An improved framework to predict river flow time series data

Hafiza Mamona Nazir[1], Ijaz Hussain[1], Ishfaq Ahmad[2], Muhammad Faisal[3,4], Ibrahim M. Almanjahie[2]

[1]*Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan*

[2]*Department of mathematics, college of science, king Khalid university, 61413 Abha, Saudi Arabia*

[3]*Faculty of Health Studies, University of Bradford, Bradford BD7 1DP, UK*

[4]*Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK*

Abstract

Due to non-stationary and noise characteristics of river flow time series data, some pre-processing methods are adopted to address the multi-scale and noise complexity. In this paper, we proposed an improved framework comprising Complete Ensemble Empirical Mode Decomposition with Adaptive Noise-Empirical Bayesian Threshold (CEEMDAN-EBT). The CEEMDAN-EBT is employed to decompose non-stationary river flow time series data into Intrinsic Mode Functions (IMFs). The derived IMFs are divided into two parts; noise-dominant IMFs and noise-free IMFs. Firstly, the noise-dominant IMFs are denoised using empirical Bayesian threshold to integrate the noises and sparsities of IMFs. Secondly, the denoised IMF's and noise free IMF's are further used as inputs in data-driven and simple stochastic models respectively to predict the river flow time series data. Finally, the predicted IMF's are aggregated to get the final prediction. The proposed framework is illustrated by using four rivers of the Indus Basin System. The prediction performance is compared with Mean Square Error, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Our proposed method, CEEMDAN-EBT-MM, produced the smallest MAPE for all four case studies as compared with other methods. This suggests that our proposed hybrid model can be used as an efficient tool for providing the reliable prediction of non-stationary and noisy time series data to policymakers such as for planning power generation and water resource management.

1   Introduction

32   The economic development of any country is directly linked to the proper management of their

33   water resources operations that can minimize the effects of various natural disasters such as floods

34   and droughts. Therefore, river flow time series prediction is an imperative task, which plays a

35   significant role for effective and appropriate water resource planning and management, early flood

36   warning, irrigation, and hydropower generation (*Yaseen et al.*, *201*8). Several algorithms have

37   been used for prediction and estimation for river flow time series data *(Aichouri et al., 2015;*

38   *Shathir & Saleh, 2016; Almasudi 2018; Gjika, Aurora & Arbesa, 2019;*). The Box-Jenkins

39   methodology technique (*Box & Jenkins,1970*) is commonly used in the literature (*Shathir & Saleh,*

40   *2016*) because it can be used in a wide class of models i.e. Autoregressive (AR), Moving Averages

41   (MA), Autoregressive Integrated Moving Averages Model (ARIMA), Seasonal ARIMA. But the

42   problem with such statistical models is that they only considered the stationary and linear behavior

43   of the process (*Di, Yang, & Wang, 2014*). However, river flow time series data is non-linear, non-

44   stationary, multi-scale and noise-corrupted (*Di, Yang, & Wang, 2014*) as stochastic nature of

45   several factors (e.g., rainfall, evaporation, and temperature). This complex non-stationary, multi-

46   scale and noise-corrupted characteristics make the prediction a challenging task. During the past

47   decades, this issue has been addressed by developing some data-driven models i.e. Artificial

48   Neural Network (ANN), model tree, Support Vector Machine (SVM), Adaptive Inference-Based

49   Neural Network (AIBNN), Extreme Learning Machine (ELM) (Ali *et al., 2018*). *Yaseen et al.*

50   *(2018)* demonstrated the viability of ELM and enhanced version of ELM method to forecast river

51   flow data as compared to other statistical models. However, the data-driven models ignore the

52   time-varying and noise characteristics of river flow processes which deprives the researcher to

53   efficiently predict such data. To address the drawback of such data-driven models, several hybrid

54   models are introduced to extract the time-varying information and reduce noises which ultimately

55   increase the prediction accuracy of data-driven models *(Toth, Brath & Montanari, 2000 ;Di, Yang,*

56   *& Wang, 2014; Su et al., 2016; Kang et al., 2017; Hadi & Tombul, 2018; Yaseen et al., 2018*).

57   The hybrid model uses data pre-processing methods such as Singular Spectrum Analysis (SSA),

58   Wavelet Analysis (WA), Empirical Mode Decomposition (EMD) and Empirical-EMD (EEMD)

59   with data-driven models also called intelligence models. An advantage of such data decomposition

60   methods is that they used not only for decomposing the data into time-frequency components but
61   also used to separate noises from data. *Wu & Chau (2011)* coupled data pre-processing techniques
62   i.e. SSA with ANN to accurately model the rainfall and river flow data. *Azadeh et al. (2011)*
63   demonstrated the ability to use the data pre-processing method to enhance the precision of data-
64   driven models. They used various data processing techniques and reported that the processed non-
65   linear data is efficiently forecasted with simple statistical and intelligent models. *Gjika, Aurora &*
66   *Arbesa (2019)* found that proper consideration of the volatility nature of non-linear data through
67   data pre-processing methods could improve the prediction quality. *Santos et al. (2019)* proposed a
68   model by coupling WA and ANN techniques. They used WA to transform the daily flow time
69   series data to enhance the precision of the ANN model. The data pre-processing methods used to
70   decompose the non-stationary and non-linear data into physical modes of time-frequency
71   components (*Han & Liu, 2009; Azadeh et al., 2011; Wang et al., 2015; Peng et al., 2017*). The
72   derived time-frequency components, also called multi-scale components (*Nazir et al., 2019*), are
73   predicted through a mixture of data-driven models accurately (*Azadeh et al., 2011*). Further, the
74   time-frequency components, are filtered out by using appropriate thresholds. The reason for using
75   a filter is to preserve significant features of original time series data while removing noises or
76   sparsity from multi-scale components. A growing body of research found on denoising the
77   extracted high and low multi-scale components in different fields by using linear and non-linear
78   thresholds. Moreover, many other methods of thresholds include Stein Unbiased Risk Estimator
79   (SURE) (*Cand, Sing-long &Trzasko, 2012; Hansen, 2017*), fixed and soft threshold (*Di, Yang &*
80   *Wang, 2014; Nazir et al., 2019*), and minimax algorithms (*Hansen, 2017*) are also used to remove
81   noises and to preserve important information from complex data. *Peng et al. (2017)* developed a
82   novel hybrid model by employing an empirical wavelet transform estimator to remove the
83   redundant noises from river flow data. Further, the denoised data are predicted through Particular
84   Swarm Optimization based-ANN model (PSO-ANN). They demonstrated the efficiency of their
85   proposed model over simple PSO-ANN model without denoising. *Di, Yang &Wang, (2014)*
86   proposed a hybrid model by considering two data pre-processing methods i.e. EMD and WA with
87   a soft and hard threshold to find the denoised time-varying information to decrease the complexity
88   of hydrological series. *Holzfuss & Kadtke (1993)* suggested that the noise reduction methods
89   coupled with the radial basis functions may enhance the quality of traditional statistical and data-
90   driven models.

However, WA, EMD, and EEMD based noise removal techniques have their own drawbacks in extracting the optimal multi-scale components. WA-based hard and soft threshold first comprised of the choice of mother wavelet, which is subjectively selected among many wavelet basis functions. The subjective selection of mother wavelet may also cause errors which decrease the performance of hard and SURE thresholds. Moreover, the EMD is a purely data-driven technique, used for pre-processing data, which is effected through its own mathematical property of mode mixing problem resulting in spurious time-frequency information. To overcome the drawback of EMD, the EEMD is introduced which is an improved version of EMD to solve the mode mixing problem. The EEMD added Gaussian white noise successively to solve noise-assisted (*Hadi & Tombul, 2018; Yaseen et al., 2018*). Although an improved EEMD has proved useful for denoising hydrological time series data (*Jiao, Guo & Ding, 2016*), it also has some drawbacks that may deprive the researchers in extracting accurate multi-scale components i.e. Intrinsic Mode Function (IMF) by simple averaging them. However, to cope up the simple averaging problem of EEMD, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) have been proposed by (*Torres et al., 2011*). The application of CEEMDAN is successfully applied to derive the time-scale components of hydrological time series data (*Antico, Schlotthauer & Torres, 2014*). *Johnstone & Silverman (2005)* have considered an empirical Bayesian approach to threshold the multi-scale components derived from WA. They argued that the empirical Bayesian method efficiently models the sparsity and noises of complex data by considering multiple priors for each level. One would hope such methods which possibly estimate thresholds that stably reflect the noises from sparse multi-scale to enhance the accuracy and reliability of prediction performance of complex river flow data.

In this study, we aimed to improve denoising stage of the hybrid model by the novel framework i.e. CEEMDAN-Empirical Bayesian Threshold estimator (EBT) estimator to optimally reduce noises from IMFs which are further used as inputs to get a precise prediction of river flow data which plays a decisive role in the accurate prediction. The principal motivation of choosing EBT is that it is purely a data-based method which deals different level of noises efficiently because of some high multi-scale components derived from CEEMDAN relatively sparse than the lower time-scale components.

2 Proposed Method

123

The proposed method is comprised of four stages such as decomposed, denoised using novel threshold, prediction, and aggregation. The CEEMDAN method is used as a decomposition tool to handle the non-linear and non-stationary data by extracting IMFs. The extracted IMFs are further divided into two parts; one part is comprised of noisiest IMFs which contains errors and sparsity and the second part is noise free IMFs. The noisiest IMF's coupled with novel Empirical Bayesian Threshold (EBT) estimator to get a ride from noises and sparsity to denoise IMF's. Then the denoised IMF's are predicted through the complex data-driven model and noise free IMF's are predicted through a simple stochastic model. Finally, all the predicted IMF's are aggregated to get the final prediction. The proposed framework of deriving multi-scale IMF's through CEEMDAN coupled with optimal denoising method i.e. EBT plays a vital role in the accurate prediction of river flow data. To our convenience, the proposed strategy is labialized as; CEEMDAN (decomposed), EBT (denoised) and Multi Models MM (data-driven and stochastic model) i.e. CEEMDAN-EBT-MM and the proposed scheme is illustrated in Fig. 1.

137

2.1 Decomposition stage

EMD: To decompose non-linear and non-stationary data, the EMD method introduced by (*Huang et al, 1998*) which decomposed data into IMF's by satisfying two conditions as follows: **(a)** The number of zero crossings and extreme, from complete data, must be equal or differ to one at most; **(b)** At all level, mean value of envelope must be zero.

