UNIVERSITY of GUELPH

Controlled Environment Systems Research Facility
Guelph, Ontario CANADA
N1G 2W1
Tel. (519) 824-4120
Fax. (519) 767-0755

MELLiSSA

Memorandum of Understanding

ECT/FG/MMM/97.012

Joint Submission of Technical Notes 50.1 and 50.2

Modelling Leafy Vegetable and Root Crop Net Carbon Exchange Responses to Carbon Dioxide Concentration and Light Intensity

Version 1

Issue: 0

Waters G., Dixon, M.A.

University of Guelph

September, 2002

Modelling Leafy Vegetable and Root Crop Net Carbon Exchange Responses to Carbon Dioxide Concentration and Light Intensity

Geoffrey C.R. Waters, Mike A. Dixon

Department of Environmental Biology, University of Guelph, CANADA

ABSTRACT

A number of curve fitting methods were evaluated for their applicability in modeling plant canopy photosynthetic dynamics and responses to environment variables. Non-Parametric and Non-Linear models were applied to long term dynamics in photosynthesis and response surfaces relating photosynthesis and yield to crop age, light intensity and CO_2 concentration. Models were developed for leafy crops including lettuce and beet. All models performed well, but computational details in the non-linear least squares algorithm made the simpler, non-parametric model forms more attractive. Applications of these modeling techniques are addressed in the context biomass production research and in the assessment of air revitalization capacity for bioregenerative life support.

INTRODUCTION

Bioregenerative life support (BLS) system design depends on accurate models of crop growth. Such models may allow for the simplified representation of crop phenology and physiology in closed environments. As such, these models are of use in the systems level modeling of any BLS dynamic which is coupled to crop growth. The European Space Agency MELiSSA program has an interest in the development of models describing nutrient uptake dynamics in higher plants in relation to their photosynthetic capacity. These models are also useful in relating crop yield to environment variables such as light and CO_2 concentration so as to establish cultural recommendations (Lasseur *et al*., 2000). As such, models which describe canopy level photosynthesis may also be used to describe nutrient dynamics when the higher plant chamber is integrated with the microbiological components of the MELiSSA loop. The work in this paper summarizes an initiative to evaluate techniques in modeling plant photosynthetic capacity for the purpose of developing cultural recommendations and developing the framework for dynamic systems models in MELiSSA-HPC integration.

Traditional plant canopy growth analysis involves the destructive harvest of samples of individual plants taken from a full canopy at successive intervals and a determination of the sample dry weights. Biomass accumulation profiles derived from dry weight data are then fitted with models having a defined functional form (Hunt, 1979; Hunt, 1982). Generally, these models have a sole predictor, which is some estimate of plant age (days after planting) or have a number of predictors associated with integrated environment variables (integrated photon flux, degree-days, integrated CO_2 exposures etc.). Destructive measures are useful when static estimates of yield are correlated with environment variables spanning a range of treatment levels (i.e. CO_2 or light intensity). However, in cases where a large range of treatment levels must be spanned it is often impractical to rely on static yield estimates alone since a large number of closed chambers and plants must be used.

While these models are predictive in nature the analyst is forced to choose among particular parametric forms which may result in over or under-fitting. This can result in a high degree of collinearity among predictors, especially if a high order polynomial of a single predictor variable is used to model a complex growth profile. Further, sampling from a full canopy induces thinning responses in remaining plants that can obscure growth profiles. Non-destructive techniques have been developed which allow for growth estimation from measures of whole plant or full canopy photosynthetic activity (Dutton *et al*., 1988). Because these methods are non-intrusive and avoid the need for replicate pairings among successive harvests, they are believed to give better results especially in cases where treatment ranges are limited (Causton, 1991). A problem of such techniques is that resultant measures of photosynthetic activity are highly variable and display more complex profiles over the course of crop development. This is a result of the fact that photosynthetic measures are first derivative estimates of destructive-biomass accumulation profiles. As such these profiles can exhibit instantaneous fluctuations in growth rate not evident in traditional profiles. The fitting of parametric models to photosynthetic profiles is expected to demand high-order polynomials (or a rigid non-linear form) which can result in model instability (Draper and Smith, 1998). Further, it is difficult to find true estimates of canopy photosynthetic variability in the absence of these models. This is a result of the fact that canopy level photosynthetic studies require considerable infrastructure (i.e. full canopy chambers) or demand experiment replication through time. This means that estimates of variability in the air revitalization capacity of bioregenerative systems (which is coupled to photosynthetic activity) is difficult to assess. However, statistical techniques of curve fitting may be harnessed to estimate photosynthetic variability from the error structure of models fitted to data collected over the period of crop growth.

