Results in Engineering 22 (2024) 102044

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

A data-driven approach to river discharge forecasting in the Himalayan region: Insights from Aglar and Paligaad rivers

Vikram Kumar [a], Selim Unal [b], Suraj Kumar Bhagat [c], Tiyasha Tiyasha [d,*]

[a] Bihar Mausam Sewa Kendra, Planning and Development Department, Government of Bihar, India
[b] Department of Metallurgical and Materials Engineering, Faculty of Engineering and Architecture, Kastamonu University, Kastamonu 37150, Turkey
[c] Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[d] Department of Civil Engineering, School of Engineering and Computing, Dev Bhoomi Uttarakhand University, Dehradun, India

ARTICLE INFO

Keywords:
Time series analysis
Holt-winters (HW) additive
Non-seasonal ARIMA
Paligaad
Simple exponential smoothing (SES)
River discharge
Forecasting

ABSTRACT

This study aims to better understand the time series forecasting of Aglar and Paligaad rivers' discharge (which has a significant impact on the Himalayan river) using advanced time series methods such as Holt-Winters (HW) additive method, Simple exponential smoothing (SES), and Non-seasonal auto-regressive integrated moving average (ARIMA) models. This study used antecedent discharge information to forecast the next event. Comprehensive statistical examinations were conducted and analyzed. The highly stochastic nature of these river discharge trends adds complexity to the forecasting efforts and requires sophisticated modeling techniques that are capable of capturing and interpreting such variability accurately. The models proposed in the current study provide a reliable forecast for the next 15 months using 31 months of recorded river discharge data. The forecast analysis shows that both the HW and non-seasonal ARIMA model results indicate exponential decay for the end of 2016 and early 2017. The HW model shows the best performance in long-term forecasting, indicating a sharp increase in spring and a small increase in discharge during fall months. However, for short-term forecasting, the non-ARIMA model should show more promising results. The results show that the proposed methodologies substantially improve the forecast accuracy of discharge for all consecutive months in perennial rivers. While the study presents promising results for forecasting the Aglar and Paligaad rivers' discharge, generalizing these findings to other river systems or different geographical regions may be problematic due to varying hydrological characteristics and environmental conditions, which may need further study.

1. Introduction

River discharge which resembles flow rate or streamflow which is the volumetric flux through river per unit of time [1]. This includes all hydrologic processes of upstream watersheds, non-linear behavior of the river's carrying capacity, and measurements that are needed to understand a river system (Kumar and Sen, 2023). Variation in river discharge depends on various factors and a steady process to investigate this [2]. Those factors are rainfall and human interference; in case of rainfall such as intensity & and duration of the storm [3], antecedent rainfall events (Kumar and Sen, 2017), geological (porosity & permeability) parameters [4], and size and shape of drainage basin [5] whereas in case of both rainfall and human factors, temperature change can be considered [6]. Moreover, human factors such as deforestation [7], and urbanization [8] can be included. Those factors are linked with the discharge such as intensity & and duration, antecedent rainfall events, and are directly proportional to the river discharge [9]. Geological properties vary with the discharge, in case of high porosity & and permeability, then discharge will be less and vice-versa [10,11]. The size of the drainage basin varies with the discharge, such as if shorter lag time then it accelerates the discharge [5]. In the case of the basin's shape, a steepy shape decreases the time for infiltration and therefore, accelerates the discharge and vice-versa [5,12]. Temperature is directly proportional to evapotranspiration and inversional to river discharge [13]. Deforestation supports more rainfall and simultaneously increases the discharge [7]. Urbanization spread tons of concrete led to more runoff and accelerated the discharge [8,14]. Forgoing factors led to the non-linearity in the river discharge and raised the chance of biasedness

* Corresponding author.
E-mail addresses: 25.vikram@gmail.com (V. Kumar), sunal@kastamonu.edu.tr (S. Unal), surajenv@gmail.com (S.K. Bhagat), tiyadas51@gmail.com, ce.tiyasha@dbuu.ac.in (T. Tiyasha).

https://doi.org/10.1016/j.rineng.2024.102044
Received 25 November 2023; Received in revised form 29 February 2024; Accepted 15 March 2024
Available online 26 March 2024

of its forecasting.

Aglar and Paaligad river watersheds are affected by both rainfall and interference of human beings on its discharge [15]. In the case of Aglar watershed, infiltration-excess (IE) dominated landscapes have been reported recently [16]. This watershed remains varied in terms of the grassed (GA) and agro-forested (AgF) hillslopes at the south and north; respectively. Moreover, both elevations of GA and AgF hillslope are reported as 1267 m and 1142 m, which add the influencing factors for the discharge. Moreover, the hydraulic variations appear significantly in the Aglar basin [17]. Nonetheless, the Paligaad watershed remains shorter in total stream length than the Aglar watershed [18]. Also, the research reported that the Paligaad has lower permeability which led to a higher runoff rate against the Aglar watershed. However, the Paligaad remains less vulnerable to erosion. To successfully manage water resources and reduce the hazards associated with floods, droughts, and changing rainfall patterns, it is also vital to determine how resilient the watershed is to climate change and to design adaptation strategies. The watershed is crucial for the Tehri Garhwal District's water supply, irrigation, hydropower production, biodiversity preservation, sustainable water management, and climate change resistance. In summary, understanding the rainfall and surface runoff of the Lesser Himalayas is essential for effective water resources management, flood mitigation, groundwater sustainability, ecosystem conservation, climate change adaptation, and infrastructure development. It provides the foundation for informed decision-making and sustainable development in the region.

A review of literature showed that the several methods for forecasting the discharge have been tried with their advancement as well as their limitation. Research proposed RBF kernel-based SVM model against persistence model and feedforward neural network (FFNN) model for forecasting discharge of Mahanadi River, India [19]. All three applied models fail for longer periods of forecasting, nonetheless, the SVM remains more accurate among them. In a recent study, researchers proposed the method i.e. linear regression-reduced error pruning tree (LR-REPTree) forecasting discharge of non-perennial river, that outperforms the contemperory GRG, LR, LR-RSS, LR-SVM, and LR-M5P models [20]. However, the research mentioned that ignoring biases coupled with overestimations and underestimations when investigating stochastic data to utilize the best investigating capacity of those models. Another research tried a wide range of input variables to forecast reservoir inflow using ANN architecture including static feed-forward neural network (FFNN), non-linear autoregressive (NAR), and nonlinear autoregressive with exogenous inputs (NARX) [21]. The research reported that NAR model showed the lower RMSE value with different input scenarios against both FFNN and NARX models. Another study applied a couple of weather (temperature and relative humidity) parameters to link up with rainfall that led to discharge of River Yazaram in Mubi town using adaptive neuro-fuzzy inference system (ANFIS) [22]. Nonetheless, short-term forecasting remains unsatisfying due to the data measured are not sufficient for the applied model investigations. Forgoing challenges in forecasting the river discharge with range of data such as short-term and long-term remains folded. Moreover, the model remains with reliable outcomes on applied in different locations. More number of input may bring another challenge of data collection that needs more effort, capital, and resources which can be avoided to minimize the time of the process. Reduce the complication of the calculation and computational time may open the door for new advanced ML algorithms for the potential sites that are vulnerable to the locality in terms of water management for agriculture, water supply, groundwater research, and more.

However, our understanding of both rivers (Aglar, and Paligaad) discharge remains limited due to a lack of trusted, comprehensive, and publicly available data. Both gauge stations remain folded in terms of discharge forecasting which plays a vital role in that area to the disposal of its tributary river. In contemporary river discharge patterns, the manifestation of antecedent flow signatures is discernible. The river's dynamics are not solely dictated by the current day's rainfall-induced discharge but are significantly influenced by the prior day's flow characteristics. Hence, the incorporation of antecedent discharge data becomes imperative to comprehensively capture the intricacies of the river's behavior [23]. This recognition underscores the necessity for considering historical discharge information in the modeling and forecasting processes, as the river's response is inherently intertwined with both immediate and antecedent flow dynamics.

The inclusion of antecedent discharge information significantly enhances the efficacy of river discharge forecasting for both Aglar and Paligaad rivers, utilizing advanced time series methods such as Holt-Winters (HW), Simple Exponential Smoothing (SES), and Non-seasonal ARIMA models. In this study, antecedent discharge data was incorporated to forecast the subsequent discharge events, particularly addressing the exponential decay challenges posed by the stochastic nature of discharge trends and various influencing factors in watershed management. The technical response emphasizes the adept utilization of antecedent discharge information, demonstrating the applicability and efficacy of the chosen advanced time series models in enhancing river discharge forecasting for complex and dynamic river systems in the lesser Himalayan in data-scare region. Also, the application of HW opens the door for producing adjustments to the observed data of wind speed forecasting [24]. The model outperforms the traditional ANN and ARIMA models in comparing the hybrid models with HW. In addition, HW has the filtering power that marginally approximates trends generated by the one-sided database that shows an elegant, moving average representation characteristics of HW [25]. However, this algorithm thirsts for a long period of database to show its best characteristics reported by the research. Moreover, the HW advised to be tested on real datasets with training models that gradually grow data sets [26]. In the case of the SARIMA model capturing information in a time series database is recommended for the different databases [27]. Another study showed that Holt-Winter's Exponential Smoothing (HWES) outperforms the ARIMA and SARIMA forecasting container throughput of major Asian ports; however, the research recommended exploring large volume datasets with high frequency that can be daily or weekly interval data to unfold the model capacity in terms of forecasting [28].

This is worth exploring the research motivation from the Scopus index literature database as keywords of 'discharge', 'forecasting', and 'Holt-Winters' using the VOSviewer algorithm. Bibliometric maps can be obtained simply and analyzed using the algorithm. Search data from the indexes are visualized. A total of 5032 documents have been filtered from 1969 to 2022, and that contain 30611 relevant keywords. fig. S1a demonstrates that India is in 4th position after the US, China, and UK for the specific research domain. Also, fig. S1b shows the research significance among the domain researchers that conducted and achieved exponential growth with 96% of R^2 values in recent years. Moreover, for the co-occurrence of the keywords categories into two groups – first (see fig. S1c) with forecasting and river discharge and second (see fig. S1d) with forecasting and applied machine learning methods. The method set 26 threshold values (value taken for clear visual Identification) for the co-occurrence keywords where 50 items touched the threshold values with were divided into two four clustering by applying the association method over the selected period which is presented by four colors (Green, Blue, cyan and Yellow). The relevant popular keywords and their co-occurrence can be seen in figs. S1c and d. There are two distinct types of keyword analysis used in fig. S1c which pertains to prediction and release, while fig. S1d relates to the application of machine learning models to discharge and forecasting. The timeline differs between figure S1c and figure S1d because significant changes were observed until 2016 for fig. S1c, whereas fig. S1d includes data until 2020 to reflect recent changes in values. In fig. S1c, there are 4 clusters observed, and the keyword "Forecasting" appears commonly in research papers related to topics like Flood forecasting, River, flood, etc. On the other hand, fig. S1d also displays 4 clusters, with "Forecasting" as the predominant keyword frequently found in research papers discussing concepts like

Holt-Winters, Time-series forecasting, forecasting methods, exponential smoothing, etc.

