Math. Model. Nat. Phenom.
Vol. 8, No. 4, 2013, pp. 112–130
DOI: 10.1051/mmnp/20138407

Development and Evaluation of Plant Growth Models: Methodology and Implementation in the PYGMALION platform

P.-H. Cournède[1*], Y. Chen[1], Q. Wu[1], C. Baey[1], B. Bayol[1]
[1] Ecole Centrale Paris, Laboratoire MAS, Digiplante - 92290 Châtenay Malabry, France

Abstract. Mathematical models of plant growth are generally characterized by a large number of interacting processes, a large number of model parameters and costly experimental data acquisition. Such complexities make model parameterization a difficult process. Moreover, there is a large variety of models that coexist in the literature with generally an absence of benchmarking between the different approaches and insufficient model evaluation. In this context, this paper aims at enhancing good modelling practices in the plant growth modelling community and at increasing model design efficiency. It gives an overview of the different steps in modelling and specify them in the case of plant growth models specifically regarding their above mentioned characteristics.

Different methods allowing to perform these steps are implemented in a dedicated platform PYGMALION (Plant Growth Model Analysis, Identification and Optimization). Some of these methods are original. The C++ platform proposes a framework in which stochastic or deterministic discrete dynamic models can be implemented, and several efficient methods for sensitivity analysis, uncertainty analysis, parameter estimation, model selection or data assimilation can be used for model design, evaluation or application.

Finally, a new model, the LNAS model for sugar beet growth, is presented and serves to illustrate how the different methods in PYGMALION can be used for its parameterization, its evaluation and its application to yield prediction. The model is evaluated from real data and is shown to have interesting predictive capacities when coupled with data assimilation techniques.

Keywords and phrases: good modelling practice, sensitivity analysis, uncertainty analysis, parameter estimation, particle filtering, data assimilation, plant growth models, LNAS model

Mathematics Subject Classification: 00A71, 92B05, 92B15, 93A30, 62F10, 62F15

1. Introduction

Mechanistic modelling of plant growth is a relatively recent research field, and has gained an increasing interest in the last decade, with the sophistication of the description of plant-environment interactions in crop models (also named agro-environmental models) (see for example APSIM [34] or STICS [6]), the development of Functional-Structural Plant Models (FSPM), which combine the description of plant structural development and the biophysical processes of photosynthesis and growth within potentially

*Corresponding author. E-mail: paul-henry.cournede@ecp.fr

© EDP Sciences, 2013

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press
Article published by EDP Sciences and available at http://www.mmnp-journal.org or http://dx.doi.org/10.1051/mmnp/20138407

complex structures (see [63], [29] for recent reviews), and at cellular level the progress made in the description of the complex processes of evolution and geometry of morphogenesis [7].

Some common characteristics can be underlined in the plant growth modelling community and the modelling approaches developed:

- complexity of the models regarding the number of interacting processes described and the number of model parameters,
- difficulty of model parameterization resulting from both model complexity (as specified above) and the costly experimental data acquisition,
- an increasing need in sophisticated computational science, mathematics and statistics to simulate and analyze these systems,
- the large variety of models that coexist in the literature with generally an absence of benchmarking between the different approaches (very few statistical comparisons of different models).

Classical numerical and statistical software solutions (MATLAB®, R - http://www.r-project.org/) that generally provide efficient and sophisticated methods for model analysis and parameter estimation do not however propose a framework which is always adequate for the development and analysis of such complex models: specific data structures (different from vectors or matrices) generally need to be developed by the user, as well as specific encapsulation methods in order to make the generic analysis and estimation methods applicable. Moreover, improving the computational performances (which is generally a necessity when facing complex biophysical models) requires heavy work, and the use of distributed computation to speed up performances is far from straightforward.

On the other hand, other software applications are dedicated to the development of complex models (see for example GroIMP [30] or OpenAlea [53] platforms more specifically in the context of plant growth models), and provide useful tools to build and simulate models, as well as interactive tools for the visualization of results. They prove to be very useful in the first steps of the modelling process, specifically the conceptual work to derive a systematic and analytic approach from the biological understanding of the biophysical processes, and the preliminary validation thanks to powerful visualization tools. However, they do not provide statistical tools for the parameterization and evaluation of models.

Software was also developed coupling both the simulation of a generic model and the implementation of statistical estimation methods for its parameterization (OptimiSTICS package for the STICS model [66] or Digiplant software for the GreenLab model [17], [18]). However, the strong imbrication of a specific model and the methods restrict the potential use and limit the full evaluation of models since they do not allow model comparison or benchmarking.

In this context, our research objective is double. First, we aim at specifying good modelling practices [61] for plant growth models, and more generally for complex biophysical dynamic models, that is to say introducing a proper mathematical methodology to help for a rigorous model development and evaluation, as well as specific methods adapted to this type of models. Second, we develop a software platform dedicated to the easy implementation of this methodology for discrete dynamic models. The platform name is PYGMALION, (Plant Growth Model Analysis, Identification and Optimization). It is developed in C++ and has the following properties:

- it implements a template class for models of dynamical systems, which is generic and simply requires from the user the specification of parameters, state variables, external variables and state functions;
- it proposes a generic and flexible data structures adapted to the observation of complex systems, and particularly biological systems: diversity of observed variables and environmental conditions, irregularity and heterogeneity of observations;
- it implements the statistical methods corresponding to the comprehensive methodology of model development (sensitivity analysis, parameter estimation, model evaluation and selection) and adapted to the specificity of plant growth dynamic models - some methods are original and some others are classical, but all are fully re-developed and adapted to the platform framework;
- it provides a framework for model simulation that makes distributed computation compatible with most expensive computational methods of sensitivity analysis and parameter estimation.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

Our objective in this paper is thus to describe a comprehensive methodology for model development and evaluation, to give an overview of the classical or original methods implemented in the PYGMALION platform for this purpose and adapted to discrete dynamic models of plant growth (and more generally of complex biophysical systems), and to apply the methodology through a typical example: the analysis and parameterization of a new model of plant growth, the LNAS (Log Normal Allocation and Senescence) model, applied here to sugar beet. The model itself is original, while remaining simple enough for the sake of paper clarity. It describes biomass production and allocation processes for crops at compartment level, and can be seen as a simplification of the GreenLab model [71] which describes the same processes at plant organ level. In the sugar beet case, the LNAS model also borrows some ecophysiological concepts from the SUCROS model [25]. The interest of the model is that the experimental data necessary for its parameterization are not too heavy and generally available in classical breeding programs. Moreover, it shows an interesting level of robustness and its relative simplicity regarding its number of state variables and parameters makes it well adapted to data assimilation. However, the methodology and the platform are also suitable for the study of more complex models (with large number of parameters and state equations, as tested for example for the NEMA model [4] or the STICS model [6]), provided of course that the necessary data are available for model evaluation.

The paper is organized as follows: Section 2 introduces the mathematical formalism and the general class of methods of interest in the different steps of model design; Section 3 introduces the LNAS model that will serve as the test case to illustrate how the platform and the implemented methods can be used in the modelling process. Sections 4, 5, 6, 7, 8 present the different steps: respectively sensitivity analysis, parameter estimation, uncertainty analysis, model selection and data assimilation for LNAS; finally Section 9 discusses the perspectives of development for this modelling framework.

2. Basic Concepts in Modelling

2.1. Good Modelling Practice

With the increasing development of modelling in all fields of science, and sometimes a lack of precautions in the way models are developed and used, some authors tried to define and promote good modelling practices, see for example [61] for environmental sciences or [14] in physiology and medicine, by proposing different steps in the modelling process, from conceptual work to model applications. An attempt was made to describe how these different steps could be translated in the context of plant growth modelling [63], even though the methodology was not applied, nor detailed regarding the mathematical steps of the modelling process.

The first step in modelling is the conceptual work that consists in first defining properly the objective of the model, the scientific hypotheses on which it relies, and the mathematical formalism to describe it. Once the equations are written (or the model structure of equations are defined), the following steps of the good modelling practice can be summarized from our point of view into three main steps:

1. Model analysis: this step corresponds to the study of the general model behavior, theoretically or more generally numerically by the way of simulation, and the mathematical properties of the model structure, specifically model identifiability, to determine the necessary experimental data for model parameterization and specification, and parameter sensitivity analysis, to assess the relative importance of the different parameters in the model.
2. Model identification: it is generally based on confronting the model to experimental data. It can include two steps, model structure identification (identify in a family of models which specific structure is the best one) and parametric identification (for a given model structure, estimate the values of model parameters).
3. Model evaluation: this step tries to check qualitatively (by checking the model behavior, its ability to simulate expected phenomena) and quantitatively (by comparing the model outputs to real data) whether the model adequately fulfills its initial objectives. It implies checking the goodness-of-fit for the experimental data set, as well as testing its predictive capacity on a validation experimental data

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

set and assessing the model uncertainty. This step may also imply comparing different models and choosing the best regarding their adequacy to the assigned objectives by the mean of selection criteria.

