M ELiSSA

Technical Note

Memorandum of Understanding 19071/05/NL/CP

MELiSSA FOOD CHARACTERIZATION: PHASE 1

TECHNICAL NOTE: 98.3.21

REVIEW OF MODELLING ISSUES RELATED TO HIGHER PLANT METABOLISM, IDENTIFICATION OF CRITICAL POINTS AND PROPOSED METHOD

prepared by/*préparé par*	Heather Maclean, Denis Dochain
reference/*reference*	Contract number
issue/*edition*	1
revision/*révision*	1
date of issue/*date d'édition*	10/01/2011
status/*état*	Final
Document type/*type de document*	Technical Note
Distribution/*distribution*	

CONFIDENTIAL DOCUMENT

APPROVAL

Title *titre*	Review of modelling issues related to higher plant metabolism, identification of critical points and proposed method	issue *issue*	1	revision *revision*	1

Author *auteur*	Heather Maclean, Denis Dochain	date *date*	07/01/2011

Reviewed by Consultant UBP	Pauline Hezard	Date *date*	27/05/2010 02/06/2010
Reviewed by (UGent) approved by (UGent) *approuvé by*	Laury Chaerle and Dominique Van Der Straeten		28/05/2010 03/06/2010

CHANGE LOG

reason for change /*raison du changement*	issue/*issue*	revision/*revision*	date/*date*

CHANGE RECORD

Issue: 1 Revision: 1

reason for change/*raison du changement*	page(s)/*page(s)*	paragraph(s)/*paragraph(s)*

TABLE OF CONTENTS

Table of Figures v

List of Tables v

List of Abbreviations vi

1 Abstract *1*

2 Model Objectives & Requirements *1*

2.1 Overall Model Objectives 1

2.2 Model Requirements 2

2.3 Objectives for a First Simple Model 2

3 Modelling Approach & Structure *2*

3.1 Approach 2

3.2 Structure 3

4 Data Sets Required versus Selected *4*

4.1 Data Expected in an Ideal Case 4

4.2 Experiments Required 7

4.3 Data Availability 10

4.4 Critical Analysis – Available versus Required Data 11

4.5 Data Sets for First Model Selection 12

5 Selection of the First Simple Model *12*

5.1 Initial Model Selection 12

5.2 Model Identifiability and Reformulation 14

5.3 A Stoichiometric Approach 16

6 Yield Identification *16*

6.1 Separating the yield identification from the identification of kinetic parameters 16

6.2 Identification of the Yields 17

7 Kinetic Model Identification *19*

7.1 Data Required for Kinetic Model Identification 19

7.2 Kinetic Model Selection 20

7.2.1 Selection of Mitochondrial Respiration Equation 20
7.2.2 Selection of Photosynthesis and Photorespiration Equations 23

7.3 Full Kinetic Model Identification and Validation on Lettuce Data 23

7.4 Identification and Validation on Table Beet Data 26

7.5 Comparison of Lettuce and Beet Models 29

8 Conclusions 30

9 Proposal of Future Work 30

10 References 34

11 Appendix A – Theoretical Yield Calculations 35

12 Appendix B – Implementation of the Model 36

12.1 Source code for yield identification and validation 36
12.1.1 M-file - yieldest.m 36
12.1.2 M-file – yield_regression.m 39

12.2 Source code for kinetic model identification and validation 41
12.2.1 M-file – call_plant_model.m 42
12.2.2 M-file – initialize_parameters.m 43
12.2.3 M-file – calculate_XV.m 45
12.2.4 M-file – ident_and_valid.m 45
12.2.5 M-file – SSE.m 46
12.2.6 M-file – validate.m 49

Table of Figures

Fig. 1 Flow chart depicting relevant plant growth processes 13
Fig. 2 Evolution of carbon dioxide consumed / oxygen produced with time. 19
Fig. 3 Validation of Monod mitochondrial respiration equation on lettuce night data 21
Fig. 4 Validation of growth + maintenance respiration equation on lettuce night data.... 22
Fig. 5 Validation of kinetic model on lettuce experimental data 24
Fig. 6 Validation of kinetic model with ‘corrected’ yield on lettuce experimental data .. 25
Fig. 7 Validation of kinetic model on separate lettuce data set 25
Fig. 8 Validation of kinetic model on full beet experimental data 26
Fig. 9 Validation of kinetic model on day 5 to harvest beet data 27
Fig. 10 Validation of kinetic model on separate beet experiment (day 5 to harvest) 28
Fig. 11 Flow chart depicting model structure for yield identification. 36
Fig. 12 Flow chart depicting model structure for kinetic model identification. 42

List of Tables

Tab. 1 Results of the yield identification 18

List of Abbreviations

Symbol	Definition
α	A constant (m^3 g^{-1}) representing the inverse of the biomass density such that the volume of the plant is proportional to biomass ($V=\alpha XV$) (selected 1000 m^3 g^{-1})
C_a	Carbon dioxide concentration in the atmosphere of the plant chamber (g m^{-3})
C_i	Carbon dioxide concentration inside the leaves (g m^{-3})
IP	Incident photon flux (µmol PAR m^{-2} s^{-1})
k	Light extinction coefficient (selected 0.7)
K_C, K_O	Michaelis-menten constant for carbon dioxide, oxygen (g m^{-3}) (selected based on [14] for spinach, K_C=0.94, K_O=487)
LA	Leaf area (m^2)
LAI	Leaf area index
LAR	Leaf area ratio (m^2 g^{-1}), a ratio of leaf area to dry mass (calculated - lettuce: 0.016 m^2 g^{-1}, beets: 0.0045 m^2 g^{-1})
O_a	Oxygen concentration in the atmosphere of the plant chamber (g m^{-3})
O_i	Oxygen concentration inside the leaves (g m^{-3})
PA	Planting area (m^2) (5 m^2)
PAR	Photosynthetically active radiation
PPWG	Plant Physiology Working Group
PQ	Photosynthetic quotient (mol O_2 produced mol / CO_2 consumed)
r	Net photosynthetic rate (g s^{-1})
r_{dc}	Diffusion resistance parameter for carbon dioxide (s)
r_{do}	Diffusion resistance parameter for oxygen (s)
r_i	Rate equation(s) (g s^{-1})
r_{mr}	Rate of mitochondrial respiration (g s^{-1})
r_{pr}	Rate of photorespiration (g s^{-1})
r_{ps}	Rate of photosynthesis (g s^{-1})
$r_{ps,\ daily\ avg}$ $r_{pr,\ daily\ avg}$	Average value of r_{ps}, r_{ps} over the previous day (g s^{-1})
t	Time (s)
u_1	Rate of CO_2 addition to the chamber for control (g s^{-1})
V	Volume of plant (m^3)
$V_{chamber}$	Volume of the plant chamber (29 m^3)
v_i	Rate constants (units vary) (identified in kinetic parameter identification)
XV	Biomass (g)
$\frac{\Delta XV_{1d}}{\Delta t_{1d}}$	Average rate of biomass production over the previous day (g s^{-1}), used as an estimate of ($r_{ps,\ daily\ avg}$ - $r_{pr,\ daily\ avg}$)
Y_i	Yields (g g^{-1}) (identified in yield identification)
z_1, z_2	Set of states used for separation of yield and kinetic parameter identification

1 Abstract

This document outlines work done on the modelling of higher plants during PHASE 1 of the MELiSSA food characterization project. Model structure and approach were chosen based on the objectives, requirements, and intended use of the model. Data requirements and availability were reviewed. The model was developed using a mass balance approach. Photosynthesis and respiration were selected as key reactions for biomass production. Reaction kinetics were chosen based on plant physiology and standard biochemical reaction knowledge. Model identification and validation were performed using closed environment crop growth data (lettuce and beet) from the University of Guelph. The model was mostly satisfactory for describing lettuce and beet growth, although some problems remain with the oxygen production prediction, and in the description of early beet development.

2 Model Objectives & Requirements

2.1 Overall Model Objectives

A general model for plant growth will be developed. It is intended that the final version of the model be used for the prediction and control of the higher plant growth chamber. The main control objective will be to provide the desired 'flow' (according to the mission scenario) of edible biomass from the plant chamber as food for the crew. The air revitalization and water purification functions associated with the total plant biomass production are also included as important chamber goals. The integration of the higher plant chamber into the MELiSSA loop is tightly connected to these parameters, as well as to the inedible plant biomass (waste) mass flow. Therefore fluxes of all nutrients and end products, whether solid, liquid or gaseous should be predicted. Specifically, the following objectives should be met:

- The model will represent plant growth in response to environmental variables.
    - Environmental parameters of interest include air temperature, hydroponic solution temperature, total and partial pressures of the main gases (water, oxygen, carbon dioxide and ethylene), light intensity, photoperiod, and nutrient concentration of the hydroponic solution.
- The model will be applicable over a wide range of environmental conditions, including extreme cases.
- The model should be adaptable for a variety of plants.
- The model will be suitable for controlling biomass production, both in terms of production rate (total and edible biomass) and associated nutritional quality.
- The model will be designed in such a way so that it can be integrated into the MELiSSA loop.
- The model should be 'updatable', so that its simple structure can be used to test new submodels for plant processes or functioning.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

2.2 Model Requirements

In order to meet these objectives, the following conditions are required:

- The model must be validated on a variety of plants. Plants chosen will depend on availability of data, but should preferably include species from different 'plant groups' (leafy plants, tuber plants, fruit plants, grain plants, etc).
- The model must be validated under a variety of environmental conditions. This should include complete cultivations at different environmental setpoints, as well as experiments with changing environmental conditions (simulated failures or other changes in environmental parameter(s), step changes in set points preferred).
- Mechanistic/explanatory descriptions will be favoured over empiricism. The model will be based on a mass balance approach, and an energy balance will be taken into account at the plant and plant chamber level.
- The model shall predict the evolution of the following over the plant life cycle:
  - Total and edible biomass production (fresh and dry weight).
  - Gas and liquid mass fluxes of carbon dioxide, oxygen, water, and nutrients.
  - In the long term: Composition of edible biomass (elemental compositions – *C, H, O, N, etc.*, nutritional compounds – *proteins, carbohydrates, lipids, etc.*). In the short term, this will need to be based on available elemental composition yield values from plant production experiments (not predicted based on a mechanistic modelling approach that considers reaction kinetics dependent on environmental conditions).
- A model structure shall be chosen to facilitate straightforward model modification and easy integration into the MELiSSA loop.
- The model shall be built on a simple basis in a first approach, with greater detail to be added as needed in later iterations.

2.3 Objectives for a First Simple Model

A simple first model should be chosen based on the reaction kinetics of key reactions. The model structure should be selected based on the objectives and intended use of the model, and also to ensure the identifiability of parameters based on available data. Validation on experimental data is required.

3 Modelling Approach & Structure

3.1 Approach

The intended use of a model is a very important consideration in selecting an appropriate modelling approach. Models used in agriculture for predicting yield and quality of crops are often semi-empirical, and apply fitted functions without considering the biological mechanisms underlying plant growth. These models have the benefit of simplicity, and can work well for simple applications. However, they are not mechanistic and therefore cannot be

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

applied 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, these models are typically over parameterized and unidentifiable, and therefore cannot be applied for control purposes.

Process based models and functional-structural models are more applicable for control. They attempt to consider a more mechanistic approach while maintaining relative simplicity. This is advantageous, as it is known that simple models are better for the development of efficient controllers. 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]. One drawback of many of these models is that they have been developed for plant growth under field conditions, and they therefore often neglect the effect of some important environmental variables (for example CO_2 and O_2 concentration). Classical mass balance models, which should fit under the umbrella of process based models, will take these factors into account. Mass balance models are aimed at capturing the main dynamical features of plant growth by considering conversion mechanisms for the most important reactions as well as mass transfer phenomena.

The model for the MELiSSA project must be applicable over a wide range of conditions, but must also be simple enough to use for control. Therefore a process based model is appropriate for this application. A mass balance approach based on the reaction kinetics of the most important plant growth processes for biomass production has been selected for its simplicity and mechanistic treatment of plant growth. This is a standard approach for chemical and biochemical systems, and should also be appropriate for plant growth systems. Environmental variables will act as inputs to the model, and should allow for prediction over a wide range of environmental conditions. The model will be developed from a first simple model towards a more detailed treatment. Focus will be placed on maintaining a mathematically unique model (all parameters identifiable) that can be validated on experimental data.

3.2 Structure

Generally, model structure can be characterized as either using a global approach, or as relying on a compartmentalized approach. A global model is one in which the growth processes are considered at the whole plant level. Organ or edible biomass could be roughly predicted using yield values (derived from experiments in comparable conditions) defined for a limited number of developmental stages. However, the development of different plant organs would not be considered separately. The model is aimed at capturing the main dynamical features of plant growth but should otherwise remain very simple. The global model is a good option for chamber control because its relative simplicity will allow a more efficient controller to be developed. A compartmentalized approach would separate the model into boxes (or modules) associated with the different organs (roots, leaves, fruit, etc.). These boxes are then connected with the appropriate interfaces to build the complete model of the plant. This approach is more complex, but allows for prediction at the organ level.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

Given the stated objectives of the model, both of these approaches have value. A global model would be more suitable for the control of the plant growth chamber. However, a higher level of detail is needed to evaluate organ development. Therefore, the first model will use a global approach to estimate total biomass. This model will then be used as a starting point for the compartmentalized approach at a later stage.

The model will be written in MATLAB. MATLAB is well suited for this modelling task because its language is easily understood, and programs can be easily modified. The model can easily be divided into small blocks of code, which will improve model clarity.

4 Data Sets Required versus Selected

Sections 4.1 and 4.2 were prepared in collaboration with Pauline Hezard (Université Blaise Pascal).

4.1 Data Expected in an Ideal Case

An ideal list of data required for model identification and validation follows. Of course these requirements should be weighed against system capabilities. An ideal frequency of measurement is supplied; however in some cases less frequent measurement should not pose a serious problem for modelling. Frequency of ‘continuous’ online measurements has been selected as once every 5 minutes, however this should be updated based on the capabilities of the system (within the range of once every 1-10 minutes will be sufficient). It is expected that, unless otherwise stated, measurement frequencies suggested should be attainable. If it is decided at a later stage that the dynamic responses of certain variables to a step point change are of interest, more frequent sampling could be considered. When specified, measurement locations are selected based on limited knowledge of chamber set up, and therefore the most practical and informative locations should be selected.

Measurements required during chamber operation include:

Carbon dioxide data

- The CO_2 concentration inside the plant growth chamber, as accurate as possible, frequency – 1/5 minutes, measurement location(s) should be representative (EnginSoft)
- Amount of CO_2 added to the chamber, as accurate as possible, frequency – 1/5 minutes.
- If significant, dissolved CO_2, frequency – 1/5 minutes, at least near inlet of the gullies (and possibly near outlet to measure change)

Oxygen data

- Oxygen concentration in the atmosphere of the plant chamber, as accurate as possible, frequency – 1/5 minutes, measurement location(s) should be representative (EnginSoft).

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

- Oxygen removed from the atmosphere of the plant chamber (if removal possible), frequency dependent on removal strategy, if continuous – 1/5 minutes, if periodical – logged upon removal
- Dissolved O_2, frequency – 1/5 minutes, at least near inlet of the gullies (and possibly near outlet to measure change)

Evapotranspiration data

- Atmospheric relative humidity (RH) and/or vapour pressure density (VPD), as accurately as possible, frequency – 1/5 minutes, measurement location(s) should be representative (EnginSoft)
- Amount of water condensed from the atmosphere, as accurately as possible, frequency – 1/5 minutes
- Water mass or volume in chamber, as accurately as possible, may include:
    - Addition of water (from HVAC condensate or external source if needed) to nutrient tank, frequency – logged at time of water addition
    - Addition of water to atmosphere through humidifier (if used), frequency – 1/5 minutes
    - Removal of water (sampling, etc.) from nutrient solution, frequency – logged at time of water removal
- No additional experiments required to estimate evaporation from closed gullies, since this can be estimated for each experiment over a timeframe of several days without plants being present, or when plants are very small and it is proven that plant transpiration is not significant.

