University of Birjand
Water Harvesting Research Vol. 7, No. 1, Spring & Summer 2024, Original paper p. 34-50
10.22077/JWHR.2024.7634.1138

Forecasting and Modeling of Dew Point Temperature in Meteorological Stations of Eastern Region of Iran Based on VAR and VAR-GARCH Models

Vahid Khoramnezhad[a], Abolfazl Akbarpour[b&*], Mohammad Nazeri Tahroudi[c]

[a]MSc Student, Department of Civil Engineering, University of Birjand, Birjand, Iran.
[b]Professor, Department of Civil Engineering, University of Birjand, Birjand, Iran.
[c]Assistant Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran.

[*]Corresponding Author, E-mail address: akbarpour@birjand.ac.ir
Received: 30 April 2024/ **Revised**: 28 May 2024/ **Accepted**: 30 May 2024

Abstract
In this research, modeling and estimation of dew point temperature values in eight meteorological stations located in the eastern regions of Iran were done. These stations, including Bam, Birjand, Iranshahr, Kerman, Mashhad, Tabas, Zabol and Zahan, are all characterized by a dry climate. First, the correlation of different weather parameters with dew point temperature was investigated and then the parameters of mean temperature, maximum temperature and minimum temperature were selected as the parameters with the highest correlation to dew point temperature. These selected parameters then incorporated into a VAR (Vector Autoregression) model as inputs for estimating dew point temperature values. This modeling approach allows us to capture the interdependencies between these variables and enhance our accuracy in predicting dew point temperature. Then the stability of the residual series of the VAR model was investigated and the residual series of this model was developed using the generalized ARCH model. The result of the development of the VAR model was the investigation of the dew point temperature in eight meteorological stations with the VAR-GARCH model. The results indicated that this combined model outperformed VAR model in both the train and test phases. Specifically, the VAR-GARCH model demonstrated higher accuracy and improved results compared to solely using a VAR model. The incorporation of GARCH allowed better modeling of the residual series, leading to an overall increase in accuracy ranging from 5% to 30% during the test phase. These findings suggest that considering both autoregressive dynamics and conditional heteroskedasticity is crucial for accurately predicting dew point temperatures. By incorporating GARCH into our modeling approach, we were able to capture additional information about volatility and further enhance our predictions.
Keywords: Autoregressive, Conditional heteroscedasticity, Linear model, Non-linear model.

1. Introduction

Research shows that the VAR-GARCH model with appropriate changes can improve the performance of time series and have conditional variance stability. Hybrid VAR-GARCH models are more accurate than vector autoregressive models. Vector autoregressive model (VAR) is one of the best and most flexible models for multivariate time series analysis. This model is actually an extended version of the univariate autoregressive model for multivariate time series. VAR model was introduced to analyze and predict the dynamic behavior of economic time series in financial markets. This model usually provides superior forecasts for those who use simple and accurate time series models. VAR model forecasts are quite flexible, as they can be conditioned on the future paths of specified potential variables. In addition to describing and predicting data, the VAR model is also used for structural inference and policy analysis. In structural analysis, certain assumptions are imposed on the structure of the studied data and the effects of unexpected shocks or innovations are summarized on the variables of the model. These effects are usually summarized by impulse response functions and prediction error variance decomposition. This model focuses on the analysis of multivariate covariance that are constant over time. VAR models in economics were introduced by Sims (1980). A technical review of VAR models can be found in Lutekpol (1991), and updated reviews of VAR techniques can be found in Watson (1994), Lutekpol (1999), and Wagner and Zha (1999).

The use of VAR models in financial data is presented in Hamilton (1994), Campbell et al. (1998), Culbertson (1996), Mills (1999) and Tsay (2005) so far, various researches have been done in the field of modeling and forecasting of weather parameters. Each of these studies have considered different and different points of view. In various studies conducted in different parts of the world, several methods have been used to study the dew point temperature and variable results have been obtained (Baguskas et al. 2016; Tomaszkiewicz et al. 2017; Aguirre-Gutiérrez et al. 2019; Lin et al. 2021).

In the study of Shahidi et al. (2020), the efficiency of a VAR model was investigated on an annual scale using the pan evaporation data of Salt Lake basin, Iran, during the period of 1996-2015. The results showed that both VAR and VAR-GARCH models have high accuracy and correlation, and the performance criteria of the model also confirm this issue. The improvement percentage of the results of the annual pan evaporation model using the VAR-GARCH model is about 4% compared to the VAR model. Due to modeling the residual series and model uncertainty reduction, the results of modeling the pan evaporation values using VAR-GARCH model are better than VAR model. However, due to the computational complexity of the GARCH model, the VAR model can also be used.

Ramezani et al. (2023) used copula-based and ARCH-based models to predict storms in the Aras River basin in northwestern Iran. They used, VAR-GARCH, copula, and Copula-GARCH models to analyze the joint frequency analysis of storms. Based on the results, the VAR-GARCH model was more accurate than the Copula and Copula-GARCH models. The VAR-GARCH model provided higher accuracy in the simulations due to considering different interruptions in the simulations and modeling the variance of the residual series. In fact, having information about a storm that has occurred in the present can accurately predict the next storm. They showed that it can be very useful in flood management and the created curves can be used as a flood warning system in the basin.

Non-precipitation water, which mainly includes fog water, dew water and water vapor absorption, plays an important role in local ecology in arid and semi-arid regions. Although the components of non-precipitating water have been studied separately in previous works, little attention has been paid to integrated properties and relationships between components. A method for identifying non-precipitating water components was developed based on a combination of lysimeter measurements and micrometeorological data. Among the researches that have been done in hydrology studies using time series, it is possible to mention the creation of univariate and bivariate models, or different artificial intelligence models, etc., also in different studies done in different places.

Also, in different researches that have been done in different parts of the world, different methods have been used to investigation the dew point temperature and different results have been obtained. However, so far there has been no research on simulating and predicting dew point temperature using VAR models, as well as developed and hybrid models that consider heteroskedasticity. This is because the effect of conditional variance modeling in multivariate simulations has not been seen in different studies.

The purpose of this research is to simulate and predict the dew point temperature in different climates of Iran using combined time series models. Combined time series is one of the newest methods for multivariate analysis of hydrological phenomena. Dew point temperature analysis using integrated multivariate time series models can lead to valuable information in hydrological applications. The main innovation of this research is also the use of time series models and conditional variance combinations to evaluate different input patterns to the simulation model. By using these models, it is possible to provide the best prediction model to simulation of dew point temperature values in different climates. This proposed approach leads to regional models for dew point temperature prediction.

Accurate prediction of dew point temperature is of particular importance in various scientific fields such as hydrology, agriculture and climatology. Because many important parameters are involved in determining and calculating dew point temperature, including temperature (minimum, maximum, mean), relative humidity, saturation vapor pressure, actual vapor pressure, and mean monthly precipitation. Therefore, it will be very efficient to determine dew point temperature using fewer parameters that can be easily measured in meteorological stations. The purpose of this research is to investigate the accuracy of the vector time series model in simulating and predicting the dew point temperature using different input patterns. Also, due to the random nature of the studied series, the investigation and modeling of the residual series increases the efficiency of the studied models, for which ARCH models are used.

