



Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com



Scale dependent prediction of reference evapotranspiration based on Multi-Variate Empirical mode decomposition



S. Adarsh*, Sulaiman Sanah, K.K. Murshida, P. Nooramol

TKM College of Engineering, Kollam 691005, Kerala, India

ARTICLE INFO

Article history:

Received 10 June 2016

Revised 23 September 2016

Accepted 18 October 2016

Available online 13 February 2017

Keywords:

Evapotranspiration

Time scale

Decomposition

MEMD

Regression

ABSTRACT

This study proposes a novel method for estimation of reference evapotranspiration (ET_o) by accounting the time scale of variability using the Multivariate Empirical Mode Decomposition (MEMD). First the ET_o and the four predictor variables such as solar radiation, air temperature, relative humidity and wind velocity are decomposed into different intrinsic mode functions (IMFs) and a residue using MEMD. To model ET_o , first the modes are modeled separately using the Stepwise Linear Regression (SLR) after identifying the significant predictors at different time scales based on the p -value statistics. Subsequently, the predicted modes are recombined to obtain ET_o at the observation scale. The method is demonstrated by predicting the monthly ET_o from Stratford station in United States. The results of the study clearly exhibited the superior performance of the proposed MEMD-SLR model when compared with that by M5 model tree, SLR and the EMD-SLR hybrid model.

© 2017 Ain Shams University. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Evapotranspiration (ET) is an important component of hydrologic cycle, which has direct and significant influence on the hydrology and agriculture of a region. The accurate estimation of ET is a challenging problem in water resources management because of the complex non-linear relationships between evapotranspiration and its controlling factors. The evapotranspiration rate from a hypothetical grass reference crop surface with specific characteristics is called reference evapotranspiration (ET_o) [1,2]. A number of methods have been proposed to model reference evapotranspiration [3–5] and most of these models are highly complex and depend on meteorological data such as temperature, solar radiation, wind speed and humidity [6]. A large number of equations are available for modeling ET_o , in which traditional physics based models (e.g., Penman-Montieth equation) or some empirical

models based on meteorological parameters (Hargreves-Samani equation or Blaney-Criddle equation) are still popular for estimation of ET_o [7,8]. In the last decade, data-driven techniques such as Artificial Neural Networks (ANN), M5 Model Tree, Support Vector Machine (SVM), Gene Expression Programming (GEP) are proven to be efficient substitute to the empirical models in determination of reference evapotranspiration [2,9–16]. Pal and Deswal [2] investigated the potential of M5 model tree in the estimation of daily reference ET_o by considering the data from Davis, California. They compared the results with those obtained FAO–56 Penman-Montieth equation and the calibrated Hargreaves-Samani relation. Then the developed model was applied to four different stations and proved the generalization capability of M5 model tree. Rahimi-khoob [16] performed the prediction of monthly ET_o of four stations in Iran using ANN and M5 model tree methods and compared the relative efficacy of these methods. It is found that M5 model tree is performing equally well with ANN and the latter method can provide simple linear relations and found to be much simple in implementation [16].

Evapotranspiration is controlled by different meteorological factors operating at different time scales [17,18]. Understanding the time scale dependent relationship between ET_o and meteorological factors is important for the accurate estimation of ET_o . To reveal the time scale dependent relationships between two or more variables Fourier spectral analysis can be used, but it performs well only for linear and stationary time series. Both evapotranspiration and most of the meteorological time series being

* Corresponding author at: Department of Civil Engineering, TKM College of Engineering, Kollam 691005, Kerala, India.

E-mail addresses: adarsh.sankaran@yahoo.in (S. Adarsh), sanah.sulaiman@yahoo.com (S. Sanah), murshidakk012@gmail.com (K.K. Murshida), mongamnoora@gmail.com (P. Nooramol).

Peer review under responsibility of Ain Shams University.



Production and hosting by Elsevier

non-linear non-stationary in characteristics, to analyze such series advanced methods such as wavelet analysis are more appropriate. Some recent studies proposed hybrid models involving discrete wavelet transform (DWT) for the prediction of evapotranspiration [19–22] while some of them attempted for the characterization of ET_o [17,18,23]. In the hybrid wavelet based models DWT is used as a data preprocessing or decomposition tool while a linear or non-linear regression methodology is used for the actual model preparation. Kisi [19] investigated the potential of Wavelet - linear regression hybrid model for the prediction of ET_o from three stations in United States while Partal [20] combined DWT and feed-forward ANN methods for reference evapotranspiration estimation. Cobaner [21] examined the efficacy of a wavelet regression (WR) model in estimating reference evapotranspiration by considering pan evaporation data as input. Gocic et al. [22] used four soft computing methods were analyzed: genetic programming (GP), SVM-FireFly Algorithm (SVM-FFA), ANN, and SVM-wavelet transform (SVM-Wavelet) hybrid model for estimation of ET_o data from Serbia. But in these studies wavelet components or recombined series were considered as inputs for preparation of model rather than preparing individual models for different time scales after identifying appropriate inputs at respective time scales. Also these studies didn't consider the exact time scale of variability of governing factors in the prediction of evapotranspiration. In the application of DWT, the time scales are assumed to be 'apriori' in powers of two (like 2, 4, 8 etc.), but the true (exact) time scales of the associated governing processes may be quite different from such pre-specified time scales. Even though the continuous wavelet transform (CWT) and wavelet packet analysis can solve this issue to a certain extent, the selection of mother wavelet function is a critical step in the analysis using CWT and it may alter the modeling results.

Huang et al. [24] proposed a purely data adaptive decomposition method namely Empirical Mode Decomposition (EMD) and their use in conjunction with the Hilbert transform resulted in a spectral analysis method called Hilbert Huang Transform (HHT). HHT is gaining popularity in characterization of time series signals in hydrology [25–29]. EMD is a purely data adaptive method, which doesn't demand the 'apriori' selection of basis function (like trigonometric function in Fourier and mother wavelet in Wavelet analyses). Also EMD decomposes a time series data into different Intrinsic Mode Functions (IMFs) by looking into the associated time scales inherent in the data. It is hypothesized that separation of variations in ET as well as in meteorological factors at similar time scales may provide enhanced information on the time scale-specific control and overall betterment in the prediction of evapotranspiration. Since its introduction, different variants of EMD evolved, out of which some were successful in improving the quality of decomposition by elimination of 'scale mixing' problems [30,31], while some of them enabled to perform decomposition of more than one variable simultaneously [32–34]. Rehman and Mandic [34] proposed a multivariate extension of EMD, which has the ability to align common time scales present within multivariate data. Here 'rotatory' modes are obtained by the decomposition in the place of oscillatory modes in the traditional EMD. Each "common time scale" is manifested in the common rotatory modes in all the variables within an n -variate data set. Such 'mode alignment' property helps to make use of similar time scales in different data sources and thus the obtained common time scales represent the true time scales of underlying processes [35,36]. This property is one of the most attractive features of MEMD, as it will simplify the additive modeling strategy adopted for complex problems [37]. Hu and Si [35] applied the MEMD method for investigating scale specific control of soil water storage in spatial scales and its prediction based on different environmental factors. According to the authors' knowledge, the prediction of ET_o based on the exact

time scale of variability has been never attempted by the researchers. This motivates the authors to frame the specific objectives of this study as follows: (i) examining the multiscaling property of reference evapotranspiration (ET_o) and its dominant meteorological parameters using MEMD; (ii) develop MEMD based stepwise linear regression (SLR) models to predict the ET_o and compare its performance with that of M5 model tree, SLR and EMD - SLR hybrid model.

