JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, D20104, doi:10.1029/2008JD011650, 2009

An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme

C. S. Draper,[1] J.-F. Mahfouf,[2] and J. P. Walker[1]

Received 21 December 2008; revised 3 March 2009; accepted 24 June 2009; published 20 October 2009.

[1] An Extended Kalman Filter (EKF) for the assimilation of remotely sensed near-surface soil moisture into the Interactions between Surface, Biosphere, and Atmosphere (ISBA) model is described. ISBA is the land surface scheme in Météo-France's Aire Limitée Adaptation Dynamique développement InterNational (ALADIN) Numerical Weather Prediction (NWP) model, and this work is directed toward providing initial conditions for NWP. The EKF is used to assimilate near-surface soil moisture observations retrieved from C-band Advanced Microwave Scanning Radiometer (AMSR-E) brightness temperatures into ISBA. The EKF can translate near-surface soil moisture observations into useful increments to the root-zone soil moisture. If the observation and model soil moisture errors are equal, the Kalman gain for the root-zone soil moisture is typically 20–30%, resulting in a mean net monthly increment for July 2006 of 0.025 m^3 m^{-3} over ALADIN's European domain. To test the benefit of evolving the background error, the EKF is compared to a Simplified EKF (SEKF), in which the background errors at the time of the analysis are constant. While the Kalman gains for the EKF and SEKF are derived from different model processes, they produce similar soil moisture analyses. Despite this similarity, the EKF is recommended for future work where the extra computational expense can be afforded. The method used to rescale the near-surface soil moisture data to the model climatology has a greater influence on the analysis than the error covariance evolution approach, highlighting the importance of developing appropriate methods for rescaling remotely sensed near-surface soil moisture data.

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

1. Introduction

[2] Soil moisture can have a strong influence on Numerical Weather Prediction (NWP) forecasts, both at short [*Baker et al.*, 2001; *Drusch and Viterbo*, 2007] and medium range [*Zhang and Frederiksen*, 2003; *Fischer et al.*, 2007]. Currently, soil moisture is initialized in most operational NWP models based on errors in short-range forecasts of low-level humidity and temperature [e.g., *Giard and Bazile*, 2000; *Hess*, 2001; *Bélair et al.*, 2003]. While these schemes can in general produce reasonable boundary layer forecasts [*Drusch and Viterbo*, 2007], they assume a causative relationship between low-level atmospheric forecast errors and local soil moisture errors. As a result, soil moisture is often adjusted to compensate for errors elsewhere in the model, resulting in soil moisture fields that are frequently unrealistic [*Seuffert et al.*, 2004; *Draper and Mills*, 2008]. The accumulation of model errors in surface variables also makes it difficult to diagnose the source of these errors. Additionally, these schemes cannot be sensibly applied to situations where the local soil moisture–atmospheric boundary layer feedback is weak; for example, during periods of strong advection, or weak radiative forcing. The effectiveness of a soil analysis based on screen-level variables is also limited by the availability of screen-level observations, which are particularly sparse across much of the Southern Hemisphere. A particularly promising approach to addressing some of the above mentioned shortcomings is the possibility of assimilating remotely sensed near-surface soil moisture into NWP models [e.g., *Seuffert et al.*, 2004; *Balsamo et al.*, 2007; *Scipal et al.*, 2008]. This approach is explored here, using an Extended Kalman Filter (EKF) to assimilate remotely sensed near-surface soil moisture into Météo-France's Aire Limitée Adaptation Dynamique développement InterNational (ALADIN) NWP model.

[3] Recent interest in the assimilation of remotely sensed near-surface soil moisture is anticipating the planned launch of the European Space Agency's Soil Moisture and Ocean Salinity (SMOS [*Kerr et al.*, 2001]) mission. SMOS is the first purpose designed soil moisture remote sensing mission, and will be followed by NASA's Soil Moisture Active Passive (SMAP [*Entekhabi et al.*, 2004]) mission. However, while SMOS and SMAP are expected to enhance the accuracy and utility of remotely sensed soil moisture data, currently orbiting microwave sensors can already provide

[1]Department of Civil and Environmental Engineering, University of Melbourne, Parkville, Victoria, Australia.
[2]Météo France, CNRS, Toulouse, France.

0148-0227/09/2008JD011650$09.00

useful soil moisture observations. For this study, near-surface soil moisture has been retrieved from passive microwave brightness temperatures observed by the Advanced Microwave Scanning Radiometer–Earth Observing System (AMSR-E). While it is difficult to quantitatively verify remotely sensed soil moisture due to the scarcity of soil moisture data at the appropriate scales [*Reichle et al.*, 2004], some encouraging comparisons have been made between soil moisture derived from AMSR-E and that from other sources. At the local scale, AMSR-E derived soil moisture has a good temporal association to in situ soil moisture data [*Wagner et al.*, 2007; *Rüdiger et al.*, 2009; *Draper et al.*, 2009], and to model data [*Rüdiger et al.*, 2009]. At the continental scale, it shows a clear response to precipitation [*McCabe et al.*, 2005; *Draper et al.*, 2009], and using a novel evaluation technique, *Crow and Zhan* [2007] showed that the assimilation of AMSR-E derived soil moisture into a simple water balance model added value to that model.

[4] In addition to recent advances in the remote sensing of soil moisture, there has also been focused development of suitable assimilation strategies for near-surface soil moisture observations. Early studies based on synthetic data showed that observation increments of near-surface soil moisture, or similarly microwave brightness temperature, can be propagated into the deeper soil layers [*Reichle et al.*, 2001; *Walker and Houser*, 2001]. Studies of single column models run over heavily instrumented field sites confirmed that such an assimilation can improve the model deep soil moisture [*Seuffert et al.*, 2004; *Muñoz Sabater et al.*, 2007]. Using remotely sensed data at the continental scale, *Drusch* [2007] used a simple nudging scheme to assimilate Tropical Rainfall Measuring Mission Microwave Imager derived near-surface soil moisture into the ECMWF Integrated Forecast System model over the southern United States, and *Scipal et al.* [2008] used the same method to assimilate European Remote Sensing scatterometer derived soil moisture globally. Both demonstrated that the nudging scheme improved the root-zone soil moisture (compared to ground data), and both recommended the development of a more sophisticated assimilation scheme. Using NASA's global Catchment Land Model (CLM), *Ni-Meister et al.* [2006] showed that an Ensemble Kalman Filter (EnKF) assimilation of near-surface soil moisture derived from Scanning Multichannel Microwave Radiometer (SMMR) generated improvements in the CLM soil moisture over Eurasia. Also using an EnKF with the CLM, *Reichle et al.* [2007] demonstrated modest improvements in the model root-zone soil moisture compared to ground data by assimilating near-surface soil moisture from SMMR and AMSR-E.

[5] A significant hurdle to the assimilation of near-surface soil moisture in NWP models has been the expense of the additional model integrations required by advanced assimilation methods. However, this expense can be reduced by assimilating the data into an off-line version of the land surface model. Using a simplified 2D-Variational assimilation approach, *Balsamo et al.* [2007] showed that the information content of different observation types (including screen-level variables and microwave brightness temperature) is similar for assimilation into either an off-line or atmospherically coupled land surface model. In the same experimental setup as used here, *Mahfouf et al.* [2009] developed a surface analysis for ALADIN, based on assimilating screen-level temperature and humidity into an off-line version of its land surface scheme, the Interactions between Surface, Biosphere, and Atmosphere (ISBA) model. *Mahfouf et al.* [2009] used a Simplified EKF (SEKF), in which a static background error was assumed at the time of each analysis, and the observation operator was a 6-h ISBA integration. In an experiment over July 2006, they showed that the dynamic Kalman gain terms for the SEKF were similar to the analytically derived coefficients used in the operational Optimal Interpolation (OI) scheme [*Giard and Bazile*, 2000]. As well as confirming the viability of the SEKF, this demonstrates that the off-line system captures the necessary surface–screen-level interactions for the assimilation of screen-level observations, since the OI coefficients were derived using the full atmospheric model [*Bouttier et al.*, 1993].

[6] This work extends that of *Mahfouf et al.* [2009] to assimilate remotely sensed near-surface soil moisture derived from AMSR-E observations into ISBA, and is a preliminary step before the combined assimilation of remotely sensed near-surface soil moisture and screen-level observations. The screen-level data assimilated by *Mahfouf et al.* [2009] were reliably available for each analysis cycle, and the use of a static background error was based on the assumption that the increase in the background error during each forecast step was balanced by the reduction from the previous analysis. This assumption is less valid in this study, since the remotely sensed data used here are not available with the same regularity, motivating the development of a full EKF. Additionally, the dynamic error covariances will be of greater importance for the future combined assimilation of screen-level observations and near-surface soil moisture, which are available at different frequencies [*Rüdiger et al.*, 2007]. The aim of this study is to determine whether an off-line assimilation of remotely sensed near-surface soil moisture is a viable method for analyzing root-zone soil moisture in ISBA, and to understand how the near-surface soil moisture increments are translated into deeper-layer increments by the Kalman filter. Additionally, the benefit of using dynamic background error covariances is tested by comparing the SEKF and EKF assimilation of near-surface soil moisture.

2. Methodology

2.1. ISBA Land Surface Scheme

[7] ISBA [*Noilhan and Mahfouf*, 1996] is the land surface scheme used in the ALADIN NWP model. The moisture and energy dynamics in ISBA are modeled using a force-restore method [*Deardorff*, 1977], with eight prognostic variables: surface temperature, mean (deep layer) surface temperature, surface water content (liquid/frozen), total (deep layer) water content (liquid/frozen), vegetation intercepted water content, and snow water content. There is free drainage from the lower boundary, and each grid is divided into a vegetated and bare soil fraction, with evaporation calculated separately from each (although there is a single heat budget). For moisture, the near-surface water reservoir (w_1) is defined as the depth from which moisture can be extracted by bare soil evaporation (~10 mm), and the total water reservoir (w_2) is defined as the depth from which moisture can be extracted through bare-soil evaporation or transpiration (0.1 to 10 m, depending on the local soil type and climate). Both soil layers are forced by precipitation and

evaporation, and transpiration is applied to w_2, with the atmospheric forcing acting more slowly on w_2. The model restore term adjusts w_1 toward an equilibrium between capillary and gravity forces, while w_2 is restored toward field capacity by gravitational drainage. For ALADIN, the soil moisture and temperature states are currently analyzed from screen-level observations of humidity and temperature, using the OI technique of *Giard and Bazile* [2000].