Trace gas data (if measurement possible with the system)

- Ethylene concentration in the atmosphere of the plant chamber, as accurately as possible, frequency – 1/hour to 1/day
- VOC concentrations in the atmosphere of the plant chamber, most critical compounds in addition to ethylene to be defined on a per crop basis, frequency – 1/day to 1/week (frequency limitation possible)
- If possible total chamber atmosphere composition using gas chromatograph or equivalent system, frequency – 1/week (frequency limitation possible), or at least as a test at vegetative maturity.

Temperature data

- Chamber air temperature, accuracy – 0.5°C minimum, frequency – 1/5 minutes, measurement location(s) should be representative (EnginSoft)
- Nutrient solution temperature, accuracy – 0.5°C minimum, frequency – 1/5 minutes, at least near inlet of the gullies (and possibly near outlet to measure change).
- Gully air temperature, accuracy – 0.5°C minimum, frequency – 1/5 minutes

Nutrient solution data

- Nutrients added (from concentrated replenishment solutions) to nutrient solution, as accurately as possible, frequency – logged at time of addition
- Nutrient solution composition, frequency – a minimum of 1/week, and also each time solution is discarded or replaced (using HPLC or ion chromatography)

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

- pH and electroconductivity (EC), as accurately as possible, frequency – 1/5 minutes, at least near inlet of the gullies (and possibly near outlet to measure change)
- *Light data*
- Light intensity, frequency – 1/5 minutes around canopy height from at least 2 sensors (1 positioned above expected canopy height so that it does not become covered by plants)
- Study of light intensity during the light/dark transitions (just after turning lights on and off) – light intensity measurements at a frequency of several points per minute (every 5 to 20 seconds) until reaching equilibrium. This should be done at least once in each experiment (once for turning lights on, once for turning lights off).

Gas flow rate data

- Gas flow rate, frequency – at least 1/week in different parts of the chamber. If possible, more frequently (1/5 minutes) (data from EnginSoft may be sufficient).

Biomass data

Biomass measurements should be taken at the seedling stage (at time of transplantation to chamber), at harvest, and at intermediate time points according to a certain frequency of measurement. The frequency of biomass measurements should depend on length of experiment and organ development. The frequency should therefore be determined on a species specific basis. This should be discussed with plant physiology working group.

- The following measurements should be made for all biomass samples:
  - Biomass fresh and dry weight – total and for each organ separately
  - Biomass composition for each organ
    - Nutritional compounds – This should include at least carbohydrate, protein, and fat content; total fibre and ash content is also useful and usually available (proximate composition determination). If possible composition with respect to different types of sugars (mono and polysaccharides), amino acids, lipids, nucleic acids, fibres, etc.
    - Elemental composition – At least C, N, P, K, and S. If possible H, O, Ca, and Mg. Mineral element composition can be determined from ash.
  - Biomass sizes – leaf area, stem and root length and diameter, branching patterns, number of buds, fruit size
  - Growth evolution (number of internodes, mean canopy height and diameter)
- Frequency examples (further discussion needed – depends on system capabilities): (frequency limitations expected)
  - Potatoes – 1/week before tuberization, 2/week after
  - Wheat – 1/week before heading, 2/week after
  - Soybean – 1/week before pod development, 2/week after
  - Leafy plants (Lettuce, Spinach, Cabbage) – 1/week
  - Onion – 1/week before bulb formation, 2/week after
  - Rice – 1/week before flowering, 2/week after
  - Tomato – 1/week before flowering, 2/week after

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

- Development steps (vegetative and reproductive development) should be noted with a date and approximate time (if possible a picture would be useful)

Note that biomass measurements could be supplemented with data from image processing software if this became available in the future. Depending on the accuracy of this technology, frequency of destructive measurements could be decreased/negotiated. It is known that biomass sampling will affect the fluid dynamics (both airflow and liquid flux through root zone) of the system and the measurement of gas exchange in the chamber (especially if many plants are removed). If proven necessary by the deviations caused on these parameters recorded in preliminary experiments, a second test could be performed with exactly the same conditions but without opening the chamber for sampling. This would allow us to obtain accurate gas measurements and keep a constant environment for the canopies.

A note on elemental mass balances:
Data to be used for the elemental mass balances will be available from nutrient uptake data, biomass composition data, and gas flux information (in the case of C, O and possibly H balances). From the "Nutrient Solution Data" outlined above, approximate net uptake rates of various elements can be calculated. Organ specific biomass composition data (measurements taken at frequencies suggested above) will be used to calculate the storage of these elements in biomass (and partitioning to different organs). Finally gas flux information is available for carbon dioxide and oxygen data. Trace gas measurements could also be used, although it is expected that this will be a relatively small pool. These measurements should provide sufficient data to complete a set of simple element balances. As stated in the Model Requirements (2.2), modelling of these components will need likely be based on yield values from plant production experiments, rather than a mechanistic approach, at least in the short term. The composition of more important elements could be linked to environmental conditions at a later stage.

Additional tests required before/after experiments:

- Gas leakage test for normal operation (CO_2, O_2 and humidity could be recorded)
- Sampling leakage test for CO_2 (and possibly O_2 and humidity) – estimate leakage during typical procedure to open door for sampling.
- Light spectrum – to be performed at beginning and end of experiment (or at least between experiments).

4.2 Experiments Required

The plant growth model should be adaptable to different types of plants by adjusting a few key parameters and constants. More realistically, certain sub-models may need to be added or removed from the overall model structure in order to apply it to very different types of plants. In any case, experimental data from a variety of crops is needed. These should include species from different crop types (leafy crops, grain crops, storage root crops, fruit crops, etc.), with at least 3 and preferably more experiments for each crop.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

Two cultivation methods have been proposed for plant growth experiments – batch and staggered cultivation. In batch experiments, all seedlings are transferred to the plant growth chamber at one time and the plants are all harvested at the same time, typically once they reach maturity. In staggered experiments, the seedlings are transferred to the plant growth chamber at a certain planting interval, so that at any time in the chamber there are plants of various ages growing. There are advantages and disadvantages to both types of experiments. For example, a closed system can be maintained for batch experiments, but not for staggered growth experiments since the chamber must be opened at a certain interval for harvest and for transfer of new seedlings. A closed chamber is valuable because it allows us to close the balance on important species in the chamber, such as CO_2, O_2 and water. It may also be more difficult to determine the effect of plant age on the gas exchange or nutrient uptake in the chamber for a staggered growth experiment, since plants of different ages are present in the chamber at any given time and it is impossible to know the contribution of the different age groups.

From a modelling perspective, the main potential benefit of using a staggered growth approach would be that it provides some information on biomass production and organ development with time. During the experiment, plants are harvested after a certain number of days (selected so that the plants can reach maturity) in the chamber. However, at the end of the experiment all plants remaining in the chamber are harvested. These plants are of different ages, and so this approach provides some useful data on biomass production, leaf area development, and organ development with time. Another benefit is that it provides an additional validation step, to show that the model can be applied to different experimental set-ups. This approach could also be potentially useful for conducting experiments in which it is attempted to test the effect of an environmental condition on growth. With this approach, it would be possible to test several set points in a single experiment since cultivation is continuous.

Both types of experiments, batch and staggered, could have value to the modelling work. Batch experiments are important for getting a complete picture of the gas exchange in the chamber and should be conducted as the primary experiment. Staggered growth experiments should also provide useful information, and have the added benefit of providing limited information on biomass production and organ development with time. Staggered growth experiments are not, however, the best method of obtaining biomass and organ development data from a modelling point of view. The staggered approach provides a lot of data on fully matured plants while only a few plants are harvested at the other time points. In addition, the conditions experienced during the growth of the plants harvested at day 10 are not exactly the same as the conditions during the first 10 days for all the other plants, and as stated previously, it is difficult to link gas exchange or nutrient uptake to plant age. It would be better (from the perspective of gathering biomass data) to start all plants at the same time and harvest a subset at selected sampling times. However, it should be noted that while batch experimental data has been worked with extensively in this work, the data from staggered experiments have not yet been used. Therefore, it is possible that we have not fully considered all possible benefits

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

and drawbacks to using a staggered approach at this time. This should be included in future work (Section 9).

To ensure model robustness to a variety of environmental factors, the plant model must be validated on data sets from experiments with different environmental conditions. This could include experiments at different setpoints (e.g. 3 experiments using 3 different CO_2 concentration setpoints) as well as experiments in which there are changing environmental conditions (step change(s) in setpoints). The range of conditions used in setpoint experiments should be within an acceptable range for plant growth and development. These ranges should be further discussed with the plant physiology working group. Experiments with changing environmental conditions could be designed to best represent expected failures in the system.

Priority should be given to experiments which vary the most important environmental variables (light, temperature, CO_2). The relative importance of environmental factors should be determined based on the sensitivity of plant growth to that factor, the likelihood of variation due to failure, and/or the likelihood of a setpoint change to support the control needs for MELiSSA integration.

Further discussion will be required before any specific recommendations can be made. However, a brief outline of suggested environmental experiments will be given below to give an idea of the types of experiments that should be considered:

Set-point experiments (long term – growth period – experiments, per species):

- Effect of CO_2 content (2 set points minimum)
- Effect of light intensity (2 set points minimum)
- Effect of humidity (2 set points minimum)

Effect of temperature (3 set points) – Day and night temperatures can have different setpoints. Priorities for testing the effects of temperature (and of the difference between day and night temperature) to be further discussed within the plant physiology working group.

Changing environmental conditions (possibility of short term experiments – at least long enough to recognize changes in biomass composition or partitioning, per species):

Experiments to be chosen based on expected scenarios in a space setting; the following is a list of some initial ideas.

- Step change in CO_2 set-point experiment (CO_2 drops to 'atmospheric' concentration of the environment in which the chamber is located (simulating a leak or chamber opening))
- Step change in light intensity set-point experiment (A step change from light to dark conditions lasting over longer time period than typical night (simulating lighting failure))

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

- Step change in air temperature set-point experiment (step increase in temperature to simulate failure in cooling mechanism if not already prevented by safety mechanisms)
- Loss of nutrient solution circulation for short periods (Test effect of water stress - shut down nutrient solution circulation for time periods thought to be tolerable by plants). Test of intermittent irrigation (5 min on / 5 min off and draining) might be useful.

4.3 Data Availability

The most useful data currently available is from closed chamber experiments at the University of Guelph. In these experiments the plants are grown in a closed environment with controlled atmospheric conditions. Lettuce and red beet experiments have been performed in these chambers and are the most complete data sets available at this time. Two types of experiments have been conducted with these plants – batch and staged growth experiments. In batch experiments (lettuce and beet) all plants in the chamber are the same age, and are harvested at the same time. In staged-growth experiments (beet), seedlings were transferred to the plant growth chamber at 10 day intervals. Therefore plants of different ages occupy the chamber at the same time. Currently, data from batch experiments is being used exclusively for modelling. However, the staged growth experiments should provide some useful information about development in the future.

The following data was available for lettuce and beet experiments:

Available temporal data (every 6 minutes) includes:

- The CO_2 concentration inside the plant growth chamber
- Amount of CO_2 added to the chamber
- Oxygen concentration in the atmosphere of the plant (data recording alternating ON for 6 hours, OFF for 6 hours since shared by 2 identical chambers)
- Atmospheric relative humidity
- Amount of water condensed from the atmosphere
- Chamber air temperature
- pH and EC
- Light intensity

Other data available:

- Total fresh and dry weight of biomass at seedling stage (before transplantation to the chamber) and at harvest.
- Fresh and dry weight of each organ at harvest.
- Leaf area at seedling stage and at harvest.
- Carbon dioxide leakage rate

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

- Nutrient solution composition every 5 days (nutrient solution was replaced every 5 days and composition was measured at the beginning and end of each 5 day period).
- Composition of harvested biomass (total and for each organ) with respect to certain elements (C, N, P, K, Ca, Mg). H, O, S data missing, as is nutritional compound data.

4.4 Critical Analysis – Available versus Required Data

The Guelph closed growth chamber data is very appropriate for the model identification and validation. Much of the data listed in Section 4.1 is available in the lettuce and beet data sets. The most important measurements for the calculation of carbon dioxide flux, oxygen flux, and evapotranspiration rates are available, as well as important environmental variables such as CO_2 and O_2 concentration, temperature, nutrient solution concentration, and light intensity. A few measurements listed under these headings were not measured; however, enough data is available to give a good estimate. For example, for carbon dioxide, we have measurements of CO_2 concentration, CO_2 added to the chamber and an estimation of the leakage rate. The dissolved CO_2 in the nutrient solution is not measured, but this could be estimated based on temperature and is, in any case, a minor factor that should not greatly affect the modelling effort. Therefore the available data will be very useful for the first model selection.

However, there are areas in which the available data is not adequate to build the ideal model. Light interception is a key factor in plant growth and it is known that leaf area largely determines this variable. The available data sets only contain measurements of leaf area at the beginning and end of the experiment. This will be a limiting factor in the model development (especially in the case of crops for which biomass (or CO_2 uptake) is not strictly proportional to leaf area development, as it is for leafy vegetables). Biomass data is also only available at the seedling and harvest stage. The biomass production rate throughout growth can be estimated from the carbon dioxide uptake rate assuming a constant yield (and therefore also constant elemental composition). This is adequate for total biomass production. However, the final model should have a compartmentalized approach so that organ development can be predicted. It will not be possible to validate such a model without data on the organ development with time (organ biomass measurements with time would be best, but more easily attainable architectural data could also be related to biomass, for example fruit size). In addition, while some compositional biomass data is available, it would be useful to have measurements with time, and also information about nutritional compounds, in order to meet model requirements.

Another significant problem with the available data is the limited number of useful data sets. Currently there are only complete data sets from a few plants available (lettuce, red beet, and durum wheat from a first experiment). Data from a wider variety of plants (e.g. MELiSSA crop list) would be useful to make sure the model is general for plant growth. In addition, the data available is at a single set of environmental conditions for each plant. These conditions were likely chosen because they represent the best growth conditions for the plants. However,

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

in order to make the model robust to changes in environmental conditions, a wider variety of experiments are needed.

4.5 Data Sets for First Model Selection

Data from batch lettuce and beet experiments were suitable for the first model selection, and were therefore used for model identification and validation. The data sets used in various stages of the identification and validation were:

Lettuce data: GW0704, GW0604, GW0204
Beet data: GW0404, GW0504.

All of the above experiments were performed in batch cultivation. Thus far, data from staggered growth experiments has not been used.
The cultivation conditions are as shown below. They are approximately the same in all experiments, with some differences due to experimental error.

Plants in chamber	- 120
Temperature	- 25°C Day / 20°C Night
Photoperiod	- 14h light / 10h dark
Relative humidity	- 70%
CO_2 concentration	- 1000ppm
Incident PAR	- ~400-500 µmol PAR m^{-2} s^{-1} during day (at canopy height, measured by sensors)

Nutrient solution composition – 1.5 mM PO_4^{3-}, 3.62 mM Ca^{2+}, 4 mM NH_4^+-N, 11.75 mM NO_3^--N, 5 mM K^+, 2 mM SO_4^{2-}, 1 mM Mg^{2+}, 0.005 mM Mn^{2+}, 0.025 mM Fe^{3+}, 0.0035 mM Zn^{2+}, 0.02 mM B^{3+}, 0.008 mM Na^+, 0.0008 mM Cu^{2+}, 0.0005 mM Mo^{6+}. Nutrient solution is replaced approximately every week.

