Received: 11 December 2017 | Accepted: 17 October 2018
DOI: 10.1002/hyp.13316

RESEARCH ARTICLE

WILEY

Evaluation of trends and predictability of short-term droughts in three meteorological subdivisions of India using multivariate EMD-based hybrid modelling

Sankaran Adarsh[1,2] | Manne Janga Reddy[1]

[1] Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai, India

[2] Department of Civil Engineering, TKM College of Engineering, Kollam, India

Correspondence
Manne Janga Reddy, Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai-400076, India.
Email: mjreddy@civil.iitb.ac.in

Abstract

This paper presented trend analysis of droughts in Kerala, Telangana, and Orissa meteorological subdivisions in India and proposed a framework for drought prediction by employing the Empirical Mode Decomposition (EMD)-based prediction models. The study used 3-month standardized precipitation index (SPI3) for drought analysis. The trend analysis of SPI3 series for the period 1871–2012 using Mann–Kendall method showed statistically significant increasing trend in Kerala and Telangana subdivisions and a decreasing trend in Orissa subdivision. In addition, the non-linear trend component extracted from EMD showed statistically significant changes in all the three subdivisions. Then, the study proposed a hybrid approach for prediction of short-term droughts by coupling multivariate extension of EMD (MEMD) with stepwise linear regression (SLR) and genetic programming (GP) methods. First, the multivariate dataset comprising the SPI3 series of current and lagged time steps are decomposed using the MEMD. Then, SLR/GP models are developed to predict each subseries of SPI3 of desired time step considering the subseries of predictor variables at the corresponding timescales as inputs. The resulting models at different timescales are recombined to obtain the SPI3 values of the desired time step. The method is applied for prediction of short-term droughts in the three subdivisions. The results obtained by the hybrid models are compared with that obtained by conventional prediction models using M5 Model Trees and GP. The rigorous performance evaluation based on multiple statistical criteria clearly exhibited the superiority of the hybrid approaches (i.e., prediction models with MEMD-based decomposition over models without decomposition) for prediction of SPI3 in three subdivisions. Further, the study found that MEMD-GP model performs marginally better than the MEMD-SLR model due to its efficacy in modelling high frequency modes.

KEYWORDS

decomposition, droughts, GP, MEMD, prediction, trends

1 | INTRODUCTION

Drought is one of the less attended but damaging natural calamities that occurs in many parts of the world annually. Indian economy is highly reliant on agricultural production, which in turn heavily relies on monsoon rainfall. The delayed onset of southwest monsoon and decline in rainfall could lead to droughts and severely impact the agricultural production and economy of India. Meteorological drought is defined as the deficiency of precipitation from expected or normal precipitation over a specific period of time (Institute of Food and

Agricultural Sciences [IFAS], 1998). Rajeevan, Pai, Dikshit, and Kelkar (2004) categorized the seasonal rainfall amount into five categories. Based on these categories, a seasonal rainfall amount, which is characterized by less than 90% of long period mean of seasonal rainfall data, is considered as drought. The reduction in rainfall may lead to crop failure, soil degradation, and even to subsequent desertification. Therefore, analysis of spatial extent of droughts and drought prediction are important for agricultural planning and management. In recent past, few studies focused on analysing the variability and risks of droughts in different parts of India (Ganguli & Janga Reddy, 2013, 2014; Janga Reddy & Ganguli, 2012; Joshi, Gupta, Suryavanshi, Adamowski, & Madramootoo, 2016; Thomas, Nayak, & Ghosh, 2015; Thomas & Prasannakumar, 2016). Over the years, many univariate and multivariate indices were proposed for modelling and risk assessment of droughts (Kao & Govindaraju, 2010; Mishra & Singh, 2011; Zargar, Sadiq, Naser, & Khan, 2011). In several studies, the characterization of drought is made based on precipitation data, and the Standardized Precipitation Index (SPI) is the most widely used drought indicator owing to the flexibility of timescale, the stability of the spatial structure, the requirement of fewer input variables, and the simplicity in its computation (Bazrafshan, Hejabi, & Rahimi, 2014). SPI of different aggregated timescales (say 3, 6, 12, and 24 months) can indicate short-term as well as long-term droughts, and the results have significant implications on different components of hydrologic cycle. The 3-month SPI can be used as a seasonal drought index to represent short-term drought, and soil moisture may be more sensitive to a 3-month SPI representing the water stress and crop failures. The medium-term drought like SPI-6 influences the streamflow variability, and the long-term drought like SPI-12 influences the groundwater component (Janga Reddy & Ganguli, 2012; NDMC, 2006). In the past, several studies have performed SPI-based drought prediction considering a pure time series approach or cause effect approach (in which other climate variables or indices are considered along with lagged time step values of SPI as inputs) using simple regression, data-driven techniques, or their hybrid variants involving wavelets (Belayneh, Adamowski, Khalil, & Ozga-Zielinski, 2014; Belayneh, Adamowski, Khalil, & Quilty, 2016; Ganguli & Janga Reddy, 2014; Morid, Smakhtin, & Bagherzadeh, 2007). Even though many SPI-based drought prediction models were developed in different parts of the world, only very few studies have focused on Indian region. Mishra, Desai, and Singh (2007) developed a hybrid drought forecast model combining autoregressive integrated moving average model and feed forward recursive multistep artificial neural network model in the Kangsabati River basin, India. Ganguli and Janga Reddy (2014) developed ensemble drought prediction models for Western Rajasthan region in India using a coupled support vector machine–copula framework. The study investigated the teleconnections of large-scale climate oscillations such as El Niño Southern Oscillation (ENSO), Indian Ocean Dipole Mode, and Atlantic Multidecadal Oscillation on regional droughts (represented by SPI-6), and separate prediction models were developed for annual and seasonal periods.

For modelling and prediction of droughts, few studies have used the wavelet transform for developing decomposition-based hybrid models (Belayneh et al., 2014, 2016), but for applying the wavelet transforms as a data preprocessing technique, the type of mother wavelet and decomposition level are to be specified “a priori,” which is somewhat a challenge for the modeller. Even though, few researchers have suggested guidelines for the selection of most appropriate wavelet type and level of decomposition, but many of them lack proper scientific reasoning, are qualitative and empirical in nature, and cannot be easily implemented in practice (Sang et al., 2016). In this context, decomposition techniques were proposed for time series decomposition and the Empirical Mode Decomposition (EMD) proposed by Huang et al. (1998) is a purely data adaptive decomposition method, which is gaining popularity for practical time series prediction problems. The method being data adaptive and evolves from the characteristics of the time series, it does not demand for the a priori selection of functional form and decomposition level. Hence, in recent times, hybrid models involving EMD are gaining popularity for hydrological prediction (Huang, Chang, Huang, & Chen, 2014; Karthikeyan & Nagesh Kumar, 2013; Napolitano, Serinaldi, & See, 2011; Wang, Xu, Chau, & Chen, 2013), but they were not applied for the prediction of droughts.

In most of the hybrid modelling procedures involving EMD, it first decomposes the time series of concern into different timescales, and then appropriate number of lagged series of individual components is considered as predictors to model the component of the target variable. The final aggregation of the predicted components provides the information on predictant. However, many of the studies did not account the use of multiple causal inputs for prediction of hydrologic variables. It was also noticed that EMD being data adaptive in nature, the decomposition of time series of different predictor variables may result in different number of orthogonal components and cause in “mode misalignment” (Huang, Su, Kareem, & Liao, 2016). This may impose difficulties for developing the predictive models in practice. The multivariate extension of EMD (MEMD) proposed by Rehman and Mandic (2010) has the ability to align “common scales” present within a given multivariate dataset, and hence, it can solve the problem of mode misalignment, and its use for practical problems in hydrology becomes easier. Few studies used the orthogonal components obtained from decomposition directly as inputs in a cause–effect modelling framework for hydrological predictions (Wang et al., 2013; Zhu et al., 2016), which does not really capture the significant information from the different process scales. By accounting these shortcomings (use of multiple input variables and ability to capture the “significant” information from different process scales), this paper proposed an alternative framework for prediction of SPI, by coupling the MEMD with genetic programming (GP)/stepwise linear regression (SLR) methods. The objectives of the present study are (a) to evaluate the long-term trends in SPI3 series of three meteorological subdivisions of India using Mann–Kendall (MK) test and EMD methods and (b) to propose an alternate framework for prediction of short-term droughts by coupling MEMD with GP/SLR methods and test its applicability for the drought prediction in three meteorological subdivisions in India.

The rest of the paper is organized as follows: In Section 2, the theoretical background of SPI, EMD and its multivariate extension, and the details of proposed methodology are presented. In Section 3, description of case study and datasets used in the study are presented. In Section 4, the details of model application and results obtained by proposed method are discussed. The main conclusions drawn from the study are presented in Section 5.

2 | MATERIALS AND METHODS

This section presents the details of SPI, EMD, MEMD, and the proposed methodology for prediction of SPI.

2.1 | Standardized Precipitation Index

Over the years, different types of drought indices have been proposed to quantify the impact of droughts (Zargar et al., 2011). These indices differ in the key meteorological or agricultural parameters considered to define the drought. For example, SPI considers only precipitation as the variable, Normalized Difference Vegetation Index considers the vegetation characteristics, Palmer Drought Severity Index considers temperature and precipitation, and Standardized Precipitation and Evaporation Index uses precipitation and evaporation as the key parameters in estimation of drought. Out of the different indices, SPI, proposed by McKee, Doesken, and Kleist (1993), is the most popular one, as it uses the easily available precipitation as only input for its estimation. SPI is commonly used to identify meteorological droughts, and its computation from the monthly precipitation series can be performed using the following steps (Janga Reddy & Ganguli, 2012).

a. Prepare the aggregated precipitation series for specified aggregation timescale (say 3 months, 6 months, 12 months, etc.)
b. Fit the cumulative distribution function (CDF) of aggregated precipitation series using the appropriate probability distribution such as Gamma distribution.
c. Modify the fitted CDF as $F_X(x) = q + (1 - q)G_X(x)$ to account the zero values, as the two parameter Gamma distribution is not defined for zero precipitation values (McKee et al., 1993), where q is the probability of zero precipitation obtained from historical records, $G_X(x)$ is the CDF of nonzero precipitation records, and $F_X(x)$ is the CDF of actual precipitation series.
d. Perform an equiprobability transformation between CDF of mixed distribution $F_X(x)$ and standard normal distribution. This transformed probability gives the SPI for given aggregation timescale, that is, $Z = \Psi^{-1}(F_X(x))$, where $\Psi^{-1}(.)$ is the inverse of the CDF.

