R Visweshwaran[1], RAAJ Ramsankaran[1*], T I Eldho[1], S Lakshmivarahan[1†, 2]

[1]Department of Civil Engineering, Indian Institute of Technology, Bombay, Mumbai, India

[2] School of Computer Sciences, University of Oklahoma, Norman, OK, 73072, USA

[†] - Visiting Professor at Indian Institute of Technology, Bombay, Mumbai, India

[*]- Correspondence: RAAJ Ramsankaran; Email id: ramsankaran@civil.iitb.ac.in

Key Points:

- The potential of assimilating only sensitive soil moisture observations on streamflow forecast was demonstrated using a novel FSM method
- Impact of temporally evolving sensitivity of initial condition on hydrological model performance during assimilation was determined
- The majority of the estimated sensitive soil moisture observations were predominantly from the monsoon season
- The sensitivity based soil moisture assimilation was effective up to 45 lead days during streamflow forecast

Abstract

The need for and the use of different data assimilation techniques to improve the quality of streamflow forecast is now well established. In this paper, the goal is to demonstrate the power of a new class of methods known as the Forward Sensitivity Method (FSM) which is based on the temporal evolution of model sensitivities with respect to the control variables consisting of initial conditions and parameters. FSM operates in two phases: The first phase provides a simple algorithm for placing observations at or near where the square of forward sensitivities attains their maximum values. Using only this selected subset of observations in a weighted least squares method, the second phase then provides an estimate of the unknown elements of the control variables. In this paper, FSM based assimilation is applied to a simple class of two parameter model in a medium-sized agriculture dominant watershed lying in the Krishna River Basin, India. Four assimilation scenarios were tested to determine the effect of assimilating only sensitive observations as well as the impact of temporally evolving initial condition sensitivity. Sensitivity results showed that observations during the monsoon time alone are enough for assimilation purposes, which has helped in reducing the computational time greatly. Assimilation and forecast results also indicated that the scenarios which assimilated only sensitive observations are better in estimating daily streamflow. From the obtained results, it is concluded that FSM based assimilation has significant potential to improve the streamflow simulations, especially in places where data availability remains a major challenge.

1 Introduction

Within the parlance of Hydrology, forecasting streamflow plays a vital role in water resource management. Based on the temporal scale, streamflow forecasting can be broadly classified into two categories: short-term and long-term forecasts. Short-term forecasting at hourly and daily scale is critical for flood warning, mitigation (Gül et al., 2010; Rogelis & Werner, 2018), and disaster management (Roulin, 2007). Long-term forecasting at monthly and seasonal scales is extensively utilized in reservoir monitoring (Rezaie-Balf et al., 2019), design of hydraulic structures (Lowe, 2006), and irrigation practices (Droogers & Bastiaanssen, 2002). However, uncertainty in hydrological predictions is still a serious concern that needs to be addressed before planning and management activities. These uncertainties may arise due to the error in the forcing variables (e.g., rainfall and temperature data), the model's inherent structural error (parameters and boundary conditions), and improper initial condition (Alvarado-Montero et al., 2017). While the quality of the input data such as the rainfall rate and temperature can be improved by deploying better observation systems, errors in model control variables consisting of the initial and boundary conditions and parameters can be corrected using the well-established tools from the theory of dynamic data assimilation (DDA) (Lewis et al., 2006; Lakshmivarahan et al., 2017). In this paper, the goal is to demonstrate the power of a new class of method called forward sensitivity method (FSM) (Lakshmivarahan & Lewis, 2010) for assimilating data into a simple conceptual two parameter model (TPM) (Xiong & Guo, 1999) in the analysis of streamflow.

DDA is the process of combining a model of a process of interest in the analysis with a finite set of relevant but noisy observations of the same process. Existing literature on DDA can be broadly classified into three classes. First is a class of sequential methods known as Kalman filtering (Kalman, 1960) and many if its extensions (Puente & Bras 1987; Evensen, 1994; Whitaker & Hamill, 2002; Sakov et al., 2012). This class of methods rests on the basic principle of best, linear unbiased estimation (BLUE), where best is in the sense minimum variance. While Kalman filter-based methods (KF) provide a natural framework for sequential dynamic data assimilation, their excessive computational requirements essentially prevented the use of this class of methods to large scale problems by operational centers around the globe. To mitigate the excessive computational demands of the classical Kalman filtering, Evensen in 1994 introduced an ensemble version of Kalman filtering (also known as EnKF). While the quality of this class of Monte-Carlo methods critically depends on the number of ensemble members, its ease of implementation combined with the fact that it does not require the development of adjoint dynamics has been responsible for its widespread use by the leading forecast centers across the world (Sabater et al., 2007; Leisenring & Moradkhani, 2011; Zhu et al., 2012; Pathiraja et al., 2016; Patil & Ramsankaran 2018). Although EnKF improved the model performance greatly, it explicitly assumed that the distribution of the ensembles is Gaussian in nature. By doing so it eliminates the importance of extreme events like flood and droughts which is of high importance in hydrological field (Carsten Montzka, 2012). To circumvent this problem, Particle filter (PF), a

non-Gaussian, non-linear filtering approach was introduced to the field of hydrology by Moradkhani et al., (2005). Although PF served as an alternative to EnKF, the computational demand to generate particles is high and its implementation become more difficult when combined with complex models. Refer to chapters 27 to 30 in Lewis et al., (2006) for a detailed analysis on various aspects of analytic approximations including the first-order or extended Kalman filter, second-order filters and ensemble based (also known as reduced rank) filters, and their computational requirements.

The second is a class of variational methods for assimilating noisy observations into the deterministic models and has come to be known as the four-dimensional variational method (4-D VAR) and it was widely used in the hydrological field (Seo et al., 2009; Lee et al., 2012; Cioaca et al., 2013, Alvarado-Montero et al., 2017; Oubanas et al., 2018). This off-line method has deep roots in optimal control theory and was developed by Le Dimet & Talagrand (1986). Conceptually, 4-D VAR is a two-stage process. The first stage computes the so called adjoint gradient of an objective function which is the weighted sum of squared errors between the model forecast and the observations. This is accomplished by developing the adjoint dynamics that run backward in time. Once the gradient is made available, it is then used in one of many known algorithms - gradient, conjugate gradient, or quasi-Newton method, (Refer to Chapter 10-12 in Lewis et al., 2006) to move towards the minimum. These two steps are then repeated until a suitable convergence criterion is satisfied. Refer to chapters 22 to 26 in Lewis et al., (2006) for more details. However, the limitation of these conventional variational approaches lies in the complexity to establish a stable adjoin model. Further, the efficiency and accuracy of the assimilation solely depend on the quality and the number of observations considered during assimilation (Cioaca et al., 2013).

Thanks to the vast improvements in the sensor and communication technologies – radars, satellites, to name a few, it is now possible to sense various field variables with ever increasing spatio-temporal frequency resulting in large volumes of data with a side effect of increased temporal and spatial correlation in the observed data. This implies more data does not mean more information. Further, assimilating large volumes of correlated data requires excessive computing power and time. Therefore, it becomes imperative to assess the efficiency and impact of assimilating only necessary observations during assimilation.

The third class of methods is based on the evolution of sensitivity of the model forecast with respect to the elements of the control variable consisting of initial, boundary conditions, and parameters and has come to be known as the forward sensitivity method (Lakshmivarahan & Lewis 2010; Tromble et al., 2016; Lakshmivarahan et al., 2017). While this method is philosophically aligned with the 4-D VAR, it does not require the backward adjoint model but requires running the sensitivity dynamics simultaneously forward with the model dynamics. This method has twin advantages. While the first two methods - 4-D VAR and ensemble filtering methods, described above are silent on the important question of

placement of observations, FSM based methodology suggests a simple algorithm for placement of observations: place the observations at or near those temporal locations where the square of the forward sensitivity attains their maximum values (Lakshmivarahan et al., 2020). This decision phase does not require the knowledge of observations. Once the number and locations of the observations are known, the data assimilation problem (using only this subset of observations at the chosen locations) can be solved using the generalized or weighted least squares method. Refer to Tromble et al. (2016), Lewis et al. (2016) and Lewis et al. (2020).

Against this backdrop, a review of the literature on the applications of data assimilation in hydrology is provided. Over the last two decades, data assimilation has been performed in various hydrological models ranging from simple conceptual lumped models to complex process-based distributed models. The efficacy of the assimilation in improving the hydrological variables (streamflow, soil moisture, and evapotranspiration) varied considerably from marginal to significant improvement. Particularly, in many studies, the efficiency of assimilating observations into a complex model is not easy and often showed marginal improvement compared to the simple models. For instance, Han et al., (2012) performed soil moisture assimilation in a complex SWAT model and the performance in simulating the streamflow was limited by the rainfall-runoff mechanism where the peak and low flows were not captured due to the incorrect precipitation. Similarly, Pan et al., (2008), assimilated SEBS observations in distributed VIC model. From the result, it was clear that the assimilation performance has optimally improved the evapotranspiration (ET) estimates but failed to update other prognostic variables such as soil moisture and streamflow. On the contrary, Xiong et al., (2019), performed EnKF assimilation in a simple lumped Two Parameter Model and showed significant improvement in the streamflow and ET predictions. Likewise, Leach et al., (2018) performed dual state-parameter assimilation on four models and concluded that the simpler conceptual GR4J-SR model outperformed others in improving the model simulations. Recently, Loizu et al., (2018) compared the assimilation efficiency between a less parameterized simple MISDc model with a physics based TOPLATS model. From the results, it was concluded that the improvement due to surface soil moisture assimilation is similar in both models. Based on these studies, it can be concluded that the simple conceptual models have strong potential to improve the simulation of hydrological variables during assimilation. Further, performing assimilation in a simple model can reduce the computational time greatly without compromising forecast quality.

Among the different variables in the hydrological field, soil moisture performs an important role in the partitioning of precipitation into overland and underground flows. It is now well understood that in performing hydrological simulations, the initial soil wetness condition prior to a rainfall event is highly significant during runoff generation. Because of this reason, soil moisture is considered as one of the important variables by the hydrological community and soil moisture assimilation had gained increasing importance over others (Kumar et

al., 2009; Sahoo et al., 2013; Alvarez-Garreton et al., 2015; Massari et al., 2015; Meng et al., 2017; Naz et al., 2018). Satellite-based soil moisture observations in particular have been extensively used for assimilation purposes (Gevaert et al., 2018; Massari et al., 2018; Patil & Ramsankaran, 2018). Over the years, various satellite soil moisture observations were made available which include Soil Moisture and Ocean Salinity mission (SMOS), Soil Moisture Active Passive (SMAP), Advanced Microwave Scanning Radiometer (AMSRE), Advanced Scatterometer (ASCAT), etc. ASCAT soil moisture which is obtained from an active microwave sensor has high spatio-temporal resolution and better accuracy (low RFI value) (Wagner et al., 2013). Also, ASCAT level 2 product was the first to deliver the data on a near real-time basis (Clement Albergel et al., 2012), which is operational till date.

In view of the limitations identified in sequential and variational assimilation methods, in this study a new assimilation strategy is proposed using the FSM to effectively assimilate only the sensitive soil moisture observations for streamflow predictions. For demonstration, few real case experiments are carried out in a moderately sized agriculture dominant watershed in India using a conceptual TPM hydrological model. The study addresses two main research questions: 1) Can the sensitivity based soil moisture assimilation help in improving hydrological simulations and forecast accuracy? and 2) How does the performance of hydrological predictions vary with different time window frames during variational assimilation? To address these two questions, four scenarios are adopted in this study and the performance of these assimilation scenarios are evaluated using five evaluation criteria namely Kling Gupta Efficiency (KGE), Percentage Bias (PBias), Root Mean Square Error (RMSE), Peak Flow Criteria (PFC), and Low Flow Criteria (LFC). For the forecasting purpose, only the scenarios whose streamflow performance showing KGE more than 0.8 during the assimilation phase are adopted. In the forecasting phase, the performance of FSM is then evaluated with nine different lead times in the future.

The remainder of this article is organized as follows: Section 2 explains the governing equations and derivation of the FSM, Section 3 briefs about the description of the model, data used, and study area adopted in this study. Later, Section 4 describes the methodological framework, different scenarios adopted during assimilation. Section 5 discusses the results of the proposed FSM assimilation during the assimilation phase and forecast phase. Finally, Section 6 summarizes the study and concludes the results with some recommendations for future work.

2 Forward Sensitivity Method: An overview

In this section, an overview of the FSM is provided for a) placement of observations and b) assimilating the observations into the model. For more details of FSM, refer to Lakshmivarahan & Lewis (2010), Lakshmivarahan et al. (2017), and Lakshmivarahan et al. (2020). For a comprehensive introduction to data assimilation refer to Lewis et al., (2006).