The modelling process is not linear and implies feedback between the different steps, for example parameter sensitivity analysis can be used in the first modelling step to check the validity of the general structure of model parameterization, but also in the third step after parameter estimation from real experimental data: sensitivity analysis is then applied with the specific model parameter distributions and can thus be used to improve the parameterization by factor fixing, or variance cutting (see [13] for more details).

2.2. Discrete Dynamic Models

The models considered in our study are Markovian models in discrete time, which take values in general state-space as described for example by [12]. Let $(t_n)_{n\in[0;N]}$ be the finite sequence of successive times corresponding to the evolution steps. We denote by $X_n \in \mathbb{R}^d$ the set of characteristic system variables at t_n ($X_n = X(t_n)$), $U_n \in \mathbb{R}^u$, the set of exogenous variables (entries, controls...) at t_n, and $P \in \mathbb{R}^p$, the vector of model parameters. As for most biological systems, X_n may not be fully accessible to observation, thus Y_n, taking values in $\mathbb{R}^{q_n}$, denotes the system observation vector at time t_n. We will denote $\mathbf{Y} = (Y_n)_{0\leq n\leq N}$. The initial density function for X_0 is μ_P, and the Markov transition density at each time step is f_{n,P,U_n}:

$$X_0 \sim \mu_P(.) \text{ and } X_{n+1}|(X_n = x) \sim f_{n,P,U_n}(.|x),\ \forall n \in [0; N-1]\ . \tag{2.1}$$

The observation Y_n depends on the state X_n and the conditional density is given by $g_{n,P}$:

$$Y_n|\,(X_n = x) \sim g_{n,P}(.|x)\ . \tag{2.2}$$

This stochastic framework also includes the deterministic discrete dynamic models, which can classically be written in the form:

$$\begin{cases} X_{n+1} = F_n(X_n, U_n, P) \\ Y_n = G_n(X_n, P) \end{cases}\ . \tag{2.3}$$

It is described in [16] how an important category of plant growth models can be set in this framework. For example, for functional-structural models that describe biomass budget during plant growth (see for example LIGNUM [52] or GREENLAB [46]), the state variables correspond to daily biomass accumulation and to masses of plant organs according to their categories, the parameters are genotype specific, and the external variables U_n correspond to environmental variables (radiation, temperature, soil water content ...).

Generally, not all the state variables can be observed experimentally (for example daily biomass production) and the experimentation being heavy (specifically when it comes to the masses of individual organs), observations are not done at all time steps. If we denote by $\mathcal{O}$ the set of all time step indexes corresponding to observation stages:

$$\mathcal{O} = \{i \in [1; N] \text{ such that } t_i \text{ is an observation time}\} \tag{2.4}$$

we then have $q_i > 0$ if and only if $i \in \mathcal{O}$ (where we recall that q_i is the dimension of Y_i). Note also that the non-zero q_i have no reason to be identical (as illustrated for example in [45] for a model of maize growth, in which at some stages individual plants were measured at organ level, and at other stages only compartment data were available, corresponding to different G_i).

2.3. Mathematical Methods Implemented in PYGMALION

In order to allow modelers to implement fully the different steps described in Section 2.1, generic numerical and mathematical methods were implemented in the PYGMALION platform to make them available to a large variety of models, provided they can be described in the general state space form described in 2.2. These methods are: simulation, parameter estimation, sensitivity analysis, uncertainty analysis, model selection and data assimilation.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

2.3.1. Simulation

It simply consists in computing the sequences $(X_n)_{0\leq n\leq N}$ and $(Y_n)_{0\leq n\leq N}$, given the initial state (or initial probability density), the parameter vector P and the external variables $(U_n)_{0\leq n\leq N}$.

2.3.2. Parameter Estimation

Given the context of a specific experimentation, characterized by the initial condition X_0 and the external variables $(U_n)_{0\leq n\leq N}$, and given $\mathbf{Y}$, the field data from experimentation, the objective of parameter estimation is to estimate the parameters P. Two general approaches are possible and are implemented in PYGMALION.

- The frequentist approach considers the existence of a true parameter vector P, and aims at approximating P from the given sample of data [24]. A classical way is maximum likelihood estimation that consists in finding $\hat{P}$ maximizing $h(P) = v(\mathbf{Y}|P)$, where $v(\mathbf{Y}|P)$ is the likelihood of the experimental data $\mathbf{Y}$ given P, and can be deduced from Equations (2.1,2.2). Under some hypotheses, the maximum likelihood estimator can be equivalent to the ordinary least-square estimator or to the generalized least-square estimator [24]. In PYGMALION, modelers can easily define their own model for the covariance matrix of the model errors. Some simple model of covariance matrices are also available. Different numerical optimization methods are implemented to solve the maximization problem (Quasi-Newton and Gauss-Newton [67] for local maximization, simulated annealing [35] and particle swarm optimization [57] for global maximization).
- The Bayesian approach considers on the contrary that the parameter vector is a random variable [51]. Starting from a prior density distribution $g(P)$ (normally resulting from past studies or expert knowledge), we update this distribution based on the experimental data $\mathbf{Y}$ to compute the posterior distribution $g(P|\mathbf{Y})$. In most cases, the posterior distribution can only be approximated. Unscented Kalman Filtering (UKF) [33] and Convolution Particle Filtering (CPF) [54] are implemented in PYGMALION for this purpose. Both methods aim at the joint estimation of parameters and hidden states. The CPF proves superior in terms of accuracy compared to UKF, even though it is more computationally expensive. CPF is a variant of the classical particle filtering based on convolution kernel approximation techniques. An improvement of this method based on multiple iterations of CPF was proposed in [15]: it takes into account the information obtained by the former iteration in order to improve the precision of the estimates. A conditional version of the method is also derived for stochastic dynamic systems : in the first place, model parameters and hidden states are jointly estimated conditionally on the noise parameters; in the second place, the latter are estimated based on the estimation of the hidden states. Both steps can be repeated until convergence.

Note that it may be interesting to use different sets of experimental data for the identification of the same system. In plant growth modelling, that corresponds for example to different experiments for a similar genotype (different locations, different cultural treatments - with and without water stress for example, different seasons ...). Let k be the number of experimental data sets, and $\mathbf{Y^j}$ the observation vector for experiment j (for all $1 \leq j \leq k$), characterized by initial condition X_0^j and external variables $\left(U_n^j\right)_{0\leq n\leq N^j}$. We thus simply choose the full observation vector $\mathbf{Y}$ as $\mathbf{Y} = \left(\mathbf{Y^1}, \ldots, \mathbf{Y^k}\right)$. Note that the observation functions $g_{n,P}$ or G_n (see Equation 2.2) may be very different for every experiment.

2.3.3. Uncertainty and Sensitivity Analysis

Given the distributions of inputs of the models (parameters, initial conditions, external variables ...), uncertainty analysis aims at computing the distribution of a model output (a state variable or an observation) [48]. Several methods exist, in PYGMALION, uncertainty analysis can be achieved by plain MonteCarlo simulations or unscented transform [33], by propagating the uncertainty in the dynamic system.

Reversely, sensitivity analysis aims at establishing the relative importance of model inputs (parameters, exogenous variables, initial conditions ...) on the output of interest [56]. A large variety of methods exist for this purpose, and can be used according to a typology of models and problems addressed [13], [47].

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

Contrary to local methods, global approaches estimate the effect on the output of a factor keeping all the others varying. Generally, global approaches use model-independent methods while not requiring assumptions of additivity or linearity, and are now widely used.

In the PYGMALION platform, the objective was not to propose a complete panel of the most classical methods but a set of efficient methods covering all the different uses of sensitivity analysis for models of different levels of complexity. For this purpose, we identified 3 methods that efficiently cover most situations.

