Comparing the Ensemble and Extended Kalman Filters for in situ soil moisture assimilation with contrasting conditions

D. Fairbairn, A. L. Barbu, J.-F. Mahfouf, J.-C. Calvet, and E. Gelati

CNRM-GAME, UMR 3589 (Météo-France, CNRS), Toulouse, France

Correspondence to: J.-C. Calvet (jean-christophe.calvet@meteo.fr)

Abstract. Two data assimilation (DA) methods are compared for their ability to produce an accurate soil moisture analysis on the Météo-France land surface model: (i) SEKF, a Simplified Extended Kalman Filter, which uses a climatological background-error covariance, (ii) EnSRF, the Ensemble Square Root Filter, which uses an ensemble background-error covariance and approximates random rainfall errors stochastically. In situ soil moisture observations at 5 cm depth are assimilated into the surface layer and 30 cm deep observations are used to evaluate the root-zone analysis on 12 sites in Southwest France (SMOSMANIA network). These sites differ in terms of climate and soil texture. The two methods perform similarly and improve on the open loop. Both methods suffer from incorrect linear assumptions which are particularly degrading to the analysis during water-stressed conditions: the EnSRF by a dry bias and the SEKF by an over-sensitivity of the model Jacobian between the surface and the root zone layers. These problems are less severe for the sites with wetter climates. A simple bias correction technique is tested on the EnSRF. Although this reduces the bias, it modifies the soil moisture fluxes and suppresses the ensemble spread, which degrades the analysis performance. However, the EnSRF flow-dependent background-error covariance evidently captures seasonal variability in the soil moisture errors and should exploit planned improvements in the model physics.

Synthetic twin experiments demonstrate that when there is only a random component in the precipitation forcing errors, the correct stochastic representation of these errors enables the EnSRF to perform better than the SEKF. It might therefore be possible for the EnSRF to perform better than the SEKF with real data, if the rainfall uncertainty was accurately captured. However, the simple rainfall error model is not advantageous in our real experiments. More realistic rainfall error models are suggested.

1 Introduction

Soil moisture has a significant influence on heat and water exchanges between the land and the
atmosphere, which makes it an important factor in Numerical Weather Prediction (NWP) (Dharssi
et al., 2011). It is also useful for a variety of other applications, including drought monitoring, crop
irrigation and water management.

An important application of data assimilation (DA) for land surface models is to assimilate ob-
served surface soil moisture to produce an analysis of root-zone soil moisture. Root-zone soil mois-
ture is usually of more interest than surface soil moisture because it has a much greater water capacity
and a far longer memory. The interest in soil moisture DA is partly driven by the wealth of satellite
data available from low-frequency microwave instruments, which can be used to retrieve global-
scale surface observations. Recent satellite launches have considerably improved coverage over the
last decade, namely the Advanced Scatterometer (ASCAT) instrument on board the METOP satel-
lites (Wagner et al., 2007), the Soil Moisture and Ocean Salinity (SMOS) Mission (Kerr et al., 2001)
and the Soil Moisture Active Passive (SMAP) Mission (Entekhabi et al., 2010). However, these in-
struments are subject to significant retrieval errors and can only measure the top few centimetres
of the soil. DA methods are used to account for the errors in the model and the observations. They
also distribute information from the surface observations to the deeper layers. In situ observations of
root-zone soil moisture do exist but are not dense enough over large areas.

A Simplified Extended Kalman Filter (SEKF, Mahfouf et al., 2009) is embedded in the surface
externalized (SURFEX) modelling platform of Météo-France (Masson et al., 2013). The SEKF sim-
plifies the Extended Kalman Filter (EKF) by assuming that errors in the prior state (the background)
are climatological (i.e. there is no flow-dependence in the errors) and uncorrelated between layers
and gridpoints. The use of a linear model in the analysis update does extract information from the
surface to the root zone. The SEKF is not yet coupled with an NWP model at Meteo-France. Instead
it is used to provide soil moisture analyses and carbon fluxes for a variety of other applications, in-
cluding hydrological models (see e.g. Draper et al., 2011), and the European Copernicus programme
(http://www.copernicus.eu/). An SEKF is currently coupled with the NWP model of the European
Centre for Medium range Weather Forecasts (ECMWF) for the assimilation of screen-level vari-
ables (de Rosnay et al., 2013) and for the assimilation of ASCAT soil moisture observations since
April 2015 (Patricia de Rosnay, personal communication). Ensemble DA methods are becoming in-
creasingly popular for land surface models (see e.g. Reichle et al., 2002; Zhou et al., 2006; Muñoz
Sabater et al., 2007; Reichle et al., 2008; Draper et al., 2012; Carrera et al., 2015). There are four
main reasons at Météo-France for developing an Ensemble Kalman Filter (EnKF) for soil moisture
assimilation: firstly, the EnKF uses a flow-dependent estimate of errors in the background, rather
than a climatological estimate. Secondly, the EnKF can stochastically represent random forcing and
model errors, which is not feasible with an SEKF. Thirdly, a diffusion-based multi-layer model
(Decharme et al., 2011) has been developed to replace the current 3 layer force-restore land surface

60 model. Increasing the number of layers would substantially increase the cost of the SEKF Jacobian calculations, which require a model integration for each prognostic variable. Finally, the EnKF may take into account background-error covariances between gridpoints, although each gridpoint is assumed to be independent in this study.

The EnKF has extensively been compared with the EKF on land surface models for assimilating
65 soil moisture observations. Experiments have been conducted with both synthetic observations (e.g. Reichle et al. (2002)) and real observations (e.g. Muñoz Sabater et al. (2007)). In most cases the EnKF delivered modest improvements over the EKF. It was not obvious beforehand which of these methods would perform better, since they both make incorrect linear assumptions in the analysis update: the EKF by using a linear model and the EnKF by using a linear combination of ensemble
70 members.

The experiments in this paper are partly motivated by studying the results of the experiments by Muñoz Sabater et al. (2007); Draper et al. (2009); Mahfouf et al. (2009). They performed comparisons of the SEKF and the EKF on previous versions of the land surface model used by Météo-France. They found that not only is the SEKF less computationally expensive than the EKF, but that
75 it's performance is slightly better. Muñoz Sabater et al. (2007) also demonstrated that a simple 1-D variational DA method (1DVar) (theoretically similar to the SEKF) performs similarly to an EnKF with a large ensemble size (≈ 200 members). Their results may seem counter-intuitive since the SEKF/1DVar methods cannot represent flow-dependent background errors, which can be estimated by the EKF and the EnKF. However, as the authors ackowledge, when they implemented the EKF
80 and EnKF they did not use accurate specifications of model errors and forcing errors. The incorrect specification of these errors leads to sub-optimal filter performance (Crow and Van Loon, 2006; Reichle et al., 2008; Crow and Van den Berg, 2010).

Various formulations of the EnKF exist, which differ in the way they perform the analysis update. This study examines an implementation called the Ensemble Square Root Filter (EnSRF, Whitaker
85 and Hamill, 2002). The EnSRF is chosen because it does not perturb the observations, which would incur sampling errors. In this paper, the EnSRF is compared with the SEKF in terms of their ability to provide an accurate deterministic soil moisture analysis. The aim of this study is to compare and analyze the performances of these DA methods by examining the impact of:

   i. random errors in the precipitation forcing;

90   ii. the Gaussian assumptions made by the DA methods;

  iii. influences of climate and soil texture;

  iv. a flow-dependent background-error covariance.

A linear rescaling technique is used in this study, which bias corrects the observations with respect to the model simulation (Calvet and Noilhan, 2000; Scipal et al., 2008). However, ensemble pertur-
95 bations can introduce additional biases as a result of the nonlinear water fluxes (Ryu et al., 2009).

3

A simple bias correction technique is also tested on the EnSRF as a means of reducing the biases caused by ensemble perturbations (Ryu et al., 2009).

Twelve grassland sites over southwest France, where in situ observations are available (the SMOS-MANIA network, Calvet et al., 2007; Albergel et al., 2008), are used to compare the methods. These sites include various climates and soil textures that can influence soil water transfers. In situ 5 cm deep soil moisture observations are assimilated into the surface layer. The performance is validated by comparing the root-zone soil moisture analysis (80 cm depth) with 30 cm deep in situ observations. The results are collected over a 3 year period (2008–2010).

The methods and materials are described in Sect. 2. In Sect. 3.1, the results of the experiments without DA are presented. The objective here is to show the physical mechanisms behind the ensemble perturbation bias, and the impact of applying a bias correction scheme. In Sect. 3.2 the DA methods are compared using a synthetic "identical twin" experiment designed to represent only random errors in the precipitation forcing. This is a test of the ability of the DA methods to represent these errors. Then in Sect. 3.3 the DA methods are tested using real in situ observations. Sect. 4 discusses the results and Sect. 5 summarises the main conclusions of the experiments.

2 Methods and materials

The ISBA-A-gs model and the atmospheric forcing are introduced in Sections 2.1 and 2.2 respectively. The DA methods are explained in Sect. 2.3. A list of the experiments that were performed is summarised in Sect. 2.4. The experimental setup for the real and synthetic experiments are given in Sect. 2.5 and 2.6 respectively. The experimental setups include a desciption of the observations and the calibration of the error representations of the DA methods. The performance diagnostics of the DA methods are given in Sect. 2.7.

2.1 ISBA-A-gs model

The experiments were all conducted on version 7.2 of SURFEX, which incorporates the "Interactions between Soil, Biosphere and Atmosphere" (ISBA) land surface model (Noilhan and Mahfouf, 1996). This model is based on the force-restore method of Deardorff (1977). The A-gs version of ISBA accounts for leaf-scale physiological processes, including the effects of carbon dioxide concentration and photosynthesis (Calvet et al., 1998). The simulated leaf biomass is used to compute the leaf area index (LAI), a key variable governing plant transpiration. The "NIT" version of the model is applied in this work, which can dynamically simulate LAI evolution (Gibelin et al., 2006). The seasonal variability in LAI has a significant impact on the soil moisture variables (Barbu et al., 2011). The three-layer version of ISBA-A-gs is used in this study (Boone et al., 1999). The three soil moisture variables are defined here with the depths used for the experiments:

4

- The surface soil moisture (WG1), with depth d_1 of 1 cm. But the effective depth is d_1/C_1, where C_1 is the surface restore coefficient, which depends on soil texture and soil moisture;

- The root-zone soil moisture (WG2), with depth d_2 of 0.8 m, which includes WG1;

- A recharge layer (WG3) with thickness d_3 of 0.2 m, which exists below WG2 (see Fig. 1).

All the variables are measured in terms of volumetric soil moisture concentration ($\mathrm{m^3\,m^{-3}}$). A diagram illustrating the soil moisture fluxes is presented in Fig. 1. The surface layer (WG1) and the root zone (WG2) layers are forced by interactions with the atmosphere and restored towards an equilibrium value. At equilibrium, the gravity forces match the capillary forces. The drainage from WG2 supplies water into a recharge zone (WG3), which conserves the total water volume.

In these experiments 12 model points were used, which are the closest points to the 12 grassland in situ observation sites (introduced in Sect. 2.5). The model points were represented by the grassland vegetation type and were located such that the nearest observation to each site was always less than 6 km away.

2.2 Forcing

The "Système d'Analyse Fournissant des Renseignements à la Neige" (SAFRAN) forcing was used, which is derived from a meso-scale analysis system with a horizontal resolution of 8 km (Durand et al., 1993). This provides values of precipitation, wind, incoming short-wave and long-wave radiation, relative humidity and air temperature, mostly derived from a surface network of weather stations. The hourly forcing values were input into the ISBA-A-gs model for the 12 gridpoints. We have adopted a version of SAFRAN that enables the additional use of 3000 climatological observing stations over France, including rain gauges (Quitana-Ségui et al., 2008; Vidal et al., 2010).

