Two-step Daily Reservoir Inflow Prediction Using ARIMA-Machine Learning and Ensemble Models

Akshita Gupta, Research Scholar, agupta@ah.iitr.ac.in
Arun Kumar, Professor, arunkumar@hre.iitr.ac.in

Key Points

- The first step of the two-step algorithm uses time series model ARIMA to forecast the monthly inflows.
- The second step for daily forecasting uses averaging and boosting ensemble models in machine learning.
- The yearly results shows an error less than 5% between actual and predicted values.

Abstract

The reservoirs play a crucial role in the development of civilisation as they facilitate the storage of water for multiple purposes like hydroelectric power generation, flood control, irrigation, and drinking water etc. In order to effectively meet these multiple purposes, the knowledge of the inflow in the reservoir is essential. Apart from the historical data, future prediction of the inflows is also necessary especially in context of climate change. A two-step algorithm for the prediction of reservoir inflow to enable meticulous planning and execution of daily reservoir operation keeping the historical variation of inflow in account has been proposed. The developed algorithm takes into account the patterns in the historic inflow data using the time series analysis along with the variability in the climatic patterns using the different predictors in the machine learning model with a small error. The first step uses time series model, Auto Regressive Integrated Moving Average (ARIMA) method to forecast the monthly inflows, which are then used as the targets in the second step for the month-wise daily forecasting of the inflows using the two types of ensemble models, namely, averaging and boosting models in machine learning. The averaging ensemble models were found to perform better than the boosting ensemble models for maximum number of months. The yearly results show an error of less than 5% between actual and predicted values for all the test cases, showing the precision in the developed algorithm.

Department of Hydro and Renewable energy (HRED), Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, 247667.

Keywords

Inflow prediction, ARIMA, Machine Learning, Ensemble models

1. Introduction

The reservoirs serve as the cornerstone in the management and development of water resources of the river basins. In earlier times, dams were built to fulfil single purpose of either water supply or irrigation. But with the development in technology and increase in population, dams have been constructed to fulfil the multiple purposes for water supply, irrigation, flood control, navigation, water quality, sediment control and energy. These multipurpose dams are very important projects especially for developing countries considering the huge investment and extensive domestic and economic benefits .

But despite the numerous advantages, the construction of new dams has been decreasing continuously after the1970s [Fig.1]. This decrease has occurred because of multiple reasons like increase in construction costs, difficulties in obtaining the clearances, long gestation periods, security concerns etc. Thus, it becomes crucial to operate the already constructed dams economically and with efficient management [*Dams*, November 2000].

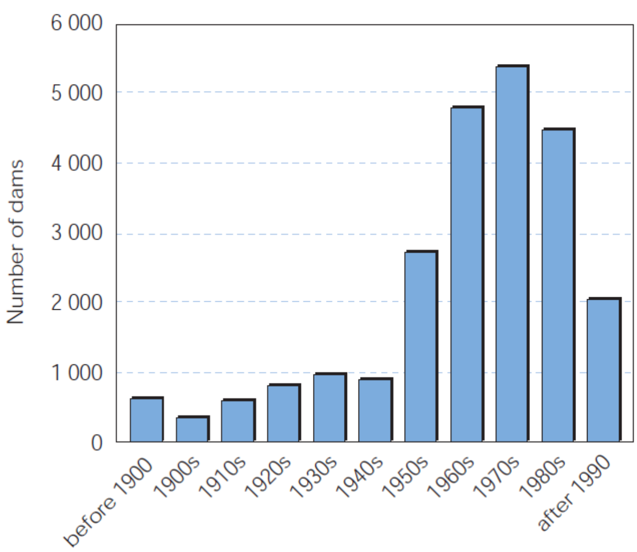


Fig.1: Decade-wise break-up of number of dams constructed [*Dams*, November 2000].

The efficient reservoir management requires the forecasting of inflows along with system variables and outputs like reservoir levels, flood damage risks, water releases, hydropower production, water supply withdrawals, water quality, navigation opportunities, and environmental flows [*Kistenmacher and Georgakakos*, 2015]. One of the major hydrological parameters required planning the construction, operation and maintenance of dams is the record of the hydrological time series data on water available in the area. Apart from past records sometimes, it becomes important to have the information on the future availability of water for planning purposes especially in way of climate change. The frequency at which the inflow data is required depends on the objectives of the study, for example, the daily inflow data is required for carrying out reservoir operation scheduling [*Ahmad and Hossain*, 2019; 2020; *S Yang et al.*, 2019] and monthly or 10-daily inflow data is enough for carrying out the reservoir planning studies [*Salas et al.*, 1985]. The horizon of inflow forecasting also depends on capacity, inflow variability, and forecast uncertainty of the given reservoir [*Zhao et al.*, 2019]. But these kinds of studies require a large time series dataset of past inflow of reservoir in order to predict or forecast the future inflows with acceptable accuracy.

The studies on reservoir inflow forecasting have been carried out for a long time. Most of the studies are carried out for monthly forecasting and only limited studies have been conducted for daily forecasting. The monthly inflow forecasts [*Bae et al.*, 2007; *Yun Bai et al.*, 2015; *Bravo et al.*, 2009; *Cigizoglu*, 2005; *Dariane and Azimi*, 2016; *Ghazali et al.*, 2019; *Jiang et al.*, 2018; *Ma et al.*, 2013; *Mohsenzadeh Karimi et al.*, 2018; *Silva Santos et al.*, 2019; *T Yang et al.*, 2017; *Yin et al.*, 2016; *Y Yu et al.*, 2017] have been carried out successfully with quite accurate results. The usage of daily forecasting [*Guimarães Santos and Silva*, 2014; *Hsu et al.*, 1995; *Londhe and Narkhede*, 2017; *Shiri et al.*, 2012] is limited to reservoir operation. Apart from this, short-term inflow forecasting have also been carried out with 1-hour ahead, 2-hour ahead, 3-hour ahead and so on [*Lin and Wu*, 2011; *Stokelj et al.*, 2002; *Z X Xu and Li*, 2002].

A number of techniques have been employed in the literature for the forecasting process. The most commonly used are the time-series analysis, the Artificial intelligence (AI) and Data mining (DI) approaches. The time series approaches including auto-regressive models like Auto-Regressive Moving Average (ARMA), Auto- Regressive Integrated Moving Average (ARIMA) [*Abdellatif et al.*, 2015; *Lohani et al.*, 2012; *Sveinsson et al.*, 2008; *W Xu et al.*, 2015; *P-S Yu and Tseng*, 2009] and Additive modelling [*Yun Bai et al.*, 2015] have been widely used in the literature. Artificial Neural Network (ANN)[*Abdellatif et al.*, 2015; *Ahmad and Hossain*, 2019; *Esmaeilzadeh et al.*, 2017; *He et al.*, 2014; *Jain et al.*, 1999; *Jothiprakash and Kote*, 2011b; *Rezaie-Balf et al.*, 2019; *Shiri et al.*, 2012; *Silva Santos et al.*, 2019; *Stokelj et al.*, 2002] has been used in significant number of studies for forecasting with the variations in the configurations [*Z X Xu and Li*, 2002]. The different ANN techniques like Multi-layer Perceptron (MLP) [*Ghazali et al.*, 2019; *Muluye and Coulibaly*, 2007], general regression neural network (GRNN) [*Cigizoglu*, 2005; *Ghazali et al.*, 2019; *Kisi and Kerem Cigizoglu*, 2007], radial basis function (RBF)[*Ghazali et al.*, 2019; *Kisi and Kerem Cigizoglu*, 2007], nonlinear autoregressive network with exogenous inputs (NARX) [*Ghazali et al.*, 2019], rotated general regression neural network (RGRNN) [*Yin et al.*, 2016], feed-forward back propagation (FFBP) [*Kisi and Kerem Cigizoglu*, 2007], time-lagged feed-forward networks (TLFN) [*Muluye and Coulibaly*, 2007; *Taghi Sattari et al.*, 2012], Bayesian neural networks (BNN) [*Muluye and Coulibaly*, 2007], multilayer feed-forward ANN [*Bravo et al.*, 2009] have been explored in the literature. The Random Forest (RF)[*Yun Bai et al.*, 2018] is found to work better than the ANN and Support Vector Regression (SVR) [*Esmaeilzadeh et al.*, 2017] for the monthly inflow forecasting [*T Yang et al.*, 2017]. A Modified Box-Cox model along with the Bayesian [*Lima et al.*, 2014; *Ma et al.*, 2013] inference of the model parameters and Markov chain Monte Carlo [*P-S Yu and Tseng*, 2009] approach for modelling the uncertainties have been developed and employed successfully for middle and long term inflow forecasting [*Q J Wang et al.*, 2009]. Other variants of AI like Fuzzy logic, clustering technique [*Nayak and Sudheer*, 2008], Adaptive Neuro-fuzzy Inference System (ANFIS)[*Bae et al.*, 2007; *Dariane and Azimi*, 2016; *He et al.*, 2014; *Lohani et al.*, 2012; *Shiri et al.*, 2012], gene expression programming (GEP) [*Mohsenzadeh Karimi et al.*, 2018; *Shiri et al.*, 2012], support vector machine (SVM)[*He et al.*, 2014; *Mohsenzadeh Karimi et al.*, 2018; *Raghavendra. N and Deka*, 2014; *Y Yu et al.*, 2017], interactive trees (IT) [*Mohsenzadeh Karimi et al.*, 2018], M5 model tree [*Esmaeilzadeh et al.*, 2017; *Jothiprakash and Kote*, 2011a; b; *Rezaie-Balf et al.*, 2019], M5-MT [*Rezaie-Balf et al.*, 2019], multivariate adaptive regression spline (MARS) [*Rezaie-Balf et al.*, 2019], Linear genetic programming (LGP)[*Jothiprakash and Kote*, 2011b], Clustered K Nearest Neighbour (CKNN) [*Akbari et al.*, 2010], extreme learning machine (ELM)[*Yaseen et al.*, 2016] has also been explored with better results than the traditional techniques.

