An Investigation into the Suitability of Gauge-Corrected Remotely Sensed Rainfall Datasets for Hydrological Modelling in the Western Ghats

Robyn HORAN[1], Jeff Smithers[1], David J. Clark[1], Mark JC Horan[1], Thomas Rodding Kjeldsen[2], and Nathan Rickards[3]

[1]University of KwaZulu-Natal
[2]University of Bath
[3]UK Centre for Ecology & Hydrology

January 23, 2024

Abstract

An accurate spatial and temporal representation of rainfall is essential for hydrological assessments and water resources management. Rainfall is monitored in India's mountainous Western Ghats region via in-situ rainfall gauging stations maintained by the Indian Meteorological Department (IMD). However, the network is sparse, and significant periods of data are missing. Furthermore, the IMD gridded rainfall dataset is known to underestimate the depth of rainfall at the high altitudes within this region. In this study, rainfall estimated by the IMD grids and from remote sensing using the CHIRPS (0.25- and 0.05-degree), MSWEP and PERSIANN datasets are compared to the IMD in-situ gauged rainfall within the Western Ghats using a point-to-pixel analysis.

The GWAVA model is utilised to determine the effect of the selected rainfall input datasets on representing wider water resources. It was found that the average ensemble provided the best representation of the in-situ gauged and catchment rainfall and a better representation than the IMD grids. It remains critical for water resources management to ensure that in-situ rainfall gauging networks are maintained. In-situ data sources of high confidence remain important for the continuous development and ground-truthing of different rainfall datasets.

Hosted file

982620_0_art_file_11704059_s5wt69.docx available at https://authorea.com/users/711305/articles/695808-an-investigation-into-the-suitability-of-gauge-corrected-remotely-sensed-rainfall-datasets-for-hydrological-modelling-in-the-western-ghats

An Investigation into the Suitability of Gauge-Corrected Remotely Sensed Rainfall Datasets for Hydrological Modelling in the Western Ghats

R. Horan[1], J.C. Smithers[1], D. Clark[1], M.J.C Horan[1], T.R. Kjeldsen[2] and N.J. Rickards[3]

[1]Centre for Water Resources Research, University of KwaZulu-Natal, Pietermaritzburg 3201, South Africa

[2]Department of Architecture and Civil Engineering, University of Bath, Bath BA2 7AY, United Kingdom

[3]UK Centre for Ecology & Hydrology, Wallingford OX10 8BB, United Kingdom

Corresponding Author: Robyn Horan, robynhoran8@gmail.com

Key Points:

- The spatial scale of the rainfall dataset does not necessarily affect the performance in the high-altitude regions of the Upper Cauvery Catchment.
- The rainfall in the Upper Cauvery Catchment does not have a distinct correlation to the altitude but correlates strongly to the aspect of the mountains.
- None of the individual remotely sensed datasets tested could be utilised with confidence in the Upper Cauvery Catchment.

Abstract

An accurate spatial and temporal representation of rainfall is essential for hydrological assessments and water resources management. Rainfall is monitored in India's mountainous Western Ghats region via in-situ rainfall gauging stations maintained by the Indian Meteorological Department (IMD). However, the network is sparse, and significant periods of data are missing. Furthermore, the IMD gridded rainfall dataset is known to underestimate the depth of rainfall at the high altitudes within this region. In this study, rainfall estimated by the IMD grids and from remote sensing using the CHIRPS (0.25- and 0.05- degree), MSWEP and PERSIANN datasets are compared to the IMD in-situ gauged rainfall within the Western Ghats using a point-to-pixel analysis.

The GWAVA model is utilised to determine the effect of the selected rainfall input datasets on representing wider water resources. It was found that the average ensemble provided the best representation of the in-situ gauged and catchment rainfall and a better representation than the IMD grids. It remains critical for water resources management to ensure that in-situ

rainfall gauging networks are maintained. In-situ data sources of high confidence remain important for the continuous development and ground-truthing of different rainfall datasets.

1. Introduction

Knowledge of the spatial and temporal distribution of rainfall is essential for hydro-climatic studies. However, many regions are subject to highly variable rainfall, and those vulnerable to climate extremes are among the most data sparse (Wambura, 2020). Many catchments, particularly in the developing world, lack sufficient rainfall records due to sparsely distributed and/or poorly maintained meteorological stations (Wilby & Yu, 2013). The development of remotely sensed technologies and methodologies to combine satellite estimates with in-situ observation data has facilitated the production of more reliable large-scale climate datasets (Hong *et al.,* 2019). These datasets are often spatially gridded and temporally complete on a regional or global scale. However, these datasets contain large uncertainties and regional bias, thus posing concern and hesitation in utilising them (Nashwan, 2020).

Hydrological models are driven by available rainfall data, and their performance is thus directly linked with the quality of these data (Wagener *et al.*, 2001). Rain gauge networks are the most trusted means for accurate point rainfall measurement. However, sparse rain gauge networks in remote areas and mountainous terrain lead to erroneous rainfall estimates when averaged over a region (Liang *et al*., 2020). Additionally, monsoonal rainfall is specifically challenging to represent as the timing of the monsoon is not consistent year-on-year, and the rainfall tends to be intense for long periods. An expanding selection of large-scale gridded rainfall datasets, both from remote sensing, reanalysis or interpolation of in-situ observations, are becoming available (Le Coz & van de Giesen, 2020). These datasets are proposed to be of value to overcome the absence of in-situ observations and provide an alternative for estimating catchment rainfall.

Southern India experiences a monsoonal rainfall pattern (Sen Roy *et al.*, 2009) with reports of significant weakening of the monsoon in recent years (Joseph & Simon, 2005; Kulkarni, 2012; Dixit *et al.,* 2014; Kumar *et al*., 2020; Swapna *et al.,* 2022). The southwest monsoon generally brings rainfall between June and October, and the northeast monsoon in November and December. In addition to the monsoon strength, timing and duration, topographic factors considerably influence the distribution and concentration of rainfall across the region (Bauer & Morrison, 2008). The estimation of catchment rainfall is complicated by the complex topography of the Western Ghats (Malik *et al.*, 2012), the large spatial and temporal variability of the annual monsoons (Daly, 2006) and the conversion of a sparse rain gauge network and proxy measurements (cloud top temperature, raindrop reflectivity, solar energy, brightness temperature, microwave emission, etc.) into quantitative rainfall estimates (Ghimire *et al*., 2018; Hong *et al.*, 2019). The seasonal nature of rainfall and the resulting streamflow generation within the region has resulted in infrastructural projects being at the forefront of water management planning over the last century (Chowdhury, 2010). The Upper Cauvery Catchment region, located in the Western Ghats, acts as the water tower of the greater catchment.

The Western Ghats act as a barrier to the southwest monsoon clouds and influence the distribution of rainfall in the region. The undulating landscape, slope and aspect of these mountains to the monsoonal winds pose many challenges to the scientific community in understanding the spatial and temporal distribution of rainfall (Venkatesh *et al*., 2021). Along the southwestern and western coasts, the Mean Annual Rainfall (MAR) can be as high as 6000 mm due to the orographic effects of the Western Ghats. In contrast, in the rain shadow on the eastern side of the Western Ghats, the rainfall is markedly reduced to a low of 300 mm (Chidambaram *et al.*, 2018). A delayed or weakened monsoon significantly influences the rainfall in the higher latitudes of the country. Both the steepness and aspect of the mountains in this region directly affect the occurrence and location of rainfall. The steep slopes of the Western Ghats in Maharashtra and Kerala result in a strong orographic effect and drier conditions on the leeward side of the range (Meunier *et al.*, 2015).

The scarce rain gauge data in the Western Ghats region has been a major impediment to scientific studies, limiting the understanding of the regional weather system (Venkatesh *et al*., 2021). The major rivers of southern India originate in this mountain range, and the livelihoods of people in this region depend on the water available (Reddy *et al.,* 2021). Many major dams and water transfers are constructed within this region to provide water for domestic, industrial, and agricultural needs (Rajesh *et al.*, 2016). Any changes in the rainfall pattern result in variations in water availability and directly impacts the livelihoods of the people and economy of the region. Rain gauge data are the primary source of historical rainfall data (Sun *et al*., 2018). Consequently, due to the sparse gauge network over the Western Ghats (and the Indian mainland), the IMD has made a significant effort to convert the available station data to a regular space-time grid (Pai *et al*., 2014). These 0.25-degree daily rainfall grids created by the IMD are the accepted rainfall dataset for India within the scientific community and are considered the rainfall standard across environmental, industrial, and operational companies within India (Singh *et al*., 2021; Buri *et al.*, 2022).

An accurate rainfall representation in India is essential for understanding the hydrological responses during the monsoon rainfall season. Satellite-derived rainfall datasets have succeeded in depicting region-specific rainfall patterns across climatologically different parts of India. Most of the published studies utilising remotely sensed data have taken place across India or in small sub-catchments near the Himalayas. The remotely sensed data are generally compared to the IMD rainfall grids and, in some cases, to the IMD gauge data. These studies have concluded that the remotely sensed data sets struggle to estimate orographic rainfall, particularly in the Western Ghats and the Himalayan foothills (Palazzi *et al*., 2013; Prakash *et al*., 2015; Shah & Mishra, 2016). Therefore, the performance of new remotely sensed datasets which have not been applied in the region needs to be assessed.

In instances where 'off-the-shelf' remotely sensed datasets do not represent the point rainfall nor the simulated catchment streamflow to an acceptable standard, it is common practice to utilise available in-situ rain gauge data to perform a bias-correction (Guo & Liu, 2016). This technique has proven effective globally (Luo *et al.*, 2020); however, it falls short in regions where in-situ rain gauge data are not available or accessible, or there is high uncertainty in the gauged measurements (Kimani *et al.*, 2018). A probable solution is utilising an average ensemble of the selected remotely sensed rainfall datasets in a similar capacity to that which is common practice in the application of global climate model (GCM) data (Noor *et al.*, 2019; Rickards *et al.*, 2020).

This study aims to provide insight into the suitability of selected remotely sensed rainfall datasets and improve the estimation of catchment rainfall by improving the fundamental understanding of rainfall in the Upper Cauvery Catchment.

a) Evaluating remotely sensed rainfall datasets not previously applied at a catchment scale in the Upper Cauvery Catchment and assessing the performance of various 'off-the-shelf' remotely sensed datasets against in-situ rain gauge data.
b) Identifying the best-performing rainfall dataset, including the IMD and remotely sensed datasets.
c) Determine whether the spatial resolution of a rainfall dataset improves the performance in the Upper Cauvery Catchment.

d) Ascertain whether an 'off-the-shelf' remotely sensed rainfall dataset is suitable for hydrological modelling within the Upper Cauvery Catchment without regional bias correction.
e) Determining whether an 'off-the-shelf' remotely sensed dataset could improve the hydrological simulations within a complex topographical region compared to the IMD gridded dataset.
f) Establish whether an ensemble could more accurately represent the catchment rainfall and the simulated streamflow than the IMD gridded rainfall data.

2. Materials and Methods

The performance of the widely used IMD (Pai *et al.,* 2014) gridded rainfall and selected remote sensing (RS) datasets not previously used in the region will be compared to the available in-situ observations. Hydrological simulations will be utilised to determine the effects of various rainfall data on water resource representation.

2.1. Catchment Description

The Cauvery Catchment (81,000 km^2) is situated in southern India (Figure 1). The diverse terrain and strong west-to-east rainfall gradient (6000 mm in the upper reaches to 300 mm on the eastern boundary) result in regionally variable surface and groundwater availability (Meunier *et al.,* 2015) and, depending on local demand patterns, is a critical and widely limiting factor for agriculture (Madhusoodhanan *et al*., 2016), with much of the irrigated agriculture dependent on groundwater abstraction from millions of wells. The catchment is primarily underlain by hard-rock aquifers (Collins *et al*., 2020). Although predominantly rural (Sreelash *et al.,* 2020), parts of the catchment have experienced considerable urban and economic growth over recent years (Gupta & Horan, 2022).

The surface water in the catchment has been affected for centuries by human influences, which have impacted the hydrological functioning of the catchment (Gupta & van der Zaag, 2008). In addition to the significant anthropogenic influence within the catchment, there are ongoing inter-state water-sharing disputes. Water disputes in the Cauvery Catchment differ from other inter-state water disputes, such as in the Krishna, Godavari and Narmada Catchments. These tend to form around the untapped potential of water resources, whereas in the Cauvery Catchment, the disputes surround the reallocation of existing water resources (Janakarajan, 2016) between the federal states of Karnataka and Tamil Nadu (Sharma *et al*., 2020). As the water-sharing agreement in the Cauvery is legally founded, the estimation and distribution of water resources throughout the catchment must be accurately understood.

The Upper Cauvery Catchment drains an area of 10 619 km^2 in the north-western region of the Cauvery Catchment (Figure 1) and constitutes 21% of the total catchment area but generates 82% of the total streamflow (Horan *et al.*, 2021a). The upper reaches of the Cauvery River lie within the Western Ghats (Figure 1: Inset 1). The Upper Cauvery Catchment drains into the Krisharaja Sagar (KRS) dam, where it is stored for domestic and agricultural use. The Western Ghats act as a critical headwater to the larger catchment and a barrier to the southwest monsoon (Chidambaram *et al.*, 2018). In the area of the Western Ghats, the soils tend to be very deep, valley bottoms are covered in dense forests, and mountain slopes are predominately grassland (Pattabaik *et al.*, 2013). As shown in Figure 1, the Upper Cauvery Catchment consists of four gauged sub-catchments (Saklesphur, Thimmanahali, Kudige and KM Vadi). The Upper Cauvery will be modelled at a 0.125-degree resolution for the period 1985-2013 due to data availability and to correspond with the pilot study (Horan *et al.*, 2021a).

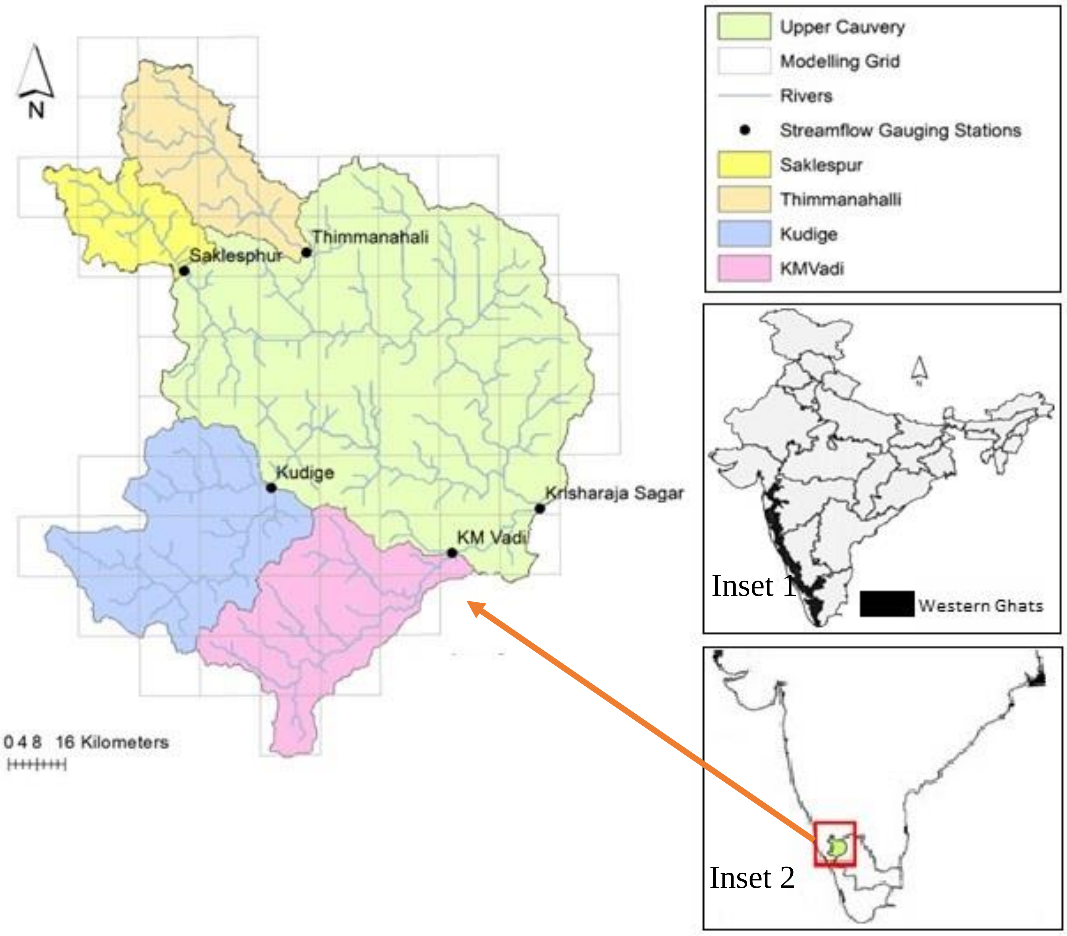


Figure 1 Inset 1: the location of the Western Ghats within India; Inset 2: the location of the Cauvery Catchment within India; Main map: Cauvery Catchment sub-catchment boundaries, modelling grid and the location of streamflow gauges used for hydrological model calibration.

