Skill assessment of GloFAS-ERA5 operational river discharge for the major Indian River Catchments

Mohd. Imaran[1], Arun Chakraborty[1], Subhasish Tripathy[2]

[1] Centre for Oceans, Rivers, Atmosphere and Land Sciences, Indian Institute of Technology, Kharagpur, 721302, India
[2] Department of Geology and Geophysics, Indian Institute of Technology, Kharagpur, 721302, India
Corresponding author: Arun Chakraborty (arunc@coral.iitkgp.ac.in)
† Subhasish Tripathy (stripathy@iitkgp.ac.in)

† Mohd. Imaran (imranrock672@gmail.com)

Key Points:

- About 60 percent of the hydrographic stations show negative bias for collected river discharge.
- Nearly 80 percent of the hydrographic stations show good skill with significant mean absolute error at few stations.
- The assessment shows a good skill for Ganges-Brahmaputra and the lowest for Pennar and Cauvery river basins.

Abstract

A significant task in river hydrology is to envisage the river's present, past, and future environments. India has some of the world's major river basins, including Ganges-Brahmaputra, Mahanadi, Krishna, and the Godavari produce an enormous amount of water as river discharge alongside turbidity into the Bay of Bengal. The revised Kling-Gupta efficiency skill score (KGESS) has been used to determine the performance of reanalysis river discharge. The skill of reanalysis discharge was found admirable for the Ganges-Brahmaputra river basin (KGESS = 0.86 > 0), with notable mean absolute error and high correlation coefficient (0.94). Furthermore, Subarnarekha, Brahmani-Baitarani, Mahanadi, Godavari, Krishna river basins, and the rivers flowing between Mahanadi and Pennar Rivers exhibit moderate to good skill. However, Pennar, Cauvery, and the rivers flowing between Pennar and Kanyakumari show the lowest skill. Approximately 60% hydrographic stations of river catchments demonstrate that reanalysis discharge is negatively biased (i.e., bias < 1). Nearly 58% hydrographic stations show lower variability (i.e., variability ratio < 1) with the median value of 0.91 and the interquartile range (0.82, 1.13). Moreover, the overall median of the Pearson correlation coefficient was 0.73 with interquartile ranges between 0.51- 0.83. The reanalysis and observed datasets show a significant change in river discharge throughout the southwest monsoon and less in the post-monsoon period. Concurrently, some hydrographic stations show a significant increase in river discharge during post-monsoon in Pennar and Cauvery River basins.

Keywords: Historical River discharge, GloFAS-ERA5 reanalysis, KGESS, Indian sub-continental basins.

1 Introduction

A significant task in river hydrology is to assess the river's historical, present, and forthcoming hydrological environments. It is usually because of sequential and spatial gaps in the overall river flow spotting network (Harrigan et al., 2020). Assessment of river discharge is significant to understand the hydrological cycle globally, which is relevant to accessing water resources. Nonetheless, skill is challenging in a zone where ground interpretations are inadequate (Seo and Lee, 2017). India is the second populous country globally, with massive freshwater supplies in cultivation and domiciliary segments (Jain et al., 2004). Quick discharge of water from artificial dams has been answerable for some major tragedies in hilly areas of the world. In the present scenario, regulatory events and failures of artificial dams have shown the need to examine the flood's anatomy and the behavior of debris dams (Stephen G. Evans., 1986). In modern times, the air temperature has unfavorably amplified across the Indian provinces (Krishna Kumar et al., 2011), which is prominent to further synchronized dry and hot extremes (Mishra et al., 2020). The fluctuations in the air temperature over the surface and escalation in precipitation inconsistency led to the weakening of dry and hot cyclical rainy season immoderations over India (Mishra et al., 2020). Change in climate consequences due to uneven rainfall patterns and runoff affect water obtainability and quality (Das et al., 2018). Most parts of the world are mainly inadequate in long-term river discharge observations. Besides, a large portion of the country's hydrometric information is not accessible continuously (Lewis et al., 2019).

Consequently, lack of interpretations is a fundamental problem in our skill to timely observe the caution of extreme events in hydrology such as droughts and floods, subsequently having implications for the systematic decline of global disaster risk (UNDRR, 2015). Concern about water accessibility in India's present and future climate scenario is dynamic for nutrition and water availability (Bharat & Mishra, 2020). In the past few decades, climate change was a reflective challenge for water accessibility and extensively affected the global community (Bharat & Mishra, 2020). A substantial increase in the mean temperature globally alongside the fluctuations in rainfall and atmospheric water stresses rigorously disturb the hydrological system altogether, ecology, changes in sea level, and yield (Arnell, 1999). Numerous hydrological products worldwide deliver the assessment of streamflow with an extensive series of compelling and operational strategies (Beck et al., 2017). The world's largest river catchments, i.e., the pooled Ganges-Brahmaputra-Meghna (GBM) delta and the Mekong River delta system, the countries located at the downstream side of these basins are predominantly vulnerable to water associated risks in the absence of upstream hydro-meteorological conditions (Sikder et al., 2019). Shallow water from these waterways gives incredible advantages. They support primary cultivation and energy creation needs for more than 690 million people (FAO, 2016), around 10^{th} of the world's human population.

Conversely, surface water's drawbacks similarly generate problems, mostly striking at the downstream portions of these catchments, which were at risk to have the world's most extraordinary floods, yet in addition to dry spells (UNEP, 2016). For instance, Bangladesh, located at the downstream segment of the GBM catchment, suffers from floods and seriously hampering its financial development. The country consists of about 80% of floodplains in the ordinary year, nearly 33% land area of Bangladesh experienced flood during the rainstorm (Brouwer et al., 2007). The significance of knowing the water cycle and evaluating its different motions by utilizing the Land Surface Model (LSM) is more intense in an ungauged and trans-border area. This information can anticipate enormous ranged catastrophes (Siddique-E-Akbor et al., 2014). The LSM outputs are useful in hydrological studies; several research studies have used these easily accessible LSMs outputs to evaluate various water cycle components. Model simulations of the Global Land Data Assimilation System (GLDAS) distinguished various water resources in the world's main river catchments (Lakshmi et al., 2018; Rodell et al., 2004). Possibly the furthermost extensive usage of GLDAS or additional worldwide available products along with Gravity, LSMs in the Southern and Southeast Asia are from the Gravity Recovery and Climate Experiment (GRACE) to classify the fluctuations in groundwater and deviations in the storage of water. Rodell et al. (2009) and Chinnasamy et al. (2015) applied soil moisture derived from GLDAS alongside GRACE to measure depletion in groundwater level over the Northern part of India. Although natural and anthropogenic activities can trigger water, the sensitivity of runoff in the sub-basin and basin scales is essential for preparing, planning, worsening the environment, and sustainable Groundwater regulation. Three methods are commonly operational, i.e., hydrological modeling, statistical methods, and climate flexibility, to predict the runoff or discharge sensitivity. Wang and Tang (2014) analyzed a physical model to evaluate hydrological understandings at the basin scale. Based on a hydrological model (Mishra and Lilhare, 2016; Vano et al., 2012), the water balance replicated by the model is applied to evaluate the identifications of runoff for diverse climatic projections.

2 Data

It is remained possible to attain useful streamflow forecasts by assembling a river directing scheme and the terrestrial surface model European Centre for Medium-Range Weather Forecasts (ECMWF) the global weather forecast system (McMillan et al., 2010). The terrestrial surface model of the ECMWF-ERA5 (Hersbach et al., 2020) coupled with the Distributed Water Balance and Flood Simulation model (LISFLOOD) (van der Knijff et al., 2010) to obtain the Global Flood Awareness System (GloFAS-ERA5) derived reanalysis of river discharge. Actual runoff (m d^{-1}) from a single cell is not associated with adjacent cells in ERA5; therefore, it is impossible to evaluate river flow ($m^3\ s^{-1}$) at the basin scale. Joining ERA5 runoff and LISFLOOD permits the connection of grid cells horizontally with runoff channels over the river network to provide river discharge.

2.1 ERA5 runoff

The scheme used in the Integrated Forecasting System (IFS) of ECMWF; Likewise, ERA5 river runoff developed using Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) terrestrial superficial model (Balsamo et al., 2009). HTESSEL records the energy and surface water fluctuations and the global advancement of soil moisture, snowpack, and soil temperature. An abundance of snowmelt and rainfall are apportioned as runoff from the surface or penetrated a four-layered soil segment (7 cm profundity for upper layer and afterward 21, 72, and 189 cm) at individually ERA5 lattice cell before depleting from the lower portion of the soil segment as subsurface flow (Balsamo et al., 2009). A revolutionary land data assimilation framework used in ERA5 to absorb expected in-situ and the satellite interpretations for land surface elements, i.e., Soil temperature, moisture of soil, the temperature of snow, snow water, and snow thickness, as delineated in (Al-Yaari et al., 2014; de Rosnay et al., 2014). By decades ERA5 proceeded with numerical weather forecast advancements in numeric, model physics, and information integrated by employing ECMWF, ERA-Interim (Dee et al., 2011). With a horizontal tenacity at the equator, i.e., 31 km, since January 2019, ERA5 runoff data accessible from 1979 to the present. ERA5 has a strong peculiarity, i.e., it's working environment, making it accessible to generate timely products. ERA5T, authorizing the development of GloFAS-ERA5 derived river discharge reanalysis regularly with the inactivity of 2 and 5 days behind close real-time.

