Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting

Kavya Johny, Maya L. Pai & S. Adarsh

Theoretical and Applied Climatology

ISSN 0177-798X

Theor Appl Climatol
DOI 10.1007/s00704-020-03177-5

ONLINE FIRST

Springer

Theoretical and Applied Climatology
https://doi.org/10.1007/s00704-020-03177-5

ORIGINAL PAPER

Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting

Kavya Johny[1,2] · Maya L. Pai[1,2] · S. Adarsh[3]

Received: 16 September 2019 / Accepted: 1 March 2020

Abstract

Forecasting of Indian summer monsoon rainfall (ISMR) is a complex problem for the hydrologists and meteorologists. The time series and data-driven methods have been used as complementary tools for forecasting ISMR against the complex physically based dynamical models due to scarcity of data and the simplicity of former approaches. The use of hybrid decomposition data–driven models is the recent improvement among the different approaches for rainfall forecasting, but these approaches differ significantly in the framework adopted. This paper presents an adaptive hybrid modelling framework so called Adaptive Ensemble Empirical Mode Decomposition-Artificial Neural Network (AEEMD-ANN) model for forecasting ISMR, which performs the forecasts adaptively as and when new information is added. The performance of the popular EEMD-ANN hybrid hindcast and forecast experiments in the prediction of All-India SMR and southwest monsoon rainfall of the state of Kerala is compared with the proposed method. The AEEMD-ANN method achieved a predictive skill of 0.78 and 0.91, respectively for rainfall predictions for Kerala and All-India. AEEMD-ANN method performed reasonably well in capturing the hydrologic extremes when compared with EEMD-ANN forecast method, with better accuracy in capturing the drought years. The proposed method is found to be successful in capturing the extreme low SWM rainfall of year 2002 for All-India and the extreme high rainfall of Kerala 2018 with an error percentage of 1.09% and 0.52%, respectively.

1 Introduction

The summer monsoon rainfall has a tremendous socio-economic impact on Indian agriculture and health fields. But the accurate forecasting of Indian summer monsoon rainfall (ISMR) is a challenging problem for the hydrologists and meteorologists because of the complex nature of Indian monsoon system. The complete understanding on the influences of external forces and the in-house fluctuations of the monsoon are not fully unravelled, even though the attempts for it have started from the time of Blandford (Blandford 1884). Accordingly, India Meteorological Department (IMD) has a long history of release of prediction of ISMR since the late nineteenth century. Because of the complex nature of Indian monsoon system, the attempts for time series models were used by researchers for prediction of ISMR a substitute for complex physically based models (Gadgil and Srinivasan 2011; Wang et al. 2015b). However, such models projected limited capacity to deal with the non-linear and non-stationary relations in data because of the stationarity in their assumption. The use of data-driven methods were introduced as alternative to identify the non-linear characteristics between predictor variables and the ISMR (Sahai et al. 2000; Singh and Borah 2013; Kashid and Maity 2012; Pai et al. 2014). Data-driven methods like artificial neural networks (ANN) have been introduced for forecasting dynamic and non-linear systems (Fahimi et al. 2017), and later on it was found that the pre-processing of the input/output datasets may improve the performance while dealing with the non-stationary datasets (Zhang et al. 2015). The hybrid models involving suitable data pre-processing and artificial intelligence techniques are found to be suitable for hydrologic predictions (Nourani et al. 2014).

✉ Kavya Johny
kavyajovita@gmail.com

Maya L. Pai
mayalpai@gmail.com

S. Adarsh
adarsh_lce@yahoo.co.in

1 Department of Computer Science and IT, School of Arts and Sciences, Kochi, India

2 Amrita Vishwa Vidyapeetham, Coimbatore, India

3 Department of Civil Engineering, TKM College of Engineering, Kollam, India

Published online: 18 March 2020

The data pre-processing techniques such as wavelet transforms, empirical mode decomposition (EMD) and singular spectrum analysis (SSA) are reported in literature for the multiscale decomposition of time series signals, which vary in the degree of complexity in setting of control parameter and candidate functional form while applying for different types of datasets. For example, the appropriate selections of signal wavelet type and level of decomposition are two issues in the wavelet transform while using as a data pre-processing tool (Maheswaran and Khosa 2012; Santos and Silva 2014). EMD propounded by Huang et al. (1998) is found to be an appropriate decomposition method, as it is data adaptive and empirical in nature. In other words, the method is capable to decompose a time series signal to a finite number of Intrinsic Mode Functions (IMFs) by merely considering the complex nature of the data without demanding any of the functional form from the modeller. Owing to its capability for analysing non-stationarity and non-linear datasets, EMD-based (or its variants) hybrid models are popular for simulation and forecasting of hydrological variables. Some of them combined EMD (or its variants) with classical time series approaches (Kisi et al. 2014; Karthikeyan and Nagesh Kumar 2013; Zhao and Chen 2015; Wang et al. 2015a), while many of them combined it with soft computing tools (Napolitano et al. 2011; Hu et al. 2013; Huang et al. 2014; Tao et al. 2017; Wang et al. 2018; Adarsh and Janga Reddy 2018a, b, 2019; Meng et al. 2019).

In developing the hybrid models, the researchers used different schemes of decomposition method. In some of the schemes, instead of original data, the decomposed components are directly given as inputs for developing the predictive models (Wang et al. 2013; Wang et al. 2018), while in a few studies, the high frequency components are omitted in developing the predictive models (Huang et al. 2014). In some of the studies the trend, periodicity and seasonal components are separated initially based on the decomposition operation and separate models are developed to forecast the target variable (Bai et al. 2015). In many literatures, the total dataset (calibration and validation) are decomposed into IMFs and residue at the first stage itself; subsequently, each component is modelled separately and the final result is obtained by summing up the model results of each IMFs and the residue. This strategy uses some future information, and hence it cannot be considered as a real forecast or prediction but more like a simulation or hindcast. This issue was discussed in detail in some of the studies (Karthikeyan and Nagesh Kumar 2013; Zhang et al. 2015; Tan et al. 2018). Karthikeyan and Nagesh Kumar (2013) studied wavelet (WA)-based and EMD-based auto-regressive integrated moving average (ARIMA) models for prediction using various case studies of monthly total streamflow and rainfall. They compared their results with the hindcast strategy by Napolitano et al. (2011), and it was reported that, compared with EMD-based method, the wavelet-based method is performing better in forecasting the data with shorter lead time such as 6 months. Zhang et al. (2015) evaluated the performances of hybrid models such as WA-ARMA, WA-ANN, EMD-ARMA, EMD-ANN, SSA-ARMA and SSA-ANN for streamflow prediction with 1 month ahead using a hindcast method and then proposed a forecast strategy. In this method, the input series are first divided into calibration part and validation part and then each data in the validation period were appended one by one into the calibration period and forecasted. Here the decomposition and forecast models are predetermined. The increase in the length of the data that is to be decomposed may lead to variation in the number of IMFs obtained. Thus this may cause significant errors in the using predetermined ANN forecast model. Therefore, to adapt to the newly appended information, both the decomposition model and forecast models have to be updated. Later on, Tan et al. (2018) introduced an AEEMD-ANN hybrid model, which performs a real forecast by updating the decomposition model based on the newly added information. Along with the modification of decomposition model, the modification of ANN models may also be needed for improving the prediction capability of the hybrid models in real forecast experiments. The EMD-based hybrid models have been used for forecasting of Indian monsoon rainfall. Iyengar and Raghu Kanth (2005) presented a hybrid strategy EMD-ANN for forecasting monsoon rainfall of India. Several atmospheric–oceanic factors such as quasi-biennial oscillation (QBO), El-Nino Southern Oscillation (ENSO), sunspot cycle and tidal forcing are evident from each IMF obtained through EMD decomposition. Janga Reddy and Adarsh (2016) performed a study on seasonal rainfall of India. The complete ensemble empirical mode decomposition with adaptive noise algorithm and normalized Hilbert transform method is used for time-frequency characterization of rainfall at a sub-divisional scale. The study also investigated the links between monsoon rainfall and the five global climate oscillations and noticed that the low frequency components follow a similarity in their trend.

Johny et al. (2019) updated the above model with ensemble empirical mode decomposition and applied over Kerala subdivision. The system could perform with a predictive skill of 0.65 for non-linear part of validation data of SWM Kerala rainfall. The study also investigated the evidence of teleconnection of Indian Ocean dipole (IOD) to SWM rainfall of Kerala using time-dependent intrinsic correlation (TDIC) analyses. Adarsh and Janga Reddy (2017) proposed multivariate EMD–based hybrid model for modelling ISMR considering multiple causal input variables determined by hydro-climatic teleconnections along with the rainfall, but they followed a hindcasting strategy for modelling. The specific objectives are (i) to investigate the efficacy of AEEMD-ANN model in forecasting AISMR and the SWM rainfall of Kerala by comparison with that by different EEMD hybrid modelling strategies and (ii) to estimate the efficiency of AEEMD-ANN model in prediction of hydrologic extremes.

2 Materials and methods

2.1 EEMD

Huang et al. (1998) introduced a data-adaptive method called empirical mode decomposition, which decomposes the signal for analysing non-linear and non-stationary data. By this operation, the time series data can be decomposed into a zero mean series (so-called IMFs) of finite frequency and a final residue. The 'mode mixing' problem (in which the presence of same frequency in different modes or making it difficult for separation of mixed modes) is one of the critical issue of EMD, and in such cases, the physical reasoning of the signals become a challenge. Ultimately, it may direct to misinterpretations. Many variants of EMD are proposed by various researchers to overcome 'mode mixing', and ensemble empirical mode decomposition (EEMD) is one such noise-assisted variant proposed by Wu and Huang (2009). The method defines the true IMF modes as the average of an ensemble of different trials; each one of them is a combination of the signal and different copies of white noise of finite amplitude.

$x_i(t) = x(t) + w(t)$, in which $x_i(t)$ is the new signal, $x(t)$ is the given data and $w(t)$ is the white noise.

The iterative steps involved in the EEMD are as follows:

(1) add different white noises to the original signal to form many new artificially generated signals;
(2) then decompose each one of the new signals using EMD to evolve IMFs;
(3) perform ensemble mean of corresponding IMFs to obtain the required final IMF.