The following section presents the theory of MEMD along with the methodology proposed in this paper to estimate ET_o . Thereafter the study area and data details are given in brief. After that the details of modeling, results and discussion are presented and subsequently the conclusion from the study is presented.

2. Materials and methods

The present study uses MEMD for decomposition of the datasets and SLR for model preparation. The algorithm of MEMD is briefly presented below. Subsequently the steps involved in modeling using the proposed MEMD-SLR hybrid method are presented.

2.1. Multivariate Empirical Mode Decomposition (MEMD)

Multivariate EMD [34] is an extension of the traditional EMD, which decomposes multiple time series in a single step after identifying the common time scales inherent in different time series of concern. The method involves (i) generation of suitable number of direction vectors; (ii) taking projections of multiple inputs along different directions in an n -dimensional space; (iii) identification of local maxima points, setting of multivariate envelope curves through them and subsequent computation of mean; (iv) iteratively extract the detail by taking the difference of the mean envelope curve from the original signal till satisfying a specified stopping criteria [38]. The fourth step will provide the IMF and considering the residue (difference between the signal and sum of all IMFs produced so far) series the steps from (ii) to be repeated till no residue can be extracted from the signal. The mathematical details of the algorithm can be found elsewhere [34–36].

2.2. Proposed methodology of modeling ET_o

The proposed methodology for modeling ET_o involves the following steps:

1. Decompose ET_o and the four predictor variables using MEMD to get different modes of specific time scale of variability.
2. Build SLR models to predict each mode of ET_o as a function of the corresponding mode of different predictor variables. SLR is a method in which the potential terms are included or removed based on their statistical significance in a regression process. More details on the algorithm can be found in [39].
3. Refine the model by discarding the modes of the predictor variables having the p -value is greater than 0.1. In this study, a mode of a parameter was added to the regression equation when the value of $p \leq 0.05$ (entrance tolerance) and was taken out from the regression equation when the value of $p \geq 0.10$ (exit tolerance).
4. Predict the modes at different time scales by the refined model.
5. Add the predicted modes to get the predicted ET_o at the measurement scale (monthly in this study).

The above built model is designated as MEMD-SLR model in this paper, while SLR model refers the stepwise linear regression model between ET_o and the meteorological parameters. The third model considered in this study is denoted as EMD-SLR, which decom-

poses ET_o alone and subsequently fits different SLR models (to predict the IMFs/residue of ET_o time series) by considering the different meteorological variables as predictors.

MEMD-SLR model takes general form $OM_{EToi} = \sum_{i=1}^4 r_i OM_{CFi}$ and $ET_o = \sum_{i=1}^N OM_{EToi}$

The EMD-SLR model takes general form $OM_{EToi} = \sum_{i=1}^4 r_i CF_i$ and $ET_o = \sum_{i=1}^N OM_{EToi}$

The SLR model takes the general form $OM_{EToi} = \sum_{i=1}^4 r_i CF_i$

where OM denote an orthogonal mode (IMF or residue), N is the total number of decomposed modes, CF is the controlling factor OM is the orthogonal mode (an IMF or residue) of the controlling factor r_i is the regression coefficient. The M5 model tree (MT) [40], is a promising data driven technique which can be used to capture non-linear predictor-predictant relationships. The technique is reported to be performing better or equally good with other data driven methods (such as ANN, SVM, GP) and at the same time it does not demand the setting of algorithm specific control parameters and can give simplified regression expressions [41]. To comment on the strength on the proposed approach the ET_o is also predicted by MT algorithm. Finally, the performance of the four models is evaluated (both for calibration and validation datasets) using different performance evaluation measures to judge the quality of overall prediction of ET_o considering the measured and predicted ET_o .

3. Study area and datasets

The California Irrigation and Management Information System (CIMIS) is a program developed in 1982 by the California Department of Water Resource and the University of California at Davis. The CIMIS manages a network of over 120 automated weather stations in the state of California. The solar radiation, air temperature, relative humidity and wind velocity are selected as the predictor variables for the study, as these four variables are reported to be most influential to ET_o process by many researchers in the past [2,42–44]. Total incoming Solar Radiation (R) (in W/m^2) measured using pyranometers at height of 2.0 m above the ground, air tem-

perature (T) (in $^{\circ}C$) measured at a height of 1.5 m above the ground using a thermistor, relative humidity (H) (in%) using humidity sensor kept along with temperature sensor, wind velocity (U) (in m/s) measured using three-cup anemometers at 2.0 m above the ground etc. used for observation. Along with the above observational data of Stratford station (36 $^{\circ}$ 09'27"N, 119 $^{\circ}$ 51'00"W) California, USA, ET_o values calculated using the CIMIS Penman method are collected from (<http://www.cimis.water.ca.gov/>) to demonstrate the applicability of proposed methodology. Data for the period 1994–2014 were used for preparing monthly scale model. The location map of the station is provided in Fig. 1.

4. Modeling, results and discussion

In this section, first the details of the decomposition of different time series by MEMD and scaling analysis of the IMF components are presented. Then the details of MEMD-SLR model are presented. It is followed by the details of EMD decomposition of ET_o time series and EMD-SLR modeling. After that, the performance evaluation of different models is made.

4.1. Multiscale characterization of ET_o time series based on MEMD

The ET_o time series along with the four meteorological parameters solar radiation, air temperature, relative humidity and wind velocity constitute the multivariate dataset for the present problem and this dataset is decomposed simultaneously into different modes using MEMD algorithm. The number of directions is chosen as 64 (as it should be greater than the dimensionality of the spatial data sets). The stopping criteria of sifting was by choosing the different tolerance parameters as 0.075, 0.75 and 0.075 respectively following the earlier studies [34,45]. The MEMD decomposition resulted in 7 IMFs and a residue. IMFs with lower order correspond to higher frequency (smaller time scale) oscillations, where as higher order IMFs corresponds to lower frequency (larger scale) oscillations. For a given multi-variate data series, the number of oscillations for a given IMF were generally the same for different variables. i.e., the periodicity of IMFs of different variables will also be nearly the same. The accurate time scales can be calculated