At least two solutions to the problems with non-destructive plant growth analysis exist. First photosynthetic profiles (or more formally Net Carbon Exchange Rates; Dutton *et al* ., 1988) may be integrated to yield non-destructive biomass accumulation profiles. This is disadvantageous because the variability masked by such a procedure has importance from a physiological perspective. Secondly, methods of non-parametric (form-free or smoother) regression may be used. These methods generate predicted values of a dependent variable from empirical data without having to assume any particular functional relationship with the independent variable. Such techniques are believed to be able to generate predicted values (and their variability) in complex profiles with greater stability than their parametric counterparts (due to collinearity problems in high order polynomials; Draper and Smith, 1998). The purpose of this paper is two fold. First, the underlying concepts of form-free regression are reviewed as well as their application in photosynthetic and traditional growth analysis and environment variable correlation analysis. Secondly, a variety of parametric and non-parametric models are applied to photosynthetic data derived from leafy vegetable production trials in closed environments. The application of non-parametric models to the description of canopy gas exchange dynamics is assessed and compared to more-complicated non-linear models. This is done with the intent of illustrating these approaches' utility in BLS biomass production studies aimed at assessing crop growth and air re-vitalization capacity under varying CO_2 and light levels.

THEORETICAL CONSIDERATIONS FOR MODEL DEVELOPMENT

GENERAL REGRESSION SMOOTHERS - Hastie and Tibishirani (1990) provide an excellent and comprehensive review of regression smoothers in the context of Generalized Additive Models. While their work provides many of the theoretical details used in this paper other sources of information include Cleveland and Devlin (1988) and Cleveland (1979). These works deal with locally weighted regression type smoothers.

In the Generalized Additive Model framework it is assumed that for response variable Y and predictor X,

$$Y = f(X) + e \qquad [1]$$

where
$$\begin{aligned} E(e) &= 0 \\ \mathrm{var}(e) &= s^2 \end{aligned} \qquad [2]$$

where ε is the error and E is the expected value. In the case of a smoother, the estimate of function f is given by s, where $s(x_o)$ is the fitted function value at x_o. There are many types of smoother functions, S, which can be used to generate fitted values. In general, the function s utilizes *neighbourhoods* around the value x_o for which a fit is obtained. For example, a running mean smoother might define a neighbourhood around x_o as x_o " k. This is to say that k points to the left and right of x_o define the neighbourhood for x_o. The fitted value $s(x_o)$ for a running mean smoother is obtained by averaging the y_i values for each of the x_i points in the neighbourhood of x_o " k. In general there are a number of ways in which $s(x_o)$ can be computed and each type of smoother differs in:

1) how the response values in each of neighbourhoods of the target value x_o are averaged and,

2) how large the neighbourhoods are defined.

Another type of smoother is the Kernel smoother. In such a smoother an explicitly defined set of weights to be applied to members of the target neighbourhood are used to define an estimate (in this case a weighted average of response variables in the neighbourhood) at each target value. The weights generally decrease in a smooth fashion (say as that defined by a normal density curve) away from the target value x_o. The reader with an interest in these and other types of smoothers are referred to Hastie and Tibishirani (1990). Particular interests in this paper are the cubic-spline and locally weighted regression type smoothers. These are the smoother types that have been used in traditional plant growth analysis (Shipley and Hunt, 1996).