In the above steps, EEMD uses specific properties of white noise such that the added white noise cancels each other producing relatively even distribution of extrema because there exists no correlation between corresponding IMFs of different series. Therefore, the mean IMFs are likely to cancel each other, which follow a dyadic filter bank property that significantly reduces the chances of mode mixing.

2.2 ANN

ANNs are data-driven models inspired by biological systems, which learn from examples and capture the underlying functional relationship between data even if it is unknown. ANNs are suitable for problem where there are enough data and they require huge amount of knowledge that are challenging to describe. A feed forward neural network has an input with some number of independent (input) neurons and an output layer with dependent (output) neurons. In addition to these two layers, there are hidden layers with some number of hidden neurons that must be determined empirically for each situation.

The functional relationship of ANN can be written as a non-linear regression model:

$$Y = f\left(x_{1,}x_{2,\ldots\ldots}x_m\right)$$

where $x_1, x_2, \ldots\ldots x_m$ are independent variables, y is the dependent variable and m is the number of inputs in a training set.

Typically, a variety of architectures are used to find a best model that optimizes the performance of the network. Different architectures can be created by using different functions and varying the number of hidden layers, hidden neurons, etc. In forward propagation, they learn the data and assign weights to the nodes, and in backward propagation, the network tries to reduce the error using gradient descent algorithm. The aforementioned to and fro process continues till it reaches a prescribed benchmark. Once the training is finished, the test data is applied to the model for evaluating the performance measure of the network.

2.3 Hybrid EEMD-ANN modelling strategies

2.3.1 Hindcast strategy

In the hindcast experiment, the complete dataset (both calibration and validation part) is decomposed into a number of IMFs and residue. Then each of these modes is partitioned into calibration and validation. The calibration data is used for developing the model for different scales. Using these models, the validation part is forecasted. As this method first decomposes the entire data and then divides into calibration and validation, it uses the future information, which is unknown at present. So it is meant as unreal forecast. The stepwise process of Hindcast experiment is shown in Fig. 1.

Algorithm

Step 1 Decompose the input series using EEMD.
Step 2 Divide the obtained IMFs and residue into calibration and validation
Step 3 Consider the calibration IMFs and develop ANN models for each IMF and residue with different previous lags and hidden neurons.
Step 4 Find the RMSE for different neurons. Select the model with minimum RMSE for each IMF and residue.
Step 5 Obtain the final forecast by summing the outputs of all ANN models.
Step 6 Save the selected ANN models.

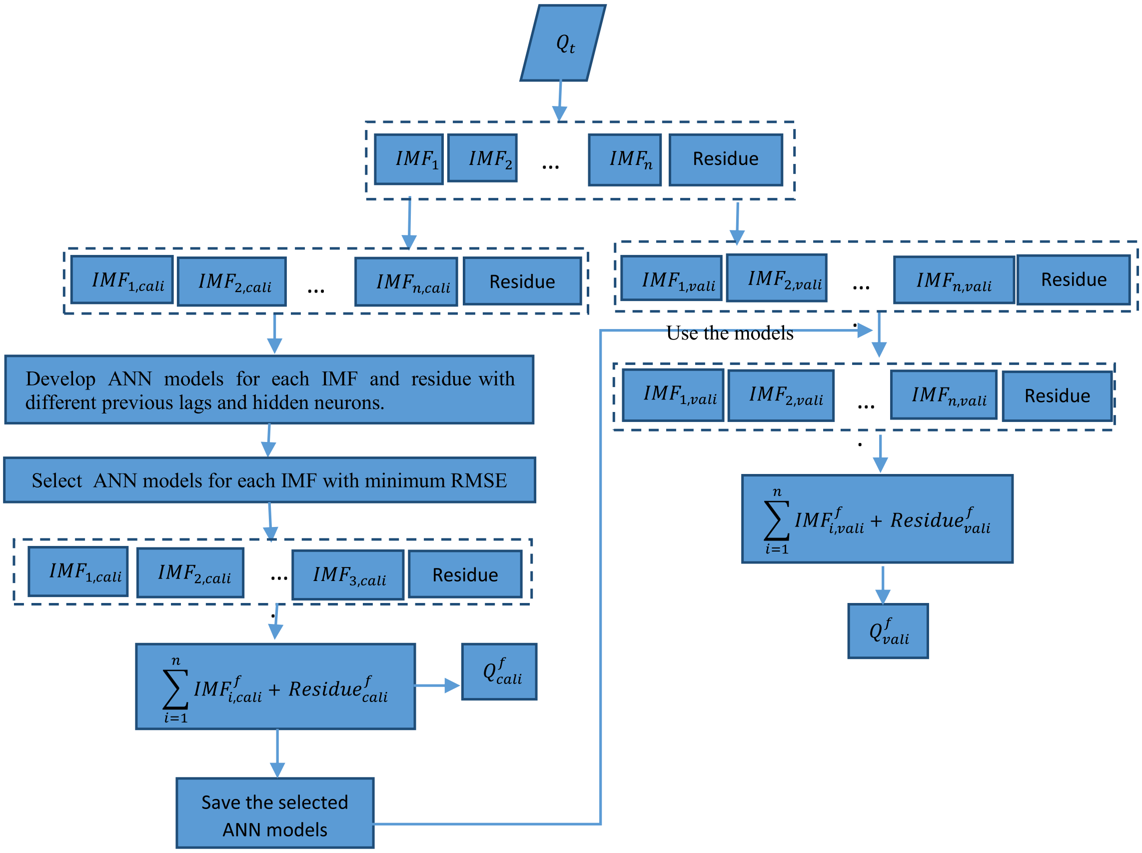

Fig. 1 Flowchart of EEMD-ANN hindcast experiment

Step 7 To forecast the subseries in the validation period use the saved models of calibration period and sum up the results of each model to get the final forecast.

2.3.2 Forecast strategy

The forecast experiment gives a real forecast when compared with that of Hindcast. In the forecast strategy, the input data is partitioned into calibration period and validation period as shown in Fig. 2. Then the data in the calibration is decomposed to produce IMFs and residue. Develop ANN models for each IMFs and residue of calibration with different lags and hidden neurons. Select the ANN models with minimum RMSE criterion. Apply these saved models for forecasting the validation period.

Algorithm

Step 1 Divide the input series into calibration and validation.
Step 2 Decompose the calibration data using EEMD.
Step 3 Consider the calibration IMFs and develop ANN models for each IMF and residue with different previous lags and hidden neurons.
Step 4 Find the RMSE for different neurons. Select the model with minimum RMSE for each IMF and residue.
Step 5 Obtain the final forecast by summing the outputs of all ANN models.
Step 6 Save the selected ANN models.
Step 7 To forecast the subseries in the validation period, append each validation data to calibration one by one during each iteration.
Step 8 Decompose the newly updated calibration period.
Step 9 Use the saved models of calibration period to forecast and sum up the results of each model to get the final forecast. Then go to step 7 until all data in validation period are appended to calibration.

2.3.3 AEEMD-ANN strategy

For a forecast experiment to be real, it should satisfy at least these two criteria: (i) it should not use any future information for forecasting the value and (ii) the decomposition model and ANN model should be well adaptive to its newly updated data. The forecast in Zhang et al. (2015) is not using the updated decomposition and forecast models, which is detrimental to the concept of the real forecast. The AEEMD-ANN model, which overcomes the downside of the forecast experiment in Zhang et al. (2015), uses an updated decomposition model and

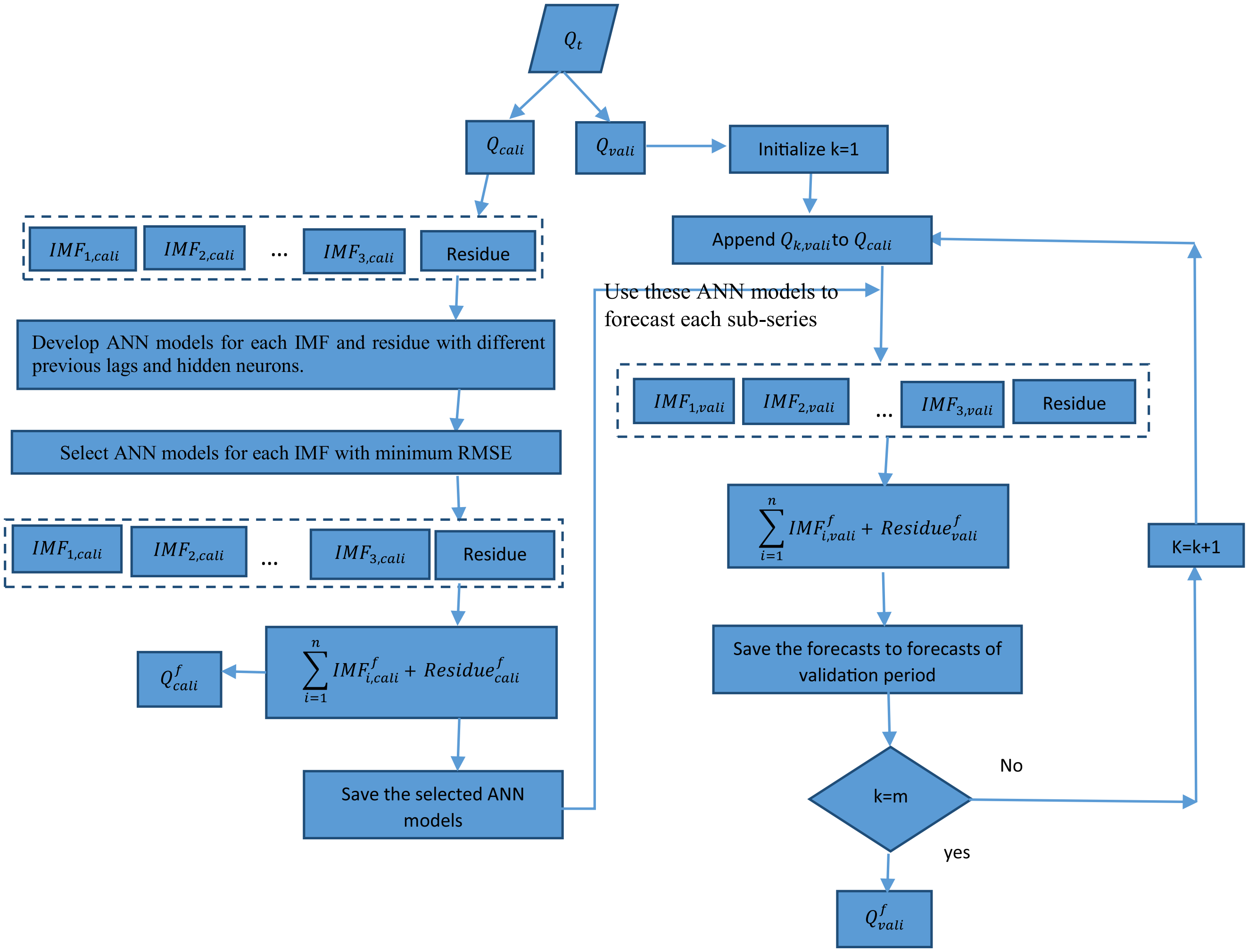

Fig. 2 Flowchart of EEMD-ANN forecast experiment

forecast model as shown in Fig. 3. The adaptivity of both models with the updation of each data is also an essential requirement. This requisite is accomplished in the proposed framework of AEEMD-ANN.

