Use of Remotely Sensed Soil Moisture Data in Hydrological Modelling

***Gunasekara KM, Gunawardhana L, and Rajapakse L**

Department of Civil Engineering, University of Moratuwa, Sri Lanka

*Corresponding author – 180206l@uom.lk

Abstract
Accurate hydrological data is crucial for understanding streamflow changes and predicting extreme events in river basins. By utilising remotely sensed soil moisture products, the estimation of soil moisture distribution at the basin scale in hydrological modelling becomes feasible, addressing practical challenges. This study investigates the integration of remotely sensed soil moisture estimates to enhance the accuracy of hydrological model simulations in the upper Peradeniya catchment, Sri Lanka. A hydrological model (ABCD model) was developed for the catchment area. Soil moisture data from NASA's soil moisture active passive (SMAP L4) were integrated with the model's estimations to improve river flow simulations. The study reveals that the integration of SMAP L4 did not significantly enhance accuracy. However, notable differences in calibrated parameters emerged, highlighting the importance of incorporating multiple inputs for calibration. These findings demonstrate the potential of remotely sensed soil moisture in enhancing hydrological model simulations. Incorporating remotely sensed soil moisture can contribute to more reliable predictions and management of water resources in river basins, aiding in sustainable development and climate change adaptation.

Keywords: ABCD model, Remote sensing, SMAP L4 products, Upper Peradeniya catchment

1 Introduction

Soil moisture was recognized as one of the fifty essential climate variables by the Global Climate Observing System in their 2010 implementation plan update. Further, soil moisture was categorized as a crucial component of the terrestrial climate system [1].

In groundwater hydrology, soil moisture plays an important role in storing water resulting from precipitation. Based on the soil moisture content, the incoming precipitation is divided among infiltration, subsurface flow and runoff which makes soil moisture important in flood simulations and management.

Soil moisture monitoring has been practised since the start of agriculture. However, with climate change, soil moisture monitoring has become a complex challenge with changes in temperature and precipitation [2]. Because of such challenges, having near real-time soil moisture data is important for productive decision-making and forecasting processes.

For large catchment areas, obtaining spatially and temporally well-distributed soil moisture data sets is beneficial. However, due to practical issues such as capital and maintenance costs of the instruments and limited spatial coverage, in-situ methods are not feasible to implement. With the evolution of satellite and computer technologies, new methods were introduced to overcome this barrier. Remote sensing and hydrological modelling that simulate soil moisture are two such technologies.

The two technologies have been used in various ways to enhance the performance of a hydrological model. Joint calibration between observed streamflow and soil moisture is a method that has successfully used remotely sensed soil moisture (RSSM) data to improve the output (river discharge) of the hydrological model [3]. The lumped hydrological model ABCD produces runoff as the output while simulating soil moisture variation internally[4]. Therefore, by using RSSM data, the performance of the ABCD model could be enhanced.

The upper Peradeniya catchment (UPC) holds significant importance as a sub-catchment within Sri Lanka's largest river basin, the Mahaweli river basin [5]. The sustainable management of water resources in this watershed is crucial due to its role in the operation of a cascade reservoir system. These reservoirs serve multiple purposes, including hydropower generation, irrigation, and domestic water supply. Given the dynamic nature of land use patterns and the impact of climate change, the hydrological behaviour of the UPC, the catchment is expected to change in the future. To address this, conducting a hydrological study of the upper Peradeniya catchment would prove beneficial to the country. A well-developed hydrological model with an appropriate technique for aggregating RSSM can be used to project the changes in the hydrological cycle over time and the subsequent impact on the downstream environment.

2 Methodology

The following methodology (Fig. 1) was followed for this study.

