Assimilation techniques of remote sensing measurements into vegetation models : overview, limits and promises

C. François, Pascale Cayrol, Laurent Kergoat, Sophie Moulin

ASSIMILATION TECHNIQUES OF REMOTE SENSING MEASUREMENTS INTO VEGETATION MODELS : OVERVIEW, LIMITS AND PROMISES

Christophe FRANÇOIS[1], Pascale CAYROL[2], Laurent KERGOAT[3], Sophie MOULIN[4]

[1]ESE, Ecologie, Systématique et Evolution ,Orsay, France

[2]CESBIO, Centre d'Etude Spatiale de la BIOsphere, Toulouse, France

[3] LET, Laboratoire d'Ecologie Terrestre, Toulouse, France

[4]INRA, Institut National de Recherche Agronomique, Avignon, France

ABSTRACT - A state of the art review of data assimilation techniques is presented, with emphasis and examples related to vegetation modelling. Easy-to-read theoretical papers have been selected in all area of data assimilation, most of them however coming from the oceanography field. Examples illustrate the different techniques. Perspectives regarding errors assessment and their use in assimilation schemes are given.

1. INTRODUCTION

Increasing efforts have been devoted to vegetation modelling during the last decades. Mechanistic models of plant canopies structure and function are now essential for i) Pastures and crops yields monitoring and forecast, ii) Carbon cycle studies and projections iii) Climate system research and, to a lesser extent, weather forecast, iv) hydrology and water ressources assesment and forecast. Common observations clearly show that plant canopies change with time, especially at the seasonal scale and also display a strong interannual variability. More rapid changes can also occur at shorter time scale, following extreme events and disturbances for instance. Common observations also reveal the strong spatial variability of terrestrial vegetation, wether induced by human impact (landscape scale), or reflecting soil, climate, or life-history variability. Vegetation growth models are expected to reproduce these temporal and spatial variabilities, which will impact plant biomass, crop yield, water and carbon exchanges. This is a challenging task, as soon as regional to continental scales are considered. At these scales, ground data related to vegetation properties, like biomass, leaf area index temporal evolution, physiological properties etc., are sparse at best, and most often nearly lacking. Remote sensing is viewed as the only way to scale models from the site-scale up to the region. Different approaches may be used to combine remote sensing and vegetation modelling: model validation, biophysical variables inversion followed by model forcing. There is also a recent trend to assimilate satellite data into models, as it is widely done in atmospheric and ocean science. The idea is to make full use of both model knowledge and data information. As far as ecosystems are concerned, the objectives are either to have the best simulation over a given period of time (*e.g.* estimation of carbon fluxes for the last decade), or to prepare a forecast by assessing the best initial conditions for the model, or the best parameters (*e.g.* short term prevision of crop yield,or floods). Because these two objectives are different, and because there are important differences in models (in terms of accuracy, robustness, errors) as well as in the satellite datasets (in terms of noise, sampling etc.), a whole suite of assimilation methods can be investigated. The objectives of this paper are to summarize or give references to the main assimilation methods suitable for vegetation-related studies, to give examples of what has been used so far, and to draw some perspectives for future works.

2. ASSIMILATION TECHNIQUES OF REMOTE SENSING DATA

In this review, recent papers have generally been preferred to classical and pioneering papers, since the formers are generally simpler and clearer, and also benefit from a larger view, including comparison between methods and links to other inverse problem issues.

Data assimilation refers to a number of techniques aimed at using optimally all available information regarding a system, generally composed of a dynamic model (imperfect), a set of measurements (with noise) and an observation model (imperfect). The objective is to obtain a

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

representation of reality (*i.e.* outputs of the model) taking into account all the features above. Alternative methods are known as 'observation model' inversion and dynamic model calibration. In the case of basic inversion of an observation model, no information from a dynamic model is taken into account, while in the case of basic dynamic model calibration, measurements and dynamic model are taken into account, but the dynamic model is supposed to be perfect in its equations during the assimilation interval and only the parameters and/or initial conditions are updated.

In both cases (observation model inversion and dynamic model calibration), measurements errors may or may not be taken into account, and observation model errors may or may not be taken into account.

We will first review assimilation schemes where no dynamic model is taken into account so that the data are assimilated independantly one from another in time and space. Then we will move towards more complete problems, for which the assimilation scheme takes a dynamic model as an additional constraint : the data must then fit with the dynamic of the model, in a certain sense related to the uncertainty of this model. Some definitions or classifications, relevant to data assimilation with a dynamic model, are given. In Appendix A the Kalman filter formalism is briefly developed since numerous allusions are made in this paper to this method.

