RESEARCH ARTICLE

Comparison of hybrid machine learning methods for the prediction of short-term meteorological droughts of Sakarya Meteorological Station in Turkey

Hatice Citakoglu[1] · Ömer Coşkun[2]

Abstract

Drought is a harmful natural disaster with various negative effects on many aspects of life. In this research, short-term meteorological droughts were predicted with hybrid machine learning models using monthly precipitation data (1960–2020 period) of Sakarya Meteorological Station, located in the northwest of Turkey. Standardized precipitation index (SPI), depending only on precipitation data, was used as the drought index, and 1-, 3-, and 6-month time scales for short-term droughts were considered. In the prediction models, drought index was predicted at $t+1$ output variable by using t, $t-1$, $t-2$, and $t-3$ input variables. Artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), Gaussian process regression (GPR), support vector machine regression (SVMR), k-nearest neighbors (KNN) algorithms were employed as stand-alone machine learning methods. Variation mode decomposition (VMD), discrete wavelet transform (DWT), and empirical mode decomposition (EMD) were utilized as pre-processing techniques to create hybrid models. Six different performance criteria were used to assess model performance. The hybrid models used together with the pre-processing techniques were found to be more successful than the stand-alone models. Hybrid VMD-GPR model yielded the best results ($NSE=0.9345$, $OI=0.9438$, $R^2=0.9367$) for 1-month time scale, hybrid VMD-GPR model ($NSE=0.9528$, $OI=0.9559$, $R^2=0.9565$) for 3-month time scale, and hybrid DWT-ANN model ($NSE=0.9398$, $OI=0.9483$, $R^2=0.9450$) for 6-month time scale. Considering the entire performance criteria, it was determined that the decomposition success of VMD was higher than DWT and EMD.

Keywords Meteorological drought · SPI · Machine learning · Hybrid models · Sakarya · Turkey

Introduction

Droughts have recently been encountered in many parts and different climate zones of the world. Droughts are classified as common and recurrent natural disasters. A clear definition of drought was made at the United Nations Combat Desertification Symposium held in Italy in 1997. It was defined as "a natural event that causes hydrological balance to be disrupted and water and soil resources to be adversely affected due to precipitation falling significantly below the recorded normal levels" (Yu et al. 2022). Drought is classified by researchers based on several different criteria. Three different types of droughts are generally used in the literature: meteorological, agricultural, and hydrological droughts. In some sources, socio-economic drought was also mentioned. Lack of precipitation is defined as "meteorological drought," decrease/lack of surface or ground water is defined as "hydrological drought," restriction of agricultural production due to lack of precipitation and surface or ground water is defined as "agricultural drought," and finally, lack of water that affects the production and consumption activities of the society is defined as "socio-economic drought." Besides these drought types, various indices have been developed by researchers to designate the frequency, severity, duration, and geographic distribution of droughts. Drought indices such as the "standardized precipitation index" (SPI) (McKee et al. 1993), "standardized precipitation evapotranspiration

✉ Hatice Citakoglu
hcitakoglu@erciyes.edu.tr

1 Department of Civil Engineering, Erciyes University, Kayseri, Turkey

2 Turkish General Directorate of State Hydraulic Works (DSI), Kayseri, Turkey

index" (SPEI) (Vicente-Serrano et al. 2010), and "Palmer drought severity index" (PDSI) (Palmer 1965) have extensively been used for monitoring meteorological droughts over the world (Salimi et al. 2021). Current and future trends of drought should be analyzed reliably, and realistic policies should be developed accordingly. In this sense, it becomes clear how important the precipitation and drought predictions are. Especially in recent years, artificial intelligence techniques have become popular tools to be used in drought prediction studies. Mishra et al. (2007) proposed a hybrid version of a linear stochastic model and a nonlinear neural network model for drought prediction in Kansabati River Basin in India. This hybrid model included autoregressive integrated moving average (ARIMA), artificial neural network (ANN), time series, and complex autocorrelation structures. Performance criteria revealed that this hybrid model predicted droughts more accurately than both the stochastic model and the artificial neural network (Mishra et al. 2007). Belayneh et al. (2016) created five different models to predict long-term drought incidences in the Awash River Basin of Ethiopia. Five models included conventional stochastic model (ARIMA), ANN, and SVM with wavelet transform. Hybrid versions with wavelet transform were superior to other stand-alone models in drought predictions (Belayneh et al. 2016). Başakın et al. (2019) tried to predict the future Palmer drought severity index values using machine learning (ML) algorithm from 116 years of precipitation data of Kayseri Region in Turkey. The accuracy of predictions made by SVM and K-nearest neighbors (KNN) algorithms was measured statistically. ML technologies contributed significantly to the solution of hydrological problems (Başakın et al. 2019). Kaur and Sood (2020) stated that there was a need to establish an automatic system that works globally because the current drought indices were not universal. Therefore, a framework model has been created for the evaluation of drought-causing parameters with the use of ANN, ANN optimized with genetic algorithm (GA), and deep neural networks (DNN) capabilities. Support vector regression (SVR) method was used to predict drought incidences in three different climate zones and three different time frames. SVR model showed a high performance in terms of accuracy, sensitivity, and originality (Kaur and Sood 2020). Fadaei-Kermani and Ghaeini-Hessaroeyeh (2020) proposed a new strategy based on fuzzy-KNN model to deal with drought monitoring. A method was presented to anticipate the most likely drought situations using the standardized precipitation index and the fuzzy-KNN methodology. The model was used to monitor droughts in Kerman, located in the southeast of Iran. In recent years, relevant area has experienced severe droughts and rainfall deficits (Fadaei-Kermani and Ghaeini-Hessaroeyeh 2020). Özger et al. (2020) used pre-processing techniques (empirical mode decomposition (EMD) and wavelet transform (WD)) to predict 1-, 3-, and 6-month

self-calibrated Palmer drought severity index (sc-PDSI) values of Adana and Antalya provinces of Turkey using stand-alone M5 model tree, adaptive neuro-fuzzy inference system (ANFIS), and support vector machines (SVM) methods. WD predictions were more accurate than EMD forecasts, and selection of proper wavelet type had a substantial impact on the results (Özger et al. 2020).

This study focused on prediction of short-term meteorological droughts in Sakarya Meteorological Station located in the northwest of Turkey by using SPI drought values at 1-, 3-, and 6-month time scales calculated from monthly precipitation data of 1960–2020 period with the use of hybrid prediction models. Five different ML methods to be used in stand-alone prediction models were ANN, ANFIS, GPR, SVM, and KNN. Besides these stand-alone models, three different pre-processing techniques (DWT, EMD, and VMD) were used. Fifteen hybrid versions were created by using three pre-processing techniques together with five stand-alone ML methods. Prediction performance of the stand-alone and hybrid versions was compared with each other. The effects of pre-processing techniques on prediction models were discussed extensively. This is the first study in which DWT, EMD, and VMD were considered together as a pre-processing technique in drought prediction research.

Materials and methods

Study area and meteorological data

In this study, SPI values calculated from the monthly precipitation data recorded between 1960 and 2020 of Sakarya Meteorology Station located in the northwest of Turkey were used. The station is located at latitude 40°76′ and longitude 30°39′, and its geographic location is shown in Fig. 1. Black Sea climate with hot summers and warm winters is dominant in Sakarya region. Precipitation is encountered in all seasons. The coldest month is January is the coldest month, and July is the hottest month. The annual average temperature is 14.6 °C, and the monthly average precipitation is 71.10 mm. The highest daily total precipitation (127.7 mm) was measured on 26 June 1999 (Republic of Turkey 2022). Change in annual total precipitation from 1960 to 2020 is presented in Fig. 2.

The standardized precipitation index

SPI, proposed by McKee et al. (1993), calculates the meteorological drought index based only on the precipitation parameter. With the use of SPI, arid or humid conditions and anomalies can be determined at a certain time scale anywhere in the world from precipitation data records. For each time scale, a functional and quantitative description
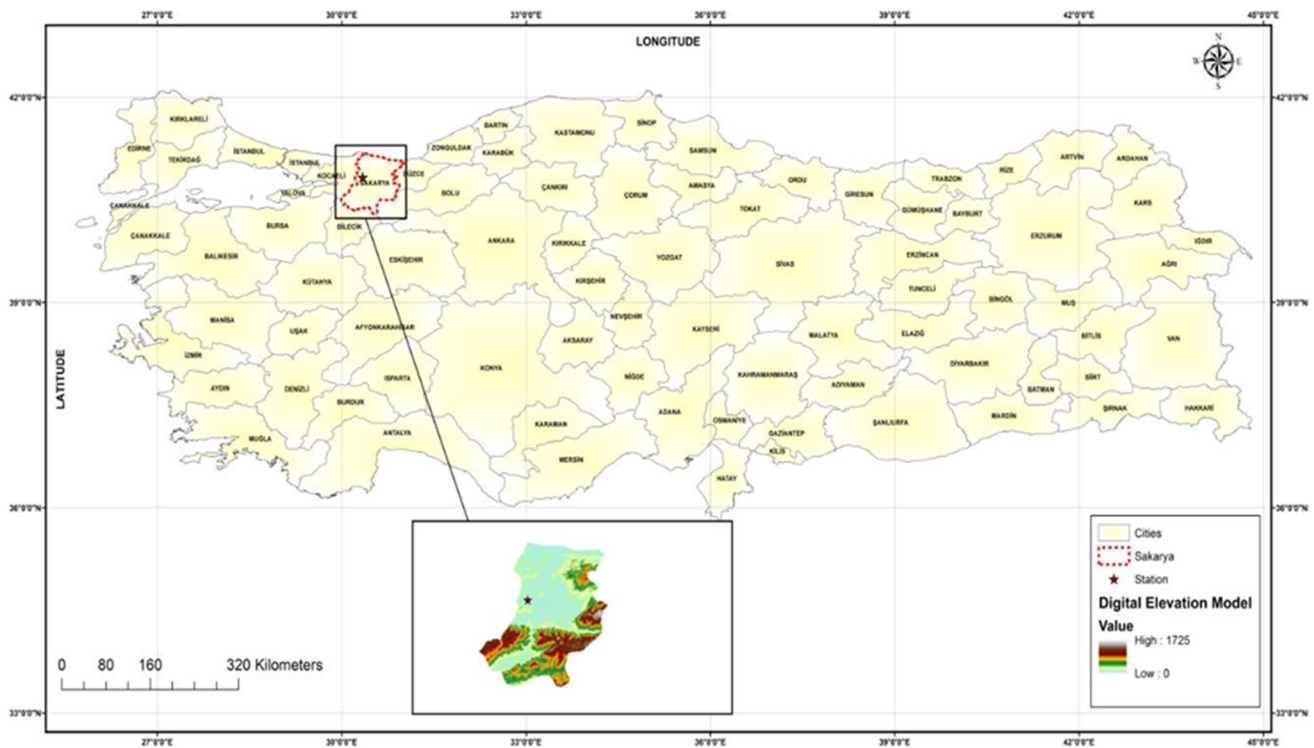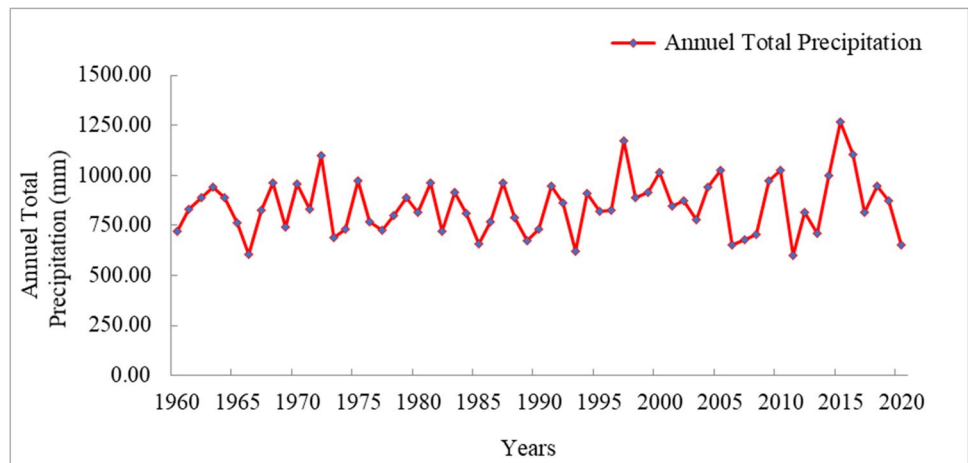
Fig. 1 Study area

Fig. 2 Annual total precipitation from 1960 to 2020 for Sakarya Meteorology Station

of drought can be constructed using the SPI as the drought index. It was initially developed according to the Gamma distribution, but later, it was found to be suitable for the normal distribution. It is calculated with the use of Eq. 1:

$$\text{SPI} = \frac{x_i - \bar{x}}{\sigma} \tag{1}$$

where x_i represents precipitation value for ith time, $\bar{x}$ represents average of precipitation values, and σ represents standard deviation (McKee et al. 1993). The advantages of SPI are as follows: it depends only on precipitation; it is easy to determine the beginning and end of meteorological drought; SPI is only about probability; it is very practical and easy to calculate; SPI provides early drought warning for different time scales; it is less complex than the Palmer drought index (PDI), since it conforms to the normal distribution; and the humid period can be followed as well as the dry period (Sevinc and Sen 2003). The classification of SPI accepted in the literature is provided in Table 1.

Calculation and classification of SPI values are related to probability distribution functions. It was initially developed

Table 1 The classification of SPI

Drought category	SPI
Extremely wet	> 2
Very wet	1.5–1.99
Moderately wet	1–1.49
Near normal	− 0.99–0.99
Moderately dry	− 1 to − 1.49
Severely dry	− 1.5 to − 1.99
Extremely dry	− 2 and less

according to the gamma distribution, but later it was found to be suitable for the normal distribution. The scale, shape, and precipitation parameters of the gamma function are found for each station and time scale. Following the prediction of gamma function, the total likelihood function is calculated. Then, the SPI drought values are calculated by transforming the total likelihood distribution into the standard normal distribution. This probability function, which fits the normal distribution, is converted to a normal random SPI with a mean of zero and a variance of one. The normal random SPI value here indicates the Z-score, which can take a positive or negative value above or below the mean.

