Advances in Water Resources 34 (2011) 1597–1615

Contents lists available at SciVerse ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states

Haksu Lee [a,b,*], Dong-Jun Seo [a,b,1], Victor Koren [a]

[a] Hydrology Laboratory, NOAA/National Weather Service, Silver Spring, MD 20910, USA
[b] University Corporation for Atmospheric Research, Boulder, CO 80307-3000, USA

ARTICLE INFO

Article history:
Received 27 September 2010
Received in revised form 29 August 2011
Accepted 30 August 2011
Available online 10 September 2011

Keywords:
Data assimilation
Distributed hydrologic modeling
Streamflow
Soil moisture

ABSTRACT

We assess the potential of updating soil moisture states of a distributed hydrologic model by assimilating streamflow and in situ soil moisture data for high-resolution analysis and prediction of streamflow and soil moisture. The model used is the gridded Sacramento (SAC) and kinematic-wave routing models of the National Weather Service (NWS) Hydrology Laboratory's Research Distributed Hydrologic Model (HL-RDHM) operating at an hourly time step. The data assimilation (DA) technique used is variational assimilation (VAR). Assimilating streamflow and soil moisture data into distributed hydrologic models is new and particularly challenging due to the large degrees of freedom associated with the inverse problem. This paper reports findings from the first phase of the research in which we assume, among others, perfectly known hydrometeorological forcing. The motivation for the simplification is to reduce the complexity of the problem in favour of improved understanding and easier interpretation even if it may compromise the goodness of the results. To assess the potential, two types of experiments, synthetic and real-world, were carried out for Eldon (ELDO2), a 795-km^2 headwater catchment located near the Oklahoma (OK) and Arkansas (AR) border in the U.S. The synthetic experiment assesses the upper bound of the performance of the assimilation procedure under the idealized conditions of no structural or parametric errors in the models, a full dynamic range and no microscale variability in the in situ observations of soil moisture, and perfectly known univariate statistics of the observational errors. The results show that assimilating in situ soil moisture data in addition to streamflow data significantly improves analysis and prediction of soil moisture and streamflow, and that assimilating streamflow observations at interior locations in addition to those at the outlet improves analysis and prediction of soil moisture within the drainage areas of the interior stream gauges and of streamflow at downstream cells along the channel network. To assess performance under more realistic conditions, but still under the assumption of perfectly known hydrometeorological forcing to allow comparisons with the synthetic experiment, an exploratory real-world experiment was carried out in which all other assumptions were lifted. The results show that, expectedly, assimilating interior flows in addition to outlet flow improves analysis as well as prediction of streamflow at stream gauge locations, but that assimilating in situ soil moisture data in addition to streamflow data provides little improvement in streamflow analysis and prediction though it reduces systematic biases in soil moisture simulation.

1. Introduction

Arguably the most important advantage of distributed models over lumped models for operational hydrology is that the former provide a means to monitor and predict streamflow and soil moisture anywhere in the basin at the spatial scale of modeling (i.e., at the grid scale or the scale of subbasin delineation). Such capability opens new doors for high-resolution hydrology and water resources products and services for a wide range of users. It has been demonstrated in recent years that calibrated distributed hydrologic models can simulate outlet streamflow at least as well as calibrated lumped models [1–3]. In operational forecasting using lumped models, however, the model states are routinely adjusted in real time by human forecaster and, if available, by automatic

* Corresponding author. Present address: NOAA/NWS/Office of Hydrologic Development, 1325 East-West Highway, Silver Spring, MD 20910, USA. Tel.: +1 301 713 0640x151; fax: +1 301 713 0963.
E-mail address: Haksu.Lee@noaa.gov (H. Lee).
[1] Present address: Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX 76019-0308, USA.

0309-1708/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2011.08.012

state updating procedures based on real-time streamflow and forcing data [4–6]. Such real-time updating significantly improves the accuracy of streamflow prediction at the outlet (see e.g. [6]) and constitutes an integral step in the operational hydrologic forecasting process. Similarly, one may expect that, for distributed models to be a routinely-used tool in real-time hydrologic prediction, real-time adjustment of distributed model states is necessary.

Regardless of the choice of the models used, analysis and prediction of streamflow and soil moisture are subject to various uncertainties in the observations and in the model structure, parameters and initial conditions. Compared to that for lumped models, state updating for distributed hydrologic models is particularly challenging due not only to the large degrees of freedom (DOF) in the inverse problem, i.e., the number of control or state variables involved can be very large compared to the information content in the available real-time data, but also to the generally large sensitivity of the nonlinear model dynamics to the space–time scale of modeling [1,7,8]. The purpose of this work is to assess the potential of real-time assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models for monitoring and prediction of streamflow and soil moisture.

Data assimilation (DA) has been used in various hydrologic modelling and prediction studies and applications [9], including satellite-derived soil moisture assimilation with land surface models [10–12], assimilation of satellite-derived snow covered area with lumped models [13], streamflow assimilation with lumped models [14,5,15,16], and streamflow assimilation with distributed models [17–19], just to mention several. While many studies may be found in the literature on soil moisture assimilation [20,10,11,21,22,12] and streamflow assimilation [18,23,5,19,15], to the best of the authors' knowledge, few address assimilating both streamflow and soil moisture data into lumped [24] or, in particular, distributed models. Since soil moisture observations contain direct information (albeit at different scale) about the model soil moisture states, they may be expected to help reduce DOF in the inverse (i.e., DA) problem compared to using streamflow data alone. Note that, with the latter, many different combinations of the initial model soil moisture states may yield similar streamflow simulation results at the basin outlet.

In this paper, we design and carry out synthetic and real-world experiments to assess the effects of uncertainties in the initial model soil moisture states and real-time observations on assimilation of streamflow and in situ soil moisture data into distributed hydrologic models. In these experiments, we assume that precipitation and potential evaporation are observed perfectly. The above assumption, while less than realistic, is motivated in this first phase of the research by the need to reduce the complexity of the DA problem so that we may understand the problem better. This paper is organized as follows. Section 2 describes the models used. Section 3 describes and formulates the DA problem. Section 3 describes the solution approach to the formulated DA problem. Section 5 describes the study area and data used. Sections 6 and 7 describe the synthetic and real-world experiments, respectively, and present the results. Section 8 provides conclusions and future research recommendations.

2. Models used

Many aspects of DA depend greatly on the particulars of the models used. As such, we first describe the models used in this work in some detail. The distributed models used are part of the National Weather Service (NWS) Hydrology Laboratory's Research Distributed Hydrologic Model (HL-RDHM, [2]) operating at an hourly time step. The HL-RDHM consists of the Sacramento Soil Moisture Accounting Model (SAC,[25,2]), the Antecedent Precipitation Index Model (API, [26]), the SNOW-17 model [27] and the kinematic-wave routing models for hillslope and channel flows [2]. The models operate at the Hydrologic Rainfall Analysis Project (HRAP) grid scale ($\sim$16 km^2) [28,29]. Of the four models in HL-RDHM, only the SAC and kinematic-wave routing models are used in this work.

The SAC is a conceptual soil moisture accounting model [25] which calculates fast and slow runoffs from "buckets" in two vertical layers. Kinematic-wave routing models calculate streamflow at the outlet of each HRAP grid with runoff calculated from the SAC. Cell-to-cell channel routing is performed based on the channel connectivity map derived by the COTAT (Cell Outlet Tracing with an Area Threshold) algorithm [2,30]. The NEXRAD-based

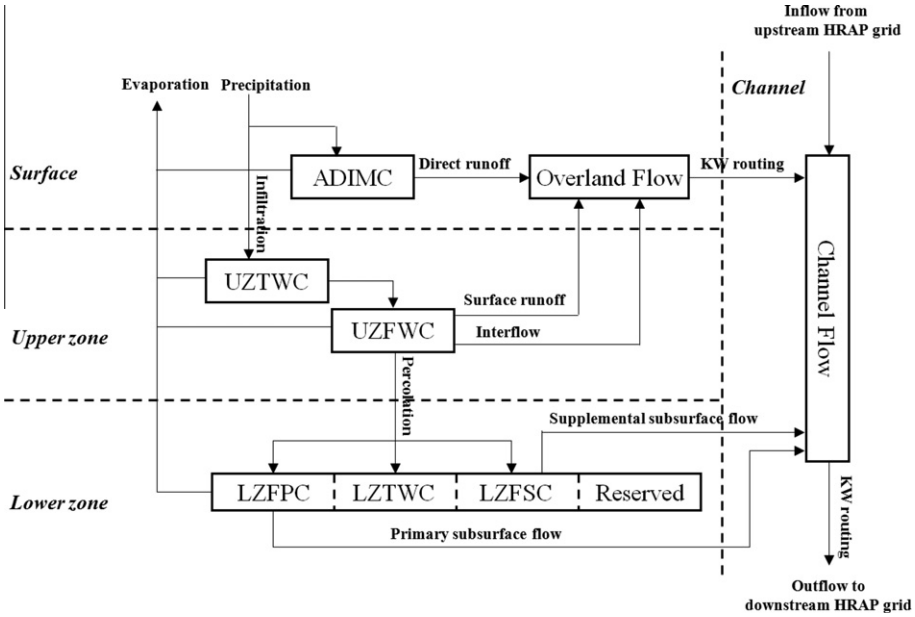

Fig. 1. Schematic of the gridded SAC and kinematic-wave (KW) routing models where ADIMC, UZTWC, UZFWC, LZTWC, LZFSC, and LZFPC, denote additional impervious area water content, upper zone tension water content, upper zone free water content, lower zone tension water content, lower zone supplemental free water content, and lower zone primary free water content, respectively.

multi-sensor precipitation [31–34] and monthly climatology of potential evaporation (PE) on the HRAP grid are used as hydrometeorological forcing to the SAC. Koren et al. [35] developed a procedure for deriving a priori estimates of the SAC parameters from the soil data such as STATSGO2 [36] and SSURGO [37]. In this work, those derived from the STATSGO2 data are used. Hillslope and channel routing parameters were estimated from the digital elevation model and channel hydraulic data [2]. These a priori estimates may be refined or optimized manually or via an optimization tool [2]. In this work, we used the manually-optimized parameter values developed from the Distributed Model Intercomparison Project (DMIP, [38]). Fig. 1 shows the schematic of the gridded SAC and kinematic-wave routing models. Further details on the model structure of HL-RDHM and the estimation of a priori SAC and routing parameters may be found in Koren et al. [2]. Recently, Koren [39] and Koren et al. [40] extended SAC by incorporating heat transfer (HT) dynamics. The extended model, referred to as SAC-HT, allows mapping of the tension and free water storages of SAC into soil moisture states in the vertical [40]. This part of the SAC-HT is used in this work to map SAC states into vertical distribution of soil water content and to assimilate in situ soil moisture data.

3. Problem formulation

Our assimilation problem may be described in generality as follows. Given the a priori knowledge of the initial conditions of the model states and model errors, real-time observations of streamflow, gridded precipitation and PE and in situ soil moisture, update the model soil moisture states at the beginning of the assimilation window (see Fig. 2) and multiplicative biases for the observed precipitation and PE within the assimilation window. The length of the assimilation window is approximately the time scale of the fast runoff reaching the outlet from the most upstream location in the basin (see [5] for rationale). Note that adjusting multiplicative biases for the observed precipitation and PE amounts to adjusting the model soil moisture states within the assimilation window, which results in updated initial conditions of model soil moisture valid at the prediction time, denoted as K in Fig. 2. The above may be formulated as the following nonlinear constrained least-squares minimization problem:

$$\begin{aligned}\text{Minimize } J_K = &\frac{1}{2}\sum_{k=K-L+1}^{K}\left[\mathbf{Z}_{Q,k}-\mathbf{H}_{Q,k}\left(\mathbf{X}_{S,K-L},\mathbf{X}_{P,K-L+1},\ldots,\mathbf{X}_{P,k},\right.\right.\\ &\left.\mathbf{X}_{E,K-L+1},\ldots,\mathbf{X}_{E,k},\mathbf{X}_{W,K-L+1},\ldots,\mathbf{X}_{W,k}\right)\Big]^T\mathbf{R}_Q^{-1}\\ &\left[\mathbf{Z}_{Q,k}-\mathbf{H}_{Q,k}\left(\mathbf{X}_{S,K-L},\mathbf{X}_{P,K-L+1},\ldots,\mathbf{X}_{P,k},\mathbf{X}_{E,K-L+1},\ldots,\right.\right.\\ &\left.\left.\mathbf{X}_{E,k},\mathbf{X}_{W,K-L+1},\ldots,\mathbf{X}_{W,k}\right)\right]\\ &+\frac{1}{2}\sum_{k=K-L+1}^{K}\left[\mathbf{Z}_{\theta,k}-\mathbf{H}_{\theta,k}\left(\mathbf{X}_{S,k},\mathbf{X}_{W,k}\right)\right]^T\mathbf{R}_\theta^{-1}\left[\mathbf{Z}_{\theta,k}-\mathbf{H}_{\theta,k}\left(\mathbf{X}_{S,k},\mathbf{X}_{W,k}\right)\right]\\ &+\frac{1}{2}\sum_{k=K-L+1}^{K}\left[\mathbf{Z}_{P,k}-\mathbf{H}_{P,k}\mathbf{X}_{P,k}\right]^T\mathbf{R}_P^{-1}\left[\mathbf{Z}_{P,k}-\mathbf{H}_{P,k}\mathbf{X}_{P,k}\right]\\ &+\frac{1}{2}\sum_{k=K-L+1}^{K}\left[\mathbf{Z}_{E,k}-\mathbf{H}_{E,k}\mathbf{X}_{E,k}\right]^T\mathbf{R}_E^{-1}\left[\mathbf{Z}_{E,k}-\mathbf{H}_{E,k}\mathbf{X}_{E,k}\right]\\ &+\frac{1}{2}\left[\mathbf{Z}_B-\mathbf{H}_B\mathbf{X}_{S,K-L}\right]^T\mathbf{R}_B^{-1}\left[\mathbf{Z}_B-\mathbf{H}_B\mathbf{X}_{S,K-L}\right]\\ &+\frac{1}{2}\sum_{k=K-L+1}^{K}\mathbf{X}_{W,k}^T\mathbf{R}_W^{-1}\mathbf{X}_{W,k}\end{aligned} \tag{1}$$

$$\text{subject to } \begin{cases}\mathbf{X}_{S,k}=\mathbf{M}\left(\mathbf{X}_{S,k-1},\mathbf{X}_{P,k},\mathbf{X}_{E,k},\mathbf{X}_{W,k}\right),\\ \qquad k=K-L+1,\ldots,K,\\ X_{S,j,i}^{\min}\leqslant X_{S,j,i,k}\leqslant X_{S,j,i}^{\max},\quad k=K-L,\ldots,K;\\ \quad j=1,\ldots,n_S;\quad i=1,\ldots,n_C.\end{cases} \tag{2}$$

In the above, J_K denotes the objective function value at the current hour K, $\mathbf{Z}$, $\mathbf{H}$, $\mathbf{X}$, $\mathbf{R}$, and $\mathbf{M}$ denote the observation, the structure function that relates control variables to observed variables, the control vector containing the set of variables to be adjusted (i.e., updated), the observation error covariance matrix, and the soil moisture accounting model, respectively, the subscripts Q, θ, S, P, E, B and W denote streamflow, in situ soil moisture, model soil moisture state, precipitation, PE, background (i.e., a priori or before-DA) model soil moisture state, and rainfall-runoff model structural error, respectively, and the superscript T denotes the matrix transpose. Accordingly, $\mathbf{Z}_Q$, $\mathbf{Z}_\theta$, $\mathbf{Z}_P$, $\mathbf{Z}_E$, and $\mathbf{Z}_B$ denote the observation vectors for streamflow, soil moisture at the grid scale, precipitation, PE, and SAC states at the beginning of the assimilation window, respectively. The variable L denotes the length of the assimilation window, and k denotes the hourly time index. Throughout this paper, we use bold- and light-face letters to denote vector and scalar quantities, respectively. In the above, we consider the model-generated (i.e., without-DA) SAC states to be the best estimates of $\mathbf{Z}_B$ given that $\mathbf{Z}_B$ is not observed in reality. The vector $\mathbf{X}_{S,K-L}$ denotes the five SAC sates at hour K-L, the upper zone tension water content (UZTWC), the upper zone free water content (UZFWC), the lower zone tension water content (LZTWC), the lower zone supplemental free water content (LZFSC), and the lower zone primary free water content (LZFPC) [25]. The vectors $\mathbf{X}_{P,k}$ and $\mathbf{X}_{E,k}$ denote the multiplicative adjustment factors for the observed precipitation and PE at hour k within the assimilation window, respectively. The vector $\mathbf{X}_{W,k}$ represents the model error at hour k. In Eq. (1), $\mathbf{H}_Q(\ )$ represents the gridded SAC and kinematic-wave routing models, $\mathbf{H}_P$ and $\mathbf{H}_E$ are the same as $\mathbf{Z}_P$ and $\mathbf{Z}_E$, respectively, and $\mathbf{H}_B$ is the identity matrix. In Eq. (2), $X_{S,j,i}^{\min}$ and $X_{S,j,i}^{\max}$ denote the lower and upper bounds of the jth state vector at the ith grid, $X_{S,j,i}$, and n_S and n_C denote the number of SAC states and the number of HRAP cells within the basin, respectively.

