



**HAL**  
open science

# A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface

Jennifer Pellenq, Gilles Boulet

## ► To cite this version:

Jennifer Pellenq, Gilles Boulet. A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface. *Agronomie*, 2004, 24 (4), pp.197-204. 10.1051/agro:2004017 . hal-00886008

**HAL Id: hal-00886008**

**<https://hal.science/hal-00886008>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface

Jennifer PELLENQ, Gilles BOULET\*

CESBIO (UMR 5126 UPS-CNRS-CNES-IRD), 18 avenue Edouard Belin, BPI 2801, 31401 Toulouse Cedex 4, France

(Received 9 October 2002; accepted 10 June 2003)

**Abstract** – This paper presents a methodology to test the performance of assimilation of satellite data into models for the functioning of the continental surface. This methodology applies the Kalman Ensemble Filter to modelling of plant growth and senescence in conjunction with the water and energy exchanges at the land surface. It belongs to a family of methods known in meteorology and oceanography as the Observing System Simulation Experiment (OSSE) approach. By combining information from modelling and observation, the Kalman Ensemble Filter permits corrections in real time of the simulated state of the continental surface, as well as propagation in time of the associated uncertainties. The OSSE approach may present a first step in designing a decision support system, and also in predicting the usefulness of new types of satellite data.

surface functioning / data assimilation / remote sensing / OSSE / Kalman Ensemble Filter

## 1. INTRODUCTION

Continental surface processes including water, carbon and energy cycles are particularly complex due to the numerous interactions and the large range of the temporal and spatial scales involved [2, 14]. These complex interactions are usually described by (i) dynamic models which propagate the state variables forward in time, (ii) ground or remote-sensing observations and (iii) observation models (including radiative transfer schemes) relating the observation variables to the system state variables. Each of these methods has its own strengths and weaknesses.

Dynamic models describe the evolution of the state variables through space and time in a continuous way, but face the problem of complexity: on the one hand, very sophisticated “state-of-the-art” models have been developed that aim at representing the system in a precise way. But these models require many input parameters whose estimation proves to be difficult. On the other hand, simplified schemes, accounting for only some parts of the interactions, have been developed to overcome the difficulty of applying “state-of-the-art” models for a wide range of spatial scales. For example, Soil-Vegetation-Atmosphere Transfer (SVAT) models usually consider the vegetation characteristics as an imposed function of time whereas vegetation

growth models usually oversimplify the water cycle description. This oversimplification is often integrated into a conceptual parameterisation. Parameters that cannot be derived from measurements are calibrated to fit an objective function over a given period. They are usually tested for a limited number of natural conditions (at several experimental sites) and a limited period of time (over a small number of Intensive Observation Periods).

Models are thus imperfect and accumulate (in variable proportions) uncertainties on the modelling strategy (following the degree of simplification) and on parameter specification (according to the degree of complexity of the model and the accessibility of the parameter value).

On the other hand, remote-sensing data provide repetitive and spatially distributed information about the surface, but this information needs to be interpolated in time and is usually indirectly related to the biophysical variables of interest. Ground data can also be collected at a satisfactory time sampling rate but need to be interpolated in space and time and their collection is usually time-consuming.

Moreover, measurements are always noisy, first because the measurement device is not perfect and induces a measurement error, and secondly because the scale on which the measurement has been made is rarely the one on which the observation

\* Corresponding author: Gilles.Boulet@ird.fr

is required. Interpolation of these observations implies hypotheses or requires ancillary information that induces new uncertainties.

If simulation or observation by itself cannot give an accurate description of the behaviour of our continental system, combining them should help us to monitor the land surface, and improve both predictive and retrospective capabilities of current surface process models.

The set composed of dynamic models, observations and observation models to represent a system defines a *System of Observation*.

Combining both model and observations inside the System of Observation is the task of Data Assimilation methods. These methods aim at minimising uncertainties in the estimation of a given system state in an optimal way (i.e. with statistical criteria). All of them try to reduce the discrepancy between the measured and the simulated observation by adjusting either poorly known parameters (calibration, [12]) or the system state variables. In the latter case, the information provided by the dynamic model is either ignored (inversion, [10]) or taken into account (optimal control and optimal estimation).

In a general way, the performance of Data Assimilation methods strongly depends on the assimilation protocol and the uncertainty specification. Moreover, they act *globally*, not just on one component of the system of observation, and they can affect any part of the system. Subsequently, one needs to perform several tests to find out the best Data Assimilation configuration that suits a particular goal. In meteorology and oceanography a methodology known as the *Observing System Simulation Experiment* (OSSE) approach has been developed for this purpose [1, 5, 16]. It uses optimal control or optimal estimation data assimilation techniques.