Apart from the individual techniques, the recent trend is towards the development of hybrid techniques for forecasting as these tend to provide better results than the individual techniques [*Yun Bai et al.*, 2016; *Hong*, 2008; *Raghavendra. N and Deka*, 2014]. The combination of low-frequency component using wavelet decomposition [*U. Okkan*, 2012; *Umut Okkan and Ali Serbes*, 2013; *Partal*, 2009] along with time series approach[*Jiang et al.*, 2018] as well as AI [*Dariane and Azimi*, 2016; *Esmaeilzadeh et al.*, 2017; *Guimarães Santos and Silva*, 2014; *Honorato et al.*, 2019; *Kişi*, 2009; *Londhe and Narkhede*, 2017; *Santos et al.*, 2019; *W Wang et al.*, 2009] have provided improved results than the individual techniques. The combination of conceptually different algorithms is also tested and found to give good results [*Coulibaly et al.*, 2005]. The multi-level algorithms have also been tested to fine tune the existing models [*Adarsh and Janga Reddy,* 2019; *Li et al.*, 2016; *Lin and Wu,* 2011; *Singh and Majumdar,* 2009; *Zhou et al.*, 2019]. The ensemble forecasting is another area that is being explored in the recent literature with promising results and future application prospects [*Fan et al.*, 2015; *Jeong and Kim*, 2005].

The increasing variation in the climatic pattern in recent years has resulted in the change of rainfall pattern. The studies investigating the climate change impacts on the reservoir inflows have also been undertaken to conclude that the present inflow patterns are quite different from the past patterns [*Ahmed et al.*, 2015; *P-S Yu et al.*, 2014]. So, it becomes very important to incorporate the variation in the climatic pattern in the inflow model along with the past inflow data. It is acknowledged in many studies that climate conditions significantly impact water supply and many climate phenomenon indices can be used as predictors in supporting water resources management [*Bae et al.*, 2007; *Ma et al.*, 2013; *Santos et al.*, 2019; *Smith et al.*, 1992; *T Yang et al.*, 2017]. The different climatic parameters that have been used for the prediction of inflow are precipitation, evaporation, discharge, temperature and wind speed [*Dixon and Wilby*, 2015; *Esmaeilzadeh et al.*, 2017; *McGuire et al.*, 2006; *Sveinsson et al.*, 2008].

The selection of inflow forecasting technique depends on a number of factors. The purpose of inflow forecasting is a crucial parameter to select the model to be used in forecasting of inflow. This also impacts the complexity of the required model. The selection should also depend upon the historical time series data [*Salas et al.*, 1985]. The comparison of the data-driven techniques for the different time horizons has been performed and some techniques were performing better for shorter time horizons while others were giving good results for monthly predictions [*Zhang et al.*, 2018]. The uncertainties in forecasting arise because of errors in the models being used, their parameters, and the boundary conditions and thus uncertainty analysis has been carried out in inflow forecasts [*Bourdin et al.*, 2014; *Soleimani et al.*, 2016; *P-S Yu and Tseng*, 2009].

Based on the information from the literature it was observed that the daily inflow forecasting is required for the effective reservoir operation and monthly inflow forecasting is required for the planning of reservoirs. This study attempts both of these sequentially, monthly forecasts giving targets to daily forecasting model. The monthly forecasting has been taken up using ARIMA model so as to retain the basic hydrograph profile. Subsequently, daily forecasting is done using different regression models in Machine Learning (ML) in order to incorporate the climatic variations by taking different climatic parameters as the predictors in the model.

2. Methodology

A brief description of the different techniques that have been applied in the study, along with the description of the developed algorithm has been presented in this algorithm.

2.1 Auto-Regressive Integrated Moving Average (ARIMA) model

ARIMA model is obtained by combining autoregressive and moving average models. This model has been widely applied and tested for different types of time-series [*Box and Jenkins*, 1990; *Lee*; *Salas et al.*, 1985]. Most of the real-world data consist of seasonal time series and thus the modelling of seasonal time series besides non-seasonal series is also required to be discussed. The seasonal time series modelling is known as multiplicative $ARIMA(p,d,q)(P,Q,D)_x$ model and is defined as per Equation 1.

$$\begin{aligned}&(1-\phi_1 B^x-\phi_2 B^{2x}-...-\phi_P B^{Px})(1-\varphi_1 B-\varphi_2 B^2-...-\varphi_p B^p)(1-B^x)^D(1-B)^d Z_t = \\ &(1-\Theta_1 B^x-\Theta_2 B^{2x}-...-\Theta_Q B^{Qx})(1-\theta_1 B-\theta_2 B^2-...-\theta_q B^q)\varepsilon_t\end{aligned} \tag{1}$$

where ε_t is random variable, x is periodic term, B is the difference operator given as $B(Z_t) = Z_{t-1}$, $(1-B^x)^D$ is Dth seasonal difference of x, $(1-B)^d$ is dth non-seasonal difference, p is the order of non-seasonal autoregressive model, q is the order of non-seasonal moving average model, P is the order of seasonal autoregressive model, Q is the order of seasonal moving average model, ϕ is the parameter of non-seasonal autoregressive model, φ is the parameter of non-seasonal moving average model, Θ is the parameter of seasonal autoregressive model, and θ is the seasonal moving average model [*Karamouz and Araghinejad*, 2012; *Valipour et al.*, 2013]. The determination of the order of AR and MA terms in ARIMA model is very important and is performed using Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) curves. [*Cryer and Chan*, 2008; *Mohammadi et al.*, 2005].

.

2.2 Regression model using Machine Learning (ML)

Machine learning is one of the many applications of artificial intelligence (AI) which provides the systems with an ability to learn automatically and improve from experiences without being explicitly programmed. The focus of machine learning is on developing computer programs that access data and use it to learn for themselves. The process of learning begins with observations of data, looking for the patterns in data and making better decisions in the future, based on the real time data. The primary aim is to allow the computers to learn automatically without human intervention or assistance and adjust actions accordingly [*Theobald*, 2018].

The ML algorithms are generally classified into two categories: Regression and Classification. For the prediction of data, regression models are used and the basic regression model available is Multiple Linear Regression model. A Multiple Linear Regression fits a linear model with coefficients w = (w1, …, wp) in order to minimize the residual sum of squares between the observed values in the dataset, and the targets are then predicted by the linear approximation.

