

Wavelet Packet-Genetic Programming: A New Model for Meteorological Drought Hindcasting

Ali DANANDEH MEHR¹
Mir Jafar Sadegh SAFARI²
Vahid NOURANI³

ABSTRACT

This study presents developing procedures and verification of a new hybrid model, namely wavelet packet-genetic programming (WPGP) for short-term meteorological drought forecast. To this end, the multi-temporal standardized precipitation evapotranspiration index (SPEI) has been used as the drought quantifying parameter at two meteorological stations at Ankara province, Turkey. The new WPGP model comprises two main steps. In the first step, the wavelet packet, which is a generalization of the well-known wavelet transform, is used to decompose the SPEI series into deterministic and stochastic sub-signals. Then, classic genetic programming (GP) is applied to formulate the deterministic sub-signal considering its effective lags. To characterize the stochastic component, different theoretical probability distribution functions were assessed, and the best one was selected to integrate with the GP-evolved function. The efficiency of the new model was cross-validated with the first order autoregressive (AR1), GP, and random forest (RF) models developed as the benchmarks in the present study. The results showed that the WPGP is a robust model, superior to AR1 and RF, and significantly increases the predictive accuracy of the standalone GP model.

Keywords: Drought, SPEI, wavelet packet, genetic programming, stochastic modelling.

1. INTRODUCTION

Drought is a hydrological extreme condition that can bring serious problems to the human life. It makes significant impacts on water quantity and quality, of land and soil degradation, agricultural productivity, desertification, famine, etc. Drought conditions are usually

Note:

- This paper has been received on August 15, 2019 and accepted for publication by the Editorial Board on January 7, 2020.
- Discussions on this paper will be accepted by September 30, 2021.
- <https://doi.org/10.18400/tekderg.605453>

1 Antalya Bilim University, Department of Civil Engineering, Antalya, Turkey
ali.danandeh@antalya.edu.tr - <https://orcid.org/0000-0003-2769-106X>

2 Yaşar University, Department of Civil Engineering, Izmir, Turkey
jafar.safari@yasar.edu.tr - <https://orcid.org/0000-0003-0559-5261>

3 Center of Excellence in Hydroinformatics and Faculty of Civil Engineering, Uni. of Tabriz, Tabriz, Iran
Near East University, Faculty of Civil and Environmental Engineering, Nicosia, North Cyprus
vnourani@yahoo.com - <https://orcid.org/0000-0002-6931-7060>

classified under four categories of meteorological, hydrological, agricultural, and socioeconomic droughts. Each of these groups indicates how long a dry period lasts and effects the human life/environment. To monitor/forecast meteorological drought, deviation of meteorological variables such as precipitation, evaporation, and transportation from their long-term mean are investigated using variety of existing drought indices. There are various meteorological drought indices including (but not limited to) the well-known Standardized Precipitation Index (SPI, [1]), Drought Area Index (DAI, [2]), Palmer Drought Severity Index (PDSI, [3]), and the most recently developed Standardized Precipitation Evapotranspiration Index (SPEI; [4]). While SPI and DAI are calculated using historical precipitation data, calculation of PDSI and SPEI is based on both precipitation and temperature data series.

To date, a large number of research articles have been conducted to forecast drought indices using classical regression or data-driven techniques [5-14]. For example, the Artificial Neural Network (ANN) based forecast models of 1 to 12 months lead time were developed taking the advantage of SPI and Effective Drought Index [15]. The suitability of the Adaptive Neuro-Fuzzy Inference System (ANFIS) for SPI-based drought forecasting at different time scales was tested and compared to that of feed-forward neural networks [6]. Nonlinear Aggregated Drought Index-based drought conditions were forecast using the ANN approach [16]. The SPI was forecast at Bojnurd synoptic station using ANN, ANFIS, and Support Vector Regression (SVM) [17]. The results showed that the SVM model is superior to ANFIS and ANN models. A recent study demonstrated that M5-tree and multivariate adaptive regression splines (MARS) models outperform the least square support vector machine (LSSVM) model in the forecasting of the SPI in Australia [13].

With specific attention to hydrology, Labat (2005) noted that the hybridization of data-driven models using wavelets may lead to several improvements in the analysis of global hydrological fluctuations and their natural time-varying relationships [18]. Successful applications of wavelet-based hydrological models have also been reported in [19]. In drought forecasting community, wavelets were applied to decompose/denoise different drought indices or its predictors [9-10, 20-23]. The conjunction of wavelet transform and ANNs was used to forecast PDSI and the results showed that wavelet-based models can significantly increase the accuracy of forecasts [20]. Long-term SPI was forecast using ARIMA, ANN, SVM, W-ANN, and WSVM models [21]. Among the models, the W-ANN was more accurate. A gene-wavelet model was used to forecast Palmer Modified Drought Index (PMDI) in which previous PMDI series and NINO 3.4 index were employed as the inputs [10]. The results demonstrated that the standalone genetic programming (GP) model was unable to learn non-linearity of drought series in 3-month ahead forecasts. However, the gene-wavelet model could effectively forecast 3-, 6-, and 12-month lead times. More recently, wavelet-ARIMA-ANN and wavelet-ANFIS hybrid models were used to forecast SPEI drought index in a tropical climate in Malaysia [24]. The results suggested that wavelet-ARIMA-ANN forecast SPEI-3 and SPEI-6 accurately.

In this study, we tackle the problem of hindcasting drought index through GP coupled with a wavelet-packet. Besides, autoregressive (AR), non-coupled GP, and random forest (RF) methods were used as the benchmarks. We considered two case studies from Ankara, Turkey. For this purpose, we used a 46-year long time series of monthly SPEI at each case with no additional predictors such as large scale oceanic or atmospheric variables. Two time-scales,

SPEI-3 and SPEI-6, were analyzed separately. GP setup, RF modeling, and wavelet packet denoising are fully described in a step by step procedure. Besides, the impact of different decomposition depths (3, 6, and 9) on the improvement of standalone GP was investigated. Although a number of earlier studies have assessed the impact of wavelets in drought forecasting [10], to the best of the authors' knowledge, this is the first study that uses wavelet packet to improve GP-based predictions of drought indices.

2. STUDY AREA AND DATA

2.1. Overview of the SPEI Drought Index

The *SPEI* is rather a new meteorological drought index which combines the variability of precipitation and temperature to derive drought condition. It has been offered to be more consistent for drought studies under climate change projections [25, 26]. The index method first developed by [4] where local temperature and precipitation (*P*) are used to estimate monthly potential evapotranspiration (*PET*) and deficit (*D*), i.e., water balance of the month *i*.

$$D_i = P_i - PET_i \tag{1}$$

Once *D_i* is estimated, its standardized time series is fitted to a given probability distribution function (e.g., log-logistic function as suggested by 4) and then, the SPEI time series is calculated as the standardized values of cumulative probability of deficit using Equation (2).

$$SPEI_i = W_i - \frac{2.515517 + 0.802853 W_i + 0.010328 W_i^2}{1 + 1.432788 W_i + 0.189269 W_i^2 + 0.001308 W_i^3} \tag{2}$$

$$W_i = \sqrt{-2 \ln p} \quad \text{for } p \leq 0.5 \tag{2}$$

$$W_i = \sqrt{-2 \ln(1-p)} \quad \text{for } p > 0.5$$

Where *p* is the probability of exceeding a given deficit. It is important to note that the sign of the calculated SPEI must be reversed for the cases of *p*>0.5.