2.1 From observation model inversion to advanced observation assimilation

In observation model inversion the information from the dynamic model is not taken into account (*e.g.* inversion of a radiative transfer model using satellite data to derive LAI), or simply no dynamic model is available that would predict the observed quantity (*e.g.* inversion of a radiative transfer model using satellite data to derive mean leaf inclination). In some cases however, other sources of information are available: *e.g.* a most probable value with a given uncertainty as background information, or an output from a model taken as a single information (as opposed to a model taken as a constraint over a whole period of time, see section 2.2). The advantage of using background observation (*i.e.* one advantage of data assimilation over 'observation model' inversion) is that the inverse problem is never ill-posed : the existence, stability and uniqueness of the solution is guaranteed by the background information (see Tarantola and Valette, 1982). The drawback is that the 'inverse' part of the assimilation scheme may be less efficient than what can be done with specific observation model inversions (neural network, simplex method, …) : in the case of the extended Kalman filter for instance (see Appendix A), the observation model is linearized around the background information : if the reality is far from the background information and if the observation model is highly non-linear, then the extended Kalman inversion will not be optimal, and may even by totally erroneous (Evensen, 1997b, Miller *et al.*, 1999).

If the problem is posed in least-square terms as in the variational formulation, and if the problem is solved with a direct minimization of the cost function (no linear tangent model nor its adjoint) then the problem is generally considered optimally solved regarding the non-linearity of the observation model. This is not so true, because the noise in the observation is only taken into account to balance between the different observations and the background information. Let's take an example, the estimation of the LAI with a single NDVI observation:

The least square or variational formulation is in this case :

$$J(\mathrm{LAI}) = \frac{(\mathrm{LAI} - \mathrm{LAI}_0)^2}{\sigma^2_{\mathrm{LAI}}} + \frac{(\mathrm{NDVI}_{\mathrm{obs}} - h(\mathrm{LAI}))^2}{\sigma^2_{\mathrm{NDVI}_{\mathrm{obs}}}} \quad (1)$$

where J is the cost function to be minimized with respect to the LAI and *h* is the non-linear 'observation model' simulating the NDVI from the LAI (other parameters are not considered here for sake of simplicity). LAI_0 is the background information, coming from a vegetation model for instance, and σ_{lai} its associated uncertainty. $\sigma_{\mathrm{NDVIobs}}$ is the noise associated with the observed NDVI. One can also include the error of the observation model *h*, so that Eq. (1) writes :

$$J(\mathrm{LAI}) = \frac{(\mathrm{LAI} - \mathrm{LAI}_0)^2}{\sigma^2_{\mathrm{LAI}}} + \frac{(\mathrm{NDVI}_{\mathrm{obs}} - h(\mathrm{LAI}))^2}{\sigma^2_{\mathrm{NDVI}_{\mathrm{obs}}} + \sigma^2_{\mathrm{h}}} \quad (2)$$

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

Note that here the inversion is performed using the *mean* value of $NDVI_{obs}$, and the *mean* value of $h(LAI)$: $h(LAI)$ is actually a mean value if the uncertainty in h is taken into account. This uncertainty in h arises from uncertainties in its parameters and inexactitudes in its equations, parameterizations and hypothesis that do not necessarily represent the reality. Since h is non-linear, it is well known that the inversion using the mean values may largely differ from the average of an *ensemble* of inversions performed using the noisy $NDVI_{obs}$ and $h(LAI)$ (according to their respective noise $\sigma_{NDVIobs}$ and σ_h). The word *ensemble* is the name given by Evensen (1999a) to a method introduced in 1994 and fully described in Evensen (1997b). Evensen proposed in fact an Ensemble Kalman Filter to solve the problem of predicting a dynamic model uncertainty in the case of a non-linear dynamic model, as will be seen in the next section about dynamic model. This idea has been adapted to overcome the above problems linked to non-linearities in the observation model (see Viovy *et al.*, this issue).

An ensemble scheme with the variational approach given in Eq. (2) would include the minimization of a great number of cost functions J_i given by :

$$J_i(LAI) = \frac{\left(LAI - LAI_0^i\right)^2}{\sigma_{LAI}^2} + \frac{\left(NDVI_{obs}^i - h^i(LAI)\right)^2}{\sigma_{NDVI_{obs}}^2 + \sigma_h^2} \tag{2}$$

with

$$\begin{aligned} LAI_0^i &= LAI_0 + N(0, \sigma_{LAI}) \\ NDVI_{obs}^i &= NDVI_{obs} + N(0, \sigma_{NDVIobs}) \\ h^i(LAI) &= h(LAI) + N(0, \sigma_h) \end{aligned} \tag{3}$$

where $N(0, \sigma)$ is a gaussian noise of mean 0 and standard deviation σ. Any other distribution may be chosen : for instance one may estimate that the background information LAI_0 is only known with a uniform distribution. This is an advantage of ensemble simulations: freely designing its own ensemble to come closer to the actual available information.

Each minimization will lead to an estimated LAI_i. The final estimation will then be the average of these LAI_i, and the corresponding uncertainty will be the standard deviation of the ensemble. Therefore one will not only get the best estimate of the LAI, but also its whole probability distribution function (which may not be a gaussian, so in this case the standard deviation would not be a valuable information). The price to pay for this optimal observation model inversion scheme is to estimate σ_{lai}, $\sigma_{NDVIobs}$ and σ_h. The other price is computation time, an ensemble of size O(100) or O(1000) being reasonable. This is the reason why ensembles are generally used with a Kalman filter approach, which avoid the time needed for the minimization of the cost function (the price here is the linearization of the observation model as mentioned above, see Appendix A). Research has still to be done to find a way of integrating the advances of Ensemble simulations to avoid the linearization in the observation model.