CUBIC SPLINE SMOOTHERS - The cubic spline algorithm operates by placing knots or breakpoints at each x_i. A cubic polynomial is fit in each of the regions marked by the knots such that, at each observation (knot), the polynomials have two continuous derivatives. The cubic spline-smoothing algorithm selects a fitted function s(x) from the family of functions S(x) such that the Penalized Residual Sum of Squares (PRSS) is minimized. The PRSS is given by:

$$[\sum_{i=1}^{n}(y_i - s(x_i))^2 + l \int (s''(t))^2]dt \qquad [3]$$

where s(x) is the cubic spline function. The first term of PRSS is the traditional residual sums of squares estimate. If a function was defined by minimizing this term alone, the result would be an interpolator of the mean values of y. The second term is present to penalize curvature in the model (represented by the second derivative of the spline function). The co-efficient 8 of this term represents span, and can be thought of as the distance between knots (expressed either as a percentage of, or the absolute number of points in the neighbourhood of the target x_o. The smaller the 8, the less dominant the penalty term plays and the greater the curvature to the fitted function. As 8 approaches infinity, the larger the penalty term, and the second derivatives are forced to zero. This means that the fitted line is equivalent to the least squares linear regression line. The best model, then, is one which shows the general trend in the data but which ignores random fluctuations around this trend (Shipley and Hunt, 1996). The choice of 8 is also significant since it defines equivalent degrees of freedom measures used to compare different smoothers.

LOCAL WEIGHTED REGRESSION (LOESS) SMOOTHERS - For a given ordered set of x_o (usually harvest dates in traditional growth analysis or Pn measurement dates in non-destructive analysis) an estimate of y_o is determined using weighted linear or quadratic regression. In this case the regression is local because the fitted y_o is determined after assigning weights based on the tri-cube weight function. If k nearest neighbours are used, then the k nearest neighbours to x_o can be denoted as $N(x_o)$. The value of) (x_o) is first determined:

$$\Delta x_o = \max_{N(x_O)} |x_o - x_i| \qquad [4]$$

This is the distance of the furthest near neighbour from x_o. Weights, w_i are assigned to each near-neighbour in $N(x_o)$ according to the tri-cube weighted function:

$$W\left(\frac{|x_o - x_i|}{\Delta(x_o)}\right) \qquad [5]$$

where $W(g) = \left\{(1 - g^3)^3\right\}$ if 0#g<1

or $W(g) = 0$, otherwise [6]

$s(x_o)$ is computed from the weighted least squares fit of y on x in $N(x_o)$ using the weights computed in Eqn. [6]. This means that the furthest neighbour is assigned a weight of 0. It is easy to see that the local weighted regression smoother is a special type of Kernel smoother with weights determined by the tri-cube function instead of those defined by a normal density curve. In the local weighted regression procedure of S-Plus (loess.smooth), 8 is defined as the percent of nearest neighbours and is used to determine k for neighbourhood definition (Data Analysis Products, 1999).

APPLICATION OF NON-PARAMETRIC MODELS TO PHOTOSYNTHESIS RESPONSE AND RESPONSE SURFACES

An application of non-parametric regression is in the fitting of surfaces describing canopy photosynthetic responses to light intensity, crop age and CO_2 concentration. Data were collected in eight sealed environment chambers located in research greenhouses and laboratories at the University of Guelph using Lettuce (*Lactuca sativa* cv. Bella-green) and Beet (*Beta vulgaris* cv. Detroit Medium Red) to determine their responses to environment variables. These experiments were conducted over a period from June 1998 to August 2001. Greenhouse chambers were lit with natural sunlight and photoperiod and employed a flow through gas analysis system that allowed for CO_2 enrichment and exposure of study plants to concentrations ranging from ambient (350 μL L) to 1300 μL L. Chamber design and operational specifications are the subject of another technical note and are therefore not presented here. Each of the eight greenhouse chambers were stocked with four plants of either lettuce, or beet plants that were left in the chambers from the seedling stage (approximately 21 DAP) to harvestable stage (65 DAP for lettuce, 56 DAP for beet). Net Carbon Exchange data were recorded in each of the eight chambers at 15-minute intervals over the course of crop development. This allowed for the generation of a response profile relating net-photosynthesis during day-light hours (Pn) to ambient light intensity (short term fluctuations) at each observation and crop age (DAP 40 to 65 only) in a single chamber. Yield responses to four CO_2 concentrations (with the exception of lettuce) of 350 ppm, 700 ppm, 1000 ppm or 1300 ppm were developed using two replicates (two chambers) per CO_2 treatment for Beet.