The objectives of this research paper are as follows: Firstly, the study aims to utilize the observed discharge database of the Aglar and Paligaad watersheds for further analysis and study. This database will provide the foundation for conducting research and investigating various aspects related to water discharge in these watersheds; secondly, the goal is to perform a descriptive analysis of the discharge database and present the findings which involves examining the patterns and characteristics of the discharge data, such as average values, variability, and trends; additionally, advanced time series forecasting methods such as Holt-Winters' additive method, Simple exponential smoothing, and Non-seasonal ARIMA models will be applied to the discharge data. These forecasting methods will allow for the prediction of future discharge values based on historical data. The performance of these forecasting methods will be evaluated and compared using statistical evaluators; Moreover, the results obtained from different models will be compared to identify any significant differences. he major outcomes of this research will be compared against the findings of published work in similar domains. This comparison will help confirm the validity and significance of the research results and provide context for the current study.

Lastly, potential areas for future studies and research in this field will be suggested. These recommendations will help guide further investigations and contribute to the advancement of knowledge in water discharge forecasting and analysis. This research aims to apply the untouched observed discharge database of the Aglar and Paligaad watersheds. Descriptive analysis of the database examined and presented. Also, the correlation with rainfall of the period has been logically presented. In addition, this study aims to apply advanced time series forecasting methods i.e. Holt-Winters' (HW) additive method, Simple exponential smoothing (SES), and Non-seasonal ARIMA models. These methods have been evaluated using the latest statistical evaluators, and compare these results within models as well as between the models. In addition, this research compares the major outcomes against published work. Moreover, this research compares those results between the stations and lists out the possible real meaning and the possible reasons for those peak values. Also, it lists the benefits of those applied methods, weaknesses, and future studies.

This section follows with materials and methods with the detail of data collection and its statistical analysis, and all three applied time series model with their mathematical expression of basic concepts along with data applied suitability. The next section described the modeling development and forecasting metrics including the software and package information and the applied model optimization reporting. This proceeded with application results and their analysis in a comprehensive way. An individual section dedicated to discussion of those major outcomes from applied time series models in line within the model and the station as well as between the models and between the stations. Also, the previously published research has been compared logically to set the edge of the art of technology. In the conclusion section all major outcomes, their advantages, and weaknesses have been reported

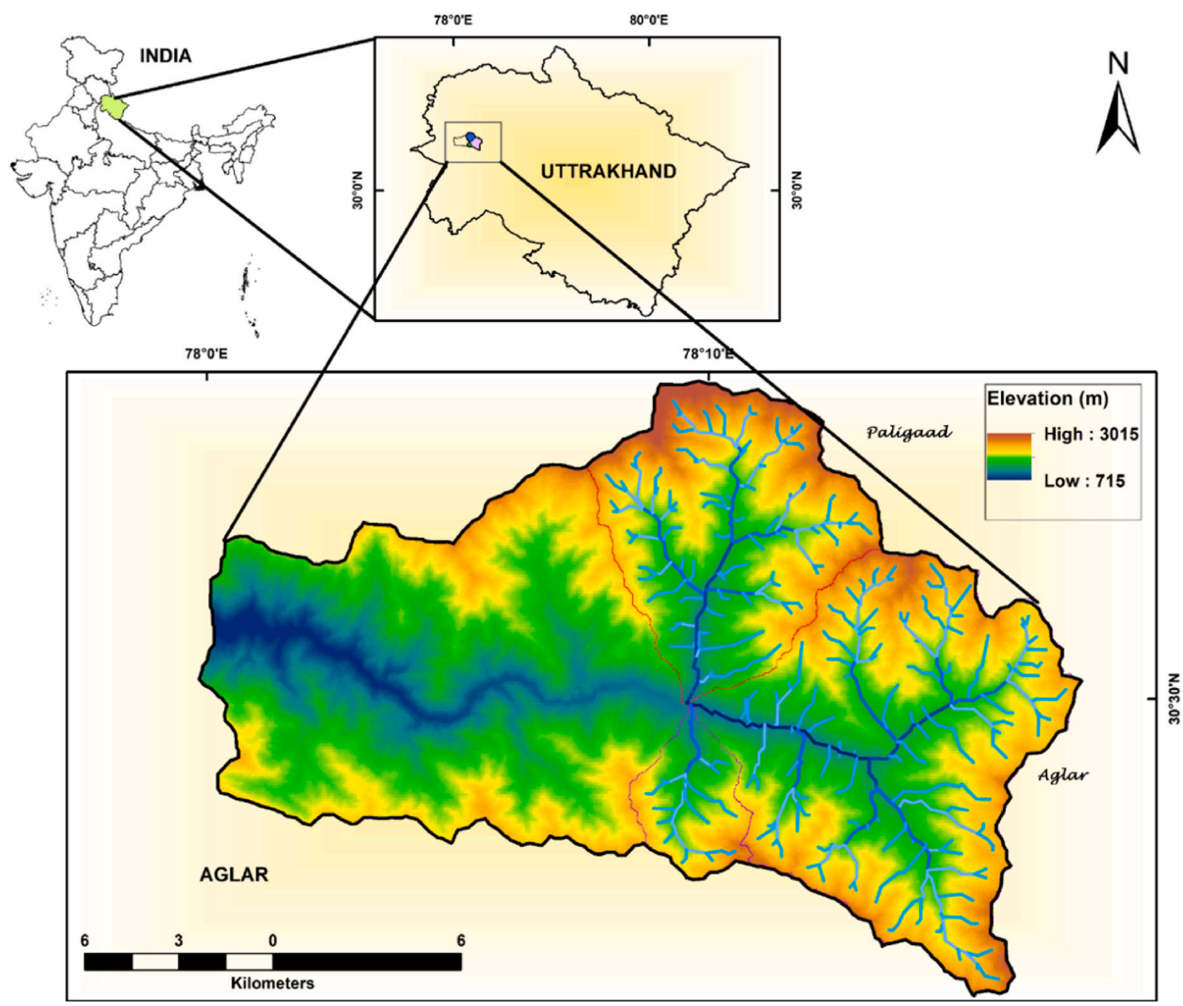

Fig. 1. Location map of the study area concerning Uttrakhand State and India. The inset blue layer represents the lowest elevation which is one of the greatest geological drainage of the area. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

scientifically.

2. Materials and MethodsMethodology

2.1. Study area and data collection and statistical analysis description

There are two river names viz. Aglar and Paligaad have been selected for the study (see Fig. 1). Fig. 1 exhibits the high and low elevations of the study catchment area. Both rivers are the major tributaries of the river Yamuna, which is part of the Himalayan watershed. It flows through the Tehri Garhwal district of Uttarakhand, known for its lush forests, meadows, and snow-capped mountains. The river is fed by rainfall, and glacial meltwater, which contribute to its flow throughout the year. In terms of water resources, the watershed is crucial as it supports the region's water supply demands by acting as a vital source of water for domestic and agricultural uses [29,30]. Stream discharge from the watershed must be accessible for agriculture to be supported, crop yields to be increased, and food security to be maintained. The watershed contains a wealth of variety and various ecosystems, making its maintenance crucial for preserving the ecological balance and aiding in species conservation. Monitoring water availability, controlling water consumption, and implementing conservation measures are all essential for sustainable development (Kumar and Sen, 2018).

A detailed statistical examination has been performed for the collected data from both stations. Table 1 demonstrated the descriptive analysis including Mean, Standard Error, Median, Standard Deviation, Sample Variance, Kurtosis, Skewness, Range, Minimum, Maximum, Sum, Count, and Confidence Level (95.0%) for the discharge database for both stations. A more than double time variation in mean value can be seen between Aglar (lesser) and Paligaad (higher) stations. In terms of standard deviation the database of Paligaad is more than six times higher against Aglar station. In the case of kurtosis and skewness, the variations between the databases appear significantly.

Moreover, fig. 3a compares the database for both stations in terms of scatter plot and box plot side by side. As per the box plot in fig. S2a, the Paligaad station database is as highly stochastic as decent at Aglar station. There are three peak events in the case of both stations, however, at Paligaad station, the discharge is visibly higher. However, in Figure S2b and Figure S2c a line ad station, the discharge is visibly higher. However, in figure S2b and figure S2c) a line plot has been drawn to exhibit the pattern of the relationship between the discharge and rainfall for Aglar and Paligaad stations, respectively. These two figures include regression equations along with correlation coefficient values. At both stations, rainfall trends declined over the period, however, this trend remains opposite in the case of discharge, but not in a similar ratio because of the land use pattern irrespective of the same rainfall. Both Table 1 and Fig. 3 reveal the data stochasticity that needs advanced empirical investigation. The presence of numerous outliers and higher peak values in Paligaad River discharge data, compared to Aglar River, is attributed to Paligaad's unstable nature and limited data collection during high flows. This uncertainty leads to rapid fluctuations, and data scarcity during extreme events hinders accurate representation. These factors impact forecasting models, necessitating advanced hydrological models, data augmentation strategies, and rigorous calibration to address unique characteristics and improve accuracy in capturing extreme conditions.

Table 1
Descriptive analysis of collected discharge database at both stations i.e. Aglar and Paligaad.

Descriptive Analysis	Aglar	Paligaadsade
Mean	0.458948	0.965876
Standard Error	0.006847	0.043269
Median	0.458948	0.50238
Standard Deviation	0.207445	1.31099
Sample Variance	0.043034	1.718695
Kurtosis	3.99643	14.08072
Skewness	1.329531	3.106372
Range	1.634834	13.1304
Minimum	0.186406	0.114885
Maximum	1.82124	13.24528
Confidence Level (95.0%)	0.013437	0.084918

2.2. Applied models

2.2.1. Holt-Winters Additive Model

Holt-Winters model is one of the most used methods for time series analysis with seasonal and trend patterns [31]. Tackling the constant seasonal variation Holt-Winters Additive model remains the best choice for scientists of various science and technology fields [32]. The working strategy of the additive model is a multiplicative model apart from seasonality that is to be considered additive. Therefore, the sum of the baseline, trend, and seasonality components of each data element is designed known as forecasted value. This model applies a triple exponential smoothing technique for using three different components. α is level β is the trend gamma is the seasonal component which can be controlled using the forecast weight parameters [33].

The Holt Winters Additive method parameters can be evaluated as the level estimation as Eqn. (1)

$$L_t = \alpha(Y_t - s_{t-m}) + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{1}$$

the trend equation is as Eqn. (2)

$$b_t = \beta(Y_t - L_t) + (1-\beta)b_{t-1} \tag{2}$$

and the seasonal component is as Eqn. (3)

$$s_t = \gamma(Y_t - L_t) + (1-\gamma)s_{t-m} \tag{3}$$

where Y_t is the recorded value the river discharge, L_{t-1}, b_{t-1} are level and trend at time t-1, m is the length of seasonal cycle, s_{t-m} is the seasonal factor at time t-m. The forecast is given as Eqn. (4)

$$\widehat{Y}_{t+} = L_t + kb_t + s_{t+k-m} \tag{4}$$

where k is the step ahead where forecast is calculated and the smoothing parameters (, β, γ) take values between 0 and 1.

2.2.1.1. Simple exponential smoothing (SES).

Time series exponential smoothing methods have two types with trend and/or seasonal pattern and without them. Those predictions are the weighted averages of past observations with the weights decaying exponentially as the observations get older [34]. Moreover, the popular forecasting method is the SES model which can be without any trend or seasonality [35]. The SES methods are less complicated but robust methods of forecasting. This is also recommended for the small-size database for higher accuracy in short term forecasting [36]. They are broadly applied in science, engineering, and business for forecasting demand for their different challenges [37]. Similar to the HW model, we can use to calculate the forecasting by applying as Eqn. (5) [38,39].

$$\widehat{Y}_{t+1} = \alpha Y_t + (1-\alpha)Y_t \tag{5}$$

where $\widehat{Y}_{t+1}$ is the forecast at time t-1, Y_t is the recorded value of the river discharge and α is the smoothing factor which is restricted between 0 and 1.

