5

Bayesian methods for updating crop-model predictions, applications for predicting biomass and grain protein content

David Makowski[#], Marie-Hélène Jeuffroy[#] and Martine Guérif[##]

Abstract

Crop models can be used for predicting crop quality and for optimizing agricultural practices, but the errors of prediction of these models are often important. The objective of this paper is to describe several methods for improving the accuracy of crop-model predictions with real-time measurements. First, we present a simple linear dynamic crop model simulating winter-wheat biomass production and we show how the Kalman-filter method can be used for updating model predictions. Second, we describe a general framework for updating complex nonlinear dynamic crop models. Finally, we present a case study with a nonlinear crop model predicting winter-wheat grain quality. The results presented in this paper show that Bayesian methods are useful for improving the reliability of the information provided to stakeholders by crop models but that additional research is required for implementing these methods with complex models.

Keywords: data assimilation; dynamic crop model; grain protein content; Kalman filter; prediction

Introduction

Numerous dynamic crop models have been developed for simulating crop growth in function of environmental factors (soil characteristics, climate) and of agricultural practices (date of sowing, nitrogen fertilization…). Some of these models can be used for predicting crop yields and crop quality before harvest. Some models can also be used to determine optimal agricultural practices in order to satisfy economic, environmental or quality objectives. For example, AZODYN (Jeuffroy and Recous 1999) is a dynamic model developed for predicting winter-wheat yield and grain protein content in function of the dates and rates of nitrogen-fertilizer applications and of field characteristics. This model is currently used to improve nitrogen-fertilizer management and to predict yields and grain quality.

Crop models can provide useful information to farmers' advisers and collecting firms, but an important limitation of these models is that their errors of prediction are often large due to uncertainties in equations, in parameters, and in input variables. A promising way for improving crop-model performances is to update model

---

[#] UMR d'agronomie INRA/INA-PG, B.P. 01, 78850 Thiverval-Grignon, France. E-mail: makowski@grignon.inra.fr
[##] Unité CSE, INRA, Domaine Saint-Paul, 84914 Avignon cedex 9, France

predictions with real-time measurements of crop and soil variables like leaf chlorophyll contents, soil water potentials, or biomass measurements. Such measurements can be obtained by farmers or by collecting firm operators.

A natural approach for updating crop-model predictions is to use a Bayesian method (Figure 1). The first step is to describe the crop-model errors by using a statistical model. The crop model and the statistical model are used to derive a prior distribution of the model state variables (biomass, nitrogen uptake…) at the time of the measurement. This prior distribution represents a first guess of the state-variable values and takes into account uncertainty in model prediction. The second step is to compute a posterior distribution from the prior distribution, the measurements, and the measurement errors. The posterior distribution is then used to derive an updated prediction of the model state variables. This two-step procedure can be repeated with several measurements obtained at different dates. Methods have recently been developed for complex nonlinear models (Burger, Van Leeuwen and Evensen 1998; Anderson and Anderson 1999; Doucet, Godsill and Andrieu 2000) and have been applied to ecosystem and hydrological models (Allen, Eknes and Evensen 2002; Margulis et al. 2002).
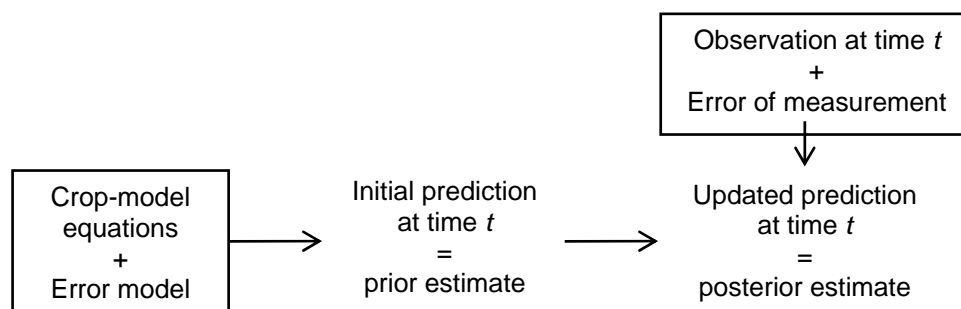

Figure 1. General framework for updating crop-model predictions

The objective of this paper is to introduce several methods for improving crop-model predictions. First, we present a simple application of the Kalman-filter method to a linear dynamic crop model predicting only one state variable, namely winter-wheat biomass. We show how the model predictions can be sequentially updated by using several measurements, and we study the sensitivity of the results to the variance of the model errors. Second, we introduce several methods for updating predictions of complex nonlinear crop models. One of these methods is applied to a nonlinear model predicting an important winter-wheat quality criterion, grain protein content. In this case study we show how the model predictions can be updated by using a chlorophyll-content measurement obtained at flowering and we study the sensitivity of the results to the value of the measurement.

