developing the deterministic model was the application of the yield conversion factor found in literature (Enoch, 1978). Examining the assumptions involved in applying the conversion factor more carefully (***Figure 6c***) two of the main parameters on which the factor is based are quite variable. First, the assumption that half the dry matter of a tomato is partitioned to yield will strongly affect the magnitude of the conversion factor. For example, if the dry matter partitioning for the tomato is actually closer to 60% in practise, rather than the assumed 50%, a 17% increase in the yield conversion factor results (factor changes value from 7 to 8.18). Similarly, from literature the dry matter content for tomatoes was found to vary between 5 to 10% at different stages of fruit development (Ho and Hewitt, 1986). If the assumed value for dry matter content changed from 5% to 10%, the yield conversion factor would change from a value of 7 to 3.4. Further, if both assumptions are slightly inaccurate then the changed yield conversion factor would produce significantly different results.

Changes in temperature are known to strongly affect the partitioning of assimilates in plants (Barendse, 1993) as well as development time of tomatoes (de Koning, 1993b). Assuming a constant or averaged dry matter content for the entire growth season in order to apply the yield conversion factor may only lead to predictions that apply to certain periods of fruit development. The yield conversion factor assumes that by averaging the high and low percentage periods of dry matter partitioning throughout a growing season a generalized yield prediction can be made. As a result of the many assumptions, sources of error, and possibilities for variations in the yield conversion factor, the low error value (RMSE = 0.3618 $kg/m^2$) of the deterministic model may not prove overall prediction accuracy.

The empirical model based on multiple regressions of historical data produced reliable and useful predictions for the 1994 growing season. Beginning with a light-only based model and expanding to a model which incorporated temperature in terms of heat units, the empirical model gradually reduced its prediction error. The best, most useful empirical model developed, involved using cumulations of 3, 6, and 9-weeks of light and temperature data to predict yields 4-weeks ahead (***Equation 4.1, Figure 17***). Error analysis of this approach produced an RMSE value of 0.45 kg/m$^2$, well within the required validation guideline of 0.8 kg/m$^2$ established for this study.

The advantages of the empirical approach include the straight-forward method of development and reliable prediction results for general yield trends. Using regression equations derived from historical data to relate environmental variables to actual yield had the advantage of not requiring plant physiology assumptions or conversion factors.

Based on previous trends of production for a specific grower and site, the regression equations were developed and analysed for their prediction accuracy. The prediction equation required only two parameters, temperature and light on a weekly basis to produce results. The empirical approach was much simpler to apply since less stringent environmental monitoring would be required for obtaining weekly averaged values for only light and temperature data. From the graphical plots, the empirical approach was able to adequately predict tomato yield trends for the first 25 weeks with fairly good accuracy. Difficulty in providing close predictions was encountered during peak yield periods (June-July) most likely due to changes in plant partitioning and shortened development times induced by warmer temperatures (Barendse, 1993) mentioned earlier.

Disadvantages to the empirical method would be that the fact that equations derived for the data sets would only be site or grower-specific, and the inability to predict for periods. Since the empirical method employs regression techniques for a specific set of data, the equations derived for one grower could not be applied to other sites or greenhouses. Equations or methods that produced the best-fitting predictions for a particular grower may not necessarily produce good forecasts for other growers. For each grower, a new set of regression equations would have to be produced, and new analysis of results would have be performed to identify the best equations. As a result, the methodology of developing the model actual makes the empirical model the least flexible of the three models compared in this study. However, comparatively, the predictions from the empirical model are more reliable than those from the deterministic model because the method used its development requires fewer assumptions.

The final model examined in this study was the Agassiz Expert System model which employed trend recognition techniques to make yield predictions. The one-week ahead prediction model (PRE-1) produced an RMSE value of 0.401 kg/m$^2$ and was considered a valid model by the validation guideline of 0.8 kg/m$^2$. The plotted results of the one-week ahead model versus actual yield (***Figure 18***) demonstrated the ability of the expert approach to provide the good variability to predict high and low peaks and during high productivity periods (weeks 20 to 35).

As a comparison model, the Agassiz Expert System model showed the most flexibility. The expert system could be applied to new sites or growers without making major changes to the model itself, other than adjusting weighting factors applied and training the model with new sets of data. Further, the expert system would be the most user-friendly model to apply for yield prediction as it quickly provides an understandable range of predictions (PRE-1, PRE-2, PRE-3, PRE-4) without requiring the user to have detailed knowledge of plant physiology.

In order to summarize the weaknesses and strengths of each approach, a three-point ranking system was devised to determine the best overall model. The best model in each category is ranked with a value of 1 (for first / best) and the worst in each category is ranked with a value of 3 (for last / worst). The model with the ***overall lowest total score*** out of a maximum of 15 points for 5 categories will be identified as the ***best overall prediction model.***

The categories used for comparison included: RMSE error for the entire annual growing period, RMSE error for the most productive period (Weeks 20 to 35 of the annual growing season), flexibility of application for individual grower sites, application flexibility for the general case, and user-friendliness of the model (***Table 3***).

When the root-mean-square-error was calculated and compared for each model the deterministic model reported the lowest error (0.3818 $kg/m^2$) and was ranked with a value of 1. The Agassiz expert system was ranked a close second for overall RMSE (0.401 $kg/m^2$), while the empirical model reported the greatest error of the comparison group (0.45 $kg/m^2$) and was ranked with a value of 3. Although the deterministic model reported the lowest overall RMSE for the year of predictions it must also be noted that during part of the growing season, incomplete data sets led to zero predictions for certain intervals. As a result, the deterministic model required extrapolations between known prediction points to provide a full year of predictions. Further, the known variability in fruit water content and dry matter partitioning throughout fruit development may require the application of a variable instead of a constant yield conversion factor. The advantage of the expert system approach was that even with missing or ambiguous data sets it had the ability to make predictions based on analysed trends of historical data.

Finally, the empirical model (***Figure 17***) produced the best predictions for the first 25 weeks of the growing season but had difficulty predicting the high productivity periods. Based on numerical analysis alone, the deterministic model had the best overall RMSE results but in terms of consistency, the empirical and Agassiz expert system may be more reliable (***Table 3***).

For a better comparison of the results during the peak productivity period, a separate root-mean-square-error analysis was performed between weeks 20 to 35 of the growing season. The results are shown in ***Table 3*** with the best results being reported for the deterministic model, followed by the empirical, and Agassiz expert system (ranked 1, 2, and 3, respectively). Although numerically the error results were the best for the deterministic method, the expert system actually had the best prediction of variability during this period, in terms of high and low peaks. The deterministic produced good results (compared with actual recorded yields) during this period as well but required a complete set of hourly data to produce these predictions. The feasibility of such strict monitoring of environmental parameters and extrapolation of results in cases of missing data is questioned if this method were to be actually applied in-situ. The empirical model had fairly good results as well, but had less ability to predict for the general trends of yield productivity when results were analysed graphically (***Figure 17***).

# Table 3: Comparison of Yield Predictive Models

| PREDICTIVE MODEL | OVERALL RMSE | RANK | Wk 20 to 35 RMSE | RANK | Application Flexibility | User Friendly | OVERALL SCORE |
|---|---|---|---|---|---|---|---|
| | (kg/m$^2$) | (1 - 3) | (kg/m^2) | (1 - 3) | Rank: (1 - 3) Site / General | (1- 3) | Score/Rank |
| DETERMINISTIC | 0.3818 | 1 | 0.170 | 1 | 3 / 2 | 3 | 10 / **2** |
| EMPIRICAL | 0.45 | 3 | 0.233 | 2 | 2 / 3 | 2 | 12 / **3** |
| AGASSIZ EXPERT SYSTEM | 0.401 | 2 | 0.247 | 3 | 1 / 1 | 1 | 9 / **1** |

*** Deterministic Model: 2 Week Cumulative $P_{net}$ based model - 6 weeks ahead**
*** Empirical Model : 3 Light & 3 Heat Variables - 4 weeks ahead**
*** Agassiz Expert Sys. : Trend Recognition Based - 1 week ahead**

In terms of flexibility of application on a site-specific basis, the Agassiz expert system was ranked highest because of its well-defined model structure (***Table 3***). The Agassiz model could be easily applied to new sites without any major modification to the model itself. New historical data for each site would be required to train the model, but the model itself could be easily applied to new sites with minor difficulties. The empirical approach was ranked second because it is based purely on site-specific data and the model is customized for particular growers. The deterministic model using first principle equations of net photosynthesis was ranked third because it applied generalized equations regardless of a specific site. Based on environmental variables alone, the deterministic model would make predictions without considering any yield trends from previous years.

When the flexibility of applying the various models to general cases was analysed, the Agassiz expert system was ranked first because it would be able to accommodate the greatest variations in available data. While the deterministic and empirical models require complete data sets and a more rigid environmental monitoring scheme, the expert model would be reliable even in cases of missing data sets. For actual field application, it is believed that providing reasonable predictions regardless of missing data sets would be extremely beneficial to the greenhouse grower and management staff.

Comparing user-friendliness of the three models, a number of factors were considered.

First, the prediction results would be considered user-friendly if they were based on very few assumptions, and could be easily interpreted without extensive knowledge or thought of plant physiology. It was also decided that yield predictions one-week ahead of time were sufficient for most greenhouse management decisions. The Agassiz expert system was the only model that had the advantage of providing an immediate 1, 2, 3, and 4 week ahead range of yield predictions and produced some of the best results from its 1-week ahead forecasts. Generally, larger management decisions for capital cost expenditures would be dependent on analysing annual productivity trends rather than weekly yield patterns. Consequently, the best use for the yield prediction models would be to prepare adequate equipment and manpower for harvests and to maximize overall crop productivity.

As the goal of the prediction models was to purely to provide good tomato crop yield predictions, the model requiring the least manipulation or customization by the user was considered the best model. The empirical model based on multiple trials of linear regression would require an experienced technician or researcher to properly fit the data for each individual grower site. To properly apply the deterministic model approach a strict environmental monitoring scheme would be required, which may not be a cost-effective option for many small-scale growers. However, the deterministic model would have the advantage of being applied effectively in general cases if predictions were properly adjusted with variable yield conversion factors throughout different stages of fruit development. As well, in situations where predictions are required fairly far in advance, the deterministic model may be favored for its 6-week-ahead predictions.

Similarly, the empirical model has the advantage of predictions made 4 weeks ahead but requires many regression trials to properly fit each growers' historical yield data. Comparatively, the Agassiz expert system would require the least customization of all three models to provide reliable and cost-effective yield predictions.

Based on the overall scores reported for each model, the Agassiz Expert system was ranked first, with the deterministic model a close second. The empirical model was ranked lower than the other two models due to its lack of flexibility in actual application and the need for customization by skilled technicians or researchers. For cases when predictions are required far in advance (i.e. 6 weeks) then the deterministic model may be applied effectively. However, for the general case, it is believed that the range of 1, 2, 3, and 4 week predictions offered by the Agassiz expert system would be sufficient for most end-users.

Overall, the best prediction approach was determined to be the Agassiz expert system model, favoured for its simplicity, reliable and accurate predictions, flexibility in application, and user-friendliness.

## 4.0 Conclusions

In this study, two mathematical models were successfully adapted (deterministic model), developed (empirical model) and validated to predict greenhouse tomato yield in terms of quantity (kg/m$^2$).

The final empirical math model developed was based on 3, 6, and 9 week cumulations of light and temperature (in heat units) to predict tomato yields 4 weeks ahead of time.

Error analysis of the empirical method produced the largest root-mean-square-error value of 0.45 kg/m$^2$ for the entire growing season. Predictions during the peak productivity periods (weeks 20 to 35) were fairly good using this approach (RMSE = 0.233 kg/m$^2$). Weaknesses of the empirical model included its site-specific nature, and difficulty in applying to general grower cases.

The final deterministic math model based on first principle equations of net photosynthesis ($P_{net}$) predicted yields 6 weeks ahead using a yield conversion factor and two week cumulations of $P_{net}$. From error analysis, the most accurate predictions were achieved with the deterministic model, with an overall RMSE value of 0.381 kg/m$^2$ for the entire growing season. The deterministic approach was also fairly flexible for general applications but there were a number of weaknesses identified with this approach. The application of a constant conversion factor was based on assumptions that fruit moisture contents and dry matter partitioning remained constant throughout the tomato development period. Ideally, a variable conversion factor reflecting changes in crop growth could be developed and applied to this model to improve prediction accuracy.

Another weakness identified with the deterministic model was the need for an extensive understanding of plant physiological processes to successfully apply the equations used to make the predictions. Further, the requirement for a continuous and complete set of hourly data to make predictions may not be time or cost-effective for many greenhouse growers. In terms of user-friendliness, the deterministic model was ranked last for its large number of plant parameters and assumptions and its inability to provide predictions without a comprehensive set of hourly environmental data.

The predictions from the Agassiz expert system model were used for comparison with the two mathematical models developed. With an overall RMSE value of 0.401 $kg/m^2$ for the entire growing season, the expert system was considered a valid model.

When an overall comparison was made with the two math models developed in this study, the expert model was top-ranked in terms of total score (***Table 3***). Although ranked third for prediction accuracy during the peak productivity periods it became evident that the expert system had the greatest potential for predicting yield variations during these intervals (***Figure 18***). The expert system would be capable of improving its predictive ability with further training with site-specific, historical data sets, to become a fully customized management tool. Another advantage of the Agassiz model was its ability to provide predictions even for cases of missing or ambiguous data.

When the three tomato yield prediction models were compared, the Agassiz expert system was identified as the best overall model for its prediction accuracy, ease of application to both site-specific and general conditions, and overall user-friendliness.

## 5.0 Recommendations

From the results of this study the inherent weaknesses and strengths of each modeling approach became clearly evident.

The empirical modeling approach using mathematical regression techniques was effective for providing accurate yield predictions for a site-specific case, but these site-specific equations could not be readily applied to other greenhouses. In order to successfully apply the approach developed in this study to other sites, the same methodology would have to be repeated to identify the best empirical model. Regression equations with new coefficient values would need to be derived for each individual site and compared with a similar error analysis method. However, an advantage of the empirical approach was that it attempted to predict trends in yield productivity using weekly environmental data. By simply applying the methodology outlined in this study, a grower with a full year of weekly historical data and a spreadsheet program could establish yield predictive equations for their current growing season. Greenhouse growers may favor the empirical modeling approach for its consistent predictability, straight-forward methodology, and low cost for implementation.

The deterministic model applying first principle equations of photosynthesis had the advantage of being applicable to all general cases of tomato crop production. Using a generalized yield conversion factor (Enoch, 1978) and cumulations of net photosynthesis activity, the deterministic model attempted to simulate actual tomato crop productivity. However, in order to improve the accuracy of predictions during different stages of crop growth, the development of a variable yield conversion factor may be necessary. The current yield conversion factor is based on fixed assumptions of dry matter partitioning and moisture content. It is suggested that a yield conversion factor should be developed to account for plant physiological changes in partitioning ratios throughout the various growth phases of the tomato fruit. Modifying the method found in literature (Enoch, 1978) average values for tomato moisture content and partitioning on a weekly or bi-weekly basis could be used to calculate a new series of yield conversion factors for weekly intervals to convert $P_{net}$ values. Further research into the plant physiology of the tomato crop at different developmental stages may be required to properly apply the deterministic model.

The Agassiz expert system used for comparison in this study was identified as a successful method for yield prediction. The structure of the model would allow its application to both site-specific and generalized cases with minimal modification to the model itself. In actual practise, the end-user would only be required establish a historical database and input new environmental data weekly, to predict yields 1, 2, 3, and 4 weeks ahead of time. The success of the expert system is dependent on the data with which it is trained. As a result, it is recommended that for the initial training of the expert system, historical data sets with the most complete sets of parameters and characteristic yields be used as input. Ideally, the more sets of data the model is trained with, the greater its capacity to predict for trends during peak productivity periods.

A recommended improvement to the Agassiz expert model would be to include an algorithm within the model itself, for an on-going error analysis with each week of predictions. Further, when training the system for predictions, error analysis for each predict ahead model should be performed (PRE-1, PRE-2, PRE-3, PRE-4) for each additional year of historical data. The model could be further integrated to provide graphical analysis of actual and predictive yields to allow end-users to monitor predictive performance visually. A report generation program within the model including yield forecasts, weekly error analyses, and graphical plots would be a useful management tool for greenhouse growers. Ideally, a generalized expert system model for tomato crops could be developed based on generalized sets of data from a variety of greenhouse growers. With a generalized yield predictive model, the end-user may only need to customize the expert model by entering one or two years of historical data to establish good predictions for their site-specific case.

Once the Agassiz expert system is fine-tuned to provide accurate predictions for a site-specific case, implementation of yield predictions to an on-line computer climate control system may be feasible. Using computer algorithms related to expected yield values, adjustments may be made to environmental settings to improve tomato crop productivity via the climate control system. Another possibility for the expert model would be to present the end-user with a software driven option to view the yield forecast if modifications are made to environmental settings. Ideally, the greenhouse grower would be allowed to experimentally change the weekly environmental settings within the software program and view the change in productivity if those changes were actually made.

Further, once the expert system software is advanced enough to allow experimental changes to parameters, additional menu options could be added to the model itself. If the user desired to increase or decrease yield productivity, the model could be programmed to provide suggestions to achieve this effect. Eventually, the expert system could be user-customized with default settings to maximize yield during selected intervals and become a companion program to the main climate control algorithm.

Great potential exists for the expert system approach for yield prediction. A similar method could be applied to other greenhouse crops to develop crop and site-specific models for each individual grower. With the encouraging results found in this study, further research and development of expert system modeling is recommended for greenhouse tomato yield prediction.

## 6.0 Bibliography

Acta Horticulturae. 1992. First International Workshop on Sensors in Horticulture. No.304 - March 1992. International Society of Horticultural Science. Netherlands.

Acta Horticulturae. 1979. Symposium on Computers in Greenhouse Climate Control. No. 106 International Society of Horticultural Science. Netherlands.

Acta Horticulturae. 1974. Symposium on Water Supply Under Glass and Plastics. No.35 - January 1974. International Society of Horticultural Science. Netherlands.

Acock, B. D.A. Charles-Edwards, D.J. Fitter, D.W. Hand, L.J. Ludwig, J. Warren-Wilson and A.C. Withers. 1978. The contribution of leaves from different levels within a tomato crop to canopy net photosynthesis: an experimental examination of two canopy models. J. Exp. Bot. 29 (111): 815-827

Alder, H.L, and E.B. Roessler. 1960. Introduction to Probability and Statistics. W.H. Freeman and Company. San Francisco.

Anon. 1990. "A Controlled Summer Climate for a Better Tomato" Groeten en Fruit 45(39): 36-37.

Atherton, J.G. and J. Rudich. 1986. The Tomato Crop - A Scientific Basis of Improvement. Chapman and Hall. New York.

ASAE Standards 1992. "Guidelines for Measuring and Reporting Environmental Parameters for Plant Experiments in Growth Chambers" p.505-508.

Barendse, M. 1993. Temperature is Distributor of Assimilates. Groeten & Fruit Vol. 3(20): 21-22.

B.C.F.A. (British Columbia Federation of Agriculture) 1994. Environmental Guidelines for Greenhouse Growers in British Columbia. B.C. Ministry of Agriculture, Fisheries, and Food. Abbotsford, B.C.

Berenson, M.L, Levine, D.M, and D. Rindskopf. 1988. Applied Statistics - A First Course. Prentice-Hall. New Jersey.

Bochereau, L., P. Bourgine, and B. Palagos. 1992. A method of prediction by combining data analysis and neural networks: application to prediction of apple quality using near infra-red spectra. J. Agric. Eng. Res. 51 (3): 207-216

Chun, C. et al. "Control of Root Environment for Hydroponic Lettuce Production - Rate of root respiration under various dissolved oxygen concentrations." ASAE, June 20-23, 1993. Paper No. 934040

Clapham, Sidney. 1974. The greenhouse book. Newton Abbot. Vermont.

Clegg, P. and D. Watkins. 1978. The Complete Greenhouse Book. Garden Way Publishing. Vermont.