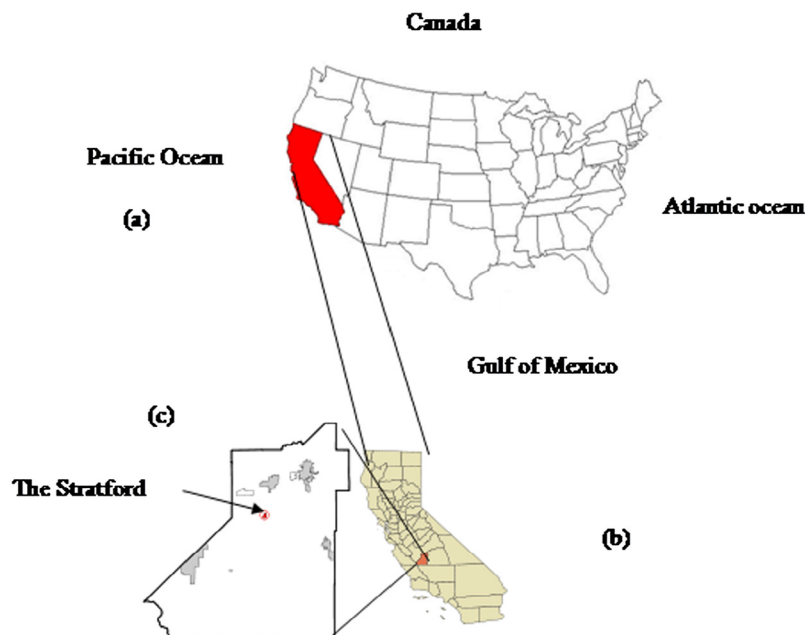


Fig. 1. Location map of Stratford station in the state USA (from right to left (a) USA; (b) California state; (c) Kings county).

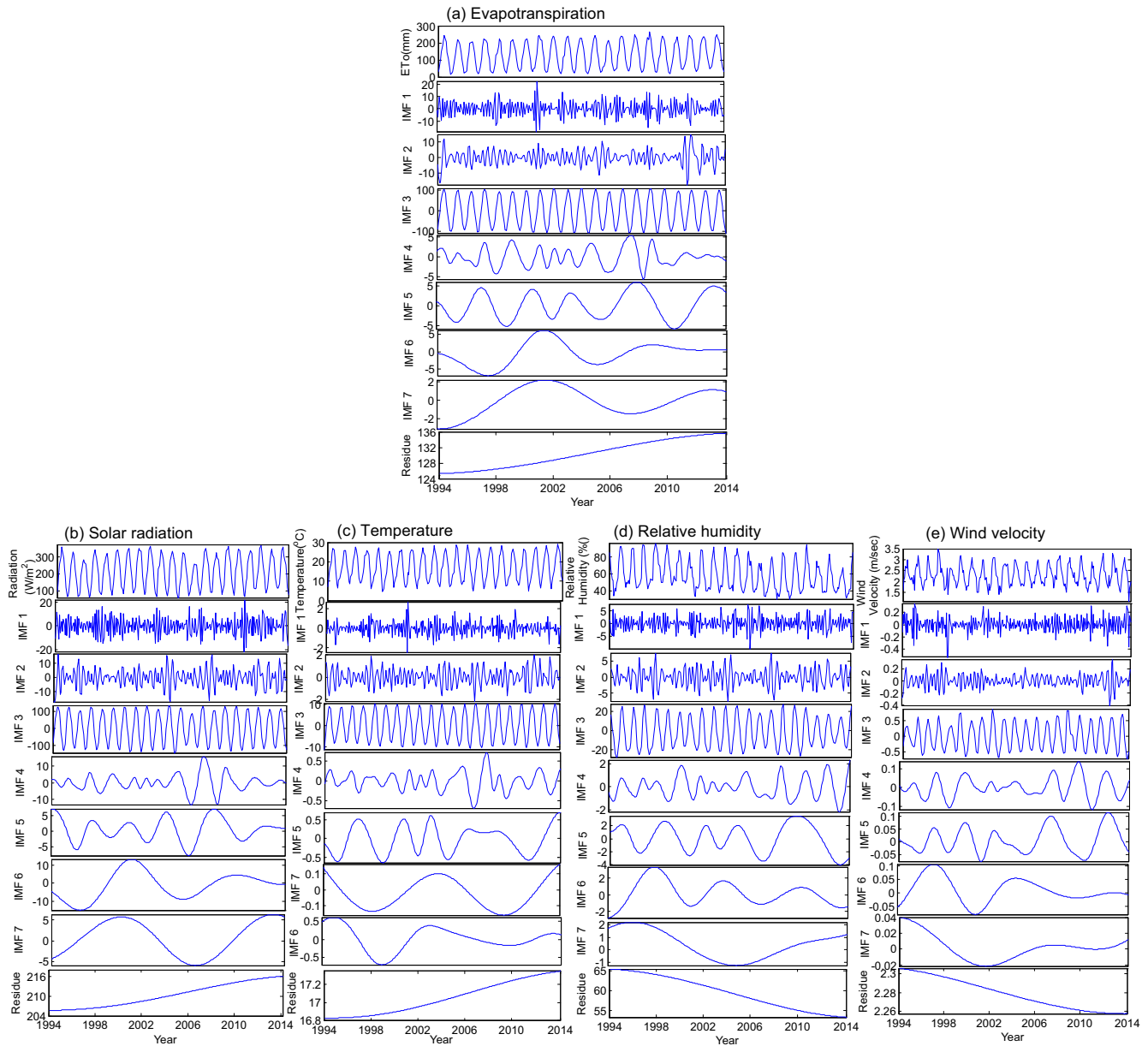


Fig. 2. Multiscale decomposition of different monthly time series, (a) evapotranspiration; (b) solar radiation; (c) surface temperature; (d) relative humidity; (e) wind velocity.

from the instantaneous frequencies of IMFs through Hilbert transform of the IMFs. The modes are provided in Fig. 2.

The mean periodicity of IMFs and residue is estimated by counting the number of zero crossings [46] and it is found that, the mean periodicity ranges from 2.7 to 252. It is to be noted that the residue has a monotonic trend in most of the series and the periodicity is referred as equal to the data length, for illustration purpose. The decomposition clearly detected annual periodicity in IMF3 (~12 months). The obtained periodicity of different IMFs is presented in Table 1.

From Table 1, it is noted that the obtained time scales for different variables are comparable for the lower order time scales, however larger differences in time scales are observed for the higher order (low frequency IMFs). This implied that exactly common time scales did not present among all variables at larger time scales and this may be due to the limited number of cycles in the higher order IMFs. The time scales of all the variables corresponding to a particular IMF were averaged to represent the time scale of that IMF. In the discrete wavelet based prediction models [20–23],

the true (exact) time scales are not identified; rather the time scales are arbitrarily assumed in powers of 2. For example, it is well known that the annual time scale is present in the ET_o and meteorological data used for the study and the MEMD method identifies this scale correctly (in IMF3), where the DWT considers either 8 ($=2^3$) or 16 ($=2^4$) months (for monthly data), which is quite different from the actual time scale. The percentage variance contributed by each IMF or residue is calculated as the percentage variance of each IMF or residue over the total variance of the data series. This information is also provided in Table 1. Pearson coefficient of correlation is calculated to explore the linear relationships among IMFs or residue of ET_o with different meteorological parameters and these are presented in Table 2. From Table 2, a strong positive correlation is noticed between the modes of ET_o and that of solar radiation. A weak positive correlation is noticed between the modes of ET_o and temperature at most of the time scales while it is strong for the annual mode and residue. Strong negative correlation exists between the different modes of relative humidity and ET_o . A weak positive correlation is noticed between the different modes of ET_o

Table 1

Periodicity of modes of ET_o (T in Months) and meteorological parameters along with the % variability explained (VE) by the modes. MN is the mode number and MP is the mean period in months.