2.2 LISFLOOD derived river discharge

Presently river discharge is not intended in HTESSEL. Instead, the surface and sub-surface runoff acquired from HTESSEL terrestrial surface model joined with an essential worldwide variety of LISFLOOD, an aerially dispersed based on a grid, the hydrological and river model. Standardization of the Global Flood Awareness System GloFAS v2.1 utilizing everyday river flow records (Hirpa et al., 2018) is available but concisely reduced here for the condition. HTESSEL derived runoff from the sub-surface applied as input into the groundwater module LISFLOOD, which comprises two analogous undeviating reservoirs that hold and successively carry water towards the river network with delay in time. Speedy groundwater and subsurface flow designated for the higher zone. However, the lesser zone signifies sluggish groundwater flow, which produces base flow. The time for the higher zone assigned a value of 10 days by default with a lower and upper bound of 3 days and 40 days respectively in the course of standardization, and the time for the lesser zone was assigned a value of 200 days as the default value with a lower and upper bound of 40 days and 500 days correspondingly. The LISFLOOD river channel routing module uses surface runoff as input from HTESSEL. During this two-stage procedure, runoff from the surface for an individual cell is initially directed to the adjoining downstream river passage cell. Subsequently, the water through the channel is guided over the river system following the kinematic wave process. River routing and groundwater factors in GloFAS have been standardized despite daily river flow interpretations for 1287 basins altogether (Hirpa et al., 2018). LISFLOOD can characterize structures that can strictly modify the scheduling and river discharge volume, i.e., Lagoons, reservoirs, and uses of water by a human being (Burek et al., 2013). The major lakes having a sum of 463 (area of the surface > 100 km^2), and also 667 of the giant reservoirs have been integrated into the GloFAS (Zajac et al., 2017). River discharge reanalysis data have been created since 1st January 1979 to immediate real-time by GloFAS-ERA5 when the LISFLOOD model integrated alongside the daily HTESSEL surface and subsurface runoff. The runoff fields obtained from ERA5 rationalized utilizing the modest nearest neighbor technique from the inherent ERA5 to the GloFAS lattice. Escaping the necessity for a lengthy spin-up time, LISFLOOD computes a stable state storing volume aimed at the lesser groundwater region throughout an extensive period called "pre-run" and hence decreases the spin-up time of the lower zone (Burek et al., 2013). For this reason, a one-year spin-up time has been given to LISFLOOD by utilizing the initial output obtained from ERA5 for the year 1978.

2.3 Observed data

Daily river discharge data acquired for 86 hydrographic stations at which CWC (Central Water Commission, India; http://cwc.gov.in/hydro-meteorological-observation) continuously gauging the streamflow in major river basins of Indian-subcontinent, India-WRIS portal (https://indiawris.gov.in/wris/#/RiverMonitoring), and also from Global Runoff Data Centre (GRDC) (https://grdc.bafg.de).

2.4 Study area

This study focuses on the major river basins of India comprising of the Ganges, Brahmaputra, Mahanadi, Subarnarekha, Brahmani-Baitarani, Krishna, Godavari, Pennar, and Cauvery Rivers flowing in the east direction. The river basins in India have enormous variations in geographical extent. For example, Ganges, Indus, and Brahmaputra river basins have a span larger than 1 million km^{2,} while river catchments are of nominal size along the coastline. We have chosen the sub-catchments and the main subcontinental river catchments to recognize the spatial inconsistency for runoff from 1979-Present in India. Various conditions applied to elect locations for the evaluation are as follows:

a) A minimum of 4 - 5 years of daily discharge data availability during 1979 – 2018.
b) We selected the hydrographic station in the observed dataset, which matches the Glofas grid cell with the longest historical records.
c) Stations close to the river mouth are also retained based on the availability of data to observe severe fluctuations.
d) In addition, the location of hydrographic stations overlaps on the digital elevation map with major and minor basins of India are presented in Figure 1.

3 Methods

This study attempts to understand the seasonal, annual variations and the performance of the reanalysis river discharge. Especially for the major river basins of the eastern region of India, by equating modeled reanalysis with those in situ river discharge. Performance metrics in hydrological modeling are essential. The performance metrics are estimated based on the dissimilarities between the simulated and observed river discharge at the basin outlet. Analysis of large samples exhibits significant sampling ambiguity in the estimators of NSE and KGE′ (Clark et al., 2021). The methodology used in this study is the advanced Kling-Gupta efficiency (KGE′; Gupta et al., 2009; Kling et al., 2012) for evaluation of the hydrological metrics. For the skill assessment of the hydrological datasets, statistical analysis is practical, such as the correlation coefficient, bias, and variability between the observed and simulated datasets. KGE′ is advanced in usage because of the standard expression of metrics in hydrology (Beck et al., 2017; Harrigan et al., 2018; Lin et al., 2019). Correspondingly, its ability to simply disintegrate into its three components which are significant to examine hydrological dynamics: chronological errors in bias, Pearson correlation, and variability ratio:

$$\text{KGE}' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{1}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{2}$$

$$\gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{3}$$

where Pearson correlation coefficient indicated by r among reanalysis and observed river discharge data, bias ratio by β, variability ratio by γ, the mean by μ, and the standard deviation of river discharge by σ. The KGE′ and the disintegrated components (i.e., bias ratio, correlation, and the variability ratio) are unitless with an ideal value of 1. The performance of the dataset equated against a simple benchmark (observed data) to examine the skill of GloFAS-ERA5 reanalysis derived river discharge (Knoben et al., 2019). The benchmark is required as a minimum reference to evaluate the simulated hydrological data. This study used KGE′ as a skill score, KGESS to estimate the skill of GloFAS-ERA5 derived reanalysis river discharge data against the benchmark, specified such as:

$$\text{KGESS} = \frac{KGE'_{reanalysis} - KGE'_{bench}}{KGE'_{perf} - KGE'_{bench}} \tag{4}$$

The value of KGE′ is calculated for the GloFAS-ERA5 derived reanalysis against observed discharge and presented by $KGE'_{reanalysis}$, KGE'_{bench} is the KGE′ value for the perceived mean, standard flow against observed, i.e., KGE′ ($\overline{Q}_{obs}$) = $1 - \sqrt{2} \approx -0.41$ given by Knoben et al., (2019), and the value of KGE′ for the perfect simulation, i.e., 1, is presented by KGE'_{perf}. If KGESS = 0, the reanalysis river discharge is poor than the mean flow benchmark and has no skill. KGESS > 0 shows that the reanalysis is skillful, while KGESS < 0 shows that the reanalysis is inferior to the benchmark and has unfavorable skill.

4 Results

The reanalysis product of river discharge has been taken from GloFAS-ERA5 v2.1 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview). We have assessed the reanalysis product against the in situ river discharge to examine the hydrologic metrics of the selected station utilizing the hydrostats function.

4.1 Overall performance and Disintegration of KGE′ into Bias, Variability, and Correlation

A unique advantage of KGE′ is that it can disintegrate into its three important factors, i.e., bias, variability, and correlation, so the performance of GloFAS-ERA5 reanalysis can assess against the observed data as good or poor skill. The KGESS was used to achieve a skill score for monthly reanalysis and observed river discharge at each selected hydrographic station. The hydrographic stations show a positive Pearson correlation coefficient with a median value of 0.73 and interquartile ranges between 0.51-0.83. The bias ratio was calculated; and observed that with a median value of 1.01 and interquartile range (0.05, 1.39), the high bias restrict to a few locations in Krishna and Godavari River basins (Figure 4). The rest of the hydrographic stations show low bias. Almost 60% of hydrographic stations show the reanalysis discharge is negatively biased (i.e., bias < 1). Nearly 58% of hydrograph stations show lower variability (i.e., variability ratio < 1), the median value of 0.91, and the interquartile range (0.82, 1.13), as shown in Figure 3 (iii). The results show that about 80% of the total hydrographic stations are skillful with a median KGESS (KGE′) 0.26 (0.797) and an interquartile range of 0.18, 0.73 (-0.147, 0.63), respectively. The poorest performing location (KGESS value presented by a dark red dot in Figure 5e) is predominantly due to a considerable bias in dryer tributaries of Krishna and Godavari Rivers. Furthermore, the significance of the average magnitude of errors is also important for over or underestimation in dry rivers. A significant error can yield a high proportion of bias (i.e., bias ratio) as we found very high mean absolute error at Hardinge Bridge and Bahadurabad hydrographic stations for Ganges and Brahmaputra River basins, respectively in Figure 3 (iv).

