Water Resources Management (2026) 40:188
https://doi.org/10.1007/s11269-026-04568-5

Uncertainty Reduction in Near Real-time Satellite Precipitation Estimates by Integrating Soil Moisture and Potential Evapotranspiration Using a Machine Learning Approach

Susantha Wanniarachchi[1,2] · Ranjan Sarukkalige[1] · H. A. Prasantha Hapuarachchi[3] · Pattiyage I.A. Gomes[2] · Upaka Rathnayake[4]

Received: 5 August 2025 / Accepted: 3 March 2026 / Published online: 13 March 2026

Abstract
Near-real-time (NRT) satellite precipitation data inherits complex and random errors due to various reasons. The primary objective of this research is to utilize satellite-based precipitation data for hydrological modelling in ungauged areas. The novelty of this study lies in the development of a hybrid stacking-based machine learning framework that integrates hydrologically meaningful predictors: root-zone soil moisture, potential evapotranspiration (PET), and their time-lagged representations to reduce uncertainty in near-real-time satellite precipitation (GSMaP-NRT). Unlike conventional bias-correction approaches that rely primarily on statistical adjustment between satellite and gauge rainfall, this study incorporates physically relevant catchment-state variables to improve predictive skill, with a focus on the Ovens River basin in Australia. A calibrated GR4H hydrological model was used to simulate the runoff of the catchment. Six objective functions were used to evaluate the performance of the approach. The results demonstrate that stacking machine learning algorithms reduces the Mean Absolute Error of GSMaP-NRT satellite precipitation data by 36% and the corresponding modelled streamflow error by 44% for lower precipitation events (<2 mm/hour). All six objective functions achieved optimal performances within the low precipitation events. However, RMSE remained high for intermediate and heavy precipitation events. The model-estimated major streamflow peaks for the years 2010 and 2016, based on gauged precipitation and ML-corrected satellite precipitation, are 41% and 48% lower than the observed streamflow peaks, respectively. The reasons were the inability of the GR4H model to capture the perfect initial conditions and the x_4 time parameter during the calibration process.

Keywords Machine learning · Satellite precipitation · GSMaP-NRT · Hydrological simulation · GR4H

Extended author information available on the last page of the article

1 Introduction

Catchment precipitation is a key forcing parameter in studies that use hydrological simulation approaches to predict floods (Wang et al. 2021). However, the non-linear behaviour of streamflow is governed by hydro-meteorological factors and anthropogenic activities, making accurate flood modelling more challenging (Madhushani et al. 2024). The primary real-time catchment precipitation data sources are telemetry rain gauges, weather radar, and satellites. The Bureau of Meteorology Australia collects weather data through automatic weather stations, meteorological satellites, and drifting ocean buoys, each with varying reporting frequencies and observation times (Bureau of Meteorology 2024). The distribution of weather stations is uneven due to socioeconomic and geographical factors (Spassiani et al. 2023), leading to issues such as the inability of point measurements to represent precipitation's spatial distribution accurately. In addition, there can be systematic and calibration errors (Amini et al. 2024). Even though a dense rain gauge network reduces uncertainties in interpolated precipitation data, this can be challenging because the technical and operational costs increase with regular monitoring and maintenance to achieve finer resolution and more reliable data (Chua et al. 2022; Li et al. 2021). Also, only limited near-real-time ground-based observations are available for most regions (Kumar et al. 2019). As a potential solution, many satellite precipitation products (SPPs) have emerged recently, with some offering high spatial resolution and others high temporal resolution (Chua et al. 2022; Li et al. 2021; Singh and Singh 2024) (Table 1).

SPPs need to be frequently calibrated to remove inherent algorithmic assumptions, atmospheric uncertainties, and sensor biases. These uncertainties arise from multiple sources, including simplified representations of hydrometeor types in microwave-based products, reliance on cloud-top temperature in infrared-based estimates, and the inability of satellites to capture precipitation over complex terrain accurately. Moreover, these uncertainties are highly non-linear and vary across seasons, precipitation intensities, topographic conditions, and geographic regions (Kumar et al. 2019; Massari et al. 2018; Rachidi et al. 2023; Wang et al. 2021). These complexities cannot be comprehensively addressed through conventional statistical-based bias correction approaches, which often fail in addressing situations where there are conditions of extreme rainfall events and conditions of mountainous terrain (Hinge et al. 2022; Nguyen et al. 2024; Priyambodoho et al. 2021).

While gauge-corrected SPPs are readily available, not all of them are provided on a near-real-time (NRT) mode, limiting their effective use in operational flood forecasting and disaster management applications (Japan Aerospace Exploration Agency 2024). Near-real-time precipitation estimates are often required for short-term hydrological forecasting, early warning systems, and emergency response (Bi et al. 2023; Fowdur and Nazir 2022). Consequently, NRT-SPPs such as PERSIANN-CCS, IMERG-E, GSMaP-NRT, CMORPH, and HydroEstimator have been widely used in previous studies (Ali et al. 2020; Llauca et al. 2021). However, their direct application remains challenging due to persistent uncertainties, highlighting the need for improved correction strategies before operational use (Wang et al. 2021).

Recent studies have demonstrated the potential of machine learning (ML) techniques to overcome some of these limitations by capturing non-linear relationships between precipitation and auxiliary variables (Generoso et al. 2023; Zhao et al. 2025). For example, Kumar et al. (2019) used Support Vector Regression to improve 3B42RT precipitation estimates

Table 1 Details of the Satellite Precipitation products (SPPs)

SPP	Short Name	Data period	Resolution	Reference
Climate Hazards Group Infrared Precipitation with Stations	CHIRPS	1981 to the present	0.05^0/ 5 km/ Global/ six-hourly	Climate Hazards Centre (University of California 2025)
Climate Prediction Centre (CPC) Merged Analysis of Precipitation	CMAP	1979 to 2008	$2.5^0 \times 2.5^0$ / 275 km / Global/ monthly and pentad	Climate Prediction Centre (National Weather Service 2006)
CPC morphing technique	CMORPH	1998 to the present	$0.073^0 \times 0.073^0$ / 8 km/ Global/ half hourly	National Centres for Environmental Information (National Center for Atmospheric Research 2025)
Global Precipitation Climatology Project	GPCP	1979 to the present	$2.5^0 \times 2.5^0$ / 275 km / Global/ monthly and daily	NOAA (NOAA 2025)
Integrated Multi-satellite Retrievals for GPM version 6	IMERG	2000 to the present	$0.1^0 \times 0.1^0$ / 10 km / Global/ half-hourly	NASA (NASA 2020)
Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks	PERSIANN-CCS/ CDR	1983 to the present	$0.25^0 \times 0.25^0$ / 25 km / Global/ daily	NOAA NCDC (NOAA 2025)
TRMM Multi-satellite Precipitation Analysis	TMPA 3B42	1998 to 2019	$0.25^0 \times 0.25^0$ / 25 km / Latitudes 50^0 to -50^0/ Three hourly	NASA (NASA 2020)
TerraClimate		1958 to 2023	0.042^0/ 4 km/ Global/ monthly	(Abatzoglou 2021)
Tropical Applications of Meteorology Using Satellite	TAMSAT	1983 to the present	0.0375^0/ 4 km/ Only for Africa/ monthly	(University of Reading 2025)
Global Satellite Mapping of Precipitation (NRT, MVK, Gauge V7)	GSMaP-NRT	2008 to the present	$0.1^0 \times 0.1^0$ / 10 km / Global/ hourly	(Japan Aerospace Exploration Agency 2024)

using Advanced Scatterometer (ASCAT) soil moisture, and some positive enhancements were recorded for light and moderate rainfall, but limited success for heavy precipitation. This predictive performance has been further enhanced with the development of ensemble and deep learning techniques (Fowdur and Nazir 2022; Zhang et al. 2023), yet many of these models function as black boxes, with limited physical interpretability (Rachidi et al. 2023). More recent efforts incorporating explainable artificial intelligence (XAI) have begun addressing this issue, primarily in streamflow prediction rather than precipitation correction (Madhushani et al. 2024).