Stand-alone models

Artificial neural networks (ANNs)

ANNs imitate nerve cells, the smallest processing unit of the brain, the center of human nervous system, and enable computers to gain the ability of human learning. ANNs have the ability to learn and assist the concept of artificial intelligence. ANNs make generalization through self-training with input parameters presented to it and, following this generalization, generate an output parameter against the presented input parameters (Citakoglu 2017). ANNs are comparable to neurons of the human brain in that they are parallel computing systems. Activation (transfer) function, weights, and nodes are three aspects that define them. Linear, logistic, and tangent activation functions are common in engineering applications. To get an output, each neuron multiplies each input by its connectivity weight, adds the products, and then transfers the sum through a transfer function. The transfer function is commonly a sigmoid function, which is a differentiable, constantly growing S-shaped curve. The output function (yi) has a range of 0 to 1, but the inputs can have an infinite range of values. The y_i from the jth neuron in a layer is calculated using the threshold function:

$$y_i = f\left(\sum w_{ji}.x_i\right) = \frac{1}{1 + e^{-(\sum w_{ji}.x_i)}} \tag{2}$$

where x_i is the value of the ith neuron in the previous layer, w_{ji} is the weight of relationship connecting the jth neuron in a layer to the ith neuron in previous layer, and $f()$ is a transfer function, which is the rule for mapping the neuron's summed input to its output and can be used to introduce nonlinearity into the network design if chosen correctly (Haykin 1998).

Adaptive neuro-fuzzy inference system

ANFIS is a combined method of artificial neural networks and fuzzy logic. It was developed by Jang in the 1990s. ANFIS is based on Takagi–Sugeno-type fuzzy inference system, which is used to predict chaotic time series through nonlinear functions. By identifying the input structure, ANFIS exploits the learning capacity of ANNs to describe the input–output relationship and builds fuzzy rules. System results are acquired with the use of thinking and reasoning ability fuzzy logic. In this method, membership functions and fuzzy rules created as "if–then" are taken into account while finding output values against the input values presented to the adaptive network. The learning algorithm used by ANFIS to optimize the input and output values is a hybrid learning algorithm that combines the least-squares approach and the backpropagation learning algorithm (Jang 1993; Citakoglu 2015).

The Takagi–Sugeno-type if–then fuzzy rule-based ANFIS structure with x and y inputs and z outputs is mathematically shown as follows:

$$Rule\ 1 = If\ \left(x\ is\ A_1\right)\ and\ \left(y\ is\ B_1\right)\ Then\ \left(f_1 = p_1 x + q_1 y + r_1\right) \tag{3}$$

$$Rule\ 2 = If\ \left(x\ is\ A_2\right)\ and\ \left(y\ is\ B_2\right)\ Then\ \left(f_2 = p_2 x + q_2 y + r_2\right) \tag{4}$$

where x and y indicate the inputs, A_i and B_i symbolize the fuzzy sets, f_i denotes the outputs inside the fuzzy region denoted by the fuzzy rule, and p_i, q_i, and r_i denote the design parameters discovered during the training phase. Six layers of ANFIS model required to realize these two rules are briefly explained below:

Layer 1: this is the input layer, which just fixes the system input variable.

Layer 2: the is the fuzzification layer, used to define the membership grades of each input set using the fuzzy membership $\left(\mu_{A_i}(x)\right)$ function as follows:

$$\mu_{A_i}(x) = exp\left[-\frac{1}{2}\left(\frac{x_1 - c_i}{\sigma_i}\right)^2\right] \tag{5}$$

where the parameters $\left\{\sigma_i, c_i\right\}$ make up a premise parameter set, $\mu_{A_i}(x)$ is the fuzzy membership function expressed at the ith fuzzy set A_i, and x_1 is the ith the input.

Layer 3: this is the rule layer, output nodes are the rule's firing strength, which is stated as the product of membership grades in the following way:

$$w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2) \tag{6}$$

Layer 4: this layer (w_i) contains normalized firing strengths. The output of the ith node equals the ratio of the ith rule's firing strength to the sum of all rule's firing strengths, as follows:

$$\overline{w_i} = \frac{w_1}{w_1 + w_2 + w_3 + w_4} \tag{7}$$

where w_1 is the first rule, w_2 is the second rule, w_3 is the third rule, and w_4 is the fourth rule.

Layer 5: the ith rule's weighted output value is computed as follows in this layer:

$$N_i = \overline{w_i} f_i = \overline{w_i}(p_i x_1 + q_i x_2 + r_i) \tag{8}$$

where $\{p, q, r\}$ is a set of resultant parameters. They are determined by using least-squares method, x_1 and x_2 are inputs.

Layer 6: this is the summation layer. This layer calculates the overall output by summing all of the incoming signals as described as follows:

$$y = \sum_i \overline{w} f_i \tag{9}$$

Gaussian process regression (GPR)

GPR is an ML method that combines a set of random variables in which several variables have a multivariate Gaussian distribution. The GPR model, in which adjoining observations transmit information to each other, has been used frequently in prediction studies in recent years. The GPR is based on the probability theorem, which can make predictions on unknown input data and provide prediction precision, which greatly increases the significance of predictions (Sihag et al. 2018; Citakoglu 2021). Assuming certain x_i and y_i from a particular process and $y_i = f.(x_i)$, GPR's mean (x_i) and covariance function $k(x_i, x_j)$ are therefore entirely specified and distributed over the function $f(x_i)$ as follows:

$$f(x_i) \sim GP\big(\mu(x_i), k_\theta(x_i, x_j)\big). \tag{10}$$

A vector of parameter $\theta = [\theta1, \ldots, \theta z]$ is usually used in covariance functions. The simplest way to optimize these parameters of a data set $w = (x, y)$ is to maximize the log-marginal likelihood log log $p(y \mid x)$:

$$\theta = \arg\max\left(-\frac{1}{2}y^T(K_\theta + \sigma^2 I)^{-1}y - \frac{1}{2}\log\big(\big|K_\theta + \sigma^2 I\big| - \frac{N}{2}\log(2\pi)\right) \tag{11}$$

where $x = [x_1, \ldots, x_N]^T$ and $y = [y_1, \ldots, y_N]^T$ are vectors with N observed data with Gaussian noise of variance σ^2, $K_\theta = k_\theta(x, x)$ is a $N \times N$ covariance matrix of the training data set, and I is a $N \times N$ identity matrix.

GPR model discovers the predictive distribution of the related output y_* based on the data set w and a fresh input x_*. The Gaussian description of the predictive distribution of y_* over w is

$$P(y_* \mid (x, y), x_*) = N(\mu_*(x_*), \sigma_*^2(x_*)) \tag{12}$$

with mean $\mu_*(x_*)$ and covariance $\sigma.^2{}_*(x_*)$ given by Rasmussen and Williams (2006)

$$\mu_*(x_*) = k_*^T(K_\theta + \sigma^2 I)^{-1}y \tag{13}$$

$$\sigma_*^2(x_*) = k_{**} - k_*^T(K_\theta + \sigma^2 I)^{-1}k_* \tag{14}$$

where $k_* = [k(x_1, x_*), \ldots, k(x_N, x_*)]^T$ is an $N \times 1$ covariance vector between the test and training data sets and $k_{**} = k(x_*, x_*)$ is the autocovariance of test data set. The user must choose a suitable covariance function to obtain the predictive mean and covariance. Further details on GPR can be found in Rasmussen and Williams (2006).

Support vector machines

SVM, created by Cortes and Vapnik in 1995, is an ML technology based on structural risk minimization (Cortes and Vapnik 1995). SVM models are classified into two categories: (a) support vector regression models and (b) support vector machine classifier models. While the former one is used to address data classification and prediction issues, the later one is used to tackle prediction problems. The goal of regression is to find a hyperplane that fits the provided data. The inaccuracy of each location on this hyperplane is determined by its distance from any other point on the hyperplane. The least-squares approach is the best method for linear regression. However, when dealing with regression problems, using the least-squares estimator in the presence of outlier data may be impractical; as a result, the processor will perform poorly. Then, a robust estimator that is not sensitive to modest model modifications should be built. In fact, the following is how a penalty function (ε) is defined:

$$L^\varepsilon(x, y, f)\begin{cases} 0, & if |y - f(x)| \le \varepsilon \\ |y - f(x)| - \epsilon & otherwise \end{cases} \tag{15}$$

The training data sets are $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}\}$, and the class of the function is as $f(x) = \{w^T x + b, w \in R, b \in R\}$. If the data differ from the value of, a deficiency variable must have been created based on the deviation value. The minimization is specifically defined according to the penalty function:

$$\text{Minimize} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \text{ Subject to} \begin{cases} f(x) - y_i \leq \varepsilon + \xi i \\ f(x) - y_i \leq \varepsilon + \xi_i^* \quad i = 1, 2, 3.., n \\ \xi i, \xi_i^* \geq 0 \end{cases} \quad (16)$$

where $||w||^2$ is the norm of weight vector and ξ_i and ξ_i^* are auxiliary deficiency variables and C is the coefficient of complexity equilibrium between the machine and the number of indivisible points achieved by trial and error (Kisi and Cimen 2011).

K-nearest neighbors

KNN is a non-parametric ML method for classification and regression issues. The KNN algorithm is largely based on the data mining technique and is thus particularly successful in classification problems. This model is fed from a training set and uses this training set to categorize objects. It classifies an unknown sample as per the known classification of its neighbors (Fadaei-Kermani et al. 2017). In general, to evaluate any methods, there are three important aspects: ease of outputs interpretation, time of calculation, and power of prediction. The KNN classification output is an object, which is categorized by most of its neighbors. The k is a positive integer, which is typically small; if $k=1$, then that object is simply allocated to its nearest neighbor. Selecting the number of KNN has a large impact on the quality of the model. Prediction errors increase as the k number decreases. On the other hand, too many KNN can lead to modeling results with the so-called error of overfitting (Kutyłowska 2018).

Developing a method and relation to determine the distance between the training data and the testing data is the first step in using this model, and the distance between the training data and the testing data is commonly calculated using Euclidean distance:

$$D(x, p) = \sqrt{(x-p)^2} \quad (17)$$

where x is the new point and p is the learning example. After Euclidean distance is determined, the data are arranged in ascending order relative to the sample data, based on minimum distance and maximum distance. The next step is to find the number of neighbors (k). The efficiency of this method is significantly dependent on the quality of the selection of the closest sample on the reference data (Akbar Jalilzadnezamabad 2019).

Pre-processing techniques

Discrete wavelet transform (DWT)

Wavelet transform (WT) is a pre-processing technique used to determine the periodic and characteristic structure of data. WT is a mathematical operation that decomposes a signal into a set of fundamental functions, or wavelets, for a more concise description of the original data set without losing their features. WT is largely based on the Fourier transform that transports the series from the time domain to the frequency domain. Unlike the Fourier transform, the WT allows the user to calculate both the signal's low- and high-frequency components at each time interval (Khanmohammadi et al. 2022). Wavelets are the most important parameters WT, and Morlet, Haar, Coiflets, Mexican hat type, and Daubechies wavelets are the main wavelets commonly used in wavelet analysis. For a function to be a wavelet, the mean value should be zero, and its duration should be limited. Therefore, the main wavelet must satisfy these two conditions:

$$\int_{-\infty}^{+\infty} \psi(x)dx = 0 \quad (18)$$

$$\int_{-\infty}^{+\infty} \psi^2(x)dx = 1 \quad (19)$$

Whole scale range and too large stack of data increase the processing time. By overlapping wavelets, to reduce the intensity of information with certain scale groups of the conversion process, discrete wavelet transform (DWT) application has been developed. DWT, also known as Mallat algorithm, by simplified decomposition application, reduces transaction volume and at the same time provides the necessary steps. The most used scale step is the binary scale and the time step. A fixed scale value is set as a_0. The position of the wave on the time axis is given by b_0. When the values 2 and 1 are chosen in that order, the wavelet equation is

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a_0^m}}\Psi\left(\frac{x - a_0 b_0 n}{a_0^m}\right) = \frac{1}{\sqrt{2^m}}\Psi(2^m x - n) \quad (20)$$

In case of DWT, the time series size $N = 2\,M$ must be equal to two times. M is the number of steps of the $x(t)$ series. Therefore, for the current time series conversion where its dose not have the appropriate size, it can be expanded into an $x(t)$ series if necessary:

$$DWT(x) = \frac{1}{\sqrt{2^m}}\sum_{t=0}^{N-1} X(t)\Psi(2^m x - n) \quad (21)$$

The largest wavelet scale is obtained at 2^M, with a total conversion step of M is $1 < m < M$, and a single conversion coefficient is obtained by multiplying the whole series by the largest wavelet scale. In the following steps, the wavelet width for each shrinking m value is reduced by half, and the number that represents the series is doubled. Signal subject

to DWT without any loss can be divided into approximate (*A*) and detail (*D*) components. While the approximate signal (*A*) represents low frequency, the detail signal (*D*) represents high frequency. The approximate signal (*A*) is obtained as output in each step, and in the next step, it is considered as an input signal, and it is divided into approximate and detail components. This procedure is repeated until the required resolution is achieved. The following equations represent the sequential decomposition process with DWT:

$$x(t) = A1 + D1 \tag{22}$$

$$x(t) = A2 + D2 + D1 \tag{23}$$

$$x(t) = A_n + \sum_{i=1}^{n} D_i \tag{24}$$

In the final step, the original initial signal is first passed through a high-pass filter and then a low-pass filter. Outputs (*y*) are the approximate (from the low-pass filter) and detail signals (from the high-pass filter). Conversion coefficients for high-pass filter *h*(*k*) and low-pass filter *l*(*k*) are obtained by convolution of the signal with filters. Mathematically, this process is expressed as

$$y_{high} = \sum_{t=0}^{N-1} X(t)h(2k - t) \tag{25}$$

$$y_{low} = \sum_{t=0}^{N-1} X(t)l(2k - t) \tag{26}$$

Empirical mode decomposition

EMD is a signal decomposition approach described by Huang et al. (1998) that entirely empirically and data adaptively decomposes a time series signal into multiple oscillatory modes with particular periodicity. EMD decomposes any given data into intrinsic mode functions (IMF), which are simple oscillations as using shifting process (Huang et al. 1998). The following phases make up the sifting portion of the IMF algorithm:

i. Local maximum and lowest values of the data set are determined.
ii. The upper ($x_{upper}(n)$) and lower ($x_{lower}(n)$) envelopes are created by combining the local maximum and minimum values.
iii. Average of the upper and lower envelopes is calculated as follows:

$$a(n) = \frac{x_{upper}(n) + x_{lower}(n)}{2} \tag{27}$$

iv. The original data set is deducted from the estimated mean value:

$$h(n) = x(n) - a(n) \tag{28}$$

v. If *h*(*n*) meets the criteria for becoming an IMF, the IMF will be $\varphi(n) = h(n)$. If not, go back to the first step and continue the process until the prerequisites are met.