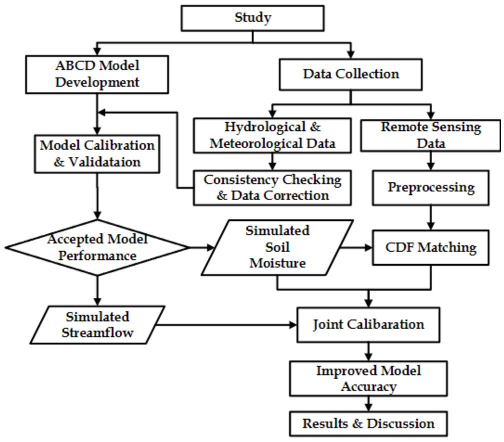

Figure 1: Methodology flowchart

2.1 Study area & data sets

The UPC covers an area of 1,168 km² and is located between the Kandy and Nuwara Eliya Districts of the Central Province in Sri Lanka (see Fig. 2). The study area is surrounded by a mountainous landscape, with notable mountains including Piduruthalagala Mountain, which marks the origin of the Mahaweli river and is the highest in Sri Lanka.

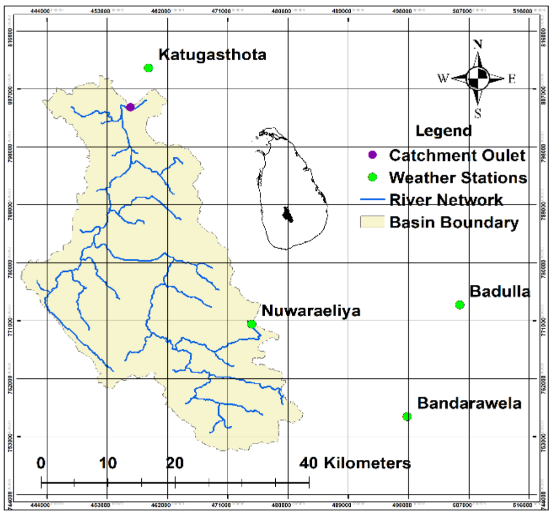

Figure 2: Study area

The UPC is situated upstream from the Peradeniya hydrometric station, spanning a length of 99.6 km along the river. The average ground slope in the catchment is 22.4%, with the steepest estimated slope reaching 99.98%. The study area has an average altitude of 595 m above mean sea level (M.S.L.), ranging from a minimum altitude of 478 m to a maximum altitude of 2519 m.

The daily mean for maximum temperature is 22.8ºC while the daily mean for minimum temperature is 14.5ºC. The 16-year mean average annual temperature is roughly 18.6ºC, reaching a minimum monthly mean of 17.4ºC in January and a maximum monthly mean of 19.8ºC in May. Shuttle Radar Topography Mission Digital Elevation Model (DEM) files with 90 m × 90 m resolution from the United States Geological Survey Earth Explorer were used and ArcGIS 10.8.2 was used to delineate the catchment area.

RSSM product, soil moisture active passive (SMAP) L4 root zone soil moisture from The National Aeronautics and Space Administration (NASA) Earth data was used. SMAP satellite was launched by NASA on 31st January 2015, to measure SM and freeze/thaw at a global scale using an L-band (1.40 GHz) radiometer and an L-band radar [6]. SMAP provides data at a spatial resolution of 36 km since SMAP radar stopped transmitting on 7th July 2015. SMAP L4 offers data from 2015/03/31 onwards [6]. SMAP L4 provides two soil moisture data sets, near-surface (0 – 5 cm) and root zone (0 – 100 cm) soil moisture.

SMAP L4 root zone soil moisture (SMAPL4_RZ_SM) dataset is produced by assimilating SMAP L-band brightness temperature observations into the NASA catchment land surface model. The assimilation has resulted in soil moisture estimates at a spatial resolution of 9 km and a temporal resolution of 3 hours with a data latency of 3 days [6].

Studies [7], and [8] have found that SMAP L4 near-surface soil moisture products have the highest accuracy compared to other RSSM products. ArcGIS 10.8.2 was used to create soil moisture maps of the catchment area from the SMAP L4 root zone soil moisture product as a soil moisture value that represents a greater depth than the near-surface zone product.