5 Selection of the First Simple Model

5.1 Initial Model Selection

The literature was reviewed for existing plant growth models and for other relevant modelling and identification strategies. In addition, previous work done for the MELiSSA project was reviewed. Based on the information gathered, and from a general knowledge of plant physiology, an outline of relevant plant growth processes and reactions was developed (Fig. 1).

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

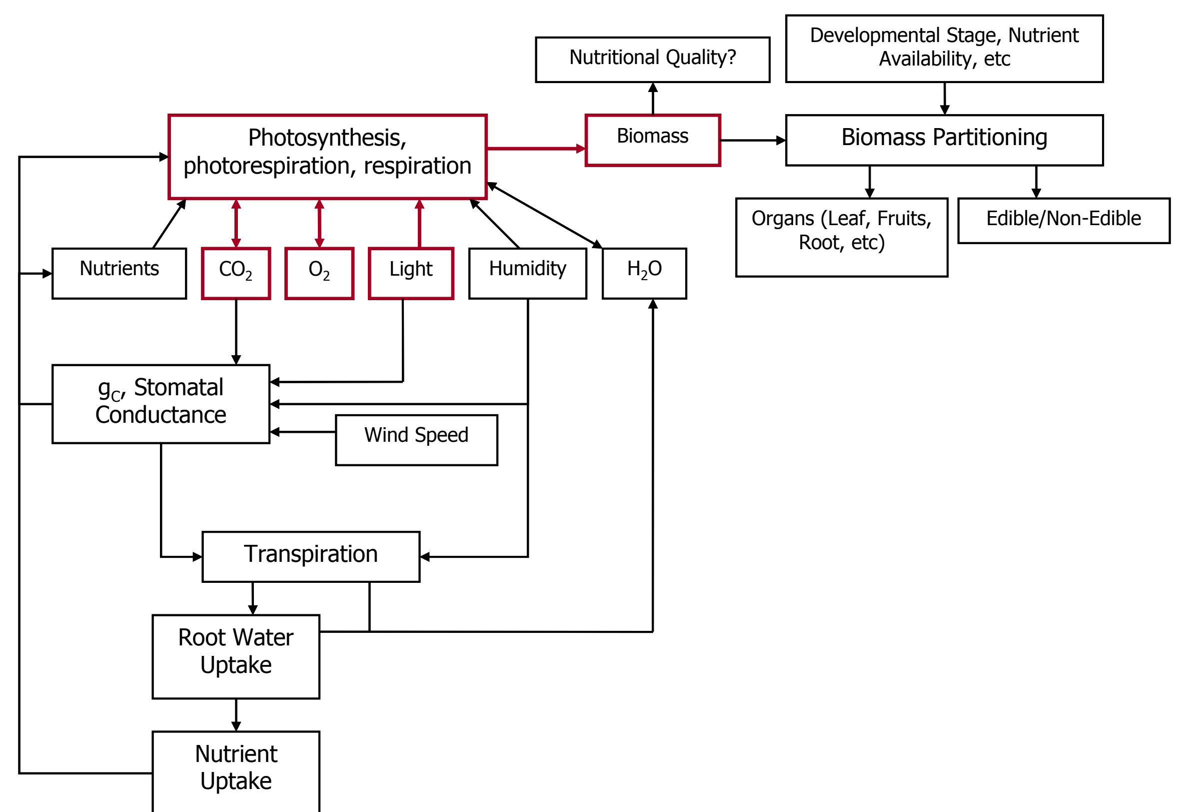

Fig. 1 Flow chart depicting relevant plant growth processes. In red are the selected processes and factors for the first model identification.

The processes and factors considered in the current version of the model are highlighted in red in Fig. 1. Photosynthesis, photorespiration, and mitochondrial (or 'dark') respiration were selected as the reactions that should be considered in the first model. This selection was made because these are the main reactions that influence the production of biomass, as well as the exchange of carbon dioxide and oxygen. At this stage in the model development, only total biomass production will be considered, and therefore biomass partitioning is not included in the model. In addition, it was assumed that water and nutrients are at sufficient levels so as not to influence the rate of photosynthesis. Therefore, for simplicity, in the first model selection we have neglected transpiration and root water uptake. These processes should be included in a later version of the model to complete the water and nutrient balances and to extend the model's applicability to drought and nutrient poor conditions that may result from failures.

Based on these assumptions, a simple mass balance model was derived for photosynthetic biomass production. The general reaction scheme is shown as (1) - (3), where (1) is photosynthesis, (2) is photorespiration and (3) is mitochondrial respiration.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

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

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

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

From the general reaction scheme ((1) – (3)), and considering the transfer of gases to and from the plant, a mass balance model ((4) - (6)) was derived for biomass (XV, g), carbon dioxide concentration in the leaf (C_i, g m^{-3}), and oxygen concentration in the leaf (O_i, g m^{-3}). Carbon dioxide and oxygen measurements are typically given in concentrations, while biomass is measured as a mass. The differential equations were therefore formulated with the states in their appropriate forms. No data is generally available on the volume of the plants throughout growth, therefore the assumption was made that volume should be approximately proportional to the biomass ($V=\alpha XV$). This assumption was used to eliminate a volume differential equation that was proportional to the biomass balance.

$$\frac{dXV}{dt} = Y_1 r_1 - Y_2 r_2 - Y_3 r_3 \quad (4)$$

$$\frac{dC_i}{dt} = \frac{-r_1 + r_2 + r_3}{\alpha XV} + \left(\frac{C_a - C_i}{r_{dc}}\right) - \frac{C_i}{XV}\left(Y_1 r_1 - Y_2 r_2 - Y_3 r_3\right) \quad (5)$$

$$\frac{dO_i}{dt} = Y_4 r_1 - Y_5 r_2 - Y_6 r_3 + \left(\frac{O_a - O_i}{r_{do}}\right) - O_i \alpha\left(Y_1 r_1 - Y_2 r_2 - Y_3 r_3\right) \quad (6)$$

In these equations XV is biomass (g), C_i and O_i are CO_2 and O_2 concentrations in the leaves (g m^{-3}), C_a and O_a are CO_2 and O_2 concentrations in the atmosphere of the chamber (g m^{-3}), r_i are the rate equations to be defined (g s^{-1}), Y_i are the yields (g g^{-1}), r_{dc} and r_{do} are diffusion resistance parameters for carbon dioxide and oxygen respectively (s), and α is a constant (m^3 g^{-1}) representing the inverse of the biomass density such that the volume of the plant is proportional to the biomass ($V=\alpha XV$).

5.2 Model Identifiability and Reformulation

The question of the identifiability of a model asks: On the basis of the structure of the model and the quality of the data available can a unique value be given to each of the model parameters [8]? To determine the answer to this question the structural and practical identifiability should be evaluated. A model is structurally identifiable if, given ideal measurements, all of the parameters can be identified uniquely. Therefore, this is based only on the structure of the model and not on the data provided. Conversely, practical identifiability examines whether we have the appropriate data to identify all of the parameters.

There are several tests for evaluating the structural identifiability of a model. The Taylor series method was used in this model development. This method makes use of successive

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

derivatives of the differential equations to gain information about the parameters. For a limited number of derivatives, the system of equations should be solvable. Based on this analysis and also on an examination of the structure of the model, it was found that in its current form the model is not structurally identifiable. Even if some parameters (resistances, α) can be selected based on our knowledge of plant growth, this still leaves 6 yield parameters and at least 3 kinetic parameters associated with the 3 rate equations, which can not all be identified uniquely.

Therefore, to ensure the structural identifiability, the model was reformulated. 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 problem. The model was reformulated to treat photosynthesis and photorespiration and mitochondrial respiration as a single stoichiometrically reversible reaction. The reversible reaction will generally proceed in the forward direction during the day (CO_2 consumed, O_2 produced) and in the reverse direction at night (CO_2 produced, O_2 consumed). By reformulating the model in this way, we are making the assumption that the stoichiometry is the same for all three reactions. Practically, it is unlikely that the stoichiometry will be exactly the same, since the reactions take different pathways. However, for a first model selection, we will assume that the stoichiometry is close enough.

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

$$CO_2 + H_2O \xleftrightarrow{Light^*} Biomass + O_2 \qquad (7)$$

* Note that a light environment is only required for photosynthesis and photorespiration reactions. This will be reflected in the rate equations. Mitochondrial respiration gives a net production of energy, although of course some activation energy (metabolic energy sources, e.g. ATP etc.) will be required.

A new mass balance model was also written to correspond to the updated reaction scheme. In writing this new set of equations, the simplifying assumption was made that the carbon dioxide and oxygen concentrations inside the leaves will be approximately equal to their concentrations in the atmosphere of the plant chamber. Therefore, instead of considering a balance on the gases inside the plant, we can consider a more simple balance on the chamber. 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 [9]. Therefore based on the updated reaction scheme and this new assumption, the mass balance equations were written as follows:

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

$$\frac{dXV}{dt} = Y_1 r \tag{8}$$

$$\frac{dC_i}{dt} = \frac{dC_a}{dt} = \frac{-r}{V_{chamber}} + \frac{u_1}{V_{chamber}} \tag{9}$$

$$\frac{dO_i}{dt} = \frac{dO_a}{dt} = \frac{Y_2 r}{V_{chamber}} \tag{10}$$

Where $V_{chamber}$ is the volume of the plant growth chamber (29 m^3), and u_1 is the rate of CO_2 addition to the chamber for control (g s^{-1}). The carbon dioxide concentration in the chamber was controlled to maintain a minimum concentration of 1000 ppm. It should also be noted that Ci and Oi could be thought of as virtual and not corresponding to the concentration of the species in a particular organelle. It is obvious that this form of the model is much simpler and will be much easier to identify. There are now only two unknown yield parameters, and one rate equation to be defined. This model was therefore used for yield and kinetic parameter identification.

5.3 A Stoichiometric Approach

In writing the aforementioned reaction schemes ((1) – (3) and (7)), no assumptions about the composition of the biomass were made. Alternatively, stoichiometric equations could be written by making some assumption about the typical biomass composition (or from elemental composition analysis of biomass). For example, considering only carbon, hydrogen, and oxygen (and therefore neglecting nitrogen, etc.) the biomass composition is often represented as $C_6H_{12}O_6$. Based on this assumption, a stoichiometric reaction (11) could be written as an alternative to the reaction shown in (7).

$$6CO_2 + 6H_2O \xleftrightarrow{Light*} C_6H_{12}O_6 + 6O_2 \tag{11}$$

The more general form of the reaction (7) will be used in the model development. However, the stoichiometric reaction (11) could be used to calculate theoretical yields. This idea will be further developed in Section 6.2, and a comparison will be made with model identification results.

6 Yield Identification

6.1 Separating the yield identification from the identification of kinetic parameters

The identification problem can be further simplified by separating the identification 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 document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

This is particularly 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 [10]. The method makes use of a state transformation based on the structure of the model (12). By applying this transformation, the initial model can be transformed into one which does not depend on the reaction kinetics (13).

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} Y_1 Vchamber \\ Y_2 \end{bmatrix} C_i + \begin{bmatrix} XV \\ O_i \end{bmatrix} \quad (12)$$

$$\frac{d}{dt}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} Y_1 u_1 \\ \dfrac{Y_2 u_1}{V_{chamber}} \end{bmatrix} \quad (13)$$

In these equations, z_1 and z_2 are the new states which are chosen (12) in such a way as to eliminate the reaction kinetics from the transformed model. Therefore, using (12) and (13), we can derive simple equations which can be used for the yield identification without any prior knowledge of the reaction rates ((14)-(15)).

$$\frac{dXV}{dt} = Y_1\left(u_1 - V_{chamber}\frac{dC_i}{dt}\right) \quad (14)$$

$$\frac{dO_i}{dt} = Y_2\left(\frac{u_1}{V_{chamber}} - \frac{dC_i}{dt}\right) \quad (15)$$

6.2 Identification of the Yields

The yield identification was performed using only initial and final measurements, since temporal data was not fully available. Based on this, all the data needed for the yield identification was either measured or could be calculated directly from lettuce experimental data. The integrated forms of (14) and (15) were used, and the yields were identified based on three available data sets using a least squares method. (For source code see Appendix B – Section 12.1)

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 Tab. 1.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

Tab. 1 Results of the yield identification

	Y_1 (g XV g CO_2^{-1})	Y_2 (g O_2 g CO_2^{-1})
Trial 1	0.631 [-0.642, 1.905]	0.363 [-0.320, 0.894]
Trial 2	0.604 [-1.229, 2.437]	0.323 [-0.468, 1.389]
Trial 3	0.532 [0.045, 1.019]	0.302 [0.181, 0.807]
'best' estimate	0.585 [0.259, 0.911]	0.329 [0.142, 0.516]
Theoretical **	0.718	0.727

** Calculations are shown in Appendix A (Section 11)

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 suggests 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 'best estimate' row of Tab. 1.

An alternative way to calculate the yields, which is sometimes used in microbiological models, is to consider the stoichiometry of the reactions. For example, if we assume that for every mole of CO_2 taken up, a mole of carbon is incorporated into biomass (as in (11)), we can calculate a theoretical yield based on the known final composition of biomass (0.38 gC gXV^{-1} from tissue analysis on three Guelph data sets). Similarly, an assumption can be made about the value of the photosynthetic quotient (moles O_2 produced/moles CO_2 consumed) to calculate a theoretical Y_2. The calculated theoretical yields are shown in the final row of Tab. 1. Calculations are shown in Appendix A (Section 11).

The theoretical yields do not match the identified yields well. The theoretical Y_2 is based on an assumption that the photosynthetic quotient (PQ) should be approximately 1 molO_2 molCO_2^{-1}, as has been reported to be fairly typical of plants [11]. However, when the photosynthetic quotient is calculated from experimental data, its value over the full experiment is approximately 0.5 molO_2 molCO_2^{-1}, which would give a theoretical yield of 0.36gO_2 gCO_2^{-1}, a value much closer to the identified value. Based on the experimental data (Fig. 2), the photosynthetic quotient does not seem to be a constant at all. Instead, the PQ is close to 1 at the beginning of the experiment, but then decreases to a smaller value when CO_2 consumption outpaces O_2 production. There are many possible reasons for this decrease in the photosynthetic quotient, including a changing composition of biomass, changing metabolism, microbial growth, or changing ratios of photosynthesis to respiration reactions (if these reactions cannot truly be considered stoichiometrically reversible). Of course, this also indicates that the yield of oxygen on carbon dioxide will also change with time, and it is likely that this is also the cause of the offset between identified and theoretical Y_1. Therefore, in

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

future work (detailed in Section 9), methods of dealing with this problem will be considered. While theoretical yields are important to consider in metabolic flux models, they are less useful here, where a much simplified metabolism is used.

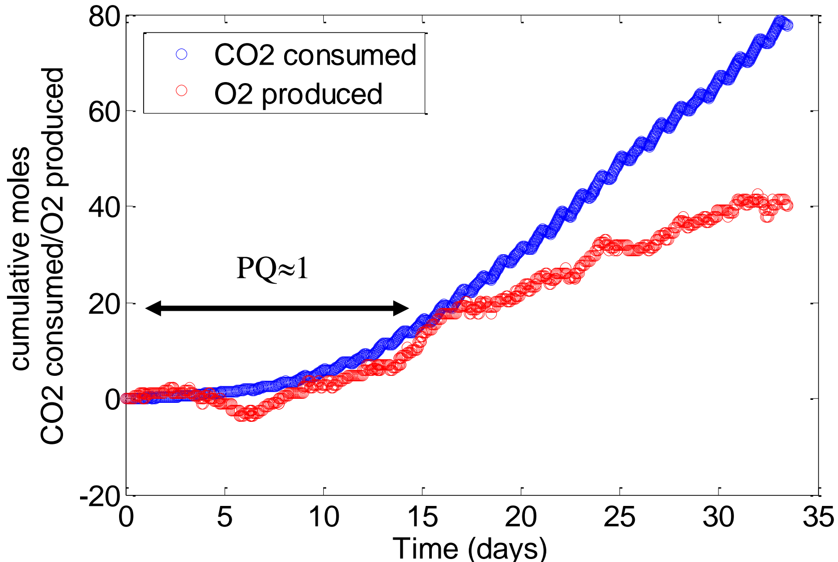

Fig. 2 Evolution of carbon dioxide consumed / oxygen produced with time for Beets 0504 experiment. (PQ = photosynthetic quotient)

For this work, identified yield values represented the experimental data well, and therefore these were taken as the true values and set as constants for the kinetic parameter identification.

