



**HAL**  
open science

## From near-surface to root-zone soil moisture using different assimilation techniques

Joaquin Munoz-Sabater, Lionel Jarlan, Jean-Christophe Calvet, F. Bouyssel,  
Patricia de Rosnay

► **To cite this version:**

Joaquin Munoz-Sabater, Lionel Jarlan, Jean-Christophe Calvet, F. Bouyssel, Patricia de Rosnay.  
From near-surface to root-zone soil moisture using different assimilation techniques. *Journal of Hydrometeorology*, 2007, 8 (2), pp.194-206. 10.1175/JHM571.1 . hal-00293340

**HAL Id: hal-00293340**

**<https://hal.science/hal-00293340>**

Submitted on 10 Jun 2021

**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.

## From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques

JOAQUÍN MUÑOZ SABATER, LIONEL JARLAN, JEAN-CHRISTOPHE CALVET, AND FRANÇOIS BOUYSSSEL

*GAME/CNRM, Météo-France, CNRS, Toulouse, France*

PATRICIA DE ROSNAY

*Centre d'Etudes Spatiales de la Biosphère, Toulouse, France*

(Manuscript received 11 January 2006, in final form 24 August 2006)

### ABSTRACT

Root-zone soil moisture constitutes an important variable for hydrological and weather forecast models. Microwave radiometers like the L-band instrument on board the European Space Agency's (ESA) future Soil Moisture and Ocean Salinity (SMOS) mission are being designed to provide estimates of near-surface soil moisture (0–5 cm). This quantity is physically related to root-zone soil moisture through diffusion processes, and both surface and root-zone soil layers are commonly simulated by land surface models (LSMs). Observed time series of surface soil moisture may be used to analyze the root-zone soil moisture using data assimilation systems. In this paper, various assimilation techniques derived from Kalman filters (KFs) and variational methods (VAR) are implemented and tested. The objective is to correct the modeled root-zone soil moisture deficiencies of the newest version of the Interaction between Soil, Biosphere, and Atmosphere scheme (ISBA) LSM, using the observations of the surface soil moisture of the Surface Monitoring of the Soil Reservoir Experiment (SMOSREX) over a 4-yr period (2001–04). This time period includes contrasting climatic conditions. Among the different algorithms, the ensemble Kalman filter (EnKF) and a simplified one-dimensional variational data assimilation (1DVAR) show the best performances. The lower computational cost of the 1DVAR is an advantage for operational root-zone soil moisture analysis based on remotely sensed surface soil moisture observations at a global scale.

### 1. Introduction

The accuracy of short-term to seasonal weather predictions depends on a good initialization of several surface variables of slow variation in the coupled land surface–atmosphere system. Among these variables, root-zone soil moisture is of prime importance.

Root-zone soil moisture plays a vital role in the regulation of water and energy budgets at the soil–vegetation–atmosphere interface through evaporation processes of the uppermost surface soil layer and plant transpiration (Shukla and Mintz 1982). If the initialization of this variable is not accurate, significant drifts of the temporal evolution of the surface state variables

may develop and may consequently cause a degradation of the weather forecast (Beljaars et al. 1996; Dirmeier 2000; Koster and Suarez 2003).

Land surface models (LSMs) aim to describe the continental lower boundary conditions for numerical weather prediction (NWP) models (i.e., water and energy exchanges). They are now able to simulate the main processes of the surface functioning (e.g., soil water dynamics, vegetation–hydrology interaction, water and energy fluxes) but are still limited by several constraints: 1) the need for a high amount of input data (soil and vegetation characteristics) that cannot be provided accurately at large scales, 2) the incompatibility between the relatively low spatial scale (~0.1–1 km) of surface and hydrological processes (in particular runoff, subsurface flow), and the grid scale of NWP models (~10–100 km), and 3) the meteorological forcing errors, especially for rainfall, which has the most significant influence on soil moisture variability. These constraints have an effect on the simulation of soil moisture

---

*Corresponding author address:* Jean-Christophe Calvet, Météo-France/CNRM/GMME/MC2, 42, Avenue Gaspard Coriolis, Toulouse CEDEX 1, 31057, France.  
E-mail: jean-christophe.calvet@meteo.fr

evolution within the LSM and may adversely affect the quality of the weather predictions.

A possible solution to improve simulations of LSM is to assimilate observations sensitive to soil moisture by using data assimilation schemes. Operational optimal interpolation systems for NWP models have been developed (Giard and Bazile 2000; Douville et al. 2000) with the aim to analyze soil moisture by incorporating air temperature and humidity observations at a height of 2 m over the soil surface. Within the framework of the European Land Data Assimilation System (ELDAS) project, Balsamo et al. (2004) have tested a simplified variational system. Nevertheless, air temperature and humidity are quite indirectly linked to soil moisture. A more direct source of information is provided by L-band microwave remote sensing, which links the observed brightness temperature ( $T_B$ ) to the surface soil moisture (top 0–5 cm). These observations show a large sensitivity to soil moisture variations (Eagleman and Lin 1976; Wigneron et al. 2002), and they could be included in NWP systems by assimilating  $T_B$  directly, which requires a radiative transfer model, or by assimilating derived soil moisture products. The potential of the analysis of root-zone soil moisture ( $w_2$ ) from surface soil moisture observations ( $w_g$ ) was highlighted by Calvet et al. (1998a) and Calvet and Noilhan (2000). Several authors have already conducted the analysis of  $w_2$  at a local scale using observations of microwave brightness temperatures (Houser et al. 1998), synthetic observations (Reichle et al. 2002), or soil surface moisture retrievals from the synthetic aperture radar on board the European Remote Sensing (ERS) satellites (François et al. 2003). Global-scale  $w_g$  products also exist like those provided by the Advanced Microwave Scanning Radiometer for the Earth Observing System (EOS) (AMSR-E) sensor (Njoku et al. 2003), on board the National Aeronautics and Space Administration's (NASA) *Aqua* satellite and those derived from the *ERS-1* and *ERS-2* scatterometers (Wagner et al. 2003). The future Soil Moisture and Ocean Salinity (SMOS) satellite of the European Space Agency (ESA), planned to be launched in 2008 (Kerr et al. 2001), will provide an estimation of the soil moisture in L band at a global scale, with a sampling time of around 3 days at the equator and with a spatial resolution compatible with NWP models. If we want to take advantage of this vast amount of available data, assimilation systems have to be developed and integrated within the NWP models (Seuffert et al. 2003; Balsamo et al. 2006). Within this context, this study aims to analyze  $w_2$  by assimilating  $w_g$  measurements. The dataset used in the present study comes from the Surface Monitoring of the Soil Reservoir Experiment (SMOSREX)

campaign (De Rosnay et al. 2006) over a fallow ground area, which is similar to that used by Calvet et al. (1999) (MUREX: monitoring the usable soil reservoir experimentally). The dataset comprises four years of measurements (2001–04), during which the area underwent very contrasting climatic conditions. In particular the severe drought of the summer of 2003 over western Europe is well represented. This study follows the work of Calvet and Noilhan (2000) by adding a comparison between several assimilation methods.

Following this introduction, section 2 presents the experimental site, the dataset, the newest version of the Interaction between Soil, Biosphere, and Atmosphere scheme (ISBA) LSM ISBA-A-gs, and the assimilation methods employed in this study [extended Kalman filter (EKF), ensemble Kalman filter (EnKF), simplified one-dimensional variational data assimilation (1DVAR), and tuning variational (T-VAR)]. Furthermore, the methodology to estimate the model error is presented, and the implementation of the four assimilation schemes is described. In section 3, the results for each assimilation method are shown, and a sensitivity study (to model and observation error) is carried out for the best performing method. Finally, section 4 summarizes the main conclusions and prospects.

## 2. Methodology

### a. Experimental site and dataset

The SMOSREX site is situated within the ONERA (French National Aerospace Research Establishment) center of Fauga-Mauzac, located 40 km at the south of Toulouse (43°23'N, 1°17'E, 188-m altitude). SMOSREX is a field-scale experiment, operative since 2001 with measurements similar to those of MUREX (Calvet et al. 1999). The experimental dataset is described below.