2.2.1.2. Non-seasonal ARIMA model.

The ARIMA (Auto-Regressive Integrated Moving Average) model is widely used for time series forecasting [40–42]. In the case of the Non-seasonal ARIMA model, a combined impact of "autoregressive terms, moving average terms, and differencing operations" was measured. Non-seasonal ARIMA models have been performed well in a wide range of environmental science and engineering data including climate change mitigation [43]. Three key elements are used in 'ARIMA (p,d,q)' model where p is the number of autoregressive lags, d is the number of non-seasonal differences and q is the number of moving-average terms. An autoregressive model can be used to forecast a time series data at time t, this can be written as Eqn. (6) [44]:

$$x_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \epsilon_t \tag{6}$$

where x_t is the value of the time series variable at time t, c is a constant value, φ is the autocorrelation constant, p is a constant which determines the number of samples used in calculation and ϵ is an error or white noise term.

A Moving average model forecasts using one or multiple precious error values at time t. It can be written as Eqn. (7) [45].

$$x_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \tag{7}$$

where θ is the autocorrelation of error values, q is the number of observations.

For the stationary time series AR (p) and MA (q) models can be represented by combining two models as Eqn. (8) [46].

$$Y_t = \xi + \sum_{i=1}^{p} \varphi_i Y_{t-p} + \sum_{j=0}^{q} \theta_j \epsilon_{t-j} \tag{8}$$

where ξ is a constant and Y_t is the value of the series at time t.

To apply ARIMA model, the time series must be stationary. For non-stationary time series, the time series needs to be modified to stationary time series. A differencing (integration) term is used in this process. In this case, the difference between precious measurements is calculated as $x_t - x_{t-1}$. Order of differencing term is given by d and represented by Eqn. (9) [47]

$$\varphi_p(B) \nabla^d x_t = \theta(B) \epsilon_t + c \tag{9}$$

Where, B is the distance operator representing how many time series are shifted as $Bx_t = x_{t-1}$ [48].

This proposed research includes combining several stages (see fig. S3) from data collection at both selected rivers such as Aglar and Paligaad, data transformation to the time series suitablity for the further function acceptibility for forecasting calculations. All three models for forecasting crossed through optimization stages. All performance stanges have been passed using applied perforamcne evaluators, residual calculations, and through autocorrelation factor (ACF) and Ljung-Box calculation. The higher error event where used to recycle the process and the least one recorded is used for the further results reportation.

3. Modeling development and forecasting metrics

This research was established using continuous monthly discharge measurements to forecast discharge for both stations i.e., Aglar and Paligaad. It utilizes the constructed model for comparative study. Also, its performance has been calculated using several performance parameters as well as with the Ljung-Box test in R version 4.2.3 (2023-03-15) software.

3.1. Optimization of forecasting models

All three applied time series model have been performed using RStudio version 4.2.3 (2023-03-15) software. Also, the performance evaluators along with Ljung-Box test have been performed with the same software. Moreover, the Residual density plot and ACF plot have drawn using the software.

To make HW additive forecasts method using *HoltWinters* function from *forecast* and *TTR* R package [49,50]. In developing HW additive methods, parameters have been estimated as $\alpha = 0.99$ and β and γ were set as FALSE for simple smoothing. Only the level parameter was taken into account and the value of α is large i.e. close to 1 for both stations, which means more weight is on the recent discharge observations and less on the earlier discharge data [51]. All the weight were given on to recent data.

To make SES forecasts method using *ses* function from *forecast* and *TTR* R package [49,50]. Here, h set as 15 that means 15 number of months are to be forecasted. Confidence level band set for prediction intervals between 80 and 95. *Initial* argument set as optimal that optimized along with the smoothing parameters. α and β were set as FALSE for simple smoothing.

To make Non-seasonal ARIMA forecasts method using *auto.arima* function from *forecast* R package [49,52]. Set the seasonal argument as FALSE to calculate the simple approach. Algorithm architecture for both station database is follows ARIMA (1,0,1) model with non-zero mean, where 1 is the number of autoregressive terms, 0 is the number of nonseasonal differences needed for stationarity, and. 1 is the number of lagged forecast errors in the prediction equation. Moreover, in case of Aglar discharge the Non-seasonal ARIMA follows the autoregressive investigation as per Eqn. (6), Eqn. (10) was calculated:

$$x_t = c + 0.5932 Y_{t-1} + 0.8750 Y_{t-2} + \epsilon_t \tag{10}$$

Where $c = 0.5798 * (1\text{–}0.5932) = 0.236$ and ϵ_t is white noise with a standard deviation of $0.012 = \sqrt{0.0001439}$; and in case of Paligaad discharge the Non-seasonal ARIMA follows the autoregressive investigation as per Eqn. (6), Eqn. (11) was calculated:

$$x_t = c + 0.5152 Y_{t-1} + 0.5579 Y_{t-2} + \epsilon_t \tag{11}$$

Where $c = 0.7714 * (1\text{–}0.5152) = 0.374$ and ϵ_t is white noise with a standard deviation of $0.093 = \sqrt{0.008725}$.

3.2. The model performance evaluators

The forecasted models were examined using the performance evaluators which is mean error (ME), root mean square (RMSE), mean absolute error (MAE), mean percentage error (MPE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE) and autocorrelation factor at lag 1 (ACF1). $\widehat{Y}_t$ is the forecasted value at time t and Y_t is the actual value at time t. All the above evaluators have been calculated using *accuracy* function with round argument to demonstrate up to two decimals.

Mean Error (ME) was calculated as Eqn. (12):

$$ME = \frac{\sum_{t=1}^{n} \left(\widehat{Y}_t - Y_t \right)}{n} \tag{12}$$

Root Mean Square Error (RMSE) was calculated as Eqn. (13):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} \left(Y_t - \widehat{Y}_t \right)^2}{n}} \tag{13}$$

Mean Absolute Error (MAE) was calculated as Eqn. (14):

$$MAE = \frac{\sum_{t=1}^{n} \left| Y_t - \widehat{Y}_t \right|}{n} \tag{14}$$

Mean Percentage Error (MPE) was calculated as Eqn. (15):

$$MPE = \frac{1}{n} \sum_{t=1}^{n} \frac{\left(\widehat{Y}_t - Y_t \right)}{Y_t} \tag{15}$$

Mean Absolute Percentage Error (MAPE) was calculated as Eqn. (16):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \widehat{Y}_t\right|}{Y_t} \tag{16}$$

Mean Absolute Scaled Error (MASE) was calculated as Eqn. (17):

$$MASE = \frac{MAE}{Q} \tag{17}$$

and Autocorrelation of Residuals at Lag1 (ACF1) was calculated as Eqn. (18):

$$Q = \frac{1}{n-1}\sum_{j=2}^{n}|Y_t - Y_{t-1}|. \tag{18}$$

Linear correlation between two variables can be found by Pearson's correlation coefficient, which was calculated as Eqn. (19)

$$r = \frac{n\sum_{t=1}^{n} Y_t\widehat{Y}_t - \sum_{t=1}^{n} Y_t \sum_{t=1}^{n}\widehat{Y}_t}{\sqrt{\left(n\sum_{t=1}^{n} Y_t^2 - \left(\sum_{t=1}^{n} Y_t\right)^2\right)\left(n\sum_{t=1}^{n}\widehat{Y}_t^{\,2} - \left(\sum_{t=1}^{n}\widehat{Y}_t\right)^2\right)}}. \tag{19}$$

Autocorrelation uses the time series data and measures the correlation between a variable and its previous values using its lagged values. For two time series values, ACF1 was calculated as Eqn. (20)

$$ACF1 = \frac{(Y_1 - \overline{Y})(Y_0 - \overline{Y})}{(Y_1 - \overline{Y})^2} \tag{20}$$

where Y_1 is the data at time 1, Y_0 is the data at time 0, $\overline{Y}$ is the sample mean of the time series which was calculated as Eqn. (21)

$$\overline{Y} = \frac{\sum_{t=1}^{n}(Y_t)}{n}. \tag{21}$$

3.3. Ljung-Box test

The Ljung-Box test determines whether the time series is white noise. The Ljung-Box test statistics is calculated by Eqn. (22) [53,54].

$$Q_H = n(n+2)\sum_{h=1}^{H}(n-h)^{-1}\widehat{\rho}_{n,h}^{\,2} \tag{22}$$

where $\widehat{\rho}_{n,h}$ is the autocorrelation at lag h that was calculated as Eqn. (23),

$$\widehat{\rho}_{n,h} = \frac{\sum_{t=1}^{n-h}(Y_{t+h} - \overline{Y})(Y_t - \overline{Y})}{\sum_{t=1}^{n-h}(Y_t - \overline{Y})^2}, \tag{23}$$

h is the number of lags being tested. A p-value determines how a null hypothesis is likely and it measures the difference between data and forecast. if the p-value is less than 0.05, the data is not random and not white noise. If the p-value is less than 0.05, the data is related to the white noise.

The residual was calculated using *checkresiduals* function in the R software and followed by Ljung-Box test that was performed using Box. test function for residuals of all three applied models where the argument *lag* = 15 and *type* = 'Ljung-Box' were set.

Moreover, the histogram density plot for three residual of applied time series models was drawn using *plotForecastErrors* function in R software. In addition, ACF plots for the residual have drawn using *acf* function where the argument 'lag.max' was set as 20.

4. Results and analysis

Discharge forecasting is a highly extensive method for various watershed management engineering processes. The current research was dedicated to the implementation of exploring optimized and more robust soft computing forecasting models for discharge forecasting for two different watersheds with different land use and soil characteristics. This section covers the results and discussion of the developed forecasting models. For each set of forecasting performance, the best model i.e., Non-seasonal ARIMA method for long-term and short-term analysis; Holt-Winters Additive method was performed well for long term forecasting; and remains average performance by SES method. All three models for both station discharge database is discussed below in detail.

4.1. Holt-Winters' additive method

Holt-Winters model is a commonly used method for time series analysis. The Holt-Winters model was performed on the data to forecast river discharge. The model was applied for both rivers viz. Aglar and Paligaad. Then, we conducted a residual noise check was using conducted using the Box-Ljung test to ensure no white noise contribution in the to our data. The density of residual errors was found as a Gaussian distribution with a zero-mean resulting in ideal forecast errors. Autocorrelation Function (ACF) was plotted using the residual errors to check for any significant autocorrelation. Finally, the model was used for forecasting analysis.

Holt-Winters method was performed to estimate the model parameters using a smoothing factor component considering a stationary time series. In this model, we only included the alpha parameter was included and ignored trend and seasonal components. Alpha parameter, smoothing factor, and coefficient are found to be 0.61 for Aglar and 0.86 for Paligaad. Paligaad station has higher coefficient that means for the higher forecast value than Aglar station. The density of residual errors has a Gaussian shape around zero, see Fig. 2a and b. The p-value of both Box-Ljung tests and the Ljung-Box test is found to be much less than 0.05 (for Algar 5.458E-10, 1.146E-5, and Paligaad 0.0002869 and 0.0363). This indicates the white noise has no contribution to the Holt-Winters model and the model is suitable to use for the forecast.