2.2. Rainfall Data

2.2.1. In-situ Rain Gauge Data

The IMD provided daily in-situ rain gauge data for 21 gauges in the Upper Cauvery Catchment (Figure 2; Table 5 in the Appendix). The data records were inconsistent between gauging stations, and thus a period of 1985 to 2013 was selected as the majority of the gauges had data available for this period. There were, however, significant gaps within the remaining data. In this study, no effort was made to infill these gaps as the gauges were not deemed close enough to each other, and due to the complex topography, no meaningful relationships could be drawn. The available data was compared to the gridded datasets using a point-to-pixel analysis.

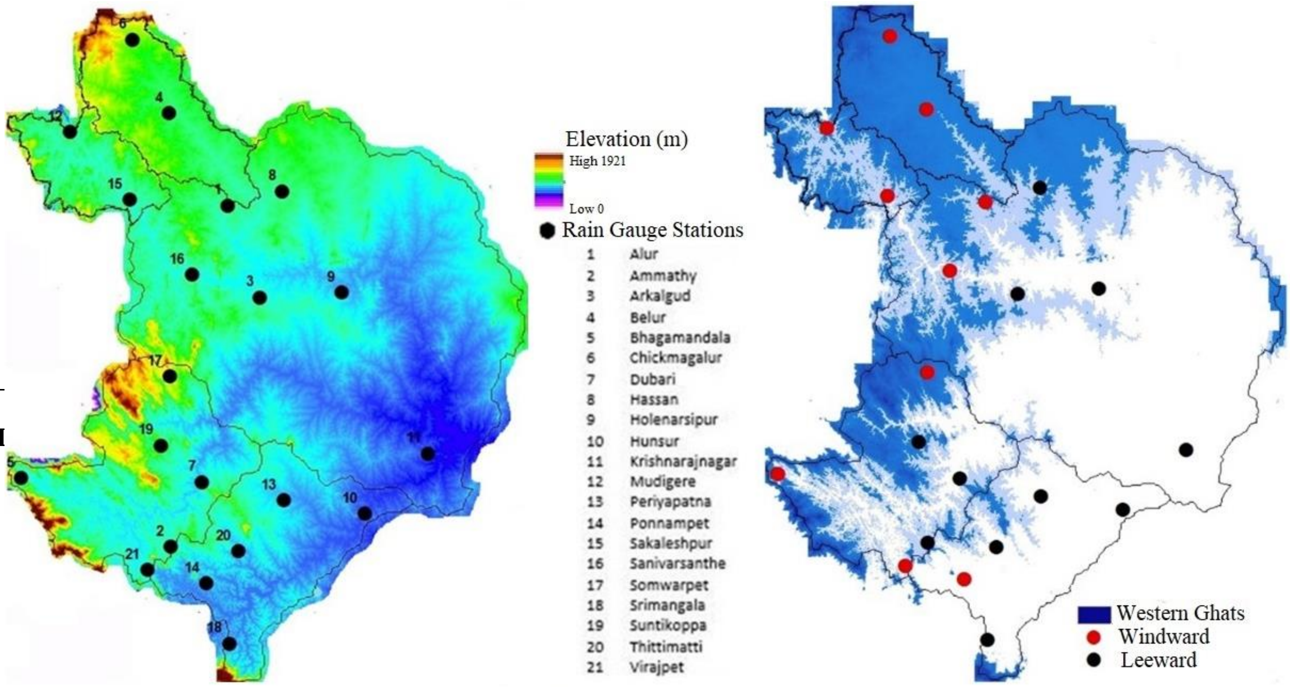


Figure 2 The location of rain gauges and elevation (left) and the demarcation of the Western Ghats within the Upper Cauvery Catchment and windward and leeward positioned gauges (right) within the Upper Cauvery Catchment.

2.2.2 Gridded Rainfall Data

Several remotely sensed rainfall datasets were considered for this study (Table 6 in the Appendix). As summarised in Table 1, only four remotely sensed rainfall datasets met all five of the following criteria, at the time of publication, and thus were selected for this study.

1. Not been explicitly applied within the Upper Cauvery Catchment
2. A spatial resolution of not more than 0.25 degree (IMD grid size)
3. Temporal coverage between 1985 and 2013 (period of available observed rainfall and streamflow data)
4. If a reanalysis dataset, it did not make use of the IMD gridded rainfall data within the compilation
5. The datasets must have undergone a degree of bias correction with gauged rainfall within the dataset production methodology

Table 1 The rainfall datasets utilised in this study, including the methodology, spatial and temporal coverage and resolution, their application in India and reference source.

							Ghats	
Indian Meteorological Department (IMD)	Gauges	India	1901-2014	0.25	Daily	✓	✓	Pai *et al.*, 2014
Climate Hazards Group InfraRed Rainfall with Station data (CHIRPS)	Infrared Gauge	50°N - 50°S	1981- NRT	0.25°	Daily	✓	✓	Funk *et al.*, 2015
CHIRPS v2.0	Infrared Gauge	Global	1981 -NRT	0.05°	Daily	×	×	Funk *et al.*, 2015
Multi-Source Weighted-Ensemble Rainfall (MSWEP) v2.0	Infrared Microwave Gauges	Global	1979- NRT	0.1°	3 hour	✓	✓	Beck *et al.*, 2017
PERSIANN-Climate Data Record (CDR)	Infrared Gauge	60°N - 60°S	1983-2016	0.25°	6 hour	✓	✓	Ashouri *et al.*, 2015

i) IMD

The IMD has developed a daily rainfall dataset at a 0.25-degree grid size over the Indian mainland for the period from 1901- to 2013 based on a network of 6 955 rain gauge stations (Rajeevan & Bhate, 2009; Pai *et al.*, 2014). The IMD gridded rainfall dataset uses these gauges and the simplest form of inverse distance weighted (IDW) interpolation (Shepard, 1968) to estimate a spatial representation of rainfall. Spatial interpolation uses gauging stations with known values to estimate rainfall at points without available data (Li & Heap, 2008). In the IDW method, the rain gauging points are weighted such that the influence of one gauge relative to another declines with distance from the point of unknown rainfall. Weighting is assigned to gauging points using a weighting coefficient that controls the weighting influence—the greater the weighting coefficient, the less effect the gauge will have. The quality of the interpolation can decrease if the distribution of gauging stations is uneven. The maximum and minimum values in the interpolated surface can only occur at sample gauging points. To speed up the computation, only rainfall data from a few of the nearest neighbour stations (minimum of 1 station and a maximum of 4 stations) within a radial distance of 1.5-degrees (166 km^2) around the grid point was used in the IDW interpolation. In the mountainous regions of India, there is a low density of rain gauge stations (approximately 1 station for every 460 km^2) and highly variable rainfall. Thus, the spatial variability of rainfall may not be captured adequately using the IDW methodology. In addition, the maximum rainfall can occur only at gauging points; the rainfall in ungauged areas may be systematically underestimated, especially in the Western Ghats, where the rainfall varies from 600 mm to 5000 mm within 50–100 km.

The IMD gridded daily rainfall data were obtained from the IMD in New Delhi. The data was provided at a 0.25-degree scale in comma-separated values format for the peninsula region of India. Using the coordinates of the in-situ rain gauges, the relevant grids were identified, and the data were extracted using an R statistical software script. The 0.25-degree data were resampled to the 0.125- degree modelling grids, whereby the finer grids retained the value of the 0.25-degree grid cell they overlay.

ii) CHIRPS 0.25- and 0.05-degree

The Climate Hazards Group Infrared Rainfall with Stations (CHIRPS) dataset is available at a spatial resolution of 0.25- and 0.05- degrees across a latitude band of 50ºS–50ºN from 1981 to the present (Funk *et al.*, 2015). CHIRPS utilises high-resolution infrared Cold Cloud Duration (CCD) observations interpolated with a global 0.05° monthly rainfall archive, Climate Hazards Rainfall Climatology (CHPclim), and historical station data from several public data streams and private archives. Monthly rainfall estimates are produced at a 0.25° scale using the CCD observations and TRMM 3B42 rainfall data. These are downscaled to 0.05º. The 0.25° and 0.05° datasets are corrected using the long-term means and CHPClim data. The corrected datasets are blended with available station data to produce the published datasets. Station data is obtained from Global Historical Climatology Network (GHCN), Global Summary of the Day (GSOD), Global Telecommunications System (GTS), Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) and national meteorological agencies in Central and South America and sub-Saharan Africa (Funk *et al.*, 2015).

The 0.05- and 0.025-degree daily rainfall data were downloaded using the in-built chirps R package (https://cran.r-project.org/web/packages/chirps/index.html). The data was provided at a 0.05-and 0.025- degree scale in NetCDF format. Using the coordinates of the in-situ rain gauges, the relevant data were identified and extracted using an R statistical software script. Furthermore, the 0.05- degree data was clipped and aggregated to the 0.125-degree modelling grid using R statistical software. Each 0.125-degree grid was assigned the daily mean of the CHIRPS grids that it fell within. The 0.25-degree data was disaggregated to the 0.125- degree modelling grids, whereby the finer grids retained the value of the 0.25-degree grid cell they overlay. The 0.125-degree datasets were output as a comma-separated values file.

iii) MSWEP

Multi-Source Weighted Ensemble Rainfall (MSWEP) is a global rainfall dataset available from 1979–2015 at a temporal resolution of three hours and a spatial resolution of 0.25°. The dataset is derived from several data sources, including 13 762 rain gauges, satellites, and atmospheric reanalysis models. The long-term mean is derived from the CHPclim dataset, corrected for orographic effects, and then downscaled to a monthly timestep using multiple satellite rainfall datasets. The monthly rainfall is then downscaled to a daily resolution using the CPC Unified rainfall gauged dataset and the area weighting technique. Available three-hourly satellite rainfall estimates are utilised to further downscale the daily resolution rainfall to three-hour MSWEP data. MSWEP undergoes a long-term bias correction using both rainfall (CHPclim and PRISM) and streamflow data (GAGES-II and GRDC) (Beck *et al.*, 2017).

MSWEP daily rainfall data were obtained from the GloH2O (http://www.gloh2o.org/mswep/). The data was provided at a 0.1-degree scale in NetCDF format. Using the coordinates of the in-situ rain gauges, the relevant data were identified and extracted using an R script. Furthermore, using R statistical software, the data was clipped and aggregated to the 0.125-degree modelling grid. Each 0.125-degree grid was assigned the daily mean of the MSWEP grids that it fell within. The 0.125-degree dataset was output as a comma-separated values file.

iv) PERSIANN-CDR

The Rainfall Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR) is available from 1983 to the present at a daily 0.25° resolution. The dataset covers between 60°N and 60°S. PERSIANN-CDR uses a modified PERSIANN algorithm that inputs infrared imagery from GEO satellites into an ANN model and includes gauge measurements from the contiguous United States (CONUS) to estimate global surface rainfall rates from satellite-based infrared measurements (Ashouri *et al.*, 2015). PERSIANN-CDR uses the National Centres for Environmental Prediction (NCEP) Stage IV hourly rainfall to train the ANN model. Bias correction is undertaken on a monthly scale using the Global Rainfall Climatology Project (GPCP) monthly 2.5º rainfall data (Nguyen *et al.*, 2018).

PERSIANN daily rainfall data were obtained from the National Centers for Environmental Information (https://www.ncei.noaa.gov/datasets/climate-data-records/ rainfall-persiann). The data was provided at a 0.25-degree scale in NetCDF format. Using the coordinates of the in-situ rain gauges, the relevant data were identified and extracted using an R script. Furthermore, using R statistical software, the data was clipped, and the 0.25-degree data were resampled to the 0.125- degree modelling grids, whereby the finer grids retained the value of the 0.25-degree grid cell they overlay. The 0.125-degree dataset was output as a comma-separated values file.

v) Ensemble

An ensemble uses the variation of input data, analysis, and methodologies of its component members and tends to be less prone to systematic biases and errors. An ensemble rainfall combines multiple rainfall datasets to create a single dataset. In regions where in-situ rainfall gauge measurements may not be available, an ensemble of selected remotely sensed rainfall datasets may provide a better and more consistent representation of the rainfall than the individual datasets. An ensemble can be applied in rainfall studies to reduce errors with an optimal bias (Baker & Ellison, 2008). Although most published studies utilise an ensemble when applying future GCM predictions, an example of a published study (Cornes *et al*., 2018) has used the concept to improve estimates of historical rainfall. Cornes *et al*. (2018) found that utilising an ensemble of gridded rainfall datasets improved uncertainty estimates compared to individual datasets across Europe.

An average ensemble was determined utilising the 0.125-degree re-gridded CHIRPS 0.05- and 0.25- degree datasets, MSWEP dataset and PERSIANN dataset from 1985-2013. The daily rainfall for each 0.125-degree grid was averaged with equal weighting to produce a single daily time series for each grid (Equation 5.1).

$$\bar{x} = \frac{\sum x}{n} \qquad (5.1)$$

where $\bar{x}$ is the mean, x is the values, and n is the number of x values in the dataset.

A median ensemble was determined utilising the median of the 0.125-degree re-gridded CHIRPS 0.05- and 0.25- degree datasets, MSWEP dataset and PERSIANN dataset from 1985-2013.

Four weighted ensembles were determined using the 0.125-degree resampled CHIRPS 0.25- and 0.05- degree, MSWEP and PERSIANN datasets from 1985-2013.

$$C_{25}\,Weighted\ average = \frac{2x_{C25} + x_{C05} + x_M + x_P}{5} \tag{5.2}$$

$$C_{05}\,Weighted\ average = \frac{x_{C25} + 2x_{C05} + x_M + x_P}{5} \tag{5.3}$$

$$M\ Weighted\ average = \frac{x_{C25} + x_{C05} + 2x_M + x_P}{5} \tag{5.4}$$

$$P\ Weighted\ average = \frac{x_{C25} + x_{C05} + x_M + 2x_P}{5} \tag{5.5}$$

Where *x* is the value in the CHIRPS 0.25-degree ($C25$), CHIRPS 0.05-degree ($C05$), MSWEP (M) and PERSIANN (P) datasets.

For each of the ensembles, using R statistical software and the coordinates of the in-situ rain gauges, the relevant data were identified and extracted for a point-to-pixel evaluation. The average ensemble was consolidated into a comma-separated values file for input to GWAVA.

2.3 Model Selection

Several hydrological modelling studies have been carried out in the headwater sub-catchments of the Cauvery. These include using the auto-regressive moving average time series (ARIMA) model (Maheswaran & Khosa, 2012), an artificial neural network (ANN) model (Maheswaran & Khosa, 2012; Patel & Ramachandran, 2015), a support vector regression (SVR) model (Patel & Ramachandran, 2015), the Water Evaluation And Planning (WEAP) model (Bhave *et al*., 2018), GWAVA (Horan *et al*., 2021a), the Soil and Water Assessment Tool (SWAT) (Kumar & Nandagiri, 2018; Horan *et al*., 2021a; Wable *et al*., 2021) and the Variable Infiltration Capacity (VIC) model (Gowri *et al*., 2021; Horan *et al*., 2021a).

The above-listed model applications within this region have not been highly successful in representing the sub-catchments; however, they provide useful scientific lessons and the identification of various shortfalls. The applications by Maheswaran and Khosa (2012), Patel and Ramachandran (2015), Bhave *et al*. (2018), Horan *et al.* (2021a, b, c), and Gowri *et al.*

(2021) utilised the IMD 0.25-degree daily rainfall grids (Pai *et al.*, 2014) as the source of rainfall data. Bhave *et al*. (2018) and Horan *et al.* (2021a) noted that a limitation of their work was the restricted availability of some specific input data, particularly observed rainfall. Kumar and Nandagiri (2018) and Wable *et al.* (2021) utilised the data from ten and sixteen rain gauges for simulations in the headwater sub-catchments using the SWAT model, respectively produced significantly better results than the studies carried out using the IMD gridded rainfall data. The ability of the SWAT model to simulate daily streamflow was reasonably good, with better low-flow than high-flow simulations. Both Kumar and Nandagiri (2018) and Wable *et al.* (2021) point to rainfall estimation in complex topography as a large source of uncertainty within the modelling exercise.

This study used an improved version of the GWAVA model (Meigh *et al.,* 1999; Horan *et al.*, 2021c). GWAVA is a large-scale gridded water resource model that accounts for natural hydrological processes (soils, land-use, and lakes), using a conceptual rainfall-runoff model and anthropogenic stresses (groundwater abstraction, irrigation, domestic and industrial demands, dam storage, and water transfers) via a demand-driven routine (Meigh *et al*., 1999). The model can be run at a daily or monthly time scale across modelled areas greater than 1000 km^2 and is adaptable to the data availability of the region. GWAVA was developed primarily for use in large, data-scarce regions.