2. Materials and Methods

2.1. Case study

The study area of this research is the provinces of South Khorasan, Razavi Khorasan, Kerman, and Sistan and Baluchestan located in eastern part of Iran. In this study, dew point temperature values were modeled and predicted using meteorological data from mentioned stations in eastern Iran. The studied stations in this research are Bam, Birjand, Iranshahr, Kerman, Mashhad, Tabas, Zabol, and Zahedan that shown in Figure 1.

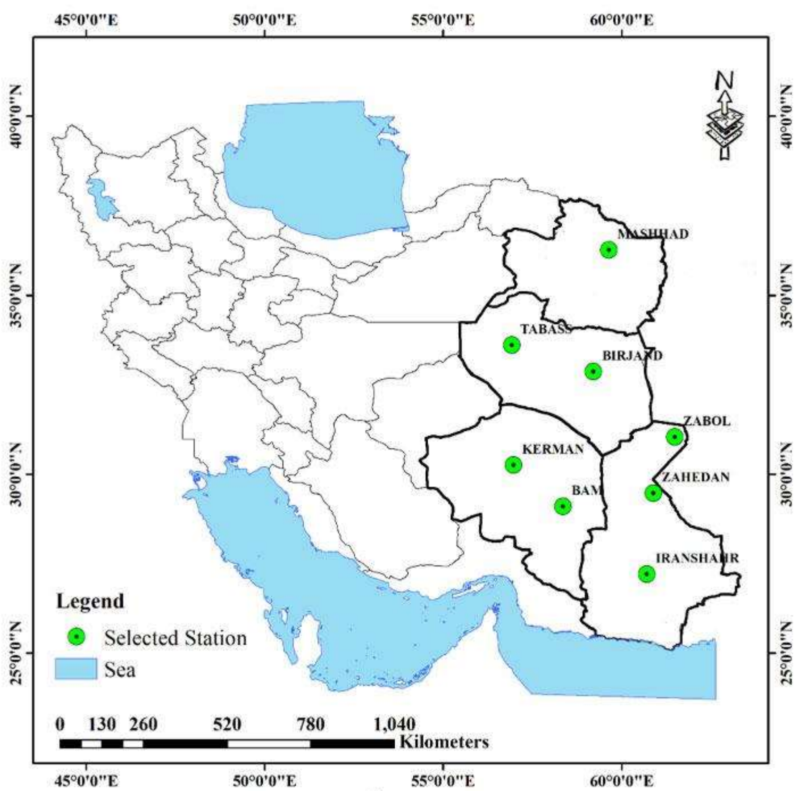

Fig. 1. Location of the studied stations in eastern Iran

Examining the statistical period of synoptic stations in the country demonstrated that although the number of these stations is high, fewer of them have a long-term statistical period that is suitable for studying climate change provides the names and specifications of the stations whose long-term statistical period had the desired characteristics. Although statistics from some stations are available from 1951 AD, the statistical period of 1983-2021 was studied to cover more stations and also to eliminate data inhomogeneity in the early years of the stations.

2.2. Dew point temperature

In the Bam station, the highest dew point values are in July (July 10 to August 10) and the lowest dew point is calculated in December (December 10 to January 10). In the stations of Birjand, Iranshahr, Kerman, Mashhad and Tabas, similar to Bam station, the highest dew point temperature is calculated in July (July 10 to August 10). In Zabol and Zahedan stations, the highest dew point temperature was recorded in August (August 11 to September 12). The lowest calculated dew point for the stations of Bam, Birjand, Iranshahr, Kerman, Tabas, Zabol and Zahedan was recorded in November (November 10 to December 10). While the lowest dew point temperature for the stations of Mashhad occurred in January (January 10 to February 10). All data were calculated for the period 1983 to 2021. The minimum and maximum dew point temperature values calculated for all stations are given in Table 1.

Table 1. Maximum and minimum dew point of the studied stations

Station	Mean dew point temperature	Maximum dew point temperature
Bam	16	15.85
Birjand	16	17.76
Iranshahr	25.8	23.83
Kerman	15	17.03
Mashhad	15.7	27.69
Tabas	20	15.88
Zabol	20	19.89
Zahedan	14.95	27.21

First, the correlation of various data (such as evapotranspiration, sunshine hours, wind speed, average humidity, maximum and minimum temperature and also mean temperature) with dew point temperature was measured and then 3 parameters (maximum temperature, minimum temperature, mean temperature) were selected as model inputs with the highest correlation coefficient. In this research, VAR method was used first to simulating the dew point temperature values. Then the residual series was developed with the ARCH model and VAR-GARCH model was produced. Dew point temperature modeling was performed using these 2 models and the results were finally analyzed and compared. The validation of the models and their efficiency were investigated in terms of root mean square error (RMSE) and Nash-Sutcliffe efficiency. In this research, VAR and hybrid VAR-GARCH model are used to simulate and model dew point temperature in different stations in eastern Iran. Also, maximum, minimum, and mean temperature data are used on a monthly scale.

2.3. Vector Autoregressive (VAR) Model

The vector autocorrelation (VAR) model is a statistical technique used to establish a linear relationship between multiple variables. It employs a self-correlated integrated model, where all the variations are incorporated simultaneously. Each value in the VAR model is explained by an equation that takes into account its own variation as well as the variations from other models, along with an error term. Understanding the forces at play in VAR modeling requires a significant amount of knowledge, as there are no pre-existing structural models with the necessary equations. Despite this complexity, the VAR model is widely used in econometrics and efficiency estimations, and it has been economically validated. However, there have been no studies conducted on this subject in our country.

If $Y_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ represents the vector (n × 1) of the time series variables, then the VAR (p) model with a p-year base delay is as follows (Salas, 1980):

$$Y_t = c + \Pi_1 Y_{t-2} + ... + \Pi_p Y_{t-p} + \varepsilon_t, \quad t = 1, ..., T \tag{1}$$

where, Πi is equal to the coefficient (n × n) of the matrix and ε_t is equal to the matrix (n × 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:

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

Or

$$\begin{aligned} y_{1t} &= c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} \\ &+ \pi_{11}^2 y_{1t-2} + \pi_{12}^2 y_{2t-2} + \varepsilon_{1t} \\ y_{2t} &= c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} \\ &+ \pi_{21}^2 y_{1t-1} + \pi_{22}^2 y_{2t-1} + \varepsilon_{2t} \end{aligned} \tag{3}$$

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 as follow (Shahidi et al. 2020):

$$\Pi(L)Y = c + \varepsilon_t \quad (4)$$

where $\Pi(L) = I_n - \Pi_1 L - ... - \Pi_1 L^p$. Now if the value of the determinant value of $(I_n - \Pi_1 z - ... - \Pi_p z^p)$ 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. 2) is constant covariance, then the mean is given by:

$$F = \begin{pmatrix} \Pi_1 & \Pi_2 & .... & \Pi_n \\ I_n & 0 & .... & 0 \\ 0 & . & 0 & : \\ 0 & 0 & I_n & 0 \end{pmatrix} \quad (5)$$

$$\mu = (I_n - \Pi_1 - ... - \Pi_p)^{-1} c \quad (6)$$

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

$$\begin{aligned} Y_t - \mu &= \Pi_1(Y_{t-1} - \mu) + \Pi_2(Y_{t-2} - \mu) \\ &+ ... + \Pi_p(Y_{t-p} - \mu) + \varepsilon_t \end{aligned} \quad (7)$$

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:

$$\begin{aligned} Y_t &= \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + ... + \Pi_p Y_{t-p} \\ &+ \Phi D_t + G X_t + \varepsilon_t \end{aligned} \quad (8)$$

where, D_t is the matrix (1×1) of the definite components, X_t is equal to the matrix (m × 1) of the external variables and, Φ and G is also the matrix of the model parameters (Shahidi et al. 2020).

