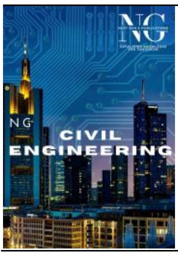

NG Civil Engineering
ISSN (online): 3107-569X
Volume 1, Issue 4 (October-December), 2025, (15-21)

NEXT GEN ACADEMIC SERVICES
CATALYZING KNOWLEDGE FOR TOMORROW

Research Article *Open Access*

Daily River Flow Forecasting in the Baitarani River, Odisha

Arpan Pradhan

Department of Civil Engineering, CHRIST (Deemed to be University), Bangalore, 560074, Karnataka, India

Corresponding author email ID: dr.arpanpradhan@gmail.com

HIGHLIGHTS

- ARIMA successfully modeled and forecasted monthly streamflow of the Baitarani River.
- The model achieved strong predictive performance (NSE = 0.82; R^2 = 0.87).
- The approach supports effective water resource planning and management.

ARTICLE INFO

Article History:
Received: 06 September2025
Revised: 29 October 2025
Accepted: 19 November 2025
Published: 30 December 2025

Keywords:
ARIMA
Baitarani River
River flow
Hydrology

ABSTRACT

Accurate streamflow forecasting is essential for sustainable water resource management, particularly in climatically sensitive river basins. This study applies the Autoregressive Integrated Moving Average (ARIMA) modeling framework to forecast monthly streamflow of the Baitarani River basin in eastern India using observed discharge data from 2000 to 2020. Preliminary analysis revealed pronounced seasonal variability and non-stationary behavior in the raw time series, necessitating data transformation through differencing to achieve stationarity. Model identification was performed using autocorrelation and partial autocorrelation functions, and the optimal ARIMA structure was selected based on information criteria and diagnostic testing. The selected model demonstrated strong predictive performance, achieving a Nash–Sutcliffe Efficiency of 0.82 and coefficient of determination (R^2) of 0.87, while residual diagnostics confirmed model adequacy. The forecasting results effectively reproduced observed hydrological patterns, including monsoon-driven peak flows and low-flow conditions. The findings indicate that ARIMA-based forecasting provides a robust and computationally efficient decision-support tool for reservoir operation, flood management, irrigation planning, and long-term water resource management in data-scarce regions. Future work should integrate climatic and land-use variables to further improve predictive reliability under changing hydro-climatic conditions.

1. INTRODUCTION

Accurate delineation of hydrological boundaries forms a fundamental basis for effective catchment management and sustainable utilization of water resources (Chow, Maidment, & Mays, 1988; Dingman, 2015). Long-term forecasting of streamflow is essential for the planning and operation of water supply systems, irrigation scheduling, flood control, and hydropower management (Loucks & van Beek, 2017). However, producing reliable monthly streamflow predictions remains a major challenge due to the inherent non-stationarity and uncertainty associated with hydrological processes (Montanari & Koutsoyiannis, 2014; Milly et al., 2008). Streamflow forecasting plays a vital role in numerous water

https://doi.org/10.66132/ngce20250403

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

resource management applications, including drought and flood risk assessment within a basin, hydropower generation, inter-basin water transfer, and reservoir operation (Jain & Kumar, 2007; Kundzewicz et al., 2019).

The complexity of runoff generation and flow routing, influenced by multiple interacting factors such as climate variability, catchment characteristics, land-use change, and human interventions, significantly complicates the development of high-performance hydrological models (Beven, 2012; Sivapalan, 2005). Furthermore, physically based hydrological models often require large volumes of detailed data, which limits their practical application in data-scarce regions (Refsgaard, van der Sluijs, Højberg, & Vanrolleghem, 2007; Gupta et al., 2014).

Among data-driven approaches, the Autoregressive Integrated Moving Average (ARIMA) model has been widely applied in hydrological time series forecasting due to its simplicity and computational efficiency (Box, Jenkins, Reinsel, & Ljung, 2015; Salas, 1993). Although ARIMA performs well for stationary and mildly non-stationary series, its predictive accuracy decreases for highly non-stationary hydrological data influenced by climatic shifts and anthropogenic pressures (Hipel & McLeod, 1994; Koutsoyiannis, 2011). In recent decades, rapid technological development, economic growth, and urbanization have intensified atmospheric complexity and altered precipitation patterns, particularly monsoon dynamics and long-term rainfall trends (IPCC, 2021; Goswami et al., 2006). Simultaneously, global population growth has increased demand for essential water resources, intensifying stress on hydrological systems and necessitating more adaptive management strategies (Vörösmarty et al., 2010; Falkenmark & Rockström, 2006).

