

Investigation of denoising effects on forecasting models by statistical and nonlinear dynamic analysis

Farhang Rahmani and Mohammad Hadi Fattahi

ABSTRACT

In this study, the denoising effect on the performance of prediction models is evaluated. The 13-year daily data (2002–2015) of hydrological time series for the sub-basin of Parishan Lake of the Helle Basin in Iran were used to predict time series. At first, based on observational precipitation and temperature data, the prediction was performed, using the ARIMA, ANN-MLP, RBF, QES, and GP prediction models (the first scenario). Next, time series were denoised using the wavelet transform method, and then the prediction was made based on the denoised time series (the second scenario). To investigate the performance of the models in the first and second scenarios, nonlinear dynamic and statistical analysis, as well as chaos theory, was used. Finally, the analysis results of the second scenario were compared with those of the first scenario. The comparison revealed that denoising had a positive impact on the performance of all the models. However, it had the least influence on the GP model. In the time series produced by all the models, the error rate, embedding dimension needed to describe the attractors in dynamical systems and entropy decreased, and the correlation and autocorrelation increased.

Key words | artificial neural network, autoregressive integrated moving average, chaos, grid partitioning ANFIS, quadratic exponential smoothing, radial basis function

Farhang Rahmani
Mohammad Hadi Fattahi (corresponding author)
Department of Civil Engineering,
Marvdasht Branch Islamic Azad University,
Marvdasht,
Iran
E-mail: fattahi.mh@gmail.com

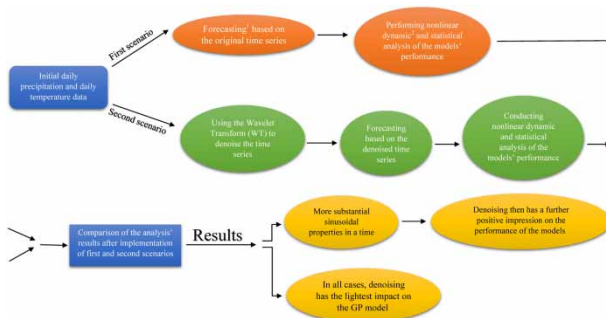
HIGHLIGHTS

- Conducting nonlinear dynamic and statistical analyses, as well as a chaotic analysis of the performance of the models.
- Performing nonlinear dynamic and chaotic analyses of denoising influences on the performance of ARIMA, QES, GP, RBF, and ANN-MLP
- Carrying out a statistical analysis of denoising impacts on the performance of forecasting models and comparing the results

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/wcc.2020.014

GRAPHICAL ABSTRACT



INTRODUCTION

Today, different methods are used to predict a time series. Each prediction model has a set of positive features and a number of weaknesses and can also be used to generate a specific type of time series. For example, the ANN-based prediction model produces a time series with nonlinear properties. All of these models use historical data to predict the future with a certain quality. Prediction models, such as autoregressive integrated moving average (ARIMA), artificial neural network (ANN), grid partitioning (GP) artificial neural-fuzzy inference system (ANFIS), and radial basis function (RBF) are used in forecasting hydrological time series, the results of which indicate that the predicted time series are sufficiently accurate and are suitable for forecasting hydrological time series (Bloomfield 1992; Khotanzad *et al.* 1996; Hayati & Mohebi 2007; Awad *et al.* 2009; Castellanos & James 2009; Xia *et al.* 2010; Babu & Reddy 2012; Rezaeian *et al.* 2012; Darji *et al.* 2015; Fattahi 2016; Al-Mukhtar 2019; Hussain *et al.* 2019; Nadiri *et al.* 2019; Niromandfard *et al.* 2019; Sanikhani *et al.* 2019).

Time series are not always definite, and in hydrology, in particular, they are considered as stable random series. If a time series is considered as a definite time series with colored or white noise, then the time series can be predicted. A better model for the time series can be determined when the noise in a time series is minimized very accurately (by a process called denoising). There are two common and convenient tools for denoising: Fourier transform (FT) and wavelet transform (WT). Due to the lower limitations and better performance of WT, in this study, this model was used for denoising. The WT function can be described as a

time series decomposed by WT into several time series with different scales (Rioul & Vetterli 1991). Using WT, the original time series can be examined with varying accuracy. WT can, therefore, be considered a Multiresolution Analysis (Alrumaih *et al.* 2002). In other words, signal decomposition causes the signal to change from the time domain to the time-scale domain, which can describe local features in the time and frequency domain very well and in detail (Guo *et al.* 2000).

In this study, prediction models including quadratic exponential smoothing (QES), autoregressive integrated moving average (ARIMA), artificial neural network–multi-layer perceptron (ANN-MLP), RBF based on neural network (RBF), and grid partitioning ANFIS (GP) were used to predict the time series of daily precipitation, daily maximum and minimum temperatures. The inherent characteristics of prediction models have had a significant impact on the selection of models for comparison in this study (e.g., the input information for the QES model for training is less than other models, and less time is needed to model (Duan & Niu 2018).

In this study, we investigate the effect of denoising on the performance of prediction models, using nonlinear statistical and dynamic analysis, as well as chaos theory. In statistical analysis, R^2 and RMSE were used to estimate the correlation and error of the predicted time series, respectively. In the nonlinear dynamic analysis and the use of chaos theory, the used tools were Approximate Entropy (ApEn) for evaluating the order and predictability of fluctuations in time series, Correlation Dimension for determining the degree of complexity of a nonlinear dynamical system with the help of determining

the embedding dimension needed, the Autocorrelation for evaluating the periodicity and the irregularity in the time series, and finally, the Surrogate Data, Phase Space Reconstruction and Method of Delays were used to investigate the nature of the irregularity in the time series. The steps are as follows: In the first scenario, the time series of daily precipitation, daily maximum, and minimum temperatures are predicted. Then in the second scenario, the predicted time series are denoised with WT and again predicted using prediction models. In the next step, the predicted time series in the first and second scenarios are subjected to non-linear, statistical, and chaotic dynamical analysis. Finally, by comparing the analysis results of the second scenario with those of the first scenario, it becomes clear whether prediction based on the denoised time series improves the performance of the prediction models.

In the following, in the Material and methods section, theoretically, analytical and predictive methods are discussed, and the results of the first scenario are presented; in the Results and discussion section, the results from denoising are compared with those of the first scenario. Finally, in the Conclusions section, the final result of the research will be stated.

MATERIALS AND METHODS

Area study data

The data utilized in this study relate to Parishan Lake. Parishan wetland permanent freshwater lake is fed by

springs and seasonal streams located in Iran, Fars province, 12 km southeast of Kazeroun between the mountains of Famur. The annual precipitation average of this basin is 450 mm and the annual temperature average is 22.2 °C and it also has to be considered that the annual evaporation average is 2,470 mm. This region is located between 51°25'25 and 51°43'50 longitude and 29°22'25 and 29°27'40 latitude. It is considered as a part of the hydrological Shapur–Dalakey river basin. Sub-basins of Parishan Lake and Shapour River are both located in the Hele basin. The location of the Parishan lake sub-basin located in the Hele basin is shown in Figure 1.

In the present study, daily precipitation and daily maximum and minimum temperatures data have been applied to be forecasted by prediction models. All data relate to 2002–2015 and daily time series were forecasted for 13 years.

In this section about prediction methods, statistical and nonlinear dynamic analysis, and then WT are going to be discussed. The nonlinear dynamic analysis was calculated by MATLAB codes provided by Kugiuntzis & Tsimpiris (2010).