2.2 | EMD and its multivariate extension

EMD is a data-adaptive decomposition algorithm that separates a time series into a set of high and low frequency oscillatory modes of specific periodicity. The process of extracting the Intrinsic Mode Functions (IMFs) from a time series $X(t)$ (which is called as "sifting" process) consists the following steps:

1. Identify all extrema (maxima and minima) of the signal $X(t)$.
2. Connect these maxima points to construct an upper envelope ($E_{\max}(t)$) and minima points to construct a lower envelope ($E_{\min}(t)$) using suitable interpolation function (say, cubic spline).
3. Compute the mean of the upper and lower envelope, $m(t)$.
4. Calculate the difference time series $d(t) = X(t)$-$m(t)$.
5. Let $d(t)$ be the new signal and repeat Steps 1 to 4 until $d(t)$ becomes a zero-mean series with no riding waves (i.e., there are no negative local maxima and positive local minima) with smoothened amplitudes.

To satisfy Step 5, an appropriate criterion is to be applied to stop the sifting iterations in order to guarantee that the IMF retains enough physical sense of both amplitude and frequency modulations (Huang & Wu, 2008). A number of stopping criteria have been reported in the literature (Huang et al., 1998; Huang & Wu, 2008). One popular criteria is the modified Cauchy-type stopping criterion (Huang & Wu, 2008) computed from two consecutive sifting results as

$$\frac{\sum_{t=0}^{T} \left| d_{k,i-1}(t) - d_{k,i}(t) \right|^2}{\sum_{t=0}^{T} \left| d_{k,i-1}(t) \right|^2} \le \xi, \tag{1}$$

where "k" is the index for IMF, "i" is the index for iteration of the sifting operation (to get kth IMF), T is the data length, $d_{k,i}(t)$ is the deviation of the original time series from the mean in the ith iteration to evolve the kth IMF, and "ξ" is the tolerance value specified by the user (normally, 0.2–0.3 as per Huang & Wu, 2008).

6. On satisfying (a) the zero-mean condition and (b) number of zero crossings in the series and total number of extrema points differ at the most by one, $d(t)$ can be designated as the first intrinsic mode function IMF1.
7. Compute the residue $R_1(t)$ by subtracting IMF1 from original signal (i.e., $R_1(t) = X(t) - IMF_1(t)$) and is used as new signal. The sifting process is repeated upon $R_1(t)$ to get IMF2. Similarly, the higher oscillatory modes can be obtained by treating the residue ($R_k(t)$) as the signal ($X(t)$), iteratively, where kth residue is defined as $R_k(t) = X(t) - \sum_{j=1}^{k} IMF_j(t)$. The process will be continued till the resulting residue is a monotonic function or a function having only one extrema.

Then the original signal can be reconstructed as

$$X(t) = \sum_{k=1}^{K} [IMF_k(t)] + R_K(t),$$

where K is the number of decomposed IMFs.

EMD being empirical in characteristics, the number of modes obtained by this procedure will be not only decided based on the data length alone but also on the characteristics of the dataset (Huang, Schmitt, Lu, & Liu, 2009). Like other decomposition algorithm, EMD also acts as a filter bank, but unlike the popular discrete wavelets, it need not be truly "dyadic." The maximum number of modes can be expected to be $\log_2(N)$, where N is the data length (Flandrin, Rilling, & Gonçalvés, 2004). EMD is basically a univariate process, when multiple time series are involved, the process is to be separately applied upon the individual series. But such application does not guarantee equal number of modes for all series (referred as "mode misalignment" problem), as the decomposition is dependent on characteristics of the dataset, which may induce difficulties in developing regression models connecting the oscillatory modes of predictors and predictant in hybrid decomposition models.

Rehman and Mandic (2010) presented multivariate EMD as an extension to the conventional EMD, which decomposes multiple time series simultaneously after identifying the common scales inherent in different time series of concern. In this method, multiple envelops are produced by taking projections of multiple inputs along different directions in an *m*-dimensional space.

Let $V(t) = \{v_1(t), v_2(t), \ldots, v_m(t)\}$ be the *m* vectors as a function of time *t*, and $X^{\varphi_k} = \{x_1{}^k, x_2{}^k, \ldots, x_m{}^k\}$ denote the direction vector along different directions given by angles $\varphi_k = \{\varphi_1{}^k, \varphi_2{}^k, \ldots \varphi_{m-1}{}^k\}$ in a direction set *X* ($k = 1, 2, 3, \ldots, K$; *K* is the total number of directions). Different schemes are available for generating the direction vectors and Hammersley–Halton sampling sequence, which uses prime numbers is the most popular among them (Huang et al., 2016; Rehman & Mandic, 2010).

The IMFs of *m* temporal datasets can be obtained by the following steps:

1. Generate a suitable set of direction vectors by sampling on a (m-1) unit hypersphere.
2. Calculate the projection $p^{\varphi_k}(t)$ of the datasets *V*(*t*) along the direction vector X^{φ_k} for all *k*.
3. Find temporal instants $t_i^{\varphi_k}$ corresponding to the maxima of projection for all *k*.
4. Interpolate $[t_i^{\varphi_k}, V(t_i^{\varphi_k})]$ to obtain multivariate envelop curves $e^{\varphi_k(t)}$ for all *k*.
5. The mean of envelope curves (*M*(*t*)) is calculated by $M(t) = \frac{1}{K}\sum_{k=1}^{K} e^{\varphi_k}(t)$.
6. Extract the "detail" *D*(*t*) using $D(t) = V(t) - M(t)$. If *D*(*t*) fulfils the stoppage criterion for a multivariate IMF, apply the above procedure from Step 1 onwards upon the residue series (i.e., $V(t)-D(t)$). Otherwise, repeat the Steps 2–6 upon the series *D*(*t*).

It can be noted that the rotational modes appear as the counterparts of the oscillatory modes in EMD, and the stoppage criteria proposed for EMD can be used in the implementation of MEMD (Huang & Wu, 2008).

2.3 | MEMD-based hybrid models for prediction of SPI3

The present study used MEMD for the decomposition of the datasets and the SLR/GP for building the regression models for each of the components. The proposed methodology involves the following steps:

1. Decompose the multivariate dataset comprising the output (predictant) variable and different inputs (predictor variable) using MEMD to get different orthogonal subseries, each one is associated with specific timescale of variability.
2. Build SLR/GP models to predict subseries mode as a function of the corresponding subseries of different input variables, after identifying the insignificant components (based on *P*-value statistic) and discarding such components.
3. Predict the modes of output variable at different timescales by the refined models.
4. Combine the predicted modes to get the output at observation scale.

In the developed MEMD–SLR/GP hybrid model, the *i*th orthogonal mode of SPI is a function of *i*th orthogonal mode of different predictor variables PV_1, PV_2, …, PV_N, that is,

$$OM_{SPIi} = f(OM_{PV1i}, OM_{PV2i}, OM_{PV3i}, \ldots, OM_{PVNi}), \quad (2)$$

and finally, the output at the measurement scale (SPI) can be obtained as

$$O = \sum_{i=1}^{M} OM_{SPIi},$$

where *OM* denotes an orthogonal mode (an IMF or residue), *M* is the total number of decomposed modes, and *O* is the output variable.

2.4 | Stepwise linear regression

The SLR scheme of linear regression adds or removes the terms from multiple linear regression based on their statistical significance in a regression process. The method begins with an initial model and then compares the explanatory power of incrementally larger and smaller models. At each step, the *P* value of an *F* statistic is computed to test models with and without a potential term. If a term is not currently in the model, the null hypothesis is that the term would have a zero coefficient if added to the model. If there is sufficient evidence to reject the null hypothesis, the term is added to the model. In this study, an IMF of an input was added to the regression equation when the value of *P* is less than 0.05 and was taken out from the regression equation when the value of *P* is greater than 0.10.

2.5 | GP and model tree

GP proposed by Koza (1992) belongs to the family of data driven techniques, which provide a non-linear regression connecting the input variables and output variable of a prediction problem without specifying the functional form a priori. It follows a "tree"-based modelling strategy based on the principle of genetics and theory of evolution. In GP modelling, a function set comprising basic arithmetic operators (+, −, and *) and conditional statement ("if-else") are used. Further, to assess the efficacy of the method, M5 model tree based models are also developed to predict SPI3 by considering the same input parameters. Model tree does not involve any algorithm specific control parameter, which splits the domain to a set of subdomains based on standard deviation reduction criteria, and separate linear models are developed for each subdomain. More details on the M5 model tree algorithm can be found elsewhere (Quinlan, 1992; Witten, Frank, & Hall, 2005).

For MEMD method, the study used the MATLAB-based EMD package provided by Danilo Mandic (https://www.commsp.ee.ic.ac.uk/~mandic/research/emd.htm) with appropriate modifications to the codes. For the implementation of model tree algorithm, the authors used M5 Prime lab (http://www.cs.rtu.lv/jekabsons/regression.html) in MATLAB, and for the implementation of GP method, the authors used the MATLAB-based GP Lab software (http://gplab.sourceforge.net/). To develop the hybrid models, the codes are written in MATLAB programming and used in the study.