The complete EMD procedure is defined as:

1. Find all local maxima and minima from data $y(t), (t = 1,2,..,N)$. Use cubic splines interpolation to find an upper envelope from maxima $e_{\max(t)}$ and lower envelope from minima $e_{\min(t)}$.

2. Calculate the average of upper and lower envelope $m(t) = (e_{\max(t)} + e_{\min(t)})/2$. Take the deviations between original time series data and calculated envelope mean as:

$$g(t) = y(t) - m(t) \qquad (1)$$

149   3. Match the requirements of $g(t)$ which is defined in **(a)** and **(b)** as an IMF, if the conditions

150   are satisfied then mark this $g(t)$ as i^{th} IMF.

151   4. In the next step replace original time series $y(t)$ by $r(t) = y(t) - g(t)$, if $g(t)$ does not

152   meet the **(a)** and **(b)** then just replace $y(t)$ with $g(t)$.

153   5. The process of (1-4) is repeated until no IMF is being extracted from the residue $r(t)$.

154   In the end, original time series data will be written as the summation of all extracted IMF's and

155   residue as:

$$y(t) = \sum_{i}^{l} g_i(t) + r(t) \tag{2}$$

156   where $g_i(t)$ is the ith IMF and $r(t)$ is the trend of the signal. However, EMD techniques have two

157   drawbacks. The one is the endpoint problem that is the extreme values of two ends of the series

158   can't be determined properly which distort the IMF's and the other is mode mixing aliasing in

159   which same IMF contains even more than two frequencies.

160

161   *EEMD:* To resolve the mode mixing problem of EMD, (*Wu & Huang*, 2004) proposed EEMD.

162   The EEMD decompose the non-linear signals into IMF's as follows:

163   a) Add a white Gaussian noise series to the original data set as follows:

$$y_2(t) = y(t) + n(t) \tag{3}$$

164   b) Decompose the new $y_2(t)$ with EMD and obtains the IMFs.

165   c) Repeat step (a) and (b) m^{th} time i.e. ($j = 1,2,...,m$) with different white noises to get IMF's

166   from new series.

167   d) Find the ensemble means of all IMFs obtained as m^{th} ensemble time as follows: where

168        $k = 1,2,..,K$ is k^{th} IMF.

$$\overline{imf}_k = \sum_{j=1}^{m} IMF_{jk}^m / m \tag{4}$$

169   where $k = 1,2,..,K$ is k^{th} IMF.

170   *CEEMDAN:* Although the mode mixing problem is alleviated with EEMD, taking the simple

171   averages of IMF's without considering the independent addition of white noises could not

172   completely remove noises from IMF's. To tackle the simple averaging problem of EEMD,

173 CEEMDAN function is introduced by (*Torres et al., 2011*). Here in our study, CEEMDAN is used

174 to decompose the river flow data which is briefly described as follows:

175 1) The CEEMDAN add Gaussian noises like EEMD in signals as follows:

$$y_3(t) = y(t) + w_0 n^j(t) \tag{5}$$

176

177 where w_0 is the amplitude of the added white noises and $(j = 1,2,...,m)$. Find the first IMF

178 using simple EEMD defined as:

$$\widetilde{IMF}_1 = \sum_{j=1}^{m} IMF_{j1}^m / m \tag{6}$$

179

180 2) Compute the deviation of original signals from the first IMF as:

$$r_1(t) = y(t) - \overline{imf}_1 \tag{7}$$

181 3) Decompose $r_1(t) + w_0 n^j(t)$ to get first IMF and find the second IMF as:

$$\widetilde{IMF}_2 = \sum_{j=1}^{m} IMF_{j1}^m / m \tag{8}$$

182 4) Repeat the (2-3) until stoppage criteria are met and residual contains not more than two

183 extremes. Finally, the residual is defined as:

$$R(t) = y(t) - \sum_{k=1}^{K} \widetilde{IMF}_k \tag{9}$$

184 However, the selection of a number of ensemble and amplitude of white noise is still an

185 open challenge but here in our study we used a number of ensemble members as 100 and

186 standard deviation of white noise is settled as 0.2 according to (*Di, Yang & Wang, 2014*).

187

188 *Identification of noisiest IMFs:* After deriving the IMF's, next step is to screen out the noise only

189 IMF's and noise free IMFs (*Wei et al., 2016*). Two types of IMF's are derived from

190 CEEMDAN/EEMD; the IMF's contains high frequency which is corrupted with sparsity and

191 noises (*Wei et al., 2016*) and the second part of IMF's comprised on low frequencies which are

192 free from noises (*Wei et al., 2016*). To get numerical validation, cross-correlation is calculated

193 between all extracted IMF's and original river flow data. The low cross-correlation implies that

194 the high-frequency IMF's are overwhelmed with noises. Then, the noise only IMF's are further

195   denoised through the appropriate threshold to get the important features from them and to get rid
196   of noises.

197

2.2 Denoising stage

199

200   After decomposing the non-linear and non-stationary data into IMF components, an appropriate
201   estimator is chosen to remove noises and sparsities from extracted IMFs. The reason for selecting
202   an appropriate estimator is to find an optimal value of threshold as the highest threshold value
203   would lead to biases, whilst the lowest threshold value would increase the noise variance. The
204   IMFs which are extracted are mostly empirical Bayesian threshold estimator is adapted to denoised
205   the noisiest IMFs. Later, the existing soft and hard threshold and improved threshold functions are
206   used for comparison purposes. Details of all estimators are given below:

207   *CEEMDAN/EEMD-based Empirical Bayesian Threshold:* To estimate the sparseness and noises
208   from decomposed IMF's (*Wei et al., 2016*), an Empirical Bayes Threshold (EBT) inspired from
209   wavelet denoising (*Johnstone & Silverman, 2005; Jansen & Bultheel, 2014*) is used. For the
210   successful application of EBT, first, all the data is scaled transformed to efficiently select the prior
211   distribution for noises and sparsities. After the scaled transformation data follows $N(\theta_i,1)$. Then a
212   mixture of priors for θ_i are considered as follows:

$$f_{prior}(\theta) = (1-w)\delta_0(\theta) + w\gamma(\theta) \tag{10}$$

213   where $\delta_0(\theta)$ is a zero part of scaled data and $\gamma(\theta)$ is a density of non-zero part. The density of
214   prior should be chosen in such a way that it must belong to a family of distributions whose tails
215   decays at polynomial rates. The parameters and weights of a mixture of prior distributions are
216   are estimated through maximum likelihood approach. The reason of using this method to estimate
217   unknowns is that it estimates weights and parameters to be proportional to the likelihood function
218   evaluated at the estimators based on data (*Hossain, Kozubowski & Podgórski, 2018*).
219   Finally, the posterior median $\tilde{\theta}_i(imf,w)$ is calculated from a mixture of prior distribution is given
220   as follows:

$$\tilde{F}_1(\mu \,|\, imf) = \int_{\mu}^{\infty} f_1(\mu \,|\, imf)d\mu \tag{11}$$

221   which is used as a threshold rule for $\tilde{\mu}$ given data (*Johnstone & Silverman, 2005*). In general, an

222   estimation rule comprised on $\eta(imf,t)$ defined for all $t > 0$, is a thresholding rule if and only if

223   for all $t > 0$, $\eta(imf,t)$ is an antisymmetric and increasing function of data and $\eta(imf,t) = 0$ if and

224   only if $|imf| \leq t$ where t is defined as a median value which is calculated through the Eq. (11).

225   *Other traditional threshold estimators:* To compare the EBT estimator with other non-linear

226   estimators to suppress the noises and sparsities from noisiest IMFs, Soft Threshold (ST), Hard

227   Threshold (HT) and Improved Threshold Function (ITF) are used which are most widely used in

228   literature (*Jeng et al., 2007; Candes, Sing-Long & Trzasko, 2012; Jansen & Bultheel, 2014*) listed

229   as follows, respectively;

$$IMF'_{t,k} = \begin{cases} IMF_{t,k} & |IMF_{t,k}| \geq T_k \\ 0 & |IMF_{t,k}| < T_k \end{cases} \tag{12}$$

230

$$IMF'_{t,k} = \begin{cases} sgn(IMF_{t,k})(|IMF_{t,k}| - T_k) & |IMF_{t,k}| \geq T_k \\ 0 & |IMF_{t,k}| < T_k \end{cases} \tag{13}$$

231   and

$$IMF'_{t,k} = \begin{cases} sgn(IMF_{t,k})\left(\dfrac{|IMF_{t,k}| - T_k}{\exp\left\{3\alpha\left(\dfrac{IMF_{t,k} - T_k}{T_k}\right)\right\}}\right) & |IMF_{t,k}| \geq T_k \\ 0 & |IMF_{t,k}| < T_k \end{cases} \tag{14}$$

232   where, T_k is the threshold value calculated as $T_k = a\sqrt{2E_k \ln(N)}$, where $k = 1,2,.....,K$ and a is

233   constant takes the values between 0.4 to 1.4 with a step of 0.1 and $E_k =$

234    $median(|IMF_{t,k}|, t = 1,2,...,N)/0.6745$ is median deviation of kth IMF.

235   **2.3 Prediction and aggregation**

236   In the prediction stage, the decomposed-denoised IMF's of river flow data are predicted through

237   some data-driven and statistical models. Specifically, the denoised IMFs are predicted through

238   data-driven model whenever noise-free IMFs and residual are predicted through a simple

239   stochastic model. To train model, 70%, 80%, and 90% data are used remaining 30%, 20% and

240   10% data used to test the accuracy of models. The selected models are briefly described as follows:

241   *The denoised-IMF prediction with the neural network:* A data-driven techniques have been proved

242   a powerful tool to model the complex non-linear data (*Campisi-Pinto, Adamowski & Oron, 2012*).

