Dual correction of rainfall and root zone soil moisture estimates for improving streamflow simulations

Ramesh, V., Patil, A. ORCID: https://orcid.org/0000-0001-9804-5925 and Ramsankaran, R. ORCID: https://orcid.org/0000-0001-8602-1934 (2025) Dual correction of rainfall and root zone soil moisture estimates for improving streamflow simulations. ARC Geophysical Research, 1. 15. ISSN 3067-6711 doi: 10.5149/arc-gr.1662 Available at https://centaur.reading.ac.uk/127382/

ARC Alliance

ARC Geophysical Research

GR Geophysical Research

Dual Correction of Rainfall and Root Zone Soil Moisture Estimates for Improving Streamflow Simulations

Visweshwaran R[*1], Amol Patil[2], and RAAJ Ramsankaran[*3]

[1]Department of Meteorology, University of Reading, UK

[2]Senior Data Scientist, Reliance Industries Limited, India

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

Abstract

Satellite-based precipitation and soil moisture products are often associated with uncertainties, rendering them less reliable for hydrological applications. The present study proposes a dual correction scheme employing satellite-based soil moisture estimates to update satellite-based rainfall and modelled soil moisture states. First, the artificial neural network (ANN) was utilised to correct TRMM 3B42RT rain rate estimates using Advanced Scatterometer (ASCAT) soil moisture observations. Subsequently, the same observations were scaled to root-zone level using the Soil Moisture Analytical Relationship and assimilated into the Soil and Water Assessment Tool model through the ensemble Kalman filter (EnKF) technique. The correction to the 3B42RT rainfall was evaluated using observed rainfall data, whereas the modelled streamflow and soil moisture were assessed under three correction schemes: sole rainfall correction (forcing correction), sole soil moisture assimilation (state correction), and combined forcing and state correction (dual correction). The results demonstrated that the ANN-based rainfall correction technique improved the 3B42RT rainfall, with an average reduction in RMSE of 7.5 mm and a 10% improvement in NSE. The streamflow evaluation revealed that the forcing correction primarily enhanced the quick-flow component of simulated streamflow, with an assimilation efficiency of 17.3%, whereas the state correction scheme improved the base-flow component (assimilation efficiency of 21.9%). The dual correction combined the benefits of both schemes to achieve an assimilation efficiency of 28.9% and an NSE improvement of 0.274 over the open-loop simulation. The forecasting performance indicated that the dual correction strategy provided maximum improvement of up to two-three lead days in the selected Wyra catchment, with an NSE increase of 0.16 compared to the open-loop forecast. Overall, the dual correction strategy based on ANN and the EnKF promotes the use of satellite-based rainfall and soil moisture data for hydrological applications.

Keywords: Data Assimilation, Advanced Scatterometer, Artificial Neural Network, Ensemble Kalman Filter, Soil and Water Assessment Tool, Soil Moisture Analytical Relationship

*Corresponding Author

1 Introduction

Accurate streamflow forecasting is essential for a variety of water resource-related applications, including flood warning and mitigation [55], disaster management [56], climate change impact [62, 64], and irrigation practices [21]. Over the years, various improvements have been made in the hydrological domain, enabling models to mimic real-world scenarios using appropriate physics-based methods. However, accurate streamflow forecasting remains challenging due to the unavailability of good-quality input datasets and assumptions or limitations in model physics [3].

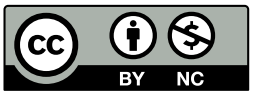
According to [10], one of the key unresolved challenges in hydrology is identifying and minimising the various sources of errors in hydrological model simulations. These errors can be broadly classified into three categories: systematic and random errors in forcing variables, incorrect assumptions within the underlying model structure, and improper model parametrisation [54]. These errors collectively propagate, resulting in uncertain model outputs. Uncertainties in model outputs are a greater concern for ungauged catchments because the only available data for modelling these catchments comes from satellite-based products or global (or regional) weather forecasts, both of which possess errors. A rational approach could be to identify useful information from all available datasets and integrate it optimally for land surface and hydrological modelling.

In this regard, data assimilation has emerged as a promising method of integrating models with real-world observations. Data assimilation approaches account for modelling errors by constraining the model states and parameters to achieve the Best Linear Unbiased Estimate (BLUE) with observations, thus helping to improve model predictions. The application of data assimilation in the hydrological field is well-studied and utilised [11, 15, 27]. Ensemble-based approaches, particularly the Ensemble Kalman Filter (EnKF), have demonstrated their ability to handle non-linear hydrological models [58, 70].

Among satellite-based meteorological fields, rainfall estimations are available on a near real-time basis with spatial resolutions ranging from $0.1^\circ \times 0.1^\circ$ to $0.25^\circ \times 0.25^\circ$ and high temporal resolution ($\sim$ 3 hours) [30, 31]. However, they are often associated with errors, including underestimation of light rainfall, timing mismatches, and false detection rates [22, 38, 71]. Moreover, their performance tends to degrade at shorter time scales (e.g. sub-daily), due to sampling limitations, insufficient temporal sampling by microwave sensors, and smoothing effects inherent in retrieval algorithms [53, 69]. In addition, topographic or microclimate influences, and sparse gauge networks hamper bias correction [59]. In many regions, the density of rain gauges is insufficient to resolve within-pixel heterogeneity, making gauge-based correction spatially uneven.

Similar to rainfall, soil moisture is a key control variable in hydrological modeling controlling infiltration, partitioning of rainfall into runoff or soil storage, and subsequently influencing hydrologic response. Therefore, accurate characterisation of soil moisture is essential for realistic hydrologic predictions. To improve model state estimation, ground-based soil moisture observations have been assimilated into models using a variety of data assimilation techniques (e.g. ensemble Kalman filters, particle filters) [40, 46]. However, ground sensors are only available at discrete locations, with limited spatial representativeness, and are often unavailable in remote or ungauged regions [12].

E-mail addresses: visweshwaran.ramesh@gmail.com; ramsankaran@civil.iitb.ac.in
doi:10.5149/ARC-GR.1662



Satellite-based soil moisture, on the other hand, provides globally available, near-daily hydrological observation that has shown a strong correlation with in-situ measurements [2, 61]. Over the past two decades, satellite-based soil moisture data assimilation has been extensively used to improve hydrological simulations [44, 65, 66]. Nevertheless, the satellite soil moisture product is also subject to a few limitations. They provide information at surface or near-surface level upto few centimeters of soil depth leading to poor representation of deeper hydrologic states [42]. To effectively utilise the available surface soil moisture information, they are often translated into subsurface through several translation techniques. the exponential filter [67], linear relationship [2], and Soil Moisture Analytical Relationship (SMAR) [42]. Past studies have shown that SMAR-based estimations, along with EnKF assimilation, have the potential to improve streamflow simulations [8, 52]. Therefore, the SMAR technique was also applied in the current study.

Given the uncertainties in near-real-time satellite-based rainfall products and the independently available satellite-based soil moisture information, few studies [13, 16, 17] have used soil moisture as proxy information to estimate or correct satellite-based rainfall accumulations. Among these, [13] developed an algorithm, SM2RAIN, to estimate rainfall accumulations based on changes in satellite-based soil moisture information. These rainfall estimates can then be combined with satellite-based rainfall to minimise overall errors [45].

In a separate effort, [16, 17] developed the Soil Moisture Analysis Rainfall Tool (SMART) algorithm to directly correct antecedent rainfall using satellite-based soil moisture within a data assimilation framework. This method was employed by [4] in a semi-distributed hydrological modelling scheme for dual rainfall and soil moisture correction, demonstrating promising potential for using satellite-based datasets in ungauged catchments. One difficulty in inferring rainfall from soil moisture is that evapotranspiration and surface runoff are not accounted for. In practice, a given change in soil moisture can result from many different combinations of rainfall, evapotranspiration, surface runoff and percolation, so there is no one-to-one mapping from soil moisture change to rainfall. The relationship is also highly non-linear and time-dependent since it depends on antecedent soil moisture conditions, rainfall intensity, soil properties and vegetation phenology. In addition, satellite products sense only the upper few centimetres of the soil column, introducing depth mismatches with catchment rainfall and modelled profile moisture. Therefore, different strategies should be explored for capturing the dynamic relationship between changes in soil moisture and antecedent precipitation.

In the past, studies have demonstrated the applicability of data-driven approaches to solving highly non-linear water resources problems [8, 35, 36]. These approaches have shown promising results in modelling antecedent rainfall using changes in surface soil moisture observations. While various sophisticated data-driven methods exist, artificial neural networks (ANNs) with simple structure have shown positive results in improving hydrological predictions [14, 53]. As a result, ANN and nudging-based approach is proposed in this work for correcting satellite-based rainfall observations using satellite-based soil moisture products.

In this study, we expand upon the dual correction framework using satellite soil moisture information, as initially proposed by [17] and later eloborated by [4] for flood applications, by applying it over a unique semi-arid catchment in southern India. This region is characterised by a highly seasonal rainfall pattern concentrated over a four-month period due to the southeast Indian summer monsoon, often resulting in flash flooding. For modelling purposes, the well-established physics-based Soil and Water Assessment Tool (SWAT) hydrological model was employed. Real-time rainfall estimates from the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) were used as meteorological forcings.

In this context, the paper addresses two research questions:

1. How effective is a data-driven approach in reducing errors in satellite-based rainfall products using satellite-based soil moisture?

2. To what extent can streamflow and soil moisture simulations be improved by a hybrid ANN–EnKF assimilation framework that corrects both satellite rainfall and modelled soil moisture states?

To address these questions, three correction schemes were evaluated: (1) forcing correction, (2) state correction, and (3) dual correction of both rainfall and model-simulated soil moisture. The Advanced Scatterometer (ASCAT) [68] Level 2 (H113) soil moisture product was utilised for these correction schemes.

The remaining sections of the paper are structured as follows: Section 2 describes the study area and datasets used. Section 3 outlines the methodological framework. Section 4 presents the results. Finally, Section 5 summarises the conclusions.

2 Study Area and Datasets

The Wyra River catchment, located in the eastern part of the Krishna River basin in India, was chosen as the study region (Figure 1). The study area is situated on the Deccan Plateau in southern India, with a catchment area of 1650 km^2, extending up to the Madhira gauging station. The Wyra River catchment has diverse topography, with altitudes ranging from 49 m to 783 m above mean sea level. The catchment received an average annual rainfall of 1089 mm during the study period, with a minimum of 650 mm in 2007 and a maximum of 1939 mm in 2010. The southwest monsoon, occurring from June through September, accounts for 74% of the total rainfall. The most common soil types are clayey loam and clay, with agriculture accounting for 70% of the land area, followed by dense forest (14%).