Algorithm

Step 1 Divide the input series into calibration and validation.

Step 2 Decompose the calibration data using EEMD.

Step 3 Consider the calibration IMFs and develop ANN models for each IMF and residue with different previous lags and hidden neurons.

Step 4 Find the RMSE for different neurons. Select the model with minimum RMSE for each IMF and residue.

Step 5 Perform one-period-ahead forecast for each IMFs and obtain the final forecast by summing the outputs of all ANN models.

Step 6 Save the forecasts to the forecasts of validation period

Step 7 To forecast the subseries in the validation period, append each validation data to calibration one by one during each iteration.

Step 8 Decompose the newly updated calibration period and then repeat steps from 3 to 8 until all data in validation period are appended to calibration.

The proposed AEEMD-ANN strategy overcomes the limitations of forecast method. The AEEMD-ANN model in Tan et al. (2018) also followed a strategy better than the forecast experiment in Zhang et al. (2015). But the model performs EEMD decomposition and forecasts using the selected ANN models for each IMF, which are saved for further use. The model then performs a normal forecast for the same period of input. The obtained forecasted values then applied to the saved models for next immediate one-period-ahead forecasts. Instead, the proposed model under the study follows the procedure that each validation data appended to the calibration data is taken as a single sample, performed the EEMD decomposition and is forecasted ahead directly. Thus, it avoids two subsequent forecasts by exploiting single direct one-period-ahead forecasts. Henceforth, the methodology is to forecast for the next year '$(y+1)$' based upon the observed values of previous years up to 'y'.

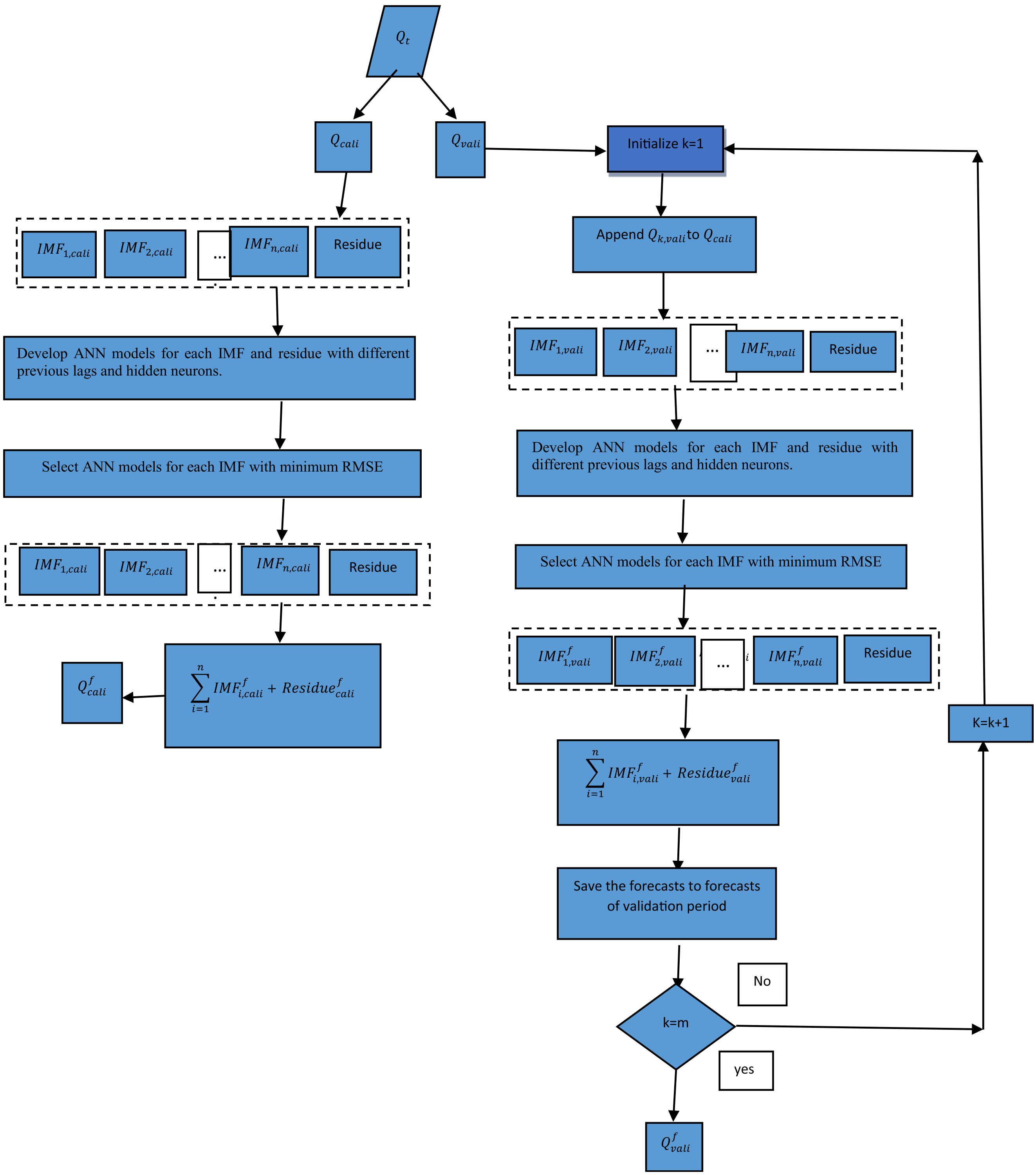

Fig. 3 Flowchart of Adaptive EEMD-ANN hybrid modelling

3 Study area and datasets

Based on homogeneity of rainfall, the Indian Institute of Tropical Meteorology (IITM) Pune fixed the boundaries of 36 meteorological subdivisions in India. The area-weighted precipitation of these subdivisions at monthly scale has been prepared first in 1994 by Parthasarathy (Parthasarathy et al. 1994) considering a minimum of one representative rain gauge station per district. This database has been updated many times by the researchers of IITM Pune, and the recent version is available at Kothawale and Rajeevan (2017). The monthly rainfall data of Kerala meteorological subdivision

(8.29° N–11.52° N Lat and 75.22° E–76.57° E Lon.) for 144 years (1871–2014 period) and the monthly data of All-India spatial region (8.5° N–37.5° N Lat and 68.5° E–97.5° E Lon.) for 146 years (1871–2016) are extracted from the website of IITM Pune (http://www.tropmet.res.in). Then the southwest monsoon (SWM) seasonal rainfall of Kerala subdivision and AISMR are computed by aggregating the monthly rainfalls of June, July, August and September months. The rainfall data from 62 rain gauge stations in Kerala for June to September months for the recent years (2017 and 2018) are collected from India Meteorological Department (IMD) Thiruvananthapuram.

4 Results and discussion

In this study, first EEMD is performed on AISMR data and SWM rainfall data of Kerala meteorological subdivision. The control parameters like number of sifting iterations, ensemble number and noise are selected as 1000, 900 and 0.04, respectively. The EEMD in MATLAB platform (available at http://perso.ens-lyon.fr/patrick.flandrin/eemd.html) is used with necessary modifications. The results of EEMD decomposition of AISMR data and SWM rainfall data of Kerala are provided in Fig. 4 a and b, respectively. The modes obtained by decomposition are analysed in detail to get useful physical insights in terms of rainfall processes and periodicity. The average period of each IMF is estimated by the zero crossing method by determining the number of extrema and zero crossings (Huang et al. 2009) . The percentage variability and mean period expressed by each modes of both the AISMR and SWM rainfall of Kerala are presented in Table 1.While the period of each oscillation is increasing from IMF_1 to IMF_n, the frequency of oscillation/number of cycles per signal is decreasing in each IMF. The first five modes, which comprise sufficient cycles (hence their periodicity estimate are reliable), are given in Table 1. Subsequently, the hindcast, forecast and adaptive forecast strategies are invoked for prediction of the AISMR and SWM rainfall of Kerala subdivision.