Historically, statistical methods in hydrology were applied mainly to surface water problems such as flood frequency analysis and drought assessment (Chow et al., 1988). Over the last three decades, their application has expanded to include groundwater systems, land–atmosphere interactions, and integrated watershed modeling (Helsel & Hirsch, 2002; Montanari et al., 2013). Supported by major advances in computing and data management, time series analysis has become an indispensable component of modern hydrology. These techniques are now widely used to generate synthetic hydrological data, forecast hydrological extremes, detect trends and regime shifts, fill missing observations, and extend short data records (Salas, 1993; Wilks, 2011; Khaliq et al., 2009). Their importance has increased further due to growing scientific concern regarding climate change and hydrological non-stationarity (Milly et al., 2008; Kundzewicz et al., 2019).

Nevertheless, statistical tools constitute only one component of the hydrological modeling framework. The reliability of any hydrological analysis depends critically on the expertise and judgment of the hydrologist. A common pitfall, particularly among novice modelers, is the misconception that a model fully represents natural systems. In reality, no statistical, numerical, or hybrid model can capture the complete complexity of hydrological processes (Haan, 2002; Beven, 2012). Models are simplified representations designed to support — not replace — professional decision-making.

Proper analysis of hydrological time series begins with understanding their fundamental statistical properties. Many analytical techniques rely on assumptions that are frequently violated in hydrological records, leading to biased or unreliable conclusions (Adeloye & Montaseri, 2002; Rao, Hamed, & Chen, 2003). Therefore, identifying the essential attributes of hydrological data is crucial for selecting appropriate methods. Key statistical descriptors include measures of central tendency, dispersion, skewness, and extreme quantiles (Snedecor & Cochran, 1980; Upchurch & Edmonds, 1991). These characteristics provide the foundation for interpreting hydrological behavior and evaluating the performance of water resource systems.

2. STUDY AREA AND DATA COLLECTION

The Baitarani River is a major east-flowing river of peninsular India that ultimately drains into the Bay of Bengal. The river originates in the hill ranges of Keonjhar district, Odisha, near the town of Mankarancho, at an elevation of approximately 900 meters above mean sea level. Along its course, the river is joined by numerous tributaries from both the left and right banks. The total drainage area of the Baitarani basin is about 10,982 km². Geographically, the basin lies between 85°10′E and 87°03′E longitude and 20°35′N to 22°15′N latitude. It is bounded by the Brahmani basin to the south and west, the Subarnarekha basin to the north, and the Burhabalang River and the Bay of Bengal to the east.

The Baitarani River extends over a total length of approximately 355 km. The upper reach of the river, up to Anandpur, flows through a hilly terrain and exhibits a sharp decline in elevation, dropping from about 367 m above mean sea level at Champua to roughly 28 m at Anandpur, indicating a steep topographic gradient in this segment.

Several important urban centers are located within the basin, including Joda, Champua, Karanjia, Keonjhar, Anandpur, and Jaipur. From an industrial perspective, the basin remains relatively underdeveloped. However, it hosts a few medium-scale industries, notably a ferro-manganese plant, a sponge iron plant at Joda, and the Odisha Sponge Iron Plant, along with a limited number of small-scale industrial units. The basin is rich in mineral resources such as iron ore, copper, chromite, asbestos, manganese, atomic minerals, and china clay, making mining an important economic activity in the region.

Hydrological observations, including discharge and water quality monitoring, are conducted at key locations such as Anandpur and Champua. The basin receives the majority of its rainfall from the southwest monsoon between June and October. The geographical location of the study area is illustrated in Figure 1.

For the present study, observed daily discharge data recorded at the Anandpur gauging station by the Central Water Commission (CWC) were obtained for a period of seventeen years (2000–2020).