Using autocorrelation (ACF) of the Holt-Winters residuals, we determined the significant residual autocorrelation was determined at lag 2, 3 11, 12, and 13 for Aglar and lag 2 and 12 for Paligaad, shown in Fig. 2c and d. The Ljung-Box statistics test was applied to the residuals to find whether residual white noise contributes. The p-value is 1.146E-5 and 0.0363 for Algar and Paligaad respectively. The results show the p-values are lower than 0.05.

4.2. Simple exponential smoothing

Simple exponential smoothing is one of the easiest models to apply to forecasting problems. For the longer term, simple exponential smoothing (SES) can be used to forecast the river discharge [55]. The error parameters, the density of residual errors, and autocorrelation factors were extracted before the model fitting. Like HW results, the MAPE of Paligaad is higher than Algar. Also, the band Lo 95 and Hi 95 remains higher than band Lo 80 and Hi 80 for both stations that shows the feasibility of the method applied. Other error parameters are low as they show a good fitting. However, the density of residuals for both Algar and Paligaad are Gaussian around zero with multiple shoulder peaks, shown in Fig. 3a and b. This shows a white noise contribution to the analysis. Autocorrelation of residuals has lags at 2, 4, and 5, shown in Fig. 3c. The autocorrelation factor of residuals for Paligaad does not show any lags, see Fig. 3d. The fitting of the SES model has no significant forecast for both rivers even though MAPE and MAE values are low, see Fig. 5c and d.

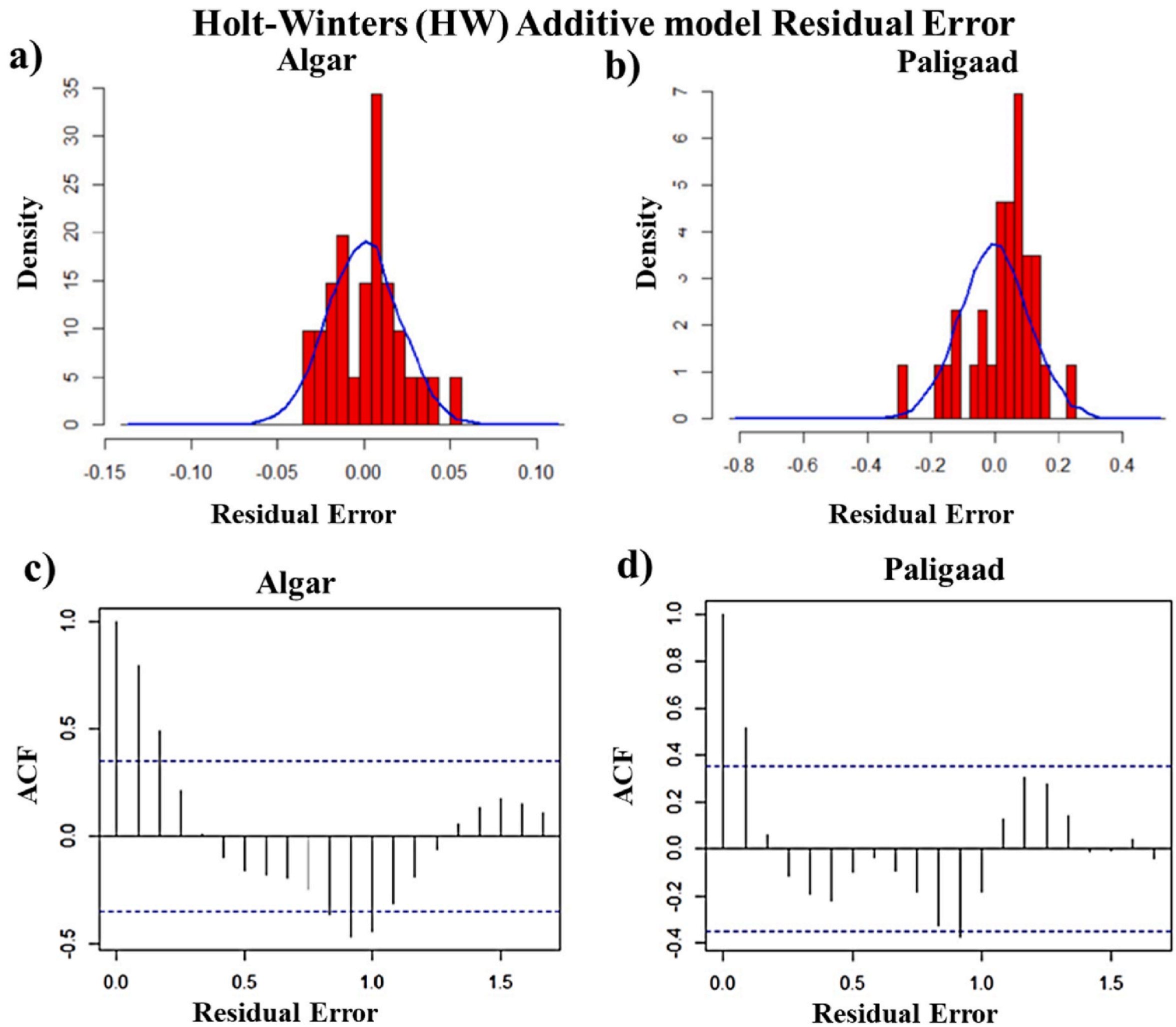

Fig. 2. a) Density plots for Holt-Winters (*plots for* HW) Additive model residuals for Algar*,* and b) Paligaad*; and* c) Autocorrelation (ACF) plots of the residual errors for Algar*,* and d) for Paligaad.

4.3. Non-seasonal ARIMA model

ARIMA model is not suitable for nonlinear patterns and relationships; therefore, non-ARIMA model can be used for river discharge forecasting problems [56]. For this model, ARIMA (1,0,1) model with non-zero mean was used to acquire the forecast. AR and MA coefficients were estimated along with other error parameters. As per Eqns. (10) and (11) *c* is not zero and d = 0 (the integrated part) that led the meaning of the long-term forecasts go to the mean of the data. In addition, the long-term forecast standard deviation will go to the standard deviation of the historical data since the d = 0. Moreover, the density of residuals for both Algar and Paligaad are shrinking near zero with multiple shoulder peaks, see Fig. 4a and b. This shows a white noise contribution to the analysis. Because of forecasting the discharge for both rivers, the fitting exponentially decays, see Fig. 5e and f. ACF of the data does not show any lags (see Fig. 4c and d). For both rivers, the residuals do not have any seasonal pattern and the ACF of the residuals do not show any lags apart from 1. If we check Box-Ljung and Ljung-Box tests, showed much higher the p-values are much higher than 0.05.

4.4. Comparison of the methods

The fitting precision indexes RMSE, MAE, and MAPE (see Table 2) of the HW, simple exponential smoothing, and ARIMA model were calculated, and the next month was forecasted. The performance of the models can be compared using the residuals, density of residual errors, ACF plots, Ljung-Box, and Box Ljung tests. The proposed forecasting methodologies introduce substantial advancements in enhancing the accuracy and reliability of discharge forecasts for consecutive months in perennial rivers, particularly from a hydrological perspective. The integration of antecedent discharge information plays a pivotal role by providing a temporal context that aids in discerning the influence of past conditions on current discharge patterns, thereby refining forecast accuracy. The methodologies are adept at robustly handling outliers and high peak values commonly found in perennial rivers, reinforcing the models' resilience to extreme values and enhancing overall reliability. Emphasizing long-term forecasting proficiency, particularly with the Holt-Winters model,. Proposed forecasting methodologies use mean absolute percentage error (MAPE) to define the overall accuracy. All MAPE values are less than 10 (except the HW of Paligaad), considered a threshold for accuracy. In the literature, the MAPE values have been reported to be much higher for discharge modeling and forecasting of the Paranaiba River [57]. The lowest MAPE value we have obtained is 1.51, showing a significant improvement in the accuracy.

For short-term forecasting, the non-ARIMA model could be useful. However, the HW model performs better than other models in the long-term forecast. Short-term prediction for the HW model shown in Fig. 5a and b and the non-seasonal ARIMA model shown in Fig. 5e and f perform similarly as they exponentially decay. However, for longer-term forecasts, SES and non-seasonal ARIMA models have constant values, which

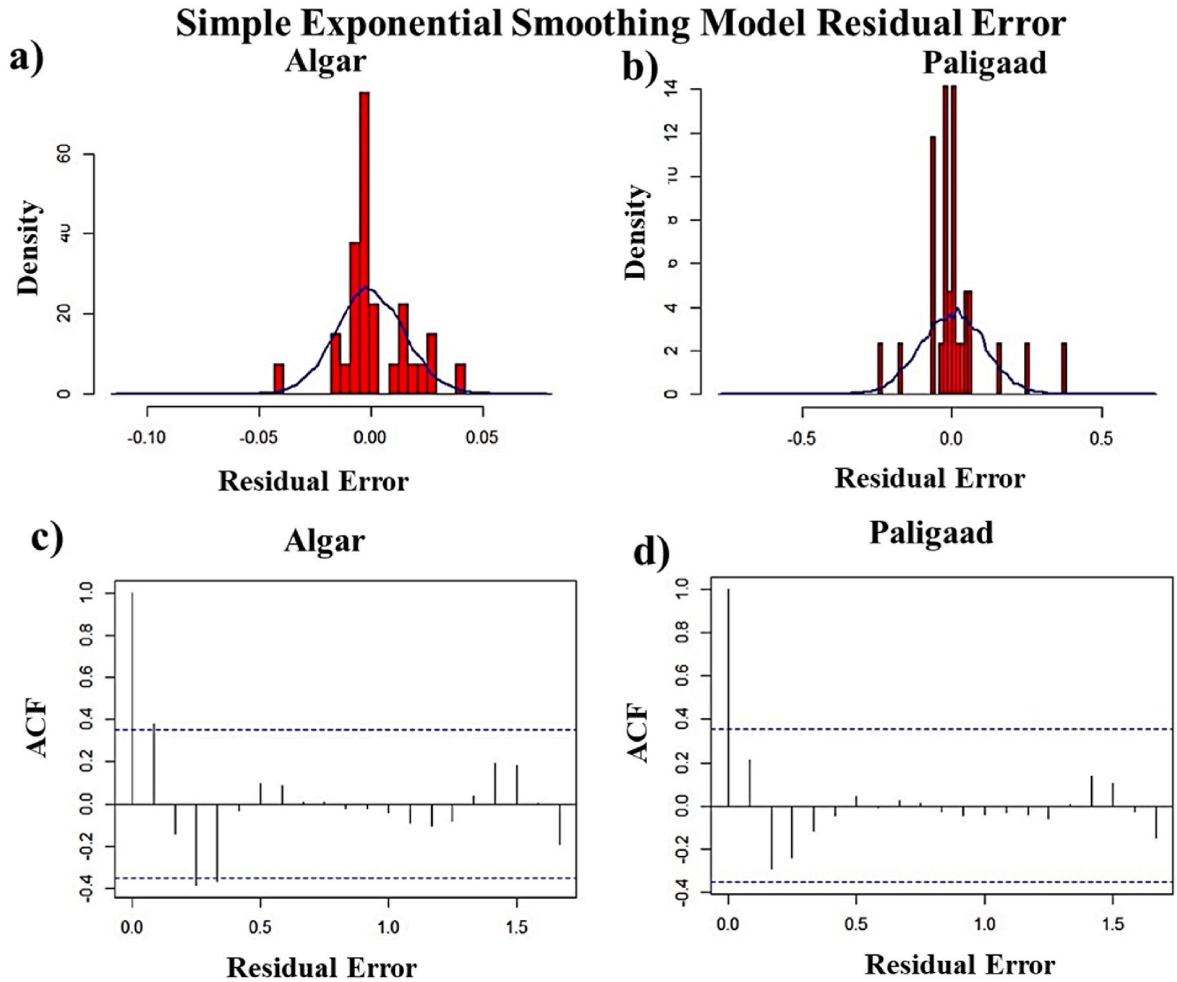

Fig. 3. a) Density plot of Simple Exponential Smoothing Model (SES) Model residuals for Algar, and b) Paligaad; and c) Autocorrelation (ACF) plots of the residual errors for Algar, and d) for Paligaad.

is not expected, see Fig. 5c, d, 5e, and 5f. The HW model performs well even after longer periods, as shown in Fig. 5a and b. The observed exponential decay in the forecast results for the end of 2016 and early 2017 can be attributed to distinct hydrological and climatic factors.

a) The decay may signify a response to the region's seasonal precipitation dynamics, where a reduction in river discharge is associated with diminished rainfall during the end of 2016 and early 2017. Seasonal variations in precipitation directly impact river flow, influencing the discharge patterns observed in the forecast results. Variations in groundwater levels exert a significant influence on river discharge during the post monsoon. The observed decay hints at fluctuations in groundwater contributions, with lower groundwater levels exerting influence over the broader discharge dynamics during the said duration.
b) Changes in vegetation dynamics also play a contributory role in the decay phenomenon. Colder months witness diminished transpiration from vegetation, potentially leading to an overall reduction in river discharge. This accentuates the significance of terrestrial ecosystems in molding the hydrological behavior observed in the forecast.

