

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 124, No. 4, October – December, 2020, pp. 463–482

Application of vector autoregressive models to estimate pan evaporation values at the Salt Lake Basin, Iran

Ali Shahidi*, Yousef Ramezani, Mohammad Nazeri-Tahroudi, and Saeedeh Mohammadi

Department of Water Engineering
Faculty of Agriculture
University of Birjand, Birjand, Iran

**Corresponding author E-mail: ashahidi@birjand.ac.ir*

(Manuscript received in final form November 11, 2019)

Abstract— Thousands of billions of cubic meters of fresh water collected at great expense are evaporated annually from dams, and salts of evaporating water reduces water quality. In this study, the efficiency of the vector autoregressive model called VAR model has been examined on an annual scale using pan evaporation data in the salt lake basin, Iran, during the statistical period of 1996–2015. Since hydrologic modeling is concerned with the accuracy and efficiency of the model, therefore, we must try to evolve and improve the results of the models. In this study, VAR multivariable time series and nonlinear GARCH models have been used. The results of linear and nonlinear hybrid models in modeling the annual and monthly pan evaporation values of studied stations at the basin area of the salt lake indicated, that the pan evaporation values in the annual scale have the best fit with hybrid models. The results of the study of the accuracy of these models in modeling the pan evaporation values indicated, that the VAR-GARCH hybrid models have a high accuracy relative to the vector models and have been able to model the pan evaporation values with good accuracy and with the lowest error rate. Of the two models that have both annual nature (VAR and VAR-GARCH), the best model can be selected based on the estimation of the error values. In this study, we first examine the accuracy of the relatively new vector autoregressive model. The results of the estimation of error and efficiency of the model indicated the acceptable accuracy of this model in estimating the pan evaporation values in the annual scale. The 95% confidence interval confirmed the simulation results of the calibration step. Overall, the results showed that both VAR and VAR-GARCH models have high accuracy and correlation, and the model's performance criterion also confirms this. The percentage of improvement in the results from the model of the pan evaporation values in the annual scale using the VAR-GARCH model is about 4% relative to the VAR model. However, due to modeling the random section and reducing the uncertainty of the model, the results of modeling the pan evaporation values using the VAR-GARCH model are better than the VAR model. But due to the complexity of calculating the GARCH model, the VAR model can also be used.

Key-words: autoregressive conditional heteroscedasticity model, potential evapotranspiration, Salt Lake, single-variable model

1. Introduction

Since Iran has a dry and semi-arid climate, estimating and modeling hydrologic and meteorological parameters is important for planning and managing water resources. In the meantime, various prediction methods have tried to determine the relationship between independent and dependent variables, and many conceptual and statistical models have been used to predict climate variables. Time series models, as a mathematical-physical model, have a great ability to model linear and nonlinear phenomena. The time series model consists of two main parts, including random component and algebraic (deterministic) component of the model, where the algebraic component of the model is obtained using observational and random components using different stochastic methods. Therefore, the structure of time series models can be adapted to the structure of the hydrological series if the selection of the model and its calculations are correct (Salas, 1993). Thomas and Fiering (1962), Yevjevich (1963), Roesner and Yevjevich (1966) used autoregressive models for modeling the annual and seasonal series of river flows. Since then, a lot of research has been done to develop and extend concepts of the time series models with modifying and correcting models. Many of the processes in the natural systems are nonlinear, although certain aspects of these systems may be closer to the linear process than others. However, the nature of non-linearity is not clear to us (Tsonis, 2001). Nonlinear models have been used more for statistical, economical, and mathematical researches. These include references such as Priestley (1988) and Tong (1990). Most of these models have been used to modeling and predicting the economic time series (Franses and Van Dijk, 2000). Wang et al. (2006) used the combination of the ARMA and GARCH¹ models to fit the variance and daily average of the Yellow River flow in China. The results showed that the ARMA-GARCH model offers very useful results in daily river flow modeling. Caiado (2007) examined the performance of time series one-parameter models in predicting the amount of water consumed in Spain in daily and weekly scales from 2001 to 2006. In this research, ARIMA and GARCH models were fitted on a series of observational data, and the performance of these models was evaluated and confirmed. In the meantime, combined models are used in order to improve the prediction results. Ghorbani et al. (2018) used the hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model to predict the pan evaporation in the northern part of Iran. Results show that an optimal MLP-FFA model outperforms the MLP and SVM model for both tested stations. Ashrafzadeh et al. (2018) estimated the daily pan evaporation using neural networks and meta-heuristic approaches at two weather stations (Anzali and Astara) in the northern part of Iran. The results indicated that converting the simple multilayer perceptron with firefly algorithm makes it a powerful hybrid model for estimating pan evaporation.

¹ Generalized Autoregressive Conditional Heteroscedasticity

Regarding the increasing number of models simulating a various hydrological parameters, linear time series models need to be upgraded by combining these models with nonlinear models, but the accuracy of this combination must be measured. Also, considering the randomness of the time series models, their use in modeling parameters such as hydrological parameters, that themselves have a random nature, seems to be better. Since linear models have generally been found to be univariate in the survey, using multivariate models seems to be necessary due to correlation between hydrological variables. On the other hand, the VAR model is a random process used to create linear dependencies between multiple time series. The VAR model uses an integrated autoregressive model using several parameters. All variables in the VAR model are simultaneously entered into the model, where each variable explains an equation whose evolution is based on delay values of different model variables and an error value. VAR modeling requires a very high level of knowledge in order to found forces affecting a variable, as much as they do not have structural models with simultaneous equations. This model has high efficiency in econometrics, in estimating and predicting the economic parameters. But so far, no studies have been done in the water field. In this study, the efficiency of this model (VAR) is investigated using pan evaporation data in the Salt Lake Basin stations in Iran in the statistical period of 1996–2015 on an annual scale. Also, to study the efficiency of the VAR model, the nonlinear GARCH model has been used.

2. Materials and methods

2.1. Study areas and data

Iran with an area of over 1,648,000 square kilometers has been located in the northern hemisphere and on the Asian continent. The climate of Iran has almost four seasons in all its parts, and in general, one year can be divided into two cold and two hot seasons. Iran with an average annual rainfall of 62.1–344.8 mm has been located between the two meridians of 44° and 64° east and two orbits of 25° and 40° north. About 94.8 percent of the country's surface is in arid and semi-arid regions with low rainfall and high evapotranspiration.

In this study, pan evaporation data from stations in the Salt Lake Basin (Qom, Qazvin, Hamedan, Arak, Tehran, and Karaj stations) have been used in the annual period of 1996–2015. Since the objective model is multivariate, adjacent station data were also used. The specifications and the position of the stations are presented in *Table 1* and *Fig. 1*, respectively.

