Hybrid point and interval prediction approaches for drought modeling using ground-based and remote sensing data

Kiyoumars Roushangar\textsuperscript{a,b,*}, Roghayeh Ghasempour\textsuperscript{a}, V. S. Ozgur Kirca\textsuperscript{c} and Mehmet Cüneyd Demirel\textsuperscript{c}

\textsuperscript{a}Department of Water Resource Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran
\textsuperscript{b}Center of Excellence in Hydroinformatics, University of Tabriz, Tabriz, Iran
\textsuperscript{c}Division of Hydraulics, Department of Civil Engineering, Istanbul Technical University, Istanbul, Turkey
*Corresponding author. E-mail: kroshangar@yahoo.com

ABSTRACT

Drought as a severe natural disaster has devastating effects on the environment; therefore, reliable drought prediction is an important issue. In the current study, based on lower upper bound estimation, hybrid models including data preprocessing, permutation entropy, and artificial intelligence (AI) methods were used for point and interval predictions of short- to long-term series of Standardized Precipitation Index in the Northwest of Iran. Ground-based and remote sensing precipitation data were used covering the period of 1983–2017. In the modeling process, first, the data processing capability via variational mode decomposition (VMD), ensemble empirical mode decomposition, and permutation entropy (PE) was investigated in drought point prediction. Then, interval prediction was applied for tolerating increased uncertainty and providing more details for practical operation decisions. The simulation results demonstrated that the proposed integrated models could achieve significantly better performance compared to single models. Hybrid PE models increased the modeling accuracy up to 40 and 55%. Finally, the efficiency of developed models was verified for Normalized Difference Vegetation Index prediction. Results demonstrated that the proposed methodology based on remote sensing data and VMD–PE–AI approaches could be successfully used for drought modeling, especially in limited or non-gauged areas.

Key words: artificial intelligence, data preprocessing, drought, permutation entropy, remote sensing, variational mode decomposition

HIGHLIGHTS

• Apply ground- and remote sensing-based data to point and interval predictions of drought series.
• Propose new model structures via ensemble empirical mode decomposition, variational mode decomposition, permutation entropy, and artificial intelligence approaches for drought modeling.
• The possibility of using the proposed hybrid models for the Normalized Difference Vegetation Index series forecasting was assessed.
• The obtained results showed the high ability of developed models in drought indices modeling.
INTRODUCTION

One of the most complex phenomena among all extreme climate events is drought (Ghulam et al. 2007). Drought has serious negative effects on different parts of human life (Hayes et al. 1999). Some of these effects are water supply shortage, low agricultural produce, reduced soil moisture, economic losses, migration, and famine (Von Hardenberg et al. 2001; Zhou et al. 2018). These widespread negative effects can be seen anywhere in the world (Mann & Gleick 2015; Marvel et al. 2019). An accurate understanding of spatial and temporal variation in rainfall and drought will provide a good insight into the arranging and management of drought subordinate exercises (Basalirwa et al. 1999; Nischitha et al. 2014). Among hydro-meteorological variables (i.e. precipitation, evapotranspiration, soil moisture, runoff, reservoir level, streamflow, and temperature), precipitation is the most appropriate variable to monitor the meteorological drought. The use of drought index is a common practical method to monitor the drought phenomenon. Therefore, different indices have been developed for precise applications (Kwak et al. 2016; Mukherjee et al. 2018) such as Surface Water Supply Index, Palmer Drought Severity Index, Standardized Runoff Index, Evapotranspiration Deficit Index, Standardized Precipitation Evapotranspiration Index, and Standardized Precipitation Index (SPI). Li et al. (2015) suggested the SPI as an ideal tool in drought severity characterization, since it can show both short- and long-term effects of drought. Different time scales of drought index makes it possible to investigate the impacts of rainfall shortages on different parts of water resources. At shorter timescales, the SPI is related to soil moisture, whereas at longer timescales, it shows the groundwater and reservoir storage conditions.
On the other hand, the traditional drought indices are usually obtained from hydro-meteorological data from stations, the spatial resolution of which does not necessarily meet the drought monitoring requirements in large-scale areas. According to Yilmaz et al. (2008), spatially and temporally continuous information can be provided by data obtained from meteorological satellites. Therefore, the use of satellite data can compensate the limitation of point-based observation data for hydro-meteorological variables and makes it possible to monitor drought conditions in non-gauged regions. Meteorological satellites can acquire multitemporal, multispectral, continuous, and complete data. Sensors used to measure light reflections from surfaces are mounted to an aircraft, drone, or even a satellite. This positioning of the sensors allows them to cover a large area in a short time, which renders remote sensing a fast process. Remote sensors are also able to scan and create maps of inaccessible areas, and they provide more homogeneous data quality compared to ground observations. However, raw satellite datasets have some disadvantages; for example, the amount of cloudiness affects the accuracy of the results. Therefore, satellite-based datasets need to be examined and verified by comparison with ground-based data.

The Normalized Difference Vegetation Index (NDVI) is an alternative remote sensing drought index, which is successfully used for monitoring the dynamic variations in vegetation cover (Chen et al. 2010). In recent years, many researchers have tried to use satellite data as an efficient tool to measure and monitor drought-related variables on a global scale. Although the use of satellite products has many benefits in drought analysis, there are still significant challenges to boost the reliability and efficiency of various approach outcomes (Brown et al. 2015). So far, numerous models have been developed to enhance the efficiency of drought predictions (Mishra et al. 2007; Mehr et al. 2014; Geng et al. 2016; Sakizadeh et al. 2019). Time-series, regression, physical-based, and artificial intelligence (AI) methods are some examples. In recent decades, AI methods [e.g., support vector machine (SVM), artificial neural networks (ANNs), genetic programming, adaptive neuro-fuzzy inference system (ANFIS), feed-forward neural network (FFNN), and Gaussian process regression (GPR)] have been widely used to investigate the complex hydraulic and hydrologic phenomena (Yin et al. 2017). Side weir discharge coefficient modeling (Azamathulla et al. 2016), rainfall–runoff modeling (Chadalawada et al. 2017), pan evaporation estimation (Wu et al. 2019), groundwater-level prediction (Sakizadeh et al. 2019), forecasting different types of drought simultaneously (Aghelpour et al. 2020a), and river discharge relationship modeling (Roushangar et al. 2021) are some examples. Deo et al. (2017) investigated the potential of several AI models in the SPI series forecasting. The obtained results highlighted the importance of periodicity in drought forecasting. Kisi et al. (2019) investigated the accuracy of novel heuristic methods in forecasting various time scales of the SPI series in a semi-arid environment and compared the obtained results with the classical ANFIS method. Mohamadi et al. (2020) used integrated machine learning models with a nomadic people optimization (NPA) algorithm to forecast meteorological droughts in Iran. The general results indicated that the NPA and wavelet coherence analysis are useful tools for modeling drought indices. Aghelpour et al. (2020b) used meta-heuristic models to model different drought indices. The applied models led to desirable predictions in arid, semi-arid, and Mediterranean climates, while humid climate provided the weakest predictions. Aghelpour et al. (2020c) investigated the snow cover of the Northern and Southern slopes of Alborz Mountains in Iran by considering two issues: (1) estimating the snow cover area and (2) investigating the effects of droughts on snow cover. The results showed that the effects of a drought event on the snow cover area would remain up to 5 (or 6) months in the region.