Amongst such assimilation techniques, the Kalman Filter can account for information which comes from observation and modelling as well as their respective uncertainty in a simple manner. One assumes in general that the parameters are known a priori and uncertainty comes from the state variables and corresponding observations. The method also allows for the dynamic model errors to be taken into account by adding a modelling noise in the state-space equation:

$$state_{t+dt} = f(state_t, parameters, forcing) + noise \quad (1)$$

By making it possible to study the impact of the assimilation of an observation on the results of a dynamic model, the OSSE approach enables the following questions to be answered:

- 1– Does the assimilation of a particular observation improve all components of the dynamic model's simulations? (“state alteration” problem);
- 2– Can calibration errors be compensated by tuning state variables online instead of parameters? (“calibration complement” problem);
- 3– What is the optimum repetitivity of measurement in order to have a realistic interpolation of a certain variable through the process model that is close enough to the noisy observations? (“observation design” problem);

- 4– What is the most suitable assimilation strategy (noise specification, assimilation method, etc.) to fulfil a particular goal? (“filter design” problem).

In this paper, an illustration of the OSSE approach is presented in the context of plant growth and senescence in conjunction with the water and energy exchanges at the land surface. The interest of assimilating remote-sensing data such as radiative surface temperature and LAI into the modelling is studied through a simple example where the following questions are considered. If midday nadir radiative surface temperature and daily Normalised Difference Vegetation Index (NDVI) are assimilated with given measurement errors, is the simulation of soil moisture and latent heat flux improved? Or are some of the state variables such as soil moisture altered? For instance, how does the filter discriminate between affecting positive innovations to LAI and soil moisture when radiative temperature is largely underestimated?

In the first section, the special features of SVAT and growth modelling regarding uncertainty will be developed. Then the Kalman Filter theory will be briefly described in order to introduce the OSSE in a third section. Finally, an example of application will be given to illustrate the relevance of the method in the context of the functioning of the continental surface.

## 2. MODEL SPECIFICITIES REGARDING UNCERTAINTY

Vegetation growth modelling consists of evaluating production (e.g. net photosynthesis) and losses (e.g. senescence) of biomass, as a function of radiation conditions, and nutrient and water availability. The dynamic equation (time evolution) for green biomass ( $B$ ) can then be written in the general form of:

$$\frac{\partial B}{\partial t} = NetPhotosynthesis(B) - Senescence(B) \quad (2)$$

On the other hand, SVAT schemes estimate soil moisture ( $\theta$ ) evolution by computing the water balance between rain or irrigation supply and drainage, runoff and evaporation losses for a soil of depth  $d$ :

$$d \frac{\partial \theta}{\partial t} = RainAndIrrigation - Losses(\theta) \quad (3)$$

Note that both equations (1) and (2) are linked through the water availability, as plant growth requires water and the water cycle is influenced by this uptake.

The major difference between these equations comes from the fact that the soil moisture evolution resulting from the water balance equation results from a “negative feedback”, whereas the biomass time-series obtained from the growth and senescence equation has, to a certain extent, a “positive feedback”. Indeed, in the SVAT scheme, if there is a positive soil moisture error (overestimation), losses are overestimated as well, and the overestimation decreases with time. If the error is negative, losses are underestimated and compensate the underestimation. In all cases then, errors tend to smooth out with time and soil moisture oscillates around a “long-term equilibrium”. On the other hand, an error in the initial biomass estimate of the growth

model can produce very different behaviours according to the large uncertainties on both photosynthesis (e.g. growth efficiency and stress functional) and senescence (e.g. proportionality factor between senescence and actual biomass) parameters, especially for small soil covers. The system does not compensate biases and a small deviation or error on the estimates can lead to significant changes. This distinction between both state-space equations implies the following: in biomass modelling the impact of initial condition errors is crucial to the prediction of the whole seasonal cycle, whereas from the strict water resources point of view, the “memory” of the system is short. But in the coupled case, the initial water content will have a critical impact on the biomass evolution, and in turn on the seasonal water balance.

Concerning the soil-plant-atmosphere continuum, observable quantities mostly come from: (i) point measurements of the surface characteristics (soil water content, vegetation height, soil texture, root zone depth, etc.), (ii) point measurements of meteorological forcing (precipitation rate, wind speed, atmospheric radiation, etc.), (iii) integrated ground measurements (stream flow, precipitation by radar signal, etc.) and (iv) multi-resolution multi-spectral remote-sensing data (NDVI, radiative temperature, and some information on soil moisture using microwaves). This last type of information is certainly the most promising data source in this *System of Observation*.