#### 1) AUTOMATIC MEASUREMENTS

A meteorological station is providing continuous measurements, every 30 min, of precipitation, atmospheric pressure, wind speed and direction, air humidity, air temperature, and incident and emitted solar and infrared radiation. Deep and surface soil temperature and soil moisture are monitored continuously on a half-hourly basis. To obtain a representative estimate of  $w_g$ , four probes (ThetaProbe, Delta T Devices) are vertically installed at different locations within the area, providing a measurement over the top 6 cm of the soil layer. Daily mean values  $w_g^j$  are obtained by averaging the four probe measurements. The uncertainty in  $w_g^j$  is given by the standard deviation of these measurements. Here  $w_2$  is obtained by calculating an average bulk soil

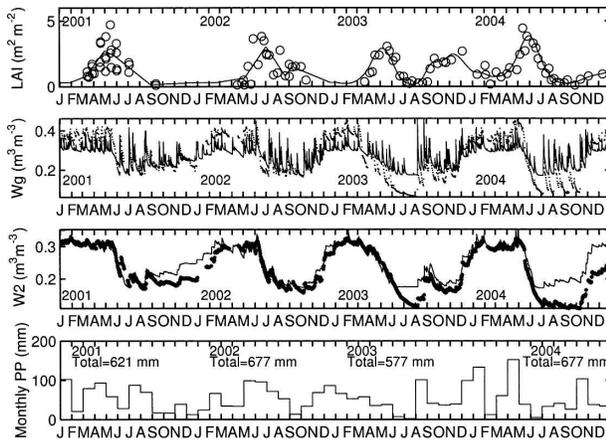


FIG. 1. (from top to bottom) LAI,  $w_g$ ,  $w_2$ , and monthly precipitation for 2001–04 as measured over the SMOSREX site. The interpolated LAI (solid line) is used to drive the ISBA-A-gs simulations and is superimposed on the observations (circles). Note that the LAI interpolation technique may change from one year to another, depending on data quality and frequency. The  $w_g$  and  $w_2$  simulations of ISBA-A-gs (solid line) are superimposed on the observations (dots).

water content from these surface probes and three ThetaProbe profiles set up at the same locations (1–3 m apart). These profiles consist of soil moisture sensors installed vertically at the surface (0–6 cm) and horizontally at depths of 10, 20, 30, 40, 50, 60, 70, 80, and 90 cm. Our goal, here, is to use the information provided by the experimental data as much as possible in order to define prescribed observation error statistics coherent with the field observations. For lack of a sufficient sampling of the plot, the spatial averaging is replaced by a temporal one (ergodicity principle). From the individual measurements, a mean and a standard deviation are computed on a daily time step. The daily standard deviation averaged over the year 2001 is assumed to be equal to the observation error:  $\sigma(w_g^{\text{OBS}}) = 0.03 \text{ m}^3 \text{ m}^{-3}$  and  $\sigma(w_2^{\text{OBS}}) = 0.02 \text{ m}^3 \text{ m}^{-3}$ . These errors are attributed to the subsequent years (2002–04).

## 2) MANUAL MEASUREMENTS

Measurements of the vegetation characteristics [leaf area index (LAI), green and dry biomass and height of the canopy] were carried out every two weeks from spring to autumn. Figure 1 shows the in situ observations of the LAI, the root-zone soil moisture, and the monthly accumulated precipitation for the four years (2001–04). It can be observed that 2003 was a particularly dry year, with a yearly accumulated precipitation of less than 600 mm. Unlike the other years, 2003 shows an atypical double cycle of LAI, with a first maximum in spring and another one at the beginning of the winter

TABLE 1. Main soil and vegetation parameters used in the ISBA-A-gs model over the SMOSREX site.

Soil parameters			
Parameter	Symbol	Unit	Value
Soil root depth	$d_2$	cm	95
Sand content	SAND	%	32.0
Clay content	CLAY	%	22.8
Field capacity	$w_{fc}$	$\text{m}^3 \text{ m}^{-3}$	0.30
Wilting point	$w_p$	$\text{m}^3 \text{ m}^{-3}$	0.17
Vegetation parameters			
Parameter	Symbol	Unit	Value
Mesophyll conductance	$g_m$	$\text{mm s}^{-1}$	0.56
Critical extractable soil moisture	$\theta_c$	%	50
Plant response to water stress	—	—	Drought tolerant

season. Precipitation is quite irregularly distributed during 2004, with a wet spring and a very dry summer. That causes a rapid growth of the vegetation and a marked senescence during the dry period, with  $w_2$  reaching values below wilting point ( $w_p$ ) during all the summer season and part of autumn.

In Table 1, a list of the most relevant characteristics of the soil over the SMOSREX site for ISBA-A-gs is provided. The soil is a loam characterized by its texture and density, which were determined in the laboratory. The wilting point and the field capacity parameters were derived from the clay content observations, by using the relationships given by Noilhan and Mahfouf (1996).

### b. Land surface model: ISBA-A-gs

The ISBA model was first developed by Noilhan and Planton (1989) and further improved by Noilhan and Mahfouf (1996) to describe the surface processes in weather and climate models.

The ISBA model uses the equations of the force–restore method (Deardorff 1977, 1978) to describe the evolution of five surface state variables: surface temperature ( $T_s$ ), mean surface temperature ( $T_2$ ), surface soil volumetric moisture ( $w_g$ ), total soil volumetric moisture ( $w_2$ ), and canopy interception reservoir ( $W_s$ ), together with the surface energy fluxes ( $LE$ ,  $H$ ,  $G$ ). The model was modified in order to account for the effect of the atmospheric carbon dioxide concentration on the stomatal aperture. This new version of ISBA was called ISBA-A-gs (Calvet et al. 1998b; Gibelin et al. 2006). The net assimilation of  $\text{CO}_2$  is used to predict the vegetation biomass and the LAI. However, a study of the impact of using an interactive LAI on the  $w_2$  analyses is

beyond the scope of this paper and, in this study, the LAI is prescribed from measurements. In Fig. 1, the ISBA-A-gs control simulation of  $w_2$  with the prescribed interpolated LAI (solid lines) is superimposed over the observations. The photosynthesis parameters (see Table 1) are the same as those prescribed during the MUREX campaign (Calvet and Soussana 2001). In section 3 these results are discussed.

### c. Assimilation methodologies

#### 1) DERIVED FROM KALMAN FILTERS (KF)

The well-known Kalman filter (Kalman 1960) is an optimal sequential assimilation method extensively used in various environmental problems. It is based on a minimum variance analysis scheme (see, e.g., Teunissen 2000). In such a scheme the best linear unbiased estimators (BLUE) yield at time  $i$  of the measurements for the variables to be analyzed ( $w_g$  and  $w_2$  in this study, for the KF case, hereafter embedded in the state vector  $\mathbf{x}_i$ ) and for the associated prognostic state variance–covariance matrix  $\mathbf{P}_i$  the following equations:

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i[\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f], \quad (1)$$

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i\mathbf{H})\mathbf{P}_i^f, \quad (2)$$

with

$$\mathbf{K}_i = \mathbf{P}_i^f\mathbf{H}^T[\mathbf{H}\mathbf{P}_i^f\mathbf{H}^T + \mathbf{R}_i]^{-1}, \quad (3)$$

where the superscripts ( $^f$ ) and ( $^a$ ) refer to the point in time just before and after the analysis, respectively;  $\mathbf{y}_i$  is the observation vector at time  $i$  ( $w_g^{\text{obs}}$  in this study);  $\mathbf{R}_i$  the associated variance–covariance error matrix;  $\mathbf{I}$  the identity matrix; and  $\mathbf{K}_i$  is called the Kalman gain. In the standard KF, a linear relationship is assumed as follows:

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i + u_i, \quad (4)$$

where  $\mathbf{H}$  is the observation operator, and  $u_i$  is a function accounting for the uncertainties of the measurements and the observation model, given the variance–covariance  $\mathbf{R}_i$ .