Let $y(k)$ R^n denotes a n-dimensional, real vector representing the state of a deterministic, non-linear dynamic model given by

$$y(k+1) = f(y(k), \alpha) \ (1)$$

where $\epsilon \ \mathrm{R}^p$ is a real vector of parameters and $f : \mathbb{R}^n \ \mathrm{X}\mathbb{R}^p \rightarrow \mathbb{R}^n$ defines the vector field of the dynamics, with $y(0)$ $\mathbb{R}^n$ as the initial condition. The solution $y(k) = y\left(k, y_o, \alpha\right)$ depends on y_o and α. Since, y_o and α control the evolution of the solution to Equation 1, together they are called the Control Vector (CV) where $CV = \left({y_o}^T, \alpha^T\right)^T$ $\mathbb{R}^{n+p}$.

It is assumed that this model is faithful to the process being modelled and the control vector is not known in advance. The main goal is to estimate this unknown control vector. To this end, it is assumed that we have access to a finite set of observations, $z(k)$given by

$$z(k) = h\left(\overline{y}(k)\right) + \eta(k) \ (2)$$

where $\overline{y}(k)$ is the unknown true state of the process, and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a function otherwise known as observation operator that relates the model state to the observation. If the states are directly observable, then $h(y) = y$, that is, h is the identity function. $\eta(k)$ in Equation 2 is the unavoidable additive observation noise and it is assumed that $\eta(k)$ is temporally uncorrelated with

$$E\left(\eta(k)\right) = 0, \ cov\left(\eta(k)\right) = R^{\mathrm{m}\ xn}, \text{ and}$$

$$\mathrm{cov}\left(\eta\left(k_1\right), \eta\left(k_2\right)\right) = 0 \ for \ k_1 \neq k_2 \ (3)$$

where the matrix R is known and is symmetric and positive definite. The inverse problem of interest may be stated as follows: given a finite set

$$S = \left\{z\left(k_i\right) \ for \ 1 \leq i \leq N\right\} \ (4)$$

of noisy observation, our goal is to estimate the unknown control vector. While there are number of methods to solve this problem (Lewis et al., 2006), in this work we concentrate on solving this problem using the FSM as described in Lakshmivarahan & Lewis, (2010) and Lakshmivarahan et al. (2017). Define

$$U(k) = \frac{\partial y(k)}{\partial y(0)} = \left[\frac{\partial y_i(k)}{\partial y_j(0)}\right] \quad \mathrm{R}^{\mathrm{n}\ xn} \ (5)$$

is the n x n matrix of forward sensitivities of the solution y(k) at time k with respect to the initial condition $y(0)$. Similarly, let

$$V(k) = \frac{\partial y(k)}{\partial \alpha} = \left[\frac{\partial y_i(k)}{\partial \alpha_j}\right] \quad \mathrm{R}^{\mathrm{n\ x}\ p} \ (6)$$

be n x p matrix of sensitivities of y(k) with respect to α. By direct differentiation of the model Equation 1 with respect to $y(0)$, we get a linear, time-varying difference equation

$$U(k+1) = D_f(k).\mathrm{U}(k) \ (7)$$

where

$$D_f(k) = \frac{\partial f(k)}{\partial y(k)} = \left[\frac{\partial f_i(k)}{\partial y_j(k)}\right] \epsilon \ \mathrm{R}^{\mathrm{m}\ xn} \ (8)$$

is the model Jacobian evaluated along the solution $y(k)$, with $\mathrm{U}(0) = \mathrm{I_n}$, the identity matrix as its initial condition. Likewise, by differentiating the model Equation 1 with respect to the parameter α, we get

$$V(k+1) = D_f(k)V(k) + D_f^{\alpha}(k) \ (9)$$

where

$$D_f^{\alpha}(k) = \frac{\partial f(k)}{\partial \alpha} = \left[\frac{\partial f_i(k)}{\partial \alpha_j}\right] \epsilon \ \mathrm{R}^{\mathrm{m}\ xp} \ (10)$$

is the model Jacobian with respect to the parameter α and $\mathrm{V}(0) = 0$, the zero matrix as the initial condition. The first step is to solve the model Equation 1 and sensitivity Equations 7 and 9 simultaneously so that we have $y(k)$, $U(k)$, and $V(k)$ for k$\geq$0.

2.1 Placement of observations

From the definition of $U(k)$ it follows that the $\mathrm{j^{th}}$ column of $U_j(k)$ is the sensitivity of the solution vector $y(k)$ with respect to the $\mathrm{j^{th}}$ component $y_j(0)$ of the initial condition. Likewise,$V_j(k)$ is the $\mathrm{j^{th}}$ column of $V(k)$ denotes the sensitivity of the solution vector $y(k)$ with respect to the $\mathrm{j^{th}}$ component α_j of the parameter vector .

Following Lakshmivarahan et al. (2020), for placing the observations, we suggest a couple of strategies. The first approach is to plot the square of the norm of each column of the U and V matrices separately and locate the observations where these are maximum. We plot the square since the sensitivity can be both positive and negative, but its square is always positive. The second way would be to plot the sum of the squares of the elements of the matrices U in one plot and the sum of the squares of the elements of V in another plot and place the observations where these plots achieve the maximum value. These strategies ensure that the observability Gramian is positive definite there by ensuring a unique local minimum (Lakshmivarahan et al. 2020). In this study, the second strategy is adopted where CV $\mathbb{R}^{n+p}$ has n+p unknown components and we can place the (n + p) observations when the sum of the square of the norm of the n column vector, $U_j(k)$ of $U(k)$ and p column vector,$V_j(k)$ of $V(k)$ attain their maximum value when added together. Once the placement problem is solved, we can move on to assimilate these (n + p) observations into the model using FSM.

2.2 FSM method for data assimilation

Let $\overline{\mathrm{CV}}$ be the unknown true control and let CV be the current operating point from which model forecast is generated. Thus

$$\delta CV = \ \overline{\mathrm{CV}} - CV = (\ \mathrm{y}_o{}^T, \ \ {}^T)^T \ \ \mathbb{R}^{n+p} \ (11)$$

be the error in the control. Recall that our goal is to estimate this error CV. This initial control error induces an error $\delta y(k)$ in the solution $y(k)$ given by

$$\delta y(k) = \overline{y}(k) - y(k) \quad (12)$$

Using the definition of $U(k)$ and $V(k)$, from first principles(Lakshmivarahan & Lewis, 2010) it follows that

$$\delta y(k) = U(k)\delta y_0 + V(k)\delta\alpha \quad (13)$$

Now consider the difference between an observation $z(k)$ and it's model counterpart

$$e(k) = z(k) - h\left(y(k)\right) = h\left(\overline{y}(k)\right) - h\left(y(k)\right) + \eta(k) \quad (14)$$

called the innovation or forecast error. Expressing $\overline{y}(k) = y(k) + \delta y(k)$ using Equation 12 and expanding $h\left(\overline{y}(k)\right)$ in the first-order Taylor series, we get

$$h\left(\overline{y}(k)\right) = h\left(y(k)\right) + D_y(h)\delta y(k) \quad (15)$$

where

$$D_h(k) = \frac{\partial h(y(k))}{\partial y(k)} = \left[\frac{\partial h_i(y(k))}{\partial y_j(k)}\right] \epsilon \ \mathrm{R}^{\mathrm{m}\ xn} \quad (16)$$

is the Jacobian of h. Now substituting Equation 15 in 14 and simplifying, it follows

$$e(k) = D_h(k)\left[U(k)\delta y_0 + V(k)\delta\alpha\right] \quad (17)$$

Now substituting Equation 13 in 17, we get after simplifying

$$e(k) = H(k) \ \ \mathrm{CV} \quad (18)$$

where

$$H(k) = \begin{bmatrix} H_1(k) & H_2(k) \end{bmatrix} \epsilon \ \mathrm{R}^{\mathrm{m}\ xn+p}$$

$$H_1(k) = D_h(k)\ U(k) \epsilon \ \mathrm{R}^{\mathrm{m}\ xn}$$

$H_2(k) = D_h(k)\ V(k) \epsilon \ \mathrm{R}^{\mathrm{m\ x\ }p}$ and

$$\delta CV = \begin{bmatrix} \delta y_0 \\ \delta\alpha \end{bmatrix} \quad (19)$$

Recall that there are N $\geq$ n + p observations at times $k_1,\ k_2,\ ..\ ,\ k_N$ and the structure of Equation 18 is additive in k. Consequently, combining the forecast errors at these N observations times, we get a linear relation connecting the sum of forecast errors and the unknown δe as

$$E = \mathrm{H} \ \ \mathrm{CV} \quad (20)$$

where $E = \sum_k e(k) \epsilon \ \mathrm{R}^m$ and $H = \sum_k H(k) \epsilon \ \mathrm{R}^{\mathrm{m}\ xn+p}$. We can readily solve Equation 20 using the standard weighted least square method refer to chapter 5 of Lewis et al., (2006) to obtain.

$$\delta \mathrm{CV}_{\mathrm{LS}} = \left(H^T R^{-1} H\right) H^T R^{-1} E \quad (21)$$

3 Study area and data used

3.1 Study area

Marol watershed of Krishna river basin, India was selected as the study area. The watershed encompasses a total area of 5092 km^2 up to the Marol gauging station that extends from 74° 48' E to 75° 36' E and 14° 5' N to 15° 7' N. Figure 1 shows the geographic location of the study site and ASCAT soil moisture observation grids. The major river flowing here is Varada and it has a total length of 188 km. The terrain is moderately flat and lies on the leeward side of Western Ghats. The Western Ghats is a long mountain range, which extends across the southern peninsula of India. The monsoon (rainfall) season extends from the first week of June till November and the catchment receives an average rainfall of over 1500 mm annually. Agriculture covers more than 75% of land use followed by deciduous forest (Patil & Ramsankaran, 2017). According to FAO, soil texture is mostly dominated by clay loam and sandy clay loam

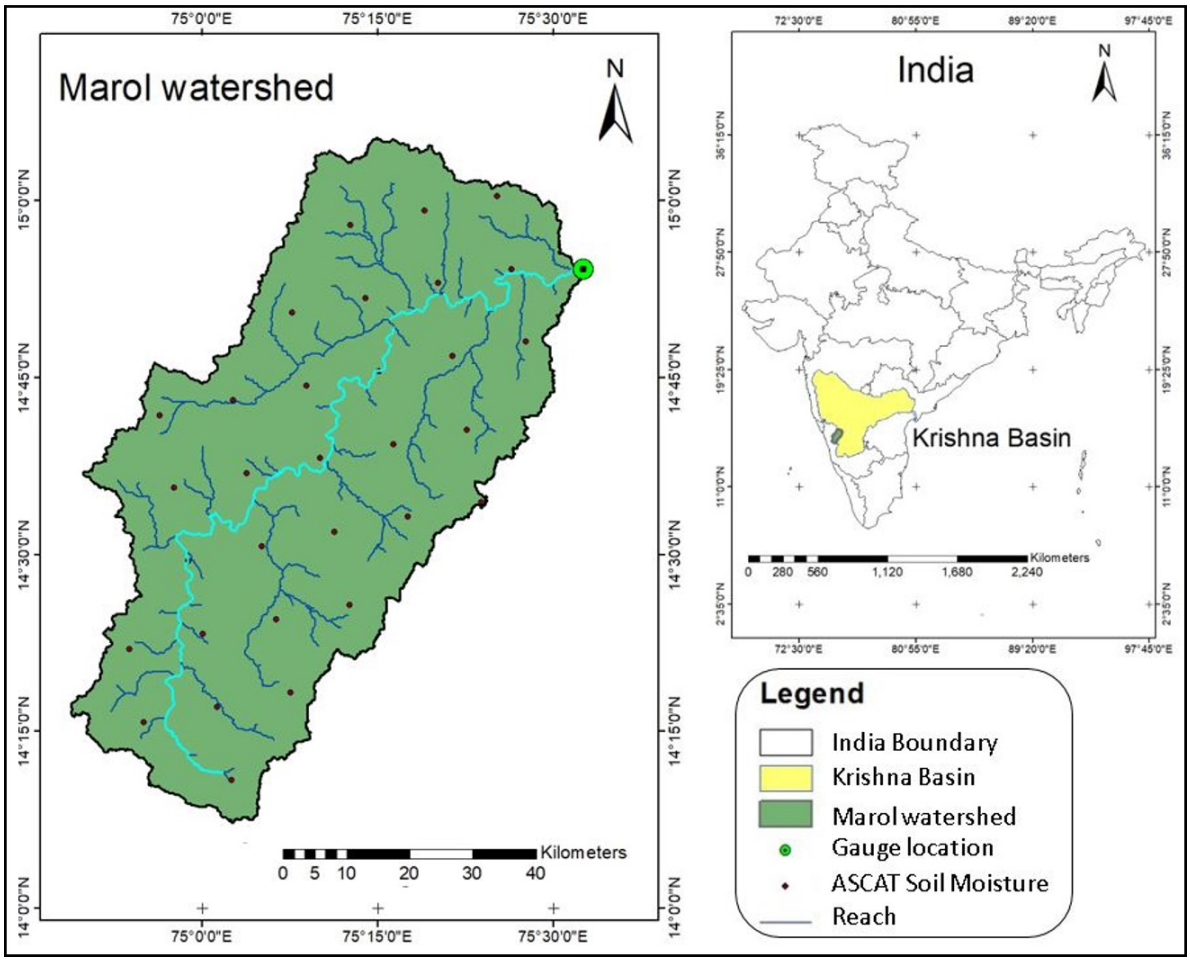

Figure 1: Study area map showing the location of ASCAT surface soil moisture observation points, stream gauging station, and stream networks of the Marol watershed.

