Hydrol. Earth Syst. Sci. Discuss., 12, 7353–7403, 2015
www.hydrol-earth-syst-sci-discuss.net/12/7353/2015/
doi:10.5194/hessd-12-7353-2015
© Author(s) 2015. CC Attribution 3.0 License.

Hydrology and Earth System Sciences
Open Access
Discussions

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

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

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

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

Received: 14 June 2015 – Accepted: 17 July 2015 – Published: 5 August 2015

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Two data assimilation methods are compared for their ability to produce a deterministic 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 forcing errors stochastically. The accuracy of the deterministic analysis is measured on 12 sites with in situ observations and various soil textures in Southwest France (SMOSMANIA network). In the experiments with real observations, 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 sandy soils than clay soils because sandy soils are less sensitive to perturbations in the initial conditions. A simple bias correction technique is tested on the EnSRF. Although this reduces the bias, it also 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 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. But in the real experiments the same rainfall error specification does not improve the EnSRF analysis. It is likely that the actual rainfall errors are underestimated and that other sources of errors could limit the usefulness of this information. More comprehensive ways of representing the rainfall errors are suggested, which might improve the EnSRF performance.

HESSD

12, 7353–7403, 2015

1 Introduction

Soil moisture has an important 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 important for a variety of other applications, including drought monitoring, crop irrigation and water management.

The main objective of data assimilation (DA) for land surface models is to assimilate observed surface soil moisture to produce an analysis of root-zone soil moisture. Root-zone soil moisture 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 satellites (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 instruments 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 simplifies 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. 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. Instead it is used to provide soil moisture analyses and carbon fluxes for a variety of other applications, including hydrological models (see e.g. Draper

HESSD
12, 7353–7403, 2015

et al., 2011), and the European Copernicus programme (http://www.copernicus.eu/). An SEKF is currently operational at the European Centre for Medium range Weather Forecasts (ECMWF) for their land surface model (de Rosnay et al., 2013).

Ensemble DA methods are becoming increasingly popular for land surface models (see e.g. Carrera et al., 2015; Draper et al., 2012; Reichle et al., 2008, 2002; Muñoz Sabater et al., 2007; Zhou et al., 2006). There are three 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. Finally, a diffusion-based multi-layer model (Decharme et al., 2011) has been developed to replace the current 3 layer force-restore land surface model. The use of the multi-layer model will have implications for the computational costs of the DA methods over a large domain. Given that the SEKF requires a model Jacobian calculation for each prognostic layer, introducing multiple layers will substantially increase the computational burden of these calculations. It may in fact be more cost-effective to implement an EnKF method.

The EnKF has extensively been compared with the EKF on land surface models for assimilating 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 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 it's performance is slightly better. Muñoz Sabater et al. (2007) also demonstrated that a simple 1-D variational DA

HESSD
12, 7353–7403, 2015

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 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 den Berg, 2010; Reichle et al., 2008; Crow and Van Loon, 2006).

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 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 soil moisture analysis. Although the EnSRF necessarily uses an ensemble of model trajectories, the experiments focus on the deterministic analysis from the ensemble mean. 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 forcing;

ii. the Gaussian assumptions made by the DA methods;

iii. different soil textures;

iv. a flow-dependent background-error covariance.

The cumulative distribution function (CDF) matching technique is used in this study, which bias corrects the observations with respect to the model simulation (Calvet and Noilhan, 2000; Reichle et al., 2004; Drusch et al., 2005). However, ensemble perturbations can introduce additional biases as a result of the nonlinear water fluxes (Ryu et al., 2009). 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).

HESSD
12, 7353–7403, 2015

Twelve grassland sites over southwest France, where in situ observations are available (the SMOSMANIA network, Calvet et al., 2007; Albergel et al., 2008), are used to compare the methods. These sites include various soil textures that can influence soil water transfers. Only surface soil moisture observations are assimilated. 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. The results of the DA experiments are presented in Sect. 3.2. In Sect. 3.2.1 the DA methods are compared using a synthetic 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. 3.2.2 the DA methods are tested using real in situ observations. Section 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 Sects. 2.1 and 2.2 respectively. The observations are described in Sect. 2.3. The DA methods are explained in Sect. 2.4. The performance diagnostics of the DA methods are given in Sect. 2.5. Finally, the experimental setup for the DA methods is presented in Sect. 2.6.

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 Dear-

HESSD
12, 7353–7403, 2015

dorff (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:

- 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 ($m^3\,m^{-3}$). 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. In the operational setup each grid-cell is split into 12 so-called "patches" corresponding to different surfaces, but for these experiments only grassland was present. The model points are such that the nearest observation to each site is always less than 6 km away.

HESSD
12, 7353–7403, 2015

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 Observations

2.3.1 Experiments with real 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 sites are part of the Soil Moisture Observing System Meteorological Automatic Network Integrated Application (SMOSMANIA) network (Calvet et al., 2007; Albergel et al., 2008). The observation sites are spaced approximately 45 km apart.

The SMOSMANIA network was selected partly because of the large variability in the soil textures between the different sites. In particular, the Sabres site and the Narbonne site represent opposite extremes of soil texture. This is linked to the ratio of the sand and clay contents. The Sabres soil has a high sand to clay ratio, while the Narbonne soil has a low sand to clay ratio. The Sabres soil was measured as having about 932 g Kg^{-1} of sand and 39 g Kg^{-1} of clay at 5 cm depth (with similar levels for deeper layers), while the Narbonne soil was measured as having 262 g Kg^{-1} of sand and 464 g Kg^{-1} of clay at 5 cm depth (Albergel et al., 2008). The other station soil types have combinations of sand and clay between these values (Albergel et al., 2008). The stations also contain

HESSD
12, 7353–7403, 2015

silt, which is an intermediate soil texture. 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. Our experiments demonstrate how this can influence 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 h assimilation window (06:00 UTC). The WG2 variable was represented by observations at 30 cm depth and these observations were used for verification only. 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. Draper et al. (2011) demonstrated that 5 cm deep SMOSMANIA observations are well correlated with the WG1 model simulation for a previous version of ISBA-A-gs. In this study, it was found that 30 cm deep observations were also well correlated with the WG2 simulation for a depth $d_2 = 0.8$ m. The average anomaly correlation coefficients (ACC, defined in Sect. 2.5) for the sites between the simulated and observed WG1 (WG2) variable is 0.53 (0.68). Table S1.1 in Supplement 1 shows the observed and modelled clay percentage, and the WG1 and WG2 ACC for each site.

A CDF matching technique was implemented in order to remove the bias from the observations relative to the model simulation. This transforms the observations such that the first and second moments match those of the model. The model climatology over the period 2007–2010 was used to scale the observations over the period 2008–2010. There were significant biases in the in situ observations (relative to the model simulation) prior to CDF matching. The CDF matching reduced the site-averaged WG1 (WG2) RMSE between the model simulation and the observations from 0.089 (0.062) $m^3\,m^{-3}$ to 0.060 (0.025) $m^3\,m^{-3}$, 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 gravimetric reference observations (Calvet et al., 2007). Overall, the RMSE after calibration was 0.038 $m^3\,m^{-3}$. This error includes instrumental errors and local representativeness errors (gravimetric samples were collected a few meters

HESSD
12, 7353–7403, 2015

around the probes). However, other errors of representivity were also likely in this study, since the observations were assumed to represent 8 km × 8 km model pixels. The observation error (σ^o) for WG1 was tested with values of $\lambda^o(w_{fc} - w_{wilt})\,m^3\,m^{-3}$, with λ^o equal to 0.5 or 0.35. 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 $(w_{fc} - w_{wilt})$ the two values of λ^o correspond to site averaged σ^o values of 0.030 and 0.044 $m^3\,m^{-3}$, which are either side of the RMSE measured by Calvet et al. (2007). The scaling by $(w_{fc} - w_{wilt})$ is based on the assumption of a linear relationship between the dynamic range of soil moisture values and the observation errors (Mahfouf et al., 2009).