[8] The EKF assimilation uses an off-line version of ISBA within the Surface Externalized (SURFEX) environment. In SURFEX the atmospheric forcing is applied at the first atmospheric model layer (17 m); this is higher than most off-line land surface models, to enable off-line assimilation of screen-level observations. For this experiment ISBA has been run in an environment that resembles the operational ALADIN model as closely as possible. One month of hourly forcing fields (precipitation, temperature, specific humidity, pressure, wind components, and short- and long-wave radiation) has been generated from ALADIN, and interpolated onto the ISBA time step (300 s). Eventually the surface analysis from SURFEX will be semicoupled to the NWP model, so that ALADIN is updated with the soil moisture analyses, and the SURFEX forcing supplied from the updated atmospheric forecast. However, for this initial investigation static forcing has been used, neglecting feedback between the soil moisture updates and the atmospheric forecasts. For further details of SURFEX, and how it would be coupled to the NWP for a land surface assimilation, refer to *Mahfouf et al.* [2009].

2.2. Soil Moisture From AMSR-E

[9] Near-surface soil moisture retrieved from AMSR-E brightness temperatures has been provided by the Vrije Universiteit Amsterdam (VUA) in collaboration with NASA-GSFC [*Owe et al.*, 2007]. C-band AMSR-E data is used here, since *Njoku et al.* [2005] showed that C-band Radio Frequency Interference (RFI) is not widespread across Europe (with the exception of isolated pockets over some urban areas). The descending AMSR-E data (approximate overpass time: 0130 LST) is used, since the nighttime soil moisture retrievals are more accurate [*Owe et al.*, 2001; *Draper et al.*, 2009]. The resolution of C-band AMSR-E data is 45 × 75 km [*Njoku et al.*, 2003], however the swath is oversampled at approximately every 5 km, and (level 2 and 3) C-band data is typically reported on a 0.25° grid, which is thought to approximate the scale of the information in the signal. The ALADIN France model has an irregular (stretched) grid, which covers most of Europe with resolution ∼9.5 km. Rather than disaggregating the 0.25° AMSR-E data, the level 1 swath data have been regridded onto the ALADIN grid using a nearest neighbor approach.

[10] AMSR-E provides global cover in less than two days [*Njoku et al.*, 2003], with coverage decreasing toward the equator. For July 2006 the daily coverage over Europe is reduced from nearly 100% at 58°N, to 70% at 33°N. AMSR-E soil moisture data must be screened to remove data contaminated by RFI or open water, or where dense vegetation or frozen ground cover conceals the near-surface microwave signal. RFI contamination has been identified based on the RFI index of *Li et al.* [2004], which is provided by VUA with the soil moisture data. The only region shown as having significant RFI is Italy, and roughly 50% of the data over the Italian peninsula has been removed (in contrast to *Njoku et al.* [2005], who found limited C-band RFI over Europe, and X-band RFI over Italy in 2003, also using the index of *Li et al.* [2004]). Frozen ground cover is identified and removed during the moisture retrieval, although this is not expected to be significant in July. For vegetation, the VUA-NASA retrieval algorithm partitions the passive microwave signal into soil moisture and vegetation optical depth [*Owe et al.*, 2001]. The vegetation optical depth is linearly proportional to the vegetation water content, and the sensitivity of the microwave brightness temperature to soil moisture decreases with increasing vegetation optical depth [e.g., *de Jeu et al.*, 2008]. *Owe et al.* [2001] show that the soil moisture sensitivity is quite low for optical depths above about 0.75, and a mean monthly optical depth threshold of 0.8 has been adopted to screen out densely vegetated regions, following *de Jeu et al* [2008].

2.3. Extended Kalman Filter

[11] The state forecast and update equations for the EKF are:

$$\mathbf{x}_t^b = \mathcal{M}_{t^-}\left[\mathbf{x}_{t^-}^a\right] \tag{1}$$

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}_t\left[\mathbf{y}_{t+6}^o - \mathcal{H}\left(\mathbf{x}_t^b\right)\right] \tag{2}$$

where

$$\mathbf{K}_t = \mathbf{B}_t^f\mathbf{H}_t^T\left(\mathbf{H}_t\mathbf{B}_t^f\mathbf{H}_t^T + \mathbf{R}\right)^{-1} \tag{3}$$

$\mathbf{x}_t$ indicates the model state at the time of the analysis, t, and the superscripts a and b indicate the analysis and background, respectively. $\mathbf{y}_{t+6}^o$ is the observation vector (6 h after the analysis time). $\mathcal{M}_{t^-}$ is the nonlinear state forecast model (ISBA) from the time of the previous analysis, t^-, $\mathbf{K}$ is the Kalman gain, and $\mathbf{B}$ and $\mathbf{R}$ are the covariance matrices of the background and observation errors. $\mathcal{H}$ is the nonlinear observation operator, and $\mathbf{H}$ is its linearization. Here, the state variable consists of the superficial soil moisture (w_1) and the total soil moisture (w_2), and the observation operator is a 6-h integration of ISBA from time t. The AMSR-E near-surface soil moisture observations are assumed to occur at the end of the assimilation window, at 0000 UTC, and the quantity observed by AMSR-E is taken to be equivalent to the model w_1 (both represent the soil moisture in approximately the uppermost 10 mm of soil). The model update is made 6 h before the observation time (in contrast to *Mahfouf et al.* [2009] who add the increment at the observation time).

[12] The linearization of $\mathcal{H}$ is obtained by finite differences, using a first-order Taylor expansion about $\mathbf{x}$. For each analysis cycle, this requires an additional (perturbed) 6-h model integration for each element of the state vector. For the *ith* observation, and the *jth* element of the control vector:

$$H_{ij,t} = \frac{\mathcal{H}\left(\mathbf{x}_t + \delta x_{j,t}\right)_i - \mathcal{H}(\mathbf{x}_t)_i}{\delta x_j} \tag{4}$$

[13] The background error covariance matrix undergoes an analogous forecast and analysis cycle:

$$\mathbf{B}_t^f = \mathbf{M}_{t^-}\mathbf{B}_{t^-}^a\mathbf{M}_{t^-}^T + \mathbf{Q} \tag{5}$$

$$\mathbf{B}_t^a = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{B}_t^f \tag{6}$$

[14] In the forecast step (equation (5)), the previous analysis, $\mathbf{B}_{t\_}^a$, is forecast forward in time by the tangent linear of the state forecast model, $\mathbf{M}$, and the forecast error covariance matrix, $\mathbf{Q}$, is added to account for errors in the model forecast, giving the background error matrix forecast, $\mathbf{B}_t^f$. The model state analysis decreases the model error, and $\mathbf{B}$ is reduced by an analysis step (equation (6)). The linearization of $\mathcal{M}$ is obtained by the same finite difference method used for $\mathbf{H}$. The linearization of $\mathcal{M}$ is made affordable by the assumption that there is no horizontal correlation in the model errors. The 24-h model Jacobian is estimated as the product of four 6-h Jacobians ($\mathbf{M}_{t\rightarrow t+24} = \Pi\mathbf{M}_{t+18\rightarrow t+24}\mathbf{M}_{t+12\rightarrow t+18}\mathbf{M}_{t+6\rightarrow t+12}\mathbf{M}_{t\rightarrow t+6}$), to reduce the potential for nonlinearities.

[15] An alternative, and more common EKF formulation for the assimilation of near-surface soil moisture retrievals is to make the update at the time of the observations, and use $\mathcal{H}$ = (1 0) rather than including the model in $\mathcal{H}$ [e.g., *Walker and Houser*, 2001; *Reichle et al.*, 2002; *Muñoz Sabater et al.*, 2007]. This form of the EKF is analytically the same as that used here, except for the timing of the addition of $\mathbf{Q}$ (see Appendix A). While the $\mathcal{H}$ = (1 0) method avoids the additional integrations required here to linearize $\mathcal{H}$, it is dependent on the observed variable being included in the state vector. If the assimilation of remotely sensed soil moisture proves useful, it is intended that it be combined with the assimilation of screen-level observations. The screen-level observations cannot be sensibly included in the update vector, since they are not prognostic within SURFEX: they are diagnosed by interpolating the humidity and temperature between the ISBA surface and the (prescribed) value at the first atmospheric model layer. For consistency with *Mahfouf et al.* [2009] and future studies, the EKF form as initially described (ISBA included in $\mathcal{H}$) is used here.

[16] The error correlations for the AMSR-E and ALADIN soil moisture have been set based on the assumption that the standard deviation of the observed and modeled soil moisture errors are equal. The variance of the difference between the (rescaled) AMSR-E and ALADIN near-surface soil moisture (w_1) is 0.0061, which gives an error standard deviation of 0.055 m^3 m^{-3} for each, assuming that the observation and model error are independent and unbiased. This agrees closely with the error estimates that have been made for AMSR-E (see, for example, Table 1 of *de Jeu et al* [2008]), and the observation error standard deviation has been set at this value. Only the diagonal entries of the model error matrices ($\mathbf{B}$ and $\mathbf{Q}$) have been prescribed (i.e., error cross correlations have not been applied), and following *Mahfouf et al.* [2009] the model soil moisture errors are assumed to be proportional to the soil moisture range (the difference between the volumetric field capacity (w_{fc}) and the wilting point (w_{wilt}), calculated as a function of soil type, as given by *Noilhan and Mahfouf* [1996]). For $\mathbf{B}$, the initial model error standard deviations for both w_1 and w_2 have been set to 0.6 $\times$ ($w_{fc} - w_{wilt}$), which converts to a mean volumetric error standard deviation of 0.052 m^3 m^{-3}, slightly lower than that used for the observation error. The magnitude of the diagonal elements of $\mathbf{Q}$ were selected to minimize long-term tendencies in $\mathbf{B}$, on the assumption that $\mathbf{B}$ should not dramatically increase or decrease over time. Using this method values of 0.3 $\times$ ($w_{fc} - w_{wilt}$) and 0.2 $\times$ ($w_{fc} - w_{wilt}$) were chosen for the w_1 and w_2 error standard deviations, respectively. The EKF assimilation is compared to a SEKF assimilation of soil moisture, which neglects the evolution of the background error (equations (5) and (6)), and assumes that $\mathbf{B}$ is constant at the start of each analysis cycle (some dependence on the conditions of the day is introduced through the use of the model in $\mathcal{H}$). For the SEKF analysis, the same $\mathbf{R}$ was used, and $\mathbf{B}$ was set at the same initial value as was used for the EKF.