The low-data requirement of the GWAVA model, with published applications in southern Africa (Meigh *et al.*, 1999), West Africa (Meigh & Tate, 2002; Meigh *et al*., 2005; Rameshwaran *et al*., 2017; Rickards *et al*., 2019), South America (Ekstrand *et al*., 2008), Europe (Dumont *et al.*, 2012; Johnson *et al.*, 2015; Williams *et al*., 2015), China (Lui *et al*., 2015) and India (Rickards *et al*., 2020) and a successful pilot study within the Upper Cauvery Catchment (Horan *et al*., 2021a), makes it suitable for application in southern India. The GWAVA model has been updated to better represent small-scale runoff harvesting interventions (Horan *et al*., 2021b), groundwater abstraction, artificial recharge, and regulated dam releases (Horan *et al*., 2021c). These updates are based largely on field data, the principles of the AMBHAS-1D (Tomer *et al*., 2012) groundwater model and the Hanasaki dam routine (Hanasaki *et al*., 2006).

GWAVA simulates the local runoff from each grid cell using a lumped conceptual, Probability Distributed rainfall-runoff Model (PDM) (Moore, 1985). The PDM is used to simulate the spatial variations in soil moisture by means of a probability distribution (Moore, 2007). The PDM utilises a 'bucket' approach, allocating the rainfall amongst various 'buckets' to determine the partitioning of water into the components of the water balance (UKCEH, 2020). Figure 3 illustrates the model configuration.

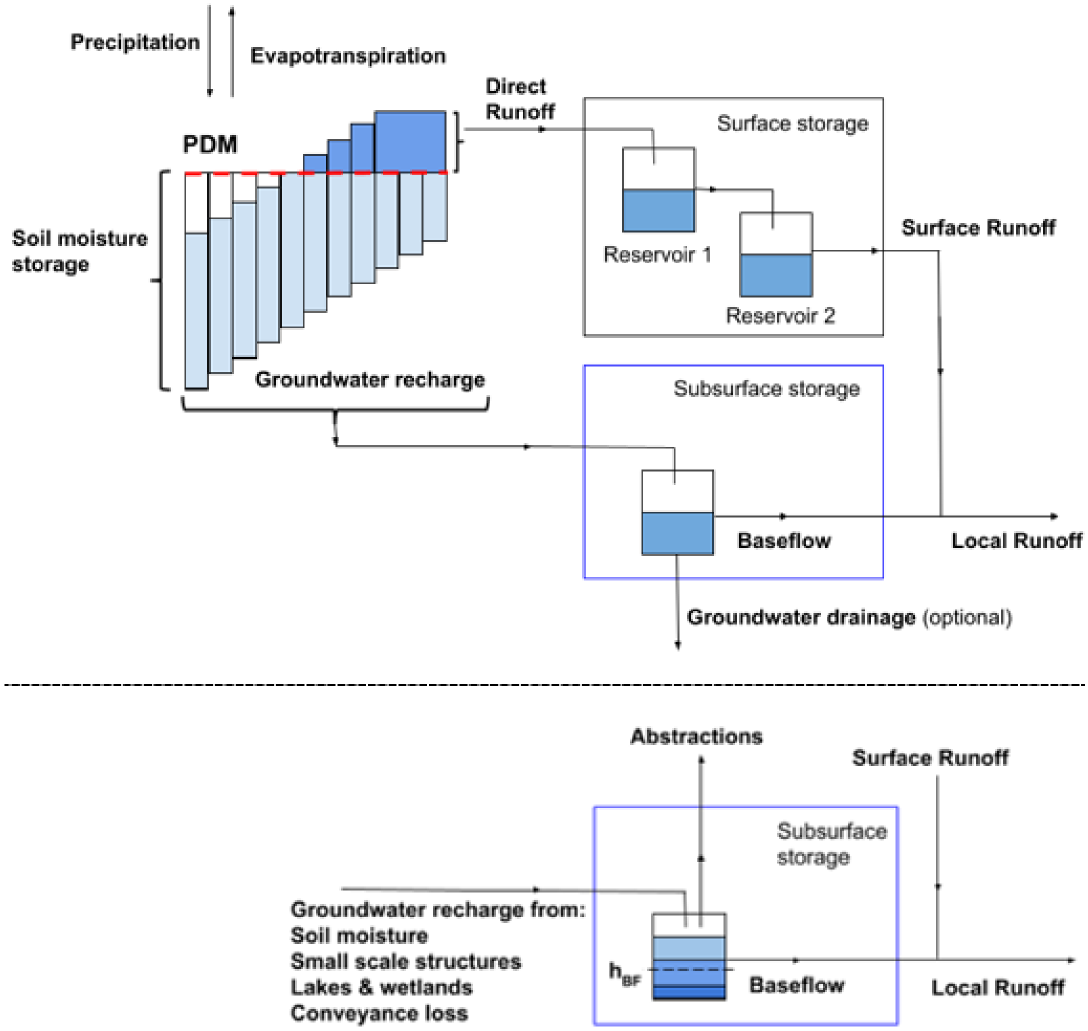


Figure 3 Schematic of the rainfall-runoff model, including the configuration of the probability distributed model (PDM) (UKCEH, 2020).

2.4. Model Application

2.4.1. Input Data

Input data were collected from several sources and extracted from global and regional datasets. The sources and details of the data used in this modelling exercise are summarised in Table 7 in the Appendix.

2.4.2. Model Setup

The Upper Cauvery Catchment was modelled using a gridded configuration with a spatial resolution of 0.125 degrees (Figure 1) using the GWAVA 5.1 model (Horan *et al*., 2021c) forced by various rainfall input datasets:

a) IMD daily rainfall gridded data
b) 0.25- degree CHIRPS daily rainfall data
c) 0.05- degree CHIRPS daily rainfall data
d) 0.1- degree MWESP daily rainfall data
e) 0.25- degree PERSIANN daily rainfall data
f) 0.125-degree ensemble rainfall data

The domestic, irrigation, industrial and livestock demand, large-scale water transfers, hydropower dams, irrigation dams, and agriculture within the irrigation and rural areas were included.

2.4.3. Model Calibration

Five streamflow gauges were used to calibrate the GWAVA model in the Upper Cauvery Catchment using the IMD gridded rainfall dataset (Figure 1). It was then assumed that these calibration parameters would be reasonable for the remotely sensed rainfall datasets. The simulated streamflow was calibrated against the observed streamflow using the SIMPLEX auto-calibration routine. This calibration routine utilises five parameters; (i) a surface routing parameter, (ii) a groundwater routing parameter, (iii) a Probability Distributed Model (PDM) parameter that describes spatial variation in soil moisture capacity, (iv) groundwater initializing depth parameter, and (v) a multiplier to adjust rooting depths. The calibration gauges were selected based on the completeness and temporal coverage of the data and the size of the sub-catchment. The observed streamflow data were deemed sufficient when it had at least five consecutive years of data available from 1981 until 2010.

2.4.4. Evaluation

Due to the high variability of rainfall and streamflow in the Upper Cauvery Catchment, the Kling-Gupta Efficiency (KGE) was used to determine the ability of the rainfall dataset and GWAVA to represent the temporal characteristics of the rainfall and streamflow against the observed data. The Root Mean Squared Error (RMSE) was used to determine the accuracy of the rainfall datasets compared to the observed values. The bias was used to evaluate the ability of the rainfall datasets and GWAVA to estimate the total volume of streamflow across the modelling period.

i) Kling-Gupta Efficiency (KGE)

The KGE (Gupta *et al.*, 2009) is based on correlation, variability bias and mean bias and is calculated (Equation 5.3) as:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_s}{\sigma_o} - 1\right)^2 + \left(\frac{\mu_s}{\mu_o} - 1\right)^2} \qquad (5.3)$$

where r is the correlation coefficient between the monthly simulated and observed data, σ_o is the standard deviation of monthly observation data, σ_s is the standard deviation of the monthly simulated data, μ_o is the mean of monthly observation data, and μ_s is the mean of monthly simulated data.

The KGE indicates the overall performance of the model. The metric allows some perceived shortcomings with the Nash-Sutcliffe Efficiency (NSE) metric to be overcome and has become increasingly popular for evaluating hydrological model skill. A KGE of one indicates perfect agreement between simulations and observations. However, there are many opinions about where the differentiation of ‘good’ and ‘poor’ model performance thresholds lie within the KGE scale. Negative KGE values do not always imply that the model performs worse than the mean flow benchmark. For this study, and to compare model performance, a KGE score of less than 0.2 was deemed poor, between 0.2 and 0.6 as fair and above 0.6 as good.

ii) Root Mean Squared Error

The root-mean-square error (RMSE) is a measure of accuracy and a frequently used measure of the differences between the simulated and observed values (Equation 5.4). The RMSE represents the square root of the second sample moment of the differences between predicted and observed values or the quadratic mean of these differences.

$$RMSE = \sqrt{\frac{\sum(y_s - y_o)^2}{n}} \tag{5.4}$$

where y_o is the monthly observed data value, y_s is the monthly simulated data value, and n is the number of samples.

iii) Bias

The bias is the average tendency of the simulated data to over-or underestimate the observed data (Equation 5.5). The optimal value for the bias is zero. Positive values indicate a model underestimation, and negative values indicate an overestimation. When assessing a model's ability to simulate streamflow, the bias indicates the ability of the model to predict the overall streamflow volume across the modelling period. A bias of between -10 and 10% is considered acceptable.

$$Bias = \frac{\sum y_o - y_s}{\sum y_o} \times 100 \tag{5.5}$$

where y_o is the monthly observation data, and y_s is the monthly simulated data.

The performance of each rainfall dataset and the streamflow generated by each rainfall dataset are ranked from best to worst performing and given a score from one to five. The best performing was assigned a one and the worst a five. The performance was evaluated across the KGE, RMSE and bias statistics within each sub-catchment. Each rainfall dataset was ranked across the individual sub-catchments and the whole Upper Cauvery Catchment to determine the spatial performance across the region and whether a dataset performs better than the IMD grids in the Upper Cauvery Catchment.

3. Results

3.1. Performance of the Rainfall Estimated by the Selected Datasets

Compared to the monthly observed rainfall values in Figure 4, the graphs pertaining to the IMD grids, CHIRPS and MSWEP illustrate a notable scatter above and below the 1:1 line, provide a good fit at lower magnitude events and underestimate at higher magnitude events. PERSIANN overestimates the rainfall depth during lower magnitude events but significantly underestimates the rainfall depth at mid-to-high magnitude rainfall events. The IMD grids present the highest R^2 value of the individual rainfall datasets.

Six ensemble techniques were investigated for use in the Upper Cauvery Catchment. The various methodologies provide similar results regarding the depth of rainfall across events of varying magnitude. As expected, the ensembles produce similar results that fit well to the 1:1 line at lower magnitude events. The clustering around the 1:1 line is more pronounced in the ensembles than in the individual datasets. At high-magnitude events, like individual datasets, the ensembles underestimate the rainfall depth. The degree to which PERSIANN underestimates the high-magnitude events affects the ensembles at these magnitudes. The average ensemble presents a higher R^2 value than the IMD grids.

Using the KGE, RMSE and bias statistics, all the ensembles performed more accurately than the individual rainfall datasets, as shown in Table 2. Although the median, CHIRPS, MSWEP and PERSIANN weighted ensembles produced good KGE scores, the bias was higher than the average ensemble.

As evident from Table 2, the various ensemble methodologies produced the most accurate overall representation (KGE) of the observed rainfall with the lowest margin of error (RMSE), followed by IMD and CHIRPS 0.25-degree, CHIRPS 0.05-degree, PERSIANN and MSWEP. PERSIANN and MSWEP, however, provide the best representation of the overall depth of rainfall across the Upper Cauvery Catchment, followed by the average ensemble, CHIRPS 0.05- degree, CHIRPS 0.25-degree and IMD. The average ensemble provided the best performance of the ensemble methodologies and all the rainfall datasets utilised (Table 2).

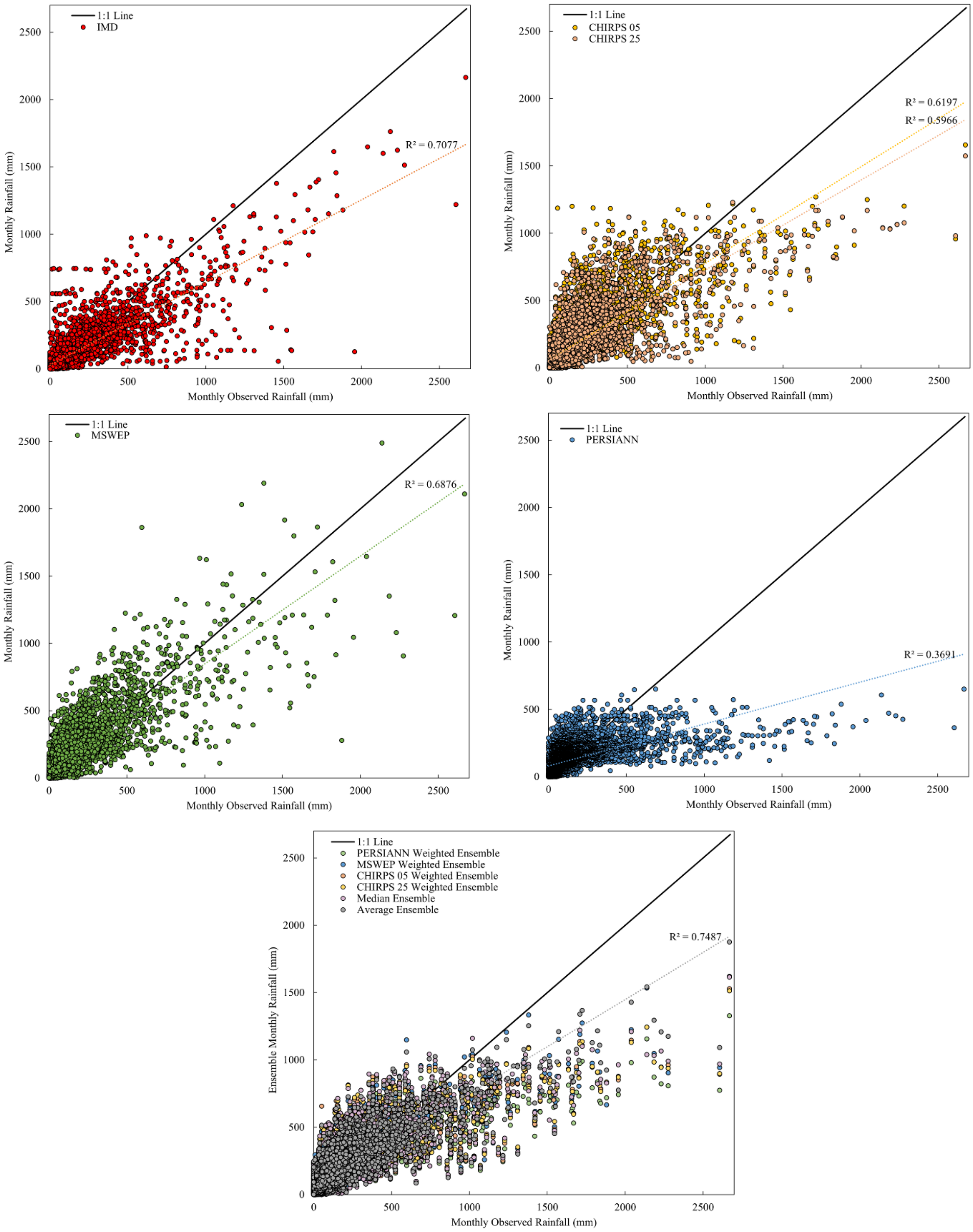


Figure 4 The monthly in-situ observed rainfall against the monthly rainfall from the gridded rainfall datasets (left) and the various ensembles (right)

Table 2 The average KGE, RMSE and bias value (V) when utilising the various rainfall datasets and ensemble techniques across the Upper Cauvery Catchment compared to the monthly observed values. A score (S) is assigned from the best-performing dataset from 1(best) to 11 and these are summed to indicate the overall best-performing dataset.

Metric	**IMD**		**CHIRPS 25**		**CHIRPS 05**		**MSWEP**		**PERSIANN**		**Average ensemble**		**Median ensemble**		**CHIRPS 0.25-degree weighted ensemble**		**CHIRPS 0.05-degree weighted ensemble**		**MSWEP weighted ensemble**		**PERSIANN weighted ensemble**	
	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S
KGE	0.54	7	0.45	8	0.4	9	0.13	10	0.21	11	0.74	1	0.72	2	0.69	5	0.7	4	0.71	3	0.64	6
Bias	-20.4	11	12.5	10	9.6	9	1.9	2	0.4	1	3.5	3	7.6	5	8.5	6	9.3	8	9.1	7	4.9	4
RMSE	129.3	5	148.9	8	152.4	9	161.7	10	204.1	11	120.4	1	127.3	3	129.4	6	129.3	4	124.1	2	134.6	7
Score	**23**		**26**		**27**		**22**		**23**		**5**		**10**		**17**		**16**		**12**		**17**	

As shown in Figure 5, the central tendency of the data from across the year is similar between datasets. The rainfall distribution presents a negative skewness, with the median shifted towards the lower quartile. Considering the nature of rainfall in this region, this is expected as there are a high proportion of days without rainfall. The overall ability of the remotely sensed datasets to represent the distribution of rainfall is fairly accurate when considering the 10^{th} and 90^{th} percentiles, the medians and the interquartile ranges (Table 8 in the Appendix).