Since most of the hydrological time series are nonstationary, trendly, or contain seasonal fluctuations, the signal preprocessing methods are commonly used by researchers especially in the past three decades (Kumar & Foufoula 1993; Adamowski et al. 2009; Roushangar & Alizadeh 2018). These methods have been used for breaking down and excavating complex, periodic, and irregular hydrological time series (Roushangar & Alizadeh 2018). Wavelet transform (WT) analysis is a common method for time-series analysis, and many researchers have used it to decompose signals in different fields. Beside the WT, the empirical mode decomposition (EMD) and variational mode decomposition (VMD) methods have also been used recently for signal decomposition. These methods are particularly suitable for decomposition of nonlinear and nonstationary signals (Huang et al. 1998). According to Abdoos (2016), the correlation frequency band and modal components can be found by the use of the variational mode; so, the VMD has better anti-noise performance compared to the EMD.

In the present study, point and interval predictions of short- to long-term drought were performed by using ground- and remote sensing-based datasets. In this regard, the FFNN, SVM, and GPR models were applied for point and prediction intervals (PIs) of the SPI series at 1-month (short-term), 9-month (mid-term), and 24-month (long-term) timescales. SVM and GPR are relatively new and promising methods based on different
kernel types, which are based on the statistical learning theory initiated. Such methods can model nonlinear decision boundaries, and moreover, they offer many alternative kernels to choose from. A kernel-based (KB) approach is capable of adapting itself to predict any variable of interest via sufficient inputs and is fairly robust against over-fitting, especially in high-dimensional space. FFNN is a commonly encountered classical type of ANN that has been applied to many diverse fields. In this type of network, the information moves in only one direction, forward from the input nodes to the output nodes, through the hidden nodes (if any). The training of this method is fast, it has high accuracy, and the probability of occurrence of data over-training in this method is relatively less. PIs are prevalent measures to investigate the uncertainty of prediction. PIs assimilate the accuracy of predicted values vs. the measured values (Khosravi et al. 2011). Since the uncertainties are considered in PIs, a bound (envelope) for the variables of concern can be obtained, which makes PIs a useful instrument to address decision-making challenges and the management of critical issues arising from design and adaptation plans in the context of climate change. In the present research, lower upper bound estimation (LUBE) approach was used for constructing PIs for GPR-based modeling of drought series. The precipitation data from meteorological stations and two PERSIANN-CDR and CMAP satellites were used for the SPI series modeling. During modeling processes, data of 12 stations located in the Northwest of Iran were used. Iran, located in the West of Asia, has diverse climatic characteristics. However, in most regions, arid and semi-arid climates are predominant, and therefore, drought is a major concern in the country. For calculating the SPI time series, the rainfall data in the period of 1983–2017 was used. The accuracy and homogeneity of data were checked and confirmed. In the proposed modeling, first, the EEMD and VMD methods were utilized for decomposition of the original time series into several subseries named intrinsic mode functions (IMFs). Then, the arrangement entropy of each IMF was calculated. Based on the closeness of these entropy values, the IMFs were combined to achieve a recombinant subseries. The recombinant subseries were inserted into AI methods to the point prediction of the drought time series. However, due to the difficulties in accurate point prediction, the LUBE method, which can tolerate increased uncertainty, was applied to construct PIs for drought time series. As such, the present study aims at proposing a novel hybrid system for short- to long-term SPI time-series point and interval predictions by combining data preprocessing, permutation entropy (PE), and AI approaches using ground- and satellite-based data. Also, for verifying the proposed hybrid model performance, the capability of the proposed methods was investigated for the NDVI time-series prediction.

**MATERIALS AND METHODS**

**Characteristics of study area and data used**

In the current study, East Azerbaijan province in the Northwest of Iran was selected as the study area. The annual precipitation variability is high in the Northwest regions of Iran; therefore, drought is one of the most frequently encountered climate-related disasters. The average annual rainfall in the selected area varies from 198 to 845 mm. Most of this precipitation is received by the mountainous areas. The mean monthly precipitation in the studied 35-year statistical period varies between 0 and 159.4 mm. Based on this monthly precipitation data, May, April, March, and October come out to be the rainiest months of the year, whereas August, July, and September are the driest months. Also, in summer, the prevailing concentration of rain gauge stations is strongly irregular in the selected area. In this study, as seen from Figure 1, 12 synoptic stations were used to assess drought modeling. The datasets used were obtained from the meteorology portal of Iran (http://www.weather.ir). The considered stations had records of 35 years, and they showed a desirable spatial distribution over the study area. The Thiessen polygon method was applied for calculating the spatial average precipitation of the region; then, these spatially averaged precipitation values were used for calculating the region SPI time series.

In addition to the ground-based data, two remote sensing observation products were used to model the SPI time series: (1) the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR) and (2) CPC Merged Analysis of Precipitation (CMAP). According to Xie & Arkin (1996), the CMAP is a group of rainfall dataset, which can be generated via gauge data analysis and satellite-based rainfall estimations. The PERSIANN-CDR is a valuable precipitation dataset, which is based on rainfall prediction via ANN algorithm and infrared satellite images (Ashouri et al. 2015). Table 1 shows the characteristics of spatio-temporal remotely sensed products used in this study. Additionally, the NDVI series, which can be obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS)
on the National Aeronautics and Space Administration (NASA) Earth Observation Satellite (EOS) Terra (https://modis.gsfc.nasa.gov/), was used for predicting the vegetation patterns.

**Standard Precipitation Index (SPI)**

The impacts of a water deficit are complex functions of water source and water use. The time scale over which precipitation deficits accumulate becomes extremely important and separates different types of drought. As described by McKee et al. (1993), some of the practical issues that are important in drought analysis are time scale, probability (the expected frequency of an event), and precipitation deficit. Addressing these issues, McKee et al. (1993) developed a parameter termed as SPI, where the standardization is based on probability. The SPI quantifies observed precipitation as a standardized departure from a selected probability distribution function that models the raw precipitation data. The raw precipitation data are typically fitted to a gamma or a Pearson Type-III distribution and then transformed to a normal distribution. The SPI values can be interpreted as the number of standard deviations by which the observed anomaly deviates from the long-term mean (Mirabbasi et al. 2013).

The SPI index has statistical consistency and gives a representation of abnormal wetness or dryness. The SPI was designed to quantify the precipitation deficit for multiple timescales, and it is capable of showing both short- and long-term drought effects in variable timescales of precipitation anomalies. The use of different time scales allows the effects of a precipitation deficit on different water resources components (groundwater, reservoir storage, soil moisture, and streamflow) to be investigated (Tsakiris & Vangelis 2004; Mirabbasi et al. 2013). For example, soil moisture conditions respond to precipitation anomalies on a relatively short scale, whereas

---

**Table 1** | Information of the used satellite products

| Satellite product | Date range | Spatial resolutions | Temporal resolutions | Input data |
|-------------------|------------|---------------------|----------------------|------------|
| PERSIANN-CDR      | 1983–2017  | 0.25° × 0.25°       | Daily                | Multiple satellite microwave sensors |
| CMAP              | 1979–2017  | 2.5° × 2.5°         | Monthly              | Gauge analysis; microwave and infrared observations from polar orbiting and geostationary satellites |
| MODIS             | 2000–2017  | 1 km                | 16 days              | HMODIS     |

---

![Figure 1](http://iwaponline.com/hr/article-pdf/doi/10.2166/nh.2021.028/905169/nh2021028.pdf) Location of the selected stations for drought modeling.
groundwater, streamflow, and reservoir storage reflect the longer-term precipitation anomalies. The SPI values can be positive or negative: positive SPI values show wet conditions, and negative SPI values show dry conditions. The descriptive features of SPI are listed in Table 2.