While it may not be feasible to quantify connectivity and influence of the runoff behavious across all spatiotemporal scales, certainly, emphasizing the utilization of available data on rainfall discharge for both rivers, more localized datasets become crucial. However, Integrating climatic variables like precipitation patterns, and temperature fluctuations, and through understanding of rainfall-runoff relationships within shorter time frames offers valuable insights. Short-term datasets, although limited, can be highly informative, especially when accurately capturing local climatic variations and their direct influence on discharge dynamics. Focusing on these variables allows for a more concentrated analysis, leveraging available data efficiently to enhance the accuracy of discharge forecasting models. Additionally, the inclusion of localized and short-term data helps develop more contextually relevant models tailored to the specific hydrological behaviours of the region, compensating for the absence of certain climatic variables like snowmelt and orographic effects. This approach enables a more precise understanding of the hydrological processes and contributes to improved forecasting accuracy, essential for effective water resource management in regions without snow dynamics.

MAPE, RMSE, and MAE values are calculated to analyze the accuracy and efficiency of the forecast, see Table 2 [58]. The highest MAPE was given by the HW model of the Paligaad and the lowest MAPE was the Non-Seasonal ARIMA model of the Algar. The lowest RMSE was in the non-seasonal ARIMA model of Algar and the highest RMSE was in the HW and SES model of the Paligaad, see Table 2. In a MAPE analysis, values less than 10 are called a highly accurate forecast [58,59]. In this case, all ofall the MAPE values are less than 10 except the HW of Paligaad. Even though, the MAPE values of the HW model are higher, both Ljung-Box tests of the HW model perform better than other models, see Table 3. This indicates that autocorrelation exists in the HW model, and the time series is not random.

Using the HW additive model, the minimum MAPE value of PM2.5 and NO2 forecast in Nagasaki has been found as 10.1772 and RMSE is

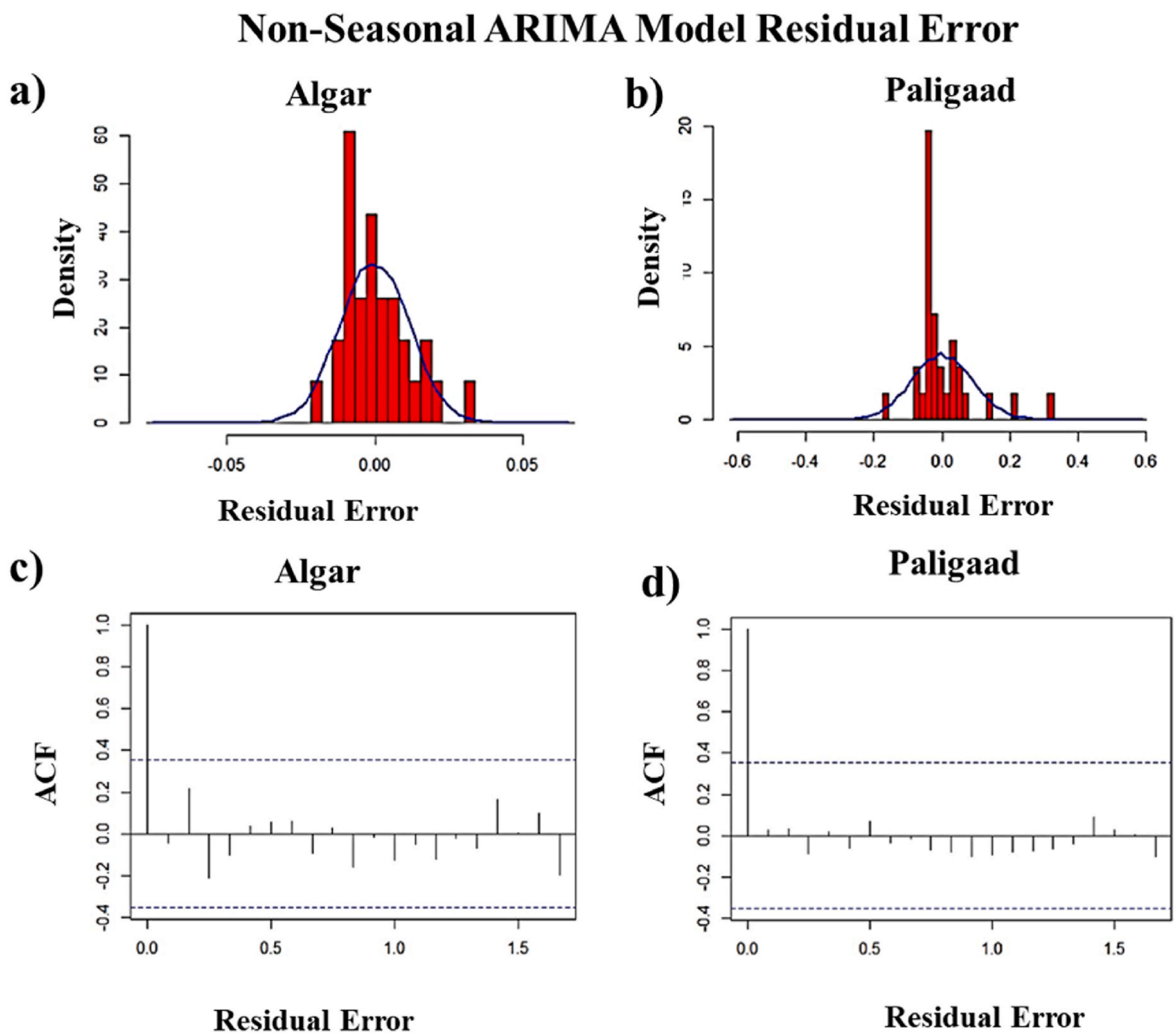

Fig. 4. a) Density plot of Non-Seasonal ARIMA Model residuals for Algar, and b) Paligaad; and c) Autocorrelation (ACF) plots of the residual errors for Algar, and d) for Paligaad.

0.4442 [60]. They have also found MAPE and RMSE values 23.9933 and 2.4873 respectively using ARIMA (1,0,1) with non-mean zero [60]. In a surface ozone layer study, the HW model shows MAPE, RMSE, and MAE values of 6.57, 0.85, and 0.65 respectively [58]. Our HW statistics are comparable with the literature with its MAPE and RMSE values. Non-seasonal ARIMA model performs also well and has much lower MAPE values than the literature. One of the drawbacks of this study is to forecast longer periods. The results show that the model could perform better in the short-term forecast see Fig. 5e and f. In the HW model Algar River discharge forecast in Fig. 5a, the forecast has a slight increase in 2017 and 2018, and it also shows seasonal increases and decreases in contrast to the non-seasonal ARIMA, see Fig. 5e. The ARIMA model had a larger value in 2017 than in 2018.

The comparative analysis between the Holt-Winters (HW) model and the Non-seasonal ARIMA model provides crucial insights into their respective suitability for long-term river discharge forecasting in the Himalayan region (See Fig. 5). The HW model demonstrates proficiency in capturing and forecasting trends influenced by distinct seasonal patterns, particularly relevant to the region's monsoonal variations. Its simultaneous consideration of trend and seasonality enhances its efficacy in scenarios where both components significantly contribute to discharge dynamics. On the other hand, the Non-seasonal ARIMA model exhibits adaptability to irregular discharge patterns and time-varying conditions, making it applicable in situations deviating from regular seasonal cycles. The performance comparison sheds light on long-term forecasting precision, robustness in the presence of anomalies, computational efficiency, and alignment with the unique characteristics of Himalayan River discharge data. Understanding these aspects is pivotal for selecting a forecasting model tailored to the complex hydrological dynamics of the Himalayan region.

During the figure analysis the result showed few discrepancies in discharge predictions of the Aglar and Paligaad rivers which can be explained by several factors.

a) Model performance: The accuracy and reliability of the forecasting models used can vary, which can lead to discrepancies in predictions.
b) Temporal context: The integration of antecedent discharge information, which considers past conditions, can significantly influence current discharge patterns, and improve forecast accuracy. Differences in the utilization of this temporal context in the models can result in discrepancies in the predictions.
c) Handling of extreme values: As discussed in previous section these perennial rivers exhibited outliers and high peak values, which can impact the accuracy of the forecasts. The ability of the models to handle these extreme values can vary, affecting the overall reliability of the predictions.
d) Short-term vs long-term forecasting: Different models may perform better in either short-term or long-term forecasts. The selection of the appropriate model for the specific forecasting horizon can lead to differences in the predictions.

