Proceedings of the 11th International Symposium on
Computer Applications in Biotechnology (CAB 2010),
Leuven, Belgium, July 7-9, 2010
Julio R. Banga, Philippe Bogaerts, Jan Van Impe,
Denis Dochain, Ilse Smets (Eds.)



# Identification of simple mass balance models for plant growth - Towards food production on manned space missions

**Heather Maclean** * **Denis Dochain** *,⋆ **Geoff Waters** **
**Mike Dixon** ** **Laury Chaerle** ***
**Dominique Van Der Straeten** ***

* *CESAME, Université catholique de Louvain, 4-6 avenue G. Lemaître 1348 Louvain-la-Neuve, Belgium*
** *Controlled Environment Systems Research Facility, School of Environmental Sciences, University of Guelph, 50 Stone Road East, Guelph, Ontario, N1G 2W1, Canada*
*** *Unit Plant Hormone Signaling and Bio-Imaging, Department of Physiology, Ghent University, Ledeganckstraat 35, 9000 Gent, Belgium*

**Abstract:** This paper presents a simple mass balance model for plant growth. This work is a first step in the development of a model intended to enable the prediction and control of a plant production chamber for MELiSSA, a regenerative life support system project developed by the ESA. Photosynthesis and respiration were selected as key reactions for biomass production. Considering these reactions, the model was developed using a mass balance approach. Reaction kinetics were chosen based on plant physiology and standard biochemical reaction knowledge. The identification and validation of yield and kinetic parameters were performed using data from lettuce and beet experiments in a closed plant chamber. The model adequately predicts lettuce growth, and predicts beet biomass and carbon dioxide flux well after an initial acclimation phase. The oxygen prediction could be improved and should be the subject of further study.

Keywords: Plant growth modelling, identification

## 1. INTRODUCTION

### *1.1 State of the Art*

Plant growth models have been developed for the purpose of increasing our knowledge of plants, improving agricultural practices, as a tool in landscaping, and for the purpose of optimization and control. Modelling efforts have taken several different approaches depending on the intended use of the model. Many plant growth models are empirical, and apply fitted functions without considering the biological mechanisms underlying plant growth. These models have the benefit of simplicity, but are not mechanistic and therefore cannot be applied to a variety of species or over a wide range of conditions [1]. In contrast, complex metabolic models give a more complete description of reactions taking place within the cells, and are useful tools for studying plant development [2][3]. However, because of the large complexity, these models are typically over-parameterized and unidentifiable, making them unsuitable for prediction and control purposes.

Process based models and functional-structural models attempt to bridge this gap. These models consider at least some plant processes and interactions between the plant and the environment. Process based models typically refer to those models that do not take plant morphology into account [4][5], while functional structural models generally include an empirical view of plant architecture [6][7]. These models work well under certain environmental conditions, but they are usually developed for plant growth under field conditions, and therefore neglect the effect of some important environmental variables (for example carbon dioxide and oxygen concentration). More mechanistic models, based on the reaction kinetics of the most important processes, should be applicable over a wider range of condtions.

### *1.2 The MELiSSA Project*

The MELiSSA (Micro-Ecological Life Support System Alternative) project aims to develop technology for a future regenerative life support system for long term manned space missions. Developed by the European Space Agency, the concept is to use microorganisms and plants to regenerate the atmosphere, recycle water, and to produce food for the crew on such missions. An important part of the MELiSSA loop is the growth of higher plants in a controlled greenhouse environment for the production of food and oxygen from 'waste' carbon dioxide. A model of plant growth is required for the prediction and eventual control of this compartment. The model must be applicable at normal operating conditions, as well as during failure and stress conditions. The main control objective will be to provide a certain desired 'flow' of biomass from

⋆ Honorary Research Director FNRS, Belgium. e-mail : denis.dochain@uclouvain.be, fax : +3210472180

the plant chamber as food for the crew. Both quantity and nutritional requirements will need to be maintained.