Table 1. Specifications of the stations in the statistical period 1996–2015 (mm per year)

Station	Min	Max	Mean	STD
Arak	1762.10	2268.70	1959.24	126.19
Hamedan	1415.30	1941.80	1668.19	144.79
Karaj	1601.30	2162.90	1932.77	175.22
Qazvin	1373.70	1805.50	1619.51	116.48
Qom	2453.60	2920.70	2650.87	133.30
Tehran	20.00	2690.40	2175.57	775.07

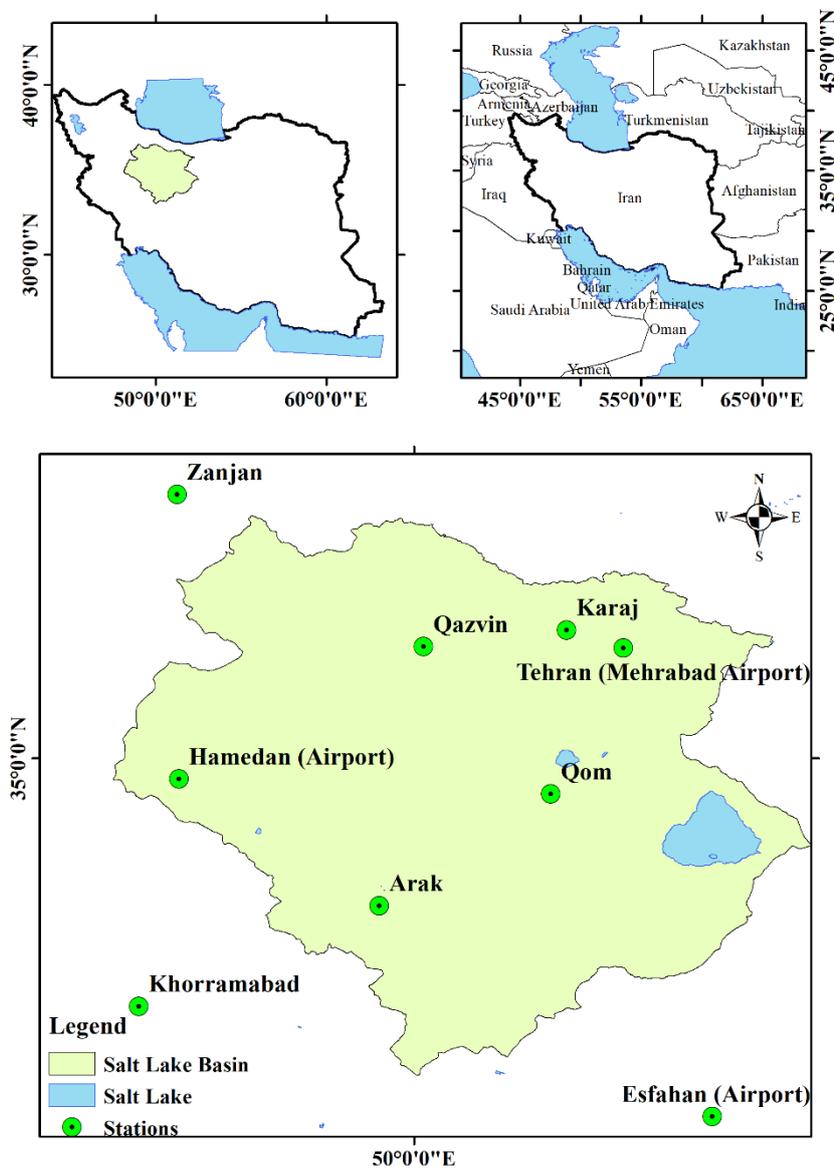


Fig. 1. Location of study areas in Iran.

2.2. Time series analysis

Time series modeling is performed on static random data. Therefore, certain components of the series should be removed and the series become static. The definitive components of a time series include trends and periods. In addition to statics, the series must also follow the normal distribution. To determine the trend in this study, the modified Mann-Kendall test was used (*Kendall, 1938; Mann, 1945; Khalili et al, 2016; Tahroudi et al, 2019b; Khozaymehnejad and Tahroudi, 2019*). After evaluation and deleting the trend (if any), the standardized and normal data will be prepared for use in the above models.

2.3. ARCH models

This model was first presented in economic studies by *Engle (1982)* and was the first to provide a systematic framework for modeling fluctuations. The main idea of the ARCH models is that (a) the modified average investment return is distinct but dependent and (b) the model is dependent and can be described by a simple quadratic function of the values before it. In summary, the ARCH model is assumed to be:

$$\varepsilon_t = \sigma_t z_t \quad \text{and} \quad \sigma_t^2 = a_0 + \sum_{i=1}^m b_i \varepsilon_{t-i}^2, \quad (1)$$

where σ_t^2 is the conditional variance, ε_t is the error term or the remainder of the model with mean value of zero and variance of 1, $a_0 \geq 0, b_i \geq 0$ are the model parameters, m is equal to the order of the model, and z_t is also the time series of the desired parameter (*Engle, 1982*).

2.4. ARCH model structure

To better understand the model, the structure of the ARCH model (1) was considered.

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = a_0 + a_1 a_{t-1}^2, \quad (2)$$

where $a_1 \geq 0, a_0 \geq 0$. First of all, the conditional mean a_t must be zero, because:

$$E(a_t) = E[E(a_t | F_{t-1})] = E[\sigma_t E(\varepsilon_t)]. \quad (3)$$

Then, the conditional variance is obtained from the following equation:

$$Var(a_t) = E(a_t^2) = E[E(a_t^2 | F_{t-1})] = E[a_0 + a_1 a_{t-1}^2] = a_0 + a_1 E(a_{t-1}^2). \quad (4)$$

Since, according to $E(a_t = 0)$ and $Var(a_t) = E(a_{t-1}) = E(a_{t-1}^2)$, a_t is a static and fixed trend, we will have:

$$Var(a_t) = a_0 + a_1 Var(a_t), \quad (5)$$

$$Var(a_t) = \frac{a_0}{(1-a_1)}. \quad (6)$$

Since the variance of a_t should be positive, the range of a_1 should be between 0 and 1.

In some applications, values above (a_t) should also exist, and so, a_1 should provide some extra moments. For example, in studying the behavior of sequences, it is necessary to limit the fourth moment (a_t) . Assuming that ε_t is normal, we will have the following equation (Engle, 1982):

$$E[E(a_t^4 | F_{t-1})] = 3[E(a_t^2 | F_{t-1})]^2 = 3E(a_0 + a_1 a_{t-1}^2)^2. \quad (7)$$

So:

$$E(a_t^4) = E[E(a_t^4 | F_{t-1})] = 3E(a_0 + a_1 a_{t-1}^2)^2 = 3E(a_0^2 + 2a_0 a_1^2 a_{t-1}^2 + a_1^2 a_{t-1}^4). \quad (8)$$

If a_t is considered as the fourth constant and $m_4 = E(a_t^4)$, then:

$$m_4 = 3E(a_0^2 + 2a_0 a_1^2 Var(a_t) + a_1^2 m_4) = 3a_0^2(1 + 2\frac{a_1}{1-a_1}) + 3a_1^2 m_4. \quad (9)$$

Eventually:

$$m_4 = \frac{3a_0^2(1+a_1)}{(1-a_1)(1-3a_1^2)}. \quad (10)$$

2.5. GARCH model

Although the ARCH model is simple, it often requires a lot of parameters to obtain the proper modeling process. For this reason, we have to look for alternative models (Moffat et al., 2017). Bollerslev (1992) proposed the developed ARCH model as follows:

$$\alpha_t = \sigma_t e_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \alpha_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (11)$$

where, e_t is equal to a random series with mean value of zero and variance of one. In fact, the EGARCH model is a natural logarithmic model of the GARCH method, which was presented by *Nelson* (1991).