3.2 Data used

Details of the input forcing data, soil moisture observations, and streamflow gauge data used for model simulations, assimilation and validation are given in Table 1. Among the input forcing data, humidity, wind speed and solar radiation for estimation of potential evapotranspiration were obtained from NCEP-Climate Forecast System Reanalysis data (Dile & Srinivasan, 2014; Fuka et al., 2014; Saha et al., 2010). Minimum and maximum temperature at $1^\circ \times 1^\circ$ spatial resolution were obtained from India Metrological Department (IMD) (Pai et al., 2014). Daily rainfall data was obtained from Karnataka State Natural Disaster Monitoring Centre (KSNDMC) at the taluk (local administrative unit)

level. A total of 11 years of data from January 2000 till December 2011 were used for the study. Data from the beginning of the year 2000 up to May 2007 was used for training the model except the year 2005. It shall be noted that the year 2005 was not considered because of the poor quality of the dataset.

For assimilation purposes, ASCAT level 2 surface soil moisture data was procured for the period June 2007 till December 2011. This data was obtained from EUMETSAT mounted on the Metop platform. For determining the observation error rate, 'sm_noise' flag data was used in the study. For validation of the model during open-loop and assimilation simulations, Central Water Commission (CWC), Government of India, gauge streamflow data at Marol gauging station was used. It shall be noted that all the data were obtained on a daily scale.

Table 1: Description of the dataset used for the current study.

Dataset	**Purpose**	**Spatial Resolution**	**Unit**
Humidity, Wind speed and Solar radiation	Model forcing	$^\circ \times 0.25^\circ$	(-), (m/sec) and ($^\circ$c)
Temperature		$^\circ \times 1^\circ$	($^\circ$c)
Rainfall		Point	(mm/day)
Surface soil moisture	Assimilation	$^\circ$	(relative percentage)
Streamflow	Validation	Point	(cumecs)

4 Methodology

The following section gives a brief description of the model, steps involved in the preparation of profile soil moisture observation, and sensitivity based assimilation runs. An overall schematic representation of the FSM based assimilation in the TPM is presented in Figure 2. The methodology is divided into four stages. First stage deals with the preparation of profile level soil moisture data that can be directly used for assimilation purposes. The second stage details the estimation of model sensitivities and the identification of sensitive observations for assimilation purpose. The third stage involves the actual assimilation of soil moisture into the model by correcting the control variables. Finally, the last stage deals with the estimation and forecasting of profile soil moisture and streamflow for different lead times.

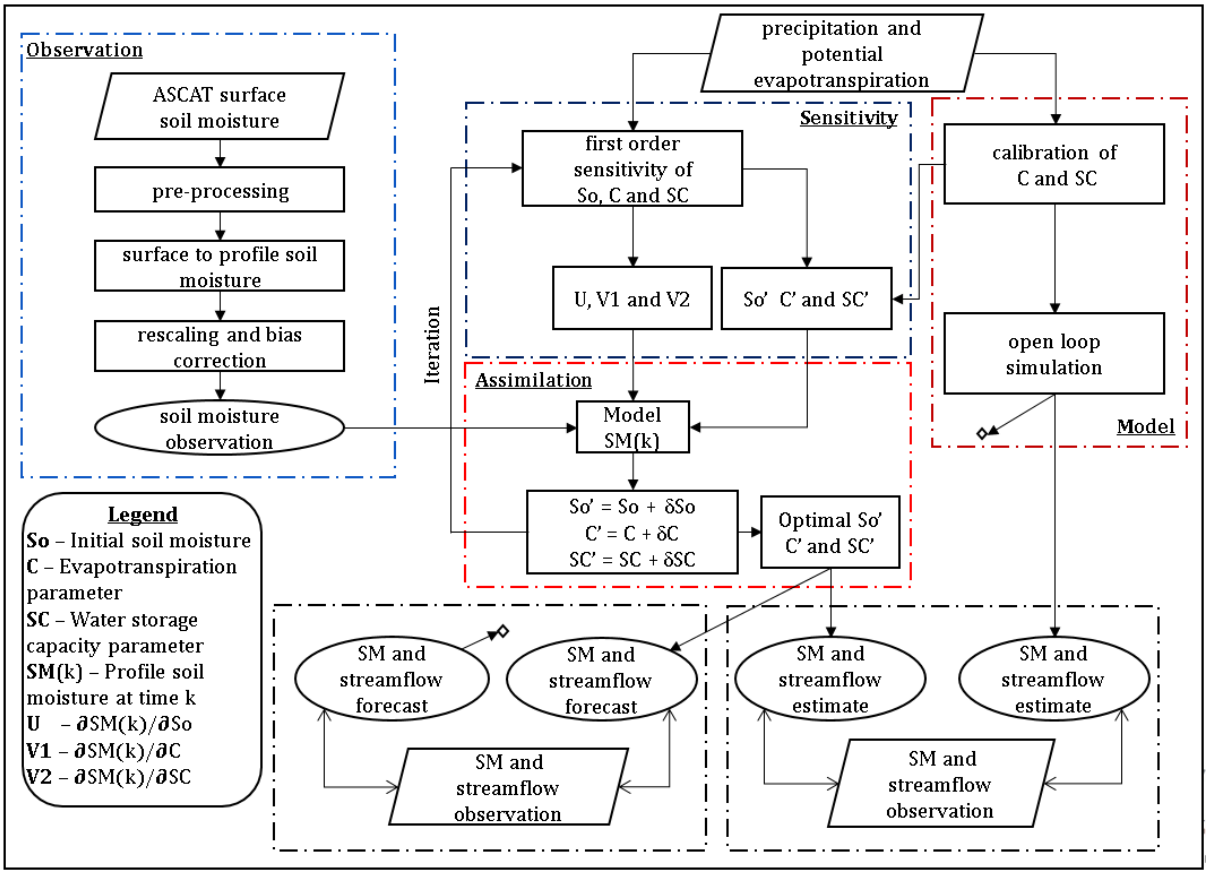

Figure 2: Schematic of the overall FSM assimilation framework used for streamflow and profile soil moisture forecasting

4.1 Two Parameter Model

TPM is a conceptual, lumped, water balance model. It was originally developed by Xiong & Guo, (1999). The two main inputs for running the model are Precipitation (P) and Potential Evapotranspiration (PET). The model estimate Actual Evapotranspiration (AET) based on the ratio of P and PET as shown in Equation 22

$$\mathrm{AET}(k) = C \ . \ \mathrm{PET}(k) \ \tanh\left(\frac{P(k)}{PET(k)}\right) \quad (22)$$

where C is the first model parameter that specifies the variations in the evapotranspiration over a catchment area. Streamflow (Q) from a catchment is then estimated from the Soil Moisture (SM) content available over the area as given in Equation 23.

$$Q(k) = SM(k) \ \tanh\left(\frac{SM(k)}{\mathrm{SC}}\right) \quad (23)$$

Here SC is the second model parameter and it reflects the amount of water storage capacity over the catchment. The details of the model parameters, their ranges, and units are presented in Table 2. The amount of soil moisture available at each time step is given by $\mathrm{SM}(k) = SM(k-1) + P(k) - AET(k)$. Hence, Equation 23 becomes

$$Q(k) = [\mathrm{SM}(k-1) + P(k) - AET(k)] \ \tanh\left(\frac{\mathrm{SM}(k-1)+P(k)-AET(k)}{\mathrm{SC}}\right) \quad (24)$$

The soil moisture at the end of each time step is then updated as shown in Equation 25

$$\text{`` } \mathrm{SM}(k) = SM(k-1) + P(k) - AET(k) - Q(k) \quad (25)$$

Table 2: Nominal range of the Two Parameter Model.

Parameter	**Description**	**Range (Unit)**
C	Evapotranspiration parameter	0.2 to 2 (-)
SC	Water storage capacity	100 – 2000 (mm)

4.2 Preparation of profile soil moisture observation

The raw ASCAT observation obtained cannot be used directly for assimilation purposes because it does not exactly represent the soil moisture that is simulated by the model. This section describes the steps involved in the preparation of soil moisture observation prior to assimilation.

4.2.1 Pre-processing of ASCAT data

ASCAT H111 product provided at 12.5km swath geometry was first masked to the bounding boundary of the study area. The surface soil moisture observations and corresponding 'sm_noise' data representing the error in the observation are then interpolated using the Thiessen polygon method and weighted averaged to the whole basin.

4.2.2 Surface to profile soil moisture

ASCAT soil moisture product derived in swath geometry represents only the top 5 cm of the soil layer while the model simulates the soil water storage at the profile level. Therefore, the data needs to be converted to the profile level before assimilating into the model. Over the years, several studies have been conducted to convert the surface to profile soil moisture (Houser et al., 1998; Sabater et al., 2007; Tobin et al., 2017) and exponential filter stands out over others due to its applicability in different soil types and land cover. Though it was initially proposed for ERS Scatterometer by Wagner et al., (1999), it was later applied for other soil moisture datasets (Albergel et al., 2008; Brocca et al., 2012; Wang et al., 2017) and the resultant profile soil moisture estimate showed good agreement with the in-situ observations.

The soil water index (SWI) estimated by the exponential filter (Equation 26)

represents the moisture condition at the root zone level transferred from the surface soil moisture time series. The water level at the bottom layer is filled based on the antecedent soil conditions at the top layer. The model reasonably assumes an exponential relationship between the two layers with more recent events showing a higher contribution to the SWI value than the previous one.

$$\mathrm{SWI}\,(k_n) = \frac{\sum_i^n \mathrm{SSM}(k_i).e^{\frac{-(k_n-k_i)}{T}}}{\sum_i^n e^{\frac{-(k_n-k_i)}{T}}} k_i \leq k_n \quad (26)$$

here $\mathrm{SWI}\,(k_n)$ is the soil water index estimated at the time k_n and value usually ranges between 0 and 1, $\mathrm{SSM}\,(k_i)$ is the remotely sensed surface soil moisture at the time k_i and T is the only parameter in this model representing the characteristics time length (in days). T takes into account the depth of the second layer and physical properties of the soil such as hydraulic property and pseudo diffusivity. Due to the exponential weightage function, more significance will be given to the recent observations. That is, SSM observed on the day k_n receives more weight than the observation at the time k_{n-1}. Usually, the value of T is determined using the field measurements or assumed meaningfully. Based on the previous study by Albergel et al., (2008) and the absence of field data in the current study T was assumed a value of 10 days.

4.2.3 Bias correction

The soil moisture thus obtained still shows the climatological difference with the model simulated soil moisture due to the differences in their representation. Therefore, it is essential to do bias correction prior to assimilation. Although a higher-order approach (CDF) is available, in this study, SWI-ASCAT was bias-corrected using a mean-variance method (Kumar et al., 2012; Massari et al., 2015) to match the model simulated profile soil moisture time series as shown in Equation 27. This is because, for a limited time span of 4 years, construction of CDF would be suboptimal and rescaling through a simpler method has proved sufficient (Lievens et al., 2015). The Mean-variance approach is regarded as the second-order moment correction as it shifts the overall mean of the observation to match the model time series. Then, it normalizes the obtained data based on the ratio of respective standard deviations.

$$\mathrm{SM}_{\mathrm{scaled}} = \overline{\mathrm{SM}}_{\mathrm{ASCAT}} + \left(\mathrm{SM}_{\mathrm{model}} - \overline{\mathrm{SM}}_{\mathrm{model}}\right) \; . \; \frac{\sigma_{\mathrm{ASCAT}}}{\sigma_{\mathrm{model}}} \quad (27)$$

4.3 First-order sensitivity equations

The obtained bias-corrected profile soil moisture observation is comparable with the model simulated soil moisture and can be used for assimilation purposes directly without the use of an observation operator (h). Therefore, in Equation 17, the Jacobian of h with respect to the model function becomes an *identity matrix* (i.e. $D_y(h) = I$). In addition, the model parameters C and SC and the initial condition S_o are the three driving control variables in the sensitivity vector. Successively, Equation 17 becomes

$$\mathrm{y} = [U(k)\delta S_0 + V_1(k)\delta C + V_2(k)\ \mathrm{SC}] = \begin{bmatrix} U(k) & V_1(k) & V_2(k) \end{bmatrix} \begin{bmatrix} \delta S_0 \\ \delta C \\ SC \end{bmatrix} \quad (28)$$

where $U(k) = \left[\frac{\partial SM(k)}{\partial S_o}\right]_{k=0} = 1$,$V_1(k) = \left[\frac{\partial SM(k)}{\partial C}\right]_{k=0}$and $V_2(k) = \left[\frac{\partial SM(k)}{\partial SC}\right]_{k=0} = 0$

On solving the first-order derivative of the model with respect to the control variables we get Equations 29-31.

$$U(k) = U(k-1)\ \left[1 - tanh\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right) + \frac{\mathrm{Nr}}{\mathrm{SC}}\ \mathrm{x}\ \left(1 - tanh\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right)^2\right)\right] \quad (29)$$

$$V_1(k) = \left[V_1(k-1) - PET(k)\ \tanh\left(\frac{P(k)}{\mathrm{PET}(k)}\right)\right]\left[1 - tanh\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right) + \frac{\mathrm{Nr}}{\mathrm{SC}}\ \left(1 - tanh\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right)^2\right)\right] \quad (30)$$

$$V_2(k) = V_2(k{-}1)\left[1 - tanh\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right) + \frac{\mathrm{Nr}}{\mathrm{SC}}\ \left(1 - tanh\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right)^2\right)\right] + \left[\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right)^2\left(1 - tanh\left(\frac{\mathrm{Nr}}{\mathrm{SC}}\right)^2\right)\right] \quad (31)$$

where $\mathrm{Nr} = \mathrm{SM}(k-1) + P(k) - AET(k)$.