7 Kinetic Model Identification

7.1 Data Required for Kinetic Model Identification

Temporal data for biomass, light, carbon dioxide and oxygen concentration in the plant chamber is needed for the kinetic model identification. This data was available for carbon dioxide but only periodically for oxygen. Therefore, average oxygen concentrations were calculated for the periods over which the data recording was off. Biomass measurements were only available at the end of the experiments. However, temporal biomass data was generated using the transformation that was used for the yield identification (14). Light intercepted and available for photosynthesis and photorespiration (μmol PAR m^{-2} s^{-1}) was quantified using a standard method analogous to the Beer-Lambert law for diffusion through a murky medium:

$$\text{Light intercepted} = IP\left(1-\exp\left(-k\,LAI\right)\right) \tag{16}$$

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

$$LAI = \frac{LA}{PA} \tag{17}$$

$$LA = LAR \times XV \tag{18}$$

where IP is the incident photon flux (µmol PAR m^{-2} s^{-1}), k is the extinction coefficient (0.7, chosen within a feasible range), LAI is the leaf area index, LA is the leaf area (m^2), PA is the planting area (5 m^2), and LAR is the ratio of leaf area to biomass dry weight, or leaf area ratio (0.016 m^2 g^{-1} calculated from data).

7.2 Kinetic Model Selection

7.2.1 Selection of Mitochondrial Respiration Equation

The rate under consideration is a net photosynthetic rate (19) which should include a term for photosynthesis (r_{ps}), photorespiration (r_{pr}), and mitochondrial respiration (r_{mr}).

$$r = r_{ps} - r_{pr} - r_{mr} \tag{19}$$

Photosynthesis and photorespiration both require a light environment for their reactions to proceed. Therefore, the rates of these reactions must be zero at night, and any changes in the carbon dioxide concentration must be due to mitochondrial respiration. This can be used to simplify the kinetic model selection process. By considering night data separately, a rate of mitochondrial respiration can be selected without considering the influence of light dependent reactions.

Several models for mitochondrial respiration were tested. Based on the reaction scheme, it was initially proposed that mitochondrial respiration should be a function of oxygen concentration inside the plant. Two rate equations were proposed – a simple first order rate law and one that assumed Monod kinetics.

$$r_{mr} = v_3 O_i \tag{20}$$

$$r_{mr} = \frac{v_3 O_i}{K_O + O_i} \tag{21}$$

In the above equations, v_3 is a rate constant to be identified, and K_O is a Michaelis-Menten constant for oxygen taken from the literature (g m^{-3}). An identification and validation was performed on night data using a least squares optimization. In the validation step, biomass, carbon dioxide, and oxygen were all reinitialized to their true values at the beginning of each night. Fig. 3 shows the night-time real and predicted carbon dioxide concentration using the Monod equation (21) to represent mitochondrial respiration.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

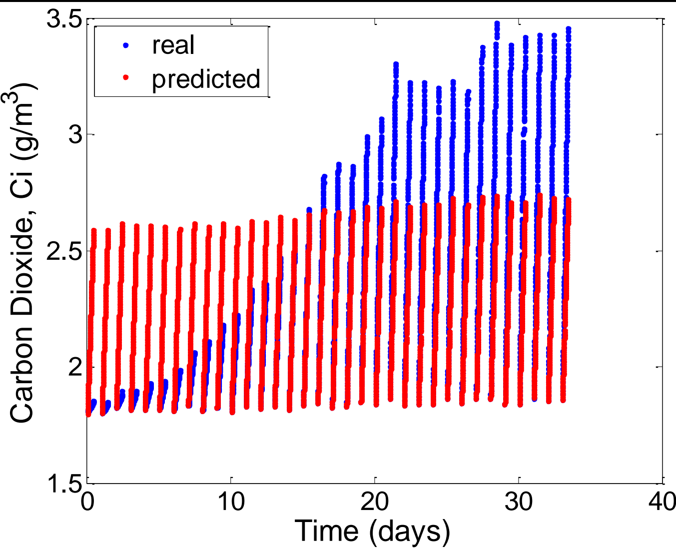

Fig. 3 Validation of Monod mitochondrial respiration equation (21) on lettuce night data, showing predicted and real CO_2 versus time.

Fig. 3 shows that the model is not adequate to represent the data. Similar results were obtained using the simple first order equation (20). These results suggest that oxygen is not the limiting substrate in mitochondrial respiration, and therefore does not determine the rate.

It is generally thought that mitochondrial respiration is affected by both the energy demands of the plant and the rate of supply of the carbon substrates produced through photosynthesis (often called carbon assimilates) [12]. In the current model, mitochondrial respiration has been divided into two components, as has been suggested in the literature [13]: growth respiration which is proportional to the photosynthetic rate (and can be thought of as representing the rate of production of assimilates), and maintenance respiration which is proportional to the total biomass. The model was therefore updated to represent mitochondrial respiration in this way (22).

$$r_{mr} = v_3(r_{ps,dailyavg} - r_{pr,dailyavg}) + v_4 XV \qquad (22)$$

In this case, v_3 (no units) and v_4 (s^{-1}) are rate constants, and $r_{ps,\ daily\ avg}$ and $r_{pr,\ daily\ avg}$ represent the average values of r_{ps} and r_{pr} over the previous day (g s^{-1}). The growth respiration term was represented by the difference between the average rate of photosynthesis and photorespiration to approximate the availability of assimilates for mitochondrial respiration. It was not possible to test this rate equation on night data alone, because the average values of r_{ps} and r_{pr} over the previous day clearly depend on daytime data. However, the growth respiration term can be approximated as an average rate of biomass accumulation over the previous day ($\Delta XV_{1d}/\Delta t_{1d}$), since this biomass accumulation will follow essentially the same trend (23). This

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

approximation was used to test the concept on night data. Fig. 4 shows the results on carbon dioxide data.

$$r_{mr} = v_3\left(\frac{\Delta XV_{1d}}{\Delta t_{1d}}\right) + v_4 XV \tag{23}$$

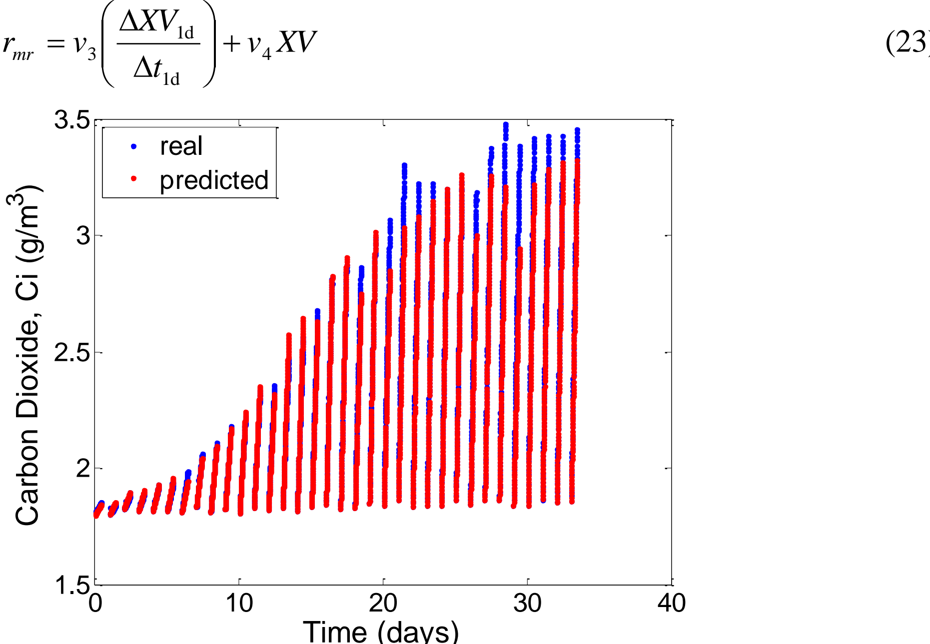

Fig. 4 Validation of growth + maintenance respiration equation (23) on lettuce night data. Showing predicted and real CO_2 versus time.

The predicted values follow experimental data quite well, suggesting that this mitochondrial respiration equation is adequate for representing night data. It should be noted that the maintenance respiration term (proportional to total biomass, v_4XV) was very small compared to the growth respiration term. The removal of this term had very little effect on the fit of the model, and therefore it was not included in the mitochondrial respiration equation selected for further testing (24).

$$r_{mr} = v_3\left(r_{ps,dailyavg} - r_{pr,dailyavg}\right) \tag{24}$$

The difference between the average rate of photosynthesis and photorespiration over the previous day was chosen over the simplification of using an average biomass accumulation rate (as in (23)) because the former is more directly related to the rate of production of carbon assimilates, which are assumed to be the limiting substrates in the respiration reaction. The validity of the respiration rate equation will be further tested in the full kinetic model identification and validation.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

7.2.2 Selection of Photosynthesis and Photorespiration Equations

The reaction scheme for photosynthesis suggests that its rate should be determined largely by the concentration of carbon dioxide and the available light energy. Based on this, Monod kinetics were chosen to represent the rate (25). Similarly photorespiration was defined by Monod kinetics considering oxygen and intercepted light as the most important factors (26).

$$r_{ps} = \frac{v_1 C_i}{K_C + C_i} \textit{Light intercepted} \tag{25}$$

$$r_{pr} = \frac{v_2 O_i}{K_O + O_i} \textit{Light intercepted} \tag{26}$$

In the above equations, v_1 (m^2 g μmol $PAR^{-1}s^{-1}$), v_2 (m^2 g μmol $PAR^{-1}s^{-1}$) are rate constants to be identified, K_C and K_O are Michaelis-Menten constants for carbon dioxide and oxygen which were estimated from the literature (g m^{-3}) [14], and all other constants and variables are as described previously.

7.3 Full Kinetic Model Identification and Validation on Lettuce Data

The full kinetic model that has been selected is summarized below (all equations have been previously described in the text, and the numbering follows accordingly):

$$r = r_{ps} - r_{pr} - r_{mr} \tag{19}$$

$$r_{ps} = \frac{v_1 C_i}{K_C + C_i} \textit{Light intercepted} \tag{25}$$

$$r_{pr} = \frac{v_2 O_i}{K_O + O_i} \textit{Light intercepted} \tag{26}$$

$$r_{mr} = v_3 (r_{ps,dailyavg} - r_{pr,dailyavg}) \tag{24}$$

The identification of kinetic parameters (v_1, v_2 and v_3) was solved using a least squares optimization technique (see Appendix B for code, Section 12.2). The identification and validation were initially performed using full experimental data. The results of the validation are shown in Fig. 5.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

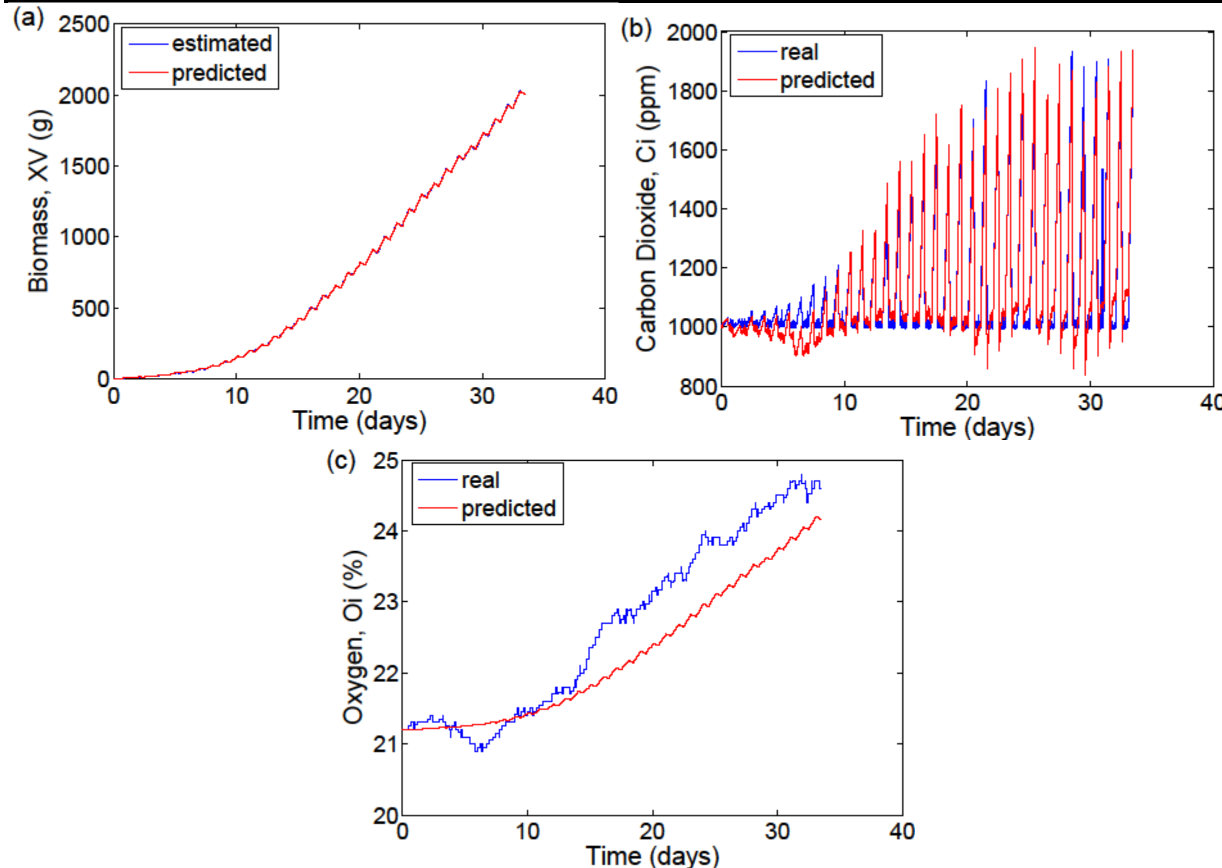

Fig. 5 Validation of kinetic model on lettuce experimental data, showing predicted and real (or estimated) (a) biomass (b) CO_2 and (c) O_2 versus time.

The validation shows that biomass and carbon dioxide are predicted quite accurately. There is some offset in the oxygen prediction, however it follows approximately the right trend. It is expected that this offset is due in part to some error in the estimated yields. In this case, the 'average' yield that was identified based on three lettuce experiments underestimates the total oxygen produced, and this causes the prediction to be offset from the true values. This hypothesis was tested by repeating the identification and validation procedure using a yield identified on this data set alone, and which would therefore represent the overall yield of this experiment perfectly. The results for oxygen (Fig. 6) show that the yield was partially, but not fully responsible for the error in the oxygen prediction. Despite this problem, the model represents the experimental data quite well.