The meteorological and hydrological data for the study were obtained from stations near the study area owned and operated by the Irrigation Department of Sri Lanka (see Fig. 2). Daily streamflow data were collected from the river gauging station at Peradeniya, covering the period from October 1984 to September 2020. Daily atmospheric temperature data were collected from three meteorological stations Katugasthota, Nuwara Eliya and Badulla, providing maximum and minimum temperature records from January 1996 to September 2021. Daily rainfall data was collected from the three meteorological stations Katugasthota, Nuwara Eliya, and Bandarawela, covering the period from January 2005 to December 2020.

2.2 Data processing and analysis

Multivariate imputation by chained equations (MICE) is a method of imputing missing values in a dataset with multiple variables while preserving the correlations between variables [9]. There are various methods to perform MICE and for this study, predictive mean matching was used. The visual observation and manual data checking of hydrological and metrological data suggest that on a particular day, there was at least one data station without missing data. This high level of completeness enhances the suitability of the datasets for further analysis and imputation using the MICE method. The highest percentage of missing data was 0.99% recorded for minimum temperature at the Katugastota station. The missing data were imputed by running MICE in R programming language. After replacing the missing data, the potential evapotranspiration was derived from the temperature data using the Hargraves method. The rainfall data from the stations were interpolated using the Thiessen polygon method. The Thiessen polygon were created using ArcGIS 10.8.2.

2.3 Selection of model

The ABCD model (Fig. 3) accepts precipitation, potential evapotranspiration (PE), initial soil moisture and initial groundwater storage as inputs. The model can be used at various temporal resolutions and the daily temporal resolution was chosen for this study.

For the model simulation, arbitrary values were used for initial soil moisture and groundwater storage, which will not cause any problems when the model is run for a longer time, as the effect of the initial conditions on the model performance will be insignificant.

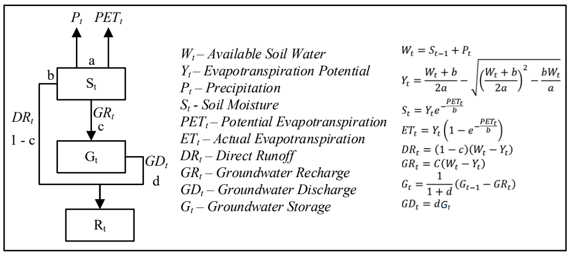

Figure 3: ABCD model

In Fig. 3, Parameter 'a' shows the runoff tendency to occur before the soil is fully saturated. Parameter 'b' represents the upper limit of the sum of monthly soil moisture storage and actual evapotranspiration. Parameter 'c' controls the water input to the aquifers. The average groundwater residence time equals the reciprocal of the parameter 'd' [4].

2.4 Preprocessing of soil moisture data

2.4.1 Spatial and temporal aggregation of soil moisture

SMAPL4_RZ_SM datasets are provided as raster data, and fifteen data cell centres were located within the catchment area. To use the SMAPL4_RZ_SM data with the ABCD lump model, it was necessary to aggregate the cell values into a single representative value for the entire catchment area (spatial aggregation). The inverse distance weighting (IDW) method was selected for this purpose.

The SMAPL4_RZ_SM datasets offer a temporal resolution of 3 hours, resulting in eight spatially aggregated soil moisture values per day. Since the ABCD model of this study operates on a daily scale, the spatially aggregated soil moisture values needed to be aggregated accordingly (temporal aggregation). The simple average method was used to accomplish this, as it has previously been used with other soil moisture products [10]. The spatially and temporally aggregated soil moisture values are then multiplied from 1000 mm (SMAPL4_RZ_SM captures soil moisture up to 1 m depth), converting the unit of soil moisture from volumetric units to mm. More than 14,600 raster files had to be processed and the Model Builder feature of ArcGIS 10.8.2 was used to automate the process.

2.4.2 Rescaling and cumulative distribution function matching of soil moisture

A rescaling procedure is needed to remove systematic differences or biases between the raw SMAPL4_RZ_SM product and the ABCD simulated soil moisture. For this rescaling cumulative distribution function (CDF) matching method [3] was chosen, and to implement this method, a MATLAB algorithm was used. The SMAPL4_RZ_SM dataset captures soil moisture up to 100 cm depth. However, the model simulates soil moisture up to aquifer depth. Therefore, a direct comparison is unreasonable.