Feature selection and hyperparameter optimization have been identified as critical components in ML-based precipitation correction. Several studies highlight the importance and strong relationship between soil moisture and precipitation, driven by the persistence of soil moisture signals following rainfall events (Kumar et al. 2019; Massari et al. 2018). Recent studies have also shown that auxiliary variables such as soil moisture and PET have proven useful for drought monitoring and have shown successful application for hydrological stud-

ies (Li et al. 2024; Parajuli et al. 2024). However, most of the studies have been carried out for daily or monthly aggregated time scales. This is inadequate for identifying the short-term variability and quick response that is essential for flood forecasting.

Therefore, a critical research gap remains in developing an operational, high-temporal-resolution machine learning framework capable of reducing uncertainties in near-real-time satellite precipitation products, particularly for extreme rainfall events in flood-prone, data-scarce regions.

To address this gap, the present study evaluates the potential of a stacking-based machine learning approach that integrates Extreme Gradient Boosting Regression, Random Forest, and Multiple Linear Regression models with satellite-derived soil moisture and PET data at an hourly resolution. By combining tree-based and linear models, the proposed framework aims to enhance both predictive accuracy and robustness. The methodology is applied to GSMaP-NRT precipitation estimates over an Australian river basin prone to frequent flooding, to improve the reliability of NRT satellite precipitation for operational flood forecasting.

2 Study Area and Datasets

2.1 Study Area

The Ovens Basin is in north-east Victoria in Australia. The catchment covers 7985 km^2 and is part of the Murray-Darling Basin (MDB). It contributes approximately 6% of the total runoff of the MDB (CSIRO 2008). Generally, catchment precipitation varies dramatically from the south to the north, ranging from southern alpine areas (1600 mm) to lower floodplains (550 mm). Even though rain is the main form of precipitation in the basin, a major component of winter precipitation falls as snow in the alpine region above 1400 m. The average precipitation is 1196 mm/ year, and the average annual streamflow is 1775 GL at Peechelba (Murray-Darling Basin Authority 2023). Nearly 65% of the average annual precipitation occurs in winter from June to August. However, the southern alpine region receives erratic precipitation during the summer due to the orographic effects (Goulburn-Murray 2024).

Water management in the Ovens basin is extremely important, as the city of Wangaratta contributes to the national economy through agriculture and serves as an industrial manufacturing hub. The year 1993 was the highest recorded flood in Wangaratta over the last 100 years, causing severe economic losses. There were three other major flood events in 2010, 2016, and 2020. Hume Dam on the Murray River, with a capacity of 1417 GL, located upstream of the Ovens basin (Fig. 1), plays a major role in flood control in the lower MDB and in maintaining downstream water security. Accurate Ovens streamflow forecasts enable efficient control of Hume Dam water releases.

2.2 Data

Since the accuracy of the machine learning approaches will be evaluated through one-to-one rainfall comparisons and hydrological model output comparisons, this study utilized multiple hydrological datasets to train the machine learning models and to develop and calibrate the hydrological model.

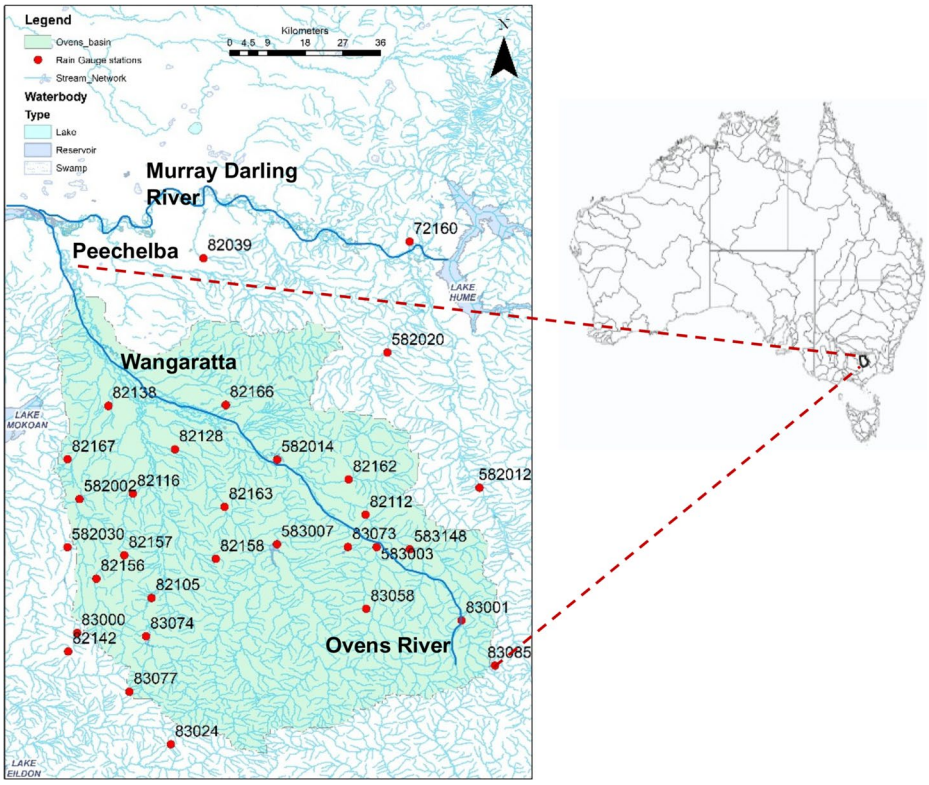

Fig. 1 Stream network and the rain gauging stations of the Ovens basin

2.2.1 Gauged Precipitation Data and PET Data

There were thirty-four rain gauge stations in the Ovens basin. The basin was subdivided into fifty-three sub-areas for model development. Hourly resolution gauged precipitation data, sub-area average precipitation data prepared using the inverse distance weighted (IDW) method, and daily resolution sub-area average potential evapotranspiration (PET) data were collected from the Bureau of Meteorology, Australia, from 2007 to 2017 period (Bureau of Meteorology 2024). Assuming diurnal variation in PET (Emmanuel et al. 2023), a sinusoidal function (Eq. 1) was used to disaggregate daily-resolution data into hourly data for analysis.

$$Hourly\ PET(h) = A.\sin\left(\frac{\pi(\mathrm{h}-6)}{12}\right) \tag{1}$$

Where **A** is a scaling factor so that the sum of hourly PET = daily PET, **h** is the hour in 24-hour format (6 to 17 for 6 AM to 5 PM), and PET = 0 for all hours not in [6, 18].

2.2.2 Gauged Streamflow Data and Geographic Data

Hourly-resolution streamflow data for 15 gauging stations were collected from the Bureau of Meteorology, Australia, for the 2007 to 2017 period (Bureau of Meteorology 2024). Geo-

graphic data, such as surface stream network, surface cartography, surface catchments, and groundwater mapping, were obtained from the National Geofabric version 3.3 (Geoscience Australia 2023) provided by the Bureau of Meteorology, Australia. A one-second SRTM-derived Digital Elevation Model (DEM) was downloaded from Geoscience Australia (Gallant et al. 2011). The surface stream network from the Geofabric dataset was validated against the DEM flow directions in ArcGIS 10.8.2.