Following Equation 20, the obtained sensitivity vector $H(k) = [U(k) \quad V_1(k) \quad V_2(k)]$ was used as the driving force to perturb the control vector $\delta e = \begin{bmatrix} \delta S_0 & \delta C & \mathrm{SC} \end{bmatrix}^T$ such that the new control vector after few iterations renders forecast error (E) purely random.

4.4 Assimilation of sensitive observations

4.4.1 Calibration and assimilation phase

For initializing the assimilation in the first iteration, the model parameters and the initial condition (i.e. control vector) can be assumed meaningfully or through using calibrated values. For this purpose, the model was calibrated from 1$^{\mathrm{st}}$ January 2000 to 31$^{\mathrm{st}}$ May 2007 with KGE as the evaluation criteria. Later, assimilation was performed for the four monsoon years from 1$^{\mathrm{st}}$ June 2007 to 30$^{\mathrm{th}}$ June 2011 to estimate streamflow. To understand the effect of assimilating only sensitive soil moisture observations and to determine the impact of temporally evolving sensitivity of initial condition during assimilation, four different assimilation scenarios were adopted in this study. Further, the simulations from each scenario were compared against the open-loop simulation (model run without assimilation) to assess their performance in hydrological simulations. Description of the adopted four scenarios is as follows:

Scenario 1: Assimilation of all the observations considering the whole time period from 1$^{\mathrm{st}}$ June 2007 till 30$^{\mathrm{th}}$ June 2011.

Scenario 2: Assimilation of only the sensitive observations considering the whole time period from 1$^{\mathrm{st}}$ June 2007 till 30$^{\mathrm{th}}$ June 2011.

Scenario 3: Assimilation of all the observations considering an annual time window at a time.

Scenario 4: Assimilation of only the sensitive observations considering an annual time window at a time.

In the 2nd and 4th scenarios, a threshold value (Th) was determined to differentiate sensitive and non-sensitive observations as shown in Equation 32. The observations are considered to be sensitive in nature for those time steps when the cumulative sensitivity (i.e.$U(k)^2 + V1(k)^2 + V2(k)^2$) is more than this threshold value. That is, for a given time step when $U(k)^2 + V1(k)^2 + V2(k)^2 \geq Th$ then the corresponding soil moisture observation SM(k) was classified as sensitive observation

$$Th = A * \frac{\sum_{k=1}^{N}\left(U(k)^2 + V1(k)^2 + V2(k)^2\right)}{N} \quad (32)$$

here N is the length of the time window taken. For the 2nd and 4th scenarios, the N value is 1491 and 365 respectively. 'A' is the first assimilation parameter and the value of the A was assumed to range between 0.1 and 2.5. For a minimum value of A, Th will be at least greater than 10 percent of the mean cumulative sensitivity. Also, the sum of the squared sensitive value was adopted because the magnitude of each sensitivity is more important than nature (sign) of the sensitivity. In addition, to determine the impact of initial condition sensitivity during assimilation different time window frames were adopted. Especially, in scenarios 3 and 4, assimilation was performed four times separately considering one year at a time. The soil moisture simulated at the end of each year was used as the antecedent soil moisture (initial condition) for the subsequent year.

The 'sm_noise' flag procured along with ASCAT observation ranges between 0.05 and 0.15 values during the assimilation phase. However, this data does not represent the observation error variance (R) as in Equation 21 correctly. This is because the ASCAT soil moisture represents moisture at the surface level in the relative percentage unit. Whereas, the final soil moisture used for assimilation is a bias-corrected product at the profile level. Moreover, considering the error induced during surface to profile conversion and in bias correction, observation error variance will vary more than the observed 'sm_noise' value. Loizu et al., (2018) and Massari et al., (2015) stated that the calibration of the observational error is necessary for each catchment area and model structure when ground validation is impossible and this might improve the performance of assimilation optimally. Therefore, 'R' (refer to Equation 21) was taken as the second assimilation parameter. Considering the 'sm_noise' flag range and possible error caused during the data preparation, the error variance value of the final observation was assumed to range between 0.05 and 0.35 representing a 5 to 35 percent error rate.

To assess the performance of the streamflow simulations during open-loop and four assimilation scenarios, RMSE, PBIAS, and KGE performance criteria were determined. Among these, KGE was used as the objective function to determine

the assimilation parameters (A and R). This is because, KGE takes into account three components to determine the differences between model simulations and observations namely, variability error, bias error, and linear correlation (Gupta et al., 2009) as shown in Equation 33.

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\mathrm{SD}_{\mathrm{sim}}}{\mathrm{SD}_{\mathrm{obs}}} - 1\right)^2 + \left(\frac{\mu_{\mathrm{sim}}}{\mu_{\mathrm{obs}}} - 1\right)^2} \quad (33)$$

where r is the linear correlation, $\mathrm{SD}_{\mathrm{sim}}$ and $\mathrm{SD}_{\mathrm{obs}}$ are the standard deviations and μ_{sim} and μ_{obs} are the mean of the model simulations and observations respectively. KGE value greater than 0.5 is considered a good model performance (Knoben et. al., 2019). Along with streamflow, the estimated profile soil moisture was also analyzed to understand the model behavior during FSM assimilation to know specifically how the corrected soil moisture state controls the streamflow estimates.

The assimilation process was performed repeatedly for different pairs of A and R values. For each pair, control variables were perturbed (corrected) up to ten iterations and the corresponding KGE value of streamflow simulation was recorded. Finally, the assimilation parameter combination for which the streamflow simulation showed the best agreement with the observation (i.e. highest KGE) was taken as the 'base simulation'. It should be noted that since 1$^{\mathrm{st}}$ and 3$^{\mathrm{rd}}$ scenarios use all the soil moisture observations for assimilation, the above process was performed with only one assimilation parameter R.

4.4.2 Forecast phase

To understand the post assimilation effect on the model behavior, the model was made to run in forecast mode without any further assimilation. For this purpose, streamflow forecasting was performed with nine different lead times (1,2,7,15,30,45,60,75 and 90 days). Out of the four assimilation scenarios adopted, those scenarios whose streamflow performance with KGE value more than 0.8 during the assimilation phase were only analyzed in the forecasting phase. The forecasting phase extends from 1$^{\mathrm{st}}$ July 2011 to 31$^{\mathrm{st}}$ December 2011 and it has a total of 95-forecasting simulations within this period for each shortlisted scenario. During forecasting, the model and the assimilation parameters obtained at the end of the assimilation phase were used with no further change. The first forecasting period starts from 1$^{\mathrm{st}}$ July 2011 till 29$^{\mathrm{th}}$ September 2011(90 days). During this period, the control variables obtained at the end of 30$^{\mathrm{th}}$ June 2011 were taken and only precipitation and potential evapotranspiration were used to run the model. Similarly, the second forecasting period starts from 2$^{\mathrm{nd}}$ July 2011 and ends on 30$^{\mathrm{th}}$ September 2011. The control variables obtained at the end of 1$^{\mathrm{st}}$ July 2011 were used. The same procedure was repeated 95 times till the end of the forecasting phase. Likewise, profile soil moisture was also forecasted 95 times during the forecasting period for each shortlisted scenario. Therefore, at the end of all the analyses, there were 95 profile soil moisture and streamflow forecasts with nine different lead times for each shortlisted scenario.

Similar to the performance assessment during assimilation, forecasting ability for profile soil moisture and streamflow simulations was assessed using RMSE and PBIAS. In addition, to determine the nature and distribution of streamflow during the forecast, peak flow criteria and low flow criteria were also determined following Coulibaly et al.,(2001) and Samuel et al., (2014).

$$\mathrm{PFC} = \frac{\left(\sum_{i=1}^{\mathrm{PL}} (X_{Obs,i} - X_{Sim,i})^2 X_{Obs,i}^2\right)^{\frac{1}{4}}}{\left(\sum_{i=1}^{\mathrm{Ti}} X_{Obs,i}\right)^{\frac{1}{2}}} \quad (34)$$

$$\mathrm{LFC} = \frac{\left(\sum_{i=1}^{\mathrm{TL}} (X_{Obs,i} - X_{Sim,i})^2 X_{Obs,i}^2\right)^{\frac{1}{4}}}{\left(\sum_{i=1}^{\mathrm{Ti}} X_{Obs,i}\right)^{\frac{1}{2}}} \quad (35)$$

where PL is the number of peak flows greater than 33.3% observed flows and TL is the number of low flows less than 66.6% observed flows taken during the forecasting phase.$X_{Sim,i}$ and $X_{Obs,i}$ are the forecasted and observed streamflow. A PFC or LFC equal to 0 represents a perfect fit.

5 Results and Discussions

5.1 Sensitivity results

The sensitivity evolution of the control variables for all the four tested scenarios was derived using Equations 29, 30 and 31 on the calibrated model. The details of model calibration are explained in Section 5.2. The assimilation parameters of 'base simulation' obtained (i.e., the streamflow simulation having the best agreement with the observation) are given in Table A1. For illustration purposes, the sensitivity plots (U1, V1 and V2) representing the best iteration run during the 'base simulation' are displayed in Figures 3(a-d). To see the nature of changes in the sensitivity value during different iterations, supplementary plots are presented in Figure A1 of Appendix A. From Figure A1 it is understood that the iterations are identical and parallel in nature. Also, with each iteration, the model values moved closer towards the streamflow observations.

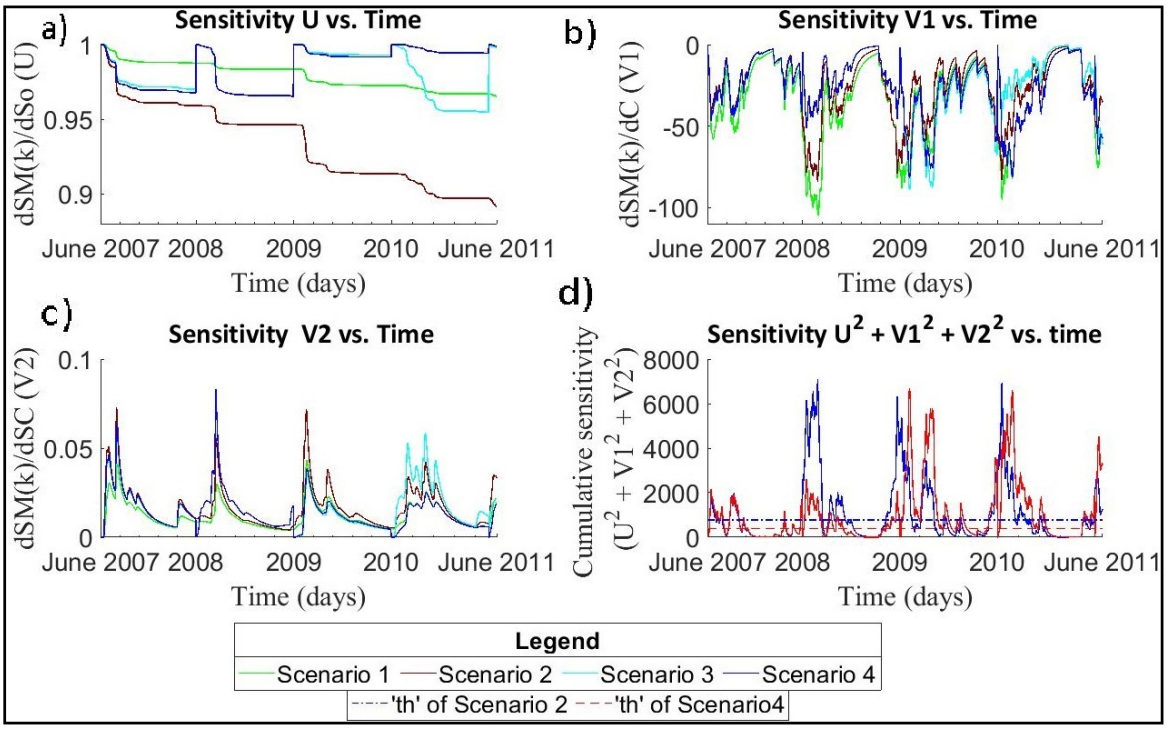

Figure 3: First-order sensitivity evolution of the soil moisture with respect to a) initial condition: U, b) parameter 'C': V1, c) Parameter 'SC': V2, and d) cumulative sensitivity ($U1^2+V1^2+V2^2$)