The analyzed state [Eq. (1)] is equal to the forecasted state plus the analysis increment, which is equal to the innovation vector (the difference between observations  $\mathbf{y}_i$  and the associated simulations in the observation space  $\mathbf{H}\mathbf{x}_i^f$ ) multiplied by the Kalman gain  $\mathbf{K}_i$ ;  $\mathbf{K}_i$  accounts for the errors in the observations and the prognostic state (the correction will be higher as more confidence is given to the observations). During the propagation step the system evolves according to the linear dynamics of the system:

$$\mathbf{x}_{i+1}^f = \mathbf{M}\mathbf{x}_i^a + w_i, \quad (5)$$

where  $\mathbf{M}$  is the prognostic model operator and  $w_i$  groups all modeled uncertainties (assumed normal distributed with zero mean and covariance equal to  $\mathbf{Q}_i$ ). Finally, for the propagation law of variances, the forecast error covariance matrix  $\mathbf{P}$  will evolve according to

$$\mathbf{P}_{i+1}^f = \mathbf{M}\mathbf{P}_i^a\mathbf{M}^T + \mathbf{Q}_i. \quad (6)$$

#### (i) Extended Kalman filter

Equations (1)–(6) describe the system propagation and the optimal updating completely, for linear problems. However, in most cases, physical systems are nonlinear and, in contrast to Eq. (5), the model system propagates in time using the full nonlinear dynamics of the system:

$$\mathbf{x}_{i+1}^f = M(\mathbf{x}_i^a) + w_i, \quad (7)$$

where  $M()$  is a nonlinear operator that groups all the model equations. The EKF equations differ from those of the standard KF in the way that the system is locally linearized around the forecasted vector  $\mathbf{x}_i^f$  at the time  $i$  of the observations. Here  $\mathbf{M}$  of Eq. (6) becomes the Jacobian of the prognostic model:

$$\mathbf{M}_i = \left. \frac{\partial M(\mathbf{x}_i)}{\partial \mathbf{x}_i} \right|_{\mathbf{x}_i = \mathbf{x}_i^f}. \quad (8)$$

#### (ii) Ensemble Kalman filter

One of the main drawbacks of the EKF is the time-consuming process of propagating the variance–covariance matrix  $\mathbf{P}_i$  when dealing with systems with a large number of state variables such as the one of NWP models (typically around  $10^6$ ). The EnKF (Evensen 1994; Burgers et al. 1998) circumvents this problem by using an ensemble of  $j$  state vectors, each of which represents a potential model trajectory.

Equations are those of the standard KF [Eqs. (1)–(6)], but now the state variance–covariance matrix  $\mathbf{P}_i$  is estimated throughout the statistics of the ensemble:

$$\mathbf{P}_i^f \approx \mathbf{P}_{e,i}^f = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{x}_{j,i}^f - \bar{\mathbf{x}}_i^f)^T (\mathbf{x}_{j,i}^f - \bar{\mathbf{x}}_i^f), \quad (9)$$

where  $\bar{\mathbf{x}}_i^f$  denotes the average of the ensemble. Moreover, observations are considered as random variables too, and an ensemble of observations is created at each observation time step by artificially perturbing them with Gaussian noise, following Burgers et al. (1998).

The state covariance matrix is implicitly propagated by the ensemble and, unlike the EKF, no linear approximation is involved. The mean of the ensemble is considered to be the most probable assimilated state and the dispersion of the ensemble will be an approxi-

mation of second moment of the model potential trajectory distribution. When the size of the ensemble tends to infinite, the ensemble  $\mathbf{P}_{e,i}^f$  matrix will converge to  $\mathbf{P}_i^f$  (Evensen 2003).

## 2) DERIVED FROM VARIATIONAL METHODS

In the variational method, instead of updating the system each time an observation is available, an assimilation window with a predefined length in time is used. The simulations are adjusted to the observations that are available within the assimilation window (at observations times) by minimizing a cost function  $J$ , with respect to a background information  $\mathbf{x}^b$ . Both the initial state and the model trajectory within the assimilation window are updated. The general form of  $J$  is given by

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \frac{\mathbf{B}^{-1}}{2} (\mathbf{x} - \mathbf{x}^b) + [\mathbf{y} - H(\mathbf{x})]^T \frac{\mathbf{R}^{-1}}{2} \times [\mathbf{y} - H(\mathbf{x})] = J_b(\mathbf{x}) + J_o(\mathbf{x}). \quad (10)$$

The cost function in Eq. (10) has two terms: the background term  $J_b(\mathbf{x})$ , which measures the distance between the state vector  $\mathbf{x}$  and the a priori state  $\mathbf{x}^b$  (weighted by the background error matrix  $\mathbf{B}$ ), and the observation term  $J_o(\mathbf{x})$ , which accounts for the distance between the vector of observations during the assimilation window,  $\mathbf{y}$ , and the simulations weighted by the observation error matrix  $\mathbf{R}$ . The subscript  $i$  has been omitted in Eq. (10) as, in contrast to sequential methods, all the observations available within the assimilation window are considered for variational methods. The projection of the state vector in the observation space is done through the observation operator  $H()$ , which is often nonlinear and includes the integration over time through the model operator  $M()$ . The minimization of  $J$  is generally computed by applying the descent gradient method for which the adjoint and the tangent linear models are needed. Building these models is usually a time-consuming task. In this study, as described below, a numerical linearization is used in order to avoid the use of the adjoint and tangent linear models.

### (i) Simplified 1DVAR

The simplified 1DVAR is an assimilation method developed by Balsamo et al. (2004) and adapted to our study. The observation operator is numerically linearized by perturbing the initial conditions (i.e., the state vector at the beginning of the assimilation window). Under this tangent linear hypothesis the observation operator  $H()$  can be developed by a first-order expansion:

$$H(\mathbf{x} + \Delta\mathbf{x}) \approx H(\mathbf{x}) + H(\Delta\mathbf{x}). \quad (11)$$

The minimum of the cost function is given by  $\nabla J = 0$  and, with the hypothesis that errors follow a normal distribution, it takes the following form:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y} - H(\mathbf{x}^b)] \quad (12)$$

with

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1}. \quad (13)$$

Note that the 1DVAR analyses in this study only concern  $w_2$  ( $1 \times 1$  control vector), whereas the KF analyses concern  $w_2$  and  $w_g$  ( $2 \times 1$  control vector). Indeed, a control vector of two variables does not imply higher computational costs for the KF, whereas the inclusion of the  $w_g$  analysis in the 1DVAR would require an extra run.

### (ii) Variational tuning method

T-VAR is a simplified suboptimal variational method introduced by Calvet and Noilhan (2000). It has the ability to retrieve soil moisture estimates of the deep reservoir by using a window of 10 days with four independent observations (if no missing values) of  $w_g$ . They are globally adjusted to the model estimates by a systematic exploration of all the potential initial values of the root-zone soil moisture. Here, the control vector is composed of  $w_2$ , only ( $w_g$  is not analyzed). The retrieved value of  $w_2$  corresponds to the minimum of the root-mean-square error (RMSE), which is in fact the cost function  $J$  to be minimized, without a background term and with  $\mathbf{R} = \mathbf{I}$ . Although a minimization is performed by systematic search of model initial state that best-fit the observations, there is no optimal use of error statistics.

## 3) METHODOLOGICAL DISCUSSION

In this study, three methods over four rely on the linear least squares theory: EKF, EnKF, and 1DVAR. Besides this apparent similarity, differences in the analysis calculation exist. First, EKF and 1DVAR rely on the local linearization of the model equations whereas the full nonlinear system dynamics is accounted for by the EnKF. The linearization of the model is valid if the time step between two observations is smaller than the correlation length of the state variables. This is not always the case, particularly for the rapid variations affecting surface soil moisture. In the case of highly nonlinear models, the filter diverges from optimality and may become unstable (Miller et al. 1999). For example, in our case the linear hypothesis may not work when strong precipitation or evaporation

TABLE 2. Definition of the background error matrix  $\mathbf{P}$  and the observation error matrix  $\mathbf{R}$ , for four assimilation schemes: EnKF, EKF, a simplified 1DVAR, and T-VAR.