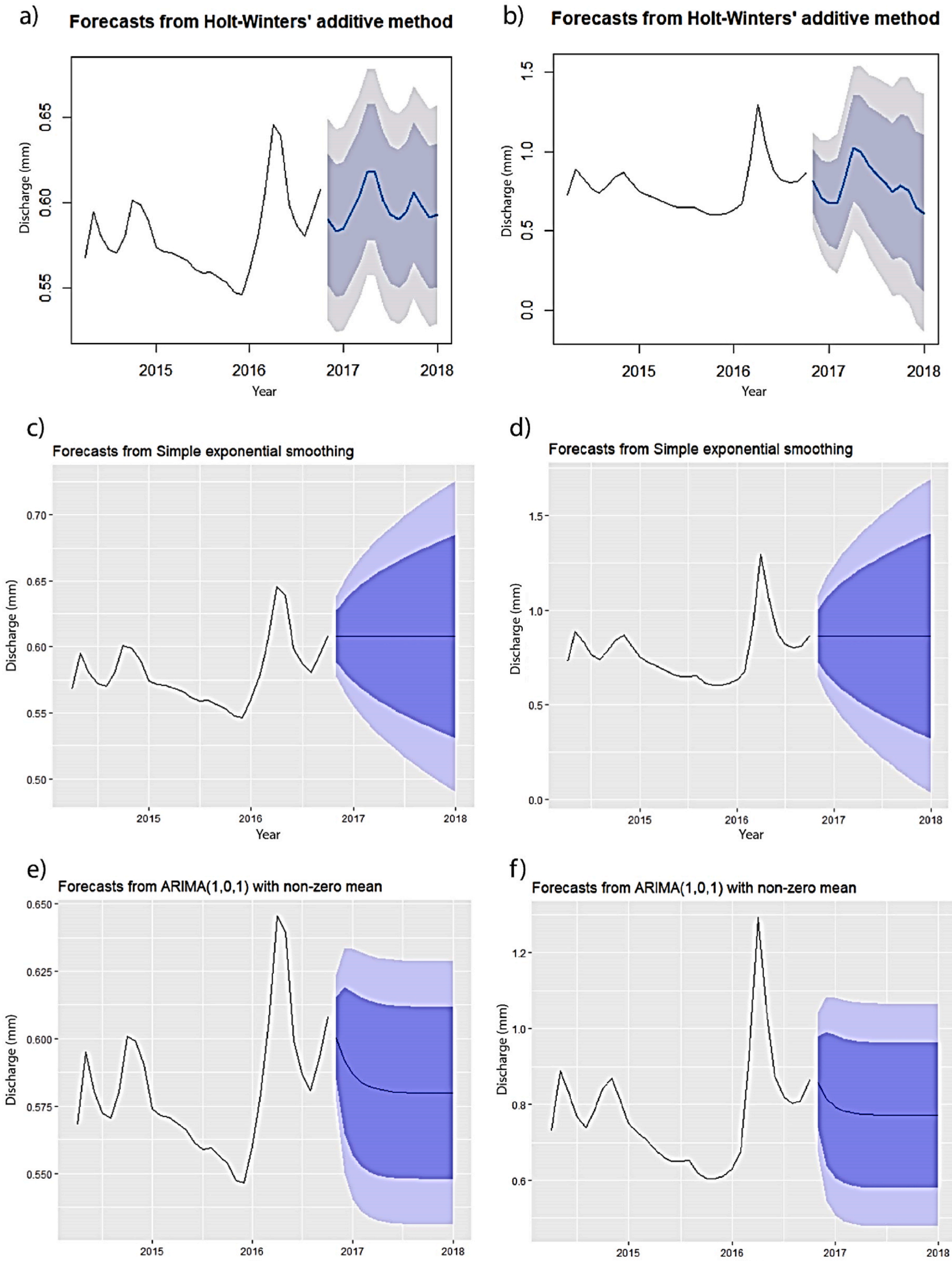

Fig. 5. Forecasting by Holt-Winters (HW) additive model for a) Aglar discharge and b) Paligaad discharge for next 15 months; forecasting by Simple Exponential Smoothing (SES) model for c) Aglar discharge and d) Paligaad discharge for next 15 months; forecasting by Non-seasonal ARIMA (1,0,1) model for cd) Aglar discharge and ed) Paligaad discharge for next 15 months.

e) Hydrological and climatic factors: The discharge of rivers is influenced by various hydrological and climatic factors, such as seasonal precipitation dynamics, groundwater levels, and vegetation dynamics. Variations in these factors can result in different discharge patterns, leading to discrepancies in the predictions.

f) Connectivity and runoff behavior: The connectivity and influence of runoff behavior across different spatiotemporal scales can also contribute to the discrepancies in discharge predictions.

Overall, the discrepancies in discharge predictions of the Aglar and

Table 2
Model fit statistics for river data analysis.

River	Model	Model Fit Statistics						
		Mean Error (ME)	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	Mean Percentage Error (MPE)	Mean Average Percentage Error (MAPE)	Mean Absolute Scaled Error (MASE)	Auto correlation of Errors at Lag 1 (ACF1)
Algar	Holt-Winters additive model	0	0.02	0.02	0.13	2.91	0.52	0.79
	Simple Exponential Smoothing	0	0.02	0.01	0.19	1.78	0.32	0.38
	Non-Seasonal ARIMA	0	0.01	0.01	0.05	1.51	0.27	−0.04
Paligaad	Holt-Winters additive model	0.02	0.11	0.09	2.67	11.39	0.43	0.52
	Simple Exponential Smoothing	0	0.11	0.06	−0.04	7	0.31	0.21
	Non-Seasonal ARIMA	0	0.09	0.06	−0.81	7.07	0.29	0.03

Table 3
Ljung-Box test results for both rivers.

River	Model	Box-Ljung Test	Ljung-Box Test
		p-value	p-value
Algar	Holt-Winters additive model	5.46E-10	1.15E-05
	Simple Exponential Smoothing	0.2305	0.0108
	Non-Seasonal ARIMA	0.9386	0.4162
Paligaad	Holt-Winters additive model	0.00029	0.0363
	Simple Exponential Smoothing	0.9291	0.2988
	Non-Seasonal ARIMA	0.9996	0.9548

Paligaad rivers can be attributed to a combination of model performance, temporal context, handling of extreme values, forecasting horizon, hydrological and climatic factors, and runoff behavior. Further research and assessment of these factors can help improve the accuracy and reliability of discharge forecasts for perennial rivers.

5. Discussion

The objective of this research paper is to utilize the observed discharge database of the Aglar and Paligaad watersheds in order to further analyze and study water discharge in these areas. By utilizing this database, the research aims to investigate various aspects related to water discharge and provide a foundation for further research based on historical data, advanced time series forecasting methods including Holt-Winters' additive method, Simple exponential smoothing, and Non-seasonal ARIMA models will be applied. The performance of these forecasting methods will be evaluated and compared using statistical evaluators. The results obtained from different models will be compared to identify any significant differences, and will be compared to findings from published work in similar domains to provide context and validate the research outcomes. Our statistical parameters can be compared with the previously published river discharge literature to demonstrate how well our results. Using LSTM Models, the river discharge forecast of the Humber River has the 3.98 % lowest error rate for a 12-h prediction [61]. For the Delaware River discharge forecast, the mean absolute percentage error and RMSE were found as 0.92 % and 0.037, respectively, using the same model [62]. However, 80 years of data have been used [63] to perform only a week's forecast, which is a much shorter range compared to our forecasted result. Slightly higher MAPE values have been reported for the same river using Convolutional Neural Networks (CNN) and Multilayer Perceptron (MLP), which are 2.17 % and 2.95 %, respectively [63]. For the short-term forecast of river discharge, the MAPE of River Yazaram is 4.64 % and a correlation coefficient of 0.13 [22]. In a study for river discharge of the Kashmir region of the Indian Himalayas-River Jhelum, the mean squared error (MSE) and RMSE have been shown as 0.2% and 6.61 by the Levenberg-Marquardt algorithm [64]. MAPE values reported in the forecast of the daily discharge time series of the Bow River result in values as low as 5.92 and 6.07 by ARIMA and Group Method of Data Handling (GMDH) Models, respectively [65]. In the results, the MAPE of the Algar River has the lowest error parameter, which reaches down to 1.51 with the non-seasonal ARIMA Model, and the RMSE is 0.01. This shows that the results are as good as the latest approaches, such as CNN and MLP models. In the preceding discussion, it is evident that indices are good measures in understanding complex watershed behaviours and drawing comparisons across contrasting landscapes. Indices such as these may serve as a general thumb rule for watershed managers to assess the status of target landscapes which in turn helps define watershed typologies and attach management protocols to each typology.

Understanding the rainfall-discharge responses is crucial for water resource management and flood control. Two watersheds, Paligaad and Aglar, have contrasting responses to rainfall events. Paligaad exhibits a flashy response, reacting rapidly to precipitation events with increased water flow. This is due to factors like steep terrain, limited soil infiltration, and lack of vegetation. On the other hand, Aglar shows a dampening response, responding smoothly and gradually to rainfall, attributed to gentle slopes, enhanced soil infiltration, and abundant vegetation. The flashy response of the Paligaad watershed indicates its sensitivity to variations in rainfall. Even small changes in rainfall intensity can lead to significant fluctuations in water flow, affecting water availability, flooding, and erosion. In contrast, the Aglar watershed dampening response makes it less sensitive to changes in rainfall. In Paligaad, frequent and severe flood events may necessitate measures like retention ponds, flood barriers, or artificial wetlands. In Aglar, the dampening response helps maintain a consistent water flow, benefiting water supply for various needs, agriculture, and ecosystem health.

The High Drainage Density of the Paligaad watershed (0.32) generally has shorter response times to rainfall events. Due to the well-connected network of streams and channels, the water can quickly drain from the catchment and contribute to the streamflow (fig. S4). As a result, these watersheds exhibit a more rapid increase in discharge following rainfall. Whereas the Agalr watershed having a low drainage density (0.21) has longer response times to rainfall. The limited number of streams and longer flow paths delay the movement of water through the watershed, resulting in a more delayed response in streamflow after rainfall events (fig. S4).

The different responses also influence land use decisions. Areas with flashy responses (Paligaad) might be suitable for recreational purposes, while those with dampening responses could be ideal for agriculture or water supply. Tailored conservation efforts can involve reforestation in Paligaad to reduce runoff and erosion and groundwater recharge initiatives in Aglar.

The differences in forecasting accuracy between the Aglar and Paligaad rivers' discharge can have significant impacts on the socio-economic and environmental aspects of the communities and ecosystems that rely on these rivers. Inaccurate forecasts can affect water-dependent industries like agriculture and fisheries, leading to suboptimal water resource utilization and impacting crop yields and livelihoods. Furthermore, fluctuations in river discharge can disrupt the habitat of aquatic species and lead to inadequate water availability for maintaining ecological balance, affecting biodiversity and overall ecosystem health. These discrepancies also influence water management decision-making, potentially leading to challenges in planning for floods or water scarcity and impacting disaster preparedness [66]. Additionally, poor predictions can impact the design and efficiency of water-related infrastructure along the riverbanks, while exacerbating social inequities and vulnerabilities within communities. To address these implications, there is a need for continual improvement in modeling approaches and the development of adaptive strategies, ultimately contributing to more resilient and sustainable water resource management [67].

The accuracy of the discharge forecast might be improved by advanced machine learning techniques such as deep learning. Dynamical predictions can be performed using real-time data, and flood risks can be forecasted (Latt and Wittenberg 2014). These models provide free open-source forecasts for both developed and developing countries. More models can be improved. Watershed management policies can be governed accordingly. Various models should be used to build more accurate models to identify how the models behave, and precautions should be taken using those models. Future studies could use a large set of time series data to perform a better forecast, resulting in a better understanding of drought periods. Flow monitoring and discharge forecasts for unusual activities in river discharge and skilled personnel to develop forecasting models are required in developed and developing nations (Costa et al., 2023).

An increase in the discharge would significantly impact watershed management strategies, water allocation, reservoir operations, and drought mitigation, ([68] Kumar et al., 2023; Stakhiva and Stewart 2010). Agricultural products can be chosen based on when the water is needed in the spring and fall months. Waters might be stored depending on the discharge changes and the forecast. Precautions can be taken for erosion, flooding risk, climate change, and groundwater research. Developing countries might use this model to improve low-cost strategies, monitoring stations, and warning systems. The model can be used to avoid flood risks and water supply shortages without any costly gauges (Latt and Wittenberg 2014).

6. Conclusion

The current study, we have explored the use of baseline data for discharge modeling as a potential tool for forecasting assessment, aimed at providing 15-month forecasting consecutively that is accessible and implementable by water managers seeking quick hydrological assessments. For this study time series forecasting of Paligaad and Aglar river discharges has been taken from 2014 to 2017. The residual components of non-stationary river discharge data were extracted and examined. Three widely used statistical forecasting models, which are HW additive, SES, and non-seasonal ARIMA models were performed to anticipate 15 months ahead of river discharges. The model parameters were investigated with statistical error parameters. The lowest MAPE was observed with the Non-seasonal ARIMA model showing an exponential drop. However, the HW model performs well in long-term forecasting representing rises and falls in the spring and fall months. Aglar station data showed a higher forecast coefficient value than Paligaad station data. This study shows statistical models can be used to forecast river discharges.