From a probability density point of view, the disintegrated component of KGE′, i.e., correlation, shows an almost normal distribution with a mean value of $\mu_1 = 0.67$. The bias ratio shows a right-skewed distribution with a mean value $\mu_2 = 0.86$, suggesting that the reanalysis discharge has low bias. The variability ratio also exposes right-skewed distribution with mean value $\mu_3 = 1.12$, suggesting the presence of low variability in the GloFAS-ERA5 reanalysis river discharge. The Kling-Gupta efficiency shows an almost normal distribution with a mean value $\mu_1 = 0.15$. The Kling-Gupta efficiency skill score shows a left-skewed distribution with a mean value $\mu_2 = 0.4$ as shown in Figure 6 (a-d), respectively.

4.2 Performance by Basin area

The skillfulness of GloFAS-ERA5 derived reanalysis river discharge grouped into eight major and two minor river basins among east flowing rivers of the Indian subcontinent. The median of the correlation coefficient was found to be 0.94, 0.68 with an interquartile range (0.94-0.95, 0.64-0.75) for Ganges-Brahmaputra and Subarnarekha basins, respectively. The median bias ratio was found to be 1.11, 0.06 with an interquartile range (1.07-1.14, 0.009-0.91) for Ganges-Brahmaputra, Subarnarekha basins, respectively. The variability ratio has the median value of 0.93, 1.13 with an interquartile range (0.81-1.01, 0.97-1.29) for Ganges-Brahmaputra and Subarnarekha basins, respectively. Median value of KGESS = 0.86, 0.26 (KGE′ = 0.81, -0.03) was found for the Ganges-Brahmaputra, Subarnarekha River basins with an interquartile range 0.82-0.90, 0.24-0.76 (0.75-0.86, -0.06-0.67), respectively, for rest of the basins the hydrologic metrics have been given in Table 1. Moreover, skill is lowest for Cauvery, Pennar River basin, and the rivers flowing between Pennar and Kanyakumari while good for Ganges-Brahmaputra River basins. Also, moderate to good skill observed for Subarnarekha, Brahmani-Baitarani, Mahanadi, Godavari, Krishna, and the rivers flowing between Mahanadi and Pennar River basins. Usually, skill varies according to the size or area of the basin, and the results are also analogous, as suggested by Harrigan et al. (2020). The disintegrated component of KGE′ in Figure 7 (a-d) and the performance of major and minor River basins in India by KGESS have shown in Figure 7 (e).

4.3 Statistical distribution by Quantile-Quantile (Q-Q) plot

The significance of statistical distribution is essential in hydrological and hydro-meteorological studies of river discharge. For instance, in intensity–duration–frequency (IDF) networks, storm designation, and precise assessment of rainfall are the primary input to numerous hydro-meteorological usage (Maghsood et al., 2020). The quantile-quantile (Q-Q) plot helps to determine whether the two datasets, i.e., observed and GloFAS-ERA5 reanalysis, have similar distribution patterns. The strategy is led by plotting quantiles of the two datasets versus each other and contrasting the plot with a 45˚ reference line (1:1). Likewise, the Q-Q plot is a dispersed plot, with the data quantiles falling roughly along the reference line (Figure 3), representing a typical distribution for the two datasets. In reality, the more significant evidence for rejecting the presumption of common distribution is moving away from the reference line. The quantiles of a dataset must be the points underneath which a specific extent of the information lies. For instance, in an exemplary standard typical likelihood assumption with a mean of 0, the 0.5 quantile (or 50th percentile), 0 implies that a large portion of the data does not surpass 0. There are additionally aware techniques, such as the chi-square and Kolmogorov–Smirnov 2-example tests, which evaluate if two arrangements of quantiles follow a similar distribution. Nonetheless, the Q-Q plot is good as it better understands the distinction between two datasets than systematic schemes. The Q-Q plot is simply a visual check instead of confirmation, and it supports noting if the hypothesis is feasible or not. It marks the data points of those quantiles that cause the violation of the uncertainty. The Q-Q plot shows the under or overestimation of a dataset without any stretch. The GloFAS-ERA5 reanalysis contrasted with the observed river discharge between percentiles of the datasets. Also, numerous distributional perspectives, remembering shifts for the area, changes in scale, change in consistency, tail conduct, and the presence of exceptions, can be observed. The deportment of the tail of the Q - Q plot can be significant for extreme events (e.g., floods and droughts) studies in hydrology.

4.4 Seasonal variations in river discharge

Monsoon in India, particularly the southwest monsoon, has a massive impact on river discharge. The monsoon is a consequence of a complex interaction among the ocean, atmosphere, and land. Although the consistency of monsoons is closely certain, its interannual variability is of critical worry regarding drought, normal, and flood years. The entire country receives almost 75% of the rainfall during this period. The intense precipitation during summer leads to high magnitude floods, although the rivers deliver only a low base stream during the dry winters. For numerous rivers in the sub-humid to subtropics, adjoining the rainy season realm, precipitation throughout the rainy spell is also the key source of surface water rejuvenation (Plink-Björklund, 2015). The hydrographic stations, which have prolonged historical river discharge and proximity to the Bay of Bengal analyzed to understand the large-scale variations in river discharge for the major river basins of India. The observed and reanalysis of daily river discharge have been reformed into monthly records to perceive the monthly or seasonal variations in the major river basins of India. The river discharge was found low at the selected hydrographic stations during the winter season (December-January-February). Throughout the pre-monsoon (March-April-May) period, changes in river discharge were insignificant at the selected hydrographic stations. Both reanalysis and in situ discharge show a severe change in river discharge during the southwest monsoon at each hydrographic station. Furthermore, very high fluctuations in river discharge through the southwest monsoon period have been reported at Hardinge Bridge and Bahadurabad hydrographic station in Ganga and Brahmaputra river basins, respectively. In contrast, Krishna, Godavari, and Pennar rivers show a moderate to lower variability in rivers discharge. The post-monsoon or northeast monsoon (October-November) period shows less river discharge or base streamflow than the southwest. Furthermore, the Ganga, Brahmaputra River shows less variability. In contrast, Krishna, Pennar, and Cauvery rivers show a significant change in river discharge due to the receding of the southwest monsoon into the northeast monsoon. Intra-annual or seasonal variations in river discharges have been displayed in Figure 8 for various river basins at the selected hydrographic stations.

5 Discussion

As per the region, the KGESS value shifts in light of the tendency in gauge river discharge. On the one hand, river basins and hydrological stations with positive bias in the gauge simulation significantly improved the skill score. Then again, those with negative bias showed the least skill score (Hirpa et al., 2018). This study focused on the regional scale to know the skill assessment of reanalysis against in situ river discharge for India's major and minor river catchments.

5.1 Importance of bias term

As deliberated by Santos et al. (2018), the importance of the bias term in KGE′, $\beta = \mu_s/\mu_o$ can prompt extremely large values of β (and thus low KGE′ scores) when μ_o is small. Such issues with intensified β values are conceivably more expressed for factors where μ_o crosses zero (e.g., log-changed streams, temperature) on the grounds that μ_o could be small. Negative bias has also been observed at several hydrographic stations. At the same time, only a few show high bias value because a small number of errors are too associated with the KGE′ metrics. Referring to shortcomings of the NSE as support, a piece of the local community has changed to using KGE′ over NSE (Clark et al., 2021) and opposed that this did not take care of; however, it just changed the issues identified with framework scale measurements. Moreover, the results are analogous with long-term river discharge reanalysis globally (Harrigan et al., 2020).

5.2 Significance of Mean Absolute Error (MAE)

The importance of the large mean absolute error is also noticeable at few hydrographic stations located in the Ganges-Brahmaputra River catchments. Our results are also similar to those by Harrigan et al. (2020) for the major river catchments of the world. The significance of magnitude of mean absolute error is essential for over or underestimation in dry rivers. Large values of MAE can yield a high proportion of bias (i.e., bias ratio) as we found very high mean absolute error at Hardinge Bridge and Bahadurabad hydrographic stations for Ganges and Brahmaputra River basins, respectively in Figure 3 (ii).

5.3 Importance of benchmark

The dataset's advancement would look at the output score against a specific benchmark that can guide which type of model performance could be normal (Seibert et al., 2019) and choose whether the model is skillful. The valuable strength of the model must be clear from the consideration of benchmark and skill approval of the model with the end goal such that the modeler can settle on an advanced result (Bettina Schaefli, 2010). Characterizing such benchmarks is unclear since it depends on the interchange between our current hydrologic understanding, the accessibility and the nature of observations, the decision of model construction, and boundary conditions. In any case, describing the advanced benchmarks will permit more strong assessments of model implementation (Abramowitz, 2012). The most effective method to define a relative structure inside hydrology is an open query into the local hydrological study.