4.1 Hindcast experiment

In the hindcast experiment, first the IMFs and residue resulted by the EEMD of SWM rainfall series of Kerala and India are divided into calibration and validation. The rainfall data from 1871 to 1970 are considered for calibration and from 1971 to 2014 data for validation period. ANN forecast model is built using the calibration dataset. The number of hidden neurons in ANN model for each IMF is determined using trial and error

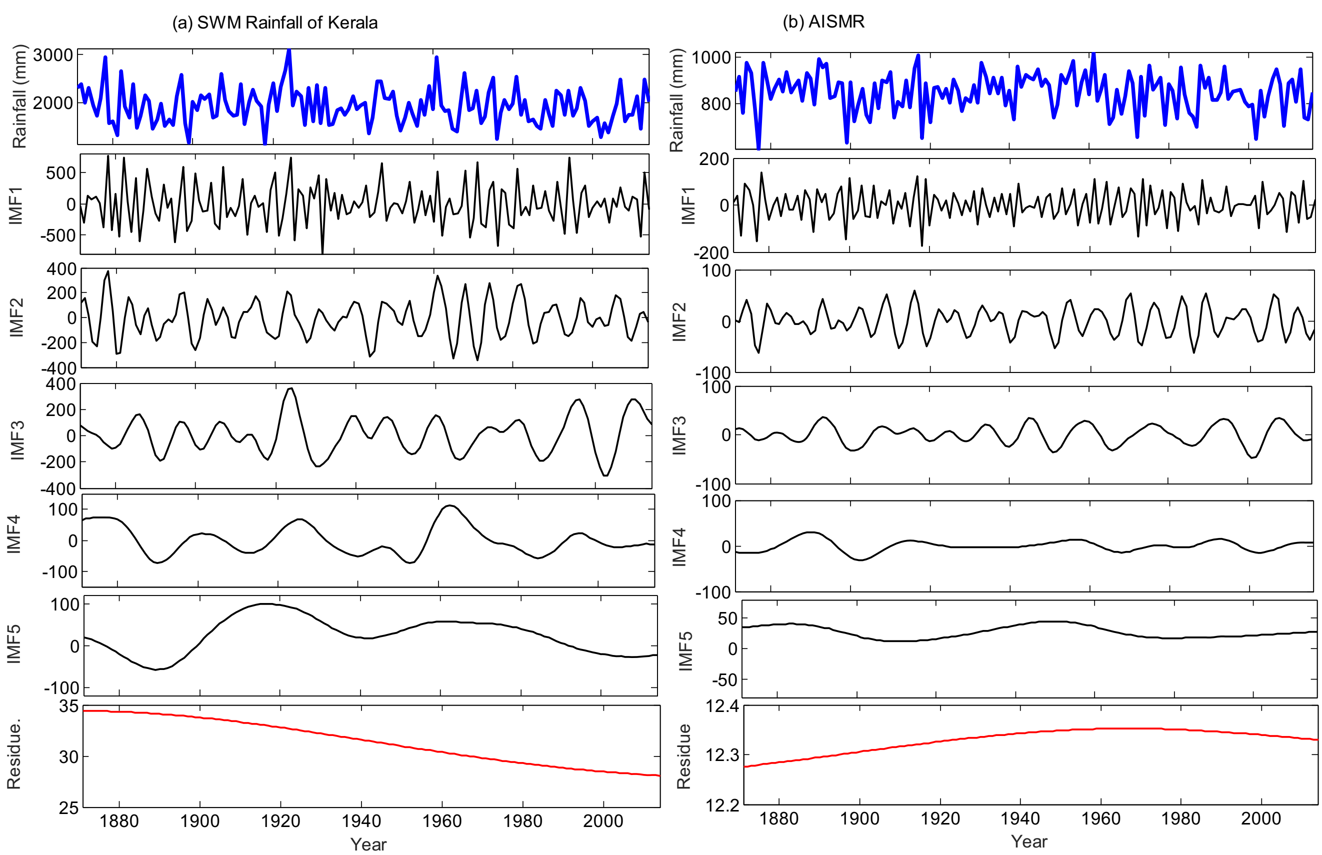

Fig. 4 Modes obtained by EEMD of rainfall data. **a** SWM Rainfall of Kerala; **b** AISMR

Table 1 The mean period (in years), variance of modes of SWM Kerala rainfall and AISMR

Mode number	Kerala		India	
	Mean period (years)	Variability (%)	Mean period (years)	Variability (%)
IMF1	2.83	62.89	2.71	59.21
IMF2	5.80	15.46	6.00	10.36
IMF3	11.13	9.24	12.40	5.46
IMF4	20.33	6.35	21.00	1.95
IMF5	50.67	1.02	88.00	1.84
Residue	–	0.37	–	0.54

method. The number of hidden nodes from two to twenty is tried, and RMSE is calculated for each trial. The best model with minimum RMSE is selected for respective IMFs. The number of hidden neurons is different for different IMFs. The selected models are applied for validation datasets and forecasted the values. The final forecast of a year is obtained by summing the outputs of all ANN models.

4.2 Forecast experiment

The SWM rainfall data of Kerala and India are divided into calibration (1871–1970) and validation (1971–2014) period. The calibration data are decomposed using EEMD into different IMFs and a final residue. These IMFs are then used to model ANN forecast model. The best models are selected using minimum RMSE. Save the models to forecast using validation period. During forecasting validation period, each value of validation dataset is added one by one to calibration data on each iteration. Then the updated data is decomposed again and performed ANN forecast using the saved models. The final forecast of a year is obtained by summing the outputs of all ANN models. This process is repeatedly performed till all data in the validation period is appended to calibration completely.

4.3 Adaptive forecast experiment

In adaptive forecast experiment, at first the data is divided into calibration and validation datasets. The calibration period is considered first and decomposed using EEMD. ANN models are developed for the IMFs and residue; each model is for forecasting the individual mode separately. Hidden neurons for each model are selected using trial and error. The number of hidden nodes from two to twenty is tried and RMSE is calculated for each trial. The best model with minimum RMSE is selected for respective IMFs. The number of hidden neurons is different for different IMFs. The sum of outputs of all ANN models is used to calculate the final forecast. When considering the validation period, one by one, data from validation is appended to calibration data. This updated calibration data is again decomposed and forecasted using ANN models. Apart from hindcast and forecast model, here the ANN models are again selected using trial and error for hidden neurons from two to twenty and by summing up the outputs of all the models the final forecast is obtained. This process is iteratively done to get the forecast for all the years.

4.4 Comparison of predictions by the three hybrid models

In the exercise of assessing the forecasting ability of different hybrid models, the simple ANN is also used for forecasting AISMR and summer monsoon rainfall of Kerala. A lag of 5 years of data is used for forecasting (Sahai et al. 2000; Singh and Borah 2013). A neural network of a single hidden layer is used for forecasting. The number of neurons in hidden layers is determined by the following equation: $HL_{\text{nodes}} = IL_{\text{nodes}} + 1$, where HL_{nodes} and IL_{nodes} represent the number of neurons in hidden and input layers, respectively. Therefore, a neural network consisting of an input layer with five neurons, a hidden layer with 10 neurons and an output layer with a single neuron are used as the model. The number of hidden neurons is selected using trial and error method. Sigmoid transfer function and the Levenberg–Marquardt (LM) algorithm are used for the ANN model, considering the rainfall datasets of 1871–1970 of both Kerala and India. The validation is done using datasets of 1971–2014 and 1971–2016 for Kerala and India, respectively. The correlation between observed and predicted SWM rainfall of calibration and validation are 0.34 and 0.18 for AISMR data and 0.63 and 0.23 for data of Kerala subdivision. It shows that simple ANN results in poor correlation than that of hindcast, forecast and adaptive forecast and the hybrid decomposition models perform much better than ANN. The performances of hindcast, forecast and adaptive forecast of SWM Kerala rainfall and All-India SWM rainfall are compared by computing the values of different statistical performance evaluation measures. Various statistical measures like correlation coefficient (R), mean absolute error (MAE), index of agreement (IA) and normalized root mean square error (NRMSE) are used as performance evaluation measures. MAE is the mean of differences between observed and its corresponding predicted value. NRMSE is the normalized RMSE which is computed as $\mathrm{NRMSE} = \frac{\sqrt{\frac{\sum_{j=1}^{m}(x_o - x_p)^2}{n}}}{(x_{\max} - x_{\min})}$ in which $x_o$is the observed rainfall, x_p is the predicted rainfall and $x_{\max}$ and $x_{\min}$ are maximum and minimum of observed values, respectively. An NRMSE value of greater than 1 means poor model performance. Index of agreement 'IA' measures model performance by comparing model predictions with observations. The index of agreement (Kashid and Maity 2012) is computed as $\mathrm{IA} = 1 - \frac{\sum_{t=1}^{K}|(x_p^t - x_o^t)|}{\sum_{t=1}^{K}(|(x_p^t - \overline{x_o})| + |(x_o^t - \overline{x_o})|)}$ where $\overline{x_o}$ is the estimated mean of the observations x_o^t is the observed value at time t and x_p^t is the predicted value at time t. IA closer to 1 means the more accurate the model is. The results of performance evaluation of the three hybrid models are presented in Table 2. Also, for a better comparison, the scatter plots of predictions by the three hybrid methods are presented in Fig. 5.

From Table 2, it is noted that the hindcast method display best accuracy in predictions in terms of highest correlation (0.81 and 0.85), least error (MAE of 174.9 and 41.35; NRMSE of and 0.18 and 0.18) and highest index of agreement (0.68 and 0.72), respectively, for predictions of SWM rainfall of Kerala and the AISMR. This level of accuracy is anticipated and in agreement with some of the past studies (Adarsh and Janga Reddy 2017). But it is to be noted that the hindcast experiment is quite unrealistic as it considers the future information (rainfall) value in the predictions. On comparing the performance of EEMD-ANN forecast experiment with that of AEEMD-ANN method, it is noted that the predictive accuracy of validation data of SWM rainfall of Kerala is better than EEMD-ANN forecast method (i.e., an R value of 0.78, NRMSE of 0.27 and IA of 0.62 against 0.63, 0.26 and 0.58 by the latter method). Identical comment holds good for the AISMR predictions with higher correlation (0.91 against 0.83), and IA (0.76 against 0.73) and lesser error measures (MAE of 34.3 against 39.8; 0.14 against 0.17) noticed for AEEMD-ANN than the EEMD-ANN forecasts. Among the three hybrid models, the adaptive forecast method performs reasonably better than that of forecast model. The scatter plots provided in Fig. 5 show that there exists a good agreement between the predicted and observed rainfall data for the adaptive method than the forecast strategy. The coefficient of determination (R^2) of predictions by adaptive method are 0.61 and 0.83 for SWM rainfall of Kerala and AISMR, respectively, which is better than 0.40 and 0.69 of forecast strategy and closer to 0.66 and 0.72 of the hindcast method. It is clear that the predictions by the AEEMD-ANN method are close to the ideal fit when compared with predictions by forecast method during validation.