During the monsoon (June-September), the data demonstrate a wider variability of data from the median and a relatively large interquartile range (Figure 5; Table 8 in the Appendix) is presumably associated with the variable timing of the onset and the strength of the monsoon. Although the data still demonstrates a positive skewness, it is not as prominent as when considering the rainfall across the year. The 'drizzle day' nature of remotely sensed datasets is evident in the representation of the 10^{th} percentile. 'Drizzle day' nature is caused by the remotely sensed data consisting of spatial means rather than point estimates, which can result in a smaller number of no-rain days when spatial estimates are compared to observed gauge data. The observed and IMD datasets present the 10^{th} percentile of zero, whilst the remotely sensed datasets vary between 10-45mm. The ability of the remotely sensed datasets to represent the distribution of rainfall for the monsoon season is more varied. The median and interquartile range values of the remotely sensed datasets are greater than that of the observed and IMD (Table 8 in the Appendix). The IMD data represents the lower distribution but not, the higher distribution well. PERSIANN presents a small interquartile range suggesting that the rainfall values are clustered around the median and do not represent the high or low quartiles well. The average and median ensembles provide the closest representation of the 10^{th} and 90^{th} percentiles and the interquartile range of observed rainfall, especially in the monsoon season. (Figure 5; Table 8 in the Appendix). Considering the 10^{th} and 90^{th} percentiles, the interquartile range and the R^2 value, the average ensemble was selected as the most accurate and will be used.

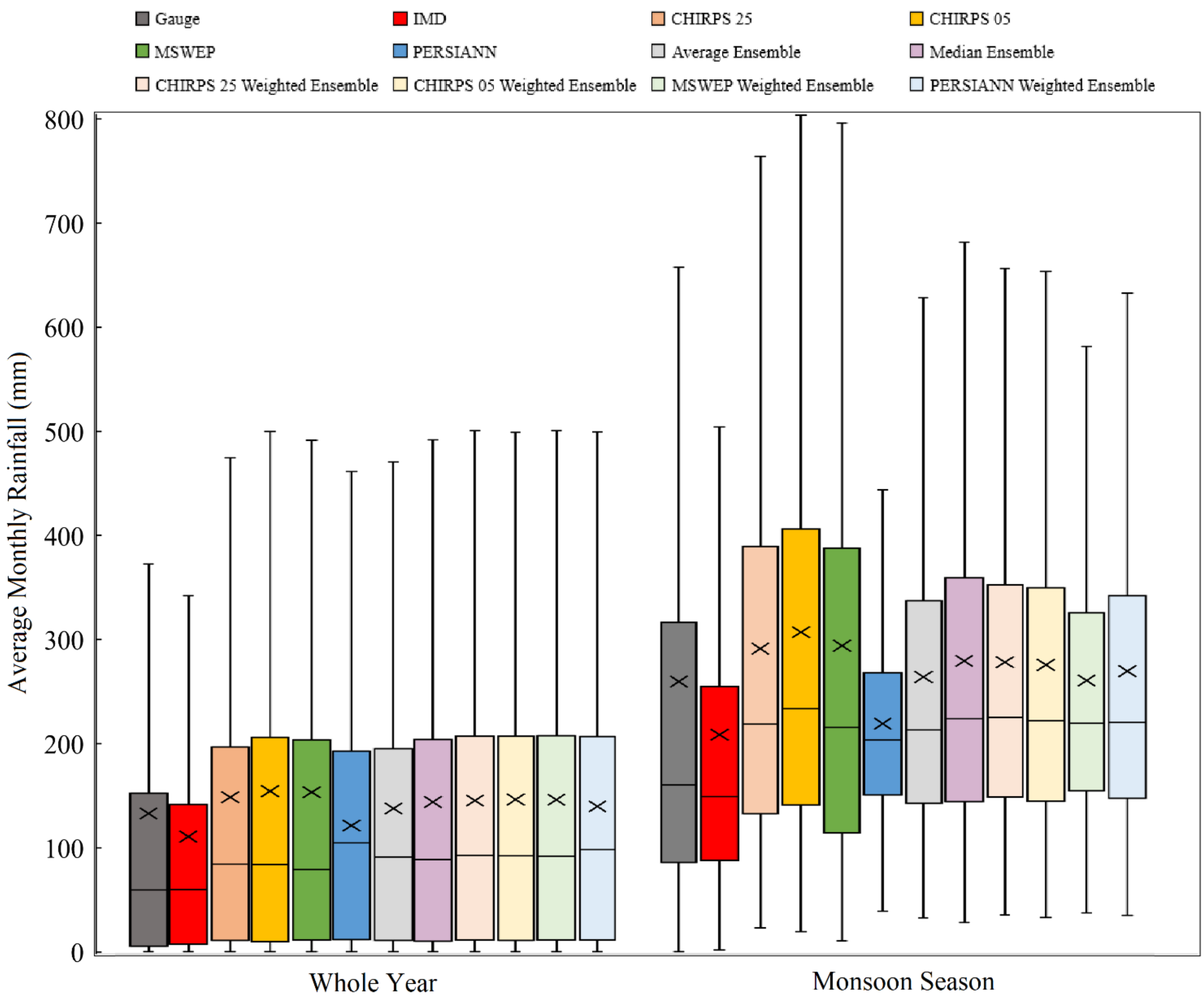


Figure 5 The range of average monthly rainfall produced by each rainfall dataset across the period of 1985 until 2013 and within the monsoon season. The whiskers represent the 10th and 90th percentiles, the line within the box represents the median and the 'X' represents the average.

Figure 6 illustrates that the estimation of rainfall by large-scale remotely sensed datasets within the Upper Cauvery Catchment is variable. The IMD grids underestimate the rainfall systematically across the Upper Cauvery Catchment, and the underestimation is particularly prevalent within the rain shadow.

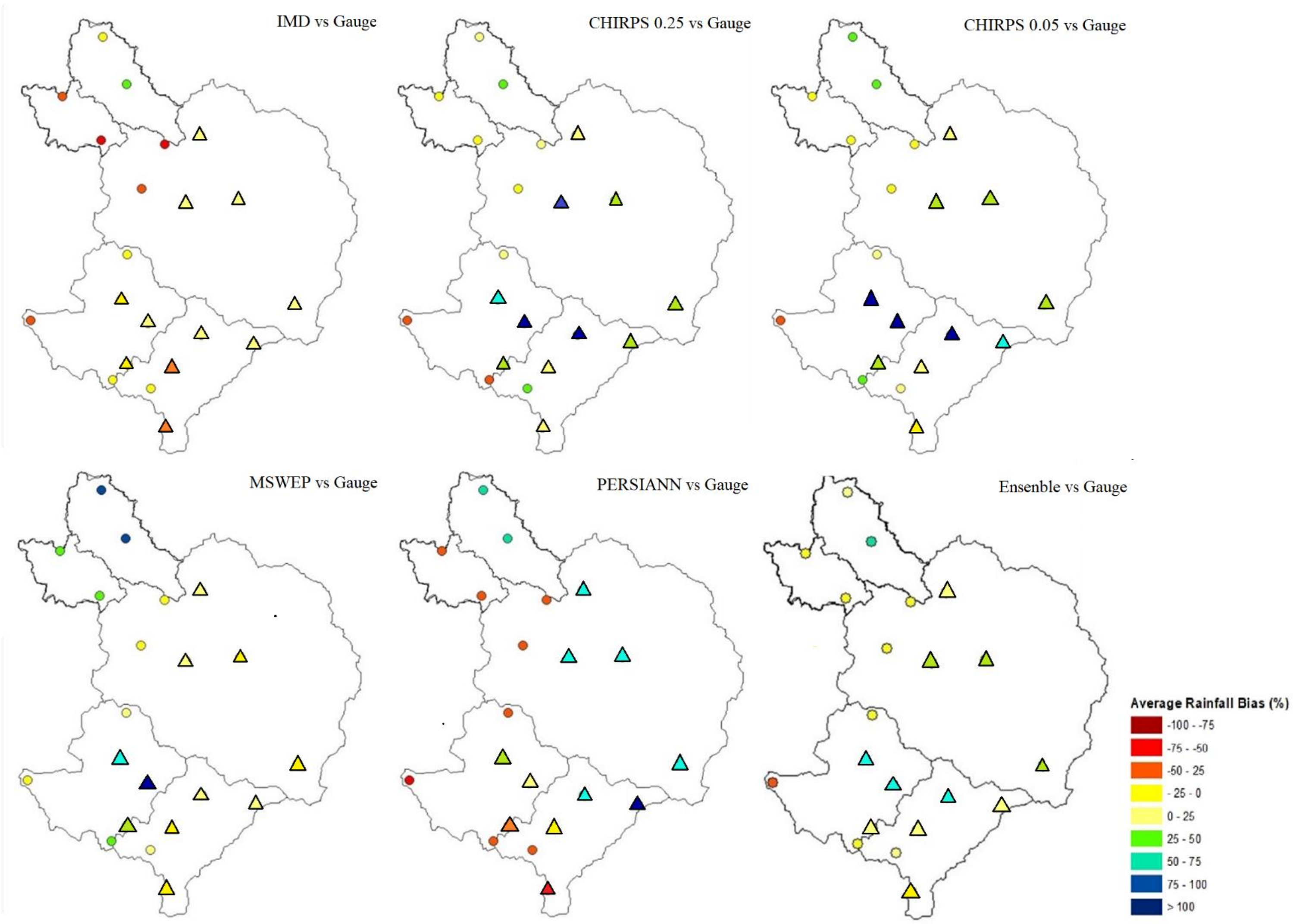


Figure 6 Average monthly rainfall bias (%) from 1985- 2013 between the rainfall datasets (IMD grids, CHIRPS 0.25-and 0.05- degree, MSWEP, PERSIANN and the average ensemble) and the station gauge data. The windward gauges are denoted as a circle and the leeward gauges as a triangle.

At lower altitudes, the CHIRPS datasets overestimate the rainfall but underestimate it at higher altitudes (Figure 6; Figure 7). In the rainshadow, CHIRPS demonstrates a decrease in rainfall with altitude (Figure 7). The performance of the CHIRPS datasets is not dependent on the spatial scale (Figure 2; Figure 5 and Figure 6). The results at both 0.05- and 0.25-degree datasets are similar and, therefore, reflect the methodology rather than the scale at which they are published. Although CHIRPS is published daily, regression slopes and rainfall anomalies are produced at a pentadal (five-year) resolution (Funk *et al.*, 2015). Within the Upper Cauvery Catchment, inter- and intra- annual rainfall and monsoonal conditions vary year on year; therefore, a pentadal methodology is unlikely to fully capture the extreme rainfall. Furthermore, the gauge correction is undertaken at a 1.5-degree scale (Funk *et al*., 2015). Due to the high rainfall variability and topography in this mountainous region and a sparse rain gauge network (Venkatesh *et al.*, 2021), it is probable that although gauge correction has occurred, it is not at a resolution fine enough to be effective.

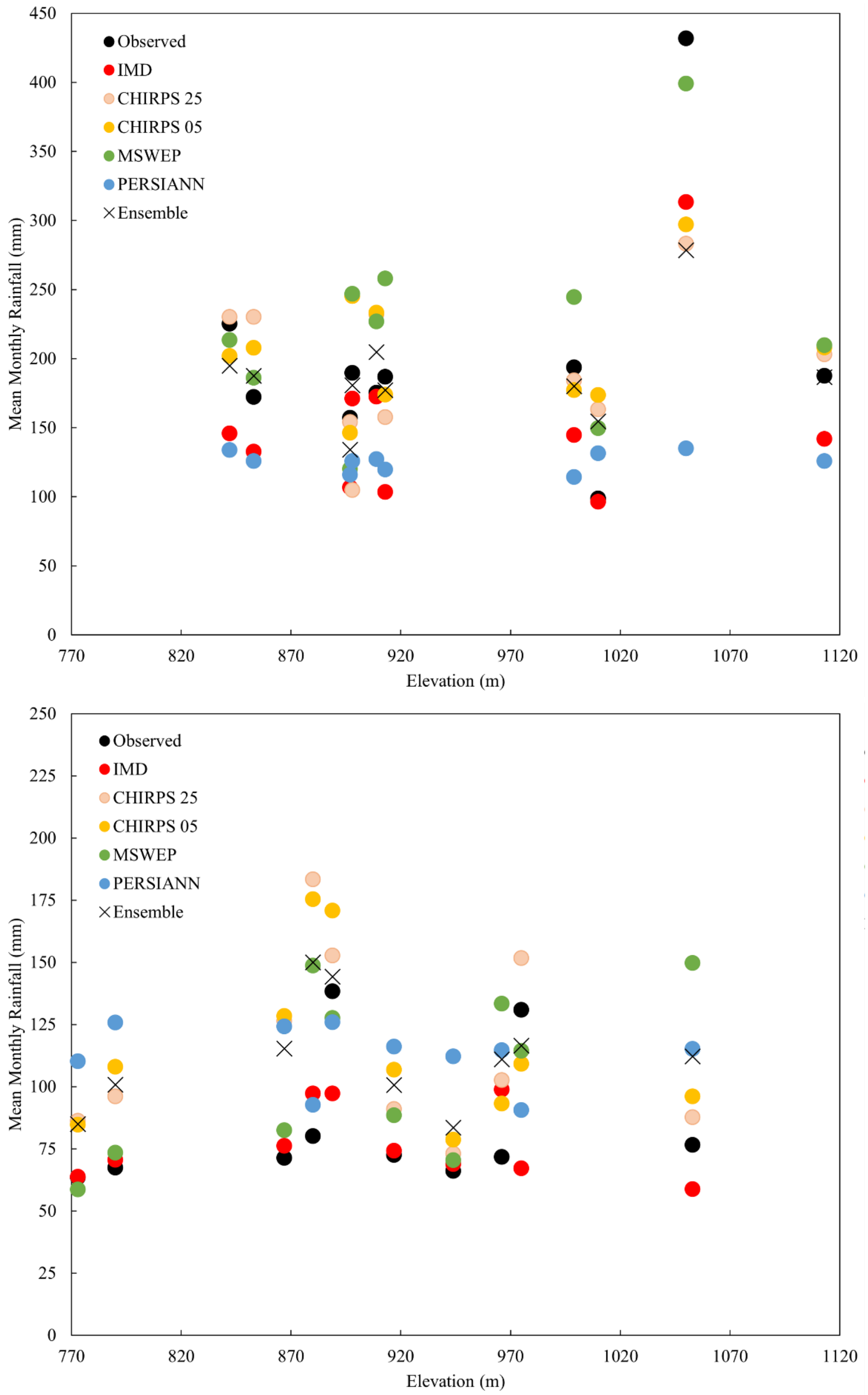


Figure 7 The mean monthly rainfall from 1985 – 2013 provided by each rainfall dataset (IMD grids, CHIRPS 0.25-and 0.05- degree, MSWEP, PERSIANN and the average ensemble) compared with the observed values across the elevation of the windward slope (top) and in the rain shadow (bottom) across the Upper Cauvery Catchment.

MSWEP overestimates the mean rainfall, particularly in the rainshadow (Figure 7). In agreement with the results reported by Prakash *et al*. (2019) and Bhattacharyya *et al*. (2022) across the Western Ghats, in the Upper Cauvery Catchment, MSWEP overestimates the rainfall in the rain shadow and underestimates the rainfall on the windward slopes compared to the in-situ gauge data (Figure 6; Figure 7). Furthermore, similar to Prakash *et al*. (2019) but in contradiction to the results of Liu *et al*. (2019) in Tibet, MSWEP overestimates the rainfall compared with the in-situ gauge data (Figure 6). Considering MSWEP is derived from multiple satellite sources and published at a 0.1-degree resolution, it is surprising that the performance of this dataset was not better in this region. MSWEP is generated through a complex multi-step process, and the long-term mean is corrected for orographic influence but not gauge under-catch. The underestimation of the rainfall on the windward slope could be explained by the lack of consideration for gauge under-catch, specifically in this high altitude and intense rainfall region. However, inverse-distance weighting is utilised via gauges to correct the monthly merged dataset. Inverse-distance weighting is not the most suitable methodology for gauge correction in this region as the gauging network is sparse (Section 2.2.2.).