This method is repeated until a specific number of IMF has been obtained. The residual is the data set that remains after the decomposition process is completed. As a result, the data set is divided into two components: IMFs and residuals (Kisi et al. 2014; Özger et al. 2020; Latifoğlu 2022).

Variational mode decomposition (VMD)

VMD is a non-recursive and adaptive time–frequency analysis method created by Dragomiretskiy and Zosso (2014). VMD is a multiple adaptive band generalization of the original Wiener filter. By decomposing a one-dimensional input signal into a defined number of modes, the VMD approach addresses mode decomposition as an optimization issue. By adding up the *K* number of decomposition modes, the signal is entirely replicated:

$$H(t) = \sum_{k=1}^{K} u_k(t) \tag{29}$$

where *K* is the total number of modes; *k* is the index of modes; $u_k(t)$ is the kth mode, and it is an amplitude-modulated-frequency-modulated signal with the following formula:

$$u_k(t) = A_k(t) \cos\left(\varphi_k(t)\right) \tag{30}$$

where $A_k(t)$ and $\varphi_k(t)$ are the time-dependent envelope and the phase of the *k*th mode, respectively. The corresponding instantaneous frequency (*t*) of the *k*th mode is non-negative and needs to change gradually according to the phase. It could be calculated using the following formula:

$$\omega_k(t) = \frac{\partial \varphi_k(t)}{\partial_t} \tag{31}$$

Decomposition of time series can be stated as a constrained variational problem with the following objective function:

$$\min_{\{u_k\}\{\varpi_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta_t + \frac{j}{\pi t}\right) * u_k(t) \right] e^{-j\varpi_k t} \right\|_2^2 \right\}, s.t. \sum_{k=1}^{K} u_t = H(t) \tag{32}$$

where $\{uk\} = \{u_1, \ldots, u_K\}$ and $\{\varpi k\} = \{\varpi_1, \ldots, \varpi_K\}$ are the sets of all modes and their related center frequencies; $\delta(t)$ is the Dirac function; * represents the convolution; and $j = \sqrt{-1}$. Detailed theoretical information about VMD can be found in Dragomiretskiy and Zosso (2014).

Comparison of model performances

Six different criteria were used to compare model performances in the prediction of SPI values: mean square error (*MSE*), root mean square error (*RMSE*), mean absolute error (*MAE*), Nash–Sutcliffe efficiency (*NSE*), overall index of model performance (*OI*), and determination coefficient (R^2). *MSE*, *RMSE*, and *MAE* values close to 0 and R^2 value close to 1 indicate that the predicted value converged strongly to the original data. The *NSE* takes values between $-\infty$ and 1. *NSE* values of < 1 are ideal, as this indicates a 100% success rate. Low prediction success is indicated by *NSE* values of between 0.3 and 0.5, acceptable prediction success is indicated by *NSE* values of between 0.5 and 0.7, great prediction success is indicated by *NSE* values of between 0.7 and 0.9, and outstanding prediction success is indicated by *NSE* values of between 0.9 and 1 (Nash and Sutcliffe 1970). The normalized root mean square error and model efficiency indicators are combined in the *OI* criterion. *OI* can take the following values: $+1$, with 1 indicating a perfect model that predicts the same values as the measured ones (Citakoglu 2015). All these performance measures are calculated with the following equations:

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^{N} \left(\text{SPI}_\text{p} - \text{SPI}_\text{c} \right)^2 \tag{33}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left(\text{SPI}_\text{p} - \text{SPI}_\text{c} \right)^2} \tag{34}$$

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^{N} \left| \text{SPI}_\text{p} - \text{SPI}_\text{c} \right| \tag{35}$$

$$\text{NSE} = 1 - \left[\frac{\sum_{n=1}^{N} \left(\text{SPI}_\text{c} - \text{SPI}_\text{p} \right)^2}{\sum_{n=1}^{N} \left(\text{SPI}_\text{c} - \overline{\text{SPI}_\text{c}} \right)^2} \right] \tag{36}$$

$$OI = \frac{1}{2} \left(2 - \frac{RMSE}{SPI_{cmax} - SPI_{cmin}} - \frac{\sum_{n=1}^{N} (SPI_c - SPI_p)^2}{\sum_{n=1}^{N} (SPI_c - \overline{SPI_c})^2} \right) \tag{37}$$

$$\text{R}^2 = 1 - \left[\frac{\sum_{n=1}^{N} \left(\text{SPI}_\text{c} - \text{SPI}_\text{p} \right)^2}{\sum_{n=1}^{N} (SPI_p - \overline{\text{SPI}_\text{p}})^2} \right] \tag{38}$$

where N is the number of data in the available series, SPI_p is the prediction value obtained from the models, and SPI_c is the drought value calculated from monthly precipitation data. $\overline{\text{SPI}_\text{p}}$ and $\overline{\text{SPI}_\text{c}}$ are the averages of predicted and calculated values; SPI_{cmax} and SPI_{cmin} are the minimum and maximum values of the calculated series.

Model development

Models were developed in this study to predict short-term meteorological droughts of the Sakarya Meteorological Station, located in the northwest of Turkey, with the use of monthly precipitation data of $1960 - 2020$ period and hybrid ML methods. Initially, drought index (SPI) values were calculated from precipitation data. SPI drought values were calculated with the use of Drought Indices Calculator (DrinC) software (Tigkas et al. 2015). In present prediction models, different lag times (t, $t-1$, $t-2$, $t-3$… etc.) of SPI drought values for 1-, 3-, and 6-month time scales

Table 2 Preliminary test results

	Test statistics values	H0 hypothesis	
Standard normal homogeneity test	$t = 10.500$	Rejection	Homogeneity
Mann–Whitney test	$q = 0.158$	Rejection	Homogeneity
Phillips–Perron test	$q = -10.657$	Rejection	Stationarity
Augmented Dickey–Fuller test	$q = -9.268$	Rejection	Stationarity
Von Neumann's test	$Q = 1.797$	Acceptance	Independence

**Table 3** Statistical parameters of SPI drought time series

|	Mean value	Minimum value	Maximum value	Standard deviation	Coefficient of skewness	Coefficient of kurtosis
1-month SPI	0.00	-3.92	2.83	0.99	-0.42	0.41
3-month SPI	0.00	-3.17	3.14	1.00	0.07	0.35
6-month SPI	0.00	-3.70	3.69	1.01	-0.16	2.23

were considered as input variables and $t+1$ lag time as output variable. Model success is designated by the relationships between the lag times and optimum number of input variables. Autocorrelation was applied to input variables to reveal relationship between lag times and input variables. Autocorrelation conceptually refers to the cooperation relationship between the value of a series in any period and the value of the previous or next period (Başakın et al. 2022). To determine whether there is autocorrelation in SPI drought time series, the codes of the graphic method in MATLAB 2021b software were used. After examining the presence of autocorrelation, the original drought data were divided into sub-series with DWT to determine the optimum input variable, and these sub-series were used as input data as training and test data in the ANFIS prediction model. DWT can comprehensively decompose data on a time scale, has a wide variety of wavelet families and band levels, and is more commonly used in the literature as compared to EMD and VMD. Many different wavelet families and their versions are used in the process of decomposing time series into sub-series with wavelet transform. It is difficult to make a definite opinion regarding the decomposition performance of wavelet families. Compared to most studies in the literature, in this study, analyses were made by diversifying both the wavelet family, the band level, and the input variables. The aim here is to evaluate the optimum input variable more comprehensively and accordingly to increase the performance of the prediction models. The reason why the ANFIS method is considered together with the DWT is that the ANFIS method makes fast analyses. In this study, Haar, symlets (sym3), coiflets (coif2), biorthogonal (bior1.3), reverse biorthogonal (rbio1.3), discrete approximation of Meyer (dmey), Fejer-Korovkin (fk4), and Daubechies (db40) wavelet families were taken into consideration. In addition, 3, 4, 5, 6, and 7 wavelet band levels and five lag times (t, $t-1$, $t-2$, $t-3$, $t-4$, and $t-5$ input variables) were used. The decomposition process with DWT was completed using the codes in MATLAB 2021b software. Of the input data divided into sub-series, 75% (1960−2005) was used as training data and 25% (2006−2020) as test data. In the ANFIS prediction model, Gaussian (gaussmf), and triangular (triangular-trimf) functions were used as membership functions of the network, and constant and linear functions were used as output functions. Predictions were obtained with iterations of between 1 and 10 by changing the number of two−three memberships in four different combinations. The input variable that gave the best results in the DWT-ANFIS hybrid prediction model was also used in all hybrid versions. In addition, the wavelet family and wavelet band levels to be used in the hybrid versions with DWT were also determined at this stage.
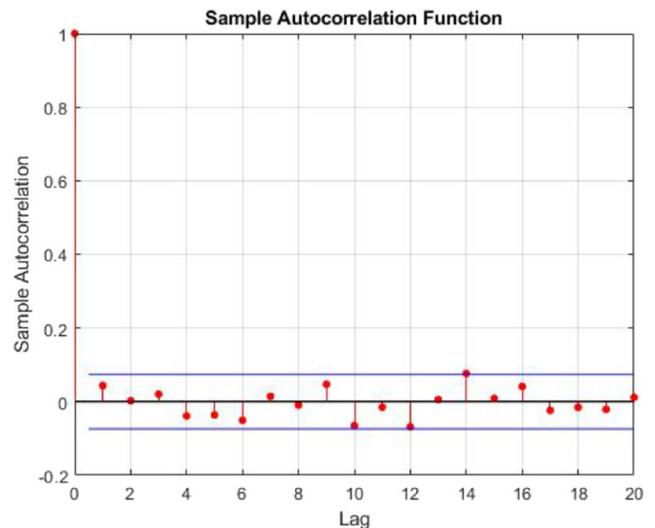
a) 1-month
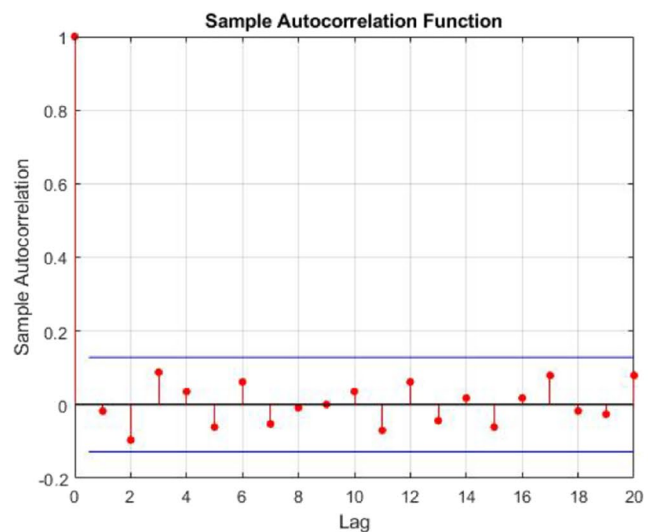
b) 3-month
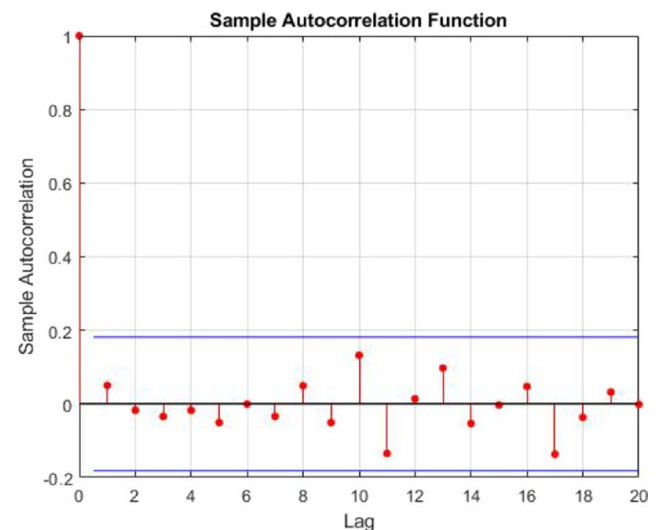
c) 6-month

Table 4 Comparison of DWT-ANFIS hybrid model results for 1-month time scale

Wavelet family	Wavelet band levels	Train R^2	Train RMSE	Test R^2	Test RMSE	Rank
Haar	3	0.2987	0.9678	0.5102	0.6891	8
	4	0.3001	0.9944	0.5090	0.6925	
	5	0.3028	1.0107	0.5098	0.6914	
	6	0.2998	1.0244	0.5066	0.6968	
	7	0.3009	1.0345	0.5067	0.6948	
Symlets (sym3)	3	0.7763	0.4684	0.7487	0.4947	5
	4	0.7768	0.4680	0.7600	0.4833	
	5	0.7768	0.4679	0.7663	0.4770	
	6	0.7769	0.4678	0.7640	0.4795	
	7	0.7769	0.4678	0.7614	0.4819	
Coiflets (coif2)	3	0.8202	0.4200	0.7468	0.5008	4
	4	0.8213	0.4188	0.7753	0.4759	
	5	0.8215	0.4186	0.7789	0.4691	
	6	0.8215	0.4187	0.7781	0.4709	
	7	0.8214	0.4187	0.7779	0.4667	
Biorthogonal (bior1.3)	3	0.7081	0.5368	0.6944	0.5527	7
	4	0.7085	0.5362	0.6914	0.5553	
	5	0.7090	0.5358	0.6889	0.5617	
	6	0.7089	0.5359	0.6914	0.5592	
	7	0.7088	0.5360	0.6899	0.5617	
Reverse biorthogonal (rbio1.3)	3	0.3530	0.9986	0.6946	0.5462	6
	4	0.3589	1.0303	0.6934	0.5475	
	5	0.3664	1.0276	0.6954	0.5451	
	6	0.3662	1.0280	0.6914	0.5497	
	7	0.3651	1.0233	0.6913	0.5500	
Discrete approximation of Meyer (dmey)	**3**	0.5799	0.9927	0.8537	0.3796	1
	4	0.5917	1.0261	0.8534	0.3798	
	5	0.5817	1.0533	0.8533	0.3799	
	6	0.5832	1.0711	0.8533	0.3799	
	7	0.5881	1.0790	0.8531	0.3800	
Fejer-Korovkin (fk4)	3	0.5654	0.9940	0.8368	0.4013	2
	4	0.5789	1.0282	0.8411	0.3965	
	5	0.5761	1.0526	0.8420	0.3959	
	6	0.5786	1.0676	0.8418	0.3959	
	7	0.5744	1.0770	0.8417	0.3965	
Daubechies (db40)	3	0.5715	0.9870	0.8301	0.4067	3
	4	0.5767	1.0020	0.8310	0.4050	
	5	0.5744	0.9912	0.8309	0.4050	
	6	0.5701	0.9846	0.8313	0.4046	
	7	0.5702	0.9846	0.8310	0.4051	

Following the determination of optimum input variable with the DWT-ANFIS model, the original SPI drought time series were pre-processed using EMD and VMD techniques. The original drought series were divided into four sub-series (2D, 3D, 4D, 5D) in both methods using the codes in MAT-LAB 2021b software. The internal mode function (IMF) and residual components specific to these techniques were obtained in each sub-series. These components were used as input data in the ANFIS prediction model as in DWT in order to determine the optimum EMD and VMD band levels.