243   The Multi-Layer Perceptron (MLP) which is the most popular sub-model of NN (*Talaee, 2014;*

244   *Ali et al., 2017*) consists of three layers of nodes used here for the prediction of denoised IMFs of

245   river flow data. The complete layout of MLP is given in (*Talaee, 2014*). The structure of MLP

246   comprised of NN structure. Three nodes are included in MLP as an input layer, hidden layer, and

247   an output layer. First, the single output is calculated by using linear combinations of inputs which

248   are further transferred to some non-linear activation functions mathematically defined as by

$$y = \varphi\left(\sum_{i}^{n} w_i x_i + b\right) \tag{15}$$

249   where w_i are the weights of inputs, x_i are the inputs, b is the biased value for each layer and φ is

250   a non-linear activation function which supplies output to the next layer. The mostly used activation

251   function i.e. logistic function is used as activation function defined as follows:

$$\frac{1}{(1 + e^{-\sum_{i}^{n} w_i x_i + b})} \tag{16}$$

252   For the optimization of neurons, the supervised learning algorithms called back-propagation and

253   forward-propagation can be used.

254

255   *The noise-free IMF prediction with ARIMA model:* To predict the noise-free IMF's and residual,

256   an autoregressive moving average model (ARMA) is selected which is described as follows:

$$IMF_t^{\,k} = \alpha_1 IMF_{t-1}^{\,k} + \ldots + \alpha_p IMF_{t-p}^{\,k} + \varepsilon_t^{\,k} + \beta_1 \varepsilon_{t-1}^{\,k} + \ldots + \beta_q \varepsilon_{t-q}^{\,k} \tag{17}$$

257   where, $IMF_t^{\,k}$ is the kth IMF, $IMF_{t-p}^{\,k}$ is p^{th} lag value of kth IMF $\varepsilon_t^{\,k}$ is the residual of kth IMF, ,

258   p and q are autoregressive and moving average lags. Moreover, in some cases time series data is

259   not stationary. To make such stationary, differences at an appropriate degree are used (*Box &*

260   *Jenkins, 2012*). If such a situation occurs, then the model is known as ARIMA (p, d, q) where d

261   is the differenced value used to make the non-stationary data stationary.

262   **3   Case study and experimental design**

263   *Selection of Study Area*

264   In this research, the largest water system of Pakistan is considered for the application of the

265   proposed strategy. The Indus Basin System (IBS) is the largest river of Pakistan and it plays an

266   important role specifically in power generation and irrigation department. The major tributaries of

267   IBS i.e. River Jhelum, River Chenab, and River Kabul are selected for the present study. Pakistan

268   is facing a large amount of frequent river flooding each year due to monsoon rain and melting

269   snow or glacier. As in Pakistan, glacier-covered 13,680 km^2 area which is estimated 13% of the

270   mountainous areas of Upper Indus Basin (UIB). Melted water from these 13% areas adds the

271   significant contribution of water in these rivers. which leads to complex characteristics in river

272   flow data. Therefore, for sustainable economic development and efficient water resources

273   planning, there is a need for analyzing such complex characteristics and predict the behaviors of

274   river flow data at IBS and its tributaries.

275

276   *Data*

277   To investigate the improved framework, four daily river flow data comprised on (1st-January to

278   19th-June) for the period of 2015-2018 is used in this study. The main river flow of Indus at Tarbela

279   is considered with its three principal tributaries: Jhelum at Mangla, Chenab at Marala and Kabul

280   at Nowshera. Data is measured in 1000 ft/s. The described river flow data was obtained from the

281   website of Pakistan Water and Power Development Authority (WAPDA).

282   ***Evaluation criteria***

283   *Evaluation of noise reduction methods:* The performances of denoised series needs comprehensive

284   evaluation after using appropriate noise reduction threshold methods. In this research, to check the

285   performances of CEEMDAN/EEMD-ST CEEMDAN/EEMD-HT, CEEMDAN/EEMD-ITF, and

286   proposed framework i.e. CEEMDAN/EEMD-EBT, Signal-to-Noise Ratio (SNR), Mean Square

287   Error (MSE) and Mean Absolute Error (MAE) (*Nazir et al., 2019*) are employed which are given

288   as follows respectively;

$$SNR = 10log_{10}(\frac{\sum_{t=1}^{N}(y_{ot})^2}{\sum_{t=1}^{N}(y_{pt} - y_{ot})^2}) \qquad (18)$$

$$MSE = \sqrt{\frac{\sum_{t=1}^{N}(y_{ot} - y_{pt})^2}{N}} \qquad (19)$$

289

290   and

$$MAE = \frac{\sum_{t=1}^{N}|y_{ot} - y_{pt}|}{N} \qquad (20)$$

291

292   where y_{ot} is the tth observed value and y_{pt} is the tth predicted value. Moreover, the performance

293   of proposed model i.e. CEEMDAN-EBT-MM and all other models including CEEMDAN/EEMD-

294   ST-MM, CEEMDAN/EEMD-HT-MM, CEEMDAN/EEM-ITF-MM and EEMD-ITF-MM are

295   compared by using three popular statistical measures: MSE, MAE defined in Eq. (19-20) and Mean

296   Absolute Percentage Error (MAPE) given as follows;

297

$$MAPE = \frac{|y_{mo} - y_{mp}|}{|y_{mo}|} * 100 \qquad (21)$$

298

299   where, y_{ot} and y_{pt} is defined above, y_{mo} is the mean value of observed values, y_{mp} is the mean

300   value of predicted values and y_{sp} is the standard deviation of predicted values.

301

302   **4   Results**

303 This section provides results of proposed CEEMDAN-EBT-MM model and benchmark models

304 i.e. CEEMDAN/EEMD-ST-MM, CEEMDAN/EEMD-HT-MM, CEEMDAN/EEM-ITF-MM, and

305 EEMD-ITF-MM in steps as follows:

306

307 *Results of decomposition stage:* First, to confirm the non-stationarity of river flow data,

308 Augmented Dickey Fuller (ADF) unit root test (*Khalili et al., 2013*) is used for all selected case

309 studies. The test is applied to data by taking the log in order to confirm non-stationarity. Results

310 of ADF showed that all selected river flow data i.e. Indus river flow, Jhelum river flow, Chenab

311 river flow, and Kabul river flow is significantly non-stationary with p-value as 0.3353, 0.4135,

312 0.333 and 0.414 respectively. Then, the non-linear and non-stationary data decomposed into

313 different time scale oscillation called IMFs to reduce the non-stationarity by extracting the time-

314 varying characteristics of daily river flow data from all selected four stations. The CEEMDAN

315 decomposition technique is used to extract IMF's of river flow data. All selected river flow data is

316 decomposed into thirteen IMFs and one residual. The starting IMF's represents the highest

317 frequencies whereas the last half IMF's showed the low frequencies and residual represents the

318 overall trend. The decomposed results of Indus River system is depicted in Fig. 2. The amplitude

319 of white noise is set 0.2 as in (*Di, Yang & Wang, 2014*) and numbers of ensemble members are

320 settled as 1000. By inspecting the IMFs, it is noticed that each IMF component represents the

321 oscillation characteristic Further, the cross-correlation method is employed to find the noisiest

322 IMFs from all thirteen IMFs. To do this, first, the decomposed IMFs are further divided into two

323 parts to find the noisiest IMFs by using the cross-correlation between IMFs and original river flow

324 data. The low correlation implies the more uncertainty present in IMFs. The first ten IMFs showed

325 the least correlation with original river flow data indicated that these IMFs are overwhelmed with

326 noises. The graph of cross-correlation between Indus river flow data and first IMF and between

327 Indus river flow and eleventh IMF is depicted in left and right corner of Fig. 3, which shows that

328 the first IMF is filled with noises with low correlation at all lags and eleventh IMF is free from

329 noises as it showed 0.75 correlation not only at lag zero but also at other lag values. For all four

330 rivers, first tenth IMFs are characterized as noisiest IMFs and last three IMFs are labeled as noise

331 free IMFs.

332 *Results of denoising stage:* the next step is to denoise the noisiest and sparse IMFs.

333 .

334   To eliminate noises from IMFs, EBT estimator is used as a filter which assumes a mixture of prior

335   as defined in Eq. (10) for each IMF separately by considering the nature of IMFs. First, the scaled

336   transformation is applied to get normal distribution so that each IMF follows $N(\theta_i, 1)$. According

337   to the nature of IMFs as depicted in Fig. 4, it is known that most of the coefficients in all IMFs are

338   zero and some are non-zero in which fewer coefficients are either very low or high. By looking

339   (Figure 4) both zero and non-zero part of IMFs, a mixture of an atom of probability at zero and

340   multiple distributions are considered for non-zero part (*Johnstone & Silverman, 2005)*. Among all

341   of them, Laplace distribution is configured out as prior distribution of θ_i with maximum SNR.

342   Finally, the important coefficients of IMFs are preserved with posterior median threshold estimator

343   described in Eq. (11) by attaining highest SNR value with minimum MSE and MAE given in

344   Table 1 for all selected case studies. For comparison purpose, the conventional denoising methods

345   as ST, HT and ITF are implemented on all river flow data. We observed that SNR of CEEMDAN-

346   EBT based decomposed and denoised method is larger than all other CEEMDAN/EEMD-ST,

347   CEEMDAN/EEMD-HT and CEEMDAN/EEMD-ITF based methods as they don't consider the

348   sparsity and magnitude of noises separately to remove noises from data. The reconstructed

349   denoised graph of   proposed   CEEMDAN/EEMD-EBT   and   bench   mark   models

350   CEEMDAN/EEMD-HT,CEEMDAN/EEMD-ST and CEEMDAN/EEMD-ITF for Indus and

351   Jhelum are shown in Fig. 5/Fig. 6 respectively. From Fig. 5 and 6, it is shown that

352   CEEMDAN/EEMD-ST over estimated noises for Jhelum, Chenab and Kabul river flow and, the

353   performance of CEEMDAN/EEMD ITF is worst for Indus river flow however, the proposed

354   CEEMDAN-EBT based model shown optimal performance for all case studies.

355   *Results of prediction stage:* The decomposed and denoised IMFs of all selected case studies are

356   further predicted through data-driven and statistical model. The denoised IMFs are predicted

357   through MLP-neural network model. Training is performed by using forward and back-

358   propagation by setting the learning rate parameter between 0.1 to 1. The back-propagation method

359   with the optimal learning rate is selected to test model. The second part of decomposed IMFs

360   comprised of noise-free IMFs and residual which are predicted through simple traditional

361   statistical model i.e. ARIMA (p, d, q) for all case studies. The river flow data of all four rivers are

362   splitted, 70%, 80% and 90% for the training set and 30%, 20% and 10% for testing set. The results

363   achieved by such splitting criteria are not much deviated from each other so the only values of

