Ecological Modelling 290 (2014) 165–177

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Data assimilation to reduce uncertainty of crop model prediction with Convolution Particle Filtering

Yuting Chen*, Paul-Henry Cournède

Laboratoire MAS, Ecole Centrale Paris, Grande Voie des Vignes, 92290 ChÃćtenay-Malabry, France

ARTICLE INFO

Article history:
Available online 19 February 2014

Keywords:
Parameter estimation
Data assimilation
Dynamic crop model
Convolution Particle Filtering
Uncertainty analysis
Sugar beet
Winter wheat
LNAS
STICS
Yield prediction

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.

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.
E-mail addresses: yuting.chen@ecp.fr (Y. Chen), paul-henry.cournede@ecp.fr (P.-H. Cournède).

0304-3800/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2014.01.030

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. Moreover, 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, etc.), 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 with deterministic models as well, 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.