2.3.2 Synthetic experiments

For the synthetic experiments the in situ observations were not used. Instead the synthetic WG1 observations were obtained from a model simulation with perturbed precipitation forcing (described in Sect. 2.6.2) and the addition of a random normally distributed observation error. Prior to DA, the RMSE of the model simulation in the synthetic experiments (compared with the synthetic observations) was roughly 10 % of the size of the model RMSE in the real experiments (compared with the real observations). Therefore the observation error for the synthetic experiments was specified to be 10 % of the higher value used in the real experiments ($\sigma^o = 0.05\,(w_{fc} - w_{wilt})$).

2.4 DA methods

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

$$\boldsymbol{x}^b(t_i) = M(\boldsymbol{x}^a(t_{i-1})), \quad (1)$$

HESSD
12, 7353–7403, 2015

where M is the (nonlinear) ISBA-A-gs model. The Kalman Filter analysis update is:

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

where $\boldsymbol{y}^{\mathrm{o}}$ is the assimilated observation and $\boldsymbol{y} = H(\boldsymbol{x})$ is the model predicted value of the observation. The nonlinear observation operator (H) transforms the model into observation space, which is simply a direct insertion i.e. $H = (1,0)$. This transformation indicates that only WG1 observations are assimilated. The model state and the observations are weighted using the Kalman gain (**K**):

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

where **H** is the linearized observation operator, **B** is the background-error covariance matrix and **R** is the observation-error covariance matrix. These matrices measure the expected errors and the covariances are a measure of the correlations in the errors between the different variables (i.e. between WG1 and WG2). The **R** matrix is assumed to be diagonal i.e. covariances equal to zero. Given that each point used in the experiments was independent, none of the DA methods could take into account background error-covariances between different gridpoints.

In the experiments the observation was present at the end of the assimilation window, so the background-error covariance needed to be propagated from the beginning to the end of the window. This is implicitly calculated via **H** for the SEKF (Sect. 2.4.1) and via the ensemble perturbations for the EnSRF (Sect. 2.4.2). A summary of the DA methods is given in Table 1.

2.4.1 SEKF

The Simplified Extended Kalman Filter (SEKF, Mahfouf et al., 2009) is based on the EKF (Jazwinski, 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

HESSD
12, 7353–7403, 2015

(Mahfouf et al., 2009). For this reason the SEKF uses the linear observation operator **H** to relate the observed quantity to the prognostic variables. Following the notation of Mahfouf (2010), there are two steps in the calculation of **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 (**M**). This linear model is a finite difference approximation between a perturbed and reference nonlinear model simulation:

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

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:

$$\mathbf{H}^{kl} = \mathbf{H}_1^k \mathbf{M}^l. \tag{5}$$

This formulation of **H** implicitly propagates the **B** matrix from the start of the assimilation window to the time of the observation at the end of the window ($\mathbf{HBH}^{\mathrm{T}} = \mathbf{H}_1\mathbf{MBM}^{\mathrm{T}}\mathbf{H}_1^{\mathrm{T}}$). 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 **M** is satisfactory but large enough to not incur significant computational round-off errors. The linear approximation can be measured 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), with values close to zero indicating linear behavior and values larger than one indicating very nonlinear behaviour. Duerinckx (2015); Draper et al. (2009) have demonstrated that the linear assumption is generally acceptable for other versions of ISBA, but in both studies the linear assumption did occasionally break down and was harmful to the analysis. Following Draper et al. (2009), we use a perturbation size of $10^{-4}(w_{\mathrm{fc}} - w_{\mathrm{wilt}})$ and a 24 h assimilation window. Draper et al. (2009)

HESSD
12, 7353–7403, 2015

demonstrated that during very nonlinear periods the $\mathbf{H}^{12}$ values can diverge, so it was necessary for them to impose a maximum limit on $\mathbf{H}^{12}$, which they set equal to 1.0. The same perturbation size and limit were used in our experiments.

The validity of the 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:

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

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 $(\sigma^{\mathrm{o}})^2$.

2.4.2 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. Each ensemble member is propagated using the nonlinear model. The deterministic analysis comes from the ensemble mean. The ensemble background-error covariance is defined as:

$$\mathbf{P}^{\mathrm{b}} = \frac{1}{m-1}\mathbf{X}^{\mathrm{b}}(\mathbf{X}^{\mathrm{b}})^{\mathrm{T}}. \tag{7}$$

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

$$\mathbf{X}^{\mathrm{b}} = \frac{1}{\sqrt{m-1}}\left[\delta\boldsymbol{x}^{\mathrm{b}}_{1}\ldots\delta\boldsymbol{x}^{\mathrm{b}}_{m}\right], \tag{8}$$

HESSD
12, 7353–7403, 2015

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

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 defined by:

$$\delta \boldsymbol{x}_j^{\mathrm{a}} = \delta \boldsymbol{x}_j^{\mathrm{b}} - \alpha \mathbf{K} \boldsymbol{y}_j, \tag{9}$$

where

$$\alpha = \frac{1.0}{1.0 + \sqrt{\frac{R}{HP^{\mathrm{b}}H^{\mathrm{T}} + R}}}. \tag{10}$$

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:

$$\mathbf{K}_{\mathrm{WG2}} = \mathbf{X}^{\mathrm{b}}_{\mathrm{WG1}}(\mathbf{X}^{\mathrm{b}}_{\mathrm{WG2}})^{\mathrm{T}} \left(\mathbf{X}^{\mathrm{b}}_{\mathrm{WG1}}(\mathbf{X}^{\mathrm{b}}_{\mathrm{WG1}})^{\mathrm{T}} + (\sigma^{\mathrm{o}})^2 \right)^{-1}, \tag{11}$$

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

2.5 Performance diagnostics

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:

HESSD
12, 7353–7403, 2015

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

– ACC: $\frac{\sum_{i=0}^{N}(H(\boldsymbol{x}_i)-\overline{H(\boldsymbol{x})})(\boldsymbol{y}_i^{\mathrm{o}}-\overline{\boldsymbol{y}^{\mathrm{o}}})}{\sqrt{\sum_i(H(\boldsymbol{x}_i)-\overline{H(\boldsymbol{x})})^2(\boldsymbol{y}_i^{\mathrm{o}}-\overline{\boldsymbol{y}^{\mathrm{o}}})^2}}$;

– Bias: $\sum_{i=0}^{N}\frac{\boldsymbol{y}_i^{\mathrm{o}}-H(\boldsymbol{x}_i)}{N}$.

