Identifiability Analysis Report

May 7, 2025

1 Introduction

The goal was to determine which parameters in the lettuce growth model are most identifiable and suitable for estimation. To do this, a sensitivity-based identifiability analysis was performed, combining both sensitivity and correlation methods. These steps helped distinguish which parameters should be estimated and which could be fixed, based on their influence on the model output and their degree of correlation with other parameters. After this, a structural analysis of those parameters was done using GenSSI matlab toolbox to assess whether those parameters can, in principle, be uniquely estimated from the model output.

2 Identifiability Analysis

Identifiability analysis is the first step in determing unknown parameters in ODE models. There are several methods to evaluate identifiability in nonlinear ODE models: structural identifiability analysis, practical identifiability analysis and sensitivity-based identifiability analysis. The methods discussed in this section are well described in [2].

2.1 Sensitivity-Based Identifiability Analysis

Sensitivity analysis

A sensitivity analysis was conducted to evaluate the influence of individual model parameters on the total dry weight.

To do this, a nominal simulation was performed over the 40-day period using parameters from Van Henten [1] and initial conditions $X_{nsdw} = 0.5\,\mathrm{g\,m^{-2}}$, $X_{sdw} = 1.0\,\mathrm{g\,m^{-2}}$. After this, each model parameter θ_k was perturbed individually from its nominal value by a perturbation factor of 5% ($p_{pert} = 0.05$). The perturbed output $DW_{pert,k}(t_{eval})$ was recorded for each perturbed parameter.

From these simulations, several time-dependent sensitivity measures were computed for each parameter θ_k at each time point t_{eval}:

- **Absolute Sensitivity**:

$$\frac{\partial DW(t_{eval})}{\partial \theta_k} \approx \frac{DW_{pert,k}(t_{eval}) - DW_{nom}(t_{eval})}{\Delta \theta_k}$$

- **Relative Sensitivity ($rrSI_k$)**: index that relates the relative change in the output DW to the relative perturbation (p_{pert}) applied to the parameter θ_k:

$$rrSI_k(t_{eval}) = \frac{(DW_{pert,k}(t_{eval}) - DW_{nom}(t_{eval}))/DW_{nom}(t_{eval})}{p_{pert}}$$

  The time evolution of $rrSI_k(t_{eval})$ for all parameters are visualized in Figure 2.

- **Scaled Sensitivity (s_k)**: it is a normalized sensitivity value, calculated to serve as the basis for deriving aggregate measures of overall parameter influence. This scaled sensitivity $s_k(t_{eval})$

was computed by scaling by scaling the absolute sensitivity functions with factors related to the parameter magnitudes and typical output values:

$$s_k(t_{eval}) = \left(\frac{\partial DW(t_{eval})}{\partial \theta_k} \right) \frac{\theta_{k,0}}{\text{mean}(DW_{nom}(t_{eval}))}$$

In this formulation, the scaling factor for the parameter is its nominal value $\theta_{k,0}$, and the scaling factor for the output is the mean of the entire nominal output trajectory, $\text{mean}(DW_{nom}(t_{eval}))$.

- **Parameter Importance Indices (ρ_k)**: two overall parameter importance indices were calculated:
  - The root-mean-square importance index:

$$\rho_k^{msqr} = \sqrt{\frac{1}{N_{t_{eval}}} \sum_{i=1}^{N_{t_{eval}}} [s_k(t_{eval,i})]^2}$$

  - The mean-absolute importance index:

$$\rho_k^{mabs} = \frac{1}{N_{t_{eval}}} \sum_{i=1}^{N_{t_{eval}}} |s_k(t_{eval,i})|$$

The ρ_k^{msqr} values are visualized in a bar plot in Figure 1.