3. Results

3.1. Scaling the Observations to the Model Climatology

[17] Since the soil moisture quantity observed by remote sensors differs from that defined in models, soil moisture data must be rescaled before assimilation, so as to be consistent with the model climatology [*Reichle et al.*, 2004]. Here, the AMSR-E data are rescaled by matching its Cumulative Distribution Function (CDF [*Reichle and Koster*, 2004; *Drusch et al.*, 2005]) to that of the superficial soil moisture forecast by ALADIN for 0000 UTC each day (this is the 6-h forecast from 1800 UTC, which provides the first guess for the operational soil moisture analysis). Ideally, a long data set is used to sample the model and observation climatology and the CDF matching is performed on as localized a scale as possible, however for this study only 1 year of ALADIN soil moisture fields are available. *Reichle and Koster* [2004] demonstrated that the CDF matching operator can be estimated from 1 year of data by using spatial averaging to compensate for the reduced temporal sample size, and the CDF-matching operator has been estimated here using a one-degree window around each grid cell.

[18] CDF matching is based on the assumption that the differences between the model and observations are stationary, however for ALADIN and AMSR-E this is not the case. For example, Figure 1 shows a time series of the AMSR-E data before and after the CDF matching at a location in northern France. While both the AMSR-E and ALADIN time series have a similar range of short-term (up to several days) variability with amplitude between 0.1 and 0.2 m^3 m^{-3}, the seasonal cycle in the AMSR-E data has a greater magnitude (>0.2 m^3 m^{-3}) than that in the ALADIN data ($\sim$0.1 m^3 m^{-3}). To compensate for the variance generated by the enhanced seasonal cycle in the AMSR-E data, the CDF matching has overly dampened the short-term variability, resulting in a lessened response to rain events in the CDF-matched time series. Additionally, at the seasonal to monthly scale there are biases in the CDF-matched time series (e.g., around day 100 in Figure 1).

[19] To avoid these problems, the CDF matching has been repeated using seasonally bias-corrected AMSR-E data, generated by subtracting the observation-model difference in the 31-day moving average. The resulting time series in Figure 1 has retained an appropriate response to

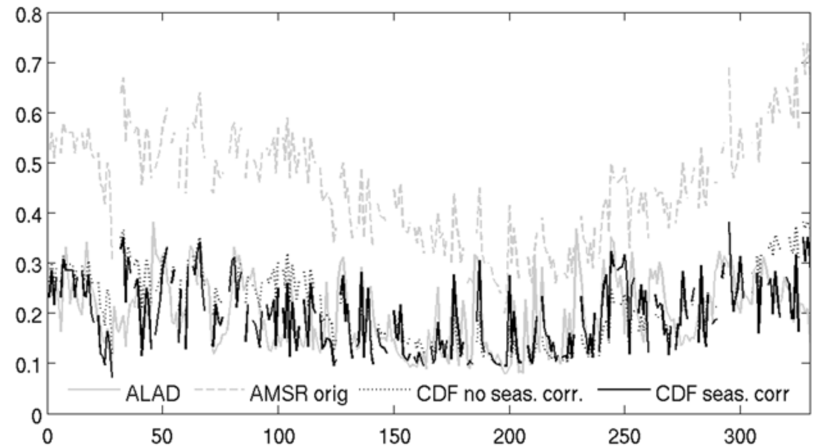

Figure 1. Time series of near-surface soil moisture (m^3 m^{-3}) for a grid cell in France (47.30 E/0.06 N) over 2006, from ALADIN (gray, solid), the original AMSR-E data (gray, dashed), and the seasonally corrected (black, solid) and nonseasonally corrected (black, dotted) CDF-matched AMSR-E data.

precipitation, and the monthly biases are reduced. For this experiment (July 2006) the mean monthly bias is reduced from -0.014 m^3 m^{-3} in the initial CDF-matched data to 0.001 m^3 m^{-3} in the seasonally corrected and CDF-matched data, compared to 0.14 m^3 m^{-3} in the original data. If a longer data set were available the difference in the seasonal cycles could be removed based on the climatological seasonal cycles, which would retain any seasonal bias anomalies in the observations. With just 1 year of data seasonal bias anomalies cannot be detected (regardless of the method used to rescale), and the approach used here is necessarily conservative, assuming that the ISBA 2006 seasonal cycle was correct.

[20] The mean monthly RMSD between the CDF-matched (and seasonally corrected) AMSR-E data and the ALADIN w_1 is 0.007 m^3 m^{-3}. Figure 2 shows that the RMSD is relatively large (>0.09 m^3 m^{-3}) over most of the Italian peninsula, where much of the data was removed due to RFI, suggesting that the remaining data is of poor quality (all data were rejected at locations with less than 100 observations over 2006, and the poor match is unlikely to be due to reduced data coverage). The RMSD is also relatively high in many locations adjacent to regions screened for dense vegetation, most likely due to increased error in the AMSR-E data due to vegetation interference. Additionally, the higher RMSD over the Alps and the Pyrenees could be due to inaccuracies in the model and/or the data, since both have known problems in regions of steep terrain [*Rüdiger et al.*, 2009].

3.2. Tangent-Linear Approximation

[21] The magnitude of the perturbations used to estimate **M** was chosen by examining the difference between the Jacobians estimated using positive and negative perturbations for a range of magnitudes, following *Walker and Houser* [2001] and *Balsamo et al.* [2004]. On the basis of this method, a perturbation of $10^{-4} \times (w_{fc} - w_{wilt})$ was selected for estimating **M**, and also **H** (the linearization of $\mathcal{H}$ is not discussed here, since linearity over 24 h strongly suggests linearity over 6 h). The difference between the Jacobians estimated with the positive and negative perturbations gives a measure of the nonlinearity of $\mathcal{M}$ for perturbations of that size. Scatterplots of the Jacobian terms estimated with positive and negative perturbations of magnitude $10^{-4} \times (w_{fc} - w_{wilt})$ for the analysis cycle on 1 July 2006 show virtually all of the points aligned along the one-to-one line (not shown), consistent with $\mathcal{M}$ being well approximated by **M** within the range of the applied perturbation. This is confirmed by the statistics in Table 1, which show little difference between the mean, standard deviation, and extreme values for the Jacobians estimated with the positive and negative perturbations. The extreme sensitivity causing the very large maximum values in Table 1 for perturbed w_1 is quite rare, and less than 0.2% of the grid cells have a $\partial w_2(t+24)/\partial w_1(t)$ or $\partial w_1(t+24)/\partial w_1(t)$ greater than 10. *Mahfouf et al.* [2009] used the same perturbation size to estimate the Jacobians for ISBA over 6 h, yet the occurrence of nonlinearities as observed by *Mahfouf et al.* [2009] does not occur here, since the (dissipative) land-surface component of the model is less prone to the nonlinearities that can occur in the atmosphere.

[22] The above analysis indicates that $\mathcal{M}$ is well approximated by **M** within the range of the very small perturbations

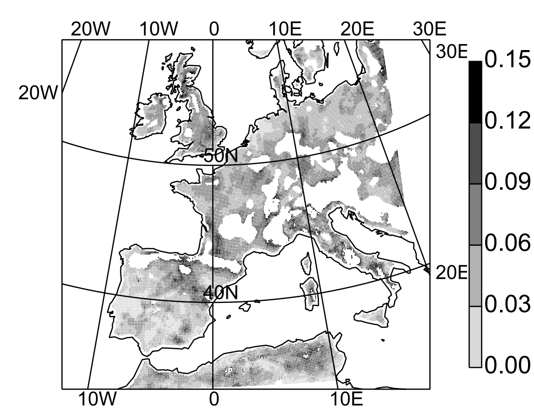

Figure 2. Root-mean-square difference (m^3 m^{-3}) between the CDF-matched AMSR-E near-surface soil moisture and ALADIN w_1 over 2006. White indicates that no AMSR-E soil moisture data are available.

Table 1. Statistics of 24-h Jacobian Terms From 1800 UTC on 1 July 2006[a]

		Mean	SD	Minimum	Maximum
$\partial w_1(t+24)/\partial w_1(t)$	+ve	0.25	1.9	−0.11	189
	−ve	0.26	1.9	−0.11	189
	lrg	0.30	1.9	−0.10	67
$\partial w_2(t+24)/\partial w_1(t)$	+ve	−0.0024	0.0045	−0.15	0.0050
	−ve	−0.0024	0.0045	−0.15	0.0045
	lrg	−0.0025	0.0055	−0.19	0.0
$\partial w_1(t+24)/\partial w_2(t)$	+ve	0.60	0.96	−2.4	120
	−ve	0.60	0.97	−0.37	120
	lrg	0.60	0.53	−0.15	23
$\partial w_2(t+24)/\partial w_2(t)$	+ve	0.95	0.054	0.21	1.0
	−ve	0.95	0.053	0.42	1.0
	lrg	0.96	0.044	0.00	1.0

[a]Estimated using a perturbation size of $+10^{-4} \times (w_{fc} - w_{wilt})$ (positive), $-10^{-4} \times (w_{fc} - w_{wilt})$ (negative), and $+10^{-1} \times (w_{fc} - w_{wilt})$ (large). Units are in (%/%).

that were applied, however this does not guarantee that **M** approximates $\mathcal{M}$ well when applied to the errors in **B**, since these errors are typically much larger than the applied perturbations. To test the potential error generated when **M** is used to propagate **B**, the model Jacobians estimated using perturbations with magnitude similar to the expected model error ($10^{-1} \times (w_{fc} - w_{wilt}) \sim O(10^{-2})$) have been compared to the above estimates. Table 1 shows that the mean Jacobian estimates for this larger perturbation are very similar to those based on the smaller perturbations, although the distribution of values about the mean is different, with differences in their extreme values and variances. The difference between the Jacobians estimated with the smaller and larger perturbation is greater than 0.1 for 4% ($\partial w_1(t+24)/\partial w_1(t)$), 0% ($\partial w_2(t+24)/\partial w_1(t)$), 23% ($\partial w_1(t+24)/\partial w_2(t)$), and 10% ($\partial w_2(t+24)/\partial w_2(t)$) of the grid cells, indicating that the larger perturbation is outside the model's linear regime in more instances (for the positive and negative perturbations considered above the difference was greater than 0.1 for less than 1% of the grids for all of the Jacobian terms). However, the Jacobian estimates compare favorably over the majority of grid cells, indicating that $\mathcal{M}$ estimated with perturbations of $10^{-4} \times (w_{fc} - w_{wilt})$ leads to an acceptable approximation of nonlinear $\mathcal{M}$ for propagating **B** forward 24 h.