A simple example

A dynamic model simulating wheat biomass

We describe here a simple dynamic crop model that will be used in the next section to illustrate the principle of the Kalman-filter method. Our crop model has a single-state variable representing above-ground winter-wheat biomass. This state variable is simulated on a daily basis in function of the daily temperature and the daily incoming radiation according to the classical efficiency approach (Varlet-Grancher et al. 1982). The biomass at time $t+1$ is linearly related to the biomass at time t as follows:

$$U_{t+1} = U_t + E_b E_{i\max} \left[1 - e^{-K\,LAI_t} \right] PAR_t + \varepsilon_t \qquad (1)$$

where t is the day number since sowing, U_t is the true above-ground plant biomass on day t, PAR_t is the incoming photosynthetically active radiation on day t, LAI_t is the leaf-area index on day t and ε_t is a random term representing the model error. The crop biomass at sowing is set equal to zero: $U_1 = 0$. LAI_t is calculated in function of the cumulative degree-days (over a basis of 0°C) from sowing until day t, noted T_t, as follows (Baret 1986):

$$LAI_t = L_{\max} \left\{ \frac{1}{1 + e^{-A[T_t - T_{s1}]}} - e^{B[T_t - T_{s2}]} \right\}.$$

Parameter T_{s2} is set equal to $T_{s2} = \dfrac{1}{B} \log \left[1 + \exp(A \times T_{s1}) \right]$ in order to have $LAI_1 = 0$.

The model includes two input variables $X_t = [T_t, PAR_t]'$ and seven parameters $\theta = \left(E_b, E_{i\max}, K, L_{\max}, A, B, T_{s1} \right)'$. E_b is the radiation use efficiency which expresses the biomass produced per unit of intercepted radiation, $E_{i\max}$ is the maximal value of the ratio of intercepted to incident radiation, K is the coefficient of extinction of radiation, $L_{\max}$ is the maximal value of LAI, T_{s1} defines a temperature threshold, and A and B are two additional parameters. The parameter values were determined in previous studies (Baret 1986).

On a practical point of view, model (1) can be used either for agronomical diagnosis or for estimating crop yield from biomass predictions. With this crop model, the crop biomass predicted at time t_P is simply:

$$\hat{U}_{t_P} = \sum_{t=1}^{t_p - 1} E_b E_{i\max} \left[1 - e^{-K\,LAI_t} \right] PAR_t \qquad (2)$$

We suppose that 10 measurements of biomass, $m_{t1}, ..., m_{t10}$, are made at different times before harvest on the site-year of interest. In practice, values of m_t can be derived from plant samples or from remote-sensing data. We assume that each measurement m_t is related to the biomass U_t by

$$m_t = U_t + \tau_t \qquad (3)$$

where τ_t is a random term representing measurement errors. In the next section we show how such measurements can be used to improve the accuracy of biomass predictions.

Standard Kalman filter

The Kalman filter (Kalman 1961) is a method for updating sequentially model predictions with measurements obtained at different dates. This method is presented here within a Bayesian framework with the simple model described in the previous section. The general idea is to compute the posterior distribution of the state variable U_t:

$$P\left[U_t | M_t \right] = P\left[U_t | m_t, M_{t-1} \right]$$

where M_{t-1} is the set of all the measurements taken at/or before time t-1, m_t is a new measurement obtained at time t, and M_t is the set including both M_{t-1} and m_t. The probability distribution $P\left[U_t | M_t \right]$ can be used to predict the state of the model or to

provide the prior distribution at the time of the next available observation. Applying Bayes' rule to $P[U_t | M_t]$ gives $P[U_t | M_t] = c\, P[m_t | U_t, M_{t-1}] P[U_t | M_{t-1}]$ where c is a constant independent of U_t. With independent measurements we obtain:

$$P[U_t | M_t] = c\, P[m_t | U_t] P[U_t | M_{t-1}]\qquad(4)$$

Eq.(4) shows that the posterior distribution depends on two terms, namely $P[U_t | M_{t-1}]$ and $P[m_t | U_t]$. The first term, $P[U_t | M_{t-1}]$, represents the prior distribution of the state U_t calculated from earlier observations. This is a first guess distribution for the dry matter at time t before making the new observation m_t. The second term, $P[m_t | U_t]$, depends on the new measurement and evaluates how likely it is to obtain m_t if the true dry matter is equal to U_t. To apply the Kalman filter to our simple model, it is necessary to specify the distribution of the error terms associated with the model and with the measurements, ε_t and τ_t. We assume further that $\varepsilon_t \sim N(0, Q)$ and $\tau_t \sim N[0, R_t]$.