2.6. Vector autoregressive models

VAR is one of the most successful and flexible models for analyzing multivariate series. This model is, in fact, a model extended from the uniform autoregressive model of multi-variable time series. The VAR model was introduced to describe the dynamic behavior of the economic and financial series and their prediction. This model often provides superior predictions for those who use similar and accurate time series models. VAR model predictions are quite flexible, since they can bet on the future path of potential variables. In addition to describing and forecasting data, the VAR model is also used for structural inferences and analysis policies. In structural analysis, specific hypotheses are imposed on the structure of the data under investigation, and the effects of unexpected shocks or innovations are summed up with the variables specified on the model variables. These effects are usually summarized with impact reaction and predicted error variance analysis functions. This model focuses on the analysis of constant covariance multivariate. VAR models in economics have been introduced by *Sims* (1980). The technical review of VAR models can be found in the *Lütkepohl* (1999) study, and updated VAR techniques are described in researches conducted by *Watson* (1994), *Lütkepohl* (1999), and *Waggoner and Zha* (1999). The use of VAR models for financial information has been given in the studies carried out by *Hamilton* (1994), *Campbell et al.* (1997), *Cuthbertson and Nitzsche* (1996), *Mills* (1999), and *Tsay* (2001).

If $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ represents the vector ($n \times 1$) of the time series variables, then the VAR (p) model with a p -year base delay is as follows:

$$Y_t = c + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \varepsilon_t, \quad t = 1, \dots, T, \quad (12)$$

where Π_i is equal to the coefficient ($n \times n$) of the matrix and ε_t is equal to the matrix ($n \times 1$) of the white noise values with mean value of zero (non-dependent or independent) with constant covariance matrix Σ . For example, the equation of the two-variable VAR model is as follows:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \pi_{11}^2 & \pi_{12}^2 \\ \pi_{21}^2 & \pi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (13)$$

or

$$\begin{aligned} y_{1t} &= c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} + \pi_{11}^2 y_{1t-2} + \pi_{12}^2 y_{2t-2} + \varepsilon_{1t} \\ y_{2t} &= c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} + \pi_{21}^2 y_{1t-2} + \pi_{22}^2 y_{2t-2} + \varepsilon_{2t} \end{aligned} \quad (14)$$

where $\text{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12}$ for $t = s$, otherwise it is zero. Note that each equation has a similar regression of the remainder of y_{1t} and y_{2t} . Hence, the VAR (p) model is just an indirect regression model with remaining variables and definitive terms as common regressions. From a user's perspective, the VAR (p) model is written as:

$$\Pi(L)Y = c + \varepsilon_t, \quad (15)$$

where $\Pi(L) = I_n - \Pi_1 L - \dots - \Pi_p L^p$. Now if the value of the determinant value of $(I_n - \Pi_1 z - \dots - \Pi_p z^p)$ is zero, then the VAR (p) will be static.

If the eigenvalues of a composite matrix have a modulus of less than one, it is outside the complex unit loop (with a modulus greater than one), or equivalent, if the eigenvalues of the composite matrix have a modulus less than one. It is assumed that the process in the past has been initiated from infinite value, then it is a stable process of VAR (p) with constant mean variance and covariance. If Y_t in Eq.(13) is constant covariance, then the mean is given by:

$$F = \begin{pmatrix} \Pi_1 & \Pi_2 & \dots & \Pi_n \\ I_n & 0 & \dots & 0 \\ 0 & . & 0 & : \\ 0 & 0 & I_n & 0 \end{pmatrix} \quad (16)$$

$$\mu = (I_n - \Pi_1 - \dots - \Pi_p)^{-1} c. \quad (17)$$

After the adjusted mean of the VAR (p) model:

$$Y_t - \mu = \Pi_1(Y_{t-1} - \mu) + \Pi_2(Y_{t-2} - \mu) + \dots + \Pi_p(Y_{t-p} - \mu) + \varepsilon_t. \quad (18)$$

The basic VAR (p) model may be very limited to show the main characteristics of the data. Specifically, other conditions of determinism such as a linear time trend or seasonal variables may be used to display data correctly. Additionally, random variables may also be required. The general form of the VAR (p) model with definitive terms and external variables is as follows:

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \dots + \Pi_p Y_{t-p} + \Phi D_t + G X_t + \varepsilon_t, \quad (19)$$

where D_t is the matrix (1×1) of the definite components, X_t is equal to the matrix ($m \times 1$) of the external variables, and Φ and G are also matrix of the model parameters.

2.7. Model performance

In order to evaluate the performance of the model, two Nash-Sutcliff and root mean square error criteria were used. Lower RMSE and higher Nash-Sutcliff coefficients represent the higher accuracy of the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Q}_i - Q_i)^2}{n}}. \quad (20)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (\hat{Q}_i - Q_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q}_i)^2}. \quad (21)$$

In the above relations, Q_i , \hat{Q}_i , and \bar{Q}_i are the observational computational, and mean values of the observational values respectively, and n is the number of data (Akbarpour et al., 2020; Tahroudi et al., 2019a).

3. Results and discussion

At first, preliminary results of time series including trend, randomness of data, and data normal survey have been presented. Then the results of the vector multivariate annual model of time series have been presented in the modeling of annual pan evaporation rates. After reviewing, correcting, and completing the data, the trend of data changes was studied for modeling and initial data analysis. By eliminating the trend of time series, data changes are considered to be constant over time, and this increases the modeling accuracy in ARMA family models. The results of the trend of pan evaporation changes and slope of trend line are presented in the two annual and monthly scales are in *Tables 2* and *3*.