CRC Handbook of Tables for Probability and Statistics. 1966. The Chemical Rubber Company.

de Koning, A.N.M. 1993a. Growth of a Tomato Crop (Measurements for Model Validation). Acta Hort. Vol. 328: 141-146.

de Koning, A.N.M. 1993b. Tomato development time - Research/Tomato. Groeten & Fruit. Vol. 3(11): 17

Enoch, H.Z. 1978. A Theory for Optimalization of Primary Production in Protected Cultivation: I. Influence of Aerial Environment Upon Primary Plant Production. Acta Hort. Vol. 76: 31-57

Freedman, D., Pisani, R., Purves, R., and Ani Adhikari. 1991. Statistics - Second Edition. W.W. Norton and Company. New York.

Giacomelli, G.A. et al. December 18-21, 1990. "Plant Growth Within an Enhanced Environment Greenhouse" ASAE. Paper No. 904531.

Gould, Wilbur A. 1983. Tomato Production, Processing, and Quality Evaluation. AVI Publishing Company Inc. Connecticut.

Graaf de, R. "Automation of the Water Supply of Glasshouse Crops by Means of Calculating the Transpiration and Measuring the Amount of Drainage Water" Acta Horticulturae No. 229. Dec. 1988.

Hack, G.R. 1989. "On-Line Measurement of Plant Growth in the Greenhouse" Acta Horticulturae 248: 337-342

Hand, D.W. 1973. "A null balance method for measuring crop photosynthesis in an airtight daylit controlled-environment cabinet" Agricultural Meteorology 12: 259-270.

Hashimoto, Y. et al. 1985. "Some Speaking Plant Approaches to the Synthesis of Control Systems in the Greenhouse" Acta Horticulturae 174: 219-225.

Hashimoto, Y. et al. 1993. The Computerized Greenhouse - Automatic Control Application in Plant Production. Academic Press Inc. California.

Hazelmere Greenhouses Ltd. 18782 - 16th Ave., Surrey, B.C.

Ho, L.C. and J.D. Hewitt. 1986. "Fruit Development." (Article within the book) The Tomato Crop. Atheron, J.G. and J. Rudich. Chapman and Hall. New York.

Jones, J.W. 1991. Crop growth, development, and production modelling. In: Proceedings of the symposium on automated agriculture for the 21st Century. December 16-17. Chicago, IL. pp. 447-457

Jones, J.W. et al. 1991. A Dynamic Tomato Growth and Yield Model (TOMGRO). Trans. ASAE 34(2): 663-672

Jones, P., J.W. Jones, and Y. Hwang. 1990. Simulation for determining greenhouse temperature setpoints. Trans. ASAE 33(5): 1722-1728

Kozak, Antal. 1995. U.B.C. Department of Forestry. Personal communication with Dr. Kozak validating the RMSE method as analysis tool and basis for comparing yield predictive models.

Lacroix, R., R. Kok, and L. Gauthier. 1993. Simulation-based determination of greenhouse temperature setpoints. Paper no. 934043. ASAE/CSAE Summer Meeting, Spokane, June 20-23, 1993. 15 pp.

Lau, A.K. and Staley, L.M. "Solar radiation transmission and capture in Greenhouses" Canadian Agricultural Engineering, p.205-213. March 1989.

Lau, Anthony K. 1993 Notes/File on greenhouse management, BIOE 456 course work, Fall 1993.

Levine, R.I, Drang, D.E., and B. Edelson. 1990. AI and Expert Systems - A Comprehensive Guide. Mc-Graw Hill Publishing Company. New York.

Lin, Wei C. (Ph.D.) 1994. Post Harvest Physiologist - Agassiz Research Station. Agriculture Canada Research Branch. P.O. Box 1000, Agassiz, B.C. V0M 1A0

Lin, W.C. and J.C.W. Keng. 1989. "Use of Digital Image Processing for Non-Destructive Measurement of Cucumber Fruit Color" CSAE. Paper No.PNR89-304. Sept 24-25/89.

McAvoy, R.J. et al. 1989. The Effect of Total Available Photosynthetic Photon Flux on Single Truss Tomato Growth and Production. J. of Hort. Science 64(3): 331-338.

Ministry of Agriculture, Fisheries, and Food (Province of B.C.). 1993. Greenhouse Vegetable Production Guide - 1993/94.

Ohta, H. et al. "Measurement of Photosynthetic and Transpiration Rates under low total pressures" ASAE, June 20-23, 1993. Paper No. 934009.

Quattro-Pro Spreadsheet Program Version 4. 1992. Borland International Inc.

Rehbein, D.A., S.M. Maze, and J.P. Havener. 1992. The Application of Neural Networks in the Process Industry. ISA Transactions 31(4): 7-13

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

Systat Version 5.03 - Statistical Package. 1991. Systat, Inc.

Stangehellini, C. 1985. Transpiration and temperature of greenhouse crops in relation to internal and external resistances. Acta Hort No.174: 87-95

Stangehellini, C. and W. Th.M. van Meurs. 1992. Environmental Control of Greenhouse crop transpiration. J. Agric. Eng. Research 51(4): 297 - 311

Takakura, T. 1993. Climate under Cover - Digital Dynamic Simulation in Plant Bio-Engineering. Kluwer Academic Publishers. Netherlands.

Takakura, Tadashi. ASAE. Dec 16-17, 1991. Proceedings of the 1991 Symposium on "Automated Agriculture for the 21st Century." Chicago, IL.

Tanaka, A. Fujita, K, and K. Kikuchi. 1974. Nutri-physiological studies on the tomato plant. III. Photosynthetic rate of individual leaves in relation to the dry matter production of plants. Soil Sci., Pl. Nutri., 20, 173-83.

Tang, Winston C. 1995. The Prediction and Validation of Greenhouse Tomato Yield using Mathematical Models and Expert Systems.
Contact Address: 5051 Quebec Street, Vancouver, B.C., V5W 2N3
Contact Ph. Number: (604) 325-3847

Willits, D.H. 1993. "The Effect of Evaporative Cooling on the Efficiency of External Greenhouse Shade Cloths". ASAE. Paper No. 93-4042.

Wolf, S., J.Rudich, A. Marani and Y. Rekah. 1986. Predicting harvesting date of processing tomatoes by a simulation model. J. Amer. Soc. Hort. Sci. 111(1): 11-16

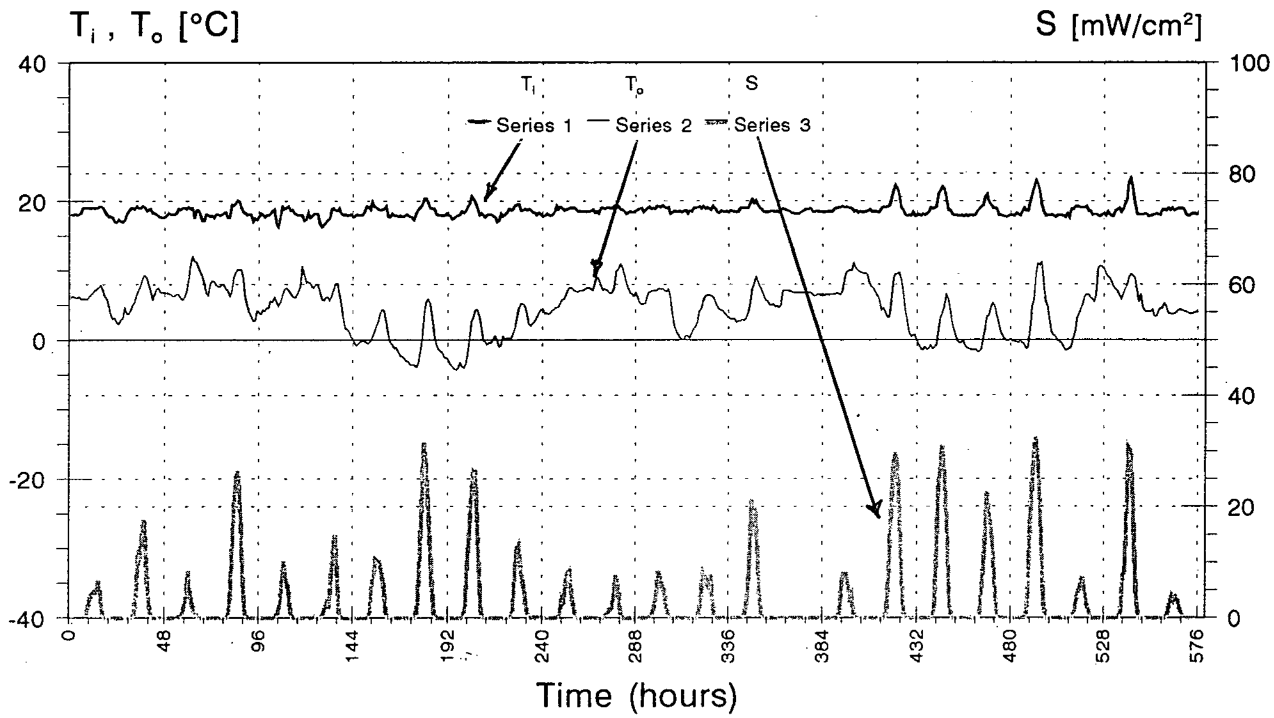
Yang, X. et al. "Dynamic Modelling of the Microclimate of a Greenhouse Cucumber Row-Crop Part I. Theoretical Model" ASAE. Vol. 33(5) Sept-Oct/1990. p.1701-1709

Yang, X. et al. "Dynamic Modelling of the Microclimate of a Greenhouse Cucumber Row-Crop Part II. Validation and Simulation." ASAE. Vol. 33(5) Sept-Oct/1990 p.1710-1716.

# APPENDIX A

## Inside Temperature ($T_i$), Outside Temperature ($T_o$), & Solar Intensity (S)

Data Set: January 1 to January 24 (920109)

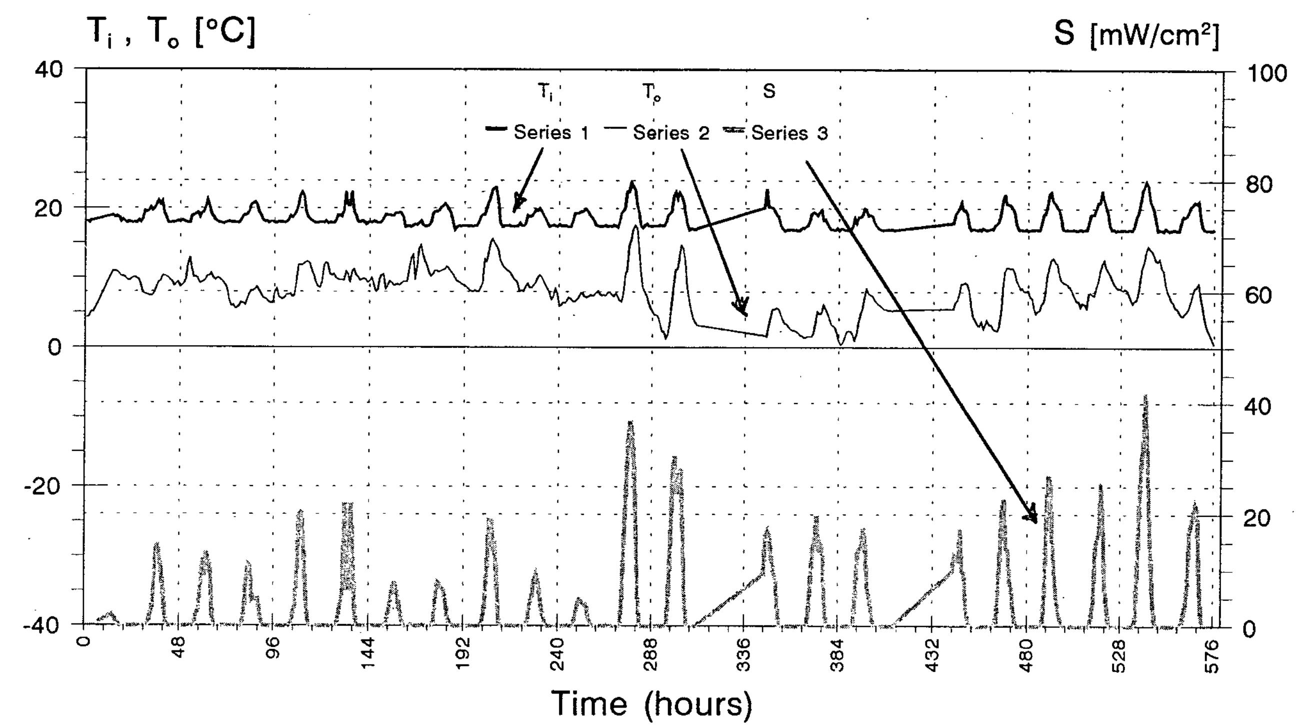



24 days: Hours 0 to 576

Figure A-1 Inside & Outside Temperature & Solar Intensity Plots - January 1992

Inside Temperature ($T_i$), Outside Temperature ($T_o$), & Solar Intensity (S)

Data Set: January 25 to February 17 (920109)

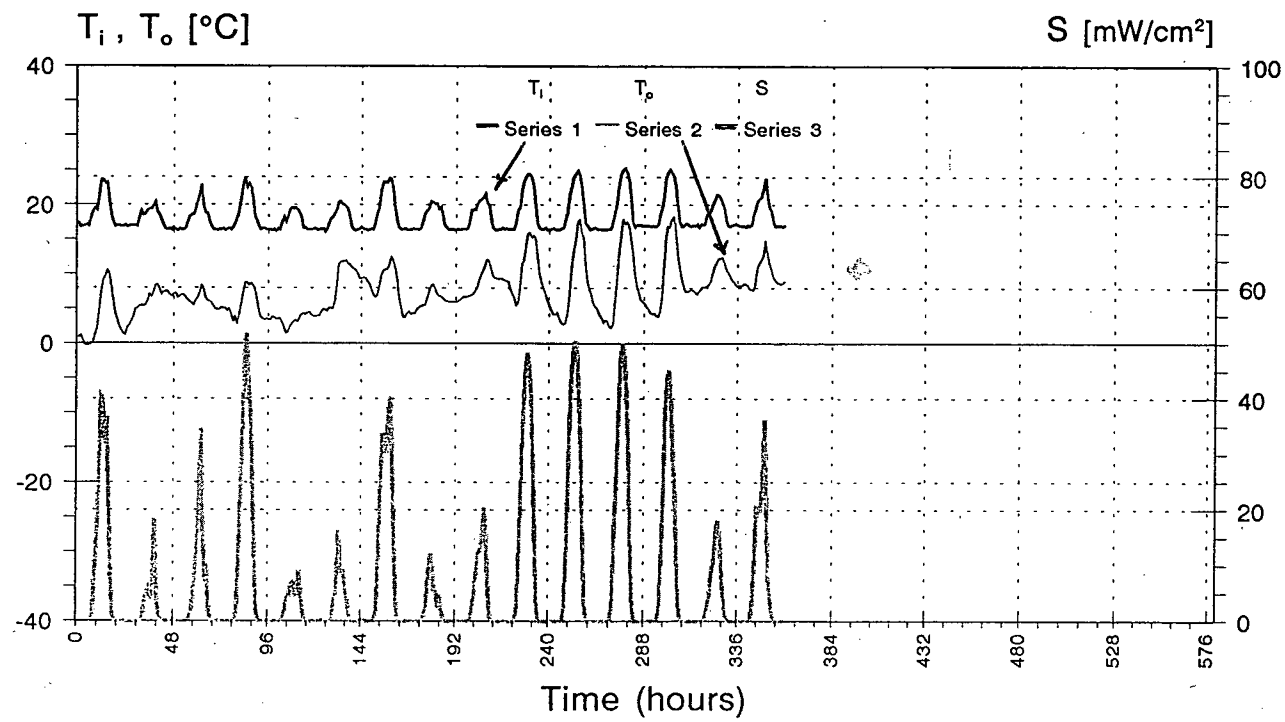



24 days: Hours 0 to 576

Figure A-2 Greenhouse Data - February Plots

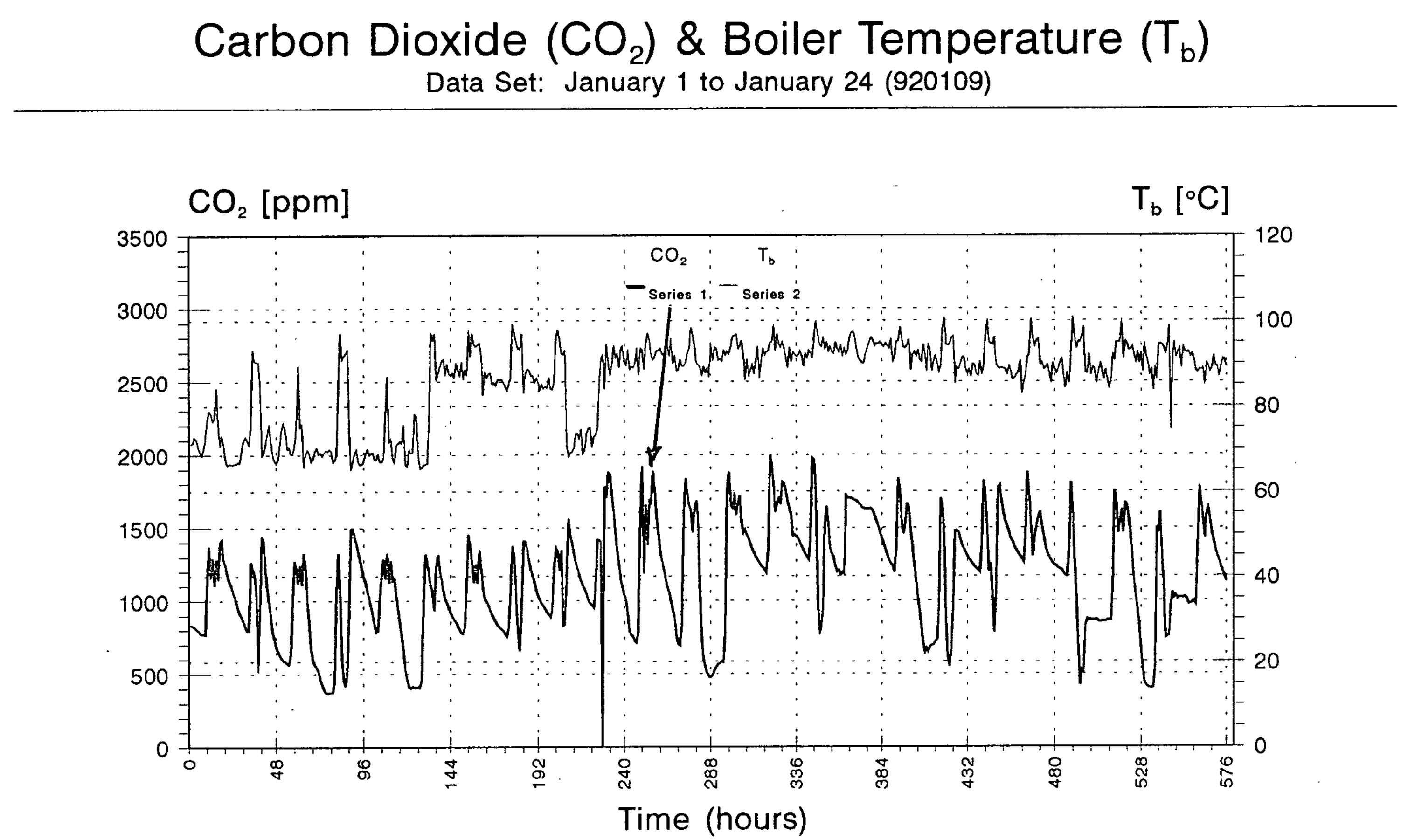
# Inside Temperature ($T_i$), Outside Temperature ($T_o$), & Solar Intensity (S)

Data Set: February 18 to March 5 (920109)