**Methods used in data preprocessing**

EMD is a popular and widely used method for time-series processing (Wang et al. 2015). This approach can be followed for processing the nonlinear and nonstationary signals, and for decomposition of these signals into several components terms as inherent mode functions (IMFs). EMD is capable of determining the signal instantaneous frequency. According to Lei et al. (2009), in the process of obtaining the frequency components of a signal by decomposition, first, the components with higher frequency shall be separated, and then the process will continue until the lowest frequency component remains. The ensemble EMD (EEMD) solves the mode mixing issue of EMD via the noise-added data analysis approach (Wu & Huang 2009). This method can be summarized as follows:

1. A random white noise signal is added to the original signal \( x(t) \), and then, the new signal is decomposed into several IMF series.
2. If the number of added white noises is less than the number of trials, step 1 should be repeated; otherwise, step 3 should be performed.
3. For obtaining the ensemble IMF, the average of the sum of all IMFs is computed \( I_j(t) \).
4. Finally, the original signal can be obtained as \( x(t) = \sum_{j=1}^{R} I_j(t) \).

VMD is another relatively new time-series processing method developed to decompose the time series into a sequence of discrete sub-modules with a special frequency band (He et al. 2019). In the VMD method, each IMF is assumed as a finite bandwidth with a different center frequency, and the aim is the minimization of the sum of the estimated bandwidths for each IMF. The detailed description is as follows:

\[
\min_{u_k(t),\delta_k} \left\{ \sum_{k=1}^{K} \left\| \frac{\partial}{\partial t} \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) e^{-jw_k t} \right\|^2 \right\}
\]

subject to \( \sum_{k=1}^{K} u_k(t) = f(t) \)

where \( K, \delta, j, \delta(t), \otimes, u_k(t), w_k \), and \( f(t) \) are the number of modes, partial derivative for time \( t \), Dirac distribution, imaginary unit, convolution operator, \( t \)th data in the \( 4 \)th decomposed mode, center frequency of the \( 4 \)th decomposed mode, and \( 4 \)th data in the original signal, respectively. In general, due to complex features (such as convexity or concavity) in objective and constraint functions in Equation (1), directly detecting the solution of constrained optimization problem is difficult. For reducing the difficulty of solution, He et al. (2019) combined the quadratic penalty function guaranteeing the reconstruction fidelity and the Lagrange multiplier, strictly enforcing constraints for transforming the problem to an unconstrained optimization problem, as shown below:

\[
Z(u_k(t), w_k, \lambda) = a \left\{ \sum_{k=1}^{K} \left\| \frac{\partial}{\partial t} \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) e^{-jw_k t} \right\|^2 \right\} + \left\| \sum_{k=1}^{R} u_k(t) - f(t) \right\|^2 + \lambda \langle t, f(t) - \sum_{k=1}^{K} u_k(t) \rangle
\]

**Table 2** | Classification of drought based on the SPI

| SPI values | Class             | SPI values | Class          |
|------------|-------------------|------------|----------------|
| >2         | Extremely wet     | –1 to –1.49| Moderately dry |
| 1.5–1.99   | Very wet          | –1.5 to –1.99| Severely dry  |
| 1.0–1.49   | Moderately wet    | <–2        | Extremely dry  |
| –0.99 to 0.99 | Near normal    |            |                |
where $Z$, $a$, and $\lambda$ are augmented Lagrange function, penalty parameter, and Lagrange multiplier, respectively. In Equation (2), for optimizing the modified unconstrained problem, the alternate direction method of multiplier method can be used (He et al. 2019). The corresponding solution can be described as below:

$$
\mathbf{1}^{-1}(\mathbf{w}) = \frac{\tilde{f}(\mathbf{w}) - \sum_{i=1}^{K} \mathbf{\pi}(\mathbf{w}) + 0.5\lambda^{\mathbf{2}}(\mathbf{w})}{1 + 2a(w - w_0^{\mathbf{2}})} ; \quad w_k^{\mathbf{2}} = \int_{0}^{\infty} w \pi^{\mathbf{2}}(\mathbf{w})^2 \, dw
$$

(3)

where $\tilde{f}(\mathbf{w})$, $\mathbf{\pi}(\mathbf{w})$, $\lambda(\mathbf{w})$, and $\pi^{\mathbf{2}}(\mathbf{w})$ are Fourier transforms of $f(t)$, $u_i(t)$, $\lambda(t)$, and $u_k^{\mathbf{2}}(t)$, respectively; and $n$ is the number of iterations.

After data decomposition with EEMD and VMD, too many IMF components are obtained, which are needed to be reorganized or regrouped. In this study, PE algorithm was used to carry out IMF reorganization. Based on the Shannon entropy form, the PE $H_p(m)$ of the $k$th different symbol sequence of time series $X(i)$ can be expressed as follows:

$$
H_p(m) = -\sum_{j=1}^{k} p_j \ln p_j
$$

(4)

Based on the above equation, when $P_j$ is $1/m!$, $H_p(m)$ has the maximum value $\ln(m!)$. Standardize $H_p(m)$ for convenience.

$$
H_{PE}(m) = \frac{H_p(m)}{\ln(m!)}
$$

(5)

where $0 \leq H_{PE}(m) \leq 1$. The value of the $H_{PE}$ indicates the time-series randomness and complexity. The larger the $H_{PE}$ value, the stronger the randomness will be and vice versa.

**Data-driven methods for the prediction of drought indices**

**Feed-forward neural network**

The ANN, as a data-driven method, aims at finding a nonlinear relationship between data series of inputs and output. An ANN is based on a connected node collection called neurons (Najafi et al. 2018). These neurons are linked to certain nodes in their neighborhood with varying connectivity coefficients which show the connection strengths. The three-layer FFNN (i.e. input, hidden, and target layers) is one of the most common ANN algorithms, and it has been widely applied for problems related to water resources engineering. The FFNN is considered as a classical type of ANN. In this method, the information transfer is only in one direction: from the input nodes and output nodes, through hidden nodes (if any). The hidden neuron sums up the weights of input connections. For selecting the appropriate information to be passed on to the next neuron, the weighted summation is passed through an activation function in the hidden layer (for more details, see Tayfur 2012).

**KB methods**

Among data-driven techniques, KB methods such as GPR and SVM are considered as relatively innovative and significant techniques in terms of various kernel types and the statistical learning theory (Roushangar et al. 2019). These models can adapt themselves for predicting any parameter of interest by adequate inputs, and numerous kernel-type alternatives exist in this regard (Babovic 2009). Nevertheless, the proper selection of the kernel type is the most crucial step in the GPR and SVM methods because of its direct effect on the precision of classification and training. According to Smola (1996), there are various kernel functions for selecting. For instance, in the SVM, the linear, polynomial, sigmoid functions, and radial basis function (RBF) and in the GPR, the polynomial, normalized polynomial, RBF, and Pearson kernels can be selected based on the nature of the phenomenon to be dealt with.

For assessing the performance and capability of the utilized AI methods, three criteria were used, including determination coefficient (DC), correlation coefficient ($R$), and root-mean-square errors (RMSE), which can be
formulated, respectively, as follows:

\[
DC = \frac{1}{C_0} \sum_{i=1}^{N} (S_{Pi} - S_{Pp})^2 
\]

\[
R = \frac{1}{C_2} \left( \frac{1}{C_0} \sum_{i=1}^{N} (S_{Pi} - S_{Pp})^2 \right) \left( \frac{1}{C_0} \sum_{i=1}^{N} (S_{Pi} - S_{Pp})^2 \right)
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (S_{Pi} - S_{Pp})^2}
\]

where \(S_{Pi}, S_{Pp}, S_{Po}, \) and \(S_{Pp}\) are the observed, predicted, mean observed, and mean predicted values, respectively. ND is the number of data.