Table 2. Results of the slope of trend line in the statistical period of 1996-2015 in annual and monthly scales

Station	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
Arak	0.00	0.00	-0.49	-0.27	-0.36	-0.08	1.03	0.01	-0.32	-1.42	-2.45	0.00	-9.69
Hamedan	0.00	0.00	0.00	3.62	1.53	1.57	2.65	1.54	1.71	-0.28	-0.27	0.00	11.09
Karaj	0.00	0.00	0.00	2.00	4.70	5.03	6.65	3.05	3.02	1.75	1.16	0.00	12.89
Qazvin	0.00	0.00	0.00	0.44	-0.35	-0.62	2.90	0.32	-0.37	-0.49	-0.99	0.00	0.9
Qom	0.00	0.00	8.61	1.07	-1.43	-1.26	0.28	-1.73	-2.58	-1.69	-1.45	0.17	-6.40
Tehran	0.00	0.00	3.50	4.60	0.67	2.20	0.00	-3.20	-4.03	-2.76	-2.39	-0.62	-6.09

Table 3. Results of the Z statistics of modified Mann-Kendall test in the statistical period of 1995–2015 annual and monthly scales

Station	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
Arak	1.39	-0.63	-0.33	-0.23	-0.16	-0.03	0.62	0.00	-0.36	-0.75	-1.20	-0.07	-1.22
Hamedan	-0.68	-0.69	1.80	1.65	1.07	1.14	1.98	1.91	1.52	-0.1	-0.36	-1.21	1.27
Karaj	0.50	0.00	-1.51	0.63	1.40	1.26	3.15	2.04	1.52	0.75	0.65	0.65	1.10
Qazvin	1.04	-0.87	-0.79	0.36	-0.29	-0.36	1.59	0.23	-0.23	-0.36	-0.29	-1.19	0.03
Qom	-0.16	2.66	3.10	0.29	-0.94	-1.40	0.10	-0.98	-2.24	-1.14	-1.14	0.23	-1.40
Tehran	-1.08	0.33	1.95	1.54	0.11	0.77	0.06	-0.75	-1.33	-1.04	-1.56	-0.93	-0.40

The variation trend of annual pan evaporation rates at the stations showed that the variation of this parameter in the studied area is a combination of increasing and decreasing trends. The results of the annual variation of the stations showed that the north and northwest areas of the studied basin had a decreasing trend, and the border areas have experienced a significant incremental trend in the annual pan evaporation values during the statistical period. These results indicate an increase in the pan evaporation during the statistical period in these areas, which can be attributed to the increase in temperature and climate change, as well as global warming. On the other hand, *Tabari and Talaee (2011)*, *Saboohi et al. (2012)*, *Kousari et al. (2013)*, *Zamani et al. (2018)*, and *Khalili et al. (2016)* showed that there is an increasing trend in

temperature in Iran, especially in the cold months of the year, therefore, temperature variations can be considered as one of the reasons for the decrease or increase of evaporation in the study area. In the meanwhile, the boundary and southern stations of the study area have experienced a decreasing trend in the annual pan evaporation rates in the studied statistical period that the trend of changes in the values of this parameter on the southern border is reduced and significant. In general, the results of the study of the changes in the values of the parameter in the studied period showed that the southern half of the study area experienced decreasing trend, while the northern part of the study area experienced incremental trend. Regarding the slope of the trend line, the maximum incremental changes in annual pan evaporation rates is related to the Karaj station, and the highest decreasing changes in the parameter values are at the annual scale is associated with the Arak station. It should be noted that the stations where the process of evaporation changes from the mattress in them had decreasing or significant increase. Using the 3-year moving average, the trend of annual and monthly changes of these data changed from significant to non-significant. Regarding the length of the statistical period, the significance level of 1% was considered as the base level. The randomness of the data was also evaluated using the Wald-Wolfowitz test. The results showed that the pan evaporation rates at the stations at the significance level of 1% and 5% were randomized. After reviewing the existing data, normalization methods were used to normalize the data, and the results are presented in *Table 4*. Also, to test the normality of the time series of the data, the skewness coefficient test was used. The results of the investigation of the normality of the data under investigation after fitting them with normal distribution functions showed, that based on the skewness test, the normalized data are in the confidence range of normality. After reviewing the normalization functions and ensuring that the data are normal, normal data were standardized and fitted by multivariable vector models in the annual scale.

Table 4. Results of normalization methods in normalizing the data

Station	Initial skewness coefficient	Secondary skewness coefficient	Distribution coefficient	Normalized distribution
Arak	0.69	0.08		Gamma
Hamedan	0.13	0.00	0.19	Box-Cox
Karaj	-0.36	-0.25	2.00	Box-Cox
Qazvin	-0.33	-0.19	2.00	Box-Cox
Qom	0.34	0.04		Gamma
Tehran	-1.62	0.12		Gamma

4. Results of modeling the pan evaporation values using VAR model

After the preliminary review of the data in this study, we examined the number of permitted delays and valid data to predict and interfere with the effective parameters in the data combination. For example, the results of examining the effective parameters and the number of permitted delays are presented in *Table 5*. 95% confidence intervals were used in the calculations in order to approve the permitted number of delays and the number of effective parameters.

Table 5. Results of investigation of correlation and number of delay of Arak station with other stations

Station	Lags	Coefficient	Standard error	z	P> z	95% confidence intervals		Results
Arak	L1.	0.279	0.189	1.470	0.140	-0.092	0.651	Reject
	L2.	0.495	0.179	0.770	0.806	0.145	0.845	Accept
Hamedan	L1.	0.006	0.180	0.030	0.973	-0.347	0.359	Accept
	L2.	0.341	0.137	2.480	0.013	0.072	0.611	Reject
Karaj	L1.	-0.261	0.178	-1.470	0.141	-0.610	0.087	Reject
	L2.	-0.068	0.182	-0.370	0.370	-0.425	0.289	Accept
Qazvin	L1.	-0.543	0.240	-2.260	0.024	-1.014	-0.072	Reject
	L2.	0.124	0.265	0.470	0.369	-0.369	0.644	Accept
Qom	L1.	0.605	0.186	3.250	0.001	0.240	0.969	Reject
	L2.	-0.183	0.223	0.120	0.512	-0.620	0.245	Accept
Tehran	L1.	0.085	0.042	0.101	0.545	0.002	0.186	Accept
	L2.	0.037	0.039	0.940	0.346	-0.040	0.114	Reject
Total		-217.212	707.080	-0.310	0.759	-1603.063	1168.640	Accept

According to the results of the study of the contribution of pan evaporation data in modeling and predicting the mentioned values, it can be seen that in the prediction of pan evaporation values of Arak station (*Table 5*), Arak, Karaj, Qazvin and Qom stations contribute with the second delay and Tehran station with the first delay. In the modeling and prediction of pan evaporation values of the Hamadan station, the pan evaporation values of Arak, Hamedan, and Tehran stations contributed with the first delay, and pan evaporation values of Karaj, Qazvin, and Qom stations contribute with the second delay in modeling and predicting pan evaporation. Similarly, stations of Arak, Hamedan, Karaj, and Qazvin contribute with the first delay and Qom and Tehran stations with the second delay in predicting pan evaporation rates of the Karaj Station. The results of the study on the participation of pan evaporation values in predicting and modeling this parameter at Tehran station showed, that both delays contributed in the stations of Arak,

Hamadan, Karaj, Qazvin, and Qom. However, the pan evaporation values of Tehran station will not contribute to the modeling and prediction of this parameter.