3 | STUDY AREA AND DATASETS

In this study, three sensitive meteorological subdivisions in India, namely, Kerala, Telangana, and Orissa are considered for analysing the trends and prediction of short-term droughts. The location map of the study area is shown in Figure 1.

The state of Kerala (8°N–12°N, 74°E–77°E), popularly known as "gateway of Indian monsoon," is one of the highest rainfall receiving regions in India. Kerala is having a coastal belt of more than 700 km, and the climate of Kerala is highly influenced by the circulations in Arabian Sea and Indian Ocean. In the past, many researchers have investigated the trends of rainfall datasets in Kerala, and reported a reduction in monsoon rainfall in Kerala (Adarsh & Janga Reddy, 2015; Krishnakumar, Rao, & Gopakumar, 2009; Thomas & Prasannakumar, 2016), and also, many of the districts of Kerala were declared as drought-affected areas by the state government in the recent past (during the years of 2012–2017). However, there were some exceptional heavy rainfalls like in the year 2018. One of the major reasons for the spatio-temporal variability of the monsoon rainfall in India is identified to be the interaction of synoptic disturbances developed over the Bay of Bengal and its movement towards a northwesterly direction along the monsoon trough (Mohapatra, Mohanty, & Behera, 2003). The Orissa subdivision (17°–22°N latitude and 81°–87°E longitude) falls in the weather forming zone of Bay of Bengal (within 15°–25°N latitude belt), which makes it more vulnerable to the changes in climate. The vulnerability of the state of Orissa is evident from the frequent occurrence of climate extremes in the form of droughts, floods, super cyclone, and heat waves in the past few decades. Therefore, the Orissa subdivision is considered to be a sensitive region for the hydroclimatic studies. Telangana subdivision (15°N–19°N, 77°E–82°E) is one of the low rainfall receiving regions (semiarid characteristics) in the peninsular India and is not having any oceanic proximity, and the region is also declared as drought affected region by state government several times in the recent past.

The monthly rainfall data of Kerala, Telangana, and Orissa subdivisions for the period 1871–2012 was collected from Indian Institute of Tropical Meteorology Pune (htpp://www.tropmet.res.in) and used for determining SPI3 values of the three regions. In the past, several researchers have used these datasets for spatio-temporal trend analysis of rainfall and droughts over different regions in India (Adarsh & Janga Reddy, 2015; Ganguli & Janga Reddy, 2013; Joshi et al., 2016; Thomas & Prasannakumar, 2016). The time series of the monthly rainfall data of the three subdivisions are presented in Figure 2.

4 | RESULTS AND DISCUSSION

4.1 | Trend analysis of SPI3 series

In the present study, for representing the short-term drought, SPI3 is selected, and the time series of SPI3 of the three regions are prepared

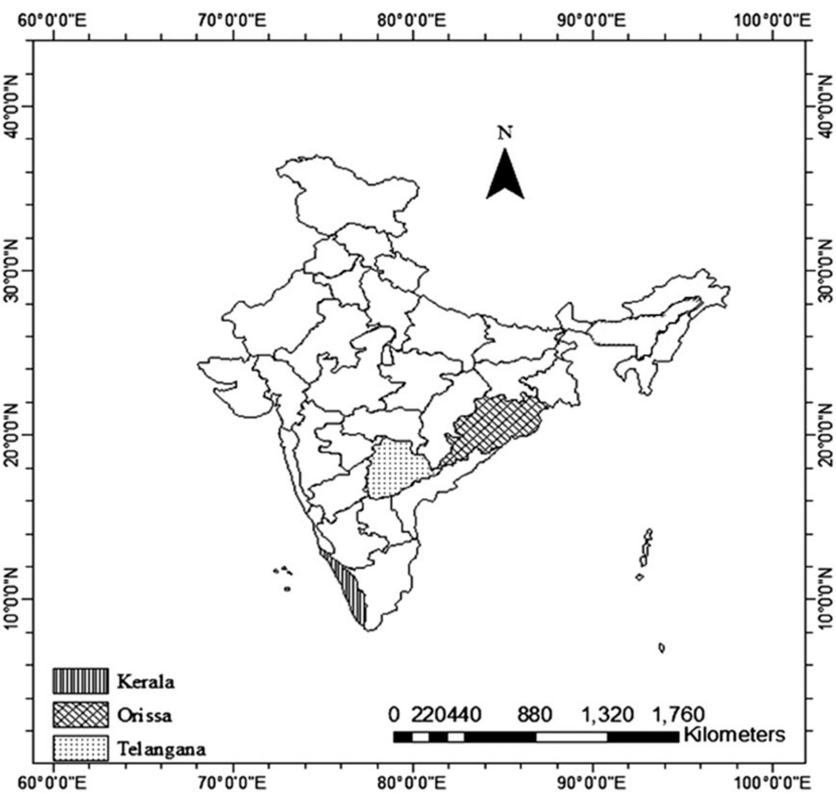

FIGURE 1 Location map of three meteorological subdivisions in India

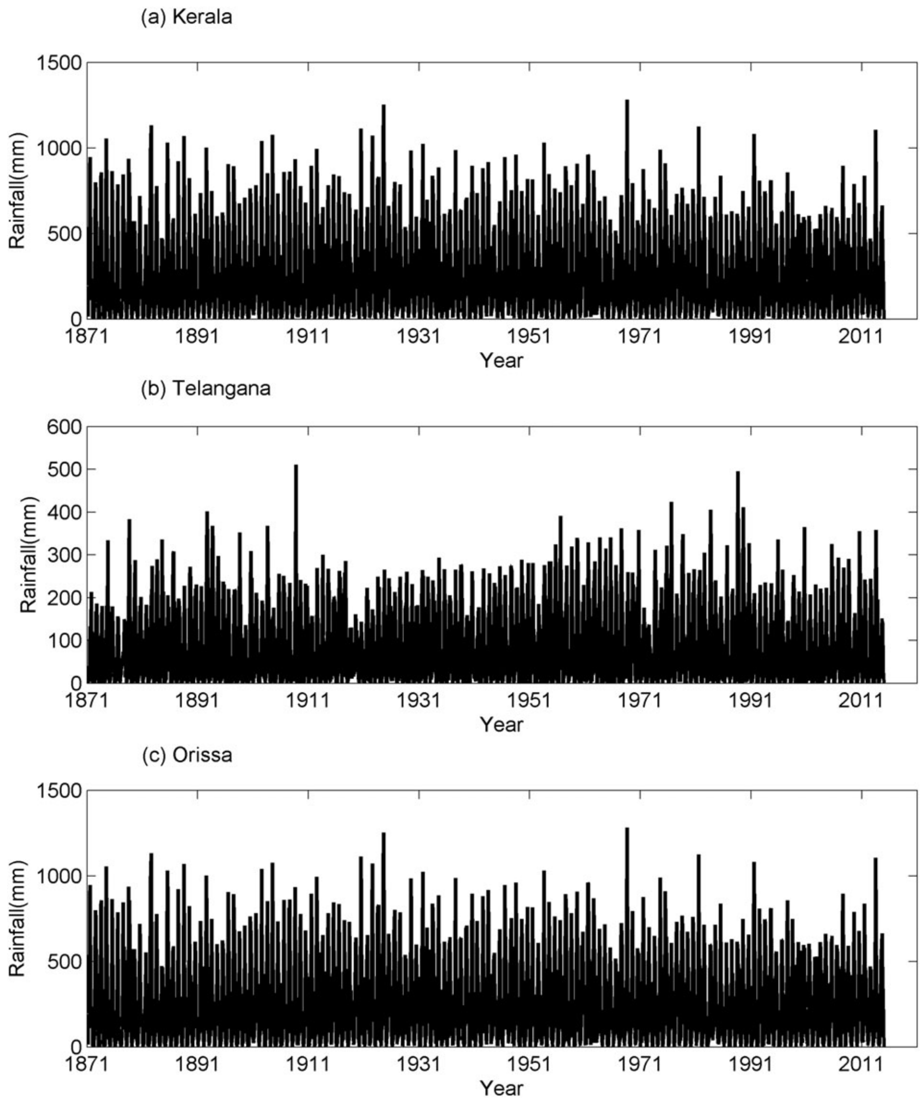

FIGURE 2 Time series plots of monthly rainfall data of three subdivisions (a) Kerala, (b) Telangana, and (c) Orissa

from monthly rainfall datasets of 1871–2012 period. To detect the trends, the modified MK test (Hamed & Rao, 1998) was applied, and the MK test statistics of the three SPI3 series are computed at 5% significance level. The *Z* values are found to be 1.961, 3.241, and −0.162, respectively, for Kerala, Telangana, and Orissa subdivisions. Hence, significantly increasing trend noticed in the SPI3 series of Kerala and Telangana subdivisions, whereas a decreasing trend (not significant) is noticed in the SPI3 series of Orissa subdivision, as the *Z* values of the SPI time series of Kerala and Telangana subdivisions equals or exceed the critical value of 1.96 at 5% significance level. Further, the linear trend in these series is estimated and depicted in Figure 3. It is noticed that the intercept values (upper panels of Figure 3) in all the SPI3 series is closer to zero, which infers that the possibility of non-linear trends cannot be ignored. Few recent studies reported that the non-linear trend is significant in many hydrometeorological datasets, and EMD method is a useful method for detecting such non-linear trends (Carmona & Poveda, 2014; Sang, Wang, & Liu, 2014).

The EMD method can portray the shape of the non-linear trend, and its statistical significance can be evaluated by the statistical significance test proposed by Wu and Huang (2005). The main steps involved in the test are (i) the computation of energy density of modes and its normalization by considering first IMF as the reference IMF; (ii) generation of white noise series by Monte Carlo simulations, its EMD, and computation of the "spread function"; and (iii) estimation of the confidence band of spread function of white noise at a selected significance level, based on step (ii). Then, the points with coordinates of mean normalized energy and mean period of modes are located in the 2D plane corresponding to all modes. This enables a comparison of the energy level of different modes with the spread function of white noise. The modes of original time series that have their energy densities located above the upper significance line of the white noise series can be considered to be statistically significant at the selected significance level. More details of the test can be found elsewhere (Wu & Huang, 2005).