The above formulation is based on the following observation equations for the control vectors:

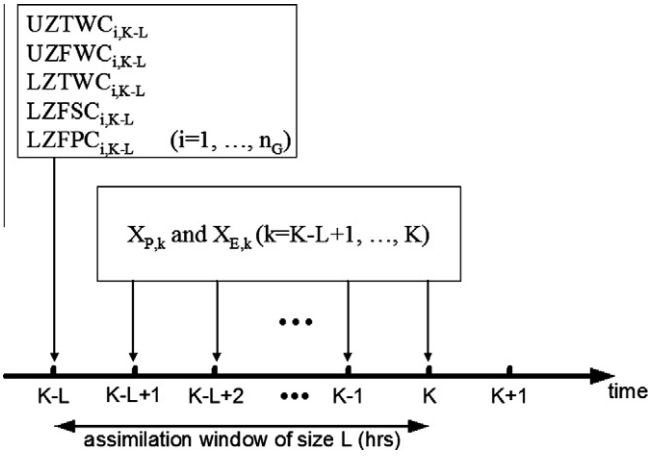

Fig. 2. Schematic of the assimilation window and the control vector. The control vector is composed of five SAC states shown above at the beginning of the assimilation window (UZTWC$_{i,K-L}$, UZFWC$_{i,K-L}$, LZTWC$_{i,K-L}$, LZFSC$_{i,K-L}$, LZFPC$_i$, $K-L$) and the adjustment factors to precipitation and potential evaporation (PE), $X_{P,k}$ and $X_{E,k}$, respectively, where $i=1,\ldots,n_G$, $k=K-L+1,\ldots,K$, and UZTWC, UZFWC, LZTWC, LZFSC, and LZFPC stand for the upper-zone tension water content (mm), the upper-zone free water content (mm), the lower-zone tension water content (mm), the lower-zone supplemental free water content (mm) and the lower-zone primary free water content (mm), respectively. In the above, n_G, L, and K denote the number of HRAP cells in the basin, the size of the assimilation window, and the current hour, respectively. In this work, $X_{P,k}$ and $X_{E,k}$ are excluded from the control vector.

$$\mathbf{Z}_P = \mathbf{H}_P\mathbf{X}_P + \mathbf{V}_P \quad (3)$$
$$\mathbf{Z}_E = \mathbf{H}_E\mathbf{X}_E + \mathbf{V}_E \quad (4)$$
$$\mathbf{Z}_B = \mathbf{H}_B\mathbf{X}_{S,K-L} + \mathbf{V}_B \quad (5)$$
$$\mathbf{Z}_Q = \mathbf{H}_Q(\mathbf{X}_{S,K-L}, \mathbf{X}_P, \mathbf{X}_E, \mathbf{X}_W) + \mathbf{V}_Q \quad (6)$$
$$\mathbf{Z}_\theta = \mathbf{H}_\theta(\mathbf{X}_S, \mathbf{X}_W) + \mathbf{V}_\theta, \quad (7)$$

where $\mathbf{V}_P$, $\mathbf{V}_E$,$\mathbf{V}_B$, $\mathbf{V}_Q$, and $\mathbf{V}_\theta$ denote the measurement error vectors for precipitation, PE, background model state, streamflow, and in situ soil moisture at the model grid scale, respectively. Following the notational convention described above, $\mathbf{X}_P$, $\mathbf{X}_E$, and $\mathbf{X}_W$ in Eqs. (3)–(7) denote the control vectors of generally time-varying biases in the precipitation and PE data, and the model error, respectively, whereas the control vectors for bias at hour k within the assimilation window are denoted as $\mathbf{X}_{P,k}$, $\mathbf{X}_{E,k}$, and $\mathbf{X}_{W,k}$ respectively. To specify the a priori or background values for the control vectors, we assume no biases in precipitation ($\mathbf{X}_{P,k} = 1.0$) and PE ($\mathbf{X}_{E,k} = 1.0$), no model structural error ($\mathbf{X}_{W,k} = 0$), and $\mathbf{X}_{S,K-L} = \mathbf{Z}_B$. In the above, $\mathbf{X}_W$ represents the error in the total channel inflow (TCI), which combines the surface and groundwater runoffs from the SAC [6].

Because soil moisture at the model grid is not observed in reality, we upscale the point-scale in situ soil moisture observations to those at the model grid scale via probability matching; Appendix B describes this procedure and how the microscale variability of soil moisture is accounted for in prescribing the variance of $\mathbf{V}_\theta$. Eqs. (1) and (2) are a least-squares constrained nonlinear minimization problem with an objective function made of linearly weighted penalty terms. It is worth noting that Eq. (1) is arrived at from the very large static Fisher estimation problem [41] of estimating the control vectors from all available observations and a priori information under the assumption that the observation errors associated with $\mathbf{Z}_Q$, $\mathbf{Z}_\theta$, $\mathbf{Z}_P$, $\mathbf{Z}_E$ and $\mathbf{Z}_B$ are independent of one another. It is very important to note that, in this work, we assume no structural or parametric uncertainty in the model. The motivation for this assumption in the synthetic experiment is to assess the upper bound of the performance of DA under the idealized conditions. With the assumption, we drop the model error terms and apply the models as a strong constraint [42] in the data assimilation procedure. Below, we describe how Eqs. (1) and (2) are solved under the above and additional simplifying assumptions.

4. Solution approach

Eqs. (1) and (2) may be solved via variational or sequential DA techniques. Here we use variational assimilation (VAR, [43]). For Kalman filter interpretation of the above formulation, the reader is referred to Jazwinski [44]. With VAR, the minimization problem of Eqs. (1) and (2) is solved numerically using gradient-based optimization algorithms in which the gradient of the objective function with respect to the control vector is evaluated by the adjoint method [45]. Additional details on VAR and the adjoint method can be found in [45,43,9] and references therein. The primary motivation for using VAR in this work is that it is a general gradient-based nonlinear minimizer in the least squares sense and hence can handle nonlinear observation equations as well as nonlinear model dynamics to the extent that the tangent linear model can accurately evaluate gradients (i.e., as long as the model dynamics does not have discontinuities within the assimilation window [46]). Ensemble Kalman filter [47], on the other hand, is optimal in the mean squared error sense, only if the observation equations are linear. In theory, one could use particle filtering [48], which is free of distributional or linearity assumptions. In reality, however, computational burden is prohibitively large for particle filters to be practical for large-dimensional DA problems [49].

As written in Eqs. (1) and (2), the minimization problem is difficult to solve in practice because we usually do not have enough available information to model the space–time structure of the observational error covariance terms. Here, we assume that the observation errors are independent of one another and time-invariant (see [5] for justification), which renders the observation error covariance matrices $\mathbf{R}$ in Eq. (1) diagonal and static. The use of the above assumptions is justified by the following reasons. First, we have little information on the spatiotemporal statistical properties of the observation error for streamflow and grid-scale soil moisture. As such, modeling of the spatiotemporal correlation structures would have been at best ad hoc. The second is that the observation errors for streamflow and soil moisture may be considered independent as there is no physical or statistical reason to postulate that they are correlated, causally or statistically. Even if we could accurately model the spatiotemporal correlation structure of the observation errors of streamflow and that of soil moisture, its primary impact would be on determination of the relative magnitude of the first two terms in Eq. (8) below, and not necessarily on the state identifiability. While we acknowledge that the relative magnitude of the two terms in Eq. (8) may be specified more accurately by modeling the spatiotemporal correlation structures, the impact of such modeling is only of second-order importance and does not justify the increase in computational burden. The third is that, even though the relative magnitude of the first two terms in Eq. (8) may not be very accurate due to the independence assumption, the range of observation error variance values used in the sensitivity analysis as carried out in this study is likely to encompass the true relative magnitude of the two terms with the independence assumption relaxed or even lifted. As such, the performance bound obtained from this study is likely to delineate reasonably well the outcome obtainable with the regularization effects associated with non-diagonal observational covariance. Also, as noted in the Introduction Section, we limit ourselves here to updating the model soil moisture states at the beginning of the assimilation window without adjusting the observed precipitation and PE within the assimilation window (see Fig. 2). Then, Eqs. (1) and (2) may be simplified to:

$$\begin{aligned}\text{Minimize}\quad J_K = {} & \frac{1}{2}\sum_{k=K-L+1}^{K}\sum_{l=1}^{n_Q}\left[Z_{Q,l,k} - H_{Q,l,k}(\mathbf{X}_{S,K-L})\right]^2\sigma_{Q,l}^{-2} \\ & + \frac{1}{2}\sum_{k=K-L+1}^{K}\sum_{i=1}^{n_C}\sum_{m=1}^{n_D}\left[Z^P_{\theta,i,m,k} - H_{\theta,i,m,k}(\mathbf{X}_{S,k})\right]^2\sigma_{\theta,i,m}^{-2} \\ & + \frac{1}{2}\sum_{j=1}^{n_S}\sum_{i=1}^{n_C}\left[Z_{B,j,i,K-L} - X_{S,j,i,K-L}\right]^2\sigma_{B,j,i}^{-2} \end{aligned} \quad (8)$$

$$\text{subject to}\quad \begin{cases} \mathbf{X}_{S,k} = \mathbf{M}(\mathbf{X}_{S,k-1}), & k = K-L+1,\ldots,K, \\ X^{\min}_{S,j,i} \leqslant X_{S,j,i,k} \leqslant X^{\max}_{S,j,i}, & k = K-L,\ldots,K; \\ j = 1,\ldots,n_S; \quad i = 1,\ldots,n_C, & \end{cases} \quad (9)$$

In the above, n_Q denotes the number of stream gauge stations, $Z_{Q,l,k}$ denotes the streamflow observation at the lth gauge station at hour k, $Z^P_{\theta,i,m,k}$ denotes the pseudo in situ soil moisture observation at the HRAP scale (see below for explanation) at the ith grid, mth depth, and hour k, and $Z_{B,\,j,\,i,\,K-L}$ denotes the background model soil moisture state associated with the jth state variable and ith cell at the beginning of the assimilation window, $H_{Q,l,k}(\ )$ denotes the observation equation that maps the model soil moisture at the beginning of the assimilation window, $\mathbf{X}_{S,K-L}$, to streamflow at the lth gauge station and hour k, $H_{\theta,i,m,k}(\ )$ denotes the observation equation that maps $\mathbf{X}_{S,k}$ to soil moisture at the HRAP scale at the ith grid, the mth depth, and hour k, where $\mathbf{X}_{S,k}$ denotes the SAC states at hour k, $\sigma_{Q,l}$ denotes the standard deviation of the streamflow observation error at the lth stream gauge location, $\sigma_{\theta,i,m}$ denotes the standard

deviation of the observation error associated with the in situ soil moisture at the HRAP scale at the ith grid and the mth depth, $\sigma_{B,j,i}$ denotes the standard deviation of the error associated with the jth background model state at the ith grid. In Eq. (8), the pseudo in situ soil moisture observation $Z^{P}_{\theta,i,m,k}$ is an estimate of the in situ soil moisture at the HRAP scale as obtained from the in situ soil moisture observation at the point scale via probability matching (see Appendix B for details).

Eqs. (8) and (9) may be solved numerically by gradient-based nonlinear minimization. We evaluated gradients in this work using the adjoint code generated from Tapenade (http://tapenade.inria.fr:8080/tapenade/index.jsp). The gradients were verified against those obtained from the tangent linear code generated also from Tapenade. We tested two minimization algorithms, Fletcher–Reeves–Polak–Ribiere minimization (FRPRMN) and the Broyden–Fletcher–Goldfarb–Shanno variant of Davidon–Fletcher–Powell minimization (DFPMIN) [50]. The algorithms FRPRMN and DFPMIN implement conjugate gradient and quasi-Newton methods, respectively. In comparative testing using Eqs. (8) and (9), both algorithms required similar numbers of function evaluations, ranging from 15 to 150 (mostly around 20), for each assimilation run with similar minimization results. Throughout this work, we used FRPRMN.

5. Study area and data used

The basin chosen for this work is Eldon (ELDO2), a 795-km^2 headwater catchment (see Fig. 3) located near the Oklahoma (OK)-Arkansas (AR) border in the southern plains of the US. The basin is within the coverage of the WSR-88D's (Weather Surveillance Radar – 1988 Doppler version) at Springfield, MO, Inola, OK, and Ft. Smith, AR. The basin is an agricultural area with gently-rolling to hilly topography [6]. The hourly streamflow data from the United States Geological Survey (USGS) are available at two interior locations, Christie (65 km^2) and Dutch (105 km^2), and at the outlet, Eldon. Mean annual precipitation for ELDO2 from 1997 to 2002, the period of study in this work, is 1,232 mm. Mean annual surface runoff for the same period is approximately 388, 199 and 451 mm for Eldon, Christie and Dutch, respectively. Within the catchment, there is an Oklahoma Mesonet soil moisture measurement site at Westville, OK [51]. A Campbell Scientific Inc. (CSI) 229-L soil moisture sensor measures temperature changes in the soil from which soil moisture may be estimated [52]. For this study, we used the hourly soil moisture data at two different depths, 25 and 60 cm. The size of the assimilation window used for ELDO2 is 36 h, which is the same as the duration of unit hydrograph estimated for this basin. Fig. 3a shows the elevation map of the study area with 50 m-interval contours, the stream gauge locations, the soil moisture measurement location, and the main channel network. Fig. 3(b) shows the soil types in the basin which are mainly sandy loam and silty loam. Fig. 3(c) shows the channel connectivity map derived by the Cell Outlet Tracing with an Area Threshold (COTAT) algorithm [30]. To assess the potential of assimilating streamflow and soil moisture data into distributed hydrologic models, we performed two types of experiments, synthetic and real-world. They are described in the following two sections.