The time t_i is the daily time, with t_0 equal to 1 January 2008 and t_N equal to 1 January 2011. The ACC climatologies ($\overline{H(\boldsymbol{x}^{\mathrm{a}})}$ and $\overline{\boldsymbol{y}^{\mathrm{o}}}$) are calculated as a 3 month moving average, which takes into account seasonal variability.

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. Moreover, Muñoz Sabater et al. (2007) already estimated the computational cost of similar algorithms 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. In the DA experiments, the performance of the model state $\boldsymbol{x}$ is always measured against independent observations of WG2. The diagnostics are averaged over the period 2008–2010.

2.6 Experimental setup

The DA methods are compared over the period 2008–2010. The observation and background error calibrations are important for optimising the performance of DA methods

HESSD
12, 7353–7403, 2015

(Reichle et al., 2008). The SEKF calibration is described in Sect. 2.6.1 and the EnSRF calibration in Sect. 2.6.2. The ensemble bias correction technique is defined in Sect. 2.6.3. Finally, a list of the experiments that were performed is summarised in Sect. 2.7.

2.6.1 SEKF calibration

The represention of model and forcing errors within DA methods can significantly influence their performance. 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 variances (σ^{b}) for WG1 and WG2 were tuned to produce the best ACC. The WG1 and WG2 background-error standard deviations were tuned with sizes:

$$\sigma^{b}_{WG1} = \lambda^{b}_{1} \times (w_{fc} - w_{wilt}),$$
$$\sigma^{b}_{WG2} = \lambda^{b}_{2} \times (w_{fc} - w_{wilt}), \quad (12)$$

with λ^{b}_{1} and λ^{b}_{2} varying between 0.0 and 0.5, in steps of 0.05. In the synthetic experiments, the σ^{b} values were calibrated with values a 10th of the size of those above (as explained in Sect. 2.3.2). The background error variances were scaled by ($w_{fc} - w_{wilt}$) for each site. This assumes that a linear relationship exists between the range of possible soil moisture values and the background errors (Mahfouf et al., 2009).

2.6.2 EnSRF calibration

The background errors in the EnSRF are captured by the ensemble background-error covariance matrix ($\mathbf{P}^{b}$), but this does not include the contribution from model errors and forcing errors. 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 our experiments, Gaussian noise ϵ was added to

HESSD
12, 7353–7403, 2015

each ensemble member after the daily analysis update. The values of ϵ for WG1 and WG2 were tuned to produce the largest ACC, with sizes:

$$\epsilon_{WG1} = \lambda_1^b \times (w_{fc} - w_{wilt}),$$
$$\epsilon_{WG2} = 0.1\lambda_2^b \times (w_{fc} - w_{wilt}), \quad (13)$$

with λ_1^b and λ_2^b varying between 0.0 and 0.5, in steps of 0.05. Time-correlated additive inflation was implemented using a 1st order auto-regressive 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 (Mahfouf, 2007; Muñoz Sabater et al., 2007; Reichle et al., 2002) and is consistent with the longer time correlations of the WG2 variable compared with WG1.

The ensemble background-error covariance matrix was further modified to allow for random errors in the precipitation forcing. Following Muñoz Sabater et al. (2007), the precipitation was perturbed using Gaussian noise with a standard deviation equal to 50 % of the precipitation ($\epsilon_{Pr} = 50\,\%\,Pr$). The pdf was truncated in order to prevent negative precipitation values. The precipitation forcing was perturbed for each ensemble member and for each hour of the 24 h assimilation window. 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. In the real experiments, the performance of the EnSRF with perturbed precipitation was compared with the performance without perturbed precipitation.

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.

2.6.3 EnSRF bias correction

A bias correction technique was tested on the ensemble DA methods as a means of correcting the biases caused by the ensemble perturbations themselves (Ryu et al.,

HESSD
12, 7353–7403, 2015

2009). The perturbation bias correction uses the forecast from the previous analysis ensemble mean as an anchor to modify the background ensemble:

$$\boldsymbol{x}_j^{\mathrm{b}}(t_i) \leftarrow \boldsymbol{x}_j^{\mathrm{b}}(t_i) - \overline{\boldsymbol{x}}_j^{\mathrm{b}}(t_i) + M(\overline{\boldsymbol{x}}^{\mathrm{a}}(t_{i-1})), \quad \text{for } j = 1, \ldots, m. \tag{14}$$

Equation (14) 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.

2.7 Experiment list

The list of experiments is displayed in Table 2. 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. 14) was then tested on this ensemble, which is labelled as $\mathrm{Ens}_{\mathrm{bc}}$. Thereafter the synthetic and real DA experiments are denoted by the subscripts S and R respecively. For each experiment the calibrated error variances are specified. For the real experiments the EnSRF was tested with three different configurations: $\mathrm{EnSRF}_{\mathrm{R1}}$ is the baseline EnSRF without perturbed precipitation forcing nor bias correction. The $\mathrm{EnSRF}_{\mathrm{R2}}$ and $\mathrm{EnSRF}_{\mathrm{R3}}$ experiments include perturbed precipitation forcing and bias correction respectively. Note that in the synthetic experiments, the $\mathrm{EnSRF}_{\mathrm{S}}$ was designed to capture the precipitation forcing errors perfectly. The same precipitation error specification was used to estimate the precipitation errors in the real experiments ($\mathrm{EnSRF}_{\mathrm{R2}}$).

3 Results

3.1 Investigating the perturbation bias

An ensemble of model trajectories was perturbed by adding random perturbations to WG2 with standard deviation 0.025 $\mathrm{m}^3\,\mathrm{m}^{-3}$. This is a similar order of magnitude to the

HESSD
12, 7353–7403, 2015

values used to calibrate the ensemble background-error covariance in the real experiments (see Sect. 2.6.2). The result for the perturbed 20 member ensemble is defined as experiment Ens in Table 2. 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. 2.5. A large bias of of $-0.0049\,m^3\,m^{-3}$ has been introduced, which is not expected because the perturbations are randomly sampled with zero mean. The origin of this bias is investigated by linking the physical processes that underpin the bias to changes in the ensemble spread.

The site-averaged (averaged over the 12 stations) and time-averaged water content of the total reservoir (WG1 + WG2 + WG3) is 243 mm for the open loop and 239 mm for Ens. This water loss of 4 mm (of Ens compared with the open loop) represents a small fraction of the total reservoir. However, this dry bias is strongly dependent on the soil texture and the season: the sites with clay soils have a greater dry bias than the sites with sandy soils, particularly during the summer period. It therefore makes sense to compare the results for Sabres with Narbonne, given that Sabres is very sandy while Narbonne has a high clay content. Indeed, between July and September, the average bias for Narbonne (Sabres) is 13 (3) mm. Between January and March the average bias for Narbonne (Sabres) is 7 (3) mm.

The impact of the perturbed ensemble on each individual layer is demonstrated by Fig. 2, 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 $m^3\,m^{-3}$ 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

HESSD
12, 7353–7403, 2015

evapotranspiration (evaporation + transpiration) and drainage are shown in Fig. 2d and e 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 peaks in May and June (Fig. 2d). In contrast to evapotranspiration, the drainage is most active during the winter and is absent during the late summer/early autumn period (Fig. 2e); 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. 2d). This effect is also clearly evident in Fig. 2f, 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. 2e). The annually averaged discrepancy between the Ens and open loop total soil water depletion is about 4 mm, 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. 3a 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

HESSD
12, 7353–7403, 2015

spread. The Narbonne soil has a much larger ensemble spread than the Sabres soil, particularly in autumn.

