Data assimilation to reduce uncertainty of crop model prediction with convolution particle filtering

Yuting Chen, Paul-Henry Cournède

Data Assimilation to reduce Uncertainty of Crop Model Prediction with Convolution Particle Filtering

Yuting CHEN[a,*], Paul-Henry COURNEDE[a]

[a] *Laboratoire MAS, Ecole Centrale Paris, Grande Voie des Vignes, 92290 Châtenay-Malabry, France*

Abstract

A three-step data assimilation approach is proposed in this paper to enhance crop model predictive capacity in various environmental conditions. The most influential parameters are first selected by global sensitivity analysis and then estimated in a Bayesian framework. The posterior distribution of the estimation step is then considered as prior information for data assimilation. In this last step, a filtering method is sequentially applied to update state and parameter estimates, with the purpose of improving model prediction and assessing the prediction uncertainty.

The estimation and assimilation steps are based on the Convolution Particle Filtering, whose features make it particularly suitable for data assimilation in crop models: the method is easy to adapt to any general state-space models (both probabilistic and deterministic ones) with very few tuning parameters, no approximation needs to be made for nonlinear models, and it remains robust in situations with irregular and sparse datasets.

With the aim of illustrating the robustness and adaptive capacity of the proposed approach, its predictive performance is evaluated with two crop models, the STICS model for winter wheat and the LNAS model for sugar beet. The two models are built with different perspectives. STICS is deterministic and provides a very detailed description of the ecophysiological processes driving crop-environment interactions, while LNAS is designed to describe only the essential ecophysiological processes of plant biomass budget in a probabilistic framework, so as to put emphasis on the uncertainty assessment.

In order to evaluate the approach, five datasets obtained in various experimental conditions were used for the sugar beet LNAS model, and three datasets for the winter wheat STICS model. In both studies, one dataset was used for *a priori* parameter estimation and the others were used to test the model predictive capacity, both with and

without data assimilation. The CPF-based data assimilation approach showed promising predictive capacity and provided robust and reduced credibility intervals in various test configurations (different years for calibration and prediction by assimilation, different experimental sites, different cultivars, different crop densities, different levels of water stresses), which suggests that the combination of such an approach with both types of crop models (simple probabilistic model or complex deterministic model) is quite reliable and can therefore be regarded as a potential tool for yield prediction applications in agriculture.

Keywords: parameter estimation, data assimilation, dynamic crop model, convolution particle filtering, uncertainty analysis, sugar beet, winter wheat, LNAS, STICS, yield prediction.

---

1. Introduction

To improve the predictive capacity of plant growth models in various environments has been a long-standing challenge. A common idea is to enrich the mechanistic description of plant ecophysiology (Yin and Struik, 2010). With this purpose, particular efforts have been made to take into account abiotic stresses regarding temperature (Fowler et al., 2003), water (Tardieu, 2003), or Nitrogen (Bertheloot et al., 2011). Some advanced agro-environmental models even aim at addressing the full diversity of environmental variations, like STICS (Brisson et al., 2003) or APSIM (Keating et al., 2003). However, the complexity of the interaction between processes can make the task rather difficult, particularly in the case when several stresses are involved (Mittler, 2006). As described by Yin and Struik (2010), the tendency is still to complicate the mechanistic description of biophysical processes, even by linking ecophysiology to "omics" sciences as an attempt for the full comprehension of the regulatory networks from which plant robustness and plasticity is supposed to emerge (Hirai et al., 2004). This direction is clearly leading the way to great advances in research, especially in extending our understanding of how

*Corresponding author. Tel.: (33) 1 41 13 15 59
Email addresses: yuting.chen@ecp.fr (Yuting CHEN), paul-henry.cournede@ecp.fr (Paul-Henry COURNEDE)

genotype leads to phenotype (Buck-Sorlin and Bachmann, 2000; Hammer et al., 2006; Yin and Struik, 2010).

However, the more complex the models are, the more troublesome their parameterization and the assessment of the estimate uncertainty become (Ford and Kennedy, 2011; Chen and Cournède, 2012), specifically due to the costly experimentation and the great number of unknown parameters to consider. Likewise, local environmental conditions (in terms of climatic and soil variables, as well as biotic stresses) and initial conditions in specific fields are also very delicate to characterize. Consequently, it may raise important issues regarding the identifiability of the parameters, the assessment of the confounding noises and the propagation of uncertainty and errors related to both parameters and inputs of these dynamic models. Failing to address these issues may finally result in poor predictions of plant-environment interactions in real situations, that is to say the opposite of the pursued objective.

Under these circumstances, an alternative pragmatic approach has been proposed for the purpose of crop growth prediction in specific farming conditions: the combination of a simplified crop model and sequential data assimilation technique to update the model variables and / or parameters from observed data in the early stages of growth (Bouman, 1992; Delécolle et al., 1992; Maas, 1988; Moulin et al., 1998). This approach was particularly studied allowing the progress in deriving biophysical and biochemical canopy state variables from optical remote sensing (Dorigo et al., 2007), which may potentially give way to crop production forecast at large scales (Moran et al., 1997) and thus be considered as a tool for decision support (Gabrielle et al., 2002; Houlès et al., 2004).

The conventionally used strategy is to consider reference models like SUCROS (Guérif and Duke, 1998, 2000; Launay and Guérif, 2005) or CERES (Dente et al., 2008) as the framework to integrate the remotely sensed observations. Several methods were developed in this perspective (see Dorigo et al. (2007) for a review). The forcing method consists in replacing a state variable of the model by the observed data, for instance the Leaf Area Index (LAI) in (Delécolle et al. (1992); Dente et al. (2008)). One important drawback is that generally a considerable part of the model state variables cannot be or are not observed and thus cannot be updated simultaneously at each time step. More-

over, the method does not take into account the observation error, which should not be neglected considering the general lack of accuracy of remote sensing data. Another possibility is to use the available observation data to recalibrate some model parameters and / or initial states that may presumably vary with local conditions (Bouman, 1992; Guérif and Duke, 2000; Launay and Guérif, 2005). The main limitation of this method is that it requires sufficient data to perform the calibration, while we would prefer to benefit directly from the data assimilation technique based on the early growth stages with regular updates when observation data are available. Besides, the global approach of this calibration step usually fails to capture and to maintain the system dynamics.

In other research domains, data assimilation problems have been commonly reformulated and studied with a Bayesian probabilistic perspective, which allows the sequential estimation of model states and parameters simultaneously (Van Leeuwen and Evensen, 1996; Jazwinski, 1970) in the framework of generalized state-space models. It permits us to circumvent the above issues. In the light of these former applications, the first attempt to adapt a relatively simple crop model into this perspective was made by Makowski et al. (2004). The method implementation relies on a probabilistic framework of crop model which is used to derive prior distributions of the model state variables and parameters at time steps with available observations while taking into account uncertainty in model prediction. Conditionally to the experimental observations and the observation model error, posterior distributions are deduced according to Bayes' law. An updated prediction of the model state variables can thus be inferred. The procedure is repeated at all measurement dates. Classical filtering methods used for this purpose are Ensemble Kalman Filter (see Evensen (2006) for the general presentation of the method, or Jones and Graham (2006) for an application in the context of crop models) or Particle Filter (see for example Kitagawa (1996) for the general concepts or Naud et al. (2007) for an application in the context of crop models).

Nonetheless, one of the difficulties to implement this approach comes from the fact that it requires the plant growth model described in a probabilistic framework, as a hidden Markov model (Cappé et al., 2005). The classical and complex crop models (like STICS (Brisson et al., 1998), APSIM (Keating et al., 2003), CERES (Jones and Kiniry, 1986), etc.) were not built in this perspective and their stochastic reformulation is therefore

far from straightforward: the large number of involved processes may potentially lead to a drastic increase in the number of parameters to model process errors. One simple solution to circumvent this problem is to only consider observation errors (Guérif et al., 2006), but it may hinder a proper update of hidden state variables.

In this context, the objective of this paper is to propose an alternative approach to crop yield prediction with data assimilation, which would further be robust, efficient and adapted to the specific characteristics of crop models (nonlinear dynamics, restricted and irregular observation data).

Although the literature on filtering methods is considerably rich (Extended, Unscented, Ensemble Kalman Filter or Particle Filter ...), the Convolution Particle Filter (CPF) (Campillo and Rossi, 2009; Rossi and Vila, 2006) which can be regarded as a generalization of the regularized particle filter proposed by Musso and Oudjane (1998), stands out for its attractive features regarding the challenges raised by parameter estimation and data assimilation of crop models. Firstly, the method is not only rather easy to adapt (with very few tuning parameters), but also robust in terms of convergence since it circumvents the classical problem of potential sample degeneracy in particle filters. This property is valuable in real situations for which irregular or heterogeneous field data are available. Moreover, it does not rely on the Gaussian assumption of distributions as the Kalman Filter-based algorithms, and is thus adapted to the potentially highly nonlinear plant / crop models. When these models are formalized as general state-space hidden Markov models, CPF can achieve a proper evaluation of model uncertainty. Another interesting feature is that it works as well with deterministic models, which makes the method straightforwardly adaptable to the classical and widely used crop models.

Therefore, in this paper, a three-step data assimilation approach based on the Convolution Particle Filtering is proposed and tested based on real experimental data. The most influential parameters are first selected and estimated in a Bayesian framework from a calibration data set. The obtained estimation along with the evaluated uncertainty is considered as prior information for the data assimilation step. With the purpose of improving model prediction and assessing the prediction uncertainty, the filtering method is sequentially applied again to update state and parameter estimates on a second data set.

To illustrate the robustness of the proposed data assimilation approach, we applied it to two models of different types. The first one is the LNAS (Log-normal allocation and senescence) model for sugar beet, describing biomass budget during crop growth, with the particularity of being fully built in a probabilistic perspective (Chen and Cournède, 2012; Cournède et al., 2013) for the purpose of data assimilation. Based on the analysis of Delécolle et al. (1992), the model describes only the major ecophysiological processes (at least in terms of Carbon economy): biomass production, biomass allocation, senescence and leaf surface development. Such a simplification allows an easier representation of the model errors without increasing significantly the number of parameters.