6. Synthetic experiment

The motivation for the synthetic experiment is to improve understanding of the DA problem described in Section 3 in an idealized controlled environment so that we may be able to make meaningful attribution and interpretation of the results. The experiment was designed in particular to help answer the following

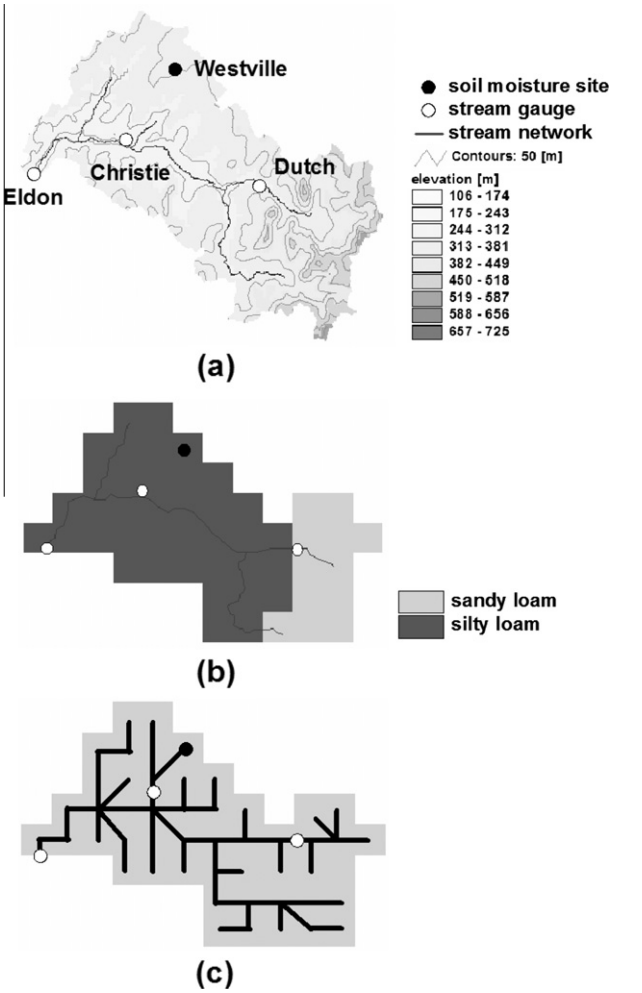

Fig. 3. (a) Digital elevation map, 50 m-interval elevation contours, main channel network, location of the soil moisture measurement site, and stream gauge locations, (b) soil map, and (c) channel connectivity map of ELDO2.

questions: (1) What is the value of assimilating streamflow data, at the outlet only and both at the outlet and interior locations, for analysis and prediction of streamflow and soil moisture at some interior location? (Figs. 4–10), (2) How does the value of DA vary according to the drainage area (Figs. 8 and 10), amount of uncertainty in the initial model soil moisture states (Figs. 4 and 5), and prediction lead time (Figs. 8 and 10)?, (3) What is the value of assimilating in situ soil moisture data in addition to streamflow data? (Figs. 4–10), 4) How accurate do the data have to be to benefit from DA? (Figs. 4–7).

6.1. Experiment design

Once the time period for the experiment was chosen (see below), the synthetic experiment consisted of the following steps: (1) Run the gridded SAC and kinematic-wave routing models using the available real-world forcing (i.e., precipitation and PE) data, and assume that the forcing data and the simulated streamflow and soil moisture represent the truth, (2) Add noise to the assumed "true" observations and initial soil moisture states to mimic uncertainty in the data and in the model initial soil moisture states; the results are then assumed to represent plausible realizations of real-world observations and model initial soil moisture states, (3) Repeat Step 2 100 times, (4) Calculate the observation error variances in Eq. (8) from Step 3, (5) Perform

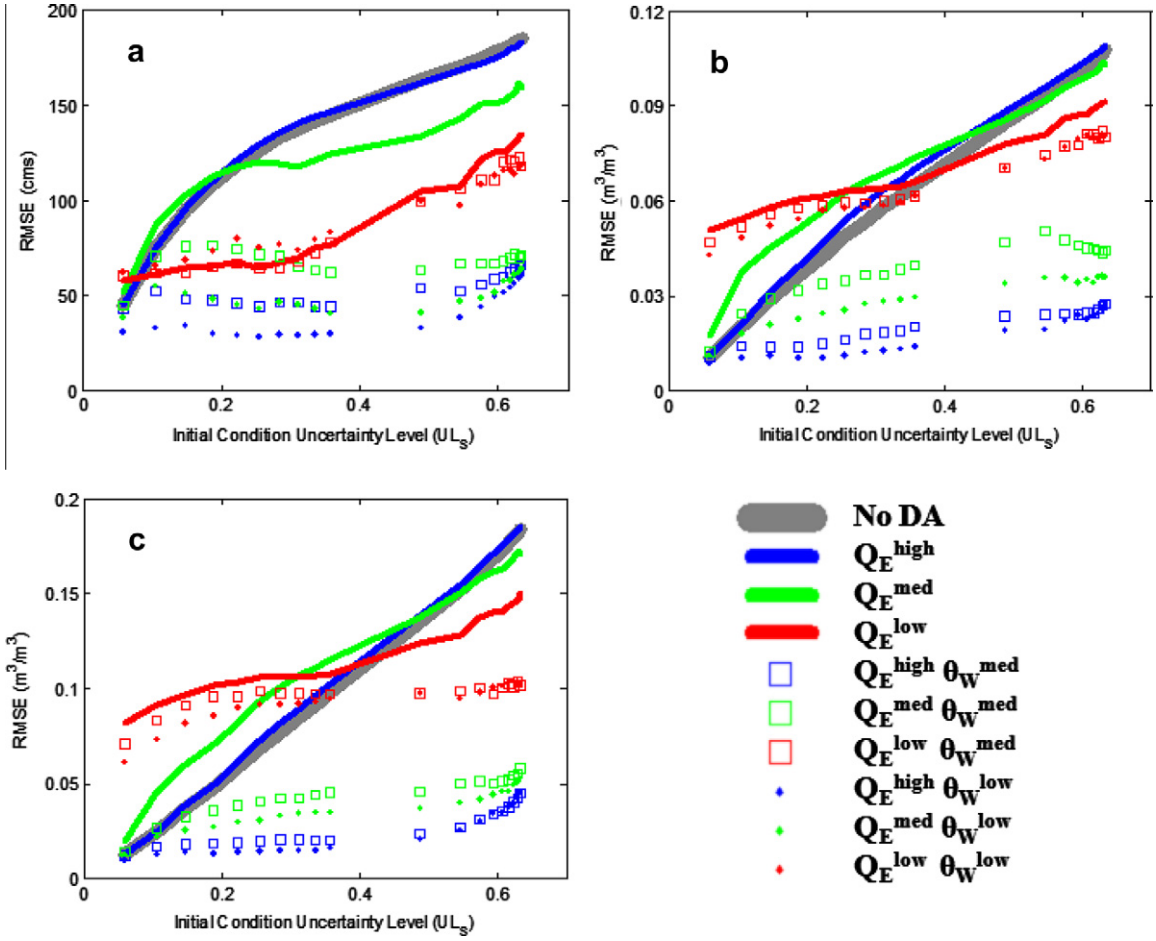

Fig. 4. The RMSE, as a function of the uncertainty level in the initial conditions (ULs), of model-simulated (a) streamflow, (b) soil moisture at 25 cm depth and (c) soil moisture at 60 cm depth with and without different combinations of data assimilated. In the above, Q_E denotes the streamflow observations at Eldon, and θ_W denotes the in situ soil moisture observations at Westville. Superscripts "low", "med" and "high" denote small, medium and large levels of uncertainties, respectively (see also Table 1).

DA as described in Section 4 for each of the 100 realizations in Step 3, and (6) Evaluate the streamflow and soil moisture results by comparing the prior, i.e., before- or without-DA, and the posterior, i.e., after- or with-DA, ensembles. Note that Step 4 allows accurate specification of the observation error variances for streamflow, soil moisture, and model initial soil moisture states. In Step 6, the prior are the model simulations generated in Step 2 with noise-added model initial soil moisture states.

To generate ensembles of the SAC states at the beginning of the assimilation window and those of the in situ soil moisture and streamflow observations within the assimilation window, we used the three synthetic error models described in Appendix A. In this work, the observation error variances for soil moisture and SAC states were assumed to be homoscedastic whereas that for streamflow was assumed to be heteroscedastic (see Appendix A). The amount of noise added is controlled by the three parameters, σ_θ, C_Q, and C_S (see Appendix A for details). Based on visual examination of the perturbed observations and soil moisture states, we chose 2, 3 and 3 varying levels of uncertainty for σ_θ, C_Q, and C_S, respectively (see Table 1 for summary). Particular care was taken in estimating the observation error variance for in situ soil moisture data, which is described in detail in Appendix B. The interested reader is kindly advised to read Appendices A and B before proceeding.

To express the amount of uncertainty added through the perturbation in more easily understandable terms, we define the uncertainty levels for the initial model soil moisture states (UL_S), for the in situ soil moisture observations (UL_θ), and for the streamflow observations (UL_Q) as follows:

$$UL_S = \frac{1}{n_S n_C}\sum_{j=1}^{n_S}\sum_{i=1}^{n_C}\frac{1}{X_{S,j,i}^{\max}}\sqrt{\frac{1}{n_E}\sum_{l=1}^{n_E}\left(X_{S,j,i,K-L}^{l}-X_{S,T,j,i,K-L}\right)^2}, \tag{10}$$

$$UL_\theta = \frac{1}{nL}\sum_{k=K-L+1}^{K}\sqrt{\frac{1}{n_E}\sum_{l=1}^{n_E}\left(\theta_k^l-\theta_{T,k}\right)^2}, \tag{11}$$

$$UL_Q = \frac{1}{L}\sum_{k=K-L+1}^{K}\frac{1}{Q_{T,k}}\sqrt{\frac{1}{n_E}\sum_{l=1}^{n_E}\left(Q_k^l-Q_{T,k}\right)^2}, \tag{12}$$

In the above, n_E denotes the number of ensemble traces, $X_{S,j,i,K-L}^{l}$ denotes the lth ensemble trace of the jth SAC state at the ith cell at the beginning of the assimilation window (i.e., k = K-L), θ_k^l denotes the lth ensemble trace of the synthetically-generated depth-specific soil moisture at hour k, and Q_k^l denotes the lth ensemble trace of the synthetically-generated streamflow at hour k, n denotes porosity (0.47 at Westville), $X_{S,j,i}^{\max}$ denotes the upper bound of the jth SAC state at the ith cell, and the subscript T signifies that the variable subscripted is the model-simulated assumed truth. As defined above, UL_S represents the uncertainty in the initial model soil moisture states normalized by $X_{S,j,i}^{\max}$, UL_θ represents the uncertainty in the in situ soil moisture observations within the assimilation window normalized by porosity, and UL_Q represents the uncertainty in the streamflow observations within the assimilation window normalized by the true streamflow $Q_{T,k}$. The uncertainty level UL_S is

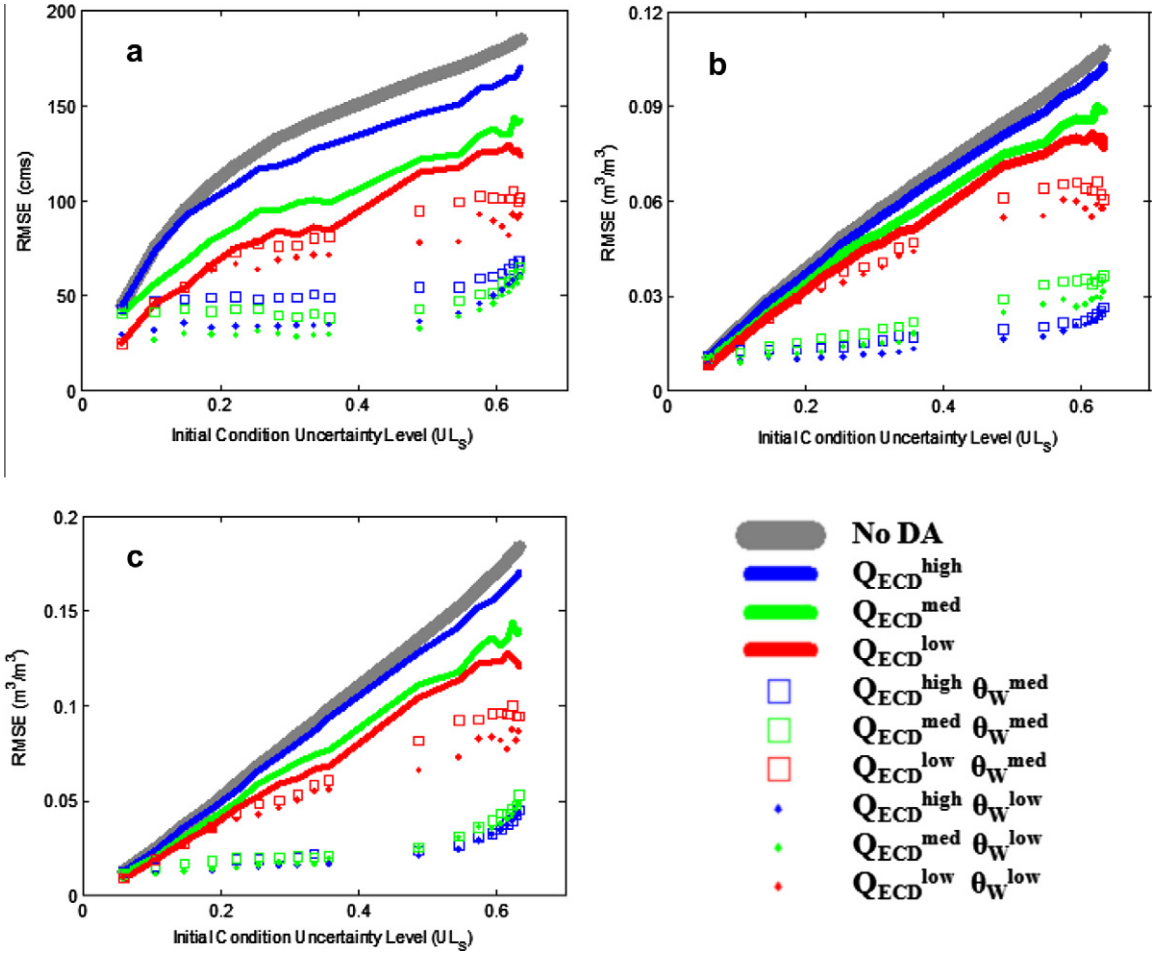

Fig. 5. Same as Fig. 4, but for the cases of simultaneously assimilating streamflow observations at Eldon, Christie and Dutch. In the above, Q_{ECD} denotes the streamflow observations at Eldon, Christie and Dutch collectively.

calculated for the five SAC states, UZTWC, UZFWC, LZTWC, LZFSC and LZFPC, for all cells within the basin.

The synthetic experiment described above requires solving the minimization problem in Eqs. (8) and (9) 100 times for each combination of the uncertainties prescribed. Due to excessive computational burden, we could not choose a lengthy simulation period or multiple events. For the synthetic experiment, we chose the June 20–26, 2000, event which produced a peak flow of 1,549 cms, the largest observed flow between 1997 and 2002.

6.2. Results

Our primary interest in this work was in assessing the potential of DA in the context of single-valued prediction. As such, here we use performance measures for single-valued prediction instead of those for ensemble or probabilistic forecasts [53–55]. Throughout the rest of this paper, streamflow at Eldon is denoted by Q_E, and streamflows at Eldon, Christie and Dutch are denoted collectively by Q_{ECD} for notational brevity. Soil moisture at Westville is denoted by θ_W. The prediction time (K in Fig. 2) used to generate the results in Figs. 4–9 corresponds to Hour 36 in Fig. 9, which is in the lower part of the falling limb of the hydrograph where the effects of precipitation forcing and fast runoff have worn off. The analysis results in the above figures hence represent the DA-aided model simulation over the assimilation window of hours 1–36 in Fig. 9, which includes all of the rising limb and the upper part of the falling limb of the hydrograph.