The Standardized Regression Coefficients method:

The Standardized Regression Coefficients (SRC) are computed from the linear approximation of the model based on Monte Carlo simulations. The method can be viewed as an interesting trade-off between local and global methods, regarding the accuracy of the analysis and the computing cost. Another important index produced by this method is the model coefficient of determination, R^2, which represents the fraction of the output variance explained by the linear regression model itself. It provides an indicator of the degree of nonlinearity of the model, thus representing the level of interaction between parameters and how this interaction contributes to the variance of the output. When $R^2 = 1$, the system is linear and the SRCs can totally explain the variance of the output affected by each factor. Even when models are moderately nonlinear (i.e. > 0.9), the SRCs can provide valid qualitative information, but they are no longer reliable when R^2 gets smaller. If the evaluation cost of the model is important, the SRC method can be used as a first step before resorting to more complex methods if necessary.

Sobol's variance-based method:

The basic concept of this method is to decompose the output variance into the contributions imputable to each input factor, and each combination of input factors [58], [31]. The different types of sensitivity indices that it estimates can fulfill different objectives of sensitivity analysis: factor prioritization, factor fixing, variance cutting or factor mapping [56]. It is a very informative method but potentially computationally expensive [27]. For a given factor X_i, the value of first-order Sobol's index S_i indicates whether a factor is mainly influent, while an important difference between ST_i (Total order effect) and S_i flags an important role of interactions for that factor regarding the output. In this case, inspection of the second order index S_{ij} will help characterize the influence of the interaction of other factors X_j with X_i, $i \neq j$.

Different estimates exist to compute these indexes based on Monte Carlo simulations. In PYGMALION, we use an efficient improvement of Homma-Saltelli's method [31] which allows a control on the estimation error with respect to the number of samples and whose implementation details can be found in [69].

Morris's Method:

[49] proposed an effective screening sensitivity measure to identify the few important factors in models with many factors. The method is based on computing for each input a number of incremental ratios, namely elementary effects, which are then averaged to assess the overall importance of the input. Some improvements of the method were proposed in [11]. Even though the method may lead to errors (specifically in the case of non-monotonic models), it is particularly well-suited when the number of uncertain factors is high and/or the model is expensive to compute and Sobol's method is too prohibitive. For this reason, the method is also implemented in PYGMALION.

2.3.4. Model Selection

For the evaluation and comparison of models, several criteria are computed in the parameter estimation process (AIC, BIC, MDL criteria [8]) and when an independent data set is available for model validation, the Mean Square Error of Prediction (MSEP) can also be computed to measure the predictive capacity of a model [1].

2.3.5. Data Assimilation

If a dynamic model is used to predict some variables, the idea of data assimilation is to incorporate dynamically observations into the model when they are available, in order to improve the model predictive capacity. For this purpose, the method proposed in PYGMALION derives from the filtering approach

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

that can also be used for parameter estimation. Given initial probability density functions for parameters and initial state, we deduce the dynamic evolution of the probability density function of state variables, giving the model prediction. These density functions of parameters and state variables can be updated at each time t_n for which observations Y_n on the system are available. The dynamic evolution of the updated density functions of state variables given the updated probability density functions of parameters provides the updated prediction [23]. This is straightforwardly performed with the UKF and CPF methods described in 2.3.2.

In the following, we will illustrate how these different methods can be used and taken advantage of in the plant growth modelling process.

3. LNAS Model

For the clarity of presentation we choose to illustrate the methodology on a simple model, even though some specific studies were performed in PYGMALION on more complex models (STICS [6], NEMA [4], GreenLab [71]...) with up to 82 parameters for NEMA.

In this section, a Log Normal Allocation and Senescence (LNAS) daily crop model is introduced in its stochastic version (with modelling and measurement errors). The equations are specifically derived for the sugar beet, extrapolating the individual-based model to the field level per unit surface area, with two kinds of organ compartments: foliage and root system. The equations can be adapted to other types of plants without difficulty by specifying organ compartments.

3.1. Model Description

Biomass production: $Q(t)$ is the biomass production on day t per unit surface area ($g.m^{-2}$) and is classically obtained by generalizing the Beer-Lambert law [50]: the fraction of intercepted radiation is given by $\left(1-e^{-\lambda \cdot Q_g(t)}\right)$, with λ ($g^{-1}.m^2$) a parameter and $Q_g(t)$ the total mass of green leaves on day t (in $g.m^{-2}$). The biomass production is then deduced by multiplying the fraction of intercepted radiation with the total amount of absorbed photosynthetically active radiation per unit surface area (PAR, in $MJ.m^{-2}$) and an energetic efficiency μ (in $g \cdot MJ^{-1}\cdot$):

$$Q(t) = \left(\mu \cdot PAR(t)\left(1-e^{-\lambda Q_g(t)}\right)\right) \cdot (1+\eta_Q(t)) \,, \tag{3.1}$$

with the modelling noise $\eta_Q \sim \mathcal{N}(0,\sigma_Q^2)$.

Allocation between the foliage and root compartments: The idea is to simplify the source-sink functions of the GreenLab model [71], [28] for which allocation is computed at organ level. Let $Q_f(t)$ and $Q_r(t)$ represent respectively the total foliage and root masses at day t:

$$Q_f(t+1) = \quad Q_f(t) + \gamma(t)\cdot Q(t) \,, \tag{3.2}$$

$$Q_r(t+1) = Q_r(t) + (1-\gamma(t))\cdot Q(t) \,. \tag{3.3}$$

The function γ is defined as:

$$\gamma(t) = (\gamma_0 + (\gamma_f - \gamma_0)\cdot G_a(\tau(t)))\cdot(1+\eta_\gamma(t)) \,, \tag{3.4}$$

with $\tau(t)$ the thermal time (according to the accumulated daily temperature since emergence day), G_a the cumulative distribution function of a log-normal law parameterized by its median μ_a and standard deviation s_a, and the modelling noise (process noise) $\eta_\gamma(t) \sim \mathcal{N}(0,\sigma_\gamma^2)$.

Senescence: The senescent foliage mass Q_s is a proportion of the accumulated foliage mass determined by another cumulative distribution of a log-normal law parameterized by its median μ_s and its standard deviation s_s:

$$Q_s(t) = G_s(\tau(t)-\tau_{sen})Q_f(t) \,, \tag{3.5}$$

with τ_{sen} the thermal time at which senescence starts. The green foliage mass Q_g is thus easily obtained:

$$Q_g(t) = Q_f(t) - Q_s(t) \,. \tag{3.6}$$

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

3.2. Experimental Data for Model Evaluation

The data used in this paper for the evaluation of the model were obtained from experiments conducted in 2010 by the the French Institute for Sugar Beet Research (ITB) and presented in [2]. Field experiments took place at La Selve, France, N49°34'22", E3°59'24", on a sandy loam soil. A commercial variety, Python, was sown on April 15. The final plant density was estimated at 11.82 plants per square meter (pl.m^{-2}).

Dry matter of root and leaves were collected on 50 plants at fourteen different dates:

$$\mathcal{O} = \{54, 68, 76, 83, 90, 98, 104, 110, 118, 125, 132, 139, 145, 160\},$$

where the indices correspond to days after sowing, see (2.4)). The observation vector Y_n is obtained by averaging each data on the 50 samples and extrapolated at m^2 level by multiplying by the observed density.

Daily mean values of air temperature (°C) and solar radiation (MJ.m^{-2}) were obtained from French meteorological advisory services (Météo France) near the experimental site. Thermal time was computed using a base temperature of 0°C [42].

For the test on predictive capacity and data assimilation in Section 8, we use older data obtained in 2008 in the Beauce plain near Pithiviers, France N48°10'12', E2°15'7 and presented in [43]. A different commercial cultivar, Radar, was sown on April 11. The final plant density was estimated at 10.9 pl.m^{-2}. The climatic data were also obtained from Météo France near the experimental site.

Dry matter of root and leaves were collected on 30 plants at seven different dates:

$$\mathcal{O} = \{39, 60, 67, 75, 88, 122, 158\},$$

where the indices correspond to days after sowing. The observation vector Y_n is obtained by averaging each data on the 30 samples and extrapolated at m^2 level by multiplying by the observed density.

Therefore the observation vector Y is given by the averaged leaf and root masses per m^2. We introduce some multiplicative observation noises:

$$Y|Q_b, Q_r \sim \mathcal{N}\left(\begin{pmatrix} Q_b \\ Q_r \end{pmatrix}; \begin{pmatrix} \eta_b^2 Q_b^2 & 0 \\ 0 & \eta_r^2 Q_r^2 \end{pmatrix}\right). \tag{3.7}$$

4. Sensitivity Analysis

When a model contains a large number of parameters, as it is often the case for models of plant growth in interaction with the environment, parametric identification from experimental data can be difficult or impossible: experimental observations are generally heavy to perform and the available data may not be sufficient to allow the differentiation between the parameter influences, making the model non-identifiable.