Table 1 provides details of the forcing data, land use land cover, digital elevation model, soil moisture observations, and streamflow gauging data used in this study. The TRMM Microwave Imager (TMI) is a passive microwave (PMW) sensor that serves as the foundation for a comprehensive understanding of the global water cycle. As a result, it has established a clear standard among other PMW sensors, and its error variation can be considered reasonable [8]. Therefore, the TRMM rainfall dataset was selected from among the available satellite-based rainfall products for the study period (2008 to 2012).

The TMPA 3B42RT rainfall product, a variant of the TRMM mission (hereinafter referred to as 3B42RT), was used as input rainfall for all three correction procedures in this study. As illustrated in Figure 1, the 3B42RT rainfall grid location was shifted by 0.125° relative to the India Meteorological Department (IMD) rainfall grid. Consequently, the TRMM rainfall was interpolated to reflect rainfall at all six IMD rain grids using the area-weighted averaging method.

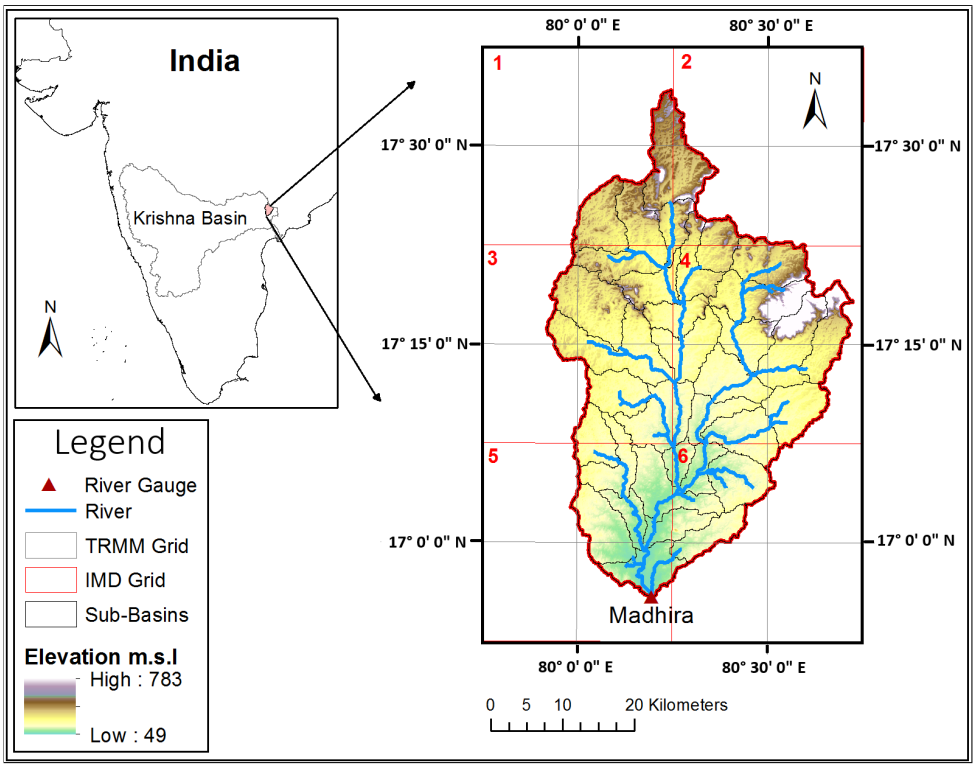


Figure 1: The location and spatial extent of the Wyra river catchment in the Krishna river basin, as well as elevation and river gauge data, are shown. The IMD and TRMM rainfall observation grids are depicted, with the IMD rain grid numbered from 1 to 6 (the elevation data was derived from 90-meter SRTM GDEM ([32]).

For data assimilation, ASCAT surface soil moisture observations were used. The acquired Level 2 product was normalised to a 40-degree incidence angle and then rescaled from 0 to 100 % to represent the soil saturation index. Along with the data, the noise level in the observations ('sm noise') was obtained, which aids in determining observational errors during assimilation. Table 1 lists the data sources, spatial resolutions, and purposes of the datasets used in this study.

3 Methodology

A description of the hydrological model setup, model calibration, and validation is provided in Section 3.1. Sections 3.2–3.4 describe the approaches used for forcing correction, state correction, and dual correction schemes, respectively. Relevant information about the use of ANN for rainfall estimation and the approach for integrating estimated rainfall with satellite-based rainfall accumulations is presented in the forcing correction scheme (Section 3.2). The implementation of the SMAR-EnKF technique is explained in the state correction scheme (Section 3.3). Finally, Section 3.4 describes the proposed dual correction scheme, which combines the forcing and state correction. Figures 2(a-c) show the flowcharts for the forcing, state, and dual correction schemes, respectively.

3.1 Model Setup and Parameterisation

The present study employs the Soil and Water Assessment Tool hydrological model. SWAT is a physics-based, semi-distributed modelling approach that efficiently simulates soil mois-

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

Data Category	**Variable**	**Resolution**	**Source**	**Remarks**
Thematic Data	Land Use	1:250,000	NRSC [49]	Derived from AWiFS optical data
	Soil	1:5,000,000	FAO HWSD V1.2 [25]	Prepared from soil survey datasets
	Topography	90 m	SRTM GDEM [32]	Interferometric SAR product
Forcing Variables	Rainfall	0.25° × 0.25°	IMD [50]	Interpolated gauge data
	Rainfall	0.25° × 0.25°	TMPA-3B42RT [30]	Active & passive microwave remote sensing
	Temperature	1° × 1°	IMD2 [60]	Interpolated gauge data
	Humidity, Wind Speed, Radiation	0.25° × 0.25°	NCEP–CFSR [57]	Reanalysis
State Variable	Soil Moisture	0.25° × 0.25°	ASCAT [6]	Active microwave remote sensing
Output	Streamflow	–	CWC India [18]	Observed river discharge data

Note: NRSC – National Remote Sensing Centre; AWiFS – Advanced Wide Field Sensor; FAO – Food and Agriculture Organization; HWSD – Harmonized World Soil Database; SRTM – Shuttle Radar Topography Mission; GDEM – Global Digital Elevation Model; IMD – India Meteorological Department; TMPA – Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis; 3B42RT – TRMM 3B42 Real-Time; IMD2 – India Meteorological Department, Version 2; NCEP – National Centers for Environmental Prediction; CFSR – Climate Forecast System Reanalysis; ASCAT – Advanced Scatterometer; CWC – Central Water Commission.

ture, evapotranspiration, streamflow, sediment flow, and the concentration levels of various nutrients in the soil at different temporal scales (daily, monthly, and annual) [7]. Previous studies have demonstrated that assimilating observations into the SWAT model improved model performance considerably [29, 51]. It should be noted that we have not considered the irrigation practices and crop phenology during SWAT simulation which can be a key source of error during simulations. A detailed description of the model is well documented in the literature [5, 48] and is therefore not discussed here.

For the current study, the Wyra catchment was divided into 33 sub-basins, with areas varying from 20 km^2 to 100 km^2. The SWAT model was calibrated using a two-stage workflow. First, following the rule-of-thumb procedures for adjusting highly sensitive parameters in SWAT as outlined by [5], a manual pre-calibration was performed to narrow the parameter search space to physically plausible ranges. Subsequently, the Sequential Uncertainty Fitting algorithm (SUFI-2) [1] was applied. This method performs a global parameter search while quantifying prediction uncertainty using the 95% prediction uncertainty (95PPU), defined by the P-factor (coverage) and R-factor (bandwidth).

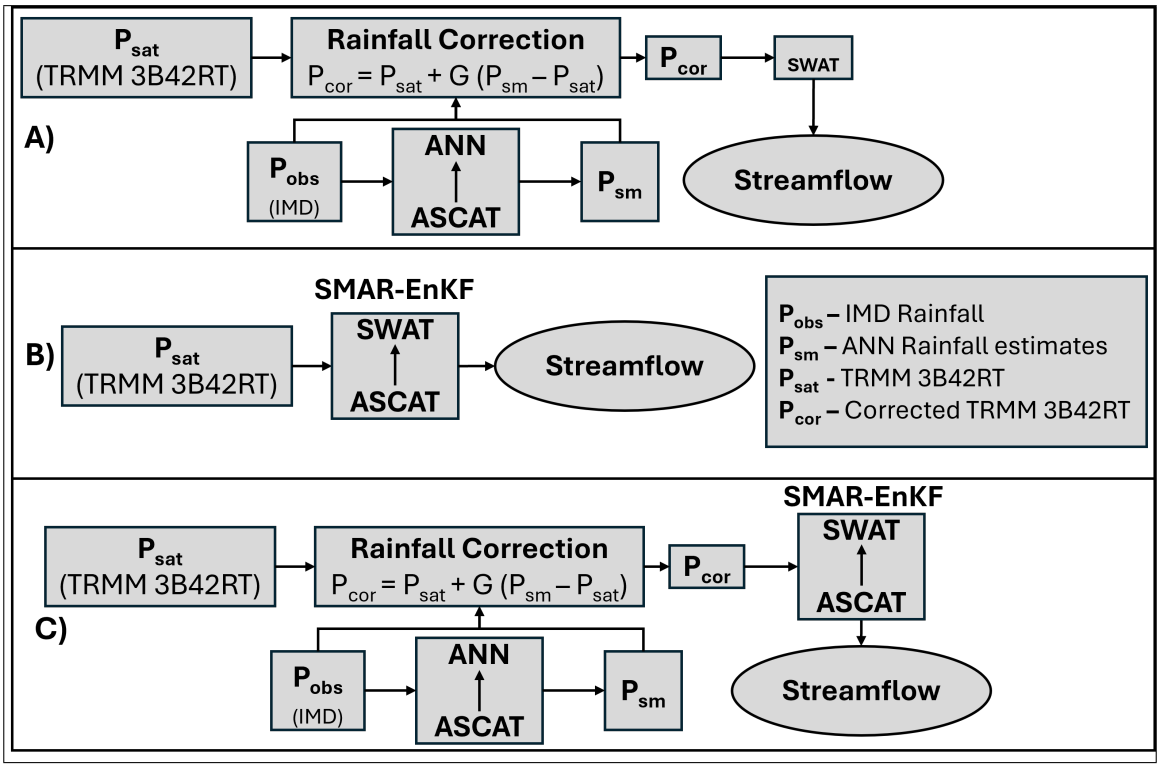


Figure 2: Flowchart of the three assimilation experiments (a: forcing correction, b: state correction, and c: dual correction) conducted in the present study.

The optimisation of model parameters was carried out with the objective of maximising the Nash-Sutcliffe Efficiency (NSE) of the simulated streamflow. Calibration was performed using gridded rainfall data from the IMD (Table 1). Further details regarding the selection and calibration of the 13 sensitive model parameters are available in previous studies [51, 52].

Following calibration, model validation was independently performed by forcing the IMD and 3B42RT rainfall datasets from June 2010 to May 2012, with a 2-years 5-months spin-up from January 2008. It should be noted that the model run during the validation period with 3B42RT rainfall forcing was considered an open-loop model run.