### *1.3 Objectives*

A first principle model of plant growth will be developed, for the prediction and control of a plant growth chamber. The model will be based on the reaction kinetics of key reactions and should be robust in that the results of failures or unexpected environmental conditions should be predictable. As a first step in development, a simple model capable of predicting the most important states for plant growth under normal operating conditions will be selected.

## 2. DATA AVAILABILITY

Data from lettuce cultivars was provided by the University of Guelph. Available temporal data (every 6 minutes) included carbon dioxide concentration in the atmosphere of the plant chamber, light data (photosynthetically active radiation, PAR, at canopy height), and data to calculate the rate of $CO_2$ addition to the chamber ($u_1$). In addition, oxygen data in the atmosphere was measured at 6 minute intervals with the data recording alternating ON for 6 hours, OFF for 6 hours due to system set up. Biomass and leaf area measurements were only available at the beginning and end of the experiment.

## 3. MODEL SELECTION

### *3.1 Selection of Important Reactions*

Plants are complicated systems. Their growth and development involves a large number of interconnected processes and reactions. Plant models can not incorporate all of these processes, and therefore those that are most important need to be selected. To simplify the selection process, only total biomass production was considered, and therefore biomass partitioning was not included at this stage. In the first model selection, only photosynthesis (1), photorespiration (2), and mitochondrial respiration (3) were considered. These reactions were selected because they are the main reactions influencing the production of biomass, as well as the exchange of carbon dioxide and oxygen between the plant and the atmosphere.

$$CO_2 + H_2O \xrightarrow{Light} Biomass + O_2 \quad (1)$$

$$Biomass + O_2 \xrightarrow{Light} CO_2 + H_2O \quad (2)$$

$$Biomass + O_2 \longrightarrow CO_2 + H_2O \quad (3)$$

From these reactions it can be deduced that carbon dioxide and oxygen concentration inside the leaves, light intercepted by the plant, total biomass produced, and water availability are the main factors influencing plant growth. Nutrient concentrations will also have an effect. However, at this stage in the model development it was assumed that water and nutrients are at sufficient levels so as not to influence the rate of photosynthesis. Therefore, for simplicity, we have neglected transpiration and root water uptake. To improve the robustness of the model, these and other factors and processes may need to be included at a later stage.

One interesting characteristic of the reactions under consideration is that the respiration reactions essentially act to reverse photosynthesis, regenerating the photosynthetic substrates, $CO_2$ and $H_2O$. This property was used to simplify the identification problem. The model was formulated to treat photosynthesis, photorespiration, and mitochondrial respiration as a single stoichiometrically reversible reaction, which will proceed in the forward direction during the day ($CO_2$ consumed, $O_2$ produced) and in the negative direction at night ($CO_2$ produced, $O_2$ consumed). By formulating the model in this way, we are making the assumption that the stoichiometry is the same for all three reactions. Theoretically, this is not an unrealistic assumption to make. However, practically it is unlikely that the stoichiometry will be exactly the same, since the reactions take different pathways. However, for the first model selection, we will assume that the stoichiometry is close enough.

If we write the model based on this assumption, the reaction scheme can be reduced to one equation:

$$CO_2 + H_2O \xleftrightarrow{Light} Biomass + O_2 \quad (4)$$

### *3.2 Mass Balance Model*

From the reaction scheme (4), a mass balance model ((5)-(7)) was derived for biomass ($XV$, g), carbon dioxide concentration inside the leaves ($C_i$, g $m^{-3}$) and oxygen concentration in the leaves ($O_i$, g $m^{-3}$). Gas concentrations inside the plant are known to vary widely, and therefore $C_i$ and $O_i$ should be considered virtual and not representative of the concentration in a particular organelle. In writing this set of equations, the simplifying assumption was made that the carbon dioxide and oxygen concentration inside the leaves will be approximately equal to their concentrations in the atmosphere of the plant chamber ($C_a$ and $O_a$ respectively, g $s^{-1}$). Therefore, instead of considering a balance on the gases inside the plant, a more simple balance on the chamber was considered. This assumption is not entirely correct, as there should be some resistance to transfer which would cause these concentrations to be different. However, the concentration in the leaves should be proportional to the concentration in the atmosphere and the trends in the data should be more or less the same assuming other factors (such as water availability) are not significant [8]. Therefore, based on the updated reaction scheme and this new assumption, the mass balance equations were written as follows:

$$\frac{dXV}{dt} = Y_1 r \quad (5)$$

$$\frac{dC_i}{dt} = \frac{dC_a}{dt} = \frac{-r}{V_{chamber}} + \frac{u_1}{V_{chamber}} \quad (6)$$

$$\frac{dO_i}{dt} = \frac{dO_a}{dt} = \frac{Y_2 r}{V_{chamber}} \quad (7)$$

where $r$ is the rate equations to be defined (g $s^{-1}$), $Y_1$ and $Y_2$ are the yields (g $g^{-1}$), $V_{chamber}$ is the volume of the plant growth chamber (29 $m^2$), and $u_1$ is the rate of $CO_2$ addition to the chamber to control the $CO_2$ concentration at a minimum value of 1000 ppm (g $s^{-1}$).

## 4. YIELD IDENTIFICATION

### *4.1 Separating Yield and Kinetic Parameter Identifications*

As the model is defined there are two unknown yield parameters ($Y_1$ and $Y_2$) to be identified and a rate equation ($r$), with its associated kinetic parameters, to be defined. This identification problem can be simplified by separating it into several steps. If the yield identification can be decoupled from the identification of the kinetic parameters, then the yield coefficients can be estimated without modelling the reaction rates. This is useful because choosing appropriate reaction kinetics is a difficult task, and by identifying the yields separately we can eliminate some of the complexity.

The separation of the yield identification from the identification of the kinetic parameters has been demonstrated [9]. The method makes use of a state transformation based on the structure of the model (8). By applying this transformation, the initial model can be transformed into one which does not depend on the reaction kinetics (9).

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} Y_1 V_{chamber} \\ Y_2 \end{bmatrix} C_i + \begin{bmatrix} XV \\ O_i \end{bmatrix} \tag{8}$$

$$\frac{d}{dt}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} Y_1 u_1 \\ \dfrac{Y_2 u_1}{V_{chamber}} \end{bmatrix} \tag{9}$$

Therefore, using (8) and (9), we can derive simple equations which can be used for the yield identification without any prior knowledge of the reaction rates ((10)-(11)).

$$\frac{dXV}{dt} = Y_1 \left( u_1 - V_{chamber} \frac{dC_i}{dt} \right) \tag{10}$$

$$\frac{dO_i}{dt} = Y_2 \left( \frac{u_1}{V_{chamber}} - \frac{dC_i}{dt} \right) \tag{11}$$

### *4.2 Identification of the Yields*

The yield identification was performed using only initial and final measurements, since temporal data was not fully available. The integrated forms of (10) and (11) were used to identify the yields based on three available data sets using a least squares method.

The identification was done in two steps. Initially, two data sets were chosen for identification, and the remaining data set was used for validation. This process was repeated using the two other combinations of data sets for identification. The yields and their associated confidence intervals for each of the three trials are shown in Table 1.

| | $Y_1$ (g $XV$ g $CO_2^{-1}$) | $Y_2$ (g $O_2$ g $CO_2^{-1}$) |
|---|---|---|
| $Trial1$ | 0.631 [-0.642, 1.905] | 0.363 [-0.167, 0.894] |
| $Trial2$ | 0.604 [-1.229, 2.437] | 0.323 [-0.743, 1.389] |
| $Trial3$ | 0.532 [0.045, 1.019] | 0.302 [-0.203, 0.807] |
| $'best' estimate$ | 0.585 [0.259, 0.911] | 0.329 [0.142, 0.516] |

Table 1. Results of the yield identification.