3.3. ISBA Jacobians

[23] The ISBA Jacobians reflect the force-restore dynamics of the model. The superficial soil layer responds rapidly to atmospheric forcing, so that a perturbation applied to w_1 is gradually reduced over 24 h. As a result the mean $\partial w_1/\partial w_1$ is reduced from 0.80 over 6 h to 0.25 over 24 h (with the Jacobians estimated from 1800 UTC on 1 July 2006). In addition to its short timescale, w_1 represents a very small physical reservoir, and cannot influence w_2 strongly, so that $\partial w_2/\partial w_1$ is insignificant (mean < 0.01 over 6 or 24 h). In contrast to w_1, the atmospheric forcing is applied more slowly to the total soil moisture, and w_2 has a timescale of 10 days. Over a comparatively short 24-h period a w_2 perturbation is largely retained (mean $\partial w_2(t+24)/\partial w_2(t)$: 0.95). The influence of w_2 on w_1 increases over time, and the mean $\partial w_1(t+6)/\partial w_2(t)$ is 0.20, increasing to 0.60 for $\partial w_1(t+24)/\partial w_2(t)$. Since w_1 does not have a strong or persistent influence on the other surface variables, its accurate analysis is less important than that of w_2. While w_1 could then be excluded from the control variable, (to reduce the number of linearization required), this would result in an underestimation of the model w_1 error, since a large component of this is due to short-lived errors (i.e., the element $\mathbf{q_{11}}$ of the **Q** matrix).

[24] The background error matrix used in each analysis is largely derived from the previous w_2 error correlations and the applied (static) **Q**, since the w_1 errors are short-lived and do not influence w_2. The important terms in the 24-h linear tangent model are then $\partial w_1(t+24)/\partial w_2(t)$ and $\partial w_2(t+24)/\partial w_2(t)$, both of which are shown in Figure 3 for a 24-h period. The Soil Wetness Index (SWI; SWI = $(w_2 - w_{wilt})/(w_{fc} - w_{wilt})$), a measure of soil water availability in the root zone, is provided in Figure 4 for comparison. Over the full diurnal cycle the ISBA moisture dynamics, and hence Jacobians, are dominated by the force component (precipitation and evapotranspiration) of its force-restore scheme. The addition of moisture from precipitation reduces the sensitivity of w_1 to w_2, and the reduced $\partial w_1(t+24)/\partial w_2(t)$ across much of southeast Europe and in smaller regions in northern Spain and along the Pyrenees was caused by rain (these locations have been excluded from the statistics given below). In the absence of precipitation, the 24-h Jacobians are most strongly influenced by evapotranspiration, specifically its sensitivity to w_2. Under dry conditions the parameterization of transpiration depends

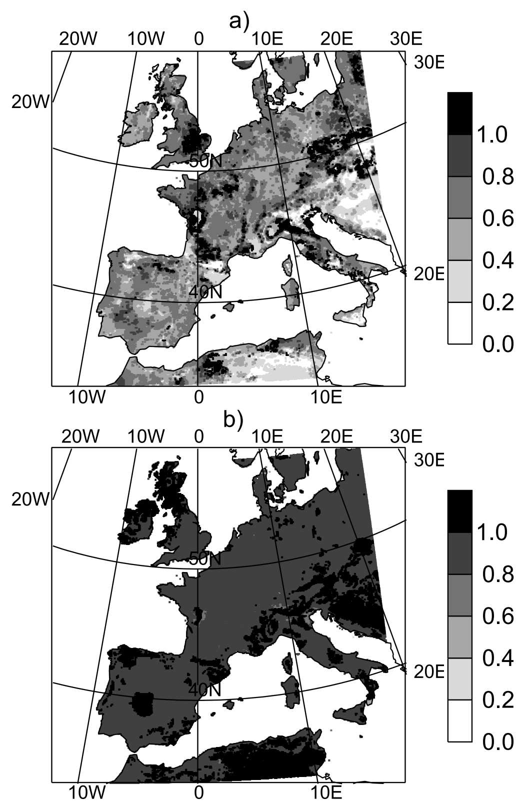

Figure 3. Jacobian terms for the forward model (**M**): (a) $\partial w_1(t+24)/\partial w_2(t)$ and (b) $\partial w_2(t+24)/\partial w_2(t)$ from 1800 UTC on 1 July to 1800 UTC on 2 July 2006. Note that $\partial w_1(t+24)/\partial w_1(t)$ and $\partial w_2(t+24)/\partial w_1(t)$ are not shown, as w_1 has little memory over 24 h.

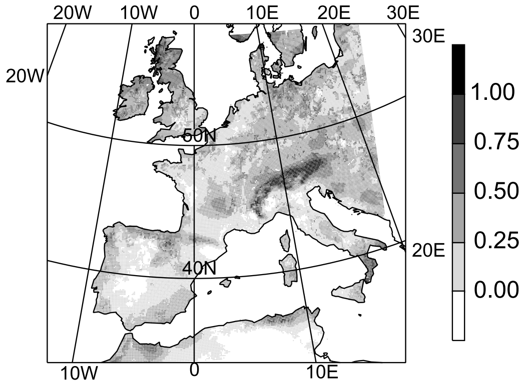

Figure 4. Surface wetness index at 1800 UTC on 2 July 2006 from the EKF analysis.

strongly on w_2, with the dependence increasing as w_2 approaches w_{wilt}. In these moisture limited conditions a small increase in w_2 generates a relatively large increase in transpiration, reducing w_2, and hence $\partial w_2(t + 24)/\partial w_2(t)$. In turn, this leads to an increase in $\partial w_1(t + 24)/\partial w_2(t)$ when the enhanced transpiration reduces the surface temperature, which reduces the depletion of w_1 by bare-ground evaporation, giving a relative increase in w_1. This mechanism has been confirmed by testing the impact on the model forecasts of switching off aspects of the model physics. It is also evidenced in Figure 3. Most obviously, where w_2 is below the wilting point, transpiration ceases and w_2 perturbations are not communicated to w_1. As a result, the regions of negative SWI in Figure 4 in North Africa, and also in Spain and France, correspond to $\partial w_2(t + 24)/\partial w_2(t)$ close to 1 (for negative SWI, the mean $\partial w_2(t + 24)/\partial w_2(t)$ is 1, compared to 0.96 across the whole domain), and reduced $\partial w_1(t + 24)/\partial w_2(t)$ (to a mean of 0.31, compared to 0.64 for the whole domain). For the rest of the domain where the SWI is positive, the Jacobians are most sensitive to moisture availability where there is sufficient vegetation present to generate substantial transpiration. Where the fractional vegetation cover in ISBA is greater than 0.5, $\partial w_2(t + 24)/\partial w_2(t)$ is reduced (mean: 0.92) and $\partial w_1(t + 24)/\partial w_2(t)$ is increased (mean: 0.79) where the SWI is below 0.25 (compared to means of 0.96 and 0.68, respectively for all grids with positive SWI and fractional vegetation greater than 0.5). In contrast, where the fractional vegetation cover is less than 0.5, there is no obvious difference in the $\partial w_2(t + 24)/\partial w_2(t)$ across all grids and across grids with SWI less than 0.25 (mean 0.96 for both), while $\partial w_1(t + 24)/\partial w_2(t)$ is slightly reduced in the drier locations (mean 0.56, compared to 0.59 for all sparsely vegetated cells). The role of vegetation can be seen in Figure 3. Over France and the UK, where the vegetation fraction is greater than 0.75, $\partial w_1(t + 24)/\partial w_2(t)$ is generally elevated (>0.8) where the SWI in Figure 4 is low (<0.25), yet in sparsely vegetated Spain (vegetation fraction < 0.5), where the SWI is similarly low there is no such relationship.

[25] The 6-h model Jacobians used in the observation operator are plotted in Figure 5. While the 24-h Jacobian terms reflect ISBA's force component, for the descending pass AMSR-E data used here the 6-h Jacobians are estimated during the night, when the forcing is weak (excepting regions of rain). In the absence of strong forcing, ISBA restores w_1 toward w_2 to achieve a balance between capillary rise and gravitational drainage. This introduces a weak nighttime sensitivity of w_1 to w_2 resulting in a mean $\partial w_1(t + 6)/\partial w_2(t)$ of 0.20 (as already noted, this is much lower than the corresponding value over 24 h). There is also less spatial variability in the 6-h nighttime Jacobians (for $\partial w_1/\partial w_2$ the variance over 6 h is 0.007, compared to 0.1 over 24 h). Owing to the absence of strong forcing and the slow timescale of the model restore term, $\partial w_1(t + 6)/\partial w_1(t)$ is reasonably high, with a mean of 0.80. Comparison of Figures 4 and 5 suggests a tendency for decreased $\partial w_1(t + 6)/\partial w_2(t)$ where the SWI is lower ($\partial w_1(t + 6)/\partial w_1(t)$ is also slightly increased in these regions, although this is not evident at the plotted scale), suggesting that capillary rise increases nonlinearly with increasing surface water availability. There is an additional influence from the soil type in ISBA (not shown), with lower clay content giving more rapid flow through the soil, corresponding to increased $\partial w_1(t + 6)/\partial w_2(t)$ and decreased $\partial w_1(t + 6)/\partial w_1(t)$.

3.4. Kalman Gain

[26] Figure 6 shows the Kalman gain for w_2 ($\mathbf{k_2}$) for the EKF and the SEKF on 2 July. The EKF $\mathbf{k_2}$ is between 0.2 and 0.4 across most of Europe, with a mean of 0.27. The SEKF gain is smaller, due to the slightly larger background errors used in this experiment, and is generally less than 0.2, with a mean of 0.12. The spatial patterns for the two gain terms differ, since they are determined by different processes. For the SEKF, the (static) **B** is evolved 6 h by **H**, and there is

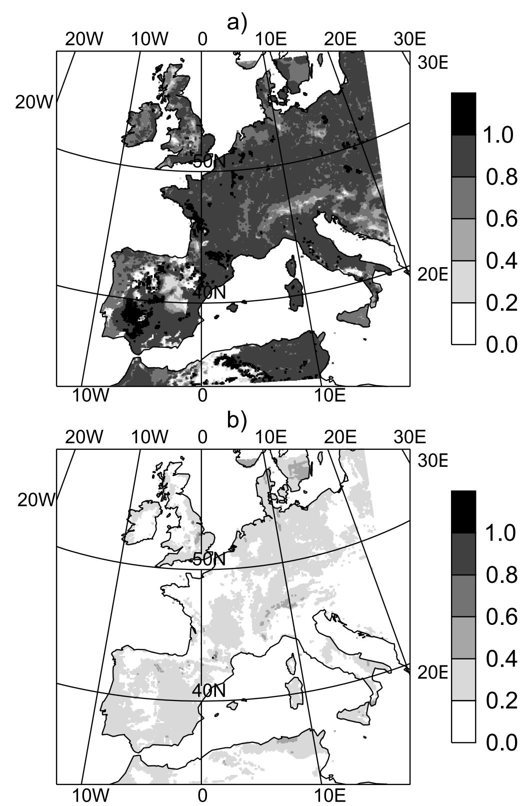

Figure 5. Jacobian terms for the observation operator (**H**): (a) $\partial w_1(t + 6)/\partial w_1(t)$ and (b) $\partial w_1(t + 6)/\partial w_2(t)$, from 1800 to 2400 UTC on 2 July 2006.