24 days: Hours 0 to 576

Figure A-3 Greenhouse Data - March 1992 Plots

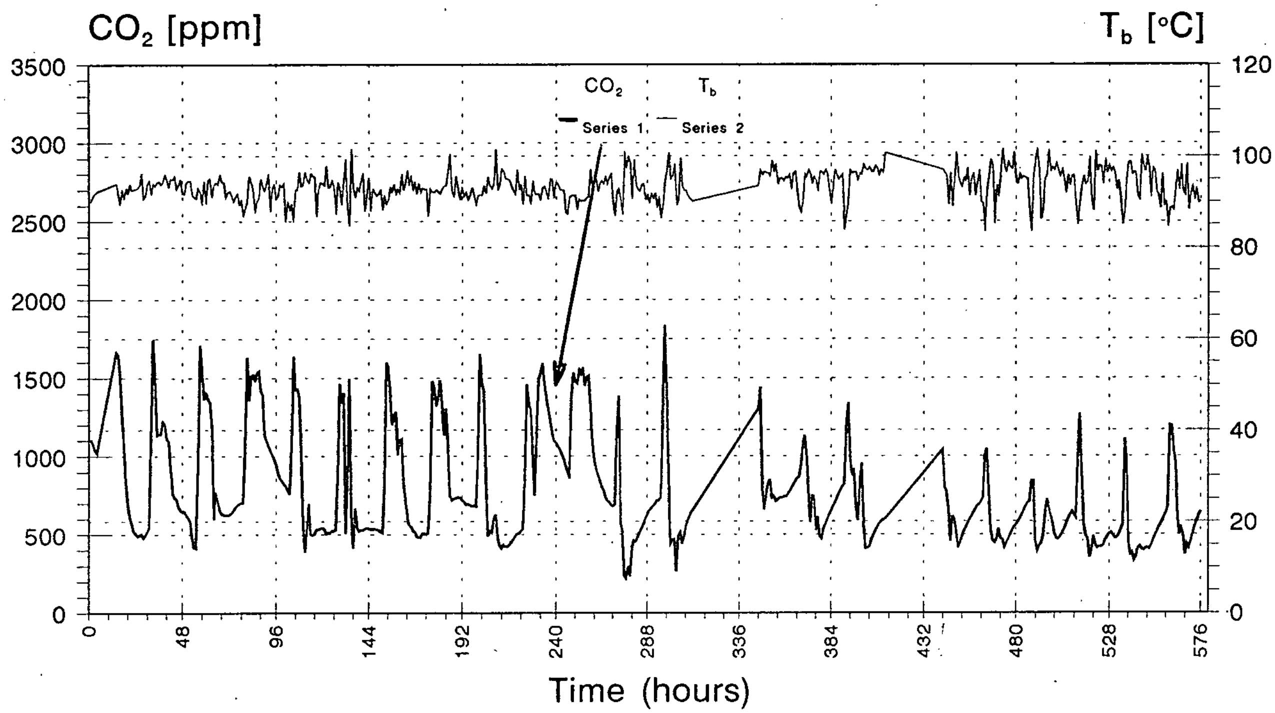



24 days: Hours 0 to 576

Figure A-4 Carbon Dioxide Plots for January 1992

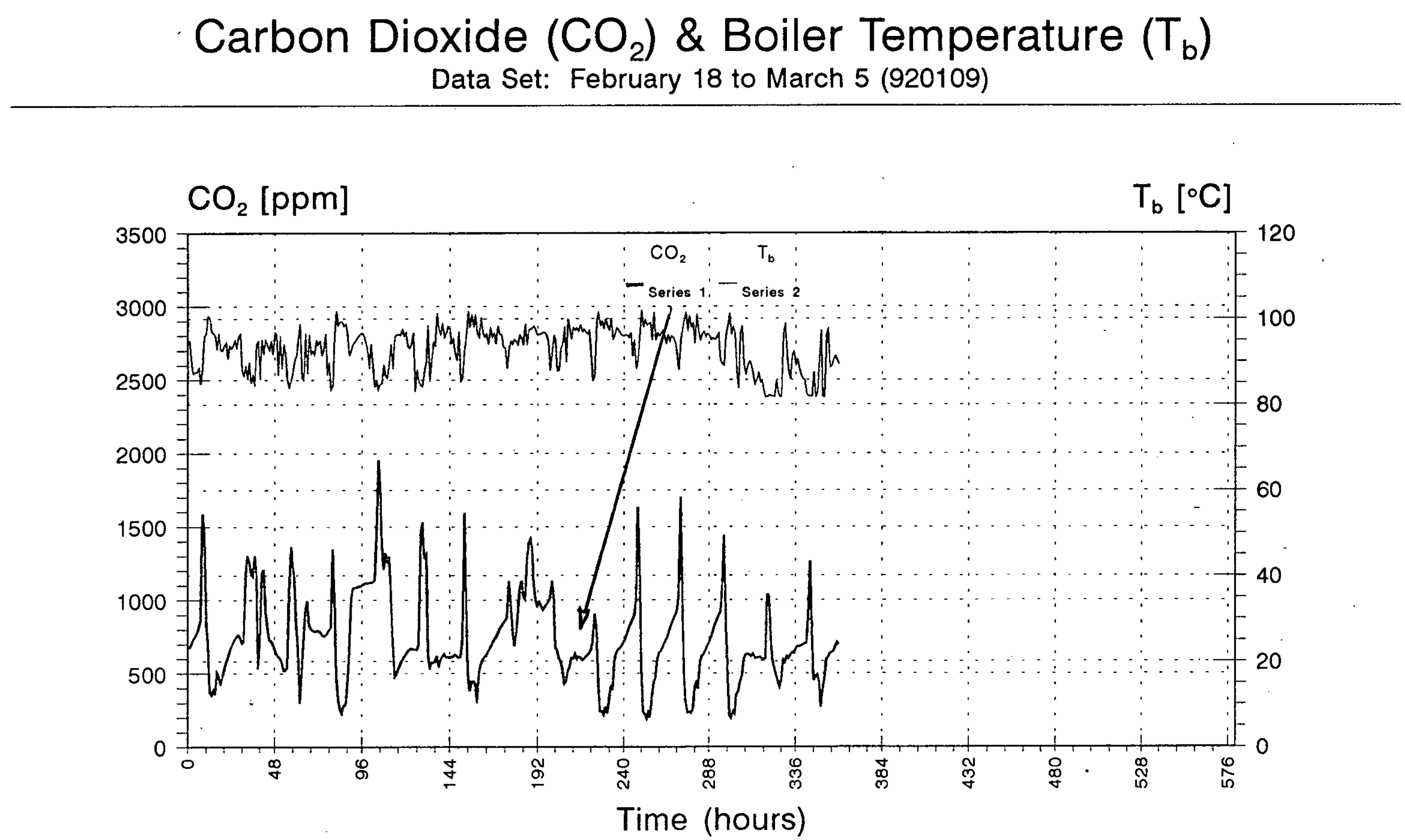
# Carbon Dioxide ($CO_2$) & Boiler Temperature ($T_b$)

Data Set: January 25 to February 17 (920109)



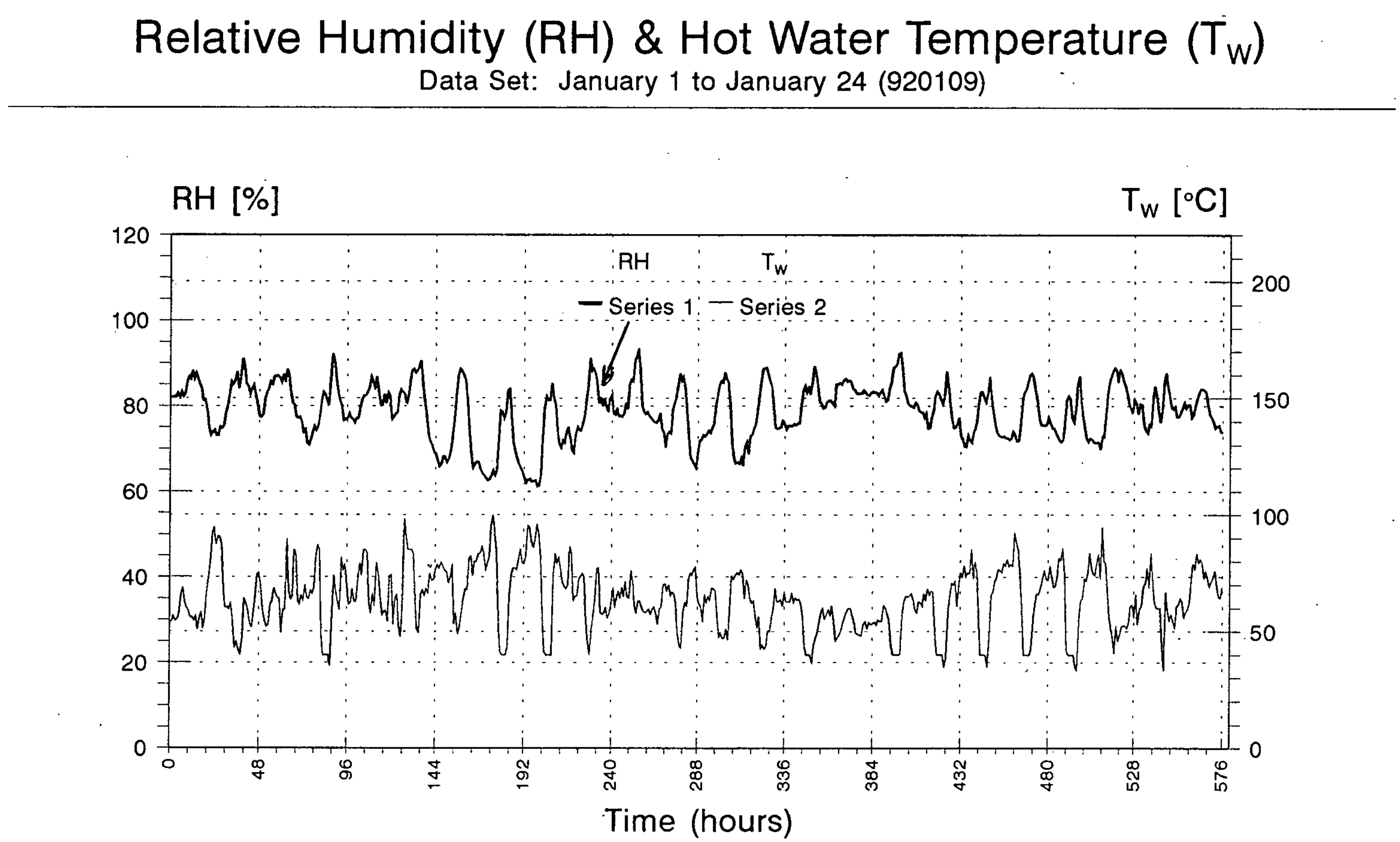

24 days: Hours 0 to 576

Figure A-5 Carbon Dioxide Plots for February 1992



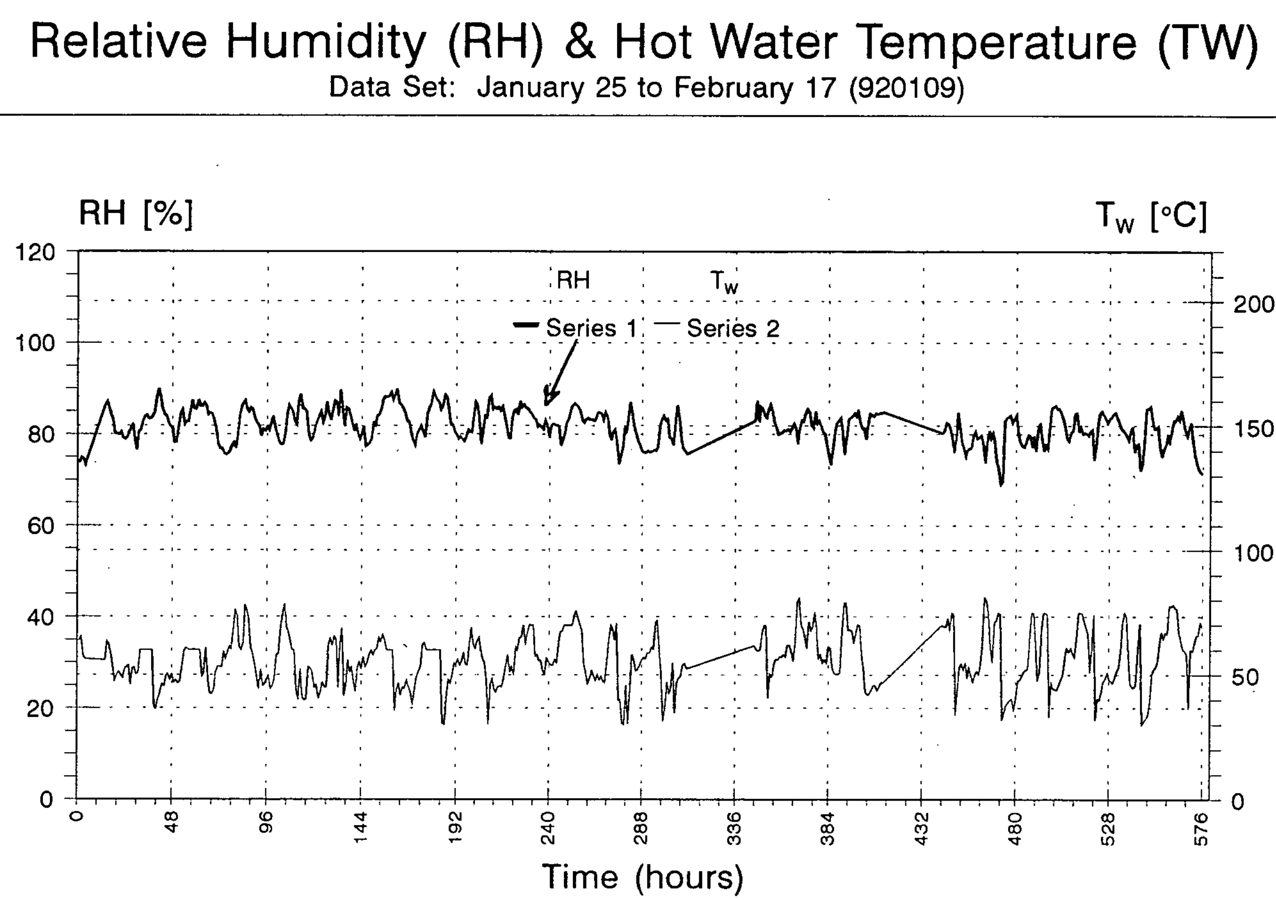

24 days: Hours 0 to 576

Figure A-6 Carbon Dioxide Plots for March 1992




24 days: Hours 0 to 576

Figure A-7 Relative Humidity Plots for January 1992

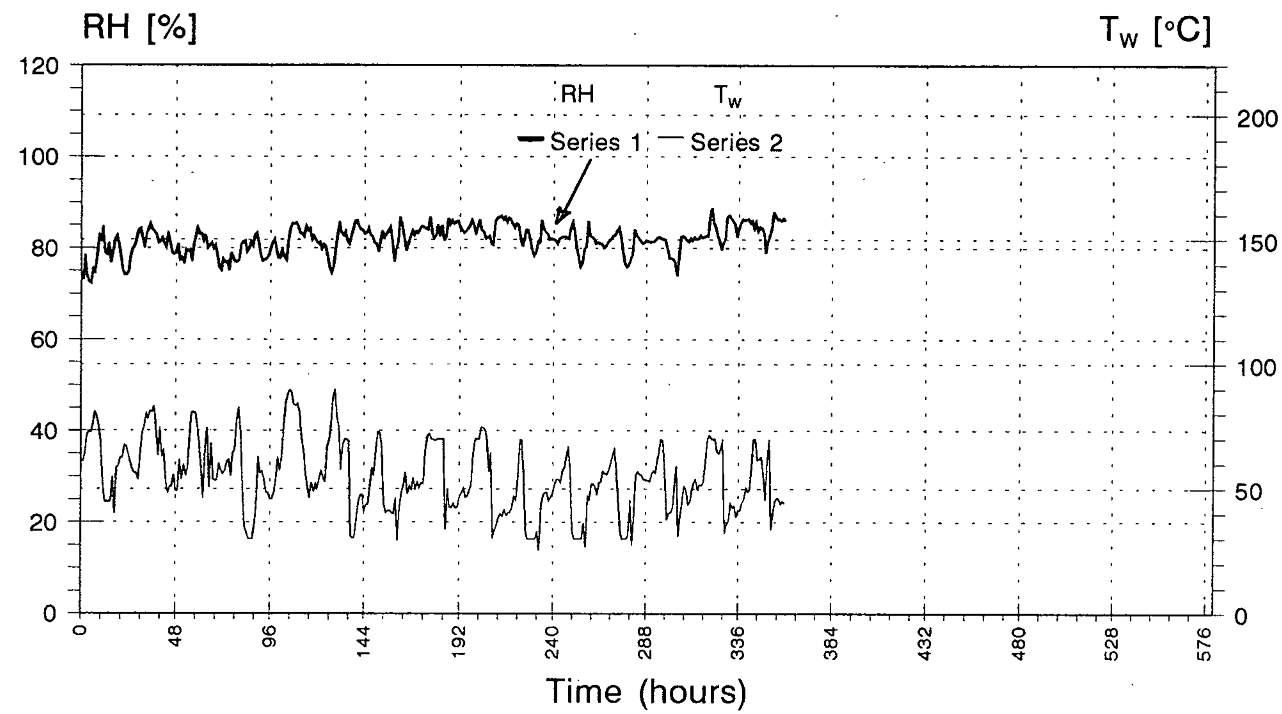



24 days: Hours 0 to 576

Figure A-8 Relative Humidity Plots for February 1992

## Relative Humidity (RH) & Hot Water Temperature ($T_W$)

Data Set: February 18 to March 5 (920109)




24 days: Hours 0 to 576

Figure A-9 Relative Humidity Plots for March 1992

# APPENDIX B

CROP YIELD MATHEMATICAL MODEL

## GROSS PHOTOSYNTHESIS (Pg)

Based on Eqn (7) in "Crop Growth, Development, and Production Modelling" by J.W. Jones

Pg = D ( (t*C*p($)) / K ) * { ln [( &*K*Io + (1 - m)* t * C *p($)) / ( &*K*Io*exp (-KL) + (1 - m)*t*C*p($) ) ] }

### For Tomato Crop / Variables & Unit Definitions:

### FOR: GROSS PHOTOSYNTHESIS (Pg)

| Term | Abbrev. | Units |
|---|---|---|
| Gross Photo. | Pg | g (CH2O) / m ^ 2*h |
| Conv. Factor | D | ----------- |
| Leaf Conduct.to CO2 | t | umol (CO2) / m ^ 2*s |
| CO2 conc. of air | C | umol(CO2) / mol (air) = ppm |
| Function of Temp. | p($) = p(theta) | Dimensionless<br>Note: " $ " = Theta (for notation) |
| Canopy L.E. Coeff. | K | Dimensionless |
| Leaf L.U. Eff. | & = alpha | umol (CO2) / umol (photon) |
| Ligh Flux Density | Io | umol (photon)/m ^ 2*s<br>** Conversion from W/m ^ 2 |
| Respiration rate @ 25*C | km | 0.0006 |
| Light T. Coeff. Leaves | m | Dimensionless |
| Leaf Area Index (LAI) | L | m ^ 2(leaf) / m ^ 2(ground) |

### FOR: p($)

p($) = [ 1 - ( 1 - { ( $h - $ ) / ( $h - $i ) } ^ 2 ]

p($) = expresses the effect of temperature on the max. rate of photo. for a single leaf
$h = temp at which leaf photosynthesis is maximum, 30*C
$i = is temp below which leaf photosynthesis is zero, 5*C
$ = measured temp = Tin = temperature inside

## RESPIRATION RATE

Based on Eqn (5) in J.W. Jones' paper

Rm = km * exp (0.0693*[$ - 25])

Rm = maintenance respiration rate, g CH2O / g tissue * h
km = respiration rate at 25*C, g CH2O / g tissue * h
$ = measured temp.

## DRY WEIGHT

Based on Eqn (6) in J.W. Jones' paper

$$dW / dt = E^* ( Pg - Rm^*W )$$

dW / dt = rate of dry wt. growth of crop, g tissue / m ^ 2*h
W = total plant dry weight, g/m ^ 2
E = conversion efficiency of CH2O to plant tissue, g (tissue) / g(CH2O)
Pg = canopy gross photosnthesis rate, g(CH2O) / m ^ 2*h
Rm = maintenance respiration rate, g CH2O / g tissue *h

Then, once dW/dt is known the new weight W(new) can be found:

$$W(new) = W(i) + dW / dt$$

W(new) = new weight, g/m ^ 2 for that HOUR
W(i) = total initial plant dry weight, g/m ^ 2
dW/dt = rate of dry weight growth of the crop ( g tissue / m ^ 2*h)

## MULTIPLICATION FACTOR: Y

Based on paper by H.Z. Enoch (Acta Hort 76, 1978 p. 48)

$$Y = [ (30 * 100) / (44 * 5) ] * X \qquad \text{or:} \quad 30/44 * X = Y * (5/100)$$

Y = multiplication factor to relate photosynthesis to yield
X = percentage of dry matter that is yield (ie . 50% = 0.50 )
30, 44 = Mol. Weights of CH2O and CO2 respectively
5/100 = represents 5% dry matter in yield.