Note that with the aim of improving the parameterization performance, especially in the case of rare or irregular experimental data, a conditional Iterative version of the Convolution Particle Filtering (ICPF) is proposed by Chen et al. (2012). In the parameter estimation step, the ICPF approach allows the estimation of the functional model parameters (in the deterministic part of the model) and both modelling and observation error distributions. Evaluation of the parameter uncertainty is also derived, which in turn provides a reliable *a priori* for data assimilation.

The second model to which the data assimilation method is applied is STICS (Simulateur mulTIdisciplinaire pour les Cultures Standard) (Brisson et al., 1998, 2008) which is a classical and generic crop model, with a rich description of crop-environment interactions. We consider in this study the STICS model for winter wheat. Contrary to LNAS, STICS is a purely deterministic crop model. It will prove interesting to see how the method also adapts to this case.

In Section 2, we first present our global approach in the framework of general state-space models: parameter screening based on Sobol sensitivity analysis, estimation of selected parameters to determine prior distributions by the Convolution Particle Filter and an iterative version of this algorithm, and finally data assimilation. The application of the method to the LNAS model of sugar beet is presented in Section 3, and to the STICS model for winter wheat in Section 4. Finally, the conclusion suggests some research perspectives.

2. A Method for Data Assimilation in Crop Models

2.1. General State-Space Models

For parameter estimation and sequential data assimilation in crop models, we rely on the statistical framework provided by the discrete nonlinear general state-space model, with a state function and an observation function as follows:

$$\begin{cases} X(t+1) = f\left(X(t), \Theta, \eta(t), t\right), \\ Y(t) = g\left(X(t), \Theta, \xi(t), t\right). \end{cases} \tag{1}$$

The evolution equation is embodied in the function f, which is time dependent. $X(t)$ represents the state variables at time t, Θ is a vector of parameters of dimension p and the modelling noise is represented with the random variables $\eta(t)$ (corresponding to model imperfections or uncertainty in the model inputs). The observation equation incorporates observations on the state variables of interest. $Y(t)$ is the output vector which is related to the state variable vector $X(t)$ through the function g. $Y(t)$ consists of variables that can be observed experimentally and usually differs from $X(t)$ (for instance, biomasses of some plant organs can be measured while the daily biomass production cannot). Measurement noises are denoted by $\xi(t)$. $(\eta(t))_t$ and $(\xi(t))_t$ are considered as sequences of independent and identically distributed random variables. Since experimental observations are usually limited due to high costs, observations are only available at irregular times. Let $(t_1, t_2, ...t_N)$ be the N measurement time steps. For all $n \in [1; N]$, we set: $X_n := X(t_n)$, $Y_n := Y\left(t_n\right)$ and $Y_{1:n} := \left(Y\left(t_1\right), Y\left(t_2\right), \ldots, Y\left(t_n\right)\right)$.

Note that only in rare occasions (Makowski et al., 2004; Chen and Cournède, 2012; Trevezas and Cournède, 2013), plant models were built by really taking into account modelling and measurement noises. Models are generally written as deterministic dynamic systems. Of course, such deterministic models can still be represented with (1), the stochastic variables being zero with probability 1.

2.2. Parameter Selection, Estimation and Data Assimilation

In this paper, we consider cases in which no satisfying prior distributions are available for the considered parameters, so that a first estimation step is performed from a full dataset to provide appropriate prior distributions. Afterwards, the obtained prior

distributions are used for data assimilation applications during which experimental data of early growth stages are assimilated to improve model prediction at later stages. Of course, the preliminary parameter estimation step can be skipped if satisfying prior distributions are available.

Under these circumstances, to achieve the prediction objective, we thus propose an approach in three steps: parameter selection, parameter estimation, data assimilation.

In parameter selection, the least influential model parameters are screened thanks to sensitivity analysis methods (Campolongo et al., 2007) based on the deterministic version of the model. The remaining parameters (most influential ones) are therefore chosen to be estimated.

In parameter estimation, a simple CPF or conditional iterative CPF approach is carried out on the calibration dataset. For the latter as explained in section 2.5 and section 2.6, the filtering process is iterated: both selected model parameters and model hidden states are estimated, while the noise parameters are supposed known. Once the convergence of all the parameter estimates is achieved, both the modelling and measurement noises are empirically evaluated from the estimation of the (deterministic) model parameters and hidden states variables. We then update the new noise parameters, and the iterative CPF approach is performed again to obtain more coherent model parameter estimations. The alternation of these two steps repeats until the convergence of both the model parameters and the noise parameters. Regarding the uncertainty assessment, parameter bootstrap is performed to provide a proper evaluation of the uncertainty in model parameters.

In data assimilation, a new comparable experimental dataset (for instance same type of crop but observed in a different year, or at a different location, or for a different genotype) with few early measurements is introduced. With the purpose of performing predictions of yield or other state variables of interest, the CPF approach is anew carried out by regarding the results of the estimation step as prior information. In this step, the probability density is represented by a large number of samples (particles) which evolve with time. Hence, after a short readjustment period while model parameters and state variables are updated based on the available measurements, the particles continue to propagate so as to forecast the system evolution and to evaluate the uncertainty related

to the state variables of interest.

2.3. Parameter Screening by Sensitivity Analysis

When a model contains a large number of parameters, as it is often the case for plant growth models, parameter estimation based on limited experimental data is usually considered to be a key issue which may affect strongly the quality of model prediction with important estimates uncertainty. Therefore, sensitivity analysis is classically applied in advance to select the most influential parameters to be estimated, whereas 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" (Campolongo et al., 2007).

With this purpose, we implement the algorithm proposed by Wu et al. (2012) to compute Sobol's indices (first order and total order) for all considered parameters, choosing as output a generalized least-squares criterion for the parameter selection.

An important issue in sensitivity analysis is the determination of the distributions representing the uncertainty in the inputs, particularly when the parameters are empirical with no explicit biological meaning, and specific methods have to be devised (Wernsdörfer et al., 2008). In the LNAS model, all the functions used represent well-known processes for sugar beet, with an important literature, so that it was relatively easy to assess proper variation intervals for the parameters. Likewise, several varieties of winter wheat have already been modelled with the STICS model so that different values were available in literature (Brisson et al., 2008), from which we could deduce intervals of variations. The question for us is the following: for each parameter, what is the reasonable range of possible values, or when calibrating our system what would be acceptable values for the parameters? In regard to such criteria, it also seems appropriate to select the uniform distribution after assessing the appropriate variation intervals. Moreover, some tests were performed by increasing the range of the least important parameters while decreasing the range of the most important ones to check the possible bias induced by the choice of the ranges, but it did not affect the importance ranking of the parameters, which is the result of interest for parameter screening.

2.4. Convolution Particle Filter for Bayesian Parameter Estimation

Particle filter is a recursive Bayesian filter based on Monte Carlo simulations (Arulampalam et al., 2002). The basic idea is the recursive approximation of the filtering distribution by a time evolving weighted sample.

Inspired by the post-Regularized Particle Filter (Oudjane and Musso, 1998), the objective of the Convolution Particle Filter (Rossi and Vila, 2006; Campillo and Rossi, 2009) is to estimate jointly the parameters and the hidden states of the dynamic system by processing the data online. When the representation of the state-space model is better described in terms of conditional distributions, its formulation as a hidden Markov model is preferred. A possible way to incorporate the parameter vector Θ in the state equation is by considering Θ_n with a constant evolution. An augmented state vector $X_n^a = (X_n, \Theta_n)$ is consequently defined which contains X_n the true hidden state at time t_n and Θ_n the vector of unknown parameters. In the following, if X represents a random variable with values in $\mathcal{X}$, then for all $x \in \mathcal{X}$, $p(x)$ denotes the probability density of X in x. The first-order hidden Markov model is characterized by the transition density $p(x_n^a|x_{n-1}^a)$ corresponding to the state equation (Quach et al., 2007), the observation density $p(y_n|x_n^a)$ corresponding to the observation equation, and the initial density $p(x_0^a)$.

In the initialization step of our implementation, the parameters are initialized from either informative distributions ($p(x_0^a)$) or non-informative distributions for all the particles. Particle weights are assigned uniformly. Each filtering step is performed recurrently in two stages and occurs only at time steps when the observation is available (Campillo and Rossi, 2009).

Prediction:

The objective is to provide a kernel estimator of $p(x_{n+1}^a, y_{n+1}|y_{0:n})$ denoted by $\hat{p}(x_{n+1}^a, y_{n+1}|y_{0:n})$. M particles $\{\tilde{x}_n^{a\,(i)}, i = 1, \ldots, M\}$ are sampled from the distribution with conditional density $\hat{p}(x_n^a|y_{0:n})$. The M particles are propagated through the evolution model until the next available measurement to obtain the predicted states $\{\tilde{x}_{n+1^-}^{a\,(i)},\ i = 1, \ldots, M\}$. The updating scheme relies directly on Bayes' law. The particle weights are calculated based on the experimental measurements and the predictions, and then normalized. The empirical kernel approximation of the probability density of (X_{n+1}^a, Y_{n+1}) conditional to $Y_{0:n}$ can accordingly be deduced using the

Parzen-Rosenblatt kernel $K^X_{h^X_M}$, with bandwidth parameter h^X_M:

$$\hat{p}(x^a_{n+1}, y_{n+1}|y_{0:n}) = \frac{1}{M}\sum_{i=1}^{M} K^X_{h^X_M}\left(x^a_{n+1} - \tilde{x}^a_{n+1^-}{}^{(i)}\right) \cdot p\left(y_{n+1}|\tilde{x}^a_{n+1^-}{}^{(i)}\right). \tag{2}$$

Correction:

The *a posteriori* form of the estimation is deduced from Bayes' law and the kernel approximation for $p(x^a_{n+1}|y_{0:n+1})$ is given by:

$$\hat{p}(x^a_{n+1}|y_{0:n+1}) = \frac{\sum_{i=1}^{M} p(y_{n+1}|\tilde{x}^a_{n+1^-}{}^{(i)}) \cdot K^X_{h^X_M}(x^a_{n+1} - \tilde{x}^a_{n+1^-}{}^{(i)})}{\sum_{i=1}^{M} p(y_{n+1}|\tilde{x}^a_{n+1^-}{}^{(i)})}. \tag{3}$$