364   80% training errors are given in Table 2. After a successful estimation of each IMF and residual,

365   the accuracy is measured with MAE, MAPE, and MSE. The training results of proposed models

366   with a comparison to all other models for all four river flow i.e. Indus flow, Jhelum flow, Chenab

367   flow, and Kabul flow are presented in Table 2. The results of the proposed model i.e. CEEMDAN-

368   EBT-MM, demonstrate its effectiveness by showing minimum MAD, MAPE and RMSE values

369   for Indus river flow. For Jhelum and Chenab, the proposed CEEMDAN-EBT-MM model shown

370   least MAE, and MAPE values and by attaining lower MAPE value for Kabul river flow

371   comparative to all other methods. The predicted graph of proposed CEEMDAN-EBT-MM model

372   and EEMD-EBT-MM with their benchmark models for Indus and Jhelum river flow are shown in

373   Fig. 7 and Fig.8 respectively.

374   ***Discussion***

375   *Decomposition and denoising results:* In order to understand the applicability of our proposed

376   CEEMDAN-EBT model the other EEMD based decomposition method and three different

377   denoising methods i.e. EEMD-HT, EEMD-ST and EEMD-ITF are employed. The evaluation of

378   denoising of proposed CEEMDAN-EBT and benchmark models are carried out using SNR, MSE

379   and MAE measures for all river flow data. From Table 1, it is clear that overall decomposition

380   based on CEEMDAN performs well with improved denoising method i.e. CEEMDAN-EBT that

381   efficiently eliminating noises by considering the mixture of priors for IMFs with highest SNR

382   value and lowest MSE and MAE values, however other existing denoising methods i.e.

383   CEEMDAN-HT, CEEMDAN-ST, and CEEMDAN-ITF performed low with low SNR value and

384   high MSE values. Moreover, comparative to CEEMDAN, the other EEMD based decomposition

385   and denoising methods i.e. EEMD-EBT, EEMD-ST, EEMD-HT, and EEMD- ITF for all case

386   studies are also performed poor as the EEMD based decomposition exhibits clear mode mixing as

387   shown in Fig. 6, where results for Indus and Jhelum river data are plotted. The MSE and MAE

388   values of EEMD-EBT, EEMD-ST, EEMD-HT, and EEMD- ITF methods having SNR values are

389   very low and MSE and MAE values high due to poor decomposition and denoising methods for

390   all four rivers data which implies that CEEMDAN has the ability to optimally extract the IMFs

391   which are further processed with optimal denoising methods to get smooth noise free IMFs as

392   shown in Fig. 5, where denoised results of Indus and Jhelum river flow are plotted. From the results

393   shown in Table 1, it is concluded that specifically our proposed strategy i.e. CEEMDAN-EBT

394 performs better than that of all other methods i.e. CEEMDAN/EEMD-ST, CEEMDAN/EEMD-

395 HT and CEEMDAN/EEMD-ITF and EEMD-EBT based decomposition and denoising methods.

396 *Final prediction model:* To verify the superiority of proposed i.e. CEEMDAN-EBT-MM strategy

397 to model the complex river flow data, we choose the CEEMDAN/EEMD-ST-MM,

398 CEEMDAN/EEMD-HT, CEEMDAN/EEMD-ITF AND EEMD-EBT-MM models to analyze the

399 prediction results of non-linear and noise-corrupted data. Our proposed prediction framework

400 based on decomposition and novel strategy of denoising performs well as compared to all other

401 decomposition and denoising methods. As shown in Table 2 and Fig. 7, the proposed model i.e.

402 CEEMDAN-EBT-MM exhibits a good prediction performance for Indus river flow with least

403 MAE, MSE and MAPE values. The MAE and MAPE values of Jhelum and Chenab river flow are

404 also better than the other models. For Kabul river flow, our proposed model showed its efficiency

405 with the least MAPE value. The other models are not consistent in terms of efficiency, their

406 prediction behaviors for each river vary as CEEMDAN-HT-MM shows effective performance for

407 Indus river flow but poor in predicting Kabul river flow. However, overall the proposed

408 CEEMDAN-EBT-MM model shows consistent and excellence prediction results which indicate

409 that with appropriate decomposition and the novel improvement of denoising technique, one can

410 overall improve the performance of existing data-driven models to handle the river flow data.

411 In this study, as the river flow of selected Indus River system exhibit seasonal persistence at

412 several multi-scales which manifested, through data decomposition method, and used to enhance

413 the prediction accuracy of river flow. Moreover, instead of overall season, Kalhoro et al. (2017)

414 shown that there is a further sub-seasonal variation i.e. high flow, called Kharif season, and low

415 flow, called Rabi season, present in Indus River system. Majority of Pakistan's population lives

416 in rural areas, having the occupation of farming, which are directly influenced from the severity

417 of both sub-seasonal variation. In future, there is a need to establish reliable sub-season

418 identification and their prediction methods that will be fruitful for efficient water allocation in

419 both seasons which can ultimately boost the economy of Pakistan.

5   Conclusion

421 In this paper, due to the non-linearity and noises complexity of the river flow data, we proposed a

422 strategy to improve the prediction of data-driven models with appropriate decomposition and a

423 novel denoising method. Our proposed denoising method improves the performance of

424 CEEMDAN based decomposition method by improving the time-scale components which
425 enhance the prediction accuracy of data-driven models. The proposed method comprised of four
426 steps such as decomposition, denoising and prediction, and aggregation. The performance of the
427 proposed CEEMDAN-EBT-MM model is evaluated using four daily river flow data of IBS. The
428 CEEMDAN-EBT-MM having the smallest MAPE values for all four case studies compared to
429 other benchmark models. The improved results suggest that the proposed hybrid model can be
430 used as an efficient tool for providing a reliable prediction of river flow data to policymakers for
431 planning power generation and water resource management.

432

433 **References**

434 Aichouri I, Hani A, Bougherira N, Djabri L, Chaffai H, Lallahem S. 2015. River Flow Model
435   Using Artificial Neural Networks. *Energy Procedia* 74:1007–14.

436 Ali GM, Kazempour R, Chau KW, Shamshirband S, Taherei Ghazvinei P. 2018. Forecasting pan
437   evaporation with an integrated artificial neural network quantum-behaved particle swarm
438   optimization model: A case study in Talesh, Northern Iran. *Engineering Applications of*
439   *Computational Fluid Mechanics* 12(1):724-737.

440 AlMasudi RK. 2018. Fitting ARIMA Models for Forecasting to Inflow of Dokan Reservoir.
441   *Journal of Babylon University*, 19(4).

442 Ali Z, Hussain I, Faisal M, Nazir HM, Hussain T, Shad MY, Mohamd Shoukry A, Hussain Gani
443   S. 2017. Forecasting drought using multilayer perceptron artificial neural network model."
444   *Advances in Meteorology*.

445 Antico A, Schlotthauer G, Torres ME. 2014. Analysis of Hydroclimatic Variability and Trends
446   Using a Novel Empirical Mode Decomposition: Application to the Paraná River Basin.
447   *Journal of Geophysical Research* 119(3):1218–33.

448 Azadeh A, Sheikhalishahi M, Tabesh M, Negahban A. 2011. The Effects of Pre-Processing
449   Methods on Forecasting Improvement of Artificial Neural Networks. *Australian Journal of*
450   *Basic and Applied Sciences* 5(6): 570–80.

451   Box GE, Jenkins GM. 1970. Time series analysis: forecasting and control. *San Francisco: Holden-*
452   *Day.*

453   Campisi-Pinto S, Adamowski J, Oron G. 2012. Forecasting Urban Water Demand Via Wavelet-
454   Denoising and Neural Network Models. Case Study: City of Syracuse, Italy." *Water*
455   *Resources Management* 26(12):3539–58.

456   Candes EJ, Sing-Long CA, Trzasko JD. 2012. Unbiased Risk Estimates for Singular Value
457   Thresholding and Spectral Estimators. *IEEE transactions on signal processing*, 61(19):4643-
458   4657.

459   Ce J, Xue-liang L, Bao-cheng T. 2012. An Improved Wavelet Threshold De-Noising Method and
460   Its Application. *In Proceedings of the 2012, 24th Chinese Control and Decision Conference,*
461   *CCDC 950-953.*

462   Di C, Yang X, Wang X. 2014. A Four-Stage Hybrid Model for Hydrological Time Series
463   Forecasting. *PloS one* 9(8).

464   Gjika E, Aurora F, Arbesa K. 2019. A Study on the Efficiency of Hybrid Models in Forecasting
465   Precipitations and Water Inflow Albania Case Study. *Advances in Sciences, Technology and*
466   *Engineering System Journals* 4(1):302-310.

467   Holzfuss J, Kadtke J. 1993. Global nonlinear noise reduction using radial basis functions.
468   *International Journal of Bifurcation and Chaos*, 3(03):589-596.

469   Hadi SJ, Tombul M. 2018. Streamflow Forecasting Using Four Wavelet Transformation
470   Combinations Approaches with Data-Driven Models: A Comparative Study. *Water*
471   *Resources Management* 32(14):4661–79.

472   Han M, Liu Y. 2009. Noise Reduction Method for Chaotic Signals Based on Dual-Wavelet and
473   Spatial Correlation. *Expert Systems with Applications* 36(6):10060–67.

474   Hansen NR. 2017. A Comment on Stein's Unbiased Risk Estimate for Reduced Rank Estimators.
475   *arXiv preprint arXiv:1708.09657.*

476   Hossain MM, Kozubowski TJ, Podgórski K. 2018. A novel weighted likelihood estimation with

477   empirical Bayes flavor. *Communications in Statistics-Simulation and Computation*, *47*(2),
478   392-412.

479   Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH. 1998. The
480   Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary
481   Time Series Analysis. *Proceedings of the Royal Society of London. Series A: Mathematical,*
482   *Physical and Engineering Sciences*, 454(1971):903-995.

483   Jansen M, Bultheel A. 2014. Empirical Bayes Approach to Improve Wavelet Thresholding for
484   Image Noise Reduction. *Journal of the American Statistical Association*, 96(454):629-639.

485   Jeng Y, Lin MJ, Chen CS, Wang YH. 2007. Noise Reduction and Data Recovery for a VLF-EM
486   Survey Using a Nonlinear Decomposition Method. *Geophysics* 72(5), 223-235.