Finally, it is worth noting that other methodologies may be proposed to investigate total discharge behavior, and that the selection of the most appropriate models is linked to a huge variety of criteria. Further studies might benefit, for instance, from a meteorological parameters integration approach before proceeding to the methods here developed. For the watershed management sector, particularly, this means applying the time series methods for each influencing parameter of discharge would be ranked as per the importance, forecasting the simulated series, and then pooling together the results into one single output. Such sector-specific studies not only would provide a more in-depth understanding of the watershed system across different countries but could also possibly further enhance the accuracy of water management policies. The study of the application of the Shapley Additive Explanation (SHAP) method to explore its wide potential such as minimizing the complexity of the applied time series model using the metric of cumulative relative variance (CRV) levels is also a logical extension of this work.

Thus, this result is crucial for effective water management, flood control, and sustainable water resource use in both watersheds. Implementing appropriate strategies based on their specific responses can mitigate flood risks, ensure a stable water supply, and support ecological well-being.

The study focuses on the potential benefits of accurate discharge forecasts for watershed management practices, particularly in soil water conservation and food production. These forecasts help optimize water management in agriculture by anticipating water availability fluctuations and aiding in irrigation planning. They also support effective soil water conservation strategies and efficient allocation of water resources. By integrating predictions into crop planning, stakeholders can ensure sustainable agricultural productivity amidst climate variability. Ultimately, these forecasting methodologies aim to secure food production by adapting to changing environmental conditions and mitigating drought risks. The study suggests that increasing discharge could significantly impact watershed management strategies, water allocation, reservoir operations, and drought mitigation, offering valuable insights for developing countries to improve low-cost strategies and monitoring systems El-Shafie et al., 2007; Kumar et al., 2023; Stakhiva and Stewart 2010). Accurate discharge forecasts facilitate optimized water management in agriculture, enabling stakeholders to anticipate fluctuations in water availability. This foresight supports timely decision-making for strategic irrigation planning, promoting efficient water use in crop cultivation (Latt and Wittenberg 2014).

Overall, we wish to acknowledge that, the technical enhancements in discharge forecasting in such data scare region with the robust methodology not only bolster accuracy and reliability but also offer targeted benefits for watershed management practices. These benefits encompass optimized agricultural water management, enhanced soil water conservation, strategic allocation of water resources, integration of forecast insights into crop planning, adaptation to climate variability, and, ultimately, the securement of food production within the watershed.

This study opens a wider range of further scope of in-depth research such as.

a) The accuracy of the discharge forecast might be improved by advanced machine learning techniques such as deep learning. Dynamical predictions can be performed using real-time data, and flood risks can be forecasted (Latt and Wittenberg 2014). These models provide free open-source forecasts for both developed and developing countries. This will couple hydrological models with climate models, integrating socio-economic variables for a comprehensive understanding of discharge dynamics. Develop probabilistic models that account for uncertainty and risk, facilitating nuanced decision-making in watershed management
b) Future studies could use a large set of time series data to perform a better forecast, resulting in a better understanding of drought periods. Flow monitoring and discharge forecasts for unusual activities in river discharge and skilled personnel to develop forecasting models are required in developed and developing nations (Costa et al., 2023). The models are open-source and free to use. It can be used in both R and Python. Developing countries might use this model to improve low-cost strategies, monitoring stations, and warning systems. The model can be used to avoid flood risks and water supply shortages without any costly gauges [69].
c) Explore high-resolution remote sensing to enhance spatial precision in discharge validation, incorporating climate change impacts and other anthropogenic changes in the region. The findings from this study have implications for rainfall-runoff modeling in data-scarce regions, particularly for perennial river systems with limited rainfall and runoff data [70]. However, it is important to consider the unique characteristics of each river system, such as hydrological, geographical, and climatic attributes [71].
d) The research findings from this study can be used to optimize water resource management in developing countries with limited resources. By adapting the methodology to specific hydrological and climatic conditions, more accurate discharge forecasts can be made for target watersheds. This allows for strategic allocation of water resources, particularly for agricultural irrigation, and supports water-dependent sectors. The research also contributes to strategic planning in the face of climate variability and promotes resilient water resource management strategies [72].

Overall, this study contributes to advancing knowledge of hydrological processes and modeling in challenging data environments.

Limitations, Implications and Recommendation of the Study*:* Defining the forecasting parameters to fit the data ideally takes much work to overcome. In the HW model, seasonal and trend components should be defined. The length of the seasonal cycle should be appropriately chosen. Similar challenges have been faced when applying the non-seasonal ARIMA model. Non-stationary data is usually a challenge to analyze. It needs to be modified to a stationary time series. Time-series shift parameter in differencing term and several observations in the Moving average should be identified. Model parameters p, d, and q are determined to fit well for the forecast. Limited historical data and more than a year of forecasting are other challenges in this forecasting.

Ethical approval

The manuscript is conducted within the ethical manner advised by the targeted journal.

Consent to participate

Not applicable.

Consent to publish

The research is scientifically consent to be published.

Funding

The research received no funds.

CRediT authorship contribution statement

Vikram Kumar: Writing – review & editing, Writing – original draft, Investigation, Funding acquisition, Formal analysis, Data curation. **Selim Unal:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Suraj Kumar Bhagat:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Tiyasha Tiyasha:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors would like to acknowledge the data used in this manuscript which was collected by VK at the Department of Hydrology, Indian Institute of Technology Roorkee (IITR) during the PhD program. The data collection was financially supported by IITR and partially by Department of Science and Technology (DST), India. The authors also acknowledge the soft computing facility support by the Ton Duc Thang University, Viet nam.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rineng.2024.102044.

References

[1] D. Feng, C.J. Gleason, P. Lin, X. Yang, M. Pan, Y. Ishitsuka, Recent changes to Arctic river discharge, Nat. Commun. 12 (2021) 6917.
[2] W. Li, L. Zhu, G. Xie, P. Hu, H.J. de Vriend, Quantification of the influencing factors for flood peak discharge increase in the Lower Yellow River, J. Hydrol. (Amst.) 613 (2022) 128329.
[3] B.W. Kerns, S.S. Chen, Compound effects of rain, storm surge, and river discharge on coastal flooding during Hurricane Irene and Tropical Storm Lee (2011) in the Mid-Atlantic region: coupled atmosphere-wave-ocean model simulation and observations, Nat. Hazards 116 (2023) 693–726.
[4] K. Pareta, Hydrological modelling of largest braided river of India using MIKE Hydro River software with rainfall runoff, hydrodynamic and snowmelt modules, Journal of Water and Climate Change 14 (2023) 1314–1338.
[5] P.-A. Garambois, K. Larnier, J. Monnier, P. Finaud-Guyot, J. Verley, A.-S. Montazem, S. Calmant, Variational estimation of effective channel and ungauged