The first point here that allowed to improve the optimality of the inversion scheme was the introduction of the background information (and associated noise). The second point was the introduction of uncertainties and noises in both observations and observation model. The technique that is used (Kalman filter or variational) is of secondary importance. One could as well use any classic inversion scheme. To illustrate this point and show that the ensemble approach is independent of the assimilation technique, we suppose that one has an inverse scheme f to obtain the LAI from a NDVI value : $LAI_{inv}=f(NDVI)$ (through a neural network for instance). Although this inversion scheme could already give an uncertainty by itself, we will not use this property here since we need to get the final uncertainty after the background information has been processed. The ensemble inversion scheme would look like the following :

$$LAI_{inv}^i = \frac{\sigma_{LAI}^2}{\sigma_{LAI}^2 + \sigma_{NDVIobs}^2} f\left(NDVI_{obs}^i\right) + \frac{\sigma_{NDVIobs}^2}{\sigma_{LAI}^2 + \sigma_{NDVIobs}^2} LAI_O^i \tag{4}$$

$$\text{where } LAI_0^i = LAI_0 + N(0, \sigma_{LAI}) \quad \text{and} \quad NDVI_{obs}^i = NDVI_{obs} + N(0, \sigma_{NDVIobs}) \tag{5}$$

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

with the same notations as Eq. (3). The result, as above, would be the probability distribution function of the LAI^i, possibly reduced to its mean and standard deviation values.

The same preoccupation - the addition of a first-guess estimate to constrain and/or improve the inversion - is of interest in many problems : in a study related to atmospheric parameter estimation, Aires *et al.* (2000) went one step further than Eq. (4) : they used a neural network to perform the inversion, but they did include the background information *into* the neural network inversion scheme, and not after the inversion as in Eq. (4). This the first study to propose this very promising approach. Furthermore their paper provides conceptual bridges between the neural network approach (with background information) and the variational approach.

A last point, related to the scale problem, can be addressed by observation model inversion. Another source of noise and uncertainty comes from the intra-pixel variability. A consequence is that the background LAI_0 itself will not have a unique and determined value, but will vary around a mean value, noted LAI_{pix}. This means again that $\overline{h(LAI_{true})}$ will differ from $h(\overline{LAI_{true}})$. The solution to take this variability into account is again to perform an ensemble method with :

$$LAI_0^i = LAI_{pix}^i + N(0, \sigma_{LAI}) \qquad \text{and} \qquad LAI_{pix}^i = LAI_{pix} + N(0, \sigma_{pix}) \tag{6}$$

Note that this is equivalent to write :

$$LAI_0^i = LAI_{pix} + N\left(0, \sqrt{\sigma_{LAI}^2 + \sigma_{pix}^2}\right) \tag{7}$$

See Viovy *et al.* (this issue) for an application of this approach with an Ensemble Kalman filter.

As a conclusion of this section, the ensemble approach provides a theoretical and practical frame to optimally solve the observation model inversion problem. It is easy to implement and very intuitive. It takes into account all non-linearities in the model and can deal with all types of error (gaussian, uniform or others). The drawback is that it needs 1000 inversions (Eq. 8) or 1000 minimization (Eq. 10) for each pixel. An approach using Look-Up tables may sometimes be used, where all calculations are performed once. The Ensemble Kalman Filter needs only 1000 runs of the direct observation model (no inversion, no minimization) but the non-linearities in the observation model are still not totally taken into account, because it relies on local linearization.

2.2 From dynamic model calibration to advanced data assimilation into dynamic model

Classical model calibration aims at adjusting a set of parameters (adequatly chosen through a sensitivity study) to fit observed data to the dynamic model (generally coupled to obervation models). We are therefore one step ahead compared to § 2.1 since an additional constraint is added through the dynamic model (over the whole simulation period, as opposed to a single isolated information as in § 2.1).

A recent review of such applications for SVAT models is presented in Olioso *et al.* (1999), the historical paper being Delecolle *et al.* (1992). Recent studies involving calibration of SVAT and vegetation models with optical and/or radar data are presented in Clevers and Van Leeuwen, 1994 (calibration of 3 parameters of a growth model with LAI estimated from optical and radar data. The calibration includes the uncertainty linked to the inversion of the observation models giving the LAI), Kergoat *et al.*, 1995 (calibration of a growth model with SPOT data and impact on the CO_2 fluxes for a natural vegetation), Cayrol *et al.*, 2000 (calibration of a coupled SVAT and growth model with optical AVHRR data and *in situ* radiometric temperatures), and Prévot *et al.*, 2001 (extensive sensitivity study of a canopy functioning model and calibration of four parameters using optical and radar data).

Uncertainties on observed data and background information about the fitted parameters have rarely been taken into account so far in the cited calibration studies, which makes a difference with advanced assimilation methods (*e.g.* Evensen, 1998 or Natvik *et al.*, 2000). Other differences between calibration and data assimilation will appear more clearly in the following paragraphs where some definitions and optimal data assimilation formulation (with sometimes sub-optimal techniques of resolution) are presented.