**Lower bound and upper bound estimation**

According to Krupnick et al. (2006), the uncertainty of any prediction model contains both uncertainties stemming from the modeling itself and those stemming from prediction. The uncertainty of modeling is due to the inputs, parameters, and structures that generally generate the uncertainty of estimation (Kasiviswanathan et al. 2013). One of the most commonly used methods in quantifying the uncertainty of a model prediction is PI. Generally, the traditional interval prediction tries to make the PI based on point forecasting. The lower and upper bounds can be computed based on the forecasting values and the level of confidence. The point forecasting accuracy plays a key role in the PI accuracy. The lower bound and upper bound estimation (LUBE) attempts to directly approximate the PI upper and lower bounds based on the set of inputs. Based on Meade & Islam (1995), a PI shows the estimation uncertainty of a random parameter future realization and accounts for more uncertainty sources.

Two measures for the quantitative assessment of constructed PIs are PI coverage probability (PICP) and mean PI width (MPIW). PICP measures the percentage of observed values that fall within the prediction band. If the prediction band is wider, it covers most of the observed values. The closer PICP is to its nominal value, the more reliable it is. PICP, described in the following equation, is the measure that shows how good the PIs are (Kasiviswanathan et al. 2013):

\[
PICP = \frac{1}{N} \sum_{i=1}^{N} c_i
\]

if \(L(X_i) < x_i < U(X_i), c_i = 1; \) else \(c_i = 0\)

in which \(L(X_i)\) and \(U(X_i)\) show the PI lower bound and upper bound corresponding to the \(i\)th sample, respectively. The ideal value of PICP could be obtained if the upper and lower bounds of PIs had extremely large values. Actually, too wide PIs are not valuable, because PIs carry no report around the observed values. Therefore, an additional criterion should be used for quantification of the width of prediction bounds as follows (Kasiviswanathan et al. 2013):

\[
MPIW = \frac{1}{N} \sum_{i=1}^{N} L(X_i) - U(X_i)
\]

For comparing the achieved PIs via various methods, the MPIW normalization is needed:

\[
NMPIW = \frac{MPIW}{R}
\]

where \(R\) shows the underlying target range (i.e. difference between the maximum and the minimum of the target amounts). The NMPIW is a dimensionless parameter and demonstrates how narrow PIs are. The consideration of these two measures (NMPIW and PICP) independently may not lead to accurate results, because they have a
contrary relationship. Therefore, both measures should be simultaneously evaluated in the procedure of PI construction.

Overall description of the proposed methodology
In this paper, hybrid models were developed for point and interval predictions of drought time series based on data preprocessing, PE, AI, and LUBE approaches. Three precipitation datasets were used in the calculation of SPI times series. The first of these was ground-based data (precipitation data from 12 synoptic stations), and the other two datasets were based on satellite precipitation data. Given that at least 30-year long data should be used in the SPI series calculations (Mishra & Singh 2010; Sahoo et al. 2015); the PERSIANN-CDR and CMAP satellites datasets were utilized for this aim. Figure 2 shows the applied methodology for data multi-preprocessing. Since in the collection process, many internal and external factors affect the raw data, the integration of unrelated information in the data may increase the difficulty of accurate prediction of the intended parameter. Therefore, at first, the original data were refined via WT. WT was used for the signals de-noising via specifying a threshold for the detail coefficients, which represent noise (Donoho 1995). Figure 3 shows de-noised signals of SPI-9 time series for three selected datasets.

Two scenarios were considered in the modeling process. According to previous studies (e.g. Mokhtarzad et al. 2017; Xu et al. 2018; Zhang et al. 2019; Li et al. 2020; Khan et al. 2020), the previous values of the SPIs, precipitation (P), relative humidity (R), and temperature (T) series are effective in drought modeling. Therefore, in the first scenario, all effective variables were considered as inputs in predicting the SPI time series in the scales of 1-, 9-, and 24-month timescales. FFNN, GPR, and SVM approaches were used for point prediction purpose. In the second scenario, only the decomposed subseries of the SPI time series were used as inputs. Directly selecting of
the original time series as input parameter may affect the capability of the single predicting models. Different types of factors affect the hydrological time series and contribute to nonlinearity and uncertainty properties of time series. This causes the time series to have different resolution components. Therefore, in this study, an attempt was made to develop practical hybrid preprocessing-based AI models to improve the drought prediction efficiency. In this regard, two data processing models, including EEMD and VMD, as well as PE, were applied for original drought series decomposition to increase the efficiency of single AI methods. For analyzing the complexity of each IMFs, component PE was used. Then, for solving the computational burden and over-decomposition issues, the IMFs with similar entropy amounts were recombined. In the next step, hybrid models were used for interval prediction based on the LUBE method. The proposed hybrid models could be described step by step as follows:

1. Calculation of short- to long-term SPI series from the selected datasets.
2. **Decomposition:** the VMD and EEMD approaches were used to decompose time series into several subseries (IMFs).
3. **Calculation of the PE of IMFs:** entropy values of each IMF were computed. Based on the closeness of entropy values, the IMFs were mixed for obtaining a recombinant subseries.
4. **Model development:** the AI models were applied to construct an approximate network.
5. **Model prediction:** combined predicted result of the developed AI models for all the subseries was considered as the final predicted result of the original drought time series. For assessing the efficiency of the models used, the datasets were divided into training, validation, and testing parts. In total, 70% of the datasets from the beginning was utilized to train the developed models, and the rest 30% of data was used for validation (15%) and testing (15%) purposes. It should be noted that in this study, all used variables (input and output parameters) were scaled between 0 and 1 (i.e. normalization) in order to eliminate the dimensions of input and output variables. Data normalization eliminates the influence of the variables with different absolute magnitudes and increases the training speed. The used normalization only rescales the data to another range, which could be de-normalized to real values after modeling. The following equation was used to normalize the utilized data in this study:

\[
x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}}
\]  

(12)

where \(x_n, x, x_{\max}, \text{ and } x_{\min}\), respectively, are the normalized value of variable \(x\), the original value, the maximum of variable \(x\), and the minimum of variable \(x\).

6. **Interval prediction:** the LUBE method was used to determine the uncertainty of predictions performed by developed methods.

**Figure 3** | The de-noised SPI signals via WT.
7. **Verifying the efficiency of developed methods**: in the final step, for verifying the developed hybrid models, the capability of the proposed methods was assessed for the NDVI series modeling.

**RESULTS AND DISCUSSION**

A comparison of the monthly precipitation time series with the long-term average value revealed that the selected region had suffered from frequent droughts. The correlation between ground-based and satellite precipitation data is shown in Figure 4(a). As it could be seen from this figure, there was a good correlation between ground-based and satellite data. Precipitation data obtained from three datasets were used to calculate the SPI time series.