On the leeward slope, PERSIANN demonstrates a decrease in rainfall with altitude (Figure 7). Similar to the results reported by Prakash *et al.* (2019), the PERSIANN rainfall was underestimated on the windward slopes and overestimated on the leeward slopes compared to the IMD grids (Figure 6; Figure 7). As in Sharannya *et al.* (2020), the rainfall was underestimated in the windward slope compared to the IMD grids. Sharannya *et al.* (2020) estimated a 10% underestimation on the windward slopes throughout the Western Ghats, whereas this study has shown an underestimation of between 25% and 50% compared to the IMD grids. In agreement with the work of Bhardwaj *et al*. (2017) in the Himalayas and Faridzad *et al*. (2018) in the high-elevation regions of the United States of America, PERSIANN consistently underestimated station rainfall depths within the Upper Cauvery Catchment (Figure 6). The coarse-scale gauge correction performed in the generation of this dataset may not capture the complex topography and subsequent variation in rainfall

When applied to the Upper Cauvery Catchment, the average ensemble provides a better point-to-pixel representation of the rainfall in the high-altitude windward regions but not in the rain shadow compared to the IMD grids (Figure 6; Figure 7). The IMD grids would be expected to perform better at the gauging points as they are generated from the IMD in-situ gauged data (Section 2.2.2.). However, in high-altitude areas, the IDW technique is known not to capture the variation in intense rainfall well (Lynch, 2003; Naoumi & Tsanis, 2004; Mair & Fares, 2011; Pingale *et al*.,2014). In the rain shadow, where the rainfall is less intense and variable, the IMD grids represent the rainfall more accurately.

In the Upper Cauvery Catchment, using CHIRPS 0.25- and 0.05- degree, MSWEP and PERSIANN datasets, the average ensemble improved the representation of monthly rainfall (Table 2; Figure 6). The ability of the average ensemble to improve the representation of catchment rainfall and simulated streamflow provides a strong case for this technique, specifically in high-altitude regions with no or low in-situ rainfall availability.

It is evident in Figure 8 that the largest root mean squared error occurs within the monsoon season, June to August, across all the rainfall datasets. PERSIANN has the greatest RMSE, followed by CHIRPS, MSWEP, IMD grids, and the ensemble. The monthly bias of the IMD data is least throughout the year, whereas MSWEP overestimates whilst CHIRPS and PERSIANN underestimates in the dry months of January and February. All the satellite-derived datasets overestimate the rainfall during the pre-monsoon season (April and May). During the monsoon season (June to September), CHIRPS and MSWEP overestimate the rainfall, while IMD and PERSIANN provide a good representation of the volume of rainfall. The ensemble provides the most accurate representation of the rainfall depth across the year (Figure 8). During the dry season, the performance of CHIRPS and MSWEP reduces. IMD has a consistently good KGE score across the year. Despite the good bias of the PERSIANN and the ensemble estimates, the KGE score between December and March is poor (Figure 8)

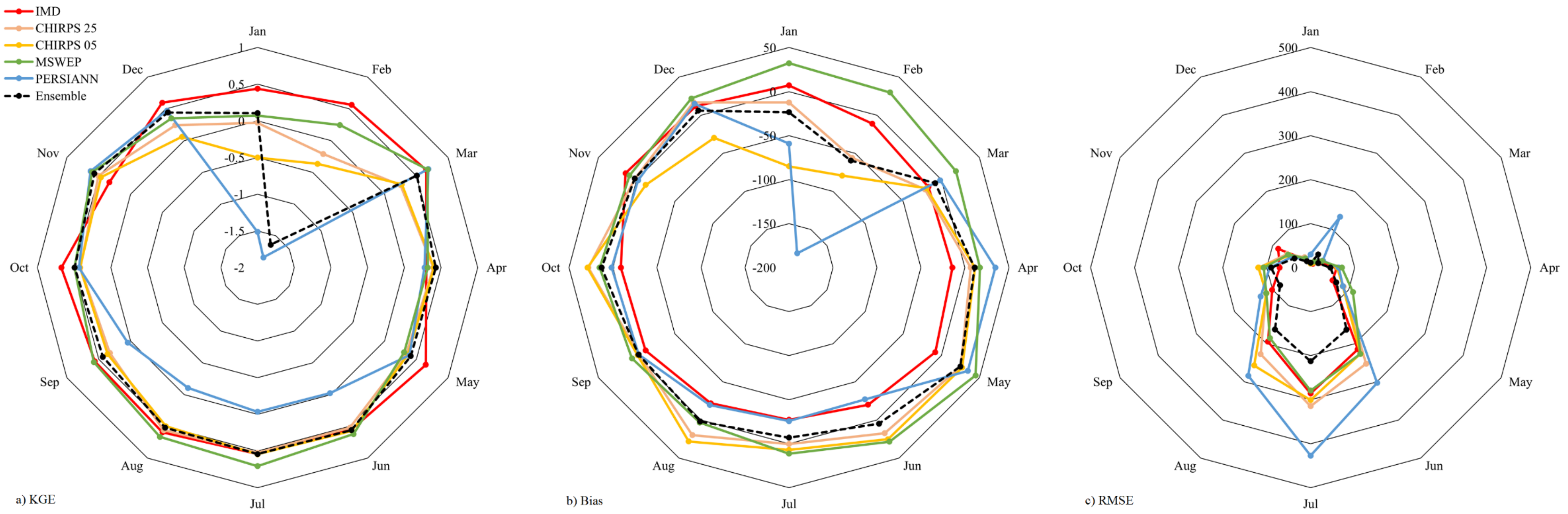


Figure 8 a) Kling-Gupta Efficiency (KGE), b) Bias in percentage and c) Root mean squared error (RMSE) of the rainfall datasets compared with the observed monthly rainfall from 1985 until 2013.

3.2. Performance of Streamflow Simulated Using the Selected Rainfall Datasets

The GWAVA model was calibrated using the observed streamflow at five gauging points using the IMD gridded rainfall. The results of the calibration are provided in Table 3. The results provided compare the GWAVA streamflow simulations using the IMD rainfall grids compared to the observed streamflow.

The monthly streamflow KGE statistics illustrate that the model was calibrated to an acceptable standard (Table 3). However, the streamflow is substantially underestimated at the Saklesphur, KM Vadi, Kudige and KRS Catchments (Figure 9). The sub-catchments with the largest rainfall RMSE produce the highest streamflow RMSE except in the case of Kudige. Thimmanahali Catchment, where the rainfall depth estimation is the most accurate, produces the most accurate simulation of the observed streamflow.

Table 3 The monthly streamflow statistics (KGE, RMSE and bias) of each calibration sub-catchment in the Upper Cauvery Catchment.

Sub- catchment	Monthly KGE	Monthly RMSE	Bias (%)
Saklesphur	0.55	40.7	-46.4
Thimmanahali	0.84	9.2	1.6
KMVadi	0.25	19.5	-33.6
Kudige	0.48	15.8	-32.3
KRS	0.47	25.6	-54.9

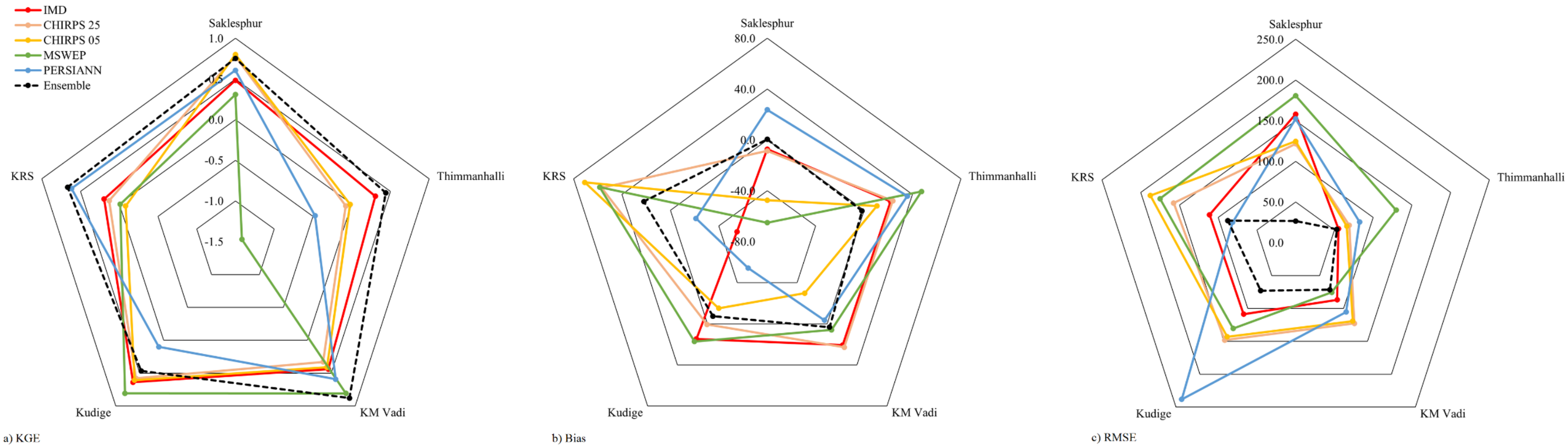
As shown in Table 4, the ensemble produces the most accurate representation of streamflow across the Upper Cauvery Catchment, followed by IMD, PERSIANN, CHIRPS 0.25-degree, MSWEP and then CHIRPS 0.05-degree. At the Saklesphur catchment CHIRPS 0.25-degree provides the most accurate simulation of streamflow, IMD at Thimmanahali and Kudige, MSWEP at KM Vadi and PERSIANN at KRS.

The accuracy of the simulated streamflow using the selected rainfall input is highly variable (Table 4; Figure 9) between the different datasets. As for the rainfall (Table 2), the ensemble provided the best KGE and RMSE scores across the Upper Cauvery Catchment, followed by the IMD grids. Regarding streamflow, PERSIANN outperforms CHIRPS and MSWEP. PERSIANN provides the lowest bias, followed by CHIRPS 0.25-degree, the ensemble, IMD, MSWEP and CHIRPS 0.05-degree (Table 4).

Table 4 Average KGE, RMSE and bias of simulated streamflow across the Upper Cauvery Catchment generated by the selected datasets

Metric	**IMD**	**CHIRPS 25**	**CHIRPS 05**	**MSWEP**	**PERSIANN**	**Ensemble**
KGE	0.46	0.13	-0.37	-0.18	0.23	0.50
Bias	-35.06	26.12	83.21	61.15	5.52	28.21
RMSE	103.98	123.82	128.06	138.30	131.93	62.37

In the monsoon season, the simulated streamflow produced using CHIRPS and MSWEP rainfall inputs was significantly overestimated compared to the observed streamflow, whereas PERSIANN and IMD underestimated the streamflow (Figure 10). The ensemble tends to overestimate the simulated streamflow during the monsoon season but provides a better representation than the individual remotely sensed dataset and the IMD grids. In the dry season, all the datasets tend to produce streamflow that underestimate compared to the observed. Of the remotely sensed datasets, PERSIANN produces simulated streamflow that best represents the observed data at KRS (Figure 9; Figure 10).



641 Figure 9 The monthly a) Kling-Gupta Efficiency (KGE) b) Bias in percentage and c) Root mean squared error (RMSE) of the simulated streamflow produced using the selected rainfall datasets (IMD, CHIRPS 0.25- and 0.05- degree, MSWEP, PERSIANN and the ensemble) compared with the observed streamflow.

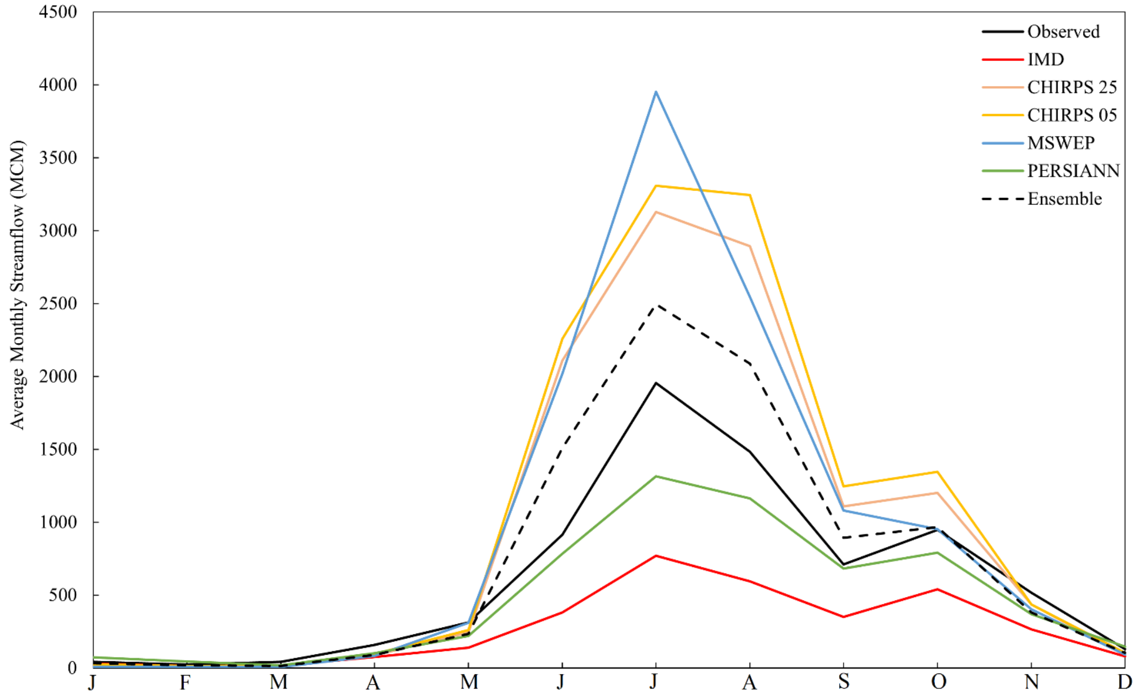


Figure 10 The monthly average streamflow in MCM at KRS simulated using the IMD, CHIRPS 0.25- and 0.05-- degree, MSWEP, PERSIANN and an ensemble rainfall dataset superimposed with the monthly average observed streamflow.

The monthly average streamflow at the entrance to KRS is of significance as approximately 82% of the total catchment streamflow is recorded at this point. Successfully simulating the temporal trend and the volume of streamflow at KRS is critical aspect to understanding and accurately representing the water resources of the greater catchment. The streamflow during the monsoon season reflects the rainfall performance across June to October, with CHIRPS, MSWEP and the ensemble overestimating, and PERSIANN and the IMD grids underestimating the volume of both streamflow and rainfall (Figure 10). However, the bias in streamflow during the monsoon season exceeds the rainfall bias of each rainfall input. The overestimation of rainfall likely causes this during the pre-monsoon period, which overestimates the filling of engineered water storage structures and groundwater stores. This results in an overestimation of the lagged baseflow contribution during the monsoon season, further increasing the over estimation of total streamflow during this period.

During the dry season, the variation in bias and KGE of the rainfall is not reflected in the streamflow (Figure 10). This could be caused by the high number of engineered water storage structures in the catchment and the intensive groundwater pumping that limits baseflow into the main channels that tend to nullify any variation of rainfall bias in the dry months between the rainfall sources. The significant underestimation of rainfall by PERSIANN from

December to March will affect the volume of water for groundwater recharge during this period. This results in an underestimated peak flow during the monsoon season, despite the overestimated rainfall in March to May, as the lagged baseflow component will be significantly underestimated.

4. Discussion

The Western Ghats region northwest of the catchment is a known area of uncertainty for the IMD rainfall data (Pai *et al*., 2014). Each 0.25-degree grid cell contains numerous terrain and gradient increments, and the grid cells span the catchment boundary. This results in an inaccurate representation of the total rainfall and distribution and the distribution of minimum and maximum temperature in this region of the catchment (Yeggina *et al*., 2020). Several studies have reported that conventional spatial interpolation techniques, such as the inverse distance weighting utilised to derive the IMD grid, do not fully account for both climatological and spatial-statistical properties of rainfall fields at high altitudes (Prudhomme & Reed, 1999; Guan *et al*., 2005; Vogel *et al*., 2015). Despite the well-reported underestimation of rainfall in high-altitude regions (Raman *et al*., 2013; Tawde & Singh, 2015; Bharti *et al*., 2016; Dahri *et al*., 2016; Bhardwaj *et al*., 2017; Li *et al*., 2018; Horan *et al.,* 2021a,b,c), the IMD grids have proven to provide one of the most accurate representations of rainfall across the Upper Cauvery Catchment (Table 2). Along with the findings of this study, where the IMD grids outperformed CHIRPS, MSWEP and PERSIANN-CDR, Bhardwaj *et al*. (2017), Yeggina *et al.* (2020) and Reddy *et al*. (2022) found that the IMD grids provided better performance than PERSIANN-CDR, TMPA-3B42 and TRMM 3B43 and MERRA within the Western Ghats.