5.4 Implementation of hydrologically significant features into a single group

According to the hydrological point of view, the performance metrics with a realistic nature like KGE′ did not provide the information about the model's lack (Gupta et al., 2008). However, KGE′ improves the NSE metric positively; Gupta et al. (2009) precisely conveyed that their purpose with KGE′ was not to plan an advanced part to implement in the model. Furthermore, they also highlighted a noticeable shortcoming of the KGE′ metric where important hydrological features are combined into a single component in the model. They used the measurements to outline that there are inherent issues with mean squared error-based methods. Also, there is no motivation to use collected measurements and study the model behavior on the individual time-step (Keith Beven and Philip Younger, 2014).

6 Conclusions

Hydrostats package has been used in this study to determine the performance of hydrological metrics. The skill evaluation of GloFAS reanalysis river discharge was analyzed by employing the revised KGESS method for the major Indian sub-continental basins, principally Ganges-Brahmaputra, Subarnarekha, Mahanadi, Brahmani and Baitarani, Mahanadi, Krishna, Godavari, Pennar, Cauvery, the rivers flowing in between Pennar and Mahanadi, and the rivers flowing in between Kanyakumari and Pennar. The GloFAS reanalysis discharge is vital in dual steps; one is scheming the flood arrivals against which real-time predictions equated to control the possibility of a flood indication. Second is the supplementary reliable hydro-meteorological settings for real-time flood and sporadic forecasts. The correlation coefficient was very high for the Ganges-Brahmaputra River basins, with a significant mean absolute error at few stations. The assessment shows that the GloFAS derived river discharge is skillful in all the river basins except some hydrographic stations. The seasonal variations have been observed in river discharge at various hydrological stations located inside several river basins of India. Both in situ and reanalysis datasets captured significant seasonal variability in river discharge during the southwest and less variations during the northeast monsoon. Moreover, the datasets which emanate from a common distribution have been marked by a 45˚ reference line. In contrast, a departure from the reference line shows that the datasets following different distribution as demonstrated by the Quantile-Quantile (Q-Q) plot at the selected hydrographic stations of various river catchments.

Data Availability Statement

Daily observed river discharge data is available at CWC (Central Water Commission, India; http://cwc.gov.in/hydro-meteorological-observation) through the India-WRIS portal (https://indiawris.gov.in/wris/#/RiverMonitoring), and Global Runoff Data Centre (GRDC) (https://grdc.bafg.de). Daily reanalysis of river discharge available at Climate Data Services-European Center for Medium-Range Weather Forecasts (CDS-ECMWF (https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview). Hydrostats function is available at (https://hydrostats.readthedocs.io/en/latest/) built under python scripts.

Acknowledgment

The author is thankful to the Ministry of Human Resources Development (MHRD) India and the Indian Institute of Technology Kharagpur India for providing an interactive research environment. The author is also thankful to the Central Water Commission (CWC) India, Global Runoff Data Centre (GRDC), for providing the in situ observed data. And also grateful to Climate Data Services-European Center for Medium-Range Weather Forecasts (CDS-ECMWF) for reanalysis of river discharge data to conduct this study. All the graphics presented in this manuscript were obtained using the python programming language.

References

Abramowitz, G. (2012). Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geoscientific Model Development*, *5*(3), 819–827. https://doi.org/10.5194/gmd-5-819-2012

Al-Yaari, A., Wigneron, J.-P., Ducharne, A., Kerr, Y., de Rosnay, P., de Jeu, R., et al. (2014). Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates. *Remote Sensing of Environment*, *149*, 181–195. https://doi.org/10.1016/j.rse.2014.04.006

Arnell, N. W. (1999). The effect of climate change on hydrological regimes in Europe: A continental perspective. *Global Environmental Change*, *9*(1), 5–23. https://doi.org/10.1016/S0959-3780(98)00015-6

Balsamo, G., Viterbo, P., Beijaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., & Scipal, K. (2009). A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, *10*(3), 623–643. https://doi.org/10.1175/2008JHM1068.1

Beck, H. E., Van Dijk, A. I. J. M., De Roo, A., Dutra, E., Fink, G., Orth, R., & Schellekens, J. (2017). Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrology and Earth System Sciences*, *21*(6), 2881–2903. https://doi.org/10.5194/hess-21-2881-2017

Bettina Schaefli, H. V. G. (2010). Do Nash values have value?, *2274*(November 2008), 2267–2274. https://doi.org/10.1002/hyp

Bharat, S., & Mishra, V. (2020). Runoff sensitivity of Indian sub-continental river basins. *Science of the Total Environment*, (xxxx), 142642. https://doi.org/10.1016/j.scitotenv.2020.142642

Brouwer, R., Akter, S., Brander, L., & Haque, E. (2007). Socioeconomic vulnerability and adaptation to environmental risk: A case study of climate change and flooding in Bangladesh. *Risk Analysis*, *27*(2), 313–326. https://doi.org/10.1111/j.1539-6924.2007.00884.x

Chinnasamy, P., Maheshwari, B., & Prathapar, S. (2015). Understanding groundwater storage changes and recharge in Rajasthan, India through remote sensing. *Water (Switzerland)*, *7*(10), 5547–5565. https://doi.org/10.3390/w7105547

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., et al. (2021). The abuse of popular performance metrics in hydrologic modeling. *Water Resources Research*. https://doi.org/10.1029/2020wr029001

Das, J., Treesa, A., & Umamahesh, N. V. (2018). Modelling Impacts of Climate Change on a River Basin: Analysis of Uncertainty Using REA & Possibilistic Approach. *Water Resources Management*, *32*(15), 4833–4852. https://doi.org/10.1007/s11269-018-2046-x

Gupta, H. V., Kling, H., Yilmaz, K. K., & 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), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., & Tanguy, M. (2018). Benchmarking ensemble streamflow prediction skill in the UK. *Hydrology and Earth System Sciences*, *22*(3), 2023–2039. https://doi.org/10.5194/hess-22-2023-2018

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., et al. (2020). GloFAS-ERA5 operational global river discharge reanalysis 1979-present. *Earth System Science Data*, *12*(3), 2043–2060. https://doi.org/10.5194/essd-12-2043-2020

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., & Dadson, S. J. (2018). Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology*, *566*(March), 595–606. https://doi.org/10.1016/j.jhydrol.2018.09.052

Hoshin V. Gupta, T. W. and Y. L. (2010). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Okt 2005 Abrufbar Uber Httpwww Tldp OrgLDPabsabsguide Pdf Zugriff 1112 2005*, *2274*(November 2008), 2267–2274. https://doi.org/10.1002/hyp

Jain, S. K., Sharma, A., & Kumar, R. (2004). Freshwater and its management in india. *International Journal of River Basin Management*, *2*(4), 259–270. https://doi.org/10.1080/15715124.2004.9635236

Keith Beven, Philip Younger, and J. F. (2014). Struggling with Epistemic Uncertainties in Environmental Modelling of Natural Hazards Vulnerability, Uncertainty, and Risk ©ASCE 2014 13, 13–22.

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424–425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

van der Knijff, J. M., Younis, J., & de Roo, A. P. J. (2010). LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science*, *24*(2), 189–212. https://doi.org/10.1080/13658810802549154

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Krishna Kumar, K., Patwardhan, S. K., Kulkarni, A., Kamala, K., Koteswara Rao, K., & Jones, R. (2011). Simulated projections for summer monsoon climate over India by a high-resolution regional climate model (PRECIS). *Current Science*, *101*(3), 312–326.

Lakshmi, V., Fayne, J., & Bolten, J. (2018). A comparative study of available water in the major river basins of the world. *Journal of Hydrology*, *567*(October), 510–532. https://doi.org/10.1016/j.jhydrol.2018.10.038

Lewis, J. S., Barani, Z., Magana, A. S., & Kargar, F. (2019). ce pte d M us pt, 0–31.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., et al. (2019). Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Maghsood, F. F., Hashemi, H., Hosseini, S. H., & Berndtsson, R. (2020). Ground validation of GPM IMERG precipitation products over Iran. *Remote Sensing*, *12*(1), 1–23. https://doi.org/10.3390/RS12010048

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., & Clark, M. (2010). Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*, *24*(10), 1270–1284. https://doi.org/10.1002/hyp.7587

Mishra, V., & Lilhare, R. (2016). Hydrologic sensitivity of Indian sub-continental river basins to climate change. *Global and Planetary Change*, *139*, 78–96. https://doi.org/10.1016/j.gloplacha.2016.01.003

Mishra, V., Thirumalai, K., Singh, D., & Aadhar, S. (2020). Future exacerbation of hot and dry summer monsoon extremes in India. *Npj Climate and Atmospheric Science*, *3*(1), 1–9. https://doi.org/10.1038/s41612-020-0113-5

Plink-Björklund, P. (2015). Morphodynamics of rivers strongly affected by monsoon precipitation: Review of depositional style and forcing factors. *Sedimentary Geology*, *323*, 110–147. https://doi.org/10.1016/j.sedgeo.2015.04.004

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). The

Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, *460*(7258), 999–1002. https://doi.org/10.1038/nature08238

de Rosnay, P., Balsamo, G., Albergel, C., Muñoz-Sabater, J., & Isaksen, L. (2014). Initialisation of Land Surface Variables for Numerical Weather Prediction. *Surveys in Geophysics*, *35*(3), 607–621. https://doi.org/10.1007/s10712-012-9207-x

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018

Seibert, J., Vis, M., Lewis, E., & Meerveld, I. Van. (2019). Seibert J , Vis MJP , Lewis E , Meerveld HJ . Upper and lower benchmarks in hydrological modeling, *32*(February), 1120–1125.