![Figure 4](http://iwaponline.com/hr/article-pdf/doi/10.2166/nh.2021.028/905169/nh2021028.pdf)

**Figure 4**: (a) The correlation between ground-based and satellite precipitation data, (b) zoning map of the selected area based on the mean yearly SPI series, and (c) the SPI series obtained from three datasets.
Figure 4(b) shows the zoning map of the mean yearly SPI values during the selected statistical period for these datasets. Based on this figure, the results of PERSIANN-CDR and CMAP satellite datasets were similar to those of ground-based stations. For all datasets, drought was more severe in the Southern parts of the region, and slight drought had occurred in the Northwest of the selected area. Therefore, the SPI time series obtained from the precipitation products of the PERSIANN-CDR and CMAP satellites were successful in drought monitoring in the selected region. Also, the results of CMAP satellite data were closer to the ground-based station results. Figure 4(c) shows the SPI time series obtained from three datasets at 1-, 9-, and 24-month timescales (SPI-1, SPI-9, and SPI-24). From these SPI time series, it was found that different drought levels occurred in the area, but in most of the cases, the drought events were slight droughts (i.e., \( 1 < \text{SPI} \leq 0 \)). The satellite datasets were used in order to investigate whether it is possible to model drought conditions using the satellite precipitation products and obtain desirable prediction accuracy via the hybrid methodology proposed in this study, especially in the areas without any stations. In the next part of the study, the SPI time series were applied for the prediction of the short-term, mid-term, and long-term droughts.

**Results of AI models for point predictions (Scenario 1)**

In this part, the capability of AI methods was investigated in modeling the SPIs via the original series (i.e. without using data processing methods). The previous values in the SPI time series, relative humidity, average temperature, and precipitation were used as inputs. For obtaining desirable modeling results, setting specific parameters of each AI approach is required, and their optimal values should be determined. For the design of KB approaches, the appropriate kind of kernel function selection should be done. According to Gill et al. (2006), there are various kernel functions for selection; however, among them, the RBF has led to better prediction results in the study of various hydrological phenomena. Also, for the FFNN modeling, various networks were tested to determine the node numbers of hidden layer, since the network topology has direct impacts on the capability of FFNN computational complexity and generalization. Appropriate activation functions for the hidden and output nodes were selected as the tangent sigmoid and pure linear functions, respectively. The backpropagation approach scheme was applied to train the model. The results of the single AI models are shown in Table 3. From the obtained results of statistical parameters (i.e. \( R, DC, \text{RMSE} \)), it can be seen that the prediction accuracy of single methods was almost the same. However, among them, the GPR approach yielded slightly better prediction results. It could be stated that modeling accuracy was deteriorated as the SPI degree reduced. The single AI methods did not provide desirable accuracy in modeling the SPI-1, while their prediction accuracy was satisfactory for the SPI-24. The reason is that the SPI-24 is the normalization of rainfall data over the last 24 months; therefore, unlike the SPI-1, it is not sensitive to data changes from 1 month to the next. In Figure 5, the scatter plots of observed vs. predicted SPI-24 data are shown for the superior model, i.e. GPR.

**Results of hybrid EEMD-AI and VMD-AI for point predictions (Scenario 2)**

The capability of preprocessing methods in enhancing the accuracy of the models was investigated in this part. In this regard, at first, the de-noised SPI time series were decomposed using EEMD and VMD methods. According to Wu & Huang (2009), the principle of EEMD is decomposition of signal to different IMFs and one residual signal. The sum of these signals will be the same original signal. In the VMD method, signal is decomposed to several IMFs. Via EEMD and VMD, the original drought time series were decomposed into a series of disjoint subseries; then, each subseries was predicted via an appropriate GPR model construction. Since the prediction results of the single GPR model were somewhat better than those of the FFNN and SVM models, this method was used in hybrid models for prediction aims. The aggregated output of all predicted subseries was considered as the final simulated output. Figure 6(a) shows the decomposition results of two preprocessing techniques for the SPI-9 obtained from ground-based data. From the results, it can be indicated that the decomposed high-frequency IMFs via the VMD method had a relatively stable trend, which was conducive for prediction. Since the final forecast result of the VMD method was an accumulation of the forecast result of each IMFs, the characteristics of VMD result helped to enhance the prediction accuracy. However, end-point impacts (large fluctuations) occurred during the EEMD decomposition process and affected all IMFs continuously. Therefore, the EEMD prediction result led to a large error.

In the second step, the EEMD, VMD, and PE (\( H_{PE} \)) were used for decomposing the original drought series into several subseries to alleviate the modeling complexity. For analyzing each IMF complexity and recombination of IMFs that have similar entropy amounts, the PE was used. The PE values for SPI-9 via all used datasets are shown
in Figure 6(b). For avoiding excessive decomposition and decreasing the scale of calculation, the IMFs with similar entropy were combined (see Figure 6(b)).

Table 4 and Figure 7 show the results of ensemble data processing models. By comparing the presented results in Tables 3 and 4, it was found that the signal processing enhanced the prediction accuracy of drought series. The ensemble methods were more efficient than single AI approaches. As it could be seen, the RMSE values of the GPR model for the test series of PERSIANN-CDR data were 0.726 for the SPI-1, 0.578 for the SPI-9, and 0.421 for the SPI-24; the error criterion values of the PE–VMD–GPR reduced to 0.393 for the SPI-1, 0.314 for the SPI-9, and 0.283 for the SPI-24. Therefore, using the ensemble methods, the short- to long-term SPI modeling was performed more accurately. Also, the predictive effect of the PE–VMD–GPR method was significantly better than others. Generally, the data preprocessing (VMD and EEMD) enhanced the modeling efficiency from 25 to 40% in testing sets. For hybrid PE methods, the efficiency improvement of the models was between 40 and 55%.

**Table 3 | Statistical parameters of the single AI models for three datasets**

| Model | Validation | Testing | Validation | Testing |
|-------|------------|---------|------------|---------|
|       | R | DC | RMSE | R | DC | RMSE | R | DC | RMSE | R | DC | RMSE | R | DC | RMSE |
| SPI 1 | Ground-based data | PERSIANN-CDR data | CMAP data |       |     |     |     |     |     |     |     |     |     |     |     |
| GPR   | 0.631 | 0.491 | 0.821 | 0.588 | 0.473 | 0.841 | 0.657 | 0.579 | 0.705 | 0.597 | 0.506 | 0.726 |
| SVM   | 0.547 | 0.426 | 0.893 | 0.539 | 0.428 | 0.965 | 0.562 | 0.494 | 0.792 | 0.538 | 0.478 | 0.795 |
| FFNN  | 0.562 | 0.437 | 0.873 | 0.525 | 0.419 | 0.973 | 0.592 | 0.515 | 0.778 | 0.558 | 0.475 | 0.797 |
|       |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SPI 9 | Ground-based data | PERSIANN-CDR data | CMAP data |       |     |     |     |     |     |     |     |     |     |     |
| GPR   | 0.837 | 0.766 | 0.701 | 0.817 | 0.720 | 0.746 | 0.829 | 0.801 | 0.504 | 0.802 | 0.727 | 0.578 |
| SVM   | 0.808 | 0.739 | 0.775 | 0.779 | 0.685 | 0.780 | 0.789 | 0.706 | 0.584 | 0.772 | 0.679 | 0.642 |
| FFNN  | 0.811 | 0.723 | 0.805 | 0.779 | 0.666 | 0.821 | 0.779 | 0.686 | 0.628 | 0.734 | 0.638 | 0.651 |
|       |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SPI 24| Ground-based data | PERSIANN-CDR data | CMAP data |       |     |     |     |     |     |     |     |     |     |     |
| GPR   | 0.918 | 0.854 | 0.488 | 0.849 | 0.815 | 0.503 | 0.906 | 0.871 | 0.395 | 0.876 | 0.816 | 0.421 |
| SVM   | 0.841 | 0.793 | 0.526 | 0.811 | 0.782 | 0.658 | 0.891 | 0.806 | 0.412 | 0.864 | 0.781 | 0.514 |
| FFNN  | 0.829 | 0.784 | 0.530 | 0.795 | 0.769 | 0.679 | 0.875 | 0.791 | 0.511 | 0.856 | 0.773 | 0.523 |
|       |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