2.4. Autoregressive Conditional Heteroskedasticity Model (ARCH)

ARCH models were first introduced by Engle (1982) for economic models and are the first models with a systematic procedure for modeling volatility. ARCH models work in two ways: (a) the mean-adjusted return on investment is separate but dependent, and (b) the model is dependent and can be described by a second-order function of the previous data. In general, the ARCH model is considered as follows:

$$\varepsilon_t = \sigma_t z_t \quad and \quad \sigma_t^2 = a_0 + \sum_{i=1}^{m} b_i \varepsilon_{t-i}^2 \quad (9)$$

Where, σ_t^2 is the conditional variance, ε_t denotes 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). To better understand the model, the structure of the ARCH (1) model was considered.

$$a_t = \sigma_t \varepsilon_t, \; \sigma_t^2 = a_0 + a_1 a_{t-1}^2 \quad (10)$$

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

$$E(a_t) = E[E(a_t \mid F_{t-1})] = E[\sigma_t E(\varepsilon_t)] \quad (11)$$

Then the conditional variance is obtained from the following equation:

$$\begin{aligned} Var(a_t) &= E(a_t^2) = E[E(a_t^2 \mid F_{t-1})] \\ &= E[a_0 + a_1 a_{t-1}^2] = a_0 + a_1 E(a_{t-1}^2) \end{aligned} \quad (12)$$

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:

$$Var(a_t) = a_0 + a_1 Var(a_t) \quad (13)$$

$$Var(a_t) = \frac{a_0}{(1 - (a_0))} \quad (14)$$

Since the variance of a_t should be positive, thus the range of a_1 should be between 0 and 1. In some applications, values above (a_t) should also exist and so, α_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 (Nazeri-Tahroudi et al., 2022):

$$E[E(a_t^4 \mid F_{t-1})] = 3[E(a_t^2 \mid F_{t-1})]^2 = 3E(a_0 + a_1 a_{t-1}^2)^2 \quad (15)$$

So:

$$E(a_t^4) = E[E(a_t^4 \mid F_{t-1})] = 3E(a_0 + a_1 a_{t-1}^2)^2 = 3E(a_0^2 + 2a_0 a_1^2 a_{t-1}^2 + a_1^2 a_{t-1}^4) \quad (16)$$

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

$$m_4 = 3E(a_0^2 + 2a_0 a_1 Var(a_t) + a_1^2 m_4) = 3a_0^2(1 + 2\frac{a_1}{1-a_1}) + 3a_1^2 m_4 \quad (17)$$

Eventually:

$$m_4 = \frac{3a_0^2(1+a_1)}{(1-a_1)(1-3a_1^2)} \quad (18)$$

2.5. Model evaluation criteria

By using two factors, the root mean square error and the Nash-Sutcliffe efficiency, it is possible to find the best model based on the minimum root mean square error (Eq.19) and the maximum Nash-Sutcliffe efficiency coefficient (Eq.20):

$$RMSE = \left[\frac{\sum_{i=1}^{n}(\hat{Q}_i - Q_i)^2}{N}\right]^{0.5} \quad (19)$$

$$CE = 1 - \frac{\sum_{i=1}^{T}(Q_i - Q'_i)^2}{\sum_{i=1}^{T}(Q_i - \bar{Q}_i)^2} \quad (20)$$

In the above equations, $\overline{Q}$، Q'، Q_i are the mean, simulation, and observation values of dew point temperature, respectively and N is the number of data (Khashei-Siuki et al., 2021).

3. Results and Discussion

As mentioned, in this study, the correlation between the studied values and the dew point temperature values of the selected model inputs was investigated at first. The results of the correlation between the mentioned values can be seen in table 2. According to table 2, the parameters that have the highest correlation coefficients with the dew point temperature at all stations studied are mean temperature, maximum temperature, and minimum temperature. These 3 parameters are used as inputs for the studied models.

The next parameters that have the highest correlation with the dew point temperature are potential evapotranspiration and sunshine hours which were not included in the simulation. Also the De Martonne method was used to study and classify the climate in the regions and studied stations. The results of the climate study of different stations are shown in table 3.

Table 2. Correlation of studied data with dew point temperature

Station	Studied parameters						
	potential Evapotranspiration	sunshine	Wind speed	Relative humidity	Mean temperature	Minimum temperature	Maximum temperature
Bam	0.62	0.53	0.18	-0.37	0.78	0.72	0.74
Zabol	0.64	0.62	0.59	-0.38	0.77	0.70	0.74
Tabas	0.50	0.47	0.24	-0.26	0.6	0.55	0.57
Mashhad	0.57	0.50	0.46	-0.37	0.83	0.61	0.65
Kerman	0.50	0.41	0.14	-0.24	0.74	0.52	0.61
Iranshahr	0.60	0.16	0.22	-0.08	0.76	0.75	0.78
Birjand	0.48	0.42	0.36	-0.21	0.69	0.57	0.64
Zahedan	0.54	0.32	-0.05	-0.17	0.67	0.55	0.62

Table 3. The De Martonne index for the studied stations

Station	mean annual rainfall	mean temperature	Climate	De Martonne index
Bam	68	16	Dry	2.615
Birjand	168.5	16	Dry	6.48
Iranshahr	105	25.8	Dry	2.93
Kerman	142	15	Dry	5.68
Mashhad	250	15.7	Dry	9.72
Tabas	80	20	Dry	2.66
Zabol	61	20	Dry	2.03
Zahedan	89	14.95	Dry	3.56

After calculation the correlation, the dew point temperature values were evaluated and simulated using a VAR model Subsequently, the random coefficient, also known as the residual series, obtained from the VAR model was further analyzed using a GARCH model to account for heteroskedasticity. This resulted in the development of a hybrid VAR-GARCH model. To train these models, 80 of the available data for each station was utilized and were then tested using the remaining 20% of the data. In other words, by comparing the predicted values generated by these models with the actual values from the remaining 20% of data, their performance at each station could be assessed. The dataset used spans from January 1983 to December 2021 equating to a total of 457 months. During train phase, out of these months, 367 months representing 80% of the data were used while comparing model outputs with actual values from the remaining 90 months. This approach allows for comprehensive evaluation and validation of both VAR and GARCH models while statistical measures such as RMSE and Nash-Sutcliffe coefficient to assess their accuracy in predicting trends or patterns.

The VAR model analyzes various time lags in order to determine the optimal lag for incorporating data into the model. This feature of the VAR model enhances the model in comparison to linear models that do not take lag into account. Based on the findings, a lag of 3 was identified as the most suitable delay for inputting data from 7 stations into the model, while for the Tabas station, a lag 2 was deemed appropriate for introducing data into the model.