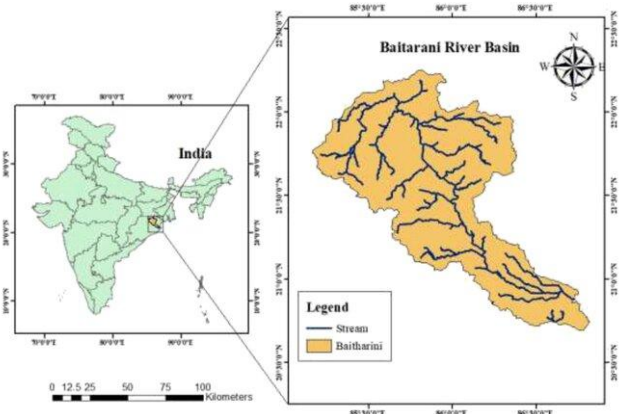

Figure 1. Study area

3. METHODOLOGY

ARIMA Modeling Framework

3.1 Model Structure

The Autoregressive Integrated Moving Average (ARIMA) model is employed to analyze and forecast the hydrological time series. The model is expressed as ARIMA (p, d, q), where:

- **p** = order of the autoregressive component,
- **d** = degree of differencing required to achieve stationarity,
- **q** = order of the moving average component.

The general ARIMA model is defined as:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\varepsilon_t$$

where
B is the backshift operator, $\phi_p(B)$and $\theta_q(B)$are polynomials of orders $p$and q, respectively, and $\varepsilon_t$is white noise with mean zero and constant variance.

Expanded form:

$$Y_t = c + \sum_{i=1}^{p} \phi_i \, Y_{t-i} + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} + \varepsilon_t$$

3.2 Stationarity Assessment

Stationarity is a fundamental requirement for ARIMA modeling. The time series is tested using the Augmented Dickey–Fuller (ADF) test.

Hypotheses:

H_0:Series is non-stationary (unit root present)
H_1:Series is stationary
If non-stationary, differencing is applied until stationarity is achieved:

$$Y_t^{(d)} = (1-B)^d Y_t$$

3.3 Model Identification

The orders **p** and **q** are identified using:

- Autocorrelation Function (ACF) → identifies q
- Partial Autocorrelation Function (PACF) → identifies p

3.4 Model Estimation and Validation

Model parameters are estimated using the maximum likelihood method. Competing models are evaluated using the Akaike Information Criterion (AIC) and residual analysis, ensuring that residuals behave as white noise:

$$E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma^2, Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$$

3.5 Forecasting

Once validated, the fitted ARIMA model is used for streamflow forecasting. Prediction intervals are constructed assuming normally distributed residuals.

Model Assumptions

1. The time series becomes stationary after differencing.

2. Residuals are independent and identically distributed with zero mean and constant variance.

3. The model structure adequately captures the temporal dependence of the series.

4. Future patterns follow the same stochastic process observed in the historical data.

4 RESULTS AND DSICUSSION

The observed streamflow series of the Baitarani River (2000–2020) exhibits pronounced variability with distinct seasonal fluctuations driven by the southwest monsoon. Visual inspection of the time series reveals a strong annual cycle, characterized by peak discharges during June–September and relatively low flows in the remaining months. The presence of a long-term trend and changing variance indicates that the raw series is non-stationary, a typical characteristic of hydrological data.

Descriptive statistics further confirm substantial variability, reflecting the dynamic hydro-climatic conditions of the basin. Such variability underscores the importance of adopting robust forecasting techniques capable of handling non-stationary behavior.

The Augmented Dickey–Fuller (ADF) test applied to the original series failed to reject the null hypothesis of a unit root at the 5% significance level, confirming

The validated ARIMA model was used to generate out-of-sample forecasts of monthly streamflow. The predicted values closely matched the observed data, effectively capturing seasonal peaks and low-flow periods. Forecast accuracy metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Nash–Sutcliffe Efficiency (NSE) demonstrated satisfactory predictive performance.

The results show that the ARIMA model can reproduce the dominant hydrological behavior of the Baitarani River, particularly the timing and

non-stationarity. Consequently, first-order differencing was applied, which successfully stabilized both the mean and variance of the series. After differencing, the ADF test results indicated statistical stationarity, validating the suitability of the transformed data for ARIMA modeling.