The "residue" obtained by the decomposition of the SPI3 series provides the non-linear trend, which is provided in the middle panels of Figure 3, and its statistical significance lines are shown in the lower panels of Figure 3. The dotted and solid lines show the upper significance lines at 1% and 5% significance levels, respectively. The plots of residue show that there is a monotonically increasing trend in SPI3 series of Kerala subdivision. The residue of SPI3 series of Telangana show an increasing trend up to ~1960, and thereafter, it

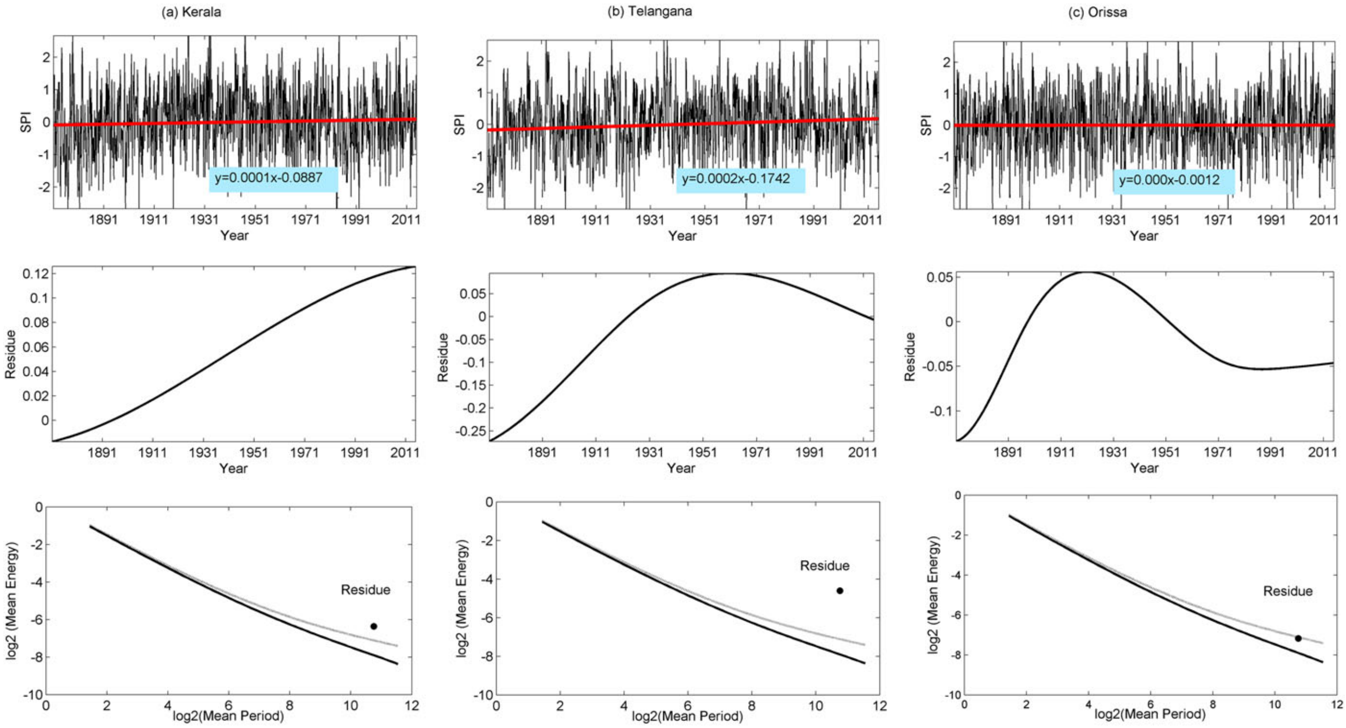

FIGURE 3 The linear trend and Empirical Mode Decomposition (EMD)-based non-linear trend of SPI3 series of different subdivisions (a) Kerala, (b) Telangana, and (c) Orissa. The upper panels shows the linear trend; middle panel shows the residue component obtained by EMD of SPI3 series; the lower panels shows the results of statistical significance test of residue component obtained by EMD of different SPI3 series, in which the dotted and solid lines show the upper significance lines at 1% and 5% significance levels, respectively

decreases, whereas SPI3 series of Orissa show a long-term declining trend (since ~1930s), and all these changes are statistically significant at 5% significance level. Moreover, the sequential change in trend of SPI3 series for three subdivisions is estimated using sequential Mann–Kendall (SQMK) method. The MK test helps in determination of trend and its statistical significance of time series, but it is to be recalled that the nature of the trend and its significance differ with the data spell considered for trend analysis. Hence, it is important to examine the temporal changes of trend and identification of trend turning points or commencement year of statistically significant trend in time series. The SQMK test proposed by Sneyers (1990) is an advancement to the MK test, which helps for performing such analysis. The SQMK method considers the ranks of original time series for estimation of sequential change in the trend. Here, the test statistic is defined based on the pairwise comparison of entities of the ranked series, and the SQMK values are computed in both progressive (from beginning to end of the series) and retrograde (vice versa) manner. The SQMK test estimates a sequence of MK values (for different time instants), and the intersection of progressive and retrograde series indicates a trend turning point. More details of this test can be found elsewhere (Adarsh & Janga Reddy, 2015; Gerstengarbe & Werner, 1999; Modarres & Sarhadi, 2009). The results of SQMK test of the SPI3 series for the three subdivisions are presented in Figure 4.

From the SQMK plot of SPI3 series of Kerala subdivision (Figure 4a), it is noticed that a trend turning point exists at 1890, and from this year, an increasing trend is noticed, and the MK value becomes significant since 1910. From the SQMK plot of SPI3 of Telangana subdivision (Figure 4b), it is clear that a significant trend prevails for more than century (1875–2012). In the SQMK plot of SPI3 series of Orissa subdivision (Figure 4c), two trend turning points are noticed—one in 1905 and the second one in 1968. The SPI3 shows an increasing trend since 1905 and became significant for more than a decade during ~1945–1960, then there exists a decreasing trend since 1980 except for a shorter time spell during 2005–2008. Also the nonparametric Pettit test (Pettitt, 1979) is invoked for detection of single change point within the selected period of study. It detected change points in 1915 for Kerala and Telangana, which is in agreement with the fact that MK values become significant since 1910. But, for the SPI3 of Orissa, Pettit test could not capture a change point at 5% significance level.

4.2 | Drought prediction using MEMD-GP/SLR framework

For the prediction of SPI3 series using MEMD-GP/SLR framework, first, the multivariate dataset is prepared considering the lagged SPI3 values as predictor variables and current SPI3 as predictant. The partial autocorrelation function (PACF) analysis of the time series is used for selecting the appropriate predictor variables. In this analysis, it is assumed that x_i as the output variable on the condition that the PACF at lag k is out of the [−1.96/√n, 1.96/√n] range, where n is the data length (and the estimate is at 95% confidence interval); x_{i-k} is one of the input variable. If none of the PACF coefficients is out of the 95% confidence interval, the previous one value x_{i-1} is selected as the input variable, and the procedure for computation of PACF

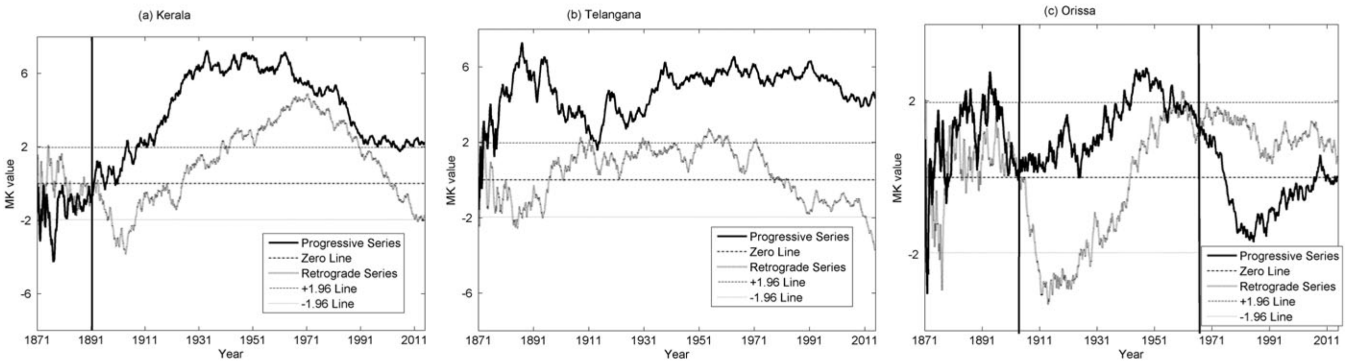

FIGURE 4 Results of SQMK test of SPI3 series of three subdivisions (a) Kerala, (b) Telangana, and (c) Orissa. The vertical bars indicate the year of trend turning points

coefficients can be found elsewhere (Huang et al., 2014). The PACF plots of SPI3 series from the three subdivisions are presented in Figure 5.