3.1. The results of modeling and simulation of dew point temperature values in the studied stations based on VAR model

Table 4 presents the comparison between the actual dew point temperature data and the model output values during the train and test phases for all studied stations. The model was trained using 80% of the data up until the month of 367, while the remaining data from the month of 367 to 457 was used for test the model's performance. The results showed that the lowest RMSE in the train phase is related to Zabol station (RMSE=1.51 ºC) and in the highest is related to Mashhad station (RMSE=2.29 ºC). In the test phase, the lowest RMSE is related to Bam station (RMSE=1.3 ºC) and the highest is related to Mashhad station (RMSE=3.06 ºC) same as train phase. Regarding the VAR model, the results of Nash-Sutcliffe efficiency showed that in the train phase, the efficiency of the VAR model in all the studied stations is more than 82%, and in the test phase, it is between 48 and 90%, and the lowest of which is related to the Mashhad station (NSE=0.48).

Table 4. The results of simulation of dew point temperature in train and test phases based on VAR model

Station	RMSE (ºC)		Nash-Sutcliffe efficiency	
	Train	Test	Train	Test
Bam	1.69	1.30	0.86	0.90
Birjand	2.02	1.68	0.82	0.82
Iranshahr	1.74	2.10	0.91	0.88
Kerman	2.02	1.70	0.86	0.76
Mashhad	2.29	3.06	0.88	0.48
Tabas	1.54	1.70	0.83	0.61
Zabol	1.51	2.34	0.90	0.64
Zahedan	2.02	1.68	0.83	0.79

3.2. The results of modeling and simulation of dew point temperature values in the studied stations based on VAR-GARCH model

Prior to fitting the GARCH model, an examination was conducted on the structural stability of the residual series using ordinary least square (OLS) regression and Cusum tests. The Cusum test employs a specific method from a generalized statistical framework to calculate the empirical volatility process. The results of this test were also validated based on the shape and confidence intervals. Consequently, the residual series exhibits the necessary stability to be incorporated into the GARCH model and, subsequently, analyze the common frequency of the residual series in VAR model in the next step. Table 5 displays the comparison between the actual dew point temperature data and the model's output values during the train and test phases by using VAR-GARCH model.

Table 5. The results of simulation of dew point temperature in train and test phases based on VAR-GARCH model

Station	RMSE (ºC)		Nash-Sutcliffe efficiency	
	Train	Test	Train	Test
Bam	1.32	0.99	0.94	0.91
Birjand	1.75	1.37	0.87	0.88
Iranshahr	1.42	1.74	0.94	0.92
Kerman	1.86	1.4	0.88	0.84
Mashhad	2.08	3.08	0.9	0.47
Tabas	1.19	1.37	0.9	0.75
Zabol	1.22	3.02	0.93	0.73
Zahedan	1.74	1.53	0.88	0.82

The results showed that the lowest amount of RMSE of VAR-GARCH model in simulation the dew point temperature in the train phase is related to Tabas (RMSE=1.19

ºC) and Zabol (RMSE=1.22 ºC) stations and the highest is related to Mashhad station (RMSE=2.08 ºC), which is similar to VAR model. In the test phase, the lowest error rate is related to Bam station (RMSE=0.99 ºC) and the highest error rate in the test phase according to the RMSE statistics is related to Mashhad station with RMSE= 3.02 degrees Celsius. The results of Nash-Sutcliffe efficiency in the train and test phases also showed that in the train phase, the efficiency of the VAR model in all the studied stations is more than 87%, which is satisfactory. But in the test phase, the efficiency of the model in Mashhad station is lower than other stations and is about 47%. But in other stations, this performance is between 73 and 92 percent.

3.3. Comparison of VAR and VAR-GARCH models

During the train phase, the VAR-GARCH model showed better performance than the VAR model in all investigated stations. In the test phase, the VAR-GARCH model was proposed as the best model in 6 out of 8 stations, namely Bam, Birjand, Iranshahr, Kerman, Tabas and Zahedan. On the other hand, the VAR model showed better performance than the VAR-GARCH model in 2 stations of Zabol and Mashhad. In fact, the RMSE difference between the two models at Mashhad station is less than 1%. In Zabol station, the VAR-GARCH model was able to show better performance in the Nash-Sutcliffe efficiency, while the VAR model provides a lower value in RMSE. In the 8 investigated stations, 6 stations improved their performance by reducing the RMSE between 5% and 30% after developing the residual series by GARCH model.

Dew point temperature values in the studied stations (Bam, Birjand, Iranshahr, Kerman, Mashhad, Tabas, Zabol, Zahedan) were estimated using two different models. VAR model and VAR-GARCH model, which is obtained from expanding the residual series of VAR model with GARCH model. The results of investigation and simulation of dew point temperature values in two train and test phases at Bam station were presented as examples in figure 2, while the remaining stations are presented in the appendix A. Based on the obtained figure, it can be seen that the closer the black dots are to the black line, the higher the data correlation. Red lines represent 95% confidence intervals. If there are more data points or black points outside the range, it indicates a higher error in the model. When comparing the two models, it was found that the VAR and VAR-GARCH models showed higher accuracy and correlation during the train phase than the other two models. In addition, the VAR-GARCH model showed the best correlation and accuracy during the test phase, which is very important. The efficiency and accuracy of the models were fully evaluated in terms of RMSE and model efficiency coefficient (Nash-Sutcliffe) in both train and test phases.

The Nash-Sutcliffe efficiency, which is a measure of model performance, was used to evaluate the models during both the train and test phases. Figure 3 shows the performance of the models during the train phase, while Figure 4 illustrates their performance in the test phase. From our analysis of Figure 3, it is clear that in the train phase, the VAR-GARCH model performs exceptionally well and outperforms all other models tested. Moving on to the test phase, Figure 4 indicates that the VAR-GARCH models exhibit comparable performance. This means that they are able to accurately predict values for this phase as well. On the other hand, it is worth noting that the VAR model displays a higher error rate compared to VAR-GARCH model. This suggests that its predictions may not be as accurate or reliable. One key advantage of VAR-GARCH models is their consistent performance across different stations. This implies that they are robust and reliable in predicting values across various locations.

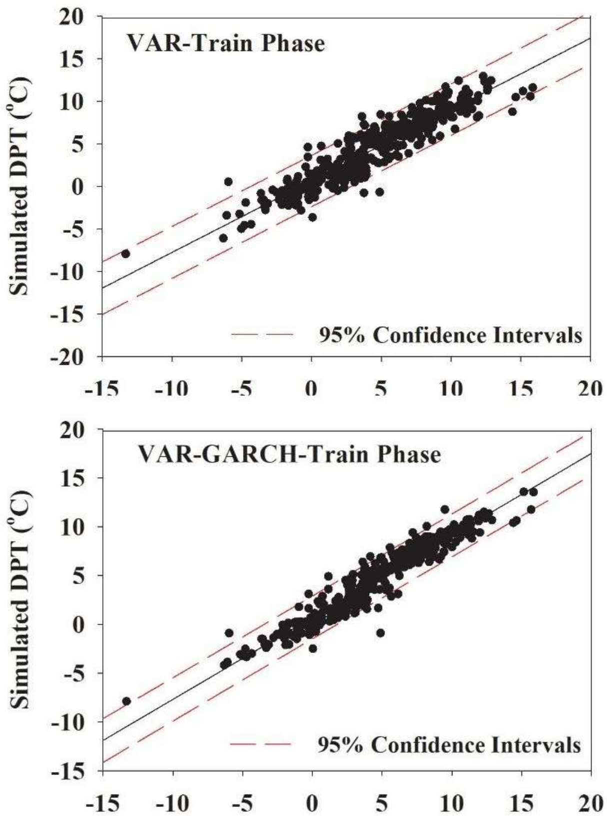

Fig. 2. The results of checking and measuring the accuracy of estimated dew point values in Bam station