The observation model (for instance, the radiative transfer scheme) that relates the available observation variables (for instance, remotely-sensed observation variables such as emissivity, temperature, reflectance, etc.) to the variables of interest (biomass and root zone soil moisture for agronomic and water management applications, or latent and sensible heat fluxes for meteorological applications) is often complex and highly non-linear. Therefore, a comprehensive assessment of the impact of measurement errors on the retrieval of the biophysical variables of interest is required. This is usually the task of many inversion methods; but these methods relate any remote-sensing variable to its corresponding biophysical quantity for a given date only. In particular, they give no information on how to interpolate this quantity in time. Combining the dynamic models and the data in a single System of Observation goes a step further by propagating in time all the information (including uncertainty) regarding the system.

The methodology to assimilate remote-sensing observations into a dynamic model presented below enables the quantification of the impact of the introduction of a given remote-sensing variable into the Observing System by taking into account the entire combination of uncertainties in the system, including the measurement error associated with this new variable.

### 3. THE KALMAN FILTER THEORY

The theory behind the Kalman filter has been fully explained in a number of books on control theory [6, 11]: it was initially developed for linear systems (Kalman Filter), extended to the case of moderately non-linear systems (Extended Kalman Filter) and later to strongly non-linear systems with the Ensemble Kalman Filter theory [5].

The Kalman filter is based on the Best Linear Unbiased Estimator whose principle is described below.

Let us suppose that we are interested in evaluating a state variable  $x$  (for example, soil moisture or biomass) which, for the sake of simplicity, is directly observable (i.e. for which direct measurements are available on some specific dates). Thus two values of  $x$  are available: a simulation  $x_{sim}$  and an observation  $x_{obs}$ .

The dynamic model propagates the simulated variable  $x_{sim}$  in time from an initial value  $x_{sim0}$ . Its associated uncertainty (variance)  $\sigma_{sim}$  is also propagated in time up to the next observation. As soon as an observation  $x_{obs}$ , with an uncertainty  $\sigma_{obs}$ , is available, the “forecast” or simulated variable  $x_{sim}^f$  is readjusted instantaneously to account for this new information. The new value of  $x$  is called the *analysed* variable and noted  $x_{sim}^a$ . This analysed state value is a combination of the forecast and observed values, weighted by their respective variances:

$$\frac{x_{sim}^a}{\sigma_{sim}^a} = \frac{x_{sim}^f}{\sigma_{sim}^f} + \frac{x_{obs}}{\sigma_{obs}} \quad (4)$$

with the analysed variance:

$$\frac{1}{\sigma_{sim}^a} = \frac{1}{\sigma_{sim}^f} + \frac{1}{\sigma_{obs}} \quad (5)$$

Equation (1) can be rewritten as

$$x_{sim}^a = x_{sim}^f + K \cdot (x_{obs} - x_{sim}^f) \quad (6)$$

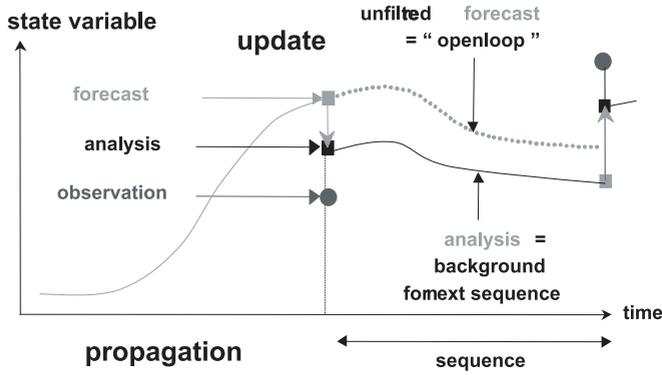
The difference  $(x_{obs} - x_{sim}^f)$  in (6) is called *innovation*. The innovation reflects the discrepancy between the predicted value  $x_{sim}^f$  and the actual measurement  $x_{obs}$ .  $K$  is the *Kalman gain* that minimises the analysed error covariance and is equal to:

$$K = \frac{\sigma_{sim}^f}{\sigma_{obs} + \sigma_{sim}^f} \quad (7)$$

The adjusted state is then taken as the new initial value for the forecast and is propagated in time up to the next available observation. Figure 1 illustrates the filter algorithm.

For moderately non-linear systems, the uncertainty  $\sigma_{sim}$  is propagated analytically using a linear approximation (the Extended Kalman Filter). When the tangent-linear model is difficult to infer manually, an automatic derivation tool can be used, but in most cases the system is strongly non-linear, and the linear approximation no longer holds; a statistical approach known as the Ensemble Kalman Filter is then preferred. In the latter case, an ensemble of possible states is generated and propagated in time, and the statistics are performed on this ensemble. The next paragraph briefly describes the Ensemble Kalman Filter (EnKF) rationale.

To initialise the simulation, an ensemble of  $N$  initial state vectors  $\{x_i(t=0) \mid i = 1, \dots, N\}$  is computed from a normal distribution with a mean equal to the first guess and covariance  $P_{ij}(t=0)$  equal to the estimated uncertainties on the variables



**Figure 1.** Schematic representation of the update of a state variable when an observation is available using the Kalman filter theory (simple case where the state variable is directly observable).