	Background error matrix $\mathbf{P}$	Observation error matrix $\mathbf{R}$
EnKF	$\mathbf{P}_i \approx ((\mathbf{x}_{j,i}^f - \overline{\mathbf{x}_{j,i}^f})^T (\mathbf{x}_{j,i}^f - \overline{\mathbf{x}_{j,i}^f}) / N - 1)$	$\mathbf{R}_i = ((\mathbf{y}_{j,i} - \overline{\mathbf{y}_{j,i}})^T (\mathbf{y}_{j,i} - \overline{\mathbf{y}_{j,i}}) / N - 1)$
1DVAR	$\mathbf{B} = \mathbf{Q}^{\text{var}}$	$\mathbf{R} = [2\sigma(w_g^{\text{obs}})]^2$
EKF	$\mathbf{P}_i^f = \mathbf{M}_i \mathbf{P}_{i-1}^a \mathbf{M}_i^T + \mathbf{Q}_i$	$\mathbf{R}_i = [2\sigma(w_g^{\text{obs}})]^2$
T-VAR	No background term	$\mathbf{R} = \mathbf{I}$

rates take place. A decoupling between both variables may then lead to inadequate corrections.

For Kalman filters, the background error covariance is sequentially updated. The information stored in the covariance matrix is propagated in time and thus extends the coherence of the assimilation beyond the assimilation time window. The propagation of the covariance information is done through the linear model for EKF whereas it is implicit thanks to a stochastic sampling of the a priori space for EnKF. In contrast to the Kalman type methods, a fixed background error matrix is assumed for the 1DVAR.

A fourth method is tested in this study that mimics the case where no a priori information is available. This method is often used on extended time windows, for example, one year, to tune parameters related to soil moisture, like field capacity. Although it is clearly inferior to the other methods, this simple algorithm permits us to assess to what extent  $w_2$  can be analyzed when both the quality of the observations is not known and a priori estimates of  $w_2$  are unavailable.

#### d. Implementation of the assimilation methods

In this subsection the practical implementation of the assimilation algorithms employed in this study and the requirement for working in a normalized space are described. A description of the model forecast and background covariance error matrices ( $\mathbf{P}$  and  $\mathbf{B}$ ) and the observation error matrix  $\mathbf{R}$  is presented for the four assimilation methods in Table 2. Note that we have set the observation error  $\mathbf{R}$  to twice the uncertainty of the observations  $[\sigma(w_g^{\text{OBS}})]$ . This step was taken because the experimental setup does not permit us to quantify the spatial representativeness error. To take this effect into account the error in the observations was inflated empirically by a factor of 2.

##### 1) NORMALIZATION OF THE STATE VARIABLES

Calvet and Noilhan (2000) pointed out the need to normalize soil moisture before any data assimilation is

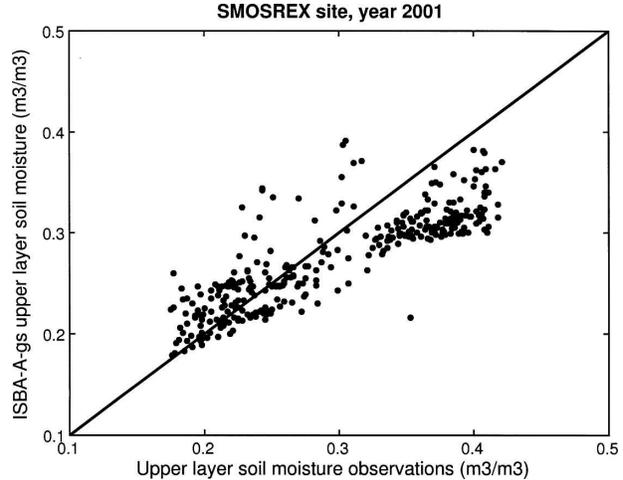


FIG. 2. Observed vs modeled surface soil moisture ( $w_g$ ) for 2001. The scatterplot is compared with the 1:1 relationship (solid line).

undertaken, because of an existing bias in the simulated  $w_g$ . Indeed,  $w_g$  is a model-dependent variable. In particular, peak values within a year of  $w_g$  are likely to vary from one model to another. In Fig. 2 the comparison between the  $w_g$  observations and the ISBA-A-gs simulations are plotted for 2001. It is shown that their relationship is far from the 1:1 line, which would be the case if the model were perfect. Therefore, all soil moisture observations for the period of 2001–04 were normalized using the maximum and minimum values observed in 2001, which was chosen as the calibration year. Similarly, all soil moisture simulations were normalized between the maximum and minimum values estimated from the model simulations for the year 2001. In this way observations and simulations can be compared in a normalized space ( $w_g$  and  $w_2$  ranging from 0 to 1).

##### 2) IMPLEMENTATION

In this subsection, the technique to estimate the model error is described, as well as the approach to apply the assimilation schemes.

###### (i) Model error

An accurate estimation of the model error is likely to be the most difficult task in the error prescription (Bouttier 1994; Reichle et al. 2002; François et al. 2003). An approximation of the model imperfections was obtained as follows: at each observation time, an ensemble of initial conditions was created for the two state variables ( $w_g$  and  $w_2$ ) from the observation values. The model was then run for each member of the ensemble, every three days (which mimics the frequency of satellite-derived  $w_g$  observations) to estimate the

forecasted error for the sequential methods and every 10 days, that is, the assimilation window duration, to estimate the background error for the 1DVAR. At the end of each time window (3 or 10 days), the dispersion of the residuals (difference between an ensemble member and the observation value) was calculated. This value is considered as the  $q_{x_i}$  term of the  $\mathbf{Q}^{\text{type}}$  [type being sequential (seq) or variational (var)] matrix at time  $i$ , and an annual evolution of this term is obtained. The year 2001 was chosen as the calibration year because of its characteristics in terms of atmospheric forcing, which resembled an average year. If the model error is stationary, the temporal evolution is close to constant. This is the case for the state variable  $w_2$  and to a lesser extent for  $w_g$ , due to its shorter temporal correlation length. Nevertheless, from here onward, the estimated forecast and background error are defined as follows:

$$\mathbf{Q}^{\text{seq}} = \begin{bmatrix} 0.05^2 (\text{m}^3 \text{m}^{-3})^2 & 0 \\ 0 & 0.02^2 (\text{m}^3 \text{m}^{-3})^2 \end{bmatrix}$$

$$\mathbf{Q}^{\text{var}} = \begin{bmatrix} 0.06^2 (\text{m}^3 \text{m}^{-3})^2 & 0 \\ 0 & 0.01^2 (\text{m}^3 \text{m}^{-3})^2 \end{bmatrix}. \quad (14)$$

$$\mathbf{M}_i = \begin{bmatrix} \frac{M_i[w_g^{(i-1)} + \Delta w_g^{(i-1)}] - M_i[w_g^{(i-1)}]}{\Delta w_g^{(i-1)}} & \frac{M_i[w_g^{(i-1)} + \Delta w_g^{(i-1)}] - M_i[w_g^{(i-1)}]}{\Delta w_2^{(i-1)}} \\ \frac{M_i[w_2^{(i-1)} + \Delta w_2^{(i-1)}] - M_i[w_2^{(i-1)}]}{\Delta w_g^{(i-1)}} & \frac{M_i[w_2^{(i-1)} + \Delta w_2^{(i-1)}] - M_i[w_2^{(i-1)}]}{\Delta w_2^{(i-1)}} \end{bmatrix}, \quad (15)$$

where  $M_i(\cdot)$  is the nonlinear full operator at time  $i$ , and  $\Delta w_g$  and  $\Delta w_2$  are the perturbations of the updated state variables  $w_g$  and  $w_2$  at the precedent time  $i - 1$ , respectively. The size of the perturbation of the initial state is critical. Theoretically, an infinitesimal perturbation in the neighborhood of the initial state vector would ensure that the linear hypothesis is fulfilled. However it may cause an adverse effect due to numerical errors. Large perturbations may also produce errors when nonlinear effects are predominant. In this study a value of  $0.05 \text{ m}^3 \text{m}^{-3}$  was chosen as perturbation of the initial state and the error variance-covariance matrix of the forecasted state is integrated in time using Eq. (6). Since the state variable  $w_g$  is directly observed, the observation operator  $\mathbf{H}$  is in this case:  $\mathbf{H} = [1 \ 0]$ . Thus, developing Eq. (3) and combining it with Eq. (1), the corrections of the forecasted state variables at the time  $i$  of the observation are given by

In Eq. (14), it is assumed that there is no correlation between the model error on  $w_g$  and the model error on  $w_2$ , by setting the nondiagonal terms to zero. Indeed, preliminary calculations (not shown) of the cross-correlation terms produced negligible values. The introduction of these values into the covariance matrix had only minor effects on the results of the assimilation.