Sensitivity analysis methods study how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model inputs [56]. As such, they can serve to quantify the influence of parameters on the model outputs corresponding to the available experimental data. Therefore, sensitivity analysis is classically used to select the parameters to be estimated among those identified as the most influential while those screened as the least influential ones can be fixed to any values in their domains. In the context of sensitivity analysis, this method is called screening or factor fixing ([11], [13]). [55] and [62] describe an application to the crop model STICS.

In the case of dynamic models with multivariate outputs, sensitivity analysis can be applied sequentially at each time step, and for every output. It results in the computation of a large number of indexes from which it can be difficult to select influential parameters, since the influence of the parameters may vary with the time step and the considered output [11]. Several strategies can be used, for example screening as non-influential factors only those that keep indexes below a given threshold for all time steps and for all outputs. [9] introduced a method based on principle component analysis adapted to multivariate

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

sensitivity analysis: sensitivity indexes are only computed on the principal components of the vector output. This method was applied in the context of crop models in [38].

In PYGMALION, we adopt a different strategy. For the purpose of qualitative model analysis, it is interesting for the modeler to get a general view of the evolution of the sensitivity indexes and of the linearity index with time, and for each output. It generally provides useful insight in terms of model diagnosis [70].

On the other hand, for the use of sensitivity analysis in the process of parameter estimation to select the parameters to estimate, we choose to compute sensitivity indexes for an output linked to the chosen estimator, like the likelihood if it is available or a generalized least-square criteria. It is important to note however that in this case the sensitivity measures depend on the available experimental data.

We illustrate these two points - sensitivity analysis for model diagnosis and for model parameterization - on the deterministic version of the LNAS model, by first computing the linearity index with the SRC method and then computing the sensitivity indexes of all parameters with Sobol's method for the generalized least-square criteria.

A crucial point for sensitivity analysis is to define an appropriate distribution for the input factors. It is not always easy when few sets of data exist in the literature, which is typically the case in the construction process of new models, as it is the case for the LNAS model. However, since it bears some similarities with the SUCROS model [25] or the GreenLab model [42], and since some parameters are related to well-known biophysical processes, we managed to give reasonable variation intervals for the parameters (see Table 1) and used the uniform law. The standard deviations of the Log-Normal laws (s_a and s_s) have however large variation intervals corresponding to high uncertainty since few data were available. Different variation intervals of comparable ranges were tested without affecting significantly the results. In Figure 1, the evolution of the linearity index with time is shown for the variables Q_g (green

Parameter	Unit	Variation interval (uniform distribution)
μ	$g.MJ^{-1}$	$[3; 4]$
λ	g^{-1}	$[0.01; 0.014]$
γ_0	-	$[0.7; 1]$
γ_f	-	$[0.; 0.3]$
μ_a	-	$[400; 800]$
s_a	-	$[200; 2000]$
μ_s	-	$[2000; 3000]$
s_s	-	$[3000; 6000]$

TABLE 1. Variation intervals for the parameters of the LNAS model

leaf mass) and Q_r (root mass) corresponding to the experimental data available in our study. We recall here that the smaller the linearity index, the stronger the non-linearity of the dynamic system. Even though the linearity remains relatively high, we can observe a nonlinear stage of growth for both Q_g and Q_r. As observed on other plants like maize [68], this is characteristic of some transition in the biomass allocation process: in the early stage of the growth the priority is given to the vegetative part and later (after canopy closure) to the root part. This result illustrates the kind of insight and model diagnosis that sensitivity analysis can provide. To investigate further, one could try to understand the interactions at stake in this nonlinear growth stage, by computing Sobol indexes: if the difference between the total order and first order indexes of a parameter is non negligible, we may look for important secondary effect indexes involving this parameter.

Regarding the parameterization issue, we compute the first and total order indexes of Sobol's method for the single output corresponding to the Generalized Least Square criteria for a diagonal variance matrix with q distinct diagonal terms, for each biological type of observed data (in our case: $q = 2$, for root mass and leaf mass), [60], [18], as described in Section 5.1. The results are given in Figure 2. We recall here

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

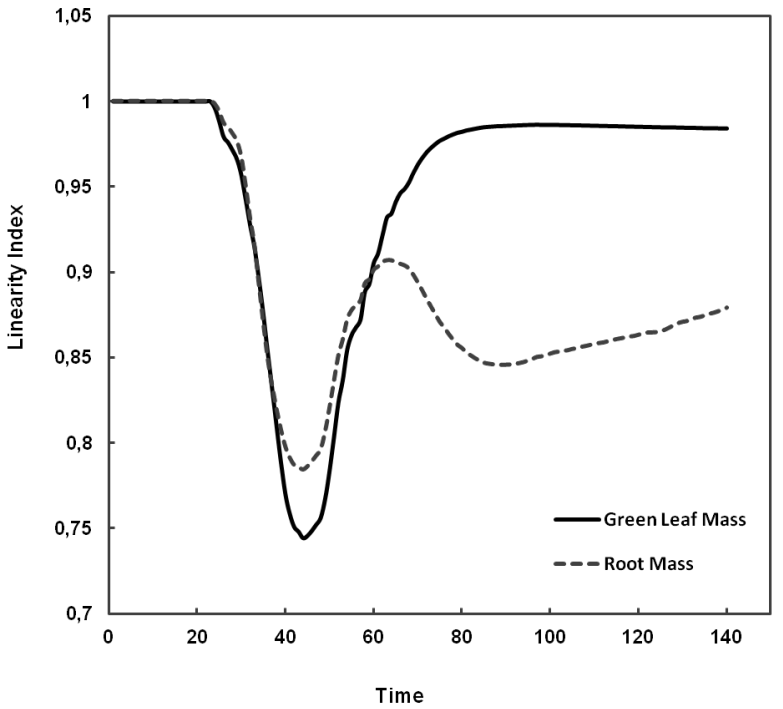


FIGURE 1. Evolution of the linearity index (y-axis) with time (x-axis) for the variables Q_g (green leaf mass) and Q_r (root mass)

that the total order effect ST_i equals the first order effect S_i plus the sum of all the interaction terms involving parameter i. From these results, we screen the parameters s_a, μ_{sen}, s_{sen} since their total order

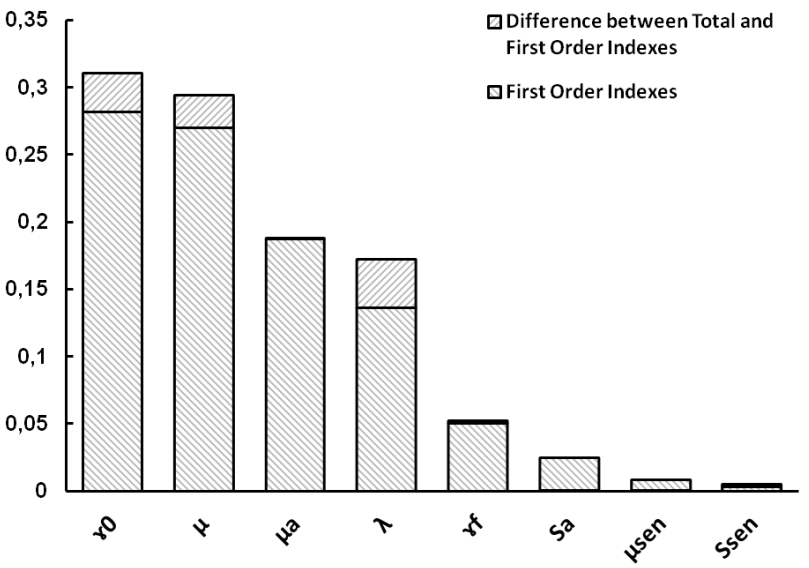


FIGURE 2. Comparison of the first and total order indexes for parameters: μ, λ, γ_0, γ_f, μ_a, s_a, μ_{sen}, s_{sen}.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

indexes are all below 0.02. We fix the parameters to their mean values in the variation intervals. For the senescence parameters, some complementary experimental data can be used to determine an appropriate empirical function. For the five other parameters, their total order effects can not be neglected. Therefore, these parameters should be estimated from experimental data. Moreover, since their first order indexes are also non negligible, no potential identifiability problem is anticipated.