Figs. 4 and 5 show the root mean squared error (RMSE) of the model-simulated, with and without DA, streamflow and soil moisture at 25 and 60 cm depths within the assimilation window as a function of the uncertainty level in the initial model soil moisture conditions (UL_S, see Eq. (10)). The RMSE is calculated using the model-simulated streamflow or soil moisture at all cells within the basin over all 100 ensemble traces and all hours within the assimilation window of 36 h. As such, it represents an average accuracy of the model-simulated streamflow or soil moisture with and without DA at some cell at some hour within the assimilation window. Note in interpreting the streamflow results of Figs. 4 and 5 that, because streamflow at downstream cells is much larger than that at upstream cells, the RMSE of streamflow is necessarily weighted more heavily toward the downstream cells.

The results for assimilating Q_E in Fig. 4 may be summarized as follows. Assimilating highly uncertain streamflow observations at Eldon (i.e., the outlet) only, or Q_E^{high}, does not improve streamflow analysis over the control (i.e., DA-less) simulation regardless of the level of uncertainty in the initial model soil moisture states. When the initial model soil moisture states have a medium level of uncertainty ($UL_S > 0.3$), assimilating streamflow data with a medium level of observational uncertainty improves streamflow analysis noticeably (see the Q_E^{med} result). When the streamflow data is accurate, DA improves streamflow analysis substantially (see the Q_E^{low} result) except when the initial conditions are already known extremely accurately ($UL_S < 0.1$). Assimilating Q_E improves soil moisture analysis somewhat only if the streamflow observations

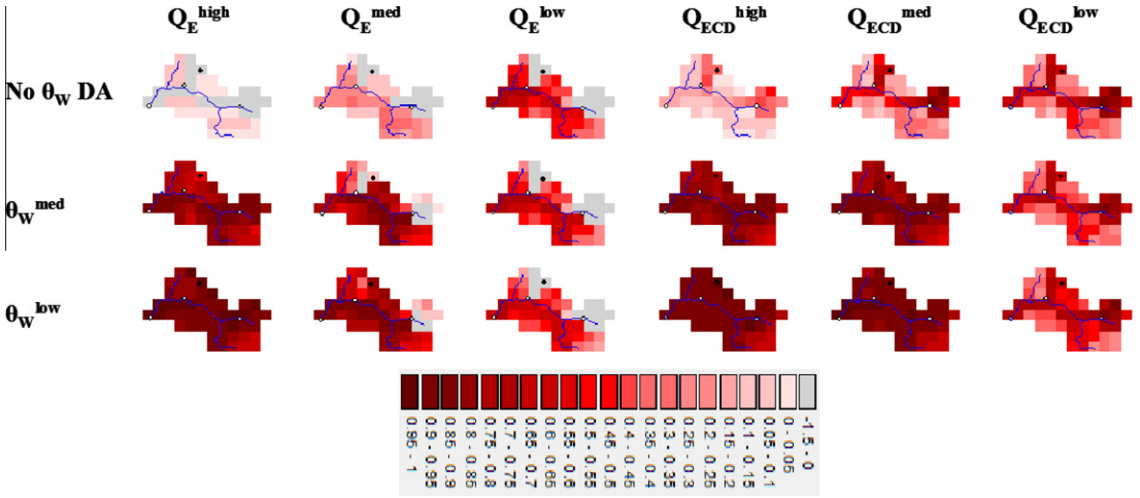

Fig. 6. The Skill Score (SS) map for streamflow analysis for UL_S (the initial condition uncertainty level) = 0.36. The 1st, 2nd and 3rd rows correspond to the assimilation results without θ_W, with θ_W^{med} and with θ_W^{low}, respectively. The superscripts, "low", "med", and "high" denote the low, medium and high levels of uncertainty, respectively, in the data (see also Table 1). In the figure, Q_E denotes streamflow at Eldon and Q_{ECD} denotes streamflow at Eldon, Christie, and Dutch collectively.

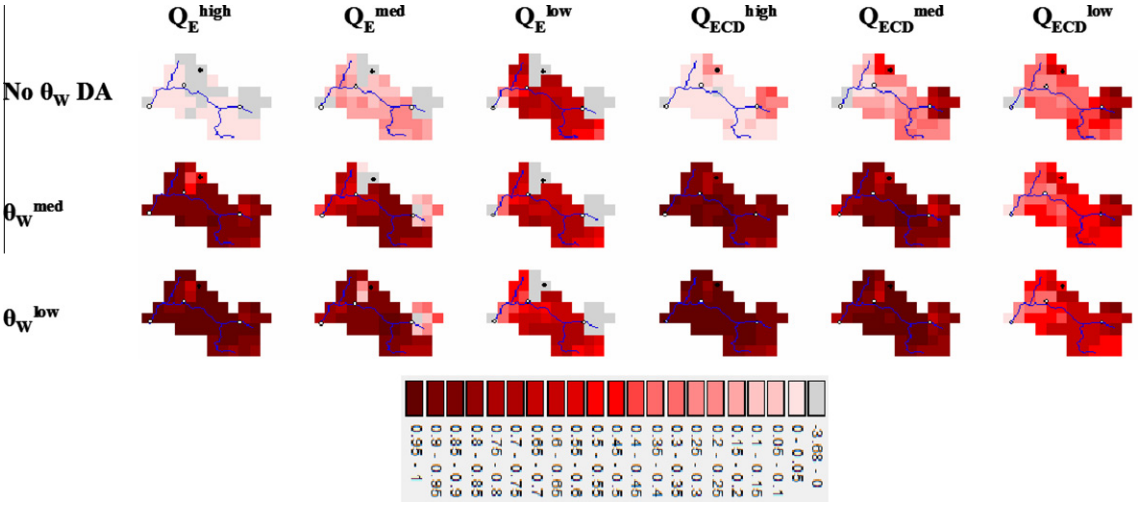

Fig. 7. Same as Fig. 6, but for soil moisture at 25 cm depth.

are highly accurate and the initial conditions are uncertain ($UL_S > 0.4$).

The results for assimilating both Q_E and θ_W in Fig. 4 may be summarized as follows. If Q_E is highly uncertain $\left(Q_E^{high}\right)$, additionally assimilating θ_w (either θ_W^{med} or θ_W^{low}) greatly improves both streamflow and soil moisture analysis. If Q_E is accurate $\left(Q_E^{low}\right)$, however, additionally assimilating θ_w (either θ_W^{med} or θ_W^{low}) provides little improvement to analysis of streamflow and soil moisture at 25 cm depth over assimilating Q_E alone. Also seen in Fig. 4 is that assimilating θ_W in addition to Q_E^{low} results in a larger RMSE than assimilating θ_W in addition to either Q_E^{med} or Q_E^{high}. The last finding is counterintuitive and suggests that the effects of assimilating Q_E and θ_W as carried out in this work may be in conflict, and that uncertainty modeling and modeling of the observation equation for in situ soil moisture may need improvement. It is also postulated that the counterintuitive results may be due to overfitting [56], i.e., the inverse problem may be too underdetermined. Our ongoing work includes reducing DOF by increasing the space–time scales of adjustment and evaluating performance of the resulting data assimilation procedure.

Note also in Fig. 4b and c that, if the initial model soil moisture states are known relatively accurately ($UL_S < 0.3$), assimilating Q_E^{low} (with or without θ_W) as carried out in this work may deteriorate the DA-less soil moisture analysis, another indication that uncertainty modeling needs improvement and/or the inverse problem may be too underdetermined. Fig. 5 shows that assimilating streamflow observations at all locations of Eldon, Christie and Dutch, or Q_{ECD}, improves streamflow analysis somewhat over assimilating Q_E alone (shown in Fig. 4) and, if the streamflow observations are accurate $\left(Q_{ECD}^{low}\right)$, also improves soil moisture analysis somewhat. Assimilating θ_W in addition to Q_{ECD}^{high} or Q_{ECD}^{med}, on the other hand, greatly improves streamflow and soil moisture analysis over assimilating Q_{ECD} alone. Similarly to assimilating θ_w in addition to Q_E (Fig. 4), assimilating θ_w in addition to Q_{ECD}^{low} yields counterintuitive results in that they produce a larger RMSE than assimilating θ_w in addition to Q_{ECD}^{med} or Q_{ECD}^{high} (See the explanations and comments above on Fig. 4). Note that Figs. 4 and 5 are based on relatively wet initial soil moisture conditions. If different initial soil moisture conditions are used, the overall picture may be different [57]. To assess this sensitivity, additional work is necessary.

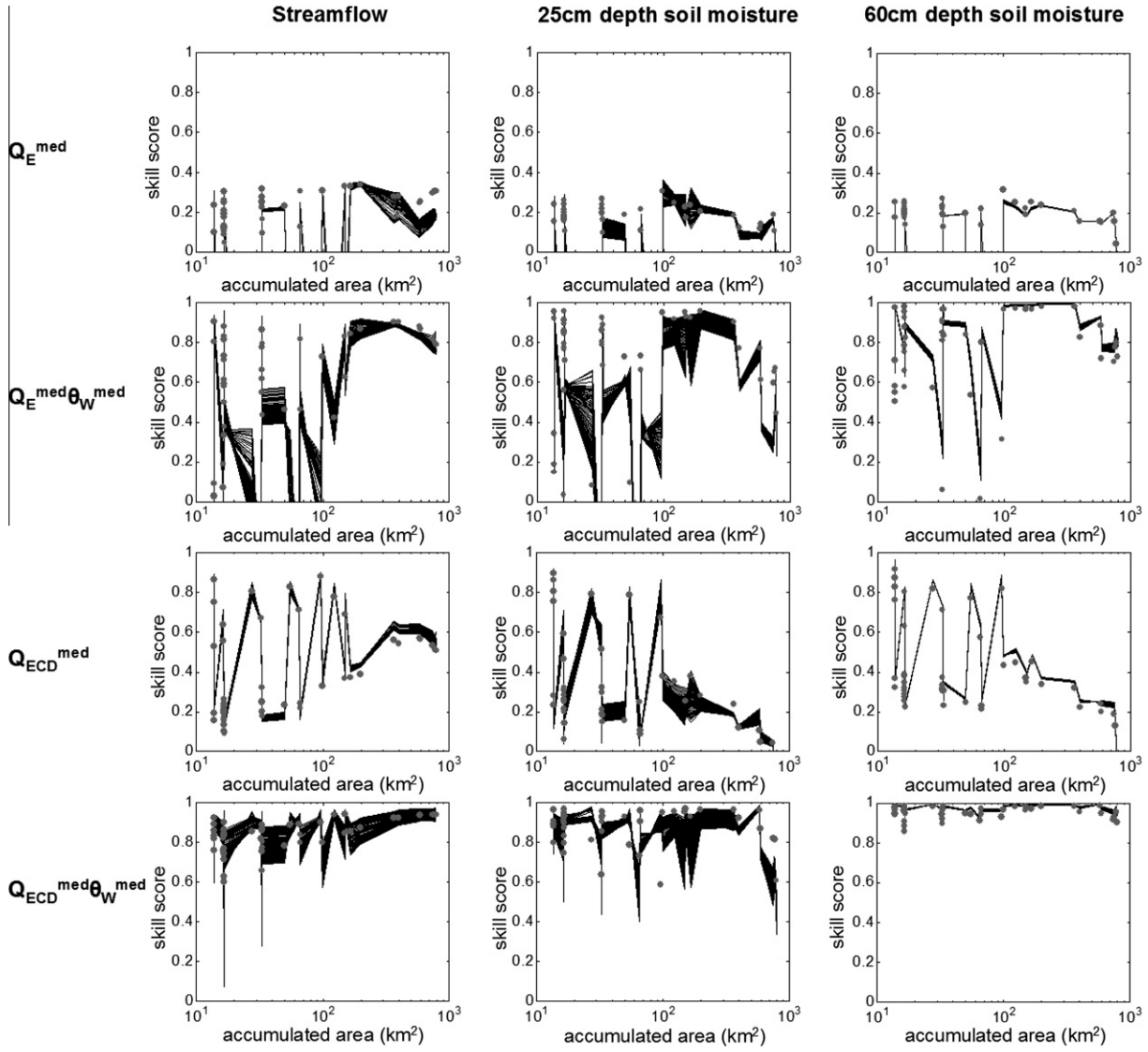

Fig. 8. The Skill Score (SS) of analysis of streamflow (1st column), soil moisture at 25 cm depth (2nd column), and soil moisture at 60 cm depth (3rd column) with assimilation of Q_E^{med} (1st row), Q_E^{med} and θ_W^{med} (2nd row), Q_{ECD}^{med} (3rd row), and both Q_{ECD}^{med} and θ_W^{med} (4th row). The superscript "med" denotes the medium level of uncertainty in the data (see also Table 1) and Q_E, Q_{ECD}, and θ_W denote streamflow at Eldon, streamflow at Eldon, Christie and Dutch collectively, and in situ soil moisture at Westville, respectively. In the above, UL_S (the initial condition uncertainty level) = 0.36 was used (see Table 1). Circles denote SS of DA-aided analysis of streamflow or soil moisture within the assimilation window. Solid lines denote SS of prediction of streamflow or soil moisture for lead times of 1 through 72 h. The assimilation window corresponds to 1–36 h (see the x-axis of Fig. 9).

Figs. 6 and 7 show the spatial pattern of the Skill Score (SS) of DA-aided analysis of streamflow and soil moisture at 25 cm depth, respectively. The SS, which measures the quality of DA-aided simulation relative to the DA-less, is defined as follows:

$$SS_i = 1 - \frac{MSE_{DA}}{MSE}, \quad i = 1, \ldots, n_C \tag{13}$$

where MSE and MSE_{DA} denote the mean squared errors (MSE) of DA-less and DA-aided analysis, respectively. The value of SS ranges from 1 (DA is perfect) to 0 (DA adds nothing) to negative (DA makes worse). Figs. 6 and 7 may be summarized as follows. As expected, assimilating Q_{ECD} improves both streamflow and soil moisture analysis over assimilating Q_E alone. Fig. 6 shows that the improvement made at Christie and Dutch propagates to the downstream cells along the main stem of the channel. Improvement in soil moisture analysis by assimilating Q_{ECD} over assimilating Q_E is seen in Fig. 7. Expectedly, the improvement is limited mainly to Christie and Dutch. The contrast in the spatial pattern of SS between streamflow and soil moisture reflects the fact that the SAC models soil moisture dynamics only in the vertical whereas the impact of streamflow assimilation is routed horizontally through the channel network.