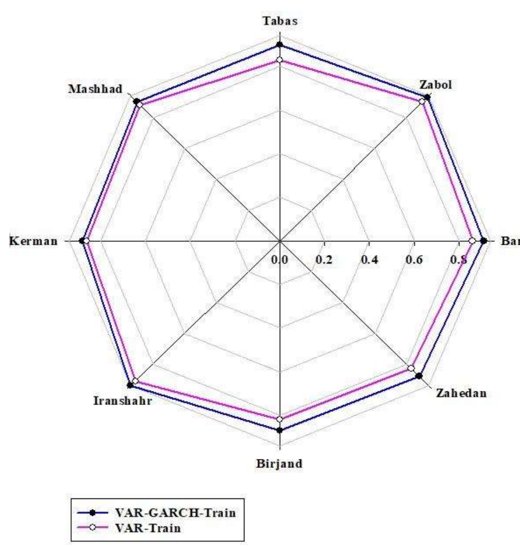

Fig. 3. NSE statistics of studied models in simulation the dew point temperature in the train phase

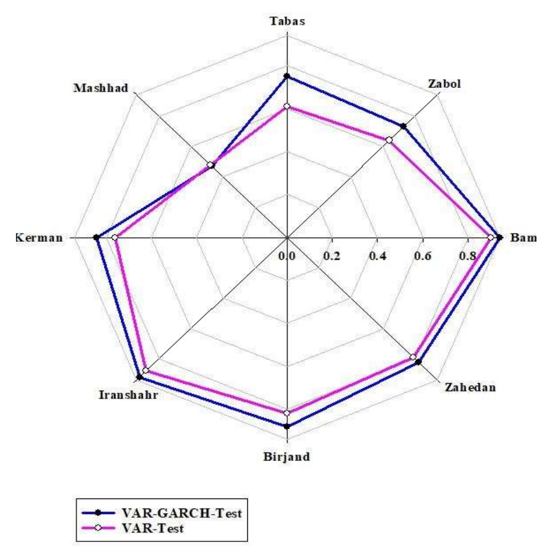

Fig. 4. NSE statistics of studied models in simulation the dew point temperature in the test phase

The results of RMSE statistic as error rate criteria in the train phase were presented in Figure 5, while Figure 6 illustrates the results in the test phase. The RMSE statistic is commonly used to evaluate the accuracy of predictive models, with lower values indicating better performance.

Figure 5 shows that the VAR-GARCH model performs better than VAR models in the train phase and provides the best performance. Also in the test phase, the VAR-GARCH model shows better performance than the VAR model, as shown in Figure 6.

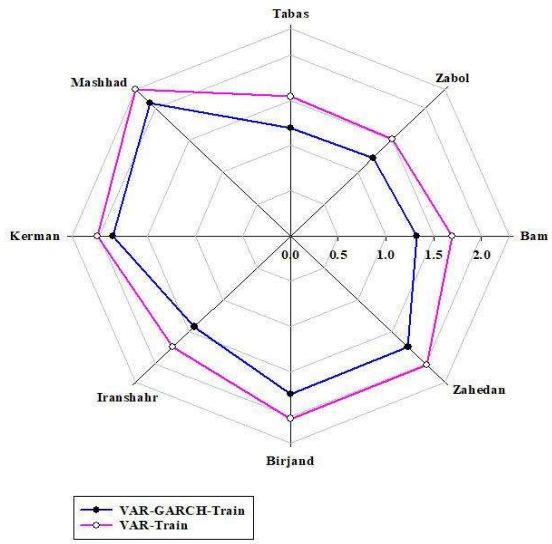

Fig. 5. RMSE statistics of studied models in simulation the dew point temperature in the train phase

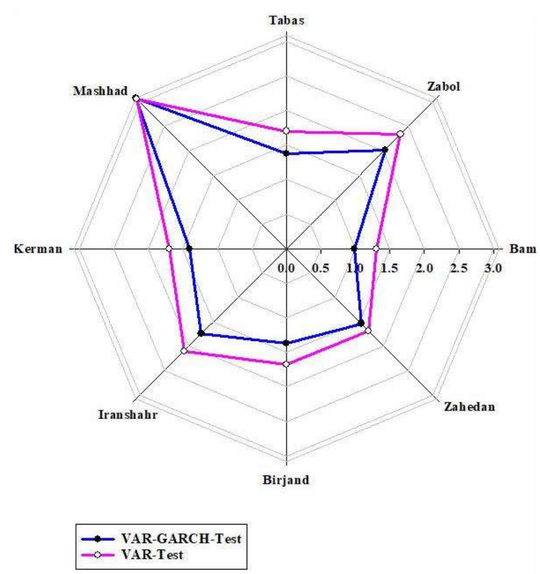

Fig. 6. RMSE statistics of studied models in simulation the dew point temperature in the test phase

However, based on the output of the Bam station in the test phase, the VAR model shows a higher error difference (nearly 30%) than the VAR-GARCH model. As a result, we can predict more reliable output from VAR-GARCH model. These outputs have an improvement of 5% to 30% compared to the same outputs from the VAR model.

In addition to the model evaluation statistics (RMSE) and (NSE), Taylor's diagram and violin plot were also used to evaluate the performance of the two studied models. Figure 7 shows the Taylor diagram, while Figure 8 shows the violin plot, which shows the similarity of the time series for the Bam station. It is evident from Figure 7 that the VAR-GARCH model shows higher certainty and correlation compared to VAR models. On the other hand, Figure 8 shows that the VAR-GARCH model successfully simulates the quartiles of the data, but has difficulty predicting the minimum and maximum dew point temperatures. In fact, none of the two models investigated in this station were able to accurately predict the minimum and maximum data values.

Overall the results showed that the VAR-GARCH model has the lowest error in simulation the dew point temperature in Bam station (RMSE: 0.99 degrees Celsius), and is considered the best model, and the VAR model is ranked next. Also, the efficiency criterion (NSE) showed that VAR-GARCH and VAR models have acceptable efficiency, but the VAR-GARCH model is the best model in bam station.