5. Parameter Identification

5.1. Generalized Least Squares

Generalized Least Squares estimation is a very classical method for the estimation of dynamic systems when model state equations are formalized in a deterministic frame (that is to say $V\left(X_{n+1}|X_n\right)=0$), see for example [24]. The underlying error terms correspond to observation errors. It was used in the context of plant dynamic systems in [18] with a simple heteroscedastic error model inspired by [60]. The covariance matrix of the observation error is a diagonal matrix directly deduced from Equation (3.7), we consider a multiplicative error, with all observation variables of the same type sharing the same observation error variance.

In the case of the LNAS model, if $\{t_1, t_2, \ldots, t_N\}$ are all the time steps corresponding to observation stages, we have:

$$\mathbf{Y} = \begin{pmatrix} Q_b\left(t_1\right) \\ Q_r\left(t_1\right) \\ Q_b\left(t_2\right) \\ Q_r\left(t_2\right) \\ \vdots \\ Q_b\left(t_N\right) \\ Q_r\left(t_N\right) \end{pmatrix} (1+\epsilon)\,, \quad \text{with } \epsilon \sim \mathcal{N}\left(0;\Sigma\right), \quad \text{and} \quad \Sigma = \begin{pmatrix} \eta_b^2 & 0 & & \ldots & & 0 \\ 0 & \eta_r^2 & 0 & & & \\ & 0 & \eta_b^2 & 0 & & \vdots \\ \vdots & & 0 & \eta_r^2 & 0 & \\ & & & \ddots & \ddots & \ddots \\ & & & & 0 & \eta_b^2 \quad 0 \\ 0 & & \ldots & & 0 & \eta_r^2 \end{pmatrix}.$$

We use an iterative conditional estimation algorithm: as suggested by [60], we take as first estimates of η_b and η_r, $\hat{\eta}_b^0$ and $\hat{\eta}_r^0$ the standard deviations computed for the experimental observations. We then compute the estimates of the functional parameters $(\hat{\gamma}_0^0, \hat{\mu}^0, \hat{\mu}_a^0, \lambda^0, \hat{\gamma}_f^0)$, conditionnally to $\hat{\eta}_b^0$ and $\hat{\eta}_r^0$, that is to say by minimizing:

$$J(P) = \left(\mathbf{Y} - Y(P)\right)^T \left(\hat{\Sigma}^0\right)^{-1} \left(\mathbf{Y} - Y(P)\right)\,,$$

where

$$Y(P) = \begin{pmatrix} Q_b\left(t_1\right)|P \\ Q_r\left(t_1\right)|P \\ \vdots \\ Q_b\left(t_N\right)|P \\ Q_r\left(t_N\right)|P \end{pmatrix}.$$

We then update the estimates of the error model parameters after convergence of the estimation for the parameters of the functional part of the model (as described in [60], [18]) and repeat the two steps (conditional estimation, update of the error model parameters) until convergence.

The results are shown in Table 2. The approximate standard deviations of the estimated values are obtained from the approximate covariance matrix of the estimates [60]:

$$\Sigma_{\hat{P}} = \left(\left(\frac{\partial Y}{\partial P}\right)^T \hat{\Sigma}^{-1} \left(\frac{\partial Y}{\partial P}\right)\right)^{-1},$$

where $\hat{\Sigma}$ is the error covariance matrix.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

Parameter	Estimated Value	Approximate Standard Deviation (Relative Error in %)
μ	3.59	0.030 (0.84%)
λ	0.0124	0.00120 (9.6%)
γ_0	0.870	0.093 (10.7%)
γ_f	0.0920	0.00466 (5.1%)
μ_a	628	84.7 (13.5%)

TABLE 2. Estimated values and approximated estimation uncertainty evaluation with the GLS estimator

The advantage of the method is its relative simplicity and its efficiency for a quick determination of reasonable estimated values, providing generally acceptable goodness-of-fit. However, the initialization of parameter values for the search algorithm may be an issue. It is implemented in the platform and can be easily adapted to any model.

In the case of dynamical models of plant growth, considering a diagonal covariance matrix for the error model of the observation vector appears as a strong limitation: plant architecture keeps the whole growth history, so it is not realistic to consider that observation errors at different times are independent.

In the frame of Generalized Least Squares approaches, other variance models for the observation errors are possible (and the platform actually allows modellers to specify their own covariance models) with non-diagonal covariance matrix inducing dependence between the error terms (see [5] for examples of more elaborate variance models for GLS estimation in ecological modelling), but that generally implies more parameters to estimate.

5.2. Iterative Convolution Particle Filtering for Hidden Markov Models

As introduced in Section 2.2, the framework of general state space hidden Markov models offers an interesting alternative to take into account error correlations between model outputs, by considering both model errors and observation errors. Several inference methods have been proposed, see for example [12] for a review.

To take into account the specificity of plant growth models (irregular observations, data scarcity, model nonlinearity), we turn to a variant of the regularized particle filtering method introduced by [41], and further improved by for example by [54], [10] (who called it 'convolution particle filter'). The idea is to simultaneously estimate hidden states and unknown parameters, both considered as an augmented state vector. A convolution kernel approximation technique is used to estimate the joint conditional probability density function of (X_k, Y_k) given $(Y)_{[1:k-1]}$ at each prediction-correction step of the filter. In the case of off-line estimation, [15] proposed a method based on the iterative version of convolution particle filtering (Iterative Convolution Particle Filter - ICPF), that can be interpreted as an alternative to smoothing methods [36]. The posterior distribution is used as the prior distribution for the next iteration of the algorithm. A conditional version was also proposed and allowed the estimation of the parameters of the error models, for both process and observation errors. With the parameterized error model, we can deduce confidence intervals for the estimated parameters with parametric bootstrap [22]: virtual experimental data (of the same type) are generated with the stochastic model, and new estimates are computed from these virtual data, allowing an evaluation of the distribution of the estimates. [15] showed that in the case of rare experimental data, such iteration of the filter increased its accuracy.

We show the results in Table 3. The standard deviations and confidence intervals are obtained from 200 estimations of the parametric bootstrap. In Table 4 are also given the parameters of the noise model estimated by the conditional version of the algorithm implemented [15].

When there exists no valuable a priori distribution to initialize the particle sampling in the parameter space, a uniform grid sampling of particles is used. However, in order to improve the convergence

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

Parameter	Estimated Value	Approximate Standard Deviation (Relative Error in %)	95% Confidence Interval
μ	3.55	0.16 (4.5%)	[3.20;3.85]
λ	0.0124	0.00086 (6.9%)	[0.0112;0.0154]
γ_0	0.92.5	0.091 (10.3%)	[0.70;1.06]
γ_f	0.0920	0.0046 (5.1%)	[0.051;0.152]
μ_a	553	86.7 (14.0%)	[474;862]

TABLE 3. Estimated values and confidence intervals for the ICPF estimator when estimating 5 parameters

Parameter	5 Estimated Parameters	6 Estimated Parameters
η_Q	0.010	0.010
η_γ	0.014	0.017
η_b	0.14	0.14
η_r	0.16	0.17

TABLE 4. Estimated values for the noise parameters when estimating 5 or 6 parameters (of the deterministic part of the model)

characteristics of the ICPF method, it is possible in PYGMALION to first compute a GLS estimation and use these results to build a more informative a priori distribution.

Figure 3 shows the fitting results, comparing the simulated and experimental data for green leaf mass Q_l and root mass Q_r.