Artificial neural network multi-layer perceptron (ANN-MLP)

Artificial neural networks (ANN) are intelligent systems that are able to discover patterns with a few prior assumptions and also learn a complex functional relationship from the data to model a phenomenon. ANN was inspired by the principles of data processing in the brain and the method based on experimental data utilized to train the ANN to precisely predict system performance. The most common

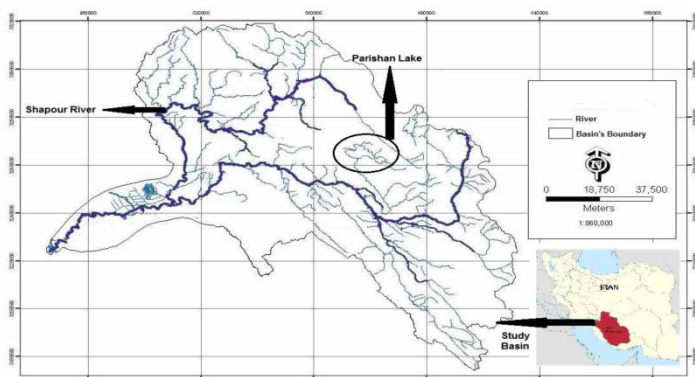


Figure 1 | Study basin of Hele and Parishan Lake sub-basin.

algorithm in ANN is multi-layer perceptron (MLP). MLP is a network with feed-forward layers consisting of one input and output layer, and a number of hidden layers. In the first step, each node includes input and computes a weighted sum. In the next step, it passes the sum through a soft non-linearity. An ANN output can be written as follows (Paoli *et al.* 2010):

$$y_i = f(x_j; w_{ij}) \quad (1)$$

where i is the index of neurons in the hidden layer, j is the index of an input to the neural network, f is the transfer function, x_j is the input, w_{ij} is the weight.

In this study, ANN-MLP was utilized to forecast minimum and maximum temperature time series which basic architecture for MLP is shown in Figure 2.

As can be noted, it is composed of a single output unit, k hidden, and n input units. T_j represents the connecting weight from the j th hidden unit to the output unit. In the same way, w_{ij} is the connection weight from the i th input unit to the j th hidden unit. The hidden layer consists of nodes connecting to both the input and output layer that can be considered as the most substantial part of a network. The set is trained to the neural network to be constructed, and the test set is also used to measure the predictive error of the model. Ultimately, the training process is utilized to find the connection weights of the networks (Pao 2007). The measured value is an anomaly if the predicted error exceeds the pre-defined threshold. To select the most acceptable model of the neural network, numbers of modeling were assessed and the model in the following detail was

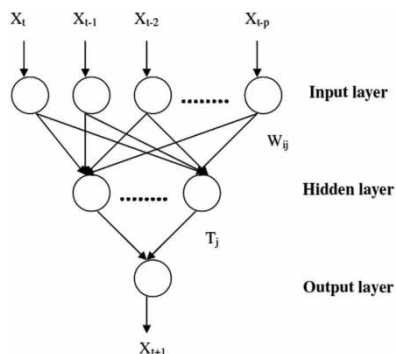


Figure 2 | An MLP feeds forward neural network for time series prediction.

preferred. The model consisted of 12 hidden layers, and the training ratio, validation ratio, and test ratio were 80, 10, and 10% respectively. The train parameter time and train parameter maximum fail were 2,000 and 1,000, respectively.

Radial basis function

In 1977–1988, the RBF ANN model was developed by Powell (1977), and Broomhead & Lowe (1988). This model consists of a general framework of an input layer with signal nodes, a hidden layer with RBF neurons which is the Gauss function, and an output layer with linear nature (see Figure 3).

There is a similarity between the number of input and output nodes of MLP and RBF neural networks. The number of input and output nodes is determined by inherent features of input and output variables and RBF networks faster learning is one of the superiorities of RBF over MLP. The output of RBF NN is calculated as:

$$Y = \sum_{r=1}^r W_r \varphi(\|X - X_r\|) \quad (2)$$

X and Y are the input and output values, respectively. $\varphi(\cdot)$ is the RBF and the hidden and output nodes are connected by weight (W). Depending on the observed input data, the center of each hidden node can be represented by X_r . Finally, $\|X - X_r\|$ represents the Euclidean distance between input and hidden nodes. A group of input nodes is represented by each hidden node containing comparable information from the input data (Rezaeian-Zadeh *et al.* 2012). The responses produced by the radial basis neurons are in a small area and if a large area is considered as an input space the radial basis neurons have to be increased to adjust the network with the necessary precision (Xia *et al.* 2010).

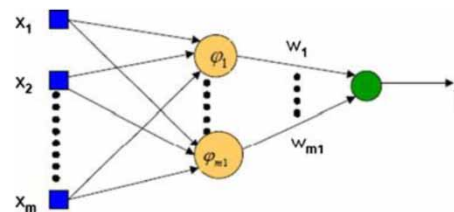


Figure 3 | Schematic diagram of the RBF ANN model.

Fuzzy inference system, GP of the antecedent variables

The artificial neural-fuzzy inference system (ANFIS) structure has two educable datasets: initial membership function parameters and polynomial parameters (result). The ANFIS structure uses a gradient descent algorithm to optimize the initial parameters and a least-squares algorithm to solve the result parameters. Each rule in the ANFIS structure is as follows:

$$\begin{aligned} \text{If } X_1 \text{ is } A_{1,j} \text{ and } X_2 \text{ is } A_{2,j} \text{ and } \dots \text{ and } X_n \text{ is } A_{n,j} \text{ then } y \\ = C_0 + C_1X_1 + \dots + C_nX_n \end{aligned} \quad (3)$$

In Equation (3), $A_{i,j}$ is the j th linguistic expression of the i th input variable of X_i . n is the number of inputs and y is the output of the model and C_i , are the output parameters, which are specified in the training process. Since every rule has a definite output (versus Fuzzy output), the total output is obtained by the weighted average. In ANFIS, sequential layers are assigned to various tasks such as creating a gradual model refinement process. The learning process consists of a forward pass and a backward pass process.

During the forward pass process, the initial parameters are fixed, and the result parameters are optimized using the least-squares algorithm. The backward pass process uses the gradient descent algorithm to modify the initial parameters of the membership functions for the input variables. The output is calculated as a weighted average of the result parameters. Each output error uses the back-propagation algorithm to adjust the foreground parameters (Guan *et al.* 2008). Figure 3 shows the structure of the adaptive neuro-fuzzy inference system. In this system, circular nodes represent fixed nodes, and square nodes represent nodes in which parameters can learn.

The ANFIS system includes five layers (Figure 4) that the first step of the process is conducted on the input by the first layer. Each node in this layer is:

$$O_{1,i} = \mu_{B_{i-2}}(y) \text{ for } i = 3, 4 \quad O_{1,i} = \mu_{A_i}(x) \text{ for } i = 1, 2 \quad (4)$$

$O_{1,i}(x)$ is necessarily a membership degree for x and y .

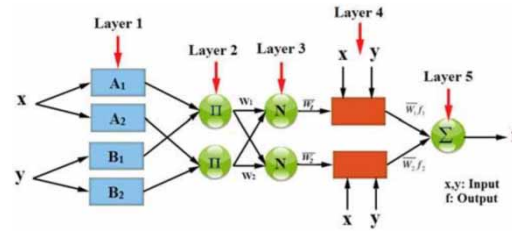


Figure 4 | Schematic diagram of the ANFIS system (Jang 1993).

Layer 5 is representing a sum of the outputs of the previous stage (layer 4). (Equation (5))

$$\sum_i \bar{W}_i f_i = \frac{\sum W_i f_i}{\sum W_i} \quad (5)$$

In Equation (5), $\bar{W}_i f_i$ is the i th node output extracted from the previous layer. A detailed description of the function of all layers is provided in the supplement file. In grid partitioning of the antecedent variables approach, every antecedent variable is suggested by independent partitions (Jang 1993). The rules section should specify the number of MFs, and the higher the number of MFs, the more efficient the ANFIS system is in describing a complex system and producing better results (Negnevitsky *et al.* 2004). The large number of inputs or MFs in the rule's premise generates a large number of Fuzzy rules. It creates a high exponential complexity in the training of the ANFIS system, which slows down or destroys the system performance, and this phenomenon is a curse of dimensionality (Bishop 1995; Jang *et al.* 1997). Based on the experience (Jang 1996; Jang *et al.* 1997), the practical number of MFs per input producing overlapping while not invoking the curse of dimensionality is two.