2.3 DA methods

The DA methods employed in this work are derived from the Kalman Filter (Kalman, 1960). The vector of prognostic variables is $\boldsymbol{x} = (\mathrm{WG1},\ \mathrm{WG2})$. The background state ($\boldsymbol{x}^{\mathrm{b}}(t_i)$) is a nonlinear model propagation of the previous analysis:

$$\boldsymbol{x}^{\mathrm{b}}(t_i) = M_{i-1}(\boldsymbol{x}^{\mathrm{a}}(t_{i-1})), \tag{1}$$

where M is the (nonlinear) ISBA-A-gs model. The analysis and the observation time (t_i) is at the end of the 24 hour assimilation window. The Kalman Filter analysis update is:

$$\boldsymbol{x}^{\mathrm{a}}(t_i) = \boldsymbol{x}^{\mathrm{b}}(t_i) + \mathbf{K}_i(\boldsymbol{y}_i^{\mathrm{o}} - \boldsymbol{y}_i), \tag{2}$$

where $\boldsymbol{y}^{\mathrm{o}}$ is the assimilated observation and $\boldsymbol{y}_i = H(\boldsymbol{x}^{\mathrm{b}}(t_i))$ is the model predicted value of the observation ($H = (1,0)$ in our case). The model state and the observations are weighted using the

Kalman gain ($\mathbf{K}$):

$$\mathbf{K}_i = \mathbf{B}_i \mathbf{H}_i^{\mathrm{T}} (\mathbf{H}_i \mathbf{B}_i \mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i)^{-1}, \tag{3}$$

where $\mathbf{H}$ is the linearized observation operator, $\mathbf{B}$ is the background-error covariance matrix and $\mathbf{R}$ is the observation-error covariance matrix. These matrices measure the expected errors and the co-variances are a measure of the correlations in the errors between the different variables (i.e. between WG1 and WG2). The $\mathbf{R}$ matrix is assumed to be diagonal i.e. covariances equal to zero. It is also assumed that each gridpoint is independent i.e. the background-error covariance is assumed to be zero between gridpoints. The observation is present at the end of the assimilation window, so the background-error covariance needs to be propagated from the beginning to the end of the window. This is implicitly calculated via $\mathbf{H}$ for the SEKF (Sect. 2.3.1) and via the ensemble perturbations for the EnSRF (Sect. 2.3.2). A summary of the DA methods is given in Table 1.

2.3.1 SEKF

The Simplified Extended Kalman Filter (SEKF, Mahfouf et al., 2009) is based on the EKF (Jazwin-ski, 1970). The SEKF simplifies the EKF by using both a diagonal and climatological background-error covariance at the start of the assimilation window.

The SEKF was originally designed to assimilate screen-level temperature and humidity, which are not prognostic variables and therefore cannot be assimilated directly (Hess, 2001; Balsamo et al., 2007). For this reason the SEKF uses the linear observation operator $\mathbf{H}$ to relate the observed quantity to the prognostic variables. Following the notation of Mahfouf (2010), there are two steps in the calculation of $\mathbf{H}$. The first step $\mathbf{H}_1$ is simply a transformation into observation space ($\mathbf{H}_1 = (1,0)$ in our case). The second step is the calculation of the Tangent linear version of the nonlinear model ($\mathbf{M}$). This linear model is a finite difference approximation between a perturbed and reference nonlinear model simulation:

$$\mathbf{M}_{i-1}^l = \frac{M_{i-1}(\boldsymbol{x}(t_{i-1}) + \Delta x_{i-1}^l) - M_{i-1}(\boldsymbol{x}(t_{i-1}))}{\Delta x_{i-1}^l}, \tag{4}$$

where Δx^l is a model perturbation applied to model variable l. Therefore the Jacobian between the observation k and the model variable l is simply:

$$\mathbf{H}_i^{kl} = \mathbf{H}_1^k \mathbf{M}_{i-1}^l. \tag{5}$$

Eq. (4) requires a model integration for each prognostic variable. This formulation of $\mathbf{H}$ implicitly propagates the $\mathbf{B}$ matrix from the start of the assimilation window to the time of the observation at the end of the window ($\mathbf{H}_i \mathbf{B}_i \mathbf{H}_i^{\mathrm{T}} = \mathbf{H}_1 \mathbf{M}_{i-1} \mathbf{B}_{i-1} \mathbf{M}_{i-1}^{\mathrm{T}} \mathbf{H}_1^{\mathrm{T}}$). Although screen-level temperature and humidity observations are not assimilated in this study, the same formulation is applied to soil moisture observations.

The Δx^l size is important – it needs to be sufficiently small that the linear approximation in deriving $\mathbf{M}$ is satisfactory but large enough to not incur significant computational round-off errors.

A measure of the nonlinearity of the nonlinear model (M) can be calculated by the magnitude of the difference between $\mathbf{H}^{kl}$ for positive and negative values of Δx^l (Walker and Houser, 2001; Balsamo et al., 2004; Draper et al., 2009), with values close to zero indicating quasi-linear model behaviour. Draper et al. (2009); Duerinckx (2015) have demonstrated that over 6 and 24 hour windows, the behaviour of the model is generally quasi-linear for other versions of ISBA. However, in both studies there were occasions when the model behaviour was highly nonlinear and in these situations the tangent linear approximation is inadequate. Following Draper et al. (2009), we use a perturbation size of $10^{-4}(w_{\text{fc}} - w_{\text{wilt}})$ and a 24 h assimilation window. Draper et al. (2009) demonstrated that during dry periods the $\mathbf{H}^{12}$ values can become much larger than 1.0, which is unrealistic and indicates highly nonlinear behaviour. For this reason we imposed a maximum limit on $\mathbf{H}^{12}$ of 1.0.

The validity of the tangent linear approximation was not tested explicitly in this study. Instead, the WG2 Kalman gain was compared before and after imposing the 1.0 Jacobian limit. It was clear when the linear assumption broke down because the WG2 Kalman gain was noticeably reduced by imposing the limit. The WG2 Kalman gain is defined by:

$$
\begin{aligned}
\mathbf{K}_{\text{WG2}} &= \mathbf{H}^{12}\mathbf{B}(\mathbf{HBH}^{\text{T}} + \mathbf{R})^{-1} \\
&= \mathbf{H}^{12}\left(\sigma^{\text{b}}_{\text{WG2}}\right)^2 \left(\left(\mathbf{H}^{11}\sigma^{\text{b}}_{\text{WG1}}\right)^2 + \left(\mathbf{H}^{12}\sigma^{\text{b}}_{\text{WG2}}\right)^2 + (\sigma^{\text{o}})^2\right)^{-1},
\end{aligned}
\tag{6}
$$

where σ^{o} and σ^{b} are the expected observation and background errors. The $\mathbf{R}$ matrix in our study is equal to the scalar $(\sigma^{\text{o}})^2$.

2.3.2 EnSRF

The EnKF (Evensen, 1994) is a way of representing the uncertainty in the prognostic variables by using an ensemble of model trajectories. This circumvents the high computational cost of explicitly storing and propagating the background-error covariance for a large model dimension. First, each ensemble member is propagated using the nonlinear model:

$$
\boldsymbol{x}^{\text{b}}_j(t_i) = M_{i-1}(\boldsymbol{x}^{\text{a}}_j(t_{i-1})), \text{ for j} = 1,...,\text{m},
\tag{7}
$$

where m is the ensemble size and j is the ensemble member The following steps then occur at the analysis time (t_i). The ensemble background-error covariance is defined as:

$$
\mathbf{P}^{\text{b}} = \frac{1}{m-1}\mathbf{X}^{\text{b}}(\mathbf{X}^{\text{b}})^{\text{T}}.
\tag{8}
$$

and the perturbation matrix (of dimension $n \times m$) is given by:

$$
\mathbf{X}^{\text{b}} = \frac{1}{\sqrt{m-1}}\left[\delta\boldsymbol{x}^{\text{b}}_1 \quad ... \quad \delta\boldsymbol{x}^{\text{b}}_m\right],
\tag{9}
$$

where $\delta\boldsymbol{x}^{\text{b}}_j = \boldsymbol{x}^{\text{b}}_j - \overline{\boldsymbol{x}}^{\text{b}}$ are the perturbations from the ensemble mean $(\overline{\boldsymbol{x}}^{\text{b}})$ and n is the model state dimension. Note that $\mathbf{B}$ in Eq. (3) is expressed as $\mathbf{P}^{\text{b}}$ for the EnSRF. The deterministic analysis is

7

calculated from the ensemble mean:

$$\overline{x}^{\mathrm{a}} = \overline{x}^{\mathrm{b}} + \mathbf{K}(y^{\mathrm{o}} - \overline{y}) \tag{10}$$

An additional step is required to avoid ensemble collapse. The traditional EnKF of Burgers et al. (1998) maintains the ensemble spread by perturbing the observations. The serial ensemble square root filter (EnSRF) was introduced by Whitaker and Hamill (2002) as a means of avoiding the sampling error from the perturbed observations. The ensemble perturbations are updated by:

$$x_j^{\mathrm{a}} = \overline{x}^{\mathrm{a}} + \delta x_j^{\mathrm{b}} - \alpha \mathbf{K} y_j, \ \text{for j =1,...,m,} \tag{11}$$

where

$$\alpha = \frac{1.0}{1.0 + \sqrt{\frac{R}{HP^{\mathrm{b}}H^{\mathrm{T}} + R}}}. \tag{12}$$

Here R and $HP^{\mathrm{b}}H^{\mathrm{T}}$ are scalars representing the variances at the observation location. The variable of interest WG2 is linked to WG1 via the Kalman gain:

$$\mathbf{K}_{\mathrm{WG2}} = \mathbf{X}_{\mathrm{WG1}}^{\mathrm{b}}(\mathbf{X}_{\mathrm{WG2}}^{\mathrm{b}})^{\mathrm{T}} \left(\mathbf{X}_{\mathrm{WG1}}^{\mathrm{b}}(\mathbf{X}_{\mathrm{WG1}}^{\mathrm{b}})^{\mathrm{T}} + (\sigma^{\mathrm{o}})^2 \right)^{-1}, \tag{13}$$

where $\mathbf{X}_{\mathrm{WG1}}\mathbf{X}_{\mathrm{WG2}}^{\mathrm{T}}$ represents the cross-product between the WG1 and WG2 ensemble perturbations. The WG1 and WG2 ensemble spreads are defined by $\mathbf{X}_{\mathrm{WG1}}^{\mathrm{b}}(\mathbf{X}_{\mathrm{WG1}}^{\mathrm{b}})^{\mathrm{T}}$ and $\mathbf{X}_{\mathrm{WG2}}^{\mathrm{b}}(\mathbf{X}_{\mathrm{WG2}}^{\mathrm{b}})^{\mathrm{T}}$ respectively.

2.4 Experiment list

The experimental setup for the DA experiments is summarised in Table 2. This includes the representation of the truth, the observations, and the data used for evaluation. A description of the real and synthetic experiments with DA are given in Sections 2.5 and 2.6 respectively.

A summary of the calibrated values and the results of each experiment are given in Table 3. The first experiment (Ens) was performed by perturbing an ensemble without DA in order to investigate the cause of the perturbation bias. The bias correction scheme (Eq. 16) was then tested on this ensemble, which is labelled as Ens$_{\mathrm{bc}}$. Thereafter the synthetic and real DA experiments are denoted by the subscripts S and R respeively. For each experiment the calibrated error variances are specified. For the real experiments the EnSRF was tested with three different configurations: EnSRF$_{\mathrm{R1}}$ is the baseline EnSRF without perturbed precipitation forcing nor bias correction. The EnSRF$_{\mathrm{R2}}$ and EnSRF$_{\mathrm{R3}}$ experiments include perturbed precipitation forcing and bias correction respectively. Bias correction was not performed in the synthetic experiments. Note that in the synthetic experiments, the EnSRF$_{\mathrm{S}}$ was designed to capture the precipitation forcing uncertainty perfectly. The same precipitation error specification was used to estimate the precipitation errors in the real experiments (EnSRF$_{\mathrm{R2}}$).