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

2.2.1 strong constraint/weak constraint

Let us assume that LAI series are directly observed (with uncertainty σ_{LAI}) and that we want to use it to update or correct a vegetation growth model. The growth model serves as a constraint. If the growth model is supposed to be perfect, and only its parameters and initial conditions are prone to errors, then we are in the *strong constraint* case. If we take into account imprecision and errors in the model, we are in the *weak constraint* case. These are the first two categories of data assimilation. Model calibration for instance relies most of time on the strong constraint hypothesis.

In the strong constraint case the model is supposed perfect during the assimilation window, which therefore must be as short as possible. Otherwise, we are trying to fit parameters with a wrong model, which would not result very efficient. The weak constraint case is more general and therefore better (more optimal) than the strong constraint case (see Evensen *et al.*, 1998).

The variational or least square formulation fits well to the strong constraint problem. Such an approach is presented in Knorr (1999) for an application with a vegetation model (see also Knorr, 1997 and Knorr, 1998). The cost function to be minimized in his case is :

$$J(x_1, x_2, x_3) = \sum_{m=1}^{12} \frac{f_{obs} - g_{sim}(x_1, x_2, x_3)}{\sigma_f^2 + \sigma_g^2} + \sum_{i=1}^{3} \frac{\left(x_i^0 - x_i\right)^2}{\sigma_{x_i}^2} \qquad (8)$$

where x_1,x_2 and x_3 are the model parameters to be determined, f_{obs} are the observed fAPAR (with associated uncertainty σ_f), g_{sim} (and associated uncertainty σ_g) the simulated fAPAR (that depends on model parameters x_1,x_2,x_3), and x_i^0 the background values of parameter x_i with uncertainty σ_{x_i} .

In Knorr (1999), x_1 is the maximum plant-available soil water, x_2 the leaf onset/shedding temperature and x_3 the fractional vegetation cover. Knorr tested his method on both local and global scales. He shows that the assimilation procedure is superior to both the control run (no assimilation) and hard updating (fAPAR is simply replaced in the model by its remote sensing estimate). Such a finding was also demonstrated in François *et al.* (1999) and Quesney (1999) in the context of Extended Kalman filtering and a hydric balance model.

The problem in Knorr application, apart from the strong constraint hypothesis, is the determination of σ_g (which may actually vary with time). The other problem is the choice of the parameters to be fitted. Knorr (1997) addresses these problems, together with the difficult problem of remote sensing data quality and recalibration before comparison with simulations.

In fact, the cost function J in Eq. (8) should include all parameters and their covariance (and not only their variance) (see Evensen 1998). The problem becomes very complicated because it is more difficult to estimate 100 variances than 3 (if the model has 100 parameters), and still more difficult to estimate the associated 50000 covariances. At this point, a sensitivity study may reduce the number of parameters of interest to 10, 5 or 3, but the variance and covariance calculation remains a problem. This problem has been solved for linear models and gaussian errors with the Kalman filter, approximately solved for non-linear models with the extended Kalman filter, and totally solved for non-linear models with non-gaussian errors with the Ensemble Kalman filter (Evensen, 1997b). In this sense the problem is easier than the inversion of non-linear observation models that is still pending (as seen previously).

Natvik *et al.* (2000) present an application of the weak constraint problem for a zero dimensional marine ecosystem model : the problem is very similar to vegetation models, which are also zero dimensional models. The method presented in this paper may therefore be transposed to vegetation models. The variational formulation is minimized using gradient descent methods (gradient steepest descent and nonlinear conjugate gradient) and these methods are fully presented and described.

2.2.2 sequential/multiple/smoother

In the *sequential* assimilation, the model is updated whenever an observation is available. If the sequential assimilation technique allows to update prognostic variables of the model, we are in the

weak constraint case : Kalman filter, Ensemble Kalman filter, weak constraint sequential variational assimilation (3D-Var). If the result of the assimilation must be an output of the model we are in the strong constraint case (only parameters are updated) : sequential parameter identification with the Kalman filter, strong constraint sequential variational assimilation. A – still simple – example of sequential data assimilation in vegetation model is presented in Cayrol (2000) : this study shows the advantage of allowing the model to deviate from its prediction. It also shows that a prior calibration is necessary before the sequential assimilation to correctly perform. The complete solution to this problem would be the weak constraint problem formulation as presented in Evensen *et al.*, 1998, (see below), allowing the determination of both the parameters and the dynamic variable LAI with a weak constraint.

In the *multiple* assimilation, all observations are processed together. This terminology (*multiple assimilation*) is not official since, for historical reasons, variational assimilation was sometimes seen as the opposite of sequential assimilation (see for instance the title of the otherwise excellent paper by Ide *et al.*, 1997), and therefore variational assimilation was employed in the sense of multiple assimilation. 3D-Var, however, is a sequential variational assimilation technique for instance (Zou *et al.*, 1997). Symmetrically, as both computer capacities and theoretical studies advances, optimal multiple assimilation schemes based on other techniques than the variational one are appearing (*e.g.* Miller *et al.*, 1999). Since assimilation history in the continental biosphere community is nearly virgin, we may use new words, so that *multiple assimilation* means the opposite of *sequential assimilation*.