As a second validation step, the identified model (with original yield estimates) was tested on a data set which was not used for identification of the kinetic parameters. This data was from a completely separate experiment; however the environmental conditions were the same. The results (Fig. 7) were quite similar to those observed using the identification data set. This reinforces the validity of the model for predicting lettuce growth.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

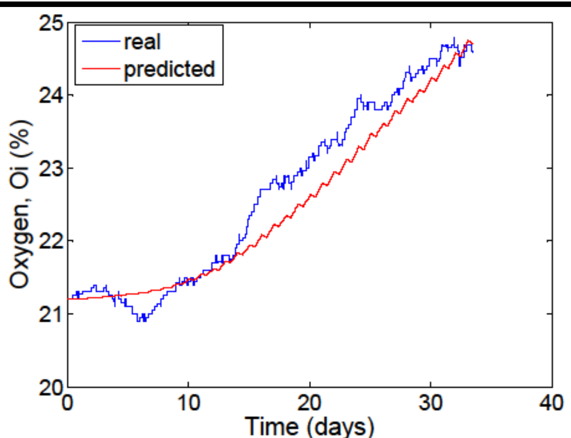

Fig. 6 Validation of kinetic model with 'corrected' yield on lettuce experimental data, showing predicted and real O_2 versus time. (yields based solely on this data set)

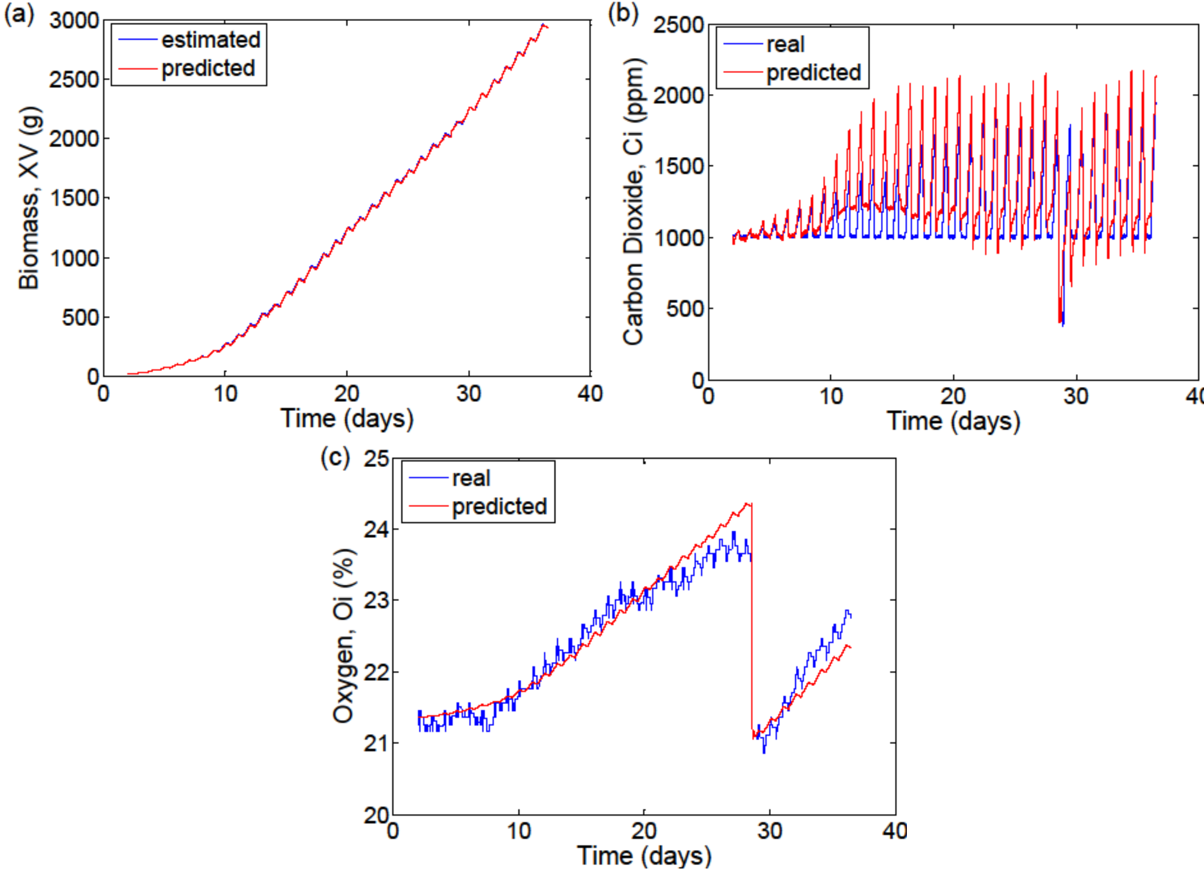

Fig. 7 Validation of kinetic model on separate lettuce data set, showing predicted and real (or estimated) (a) biomass (b) CO_2 and (c) O_2 versus time. The drop in oxygen concentration on day 28 was a result of chamber opening - the model was reinitialized on day 29 to account for this.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

7.4 Identification and Validation on Table 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, m^2g^{-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 Y_1=0.694 g XV g CO_2^{-1} [0.674, 0.713] and Y_2=0.347 g O_2 g CO_2^{-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 may 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 Fig. 8.

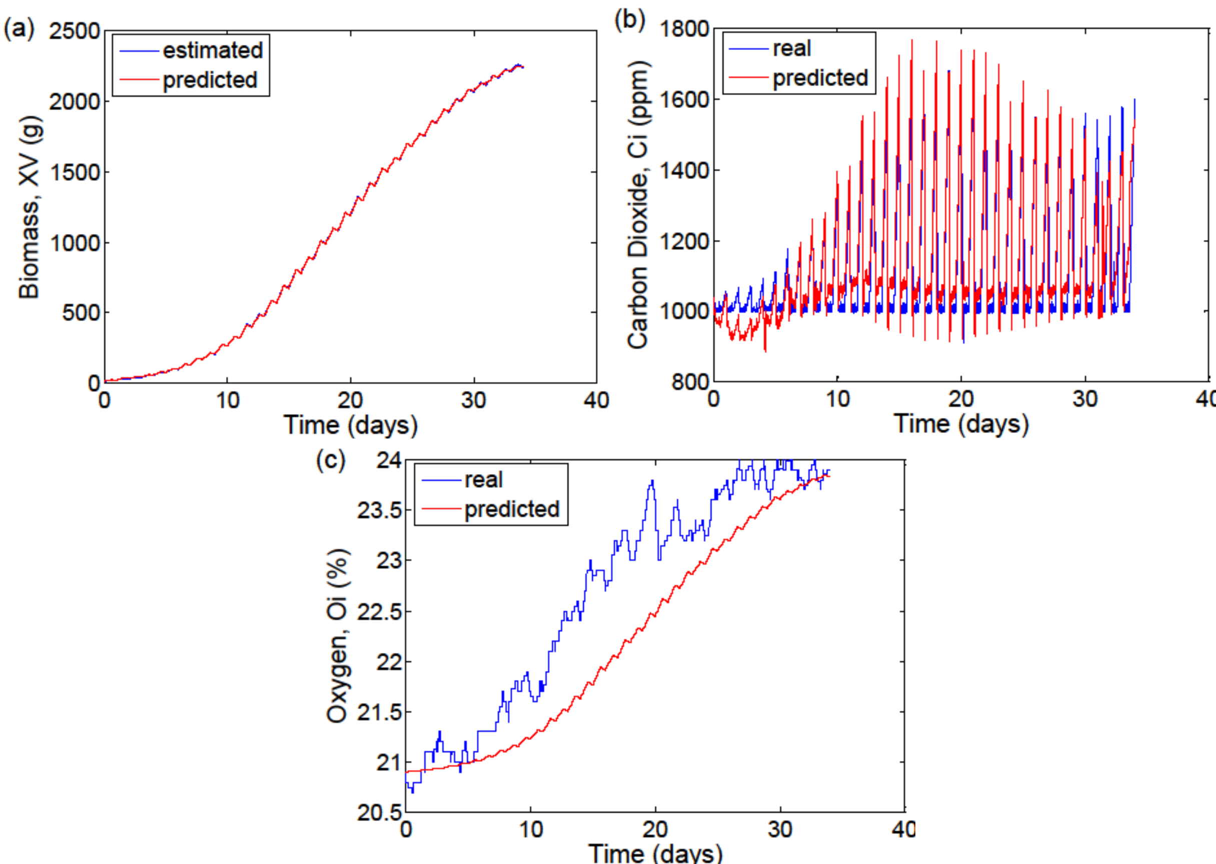

Fig. 8 Validation of kinetic model on full beet experimental data, showing predicted and real (or estimated) (a) biomass (b) CO_2 and (c) O_2 versus time.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

The model is a fairly good fit for the beet data, and especially for biomass. However, the carbon dioxide concentration is underestimated early in the experiment, and overestimated throughout most of the rest of the experiment. It was hypothesized that this could be due to 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 Fig. 8. This hypothesis was tested by attempting 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 Fig. 9.

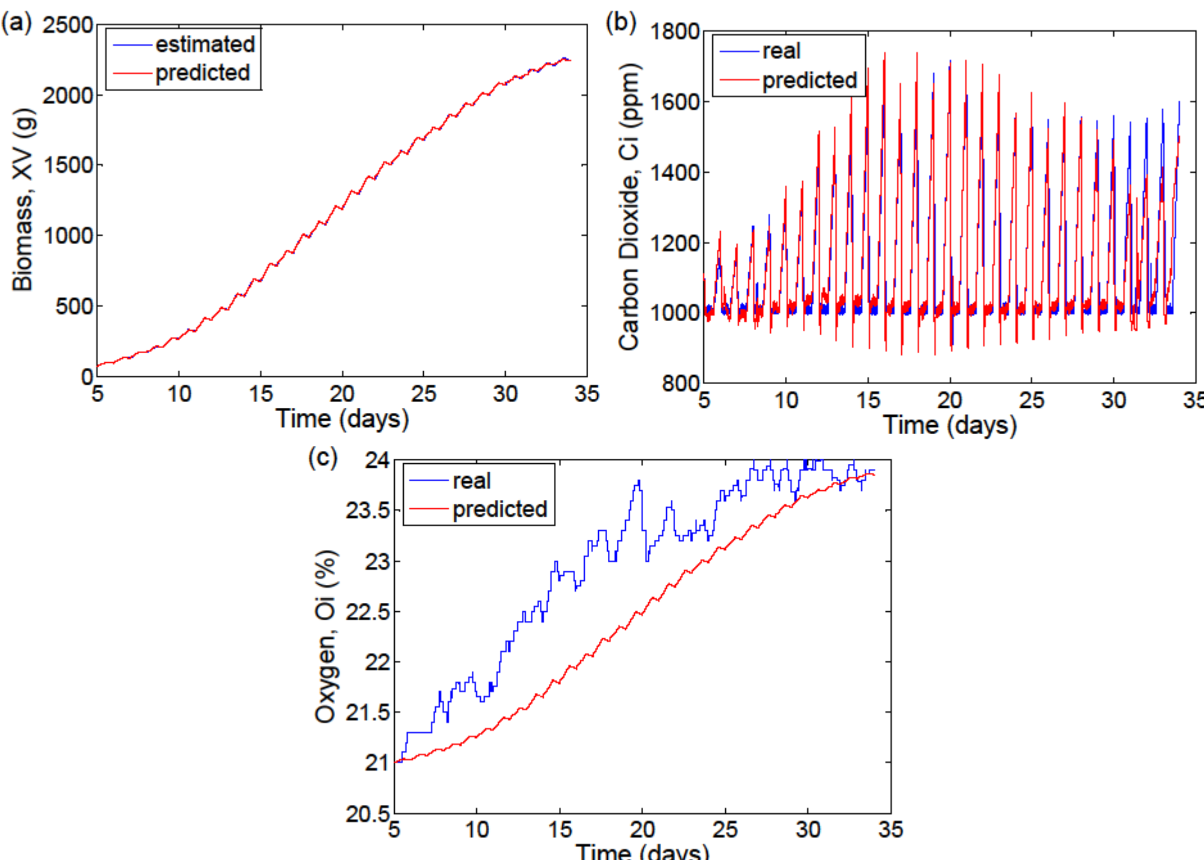

Fig. 9 Validation of kinetic model on day 5 to harvest beet data, showing predicted and real (or estimated) (a) biomass (b) CO_2 and (c) O_2 versus time.

The predicted carbon dioxide concentration fit the real data much better in this case, suggesting that an acclimation phase may indeed be required. However there are other possible reasons why the exclusion of the first 5 days would improve the model fit. Further discussion can be found in Section 7.5. This phase should be predicted by the model. Since the transfer phase is characterized by high light levels and possible root damage, it is suspected that the carbon dioxide uptake rate was lower than predicted due to stomata closure caused by water

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

limitation. A more accurate model could be achieved by including a water balance and the effects of water limitation. Fig. 9 also shows that oxygen production was underestimated throughout the experiment.

The model, which excluded day 1-5 data, 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 Fig. 10.

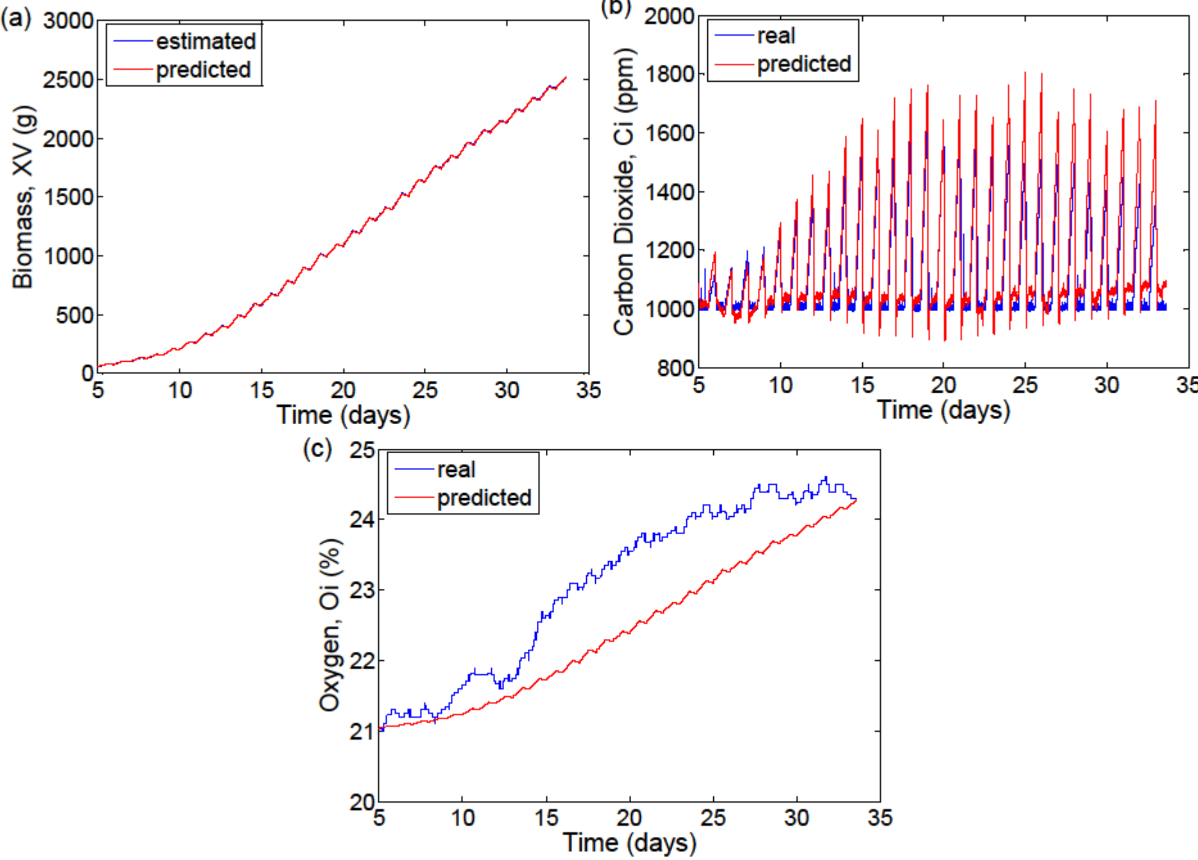

Fig. 10 Validation of kinetic model on separate beet experiment (day 5 to harvest), showing predicted and real (or estimated) (a) biomass (b) CO_2 and (c) O_2 versus time.