### (ii) EKF

The assimilation of remote sensing data into LSMs usually constitutes a low-dimensional problem in comparison with the assimilation of observations in atmospheric or oceanic models. Therefore the propagation of the model error covariance matrix is rather straightforward and methods like the EKF can be tested easily.

An initial background error covariance matrix  $\mathbf{P}_0$  is constructed using the uncertainty of the observations and assuming no initial correlation between the state variables, hence a block diagonal covariance matrix. To propagate  $\mathbf{P}$  between observations and apply the tangent linear hypothesis, a perturbation of the initial state vector (composed in the present case of  $w_g$  and  $w_2$ ) is carried out, yielding the numerical linear matrix  $\mathbf{M}_i$  [which is substituted into Eq. (6)]:

$$w_{g,i}^a = w_{g,i}^f + \frac{\mathbf{P}_i(1, 1)}{\mathbf{P}_i(1, 1) + \mathbf{R}_i} (w_{g,i}^{\text{obs}} - w_{g,i}^f) \quad (16)$$

and

$$w_{2,i}^a = w_{2,i}^f + \frac{\mathbf{P}_i(1, 2)}{\mathbf{P}_i(1, 1) + \mathbf{R}_i} (w_{g,i}^{\text{obs}} - w_{g,i}^f), \quad (17)$$

where the  $\mathbf{P}_i(1, 1)$  and  $\mathbf{P}_i(1, 2)$  terms are elements of the  $2 \times 2$   $\mathbf{P}_i$  matrix.

### (iii) EnKF

Samples of the initial ensemble are created assuming a normal distribution with a mean equal to the first observation and a variance-covariance matrix equal to  $\mathbf{R}$ . An ensemble of  $N = 100$  members is used following Evensen (2003). A rapid convergence of the forecasted  $w_2$  is observed (not shown) and the analysis ensemble tends to collapse. Physically, the explanation is that the

TABLE 3. Surface and root-zone soil moisture yearly scores of the control simulation and for the whole 2001–04 period [RMSE ( $\text{m}^3 \text{m}^{-3}$ ), bias ( $\text{m}^3 \text{m}^{-3}$ ), and skill score  $E$ ].

	2001			2002			2003			2004			2001–04		
	RMSE	Bias	$E$	RMSE	Bias	$E$	RMSE	Bias	$E$	RMSE	Bias	$E$	RMSE	Bias	$E$
$w_g$	0.05	−0.020	0.56	0.06	−0.023	0.50	0.08	0.027	0.51	0.08	0.006	0.60	0.07	−0.03	0.59
$w_2$	0.02	0.010	0.82	0.02	0.009	0.92	0.03	0.015	0.83	0.05	0.044	0.51	0.03	0.020	0.73

water loss by evaporation modulates the root-zone soil moisture: the members of the ensemble starting with a wet soil are loosing more water than the drier soils. This tends to make  $w_2$  converge to the same value on average over a year. To prevent the collapse of the ensemble, the new ensembles are built by multiplying the variance of the updated ensemble by an inflation factor. This approach is equivalent to the covariance inflation described in Anderson and Anderson (1999). Moreover the atmospheric forcing is perturbed at each model step by adding random Gaussian noise. The inflation factor was empirically calibrated by minimizing the RMSE between the analyzed and the observed  $w_2$  for the year 2001. A value of 1.35 was found and kept constant for the other years. For the perturbation of the atmospheric forcing, the following standard deviations were used:  $60 \text{ W m}^{-2}$ ,  $35 \text{ W m}^{-2}$ , 50% relative difference, 10 K,  $1 \text{ m s}^{-1}$ , and 1000 Pa, for shortwave and longwave incident radiation, precipitation, air temperature, wind speed, and surface pressure, respectively. As for the EKF, the observation operator remains  $\mathbf{H} = [1 \ 0]$ . Finally, since information about the observation error is obtained from the ThetaProbe measurements, an ensemble of normally distributed observations is created, with  $\sigma$  equal to twice the observation error of  $w_g$ .

#### (iv) Simplified 1DVAR

In this simplified version of the 1DVAR the  $\mathbf{B}$  and  $\mathbf{R}$  variance–covariance matrices are estimated once and remain unchanged for the rest of the assimilation period. After the updating step, the assimilation window slides in time until a new observation is found (minimum of three days). Since observations are assimilated more than once, this method departs from optimality in theory. Nevertheless, in this context, rather than searching for the optimality, our objective is the comparison of this assimilation approach with the T-VAR in the fairest way. For an operational application, it is recommended using sequential assimilation windows and suppressing the first observation within the assimilation window. In this study, this does not adversely affect the analyses (not shown). The background matrix has been left fixed and equal to  $\mathbf{Q}^{\text{var}}$  [Eq. (14)]. On the other hand, in consistency with the sequential methods, the

variance of the observations has been set to twice the uncertainty of the observations of  $w_g$ . The linearization of the model is done through the observation operator  $\mathbf{H}$  by perturbing the initial state of  $w_2$ . Finally, the magnitude of the perturbation has been set to the same value as for EKF.

#### (v) T-VAR

In this simple variational method no initial prescription of variance–covariance error matrices, nor background information, are needed. At each analysis step a systematic exploration of the potential initial conditions is conducted, by initializing the model with values between  $0.10 \text{ m}^3 \text{m}^{-3}$  ( $w_{2,\text{min}}$ ) and  $0.40 \text{ m}^3 \text{m}^{-3}$  ( $w_{2,\text{max}}$ ), incremented by steps of  $0.015 \text{ m}^3 \text{m}^{-3}$ . The analysis of  $w_2$  is undertaken by minimizing the cost function between observations  $w_g^{\text{obs}}$  and model estimations  $w_g^{\text{sim}}$ :

$$J(w_2) = \sqrt{\frac{1}{n} \sum_{i=1}^n (w_{i,g}^{\text{sim}} - w_{i,g}^{\text{obs}})^2}, \quad (18)$$

with  $n$  the number of measurements within the assimilation window. According to Calvet and Noilhan (2000), if the  $w_g^{\text{obs}}$  are available every 3 days, a 10-day assimilation window yields the best results for the MUREX site (i.e., using four observations).

### 3. Results and discussion

#### a. Root-zone soil moisture simulation (2001–04)

In Fig. 1 the ISBA-A-gs simulations of  $w_2$  and  $w_g$  are compared with the observations during the period 2001–04. Error statistics are given in Table 3 (RMSE, bias, and skill score). The skill score  $E$  is defined as

$$E = 1 - \frac{\sum_i (x_i^{\text{obs}} - x_i^{\text{mod/ana}})^2}{\sum_i (x_i^{\text{obs}} - \bar{x}^{\text{obs}})^2}, \quad (19)$$

where  $x$  refers to the soil moisture variables, either observed (obs), simulated (mod), or analyzed (ana).