It has to be mentioned that the MATLAB codes to forecast time series utilizing ANFIS are provided in the link <https://yarpiz.com/327/ypfz102-time-series-prediction-using-anfis>.

Autoregressive integrated moving average

In statistics and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better interpret the data or to predict

future points in the series (forecasting). The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error represents a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for 'integrated') indicates that the data values have been replaced with the difference between values and the previous values. The purpose is to generate the model to adjust the data as well as possible. Non-seasonal ARIMA models are generally denoted ARIMA(p, d, q) where parameters p , d , and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving average model. ARIMA generalized form is defined as follows (Ong *et al.* 2005):

$$\gamma(B) \cdot \mu(B^s) \cdot (1 - B)^d \cdot (1 - B)^D \cdot Y_t = \rho(B) \cdot \sigma(B^s) \cdot Z_t \quad (6)$$

where B is the backward shift operator, d is the non-seasonal order of differences, D is a seasonal order of differences, γ , μ , ρ , σ are polynomials in B and B^s .

It has to be mentioned that there is a basic hypothesis that time series data incorporate statistical stationarity implying that measured statistical properties, such as the mean, variance, and autocorrelation remain constant over time (Yuan *et al.* 2016). Although if the training data display non-stationarity, the differenced data are required for the ARIMA model to be transformed to stationarity. This is denoted as ARIMA (p, d, q) where d obtains the degree of differencing (Rosca 2011).

Previous to the analysis, since both the mean and variance of the daily time series were considered to be non-stationary, a first-ordered difference was applied to obtain stationary set data.

QES model

The exponential smoothing model is a kind of weighted average model. The exponential smoothing model has a sensitivity to historical data closing to the forecasting period and also the weight of the model is reduced by time (Taylor 2003; Corberan-Vallet *et al.* 2011). The QES model

is based on employing the single exponential smoothing model and simultaneously considering the overall trend of the curve (Olaofe 2015). To utilize the QES model for forecasting, three sets of input training data and a parameter value are required and the quadratic exponential smoothing model can be defined as follows:

$$F_{t+T} = d_t + b_t T \rightarrow b_t = \frac{a(Q'_t - Q'_t)}{1 - a} \quad (7)$$

Q_t represents the smoothing values at time t . a is the smoothing coefficient (or damping coefficient (Spelta *et al.* 2011)). T represents predictive periods and F_{t+T} is the predicted value of the time series at the time $t + T$, and d_t and b_t are the data parameters at the time T . The smoothing parameter is considered a constant between 0 and 1. a has two modes. First, when a is close to 1, the predicted values approach the current value and far from the smoothness value of the time series data. Second, when a is close to zero, the predicted values approach the smoothness value of the time series data.

Correlation dimension

Concerning the review of a deterministic or random process, the correlation dimension is presumed to be a practical tool. Although a random (a complete stochastic) process might utilize all available dimensions of the phase space, a deterministic process might consume a fraction of the dimension (it is usually much smaller than the degree of freedom of the system). The dimension is used as a measure of the complexity of a system and also as an indicator of the variables required to describe a specific system. Correlation dimension as a tool based on the phase space can be defined as an estimation of the relative frequency with which the attractor visits each covering element. As a rule, concerning the correlation dimension and fractal dimension, it is proven that fractal dimension is always bigger or equal to the correlation dimensions. For a chaotic system, fractal and correlation dimensions are approximately equal (Tsonis 1992). To deal with the correlation dimension, the trajectories in a sufficient space should be unfolded. To illustrate more about the correlation dimension, consider a set of random points on the real number line between 0

and 1. The correlation dimension of the former equals one ($d = 1$), while for a triangle embedded in three-dimensional space (or m -dimensional space), the correlation dimension will be two ($d = 2$).

If a time series $u(t)$, $t = 1, 2, \dots, N$ is supposed, it is possible to reconstruct the series with embedding dimension m by selecting a proper time delay. In the correlation integral, d represents the dimension of the space and the relationship between correlation integral and d is $C(r) \propto r^d$.

The correlation dimension d is estimated as follows by taking a logarithm to $C(r)$. Correlation dimension is shown as follows:

$$d = \lim_{r \rightarrow 0} \frac{\log C(r)}{\log(r)} \quad (8)$$

The correlation integral $C(r)$ is used by correlation dimension d to measure how many times a trajectory in the phase space occurs in a distance r of a specific point. For a bounded data set, the correlation dimension is known as the correlation exponent. If the number of points is sufficiently large, and evenly distributed, a log-log graph of the correlation integral versus r will yield an estimate of d , corresponding to a given range of embedding dimension m (in this study $m = 1, 2, 3, \dots, 20$) and radius r (e.g., $r = 0.01, 0.1, 1, \dots, 10,000$). These embedding dimensions have to be employed to test the minimum dimension required to explain a nonlinear dynamic system. A low-dimensionality and parsimonious system is desirable and larger m values are necessarily unevaluated. The assortment of radius must be sufficient to examine entire cases of differences between $u(i)$ and $u(j)$ state vectors. A highly desirable correlation dimension in a low-dimensionality chaotic system has two properties. The first characteristic is that the correlation dimension must be consistent for various values of the radius. The second feature is that the correlation dimension must be bounded by all examined embedding dimensions. Complexity in a system represents randomness and bigger values of m demonstrate complexity; therefore, the system with lower m is preferable. A system with minimal values of m is deterministic and chaos concomitant of certainty. Accordingly, in this study, lower m is investigated.

Table 1 shows that m values of daily precipitation of the ANN-MLP and RBF models equal approximately the observational data. It has results such as time series generated with the same outliers, the same degree of complexity in the system as well as the need for many variables to the model. But the ARIMA, GP, and QES models have a dynamic nature with low embedding dimension (m) and have a chaotic system. The m values of daily maximum and minimum temperatures show that all the models except the RBF model have a dynamic nature with m almost identical. However, the RBF model has a dynamic nature with more m (more complex system). The QES model is more dynamic in nature than other models and observational data. The m values of the first scenario are provided in Table 1. The results of daily precipitation demonstrated that the forecasting models based on neural networks (ANN-MLP and RBF) produced more complex systems (time series) than the observed data; however, ARIMA and QES suggested less complex systems than the observed data. Besides, the results of minimum and maximum daily temperatures indicated that all models had the same acceptable performance, except RBF (in the case of maximum daily temperatures).

Approximate entropy

To quantify the amount of regularity and the unpredictability of fluctuations in time series, approximate entropy (ApEn) can be utilized as an effective and practical technique. Moment statistics, such as mean and variance, are not able to distinguish between two series which a series (I) alternates 10 and 20 and a series (II) has either a value of 10 or 20, chosen randomly, each with probability $\frac{1}{2}$. Series (I) is perfectly regular and can be predicted with certainty, and series

Table 1 | m values in which the graphs of $\log r$ versus $\log C_m(r)$ are saturated

Time series	Forecasting models and observed data					
	Observed	ARIMA	ANN-MLP	RBF	QES	GP
Daily Pre ^a	16	5	17	17	6	5
Daily Max ^b	9	10	10	12	8	9
Daily Min ^c	10	11	11	11	10	10

^aPrecipitation; ^bmaximum temperature; ^cminimum temperature.