Other advanced and accurate techniques that are used are ensemble methods with the basic aim to combine the predictions of several base estimators that are built with given learning algorithms so as to improve generalizability as well as the robustness over a single estimator. The ensemble methods are usually of two types and are summarized as below :

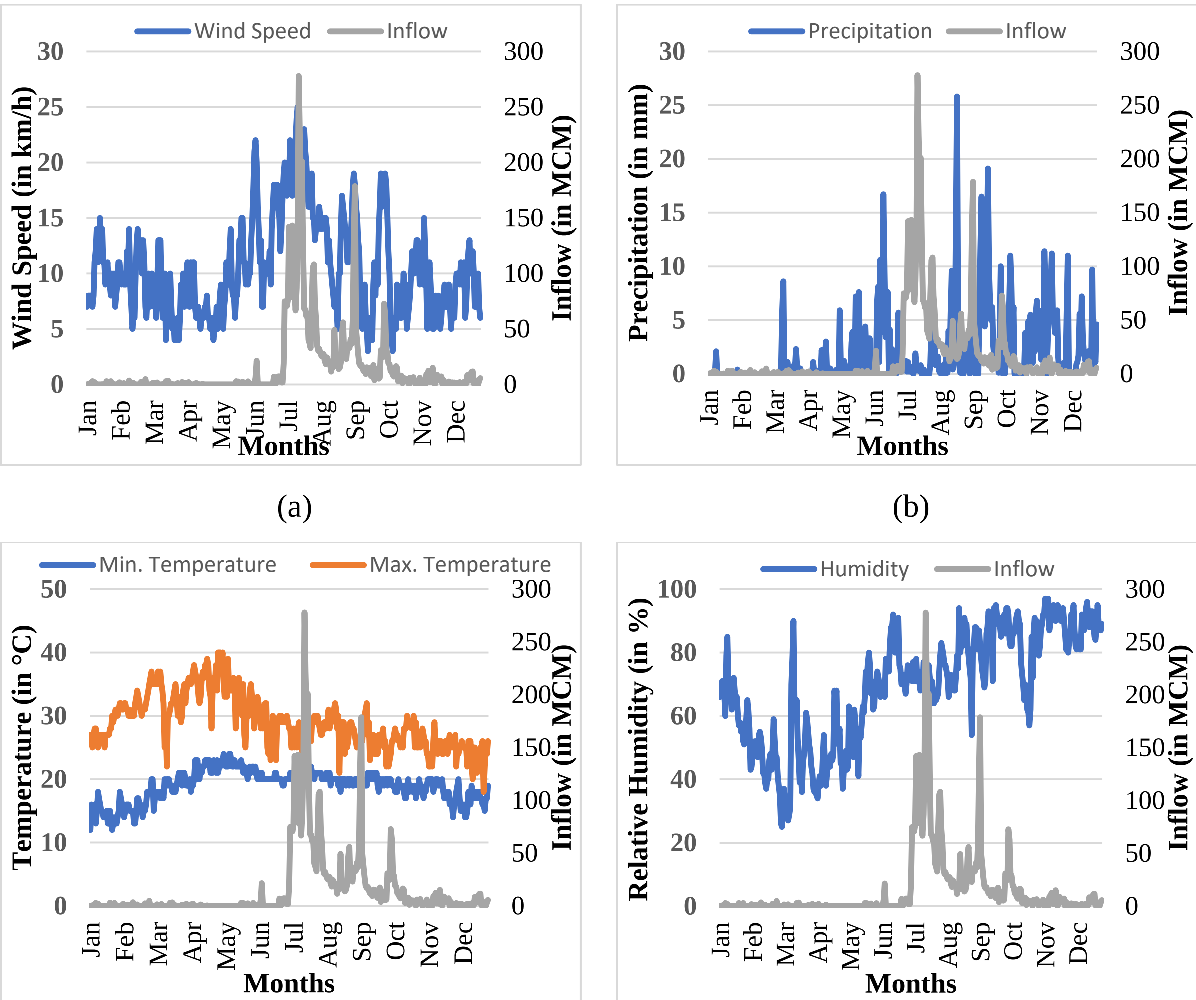
1. Averaging methods: The driving principle for these methods are building of several estimators independently and then to average their predictions. For example, Bagging methods, and Forests of randomized trees, etc.
2. Boosting methods: The base estimators are built sequentially and the aim is to reduce the bias of the combined estimator. The motivation is to combine several weak models to produce a powerful ensemble. For example, AdaBoost, and Gradient Tree Boosting, etc.

This work uses the ensemble methods both averaging and boosting to check the performance of these for inflow forecasting.

2.3 Developed Algorithm

This paper deals with the daily as well as monthly forecasting of reservoir inflow. The forecasting algorithm is divided into two stages: Monthly model and the daily ML model using ensemble methods. The monthly model forecasts the monthly inflows using the ARIMA model, based on the historical inflow time series data. The monthly targets obtained are then used in the next stage for determining the daily inflows using ML incorporating the daily weather conditions into the forecast for different years for the training purpose.

The different weather conditions viz. temperature, precipitation, wind speed, humidity, cloud cover and pressure have been considered in the study and their relation with the daily inflows are shown in Fig.2. It can be seen that each parameter has either a direct or inverse relationship with the inflow value. Apart from these other parameters that have been considered in this study are Day of week, day of month and the direction of wind. Thus, a total of nine parameters have been used in machine learning model.



(a) (b)

(c) (d)

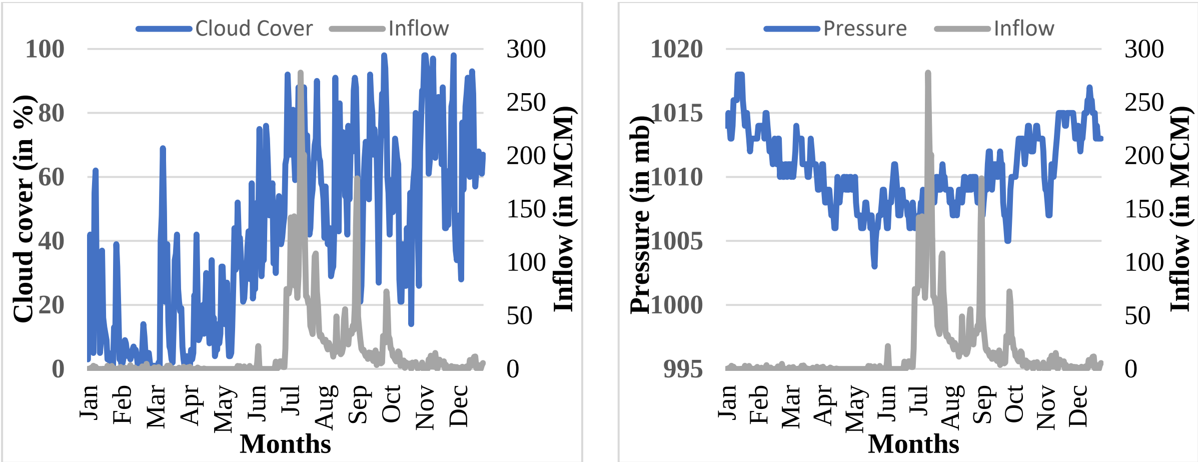


Fig.2: Daily variation of Inflow with the different climatic parameters for a typical year

In the developed algorithm different Machine learning models have been used to predict the daily inflow values on the basis of the targets set by the monthly ARIMA model. The different models used in the study are listed in Table 1. A total of 20 models have been applied for the daily inflow prediction covering both the averaging and the boosting methods. The flowchart of the developed algorithm is shown in Fig.3.

Table 1: Different models of regression used for Daily prediction

S.No.	Model	Abbreviation
Single Regression		
1.	Multiple Linear Regression	MLR
2.	Gradient Boost Regression	GBR
3.	Random Forest Regression	RFR
4.	Extra-Tree Regression	ETR
5.	Ada-Boost Regression	ABR
Bagging Regression		
6.	MLR	BRLR
7.	GBR	BRGBR
8.	RFR	BRRFR
9.	ETR	BRETR
10.	ABR	BRABR
Voting Regression		
11.	MLR, GBR, RFR	VMGR
12.	MLR, GBR, ETR	VMGE
13.	MLR, GBR, ABR	VMGA
14.	MLR, RFR, ETR	VMRE
15.	MLR, RFR, ABR	VMRA
16.	MLR, ETR, ABR	VMEA
17.	GBR, RFR, ETR	VGRE
18.	GBR, RFR, ABR	VGRA
19.	GBR, ETR, ABR	VGEA
20.	RFR, ETR, ABR	VREA