In general, the agreement is good as long as the observed  $w_2$  is above the  $w_p$ . The model overestimates  $w_2$  from September 2001 to March 2002. This may be due

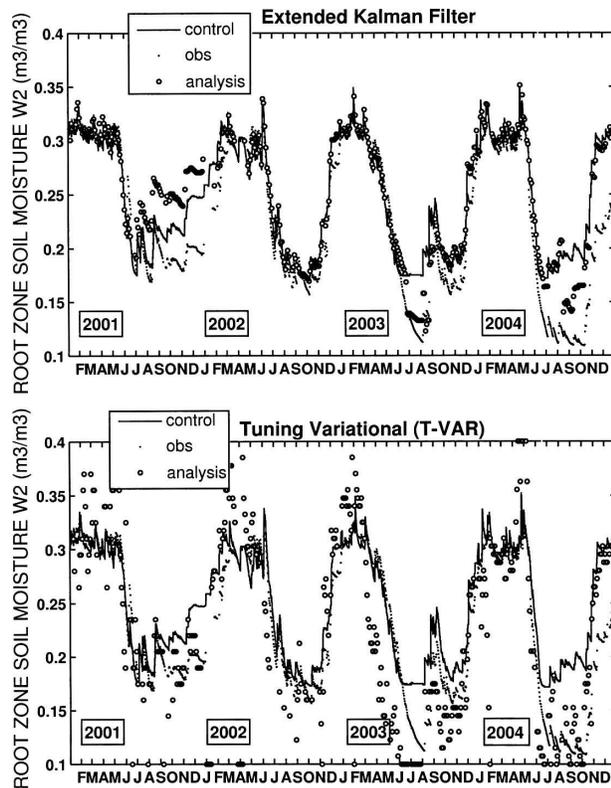


FIG. 3. Analysis of the root-zone soil moisture from surface observations using (top) an extended Kalman filter and (bottom) a tuning variational method, for 2001–04 over the SMOSREX experimental site: model control simulation (solid line), observations (dots), and analysis (circles).

to the lack of regular LAI measurements during this period and to the underestimation of LAI by the linear interpolation employed. Low values of LAI tend to decrease the root water extraction and transpiration rate, leading to an overestimation of the soil moisture with regard to the observations during this period. The model overestimates  $w_2$  also during the droughts of the summers of 2003 and 2004. In this case, the modeled  $w_2$  reaches the prescribed  $w_p$  value and root extraction stops, whereas in reality, evaporation may continue even with a soil moisture below the prescribed wilting point. Our goal is to investigate to what extent the assimilation schemes used here are able to improve the model simulation.

#### b. Analysis of the root-zone soil moisture: EKF and T-VAR

Figure 3 shows the results of the EKF and T-VAR analysis of the root-zone soil moisture. The analyses, the observations, and the free model simulations are plotted together for comparison purposes (hereafter,

all the soil moisture results are given in absolute units of  $\text{m}^3 \text{m}^{-3}$ ). By using the EKF, an enhancement is achieved with regard to the control model simulation ( $E = 0.85$  against  $E = 0.73$ ). Small improvements are achieved for the years 2003 ( $E = 0.93$ ) and 2004 ( $E = 0.66$ ), where  $w_2$  goes below the wilting point. Here, the nonlinear effects observed during the dry periods trigger large Kalman gains, and a significant correction of  $w_2$ . For 2001 and 2002, the EKF tends to degrade the model estimations (Table 4). This is a consequence of the lack of sensitivity of the state variables to a single perturbation of the surface soil moisture between two observations (3 days). This means that, except for periods of large recharge or high evaporation rates, the system behaves stable to perturbations of  $w_g$ . As a consequence, the product  $\mathbf{MPM}^T$  of Eq. (6) often acquires a very low value. Hence, the forecasted  $\mathbf{P}_i$  matrix is mainly controlled by the estimated model covariance error matrix  $\mathbf{Q}_i$  [see Eq. (6)]. In a first approximation, the correction of  $w_{2,i}^f$  [Eq. (17)] is proportional to the  $\mathbf{Q}_{12}$  term of the model error matrix, which was assumed to be zero [Eq. (14)]. Therefore, by using this method, it is expected that only small corrections of the  $w_{2,i}^f$  a priori estimate are undertaken. With the EKF, the control simulations are virtually unchanged during the first half of 2003, and since the control model fits the observations well, a very good skill score is obtained.

Therefore, even though an apparently good performance of the EKF is obtained ( $E = 0.85$  for the whole period), the corrections are small for the major part of the period. This is attributed to the force–restore scheme of ISBA, which produces a low sensitivity of the surface soil moisture to a perturbation of the root-zone soil moisture, during the days following the perturbation. Concerning T-VAR, a high scattering of the retrieved  $w_2$  is observed (Fig. 3), which deteriorates the skill score with regard to the model estimation. In general this method is able to reproduce the overall shape of the evolution of the  $w_2$  observations, but also the limitations are obvious, since no background information is used. Locally, retrieved points are found within the uncertainty of the observations, and either during the drought of 2003 or 2004, the analyzed  $w_2$  is below the wilting point, which confirms the potential of this method to retrieve information on  $w_2$  despite the lack of any background information.

#### c. Analysis of the root-zone soil moisture: EnKF and simplified 1DVAR

In Fig. 4 the results obtained with the EnKF and 1DVAR are shown. For both methods, a significant overall improvement of the control model simulations is achieved (see Tables 3 and 4); in particular the model

TABLE 4. Root-zone soil moisture analysis yearly scores and for the whole 2001–04 period [RMSE ( $\text{m}^3 \text{m}^{-3}$ ), bias ( $\text{m}^3 \text{m}^{-3}$ ), and skill score  $E$ ], using an EnKF with  $N = 10, 20, 50, 100$ , and 200 members, an EKF, a simplified 1DVAR, and a T-VAR.

	2001			2002			2003			2004			2001–04		
	RMSE	Bias	$E$	RMSE	Bias	$E$	RMSE	Bias	$E$	RMSE	Bias	$E$	RMSE	Bias	$E$
EnKF															
( $N = 10$ )	0.02	-0.001	0.91	0.05	0.043	0.15	0.07	0.052	0.12	0.08	0.064	0.05	0.06	0.040	0.25
( $N = 20$ )	0.02	0.001	0.91	0.05	0.044	0.18	0.04	-0.008	0.71	0.07	0.065	0.15	0.05	0.024	0.43
( $N = 50$ )	0.02	-0.001	0.92	0.03	0.004	0.62	0.03	-0.005	0.77	0.06	0.065	0.19	0.04	0.016	0.58
( $N = 100$ )	0.01	-0.001	0.94	0.03	0.006	0.76	0.03	-0.005	0.74	0.04	0.029	0.68	0.03	0.009	0.78
( $N = 200$ )	0.01	-0.001	0.94	0.03	0.007	0.67	0.03	0.004	0.82	0.04	0.028	0.76	0.03	0.010	0.81
1DVAR	0.02	-0.011	0.90	0.01	0.001	0.93	0.02	-0.009	0.89	0.04	0.029	0.79	0.02	0.002	0.86
EKF	0.04	0.026	0.44	0.02	0.014	0.81	0.02	0.010	0.93	0.05	0.037	0.66	0.03	0.011	0.85
T-VAR	0.04	-0.056	0.50	0.04	-0.006	0.37	0.06	-0.026	0.32	0.05	0.009	0.58	0.05	0.014	0.53

overestimation at the end of 2001 and at the beginning of 2002 is corrected. The 1DVAR shows a slightly better performance during periods where the simulations are limited by the prescribed wilting point, that is, the summers of 2003 and 2004. In our study case, the EnKF outperforms 1DVAR for 2001, which is the calibration year for the inflation factor, but the 1DVAR is better, on average, for the whole 2001–04 period (RMSE =  $0.02 \text{ m}^3 \text{m}^{-3}$  and  $E = 0.86$ , compared to RMSE =  $0.03 \text{ m}^3 \text{m}^{-3}$  and  $E = 0.78$  for an EnKF with 100 members). Furthermore, the 1DVAR analyses are smoother than those of the EnKF. The main shortcoming of the EnKF is observed close to wilting point. Indeed, the main difference between the two methods consists in the EnKF background error covariance matrix propagated by the ensemble, as opposed to the 1DVAR fixed background error. Close to wilting point, the spread of the ensemble broadens significantly (not shown) and the analysis does not match the observations. These results suggest that, over the SMOSREX site, the analyses are more stable and accurate by using a fixed background error. A possible explanation is that the normalization of  $w_g$ , performed for 2001, does not eliminate all the seasonal biases.