A rectangular parabola (non-linear, parametric) model was applied to the photosynthesis data collected for lettuce and Beet in the greenhouse chambers over the range of crop development and under a range of ambient light intensities. This was done using the *nls* function of S-Plus (Data Analysis Products, 2000). This model is similar to the model presented by Iqbal *et al*, (1996) but allows for dynamic maximum gross photosynthesis and dark respiration rates in relation to crop age. This new model has the form;

$$Pn = \frac{a\mathrm{I}(b_o + b_1 DAP)}{a\mathrm{I} + (b_o + b_1 DAP)} + (b_2 + b_3 DAP) \quad [6]$$

where α is a non-linear least squares estimate of photosynthetic efficiency, β_0 β_1 β_2 and β_3 are parameter estimates, I is the incident photosynthetic photon flux at canopy height, and DAP is days after planting. The surface response resulting from the non-linear least

Light (PPF)	DAP	Loess Regression 95% CI for Predicted Pn			Modified Rectangular Hyperbola 95% CI for Predicted Pn		
		Fit	upper	lower	fit	upper	lower
215	40	0.323	0.438	0.206	0.284	0.355	0.212
146	46	0.751	0.808	0.693	1.21	1.24	1.17
101	54	1.07	1.12	1.01	1.59	1.62	1.55
394	57	3.63	3.72	3.54	3.59	3.67	3.50
494	61	4.71	4.87	4.55	4.47	4.58	4.36

Table 1 Comparison of variability measures derived from non-linear parametric (modified rectangular hyperbola) and locally weighted (loess) regression at five selected points. Presented are the 95% confidence intervals for the predicted Pn response for both models.

squares fit for lettuce data is presented in Figure 2. The model fit the data well with a residual-sum of mean-squared error of 0.460 with df=2597 and model df=4. All parameters were significant at the p=0.05 with the exception of β_2. Parameter estimates were α=0.057, β_0 =-10.28, β_1 = 0.289, β_2 = 0.064 and β_3 = -0.226.

A loess model was fit to the same set of lettuce, single chamber data. The resulting model had equivalent degrees of freedom of 7.1, a residual standard error of 0.3812 and a multiple r^2 value of 0.89. A plot of the fitted response surface is presented in Figure 3.

A comparison of fitted values and 95% confidence intervals (using traditional inferential techniques in both cases) for prediction are presented in Table 1 for five selected points on the surface. Both the non-linear parametric model and the loess model performed well. A comparison between fitted values across the DAP and Light domains using the Modified Welch two sample t-test (for unequal variances) indicated no significant difference at p=0.05. However, the prediction variability estimates presented in Table 3 indicate that the non-linear model gave tighter estimates of canopy photosynthetic response owing to its lower model degrees of freedom. It is important to note that use of the non-parametric model avoided difficulties of having to obtain initial parameter estimates and the computational challenges associated with non-convergence in the nls algorithm. These results indicate that the loess model could perform as well as, if not better than the non-linear parametric model(s) used to describe light/crop age influences on photosynthesis.

A comparison of prediction intervals at four common surface points was made with the 95% confidence intervals for the mean of photosynthesis measures derived from the five other chambers (not used in model development). Only 50% of the predicted model responses (both the non-linear and loess) were marginally significantly different (p=0.05) from the mean responses in the other five chambers. This is a result of error, which may have been accounted for by incorporating a chamber effect. Despite these effects, the non-parametric and parametric models estimated Pn variability quite well in data extraneous to those used in their development.

The performance of the non-parametric and non-linear models indicates that such models can be applied to data collected using Beet. The parametric model presented in equation [6] was fit to Beet data collected as described above. In the initial mode formulation, however, terms for CO_2 concentration were included in the rectangular hyperbola form. The results of ANOVA indicated that CO_2 was not a significant predictor of NCER and these terms were therefore dropped from the model. This result seems a bit curious but is likely a result of two factors. First, NCER data, as described earlier, is highly variable. The limited range (as a result of control) in CO_2 concentrations to which the plants were exposed meant that the CO_2 term in the model did not account for a significant portion of variability in NCER. Secondly, the crop was grown for the majority of its life cycle in non-saturating lighting conditions. This means that the effect of CO_2 concentrations on NCER were diminished. However, an analysis of post harvest data indicated that CO_2 had a significant effect on edible and total plant biomass. The reader is reminded that small, and statistically insignificant differences in NCER owing to CO_2 treatment effects may be manifested as significant differences in yield since biomass is a result of integrated NCER over the period of crop development.