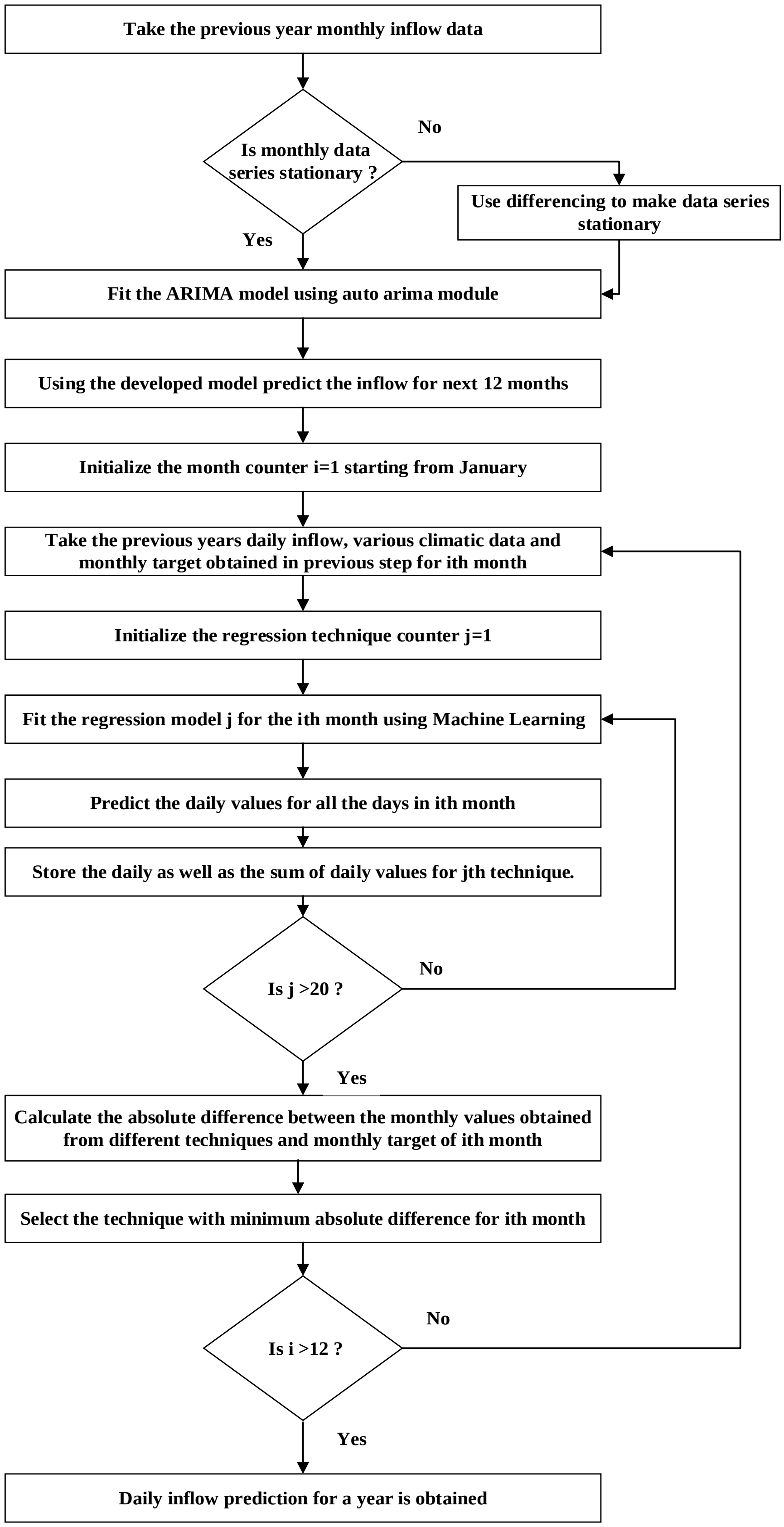


Fig.3: Flow chart of the developed algorithm

3. Problem Setting

The developed algorithm has been tested for the reservoirs in a southern state of India, Karnataka located on the Western part of the Deccan Peninsular region of India as shown in Fig.4.

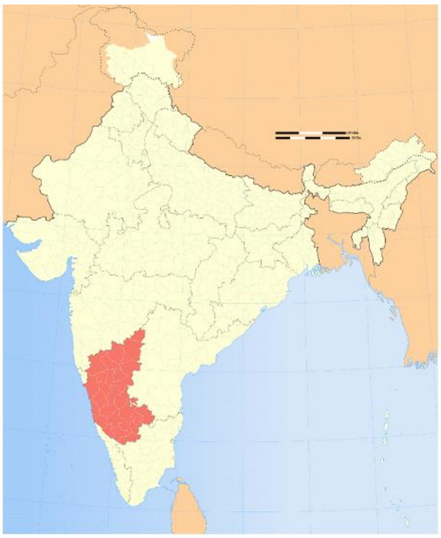


Fig.4: The index map of Karnataka

The annual rainfall in the state varies from 50 to 350 cm during monsoon season from June to September which increases significantly in the western part of the state and reaches its maximum over the coastal belt. There are three major reservoirs Linganmakki, Supa and Mani that are located in the Western part of the state with details as given in Table 2.

Table 2: Details of Linganmakki, Supa and Mani reservoirs

S.No.	Reservoir	Catchment area (sq. km)	Dam Size (m)		FRL (m)	MDDL (m)	Storage (MCM)	
			Height	Length			Gross	Live
1	Linganmakki	1992	59.13	2749	554.43	522.74	4435	4294.5
2	Supa	1057	101	332	564	494	4178	4116
3	Mani	163	59	585	594.36	572	960	881.6

FRL: Full Reservoir Level, MDDL: Maximum Draw Down Level
MCM: Million Cubic Meter

These three reservoirs are of great importance to the state because they are the first and the largest reservoirs in three cascaded hydroelectric schemes totalling to an installed capacity of

Department of Hydro and Renewable energy (HRED), Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, 247667.

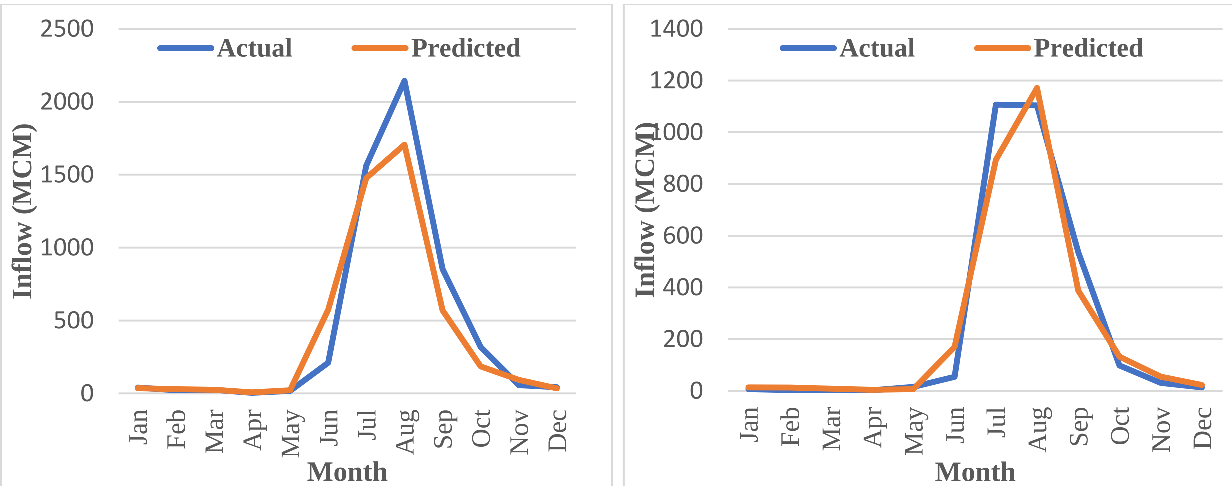
about 3200 MW. Being the first reservoirs in the respective cascades it is important that the inflows of these reservoirs should be predicted accurately for the future scheduling and planning of the reservoirs. This becomes even more crucial with the increase in solar and wind installations in the state of Karnataka as hydropower needs to be scheduled in order to balance the renewable energy penetration, taking into consideration the economic sustainability of themselves [*Gupta et al.*, 2019].

4. Results and Discussion

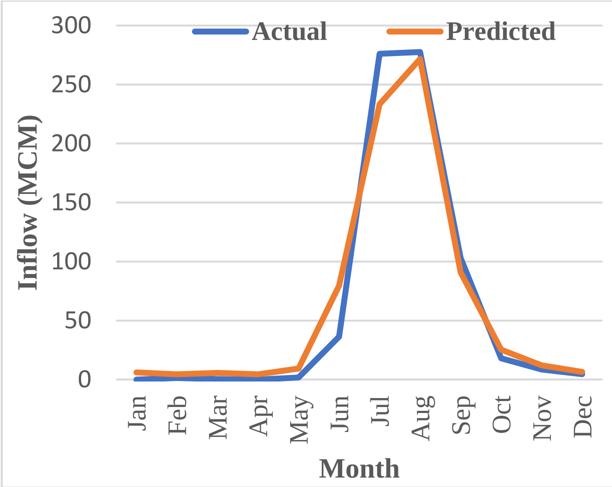
The developed algorithm has been applied to Linganmakki, Supa and Mani reservoir. The following historical data has been used in the algorithm:

1. In ARIMA model for predicting monthly targets 10 years historical inflow data has been used from 2004-14. In this data from 2004-2013 has been used as training data and the 2014 data has been used for validating and results of 2014 have been compared with the actual values.
2. In the machine learning regression models, all the daily data i.e. inflow data, minimum and maximum temperatures, wind speed and direction, humidity, cloud cover, precipitation and pressure are taken for 6 years from 2009-2014. In this data from 2009-2013 has been used as training data and the 2014 data has been used for validating and results of year 2014 haven been compared with the actual values.