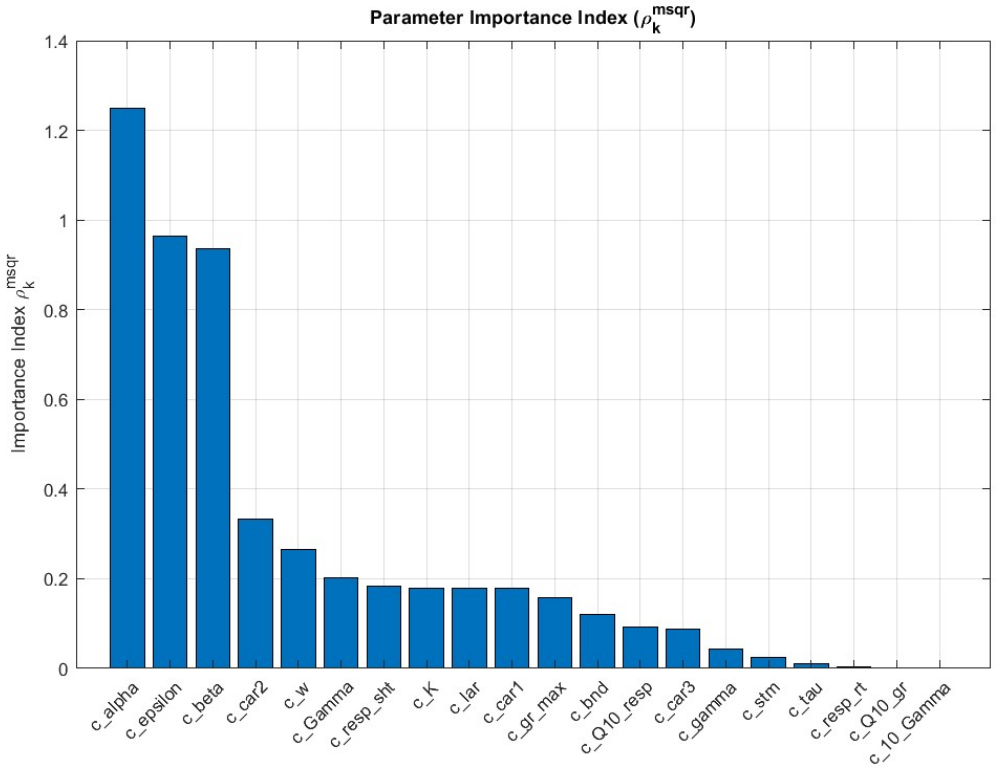

Figure 1: Parameter importance ranking based on the ρ_k^{msqr} index.

Correlation method

Another approach for practical identifiability analysis is to examine the correlations between model parameters. A strong correlation between two parameters indicates that one parameter strongly depends on another parameter and these two parameters cannot be separately estimated.

The methodology involved the following steps:

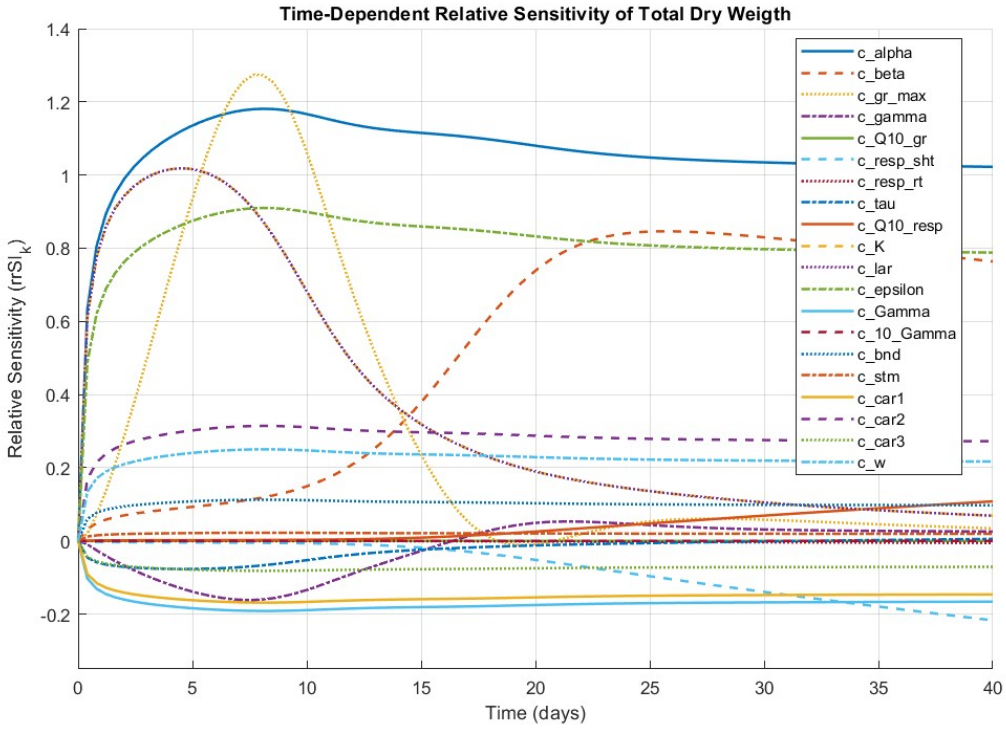

Figure 2: Relative Sensitivity of Total Dry Weigth

1. **Output trajectory:** The total dry weight trajectory, $DW(t) = X_{nsdw}(t) + X_{sdw}(t)$, was calculated over the 40 day simulation period at 100 equally spaced time points (t_{eval}). Inputs and initial conditions were the same as in the sensitivity analysis.