**Constructed PIs using the LUBE method for the SPI time series**

For assessing the capability of the LUBE method in constructing the PIs of short- to long-term SPI time series, the IMFs were used as the GPR model inputs to project the SPI values (point predictions) and related PIs. Due to the better prediction results of single GPR model, this method was used for PI prediction aim. The associated confidence level for all PIs was considered as 95%. Two criteria, PICP and NMPIW, were used to determine the optimum structure of GPR in the PI construction (with the target to maximize PICP and minimize NMPIW). Table 5 and Figure 8 show the obtained results. As can be seen, the PICP values for the hybrid PE–VMD–GPR model were up to 24, 19, and 12% higher than the single GPR, while the NMPIW values were up to 30, 26, and 32% less than the single GPR, for the SPI-1, SPI-9, and SPI-24, respectively, in comparison with other
Therefore, the hybrid PE–VMD–GPR model led to significantly better outcomes both in point and interval predictions.

Verifying the capability of developed methods for modeling NDVI time series

According to Tucker (1979), satellite products make it possible to monitor vegetation density via an index called vegetation index, which makes the use of the difference in spectral reflection of green vegetation in the near-infrared (NIR) and the red parts of the spectrum (RED) (Fensholt et al. 2009). The data provided by weather satellites like Terra can be used for obtaining the vegetation index time series over long periods. Many researchers (e.g., Nicholson et al. 1990; Liu & Juárez 2001; Karabulut 2003; Fensholt et al. 2009) assessed the relationship between precipitation and NDVI in different areas and indicated that the NDVI can be successfully used for drought monitoring. Therefore, in this study, NDVI series was further used for verifying the proposed methods. The NDVI values vary from $-1$ to $+1$, and according to Tucker (1979), higher NDVI values show the greater photosynthetic capacity of vegetation canopy. The NDVI series is defined below:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (13)$$

In this study, the NDVI time series were obtained for the period of February 2000 to December 2017 from MODIS. The 16-day composite NDVI images were used for obtaining the monthly average NDVI values. In Figure 9(a), the NDVI images of one dry-year and one wet-year are shown. As seen, the NDVI successfully showed vegetation conditions in these 2 years, and it could be used for effectively assessing and monitoring the changes in vegetation cover. In the next step, the developed methods were applied for point and interval predictions of the monthly NDVI time series. The obtained results are listed in Table 6 and shown in Figure 9(b). From the analysis, it can be seen that the hybrid models were successful in the modeling process and the PE–
VMD–GPR approach resulted in more reliable outcomes in both point and interval predictions of the monthly NDVI time series. The PICP value for the hybrid PE–VMD–GPR model was up to 15% higher than the single GPR, whereas the NMPIW value was approximately 37% less than GPR.

CONCLUSIONS

The accurate prediction of drought is crucial in understanding the drought process and improving water management practices. The current study investigated the capability of single and integrated AI methods in point and interval drought predictions, considering short-, mid-, and long-term timescales. The ground-, PERSIANN-CDR-, and CMAP-based precipitation datasets were used for the prediction of 1-, 9-, and 24-month timescale SPI time series. First, unprocessed time series were applied in the modeling process; then, these time series were decomposed into IMF subseries using EEMD and VMD. The PE was used to select the IMFs with similar entropy values and recombine them. According to the results, the single AI methods did not lead to desirable efficiency in short-term drought modeling in the selected area. For longer timescale SPI time series, the model efficiency was improved, and standalone methods were slightly better in SPI-24 modeling. It was shown that hybrid preprocessing methods enhanced the performance of standalone AI approaches approximately 25–40%, whereas hybrid PE methods enhanced the model performance between 40 and 55%. On the other hand, since point predictions via AI methods do not convey any details about the uncertainty of prediction, PI can
be an essential index to quantify the reliability of AI-based modeling of hydrological time series. In this study, PIs of the GPR-based model were obtained using the LUBE method. The NMPIW and PICP criteria were used to quantitatively evaluate the efficiency of the constructed PIs. Obtained results indicated that the VMD–PE–GPR led to more desirable results. The PICP of this hybrid model was approximately 24, 19, and 12% higher, and the NMPIW was approximately 30, 26, and 32% lower than the GPR method for SPI-1, SPI-9, and SPI-24, respectively. Furthermore, the performance of the developed models was evaluated for the NDVI series modeling. It was found that the proposed hybrid models were successful in the modeling process and provided desirable accuracy. The results obtained in this study supported the premise that hybrid data processing models had considerable potential to be used as an alternative approach for drought prediction based on ground-based and remote sensing data. For the prediction of complex short- to long-term hydrological time series, the developed hybrid models can be used. It should, however, be noted that the AI models are data-driven models, and thus, they are data sensitive. That is why studies on regions with different climatic characteristics are required to conclusively prove the advantages of the proposed models in predicting the drought time series. In addition, different types and lengths of meteorological data and satellite datasets used in this paper
Figure 7 | The observed vs. predicted SPI-24 scatter plots for the PE-VMD-GPR model.

Table 5 | Constructed PI results for three datasets of LUBE methods

| Model  | Ground-based data | PERSIANN-CDR data | CMAP data |
|--------|-------------------|-------------------|-----------|
|        | PICP   | NMP IW | PICP   | NMP IW | PICP   | NMP IW |
| SPI 1  |        |        |        |        |        |        |
| GPR    | 0.77   | 2.49   | 0.75   | 2.57   | 0.78   | 2.55   |
| EEMD-GPR | 0.91   | 2.21   | 0.90   | 2.22   | 0.89   | 2.23   |
| PE-EEMD-GPR | 0.94   | 1.78   | 0.94   | 1.82   | 0.93   | 1.64   |
| VMD-GPR | 0.94   | 1.79   | 0.95   | 1.68   | 0.95   | 1.08   |
| PE-VMD-GPR | 0.96   | 0.80   | 0.95   | 1.03   | 0.96   | 1.02   |
| SPI 9  |        |        |        |        |        |        |
| GPR    | 0.82   | 0.71   | 0.84   | 0.81   | 0.86   | 0.77   |
| EEMD-GPR | 0.93   | 0.69   | 0.91   | 0.79   | 0.91   | 0.73   |
| PE-EEMD-GPR | 0.96   | 0.55   | 0.97   | 0.51   | 0.96   | 0.54   |
| VMD-GPR | 0.96   | 0.62   | 0.93   | 0.65   | 0.95   | 0.63   |
| PE-VMD-GPR | 0.98   | 0.56   | 0.95   | 0.54   | 0.98   | 0.55   |
| SPI 24 |        |        |        |        |        |        |
| GPR    | 0.88   | 0.64   | 0.86   | 0.65   | 0.90   | 0.67   |
| EEMD-GPR | 0.95   | 0.55   | 0.94   | 0.59   | 0.92   | 0.61   |
| PE-EEMD-GPR | 0.96   | 0.49   | 0.94   | 0.48   | 0.96   | 0.49   |
| VMD-GPR | 0.97   | 0.48   | 0.95   | 0.52   | 0.97   | 0.51   |
| PE-VMD-GPR | 0.99   | 0.48   | 0.97   | 0.51   | 0.98   | 0.47   |
Figure 8 | Results of the PI construction by the LUBE method for (a) CMAP data, (b) ground-based data, and (c) PERSIANN-CDR data.