The model is again a fairly good fit for the biomass and carbon dioxide data, which reinforces it's validity for predicting beet growth excluding an acclimation phase. However, the underestimation of oxygen production in the chamber remains.

In this case, the oxygen prediction error is not a problem with the overall yield, since the predicted concentration matches the true value at the end of the experiment. Instead the prediction error may suggest that there are some unconsidered biological phenomena affecting

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

oxygen production, or that the yield between oxygen and carbon dioxide, as has been defined in this model, is not constant. As has been discussed previously (Section 6.2) the yield of oxygen on carbon dioxide is not constant throughout the experiment. This was true for both lettuce and beet data. Instead, the yield decreases throughout the experiment (seemingly in two stages), possibly due to a changing metabolism, a changing composition of biomass, microbial growth, or other factors. In future work (9), this problem must be dealt with in the model. One possible approach would be to divide the model into several developmental stages, with different yields. This approach would likely solve the problem of changing yeilds while minimizing the added complexity. A key challenge will be to identify the appropriate transition points between stages and identify these transitions based on measured data.

7.5 Comparison of Lettuce and Beet Models

Overall, when the model is applied to beet data there is less agreement between the predicted and measured states compared to its fit on lettuce data. The two main problems are in the oxygen prediction and in the fit of the model over the first 5 days of growth in the chamber. Potential causes and solutions for the oxygen prediction problem have been discussed. The limited data initially makes it seem that the oxygen prediction problem is more pronounced for beet data. One possible explanation for this would be that the yields might not change as drastically for lettuce as they do for beets. This could be explained by the more complicated plant structure of beets. For example, it is expected that the metabolism of beets will shift from primarily leaf production during early growth towards more storage root production later in the experiment, whereas lettuce would not experience such a substantial shift in metabolism. This may result in more drastic changes in the yields for beets compared to lettuce. In any case, there are clearly errors for both plants, and it is expected that improvements to the yield estimation should help to solve these errors.

The changing yields could also be responsible for the problem in fitting the model to full beet experimental data. Initially it was proposed that the lack of fit over the first 5 days could be due to an acclimation phase that the plants experience after transfer to the chamber. Although this could be true, it does not explain why the same acclimation effect is not observed for lettuce. An alternate explanation is that an underestimation of the yield of oxygen on carbon dioxide (Y_2) at the beginning of the experiment could cause the optimization procedure to overestimate the rate term (which would lead to the observed underestimation of CO_2 concentration). In this case, the fact that the prediction problem during early growth are not observed for lettuce could again be contributed to smaller changes in the yields over the experiments compared to beets.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

8 Conclusions

We have developed a simple, general model for plant growth considering photosynthesis and respiration reactions. The model is generally satisfactory for predicting lettuce growth and the growth of beets (excluding an initial 5-day period) during batch experiments. However, for both crops there remains some error between real and predicted oxygen concentrations. The reasons for this discrepancy are currently under investigation. It is hypothesized that the oxygen prediction problem, as well as the error in the prediction of early beet growth, may be caused by changing yields over the course of the experiments. This problem will be further investigated as part of the future work. The validations performed in this work utilized experimental data that was collected under the same environmental conditions as the identification dataset. Therefore, little can be said about the reliability of the model under different environmental conditions at this time. However, the model is adequate for small changes in environmental conditions (such as those observed within an experiment or between experiments with the same set-points)

9 Proposal of Future Work

The following list proposes various tasks that could be performed and/or considered as part of future work on model development.

- Find solutions for the problems caused by changing yields with time (problems include errors in oxygen prediction and possibly errors in the predictions of early beet growth)
    - Approach to be tested: Divide the model into several submodels representing subsequent developmental stages. By identifying different yields for the different phases of growth, the model predictions should be improved without adding much complexity. It is important to keep the model relatively simple so that it can be applied for control. An important issue with this approach will be to identify online indicators that could signal the transition to the next developmental stage. As an example, an online estimation of the photosynthetic quotient (based on CO_2 and O_2 measurements) could be used as an indicator to detect a change in the metabolism of the plant, however further discussion with the plant physiology working group is needed to make this selection.
    - It is also possible that we could redefine the model structure to create a model with constant yields. Possible changes could include removing the assumption that photosynthesis, photorespiration, and respiration should be considered one stoichiometrically reversible reaction. This change would increase complexity, may cause some issues with respect to identifiability of the parameters. However, it would allow the yields to vary when the relative rates of the different reactions change (photosynthesis, photorespiration, mitochondrial respiration). The first approach is preferred for simplicity.
- Consider a more detailed approach in which the development of different plant parts (organs) is taken into account. Two approaches are proposed below. Both of these

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

approaches will require data on organ growth with time that is not available for the data sets used in this report. The required data has been elaborated in Section 4.1.
    - The first approach will consider sequential developmental stages (as proposed as a solution to the problems caused by changing yields above). Organ development can also be considered using this approach. The model could be divided into submodels where each stage is dominated by growth of one or several plant parts.
    - A second approach will consider sub-models (compartments) for each organ type with the connections between organs controlled by transfer processes.
- Link biomass partitioning to metabolism. Information on organ development can be linked to growth and metabolism to provide useful information that could result in more accurate predictions. Partitioning can be included in the model according to the two approaches discussed in the previous point. Model improvements resulting could take several forms.
    - More accurate leaf area predictions, through estimation of leaf biomass development with time, could lead to more accurate prediction of the amount of light intercepted, which is a key variable in the rate equation.
    - As biomass partitioning changes during growth, the metabolism and yields could be affected. For example, during beet development it is likely that leaf growth would be a dominant sink for carbohydrates produced through photosynthesis early in development. At a later stage of growth, it is expected that partitioning would shift more towards storage root production. This could lead to changing yields for CO_2, O_2, and various nutrients as the biomass composition changes. This example would also increase the ratio of respiring tissues to photosynthesizing ones, which would also impact the metabolism. These effects should be considered for inclusion in the model.
- Expand model to include a water and energy balance. These balances are highly coupled, since water loss from the plant through transpiration is the largest source of heat loss for the plant. The inclusion of these balances will provide useful information for the larger life support system. In addition, the energy state of the plant has a large impact on many plant processes (including metabolism) and the plants' water status influences substrate concentrations available for photosynthesis and respiration through its' effect on transfer at the leaf surface (stomatal control). Therefore, it is expected that the inclusion of water and energy balances should improve model robustness and provide important insight for future model improvements. In addition, these modifications should provide the model structure necessary to predict the effects of water stress.
- Test model on different plants and under different environmental conditions. This will be important to test reliability, and also to improve and refine the model. Currently, all experimental data used for identification and validation were collected under the same environmental conditions. The model was able to account for small differences between datasets (mostly due to experimental errors or natural variability between

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

experiments), however there is so far no guarantee that this model would work well if the environmental conditions were changed more drastically. To ensure model robustness over a wide variety of environmental conditions, data under a variety of conditions (as described in Section 4.2) is needed.

- Analysis of currently available data and a data search to attempt to find other datasets that can be used for model identification and validation should be conducted. Data on organ development with time will be required before a plant organ compartment approach can be tested. Therefore, finding datasets that have dry weight measurements with time for the different organs will be a top priority.
  - Analysis of currently available data. The staggered growth experiments (available from Guelph) provide some information on organ development because at the conclusion of the experiment all plants in the chamber (including plants that have been in the chamber for different lengths of time) are harvested. This could potentially be useful for model identification, and the value of these data sets should be further analyzed and explored as part of future work. However, it should be kept in mind that the conditions will not be perfectly consistent over the full experiment. Therefore, plants harvested after 10 days (for example) were likely exposed to at least slightly different conditions than were present the first 10 days of growth of the other plants in the study (since the plants in the chamber are not all of the same age). The importance of this distinction is so far unknown; nonetheless the staged experiments should at least provide some useful insight into organ development. The model should be tested on data currently available from staggered growth experiments. Based on this work, further analysis should be done on the potential benefits and drawbacks of this cultivation technique.
  - A data search should also be conducted to provide additional datasets that can be used for model identification and validation. This may include working with existing software (DigiPlante) to provide data, or finding additional sources of experimental data (potentially from other groups working in the field).
- Model refinement. Several simplifications have been made in model development. These simplifications could be re-evaluated if adding additional information would improve model validation significantly. Several areas could be considered:
  - Consider transfer at the leaf surface. Rates of photosynthesis and respiration are currently calculated based on atmospheric concentrations in the plant chamber. Physiologically, they should depend on the concentrations at the site of the reaction (in the leaves). These concentrations could be estimated by considering transfer into and out of the leaves, controlled by a resistance to transfer that depends on stomatal control. (Some work has been done on trying to include transfer at the leaf surface, but so far this has been unsuccessful – appropriate model parameters could not be identified).
  - Discussions with the PPWG may yield other insights with respect to plant physiology and metabolism which should be considered and addressed. For

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

example, it has been suggested that certain enzyme quantities may be important in determining rates (for example Rubisco for photosynthesis). Quantification of important enzymes is not currently available in datasets, further analysis and discussion is required to determine if and how such effects can be included.

- Include nutrient balances. These will likely have to be fairly simple, at least in the short term. Modelling of nutrient composition of biomass will likely be based largely on yield values from plant production experiments. However, the composition of the most important elements could potentially be linked to metabolism at a later stage.
  - For example, it is knows that nitrogen is a very important nutrient for the plant, and nitrogen limitation strongly inhibits growth. Including the effect of nitrogen on growth and metabolism would be useful in predicting the result of stress conditions. Of course, experiments will be required on any nutrients of interest. Experiments to test the effects of individual nutrients on growth have not been included in the proposed experiments at this time (4.2). An experiment in which the nutrient solution circulation is turned off and on has been proposed, but this would not allow for any conclusions about individual nutrients. If this is necessary, more experiments should be proposed. Alternatively we could consider some average nutrient effect, where low overall nutrient concentrations could be linked to slowed growth. This should be discussed further with the plant physiology working group.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

10 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 C4 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 and 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] Dochain D. (2008). *Bioprocess Control*. ISTE Ltd, London. John Wiley & Sons, Inc., Hoboken, NJ.

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

[10] 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.

[11] Blankenship R.E. (2002). *Molecular Mechanisms of Photosynthesis*. Wiley-Blackwell, London. MPG Book Ltd., Bodmin, Cornwall.

[12] Farrar J.F. (1985). The respiratory source of CO2, *Plant, Cell and Environment*, Vol. 8, pp 427-438.

[13] Thornley J.H.M. and M.G.R. Cannell (2000). Modelling the Components of Plant Respiration: Representation and Realism, *Annals of Botany*, Vol. 85, pp 55-67.

[14] 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.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

11 Appendix A – Theoretical Yield Calculations

Calculation of theoretical yield of biomass on CO_2:

Assumption (from (11)): 1 mol CO_2 uptake = 1/6 mol biomass = 1 mol C in biomass
Carbon composition of biomass: 38% biomass dry weight (0.38 gC g biomass^{-1})

$$1\,\text{mol C in biomass}\left(\frac{12\,\text{g C}}{\text{mol C}}\right)\left(\frac{\text{g XV}}{0.38\,\text{g C}}\right) = 31.6\,\text{g XV}$$

$$1\,\text{mol CO}_2\,\text{uptake}\left(\frac{44\,\text{g CO}_2}{\text{mol CO}_2}\right) = 44\,\text{g CO}_2$$

$$\frac{1\,\text{mol C in biomass}}{1\,\text{mol CO}_2\,\text{uptake}} = \frac{31.6\,\text{g XV}}{44\,\text{g CO}_2} = 0.718\,\text{g XV g CO}_2^{-1}$$

Calculation of theoretical yield of O_2 on CO_2:

Assumption: photosynthetic quotient = 1 mol O_2 mol CO_2^{-1} [11]

$$\frac{1\,\text{mol O}_2\,\text{produced}}{1\,\text{mol CO}_2\,\text{consumed}}\left(\frac{32\,\text{g O}_2/\text{mol O}_2}{44\,\text{g CO}_2/\text{mol CO}_2}\right) = 0.727\,\text{g O}_2\,\text{g CO}_2^{-1}$$

$$\frac{1.3\,\text{mol O}_2\,\text{produced}}{1\,\text{mol CO}_2\,\text{consumed}}\left(\frac{32\,\text{g O}_2/\text{mol O}_2}{44\,\text{g CO}_2/\text{mol CO}_2}\right) = 0.945\,\text{g O}_2\,\text{g CO}_2^{-1}$$

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

12 Appendix B – Implementation of the Model

The model is written in MATLAB. A series of m-files are called in succession to execute either the yield identification and validation or the kinetic model identification and validation. The sequence is called in the command window of MATLAB by calling the first m-file (yieldest for yield identification and validation, call_plant_model for the kinetic model identification and validation). Note that the text colours are as they appear in MATLAB – comments are shown in green.

12.1 Source code for yield identification and validation

The source code for the yield identification is divided into two m-files (shown in Sections 12.1.1 to 12.1.2 below). The structure of the model is shown in Fig. 11. In this case, the program is called by starting yieldest.m. Within yieldest.m there is a command to run yield_regression.m. Once this file executes it will return certain variables back to yieldest.m, which will continue running until all commands have been executed.

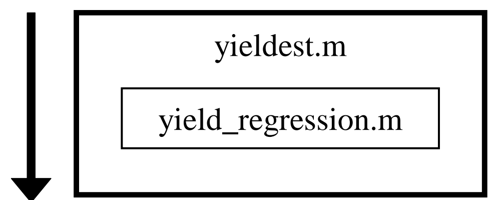

Fig. 11 Flow chart depicting model structure for yield identification.