Table 1 - Classifications of drought events using SPI and SPEI indices

Classification	SPI threshold	SPEI threshold
Moderate drought	-1.49 <SPI< -1.0	-1.42 <SPEI< -1.0
Severe drought	-2.0 <SPI< -1.5	-1.82 <SPEI< -1.43
Extreme drought (ED)	SPI ≤ -2.0	SPEI ≤ -1.83

Returning to the study of [4], there is no drought classification thresholds for the SPEI values. In some studies, the well-documented standardized precipitation index (SPI) thresholds have been used regardless of the difference in cumulative density functions (CDFs) of the SPI and

SPEI indices [e.g., 27]. Considering the inconsistency caused by different CDFs, we used the drought classes suggested by Danandeh Mehr et al. (2019) in which SPEI thresholds were attained for a set of given probability of exceeding the deficit (see Table 1). For more details about the SPEI and its classification procedure, the interested reader is referred to [26].

2.2. Observed Data and Hindcasting Scenarios

Historical precipitation and temperature measurements in the period 1971-2016 from two meteorological stations (Beypazari and Nallihan; see Figure 1) were used in this study to calculate SPEI time series at each station. The stations with a rough distance of 55 km represent the climatology of north-western Ankara, Turkey. To obtain the SPEI time series in 3-month and 6-month time resolutions, the SPEI package available in R library (<http://sac.csic.es/spei/tools.html>) is used. The package uses Thornthwaite method to calculate PET by default, and we followed this initial setup. For more details about the climatology of the study area, the reader is referred to [26].



Figure 1 - Location of meteorological stations used in this study

Drought forecasting using data-driven methods can be divided into two categories of univariate hindcasting and multivariate forecasting models. In the univariate hindcasting models, the desired SPEI index is modelled using lagged values of the index, i.e., past events. This is a fast methodology, particularly useful in the station-scale studies. In the multivariate forecasting models, one may use exogenous predictors (inputs) or ensemble precipitation and PET forecasts from meteorological authorities like European Centre for Medium-Range Weather Forecasts (ECMWF).

One important step in drought hindcasting is to determine the optimum number of lags (m) which plays a significant role on the accuracy and complexity of forecasts. Despite the fact that the optimal set of lags leads to the reliable forecasts, inadequate or even more than required lags may result in weak or complex solutions, respectively. It might be identified

either via trial and error method [28] or through the autocorrelation function (ACF) analysis of drought index [21]. However, it is a linear method and the values achieved are often too large. To prevent the presence of spurious lags that tend to confuse the training process, the optimum lags can be judged via partial autocorrelation function (PACF).

In this study, SPEI-3 (i.e., precipitation and temperature conditions over 3 months) and SPEI-6 were forecast for both meteorology stations over the lead time of 1 month. The 3-month SPEI (SPEI-3) and 6-month SPEI (SPEI-6) forecasts with a 1-month lead time are desirable for early warning and taking action against meteorological drought. The relevant forecasting scenarios can be mathematically expressed as below:

$$SPEI-3(t) = f(SPEI-3_{t-i}, \dots, SPEI-3_{t-m}, \varepsilon_t) \quad (3)$$

$$SPEI-6(t) = f(SPEI-6_{t-i}, \dots, SPEI-6_{t-m}, \varepsilon_t) \quad (4)$$

where the time index i ($=1, \dots, m$), the lag indicating the number of past months (i.e., lags), must be considered in the scenarios.

3. METHODS

3.1. The Benchmark Random Forest and Autoregressive Models

Random forest (RF) is constructed through the combination of several numbers of decision trees (DT). RF is an ensemble machine learning technique that has been widely used for classification and regression analysis [29]. Recently, RF has been applied in verity of water resources problems [30-32]. Poor performance on testing data set, i.e., overfitting is the primary problem in the application of conventional DT. The RF works out the deficiency of DT through its randomness feature [33]. Randomness elements have been introduced in a growing process by the partial selection of variables in the tree structure. Created trees are combined to generate a robust predictor based on a given dataset. The predicted values are the results of averaging the individual outputs. RF differs from the conventional DT in which the bootstrap sample is used instead of using all training data for each tree creation. The decision tree separates the data into portions by using the best splitter variable at a time. While RF changes the split selection procedure by randomly choosing the predictors. The robustness of RF comes from the combination of several trees. As basic futures of RF, there is no need to rescale the data like other data-driven techniques, and it can determine the best predictor, automatically. As it uses several trees at its structure, RF provides more accurate results and prevents overfitting. The regression tree generates a non-linear model via a collection of some linear portions of training data. The RF regression model is generated through a random vector that grows the trees. The results of such tree predictors are numerical values. RF provides numerical outputs and the training set is independently drawn from the random vector distribution [29].

Assuming a regression problem having training data of Z with N records and P variables, it is aimed to obtain prediction $\hat{f}_{rf}^B(x)$ in input x . Through bootstrap aggregation or bagging averages, variance in predication is reduced. The model is fitted for each bootstrap samples of $b = 1, 2, 3, \dots, B$. RF is created firstly by random selection of m variables from set of P

variables, then best variable is picked from m variables and finally, the node is split into two daughter nodes. This process is repeated until the minimum node size n_{min} is achieved at each terminal node to grow a random forest tree T_b on the bootstrapped data. In order to make prediction at point x , it is defined as

$$f_{rf}^{\wedge B}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \tag{5}$$

Through modeling of the random forest, it is required to set the number of trees. To this end, various number of trees should be evaluated by trial and error process.

The first order autoregressive AR1 model which is a standard linear difference equation is also used as another benchmark. It is a well-documented model that directly relates the parameter x at time k to the value of x at a previous time period, plus another variable ϵ dependent on time k .

3.2. Canonical GP

GP is one of the modern soft computing methods emerged in the last three decades. Similar to the well-known genetic algorithm, GP uses Darwinian evolutionary process to find the best solution but unlike GA, it simultaneously optimizes both the solution structure (function) and its coefficients. In other words, the functional form of the best model (solution) has not been chosen in advance. The three main evolutionary operations that map a population of random functions, aka GP, trees to a set of potential solutions include reproduction, crossover, and mutation. An example of randomly produced GP tree and its associated functional expression are illustrated in Figure 2. As shown in the figure, the function has three levels (tree depth =4) constructed from a root node level (plus function at the top), three inner nodes, and five terminal nodes comprising two variables x_1 and x_2 and two random constants linked via branches.

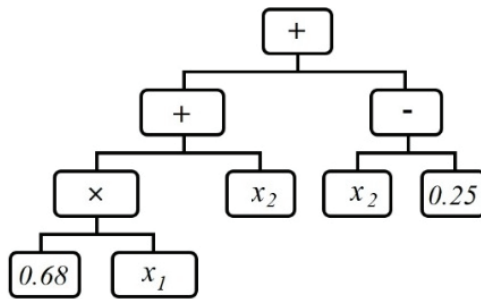


Figure 2 - Structure of GP tree representing the function $y = 0.68x_1 + x_2 + (x_2 - 0.25)$

During the training process, the three that shows the fittest function to the given problem is transferred to the next population without any alteration which is called reproduction. Crossover is the interchange of tree materials between two high performance trees called

parents. Mutation is the last evolutionary operation in which a randomly selected node (inner or terminal) is replaced with another node from the initial population that has not necessarily the same functionality with the selected node. These three operations are iterated on the GP trees till the fittest function is generated for the given problem or a function satisfies the modeler's goal. GP and its variants were successfully applied to solve variety of engineering problems [34-35]. This is more or less the whole processes in canonical GP as well as its different variants. For details its applications in hydrological studies, the reader is referred to [35].