Since all observations are considered as a whole (*i.e.* all information are considered), one may think that multiple assimilation is preferable to sequential assimilation. On a theoretical point of view this is true. For linear models with gaussian errors however, it is also proved that optimal sequential assimilation (*i.e.* Kalman filter) gives the same result than 4D-VAR assimilation at the end of the assimilation window (*e.g.* Li and Navon, 1999), so that the advantage of multiple assimilation is not so evident. Basically the advantage of sequential assimilation is that model errors are easily processed and accurately estimated (through Extended Kalman filter or, better, through Ensemble techniques, see Evensen, 1997a,b or Miller *et al.*, 1998 for wilder cases), while in multiple variational assimilation, nothing exists that allows to estimate the variances and covariances of the errors of the models (except adaptation of Kalman filter-like schemes).

A smoother is simply the combination of two filters, one integrated forward in time, and the other processing the observations backward in time, so that the trajectory is optimal not only at the end of the observation period, but also all along it (see Bennet, 1992, Tanizaki, 1996). Evensen and Van Leeuwen (2000) present a general ensemble smoother for nonlinear dynamics formulated as a sequential method (the observations can be assimilated sequentially during a forward integration). They also present a general and easy to read presentation of what is a smoother (compared to a filter).

In a terminological way, Natvik *et al.* (2000) distinguish between sequential methods and smoothing methods in the way we distinguished between sequential and multiple methods. Therefore, smoothing assimilation and multiple assimilation may be seen as synonymous, with a slight connotation of weak constraint problems for the term smoothing assimilation.

We won't go any further in the presentation of the different solutions of the strong and weak constraint problems with sequential, multiple or smoother methods, involving variational, Kalman, Ensemble or Bayes techniques, since numerous quality papers already present them (and most of them were already cited in the previous paragraphs). In a very complete and easy to read paper, Evensen *et al.* (1998) present the generalized parameter estimation problem, for both strong and weak constraint cases, including the determination of the model bias (together with the parameters estimation); The problem is formulated in terms of the Euler-Lagrange equations. An iterative solution is proposed for weakly non-linear models using the gradient descent method. In the strong constraint case, they obtain the well known "adjoint method". As an alternative the representer method is presented, which allows to solve the problem without iterations. Finally a sequential

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

scheme is presented to correct the bias of the model "on the flight". This last scheme however is only useful is one already has a complex assimilation scheme and wants an additional feature of bias estimation. If one can develop its own assimilation scheme, the bias estimation should be included in the principal scheme, either in its variational (Euler-Lagrange) or sequential form. More details about these methods can be found in Bennet (1992) for the Euler-Lagrange and representer methods, and Tanizaki (1996) for the non-linear filters and smoothers. The paper by Evensen (1998) is nicely completed by Evensen (1994a), which includes simulation-annealing methods.

An overview of different assimilation techniques, together with a unified notation is also presented in Ide *et al.* (1997). Van Leeuwen and Evensen (1996) in a fundamental paper present the data assimilation and inverse methods in terms of a probabilistic formulation that gives more insight into data assimilation basis (Bayesian statistics, Maximum-Likelihood estimator vs. Minimum-Variance estimator, …). The Maximum-Likelihood vs. Minimum-Variance problem is also clearly posed in Evensen (1997a).

A clear and nice presentation of the adjoint method to solve the strong constraint variational formulation (3D-VAR and 4D-VAR), together with links to the Kalman filter and basics in Bayes statistics is presented in the hard to find but inestimable technical note by Zou *et al.* (1997). In the same domain, the paper by Navon (1996) is helpful and includes a state of the art review of parameter estimation in oceanography (but basics are basics and also apply to vegetation models).

If we investigate the hydrological field, we find that numerous studies are related to model calibration : one very interesting recent paper by Xiong and O'Connor (2000) introduces techniques to adequately represent a n-dimension objective function (which is the Nash-Sutcliffe criterion in their case, but could be a variational cost function J) in a 2-D space.

2.3 Estimation of model error

A very important point in data assimilation method is related to the technique employed to estimate the model error : all the methods presented above include the covariance matrix of the model error, either in the observation model inversion or in the dynamical model calibration. This issue is addressed in the above-mentioned papers, but no practical method is given.

Concerning vegetation models, we should be able to refine the estimation of model error through ensemble technique : each line of model code should be associated with a process error (either uniform or gaussian), assuming that all parameters and quantities (*e.g.* evaporation) in this line are known and exact. The scheme begins with the estimation of parameter uncertainties, and goes on until the prognostic equations. Such an idea is an extension of previous works by Spear and Hornberger (1980), cited and applied in Franks *et al.* (1999). Only the parameters are affected with errors in these works however (not the equaations), and uniform errors are always retained (instead of gaussian, uniform or others).

To illustrate the method, we can take the example of hydric balance. Suppose that we are at the final hydric balance equation of the surface layer (involving precipitation, evaporation, transpiration and exchanges with a sub-layer). We should evaluate the error of this equation, assuming that precipitation, evaporation, transpiration and exchanges are correct (their respective errors have been taken into account in the preceding code lines). The error to be evaluated may include for instance lateral exchanges if they have been neglected (1D model) and all physical processes that have not been taken into account in the equation. The final step is to run the model with all errors to create an ensemble of simulations (say 1000 simulations). For each simulations, at each line code, an error will be added according to its statistical properties. Finally, one obtains the whole probability distribution function (pdf) of the model state, which represents all the available information. One can see whether the obtained pdf is gaussian or not : most of the time, it will not be, because of uniform noises that have been chosen for poorly modelized processes and poorly known parameters, and because of the non-linearities in the model.