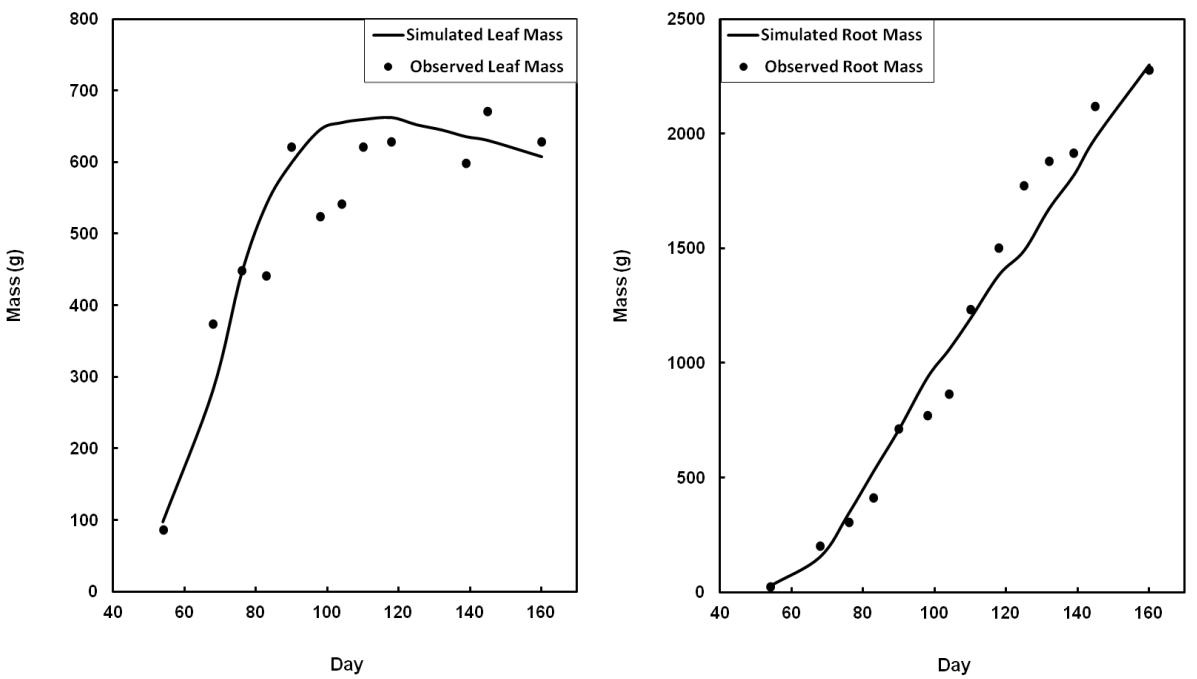


FIGURE 3. Comparison of simulated leaf mass (on the left) and root mass (on the right) with experimental data

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

In order to illustrate the interest of parameter screening as described in Section 4, we show in Table 5 the estimation results when estimating 6 parameters instead of 5 (μ, λ, μ_a, γ_0, γ_f plus s_a). We clearly see an increase of the standard errors of parameter estimates when 6 parameters are estimated, corresponding to a larger estimation uncertainty.

Parameter	Estimated Value	Approximate Standard Deviation (Relative Error in %)	95% Confidence Interval
μ	3.51	0.15 (4.1%)	[3.29;3.86]
λ	0.0124	0.00134 (10.9%)	[0.00980;0.0154]
γ_0	0.945	0.10 (11.8%)	[0.68;1.11]
γ_f	0.0980	0.010 (16.8%)	[0.060;0.123]
μ_a	551	115 (18.0%)	[439;924]
s_a	1073	688 (55.6%)	[322;3000]

TABLE 5. Estimated values and confidence intervals for the ICPF estimator when estimating 6 parameters

In Table 4, we also see that the estimated noise parameters corresponding to the standard deviations of model and observation perturbations are a bit bigger when the 6 parameters (μ, λ, μ_a, γ_0, γ_f plus s_a) were estimated.

6. Uncertainty Analysis

From the estimation results, we can perform an uncertainty analysis: what is the prediction uncertainty on the model outputs regarding the uncertainty of the inputs [19], [65]. Such analysis allows to study the predictive ability of the model: from the covariance matrix of parameter estimates obtained from the parametric bootstrap, we simulate the propagation of the uncertainty in the dynamic system. In order to avoid mixing the effects of variability (resulting from the stochastic model) and uncertainty (resulting from the inaccuracy of parameter estimation), we use the deterministic part of the dynamic model. Here we simply use the direct Monte-Carlo approach even though other more efficient techniques would be possible. As expected, the uncertainty is smaller when only the 5 most important parameters are estimated, illustrating the trade-off between bias and variance, and the importance of the preliminary sensitivity analysis to improve parameterization.

7. Model Selection

In the modelling process, the comparison between models is an important step which is often neglected. It can consist in comparing a new model to a reference one when it exists, or testing different formulations of state equations when describing new processes. Some statistical criteria can be used for model selection. The most famous one is the Akaike Information Criteria and its variants [8], which help characterize the trade-off between bias and variance when fitting a model to real data.

In order to illustrate the usefulness of model selection, we show in Table 7 the evolution of the corrected Akaike Information Criteria (AICc), when estimating from 1 to 6 parameters, according to the importance

	Q_g	Q_r
Model with 5 estimated parameters	87.3 (13.0%)	113 (4.7%)
Model with 6 estimated parameters	96.7 (13.1%)	155 (7.0%)

TABLE 6. Standard deviations of Q_l and Q_r (and relative uncertainty in %) at time 160

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

Number of estimated parameters	1	2	3	4	5	6
AICc	351.5	346.9	346.0	348.2	343.0	346.0

TABLE 7. Corrected AIC for LNAS model with 1 to 6 estimated parameters

	$Q_b\,(t_{122})$	$Q_r\,(t_{122})$	$Q_b\,(t_{158})$	$Q_r\,(t_{158})$
Real Data 2008	374	1559	380	2327
Direct Prediction (relative error in %)	525 (40.4%)	1696 (8.6%)	497 (30.7%)	2399 (3.1%)
Prediction with Data Assimilation (relative error in %)	391 (4.5 %)	1603 (2.8%)	372 (2.1%)	2284 (1.8%)

TABLE 8. Comparison of model prediction with and without data assimilation

ranking obtained with sensitivity analysis, see Figure 2. The parameters that are not estimated are fixed to their mean values in the variation intervals (given in Table 1). The model with 5 estimated parameters is preferred (lower AICc).

Moreover, we also compute the AICc for the model obtained with 5 parameters estimated with GLS (see Section 5.1), and find a value of 344.5, also bigger than 343.0, the value obtained when considering the full hidden Markov model and estimating 5 parameters. The difference is small but still illustrates the interest of the hidden Markov model approach. Other criteria can be computed in PYGMALION, like the mean-square error of prediction (MSEP) testing the predictive capacity of a model, when two independent data sets are available [64]. An example of use can be found in [2] for the benchmarking of several plant models for sugar beet yield prediction.

8. Data Assimilation

The interest of data assimilation for crop models has long been acknowledged, with the use of remote sensing to provide useful observations (like the leaf area index variable) [20, 26, 39]. However, the statistical techniques generally used remain crude, like model re-parameterization or reinitialization. Here we illustrate how the data assimilation technique can be used to improve the predictive capacity of a model.

Based on the parameter estimation results from the 2010 data, we compare the predictive capacity of the model for 2008 experiment with and without data assimilation, for the two last dates of observations t_{122} and t_{158}. The situation is interesting here because the two experiments are quite different: very different locations, different years, slightly different cultivars. The direct prediction (without data assimilation) is obtained by simulating crop growth with the parameters obtained from 2010 experiments and 2008 environmental conditions (climatic variables, density, date of emergence). For the prediction with data assimilation, we use the convolution particle filtering (with 10000 particles) to assimilate the data obtained in approximately the first half of the growth ($\mathcal{O} = \{39, 60, 67, 75, 88\}$) to update state and parameter estimation and then run the simulations until t_{158} for the set of particles. The prediction is given by the mean on all particles. The results are shown in Table 8, for the leaf mass Q_b and the root mass Q_r. The results are clearly improved with data assimilation, and it is particularly spectacular for the leaf mass prediction.

9. Conclusion and Perspectives

This paper described the different steps for a good modelling practice and the statistical and mathematical methods of model analysis and evaluation implemented in the PYGMALION platform for this purpose. An illustration of how the platform can be used in the process of model design was given with a detailed analysis of the LNAS model for sugar beet. This simple stochastic model introduced in this paper

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

describes the major processes of crop growth (biomass production, allocation and senescence), is easy to parameterize (both for the parameters of the biological laws and the parameters of the noise model) and allows a precise quantification of parameter uncertainty. By applying data assimilation techniques, we proved on a real test case its interesting predictive ability in experimental conditions that are very different from those used for model calibration. Of course further experimental data and tests in different situations should be performed to prove the interest of the model, whose principal quality for real applications should rely on its simple parameterization and its adaptation to data assimilation. An application of the model to maize is currently under investigation. Since the model can be seen as a simplification of the GreenLab model, comparison of its performance against GreenLab in terms of predictive capacity would be interesting.

The platform is currently developed and used in the laboratory of Applied Mathematics and Systems at Ecole Centrale Paris, but is also tested by a few other labs. Provided some basic knowledge in C++, it is very easy to implement models and to use the proposed methods: approximately 20 models of plant growth in interaction with the environment and variants are currently implemented in the platform: mostly in the GreenLab family [21] and variants of the STICS [6], CERES [32], SUNFLO [40] models, for different species. As such, it provides an interesting tool in the process of model comparison and benchmarking in the plant growth community as illustrated in [2]. Even if some specific classes of the platform are specifically derived for plant growth models (environment class, specific plant architectural class, some tools to handle experimental data ...) and the platform is so far exclusively used in the plant growth modeling community, its scope is actually broader and could find applications for other discrete dynamic models. A broader diffusion of the platform shall be done after further consolidation tests, and an educative version is also tested in courses on 'Mathematical Modelling' at master level. The detailed description of software conception and engineering of the platform will be published in another article.