A widely adopted approach [3] to address this concern is to employ the exponential filtering (EF) technique, introduced initially by [11], to transform the rescaled remote sensing soil moisture data ($\emptyset_{tn}{}^{CDF-RS}$) into the soil wetness index (SWI) of the root zone. The following EF is applied to develop the SWI time series:

$$\emptyset_{t_n}^{SWI}=\emptyset_{t_{n-1}}^{SWI}+K_n(\emptyset_{t_n}^{CDF\text{-}RS}-\emptyset_{t_{n-1}}^{SWI}) \quad (1)$$

Here, $\emptyset_{t_n}^{SWI}$ is the SWI value for the n^{th} day and K_n is gain time which can be denoted as follows.

$$K_n=\frac{K_{n-1}}{K_{n-1}+e^{-\left(\frac{t_n-t_{n-1}}{T}\right)}} \quad (2)$$

The parameter T is a characteristic time length that controls the smoothing degree of the $\emptyset_{t_n}^{CDF-RS}$series and the response time to the changes in the surface wetness conditions. In this study, T takes the value of 1 day as it provides the best correlation between $\emptyset^{SWI}$ and $\emptyset^{ABCD}$. For the initial conditions, K_1 and $\emptyset_{t_1}^{SWI}$ were taken as one and $\emptyset_{t_n}^{CDF-RS}$adapting [3].

The degree of agreement between the observed and simulated river flow was assessed using the Kling-Gupta efficiency coefficient (KGE) [12]. When integrating SMAPL4_RZ_SM estimations into the ABCD model, a weighted objective function considering both stream flow and soil moisture data was used. The respective equations are mentioned as follows.

$$KGE_Q= 1-\sqrt{(r_Q-1)^2+\left(\frac{\mu_{Q\text{-}ABCD}}{\mu_{Q\text{-}O}}-1\right)^2+\left(\frac{\sigma_{Q\text{-}ABCD}}{\sigma_{Q\text{-}O}}-1\right)^2} \quad (3)$$

In (3), r_Q is the correlation between observed and simulated river flows, μ_{Q-ABCD} and μ_{Q-O} are the means of the simulated and observed river flow series, and σ_{Q-ABCD} and σ_{Q-O} are the standard deviations of the simulated and observed river flow series, respectively. After the initial parameter calibration using river flow data, the KGE coefficient for soil moisture was estimated as follows.

$$KGE_{SM}=1-\sqrt{(r_\emptyset-1)^2+\left(\frac{\mu_\emptyset ABCD}{\mu_\emptyset SWI}-1\right)^2+\left(\frac{\sigma_\emptyset ABCD}{\sigma_\emptyset SWI}-1\right)^2} \quad (4)$$

In (4), r_ϕ represents the correlation between ϕ ABCD and ϕ^{SWI}, $\mu_\phi{}^{ABCD}$ and $\mu_\phi{}^{SWI}$ are the means of ϕ^{ABCD} and ϕ^{SWI}, and $\sigma_\phi{}^{ABCD}$ and $\sigma_\phi{}^{SWI}$ are the standard deviations of ϕ^{ABCD} and ϕ^{SWI}, respectively. The Weighted Objective Function that evaluates SMAPL4_RZ_SM soil moisture integration to the model is shown in (5). There, the coefficient α is the assigned weight, which ranges from 0 to 1.

$$KGE_\alpha=\alpha\times KGE_{SM}+(1-\alpha)\times KGE_Q \quad (5)$$

To assess the influence of RSSM estimates on model calibration, eleven different calibration schemes were implemented for the study catchment. Within these eleven schemes, the weight parameter α in the objective function KGE_α was varied from 0 to 1 with 0.1 increments. It is important to note that when α is set to 0, the model calibration solely relies on streamflow data, resulting in KGE_α being equivalent to KGE_Q.