487   Jiao G, Guo T, Ding Y. 2016. A New Hybrid Forecasting Approach Applied to Hydrological Data:
488   A Case Study on Precipitation in Northwestern China. *Water (Switzerland)* 8(9):365.

489   Johnstone IM, Silverman BW. 2005. Empirical Bayes Selection of Wavelet Thresholds. *Annals of*
490   *Statistics* 33(4):1700–1752.

491   Kalhoro NA, He Z, Xu D, Inam A, Muhammad F, Sohoo N. 2017. Seasonal variation of
492   oceanographic processes in Indus river estuary. *MAUSAM*, 68(4): 643-654.

493   Kang A, Tan Q, Yuan X, Lei X, Yuan Y. 2017. Short-Term Wind Speed Prediction Using EEMD-
494   LSSVM Model. *Advances in Meteorology*.

495   Khalili K, Ahmadi F, Dinpashoh Y, Fard AF. 2013. Determination of Climate Changes on
496   Streamflow Process in the West of Lake Urmia with Used to Trend and Stationarity Analysis.
497   *International Journal of Advanced Biological and Biomedical Research* 1(10):1220–35.

498   Wu Z, Huang NE. 2004. A Study of the Characteristics of White Noise Using the Empirical Mode
499   Decomposition Method. *Proceedings of the Royal Society of London. Series A:*
500   *Mathematical, Physical and Engineering Sciences* 460(2046):1597–1611.

501   Nazir HM, Hussain I, Faisal M, Shoukry AM, Gani S, Ahmad I. 2019. Development of
502   Multidecomposition Hybrid Model for Hydrological Time Series Analysis. *Complexity*.

503 Peng T, Zhou J, Zhang C, Fu W. 2017. Streamflow Forecasting Using Empirical Wavelet
504   Transform and Artificial Neural Networks. *Water* 9(6):1–20.

505 Shathir AK, Saleh LA. 2016. Best ARIMA Models for Forecasting Inflow of Hit Station. *Basrah*
506   *Journal for Engineering Science*, 16(1):62-72.

507 Santos CA, Freire PK, Silva RM, Akrami SA. 2019. Hybrid Wavelet Neural Network Approach
508   for Daily Inflow Forecasting Using Tropical Rainfall Measuring Mission Data. 24(2)
509   04018062.

510 Su H, Li H, Chen Z, Wen Z. 2016. An Approach Using Ensemble Empirical Mode Decomposition
511   to Remove Noise from Prototypical Observations on Dam Safety. *SpringerPlus* 5(1):1–14.

512 Talaee PH. 2014. Multilayer perceptron with different training algorithms for streamflow
513   forecasting. *Neural Computing and Applications*. 24(3-4):695-703.

514 Torres ME, Colominas MA, Schlotthauer G, Flandrin P. 2011. A complete ensemble empirical
515   mode decomposition with adaptive noise. *IEEE international conference on acoustics, speech*
516   *and signal processing (ICASSP)* 4144–4147.

517 Toth E, Brath A, Montanari A. 2000. Comparison of Short-Term Rainfall Prediction Models for
518   Real-Time Flood Forecasting. *Journal of Hydrology 239*(1-4), 132-147.

519 Wang WC, Chau KW, Xu DM, Chen XY. 2015. Improving Forecasting Accuracy of Annual
520   Runoff Time Series Using ARIMA Based on EEMD Decomposition. *Water Resources*
521   *Management* 29(8):2655–75.

522 Wei X, Lin R, Liu S, Zhang C. 2016. Improved EEMD Denoising Method Based on Singular
523   Value Decomposition for the Chaotic Signal. *Shock and Vibration*.

524 Wu CL, Chau KW. 2011. Rainfall − Runoff Modeling Using Artificial Neural Network Coupled
525   with Singular Spectrum Analysis. 399(3-4), 394-409.

526 Yaseen ZM, Fu M, Wang C, Mohtar WH, Deo RC, El-Shafie A. 2018. Application of the Hybrid
527   Artificial Neural Network Coupled with Rolling Mechanism and Grey Model Algorithms for
528   Streamflow Forecasting Over Multiple Time Horizons. *Water Resources Management 32*(5),

529   1883-1899.

530   Yaseen ZM, Sulaiman SO, Deo RC, Chau KW. 2018. An Enhanced Extreme Learning Machine
531   Model for River Flow Forecasting: State-of-the-Art, Practical Applications in Water
532   Resource Engineering Area and Future Research Direction. *Journal of Hydrology*.

533

534

Table 1

The statistical measures of proposed CEEMDAN-EBT and existing denoising methods for all case studies.

River Inflow	Method	SNR	MSE	MAE
Indus Inflow	CEEMDAN-HT	15.4996	0.9924	0.7905
	CEEMDAN-ST	-18.4131	3.1607	1.3692
	CEEMDAN-ITF	-15.7423	3.1645	1.3713
	CEEMDAN-EBT	**22.0440**	**1.03563**	**0.8741**
	EEMD-HT	-19.32131	49.8554	5.6131
	EEMD-ST	-36.7393	36.1929	4.4431
	EEMD-ITF	-36.4253	36.1766	4.4442
	EEMD-EBT	-14.9984	49.31019	5.6353
Jhelum Inflow	CEEMDAN-HT	4.0185	0.6706	0.7905
	CEEMDAN-ST	-21.0096	2.4233	1.1978
	CEEMDAN-ITF	-20.4669	2.4234	1.1985
	CEEMDAN-EBT	**20.7036**	**0.0012**	**0.0142**
	EEMD-HT	-12.7805	3.8168	1.5496
	EEMD-ST	-28.2531	5.7458	1.7912
	EEMD-ITF	-28.0296	5.7463	1.7927
	EEMD-EBT	9.7364	3.2357	1.4390
Chenab Inflow	CEEMDAN-HT	-4.8656	0.3252	0.4230
	CEEMDAN-ST	-17.3340	3.1607	1.3692
	CEEMDAN-ITF	-16.4194	1.1085	0.7870
	CEEMDAN-EBT	**31.6470**	**0.0004**	**0.0165**
	EEMD-HT	-8.1820	3.8706	1.5152
	EEMD-ST	-29.2759	4.9703	1.6133
	EEMD-ITF	-29.0425	4.9693	1.6148
	EEMD-EBT	-12.8752	3.5997	1.5085
Kabul Inflow	CEEMDAN-HT	10.8756	0.4823	0.4986
	CEEMDAN-ST	-22.289	2.2647	1.1272
	CEEMDAN-ITF	-21.6878	2.2649	1.1280
	CEEMDAN-EBT	**31.9856**	**0.0003**	**0.0144**
	EEMD-HT	-2.4276	5.4761	1.7961
	EEMD-ST	-32.0408	8.5625	2.0934
	EEMD-ITF	-31.8501	8.5601	2.0949
	EEMD-EBT	-10.2153	5.1512	1.7999

2   Table 1: The statistical measures of proposed CEEMDAN-EBT and existing denoising methods

3   for all case studies.

Table 2

Evaluation index of the testing prediction error of proposed models (CEEMDAN-EBT-MM) with all selected models for all four case studies.

Rivers	Method	MAE	MAPE	MSE
Indus River Inflow	CEEMDAN-HT-MM	9.5938	0.1317	13.7910
	CEEMDAN-ST-MM	8.7531	0.1413	17.2590
	CEEMDAN-ITF-MM	8.4614	0.1349	14.7238
	CEEMDAN-EBT-MM	**5.6837**	**0.1071**	**9.8704**
	EEMD-HT-MM	16.7567	5.0868	18.7316
	EEMD-ST-MM	12.7899	0.1480	17.2683
	EEMD-ITF-MM	12.7459	0.1648	17.7453
	EEMD-EBT-MM	17.4940	0.1338	15.4142
Jhelum River Inflow	CEEMDAN-HT-MM	6.9537	0.3502	0.0547
	CEEMDAN-ST-MM	6.0630	0.0577	0.3912
	CEEMDAN-ITF-MM	6.0558	0.0407	0.1949
	CEEMDAN-EBT-MM	**6.0185**	**0.0400**	0.2609
	EEMD-HT-MM	7.1364	0.1157	1.5763
	EEMD-ST-MM	8.8392	0.1227	1.7755
	EEMD-ITF-MM	6.4300	0.1451	2.4853
	EEMD-EBT-MM	7.2961	0.0414	0.2015
Chenab River Inflow	CEEMDAN-HT-MM	6.6962	0.0115	0.0140
	CEEMDAN-ST-MM	6.2191	0.0110	0.0141
	CEEMDAN-ITF-MM	6.2198	0.0217	0.0504
	CEEMDAN-EBT-MM	**6.0027**	**0.0025**	0.1934
	EEMD-HT-MM	6.9454	0.0818	0.7159
	EEMD-ST-MM	5.3718	1.0425	115.923
	EEMD-ITF-MM	6.5058	0.3943	16.678
	EEMD-EBT-MM	7.0002	0.0196	0.0410
Kabul River Inflow	CEEMDAN-HT-MM	8.5972	0.2274	7.0561
	CEEMDAN-ST-MM	7.8189	0.1676	3.8311
	CEEMDAN-ITF-MM	7.8192	0.2328	3.3163
	CEEMDAN-EBT-MM	8.5971	**0.1660**	3.7578
	EEMD-HT-MM	8.7354	0.2093	5.9963
	EEMD-ST-MM	6.6151	0.2384	0.2028
	EEMD-ITF-MM	6.6042	0.1864	0.2881
	EEMD-EBT-MM	9.0280	0.1878	4.8166

2 Table 2: Evaluation index of testing prediction error of proposed models (CEEMDAN-EBT-MM)
3 with all selected models for all four case studies.