MN	T	VE	T	VE	T	VE	T	VE	T	VE	MP
1	2.7	0.72	2.7	1.0	2.7	2.8	2.6	9.16	2.6	0.76	2.67
2	4.9	0.51	4.7	1.2	4.9	2.5	5.0	5.72	4.7	0.43	4.9
3	12	97.2	12	97	12	85.9	12.6	81.2	12	98	12.1
4	25.2	0.31	22.9	0.12	21	0.29	22.9	1.47	28	0.1	24
5	50.4	0.17	42	0.29	36	1.5	42	1.21	42	0.2	42.5
6	63	0.69	84	0.23	63	0.7	84	0.99	63	0.2	71.4
7	126	0.21	126	0.02	252	0.4	126	0.12	84	0.04	168
8	252	0.15	252	0.06	252	5.8	252	0.13	252	0.2	201.6

Table 2

Correlation coefficient between modes of monthly ET and that of different meteorological parameters

Mode of ET_o	Correlation coefficient with respective mode of			
	Radiation	Temperature	Relative humidity	Wind velocity
IMF1	0.683	0.416	-0.697	0.210
IMF2	0.583	0.197	-0.487	0.398
IMF3	0.989	0.953	-0.944	0.824
IMF4	0.859	0.134*	-0.514	0.326
IMF5	0.576	0.757	-0.710	0.519
IMF6	0.882	-0.011*	-0.529	-0.887
IMF7	0.688	0.208	-0.377	-0.893
Residue	0.998	0.996	-1.000	-0.993

* Not significant at 5% significance level. The numbers in italics indicates a change in nature of correlation.

and wind velocity, except for annual scale mode, inter-annual scale (IMF6, IMF7) and the residue. Interestingly, the inter annual modes IMF6 and IMF7 are negatively correlated with the respective modes of ET_o . This observation is remarkable, as difference in the nature of correlation is noticed among the modes for few cases (highlighted in italics in Table 2), it can be concluded that the association between the modes of ET_o and variable may not be of unique character at all time scales, i.e., relation can be positive at some of the time scales and negative in some other time scales. Moreover, it is further noticed that there exists a very strong correlation between the annual modes of meteorological parameters and ET_o (for IMF3). The correlations are 0.989, 0.953, -0.944 and 0.824 for different parameters.

4.2. Modeling of ET_o and performance evaluation

As presented earlier in the methodology section, four models were prepared. In each case, first 2/3 of the datasets (~1994 to 2007) are used for calibration and the rest of the dataset (~2008 to 2014) used for validation of the model. First, the SLR model is prepared between the ET_o and the four controlling factors.

The SLR model takes the following form

$$ET_o = -27.27 + 0.43R + 3.393T - 0.47H + 14.167U$$

$$(RMSE = 8.067, r^2 = 0.988, F = 3511.4, p\text{-value} = 0)$$

In a second exercise, the ET_o time series is decomposed into different modes using the basic EMD algorithm. The decomposition

Table 3

Mean period and percentage (%) variability explained by different modes of ET_o obtained by EMD.

IMF number	Monthly time series	
	Mean period (month)	% variability explained
IMF1	3.1	2.2
IMF2	7.2	1.87
IMF3	18	2.38
IMF4	42	2.01
Residue	252	91.5

resulted in 5 modes and the results of decomposition of ET_o time series by EMD is compared with that obtained from MEMD (previous section). The time period and percentage variability is presented in Table 3. From the periodicity values it is noted that EMD method couldn't capture the annual periodicity within the series when compared to the MEMD decomposition (Table 1). Hence MEMD is proven to be effective than EMD in decomposition of the different time series. Here maximum variability is explained by the residue. To assess the relative importance of the modes the statistical significance test of IMFs [47,48]. In this test, the energy levels of IMFs are compared with that of a white noise series generated by Monte Carlo simulation. In a plot between $\log_2(\text{Mean Period})$ and $\log_2(\text{Normalized Energy})$, the IMFs whose energy level is above the upper confidence line is considered as significant. It is customary to consider IMF1 as the reference IMF for normalizing the energy, and the residue is often excluded in the rest (as it generally possesses very high energy level compared with the IMFs). The results of statistical significance test of the ET_o time series is presented in Fig. 3.

Fig. 3 shows that all IMFs are significant (energy level falls above the 95% and 99% upper significance line) and the highest energy is for IMF3. It is well known that annual cycle is present in the dataset, but the periodicity of IMF estimated by in this case it is 18 months instead of the annual cycle (12 months). Hence it can be inferred that annual cycle is not captured well by the EMD based scale separation. SLR models were developed between the modes of ET_o at each time scale and the actual meteorological parameters. For brevity, the decomposition models are not presented herein, but it is applied for the validation dataset to get the IMFs of ET_o . Finally, the summation of predicted IMFs given the ET_o for the validation dataset.

As a third method, the MEMD-SLR model is developed. Here, individual models are prepared for each mode. The models for first three IMFs and residue only are provided herein. The fitted models are in the form:

$$IMF1_{ET_o} = 0.073 + 0.368IMF1_R + 3.61IMF1_T - 0.875IMF1_{RH} + (7.547)IMF1_U$$

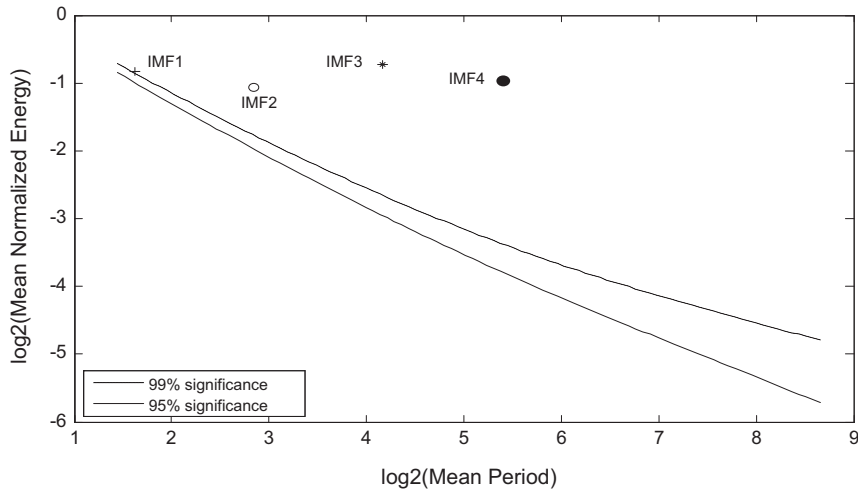


Fig. 3. Statistical significance test of IMF components obtained by EMD.