Based on Paper: Y = 7 (approx., based on X= 50% )

## YIELD VALUE #1

Based on Pg and multiplication factor "Y"

Yield Value #1 = Pg * Y

## NET PHOTOSYNTHESIS

$$Pnet = Pg - Rm^*W$$

Pnet = net photosynthesis, g (CH2O) / m ^ 2*h
Rm = maintenance resp. rate, g (CH2O) / g tissue*h
W(new) = total plant dry weight, g/m ^ 2 = W = W(i) for the first assumed weight hour
Pg = gross photosynthesis, g (CH2O) / m ^ 2*h

## YIELD VALUE #2

Based on Net Photosynthesis (Pnet) and Multiplication Factor "Y"

Yield Value #2 = Pnet * Y

## YIELD CONVERSION FROM GROWER's DATA:

Weight ( kg / m^2 ) = [ Total number of 20 pound cases * 20 / (2.2*25400) ]

** Greenhouse Area = 25,400 m^2

## LIGHT INTENSITY CONVERSION FROM W/m^2 to umol / m^2*s

Io = umol / m^2*s = ( I, W/m^2 / 0.22 ) * 0.80 * 0.45

Io = converted units from I, raw data entered as W/m^2
I = raw data from grower in units of W/m^2

## LEAF AREA INDEX (LAI) , L

Separate Column for varying LAI, L

Shaded Areas are
= Formula Blocks

$i = 5.0 and $h = 30.0

To Follow is the Spreadsheet Layout for the Deterministic Math Model

Sample Spreadsheet Column Headings for the Deterministic Model

| $I_o$ | LAI | $P_g$ | $R_m$ | $P_{net}$ | 24 Hour Yield #2 | Y Factor |
|---|---|---|---|---|---|---|
| | | | | | | |

$I_o$ = Light Flux Density (umol (photon) / $m^2$ *s )

LAI = Leaf Area Index

$P_g$ = Gross Photosynthesis ( $g(CH_2O)/m^2$ *h )

$R_m$ = Maintenance Respiration Rate ( $g(CH_2O)$ / g tissue* h)

$P_{net}$ = Net Photosynthesis ( g $(CH_2O)$ / $m^2$ * h)

24 Hour Yield #2 = Yield for a 24 hour period (kg / $m^2$ ) based on Net Photosynthesis

Y Factor = Enoch's Yield factor of 7

Sample Spreadsheet Layout for the DETERMINISTIC MATH MODEL

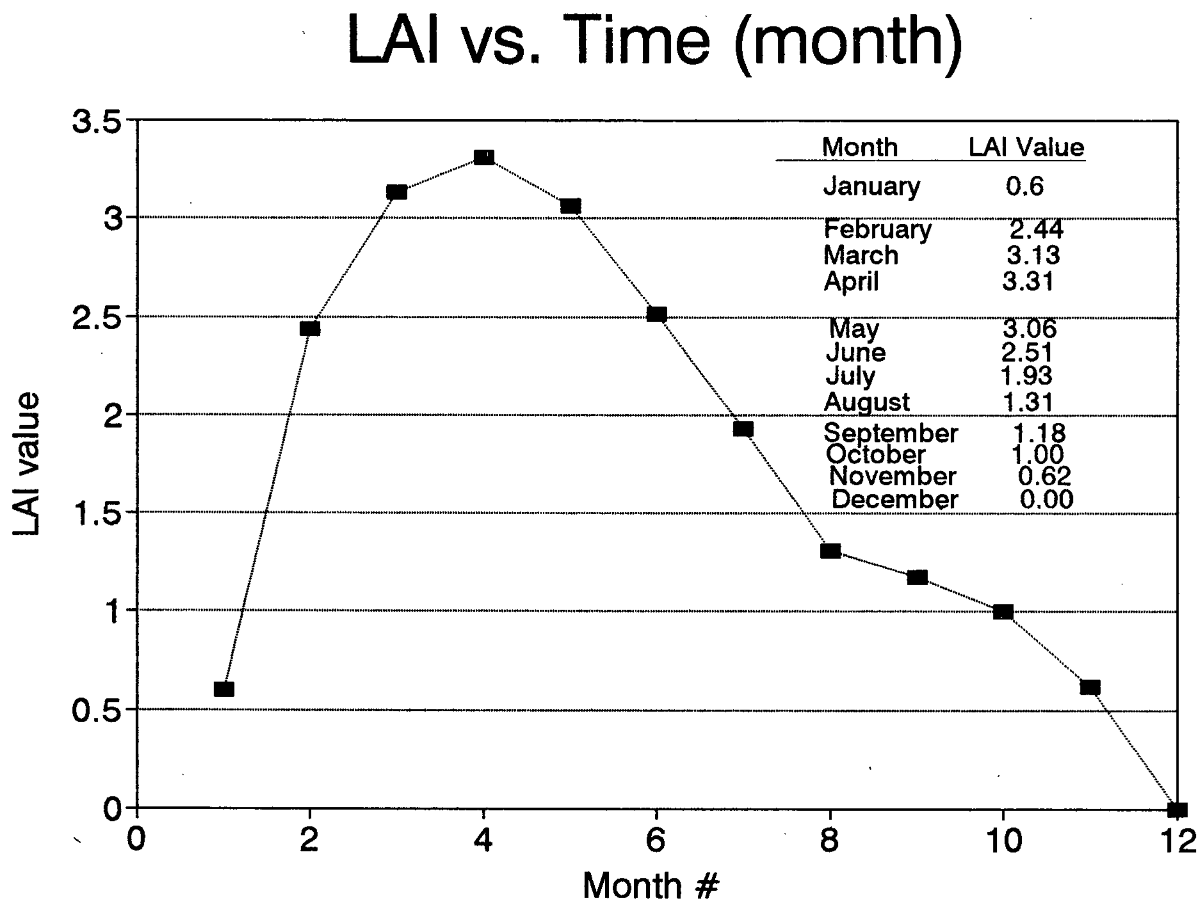



Figure A-10 Changes in Leaf Area Index (LAI) throughout the year.
Reference: de Koning. 1993

# APPENDIX C

**Summary of:**

**Total PPF vs. Crop Yield Results**

| Year | $r^2$ | Equation |
|---|---|---|
| 1987 | 0.722917 | y = 0.95 x - 35.28 |
| 1989 | 0.711196 | y = 0.99 x - 29.43 |
| 1990 | 0.618388 | y = 1.12 x - 28.23 |
| 1991 | 0.613489 | y = 1.07 x - 41.15 |
| **1992** | **0.746478** | **y = 0.99 x - 33.53** |
| 1993 | 0.672635 | y = 1.02 x - 68.31 |
| 1987 - 1993 | 0.668588 | y = 0.00025 x - 0.07858 |

**Where:**

**x** = 9 week cumulated photosynthetic photon flux (PPF)
**y** = crop yield in grams/plant

Given yield in $kg/m^2$, and plant density of 2.5 plants/$m^2$:

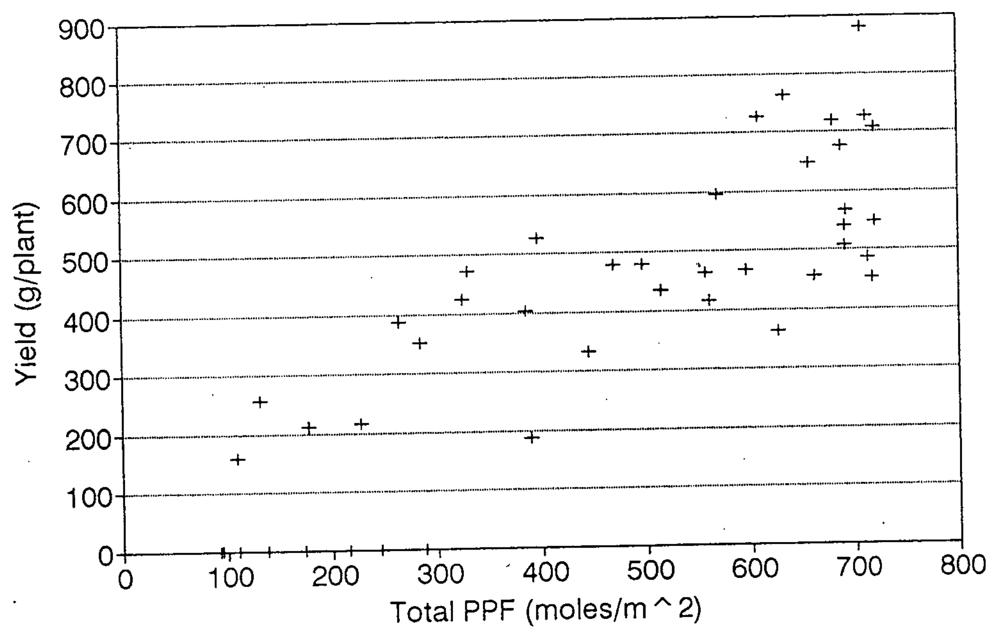
Crop Yield (grams/plant) = Yield ($kg/m^2$) *(Plants / $m^2$)* (1000 g / kg)

The best correlated data for light (9 week PPF) and crop yield (g/plant) as analysed by the $r^2$ correlation coefficient was found to be the 1992 data with a value of 0.746478.

Table A-1

## 1987: Total PPF vs. Crop Yield

[ 9 wk. summations of PPF ]




Regression Output:

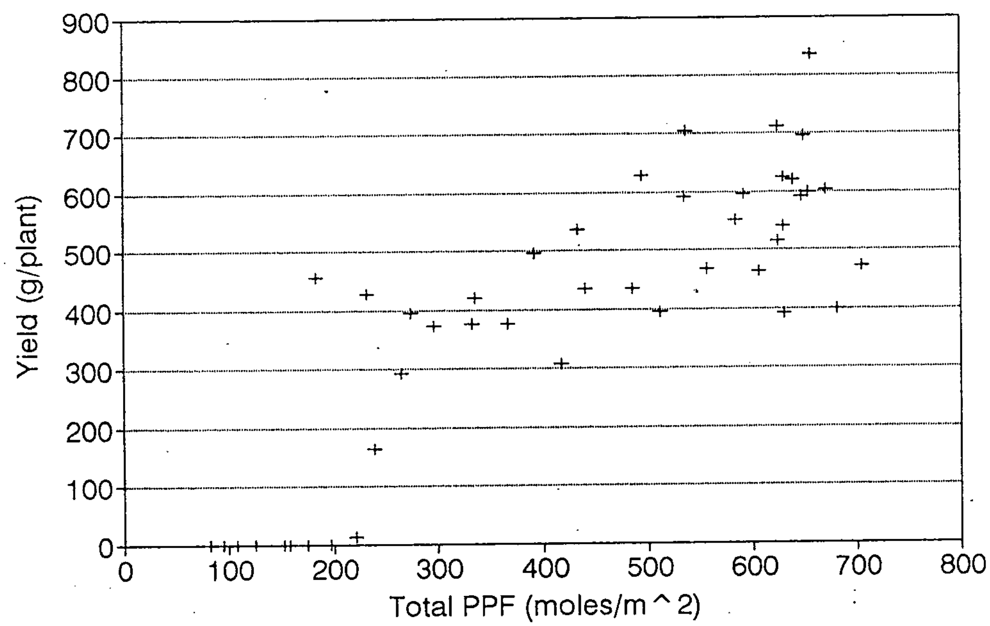
| | | |
|---|---|---|
| Constant | | -35.2756 |
| Std Err of Y Est | | 130.2987 |
| R Squared | | 0.722917 |
| No. of Observations | | 44 |
| Degrees of Freedom | | 42 |
| X Coefficient(s) | 0.95092 | |
| Std Err of Coef. | 0.090841 | |

1987 Eqn: y = 0.95092 X - 35.2756

Figure A-11

# 1989: Total PPF vs. Crop Yield

[ 9 wk. summations of PPF ]




Regression Output:

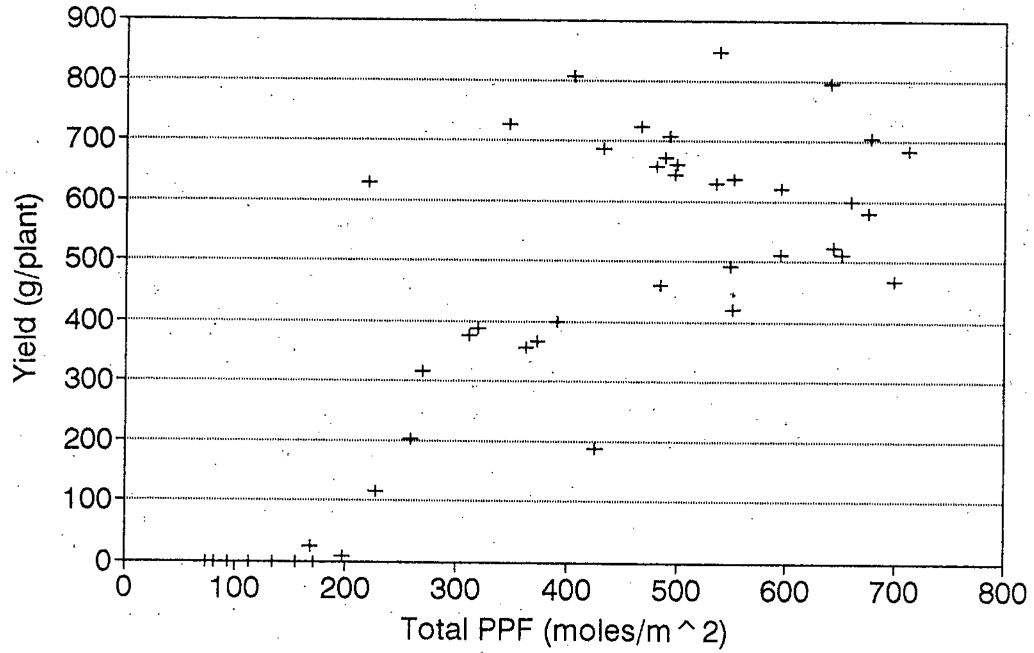
| | | |
|---|---|---|
| Constant | | -29.4388 |
| Std Err of Y Est | | 129.1234 |
| R Squared | | 0.711196 |
| No. of Observations | | 44 |
| Degrees of Freedom | | 42 |
| X Coefficient(s) | 0.996418 | |
| Std Err of Coef. | 0.097977 | |

1989 Eqn: $y = 0.996418 \cdot X - 29.4388$

Figure A-12

# 1990: Total PPF vs. Crop Yield

[ 9 wk. summations of PPF ]



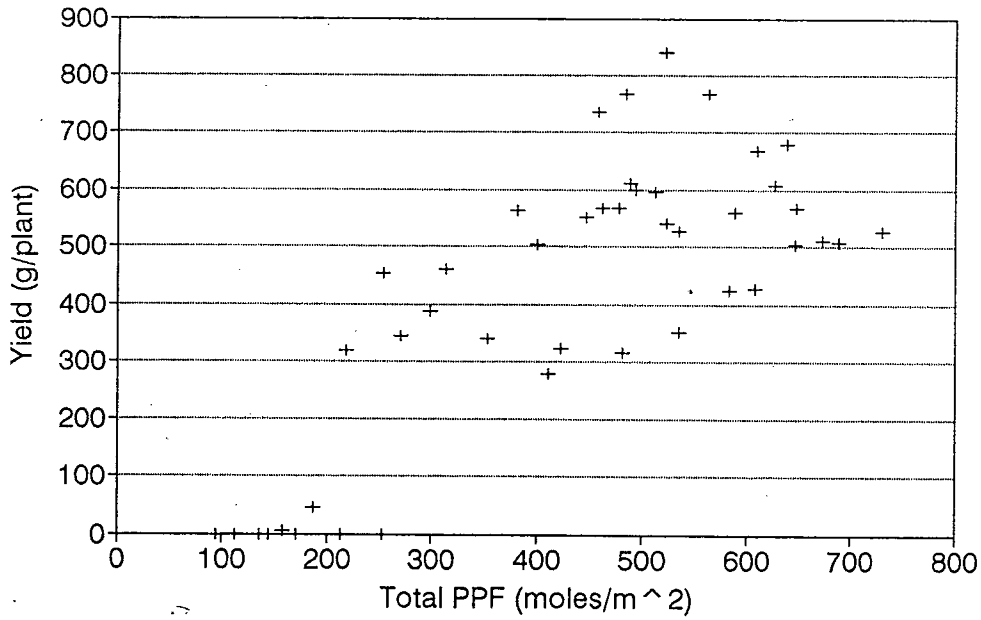

| Regression Output: | | |
|---|---|---|
| Constant | | -28.2261 |
| Std Err of Y Est | | 171.2511 |
| R Squared | | 0.618388 |
| No. of Observations | | 44 |
| Degrees of Freedom | | 42 |
| X Coefficient(s) | 1.119628 | |
| Std Err of Coef. | 0.135716 | |

1990 Eqn: $y = 1.119628\ X - 28.2261$

Figure A-13

# 1991: Total PPF vs. Crop Yield

[ 9 wk. summations of PPF ]




Regression Output:

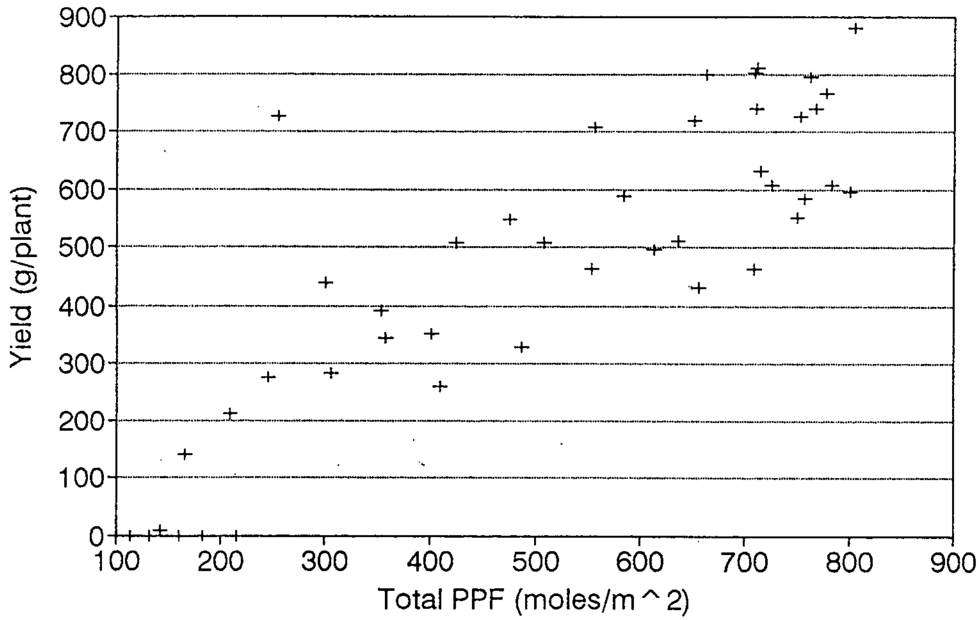
| | | |
|---|---|---|
| Constant | | -41.1495 |
| Std Err of Y Est | | 154.496 |
| R Squared | | 0.613489 |
| No. of Observations | | 44 |
| Degrees of Freedom | | 42 |
| X Coefficient(s) | 1.072094 | |
| Std Err of Coef. | 0.131306 | |

1991 Eqn: y = 1.072094 X - 41.1495

Figure A-14

# 1992: Total PPF vs. Crop Yield

[ 9 wk. summations of PPF ]




Regression Output:

| | | |
|---|---|---|
| Constant | | -33.5333 |
| Std Err of Y Est | | 137.0291 |
| R Squared | | 0.746478 |
| No. of Observations | | 44 |
| Degrees of Freedom | | 42 |
| X Coefficient(s) | 0.987691 | |
| Std Err of Coef. | 0.088817 | |

1992 Eqn: y = 0.987691 X - 33.533

Figure A-15

# 1993: Total PPF vs. Crop Yield

[ 9 wk. summations of PPF ]