3.2 Forcing Correction

In the forcing correction experiment, the satellite-based rainfall was corrected using ASCAT-retrieved soil moisture observations. An overall schematic illustration of the proposed forcing correction scheme is shown in Figure 2(a). This study investigated a data-driven approach based on an artificial neural network to estimate rainfall accumulations from soil moisture variations. The details of ANN implementation are discussed in Section 3.2.1. Once the rainfall was estimated from soil moisture variations, it was used to correct the 3B42RT dataset using the Newtonian nudging assimilation technique [44]. A description of this correction approach is provided in Section 3.2.2. The calibration of the ANN and rainfall correction technique was performed using 3B42RT rainfall and ASCAT surface soil moisture data for the years 2007, 2008, 2009, and 2015. Likewise, from January 2010 to May 2012, validation was carried out alongside the data assimilation experiments in this study.

3.2.1 Predicting Rainfall from Satellite-Based Surface Soil Moisture Using ANN

(a) Soil Moisture Data Processing : The ASCAT soil moisture products were provided in grid format. To match them with each SWAT-simulated sub-basin, an area-weighted average of the gridded soil moisture was performed. Subsequently, the soil moisture values were normalised between 0 and 1 by considering the long-term maximum and minimum soil moisture values. Using the normalised soil moisture, the change in soil moisture was computed on a daily basis. Given the non-linear relationship between rainfall and soil moisture, the soil moisture value on a given day—along with the change in soil moisture—was used to predict the rainfall accumulation for that day.

(b) Artificial Neural Network (ANN) : The ANN structure selected in this study follows a conventional design widely used in various studies including [34, 35]. It is also known as a feed-forward neural network or multilayer perceptron. An ANN consists of several hidden layers, along with an input and an output layer. While the input and output layers can be directly accessed, the hidden layers remain inaccessible. Each layer consists of multiple neurons. The output from a hidden layer neuron, ϵ_j, is given as:

$$\epsilon_j = f\left(\sum_{i=1}^{n} w_{ji}x_i + b_i\right) \tag{1}$$

where w_{ji} and b_i represent the weights and biases for a given neuron, and f is the non-linear activation function of the neuron. The final output from the entire network, γ, is given as:

$$\gamma = f\left(\sum_{j=1}^{m} w_{kj}\epsilon_j + b_o\right) \tag{2}$$

Here, w_{kj} and b_o represent the weights and bias of the output layer neuron. The network parameters are calibrated using a set of training data, $D = \{x_n, t_n\}$, by minimising the mean square error (MSE) between predictions and observations, given as:

$$E = \frac{1}{N}\sum_{n=1}^{N}(t_n - y_n)^2 \tag{3}$$

To minimise the average errors across all layers, the backpropagation method iteratively modifies the weights and biases in the network, starting at the output layer. This process continues until the output error E falls within an acceptable limit.

The network architecture employed in this study consists of three hidden layers, each containing 15, 15, and 10 neurons, respectively. The input vector x_n consists of two variables, namely, observed soil moisture (smt) and the daily change in soil moisture ($\text{smt}_{\text{ASCAT},t} - \text{smt}_{\text{ASCAT},t-1}$). Furthermore, it was assumed that a negative change in soil moisture resulted from a lack of rainfall. Hence, the algorithm was trained only on days with a positive change in soil moisture.

The ANN was trained using the Levenberg-Marquardt (LM) backpropagation algorithm [20]. The IMD gridded daily rainfall product [50] was used as observational data for training the neural network. Of the entire training dataset (2007–2009 and 2015), 60% was used for training the network, while the remaining 40% was used for cross-validation. Testing of the trained network was carried out using independent data from 2010 to 2012.

3.2.2 3B42RT Rainfall Correction

Once the rainfall was estimated using ASCAT soil moisture and ANN, the 3B42RT product was corrected using the ANN-predicted rainfall. For this purpose, the Newtonian nudging scheme was applied, following [44]:

$$P_{\mathrm{cor}}(t) = P_{\mathrm{TMPA}}(t) + G\left[P_{\mathrm{SM}}(t) - P_{\mathrm{TMPA}}(t)\right] \tag{4}$$

Here, $P_{\mathrm{cor}}(t)$ is the corrected rainfall product, and $P_{\mathrm{SM}}(t)$ represents rainfall accumulation estimates derived from ASCAT soil moisture and ANN. The weighting parameter G was calibrated to maximise the correlation between the 3B42RT and IMD rainfall data. The calibration of G was performed for the years 2007 to 2009 and 2015, as stated earlier in Section 3.2.

3.3 State Correction Scheme

In the state correction scheme, ASCAT soil moisture retrievals were assimilated into an open-loop SWAT run. This run used uncorrected 3B42RT rainfall as its input. A schematic illustration of the state correction strategy used in this study is shown in Figure 2(b). Details of the state correction scheme are discussed in the following subsections.

3.3.1 Soil Moisture Bias Correction

The inherent bias between the model-predicted variable and the corresponding observations is frequently overlooked by assimilation approaches. This issue is often referred to as 'bias-blind assimilation' [19, 39]. Such biases can result in suboptimal assimilation outcomes, particularly when using remote sensing observations that contain spatially variable biases. Therefore, the ASCAT soil moisture observations were bias-corrected using the mean-variance correction technique (Equation 5). This method matches the mean and variance of the model-predicted soil moisture on a monthly timescale. Although higher-order bias correction approaches (like CDF matching) are available, they tend to remove more inherent variability in the observations [41, 43]. Further, for a limited time span of two years, construction of CDF would be suboptimal, and adopting a simpler method has proved sufficient [41].

$$SM_{\mathrm{corrected}} = SM_{\mathrm{model}} + \frac{\sigma_{\mathrm{model}}}{\sigma_{\mathrm{ASCAT}}}(SM_{\mathrm{ASCAT}} - \overline{SM}_{\mathrm{ASCAT}}) \tag{5}$$

Here, $SM_{\mathrm{corrected}}$ is the bias-corrected soil moisture. SM_{model} is the mean SWAT-simulated soil moisture for a given month, and SM_{ASCAT} is the observed ASCAT soil moisture value. The terms σ_{model} and σ_{ASCAT} represent the standard deviations of the SWAT-simulated and observed soil moisture values for a given month, respectively.

3.3.2 Soil Moisture Analytical Relationship (SMAR)

The ASCAT soil moisture observation represents only the top few centimetres of the soil layer (0–5 cm). Hydrological models, on the other hand, simulate soil water content at much deeper levels (usually varying between 100 and 200 cm). The assimilation of top-layer soil moisture has shown limited success in improving soil moisture at deeper layers in previous studies [29, 51]. Other studies [12, 52] have shown that converting surface soil moisture to root-zone soil moisture before assimilation into hydrological models improves the accuracy of root-zone soil moisture and streamflow simulations.

Based on the previous work [52], the SMAR algorithm [42] was used in this study to convert ASCAT surface soil moisture (SSM) to root-zone soil moisture (RZSM) using the

model-simulated climatology of RZSM. SMAR estimates RZSM conditions based on a physical relationship with SSM conditions.

The equations governing this process are derived from the water balance between these two layers. Water infiltrates into RZSM only when the field capacity of SSM is reached [28]. However, horizontal/lateral water movement is typically neglected due to its insignificant contribution. The vertical water flux reaching the RZSM is summarised as:

$$y(t_j) = \begin{cases} s_1(t_j) - s_{\text{fc1}}, & \text{if } s_1(t_j) \geq s_{\text{fc1}} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Here, s_1 is the relative saturation of SSM, and s_{fc1} is the relative soil saturation of SSM at field capacity. The infiltration of water occurs instantaneously and recharges the RZSM based on:

$$s_2(t_j) = s_{\text{w2}} + \left[(s_2(t_{j-1}) - s_{\text{w2}})e^{-a(t_j - t_{j-1})}\right] + [(1 - s_{\text{w2}})by(t_j)(t_j - t_{j-1})] \tag{7}$$

Here, s_2 is the relative saturation of RZSM, and s_{w2} is the relative saturation of RZSM at the wilting point. The parameters a and b are estimated as:

$$a = \frac{v_2}{(1 - s_{\text{w2}})n_2 z_2}, \quad b = \frac{n_1 z_1}{(1 - s_{\text{w2}})n_2 z_2} \tag{8}$$

In these equations, v_2 (mm/day) is the soil water loss coefficient, which considers all water losses occurring at RZSM. The terms n_1 and n_2 represent soil porosity for SSM and RZSM, respectively, while z_1 (mm) and z_2 (mm) are the depths of SSM and RZSM.

The first parameter, a (day^{-1}), represents the water loss coefficient in RZSM. It is the ratio of water loss in RZSM to its storage. The second parameter, b, represents the diffusivity coefficient. It is based on the ratio of SSM storage to RZSM storage. Along with these two parameters, s_{w2} and s_{fc1} serve as the physical parameters of SMAR. These parameters can be measured using in-situ observations or estimated through an optimisation algorithm. In the current study, all four parameters were calibrated using the Particle Swarm Optimisation (PSO) algorithm for each SWAT sub-basin to maximise the NSE with respect to simulated RZSM.

3.3.3 Ensemble Kalman Filter (EnKF)

EnKF belongs to the family of Kalman filtering approaches proposed by [33]. Among its various variants, it is one of the most widely used sequential assimilation approaches in the hydrological community. It utilises the Monte Carlo approximation to create samples (also called ensemble members) in the distribution. EnKF was first introduced by [23] to handle non-linear models efficiently. It uses two steps to optimally ingest the observation into the model.

The first step is the 'forecasting step', where an ensemble of model simulations and observations is generated (Equation 9) to better represent their respective uncertainties. These ensembles are generated randomly by perturbing the model forcings and states:

$$X_{t+1}^{i,b} = M(X_t^{i,a}, U_t^i) + v_{t+1} \tag{9}$$

Here, $X_{t+1}^{i,b}$ represents the forecasted ensemble state. M is the non-linear SWAT model, v_{t+1} represents the Gaussian white noise of the model state, and t represents the timestep. Similarly, the ensemble generation of observations is given by:

$$Z_{t+1}^{i,b} = H_t X_t^{i,b} + w_{t+1} \tag{10}$$

Here, $Z_{t+1}^{i,b}$ represents the model-simulated observational counterpart. H_t is the observation operator that connects the observation to the model state, and w_{t+1} represents the observational error.