From the PACF plots (Figure 5), it is clear that the first four lags of SPI3 series have significant correlation for all the three regions. Thus, the models take a general functional form of

$$SPI3(t) = f(SPI3(t-1), SPI3(t-2), SPI3(t-3), SPI3(t-4)). \quad (3)$$

The study developed prediction models using MEMD-based hybrid modelling framework by keeping the first 70% of the data for calibration of the models and the remaining 30% of the data for validation of the models. The SPI3 of a generic time t along with four lagged values of SPI forms the multivariate dataset for the drought prediction modelling in all the three subdivisions. The decomposition process should be performed in such a way that it should not lead to over iteration and over decomposition, for which suitable threshold parameters can be used to control the magnitude of fluctuations of the series and the number of sifting iterations (Rilling, Flandrin, & Goncalves, 2003). The lower and upper threshold parameters θ_1 and θ_2 evaluate the relative magnitude of amplitude of the mean in comparison with the amplitude of the corresponding modes. The thresholds should be selected carefully, because imposing a too low or very high threshold may lead to overiteration and hence overdecomposition. Hence, in this selection process, the parameters are supposed to control the fluctuations in such a way that it simultaneously ensures overall small fluctuations in mean even while accounting the locally large fluctuations. In this process, the evaluation function $\varepsilon(t) = |m(t)/A(t)|$, (where $A(t)$ is mode amplitude $A(t) = (E_{max}(t) - E_{min}(t))/2$ and $m(t)$ is local mean of envelope curves) is executed until $\varepsilon(t) < \theta_1$ for a specified fraction $(1 - \alpha)$ of total duration and $\varepsilon(t) < \theta_2$ for remaining fraction. The fraction α that controls the number of "sifting" iterations is chosen as 0.075, lower and upper threshold parameters are fixed as 0.075 and 0.75, based on the guidelines available in literature (Hu & Si, 2013) to avoid the chances of any over decomposition. On applying MEMD, it decomposed the multivariate datasets into 11 modes for Kerala and Telangana subdivisions, whereas it resulted in 12 modes for the data from Orissa subdivision. As the expected number of modes is nearly log 2(*data length*) (Flandrin et al., 2004), the decomposition is appropriate for monthly time series of SPI for 142 years. A sample plot of decomposition of dataset (SPI3 series) for Orissa subdivision is provided in Figure 6.

In the first case, SLR models are prepared for prediction of different rotatory mode of SPI3 of a generic time t, considering modes of predictor variable at the same scale as inputs. The resulting regression coefficient matrix obtained by SLR methods for prediction of different modes of SPI3 of the three subdivisions Kerala, Telangana, and Orissa are presented in Table 1.

From the regression coefficients of SPI3 prediction for the three subdivisions (Table 1), it is noticed that the components of first lag of SPI3 series have a positive influence on the respective modes of SPI3, except for IMF1. For Kerala subdivision, the coefficients of second and fourth lag components are having a negative influence in the respective modes of SPI(t). Here, except for IMF1, third lag component has a positive influence on the respective components of SPI(t). In the other two subdivisions, such uniform pattern

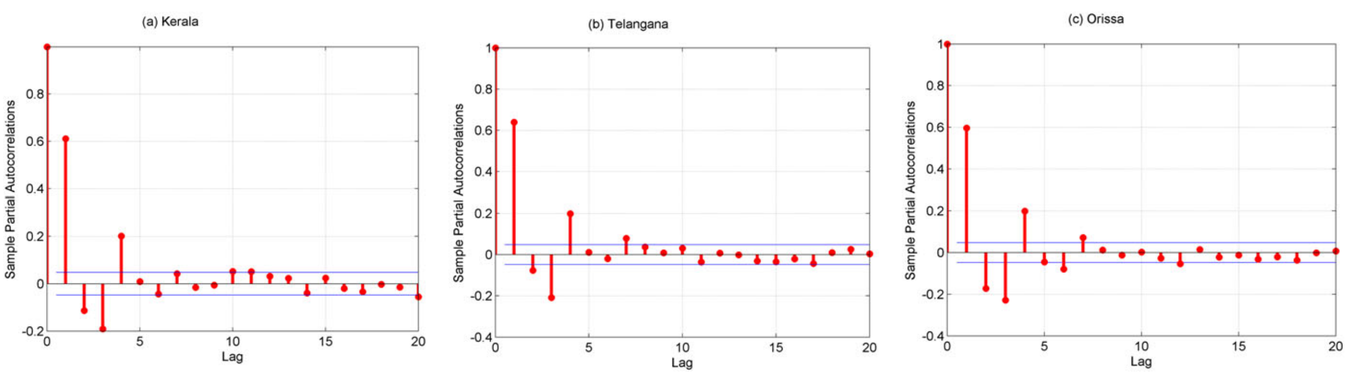

FIGURE 5 The partial autocorrelation function plots of the SPI3 series of three subdivisions: (a) Kerala, (b) Telangana, and (c) Orissa

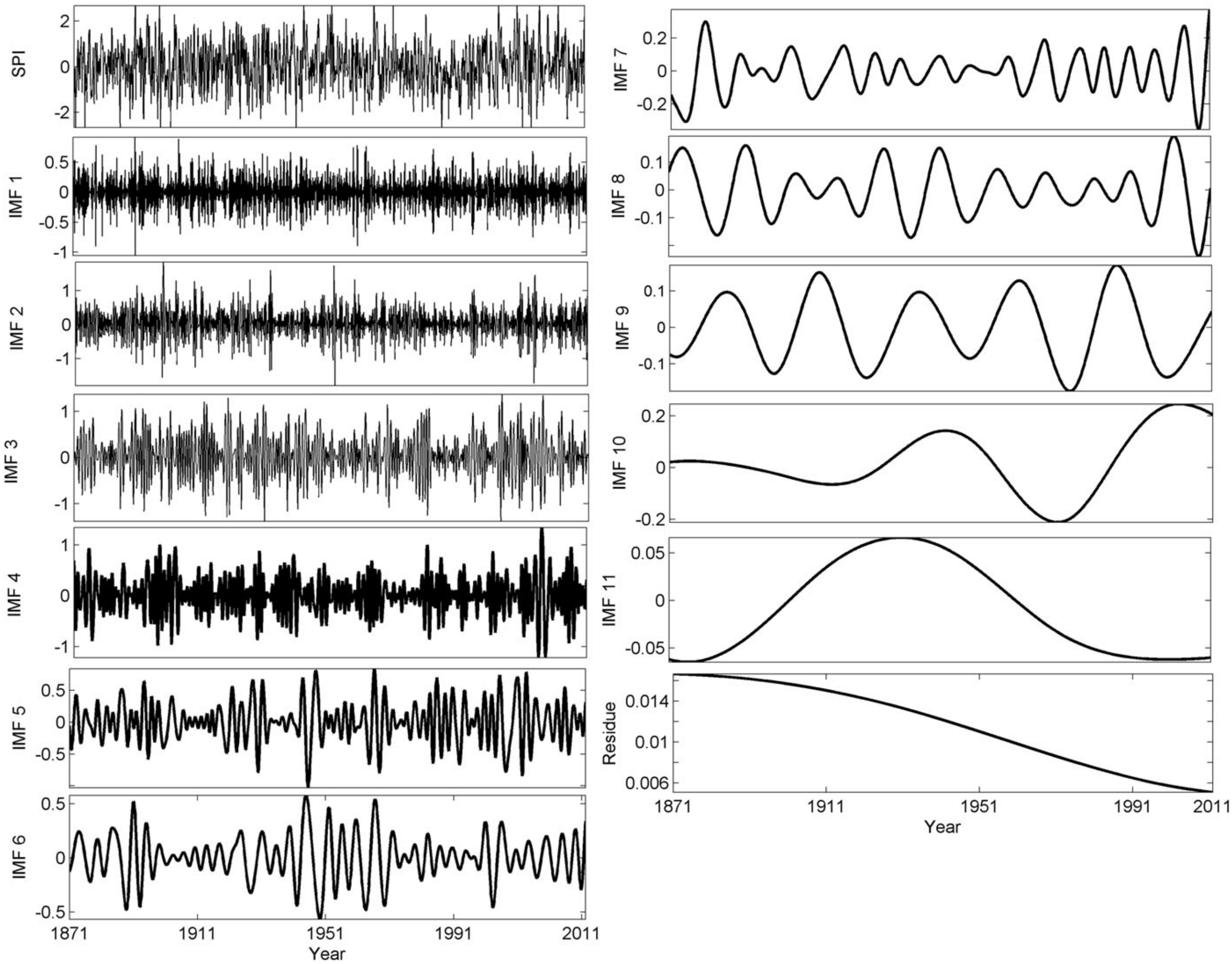

FIGURE 6 The components obtained by the decomposition of SPI3 of Orissa subdivision

is not observed except for the inference made for the coefficients of components of first lag of SPI3 series. Also, from Table 1, it is noticed that at certain process scales, certain input variables are not influential, for example, in IMF5 of SPI3 of Telangana SPI3(*t-3*) and SPI3 (*t-4*) are not influential and in IMF6 of Orissa, SPI3(*t-2*) and SPI3 (*t-3*) are not influential.

The study investigated different methods for prediction of droughts. In the SLR, purely linear models are used for prediction of components of SPI at different process scales. Next, instead of SLR, the GP is used for developing the non-linear regression models for each of the components. Also to comment on the efficacy of both of the hybrid models, a purely non-linear method (GP) and a semilinear method (M5 model tree) are adopted for prediction of SPI. The performance evaluation of different methods is carried out using the popular evaluation measures such as correlation coefficient (R), Nash–Sutcliffe efficiency (Nash & Sutcliffe, 1970), root mean square error, and mean absolute error. The results of performance evaluation of different methods are presented in Table 2. Also, the density scatter plots of SPI3 predictions by different methods are presented in Figures 7–9, respectively, for Kerala, Telangana, and Orissa subdivisions. From the visual examination of density scatter plots (Figures 7–9), it is clear that the clustering of points are along the ideal fit line for both of the MEMD-based hybrid models, and more deviation is noticed for the GP- and model tree-based model predictions.