Figure 9 | (a) Comparison of the NDVI images of one dry-year (2008) and one wet-year (2010) for the selected area and (b) results of the PIs construction by the LUBE method for NDVI modeling.
may affect the fitting and prediction accuracy of meteorological and remote sensing data models, which should be further discussed in the future.

**DATA AVAILABILITY STATEMENT**

All relevant data are included in the paper or its Supplementary Information.

**REFERENCES**

Abdoos, A. A. 2016 A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* **205**, 111–120.

Adamowski, K., Prokop, A. & Adamowski, J. 2009 Development of a new method of wavelet aided trend detection and estimation. *Hydrological Processes: An International Journal* **23** (18), 2686–2696.

Aghelpour, P., Mohammadi, B., Biazar, S. M., Kisi, O. & Sourmirinezhad, Z. 2020a A theoretical approach for forecasting different types of drought simultaneously, using entropy theory and machine-learning methods. *ISPRS International Journal of Geo-Information* **9** (12), 701.

Aghelpour, P., Bahrami-Pichaghchi, H. & Kisi, O. 2020b Comparison of three different bio-inspired algorithms to improve ability of neuro fuzzy approach in prediction of agricultural drought, based on three different indexes. *Computers and Electronics in Agriculture* **170**, 105279.

Aghelpour, P., Guan, Y., Bahrami-Pichaghchi, H., Mohammadi, B., Kisi, O. & Zhang, D. 2020c Using the MODIS sensor for snow cover modeling and the assessment of drought effects on snow cover in a mountainous area. *Remote Sensing* **12** (20), 3437.

Ashouri, H., Hsu, K. L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R. & Prat, O. P. 2015 PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society* **96** (1), 69–83.

Azamathulla, H. M., Haghiabi, A. H. & Parsaie, A. 2016 Prediction of side weir discharge coefficient by support vector machine technique. *Water Science and Technology: Water Supply* **16** (4), 1002–1016.

Babovic, V. 2009 Introducing knowledge into learning based on genetic programming. *Journal of Hydroinformatics* **11**(3–4), 181–193.

Basalirwa, C., Odiyo, J., Mngodo, R. & Mjeta, E. 1999 The climatological regions of Tanzania based on the rainfall characteristics. *International Journal of Climatology* **19** (1), 69–80.

Brown, S. C., Versace, V. L., Lester, R. E. & Walter, M. T. 2015 Assessing the impact of drought and forestry on streamflows in south-eastern Australia using a physically based hydrological model. *Environmental Earth Sciences* **74**, 6047–6063.

Chadalawada, J., Havlicek, V. & Babovic, V. 2017 A genetic programming approach to system identification of rainfall-runoff models. *Water Resources Management* **31** (12), 3975–3992.

Chen, T., De, J. R. A. M., Liu, Y., Van, D. W. G. R. & Dolman, A. J. 2010 Using satellite based soil moisture to quantify the water driven variability in NDVI: a case study over mainland Australia. *Ecological Informatics* **5**(5), 400–409.

Deo, R. C., Kisi, O. & Singh, V. P. 2017 Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and m5tree model. *Atmospheric Research* **184**, 149–175.

Donoho, D. L. 1995 De-noising by soft-thresholding. *IEEE Transactions on Information Theory* **41**(3), 613–627.

Fensholt, R., Rasmussen, K., Nielsen, T. T. & Mbow, C. 2009 Evaluation of earth observation based long term vegetation trends-intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data. *Remote Sensing of Environment* **113**(9), 1886–1898.

Geng, G., Wu, J., Wang, Q., Lei, T., He, B., Li, X., Mo, X., Luo, H., Zhou, H. & Liu, D. 2016 Agricultural drought hazard analysis during 1980–2008: a global perspective. *International Journal of Climatology* **36**(1), 389–399.

Ghulam, A., Qin, Q., Teyp, T. & Li, Z. L. 2007 Modified perpendicular drought index (MPDI): a real-time drought monitoring method. *ISPRS Journal of Photogrammetry and Remote Sensing* **62**(2), 150–164.

---

**Table 6** | Statistical parameters and constructed PI results of the hybrid GPR models for the NDVI modeling

| Model          | Validation |            | Testing |            |            |            |            |            |
|----------------|------------|------------|---------|------------|------------|------------|------------|------------|
|                | R  | DC  | RMSE | R       | DC  | RMSE | PICP | NMIPIW |
| GPR            | 0.814 | 0.802 | 0.038 | 0.814 | 0.794 | 0.041 | 0.82 | 0.29 |
| EEMD–GPR       | 0.833 | 0.855 | 0.029 | 0.832 | 0.818 | 0.033 | 0.88 | 0.21 |
| PE–EEMD–GPR    | 0.924 | 0.913 | 0.016 | 0.901 | 0.901 | 0.021 | 0.96 | 0.13 |
| VMD–GPR        | 0.901 | 0.905 | 0.018 | 0.912 | 0.892 | 0.024 | 0.91 | 0.14 |
| PE–VMD–GPR     | 0.948 | 0.951 | 0.012 | 0.922 | 0.908 | 0.017 | 0.95 | 0.11 |
Gill, M. K., Asefa, T., Kemblowski, M. W. & McKee, M. 2006. Soil moisture prediction using support vector machines 1. *JAWRA Journal of the American Water Resources Association* 42 (4), 1033–1046.

Hayes, M. J., Svoboda, M. D., Wilhite, D. A. & Vanyarkho, O. V. 1999. Monitoring the 1996 drought using the standardized precipitation index. *Bulletin of the American Meteorological Society* 80 (3), 429–438.

He, F., Zhou, J., Feng, Z. K., Liu, G. & Yang, Y. 2019. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Applied Energy* 257, 103–116.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C. & Liu, H. H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A* 454 (1971), 903–995.

Karabulut, M. 2005. An examination of relationships between vegetation and rainfall using maximum value composite AVHRR-NDVI data. *Turkish Journal of Botany* 27 (2), 93–101.

Kasiviswanathan, K. S., Cibin, R., Sudheer, K. & Chauhey, I. 2015. Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. *Journal of Hydrology* 499, 275–288.

Khan, M. M. H., Muhammad, N. S. & El-Shafie, A. 2020. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *Hydrology* 900, 125380.

Khosravi, A., Nahavandi, S., Creighton, D. & Atiya, A. F. 2011. Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks* 22 (9), 1341–1356.

Kisi, O., Gorgij, A. D., Zounemat-Kermani, M., Mahdavi-Meymand, A. & Kim, S. 2019. Drought forecasting using novel heuristic methods in a semi-arid environment. *Journal of Hydrology* 578, 124053.

Krupnick, A., Morgenstern, R., Batz, M., Nelson, P., Burtraw, D., Shih, J. S. & McWilliams, M. 2006. *Not a Sure Thing: Making Regulatory Choices under Uncertainty*. Resources for the Future, Washington, DC.

Kumar, P. & Foufoula, G. E. 1993. Multicomponent decomposition of spatial rainfall fields, 1. Segregation of large- and small-scale features using wavelet transforms. *Water Resources Research* 29 (8), 2515–2532.