The results of the parametric and non-parametric fits are presented in Figures 4 and 5, respectively. The results of comparisons made between the LOESS and the rectangular hyperbola model were similar to those described above for lettuce, with the exception of a small valley at mid-to high light ranges at crop maturity. This prediction of the LOESS model can not be explained. The modified rectangular hyperbola model performed very well with regards to allowing for variable light compensation points and shifting points of light saturation. The parametric model shows very early saturation of the crop to light at young ages followed by an increase in photosynthetic efficiency (greater slope) as the crop matures.

The relationship between crop edible, total biomass and CO_2 concentration are presented in Figures 6 and 7. A significant relationship between these harvest indices and CO_2 concentration was found ($p=0.01$).

GENERAL CONCLUSIONS

A comparison between parametric non-linear and non-parametric models used to describe the photosynthesis response surfaces indicated that either would suffice in terms of their ability to capture the main trends in a complicated NCER trajectory.

The effects of CO_2 concentration of Beet NCER were not significant but CO_2 had a significant effect on total plant and edible biomass. The non-parametric model performed well and indicated strong crop developmental stage effects on light response.

ACKNOWLEDGEMENTS.

Funding for this project has been provided by the European Space Agency (MELiSSA) , Agriculture and Agri-food Canada, the Centre for Research in Earth and Space Technology (CRESTech), and the University of Guelph.

CONTACT

Geoffrey Cloutier is a Ph.D. candidate in the Department of Plant Agriculture, University of Guelph, CANADA. His research interest is in the modeling of dynamic interactions between higher plant biomass production chambers and the MELiSSA loop.

E-mail: waters@ces.uoguelph.ca

REFERENCES

Atkinson, A.C. (1997). Plots, Transforms and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Toronto: Oxford Science Publications.

Causton, D.R. (1991). Plant Growth Analysis: The Variability of Relative Growth Rate Within a Sample. *Ann. Bot.* **67**, 137-144.

Cleveland, W.S., Devlin, S.J. (1988). Locally- Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* **83**, 829-836.

Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**, 596-610.

Cloutier, G.R., Stasiak, M.A. Cote, R., Dixon, M.A., Arnold, K.E. (1998). Physiological Aspects of Integrated Crop Production in Advanced Life Support Systems. SAE International Conference on Environmental Systems, Danvers, Mass. Paper 981561.

Data Analysis Products (1999). S-Plus 2000. Guide to Statistic, Vol. I., a division of Math Soft Inc. Seattle.

Draper, N.R., Smith, H. (1998). Applied Regression Analysis. Toronto: John Wiley & Sons Inc.

Dutton, R., Jiao, J., Tsujita, J., M., Grodzinski, B. (1988). Whole Plant CO_2 Exchange Measurement for Nondestructive Estimation of Growth. *Plant Physiol.* **86**, 355-358.

Hastie, T.J., Tibshirani R.J. (1990). Generalized Additive Models. New York: Chapman and Hall.

Hunt, R. (1979). Plant Growth Analysis: The Rationale Behind the Use of the Fitted Mathematical Function. *Ann. Bot.* **43**, 245-249.

Hunt R. (1982). Plant Growth Curves. The Functional Approach to Plant Growth Analysis. London: Edward Arnold.

Iqbal, R.M., Rao, A., Rasul, E., Wahid, A. (1996). Mathematical Models and Response Functions in Photosynthesis: An Exponential Model in *Handbook of Photosynthesis*, Pessaraki M. ed. Marcel Dekker, Inc., New York.

Lasseur, C., Dixon, M.A., Dubertret, G., Dussap, C.G., Francesco, G., Gros, J.B. (2000). MELiSSA: 10 Years of Research, Results, Status and Perspectives. SAE Technical Paper 00ICES-350.

Leger C., Politis, D.N., Romano, J.P. (1992). Bootstrap Technology and Applications. *Technometrics* **34**, 378-398.

Shipley, B., Hunt, R., (1996). Regression Smoothers for Estimating Parameters of Growth Analyses. *Ann. Bot.* **78**: 569-576.