As no observation is available before the date of the first measurement t_1, the prior distribution at time t_1 is defined simply by $P[U_{t_1}] = N\{\hat{U}_{t_1}, \sigma_{t_1}^2\}$ with

$\hat{U}_{t_1} = \sum_{t=1}^{t_1 - 1} E_b E_{i\max} \left[1 - e^{-K\, LAI_t} \right] PAR_t$ and $\sigma_{t_1}^2 = (t_1 - 1)\, Q$. The prior mean $\hat{U}_{t_1}$ represents an initial biomass prediction at time t_1. With these notations, eq.(4) becomes $P[U_{t_1} | m_{t_1}] = c\, P[m_{t_1} | U_{t_1}] P[U_{t_1}]$ with $P[m_{t_1} | U_{t_1}] = N[U_{t_1}, R_{t_1}]$. It follows that the posterior distribution at time t_1 is defined by $P[U_{t_1} | m_{t_1}] = N\{\hat{U}_{K,t_1}, \sigma_{K,t_1}^2\}$ with:

$$\hat{U}_{K,t_1} = \hat{U}_{t_1} + K_{t_1} \left[m_{t_1} - \hat{U}_{t_1} \right]\qquad(5)$$

$$\sigma_{K,t_1}^2 = \left[1 - K_{t_1} \right] \sigma_{t_1}^2\qquad(6)$$

$$\text{and } K_{t_1} = \frac{\sigma_{t_1}^2}{\sigma_{t_1}^2 + R_{t_1}} = \frac{(t_1 - 1) Q}{(t_1 - 1) Q + R_{t_1}}\qquad(7)$$

Eq.(5) shows that the expected value of the posterior distribution, $\hat{U}_{K,t_1}$, is defined as a linear combination of the initial prediction $\hat{U}_{t_1}$ and a weighted difference between the actual measurement and the initial prediction. The weight (7) depends on the errors associated with the model and with the measurements. Eq.(5) and eq.(7) show that the posterior mean $\hat{U}_{K,t_1}$ can differ strongly from the prior mean $\hat{U}_{t_1}$ when the error associated with the measurement is low compared to the error associated with the crop-model prediction. The posterior mean $\hat{U}_{K,t_1}$ represents an updated prediction of the dry matter at time t_1 and takes into account both the initial prediction and the measurement. In case of perfect measurement, we have $R_{t_1} = 0$, $K_{t_1} = 1$, and $\hat{U}_{K,t_1} = m_{t_1}$. On the contrary, when R_{t_1} is very large, K_{t_1} is near from zero and $\hat{U}_{K,t_1}$ is not very different from the initial prediction. Eq.(6) and eq.(7) show that the variance of the posterior distribution $P[U_{t_1} | m_{t_1}]$ is lower than the variance of the prior distribution $P[U_{t_1}]$: as $K_{t_1} < 1$, we have $\sigma_{K,t_1}^2 < \sigma_{t_1}^2$. This result shows that the

measurement m_{t_1} reduces the uncertainty in the predicted biomass. The extent of this reduction depends on the variance associated with the measurement. Lower R_{t_1}, lower σ^2_{K,t_1}.

Another measurement, m_{t_2}, is obtained at time t_2. The prior distribution of biomass at time t_2 is defined by:

$$P\left[U_{t_2} \middle| m_{t_1}\right] = N\left\{\hat{U}_{t_2}, \sigma^2_{t_2}\right\}$$

where the prior mean and variance are defined by

$$\hat{U}_{t_2} = \hat{U}_{K,t_1} + \sum_{t=t_1}^{t_2-1} E_b E_{i\max}\left[1 - e^{-K\ LAI_t}\right]PAR_t \quad \text{and} \quad \sigma^2_{t_2} = \sigma^2_{K,t_1} + (t_2 - t_1)\ Q, \text{ respectively.}$$

$\hat{U}_{t_2}$ is the initial prediction of biomass at time t_2. It depends on the updated prediction at time t_1 $\hat{U}_{K,t_1}$ (and, so, on the first measurement) but not on m_{t_2}. The new measurement m_{t_2} is used to compute a new posterior distribution defined by:

$$P\left[U_{t_2} \middle| m_{t_2}, m_{t_1}\right] = N\left\{\hat{U}_{K,t_2}, \sigma^2_{K,t_2}\right\}$$

where $\hat{U}_{K,t_2} = \hat{U}_{t_2} + K_{t_2}\left[m_{t_2} - \hat{U}_{t_2}\right]$, $\sigma^2_{K,t_2} = \left[1 - K_{t_2}\right]\sigma^2_{t_2}$ and $K_{t_2} = \dfrac{\sigma^2_{t_2}}{\sigma^2_{t_2} + R_{t_2}}$. $\hat{U}_{K,t_2}$ is the

updated biomass prediction at time t_2.