The yields identified in each of the trials are quite similar. However, the confidence intervals are large and there was error in the validations (not shown). This is due to the small sample size and the large error associated with these types of independent experiments. In spite of this, the similarity of all three yield estimates suggest that the identification is satisfactory, and therefore the available data is sufficient to identify the yields. Therefore, a final identification step was performed to get the 'best' possible yield estimates from the available data. In this case all three of the data sets were used for identification. The yield values are shown in the final row of Table 1. These yields were taken as the true values, and set as constants for the kinetic parameter identification.

## 5. KINETIC MODEL IDENTIFICATION

### *5.1 Data Required for Kinetic Model Identification*

Temporal data for biomass, light, carbon dioxide and oxygen concentration in the plant chamber is required for the kinetic model identification. This data was available for carbon dioxide, but only periodically for oxygen due to system set up. Therefore, average oxygen concentrations were calculated for the periods over which the data recording was off. Biomass measurements were only available at the beginning and end of experiments. However, temporal biomass data was generated using the transformation that was used for the yield identification (10). Light available for photosynthesis and photorespiration ($\mu$mol PAR $m^{-2}$ $s^{-1}$) was quantified using a standard method analogous to the Beer-Lambert law:

$$Light = IP(1 - exp(-k\ LAI)) \tag{12}$$

$$LAI = \frac{LA}{PA} \tag{13}$$

$$LA = LAR \times XV \tag{14}$$

where $IP$ is the incident photon flux ($\mu$mol PAR $m^{-2}$ $s^{-1}$), $k$ is the extinction coefficient, $LAI$ is the leaf area index, $LA$ is the leaf area ($m^2$), $PA$ is the planting area ($m^2$), and $LAR$ is the leaf area ratio ($m^2$ $g^{-1}$).

### *5.2 Kinetic Model Selection*

The rate under consideration is a net photosynthetic rate (15) which should include a term for photosynthesis ($r_{ps}$), photorespiration ($r_{pr}$), and mitochondrial respiration ($r_{dr}$). Several kinetic models were tested. The proposed kinetic model is shown in (15) to (18). From the reaction scheme, the rates of photosynthesis and photorespiration should depend on the concentrations of $CO_2$ and $O_2$ respectively, as well as the amount of light energy available. Standard Monod kinetics were chosen to represent these rates. The rate of mitochondrial respiration is said to be affected by both the energy demands of the plant and the rate of supply of the carbon substrates produced through photosynthesis [10]. In the current model, mitochondrial respiration is divided into two components as has been often done in the literature [11]: a growth respiration term proportional to the rate of production of carbon substrates (represented here by photosynthesis less photorespiration) and a maintenance respiration term proportional to total biomass. Upon identification of the kinetic parameters it was found that the maintenance respiration term was very small and did not affect the fit of

the model. Therefore this term was omitted and equation (18) simply consists of the term for growth respiration.

$$r = r_{ps} - r_{pr} - r_{dr} \tag{15}$$

$$r_{ps} = \frac{v_1 C_i Light}{K_C + C_i} \tag{16}$$

$$r_{pr} = \frac{v_2 O_i Light}{K_O + O_i} \tag{17}$$

$$r_{dr} = v_3 \left( r_{ps,daily\ avg} - r_{pr,daily\ avg} \right) \tag{18}$$

In the above equations, $v_1$ (m$^2$ g $\mu$mol PAR$^{-1}$), $v_2$ (m$^2$ g $\mu$mol PAR$^{-1}$) and $v_3$ (no units) are rate constants to be identified, $K_C$ and $K_O$ are Michaelis-Menten constants for carbon dioxide and oxygen which were estimated from literature [12] (g m$^{-3}$), and $r_{ps,daily\ avg}$ and $r_{pr,daily\ avg}$ are the average values of the rate of photosynthesis ($r_{ps}$) and photorespiration ($r_{pr}$) over the previous day (g s$^{-1}$).