Seo, J. Y., & Lee, S. Il. (2017). Total discharge estimation in the Korean Peninsula using multi-satellite products. *Water (Switzerland)*, *9*(7). https://doi.org/10.3390/w9070532

Siddique-E-Akbor, A. H. M., Hossain, F., Sikder, S., Shum, C. K., Tseng, S., Yi, Y., et al. (2014). Satellite Precipitation Data–Driven Hydrological Modeling for Water Resources Management in the Ganges, Brahmaputra, and Meghna Basins. *Earth Interactions*, *18*(17), 1–25. https://doi.org/10.1175/ei-d-14-0017.1

Sikder, M. S., David, C. H., Allen, G. H., Qiao, X., Nelson, E. J., & Matin, M. A. (2019). Evaluation of Available Global Runoff Datasets Through a River Model in Support of Transboundary Water Management in South and Southeast Asia. *Frontiers in Environmental Science*, *7*(October), 1–18. https://doi.org/10.3389/fenvs.2019.00171

Vano, J. A., Das, T., & Lettenmaier, D. P. (2012). Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *Journal of Hydrometeorology*, *13*(3), 932–949. https://doi.org/10.1175/JHM-D-11-069.1

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F., & Beck, H. (2017). The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*, *548*, 552–568. https://doi.org/10.1016/j.jhydrol.2017.03.022

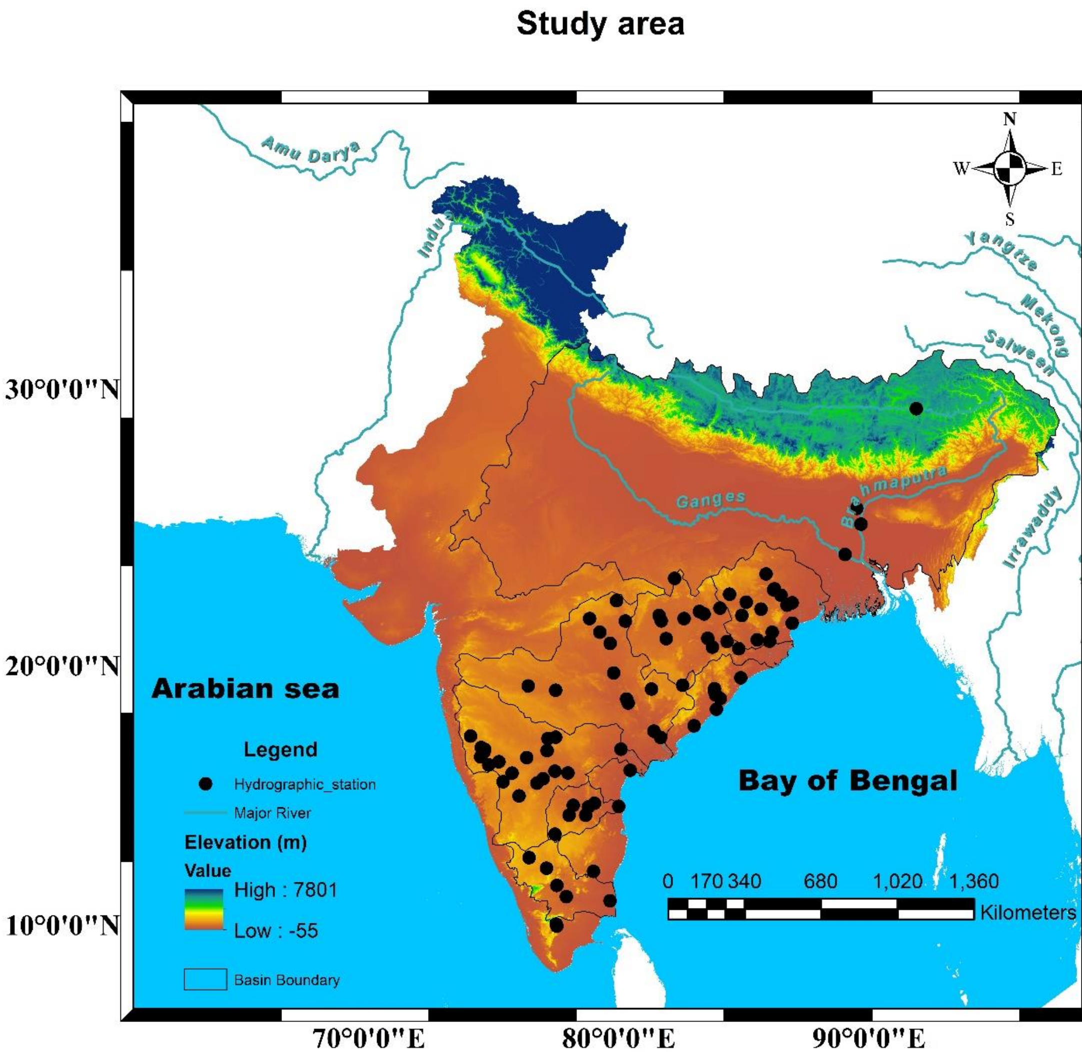


Figure 1. The study area includes the river network of major and minor river basins of India with the extent in adjacent countries and the Hydrographic stations' location and the elevation (m).

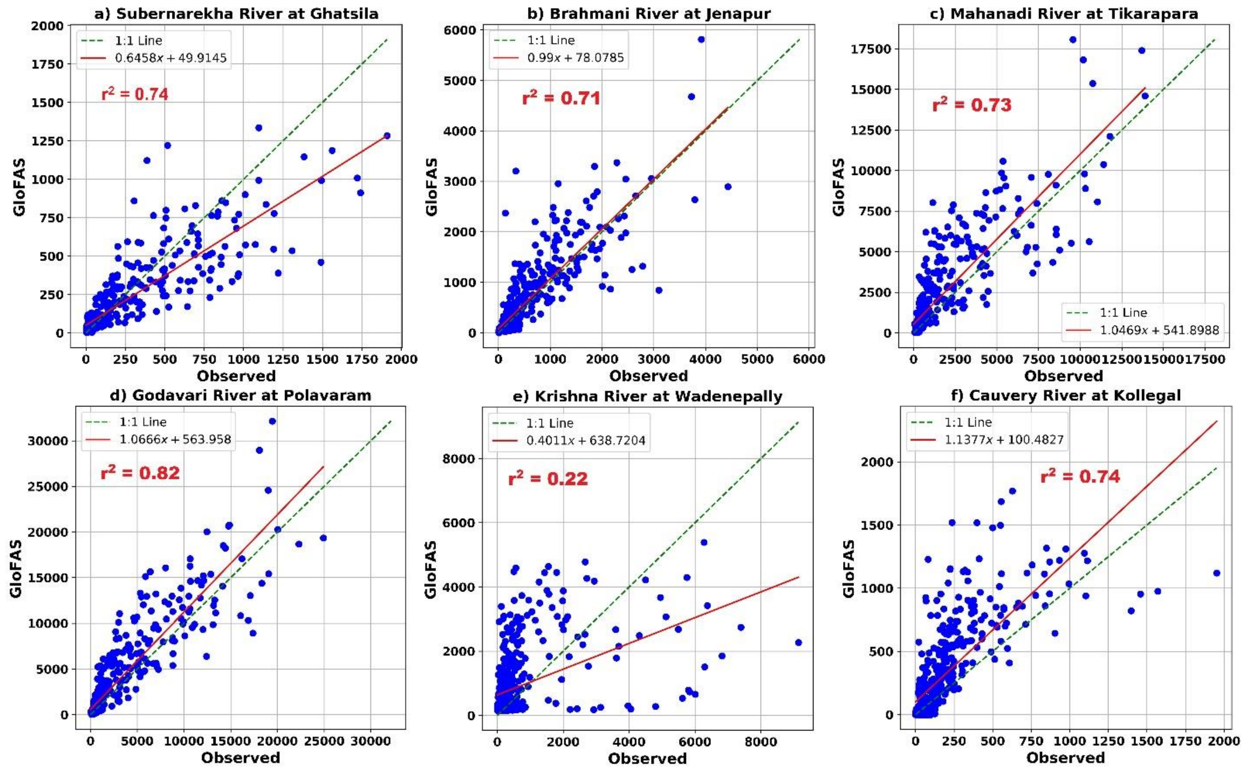


Figure 2. Linear regression between GloFAS reanalysis and observed river discharge at the respective hydrographic stations of major river basins of India alongside for the period 1979 - 2018.