According to the mentioned conditions, modeling of pan evaporation values at the stations was studied using vector autoregressive models. The modeling of the values was done with the VAR model based on the 1000-value Monte Carlo simulation. The results of the modeling of the pan evaporation values in the annual scale using the VAR model are presented in *Figs. 2 and 3*. The results of the study on the accuracy of the model in time series simulation of pan evaporation in the simulation stage are also presented in *Table 6*.

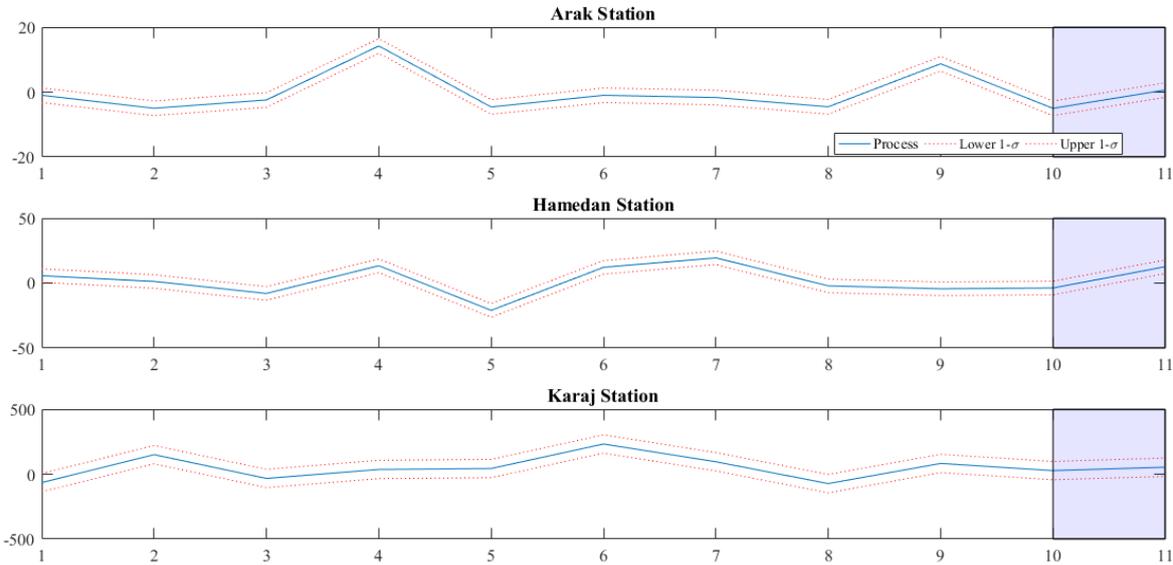


Fig. 2. Verification of simulated values by using vector autoregressive model at Arak, Hamedan, and Karaj stations.

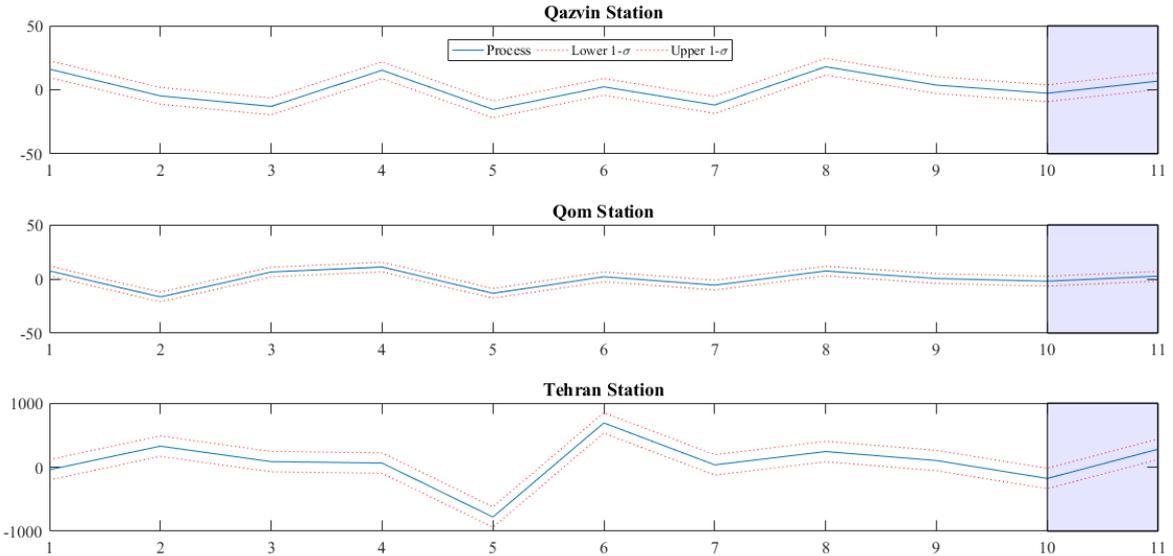


Fig. 3. Verification of simulated values by using vector autoregressive model at Qazvin, Qom, and Tehran stations.

Table 6. Results of the accuracy and efficiency of the VAR model in simulating the annual pan evaporation values

Station	Arak	Hamedan	Karaj	Qazvin	Qom	Tehran
Correlation coefficient	0.61	0.99	0.98	0.98	0.90	0.99
Nash- Sutcliffe coefficient	0.25	0.99	0.97	0.97	0.85	0.99
RMSE	11.37	6.62	14.48	13.45	25.50	21.34

The results of the study of accuracy of the model using the correlation coefficient extracted from the simulated and observed values of pan evaporation time series showed, that the accuracy of the model between the existing stations is between 61% and 99%. On average, the accuracy of the VAR model in simulating pan evaporation values of the existing stations is about 91%. The maximum accuracy of modeling and prediction of pan evaporation values is associated with the Hamedan and Tehran stations, and the minimum accuracy related to the pan evaporation time series of the Arak station in the annual scale. The results of the study of the error rate of the VAR model in simulating pan evaporation values of the studied stations in the Salt Lake Basin were estimated in annual scale using the root mean square error (RMSE). The results showed that the error values in the existing stations varied from 6.62 to 25.55 mm per year. The highest error is related to Qom station, and the lowest error rate is related to Hamedan station. The average error rate of the VAR model in simulating pan evaporation values in the verification stage at an annual scale is 15.46 mm/year. The results of the evaluation of the error values due to pan evaporation time series modeling of the stations in the annual scale showed, that all simulated cases are in the 95% confidence intervals and acceptable. Considering the range of time series variations, the accuracy of the models and their error rate are accepted and confirmed. The efficiency of the VAR model in simulating pan evaporation in the Salt Lake Basin was investigated using the Nash-Sutcliffe test (N-S). The average model efficiency for stations in the Salt Lake Basin is about 84%. Except for Arak station, other stations have efficiency more than 90%.