Once the ensemble is generated, the 'assimilation step' is performed based on the relative uncertainties in the model and observation. During this step, the observation is assimilated into the model using a nudging parameter called the Kalman gain (K), as shown in Equation 11:

$$X_{t+1}^{i,a} = X_{t+1}^{i,b} + K(Z_{t+1}^{i} - Z_{t+1}^{i,b}) \tag{11}$$

Here, $X_{t+1}^{i,a}$ represents the assimilated model state, denoted by superscript a. The K follows the best linear unbiased estimate principle to optimally adjust the observation. For a given timestep, it calculates the relative error rate in the model states with respect to the combined uncertainty in the model and observation as:

$$K = B_t H_t^T (R_t + H_t B_t H_t^T)^{-1} \tag{12}$$

In this equation, B_t represents the error covariance matrix of the model state before assimilation, and R_t represents the error covariance matrix of the observation.

3.3.4 Ensemble Simulations and SMAR-EnKF Implementation

For generating ensemble model simulations, spatially homogeneous and temporally uncorrelated multiplicative Gaussian noise with a zero mean and a standard deviation of 0.3 mm/mm was applied to the TRMM 3B42RT rainfall data. Similarly, Gaussian noise with a zero mean and standard deviations of 0.05 mm/mm and 0.10 mm/mm was added to the surface and subsurface soil moisture layers, respectively, during model simulations. This accounted for uncertainties due to model parameters. All ensemble simulations in this study were run using 100 members.

The ensemble simulations were initiated with a two-year warm-up phase to ensure an appropriate ensemble distribution of simulated soil moisture at the start of the assimilation cycle.

To represent errors in the ASCAT observations, the ASCAT soil moisture was perturbed using a zero-mean standard deviation of σ_{ascat}, given as:

$$\sigma_{\text{ascat}} = \alpha \text{ASCAT}_{\text{sm noise}} \tag{13}$$

Here, $\text{ASCAT}_{\text{sm noise}}$ represents the uncertainty information provided with Level 2 ASCAT soil moisture observations, which range from 0.01 to 0.06 in this study region. The scaling factor α was set to 3 to achieve an average σ_{ascat} of 0.1 mm/mm.

To ensure an efficient ensemble spread, the Latin hypercube sampling technique was employed to generate random numbers for all forms of sampling [26]. To maintain vertical error correlation across soil layers, the field capacity of the soil was perturbed using Gaussian noise with a zero mean and a standard deviation of 0.1 mm/mm for both surface and subsurface soil layers. The same random numbers were used, as described by [51].

Furthermore, rainfall variables were considered semi-restricted, while soil moisture variables were treated as fully restricted fields. This was done to maintain the quality of ensemble simulations, as suggested by [63].

The SMAR-EnKF state estimation technique employed in this study is based on the work of [52]. The soil moisture of the surface and deeper soil layers was estimated individually during the state estimation run. First, EnKF was used to update the surface layer soil moisture using ASCAT observations and error information. The analysis from EnKF was

then utilised to estimate subsurface soil moisture for each ensemble member using the SMAR model. This subsurface soil moisture ensemble was subsequently assimilated into the SWAT subsurface soil layer using EnKF.

3.4 Dual Correction

The dual correction scheme adopted in this study combines the forcing correction scheme and the state correction scheme, described in Sections 3.2 and 3.3, respectively. A detailed flowchart of the proposed dual correction scheme is shown in Figure 2(c). In this scheme, ASCAT soil moisture retrievals were used to correct the 3B42RT rainfall. Furthermore, the scaled root-zone soil moisture, obtained using the SMAR technique, was assimilated into SWAT to update the soil moisture states.

We assessed the performance of each of the three assimilation techniques by validating the modelled streamflow during the validation period (June 2010 to May 2012). It should be noted that independent in situ observations of evapotranspiration and other water-balance components are not available for this basin (aside from the outlet discharge) during validation period. In light of these constraints, we added a qualitative internal-consistency assessment using the modelled soil moisture at two spatial scales: first, at the outlet location where the streamflow assessment is conducted, which allows a direct comparison between the modelled soil moisture and streamflow; and second, a basin-wide assessment across the catchment to capture spatial patterns and variability.

Correlation (R), Root Mean Square Error (RMSE), Absolute Bias (Bias), Nash–Sutcliffe Efficiency, and assimilation efficiency (EFF) statistics were used for this purpose. These metrics are widely applied in evaluating the performance of hydrological models and, collectively, they assess different aspects of model performance, including the magnitude of errors, systematic bias, and overall predictive skill. Additionally, to evaluate improvements in the 3B42RT estimates before and after correction, the Probability of Detection (POD) and False Alarm Ratio (FAR) statistics were used. These statistics are calculated as:

$$\mathrm{POD} = \frac{h}{h+m} \tag{14}$$

$$\mathrm{FAR} = \frac{f}{f+h} \tag{15}$$

Here, h, m, and f denote the number of hits, misses, and false rainfall events, respectively. Hits (h) refer to the number of instances where both predicted and reference products indicate rainfall occurrence. Misses (m) refer to the number of instances where the predicted rainfall fails to capture rainfall occurrences in the reference product. False alarms (f) refer to the number of instances where the predicted rainfall reports rainfall occurrence while no rainfall is recorded in the reference product.

Among others, the NSE metric was chosen as the objective function to optimise the model. This is due to its ability to normalise performance against the variability of observed data, although it is known to be sensitive to extreme values.

4 Results and Discussion

Enhancements to the 3B42RT product using the forcing correction technique are discussed in Section 4.1. The performance of the selected assimilation strategies on model-simulated soil moisture and streamflow is then discussed in Sections 4.2 and 4.3 respectively. Later, SWAT performance during forecast period is discussed in Section 4.4.

4.1 Rainfall Correction

Figures 3(a–e) present the validation metrics (R, RMSE, NSE, POD, and FAR) for 2010–2012. From Figure 3(a), it can be seen that Pearson's correlation at all grids improved slightly after applying the correction. In contrast, the RMSE and NSE of the corrected rainfall at all grids improved considerably compared to the raw 3B42RT rainfall product (Figures 3(b–c)). The improvement in R and the reduction in RMSE demonstrate that the ANN-based rainfall correction scheme successfully reduces random errors in 3B42RT rainfall accumulations. Similarly, the improvement in NSE indicates the ability of the proposed forcing correction scheme to reduce errors in moderate to high rainfall events, as NSE is more sensitive to higher values in a time series. The probability of detecting rainfall (Figure 3d) improved consistently across all rainfall grids. At the same instance, an increase in the False Alarm Ratio was observed across all grids (Figure 3e). A primary reason for this could be noise in the satellite-derived soil moisture estimates, which triggered rainfall events erroneously classified as rainy days in the $3B42RT_{corr}$ product. Similar issues have been reported in previous studies [4]. Fig-

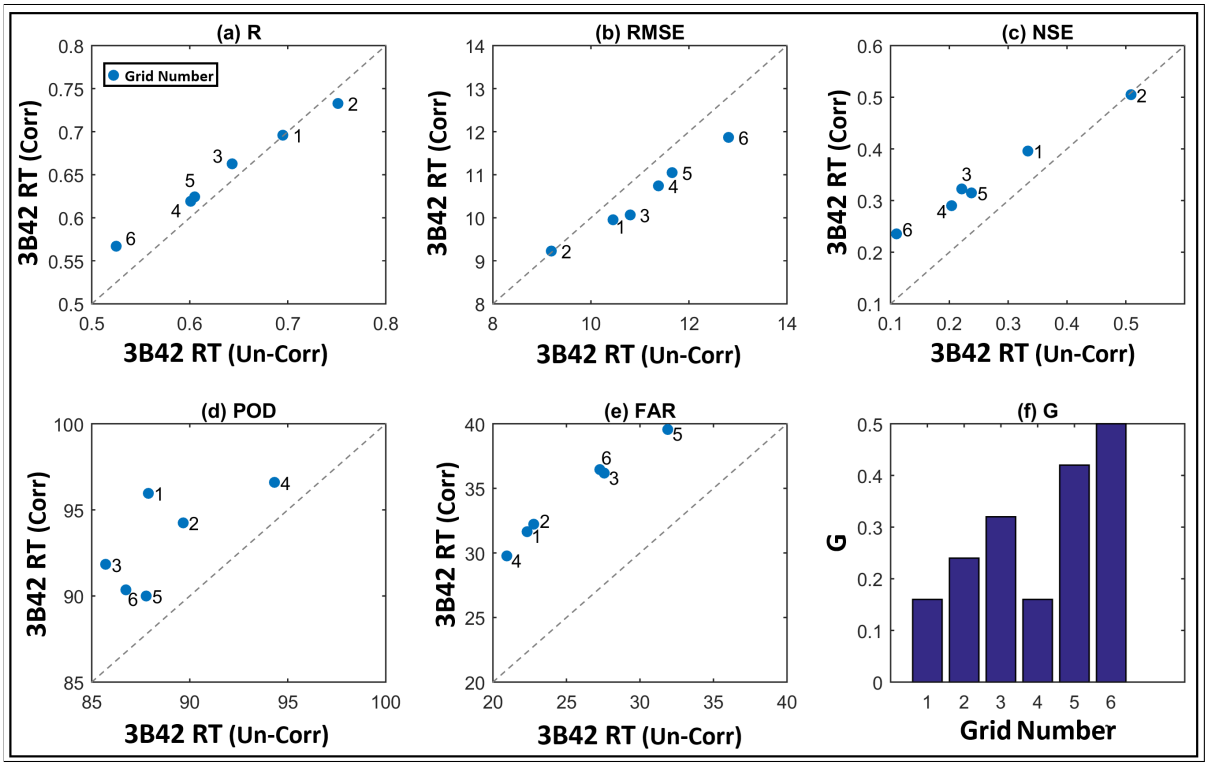


Figure 3: Statistical comparison of the uncorrected 3B42RT rainfall product versus the corrected rainfall (3B42RTcorr) with respect to IMD gridded rainfall data (reference dataset) at all six rainfall grids in the study area. (a) to (e) shows correlation (R), RMSE, NSE, POD, and FAR respectively. (f) shows the gain parameter (G) of the Newtonian nudging scheme used for correction in all six rainfall grids. Note: For Figures(a-e), the y-axis shows statistics of the corrected data and the x-axis shows the same statistics for uncorrected data.

ure 3(f) shows the calibrated gain parameter G at each of the six rainfall grids used in this study. The gain parameter G describes the relative weight assigned between the 3B42RT and ANN-derived rainfall estimates while correcting the 3B42RT datasets (Equation 11). The gain parameter G (Figure 3f) was optimised to be less than 0.5 at all rainfall grids. This implies that the nudging correction (Equation 11) gave more weight to the uncorrected TRMM 3B42RT rainfall estimates than the ANN-generated rain rate estimates in the selected study

region.

One noteworthy point is that the gain parameter in Equation 11 was calibrated to maximise correlation rather than NSE efficiency criteria, as was done for the SWAT and SMAR model calibration. This decision was made because using NSE as the objective function produced superior correction for large-magnitude rainfall accumulations (>30 mm/day) due to its sensitivity to extreme values. However, it resulted in an increase in small-magnitude unobserved rainfall events (≤2 mm/day), which were produced by ANN due to errors in ASCAT observations (results not shown here). This caused a significant degradation in correlation and FAR statistics of the corrected rainfall.