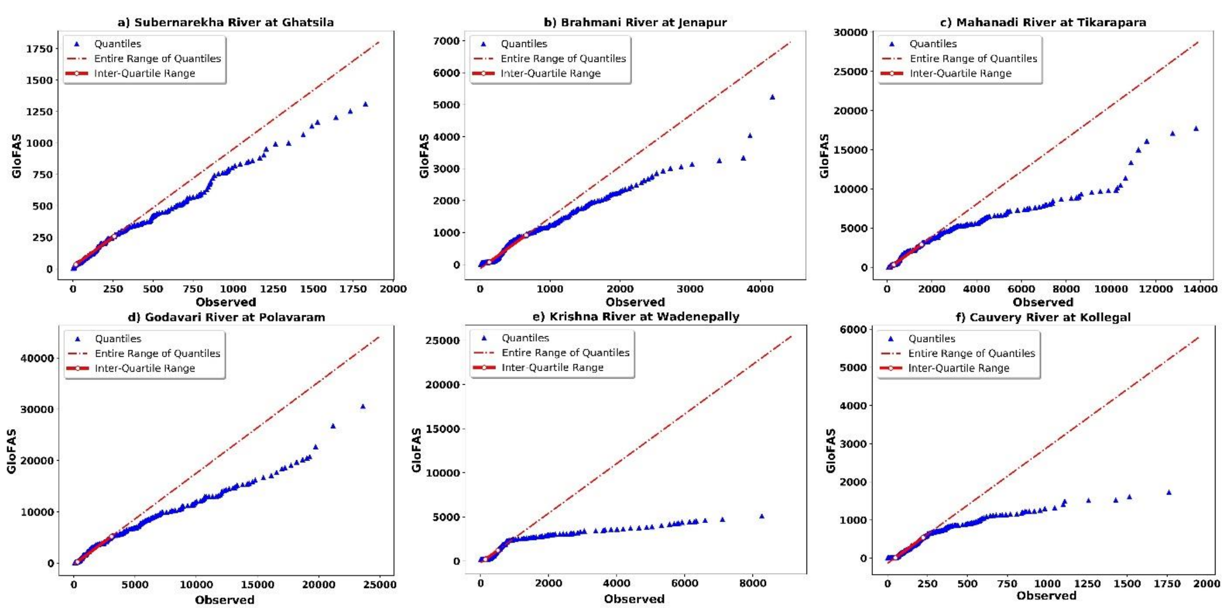
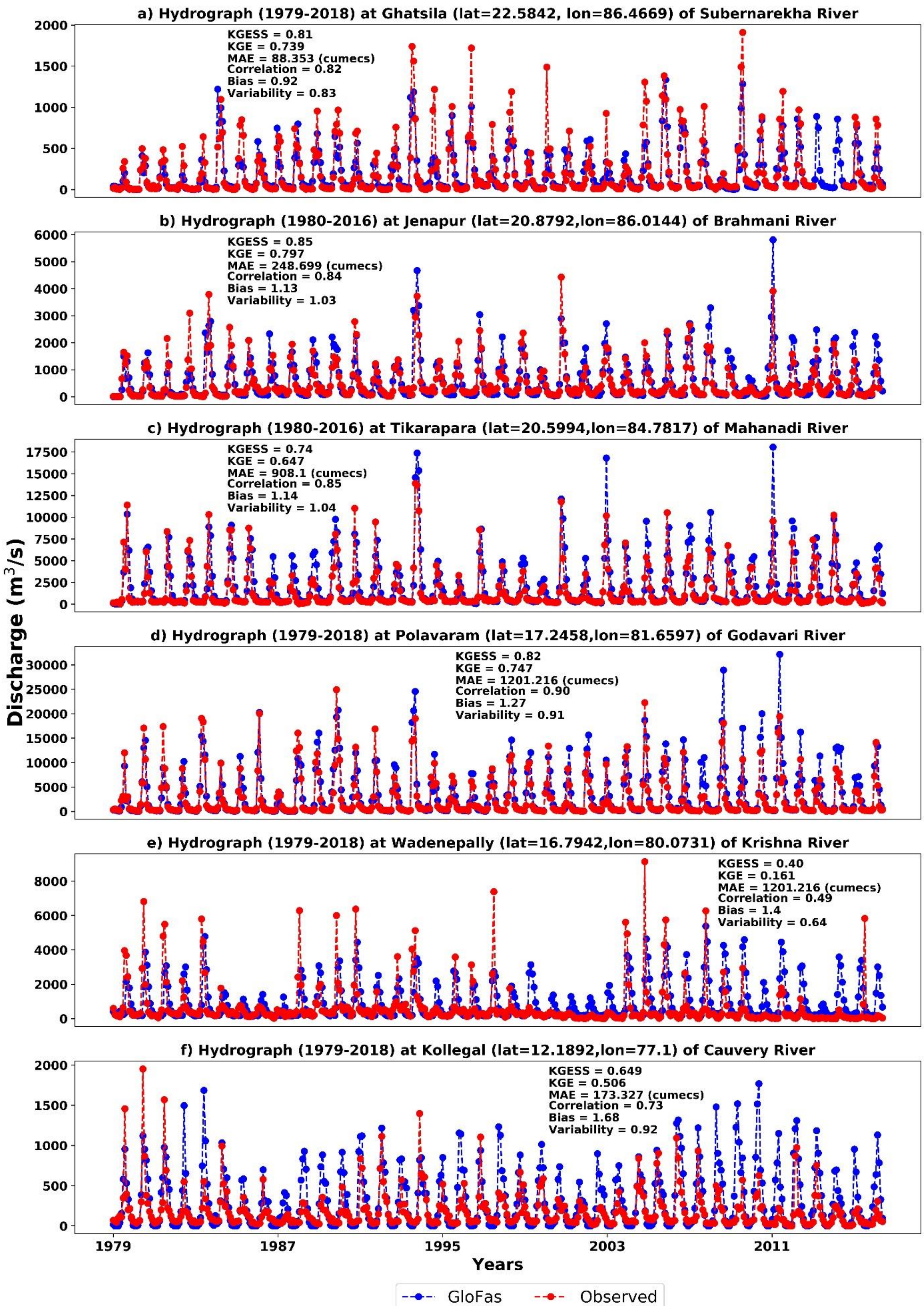


Figure 3. Quantile - Quantile (Q - Q) plots at the selected hydrographic stations with their respective river basins in between GloFAS and observed river discharge (m^3/s).


a) Hydrograph (1979-2018) at Ghatsila (lat=22.5842, lon=86.4669) of Subernarekha River
KGESS = 0.81
KGE = 0.739
MAE = 88.353 (cumecs)
Correlation = 0.82
Bias = 0.92
Variability = 0.83
b) Hydrograph (1980-2016) at Jenapur (lat=20.8792,lon=86.0144) of Brahmani River
KGESS = 0.85
KGE = 0.797
MAE = 248.699 (cumecs)
Correlation = 0.84
Bias = 1.13
Variability = 1.03
c) Hydrograph (1980-2016) at Tikarapara (lat=20.5994,lon=84.7817) of Mahanadi River
KGESS = 0.74
KGE = 0.647
MAE = 908.1 (cumecs)
Correlation = 0.85
Bias = 1.14
Variability = 1.04
d) Hydrograph (1979-2018) at Polavaram (lat=17.2458,lon=81.6597) of Godavari River
KGESS = 0.82
KGE = 0.747
MAE = 1201.216 (cumecs)
Correlation = 0.90
Bias = 1.27
Variability = 0.91
e) Hydrograph (1979-2018) at Wadenepally (lat=16.7942,lon=80.0731) of Krishna River
KGESS = 0.40
KGE = 0.161
MAE = 1201.216 (cumecs)
Correlation = 0.49
Bias = 1.4
Variability = 0.64
f) Hydrograph (1979-2018) at Kollegal (lat=12.1892,lon=77.1) of Cauvery River
KGESS = 0.649
KGE = 0.506
MAE = 173.327 (cumecs)
Correlation = 0.73
Bias = 1.68
Variability = 0.92
Discharge (m³/s)
Years
1979
1987
1995
2003
2011
GloFas
Observed

539

Figure 4. Monthly time series (Hydrograph) plot at the respective hydrographic stations of major river basins of India together with the hydrological metrics for the time period 1979 - 2018.