2.5 Experimental setup: DA with real observations

2.5.1 Observations

For the experiments with real observations, in situ observations at 12 grassland sites in Southwest France were assimilated. These experiments are hereafter referred to as "real experiments". These grassland sites are part of the Soil Moisture Observing System Meteorological Automatic Network Integrated Application (SMOSMANIA) network (Calvet et al., 2007; Albergel et al., 2008). A map of the sites is shown in Fig. 2. Also included in Fig. 2 is a bar chart of the average daily precipitation for the sites (from west to east). The observations are spaced approximately 45 km apart. The SMOSMANIA network was selected partly because of the large variability in the climatologies and soil textures between the different sites. The more westerly sites are generally wetter as they are more exposed to weather systems arriving from the Atlantic ocean. The most westerly site (Sabres) receives more than twice the average rainfall of the most easterly site (Narbonne), with a mean daily rainfall of 3.5 mm, compared with 1.5 mm for Narbonne. Sabres also has a smaller clay to sand ratio than Narbonne. For the interested reader, Table S1.1 of Supplement 1 shows modelled and observed sand and clay percentages for each site. The larger sand particles have a smaller aggregate surface area than the finer clay particles, which means that sand can hold less water than clay. This leads to faster water transfers for sandier soils. This study examines the influences of these factors on the performance of the DA methods.

The 5 cm deep observations were simulated by WG1 and were assimilated daily at the end of a 24 hour assimilation window (06 UTC). The WG2 variable was represented by in situ observations at 30 cm depth, which were used to evaluate the performance of the DA methods. It is possible to inter-compare in situ observations and model simulations, provided the observations are a good representation of the depth of the layer. The observations and the model were well correlated for our experiments. The average anomaly correlation coefficients (ACC, defined in Sect. 2.7) for the sites between the simulated and observed WG1 (WG2) variable is 0.53 (0.68). Table S1.1 in Supplement 1 shows the WG1 and WG2 ACC for each site.

A linear rescaling technique was employed in this study, which scales the observations such that the mean and the variance match that of the model (Calvet and Noilhan, 2000; Scipal et al., 2008). The rescaling was performed over 2007-2010. The results were calculated over 2008-2010 because a one year spin-up was used in the experiments. The small bias that remained, as a result of the different time periods, is not significant. There were significant biases in the in situ observations (relative to the model simulation) prior to rescaling. The rescaling reduced the site-averaged WG1 (WG2) RMSE between the model simulation and the observations from 0.089 (0.062) $\mathrm{m^3\,m^{-3}}$ to 0.060 (0.025) $\mathrm{m^3\,m^{-3}}$, without changing the ACC.

It was necessary to estimate the observation errors for the DA experiments. All the ThetaProbe sensors used to measure soil moisture in the SMOSMANIA network were calibrated using gravi-

9

metric reference observations (Calvet et al., 2007). Overall, the RMSE after calibration was $0.038\,\mathrm{m}^3\,\mathrm{m}^{-3}$. This error includes instrumental errors and local representativeness errors (gravimetric samples were collected a few meters around the probes). However, other representativeness errors were also likely in this study, since the point observations were assumed to represent $8\times8\,\mathrm{km}^2$ model pixels. The observation error standard deviation (σ^o) for WG1 was tested with values of $\lambda^\mathrm{o}(w_\mathrm{fc} - w_\mathrm{wilt})\,\mathrm{m}^3\,\mathrm{m}^{-3}$, with λ^o equal to 0.5 or 0.35. The variable λ^o is a dimensionless scaling coefficient for the observation error. The w_fc and w_wilt parameters are the field capacity and the wilting point respectively, and they depend on the soil texture and vegetation type of each site. With the scaling by $(w_\mathrm{fc} - w_\mathrm{wilt})$ the two values of λ^o correspond to site averaged σ^o values of $0.030\,\mathrm{m}^3\,\mathrm{m}^{-3}$ and $0.044\,\mathrm{m}^3\,\mathrm{m}^{-3}$, which are either side of the RMSE measured by Calvet et al. (2007). The scaling by $(w_\mathrm{fc} - w_\mathrm{wilt})$ is based on the assumption of a linear relationship between the dynamic range of soil moisture values and the errors (Mahfouf et al., 2009).

2.5.2 SEKF calibration

The representation of errors by the DA methods can significantly influence their performance (Reichle et al., 2008). The SEKF uses a climatological background-error covariance matrix. This matrix theoretically captures the total contribution from background errors and additive model/forcing errors. In this study, the SEKF background-error standard deviations (σ^b) for WG1 and WG2 were tuned to produce the best ACC, with sizes:

$$\sigma^\mathrm{b}_\mathrm{WG1} = \lambda^\mathrm{b}_1 \times (w_\mathrm{fc} - w_\mathrm{wilt}),$$
$$\sigma^\mathrm{b}_\mathrm{WG2} = \lambda^\mathrm{b}_2 \times (w_\mathrm{fc} - w_\mathrm{wilt}), \tag{14}$$

In the real experiments, λ^b_1 and λ^b_2 were tuned between 0.0 and 0.5, in steps of 0.05. The variable λ^b is a dimensionless scaling coefficient for the calibration of the background errors. The background error variances were scaled by $(w_\mathrm{fc} - w_\mathrm{wilt})$ for each site.

2.5.3 EnSRF calibration

The background errors in the EnSRF are implicitly captured by the ensemble spread. Hamill and Whitaker (2005) demonstrated that the addition of random perturbations to the model state (additive inflation) at the start of each cycle can be used to represent model errors. For the real experiments, Gaussian noise with zero mean and standard deviation ϵ was added to each ensemble member after the daily analysis update. The values of ϵ for WG1 and WG2 were tuned to produce the largest ACC:

$$\epsilon_\mathrm{WG1} = \lambda^\mathrm{b}_1 \times (w_\mathrm{fc} - w_\mathrm{wilt}),$$
$$\epsilon_\mathrm{WG2} = \lambda^\mathrm{b}_2 \times (w_\mathrm{fc} - w_\mathrm{wilt}), \tag{15}$$

with λ^b_1 varying between 0.0 and 0.5, in steps of 0.05, and λ^b_2 varying between 0.0 and 0.05, in steps of 0.005. Time-correlated additive inflation was implemented using a 1st order auto-regressive

325 model. It was decided to use time correlations of $\tau = 1$ day for ϵ_{WG1}, and $\tau = 3$ days for ϵ_{WG2}. This is similar to previous studies (Reichle et al., 2002; Mahfouf, 2007) and is consistent with the longer time correlations of the WG2 variable compared with WG1. An ensemble size of 20 members was chosen for the calibration of the additive inflation. The calibrated EnSRF was then tested with ensemble sizes ranging from 3 to 200 in order to explore the effects of sampling errors.

330 The EnSRF ensemble used hourly perturbations of the precipitation forcing in order to estimate the precipitation uncertainty. The same perturbations were also used to capture precipitation errors in the synthetic experiment and this technique is described in Sect. 2.6. The performance of the EnSRF with perturbed precipitation was also compared with the performance without perturbed precipitation.

335 2.5.4 EnSRF bias correction

A bias correction technique was tested on the EnSRF as a means of correcting the biases caused by the ensemble perturbations themselves (Ryu et al., 2009). The perturbation bias correction uses the forecast from the previous analysis ensemble mean as an anchor to modify the background ensemble:

$$\tilde{\boldsymbol{x}}_j^{\mathrm{b}}(t_i) = \boldsymbol{x}_j^{\mathrm{b}}(t_i) - \overline{\boldsymbol{x}}_j^{\mathrm{b}}(t_i) + M(\overline{\boldsymbol{x}}^{\mathrm{a}}(t_{i-1})), \quad \text{for } j = 1, \ldots, m. \tag{16}$$

340 where $\tilde{\boldsymbol{x}}_j^{\mathrm{b}}$ is the bias-corrected $\boldsymbol{x}_j^{\mathrm{b}}$. Equation (16) prevents the mean of the ensemble forecasts from becoming biased with respect to the forecast of the analysis ensemble mean. The perturbation bias correction was implemented on all three layers before the analysis update step.

2.6 Experimental setup: DA with synthetic observations

For the synthetic experiments the in situ observations were not used, although the model was used 345 for the same 12 sites. The truth was generated from a single model simulation. The WG1 observations were extracted from the truth with the addition of a random normally distributed observation error with zero mean and standard deviation equal to to 10% of the higher value used in the real experiments ($\sigma^{\mathrm{o}} = 0.05(w_{\mathrm{fc}} - w_{\mathrm{wilt}})$). The size of the observation error was small enough for the observations to have a noticeable impact on the analysis. We also tested an observation error equiv-350 alent to the value in the real experiments and the impact on the analysis was about 10 times less (not shown).

A perfect model was used for the DA. However, errors were introduced in the precipitation forcing by adding random hourly perturbations ϵ_{Pr} to the hourly precipitation accumulations (Pr):

$$Pr* = Pr + \epsilon_{\mathrm{Pr}}, \tag{17}$$

355 where $Pr*$ is the perturbed hourly precipitation. The hourly perturbations were randomly sampled from a normal distribution with standard deviation equal to 50% of the hourly precipitation and zero mean. The probability distribution function (pdf) was truncated in order to prevent negative values

of $Pr*$ and to maintain a mean of zero ($-Pr \leq \epsilon_{\text{Pr}} \leq Pr$). A single precipitation time series from Eq. (17) was generated over 2007-2010. This was used to force the model. A standard deviation

360 of 50% of the hourly precipitation was judged as an appropriate order of magnitude by Reichle et al. (2002), by comparing the magnitude of the resulting errors with the difference between two precipitation data sets. Other forcing parameters were not perturbed, since it was found through a sensitivity study that perturbing these values had little impact on the model simulations. The results for the sensitivity study are presented in Table S1.2 of Supplement 1.

365 The SEKF has no means of capturing the uncertainty in the precipitation forcing from Eq. (17) directly. Therefore it was necessary to calibrate the **B** matrix to capture the background errors that resulted from the precipitation errors. The background-error variances (σ^{b}) were calibrated with values a 10th of the values used for the real experiments (Eq. (14)), since the open loop errors in the synthetic experiments were about 10% of the errors in the real experiments.

370 In the synthetic experiments, the EnSRF ensemble members used the single precipitation time series from Eq. (17) to force the model. However, each ensemble member was then perturbed using Eq. (17) with different random seeds for each member. This enabled the EnSRF to capture the uncertainty in the precipitation forcing directly.

2.7 Performance diagnostics

375 The Root Mean Square Error (RMSE), the Anomaly Correlation Coefficient (ACC) and the Bias for the WG2 variable were used to determine the performance of the DA methods:

- RMSE: $\sqrt{\frac{\sum_{i=0}^{N}(\boldsymbol{y}_i^{\text{e}}-H(\boldsymbol{x}(t_i)))^2}{N}}$;

- ACC $= \frac{\sum_{i=0}^{N}(H(\boldsymbol{x}(t_i))'-C(H(\boldsymbol{x})'))((\boldsymbol{y}_i^{\text{e}})'-C((\boldsymbol{y}^{\text{e}})'))}{\sqrt{\sum_i(H(\boldsymbol{x}(t_i))'-C(H(\boldsymbol{x})'))^2((\boldsymbol{y}_i^{\text{e}})'-C((\boldsymbol{y}^{\text{e}})'))^2}}$;

- Bias: $\sum_{i=0}^{N}\frac{\boldsymbol{y}_i^{\text{e}}-H(\boldsymbol{x}(t_i))}{N}$.

380 The time t_i is the daily time, with t_0 equal to 01/01/2008 and t_N equal to 31/12/2010. The climatological (time) mean of $\boldsymbol{x}$ is defined as $C(\boldsymbol{x})$. The observations used for evaluation are defined as ($\boldsymbol{y}^{\text{e}}$). For the real experiments these are in situ observations of WG2 after rescaling. In the synthetic experiments these are the true values of WG2. The anomaly $\boldsymbol{x}(t_i)'$ is taken as the difference between the variable and a 3-month moving average.

385 The RMSE is a measure of both the random and systematic components of the error. The ACC represents the seasonal correlations, which are unaffected by systematic errors, while the bias measures the systematic errors. The computational cost of the DA methods was not assessed because the ensemble DA methods were not optimized to take advantage of parallel computing. Furthermore, Muñoz Sabater et al. (2007) already estimated the computational cost of similar algorithms

390 on a previous version of ISBA-A-gs. They found that the main wall-clock time constraints of the EKF and EnKF algorithms were the model simulations, rather than the DA. Indeed, in our study

the SEKF (which requires three simulations for the model Jacobian calculations) did have about the same wall-clock time as the EnSRF with 3 ensemble members.