The ARIMA model has been used to predict the monthly reservoir inflow for all the three reservoir and the results for the year have been compared with the actual monthly results in Fig 5.



(a) (b)



(c)

Fig.5: Monthly prediction using ARIMA model (a) Linganmakki (b) Supa (c) Mani reservoir

It can be observed that the predicted monthly values of the inflow obtained from the fitted ARIMA model are matching closely with the pattern of the inflow curve for the year. However, the magnitudes of the inflows do not necessarily match the actual values. This can be better visualized from the Table 3, which gives the values of the absolute and relative Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). The absolute values of RMSE and MAE for monsoon and non-monsoon periods does not give a clear picture of the situation because of the fact that in monsoon periods the inflow is very high compared to non-monsoon periods, thus absolute errors are also high. Hence, relative values are calculated based on the maximum values of the monsoon and non-monsoon periods. It shows that a better prediction is obtained for the monsoon period compared to non-monsoon.

Table 3: RMSE and MAE values of the monthly predicted values

S.No	Reservoir	Non-monsoon Period				Monsoon Period			
		RMSE		MAE		RMSE		MAE	
		Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative
1	Linganmakki	129.85	0.408	62.24	0.196	304.71	0.142	268.58	0.125
2	Supa	41.50	0.421	23.89	0.242	155.21	0.141	143.13	0.130
3	Mani	15.26	0.420	9.19	0.253	25.80	0.093	20.20	0.073

After obtaining the monthly targets from the ARIMA model the same targets are then used for the prediction of the daily values using the different Machine Learning algorithms as shown in Table 2. The daily results corresponding to the best monthly regression model for the three reservoirs are shown in Fig.6.

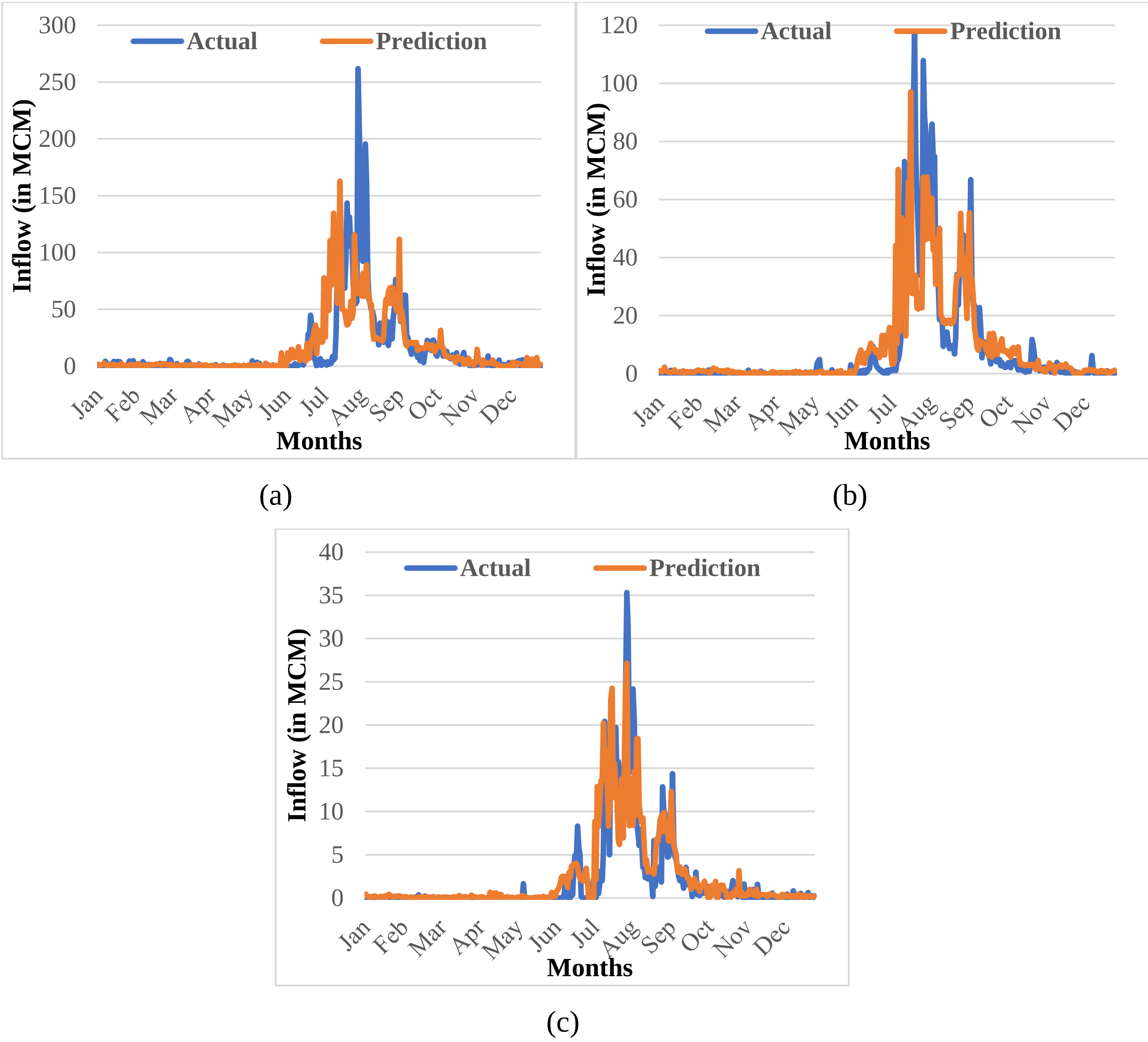


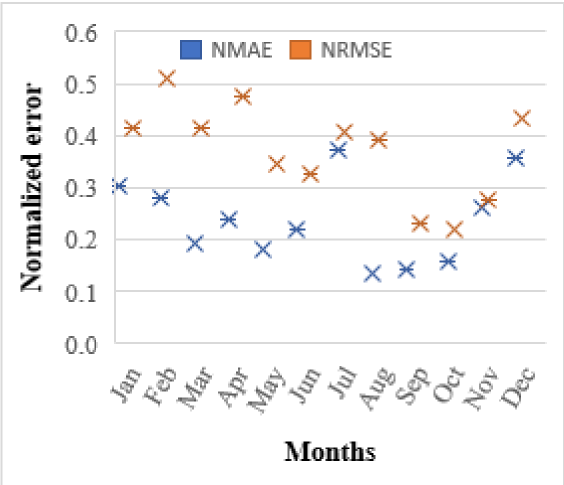
Fig.6: Daily inflow for the three reservoirs for (a) Linganmakki (b) Supa (c) Mani reservoirs

The daily results for the all the reservoirs are based on the monthly best regression technique, out of the 20 models used. The best techniques identified for the different months and for different reservoirs are illustrated in Table 4. It can be seen that for maximum number of months best technique belongs to the averaging ensemble group.