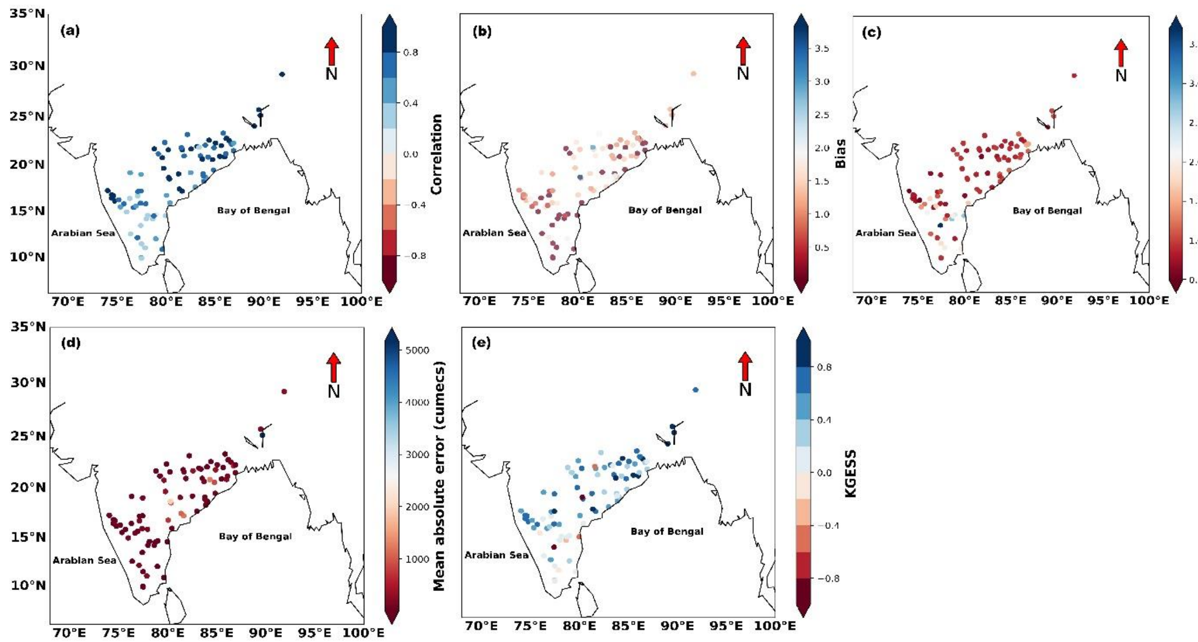


Figure 5. Spatial distribution of hydrologic metrics (a) correlation, (b) bias, (c) variability, (d) mean absolute error (MAE) (cumecs), and (e) Kling-Gupta efficiency skill score (KGESS) displayed by scatter plots for all the hydrographic stations for which the datasets collected.

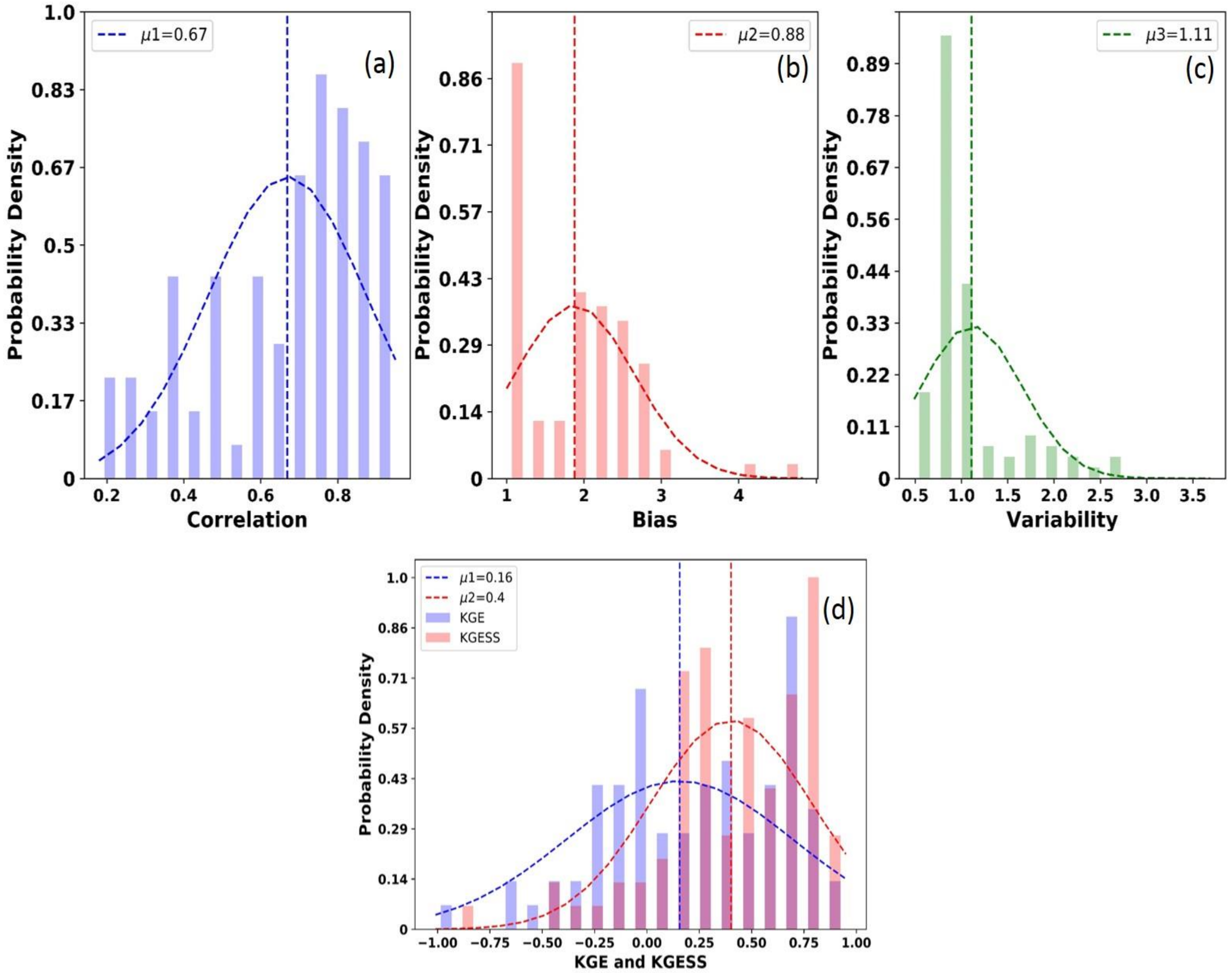


Figure 6. The Probability density of the decomposed component of KGE′ (a) Pearson correlation, (b) bias, (c) variability, and (d) Kling-Gupta efficiency (KGE′) along with Kling-Gupta efficiency skill score (KGESS).