This procedure is used sequentially with the eight other measurements obtained before harvest.

Numerical application

We present here a numerical application of the Kalman filter. First, the crop model described in section 'A dynamic model simulationg wheat biomass' is used alone for estimating biomass for a field located in the South of France (Camargue) for the year 1987. The predicted values of biomass are calculated by applying eq.(2) each day between sowing and harvest. Second, we use the Kalman filter to update sequentially the crop-model predictions with 10 biomass measurements. The posterior distribution (4) is computed at each date of measurement with the Kalman-filter equations. We obtain successively 10 posterior distributions: $P\left[U_{t_1} \middle| M_{t_1}\right]$, $P\left[U_{t_2} \middle| M_{t_2}\right]$, ..., $P\left[U_{t_{10}} \middle| M_{t_{10}}\right]$ with $M_{t_1} = \left\{m_{t_1}\right\}$, $M_{t_2} = \left\{m_{t_1}, m_{t_2}\right\}$, ..., $M_{t_{10}} = \left\{m_{t_1}, ..., m_{t_{10}}\right\}$. As shown above, the posterior distributions depend on the variances of the measurement errors R_t and on the variance of the daily model error Q. In this study the values R_t are set equal to the empirical variances calculated from replicates. Different values of Q in the range 1-10 (g/m²)² are tested successively in order to study the influence of the model error variance on the posterior distribution. The biomass at harvest is estimated with the crop model by using the expected value of the last posterior distribution $P\left[U_{t_{10}} \middle| M_{t_{10}}\right]$ as a starting point.
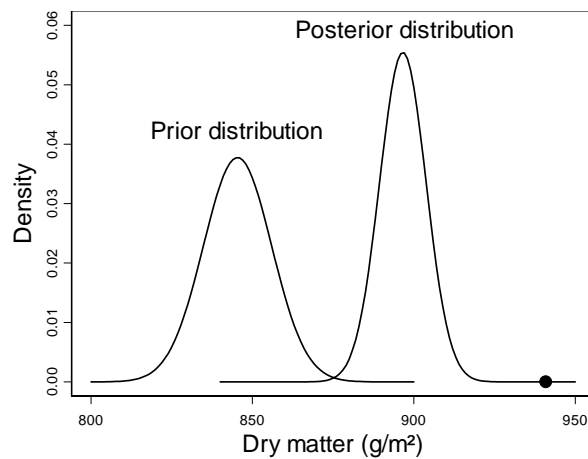
Figure 2. Prior and posterior distributions of biomass at time t_{10} obtained with Q=10 (g/m²)². The black dot indicates the value of biomass measured at time t_{10}