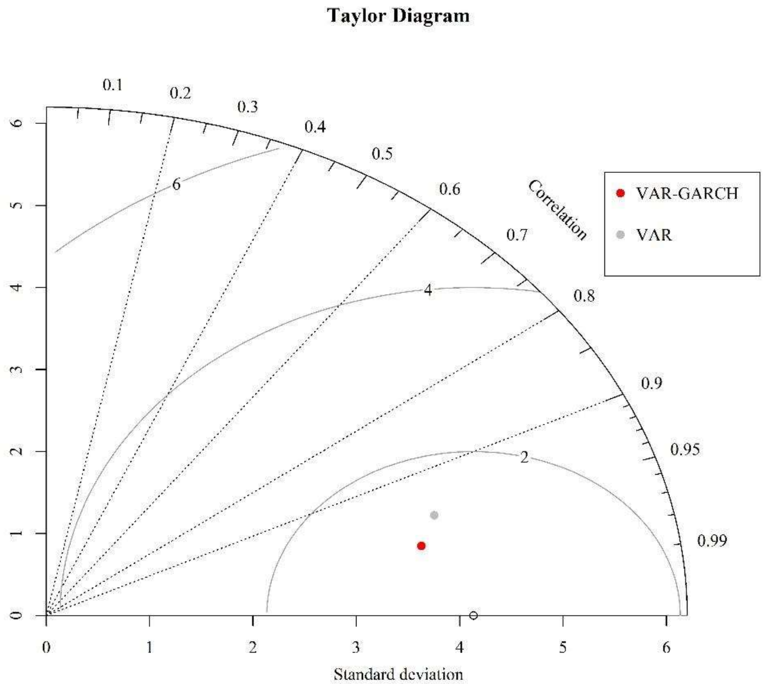

Fig. 7. Taylor diagram of Bam station in the test phase

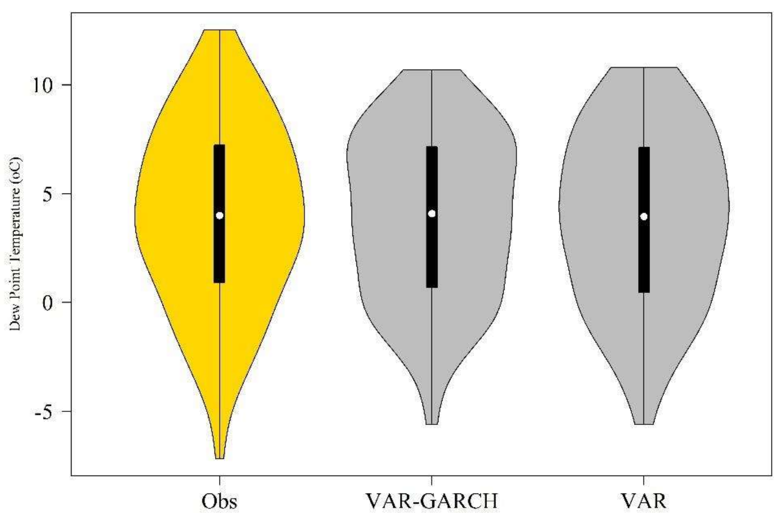

Fig. 8. Violin diagram (similarity of time series) of BAM station in the test phase

4. Conclusion

The dew point temperature values in different climates of Iran were estimated using meteorological data from 8 stations, eastern of Iran. These stations were classified as dry based on the De Martonne index. The input for estimating dew point temperature values consisted of meteorological data from 1983-2021, specifically maximum temperature, minimum temperature, and mean temperature, which had the highest correlation with dew point. Initially, VAR method was employed in this research. Subsequently, the residual series of VAR model was developed using the ARCH model, and VAR-GARCH model was generated. The accuracy of the estimated values at each step was evaluated using the RMSE and the Nash-Sutcliffe model efficiency coefficient. The evaluation of the studied models showed that the VAR-GARCH model outperformed the VAR model in both the train and test phases. This superiority can be attributed to the VAR-GARCH model's ability to model the non-linear component and Residual Series. However, it was observed that the VAR-GARCH model exhibited significant superiority over the VAR model during the train phase. The VAR-GARCH model yielded the lowest RMSE and the highest NSE for the Bam station. Specifically, this model achieved RMSE value of 0.99 degrees Celsius and a NSE of 0.94 for this station, which is considered the best model among all the stations and models examined.

The number of input parameters for the 3-variable model depends on the type of temperature. Generally, a model that can generate accurate results with fewer variable inputs tends to have higher efficiency compared to a model that relies on numerous parameters for estimation. Based on the aforementioned findings, the investigations conducted among the 2 models indicate the exceptional performance of the VAR-GARCH model during the train and test phases, along with its superiority in most stations due to its lower RMSE and higher NSE. Consequently, this model can be regarded as the best model investigated in this research. The superior performance of the VAR-GARCH model highlights its potential as a valuable tool for forecasting dew point values in dry climate regions. This research contributes to advancements in meteorology by providing more accurate prediction techniques, which can support various applications such agriculture, urban planning, and water resource management. Further research could explore alternative modeling approaches or investigate additional that may influence dew point temperatures. Additionally, expanding this analysis beyond dry climate regions could provide insights into how different climatic conditions impact dew formation and contribute to more comprehensive forecasting models. It is worth noting that this study solely on investigating this specific weather parameters' relationship with dew temperature. Future research could explore additional factors or refine existing models to further enhance accuracy and improve predictions of dew point values across different climatic conditions or geographical areas within Iran.

5. Disclosure Statement

No potential conflict of interest was reported by the authors

6. References

Aguirre-Gutiérrez, C. A., Holwerda, F., Goldsmith, G. R., Delgado, J., Yepez, E., Carbajal, N., ... & Arredondo, J. T. (2019). The importance of dew in the water balance of a continental semiarid grassland. *Journal of Arid Environments*, *168*, 26-35.

Baguskas, S. A., King, J. Y., Fischer, D. T., D'Antonio, C. M., & Still, C. J. (2016). Impact of fog drip versus fog immersion on the physiology of Bishop pine saplings. *Functional plant biology*, *44*(3), 339-350.

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., & Whitelaw, R. F. (1998). The econometrics of financial markets. *Macroeconomic Dynamics*, *2*(4), 559-562.

Culbertson, K. (1996). Quantitative Financial Economics: Stocks, Bonds and Foreign Exchange. John Wiley and Sons, Chichester.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 987-1007.

Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press, Princeton.

Khashei-Siuki, A., Shahidi, A., Ramezani, Y., & Nazeri Tahroudi, M. (2021). Simulation of potential evapotranspiration values based on vine copula. *Meteorological Applications*, *28*(5), e2027.

Lin, J., Shahzad, M. W., Li, J., Long, J., Li, C., & Chua, K. J. (2021). A robust physics-based model framework of the dew point evaporative cooler: From fundamentals to applications. *Energy Conversion and Management*, *233*, 113925.

Lutkepohl, H. (1991). Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin. *11.7* References 427

Lutkepohl, H. (1999). "Vector Autoregressions," unpublished manuscript, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

Mills, T.C. (1999). The Econometric Modeling of Financial Time Series, Second Edition. Cambridge University Press, Cambridge.

Nazeri-Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2022). Bivariate simulation of potential evapotranspiration using copula-GARCH model. *Water Resources Management*, *36*(3), 1007-1024.

Ramezani, Y., Nazeri Tahroudi, M., De Michele, C., & Mirabbasi, R. (2023). Application of copula-based and ARCH-based models in storm prediction. *Theoretical and Applied Climatology*, *151*(3), 1239-1255.

Salas, J. D. (1980). *Applied modeling of hydrologic time series*. Water Resources Publication.

Shahidi, A., Ramezani, Y., Nazeri-Tahroudi, M., & Mohammadi, S. (2020). Application of vector autoregressive models to estimate pan evaporation values at the Salt Lake Basin, Iran. *IDŐJÁRÁS/QUARTERLY JOURNAL OF THE HUNGARIAN METEOROLOGICAL SERVICE*, *124*(4), 463-482.