2.2.3 Soil Moisture Data

Daily resolution root zone soil moisture data (5 km spatial resolution) were collected from the Australian Water Outlook for the same period (Bureau of Meteorology 2024). Root zone soil moisture represents the percentage of available water content in the top 1 m of the soil profile (Guillory et al. 2023). This value was converted to soil moisture depth (mm), and daily resolution data were disaggregated to hourly resolution for the analysis. Since the soil moisture doesn't fluctuate sharply daily, unlike PET, a uniform distribution was assumed for disaggregation.

2.2.4 GSMaP-NRT Satellite Precipitation Data

Hourly GSMaP NRT precipitation data were obtained from the JAXA Global Rainfall Watch (Japan Aerospace Exploration Agency 2024). Using the point option, the satellite precipitation data corresponding to the precipitation gauging locations were downloaded based on their geographical coordinates.

3 Methodology

3.1 Workflow of the Study

The workflow of the study was divided into three key areas: data processing, machine learning model development, and performance evaluation (Fig. 2). Based on the literature analysis, possible uncertainties in satellite data were identified, followed by data preprocessing and missing data handling. The machine learning model development was primarily handled in Python. A detailed explanation of hyperparameter tuning and optimization is given below. The next section compares the performance of the GSMaP-NRT uncertainty-reduced precipitation data when applied to the calibrated GR4H model (Wanniarachchi and Sarukkalige 2023).

3.2 Data Pre-processing

The gauged precipitation data were verified by comparing them with the data archives of the water measurement information system of the Victorian Department of Environment, Land, Water, and Planning (Department of Environment, Land, Water & Planning 2023). The streamflow data were cross-checked against the Department of Environment, Land, Water, and Planning's water measurement information system archives and found some deviations, especially in the peak flow estimates for a few gauging stations. Missing data filtration was

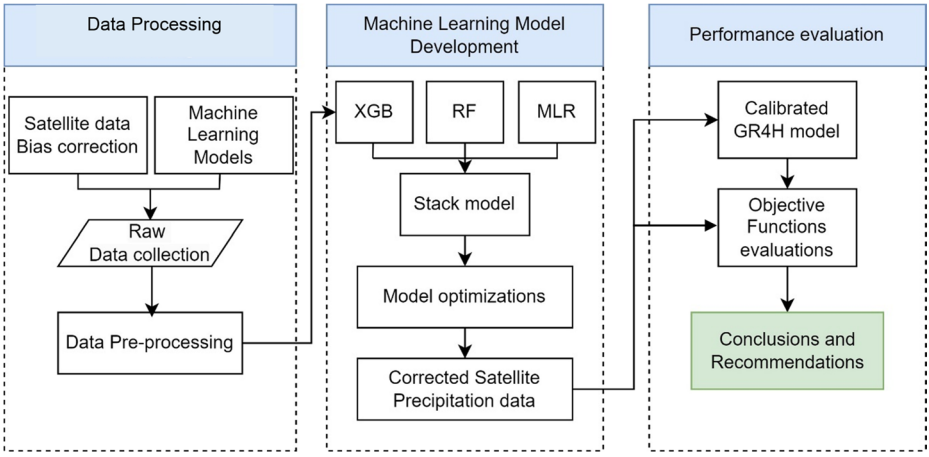

Fig. 2 Methodology flow chart

done for gauged and satellite data to maintain consistency (Wanniarachchi et al. 2026). Necessary precautions were taken to ensure all the data were at the same time standard (UTC). It is useful to analyse the ML model's performance across different precipitation classes to understand the ML stack's ability to reduce uncertainties in satellite precipitation data. Das et al. (2022) have done a similar comprehensive ML study to classify precipitation types. The precipitation data were further categorized into precipitation classes (Table 2) to evaluate the machine learning process in different precipitation classes using the Australian precipitation threshold classification (Kingston 2023).

3.3 Machine Learning Algorithms

The highly non-linear and non-linearly interacting nature of precipitation makes it inappropriate and hard to model using linear relationships. However, the precise design of predictor variables enables Multiple Linear Regression (MLR) to model specific linear relationships within the precipitation process (Generoso et al. 2023). In this study, soil moisture, potential evapotranspiration (PET), satellite-based precipitation estimates, and their time-lagged representations were used as independent variables to develop the MLR model, with observed precipitation as the target variable. The dataset was split into training and testing subsets, with 80% of the data used for model training and 20%

Table 2 Precipitation Threshold Classifications used in different regions

	American Meteorological Society	UK Meteorological Office	Indonesian Meteorological Department	Malaysia (Ziarh et al. 2021)	Australia (Kingston 2023)
No Rain (mm/hr)	-	-	0–1	-	-
Light Rain (mm/hr)	<2.5	<0.5	1–5	<2	<2.5
Moderate (mm/hr)	-	-	5–10	2–10	2.5–10
Heavy Rain (mm/hr)	>7.5	>4	>10	10–20	>10
Extreme (mm/hr)	-	-	>20	>20	-

reserved for independent performance evaluation. The concept of data split has been incorporated by employing random sampling methods to guarantee that the performance of the training and testing data contains samples of the entire spectrum of precipitation classes. However, low levels of precipitation are dominant, with heavy precipitation occurring infrequently. Care was taken to verify that heavy-precipitation events were present in both subsets. This ensured that model evaluation remained meaningful across all precipitation regimes while minimizing sampling bias.

Random Forest (RF) is an ensemble learning method that constructs multiple decision trees using bootstrap aggregation, commonly known as bagging (Barbhuiya et al. 2024). Each tree is trained independently on a randomly sampled subset of the training data (with replacement) and a random subset of predictor variables. This randomness assures the trees are different in nature; hence, each tree will pick up a different aspect of the data. In the case of regression, the final RF prediction is determined by averaging the outputs of all individual trees. This ensemble averaging reduces variance, improves robustness, and limits overfitting compared to single decision-tree models (Generoso et al. 2023; Tang et al. 2023; Zhang et al. 2023).

Extreme Gradient Boosting Regression (XGB) follows a boosting framework based on gradient descent optimization. Unlike bagging-based approaches, boosting builds decision trees sequentially, where each new tree is trained to correct the residual errors of the combined ensemble of previously built trees (Madhushani et al. 2024). At each iteration, XGB computes the gradient of the loss function with respect to the current predictions and fits a new tree to minimize this loss. Regularization terms in the objective function control model complexity by penalizing overly complex trees, thereby reducing overfitting. Additionally, the learning rate controls the contribution of each new tree to the ensemble, enabling the model to learn gradually and improve generalization when smaller learning rates are used (Kim et al. 2021).

Python libraries including *pandas*, *scikit-learn*, *matplotlib*, *pyplot*, and *seaborn* were used for data processing, modelling, and visualization. The initial hyperparameter values of the RF and XGB used for the model, as presented in Table 3, were optimized during the training period. Through the use of various cross-validation processes, the model was optimized using the XGB early stopping method, ensuring the optimal values of the param-

Table 3 Hyperparameters of Random Forest and XGB algorithms

Random Forest Hyperparameters		XGB Hyperparameters	
n_estimators	[100, 200, 300]	n_estimators	[100, 200, 300]
max_features	['sqrt', 'log2']	learning_rate	[0.01, 0.05, 0.1]
max_depth	[None, 10, 20, 30]	max_depth	[3, 6, 9]
min_sample_split	[2, 5, 10]	subsample	[0.7, 0.8, 0.9]
min_sample_leaf	[1, 2, 4]	colsample_bytree	[0.7, 0.8, 0.9]

eters were found through the stopping of the training process when the improvement in the metrics reached a point where it became negligible, providing a delicate balance between the accuracy of the model and the computational efficiency of the model.

To further enhance the predictive performance, a stacking ensemble framework is applied (Das et al. 2022). Stacking ensemble is a more powerful ensemble learning method. In the stacking ensemble approach, a set of heterogeneous base learners is learned individually. In the present study, MLR, RF, and XGB algorithms are applied as base learners. The application of MLR and RF is to ensure the diversity of the models by incorporating a linear and a tree-based learning paradigm to learn the characteristics of the precipitation process from different angles.