Further, the performance comparison among the different models is done by plotting the Taylor diagrams as shown in Fig. 6. Taylor diagram is one of the best graphical visualizations to evaluate the prediction accuracy based on several statistical indicators (Taylor, 2001). The standard deviation of observed, hindcast, forecast and AEEMD-ANN methods are 84.17 cm, 93.76 cm, 94.35 cm and 95.01 cm, respectively, for AISMR rainfall and 334.98 cm, 319.89 cm, 405.34 cm and 468.82 cm, respectively, for Kerala SWM rainfall. From Fig. 6, it is observed that AEEMD-ANN strategy makes predictions with correlation coefficient and standard deviation greater and centred root mean square deviation lesser than EEMD-ANN forecast model of Kerala and India and implies that even though the spread-out rainfall data distribution is more, the new method could predict the future rainfall values with minimum error.

In this research, we examined the potential of proposed variant of AEEMD-ANN hybrid model for performing ISMR predictions. Despite the inherent complexity and data scarcity, researchers have used different physical models for

Table 2 Performance evaluation of predictions of hindcast, forecast and adaptive forecast (AEEMD-ANN) experiments for AISMR and SWM rainfall of Kerala

Performance measure	Hindcast		Forecast		AEEMD-ANN
	Calibration	Validation	Calibration	Validation	Validation
SWM of Kerala					
R	0.85	0.81	0.76	0.63	0.78
MAE	185.28	174.86	194.98	256.53	260.53
NRMSE	0.12	0.18	0.54	0.26	0.27
IA	0.72	0.68	0.67	0.58	0.62
AISMR					
R	0.99	0.85	0.86	0.83	0.91
MAE	7.42	41.35	25.55	39.76	34.28
NRMSE	0.02	0.18	0.10	0.17	0.14
IA	0.94	0.72	0.79	0.73	0.76

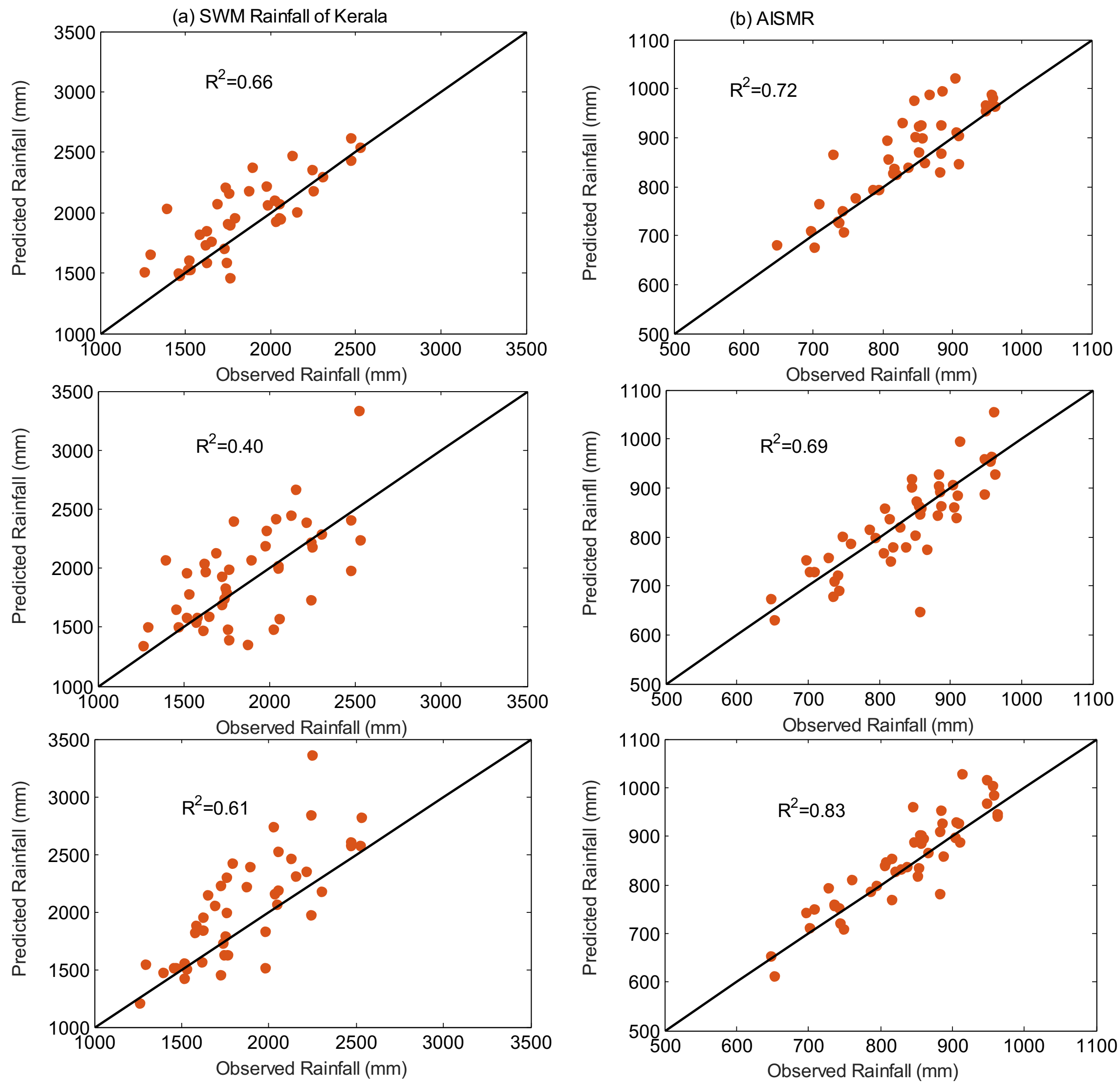

Fig. 5 Predictions by hindcast, forecast and AEEMD-ANN hybrid model for validation data. **a** SWM rainfall of Kerala; **b** AISMR. Upper panels show the results by hindcast method, middle panel show the results by forecast method and the lower panels show the results by adaptive forecast method

ISMR predictions (Wang et al., 2015b) considering the argument that it is more realistic to consider the physics of the problem in the monsoon predictions. Ensemble project multi-model ensemble (MME) hindcast of All-India rainfall index (AIRI) has resulted in a skill of 0.63 for the ISMR predictions, whereas physically based empirical model (P-E model) displayed a skill of 0.77 (Wang et al., 2015b). Hence it can be concluded that the model calibration for the present case is acceptable for predictions. On the other hand, some of the well-applauded studies, which used ANN as the prediction method and lagged rainfall values as input, resulted in a predictive skill up to 0.81(Sahai et al., 2000). Many of the recent studies that used data-driven techniques also have reported similar performance for ISMR predictions using the IITM dataset, considering lagged rainfall as the only input variable (Singh and Borah 2013; Singh et al., 2018), except for the differences in the time spells of calibration and validation. Here it is to be noted that using large number of regressors may lead to multi-collinearity and overfitting (Ghosh and Mujumdar 2007). The smaller difference between correlation values of training and validation (say < 0.2, in Table 2) and a reasonably good value of correlation (say $>$ 0.7, for the proposed variant, which follow a 'real-time' calibration operation as and when a new dataset is added) also proves that there is no chances of overfitting. The EMD-ANN hybrid model by Iyengar and Raghu Kanth (2005) has resulted in a correlation skill even up to 0.91, and the multivariate EMD–based hybrid hindcasting strategy proposed by Adarsh and Janga Reddy (2017) has resulted in a correlation skill of 0.798 for different validation time spells. In general, the inclusion of decomposition step in modelling can help in capturing information from different process (time) scales, which may result in obvious improvement in overall predictions including extremes (Huang et al. 2014; Adarsh and Janga Reddy 2017,

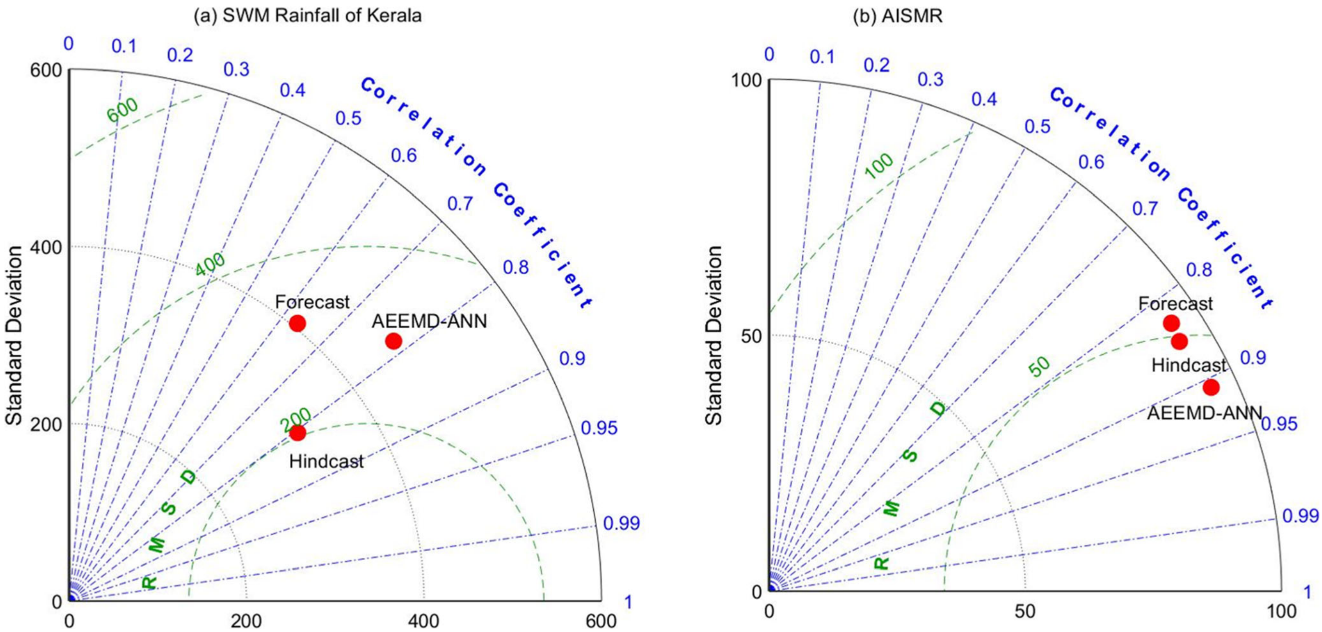

Fig. 6 Taylor diagrams showing the performance by hindcast, forecast and adaptive forecast experiments. **a** Prediction of SWM rainfall of Kerala (standard deviation = 334.98). **b** Prediction of AISMR (standard deviation = 84.17)

2018b). The most relevant input that can contribute to the rainfall component of the particular time scale will be considered in the modelling process, and this will be more effective in AEEMD-ANN model as the decomposition process is carried out in each step as and when a new input is added to get the forecasts of subsequent years, i.e., the adaptiveness enhances the capability of information capturing pertaining to different process scales. This may be the reason for improved performance of the proposed AEEMD-ANN model than the other EEMD variants following forecast strategy and the traditional ANNs. The characteristics of newly added data compared with those added in the previous steps will reflect in the performance improvement of AEEMD-ANN forecasts over the EEMD-ANN forecasts, i.e., noticeable improvement gets displayed when the characteristics of newly added data are quite different and information from the process scales are distinctly different and those gets captured effectively.