Now the seasonal changes in the bias can be related to changes in the ensemble spread and the soil texture. The Ens WG2 ensemble mean is clearly negatively biased (compared with the open loop) for Narbonne during much of the period (Fig. 3b), 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. 4a 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. 4b. 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. 3a with b).

The impact of the ensemble perturbations on drainage is most significant near the field capacity. Figure 4c and d shows 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. 2e.

The simple bias correction scheme (Eq. 14) was tested on the ensemble and the results are also shown in Table 2 (labelled as $\mathrm{Ens}_{\mathrm{bc}}$). The bias has been reduced to less

HESSD
12, 7353–7403, 2015

than a 10th of the size and the RMSE reduced by half compared with the original Ens. Figure 2a–c shows 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.

The soil water depletion is shown in Fig. 2d–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. 5 for (a) Sabres; and (b) Narbonne. The bias is much greater for Narbonne than Sabres (comparing Fig. 3a with b). 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. 3.2.2.

3.2 DA experiments

3.2.1 SEKF vs. EnSRF: synthetic random precipitation errors

The EnSRF and the SEKF were first tested with only errors in the precipitation forcing. Recall that the observations were taken from a simulation with perturbed precipitation with a small random observation error (Sect. 2.3.2). For this experiment the EnSRF used a perfect stochastic representation of the precipitation forcing errors. The SEKF cannot capture these forcing errors directly, but instead the climatological background-error variances were calibrated to produce the best performance. The SEKF used the same unperturbed precipitation as the open loop.

Firstly, the monthly and daily precipitation for the EnSRF and the open loop are described for both Sabres and Narbonne. The Sabres site is the furthest west of the 12 sites and is near the Atlantic coast, while Narbonne is the furthest east of the 12 sites and is on the Mediterranean coast. The precipitation undergoes large seasonal changes during the year. Generally the precipitation is greater during the winter than

HESSD
12, 7353–7403, 2015

the summer, but there is much inter-annual variability. Since Sabres is more exposed than Narbonne to weather systems arriving from the Atlantic, Sabres receives more than twice the average rainfall, with a mean monthly rainfall of 99 mm, compared with 44 mm for Narbonne. The EnSRF is able to capture the larger perturbations of the wetter stations with a greater ensemble spread (not shown).

The time-averaged and site-averaged WG2 RMSE of the 20 member EnSRF is labelled as $EnSRF_S$ in Table 2. 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 3. A gradual improvement in the $EnSRF_S$ is apparent in Table 3 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 and no significant bias was introduced. In fact the bias is only significant when the ensemble spread is large, which was not the case for the synthetic experiments (not shown). Therefore implementing the perturbation bias correction was not necessary.

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 2. 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 4 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. 6a–c. The open loop RMSE is greatest in the spring and autumn seasons (Fig. 6a). 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

HESSD
12, 7353–7403, 2015

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.

The monthly average WG2 Kalman gain for $EnSRF_S$ is displayed in Fig. 7. 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. 7 both before and after the 1.0 limit is imposed on the model Jacobian. In Sect. 2.4.1 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 oversensitive 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. 6a and b).

3.2.2 Real experiments

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 2. The $EnSRF_{R1}$ method does perform significantly better than the open loop (labelled as OL_R in Table 2). But a large dry bias has been introduced, which is about 25 % of the RMSE and is consistent with the size of the dry bias introduced by experiment Ens in Table 2. Note that the very small bias for OL_R is present because the CDF matching (for bias correcting the observations) was applied over the period 2007–2010, but the results are taken over 2008–2010. The $EnSRF_{R1}$ method was tested with various ensemble sizes and the results are shown in Table 3. There is no improvement beyond an ensemble size of 20 members.

Figure 3c and d is equivalent to Fig. 3a and b but instead shows the $EnSRF_{R1}$ ensemble, the open loop and the observations for 2009. During December and January, the ensemble for Narbonne (Fig. 3d) has a similar negatively skewed spread to Fig. 3b,

HESSD
12, 7353–7403, 2015

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 $SEKF_{R1}$ performance is presented in Table 2. In terms of RMSE, $SEKF_{R1}$ performs marginally better than $EnSRF_{R1}$, while $EnSRF_{R1}$ has a slightly higher ACC than $SEKF_{R1}$. The $SEKF_{R1}$ method is also affected by a negative bias, which is about half the size of the $EnSRF_{R1}$ bias. The $SEKF_{R1}$ 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 $SEKF_{R1}$ bias.

Figure 8a and b shows contour plots of the $EnSRF_{R1}$ RMSE and ACC respectively, for the range of additive perturbations used to calibrate the method. Figure 8c and d shows equivalent contour plots for the $SEKF_{R1}$. 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. The $SEKF_{R1}$ results are less noisy than the $EnSRF_{R1}$ results (Fig. 8a 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, $SEKF_{R1}$ and $EnSRF_{R1}$ in Fig. 6d–f respectively. The RMSE in all cases is highest in June and October (Fig. 6d), 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 $SEKF_{R1}$ from April to July as a result of the perturbation bias during this period (evident in Fig. 6f). The superior $EnSRF_{R1}$ ACC from July to September is explained below.

HESSD
12, 7353–7403, 2015

The WG2 Kalman gains for $EnSRF_{R1}$ and the $SEKF_{R1}$ are shown in Fig. 7b. The $SEKF_{R1}$ 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. 6d). The $SEKF_{R1}$ Kalman gain is shown in Fig. 7b 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 $SEKF_{R1}$ ACC during July, August and September (Fig. 6b).

The impact of precipitation forcing perturbations on the EnSRF is investigated. This experiment is labelled as $EnSRF_{R2}$ in Table 2. 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. 3.2.1).

Finally, the bias correction scheme is tested on the EnSRF. This experiment is labelled as $EnSRF_{R3}$ in Table 2. The large bias in Table 2 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. 5).

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

HESSD
12, 7353–7403, 2015

soil moisture texture (Sect. 4.3) and flow-dependence of the DA methods (Sect. 4.4). Section 4.5 discusses the influence of the choice of model on the results.

4.1 Stochastic precipitation error representation

The experiments in Sect. 3.2.1 were designed to assess the advantage gained by a perfect stochastic representation of random precipitation errors in the EnSRF over additive background errors in the SEKF. Clearly the EnSRF benefited from the direct representation of the errors. 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 2).

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 behavior of soil moisture dynamics, which dampens perturbations in the initial conditions. These conclusions agree with our findings. The results from our study suggest that other sources of errors may also limit the usefulness of precipitation uncertainty information. Important sources of errors in this study include the errors in the model physics, the linear assumptions made by the DA methods and the errors in representing observations over a 64 km squared grid.