The part $p(y_{n+1}|\tilde{x}^a_{n+1^-}{}^{(i)})/\sum_{i=1}^{M} p(y_{n+1}|\tilde{x}^a_{n+1^-}{}^{(i)})$ can be considered as the normalized weight $\tilde{w}^{(i)}_{n+1}$ associated to the particle $\tilde{x}^a_{n+1^-}{}^{(i)}$. When the analytic form of the observation density $p(y_{n+1}|\tilde{x}^a_{n+1^-})$ is unknown, or in the case of a deterministic model, an observation kernel can similarly be introduced (Campillo and Rossi, 2009) with another Parzen-Rosenblatt kernel $K^Y_{h^Y_M}$, associated with bandwidth parameter h^Y_M:

$$\hat{p}(x^a_{n+1}|y_{0:n+1}) = \frac{\sum_{i=1}^{M} K^X_{h^X_M}(x^a_{n+1} - \tilde{x}^a_{n+1^-}{}^{(i)}) K^Y_{h^Y_M}(y_{n+1} - \tilde{y}^{(i)}_{n+1^-})}{\sum_{i=1}^{M} K^Y_{h^Y_M}(y_{n+1} - \tilde{y}^{(i)}_{n+1^-})}. \tag{4}$$

The new set of particles $\left\{\tilde{x}^a_{n+1}{}^{(i)},\ 1 \leq i \leq M\right\}$ are then sampled from $\hat{p}(x^a_{n+1}|y_{0:n+1})$.

2.5. Iterative Convolution Particle Filter

In the case of off-line estimation with a finite number of observations, in order to determine the *a priori* distribution for data assimilation, an iterative version of CPF can be applied accordingly. The idea is to repeat the filtering process in order to provide better estimations based on the available experimental observation vector. A resampling step is therefore introduced (Arulampalam et al., 2002; Doucet et al., 2001) at the end of each iteration. The Gaussian Randomization method is applied during the reinitialization, referring to a resampling step which allows us to transform the discrete approximation of the filtering density at the end of one iteration to a continuous approximation in order to avoid sample impoverishment (Musso and Oudjane, 1998). At iteration k, the particles

$x_0^{a(i)}$ are obtained as follows: the initial state vectors $\{\tilde{x}_0{}^{(i)}, i = 1, \ldots, M\}$ are selected in the same fashion as for the classical filtering process (sampled from $p(x_0)$), and the vectors of unknown parameters $\{\tilde{\Theta}_0^{(i)}, i = 1, \ldots, M\}$ are sampled from the multivariate Gaussian distribution defined by the mean and covariance matrix of $\{\tilde{\Theta}_N{}^{(i)}, i = 1, \ldots, M\}$ at iteration $k - 1$.

Due to the stochastic nature of this method, the averaging technique (Cappé et al., 2005) (Chap.4) is carried out to smooth and to decrease the fluctuations of the estimates after a burn-in period of K iterations. If we denote $\hat{\Theta}^{(l)}$ and $\hat{x}_n^{(l)}$ the estimates of the parameters and the hidden state variables at the l-th filtering iteration respectively, then for $l > K$:

$$\bar{\hat{\Theta}}^{(l)} = \frac{1}{l-K} \sum_{j=K+1}^{l} \hat{\Theta}^{(j)} \quad \text{and} \quad \bar{\hat{x}}_n^{(l)} = \frac{1}{l-K} \sum_{j=K+1}^{l} \hat{x}_n^{(j)} \ , \tag{5}$$

if a constant number of particles are simulated.

2.6. Conditional ICPF for Uncertainty Assessment

In order to estimate the noise parameters and to evaluate the uncertainty related to the estimates of the ICPF approach, we partition Θ_n: $\Theta_n = (\Theta_1, \Theta_2)$. Θ_1 denotes the parameters from the deterministic part of the model (state equation and measurement equation) and Θ_2 denotes those of the noise model (the parameters of the distributions of η and ξ in (1)). A conditional ICPF algorithm proposed by Chen et al. (2012) is therefore implemented.

In the first place, the estimation of the hidden states and of Θ_1 is performed by considering that Θ_2 is known. In practice, small initial variances for the noises seem to ensure the convergence of the algorithm towards satisfactory estimation results for the hidden states and for Θ_1. From this first estimation of Θ_1 and of the hidden states, we can estimate the parameters of the distributions of the modelling and measurement noises Θ_2 from the results. Conditionally to the new estimated Θ_2, the ICPF approach is then carried out again to estimate Θ_1 together with the hidden states. In this way, the algorithm can be iterated in turns between the model parameters Θ_1 and the noise parameters Θ_2 until the convergence of both.

However, since the posterior distributions of the parameters are no longer representative of the estimates' uncertainty because of the successive iterations of the filtering

process, parametric bootstrap (Efron and Tibshirani, 1994) is implemented to calculate the related credibility intervals. New observation vectors are hence randomly generated with $\hat{\Theta}$ and the conditional ICPF algorithm estimation is performed for each of them. The uncertainty related to the estimate $\hat{\Theta}$ can thus be evaluated properly. Moreover, we highlight that the algorithmic uncertainty (linked to the stochastic algorithm) can also be assessed by applying the conditional ICPF approach to the same experimental data set a large number of times as presented by Chen et al. (2012). In the conditions used for our test cases, the algorithmic uncertainty can be neglected.

3. Application to the LNAS Model for Sugar Beet

In this section, we apply the CPF-based estimation and assimilation methods to the LNAS model of sugar beet growth for the prediction of crop biomass compartments in various situations.

3.1. LNAS Model of Plant Growth

For data assimilation purpose, a general scheme for agrosystem models was given by Delécolle et al. (1992) and adapted by Dorigo et al. (2007). They underlined the main eco-physiological processes and key variables to describe the plant-environment interactions. We adapt this scheme to the sugar beet case in Fig. 1. The principal processes they suggested to consider are crop development, light interception, biomass accumulation, biomass partitioning and senescence.

[Figure 1 about here.]

In the following, we recall the equations of the Log-Normal Allocation and Senescence (LNAS) model (Chen and Cournède, 2012; Cournède et al., 2013) specifically derived for the sugar beet, per unit surface area, with two organ compartments: foliage and root system. The model is built based on a daily time step and environmental variables are daily averages. Note that adaptations of the model are currently also derived for maize, wheat and sunflower.

A specificity of the model is that the main growth processes underlined in Fig. 1 are considered as stochastic processes when it appears relevant.

Interception and assimilation: $Q(t)$ is the biomass production on day t per unit surface area ($g \cdot m^{-2}$) which can be obtained by an adaptation of the Beer-Lambert law (Monteith, 1977): $\left(1 - e^{-\lambda \cdot Q_g(t)}\right)$ represents the fraction of intercepted radiation, with $Q_g(t)$ the total mass of green leaves on day t (in $g \cdot m^{-2}$) and λ a coefficient parameter (in $g^{-1} \cdot m^2$). The biomass production of the whole plant is then deduced by multiplying the total amount of absorbed photosynthetically active radiation per unit surface area (PAR, in $MJ \cdot m^{-2}$) and an energetic efficiency μ (in $g \cdot MJ^{-1}$):

$$Q(t) = \left(\mu \cdot PAR(t) \left(1 - e^{-\lambda Q_g(t)}\right)\right) \cdot (1 + \eta_Q(t)) \tag{6}$$

where we introduce the modelling noise $\eta_Q \sim \mathcal{N}(0, \sigma_Q^2)$. Since the characterization of the environmental variables and of the light interception is not accurately described by the Beer-Lambert law, the model noise appears relevant for the production equation. Despite the fact that $Q(t)$ should always be positive, we still make a normal assumption. With the multiplicative form and the levels of noise generally considered (inferior to 10%), we assure that there is no problem of positivity loss.

The parameter λ corresponds to $\lambda = k\,SLA$, where k is the Beer-Lambert extinction coefficient and SLA is the specific leaf area, so that the term $\left(1 - e^{-\lambda Q_g(t)}\right)$ can classically be rewritten as $\left(1 - e^{-k\,LAI}\right)$, with LAI the leaf area index. There is a slight difference in the formulation however, because we consider λ as an empirical constant parameter, whereas linking leaf mass to leaf surface via the SLA variable is not obvious since the SLA is known to vary during crop growth and within plants (see for example Jullien et al. (2009)), even though it is often regarded as constant in models. This simplification also allows us to avoid the differentiation between blades and petioles (which is not always easy from a botanical point of view in sugar beet), both constitute the leaf compartment.

Allocation to the foliage and root system compartments:
The description of the allocation process is a simplification of the GreenLab model (Yan et al., 2004; Guo et al., 2006) with the organ sink dynamics being described at compartment level: the proportion of biomass allocated to each compartment (foliage and root) is described by an empirical function γ:

$$Q_f(t+1) = \quad Q_f(t) + \gamma(t) \cdot Q(t), \tag{7}$$

$$Q_r(t+1) = \quad Q_r(t) + (1 - \gamma(t)) \cdot Q(t), \tag{8}$$

$$\gamma(t) = (\gamma_0 + (\gamma_f - \gamma_0) \cdot G_a(\tau(t))) \cdot (1 + \eta_\gamma(t)) \tag{9}$$

with $\tau(t)$ the thermal time, which corresponds to the accumulated daily temperature (above a threshold temperature, which is taken as 0 for the sugar beet (Lemaire et al., 2008)) since emergence, G_a the cumulative distribution function of a log-normal law parameterized by its median μ_a and standard deviation s_a (for more explicit biological meanings of the parameters), and the modelling noise (process noise) denoted by $\eta_\gamma(t) \sim \mathcal{N}(0, \sigma_\gamma^2)$. The cumulative distribution of the log-normal law is chosen for its flexibility: it allows to reproduce dynamics similar to the sigmoid-type functions often employed to describe biological processes, while having the advantage to start with a null value in zero. The transformation chosen to obtain γ is inspired by the simulation of biomass allocation to root and leaf compartments of sugar beet described by SUCROS (Spitters et al., 1989) and GreenLab (Lemaire et al., 2008).