3 Results and discussion

3.1 Rainfall shadow effect

Usually, with increasing elevation, the annual average rainfall should also increase. This is called the orographic effect of rainfall. When the annual average rainfall (from 2005 - 2020) and the elevations of the rainfall gauging stations Katugasthota, Nuwara Eliya and Bandarawela were considered, an anomaly from the rainfall orographic effect was identified. Despite Katugasthota having the lowest elevation, it has the highest annual average rainfall (Table 1).

Table 1: Annual average rainfall and elevations

Rainfall gauging station	**Elevation from mean sea level (m)**	**Annual average rainfall (mm/year)**
Katugastota	417	2527.73
Nuwara Eliya	1894	2414.36
Bandarawela	1225	2244.44

The Katugasthota and Bandarawela stations are located outside the basin and Nuwara Eliya is inside the basin. A mountain range surrounds the entire basin, and this anomaly is due to the rainfall shadowing effect.

3.2 Land use land cover of UPC

According to the land use and land cover, (LULC) extracted from the Sentinel – 2, 10 m land use/land cover time series of the world, UPC has an estimated area of 53% tree cover, 16.7% grasslands and 26.9% built-up areas. The built-up areas are more frequent in the northern part of the catchment than in other areas (see Fig. 4).

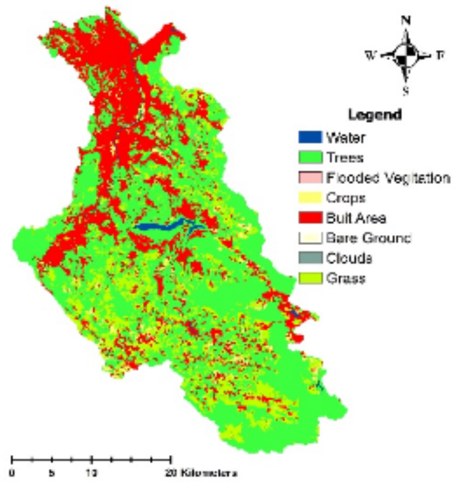

Figure 4: LULC map of UPC

3.3 CDF matching

Before performing the CDF matching, the spatially and temporally aggregated RSSM has to be converted to mm/day. The simulated soil moisture data from the ABCD model was used as reference data. The polynomial (7) of the corrected data curve (Fig. 5) is as follows.

The absolute bias after the correction was 20.4. The degree of the polynomial was chosen from trial and error. The third-degree polynomial gave the least absolute bias and the best-fitting curve with the reference data.

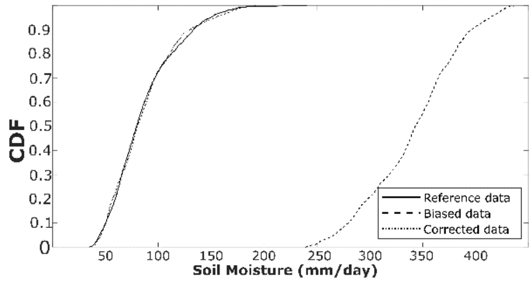

Figure 5: CDF curve

$$X_C = 1.56 \times 10^{-5}X^3 - 1.19 \times 10^{-2}X^2 + 2.34X - 293.88 \quad (7)$$

X_c - Corrected RSSM
X - Spatially and temporally aggregated RSSM soil moisture

3.4 ABCD modelling

During the calibration process using the weighted objective function, the Microsoft Excel solver evaluation mode was employed for the initial calibration. Solver evaluation mode requires ranges for the variable parameters and the ranges were provided from a, b, c and d values of previous studies [4]. Subsequently, manual fine-tuning was conducted. Table 2 presents the results obtained. N/A indicates that the objective function is not applicable.

Based on the results during the calibration period, the highest KGE_Q value is achieved for the α = 0.1 arrangement. However, it is essential to note that the gain in performance is at the fourth decimal point, indicating that the impact of this gain is insignificant. However, in this arrangement, the model successfully simulates soil moisture (KGE_{SM}) up to a satisfactory level of 0.6518 (with SWI) in Table 2, aligning with the findings of previous research conducted by [3].