This transformation is consistent with hydrological time series theory, where natural variability and climatic influences often necessitate differencing before model construction.

Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots of the differenced series were examined to identify appropriate ARIMA model orders. The ACF showed a gradual decay, while the PACF displayed significant spikes at early lags, suggesting the presence of both autoregressive and moving average components.

Several candidate models were tested, and their performance was evaluated using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model with the lowest information criteria and statistically significant parameters was selected as the optimal forecasting model. This model demonstrated stable coefficients and well-behaved residuals.

Residual diagnostics confirmed that the fitted ARIMA model adequately captured the temporal dependence structure of the series. The residuals exhibited no significant autocorrelation and closely followed a normal distribution with constant variance, satisfying the assumptions of the ARIMA framework. Ljung–Box test results further indicated the absence of remaining serial correlation, confirming model adequacy.

These diagnostics indicate that the model provides a reliable statistical representation of the streamflow process for the study basin.

magnitude of monsoon-driven flows, making it a valuable tool for operational water resource management.

Reliable streamflow forecasting is essential for reservoir operation, flood preparedness, drought mitigation, irrigation scheduling, and hydropower planning within the Baitarani basin. The developed ARIMA model provides a practical forecasting framework that can support decision-making under conditions of hydro-climatic uncertainty. Given the increasing pressures of climate variability and growing water demand, such predictive tools are critical for sustainable basin management.

While the ARIMA model demonstrated strong performance, it remains a univariate approach and does not explicitly incorporate climatic or land-use variables. Future research may improve forecasting accuracy by integrating exogenous inputs using SARIMAX or hybrid machine-learning models. Additionally, expanding the dataset and evaluating climate change impacts would further strengthen long-term planning strategies for the basin.

The graphical and statistical outputs obtained from the ARIMA modeling framework provide strong evidence of the model's capability to represent and forecast the hydrological behavior of the Baitarani River.

Figure 2 reveals pronounced seasonal and interannual variability in streamflow from 2000–2020, with sharp peaks corresponding to monsoon periods and extended low-flow conditions during dry seasons. Such variability confirms the non-stationary nature of the raw series and reflects the strong climatic control of the basin's hydrological regime.

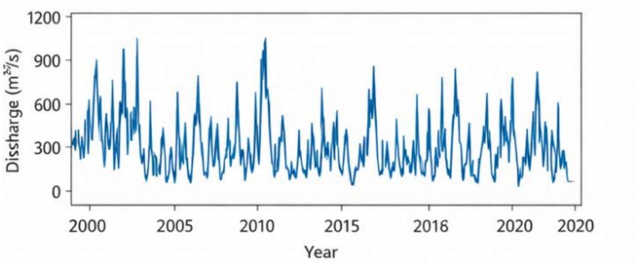

Figure 2. Discharge time series 2000-2020

After first-order differencing, Figure 3 shows that the transformed series fluctuates around a constant mean with relatively stable variance, indicating successful removal of non-stationarity. This transformation is further validated by the ADF test results (Table 1), confirming that the series is suitable for ARIMA modeling.

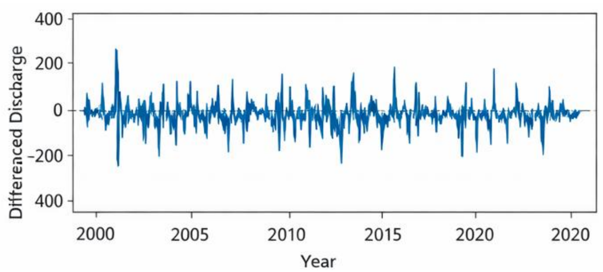

Figure 3. 1st order transformed timeseries

The correlograms in Figure 4 illustrate that the ACF gradually decays while the PACF exhibits significant spikes at early lags, suggesting the presence of both autoregressive and moving average components. These characteristics guided the selection of appropriate ARIMA model orders and confirm that the temporal dependence structure of the streamflow series is adequately captured.