Sub-series obtained with DWT, EMD, and VMD were included as input data to five different ML methods (ANFIS, ANN, SVMR, KNN, GPR) as training and test data. In addition, stand-alone prediction models were created using these

Table 5 Comparison of EMD-ANFIS hybrid model results

Model	Time scale	Inputs	Output	EMD band levels	Train		Test	
					R^2	RMSE	R^2	RMSE
EMD-ANFIS	1-month	$t, t-1, t-2, t-3$	$t+1$	2	0.940	0.447	0.882	0.582
				3	0.939	0.486	0.880	0.604
				4	0.939	0.507	0.880	0.605
				5	0.939	0.523	0.880	0.605
	3-month	$t, t-1, t-2, t-3$	$t+1$	2	0.945	0.411	0.896	0.709
				3	0.943	0.468	0.899	0.749
				4	0.942	0.501	0.900	0.740
				5	0.942	0.559	0.900	0.743
	6-month	$t, t-1, t-2, t-3$	$t+1$	2	0.935	0.406	0.942	0.703
				3	0.936	0.442	0.940	0.662
				4	0.938	0.484	0.941	0.657
				5	0.937	0.500	0.940	0.661

Table 6 Comparison of VMD-ANFIS hybrid model results

Model	Time scale	Inputs	Output	VMD band levels	Train		Test	
					R^2	RMSE	R^2	RMSE
VMD-ANFIS	1-month	$t, t-1, t-2, t-3$	$t+1$	2	0.379	0.967	0.607	0.644
				3	0.464	1.012	0.722	0.531
				4	0.627	0.982	0.857	0.377
				5	0.703	1.029	0.936	0.255
	3-month	$t, t-1, t-2, t-3$	$t+1$	2	0.572	0.907	0.442	0.964
				3	0.509	0.963	0.510	0.830
				4	0.628	0.967	0.826	0.469
				5	0.673	0.955	0.932	0.295
	6-month	$t, t-1, t-2, t-3$	$t+1$	2	0.462	0.878	0.346	1.070
				3	0.558	0.851	0.489	1.224
				4	0.661	0.904	0.846	0.514
				5	0.721	0.971	0.947	1.197

ML methods without pre-processing the original drought data. A total of twenty prediction models, fifteen of which are hybrid prediction models and five of which are stand-alone prediction models, were created, and the results of all these models were compared with each other.

In this study, the prediction model of the multi-layered ANN method was used. Levenberg–Marquardt method, which has been widely used in recent years and based on the standard numerical optimization technique, was used as a learning algorithm. Tansig and logsig were used as input functions, and tansig, logsig, and purelin functions were used as output functions. Predictions were obtained with THE number of one-ten neurons and with iterations varying between one-two hundred in six different combinations.

In SVMR method, a prediction model was developed by using kernel functions. The kernel functions used included Gaussian, polynomial, rbf, and linear. Gamma (γ), epsilon (ε), and penalty (c) parameters were obtained by trial and error. Predictions were obtained with true standardization and iterations varying between one-fifty.

In GPR method, a prediction model was created by using both kernel functions and radial basis function (RBF) to increase model performance. As kernel functions, ardmatern32, ardexponential, ardrationalquadratic, ardmatern52, ardsquaredexponential, matern32, exponential, matern52, squaredexponential, and rationalquadratic functions were used. Constant, nonlinear, and pure-quadratic functions were used as RBF functions. In GPR model, beta (β) and sigma (σ) parameters were calculated using the fully

Table 7 Performance comparison of models for 1-month time scale

Models	Train						Test					
	MSE	RMSE	NSE	OI	MAE	R^2	MSE	RMSE	NSE	OI	MAE	R^2
ANFIS	0.889	0.943	0.093	0.477	0.729	0.122	1.566	1.252	-0.628	0.069	0.921	0.002
DWT-ANFIS	0.986	0.993	− 0.005	0.604	0.764	0.580	0.145	0.380	0.850	0.889	0.298	0.854
EMD-ANFIS	0.200	0.447	0.796	0.830	0.317	0.940	0.325	0.571	0.662	0.778	0.456	0.881
VMD-ANFIS	1.058	1.029	− 0.080	0.709	0.830	0.703	0.064	0.254	0.933	0.943	0.204	0.935
ANN	0.926	0.962	0.055	0.456	0.755	0.055	1.220	1.105	-0.268	0.263	0.876	0.164
DWT-ANN	0.115	0.339	0.880	0.920	0.267	0.883	0.148	0.385	0.846	0.887	0.301	0.851
EMD-ANN	0.359	0.599	0.634	0.769	0.463	0.881	0.341	0.584	0.646	0.768	0.471	0.880
VMD-ANN	0.849	0.922	0.133	0.488	0.724	0.980	0.819	0.905	0.149	0.490	0.735	0.983
GPR	0.935	0.967	0.046	0.451	0.763	0.094	1.013	1.007	-0.053	0.380	0.809	0.054
DWT-GPR	0.135	0.367	0.863	0.892	0.291	0.863	0.143	0.378	0.852	0.891	0.297	0.861
EMD-GPR	0.360	0.600	0.633	0.765	0.461	0.881	0.348	0.590	0.638	0.764	0.474	0.879
VMD-GPR	0.065	0.255	0.933	0.941	0.201	0.934	0.063	0.251	0.935	0.944	0.205	0.937
SVMR	0.964	0.982	0.016	0.435	0.757	0.021	1.247	1.117	-0.296	0.248	0.870	0.293
DWT-SVMR	0.132	0.364	0.865	0.940	0.283	0.869	0.217	0.466	0.774	0.844	0.347	0.780
EMD-SVMR	0.372	0.610	0.620	0.769	0.470	0.878	0.344	0.586	0.643	0.767	0.471	0.875
VMD-SVMR	0.065	0.255	0.934	0.946	0.199	0.934	0.079	0.282	0.918	0.933	0.227	0.921
KNN	1.472	1.213	− 0.502	0.159	0.908	0.003	1.966	1.402	-1.043	-0.153	1.070	0.035
DWT-KNN	0.741	0.861	0.245	0.516	0.753	0.823	0.771	0.878	0.199	0.518	0.789	0.825
EMD-KNN	0.414	0.643	0.578	0.777	0.478	0.879	0.309	0.556	0.679	0.788	0.434	0.891
VMD-KNN	0.573	0.757	0.415	0.574	0.689	0.898	0.690	0.830	0.283	0.564	0.764	0.883

Table 8 Performance comparison of models for 3-month time scale

Models	Train						Test					
	MSE	RMSE	NSE	OI	MAE	R^2	MSE	RMSE	NSE	OI	MAE	R^2
ANFIS	0.834	0.913	0.110	0.479	0.728	0.153	1.999	1.414	-0.621	0.069	1.034	0.080
DWT-ANFIS	0.848	0.921	0.090	0.660	0.721	0.640	0.158	0.398	0.872	0.902	0.315	0.875
EMD-ANFIS	0.169	0.411	0.820	0.777	0.291	0.945	0.502	0.708	0.593	0.736	0.528	0.896
VMD-ANFIS	0.912	0.955	0.026	0.678	0.766	0.673	0.087	0.294	0.930	0.940	0.224	0.932
ANN	0.740	0.860	0.211	0.534	0.658	0.211	1.134	1.065	0.081	0.450	0.893	0.136
DWT-ANN	0.128	0.357	0.863	0.917	0.282	0.895	0.153	0.391	0.876	0.905	0.315	0.884
EMD-ANN	0.259	0.509	0.723	0.763	0.395	0.890	0.400	0.633	0.676	0.784	0.505	0.914
VMD-ANN	0.127	0.356	0.865	0.903	0.297	0.941	0.118	0.344	0.904	0.923	0.277	0.961
GPR	0.921	0.960	0.018	0.429	0.758	0.018	1.325	1.151	-0.074	0.365	0.916	0.034
DWT-GPR	0.135	0.367	0.863	0.890	0.291	0.863	0.155	0.394	0.874	0.904	0.322	0.876
EMD-GPR	0.262	0.512	0.720	0.764	0.395	0.889	0.404	0.636	0.672	0.782	0.506	0.911
VMD-GPR	0.062	0.249	0.934	0.941	0.193	0.942	0.058	0.241	0.953	0.956	0.196	0.957
SVMR	0.671	0.819	0.284	0.574	0.579	0.301	1.239	1.113	-0.005	0.403	0.865	0.020
DWT-SVMR	0.110	0.332	0.882	0.981	0.255	0.883	0.337	0.580	0.727	0.814	0.459	0.733
EMD-SVMR	0.268	0.518	0.714	0.827	0.394	0.887	0.221	0.470	0.821	0.871	0.349	0.934
VMD-SVMR	0.052	0.229	0.944	0.970	0.180	0.954	0.179	0.423	0.855	0.891	0.327	0.876
KNN	3.635	1.907	-2.894	-1.106	1.313	0.131	3.773	1.942	-2.059	-0.695	1.415	0.063
DWT-KNN	1.408	1.187	-0.512	0.146	1.083	0.757	1.868	1.367	-0.514	0.127	1.217	0.657
EMD-KNN	0.331	0.575	0.647	0.774	0.452	0.901	0.351	0.593	0.715	0.807	0.491	0.940
VMD-KNN	0.797	0.893	0.149	0.476	0.831	0.900	1.189	1.090	0.036	0.426	0.995	0.805

Table 9 Performance comparison of models for 6-month time scale

Models	Train						Test					
	MSE	RMSE	NSE	OI	MAE	R^2	MSE	RMSE	NSE	OI	MAE	R^2
ANFIS	0.701	0.837	0.082	0.462	0.643	0.121	3.235	1.799	− 0.920	− 0.082	1.348	0.016
DWT-ANFIS	0.761	0.872	0.000	0.643	0.697	0.637	0.109	0.330	0.935	0.945	0.276	0.937
EMD-ANFIS	0.165	0.406	0.785	0.800	0.270	0.935	0.493	0.702	0.707	0.806	0.535	0.942
VMD-ANFIS	0.942	0.971	− 0.235	0.376	0.819	0.721	1.432	1.197	0.150	0.494	0.936	0.947
ANN	0.557	0.747	0.257	0.558	0.555	0.257	2.240	1.497	− 0.329	0.234	1.119	0.113
DWT-ANN	0.107	0.327	0.860	0.867	0.253	0.864	0.102	0.319	0.940	0.948	0.256	0.945
EMD-ANN	0.238	0.488	0.688	0.738	0.376	0.884	0.565	0.752	0.665	0.781	0.571	0.936
VMD-ANN	0.098	0.314	0.871	0.906	0.244	0.874	0.238	0.487	0.859	0.897	0.387	0.865
GPR	0.691	0.831	0.078	0.461	0.655	0.200	1.665	1.290	0.012	0.419	0.940	0.020
DWT-GPR	0.085	0.291	0.888	0.899	0.233	0.890	0.114	0.338	0.932	0.943	0.291	0.934
EMD-GPR	0.243	0.493	0.681	0.725	0.381	0.889	0.641	0.801	0.620	0.756	0.597	0.926
VMD-GPR	0.069	0.262	0.910	0.922	0.209	0.910	0.158	0.398	0.906	0.926	0.324	0.914
SVMR	0.449	0.670	0.401	0.638	0.415	0.420	1.636	1.279	0.030	0.428	0.923	0.035
DWT-SVMR	0.099	0.315	0.870	0.899	0.250	0.870	0.374	0.611	0.778	0.848	0.482	0.821
EMD-SVMR	0.228	0.478	0.701	0.848	0.359	0.890	0.153	0.391	0.909	0.928	0.296	0.974
VMD-SVMR	0.053	0.230	0.931	0.943	0.189	0.933	0.256	0.506	0.848	0.890	0.400	0.881
KNN	1.909	1.382	− 1.476	− 0.368	1.079	0.040	2.934	1.713	− 0.741	0.014	1.387	0.041
DWT-KNN	0.774	0.880	− 0.018	0.594	0.781	0.785	1.212	1.101	0.281	0.566	0.932	0.748
EMD-KNN	0.303	0.550	0.604	0.821	0.450	0.913	0.342	0.585	0.797	0.859	0.515	0.954
VMD-KNN	1.108	1.053	− 0.452	0.473	0.979	0.804	1.563	1.250	0.072	0.452	1.120	0.769

independent conditional approximation method. Predictions were obtained with true standardization and iterations varying between one–one hundred.

In KNN method, the *K*-nearest neighbor's algorithm was used as the learning algorithm. The prediction model was developed using *K*-d tree functions and exhaustive functions. Euclidean, cityblock, Minkowski, and Chebyshev functions were used as *K*-d tree functions, and Spearman, Jaccard, Mahalanobis, correlation, Hamming, cosine, and seuclidean functions were used as exhaustive functions. The *k* coefficient, which shows the neighborhood relationship, was determined by trial and error, and predictions were obtained with correct standardizations.

The accuracy of the prediction models used in the study was evaluated according to *MSE*, *RMSE*, *MAE*, *NSE*, *OI*, and R^2 performance criteria. Besides these criteria, Taylor diagrams, in which the statistical values of standard deviation, correlation, and centered root mean square difference (RMSD) can be evaluated simultaneously, were used, and the model results were compared. Radar charts showing all performance criteria together were used to more easily compare the errors of five hybrid prediction models that yielded the best results. Scatter plots of the stand-alone models and the hybrid methods that yielded the best results were also created. In addition, boxplot charts that provide statistical comparison of the errors of the models were also used.