   Figure 2 shows the full prior and posterior distributions at time t_{10}=191 days (time of the last measurement) obtained with Q=10 (g/m²)². Both distributions are Normal but the two distributions are characterized by very different expected values and variances. The prior mean (845.4 g/m²) is lower than the posterior mean (896.6 g/m²) and the latter is nearer to the measured biomass value. It is important to note that the posterior mean is not strictly equal to the biomass measured at time t_{10}. This is due to the error associated with the measurement. As shown in eq.(5), the posterior mean is a weighted sum of the prior mean and of the measurement. The weight depends both on the variance of the prior distribution and on the variance associated with the measurement. The posterior distribution results from an optimal combination of the prior distribution and of the measurement. Another interesting result is that the prior variance (111.7 (g/m²)²) is much higher than the posterior variance (51.8 (g/m²)²). This result shows that the use of the measurement at time t_{10} has reduced the uncertainty in the crop-model prediction.
   Figure 3 presents the initial and updated crop-model predictions obtained between sowing and harvest. The updated predictions reported in Figure 3 (continuous lines) correspond to the expected values of the posterior biomass distributions. For example, the updated prediction at time t_{10} obtained with Q=10 is equal to the expected value of the posterior distribution presented in Figure 2. Figure 3 shows that errors of predictions are large when the crop model is not updated with measurements (dashed curve). At harvest, the initial biomass prediction (1698.2 g/m²) is much higher than the measured value (1486.7 g/m²). The biomass at harvest is more accurately predicted when the crop model is adjusted to the 10 measurements by using the Kalman filter (continuous curves). Figure 3 illustrates the influence of the size of the model error Q on the results of the Kalman filter. The model predictions are more strongly adjusted to the measurements with Q=10 than with Q=1. This is logical because high values of Q tend to increase the Kalman coefficient K_t. However, for the situation considered in Figure 3, the value of Q has almost no influence on the accuracy of the prediction at harvest. This result is not general and, in practice, the parameters describing the model errors must be chosen carefully, for instance by using a training dataset.
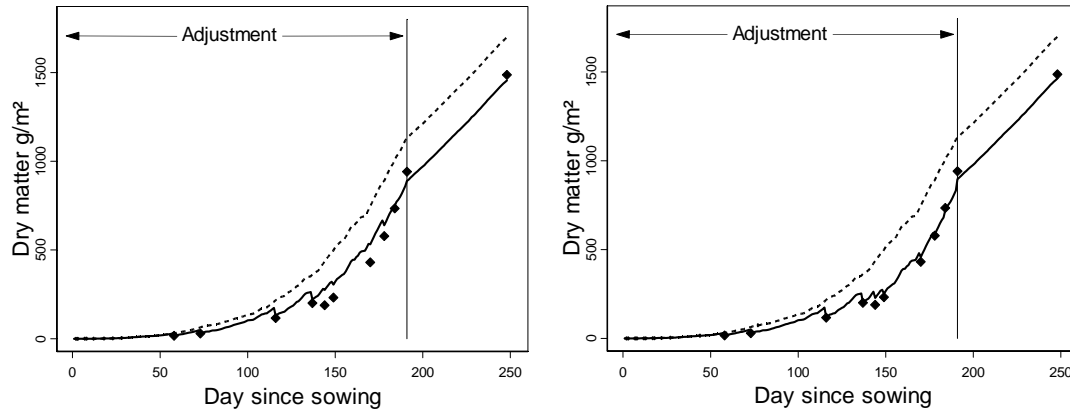
Figure 3. Initial model predictions (dashed line) and updated model predictions (continuous line). Model predictions were updated with 10 measurements by applying the Kalman-filter method with Q=1 (left) and with Q=10 (right). Black dots represent biomass measurements. The measurement obtained at time t=248 is the biomass measured at harvest

Methods for complex nonlinear crop models

General framework

A nonlinear crop model can be defined by :

$$\varphi_{t+1} = F(\varphi_t, X_t; \theta) + \varepsilon_t \tag{8}$$

where φ_t is the $(p \times 1)$ vector including the true p state variables values at time t, X_t is the vector including the input variable at time t (radiation, temperature, fertilizer dose, …), F is a nonlinear function, θ is the set of parameters, and ε_t is a vector including p error terms. There are two important differences between model (8) and the simple crop model presented in 2.1. First, model (8) includes several state variables and the errors associated with the different state variables are likely to be correlated. Second, model (8) is a nonlinear model: the state variable at time $t+1$ is a nonlinear function F of the state variable at time t.

Suppose that a measurement m_t is available at time t and that m_t is related to the state variable vector by:

$$m_t = H\varphi_t + \tau \tag{9}$$

where H is $(1 \times p)$ one-row matrix and τ is an error term. Our problem is to compute the conditional density $P(\varphi_t | m_t)$ *i.e.,* the posterior distribution of the state variables. An analytical expression of $P(\varphi_t | m_t)$ can easily be derived when the model is linear (see section 'Standard Kalman filter'). On the contrary, this is generally impossible with complex crop models. Several methods have been proposed to solve this problem. A first method is called Extended Kalman filter. The principle is to linearize eq.(8) and to apply the standard Kalman-filter method to the following model:

$$\varphi_{t+1} = F(\hat{\varphi}_t, X_t; \theta) + A_t(\varphi_t - \hat{\varphi}_t) + \varepsilon_t$$

where A_t is a $(p \times p)$ matrix of partial derivatives of F with respect to the five state variables, $\hat{\varphi}_t$ is the predicted state variable at time t, ε_t is a $(p \times 1)$ error-term vector assumed to be Normally distributed. The main drawback of this method is that the linearization has been shown to be invalid in a number of applications. The linear approximation does not always describe properly the error evolution of the model.