$$IMF2_{ET_o} = 0.014 + 0.381IMF2_R - 2.456IMF2_{RH} + 1.492IMF2_T + (14.71)IMF2_U$$

$$IMF3_{ET_o} = -2.767 + 0.468IMF3_R + 3.539IMF3_T - 0.3IMF3_{RH} + 14.097IMF3_U$$

$$Re\ sidue_{ET_o} = 160.74 + 0.221Re\ sidue_R + 2.61IMFRe\ sidue_T - 0.372Re\ sidue_{RH} - 43.696Re\ sidue_U$$

The adjusted R^2 values for the above fittings are 0.765, 0.467, 0.996 and 1.00 with RMSE values of 3.14, 4.45, 3.51 and practically zero, respectively for IMF1, IMF2, IMF3 and the final residue. For the IMFs 1 and 2, the adjusted R^2 values were the lowest while it approached unity for the low frequency IMFs and the final residue component. Also the RMSE is found to be decreasing with increase in numerical number of modes. This inferred that the relationship between ET_o and meteorological parameters is more linear at larger time scales.

In the four models presented above, the coefficients with brackets are not significant (at 5% significance level) and in the modeling exercise, these coefficients are brought to zero and subsequently the respective models of ET_o are revised. For example, in the prediction of IMF1 of ET_o , wind velocity is not influential; in the prediction of IMF2 of ET_o the relative humidity plays no role etc.

Hence it is confirmed that in the modeling exercise, the SLR is capable to detect and omit such less contributing predictors. By considering IMF2, a comparison is made between the IMF2 of ET_o and IMF2 of relative humidity (for training dataset) and it is presented in Fig. 4. From Fig. 4, it is noticed that the association is primarily negative, but positive relation between the two (with some phase shift) can also be noticed at many shorter time spells. On finding the overall correlation between these IMFs it is found that the value is only -0.0074. Hence it is inferred that the positive correlation at some time spell (window) might have cancelled by the negative correlation between the two at some other time spells. No such omissions were noticed in the low frequency IMFs and the residue.

To enable a comparison of results obtained by the proposed method, a non-linear modeling employing the M5 model tree algorithm is also made. Here, the pruned and smoothed version of M5 model tree [49] is used and number of piecewise linear models are obtained for the domain of datasets supplied. The time series plots of calibration and validation datasets are prepared and presented in Fig. 5. A detailed statistical performance evaluation of the candidate models was done by a multitude of evaluation criteria such as coefficient of determination (r^2), Nash Sutcliffe Efficiency (NSE), root mean square error (RMSE), mean absolute error (MAE), scatter index (SI), threshold statistics (TS) and the results are presented in Table 4.

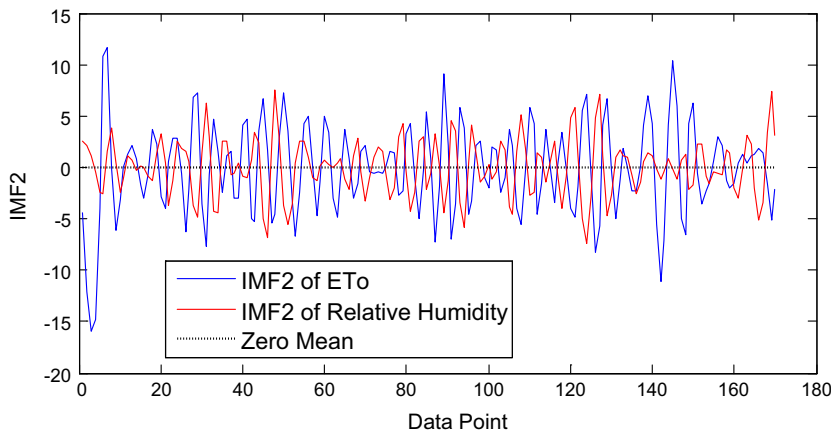


Fig. 4. Comparison of IMF2 of evapotranspiration and relative humidity.