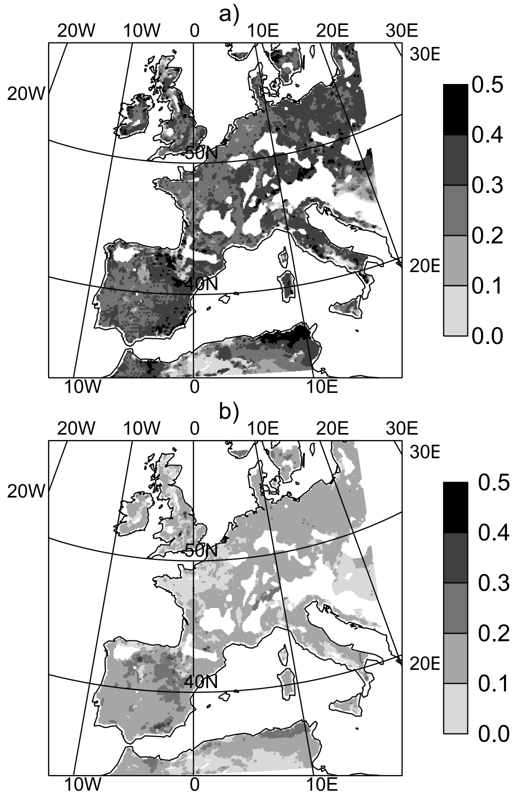

Figure 6. Kalman gain ($m^3\ m^{-3}/m^3\ m^{-3}$) for w_2 at 1800 UTC on 2 July 2006 for the (a) EKF and (b) SEKF.

a qualitative correspondence in Figure 6b (SEKF gain) and Figure 5 (**H**), with the gain being reduced where $\partial w_1(t+6)/\partial w_2(t)$ is lower (e.g., northern France and North Africa). In contrast, the EKF gain is determined by a combination of the Jacobian terms over 6 h (**H**) and 24 h (**M**), and it shows a combination of features from both. Even after a single assimilation cycle, the 24-h **M** has introduced much more fine-scale spatial heterogeneity into the EKF gain than is present in the SEKF gain.

3.5. Soil Moisture Analyses

[27] The time series in Figure 7 shows the soil moisture states for the EKF and the SEKF at four locations with contrasting conditions, together with an open-loop simulation, in which the model surface is allowed to evolve without data assimilation. The grid cells in Slovakia and France are in vegetated regions where transpiration links w_2 to w_1 ($\partial w_1(t+24)/\partial w_2(t) \sim 0.6$ in Figure 3 for both). In both cases the observations are consistently higher than the model forecast w_1, and the assimilation improves the fit between the model w_1 and the observations by adding moisture to w_2. The EKF and SEKF produce similar results, except for a few isolated large increments generated by the EKF in Slovakia. The observation increments in Slovakia are unusually large, particularly in the first part of the month where the observations (if correct) suggest a precipitation event not present in the model forcing. As a result, a large volume of water is added in Slovakia, and the difference between the w_2 SWI for the analyses and the open loop approaches 0.5 at times. The large observation increment on day 18 in France does not

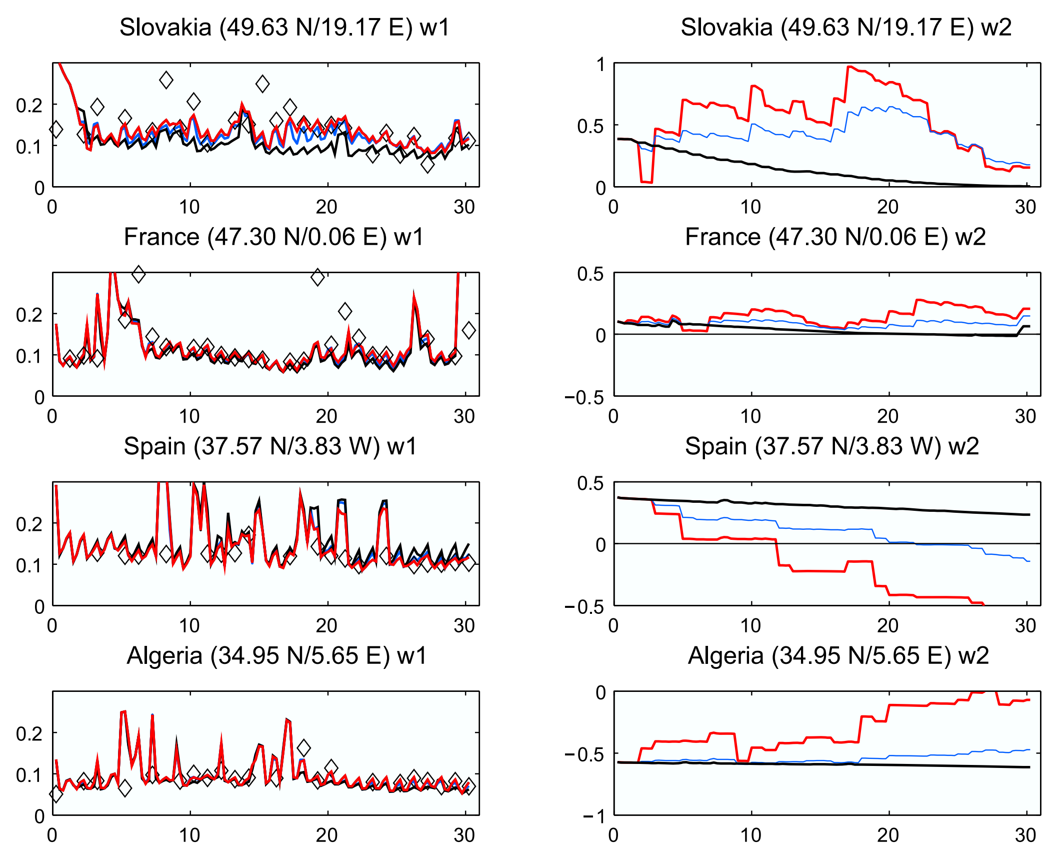

Figure 7. Time evolution of (left) w_1 ($m^3\ m^{-3}$) and (right) w_2 (as SWI) through July 2006, from the open loop (black), EKF (red), and SEKF (blue). Observations of w_1 are indicated as diamonds.

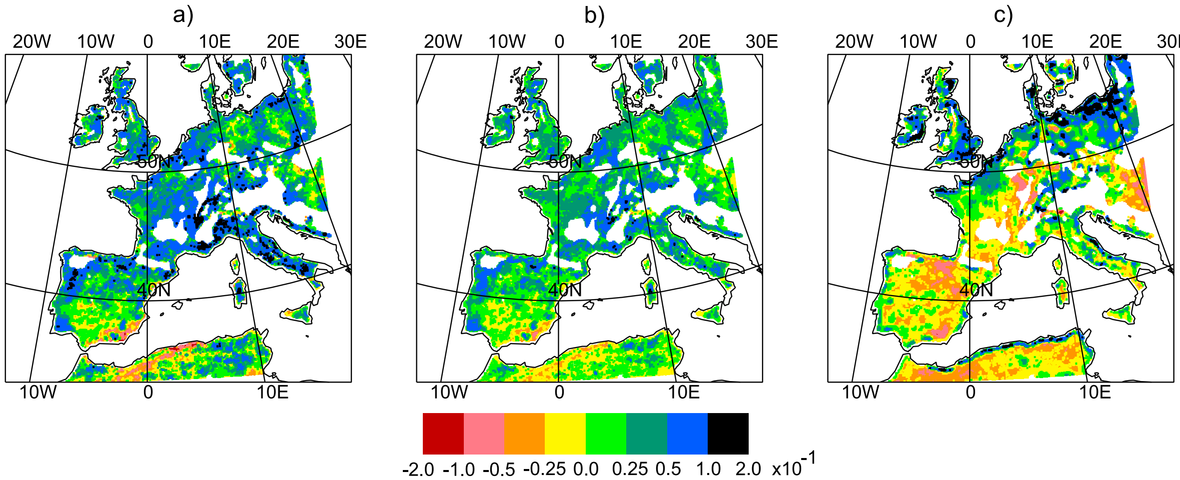

Figure 8. The net monthly w_2 increments ($m^3\ m^{-3}$) over July 2006, from the (a) EKF and (b) SEKF for assimilation of the seasonal-bias corrected AMSR-E data, and (c) EFK assimilation of the nonseasonal-bias corrected AMSR-E data.

translate into an analysis increment as it is screened out by the observation quality control, which discards data more than 0.1 $m^3\ m^{-3}$ away from the model w_1.

[28] The grid cells in Spain and Algeria both have sparse vegetation cover, and with limited transpiration the dependence of w_1 on w_2 is weaker, particularly in Algeria, where w_2 is below the wilting point, and $\partial w_1(t + 24)/\partial w_2(t)$ in Figure 3 is $\sim$0.05 (compared to $\sim$0.2 in Spain). In Spain the observations are generally lower than the model w_1, and the analysis consistently decreases w_2. In the first half of the month, despite the SWI having been decreased by 0.5, neither analysis generates substantial changes in w_1, and the analysis continues to deplete w_2 until a very large net increment is evolved. It is only after the w_2 SWI has been decreased by nearly one that the analysis generates a slight reduction in w_1 (which does give a better fit to the data). A similar situation occurs in Algeria where the observations are consistently above the model, and the analysis makes a series of positive increments to w_2, which do not affect w_1 until a large net change is accumulated. By the end of the month w_2 is approaching a SWI of 0, and w_1 shows an enhanced diurnal cycle, with greater nighttime increases. In both of these cases, since there is little transpiration to expose w_1 to the w_2 increments (compare the ratio of the net change in w_2 and w_1 to that from the previous examples), the analysis continues to make monotonic corrections to w_2 until a large (and likely erroneous) net increment has been imposed on w_2. Initially, the SEKF and EKF increments are similar in magnitude, however as the month progresses the EKF increments become larger. This is due to an inflation of the background error where transpiration is limited. Recall from Figure 3 that $\partial w_2(t + 24)/\partial w_2(t)$ approaches one in regions with little transpiration. As a result the $\mathbf{b}_{22}$ element of $\mathbf{B}$ is not decreased during the forecast step (equation (5)), and $\mathbf{B}$ gradually increases with each addition of $\mathbf{Q}$ (which is generally greater than the analysis reduction). While the $\mathbf{Q}$ used here does not generate a discernible trend in $\mathbf{B}$ across the remainder of the domain, $\mathbf{b}_{22}$ in north Africa and Spain is almost doubled within two weeks.