(II) is randomly valued which is unpredictable with acceptable certainty. ApEn was developed by Steve M. Pincus to rectify the defects and limitations by modifying an exact regularity statistic, Kolmogorov–Sinai entropy. The presence of repetitive patterns of fluctuation in a time series is a key factor leading to a more predictable time series than a time series in which such patterns are absent. ApEn demonstrates the prospect that identical patterns of observations will not be followed by additional resembling observations (Ho *et al.* 1997). A time series or process containing noticeable recurrent patterns has a relative minute ApEn; a less predictable process has a higher ApEn. There are two fixed input parameters, m and r , to ApEn be calculated, where m is an integer representing the length of the compared run of data and r is a positive real number specifying a filtering level.

Consider a time series of data $u(1), u(2), \dots, u(N)$ containing N raw data values from measurement equally spaced in time, and a sequence of vectors $x(1), x(2), \dots, x(N - m + 1)$ in R^m . $d[x(i), x(j)]$ is the distance between vectors $x(i)$ and $x(j)$ which is the maximum difference in the vectors' scalar components. The sequence $x(1), x(2), \dots, x(N - m + 1)$ is utilized for each $i \leq N - m + 1$, to $C_i^m(r)$ be constructed. $C_i^m(r)$ measures the regularity in consideration of r tolerance, or frequency of patterns identical to a given pattern of window length m . ApEn can be defined as Equation (9) (Pincus & Goldberger 1994):

$$\text{ApEn}(m, r) = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)] \quad (9)$$

where $\Phi^m(r)$ consists of N , m , and $C_i^m(r)$ elements (more detailed information is furnished in supplementary file).

From Table 2, we find that the ApEn values for all the models decrease after the daily precipitation forecast. It indicates an increase in predictability and an increase in the

iterative patterns in the time series. The highest reduction of ApEn is observed in the two models, ARIMA and GP, and the lowest decrease in ApEn is observed in the RBF and ANN-MLP models.

Both RBF and ANN-MLP models have the same performance in predicting the daily maximum and minimum temperature time series, and the ApEn values of both of these have a slight difference equal to the observational time series data. All the time series of the RBF and ANN-MLP models have significant alternation and repetitive patterns (such as the observational time series data). The ApEn values of the ARIMA, QES, and GP models are much less than the observed time series. It shows that the ARIMA, QES, and GP models present a more predictable time series than the RBF and ANN-MLP. It indicates, moreover, the frequency similarity, as well as the periodicity, are developed after prediction by the ARIMA, QES, and GP models. The results suggested that the performance of the ANN-MLP and RBF models was suitable, except that of QES. Besides, ARIMA and GP yielded the same results in the case of daily precipitation. However, QES produced inappropriate results (its results are significantly distinct from the observed time series).

Autocorrelation

Autocorrelation, also called serial correlation, is the correlation of a signal with a delayed copy of itself functioning as a delay. In other words, it is the similarity between observations as a function of the time lag between them. An analytical apparatus is provided by autocorrelation function for distinguishing between periodic and stochastic behavior of a particular signal. The autocorrelation is periodic when a signal is periodic, whereas in a stochastic signal, the

Table 2 | Approximate entropy values of daily precipitation, daily maximum, and daily minimum temperatures

Time series	Approximate entropy values of the forecasting models and observed data					
	Observed	ARIMA	ANN-MLP	RBF	QES	GP
Daily Pre ^a	0.1220	0.0001	0.1219	0.1219	0.0030	0.0001
Daily Max ^b	0.0547	0.0470	0.0547	0.0545	0.0420	0.0429
Daily Min ^c	0.0597	0.0517	0.0598	0.0597	0.0271	0.0453

^aPrecipitation; ^bmaximum temperature; ^cminimum temperature.

autocorrelation is irregular. The autocorrelation function is shown as follows:

$$c_j = \frac{1}{N} \sum_{i=1}^N u_i u_{i+j} \quad (10)$$

where to extend the times series beyond u_N , periodic boundary conditions have to be imposed (boundary conditions $\rightarrow u_{N+k} = u_k$). A simple quantitative measure of the linear correlation between data points is provided by the former function.

Figure 5 shows that, in forecasting the time series of daily precipitation, the ANN-MLP and RBF models have autocorrelation, which is approximately equal to observational time series. This indicates a significant correlation in the predicted time series (such as the observational time series) and also shows that the available spikes are equal to the observational time series. The graphs of the ARIMA and QES models confirm that the correlation and periodicity of the predicted time series are greater than those of the observational time series. The GP model does not have autocorrelation, and the spikes are not much larger than zero. The daily minimum and maximum temperature graphs (Figures 6 and 7) determine that the time series predicted by all the models are the same (have scanty differences), and are equivalent to the observational time

series. Moreover, they have an equal correlation and periodicity. The results revealed that ANN-MLP and RBF models had an acceptable performance and GP, ARIMA, and QES suggested poor performance in the cases of daily precipitation (the results are significantly different from observed time series).

Surrogate data

In 1992, the method of surrogate data in nonlinear time series analysis was introduced by Theiler *et al.* (1992). In the first step, random nondeterministic surrogate data sets are created that mean, variance, and power spectrum of the sets are identical to the experimental time series data. In the next step, the measured topological properties of the experimental time series (original time series) have to be compared with the measured topological properties of the surrogate data sets created in the prior step. If the topological properties of the experimental time series data (original time series) and the surrogate data sets are similar, then the experimental data set is not random noise.

To create a surrogate data set, the following steps should be followed. 1 – the FT of the experimental time series data is constructed; 2 – phases are randomized; 3 – FT is inverted. In other words, the experimental time series data $x(t_j)$, $j = 1, 2, \dots, N$ is inputted into a complex

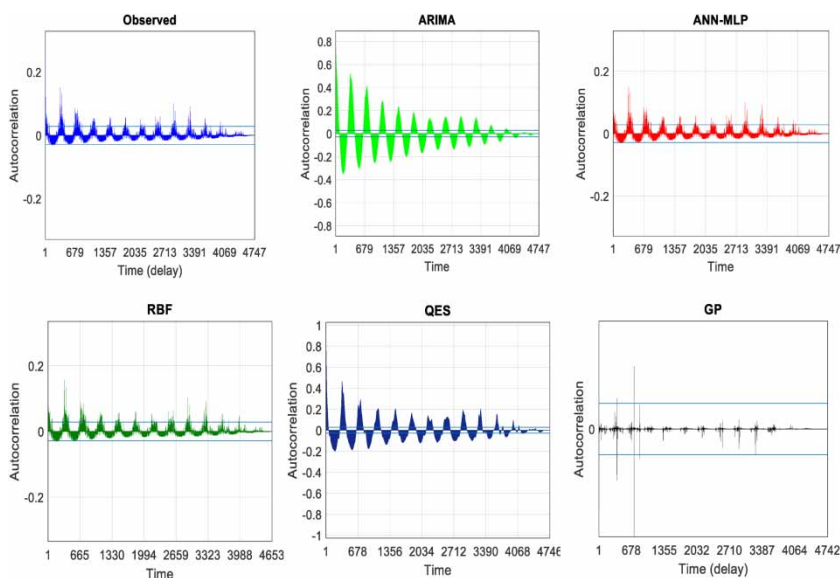


Figure 5 | Autocorrelation values of daily precipitation, observed, and forecasted time series.

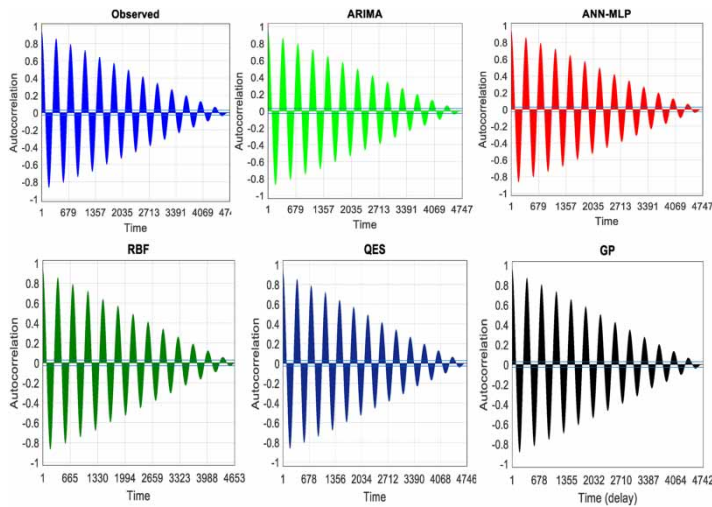


Figure 6 | Autocorrelation values of daily maximum temperatures, observed, and forecasted time series.