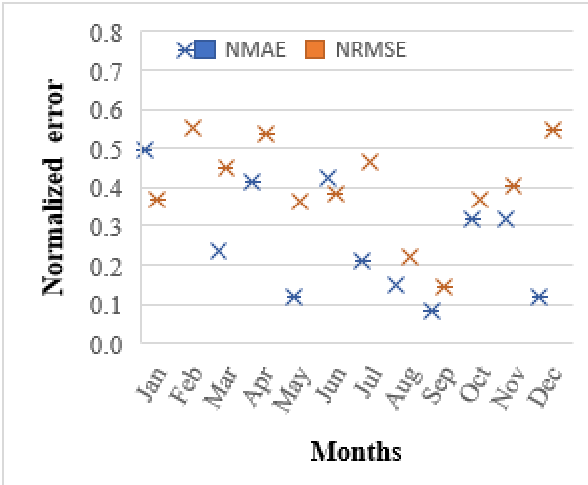
Table 4: Best techniques identified for different months

S.No.	Month	Reservoir		
		Linganmakki	**Supa**	**Mani**
1	Jan	BRGBR	VGRE	BRRFR
2	Feb	VREA	GBR	ABR
3	Mar	VGRA	BRLR	ABR
4	Apr	BRLR	VMEA	VMEA
5	May	ABR	ETR	BRETR
6	Jun	BRETR	VMRA	BRRFR
7	Jul	VGEA	ETR	MLR
8	Aug	BRRFR	ABR	RFR
9	Sep	BRABR	VGRE	VMEA
10	Oct	BRETR	VGRE	ETR
11	Nov	VMGE	BRLR	VGEA
12	Dec	ETR	MLR	VGEA

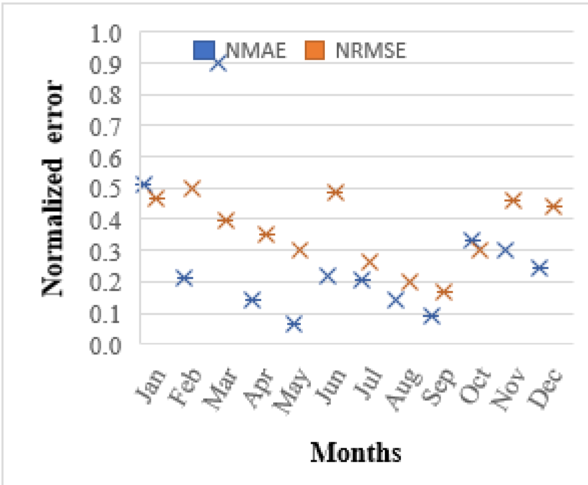
The Fig.6 shows that the daily inflow values obtained from the developed algorithm also follows the inflow trend, just like the monthly inflows. The magnitude comparison can be carried out using Fig.7, where the month-wise Normalized Root Mean Square Error (NRMSE) and the Normalized Mean Absolute Error (NMAE) for the daily inflow for all the three reservoirs are presented. It can be seen that for all the three reservoirs the NMAE values are mostly below 0.5 and that of NRMSE 0.6 for all the months.



(a)



(b)



(c)

Fig.7: Daily NRMSE and NMAE values for (a) Linganmakki (b) Supa (c) Mani reservoir

The Average NRMSE and NMAE values for the daily prediction of inflow for all the reservoirs for monsoon, non-monsoon and full year is shown in Table 5. It is quite evident that the techniques presented work a little better in case of monsoon period compared to the non-monsoon period. But the overall values of errors are quite low for the developed algorithm.

Table 5: Average NRMSE and NMAE values for the daily predicted inflow

S.No.	Reservoir	Non-monsoon period		Monsoon period		Total	
		NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
1	Linganmakki	0.379	0.243	0.342	0.217	0.370	0.237
2	Supa	0.441	0.355	0.276	0.147	0.400	0.303

3	Mani	0.413	0.325	0.209	0.144	0.362	0.280

The consolidated yearly results of the daily inflow prediction for the three reservoirs is summarized in Table 6. The percentage error between the actual and the predicted inflows is less than 5 % for all the three reservoirs. Also, the percentage error is maximum for the non-monsoon period whereas for the monsoon period the error is quite low, which is in agreement to the average NRMSE and MAE values.

Table 6: Yearly values for the daily predicted inflow

S.No.	Reservoir	Non-monsoon period			Monsoon period			Total		
		Actual (MCM)	Predicted (MCM)	Error (%)	Actual (MCM)	Predicted (MCM)	Error (%)	Actual (MCM)	Predicted (MCM)	Error (%)
1	Linganmakki	753	956	26.96	4542	4551	0.20	5295	5507	4.00
2	Supa	229	265	15.72	2747	2790	1.56	2976	3055	2.65
3	Mani	70	80	14.29	657	681	3.65	727	761	4.67

5. Conclusion

The prediction of reservoir inflow is a crucial area for planning the various activities related to reservoirs like hydropower generation, irrigation and drinking water etc. The prediction become increasingly important citing the changing climatic conditions, as the changing the patterns of rainfall alter the patterns of inflow from the historic data. This also reflects the scenario in which the models that are only based upon the historical time series of the inflow may not predict the future values accurately and a data driven model that incorporates the different climatic factors could provide better results.

This work attempts the prediction of the reservoir inflow incorporating the time series analysis as well as the data driven regression analysis. The ARIMA model for time series analysis has been used for monthly prediction of the inflows based on past 10 years data and is used to set the target for the daily prediction. These targets are then used in the daily data driven prediction model. In this the minimum and maximum temperature, humidity, wind speed and direction, pressure, precipitation, cloud cover have been used as the predictors to formulate the model and based on the daily data for these values, prediction of daily inflow for the reservoir has been carried out. A total of 20 regression models, both averaging and boosting ensemble have been fitted for each month and out of them the one giving the least MAE is selected for every month.

The developed model has been tested on the set of three reservoirs located in the state of Karnataka in India. It was observed that the prediction model is more accurate for the monsoon period as compared to the non-monsoon period, especially in the case of monthly ARIMA model. The observed relative RMSE and MAE for the non-monsoon periods are in the range of 0.4-0.45 and 0.2-0.25 respectively, whereas for the monsoon period are in between 0.1-0.15 and 0.1-0.15 respectively. Out of the regression techniques used in the study, it was observed that for maximum number of months in daily inflow prediction, averaging type ensemble methods perform better compared to the boosting methods. The daily month-wise NRMSE and NMAE values were found below 0.6 for maximum number of months. The average values for monsoon and non-monsoon months show that the prediction for monsoon period was better than that of non-monsoon period.

The total yearly results for the three reservoirs show that for the yearly values the percentage error of the predicted values is less than 5%. But in case of monsoon period the error is quite less in comparison to non-monsoon periods. This work could be extended by incorporating the use of weightage in the months to fine tune the forecasting using the time series and machine learning. Also, the uncertainty analysis in inflow prediction could also be considered in future work.

Acknowledgement

The first author would extend a sincere gratitude to the Ministry of Human Resources and Development (MHRD) for providing the financial and technical support in order to carry out the research work at the Indian Institute of Technology (IIT), Roorkee.

Data Availability Statement

The authors confirm that the datasets supporting the problem statement and findings of this study are available within the article.

References

https://www.icold-cigb.org/userfiles/files/CIGB/History%20of%20the%20WRD-PleDelliou.pdf, edited.

https://scikit-learn.org/stable/modules/ensemble.html, edited.

Department of Hydro and Renewable energy (HRED), Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, 247667.

Abdellatif, M. E., Y. Z. Osman, and A. M. Elkhidir (2015), Comparison of artificial neural networks and autoregressive model for inflows forecasting of Roseires Reservoir for better prediction of irrigation water supply in Sudan, International Journal of River Basin Management, 13(2), 203-214, doi: 10.1080/15715124.2014.1003381.

Adarsh, S., and M. Janga Reddy (2019), Multiscale Characterization and Prediction of Reservoir Inflows Using MEMD-SLR Coupled Approach, Journal of Hydrologic Engineering, 24(1), 04018059, doi: 10.1061/(asce)he.1943-5584.0001732.

Ahmad, S. K., and F. Hossain (2019), A generic data-driven technique for forecasting of reservoir inflow: Application for hydropower maximization, Environmental Modelling & Software, 119, 147-165, doi: 10.1016/j.envsoft.2019.06.008.

Ahmad, S. K., and F. Hossain (2020), Maximizing energy production from hydropower dams using short-term weather forecasts, Renewable Energy, 146, 1560-1577, doi: 10.1016/j.renene.2019.07.126.

Ahmed, S., P. Coulibaly, and I. Tsanis (2015), Improved Spring Peak-Flow Forecasting Using Ensemble Meteorological Predictions, Journal of Hydrologic Engineering, 20(2), 04014044, doi: 10.1061/(asce)he.1943-5584.0001014.

Akbari, M., P. J. v. Overloop, and A. Afshar (2010), Clustered K Nearest Neighbor Algorithm for Daily Inflow Forecasting, Water Resources Management, 25(5), 1341-1357, doi: 10.1007/s11269-010-9748-z.

Bae, D.-H., D. M. Jeong, and G. Kim (2007), Monthly dam inflow forecasts using weather forecasting information and neuro-fuzzy technique, Hydrological Sciences Journal, 52(1), 99-113, doi: 10.1623/hysj.52.1.99.

Bai, Y., Z. Chen, J. Xie, and C. Li (2016), Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models, Journal of Hydrology, 532, 193-206, doi: 10.1016/j.jhydrol.2015.11.011.