In the experiments where DA is not applied, the perturbed model simulation is measured against the unperturbed model simulation.

3 Results

3.1 Investigating the perturbation bias (no DA)

An ensemble of model trajectories was perturbed by adding Gaussian perturbations to WG2 with standard deviation $0.025 \mathrm{m^3\,m^{-3}}$ and zero mean. This is a similar order of magnitude to the values used to calibrate the ensemble spread in the real experiments (see Sect. 2.5.3). The result for the perturbed 20 member ensemble is defined as experiment Ens in Table 3. The unperturbed simulation is hereafter referred to as the open loop. The ensemble mean is compared with the open loop using the performance diagnostics defined in Sect. 2.7. A bias of $-4.9 \times 10^{-3} \mathrm{m^3\,m^{-3}}$ has been introduced. The origin of this bias is investigated by linking the physical processes that underpin the bias to changes in the ensemble spread (Sect. 3.1.1). The influences of precipitation and soil clay content on the bias are then explored in Sect. 3.1.2.

3.1.1 Influence of model physics

The site-averaged (averaged over the 12 stations) and time-averaged water content of the total reservoir (WG1 + WG2 + WG3) is $243 \, \mathrm{mm}$ for the open loop and $239 \, \mathrm{mm}$ for Ens. This water loss of $4 \, \mathrm{mm}$ (of Ens compared with the open loop) represents a small fraction of the total reservoir. The impact of the perturbed ensemble on each individual layer is demonstrated by Fig. 3, which shows the monthly, annually and site-averaged net a WG1, b WG2 and c WG3 for Ens and the open loop. The net water amount represents the concentration in $\mathrm{m^3\,m^{-3}}$ for the layer scaled by the depth of the layer (in mm). The dry bias (Ens relative to the open loop) is evident in WG2 during the entire period and it peaks between July and September. There also appears to be a dry bias in the winter in WG3, but there is no significant bias in WG1.

The seasonal water fluxes are investigated in order to explain the seasonal variabilities in the bias. Water is depleted from the reservoir via either drainage, evaporation, transpiration or surface runoff. Surface runoff is neglected in this investigation because it is relatively small compared with the other processes. The site-averaged monthly evapotranspiration (evaporation + transpiration) and drainage are shown in Fig. 3d and 3e respectively. The bare-soil evaporation is most active during summer, which corresponds with the maximum insolation. The transpiration is largest in spring and early summer, when the vegetation is most developed and before the onset of water-stressed conditions in late summer. Transpiration dominates over bare-soil evaporation, since the grassland vegetation type covers 90% of the land surface. These two processes add up to an evapotranspiration curve which

13

peaks in May and June (Fig. 3d). In contrast to evapotranspiration, the drainage is most active during the winter and is absent during the late summer/early autumn period (Fig. 3e); since drainage only occurs when the soil moisture is near the field capacity.

The main effect of the ensemble perturbations (Ens) on evapotranspiration relative to the open loop is an enhancement in July and August and then a reduction in September and October (Fig. 3d). This effect is also clearly evident in Fig. 3f, which shows the total difference in soil water depletion between Ens and the open loop. The effect of perturbing the ensemble on drainage is a slight increase relative to the open loop between February and June (Fig. 3e). The annually averaged discrepancy between the Ens and open loop total soil water depletion is about 4mm, which accounts for the dry bias in the Ens total reservoir.

It is possible to link the seasonal changes in soil water depletion to changes in the ensemble spread. The ensemble members, ensemble mean and the open loop for 2009 are shown for the Sabres site and the Narbonne site in Fig. 4a and b respectively. Also shown are the wilting point and the field capacity for the two sites. The larger water capacity of clay relative to sand explains the greater field capacity and wilting point of Narbonne compared with Sabres. During prolonged wet periods, which tend to occur in winter, the ensemble members converge because the soil reservoir reaches the field capacity. This corresponds to a reduction in the ensemble spread. Between the spring and autumn the largest fluctuations in soil moisture occur due to changes in rainfall and insolation. During this period the soil moisture simulation becomes sensitive to perturbations in the initial conditions, which is reflected by the large WG2 ensemble spread. The Narbonne soil has a much larger ensemble spread than the Sabres soil, particularly in autumn. Separate experiments have confirmed that this is related to the larger precipitation of Sabres compared with Narbonne, which acts to suppress the ensemble spread (see Sect. 3.1.2).

Now the seasonal changes in the bias can be related to changes in the ensemble spread. The Ens WG2 ensemble mean is clearly negatively biased (compared with the open loop) for Narbonne during much of the period (Fig. 4b), most especially when the open loop is near the wilting point during summer and autumn. Near the wilting point the WG2 ensemble spread becomes negatively skewed, which occurs because the negative perturbations remain almost unchanged, but the extra water from the positive perturbations is removed rapidly by transpiration. This is evidenced in Fig. 5a for Narbonne, which for clarity shows only 4 of the 20 ensemble members between June and September 2009. The evapotranspiration for the same members is shown in Fig. 5b. The evapotranspiration is very small for the open loop and for the ensemble members below the wilting point. The members above the wilting point experience strong evapotranspiration. This effect is partly linked to the phenology; under water-stressed conditions the vegetation roots readily absorb excess water that becomes available, which increases the transpiration and the LAI (not shown). The Ens negative bias is larger for Narbonne than Sabres because of the greater ensemble spread for Narbonne (compare Fig. 4a with b).

14

The impact of the ensemble perturbations on drainage is most significant near the field capacity. Fig. 5c and Fig. 5d show the soil moisture and drainage respectively for Narbonne, between February and March 2009. For clarity only 5 of the ensemble members are shown. When the ensemble members are greater than the field capacity then the drainage rapidly increases, which suppresses the ensemble spread. The ensemble members below the field capacity have a drainage near zero. This implies that when the open loop is below the field capacity, and some ensemble members are above the field capacity, the ensemble mean loses water relative to the open loop. This often occurs during the spring and autumn months, which agrees with Fig. 3e.

3.1.2 Influences of precipitation and clay content

We performed extra experiments to determine whether the differences in the ensemble perturbation bias between the sites can be partly attributed to soil clay content or to precipitation. The clay content is an important aspect of soil texture, while precipitation was the most important climate variable in our sensitivity study (Table S1.2 of Supplement 1). Firstly, we compared the average perturbation bias (normalized by the RMSE) of the experiment Ens with the average precipitation for each site. A scatter plot of the average daily precipitation against the normalized bias is shown in Fig. 6a. The linear regression line shows a strong negative correlation between the precipitation amount and the magnitude of the perturbation bias. The wettest site, Sabres, is labelled as 'S' and the driest site, Narbonne, is labelled as 'N'. These are the two sites that were compared in detail in Sect. 3.1.1.

We then performed an experiment to determine the impact of clay content on the bias. In this experiment we used the same atmospheric forcing of the wettest site (Sabres) for all the sites. This eliminates the impact of different climate on the results and leaves only differences in soil class. The clay percentage is plotted against the perturbation bias in Fig. 6b. We then repeated the experiment in Fig. 6b but instead using the same atmospheric forcing of the driest site (Narbonne) for all the sites. The results are shown in Fig. 6c. Neither Fig. 6b nor 6c show a strong correlation between the clay percentage and the bias. On the other hand, the perturbation bias for the drier climate in Fig. 6c is much greater for all the sites than for the wetter climate 6b. These results demonstrate that precipitation acts to suppress the perturbation bias, while clay content has little influence on the bias for these 12 sites.

3.1.3 Bias correction

The simple bias correction scheme (Eq. (16)) was tested on the ensemble and the results are also shown in Table 3 (labelled as Ens_{bc}). The bias has been reduced to less than a 10th of the size and the RMSE reduced by half compared with the original Ens. Fig. 3 a-c show the net soil moisture content of each layer for the bias corrected ensemble (Ens_{bc}). The bias correction has effectively removed the bias from all three layers.

15

The soil water depletion is shown in Fig. 3 d–f for Ens_{bc}. It appears that the application of the bias correction scheme has inadvertently increased the soil water depletion of Ens_{bc} relative to the open loop. A side-effect of the increase in water depletion processes is a reduction in the ensemble spread. The monthly average spread is shown in Fig. 7 for (a) Sabres; and (b) Narbonne. The bias is much greater for Narbonne than Sabres (comparing Fig. 4a with 4b). Therefore the ensemble spread is halved by the bias correction for Narbonne, but only slightly reduced for Sabres. The reduced ensemble spread has important repercussions for DA, where the ensemble spread determines the weight to give to the background. This is investigated in Sect. 3.3.

3.2 SEKF vs EnSRF: synthetic observations

The EnSRF and the SEKF were first tested with only errors in the precipitation forcing. Recall that the observations were taken from a single model simulation with a small random observation error (Sect. 2.6). For this experiment the EnSRF used a perfect stochastic representation of the precipitation uncertainty. The SEKF cannot capture the precipitation uncertainty directly, but instead the climatological background-error variances were calibrated to produce the best performance.

The time-averaged and site-averaged WG2 RMSE of the 20 member EnSRF is labelled as EnSRF_S in Table 3. The EnSRF_S RMSE is about half the size of the open loop RMSE (OL_S). The performance of EnSRF_S for various ensemble sizes is demonstrated in Table 4. A gradual improvement in the EnSRF_S is apparent in Table 4 as the ensemble size is increased from 3 to 20. The sampling error in the perturbed forcing is the cause of the larger RMSE for the smaller ensemble sizes. The EnSRF_S has an ACC close to 1.0. No significant bias was introduced because the ensemble spread was small (not shown).

The SEKF_S climatological background-error covariance needed to be calibrated in order to minimize the RMSE. The SEKF_S with the optimal calibration is labelled as SEKF_S in Table 3. The SEKF_S performs slightly better than the open loop, but not as well as the ensemble DA methods. This is expected because the SEKF_S does not capture the uncertainty in the precipitation directly, rather it uses larger variances in **B** to compensate for forcing errors. Table 5 shows the performance of the SEKF_S with various background-error covariance specifications, with the bold font showing the optimal calibration.

The monthly average performances of the open loop and the DA methods are shown in Fig. 8 a-c. The open loop RMSE is greatest in the spring and autumn seasons (Fig. 8a). The soil moisture is going through a transition from a wet to dry state in spring and from a dry to wet state in autumn, which increases its sensitivity to perturbations in the precipitation. During the winter the WG2 reservoir is close to the field capacity. During the summer the soil moisture is close to the wilting point and there is relatively little precipitation to perturb. Unlike the EnSRF_S, the SEKF_S climatological background-error covariance does not account for the seasonal variability in precipitation amounts. This is evidenced by examining the Kalman gains.

16

The monthly average WG2 Kalman gain for EnSRF$_S$ is displayed in Fig. 9. The EnSRF$_S$ Kalman gain is closely correlated with the open loop RMSE, with peaks in late spring and autumn. The SEKF$_S$ Kalman gain is plotted in Fig. 9 both before and after the 1.0 limit is imposed on the model Jacobian. In Sect. 2.3.1 it was explained that this limit is exceeded only when the model behaviour is very nonlinear, during which time the SEKF tangent linear approximation is inadequate. In contrast to the EnSRF$_S$ Kalman gain, the SEKF$_S$ Kalman gain peaks in July. By imposing the limit on the over-sensitive Jacobian, the Kalman gain is notably reduced between May and October, which shows that the SEKF$_S$ tangent linear approximation is poor during this period. This explains why the SEKF$_S$ WG2 RMSE and ACC are worse than the open loop between June and September (Fig. 8a and b).

3.3 SEKF vs EnSRF: real observations

Firstly, the performance of the calibrated EnSRF is analyzed for the baseline experiment, where only additive perturbations were applied to WG1 and WG2. This is labelled as EnSRF$_{R1}$ in Table 3. The EnSRF$_{R1}$ method does perform significantly better than the open loop (labelled as OL$_R$ in Table 3). But a dry bias has been introduced, which represents a large fraction of the RMSE (25%), and is consistent with the size of the dry bias introduced by experiment Ens in Table 3. The EnSRF$_{R1}$ method was tested with various ensemble sizes and the results are shown in Table 4. There is no improvement beyond an ensemble size of 20 members.