Given the fact that the allocation strategy is very sensitive to environmental conditions, we introduce a multiplicative perturbation, again under normal assumption.

Senescence: The senescent foliage mass Q_s is a proportion of the accumulated foliage mass based on its underlying cumulative distribution of a log-normal law characterized by median μ_s and standard deviation s_s:

$$Q_s(t) = G_s(\tau(t) - \tau_{sen}) Q_f(t) \tag{10}$$

with τ_{sen} the thermal time at which the senescence process initiates. The green foliage mass Q_g can be hence easily obtained:

$$Q_g(t) = Q_f(t) - Q_s(t). \tag{11}$$

We choose a deterministic version of the senescence equation despite the strong variations that can characterize the process. As a matter of fact, the influence of senescence in the biomass budget is due to the decrease of photosynthetic foliage in Equation (6), adding a perturbation in the senescence mass either would be of second order or could be summarized in the modelling noise η_Q.

Note that both allocation and senescence processes are driven by the thermal time. The phenological stage in Fig. 1 is thus simply represented in our model by the course of the thermal time, from emergence.

Observations: In this study, the observation variables potentially available from field measurements for parameter estimation and data assimilation are:

$$Y(t) = \begin{pmatrix} Q_g(t) \cdot (1 + \epsilon_g(t)) \\ Q_r(t) \cdot (1 + \epsilon_r(t)) \end{pmatrix} \tag{12}$$

with measurement noises: $\epsilon_g(t)) \sim \mathcal{N}(0, \sigma_g^2)$, and $\epsilon_r(t) \sim \mathcal{N}(0, \sigma_r^2)$. Again, the choice of the normal law (instead of the log-normal one for instance) does not practically generate problems of positivity loss, thanks to the multiplicative form and the low levels of noise generally considered (inferior to 10%).

The proposed framework would still be applicable when using satellite image data with LAI evaluation. In that case, LAI should be specified clearly as a state variable in the system (with an explicit formulation of the SLA variable), consequently a corresponding observation function should be defined.

3.2. Experimental Data

The data used for this study were obtained by the French institute for sugar beet research (ITB, Paris, France) in 2006, 2008 and 2010 with slightly different cultivars and in different locations with different observed densities (details of the experimental protocols can be found respectively in Lemaire et al. (2008), Lemaire et al. (2009) and Baey et al. (2013)). For the test case, the 2010 dataset is chosen for estimation since more observation points are available compared to the other four datasets (of course, in real applications, the older datasets are supposed to be used for parameter estimation). Dry matter of root and leaves were collected on 50 plants at 12 dates (in days after sowing):

$$\mathcal{O}_{2010} = \{54, 68, 76, 83, 98, 104, 110, 118, 125, 132, 145, 160\},$$

whereas for assimilation and prediction, other datasets (2006 and 2008) are used. The same type of observations were made only at 7 different dates given by

$$\mathcal{O}_{2006} = \{54, 59, 66, 88, 114, 142, 198\},$$

$$\mathcal{O}_{2008} = \{39, 60, 67, 75, 88, 122, 158\},$$

respectively. For each plant, the green foliage mass denoted by Q_g and the root compartment mass denoted by Q_r were measured. The observation vector Y_n is obtained by

averaging each data on all the samples and extrapolated at m^2 level by multiplying by the observed density.

We note that the three 2008 datasets used for assimilation correspond to different density conditions (5.4 plants$\cdot m^{-2}$, 10.9 plants$\cdot m^{-2}$, 16.4 plants$\cdot m^{-2}$) compared to the 2010 dataset (11.9 plants$\cdot m^{-2}$) and the 2006 dataset (9.6 plants$\cdot m^{-2}$).

3.3. Results

3.3.1. Parameter Screening by Sensitivity Analysis

Sobol's indices (first order and total order) of all the functional parameters of the LNAS model (without counting the noises parameters) were computed using a generalized least-squares criterion as output (see Cournède et al. (2013)). According to the sensitivity analysis result, we screened the parameters s_a, μ_{sen}, s_{sen} and fixed them to their mean values of the variation interval, as their total order indexes are all below 0.02. For the five other parameters, their total order effects cannot be neglected and they should be estimated from experimental data.

3.3.2. Parameter Estimation

Based on the sensitivity analysis results, the unknown parameter vector for the deterministic part of the model is assumed to be $\Theta_1 = (\mu, \lambda, \mu_a, \gamma_0, \gamma_f)$ and the unknown noise parameter vector is $\Theta_2 = (\sigma_Q, \sigma_\gamma, \sigma_g, \sigma_r)$. We compared the estimation performances of the classical CPF and the iterative version, ICPF. For the CPF approach, 500000 particles were initialized with non-informative prior distributions (uniform distributions), while as for the conditional ICPF approach, 8000 particles were drawn from the same prior distributions. The conditional ICPF estimation process began with the estimation of Θ_1 given Θ_2, then Θ_2 was estimated empirically based on the estimates of the hidden states. The estimation then proceeded with the new value of Θ_2 and iterated. Finally, 3 repetitions of the conditional version of ICPF were performed in our test, each of them contained 200 filtering iterations to ensure convergence. A parametric bootstrap was subsequently carried out to evaluate the estimates' uncertainty. Standard deviations and confidence intervals were hence obtained from 200 bootstrap samples. Results are given in Table 1. The estimated modelling noises are smaller with the ICPF method, which implies a better adaptation of the model. For this reason, we choose to consider

the estimates obtained with the conditional ICPF to provide the prior distribution for data assimilation. The issues related to the uncertainty assessment and the comparison between ICPF and CPF were further addressed in Chen and Cournède (2012).

[Table 1 about here.]

3.3.3. Data Assimilation with CPF

The distribution resulting from the parametric bootstrap of the ICPF estimation was used to provide the prior information in the assimilation step, in which the (classical) CPF algorithm was applied. For both the 2006 and 2008 datasets, 500000 particles were simulated, all but the last two measurements (corresponding to data until day 114 for the 2006 dataset, and until day 88 for the 2008 dataset) were used to update and correct the parameter and the state estimates. The propagation of particles through the stochastic dynamic model was carried on without any further correction until day 198 for the 2006 dataset and day 158 for the 2008 dataset. At last, the simulated values of the state variables Q_g and Q_r on day 142 and day 198 (resp. day 122 and day 158 for the 2008 dataset) given by all the particles as well as their associated weights were used to build the posterior predictive distributions.

In order to provide reference values of the prediction without assimilation, an Uncertainty Analysis (UA) was performed. 500000 particles were initialized in the same way as in the CPF approach, from the prior distribution given by the parameter estimation step, and the distribution of the model outputs of interest was approximated by propagating independently through the stochastic dynamic system all the particles. No parameter or state update is done from the experimental data of the early stages.

Based on the parameter estimation results of the 2010 dataset, we compared the predictive capacity of the model for the 2006 and the 2008 experiments, with and without data assimilation, regarding the last two observations (t_{142} and t_{198} for 2006 and t_{122} and t_{158} for 2008). Although the six experiments are quite different (different locations, in different years, and different cultivars) the proposed CPF-based approach was shown to provide fair predictions in most cases and managed to reduce the prediction errors compared to the results obtained with UA (without data assimilation), as demonstrated by Fig. 2. It is particularly spectacular for the leaf biomass (Q_g) prediction (Table 2, Table 3).

[Figure 2 about here.]

Tables 2 to 5 illustrate the prediction results of the two methods (with or without data assimilation) for different experiments. The average estimates are given with their corresponding 95% credibility intervals. According to the result based on the 2006 dataset (Table 2) and the 2008 dataset with observed density of 10.9 (Table 3), the prediction results are clearly improved both in terms of mean prediction and credibility intervals for nearly all the predictions given by the CPF-based approach compared to those provided by UA. For the result of 2008 (density 10.9), all the predictions have a relative error close or inferior to 10% which indicate that the CPF-based method had an excellent predictive performance. The proposed credibility intervals all include the real value of the last two measurements and generally narrower than those given by the UA.

On the other hand, as suggested by Fig. 2, the 95% credibility interval provided by the UA can be considered non-reliable since it does not always contain the real measurement values in several cases (*cf.* Table 4 and Table 5).

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

3.3.4. Discussion

The predicted credibility intervals provided by the CPF-based approach are nearly in all cases narrower than those of the UA. It is of course an expected result for data assimilation: to reduce the prediction uncertainty based on the available information. However, regarding the point estimations, the results of the CPF-based method were not always more accurate than those provided by UA, especially for the root mass, as shown in Table 3 for $Q_r(t_{158})$ or in Table 5 for $Q_r(t_{122})$ or $Q_r(t_{158})$. However, when it happened, the corresponding predictions of leaf biomass given by the UA were nonetheless far from the real observed values and can be considered as unreliable. This may reveal some particular plasticity in root biomass production that was not well captured by the

model, which can be regarded as a drawback of the simplified model. However, even in the few cases when the proposed method failed to improve the point prediction of some variables, it always managed to provide reasonable credibility intervals which contained the real values of the observations.

Regarding the uncertainty assessment, although the modelling noises are quite small (around 1%), which implies a good adaptation of the model, the observation noises are rather important (10% and 7%, *cf.* Table 1). Therefore, the prediction gain with data assimilation remains quite limited since the observations may not be reliable enough to update parameters and state variables, and the important observation noises used in the assimilation step may also prevent the algorithm to retrieve the most useful information out of the few available data.

Moreover, further tests (not detailed here) showed that the CPF-based data assimilation is sensitive to the observation noise, and their proper evaluation is thus crucial to improve the method accuracy and reliability. It is clearly a bottleneck for real applications since the observation noises in practice tend to vary a lot according to the experimental configurations (for example in two different years or in two different fields), so that their proper evaluation remains quite difficult.