Figure 1

The proposed CEEMDAN-EBT-MM structure to predict non-linear rivers inflow data.
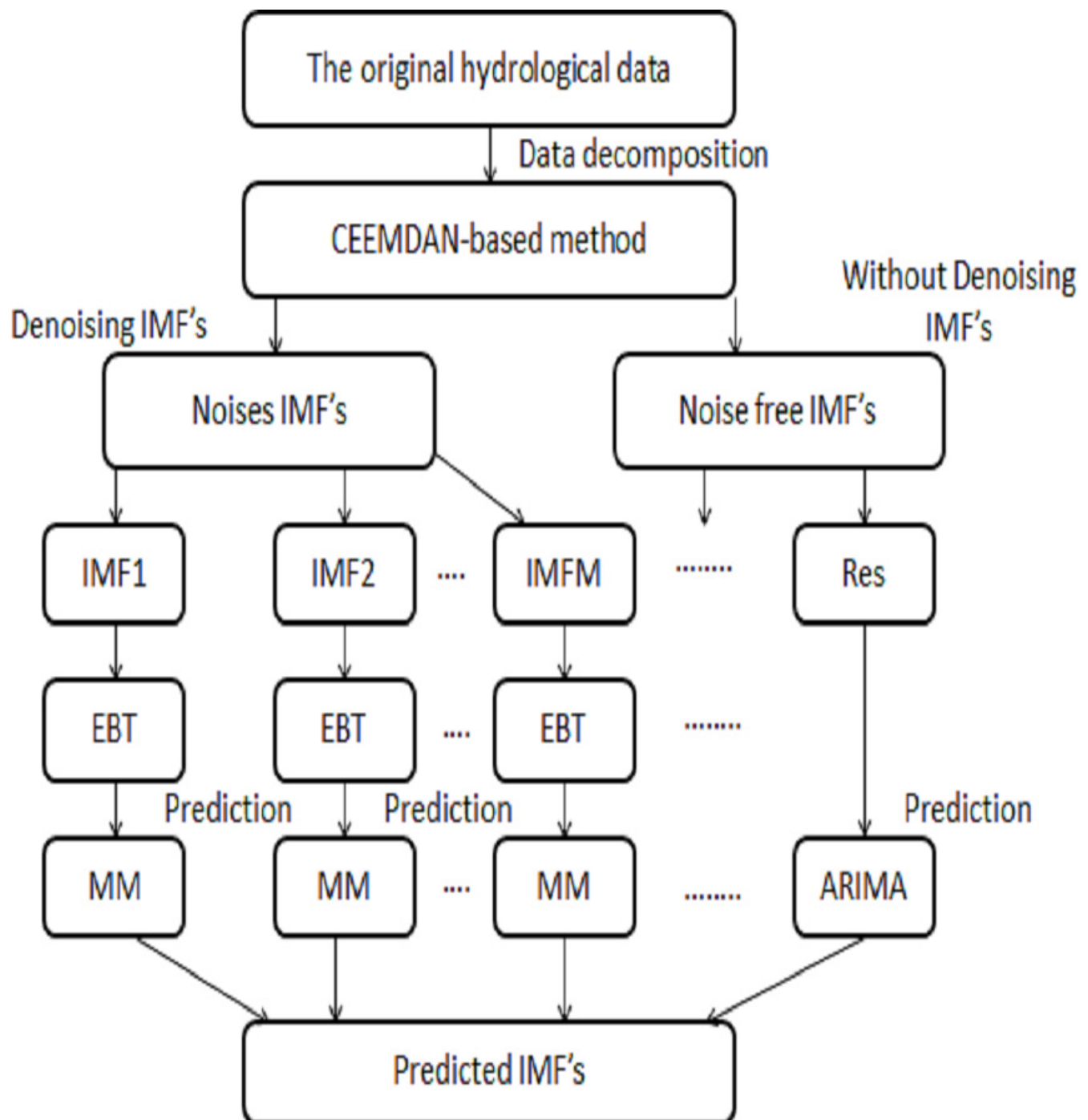
Figure 2

The CEEMDAN based decomposition of rivers inflow data.

The graph (A) shows the highest frequencies and the graph (B) shows the lower frequencies and graph (C) shows irregular variations with the trend.
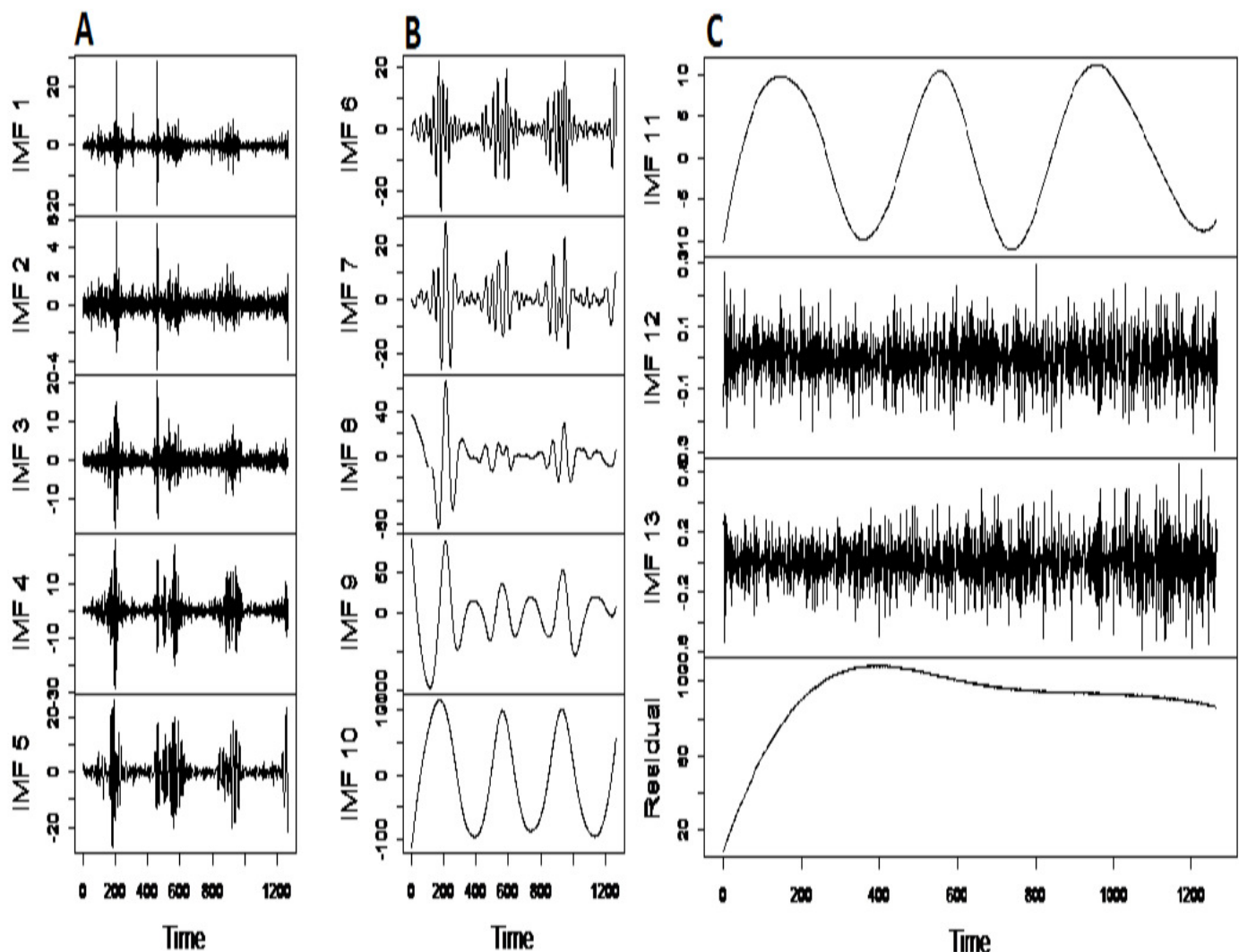
Figure 3

The cross-correlational graph

The plot (A) shows the cross-correlation between first IMF and original Indus rivers inflow data and plot (B) shows the cross-correlation between eleventh IMF and original Indus river inflow.
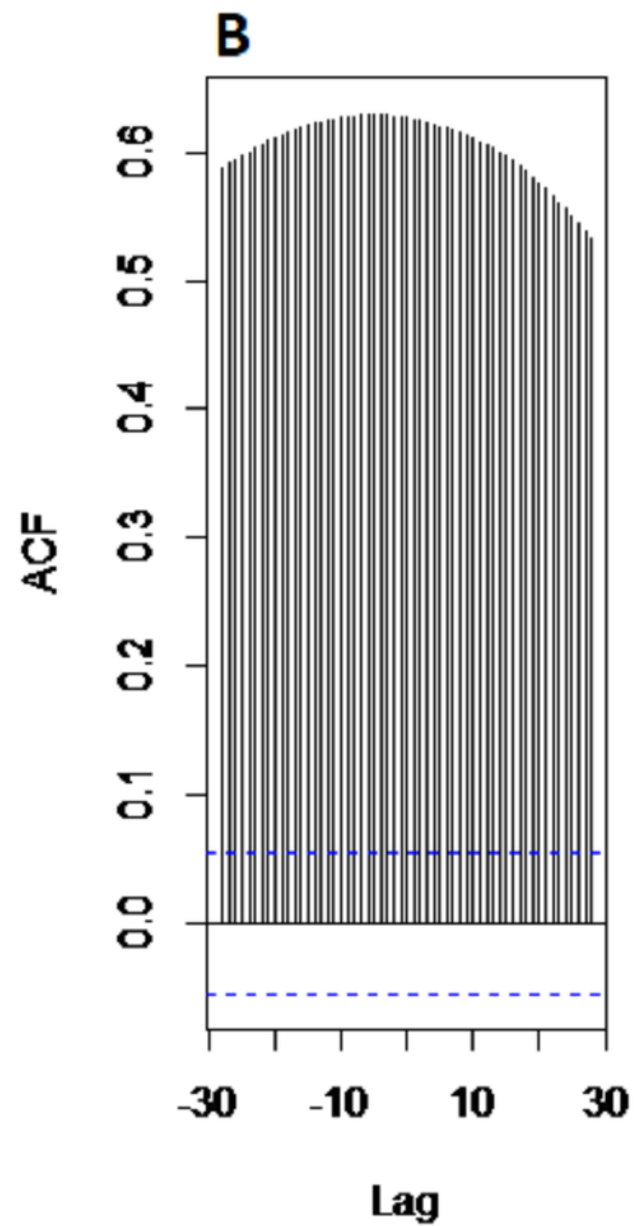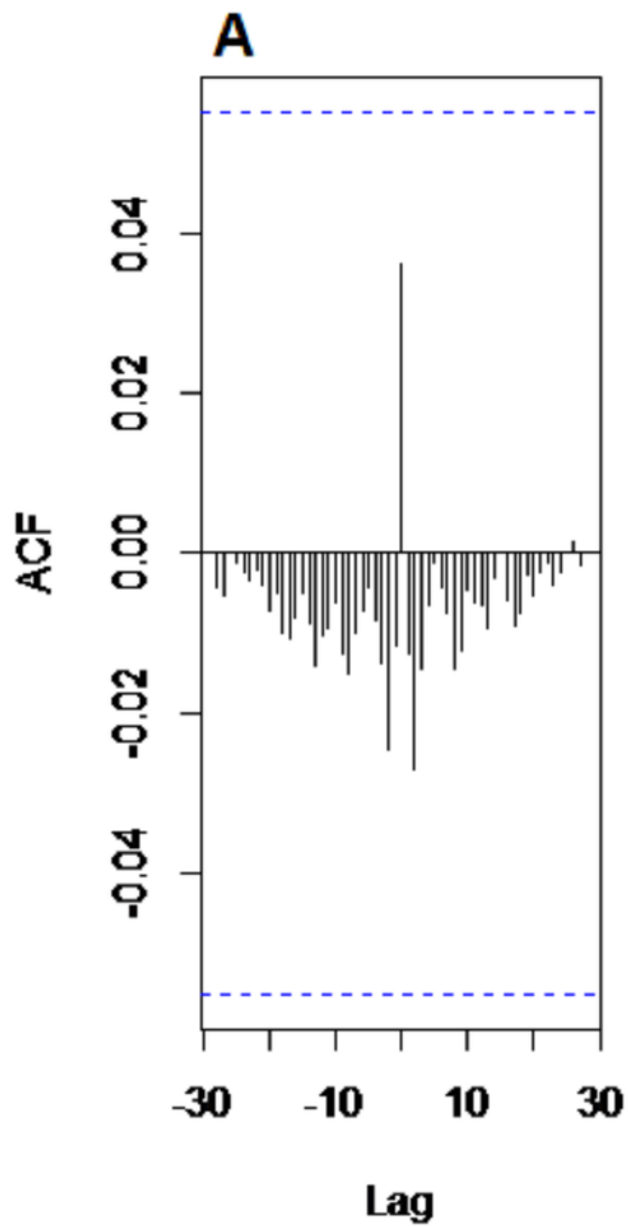
Figure 4