Rainfall across the study region was found to be highly variable (Figure 7; Table 5 in the Appendix), supporting the findings of Sharannya *et al*. (2018), Wagener *et al*. (2015) and Varikoden *et al*. (2019). Despite all the remotely sensed datasets integrating in-situ gauged data into their methodologies, there were disparities between the rainfall provided by these remotely sensed datasets and the in-situ gauged data provided by the IMD for the Upper Cauvery Catchment. In the Upper Cauvery Catchment, all the datasets tend to underestimate the average rainfall at higher altitudes and overestimate the rainfall in the rain shadow (Figure 7; Figure 8). Previous studies by Prakash *et al.* (2014) and Shah and Mishra (2016) indicated that the CHIRPS datasets underestimate the rainfall on the windward slope compared to the IMD grids. This study found that the CHIRPS datasets tend to underestimate the total volume of rainfall in the high-altitude regions and on the windward slopes, supporting previous studies. Similar results were presented by Saeidizand *et al*. (2018) in Iran and Divya and Shetty (2020) across the Western Ghats. In these studies, and similar to this study, CHIRPS did not accurately represent the rainfall in the high-altitude regions and produced an overestimation of rainfall in the lower-lying regions of the Zagros (Iran) and Western Ghats mountains. Contrary to the conclusions of Huffman *et al*. (2007), Huffman *et al*. (2010), Terzago *et al.* (2018) and Lengfeld *et al*. (2020), the finer scale rainfall datasets, i.e. CHIRPS 0.05-degree and MSWEP did not perform better than the coarser scale datasets in this region of complex topography. This might be because both datasets are produced at a coarser scale, downscaled through various methods, and are gauge-corrected using the same limited number of available rainfall gauges as the coarse-scale datasets.

It was found that the rainfall in the region does not simply increase with altitude as occurs in other mountainous regions of the world (Fowler *et al*., 1988; Al-Ahmadi & Al-Ahmadi, 2013; Morris *et al*., 2016) or decrease in the high altitudes as Singh and Mal (2014) reported in the Himalayas. In the Upper Cauvery Catchment, there does not seem to be a straightforward correlation between altitude and rainfall (Figure 7). The orographic effect on the rainfall was more evident in the Upper Cauvery Catchment (Figure 6; Figure 7), with the Western Ghats forcing the upward movements of moisture-filled air resulting in increased rainfall on the windward slope and less rainfall on the leeward (rains shadow) slope (Arora *et al*., 2006; Chang *et al*., 2014; Morris *et al*., 2016).

Several methodologies of building an ensemble of remotely sensed datasets were tested. All the ensembles tested outperformed the individual rainfall datasets. The ensemble representing the average of the remotely sensed datasets was the best-performing ensemble. Average ensembles can be effectively utilised to reduce uncertainties (Hughes, 2016). Utilising an ensemble allows for the weaknesses in one technique and/or dataset to be shadowed or compensated by the strength of others. The average ensemble accounts for the skill of each technique, maximises the available input data and provides an estimate of the range of possible outcomes. Ensembles can have higher predictive accuracy and successfully represent non-linear interactions. An ensemble can reduce the noise, bias and variance of simulations and potentially create a more in-depth understanding of the data. However, ensemble modelling results can suffer from a lack of interpretability and depend on the ensemble members' prediction accuracy. In areas with perhaps more availability of in-situ rainfall data, more complex techniques such as machine learning (Zhang *et al*., 2021), Google Earth Engine

(Banerjee *et al*., 2020) and big data merging (Hu *et al*., 2019) could be utilised to improve the representation of rainfall. In the case of the Upper Cauvery Catchment, these techniques would not have been feasible, nor would a regional bias correction, due to the sparse and missing in-situ rainfall data. The average ensemble of the chosen datasets provided a more accurate representation of the rainfall than the IMD gridded and the individually remotely sensed datasets. However, it remains critical to ensure that in-situ rainfall gauging networks are maintained and expanded as in-situ data sources of high confidence remain important for the continuous development and ground-truthing of different rainfall datasets.

In agreement with the findings of Sylla *et al.* (2013), Beck *et al.* (2017) and Dembélé *et al.* (2020), it was illustrated that there is no single rainfall dataset which provides the best representation of rainfall and streamflow across the five sub-catchments. Also, the large-scale performance for rainfall datasets is not always valid for sub-catchments in the same catchment. The average ensemble rainfall dataset also provided the most accurate simulation streamflow and, therefore, can be assumed to have accounted for the catchment rainfall most appropriately. A significant challenge in large-scale hydrological modelling is quantifying and managing the uncertainty in climate forcing and evaluation data (e.g. streamflow). Although the model was calibrated to a satisfactory standard using the observed streamflow, at some gauging points in the catchment, there is low confidence in the observed streamflow data (Srinivas & Srinivasan, 2005). Eye-witness accounts and some literature (Srinivasan *et al*., 2015) report the drying out of streams in the Upper Cauvery Catchment in the dry season, which is not reflected in the observed data. Furthermore, the model structure can exaggerate the over-and underestimation of streamflow in both dry and wet periods. The model structure allocates water to the evaporative component first, and thus, the evaporative processes are favoured in times of water stress, and streamflow is favoured in the wet season. This can result in a further underestimation of streamflow when the rainfall is underestimated and an overestimation of streamflow when the rainfall is overestimated.

5. Conclusion

CHIRPS 0.25- and 0.05- degree MWSEP and PERSIANN-CDR rainfall data were applied at a catchment scale in the Upper Cauvery Catchment for the first time alongside the IMD 0.25-degree gridded and an ensemble rainfall. The 'off-the-shelf' remotely sensed rainfall datasets provided a high variation in performance against the in-situ rain gauge data. The IMD grids provided the most accurate representation of rainfall of the individual datasets, despite underestimating the rainfall depths at high altitudes; however, the ensembles, notably the average ensemble, provided the overall best estimates. The following conclusions were drawn from this study:

a) The ensemble rainfall, notably the average ensemble, produced the most accurate representation of the rainfall, followed by IMD, CHIRPS 0.05-and 0.25-degree, MSWEP and then PERSIANN.
b) The spatial scale of the rainfall dataset does not necessarily affect the performance in the high-altitude regions of the Upper Cauvery Catchment.
c) The rainfall in the Upper Cauvery Catchment does not have a distinct correlation to the altitude but correlates strongly to the aspect of the mountains.
d) None of the individual remotely sensed datasets tested could be utilised with confidence in the Upper Cauvery Catchment.
e) The average ensemble and IMD rainfall data produced the most accurate simulation of the observed streamflow across the sub-catchments of the Upper Cauvery, followed by PERSIANN, CHIRPS 0.25-degree, MSWEP and then CHIRPS 0.05-degree.
f) PERSIANN and the average ensemble provided the most accurate simulation of observed streamflow at KRS.

This study evaluated the performance of remotely sensed rainfall datasets not applied in the Upper Cauvery Catchment previously, proposed an ensemble approach to improve rainfall estimations and applied multiple rainfall estimations within the GWAVA water resources model.

Acknowledgements

The authors knowledge funding through both the University of KwaZulu-Natal Doctoral Scholarship and the UPSCAPE (Upscaling local water management interventions to inform larger-scale decision-making in the Cauvery Basin, India) project. The authors would like to extend their gratitude to the Indian Institute of Science- Bangalore, the Ashoka Trust for Research in Ecology and The Environment (ATREE), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT for the guidance and support required to complete this research.

Open Data Statement

All input data used in this study are freely available. Details can be found in Table 6 and Table 7 in the Appendix. For access to the GWAVA code, please contact the GWAVA team at UKCEH. Simulated data are published in the NERC Environmental Information Data Centre under Horan *et al*. (2021).

References

Al-Ahmadi, K. and Al-Ahmadi, S., 2013. Rainfall-altitude relationship in Saudi Arabia. *Advances in Meteorology*, 2013.

Allen, D.M., Cannon, A.J., Toews, M.W. and Scibek, J., 2010. Variability in simulated recharge using different GCMs. *Water Resources Research*, 46(10).

Arora, M., Singh, P., Goel, N.K. and Singh, R.D., 2006. Spatial distribution and seasonal variability of rainfall in a mountainous basin in the Himalayan region. *Water Resources Management*, 20(4), pp.489-508.

Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R. and Prat, O.P., 2015. PERSIANN-CDR: Daily precipitation climate data record from multi-satellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96(1), pp.69-83.

Baker, L. and Ellison, D., 2008. Optimisation of pedotransfer functions using an artificial neural network ensemble method. *Geoderma*, 144(1-2), pp.212-224.

Banerjee, A., Chen, R., E. Meadows, M., Singh, R.B., Mal, S. and Sengupta, D., 2020. An analysis of long-term rainfall trends and variability in the Uttarakhand Himalaya using the google earth engine. *Remote Sensing*, 12(4), p.709.

Bauer, A.M. and Morrison, K.D., 2008. Water management and reservoirs in India and Sri Lanka. *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*, pp.4376-4385.

Beck, H.E., Van Dijk, A.I., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B. and De Roo, A., 2017. MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21(1), pp.589-615.

Bhardwaj, A., Ziegler, A.D., Wasson, R.J. and Chow, W.T., 2017. Accuracy of rainfall estimates at high altitude in the Garhwal Himalaya (India): A comparison of secondary precipitation products and station rainfall measurements. *Atmospheric Research*, 188, pp.30-38.

Bhardwaj, A., Ziegler, A.D., Wasson, R.J. and Chow, W.T., 2017. Accuracy of rainfall estimates at high altitude in the Garhwal Himalaya (India): A comparison of secondary precipitation products and station rainfall measurements. *Atmospheric Research*, 188, pp.30-38.

Bharti, V., Singh, C., Ettema, J. and Turkington, T.A.R., 2016. Spatiotemporal characteristics of extreme rainfall events over the Northwest Himalayas using satellite data. *International Journal of Climatology*, 36(12), pp.3949-3962.

Bhattacharyya, S., Sreekesh, S. and King, A., 2022. Characteristics of extreme rainfall in different gridded datasets over India during 1983–2015. *Atmospheric Research*, 267, p.105930.

Bhave, A.G., Conway, D., Dessai, S. and Stainforth, D.A., 2018. Water resource planning under future climate and socioeconomic uncertainty in the Cauvery River Basin in Karnataka, India. *Water Resources Research*, 54(2), pp.708-728.

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F. and Wada, Y., 2020. Development of the Community Water Model (CWatM v1. 04)–a high-resolution hydrological model for global and regional assessment of integrated water resources management. *Geoscientific Model Development*, 13(7), pp.3267-3298.

Buri, E.S., Keesara, V.R. and Loukika, K.N., 2022. Long-term trend analysis of observed gridded precipitation and temperature data over Munneru River basin, India. *Journal of Earth System Science*, 131(2), pp.1-18.

Chang, F.J., Chiang, Y.M., Tsai, M.J., Shieh, M.C., Hsu, K.L. and Sorooshian, S., 2014. Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information. *Journal of Hydrology*, 508, pp.374-384.

Chidambaram, S., Ramanathan, A.L., Thilagavathi, R. and Ganesh, N., 2018. Cauvery River. In The Indian Rivers (pp. 353-366). Springer, Singapore.

Chidambaram, S., Ramanathan, A.L., Thilagavathi, R. and Ganesh, N., 2018. Cauvery River. In The Indian Rivers (pp. 353-366). Springer, Singapore.

Ciabatta, L., Massari, C., Brocca, L., Gruber, A., Reimer, C., Hahn, S., Paulik, C., Dorigo, W., Kidd, R. and Wagner, W., 2018. SM2RAIN-CCI: A new global long-term rainfall data set derived from ESA CCI soil moisture. *Earth System Science Data*, 10(1), pp.267-280.

Chowdhury, N.T., 2010. Water management in Bangladesh: an analytical review. *Water Policy*, 12(1), pp.32-51.

Collins, S., Loveless, S.E., Muddu, S., Buvaneshwari, S., Palamakumbura, R.N., Krabbendam, M., Lapworth, D.J., Jackson, C.R., Gooddy, D.C., Nara, S.N.V. and Chattopadhyay, S., 2020. Groundwater connectivity of a sheared gneiss aquifer in the Cauvery River basin, India. *Hydrogeology Journal*, 28(4), pp.1371-1388.

Contractor, S., Donat, M.G., Alexander, L.V., Ziese, M., Meyer-Christoffer, A., Schneider, U., Rustemeier, E., Becker, A., Durre, I. and Vose, R.S., 2020. Rainfall Estimates on a Gridded Network (REGEN)–a global land-based gridded dataset of daily rainfall from 1950 to 2016. *Hydrology and Earth System Sciences*, 24(2), pp.919-943.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J. and Jones, P.D., 2018. An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17), pp.9391-9409.

Dahri, Z.H., Ludwig, F., Moors, E., Ahmad, B., Khan, A. and Kabat, P., 2016. An appraisal of precipitation distribution in the high-altitude catchments of the Indus basin. *Science of the Total Environment*, 548, pp.289-306.

Daly, C., 2006. Guidelines for assessing the suitability of spatial climate data sets. International Journal of Climatology: *A Journal of the Royal Meteorological Society*, 26(6), pp.707-721.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, D.P. and Bechtold, P., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), pp.553-597.

Dembélé, M., Hrachowitz, M., Savenije, H.H., Mariéthoz, G. and Schaefli, B., 2020. Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite data sets. *Water Resources Research*, 56(1), p.e2019WR026085.

Demirel, M.C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L. and Stisen, S., 2018. Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrology and Earth System Sciences*, 22(2), pp.1299-1315.

Divya, P. and Shetty, A., 2021. Evaluation of CHIRPS satellite rainfall datasets over Kerala, India. Trends in Civil Engineering and Challenges for Sustainability, pp.655-664.

Dixit, Y., Hodell, D.A. and Petrie, C.A., 2014. Abrupt weakening of the summer monsoon in northwest India~ 4100 yr ago. *Geology*, 42(4), pp.339-342.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H. and Wada, Y., 2016. Modelling freshwater resources at the global scale: challenges and prospects. *Surveys in Geophysics*, 37(2), pp.195-221.

Dumont, E., Williams, R., Keller, V., Voß, A. and Tattari, S., 2012. Modelling indicators of water security, water pollution and aquatic biodiversity in Europe. *Hydrological Sciences Journal*, 57(7), pp.1378-1403

Ekstrand, S., Mancinelli, C., Houghton-Carr, H., Govers, G., Debels, P., Camano, B., Alcoz, S., Filiberto, I., Gámez, S. and Duque, A., 2009. TWINLATIN: Twinning European and Latin American river basins for research enabling sustainable water resources management. Final report.

Faridzad, M., Yang, T., Hsu, K., Sorooshian, S. and Xiao, C., 2018. Rainfall frequency analysis for ungauged regions using remotely sensed precipitation information. *Journal of Hydrology*, 563, pp.123-142.

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

Food and Agriculture Organization of the United Nations, AQUASTAT. Food and Agriculture Organization of the United Nations. Available online: http//www.Fao.Org/Aquastat/Statistics/Query/index.Html?Lang=En (accessed on 19 January 2019).

Fowler, D., Cape, J.N., Leith, I.D., Choularton, T.W., Gay, M.J. and Jones, A., 1988. The influence of altitude on rainfall composition at Great Dun Fell. *Atmospheric Environment* (1967), 22(7), pp.1355-1362.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2(1), pp.1-21.

Ghimire, G.R., Krajewski, W.F. and Mantilla, R., 2018. A power law model for river flow velocity in Iowa basins. *JAWRA Journal of the American Water Resources Association*, 54(5), pp.1055-1067.

Gowri, R., Dey, P. and Mujumdar, P.P., 2021. A hydro-climatological outlook on the long-term availability of water resources in the Cauvery river basin. Water Security, 14, p.100102.

Guan, H., Wilson, J.L. and Makhnin, O., 2005. Geostatistical mapping of mountain precipitation incorporating auto-searched effects of terrain and climatic characteristics. *Journal of Hydrometeorology*, 6(6), pp.1018-1031.

Guo, R. and Liu, Y., 2016. Evaluation of satellite precipitation products with rain gauge data at different scales: Implications for hydrological applications. *Water*, 8(7), p.281.

Gupta, B. and Horan, R., 2022. Hydrological modelling to assess the impacts of socio-economic development and climate change on water resources in Cauvery Basin, India. Preprint: DOI: https://doi.org/10.21203/rs.3.rs-1393124/v1

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2), pp.80-91.

Gupta, J. and van der Zaag, P., 2008. Interbasin water transfers and integrated water resources management: Where engineering, science and politics interlock. *Physics and Chemistry of the Earth*, Parts A/B/C, 33(1-2), pp.28-40.

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, F. 2021. Evaluation of ECMWF forecasts, including the 2021 upgrade. Technical memorandum. ECMWF Technical Memoranda. https://www.ecmwf.int/node/20142

Hanasaki, N., Kanae, S. and Oki, T., 2006. A reservoir operation scheme for global river routing models. *Journal of Hydrology*, 327(1-2), pp.22-41.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. and Tanaka, K., 2008. An integrated model for the assessment of global water resources–Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*, 12(4), pp.1007-1025.

Hanasaki, N., Yoshikawa, S., Pokhrel, Y. and Kanae, S., 2018. A global hydrological simulation to specify the sources of water used by humans. *Hydrology and Earth System Sciences*, 22(1), pp.789-817.