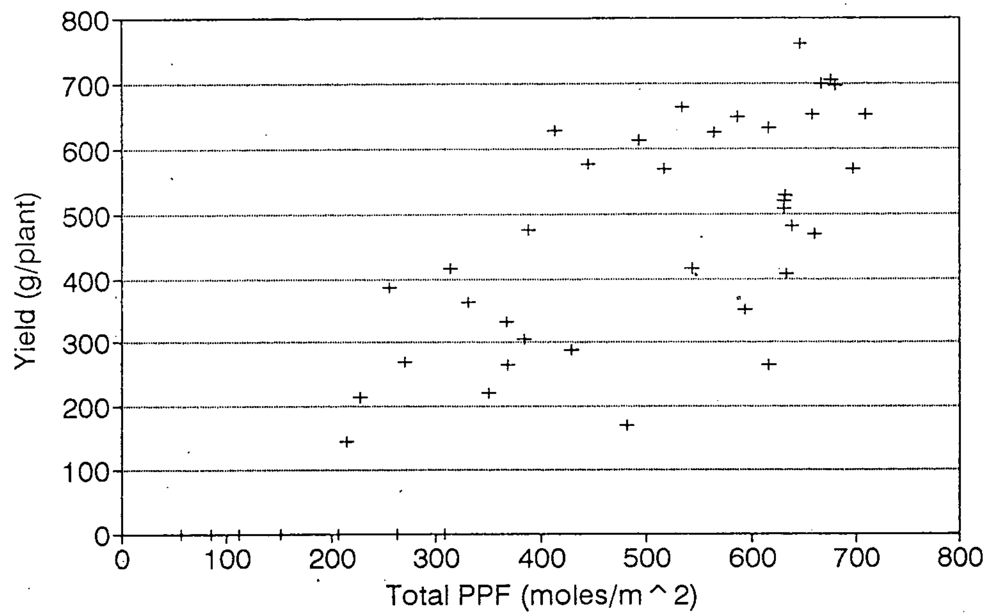



Figure A-16

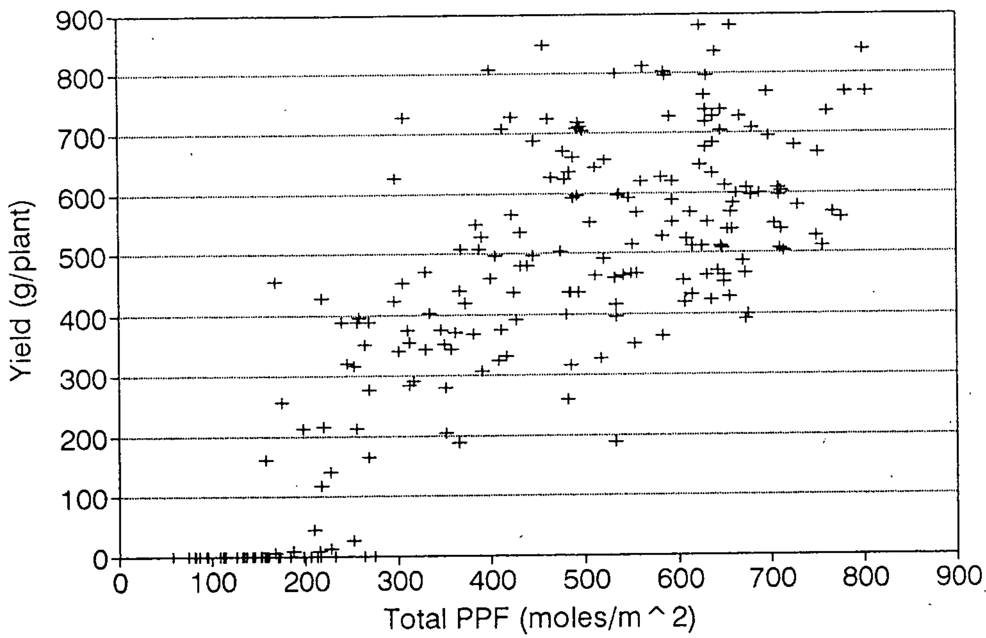
Regression Output:

| | | |
|---|---|---|
| Constant | | -68.3091 |
| Std Err of Y Est | | 136.6947 |
| R Squared | | 0.672635 |
| No. of Observations | | 44 |
| Degrees of Freedom | | 42 |
| X Coefficient(s) | 1.021312 | |
| Std Err of Coef. | 0.109941 | |

1993 Eqn: y = 1.021312 X - 68.3091

## 1987-1993: Total PPF vs. Crop Yield

[ 9 wk. summations of PPF ]




Regression Output:

| | | |
|---|---|---|
| Constant | | -0.07858 |
| Std Err of Y Est | | 0.361 |
| R Squared | | 0.668588 |
| No. of Observations | | 264 |
| Degrees of Freedom | | 262 |
| X Coefficient(s) | 0.002512 | |
| Std Err of Coef. | 0.000109 | |

ALL Eqn: y = 0.002512 X - 0.07858

Figure A-17

# APPENDIX D

## EMPIRICAL MATH MODEL REGRESSION RESULTS (Equation 1)

REGRESSION RESULTS FOR LIGHT ALONE ($CL_9$)
(Cumulative Light for 9 weeks, $W/m^2$)

$$Y_{p2} = a_0 + a_1 (CL_9) \;\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

| | $a_0$ | $a_1$ | RMSE Results for 1994 Predictions | |
|---|---|---|---|---|
| 1987 Equation......... | -0.08819 | 0.000014 | 0.489258 | |
| 1989 Equation......... | -0.0736 | 0.0000147 | 0.441116 | |
| **1990 Equation........** | **-0.07057** | **0.0000165** | **0.382832** | **(Eqn. 1.1)** |
| 1991 Equation......... | -0.10287 | 0.0000158 | 0.408559 | |
| **1992 Equation........** | **-0.08383** | **0.0000145** | **0.457382** | **(Eqn. 1.2)** |
| 1993 Equation......... | -0.17077 | 0.0000150 | 0.482722 | |
| 1987-1993 Eqn....... | -0.07858 | 0.00001479 | 0.439355 | |

**Where**: $a_0$ is the constant for predictive equation 1
$a_1$ is the coefficient for predictive equation 1 ($CL_9$)

"1992 Equation" means the regression equation based on fitting 1992 data to regression equation 1.

RMSE for 1994 Predictions: is the calculated root-mean-square-error of tomato yield predictions for 1994 using that particular year's regression equation.

For example, using the 1992 Equation of:
$Y = -0.08383 + 0.0000145\,(CL_9)$ to predict 1994 yield (with 1994 data), the root-mean-square-error of the predictions was found to be: 0.457382

Table A-2

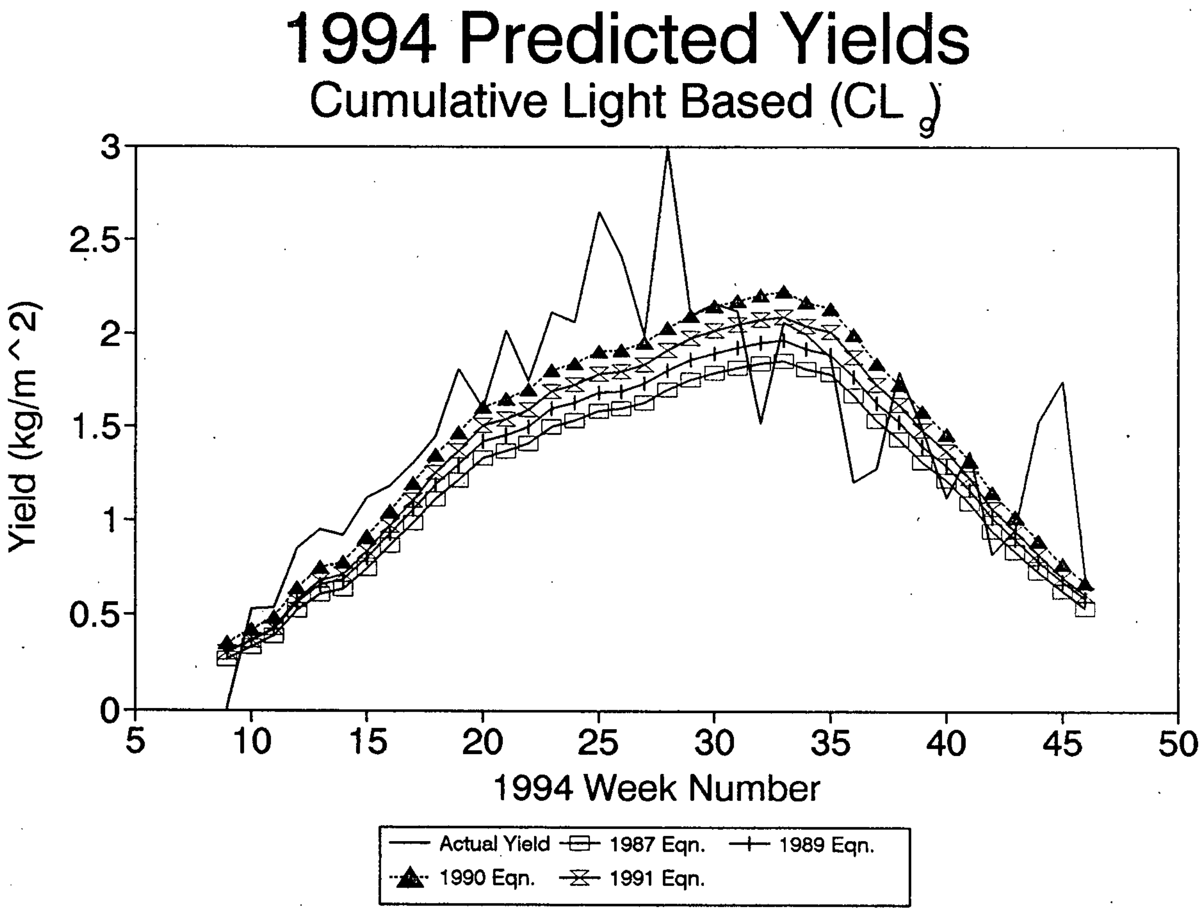



Figure A-18 Predicted yields for 1994 production based on regression equations derived from light data from 1987 to 1991.

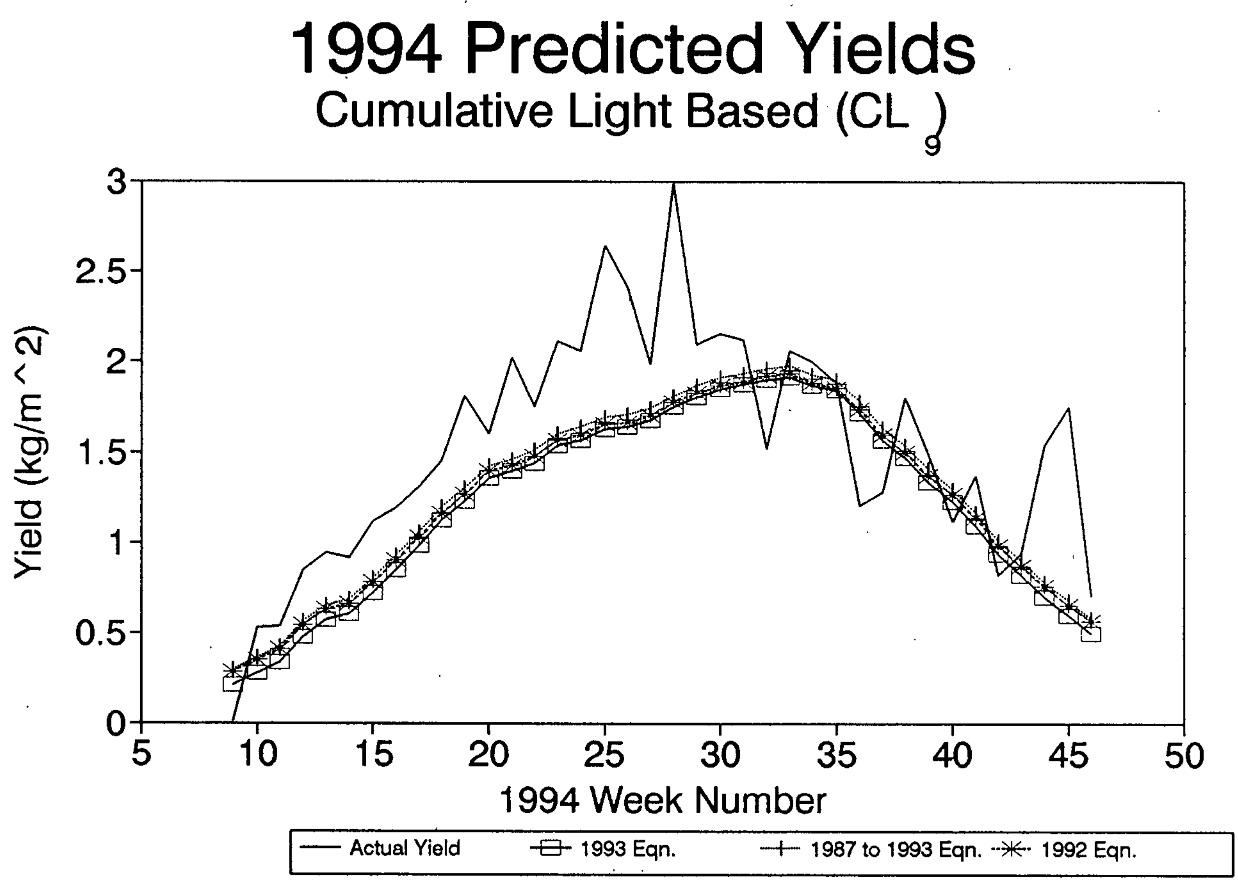



Figure A-19 Predicted yields for 1994 production based on regression equations derived from light data (1992, 1993, and '87 - '93)

## EMPIRICAL MATH MODEL REGRESSION RESULTS (Equation 2)

REGRESSION RESULTS FOR LIGHT and HEAT ($CL_9$ and $CH_9$)
(Cumulative Light for 9 weeks, $W/m^2$ and Cumulative Heat Units for 9 weeks)

$$Y_{p2} = a_0 + a_1 (CL_9) + a_2 (CH_9)\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

| | $a_0$ | $a_1$ | $a_2$ | RMSE Results for 1994 Predictions | |
|---|---|---|---|---|---|
| 1987 Equation......... | 0.52199 | 0.00001207 | -0.0043454 | 0.440142 | |
| 1989 Equation......... | 4.25232 | 0.00002708 | -0.057994 | 0.692027 | |
| **1990 Equation........** | **2.38054** | **0.00002444** | **-0.033809** | **0.542205** | **(Eqn. 2.2)** |
| 1991 Equation......... | 3.76864 | 0.00002193 | -0.0480305 | 0.527883 | |
| **1992 Equation........** | **2.39584** | **0.00001513** | **-0.0300318** | **0.353099** | **(Eqn. 2.1)** |
| 1993 Equation......... | 0.77455 | 0.00001349 | -0.0092041 | 0.425785 | |
| 1987-1993 Eqn....... | 0.770725 | 0.00001364 | -0.0080933 | 0.379579 | |

**<u>Where</u>**: $a_0$ is the constant for predictive equation 2
$a_1$ is the coefficient for predictive equation 2 ($CL_9$)
$a_2$ is the coefficient for predictive equation 2 ($CH_9$)

"1992 Equation" means the regression equation based on fitting 1992 data to regression equation 2.

RMSE for 1994 Predictions: is the calculated root-mean-square-error of tomato yield predictions for 1994 using that particular year's regression equation.

For example, using the 1992 Equation of:
$Y = 2.39584 + 0.00001513\ (CL_9) - 0.030018(CH_9)$ to predict 1994 yields (with 1994 data), the root-mean-square-error of the predictions was found to be: 0.353099

Table A-3

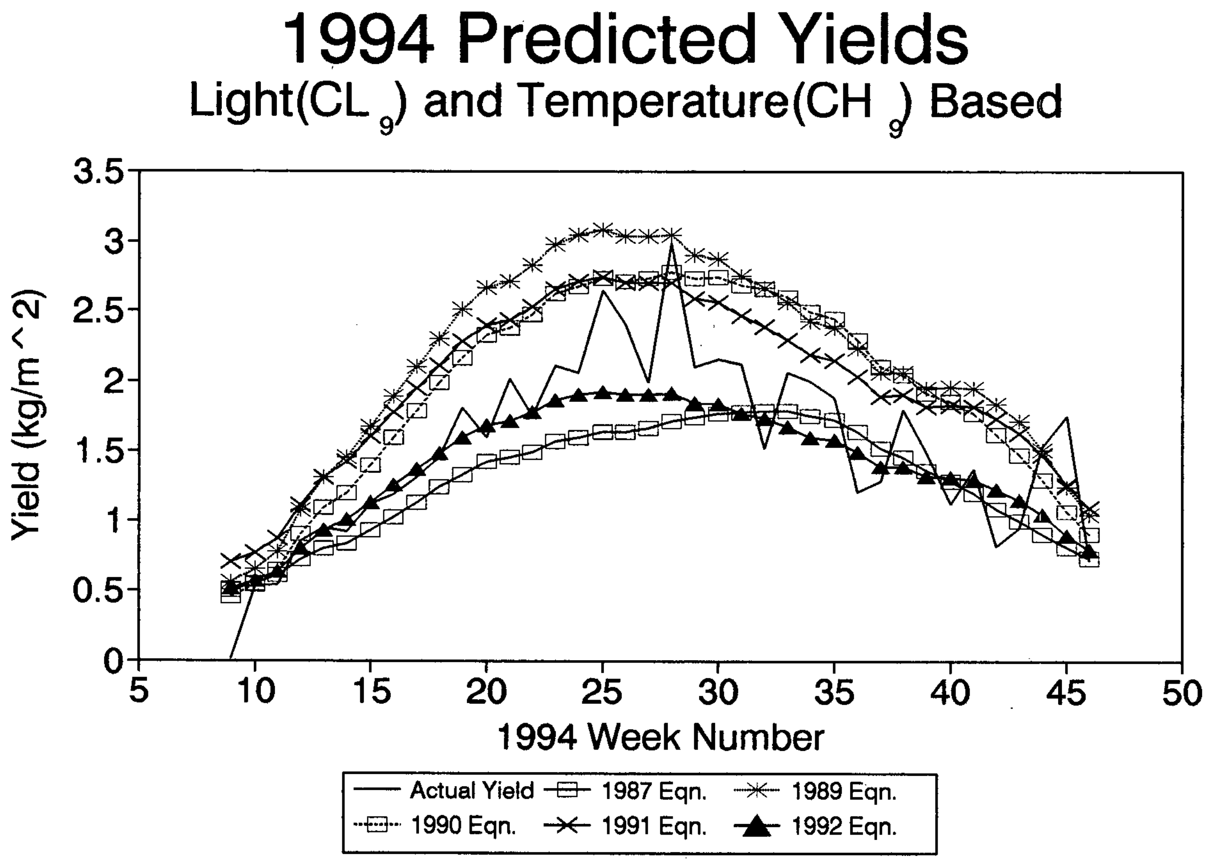



Figure A-20 Predicted yields for 1994 production based on regression equations derived from light & heat data from 1987 to 1992.