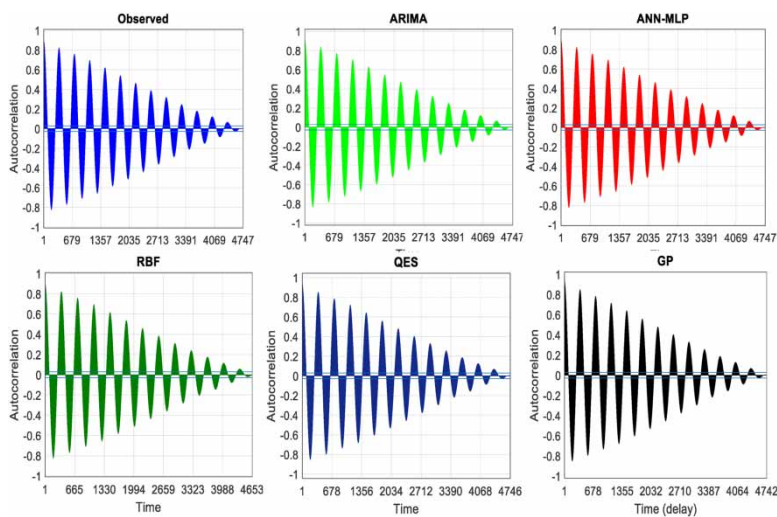


Figure 7 | Autocorrelation values of daily minimum temperatures, observed, and forecasted time series.

array (Mazaraki 1997). Then, the discrete FT is constructed and a set of random phases is constructed ($\varphi_m \in [0 \pi]$ $m = 2, 3, 4, \dots, \frac{N}{2}$), and the randomized phases are applied to the Fourier transformed data. Finally, the inverse FT can be produced by Equation (11):

$$z(n)' = x(n)' + iy(n)' = \frac{1}{N} \sum_{m=1}^N Z'_m e^{\frac{2\pi i(m-1)(n-1)}{N}} \quad (11)$$

where $Z(m)$ is the discrete FT (further details are given in supplementary file).

The daily precipitation phase space graph shows that the phase space graphs of all the models except the QES and GP models have stellar attractors, and the attractors of the QES and GP models are of a central type. The formation of trajectories in phase space indicates a degree of chaos in time series. The surrogate data graph exhibits that the irregularity in the predicted time series is identical to the observational time series, and has a small degree of dynamic properties. The phase space diagram of the time series and the surrogate data of the QES model proves that the chaos produced by this model is less (due to less phase space occupied by the trajectories), and in addition the irregularity in this time series

has a more dynamic nature compared to other models and observational time series. The graphs of the RBF and ANN-MLP have properties (attractors type and the formation of trajectories distribution) similar to the observed time series. Examination of the daily minimum and maximum temperatures graphs demonstrates that all the models except the QES model have the same performance with observational data (the differences are negligible), and the nature of irregularity in their time series is dynamically similar. The QES model has less chaos and irregularity with a more dynamic nature than other models and observational data. For example, the observational data phase space graph and the RBF prediction model for daily precipitation are presented in Figure 8, and other graphs are appended, due to their large number.

The results indicate poor performance of QES in the case of daily temperatures forecast. Besides, the ANN-MLP and RBF models have the most proper performance, respectively.

Statistical evaluation criteria

To evaluate the performance of prediction models, the coefficient of determination (R^2) and root mean square error (RMSE) were used. Equations (12) and (13) designate how these indices are calculated.

$$R^2 = \left(\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}} \right)^2 \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \tag{13}$$

where x_i is the observational data and y_i is the predicted data corresponding to the observational data. \bar{x} and \bar{y} are the

average observational and predicted data, respectively, and n is the number of data. The low RMSE value and the high R^2 coefficient indicate the higher precision of the model and its superiority to other models.

According to Table 3, for daily precipitation, the time series of all the models do not correlate with the observational time series, although the error rate is very low, and the lowest is concerning the RBF model. For daily maximum and minimum temperatures, all the models produce time series with a high correlation with observational time series. The highest correlation is related to the time series produced by the ANN-MLP and RBF models, and the lowest error for the maximum daily temperatures time series is for the ARIMA model, and concerning the minimum daily temperatures, it is for the RBF and ANN-MLP models, respectively. The most ineffective performance is associated with the GP model, which has the highest error and the slightest correlation with the observational time series. In general, ANN-MLP and RBF can be considered as the most reliable models, and in the next rank, the ARIMA and QES models are excellent in terms of statistical results, respectively. GP suggests poor performance compared with other models.

Table 3 | Statistical values of forecasted time series

		ARIMA	ANN-MLP	RBF	QES	GP
Pre	R^2	0.15	0.18	0.19	0.08	0.11
	RMSE	5.18	4.72	4.53	5.44	5.15
Max	R^2	0.96	0.99	0.99	0.94	0.91
	RMSE	2.08	3.56	5.90	2.42	3.06
Min	R^2	0.93	0.99	0.99	0.90	0.86
	RMSE	2.22	0.84	0.65	2.56	3.04

Pre, daily precipitation; Max, daily maximum temperatures; Min, daily minimum temperatures.

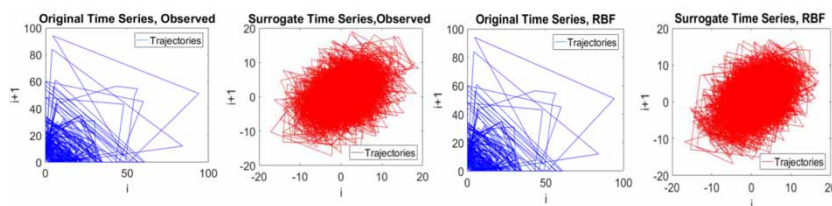


Figure 8 | Phase space reconstruction of original and surrogate daily precipitation, observed, and RBF model time series.

Denoising, WT

Wavelet coefficients correspond to details and additionally when details are small, they could be considered as noise, therefore they are dissimilar to the FT, omitted without compromising the sharp detail of the original series. In other words, without a formal mathematical point of view, most of the energy of the original series is compressed into a small number of large wavelet coefficients by WT. The thresholding excludes the noise while not eliminating the valuable series' properties. The purpose of wavelet denoising is to threshold the wavelet coefficients at every multiresolution level. The former is conducted to sift through details and delete inconsequential details (noise). Many conventional methods of noise reduction such as 'linear lowpass filtering' do not run properly for chaotic data since the signal and the noise often hold overlapping bandwidths. Most of the noise reduction methods are proposed for nonlinear time series (Billings & Lee 2003), and the number of noise reduction iterations and also embedding dimensions must be chosen by the user. Recently, a powerful approach for noise reduction has been established to utilize WT (Donoho & Johnstone 1994; Bahoura & Rouat 2001). It employs thresholds in the wavelet domain and this algorithm might also be utilized even when the dynamics are chaotic time series (Chen & Bui 2003; Azzalini & Schneider 2005).

A bedrock model for the noisy series can be specified as the following form:

$$S_t = f_t + e_t \quad t = 1, 2, 3, \dots, N \quad (14)$$

where e_t is the multivariate normal distribution process and the mean of zero and also covariance matrix Γ_t .