[29] Figure 8 shows the net soil moisture increments added by the EKF and the SEKF over July 2006. Even though the Kalman gain terms for each depend on different aspects of the model physics, the resultant analyses are similar. For both the EKF and the SEKF moisture has been added across most of the domain, except for areas in southern Spain and central North Africa (as well as some smaller isolated in northern and eastern Europe). The mean monthly net increment is 0.025 $m^3\ m^{-3}$ for the EKF and 0.018 $m^3\ m^{-3}$ for the SEKF. The EKF has a greater spread of increments, with more extreme values (both positive and negative), resulting in a larger standard deviation of the net monthly increment (0.037 $m^3\ m^{-3}$) than for the SEKF (0.023 $m^3\ m^{-3}$). Some of the very large (>0.1 $m^3\ m^{-3}$) increments for the EKF surrounding the Alps correspond to the high RMSD between the CDF-matched AMSR-E and ALADIN soil moisture in Figure 2, where there are known errors in both the model and observations (section 3.1). The increments are also large over Italy, where the coverage of AMSR-E data is limited by RFI, and the quality of the remaining data is questionable. In general, the analysis increments are relatively large compared to the dynamics of w_2, being approximately the same magnitude as the mean range of w_2 throughout July 2006 (0.02 $m^3\ m^{-3}$). In terms of the net volume of water added to the surface, the EKF added a monthly mean volume of 55 mm, while a mean of 41 mm was added by the SEKF (with a total range of approximately ±200 mm for both). This represents a substantial component of the monthly water balance, and is similar to the mean monthly volume added by precipitation (50 mm). Similarly large increments were obtained by *Mahfouf et al.* [2009] for the assimilation of screen-level observations. The large volume of water being added (or removed) is partly due to the two-layer structure of ISBA, since increments to w_2 must be applied across the total soil depth, leading to large net increments, as discussed by *Mahfouf et al.* [2009].

[30] Owing to the strong seasonal cycle in the AMSR-E–ALADIN bias, it was necessary to bias-correct the seasonal cycle in the AMSR-E data before performing the CDF matching (section 3.1). To highlight the central role of data quality to the analysis, the EKF assimilation has been repeated with the original CDF-matched (no seasonal bias correction) AMSR-E data. In this case, the resultant analyses

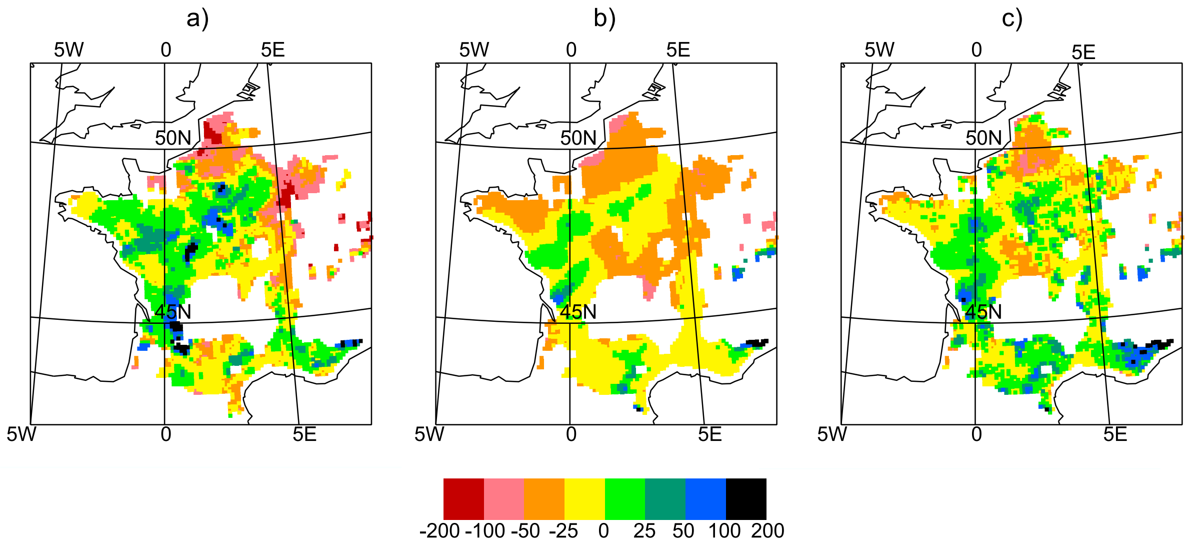

Figure 9. Change in total soil moisture storage (mm) from 1 to 31 July 2006, from (a) SIM, (b) open loop, and (c) the EKF.

differed substantially from the results obtained with the seasonally bias-corrected data. The mean absolute difference between the net monthly increments produced by assimilating the AMSR-E data with and without seasonal bias correction was 0.039 m^3 m^{-3}, compared to a mean difference between the EKF and SEKF assimilation (for the seasonally bias-corrected data) of 0.014 m^3 m^{-3}. Figure 8c shows the net monthly soil moisture increments for the assimilation of the nonseasonally bias-corrected data. In this case the quality control (removal of all data greater than 0.01 m^3 m^{-3} from the model w_1) was not applied as this resulted in most data being removed. Figure 8c is quite different from the previous two panels, and has net positive increments in northern Europe and net negative increments elsewhere, consistent with the strong negative bias over July 2006 in the nonseasonally bias corrected AMSR-E data.

[31] The current operational analysis also includes analysis of the soil temperature. While it is beyond the scope of this paper, an additional experiment was carried out with soil temperature included in the control variable. This experiment showed that total (deep layer) soil temperature analysis increments can also be obtained from near-surface soil moisture observations, and also that the inclusion of temperature has a slight effect on the soil moisture analysis, but does not alter the main findings presented here.

3.6. Comparison to SIM Water Balance

[32] While the focus here is on the mechanics of the assimilation, the resultant analyses have been reality checked by comparison to simulations from SAFRAN-ISBA-MODCOU (SIM [*Habets et al.*, 2008]). SIM is a three-layer version of ISBA forced with high-quality data [*Quintana-Seguí et al.*, 2008] over France. The soil moisture from SIM compares favorably to other estimates of soil moisture [*Rüdiger et al.*, 2009], and it can be regarded as the best available estimate of the true surface state over France. As outlined by *Mahfouf et al.* [2009], the total change in column soil moisture over a time period gives an integration of the surface-moisture inputs (precipitation), outputs (evapotranspiration, runoff), and soil moisture increments (where an assimilation is performed). Figure 9 shows the change in the total-column soil moisture over July 2006 from SIM (Figure 9a), the open loop (Figure 9b), and the EKF (Figure 9c) (the SEKF is not included, since its results are very similar to the EKF). The open loop is forced with the same ALADIN forecasts used in the EKF experiment, and the difference between the change in soil moisture from SIM and from the open-loop simulation will be predominantly due to errors in these forecasts. It is hoped that the assimilation can correct for some of the forcing errors, bringing the total change in soil moisture closer to that from SIM.

[33] In comparison to SIM (Figure 9a), the open loop (Figure 9b) has a tendency toward excessive drying and insufficient wetting, resulting in a mean monthly change in soil moisture for the open loop of -20 mm, compared to -11 mm for SIM. The open loop has generated incorrect drying (in spatial extent and magnitude) along the English Channel coast and in central France, with a region of insufficient moistening in between, associated with a low bias in the ALADIN precipitation forcing. Also, the open loop did not moisten the regions along the Atlantic Coast and south of the Alps indicated by SIM. The EKF (Figure 9c) has added moisture across most of France, increasing the mean monthly increment to -7 mm (overshooting the SIM mean). The EKF shows a general improvement in the correspondence to SIM. It has corrected the band of insufficient moistening in the north, as well as the lack of moistening along the Atlantic coast and in the southeast. However, in the east it has degraded the open loop by adding moisture where drying was correctly identified in the open loop.

4. Discussion

[34] This experiment has demonstrated that the total soil moisture in ISBA can be analyzed using an EKF from remotely sensed near-surface soil moisture observations, in this case from AMSR-E. Assuming that the background and observation errors are (approximately) equal, the EKF Kalman gain over July 2006 was typically around 20–30%, giving a mean net monthly increment of 0.025 m^3 m^{-3}

(equivalent to 55 mm of water added to the soil column). While the EKF increments are large compared to the model dynamics and water balance, they are similar in magnitude to the increments generated by Météo-France's operational OI scheme over the same period [*Mahfouf et al.*, 2009]. Comparison of the monthly water balance generated by the EKF analysis to that from SIM over France showed a general improvement compared to an open loop, although some areas were degraded. The EKF requires the linearization of the forecast model in order to propagate background errors through time. The inaccuracy introduced by the linearization has been estimated by comparing the model Jacobians (calculated using a perturbation small enough that the model is approximately linear) to the Jacobians generated by applying a perturbation of the approximate size of the expected background errors. This test indicated that the linearization provides a good approximation of the model Jacobians for use in the EKF in most instances.

[35] Since w_1 does not directly influence w_2 in ISBA, the analysis of w_2 from w_1 observations must utilize the sensitivity of w_1 to changes in w_2. The effectiveness of the assimilation is then limited by the strength of $\partial w_1(t + 24)/\partial w_2(t)$. Over the diurnal cycle $\partial w_1(t + 24)/\partial w_2(t)$ is dominated by daytime radiative forcing and the influence of w_2 on w_1 is principally determined by the transpiration physics (enhanced w_2 causes enhanced transpiration, causing decreased superficial soil temperature, giving decreased bare soil evaporation, and a relative increase in w_1). The greatest sensitivity, and hence most effective analysis of w_2, occurs where transpiration is most sensitive to w_2: in reasonably vegetated regions when w_2 is close to, but above, w_{wilt}. Conversely, where w_2 is less than w_{wilt}, or is very high (so that transpiration is not moisture limited), or where there is little vegetation, w_2 does not substantially influence w_1, and so cannot be effectively analyzed from w_1 observations.