(model). These  $N$  points are propagated in time until an observation is available (errors are implicitly propagated and are estimated by the scatter of the points). Once an observation (or a set of observations) is available, the vector of observations  $z$  is used to generate a set of  $N$  observation vectors  $\{z_i(t), i = 1, \dots, N\}$  from a normal distribution of variance equal to the estimated observation uncertainty  $\sigma_z$ . For each point of the ensemble, the state variables are then readjusted, according to the uncertainties of the observed data and of the simulated variables:

$$x_i^a = x_i^f + K_e(z_i - z_{s,i}) \quad (8)$$

where  $x_i^f$  is the *forecasted* state vector,  $x_i^a$  is the *analysed* state vector after readjustment and  $z_{s,i}$  is the simulated observation variable.

$z_{s,i}$  and  $x_i^f$  are linked through the observation model. For remote-sensing data assimilation, the observation model is the radiative transfer scheme used to link state variables (biomass and surface temperatures, respectively) to remote-sensing observations (NDVI and radiative temperatures, respectively).

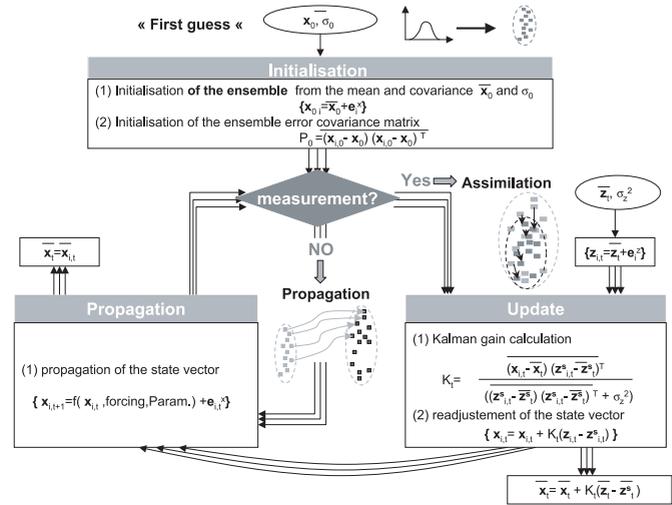
The difference  $(z_i - z_{s,i})$  in (8) is called the *measurement innovation* and reflects the discrepancy between the predicted measurement  $z_{s,i}$  and the actual measurement  $z_i$ . The expression for the *Kalman gain* that minimises the analysed error covariance is now given by [5]:

$$K_e = P/(M + O) \quad (9)$$

where  $P = (x_i^f - \bar{x}^f)(z_{s,i}^f - \bar{z}_s^f)^T$  is the simulated state-observations cross-covariance matrix,  $M = (z_{s,i}^f - \bar{z}_s^f)(z_{s,i}^f - \bar{z}_s^f)^T$  the simulated observation variance-covariance matrix and  $O = \sigma_z$  the measured observation variance-covariance matrix.

The filter algorithm is schematised in Figure 2.

The advantage of this method is that are only performed statistics only at the time of assimilation, i.e. when an observation is available, and the variance-covariance matrix does not need to be propagated in time. No adjoint model is needed, and the dynamic model can be used as it is: the filter manipulates input and output files only. This is particularly attractive for large



**Figure 2.** Schematic representation of the Ensemble Kalman Filter algorithm.

model codes. Since in our system the number of state variables is rather limited (in contrast to meteorology or oceanography applications) the Monte-Carlo method is well suited because less cumbersome.

For more information on Ensemble Kalman filter Theory see [5]. For vocabulary definitions and notations on Data Assimilation see [9].

## 4. THE OBSERVING SYSTEM SIMULATION EXPERIMENT

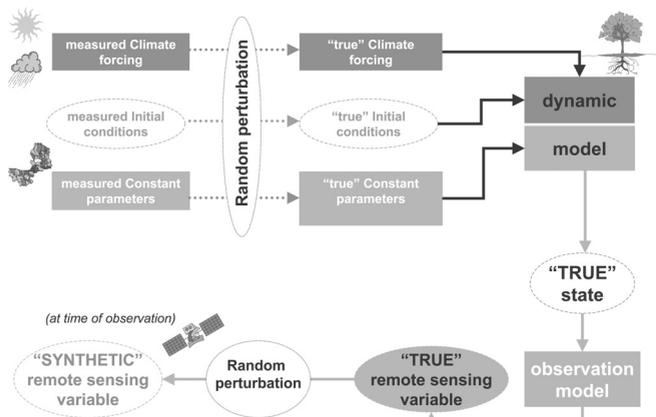
A realistic mathematical representation of the combined effect of most uncertainties into the mathematical representation of the system enables to study how the errors can be efficiently reduced for predictive assessment through data assimilation. For this purpose, the *Observing System Simulation Experiment* (OSSE) approach has been mostly developed in meteorology and oceanography [1, 16]. It was first applied to the observation of the continental surfaces in the context of future passive microwave satellite missions [4, 15]. An application in the context of agronomy is described here.