The object of the denoising is to suppress the noise part of the series S_t and in the next step recover f_t . The time series $s(t)$ could be decomposed as

$$s(t) = \sum_k cA_{K,k} \phi_{K,k}(t) + \sum_{j=1}^K \sum_k cD_{j,k} \Psi_{j,k}(t) \quad (15)$$

where K is the decomposition level, $\Phi_{K,k}(t)$ is the associated scaling function, and $\Psi_{j,k}(t)$ is generated from $\Psi(t)$. The

coefficient $cA_{K,k}$ is called the approximating coefficient and the coefficient $cD_{j,k}$ is called the detail coefficient (more details are provided in supplementary file). There are two standards of thresholding methods for denoising the series: 1 – soft thresholding; 2 – hard thresholding.

The mechanism of hard thresholding is that the wavelet coefficient is assumed zero if its amplitude is smaller than a pre-defined threshold; otherwise, it is kept unchanged. Additionally, if the absolute value of the wavelet coefficient is decreased to the threshold, we have soft thresholding. Notice that owing to the discontinuation of the hard-thresholding method at the threshold, additional oscillation might be imposed on the original series, even though the continuous soft-thresholding method can offer a smooth series. Consequently, the soft-thresholding method is utilized to reduce noise and outliers.

There are several approaches to choose the threshold. Three of the approaches are as follows.

Minimax thresholding. The minimax principle was found by the threshold. To allow minimax performance concerning mean square error against a proper procedure, a fixed threshold has to be accepted (Donoho & Johnstone 1993).

SURE thresholding. The threshold could be selected based on Stein's Unbiased Risk Estimate (SURE).

For more detail, refer to Johnstone & Silvenan (1997).

Universal thresholding. The threshold could be calculated by the following formula:

$$\lambda_j = \sigma_j \sqrt{\frac{2 \log n}{n}} \quad (16)$$

where σ_j is an estimate of the variation of wavelet coefficients of level j , and n is the number of wavelet coefficients at that level.

Subsequently, a denoised version of the original signal is reconstructed from the approximation coefficient $cA'_{K,k}$, and detail coefficients $cD'_{j,k}$, utilizing the inverse WT.

RESULTS AND DISCUSSION

A random signal produced by noise is completely separate from a random signal produced by a deterministic dynamical system. The signal generated on the basis of a deterministic dynamical system is explained by a scanty number of variables and is also of moderate complexity. A signal may appear arbitrary, but it can be justified through dynamic systems with a few degrees of freedom. Statistical analysis is appropriate to determine the median, mean, variance, standard deviation, and linear properties of a time series without expressing nonlinear time series properties. Statistical analysis cannot be employed to investigate the discrepancy between the random signal with the noise and the random signal produced by a deterministic dynamic system, as it does not reveal the difference and requires dynamic analysis, which is based on the phase space reconstruction. In the first scenario, the prediction is made based on observational data, and the following steps are executed for dynamic analysis. We first discuss the complexity of the system and the embedding dimension needed to describe the dynamical system. Then the entropy, correlation, and periodicity of the series are evaluated. Finally, the nature of irregularity is examined in the time series to determine its dynamic or random nature. Following dynamic analysis, statistical analysis is performed, and in the next step, the observational time series (daily precipitation, daily maximum and minimum temperatures) are denoised (second scenario), and the first scenario steps are repeated. The results of the first scenario are presented in the Materials and methods section, followed by the results of the second scenario and their comparison with the results of the first scenario.

Correlation dimension

Running the second scenario on the time series of daily precipitation reduces the numerical distance of m of the observational time series to the time series of the QES model, and also the ARIMA, RBF, and ANN-MLP models produce the time series with a complexity equal to the observational time series. Before the second scenario (in the case of daily minimum and maximum temperatures), the difference of m , where the graph is saturated, is high for the observational and predicted data (between 1 and 11 units

of difference), whereas, in the second scenario, the difference is very small (between 1 and 3 units). It is concluded that the models in the second scenario generate a time series with a dynamic characteristic very similar to the observational time series, and in the first scenario, the dynamic properties differ considerably between the observational and the predicted time series. The m values of the second scenario are provided in Table 4.

Applying denoising revealed signs of an enhancement in the performance of the models; however, it had a slight impression on GP outcomes.

To sum it up, the results of the chaotic analysis revealed that applying the denoising method improved the performance of all models and in a few cases, that of GP; however, it was verified that some models were less affected than others. Nevertheless, in almost all cases, it had a neutral effect on GP performance. ANN-MLP and RBF accepted more positive influences than others and yielded the most accurate performance and outcomes.

ApEn

Comparison between Tables 2 and 5 suggests that after running the second scenario, the ApEn values of the predicted

Table 4 | m values in which the graphs of Log r versus Log $C_m(r)$ are saturated (related to denoised time series)

Time series	Forecasting models and observed data					
	Observed	ARIMA	ANN-MLP	RBF	QES	GP
Daily Pre ^a	5	5	5	5	4	2
Daily Max ^b	6	6	6	5	6	6
Daily Min ^c	6	6	6	6	5	5

^aPrecipitation; ^bmaximum temperature; ^cminimum temperature.

Table 5 | Approximate entropy values of daily precipitation, daily maximum, and daily minimum temperatures of the denoised time series

Time series	Forecasting models and observed data					
	Observed	ARIMA	ANN-MLP	RBF	QES	GP
Daily Pre ^a	16	5	17	17	6	5
Daily Max ^b	9	10	10	12	8	9
Daily Min ^c	10	11	11	11	10	10

^aPrecipitation; ^bmaximum temperature; ^cminimum temperature.

time series of all the models are depreciated. Dominant impacts are observed on the performance of models QES and GP. The second scenario has a considerable effect on the accurate prediction of the model of QES (in the case of daily temperatures), while the influence on the other models is more limited. Generally speaking, denoising has a positive influence on all the models.

Autocorrelation

The results of the autocorrelation comparison of the first scenario with the second one indicate that denoising advanced the performance of all the models (the results resembled the observational time series). However, the difference between RBF results and observational data was more notable than that of other models (in the case of daily temperatures). The autocorrelation graphs of the second scenario are given in the appendix in Supplementary Materials.

Phase space reconstruction and surrogate data

Following the implementation of the second scenario, all models (in case of daily temperatures) produced promising results, and the outcomes became remarkably similar to those of the observational data. Furthermore, the performance of QES strongly developed, thereby suggesting a higher degree of dynamic property than the first scenario. The second scenario graphs are presented in the appendix. Generally speaking, denoising has a positive influence on all the models. In all cases, the performance of the RBF and ANN-MLP models was appropriate.

To sum it up, the results of the chaotic analysis revealed that applying the denoising method improved the performance of all models and in a few cases, that of GP; however, it was verified that some models were less affected than others. Nevertheless, in many cases, it had a neutral effect on GP performance. ANN-MLP and RBF accepted more positive influences than others and yielded the most accurate performance and outcomes.

Statistical comparison

After forecasting the time series of daily precipitation based on the denoised time series (second scenario), the

correlation of all the models increases, and the highest correlation extension is concerning the ARIMA model. The error rate of this model (ARIMA) is also lower than all other models and has approached zero, and all the models have a reduced error rate. However, the GP model has received the least impact. The R^2 and RMSE results for the time series of daily minimum and maximum temperatures confirm that the second scenario produces time series with high correlation and negligible error by all the models, and the most positive effect (an increase in correlation and an error reduction) is accepted by the ARIMA model. The second scenario has an imperceptibly positive impact on the GP model so that the correlation almost remains constant (compare Tables 3 and 6).

The statistical analysis revealed that denoising improved the performance of all forecasting models; however, the GP model accepted the least impact.

According to the results, the denoising theory exerted promising and positive effects on all models. Nevertheless, the GP model accepted the least impact. The studies of Papadimitriou & Bezerianos (1999), Mohanty *et al.* (2019), Frikha & Hamida (2012), Balabin *et al.* (2008), Hui & Xinxia (2010), Chen & He (2006), Zhu *et al.* (2018), and Yang *et al.* (2019) were consistent with the present results, which indicated positive impressions on ARIMA, RBF, and ANN-MLP. As far as the authors of this study are concerned, no study has been conducted on the impacts of denoising on the performance of QES and GP. Furthermore, results demonstrate a strong and close resemblance between RBF and ANN-MLP performance. Bounds & Liroyd (1990) proved the same results.