Despite using correlation as the objective function during the optimisation of parameter G, the degradation of FAR was not eliminated completely. It still produced an overall better rainfall correction across multiple evaluation criteria (Figures 3 a–d). Nonetheless, the improvement in RMSE (average decrease of 7.5 mm) and NSE (average increase of 10%) is substantial for the corrected rainfall product (P_{SM}), thereby instilling confidence in the proposed approach for improving real-time satellite-based rainfall estimates.

4.2 Soil Moisture Evaluation

Figure 4 displays the surface-layer soil moisture time series for the open-loop and other assimilation experiments. Visually, it is observed that the model captures the ASCAT observations well. Table 2 shows the same trend, with a correlation coefficient greater than 0.75 across all experiments.

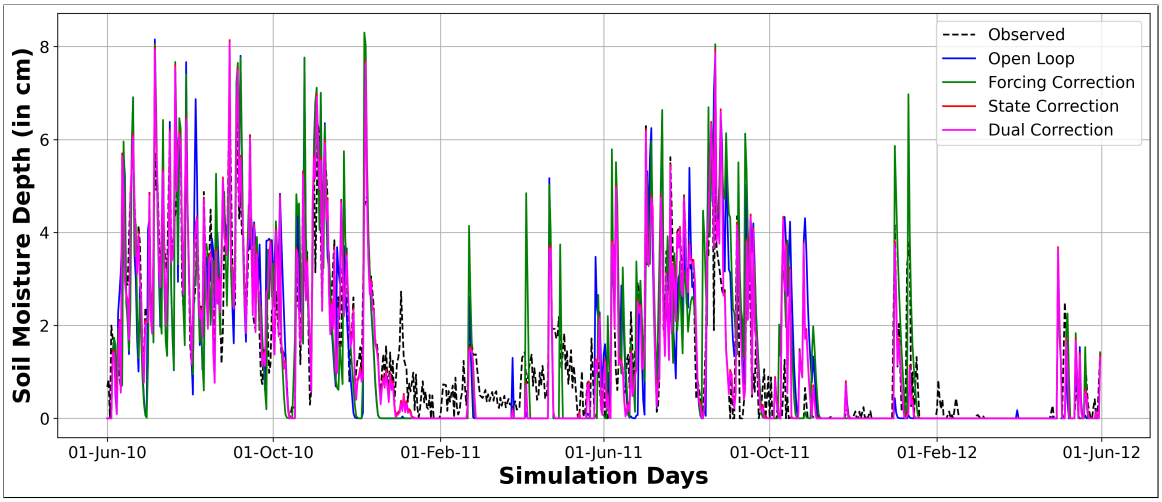


Figure 4: Comparison of model simulated soil moisture with ASCAT observation at basin outlet for all assimilation experiments.

Furthermore, the RMSE and bias are lower, indicating very good agreement with the observations (Table 2). The forcing correction does not improve the modelled soil moisture. This likely reflects the agriculture-intensive nature of the catchment, where irrigation, evapotranspiration, and soil–plant parameter uncertainties dominate the soil moisture dynamics, so precipitation correction alone yields minimal improvement. On the other hand, both state and dual corrections improve SWAT performance in simulating soil moisture. NSE values increased from 0.375 to 0.728 for the state correction and to 0.732 for the dual correction experiment. The assimilation efficiency score also shows an improvement of approximately 34% relative to the open-loop case. These results highlight the importance of a dual correction scheme for model improvement.

Table 2: Comparison of SWAT-simulated soil moisture performance indices with ASCAT observations for all assimilation experiment runs from June 2010 to May 2012.

Model Run	**R**	**RMSE**	**Bias**	**NSE**	**EFF**
	(-)	(Cm)	(Cm)		(%)
Open Loop	0.76	1.276	0.01	0.375	-
Forcing Correction	0.76	1.272	0.03	0.374	-0.03
State Correction	0.88	0.841	-0.028	0.728	34.11
Dual Correction	0.88	0.836	-0.056	0.732	34.52

Figure 5 shows box plots of all metrics across the 33 sub-basins for the experiments. Although forcing correction showed no improvement at the outlet (Figure 4), the box plots reveal a marginal improvement over the open-loop case across the catchment. The distributions are tighter, and the median NSE increases from 0.46 to 0.48, and the median bias is reduced from 0.03 to 0.01, indicating a positive effect of rainfall correction across the catchment. For the state and dual corrections, the median NSE and R improved to 0.72 and 0.88, respectively. Bias is close to zero (0.02) and RMSE is lower (0.97 cm) for both state and dual corrections, suggesting a strong impact of the dual correction on the SWAT simulations.

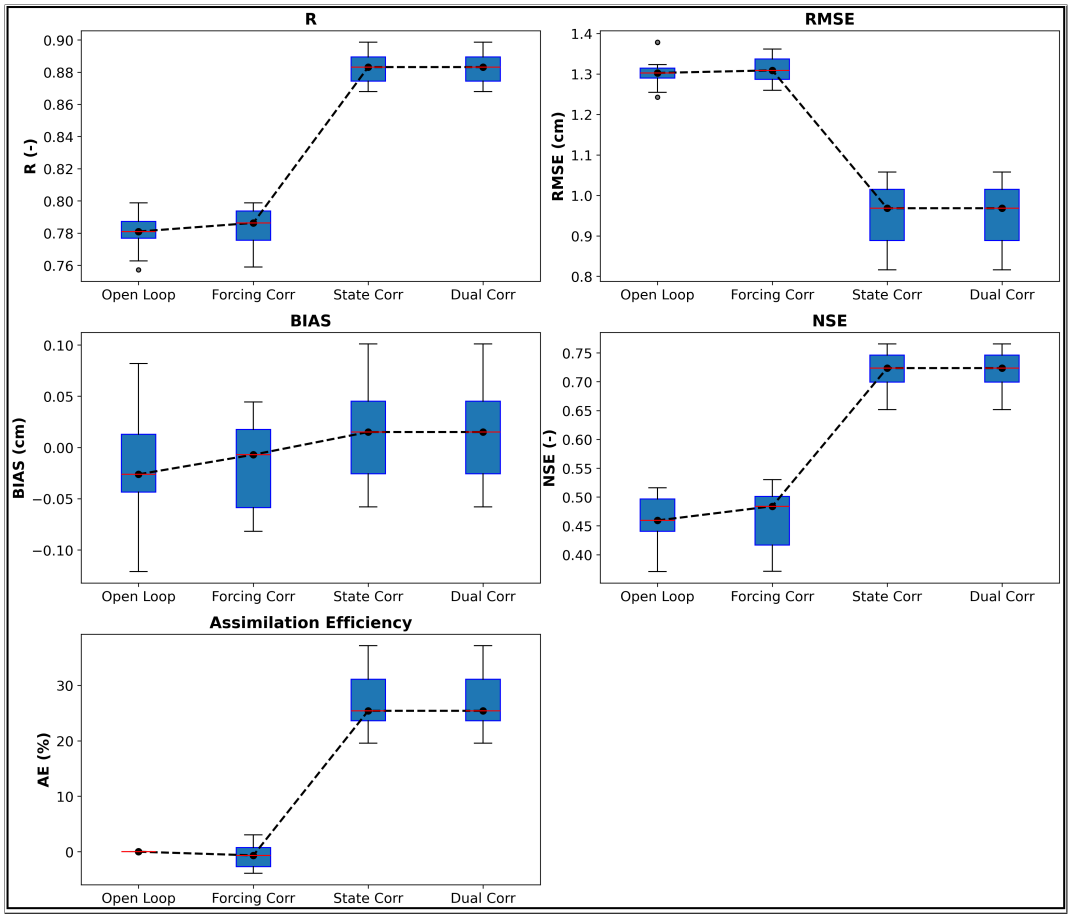


Figure 5: Boxplot showing evaluation metrics between SWAT simulated soil moisture and ASCAT observations across 33 sub-basins during different assimilation experiments.

4.3 Streamflow Evaluation

Figure 6 illustrates the comparison of streamflow data at the Madhira gauging station with simulated streamflow from all the experiments. From Figure 6, it was observed that the forcing correction scheme improved model performance over the open-loop run, particularly during high-flow conditions (as shown in the inset Figure 6a). On the other hand, the state correction and dual correction schemes were relatively less effective in improving streamflow during high-flow conditions. It was also observed that the state correction scheme, using the SMAR-EnKF strategy, suppressed peak flows during both the 2010 and 2011 monsoon (June–September) seasons.

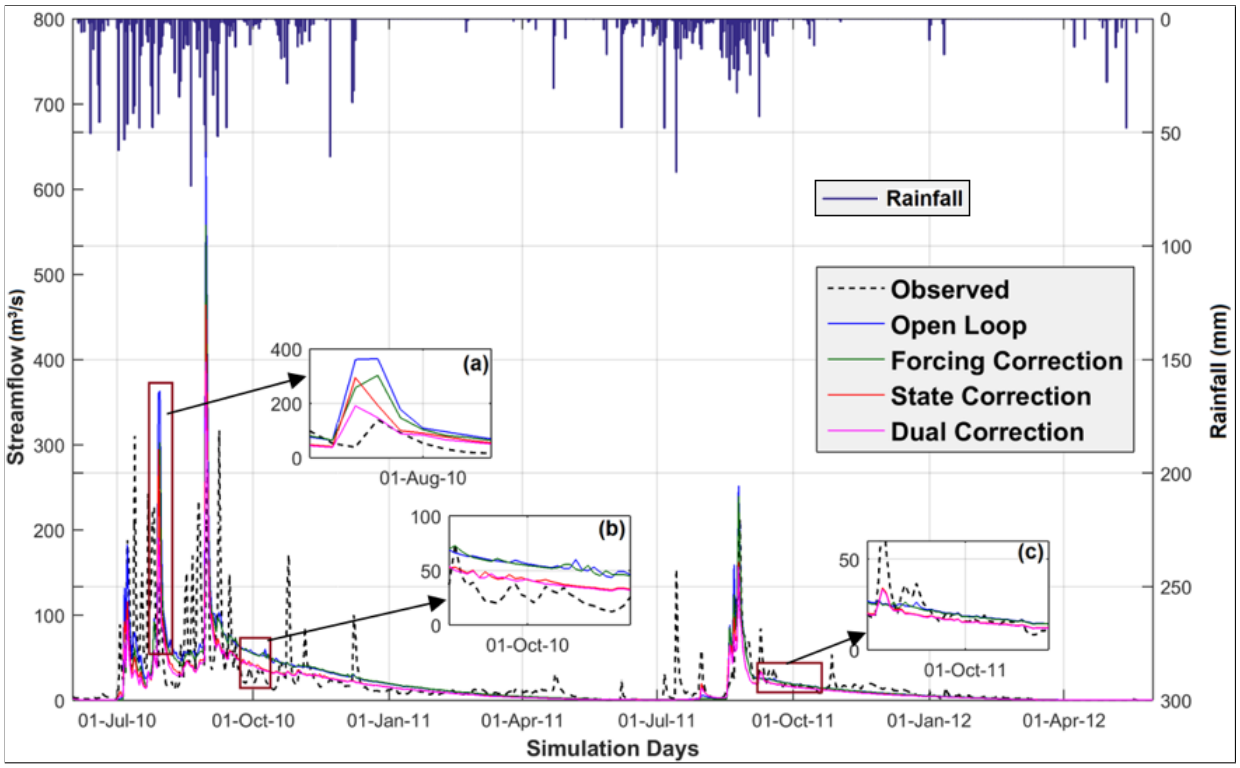


Figure 6: Comparison of model simulated streamflow with observed streamflow at basin outlet for all assimilation experiments.