OSSE aims at testing the potential of some chosen observations to constrain the dynamic model state variables. The idea is to test, before any measurement campaign has been conducted, or any sensor has been launched, the potential benefit of the assimilation of a certain variable into the system when known uncertainties are provided.

The OSSE is divided into two parts.

### 4.1. Generation of a “synthetic observation set”

In a first step, a set of “synthetic observations” for the assimilation period is generated. To do so, the given initial conditions, parameters and forcing variables are used to generate a



**Figure 3.** Schematic representation of the OSSE first step: creation of a synthetic data set.

“true” (also called “nature”) set of  $N$  initial conditions, parameters and forcing variables.

#### 4.1.1. Setting the “true inputs of the model”

Because of errors on measurement devices, heterogeneity or lack of information, the true initial conditions are not the ones that have been measured. The same applies for the parameter values and the forcing data. It is assumed that the true initial state is one of the possible states of a normal distribution with a mean value equal to the observed initial conditions and a variance equal to the associated uncertainty. The set of possible “true” initial conditions, parameters and forcing is generated by perturbing the mean values and choosing random elements in a Gaussian distribution built around the mean initial values and their associated uncertainty.

#### 4.1.2. Computing the “true state variables”

Once the “true” initial conditions, the “true” parameter values and the “true” forcing series have been determined, the dynamic model is run to give a series of “true” state variables.

#### 4.1.3. Computing the “true remote-sensing data”

These true state variables are then used as inputs by an observation model to generate “true” observation variables.

#### 4.1.4. Setting the “measured remote-sensing data”

As for the initial conditions, a stochastic noise term is added to these observation variables to account for the measurement errors and generate what are supposed to be the “measured” (synthetic) observation variables at the time of observation. This step of the synthetic data set creation is schematised in Figure 3.

#### 4.1.5. Computing reference, called the “openloop state variables”

In order to evaluate the improvement (or alteration) of the simulations due to the assimilation of the “measured” remote-sensing variable, a reference simulation was performed, called

an “openloop” (or “control”) simulation. This was done using the unperturbed data set (i.e. using the prescribed mean for initial conditions, parameters and forcing) in order to obtain series of unperturbed state variables; that is, series of simulated state variables that would be obtained if it is assumed that there is no error in the simulation.

## 4.2. Propagation – Testing the observability

In a second step, the synthetic observations are introduced into the Ensemble Kalman Filter assimilation algorithm (Fig. 2). The algorithm uses the same normal distributions as in the first step to generate a set of  $N$  initial conditions,  $N$  sets of parameters and forcing variables time-series. These sets of initial conditions, parameters and atmospheric forcing are then used to simulate  $N$  probable initial realisations (also called particles) of the system. Each of these particles is then propagated in time by the dynamic model up to the first available “synthetic” observation. At that time, each of the particles is re-adjusted using the Kalman Filter (Eq. (8)). The Ensemble is propagated again forward in time until the next observation, and so on. Some unrealistic particles can also be deleted in order to avoid non-realistic spread (for instance: senescence-only behaviours not corrected by the filter). This implies some re-sampling.

The problem of reproductibility of the experiment is posed here with respect to the Monte-Carlo assumptions: is the number of particles used to generate the ensemble sufficient? Indeed, if the Extended Kalman Filter provides one and one result only, several EnKF can lead to different results depending on the number of particles (or replicates) used in the simulation. A preliminary study must thus be performed in order to find the minimum ensemble size ensuring that a few particles have the possibility of following the same trajectory as the true state. It usually corresponds to the minimum ensemble size above whose assimilation results are identical. If this condition is not fulfilled, the solution will rarely converge towards the true state.

## 5. EXAMPLE OF APPLICATION

### 5.1. Description of the dynamic model

The complex Soil-Vegetation-Atmosphere-Transfer Scheme SiSPAT [3] has been coupled to the vegetation functioning model STEP [13] which is designed to describe the evolution of annual herbaceous vegetation in a water limited environment. State variables are the soil matric head and temperature vertical profiles (28 nodes) and the total biomass. Forcing variables are standard climate forcing at an hourly time-step. For convenience, both models run at a minute time-step. SiSPAT propagates the soil and temperature profiles through time. It estimates the stomatal resistance, leaf potential and leaf temperature for each time-step. These values are used by STEP, which estimates in turn the evolution of LAI as the total biomass divided by the Specific Leaf Area value at the end of this time-step. The new LAI value is then used by SiSPAT at the next time-step (there was no evolution of the root density profile through time in this example).