Kwak, J., Kim, S., Jung, J., Singh, V. P., Lee, D. R. & Kim, H. S. 2016. Assessment of meteorological drought in Korea under climate change. *Advances in Meteorology* 2016, 1–13.

Lei, Y., He, Z. & Zi, Y. 2009. Application of the EEMD method to rotor fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing* 25 (4), 1327–1338.

Li, X., He, B., Quan, X., Liao, Z. & Bai, X. 2015. Use of the standardized precipitation evapotranspiration index (SPEI) to characterize the drying trend in southwest China from 1982–2012. *Remote Sensing* 7 (8), 10917–10937.

Li, Z., Chen, T., Wu, Q., Xia, G. & Chi, D. 2020. Application of penalized linear regression and ensemble methods for drought forecasting in Northeast China. *Meteorology and Atmospheric Physics* 132 (1), 113–130.

Liu, W. T. & Juárez, R. N. 2001. ENSO drought onset prediction in northeast Brazil using NDVI. *Remote Sensing* 17 (6), 626–638.

Liu, W. & Juárez, R. N. 2001. NDVI data. *Journal of Hydrology* 275 (1-2), 275–288.

Mann, M. E. & Gleick, P. H. 2015. Climate change and California drought in the 21st century. *Proceedings of the National Academy of Sciences* 112 (13), 3858–3859.

Marvel, K., Cook, B. I., Bonfils, C. J., Durack, P. J., Smardon, J. E. & Williams, A. P. 2019. Twentieth-century hydroclimate changes consistent with human influence. *Nature* 569, 59–65.

McKee, T. B., Doesken, N. J. & Kleist, J. 1993. The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th Conference on Applied Climatology*, pp. 179–184.

Meade, N. & Islam, T. 1995. Prediction intervals for growth curve forecasts. *Journal of Forecasting* 14, 413–430.

Mehr, A. D., Kahya, E. & Özger, M. 2014. A gene-wavelet model for long lead time drought forecasting. *Hydrology* 517, 691–699.

Mirabbasi, R., Anagnostou, E. N., Fahrner-Fard, A., Dinpashoh, Y. & Eslamian, S. 2015. Analysis of meteorological drought in northwest Iran using the Joint Deficit Index. *Journal of Hydrology* 492), 35–48.

Mishra, A. K. & Singh, V. P. 2010. A review of drought concepts. *Hydrology* 391 (1-2), 202–216.

Mishra, A., Desai, V. & Singh, V. 2007. Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering* 12 (6), 626–638.

Mohamadi, S., Sammen, S. S., Panahi, F., Ehteram, M., Kisi, O., Mosavi, A., Ahmed, A. N., El-Shafie, A. & Al-Ansari, N. 2020. Zoning map for drought prediction using integrated machine learning models with a nomadic people optimization algorithm. *Natural Hazards* 104 (1), 537–579.

Mokhtarzad, M., Eskandari, F., Vanjani, N. J. & Arabasadi, A. 2017. Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environmental Earth Sciences* 76 (21), 1–10.

Mukherjee, S., Mishra, A. & Treberth, K. E. 2018. Climate change and drought: a perspective on drought indices. *Current Climate Change Reports* 4, 145–165.

Najafi, B., Faizollahzadeh Ardabili, S., Shamsirband, S., Chau, K. W. & Rabczuk, T. 2018. Application of ANNs, ANFIS and RSM to estimating and optimizing the parameters that affect the yield and cost of biodiesel production. *Engineering Applications of Computational Fluid Mechanics* 12 (1), 611–624.

Nicholson, S. E., Davenport, M. L. & Malo, A. R. 1990. A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. *Climatic Change* 17 (2–3), 209–241.

Nischetha, V., Ahmed, S., Varikoden, H. & Raveendevar, J. 2014. The impact of seasonal rainfall variability on NDVI in the Tunga and Bhadra river basins, Karnataka, India. *International Journal of Remote Sensing* 35 (23), 8025–8043.
Roushangar, K. & Alizadeh, F. 2018 Entropy-based analysis and regionalization of annual precipitation variation in Iran during 1960–2010 using ensemble empirical mode decomposition. *Hydroinformatics* 2 (2), 468–485.

Roushangar, K., Nourani, V. & Alizadeh, F. 2019 Corrigendum to a multiscale time-space approach to analyze and categorize the precipitation fluctuation based on the wavelet transform and information theory concept. *Hydrology Research* 49 (3), 724–743.

Roushangar, K., Chamani, M., Ghasempour, R., Azamathulla, H. M. & Alizadeh, F. 2021 A comparative study of wavelet-and empirical mode decomposition-based GPR models for river discharge relationship modeling at consecutive hydrometric stations. *Water Supply*. doi.org/10.2166/ws.2021.073.

Sahoo, A. K., Sheffield, J., Pan, M. & Wood, E. F. 2015 Evaluation of the tropical rainfall measuring mission multi-satellite precipitation analysis (TMPA) for assessment of large-scale meteorological drought. *Remote Sensing of Environment* 159, 181–193.

Sakizadeh, M., Mohamed, M. M. & Klammler, H. 2019 Trend analysis and spatial prediction of groundwater levels using time series forecasting and a novel spatio-temporal method. *Water Resources Management* 33 (4), 1425–1437.

Smola, A. J. 1996 *Regression Estimation with Support Vector Learning Machines*. Master’s Thesis, Technische Universität München, Germany.

Tayfur, G. 2012 *Soft Computing Methods in Water Resources Engineering: Artificial Neural Networks, Fuzzy Logic, and Genetic Algorithm*. WIT Press, Southampton, UK.

Tsakiris, G. & Vangelis, H. 2004 Towards a drought watch system based on spatial SPI. *Water Resources Management* 18 (1), 1–12.

Tucker, C. J. 1979 Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8 (2), 127–150.

Von Hardenberg, J., Meron, E., Shachak, M. & Zarmi, Y. 2001 Diversity of vegetation patterns and desertification. *Physical Review Letters* 87 (19), 198101–198104.

Wang, W. C., Chau, K. W., Qiu, L. & Chen, Y. B. 2015 Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. *Environmental Research* 139, 46–54.

Wu, Z. H. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise assisted data analysis method. *Advances in Adaptive Data Analysis* 1, 1–41.

Wu, L., Peng, Y., Fan, J. & Wang, Y. 2019 Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. *Hydrology Research* 50 (6), 1730–1750.

Xie, P. & Arkin, P. A. 1996 Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *Journal of Climate* 9 (4), 840–858.

Xu, L., Chen, N., Zhang, X. & Chen, Z. 2018 An evaluation of statistical, NMME and hybrid models for drought prediction in China. *Hydrology* 566, 235–249.

Yilmaz, M. T., Hunt, E. R. & Jackson, T. J. 2008 Remote sensing of vegetation water content from equivalent water thickness using satellite imagery. *Remote Sensing of Environment* 112 (5), 2514–2522.

Yin, Z., Wen, X., Feng, Q., He, Z., Zou, S. & Yang, L. 2017 Integrating genetic algorithm and support vector machine for modeling daily reference evapotranspiration in a semi-arid mountain area. *Hydrology Research* 48 (5), 1177–1191.

Zhang, R., Chen, Y. Z., Xu, L. J. & Ou, C. Q. 2019 Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. *Science of the Total Environment* 665, 338–346.

Zhou, Q., Leng, G. & Peng, J. 2018 Recent changes in the occurrences and damages of floods and droughts in the United States. *Water* 10 (9), 1109.

First received 24 March 2021; accepted in revised form 21 May 2021. Available online 18 June 2021