This method allows to obtain the exact information available from the model, together with the exact variance-covariance matrix of model errors (if relevant, *i.e.* for gaussian pdf's), provided that the errors were correctly estimated for each code line. A similar concern about error estimation in

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

SVATs is expressed by Franks and Beven (1997), and they also propose Monte-Carlo based simulations.

In most papers, this method is replaced with a less practical one : one is supposed to find out the variance and covariance of the prognostic equations (q in Eq. (12) in Evensen (1998), or **Q** in the Kalman filter, see for instance Eq. (3) in Ide *et al.*, 1997). Less code lines are concerned than in the above-proposed method (only the prognostic equations), but this is absolutely impossible to correctly guess what could be the sequential model error on one time step. Furthermore, most of the time the matrix **Q** is supposed to be diagonal (no error covariance) because this is still more impossible to guess error covariances than error variances : the consequence of the hypothesis of null covariance in **Q** is to consistently underestimate the final covariance errors in the variance-covariance matrix **P** of the model (computed with Eq. (4b) in Ide *et al.* 1997, Eq. 44 in Evensen (1994b) with the Kalman filter). The same problem occurs in the ensemble methods (Evensen, 1997b) : to compute the samples, one has to estimate the q's or **Q**. In the variational assimilation the corresponding problem is the estimation of background covariance $\mathbf{B_0}$ (Eq. 8 in Li and Navon, 1999, Eq. 11 in Ide, 1997, Eq. 2.45 (3D-VAR) or 2.60 (4D-VAR) in Zou *et al.*, 1997).

3. CONCLUSION

To conclude, we shall quote the last paragraph in Miller *et al.* (1999), about nonlinear filtering, which in our case refers to both the ensemble method and the associated method that we suggest in §2.3 to estimate the variance-covariance matrix of the model (in gaussian cases), or more generally the pdf's (by "nonlinear filtering" the Miller *et al.* refer to the use of pdf's) :

"It is not our purpose to advocate nonlinear filtering as a competitor to other data assimilation methods for operational use. At this time, it is far too resource intensive for that, but we propose it as a conceptual tool, to be used to gain insight into the performance of approximate schemes in highly nonlinear settings. The reader should be reminded that this was the status of the Kalman filter a decade ago. While this remains so to a large extent today, the most cursory glance at the literature shows application of the Kalman filter and other weak constraint schemes to problems of complexity far beyond anything envisioned by the early investigators of those methods."

In the continental biosphere domain, and more specifically in the vegetation model domain, our models are simple (0D) and do not include a huge number of variables, so that we are very favored with respect to oceanographers and meteorologists. In our case, advanced data assimilation are not far too resource intensive, and rather we have the opportunity of using the most advanced techniques for our data assimilation schemes, and also gain invaluable insight into our models by studying and predicting their errors evolution. The citation by Miller is well completed by the citation from Franks *et al.* (1999) about Monte-Carlo simulations (which applies to the error prediction scheme as discussed in §2.3, and also to assimilation with ensemble methods) :

"While [the method] contains a number of subjective elements (for example, the prior choice of parameter ranges) it forces those choices to be made explicit. A large number of computer runs is also required, particularly for models with a large number of parameters. (…) In practice, sampling efficiency has not proven to be a particular constraint, especially since Monte-Carlo simulation is well suited to parallel computer systems. Against the disadvantage of computer run times is the considerable advantage that the approach is conceptually very easy to understand and easy to implement".

APPENDIX A : THE KALMAN FILTER SCHEME

The Kalman filter (see for instance Grewal and Andrew, 1993, Zou *et al.* 1997, and Evensen papers) is a weak constraint sequential assimilation method (to be compared to the 4D-VAR which is a strong constraint multiple assimilation method; both are based on minimum variance estimation for gaussian errors, see Zou *et al.* 1997 or Li and Navon, 1999). In this appendix the general

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

formulation is given, first illustrated by the simple LAI-NDVI example (Eq. 2) completed with the inclusion of a dynamic model constraint to form the complete Kalman filter.

The equivalent of Eq. (2) for the Kalman filter would be :

$$LAI_{new} = LAI_{sim} + K_{lai} * (NDVI_{obs} - NDVI_{sim}) \quad (A1)$$

with

$$K_{LAI} = \frac{\frac{\partial h}{\partial LAI}\,\sigma^2_{LAI}}{\sigma^2_{NDVIsim} + \sigma^2_{NDVIobs}} \quad (A2)$$

and

$$\sigma^2_{NDVIsim} = \left(\frac{\partial h}{\partial LAI}\,\sigma_{LAI}\right)^2 + \left(\frac{\partial h}{\partial x_1}\,\sigma_{x_1}\right)^2 + \left(\frac{\partial h}{\partial x_2}\,\sigma_{x_2}\right)^2 \quad (A3)$$

where LAI, x_1 and x_2 are the dynamic model variables. All these variables will be corrected through an analogous of Eq. (A2) for K_{x1} and K_{x2}.