Fig. 4c and d are equivalent to Fig. 4a and b but instead show the EnSRF$_{R1}$ ensemble, the open loop and the observations for 2009. During December and January, the ensemble for Narbonne (Fig. 4d) has a similar negatively skewed spread to Fig. 4b, which indicates that the perturbation bias is present. The observations are wetter than the open loop during the summer and are therefore offsetting the perturbation bias. The opposite is true in December and January, when the observations are drier than the open loop, which causes many of the ensemble members to dip well below the wilting point.

The SEKF$_R$ performance is presented in Table 3. In terms of RMSE, SEKF$_R$ performs marginally better than EnSRF$_{R1}$, while EnSRF$_{R1}$ has a slightly higher ACC than SEKF$_R$. The SEKF$_R$ method is also affected by a negative bias, which is about half the size of the EnSRF$_{R1}$ bias. The SEKF$_R$ analysis increments themselves introduce a negative bias through the same mechanisms as the ensemble perturbation bias. But the EnSRF$_{R1}$ method is affected by the biases introduced by both the ensemble perturbations and the analysis increments, and therefore the EnSRF$_{R1}$ bias is greater than the SEKF$_R$ bias.

Fig. 10 a and b show contour plots of the EnSRF$_{R1}$ RMSE and ACC respectively, for the range of additive perturbations used to calibrate the method. Fig. 10 c and d show equivalent contour plots for the SEKF$_R$. Both performance metrics are much more sensitive to the WG2 perturbations than the WG1 perturbations, which is logical given that the WG2 layer is much thicker than the WG1 layer.

17

The SEKF$_R$ results are less noisy than the EnSRF$_{R1}$ results (Fig. 10 a and b) because the SEKF is not affected by the noise associated with the finite ensemble size of the EnSRF. The EnSRF and the SEKF were also tested with the smaller observation error of $0.35(w_{fc}-w_{wilt})$, but this did not significantly change the performance of the methods (not shown).

The monthly-averaged and station-averaged RMSE, the ACC and the bias are shown for the open loop, SEKF$_R$ and EnSRF$_{R1}$ in Fig. 8 d-f respectively. The RMSE in all cases is highest in June and October (8d), as this corresponds with the greatest fluctuations in soil moisture. This is also when the most improvement over the open loop occurs. The EnSRF$_{R1}$ RMSE is slightly degraded relative to the SEKF$_R$ in July and August as a result of the perturbation bias during this period (evident in Fig. 8f). The superior EnSRF$_{R1}$ ACC from July to September is explained below.

The WG2 Kalman gains for EnSRF$_{R1}$ and the SEKF$_R$ are shown in Fig. 9b. The SEKF$_R$ performs better with a larger WG2 Kalman gain than EnSRF$_{R1}$. The EnSRF$_{R1}$ Kalman gain shows some seasonal variability, with the largest values occurring at the same times as the open loop in June and October (Fig. 8d). The SEKF$_R$ Kalman gain is shown in Fig. 9b before and after the limit of 1.0 imposed on the Jacobian. The Kalman gain peaks in summer as a consequence of the over-sensitive model Jacobian during water-stressed conditions. This problem with the model Jacobian appears to explain why the EnSRF$_{R1}$ ACC is higher than the SEKF$_R$ ACC during July, August and September (Fig. 8b).

The impact of precipitation forcing perturbations on the EnSRF is investigated. This experiment is labelled as EnSRF$_{R2}$ in Table 3. The perturbed precipitation does not modify the analysis performance significantly compared with the unperturbed case (EnSRF$_{R1}$). A slightly smaller additive inflation is optimal with the perturbed forcing. This indicates that the perturbed forcing is having a similar effect to additive covariance inflation, but without the advantages demonstrated for the idealized experiments (Sect. 3.2).

Finally, the bias correction scheme is tested on the EnSRF. This experiment is labelled as EnSRF$_{R3}$ in Table 3. The large bias in Table 3 for the EnSRF without bias correction (EnSRF$_{R1}$) has been approximately halved by applying bias correction. The bias correction technique cannot correct biases introduced by the analysis increments. Therefore EnSRF$_{R3}$ is affected by a small negative bias. The ACC of EnSRF$_{R3}$ is degraded relative to EnSRF$_{R1}$, which is probably related to unrealistic temporal changes in the ensemble spread that occur as a result of the bias correction (see Fig. 7).

4 Discussion

The discussion focusses on the knowledge gained from the experiments, refering to the four criteria set out in the introduction. These are: the stochastic error representation of rainfall errors (Sect. 4.1), the Gaussian assumptions (Sect. 4.2) the influence of climate and texture (Sect. 4.3)

18

and flow-dependence of the DA methods (Sect. 4.4). Sect. 4.5 discusses the influence of the choice of model on the results.

4.1 Stochastic precipitation error representation

The experiments in Sect. 3.2 were designed to assess the advantage gained by a perfect stochastic representation of precipitation uncertainty in the EnSRF over additive background errors in the SEKF. Clearly the EnSRF benefited from the direct representation of the uncertainty. However, in the real experiments the same perturbations gave no advantage to the EnSRF compared with additive covariance inflation alone (compare EnSRF$_{R2}$ with EnSRF$_{R1}$ in Table 3).

Maggioni et al. (2011) demonstrated that soil moisture simulations are less sensitive to rainfall uncertainty information than the precipitation fields themselves. They attributed this loss of information to two factors: (i) The nonlinear and integrating nature of soil moisture models; and (ii) The dissipative behaviour of soil moisture dynamics, which dampens perturbations in the initial conditions. These conclusions agree with our findings. Furthermore, the transfer of precipitation uncertainty information to soil moisture depends on an accurate land surface model. Significant model errors were likely in this study (see Sect. 4.5).

It is likely that the precipitation errors in this study were also underestimated. Hossain and Anagnostou (2005) estimated that rainfall errors represent between 20 and 60% of the uncertainty in soil moisture prediction. In this study the estimated rainfall errors only made up 10% of the total open loop errors. The percentage scaling in this study could not take into account the nonstationary and intermittent nature of precipitation errors, including the probability of missed precipitation or false alarms. More comprehensive precipitation error models have been developed which can take these factors into account (see e.g. (Clark and Slater, 2006; Hossain and Anagnostou, 2006; Maggioni et al., 2011, 2012, 2014; Carrera et al., 2015)). It is planned that one of these approaches will be adopted for the Land Data Assimilation System (LDAS) at Météo-France. The calibration of the various parameters for these rainfall error models requires considerable testing.

4.2 Gaussian assumptions

In the synthetic experiments, the EnSRF was applied with a perfect stochastic representation of the precipitation uncertainty, while the SEKF used climatological background errors. It is possible to determine how well the DA methods agreed with Kalman Filter theory by comparing the pdfs of the innovations $(\mathbf{y} - \mathbf{y}^{\mathrm{o}})$ with the sum of the background and observation-error covariances $(\mathbf{HP}^{\mathrm{b}}\mathbf{H}^T + \mathbf{R})$ (Desroziers et al., 2005). Fig. 11 shows the histograms of the innovations normalized by $\sqrt{\mathbf{HP}^{\mathrm{b}}\mathbf{H}^T + \mathbf{R}}$ for the EnSRF$_S$ a and the SEKF$_S$ b for the synthetic experiments. The pdf for EnSRF$_S$ agrees very well with Kalman theory, since it has a mean of zero and it closely fits the normal distribution (the green line). On the other hand, the pdf for the SEKF$_S$ is flatter than the normal distribution and therefore agrees less well with Kalman theory. This demonstrates that without the

correct specification of forcing errors, the optimum calibration of the background-error covariance will not necessarily agree with Kalman filter theory.

In the real experiment neither method had a perfect representation of the background errors. Both methods used an average value of $\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^T + \mathbf{R}$ about 4 times larger than $(\mathbf{y} - \mathbf{y}^{\mathrm{o}})^2$ (not shown), which indicates that the Kalman filter assumptions were incorrect.

The nonlinearity problems manifested themselves in different ways for the SEKF and the EnSRF. For the SEKF, the Jacobian between the surface and the root zone became too large. This over-sensitivity is partly related to an unrealistic feature of the modelled surface energy balance, since one single surface temperature is used for bare soil and the vegetation layer (Draper et al., 2009; Mahfouf, 2014). The EnSRF instead suffered from the perturbation bias. The explanation for the perturbation bias was linked to the nonlinear behaviour of evapotranspiration and drainage. Similar problems were encountered by Ryu et al. (2009) on the NOAH land surface model used by the National Center for Environmental Prediction (NCEP), where detailed explanations are given. It should be noted that the nonlinear effects of water transfers on the ensemble perturbations are highly sensitive to the model and the model regime (including the layer depths and vegetation content). Indeed, Ryu et al. (2009) discovered a positive perturbation bias for their system, when our study discovered a negative bias.

Finally, a key assumption underpinning the EnSRF is that the ensemble size is sufficiently large to represent the errors in the background state. An ensemble size that is too small results in inbreeding, where the errors in the background state are underestimated (Houtekamer and Mitchell, 1998). We investigated the impact of ensemble size on the EnSRF WG2 RMSE (see Table 4). It was found that sampling errors were only significant with ensemble sizes less than 20 members, which is consistent with studies by Carrera et al. (2015) and Maggioni et al. (2012). However, in all these studies the EnKF analysis is calculated independently for each gridpoint. It is likely that sampling errors would be much more important for a 3D-EnKF approach, where background-error covariances between gridpoints are taken into account, due to the much larger number of degrees of freedom.

4.3 Influences of climate and soil texture

It was discovered in Sect. 3.1.2 that there was a strong correlation between average precipitation and the magnitude of the perturbation bias. There are two reasons for this: Firstly, the soil saturates rather quickly during precipitation events, which reduces the ensemble spread. This in turn reduces the perturbation bias. Secondly, more frequent precipitation events reduce the occurrence of ensemble members dipping below the wilting point in summer. This was confirmed by comparing the seasonal biases and the ensemble spread over the 12 sites (not shown). It is important to emphasize that these results are limited to the 12 sites and other climate variables have not been considered.

4.4 Flow-dependence

The flow-dependence of the DA methods was examined by comparing the WG2 Kalman gains. When the background errors are larger, the Kalman gain should increase in order to give more weight to the observations. Given the assumption that there is no temporal evolution in the observation errors, the Kalman gain and the open loop errors should peak at the same time. In the synthetic experiments, the EnSRF Kalman gains showed a similar seasonal variability to the open loop RMSE, unlike the SEKF Kalman gains. This showed that the EnSRF was able to capture seasonal variability in the background errors. The EnSRF Kalman gains showed less seasonal variability in the real experiments, possibly because the model and forcing errors were not accurately represented. This could explain why the EnSRF performed better than the SEKF in the synthetic experiments but not in the real experiments.

4.5 Land surface model

In this study, the nonlinearity issues were most prevalent when the model descended below the wilting point or ascended above the field capacity. For this reason it may seem intuitive to introduce lower and upper bounds at these thresholds. However, water can be slowly lost through the leaves by cuticular conductance below the wilting point or though incomplete closure of the stomata. Soil moisture may also temporarily increase above the field capacity before it drains out (Mahfouf and Noilhan, 1996). These features are part of the NIT version of the ISBA-A-gs model. Therefore, imposing bounds would not be realistic. Boundary problems can also affect the analysis. For example, Ryu et al. (2009) used a bounded land surface model and found the model bounds were partly responsible for the positive perturbation bias in their study.