63

The Ensemble Kalman filter is another popular method described by Burgers, Van Leeuwen and Evensen (1998). The principle is to approximate the prior and posterior distributions by an ensemble of state-variable values. First, an ensemble of N values of φ_t, φ_t^1, ..., φ_t^j, ..., φ_t^N and an ensemble of N values of m_t, m_t^1, ..., m_t^j, ..., m_t^N, are randomly generated. Second, the Kalman-filter equation is applied to each ensemble element as follows:

$$\varphi_{t,K}^j = \varphi_t^j + K_t^e \left(m_t^j - H\varphi_t^j \right) \tag{10}$$

where K_t^e is a ($p\times1$) vector defined by $K_t^e = \Sigma_t^e H' \left(H\Sigma_t^e H' + R_t \right)^{-1}$, R_t is the variance of the measurement error, and Σ_t^e is the ($p\times p$) variance−covariance matrix of N state vectors φ_t^j, $j=1,\ldots$, N. The ensemble of state variables φ_t^j, $j=1,\ldots$, N, describes the uncertainty in the state-variable values before using the measurement. In this approach, the updated crop-model prediction is set equal to the average value of the $\varphi_{t,K}^j$, $j=1$, ..., N, noted further $\bar{\varphi}_{t,K}$. Note that $\bar{\varphi}_{t,K}$ is related to the average value, $\bar{\varphi}_t$, of the initial ensemble φ_t^j, $j=1$, ..., N, by

$$\bar{\varphi}_{t,K} = \bar{\varphi}_t + K_t^e \left(\bar{m}_t - H\bar{\varphi}_t \right) \tag{11}$$

where $\bar{m}_t$ is the average value of m_t^j, $j=1$, ..., N. $\bar{\varphi}_t$ and $\bar{\varphi}_{t,K}$ can be interpreted as prior and posterior means, respectively. The attractive feature of this method is that its implementation does not require a linear approximation of the crop model. However, it is necessary to choose the value of N, to define a procedure for generating φ_t^j, $j=1,\ldots$, N, and to define another procedure for generating m_t^j, $j=1,\ldots$, N. The values of m_t^j, $j=1,\ldots$, N, can simply be generated by adding random terms to m_t: $m_t^j = m_t + \tau_t^j$ with $\tau_t^j \sim N(0, R_t)$ (Burger, Van Leeuwen and Evensen 1998). This is straightforward if R_t is known. On the contrary, there is no systematic method for choosing N and for generating the ensemble of vectors φ_t^j, $j=1,\ldots$, N. The value of N is generally selected by trial and error. For generating the ensemble of state variables, a common approach consists in calculating N vectors of state variables at each time step as follows (e.g. Allen, Eknes and Evensen 2002):

$$\varphi_{t+1}^j = F\left(\varphi_t^j, X_t; \theta \right) + \varepsilon_t^j$$

where $\varepsilon_t^j \sim N(0, Q_t)$. The procedure requires the knowledge of Q_t. As before, different values Q_t can be tested by using a training data set. Another approach is to generate randomly N values for all the uncertain elements of the crop models (parameters and input variables) (Margulis et al. 2002).

Finally, particle filter is another method to compute the posterior distribution. This method is not described here. See Doucet, Godsill and Andrieu (2000) for more details.

Application to a crop model predicting grain protein content

The AZODYN crop model (Jeuffroy and Recous 1999) is a nonlinear dynamic model simulating winter-wheat crop in function of environmental variables (characteristics of the crop at the end of winter, soil characteristics, climate) and of nitrogen fertilization (dates and rates of fertilizer applications). This model can be used to predict grain yield, soil mineral nitrogen, and grain protein content at harvest.

AZODYN is a useful tool for studying the effects of nitrogen management on crop yield, grain quality and risk of pollution by nitrate (Meynard et al. 2002). Before flowering, five state variables are simulated each day by AZODYN: nitrogen uptake (NU), dry matter (DM), nitrogen-nutrition index (NNI), leaf-area index (LAI), soil mineral nitrogen supply (SNS).

We consider chlorophyll-content measurements obtained with a chlorophyll meter. These measurements are correlated to one of the model state variables, namely nitrogen uptake, and can be easily performed by farmers, collecting-firm operators, or farmers' advisors. Here, we suppose that only one chlorophyll-content measurement is performed at flowering and that this measurement is linearly related to the model state variables as follows:

$$m_{flo} = \mu + H \varphi_{flo} + \tau \tag{12}$$

where m_{flo} and φ_{flo} are, respectively, the chlorophyll-content measurement and the (5×1) vector of the true state-variable values at flowering, μ is an intercept parameter, and H is a one-row matrix defined by $H = (\alpha, 0, 0, 0, 0)$ where α is the slope of the linear equation relating the measurement to nitrogen uptake. We assume that the error term τ is Normally distributed, $\tau \sim N(0, R)$.