If the vegetation growth model is taken as a (weak) constraint, then the background information LAI_{sim} is an output taken from the model, and its uncertainty is computed through the following equation derived from the Fokker-Planck (or forward Kolmogorov) equation (Miller *et al.*, 1999) :

$$\mathbf{P}_{k+1} = \mathbf{M}_k \mathbf{P}_k \mathbf{M}_k^T + \mathbf{Q}_k \quad (A4)$$

where $\mathbf{P}_{k+1}$ is the model error covariance matrix at time k+1, calculated from the covariance matrix $\mathbf{P}_k$ at time k, the Jacobian matrix $\mathbf{M}_k$ of the dynamic model *m*, and the sequential error matrix $\mathbf{Q}_k$.

The equation to compute the Kalman gain K whenever observations are available is :

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T \left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)^{-1} \quad (A5)$$

where **H** is the Jacobian of the observation model *h*, and **R** is the variance of the observations.

The general form of the updating equation (A1) in the Extended Kalman Filter then writes :

$$\mathbf{X}^+ = \mathbf{X}^- + \mathbf{K}\left[z - h\left(\mathbf{X}^-\right)\right] \quad (A6)$$

where $\mathbf{X}^+$ refers to *a priori* estimates, $\mathbf{X}^-$ refers to *a posteriori* (updated) estimates, z the observations and *h* the observation model. The term between brackets in Eq. (A6) represents the innovation vector. **X** may contain all prognostic variables and/or parameters of the dynamic model.

One additional equation allows to update P after the assimilation of the observation in (A6), before going back to the error propagation equation (A4) :

$$\mathbf{P}^+ = \left(\mathbf{1} - \mathbf{K}\mathbf{H}\right)\mathbf{P}^- \quad (A7)$$

The term *extended* in Extended Kalman Filter means that partial derivatives are taken to form the model matrix **M** and the observation model **H** (see Eqs. A2 and A3 for an example with *h*). These local linearizations are valid only for weakly non-linear models, as discussed above. The Ensemble method of Evensen avoids the linearization **M** of *m* (Eq. A4), but not the linearization **H** of *h*.

REFERENCES

[Aire 2000] Aires, F., Prigent, C., Rossow, W. B., and M. Rothstein, 2000, A new neural network approach including first-guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature and emissivities over land from satellite microwave observation, to be published in *.l Geophys. Res.*

[Benn 1992] Bennet, A., 1992, Inverse methods in physical oceanography, *Cambridge University Press*, 342 pp.

[Cayr 2000] Cayrol, P., 2000, Assimilation de données satellitales dans un modèle de croissance et de bilan d'énergie. Application à des zones semi-arides, Phd Thesis, University of Toulouse, 210 pp. (in French)

[Cayr 2000] Cayrol, P., Kergoat, L., Moulin, S., Dedieu, G., and A. Chehbouni, 2000, Calibrating a coupled SVAT/Vegetation growth model with remotely sensed reflectance and surface temperature. A case study for the HAPEX-Sahel grassland sites. *J. Appl. Meteor.*, in press

[Clev 1994] Clevers, J.P.G.W., and Van Leeuwen, 1994, Combined use of optical and microwave remote sensing data for crop growth monitoring, *Remote Sens. Environ.*, 56:42-51.

[Dele 1992] Delecolle, R., Maas, S.J., Guerif, M., and F. Baret, 1992, Remote Sensing and crop production models: present trends, *ISPRS J. Photogramm. Remote Sens.*, 47:145-161.

[Even 1994a] Evensen, G., 1994, Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99(C5),10143-10162.

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.

[Even 1994b] Evensen, G., 1994, Inverse Methods and Data Assimilation in Nonlinear Ocean Models, *Physica D*, 77:108-129, http://fram.nrsc.no/~geir/

[Even 1997a] Evensen, G., 1997, Advanced data assimilation for strongly nonlinear dynamics, *Mon. Weather Rev.*, 125:1342-1354, http://fram.nrsc.no/~geir/

[Even 1997b] Evensen, G., 1997, Application of ensemble integrations for predictability studies and data assimilation, in *Monte Carlo Simulations in Oceanography, Proceedings 'Aha Huliko'a Hawaiian Winter Workshop Manoa, January 14-17, 1997*, eds. Muller and Henderson, pp. 11-22, http://fram.nrsc.no/~geir/

[Even 1998] Evensen, G., 1998, Parameter estimation in dynamical models, NATO ASI, *Ocean Modeling and Parameterizations* edited by E. P. Chassignet and J. Verron, http://fram.nrsc.no/~geir/

[Even 2000] Evensen, G., and Van Leeuwen, 2000, An Ensemble Kalman Smoother for Nonlinear Dynamics, *Mon. Weather Rev.*, 128, pp. 1852-1867, http://fram.nrsc.no/~geir/

[Fran 1997] Franks, S.W. and K.J. Beven, 1997, Bayesian estimation of uncertainty in land surface – atmosphere flux predictions, *J. Geophys. Res.*, 102(D20):23991-23999.