Bai, Y., P. Wang, J. Xie, J. Li, and C. Li (2015), Additive Model for Monthly Reservoir Inflow Forecast, Journal of Hydrologic Engineering, 20(7), doi: 10.1061/(asce)he.1943-5584.0001101.

Bai, Y., Z. Sun, B. Zeng, J. Long, C. Li, and J. Zhang (2018), Reservoir Inflow Forecast Using a Clustered Random Deep Fusion Approach in the Three Gorges Reservoir, China, Journal of Hydrological Engineering 2018, 23(10), 1-15, doi: 10.1061/(ASCE)HE.1943-5584.0001694.©2018.

Bourdin, D. R., T. N. Nipen, and R. B. Stull (2014), Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system, Water Resources Research, 50(4), 3108-3130, doi: 10.1002/2014wr015462.

Box, G. E. P., and G. Jenkins (1990), Time Series Analysis, Forecasting and Control, 500 pp., Holden-Day, Inc. Bravo, J. M., A. R. Paz, W. Collischonn, C. B. Uvo, O. C. Pedrollo, and S. C. Chou (2009), Incorporating Forecasts of Rainfall in Two Hydrologic Models Used for Medium-Range

Department of Hydro and Renewable energy (HRED), Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, 247667.

Streamflow Forecasting, Journal of Hydrologic Engineering, 14(5), 435-445, doi: 10.1061//ASCE/HE.1943-5584.0000014.

Cigizoglu, H. K. (2005), Generalized regression neural network in monthly flow forecasting, Civil Engineering and Environmental Systems, 22(2), 71-81, doi: 10.1080/10286600500126256.

Coulibaly, P., M. Haché, V. Fortin, and B. Bobée (2005), Improving Daily Reservoir Inflow Forecasts with Model Combination, Journal of Hydrolgic Engineeering, 10(2), 91-99, doi: 10.1061//ASCE/1084-0699/2005/10:2/91.

Cryer, J. D., and K. S. Chan (2008), Time Series Analysis: With Applications in R, Second ed. ed., Springer, New York. Dams, W. C. o. (November 2000), Dams And Development- A New Framework For Decision-Making*Rep.*, London and Sterling, VA.

Dariane, A. B., and S. Azimi (2016), Forecasting streamflow by combination of a genetic input selection algorithm and wavelet transforms using ANFIS models, Hydrological Sciences Journal, 61(3), 585-600, doi: 10.1080/02626667.2014.988155.

Dixon, S. G., and R. L. Wilby (2015), Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study for the River Naryn, Kyrgyzstan, Hydrological Sciences Journal, 61(1), 107-122, doi: 10.1080/02626667.2015.1006227.

Esmaeilzadeh, B., M. T. Sattari, and S. Samadianfard (2017), Performance evaluation of ANNs and an M5 model tree in Sattarkhan Reservoir inflow prediction, ISH Journal of Hydraulic Engineering, 23(3), 283-292, doi: 10.1080/09715010.2017.1308277.

Fan, F. M., D. Schwanenberg, W. Collischonn, and A. Weerts (2015), Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil, Journal of Hydrology: Regional Studies, 4, 196-227, doi: 10.1016/j.ejrh.2015.05.012.

Ghazali, M., T. Honar, and M. R. Nikoo (2019), A fusion-based neural network methodology for monthly reservoir inflow prediction using MODIS products, Hydrological Sciences Journal, 63(15-16), 2076-2096, doi: 10.1080/02626667.2018.1558365.

Guimarães Santos, C. A., and G. B. L. d. Silva (2014), Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models, Hydrological Sciences Journal, 59(2), 312-324, doi: 10.1080/02626667.2013.800944.

Gupta, A., A. Kumar, and D. K. Khatod (2019), Optimized scheduling of hydropower with increase in solar and wind installations, Energy, 183, 716-732, doi: 10.1016/j.energy.2019.06.112.

He, Z., X. Wen, H. Liu, and J. Du (2014), A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region, Journal of Hydrology, 509, 379-386, doi: 10.1016/j.jhydrol.2013.11.054.

Hong, W.-C. (2008), Rainfall forecasting by technological machine learning models, Applied Mathematics and Computation, 200(1), 41-57, doi: 10.1016/j.amc.2007.10.046.

Honorato, A. G. d. S. M., G. B. L. d. Silva, and C. A. Guimarães Santos (2019), Monthly streamflow forecasting using neuro-wavelet techniques and input analysis, Hydrological Sciences Journal, 63(15-16), 2060-2075, doi: 10.1080/02626667.2018.1552788.

Hsu, N.-S., J.-T. KUO, W.-s. ChU, and Y.-J. Lin (1995), Proposed Daily Streamflow Forecasting Model for Reservoir operation, Journal of Water Resoure Planning and Management, 121(2), 132-143.

Jain, S. K., A. Das, and D. K. Srivastava (1999), Application of ANN for reservoir inflow prediction and operation, Journal of Water Resource Planning and Management, 125(5), 263-271.

Jeong, D.-I., and Y.-O. Kim (2005), Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction, Hydrological Processes, 19(19), 3819-3835, doi: 10.1002/hyp.5983.

Jiang, Z., R. Li, C. Ji, A. Li, and J. Zhou (2018), Wavelet analysis-based projection pursuit autoregression model and its application in the runoff forecasting of Li Xiangjiang basin, Hydrological Sciences Journal, 63(12), 1817-1830, doi: 10.1080/02626667.2018.1541091.

Jothiprakash, V., and A. S. Kote (2011a), Effect of Pruning and Smoothing while Using M5 Model Tree Technique for Reservoir Inflow Prediction, Journal of Hydrologic Engineering, 16(7), 563-574, doi: 10.1061/(asce)he.1943-5584.0000342.

Jothiprakash, V., and A. S. Kote (2011b), Improving the performance of data-driven techniques through data pre-processing for modelling daily reservoir inflow, Hydrological Sciences Journal, 56(1), 168-186, doi: 10.1080/02626667.2010.546358.

Karamouz, M., and S. Araghinejad (2012), Advance Hydrology, Amirkabir University of Technology Press.

Kisi, O., and H. Kerem Cigizoglu (2007), Comparison of different ANN techniques in river flow prediction, Civil Engineering and Environmental Systems, 24(3), 211-231, doi: 10.1080/10286600600888565.

Kişi, Ö. (2009), Neural Networks and Wavelet Conjunction Model for Intermittent Streamflow Forecasting, Journal of Hydrologic Engineering, 14(8), 773-782, doi: 10.1061//ASCE/HE.1943-5584.0000053.

Kistenmacher, M., and A. P. Georgakakos (2015), Assessment of reservoir system variable forecasts, Water Resources Research, 51(5), 3437-3458, doi: 10.1002/2014wr016564.

Lee, J. Univariate time series modeling and forecasting (Box-Jenkins Method) Econ 413, lecture 4. .

Li, C., Y. Bai, and B. Zeng (2016), Deep Feature Learning Architectures for Daily Reservoir Inflow Forecasting, Water Resources Management, 30(14), 5145-5161, doi: 10.1007/s11269-016-1474-8.

Lima, L. M. M., E. Popova, and P. Damien (2014), Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models, International Journal of Forecasting, 30(3), 464-476, doi: 10.1016/j.ijforecast.2013.12.009.

Lin, G.-F., and M.-C. Wu (2011), An RBF network with a two-step learning algorithm for developing a reservoir inflow forecasting model, Journal of Hydrology, 405(3-4), 439-450, doi: 10.1016/j.jhydrol.2011.05.042.

Lohani, A. K., R. Kumar, and R. D. Singh (2012), Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques, Journal of Hydrology, 442-443, 23-35, doi: 10.1016/j.jhydrol.2012.03.031.

Londhe, S. N., and S. Narkhede (2017), Forecasting stream flow using hybrid neuro-wavelet technique, ISH Journal of Hydraulic Engineering, 24(3), 275-284, doi: 10.1080/09715010.2017.1360158.

Ma, Z., Z. Li, M. Zhang, and Z. Fan (2013), Bayesian Statistic Forecasting Model for Middle-Term and Long-Term Runoff of a Hydropower Station, Journal Hydrologic Engineering 18(11), 1458-1463, doi: 10.1061/(ASCE)HE.1943-5584.

McGuire, M., A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier (2006), Use of Satellite Data for Streamflow and Reservoir Storage Forecasts in the Snake River Basin, Journal of Water Resource Planning and Management, 132(2), 97-110, doi: 10.1061//ASCE/0733-9496/2006/132:2/97.