For low-flow comparisons, different observations were made. The forcing correction did not improve low flows, as seen in Figures 6 (b–c). In contrast, both the state correction and dual correction schemes significantly enhanced low flows. A comparison of the performance statistics for all four model runs, along with a model run (IMD run) forced with reference rainfall (IMD gridded rainfall) without assimilation, is presented in Table 3. From this table, it can be observed that some negative bias was introduced during all assimilation experiments. This is acceptable because sequential data assimilation techniques do not account for model water balance, and hence, the model can exhibit deviations in long-term biases. However, the dual correction scheme outperformed the forcing correction and state correction schemes achieving the highest assimilation efficiency of 28.9%. The NSE obtained during the validation period when the model was driven using IMD rainfall was 0.3. According to [47], a streamflow NSE ≥ 0.5 is considered satisfactory at a monthly time scale, although lower values can be acceptable at shorter time scales (weekly or daily). Similarly, a NSE of 0.05 was obtained from model simulations utilising 3B42RT rainfall across the validation period. This poor performance was attributed to the fact that 3B42RT, being a real-time dataset, contains

significant errors in it [9].

In addition, the RMSE and NSE of simulated streamflow were better for the dual correction scheme, which utilised corrected TRMM rainfall and ASCAT soil moisture assimilation, compared to the model performance when driven by observed IMD rainfall without data assimilation. This level of improvement was not achieved when the forcing correction and state correction schemes were applied independently.

Table 3: Comparison of SWAT-simulated streamflow performance indices with observed streamflow for all model runs from June 2010 to May 2012.

Model Run	**R**	**RMSE**	**Bias**	**NSE**	**EFF**
	(-)	(Cumecs)	(Cumecs)		(%)
IMD	0.612	31.724	5.882	0.302	-
Open Loop	0.59	37.026	-0.01	0.05	-
Forcing Correction	0.6	33.682	-1.48	0.214	17.3
State Correction	0.578	32.719	-5.295	0.258	21.9
Dual Correction	0.61	31.227	-6.458	0.324	28.9

During high-flow conditions, the forcing correction tends to perform better because extreme rainfall events produce strong signals in soil moisture changes that the ANN can effectively capture and use to correct the satellite rainfall input. In contrast, the state correction using SMAR-EnKF underperforms in high flows because the model's internal states are pushed into regimes where the assumed physical relationships (e.g. linear response via SMAR) may be violated, leading to less reliable subsurface fluxes. Consequently, the state correction alone struggles to keep pace with the unmodelled dynamics during intense rainfall events. It was also evident that the dual correction scheme retained the superior performance observed during low flows with the state correction scheme while achieving the highest improvement in assimilation efficiency.

Additionally, the equivalent performance of the dual correction model run compared to the model run utilising observed IMD rainfall (Table 3) supports the application of satellite-based inputs for hydrological modelling in data-scarce catchments.

4.4 Forecasting Performance Evaluation

The long-term durability of the performance gains obtained from each assimilation strategy in the absence of satellite soil moisture data is analysed in this section. The forecasting skill of the SWAT hydrological model is evaluated for 0, 1, 3, 5, 10, and 15-day lead times, as presented in Table 4.

Table 4: Streamflow forecasting performance of the SWAT model during all three assimilation schemes.

Model Run	**Lead Time** (Days)	**RMSE** (Cumecs)	**Bias** (Cumecs)	**NSE**	**EFF** (%)
Forcing Correction	0	33.939	-1.507	0.210	17.2
	1	36.643	-1.111	0.079	3.5
	2	37.305	-0.894	0.045	0.0
	3	37.406	-0.731	0.040	-0.5
	5	37.476	-0.613	0.036	-0.9
	10	37.274	-0.482	0.047	0.2
	15	37.159	-0.454	0.053	0.8
State Correction	0	32.968	-5.380	0.254	21.9
	1	32.968	-5.380	0.254	21.9
	2	34.886	-4.781	0.165	12.6
	3	35.842	-4.381	0.119	7.7
	5	36.570	-3.934	0.083	3.9
	10	36.590	-3.385	0.081	3.8
	15	36.641	-3.030	0.079	3.5
Dual Correction	0	31.465	-6.560	0.321	28.9
	1	33.211	-6.229	0.243	20.8
	2	34.985	-5.416	0.160	12.1
	3	35.940	-4.934	0.114	7.2
	5	36.613	-4.370	0.080	3.7
	10	36.575	-3.772	0.082	3.9
	15	36.627	-3.363	0.080	3.6

During the forcing correction, it was observed that the assimilation efficiency declined immediately after day one. This was primarily because the corrected 3B42RT rainfall contributed mainly to improving the quick flow component of the streamflow (as discussed in Section 4.3), which had little or no effect on simulated streamflow after a one-day forecast.

During the state correction, the assimilation efficiency remained at 21.9% for one-day forecasts but declined rapidly to 3.9% by the five-day forecast.

For the dual correction scheme, the assimilation efficiency was highest during same-day forecasts ($\approx 28.9\%$). However, by the one-day forecast, EFF reduced to 20.8%, a value close to that obtained from the state correction scheme (Table 4). From the three-day forecast onward, the EFF of the dual correction scheme closely resembled that of the state correction scheme. Since the dual correction scheme integrates the benefits of both correction strategies, the highest assimilation efficiency was achieved only on the assimilation day.

It should be noted that the decay rate of assimilation efficiency in this study was higher than that reported in previous work by [51]. This could be attributed to two factors: 1. The present study used noisy satellite rainfall estimates, whereas the previous study used observed IMD rainfall. 2. The catchment area in this study was relatively smaller, resulting in a shorter concentration time.

As a result, maximum improvement in forecasting performance can be achieved for short

lead times of up to 1–2 days using the dual correction strategy. Beyond two days, the improvement in forecasted streamflow is comparable to that obtained from the state correction strategy. Comparisons with previous studies suggest that forecasting performance varies depending on catchment size, watershed characteristics, and the ratio of surface to subsurface runoff. To extend forecast skill beyond 1–2 days in this small basin, such as Wyra, future efforts should drive SWAT with short-range weather forecasts and maintain sequential updates of states and forcings throughout the forecast window.

5 Summary and Conclusions

This study assessed the capability of satellite-based soil moisture to correct both satellite rainfall estimates and modelled soil moisture states through data assimilation. Three correction strategies were tested: (i) forcing correction, where ASCAT soil moisture information was used to estimate rainfall accumulations via artificial neural networks that were then merged with 3B42RT using a Newtonian-nudging scheme; (ii) state correction, in which SMAR coupled with an Ensemble Kalman Filter (SMAR–EnKF) updated SWAT profile soil moisture; and (iii) dual correction, which combined both approaches to simultaneously correct satellite-based rainfall estimates and modelled soil moisture states using ASCAT observations.

The evaluation of the corrected 3B42RT rainfall demonstrated that the errors in the raw 3B42RT product can be effectively corrected using ASCAT soil moisture information, particularly for moderate to high rainfall events. [37] conducted a similar experiment to correct the 3B42RT product using support vector regression for a semi-arid catchment in India. Although the corrected product showed improvement for moderate rainfall values, its overall performance during high rainfall events remained poor, thereby highlighting the benefit of the ANN method for rainfall correction. The false alarm rate, on the other hand, increased slightly for the corrected rainfall, likely due to noise in the ASCAT soil moisture product used for the correction. Nevertheless, the application of ANNs in this study has demonstrated their potential for correcting satellite-based rainfall using satellite-based soil moisture information. The improvement in the performance metrics (Figure 3) reiterates that ANNs are capable of capturing complex non-linear relationships between soil moisture dynamics and antecedent rainfall. Future enhancements may involve the use of hybrid models by combining two machine learning methods [24] or by developing a more comprehensive ANN framework incorporating additional auxiliary information (including soil types and other driving variables) as predictors, to facilitate a more generalised correction approach.

Forcing correction had a marginal effect on the improvement of the modelled surface soil moisture. NSE and R rose by 5 and 0.7 % , respectively (Figure 5). On the other hand, the state-correction scheme improved the modelled soil moisture considerably, with an increase of 57.5 and 13.1 % in NSE and R, respectively. This difference is due to the agriculture-intensive nature of the catchment, which is heavily influenced by local practices and high evaporative demand. During streamflow evaluation, the forcing-correction scheme improved the accuracy of satellite-based rainfall estimates, leading to better high-flow simulations in the hydrological model, improving the NSE value by 328 per cent. In contrast, the SMAR–EnKF-based state-correction scheme had a stronger influence on subsurface runoff, improving the baseflow component in SWAT and leading to a 416 % improvement in NSE. [15] and [44] conducted similar experiments for alternate conceptual hydrological models and reported that state updates enhance baseflow while forcing correction benefits peak-flow simulation, reflecting a similar outcome to the current work.

Finally, the dual-correction scheme resulted in the best and most significant improvement in SWAT simulations. Surface soil moisture and streamflow saw NSE improvements of 57.5 and 548 % , respectively. In addition, the NSE achieved through the dual-correction strategy

slightly exceeded that of the model run forced with observed rainfall (IMD dataset). In the streamflow forecast analysis, the improvements from the forcing-correction scheme did not persist beyond 1–2 days of lead time. Conversely, the state- and dual-correction schemes maintained skill improvements for up to five days. The highest forecasting skill was observed during the dual-correction strategy for lead times of 1–2 days, with improvements of 20.8 and 12.1 % in the EFF metric. These results suggest that satellite soil moisture observations can be used effectively to constrain both model forcings and states, thereby enhancing model forecasts.