As the α value increases, the KGE_Q value during the calibration period exhibits irregular behaviour. According to previous studies [3], aggregating RSSM in joint calibration [6] during the calibration process leads to a slight degradation in KGE_Q. In the study by [3], one study area showed this irregular behaviour. One possible reason is the limitation of the conceptual lump model (ABCD) to precisely simulate the complex soil profile characteristics of the UPC. The simplified representation of soil processes in the model can result in discrepancies between the simulated and observed soil moisture (Processed RSSM) values.

Table 2: Results of simulations of all arrangements

α	Calibration			Validation
	KGE_Q	KGE_{SM}	$KGE_α$	KGE_Q
0	0.7301	0.6483	N/A	0.6247
0.1	0.7310	0.6519	0.7230	0.6306
0.2	0.7300	0.6521	0.7144	0.6277
0.3	0.7299	0.6519	0.7065	0.6267
0.4	0.7299	0.6520	0.6987	0.6274
0.5	0.7298	0.6518	0.6908	0.6262
0.6	0.7299	0.6520	0.6832	0.6275
0.7	0.7300	0.6521	0.6754	0.6278
0.8	0.7300	0.6521	0.6677	0.6281
0.9	0.7301	0.6521	0.6599	0.6279
1	0.7301	0.6522	N/A	0.6283

However, the integrated attempt of both stream flow and remotely sensed soil moisture in parameter calibration produced a reasonably good agreement with *KGE* of more than 0.65 for all cases. Additionally, uncertainties associated with input data can contribute to KGE_Q degradation. These uncertainties can arise from various sources, such as RSSM data errors or inaccuracies in the meteorological inputs used for model calibration. These uncertainties propagate throughout the modelling process and can impact the model's overall performance.

When comparing the observed and the simulated soil moisture (Fig. 6), it is observed that, at specific points, the catchment exhibits a lag in response to precipitation. This delay can be attributed to the Kothmale reservoir and dam within the UPC area. As mentioned previously, the ABCD model exhibits limitations in catchments with large water bodies, as it does not accurately simulate the storage effect of reservoirs and the impact of dams on streamflow.

However, the model has simulated streamflow up to a satisfactory level of 0.7301 when calibrated only with observed streamflow (α = 0). Flow duration curves (FDC) plotted for the observed and simulated flows (Fig. 7) show reasonably good agreement. There is no apparent difference in results for different weightages.

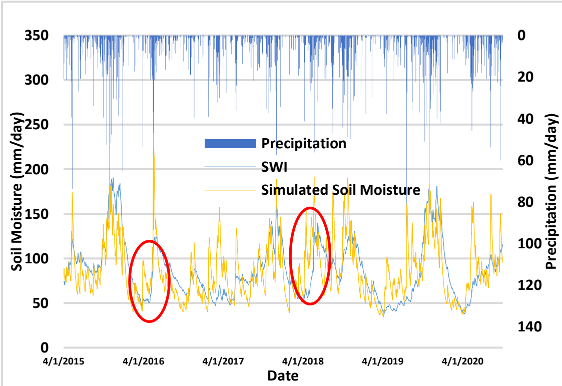

Figure 6: Soil moisture variation over time

When examining the Fig. 7, it becomes apparent that the FDC curves of the simulated streamflow data are almost identical. All simulated flows consistently underestimate extreme high flows compared to in situ extreme high flows. However, in the case of mid and low flows, the simulated flows tend to overestimate the observed flow values.