2. **Sensitivity matrix:** A sensitivity matrix $S \in R^{100 \times N_p}$ was computed, where N_p is the number of parameters. Each element S_{ij} represents the sensitivity of the total dry weight at time t_i to a change in parameter p_j:

$$S_{ij} = \frac{\partial DW(t_i)}{\partial p_j} \approx \frac{DW(t_i, p_j + \delta p_j) - DW(t_i, p_j - \delta p_j)}{2\delta p_j}$$

3. **Rank of sensitivity matrix:** theoretical derivations based on a first-order Taylor expansion of the model output show that if this matrix is rank deficient, the parameters are not locally identifiable, indicating linear dependencies among the sensitivity vectors.

   The rank of $S^T S$ was found to be 5, significantly less than the number of parameters. This suggests that the parameters are locally non-identifiable.

4. **Pairwise correlation:** The Pearson correlation coefficient matrix was computed with the MATLAB function $R = \text{corrcoef}(S)$ to measure the correlation between the sensitivity vectors of each pair of parameters. The formula to obtain each element R_{kl} of this matrix, representing the correlation between the sensitivity vector for parameter p_k (column S_k) and parameter p_l (column S_l), is:

$$R_{kl} = \frac{\sum_{i=1}^{n}(S_{ik} - \bar{S}_k)(S_{il} - \bar{S}_l)}{\sqrt{\sum_{i=1}^{n}(S_{ik} - \bar{S}_k)^2}\sqrt{\sum_{i=1}^{n}(S_{il} - \bar{S}_l)^2}}$$

   Here, S_{ik} is the sensitivity to parameter p_k at the i-th time point (t_i), $\bar{S}_k$ is the mean sensitivity for parameter p_k across all $n = 100$ time points ($i = 1, \ldots, n$), and similarly for parameter p_l. A heatmap visualizing the absolute correlation matrix $|R|$ is shown in Figure 3.

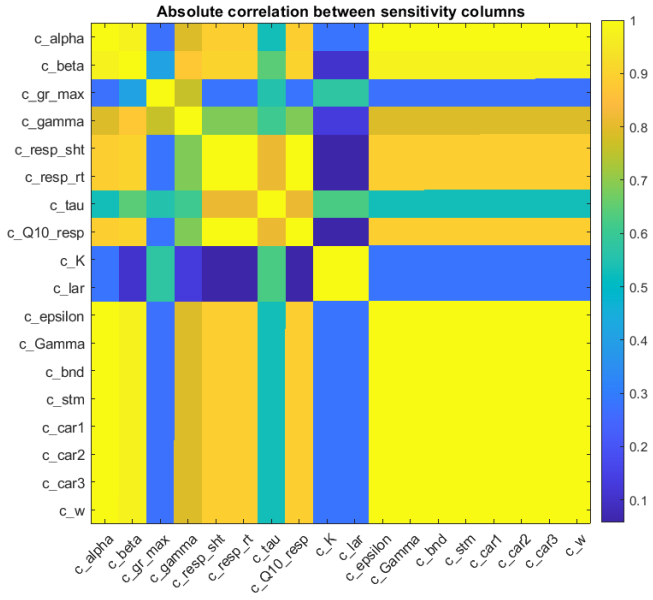

Figure 3: Heatmap of the absolute Pearson correlation coefficients ($|R|$) between parameter sensitivities.

A large number of pair of parameters exhibited very high absolute correlations and some pairs even showed perfect correlation:

- **Light interception parameters c_K vs c_{lar} ($R = 1.0000$)** .

  Because they only influence the model through their product $c_K \times c_{lar}$, their effects are perfectly substitutable, leading to $R = 1.0$. Only the value of the product $c_K \cdot c_{lar}$ can be determined from the model output.