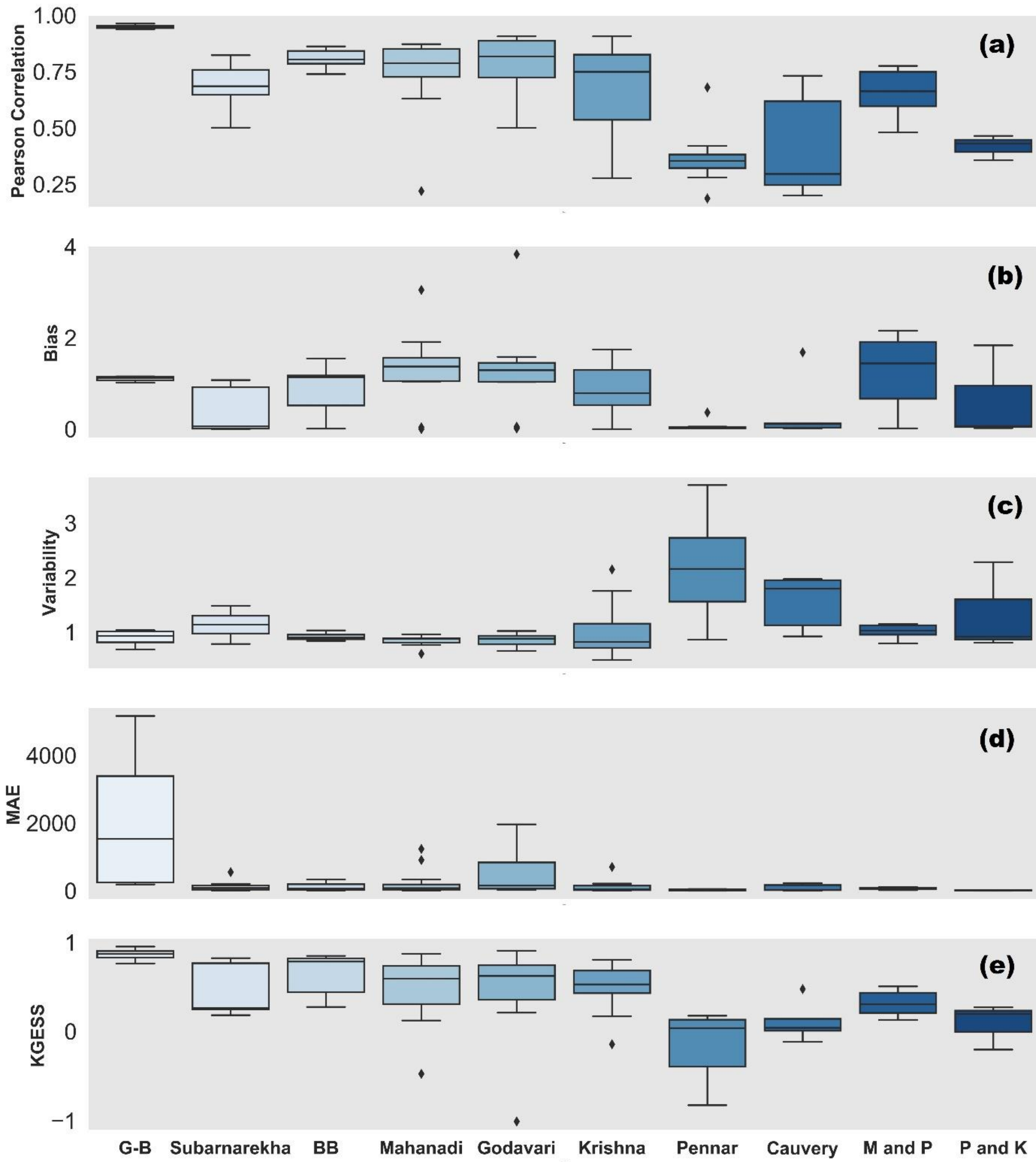


Figure 7. Basin wise performance of hydrologic metrics (a) correlation, (b) bias, (c) variability, (d) mean absolute error (MAE) (cumecs), and (e) Kling-Gupta efficiency skill score (KGESS) displayed by box plots. Where G-B (Ganges-Brahmaputra), BB (Brahmani and Baitarani), M and P (rivers flowing in between Mahanadi and Pennar), P and K (rivers flowing in between Pennar and Kanyakumari).

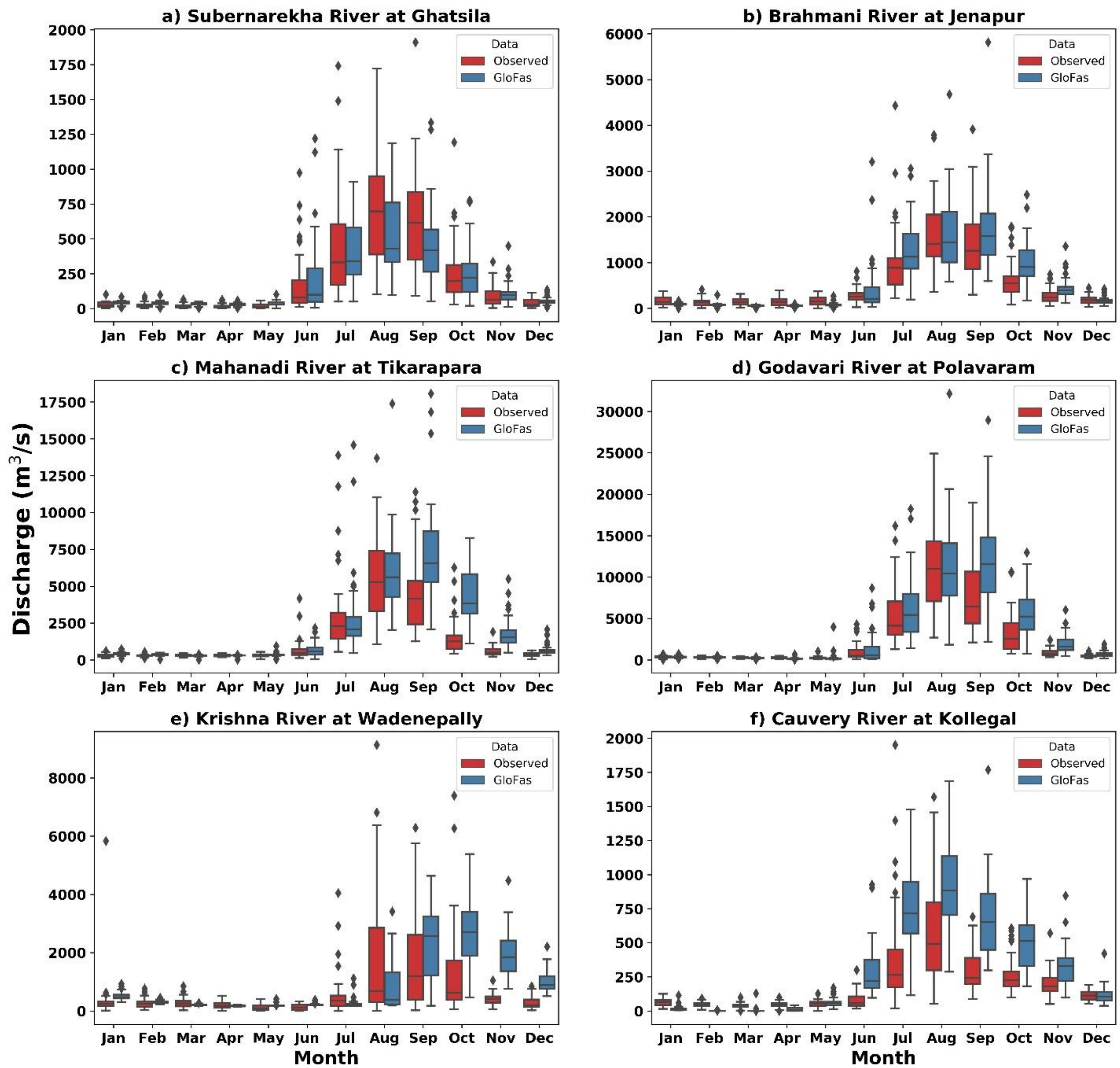


Figure 8. Seasonal changes in both datasets the GloFAS and observed river discharge at the hydrographic stations of its respective River basin are presented by Box plots.

Table 1. Hydrometrics for the major and minor river basins of India.

Sl. No.	Basin	Area (km^2)	KGE′ Median (IQR)	KGESS Median (IQR)	Pearson correlation Median (IQR)	Bias Median (IQR)	Variability Median (IQR)
1.	Ganges-Brahmaputra	1.7 million	0.81 (0.75-0.86)	0.86 (0.82-0.90)	0.94 (0.94-0.95)	1.11 (1.07-1.14)	0.93 (0.81-1.01)
2.	Subarnarekha	29,196	-0.03 (-0.06-0.67)	0.26 (0.24-0.76)	0.68 (0.64-0.75)	0.06 (0.009-1.29)	1.13 (0.97-1.29)
3.	Brahmani-Baitarani	51,822	0.69 (0.21-0.74)	0.78 (0.44-0.81)	0.8 (0.78-0.84)	1.13 (0.51-1.17)	0.9 (0.86-0.95)
4.	Mahanadi	74,970	0.42 (0.02-0.65)	0.59 (0.30-0.75)	0.78 (0.72-0.85)	1.37 (1.05-1.56)	0.87 (0.8-0.88)
5.	Godavari	3,12,812	0.46 (0.09-0.63)	0.62 (0.35-0.74)	0.81 (0.72-0.88)	1.28 (1.04-1.45)	0.88 (0.77-0.93)
6.	Krishna	2,58,948	0.33 (0.16-0.55)	0.52 (0.41-0.68)	0.74 (0.53-0.82)	0.78 (0.52-1.29)	0.82 (0.71-1.15)
7.	Pennar	50,493	-0.35 (-0.96-0.22)	0.03 (-0.39-0.13)	0.35 (0.32-0.38)	0.03 (0.01-0.04)	2.15 (1.55-2.72)
8.	Cauvery	81,155	-0.35 (-0.39-0.22)	0.04 (0.008-0.14)	0.29 (0.24-0.61)	0.10 (0.03-0.13)	1.79 (1.12-1.95)
9.	River flowing in between Mahanadi and Pennar	86,643	0.02 (-0.11-0.20)	0.28 (0.2-0.43)	0.66 (0.59-0.74)	1.44 (0.66-1.91)	1.02 (0.95-1.12)
10.	River flowing in between Pennar and Kanyakumari	1,00,139	-0.13 (-0.41-0.08)	0.19 (-0.003-0.23)	0.43 (0.39-0.44)	0.07 (0.04-0.95)	0.91 (0.86-1.6)