The sensitivity of the initial condition (Figure 3a) for the 1st and 2nd scenarios shows a constant decreasing trend throughout the assimilation phase. This shows that the effect of the initial condition deteriorates with time. Whereas in the 3rd and 4th scenarios, since the time window considered was annual, sensitivity 'U' resets to 1 and shows relatively a higher value at the beginning of each monsoon season. It can also be noted that the sensitivity decays mainly during the beginning of each monsoon period (June to October) and remains constant throughout the rest of the year for all four scenarios. This is because *U(k)* is mainly dependent on *U(k-1)* as observed from Equation 29. Therefore, during the non-monsoon time when the soil moisture remained negligible and constant, the decay rate in the sensitivity has also remained the same. On the contrary, parameters *C* and *SC* showed more sensitivity (Figures 3b and 3c) during the monsoon period and decay rapidly during the non-monsoon period (December to April). Though sensitivity V1 shows a negative value due to the (- PET × tanh(.)) term in Equation 30, it contributes more to the cumulative sensitivity (Figure 3d) compared to other sensitivities due to its higher magnitude. This shows that magnitude of the sensitivity is more important than the direction (sign) of the sensitivity. Further, it is revealed that evapotranspiration plays a major role in model dynamics. The derived cumulative sensitivity (Figure

3d) was used to obtain the sensitive observations required for assimilation in the 2nd and 4th scenarios. *Th* value represented by the horizontal dashed line differentiates the sensitive and non-sensitive observations. Parameter *A* equal to 0.75 (for scenario 2) and 0.35 (for scenario 4 during all assimilation years) yielded the best KGE value in streamflow simulations during the assimilation phase. During those time steps, when the cumulative sensitive value lies above this threshold line (i.e. *A* times the mean cumulative sensitivity), then those soil moisture observations were considered as sensitive in nature.

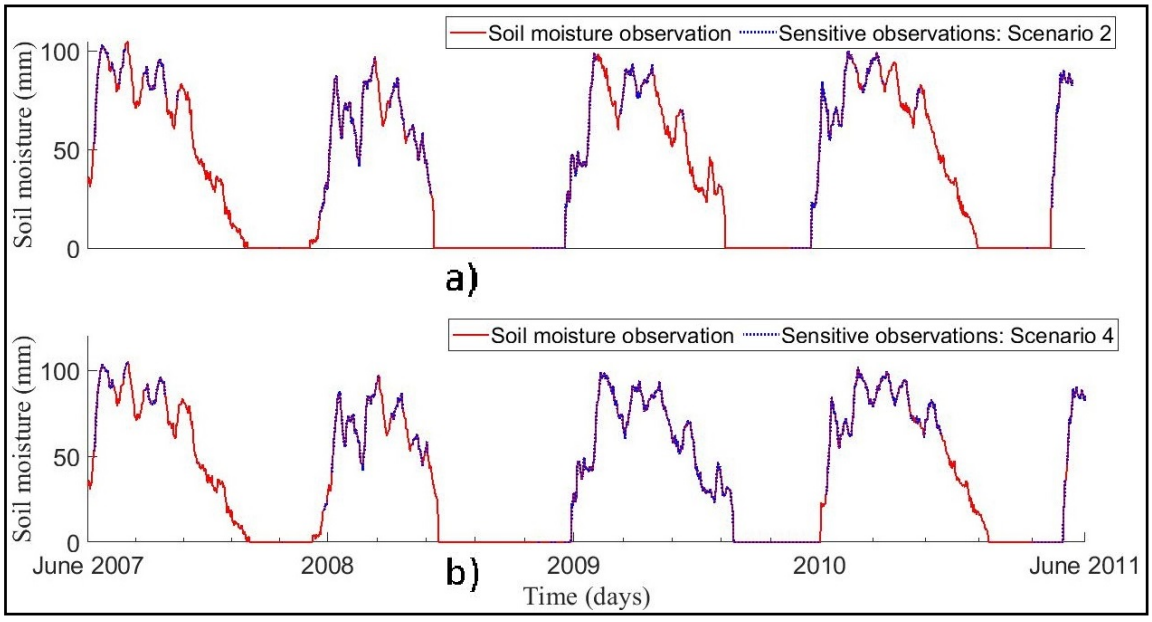

Figure 4: Identification of the sensitive observations during a) scenario 2 and b) scenario 4 selected for assimilation purpose after analyzing the cumulative sensitivity plot

Figure 4 shows the sensitive observations obtained for the 2nd and 4th scenarios. In the 2nd scenario (Figure 4a), only 558 observations were classified as sensitive out of 1491 observations that account for just 37.42 percent. Likewise, in the 4th scenario (Figure 4b), 891 observations were found sensitive accounting for only 59.75 percent of the total observations available during the assimilation phase. Since the threshold value in the 4th scenario is relatively low compared to the 2nd scenario, more observations were classified as sensitive in the 4th scenario. In both these scenarios (Figure 4), it can be noticed that the sensitive observations were predominantly from the monsoon period of the year. In addition, it is understood that the starting period of the monsoon season is more sensitive than the ending period (late November) and this is correctly reflected during the determination of cumulative sensitive (Figure 3d).

5.2 Calibration and assimilation results

To initialize the assimilation processes, the model was calibrated in a deterministic mode for different model parameter space (C and SC) with KGE as the deciding criteria. For this purpose, data of six years from January 2000 to May 2006 (except 2005) were considered to train the model as stated in Section 3.2. For the parameter set C and SC equal to 0.9 and 648 respectively, the model rendered the best streamflow simulation with KGE equal to 0.668. Also, the RMSE and PBias were estimated to be 119.73 cumecs and 24.596 mm respectively. The soil moisture simulated at the end of the calibration period was used as the initial condition during the assimilation phase. With these parameter sets and initial condition, all four assimilation scenarios were carried out. To visualize the impact of sensitivity based assimilation, time-series plots (Figures 5a-c) of the model simulations under different assimilation scenarios were compared against the open-loop simulations. Figure 5a represents the weighted average precipitation value (P) observed in the study area and Figures 5b and 5c show the time series simulation of profile soil moisture and streamflow respectively during the assimilation and open-loop case.

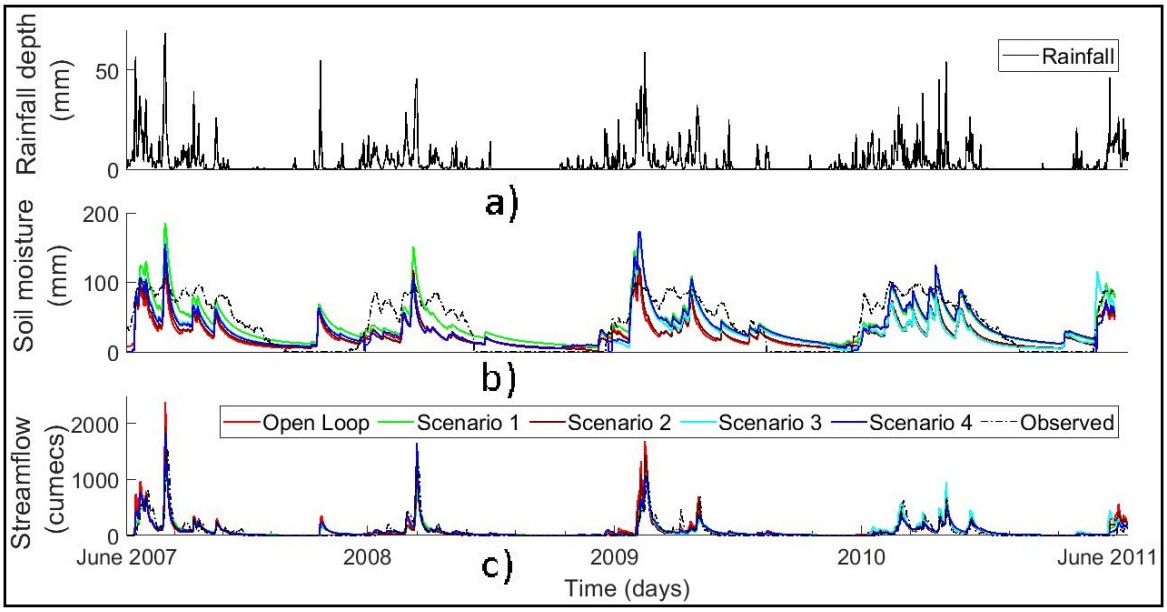

Figure 5: Time series plot of a) rainfall b) profile soil moisture and c) streamflow for open-loop and four FSM assimilation runs.

From Figure 5b it is observed that the overall trend of the simulated profile soil moisture showed better agreement with the observations during all the assimilation runs as compared to the open-loop simulation. However, the soil moisture decays slowly during the end of each monsoon due to the model's inherent error and this led to the overestimation of soil moisture during the non-monsoon period. This, in turn, affects the streamflow estimates during the non-monsoon period. Similarly, from Figure 5c, it is clear that the streamflow simulation

during all the four assimilation runs was in better agreement with the observations than the open-loop simulation. Though the model overpredicts slightly during low flows in the non-monsoon period, the overall performance has significantly improved, especially during the peak flows in the monsoon period. To further evaluate the performance of the assimilation results, different evaluation indices were calculated for profile soil moisture and streamflow simulations, and the results were plotted as a bar chart as shown in Figures 6(a-c) and 7(a-c) respectively.

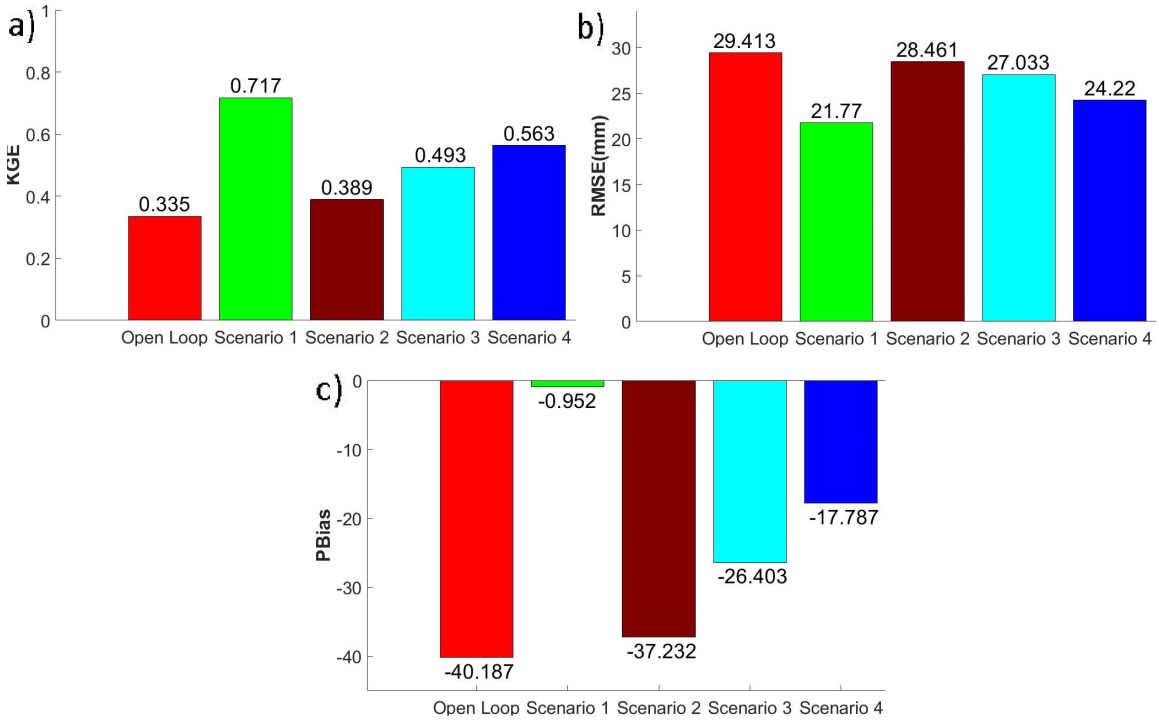

Figure 6: Performance measures of the estimated profile soil moisture for open-loop simulation and four assimilation scenarios representing a) KGE, b) RMSE and c) PBIAS values.