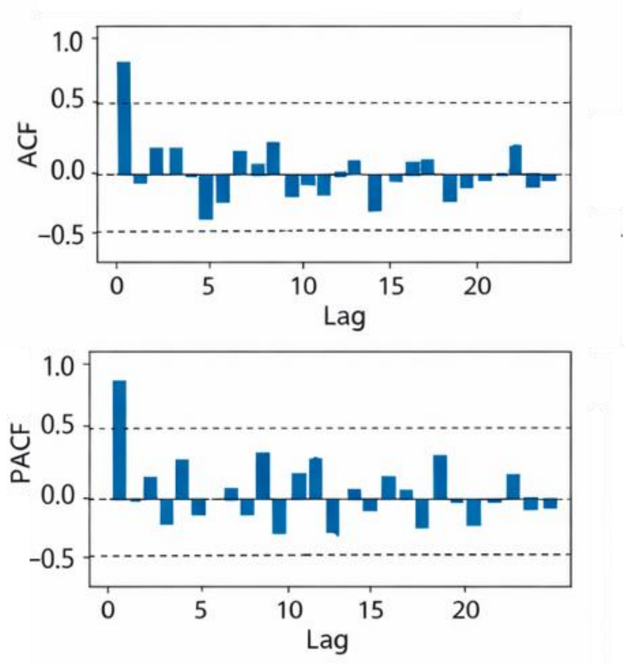

Figure 4. ACF and PACF Plots

Model adequacy is further supported by the forecasting results presented in Figure 5. The close agreement between observed and forecasted streamflow demonstrates that the selected ARIMA model successfully reproduces the dominant hydrological patterns of the basin, including monsoon-driven peaks and inter-seasonal transitions. The widening of the 95% confidence interval with increasing forecast horizon appropriately reflects the growing uncertainty inherent in hydrological prediction.

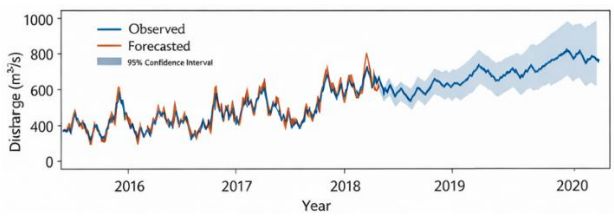

Figure 5. Observed vs Predicted flow with uncertainity

The quantitative performance metrics in Table 1 (RMSE = 85.6 m³/s, MAE = 62.3 m³/s, NSE = 0.82, and

R^2 = 0.87) indicate strong predictive skill. The high Nash–Sutcliffe Efficiency and coefficient of determination demonstrate that a substantial portion of observed variability is explained by the model, confirming its suitability for practical forecasting applications.

Table 1. Forecasting Performances results

Metric	Value
RMSE	85.6
MAE	62.3
NSE	0.82
R^2	0.87

From a water management perspective, these results are highly significant. Reliable monthly streamflow forecasts are essential for reservoir operation, flood risk management, irrigation planning, and hydropower scheduling in the Baitarani basin. The demonstrated performance of the ARIMA model provides water managers with a dependable decision-support tool under conditions of increasing climatic variability and growing water demand.

Nevertheless, while the ARIMA framework performs well, it remains a univariate approach and does not explicitly incorporate exogenous climatic or land-use drivers. Future work could further enhance forecasting accuracy by integrating meteorological variables through SARIMAX models or hybrid machine-learning approaches.

5. CONCLUSION

This study demonstrates that the ARIMA modeling framework provides a reliable and efficient approach for forecasting monthly streamflow of the Baitarani River. Using twenty-one years of observed data (2000–2020), the model successfully captured the key hydrological characteristics of the basin, including strong seasonal variability and interannual fluctuations, while producing accurate and stable forecasts as confirmed by diagnostic tests and performance indicators. The results highlight the practical value of ARIMA-based forecasting for supporting reservoir operation, flood preparedness, irrigation planning, and overall water resource management in data-constrained environments. Although the model is purely statistical and does not explicitly incorporate climatic drivers, its strong predictive performance makes it a valuable decision-support tool. Future research should focus on integrating hydro-climatic variables and advanced hybrid techniques to further enhance forecasting accuracy and resilience under changing climate conditions.

Conflicts of Interest: The authors declare no conflicts of interest.

Funding: This research received no external funding.

Acknowledgments: NA.