SIMS, C. I. A. (1980). ECONO MI ETRICA. *Econometrica*, *48*(1), 1-48.

Tomaszkiewicz, M., Abou Najm, M., Zurayk, R., & El-Fadel, M. (2017). Dew as an adaptation measure to meet water demand in agriculture and reforestation. *Agricultural and Forest Meteorology*, *232*, 411-421.

Tsay, R. S. (2005). *Analysis of financial time series*. John wiley & sons.

Waggoner, D. F., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics*, *81*(4), 639-651.

Watson, M. W. (1994). Vector autoregressions and cointegration. *Handbook of econometrics*, *4*, 2843-2915.

Appendix A: The results of evaluation the accuracy of estimated dew point temperature values in studied stations

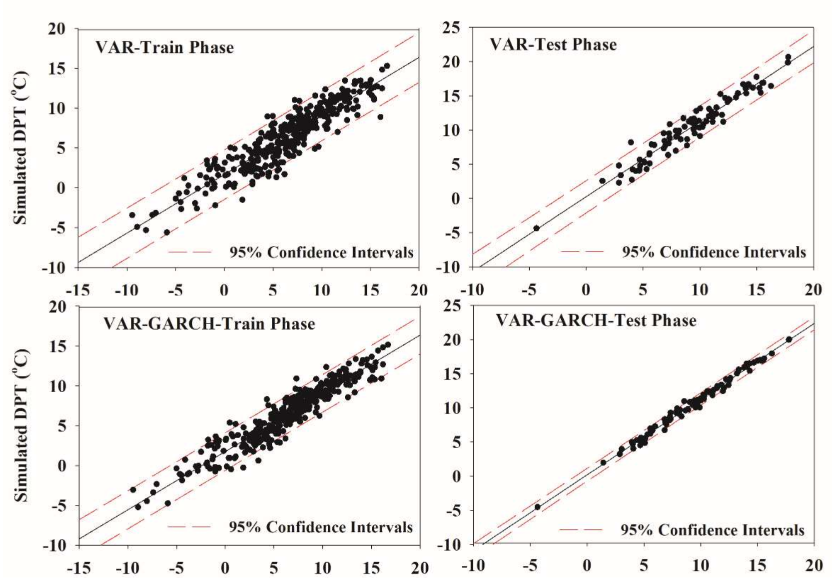

Fig. A.1. The results of evaluation the accuracy of estimated dew point temperature values in Birjand station

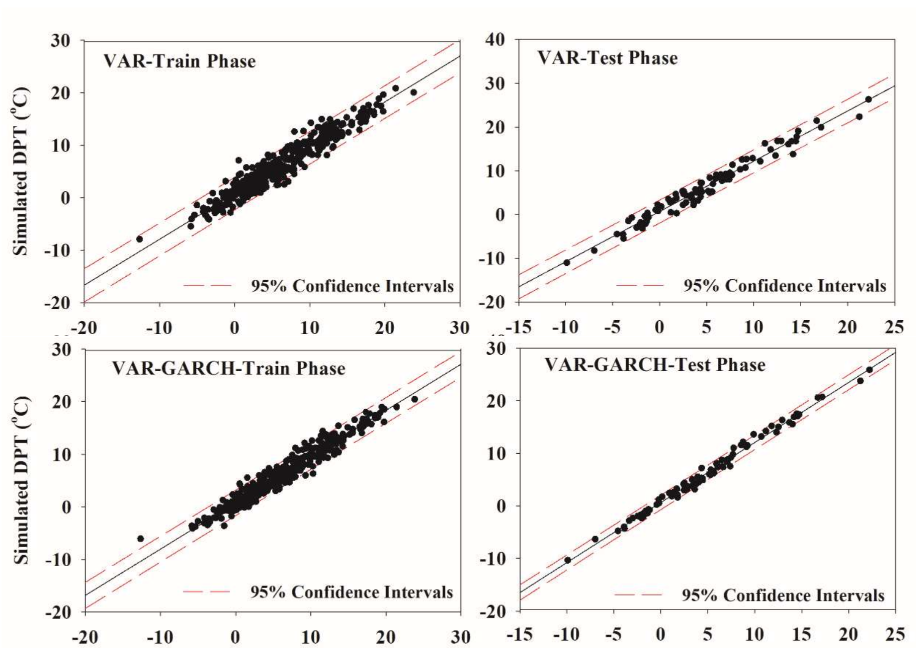

Fig. A. 2. The results of evaluation the accuracy of estimated dew point temperature values in Iranshahr station

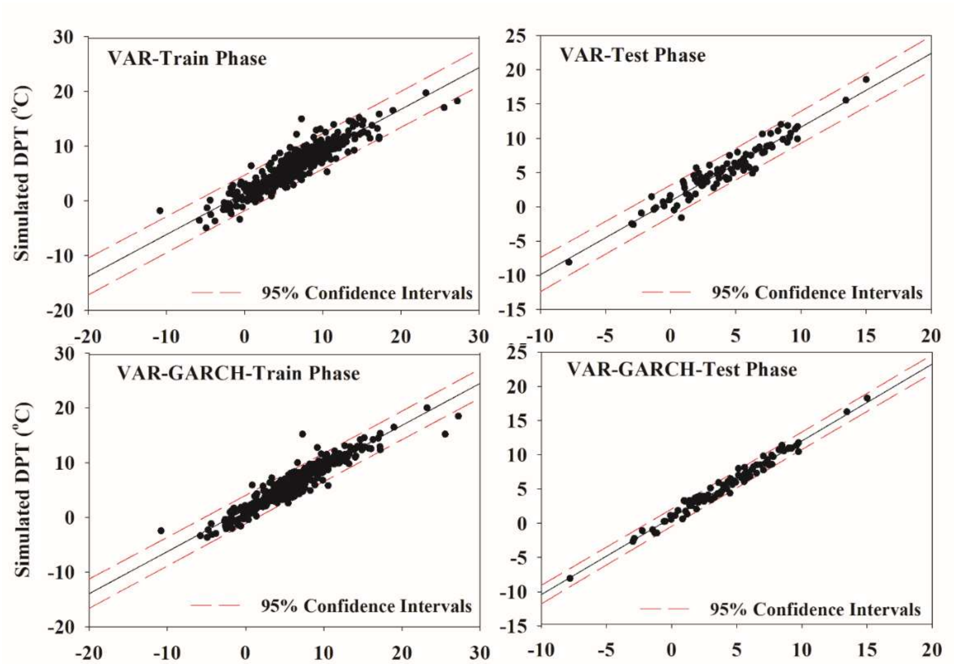

Fig. A. 3. The results of evaluation the accuracy of estimated dew point temperature values in Kerman station