Results and discussion

Before being use in drought analysis, present raw data supplied by the Turkish State Meteorological Service (MGM) were subjected to homogeneity, independence, autocorrelation, and stationary tests. Standard normal homogeneity test, Mann–Whitney test, autocorrelation test, Phillips–Perron test stationary test, augmented Dickey–Fuller test, and Von Neumann's independence tests were used to check the homogeneity, stationarity, and independence of the precipitation data. The details of all the tests used in the preliminary phase can be accessed from Haktanir and Citakoglu (2014, 2015), Yagbasan et al. (2020), Citakoglu and Minarecioglu (2021), and Demir (2022a). Test results revealed that precipitation data of Sakarya station was homogeneous, independent, and stationary (Table 2).

SPI drought values for 1-, 3-, and 6-month time scales were calculated separately in the DrinC software using the monthly precipitation values of the station between the years 1960 and 2020. Statistical parameters of the calculated SPI drought time series are given in Table 3.

The means for 1-, 3-, and 6-month SPI drought values were found to be zero, and the standard deviations were around one. This is because the SPI conformed to standard normal distribution with a mean of zero and a variance of one. The kurtosis coefficient of the SPI drought values on

Fig. 4 Scatter plots of calculated and predicted SPI values for 1-month time scale
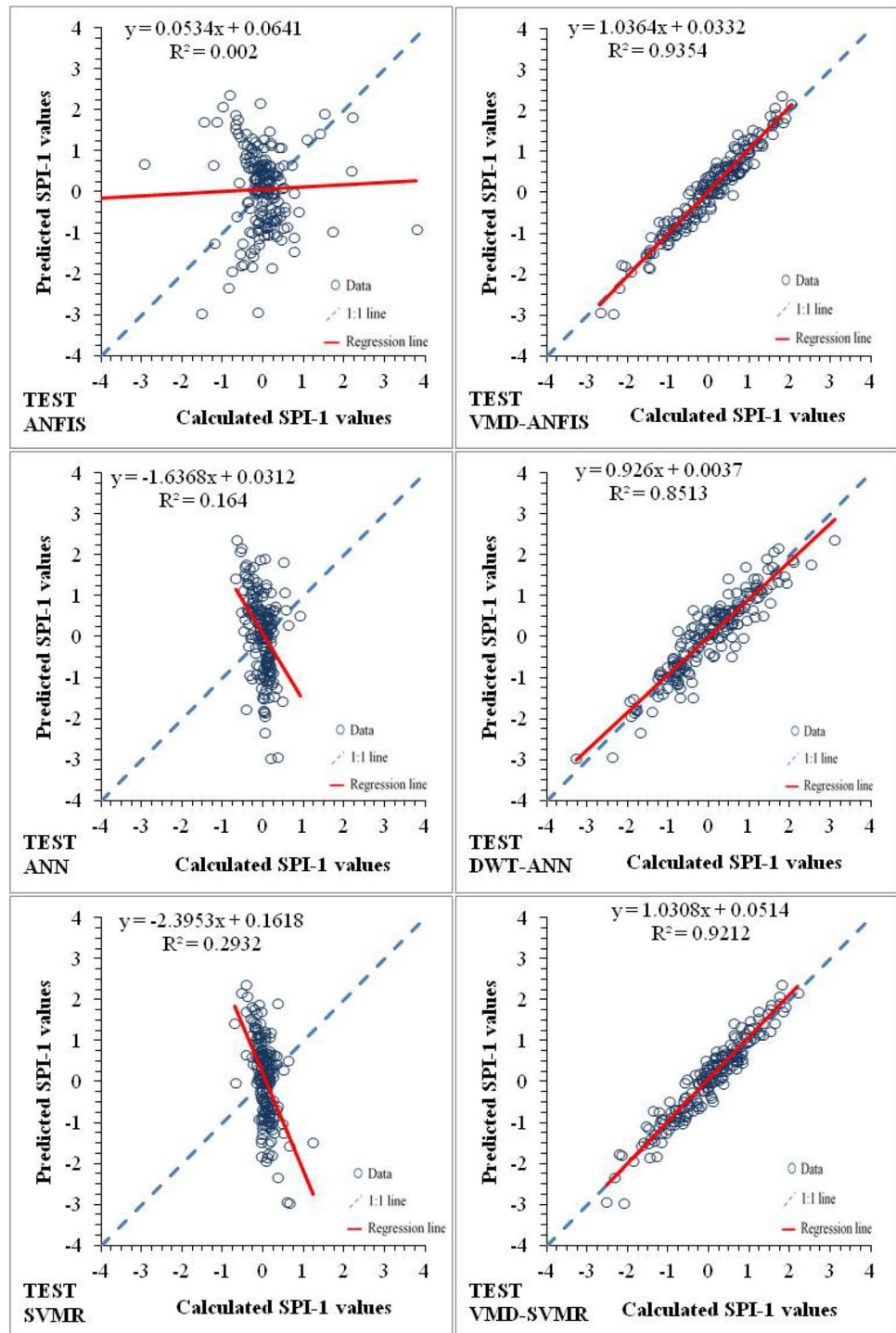

a 6-month time scale was found to be 2.23. According to this parameter, some deviation from the normal distribution was observed only in this time period. The lowest SPI drought value was calculated as −3.92 for 1-month time scale, and the highest SPI drought value was 3.69 for 6-month time scale.

Autocorrelation functions were calculated for 20 lag times, and their plots were created. Autocorrelation plots of SPI drought time series for 1-, 3-, and 6-month time scales are presented in Fig. 3.

There was no autocorrelation in the series until the 20th lag time for each time scale. All the lag times considered were within the confidence interval limits. The 10th, 12th, and 14th lag times in the SPI autocorrelation plot of the only 1-month time scale came close to the confidence interval limits.

Fig. 4 (continued)
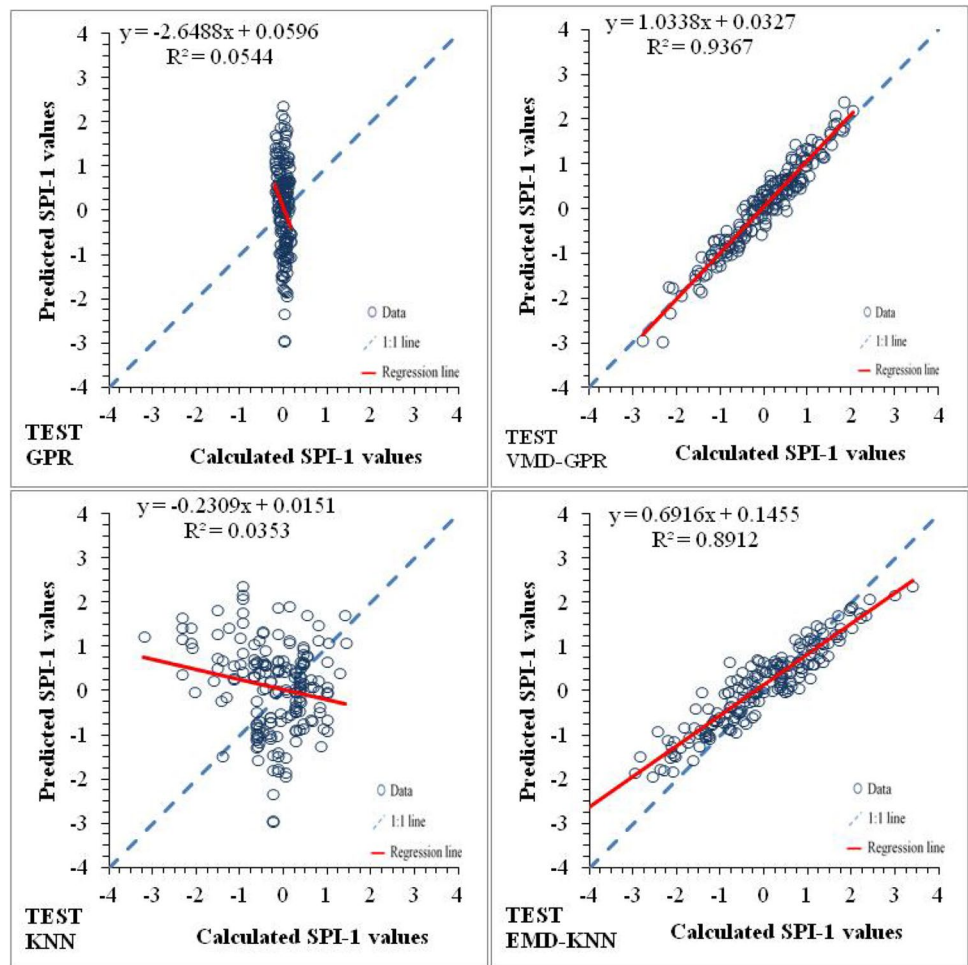
Autocorrelation plots revealed that the times up to the 20th lag time in all time scales could be used as input variables in prediction models. Comparison of the results with the use of R^2 and *RMSE* performance criteria for 1-month time scale in DWT-ANFIS hybrid model are provided in Table 4.

As can be seen in Table 4, according to R^2 and *RMSE* for 1-month DWT-ANFIS hybrid model, dmey wavelet family yielded the best prediction results. The results of fk4 and db40 wavelet families were quite close to the results of dmey. When the results of the dmey wavelet family were evaluated within themselves, the 3rd wavelet band level and the 3rd lag time (t, $t-1$, $t-2$, $t-3$ input variables) yielded the best results. These analyses were performed for all 1-, 3-, and 6-month time scales. In all time scales, dmey wavelet family and the 3rd lag time (t, $t-1$, $t-2$, $t-3$ input variables) yielded the best results. In addition, the 3rd wavelet band level for 1- and 3-month time scales and the 4th wavelet band level for 6-month time scale yielded the best prediction results. According to the results obtained from this model, in all hybrid models created with DWT, dmey was used as the wavelet family; the 3rd band level for 1- and 3-month time

scales and the 4th band level for 6-month time scale were used as the wavelet band levels. In addition, t, $t-1$, $t-2$, and $t-3$ were taken as the optimum lag times (input variables) in all hybrid models considered in this study.

Following the identification of the best wavelet family, wavelet band level and optimum input variable with the DWT-ANFIS hybrid model, EMD-ANFIS, and VMD-ANFIS hybrid prediction models were created. Comparison of the results obtained with these models according to R^2 and *RMSE* performance criteria is given in Table 5 and Table 6.

As can be seen in Tables 5 and 6, based on the test performances of the models, the 2nd band level yielded the best decomposition performance in all time scales in the EMD-ANFIS hybrid model; the 5th band level for 1- and 3-month time scales and the 4th band level for 6-month time scale yielded the best decomposition performance in the VMD-ANFIS hybrid model. Therefore, IMF and residual components decomposed to the 2nd band level in all EMD hybrid prediction models, and the 4th and 5th band levels in the VMD hybrid prediction models were used in this study. These components were used as input data and training and test data in each ML method.

Fig. 5 Scatter plots of calculated and predicted SPI values for 3-month time scale
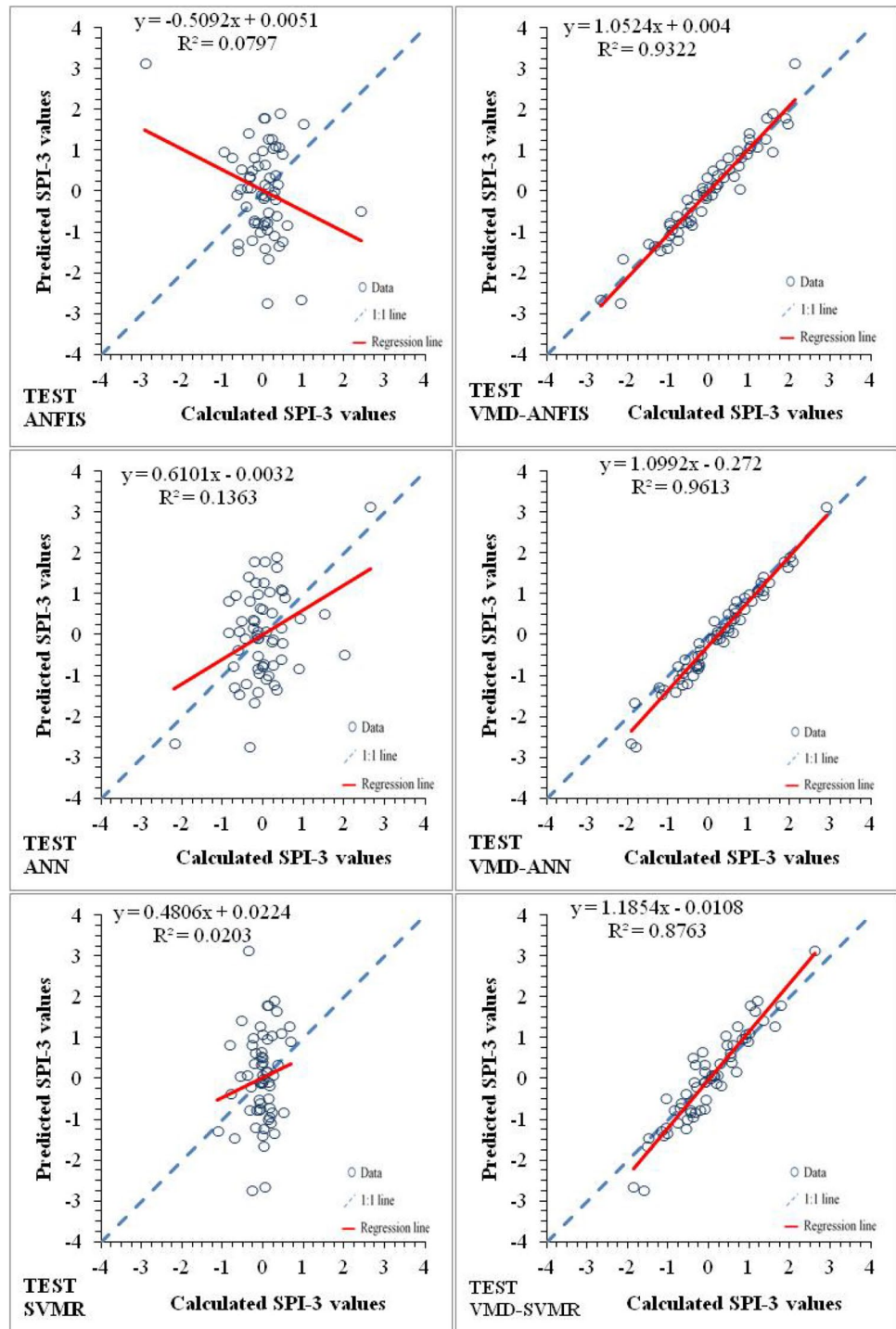
Following the completion of pre-processing techniques, prediction processes were completed by using all the stand-alone and hybrid prediction models. Comparison of the prediction results of the training and test data obtained from the models according to *MSE*, *RMSE*, *MAE*, *NSE*, *OI*, and R^2 performance criteria are given in Tables 7, 8, and 9.