[Fran 1999] Franks, S.W., Beven, K.J., and J.H.C. Gash, 1999, Multi-objective conditioning of a simple SVAT model, *Hydrology and Earth System Sciences*, 3(4):477-489.

[Fran 1999] Francois, C., Quesney, A., and C. Ottlé, 1999, Assimilation séquentielle d'observations SAR/ERS dans un modèle de bilan hydrique couplé à un modèle hydrologique : application d'un filtre de Kalman étendu, in *Actes des Ateliers de Modélisation de l'Atmosphère*, CNRM-Météo-France, Toulouse, France, 30 nov – 2 déc 1999, pp. 195-198 (in French)

[Grew 1993] Grewal, M., and A. Andrew, 1993, Kalman filtering : theory and practice, Ed. Prentice Hall, 375 pp.

[Ide 1997] Ide, K., Courtier, P., Ghil, M. and A.C. Lorenc, 1997, Unified notation for data assimilation: operational, sequential and variational, *J. Meteor. Soc. of Japan*, 75:181-189, http://uniblab.atmos.ucla.edu/tcd/MG

[Kerg 1995] Kergoat, L., Fischer, A., Moulin, S., and G. Dedieu, 1995, Satellite measurements as a constraint on estimates of vegetation carbon budget, Tellus, 47B,251-263.

[Knor 1997] Knorr, W., 1997, Satellite Remote Sensing and Modelling of the Global CO2 Exchange of Land Vegetation: A Synthesis Study, *PhD thesis*, Max-Planck-Institut für Meteorologie, Hamburg, Germany, ISSN 0938-5177, http://www.mpimet.mpg.de/~knorr.wolfgang/

[Knor 1998] Knorr W., 1998, Constraining a global mechanistic vegetation model with satellite data, in *"IEEE International Geosciency and Remote Sensing Symposium Proceedings"*, Seattle, WA, 6-10 July 1998.

[Knor 1999] Knorr W., 1999, Assimilation of Satellite Data into a Model of Vegetation and Land-Surface Processes, in *"ALPS99 Conference and Workshop"*, Meribel, France, 18-22 January 1999.

[Leeu 1996] van Leeuwen, P. J. and G. Evensen, 1996, Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation, *Mon. Weather Rev.*, 124:2898-2913, http://fram.nrsc.no/~geir/

[Li 1999] Li, Z.. and M. Navon, 1999, Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother, Submitted to *Q. J. Roy. Meteorol. Soc.*, http://www.scri.fsu.edu/~navon/

[Mill 1999] Miller, R., Carter, E., and S. Blue, 1999, Data assimilation into nonlinear stochastic models, *Tellus*, V51A, pp. 167-194, http://tangaroa.oce.orst.edu/cmg3b/cmg3b.ps

[Natv 2000] Natvik, L. J., Eknes, M. and G. Evensen, 2000, A weak constraint inverse for a zero dimensional marine ecosystem model, to be published in J. Marine. Systems., http://fram.nrsc.no/~geir/

[Navo 1996] Navon, I. M., 1996, Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, in *Computational Mechanics, Third AsianPacific Conference, Sept 1996, Seoul, Korea*, Choi, Yun and Lee Eds., TechnoPress, 17771785, www.scri.fsu.edu/~navon

[Olio 1999] Olioso, A., Chauki, H., Courault, D. and J.P. Wigneron, 1999, Estimation of evapotranspiration and photosynthesis by assimilation of remote sensing data into SVAT models, *Rem. Sens. Env.*.,68:341-356.

[Prev 2001] Prévot, L., Chauki, H., Troufleau, D., Weiss, M., Baret, F., and N. Brisson, 2001, Assimilating optical and radar data into the STICS model for wheat crops : preliminary results, to be published in to *Physics and Chemistry of the Earth*

[Ques 1999] Quesney, A., 1999, Assimilation de mesures d'humidité de surface dans un modèle hydrologique conceptuel global. Apport de la télédétection radar ERS/SAR, *Phd Thesis*, Université Paris 7, 180 pp.

[Spea 1980] Spear, R. C., and G. M., Hornberger, 1980, Eutrophication in Peel Inlet II. Identification of critical uncertainties via generalised sensitivity analysis, *Wat. Res.*, 14:43-49.

[Tani 1996] Tanizaki, H., 1996, Nonlinear filters : estimation and applications, *Springer-Verlag Eds.*, 252 pp.

[Tara 1982] Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion, Reviews of geophysics and space physics, 20(2):219-232.

[Xion 2000] Xiong, L. and K. M. O'Connor, 2000, Analysis of the response surface of the objective function by the optimum paramter curve: how good can the optimum parameter values be ?, *J. Hydro.*, 234:187-207.

[Zou 1997] Zou, X,., Vandenberghe, F., Pondeca, M. and Y.-H. Kuo, 1997, Introduction to adjoint techniques and the MM5 adjoint modeling system, *NCAR technical note*, NCAR/TN-435-STR, 110 pp.

2001, 8th international congress on Physical measurements and signatures in remote sensing, 8-12 janvier 2001, Aussois, France, CNES Ed., pp. 645-658.