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. Even if the specified rainfall errors applied in this work were correct, they would only make up 10 % of the total RMSE of the open loop. The percentage scaling used to generate the perturbations in the experiments could only estimate random errors in accumulations. It 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. Carrera et al., 2015; Maggioni et al., 2014, 2012, 2011; Clark and Slater, 2006; Hossain and

HESSD
12, 7353–7403, 2015

Anagnostou, 2006). 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 forcing errors, 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 ($\boldsymbol{y} - \boldsymbol{y}^{o}$) with the sum of the background and observation-error covariances ($\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \mathbf{R}$) (Desroziers et al., 2005). Figure 9 shows the histograms of the innovations normalized by $\sqrt{\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \mathbf{R}}$ for the $\mathrm{EnSRF_S}$ (a) and the $\mathrm{SEKF_S}$ (b) for the synthetic experiments. The pdf for $\mathrm{EnSRF_S}$ 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 $\mathrm{SEKF_S}$ 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 background-error covariance calibration that optimizes the analysis is not the same as the calibration that agrees best with the Kalman Filter theory.

In the real experiment neither method had a perfect representation of the background errors. Both methods used an average value of $\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \mathbf{R}$ about 4 times larger than $(\boldsymbol{y} - \boldsymbol{y}^{o})^{2}$ (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 behavior of evapotranspiration and drainage. The ISBA-A-

HESSD
12, 7353–7403, 2015

gs evapotranspiration and drainage functions are described by Masson et al. (2013) and Mahfouf and Noilhan (1996) respectively. 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 perurbation bias for their system, when our study discovered a negative bias. Ryu et al. (2009) also noted a reduction of the ensemble spread when applying the bias correction, which agrees with our study.

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 3). 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 1D-EnKF approach was implemented, where the analysis is calculated independently for each gridpoint. It is likely that sampling errors would be much more important for a 2D-EnKF approach, where background-error covariances between gridpoints are taken into account, due to the much larger number of degrees of freedom.

4.3 Different soil textures

It was found that the soil texture, in particular the ratio of sand to clay, had a significant influence on the model behaviour. Clay soils are more sensitive than sandy soils to perturbations in the initial conditions (Calvet and Noilhan, 2000). This occurs because water travels faster through sand and as a result the atmospheric forcing tends to dominate over perturbations in the model state. This made the clay soils more vulnerable to the nonlinearity issues encountered in this study. The increased sensitivity of clay soils

HESSD
12, 7353–7403, 2015

compared with sandy soils was reflected by a larger ensemble spread for the ensemble experiments. In particular, Narbonne had a much larger ensemble spread than Sabres (see Fig. 3).

The correlation coefficient between the modelled clay percentage and the size of the negative perturbation bias (normalized by the RMSE) for the 12 sites was calculated as 0.53 for the experiment Ens and 0.04 for the experiment $EnSRF_{R1}$. The correlation was much smaller for $EnSRF_{R1}$, which was probably because other sources of errors were also important. The list of the biases (normalized by the RMSE) and the clay percentages are shown in S1.1 of Supplement 1. It is recommended that similar experiments are performed on a much larger number of model points, in order to calculate a more reliable estimate of the correlation between the clay content and the perturbation bias.

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 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. However, land surface models are not chaotic, unlike atmospheric models. The most important contributions to soil moisture errors come from the model and atmospheric forcing errors, rather than the errors in the initial conditions. Therefore it is unlikely that the EnSRF could gain much advantage over the SEKF without having a superior representation of these errors. This would explain why the EnSRF performed better than the SEKF in the synthetic experiments but not in the real experiments.

HESSD
12, 7353–7403, 2015

4.5 Land surface model

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. At times they found that the assimilation of surface observations influenced only the top part of the root-zone with ISBA-DF. The multi-layer model captured the vertical profile of the root zone soil moisture, which was not possible with 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 in situ soil moisture DA experiments with various soil conditions. In particular, the different sites were chosen for their variability in sand and clay concentrations, 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, which is based on the force-restore method of Deardorff (1977).

Synthetic experiments were designed to assess the advantage the EnSRF could gain over the SEKF by using a perfect stochastic representation of random precipi-

HESSD
12, 7353–7403, 2015

tation forcing errors. By design the SEKF cannot represent stochastic forcing errors directly, but it may compensate for such errors by using larger background-error variances. Indeed, the EnSRF performance was superior to the SEKF, both in terms of analysis RMSE and ACC. In the real experiments, the same rainfall error specification did not improve the analysis. It is likely that the actual precipitation errors were underestimated. It is also possible that other sources of errors could have limited the transformation of precipitation uncertainty information into improved soil moisture scores. It is recommended that more sophisticated rainfall error models are tested, which should take into account the intermittent and non-stationary nature of precipitation errors.

In the real experiments the SEKF and the EnSRF performed similarly and improved on the open loop. 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. This led to excessive analysis increments. For the EnSRF, a significant dry bias was found, mainly as a result of the ensemble perturbations causing excessive evapotranspiration during dry conditions. These problems were less detrimental to sandy soils than clay soils because of the smaller sensitivity of sandy soils to perturbations in the initial conditions.

A bias correction scheme was tested on the EnSRF and although this removed the perturbation bias, it failed to improve the analysis. The bias correction led to spurious increases in drainage and evapotranspiration, induced by the wetter soil moisture states. Consequently it reduced the ensemble spread, which degraded the analysis. For this reason 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).

HESSD
12, 7353–7403, 2015

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 both vegetation and soil texture. 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.

The Supplement related to this article is available online at doi:10.5194/hessd-12-7353-2015-supplement.

Acknowledgements. This work is a contribution to the IMAGINES (grant agreement 311766) and CORE-CLIMAX (grant agreement 313085) projects, co-funded by the European Commission within the Copernicus initiative in FP7.

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. 7358, 7360

Balsamo, G., Bouyssel, F., and Noilhan, J.: A land data assimilation system for soil moisture and temperature, Q. J. Roy. Meteor. Soc., 130, 895–915, 2004. 7364

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. 7359

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. 7359

HESSD
12, 7353–7403, 2015

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the Ensemble Kalman Filter, Mon. Weather Rev., 126, 1719–1724, 1998. 7366

Calvet, J.-C. and Noilhan, J.: From near-surface to root-zone soil moisture using year-round data, J. Hydrometeorol., 1, 393–411, 2000. 7357, 7381

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. 7359, 7383

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. 7358, 7360, 7361, 7362

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. 7356, 7379, 7381, 7384

Clark, M. and Slater, A.: Probabilistic quantitative precipitation estimation in complex terrain, J. Hydrometeorol., 7, 3–22, doi:10.1175/JHM474.1, 2006. 7379

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. 7357

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. 7357

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. 7356

Deardorff, J.: A paramaterization of ground surface moisture for use in atmospheric prediction models, J. Appl. Meteorol., 16, 1182–1185, 1977. 7358, 7383

Decharme, B., Boone, A., Delire, C., and Noilhan, J.: Local evaluation of the interaction between soil biosphere atmosphere soil multilayer diffusion scheme using four pedotransfer functions, J. Geophys. Res., 116, D20126, doi:10.1029/2011JD016002, 2011. 7356, 7383, 7385

HESSD

12, 7353–7403, 2015

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. 7380

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. 7355

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. 7356, 7364, 7380

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. 7355, 7361

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. 7356

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. 7357

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. 7364

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. 7360

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. 7355

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. 7365

HESSD
12, 7353–7403, 2015

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. 7359

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. 7368

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. 7379

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. 7379

Houtekamer, P. and Mitchell, H.: Data assimilation using an Ensemble Kalman Filter technique, Mon. Weather Rev., 126, 796–811, 1998. 7381

Jazwinski, A. H.: Stochastic Processes and Filtering Theory, Academic Press, New York, USA, 150–158, 1970. 7363