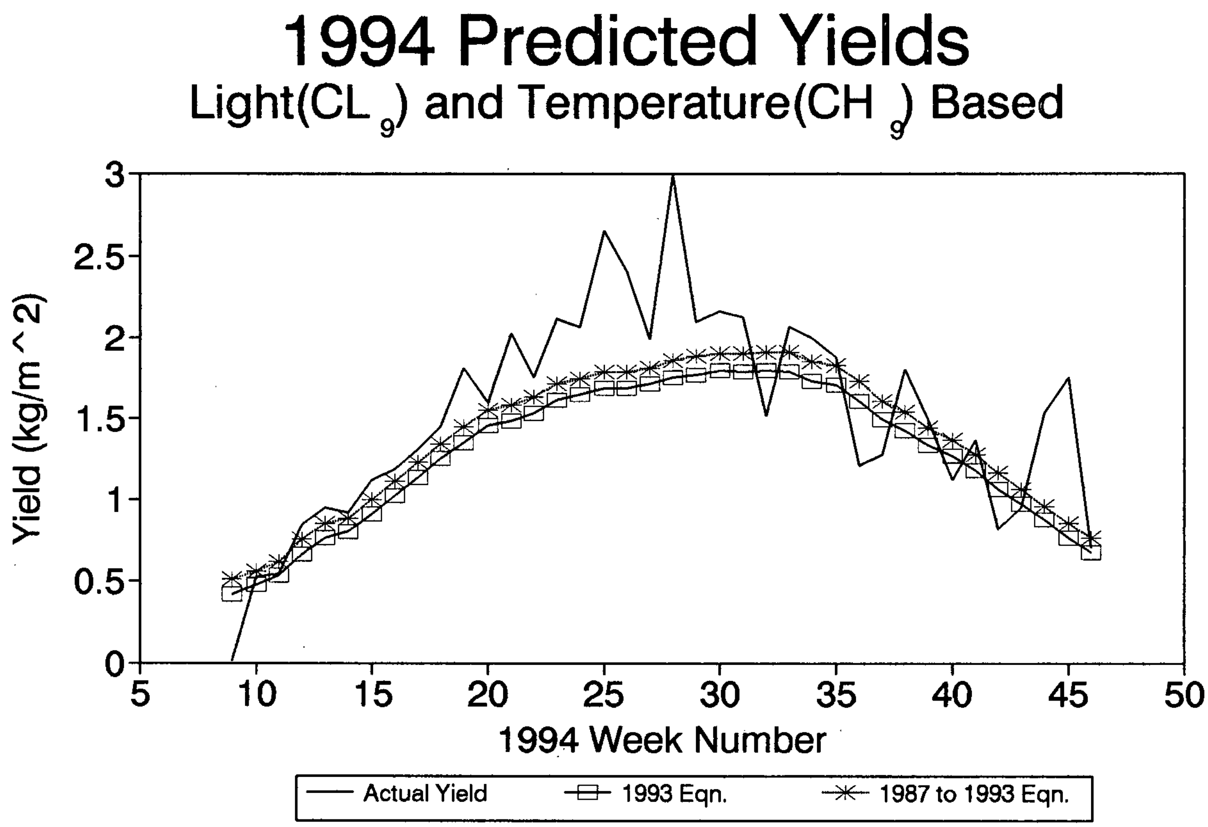



Figure A-21 Predicted yields for 1994 production based on regression equations derived from light & heat data (1993 and '87 - 93)

## 1987 STEPWISE REGRESSION RESULTS

***Year:*** *1987*
***Data Set:*** D-RYALL (MR87.SYS from Systat)

***Variables:*** VAR(3) = Yield in kg/m$^2$
VAR(4) = 9 week cumulative light readings in W/m$^2$
VAR(5) = 9 week cumulative heat units, degrees celsius
(each 9 wk. period = sum of 9 wk. avg temp *C - 90)

***Statistical Method:***
Stepwise Regression for light VAR(4) and heat units VAR(5).
All results based on dependent variable VAR(3), yield in kg/m$^2$

---

-> DEPENDENT VARIABLE VAR(3), Yield in kg/m$^2$ <-

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

**STEP # 0** R= .000 RSQUARE= .000

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 2 VAR(4) | 0.75591487 | . | . | .1E+01 | .45E+02 | 0.0000 |
| 3 VAR(5) | 0.59594473 | . | . | .1E+01 | .19E+02 | 0.0001 |

---

**STEP # 1** R= .756 RSQUARE= .571
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001021 | 0.00000152 | .076E+01 | .1E+01 | .45E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 3 VAR(5) | -0.13153800 | . | . | 0.27982 | 0.58103 | 0.4513 |

---

## 1987 STEPWISE REGRESSION RESULTS

STEP # 2 R= .761 RSQUARE= .579
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001207 | 0.00000288 | .089E+01 | 0.27982 | .18E+02 | 0.0002 |
| 3 VAR(5) | -0.00434545 | -0.00570082 | -.016E+01 | 0.27982 | 0.58103 | 0.4513 |

OUT PART. CORR

---

none

---

THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:

CONSTANT
VAR(4)
VAR(5)

---

DEP VAR: VAR(3) N: 36 MULTIPLE R: 0.761 SQUARED MULTIPLE R: 0.579
ADJUSTED SQUARED MULTIPLE R: .553 STANDARD ERROR OF ESTIMATE: 0.28657972

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 0.52199056 | 0.30523030 | 0.00000000 | . | 1.71015 | 0.09663 |
| VAR(4) | 0.00001207 | 0.00000288 | 0.89406748 | 0.2798160 | 4.18631 | 0.00020 |
| VAR(5) | -0.00434545 | 0.00570082 | -0.16279360 | 0.2798160 | -0.76225 | 0.45132 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 3.72465308 | 2 | 1.86232654 | 22.67592009 | 0.00000064 |
| RESIDUAL | 2.71022192 | 33 | 0.08212794 | | |

---

DEPENDENT VARIABLE VAR(3)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

# 1989 STEPWISE REGRESSION RESULTS

***Year:*** ***1989***
***Data Set:*** D-RYALL (MR89.SYS from Systat)

***Variables:*** VAR(3) = Yield in kg/m$^2$
VAR(4) = 9 week cumulative light readings in W/m$^2$
VAR(5) = 9 week cumulative heat units, degrees celsius
(each 9 wk. period = sum of 9 wk. avg temp *C - 90)

***Statistical Method:***
Stepwise Regression for light VAR(4) and heat units VAR(5).
All results based on dependent variable VAR(3), yield in kg/m$^2$

---

-> DEPENDENT VARIABLE VAR(3), Yield in kg/m$^2$ <-

DEPENDENT VARIABLE VAR(3)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

**STEP # 0** R= .000 RSQUARE= .000

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 2 VAR(4) | 0.76284622 | . | . | .1E+01 | .52E+02 | 0.0000 |
| 3 VAR(5) | 0.43172526 | . | . | .1E+01 | 8.47615 | 0.0061 |

---

**STEP # 1** R= .763 RSQUARE= .582
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001285 | 0.00000179 | .076E+01 | .1E+01 | .52E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 3 VAR(5) | -0.73518613 | . | . | 0.24047 | .42E+02 | 0.0000 |

---

## 1989 STEPWISE REGRESSION RESULTS

*STEP # 2* R= .899 RSQUARE= .808
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00002708 | 0.00000251 | .161E+01 | 0.24047 | .12E+03 | 0.0000 |
| 3 VAR(5) | -0.05799486 | -0.00891219 | -.097E+01 | 0.24047 | .42E+02 | 0.0000 |

OUT PART. CORR
---
none

---

***THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:***

CONSTANT
VAR(4)
VAR(5)

---

DEP VAR: VAR(3) N: 39 MULTIPLE R: 0.899 SQUARED MULTIPLE R: 0.808
ADJUSTED SQUARED MULTIPLE R: .797 STANDARD ERROR OF ESTIMATE: 0.22600488

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 4.25232655 | 0.64683641 | 0.00000000 | . | 6.57404 | 0.00000 |
| VAR(4) | 0.00002708 | 0.00000251 | 1.60766401 | 0.2404687 | .11E+02 | 0.00000 |
| VAR(5) | -0.05799486 | 0.00891219 | -0.96937144 | 0.2404687 | -6.50736 | 0.00000 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 7.73327686 | 2 | 3.86663843 | 75.70035554 | 0.00000000 |
| RESIDUAL | 1.83881545 | 36 | 0.05107821 | | |

---

# 1990 STEPWISE REGRESSION RESULTS

***Year:*** *1990*
***Data Set:*** D-RYALL (MR90.SYS from Systat)

***Variables:*** VAR(3) = Yield in kg/m$^2$
VAR(4) = 9 week cumulative light readings in W/m$^2$
VAR(5) = 9 week cumulative heat units, degrees celsius
(each 9 wk. period = sum of 9 wk. avg temp *C - 90)

***Statistical Method:***
Stepwise Regression for light VAR(4) and heat units VAR(5).
All results based on dependent variable VAR(3), yield in kg/m$^2$

---

**-> DEPENDENT VARIABLE VAR(3), Yield in kg/m$^2$ <-**

DEPENDENT VARIABLE VAR(3)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

**STEP # 0** R= .000 RSQUARE= .000

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 2 VAR(4) | 0.67698642 | . | . | .1E+01 | .31E+02 | 0.0000 |
| 3 VAR(5) | 0.53005288 | . | . | .1E+01 | .14E+02 | 0.0005 |

---

**STEP # 1** R= .677 RSQUARE= .458
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001443 | 0.00000258 | .068E+01 | .1E+01 | .31E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 3 VAR(5) | -0.28777015 | . | . | 0.16902 | 3.25039 | 0.0798 |

---

## 1990 STEPWISE REGRESSION RESULTS

*STEP #* 2 R= .709 RSQUARE= .503
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00002444 | 0.00000609 | .115E+01 | 0.16902 | .16E+02 | 0.0003 |
| 3 VAR(5) | -0.03380935 | -0.01875293 | -.052E+01 | 0.16902 | 3.25039 | 0.0798 |

OUT PART. CORR
---
none

---

THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:

CONSTANT
VAR(4)
VAR(5)

---

DEP VAR: VAR(3) N: 39 MULTIPLE R: 0.709 SQUARED MULTIPLE R: 0.503
ADJUSTED SQUARED MULTIPLE R: .476 STANDARD ERROR OF ESTIMATE: 0.43162224

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 2.38053976 | 1.27331469 | 0.00000000 | . | 1.86956 | 0.06970 |
| VAR(4) | 0.00002444 | 0.00000609 | 1.14660486 | 0.1690213 | 4.01265 | 0.00029 |
| VAR(5) | -0.03380935 | 0.01875293 | -0.51516956 | 0.1690213 | -1.80288 | 0.07978 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 6.79227052 | 2 | 3.39613526 | 18.22960904 | 0.00000340 |
| RESIDUAL | 6.70671922 | 36 | 0.18629776 | | |

# 1991 STEPWISE REGRESSION RESULTS

*Year:* *1991*
*Data Set:* D-RYALL (MR91.SYS from Systat)

*Variables:* VAR(3) = Yield in kg/m$^2$
VAR(4) = 9 week cumulative light readings in W/m$^2$
VAR(5) = 9 week cumulative heat units, degrees celsius
(each 9 wk. period = sum of 9 wk. avg temp *C - 90)

*Statistical Method:*
Stepwise Regression for light VAR(4) and heat units VAR(5).
All results based on dependent variable VAR(3), yield in kg/m$^2$

---

**-> DEPENDENT VARIABLE VAR(3), Yield in kg/m$^2$ <-**

DEPENDENT VARIABLE VAR(3)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

**STEP # 0** R= .000 RSQUARE= .000

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 2 VAR(4) | 0.65059951 | . | . | .1E+01 | .26E+02 | 0.0000 |
| 3 VAR(5) | 0.26286050 | . | . | .1E+01 | 2.67207 | 0.1108 |

---

**STEP # 1** R= .651 RSQUARE= .423
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001201 | 0.00000234 | .065E+01 | .1E+01 | .26E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 3 VAR(5) | -0.54492557 | . | . | 0.37242 | .15E+02 | 0.0005 |

---

## 1991 STEPWISE REGRESSION RESULTS

**STEP # 2** R= .771 RSQUARE= .595
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00002193 | 0.00000326 | .119E+01 | 0.37242 | .45E+02 | 0.0000 |
| 3 VAR(5) | -0.04803047 | -0.01249225 | -.068E+01 | 0.37242 | .15E+02 | 0.0005 |

OUT PART. CORR

---

none

---

**THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:**

CONSTANT
VAR(4)
VAR(5)

---

DEP VAR: VAR(3) N: 38 MULTIPLE R: 0.771 SQUARED MULTIPLE R: 0.595
ADJUSTED SQUARED MULTIPLE R: .571 STANDARD ERROR OF ESTIMATE: 0.31662425

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 3.76864580 | 0.92989756 | 0.00000000 | . | 4.05275 | 0.00027 |
| VAR(4) | 0.00002193 | 0.00000326 | 1.18780098 | 0.3724207 | 6.73468 | 0.00000 |
| VAR(5) | -0.04803047 | 0.01249225 | -0.67811428 | 0.3724207 | -3.84482 | 0.00049 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 5.14490480 | 2 | 2.57245240 | 25.66013879 | 0.00000014 |
| RESIDUAL | 3.50878204 | 35 | 0.10025092 | | |

---

# 1992 STEPWISE REGRESSION RESULTS

***Year:*** *1992*
***Data Set:*** D-RYALL (MR92.SYS from Systat)

***Variables:*** VAR(3) = Yield in kg/$m^2$
VAR(4) = 9 week cumulative light readings in W/$m^2$
VAR(5) = 9 week cumulative heat units, degrees celsius
(each 9 wk. period = sum of 9 wk. avg temp *C - 90)

***Statistical Method:***
Stepwise Regression for light VAR(4) and heat units VAR(5).
All results based on dependent variable VAR(3), yield in kg/$m^2$

---

**-> DEPENDENT VARIABLE VAR(3), Yield in kg/$m^2$ <-**

DEPENDENT VARIABLE VAR(3)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

**STEP # 0** R= .000 RSQUARE= .000

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| 2 VAR(4) | 0.78102804 | . | . | .1E+01 | .56E+02 | 0.0000 |
| 3 VAR(5) | 0.24309995 | . | . | .1E+01 | 2.26114 | 0.1414 |

---

**STEP # 1** R= .781 RSQUARE= .610
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001186 | 0.00000158 | .078E+01 | .1E+01 | .56E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| 3 VAR(5) | -0.45593461 | . | . | 0.63726 | 9.18502 | 0.0046 |

---

## 1992 STEPWISE REGRESSION RESULTS

**STEP # 2** R= .831 RSQUARE= .691
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001513 | 0.00000179 | .100E+01 | 0.63726 | .72E+02 | 0.0000 |
| 3 VAR(5) | -0.03003183 | -0.00990927 | -.036E+01 | 0.63726 | 9.18502 | 0.0046 |

OUT PART. CORR
---
none

---

**THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:**

CONSTANT
VAR(4)
VAR(5)

---

DEP VAR: VAR(3) N: 38 MULTIPLE R: 0.831 SQUARED MULTIPLE R: 0.691
ADJUSTED SQUARED MULTIPLE R: .673 STANDARD ERROR OF ESTIMATE: 0.29923584

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 2.39584015 | 0.73325947 | 0.00000000 | . | 3.26738 | 0.00244 |
| VAR(4) | 0.00001513 | 0.00000179 | 0.99584564 | 0.6372618 | 8.46172 | 0.00000 |
| VAR(5) | -0.03003183 | 0.00990927 | -0.35667547 | 0.6372618 | -3.03068 | 0.00457 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 7.01081632 | 2 | 3.50540816 | 39.14816096 | 0.00000000 |
| RESIDUAL | 3.13397316 | 35 | 0.08954209 | | |

---

# 1993 STEPWISE REGRESSION RESULTS

*Year:* *1993*
*Data Set:* D-RYALL (MR93.SYS from Systat)

*Variables:* VAR(3) = Yield in kg/m$^2$
VAR(4) = 9 week cumulative light readings in W/m$^2$
VAR(5) = 9 week cumulative heat units, degrees celsius
(each 9 wk. period = sum of 9 wk. avg temp *C - 90)

*Statistical Method:*
Stepwise Regression for light VAR(4) and heat units VAR(5).
All results based on dependent variable VAR(3), yield in kg/m$^2$

---

**-> DEPENDENT VARIABLE VAR(3), Yield in kg/m$^2$ <-**

DEPENDENT VARIABLE VAR(3)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

**STEP # 0** R= .000 RSQUARE= .000

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 2 VAR(4) | 0.67378618 | . | . | .1E+01 | .29E+02 | 0.0000 |
| 3 VAR(5) | 0.31215946 | . | . | .1E+01 | 3.77874 | 0.0600 |

---

**STEP # 1** R= .674 RSQUARE= .454
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001149 | 0.00000213 | .067E+01 | .1E+01 | .29E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 3 VAR(5) | -0.19585125 | . | . | 0.60300 | 1.35618 | 0.2523 |

---

## 1993 . STEPWISE REGRESSION RESULTS

STEP # 2 R= .689 RSQUARE= .475
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001349 | 0.00000273 | .079E+01 | 0.60300 | .24E+02 | 0.0000 |
| 3 VAR(5) | -0.00920407 | -0.00790353 | -.019E+01 | 0.60300 | 1.35618 | 0.2523 |

OUT PART. CORR
---
none

---

**THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:**

CONSTANT
VAR(4)
VAR(5)

---

DEP VAR: VAR(3) N: 37 MULTIPLE R: 0.689 SQUARED MULTIPLE R: 0.475
ADJUSTED SQUARED MULTIPLE R: .444 STANDARD ERROR OF ESTIMATE: 0.32356666

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 0.77455001 | 0.54026852 | 0.00000000 | . | 1.43364 | 0.16081 |
| VAR(4) | 0.00001349 | 0.00000273 | 0.79121212 | 0.6030001 | 4.94405 | 0.00002 |
| VAR(5) | -0.00920407 | 0.00790353 | -0.18636692 | 0.6030001 | -1.16455 | 0.25231 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 3.21974609 | 2 | 1.60987305 | 15.37673359 | 0.00001753 |
| RESIDUAL | 3.55964310 | 34 | 0.10469539 | | |

## 1987 - 1993 STEPWISE REGRESSION RESULTS

***Year:*** *1987 - 1993*
***Data Set:*** D-RYALL (MR8793.SYS from Systat)

***Variables:*** VAR(3) = Yield in kg/m$^2$
VAR(4) = 9 week cumulative light readings in W/m$^2$
VAR(5) = 9 week cumulative heat units, degrees celsius
(each 9 wk. period = sum of 9 wk. avg temp *C - 90)

***Statistical Method:***
Stepwise Regression for light VAR(4) and heat units VAR(5).
All results based on dependent variable VAR(3), yield in kg/m$^2$

---

**-> DEPENDENT VARIABLE VAR(3), Yield in kg/m$^2$ <-**

DEPENDENT VARIABLE VAR(3)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 4 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 6

**STEP # 0** R= .000 RSQUARE= .000

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 2 VAR(4) | 0.71031954 | . | . | .1E+01 | .23E+03 | 0.0000 |
| 3 VAR(5) | 0.34628936 | . | . | .1E+01 | .31E+02 | 0.0000 |

---

**STEP # 1** R= .710 RSQUARE= .505
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001190 | 0.00000079 | .071E+01 | .1E+01 | .23E+03 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 3 VAR(5) | -0.18303353 | . | . | 0.60489 | 7.76440 | 0.0058 |

---

## 1987 - 1993 STEPWISE REGRESSION RESULTS

STEP # 2 R= .722 RSQUARE= .521
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(4) | 0.00001364 | 0.00000100 | .081E+01 | 0.60489 | .19E+03 | 0.0000 |
| 3 VAR(5) | -0.00809330 | -0.00290450 | -.017E+01 | 0.60489 | 7.76440 | 0.0058 |

OUT PART. CORR
---
none

---

**THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:**

CONSTANT
VAR(4)
VAR(5)

---

DEP VAR: VAR(3) N: 227 MULTIPLE R: 0.722 SQUARED MULTIPLE R: 0.521
ADJUSTED SQUARED MULTIPLE R: .517 STANDARD ERROR OF ESTIMATE: 0.34617793

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 0.77072466 | 0.21153971 | 0.00000000 | . | 3.64340 | 0.00033 |
| VAR(4) | 0.00001364 | 0.00000100 | 0.81444264 | 0.6048929 | .14E+02 | 0.00000 |
| VAR(5) | -0.00809330 | 0.00290450 | -0.16564933 | 0.6048929 | -2.78647 | 0.00579 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 29.21550328 | 2 | 14.60775164 | .121895E+03 | 0.00000000 |
| RESIDUAL | 26.84397160 | 224 | 0.11983916 | | |

EMPIRICAL MATH MODEL

STATISTICAL RESULTS FOR EQUATIONS 3.1 and 4.1

**Light & Heat (3,6,9 week) Combined Step-Wise Regression**
**1992 Equation for Yield Prediction**

**Definition of Variables:**

**VAR(10) = Yield (kg/m^2)**
**VAR( 5) = Cumulative Light for 3 weeks (W/m^2)**
**VAR( 6) = Cumulative Light for 6 weeks (W/m^2)**
**VAR( 4) = Cumulative Light for 9 weeks (W/m^2)**
**VAR(16) = Cumulative Heat for 3 weeks (*C)**
**VAR(15) = Cumulative Heat for 6 weeks (*C)**
**VAR(14) = Cumulative Heat for 9 weeks (*C)**

## EQUATION LAYOUT:

**VAR(10) = CONSTANT + $a_1$VAR(5) + $a_2$VAR(6) + $a_3$VAR(4) + $a_4$VAR(16) + $a_5$VAR(15) + $a_6$VAR(14)**

Where 'CONSTANT' and $a_1$ to $a_6$ are to be determined.