The SEKF and the force-restore based ISBA-A-gs model are currently embedded in the SURFEX platform of Météo-France. However, the diffusion-based multi-layer model (ISBA-DF) will soon be implemented (Decharme et al., 2011). The soil moisture evolution of ISBA-DF is determined by the mixed form of the Richard's equation, rather than the force-restore method. This is more realistic than the force-restore method as it solves the heat and water diffusion equations explicitly over at least 5 layers. Parrens et al. (2014) compared the SEKF for a two-layer version of the force-store model with an 11-layer implementation of ISBA-DF. They found that the SEKF performance was enhanced by introducing multiple layers. In particular, the multi-layer model captured the vertical profile of the root zone soil moisture better than the two-layer model. It will be interesting to test the EnSRF with ISBA-DF and multiple layers. The EnSRF flow-dependent background-error covariance may be able to exploit the improved vertical correlations between the layers.

5 Conclusions

Twelve sites in Southwest France were selected for soil moisture DA experiments. The different sites were chosen, in particular, for their variability in climate and soil texture, which influence soil water transfers. The SEKF and the EnSRF DA methods were compared in terms of their ability to provide an accurate soil moisture analysis. The three-layer ISBA-A-gs land surface model (Noilhan and Mahfouf, 1996; Calvet et al., 1998) was implemented in the experiments. This model is based on the force-restore method of Deardorff (1977).

In the real experiments, 5 cm deep in situ observations were assimilated and 30 cm deep observations were used to evaluate the root-zone soil moisture analysis. The two methods performed similarly and improved on the open loop. This accords with Muñoz Sabater et al. (2007) who compared similar methods. However, the synthetic "identical twin" experiments were designed to assess the advantage the EnSRF could gain over the SEKF by using a perfect stochastic representation of precipitation uncertainty. The results clearly demonstrated an advantage in the EnSRF performance for the idealized regime (with a perfect model and small observation errors). It might therefore be possible for the EnSRF to perform better than the SEKF with real observations, if the rainfall errors were accurately captured stochastically. However, this is challenging because the actual rainfall errors are unknown. Moreover, the transfer of precipitation uncertainty information to soil moisture relies on the accuracy of the land surface model. These challenges could explain why the simple rainfall error model did not improve the EnSRF soil moisture analysis in the real experiments. It is recommended that a more realistic rainfall error model is tested on the EnSRF. This should use a lognormal distribution and take into account the intermittent and non-stationary nature of precipitation errors. The EnSRF representation of model errors in this study could also be improved. We employed the common approach of adding perturbations to the prognostic variables. Maggioni et al. (2012) demonstrated that model errors are better represented by perturbing the model parameters instead of the prognostic variables.

Both methods suffered from incorrect linear assumptions related to the nonlinear evapotranspiration and drainage functions: For the SEKF, the model Jacobian between the surface and the root-zone was over-sensitive to soil moisture perturbations during dry conditions and this led to excessive analysis increments. For the EnSRF, a significant dry bias was found, largely as a result of the ensemble perturbations causing excessive evapotranspiration near the wilting point. The perturbation bias was less detrimental to the sites with wetter climates because precipitation forces the soil moisture above the wilting point and reduces the ensemble spread.

A bias correction scheme was tested on the EnSRF. Although this removed the perturbation bias, it led to spurious increases in drainage and evapotranspiration water fluxes, induced by the wetter soil moisture states. Consequently it reduced the ensemble spread and degraded the analysis. For many applications, such as hydrology, the water fluxes can be as important as the soil moisture states.

22

Therefore it would be inappropriate to use this bias correction scheme when coupling LDAS with a hydrological model, which requires accurate drainage inputs.

A disadvantage with the EnSRF is the computational burden and sampling error associated with the ensemble. However, there is evidence to suggest that a large ensemble size is not necessary for land surface models. In this study, there was no significant advantage gained by using more than 20 members, which is consistent with studies by Carrera et al. (2015) and Maggioni et al. (2012).

Although this study was implemented on sites with various soil textures, only the grassland vegetation type was present, and the root-zone depth was relatively shallow. It will be important to test the EnSRF over the French domain, which is much more variable in terms of vegetation, soil texture and climate. The EnSRF will also be tested on a more realistic multi-layer diffusion-based model (Decharme et al., 2011). In our experiments the EnSRF flow-dependent background-error covariance was able to capture the seasonal variability in the background errors, which was not evident with the SEKF. The EnSRF covariance also has greater potential to exploit improvements in the model physics.

Finally, the SEKF and the EnSRF methods in this study were calibrated using the same observation source as the assimilated observations (albeit at different depths). Therefore we acknowledge that the results in this study will not necessarily apply to sites where these observations are not available for calibration. In the case of independent data sources, the triple colocation approach might be preferable to rescaling the observations to match the model mean and standard deviation, as this rescales datasets using three independent estimates of the same variable (see e.g. Tugrul Yilmaz and Crow (2008)).

The Supplement related to this article is available online at doi:10.5194/hess-0-1-2015-supplement.

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., and Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, Hydrol. Earth Syst. Sci., 12, 1323–1337, doi:10.5194/hess-12-1323-2008, 2008.

Balsamo, G., Bouyssel, F., and Noilhan, J.: A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model, Q. J. Roy. Meteor. Soc., 130, 895–915, 2004.

Balsamo, G., Mahfouf, J.-F., Bélair, S.and Deblonde, G.: A land data assimilation system for soil moisture and temperature: An information study, J. Hydrometeorol., 8(6), 1225–1242, doi:10.1175/2007JHM819.1,2007.

Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., Albergel, C., and Lafont, S.: Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study, Biogeosciences, 8, 1971–1986, doi:10.5194/bg-8-1971-2011, 2011.

Boone, A., Calvet, J.-C., and Noilhan, J.: Inclusion of a third soil layer in a land surface scheme using the force–restore method, J. Appl. Meteorol., 38, 1611–1630, 1999.

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the Ensemble Kalman Filter, Mon. Weather Rev., 126, 1719–1724, 1998.

Calvet, J.-C. and Noilhan, J.: From near-surface to root-zone soil moisture using year-round data, J. Hydrometeorol., 1, 393–411, 2000.

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J.-P.: An interactive vegetation SVAT model tested against data from six contrasting sites, Agr. Forest Meteorol., 126, 92, doi:10.1016/S0168-1923(98)00091-4, 1998.

Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., and Piguet, B.: In situ soil moisture observations for the CAL/VAL of SMOS: the SMOSMANIA network, in: 2007 International Geoscience and Remote Sensing Symposium, Proceedings, Barcelona, Spain, 23–28 July 2007, 1196–1199, 2007.

Carrera, M., Bélair, S., and Bilodeau, B.: The Canadian Land Data Assimilation System (CaLDAS): description and synthetic evaluation study, J. Hydrometeorol., 16, 1293–1294, doi:10.1175/JHM-D-14-0089.1, 2015.

Clark, M. and Slater, A.: Probabilistic quantitative precipitation estimation in complex terrain, J. Hydrometeorol., 7, 3–22, doi:10.1175/JHM474.1, 2006.

Crow, W. and Van den Berg, M.: An improved approach for estimating observation and model error parameters in soil moisture data assimilation, Water Resour. Res., 46, W12519, doi:10.1029/2010WR009402, 2010.

Crow, W. and Van Loon, E.: Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture, J. Hydrometerol., 7, 421–432, 2006.

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., and Isaksen, L.: A Simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF, Q. J. Roy. Meteor. Soc., 139, 1199–1213, doi:10.1002/qj.2023, 2013.

Deardorff, J.: A paramaterization of ground surface moisture for use in atmospheric prediction models, J. Appl. Meteorol., 16, 1182–1185, 1977.

Decharme, B., Boone, A., Delire, C., and Noilhan, J.: Local evaluation of the interaction between soil bio-sphere atmosphere soil multilayer diffusion scheme using four pedotransfer functions, J. Geophys. Res., 116, D20126, doi:10.1029/2011JD016002, 2011.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Q. J. Roy. Meteor. Soc., 131, 3385–3396, 2005.

Dharssi, I., Bovis, K. J., Macpherson, B., and Jones, C. P.: Operational assimilation of ASCAT surface soil wetness at the Met Office, Hydrol. Earth Syst. Sci., 15, 2729–2746, doi:10.5194/hess-15-2729-2011, 2011.

Draper, C., Mahfouf, J.-F., and Walker, J.: An EKF assimilation of AMSR-E soil moisture into the ISBA surface scheme, J. Geophys. Res., 114, D20104, doi:10.1029/2008JD011650, 2009.

Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., and Wagner, W.: Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France, Hydrol. Earth Syst. Sci., 15, 3829–3841, doi:10.5194/hess-15-3829-2011, 2011.

Draper, C., Reichle, R., Lannoy, G. D., and Liu, Q.: Assimilation of passive and active microwave soil moisture retrievals, Geophys. Res. Lett., 39, L04401, doi:10.1029/2011GL050655, 2012.

Drusch, M., Wood, M., and Gao, H.: Observation operators for the direct assimilation of TRMM Microwave Imager retrieved soil moisture., Geophys. Res. Lett, 32, L15403, doi:10.1029/2005GL023623, 2005.

Duerinckx, A., Hamdi, R., Mahfouf, J.-F., and Termonia, P.: Study of the Jacobian of an extended Kalman filter for soil analysis in SURFEXv5, Geosci. Model Dev., 8, 845–863, doi:10.5194/gmd-8-845-2015, 2015.

Durand, Y., Brun, E., Mérindol, L., Gilbert, G., Bernard, L., and Martin, E.: A meteorological estimation of relevant parameters for snow models, Ann. Glaciol., 18, 65–71, 1993.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.-C., Spencer, M. W., Thurman, S. W., Leung Tsang, and Van Zyl, J.: The Soil Moisture Active Passive (SMAP) mission, P. IEEE, 98, 704–716, 2010.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res, 99, 10143–10162, 1994.

Gibelin, A.-L., Calvet, J.-C., Roujean, J.-L., Jarlan, L., and Los, S.: 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, 2006.

Hamill, T. and Whitaker, J.: Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches, Mon. Weather Rev., 133, 3132–3147, 2005.

Hess, H.: Assimilation of screen-level observations by variational soil moisture analysis, Meteorol. Atmos. Phys., 77, 142-154, doi:10.1007/s007030370023, 2001.

Hossain, F. and Anagnostou, E. N.: Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture, Adv. Water Resour., 28, 1336–1350, doi:10.1016/j.advwatres.2005.03.013, 2005.

Hossain, F. and Anagnostou, E. N.: Assessment of a multi-dimensional satellite rainfall error model for ensemble generation of satellite rainfall data, Geosci. Remote Sens. Lett., 3, 419–423, 2006.

Houtekamer, P. and Mitchell, H.: Data assimilation using an Ensemble Kalman Filter technique, Mon. Weather Rev., 126, 796–811, 1998.

25

Jazwinski, A. H.: Stochastic Processes and Filtering Theory, Academic Press, New York, USA, 150–158, 1970.

Kalman, R. E.: A new approach to linear filtering and prediction problems, J. Basic Eng.-T. ASME, 82, 35–45, 1960.

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J.-M., Font, J., and Berger, M.: Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission, IEEE T. Geosci. Remote, 39, 1729–1735, 2001.

Maggioni, V., Reichle, R., and Anagnostou, E.: The effect of satellite rainfall error modeling on soil moisture prediction uncertainty, J. Hydrometeorol., 12, 413–428, doi:10.1175/2011JHM1355.1, 2011.

Maggioni, V., Anagnostou, E. N., and Reichle, R. H.: The impact of model and rainfall forcing errors on characterizing soil moisture uncertainty in land surface modeling, Hydrol. Earth Syst. Sci., 16, 3499–3515, doi:10.5194/hess-16-3499-2012, 2012.

Maggioni, V., Sapiano, M., and Adler, R.: An error model for uncertainty quantification in high-time-resolution precipitation products, J. Hydrometeorol., 15, 1274–1292, doi:10.1175/JHM-D-13-0112.1, 2014.

Mahfouf, J.: Advances in model physics and their relevance to satellite data assimilation, in: Proceedings of the ECMWF Seminar on the Use of Satellite Observations in NWP, available at: http://www.ecmwf.int/sites/default/files/MAHFOUF_ECMWF2014_web.pdf, last access: September 2014.

Mahfouf, J.-F.: L'analyse dans le sol a Meteo-France. Partie 1: Evaluation et perspectives a l'echelle locale, Meteo France technical report, Toulouse, France, 2007.