Kalman, R. E.: A new approach to linear filtering and prediction problems, J. Basic Eng.-T. ASME, 82, 35–45, 1960. 7362

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. 7355

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. 7379

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. 7379, 7381, 7384

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. 7379

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, avail-

HESSD
12, 7353–7403, 2015

able at: http://www.ecmwf.int/sites/default/files/MAHFOUF_ECMWF2014_web.pdf, last access: September 2014. 7380
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. 7369
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. 7364
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. 7381
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. 7355, 7356, 7362, 7363, 7364, 7368
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. 7355, 7381
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. 7356, 7367, 7369
Noilhan, J. and Mahfouf, J.-F.: The ISBA land surface parameterisation scheme, Global Planet. Change, 13, 145–159, 1996. 7358, 7383
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. 7383
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. 7360
Reichle, R., Walker, J., Koster, R., and Houser, P.: Extended vs ensemble Kalman filtering for Land Data Assimilation, J. Hydrometeorol., 3, 728–740, 2002. 7356, 7369

HESSD
12, 7353–7403, 2015

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. 7357

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. 7356, 7357, 7368

Ryu, D., Crow, W., Zhan, X., and Jackson, T.: Correcting unintended perturbation biases in hydrologic data assimilation, J. Hydrometeorol., 10, 734–750, 2009. 7357, 7369, 7381

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. 7360

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., and Kerr, Y.: Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, Nord. Hydrol., 38, 1–20, doi:10.2166/nh.2007.029, 2007. 7355

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. 7364

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002. 7357, 7366

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. 7356

HESSD

12, 7353–7403, 2015

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 **H**	NA
EnSRF	Ensemble	Implicitly via ensemble	Eq. (9)

HESSD
12, 7353–7403, 2015

Table 2. 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.

Exp.	Calibration: Obs. λ^o	WG1 λ_1^b	WG2 λ_2^b	Add. criteria	WG2 RMSE ($m^3\,m^{-3}$) $\times 10^3$	WG2 ACC	WG2 Bias ($m^3\,m^{-3}$) $\times 10^3$
Ens	–	–	0.025	–	9	0.97	−4.9
Ens_{bc}	–	–	0.025	Bias correct	4	0.99	0.6
OL_S	–	–	–	–	2.2	0.995	0.0
$EnSRF_S$	0.05	–	–	ϵ_{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	ϵ_{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
$SEKF_{R1}$	0.5	0.25	0.25	–	20.1	0.716	−2.21

HESSD
12, 7353–7403, 2015

Table 3. 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 $(m^3 m^{-3}) \times 10^3$	$EnSRF_{R1}$ WG2 RMSE $(m^3 m^{-3}) \times 10^3$	$EnSRF_S$ 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

HESSD
12, 7353–7403, 2015

Table 4. SEKF$_S$ performance for various calibrations. The WG1 (λ_1^b) and WG2 (λ_2^b) calibrated values are shown. The optimal value is in bold font.

λ_1^b	λ_2^b	WG2 RMSE ($m^3\,m^{-3}$) $\times 10^3$
0.03	0.01	2.2
0.04	**0.02**	**1.8**
0.08	0.04	2.3
0.16	0.015	2.1

HESSD
12, 7353–7403, 2015

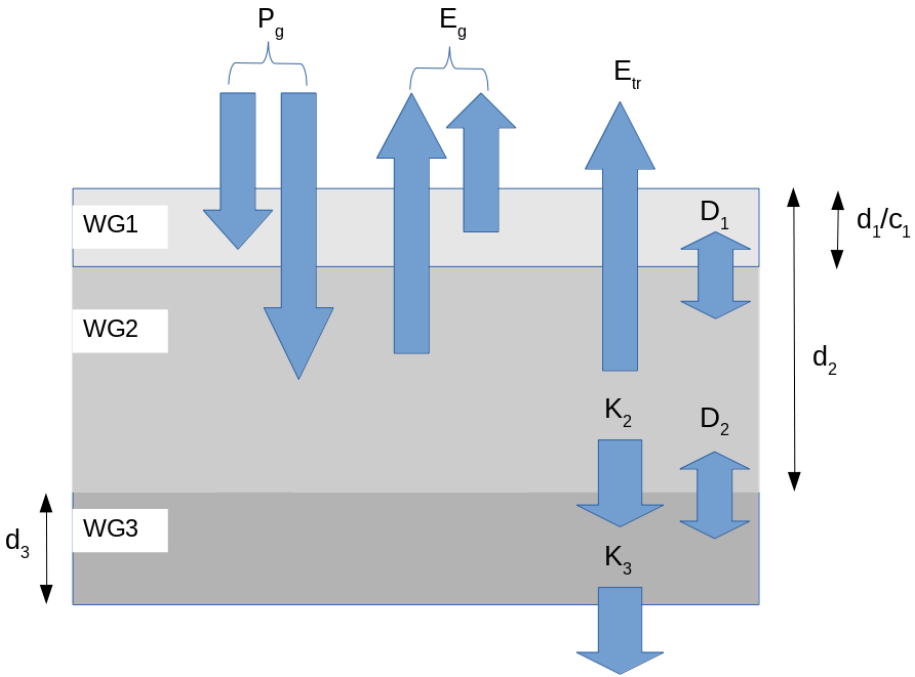

Figure 1. The soil moisture fluxes for the three-layer version of ISBA. The variables P_g, E_g and E_{tr} represent the precipitation, bare soil evaporation and transpiration respectively. The fluxes K and D represent the drainage and diffusion at the bottom of the layer.

HESSD
12, 7353–7403, 2015

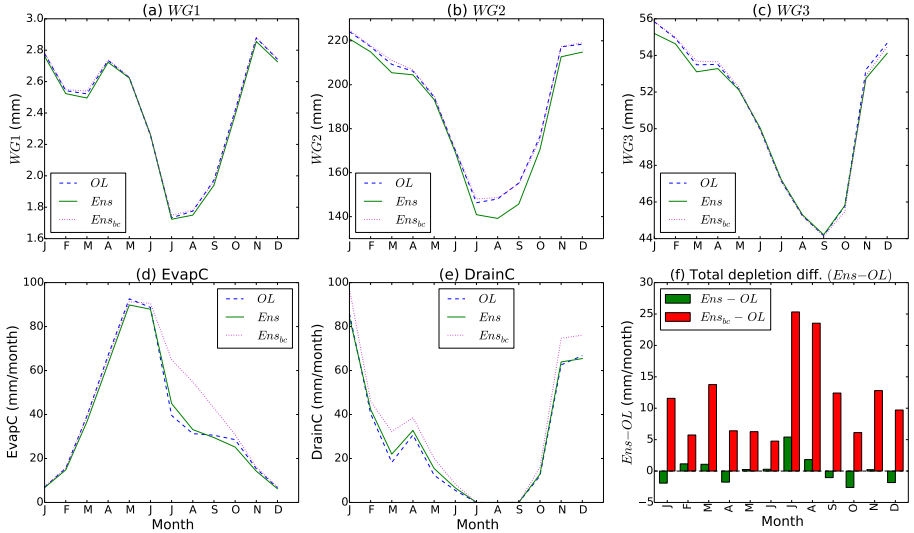

Figure 2. Monthly averaged and station averaged net water for **(a)** WG1, **(b)** WG2 and **(c)** WG3 and for the open loop, the ensemble mean (Ens) and the bias corrected ensemble mean (Ens_{bc}). Station averaged and monthly **(d)** evapotranspiration; **(e)** drainage and **(f)** total soil water depletion differences between ensemble mean and the open loop. Results are all averaged over the period 2008–2010.