Lettuce Photosynthesis

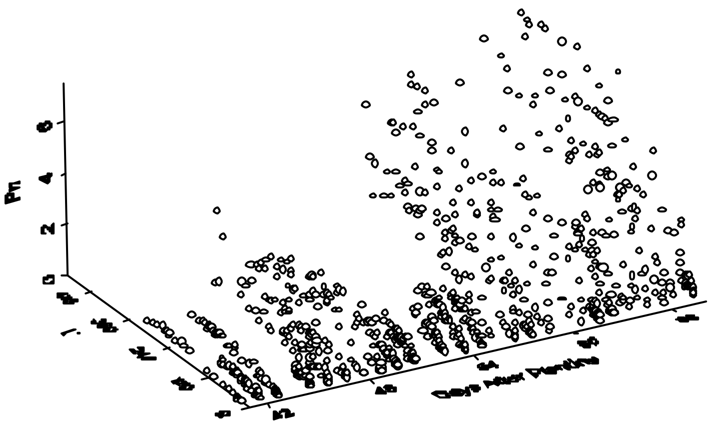

Figure 1. Photosynthesis response data derived from a closed chamber trial using lettuce under ambient photoperiod and light intensity. Pn is expressed in ppm CO_2 s^{-1}(: L CO_2 L^{-1} s^{-1}). Light Intensity (PPF) is expressed in µmol m^{-2} s^{-1} PAR.

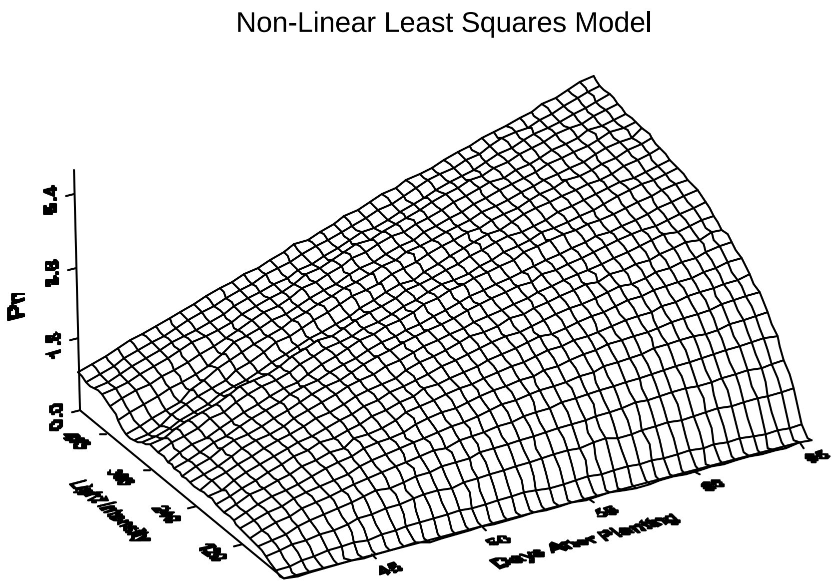

Figure 2. Non-linear response surface fit to the lettuce photosynthesis data presented in Figure 1. Fit is a result of a modified rectangular hyperbola. Pn is expressed in ppm CO_2 s^{-1}(: L CO_2 L^{-1} s^{-1}). Light Intensity (PPF) is expressed in µmol m^2 s^{-1} PAR.

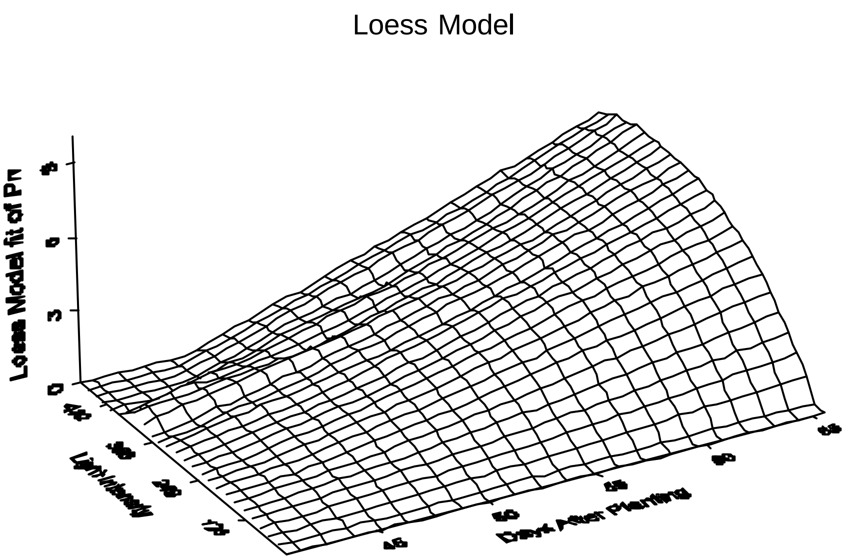

Figure 3. Loess model fit the photosynthesis the lettuce photosynthesis response data presented in Figure 1. Pn is expressed in ppm CO_2 s^{-1}(: L CO_2 L^{-1} s^{-1}). Light Intensity (PPF) is expressed in µmol m^{2} s^{-1} PAR.