12.1.1 M-file - yieldest.m

```
function yieldest
clear all
close all
clc

% Select figure locations on screen:
scrsz = get(0,'ScreenSize');
figure('Position',[1 scrsz(4)/2 scrsz(3)/2 scrsz(4)/2])
figure('Position',[scrsz(3)/2 scrsz(4)/2 scrsz(3)/2 scrsz(4)/2])
figure('Position',[1 1 scrsz(3)/2 scrsz(4)/2])
figure('Position',[scrsz(3)/2 1 scrsz(3)/2 scrsz(4)/2])

% Introduce data and select data sheets for validation/identification:
global ident_data valid_data numberidentfiles valid_time_data ident_time_data Vchamber
Vchamber=29;                    % volume of chamber (m3)

% Opens dialog box. If YES - 2 data sets used for identification, 1 for validation
% If NO - all 3 data sets used for identification
answer = inputdlg('Would you like to use 1 of the 3 data sets for validation? (yes/no)')

% The following IF structure sets up the two cases (YES or NO) that could
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
% result from the above dialog question.
if strcmp(char(answer), 'yes')==1,

    numberidentfiles=2;

    % The following opens two dialog boxes to select the identification and
    % validation files (files must be in the same folder as this m-file).
    d = dir;
    str = {d.name};
    [valid_worksheet,v] = listdlg('PromptString','Select a file for validation:',...
                'SelectionMode','single',...
                'ListString',str);

    [identification_worksheet,v2]=listdlg('PromptString','Select files for identification:',...
                'SelectionMode','multiple',...
                'ListString',str);

    % Reads the excel worksheets into matlab as arrays.
    valid_data=xlsread(char(str(valid_worksheet)),'Sheet1')
    valid_time_data=xlsread(char(str(valid_worksheet)),'Sheet2')
    ident_data(1:2,:)=xlsread(char(str(identification_worksheet(1))),'Sheet1');
    ident_time_data{1}=xlsread(char(str(identification_worksheet(1))),'Sheet2');
    ident_data(3:4,:)=xlsread(char(str(identification_worksheet(2))),'Sheet1');
    ident_time_data{2}=xlsread(char(str(identification_worksheet(2))),'Sheet2');

    % Set variables for output file (data is written to an excel file later in this program):
    if char(str(valid_worksheet))=='Lettuce GW0204.xls',
        output_position='A2';
        valid_ident=0204;
     elseif char(str(valid_worksheet))=='Lettuce GW0604.xls',
        output_position='A3';
        valid_ident=0604;
    elseif char(str(valid_worksheet))=='Lettuce GW0704.xls',
        output_position='A4';
        valid_ident=0704;
    end

elseif strcmp(char(answer), 'no')==1,
    numberidentfiles=3;
    % Reads the excel worksheets into matlab as arrays (uses all data sets in this case).
    ident_data(1:2,:)=xlsread('Lettuce GW0204.xls','sheet1');
    ident_data(3:4,:)=xlsread('Lettuce GW0604.xls','sheet1');
    ident_data(5:6,:)=xlsread('Lettuce GW0704.xls','sheet1');
    ident_time_data{1}=xlsread('Lettuce GW0204.xls','sheet2');
    ident_time_data{2}=xlsread('Lettuce GW0604.xls','sheet2');
    ident_time_data{3}=xlsread('Lettuce GW0704.xls','sheet2');
    close(figure(3))
    close(figure(4))
    % set variables for output file (data is written to an excel file later in this program):
        output_position='A5';
        valid_ident=NaN;
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
end

% Call subroutine to do regression. This calls a separate m-file "yield_regression.m".
% It returns certain data to this m-file as "yieldinfo".:
yieldinfo=yield_regression
% Separate the information returned in "yieldinfo" into separate variables (yields Y1 and Y2,
% and confidence intervals Y1conf and Y2conf)
Y1=yieldinfo(1,1)
Y2=yieldinfo(2,1)
Y1conf=[yieldinfo(1,2) yieldinfo(1,3)]
Y2conf=[yieldinfo(2,2) yieldinfo(2,3)]

% Validation - if one of the data sets was initially selected for validation,
% this code will plot the validation data along with the predicted data
% (based on the yield values calculated in the identification):
if strcmp(char(answer), 'yes')==1,
   XV_valid=[valid_data(1,2) valid_data(2,2)];          % Total biomass (gDW) accumulated
   Ca_valid=[valid_time_data(1,2) trapz(valid_time_data(:,1),valid_time_data(:,2))/valid_time_data(end,1)];
                                                         % CO2 concentration in atmosphere (g/m3)
                                                         % Note: harvest CO2 calculated using integral of CO2 over
                                                         % full experiment divided by total experiment time
   Ci_valid=Ca_valid;                                    % Assuming Ci=Ca
   u1_valid=trapz(valid_time_data(:,1),valid_time_data(:,3));  % the integral of u1 over the full experiment (g),
                                                         % where u1 is net CO2 input to the chamber.
   y_var1_valid=XV_valid(2)-XV_valid(1);                 % y variable in the XV regression  (gXV)
   x_yar1_valid=u1_valid-(Ci_valid(2)-Ci_valid(1))*Vchamber;   % x variable in the XV regression
                                                         % (gCO2)
   plotline_x_var1_valid=[0 x_yar1_valid];               % for plotting
   plotline_y_var1_valid=plotline_x_var1_valid*Y1;       % for plotting
   residXV_valid=y_var1_valid-plotline_y_var1_valid(end)  % calculates residual

   figure(3)
   hold off
   plot(x_yar1_valid,y_var1_valid,'bo')
   hold
   plot(plotline_x_var1_valid,plotline_y_var1_valid,'-r')
   xlabel('g CO2')
   ylabel('g biomass')

   Oa_valid=[ident_data(1,7) ident_data(2,7)];           % O2 concentration in the atmosphere (g/m3)
   Oi_valid=Oa_valid;                                    % Assuming Oi=Oa
   y_var2_valid=(Oi_valid(2)-Oi_valid(1))                % y variable in the O2 regression  (gO2/m3)
   x_var2_valid=u1_valid/Vchamber-(Ci_valid(2)-Ci_valid(1));   % x variable in the O2 regression  (gCO2/m3)
   plotline_x_var2_valid=[0 x_var2_valid];               % for plotting
   plotline_y_var2_valid=plotline_x_var2_valid*Y2;       % for plotting
   residO2_valid=y_var2_valid-plotline_y_var2_valid(end)  % calculates residual

   figure(4)
   hold off
   plot(x_var2_valid,y_var2_valid,'bo')
   hold
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
   plot(plotline_x_var2_valid,plotline_y_var2_valid,'-r')
   xlabel('g/m3 CO2')
   ylabel('g/m3 O2')

   % Output data to Excel file ("output_data.xls" - if one exists it will add the data in the specified position,
   % otherwise will create a new file):
   output_data=[valid_ident,Y1,Y1conf,Y2,Y2conf,residXV_valid,residO2_valid]
   xlswrite('output_data.xls',output_data,'Sheet1',output_position);

elseif strcmp(char(answer), 'no')==1,                % if no validation step was done
   % Output data to Excel file ("output_data.xls" - if one exists it will add the data in the specified position,
   % otherwise will create a new file):
   output_data=[valid_ident,Y1,Y1conf,Y2,Y2conf]
   xlswrite('output_data.xls',output_data,'Sheet1',output_position);
end
```

12.1.2 M-file – yield_regression.m

```
function val=yield_regression

global ident_data numberidentfiles ident_time_data Vchamber

%Calculation of variables for yield identification:
for m=1:numberidentfiles,
   clear Oa_recorded last5 first5 data Oa_data

   XV(m,:)=[ident_data(2*m-1,2) ident_data(2*m,2)]     % total biomass (gDW) accumulated
   ident_time_data_working=ident_time_data{m};
   Ca(m,:)=[ident_time_data_working(1,2) trapz(ident_time_data_working(:,1),ident_time_data_working(:,2))/
         ident_time_data_working(end,1)]
                                                     % The equation (Ca(m,:)=…) should be on one line
                                                     % CO2 concentration in atmosphere (g/m3)
                                                     % Note: harvest CO2 calculated using integral of CO2 over
                                                     full experiment divided by total experiment time

   u1_accum(m)=trapz(ident_time_data_working(:,1),ident_time_data_working(:,3))
                                                      % the integral of u1 over the full experiment (g), where u1
                                                     is net CO2 input to the chamber.

   Ci(m,:)=Ca(m,:)                                   % It is assumed that Ci=Ca
   Oa(m,:)=[ident_data(2*m-1,7) ident_data(2*m,7)]   % O2 concentration in the atmosphere (g/m3)
   Oi(m,:)=Oa(m,:)                                   % It is assumed that Oi=Oa

   % Treating the yield identification in the following way: y_var=Yield*x_var the variables required for the
   % regression are calculated:
   x_var1(m)=u1_accum(m)-Vchamber*(Ci(m,2)-Ci(m,1));      % x variable in the XV regression  (gCO2)
   plotline_x_var1=[0 max(x_var1)];                       % for plotting

   x_var2(m)=u1_accum(m)/Vchamber-(Ci(m,2)-Ci(m,1));      % x variable in the Oi regression (gCO2/m3)
   plotline_x_var2=[0 max(x_var2)];                       % for plotting

   y_var1(m)=(XV(m,2)-XV(m,1));                           % y variable in the XV regression (gXV).
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
  y_var2(m)=(Oi(m,2)-Oi(m,1));                          % y variable in the O2 regression (gO2/m3).

end

% Perform regression:
[Y1 Y1conf]=regress(y_var1',x_var1')
[Y2 Y2conf]=regress(y_var2',x_var2')
val=[Y1 Y1conf; Y2 Y2conf]                   % val is the data returned to the calling m-file ("yieldest.m")

% Additional variables calculated for plotting. Figures (1) and (2) show the validation on identification data sets.
y_var1_model=Y1*x_var1                                  % y_var as calculated by the model
plotline_y_var1_model=plotline_x_var1*Y1;               % for plotting
residXV = y_var1'-y_var1_model'                         % calcaultes residuals

figure(1)
hold off
plot(x_var1,y_var1,'bo')
hold
plot(plotline_x_var1,plotline_y_var1_model,'-r')
xlabel('g CO2')
ylabel('g biomass')
text(1000,2500,{'Y1=' num2str(Y1)},'HorizontalAlignment','left')

y_var2_model=Y2*x_var2                                  % y_var as calculated by the model
plotline_y_var2_model=plotline_x_var2*Y2                % for plotting
residO2=y_var2'-y_var2_model';                          % calculates residuals

figure(2)
hold off
plot(x_var2,y_var2,'bo')
hold
plot(plotline_x_var2,plotline_y_var2_model,'-r')
xlabel('g/m3 CO2')
ylabel('g/m3 O2')
text(40,40,{'Y2=' num2str(Y2)},'HorizontalAlignment','left')

sse=residXV'*residXV+residO2'*residO2                   % calculates sum of squared errors
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

12.2 Source code for kinetic model identification and validation

The source code for the kinetic model identification is divided into six m-files (shown in Sections 12.2.1 to 12.2.6 below). The structure of the model is shown in Fig. 12. The figure should be followed top to bottom. So for example, the program is started by calling call_plant_model. Call_plant_model will execute until it reaches the command to call initialize_parameters. Initialize_parameters will then execute, and upon completion, should return certain data to the call_plant_model box. Call_plant model will then continue to execute until it reaches the command for calculate_XV and so on.

Call_plant_model.m is the general file from which all other files are called. Initialize_parameters.m holds all constants and parameters that can be selected. Therefore, all regular changes to the model can be made in this file (excepting changes to model structure). In addition, this file calls the excel spreadsheets containing experimental data and defines certain variables. Ident_and_valid.m in the file in which the identification and validation is performed. The identification uses fminsearch to minimize the sum of squared errors, as calculated in SSE.m. Validate.m is called to validate a known (or previously identified) model. SSE.m and validate.m are very similar files, but are used for different purposes.

The code uses the working kinetic model that is described in Section 7.2 of the report. For simplicity, other kinetic models tested are not shown, but they could be easily incorporated using a case structure. The code is for the lettuce identification and validation. The beet code is identical, except that leaf area ratio and yields are redefined.

This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

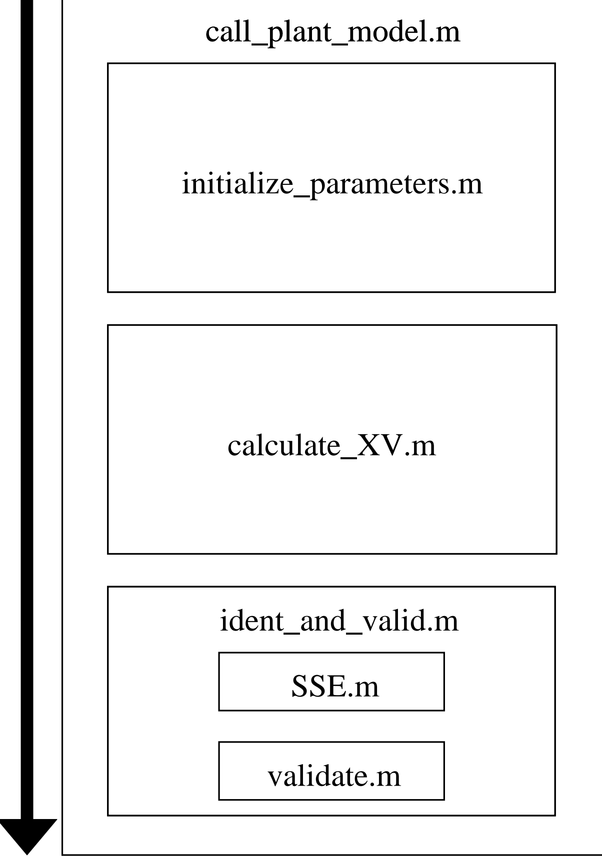

Fig. 12 Flow chart depicting model structure for kinetic model identification. The chart should be followed top to bottom.

12.2.1 M-file – call_plant_model.m

```
function call_plant_model      % General file from which I call all other files

clear all
close all
clc

% Define Global Variables:

global timedata yielddata yields T michment LAR ext_coeff plant_area reflect_coeff time Ca u1 IP Oa Oi XV_0
Ci XV Light v1 v2 v3 choose_case Vchamber u1_avg tdays
global starting_time ending_time validate_datasets v1_validate_full v2_validate_full v3_validate_full exclude
exclude_start exclude_end graphunits LA_0
global reinitialize time_ind_start time_ind_end

% Define figure locations:
scrsz = get(0,'ScreenSize');
figure(1)
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
figure('Position',[1 scrsz(4)/2 scrsz(3)/3 scrsz(4)/2])
figure('Position',[scrsz(3)/3 scrsz(4)/2 scrsz(3)/3 scrsz(4)/2])
figure('Position',[2*scrsz(3)/3 scrsz(4)/2 scrsz(3)/3 scrsz(4)/2])
figure('Position',[1 1 scrsz(3)/3 scrsz(4)/2])
figure('Position',[scrsz(3)/3 1 scrsz(3)/3 scrsz(4)/2])
figure('Position',[2*scrsz(3)/3 1 scrsz(3)/3 scrsz(4)/2])

% Initialize Parameters and Load Data:

initialize_parameters;        % This calls the m-file initialize_parameters

% Calculate biomass, Ci, Oi, and biomass data:

Ci=Ca;
Oi=Oa;
XV=calculate_XV;          % This calls the m-file calculate_XV

% Set IP to measured value if light ON, 0 if light OFF (This could be
% changed at a later stage, but works fine for now.
for i2=1:length(IP),
  if IP(i2)<100,
    IP(i2,1)=0;
  else
    IP(i2,1)=IP(i2);
  end
end

% Identify and validate model:
ident_and_valid;                    % This calls the m-file ident_and_valid
```

12.2.2 M-file – initialize_parameters.m

```
% Load Data from Excel worksheets:

% The following opens a dialog box to select the identification and
% validation dataset (file must be in the same folder as this m-file).
d = dir;
str = {d.name};
[worksheet,v] = listdlg('PromptString','Select a file for identification and validation:',...
            'SelectionMode','single',...
            'ListString',str);

yielddata=xlsread(char(str(worksheet)),'Sheet1')
timedata=xlsread(char(str(worksheet)),'Sheet2')
ind2=find(timedata(:,1)<43920,1,'last')
timedata=timedata(ind2:end,:);

if char(str(worksheet))=='Lettuce GW0604.xls',

  starting_time=216000;
  ending_time=3196800;            % Find the actual number for 0604
```

```
   % Give the option to exclude erroneous data (1=yes, 0=no):
   exclude=0;                % Do you want to exclude any time period? Should be 0 (no) or 1 (yes)
   exclude_start=2419200;
   exclude_end=2552400;
   reinitialize=1;

elseif char(str(worksheet))=='Lettuce GW0704.xls',

   starting_time=43920;
   ending_time=2937600;

   % Give the option to exclude erroneous data (1=yes, 0=no):
   exclude=0;          % Do you want to exclude any time period? Should be 0 (no) or 1 (yes)

end

% Choose starting and ending times for the identification and validation:
% This is where you can choose to select only a subset of the data, or full
% data by chooseing starting_time and ending_time respectively
   time_ind_start=starting_time;
   time_ind_end=ending_time;

% Initialize Parameters (These are the constants that you can change):