As also observed in Figs. 4–7 indicate that additionally assimilating θ_W greatly improves SS for analysis of streamflow and soil moisture over assimilating Q^{high} or Q^{med} alone, but not over assimilating Q^{low} alone regardless of the availability of interior gauge observations for assimilation. Clearly, we owe the above dramatic improvement in SS from assimilating θ_W in addition to Q^{high} or Q^{med} to the fact that, through the controlled experiment, we removed the model structural and parametric errors as well as the input errors. Also, in the controlled experiment, the in situ soil moisture observations were given full dynamic range (from residual water content to porosity) and were not subject to the microscale variability of point soil moisture within an HRAP cell. In addition, through the controlled experiment, we were able to quantify at least the univariate statistics of the observation

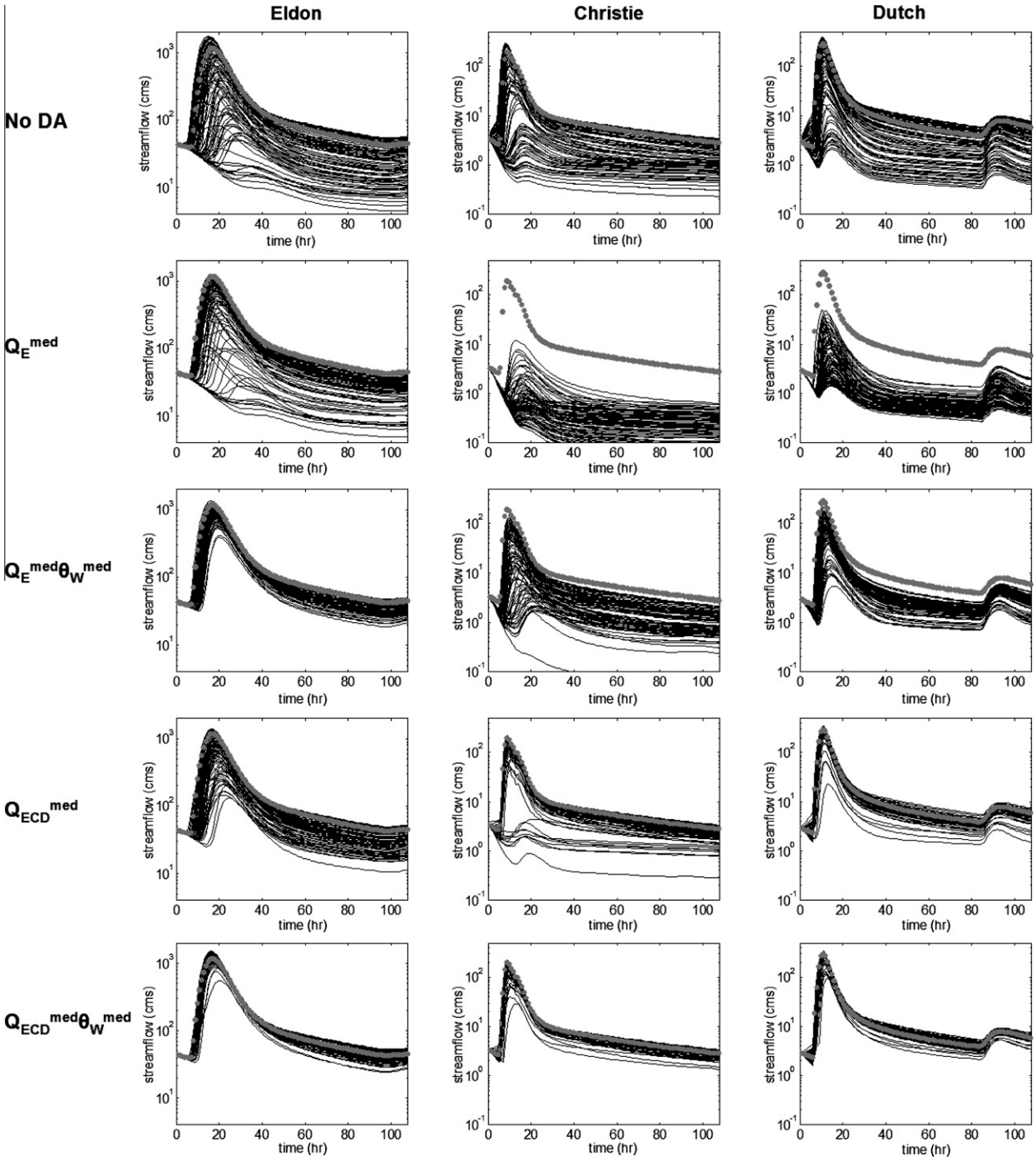

Fig. 9. Ensembles of simulated streamflow (solid lines) for Eldon (1st column), Christie (2nd column) and Dutch (3rd column) without assimilation (1st row), and with assimilation of Q_E^{med} (2nd row), $Q_E^{med}\theta_W^{med}$ (3rd row), Q_{ECD}^{med} (4th row), $Q_{ECD}^{med}\theta_W^{med}$ (5th row), respectively. The superscript "med" denotes the medium level of uncertainty in the data (see also Table 1) and Q_E, Q_{ECD}, and θ_W denote streamflow at Eldon, streamflow at Eldon, Christie and Dutch collectively, and in situ soil moisture at Westville, respectively. In the above, UL_S (the initial condition uncertainty level) = 0.36 was used (see Table 1). The assimilation window corresponds to 1 to 36 h. The dotted line denotes the assumed truth.

error variances accurately (albeit only homoscedastically for those of the background model states and soil moisture). In reality, one may expect the positive impact of assimilating streamflow or streamflow and soil moisture observations to be significantly smaller.

Fig. 8 shows the SS for analysis and prediction of streamflow (1st column), soil moisture at 25 cm depth (2nd column), and soil moisture at 60 cm depth (3rd column) from assimilating Q_E^{med} (1st row), Q_E^{med} and θ_W^{med} (2nd row), Q_{ECD}^{med} (3rd row), and Q_{ECD}^{med} and θ_W^{med} (4th row). In each plot, the x-axis denotes the cell number. The cells are numbered in an increasing order of the contributing area. As such, for streamflow (1st column), the leftmost SS values are associated with the SS of DA-aided streamflow generated from single cells, and the rightmost SS value is associated with that from the entire basin. The same cell numbering is used for the SS plots of DA-aided soil moisture (2nd and 3rd columns). In each plot, the circles denote the SS of DA-aided hourly analysis of streamflow or soil moisture within the assimilation window. The solid lines,

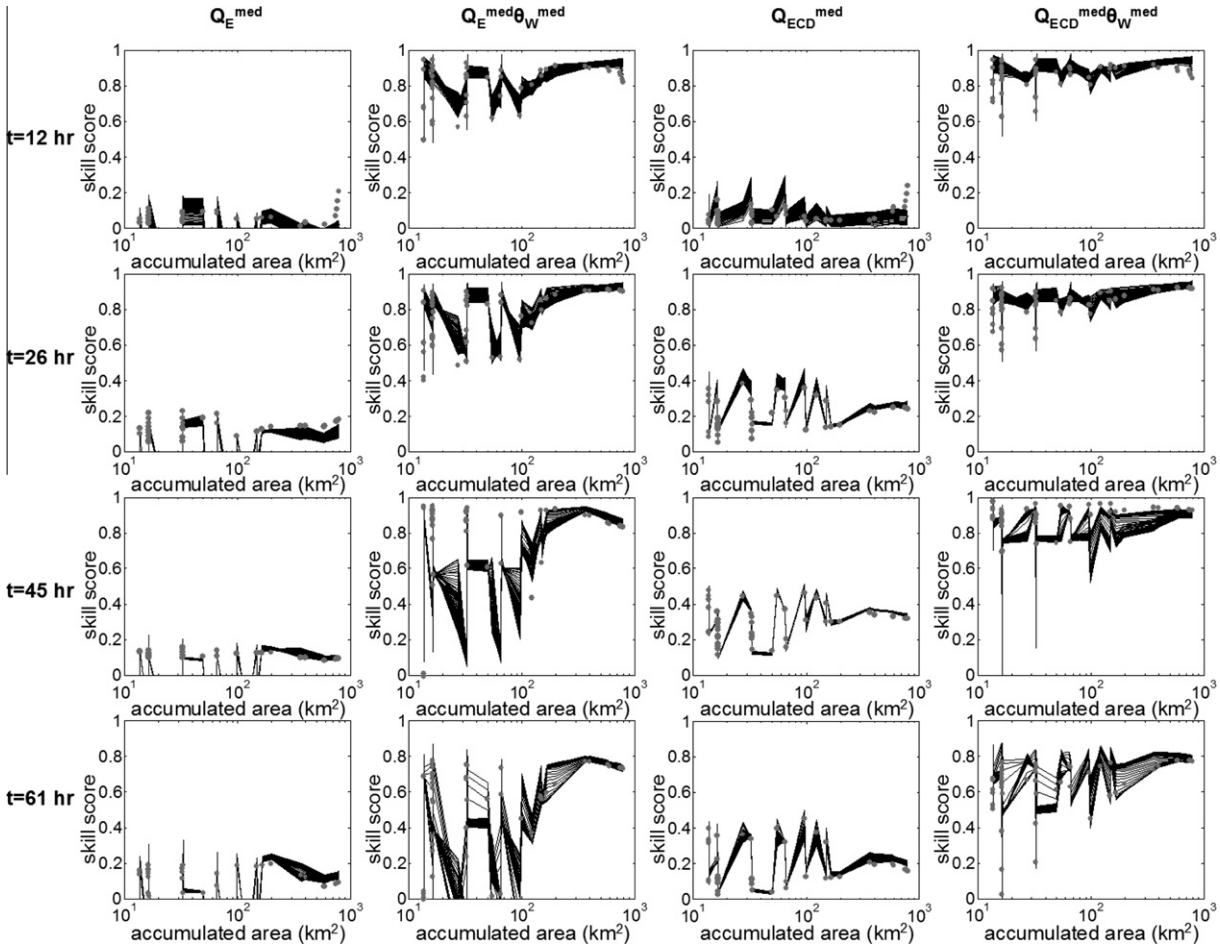

Fig. 10. The Skill Score (SS) of DA-aided analysis (dots) and prediction from 1 to 72 h of lead time (solid lines) of streamflow at all grid cells (sorted in the increasing order of the contributing area on the x-axis). The prediction times are, from the top to the bottom row, 12, 26, 45 and 61 h (see the x-axis of Fig. 9). The 1st through 4th columns correspond to assimilating Q_E^{med}, $Q_E^{med}\theta_W^{med}$, Q_{ECD}^{med}, and $Q_{ECD}^{med}\theta_W^{med}$, respectively. The superscript "med" denotes the medium level of uncertainty in the data (see also Table 1) and Q_E, Q_{ECD}, and θ_W denote streamflow at Eldon, streamflow at Eldon, Christie and Dutch collectively, and in situ soil moisture at Westville, respectively. For all cases, UL_S (the initial condition uncertainty level) = 0.59 was used (see Table 1).

Table 1
Uncertainty levels for initial SAC states, soil moisture observations and streamflow observations, respectively (see Eqs. (10)–(12)) for definitions). In the table, C_S, $\sigma_\theta$and C_Q denote the perturbation coefficients used to generate synthetic observations of the initial SAC states ($X_{S,k=K-L}$), soil moisture at Westville (θ_W) and streamflows at Eldon (Q_E), Christie (Q_C) and Dutch (Q_D), respectively (see Appendix A for details), K and L denote the current hour and the length of the assimilation window, respectively, and $[a,b]$ denotes the closed range of a perturbation coefficient between a and b.

Variables	Perturbation coefficient	Coefficient value	Uncertainty level	Symbol	Uncertainty level value
Initial SAC states ($X_{S,k=K-L}$)	C_S	[0.001,0.005]	Low	UL_S	[0.06,0.22]
		[0.006,0.01]	Medium		[026,0.36]
		[0.02,0.1]	High		[0.49,0.63]
Soil moisture observations (θ_W)	σ_θ	0.01	Low	UL_θ	0.03
		0.03	Medium		0.07
Streamflow observations (Q_E, Q_C, Q_D)	C_Q	0.01	Low	UL_Q	0.01
		0.1	Medium		0.11
		0.3	High		0.33

which connect the data points only to aid visualization, correspond to the SS of streamflow or soil moisture prediction for lead times of 1–72 h. In general, the shorter the lead time is, the higher the SS is. In Fig. 8, the four plots from top to bottom in the 1st and 2nd columns correspond to the (1,2)-nd, (2,2)-nd, (1,5)th and (2,5)th SS maps in Figs. 6 and 7, respectively. In the upper half of Fig. 8, there are cells with negative SS values which have been truncated at zero for comparison with the rest of the figure; it suggests that the DA problem may be significantly underdetermined when assimilating Q_E^{med} or $Q_E^{med}\theta_W^{med}$. Fig. 8 also shows that the SS has large variability up to a spatial scale of about 165 km^2 of the contributing area; it suggests that the impact of underdeterminedness is greater for upstream locations. Fig. 8 may be summarized as follows. Assimilating Q_E (see the Q_E^{med} result) has significant positive impact on analysis and prediction of streamflow and soil moisture. The impact tends to be larger at downstream locations for which the contributing areas are larger. Assimilating θ_W in addition to Q_E (see the $Q_E^{med}\theta_W^{med}$ result) greatly improves analysis and prediction of

streamflow and soil moisture at all ranges of contributing area over assimilating Q_E alone (see the Q_E^{med} result). Compared to assimilating Q_E only (see the Q_E^{med} result), assimilating Q_{ECD} (see the Q_{ECD}^{med} result) provides significant positive impact for both analysis and prediction of streamflow and soil moisture. Assimilating θ_W in addition to Q_{ECD} (see the $Q_{ECD}^{med}\theta_W^{med}$ result) has a very large positive impact for analysis and prediction of streamflow and soil moisture. It also greatly reduces cell-to-cell variability of skill in the DA-aided analysis and prediction in the upstream areas, an indication that, under the idealized conditions, soil moisture observations may reduce the under-determinedness of the inverse problem significantly.

Fig. 9 shows the 100 ensemble traces of perturbed (solid lines in the 1st-row plots) and DA-aided (solid lines in the 2nd- to 5th-row plots) hydrographs at the three stream gauge locations of Eldon (1st column), Christie (2nd column) and Dutch (3rd column) for UL_S of 0.36 (see Eq. (10) and Fig. 5 for interpretation). These plots provide a visual sense of the quality of the DA-aided analysis that the statistical performance measures and skill scores may not. Fig. 9 may be summarized as follows. Assimilating Q_E^{med} only (2nd row) does not improve streamflow simulation at Christie and Dutch. Assimilating θ_W^{med} in addition to Q_E^{med} (3rd row) improves streamflow simulation at Christie and Dutch as well as at the outlet. Compared to assimilating Q_E^{med} only, assimilating Q_{ECD}^{med} (4th row) improves streamflow simulations at Christie and Dutch substantially and at the outlet noticeably. Assimilating both Q_{ECD}^{med} and θ_W^{med} (5th row) further improves streamflow simulation.

The results presented above are limited to analysis over the assimilation window ending at hour 36 and prediction out to Hour 108 in Fig. 9. Because DOF associated with the DA problem vary according to the state of the system, one may expect the performance of DA to vary accordingly. To assess this variability, we made the same analysis and prediction runs as those seen in Fig. 9 using different points on the hydrograph as the beginning of the assimilation window. Fig. 10 shows the SS of DA-aided hourly analysis and prediction of streamflow over the 36- and 72-h assimilation window and prediction horizon, respectively, at all cells at the prediction times of Hours 12, 26, 45 and 61 for UL_S of 0.59 (this corresponds to a high level of uncertainty in the initial model soil moisture states; see Table 1). The four prediction times correspond to the early part of the rising limb, the upper part of the falling limb (controlled by the lower-zone supplemental free water), the mid- to low part of the falling limb (controlled by the lower-zone primary free water), and the lower part of the falling limb of the hydrograph, respectively (see Fig. 9). We may summarize Fig. 10 as follows. Assimilating Q_E^{med} consistently has positive impact on streamflow analysis for downstream cells. Assimilating θ_W^{med} in addition to Q_E^{med} greatly increases the SS for analysis and prediction of streamflow over all ranges of the contributing area. The positive impact is greater when the rising limb or the upper part of the falling limb is being predicted. Assimilating Q_{ECD}^{med} consistently has significant positive impact across all ranges of the contribution area. Assimilating θ_W^{med} in addition to Q_{ECD}^{med} has a very large positive impact, particularly for prediction during the rising limb or the early part of the falling limb of the hydrograph. These observations in Fig. 10 suggest that additionally assimilating soil moisture data improves skill most significantly during the rising limb of the hydrograph, and that the margin of improvement decreases as the prediction time moves toward the lower part of the receding limb.

7. Real-world experiment

In the real world, the models usually have significant structural and parametric errors. To rigorously assess the value of assimilating streamflow and soil moisture data in the real-world conditions, large-scale experiments using multiple basins would be necessary. Such experiments are, however, a very large challenge. In the US, there are only a small number of headwater basins that are unregulated and instrumented for streamflow and in situ soil moisture observations with sufficient density and quality. We are currently identifying such basins and collecting data for multi-basin evaluation of the DA procedure developed in this work, the results of which will be reported in the near future. Here, we carried out exploratory evaluation to aid the design of such an experiment and to gain glimpse into the performance of DA in the real-world conditions.