3.3. Wavelet Packet Transform

The conventional Fourier transform was extended as wavelet transform that provides a multi-resolution analysis of signals. Wavelet packets (WP) are the generalization of wavelet transform capable of providing better frequency localization of signals [36]. The WP provides extensive decomposition over classic wavelet transform. The foremost difference between wavelet transform and WP transform is the way of the decomposition of a signal. Whereas classic wavelet transform decomposes only the low pass (approximation subspace) signals, the WP transform effectively decomposes both the low pass and high pass (detail subspace) components for the production of high-frequency resolution. Indeed, WPs are ways of mixing and matching wavelet filter banks in tree-like structures to create arbitrary time-frequency tiling. Recently WP has been compared with discrete wavelet transform [36].

3.4. Structure of the Proposed WPGP Model

The proposed WPGP model is a hybrid evolutionary model (Figure 3) that intends for the fine-tuning of the denoised SPEI signals attained from WP decomposition. As illustrated in Figure 4, the model includes three phases of data pre-processing (denoising), modeling (GP mapping and random noise creation), and data post-processing phase (noise injecting). In the first phase, SPEI time series at desired time horizons (here SPEI-3 and SPEI-6) are decomposed using WP transformation procedure. The result is a denoised SPEI time series plus a noise signal. The WPs can be used for various expansions of the original SPEI time series. Thus, an important task in this phase is to determine the most appropriate level of decomposition for the given signal. Either an entropy-based criterion [37] or Brute-force search method can be used to optimal decomposition selection.

In the modeling phase, the denoised SPEI time series and an optimum number of its lags are used respectively as target and predictor vectors to construct a predictive model. To create an explicit model, the GP could be an ideal option. As it is also capable of identifying the best predictors (among a set of input vectors), the well-established ACF analysis can be used to select required lags, i.e., input vectors. Since the GP deals with deterministic time series in this phase, a range of precise predictive models/solutions may be evolved. However, such solutions need to be modified using the stochastic component (noise) of the original SPEI time series. To this end, an appropriate probability distribution function (PDF) that perfectly represents the noise signal pattern is added to the evolved deterministic model. The identification of the best PDF for the noise signals is an important issue in this phase. In the present study, different unbounded distribution models were fitted to the noise signals, and

Kolmogorov–Smirnov (K-S) goodness-of-fit test was utilized in order to obtain the PDF so that it perfectly describes the likelihood of the behavior of the noise at each decomposition level. The assessed distribution models include Four-parameter Johnson SB, Beta, Wakeby, and Burr distributions, three-parameter Weibull, Log-Logistic, generalized extreme value, and two-parameter Normal distributions. The distribution parameters were estimated by the maximum likelihood estimation method in all the models.

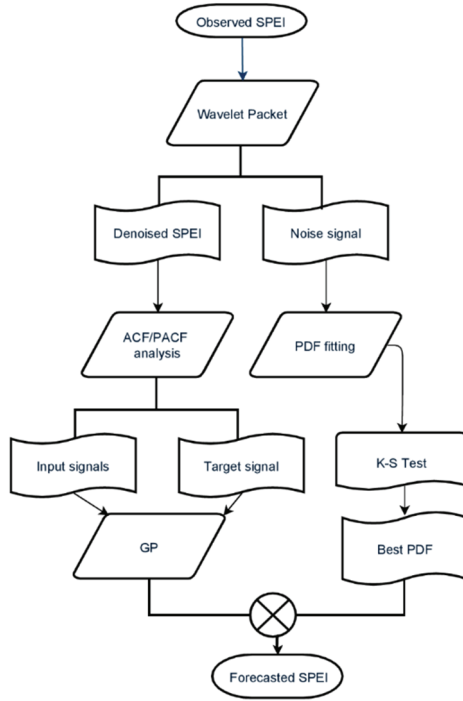


Figure 3 - The proposed WPGP model for SPEI hindcasting

It is worth mentioning that GP solutions need to control in terms of program size (so-called complexity), as the evolved models could be of illogical complexity that may result in over-fitting problem [35]. To avoid such a problem, we selected the best solution at potential validation data sets instead of the best in the training data set that guarantees both parsimony and generalizability of the best solutions. The idea was taken from the distinctive feature of GP that produces a population of solutions (potential models) instead of improving a single solution at each iteration for a given problem.

As previously mentioned, the last phase is the reconstruction of the forecasted SPEI through the accumulation of the GP-evolved model and the perfect PDF. For stochastic processes, it's logistic to provide a range of future forecasts instead of a deterministic value. To this end, maximum and minimum values of a set of random numbers (noises) that follow the distribution of the obtained perfect PDF generated at each forecasting scenario can be considered as the lower and upper bounds of SPEI forecasts.

4. PERFORMANCE EVALUATION METRICS

Nash-Sutcliffe efficiency (NSE) and root mean squared error (RMSE) measures were used in the present study.

$$NSE = 1 - \frac{\sum_{i=1}^n (X_i^{obs} - X_i^{pre})^2}{\sum_{i=1}^n (X_i^{obs} - X_{mean}^{obs})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i^{obs} - X_i^{pre})^2}{n}} \quad (7)$$

where X_i^{obs} = observed value of X (here SPEI), X_i^{pre} = predicted value X_{mean}^{obs} = mean value of observed data, and n is the number of observed data.

5. RESULTS AND DISCUSSION

5.1. RF and Baseline GP results

To develop the benchmark of RF and GP models at each station, the best input vectors were determined using the ACF and PACF analysis in each scenario (Figure 4). The figure shows more or less the same autocorrelation pattern for a given index at both stations. Thus, the forecasting scenarios were structured as below:

$$SPEI-3(t) = f(SPEI-3_{t-1}, SPEI-3_{t-2}, SPEI-3_{t-4}, SPEI-3_{t-7}, SPEI-3_{t-10}, \varepsilon_t) \quad (8)$$

$$SPEI-6(t) = f(SPEI-6_{t-1}, SPEI-6_{t-2}, SPEI-6_{t-3}, SPEI-6_{t-4}, SPEI-6_{t-7}, SPEI-6_{t-13}, \varepsilon_t) \quad (9)$$

The RF and GP structures were trained to minimize the *MSE* between the outputs from the model and the targets in the data set. Determination of the optimum number of trees in RF structure is of importance to construct the best RF model. Therefore, several number of trees are examined through trial and error procedure. To this end, RF models are constructed adjusting the maximum 200 number of trees. Results for two meteorological stations of Beypazari and Nallihan for both indexes of SPEI-3 and SPEI-6 are shown in Figure 5 at training stage. It is seen from Figure 6 that there is noticeable reduction in *MSE* until 20 number of trees for all cases, however, for higher number of trees no significant changes are seen. The optimum number of trees are shown in Figure 5 by red vertical lines where for SPEI-3 and SPEI-6 at both stations, they are found to be 60 and 70, respectively.