The CEEMDAN based first two IMFs for Indus river inflow.

The CEEMDAN based first two IMFs (A and B) for Indus river inflow.
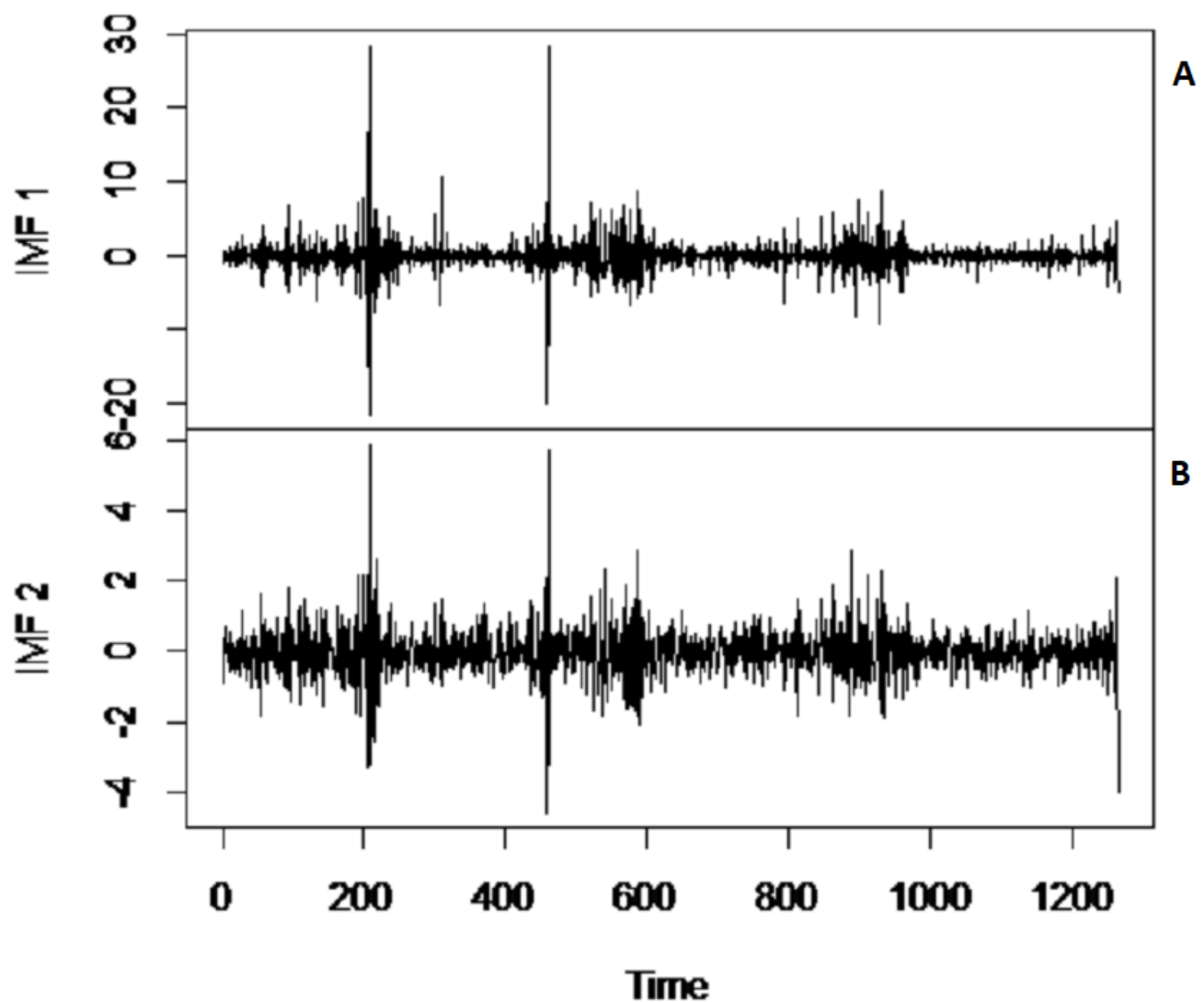
Figure 5

The CEEMDAN based decomposed and denoised series of proposed EBT and existing ST, HT and ITF of Indus (left) and Jhelum (right) rivers inflow.

The CEEMDAN based decomposed and denoised series of proposed EBT and existing ST, HT and ITF of Indus (A) and Jhelum (B) rivers inflow.
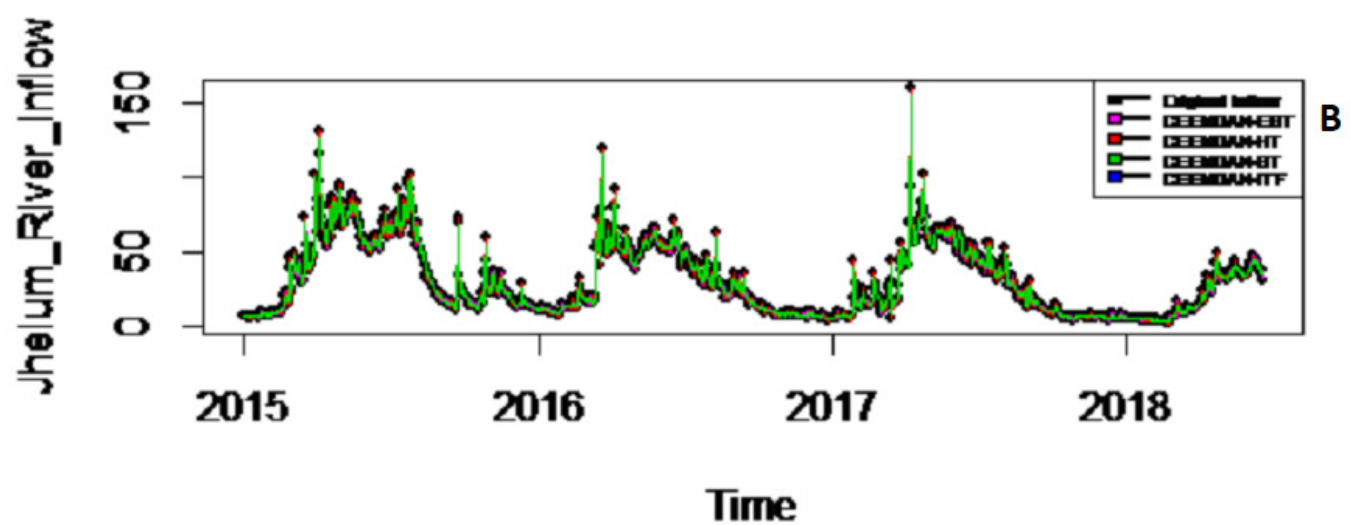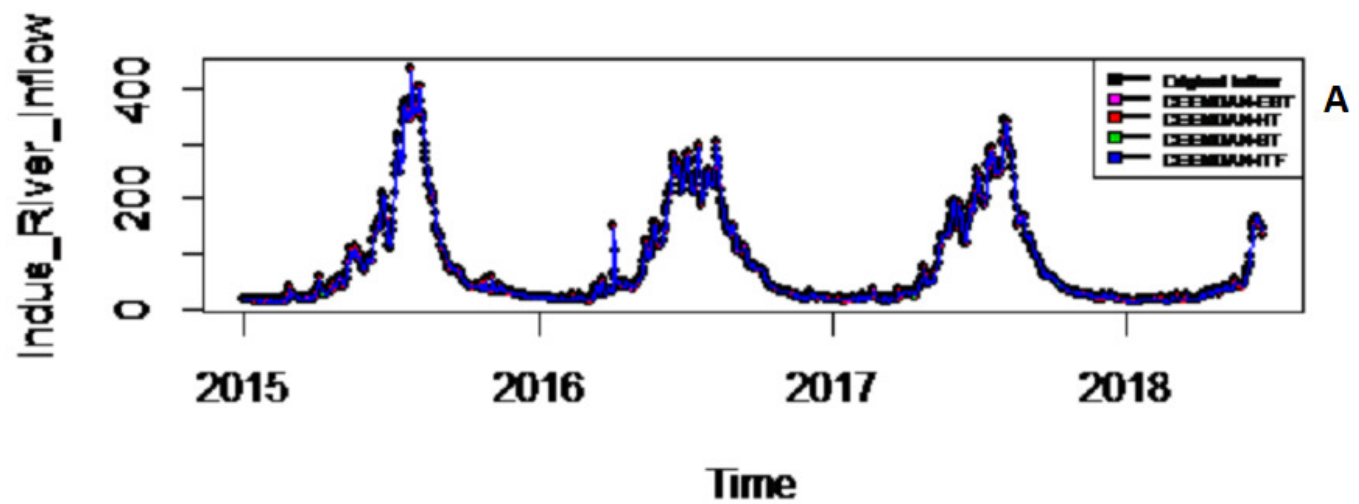
Figure 6

The EEMD based decomposed and denoised series of Indus (A) and Jhelum (B) rivers inflow.