7.1. Experiment design

As in the synthetic experiment, we assumed perfectly known precipitation and PE forcing. Though less than realistic, this assumption was made to facilitate comparisons with the synthetic experiment. The simulation period used was September 5, 2001, through September 5, 2002. We chose 1 and 10 $(cms)^2$ as reasonable lower and upper bounds, respectively, for the homoscedastic variance of the streamflow observation error, $\sigma_{Q,l}^2$ in Eq. (8). The observation error variance for the in situ soil moisture data, $\sigma_{\theta,i,m}^2$ in Eq. (8), was specified as described in Appendix B. The background model state error variance, $\sigma_{B,j,i}^2$ in Eq. (8), was assumed to be spatially homogeneous, and specified by the sample variance of the corresponding simulated SAC states for the 1-yr period above. This strategy assumes that the uncertainty in the model-simulated soil moisture states at any time is comparable to the mean variability of the model-simulated soil moisture over the 1-yr period. Because it uses the same mean error variance for all situations, the above strategy is likely to under-/overprescribe the amount of uncertainty in the initial model soil moisture states in the storm/inter-storm periods, resulting possibly in under-/overadjustment of the model soil moisture states. In evaluating the performance of DA in the real-world experiment, we also assumed that the streamflow and in situ soil moisture data are free of observational errors. This practice in the evaluation process is clearly at odds with prescribing uncertainty for streamflow and in situ soil moisture observations in the assimilation process. While unsatisfactory, the above is unavoidable given that the information on the statistical properties of the real observational errors is lacking.

7.2. Results

Fig. 11 shows the RMSE of simulated streamflow at Eldon, Christie, and Dutch as a function of lead time. The shaded area and the two dotted lines represent the RMSE bounds obtained from assimilating streamflow data with $\sigma_{Q,l}^2$ of 1 and 10 $(cms)^2$, respectively. For each choice of $\sigma_{Q,l}^2$, the upper bound of RMSE is associated with the lower bounds of the standard deviation of the total error in soil moisture observation of 0.036 m^3/m^3 at 25 cm and 0.043 m^3/m^3 at 60 cm of depth (see Table B.1). The lower bound of RMSE of the shaded area is associated with a hypothetically large standard deviation of the total error in soil moisture observation of 0.25 m^3/m^3at both 25 and 60 cm depths. This value is over a half of the porosity at Westville and represents nearly non-informative soil moisture observation. The following observations may be made in Fig. 11. Assimilating streamflow data at the outlet improved analysis and prediction of streamflow at the outlet as expected, but improved little at the interior locations (mid- and bottom-left plots). Similar results have been obtained also by [17]. Assimilating both the streamflow data at the outlet and the soil moisture data at Westville tended to deteriorate streamflow analysis and prediction at the interior locations compared to

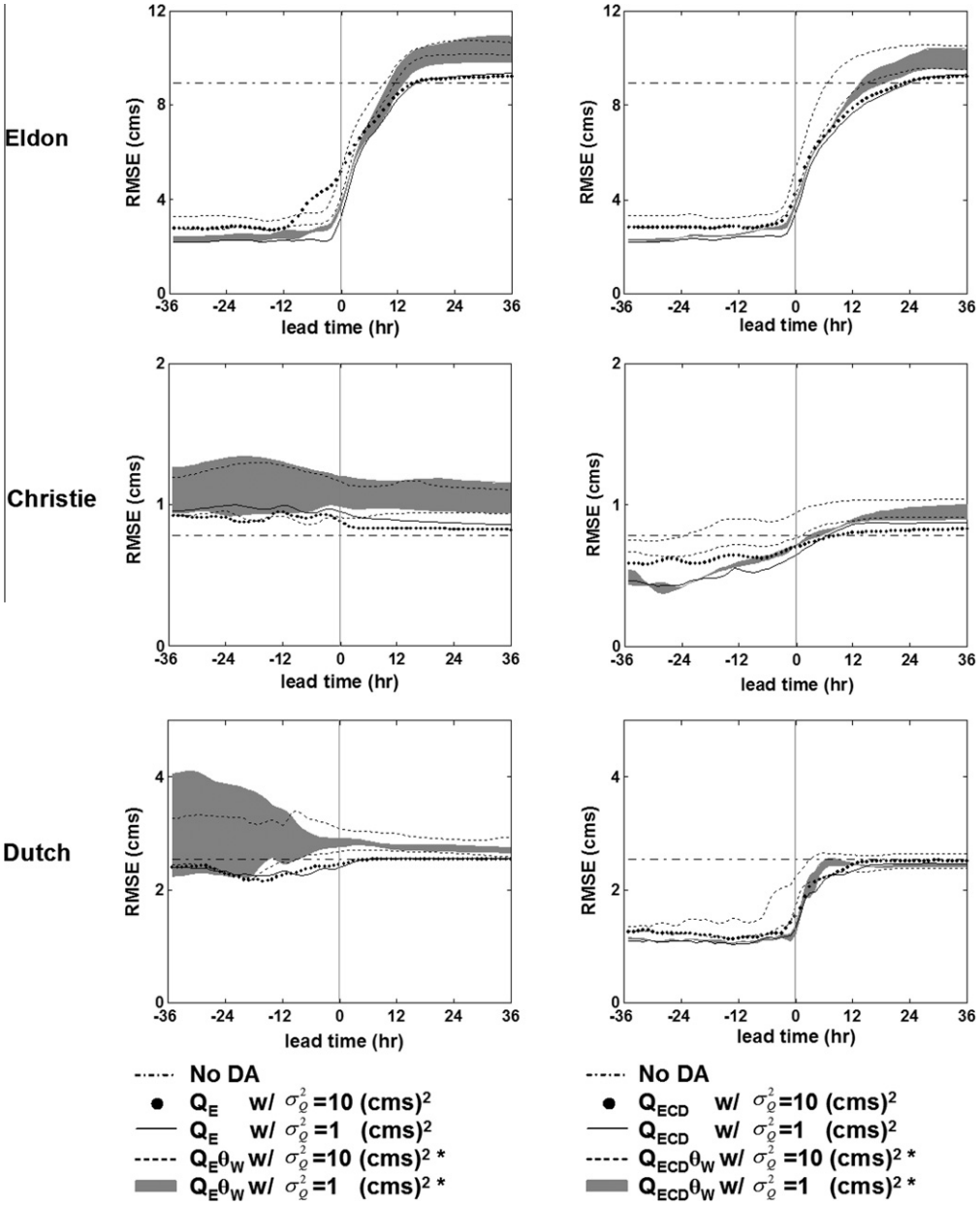

Fig. 11. The RMSE of simulated streamflow at all stream gauge locations. In the above, Q_E, Q_{ECD}, and θ_W denote streamflow at Eldon, streamflow at Eldon, Christie and Dutch collectively, and in situ soil moisture at Westville, respectively. Values of 1 and 10 $(\text{cms})^2$ were used for the variance of the streamflow observation error $\left(\sigma_{Q}^2\right)$. $*$ The lower small-dotted line and the lower bound of the shaded area were obtained by using 0.25 m^3/m^3 as the standard deviations of the soil moisture measurement error at both 25 and 60 cm depths. The upper small-dotted line and the upper bound of the shaded area were obtained by using 0.036 and 0.043 m^3/m^3 as the standard deviation of the soil moisture measurement error at 25 and 60 cm depth, respectively.

assimilating streamflow data at the outlet only (mid- and bottom-left plots). In particular, assimilating soil moisture observations at Westville did not improve the streamflow results at Christie (65 km^2), whose drainage area includes Westville (see Fig. 3(a)). As expected, assimilating streamflow data at the outlet and interior locations improved analysis and prediction of streamflow at the stream gauge locations (right plots) and also improved streamflow prediction at the outlet compared to assimilating streamflow data at the outlet alone (top plots). However, assimilating both the soil moisture data and the streamflow observations at the outlet and interior locations tended to deteriorate analysis and prediction of streamflow at the gauge locations compared to assimilating only the streamflow data at all gauge locations (right plots). Fig. 12 shows the scatter plots of soil moisture analysis at the prediction time (i.e., lead time of 0 h) from assimilating streamflow observations at all gauge locations (left plots) and assimilating both the soil moisture data and the streamflow observations at all locations (right plots) with $\sigma_{Q,I}^2 = 10\ (\text{cms})^2$. As expected, assimilating soil moisture observations improved soil moisture analysis at the measurement location, particularly at 60 cm. However, assimilating all available streamflow observations showed little impact on soil moisture analysis at Westville (left plots). Fig. 13 shows soil moisture and streamflow at 6-h lead time for October 9–13, 2001. Assimilating the soil moisture data in addition to the streamflow observations improves 60 cm depth soil moisture simulations at Westville in the early part of the rising limb. The increased soil moisture conditions, however, result in overestimating the peak flows, particularly at Eldon and Dutch. Note in the figure that the

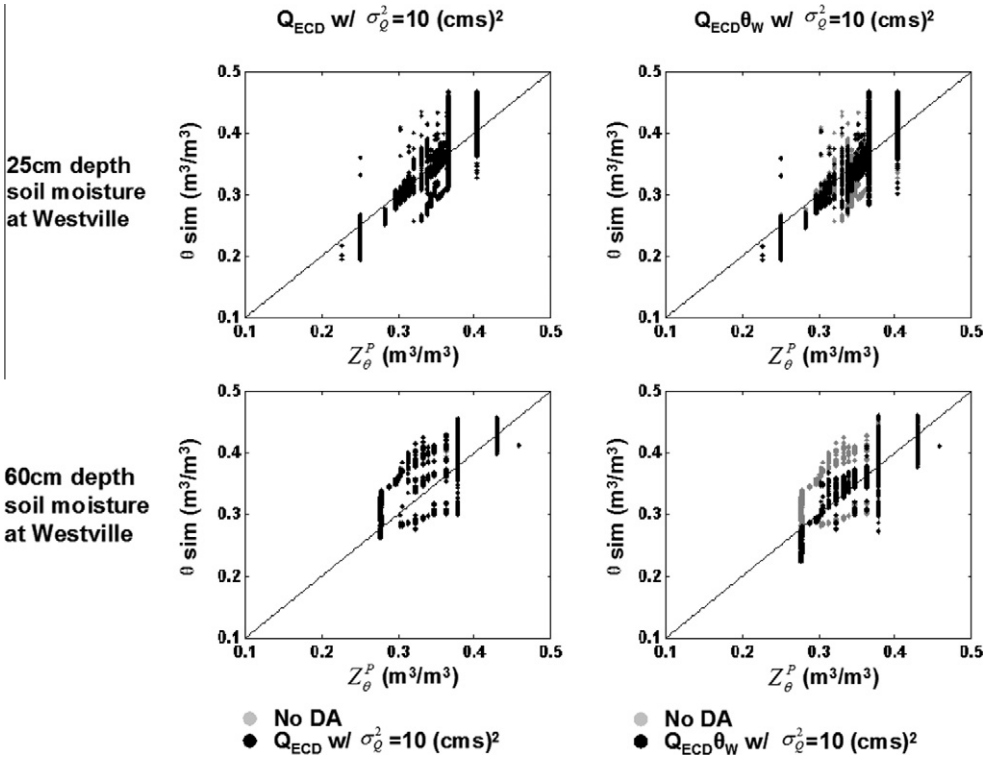

Fig. 12. Scatter plots of the pseudo soil moisture observations $\left(Z_\theta^P\right)$ and the simulated soil moisture (θ_{sim}) at 25 cm- and 60 cm-depths at Westville for lead time of 0 h. The left plots were generated by assimilating streamflow at Eldon, Christie and Dutch (Q_{ECD}) with $\sigma_Q^2 = 10\ (\text{cms})^2$, and the right plots by assimilating both in situ soil moisture at Westville (θ_W) and Q_{ECD} with $\sigma_Q^2 = 10\ (\text{cms})^2$. In the above, σ_Q^2 denotes the error variance of streamflow observations and Z_θ^P denotes the pseudo soil moisture observations at the HRAP scale converted from the point-scale in situ soil moisture data via probability matching. The grey and black dots denote the DA-less and DA-aided results, respectively.

25 cm depth soil moisture observations respond to precipitation about 2 h later than those at 60 cm depth. This is due to the effects of preferential flow following cracks formed in the soil after very dry conditions (Brad Illston, personal communications) which is not modelled in the SAC. The constant values of soil moisture observations exceeding 0.35 in Fig. 13 are due to the limited dynamic range, i.e., (wilting point, field capacity), of the soil moisture sensor (see Appendix B for details). These results, jointly with those from the synthetic experiment, suggest that a combination of structural and parametric errors in the hydrologic models, less than accurate uncertainty modeling, and observational uncertainties associated with the in situ soil moisture data used in this work is responsible for the lack of improvement by DA in this exploratory real-world experiment. We note here that modelling of the measurement error variance for soil moisture at the HRAP scale is particularly challenging; while probability matching used in this work reduces statistical biases in situ soil moisture observations [58,59] relative to the (unknown) soil moisture at the HRAP scale, we do not know the absolute accuracy of the soil moisture at the HRAP scale estimated in this way.

8. Summary, conclusions and future research recommendations

We assess the potential of updating soil moisture states of the Sacramento model [25] in the National Weather Service (NWS) Hydrology Laboratory's Research Distributed Hydrologic Model (HL-RDHM, [2]) via variational assimilation of streamflow and in situ soil moisture data for high-resolution analysis and prediction of streamflow and soil moisture. Assimilating streamflow and soil moisture data into distributed hydrologic models is new and particularly challenging due not only to large dimensionality of the inverse problem but also to nonlinearity and scale dependence in uncertainty propagation. To improve understanding of the problem and to assess the upper bound of the performance of the data assimilation (DA) procedure developed in this work under idealized conditions, we first designed and carried out a synthetic experiment. The assumptions for the idealized conditions include no structural and parametric uncertainties in the hydrologic models, perfectly known precipitation and potential evaporation (PE), in situ soil moisture observations with full dynamic range (i.e., from residual soil water content to porosity), and accurately estimated univariate statistics of the observation errors in streamflow and in situ soil moisture. To assess the performance of DA under more realistic conditions, we also designed and carried out an exploratory real-world experiment. The assumptions for the real-world experiment include perfectly known precipitation and PE and, for validation purposes only, observation error-free streamflow and in situ soil moisture (i.e., within the dynamic range between wilting point and field capacity) data. In both experiments, the control variables, i.e., the variables to be adjusted (or updated) via DA, are the SAC model states at the beginning of the assimilation window at each HRAP cell in the basin. The motivation for the simplification in this first phase of the research is to reduce the complexity of the problem in favour of improved understanding and easier interpretation even if it may compromise the goodness of the results. The basin used in the experiments is Eldon (ELDO2), a 795-km^2 headwater catchment located near the Oklahoma (OK) and Arkansas (AR) border in the southern plains of the U.S. The basin has two interior stream gauges at Christie (65 km^2) and Dutch (105 km^2) and an in situ Oklahoma Mesonet soil moisture measurement site at Westville, Oklahoma.