As useful feature of RF technique, it examines the importance of input variables on computing the output. Results of four RF models for two meteorological stations of Beypazari and Nallihan and for SPEI-3 and SPEI-6 indexes are given as pie charts in Figure 6. It is seen in Figure 6 that RF provides similar results for Beypazari and Nallihan stations

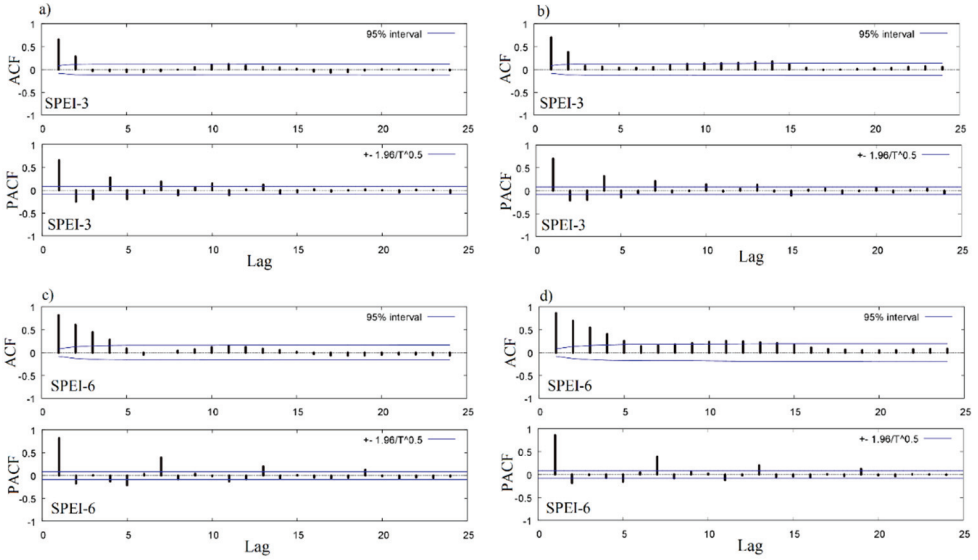


Figure 4 - Autocorrelation function (ACF) and Partial autocorrelation function (PACF) of original SPEI-3 (a and b) and SPEI-6 (c and d) time series at Bepazari (left column) and Nallihan (right column) meteorology stations

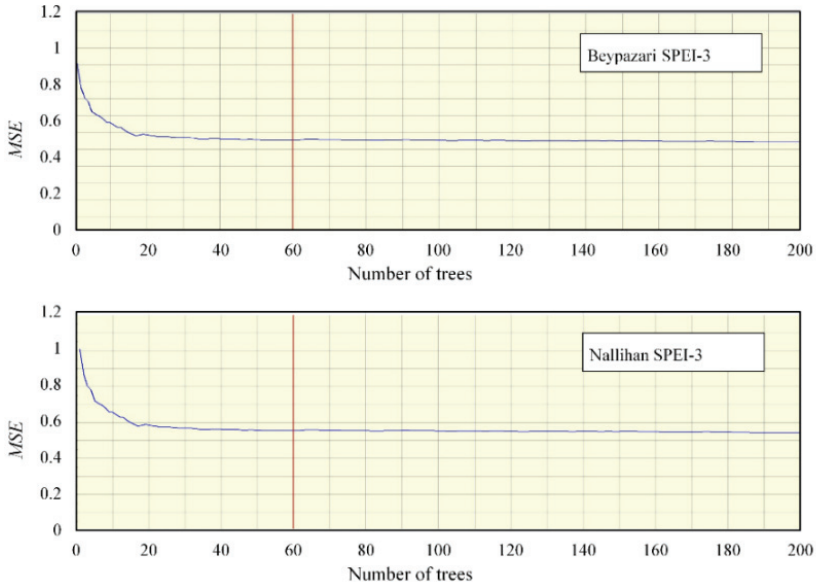


Figure 5 - Performance of RF models with different number of trees

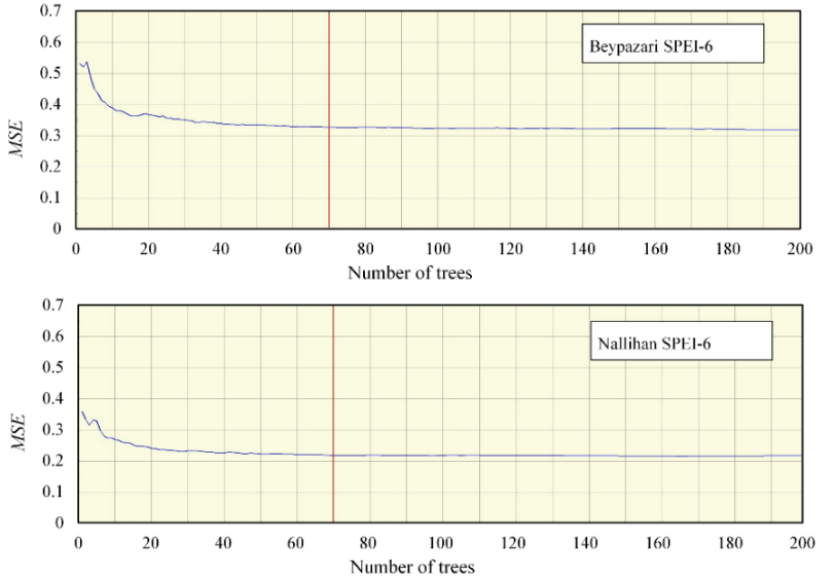


Figure 5 - Performance of RF models with different number of trees (continue)

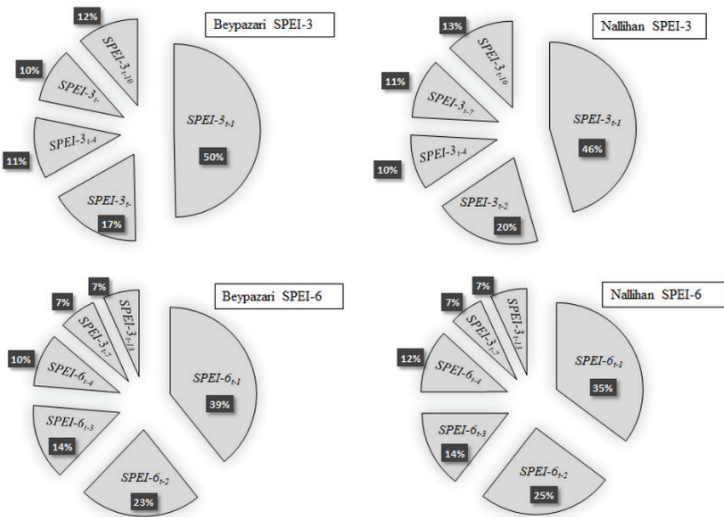


Figure 6. Variable importance at the RF models

considering two different SPEI-3 and SPEI-6 indexes. It is found that SPEI-3_{t-1} and SPEI-3_{t-2} have greatest effect while, SPEI-3_{t-4}, SPEI-3_{t-7} and SPEI-3_{t-10} are less important in RF models output. Similar results are achieved for SPEI-6 where, SPEI-6_{t-1} and SPEI-6_{t-2} are found as most important variables and SPEI-6_{t-7} and SPEI-6_{t-13} have the lower contribution in model

output. It is worthy to mention that the degree of importance of input variables is also related to selected SPEI index. It is seen in Figure 6 that t-1 lag in SPEI-3 is more important than SPEI-6. On the other hand, the degree of importance of t-2 lag is increased in SPEI-6 in comparison with SPEI-3 index results.