After verifying the accuracy and efficiency of the VAR model in estimating the pan evaporation values in the salt lake basin, the remaining series values of the VAR model resulting from the modeling of the parameters in the annual scale using the nonlinear GARCH model were investigated and fitted. After combining the vector autoregressive model with the GARCH model, the hybrid model of VAR-GARCH was formed. The results of the study and comparison of the two VAR and VAR-GARCH models in modeling and estimating the pan evaporation values in the annual scale in the catchment area of the Salt Lake in the statistical period are presented in *Table 7*.

Table 7. Comparison of the error rate and efficiency of the two studied models

Station	Percentage of getting better the error of hybrid model	Percentage of getting better the performance of hybrid model	RMSE of hybrid model (mm/year)	N-S of hybrid model (%)	Correlation coefficient of hybrid model
Arak	3.38	-1.40	10.98	95	0.99
Hamedan	3.97	1.32	6.35	99	0.99
Karaj	7.78	3.20	13.35	92	0.94
Qazvin	1.40	2.08	13.26	99	0.99
Qom	5.98	0.80	23.97	92	0.96
Tehran	3.28	0.40	20.64	98	0.97

After verifying the accuracy of the VAR-GARCH model in simulating pan evaporation values on an annual scale, this parameter was simulated using pan evaporation values of adjacent stations and predicted for 5 years (2015–2020). The results of the prediction of pan evaporation values are presented in *Figs. 4* and *5* using the VAR-GARCH model.

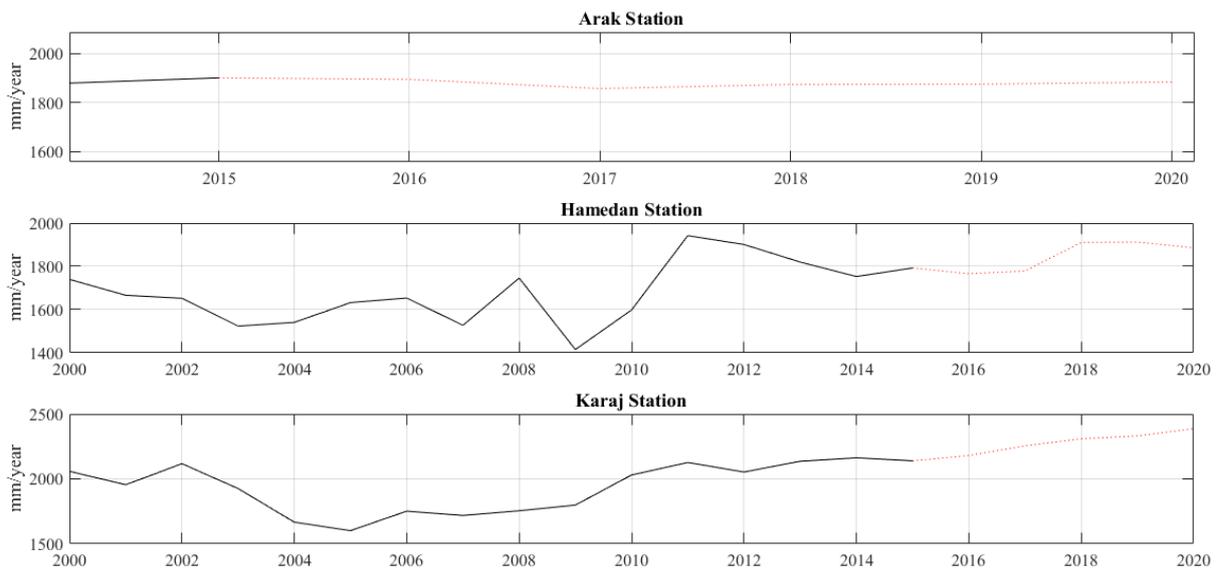


Fig. 4. Prediction of pan evaporation values in the period of 1996-2020 at Arak, Hamadan, and Karaj stations using the VAR-GARCH model.

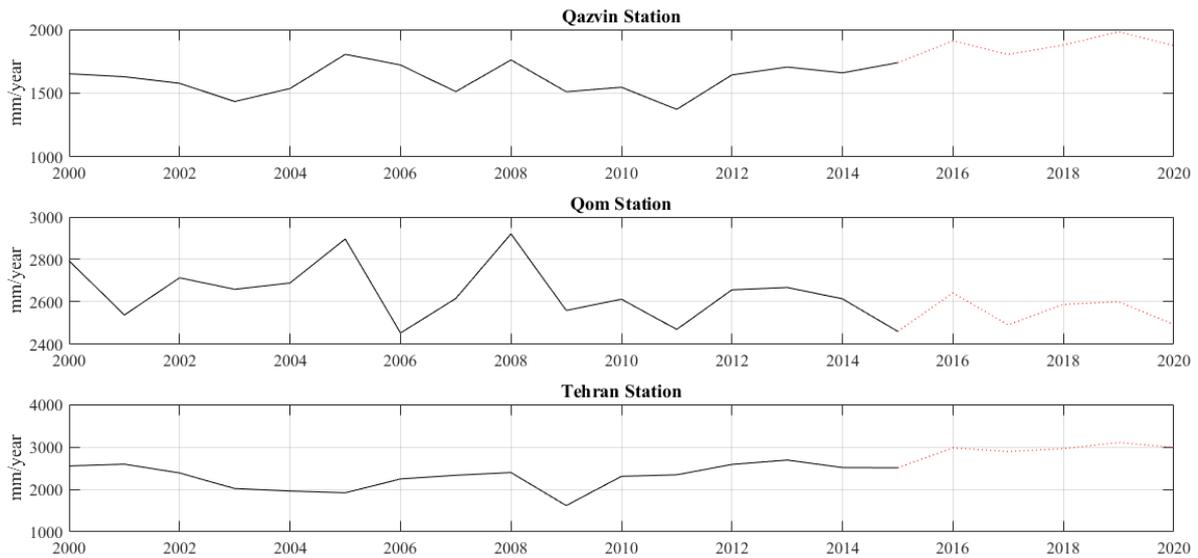


Fig. 5. Prediction of pan evaporation values in the period of 1996-2020 at Qazvin, Qom, and Tehran stations using the VAR-GARCH model.

The results of modeling pan evaporation values using multivariate vector time-series models showed, that these models have high ability to model these values under the influence of pan evaporation of other stations. As it can be seen from the results, the VAR and VAR-GARCH models have been able to simulate and predict pan evaporation values. It is clear, that at all stations the correlation between observational and computational data and the performance of the model at all stations are high. The results of the modeling of the monthly pan evaporation values under the influence of annual pan evaporation values of other adjacent stations showed that multivariable vector hybrid models at the Tehran station located in the northeast of the basin have lower accuracy and lower efficiency.