4. Application to the STICS Model for Winter Wheat

To prove that the proposed approach is generic and adaptable to various models, and robust enough to confront different situations, another study is conducted with an application to the complex crop model STICS. Rather different from LNAS, STICS is deterministic, and focuses on the detailed description of the crop-environment interactions. As for the LNAS model, the proposed approach was applied to the STICS model of wheat growth in interaction with the water resource, first for model calibration and then for prediction of crop biomass compartments, LAI and the average water content in soil with data assimilation. This application clearly demonstrates the strength of the Convolution Particle Filtering approach which can be applied straightforwardly to deterministic models.

4.1. Principles of the STICS Model

The STICS model focuses on the crop-soil system and has already been applied to various crops. It is divided into several modules, each representing different plant growth mechanisms (Brisson et al., 2008). Among them, the development module is in charge of the evolution of LAI and root compartment, and in the meantime defines the harvested organ filling phase. Fig. 3 illustrates the main processes involved in the STICS model adapted to the wheat crop.

Contrary to LNAS which is based on a biomass budget, the growth in STICS is driven by an empirical law for the LAI growth. Three phases are involved, the first phase (from emergence to the maximal LAI point) is approached by a logistic function with the hypothesis that the ratio between blade and petioles is constant, followed by a stabilized phase and a senescent phase of linearly decreasing LAI. Several stress factors limit the potential daily increase in LAI. The daily biomass production is then computed as a quadratic function of the intercepted radiation, given by the Beer-Lambert law. Hence, crop total biomass results from the accumulation of the daily biomass increase, and the final grain biomass is obtained through an harvest index.

[Figure 3 about here.]

In our study, the field experimentations were conducted without Nitrogen stress, but in light water stress conditions. Therefore, the soil characteristics are taken into account to compute the water balance of the plant-soil-atmosphere system and thus to estimate several water stress indices impacting plant growth at different levels. For this purpose, the water contents in three soil layers are calculated.

In consequence, the above model can be divided into two sub-models, one concerns the plant system with state variables at time t denoted $X_p(t)$ and the other the soil system with state variables denoted $X_s(t)$. Plant growth is described by the function f_p. The root compartment growth is directly affected by the soil temperature and water content. Likewise, the soil water content determines several water stress indices impacting LAI development, biomass production and senescence. On the other hand, water transfers including evaporation and plant transpiration are calculated in the soil system by the function f_s:

$$\begin{cases} X_p(t+1) = f_p\left(X_p(t), X_s(t), E_p(t), \Theta_p\right), \\ X_s(t+1) = f_s\left(X_s(t), X_p(t), E_s(t), \Theta_s\right). \end{cases} \tag{13}$$

Functions f_p and f_s are detailed in (Brisson et al., 2008). Given the complexity of the STICS model, a large number of parameters are involved, some of which may be species dependent or genotype dependent.

4.2. Experimental Data

The data used for this study were obtained by INRA (Institut National de Recherche Agronomique) in the context of the Aquateam project (http://www.projet-aquateam.org/) whose objective is to develop decision aid tools for crop irrigation. The experiments were carried out at Villamblain (France). The growth of two commercial varieties of winter wheat were monitored: *Raffy* in two experimental campaigns, 2011-2012 (sowing date: 25 October 2011; harvesting date: 25 July 2012) and 2012-2013 (sowing date: 29 October 2012; harvesting date: 30 July 2013) and *Numeric* in one experimental campaign, 2012-2013 (sowing date: 29 October 2012; harvesting date: 30 July 2013). In our study, the 2012 dataset of variety *Raffy* was used for parameter estimation. For the sake of clarity for the readers already familiar with the STICS model, we use the classical notations for state variables and parameters recalled extensively in Brisson et al. (2008). Dry matter of green leaves (denoted *mafv*, g$\cdot m^{-2}$), above-ground dry matter (denoted *masec*, g$\cdot m^{-2}$), soil averaged water content (denoted *hur*, mm/unit area) and dry matter of grain yield (denoted *magrain*, g$\cdot m^{-2}$) were measured and collected at different dates in 2012 (in days after sowing):

$$\mathcal{O}_{2012}^{mafv} = \{155, 185\}, \quad \mathcal{O}_{2012}^{magrain} = \{269\},$$

$$\mathcal{O}_{2012}^{masec} = \{155, 185, 213, 239, 269\},$$

$$\mathcal{O}_{2012}^{hur} = \{155, 171, 178, 192, 203, 219, 234, 247, 260\}.$$

For data assimilation and prediction, the two 2013 datasets corresponding to both varieties were used. Additional measurements of LAI (denoted *LAI*, m^2 leaf $\cdot m^{-2}$ soil) were available for the 2013 dataset (obtained with the SunScan Canopy Analysis System of Delta-T Devices), while the dry matter of green leaves was not measured. The

observation dates are given as follows:

$$\mathcal{O}_{2013}^{hur} = \{21, 79, 114, 155, 170, 182, 198, 210, 226, 240, 255, 275\},$$

$$\mathcal{O}_{2013}^{magrain} = \{266\}, \quad \mathcal{O}_{2013}^{LAI} = \{162, 175, 191, 203, 212, 219, 233\}.$$

Daily mean values of air temperature, solar radiation, potential evapotranspiration, and total daily rainfall were obtained from French meteorological advisory services (MÃl'tÃl'o France) 3 km away from the experimental site.

4.3. Results

4.3.1. Parameter Screening by Sensitivity Analysis

Since STICS has a large number of parameters, and regarding the reduced experimental data sets that are available for its parameterization, a preliminary parameter selection was conducted in the first place. 16 variety-dependent parameters that are relevant to the concerned model state equations were selected to perform the sensitivity analysis, for the output chosen as a generalized least-square criterion based on the 2012 dataset.

According to Sobol total order indices, the four most influential parameters EFFICIENCE, STAMFLAXV, UDLAIMAXP and VLAIMAXP were selected for the next calibration step. Their definitions are given by Table 6. The other parameters (with total order indices below 0.05) were hence fixed to the mean values of their variation intervals (deduced from literature (Brisson et al., 2008)).

[Table 6 about here.]

4.3.2. Parameter Estimation

In the deterministic case, the conditional ICPF approach coupled with the parametric bootstrap used for the LNAS model is not possible. We thus estimated the unknown parameter vector $\Theta = (EFFICIENCE, STAMFLAXV, UDLAIMAXP, VLAIMAXP)$ with the classical CPF. 100000 particles were drawn from the distributions used for sensitivity analysis and obtained from the literature (Brisson et al., 2008). Means and standard deviations of all the four parameters are thus obtained based on the posterior distribution provided by the population of the particles and their associated weights. The results are presented in Table 7.

[Table 7 about here.]

4.3.3. Data Assimilation with CPF

Data assimilation was subsequently performed using as prior distributions the posterior distributions of the parameters provided by the estimation step. 100000 particles were simulated for the two 2013 datasets. The recalibration was carried out based on the first seven observations of soil water content *hur* and the first three observations of *LAI*, corresponding for both cases to data obtained before Day 199 after sowing. The values of these state variables along with the grain yield were then simulated for all particles until the end of the growing season (Day 275 after sowing), which allows us to build the posterior distribution of the prediction.

Uncertainty Analysis (UA) was also performed in this study to provide reference values for the prediction with 100000 particles as well initialized in the same way as for the prediction with assimilation.

Fig. 4 and Fig. 5 illustrate the model prediction for the LAI variable, with and without data assimilation. The assimilation step has clearly enhanced the predictive capacity of the model. The results are detailed in Table 8 for the variety *Raffy* and Table 9 for the variety *Numeric*: the predictions and relative prediction errors as well as the prediction uncertainty (given by the credibility intervals) are given for the soil water content *hur*, *LAI*, and grain yield *magrain*, with and without data assimilation. The relative error of prediction was greatly reduced when taking advantage of the early data by assimilation, and the prediction uncertainty was also significantly decreased in almost all the cases.

[Figure 4 about here.]

[Figure 5 about here.]

[Table 8 about here.]

[Table 9 about here.]

4.3.4. Discussion

As for the LNAS model for sugar beet, the mean prediction and prediction credibility intervals are generally greatly improved with data assimilation compared to simple

prediction by uncertainty analysis.

A noteworthy point concerns the special climatic conditions of year 2013 for the winter wheat data, since severe stresses due to heavy rain and frost in winter were observed which resulted in a far lower plant density compared to 2012. However, since there was more space available for each plant, a compensation phenomenon was observed. The LAI curve got delayed but finally nearly caught up with what would have been observed with the regular climate conditions, and therefore grain yield was not seriously influenced. However, the prediction performed without assimilation was not able to capture this compensation phenomena and thus failed to provide a reasonable prediction for the LAI and for the final yield (*cf.* variable *magrain* in Table 9 and Table 8), while the proposed approach has remarkably tackled this issue by updating properly the parameters and hidden variables based on the available data. As a result, the satisfactory predictions demonstrated its robustness in case of extreme weather scenarios.

On the other hand, the prediction improvement is less impressive with the soil water content variable, probably due to the fact that there was no soil parameter selected by sensitivity analysis, and only light water stress conditions: the influence of the crop parameters updated in the data assimilation is limited, so that the difference with the case without assimilation is not so apparent (while still to the advantage of the assimilation method). In the light of this example, we may also consider a more subtle selection of parameters in the first step of our approach to improve the prediction of a specific output variable of interest when needed.

Finally, it is important to underline that when the proposed approach is applied to a deterministic model, the resulting uncertainty assessment is less rigorous in both the estimation and assimilation steps. In this case, the posterior distributions obtained with the CPF approach are extensively influenced by the tuning parameters of the algorithm, especially the choice of the kernel functions and their bandwidth parameters, so that the credibility intervals computed should be regarded as contextual approximations. This directly results from the fact that the model does not have a probabilistic framework, which of course restricts the validity of the statistical analysis.

5. Conclusion

In this paper, we detailed CPF-based methods for parameter estimation and data assimilation, and applied it to the LNAS stochastic crop model for sugar beet and the STICS model for winter wheat. The methods can explicitly account for different sources of uncertainty during the parameter estimation step, which provides prior distributions for the following step of data assimilation. Both state variables and parameters are then updated to improve model prediction and prediction uncertainty.