Table 2 - Parameter setting for standalone GP runs

Parameter	Value
Functions	+, -, *, /
Maximum generation	250
Initialization	Half and Half
Initial Population	500
Mutation rate	5 %
Crossover rate	90 %
Reproduction rate	20 %
Maximum depth	6

Table 3 - Performance metrics of the best RF and GP models

Model	Station	index	RMSE		NSE	
			Train	Test	Train	Test
AR1	Bey pazari	SPEI-3	0.712	0.837	0.448	0.362
		SPEI-6	0.575	0.664	0.642	0.576
	Nallihan	SPEI-3	0.727	0.880	0.391	0.322
		SPEI-6	0.682	0.864	0.457	0.333
RF	Bey pazari	SPEI-3	0.737	0.839	0.409	0.359
		SPEI-6	0.564	0.658	0.656	0.582
	Nallihan	SPEI-3	0.669	0.821	0.486	0.409
		SPEI-6	0.469	0.618	0.745	0.652
GP	Bey pazari	SPEI-3	0.670	0.831	0.511	0.371
		SPEI-6	0.529	0.637	0.697	0.610
	Nallihan	SPEI-3	0.628	0.806	0.547	0.430
		SPEI-6	0.449	0.576	0.765	0.700

GPdotNetV5.0, the open source software frame work [38], was used in this study to develop standalone GP-based SPEI forecasting models. The software has been successfully applied in variety of hydrological modelling tasks in the recent studies [e.g., 39-40]. The evolutionary parameters adopted for GPdotNetV5.0 setup were tabulated in Table 2. Among the hundreds

of standalone GP models evolved at each scenario, the best models were selected with respect to both accuracy and complexity of the potential solution at validation period. The complexity of each model is calculated after finding and eliminating the introns in GP trees.

It is worth to mention that a set of random floating-point numbers in the range $[0, 1]$ were used in terminal set given that the noise term is not separated in standalone GP models. Table 3 presents the performance metrics of the best RF and GP models at each station. The table also includes performance of the AR1 models developed for each index. The results show that GP models produce better forecasts than both RF and AR1 in terms of all error measures. However, they are not accurate enough, particularly during testing period.

Table 3 clearly shows that the models trained for SPEI-6 achieved higher accuracy than the models trained for SPEI-3. The reason behind it may be due to the smoother time series in SPEI-6 as a result of averaging of temperature and precipitation over six months period.

5.2. The WPGP Results

To forecast drought in a regional- or even catchment-scale, modelers typically deal with drought time series from several meteorology stations. It is not a good plan to denoise each of them individually. Thus in the first step, we create a matrix of such time series (here two dimensional real data) before using WP decomposition. In each hindcasting scenario, first a concatenated signal was made and then, the denoised signal in concatenated form was created. Now, the modeler can subtract denoised signal from original SPEI series to obtain the remaining noise signal. Figure 7 exhibits the concatenated SPEI-3 and SPEI-6 signals decomposed at depths 3, 6, and 9 with Debauches (db4) wavelet packets. Bearing in mind that db4 is commonly used to decompose hydrological time series [41], one may use any other type of WPs to this end. As previously mentioned, the optimal subtree of an initial wavelet packet tree can be obtained with respect to an entropy type criterion. In this study, the optimal subtree at each depth was computed considering Shannon entropy criterion that perhaps yields to a smaller tree than the initial one. Figure 8 provided an example of the initial and best WP trees at depth 3.

To select the best WP tree depth (i.e., deterministic part of SPEI series) at each meteorology station, the noise signals were suppressed and the denoised versions of SPEI series were modeled using GP. As illustrated in Figure 7, the inputs are the optimum number of lagged denoised signals having the highest correlation with the denoised SPEI at current time (i.e., 1-month ahead hindcasting scheme). Table 4 presented accuracy of the WPGP models at different depths.

The modeling results showed that different decomposition depths had different impacts on the original SPEI signal. The best deterministic model was found at depth 6. Therefore, the associated noise signals (see Figure 9 left column) were used to determine the perfect noise PDF at each station/time horizon. The results of goodness-of-fit tests applied to identify the PDF of the noise signals were given in Table 5 and the best distribution model was compared to normal distribution in Figure 9 (right column). Out of 8 empirical models, the K-S test showed that the Johnson SB distribution is the best representor for noise signal in SPEI-3. Therefore, the relevant forecasting scenarios can be mathematically expressed as Equations 10-12.

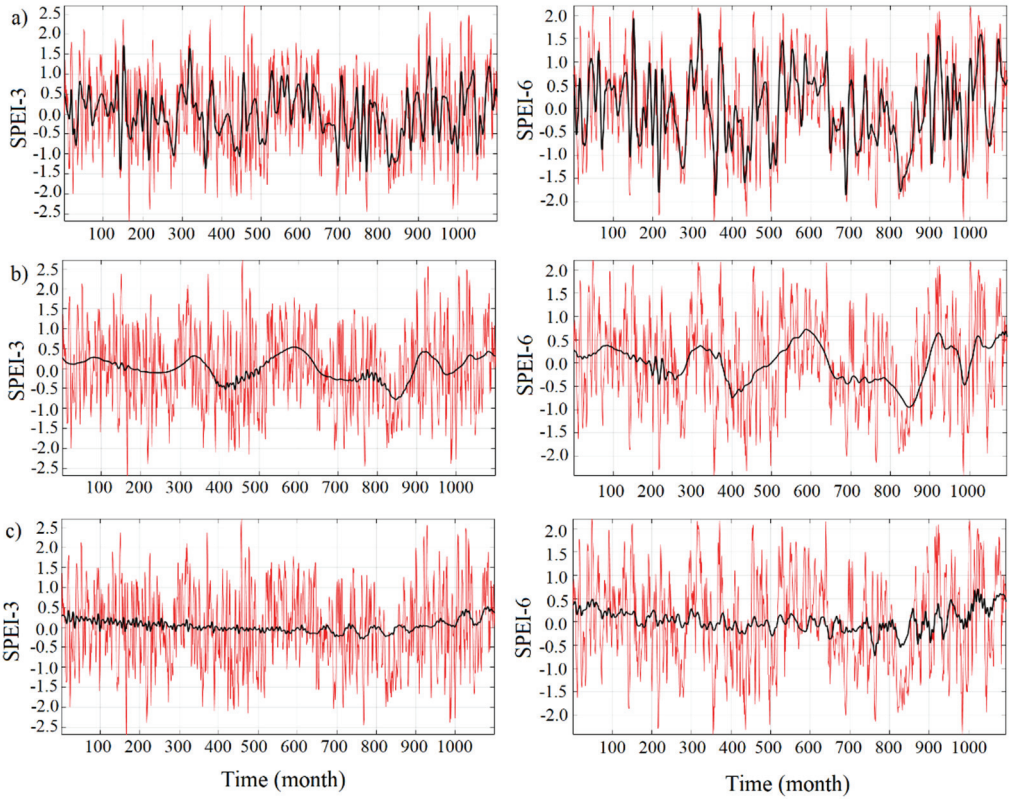


Figure 7 - Concatenated original SPEI and denoised SPEI signals at decomposition depth of (a) 3, (b) 6, and (c) 9.