Table 6 | Statistics results of forecasting models, forecasted based on the denoised time series

		ARIMA	ANN-MLP	RFB	QES	GP
Pre	R^2	0.17	0.19	0.20	0.09	0.12
	RMSE	3.22	4.11	4.12	5.12	4.36
Max	R^2	0.99	0.99	0.99	0.95	0.91
	RMSE	0.75	2.95	2.95	1.75	1.95
Min	R^2	0.98	0.99	0.99	0.92	0.87
	RMSE	0.80	0.84	0.85	0.83	0.88

Pre, daily precipitation; Max, daily maximum temperatures; Min, daily minimum temperatures.

CONCLUSION

Statistical analysis, nonlinear dynamic analysis, and chaos theory were adopted to investigate the effect of denoising on the performance of prediction models. Positive impact on the prediction model means that the prediction model generates time series by nonlinear, statistical, and chaotic dynamic properties (including entropy, correlation dimension, autocorrelation, error, correlation, periodicity, and irregularity in time series) equal to or very close to the observational time series. However, the negative impact on the prediction model means that the time series is produced by the prediction model with nonlinear, statistical, and chaotic dynamic properties with a significant disagreement between the observational time series and the forecasted time series. The results of the nonlinear and chaotic dynamic analysis revealed that prediction based on the denoised time series of daily precipitation has a positive effect on all the models, particularly RBF and ANN-MLP models, although it has the most limited influence on the GP model. Forecasting based on the denoised time series of daily minimum and maximum temperatures has a positive effect on all the models. RBF and ANN-MLP models generate a time series with a dynamic and statistical characteristic very comparable to the observational time series. To sum up, with more substantial sinusoidal characteristics in a time series, denoising then has a further positive impression on the performance of all the models, of which the ANN-MLP and RBF performed most highly among the five models. GP accepted less positive influences than the four other models. Therefore, it is recommended that the time series be denoised before prediction depending on the intensity of the sinusoidal property of the time series, and then the prediction is performed.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Al-Mukhtar, M. 2019 [Random forest, support vector machine, and neural networks to modelling suspended sediment in Tigris](#)

- River-Baghdad. *Environmental Monitoring and Assessment* **191**. Article 673. <https://doi.org/10.1007/s10661-019-7821-5>
- Alrumaih, R. M., Farge, M. & Al-Fawzan, M. A. 2002 [Time series forecasting using wavelet denoising an application to Saudi Stock Index](#). *Journal King Saud University* **14** (2), 221–234.
- Awad, M., Pomares, H., Rojas, I., Salameh, O. & Hamdon, M. 2009 Prediction of time series using RBF neural networks: a new approach of clustering. *The International Arab Journal of Information Technology* **6** (2), 138–144.
- Azzalini, M. F. & Schneider, K. 2005 [Nonlinear wavelet thresholding: a recursive method to determine the optimal denoising threshold](#). *Journal of Applied and Computational Harmonic Analysis* **18** (2), 177–185.
- Babu, C. N. & Reddy, B. E. 2012 Predictive data mining on average global temperatures using variants of ARIMA models. In: *IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM-2012)*. pp. 256–260.
- Bahoura, M. & Rouat, J. 2001 [Wavelet speech enhancement based the Teager energy operator](#). *IEEE Signal Processing Letters* **8** (1), 10–12.
- Balabin, R. M., Safieva, R. Z. & Lomakina, E. I. 2008 [Wavelet neural network \(WNN\) approach for calibration model building based on gasoline near infrared \(NIR\) spectra](#). *Chemometrics and Intelligent Laboratory Systems* **93**, 58–62. doi:10.1016/j.chemolab.2008.04.003.
- Billings, S. A. & Lee, K. L. 2003 [A smoothing algorithm for nonlinear time series](#). *International Journal of Bifurcation and Chaos* **14** (3), 1037–1051.
- Bishop, C. M. 1995 *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, pp. 5–10.
- Bloomfield, P. 1992 [Trends in global temperatures](#). *Climate Change* **21**, 1–16.
- Bounds, D. G. & Lloyd, P. J. 1990 [A comparison of neural network and other pattern recognition approaches to the diagnosis of low back disorders](#). *Neural Network* **3**, 583–591.
- Broomhead, D. S. & Lowe, D. 1988 Multivariable functional interpolation and adaptive networks. *Complex Systems* **2**, 321–355.
- Castellanos, F. & James, N. 2009 Average hourly wind speed forecasting with ANFIS. In: *11th Americas Conference on Wind Engineering*, Puerto Rico.
- Chen, G. Y. & Bui, T. D. 2003 [Multiwavelets denoising using neighboring coefficients](#). *IEEE Signal Processing Letters* **10** (7), 211–214. doi: 10.1109/LSP.2003.811586.
- Chen, H. & He, S. 2006 [Identification of the hydraulic AGC system of cold rolling mill using RBF neural network](#). In: *6th World Congress on Intelligent Control and Automation*, Dalian, pp. 1383–1387. doi:10.1109/WCICA.2006.1712574.
- Corberan-Vallet, A., Bermudez, J. D. & Vercher, E. 2011 [Forecasting correlated time series with exponential smoothing models](#). *International Journal of Forecasting, Elsevier* **27** (2), 252–265.
- Darji, M. P., Dabhi, V. K. & Prajapati, H. B. 2015 [Precipitation forecasting using neural network: a survey](#). In: *International*