[36] A major motivation for using remotely sensed near-surface soil moisture in NWP is the expectation that it will provide a more direct observation of total soil moisture than screen-level observations do, since the latter rely on the model flux parameterizations to link the surface state to the screen-level atmosphere. However it has been shown here that for ISBA the link between the near-surface soil moisture observations and the deeper soil moisture is still provided by transpiration. This result is derived from the model physics, and it is expected that many other models, such as multilayer models with more substantial surface layers and more explicit drainage, will provide a more direct relationship between the near-surface and deeper soil moisture.

[37] During the nighttime, when the surface forcing is weak, the sensitivity of w_1 to w_2 in ISBA is determined by the model restore term, representing the balance between capillary rise and gravitational drainage. Since the descending AMSR-E data used here are observed at night, the gain terms are influenced by the nighttime dynamics through the observation operator (this also occurs for the alternate EKF formulation discussed in Appendix A, due to a diurnal cycle in **B**). While a nighttime assimilation has the theoretical advantage of utilizing a more direct physical link between w_2 and w_1, it leads to problems where the nighttime model Jacobians differ from those across the full diurnal cycle. For example, due to the absence of transpiring vegetation in Spain and Algeria in Figure 7, w_1 is only very weakly influenced by w_2 over the full diurnal cycle, however, there is a short-lived stronger sensitivity at night, which generates w_2 analysis increments from the w_1 observations. These w_2 increments do not influence the w_1 forecast for the next day (this suggests that the w_1 observation increments were not caused by w_2 errors), and since the w_1 observation increments are mostly monotonic, a large and likely erroneous net w_2 increment is generated over time. For the EKF this situation was exacerbated in this study by the inflation of **B** in these areas, and the (simplistic) error correlations used here will be refined in the future.

[38] It was assumed here that the background error standard deviations for w_1 and w_2 were both equal to the observation error standard deviation. However, in reality $\mathbf{b_{22}}$ will be lower than $\mathbf{b_{11}}$, since w_1 has more rapid dynamics and is more susceptible to forcing errors. *Muñoz Sabater et al.* [2007] compared soil moisture from ISBA forced with observations to in situ data from 2001 to 2004, and obtained a RMS error of 0.07 v/v for w_1 and 0.03 for w_2 (the contrast in the errors would likely be enhanced by the use of NWP forcing). The use of a smaller background error for w_2 would reduce some of the excessive model increments obtained in this study. Additionally, **Q** was chosen here in an attempt to generate stationary **B** (within the assumed structure of **Q**, proportional to $(w_{fc} - w_{wilt})$), however such a value could not be found across the entire European domain. With the chosen **Q**, the background error grew rapidly in dry and sparsely vegetated regions, including much of North Africa and Spain (resulting in large increments after several weeks in Figure 7). This suggests that **Q** should be lower in these regions. Intuitively, this is sensible: since there is little transpiration, w_2 does not vary greatly, and excepting a precipitation forcing error, the additive forecast error (**Q**) should be low compared to locations with substantial transpiration.

5. Conclusion

[39] This work is the first continental scale study to assimilate remotely sensed near-surface soil moisture into the ISBA model, and it is also the first study to contrast the assimilation of remotely sensed soil moisture using dynamic and static model error covariances. It is demonstrated that useful increments to the total soil layer in ISBA can be generated from near-surface soil moisture observations, in this case derived from AMSR-E. The spatially averaged net monthly increment for the EKF over ALADIN's European domain was 0.025 $m^3\ m^{-3}$, using approximately equal model and observation soil moisture errors. The assimilation was performed over July 2006, using both an EKF and a SEKF (in which the background error at the time of each analysis was assumed constant). While the Kalman gain terms for the SEKF and the EKF are determined by different physical processes, their resultant soil moisture analyses are similar (recall that horizontal error correlations were neglected in this study). Since performing the analysis in an off-line environment makes it computationally feasible, the EKF is suggested for future work with ISBA, although this study suggests that the SEKF provides an acceptable approximation (which is cheaper to compute and easier to implement). The difference between the two may well increase over a longer time period, particularly since the one-month period used in this experi-

ment is only three times the 10-day timescale of w_2, and subsequently $\mathbf{b_{22}}$. The model itself is used as the observation operator, which in combination with the increased height of the atmospheric forcing in SURFEX enables the assimilation of screen-level observations, as demonstrated by *Mahfouf et al.* [2009]. The next stage of this work will be to investigate whether the EKF assimilation of remotely sensed soil moisture can be usefully combined with that of screen-level observations.

[40] While the focus here is on the design of the assimilation, the analyses are ultimately limited by the quality of the ingested data. In this experiment it was necessary to remove the seasonal bias between the AMSR-E and ALADIN soil moisture before using CDF-matching to rescale the AMSR-E data to the model's climatology. The profound difference in the soil analyses generated by including and excluding this seasonal correction was far greater than the difference between the EKF and the SEKF. This highlights the importance of the observation rescaling technique to soil moisture data assimilation. The relatively short period of available data (1 year), combined with the nonstationarity of the model-observation bias presented particular difficulties in rescaling the observations in this study. The issue of obtaining sufficient data to sample the model-observation climatology for rescaling presents a serious challenge for land-surface assimilation, particularly within NWP modeling, where frequent model changes are made, but also for the broader land surface community when using data form the early years of satellite missions.

Appendix A: Forecast Model as the Observation Operator

[41] The classic formulation for the EKF assimilation of near-surface soil moisture is to make the model update at the observation time, and use an observation operator of $\mathcal{H}$ = (1 0) (for the state vector, $\mathbf{x} = (w_1\ w_2)^T$, used here). To allow for the eventual assimilation of both near-surface soil moisture and screen-level atmospheric observations, a 6-h model forecast of the observation equivalent has been used for the observation operator in this study, with the analysis made 6 h before the observation time. It is shown below that for the assimilation of near-surface soil moisture this approach differs from the classic EKF only in the timing of the addition of $\mathbf{Q}$.

[42] For an observation at t = 24, the classic observation operator is written $\hat{\mathcal{H}}(\mathbf{x}^b_{24})$ = (1 0) $(\mathbf{x}^b_{24})$, while the present version is $\mathcal{H}(\mathbf{x}^b_{18})$ = (1 0) $\mathcal{M}_{18\to 24}(\mathbf{x}^b_{18})$. In both cases the result is the forecast w_1 at t = 24, $w_{1,24}$. $\mathbf{Q}$ is neglected for the time being, and $\mathbf{B}$ is expressed as a function of its value at time 0. This gives equations (A1) and (A2) for the classic and current EKF, respectively:

$$\mathbf{x}^a_{24} - \mathbf{x}^b_{24} = \mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 24}\hat{\mathbf{H}}^T \cdot \left(\hat{\mathbf{H}}\mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 24}\hat{\mathbf{H}}^T + \mathbf{R}\right)^{-1}\left(y^o_{24} - w_{1,24}\right) \quad \text{(A1)}$$

$$\mathbf{x}^a_{18} - \mathbf{x}^b_{18} = \mathbf{M}_{0\to 18}\mathbf{B}_0\mathbf{M}^T_{0\to 18}\mathbf{H}^T \cdot \left(\mathbf{H}\mathbf{M}_{0\to 18}\mathbf{B}_0\mathbf{M}^T_{0\to 18}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(y^o_{24} - w_{1,24}\right) \quad \text{(A2)}$$

Applying $\mathbf{M}_{18\to 24}$ to equation (A2) carries it forward 6 h, giving:

$$\begin{aligned}\mathbf{M}_{18\to 24}\left(\mathbf{x}^a_{18} - \mathbf{x}^b_{18}\right) &\simeq \mathbf{x}^a_{24} - \mathbf{x}^b_{24} \\ &= \mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 18}\mathbf{H}^T\left(\mathbf{H}\mathbf{M}_{0\to 18}\mathbf{B}_0\mathbf{M}^T_{0\to 18}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(y^o_{24} - w_{1,24}\right)\end{aligned} \quad \text{(A3)}$$

Substituting $\mathbf{H} = \hat{\mathbf{H}}\mathbf{M}_{18\to 24}$ into equation (A3) produces equation (A1), hence the two forms of the EKF are equivalent if $\mathbf{Q}$ is neglected.

[43] If $\mathbf{Q}$ is included, equations (A1) and (A3) become, respectively:

$$\begin{aligned}\mathbf{x}^a_{24} - \mathbf{x}^b_{24} = &\left(\mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 24}\hat{\mathbf{H}}^T + \mathbf{Q}\hat{\mathbf{H}}^T\right) \\ &\times \left(\hat{\mathbf{H}}\mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 24}\hat{\mathbf{H}}^T + \hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^T + \mathbf{R}\right)^{-1} \\ &\cdot \left(y^o_{24} - w_{1,24}\right)\end{aligned} \quad \text{(A4)}$$

and

$$\begin{aligned}\mathbf{x}^a_{24} - \mathbf{x}^b_{24} &= \left(\mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 18}\mathbf{H}^T + \mathbf{M}_{18\to 24}\mathbf{Q}\mathbf{H}^T\right) \\ &\times \left(\mathbf{H}\mathbf{M}_{0\to 18}\mathbf{B}_0\mathbf{M}^T_{0\to 18}\mathbf{H}^T + \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(y^o_{24} - w_{1,24}\right) \\ &= \left(\mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 24}\hat{\mathbf{H}}^T + \mathbf{M}_{18\to 24}\mathbf{Q}\mathbf{M}^T_{18\to 24}\hat{\mathbf{H}}^T\right) \\ &\times \left(\hat{\mathbf{H}}\mathbf{M}_{0\to 24}\mathbf{B}_0\mathbf{M}^T_{0\to 24}\hat{\mathbf{H}}^T + \hat{\mathbf{H}}\mathbf{M}_{18\to 24}\mathbf{Q}\mathbf{M}^T_{18\to 24}\hat{\mathbf{H}}^T + \mathbf{R}\right)^{-1} \\ &\cdot \left(y^o_{24} - w_{1,24}\right)\end{aligned} \quad \text{(A5)}$$

[44] The difference between equations (A4) and (A5) is in the timing of the addition of $\mathbf{Q}$. In the latter $\mathbf{Q}$ is added to $\mathbf{B}$ at t = 18, and $\mathbf{B}$ is evolved forward 6 h before the Kalman gain is calculated. This result has been confirmed by comparing the analyses generated by the two methods (using the same $\mathbf{Q}$ s, $\mathbf{R}$, and initial $\mathbf{B}$). The differences are limited to the magnitude of the analysis increments, with the increments being larger (yet showing the same spatial pattern) when the model is used as the observation operator.