The current efforts to improve the platform first concern the implementation of new specific analysis methods. Regarding sensitivity analysis, some works are currently done to extend Sobol's method to correlated inputs and to group analysis [70], a method particularly suited to models of complex systems characterized by several sub-models. In terms of parameter estimation, a more general approach to maximum likelihood estimation is also developed based on a stochastic variant of a generalized EM (Expectation-Maximization) algorithm as described in [59]. An objective is also to be able to handle mixed-effects models in order to describe plant populations and estimate the interindividual variability as illustrated in [3]. Stochastic EM algorithms can also be used for this purpose [37].

From a computational point of view, our objective is to provide a stronger parallel implementation for distributed computing. So far, only the bootstrap algorithm used to compute the confidence interval of estimated parameters (see Section 5.2) and Sobol algorithm (see Section 2.3.3) can benefit from distributed computing via MPI (Message Passing Interface) or OpenMP (Open Multiprocessing) approaches. The objective is to enable the possibility of distributed computation at more elementary levels (each time step for example) which would thus benefit to more methods, specifically the costly particle-based methods like convolution particle filtering (see Section 5.2) or sequential methods for stochastic EM.

More generally, the platform should evolve towards an EDSL (embedded domain-specific languages), with the objective to ease the development of models, the calls of analysis methods and the use of parallel architectures for computation.

Acknowledgements. We are grateful to our research colleagues at ITB (French Institute for Sugar Beet Research) for providing experimental data for LNAS model evaluation.

References

[1] D.R. Anderson. Model based inference in the life sciences. Springer, 2008.

[2] C. Baey, A. Didier, S. Li, S. Lemaire, F. Maupas, P.-H. Cournède. *Evaluation of the predictive capacity of five plant growth models for sugar beet.* 4th international symposium on Plant Growth and Applications (PMA12), Shanghai, China, IEEE, 2012.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

[3] C. Baey, A. Didier, S. Lemaire, F. Maupas, P.-H. Cournède. *Modelling the interindividual variability of organogenesis in sugar beet populations using a hierarchical segmented model*. Ecological Modelling, 263 (2013), 56–63.

[4] J. Bertheloot, P.-H. Cournède, B. Andrieu. *NEMA, a functional-structural model of N economy within wheat culms after flowering: I. Model description*. Annals of Botany, 108 (2011), No. 6, 1085–1096.

[5] B.M. Bolker. Ecological models and data in R. Princeton University Press, 2008.

[6] N. Brisson, C. Gary, E. Justes, R. Roche, B. Mary, D. Ripoche, D. Zimmer, J. Sierra, P. Bertuzzi, P. Burger, F. Bussière, Y.M. Cabidoche, P. Cellier, P. Debaeke, J.P. Gaudillère, C. Hénault, F. Maraux, B. Seguin, H. Sinoquet. *An overview of the crop model STICS*. European Journal of Agronomy, 18 (2003), 309–332.

[7] V. Brukkin, N. Morozova. *Plant growth and development - basic knowledge and current views*. Mathematical Modelling of Natural Phenomena, 6 (2011), No. 2, 1–53.

[8] K.P. Burnham, D.R. Anderson. Model selection and multimodel inference: a practical information-theoretic approach. 2nd edition, Springer Verlag, 2002.

[9] K. Campbell, M.D. McKay, and B.J. Williams. *Sensitivity analysis when model outputs are functions*. Reliability Engineering and System Safety, 91 (2006), No. 10-11, 1468–1472.

[10] F. Campillo, V. Rossi. *Convolution particle filter for parameter estimation in general state-space models*. IEEE Transactions on Aerospace and Electronic Systems, 45 (2009), No. 3, 1063–1072.

[11] F. Campolongo, J. Cariboni, A. Saltelli. *An effective screening design for sensitivity analysis of large models*. Environmental Modelling and Software, 22 (2007), 1509–518.

[12] O. Cappé, E. Moulines, T. Rydén. Inference in hidden Markov models, Springer, New York, 2005.

[13] J. Cariboni, D. Gatelli, R. Liska, A. Saltelli. *The role of sensitivity analysis in ecological modelling*. Ecological Modelling, 203 (2007), 167–182.

[14] E.R. Carson, C. Cobelli. *Modelling methodology for physiology and medicine*. Academic Press, San Diego (US), 2001.

[15] Y. Chen, B. Bayol, C. Loi, S. Trevezas, P.-H. Cournède. *Filtrage par noyaux de convolution itératif*. Actes des 44èmes Journées de Statistique, JDS2012, Bruxelles 21-25 Mai 2012.

[16] P.-H. Cournède. Dynamic system of plant growth. HDR Thesis, University of Montpellier II, 2009.

[17] P.-H. Cournède, M.Z. Kang, A. Mathieu, J.-F. Barczi, H.P. Yan, B.G. Hu, P. de Reffye. *Structural factorization of plants to compute their functional and architectural growth*. Simulation, 82 (2006), No. 7, 427–438.

[18] P.-H. Cournède, V. Letort, A. Mathieu, M.Z. Kang, S. Lemaire, S. Trevezas, F. Houllier, P. de Reffye. *Some parameter estimation issues in functional-structural plant modelling*. Math. Model. Natural Phenom., 6 (2011), No. 2, 133–159.

[19] D.C. Cox, P. Baybutt. *Methods for uncertainty analysis: a comparative survey*. Risk Analysis, 1 (1981), No. 4, 251–258.

[20] L. Dente, G. Satalino, F. Mattia, M. Rinaldi. *Assimilation of leaf area index derived from ASAR and MERIS data into CERES-wheat model to map wheat yield*. Remote Sensing of Environment, 112 (2008), No. 4, 1395–1407.

[21] P. de Reffye, E. Heuvelink, D. Barthélémy, P.-H. Cournède. *Plant growth models*. Ecological Models, Vol. 4 of Encyclopedia of Ecology (5 volumes) (S.E. Jorgensen and B. Fath, eds.), Elsevier, Oxford, 2008, pp. 2824–2837.

[22] B. Efron, R.J. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall / CRC Monographs on Statistics and Applied Probability, 1994.

[23] G. Evensen. Data assimilation: The ensemble Kalman filter. Springer, 2009.

[24] G.C. Goodwin, R.L. Payne. Dynamic system identification: Experiment design and data analysis. Academic Press, New York, 1977.

[25] M. Guérif, C. Duke. *Calibration of the sucros emergence and early growth module for sugar beet using optical remote sensing data assimilation*. European Journal of Agronomy, 9 (1998), 127–136.

[26] M. Guérif, C. Duke. *Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation*. Agriculture, Ecosystems and Environment, 81 (2000), No. 1, 57–69.

[27] J.C. Helton, J.D. Johnson, C.J. Salaberry, C.B Storlie. *Survey of sampling based methods for uncertainty and sensitivity analysis*. Reliability Engineering and System Safety, 91 (2006), 1175–1209.

[28] Y. Guo, Y.T. Ma, Z.G. Zhan, B.G. Li, M. Dingkuhn, D. Luquet, P. de Reffye. *Parameter optimization and field validation of the functional-structural model Greenlab for Maize*. Annals of Botany, 97 (2006), 217–230.

[29] Y. Guo, T. Fourcaud, M. Jaeger, X.P. Zhang, B.G. Li. *Plant growth and architectural modelling and its applications*. Annals of Botany, 107 (2011), 723–727.

[30] R. Hemmerling, O. Kniemeyer, D. Lanwert, G. Buck-Sorlin, W. Kurth. *The rule based language XL and the modeling environment GroIMP illustrated with simulated tree competition*. Functional Plant Biology 35 (2008), No. 10, 739–750.

[31] T. Homma, A. Saltelli. *Importance measures in global sensitivity analysis of nonlinear models*. Reliability Engineering and System Safety, 52 (1996), 1–17.

[32] C.A. Jones, J.R. Kiniry. *Ceres-Maize: A simulation model of Maize growth and development*. Texas A&M University Press, 1986.

[33] S. Julier, J. Uhlmann, H.F. Durrant-Whyte. *A new method for the nonlinear transformation of means and covariances in filters and estimators*. IEEE Transactions on Automatic Control, 45 (2000), No. 3, 477–482.