[4] evaluated dual forcing and state correction in data-scarce catchments and reported sustained improvements in forecasts at short lead times, echoing our findings. Additionally, we conducted a similar study on the same basin to assess state correction using the SMAR-EnKF approach on SWAT, with an elongated evaluation period [52]. From that work, it was evident that the improvement in streamflow simulation was limited when only the model state was updated. For instance, the EFF for the best-performing SMAR-EnKF scenario was 24.5 % (Table 4, [52]), which is notably lower than the EFF achieved through dual correction in this study (28.9 % , Table 3). Taken together, these results strongly support the broader adoption of dual correction strategies in operational flood forecasting and short-term water allocation planning.

During the forecast period, we acknowledge mismatches between the model and observations, particularly beyond 5 days' lead time. A likely contributor is the reliance on a single soil-moisture sensor (ASCAT) and a single satellite rainfall product (3B42RT). The performance of the forcing correction scheme is also constrained by the simplified representation of irrigation and crop phenology. This is an important limitation in the Wyra catchment, where ≈70 % of land use is agricultural. In light of these factors, future work should test multi-sensor assimilation (e.g., SMAP/Sentinel-1 soil moisture, gauge-merged rainfall). In addition, evapotranspiration information should be assimilated alongside soil moisture within a multivariate framework to improve water-balance closure.

Despite these limitations, the proposed hybrid ANN–EnKF framework allows for correction of both forcing and model state errors using satellite-derived information. It demonstrates a balanced combination of physics-informed modeling and data-driven adjustment. In addition, this approach is dynamic, and suitable for application in data-scarce environments supporting flood-forecasting applications.

Acknowledgements

The authors would like to thank the India Meteorological Department (IMD) for providing the rainfall data used in this study and the National Center for Environmental Prediction (NCEP) for supplying the humidity and wind speed data. We also acknowledge the National Remote Sensing Centre (NRSC) for the soil and land-use thematic data, the Food and Agriculture Organization (FAO) for global soil information, and the SRTM-GDEM for elevation data. The authors further acknowledge EUMETSAT for providing the ASCAT Level-2 soil moisture product and the Central Water Commission (CWC), Government of India, for making available the gauge-based streamflow observations. We also thank the two anonymous reviewers and the editor, Dr.Tamlin Pavelsky, for their constructive comments and time, which helped improve the quality of the paper.

Code Availability

The code used for the ensemble simulations in SWAT and the SMAR–EnKF implementation for the data assimilation experiments is available at: https://github.com/amolpatil771/DualCorrection

Data Availability

The meteorological datasets used in this study are publicly available from the India Meteorological Department (https://mausam.imd.gov.in) and the IMD Hydromet CRIS portal (https://hydro.imd.gov.in/hydrometweb/). Soil and land-use thematic datasets were obtained from the NRSC Bhuvan portal (https://bhuvan-app3.nrsc.gov.in/data/download/index.php), while global soil information was accessed through the FAO Soils Portal (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/en/). Elevation data were sourced from the SRTM-GDEM product (https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003). The ASCAT Level-2 soil moisture product is available from EUMETSAT (https://doi.org/10.15770/EUM_SAF_H_0004), and streamflow observations were obtained from the Central Water Commission of India (https://indiawris.gov.in/wris). The SWAT input files and SWAT-simulated soil moisture and streamflow outputs generated as part of this study are also provided in the same repository: https://github.com/amolpatil771/DualCorrection

Author Contributions

AP and RR devised the experiments. AP and VR ran the experiments. AP prepared the initial manuscript. VR and RR edited the manuscript. RR supervised the work.

References

[1] K. C. Abbaspour, C. Johnson, and M. T. Van Genuchten. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose zone journal*, 3(4):1340–1352, 2004. URL https://doi.org/10.2113/3.4.1340.

[2] C. Albergel, C. Rüdiger, D. Carrer, J.-C. Calvet, N. Fritz, V. Naeimi, Z. Bartalis, and S. Hasenauer. An evaluation of ascat surface soil moisture products with in-situ observations in southwestern france. *Hydrology and Earth System Sciences*, 13(2):115–124, 2009. URL https://doi.org/10.5194/hess-13-115-2009.

[3] R. Alvarado-Montero, D. Schwanenberg, P. Krahe, P. Helmke, and B. Klein. Multiparametric variational data assimilation for hydrological forecasting. *Advances in Water Resources*, 110:182–192, 2017. ISSN 0309-1708. URL https://doi.org/10.1016/j.advwatres.2017.09.026.

[4] C. Alvarez-Garreton, D. Ryu, A. W. Western, W. T. Crow, C. H. Su, and D. R. Robertson. Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments. *Water Resources Research*, 52:5357–5375, 2016. URL https://doi.org/10.1002/2015WR018429.

[5] J. G. Arnold, D. N. Moriasi, P. W. Gassman, K. C. Abbaspour, M. J. White, R. Srinivasan, C. Santhi, R. D. Harmel, A. Van Griensven, M. W. Van Liew, N. Kannan, and

M. K. Jha. Swat: Model use, calibration, and validation. *Transactions of the ASABE*, 55:1491–1508, 2012. URL https://doi.org/10.13031/2013.42256.

[6] Z. Bartalis, W. Wagner, V. Naeimi, S. Hasenauer, K. Scipal, H. Bonekamp, J. Figa, and C. Anderson. Initial soil moisture retrievals from the metop-a advanced scatterometer (ascat). *Geophysical Research Letters*, 34(20), 2007. URL https://doi.org/10.1029/2007GL031088.

[7] S. Behera and R. K. Panda. Evaluation of management alternatives for an agricultural watershed in a sub-humid subtropical region using a physical process based model. *Agriculture, Ecosystems Environment*, 113:62–72, 2006. URL https://doi.org/10.1016/j.agee.2005.08.032.

[8] M. A. E. Bhuiyan, E. N. Anagnostou, and P. E. Kirstetter. A nonparametric statistical technique for modeling overland tmi (2a12) rainfall retrieval error. *IEEE Geoscience and Remote Sensing Letters*, 14:1898–1902, 2017. URL https://doi.org/10.1109/LGRS.2017.2728658.

[9] M. M. Bitew and M. Gebremichael. Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research*, 47(6), 2011. URL https://doi.org/10.1029/2010WR009917.

[10] G. Blöschl, M. F. Bierkens, A. Chambel, C. Cudennec, G. Destouni, A. Fiori, J. W. Kirchner, J. J. McDonnell, H. H. Savenije, M. Sivapalan, et al. Twenty-three unsolved problems in hydrology (uph)–a community perspective. *Hydrological sciences journal*, 64 (10):1141–1158, 2019. URL https://doi.org/10.1080/02626667.2019.1620507.

[11] L. Brocca, F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer. Improving runoff prediction through the assimilation of the ascat soil moisture product. *Hydrology and Earth System Sciences*, 14:1881–1893, 2010. URL https://doi.org/10.5194/hess-14-1881-2010.

[12] L. Brocca, S. Hasenauer, T. Lacava, F. Melone, T. Moramarco, W. Wagner, W. Dorigo, P. Matgen, J. Martínez-Fernández, P. Llorens, et al. Soil moisture estimation through ascat and amsr-e sensors: An intercomparison and validation study across europe. *Remote Sensing of Environment*, 115(12):3390–3408, 2011. URL https://doi.org/10.1016/j.rse.2011.08.003.

[13] L. Brocca, L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani. Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *Journal of Geophysical Research: Atmospheres*, 119(9):5128–5141, 2014. URL https://doi.org/10.1002/2014JD021489.

[14] R. Chadwick and D. Grimes. An artificial neural network approach to multispectral rainfall estimation over africa. *Journal of Hydrometeorology*, 13(3):913 – 931, 2012. doi: 10.1175/JHM-D-11-081.1. URL https://journals.ametsoc.org/view/journals/hydr/13/3/jhm-d-11-081_1.xml.

[15] F. Chen, W. T. Crow, and D. Ryu. Dual forcing and state correction via soil moisture assimilation for improved rainfall-runoff modeling. *Journal of Hydrometeorology*, 15: 1832–1848, 2014. URL https://doi.org/10.1175/JHM-D-14-0002.1.

[16] W. Crow, M. van Den Berg, G. Huffman, and T. Pellarin. Correcting rainfall using satellite-based surface soil moisture retrievals: The soil moisture analysis rainfall tool

(smart). *Water Resources Research*, 47(8), 2011. URL https://doi.org/10.1029/2011WR010576.

[17] W. T. Crow, G. J. Huffman, R. Bindlish, and T. J. Jackson. Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals. *Journal of Hydrometeorology*, 10:199–212, 2009. URL https://doi.org/10.1175/2008JHM986.1.

[18] CWC. Integrated hydrological data book (non-classified river basins). Technical report, Hydrological Data Directorate, Information System Organisation, Water Planning Projects Wing, Central Water Commission, Govt. of India, New Delhi., 2012.

[19] D. P. Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613):3323–3343, 2005. URL https://doi.org/10.1256/qj.05.137.

[20] H. Demuth, M. Beale, and M. Hagan. *Neural network toolbox*. Mathworks, 1994.

[21] P. Droogers and W. Bastiaanssen. Irrigation performance using hydrological and remote sensing modeling. *Journal of Irrigation and Drainage Engineering*, 128(1):11–18, 2002. URL https://doi.org/10.1061/(ASCE)0733-9437(2002)128:1(11).

[22] Z. Duan, J. Liu, Y. Tuo, G. Chiogna, and M. Disse. Evaluation of eight high spatial resolution gridded precipitation products in adige basin (italy) at multiple temporal and spatial scales. *Science of the Total Environment*, 573:1536–1553, 2016. URL https://doi.org/10.1016/j.scitotenv.2016.08.213.

[23] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99: 10143–10162, 1994. URL https://doi.org/10.1029/94jc00572.

[24] H. Farman, Q. Abbas, S. Ahmad, A. Alshahrani, S. Jan, T. A. Syed, et al. A comparative study of deep learning and modern machine learning methods for predicting australia's precipitation. *International Journal of Advanced Computer Science & Applications*, 16 (4), 2025. URL https://doi.org/10.14569/ijacsa.2025.0160493.

[25] G. Fischer, F. Nachtergaele, S. Prieler, H. Van Velthuizen, L. Verelst, and D. Wiberg. Global agro-ecological zones assessment for agriculture (gaez 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy*, 10, 2008.

[26] A. N. Flores, D. Entekhabi, and R. L. Bras. Reproducibility of soil moisture ensembles when representing soil parameter uncertainty using a latin hypercube–based approach with correlation control. *Water Resources Research*, 46(4), 2010. URL https://doi.org/10.1029/2009WR008155.

[27] C. Francois, A. Quesney, and C. OttlÃ©. Sequential assimilation of ers-1 sar data into a coupled land surface-hydrological model using an extended kalman filter. *Journal of Hydrometeorology*, 4:473–487, 2003. URL https://doi.org/10.1175/1525-7541(2003)4<473:SAOESD>2.0.CO;2.