Mohammadi, K., H. R. Eslami, and D. Dardashti (2005), Comparison of Regression, ARIMA and ANN Models for Reservoir Inflow Forecasting using Snowmelt Equivalent (a Case study of Karaj) Journal of Agriculture, Science and Technology, 7, 17-30.

Mohsenzadeh Karimi, S., S. Karimi, and M. Poorrajabali (2018), Forecasting monthly streamflows using heuristic models, ISH Journal of Hydraulic Engineering, 1-6, doi: 10.1080/09715010.2018.1516575.

Muluye, G. Y., and P. Coulibaly (2007), Seasonal reservoir inflow forecasting with low-frequency climatic indices: a comparison of data-driven methods, Hydrological Sciences Journal, 52(3), 508-522, doi: 10.1623/hysj.52.3.508.

Nayak, P. C., and K. P. Sudheer (2008), Fuzzy model identification based on cluster estimation for reservoir inflow forecasting, Hydrological Processes, 22(6), 827-841, doi: 10.1002/hyp.6644.

Okkan, U. (2012), Wavelet neural network model for reservoir inflow prediction, Scientia Iranica, 19(6), 1445-1455, doi: 10.1016/j.scient.2012.10.009.

Okkan, U., and Z. Ali Serbes (2013), The combined use of wavelet transform and black box models in reservoir inflow modeling, Journal of Hydrology and Hydromechanics, 61(2), 112-119, doi: 10.2478/johh-2013-0015.

Partal, T. (2009), River flow forecasting using different artificial neural network algorithms and wavelet transform, Canadian Journal of Civil Engineering, 36(1), 26-38, doi: 10.1139/l08-090.

Raghavendra. N, S., and P. C. Deka (2014), Support vector machine applications in the field of hydrology: A review, Applied Soft Computing, 19, 372-386, doi: 10.1016/j.asoc.2014.02.002.

Rezaie-Balf, M., S. R. Naganna, O. Kisi, and A. El-Shafie (2019), Enhancing streamflow forecasting using the augmenting ensemble procedure coupled machine learning models: case study of Aswan High Dam, Hydrological Sciences Journal, 64(13), 1629-1646, doi: 10.1080/02626667.2019.1661417.

Salas, J. D., J. W. Delleur, and V. Yevjevich (1985), Applied Modeling of Hydrologic Time Series, 498 pp., Water Resources Publications.

Santos, C. A. G., P. K. M. M. Freire, R. M. da Silva, and S. A. Akrami (2019), Hybrid Wavelet Neural Network Approach for Daily Inflow Forecasting Using Tropical Rainfall Measuring Mission Data, Journal of Hydrologic Engineering, 24(2), 1-13, doi: 10.1061/(ASCE).

Shiri, J., Ö. Kişi, O. Makarynskyy, A.-A. Shiri, and B. Nikoofar (2012), Forecasting daily stream flows using artificial intelligence approaches, ISH Journal of Hydraulic Engineering, 18(3), 204-214, doi: 10.1080/09715010.2012.721189.

Silva Santos, K. M., A. B. Celeste, and A. El-Shafie (2019), ANNs and inflow forecast to aid stochastic optimization of reservoir operation, Journal of Applied Water Engineering and Research, 7(4), 314-323, doi: 10.1080/23249676.2019.1687017.

Singh, K., and M. Majumdar (2009), A modified two-level estimator for real time forecasting, Hydrological Sciences Journal, 38(5), 417-430, doi: 10.1080/026266693099492691.

Smith, J. A., G. N. Day, and M. D. Kane (1992), Non-Parametric Framework for Long-range streamflow forecasting, Journal of Water Resouce Planning and Management, 118(1), 82-92.

Soleimani, S., O. Bozorg-Haddad, and H. A. Loáiciga (2016), Reservoir Operation Rules with Uncertainties in Reservoir Inflow and Agricultural Demand Derived with Stochastic Dynamic Programming, Journal of Irrigation and Drainage Engineering, 142(11), 04016046, doi: 10.1061/(asce)ir.1943-4774.0001065.

Stokelj, T., D. Paravan, and R. Golob (2002), Enhanced Artificial Neural Network Inflow Forecasting Algorithm for Run-of-River Hydropower Plants, Journal of Water Resoure Planning and Management, 128(6), 415-423, doi: 10.1061//ASCE/0733-9496/2002/128:6/415.

Sveinsson, O. G. B., U. Lall, V. Fortin, L. Perrault, J. Gaudet, S. Zebiak, and Y. Kushnir (2008), Forecasting Spring Reservoir Inflows in Churchill Falls Basin in Québec, Canada, Journal of Hydrologic Engineering, 13(6), 426-437.

Taghi Sattari, M., K. Yurekli, and M. Pal (2012), Performance evaluation of artificial neural network approaches in forecasting reservoir inflow, Applied Mathematical Modelling, 36(6), 2649-2657, doi: 10.1016/j.apm.2011.09.048.

Theobald, O. (2018), Machine learning for absolute beginners : a plain English introduction.

Valipour, M., M. E. Banihabib, and S. M. R. Behbahani (2013), Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir, Journal of Hydrology, 476, 433-441, doi: 10.1016/j.jhydrol.2012.11.017.

Wang, Q. J., D. E. Robertson, and F. H. S. Chiew (2009), A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, Water Resources Research, 45(5), doi: 10.1029/2008wr007355.

Wang, W., J. Jin, and Y. Li (2009), Prediction of Inflow at Three Gorges Dam in Yangtze River with Wavelet Network Model, Water Resources Management, 23(13), 2791-2803, doi: 10.1007/s11269-009-9409-2.

Xu, W., J. Zhao, T. Zhao, and Z. Wang (2015), Adaptive Reservoir Operation Model Incorporating Nonstationary Inflow Prediction, Journal of Water Resources Planning and Management, 141(8), doi: 10.1061/(asce)wr.1943-5452.0000502.

Xu, Z. X., and J. Y. Li (2002), Short-term inflow forecasting using an artificial neural network model, Hydrological Processes, 16(12), 2423-2439, doi: 10.1002/hyp.1013.

Yang, S., D. Yang, J. Chen, and B. Zhao (2019), Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model, Journal of Hydrology, 579, 124229, doi: 10.1016/j.jhydrol.2019.124229.

Yang, T., A. A. Asanjan, S. Sorooshian, E. Welles, X. Gao, and X. Liu (2017), Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information, Water Resource Research 53, 2786–2812, doi: 10.1002/.

Yaseen, Z. M., O. Jaafar, R. C. Deo, O. Kisi, J. Adamowski, J. Quilty, and A. El-Shafie (2016), Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq, Journal of Hydrology, 542, 603-614, doi: 10.1016/j.jhydrol.2016.09.035.

Yin, S., D. Tang, X. Jin, W. Chen, and N. Pu (2016), A combined rotated general regression neural network method for river flow forecasting, Hydrological Sciences Journal, 61(4), 669-682, doi: 10.1080/02626667.2014.944525.

Yu, P.-S., and T.-Y. Tseng (2009), A model to forecast flow with uncertainty analysis, Hydrological Sciences Journal, 41(3), 327-344, doi: 10.1080/02626669609491506.

Yu, P.-S., T.-C. Yang, C.-M. Kuo, J.-C. Chou, and H.-W. Tseng (2014), Climate change impacts on reservoir inflows and subsequent hydroelectric power generation for cascaded hydropower plants, Hydrological Sciences Journal, 59(6), 1196-1212, doi: 10.1080/02626667.2014.912035.

Department of Hydro and Renewable energy (HRED), Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, 247667.

Yu, Y., P. Wang, C. Wang, J. Qian, and J. Hou (2017), Combined Monthly Inflow Forecasting and Multiobjective Ecological Reservoir Operations Model: Case Study of the Three Gorges Reservoir, Journal of Water Resources Planning and Management, 143(8), doi: 10.1061/(asce)wr.1943-5452.0000786.

Zhang, Z., Q. Zhang, and V. P. Singh (2018), Univariate streamflow forecasting using commonly used data-driven models: literature review and case study, Hydrological Sciences Journal, 63(7), 1091-1111, doi: 10.1080/02626667.2018.1469756.

Zhao, Q., X. Cai, and Y. Li (2019), Determining Inflow Forecast Horizon for Reservoir Operation, Water Resources Research, doi: 10.1029/2019wr025226.

Zhou, W., Z. Yang, P. Liu, F. Bai, and C. Zheng (2019), Estimation of reservoir inflow with significant lateral inflow by using the adjoint equation method, Journal of Hydrology, 574, 360-372, doi: 10.1016/j.jhydrol.2019.04.047.