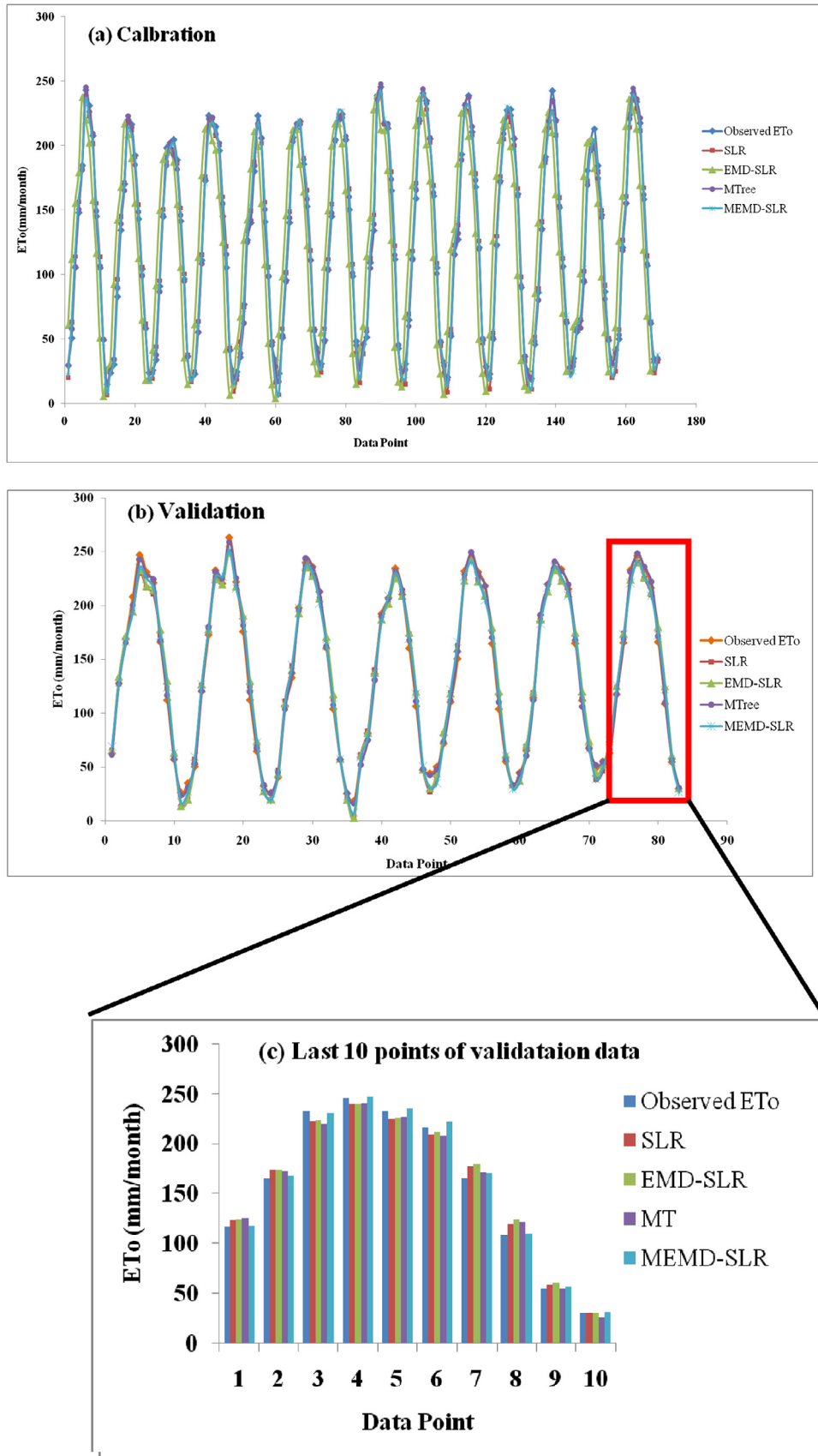


Fig. 5. Time series plots between observed ET_0 and ET_0 predicted by different models; (a) calibration data; (b) validation; (c) bar plots of last 10 points in validation data.

Table 4
Performance evaluation of different models.

Evaluation criteria	Calibration data				Validation data			
	SLR	EMD-SLR	MT	MEMD-SLR	SLR	EMD-SLR	MT	MEMD-SLR
r^2	0.988	0.987	0.995	0.992	0.988	0.986	0.989	0.990
NSE	0.988	0.987	0.992	0.992	0.988	0.986	0.990	0.990
RMSE	7.977	8.370	5.923	6.704	8.326	9.100	8.134	7.612
MAE	6.779	7.089	5.346	5.525	7.127	7.700	7.010	6.443
SI	0.064	0.067	0.048	0.054	0.061	0.100	0.059	0.055

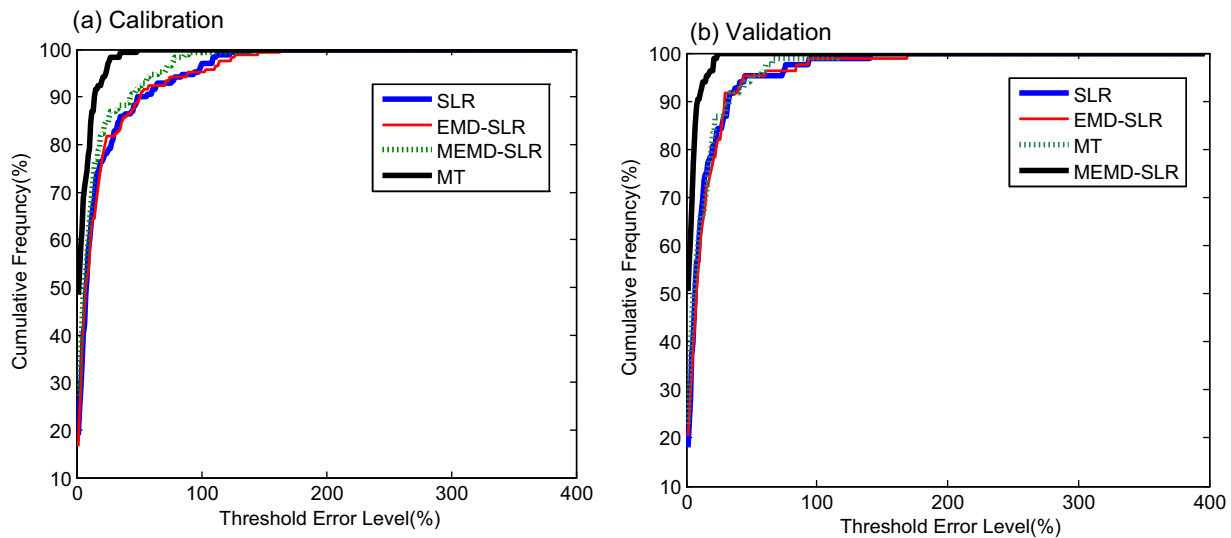


Fig. 6. Distribution of prediction error across different error thresholds for all the models (a) calibration data, (b) validation data.

An improvement in performance of MEMD-SLR model is noticed when compared with other methods and it is more apparent while considering the different error statistics. For example, RMSE of MEMD-SLR model for validation is 8.6% better than that of SLR model and 16.35% better than that of EMD-SLR. Similarly, the MAE of predictions by MEMD-SLR model (6.443) is 9.5% better than that of SLR model (7.12) and 16.3% better than that of EMD-SLR model. On considering the results by non-linear fitting using M5 Model tree (MT) algorithm, the performance is best for the calibration data. But on comparing the results of validation data, it is noticed that MEMD-SLR hybrid model performs slightly better than MT. Unlike for the case of ANN/SVM, MT is free from algorithm specific parameters and therefore the possibility of over fitting can be safely ruled out for MT. Hence it can be concluded that the scale specific decomposition shows better generalization capability than other methods considered in this study. It is observed that the performance of EMD based SLR model is not found to be better than that of SLR model, particularly in terms of error statistics. This may be because of the fact that the four parameters chosen are strongly correlated with the ET_o but not so on considering the modes (IMFs/Residue) of ET_o . The plots of threshold statistics evaluation for different cases are given in Fig. 6. From Fig. 6, it is evident that MEMD-SLR model is better than the SLR, EMD-SLR and MT (capturing of nearly 100% frequency happens at lower error level) for the validation data. Overall performance evaluation indicated that prediction of ET_o using MEMD-SLR is a robust alternative to the predictions of ET_o based on the original data and the prediction based on decomposition of ET_o alone.

From this study it is noted that the proposed MEMD-SLR method can improve the overall prediction of ET_o , yet it can be observed that the improvement is less than 10% for the best per-

forming model. Some of the researchers investigated about this 'marginal gain in skill' in performance by the different methods when compared with the linear regression method [20,43] and the results are in similar lines on comparing with that reported in some of the other studies also [12].

In all of the above stated studies the same 4 variables from the CIMIS database are considered as inputs and these parameters are strongly correlated with ET_o . On applying SLR directly upon meteorological parameters, no omission of parameters is noticed (means all the 4 parameters are relevant for SLR model prepared), it becomes equivalent to multiple linear regression. Further, it is well understood that linear regression assumes that the data has normal distribution [20]. Here the skewness of the ET_o time series is found to be 0.048, which is close to zero. This may one reason for the satisfactory degree of performance of SLR for the estimation of ET_o . However any improvement in the prediction performance is appreciable in the modeling of complex process of ET_o and the proposed clearly identified that at all time scales all of the predictor variables may not be influencing the process. Identifying such variables and omitting them improves the modeling accuracy, which is well reflected in the present modeling exercise. If more such cases of omissions are present, it is expected that the time scale based modeling can improve the modeling accuracy considerably, which will essentially depends on the type of problem we are addressing and the predictor dataset we are choosing. Hence the proposed time scale based modeling strategy involving MEMD and SLR is rather innovative and the methodology presented herein being general one; it can be applied for any other combination of the input datasets and at any other location. The identification and exclusion of such less influential parameters at different time scale essentially improve the modeling accuracy, and such an exercise

cannot be performed by applying linear/non-linear regression tools directly upon the input parameters.