The results of the vector-hybrid models showed that using these models improved the error of modeling of the pan evaporation values of the Arak, Hamedan, Karaj, Qazvin, Qom, and Tehran stations on average by 4, 4, 8, 1, 6, and 3% respectively. The results of the accuracy of the studied models showed that vector-hybrid models provide better results than vector time series models. Also, due to the large effect of the pan evaporation parameter of the adjacent stations and the interference of this parameter, pan evaporation values of each station were modeled well. For this reason, by interfering with the parameters associated with the data used in modeling, the accuracy of modeling and analysis can be greatly increased. The results obtained by *Camacho et al.* (1985) showed the superiority of multivariate models compared to single-variable models. Also, the results showed that among two hybrid and multivariate models, hybrid models provide better fitness and less error than multivariate

models, although the accuracy of multivariate models is acceptable, which is consistent with the studies conducted by *Tesfaye et al.* (2006) in the modeling of seasonal flow discharge of the British Frieser River. On the other hand, by adding nonlinear models to linear time series models, the model's uncertainty can be partially eliminated. The results showed that the hybrid model was more accurate than the linear time series model, which is consistent with *Wang et al.* (2005), *Caiado* (2007), and *Laux* (2011). Since the autoregressive model is an annual and single-variable model, using the VAR model as an alternative to the AR model is the best option. Because in addition to using multivariate mode, other effective parameters are introduced with appropriate delays.

5. Conclusion

Using monthly and annual pan evaporation time series data, the variation trend of the parameter was investigated in monthly and annual scales. The data of the evaporation gauge stations of the center provinces of Iran located in the catchment area of the Salt Lake during the period of 1996–2015 were analyzed using a modified Mann-Kendall test. The results of this study showed that the southern and eastern stations of the studied area have a decreasing trend and the northern and western stations have an increasing trend. On the annual scale, the highest incremental trend happened at the Karaj station based on the trend line slope, which was about 257 units during the 20-year-long statistical period. According to the results of the study, the Arak station has experienced the most evaporation reduction during the 20-year-long statistical period. These decreasing trends were around 200 units over the past 20 years. On a monthly scale, the results of the study of the pan evaporation data showed, that in all studied months, the trend of changes in the parameter values in the studied period in the north and northwest parts of the study area and the Karaj, Hamedan, and Qazvin stations have an unreasonable and significant increase, indicating a decrease in humidity, an increase in evaporation and temperature in the area. In general, the results indicate an increase in the pan evaporation rate in the north and northwest basin in all months, which indicates warming and increasing temperature in these areas. As the temperature increases and the humidity decreases, the pan evaporation rate increases.

The conditional nonconformity (variable time variance or oscillation) is usually ignored in the context of meteorological variable modeling. The present study shows that although the VAR approach is sufficient to model the conditional average of pan evaporation time series, the ARCH effect will still improve the results. Identification of ARCH effects and the inability of the VAR method to eliminate the effect of conditional variance are out of the scope of this study. However, the presence of the ARCH effect in the pan evaporation rates may be due, in part, to fluctuations in variance reported by *Wang et al.* (2005).

Other factors which may prove the effects of ARCH in hydrological periods are fluctuations in air temperature, effective factor for snow drops, and evapotranspiration and precipitation changes. Similar reports have been presented by *Modarres* and *Ouarada* (2013) on the reasons for this work. In these reports, conditional non-correlation may be due to climate factors that affect the hydrological series variance changes. The VAR-GARCH approach demonstrated the ability to model conditional variance. On the other hand, this approach improves the performance of multi-criteria error estimation. This result can be one of the important aspects of pan evaporation modeling in areas, where some of the evaporation stations are located within a region with climatic variation. As a result, this study shows that the VAR-GARCH model with a suitable change can increase the performance of time series and have a conditional nonconformity stability. Regarding the structure of conditional-covariance variance and correlation, *Modarres* and *Ouarada* (2013) stated that this is a physical feature of the watershed that may affect the existence of the variance of the time variable in the remainder. The results of linear and nonlinear hybrid vector models in modeling the annual and monthly pan evaporation rates at the site of evaporation stations studied at the catchment area of the Salt Lake showed, that pan evaporation values of the studied stations have the best fitness in annual scale with hybrid models. The results of the study of the accuracy of these models in modeling the pan evaporation values of the stations showed, that the VAR-GARCH hybrid models have a high accuracy relative to the vector models and have been able to model the pan evaporation rates with the lowest error rate. Of the two models that both have annual nature (VAR and VAR-GARCH), the best model can be selected based on the estimation of the error values. In this study, we first looked at the accuracy of the relatively new autoregressive vector model called VAR. The results of the estimation of error and the efficiency of the model indicated the acceptable accuracy of this model in estimating the pan evaporation values in the annual scale. The 95% confidence interval confirmed the simulation results of the calibration step. Generally, according to the range of data variations, as well as the computational errors and accuracy, we can see the appropriate performance of this model. The improvement percentage in the results of modeling pan evaporation rates in the annual scale using the VAR-GARCH model is about 4% relative to the VAR model. However, due to modeling the random section and reducing the uncertainty of the model, the results of modeling the pan evaporation rates using the VAR-GARCH model are better than the VAR model. However, considering the complexity of the GARCH model calculations, we can ignore the 4% improvement of the model. However, this model is presented for a salt lake watershed, and more research is needed in different climates for general conclusion, and generalizing it to all areas.