[45] **Acknowledgments.** The authors thank Patrick Le Moigne and Valéry Masson for assistance with technical aspects of SURFEX, Karim Bergaoui for coding the first version of the assimilation system, and Thomas Holmes for providing the AMSR-E soil moisture data, together with advice regarding its use. We also acknowledge three anonymous reviewers for their comments and insights. Clara Draper is funded by an Australian Postgraduate Scholarship, and an eWater CRC top-up scholarship, and this work was accomplished with support from the Fondation de Coopération Scientifique Sciences et Technologies pour l'Aéronautique et l'Espace, and a University of Melbourne PORES Scholarship.

References

Baker, R., B. Lynn, A. Boone, W.-K. Tao, and J. Simpson (2001), The influence of soil moisture, coastline curvature, and land-breeze circulations on sea-breeze initiated precipitation, *J. Hydrometeorol.*, *2*(2), 193–211.

Balsamo, G., F. Bouyssel, and J. Noilhan (2004), A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model, *Q. J. R. Meteorol. Soc.*, *130*(598A), 895–915, doi:10.1256/qj.02.215.

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

Bélair, S., L. Crevier, J. Mailhot, B. Bilodeau, and Y. Delage (2003), Operational implementation of the ISBA land surface scheme in the

Canadian regional weather forecast model. Part I: Warm season results, *J. Hydrometeorol.*, *4*(2), 352–370.
Bouttier, F., J.-F. Mahfouf, and J. Noilhan (1993), Sequential assimilation of soil moisture from atmospheric low-level parameters. Part II: Implementation in a mesoscale model, *J. Appl. Meteorol.*, *32*, 1352–1364.
Crow, W., and X. Zhan (2007), Continental-scale evaluation of remotely sensed soil moisture products, *IEEE Geosci. Remote Sens. Lett.*, *4*(3), 451–455, doi:10.1109/LGRS.2007.896533.
Deardorff, J. (1977), A parameterization of the ground surface moisture content for use in atmospheric prediction models, *J. Appl. Meteorol.*, *16*(11), 1182–1185.
de Jeu, R., W. Wagner, T. Holmes, A. Dolman, N. van de Giesen, and J. Friesen (2008), Global soil moisture patterns observed by space borne radiometers and scatterometers, *Surv. Geophys.*, *29*, 399–420.
Draper, C., and G. Mills (2008), The atmospheric water balance over the semiarid Murray-Darling River basin, *J. Hydrometeorol.*, *9*(3), 521–534, doi:10.1175/2007JHM889.1.
Draper, C., J. Walker, P. Steinle, R. de Jeu, and T. Holmes (2009), An evaluation of AMSR-E derived soil moisture over Australia, *Remote Sens. Environ.*, *113*, 703–710, doi:10.1016/j.rse.2008.11.011.
Drusch, M. (2007), Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set, *J. Geophys. Res.*, *112*, D03102, doi:10.1029/2006JD007478.
Drusch, M., and P. Viterbo (2007), Assimilation of screen-level variables in ECMWF's Integrated Forecast System: A study on the impact on the forecast quality and analyzed soil moisture, *Mon. Weather Rev.*, *135*(2), 300–314, doi:10.1175/MWR3309.1.
Drusch, M., E. Wood, and H. Gao (2005), Observation operators for the direct assimilation of TRMM Microwave Imager retrieved soil moisture, *Geophys. Res. Lett.*, *32*, L15403, doi:10.1029/2005GL023623.
Entekhabi, D., et al. (2004), The hydrosphere state (Hydros) satellite mission: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *IEEE Trans. Geosci. Remote Sens.*, *42*(10), 2184–2195, doi:10.1109/TGRS.2004.834631.
Fischer, E., S. Seneviratne, P. Vidale, D. Luthi, and C. Schär (2007), Soil moisture–atmosphere interactions during the 2003 European summer heat wave, *J. Clim.*, *20*(20), 5081–5099, doi:10.1175/JCLI4288.1.
Giard, D., and E. Bazile (2000), Implementation of a new assimilation scheme for soil and surface variables in a global NWP model, *Mon. Weather Rev.*, *128*(4), 997–1015.
Habets, F., et al. (2008), The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, *J. Geophys. Res.*, *113*, D06113, doi:10.1029/2007JD008548.
Hess, H. (2001), Assimilation of screen-level observations by variational soil moisture analysis, *Meteorol. Atmos. Phys.*, *77*, 145–154, doi:10.1007/s007030170023.
Kerr, Y., P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger (2001), Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1729–1735, doi:10.1109/36.942551.
Li, L., E. Njoku, E. Im, P. Chang, and K. Germain (2004), A preliminary survey of radio-frequency interference over the US in Aqua AMSR-E data, *IEEE Trans. Geosci. Remote Sens.*, *42*(2), 380–390, doi:10.1109/TGRS.2003.817195.
Mahfouf, J.-F., K. Bergaoui, C. Draper, C. Bouyssel, F. Taillefer, and L. Taseva (2009), A comparison of two off-line soil analysis schemes for assimilation of screen-level observations, *J. Geophys. Res.*, *114*, D08105, doi:10.1029/2008JD011077.
McCabe, M., E. Wood, and H. Gao (2005), Initial soil moisture retrievals from AMSR-E: Multiscale comparison using in situ data and rainfall patterns over Iowa, *Geophys. Res. Lett.*, *32*, L06403, doi:10.1029/2004GL021222.
Muñoz Sabater, J., L. Jarlan, J.-C. Calvet, F. Bouyssel, and P. de Rosnay (2007), From near-surface to root-zone soil moisture using different assimilation techniques, *J. Hydrometeorol.*, *8*(2), 194–206, doi:10.1175/JHM571.1.
Ni-Meister, W., P. Houser, and J. Walker (2006), Soil moisture initialization for climate prediction: Assimilation of Scanning Multifrequency Microwave Radiometer soil moisture data into a land surface model, *J. Geophys. Res.*, *111*, D20102, doi:10.1029/2006JD007190.
Njoku, E., T. Jackson, V. Lakshmi, T. Chan, and S. Nghiem (2003), Soil moisture retrieval from AMSR-E, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 215–229.
Njoku, E., P. Ashcroft, T. Chan, and L. Li (2005), Global survey of statistics of radio-frequency interference in AMSR-E land observations, *IEEE Trans. Geosci. Remote Sens.*, *43*(5), 938–947, doi:10.1109/TGRS.2004.837507.
Noilhan, J., and J.-F. Mahfouf (1996), The ISBA land surface parameterisation scheme, *Global Planet. Change*, *13*, 145–159.
Owe, M., R. de Jeu, and J. Walker (2001), A methodology for surface soil moisture and vegetation optical depth retrieval using the Microwave Polarization Difference Index, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1643–1654, doi:10.1109/36.942542.
Owe, M., R. de Jeu, and T. Holmes (2007), Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.*, *113*, F01002, doi:10.1029/2007JF000769.
Quintana-Seguí, P., P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C. Canellas, L. Franchisteguy, and S. Morel (2008), Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, *J. Appl. Meteorol. Climatol.*, *47*(1), 92–107, doi:0.1175/2007JAMC1636.1.
Reichle, R., and R. Koster (2004), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, *31*, L19501, doi:10.1029/2004GL020938.
Reichle, R., D. McLaughlin, and D. Entekhabi (2001), Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1708–1718, doi:10.1109/36.942549.
Reichle, R., J. Walker, R. Koster, and P. Houser (2002), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, *3*(6), 728–740.
Reichle, R., R. Koster, J. Dong, and A. Berg (2004), Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation, *J. Hydrometeorol.*, *5*(3), 430–442.
Reichle, R., R. Koster, P. Liu, S. Mahanama, E. Njoku, and M. Owe (2007), Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR), *J. Geophys. Res.*, *112*, D09108, doi:10.1029/2006JD008033.
Rüdiger, C., J.-C. Calvet, J.-F. Mahfouf, L. Jarlan, G. Balsamo, and J. Muñoz Sabater (2007), Assimilation of land surface variables for model initialisation at Météo-France, in *Proceedings of MODSIM 2007 International Congress on Modelling and Simulation*, pp. 1716–1722, Modell. and Simul. Soc. of Aust. and N. Z., Canberra, Australia.
Rüdiger, C., J.-C. Calvet, C. Gruhier, T. Holmes, R. de Jeu, and W. Wagner (2009), An intercomparison of ERS-Scat and AMSR-E soil moisture observations with model simulations over France, *J. Hydrometeorol.*, *10*(2), 431–447, doi:10.1175/2008JHM997.1.
Scipal, K., M. Drusch, and W. Wagner (2008), Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, *Adv. Water Resour.*, *31*(8), 1101–1112, doi:10.1016/j.advwatres.2008.04.013.
Seuffert, G., G. Wilker, P. Viterbo, M. Drusch, and J.-F. Mahfouf (2004), The usage of screen-level parameters and microwave brightness temperature for soil moisture analysis, *J. Hydrometeorol.*, *5*(3), 516–531.
Wagner, W., V. Naeimi, K. Scipal, R. de Jeu, and J. Martínez-Fernández (2007), Soil moisture from operational meteorological satellites, *Hydrogeol. J.*, *15*(1), 121–131, doi:10.1007/s10040-006-0104-6.
Walker, J., and P. Houser (2001), A methodology for initializing soil moisture in a global climate model: Assimilation of near surface soil moisture observations, *J. Geophys. Res.*, *106*(D11), 11,761–11,774.
Zhang, H., and C. S. Frederiksen (2003), Local and nonlocal impacts of soil moisture initialization on AGCM seasonal forecasts: A model sensitivity study, *J. Clim.*, *16*(13), 2117–2137.

C. S. Draper and J. P. Walker, Department of Civil and Environmental Engineering, University of Melbourne, Parkville, Vic 3010, Australia. (c.draper@civenv.unimelb.edu.au; j.walker@unimelb.edu.au)

J.-F. Mahfouf, Météo France, CNRS, CNRM/GAME, 42 avenue Gaspard Coriolis, F-31057 Toulouse CEDEX, France. (jean-francois.mahfouf@meteo.fr)