From Table 2, it is noticed that the correlation statistics are much higher (>0.95) in all cases for predictions using both of the hybrid MEMD-SLR and MEMD-GP methods when compared with that by other methods, and also, the error statistics are considerably low. Also, it is noticed that the MEMD-GP hybrid model performs marginally better than MEMD-SLR hybrid model. For example, the resulting R, Nash–Sutcliffe efficiency, and root mean square error values for predictions of SPI3 for validation data of Kerala subdivision are 0.982, 0.96, and 0.141, respectively, by MEMD-GP model against the 0.978, 0.953, and 0.161 by the MEMD-SLR model. It is to be noted that only few regression coefficients are brought to zero during the MEMD-SLR model (Table 1), which infers all of the input parameters are influential at different timescales in the prediction of SPI3, which is reflected in the improved performance of MEMD-GP model. Further, to examine this, the R^2 statistics of prediction of different IMFs of SPI3 series of the subdivisions by using the two hybrid models are presented in Table 3.

From Table 3, it can be noticed that both of the models perform well in prediction of higher order modes (from Mode 3 onwards) with high R^2 statistics (>0.96) for the validation data in all cases. This high

TABLE 1 Regression coefficient matrix for SPI3 prediction of Kerala, Telangana, and Orissa subdivisions

Subdivision	Mode number of SPI3(*t*)	Values of regression coefficients for mode number of SPI3(*t-1*)	SPI3(*t-2*)	SPI3(*t-3*)	SPI3(*t-4*)
Kerala	IMF1	−1.421	−1.312	−1.079	−0.506
	IMF2	0.865	−1.549	0.607	−0.650
	IMF3	2.021	−2.097	1.044	−0.365
	IMF4	2.019	−1.525	0.453	−0.104
	IMF5	2.033	−1.379	0.411	−0.125
	IMF6	2.284	−1.654	0.351	**−0.024**
	IMF7	1.616	−0.840	0.663	−0.402
	IMF8	1.445	−0.566	0.350	−0.214
	IMF9	2.215	−2.123	0.908	**−0.049**
	IMF10	5.105	−6.542	3.716	−1.348
	Residue	3.581	−3.981	1.800	−0.428
Telangana	IMF1	−1.720	−1.699	−1.236	−0.495
	IMF2	0.637	−1.429	0.422	−0.605
	IMF3	2.001	−2.319	1.341	−0.506
	IMF4	2.074	−1.525	0.320	**0.031**
	IMF5	1.880	−0.939	**−0.001**	**0.004**
	IMF6	1.742	−0.863	0.279	−0.191
	IMF7	1.502	−0.272	−0.234	**−0.010**
	IMF8	1.566	−0.880	0.350	**0.018**
	IMF9	1.544	−0.769	0.390	−0.184
	IMF10	0.229	2.826	−1.352	−0.771
	Residue	0.162	2.937	−1.303	−0.860
Orissa	IMF1	−1.871	−1.990	−1.469	−0.586
	IMF2	0.908	−1.521	0.627	−0.586
	IMF3	1.947	−1.896	0.816	−0.245
	IMF4	2.107	−1.574	0.355	**0.017**
	IMF5	1.972	−1.000	−0.209	0.193
	IMF6	1.277	**−0.176**	**−0.001**	−0.087
	IMF7	1.576	−0.737	−0.064	0.110
	IMF8	1.561	−1.284	0.870	−0.175
	IMF9	1.077	0.708	−1.879	1.083
	IMF10	3.414	−2.512	4.891	−4.759
	IMF11	1.692	−1.183	3.226	−2.737
	Residue	0.429	0.701	−3.747	3.588

Note. The bold numbers show that these coefficients are not significant at 5% significance level.

TABLE 2 Performance evaluation of different models for SPI3 predictions

		MEMD-SLR		MEMD-GP		MT		GP	
Subdivision	PEC	C	V	C	V	C	V	C	V
Kerala	*R*	0.978	0.978	0.983	0.982	0.737	0.640	0.679	0.639
	NSE	0.954	0.953	0.961	0.960	0.538	0.408	0.461	0.405
	RMSE	0.211	0.209	0.183	0.188	0.680	0.751	0.734	0.753
	MAE	0.160	0.161	0.139	0.145	0.531	0.594	0.576	0.587
Telangana	*R*	0.983	0.984	0.985	0.985	0.751	0.632	0.700	0.662
	NSE	0.968	0.965	0.970	0.966	0.561	0.393	0.489	0.435
	RMSE	0.183	0.170	0.172	0.172	0.652	0.743	0.702	0.717
	MAE	0.137	0.130	0.128	0.128	0.499	0.578	0.547	0.551
Orissa	*R*	0.979	0.976	0.982	0.977	0.740	0.651	0.677	0.664
	NSE	0.958	0.953	0.967	0.961	0.544	0.423	0.458	0.438
	RMSE	0.200	0.217	0.186	0.213	0.658	0.763	0.718	0.754
	MAE	0.151	0.161	0.141	0.154	0.511	0.593	0.552	0.587

Note. PEC: Performance Evaluation Criteria; C: calibration; V: validation; GP: genetic programming; MT: model tree; RMSE: root mean square error; NSE: Nash–Sutcliffe efficiency; MAE: mean absolute error.

R^2 values also infer that the process are more deterministic at higher timescales, and both of the models are equally capable in capturing the variability of the higher order modes. But in the prediction of the two lower order modes, relatively lesser R^2 statistics is noticed, and for the prediction of these modes, the MEMD-GP method is relatively better when compared with the MEMD-SLR method, which may be the reason for marginal improvement in performance of overall prediction for SPI3 using the former method when compared with the latter one. Thus, it is evident that MEMD-based hybrid models are effective in modelling short-term droughts when compared with the conventional models (semilinear model tree approach and non-linear GP approach).

It is to be noted that the presented modelling strategy is general one and applicable for prediction of any hydrologic variable

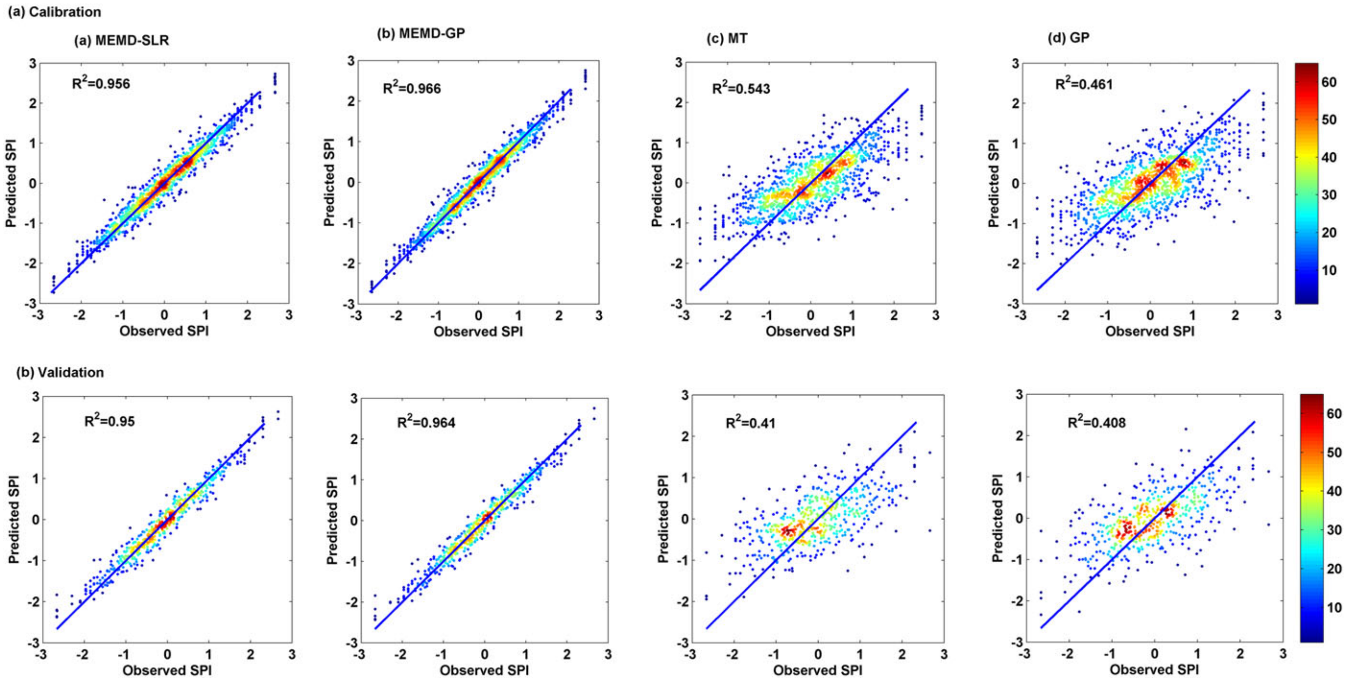

FIGURE 7 Comparison of SPI3 predictions by MEMD-SLR, MEMD-GP, M5 model tree, and GP methods with the observed data of Kerala subdivision during calibration and validation periods. Upper panels show the density scatter plots for calibration dataset, and lower panels show the density scatter plots for validation dataset. In colour bar, red (blue) colour indicates higher (lower) density of points. MEMD: multivariate extension of Empirical Mode Decomposition; SLR: stepwise linear regression; GP: genetic programming