The assimilation in the transition period between 2001 and 2002 and the drought of 2003 shows a much better performance of the 1DVAR, resulting in better yearly skill scores for 2002, 2003, and 2004 (Table 4). On the other hand, the EnKF performs better than the 1DVAR in 2004. To understand how the model estimations are corrected by the 1DVAR method, Fig. 5 shows an example of the temporal evolution of  $w_2$  along with the gain components for the year 2003 (Figs. 5a and 5b, respectively). In our case the gain of the 1DVAR is a vector of four components (one for each observation within the assimilation window). The beginning of the assimilation time window coincides with

the first available observation [see section 2d(2)iv]; therefore, the first component of the gain is negligible with regard to the other three. Then in Fig. 5 (bottom) only the last three components of the gain are considered. It is observed that with the soil moisture at field capacity the greatest term of the gain is generally the

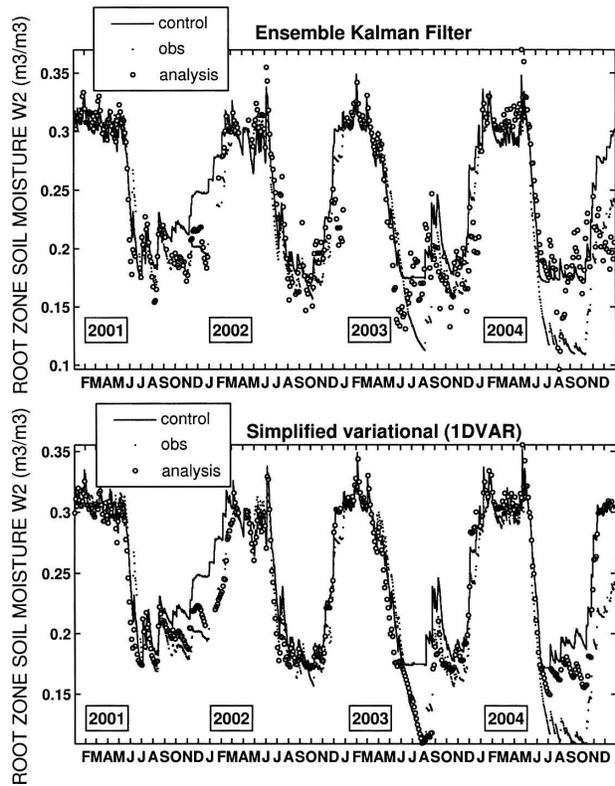


FIG. 4. Analysis of the root-zone soil moisture from surface observations using (top) an ensemble Kalman filter and (bottom) a simplified 1DVAR method, for 2001–04 over the SMOSREX experimental site: model control simulation (solid line), observations (dots), and analysis (circles).

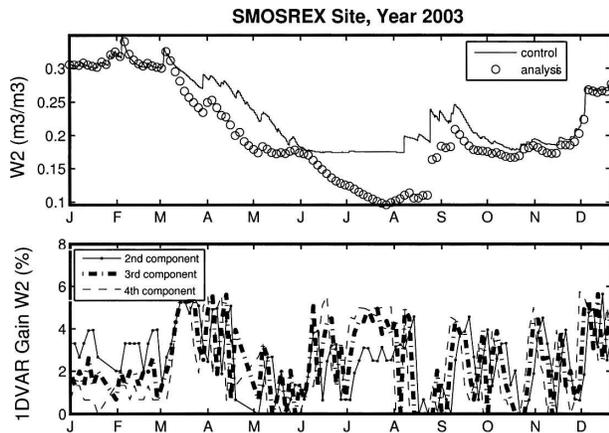


FIG. 5. Gain components for the simplified 1DVAR method compared with the yearly evolution of  $w_2$  in 2003. (top) The control simulation (solid line) is superimposed on the analyses (circles). (bottom) The evolution of the gain components (multiplied by 100) at each analysis step.

second term, that is, the one corresponding to the difference between the second observation (within the assimilation window) and the model estimate. For the following observations, small corrections of  $w_2$  were undertaken, of the order of 5%. Nonlinearities are more significant during the rest of the hydrological cycle, when no privileged gain component is found. In that case, the tangent linear model may depart from the real model trajectory and important deviations with regard to the observations are found. Moreover the innovation term is also larger, which indicates a decoupling between  $w_g$  and  $w_2$ , leading sometimes to inaccurate corrections.

#### d. Sensitivity to different levels of prescribed errors

The performance of an assimilation method is very dependent on accurate prescription of the error statistics. Over the SMOSREX experimental site, continuous observations of both  $w_g$  and  $w_2$  are available. We have taken advantage of all these sources of information to define errors, coherent with the terrain observations. A sensitivity study permits us to assess how the system behaves with regard to background and observation errors. Figure 6 shows the contour lines of regions of the same performance for different values of the observation and forecasted/background errors for the simplified 1DVAR. A maximum of efficiency is obtained for an observation error of  $0.07 \text{ m}^3 \text{ m}^{-3}$  and a forecasted error equal to the estimated model error  $\mathbf{Q}^{\text{var}}$  with a skill score close to 0.90 and a RMSE of  $0.02 \text{ m}^3 \text{ m}^{-3}$ . Even though this sensitivity study is specific to this experimental site, it is important to note that around this peak of efficiency a broad region is found

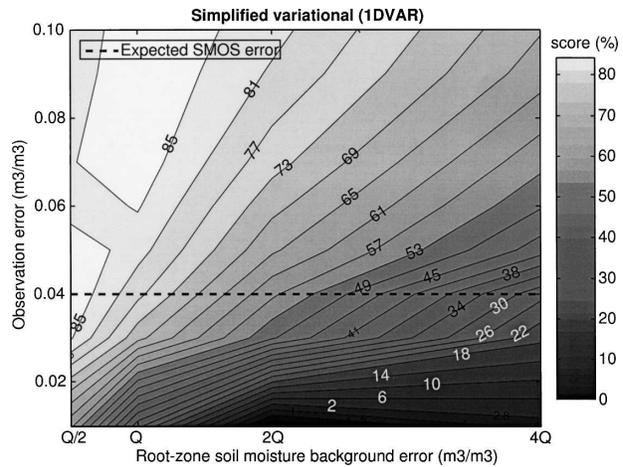


FIG. 6. Skill score for different error levels in observations ( $\mathbf{R}$ ) and forecasted state variables ( $\mathbf{P}$ ) for the simplified 1DVAR. Note that the  $\mathbf{Q}$  values in the abscises axis are those in Eq. (14). The dashed line superimposed on the figures corresponds to the performance with an observation error corresponding to the SMOS satellite specification.

where the skill score is higher than 0.8 and the RMSE is lower than  $0.03 \text{ m}^3 \text{ m}^{-3}$ , confirming thus the skill of a method that uses several observations (for each analysis) to generate correct gains. However at the boundaries, that is, very small observation errors or large forecasted state errors, a much larger RMSE and a sharp drop in performance is found. Monte Carlo-based methods, like the EnKF, could partially correct and counteract large deviations from the true state by using an ensemble of model trajectories.

Furthermore, within the range of the expected SMOS observation errors, of about  $0.04 \text{ m}^3 \text{ m}^{-3}$ , a good performance (around or higher than 0.8) of the system is observed for a wide range of background model error. This shows the potential use of the SMOS data to obtain a spatialized information on the root-zone soil moisture.

#### e. Processing time

To finish this analysis the importance of computational time of an assimilation algorithm in an operational system has to be emphasized, in particular when run over large areas. In Table 5, the total processing times for a whole year and for the four assimilation algorithms are compared. The runs were performed on the same platform, an Intel Pentium IV processor with a 2.40-GHz CPU. The simplified 1DVAR appears to be a good trade-off between computing time and the quality of the results. It can be seen (Table 5) that by using an ensemble of around 10 members the EnKF and the simplified 1DVAR are comparable in terms of comput-

TABLE 5. Computer processing time for a whole year (in seconds) for the EnKF, EKF, simplified 1DVAR, and a T-VAR. The computing time for the EnKF is shown for ensembles of  $N = 10, 20, 50, 100,$  and  $200$  members.

EnKF					EKF	1DVAR	T-VAR
10	20	50	100	200			
81	146	353	738	1358	36	76	300

ing time. Nevertheless, the statistics of the ensemble would be of lesser quality and, consequently, the quality of the retrievals of  $w_2$ . To get closer to a performance similar to the 1DVAR, an ensemble of around 200 members is necessary for the EnKF, thus increasing the EnKF processing time.