In the stacking mechanism, the predictions from the base learners were obtained through a form of cross-validation, where the data from the training set was utilized. This meant that the data from the training set was divided into k folds, and for each fold, the base learners were trained on k-1 folds and then used to predict the remaining fold. This ensured the prediction from each of the base learners used for the meta-model was out-of-sample, preventing data leakage. The cross-validated data from each of the base learners was then used for the formation of the meta-model.

For instance, the meta-model was designed in a way that enabled it to learn the optimal combination of the base learner combinations in a way that best approximated the observed precipitations. By doing so, the stacking learning framework effectively reduced the bias and variance, thereby yielding a better performance in terms of generality. The performance of the stacked learning model was then evaluated on the independent set of data, which was never used when learning the base learners and the meta-model.

3.4 GR4H Hydrological Model

The GR4J four-parameter parsimonious model, which relates runoff (Q) to rainfall (P) and evapotranspiration (ET) in daily resolution (Perrin et al. 2003), was identified as a potential hydrological model. The GR4H model is a variant of GR4J, having the same model structure but operating with hourly data (Ayzel and Heistermann 2021; Basri et al. 2019; Han et al. 2019). Wanniarachchi and Sarukkalige (2023) provided the details of the model schematic, model parameters, units and the calibration procedure.

3.5 Evaluation of Model Efficiency

Objective functions are essential for measuring prediction accuracy because they are sensitive to outliers and can capture overall variability (Lin et al. 2024). The choice of an objective function is a critical step in the model calibration process, as it directly influences the behaviour of the optimized model (Huo and Liu 2020; Lin et al. 2024).

The Nash-Sutcliffe Efficiency (NSE) provides a normalized measure of model performance relative to a benchmark, typically the mean of the observed parameter (Lin et al. 2024). While NSE evaluates the model's ability to replicate the full range of streamflow, it tends to favour peak flows. In contrast, NSE(log) places more emphasis on low flows (Huo and Liu 2020). Root Mean Squared Error (RMSE) reflects the sum of squared errors, giving greater weight to larger errors due to squaring, which makes RMSE less robust in the pres-

ence of outliers (Huo and Liu 2020). The Kling-Gupta Efficiency (KGE) integrates multiple performance aspects, including correlation, bias, and variability, into a single metric, and a few recent studies (Lin et al. 2024) considered that KGE to provide a more balanced evaluation of the accuracy and reliability of streamflow predictions, offering valuable insights for model refinement and decision-making.

In this study, the ML algorithms were optimized with a grid search approach, and optimum hyperparameters were identified for each algorithm. The performance of the stack approach of the ML models was initially validated by comparing the corrected satellite precipitation with gauge precipitation. Further analysis of the corrected satellite precipitation was conducted using the calibrated GR4H model. While keeping the model parameters and other input data constant, only the precipitation input has been changed to observe the model output. The behaviour of the corrected satellite precipitation was analysed in terms of the GR4H model-generated streamflow using performance measures such as correlation coefficient (R^2), NSE, NSE(log), RMSE, Bias, and KGE. The mathematical expressions of these six objective functions are provided in Eqs. (2)–(7), respectively:

$$R^2 = 1 - \frac{\sum_{I=1}^{n} \left(Q_{obs,i} - Q_{stat\_model,i}\right)^2}{\sum_{i=1}^{n} \left(Q_{obs,i} - \overline{Q_{obs}}\right)^2} \tag{2}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(Q_{obs,i} - Q_{sim,i}\right)^2}{\sum_{i=1}^{n} \left(Q_{obs,i} - \overline{Q_{obs}}\right)^2} \tag{3}$$

$$NSE(\log) = 1 - \frac{\sum_{i=1}^{n} \left(\log Q_{obs,i} - \log Q_{sim,i}\right)^2}{\sum_{i=1}^{n} \left(\log Q_{obs,i} - \overline{\log Q_{obs}}\right)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(Q_{obs,i} - Q_{sim,i}\right)^2}{N}} \tag{5}$$

$$Bias = \frac{\sum_{i=1}^{n} Q_{sim}}{\sum_{i=1}^{n} Q_{obs}} - 1 \tag{6}$$

$$\begin{gathered} KGE = 1\sqrt{\left((r-1)^2 + (a-1)^2 + (\beta-1)^2\right)} \\ Where;\ r = \frac{1}{N}\sum_{i=1}^{n} \frac{(Q_{obs,i} - \overline{Q}_{obs})(Q_{sim,i} - \overline{Q}_{sim})}{\sigma_{obs}\sigma_{sim}} \\ \alpha = \frac{\sigma_{sim}}{\sigma_{obs}},\ \beta = \frac{\overline{Q}_{sim}}{\overline{Q}_{obs}} \end{gathered} \tag{7}$$

Where r is the Pearson product-moment correlation coefficient, $\bar{Q}$ and σ are the mean and the standard deviation of the observed and simulated time series, respectively. α is the ratio between the simulated and observed standard deviation values, and β is the ratio between the simulated and observed mean values. When r, α, and β tend towards 1, the KGE criterion is maximized (Garcia et al. 2017).

4 Results and Discussion

This section first compares the raw satellite precipitation with gauge observations to identify systematic discrepancies. The behaviour of the ML-corrected satellite precipitation is then examined across three precipitation classes, followed by an evaluation of its implications for streamflow simulation.

The scatter plot of the sub-area weighted average gauged precipitation with the GSMaP-NRT raw precipitation data of the Ovens basin (Fig. 3) shows substantial dispersion from the 1:1 line. GSMaP-NRT data generally underestimates most of the gauged precipitation records. A limited number of cases exhibit overestimation, where satellite values exceed the corresponding gauge measurements. This pattern remains evident at the hourly resolution, suggesting that the discrepancy is not solely a temporal-aggregation effect.

After training the ML stacking model, the relationship between corrected GSMaP-NRT precipitation and gauge precipitation improved (Fig. 4a). Compared with Fig. 3, the corrected data points are more closely aligned with the 1:1 line, as reflected in the objective functions presented in Table 4. The correction procedure also reduces the occurrence of abnormally high satellite precipitation values. Nevertheless, the highest precipitation peaks remain underestimated, indicating that the bias correction is more effective for moderate ranges than for extremes.

The distribution of the corrected precipitation data was further analyzed using Kernel density estimation (KDE) and shown in Fig. 4(b). The KDE shows an elongated concentration along the diagonal, indicating stronger agreement between corrected satellite and gauge precipitation within the dominant low-intensity regime. The intermediate (2–6 mm/hour) and heavy (>6 mm/hour) precipitation classes are less represented in the density plot; therefore, the model performance in these regimes is evaluated separately in Fig. 5.

Figure 5 indicates that the intermediate precipitation class exhibits improved agreement after correction, although the dispersion remains larger than in the low-intensity class.