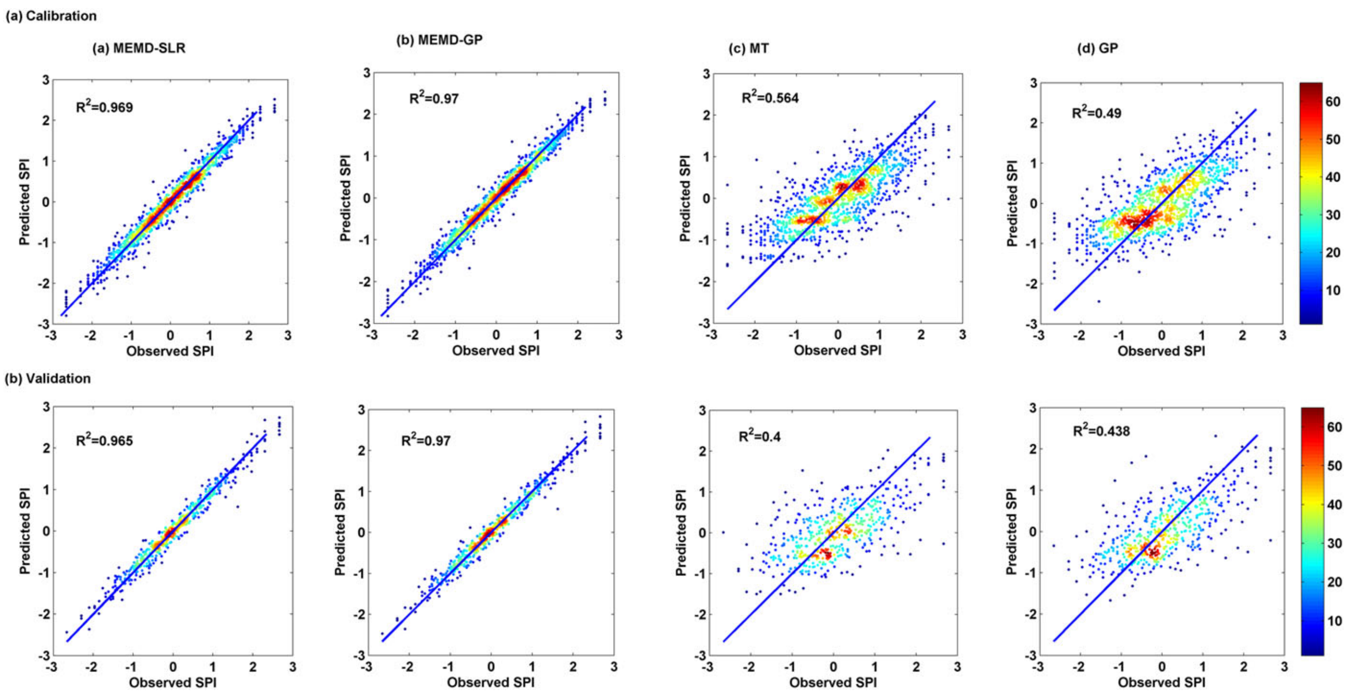

FIGURE 8 Comparison of SPI3 predictions by MEMD-SLR, MEMD-GP, M5 model tree, and GP methods with the observed data of Telangana subdivision during calibration and validation periods. Upper panels show the density scatter plots for calibration dataset, and lower panels show the density scatter plots for validation dataset. In colour bar, red (blue) colour indicates higher (lower) density of points. MEMD: multivariate extension of Empirical Mode Decomposition; SLR: stepwise linear regression; GP: genetic programming

considering any input combinations. Thus, the presented approach is robust, accounts multiple inputs, and able to capture the information from specific timescales. In the present study, the methodology followed a pure time series approach, in which only lagged SPI values are considered as model inputs. However, it is evident that the climatic variables such as ENSO can influence the droughts in India (Ganguli & Janga Reddy, 2014). The MEMD-based methodology is also very much flexible in including such climatic variables in addition

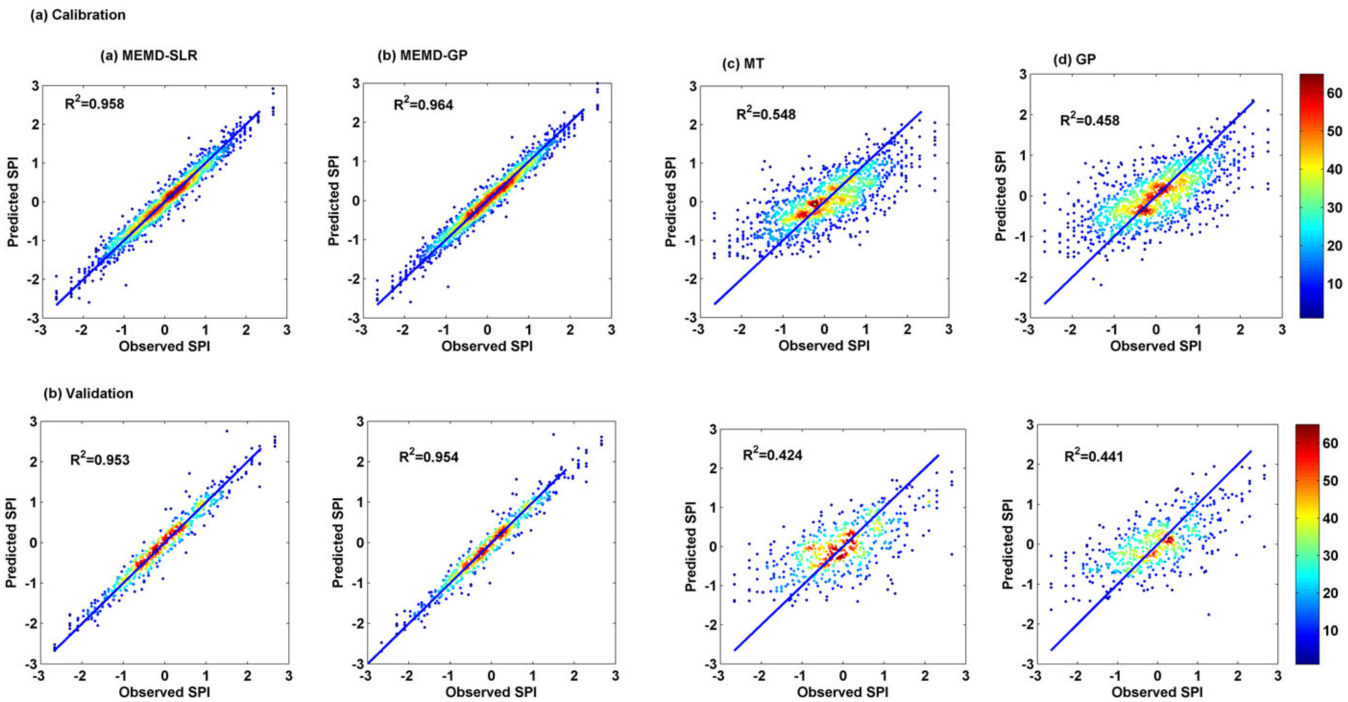

FIGURE 9 Comparison of SPI3 predictions by MEMD-SLR, MEMD-GP, M5 model tree, and GP methods with the observed data of Orissa subdivision during calibration and validation periods. Upper panels show the density scatter plots for calibration dataset, and lower panels show the density scatter plots for validation dataset. In colour bar, red (blue) colour indicates higher (lower) density of points. MEMD: multivariate extension of Empirical Mode Decomposition; SLR: stepwise linear regression; GP: genetic programming

TABLE 3 Comparison of R^2 of different MEMD-SLR and MEMD-GP models in prediction of orthogonal modes of SPI3 series of the three subdivisions

	Kerala				Telangana				Orissa			
	MEMD-SLR		MEMD-GP		MEMD-SLR		MEMD-GP		MEMD-SLR		MEMD-GP	
IMF number	C	V	C	V	C	V	C	V	C	V	C	V
IMF1	0.84	0.78	0.85	0.89	0.89	0.76	0.88	0.84	0.89	0.76	0.89	0.87
IMF2	0.88	0.87	0.90	0.90	0.89	0.88	0.91	0.89	0.90	0.88	0.90	0.90
IMF3	0.97	0.95	0.97	0.97	0.97	0.98	0.96	0.95	0.98	0.97	0.98	0.97
IMF4	0.98	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.98	0.98	0.98	0.99
IMF5	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
IMF6	1.00	0.99	1.00	1.00	0.99	0.97	1.00	0.98	0.99	0.98	0.99	0.99
IMF7	1.00	0.98	1.00	0.99	0.99	0.98	1.00	1.00	0.99	0.99	0.99	0.99
IMF8	1.00	1.00	1.00	0.99	1.00	0.96	1.00	1.00	1.00	0.96	1.00	0.99
IMF9	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00
IMF10	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	0.99
Residue/IMF11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Residue									1.00	1.00	1.00	1.00

Note. C: calibration; V: validation; MEMD: multivariate extension of Empirical Mode Decomposition; SLR: stepwise linear regression; GP: genetic programming.

to the SPI values, only the size of multivariate dataset increases. The indices (or lagged series) representing the phenomenon like ENSO are to be decomposed first, by including such series also as one of the variable in addition to the lagged SPI values for prediction of SPI at a generic time period "*t*." The MEMD is capable of identifying the common scales inherent in the different variables of a multivariate dataset and helps to build the linear or non-linear regression models at a specified process scales by considering the EMD profiles (IMF component or residue) of such indices as inputs.

The better performance of MEMD hybrid models in prediction of drought may be attributed to its capability to identify the less significant predictors in different timescales and subsequent exclusion, which cannot be executed by conventional models. However, more experiments on the partitioning of the datasets calibration and validation, the treatment of issues such as "end effect" through extension methods, and so on are to be solicited to apply the method for real time forecasting of droughts.

5 | CONCLUSIONS

In this study, first nonparametric modified MK test and EMD methods are applied for extracting the long term trends (1871–2012) in SPI3 series of Kerala, Telangana, and Orissa subdivisions in India. The trend analysis using MK method displayed significant increase in the SPI3 of Kerala and Telangana meteorological subdivisions, whereas that of Orissa subdivision showed an insignificant decreasing trend. The EMD method portrayed the non-linear trend in the SPI3 series of the three subdivisions, and the trend is found to be statistically significant in the past century. The SQMK test of the SPI3 series of Telangana showed a consistent statistically significant increasing trend during the study period (1871–2012), whereas an early trend turning point is detected in 1890s in the SPI3 series of Kerala subdivision. Two trend turning points are noticed (in 1910 and 1965) in the SPI3 series of Orissa subdivision, and it has a declining trend since 1980s. Moreover, this study proposed two hybrid models involving MEMD and GP/SLP methods for prediction of short-term drought index SPI3. MEMD is used for separation of multiple causal input variables into equal number of orthogonal modes each associated with specific timescales. The results obtained by the MEMD-based hybrid models involving GP and SLR showed substantial improvement in prediction of SPI3 for all the three subdivisions when compared with GP and M5 model tree models, which is attributed to the plausible inclusion of appropriate inputs at different timescales. Further, it is noted that the MEMD-GP hybrid model performs marginally better than MEMD-SLR model due to its skill in capturing the variability of high frequency modes.