#### 4. Summary and conclusions

Four assimilation methods were implemented over an experimental site in southwestern France (SMOSREX) and the analyzed soil moisture results were compared. Three methods were based on least squares principles (EKF, EnKF, and 1DVAR) and one was a simple tuning method (T-VAR). The assimilation approaches and their practical implementation in the ISBA-A-gs land surface model were described and discussed. The multiyear SMOSREX dataset (2001–04) allowed us to assess the performance of the assimilation methods in contrasting conditions. In particular, marked droughts were observed during the summers of 2003 and 2004, for which the observed root-zone soil moisture was lower than the wilting point of the control simulation of ISBA-A-gs. The difficulty of the model to adequately reproduce the drought in 2003 and 2004 offered a good test for the assimilation schemes. In general, the four methods provided satisfactory results. The best performance was shown by the 1DVAR, with a skill score of 0.86, improving the control simulations (skill score of 0.73). Finally, a sensitivity study of the 1DVAR performance to different levels of background and observation errors was conducted.

The analysis results over SMOSREX show that

- The  $w_2$  analyses were improved by using a background/a priori information: the 1DVAR outperformed the T-VAR, which did not use any background term.
- The EnKF, which propagates the covariance through the sampling of several model trajectories, was more efficient than the EKF, which relies on the tangent linearization of the model equations to propagate the covariance information.
- The EnKF is a promising technique to deal with high

nonlinear systems, but over the SMOSREX site, it was outperformed by the 1DVAR. This result was attributed to 1) the limited nonlinearity of the system, which could have prevented the expression of the added value of the EnKF; 2) the difficulty in “tuning” the algorithmic parameters of the EnKF, such as the inflation factor, which, in this study, was required to prevent the ensemble from collapsing.

It is also important to remark that, in this study, the limited performance of the EKF (despite its apparently good behavior) could be related to the functioning of the ISBA-A-gs model (and models relying on the force–restore scheme) rather than to the assimilation method itself. Indeed, the force–restore approach presents a low sensitivity of the KF state variables to a perturbation of the surface soil moisture for the days following the date of the perturbation.

A sensitivity study showed that the 1DVAR method leads to good performance for a large range of background and observation errors. Moreover, using  $0.04 \text{ m}^3 \text{ m}^{-3}$  as a prescribed error in the  $w_g$  observations, that is, the error level that is expected from the SMOS satellite (shown as a dashed line in Fig. 6), good results are also obtained provided that variances in the forecasted state variables do not exceed the estimated model error.

Finally, with the lower computing time, the 1DVAR is a good alternative to the EnKF for the development of an operational data assimilation system aiming to analyze root-zone soil moisture from surface soil moisture observations. Nevertheless, this promising method needs to be tested at other experimental sites representing different geoclimatic environments, and further research is needed before the implementation of a full 2D application, in particular concerning the spatial correlation of background errors (Reichle and Koster 2003).

*Acknowledgments.* This study was cofunded by the European Commission within the GMES initiative in FP6, in the framework of the geoland integrated GMES project on land cover and vegetation. The authors thank Dr. Sébastien Masart (CERFACS, Toulouse), Dr. Christoph Rüdiger (CNRM, Toulouse), and Dr. Gianpaolo Balsamo (ECMWF, Reading) for fruitful discussions, as well as the anonymous reviewers, for their helpful comments.

#### REFERENCES

- Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.

- Balsamo, G., F. Bouysse, and J. Noilhan, 2004: A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model. *Quart. J. Roy. Meteor. Soc.*, **130A**, 895–915.
- , J.-F. Mahfouf, S. Belair, and G. Deblonde, 2006: A global root-zone soil moisture analysis using simulated L-band brightness temperature in preparation for the Hydros satellite mission. *J. Hydrometeorol.*, **7**, 1126–1146.
- Beljaars, A. C. M., P. Viterbo, M. J. Miller, and A. K. Betts, 1996: The anomalous rainfall over the United States during 1993: Sensitivity to land surface parameterization and soil moisture anomalies. *Mon. Wea. Rev.*, **124**, 362–383.
- Bouttier, F., 1994: Sur la prévision de la qualité des prévisions météorologiques (On the assessment of the weather forecast quality). Ph.D dissertation, Université Paul Sabatier III, 240 pp. [Available from Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse CEDEX, France.]
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Calvet, J.-C., and J. Noilhan, 2000: From near-surface to root-zone soil moisture using year-round data. *J. Hydrometeorol.*, **1**, 393–411.
- , and J. F. Soussana, 2001: Modelling CO<sub>2</sub> enrichment effects using an interactive vegetation SVAT scheme. *Agric. For. Meteorol.*, **108**, 129–152.
- , J. Noilhan, and P. Bessemoulin, 1998a: Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: A feasibility study based on field measurements. *J. Appl. Meteorol.*, **37**, 371–386.
- , —, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J.-P. Wigneron, 1998b: An interactive vegetation SVAT model tested against data from six contrasting sites. *Agric. For. Meteorol.*, **92**, 73–95.
- , and Coauthors, 1999: MUREX: A land-surface field experiment to study the annual cycle of the energy and water budgets. *Ann. Geophys.*, **17**, 838–854.
- Deardorff, J. W., 1977: A parameterization of ground-surface moisture content for use in atmospheric prediction models. *J. Appl. Meteorol.*, **16**, 1182–1185.
- , 1978: Efficient prediction of ground temperature and moisture with inclusion of a layer of vegetation. *J. Geophys. Res.*, **83**, 1889–1903.
- De Rosnay, P., and Coauthors, 2006: SMOSREX: A long term field campaign experiment for soil moisture and land surface processes remote sensing. *Remote Sens. Environ.*, **102**, 377–389.
- Dirmeyer, P. A., 2000: Using a global soil wetness dataset to improve seasonal climate simulation. *J. Climate*, **13**, 2900–2922.
- Douville, H., P. Viterbo, J.-F. Mahfouf, and A. C. M. Beljaars, 2000: Evaluation of the optimal interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Wea. Rev.*, **128**, 1733–1756.
- Eagleman, J. R., and W. C. Lin, 1976: Remote sensing of soil moisture by a 21 cm passive radiometer. *J. Geophys. Res.*, **81**, 3660–3666.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
- François, C., A. Quesney, and C. Ottlé, 2003: Sequential assimilation of ERS-1 SAR data into a coupled land surface–hydrological model using an extended Kalman filter. *J. Hydrometeorol.*, **4**, 473–487.
- Giard, D., and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997–1015.
- Gibelin, A.-L., J.-C. Calvet, J.-L. Roujean, L. Jarlan, and S. Los, 2006: Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products. *J. Geophys. Res.*, **111**, D18102, doi:10.1029/2005JD006691.
- Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich, 1998: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resour. Res.*, **34**, 3405–3420.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.*, **82**, 35–45.
- Kerr, Y. K., P. Waldteufel, J.-P. Wigneron, J. M. Martinuzzi, J. Font, and M. Berger, 2001: Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1729–1735.
- Koster, R. D., and M. J. Suarez, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeorol.*, **4**, 408–423.
- Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167–194.
- Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, 2003: Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sens.*, **41**, 215–229.
- Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549.
- , and J.-F. Mahfouf, 1996: The ISBA land surface parameterization scheme. *Global Planet. Change*, **13**, 145–159.
- Reichle, R. H., and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeorol.*, **4**, 1229–1242.
- , J. P. Walker, R. D. Koster, and P. R. Houser, 2002: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.*, **3**, 728–740.
- Seuffert, G., H. Wilker, P. Viterbo, J.-F. Mahfouf, M. Drusch, and J.-C. Calvet, 2003: Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data. *Geophys. Res. Lett.*, **30**, 1498, doi:10.1029/2003GL017128.
- Shukla, J., and Y. Mintz, 1982: Influence of land-surface evapotranspiration on the Earth's climate. *Science*, **215**, 1498–1501.
- Teunissen, P. J. G., 2000: *Adjustment Theory: An Introduction*. Delft University Press, 194 pp.
- Wagner, W., K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf, 2003: Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *J. Geophys. Res.*, **108**, 4611–4626.
- Wigneron, J.-P., A. Chanzy, J.-C. Calvet, A. Olioso, and Y. Kerr, 2002: Modeling approaches to assimilating L-band passive microwave observations over land surfaces. *J. Geophys. Res.*, **107**, 4219, doi:10.1029/2001JD000958.