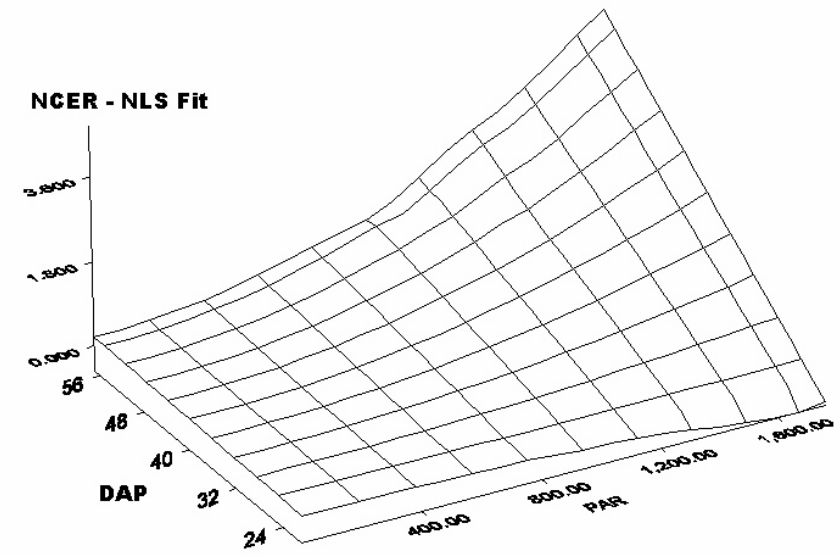

Figure 4. Non-linear response surface fit to the Beet NCER data. Fit is a result of a modified rectangular hyperbola. NCER is expressed in $\mu mol\ CO_2\ s^{-1}$ Light Intensity (PAR) is expressed in $\mu mol\ m^{-2}\ s^{-1}$ PAR and DAP is Days After Planting.

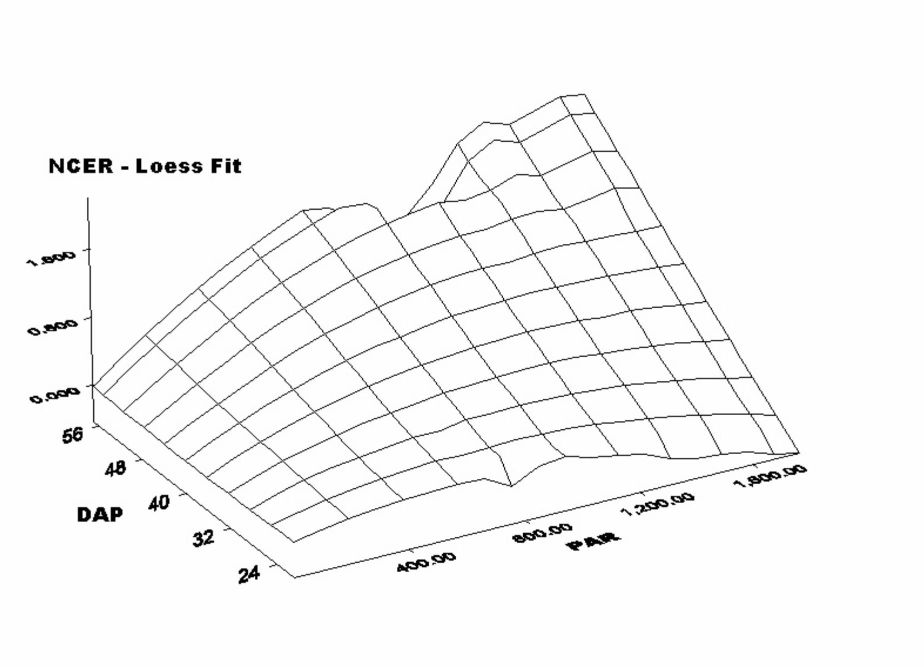

Figure 5. LOESS response surface fit to the Beet NCER data. Fit is a result of a modified rectangular hyperbola. NCER is expressed in µmol CO_2 s^{-1} Light Intensity (PAR) is expressed in µmol m^{-2} s^{-1} PAR and DAP is Days After Planting.