From profile soil moisture simulation results (Figure 6), it is understood that the performance of all the four assimilation scenarios was superior to the open-loop performance. Though marginal improvement is seen while assimilating only sensitive observations, the effect on soil moisture estimate is significant when all the observations were used during 1st scenario. This is obvious as the profile soil moisture is bound to perform better when ingested with all the available observations. The soil moisture results of the 1st scenario showed a significant improvement where the KGE value has doubled from 0.335 to 0.717 and PBIAS value got diminished from -40.18 percent to just -0.95 percent.

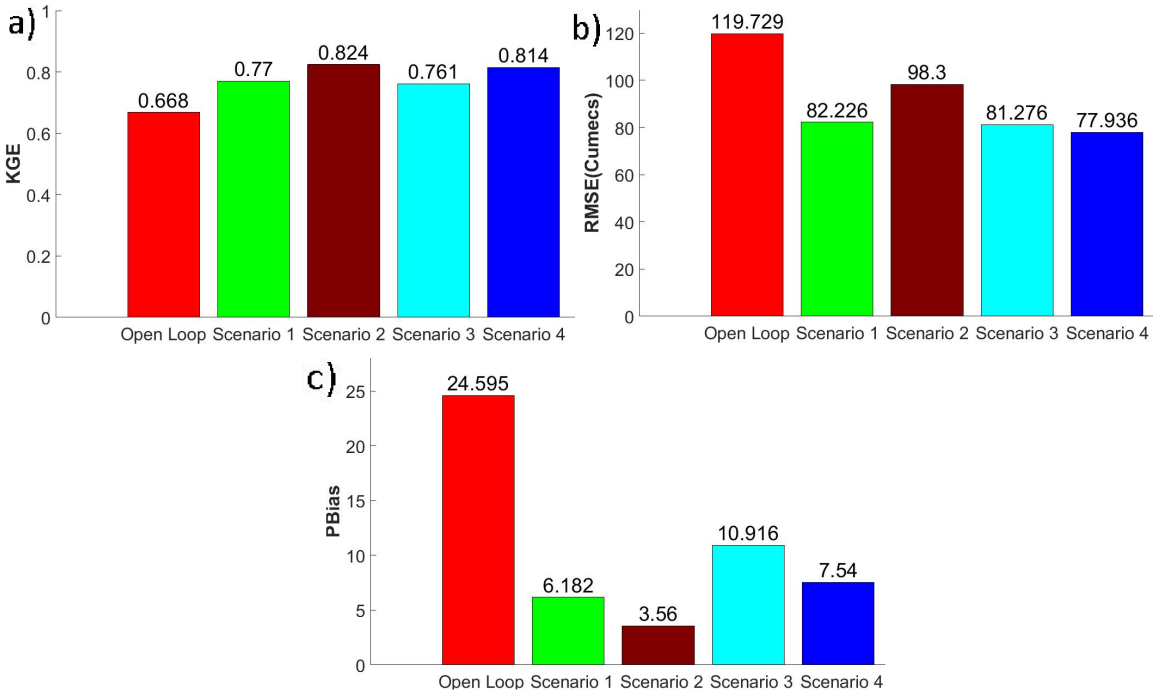

Figure 7: Performance measures of the estimated streamflow for open-loop simulation and four assimilation scenarios representing a) KGE, b) RMSE and c) PBIAS values.

From streamflow simulation results (Figure 7), it is clear that all four assimilation scenarios showed improvement against the open-loop case. Especially, 2nd and 4th scenarios gave the best results with KGE values (Figure 7a) equal to 0.824 and 0.814 respectively. This shows that the effect of assimilation is more profound when only sensitive observations were used. Similarly, the RMSE value (Figure 7b) reduced greatly from 119.73 cumecs during open-loop to 82.226 and 77.936 cumecs during the 2nd and 4th scenarios respectively. Also, PBias value (Figure 7c) contracted by more than 4 times from 24.59 percent during open-loop to just 3.56 percent in the 2nd scenario and 7.54 percent in the 4th scenario. The profile soil moisture results (Figure 6) and streamflow results (Figure 7) showed contrary behavior because the profile soil moisture estimate is not translated properly to streamflow estimate. From Figure 7, especially when comparing either scenario 1 and 3 or scenario 2 and 4, it is clear that the effect of initial condition sensitivity on model performance was marginal compared to the model parameters sensitivity, which is obvious because the magnitude of the sensitivities as shown in Figures 3a-c showed a similar trend.

From these results (Figures 6 and 7), it can be concluded that the FSM based soil moisture assimilation has a significant impact on the hydrological simulations. It is also proved that with less than 40 percent of data used for assimilation,

it is possible to have better hydrological predictions. The computational time has also reduced from 12 hours while assimilating all the observations to just 5 hours when assimilating only sensitive observations when using a 3GHz computer processor speed. Also, all the sensitive observations obtained were from the monsoon period of the year and this justifies the proper temporal placement of observations for effective assimilation. The 2^{nd} and 4^{th} scenarios that showed better streamflow estimates with a KGE value of more than 0.8 were used in the forecasting phase to check for its forecasting capability.

5.3 Forecasting results

To assess the forecasting ability of the FSM based soil moisture assimilation, streamflow was forecasted 95 times (between 1^{st} July to 31^{st} December 2011) with nine different lead times. For this purpose, only the 2^{nd} and 4^{th} scenarios were adopted for this analysis. Here, along with streamflow, profile soil moisture was also forecasted to understand how the effect of FSM assimilation translated from profile soil moisture forecast to streamflow forecast.

5.3.1 Soil moisture forecast

To analyze the trend between open-loop case and assimilation scenarios, the time series plot for the first forecasting run from 1^{st} July 2011 till 29^{th} September 2011 is presented in Figures 8(a-b) as an illustration. From Figure 8a, it is seen that the performance of the soil moisture forecast was better during the 2^{nd} scenario than the 4^{th} scenario and open-loop simulations. Further, a clear under biased trend is observed in all the three soil moisture forecast. This is because the main objective during the assimilation phase was to improve the streamflow estimates and accordingly the model's control variables were corrected to match the streamflow distribution rather than soil moisture. Also, from Figure 8a we can see that the magnitude of the soil moisture reduced after assimilation such that the corresponding streamflow value in Figure 8b (for both the scenarios) moved closer towards the streamflow observation. Yet, the differences between the soil moisture observation and model simulations after assimilation have diminished and showed improvement. Figures 8(a-b) represents only one forecast run and it may not be a correct representation of all the hydrological forecast. Therefore, to assess the overall performance during all the 95-forecasting runs, the mean value of each of the evaluation criteria was evaluated and plotted as a point plot in Figures 9(a-f). Among them, Figures 9a and 9b represent the mean statistics of soil moisture forecast results.

The soil moisture forecast results reveal that the assimilation showed improvement during both 2^{nd} and 4^{th} scenarios against open-loop performance. In particular, the PBias and RMSE values reduced significantly up to 14 and 75 lead days for the 4^{th} scenario and 2^{nd} scenario respectively. Note that the 4^{th} scenario performed poorer than open-loop during the 1^{st} forecast simulation (see Figure 8a). This justifies the importance of considering the mean of 95 simulations. Also, the forecast results (Figures 9a and b) for both assimilation scenarios performed better only during the shorter lead time. Later, it diverges

and performed poorly for longer lead times. This is partially because the noise in the input data is not filtered out properly to forecast the profile soil moisture due to the model's inherent structural error.

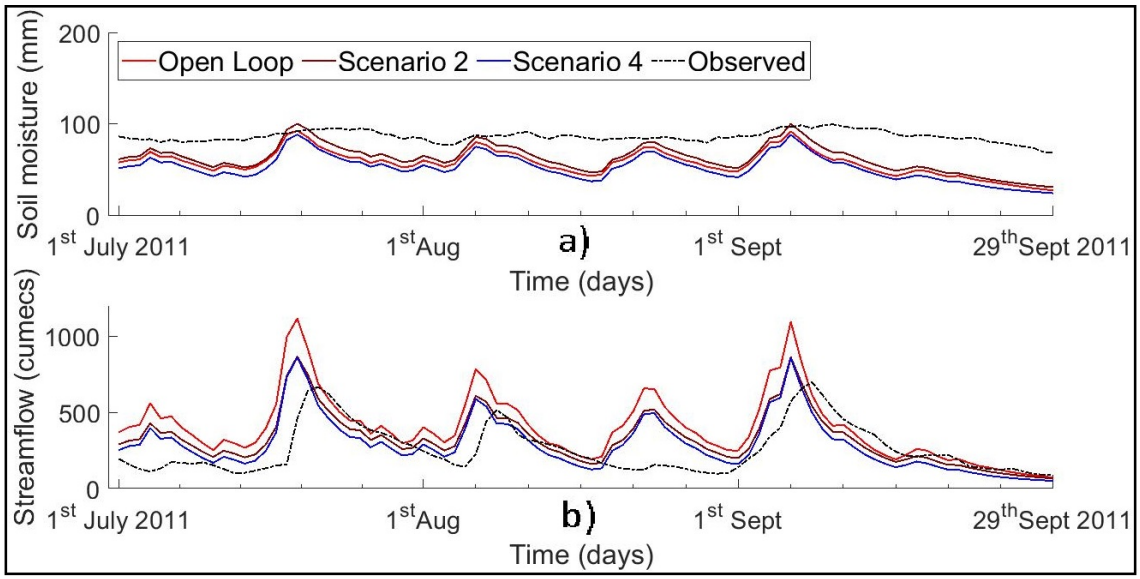

Figure 8: Time series plot of the first 90 lead days forecast model run showing a) profile soil moisture and b) streamflow from July 1-2011 to September-29-2011.

5.3.2 Streamflow forecast

Figure 8b displays the first streamflow forecast up to 90-days lead time. It reveals that the overall performance of the assimilation is in better agreement with the observations. When compared to the open-loop simulation, assimilation scenarios have better captured the overall distribution of the streamflow. Especially, the 4^{th} scenario is showing better performance during the peak flows as compared to the 2^{nd} scenario.

Figures 9c and 9d represent the mean PBias and RMSE results of the streamflow respectively. As concluded from the time series plots (refer to Figure 8b), we can see that the performances of both the assimilation forecasts were much superior than the open-loop forecast. Further, to understand the effect of assimilation during the peak and low flows, mean PFC and LFC were evaluated for all the lead times and plotted in Figures 9e and 9f. It should be noted that the PFC and LFC value closer to 0 implies the forecast is more accurate like PBias and RMSE.

The assimilation scenarios have well captured the extreme events during the streamflow forecast and the effect is prominent till the 45-days lead time. But, for a longer lead time (60 to 90 days) assimilation performances are no better

than the open-loop results.

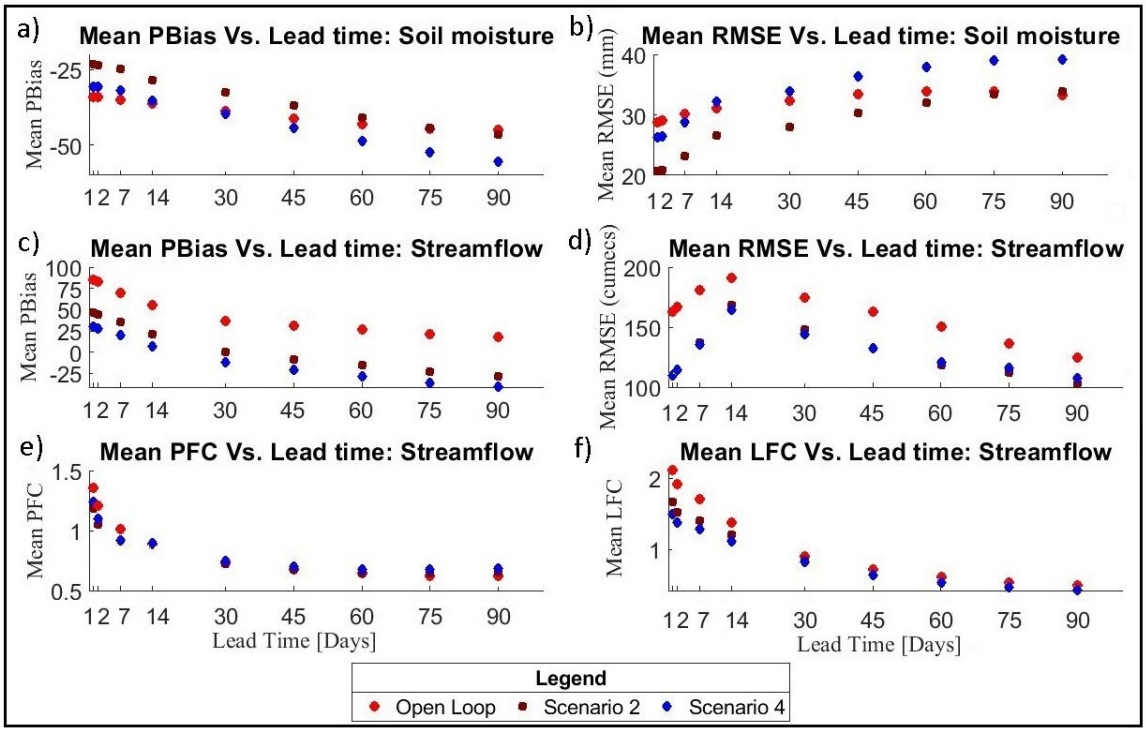

Figure 9: Mean values of PBIAS, RMSE, PFC, and LFC evaluated for 95 simulations of a-b) profile soil moisture, and c-f) streamflow forecasts represented for nine different lead times.