**Step-Wise Regression Results (Based on 1992 Data)**

DEPENDENT VARIABLE VAR(10) = Four Week Ahead Actual Yield

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

FORWARD STEPWISE WITH FIRST 8 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS= 14

**STEP # 0 R= .000 RSQUARE= .000**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| OUT | PART. CORR | | | | | |
| 2 VAR(5) | 0.83258920 | . | . | .1E+01 | .88E+02 | 0.0000 |
| 3 VAR(6) | 0.78185025 | . | . | .1E+01 | .61E+02 | 0.0000 |
| 4 VAR(4) | 0.68540368 | . | . | .1E+01 | .35E+02 | 0.0000 |
| 5 VAR(16) | 0.60663037 | . | . | .1E+01 | .23E+02 | 0.0000 |
| 6 VAR(15) | 0.56819254 | . | . | .1E+01 | .19E+02 | 0.0001 |
| 7 VAR(14) | 0.53326428 | . | . | .1E+01 | .15E+02 | 0.0003 |

**STEP # 1 R= .833 RSQUARE= .693**
**TERM ENTERED: VAR(5)**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(5) | 0.00003932 | 0.00000419 | .083E+01 | .1E+01 | .88E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |

---

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 VAR(6) | -0.07246533 | . | . | 0.09069 | 0.20060 | 0.6568 |
| 4 VAR(4) | -0.11529719 | . | . | 0.25680 | 0.51196 | 0.4787 |
| 5 VAR(16) | 0.11202455 | . | . | 0.54637 | 0.48294 | 0.4913 |
| 6 VAR(15) | 0.10239134 | . | . | 0.60371 | 0.40261 | 0.5295 |
| 7 VAR(14) | 0.06869768 | . | . | 0.63510 | 0.18019 | 0.6736 |

**STEP # 2 R= .834 RSQUARE= .695**
**TERM ENTERED: VAR(6)**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(5) | 0.00004532 | 0.00001405 | .096E+01 | 0.09069 | .10E+02 | 0.0026 |
| 3 VAR(6) | -0.00000332 | -0.00000742 | -.013E+01 | 0.09069 | 0.20060 | 0.6568 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 4 VAR(4) | -0.12780013 | . | . | 0.03675 | 0.61435 | 0.4381 |
| 5 VAR(16) | 0.12679220 | . | . | 0.53004 | 0.60454 | 0.4418 |
| 6 VAR(15) | 0.12233083 | . | . | 0.57265 | 0.56211 | 0.4582 |
| 7 VAR(14) | 0.09128331 | . | . | 0.59089 | 0.31090 | 0.5805 |

**STEP # 3 R= .837 RSQUARE= .700**
**TERM ENTERED: VAR(4)**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(5) | 0.00003448 | 0.00001977 | .073E+01 | 0.04630 | 3.04195 | 0.0894 |
| 3 VAR(6) | 0.00001098 | 0.00001972 | .044E+01 | 0.01298 | 0.31025 | 0.5809 |
| 4 VAR(4) | -0.00109713 | -0.00139975 | -.037E+01 | 0.03675 | 0.61435 | 0.4381 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 5 VAR(16) | 0.12114063 | . | . | 0.52853 | 0.53617 | 0.4688 |
| 6 VAR(15) | 0.11464885 | . | . | 0.56987 | 0.47950 | 0.4931 |
| 7 VAR(14) | 0.09119794 | . | . | 0.59087 | 0.30193 | 0.5861 |

**STEP # 4 R= .839 RSQUARE= .704**
**TERM ENTERED: VAR(16)**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(5) | 0.00003387 | 0.00001991 | .072E+01 | 0.04622 | 2.89251 | 0.0976 |
| 3 VAR(6) | 0.00000930 | 0.00001998 | .037E+01 | 0.01281 | 0.21668 | 0.6444 |
| 4 VAR(4) | -0.00104205 | -0.00141062 | -.035E+01 | 0.03664 | 0.54570 | 0.4649 |
| 5 VAR(16) | 0.00831282 | 0.01135264 | .09129784 | 0.52853 | 0.53617 | 0.4688 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 6 VAR(15) | -0.00178971 | . | . | 0.05454 | 0.00011 | 0.9916 |
| 7 VAR(14) | -0.04931388 | . | . | 0.09570 | 0.08532 | 0.7719 |

**STEP # 5 R= .839 RSQUARE= .704**
**TERM ENTERED: VAR(15)**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(5) | 0.00003382 | 0.00002069 | .072E+01 | 0.04405 | 2.67310 | 0.1110 |
| 3 VAR(6) | 0.00000933 | 0.00002045 | .037E+01 | 0.01257 | 0.20809 | 0.6511 |
| 4 VAR(4) | -0.00104297 | -0.00143331 | -.035E+01 | 0.03650 | 0.52950 | 0.4717 |
| 5 VAR(16) | 0.00868756 | 0.03721875 | .09541359 | 0.05058 | 0.05448 | 0.8168 |
| 6 VAR(15) | -0.00022289 | -0.02105054 | .00416809 | 0.05454 | 0.00011 | 0.9916 |

OUT PART. CORR

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | VAR(6) | -0.07246533 | . | . | 0.09069 | 0.20060 | 0.6568 |
| 4 | VAR(4) | -0.11529719 | . | . | 0.25680 | 0.51196 | 0.4787 |
| 5 | VAR(16) | 0.11202455 | . | . | 0.54637 | 0.48294 | 0.4913 |
| 6 | VAR(15) | 0.10239134 | . | . | 0.60371 | 0.40261 | 0.5295 |
| 7 | VAR(14) | 0.06869768 | . | . | 0.63510 | 0.18019 | 0.6736 |

**STEP # 2 R= .834 RSQUARE= .695**
**TERM ENTERED: VAR(6)**

| | VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|---|
| IN | | | | | | | |
| 1 | CONSTANT | | | | | | |
| 2 | VAR(5) | 0.00004532 | 0.00001405 | .096E+01 | 0.09069 | .10E+02 | 0.0026 |
| 3 | VAR(6) | -0.00000332 | -0.00000742 | -.013E+01 | 0.09069 | 0.20060 | 0.6568 |
| OUT | | PART. CORR | | | | | |
| 4 | VAR(4) | -0.12780013 | . | . | 0.03675 | 0.61435 | 0.4381 |
| 5 | VAR(16) | 0.12679220 | . | . | 0.53004 | 0.60454 | 0.4418 |
| 6 | VAR(15) | 0.12233083 | . | . | 0.57265 | 0.56211 | 0.4582 |
| 7 | VAR(14) | 0.09128331 | . | . | 0.59089 | 0.31090 | 0.5805 |

**STEP # 3 R= .837 RSQUARE= .700**
**TERM ENTERED: VAR(4)**

| | VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|---|
| IN | | | | | | | |
| 1 | CONSTANT | | | | | | |
| 2 | VAR(5) | 0.00003448 | 0.00001977 | .073E+01 | 0.04630 | 3.04195 | 0.0894 |
| 3 | VAR(6) | 0.00001098 | 0.00001972 | .044E+01 | 0.01298 | 0.31025 | 0.5809 |
| 4 | VAR(4) | -0.00109713 | -0.00139975 | -.037E+01 | 0.03675 | 0.61435 | 0.4381 |
| OUT | | PART. CORR | | | | | |
| 5 | VAR(16) | 0.12114063 | . | . | 0.52853 | 0.53617 | 0.4688 |
| 6 | VAR(15) | 0.11464885 | . | . | 0.56987 | 0.47950 | 0.4931 |
| 7 | VAR(14) | 0.09119794 | . | . | 0.59087 | 0.30193 | 0.5861 |

**STEP # 4 R= .839 RSQUARE= .704**
**TERM ENTERED: VAR(16)**

| | VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|---|
| IN | | | | | | | |
| 1 | CONSTANT | | | | | | |
| 2 | VAR(5) | 0.00003387 | 0.00001991 | .072E+01 | 0.04622 | 2.89251 | 0.0976 |
| 3 | VAR(6) | 0.00000930 | 0.00001998 | .037E+01 | 0.01281 | 0.21668 | 0.6444 |
| 4 | VAR(4) | -0.00104205 | -0.00141062 | -.035E+01 | 0.03664 | 0.54570 | 0.4649 |
| 5 | VAR(16) | 0.00831282 | 0.01135264 | .09129784 | 0.52853 | 0.53617 | 0.4688 |
| OUT | | PART. CORR | | | | | |
| 6 | VAR(15) | -0.00178971 | . | . | 0.05454 | 0.00011 | 0.9916 |
| 7 | VAR(14) | -0.04931388 | . | . | 0.09570 | 0.08532 | 0.7719 |

**STEP # 5 R= .839 RSQUARE= .704**
**TERM ENTERED: VAR(15)**

| | VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|---|
| IN | | | | | | | |
| 1 | CONSTANT | | | | | | |
| 2 | VAR(5) | 0.00003382 | 0.00002069 | .072E+01 | 0.04405 | 2.67310 | 0.1110 |
| 3 | VAR(6) | 0.00000933 | 0.00002045 | .037E+01 | 0.01257 | 0.20809 | 0.6511 |
| 4 | VAR(4) | -0.00104297 | -0.00143331 | -.035E+01 | 0.03650 | 0.52950 | 0.4717 |
| 5 | VAR(16) | 0.00868756 | 0.03721875 | .09541359 | 0.05058 | 0.05448 | 0.8168 |
| 6 | VAR(15) | -0.00022289 | -0.02105054 | .00416809 | 0.05454 | 0.00011 | 0.9916 |
| OUT | | PART. CORR | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 7 VAR(14) | -0.11478835 | . | . | 0.01651 | 0.45398 | 0.5050 |

**STEP # 6 R= .841 RSQUARE= .708**
**TERM ENTERED: VAR(14)**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(5) | 0.00003552 | 0.00002100 | .075E+01 | 0.04342 | 2.86024 | 0.0999 |
| 3 VAR(6) | 0.00000548 | 0.00002139 | .022E+01 | 0.01167 | 0.06558 | 0.7994 |
| 4 VAR(4) | -0.00066667 | -0.00154883 | -.022E+01 | 0.03176 | 0.18527 | 0.6696 |
| 5 VAR(16) | -0.00112719 | -0.04024149 | .01237968 | 0.04395 | 0.00078 | 0.9778 |
| 6 VAR(15) | 0.03107864 | 0.05107220 | .058E+01 | 0.00941 | 0.37030 | 0.5469 |
| 7 VAR(14) | -0.02212007 | -0.03282987 | -.049E+01 | 0.01651 | 0.45398 | 0.5050 |

| OUT | PART. CORR |
|---|---|
| none | |

**DEP VAR: VAR(10) N: 41 MULTIPLE R: 0.841 SQUARED MULTIPLE R: 0.708**
**ADJUSTED SQUARED MULTIPLE R: .657 STANDARD ERROR OF ESTIMATE: 0.38140815**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 0.34917888 | 0.62100805 | 0.00000000 | . | 0.56228 | 0.57762 |
| VAR(5) | 0.00003552 | 0.00002100 | 0.75201265 | 0.0434212 | 1.69123 | 0.09994 |
| VAR(6) | 0.00000548 | 0.00002139 | 0.21962994 | 0.0116716 | 0.25608 | 0.79943 |
| VAR(4) | -0.00066667 | 0.00154883 | -0.22379827 | 0.0317579 | -0.43044 | 0.66960 |
| VAR(16) | -0.00112719 | 0.04024149 | -0.01237968 | 0.0439517 | -0.02801 | 0.97782 |
| VAR(15) | 0.03107864 | 0.05107220 | 0.58119010 | 0.0094117 | 0.60852 | 0.54688 |
| VAR(14) | -0.02212007 | 0.03282987 | -0.48579463 | 0.0165150 | -0.67378 | 0.50501 |

**ANALYSIS OF VARIANCE**

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 11.99853611 | 6 | 1.99975602 | 13.74665598 | 0.00000007 |
| RESIDUAL | 4.94605413 | 34 | 0.14547218 | | |

DURBIN-WATSON D STATISTIC 1.510
FIRST ORDER AUTOCORRELATION .240

## Final Regression Equation:

**YIELD = 0.3491788 + 0.00003552 ($CL_3$) + 0.00000548 ($CL_6$) - 0.00066667 ($CL_9$) - 0.00112719 ($CH_3$) + 0.03107864 ($CH_6$) - 0.02212007 ($CH_9$)**

Where $CL_a$ and $CH_a$ are the Cumulative Light and Cumulative Heat for 'a' weeks respectively.

EMPIRICAL MATH MODEL

STATISTICAL RESULTS FOR 'ALTERNATIVE REGRESSION METHOD'

**Light & Heat (3,6,9 week) Combined Step-Wise Regression**
**1992 Equation for Yield Prediction (4 weeks ahead yield)**

---

**Definition of Variables:**

**VAR(9) = 4 weeks ahead Yield (kg/m^2)**
**VAR(3) = Cumulative Light for 3 weeks (W/m^2)**
**VAR(4) = Cumulative Light for 6 weeks (W/m^2)**
**VAR(5) = Cumulative Light for 9 weeks (W/m^2)**
**VAR(6) = Cumulative Heat for 3 weeks (*C)**
**VAR(7) = Cumulative Heat for 6 weeks (*C)**
**VAR(8) = Cumulative Heat for 9 weeks (*C)**

**EQUATION LAYOUT:**

$$VAR(9) = CONSTANT + a_1 VAR(3) + a_2 VAR(4) + a_3 VAR(5) + a_4 VAR(6) + a_5 VAR(7) + a_6 VAR(8)$$

Where 'CONSTANT' and $a_1$ to $a_6$ are to be determined.

---

**Step-Wise Regression Results (Based on 1992 Data)**

```
DEPENDENT VARIABLE      VAR(9)

MINIMUM TOLERANCE FOR ENTRY INTO MODEL =   .010000

FORWARD STEPWISE WITH FIRST  9 VARIABLES FORCED IN MODEL
ALPHA-TO-ENTER= .150 AND ALPHA-TO-REMOVE= .150 MAX # STEPS=     14

STEP #     0 R=  .000 RSQUARE=   .000

VARIABLE            COEFFICIENT    STD ERROR  STD COEF TOLERANCE    F     'P'

IN
---
  1 CONSTANT

OUT             PART. CORR
---
  2 VAR(3)          0.81997755   .           .     .1E+01  .78E+02  0.0000
  3 VAR(4)          0.71710795   .           .     .1E+01  .40E+02  0.0000
  4 VAR(5)          0.58652376   .           .     .1E+01  .20E+02  0.0001
  5 VAR(6)          0.55901034   .           .     .1E+01  .17E+02  0.0002
  6 VAR(7)          0.50429635   .           .     .1E+01  .13E+02  0.0009
  7 VAR(8)          0.45908945   .           .     .1E+01  .10E+02  0.0029

STEP #     1 R=  .820 RSQUARE=   .672
TERM ENTERED: VAR(3)

VARIABLE            COEFFICIENT    STD ERROR  STD COEF TOLERANCE    F     'P'

IN
---
  1 CONSTANT
  2 VAR(3)          0.00003977  0.00000450  .082E+01  .1E+01  .78E+02  0.0000
```

| OUT | PART. CORR | | | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| 3 VAR(4) | -0.34723606 | . | . | 0.09725 | 5.07284 | 0.0303 |
| 4 VAR(5) | -0.37278569 | . | . | 0.27466 | 5.97175 | 0.0194 |
| 5 VAR(6) | 0.03765418 | . | . | 0.56170 | 0.05253 | 0.8200 |
| 6 VAR(7) | -0.04839489 | . | . | 0.58919 | 0.08686 | 0.7699 |
| 7 VAR(8) | -0.08879462 | . | . | 0.62907 | 0.29404 | 0.5909 |

STEP # 2 R= .844 RSQUARE= .712
TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | 0.00006913 | 0.00001372 | .143E+01 | 0.09725 | .25E+02 | 0.0000 |
| 3 VAR(4) | -0.00001641 | -0.00000729 | -.064E+01 | 0.09725 | 5.07284 | 0.0303 |
| OUT | PART. CORR | | | | | |
| 4 VAR(5) | -0.14479167 | . | . | 0.03967 | 0.77089 | 0.3858 |
| 5 VAR(6) | 0.09267208 | . | . | 0.55079 | 0.31185 | 0.5800 |
| 6 VAR(7) | 0.02907883 | . | . | 0.56169 | 0.03047 | 0.8624 |
| 7 VAR(8) | 0.00493994 | . | . | 0.58369 | 0.00088 | 0.9765 |

STEP # 3 R= .847 RSQUARE= .718
TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | 0.00005730 | 0.00001927 | .118E+01 | 0.04965 | 8.84633 | 0.0052 |
| 3 VAR(4) | -0.00000079 | -0.00001923 | .03076607 | 0.01405 | 0.00170 | 0.9674 |
| 4 VAR(5) | -0.00000705 | -0.00000803 | -.039E+01 | 0.03967 | 0.77089 | 0.3858 |
| OUT | PART. CORR | | | | | |
| 5 VAR(6) | 0.08444639 | . | . | 0.54852 | 0.25138 | 0.6192 |
| 6 VAR(7) | 0.01526350 | . | . | 0.55640 | 0.00816 | 0.9286 |
| 7 VAR(8) | 0.00355501 | . | . | 0.58364 | 0.00044 | 0.9833 |

STEP # 4 R= .848 RSQUARE= .720
TERM ENTERED: VAR(6)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | 0.00005673 | 0.00001950 | .117E+01 | 0.04948 | 8.46182 | 0.0063 |
| 3 VAR(4) | -0.00000189 | -0.00001956 | .07355506 | 0.01387 | 0.00938 | 0.9234 |
| 4 VAR(5) | -0.00000679 | -0.00000814 | -.038E+01 | 0.03951 | 0.69702 | 0.4095 |
| 5 VAR(6) | 0.00665415 | 0.01327160 | .06055901 | 0.54852 | 0.25138 | 0.6192 |
| OUT | PART. CORR | | | | | |
| 6 VAR(7) | -0.20307304 | . | . | 0.05696 | 1.46242 | 0.2349 |
| 7 VAR(8) | -0.15773190 | . | . | 0.12118 | 0.86748 | 0.3582 |

STEP # 5 R= .855 RSQUARE= .731
TERM ENTERED: VAR(7)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | 0.00005066 | 0.00002001 | .104E+01 | 0.04637 | 6.40679 | 0.0162 |
| 3 VAR(4) | 0.00000303 | 0.00001985 | .012E+01 | 0.01328 | 0.02336 | 0.8794 |
| 4 VAR(5) | -0.00000791 | -0.00000814 | -.044E+01 | 0.03900 | 0.94485 | 0.3379 |
| 5 VAR(6) | 0.05386681 | 0.04120735 | .049E+01 | 0.05616 | 1.70881 | 0.1999 |
| 6 VAR(7) | -0.03271546 | -0.02705307 | -.045E+01 | 0.05696 | 1.46242 | 0.2349 |

| OUT | PART. CORR | | | | | |
|---|---|---|---|---|---|---|
| 7 VAR(8) | 0.03349897 | . | . | 0.03169 | 0.03707 | 0.8485 |

STEP # 6 R= .855 RSQUARE= .732
TERM ENTERED: VAR(8)

| VARIABLE IN | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | 0.00005016 | 0.00002047 | .103E+01 | 0.04561 | 6.00256 | 0.0198 |
| 3 VAR(4) | 0.00000414 | 0.00002094 | .016E+01 | 0.01229 | 0.03906 | 0.8445 |
| 4 VAR(5) | -0.00000853 | -0.00000886 | -.047E+01 | 0.03380 | 0.92609 | 0.3429 |
| 5 VAR(6) | 0.05684290 | 0.04456955 | .052E+01 | 0.04940 | 1.62658 | 0.2111 |
| 6 VAR(7) | -0.04159460 | -0.05366335 | -.057E+01 | 0.01490 | 0.60078 | 0.4438 |
| 7 VAR(8) | 0.00611141 | 0.03174018 | .09750775 | 0.03169 | 0.03707 | 0.8485 |

| OUT | PART. CORR |
|---|---|
| none | |

DEP VAR: VAR(9) N: 40 MULTIPLE R: 0.855 SQUARED MULTIPLE R: 0.732
ADJUSTED SQUARED MULTIPLE R: .683 STANDARD ERROR OF ESTIMATE: 0.36795680

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 0.46699163 | 0.70285733 | 0.00000000 | . | 0.66442 | 0.51104 |
| VAR(3) | 0.00005016 | 0.00002047 | 1.03422418 | 0.0456141 | 2.45001 | 0.01976 |
| VAR(4) | 0.00000414 | 0.00002094 | 0.16075081 | 0.0122872 | 0.19764 | 0.84454 |
| VAR(5) | -0.00000853 | 0.00000886 | -0.47190541 | 0.0338014 | -0.96234 | 0.34288 |
| VAR(6) | 0.05684290 | 0.04456955 | 0.51732383 | 0.0494018 | 1.27538 | 0.21108 |
| VAR(7) | -0.04159460 | 0.05366335 | -0.57251194 | 0.0148984 | -0.77510 | 0.44380 |
| VAR(8) | 0.00611141 | 0.03174018 | 0.09750775 | 0.0316941 | 0.19255 | 0.84850 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 12.18925463 | 6 | 2.03154244 | 15.00486965 | 0.00000003 |
| RESIDUAL | 4.46794287 | 33 | 0.13539221 | | |

## Final Regression Equation:

**YIELD = 0.46699163 + 0.00005016($CL_3$) + 0.00000414 ($CL_6$) - 0.00000853 ($CL_9$) + 0.05684290 ($CH_3$) - 0.04159460 ($CH_6$) + 0.00611141 ($CH_9$)**

Where $CL_a$ and $CH_a$ are the Cumulative Light and Cumulative Heat for 'a' weeks respectively.