The main conclusions from the synthetic experiment are as follows. In general, the performance of the assimilation procedure

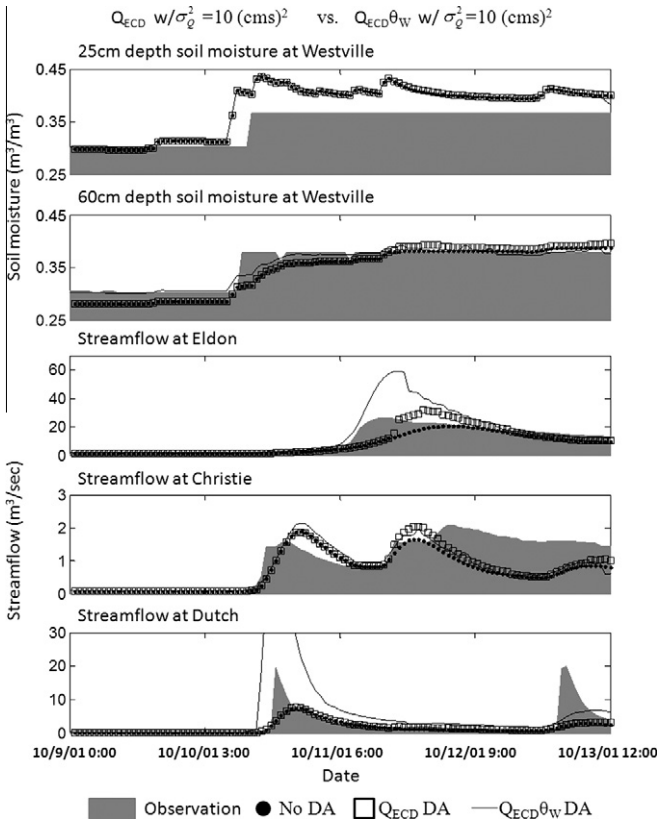

Fig. 13. Soil moisture and streamflow at 6-h lead time for October 9–13, 2001. The shaded area in the top two plots represents the probability-matched soil moisture observations at Westville (see Appendix B for details).

exhibits the expected sensitivity to the amount of uncertainty in the initial model soil moisture states and in the streamflow and in situ soil moisture data; the more uncertain the initial states and/or the more accurate the observations are, the larger the margin of improvement by DA is. A departure from this pattern was observed, however, when both the streamflow and the in situ soil moisture observations assimilated were very accurate, for which the performance of DA was significantly lower than expected. The above findings suggest that, under the idealized conditions, the DA procedure developed is generally capable of assimilating streamflow and in situ soil moisture data into distributed hydrologic models, but that uncertainty modeling needs improvement. While assimilating streamflow observations alone did not, in general, improve analysis and prediction of soil moisture, additionally assimilating in situ soil moisture observations provided very significant improvement under the idealized conditions. The above result indicates that assimilating streamflow and/or in situ soil moisture data, as carried out in this work, may pose a significantly underdetermined inverse problem, and suggests that the dimensionality of the control vector may need to be reduced to improve underdeterminedness. The main conclusions from the real-world experiment are as follows. In the real-world conditions considered in this work, assimilating streamflow or streamflow and in situ soil moisture data provided little improvement. Comparisons with the synthetic experiment suggests that a combination of structural and parametric errors in the hydrologic models, less than accurate modeling of scale-dependent and heteroscedastic uncertainties, and large observational uncertainty and microscale variability in the in situ soil moisture data is primarily responsible for the lack of improvement by DA. Additional research is necessary to ascertain the relative importance of these contributing factors. The DA procedure described in this work assumes that the models are capable of skilfully simulating the joint dynamics of streamflow and soil moisture over the catchment at the HRAP and 1-h scales. If this assumption is not met, one may not expect simultaneous assimilation at that scale of streamflow and soil moisture observations to be effective [60]. Hence, the future work should also include diagnosis of the realism of the model dynamics, as observed through streamflow and soil moisture data, at different temporal scales of aggregation. Such investigation is also likely to yield a more effective strategy for assimilating soil moisture data.

Acknowledgments

This work is supported by the Advanced Hydrologic Prediction Service (AHPS) program of the National Weather Service (NWS), by the Climate Predictions Program for the Americas (CPPA) of the Climate Program Office (CPO), both of the National Oceanic and Atmospheric Administration (NOAA), and by the Integrated Systems Solutions component of the NASA Decision Support through Earth Science Results Cooperative Agreement Notice (CAN) NN-H-04-Z-YO-010-Cof the National Aeronautic and Space Administration (NASA). These supports are gratefully acknowledged. The authors thank the Oklahoma Climatological Survey (OCS) for providing the in situ soil moisture data used in this work, and for Brad Illston of OCS for very helpful discussions. The first author is grateful to Ziya Zhang for providing help with HL-RDHM, to Fekadu Moreda for sharing knowledge and experience with the Oklahoma Mesonet data, and to Yuqiong Liu for critical review. We are also grateful to the four anonymous reviewers for many very helpful comments.

Appendix A. Error models for synthetic observations

To generate ensemble traces of synthetic soil moisture observations at the HRAP scale at Westville, we used:

$$\theta_k = \theta_{T,k} + \sigma_\theta w_k. \tag{A.1}$$

In the above, θ_k denotes the synthetically-generated depth-specific soil moisture observation at the HRAP scale at hour k, $\theta_{T,k}$ denotes the assumed true (as generated by HL-RDHM) soil moisture at the HRAP scale at hour k, where the subscript T denotes the model-generated assumed truth, σ_θ denotes the standard deviation of the synthetic soil moisture observation at the HRAP scale given the true soil moisture at the HRAP scale, and w_k denotes the temporally-correlated random noise of $N(0,1)$ at hour k. We modelled w_k in Eq. (A.1) as AR (1) with a decorrelation time of 10 h. Eq. (A.1) was used to generate synthetic soil moisture observations at 25 cm and 60 cm depths at the HRAP scale for the grid containing Westville. Perfect correlation was assumed between w_k at 25 cm depth and that at 60 cm depth based on the observation that the cross-correlation between the Oklahoma Mesonet soil moisture observation at 25 cm depth and that at 60 cm depth for the period of 1997–2002 was 0.97. It is reminded here that, unlike the real-world soil moisture observations from the Oklahoma Mesonet whose dynamic range is limited between wilting point and field capacity, the synthetic observations from Eq. (A.1) assume the full dynamic range from residual soil water content to porosity.

To generate synthetic streamflow observations we used:

$$Q_k = Q_{T,k} + \sigma_Q w_k, \tag{A.2}$$

where Q_k denotes the synthetically-generated streamflow observation at hour k, $Q_{T,k}$ denotes the assumed true streamflow at time k, σ_Q denotes the standard deviation of the synthetic streamflow observation given the true streamflow, and w_k denotes the temporally-correlated random noise of $N(0,1)$ at hour k. To specify σ_Q in

Eq. (A2), we used $\sigma_Q = C_Q\, Q_{T,k}$ where C_Q denotes the multiplicative factor (see Table 1). The random noise w_k in Eq. (A.2) was modeled as AR (1) with a decorrelation time of 10 h. Eq. (A.2) was used to generate synthetic streamflow observations at Eldon, Christie and Dutch.

To generate synthetic SAC states of UZTWC, UZFWC, LZTWC, LZFSC and LZFPC (see Section 3 for explanation) at the beginning of the assimilation window, we used:

$$X_{S,j,i,K-L} = X_{S,j,i,T,K-L} + X_{S,j,i}^{\max}(\exp(\alpha) - 1), \tag{A.3}$$

where

$$\alpha = -0.5\ln\left(1 + \left(C_S X_{S,j,i}^{\max}\right)^2\right) + \beta\sqrt{\ln\left(1 + \left(C_S X_{S,j,i}^{\max}\right)^2\right)}, \tag{A.4}$$

In Eqs. (A.3) and (A.4), $X_{S,j,i,K-L}$ denotes the synthetic observation of the jth SAC state, $X_{S,j}$, at the ith grid at the beginning of the assimilation window (i.e., $k = K - L$), K denotes the current hour, L denotes the size of the assimilation window (hrs), $X_{S,j,i}^{\max}$ denotes the upper bound of $X_{S,j,i}$, $X_{S,j,i,T,K-L}$ denotes the assumed true SAC state (as generated by HL-RDHM) at $k = K - L$, α denotes the random noise of $N\left(-0.5\ln\left(1 + (C_S X_{S,j,i}^{\max})^2\right), \ln\left(1 + \left(C_S X_{S,j,i}^{\max}\right)^2\right)\right)$ where C_S specifies the amount of perturbation (see Table 1). Note that exp (α) in Eq. (A.3) follows $LN\left(1, \left(C_S X_{S,j,i}^{\max}\right)^2\right)$ where LN ( ) denotes the lognormal distribution. In Eq. (A.4), β denotes the spatially-correlated random noise of $N(0,1)$. We used the Turning Bands Method (TBM, [61]) with the exponential correlation function and a decorrelation length of 10 HRAP cells to generate traces of β. With the above, $X_{S,j,i,K-L}$ in Eq. (A.3) follows $Ln\left(X_{S,j,i,T,K-L}, C_S^2\left(X_{S,j,i}^{\max}\right)^4\right)$. Eqs. (A.3) and (A.4) were used to generate the synthetic observations of the SAC states at the beginning of the assimilation window.

Appendix B. Observation model for in situ soil moisture and estimation of its observation error variance

To assimilate the in situ soil moisture observations at Westville in the Oklahoma Mesonet into the gridded SAC, we need an observation equation that relates the point-scale in situ observations at a single location to the grid-scale model soil moisture states at all cells. Recall in the DA formulation that the SAC states are adjusted at each cell whereas the soil moisture observations are available only within a single cell, and that only the vertical soil moisture dynamics is modeled in SAC. Note also that variational assimilation does not model propagation of the second-order moment of the state variables, and hence does not provide information on the spatial dependence among grid-specific model soil moisture. As such, direct variational assimilation into gridded SAC of the soil moisture observations at Westville would result in adjustment of the SAC states only at that cell. Below, we describe how we formulate Eq. (7) to spread the effects of assimilating soil moisture observations at Westville to all cells. The soil moisture observation equation, Eq. (7), is rewritten as Eq. (B.1):

$$Z_{\theta i,m,k}^P = H_{\theta i,m,k}(\mathbf{X}_{S,k}) + V_{\theta i,m,k}^P \quad \text{for } i = 1,\ldots,n_C;\ m = 1,\ldots,n_D;\ k = K-L+1,\ldots,K, \tag{B.1}$$

where $Z_{\theta i,m,k}^P$, $H_{\theta i,m,k}(\mathbf{X}_{S,k})$ and $V_{\theta i,m,k}^P$ denote the pseudo (see below for explanation) soil moisture observation at the HRAP scale, the observation equation that maps the model soil moisture states $\mathbf{X}_{S,k}$ to in situ soil moisture at the HRAP scale, and the zero-mean observation error associated with $Z_{\theta i,m,k}^P$ at the ith cell, mth depth and k-th hour in the assimilation window, respectively, n_C denotes the number of cells in the basin, n_D denotes the number of depths where the in situ soil moisture observations are available, K denotes the current hour, and L denotes the size of the assimilation window (hrs). The pseudo observation $Z_{\theta i,m,k}^P$ at the HRAP cell containing Westville is obtained by probability-matching the point-scale in situ soil moisture observations with the SAC-simulated soil moisture at the HRAP cell containing Westville (see below for details).

To propagate the pseudo observation $Z_{\theta i,m,k}^P$ at the HRAP cell containing Westville to other cells, one may model the spatial correlation function of the soil moisture observation error in the observation equation and solve the resulting Fisher estimation problem [41]. As shown below, however, the soil moisture observation error is a combination of a number of different types of errors of which we do not know the spatial correlation structures very well. For this reason, we assume in this work that the zero-mean soil moisture observation error is spatially independent so that $E\left[V_{\theta i,m,k}^P V_{\theta j,m,k}^P\right] = 0$ holds for $i \neq j$. We then assign $Z_{\theta i,m,k}^P$ at the cell containing Westville to all other HRAP cells as pseudo observations, and prescribe the error variances associated with them in such a way that they reflect the spatial variability of soil moisture. Such an approach has also been used in rainfall estimation [62]. Below, we describe how the variance of $V_{\theta i,m,k}^P$, or $\sigma_{\theta,i,m,k}^2$, is estimated.

To estimate $\sigma_{\theta,i,m,k}^2$ for all HRAP bins in the basin we identified five different types of uncertainty in the in situ soil moisture observations from the Oklahoma Mesonet in reference to the true soil moisture at the HRAP scale at Westville: (1) uncertainty in estimating soil moisture, θ, from temperature change, ΔT, due to the limited accuracy in measuring ΔT and in converting ΔT to θ, (2) uncertainty due to the limited numerical precision in the soil moisture data, (3) uncertainty due to point-vs.-HRAP scale difference

Table B.1
Error standard deviations of the pseudo soil moisture observations at the HRAP scale and their estimated values (see Appendix B for details). In the table, θ denotes in situ soil moisture (m^3/m^3) estimated by the CSI 229-L soil moisture sensor at Westville.

Error type	Standard deviation (m^3/m^3)	
	25 cm-Depth soil moisture	60 cm-Depth soil moisture
ΔT-to-θ conversion error $0.25 \leqslant \theta \leqslant 0.42$ at 25 cm depth $0.21 \leqslant \theta \leqslant 0.36$ at 60 cm depth	0.008	0.007
Data numerical precision error	0.010	0.010
Data representativeness error due to scale difference between point and HRAP grid	0.033	0.037
Spatial variability error due to inter-grid variability of soil moisture at HRAP scale	0.008–0.083	0.006–0.075
Total error $0.25 \leqslant \theta \leqslant 0.42$ at 25 cm depth $0.21 \leqslant \theta \leqslant 0.36$ at 60 cm depth	0.036–0.090	0.043–0.086

(i.e., microscale variability of soil moisture), (4) uncertainty due to spatial variability of soil moisture at the HRAP scale within the basin, and (5) uncertainty due to the limited dynamic range of the sensor. Below we describe how these uncertainties are estimated or accounted for in the assimilation process.

The CSI 229-L soil moisture sensor used in the Oklahoma Mesonet measures the temperature change in the thermocouple [63]. The measured temperature difference, ΔT, is used to estimate the matric potential of the soil via the van Genuchten equation [64,65], which is then converted to soil moisture, θ, via the van Genuchten equation [63,65]. According to Flint et al. [64], over the range of -0.01 to -35 MPa, the mean percentage error of matric potential estimated from the measurements of ΔT against the measured matric potential is about 22.8%. To estimate the variance of the error associated with converting ΔT to θ, we designed a small numerical experiment to generate via the van Genuchten equation soil moisture estimates with 22.8% of the assumed true soil moisture as the error standard deviation given different values of matric potential within the range observed in the Westville data. We then calculated the sample variance of the difference between the above-generated soil moisture estimates and the assumed true soil moisture. The sample variance then represents that uncertainty associated with the ΔT-to-θ conversion. Note that the above procedure implicitly assumes that the matric potential-to-soil moisture relationship is known perfectly. This assumption is reasonable because we are only interested in the uncertainty spread of the conversion rather than the absolute accuracy of the conversion itself.

For the data precision error, the numerical precision of the Westville data, 0.01 m^3/m^3, was assumed as its standard deviation.

To reconcile the difference in spatial scale between the soil moisture observations at point scale and the modeled soil moisture at the HRAP scale [58,59], we used probability matching as described below. We used the in situ soil moisture observations at Westville and the HL-RDHM simulations of soil moisture for the period of 1997–1999 to derive the empirical cumulative distribution functions (CDF) at each depth. The in situ soil moisture data were then converted to pseudo soil moisture observations at the HRAP scale in the model soil moisture space via probability matching. For the period of 2000–2002, the pseudo soil moisture observations at the HRAP scale in the model space obtained from the in situ soil moisture data via probability matching were compared to the model soil moisture. The difference between the two may be attributed to the point-vs.-HRAP scale difference and the sampling variations between the two periods. Assuming that the latter is relatively small, the sample variance of the above difference may be considered representative of the uncertainty due to the representativeness error.