### *5.3 Identification and Validation of the Kinetic Model on Lettuce Data*

The identification of kinetic parameters ($v_1$, $v_2$ and $v_3$) was solved using a least squares optimization. The results of the validation on lettuce data are shown in Figure 1.

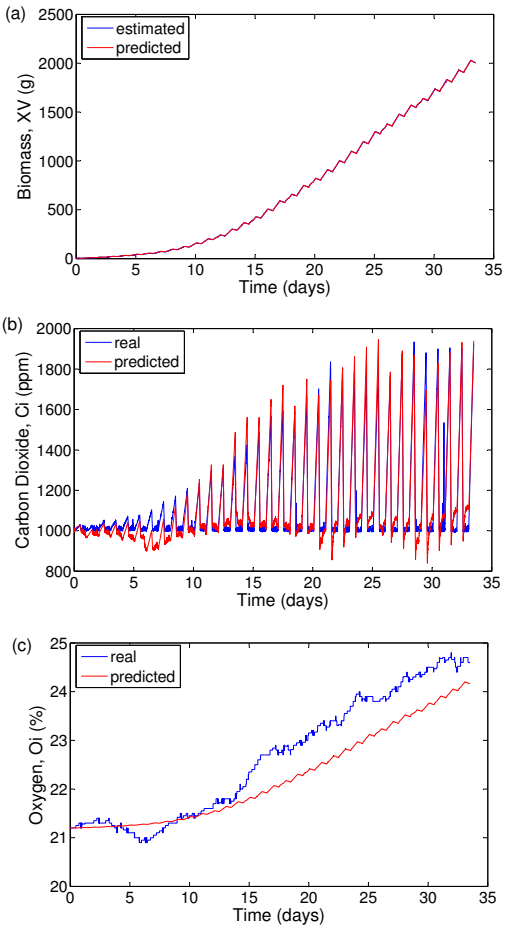


Fig. 1. Validation of kinetic model identification on lettuce data, showing predicted and real (or estimated) (a) biomass (b) $CO_2$ and (c) $O_2$ versus time.

The validation shows that the biomass and carbon dioxide predictions fit the data well. There is some offset between the oxygen prediction and the measured data. This offset is partially due to the error in the identification of the yields (specifically $Y_2$). In this case, the yield underestimates the total oxygen produced in the experiment, and this causes the prediction to be offset from the true values. Despite this problem, the model describes the data fairly well.

As a second validation step, the identified model was tested on a data set which was not used for identification of the kinetic parameters. The results (Figure 2) were quite similar to those observed using the identification data set. This reinforces the validity of the model for predicting lettuce growth.

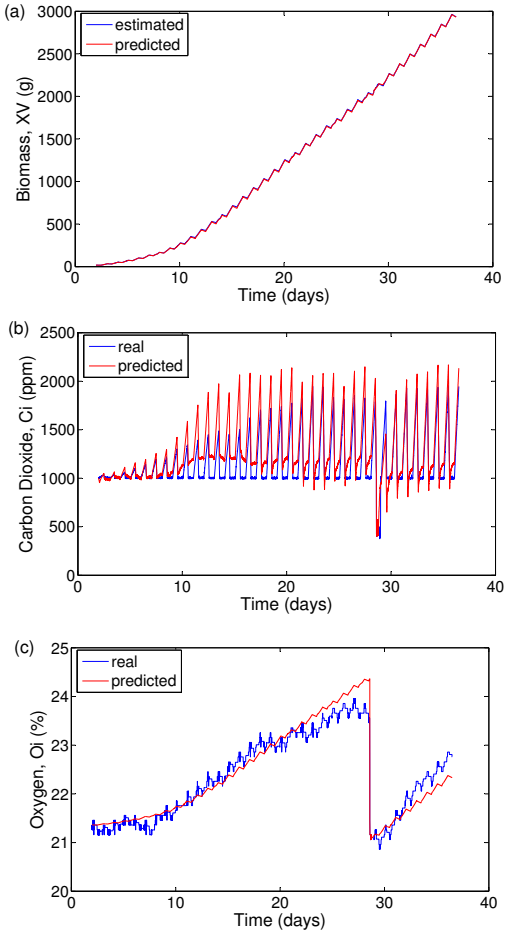


Fig. 2. Validation of kinetic model on an independent lettuce data set, showing predicted and real (or estimated) (a) biomass (b) $CO_2$ and (c) $O_2$ versus time. The drop in oxygen concentration on day 28 was a result of chamber opening - the model was reinitialized to account for this.