# APPENDIX E

## EMPIRICAL MATH MODEL (**HUMIDITY TRIAL**)

## STEP-WISE STATISTICAL RESULTS FOR:

Light & Heat & Humidity (3, 6, 9 week) Combined Step-Wise Regression
1992 Equation for Tomato Yield Prediction

---

Definition of Variables:

VAR( 9) = Yield (kg / m^2)
VAR( 3) = Cumulative Light for 3 weeks ( W / m^2)
VAR( 4) = Cumulative Light for 6 weeks ( W / m^2)
VAR( 5) = Cumulative Light for 9 weeks ( W / m^2)
VAR( 6) = Cumulative Heat for 3 weeks ( *C)
VAR( 7) = Cumulative Heat for 6 weeks ( *C)
VAR( 8) = Cumulative Heat for 9 weeks ( *C)
VAR(19)= Humidity, Vapor Pressure Deficit Avg. for 3 weeks (kPa)
VAR(20)= Humidity, Vapor Pressure Deficit Avg. for 6 weeks (kPa)
VAR(21)= Humidity, Vapor Pressure Deficit Avg. for 9 weeks (kPa)

EQUATION LAYOUT:

$$VAR(9) = CONSTANT + a_1VAR(3) + a_2VAR(4) + a_3VAR(5) + a_4VAR(6) + a_5VAR(7) + a_6VAR(8) + a_7VAR(19) + a_8VAR(20) + a_9VAR(21)$$

---

**STEP # 1 R= .820 RSQUARE= .673**
TERM ENTERED: VAR(3)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | 0.00003767 | 0.00000405 | .082E+01 | .1E+01 | .86E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 3 VAR(4) | 0.52099545 | . | . | 0.07492 | .15E+02 | 0.0003 |
| 4 VAR(5) | 0.51599741 | . | . | 0.21251 | .15E+02 | 0.0004 |
| 5 VAR(6) | 0.41882792 | . | . | 0.46353 | 8.72209 | 0.0052 |
| 6 VAR(7) | 0.34316870 | . | . | 0.52028 | 5.47287 | 0.0243 |
| 7 VAR(8) | 0.31997085 | . | . | 0.57098 | 4.67641 | 0.0365 |
| 8 VAR(19) | 0.28913879 | . | . | 0.98058 | 3.74035 | 0.0600 |
| 9 VAR(20) | 0.07158819 | . | . | 0.94214 | 0.21120 | 0.6483 |
| 10 VAR(21) | 0.08411291 | . | . | 0.94064 | 0.29214 | 0.5918 |

**STEP # 2 R= .873 RSQUARE= .762**

TERM ENTERED: VAR(4)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00001041 | -0.00001279 | -.023E+01 | 0.07492 | 0.66273 | 0.4203 |
| 3 VAR(4) | 0.00002623 | 0.00000671 | .109E+01 | 0.07492 | .15E+02 | 0.0003 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 4 VAR(5) | 0.10479260 | . | . | 0.03056 | 0.44414 | 0.5090 |
| 5 VAR(6) | 0.37222300 | . | . | 0.44360 | 6.43334 | 0.0152 |
| 6 VAR(7) | 0.26183291 | . | . | 0.48968 | 2.94410 | 0.0939 |
| 7 VAR(8) | 0.23000082 | . | . | 0.53551 | 2.23421 | 0.1428 |
| 8 VAR(19) | 0.37526138 | . | . | 0.97720 | 6.55608 | 0.0143 |
| 9 VAR(20) | 0.07520101 | . | . | 0.94195 | 0.22749 | 0.6360 |
| 10 VAR(21) | 0.00796328 | . | . | 0.91989 | 0.00254 | 0.9601 |

**STEP # 3 R= .874 RSQUARE= .764**

TERM ENTERED: VAR(5)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00000220 | -0.00001783 | .04782375 | 0.03909 | 0.01517 | 0.9026 |
| 3 VAR(4) | 0.00001524 | 0.00001782 | .063E+01 | 0.01077 | 0.73185 | 0.3974 |
| 4 VAR(5) | 0.00000493 | 0.00000739 | .029E+01 | 0.03056 | 0.44414 | 0.5090 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 5 VAR(6) | 0.36826236 | . | . | 0.44176 | 6.11890 | 0.0178 |
| 6 VAR(7) | 0.25623690 | . | . | 0.48704 | 2.74058 | 0.1059 |
| 7 VAR(8) | 0.21882036 | . | . | 0.52530 | 1.96132 | 0.1693 |
| 8 VAR(19) | 0.36258689 | . | . | 0.88100 | 5.90342 | 0.0198 |
| 9 VAR(20) | 0.05869590 | . | . | 0.91519 | 0.13483 | 0.7155 |
| 10 VAR(21) | -0.01704183 | . | . | 0.86984 | 0.01133 | 0.9158 |

**STEP # 4 R= .892 RSQUARE= .796**

TERM ENTERED: VAR(6)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00000697 | -0.00001690 | -.015E+01 | 0.03858 | 0.17012 | 0.6823 |
| 3 VAR(4) | 0.00001438 | 0.00001678 | .060E+01 | 0.01077 | 0.73421 | 0.3968 |
| 4 VAR(5) | 0.00000382 | 0.00000697 | .023E+01 | 0.03044 | 0.29924 | 0.5875 |
| 5 VAR(6) | 0.01910167 | 0.00772209 | .027E+01 | 0.44176 | 6.11890 | 0.0178 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 6 VAR(7) | -0.40366647 | . | . | 0.03388 | 7.39734 | 0.0098 |
| 7 VAR(8) | -0.33029798 | . | . | 0.08036 | 4.65334 | 0.0374 |
| 8 VAR(19) | 0.44322527 | . | . | 0.86712 | 9.29007 | 0.0042 |
| 9 VAR(20) | 0.43539595 | . | . | 0.51259 | 8.88867 | 0.0050 |
| 10 VAR(21) | 0.47428924 | . | . | 0.33216 | .11E+02 | 0.0020 |

**STEP # 5 R= .911 RSQUARE= .829**
TERM ENTERED: VAR(7)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00001305 | -0.00001582 | -.028E+01 | 0.03781 | 0.68038 | 0.4146 |
| 3 VAR(4) | 0.00001548 | 0.00001556 | .064E+01 | 0.01076 | 0.98936 | 0.3262 |
| 4 VAR(5) | 0.00000457 | 0.00000647 | .027E+01 | 0.03038 | 0.49934 | 0.4841 |
| 5 VAR(6) | 0.09029463 | 0.02713664 | .127E+01 | 0.03073 | .11E+02 | 0.0020 |
| 6 VAR(7) | -0.03759139 | -0.01382135 | -.099E+01 | 0.03388 | 7.39734 | 0.0098 |
| CUT | PART. CORR | | | | | |
| --- | | | | | | |
| 7 VAR(8) | 0.07802977 | . | . | 0.01576 | 0.22666 | 0.6368 |
| 8 VAR(19) | 0.62193128 | . | . | 0.80748 | .23E+02 | 0.0000 |
| 9 VAR(20) | 0.44746344 | . | . | 0.51031 | 9.26293 | 0.0043 |
| 10 VAR(21) | 0.39482351 | . | . | 0.29592 | 6.83292 | 0.0129 |

**STEP # 6 R= .911 RSQUARE= .831**
TERM ENTERED: VAR(8)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00001474 | -0.00001638 | -.032E+01 | 0.03603 | 0.81047 | 0.3738 |
| 3 VAR(4) | 0.00001801 | 0.00001660 | .075E+01 | 0.00965 | 1.17767 | 0.2849 |
| 4 VAR(5) | 0.00000341 | 0.00000698 | .020E+01 | 0.02666 | 0.23875 | 0.6280 |
| 5 VAR(6) | 0.09749132 | 0.03130811 | .137E+01 | 0.02357 | 9.69657 | 0.0036 |
| 6 VAR(7) | -0.05104955 | -0.03152920 | -.134E+01 | 0.00665 | 2.62155 | 0.1139 |
| 7 VAR(8) | 0.00704359 | 0.01479472 | .026E+01 | 0.01576 | 0.22666 | 0.6368 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 8 VAR(19) | 0.62017524 | . | . | 0.80438 | .23E+02 | 0.0000 |
| 9 VAR(20) | 0.44201762 | . | . | 0.49757 | 8.74159 | 0.0055 |
| 10 VAR(21) | 0.41669313 | . | . | 0.28649 | 7.56419 | 0.0093 |

**STEP # 7 R= .946 RSQUARE= .896**
TERM ENTERED: VAR(19)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| --- | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00003216 | -0.00001353 | -.070E+01 | 0.03338 | 5.64930 | 0.0229 |
| 3 VAR(4) | 0.00003744 | 0.00001382 | .155E+01 | 0.00881 | 7.33993 | 0.0103 |
| 4 VAR(5) | -0.00000444 | -0.00000579 | -.026E+01 | 0.02448 | 0.58794 | 0.4482 |
| 5 VAR(6) | 0.12470081 | 0.02555106 | .176E+01 | 0.02238 | .24E+02 | 0.0000 |
| 6 VAR(7) | -0.05877324 | -0.02512748 | -.155E+01 | 0.00662 | 5.47093 | 0.0250 |
| 7 VAR(8) | 0.00357863 | 0.01178866 | .013E+01 | 0.01570 | 0.09215 | 0.7632 |
| 8 VAR(19) | 0.69266441 | 0.14602588 | .028E+01 | 0.80438 | .23E+02 | 0.0000 |
| OUT | PART. CORR | | | | | |
| --- | | | | | | |
| 9 VAR(20) | -0.06834448 | . | . | 0.20402 | 0.16425 | 0.6877 |
| 10 VAR(21) | -0.37137923 | . | . | 0.06060 | 5.59960 | 0.0236 |

**STEP # 8 R= .947 RSQUARE= .896**

TERM ENTERED: VAR(20)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00003287 | -0.00001380 | -.072E+01 | 0.03284 | 5.67163 | 0.0228 |
| 3 VAR(4) | 0.00003840 | 0.00001418 | .159E+01 | 0.00856 | 7.33172 | 0.0104 |
| 4 VAR(5) | -0.00000448 | -0.00000586 | -.027E+01 | 0.02448 | 0.58473 | 0.4496 |
| 5 VAR(6) | 0.12635311 | 0.02617240 | .178E+01 | 0.02184 | .23E+02 | 0.0000 |
| 6 VAR(7) | -0.06235919 | -0.02691998 | -.164E+01 | 0.00590 | 5.36601 | 0.0265 |
| 7 VAR(8) | 0.00442500 | 0.01210936 | .016E+01 | 0.01524 | 0.13353 | 0.7170 |
| 8 VAR(19) | 0.76449037 | 0.23073661 | .031E+01 | 0.32983 | .11E+02 | 0.0022 |
| 9 VAR(20) | -0.18221414 | -0.44960168 | .04886636 | 0.20402 | 0.16425 | 0.6877 |
| OUT | PART. CORR | | | | | |
| 10 VAR(21) | -0.36916807 | . | . | 0.05480 | 5.36484 | 0.0267 |

**STEP # 9 R= .954 RSQUARE= .910**

TERM ENTERED: VAR(21)

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | 'P' |
|---|---|---|---|---|---|---|
| IN | | | | | | |
| 1 CONSTANT | | | | | | |
| 2 VAR(3) | -0.00003803 | -0.00001320 | -.083E+01 | 0.03190 | 8.29517 | 0.0068 |
| 3 VAR(4) | 0.00003563 | 0.00001342 | .148E+01 | 0.00850 | 7.04483 | 0.0120 |
| 4 VAR(5) | 0.00000361 | 0.00000654 | .021E+01 | 0.01749 | 0.30470 | 0.5846 |
| 5 VAR(6) | 0.13616252. | 0.02503952 | .192E+01 | 0.02121 | .30E+02 | 0.0000 |
| 6 VAR(7) | -0.06272876 | -0.02538418 | -.165E+01 | 0.00590 | 6.10671 | 0.0186 |
| 7 VAR(8) | -0.01112060 | -0.01324475 | -.041E+01 | 0.01132 | 0.70497 | 0.4070 |
| 8 VAR(19) | 1.29369637 | 0.31549768 | .053E+01 | 0.15685 | .17E+02 | 0.0002 |
| 9 VAR(20) | 0.13726317 | 0.44581701 | .03681137 | 0.18449 | 0.09480 | 0.7600 |
| 10 VAR(21) | -2.61590639 | -1.12938969 | -.051E+01 | 0.05480 | 5.36484 | 0.0267 |
| OUT | PART. CORR | | | | | |
| none | | | | | | |

THE SUBSET MODEL INCLUDES THE FOLLOWING PREDICTORS:

CONSTANT
VAR(3)
VAR(4)
VAR(5)
VAR(6)
VAR(7)
VAR(8)
VAR(19)
VAR(20)
VAR(21)

**DEP VAR: VAR(9) N: 44 MULTIPLE R: 0.954 SQUARED MULTIPLE R: 0.910 ADJUSTED SQUARED MULTIPLE R: .887**
**STANDARD ERROR OF ESTIMATE: 0.22643645**

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | 1.35770256 | 0.84447452 | 0.00000000 | . | 1.60775 | 0.11714 |
| VAR(3) | -0.00003803 | 0.00001320 | -0.82807586 | 0.0319035 | -2.88013 | 0.00683 |
| VAR(4) | 0.00003563 | 0.00001342 | 1.47884509 | 0.0084953 | 2.65421 | 0.01200 |
| VAR(5) | 0.00000361 | 0.00000654 | 0.21434412 | 0.0174902 | 0.55199 | 0.58456 |
| VAR(6) | 0.13616252 | 0.02503952 | 1.91743544 | 0.0212117 | 5.43791 | 0.00000 |
| VAR(7) | -0.06272876 | 0.02538418 | -1.65176495 | 0.0059029 | -2.47118 | 0.01864 |
| VAR(8) | -0.01112060 | 0.01324475 | -0.40521191 | 0.0113229 | -0.83962 | 0.40698 |
| VAR(19) | 1.29369637 | 0.31549768 | 0.53170290 | 0.1568509 | 4.10049 | 0.00024 |
| VAR(20) | 0.13726317 | 0.44581701 | 0.03681137 | 0.1844946 | 0.30789 | 0.76004 |
| VAR(21) | -2.61590639 | 1.12938969 | -0.50810503 | 0.0548028 | -2.31621 | 0.02671 |

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| REGRESSION | 17.69865677 | 9 | 1.96651742 | 38.35351215 | 0.00000000 |
| RESIDUAL | 1.74329777 | 34 | 0.05127346 | | |

DURBIN-WATSON D STATISTIC 2.100
FIRST ORDER AUTOCORRELATION -.083

RESIDUALS HAVE BEEN SAVED

## Final Regression Equation for Light, Heat, & Humidity (3, 6, 9 wk.)

**Yield (kg/ m²) =** $1.3577 - 0.00003803\,(CL_3) + 0.00003563\,(CL_6) + 0.00000361\,(CL_9) + 0.1361\,(CH_3) - 0.0627\,(CH_6) - 0.0111\,(CH_9) + 1.293\,(VPD_3) + 0.1372\,(VPD_6) - 2.6159\,(VPD_9)$

Where $CL_a$ , $CH_a$ , and $VPD_a$ are the Cumulative Light, Cumulative Heat, and Average Vapor Pressure Deficit (VPD) for 'a' weeks respectively.

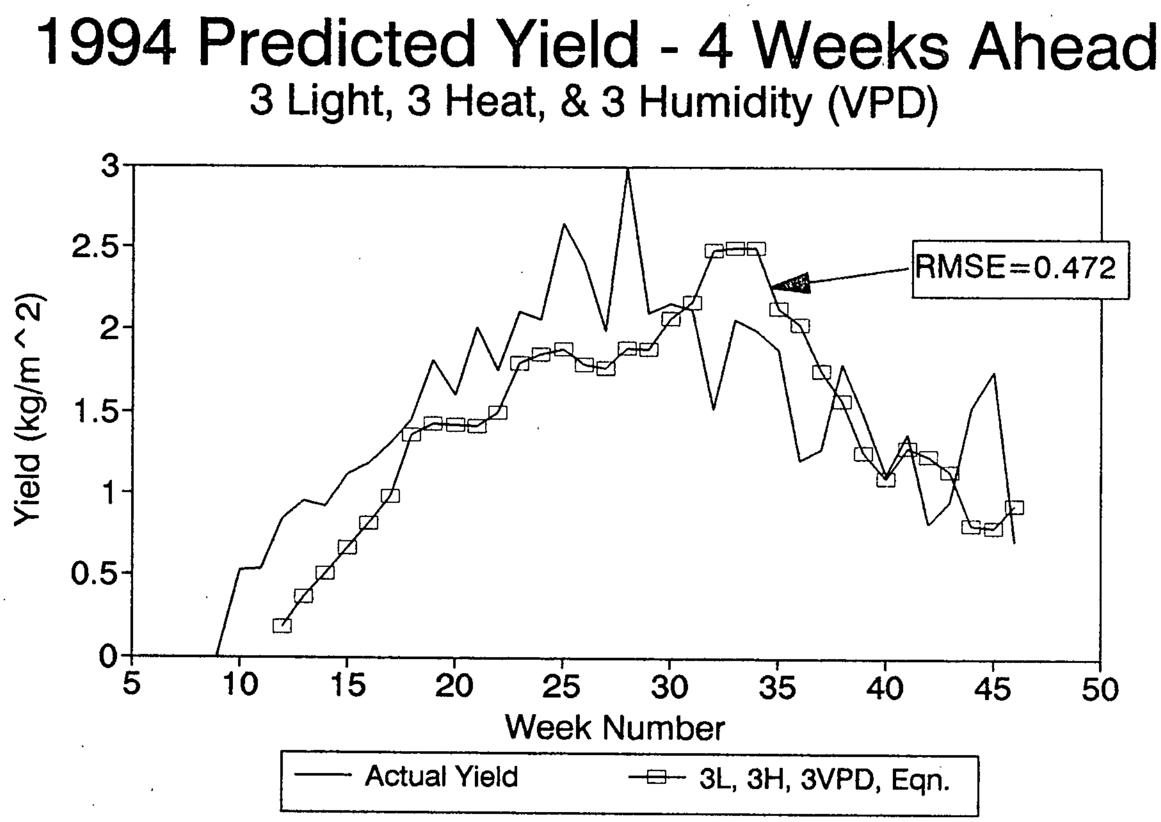



Figure A-22 Predicted yield for 4 weeks ahead using 3 light, 3 heat, 3 humidity variables. (CL3,6,9, CH3,6,9, VPD 3,6,9)