Two prevailing conditions for a real forecast model could be (i) future information must not be used in the decomposition–ensemble framework and (ii) both decomposition and forecast models should be highly adaptive (Tan et al. 2018). The reuse of calibration decomposition model for validation purpose may induce error in the forecast experiment. This could be the reason for the poor performance of existing forecasting models, while the use of future information is obvious in hindcast models, as the decomposition of the complete dataset (calibration and validation) is carried out at the first step itself. As a result, the method should make some alternations for the model to adapt with the validation data. Therefore, the proposed strategy updates both the decomposition model and ANN model each time a new value is inputted into the system. Following the decomposition process, the ANN model with minimum RMSE is generated for each IMF with a rigorous trial and error procedure by varying the number of neurons between 2 and 20. This also helps to capture the characteristics of the updated signal, which is almost impossible in the hindcast and forecast strategies. Overall, this study proposed a practical real forecasting strategy for rainfall predictions circumventing the possible accuracy reduction resulting from calibration decomposition model during implementation reported in some of the earlier variants (Karthikeyan and Nagesh Kumar 2013; Zhang et al. 2015; Tan et al. 2018) and accepting the application of hindcasting operation (Napolitano et al. 2011; Adarsh and Janga Reddy 2017) as a suitable strategy for simulation studies. In short, it can be concluded that (i) the use of rainfall (and the lagged values) is a complementary option in ISMR predictions, in the absence of sufficient data inputs for performing the physically based models; (ii) the high predictive skills and smaller difference between the correlation coefficient for calibration and validation data rules out the chances of overfitting in the modelling; (iii) prediction skills can be improved by hybridizing the ANN model by using the decomposition as a preprocessing step, provided the strategy of operation (hindcast or forecast) is crucial in type of application (simulation or prediction); (iv) decomposition methods provide a supplementary feature to capture the significant information from different process scales and this is more evident and effective in the adaptive EEMD (AEEMD) method. From the aforementioned reasons, it could be well defined that the new adaptive framework performing real forecast is the best suited methodology for EMD-based hybrid models. From the results of the study, it is clear that the AEEMD-ANN method performs better than forecast method for AISMR and that for Kerala AEEMD-ANN could make a real forecast equally good with the EEMD-ANN method. This could be attributed to the characteristics of newly added information in each forecast. If data of similar characteristics are added, the difference in performance between the two methods will be marginal and vice versa. It is noted that the performance is relatively better for AISMR predication than the SWM rainfall predictions of Kerala in which the AISMR data is from a larger spatial extent. Analysing the local regional processes and identifying its varying periodicity may help to improve the performance. However, to corroborate this observation, more investigation with similar exercises has to be conducted for other meteorological subdivisions of different climatic conditions in India.

4.5 Prediction of extreme rainfall

One possible reason for the improved performance AEEMD-ANN could be the ability of the models to capture the extreme floods and droughts better than the forecast method, i.e., in addition of new values pertaining to extremes (like low or high value), the potential of decomposition may be more noticeable compared with the EEMD-ANN. To investigate this, the performance of the models in capturing the extremes are analysed separately. The Drought Research Unit (DRU) of India is conducting studies on different aspects of the drought and identified different drought years. The predicted values using hindcast, forecast and adaptive forecast are classified into drought and wet years of seven indexes, namely, Extremely Wet (EW), Very Wet (VW), Moderately Wet (MW), Near Normal (NN), Moderately Dry (MD), Severely Dry (SD) and Extremely Dry (ED) after standardizing the precipitation data and is shown in Table 3. From Table 3, it is noticed that considering the drought years, except for one case (out of eight), AEEMD-ANN identified the same drought state while forecast could capture only one case correctly. But, on considering the five flood years, both AEEMD-ANN and the EEMD-ANN display fair degree of accuracy in identifying the state.

Based on the above classification, observed data of rainfall of Kerala has a ratio of 8:5 drought and flood years. When compared with forecast, adaptive forecast (AEEMD-ANN) could make predictions better, though the percentage of extremes varies lesser with that of respective observed values. Considering the different lags for different IMFs, the hindcast method could make prediction from the year 1977. So the classification for the years from 1970 to 1976 cannot be done for the hindcast. The results obtained on prediction using the three methods and % error in prediction of extremes is summarized in Table 4.

From Table 4, it is noticed that the error percentage in the prediction of extreme drought of 2002 was 20% by the AEEMD-ANN model, while it was 16.2% by the forecast method, and the extreme flood of 1981 was predicted with an error % of 11.55 by the AEEMD-ANN method, while the prediction by forecast displayed a comparable error percentage of 11.2%. Dropping this extreme case of drought, it is noted that the droughts were predicted remarkably well by the AEEMD-ANN method (with an average error % of 6) in comparison with the EEMD-ANN forecast model (displayed an average error % of 17.4). Also except for two cases out of eight (with a success rate of 75%), the predictions by AEEMD-ANN method was much accurate than that by forecast method. On considering the flood years two out of five cases, the AEEMD-ANN method was found to be more accurate than forecast method, while in one case (of 1981), the accuracy was comparable. The results reveal that the AEEMD-ANN is capable of predicting the drought years in a better way than flood years than by a forecast model, while in prediction of flood years, it displays a mixed response.

In addition to this, predictions of Kerala rainfall for years from 2015 to 2018 are performed using AEEMD-ANN method, and the results are presented in Table 5. Out of four years, AEEMD-ANN could predict the drought states of extreme drought in 2015, near normal condition of 2017 and the flood of 2018 correctly. The predictions of these four years are presented as bar graph in Fig. 7a. Figure 7 a shows that AEEMD-ANN hybrid model could predict the rainfall magnitudes of the recent years reasonably well including the extreme flood of 2018 and except in 2016 (Fig. 7a). The percentage departure of the observed and predicted values from long term mean

Table 3 Drought and flood years in validation data (1971–2014) of Kerala and prediction state by different models (ED-Extremely Dry>SD-Severely Dry>MD-Moderately Dry; NN- Near Normal; EW-Extremely Wet>VW-Very Wet>MW-Moderately Wet)

Year	Observed	Hindcast	Forecast	AEEMD-ANN
Drought years				
1976	SD	SD	MD	SD
1986	MD	MD	NN	MD
1987	MD	SD	NN	MD
1990	MD	MD	NN	MD
1999	MD	MD	NN	MD
2002	SD	MD	MD	MD
2004	MD	NN	NN	MD
2012	MD	SD	MD	MD
Wet years				
1975	VW	NA	EW	MW
1981	VW	VW	NN	VW
1997	MW	MW	NN	NN
2007	VW	WE	NN	MW
2013	VW	VW	MW	MW

Table 4 Error percentage on predicted values on rainfall extremes (drought and flood) of Kerala by hindcast, forecast and adaptive forecast methods. The italicized values show the extreme among the drought and flood years

Year	Observed	AEEMD-ANN	Error %	Forecast	Error %	Hindcast	Error %
Drought years							
1976	*1260.6*	*1212.03*	*3.85*	*1345.76*	*6.75*	*1511.02*	*19.86*
1986	1528.6	1501.29	1.78	1780.47	16.47	1529.40	0.052
1987	1456.5	1515.56	4.05	1650.33	13.30	1500.56	3.02
1990	1517.1	1426.82	5.95	1963.05	29.39	1611.88	6.24
1999	1516.0	1551.82	2.36	1580.16	4.23	1525.80	0.64
2002	1292.2	1550.84	20.01	1501.47	16.19	1655.02	28.07
2004	1390.1	1472.04	5.89	2073.65	49.17	2033.99	46.31
2012	1466.8	1520.02	3.62	1502.97	2.46	1477.69	0.74
Flood years							
1975	2521.5	2581.28	2.37	3334.65	32.24	NA	NA
1981	*2526.3*	*2818.14*	*11.55*	*2243.84*	*11.18*	*2545.03*	*0.74*
1997	2303.2	2174.78	5.57	2286.45	0.72	2298.11	0.22
2007	2470.7	2610.80	5.67	1982.38	19.76	2621.90	6.11
2013	2470.0	2576.95	4.33	2408.78	2.47	2431.71	1.55

of 1871 to 2018 period (1920.96 mm) of these four years (2015–2018) is given in Fig. 7 b and c. It is noted that there is a close matching in percentage departure (− 24.17% against − 23.52% for observed value) in the year 2018, which was a year of hydrologic extreme. It should be noted that the AEEMD-ANN method could make the prediction of flood year 2018 with an error percentage of 0.52%.