anabranching river discharge from multi-satellite water heights of different spatial sparsity, J. Hydrol. (Amst.) 581 (2020) 124409.
[6] W. Ren, L. Tian, L. Shao, Temperature and precipitation control the seasonal patterns of discharge and water isotopic signals of the Nyang River on the southeastern Tibetan Plateau, J. Hydrol. (Amst.) 617 (2023) 129064.
[7] J.A. Posada-Marín, J.F. Salazar, River flow response to deforestation: contrasting results from different models, Water Secur 15 (2022) 100115.
[8] A. Omer, X. Yuan, A. Gemitzi, Transboundary Nile basin dynamics: land use change, drivers, and hydrological impacts under socioeconomic pathways, Ecol. Indicat. 153 (2023) 110414.
[9] M. Jahandideh-Tehrani, H. Zhang, F. Helfer, Y. Yu, Review of climate change impacts on predicted river streamflow in tropical rivers, Environ. Monit. Assess. 191 (2019) 1–23.
[10] E. V Arntzen, D.R. Geist, P.E. Dresel, Effects of fluctuating river flow on groundwater/surface water mixing in the hyporheic zone of a regulated, large cobble bed river, River Res. Appl. 22 (2006) 937–946.
[11] J. Agnihotri, A. Behrangi, A. Tavakoly, M. Geheran, M.A. Farmani, G. Niu, Higher Frozen soil permeability represented in a hydrological model improves spring streamflow prediction from river basin to Continental scales, Water Resour. Res. 59 (2023) e2022WR033075.
[12] S. Soni, Assessment of morphometric characteristics of Chakrar watershed in Madhya Pradesh India using geospatial technique, Appl. Water Sci. 7 (2017) 2089–2102.
[13] M.T.H. van Vliet, W.H.P. Franssen, J.R. Yearsley, F. Ludwig, I. Haddeland, D. P. Lettenmaier, P. Kabat, Global river discharge and water temperature under climate change, Global Environ. Change 23 (2013) 450–464.
[14] S. Han, L. Slater, R.L. Wilby, D. Faulkner, Contribution of urbanisation to non-stationary river flow in the UK, J. Hydrol. (Amst.) 613 (2022) 128417.
[15] S. Sharma, P.K. Joshi, C. Fürst, Unravelling net primary productivity dynamics under urbanization and climate change in the western Himalaya, Ecol. Indicat. 144 (2022) 109508.
[16] A. Nanda, S. Sen, A complex network theory based approach to better understand the infiltration-excess runoff generation thresholds, J. Hydrol. (Amst.) 603 (2021) 127038.
[17] G. Jeelani, S.A. Lone, A. Lone, R.D. Deshpande, Groundwater resource protection and spring restoration in Upper Jhelum Basin (UJB), western Himalayas, Groundw Sustain Dev 15 (2021) 100685.
[18] V. Kumar, S. Sen, P. Chauhan, Geo-morphometric prioritization of Aglar micro watershed in Lesser Himalaya using GIS approach, Model Earth Syst Environ 7 (2021) 1269–1279.
[19] S.K. Aggarwal, A. Goel, V.P. Singh, Stage and discharge forecasting by SVM and ANN techniques, Water Resour. Manag. 26 (2012) 3705–3724.
[20] D.K. Vishwakarma, A. Kuriqi, S.A. Abed, G. Kishore, N. Al-Ansari, K. Pandey, P. Kumar, N.L. Kushwaha, A. Jewel, Forecasting of stage-discharge in a non-perennial river using machine learning with gamma test, Hel iyon 9 (2023).
[21] P.P. Hadiyan, R. Moeini, E. Ehsanzadeh, Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir, Sustainable Computing: Informatics and Systems 27 (2020) 100401.
[22] A. Danladi, M. Stephen, B.M. Aliyu, G.K. Gaya, N.W. Silikwa, Y. Machael, Assessing the influence of weather parameters on rainfall to forecast river discharge based on short-term, Alex. Eng. J. 57 (2018) 1157–1162.
[23] W. Ghani, S. Rasool Shah, B. Kumar, Higher-order turbulent characteristics of sinuous alluvial river, Results in Engineering 20 (2023) 101508, https://doi.org/10.1016/j.rineng.2023.101508.
[24] H. do Nascimento Camelo, P.S. Lucio, J.B.V.L. Junior, P.C.M. de Carvalho, D. von G. dos Santos, Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks, Energy 151 (2018) 347–357.
[25] R. Alfaro, M. Drehmann, The Holt–Winters filter and the one-sided HP filter: a close correspondence, Econ. Lett. 222 (2023) 110925.
[26] R. Rushton, O. Lorraine, J. Tiong, M. Karim, R. Dixon, W. Greenshields, R. Marotti, N.A. Bretaña, Forecasting inventory for the state-wide pharmaceutical service of South Australia, Procedia Comput. Sci. 219 (2023) 1257–1264.
[27] X. Gao, Z. Gong, Q. Li, G. Wei, Model selection with decision support model for US natural gas consumption forecasting, Expert Syst. Appl. 217 (2023) 119505.
[28] Z.H. Munim, C.S. Fiskin, B. Nepal, M.M.H. Chowdhury, Forecasting container throughput of major Asian ports using the Prophet and hybrid time series models, The Asian Journal of Shipping and Logistics 39 (2023) 67–77.
[29] V. Kumar, S. Sen, Assessment of spring potential for sustainable agriculture: a case study in lesser Himalayas, Appl. Eng. Agric. 36 (2020) 11–24.
[30] V. Kumar, S. Sen, Analysis of spring discharge in the lesser Himalayas: a case study of Mathamali spring, Aglar watershed, Uttarakhand, in: Water Resources Management: Select Proceedings of ICWEES-2016, Springer, 2018, pp. 321–338.
[31] P.R. Winters, Forecasting Sales by exponentially weighted moving averages, Manage Sci 6 (1960) 324–342, https://doi.org/10.1287/mnsc.6.3.324.
[32] D. Dinis, A. Barbosa-Povoa, Â.P. Teixeira, Enhancing capacity planning through forecasting: an integrated tool for maintenance of complex product systems, Int. J. Forecast. 38 (2022) 178–192.
[33] C.C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, Int. J. Forecast. 20 (2004) 5–10, https://doi.org/10.1016/j.ijforecast.2003.09.015.
[34] M.S. Omar, H. Kawamukai, Prediction of NDVI using the Holt-Winters model in high and low vegetation regions: a case study of east Africa, Sci Afr 14 (2021) e01020.
[35] K.F. Pritularga, I. Svetunkov, N. Kourentzes, Shrinkage estimator for exponential smoothing models, Int. J. Forecast. 39 (2022) 1351–1365.
[36] W. Jiang, X. Wu, Y. Gong, W. Yu, X. Zhong, Holt–Winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption, Energy 193 (2020) 116779, https://doi.org/10.1016/j.energy.2019.116779.
[37] T.M. Baykal, H.E. Colak, C. Kılınc, Forecasting future climate boundary maps (2021–2060) using exponential smoothing method and GIS, Sci. Total Environ. 848 (2022) 157633.
[38] B. Billah, M.L. King, R.D. Snyder, A.B. Koehler, Exponential smoothing model selection for forecasting, Int. J. Forecast. 22 (2006) 239–247, https://doi.org/10.1016/j.ijforecast.2005.08.002.
[39] R.D. Snyder, Further Developments in the Estimation of Dynamic Linear Statistical Models, Department of Econometrics and Operations Research, Faculty of Economics and, 1985.
[40] V. Papastefanopoulos, P. Linardatos, S. Kotsiantis, COVID-19: a comparison of time series methods to forecast percentage of active cases per population, Appl. Sci. 10 (2020) 1–15, https://doi.org/10.3390/app10113880.
[41] S.D. Chicas, J.G. Valladarez, K. Omine, V. Sivasankar, S. Kim, Spatiotemporal distribution, trend, forecast, and influencing factors of transboundary and local air pollutants in Nagasaki Prefecture, Japan, Sci. Rep. 13 (2023), https://doi.org/10.1038/s41598-023-27936-2.
[42] C. Xie, H. Wen, W. Yang, J. Cai, P. Zhang, R. Wu, M. Li, S. Huang, Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model, Sci. Rep. 11 (2021) 1–8, https://doi.org/10.1038/s41598-021-81100-2.
[43] X. Li, C. Zhang, W. Li, R.O. Anyah, J. Tian, Exploring the trend, prediction and driving forces of aerosols using satellite and ground data, and implications for climate change mitigation, J. Clean. Prod. 223 (2019) 238–251.
[44] R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, OTexts, 2018.
[45] A.L. Schaffer, T.A. Dobbins, S.A. Pearson, Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions, BMC Med. Res. Methodol. 21 (2021) 1–12, https://doi.org/10.1186/s12874-021-01235-8.
[46] A.B. Alemu, U.J. Parakash Raju, A.M. Seid, B. Damtie, Comparative study of seasonal autoregressive integrated moving average and Holt-Winters modeling for forecasting monthly ground-level ozone, AIP Adv. 13 (2023) 035303, https://doi.org/10.1063/5.0132812.
[47] S.G. Gocheva-Ilieva, A.V. Ivanov, D.S. Voynikova, D.T. Boyadzhiev, Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach, Stoch. Environ. Res. Risk Assess. 28 (2014) 1045–1060, https://doi.org/10.1007/s00477-013-0800-4.
[48] K. Taneja, S. Ahmad, K. Ahmad, S.D. Attri, Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach, Atmos. Pollut. Res. 7 (2016) 585–596, https://doi.org/10.1016/j.apr.2016.02.004.
[49] R.J. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, Package 'forecast,' online]. https://Cran.r-Project.Org/Web/Packages/Forecast/Forecast.Pdf, 2020.
[50] J. Ulrich, TTR: technical trading rules. R Package, 2021, version 0.24.3.
[51] C. Lidiema, Modelling and forecasting Inflation rate in Kenya using SARIMA and holt-winters triple exponential smoothing, Am. J. Theor. Appl. Stat. 6 (2017) 161, https://doi.org/10.11648/j.ajtas.20170603.15.
[52] A. Zeileis, C. Kleiber, S. Jackman, Regression models for count data in R, J. Stat. Software 27 (2008) 1–25.
[53] G.M. Ljung, G.E.P. Box, On a measure of lack of fit in time series models, Biometrika 65 (1978) 297–303, https://doi.org/10.1093/biomet/65.2.297.
[54] H. Hassani, M.R. Yeganegi, Selecting optimal lag order in Ljung–Box test, Phys. Stat. Mech. Appl. 541 (2020) 123700, https://doi.org/10.1016/j.physa.2019.123700.
[55] A. Kabovic, M. Kabovic, S. Bostjancic Rakas, V. Timcenko, The influence of the input parameters variation of the non-seasonal ARIMAX model on the accuracy of meteorological parameters forecasting. 2022 30th Telecommunications Forum, TELFOR 2022 - Proceedings, 2022, pp. 2–5, https://doi.org/10.1109/TELFOR56187.2022.9983670.
[56] V. Papastefanopoulos, P. Linardatos, S. Kotsiantis, COVID-19: a comparison of time series methods to forecast percentage of active cases per population, Appl. Sci. 10 (2020) 1–15, https://doi.org/10.3390/app10113880.
[57] G.E. de M. e Costa, F.C.M. de Menezes Filho, F.A. Canales, M.C. Fava, A.R. A. Brandão, R.P. de Paes, Assessment of time series models for mean discharge modeling and forecasting in a sub-basin of the paranaíba river, Brazil, Hydrology 10 (2023) 208, https://doi.org/10.3390/hydrology10110208.
[58] A.B. Alemu, U.J. Parakash Raju, A.M. Seid, B. Damtie, Comparative study of seasonal autoregressive integrated moving average and Holt-Winters modeling for forecasting monthly ground-level ozone, AIP Adv. 13 (2023) 035303, https://doi.org/10.1063/5.0132812.
[59] Y.J. Puah, Y.F. Huang, K.C. Chua, T.S. Lee, River catchment rainfall series analysis using additive holt–Winters method, J. Earth Syst. Sci. 125 (2016) 269–283, https://doi.org/10.1007/s12040-016-0661-6.
[60] S.D. Chicas, J.G. Valladarez, K. Omine, V. Sivasankar, S. Kim, Spatiotemporal distribution, trend, forecast, and influencing factors of transboundary and local air pollutants in Nagasaki Prefecture, Japan, Sci. Rep. 13 (2023), https://doi.org/10.1038/s41598-023-27936-2.
[61] Y. Zhang, Z. Gu, J.V.G. Thé, S.X. Yang, B. Gharabaghi, The discharge forecasting of multiple monitoring station for Humber River by hybrid LSTM models, Water (Switzerland) 14 (2022), https://doi.org/10.3390/w14111794.

[62] M.A. Al Mehedi, M. Khosravi, M.M.S. Yazdan, H. Shabanian, Exploring temporal dynamics of River Discharge using univariate long short-term memory (LSTM) recurrent neural network at east branch of Delaware river, Hydrology 9 (2022), https://doi.org/10.3390/hydrology9110202.
[63] M.A. Al Mehedi, M. Khosravi, M.M.S. Yazdan, H. Shabanian, Exploring temporal dynamics of River Discharge using univariate long short-term memory (LSTM) recurrent neural network at east branch of Delaware river, Hydrology 9 (2022), https://doi.org/10.3390/hydrology9110202.
[64] R. Tabbussum, A.Q. Dar, Comparative analysis of neural network training algorithms for the flood forecast modelling of an alluvial Himalayan river, J Flood Risk Manag 13 (2020) e12656.
[65] H. Bonakdari, A.D. Binns, B. Gharabaghi, A comparative study of linear stochastic with nonlinear daily River Discharge forecast models, Water Resour. Manag. 34 (2020) 3689–3708, https://doi.org/10.1007/s11269-020-02644-y.
[66] Z.M. Yaseen, Flood hazards and susceptibility detection for Ganga river, Bihar state, India: employment of remote sensing and statistical approaches, Results in Engineering 21 (2024) 101665, https://doi.org/10.1016/j.rineng.2023.101665.
[67] D. Mehta, J. Dhabuwala, S.M. Yadav, V. Kumar, H.M. Azamathulla, Improving flood forecasting in Narmada river basin using hierarchical clustering and hydrological modelling, Results in Engineering 20 (2023) 101571, https://doi.org/10.1016/j.rineng.2023.101571.
[68] A.H. El-Shafie, An application of different artificial intelligences techniques for water quality prediction, Int. J. Phys. Sci. 6 (2011) 5298–5308, https://doi.org/10.5897/IJPS11.1180.
[69] Z.Z. Latt, H. Wittenberg, Improving flood forecasting in a developing country: a comparative study of stepwise multiple linear regression and artificial neural network, Water Resour. Manag. 28 (2014) 2109–2128, https://doi.org/10.1007/s11269-014-0600-8.
[70] M.T. Islam, K. Yoshida, S. Nishiyama, K. Sakai, Mutual validation of remote hydraulic estimates and flow model simulations using UAV-borne LiDAR and deep learning-based imaging techniques, Results in Engineering 20 (2023), https://doi.org/10.1016/j.rineng.2023.101415.
[71] M. Gohil, D. Mehta, M. Shaikh, An integration of geospatial and fuzzy-logic techniques for multi-hazard mapping, Results in Engineering 21 (2024) 101758, https://doi.org/10.1016/j.rineng.2024.101758.
[72] H. Kamyab, T. Khademi, S. Chelliapan, M. SaberiKamarposhti, S. Rezania, M. Yusuf, M. Farajnezhad, M. Abbas, B. Hun Jeon, Y. Ahn, The latest innovative avenues for the utilization of artificial Intelligence and big data analytics in water resource management, Results in Engineering 20 (2023) 101566, https://doi.org/10.1016/j.rineng.2023.101566.