- **Photosynthesis parameters c_α vs c_ϵ, c_Γ, c_{bnd}, c_{stm}, c_{car1}, c_{car2}, c_{car3}, c_w (all $|R| \approx 1.0000$)**.
  the parameters related to CO_2 transport and assimilation are highly correlated. This result can be motivated by looking at the model equations: those parameters interact with one another through a sequence of calculations involving multiplications and combined terms. This means that a change in one parameter early in the sequence can be cancelled out or compensated for by adjusting another parameter later in the sequence, since they results of a sequence of multiplications

- **Growth efficiency parameters c_β vs c_α, c_ϵ, ... ($|R| \approx 0.9769$)**.
  c_β is highly correlated with the photosynthesis parameters. The growth efficiency parameter c_β determines the fraction of assimilated carbon lost to growth respiration:

$$\frac{dX_{\text{nsdw}}}{dt} = \cdots - \frac{1 - c_\beta}{c_\beta} r_{gr} X_{sdw}$$

  It is highly correlated with the photosynthesis parameters because the growth depends strongly on both the *rate* of carbon intake (governed by $c_\alpha, \epsilon, g_{CO_2}$) and the *efficiency* (governed by c_β) with which that carbon is used for structural growth. This means that increasing the photosynthesis rate has a similar positive effect on net growth as increasing the efficiency c_β (which reduces the loss term), which explain their high correlation.

Similar considerations can be made for the other parameters with high correlations.

5. **Total correlation:** A total correlation index (TC_i) was calculated for each parameter p_i by summing the squared correlation coefficients with the other parameters p_j, considering only

pairs where the absolute correlation $|R_{ij}|$ exceeded a threshold $\delta_{tc} = 0.95$:

$$TC_i = \sum_{j \neq i, |R_{ij}| \geq \delta_{tc}} R_{ij}^2$$

These results reinforced the findings from the pairwise analysis. The parameters with the highest values of TC were, as expected, those embedded within the complex sequence of multiplications and dependencies in the photosynthesis calculations. The results, sorted by TC value, are shown in Table 1.

Table 1: Total correlation index (TC) for each parameter, based on pairs with $|R_{ij}| \geq 0.95$.

Parameter	Total Correlation (TC)
c_{car2}	8.9543
c_{ϵ}	8.9543
c_{α}	8.9543
c_{w}	8.9543
c_{car1}	8.9543
c_{bnd}	8.9543
c_{car3}	8.9543
c_{stm}	8.9543
c_{Γ}	8.9543
c_{β}	8.5891
c_{resp_rt}	2.0000
c_{resp_sht}	2.0000
c_{Q10_resp}	2.0000
c_{K}	1.0000
c_{lar}	1.0000
c_{gr_max}	0.0000
c_{γ}	0.0000
c_{τ}	0.0000

2.2 Structural Identifiability Analysis

Structural Identifiability verify system identifiability by analysing the system structure. The basic idea of this approach is to directly use the identifiability definition to verify parameter identifiability. It is used to investigate whether parameters can take on unique, finite, or infinite (thus termed structurally unidentifiable) values for the given model outputs

GenSSI toolbox

The structural identifiability analysis has been performed using GenSSI [3], a software toolbox for structural identifiability analysis of linear and non-linear ODE models. It couples the generating series approach with identifiability tableaus. Using Lie derivatives of the ODE model, a system of equations is generated, the solvability properties of which provide information about global and local structural identifiability as well as non-identifiability.

GenSSI allows for the analysis of the structural identifiability of a subset of parameters, and analyzing all of them can be computationally expensive. A subset of parameters has been chosen based on the outcomes of the previous sensitivity-based analysis. Additionally, as shown in (Table **??**) and the results from the Correlation Method, particularly the low rank of $S^T S$ (Rank 5) and the numerous high pairwise correlations (Figure 3), it was evident that estimating all parameters of the original model is infeasible.

Parameter selection for structural analysis

The parameter selected were: c_{α}, c_{β}, c_{resp_sht}, c_{gr_max}. This selection prioritized parameters that demonstrated significant sensitivity and represented key physiological processes, while parameters that

remained less sensitive or potentially still confounded in the model should be fixed to their nominal values for subsequent analysis.

A structural identifiability analysis using GenSSI was performed and confirmed that the selected parameters are structurally globally identifiable. The script used for the structural identifiability analysis, which includes detailed explanations of the procedures performed, is located in the `structural_analysis\script` folder. The corresponding output file can be found in the `structural_analysis\results` folder.

References

[1] Van Henten, E. J. (1994). Sensitivity analysis of a dynamic growth model of lettuce.

[2] Miao, H., Xia, X., Perelson, A. S., Wu, H. (2021). On Identifiability of nonlinear ODE models and applications in viral dynamics.

[3] Ligon, T. S., Fröhlich, F., Chis, O. T., Banga, J. R., Balsa-Canto, E., Hass, H. (2018). GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models.