[34] B.A. Keating, P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J.P. Dimes, M. Silburn, E. Wang, S. Brown, K.L. Bristow, S. Asseng, S. Chapman, R.L. McCown, D.M. Freebairn, C.J. Smith. *An overview of APSIM, a model designed for farming systems simulation*. European Journal of Agronomy, 18 (2003), No. 3-4, 267–288.

[35] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. *Optimization by Simulated Annealing*. Science, 220 (1983), No. 4598, 671–680.

[36] G. Kitagawa. *Monte Carlo filter and smoother for non-gaussian nonlinear state space models*. Journal of Computational and Graphical Statistics, 5 (1996), No. 1, 1–25.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

[37] E. Kuhn, M. Lavielle. *Maximum likelihood estimation in nonlinear mixed effects models*. Computational Statistics and Data Analysis, 49 (2005), No. 4, 1020–1038.

[38] M. Lamboni, H. Monod, D. Makowski. *Multivariate global sensitivity analysis for dynamic crop models*. Field Crops Research, 113 (2009), 312–320.

[39] M. Launay, M. Guérif. *Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications*. Agriculture, ecosystems and environment, 111 (2005), 321–339.

[40] J. Lecoeur, R. Poiré-Lassus, A. Christophe, B. Pallas, P. Casadebaig, P. Debaeke, F. Vear, L. Guilioni. *Quantifying physiological determinants of genetic variation for yield potential in sunflower. SUNFLO: a model-based analysis*. Functional Plant Biology, 38 (2011), 246–259.

[41] F. Legland, C. Musso, N. Oudjane. *An analysis of regularized interacting particle methods for nonlinear filtering*. 3rd IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, Prague, 1998.

[42] S. Lemaire, F. Maupas, P.-H. Cournède, P. de Reffye. *A morphogenetic crop model for sugar-beet (Beta Vulgaris l.)*. Crop Modeling and Decision Support, (W. Cao, J. White, E. Wang, eds.), Springer, 2009, pp 116–129.

[43] S. Lemaire, F. Maupas, P.-H. Cournède, J.-M. Allirand, P. de Reffye, B. Ney. *Analysis of the density effects on the source-sink dynamics in sugar-beet growth*. 3rd international symposium on Plant Growth and Applications(PMA09), Beijing, China (B.-G. Li, M. Jaeger, Y. Guo, eds.), IEEE Computer Society (Los Alamitos, California), Novem. 9-12 2009.

[44] C. Loi, P.-H. Cournède. *Generating functions of stochastic L-systems and application to models of plant development*. Discrete Mathematics and Theoretical Computer Science Proceedings, AI (2008), 325–338.

[45] Y. Ma, M.P. Wen, Y. Guo, B.G. Li, P.-H. Cournède, P. de Reffye. *Parameter optimization and field validation of the functional-structural model GreenLab for maize at different population densities*. Annals of Bot., 101 (2008), 1185–1194.

[46] A. Mathieu, P.-H. Cournède, V. Letort, D. Barthélémy, P. de Reffye. *A dynamic model of plant growth with interactions between development and functional mechanisms to study plant structural plasticity related to trophic competition*. Annals of Botany, 103 (2009), 1173–1186.

[47] H. Monod, C. Naud, D. Makowski. *Uncertainty and sensitivity analysis for crop models*. Working with Dynamic Crop Models (D. Wallach, D. Makowski, J.W. Jones, eds.), Elsevier, 2006, pp. 55–100.

[48] M.G. Morgan, M. Henrion, M. Small. Uncertainty. Cambridge University Press, 1990.

[49] M.D. Morris. *Factorial sampling plans for preliminary computational experiments*. Technometrics, 33 (1991), 161–174.

[50] T. Nilson. *A theoretical analysis of the frequency of gaps in plant stands*. Agricult. and Forest Meteorol., 8 (1971), 25–38.

[51] A. O'Hagan, J.J. Forster. Kendall's advanced theory of statistics: Bayesian inference. Arnold, London, 2nd edit., 2004,

[52] J. Perttunen, R. Sievänen, E. Nikinmaa, H. Salminen, H. Saarenmaa, J. Vakeva. *Incorporating Lindenmayer systems for architectural development in a functional-structural tree model*. Ecological Modelling, 181 (2005), 479–491.

[53] C. Pradal, S. Dufour-Kowalski, F. Boudon, C. Fournier, C. Godin. *OpenAlea: a visual programming and component-based software platform for plant modelling*. Functional Plant Biology, 35 (2008), No. 10, 751–760.

[54] V. Rossi, J.-P. Vila. *Nonlinear filtering in discrete time: A particle convolution approach*. Annales de l'Institut de Statistique de l'Université de Paris, 50 (2006), No. 3, 71–102.

[55] F. Ruget, N. Brisson, R. Delécolle, R. Faivre. *Sensitivity analysis of a crop simulation model, STICS, in order to choose the main parameters to be estimated*. Agronomie, 22 (2002), 133–158.

[56] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola. Global sensitivity analysis. The primer ed., John Wiley&Sons, 2008.

[57] Y.H. Shi, R. Eberhart. *A modified particle swarm optimizer*. Evolutionary Computation Proceedings (IEEE World Congress on Computational Intelligence) (K.R. Belew, L.B. Booker, eds.), Morgan Kaufmann, 1998, pp. 69–73.

[58] I. Sobol. *Sensitivity analysis for non-linear mathematical models*. Math. Model. Comput. Experim., 1 (1993), 407–414.

[59] S. Trevezas, P.-H. Cournède. *A sequential Monte Carlo approach for MLE in a plant growth model*. Journal of Agricultural, Biological, and Environmental Statistics, 18 (2013), No. 2, 250–270.

[60] W. Taylor. *Small sample properties of a class of two-stage Aitken estimator*. Econometrica, 45 (1977), No. 2, 497–508.

[61] R.H. Van Waveren, S. Groot, H. Scholten, F. Van Geer, H. Wosten, R. Koeze, J. Noort. *Good modelling practice handbook*. Tech. Report 99-05, STOWA, Utrecht, RWS-RIZA, Lelystad, The Netherlands, 1999.

[62] H. Varella, S. Buis, M. Launay, and M. Guérif. *Global sensitivity analysis for choosing the main soil parameters of a crop model to be determined*. Agricultural Sciences, 3 (2012), 949–961.

[63] J. Vos, J.B. Evers, G.H. Buck-Sorlin, B. Andrieu, M. Chelle, P.H.B. de Visser. *Functional-structural plant modelling: a new versatile tool in crop science*. Journal of Experimental Botany, 61 (2010), No. 8, 2101–2115.

[64] D. Wallach, B. Goffinet. *Mean Squared Error of Prediction in Models for Studying Ecological and Agronomic Systems*. Biometrics, 43 (1987), No. 3, 561–573.

[65] D. Wallach, B. Goffinet, J.-E. Bergez, P. Debaeke, D. Leenhardt, J.-N. Aubertot. *The effect of parameter uncertainty on a model with adjusted parameters*. Agronomie, 22 (2002), 159–170.

[66] D. Wallach, S. Buis, P. Lecharpentier, J. Bourges, P. Clastre, M. Launay, J.-E. Bergez, M. Guérif, J. Soudais, E. Justes. *A package of parameter estimation methods and implementation for the STICS crop-soil model*. Environmental Modelling and Software, 26 (2011), 386–394.

[67] E. Walter, L. Pronzato. Identification de modèles paramétriques. Masson, Paris, 2006.

[68] Q. Wu, P.-H. Cournède. *Sensitivity analysis of Greenlab model for Maize*. 3rd international symposium on Plant Growth and Applications(PMA09), Beijing, China (B.G. Li, M. Jaeger, Y. Guo, eds.), IEEE, November 9-12 2009.

[69] Q. Wu, P.-H. Cournède, A. Mathieu. *An efficient computational method for global sensitivity analysis and its application to tree growth modelling* Reliability Engineering and System Safety, 107 (2012), 35–43.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press

[70] Q. Wu, P.-H. Cournède. *A comprehensive methodology of global sensitivity analysis for complex mechanistic models: An application to plant growth*. Submitted, (2013).

[71] H.P. Yan, M.Z. Kang, P. de Reffye, M. Dingkuhn. *A dynamic, architectural plant model simulating resource-dependent growth*. Annals of Botany, 93 (2004), 591–602.

https://doi.org/10.1051/mmnp/20138407 Published online by Cambridge University Press