Table 4 - Efficiency results of one-month a head WPGP models at different depth

Station	Drought index	WP Depth	Best model	Training		Testing	
				NSE	RMSE	NSE	RMSE
Beypazari	SPEI-3	3	$x_4 + 0.927(x_4 - x_3) + e$	0.983	0.072	0.993	0.044
		6	$x_4 - (((x_1 \times x_2) / (x_3 + x_2)) \times 0.84) + 0.84 \times (x_3 - x_4) + e$	0.999	0.004	0.992	0.023
		9	$(x_4 / 0.777) - ((x_2 - x_1) + (0.225 \times ((x_2 - 0.025) \times 0.854))) + e$	0.844	0.04	0.731	0.021

Table 4 - Efficiency results of one-month ahead WPGP models at different depths (continue)

Beypezari	SPEI-6	3	$x_4 - 0.9016(x_3 - x_4) + e$	0.987	0.086	0.992	0.067
		6	$((0.8776(x_4 - x_3)) + x_4) - (((x_3 \times (x_2 - x_1)) \times x_4) / (x_2 / x_4)) + e$	0.997	0.011	1.00	0.006
		9	$x_4 - (((((x_3 - 0.3157) + (x_3 - 0.9395)) \times ((x_2 / 0.9395) - x_4)) + 0.785) \times (x_3 - x_4)) + e$	0.987	0.017	0.990	0.011
Nalihan	SPEI-3	3	$2x_4 - x_3 - 0.0447(2x_4 - x_3) + e$	0.991	0.059	0.989	0.062
		6	$x_4 - (0.817(x_3 - x_4)) + e$	0.999	0.013	1.00	0.003
		9	$((1.367 - x_4 \times x_1) (x_1 - x_2)) (0.8085 - x_4) + (((x_4 - x_1) ((0.559 - x_4) / (1.118 - x_2))) + x_4 + e$	0.982	0.012	0.989	0.016
	SPEI-6	3	$x_4 - (0.8357((x_3 / 0.966) - x_4)) + e$	0.995	0.054	0.994	0.0962
		6	$x_4 + ((x_4 / (x_2 \times x_4)) / (x_2(x_4 - x_3))) + e$	1.000	0.003	0.999	0.009
		9	$x_4 - (0.745((x_3 - x_4))) + e$	0.975	0.031	0.974	0.041

$x_1 = \text{SPEI}_{t-4}$, $x_2 = \text{SPEI}_{t-3}$, $x_3 = \text{SPEI}_{t-2}$, and $x_4 = \text{SPEI}_{t-1}$

$$\text{SPEI-3}(t) = \text{SPEI-3}_{t-1} - (((((\text{SPEI-3}_{t-4} \times \text{SPEI-3}_{t-3}) / (\text{SPEI-3}_{t-2} + \text{SPEI-3}_{t-3})) \times 0.84) + 0.84) \times (\text{SPEI-3}_{t-2} - \text{SPEI-3}_{t-1})) + \text{SPEI-3}_N \text{ for Beypezari station (10)}$$

$$\text{SPEI-3}(t) = \text{SPEI-3}_{t-1} - (0.817(\text{SPEI-3}_{t-2} - \text{SPEI-3}_{t-1})) + \text{SPEI-3}_N \text{ for Nalihan station (11)}$$

Where SPEI-3_N represents the noise value extracted from Johnson SB distribution with shape parameters of $\gamma=0.32655$ and $\delta=1.9162$, the scale parameter of $\lambda=7.852$, and the location parameter of $\xi=-3.6218$ at Beypezari station and $\gamma=0.08172$, $\delta=1.8022$, $\lambda=7.002$, and $\xi=-$

3.4173 at Nallihan station. Considering these parameters the Johnson SB distribution is expressed as below:

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi z(1-z)}} \exp\left(-\frac{1}{2}\left(\gamma + \delta \cdot \ln\left(\frac{z}{1-z}\right)\right)^2\right) \text{ and } z = (SPEI-3_{(t-1)} - \xi)/\lambda \quad (12)$$

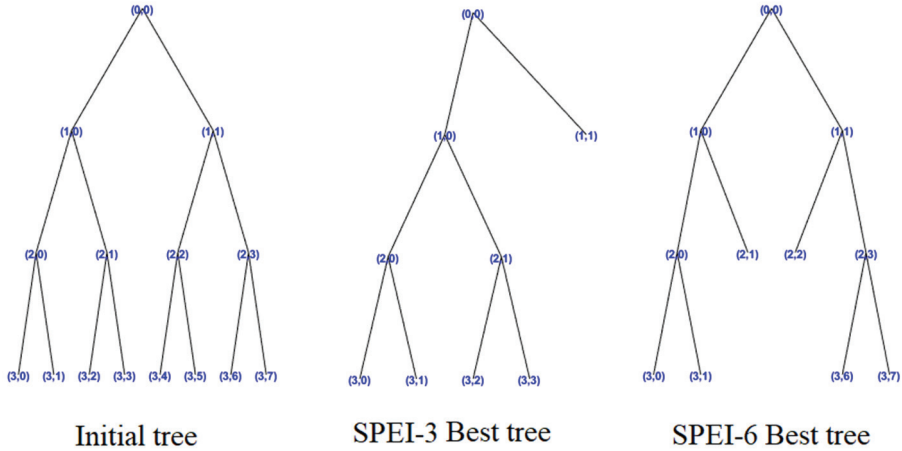


Figure 8 - Example of initial and best wavelet packet tree at depth 3

Table 5 - Summary of Kolmogorov-Smirnov Goodness of Fit test

Distribution	SPEI-3				SPEI-6			
	Beypazari		Nallihan		Beypazari		Nallihan	
	Statistic	Rank	Statistic	Rank	Statistic	Rank	Statistic	Rank
<u>Johnson SB</u>	0.023	1	0.015	1	0.022	3	0.022	3
<u>Beta</u>	0.025	2	0.016	3	0.022	4	0.022	2
<u>Wakeby</u>	0.025	3	0.021	6	0.022	5	0.026	5
<u>Burr (4P)</u>	0.026	4	0.018	4	0.021	2	0.023	4
<u>Weibull (3P)</u>	0.026	5	0.018	5	0.021	1	0.022	1
<u>Gen. Extreme Value</u>	0.026	6	0.015	2	0.024	7	0.024	6
<u>Normal</u>	0.031	7	0.024	7	0.028	9	0.026	7
<u>Log-Logistic (3P)</u>	0.045	8	0.034	8	0.036	11	0.033	8

Considering the SPEI-6, Weibull probability function is the best distribution model. Therefore, the relevant forecasting scenarios can be mathematically expressed as Equations 13-15.

$$SPEI-6(t) = ((0.8776(SPEI-6_{t-1} - SPEI-6_{t-2})) + SPEI-6_{t-1}) - (((SPEI-6_{t-2} \times ((SPEI-6_{t-3} - SPEI-6_{t-4}) \times SPEI-6_{t-1}) / (SPEI-6_{t-3} / SPEI-6_{t-1})) + SPEI-6_N \text{ for Beypazari station} \quad (13)$$

$$SPEI-6(t) = SPEI-6_{t-1} + ((SPEI-6_{t-1} / (SPEI-6_{t-3} \times SPEI-6_{t-1})) \times (SPEI-6_{t-3} \times (SPEI-6_{t-1} - SPEI-6_{t-2}))) + SPEI-6_N \text{ for Nallihan station} \quad (14)$$

Where $SPEI-6_N$ represents the noise value extracted from Weibull distribution with the shape parameter of $\alpha=3.5324$, the scale parameter of $\beta=3.2345$, and the location parameter of $\gamma=-2.9211$ at Beypazari station as well as $\alpha=3.4774$, $\beta=2.8798$, and $\gamma=-2.5827$ at Nallihan station. Considering these parameters, Weibull distribution is expressed as below:

$$f(x) = \frac{\alpha}{\beta} \left(\frac{SPEI-6_{(t-1)} - \gamma}{\beta} \right)^{\alpha-1} \exp\left(-\left(\frac{SPEI-6_{(t-1)} - \gamma}{\beta}\right)^\alpha\right) \quad (15)$$

It is also evident from the table that the best distribution model varies among both stations and time scales. This means that if a distribution model is superior for a station, it is not necessarily an effective model for the adjacent station. Therefore, instead of using a single distribution model, a comparative analysis among different models is crucial when using WPGP at multi-station studies.