For the uncertainty due to spatial variability of soil moisture at the HRAP scale, we compared the model-simulated soil moisture at the HRAP cell containing Westville to that at each of the other HRAP cells in the basin for the period of 1997–2002. We then considered the sample variance of the inter-grid difference to represent the added uncertainty in the pseudo soil moisture observation due to spatial (i.e., cell-to-cell) variability of cell-averaged soil moisture.

The dynamic range of soil moisture observation is estimated directly from the Oklahoma Mesonet data. At Westville, the observations are bounded by wilting point and field capacity of 0.25 and 0.42 m^3/m^3 at 25 cm and 0.21 and 0.36 m^3/m^3 at 60 cm, respectively. In the data assimilation process for the real-world experiment, whenever the observed soil moisture hit either bound, we prescribed a very large error variance to the observation to reflect that the actual soil moisture may be greater or smaller than the upper or the lower bound, respectively. In the synthetic experiment, on the other hand, no such bounds were assumed and the assumed "true" soil moisture observation at the HRAP scale (see Appendix A) ranged from residual soil moisture content to porosity.

Table B.1 summarizes the individual error standard deviations for the first four uncertainty sources contributing to the total error standard deviation associated with the pseudo soil moisture observations at the HRAP scale at both depths, i.e., $Z^P_{\theta i,m,k}$ in Eq. (B.1). Assuming independence among the different errors considered, we obtain the total error variance for $V^P_{\theta i,m,k}$ in Eq. (B.1) by summing all contributing error variances. The lower bounds of the total error standard deviation in Table B.1 correspond to that for $V^P_{\theta i,m,k}$ at the HRAP cell containing Westville. It is worth nothing that they are close to the maximum error bound of 0.05 cm^3/cm^3 reported by Illston et al. [63] for soil moisture estimates from the Oklahoma Mesonet. In the synthetic experiment, the first three errors in Table B.1 and the cell-to-cell variability of soil moisture at the HRAP scale are accounted for via Eq. (A1) and via the observation equation as described above, respectively. The first three errors in Table B.1 amount collectively to an error standard deviation of 0.036 m^3/m^3 at 25 cm depth and 0.043 m^3/m^3 at 60 cm depth. Accordingly, the low level of uncertainty of σ_θ of 0.01 (see Table 1) represents an idealized situation where the soil moisture observations are free from errors due to ΔT-to-θ conversion or microscale variability. The medium level of uncertainty of σ_θ of 0.03, on the other hand, represents an uncertainty level close to the collective uncertainty due to the first three errors in Table B.1. It is reminded,

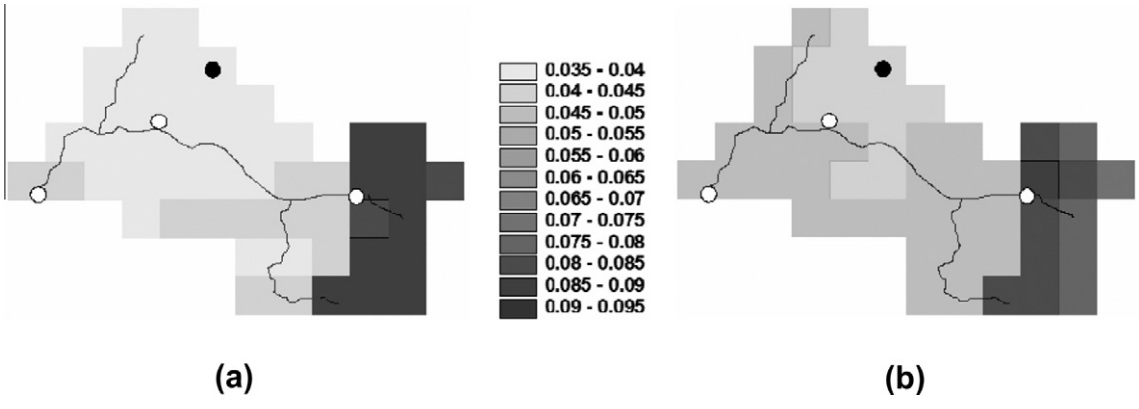

Fig. B.1. The maps of the error standard deviation (m^3/m^3) of the pseudo soil moisture observations at the HRAP-grid scale for (a) 25 cm- and (b) 60 cm-depth soil moisture. The pseudo soil moisture observations are probability-matched in situ soil moisture data at Westville (see Appendix B for details). The white and black circles denote the stream gauge and soil moisture measurement locations, respectively, and the solid line depicts the channel network.

however, here that, in the synthetic experiment, the dynamic range of soil moisture observation is assumed to be unlimited whereas, in reality, that of the Oklahoma Mesonet soil moisture observations is bounded by wilting point and field capacity. Hence, the medium level of uncertainty in Table 1 is likely to be an underestimate.

In the real-world experiment, the total error standard deviation estimates in Table B.1 were used. Fig. B.1 shows the maps of $\left(V^{P}_{\theta i,m,k}\right)^{1/2}$ used in the real-world experiment (see Section 7) as estimated from the above steps.

References

[1] Carpenter TM, Georgakakos KP, Sperflage JA. On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use. J Hydrol 2001;253:169–93.

[2] Koren V, Reed S, Smith M, Zhang Z, Seo D-J. Hydrology Laboratory Research Modeling System (HL-RMS) of the US National Weather Service. J Hydrol 2004;291:297–318.

[3] Reed S, Koren V, Smith M, Zhang Z, Moreda F, Seo D-J. DMIP participants overall Distributed Model Intercomparison Project results. J Hydrol 2004;298:27–60.

[4] Georgakakos KP, Rajaram H, Li SG. On improved operational hydrologic forecasting of streamflows. Report 325. Iowa City, IA: Iowa Institute of Hydraulic Research; 1988. p. 162.

[5] Seo D-J, Koren V, Cajina N. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. J Hydrometeorol 2003;4:627–41.

[6] Seo D-J, Cajina L, Corby R, Howieson T. Automatic state updating for operational streamflow forecasting via variational data assimilation. J Hydrol 2009;367:255–75.

[7] Koren VI, Finnerty BD, Schaake JC, Smith MB, Seo D-J, Duan Q-Y. Scale dependencies of hydrologic models to spatial variability of precipitation. J Hydrol 1999;217:285–302.

[8] Winchell M, Gupta HV, Sorooshian S. On the simulation of infiltration- and saturation excess runoff using radar-based rainfall estimates: effects of algorithm uncertainty and pixel aggregation. Water Resour Res 1998;34(10):2655–70.

[9] Liu Y, Gupta HV. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour Res 2007;43:W07401. doi:10.1029/2006WR005756.

[10] De Lannoy GJM, Houser PR, Pauwels VRN, Verhoest NEC. State and bias estimation for soil moisture profiles by an ensemble Kalman filter: effect of assimilation depth and frequency. Water Resour Res 2007;43:W06401. doi:10.1029/2006WR005100.

[11] De Lannoy GJM, Reichle RH, Houser PR, Pauwels VRN, Verhoest NEC. Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resour Res 2007;43:W09410. doi:10.1029/2006WR005449.

[12] Reichle RH, McLaughlin DB, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. Mon Weather Rev 2002;130:103–14.

[13] Clark MP, Slater AG, Barrett AP, Hay LE, McCabe GJ, Rajagopalan B, et al. Assimilation of snow covered area information into hydrologic and land-surface models. Adv Water Resour 2006;29:1209–21.

[14] Pauwels VRN, De Lannoy GJM. Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge. J Hydrometeorol 2006;7(3):458–77.

[15] Vrugt JA, Diks CGH, Gupta HV, Bouten W, Verstaten JM. Improved treatment of uncertainty in hydrologic modelling: combining the strengths of global optimization and data assimilation. Water Resour Res 2005;41:W01017. doi:10.1029/2004WR003059.

[16] Weerts AH, Serafy YH. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall–runoff models. Water Resour Res 2006;42:W09403. doi:10.1029/2005WR004093.

[17] Clark MP, Rupp DE, Woods RA, Zhend X, Ibbitt RP, Slater AG, et al. Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. Adv Water Resour 2008;31:1309–24.

[18] Ibbitt RP, Clark MP, Woods RA, Zheng X, Slater AG, Rupp DE, et al. Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. In: AGU fall meeting, San Francisco; 2007.

[19] Seo D-J, Koren V, Cajina N. Real-time assimilation of radar-based precipitation data and streamflow observations into a distributed hydrological model. In: Proceedings of symposium HS03 held during IUGG2003 at Sapporo, July 2003. Weather radar information and distributed hydrological modelling, vol. 282. IAHS Publisher; 2003. p. 138–42.

[20] Brocca L, Melone F, Moramarco T, Wagner W, Naeimi V, Bartalis Z, et al. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol Earth Syst Sci 2010;14:1881–93.

[21] Francois C, Quesney A, Ottlé C. Sequential assimilation of *ERS-1* SAR data into a coupled land surface-hydrological model using an extended Kalman filter. J Hydrometeorol 2003;4:473–87.

[22] Komma J, Bloschl G, Reszler C. Soil moisture updating by ensemble Kalman filtering in real-time flood forecasting. J Hydrol 2008;357:228–42.

[23] Moradkhani H, Sorooshian S, Gupta HV, Houser PR. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 2005;28:135–47.

[24] Aubert D, Loumagne C, Oudin L. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. J Hydrol 2003;280:145–61.

[25] Burnash RJ, Ferral RL, McGuire RA. A generalized streamflow simulation system: conceptual modeling for digital computers. US Department of Commerce National Weather Service and State of California Department of Water Resources; 1973.

[26] Anderson EA. Continuous API Model. National weather service river forecasting system users manual. <http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/htm/xrfsdocwpd.php>; 1994.

[27] Anderson EA. National weather service river forecast system-snow accumulation and ablation model. Technical Memo. Silver Spring, MD: NOAA; 1973. p. 217.

[28] Greene DR, Hudlow MD. Hydrometeorologic grid mapping procedures. In: AWRA international symposium on hydrometeorology, Denver, CO; 1982.

[29] Reed SM, Maidment DR. Coordinate transformations for using NEXRAD data in GIS based hydrologic modelling. J Hydrol Eng 1999;4:174–83.

[30] Reed SM. Deriving flow directions for coarse-resolution (1–4 km) gridded hydrologic modelling. Water Resour Res 2003;39(9):1238. doi:10.1029/2003WR001989.

[31] Fulton RA, Breidenbach JP, Seo D-J, Miller DA. WSR-88D rainfall algorithm. Weather Forecast 1998;13:377–95.

[32] Seo D-J. Real-time estimation of rainfall fields using radar rainfall and rain gauge data. J Hydrol 1998;208:37–52.

[33] Seo D-J, Breidenbach JP, Johnson ER. Real-time estimation of mean field bias in radar rainfall data. J Hydrol 1999;223:131–47.

[34] Young CB, Bradley AA, Krajewski WF, Kruger A. Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic forecasting. J Hydrometeorol 2000;1:241–54.

[35] Koren V, Smith M, Wang D, Zhang Z. Use of soil property data in the derivation of conceptual rainfall-runoff model parameters. In: Proceedings of the 15th conference on hydrology. Long Beach, CA: AMS; 2000. p. 103–6.

[36] Natural Resources Conservation Service, United States Department of Agriculture, US General Soil Map (STATSGO2). Available at: <http://soildatamart.nrcs.usda.gov>; 2006.

[37] Natural Resources Conservation Service, United States Department of Agriculture, Soil Survey Geographic (SSURGO) Database. Available at: <http://soildatamart.nrcs.usda.gov>; 2004.

[38] Smith MB, Seo D-J, Koren VI, Reed SM, Zhang Z, Duan Q, et al. The Distributed Model Intercomparison Project (DMIP): motivation and experiment design. J Hydrol 2004;298:4–26.

[39] Koren V. Parameterization of frozen ground effects: sensitivity to soil properties. In: Proceedings of symposium S7 at Foz do Iguacu, Brazil, April 2005. Prediction in ungauged basins: promise and progress, vol. 303. IAHS Publisher; 2006. p. 125–33.

[40] Koren V, Fekadu M, Reed S, Smith M, Zhang Z. Evaluation of grid-based distributed hydrological model over a large area. In: Proceedings of symposium S7 at Foz do Iguacu, Brazil, April 2005. Prediction in ungauged basins: promise and progress, vol. 303. IAHS Publisher; 2006. p. 47–56.

[41] Schweppe FC. Uncertain dynamic systems. Prentice-Hall; 1973. p. 563.

[42] Zupanski D. A general weak constraint applicable to operational 4DVAR data assimilation systems. Mon Weather Rev 1997;125:2274–92.

[43] Li Z, Navon IM, Zhu Y. Performance of 4D-Var with different strategies for the use of adjoint physics with the FSU global spectral model. Mon Weather Rev 2000;128:668–88.

[44] Jazwinski AH. Stochastic processes and filtering theory. Academic Press; 1970.

[45] Lewis JM, Lakshmivarahan S, Dhall SK. Dynamic data assimilation: a least squares approach. Cambridge University Press; 2006.

[46] Zhang S, Zou X, Ahlquist JE. Examination of numerical results from tangent linear and adjoint of discontinuous nonlinear models. Mon Weather Rev 2001;129:2791–804.

[47] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 1994;99(10):143-10,162.

[48] Gordon NJ, Salmond DJ, Smith AFM. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc F 1993;140:107–13.

[49] Pham DT. Stochastic methods for sequential data assimilation in strongly nonlinear systems. Mon Weather Rev 2001;129:1194–207.

[50] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in fortran. 2nd ed. Cambridge University Press; 1992.

[51] Brock FV, Crawford KC, Elliot RL, Cuperus GW, Stadler SJ, Johnson HL, et al. The Oklahoma Mesonet: a technical overview. J Atmos Ocean Technol 1995;12:5–19.

[52] Illston BG, Caldwell JC, Bodnar SG. Representativeness of soil moisture conditions in central Oklahoma during the enhanced drying phase. In: Proceedings of 15th symposium on global change and climate variations. Seattle, Washington: AMS; 2004.

[53] Hersbach H. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast 2000;15:559–70.
[54] Jolliffe IT, Stephenson DB. Forecast verification: a practitioner's guide in atmospheric science. Chichester: John Wiley and Sons; 2003.
[55] Wilks DS. Statistical methods in the atmospheric science. Academic Press; 1995.
[56] Xie X, Zhang D. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Adv Water Resour 2010;33:678–90.
[57] Zehe E, Blöschl G. Predictability of hydrologic response at the plot and catchment scales: role of initial conditions. Water Resour Res 2004;40:W10202. doi:10.1029/2003WR002869.
[58] Choi M, Jacobs JM. Temporal variability corrections for Advanced Microwave Scanning Radiometer E (AMSR-E) surface soil moisture: case study in Little River region, Georgia, US. Sensors 2008;8:2617–27.
[59] Reichle RH, Koster RD. Bias reduction in short records of satellite soil moisture. Geophys Res Lett 2004;31:L19501. doi:10.1029/2004GL020938.
[60] Parajka J, Naeimi V, Blöschl G, Komma J. Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria. Hydrol Earth Syst Sci 2009;13:259–71.
[61] Mantoglou A, Wilson JL. The turning bands methods for simulation of random fields using line generation by a spectral method. Water Resour Res 1982;18(5):1379–94.
[62] Seo D-J, Breidenbach JP. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. J Hydrometeorol 2002;3:93–111.
[63] Illston BG, Basara JB, Fisher DK, Elliott R, Fiebrich CA, Crawford KC, et al. Mesoscale monitoring of soil moisture across a statewide network. J Atmos Ocean Technol 2008;25:167–82.
[64] Flint AL, Campbell GS, Ellett KM, Calissendorff C. Calibration and temperature correction of heat dissipation matric potential sensors. Soil Sci Soc Am J 2002;66:1439–45.
[65] van Genuchten MTh. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 1980;44:892–8.