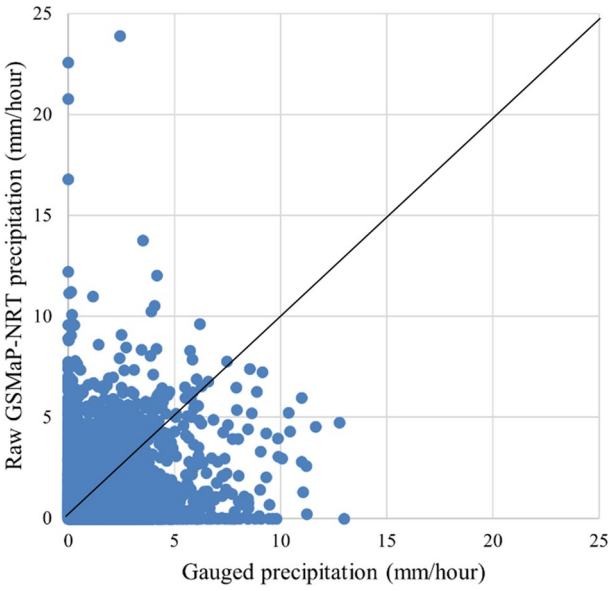

Fig. 3 Scatter plot of GSMaP-NRT Precipitation Vs Gauged Precipitation

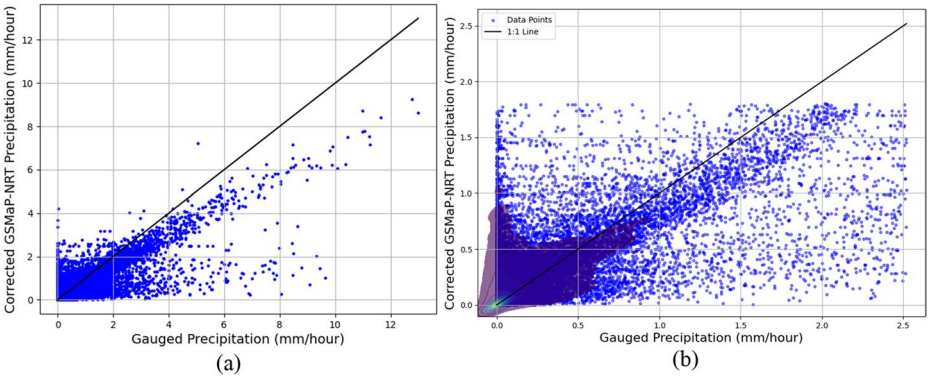

Fig. 4 Scatter plot (**a**) and the density plot (**b**) of GSMaP-NRT Corrected Precipitation Vs Gauged Precipitation

Table 4 Behaviour of objective functions in Satellite Precipitation bias correction and Streamflow predictions

Compared with the Gauged Precipitation data	NSE	NSE(log)	R^2	BIAS	RMSE	KGE
Raw GSMaP-NRT precipitation	-0.301	-0.321	0.554	-0.035	0.589	0.586
Corrected GSMaP-NRT precipitation	0.349	0.371	0.842	0.000	0.403	0.417
Corrected GSMaP-NRT precipitation (Lower class)	0.476	0.513	0.922	0.002	0.170	0.556
Corrected GSMaP-NRT precipitation (Intermediate class)	0.164	0.163	0.836	0.027	0.802	0.156
Corrected GSMaP-NRT precipitation (Heavy class)	0.026	0.034	0.731	-0.133	1.391	-0.178
Compared with the Gauged Streamflow at the catchment outlet	**NSE**	**NSE(log)**	**R^2**	**BIAS**	**RMSE**	**KGE**
Simulated Streamflow with the gauged precipitation	0.814	0.21	0.83	0.001	0.019	0.865
Simulated Streamflow with raw GSMaP-NRT precipitation	0.245	-0.365	0.356	-0.013	0.039	0.491
Simulated Streamflow with corrected GSMaP-NRT precipitation	0.766	0.069	0.795	-0.013	0.022	0.826
Simulated Streamflow with corrected GSMaP-NRT precipitation vs. simulated streamflow with the gauged precipitation	0.948	0.953	0.954	-0.002	0.008	0.934

In contrast, the heavy precipitation class shows a weak or negative trend, and several corrected values still fall below the 1:1 line. This behaviour coincides with the limited number of heavy rainfall observations (approximately 0.4% of the dataset). Given this strong class imbalance, the model is primarily trained on low-intensity events, which likely contributes to reduced performance in the heavy precipitation regime. These findings suggest that the stacking model is more stable when trained on the full dataset rather than separately within sparsely populated classes. Consistent with the graphical evidence, the objective functions (Table 4) indicate stronger performance in the low-precipitation class, with higher NSE, R^2, and KGE values than in intermediate and heavy classes. The heavy precipitation class exhibits larger RMSE and a persistent negative

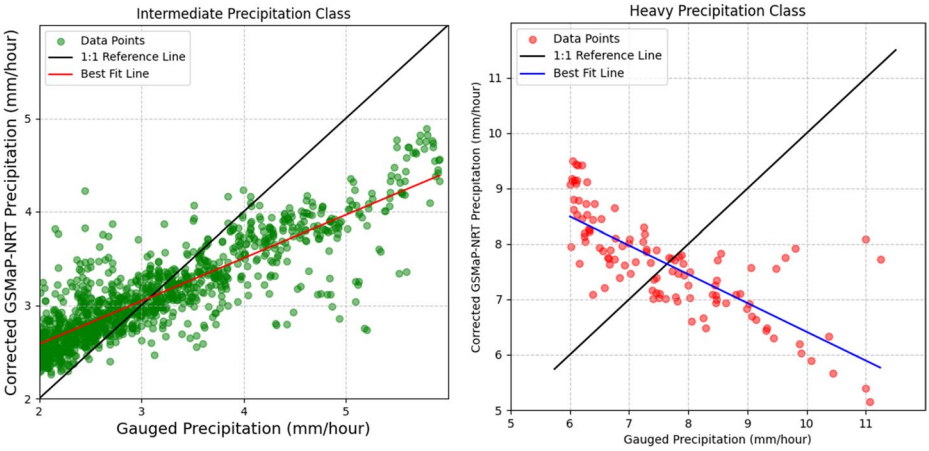

Fig. 5 Scatter plot of GSMaP-NRT Corrected Precipitation Vs Gauged Precipitation for Intermediate and Heavy precipitation classes

bias. Because extreme rainfall events are rare yet hydrologically critical, these residual errors remain a key limitation of the bias-correction framework.

The scatter plot comparing observed streamflow at the catchment outlet with GR4H-simulated streamflow using gauge precipitation input (Fig. 6a) shows that the model captures the general trend but underestimates peak flows by approximately 40%. This discrepancy is consistent with previously reported uncertainties associated with spatial precipitation interpolation, parameter estimation, and model structural limitations (Moges et al. 2021; Panchanathan et al. 2024). When corrected GSMaP-NRT precipitation is used as input (Fig. 6b), the overall correlation and pattern remain similar to the simulation driven by gauge precipitation. In both cases, the model reproduces the general variability but shows

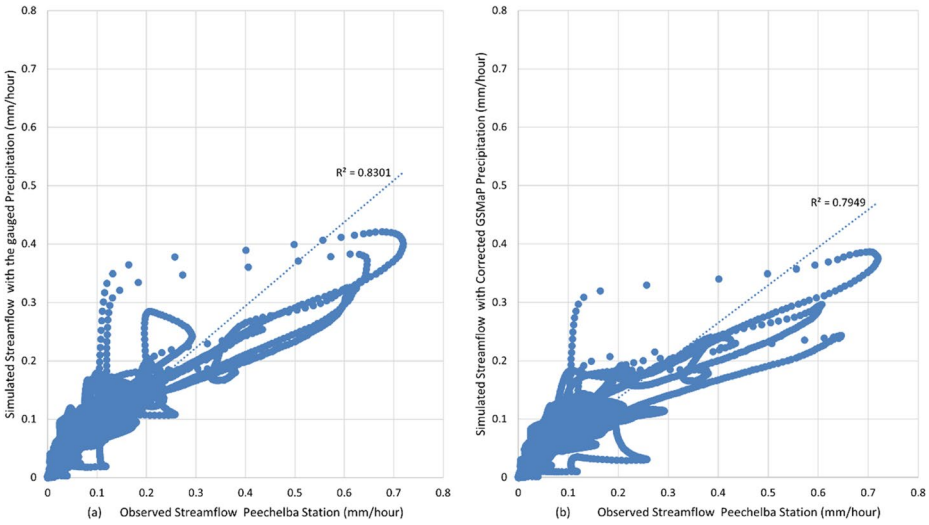

Fig. 6 Scatter plot of observed streamflow vs the simulated streamflow for Peechelba station

greater dispersion at intermediate and high flows. This behaviour may be related to the calibration strategy of the GR4H model, where parameter optimization based on NSE(log) tends to prioritize overall water balance and low-to-moderate flows rather than peak flows.