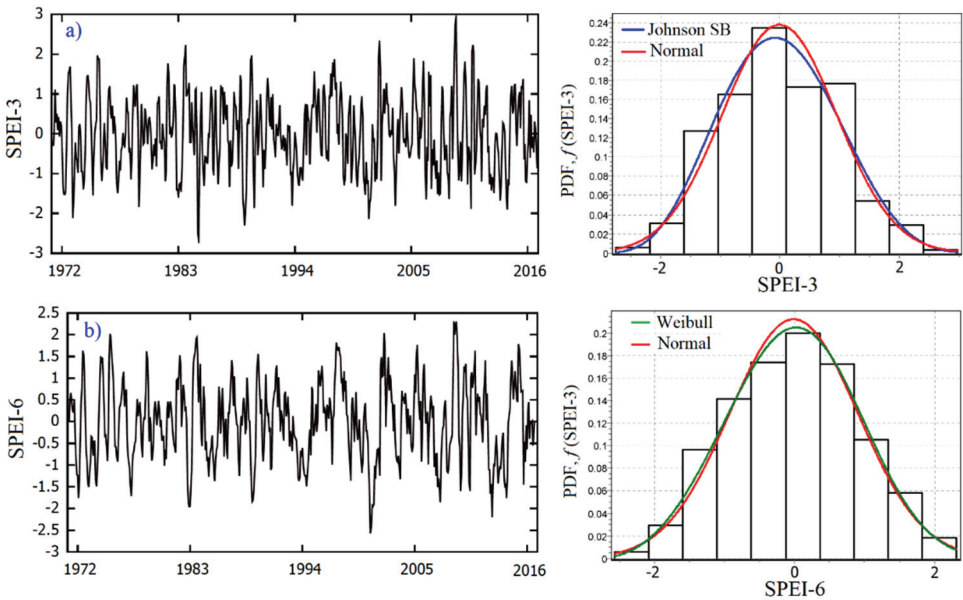


Figure 9 - Noise signals (left) and histogram with theoretical PDFs for noise signals of SPEI at Beypazari (a and b) and Nallihan (c and d) stations

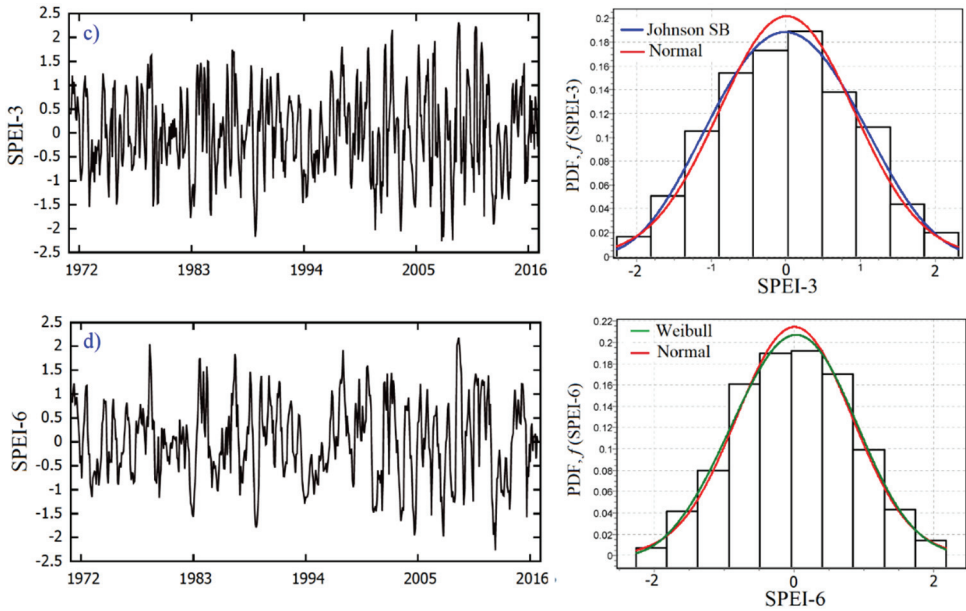


Figure 9 - Noise signals (left) and histogram with theoretical PDFs for noise signals of SPEI at Bey pazari (a and b) and Nallihan (c and d) stations (continue)

6. CONCLUSIONS

In this study, a novel hybrid data-driven model, namely WPGP, has been introduced for 1-month ahead hindcasting of meteorological drought. The model integrates WP denoising technique with symbolic regression ability of GP to model SPEI time series. Using historical SPEI series from two meteorological stations, efficiency of the proposed WPGP model was compared to those of three benchmarks: the standalone GP, RF, and AR1 models. The WPGP uses antecedent denoised SPEI signal instead of their original time series and provides explicit predictive models under parsimony pressure. To use the model for future forecasts, a PDF with the similar pattern of noise signals is added to the evolved models. Indeed, the WPGP is an evolutionary model that assimilates the capabilities of standalone GP and WP to yield a better solution through the elimination of noise and reconstruction of predictive signals and good noises. Based on the forecasting results, the proposed WPGP model was found to have a higher ability than its baseline GP as well as the benchmark RF and AR1. The proposed methodology resulted in the acquisition of a more robust model that yields significant improvement in terms of RMSE and NSE metrics. From a model developing view, it is worthy to remind that the WPGP is a hybrid but explicit approach which means a mathematical expression is provided for future predictions in each scenario. It must be noted that the methodology presented is general and can be applied to similar problems dealing with other drought indices. The present study was confined to 1-month ahead hindcasting scenarios which use past SPEI series as the only input information. For many implementations in water resources management, the future studies can comprise improving the offered model so that it is able to do forecasts with longer lead-times such as seasonal, or

even longer. To this end, implementation of exogenous inputs or ensemble precipitation and PET forecasts from ECMWF would be informative. Another suggestion for the future studies may be the use of frequency-based estimations to predict chaotic features of drought time series as those studies accomplished for streamflow series [42].

Symbols

γ	: shape parameters 0
ξ	: location parameter
λ	: scale parameter
ACF	: autocorrelation function
ANN	: artificial neural network
ANFIS	: adaptive neuro-fuzzy inference system
db	: Debauches
DAI	: drought area index
DT	: decision trees
GP	: genetic programming
LSSVM	: least square support vector machine
MARS	: multivariate adaptive regression splines
PACF	: partial autocorrelation function
PDF	: probability distribution function
PDSI	: Palmer drought severity index
RF	: random forest
SPEI	: standardized precipitation evapotranspiration index
SPI	: standardized precipitation index
SVR	: support regression vector
WP	: wavelet packets
WPGP	: wavelet packet-genetic programming

References

- [1] McKee, T.B., Doesken, N.J., and Kleist, J., (1993). The relationship of drought frequency and duration to time scales. In Proceedings of the International 8th Conference on Applied Climatology. American Meteorological Society, Anaheim, CA, USA, 17–22 January. pp. 179–184.