As can be seen in Tables 7, 8, and 9, the prediction performances of the test data were quite low in all stand-alone models that were not subjected to pre-processing techniques. The prediction success of hybrid models created by incorporating DWT, EMD, and VMD pre-processing techniques into these models has increased significantly. When the methods were evaluated within

Fig. 5 (continued)
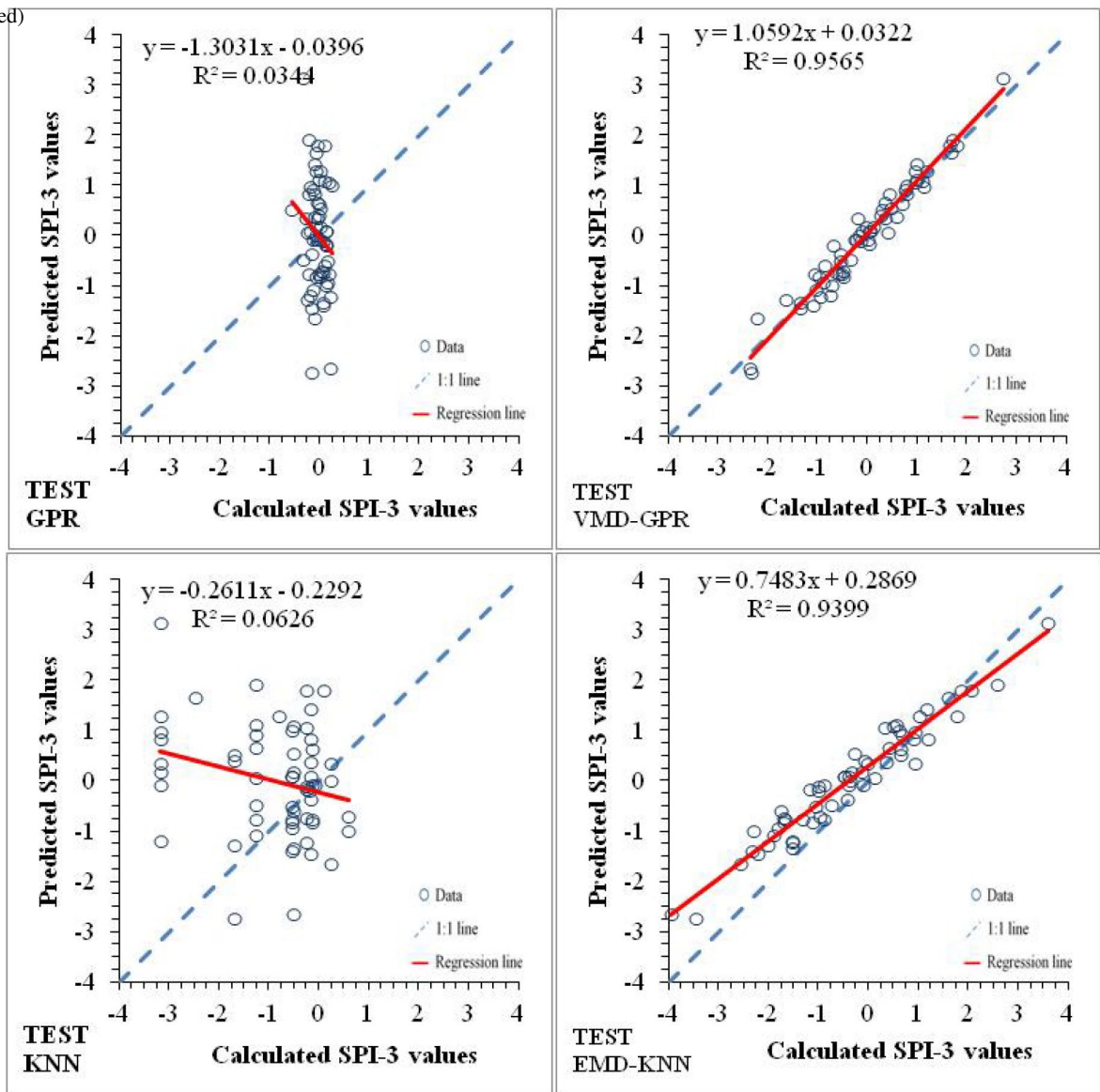
themselves, for 1-month time scale, VMD-ANFIS hybrid model in ANFIS models, DWT-ANN hybrid model in ANN models, VMD-GPR hybrid model in GPR models, VMD-SVMR hybrid model in SVMR models, and EMD-KNN hybrid model in KNN models yielded the best results. For 3-month time scale, VMD-ANFIS hybrid model in ANFIS models, VMD-ANN hybrid model in ANN models, VMD-GPR hybrid model in GPR models, VMD-SVMR hybrid model in SVMR models, and EMD-KNN hybrid model in KNN models yielded the best results. For 6-month time scale. DWT-ANFIS hybrid model in ANFIS models, DWT-ANN hybrid model in ANN models, DWT-GPR hybrid model in GPR models, EMD-SVMR hybrid model in SVMR models, and EMD-KNN hybrid model in KNN models yielded the

best results. When all hybrid models were examined, according to the best results of each method, ANFIS and GPR methods for 1-month and 3-month time scales were compatible with VMD pre-processing method, while these methods were compatible with DWT pre-processing method for 6-month time scale. While the SVMR method was compatible with the VMD pre-processing method for 1-month and 3-month time scales, it was found to be compatible with the EMD pre-processing method for 6-month time scale. In ANFIS, GPR, and SVMR methods, it was seen that pre-processing method changed with the growth of time scales. It was determined that the KNN method was compatible with the EMD pre-processing method for all time scales.

Fig. 6 Scatter plots of calculated and predicted SPI values for 6-month time scale
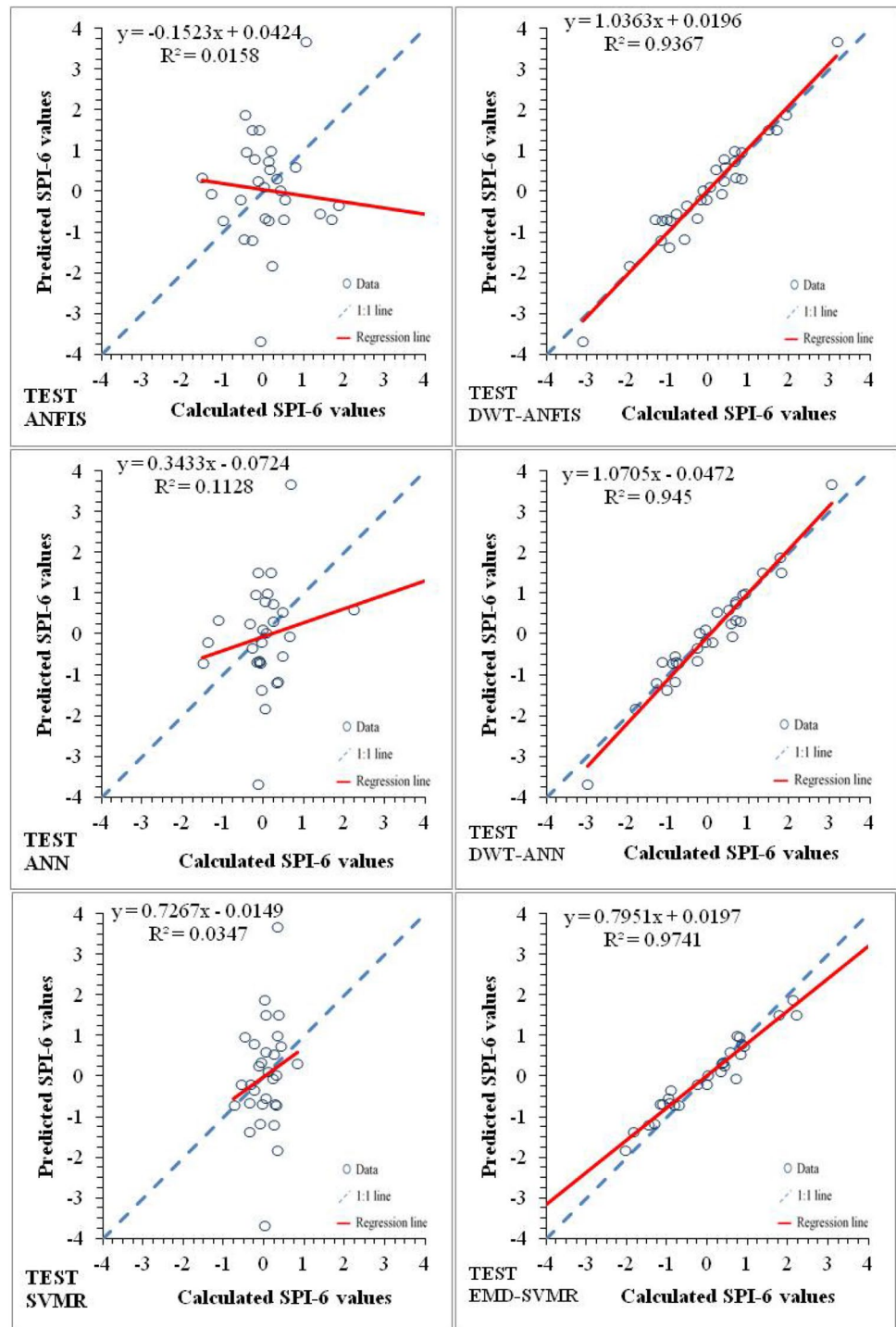
When all the hybrid models in Tables 7, 8, and 9 were examined, it was understood that all the hybrid methods obtained by the KNN method according to the performance criteria were considerably weaker than the other hybrid methods. In general, DWT-KNN and VMD-KNN hybrid models yielded higher *MSE*, *RMSE*, and *MAE* values than the stand-alone models for all time scales. As can be understood from these results, it was seen that hybrid

KNN models obtained by DWT and VMD pre-processing methods did not yield predictions as effectively as the other hybrid models. According to the performance criteria, the EMD-KNN model was a slightly more successful hybrid model.

As can be seen in Tables 7, 8, and 9, the model results of the test data yielded better results than the training data in six models out of twenty models belonging to each time

Fig. 6 (continued)
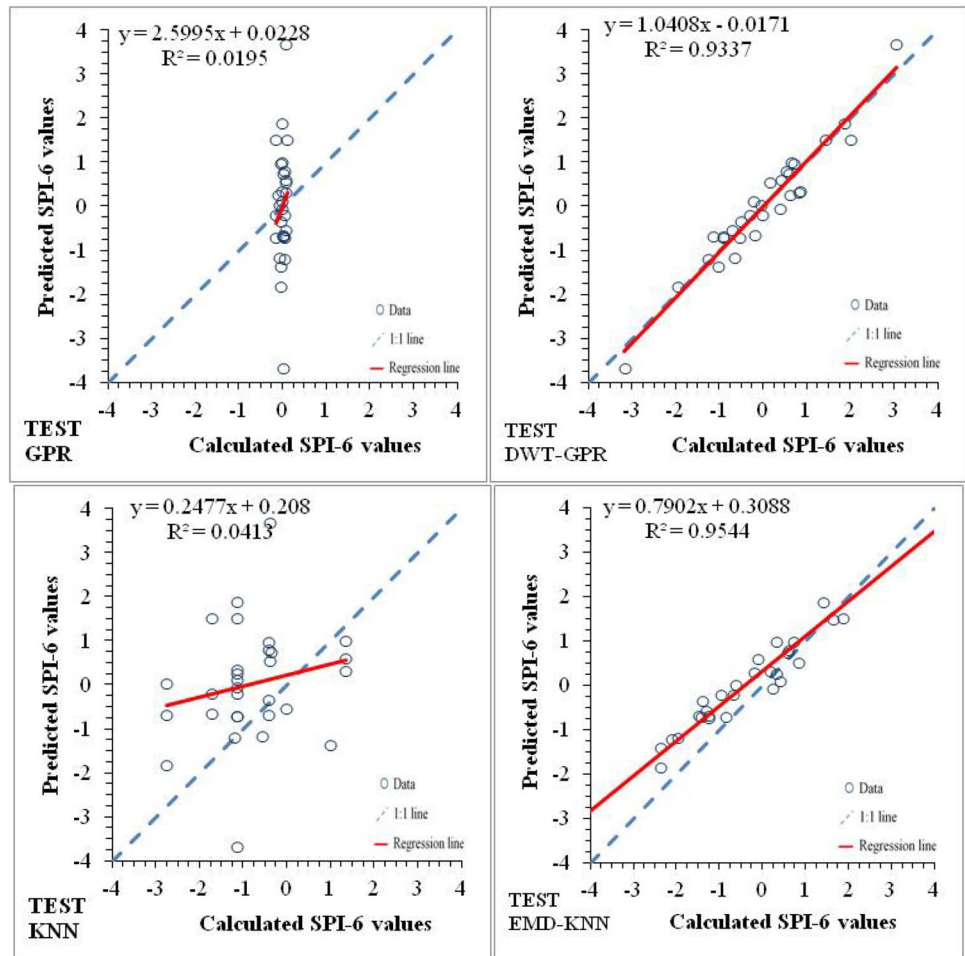
scale. However, in many prediction studies with machine learning, the performance of training data generally yielded superior results than the test data. In this study, most of the models in which the test data were superior to the training data were hybrid models using VMD. This is because the VMD pre-processing technique removed the boundary effect (Zuo et al. 2020; Purohit et al. 2021; Latifoğlu 2022).