A one-to-one comparison between gauge precipitation and corrected GSMaP-NRT precipitation (Fig. 4a) already indicated underestimation of peak rainfall. Therefore, some underestimation of peak streamflow is expected in the hydrological simulations. Despite this, Fig. 7 shows that the streamflow simulated with corrected satellite precipitation is highly consistent with that driven by gauge precipitation ($R^2 = 0.954$). The clustering of points near the 1:1 line indicates that both precipitation inputs produce comparable streamflow dynamics, although increased dispersion is observed at intermediate flows. This pattern is also visible in Fig. 6. These results suggest that the corrected GSMaP-NRT precipitation provides a hydrologically consistent input for the GR4H model. However, comparison with observed streamflow (Fig. 6) confirms that major peaks remain underestimated. The hydrograph comparison over the whole analysis period (Fig. 8) further illustrates this limitation.

During the largest recorded flood event in the Ovens basin (2007–2017), with an observed peak of 1236.4 m³/s, the simulated peak improved substantially after precipitation correction from 77.34 m³/s (raw satellite input) to 639.36 m³/s. The simulation driven by gauge precipitation produced a peak of 721.86 m³/s, which is still considerably lower than the observed value. These results indicate that both precipitation uncertainty and hydrological model limitations contribute to the underestimation of peak flows.

At the cumulative scale, GSMaP-NRT precipitation accumulates to 7,985.9 mm over the study period, while the gauge estimates accumulate to 11,099.65 mm, validating the previous findings of underestimation. After ML-based correction using soil moisture and PET

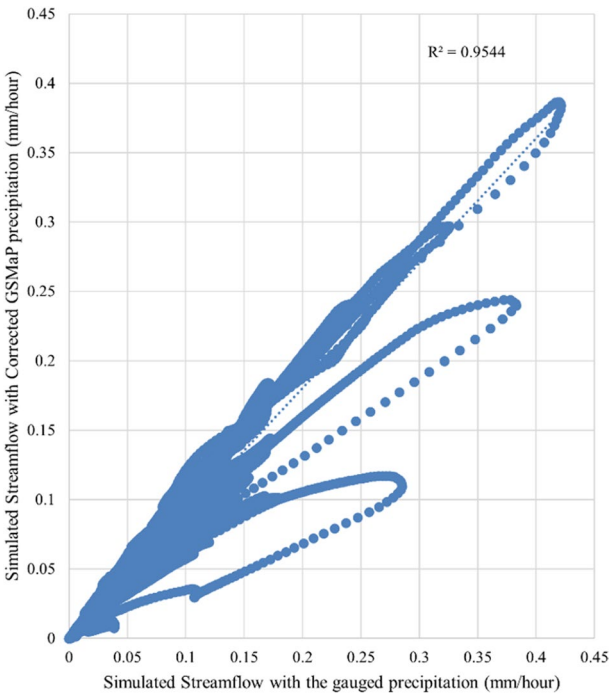

Fig. 7 Scatter plot of simulated streamflow with the gauged precipitation Vs the corrected GSMaP precipitation

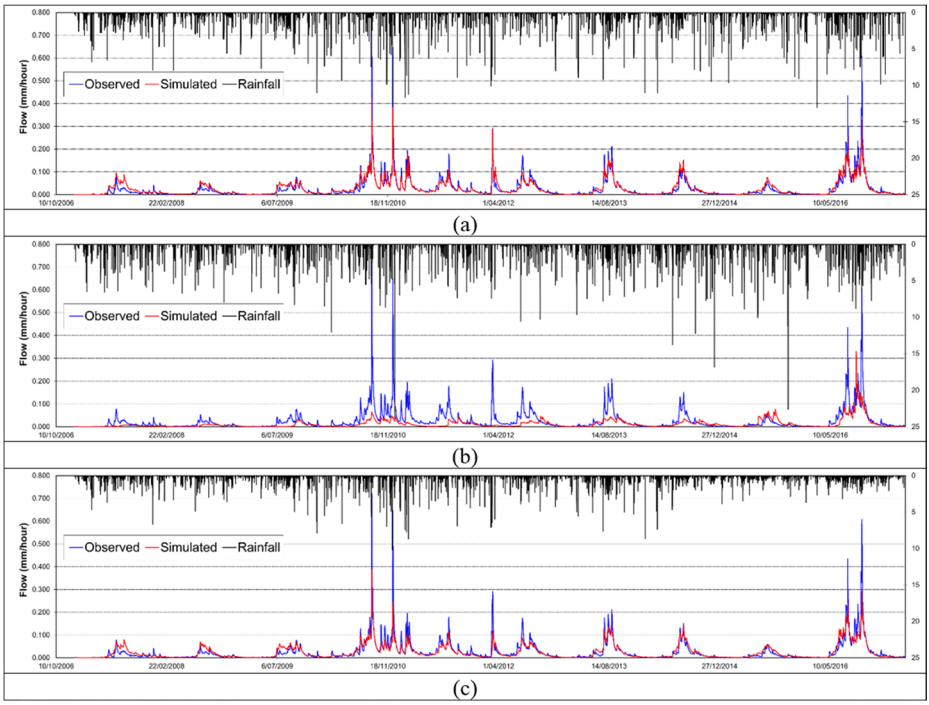

Fig. 8 Observed and simulated Streamflow hydrographs for the 2007-2017 period using precipitation data:(**a**) gauged; (**b**) GSMaP-NRT; (**c**) GSMaP-NRT corrected

as supplementary variables, the cumulative precipitation increases to 11,197 mm, closely matching the gauge-derived volume while maintaining a realistic temporal distribution.

Figure 8 shows that insufficient antecedent and event precipitation in the raw satellite product leads to reduced simulated peak flows, particularly during the September 2010 and September 2016 events. Differences between similar streamflow peaks despite varying gauge-derived precipitation suggest additional uncertainty in spatial rainfall representation. The uneven distribution of rain gauges, especially in the southeastern alpine region (Fig. 1), may cause interpolation errors and non-representative catchment averages. An unusually high satellite estimates further indicate residual sensor and atmospheric uncertainties.

Overall, these findings highlight that while the ML stacking framework substantially reduces systematic bias in GSMaP-NRT precipitation and improves hydrological consistency, challenges remain in accurately representing extreme precipitation and peak flood responses. Also, it established the need for bias-correction approaches that explicitly account for class imbalance, spatial sampling limitations, and event-scale dynamics.

5 Limitations and Future Work

The major challenge in this study was the higher-resolution data, leading to a significant temporal mismatch between satellite estimates and gauged data. To be able to overcome this limitation, all model features were tested up to a 3-hour time lag. It is recommended,

however, that further testing at coarser resolutions and over longer timescales should be carried out to examine the robustness of the trained machine learning model across catchments. The performance in representing heavy rainfall events may be improved, while event-based or stratified training strategies may be considered for future work, along with an investigation into the robustness of the models across a wide range of climatic and hydrological regimes. This approach will help ensure consistency in the findings and allow the application of the GSMaP-NRT precipitation data with reduced uncertainty for data-scarce regions. Considering the incorporation of explainable machine learning (XAI) into future works is necessary for interpreting and justifying the selection of hydrological parameters used in the model training.