[28] W. H. Green and G. A. Ampt. Studies on soil physics. *Journal of Agricultural Science*, 4:1–24, 1911. URL https://doi.org/10.1017/S0021859600001441.

[29] E. Han, V. Merwade, and G. C. Heathman. Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. *Journal of Hydrology*, 416:98–117, 2012. URL https://doi.org/10.1016/j.jhydrol.2011.11.039.

[30] G. J. Huffman, D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker. The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hydrometeorology*, 8(1):38–55, 2007. URL https://doi.org/10.1175/JHM560.1.

[31] G. J. Huffman, D. T. Bolvin, E. J. Nelkin, and J. Tan. Integrated multi-satellite retrievals for gpm (imerg) technical documentation. *NASA/GSFC Code*, 612(47):2019, 2015.

[32] A. Jarvis, H. I. Reuter, A. Nelson, E. Guevara, et al. Hole-filled srtm for the globe version 4. *available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org)*, 15: 25–54, 2008.

[33] R. E. Kalman. A new approach to linear filtering and prediction problems. 1960. URL https://doi.org/10.1115/1.3662552.

[34] M. S. Khan and P. Coulibaly. Bayesian neural network for rainfall-runoff modeling. *Water Resources Research*, 42:1–18, 2006. URL https://doi.org/10.1029/2005WR003971.

[35] K. C. Kornelsen and P. Coulibaly. Root-zone soil moisture estimation using data-driven methods. *Water Resources Research*, 50:2946–2962, 2014. URL https://doi.org/10.1002/2013WR014127.

[36] A. Kumar, A. Dasgupta, S. Lokhande, and R. Ramsankaran. Benchmarking the indian national cartodem against srtm for 1d hydraulic modelling. *International Journal of River Basin Management*, 17(4):479–488, 2019. URL https://doi.org/10.1080/15715124.2019.1606816.

[37] A. Kumar, R. Ramsankaran, L. Brocca, and F. Munoz-Arriola. A machine learning approach for improving near-real-time satellite-based rainfall estimates by integrating soil moisture. *Remote Sensing*, 11(19), 2019. ISSN 2072-4292. URL https://www.mdpi.com/2072-4292/11/19/2221.

[38] B. Kumar, K. C. Patra, and V. Lakshmi. Daily rainfall statistics of trmm and cmorph: A case for trans-boundary gandak river basin. *Journal of Earth System Science*, 125: 919–934, 2016. URL https://doi.org/10.1007/s12040-016-0710-1.

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

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

[41] H. Lievens, S. K. Tomer, A. Al Bitar, G. J. De Lannoy, M. Drusch, G. Dumedah, H.-J. H. Franssen, Y. H. Kerr, B. Martens, M. Pan, et al. Smos soil moisture assimilation for improved hydrologic simulation in the murray darling basin, australia. *Remote Sensing of Environment*, 168:146–162, 2015. URL https://doi.org/10.1016/j.rse.2015.06.025.

[42] S. Manfreda, L. Brocca, T. Moramarco, F. Melone, and J. Sheffield. A physically based approach for the estimation of root-zone soil moisture from surface measurements. *Hydrology and Earth System Sciences*, 18:1199–1212, 2014. URL https://doi.org/10.5194/hess-18-1199-2014.

[43] C. Massari, L. Brocca, A. Tarpanelli, and T. Moramarco. Data assimilation of satellite soil moisture into rainfall-runoff modelling: A complex recipe? *Remote Sensing*, 7(9): 11403–11433, 2015. URL https://doi.org/10.3390/rs70911403.

[44] C. Massari, S. Camici, L. Ciabatta, and L. Brocca. Exploiting satellite-based surface soil moisture for flood forecasting in the mediterranean area: State update versus rainfall correction. *Remote Sensing*, 10(2):292, 2018. URL https://doi.org/10.3390/rs10020292.

[45] C. Massari, L. Brocca, T. Pellarin, G. Abramowitz, P. Filippucci, L. Ciabatta, V. Maggioni, Y. Kerr, and D. Fernández-Prieto. A daily/25 km short-latency rainfall product for data scarce regions based on the integration of the gpm imerg early run with multiple satellite soil moisture products. *Hydrol. Earth Syst. Sci*, 24:2687–2710, 2020. URL https://doi.org/10.5194/hess-24-2687-2020.

[46] H. Moradkhani, K. L. Hsu, H. Gupta, and S. Sorooshian. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41:1–17, 2005. URL https://doi.org/10.1029/2004WR003604.

[47] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3):885–900, 2007. URL https://doi.org/10.13031/2013.23153.

[48] S. L. Neitsch, J. G. Arnold, J. R. Kiniry, and J. R. Williams. Soil and water assessment tool theoretical documentation version 2009. Technical report, Texas Water Resources Institute, 2011.

[49] NRSC. National land use and land cover mapping using multi-temporal awifs data. Technical report, National Remote Sensing Centre, Indian Space Research Organisation, India, 2008.

[50] D. S. Pai, L. Sridhar, M. Rajeevan, O. P. Sreejith, N. S. Satbhai, and B. Mukhopadhyay. Development of a new high spatial resolution (0.25 Ã— 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, 2014. URL https://doi.org/10.54302/mausam.v65i1.851.

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

[52] A. A. Patil and R. Ramsankaran. Improved streamflow simulations by coupling soil moisture analytical relationship in enkf based hydrological data assimilation framework. *Advances in Water Resources*, 121:173–188, 2018. URL https://doi.org/10.1016/j.advwatres.2018.08.010.

[53] E. PeinÃ³, J. Bech, M. Udina, and F. Polls. Disentangling satellite precipitation estimate errors of heavy rainfall at the daily and sub-daily scales in the western mediterranean. *Remote Sensing*, 16(3), 2024. ISSN 2072-4292. doi: https://doi.org/10.3390/rs16030457. URL https://www.mdpi.com/2072-4292/16/3/457.

[54] G. Piazzi, G. Thirel, C. Perrin, and O. Delaigue. Sequential data assimilation for streamflow forecasting: Assessing the sensitivity to uncertainties and updated variables of a conceptual hydrological model at basin scale. *Water Resources Research*, 57:1–28, 2021. URL https://doi.org/10.1029/2020WR028390.

[55] M. C. Rogelis and M. Werner. Streamflow forecasts from wrf precipitation for flood early warning. *Hydrology and Earth System Sciences*, 22(1):853–870, 2018. URL https://doi.org/10.5194/hess-22-853-2018.

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

[57] S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, et al. The ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1058, 2010. URL https://doi.org/10.1175/2010BAMS3001.1.

[58] J. Samuel, P. Coulibaly, G. Dumedah, and H. Moradkhani. Assessing model state and forecasts variation in hydrologic data assimilation. *Journal of Hydrology*, 513:127–141, 2014. URL https://doi.org/10.1016/j.jhydrol.2014.03.048.

[59] H. Seyyedi, E. N. Anagnostou, P. E. Kirstetter, V. Maggioni, Y. Hong, and J. J. Gourley. Incorporating surface soil moisture information in error modeling of trmm passive microwave rainfall. *IEEE Transactions on Geoscience and Remote Sensing*, 52:6226–6240, 2014. URL https://doi.org/10.1109/TGRS.2013.2295795.

[60] A. Srivastava, M. Rajeevan, and S. Kshirsagar. Development of a high resolution daily gridded temperature data set (1969–2005) for the indian region. *Atmospheric Science Letters*, 10(4):249–254, 2009. URL https://doi.org/10.1002/asl.232.

[61] C.-H. Su, D. Ryu, R. I. Young, A. W. Western, and W. Wagner. Inter-comparison of microwave satellite soil moisture retrievals over the murrumbidgee basin, southeast australia. *Remote Sensing of Environment*, 134:1–11, 2013. URL https://doi.org/10.1016/j.rse.2013.02.016.

[62] S. Subbarayan, Y. M. Youssef, L. Singh, D. Dabrowska, N. Alarifi, R. Ramsankaran, R. Visweshwaran, and A. M. Saqr. Soil and water assessment tool-based prediction of runoff under scenarios of land use/land cover and climate change across indian agro-climatic zones: Implications for sustainable development goals. *Water*, 17(3):458, 2025. URL https://doi.org/10.3390/w17030458.

[63] M. Turner, J. Walker, and P. Oke. Ensemble member generation for sequential data assimilation. *Remote Sensing of Environment*, 112(4):1421–1433, 2008. URL https://doi.org/10.1016/j.rse.2007.02.042.

[64] R. Visweshwaran, R. Ramsankaran, T. Eldho, and M. K. Jha. Hydrological impact assessment of future climate change on a complex river basin of western ghats, india. *Water*, 14(21):3571, 2022. URL https://doi.org/10.3390/w14213571.

[65] R. Visweshwaran, R. Ramsankaran, T. Eldho, and S. Lakshmivarahan. Sensitivity-based soil moisture assimilation for improved streamflow forecast using a novel forward sensitivity method (fsm) approach. *Water Resources Research*, 58(1):e2021WR031092, 2022. URL https://doi.org/10.1029/2021WR031092.

[66] R. Visweshwaran, R. Ramsankaran, and T. Eldho. Improving modelled streamflow using time-varying multivariate assimilation of remotely sensed soil moisture and in-situ streamflow observations. *Advances in Water Resources*, 186:104676, 2024. URL https://doi.org/10.1016/j.advwatres.2024.104676.

[67] W. Wagner, G. Lemoine, and H. Rott. A method for estimating soil moisture from ers scatterometer and soil data. *Remote sensing of environment*, 70(2):191–207, 1999. URL https://doi.org/10.1016/S0034-4257(99)00036-X.

[68] W. Wagner, S. Hahn, R. Kidd, T. Melzer, Z. Bartalis, S. Hasenauer, J. Figa-Saldana, P. De Rosnay, A. Jann, S. Schneider, et al. The ascat soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*, 2013. URL https://doi.org/10.1127/0941-2948/2013/0399.

[69] F. Wang and D. Tian. Hourly evaluation of eight gridded precipitation datasets over the contiguous united states: Intercomparison of satellite, radar, reanalysis, and merged products. *Journal of Hydrometeorology*, 2025. doi: 10.1175/JHM-D-25-0063.1. URL https://journals.ametsoc.org/view/journals/hydr/aop/JHM-D-25-0063.1/JHM-D-25-0063.1.xml.

[70] M. Xiong, P. Liu, L. Cheng, C. Deng, Z. Gui, X. Zhang, and Y. Liu. 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, 2019. URL https://doi.org/10.1016/j.jhydrol.2018.11.038.

[71] B. Yong, D. Liu, J. J. Gourley, Y. Tian, G. J. Huffman, L. Ren, and Y. Hong. Global view of real-time trmm multisatellite precipitation analysis: Implications for its successor global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 96(2):283–296, 2015. URL https://doi.org/10.1175/BAMS-D-14-00017.1.