As seen in Table 7, although the VMD-ANFIS and VMD-GPR models yielded close results for 1-month time scale, VMD-GPR was chosen as the best hybrid model because it had the lowest *MSE* and *RMSE* and high *NSE*, *OI*, and R^2 values. The *MSE*, *RMSE*, *NSE*, *OI*, and R^2 values of the VMD-GPR hybrid model of 1-month time scale were found to be 0.063, 0.251, 0.9345, 0.9438, and 0.9367, respectively. As can be seen in Table 8, when all hybrid models were examined for 3-month time scale, the VMD-GPR hybrid model yielded the best result according to all performance criteria. The *MSE*, *RMSE*, *NSE*, *OI*, and R^2 values of the VMD-GPR hybrid model were found to be 0.0582, 0.2413, 0.9528, 0.9559, 0.1962, and 0.9565, respectively. As seen in Table 9, DWT-ANFIS and

DWT-ANN models yielded close results for 6-month time scale, and DWT-ANN was determined as the best hybrid model since it yielded the lowest *MSE*, *RMSE*, and *MAE* and high *NSE*, *OI*, and R^2 values.

In Figs. 4, 5, and 6, the calculated and predicted SPI drought data of the hybrid model that yielded the best results, together with the stand-alone model of each method, were given.

As seen in Figs. 4, 5, and 6, most of the stand-alone models for all time scales did not fall above the $y = x$ (45°) line. As can be seen from these plots, it was understood that stand-alone models failed to predict drought. In hybrid models, on the other hand, it was seen that the relationship between the calculated and predicted SPI drought values was on the $y = x$ (45°) linear curve. In other words, hybrid prediction models seemed to be successful. As can be seen in Fig. 4, except for the EMD-KNN model, other hybrid models were quite successful for 1-month time scale. In the EMD-KNN hybrid model, the $y = x$ (45°) linear curve and the regression line did not coincide. As can be

Fig. 7 Taylor diagrams of SPI test data for **a** 1-, **b** 3-, and **c** 6-month time scales

a) 1-month

b) 3-month

c) 6-month
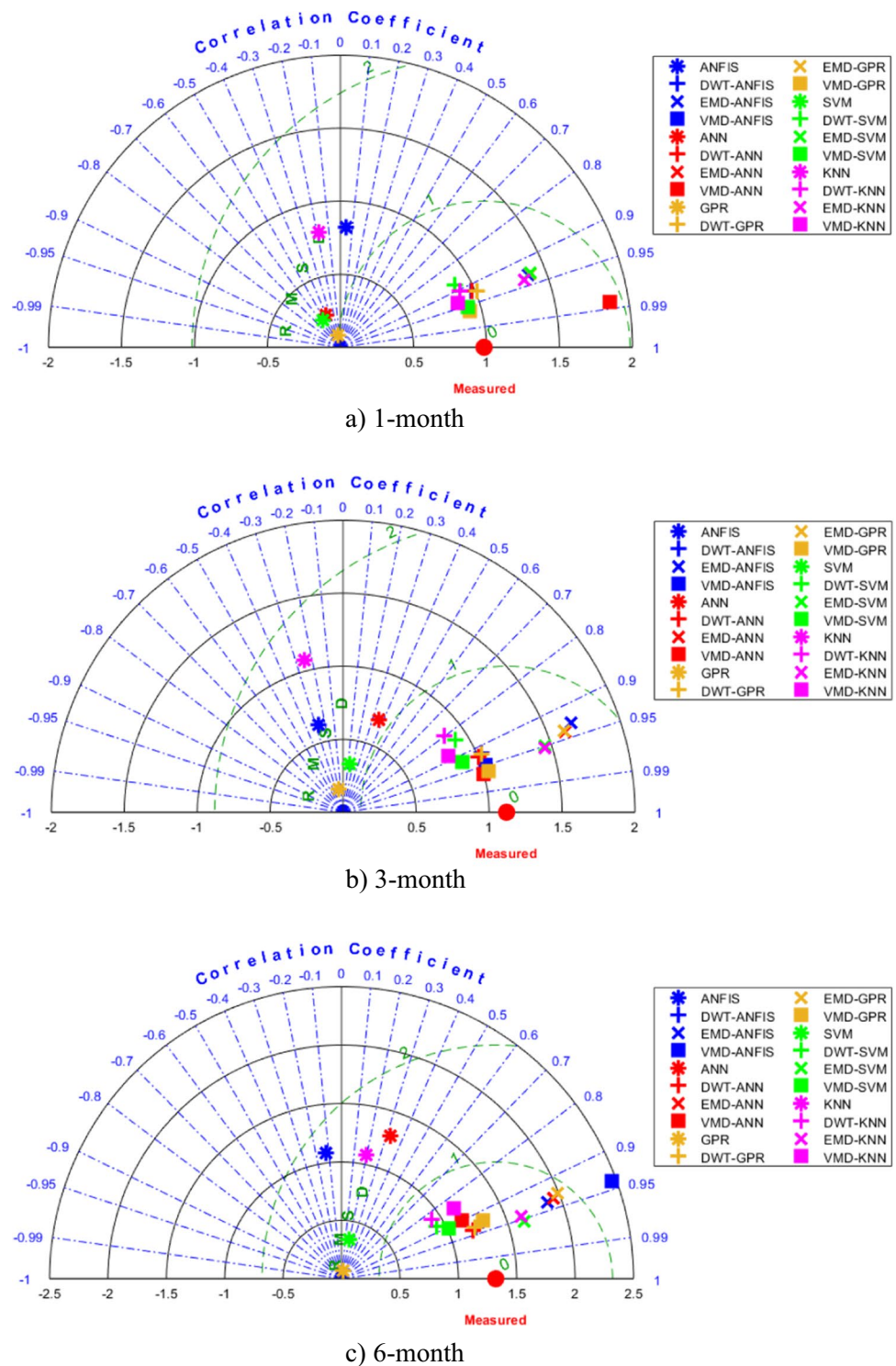
seen in Fig. 5, the linear curve of $y = x$ (45°) and the regression line did not coincide in the hybrid models of VMD-SVMR, VMD-ANN, and EMD-KNN for 3-month time scale. As can be seen in Fig. 6, the linear curve of $y = x$ (45°) and the regression line did not coincide in

the hybrid models of EMD-SVMR and EMD-KNN for 6-month time scale.

Besides the scatter plots, Taylor diagrams, in which all the models were together and which were created to compare the model results of the test data, are presented
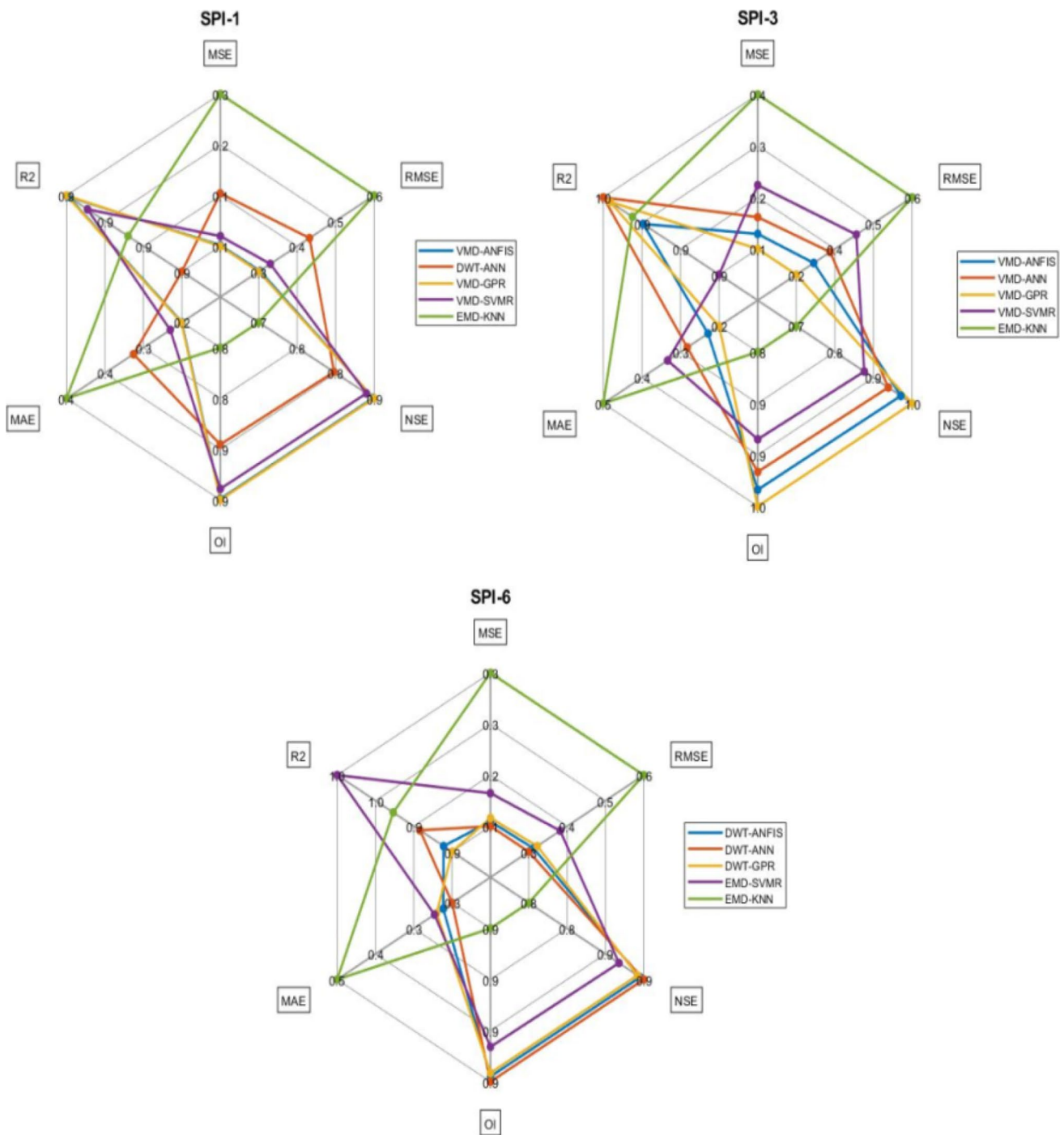
Fig. 8 Radar diagrams of the models that yielded the best five results in the SPI test data for 1-, 3-, and 6-month time scales

in Fig. 7. Taylor diagram examines the statistical errors of all models and their agreement with the reference data (Citakoglu 2021; Demir 2022b). As can be seen in Fig. 7, the hybrid models of VMD-ANFIS, VMD-GPR, and VMD-SVM yielded very close results for 1-month time scale. The VMD-ANN and VMD-GPR hybrid models yielded quite close results for 3-month time scale.

The DWT-ANN, DWT-GPR, and VMD-GPR hybrid models yielded similar results for 6-month time scale. When Taylor diagrams were examined in more detail, it was understood that VMD-GPR hybrid models yielded the best results for 1- and 3-month time scales, and the DWT-ANN hybrid model yielded the best results for 6-month time scale.

(a) Boxplot diagram of all models

(a) Boxplot diagram of all models

(b) Error boxplot diagram of models

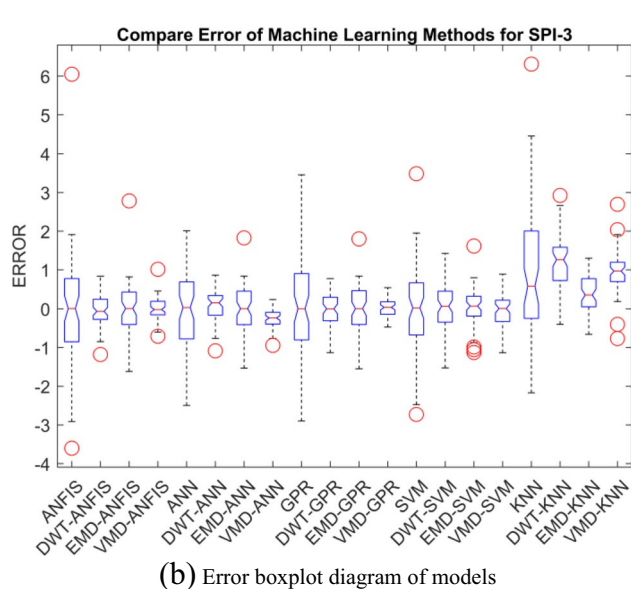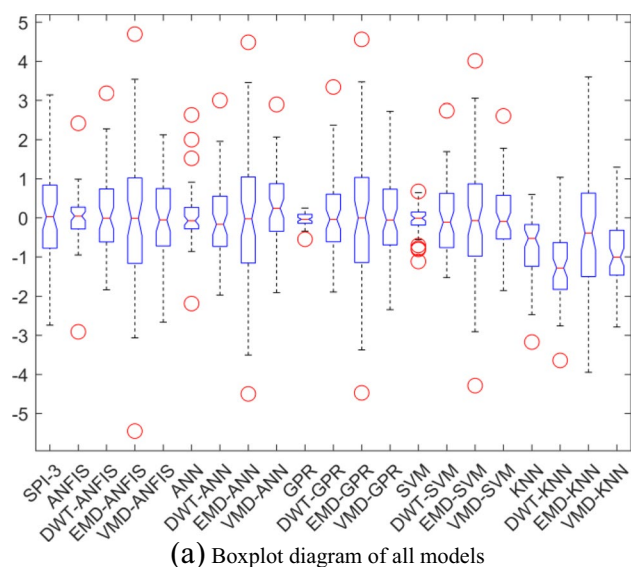(b) Error boxplot diagram of models
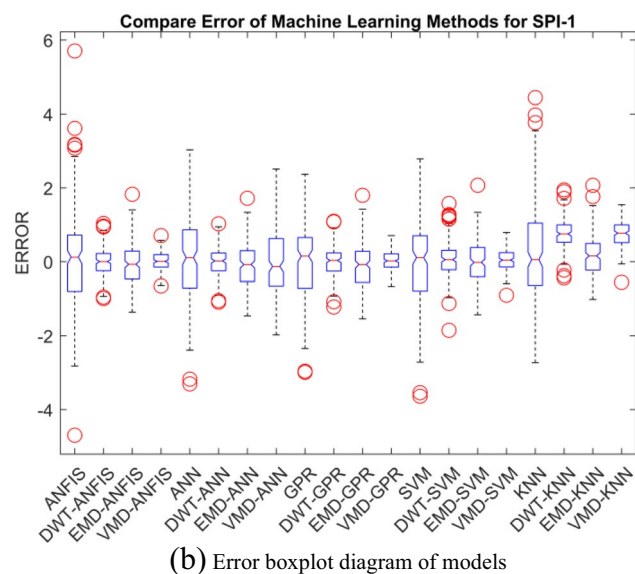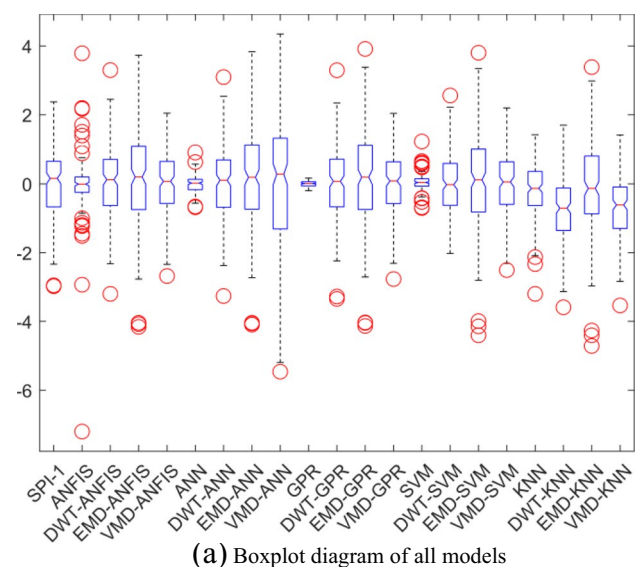
Fig. 9 Boxplot diagrams of all models and errors for 1-month time scale. **a** Boxplot diagram of all models. **b** Error boxplot diagram of models

Fig. 10 Boxplot diagrams of all models and errors for 3-month time scale. **a** Boxplot diagram of all models. **b** Error boxplot diagram of models

Radar diagrams with all performance criteria are presented in Fig. 8 to more easily compare the best results of the prediction models for each time scale. As can be seen in Fig. 8, VMD-GPR hybrid models for 1- and 3-month time scales and DWT-ANN hybrid models for 6-month time scale yielded the best prediction results according to all performance criteria.