- Conference on Advances in Computer Engineering and Applications, Ghaziabad*. pp. 706–713, doi:10.1109/ICACEA.2015.7164782.
- Donoho, D. & Johnstone, I. 1993 [Adapting to unknown smoothing via wavelet shrinkage](#). *Journal of the American Statistical Association* **90** (432), 1200–1224. Available from: <https://www.jstor.org/stable/2291512>
- Donoho, D. & Johnstone, I. 1994 [Ideal spatial adaptation by wavelet shrinkage](#). *Biometrika* **81** (3), 425–455. Available from: <https://www.jstor.org/stable/2337118>
- Duan, G. & Niu, R. 2018 Lake area analysis using exponential smoothing model and long time-series landsat images in Wuhan, China. *Sustainability* **10** (149), 1–16.
- Fattahi, H. 2016 [Indirect estimation of deformation modulus of an in situ rock mass: an ANFIS model based on grid partitioning, fuzzy c-means clustering and subtractive clustering](#). *Geosciences Journal* **20** (5), 681–690.
- Frikha, M. & Hamida, A. B. 2012 [A comparative survey of ANN and hybrid HMM/ANN architectures for robust speech recognition](#). *American Journal of Intelligent Systems* **2** (1), 1–8. doi:10.5923/j.ajis.20120201.01.
- Guan, J., Zurada, J. & Levitan, A. S. 2008 An adaptive neuro-fuzzy inference system based approach to real estate property assessment. *Journal of Real Estate Research* **30**, 395–422.
- Guo, D. F., Zhu, W. H., Gao, Z. -M. & Zhang, J. Q. 2000 [A study of wavelet thresholding denoising, WCC 2000 – ICSP 2000](#). In: *2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000*, Beijing, China, Vol. 1, pp. 329–332. doi: 10.1109/ICOSP.2000.894502.
- Hayati, M. & Mohebi, Z. 2007 Application of artificial neural networks for temperatures forecasting. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering* **1** (4), 654–658. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.193.3451&rep=rep1&type=pdf>
- Ho, K. K., Moody, G. B., Peng, C. K., Mietus, J. E., Larson, M. G., Levy, D. & Goldberger, A. L. 1997 [Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics](#). *Circulation* **96** (3), 842–848. doi:10.1161/01.cir.96.3.842.
- Hui, S. & Xinxia, L. 2010 [Multi-scale RBF prediction model of runoff based on EMD method](#). In: *Third International Conference on Information and Computing*, Wuxi, pp. 296–299, doi: 10.1109/ICIC.2010.260.
- Hussain, A., Singh, J. K., Kumar, A. R. & Harne, K. R. 2019 [Precipitation-runoff modeling of Sutlej River Basin \(India\) using soft computing techniques](#). *International Journal of Agricultural and Environmental Information Systems (IJAEIS)* **10** (2), 1–20. doi:10.4018/IJAEIS.2019040101.
- Jang, J. S. R. 1993 ANFIS: Adaptive-network-based fuzzy inference system. In *IEEE Transactions on Systems, Man and Cybernetics*. Vol. 23, pp. 665–685.
- Jang, J. S. R. 1996 Input selection for ANFIS Learning. In: *Proceedings of the 5th IEEE International Conference on Fuzzy Systems*. Vol. 2, pp. 1493–1499. <http://dx.doi.org/10.1109/FUZZY.1996.552396>.
- Jang, J.-S. R., Sun, C. T. & Mizutani, E. 1997 *Neuro-Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligence*, 1st edn. Pearson.
- Johnstone, L. M. & Silvennan, B. W. 1997 [Wavelet thresholding estimators for data with correlated noise](#). *Journal of the Royal Statistical Society, Series B* **59** (2), 319–351. <https://doi.org/10.1111/1467-9868.00071>
- Khotanzad, A., Davis, M. H. & Abaye, A. 1996 [An artificial neural network hourly temperatures forecaster with applications in load forecasting](#). *IEEE Transactions on Power Systems* **11** (2), 870–876. doi: 10.1109/59.496168.
- Kugiumtzis, D. & Tsimpiris, A. 2010 [Measures of analysis of time series \(MATS\): a matlab toolkit for computation of multiple measures on time series data bases](#). *Journal of Statistical Software* **33** (5), 1–30.
- Mazaraki, J. 1997 *Dynamical Methods for Analysing and Forecasting Chaotic Data*. Honours Thesis, Applied Mathematics, University of New South Wales.
- Mohanty, K. K., Barik, P. K., Barik, R. C. & Bhuyan, K. C. 2019 [An efficient prediction of diabetic from retinopathy using machine learning and signal processing approach](#). In: *International Conference on Information Technology (ICIT)*, Bhubaneswar, India, pp. 103–108. doi:10.1109/ICIT48102.2019.00025.
- Nadiri, A. A., Naderi, K., Khatibi, R. & Gharekhani, M. 2019 [Modelling groundwater level variations by learning from multiple models using fuzzy logic](#). *Hydrological Sciences Journal* **64** (2), 210–226. doi:10.1080/02626667.2018.1554940.
- Negnevitsky, M., Potter, C. W. & Ringrose, M. 2004 Short term wind forecasting techniques for power generation. In: *Australasian Universities Power Engineering Conference*.
- Niromandfard, F., KhasheiSiuki, A. & Shahidi, A. 2019 Evaluation of the neuro-fuzzy and hybrid wavelet-neural models efficiency in river flow forecasting (case study: Mohammad Abad Watershed). *Journal of Waste Management and Reprocessing*. **10** (19), 211–221. Available from: <http://jwmr.sanru.ac.ir/article-1-931-en.html>
- Olaofe, Z. O. 2015 [Wind energy predictions of small-scale turbine output using exponential smoothing and feed-forward neural network](#). *International Journal of Energy Engineering* **5**, 28–42.
- Ong, C. S., Huang, J. J. & Tzeng, G. H. 2005 [Model identification of ARIMA family using genetic algorithms](#). *Applied Mathematics and Computation* **164** (3), 885–912. <http://dx.doi.org/10.1016/j.amc.2004.06.044>.
- Pao, H. 2007 *Forecasting Electricity Market Pricing Using Artificial Neural Networks*.
- Paoli, C., Voyant, C., Muselli, M. & Nivet, M. L. 2010 [Forecasting of preprocessed daily solar radiation time series using neural networks](#). *Solar Energy, Elsevier* **84** (12), 2146–2160.

- Papadimitriou, S. & Bezerianos, A. 1999 Nonlinear analysis of the performance and reliability of wavelet singularity detection based denoising for Doppler ultrasound fetal heart rate signals. *International Journal of Medical Informatics* **53** (1), 43–60. [https://doi.org/10.1016/S1386-5056\(98\)00102-6](https://doi.org/10.1016/S1386-5056(98)00102-6).
- Pincus, S. M. & Goldberger, A. L. 1994 Physiological time-series analysis: what does regularity quantify? *The American Journal of Physiology*. **266** (4), 1643–1656.
- Powell, M. J. D. 1977 Restart procedures for the conjugate gradient method. *Mathematical Programming* **12** (1), 241–254.
- Rezaeian-Zadeh, M., Zand-Parsa, Sh., Abghari, H., Singh, V. P. & Zolghadr, M. 2012 Hourly air temperatures driven using multi-layer perceptron and radial basis function networks in arid and semi-arid regions. *Theoretical and Applied Climatology* **109**, 519–528.
- Rioul, o. & Vetterli, M. 1991 Wavelets and signal processing. IEEE SP Magazine, 14–38.
- Rosca, R. E. 2011 Stationary and non-stationary time series. *The USV Annals of Economics and Public Administration* **10**, 177–186.
- Sanikhani, H., Kisi, O., Maroufpoor, E. & Mundher Yaseen, Z. 2019 Temperatures-based modeling of reference evapotranspiration using several artificial intelligence models: application of different modeling scenarios. *Theoretical and Applied Climatology* **135**, 449–462. <https://doi.org/10.1007/s00704-018-2390-z>.
- Spelta, C., Previdi, F., Savaresi, S. M., Bolzern, P., Cutini, M., Bisaglia, C. & Bertinotti, S. A. 2011 Performance analysis of semi-active suspensions with control of variable damping and stiffness. *Vehicle System Dynamics* **49**, 237–256.
- Taylor, J. W. 2003 Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting* **19**, 715–725.
- Theiler, J., Galdrikian, B., Longtin, A., Eubank, S. & Farmer, J. D. 1992 Using Surrogate Data to Detect Nonlinearity in Time Series. United States. Available from: <https://www.osti.gov/biblio/5299382-using-surrogate-data-detect-nonlinearity-time-series>
- Tsonis, A. A. 1992 *Chaos. From Theory to Applications*. Springer, New York. DOI: 10.1007/978-1-4615-3360-3.
- Xia, C., Wang, J. & McMenemy, K. 2010 Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *Electrical Power and Energy Systems* **32**, 743–750.
- Yang, H., Zhang, Y., Zhang, L. & Xu, T. 2019 Research on floating car speed short-time prediction with wavelet-ARIMA under data missing. In: *Green Intelligent Transportation Systems. GITSS 2017. Lecture Notes in Electrical Engineering*, Vol 503 (W. Wang, K. Bengler & X. Jiang, eds). Springer, Singapore. https://doi.org/10.1007/978-981-13-0302-9_29
- Yuan, C., Liu, S. & Fang, Z. 2016 Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. *Energy* **100** (C), 384–390. doi: 10.1016/j.energy.2016.02.001.
- Zhu, J., Xu, W., Jin, H. & Sun, H. 2018 Prediction of urban rail traffic flow based on multiply Wavelet-ARIMA model. In: *International Conference on Green Intelligent Transportation System and Safety, GITSS 2016: Green Intelligent Transportation Systems*. Springer, Singapore, Vol. 419, pp. 561–576. https://doi.org/10.1007/978-981-10-3551-7_44.

First received 18 January 2020; accepted in revised form 26 August 2020. Available online 12 October 2020