Observed data of rainfall of All-India has a ratio of 8:6 drought and flood years. The prediction states of drought and flood years in validation data (1971–2014) of AISMR by all three models are presented in Table 6. From Table 6, it is noticed that on considering the drought years, in six cases (out of eight), AEEMD-ANN identified the same drought state against 3 cases (out of eight) by EEMD-ANN method. Similarly, on considering the six flood years, in three cases, the AEEMD-ANN identified the drought state, whereas the EEMD-ANN method identified only one case correctly. Further, the predictions and % error in predictions of drought and flood year rainfall magnitude are presented in Table 7. From Table 7, it is noticed that out of the eight cases

Table 5 Predicted rainfall state Kerala for the years 2015–2018

Year	Observed interpretation	Predicted interpretation
2015	MD	MD
2016	SD	NN
2017	NN	NN
2018	*VW*	*VW*

Note: The category in italics indicate that the proposed method could predict the extreme flood of Kerala 2018

considered, except for three cases, the error percentage in the prediction of drought by AEEMD-ANN is less than that by forecast method (with a success rate of 62%). The average error in predictions is 4% and 5.2% by the AEEMD-ANN method and EEMD-ANN forecast model for the drought years, respectively, while the values are comparable (3.52% and 3.62%) for the two methods for the flood years. Even though the average error is comparable, the individual years the prediction of flood years by the AEEMD-ANN method is not always encouraging. It is also to be highlighted that the model could forecast the well-debated drought in India of the year 2002 with excellent accuracy of (i.e., with an error of only 1.09%). From this, it is understood that AEEMD-ANN adaptive forecast, which is a real forecast, can make better prediction than forecast experiment for prediction of extremes of AISMR. It is to be noted that the above discussion focused mainly on the comparison between forecast and AEEMD-ANN method as the hindcast experiment is not a 'real' forecast, and accordingly it showed inconsistencies in prediction accuracy of extremes. In short, by both analysis (for SWM rainfall of Kerala and AISMR), it can be concluded that AEEMD-ANN model display commendable accuracy in forecasting extreme events especially the drought.

To determine the efficiency on prediction of all the three models such as hindcast, forecast and adaptive forecast, a bar graph (Fig. 8) is used to represent the amount of rainfall predicted by each model against the observed values for extreme years of SWM rainfall of All-India and Kerala. The extreme events such as ED, SD, EW and VW are selected based on the extremity. Figure 8 a represents the predictions made by hindcast, forecast and AEEMD-ANN for drought years

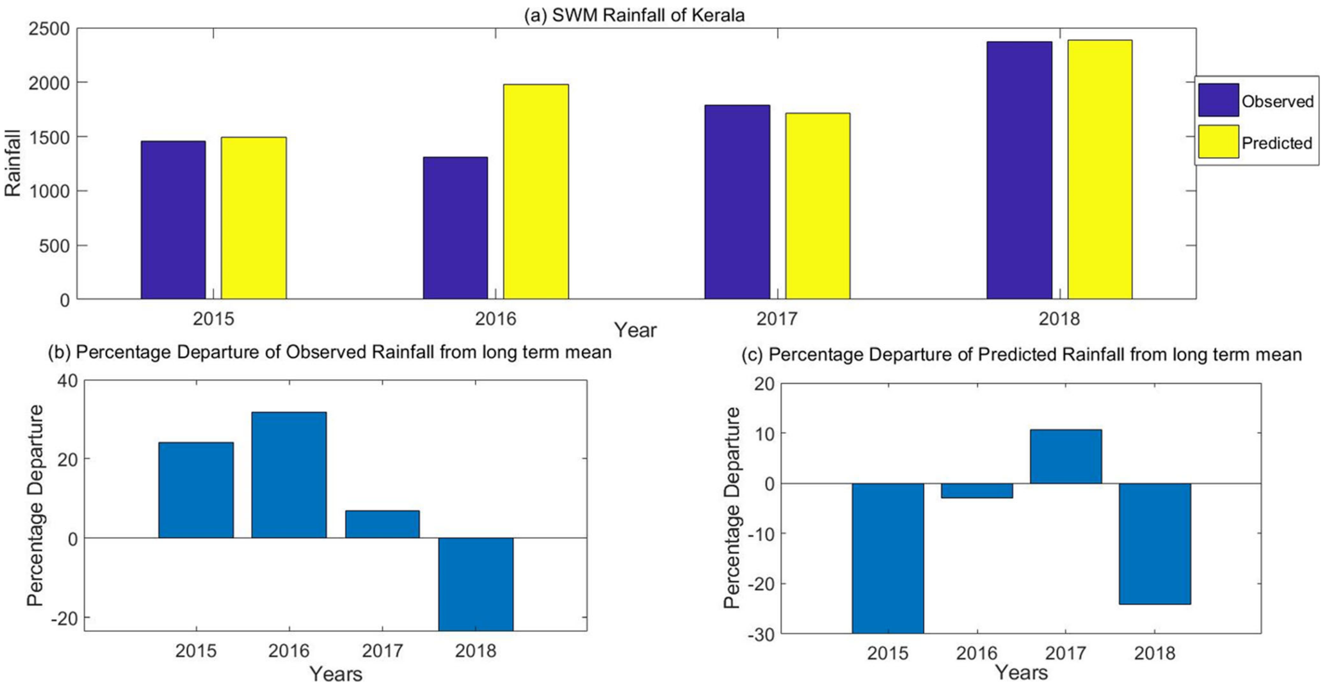

Fig. 7 Predictions of SWM rainfall of Kerala for recent years 2015–2018. **a** Comparison of rainfall predictions by AEEMD-ANN with the observed values; **b** Percentage departures of observed values from long term mean; **c** Percentage departures of AEEMD-ANN predictions from long term mean

(1976, 2002) and flood years (1981, 2007, 2013) for Kerala. The VW event of 1975 is not considered as hindcast method makes no prediction for the year. Figure 8 b represents the predictions made by hindcast, forecast and AEEMD-ANN

Table 6 Drought and flood years in validation data (1971–2014) of AISMR and prediction state by different models (ED-Extremely Dry>SD-Severely Dry>MD-Moderately Dry; NN-Near Normal; EW-Extremely Wet>VW-Very Wet>MW-Moderately Wet)

Year	Observed	Hindcast	Forecast	AEEMD-ANN
Drought years				
1972	ED	NA	ED	ED
1974	MD	NA	NN	SD
1979	MD	MD	MD	MD
1982	MD	MD	SD	MD
1986	MD	SD	SD	MD
1987	SD	SD	NN	MD
2002	ED	SD	SD	ED
2004	MD	MD	MD	MD
Flood years				
1975	VW	NA	MW	MW
1983	VW	MW	MW	VW
1988	VW	MW	EW	NN
1994	VW	MW	MW	MW
2007	MW	NN	NN	MW
2013	MW	MW	MW	MW

for drought years (1987, 2002) and flood years (1983, 1988, 1994) for All-India. For All-India ED of 1972 cannot be predicted by hindcast method as the lagged values are considered as inputs. Therefore, it is omitted from the bar graph. From Fig. 8, it can be inferred that out of five extreme events, the AEEMD-ANN could predict four extreme events better than EEMD-ANN forecast model for All-India and three extreme events, for Kerala.

The percentage departures for all the three methods hindcast, forecast and AEEMD-ANN are compared with the long term mean of 1871–2014 (1926.22 mm for SWM of Kerala and 948.17 for AISMR) and presented in Fig. 8 c and d. Figure 8 c and d show that the percentage departures of AEEMD-ANN predictions of extreme cases are closely matching with that of observed, while that of EEMD-ANN forecasts show significant deviation in all cases. This is more apparent in the case of predictions of extreme rainfalls of Kerala (see − 16.5% against − 31% for the observed data in the year 1981; − 3% against − 28.3% for the observed data in the year 2007, etc.). This close matching by AEEMD-ANN predictions is also found to be true in the case of extreme floods of Kerala in 2018, as discussed (see Fig. 7).Thus, especially in the predictions in extreme years, the percentage departures (from long term mean) by AEEMD-ANN forecasts are found to be closely matching with that of observed than the EEMD-ANN forecasts.

The proposed AEEMD-ANN method does not use any future information for the predictions unlike that of

Table 7 Error percentage on predicted values on rainfall extremes (drought and flood) of AISMR by the hindcast, forecast and adaptive forecast methods. The italicized values show the extreme among the drought and flood years

Year	Observed rainfall (mm)	AEEMD-ANN	Error (%)	Forecast	Error (%)	Hindcast	Error (%)
Drought years							
1972	652.80	613.33	6.04	629.81	3.52	NA	NA
1974	748.00	709.04	5.20	800.55	7.02	NA	NA
1979	707.70	749.25	5.87	728.62	2.95	764.41	0.08
1982	735.40	759.56	3.28	677.73	7.84	729.27	0.01
1986	743.00	720.50	3.02	690.44	7.07	708.63	0.04
1987	697.10	742.88	6.56	753.50	8.09	710.27	0.02
2002	*646.80*	*653.86*	*1.09*	*673.73*	*4.16*	*680.96*	*0.05*
2004	741.70	753.06	1.53	750.29	1.15	750.29	0.01
Flood years							
1971	886.8	946.71	1.66	928.44	3.55	NA	NA
1975	*962.7*	*1004.66*	*5.13*	*953.66*	*0.20*	*988.58*	*3.45*
1983	955.6	942.37	1.98	1054.45	9.66	964.22	0.28
1994	957.5	984.44	2.81	964.72	0.75	980.83	2.43
2007	947.1	1015.99	7.27	887.11	6.33	954.49	0.78
2013	947.3	968.95	2.28	958.45	1.17	965.75	− 1.94

hindcast experiment. It follows a recalibrating strategy while making the forecasts as the characteristics of the existing (calibration) model does not remain the same on addition of new datasets for the forecasts of successive time steps. The AEEMD-ANN method display reasonable accuracy in ISMR predictions could capture the years of hydrologic extremes in general and the drought years in particular. Hence the proposed method is robust, theoretically justified and more realistic when compared with the EEMD-ANN hybrid hindcast and forecast experiments.