Hong, Y., Gochis, D., Cheng, J.T., Hsu, K.L. and Sorooshian, S., 2007. Evaluation of PERSIANN-CCS rainfall measurement using the NAME event rain gauge network. *Journal of Hydrometeorology*, 8(3), pp.469-482.

Hong, K.Y., Pinheiro, P.O., Minet, L., Hatzopoulou, M. and Weichenthal, S., 2019. Extending the spatial scale of land use regression models for ambient ultrafine particles using satellite images and deep convolutional neural networks. *Environmental Research*, 176, p.108513.

Horan, R., Gowri, R., Wable, P.S., Baron, H., Keller, V.D., Garg, K.K., Mujumdar, P.P., Houghton-Carr, H. and Rees, G., 2021a. A comparative assessment of hydrological models in the Upper Cauvery catchment. *Water*, 13(2), p.151.

Horan, R., Wable, P.S., Srinivasan, V., Baron, H.E., Keller, V.J., Garg, K.K., Rickards, N., Simpson, M., Houghton-Carr, H.A. and Rees, H.G., 2021b. Modelling small-scale storage interventions in semi-arid India at the basin scale. *Sustainability*, 13(11), p.6129.

Horan, R., Rickards, N.J., Kaelin, A., Baron, H.E., Thomas, T., Keller, V.D., Mishra, P.K., Nema, M.K., Muddu, S., Garg, K.K. and Pathak, R., 2021c. Extending a large-scale model to better represent water resources without increasing the model's complexity. *Water*, 13(21), p.3067.

Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y. and Li, L., 2019. Rainfall spatial estimations: A review from spatial interpolation to multi-source data merging. *Water*, 11(3), p.579

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P. and Stocker, E.F., 2007. The TRMM multi-satellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1), pp.38-55.

Huffman, G.J., Adler, R.F., Bolvin, D.T. and Gu, G., 2009. Improving the global rainfall record: GPCP version 2.1. *Geophysical Research Letters*, 36(17).

Huffman, G.J., Adler, R.F., Bolvin, D.T. and Nelkin, E.J., 2010. The TRMM multi-satellite precipitation analysis (TMPA). In Satellite rainfall applications for surface hydrology (pp. 3-22). Springer, Dordrecht.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K.L., Joyce, R.J., Kidd, C., Nelkin, E.J., Sorooshian, S., Stocker, E.F., Tan, J. and Wolff, D.B., 2020. Integrated multi-satellite retrievals for the global rainfall measurement (GPM) mission (IMERG). In Satellite rainfall measurement (pp. 343-353). Springer, Cham.

Hughes, D.A., 2016. Hydrological modelling, process understanding and uncertainty in a southern African context: lessons from the northern hemisphere. *Hydrological Processes*, 30(14), pp.2419-2431.

Janakarajan, S., 2016. The Cauvery Water Dispute: Need for a Rethink. *Economic and Political Weekly*, pp.10-15.

Johnson, A.C., Keller, V., Dumont, E. and Sumpter, J.P., 2015. Assessing the concentrations and risks of toxicity from the antibiotics ciprofloxacin, sulfamethoxazole, trimethoprim and erythromycin in European rivers. *Science of the Total Environment*, 511, pp.747-755.

Joseph, P.V. and Simon, A., 2005. The weakening trend of the southwest monsoon current through peninsular India from 1950 to the present. *Current Science*, pp.687-694.

Joseph, R., Smith, T.M., Sapiano, M.R. and Ferraro, R.R., 2009. A new high-resolution satellite-derived rainfall dataset for climate studies. *Journal of Hydrometeorology*, 10(4), pp.935-952.

Kimani, M.W., Hoedjes, J.C. and Su, Z., 2018. Bayesian bias correction of satellite rainfall estimates for climate studies. *Remote Sensing*, 10(7), p.1074.

Knapp, K.R. and Wilkins, S.L., 2018. Gridded satellite (GridSat) GOES and CONUS data. *Earth System Science Data*, 10(3), pp.1417-1425.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H. and Miyaoka, K., 2015. The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan*. Ser. II, 93(1), pp.5-48.

Kulkarni, A., 2012. Weakening of Indian summer monsoon rainfall in a warming environment. *Theoretical and Applied Climatology*, 109(3), pp.447-459.

Kumar, P.V., Naidu, C.V. and Prasanna, K., 2020. Recent unprecedented weakening of Indian summer monsoon in a warming environment. *Theoretical and Applied Climatology*, 140(1), pp.467-486.

Le Coz, C. and van de Giesen, N., 2020. Comparison of rainfall products over sub-Saharan Africa. *Journal of Hydrometeorology*, 21(4), pp.553-596.

Lengfeld, K., Kirstetter, P.E., Fowler, H.J., Yu, J., Becker, A., Flamig, Z. and Gourley, J., 2020. Use of radar data for characterizing extreme precipitation at fine scales and short durations. *Environmental Research Letters*, 15(8), p.085003.

Li, H., Haugen, J.E. and Xu, C.Y., 2018. Precipitation pattern in the Western Himalayas revealed by four datasets. *Hydrology and Earth System Sciences*, 22(10), pp.5097-5110.

Li, J. and Heap, A.D., 2008. Spatial interpolation methods: a review for environmental scientists. *Geoscience Australia, Record*. Geoscience Australia, Canberra.

Liang, Z., Liu, H., Zhao, Y., Wang, Q., Wu, Z., Deng, L. and Gao, H., 2020. Effects of rainfall intensity, slope angle, and vegetation coverage on the erosion characteristics of Pisha sandstone slopes under simulated rainfall conditions. *Environmental Science and Pollution Research*, 27(15), pp.17458-17467.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. and Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research*, 41(3-4), pp.295-319.

Liu, X., Keller, V., Dumont, E.L., Shi, J. and Johnson, A.C., 2015. The risk of endocrine disruption to fish in the Yellow River catchment in China was assessed using a spatially explicit model. *Environmental Toxicology and Chemistry*, 34(12), pp.2870-2877.

Liu, J., Shangguan, D., Liu, S., Ding, Y., Wang, S. and Wang, X., 2019. Evaluation and comparison of CHIRPS and MSWEP daily-precipitation products in the Qinghai-Tibet Plateau during the period of 1981–2015. *Atmospheric Research*, 230, p.104634.

Luo, X., Fan, X., Ji, X. and Li, Y., 2020. Evaluation of corrected APHRODITE estimates for hydrological simulation in the Yarlung Tsangpo–Brahmaputra River Basin. *International Journal of Climatology*, 40(9), pp.4158-4170.

Lynch, S.D., 2004. Development of a raster database of annual, monthly and daily rainfall for southern Africa: WRC Report No. 1156/1/04.

Madhusoodhanan, C.G., Sreeja, K.G. and Eldho, T.I., 2016. Climate change impact assessments on the water resources of India under extensive human interventions. *Ambio*, 45(6), pp.725-741.

Maheswaran, R. and Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. *Computers & Geosciences*, 46, pp.284-295.

Mair, A. and Fares, A., 2011. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering*, 16(3), pp.371-383.

Malik, N., Bookhagen, B., Marwan, N. and Kurths, J., 2012. Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Climate dynamics*, 39(3), pp.971-987.

Meigh, J.R. and Tate, E.L., 2002. The Gwava Model-Development of A Global-scale Methodology To Assess The Combined Impact of Climate and Land Use Changes. In EGS General Assembly Conference Abstracts (p. 1276).

Meigh, J.R., Folwell, S. and Sullivan, C., 2005. Linking water resources and global change in West Africa: options for assessment. Seventh IAHS scientific assembly, Foz do Iguaçu, Brazil, pp.297-306.

Meigh, J.R., McKenzie, A.A. and Sene, K.J., 1999. A grid-based approach to water scarcity estimates for eastern and southern Africa. *Water Resources Management*, 13(2), pp.85-115.

Meunier, J.D., Riotte, J., Braun, J.J., Sekhar, M., Chalié, F., Barboni, D. and Saccone, L., 2015. Controls of DSi in streams and reservoirs along the Kaveri River, South India. *Science of the Total Environment*, 502, pp.103-113.

Moore, R. J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30(2), pp.263-297.

Moore, R.J., 2007. The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, 11(1), pp.483-499.

Morris, F., Toucher, M.W., Clulow, A., Kusangaya, S., Morris, C. and Bulcock, H., 2016. Improving the understanding of rainfall distribution and characterisation in the Cathedral Peak catchments using a geo-statistical technique. *Water SA*, 42(4), pp.684-693.

Naoum, S. and Tsanis, I.K., 2004. Ranking spatial interpolation techniques using a GIS-based DSS. *Global Nest*, 6(1), pp.1-20.

NASA Jet Propulsion Laboratory (JPL), NASA Shuttle Radar Topography Mission Global 1 Arc Second Number, National Aeronautics and Space Administration, U.S. Government, NASA. Pasadena, CA, USA. 2013. Available online: https://www2.jpl.nasa.gov/srtm/ (accessed on 20 October 2018).

Nashwan, M.S. and Shahid, S., 2020. A novel framework for selecting general circulation models based on the spatial patterns of climate. *International Journal of Climatology*, 40(10), pp.4422-4443.

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D., Ashouri, H. and Thorstensen, A.R., 2018. The PERSIANN family of global satellite precipitation data: A review and evaluation of products. *Hydrology and Earth System Sciences*, 22(11), pp.5801-5816.

Nijssen, B., Lettenmaier, D.P., Liang, X., Wetzel, S.W. and Wood, E.F., 1997. Streamflow simulation for continental-scale river catchments. *Water Resources Research*, 33(4), pp.711-724.

Noor, M., Ismail, T., Shahid, S., Nashwan, M.S. and Ullah, S., 2019. Development of multi-model ensemble for projection of extreme rainfall events in Peninsular Malaysia. *Hydrology Research*, 50(6), pp.1772-1788.

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

Palazzi, E., Von Hardenberg, J. and Provenzale, A., 2013. Precipitation in the Hindu-Kush Karakoram Himalaya: observations and future scenarios. *Journal of Geophysical Research: Atmospheres*, 118(1), pp.85-100.

Patel, S.S. and Ramachandran, P., 2015. A comparison of machine learning techniques for modelling river flow time series: the case of upper Cauvery River basin. *Water Resources Management*, 29(2), pp.589-602.

Pattanaik, J.K., Balakrishnan, S., Bhutani, R. and Singh, P., 2013. Estimation of weathering rates and CO2 drawdown based on solute load: Significance of granulites and gneisses dominated weathering in the Kaveri River basin, Southern India. *Geochimica et Cosmochimica Acta*, 121, pp.611-636.

Pingale, S.M., Khare, D., Jat, M.K. and Adamowski, J., 2014. Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan, India. *Atmospheric Research*, 138, pp.73-90.

Portmann, F.T., Siebert, S. and Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modelling. *Global Biogeochemical Cycles*, 24(1).

Prakash, S., 2019. Performance assessment of CHIRPS, MSWEP, SM2RAIN-CCI, and TMPA precipitation products across India. *Journal of Hydrology*, 571, pp.50-59.

Prudhomme, C. and Reed, D.W., 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in Scotland. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 19(12), pp.1337-1356.

Rajeevan, M. and Bhate, J., 2009. A high-resolution daily gridded rainfall dataset (1971–2005) for mesoscale meteorological studies. *Current Science*, pp.558-562.

Rajesh, S.V.J.S.S., Rao, B.P. and Niranjan, K., 2016. Inter-Basin Water Transfer Impact Assessment on the Environment of Pennar to Cauvery Link Canal. *IRA-International Journal of Technology & Engineering* Vol. 03 no. 03, 3, pp.175-194.

Raju, B.K. and Nandagiri, L., 2018. Assessment of variable source area hydrological models in humid tropical watersheds. *International Journal of River Basin Management*, 16(2), pp.145-156.

Raman, A.T.V.R., Gurugnanam, B. and Arunkumar, M., 2013. One decade hydro-meteorological data assessment through statistics, Dindigul district, Tamil Nadu, South India. *International Journal of Geomatics and Geosciences*, 3(3), pp.659-667.

Rameshwaran, Ponnambalam; Bell, Vicky; Davies, Helen; Houghton-Carr, Helen; Kay, Alison; Miller, James; Rickards, Nathan; Bologo-Traore, Maimouna; Diarra, Abdoulaye; Gnenakantanhan, Coulibaly; Tazen, Fowe; Traore, Karim. 2017 Hydrological research for AMMA-2050. [Poster] In: Future Climate for Africa Mid-Term Conference 2017, Cape Town, South Africa, 4–7 Sept 2017. (Unpublished)

Reddy, B.S.N. and Pramada, S.K., 2022. Suitability of different precipitation data sources for hydrological analysis: a study from Western Ghats, India. *Environmental Monitoring and Assessment*, 194(2), pp.1-20.

Reddy, S., Gupta, H., Reddy, D.V. and Kumar, D., 2021. The suitability of surface waters from small west-flowing rivers for drinking, irrigation, and aquatic life from a global biodiversity hotspot (Western Ghats, India). *Environmental Science and Pollution Research*, 28(29), pp.38613-38628.

Rickards, Nathan; Kaelin, Alexandra; Houghton-Carr, Helen. 2019 Implications of climate change and anthropogenic activity for the water security of West African river basins. [Poster] In African Climate Risks Conference 2019, Addis Ababa, Ethiopia, 7-9 Oct 2019. (Unpublished)

Rickards, N., Thomas, T., Kaelin, A., Houghton-Carr, H., Jain, S.K., Mishra, P.K., Nema, M.K., Dixon, H., Rahman, M.M., Horan, R. and Jenkins, A., 2020. Understanding future water challenges in a highly regulated Indian river basin—modelling the impact of climate change on the hydrology of the Upper Narmada. *Water*, 12(6), p.1762.

Robinson, T.P., Wint, G.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S.I. and Gilbert, M., 2014. Mapping the global distribution of livestock. *PloS one*, 9(5), p.e96084.

Roy, P.S., Meiyappan, P., Joshi, P.K., Kale, M.P., Srivastav, V.K., Srivasatava, S.K., Behera, M.D., Roy, A., Sharma, Y., Ramachandran, R.M. and Bhavani, P., 2016. Decadal land use and land cover classifications across India, 1985, 1995, 2005. ORNL DAAC.

Saeidizand, R., Sabetghadam, S., Tarnavsky, E. and Pierleoni, A., 2018. Evaluation of CHIRPS rainfall estimates over Iran. *Quarterly Journal of the Royal Meteorological Society*, 144, pp.282-291.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.T., Chuang, H.Y., Iredell, M. and Ek, M., 2014. The NCEP climate forecast system version 2. *Journal of Climate*, 27(6), pp.2185-2208.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year high-resolution global dataset of meteorological forcings for land surface modelling, J. *Climate*, 19, 3088–3111, https://doi.org/10.1175/JCLI3790.1, 2006.

Sen Roy, S., 2009. A spatial analysis of extreme hourly precipitation patterns in India. International Journal of Climatology: *A Journal of the Royal Meteorological Society*, 29(3), pp.345-355.

Shah, H.L. and Mishra, V., 2016. Uncertainty and bias in satellite-based precipitation estimates over Indian subcontinental basins: Implications for real-time streamflow simulation and flood prediction. *Journal of Hydrometeorology*, 17(2), pp.615-636.

Sharannya, T.M., Al-Ansari, N., Deb Barma, S. and Mahesha, A., 2020. Evaluation of satellite precipitation products in simulating streamflow in a humid tropical catchment of India using a semi-distributed hydrological model. *Water*, 12(9), p.2400.

Sharannya, T.M., Mudbhatkal, A. and Mahesha, A., 2018. Assessing climate change impacts on river hydrology–A case study in the Western Ghats of India. *Journal of Earth System Science*, 127(6), pp.1-11.

Sharma, A., Schweizer, V. and Hipel, K.W., 2020, October. Analyzing the Cauvery river dispute using a system of systems approach. In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 3969-3975). IEEE.

Shepard, D., 1968, January. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM national conference (pp. 517-524).

Singh, R.B. and Mal, S., 2014. Trends and variability of monsoon and other rainfall seasons in Western Himalaya, India. *Atmospheric Science Letters*, 15(3), pp.218-226.

Singh, S., Mall, R.K. and Singh, N., 2021. Changing Spatio-temporal trends of heat wave and severe heat wave events over India: An emerging health hazard. *International Journal of Climatology*, 41, pp.E1831-E1845.

Sorooshian, S., Hsu, K.L., Gao, X., Gupta, H.V., Imam, B. and Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, 81(9), pp.2035-2046.

Sreelash, K., Mathew, M.M., Nisha, N., Arulbalaji, P., Bindu, A.G. and Sharma, R.K., 2020. Changes in the Hydrological Characteristics of Cauvery River draining the eastern side of southern Western Ghats, India. *International Journal of River Basin Management*, 18(2), pp.153-166.

Srinivas, V.V. and Srinivasan, K., 2005. Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows. *Journal of Hydrology*, 302(1-4), pp.307-330.