HESSD
12, 7353–7403, 2015

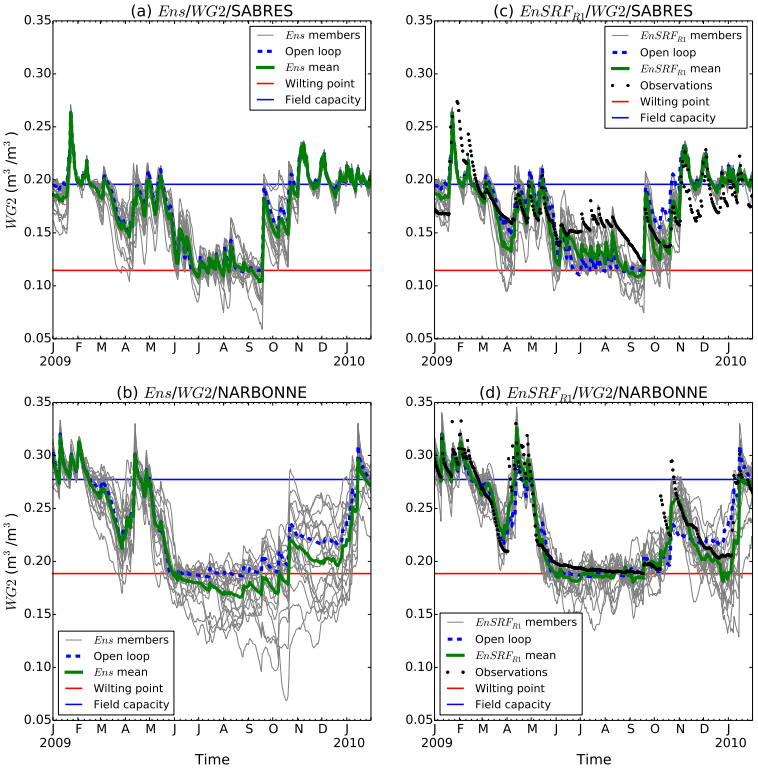

Figure 3. WG2 ensemble members, open loop and the ensemble mean for the sites (top) Sabres; and (bottom) Narbonne, over 2009. **(a)** and **(b)** show the results for experiment Ens (no DA), while **(c)** and **(d)** show the analysis ensemble members for the the experiment $EnSRF_{R2}$ (with observations). The field capacity and the wilting point are also shown for each site.

HESSD
12, 7353–7403, 2015

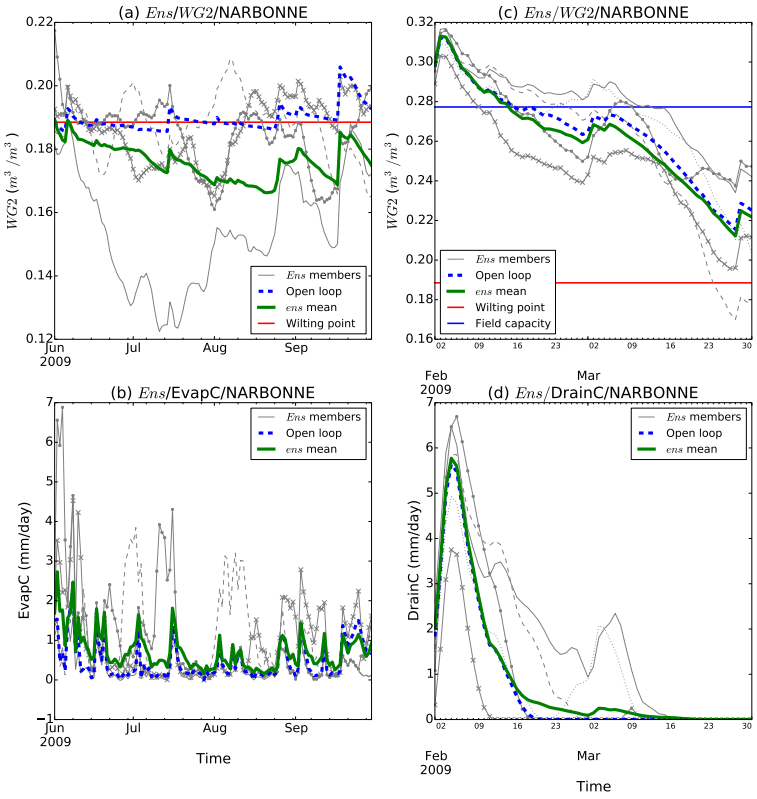

Figure 4. Four of the Ens ensemble members, open loop and the Ens ensemble mean for **(a)** WG2; and **(b)** Evapotranspiration; for Narbonne between June and September 2009. **(c)** and **(d)** are equivalent to **(b)** and **(c)** but instead show WG2 and drainage respectively, for February and March 2009.

HESSD
12, 7353–7403, 2015

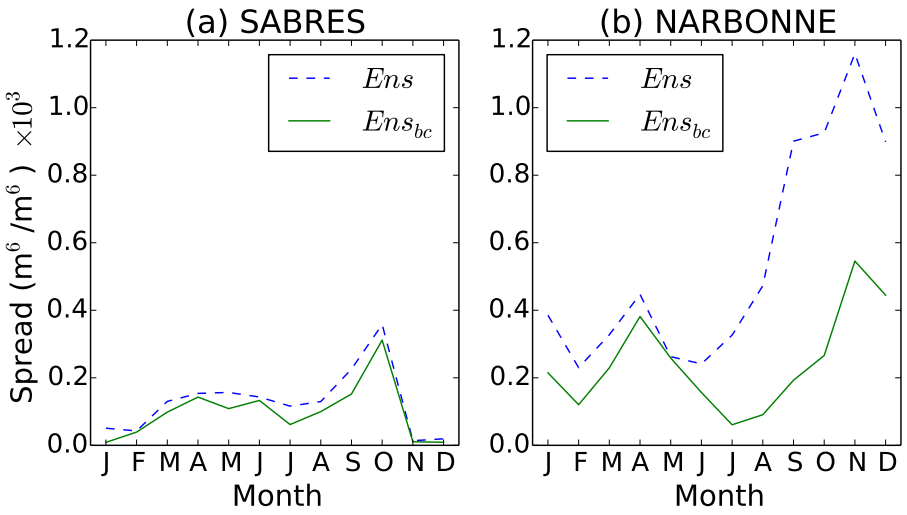

Figure 5. WG2 monthly averaged Ens and Ens_{bc} ensemble spread, for **(a)** Sabres; and **(b)** Narbonne. The results are averaged over 2008–2010.