On the contrary, the overall performance of the assimilation scenarios for streamflow forecast (Figures 9c-f) was poor during the shorter lead times (1 to 45 days) and performed better during the longer lead time (60 to 90 days). In particular, the 2nd scenario showed better results compared to the 4th scenario in all evaluation criteria. This is because, unlike sequential assimilation, FSM assimilation tries to correct the model trajectory to match the peaks and low flows of the observations within the assumed time window. Since, in this study, the time window considered was annual for the 4th scenario and four years for the 2nd scenario, control variables are optimized to perform better for longer lead times in the 2nd scenario.

6 Summary and Conclusions

The study proposed a new class of assimilation of soil moisture into a conceptual hydrological model using the FSM approach. FSM uses the sensitivity of the model solution to identify and subset the sensitive observations that have a stronger impact on the estimates of initial condition and model parameters. To evaluate its efficiency, the model is again ingested with all the soil moisture

observations to check for its performance. Also, the study analyzed the impact of temporally evolving sensitivity of initial condition by adopting different time window frames during assimilation. Subsequently, four scenarios were adopted in this study for estimation and forecasting of profile soil moisture and streamflow. For this purpose, KGE, PBias and RMSE were evaluated. Later, to check the performance of streamflow forecasts during peak flows, PFC and LFC were determined.

From the results, it was concluded that during assimilation, the scenarios that assimilated only sensitive observations (2^{nd} and 4^{th}) performed best with the KGE value showing more than 0.8 for the streamflow simulations. Specifically, with less than 30 percent of the observations used in the 2^{nd} scenario, it showed the best performance than the open-loop and other assimilation scenarios. This proves that sensitivity based assimilation has a strong impact on improving hydrological simulations such as streamflow. Moreover, FSM assimilation helped in understanding the decisive placement of the observation during assimilation. For instance, from the time evolution of the sensitivity, it was clear that the observations during the monsoon time alone can contribute much during the assimilation. This has reduced the huge burden on the use of all observations during the assimilation with much less computational time (i.e. 12 hours of computational time when all the observations were assimilated, whereas it was only 5 hours while assimilating only sensitive observations using a 3GHz computer processor speed). However, the impact of adopting different time window frames is negligible. It is understandable because the sensitivity results showed that the parameter sensitivity had a strong influence on the profile streamflow simulations than the initial condition. Profile soil moisture simulation showed a better estimate during all the assimilation scenarios when compared against the open-loop case. Especially, the best result was observed when all the observations were assimilated during the 1^{st} scenario. During the forecasting phase, the 2^{nd} and 4^{th} scenarios showed only marginal improvement up to 75 and 14 lead days respectively. Comparatively, the soil moisture performance is only marginal, unlike streamflow results. This is because, in this study, the model was trained to accurately estimate streamflow during the assimilation phase, which might have degraded the performance of the soil moisture forecast. On the other hand, streamflow simulation showed significant improvement during all the assimilation scenarios when compared with the open-loop case. Also, during the forecasting phase, the streamflow forecast showed better performance for 2^{nd} and 4^{th} scenarios than the open-loop run up to 45 lead days but for the longer lead times (60 to 90 days), the improvement was only marginal. However, the overall performance of the assimilation scenarios during the streamflow forecast was better for the longer lead times than the shorter lead times. This is because during assimilation the model trajectory is smoothened to reduce the sum of squared difference considering the whole time window.

With new satellite missions being launched, the amount of data is expected to increase by orders of magnitude. In such a scenario, the FSM based strategy would play a vital role in the optimal selection of appropriate observations dur-

ing the assimilation process and the researchers can gain insight on when and what to assimilate with the reduced computational burden. Although the current work identified and effectively assimilated sensitive observations in the time domain, further research is needed to extend this work by including spatial heterogeneity using distributed hydrological models to identify spatially sensitive locations. Once identified, the observations from only these sensitive locations can be leveraged to enhance assimilation efficiency, especially in regions where data availability remains a challenge.

Acknowledgments

The authors would like to thank India Metrological Department (IMD) for providing temperature data and National Center for Environmental Prediction (NCEP) for providing humidity and wind speed data. The authors also acknowledge EUMETSAT (http://doi.org/10.15770/EUM_SAF_H_0004) for providing level 2 processed ASCAT data. Gratitude is extended to Karnataka State Natural Disaster Monitoring Centre (KSNDMC), India for sharing the input rainfall data at taluk level (administrative unit) and Central Water Commission (CWC), Government of India (https://indiawris.gov.in/wris) for gauge streamflow data.

Data availability

The authors do not have the right to upload the data publicly. Interested readers can approach the concerned organization and obtain the data.

Conflict of interest statement

No potential conflict of interest was reported by the authors.

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., et al. (2008). From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences*, 12(6), 1323–1337. https://doi.org/10.5194/hess-12-1323-2008Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., K et al. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment*, 118, 215–226. https://doi.org/10.1016/j.rse.2011.11.017Alvarado-Montero, R., Schwanenberg, D., Krahe, P., Helmke, P., & Klein, B. (2017). Multi-parametric variational data assimilation for hydrological forecasting. *Advances in Water Resources*, 110, 182–192. https://doi.org/10.1016/j.advwatres.2017.09.026Alvarez-Garreton, C., Ryu, D., Western, A. W., Su, C. H., Crow, W. T., Robertson, D. E., & Leahy, C. (2015). Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: Comparison between lumped and semi-distributed schemes. *Hydrology and Earth System Sciences*, 19(4),

1659–1676. https://doi.org/10.5194/hess-19-1659-2015Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., & Hahn, S. (2012). Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 50(7), 2542–2555. https://doi.org/10.1109/TGRS.2011.2177468Calvet, J.C., Noilhan, J. and Bessemoulin, P. (1998). Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: A feasibility study based on field measurements. Journal of Applied Meteorology and Climatology, 37(4), pp.371-386. https://doi.org/10.1175/1520-0450(1998)037<0371:RTRZSM>2.0.CO;2Cioaca, A., Sandu, A., & de Sturler, E. (2013). Efficient methods for computing observation impact in 4D-Var data assimilation. *Computational Geosciences*, 17(6), 975–990. https://doi.org/10.1007/s10596-013-9370-2Coulibaly, P., Anctil, F., & Bobee, B. (2001). Multivariate reservoir inflow forecasting using temporal neural networks. *Operations Research*, 6(5), 367–376. https://doi.org/10.1061/(ASCE)1084-0699(2001)6:5(367) Dechant, C. M., & Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, 48(4), 1–15. https://doi.org/10.1029/2011WR011011Dile, Y. T., & Srinivasan, R. (2014). Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: An application in the blue nile river basin. *Journal of the American Water Resources Association*, 50(5), 1226–1241. https://doi.org/10.1111/jawr.12182Droogers, P., & Bastiaanssen, W. (2002). Irrigation Performance using Hydrological and Remote Sensing Modeling. *Journal of Irrigation and Drainage Engineering*, 128(1), 11–18. https://doi.org/10.1061/(asce)0733-9437(2002)128:1(11)Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143–10162. https://doi.org/10.1029/94jc00572Fuka, D. R., Walter, M. T., Macalister, C., Degaetano, A. T., Steenhuis, T. S., & Easton, Z. M. (2014). Using the Climate Forecast System Reanalysis as weather input data for watershed models. *Hydrological Processes*, 28(22), 5613–5623. https://doi.org/10.1002/hyp.10073Gejadze, I., & Malaterre, P. O. (2017). Discharge estimation under uncertainty using variational methods with application to the full Saint-Venant hydraulic network model. International Journal for Numerical Methods in Fluids, 83(5), 405–430. https://doi.org/10.1002/fld.4273Gevaert, A. I., Renzullo, L. J., Van Dijk, A. I. J. M., Van Der Woerd, H. J., Weerts, A. H., & De Jeu, R. A. M. (2018). Joint assimilation of soil moisture retrieved from multiple passive microwave frequencies increases robustness of soil moisture state estimation. *Hydrology and Earth System Sciences*, 22(9), 4605–4619. https://doi.org/10.5194/hess-22-4605-2018Gül, G. O., Harmancioğlu, N., & Gül, A. (2010). A combined hydrologic and hydraulic modeling approach for testing efficiency of structural flood control measures. *Natural Hazards*, 54(2), 245–260. https://doi.org/10.1007/s11069-009-9464-2Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Impli-

cations for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003Han, E., Merwade, V., & Heathman, G. C. (2012). Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. *Journal of Hydrology*, 416, 98–117. https://doi.org/10.1016/j.jhydrol.2011.11.039Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, 34(12), 3405–3420. http://dx.doi.org/10.1029/1998WR900001

Jay-Allemand, M., Javelle, P., Gejadze, I., Arnaud, P., Malaterre, P.-O., Fine, J.-A., & Organde, D. (2019). On the potential of variational calibration for a fully distributed hydrological model: application on a Mediterranean catchment. *Hydrology and Earth System Sciences Discussions*, 24(11), 5519-5538. https://doi.org/10.5194/hess-24-5519-2020

Kalman, R.E.,. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluid Engineering*, 82(1), 35-45. https://doi.org/10.1115/1.3662552

Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., & Peters-Lidard, C. D. (2009). Role of subsurface physics in the assimilation of surface soil moisture observations. *Journal of Hydrometeorology*, 10(6), 1534–1547. https://doi.org/10.1175/2009JHM1134.1

Kumar, S. V., Reichle, R. H., Harrison, K. W., Peters-Lidard, C. D., Yatheendradas, S., & Santanello, J. A. (2012). A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resources Research*, 48(3), 1–16. https://doi.org/10.1029/2010WR010261

Lakshmivarahan, S., & Lewis, J. M. (2010). Forward Sensitivity Approach to Dynamic Data Assimilation. *Advances in Meteorology*, 2010, 1–12. https://doi.org/10.1155/2010/375615

Lakshmivarahan, S., Lewis, J. M., & Jabrzemski, R. (2017). Forecast error correction using dynamic data assimilation. *Springer Atmospheric Sciences*. Switzerland: Springer. https://doi.org/10.1007/978-3-319-39997-3

Lakshmivarahan, S., Lewis, J. M., & Hu, J. (2020). On Controlling the Shape of the Cost Functional in Dynamic Data Assimilation: Guidelines for Placement of Observations and Application to Saltzman's Model of Convection. *Journal of Atmospheric Sciences*, 77(8), 2969-2989. https://doi.org/10.1175/JAS-D-19-0329.1

Le Dimet, F. X., & Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus, Series A,* 38 A(2), 97–110. https://doi.org/10.3402/tellusa.v38i2.11706

Leach, J. M., Kornelsen, K. C., & Coulibaly, P. (2018). Assimilation of near-real time data products into models of an urban basin. *Journal of Hydrology*, 563, 51–64. https://doi.org/10.1016/j.jhydrol.2018.05.064

Lee, H., Seo, D. J., Liu, Y., Koren, V., McKee, P., & Corby, R. (2012). Variational assimilation of streamflow into operational distributed hydrologic models: Effect of spatiotemporal scale of adjustment. *Hydrology and Earth System Sciences*, 16(7), 2233–2251. https://doi.org/10.5194/hess-16-2233-2012

Leisenring, M., & Moradkhani, H. (2012). Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. Journal of Hydrology, 468–469, 268–282. https://doi.org/10.1016/j.jhydrol.2012.08.049

Lewis, J. M., Lakshmivarahan, S., & Dhall, S. (2006). Dynamic Data Assimilation, A Least Squares Approach. *Mathematics of Computation.*(Vol 104) New York, NY: Cambridge University Press.

Lewis, J. M., Lakshmivarahan, S., Hu, J., Edwards, R., Robert, A., Thompson, R. L., & Corfidi, S. F. (2016). Ensemble Forecasting of Return Flow over the Gulf of Mexico. *Electronic Journal of Severe Storms Meteorology*, 11(4), 1–26.

Lewis, J. M., Lakshmivarahan, S., Hu, J., & Rabin, R. (2020). Placement of Observations to Correct Return-Flow Forecasts, *15*(4), 1–20.