## 5.2. Strategic choices regarding uncertainty specification

The assumption made here is that models are not perfect, but that there is at least one parameter set that reproduces satisfactorily the natural conditions. Therefore no model noise was added. The spread of the ensemble of points arose only from initial conditions and parameter sets. The spread of the ensemble represents the growth or collapse of the simulated variance, i.e. decrease or increase (respectively) in the associated precision or confidence in the model outputs. The diverging (resp. converging) behaviours of the growth (resp. water balance) equations have an important impact on the simulated covariances: for the biomass, chaotic properties will ensure a divergence of the ensemble, whereas for the soil moisture (in the absence of rainfall errors or additional incremental noise) the ensemble will eventually converge towards the equilibrium soil moisture. Thus for biomass, there is always a way to specify the parameter uncertainty so that the simulated variance is large enough to correct the state variables, whereas for the water balance monitoring the simulated variance tends to be too low. If parameters (saturated soil moisture, for instance) are erroneous, the filter won't perform any correction and will simply propagate the error through time.

Since our assumption is that calibration errors exceed all other model errors, parameter uncertainty should sustain the spread of the ensemble through time. As above, if the ensemble does not spread, no assimilation is performed. One can add an incremental noise term to correct insufficient spread, for soil moisture, for instance. This noise sums up all the remaining unexplained uncertainties. Adding model noise will give more flexibility to the assimilation and give observations closer to the true, but the model will generally lose at the same time its integrity (i.e. some of the interdependence between variables). A compromise must be found between fitting model outputs to the observations and keeping a model-consistent signal (i.e. smoothing out sampling problems or measurement errors in the observations). Incremental noise must be treated as a parameter and as such obtained by calibration for a dedicated calibration period.

We must note that the method presented here is not the “classical” OSSE approach. In meteorology and oceanography, the model used to generate the “true” time-series is a complex, exhaustive, physically-based and “state-of-the-art” model, while the model used for data assimilation (and in the future for multiple-scale data assimilation) is a rather simple model. In our case, and again because parameter uncertainty is such an issue for most land-surface process models, we suppose that any “natural” evolution of the surface can be explained by the model using an appropriate set of parameters. This is certainly overoptimistic and the OSSE results should be interpreted with this restriction in mind.

## 5.3. Results

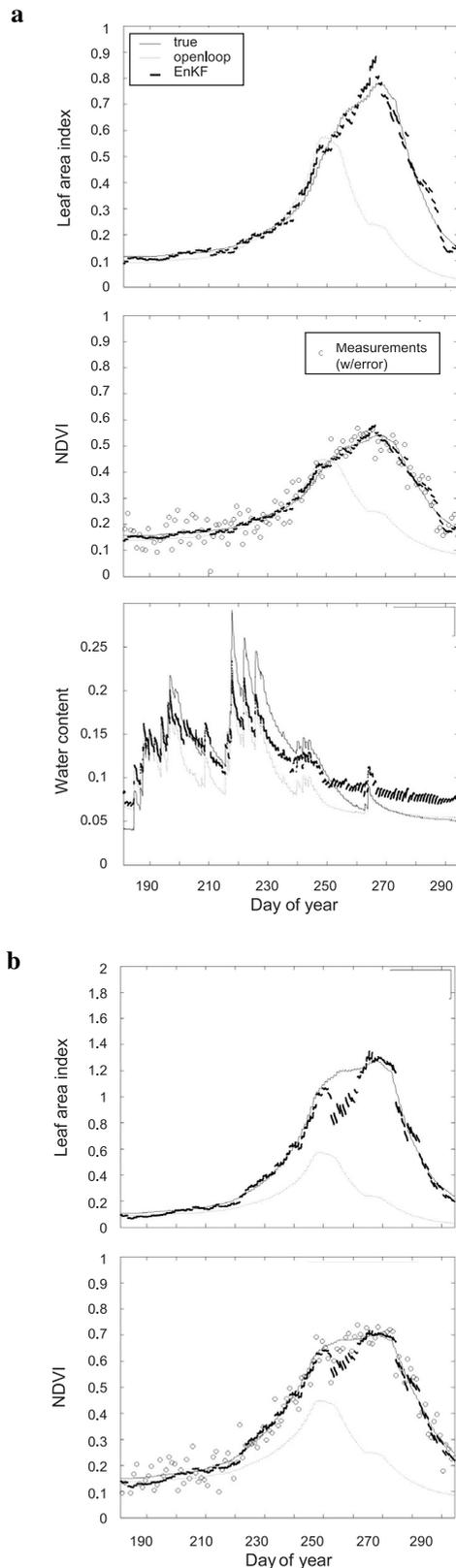
The coupled model was applied during the entire growing season to an annual grassland site in a mixed herbaceous field during the SALSA experiment [7]. In order to use a realistic basis for the OSSE, the extensive data set (biomass evolution, soil moisture, temperature, and heat flux profiles as well as tur-