HESSD
12, 7353–7403, 2015

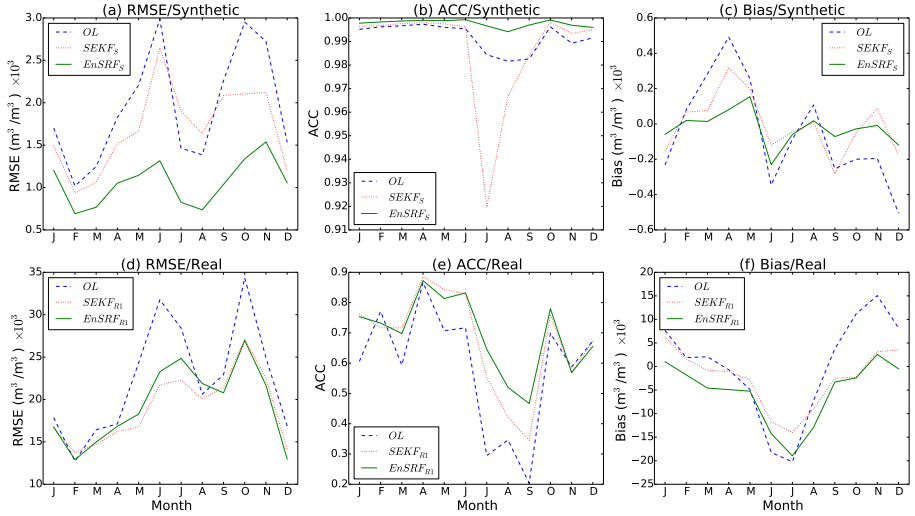

Figure 6. Site averaged and monthly averaged **(a)** WG2 RMSE; **(b)** WG2 ACC; and **(c)** WG2 bias; for the open loop, $SEKF_S$ and $EnSRF_S$ for the synthetic experiments. Plots **(d–f)** are equivalent to plots **(a–c)** but instead show the results for the open loop, $SEKF_R$ and $EnSRF_{R1}$ for the real experiments. The results are averaged over 2008–2010.

HESSD
12, 7353–7403, 2015

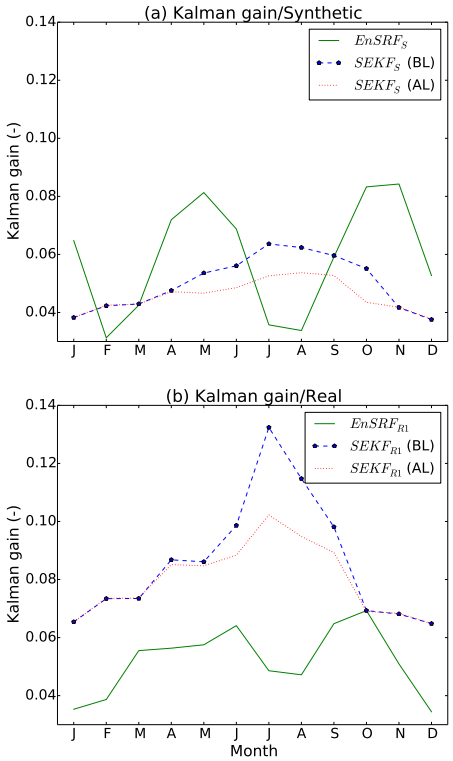

Figure 7. Station and monthly averaged WG2 Kalman gains for **(a)** $EnSRF_S$ and the $SEKF_S$ before (BL) and after (AL) the Jacobian limit of 1.0 is imposed. Plot **(b)** is equivalent to plot **(a)** but instead shows the results for $EnSRF_{R1}$ and $SEKF_R$. The results are averaged over 2008–2010.

HESSD
12, 7353–7403, 2015

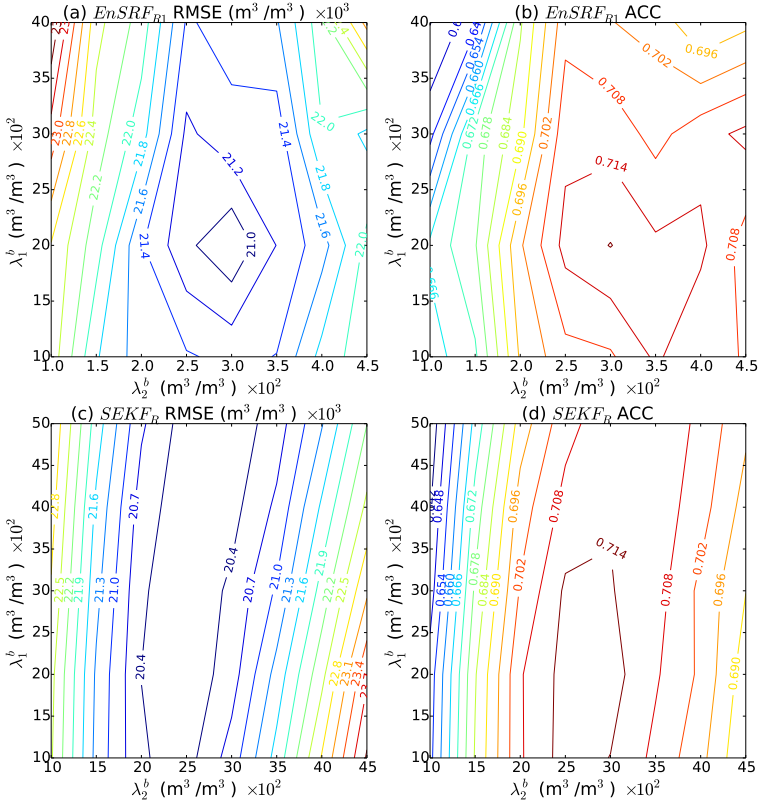

Figure 8. Contour plot showing the $EnSRF_{R1}$ **(a)** WG2 RMSE; and **(b)** WG2 ACC; for the range of additive inflation values used to calibrate the $EnSRF_{R1}$. Plots **(c)** and **(d)** show the equivalent contour plots to **(a)** and **(b)** for $SEKF_R$, with the range of background-error variances used to calibrate $SEKF_R$. The results are averaged over 2008–2010 and over the 12 sites.

HESSD
12, 7353–7403, 2015

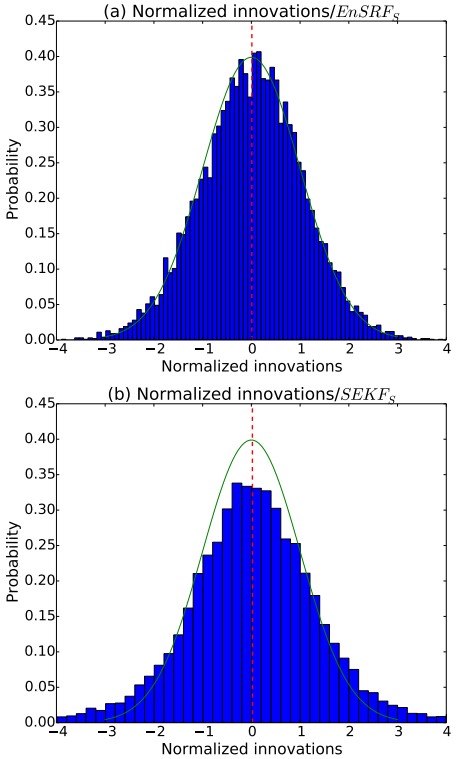

Figure 9. Histogram of normalized innovations for the 20 member $EnSRF_S$ **(a)** and the $SEKF_S$ **(b)** for the synthetic experiments. The green line shows the normal distribution and the red dashed line shows the mean.

HESSD
12, 7353–7403, 2015