Lievens, H., Tomer, S. K., Al Bitar, A., De Lannoy, G. J. M., Drusch, M., Dumedah, G., et al. (2015). SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sensing of Environment*, 168, 146–162. https://doi.org/10.1016/j.rse.2015.06.025

Loizu, J., Massari, C., Álvarez-Mozos, J., Tarpanelli, A., Brocca, L., & Casalí, J. (2018). On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. *Advances in Water Resources*, 111, 86–104. https://doi.org/10.1016/j.advwatres.2017.10.034

Lowe, S. A. (2006). Sanitary sewer design using EPA storm water management model (SWMM). *Computer Applications in Engineering Education*, 18(2), 203–212. https://doi.org/10.1002/cae.20124

Li, N., Kinzelbach, W., Li, H., Li, W., & Chen, F. (2021). Improving parameter and state estimation of a hydrological model with the ensemble square root filter. *Advances in Water Resources*, 147, 103813. https://doi.org/10.1016/j.advwatres.2020.103813

Lü, H., Yu, Z., Zhu, Y., Drake, S., Hao, Z., & Sudicky, E. A. (2011). Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation. *Advances in Water Resources*, 34(3), 395–406. https://doi.org/10.1016/j.advwatres.2010.12.005

Massari, C., Brocca, L., Tarpanelli, A., & Moramarco, T. (2015). Data assimilation of satellite soil moisture into rainfall-runoffmodelling: A complex recipe?. *Remote Sensing*, 7(9), 11403-11433. https://doi.org/10.3390/rs70911403

Massari, C., Camici, S., Ciabatta, L., & Brocca, L. (2018). Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean

area: State update versus rainfall correction. *Remote Sensing*, 10(2), 292. https://doi.org/10.3390/rs10020292

McLaughlin, D. (1995). Recent developments in hydrologic data assimilation. *Reviews of Geophysics,* 33(2 S), 977–984. https://doi.org/10.1029/95RG00740

Meng, S., Xie, X., & Liang, S. (2017). Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. *Journal of Hydrology*, 550, 568–579. https://doi.org/10.1016/j.jhydrol.2017.05.024

Montzka, C., Pauwels, V. R. N., Franssen, H. H., Han, X., & Vereecken, H. (2012). Multivariate and Multiscale Data Assimilation in Terrestrial Systems: A Review. *Sensors*, 12(12), 16291-16333.https://doi.org/10.3390/s121216291

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Naz, B. S., Kurtz, W., Montzka, C., Sharples, W., Goergen, K., Keune, J., et al. (2018). Improving soil moisture and runoff simulations over Europe using a high-resolution data-assimilation modeling framework. *Hydrology and Earth System Sciences Discussions*, (March), 1–32. https://doi.org/10.5194/hess-2018-24

Oubanas, H., Gejadze, I., Malaterre, P. O., Durand, M., Wei, R., Frasson, R. P. M., & Domeneghetti, A. (2018). Discharge Estimation in Ungauged Basins Through Variational Data Assimilation: The Potential of the SWOT Mission. *Water Resources Research*, 54(3), 2405–2423. https://doi.org/10.1002/2017WR021735

Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., & Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25× 0.25) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam*, 65(1), 1–18.

Pan, M., Wood, E. F., Wójcik, R., & Mccabe, M. F. (2008). Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation, *Remote Sensing of Environment*, 112(4), 1282–1294. https://doi.org/10.1016/j.rse.2007.02.039

Pathiraja, S., Marshall, L., Sharma, A., & Moradkhani, H. (2016). Hydrologic modeling in dynamic catchments: A data assimilation approach. *Journal of the American Water Resources Association,* 52(5), 3350–3372.https://doi.org/10.1002/2015WR017192

Patil, A. A., & Ramsankaran, R. (2017). Improving streamflow simulations and forecasting performance of SWAT model by assimilating remotely sensed soil moisture observations. *Journal of Hydrology*, 555, 683–696. https://doi.org/10.1029/2019JD031369

Patil, A. A., & Ramsankaran, R. (2018). Improved streamflow simulations

by coupling soil moisture analytical relationship in EnKF based hydrological data assimilation framework. *Advances in Water Resources*, 121, 173–188. https://doi.org/10.1016/j.advwatres.2018.08.010

Puente, C. E., & Bras, R. L. (1987). Application of nonlinear filtering in the real time forecasting of river flows. Water Resources Research, 23(4), 675–682. https://doi.org/10.1029/WR023i004p00675

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2001). Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications. IEEE Transactions on Geoscience and Remote Sensing, 39(8), 1708–1718. https://doi.org/10.1109/36.942549

Rezaie-Balf, M., Naganna, S. R., Kisi, O., & El-Shafie, A. (2019). Enhancing streamflow forecasting using the augmenting ensemble procedure coupled machine learning models: case study of Aswan High Dam. *Hydrological Sciences Journal*, 64(13), 1629–1646. https://doi.org/10.1080/02626667.2019.1661417

Rogelis, M. C., & Werner, M. (2018). Streamflow forecasts from WRF precipitation for flood early warning. *Hydrological Earth System Science*, 22(1), 853–870. https://doi.org/10.5194/hess-22-853-2018

Roulin, E. (2007). Skill and relative economic value of medium-range hydrological ensemble predictions. *Hydrology and Earth System Sciences*, 11(2), 725–737. https://doi.org/10.5194/hess-11-725-2007

Sabater, J. M., Jarlan, L., Calvet, J. C., Bouyssel, F., & De Rosnay, P. (2007). From near-surface to root-zone soil moisture using different assimilation techniques. *Journal of Hydrometeorology*, 8(2), 194–206. https://doi.org/10.1175/JHM571.1

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8), 1015–1057. https://doi.org/10.1175/2010BAMS3001.1

Sahoo, A. K., De Lannoy, G. J. M., Reichle, R. H., & Houser, P. R. (2013). Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA. *Advances in Water Resources*, 52, 19–33. https://doi.org/10.1016/j.advwatres.2012.08.007

Sakov, P., Oliver, D. S., & Bertino, L. (2012). An iterative EnKF for strongly nonlinear systems. Monthly Weather Review, 140(6), 1988–2004. https://doi.org/10.1175/MWR-D-11-00176.1

Salamon, P., & Feyen, L. (2009). Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. Journal of Hydrology, 376(3–4), 428–442. https://doi.org/10.1016/j.jhydrol.2009.07.051

Samuel, J., Coulibaly, P., Dumedah, G., & Moradkhani, H. (2014). Assessing model state and forecasts variation in hydrologic data assimilation. *Journal of*

Hydrology, 513, 127–141. https://doi.org/10.1016/j.jhydrol.2014.03.048

Seo, D. J., Cajina, L., Corby, R., & Howieson, T. (2009). Automatic state updating for operational streamflow forecasting via variational data assimilation. *Journal of Hydrology*, 367(3–4), 255–275. https://doi.org/10.1016/j.jhydrol.2009.01.019

Seo, D.J., Koren, V. and Cajina, N. (2003). Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. Journal of Hydrometeorology, 4(3), pp.627-641. https://doi.org/10.1175/1525-7541(2003)004<0627:RVAOHA>2.0.CO;2

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. Hydrological Sciences Journal, 61(13), 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Shutyaev, V., Le Dimet, F. X., & Parmuzin, E. (2018). Sensitivity analysis with respect to observations in variational data assimilation for parameter estimation. *Nonlinear Processes in Geophysics*, 25(2), 429–439. https://doi.org/10.5194/npg-25-429-2018

Tobin, K. J., Torres, R., Crow, W. T., & Bennett, M. E. (2017). Multi-decadal analysis of root-zone soil moisture applying the exponential filter across CONUS. *Hydrology and Earth System Sciences Discussions*, 21(9), 4403-4417. https://doi.org/10.5194/hess-2017-118

Tromble, E. M., Lakshmivarahan, S., Kolar, R. L., & Dresback, K. M. (2016). Application of the forward sensitivity method to a GWCE-based shallow water model. *Journal of Marine Science and Engineering*, 4(4), 73. https://doi.org/10.3390/jmse4040073

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sensing of Environment*, 70(2), 191–207. https://doi.org/10.1016/S0034-4257(99)00036-X

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., et al. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*, 22(1), 5–33. https://doi.org/10.1127/0941-2948/2013/0399

Wang, T., Franz, T. E., You, J., Shulski, M. D., & Ray, C. (2017). Evaluating controls of soil properties and climatic conditions on the use of an exponential filter for converting near surface to root zone soil moisture contents. *Journal of Hydrology*, 548, 683–696. https://doi.org/10.1016/j.jhydrol.2017.03.055

Weerts, A. H., & El Serafy, G. Y. H. (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research,* 42(9), 1–17. https://doi.org/10.1029/2005WR004093

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Xiong, L., & Guo, S. (1999). A two-parameter monthly water balance model and its application. *Journal of Hydrology*, 216(1–2), 111–123. https://doi.org/10.1016/S0022-1694(98)00297-2

Xiong, M., Liu, P., Cheng, L., Deng, C., Gui, Z., Zhang, X., & Liu, Y. (2019). Identifying time-varying hydrological model parameters to improve simulation efficiency by the ensemble Kalman filter: A joint assimilation of streamflow and actual evapotranspiration. *Journal of Hydrology*, 568, 758–768. https://doi.org/10.1016/j.jhydrol.2018.11.038

Zhu, Y., Chen, X., Fu, X., Hou, T., Lü, H., Jia, Y., et al. (2012). The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method. *Journal of Hydrology*, 480, 102–114. https://doi.org/10.1016/j.jhydrol.2012.12.011

Appendix A

Table A1 shows the calibrated assimilation parameter values of the 'base simulation' run. For scenarios 3 and 4, four separate values represent the assimilation performance during each year taken successively four times.

Table A1: Calibrated parameter values of the base simulation run for all the four assimilation scenarios.

Scenario	**Year**	**R**	**A**
1	2007 -2011	0.25	-
2	2007 -2011	0.335	0.75
3	2007- 2008	0.35	-
	2008- 2009	0.05	
	2009 - 2010	0.075	
	2010 - 2011	0.05	
4	2007- 2008	0.325	0.35
	2008- 2009	0.05	0.35
	2009 - 2010	0.065	0.35
	2010 - 2011	0.35	0.35

Figure A1 shows the different iterations of the 'base simulation' run. The first column represents the sensitivity during the 1st scenario, the second column represents the 2nd scenario, the third column represents the 3rd scenario, and the last column represents the 4th scenario.

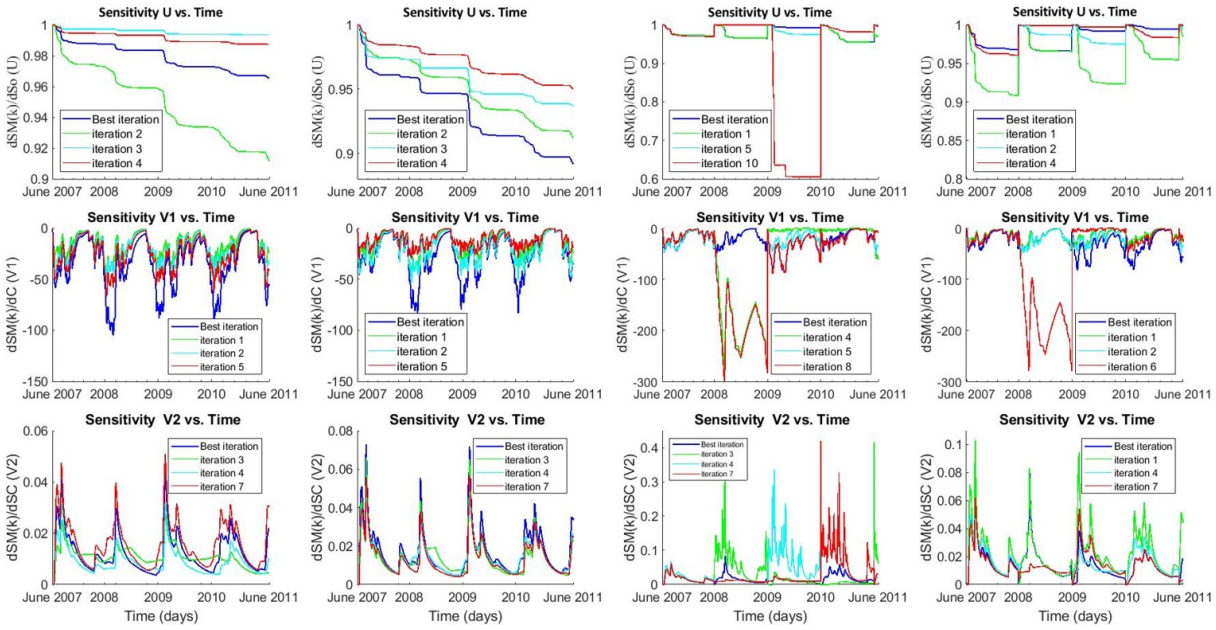

Figure A1: Evolution of sensitivity of initial condition:(U), parameter C: (V1), and parameter SC: (V2) displaying different iterations for the base simulation run