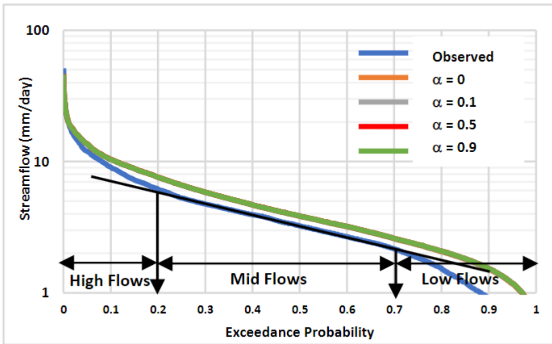

Figure 7: FDC curve of UPC

4 Conclusion

The findings of this study suggest that the ABCD model can simulate soil moisture to an acceptable level, even only when calibrated with streamflow data. The incorporation of RSSM has a limited impact on improving streamflow simulation performance, with only a slight variation observed. However, when considering simulated soil moisture, there is an overall tendency for slight performance improvement. The joint calibration function indicates a degradation in the model performance when incorporating RSSM. However, it is essential to note that including reliable data sources reduces the number of uncertain simulated components of the model, potentially increasing its reliability and bringing the simulations closer to real-world conditions. The changes observed in the SWI and ABCD simulated soil moisture variation patterns over time support this notion. Thus, the joint calibration approach could serve as an alternative to traditional calibration methods. In future studies, selecting catchments with no large water bodies or employing models capable of handling such situations is recommended. approaches such as data assimilation and multi objective calibration can also be explored. Other RSSM products with extended temporal coverage could also be explored. The examination of the impact of spatial resolution, especially with the availability of finer-resolution products such as Sentinel 1, could further enhance the understanding of this approach.

References

[1] World Meteorological Organization, "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 update) - Executive summary United Nations Environment Programme International Council for Science," 2010. [Online]. Available:http://www.wmo.int/pages/prog/gcos/Publications/gcos-92_GIP.pdf

[2] S. I. Seneviratne et al., "Investigating Soil Moisture-Climate Interactions in a Changing Climate: A Review," *Earth Sci Rev*, vol. 99, no. 3–4, pp. 125–161, 2010, doi: 10.1016/j.earscirev.2010.02.004.

[3] L. Xiong and L. Zeng, "Impacts of introducing remote sensing soil moisture in calibrating a distributed hydrological model for streamflow simulation," *Water (Switzerland)*, vol. 11, no. 4, Apr. 2019, doi: 10.3390/w11040666.

[4] P. G. Marinou, E. G. Feloni, O. Tzoraki, and E. A. Baltas, "An Implementation of a Water Balance Model in the Evrotas Basin," *European Water*, vol. 57, no. January, pp. 147–154, 2017.

[5] T. Hewawasam, "Effect of Land Use in The Upper Mahaweli Catchment Area on Erosion, Landslides and Siltation in Hydropower Reservoirs of Sri Lanka," J Natl Sci Found, vol. 38, no. 1, pp. 3–14, 2010.

[6] R. H. Reichle et al., "Assessment of the SMAP Level-4 surface and root-zone soil moisture product using in situ measurements," *J Hydrometeorol*, vol. 18, no. 10, pp. 2621–2645, Oct. 2017, doi: 10.1175/JHM-D-17-0063.1.

[7] M. El Hajj et al., "Evaluation of SMOS, SMAP, ASCAT and Sentinel-1 Soil Moisture Products at Sites in Southwestern France," *Remote Sens (Basel)*, vol. 10, no. 4, pp. 1–17, 2018, doi: 10.3390/rs10040569.

[8] R. L. Ray, A. Fares, Y. He, and M. Temimi, "Evaluation and Inter-comparison of Satellite Soil Moisture Products Using In situ Observations Over Texas, U.S.," *Water (Switzerland)*,

vol. 9, no. 6, 2017, doi: 10.3390/w9060372.

[9] S. Van Buuren and K. Groothuis-Oudshoorn, "*Journal of Statistical Software mice: Multivariate Imputation by Chained Equations in R*," 2011.

[10] R. H. Reichle and R. D. Koster, "Bias reduction in short records of satellite soil moisture," *Geophys Res Lett*, vol. 31, no. 19, Oct. 2004, doi: 10.1029/2004GL020938.

[11] W. Wagner, G. Lemoine, and H. Rott, "A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data," *Remote Sens Environ*, vol. 70, no. 2, pp. 191–207, 1999, doi: doi.org/10.1016/S0034-4257(99)00036-X.

[12] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling," *J Hydrol (Amst)*, vol. 377, no. 1–2, pp. 80–91, Oct. 2009, doi: 10.1016/j.jhydrol.2009.08.003.