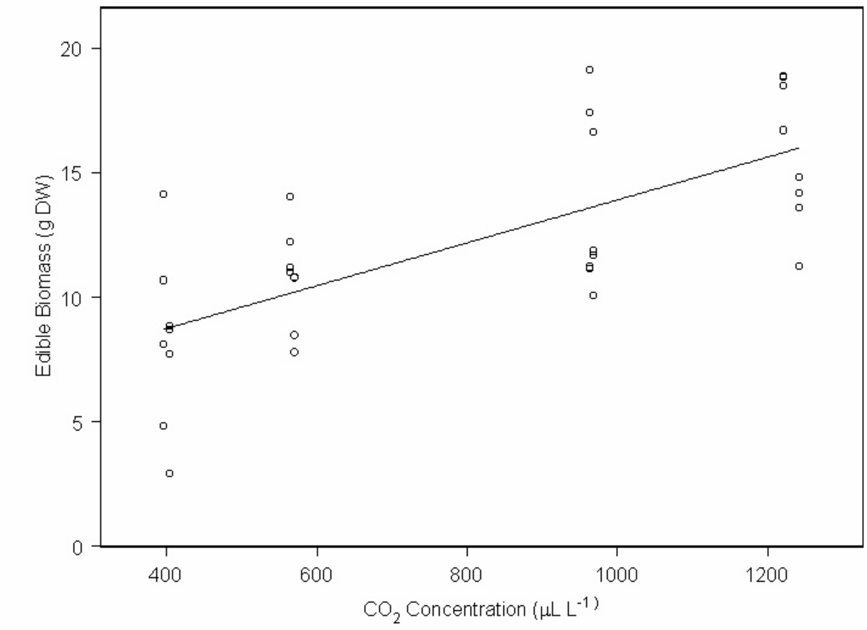

Figure 6. Beet Edible Biomass (g Dry Weight (DW)) relationship with CO_2 concentration in the chambers. The fitted line is a result of a LOESS regression with df=2.

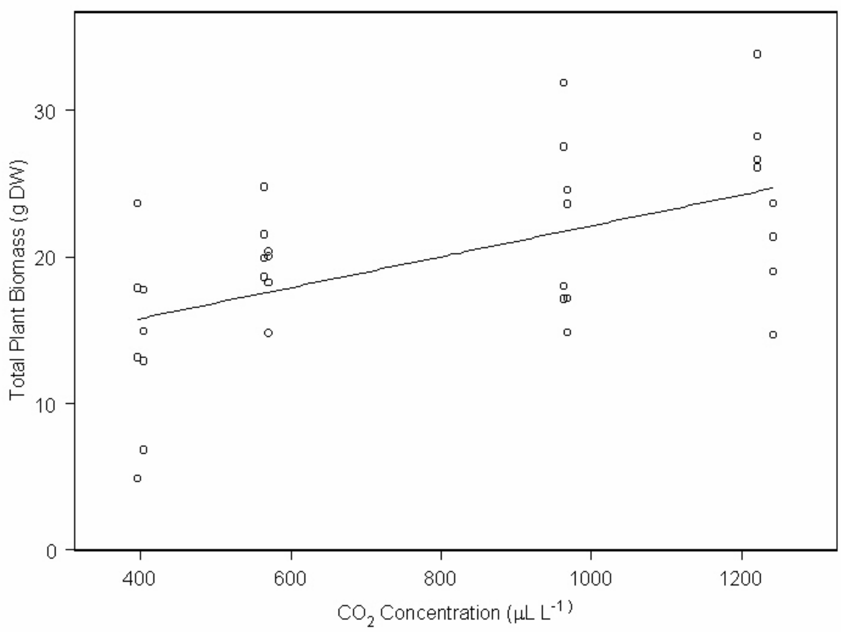

Figure 7. Total Beet Biomass (g Dry Weight (DW)) relationship with CO_2 concentration in the chambers. The fitted line is a result of a LOESS regression with df=2.