Various experimental conditions were considered for the test cases, including real datasets from different years, in different locations, with different crop densities, different cultivars, in different stress conditions... In most cases the proposed approach was able to provide fair predictions after a short period of adjustment and managed to reduce significantly the prediction errors. Such performances raise interest for future applications.

For a stochastic model like LNAS, a proper uncertainty assessment with the conditional ICPF approach in the estimation step is possible. Based on these results, the performance of the prediction with data assimilation is greatly improved both in terms of mean prediction error and robustness of the provided credibility interval compared to a simple uncertainty analysis.

Thanks to the kernel-based approach in the Convolution Particle Filtering, the method can also cope with deterministic models, which is a very interesting feature in crop modelling, since a lot of broadly used classical models are not formulated as hidden Markov models. In this case, however, if the mean predictions are fair enough, the credibility intervals should be seen as indicative, as a proper uncertainty assessment cannot be achieved without a proper probabilistic framework. More generally, in applicative contexts in agriculture, the specific local field conditions (like soil properties, or Nitrogen budget) are not easy to evaluate, so that the characterization of appropriate prior parameters and inputs for a complex model like STICS is a delicate task, involving great efforts and a potentially important risk to deteriorate the predictive capacity of the method. On the contrary, a simple model built with a minimal set of ecophysiological processes (Delécolle et al., 1992) coupled with data assimilation should be able to adjust and capture crop dynamics corresponding to very diverse environmental situations.

As underlined above, our method implies a parameter estimation step, in order to determine the prior distribution for the assimilation step. Of course, if this prior knowledge is already available, then the estimation step can be skipped. Moreover, this preliminary estimation step corresponds to an off-line estimation, for which other methods than CPF could be used, such as Kalman filters-based algorithms or Markov chain Monte-Carlo (MCMC) algorithms.

We have already tested the Unscented Kalman Filter (Julier et al., 2000) and Ensemble Kalman Filter (Evensen, 2006) with the LNAS model for sugar beet in the same test condition and compared their performances to the CPF approach. Our results showed that Kalman filter-based approaches suffer from the important nonlinearity of the model, contrary to the CPF-based methods. A clear advantage of the latter was shown, which confirms the conclusion of Rossi (2004) obtained for different models.

Likewise, some preliminary results suggest the superiority of the CPF approach compared to the classical implementation of the Markov chain Monte Carlo method for the estimation step (Adaptive Metropolis (Gelman et al., 1996; Haario et al., 1999, 2001) with multivariate random walk, based on the famous Metropolis-Hasting algorithm (Metropolis et al., 1953)). As a matter of fact, the credibility intervals obtained for the prediction with assimilation appear less robust when the preliminary estimation was achieved with MCMC than with the CPF-based algorithm. It may be caused by some convergence problems of the MCMC approach in the case of sparse datasets as already pointed out by Geyer (1992). A solution to improve the mixture properties as well as the convergence rate, according to Campillo et al. (2009), could involve making different copies of the same Markov chain and letting them interact with each other. Note that this idea corresponds to one of the key step of the Convolution Particle Filter, the resampling, which allows to preserve the variability of updated state variables and parameters. The acceptance probability can therefore be regarded as a sort of weight associated to each particle. Moreover, when no satisfying prior distribution is available, iterating the filtering step in our algorithm is an efficient way to refine the quality of the estimates by taking advantage of the off-line estimation. We are currently testing the interacting MCMC algorithm proposed by Campillo et al. (2009).

The CPF-based estimation and assimilation algorithms are implemented in the PYG-

MALION modelling platform (Cournède et al., 2013) with the aim of making it available to a large set of plant models implemented in the platform, and thus to carry out model comparison and selection with the precise objective of model adaptation for data assimilation. Likewise, the influence of the estimation and assimilation algorithms on the performance of the method can be further tested in future studies.

Generally speaking, the assessment of uncertainty appears as a crucial point for the usefulness of plant models (Ford and Kennedy, 2011). How do parameters vary across years, different locations or genotypes? How to distinguish and to identify the different sources of variations and uncertainty? How to extract most of the information from the available data while still addressing the uncertainty in an appropriate way in order to provide reliable credibility intervals? These issues should be considered crucial and properly addressed for the new generation of plant models.

Acknowledgements

The authors are grateful to the research colleagues at ITB and INRA for providing the experimental data used in this paper. They would also like to thank the two anonymous reviewers for their useful remarks and their suggestions for future research perspectives.

References

Arulampalam, M., Maskell, S., Gordon, N.and Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. In: Connection between forest resources and wood quality: modelling approaches and simulation software. No. 50(2). IEEE Trans. Signal Proces., p. 1740188.

Baey, C., Didier, A., Lemaire, S., Maupas, F., Cournède, P.-H., 2013. Parametrization of five classical plant growth models applied to sugar beet and comparison of their predictive capacity on root yield and total biomass. Ecological Modelling.

Bertheloot, J., Cournède, P.-H., Andrieu, B., 2011. Nema, a functional-structural model of n economy within wheat culms after flowering: I. model description. Annals of Botany In press.

Bouman, B., 1992. Linking physical remote sensing models with crop growth simulation models, applied for sugar beet. International Journal of Remote Sensing 13 (14), 2565–2581.

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

Brisson, N., Launay, M., Mary, B., Beaudoin, N. (Eds.), 2008. Conceptual Basis, Formalisations and Parameterization of the Stics Crop Model. Éditions QUAE, Versailles, France.

Brisson, N., Mary, B., Ripoche, D.and Jeuffroy, M. H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J., Meynard, J. M., Delecolle, R., 1998. Stics : a generic model for the simulation of crops and their water and nitrogen balances. i. theory, and parameterization applied to wheat and corn. Agronomie 18, 311–346.

Buck-Sorlin, G., Bachmann, K., 2000. Simulating the morphology of barley spike phenotypes using genotype information. Agronomie 20, 691–702.

Campillo, F., Rakotozafy, R., Rossi, V., 2009. Parallel and interacting Markov chain Monte Carlo algorithm. Mathematics and Computers in Simulation 79, 3424–3433.

Campillo, F., Rossi, V., 2009. Convolution Particle Filter for Parameter Estimation in General State-Space Models. IEEE Transactions in Aerospace and Electronics. 45 (3), 1063–1072.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environmental Modelling and Software 22, 1509–1518.

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. Springer, New York.

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

Chen, Y., Cournède, P.-H., 2012. Assessment of parameter uncertainty in plant growth model identification. In: Kang, M., Dumont, Y., Guo, Y. (Eds.), Plant growth Modeling, simulation, visualization and their Applications (PMA12).

Cournède, P.-H., Chen, Y., Wu, Q., Baey, C., Bayol, B., 2013. Development and evaluation of plant growth models: Methodology and implementation in the PYGMALION platform. Mathematical Modelling of Natural Phenomena 8, 112–130.

Delécolle, R., Maas, S., Guérif, M., Baret, F., 1992. Remote sensing and crop production models: present trends. ISPRS Journal of Photogrammetry and Remote Sensing 47 (23), 145 – 161.

Dente, L., Satalino, G., Mattia, F., Rinaldi, M., 2008. Assimilation of leaf area index derived from ASAR and MERIS data into CERES-wheat model to map wheat yield. Remote Sensing of Environment 112 (4), 1395 – 1407.

Dorigo, W., Zurita-Milla, R., de Wit, A., Brazile, J., Singh, R., Schaepman, M., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation and Geoinformation 9 (2), 165 – 193.

Doucet, A., De Freitas, N., Gordon, N., 2001. Sequential Monte Carlo methods in practice. Springer-Verlag, New-York.

Efron, B., Tibshirani, R., 1994. An Introduction to the Bootstrap. Chapman & Hall/CRC Monographs on Statistics and Applied Probability.

Evensen, G., 2006. Data assimilation: The ensemble Kalman Filter. Springer.

Ford, E. D., Kennedy, M. C., 2011. Assessment of uncertainty in functional-structural plant models. Annals of Botany 108 (6), 1043–1053.

Fowler, D. B., Limin, A. E., Ritchie, J. T., 2003. Low-temperature tolerance in cereals: Model and genetic interpretation. Crop Science 39 (3), 626–633.

Gabrielle, B., Roche, R., Angas, P., Cantero-Martinez, C., Cosentino, L., Mantineo, M., Langen-Siepen, M., Hénault, C., Laville, P., Nicoullaud, B., Gosse, G., 2002. A priori parameterisation of the ceres soil-crop models and tests against several european data sets. Agronomie 22-2, 119–132.

Gelman, A., Roberts, G., Gilks, W., 1996. Efficient Metropolis jumping rules. Bayesian Statistics V, 599–608.

Geyer, C., 1992. Practical Markov chain Monte Carlo (with discussion). Statistical Science 7 (4), 473–482.

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

Guérif, M., Duke, C., 2000. Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation. Agriculture, ecosystems & environment 81 (1), 57–69.

Guérif, M., Houlès, V., Makowski, D., Lauvernet, C., 2006. Data assimilation and parameter estimation for precision agriculture using the crop model stics. In: Wallach, D., Makowski, D., Jones, J. (Eds.), Working with Dynamic Crop Models. Elsevier, pp. 391–398.

Guo, Y., Ma, Y., Zhan, Z., Li, B., Dingkuhn, M., Luquet, D., de Reffye, P., 2006. Parameter optimization and field validation of the functional-structural model Greenlab for maize. Annals of Botany 97, 217–230.

Haario, H., J., S., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. Comput. Statist. 14, 375–395.

Haario, H., J., S., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7, 223–242.

Hammer, G., Cooper, M., Tardieu, F., Welch, S., Walsh, B., Van Eeuwijk, F., Chapman, S., Podlich, D., 2006. Models for navigating biological complexity in breeding improved crop plants. Trends in Plant Science 11 (12), 587–593.

Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., Saito, K., 2004. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 101 (27), 10205–10210.