Mahfouf, J.-F.: Assimilation of satellite-derived soil moisture from ASCAT in a limited-area NWP model, Q. J. Roy. Meteor. Soc., 136, 784–798, doi:10.1002/qj.602, 2010.

Mahfouf, J.-F. and Noilhan, J.: Inclusion of gravitational drainage in a land surface scheme based on the force-restore method, J. Appl. Meteorol., 35, 987–992, 1996.

Mahfouf, J.-F., Bergaoui, K., Draper, C., Bouyssel, C., Taillefer, F., and Taseva, L.: A comparison of two off-line soil analysis schemes for assimilation of screen-level observations, J. Geophys. Res, 114, D08105, doi:10.1029/2008JD011077, 2009.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6, 929–960, doi:10.5194/gmd-6-929-2013, 2013.

Muñoz Sabater, J., Jarlan, L., Calvet, J.-C., and Boyssel, F.: From near-surface to root-zone soil moisture using different assimilation techniques, J. Hydrometeorol., 8, 194–206, doi:10.1175/JHM571.1, 2007.

Noilhan, J. and Mahfouf, J.-F.: The ISBA land surface parameterisation scheme, Global Planet. Change, 13, 145–159, 1996.

Parrens, M., Mahfouf, J.-F., Barbu, A. L., and Calvet, J.-C.: Assimilation of surface soil moisture into a multilayer soil model: design and evaluation at local scale, Hydrol. Earth Syst. Sci., 18, 673–689, doi:10.5194/hess-18-673-2014, 2014.

885   Quitana-Ségui, P., Moigne, P. L., Durand, P., Martin, E., Baillon, F. H. M., Canellas, C., and Franchistéguy, L.: Analysis of near-surface atmospheric variables: Validation of SAFRAN analysis over France, J. Appl. Meteorol. Clim., 47, 92–107, 2008.

Reichle, R., Walker, J., Koster, R., and Houser, P.: Extended vs ensemble Kalman filtering for Land Data Assimilation, J. Hydrometeorol., 3, 728–740, 2002.

890   Reichle, R., Koster, R., Dong, J., and Berg, A.: Global soil moisture from satellite observations, land surface models, and ground data: implications for data assimilation, J. Hydrometeorol., 5, 430–442, 2004.

Reichle, R., Crow, W., and Keppenne, C.: An adaptive Ensemble Kalman Filter for soil moisture data assimilation, Water Resour. Res., 44, WO3243, doi:10.1029/2007WR006357, 2008.

Ryu, D., Crow, W., Zhan, X., and Jackson, T.: Correcting unintended perturbation biases in hydrologic data 895   assimilation, J. Hydrometeorol., 10, 734–750, 2009.

Scipal, K., Drusch, M., Wagner, W.: Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, Adv. Water. Resour., 31, 1101–1112, doi:10.1016/j.advwatres.2008.04.013, 2008.

Tugrul Tilmaz, M., and Crow., W.T.: The Optimality of Potential Rescaling Approaches in Land Data Assimi-900   lation, J. Hydrometeorol., 14, 650-660, doi:10.1175/JHM-D-12-052.1, 2012.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric analysis over France with the Safran system, Int. J. Climatol., 30, 1627–1644, doi:10.1002/joc.2003, 2010.

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., and Kerr, Y.: Operational 905   readiness of microwave remote sensing of soil moisture for hydrologic applications, Nord. Hydrol., 38, 1–20, doi:10.2166/nh.2007.029, 2007.

Walker, J. and Houser, P.: A methodology for initializing soil moisture in a global climate model: Assimilation of near surface soil moisture observations, J. Geophys. Res., 11, 761–774, 2001.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather 910   Rev., 130, 1913–1924, 2002.

Zhou, Y., McLaughlin, D., and Entekhabi, D.: Assessing the performance of the Ensemble Kalman Filter for land surface data assimilation, Mon. Weather Rev., 134, 2128–2142, 2006.

Table 1. Table summarising the different methods. Cov stands for covariance matrix.

Method	Background Cov source	Cov propagation	Maintaining ensemble Spread
SEKF	Climatological	Implicitly via $\mathbf{H}$	NA
EnSRF	Ensemble	Implicitly via ensemble	Eq. (11)

Table 2. Table showing the experimental setup for the synthetic and real DA experiments.

DA component	Synthetic DA experiment	Real DA experiment
Truth	Model run	Unknown
Model	Model run + Eq. (17)	Model run
Assimilated obs.	WG1: Model run + obs. error	WG1: 5 cm depth in situ obs. + linear rescaling
EnSRF calibration	Eq. (17)	Eq. (15) + Eq. (17)
SEKF calibration	Eq. (14)	Eq. (14)
Validation data	WG2: Truth simulation	WG2: 30 cm depth in situ obs. + linear rescaling

Table 3. Table showing the calibrations and the performances of the experiments (averaged over 2008–2010 and averaged over all sites). The open loop is abbreviated to OL. The first two experiments were perturbed ensembles performed without DA and compared with the open loop. The synthetic and real experiments (denoted by subscripts S and R) were compared against the synthetic and real observations respectively. The variable Pr stands for hourly accumulated precipitation and ϵ_{Pr} is the standard deviation of the hourly precipitation perturbation (defined in Eq. (17)).

Exp.	Calibration: Obs. λ^o	WG1 λ_1^b	WG2 λ_2^b	Add. criteria	WG2 RMSE $(\mathrm{m^3\,m^{-3}})\times10^3$	WG2 ACC	WG2 Bias $(\mathrm{m^3\,m^{-3}})\times10^3$
Ens	–	–	0.025	–	9	0.97	−4.9
Ens$_{bc}$	–	–	0.025	Bias correct	4	0.99	0.6
OL$_S$	–	–	–	$\epsilon_{Pr}=50\% \, Pr$	2.2	0.995	0.0
EnSRF$_S$	0.05	–	–	$\epsilon_{Pr}=50\% \, Pr$	1.1	0.999	0.02
SEKF$_S$	0.05	0.04	0.02	–	1.8	0.996	0.01
OL$_R$	–	–	–	–	24.7	0.607	0.03
EnSRF$_{R1}$	0.5	0.2	0.03	–	20.8	0.720	−5.32
EnSRF$_{R2}$	0.5	0.1	0.03	$\epsilon_{Pr}=50\% \, Pr$	21.2	0.722	−5.82
EnSRF$_{R3}$	0.5	0.25	0.035	Bias correct	21.3	0.690	−2.79
SEKF$_R$	0.5	0.25	0.25	–	20.1	0.716	−2.21

Table 4. Site-averaged and time-averaged WG2 performances for EnSRF$_S$ and EnSRF$_{R1}$ for various ensemble sizes. The calibrated EnSRF is shown in bold font.

Ens. size	EnSRF$_S$ WG2 RMSE $(\mathrm{m^3\,m^{-3}}) \times 10^3$	EnSRF$_{R1}$ WG2 RMSE $(\mathrm{m^3\,m^{-3}}) \times 10^3$	EnSRF$_S$ WG2 ACC	EnSRF$_{R1}$ WG2 ACC
3	1.6	24.2	1.00	0.647
6	1.4	22.5	1.00	0.687
20	**1.1**	**20.8**	**1.00**	**0.720**
50	1.1	20.9	1.00	0.719
200	1.1	20.9	1.00	0.719

Table 5. SEKF$_S$ performance for various calibrations of the background-error scaling coefficients (λ_1^b and λ_2^b). The optimal value is in bold font.

λ_1^b	λ_2^b	WG2 RMSE $(\mathrm{m^3\,m^{-3}}) \times 10^3$
0.03	0.01	2.2
0.04	**0.02**	**1.8**
0.08	0.04	2.3
0.16	0.015	2.1
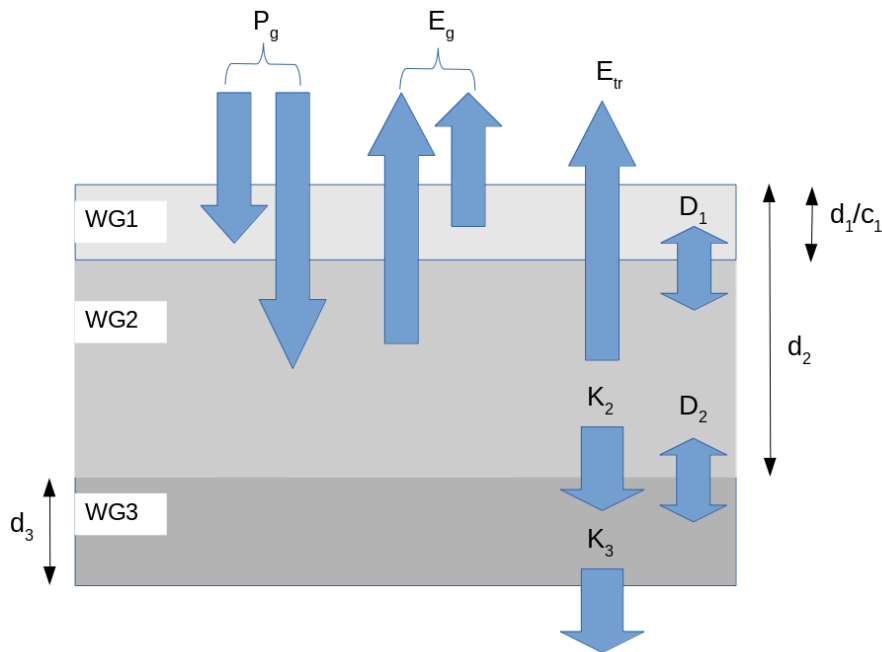
Figure 1. The soil moisture fluxes for the three-layer version of ISBA. The variables P_g, E_g and E_{tr} represent the precipitation, bare soil evaporation and transpiration respectively. The fluxes K and D represent the drainage and diffusion at the bottom of the layer.
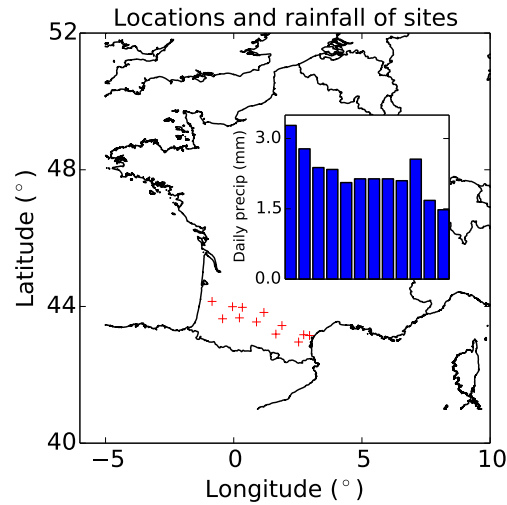
Figure 2. Location of the 12 sites used for the experiments (red crosses), selected from the SMOSMANIA network. Also shown is a bar chart of the daily average rainfall for each site (plotted from west to east). The rainfall is measured from the SAFRAN forcing database over the analysis period (2008-2010).
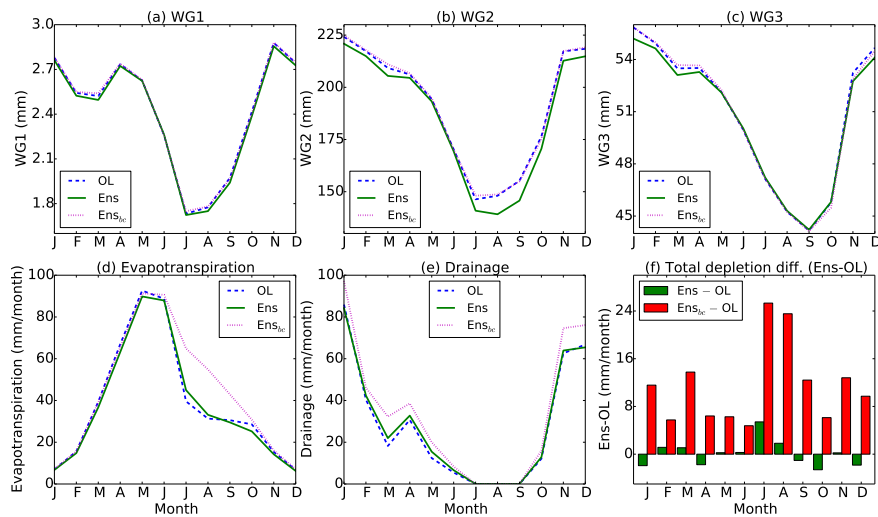