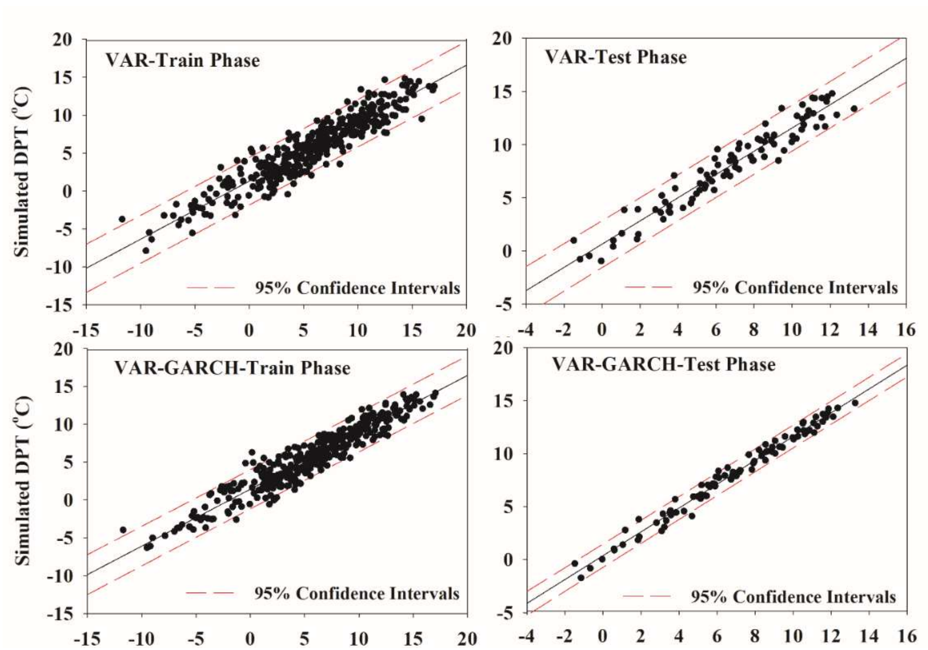

Fig. A. 4. The results of evaluation the accuracy of estimated dew point temperature values in Mashhad station

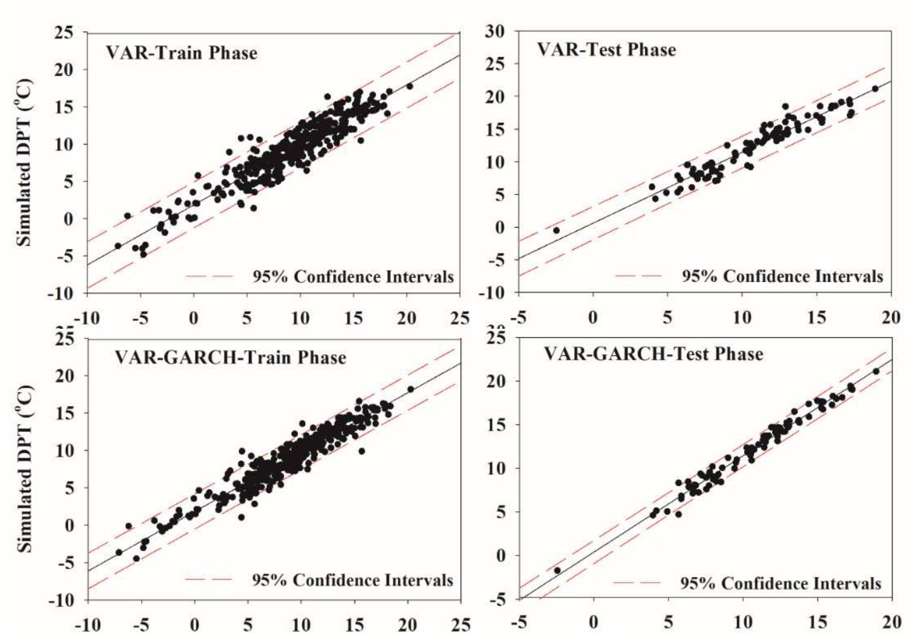

Fig. A. 5. The results of evaluation the accuracy of estimated dew point temperature values in Tabas station

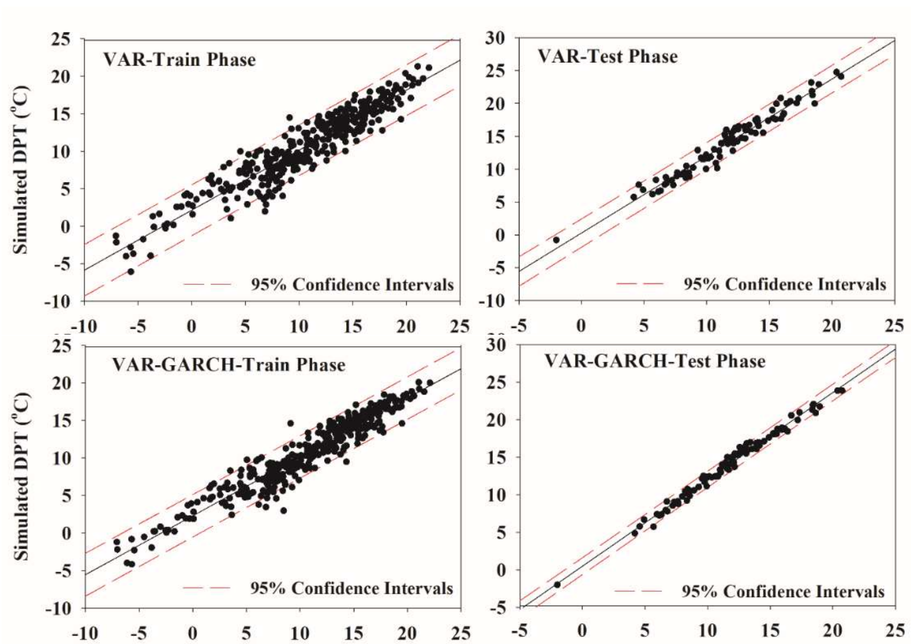

Fig. A. 6. The results of evaluation the accuracy of estimated dew point temperature values in Zabol station

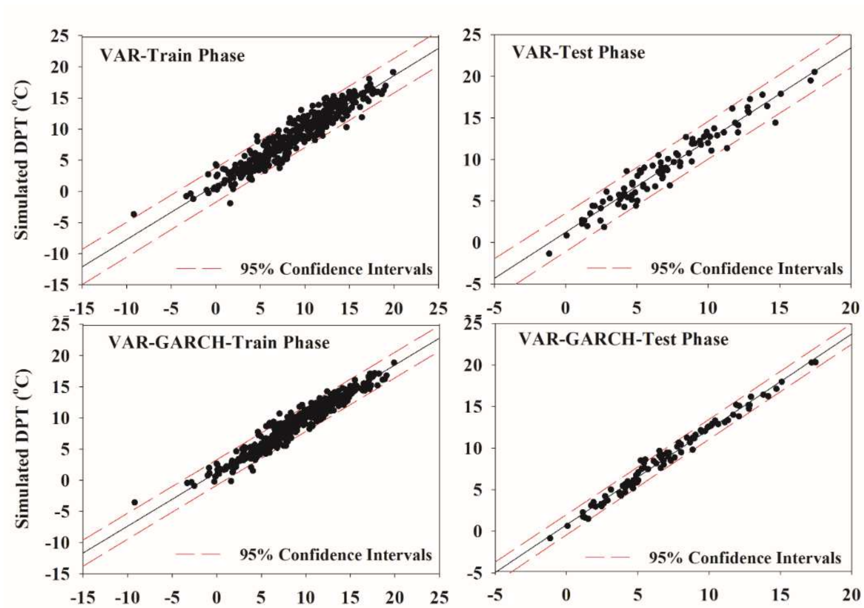

Fig. A. 7. The results of evaluation the accuracy of estimated dew point temperature values in Zahedan station