Houlès, V., Mary, B., Guérif, M., Makowski, D., Justes, E., 2004. Evaluation of the ability of the crop model STICS to recommend nitrogen fertilisation rates according to agro-environmental criteria. Agronomie 24, 339–349.

Jazwinski, A., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York.

Jones, C., Kiniry, J., 1986. CERES-MaÃŕze : A simulation model of maÃŕze growth and development. Texas A&M University Press.

Jones, J., Graham, W., 2006. Application of extended and ensemble kalman filters to soil carbon estimation. In: Wallach, D., Makowski, D., Jones, J. (Eds.), Working with Dynamic Crop Models. Elsevier, pp. 55–100.

Julier, S., Uhlmann, J., Durrant-Whyte, H., 2000. A New Method for the Non-Linear Transformation

of Means and Covariances in Filters and Estimators. IEEE Transaction on Automatic Control 45 (3), 477–482.

Jullien, A., Allirand, J.-M., Mathieu, A., Andrieu, B., Ney, B., 2009. Variations in leaf mass per area according to n nutrition, plant age, and leaf position reflect ontogenetic plasticity in winter oilseed rape (brassica napus l.). Field Crops Research 114 (2), 188 – 197.

Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., Smith, C., 2003. An overview of apsim, a model designed for farming systems simulation. European Journal of Agronomy 18 (3-4), 267–288.

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics 5 (1), 1–25.

Launay, M., Guérif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agriculture, ecosystems & environment 111, 321–339.

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

Lemaire, S., Maupas, F., Cournède, P.-H., de Reffye, P., 2008. A morphogenetic crop model for sugar-beet (*beta vulgaris l.*). In: International Symposium on Crop Modeling and Decision Support: ISCMDS 2008, April 19-22, 2008, Nanjing, China.

Maas, S. J., 1988. Using satellite data to improve model estimates of crop yield. Agronomy Journal 80 (4), 655–662.

Makowski, D., Jeuffroy, J., Guérif, M., 2004. Bayesian methods for updating crop-model predictions, applications for predicting biomass and grain protein content. Wageningen UR Frontis Series.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21 (1087-1091).

Mittler, R., 2006. Abiotic stress, the field environment and stress combination. Trends in Plant Science 11 (1), 15 – 19.

Monteith, J., 1977. Climate and the efficiency of crop production in Britain. Proceedings of the Royal Society of London B 281, 277–294.

Moran, M., Inoue, Y., Barnes, E., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. Remote Sensing of Environment 61 (3), 319 – 346.

Moulin, S., Bondeau, A., Delecolle, R., 1998. Combining agricultural crop models and satellite observations: From field to regional scales. International Journal of Remote Sensing 19 (6), 1021–1036.

Musso, C., Oudjane, N., 1998. Regularization schemes for branching particle systems as a numerical solving method of the nonlinear filtering problem. In: Proceedings of the Irish Signals and Systems Conference.

Naud, C., Makowski, D., Jeuffroy, M.-H., 2007. Application of an interacting particle filter to improve

nitrogen nutrition index predictions for winter wheat. Ecological Modelling 207 (24), 251 – 263.

Oudjane, N., Musso, C., 1998. Regularized particle schemes applied to the tracking problem. In: International Radar Symposium, Munich, Proceedings.

Quach, M., Brunel, N., d'Alché Buc, F., 2007. Estimating parameters and hidden variables in non-linear state-space models based on odes for biological networks inference. Bioinformatics 23 (23), 3209–3216.

Rossi, V., 2004. Filtrage non linéaire par noyaux de convolution: Application à un procédé de dépollution biologique. Ph.D. thesis, Ecole National Supérieure Agronomique de Montpellier.

Rossi, V., Vila, J.-P., 2006. Nonlinear filtering in discrete time: A particle convolution approach. Ann. Inst. Stat. Univ. Paris 3, 71–102.

Spitters, C., Van Keulen, H., Van Kraalingen, D., 1989. A simple and universal crop growth simulator: SUCROS87. PUDOC, Wageningen.

Tardieu, F., 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends in Plant Science 8 (1), 9–14.

Trevezas, S., Cournède, P.-H., 2013. Journal of Agricultural, Biological, and Environmental Statistics.

Van Leeuwen, P., Evensen, G., 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. Monthly Weather Review 124, 2898–2913.

Wernsdörfer, H., Rossi, V., Cornu, G., Oddou-Muratorio, S., Gourlet-Fleury, S., 2008. Impact of uncertainty in tree mortality on the predictions of a tropical forest dynamics model. Ecological Modelling 218 (3), 290–306.

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

Yan, H., Kang, M., De Reffye, P., Dingkuhn, M., 2004. A dynamic, architectural plant model simulating resource-dependent growth. Annals of Botany 93, 591–602.

Yin, X., Struik, P. C., 2010. Modelling the crop: from system dynamics to systems biology. Journal of Experimental Botany 61 (8), 2171–2183.

List of Figures

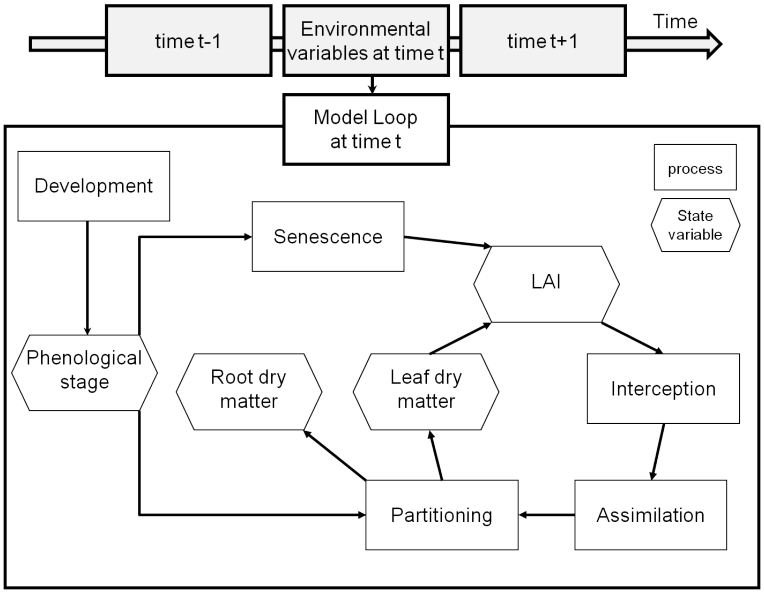

Figure 1: General scheme of an agroecosystem model for sugar beet (adapted from Delécolle et al. (1992) and Dorigo et al. (2007)).

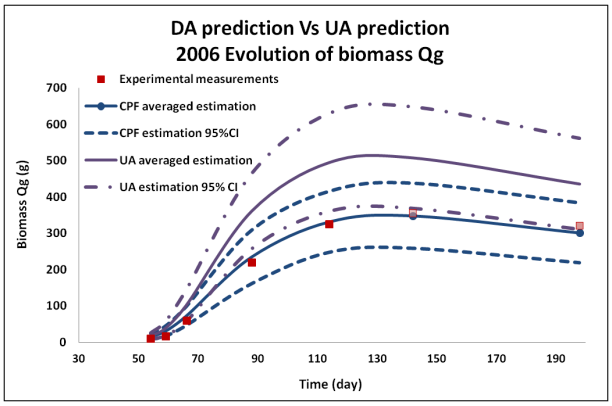

Figure 2: Comparison of the predictions for Q_g of the LNAS model given by the CPF approach with ICPF's estimates and predictions given by the uncertainty analysis (UA) based on the 2006 dataset. The red squares correspond to the assimilated experimental data while the pink squares represent the data used for validation.

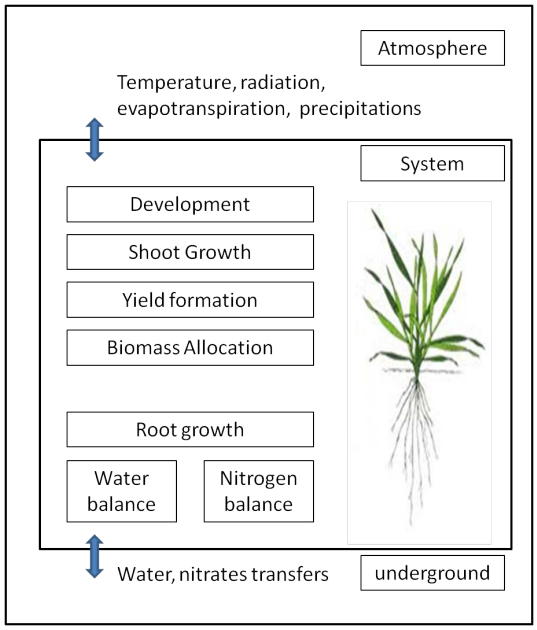

Figure 3: General scheme of the STICS model for winter wheat (adapted from Brisson et al. (2008)).

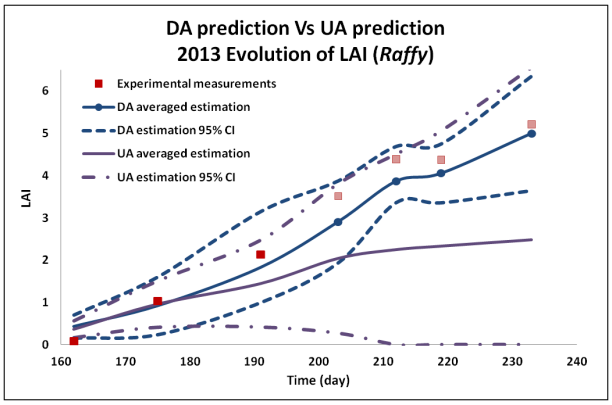

Figure 4: Comparison of the predictions of LAI obtained with CPF-based assimilation and by the uncertainty analysis (without assimilation) based on the 2013 dataset for variety *Raffy*. The red squares correspond to the assimilated experimental data while the pink squares represent the data used for validation.