### *5.4 Identification and Validation of the Kinetic Model on Beet Data*

In order to achieve a general model of plant growth that can be applied to a variety of plants with minimal adjustments, the model must be tested on a wide range of species. Table beet data from the University of Guelph was therefore used to identify and validate the model. The

experiments were performed in the same growth chambers as the lettuce experiments, and the data recorded is very similar. The model was not adjusted in any way except to redefine the leaf area ratio (ratio of leaf area to biomass, $LAR$, $m^2\ g^{-1}$), and to re-identify yields and kinetic parameters. The yield identification was performed using two data sets. The yields and confidence intervals estimated were $Y1$=0.694 g $g^{-1}$ [0.674, 0.713] and $Y2$=0.347 g $g^{-1}$ [0.268, 0.427]. Despite the small number of experiments, the confidence intervals on the yields are fairly small, suggesting that the values obtained should be adequate. These yields were therefore taken as constants, and the kinetic parameter identification was performed using a least squares optimization technique. The validation is shown in Figure 3.

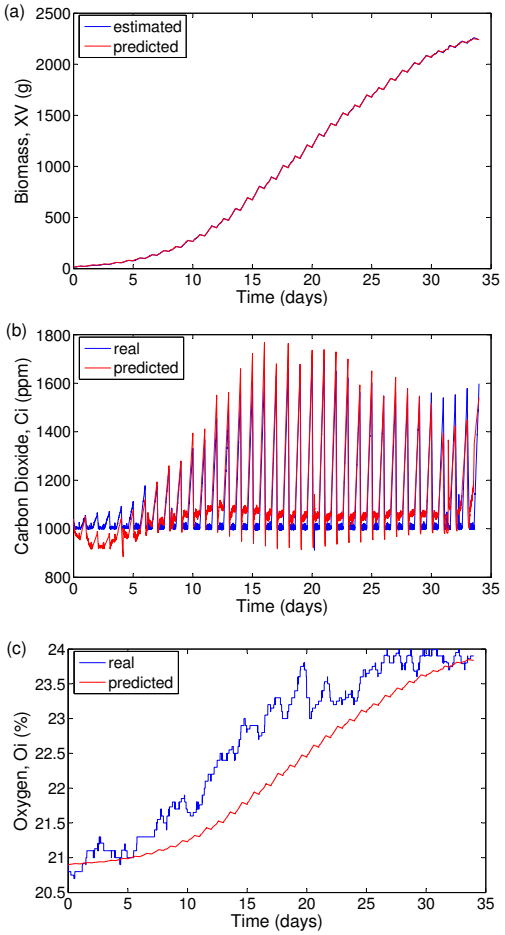


Fig. 3. Validation of kinetic model identification on beet data, showing predicted and real (or estimated) (a) biomass (b) $CO_2$ and (c) $O_2$ versus time.

The model is a fairly good fit for the beet data. However, the carbon dioxide concentration is underestimated early in the experiment. It was hypothesized that this could be due to the effects of transplanting the seedlings in the chamber. During transfer, the plants experience root damage and may need some adaptation time to adjust to the high light levels used in the chamber. This could cause the net growth rate to be lower than predicted and could explain the lack of fit in Figure 3.

To test the validity of the model for predicting beet growth excluding the presumed acclimation phase, it was attempted to identify and validate the model on the data collected from day 5 to the end of the experiment. The results of the validation are shown in Figure 4.