6 Conclusions

In this research, the Random Forest (RF), Multiple Linear Regression (MLR), and Extreme Gradient Boosting (XGB) models were used. Root-zone soil moisture and PET were used in the models to reduce the uncertainty in GSMaP-NRT precipitation estimates for the Ovens Basin. Model performance was evaluated using six objective functions as well as the calibrated GR4H hydrological model simulations.

The results show that raw GSMaP-NRT precipitation exhibits a negative bias, high RMSE, and negative NSE values. The machine-learning-based correction substantially reduces the bias; however, RMSE remains comparatively high for the intermediate and heavy precipitation classes. All six objective functions achieved their optimal ranges primarily under low precipitation conditions, indicating that the proposed framework performs more reliably for frequent, low-intensity rainfall events than for extreme rainfall events.

Streamflow simulations using the GR4H model show that peak discharge estimated with bias-corrected satellite precipitation remains 48% underestimated, compared with 41% with gauged precipitation. However, this represents a significant improvement over the 93% underestimation observed with raw GSMaP-NRT data. These inconsistencies suggest that remaining errors are influenced not only by limitations in precipitation correction but also by satellite retrieval uncertainty, the strong imbalance in extreme rainfall events in the dataset, and the hydrological model's sensitivity to peak rainfall events.

Despite these limitations, the strong agreement ($R^2 = 0.9544$) between simulations driven by gauged precipitation and corrected satellite precipitation indicates that the proposed machine-learning framework can improve the consistency of near-real-time satellite rainfall estimates. The addition of root-zone soil moisture, PET, and their time-lagged representations appears to improve the model's ability to reproduce major hydrological signals, particularly for low to moderate precipitation conditions.

Overall, the application of the stacking-based machine learning framework together with the GR4H hydrological model demonstrates the potential of integrating data-driven correction with process-based modelling for applications where ground observations are limited. However, the results should be interpreted with caution, due to strong class imbalance (with most samples from low-rainfall scenarios), the temporal disaggregation assumptions, and the limited representation of extreme events may affect the generalizability of the approach.

Acknowledgements The authors are thankful for the support from the Bureau of Meteorology of Australia and the Japan Aerospace Exploration Agency for providing the weather data.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Susantha Shameera Wanniarachchi. The first draft of the manuscript was written by Susantha Shameera Wanniarachchi and Ranjan Sarukkalige; all authors commented and made corrections on previous versions of the manuscript. The software, model development, and validation were performed by Susantha Shameera Wanniarachchi and H.A.Prasantha Hapuarachchi. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. The authors declare that no funds, grants, or other financial support were received during the preparation of this manuscript.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Abatzoglou J (2021) Climatology Lab, TerraClimate. https://www.climatologylab.org/terraclimate.html. Accessed 22 04 2025

Ali S, Cheema MJM, Bakhsh A, Khaliq T (2020) Near Real-time Flood Forecasting in the Transboundary Chenab River using Global Satellite Mapping of Precipitation. Pakistan J Agricultural Sci 57(5):1327–1335

Amini A, Dolatshahi M, Kerachian R (2024) Real-time rainfall and runoff prediction by integrating BC-MODWT and automatically-tuned DNNs: Comparing different deep learning models. J Hydrol 631(130804):1–14

Ayzel G, Heistermann M (2021) The effect of calibration data length on the performance of a conceptual hydrological model versus LSTM and GRU: A case study for six basins from the CAMELS dataset. Comput Geosci 149(104708)

Barbhuiya S, Manekar A, Ramadas M (2024) Performance evaluation of ML techniques in hydrologic studies: Comparing streamflow simulated by SWAT, GR4J, and state-of-the-art ML-based models. J Earth Syst Sci 133(136):1–16

Basri H, Sidek LM, Saad S, Sammen A, Razad AZ, Pokhrel P (2019) Hydrological Modelling of Surface Runoff for Temengor Reservoir Using GR4H Model. Int J Civil Eng Technol 10(7):29–40

Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q (2023) Accurate medium-range global weather forecasting with 3D neural networks. Nat Volume 619:533–538

Bureau of Meteorology (2024) Australian Government Bureau of Meteorology. http://www.bom.gov.au/climate/data/stations/about-weather-station-data.shtml. Accessed 05 05 2024

Bureau of Meteorology (2024) Australian Water Outlook. https://awo.bom.gov.au/products/forecast/soilMoisture-rootZone/9,-36.698,146.891/riv,-36.732,146.594/a/m/2023-10. Accessed 30 07 2024

Chua ZW, Evans A, Kuleshov Y, Watkins A, Choy S, Sun C (2022) Enhancing the Australian Gridded Climate Dataset rainfall analysis using satellite data. Sci Rep 12(20691)

CSIRO (2008) Water availability in the Ovens. A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project. CSIRO, Australia., s.l., pp 25–26

Das S, Wang Y, Gong J, Ding L, Munchak SJ, Wang C, Wu DL, Liao L, Olson WS, Barahona DO (2022) A Comprehensive Machine Learning Study to Classify Precipitation Type over Land from Global Precipitation Measurement Microwave Imager (GPM-GMI) Measurements. Remote sensing 1(0): 1–24

Department of Environment, Land, Water &, Planning (2023) Water Monitoring. https://data.water.vic.gov.au/. Accessed 10 9 2023

Emmanuel R, Jalal M, Ogunfuyi S, Maharoof N, Zala M, Perera N, Ratnayake R (2023) Urban Heat Risk: Protocols for Mapping and Implications for Colombo, Sri Lanka. Atmosphere 14(343):1–35

Fowdur TP, Nazir RMNUDI (2022) A real-time collaborative machine learning based weather forecasting system with multiple predictor locations. Array 14(100153):1–13

Gallant J, Tickle P, Wilson N, Inskeep C, Dowling T, Read A (2011) 1 second SRTM Derived Digital Elevation Models User Guide. Geoscience Australia, Canberra ACT 2601: Geoscience Australia: 5–11

Garcia F, Folton N, Oudin L (2017) Which objective function to calibrate rainfall–runoff models for low-flow index simulations? Hydrol Sci J 62(7):1149–1166

Generoso TN, da Silva DD, Amorim RSS, Rodrigues LN, Althoff D, dos Santos EP (2023) Forecasting of daily streamflows downstream from reservoirs with streamflow regularization using machine learning methods. J S Am Earth Sci 130(104583):1–13

Geoscience Australia (2023) Digital Atlas of Australia- Australian Hydrological Geospatial Fabric - Water Bodies. Geoscience Australia, Canberra

Goulburn-Murray W (2024) Goulburn-Murray Water-Ovens Basin. https://www.g-mwater.com.au/water-operations/catchments/ovensbasin. Accessed 30 07 2024

Guillory L, Pudmenzky C, Nguyen-Huy T, Cobon D, Stone R (2023) A drought monitor for Australia. Environ Model Softw 170(105852):1–14

Han S, Coulibaly P, Biondi D (2019) Assessing Hydrologic Uncertainty Processor Performance for Flood Forecasting in a Semiurban Watershed. J Hydrol Eng 24(9):1–18

Hinge G, Hamouda MA, Long D, Mohamed MM (2022) Hydrologic utility of satellite precipitation products in flood prediction: A meta-data analysis and lessons learnt. J Hydrol 612(Part A).