bulent fluxes and radiative surface temperature time-series) was used to evaluate the sensitivity and then calibrate the value of the main parameters. This realistic simulation obtained using this rough calibration was kept as an “openloop” configuration. For our application, the chosen parameters for uncertainty assessment are the soil hydraulic conductivity and retention curve parameters, the minimum stomatal resistance, the growth efficiency, the wilting point leaf potential and the senescence factor. Very large uncertainty ranges were defined for all the above-mentioned parameters, as well as the initial values of the state variables. This choice was made deliberately according to the OSSE philosophy in order to generate very contrasting “true” behaviours of our system. This large uncertainty was kept in the specification of the parameter noise for the filter, indicating that very little information is accessible for estimating the initial state and the parameter values of our system.

Since both remote-sensing observations of reflectances and radiative temperatures are likely to be the most common remote-sensing variables to be assimilated in the near future, the OSSE is illustrated by the assimilation of midday (12 h 30) nadir radiative surface temperature and NDVI. NDVI is related to LAI with a simple exponential function [8]. Prescribed observation errors (standard deviations) are 1 °K for the radiative temperature, and 0.05 for NDVI. Results are shown in Figure 4a for an average synthetic “true” run, and in Figure 4b for a peculiar case.

In the first example (Fig. 4a), which is similar to most synthetic cases that were generated for this example, both the “true” and “openloop” time-series are very similar, and the amplitude of the true LAI is mostly related to wetter initial conditions. In that case, LAI is well retrieved by the filter, and water content is efficiently corrected only during the first long dry-down (DOY 200-210). At that time, the soil is mostly bare and the filter efficiently assimilates temperature information to correct the soil moisture. However, in the second dry-down after senescence at the end of the simulation period, the soil is again mostly bare but the filter performs very poorly, indicating that the error-covariance is certainly inaccurate, and that an additional incremental noise should be implemented. In the second example, the “true” time-series is related to a combination of parameters that induces a vegetation less sensitive to the dry conditions. Therefore the phase of maximum growth is sustained throughout the dry conditions that prevail at the end of the simulation (cf. the plateau of the “true” LAI time-series, Fig. 4b). In that case, the generally dry conditions force senescence for most of the other parameter sets and NDVI uncertainty is too large to correct the LAI adequately (Fig. 4b). Most of the measurement errors are negative and thus the “measured” NDVI values are compatible with these very dry conditions. Apart from this peculiar case, and for most synthetic “true” cases, LAI is very well retrieved with the daily NDVI, while water content is usually little corrected by the filter. Cases of soil moisture error increase are numerous. The best soil moisture corrections are performed at the beginning of the growing season (low LAI) when temperature differences are very well correlated to soil moisture biases.

Except for cases similar to Figure 4b where drought conditions are consistent with observation noise and lead to an underestimation of LAI, small errors on the vegetation growth model can be efficiently reduced by the Ensemble Kalman Filter.



**Figure 4.** LAI, NDVI and root zone water content time-series for two synthetic cases (4a and 4b); in each plot the “true”, “measured”, “openloop” and “filtered” (“EnKF”) time-series are shown.

## 6. CONCLUSION

Land surface processes are particularly complex due to the strong non-linearity of the processes and the large range of time/space scales involved. Therefore, dynamic models (continuous in time and space) can only give an approximation of reality whereas observations, even if they are almost certain, are sparse and discontinuous in time.

From these considerations, it appears that current monitoring frameworks should focus on two important research topics: (i) an adequate modelling strategy which is able to simulate a given set of remote-sensing variables – this concerns dynamic models and observation models that relate the observation variables to the system state variables – and (ii) an adequate *data assimilation* method to control the dynamic model with the set of remote-sensing information.

In this paper, a method called the *Observing System Simulation Experiment* approach, which assesses the quality of an observing system, was applied to plant growth and water budget modelling. It is designed to test the added value of new remote-sensing products for a given purpose, and to test the efficiency of a particular assimilation framework (including, for instance, the impact of poorly or unknown uncertainties, like the model error). Future improvements of the method include the use of other data assimilation methods such as the Kalman smoother, which could be better suited for assimilation of observations with a short fluctuation period (like temperature) to retrieve variables whose evolution has a long characteristic timescale (such as biomass or soil moisture). Since in land-surface applications parameter estimation is very critical, online parameterisation can also be performed online by considering the parameters as state variables whose time evolution is either constant or show a strong auto-correlation.