The MEMD based decomposition being data adaptive, completely avoids the tedious process of selection of number of decomposition levels and basis function type. This is another advantage of the proposed method over DWT based hybrid models. Also it is very simple in implementation compared with EMD, as the decomposition of all the variables (input and output data) are carried out simultaneously. Another inherent problem which can be expected in traditional EMD based hybrid decomposition models [50–53] is the number of decomposition modes evolved. The number of decomposition modes evolved may not be necessarily same for the different predictor variables and the predictant. This will offer difficulties in modeling stage and constrain the modeler from using multiple inputs in modeling stage. Hence in most of the past studies [50–52] the concerned dependent time series is considered, it is decomposed and appropriate number of lagged values IMFs at the corresponding time scale is considered and the final summation of predicted IMFs will provided the desired predictions at the measurement scale. Eventhough some of the very recent study [53] made an attempt to consider the lagged value of multiple inputs, the final model is prepared by considering the IMFs of independent variables as input and the original dependent variable as output. In such a case the important step of 'identification of appropriate inputs at different scales from multiple independent variables' is not present. In the proposed MEMD-SLR method such issues are addressed effectively and it can be used as a promising alternative strategy for modeling complex hydrologic time series signals.

5. Conclusions

This paper proposed a new method to reveal the time scaling characteristics of ET_o time series and the four meteorological parameters by employing the MEMD and based on which, a novel method is proposed for prediction of ET_o .

The important conclusions from the study are:

- The MEMD based decomposition is purely data adaptive, capable to extract the exact time scale of variability, maintain same number of modes for all variables and simple to implement.
- The decomposition by MEMD could detect the annual cycle inherent in different time series in a better way than the traditional EMD method and when multiple time series plays a role to the process this is significant.
- The performance evaluation of predictions for a validation dataset based on a multitude of statistical criteria revealed that the proposed MEMD-SLR model offers better predictive capabilities than the M5 Model tree and SLR models (in which the meteorological parameters are used directly) and the EMD - SLR hybrid model (in which decomposition of ET_o dataset alone is done). Hence the proposed model is a robust alternative for accurate estimation of ET_o accounting the time scale of variability.

Acknowledgements

The authors like to express their acknowledgements to Dr Danilo P Mandic for providing the basic source code of MEMD algorithm at <http://www.commsp.ee.ic.ac.uk/~mandic/research/emd.htm> with the intension of promoting non commercial scientific research. Also, the authors express their acknowledgements to Prof Bing Cheng Si, College of Agriculture and Bioresources, University of Saskatchewan Canada, and Dr Wei Hu, Post Doctoral Fellow, College of Agriculture and Bioresources, University of Saskatchewan

Canada for scientific discussions on the methodology during its implementation for evapotranspiration modeling.

References

- [1] Parasuraman K, Elshorbagy A, Carey CK. Modelling the dynamics of the evapotranspiration process using genetic programming. *Hydrol Sci J* 2009;52(3):563–78.
- [2] Pal P, Deswal S. M5 model tree based modelling of reference evapotranspiration. *Hydrol Process* 2009;23(10):1437–43.
- [3] Jensen M, Burman RD, Allen RG. Evapotranspiration and irrigation water requirements. In: ASCE manuals and reports on engineering practice no. 70. New York: ASCE; 1990.
- [4] Smith M. Report on the expert consultation on revision of FAO methodologies for crop water requirements. Rome, Italy: FAO; 1992.
- [5] Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration, guidelines for computing crop water requirements. In: FAO irrigation and drainage, paper no. 56, Rome: Food and Agriculture Organization of the United Nations; 1998.
- [6] Kumar M, Raghuvanshi N, Singh R, Wallender W, Pruitt W. Estimating evapotranspiration using artificial neural network. *J Irrig Drain Eng* 2002;128(4):224–33.
- [7] Nandagiri L, Kovoov GM. Performance evaluation of reference evapotranspiration equations across range of Indian climates. *J Irrig Drain Eng* 2006;132(3):238–49.
- [8] Kisi O. Comparison of different empirical methods for estimating daily reference evapotranspiration in Mediterranean climate. *J Irrig Drain Eng* 2014;140(1):1–8.
- [9] Sudheer KP, Gosain AK, Ramasastri KS. Estimating actual evapotranspiration from limited climatic data using neural computing technique. *J Irrig Drain Eng* 2003;129(3):214–8.
- [10] Trajkovic S, Todorovic B, Stankovic M. Forecasting reference evapotranspiration by artificial neural networks. *J Irrig Drain Eng* 2003;129(6):454–7.
- [11] Kisi O. Generalized regression neural networks for evapotranspiration modeling. *Hydrol Sci J* 2006;51(6):1092–105.
- [12] Kisi O. Least squares support vector machine for modeling daily reference evapotranspiration. *Irrig Sci* 2013;31(4):611–9.
- [13] Shiri J, Sadreddini AA, Nazemi AH, Kisi O, Landeras G, Fard AF, Marti P. Generalizability of gene expression programming-based models for estimating daily reference evapotranspiration in coastal stations of Iran. *J Hydrol* 2014;508:1–11.
- [14] Kisi O, Cimen M. Evapotranspiration modelling using support vector machines. *Hydrol Sci J* 2009;5(5):918–28.
- [15] Kisi O, Guven A. Evapotranspiration modeling using linear genetic programming technique. *J Irrig Drain Eng* 2010;136(10):715–23.
- [16] Rahimikhoob A. Comparison between M5 model tree and neural networks for estimating reference evapotranspiration in an arid environment. *Water Resour Manage* 2014;28(3):657–69.
- [17] Izadifar Z, Elshorbagy A. Modeling and analysis of actual evapotranspiration: data driven modeling and wavelet analysis. Germany: VDM Verlag, Dr Muller Saarbrucken; 2010.
- [18] Ding R, Kang S, Vargas R, Hao X. Multiscale spectral analysis of temporal variability in evapotranspiration over irrigated cropland in an arid region. *Agric Water Manage* 2014;130:79–89.
- [19] Kisi O. Evapotranspiration modeling using a wavelet regression model. *Irrig Sci* 2011;29(3):241–52.
- [20] Partal T. Modelling evapotranspiration using discrete wavelet transform and neural networks. *Hydrol Process* 2009;23:3545–55.
- [21] Cobaner M. Reference evapotranspiration based on Class A pan evaporation via wavelet regression technique. *Irrig Sci* 2013;31:119–34.
- [22] Gocic M, Motamodi S, Shamshirband S, Perkovic D, Sudheer Ch, Hashim R, Arif M. Soft computing approaches for forecasting reference evapotranspiration. *Comput Electr Agric* 2009;113:164–73.
- [23] De Sousa AL, Castro NMR, Canales FA, Louzada JAS, Vitorino MI, DeSouza EB. Multiscale variability of the evapotranspiration in eastern Amazonia. *Atmos Sci Lett* 2010;11:192–8.
- [24] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc Roy Soc London, Ser A* 1998;454:903–95.
- [25] Huang Y, Schmitt FG, Lu Z, Liu Y. Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis. *J Hydrol* 2009;454:103–11.
- [26] Kuai KZ, Tsai CW. Identification of varying time scales in sediment transport using the Hilbert-Huang Transform method. *J Hydrol* 2012;420–421:245–54.
- [27] Massei N, Fournier M. Assessing the expression of large scale climatic fluctuations in the hydrologic variability of daily Seine river flow (France) between 1950–2008 using Hilbert Huang Transform. *J Hydrol* 2012;448–449:119–28.
- [28] Adarsh S, Janga Reddy M. Multiscale analysis of suspended sediment concentration data from natural channels using the hilbert-huang transform. *Aquat Procedia* 2015;4:780–8.