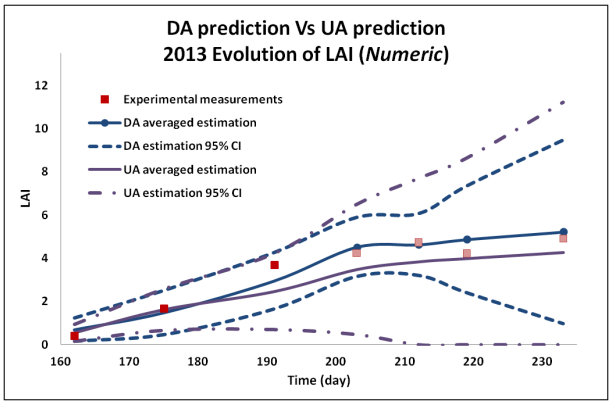

Figure 5: Comparison of the predictions of LAI obtained with CPF-based assimilation and by the uncertainty analysis (without assimilation) based on the 2013 dataset for variety *Numeric*. The red squares correspond to the assimilated experimental data while the pink squares represent the data used for validation.

List of Tables

Parameter	ICPF		CPF	
	Estimates	Std.	Estimates	Std.
μ	3.55	0.16	3.50	0.01
λ	56.6	3.9	57.7	7.7
γ_0	0.925	0.091	0.864	0.104
γ_f	0.104	0.027	0.099	0.013
μ_a	553.9	86.5	678.7	26.1
σ_Q	0.011	-	0.021	-
σ_γ	0.013	-	0.080	-
σ_g	0.098	-	0.102	-
σ_r	0.070	-	0.072	-

Table 1: Estimated values and approximated standard deviations provided by the conditional ICPF approach and the classical CPF approach for the 5 functional parameters and 4 noise parameters of the LNAS model based on the 2010 experimental data.

	Real Data 2006	DA estimates (relative error in %)	95% CI	UA estimates (relative error in %)	95% CI
$Q_b\,(t_{142})$	355.2	348.1 (2.0%)	[258.7; 437.4]	507.8 (43.0%)	[368.4; 647.3]
$Q_b\,(t_{198})$	320.6	301.3 (6.0%)	[219.0; 383.6]	435.7 (35.9%)	[384.3; 560.7] *
$Q_r\,(t_{142})$	1459.2	1716.2 (17.6%)	[1427.9; 2004.5]	1930.7 (32.3%)	[1603.0; 2258.4] *
$Q_r\,(t_{198})$	2400.0	2644.3 (10.2%)	[2209.4; 3079.2]	2942.9 (22.6%)	[2455.0; 3430.7] *

Table 2: Comparison of the prediction results of the LNAS model with and without data assimilation based on the 2006 dataset (density 9.6) according to estimated noises. *: the predicted credibility interval does not contain the real observed data.

	Real Data 2008	DA estimates (relative error in %)	95% CI	UA estimates (relative error in %)	95% CI
$Q_b\,(t_{122})$	373.5	417.8 (11.9%)	[314.8; 520.8]	527.1 (41.1%)	[380.8; 673.5]
$Q_b\,(t_{158})$	380.6	399.2 (4.9%)	[294.8; 503.5]	502.9 (32.1%)	[359.8; 646.0]
$Q_r\,(t_{122})$	1559.1	1531.1 (1.8%)	[1275.5; 1786.7]	1656.4 (6.2%)	[1373.8; 1938.9]
$Q_r\,(t_{158})$	2327.7	2192.9 (5.8%)	[1831.2; 2554.6]	2352.4 (1.1%)	[1960.4; 2744.4]

Table 3: Comparison of the prediction results of the LNAS model with and without data assimilation based on the 2008 dataset (density 10.9) according to estimated noises.

	Real Data 2008	DA estimates (relative error in %)	95% CI	UA estimates (relative error in %)	95% CI
$Q_b\,(t_{122})$	318.6	426.5 (33.9%)	[317.8; 535.3]	552.8 (73.5%)	[403.1; 702.5]*
$Q_b\,(t_{158})$	385.5	409.5 (6.2%)	[300.0; 519.1]	523.6 (35.2%)	[378.2; 669.0]
$Q_r\,(t_{122})$	1319.5	1539.5 (16.7%)	[1302.0; 1777.0]	1689.2 (28.0%)	[1401.5; 1977.0]*
$Q_r\,(t_{158})$	2368.5	2260.9 (4.5%)	[1934.0; 2587.8]	2416.0 (2.0%)	[1983.4; 2848.7]

Table 4: Comparison of the prediction results of the LNAS model with and without data assimilation based on the 2008 dataset (density 16.4) according to estimated noises. *: the predicted credibility interval does not contain the real observed data.

	Real Data 2008	DA estimates (relative error in %)	95% CI	UA estimates (relative error in %)	95% CI
$Q_b\,(t_{122})$	297.6	383.1 (28.7%)	[283.4; 482.8]	483.0 (62.3%)	[342.1; 624.0]*
$Q_b\,(t_{158})$	408.1	376.5 (7.7%)	[300.0; 480.8]	467.5 (14.6%)	[328.5; 606.6]
$Q_r\,(t_{122})$	1551.4	1482.8 (4.4%)	[1232.5; 1733.1]	1590.7 (2.5%)	[1316.9; 1864.5]
$Q_r\,(t_{158})$	2535.0	2157.8 (14.9%)	[1779.6; 2536.0]	2285.6 (9.8%)	[1886.3; 2684.8]

Table 5: Comparison of the prediction results of the LNAS model with and without data assimilation based on the 2008 dataset (density 5.4) according to estimated noises. *: the predicted credibility interval does not contain the real observed data.

EFFICIENCE: maximum radiation use efficiency ($g \cdot MJ^{-1}$)
STAMFLAXV: duration between the day of stage maximal of leaf growth (the end of juvenile phase) and the day of the stage maximal of leaf area index (deg C.days)
UDLAIMAXP: maximal daily relative development of LAI (no unit)
VLAIMAXP: daily relative development of LAI at the inflexion point (no unit)

Table 6: Definition of the selected parameters for the calibration of STICS.

Parameter	Prior Distribution	CPF Estimates	Std.
EFFICIENCE	$\mathscr{N}(4.00, 0.20^2)$	3.62	0.19
STAMFLAXV	$\mathscr{U}(250, 350)$	277.05	35.88
UDLAIMAXP	$\mathscr{N}(3.00, 0.5^2)$	2.89	0.67
VLAIMAXP	$\mathscr{N}(2.20, 0.2^2)$	1.91	0.14

Table 7: Estimated values and approximated standard deviations obtained by CPF for the 4 selected parameters of the STICS winter wheat model, based on the 2012 experimental data.

	Real Data 2013	DA estimates (relative error in %)	95% CI	UA estimates (relative error in %)	95% CI
$hur\,(t_{210})$	0.296	0.311 (5.2%)	[0.298; 0.324]	0.308 (4.0%)	[0.304; 0.312]
$hur\,(t_{226})$	0.296	0.305 (3.2%)	[0.292; 0.318]	0.303 (2.3%)	[0.296; 0.310]
$hur\,(t_{240})$	0.293	0.304 (3.7%)	[0.291; 0.317]	0.301 (2.8%)	[0.294; 0.308]
$hur\,(t_{255})$	0.247	0.266 (7.8%)	[0.253; 0.279]	0.273 (10.6%)	[0.255; 0.292]
$hur\,(t_{275})$	0.252	0.253 (0.4%)	[0.238; 0.268]	0.268 (6.5%)	[0.233; 0.304]
$LAI\,(t_{203})$	3.51	2.91 (17.2%)	[1.94; 3.87]	2.04 (41.9%)	[0.271; 3.809]
$LAI\,(t_{212})$	4.39	3.87 (11.9%)	[3.37; 4.69]	2.25 (49.8%)	[0.000; 4.512]
$LAI\,(t_{219})$	4.37	4.06 (7.2%)	[3.36; 4.75]	2.33 (46.7%)	[0.000; 5.062]
$LAI\,(t_{233})$	5.21	5.00 (4.0%)	[3.65; 6.35]	2.48 (52.4%)	[0.000; 6.557]
$magrain\,(t_{266})$	829.933	769.700 (7.3%)	[613.176; 926.224]	449.111 (45.9%)	[41.079; 857.143]

Table 8: Comparison of the prediction results of the STICS model with and without data assimilation based on the 2013 dataset for variety *Raffy*.

	Real Data 2013	DA estimates (relative error in %)	95% CI	UA estimates (relative error in %)	95% CI
$hur\,(t_{198})$	0.276	0.272 (1.7%)	[0.260; 0.283]	0.290 (5.0%)	[0.286; 0.295]
$hur\,(t_{210})$	0.274	0.303 (10.3%)	[0.284; 0.322]	0.308 (12.2%)	[0.303; 0.313]
$hur\,(t_{226})$	0.289	0.297 (2.6%)	[0.278; 0.316]	0.303 (4.7%)	[0.295; 0.310]
$hur\,(t_{255})$	0.239	0.262 (9.7%)	[0.240; 0.283]	0.272 (14.0%)	[0.253; 0.291]
$hur\,(t_{275})$	0.257	0.260 (1.2%)	[0.229; 0.291]	0.267 (4.0%)	[0.233; 0.302]
$LAI\,(t_{203})$	4.235	4.514 (6.6%)	[3.144; 5.883]	3.482 (17.8%)	[0.471; 6.494]
$LAI\,(t_{212})$	4.735	4.627 (2.3%)	[3.192; 6.062]	3.842 (18.9%)	[0.000; 7.693]
$LAI\,(t_{219})$	**4.225**	4.867 (15.2%)	[2.394; 7.340]	3.994 (5.5%)	[0.000; 8.638]
$LAI\,(t_{233})$	4.910	5.215 (6.2%)	[0.969; 9.462]	4.274 (13.0%)	[0.000; 11.214]
$magrain\,(t_{266})$	819.38	804.33 (1.8%)	[611.68; 996.98]	550.58 (32.8%)	[128.97; 972.20]

Table 9: Comparison of the prediction results of the STICS model with and without data assimilation based on the 2013 dataset for variety *Numeric*.