Figure 3. Monthly averaged and station averaged net water for **a** WG1, **b** WG2 and **c** WG3 and for the open loop, the ensemble mean (Ens) and the bias corrected ensemble mean (Ens$_{bc}$). Station averaged and monthly **d** evapotranspiration; **e** drainage and **f** total soil water depletion differences between ensemble mean and the open loop. Results are all averaged over the period 2008–2010.
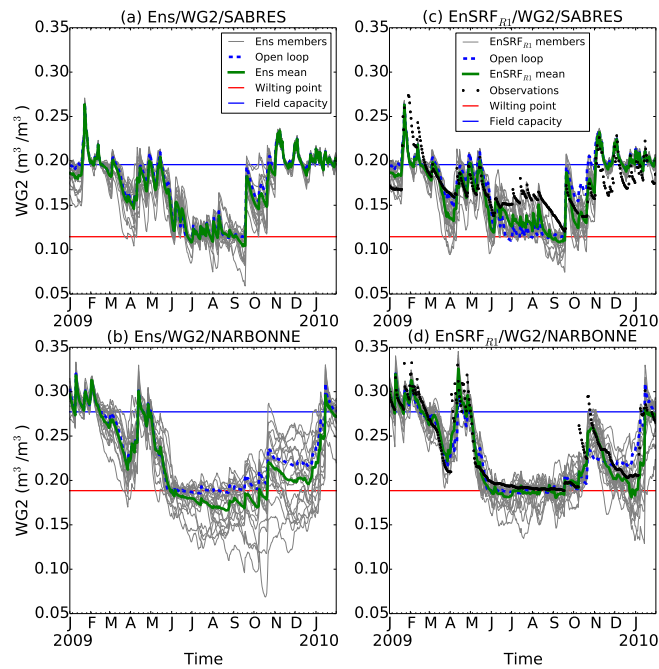
Figure 4. WG2 ensemble members, open loop and the ensemble mean for the sites (top) Sabres; and (bottom) Narbonne, over 2009. **a** and **b** show the results for experiment Ens (no DA), while **c** and **d** show the analysis ensemble members for the the experiment EnSRF$_{R2}$ (with observations). The field capacity and the wilting point are also shown for each site. The legends for the upper plots apply to the lower plots.

31
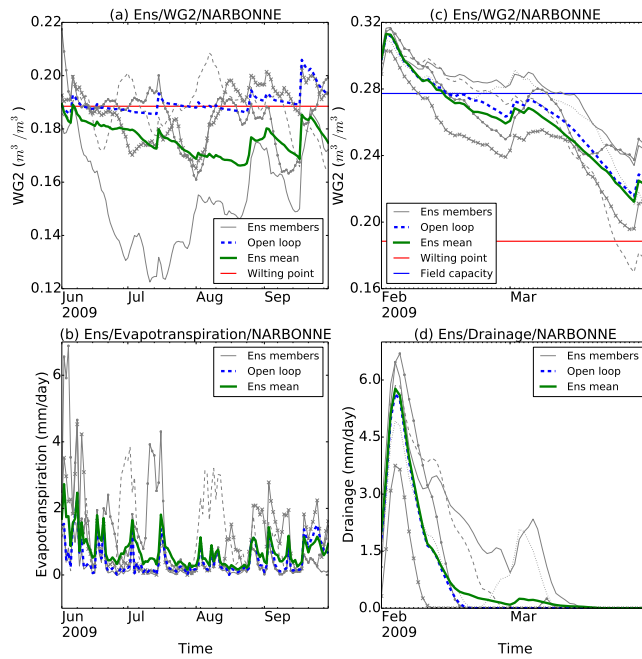
Figure 5. Four of the Ens ensemble members, open loop and the Ens ensemble mean for **a** WG2; and **b** Evapotranspiration; for Narbonne between June and September 2009. **c** and **d** are equivalent to **b** and **c** but instead show WG2 and drainage respectively, for February and March 2009.
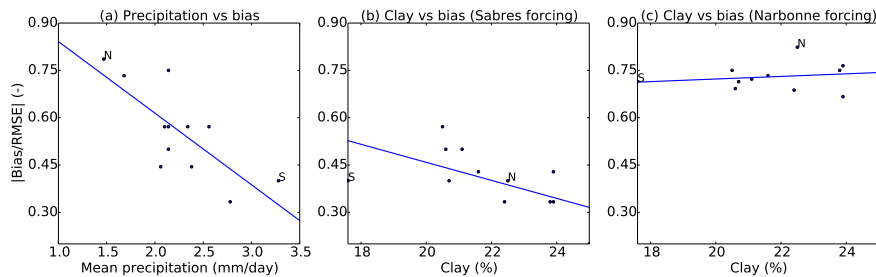

Figure 6. Average perturbation bias normalized by the RMSE against **a** average daily precipitation for each site; and **b** against clay percentage for each site, with the climatological forcing of the Sabres site applied to all the sites; and **c** same as **b** but the climatological forcing of the Narbonne site has been applied to all the sites. The Sabres and Narbonne sites are labelled as 'S' and 'N' respectively. The line of best fit (linear regression) is shown for each plot.
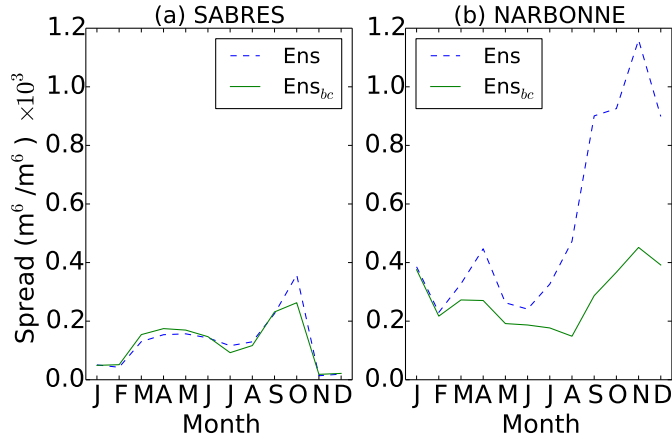
Figure 7. WG2 monthly averaged Ens and Ens_bc ensemble spread, for **a** Sabres; and **b** Narbonne. The results are averaged over 2008–2010.
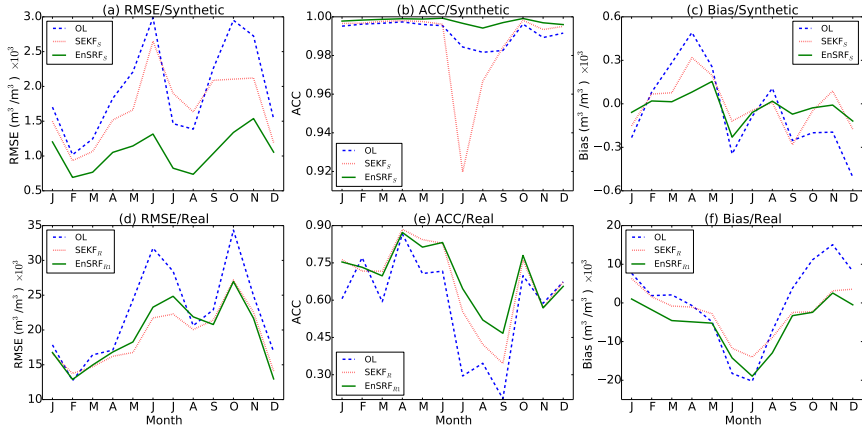

Figure 8. Site averaged and monthly averaged **a** WG2 RMSE; **b** WG2 ACC; and **c** WG2 bias; for the open loop, SEKF_S and EnSRF_S for the synthetic experiments. Plots **(d–f)** are equivalent to plots **(a–c)** but instead show the results for the open loop, SEKF_R and EnSRF_R1 for the real experiments. The results are averaged over 2008–2010.
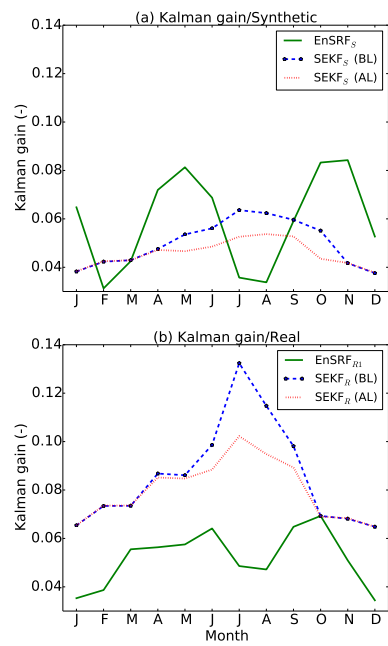
Figure 9. Station and monthly averaged WG2 Kalman gains for **a** EnSRF$_S$ and the SEKF$_S$ before (BL) and after (AL) the Jacobian limit of 1.0 is imposed. Plot **b** is equivalent to plot **a** but instead shows the results for EnSRF$_{R1}$ and SEKF$_R$. The results are averaged over 2008–2010.
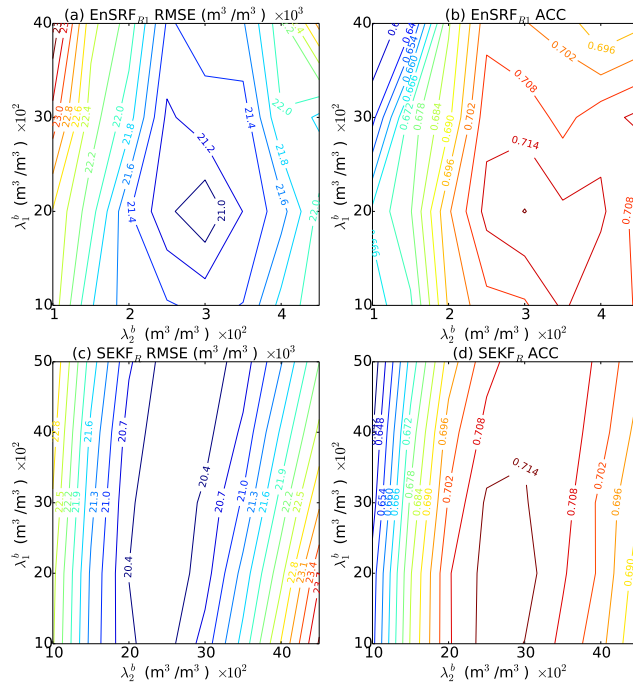
Figure 10. Contour plot showing the EnSRF$_{R1}$ **a** WG2 RMSE; and **b** WG2 ACC; for the range of additive inflation values used to calibrate the EnSRF$_{R1}$. Plots **c** and **d** show the equivalent contour plots to **a** and **b** for SEKF$_R$, with the range of background-error variances used to calibrate SEKF$_R$. The results are averaged over 2008–2010 and over the 12 sites.
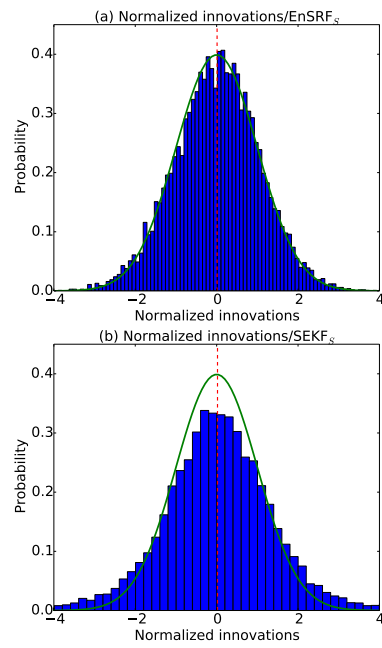
Figure 11. Histogram of normalized innovations for the 20 member EnSRF$_S$ **a** and the SEKF$_S$ **b** for the synthetic experiments. The green line shows the normal distribution and the red dashed line shows the mean.