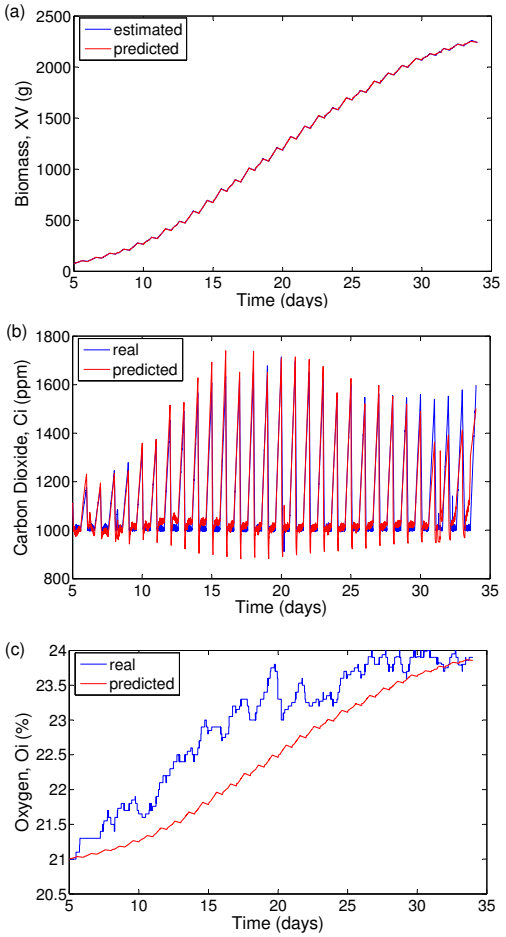


Fig. 4. Validation of kinetic model identification on day 5 - harvest beet data, showing predicted and real (or estimated) (a) biomass (b) $CO_2$ and (c) $O_2$ versus time.

The predicted carbon dioxide concentration fit the real data much better in this case suggesting that an acclimation phase may indeed be required. Figure 4 also shows that the biomass prediction was successful but that oxygen production was underestimated throughout the experiment. This is not a problem with the yield, since the predicted concentration matches the true value at the end of the experiment. Instead, the prediction error may suggest that there is some unconsidered biological phenomena affecting oxygen production. This discrepancy should be further investigated.

The model was validated on a separate beet experiment not used for identification of the kinetic parameters in order to further test the model's validity. The experiment was conducted under the same environmental conditions as the identification data set. The results of the validation are shown in Figure 5.

The model is again, a good fit for the biomass and carbon dioxide data, which reinforces it's validity for predicting

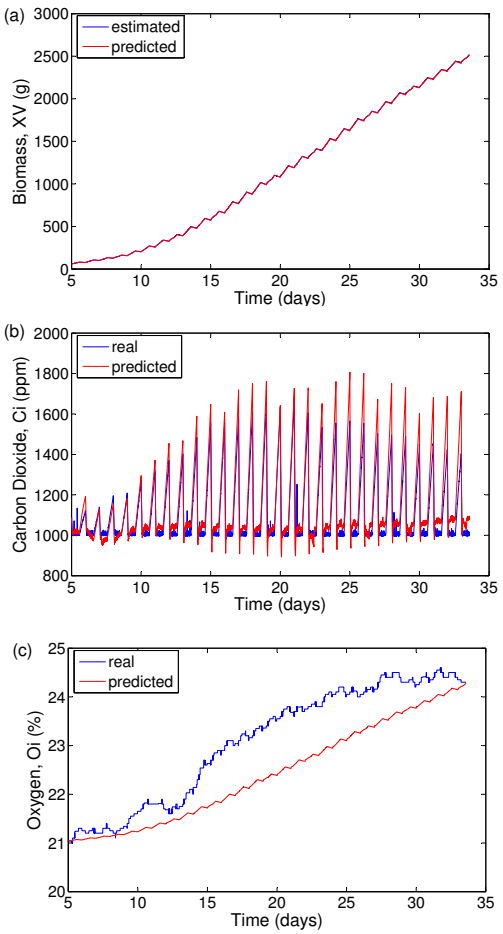


Fig. 5. Validation of kinetic model on an independent beet data set (day 5 - harvest), showing predicted and real (or estimated) (a) biomass (b) $CO_2$ and (c) $O_2$ versus time.

beet growth excluding an acclimation phase. However, the underestimation of oxygen production in the chamber remains.