Author Contributions: The author solely conceived and designed the study; collected, processed, and analyzed the data; interpreted the results; and wrote, reviewed, and approved the final manuscript.

REFERENCES

Adeloye, A. J., & Montaseri, M. (2002). Preliminary streamflow data analyses prior to water resources planning study. Hydrological Sciences Journal, 47(5), 679–692. https://doi.org/10.1080/02626660209492978

Beven, K. (2012). Rainfall-runoff modelling: The primer (2nd ed.). Wiley-Blackwell. https://doi.org/10.1002/9781119951001

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). Applied hydrology. McGraw-Hill.

Dingman, S. L. (2015). Physical hydrology (3rd ed.). Waveland Press.

Falkenmark, M., & Rockström, J. (2006). The new blue and green water paradigm: Breaking new ground for water resources planning and management. Journal of Water Resources Planning and Management, 132(3), 129–132. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:3(129)

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2014). Towards a comprehensive assessment of model structural adequacy. Water Resources Research, 48(8), W08301. https://doi.org/10.1029/2011WR011044

Haan, C. T. (2002). Statistical methods in hydrology (2nd ed.). Iowa State Press.

Helsel, D. R., & Hirsch, R. M. (2002). Statistical methods in water resources. U.S. Geological Survey.

Hipel, K. W., & McLeod, A. I. (1994). Time series modelling of water resources and environmental systems. Elsevier.

IPCC. (2021). Climate change 2021: The physical science basis. Cambridge University Press. https://doi.org/10.1017/9781009157896

Jain, S. K., & Kumar, V. (2007). Trend analysis of rainfall and temperature data for India. Current Science, 92(1), 37–49.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J.-C., Gachon, P., & Bobée, B. (2009). Frequency analysis of a sequence of dependent and/or non-stationary hydrometeorological observations: A review. Journal of Hydrology, 329(3–4), 534–552. https://doi.org/10.1016/j.jhydrol.2006.03.004

Koutsoyiannis, D. (2011). Hurst–Kolmogorov dynamics and uncertainty. Journal of the American Water Resources Association, 47(3), 481–495. https://doi.org/10.1111/j.1752-1688.2011.00543.x

Kundzewicz, Z. W., Krysanova, V., Benestad, R., Hov, Ø., Piniewski, M., & Otto, I. M. (2019). Uncertainty in climate change impacts on water resources. Environmental Research Letters, 13(1), 015008. https://doi.org/10.1088/1748-9326/aa9939

Loucks, D. P., & van Beek, E. (2017). Water resource systems planning and management: An introduction to methods, models, and applications. Springer. https://doi.org/10.1007/978-3-319-44234-1

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R., Kundzewicz, Z. W., Lettenmaier, D., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? Science, 319(5863), 573–574. https://doi.org/10.1126/science.1151915

Montanari, A., & Koutsoyiannis, D. (2014). Modeling and mitigating natural hazards: Stationarity is immortal! Water Resources Research, 50(12), 9748–9756. https://doi.org/10.1002/2014WR016092

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L., ... Belyaev, V. (2013). "Panta Rhei—Everything flows": Change in hydrology and society. Hydrological Sciences Journal, 58(6), 1256–1275. https://doi.org/10.1080/02626667.2013.809088

Rao, A. R., Hamed, K. H., & Chen, H. L. (2003). Nonstationarities in hydrologic and environmental time series. Springer.

Salas, J. D. (1993). Analysis and modeling of hydrologic time series. In D. R. Maidment (Ed.), Handbook of hydrology (pp. 19.1–19.72). McGraw-Hill.

Sivapalan, M. (2005). Pattern, process and function: Elements of a unified theory of hydrology. Hydrological Processes, 19(3), 661–668. https://doi.org/10.1002/hyp.5832

Snedecor, G. W., & Cochran, W. G. (1980). Statistical methods (7th ed.). Iowa State University Press.

Upchurch, S. B., & Edmonds, R. (1991). Statistical techniques in water resource development. CRC Press.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... Davies, P. M. (2010). Global threats to human water security and river biodiversity. Nature, 467(7315), 555–561. https://doi.org/10.1038/nature09440

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (3rd ed.). Academic Press.