Huo J, Liu L (2020) Evaluation Method of Multiobjective Functions' Combination and Its Application in Hydrological Model Evaluation. Comput Intell Neurosci 2020:1–23

Japan Aerospace Exploration Agency (2024) Jaxa Global Rainfall Watch. https://sharaku.eorc.jaxa.jp/GSMaP/index.htm. Accessed 5 5 2024

Kim JH, Kim J, Lee G, Park J (2021) Machine Learning-Based Models for Accident Prediction at a Korean Container Port. Sustainability 13(16):1–14

Kingston C (2023) Change to Rainfall Intensity Thresholds used in Bureau of Meteorology Surface Observations (METAR, SPECI and AWIS). https://www.avsef.gov.au/consultations/change-rainfall-intensity-thresholds-used-bureau-meteorology-surface-observations-metar-speci-and-awis#:~:text=Light%20(%2D)%20intensity%20is%20less,for%20frozen%20precipitation%20or%20drizzle. Accessed 18 06 2024

Kumar A, Ramsankaran R, Brocca L, Munoz-Arriola F (2019) A Machine Learning Approach for Improving Near-Real-Time Satellite-Based Rainfall Estimates by Integrating Soil Moisture. Remote Sens 11(2221):1–20

Li X, Chen S, Liang Z, Huang C, Li Z, Hu B (2021) Performance Assessment of GSMaP and GPM IMERG Products during Typhoon Mangkhut. Atmosphere 12(134):1–13

Li Y, Yan H, Chen L, Huang M, Shou W, Zhu L, Zhao L, Xing Y (2024) Performance and uncertainties of five popular satellite-based precipitation products in drought monitoring for different climate regions. J Hydrol 628(130562):1–17

Lin Y, Wang D, Zhu J, Sun W, Shen C, Shangguan W (2024) Development of objective function-based ensemble model for streamflow forecasts. J Hydrol 632(130861)

Llauca H, Lavado-Casimiro W, León K, Jimenez J, Traverso K, Rau P (2021) Assessing Near Real-Time Satellite Precipitation Products for Flood Simulations at Sub-Daily Scales in a Sparsely Gauged Watershed in Peruvian Andes. Remote Sens 13(4)

Madhushani C, Dananjaya K, Ekanayake IU, Meddage DPP, Kantamaneni K, Rathnayake U (2024) Modeling streamflow in non-gauged watersheds with sparse data considering physiographic, dynamic climate, and anthropogenic factors using explainable soft computing techniques. J Hydrol 631(130846):1–19

Massari C, Camici S, Ciabatta L, Brocca L (2018) Exploiting Satellite-Based Surface Soil Moisture for Flood Forecasting in the Mediterranean Area: State Update Versus Rainfall Correction. Remote Sens 10(2):1–21

Moges E, Demissie Y, Larsen L, Yassin F (2021) Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. Water 13(1):1–28

Murray-Darling Basin Authority (2023) Murray-Darling Basin Authority-Ovens catchment. https://www.mdba.gov.au/basin/catchments/southern-basin-catchments/ovens-catchment. Accessed 17 09 2023.

NASA (2020) Global Precipitation Measurement. https://gpm.nasa.gov/data/imerg. Accessed 30 05 2025
National Weather Service (2006) CMAP. https://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cmap.shtml#:~:text=The%20CPC%20Merged%20Analysis%20of,algorithms%20(infrared%20and%20microwave). Accessed 30 05 2025
National Center for Atmospheric Research UCAR (2025) NOAA CPC Morphing Technique (CMORPH) Global Precipitation Analyses. https://rda.ucar.edu/datasets/ds502.0/. Accessed 30 05 2025
Nguyen NY, Anh TN, Nguyen HD, Dang DK (2024) Quantile mapping technique for enhancing satellite-derived precipitation data in hydrological modelling: a case study of the Lam River Basin, Vietnam. J Hydroinformatics 26(8):2026–2044
NOAA (2025) Global Precipitation Climatology Project (GPCP) Clearinghouse. https://www.ncei.noaa.gov/products/global-precipitation-climatology-project. Accessed 30 05 2025
Panchanathan A, Ahrari A, Ghag KS, Mustafa S, Haghighi AT, Kløve B, Oussalah M (2024) An overview of approaches for reducing uncertainties in hydrological forecasting: Progress and challenges. Earth Sci Rev 258(104956)
Parajuli A, Parajuli R, Banjara M, Bhusal A, Dahal D, Kalra A (2024) Application of Machine Learning and Hydrological Models for Drought Evaluation in Ungauged Basins Using Satellite-Derived Precipitation Data. Climate 12(190):1–22
Perrin C, Michel C, Andréassian V (2003) Improvement of a parsimonious model for streamflow simulation. J Hydrology Volume 279:275–289
Priyambodoho BA, Kure S, Yagi R, Januriyadi NF (2021) Flood inundation simulations based on GSMaP satellite rainfall data in Jakarta, Indonesia. Progress Earth Planet Sci 8(34):1–17
Rachidi S, Mazoudi EHE, El Alami J, Jadoud M, Er-Raki S (2023) Assessment and Comparison of Satellite-Based Rainfall Products: Validation by Hydrological Modeling Using ANN in a Semi-Arid Zone. Water 15(11):1–18
Singh AK, Singh V (2024) Assessing the accuracy and reliability of satellite-derived precipitation products in the Kosi River basin (India). Environ Monit Assess 196(671):1–28
Spassiani AC, Mason MS, Cheng VYS (2023) An Australian convective wind gust climatology using Bayesian hierarchical modelling. Nat Hazards Volume 118:2037–2067
Tang Y, Sun Y, Han Z, Soomro SEH, Wu Q, Tan B, Hu C (2023) flood forecasting based on machine learning pattern recognition and dynamic migration of parameters. J Hydrology: Reg Stud 47(101406):1–13
University of California SB (2025) CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations. https://www.chc.ucsb.edu/data/chirps. Accessed 30 05 2025
University of Reading (2025) Tropical Applications of Meteorology using Satellite and ground-based observations. https://research.reading.ac.uk/tamsat/. Accessed 20 04 2025
Wang S, Zhang K, Chao L, Li D, Tian X, Bao H, Chen G, Xia Y (2021) Exploring the utility of radar and satellite-sensed precipitation and their dynamic bias correction for integrated prediction of flood and landslide hazards. J Hydrol 603(126964):1–15
Wanniarachchi S, Sarukkalige R (2023) Enhancing the Flood Modeling Accuracy Under Varying Spatial Scales Finer Resolution Weather data. Kandy, pp 13–24
Wanniarachchi S, Sarukkalige R, Hapuarachchi HAP, Gomes PIA, Rathnayake U (2026) Enhancing the effectiveness of satellite precipitation products with topographic and seasonal bias correction. J Hydrol 665(134688). https://doi.org/10.1016/j.jhydrol.2025.134688
Zhang T, Liang Z, Wang H, Wang J, Hu Y, Li B (2023) Merging multisatellite precipitation products using stacking method and the censored-shifted gamma ensemble model output statistics in china's Beimiaoji basin. J Hydrol 618(129263):1–11
Zhao B, Gao D, Xu L (2025) Leakage risk prediction of hydrogen blended natural gas station based on improved grey support vector machine optimization by Pelican optimization algorithm. International Journal of Hydrogen Energy 163(150826). Retrieved from https://doi.org/10.1016/j.ijhydene.2025.150826
Ziarh GF, Shahid S, Ismile TB, Asaduzzaman M, Dewan A (2021) Correcting bias of satellite rainfall data using physical empirical model. Atmos Res 251(105430):1–15

Authors and Affiliations

Susantha Wanniarachchi[1,2] · Ranjan Sarukkalige[1] · H. A. Prasantha Hapuarachchi[3] · Pattiyage I.A. Gomes[2] · Upaka Rathnayake[4]

✉ Susantha Wanniarachchi
s.wanniarachchi@postgrad.curtin.edu.au

[1] School of Civil and Mechanical Engineering, Curtin University, Perth, Western Australia, Australia

[2] Department of Civil Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

[3] The Bureau of Meteorology, Melbourne, Australia

[4] Department of Civil Engineering and Construction, Atlantic Technological University, Sligo, Ireland