- [2] Bhalme, H.N., and Mooley, D. A. (1980). Large-scale droughts/floods and monsoon circulation. *Monthly Weather Review*, 108(8), 1197-1211.
- [3] Palmer, W.C. (1965). *Meteorological Drought*, Weather Bureau Research Paper No. 45, U.S. Department of Commerce, Washington, D.C.
- [4] Vicente-Serrano, S.M., Beguería, S., and López-Moreno, J.I. (2010). A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7), 1696-1718.
- [5] Mishra, A. K., & Desai, V. R. (2005). Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment*, 19(5), 326-339.
- [6] Bacanlı, U. G., Firat, M., & Dikbas, F. (2009). Adaptive neuro-fuzzy inference system for drought forecasting. *Stochastic Environmental Research and Risk Assessment*, 23(8), 1143-1154.
- [7] Keskin, M. E., Terzi, O., Taylan, E. D., & Küçükyaman, D. (2009). Meteorological drought analysis using data-driven models for the Lakes District, Turkey. *Hydrological sciences journal*, 54(6), 1114-1124.
- [8] Durdu, Ö. F. (2010). Application of linear stochastic models for drought forecasting in the Büyük Menderes river basin, western Turkey. *Stochastic Environmental Research and Risk Assessment*, 24(8), 1145-1162.
- [9] Özger, M., Mishra, A. K., & Singh, V. P. (2012). Long lead time drought forecasting using a wavelet and fuzzy logic combination model: A case study in Texas. *Journal of Hydrometeorology*, 13(1), 284-297.
- [10] Danandeh Mehr, A., Kahya, E., & Özger, M. (2014). A gene-wavelet model for long lead time drought forecasting. *Journal of Hydrology*, 517, 691-699.
- [11] Bazrafshan, O., Salajegheh, A., Bazrafshan, J., Mahdavi, M., & Fatehi Maraj, A. (2015). Hydrological drought forecasting using ARIMA models (case study: Karkheh Basin). *Ecopersia*, 3(3), 1099-1117.
- [12] Karavitis, C. A., Vasilakou, C. G., Tsesmelis, D. E., Oikonomou, P. D., Skondras, N. A., Stamatakos, D., ... & Alexandris, S. (2015). Short-term drought forecasting combining stochastic and geo-statistical approaches. *European Water*, 49, 43-63.
- [13] Deo, R. C., Tiwari, M. K., Adamowski, J. F., & Quilty, J. M. (2017). Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stochastic environmental research and risk assessment*, 31(5), 1211-1240.
- [14] Katip, A. (2018). Meteorological Drought Analysis Using Artificial Neural Networks for Bursa City, Turkey. *Applied Ecology and Environmental Research*, 16(3), 3315-3332.
- [15] Morid, S., Smakhtin, V., & Bagherzadeh, K. (2007). Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(15), 2103-2111.

- [16] Barua, S., Ng, A. W. M., & Perera, B. J. C. (2012). Artificial neural network–based drought forecasting using a nonlinear aggregated drought index. *Journal of Hydrologic Engineering*, 17(12), 1408-1413.
- [17] 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), 729.
- [18] Labat, D. (2005). Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology*, 314(1-4), 275-288.
- [19] Nourani, V., Baghanam, A. H., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. *Journal of Hydrology*, 514, 358-377.
- [20] Kim, T. W., & Valdés, J. B. (2003). Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal of Hydrologic Engineering*, 8(6), 319-328.
- [21] Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508, 418-429.
- [22] Maity, R., & Suman, M. (2019). Predictability of Hydrological Systems Using the Wavelet Transformation: Application to Drought Prediction. In *Hydrology in a Changing World* (pp. 109-137). Springer, Cham.
- [23] Gyamfi, C., Amaning-Adjei, K., Anornu, G. K., Ndambuki, J. M., & Odai, S. N. (2019). Evolutional characteristics of hydro-meteorological drought studied using standardized indices and wavelet analysis. *Modeling Earth Systems and Environment*, 5(2), 455-469.
- [24] Soh, Y. W., Koo, C. H., Huang, Y. F., & Fung, K. F. (2018). Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia. *Computers and electronics in agriculture*, 144, 164-173.
- [25] Ahmadalipour, A., Moradkhani, H., and Demirel, M.C. (2017). A comparative assessment of projected meteorological and hydrological droughts: Elucidating the role of temperature. *Journal of Hydrology*, 553, 785-797.
- [26] Danandeh Mehr, A., Sorman, A. U., Kahya, E., & Hesami Afshar, M. (2019). Climate change impacts on meteorological drought using SPI and SPEI: case study of Ankara, Turkey. *Hydrological Sciences Journal*, DOI: 10.1080/02626667.2019.1691218
- [27] Meresa, H. K., Osuch, M., and Romanowicz, R. (2016). Hydro-meteorological drought projections into the 21-st century for selected Polish catchments. *Water*, 8(5), 206.
- [28] Khan, M., Muhammad, N., & El-Shafie, A. (2018). Wavelet-ANN versus ANN-based model for hydrometeorological drought forecasting. *Water*, 10(8), 998.
- [29] Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.

- [30] Chen, J., Li, M., & Wang, W. (2012). Statistical uncertainty estimation using random forests and its application to drought forecast. *Mathematical Problems in Engineering*, 2012.
- [31] Yu, P.S., Yang, T.C., Chen, S.Y., Kuo, C.M., Tseng, H.W., 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of hydrology* 552, 92-104.
- [32] Zhao, W., Sánchez, N., Lu, H., Li, A., (2018). A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *Journal of hydrology* 563, 1009-1024.
- [33] Sadler, J.M., Goodall, J.L., Morsy, M.M., Spencer, K., (2018). Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. *Journal of hydrology* 559, 43-55.
- [34] Şarlak, N., & Güven, A. (2016). Global güneş radyasyon tahmini: Gaziantep uygulaması. *Teknik Dergi*, 27(3), 7561-7568.
- [35] Danandeh Mehr, A., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A. M., & Yaseen, Z. M. (2018). Genetic programming in water resources engineering: A state-of-the-art review. *Journal of hydrology* 566, 643-667.
- [36] Hu, J., Liu, B., & Peng, S. (2019). Forecasting salinity time series using RF and ELM approaches coupled with decomposition techniques. *Stochastic Environmental Research and Risk Assessment*, 1-19.
- [37] Coifman, R.R.; M.V. Wickerhauser, (1992), "Entropy-based algorithms for best basis selection," *IEEE Trans. on Inf. Theory*, vol. 38, 2, pp. 713–718
- [38] Hrnjica, B., & Danandeh Mehr, A. (2018). *Optimized Genetic Programming Applications: Emerging Research and Opportunities: Emerging Research and Opportunities*. IGI Global. DOI: 10.4018/978-1-5225-6005-0
- [39] Rahmani-Rezaeieh, A., Mohammadi, M., & Danandeh Mehr, A. (2019). Ensemble gene expression programming: a new approach for evolution of parsimonious streamflow forecasting model. *Theoretical and Applied Climatology*, 1-16.
- [40] Danandeh Mehr, A., & Safari, M. J. S. (2020). Multiple genetic programming: a new approach to improve genetic-based month ahead rainfall forecasts. *Environmental Monitoring and Assessment*, 192(1), 25.
- [41] Danandeh Mehr, A., Kahya, E., & Olyaie, E. (2013). Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *Journal of Hydrology*, 505, 240-249.
- [42] Dikbaş, F. (2016). Büyük Menderes Akımlarının Frekans Tabanlı Tahmini. *Teknik Dergi*, 27(1), 7325-7343.