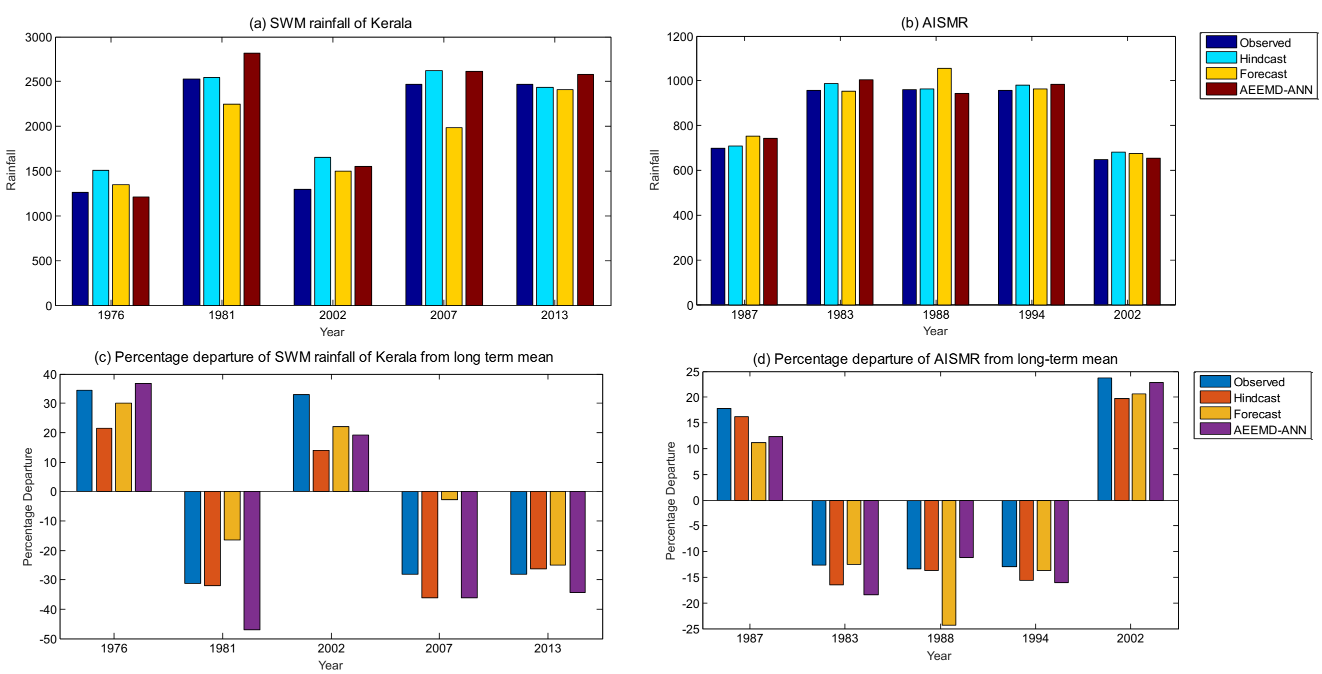

Fig. 8 Comparison of prediction of monsoon rainfall by the hindcast, forecast and AEEMD-ANN hybrid model with the observed data. **a** SWM rainfall of Kerala; **b** AISMR; **c** percentage departures of SWM rainfall of Kerala from long term mean; **d** percentage departures of AISMR from long term mean

However, more experiments need to be solicited in subsequent studies by considering the issues such as end effect, propagation of errors and data partitioning.

5 Conclusion

In this study, an adaptive EEMD-ANN hybrid model strategy is proposed for forecasting SWM rainfall in Kerala meteorological subdivision and for SWM rainfall of India. In the proposed framework of AEEMD-ANN model, the new dataset based on available forecast are appended on the existing datasets, and one-step-ahead forecasts are made using ANN. From the different statistical performance evaluation measures of the results, it is found that EEMD improves the accuracy of predictions by simple ANN model. Among the three hybrid models, AEEMD-ANN model performs better or equally well with the EEMD-ANN forecast model in different rainfall datasets. The AEEMD-ANN is more successful in capturing the drought years than the flood years, where it displayed remarkable accuracy than the EEMD-ANN forecast model. In short, the proposed framework can not only forecast the complex summer monsoon over India reasonably well but it can also capture the hydrologic extremes.

Acknowledgements The authors sincerely thank the Indian Institute of Tropical Meteorology Pune for making the data available in the public domain, benefitting non-commercial scientific research**.** The authors also thank India Meteorological Department (IMD) for providing the Kerala SWM data for the years 2017 and 2018 for supporting the work. The authors are also grateful to Amrita School of Arts and Sciences for the supports and research facilities provided.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Adarsh S, Janga Reddy M (2019) Evaluation of trends and predictability of short-term droughts in three meteorological subdivisions of India using multivariate EMD based hybrid modeling. Hydrol Process 33(1):130–143

Adarsh S, Janga Reddy M (2018a) Developing hourly intensity duration frequency curves for urban areas in India using multivariate empirical mode decomposition and scaling theory. Stochastic Environ. Res. Risk Assess. 32(6):1889–1902

Adarsh S, Janga Reddy M (2018b) Multiscale characterization and prediction of reservoir inflows using MEMD-SLR coupled approach. J Hydrol Eng 24:04018059. https://doi.org/10.1061/(asce)he.1943-5584.0001732

Adarsh S, Janga Reddy M (2017) Multiscale characterization and prediction of monsoon rainfall in India using Hilbert–Huang transform and time-dependent intrinsic correlation analysis. Meteorog Atmos Phys 130:667–688

Blandford HF (1884) On the connection of the Himalayan snowfall and season of droughts in India. Proc Royal Soc London 37:3–22

Bai Y, Wang P, Xie J et al (2015) Additive model for monthly reservoir inflow forecast. J Hydrol Eng 20:04014079. https://doi.org/10.1061/(asce)he.1943-5584.0001101

Fahimi F, Yaseen ZM, El-shafie A (2017) Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theoret Appl Climatol 128:875–903

Gadgil S, Srinivasan J (2011) Seasonal prediction of the Indian monsoon. Curr. Sci. 100:343–353

Ghosh S, Mujumdar PP (2007) Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment. Water Resour Res, AGU 43:W07405. https://doi.org/10.1029/2006WR005351

Hu J, Liu J, Liu Y, Gao C (2013) EMD-KNN model for annual average rainfall forecasting. J Hydrol Engng 18:1450–1457

Huang NE, Shen Z, Long SR et al (1998) The empirical mode decompositionand the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc R Soc London Series A454: 903–995

Huang NE, Wu Z, Long SR, Arnold KC, Chen Z, Blank K (2009) On instantaneous frequency. Adv Adaptive Data Anal 1(2):177–229

Huang S, Chang J, Huang Q, Chen Y (2014) Monthly streamflow prediction using modified EMD-based support vector machine. J Hydrol 511:764–775

Iyengar RN, Raghu Kanth TSG (2005) Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall. Meteorol Atmos Phys 90:17–36

Janga Reddy M, Adarsh S (2016) Time–frequency characterization of sub-divisional scale seasonal rainfall in India using the Hilbert–Huang transform. Stochastic Environ. Res. Risk Assess. 30(4): 1063–1085

Johny K, Pai M, Adarsh S (2019) Empirical forecasting and Indian Ocean dipole teleconnections of south–west monsoon rainfall in Kerala. Meteorol Atmos Phys 1–11 131(4):1055–1065

Karthikeyan L, Nagesh Kumar D (2013) Predictability of non-stationary time series using wavelet and EMD based ARMA models. J Hydrol 502:103–119

Kashid S, Maity R (2012) Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using genetic programming. J Hydrol 454-455:26–41

Kisi O, Latifoğlu L, Latifoğlu F (2014) Investigation of empirical mode decomposition in forecasting of hydrological time series. Wat Resour Manage 28(12):4045–4057

Kothawale DR, Rajeevan M (2017) Monthly, seasonal and annual rainfall time series for all-India, homogeneous regions and meteorological subdivisions: 1871–2016. Research Report No. RR-138. Indian Institute of Tropical Meteorology (IITM), Pune

Maheswaran R, Khosa R (2012) Comparative study of different wavelets for hydrologic forecasting. Comp Geosci 46:284–295

Meng E, Huang S, Huang Q et al (2019) A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. J Hydrol 568:462–478

Napolitano G, Serinaldi F, See L (2011) Impact of EMD decomposition and random initialisation of weights in ANN hindcasting of daily stream flow series: an empirical examination. J Hydrol 406:199–214

Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. J Hydrol 514:358–377

Pai M, Pramod K, Balchand A (2014) Long range forecast on south west monsoon rainfall using artificial neural networks based on clustering approach. Int J Inform Technol Comp Sci 6:1–8

Parthasarathy B, Munot AA, Kothawale DR (1994) All-India monthly and seasonal rainfall series: 1871–1993. Theor. Appl. Climatol. 49(4):217–224

Sahai A, Soman M, Satyan V (2000) All India summer monsoon rainfall prediction using an artificial neural network. Clim Dyn 16:291–302

Santos GCA, Silva GBLD (2014) Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models. Hydrol Sci J59:312–324

Singh P (2018) Indian summer monsoon rainfall (ISMR) forecasting using time series data: a fuzzy-entropy-neuro based expert system. Geosci Front 9(4):1243–1257

Singh P, Borah B (2013) Indian summer monsoon rainfall prediction using artificial neural network. Stochastic Environ. Res. Risk Assess. 27(7):1585–1599

Tan QF, Lei XH, Wang X et al (2018) An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. J Hydrol 567:767–780

Tao L, He X, Wang R (2017) A hybrid LSSVM model with empirical mode decomposition and differential evolution for forecasting monthly precipitation. J Hydrometeorol 18:159–176

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res.: Atmos. 106:7183–7192

Wang W-C, Xu D-M, Chau K-W, Chen S (2013) Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. J Hydroinf 15(4):1377–1390

Wang WC, Chau KW, Xu DM, Chen X-Y (2015a) Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Wat Resour Manage 29:2655–2675

Wang B, Xiang B, Li J, Webster PJ, Rajeevan MN, Liu J, Ha K-J (2015b) Rethinking Indian monsoon rainfall prediction in the context of recent global warming. Nat. Commun. 6:7154

Wang ZY, Qiu J, Li FF (2018) Hybrid models combining EMD/EEMD and ARIMA for long-term streamflow forecasting. Water 10:853

Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advadapt Data Anal 1:1–41

Zhang X, Peng Y, Zhang C, Wang B (2015) Are hybrid models integrated with data pre-processing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J Hydrol 530:137–152

Zhao X-h, Chen X (2015) Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting. Wat Resour Manage 29(8):2913–2926