- [29] Janga Reddy M, Adarsh S. Time frequency characterization of subdivisional scale seasonal rainfall in India using Hilbert Huang transform. *Stoch Environ Res Risk Assess* 2016;30(4):1063–85.
- [30] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Centre for ocean-land-atmospheric studies technical report.193, Centre for Ocean-Land-Atmos. Stud., Calverton, Md.; 2005. p. 1–51. <ftp://grads.iges.org/pub/ctr/ctr_193.pdf>.
- [31] Torres ME, Colominas MA, Schlotthauer G, Fladrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: IEEE international conference on acoustic speech and signal processing, Prague. p. 4144–7.
- [32] Rilling G, Flandrin P, Goncalves P, Lilly JM. Bi-variate empirical mode decomposition. *IEEE Signal Process Lett* 2007;14:936–9.
- [33] Rehman N, Mandic DP. Empirical mode decomposition for tri-variate signals. *IEEE Trans Signal Process* 2007;58:1059–68.
- [34] Rehman N, Mandic DP. Multivariate empirical mode decomposition. *Proc Royal Soc Series A* 2010;466(2117):1291–302.
- [35] Hu W, Si BC. Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition. *Geoderma* 2013;193–194:180–8.
- [36] Huang G, Su Y, Kareem A, Liao H. Time-frequency analysis of non-stationary process based on multivariate empirical mode decomposition. *J Eng Mech ASCE* 2016;142(1). doi: [http://dx.doi.org/10.1061/\(ASCE\)EM.1943-7889.0000975](http://dx.doi.org/10.1061/(ASCE)EM.1943-7889.0000975).
- [37] Bai Y, Wang P, Xie J, Li J, Li C. Additive model for monthly reservoir inflow forecast. *J Hydrol Eng* 2015. doi: [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0001101.04013079](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001101.04013079).
- [38] Huang NE, Wu MCL, Long SR, Shen SSP, Qu W, Gloersen P, Fan KL. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc Roy Soc Ser A* 2003;459:2317–45.
- [39] Draper NR, Smith H. *Applied regression analysis*. Hoboken, NJ: Wiley-Interscience; 1998.
- [40] Quinlan JR. Learning with continuous classes. In: *Proceedings of australian joint conference on artificial intelligence*. Singapore: World Scientific Press; 1992. p. 343–8.
- [41] Singh KK, Pal M, Singh VP. Estimation of mean annual flood in Indian catchments using back-propagation neural network and M5 Model tree. *Water Res Manage* 2010;24:2007–19.
- [42] Kisi O, Kilic Y. An investigation on generalization ability of artificial neural networks and M5 model tree in modeling reference evapotranspiration. *Theor Appl Climatol* 2016;126(3):413–25.
- [43] Ayttek A, Guven A, Yuce MI, Aksoy H. An explicit neural network formulation for evapotranspiration. *Hydrol Sci J* 2008;53(4):893–904.
- [44] Tabari SH, Kisi O, Ezani A, Talaee PH. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid high land environment. *J Hydrol* 2012;444–445:78–89.
- [45] Rilling G, Fladrin P, Goncalves P. On empirical mode decomposition and its algorithms. IEEE-EURASIP workshop non linear signal and image processing; 2003.
- [46] Barnhart BL, Eichinger WL. Empirical mode decomposition applied to solar irradiance, global temperature, sunspot number and CO₂ concentration data. *J Atmos Sol – Terr Phys* 2011;73:1771–9.
- [47] Wu Z, Huang NE. A study on the characteristics of white noise using the empirical mode decomposition method. *Proc Roy Soc Ser A* 2004;460:1597–611.
- [48] Wu Z, Huang NE. Statistical significance test of intrinsic mode functions. In: Norden E Huang (NASA Goddard Space Flight Center, USA), Samuel S P Shen (University of Alberta, Canada), editors. *Hilbert huang transform and its applications*. Singapore: World Scientific Publishing; 2005.
- [49] Jothiprakash V, Kote AS. Effect of pruning and smoothing while using M5 Model Tree technique for reservoir inflow prediction. *J Hydrol Eng* 2011;16(7):563–74.
- [50] Napolitano G, Serinaldi F, See L. Impact of EMD decomposition and random initialization of weights in ANN hindcasting of daily streamflow series: an empirical examination. *J. Hydrol.* 2011;406(3–4):199–214.
- [51] Huang S, Cheng J, Huang Q, Chen Y. Monthly streamflow prediction using modified EMD-based support vector machine. *J Hydrol* 2011;511:764–75.
- [52] Karthikeyan L, Kumar DN. Predictability of non-stationary time series using Wavelet and EMD based ARMA models. *J. Hydrol.* 2013;502:103–19.
- [53] Zhu S, Zhou J, Ye L, Meng C. Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper

reaches of Yangtze River, China. *Environ Earth Sci* 2016. doi: <http://dx.doi.org/10.1007/s12665-016-5337-2016>.



resources systems, application of data driven techniques for hydrologic modeling and hydro climatology (Link: <http://tkmce.ac.in/user-profile?user=307>).

S. Adarsh completed his B.Tech degree in Civil Engineering in the year 2003 from TKM College of Engineering Kollam, India. He joined as an Assistant professor in the Civil Engineering department of the same institute in 2004. He completed M.Tech in Water Resources Engineering from IIT Bombay in 2009. He is currently about to complete his Ph.D from IIT Bombay on multiscale analysis of hydrologic time series. He published 25 papers in international journals of high reputation and many in conferences and undertook three research projects so far. His areas of research interest include hydrologic time series analysis, water



Sulaiman Sanah was a graduate student at TKM college of Engineering Kollam, Kerala India and received her B. Tech degree in Civil Engineering in 2015.



K.K. Murshida was a graduate student at TKM college of Engineering Kollam, Kerala India and received her B. Tech degree in Civil Engineering in 2015.



P. Nooramol was a graduate student at TKM college of Engineering Kollam, Kerala India and received her B. Tech degree in Civil Engineering in 2015.