References

- Akbarpour, A., Zeynali, M.J., and Tahroudi, M.N., 2020: Locating optimal position of pumping Wells in aquifer using meta-heuristic algorithms and finite element method. *Water Res. Manag* 34, 21–34. <https://doi.org/10.1007/s11269-019-02386-6>
- Ashrafzadeh, A., Malik, A., Jothiprakash, V., Ghorbani, M.A., and Biazar, S.M., 2018: Estimation of daily pan evaporation using neural networks and meta-heuristic approaches. *ISH J. Hydraulic Engin.*, 1–9. <https://doi.org/10.1080/09715010.2018.1498754>
- Bollerslev, T., Chou, R.Y., and Kroner, K.F., 1992: ARCH modeling in finance. A selective review of the theory and empirical evidence. *J. Econometrics* 52, 5–59. [https://doi.org/10.1016/0304-4076\(92\)90064-X](https://doi.org/10.1016/0304-4076(92)90064-X)
- Caiado, J., 2007: Forecasting water consumption in Spain using univariate time series models. Munich Personal RePEc Archive, no: 6610.
- Camacho, F., McLeod, A.I., and Hipel, K.W., 1985: Contemporaneous autoregressive - moving average (CARMA) modeling hydrology. *Water Res. Bull.* 21, 709–720. <https://doi.org/10.1111/j.1752-1688.1985.tb05384.x>
- Campbell, J.Y., Lo, A.W., and MacKinlay, A.C., 1997: The econometrics of financial markets (Vol. 2.) Princeton, NJ: Princeton University press. 149–180. <https://doi.org/10.1515/9781400830213>
- Cuthbertson, K. and Nitzsche, D., 1996: Quantitative financial economics: stocks, bonds and foreign exchange. John Wiley & Sons.
- Engle, R.F., 1982: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: J. Econometric Soc.* 50, 987–1007. <https://doi.org/10.2307/1912773>
- Franses, P.H. and Van Dijk, D., 2000: Non-linear time series models in empirical finance. Cambridge University Press. <https://doi.org/10.1017/CBO9780511754067>
- Ghorbani, M.A., Deo, R.C., Yaseen, Z.M., Kashani, M.H., & Mohammadi, B., 2018: Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theor. Appl. Climatol.* 133, 1119–1131. <https://doi.org/10.1007/s00704-017-2244-0>
- Hamilton, J.D., 1994: Time series analysis (Vol. 2). Princeton, NJ: Princeton Univ. Press. 690–696.
- Kendall, M.G., 1938: A new measure of rank correlation. *Biometrika* 30, 81–93. <https://doi.org/10.1093/biomet/30.1-2.81>
- Khalili, K., Tahoudi, M.N., Mirabbasi, R., and Ahmadi, F., 2016: Investigation of spatial and temporal variability of precipitation in Iran over the last half century. *Stoch. Environ. Res. Risk Assess.* 30, 1205–1221. <https://doi.org/10.1007/s00477-015-1095-4>
- Khozeymehzad, H. and Tahroudi, M.N., 2019: Annual and seasonal distribution pattern of rainfall in Iran and neighboring regions. *Arabian J. Geoscie.* 12, 271. <https://doi.org/10.1007/s12517-019-4442-9>
- Kousari, M.R., Ahani, H., and Hendi-zadeh, R., 2013: Temporal and spatial trend detection of maximum air temperature in Iran during 1960–2005. *Glob. Planet. Change* 111, 97–110. <https://doi.org/10.1016/j.gloplacha.2013.08.011>
- Laux, P., Vogl, S., Qiu, W., Knoche, H.R., and Kunstmann, H., 2011: Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. *Hydrol. Earth Syst. Sci.* 15, 2401–2419. <https://doi.org/10.5194/hess-15-2401-2011>
- Lütkepohl, H., 1999: Vector autoregressions. A companion to theoretical econometrics, 678–699. <https://doi.org/10.1002/9780470996249.ch33>
- Mann, H.B., 1945: Nonparametric tests against trend. *Econometrica* 13, 245–259. <https://doi.org/10.2307/1907187>
- Mills, T.C., 1999: The Econometric Modeling of Financial Time Series, Second Edition. Cambridge University Press, Cambridge.
- Modarres, R. and Ouarda, T.B., 2013: Generalized autoregressive conditional heteroscedasticity modelling of hydrologic time series. *Hydrol. Proc.* 27, 3174–3191. <https://doi.org/10.1002/hyp.9452>

- Moffat, I.U., Akpan, E.A., and Abasiokwere, U.A. 2017: A time series evaluation of the asymmetric nature of heteroscedasticity: an EGARCH approach. *Int. J. Stat. Appl. Math.* 2 (6), 111–117.
- Nelson, D.B., 1991: Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*: 59, 347–370. <https://doi.org/10.2307/2938260>
- Priestley, M.B., 1988: Non-linear and non-stationary time series analysis. Academic Press, London.
- Roesner, L.A. and Yevjevich, V.M., 1966: Mathematical models for time series of monthly precipitation and monthly runoff. *Hydrol. Papers* 15, 28–29.
- Saboohi, R., Soltani, S., and Khodaghali, M., 2012: Trend analysis of temperature parameters in Iran. *Theor. Appl. Climatol.* 109, 529–547. <https://doi.org/10.1007/s00704-012-0590-5>
- Salas, J.D., 1993: Analysis and modeling of hydrologic time series. *Handbook of hydrology* 19, 1–72.
- Sims, C.A. 1980: Macroeconomics and reality. *Econometrica* 48, 1–48. <https://doi.org/10.2307/1912017>
- Tabari, H. and Talaee, P.H. 2011: Temporal variability of precipitation over Iran: 1966–2005. *J. Hydrol.* 396, 313–320. <https://doi.org/10.1016/j.jhydrol.2010.11.034>
- Tahroudi, M.N., Khalili, K., Ahmadi, F., Mirabbasi, R., and Jhajharia, D., 2019a: Development and application of a new index for analyzing temperature concentration for Iran's climate. *Int. J. Environ. Sci. Technol.* 16(6), 2693–2706. <https://doi.org/10.1007/s13762-018-1739-2>
- Tahroudi, M.N., Ramezani, Y., and Ahmadi, F., 2019b: Investigating the trend and time of precipitation and river flow rate changes in Lake Urmia basin, Iran. *Arabian J. Geosci.* 12, 219.
- Tesfaye, Y.G., Meerschaert, M.M., and Anderson, P.L., 2006: Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resour. Res.* 42, W01419. <https://doi.org/10.1029/2004WR003772>
- Thomas, H.A., and Fiering, M.B., 1962: Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In (Eds. *Mass et al.*) *Design of Water Resource Systems*, Harvard University Press, Cambridge 459–493. <https://doi.org/10.4159/harvard.9780674421042.c15>
- Tong, H., 1990: Non-linear time series: a dynamical system approach. Oxford University Press.
- Tsay, R.S., 2001: Analysis of financial time series (Vol. 543). John Wiley & Sons.
- Tsonis, A.A., 2001: Probing the linearity and nonlinearity in the transitions of the atmospheric circulation. *Nonlinear Proc. Geophys.* 8, 341–345. <https://doi.org/10.5194/npg-8-341-2001>
- Waggoner, D.F. and Zha, T., 1999: Conditional forecasts in dynamic multivariate models. *Rev. Economics Stat.* 81, 639–651. <https://doi.org/10.1162/003465399558508>
- Wang, W., Van Gelder, P.H.A.J.M., and Vrijling, J.K., 2005: Testing and modeling autoregressive conditional heteroskedasticity of streamflow processes. *Nonlin. Process. Geophys.* 12, 55–66. <https://doi.org/10.5194/npg-12-55-2005>
- Wang, W., Vrijling, J.K., Van Gelder, P.H., and Ma, J., 2006: Testing for nonlinearity of streamflow processes at different timescales. *J. Hydrol.* 322, 247–268. <https://doi.org/10.1016/j.jhydrol.2005.02.045>
- Watson, M.W., 1994: Vector autoregressions and cointegration. *Handbook of econometrics*, 4, 2843–2915. [https://doi.org/10.1016/S1573-4412\(05\)80016-9](https://doi.org/10.1016/S1573-4412(05)80016-9)
- Yevjevich, V., 1963: Fluctuations of wet and dry years. Part I: Research data assembly and mathematical models. *Hydrology Paper* 1.
- Zamani, R., Mirabbasi, R., Nazeri, M., Meshram, S.G., and Ahmadi, F., 2018: Spatio-temporal analysis of daily, seasonal and annual precipitation concentration in Jharkhand State, India. *Stoch. Environ. Res. Risk Ass.* 32, 1085–1097. <https://doi.org/10.1007/s00477-017-1447-3>