The Ensemble Kalman filter is implemented with AZODYN. The objective is to study the sensitivity of the model predictions to the value of the measurement m_{flo} obtained at flowering. We consider a particular site-year (Grignon-1995) and we use AZODYN to predict the characteristics of a wheat crop at harvest, specifically yield and grain protein content. The crop model is first run without updating the simulations with any chlorophyll-content measurement. Next, the model is run a second time and the Ensemble Kalman-filter method is used to update the five state variables of AZODYN with a single chlorophyll-content measurement performed at flowering. Yield and grain protein content at harvest are then computed from the updated state variables.

To apply the Ensemble Kalman-filter method, we generate $N=150$ state-variable vectors at time $t=1$. Each vector contains five elements representing particular realizations of the five state variables. The N vectors are generated by sampling N values for each of seven uncertain parameters and input variables. Parameter and input-variable distributions were specified from expert knowledge. The N state-variable vectors are propagated forward in time until the date of measurement (flowering). Figure 4 shows the evolution of ten realizations of two state variables, nitrogen uptake and soil nitrogen supply, between the end of winter and flowering.
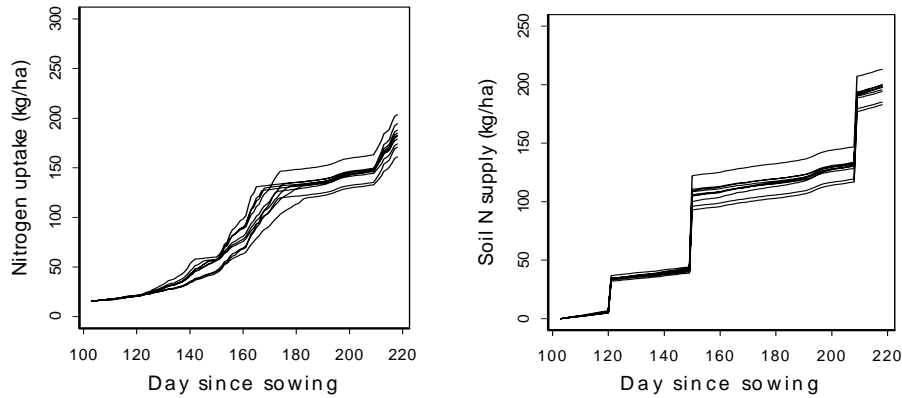
Figure 4. Ten realizations of nitrogen uptake and soil nitrogen supply between the end of winter and flowering for Grignon-1995. Nitrogen fertilizer was applied at days 121, 150, and 209

The (5×5) variance−covariance matrix of the model errors, Σ_t^e, is computed at t=flowering from the 150 state-variable vectors obtained at the date of measurement. Then eq.(10) is applied to each vector with $\mu = 30.26$, $\alpha = 0.083$, and $R = 8.11$, these parameters being estimated by linear regression on 42 measurements of nitrogen uptake and chlorophyll content obtained in Grignon at flowering. The (5×1) vector K_t^e is computed at flowering from Σ_t^e, α, and R as explained in 3.1, and we obtain $K_t^e = \left[K_{NU}, K_{DM}, K_{LAI}, K_{NNI}, K_{SNS}\right]^{'} = \left[2.67, 384.7, 0.13, -0.005, 1.99\right]^{'}$, where K_{NU}, K_{DM}, K_{LAI}, K_{NNI}, K_{SNS} are the coefficients associated with the five state variables. Note that all the state variables are updated by using a single chlorophyll-content measurement, even if only nitrogen uptake is explicitly related to chlorophyll content. For example, at flowering, the updated value of soil N supply is equal to $\bar{\varphi}_{SNS} + K_{SNS}\left[m_{flo} - \mu - \alpha\,\bar{\varphi}_{NU}\right]$, where m_{flo} is the measurement performed at flowering, $\bar{\varphi}_{NU}$ and $\bar{\varphi}_{SNS}$ are, respectively, the means of the 150 nitrogen-uptake and soil nitrogen-supply values obtained at flowering before correction. K_{SNS} depends on Σ^e and notably on the covariance between nitrogen uptake and soil nitrogen supply.
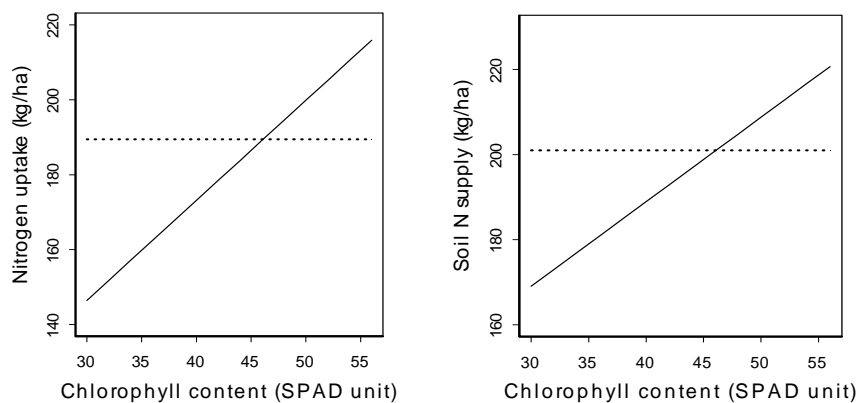