**Acknowledgements:** The authors thank Dara Entekhabi, MIT, and Steve Margulis, UCLA, for providing guidance and codes for the Ensemble Kalman Filter and the OSSE implementation.

## REFERENCES

- [1] Arnold C.P. Jr., Dey C.H., Observing-system simulation experiments: Past, present, and future, *Bull. Am. Meteor. Soc.* 67 (1986) 687–695.
- [2] Bloeschl G., Sivapalan M., Scale issues in hydrological modelling: a review, *Hydrol. Process.* 9 (1995) 251–290.
- [3] Braud I., Dantas-Antonino A.C., Vauclin M., Thony J.L., Ruelle P., A Simple Soil-Plant-Atmosphere Transfer model (SiSPAT), development and field verification, *J. Hydrol.* 166 (1995) 231–260.
- [4] Crow W.T., Drusch M., Wood E.F., An observation system simulation experiment for the impact of land surface heterogeneity on AMSR-E soil moisture retrievals, *IEEE Trans. Geosci. Remote Sens.* 39 (2001) 1622–1632.
- [5] Evensen G., Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, *J. Geophys. Res.* 99 (1994) 10143–10162.
- [6] Gelb A., *Applied Optimal Estimation*, MIT Press, 1974.
- [7] Goodrich D.C., Chehbouni A., Goff B., MacNish B., Maddock T., Moran S., Shuttleworth W.J., Williams D.G., Watts C., Hipps L.H., Cooper D.I., Schieldge J., Kerr Y.H., Arias H., Kirkland M., Carlos R., Cayrol P., Kepner W., Jones B., Avissar R., Begue A., Bonnefond

- J.-M., Boulet G., Branat B., Brunel J.P., Chen L.C., Clarke T., Davis M.R., DeBruin H., Dedieu G., Elguero E., Eichinger W.E., Everitt J., Garatuza-Payan J., Gempko V.L., Gupta H., Harlow C., Hartogensis O., Helfert M., Holifield C., Hymer D., Kahle A., Keefer T., Krishnamoorthy S., Lhomme J.-P., Lagouarde J.-P., Lo Seen D., Luquet D., Marsset R., Monteny B., Ni W., Nouvellon Y., Pinker R., Peters C., Pool D., Qi J., Rambal S., Rodriguez J., Santiago F., Sano E., Schaeffer S.M., Schulte M., Scott R., Shao X., Snyder K.A., Sorooshian S., Unkrich C.L., Whitaker M., Yucel I., Preface paper to the Semi-Arid Land-Surface-Atmosphere (SALSA) Program special issue, *Agric. For. Meteorol.* 105 (2000) 3–20.
- [8] Goward S.N., Xue Y., Czajkowski K.P., Evaluating land surface moisture conditions from the remotely sensed temperature/vegetation index measurements. An exploration with the simplified biosphere model, *Remote Sens. Environ.* 79 (2002) 225–242
- [9] Ide K., Courtier P., Ghil M., Lorenc A., Unified notation for data assimilation: Operational, sequential and variational, *J. Meteor. Soc. Jpn* 75 (1997) 181–189.
- [10] Jacquemoud S., Bacour C., Poilvé H., Frangi J.-P., Comparison of four radiative transfer models to simulate plant canopies reflectance; direct and inverse mode, *Remote Sens. Environ.* 74 (2000) 471–481.
- [11] Jaswinski A.H., *Stochastic processes and filtering theory*, Academic Press, San Diego, 1970.
- [12] Kuczera G., Parent E., Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.* 211 (1998) 69–85.
- [13] Mougin E., Lo Seen D., Rambal S., Gaston A., Hiernaux P., A regional sahelian grassland model to be coupled with multispectral satellite data. I: Model description and validation, *Remote Sens. Environ.* 52 (1995) 181–193.
- [14] Raupach M.R., Baldocchi D.D., Bolle H.-J., Dümenil L., Eugster W., Meixner F.X., Olejnik J.A., Pielke R.A. Sr., Tenhunen J.D., Valentini R., How is the atmospheric coupling of land surfaces affected by topography, complexity in landscape patterning, and the vegetation mosaic? in: Tenhunen J.D., Kabat P. (Eds.), *Integrating hydrology, ecosystem dynamics, and biogeochemistry in complex landscapes*, Dahlem Workshop Series, Wiley, Chichester, 1999, pp. 177–196.
- [15] Reichle R., McLaughlin D.B., Entekhabi D., Hydrologic data assimilation with the Ensemble Kalman filter, *Mon. Weather Rev.* 130 (2002) 103–114.
- [16] Rohaly G.D., Krishnamurti T.N., An observing system simulation experiment for the Laser Atmospheric Wind Sounder (LAWS), *J. Appl. Meteor.* 32 (1993) 1453–1471.