ORCID

Manne Janga Reddy http://orcid.org/0000-0001-7003-8651

REFERENCES

Adarsh, S., & Janga Reddy, M. (2015). Trend analysis of rainfall in four meteorological subdivisions of southern India using non-parametric methods and wavelet analysis. *International Journal of Climatology*, *35*(6), 1107–1124. https://doi.org/10.1002/joc.4042

Bazrafshan, J., Hejabi, S., & Rahimi, J. (2014). Drought monitoring using the multivariate standardized precipitation index (MSPI). *Water Resources Management*, *28*(4), 1045–1060. https://doi.org/10.1007/s11269-014-0533-2

Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash river basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, *508*, 418–429. https://doi.org/10.1016/j.jhydrol.2013.10.052

Belayneh, A., Adamowski, J., Khalil, B., & Quilty, J. (2016). Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmospheric Research*, *172–173*, 37–47.

Carmona, A. M., & Poveda, G. (2014). Detection of long-term trends in monthly hydro-climatic series of Colombia through Empirical Mode Decomposition. *Climatic Change*, *123*(2), 301–313. https://doi.org/10.1007/s10584-013-1046-3

Flandrin, P., Rilling, G., & Gonçalvés, P. (2004). Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters*, *11*(2), 112–114. https://doi.org/10.1109/LSP.2003.821662

Ganguli, P., & Janga Reddy, M. (2013). Evaluation of trends and multivariate frequency analysis of droughts in three meteorological subdivisions of Western India. *International Journal of Climatology*, *34*(3), 911–928.

Ganguli, P., & Janga Reddy, M. (2014). Ensemble prediction of regional droughts using climate inputs and the SVM–copula approach. *Hydrological Processes*, *28*(19), 4989–5009. https://doi.org/10.1002/hyp.9966

Gerstengarbe, F. W., & Werner, P. C. (1999). Estimation of the beginning and end of recurrent events within a climate regime. *Climate Research*, *11*, 97–107. https://doi.org/10.3354/cr011097

Hamed, K. H., & Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, *204*, 182–196. https://doi.org/10.1016/S0022-1694(97)00125-X

Hu, W., & Si, B.-C. (2013). Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition. *Geoderma*, *193-194*, 180–188. https://doi.org/10.1016/j.geoderma.2012.10.021

Huang, G., Su, Y., Kareem, A., & Liao, H. (2016). Time-frequency analysis of non-stationary process based on multivariate empirical mode decomposition. *Journal of Engineering Mechanics 142*(1), https://doi.org/10.1061/(ASCE)EM.1943-7889.0000975, 04015065.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., … Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proceedings of Royal Society London, Series A*, *454*, 903–995.

Huang, N. E., & Wu, Z. (2008). A review on Hilbert Huang Transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, *46*(2). https://doi.org/10.1029/2007RG000228

Huang, S., Chang, J., Huang, Q., & Chen, Y. (2014). Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*, *511*, 764–775. https://doi.org/10.1016/j.jhydrol.2014.01.062

Huang, Y., Schmitt, F. G., Lu, Z., & Liu, Y. (2009). Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis. *Journal of Hydrology*, *373*, 103–111. https://doi.org/10.1016/j.jhydrol.2009.04.015

IFAS (1998). *Extreme heat and drought: The disasters handbook*. Orlando: University of Florida.

Janga Reddy, M., & Ganguli, P. (2012). Application of copulas for derivation of drought severity-duration-frequency curves. *Hydrological Processes*, *26*, 1672–1685. https://doi.org/10.1002/hyp.8287

Joshi, N., Gupta, D., Suryavanshi, S., Adamowski, J., & Madramootoo, C. A. (2016). Analysis of trends and dominant periodicities in drought variables in India: A wavelet transform based approach. *Atmospheric Research*, *182*, 200–220. https://doi.org/10.1016/j.atmosres.2016.07.030

Kao, S. C., & Govindaraju, R. S. (2010). A copula-based joint deficit index for droughts. *Journal of Hydrology*, *380*(2010), 121–134. https://doi.org/10.1016/j.jhydrol.2009.10.029

Karthikeyan, L., & Nagesh Kumar, D. (2013). Predictability of non-stationary time series using wavelet and EMD based ARMA models. *Journal of Hydrology*, *502*, 103–119. https://doi.org/10.1016/j.jhydrol.2013.08.030

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, Massachusetts: The MIT Press.

Krishnakumar, K. N., Rao, G. S. L. H. V. P., & Gopakumar, C. S. (2009). Rainfall trends in twentieth century over Kerala, India. *Atmospheric Environment*, *43*, 1940–1194. https://doi.org/10.1016/j.atmosenv.2008.12.053

McKee, T.B., Doesken, D.J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Proceedings of the 8th

conference on applied climatology, vol. 17, American Meteorological Society, Boston, MA, USA, pp. 179–184

Mishra, A. K., Desai, V. R., & Singh, V. P. (2007). Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering*, *12*(6), 626–638. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:6(626)

Mishra, A. K., & Singh, V. P. (2011). Drought modeling—A review. *Journal of Hydrology*, *403*(1–2), 157–175. https://doi.org/10.1016/j.jhydrol.2011.03.049

Modarres, R., & Sarhadi, A. (2009). Rainfall trends analysis of Iran in the last half of the twentieth century. *Journal of Geophysical Research*, *114*, D03101. https://doi.org/10.1029/2008JD010707

Mohapatra, M., Mohanty, U. C., & Behera, S. (2003). Spatial variability of daily rainfall over Orissa, India, during the southwest summer monsoon season. *International Journal of Climatology*, *23*, 1867–1887.

Morid, S., Smakhtin, V., & Bagherzadeh, K. (2007). Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology*, *27*, 2103–2111. https://doi.org/10.1002/joc.1498

Napolitano, G., Serinaldi, F., & See, L. (2011). Impact of EMD decomposition and random initialization of weights in ANN hindcasting of daily stream flow series: An empirical examination. *Journal of Hydrology*, *406*, 199–214. https://doi.org/10.1016/j.jhydrol.2011.06.015

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models, Part I—A discussion of principles. *Journal of Hydrology*, *10*(282–290), 1970.

NDMC (2006). Monitoring drought. The standardized precipitation index. Interpretation of SPI maps. National Climatic Data Center. http://www.drought.unl.edu/monitor/interp.htm

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, *28*, 126–135. https://doi.org/10.2307/2346729

Quinlan, J. R. (1992). Learning with continuous classes. In *Proceedings of Australian Joint Conference on Artificial Intelligence* (pp. 343–348). Singapore: World Scientific Press.

Rajeevan, M., Pai, D. S., Dikshit, S. K., & Kelkar, R. R. (2004). IMD's new operational models for long-range forecast of southwest monsoon. *Current Science*, *86*(3), 422–431.

Rehman, N., & Mandic, D. P. (2010). Multivariate empirical mode decomposition. *Proceedings of the Royal Society Series a*, *466*(2117), 1291–1302. https://doi.org/10.1098/rspa.2009.0502

Rilling G., Flandrin P., & Goncalves P. (2003). On empirical mode decomposition and its algorithms, In *Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado (Italy), pp. 8–11.

Sang, Y., Singh V.P., Sun F., Chen Y., Liu Y., & Yang M. (2016). Wavelet-based hydrological time series forecasting. *Journal of Hydrologic Engineering*, *21(5)*, 06016001, pp. 1–5, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001347.

Sang, Y. F., Wang, Z., & Liu, C. (2014). Comparison of the MK Test and EMD method for trend identification in hydrological time series. *Journal of Hydrology*, *510*, 293–298. https://doi.org/10.1016/j.jhydrol.2013.12.039

Sneyers, R. (1990). On the statistical analysis of series of observations, Tech. Note 143, 192 pp., Geneva, Switzerland

Thomas, J., & Prasannakumar, V. (2016). Temporal analysis of rainfall (1871–2012) and drought characteristics over a tropical monsoon-dominated State (Kerala) of India. *Journal of Hydrology*, *534*, 266–280. https://doi.org/10.1016/j.jhydrol.2016.01.013

Thomas, T., Nayak, P.C., & Ghosh, N.C. (2015). Spatio-temporal analysis of drought characteristics in the Bundelkhand region of Central India using the standardized precipitation index. *Journal of Hydrologic Engineering*, *20*(11), 05015004, pp. 1–12, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001189.

Wang, W. C., Xu, D. M., Chau, K. W., & Chen, S. (2013). Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *Journal of Hydroinformatics*, *15*(4), 1377–1390. https://doi.org/10.2166/hydro.2013.134

Witten, I. H., Frank, E., & Hall, M. A. (2005). *Data Mining: Practical machine learning tools and techniques*. San Francisco, CA, USA: Morgan Kaufmann.

Wu Z., & Huang N.E. (2005). Statistical significance test of intrinsic mode functions. In *Hilbert-Huang transform and its applications*. Edited by: Norden E Huang (NASA Goddard Space Flight Center, USA), Samuel S P Shen (University of Alberta, Canada). World Scientific Publishing Singapore.

Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. O. (2011). A review of drought indices. *Environmental Reviews*, *19*, 333–349. https://doi.org/10.1139/a11-013

How to cite this article: Adarsh S, Janga Reddy M. Evaluation of trends and predictability of short-term droughts in three meteorological subdivisions of India using multivariate EMD-based hybrid modelling. *Hydrological Processes*. 2018;1–14. https://doi.org/10.1002/hyp.13316