## 6. CONCLUSIONS

We have developed a simple, general model for plant growth considering photosynthesis and respiration reactions. The model is adequate for predicting lettuce growth, and predicts biomass and $CO_2$ development quite well for beets excluding an acclimation phase at the beginning of the experiments. There remains some error in the oxygen prediction for beets. Further work is needed to determine the source of this error and to improve the model. It is possible that important biological processes have not been fully considered.

The next step in the model development will be to consider a more detailed approach in which the development of different plant parts (organs) is taken into account. In addition, the model should be expanded to include other important processes, such as transpiration and root water uptake. Finally, the model should be tested on different plants and under different environmental conditions to test reliability and to improve and refine the model.

**Acknowledgements :** Technical support for Guelph provided by Danuta Gidzinski. This research was funded by the European Space Agency. This paper presents research results of the Belgian Programme on Inter-University Poles of Attraction initiated by the Belgian State, Prime Minister's office for Science, Technology and Culture. The scientific responsibility rests with its authors.

## REFERENCES

[1] Thornley J.H.M and I.R. Johnson (2000). *Plant and Crop Modelling - A Mathematical Approach to Plant and Crop Physiology*. The Blackburn Press, Caldwell, NJ.

[2] Zhu X-G., de Sturler, E., and S.P. Long (2007). Optimizing the Distribution of Resources between Enzymes of Carbon Metabolism Can Dramatically Increase Photosynthetic Rate: A Numerical Simulation Using an Evolutionary Algorithm, *Plant Physiology*, Vol. 145, pp 513-526.

[3] Laisk A. and G.E. Edwards (2000). A mathematical model of $C_4$ photosynthesis: The mechanism of concentrating $CO_2$ in NADP-malic enzyme type species, *Photosynthesis Research*, Vol. 66, pp 199-224.

[4] Heuvelink, E. (1999). Evaluation of a Dynamic Simulation Model for Tomato Crop Growth and Development, *Annals of Botany*, Vol. 83, pp 413-422.

[5] Marcelis, L.F.M., Heuvelink, E., and J. Goudriaan (1998). Modelling biomass production and yield of horticultural crops: a review, *Scientia Horticulturae*, Vol. 74, pp 83-111.

[6] Cournède P.H., Kang M.Z., Mathieu A., Barczi J.F., Yan H.P., Hu B.G., and P. de Reffye (2006). Structural Factorization of Plants to Compute Their Functional adn Architectural Growth, *Simulation*, Vol. 82, pp 427-438.

[7] Allen M.T., Prusinkiewicz P., and T.M. DeJong (2005). Using L-systems for modeling source-sink interactions, architecture and physiology of growing trees: the L-PEACH model, *New Phytologist*, Vol. 166, pp 869-880.

[8] Wong S.C., Cowan I.R., and G.D. Farquhar (1979). Stomatal conductance correlates with photosynthetic capacity, *Nature*, Vol. 282, pp 424-426.

[9] Chen, L. (1996). Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown. *Mathematical Biosciences*, Vol. 132, pp 35-67.

[10] Cannell M.G.R. and J.H.M. Thornley (2000). Modelling the Components of Plant Respiration: Some Guiding Principles, *Annals of Botany*, Vol. 85, pp 45-54.

[11] McCree, K.J. (1970). An equation for the respiration of white clover plants grown under controlled conditions. In: Setlik I, ed. *Prediction and measurement of photosynthetic productivity*. Wageningen: Pudoc, pp 221-229.

[12] von Caemmerer S. and G.D. Farquhar (1981). Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. *Planta*, Vol. 153, pp 376-387.