Figure 5. Sensitivity of the updated values of nitrogen uptake and soil N supply at flowering (continuous lines) to chlorophyll-content measurement. Dotted lines indicate the initial model predictions

Figure 5 shows the initial and updated values of crop nitrogen uptake at flowering and cumulated soil N supply from sowing to flowering for different values of the measurement m_{flo} taken in the range 30-55. Each value reported in Figure 5 results from averaging 150 individual values. Compared to the initial state-variable values (dotted lines), the updated values (continuous lines) are either increased or decreased depending on the value of the measurement. Figure 6 shows the resulting effects on two variables simulated by AZODYN at harvest, yield and grain protein content. Higher chlorophyll content, higher yield but lower grain protein content.
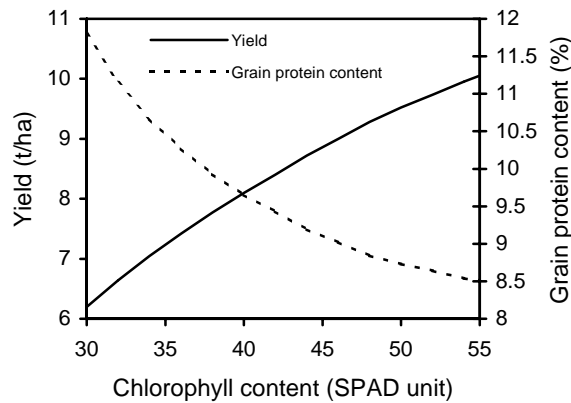

Figure 6. Updated values of yield and grain protein content obtained at harvest in function of chlorophyll-content measurement

Conclusion

The Kalman-filter method is an attractive method for updating crop-model state variables with real-time measurements. Updated values of model state variables can be used either to predict crop characteristics or to optimize farmers' practices in order to achieve economic, quality or environmental objectives. The Kalman filter can be easily applied to simple crop models but its application to complex models is not straightforward. Several methods have recently been developed in order to update predictions of complex nonlinear models, and an application of one of these methods has been presented in this paper. However, we think that additional research is required to demonstrate the operational capabilities of these methods. Their performances depend on several key elements: the variance−covariance matrix of the model errors, the relationship between measurements and state variables, and the number of measurements. Different approaches can be used to estimate the variance−covariance matrix and to relate state variables to measurements. The interest of these approaches will be evaluated in further studies.

References

Allen, J.I., Eknes, M. and Evensen, G., 2002. An ensemble Kalman filter with a complexe marine ecosystem model: hincasting phytoplankton in the Cretan sea. *Annales Geophysicae,* 20, 1-13.

Anderson, J.L. and Anderson, S.L., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review,* 127 (12), 2741-2758.

Baret, F., 1986. *Contribution au suivi radiométrique de cultures de céréales.* Unpublished PhD thesis, Université Orsay.

Burger, G., Van Leeuwen, P.J. and Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review,* 126, 1719-1724.

Doucet, A., Godsill, S. and Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing,* 10, 197-208.

Jeuffroy, M.H. and Recous, S., 1999. Azodyn: a simple model simulating the date of nitrogen deficiency for decision support in wheat fertilization. *European Journal of Agronomy,* 10 (2), 129-144.

Kalman, R.E., 1961. A new approach to linear filtering and prediction theory. *Transactions of the ASME. Journal of Basic Engineering,* 83, 95-108.

Margulis, S.A., McLaughlin, D., Entekhabi, D., et al., 2002. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment. *Water Resources Research,* 38 (12), 35/1-18.

Meynard, J.-M., Cerf, M., Guichard, L., et al., 2002. Which decision support tools for the environmental management of nitrogen? *Agronomie,* 22 (7/8), 817-829.

Varlet-Grancher, C., Bonhomme, R., Chartier, M., et al., 1982. Efficience de la conversion de l'énergie solarie par un couvert végétal. *Acta Oecologica, Oecologia Plantarum,* 3 (1), 3-26.