% Do you want to use selected parameters for validation steps?
% If you're using chosen parameters to validate the data sets, set to 1, if
% you're using parameters from identification, set to 0.
validate_datasets=1;
v1_validate_full=7.7319E-05;          % Parameters that will be used to validate
v2_validate_full=1.0294E-04;
v3_validate_full=0.3890;

graphunits=1;            % Note, 0 will put CO2 and O2 units in g/m3, 1 will put them in easier to read units
                         % ppm and % respectively
yields=[0.585 0.329];    % Input yields from the yield identification.
T=360;                   % sampling period (s)
michment=[0.94 487];     % These were calculated based on a paper by Farquhar et al for spinach
ext_coeff=0.7;           % Extinction coefficient for intercepted light calculation
reflect_coeff=0;         % Reflection coefficient for intercepted light calculation (not used)
plant_area=5;            % planting area (m2)
Vchamber=29;             % volume of plant growth chamber (m3)
LAR=0.016;               % m2/gDW average 0604 and 0704 data sets


% Identify the variables within the data sets:

time=timedata(:,1)-timedata(1,1);
Ca=timedata(:,10);       % Carbon dioxide in the atmosphere of the plant chamber (g/m3)
u1=timedata(:,3);        % net rate of CO2 additions to the chamber (addition-leakage) (g/s)
IP=timedata(:,6);        % incident photon flux
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
Oa_recorded=timedata(:,9);
% O2 data recorded on 6 hours, off 6 hours, I averaged when there is no data:
for t=1:length(Oa_recorded),
   if Oa_recorded(t)==0,
      last5=find(Oa_recorded(1:t-1),5,'last');
      first5=find(Oa_recorded(t:end),5,'first');
      data=[Oa_recorded(last5); Oa_recorded(t+first5-1)];
      Oa(t)=mean(data);
   else
      Oa(t)=Oa_recorded(t);
   end
end
Oa=Oa';

% Initial Conditions:
XV_0=yielddata(1,2);
LA_0=yielddata(1,5);

% The following will be saved and returned to call_plant_model:
save timedata yielddata yields T michment LAR ext_coeff reflect_coeff plant_area time Ca u1 IP Oa XV_0
Vchamber starting_time ending_time validate_datasets v1_validate_full v2_validate_full v3_validate_full
exclude exclude_start exclude_end graphunits LA_0 reinitialize time_ind_start time_ind_end
```

12.2.3 M-file – calculate_XV.m

```
function XV=calculate_XV

global time XV_0 u1 yields Ci Vchamber

% Calculate integral of u1 with time:
u1_integrated=cumtrapz(time,u1);
% Calculate XV according to yield equation:
XV=yields(1)*(u1_integrated-Vchamber*(Ci-Ci(1)))+XV_0;

% Plot XV versus time:
figure(1)
hold off
plot(time,XV,'b.')
hold
xlabel('time (s)')
ylabel('biomass (g)')
```

12.2.4 M-file – ident_and_valid.m

```
function ident_and_valid

global XV Ci Light michment Oi T u1 time yields timedata choose_case Vchamber starting_time ending_time
global v1_validate_full v2_validate_full v3_validate_full validate_datasets exclude exclude_start exclude_end
global ext_coeff LAR plant_area LA_0 XV_0 reinitialize XV_1 Ci_1 Oi_1 IP_1 time_1 rate_day1_1 u1_1
global Light_1 v4_select time_ind_start time_ind_end graphunits IP

% Use full experimental data for plotting, define:
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
ind_full_start=find(time(:,1)<(starting_time-timedata(1,1)),1,'last');
ind_full_finish=find(time(:,1)<(ending_time-timedata(1,1)),1,'last');
time_full=time(ind_full_start:ind_full_finish);
XV_full=XV(ind_full_start:ind_full_finish);
Ci_full=Ci(ind_full_start:ind_full_finish);
Oi_full=Oi(ind_full_start:ind_full_finish);

% Select starting and ending times for identification and validation (depending on whether data is excluded):
ind_start=find(time(:,1)<(time_ind_start-timedata(1,1)),1,'last');
ind_finish=find(time(:,1)<(time_ind_end-timedata(1,1)),1,'last');

% Select the data that falls within the chosen times:
time_1=time(ind_start:ind_finish);
u1_1=u1(ind_start:ind_finish);
Ci_1=Ci(ind_start:ind_finish);
Oi_1=Oi(ind_start:ind_finish);
XV_1=XV(ind_start:ind_finish);
IP_1=IP(ind_start:ind_finish);
LA_1=LA_0+(XV_1-XV_0)*LAR;                      % This is just used for an estimation of an initial rate
Light_1=IP_1.*(1-exp(-ext_coeff*LA_1/plant_area));   % This is just used for an estimation of an initial rate

% if we're using known parameters to validate data set, use parameters
% defined in identify_parameters.m
if validate_datasets==1
   set_parameters=[v1_validate_full;v2_validate_full;v3_validate_full];
   validate_set_parameters(set_parameters);
% if we're identifying parameters based on the data set, define starting
% values of parameters, and run fminsearch. Returns least squared estimates
% of the parameters.
else
   param_init=[3.0863E-05;2.9126E-05;0.8167];
   b_1=fminsearch('SSE',param_init);

   v1_full=b_1(1)
   v2_full=b_1(2)
   v3_full=b_1(3)

   set_parameters=[v1_full;v2_full;v3_full];
   validate_set_parameters(set_parameters);
end
```

12.2.5 M-file – SSE.m

```
function val=SSE(param)

global michment T yields Vchamber LAR plant_area ext_coeff LA_0 XV_0 reinitialize
global XV_1 Ci_1 Oi_1 time_1 IP_1 u1_1 Light_1 graphunits

% Define Michaelis-Menten constants individually:
Kc=michment(1);
Ko=michment(2);
```

```
% Define kinetic parameters individually:
v1=param(1);
v2=param(2);
v3=param(3);

% Calculate rate photosynthesis less photorespiration for day 1
rate_day1_1=mean([v1*Light_1(1:241,1).*Ci_1(1:241,1)./(Kc+Ci_1(1:241,1))-
v2*Light_1(1:241,1).*Oi_1(1:241,1)./(Ko+Oi_1(1:241,1))]);

% Preallocate size of vectors for speed
LAmodel=zeros(size(XV_1));
Light_intercepted=zeros(size(XV_1));
rp=zeros(size(XV_1));
rmr=zeros(size(XV_1));
r1_1d=zeros(size(XV_1));
XVmodel=zeros(size(XV_1));
Cimodel=zeros(size(XV_1));
Oimodel=zeros(size(XV_1));

% Initialize states:
XVmodel(1,1)=XV_1(1);
Cimodel(1,1)=Ci_1(1);
Oimodel(1,1)=Oi_1(1);

for t=1:length(XV_1)-1,

    % Calculate leaf area and light intercepted:
    if t==1
    LAmodel(t,1)=(XVmodel(t,1)-XV_0)*LAR+LA_0;
    else
    LA_inc=(XVmodel(t,1)-XVmodel(t-1,1))*LAR;
    LAmodel(t,1)=LAmodel(t-1,1)+LA_inc;
    end
    Light_intercepted(t,1)=IP_1(t)*(1-exp(-ext_coeff*LAmodel(t,1)/plant_area));

    % If model must be reinitialized at some point (chamber door
    % opening etc)
    if reinitialize==1
    if (time_1(t))>2471000 && (time_1(t))<2488400
       Oimodel(t,1)=Oi_1(t,1);
       Cimodel(t,1)=Ci_1(t,1);
       XVmodel(t,1)=XV_1(t,1);
    end
    end

    % Calculate rate of photosynthesis less photorespiration for dark
    % respiration equation:
    if t==1,
       rate=mean(rate_day1_1);
    elseif t<=240,
       rate_day1_1(t-1,1)=rp(t-1,1);
```

```
        rate=mean(rate_day1_1);
    else
        rate=mean(rp((t-240):(t-1),1));
    end

    % Calculate rate equation and states:
    rp(t,1)=v1*Light_intercepted(t,1)*Cimodel(t)/(Kc+Cimodel(t,1))-
v2*Light_intercepted(t,1)*Oimodel(t,1)/(Ko+Oimodel(t,1));
    rmr(t,1)=v3*rate;
    r1_1d(t)=rp(t,1)-rmr(t,1);
    XVmodel(t+1,1)=XVmodel(t,1)+T*yields(1)*r1_1d(t,1);
    Cimodel(t+1,1)=Cimodel(t,1)+T*u1_1(t,1)/Vchamber-T*r1_1d(t,1)/Vchamber;
    Oimodel(t+1,1)=Oimodel(t,1)+T*yields(2)*r1_1d(t,1)/Vchamber;

end

% Calculate residuals:
residXV=XVmodel-XV_1;
residCi=Cimodel-Ci_1;
residOi=Oimodel-Oi_1;

% Put Ci and Oi in appropriate units:
if graphunits==1,
  Ci_graph=Ci_1*Vchamber*1000000/(44*1185.906);        % Note: 1000000 micromol/mol, 100 puts it in
                                                        % percent, 44g/mol, 1185.906 is the moles of air in
                                                        % the chamber
  Oi_graph=Oi_1*Vchamber*100/(32*1185.906);
  Cimodel_graph=Cimodel*Vchamber*1000000/(44*1185.906);
  Oimodel_graph=Oimodel*Vchamber*100/(32*1185.906);
else
  Ci_graph=Ci_1;
  Oi_graph=Oi_1;
  Cimodel_graph=Cimodel;
  Oimodel_graph=Oimodel;
end


figure(2)
hold off
plot(time_1/(24*3600),XV_1,'-b')
hold
plot(time_1/(24*3600),XVmodel,'-r')
xlabel('Time (days)')
ylabel('Biomass, XV (g)')
legend('estimated','predicted','location','NorthWest')

figure(3)
hold off
plot(time_1/(24*3600),Ci_graph,'-b')
hold
plot(time_1/(24*3600),Cimodel_graph,'-r')
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
xlabel('Time (days)')
if graphunits==1
   ylabel('Carbon Dioxide, Ci (ppm)')
else
   ylabel('Carbon Dioxide, Ci (g/m3)')
end
legend('real','predicted','location','NorthWest')

figure(4)
hold off
plot(time_1/(24*3600),Oi_graph,'-b')
hold
plot(time_1/(24*3600),Oimodel_graph,'-r')
xlabel('Time (days)')
if graphunits==1
   ylabel('Oxygen, Oi (%)')
else
   ylabel('Oxygen, Oi (g/m3)')
end
legend('real','predicted','location','NorthWest')

% Calculate sum of squared errors: (This is the value that will be
% minimized)
val=residXV'*residXV+residCi'*residCi+residOi'*residOi
```

12.2.6 M-file – validate.m

```
function validate(set_param)

global michment T yields Vchamber LAR plant_area ext_coeff LA_0 XV_0 reinitialize
global XV_1 Ci_1 Oi_1 time_1 IP_1 u1_1 Light_1 graphunits

% Define Michaelis-Menten constants individually:
Kc=michment(1);
Ko=michment(2);

% Define kinetic parameters individually:
v1=set_param(1);
v2=set_param(2);
v3=set_param(3);

% Calculate rate photosynthesis less photorespiration for day 1
rate_day1_1=mean([v1*Light_1(1:241,1).*Ci_1(1:241,1)./(Kc+Ci_1(1:241,1))-
v2*Light_1(1:241,1).*Oi_1(1:241,1)./(Ko+Oi_1(1:241,1))]);

% Preallocate size of vectors for speed
LAmodel=zeros(size(XV_1));
Light_intercepted=zeros(size(XV_1));
rp=zeros(size(XV_1));
rmr=zeros(size(XV_1));
r1_1d=zeros(size(XV_1));
XVmodel=zeros(size(XV_1));
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
Cimodel=zeros(size(XV_1));
Oimodel=zeros(size(XV_1));

% Initialize states:
XVmodel(1,1)=XV_1(1);
Cimodel(1,1)=Ci_1(1);
Oimodel(1,1)=Oi_1(1);

for t=1:length(XV_1)-1,

    % Calculate leaf area and light intercepted:
    if t==1
    LAmodel(t,1)=(XVmodel(t,1)-XV_0)*LAR+LA_0;
    else
    LA_inc=(XVmodel(t,1)-XVmodel(t-1,1))*LAR;
    LAmodel(t,1)=LAmodel(t-1,1)+LA_inc;
    end
    Light_intercepted(t,1)=IP_1(t)*(1-exp(-ext_coeff*LAmodel(t,1)/plant_area));

    % If model must be reinitialized at some point (chamber door
    % opening etc)
    if reinitialize==1
    if (time_1(t))>2471000 && (time_1(t))<2488400
      Oimodel(t,1)=Oi_1(t,1);
      Cimodel(t,1)=Ci_1(t,1);
      XVmodel(t,1)=XV_1(t,1);
    end
    end

    % Calculate rate of photosynthesis less photorespiration for dark
    % respiration equation:
    if t==1,
      rate=mean(rate_day1_1);
    elseif t<=240,
      rate_day1_1(t-1,1)=rp(t-1,1);
      rate=mean(rate_day1_1);
    else
      rate=mean(rp((t-240):(t-1),1));
    end

    % Calculate rate equation and states:
    rp(t,1)=v1*Light_intercepted(t,1)*Cimodel(t)/(Kc+Cimodel(t,1))-
v2*Light_intercepted(t,1)*Oimodel(t,1)/(Ko+Oimodel(t,1));
    rmr(t,1)=v3*rate;
    r1_1d(t)=rp(t,1)-rmr(t,1);
    XVmodel(t+1,1)=XVmodel(t,1)+T*yields(1)*r1_1d(t,1);
    Cimodel(t+1,1)=Cimodel(t,1)+T*u1_1(t,1)/Vchamber-T*r1_1d(t,1)/Vchamber;
    Oimodel(t+1,1)=Oimodel(t,1)+T*yields(2)*r1_1d(t,1)/Vchamber;

end
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
% Put Ci and Oi in appropriate units:
if graphunits==1,
   Ci_graph=Ci_1*Vchamber*1000000/(44*1185.906);        % Note: 1000000 micromol/mol, 100 puts it in
                                                        % percent, 44g/mol, 1185.906 is the moles of air in
                                                        % the chamber
   Oi_graph=Oi_1*Vchamber*100/(32*1185.906);
   Cimodel_graph=Cimodel*Vchamber*1000000/(44*1185.906);
   Oimodel_graph=Oimodel*Vchamber*100/(32*1185.906);
else
   Ci_graph=Ci_1;
   Oi_graph=Oi_1;
   Cimodel_graph=Cimodel;
   Oimodel_graph=Oimodel;
end

figure(5)
hold off
plot(time_1/(24*3600),XV_1,'-b')
hold
plot(time_1/(24*3600),XVmodel,'-r')
xlabel('Time (days)')
ylabel('Biomass, XV (g)')
legend('estimated','predicted','location','NorthWest')
text(-5,2200,'(a)')

figure(6)
hold off
plot(time_1/(24*3600),Ci_graph,'-b')
hold
plot(time_1/(24*3600),Cimodel_graph,'-r')
xlabel('Time (days)')
if graphunits==1
   ylabel('Carbon Dioxide, Ci (ppm)')
else
   ylabel('Carbon Dioxide, Ci (g/m3)')
end
legend('real','predicted','location','NorthWest')
text(-5,1900,'(b)')

figure(7)
hold off
plot(time_1/(24*3600),Oi_graph,'-b')
hold
plot(time_1/(24*3600),Oimodel_graph,'-r')
xlabel('Time (days)')
if graphunits==1
   ylabel('Oxygen, Oi (%)')
else
   ylabel('Oxygen, Oi (g/m3)')
end
legend('real','predicted','location','NorthWest')
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP

```
text(-5,24.7,'(c)')

% Calculate residuals and SSE:
residXV=XVmodel-XV_1;
residCi=Cimodel-Ci_1;
residOi=Oimodel-Oi_1;
val=residXV'*residXV+residCi'*residCi+residOi'*residOi
```


This document is confidential property of the MELiSSA partners and shall not be used, duplicated, modified or transmitted without their authorization
Memorandum of Understanding 19071/05/NL/CP