The EEMD based decomposed and denoised series of Indus (A) and Jhelum (B) rivers inflow.
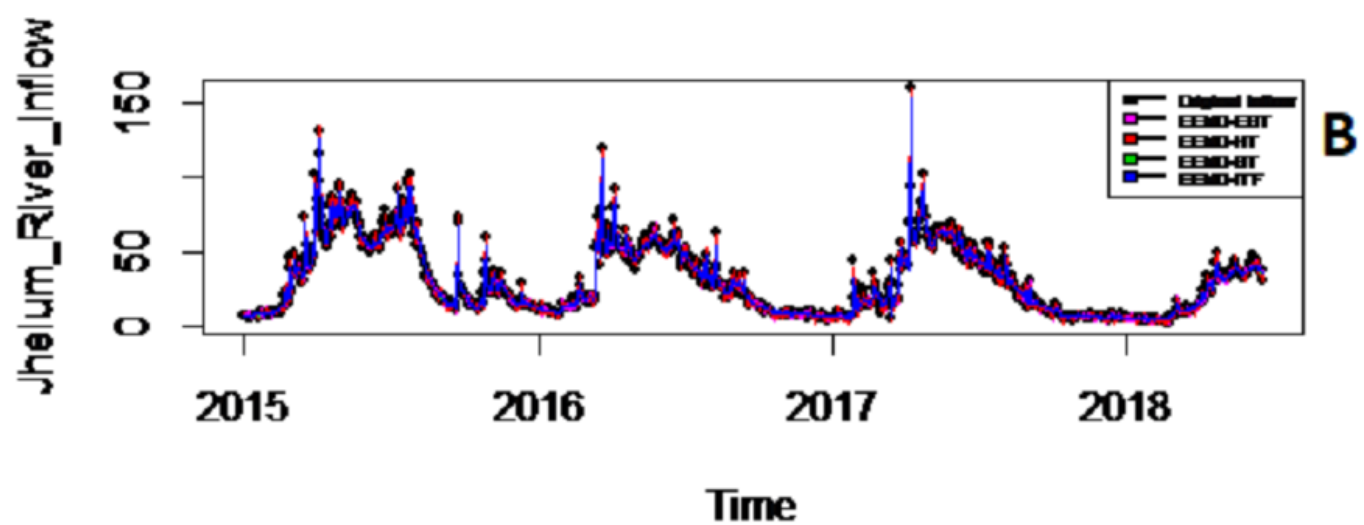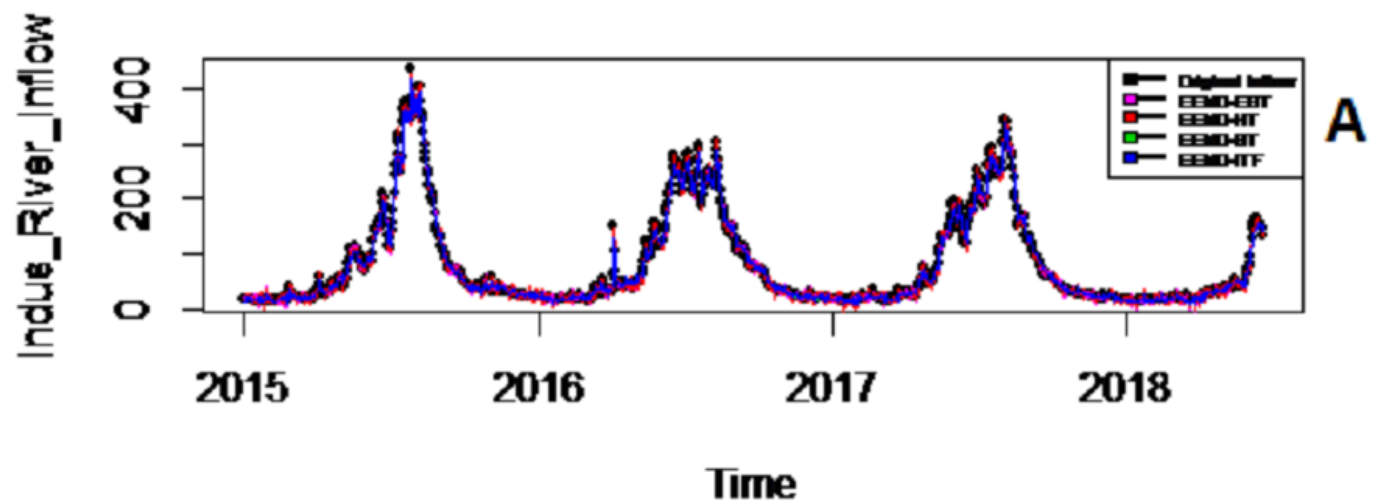
Figure 7

Prediction results of Indus (A) and Jhelum (B) river inflow using proposed CEEMDAN-EBT-MM with a comparison to all benchmark models.

Prediction results of Indus (A) and Jhelum (B) river inflow using proposed CEEMDAN-EBT-MM with a comparison to all benchmark models.
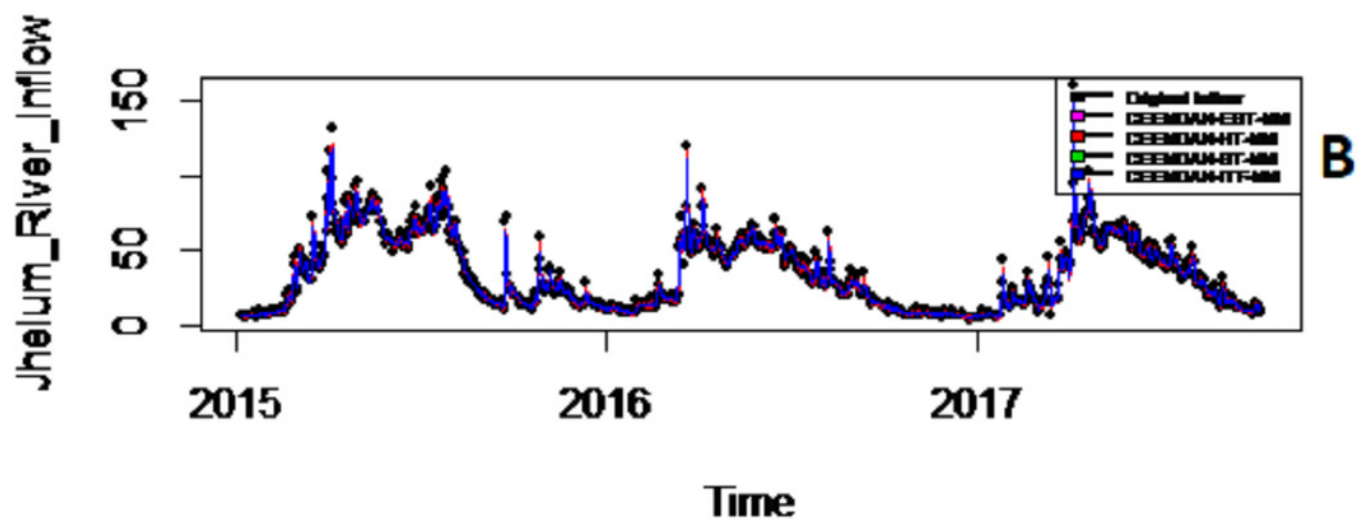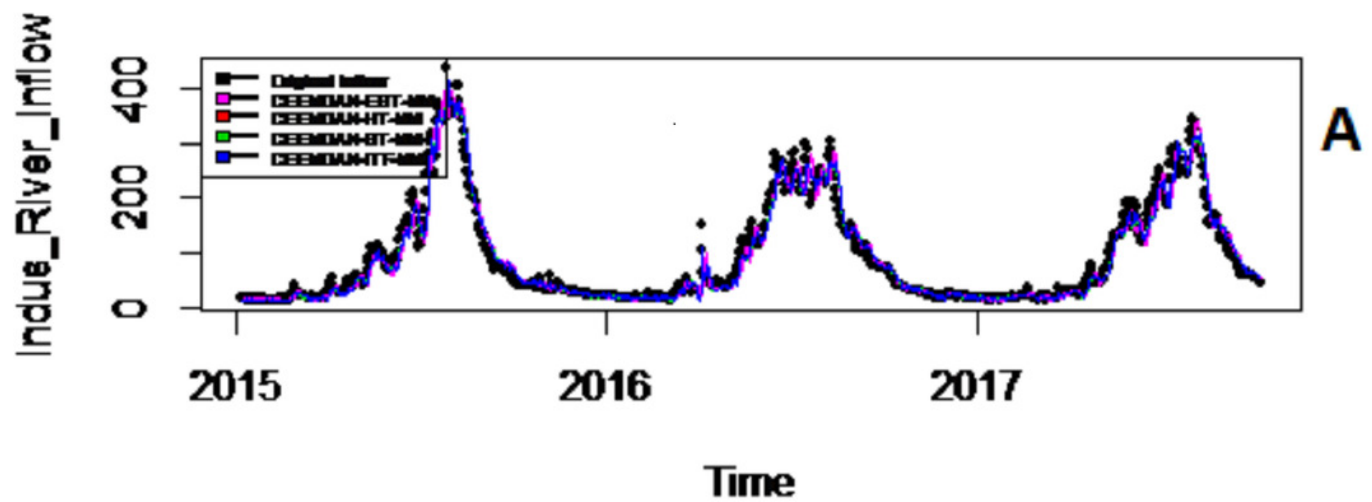
Figure 8

Prediction results of Indus and Jhelum river inflow using EEMD-EBT-MM with a comparison to all benchmark models.