Boxplot diagrams of stand-alone and hybrid models and error boxplot diagrams of these models were also drawn and presented in Figs. 9, 10, and 11. As can be seen in Figs. 9, 10, and 11, the stand-alone models did not resemble the original SPI-1, SPI-3, and SPI-6 data. According to boxplot diagrams, the highest error values

were given by the stand-alone models. As can be seen in Fig. 9, both model boxplot and error boxplot diagrams of VMD-ANFIS, VMD-GPR, and VMD-SVMR models were similar for 1-month time scale. As can be seen in Fig. 10, both model boxplot and error boxplot diagrams of VMD-ANN and VMD-GPR hybrid models were similar for 3-month time scale. As can be seen in Fig. 11, both model boxplot and error boxplot diagrams of the DWT-ANN, DWT-GPR, and VMD-GPR hybrid models were similar for 6-month time scale. As can be seen from these results, the boxplot diagram results overlapped with the Taylor diagram results.
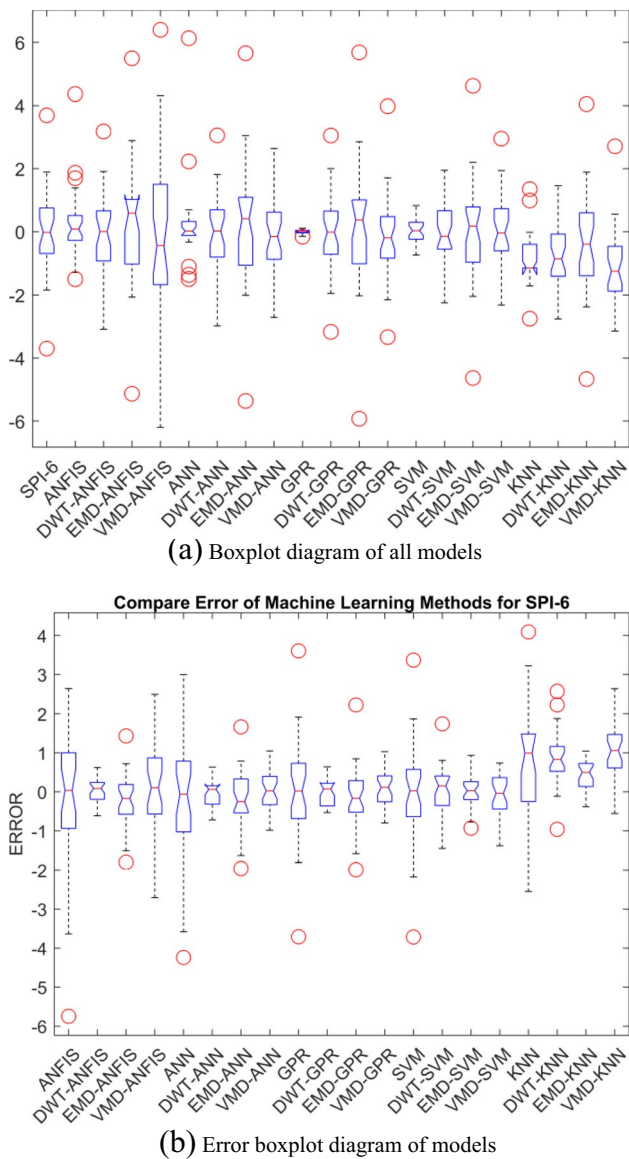
(a) Boxplot diagram of all models

(b) Error boxplot diagram of models

Fig. 11 Boxplot diagrams of all models and errors for 6-month time scale. **a** Boxplot diagram of all models. **b** Error boxplot diagram of models

Conclusion

In this study, twenty different prediction models were implemented to predict the future values of the SPI drought indices for 1-, 3-, and 6-month time scales of Sakarya Meteorology Station, located in the northwest of Turkey. SPI drought indices were calculated with the use of precipitation data of station between 1960 and 2020 period. Fifteen hybrid models were created by incorporating DWT, EMD, and VMD pre-processing techniques into five different ML methods (ANFIS, ANN, SVMR, KNN, GPR). In addition, five stand-alone models were created without pre-processing. The primary objective of

this study was to compare the effects of different pre-processing methods on different ML methods in the prediction of short-term SPI drought indices. The following conclusions were drawn from the present findings and analyses:

- The dmey was the most appropriate discrete wavelet family in the prediction of SPI drought indices for three different time scales.
- As for discrete wavelet transform, the 3rd band level yielded the best results. Increasing band levels did not improve the performance of prediction models.
- DWT-ANFIS model revealed t, $t-1$, $t-2$, and $t-3$ lag times input variables for the prediction of the SPI drought indices of Sakarya Meteorology Station.
- Models with pre-processing techniques performed better than stand-alone models.
- Hybrid models with KNN were not as successful as the other hybrid models.
- As the time scale increases, different pre-processing techniques yielded better predictions.
- According to performance criteria, scatter plots and Taylor and boxplot diagrams, VMD pre-processing technique yielded better results than DWT and EMD pre-processing techniques.
- The VMD-GPR hybrid model for 1- and 3-month time scales and the DWT-ANN hybrid model for 6-month time scale yielded the best predictions for SPI drought indices.

Present findings are expected to provide important contributions to decision-makers for short-term drought incidents to be encountered in Sakarya, one of the most important industrial provinces of Turkey.

Declarations

Ethics approval The authors paid attention to the ethical rules in the study. There is no violation of ethics.

Consent to participate Not applicable.

References

Akbar Jalilzadnezamabad (2019) Forecasting palmer drought severity index using hybrid wavelet-heuristic models. Istanbul Technical University

Başakın EE, Ekmekcioğlu Ö, Çıtakoğlu H, Özger M (2022) A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. Neural Comput Appl 34. https://doi.org/10.1007/s00521-021-06424-6

Başakın EE, Ekmekcioğlu Ö, Ozger M (2019) Drought analysis with machine learning methods. Pamukkale Univ J Eng Sci 25:985–991. https://doi.org/10.5505/pajes.2019.34392

Belayneh A, Adamowski J, Khalil B (2016) Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustain Water Resour Manag 2:87–101. https://doi.org/10.1007/s40899-015-0040-5

Citakoglu H (2015) Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. Comput Electron Agric 118. https://doi.org/10.1016/j.compag.2015.08.020

Citakoglu H (2017) Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. Theor Appl Climatol 130. https://doi.org/10.1007/s00704-016-1914-7

Citakoglu H (2021) Comparison of multiple learning artificial intelligence models for estimation of long-term monthly temperatures in Turkey. Arab J Geosci 14. https://doi.org/10.1007/s12517-021-08484-3

Citakoglu H, Minarecioglu N (2021) Trend analysis and change point determination for hydro-meteorological and groundwater data of Kizilirmak basin. Theor Appl Climatol 145. https://doi.org/10.1007/s00704-021-03696-9

Cortes C, Vapnik V (1995) Support vector networks. Mach Learn 20:293–297

Demir V (2022a) Trend analysis of lakes and sinkholes in the Konya Closed Basin, in Turkey. Nat Hazards. https://doi.org/10.1007/s11069-022-05327-6

Demir V (2022b) Enhancing monthly lake levels forecasting using heuristic regression techniques with periodicity data component: application of Lake Michigan. Theor Appl Climatol 148:915–929. https://doi.org/10.1007/s00704-022-03982-0

Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. IEEE Trans Signal Process 62. https://doi.org/10.1109/TSP.2013.2288675

Fadaei-Kermani E, Barani GA, Ghaeini-Hessaroeyeh M (2017) Drought monitoring and prediction using K-nearest neighbor algorithm. J AI Data Min 5:319–325. https://doi.org/10.22044/JADM.2017.881

Fadaei-Kermani E, Ghaeini-Hessaroeyeh M (2020) Fuzzy nearest neighbor approach for drought monitoring and assessment. Appl Water Sci 10:130. https://doi.org/10.1007/s13201-020-01212-4

Haktanir T, Citakoglu H (2015) Closure to "Trend, independence, stationarity, and homogeneity tests on maximum rainfall series of standard durations recorded in Turkey" by Tefaruk Haktanir and Hatice Citakoglu. J Hydrol Eng 20. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001246

Haktanir T, Citakoglu H (2014) Trend, independence, stationarity, and homogeneity tests on maximum rainfall series of standard durations recorded in Turkey. J Hydrol Eng 19:05014009. https://doi.org/10.1061/(asce)he.1943-5584.0000973

Haykin S (1998) Neural networks: a comprehensive foundation

Huang NE, Shen Z, Long SR et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc London Ser A Math Phys Eng Sci 454:903–995. https://doi.org/10.1098/rspa.1998.0193

Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23:665–685

Kaur A, Sood SK (2020) Deep learning based drought assessment and prediction framework. Ecol Inform 57:101067. https://doi.org/10.1016/j.ecoinf.2020.101067

Khanmohammadi N, Rezaie H, BehmaneshJavad, Khanmohammadi N (2022) Investigation of drought trend on the basis of the best obtained drought index. Water Resour Manag. https://doi.org/10.1007/s11269-022-03086-4

Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J Hydrol 399:132–140. https://doi.org/10.1016/j.jhydrol.2010.12.041

Kisi O, Latifoğlu L, Latifoğlu F (2014) Investigation of empirical mode decomposition in forecasting of hydrological time series. Water Resour Manag 28:4045–4057. https://doi.org/10.1007/s11269-014-0726-8

Kutyłowska M (2018) Application of K-nearest neighbours method for water pipes failure frequency assessment. E3S Web Conf 59:00021. https://doi.org/10.1051/e3sconf/20185900021

Latifoğlu L (2022) The performance analysis of robust local mean mode decomposition method for forecasting of hydrological time series. Iran J Sci Technol Trans Civ Eng. https://doi.org/10.1007/s40996-021-00809-2

McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales, in: Proc. 8th Conf. on Applied Climatology, Anaheim, California, 179–184. In: Eighth Conference on Applied Climatology. CA. American Meteorological Society, Boston, pp 17–22

Mishra AK, Desai VR, Singh VP (2007) Drought forecasting using a hybrid stochastic and neural network model. J Hydrol Eng 12:626–638. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:6(626)

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I — a discussion of principles. J Hydrol 10:282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Özger M, Başakın EE, Ekmekcioğlu Ö, Hacısüleyman V (2020) Comparison of wavelet and empirical mode decomposition hybrid models in drought prediction. Comput Electron Agric 179:105851. https://doi.org/10.1016/j.compag.2020.105851

Palmer WC (1965) Meteorological drought. Weather Bureau, Washington, DC

Purohit SK, Panigrahi S, Sethy PK, Behera SK (2021) Time series forecasting of price of agricultural products using hybrid methods. Appl Artif Intell 35:1388–1406. https://doi.org/10.1080/08839514.2021.1981659

Rasmussen CE, Williams CK (2006) Gaussian processes for machine learning (adaptive computation and machine learning). The MIT Press, Cambridge

Republic of Turkey M of A and FGD of M (2022) https://mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx

Salimi H, Asadi E, Darbandi S (2021) Meteorological and hydrological drought monitoring using several drought indices. Appl Water Sci 11:11. https://doi.org/10.1007/s13201-020-01345-6

Sevinc AS, Sen Z (2003) Spatio-temporal drought analysis in the Trakya region, Turkey. Hydrol Sci J 48:809–820. https://doi.org/10.1623/hysj.48.5.809.51458

Sihag P, Jain P, Kumar M (2018) Modelling of impact of water quality on recharging rate of storm water filter system using various

kernel function based regression. Model Earth Syst Environ 4:61–68. https://doi.org/10.1007/s40808-017-0410-0

Tigkas D, Vangelis H, Tsakiris G (2015) DrinC: a software for drought analysis based on drought indices. Earth Sci Informatics 8:697–709. https://doi.org/10.1007/s12145-014-0178-y

Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J Clim 23:1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Yagbasan O, Demir V, Yazicigil H (2020) Trend analyses of meteorological variables and lake levels for two shallow lakes in Central Turkey. Water 12:414. https://doi.org/10.3390/w12020414

Yu Y, Campo J, Orimoloye IR et al (2022) Academic editors: María Drought: a common environmental disaster. . https://doi.org/10.3390/atmos13010111

Zuo G, Luo J, Wang N et al (2020) Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. J Hydrol 585:124776. https://doi.org/10.1016/j.jhydrol.2020.124776

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.