Srinivasan, V., Thompson, S., Madhyastha, K., Penny, G., Jeremiah, K. and Lele, S., 2015. Why is the Arkavathy River drying? A multiple-hypothesis approach in a data-scarce region. *Hydrology and Earth System Sciences*, 19(4), pp.1905-1917.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. and Hsu, K.L., 2018. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1), pp.79-107.

Sutanudjaja, E.H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J.H., Drost, N., Van Der Ent, R.J., De Graaf, I.E., Hoch, J.M., De Jong, K. and Karssenberg, D., 2018. PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, 11(6), pp.2429-2453.

Swapna, P., Sreeraj, P., Sandeep, N., Jyoti, J., Krishnan, R., Prajeesh, A.G., Ayantika, D.C. and Manmeet, S., 2022. Increasing frequency of extremely severe cyclonic storms in the north Indian Ocean by anthropogenic warming and southwest monsoon weakening. *Geophysical Research Letters*, 49(3), p.e2021GL094650.

Sylla, M.B., Giorgi, F., Coppola, E. and Mariotti, L., 2013. Uncertainties in daily rainfall over Africa: assessment of gridded observation products and evaluation of a regional climate model simulation. *International Journal of Climatology*, 33(7), pp.1805-1817.

Tawde, S.A. and Singh, C., 2015. Investigation of orographic features influencing spatial distribution of rainfall over the Western Ghats of India using satellite data. *International Journal of Climatology*, 35(9), pp.2280-2293.

Terzago, S., Palazzi, E. and von Hardenberg, J., 2018. Stochastic downscaling of precipitation in complex orography: A simple method to reproduce a realistic fine-scale climatology. *Natural Hazards and Earth System Sciences*, 18(11), pp.2825-2840.

Tomer, S.K., Al Bitar, A., Sekhar, M., Zribi, M., Bandyopadhyay, S. and Kerr, Y., 2016. MAPSM: A Spatio-temporal algorithm for merging soil moisture from active and passive microwave remote sensing. *Remote Sensing*, 8(12), p.990.

UK Centre for Ecology and Hydrology (UKCEH). GWAVA: Global Water Availability Assessment Model Technical Guide and User Manual; Technical Report; UK Centre for Ecology and Hydrology: Wallingford, UK, 2020.

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K.I., Aonashi, K., Inoue, T., Takahashi, N., Iguchi, T., Kachi, M. and Oki, R., 2009. A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *Journal of the Meteorological Society of Japan*. Ser. II, 87, pp.137-151.

Varikoden, H., Revadekar, J.V., Kuttippurath, J. and Babu, C.A., 2019. Contrasting trends in southwest monsoon rainfall over the Western Ghats region of India. *Climate Dynamics*, 52(7), pp.4557-4566.

Venkatesh, B., Nayak, P.C., Thomas, T., Jain, S.K. and Tyagi, J.V., 2021. Spatio-temporal analysis of rainfall pattern in the Western Ghats region of India. *Meteorology and Atmospheric Physics*, 133(4), pp.1089-1109.

Venkatesh, B., Nayak, P.C., Thomas, T., Jain, S.K. and Tyagi, J.V., 2021. Spatio-temporal analysis of rainfall pattern in the Western Ghats region of India. *Meteorology and Atmospheric Physics*, 133(4), pp.1089-1109.

Vogel, B. and Henstra, D., 2015. Studying local climate adaptation: A heuristic research framework for comparative policy analysis. *Global Environmental Change*, 31, pp.110-120.

Wable, P.S., Garg, K.K., Nune, R., Venkataradha, A., Srinivasan, V., Ragab, R., Rowan, J., Keller, V., Majumdar, P., Rees, G. and Singh, R., 2022. Impact of agricultural water management interventions on upstream-downstream trade-offs in the upper Cauvery catchment, southern India: a modelling study. *Irrigation and Drainage*, 71(2), pp.472-494.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V. and Sorooshian, S., 2001. A framework for the development and application of hydrological models. *Hydrology and Earth System Sciences*, 5(1), pp.13-26.

Wagner, P.D., Reichenau, T.G., Kumar, S. and Schneider, K., 2015. Development of a new downscaling method for hydrologic assessment of climate change impacts in data-scarce regions and its application in the Western Ghats, India. *Regional Environmental Change*, 15(3), pp.435-447.

Wambura, F.J., 2020. Potential of rainfall data hybridization in a data-scarce region. *Scientific African*, 8, p.e00449.

Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J. and Viterbo, P., 2014. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 50(9), pp.7505-7514.

Wilby, R.L. and Yu, D., 2013. Rainfall and temperature estimation for a data-sparse region. *Hydrology and Earth System Sciences*, 17(10), pp.3937-3955.

Williams, R., Neal, C., Jarvie, H., Johnson, A., Whitehead, P., Bowes, M., Jenkins, A., 2015. Water Quality, in Progress in Modern Hydrology: Past, Present and Future, First Edition, Rodda, J C, Robinson, M (eds), Wiley and Sons Ltd.

Yeggina, S., Teegavarapu, R.S. and Muddu, S., 2020. Evaluation and bias corrections of gridded precipitation data for hydrologic modelling support in Kabini River basin, India. *Theoretical and Applied Climatology*, 140(3), pp.1495-1513.

Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y. and Ge, Y., 2021. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. *Journal of Hydrology*, 594, p.125969.

Appendix

Table 5 Analysis of the available in-situ rainfall data within the Upper Cauvery

Gauging station	X co-ord	Y co-ord	Start Date	End Date	Missing days	Total Days	Missing %	Mean	Standard Dev	Total
Alur	12.97	75.98	01/1979	12/2013	315	12784	2.00	4.33	12.69	44082.7
Ammathy	12.23	75.85	01/1979	11/2013	615	12753	5.00	5.84	16.46	60267.0
Arkalgud	12.77	76.05	01/1979	12/2013	92	12784	1.00	2.38	7.70	25212.9
Belur	13.17	75.85	01/1979	12/2013	158	12784	1.00	2.36	8.12	24952.7
Bhagamandala	12.38	75.52	01/1979	12/2013	2103	12784	16.00	15.27	33.89	131053
Chickmagalur	13.33	75.77	01/1981	12/2009	1826	10592	17.00	2.45	8.10	18361.4
Dubari	12.37	75.92	01/1979	12/2009	614	11323	5.00	2.73	8.17	24915.0
Hassan	13	76.1	01/1979	12/2013	370	12784	3.00	2.01	7.84	22621.2
Holenarsipur	12.78	76.23	01/1979	12/2013	1166	12784	9.00	2.20	7.55	21042.9
Hunsur	12.3	76.28	/01/1981	12/2013	2203	12022	18.00	2.12	7.90	18547.4
Krishnarajnagar	12.67	76.48	01/1979	12/2013	731	12784	6.00	2.16	7.77	21320.6
Mudigere	13.13	75.63	01/1979	12/2013	909	12784	7.00	6.11	17.02	62511.9
Periyapatna	12.33	76.1	01/1979	12/2013	975	12784	8.00	2.31	7.24	23449.4
Ponnampet	12.15	75.93	01/1979	12/2013	785	12784	6.00	5.52	16.57	57831.4
Sakaleshpur	12.95	75.78	01/1989	12/2013	3537	9131	39.00	5.90	15.21	34359.2
Sanivarsanthe	12.82	75.9	01/1979	12/2013	1281	12753	10.00	4.78	12.90	49779.3
Somwarpet	12.6	75.85	01/1981	12/2013	6493	12053	54.00	5.69	15.55	25319.6
Srimangala	12.02	75.98	01/1979	12/2013	827	12753	6.00	6.96	21.41	74981.2
Suntikoppa	12.45	75.83	01/1979	12/2013	859	12753	7.00	3.96	10.43	41480.8
Thittimatti	12.22	76	01/1979	12/2013	1678	12753	13.00	4.11	12.26	42207.0
Virajpet	12.18	75.8	01/1979	12/2013	402	12784	3.00	6.09	16.29	66006.8

Table 6 Non-exhaustive list of spatial and temporal considerations of available satellite rainfall products

Dataset	**Methodology**	**Spatial coverage**	**Temporal coverage**	**Spatial resolution**	**Temporal resolution**	**Application in India**	**Application in WGs**	**Reference**
Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)	Infrared Gauge	50°N - 50°S	1981- NRT	0.25°	Daily	✓	✓	Funk *et al.*, 2015
CHIRPS v2.0	Infrared Gauge	Global	1981 -NRT	0.05°	Daily	×	×	Funk *et al.*, 2015
CICS High-Resolution Optimally Interpolated Microwave Precipitation from Satellites (CHOMPS)	Microwave	Global	1998-2007	0.25°	Daily	×	×	Joseph *et al*., 2009
CPC MORPHing technique (CMORPH) v1.0	Microwave	60°N - 60°S	1998- NRT	0.25°	3 hour	✓	✓	Joyce *et al.*, 2004
European Re-analysis (ERA)-Interim	Reanalysis	Global	1979- 2017	0.75°	3 hour	✓	✓	Dee *et al.*, 2011
European Re-analysis (ERA) 5	Reanalysis	Global	1979-NRT	0.14°	Hourly	✓	✓	Haiden *et al*., 2021
Global Precipitation Climatology Project (GPCP)-1DD v2.1	Microwave Infrared Gauge	Global	1996-2015	1°	Daily	✓	✓	Huffman *et al*., 2009
Gridded Satellite (GridSat) v1.0	Microwave Infrared	50°N - 50°S	1983-2016	0.01°	3 hour	×	×	Knapp & Wilkins, 2018

Table 6 Cont…

Dataset	**Methodology**	**Spatial coverage**	**Temporal coverage**	**Spatial resolution**	**Temporal resolution**	**Application in India**	**Application in WGs**	**Reference**
Global Satellite Mapping of Precipitation (GSMaP) v6	Microwave Infrared	60°N - 60°S	2000- NRT	0.01°	Hourly	✓	✓	Ushio *et al.*, 2009
Integrated Multi-satellitE Retrievals for GPM (IMERG)	Microwave	60°N - 60°S	2014-NRT	0.1°	½ hour	✓	✓	Huffman *et al.*, 2020
JRA-55	Reanalysis	Global	1959 - NRT	0.56°	3 hour	×	×	Kobayashi *et al.*, 2015
Multi-Source Weighted-Ensemble Precipitation (MSWEP) v2.0	Infrared Microwave Gauges	Global	1979- NRT	0.1°	3 hour	✓	✓	Beck *et al.*, 2017
National Centers for Environmental Prediction- Climate Forecast System Reanalysis (NCEP-CFSR)	Reanalysis	Global	1979-2010	0.31°	Hourly	✓	✓	Saha *et al.*, 2010
Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN)	Infrared	60°N - 60°S	2000-NRT	0.25°	Hourly	✓	✓	Sorooshian *et a*l., 2000
PERSIANN- Cloud Classification System (CCS)	Infrared	60°N - 60°S	2003-NRT	0.04°	Hourly	✓	✓	Hong *et al.*, 2004
PERSIANN- Climate Data Record (CDR)	Infrared Gauge	60°N - 60°S	1983-2016	0.25°	6 hour	✓	✓	Ashouri *et al.*, 2015

Table 6 Cont…

Dataset	**Methodology**	**Spatial coverage**	**Temporal coverage**	**Spatial resolution**	**Temporal resolution**	**Application in India**	**Application in WGs**	**Reference**
Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN)	Infrared	60°N - 60°S	2000-NRT	0.25°	Hourly	✓	✓	Sorooshian *et a*l., 2000
Global Meteorological Forcing Dataset for land surface modelling (PGF)	Gauge, Reanalysis	Global	1948-2012	0.25°	3 hour	✓	✓	Sheffield *et al.*, 2006
Rainfall Estimates on a Gridded Network (REGEN)	Gauge	Global	1950 - 2016	1°	Daily	✓	✓	Contractor *et al.*, 2020
Soil Moisture to Rain -Advanced SCATterometer (SM2RAIN- ASCAT)	Microwave Infrared	Global	2007-2021	0.5°	Daily	✓	✓	Ciabatta *et al.*, 2018
Multi-satellite Precipitation Analysis (TMPA) 3B42RT v7	Microwave	50°N - 50°S	2000-NRT	0.25°	3 hour	✓	✓	Huffman *et al.*, 2007
Tropical Rainfall Measuring Mission (TRMM)-3B42 v7	Microwave Gauge	50°N - 50°S	1997- 2019	0.25°	3 hour	✓	✓	Huffman *et al.*, 2010
WFDEI-CRU	Reanalysis	Global	1979-2015	0.5°	3 hour	✓	✓	Weedon *et al.*, 2014

Table 7 The spatial and temporal resolutions, periods and sources of the input data used in the setup of GWAVA in the Cauvery Catchment

Input Data	**Spatial Resolution**	**Temporal Resolution**	**Time Period**	**Source**
Maximum temperature	0.25 degree	Daily	1951-2016	Indian Meteorological Department (Pai *et al*., 2012)
Minimum Temperature	0.25 degree	Daily	1951-2016	Indian Meteorological Department (Pai *et al*., 2012)
Streamflow gauged data	Catchment	Daily	1971-2014	India-WRIS
Dam Characteristics	Catchment		2018	India-WRIS
Dam inflow and outflow data	Catchment	Monthly	1974-2017	India-WRIS
Dam storage	Catchment	Daily	200-2010	India-WRIS
Water transfers	Catchment	Annual	2008	Ashoka Trust for Research in Ecology and the Environment
Tanks	Catchment		2019	Waterbodies dataset (ATREE)
Check Dams	Karnataka		2006-2012	Structural Investment Report, Watershed Development Department
Farm Bunds	Karnataka		2006-2012	Structural Investment Report, Watershed Development Department
Groundwater levels	District	Monthly	1990-2017	Central Ground Water Board, India
Elevation	0.003 degree		2000	NASA Shuttle Radar Mission Global 1 arc second V003 (NASA Jet Propulsion Laboratory, 2013)
Geology	Asia			United States Geological Survey
Specific yield	India			Central Ground Water Board, India
Soil type	0.008 degree		1971-1981	Harmonized World Soil Database v1.2 (Fischer *et al*., 2008)
Soil properties	Global		2010	Table 2- Allen *et al*. (2010)
Input Data	**Spatial**	**Temporal**	**Time**	**Source**

	Resolution	**Resolution**	**Period**	
Land Cover Land Use	0.001 degree		2005	Decadal land use and land cover across India 2005 (Roy *et al*., 2016)
Crops	Taluk*		2000	National Remote Sensing Centre (NRSC)
Total and Rural Population	Village		2001	Census of India 2001 (http://sedac.ciesin.columbia.edu/data/set/india-india-village-level-geospatial-socio-econ-1991-2001)
Livestock	0.05 degree		2005	CGIR Livestock of the World v2 (Robinson *et al.*, 2014)
Conveyance losses	Village		2011	Household & Irrigation Census 2011- Town and Village directory (https://censusindia.gov.in/DigitalLibrary/TablesSeries2001.aspx
Return flow	Village		2011	Household & Irrigation Census 2011- Town and Village directory (https://censusindia.gov.in/DigitalLibrary/TablesSeries2001.aspx)
Irrigation efficiency	Continental		1986	Irrigation and Drainage Paper (FAO) No 1
Surface-water fraction	Village		2011	Household & Irrigation Census 2011- Town and Village directory (https://censusindia.gov.in/DigitalLibrary/TablesSeries2001.aspx)
Industrial demand	Karnataka		Currently unknown	Industrial Plot Information System- Karnataka Industrial Area Development Board (https://http://164.100.133.168/kiadbgisportal/)
Livestock demand	India		2006	CGIR Livestock of the World v2 (Robinson *et al.*, 2014)
Domestic demand	Village		2001	Household & Irrigation Census 2011- Town and Village directory (https://censusindia.gov.in/DigitalLibrary/TablesSeries2001.aspx)

Table 8 Statistical analysis of the distribution of rainfall values produced by each rainfall dataset during the whole year as well as the monsoon season.

	Whole year			**Monsoon Season**		
Dataset	**10th Percentile**	**90th Percentile**	**Interquartile Range**	**10th Percentile**	**90th Percentile**	**Interquartile Range**
Gauge	**0.0**	**346.1**	**146.8**	**40.0**	**589.7**	**230.5**
IMD	**0.0**	272.1	**134**	**45.5**	436.2	167.1
CHIRPS 25	1.1	**404.0**	185.9	86.8	624.0	**256.4**
CHIRPS 05	0.0	420.8	196.2	88.1	652.1	265.4
MSWEP	0.9	409.9	192.2	65.6	627.8	273.5
PERSIANN	0.5	275.4	180.7	108.6	347.3	117.4
Average Ensemble	**1.0**	**349.6**	**184.2**	**92.3**	**544.2**	**215.1**
Median Ensemble	1.6	370.8	193.8	94.8	552.7	194.6
CHIRPS 25 Weighted Ensemble	1.9	370.0	195.9	98.5	534.2	203.5
CHIRPS 05 Weighted Ensemble	1.5	377.9	196.2	94.4	528.8	205.3
MSWEP Weighted Ensemble	1.8	373.8	196.1	104.5	468.5	170.9