Relative Humidity Forecasting as an Influential Factor in Hydro-climatology

Yaser Sabzevari[1,2*], Saeid Eslamian[1], Saeid Okhravi[2]

1- Department of Water Science and Engineering, College of Agriculture, Isfahan University of Technology, 8415683111, Iran
2- Institute of Hydrology, Slovak Academy of Sciences, Dúbravská cesta 9, 84104, Bratislava, Slovakia
(*corresponding author E-mail: Y.sabzevari@ag.iut.ac.ir, sabzevari@uh.savba.sk)

Abstract

This study explores the forecasting of relative humidity (RH) as a significant factor in hydro-climatology using the ARIMA(0,1,1) model. The research utilizes historical RH data from 1993 to 2022, coupled with the Mann-Kendall test for trend analysis and time series techniques, to evaluate both past and future RH trends. The results indicate that while RH experienced a decreasing trend over the past three decades, this trend was not statistically significant because of Z statistical equal= -0.01, which is lower than the threshold value of 1.96. Further analysis using the ARIMA model forecasts a slight increase in RH in the coming years. These findings suggest that, despite the absence of significant past trends, RH may show modest upward changes, which could influence future hydro-climatic conditions. The study highlights the importance of incorporating RH forecasting into hydro-climatological models for more accurate predictions of climate behavior and water resource management. Additionally, the research underscores the utility of time series analysis in understanding long-term climate variables and their implications for environmental planning.

Keywords: Relative Humidity, Mann-Kendall, Time series, ARIMA.

Introduction

Climate change is one of the major environmental challenges of the present era [1]. Among the key climatic parameters, atmospheric humidity plays a crucial role in defining weather conditions, agriculture, human health, and various economic activities [2]. Accurate prediction of relative humidity can significantly improve decision-making processes in water resource management, the control of humidity-related diseases, and the optimization of industrial operations [3]. Several approaches have been developed for predicting relative humidity, including physical models, statistical methods, and artificial intelligence techniques [4]. Among these, time series analysis has gained considerable attention as one of the most widely used statistical methods for climate data prediction. This method is highly effective due to its ability to model recurring patterns and identify past and future trends in climatic data [5].
Time series models, such as Autoregressive Moving Average (ARMA) and Seasonal Autoregressive Moving Average (SARIMA), have been successfully applied to predict climatic variables such as temperature, precipitation, and humidity [6]. Prediction of climatic parameters, including atmospheric humidity, has always attracted the attention of researchers

in fields such as meteorology, agriculture, and water resource management. In this context, time series analysis serves as an effective tool for modeling and forecasting complex climatic behaviors [5]. Classical time series models, such as ARMA and SARIMA, were among the earliest methods used for climate data prediction. Due to their simple structure and ability to analyze seasonal patterns, these models have been widely applied in many early studies [6].
Zhang et al. [4] applied the SARIMA model to predict relative humidity in coastal regions of China and achieved acceptable accuracy for short-term intervals. Kurnianingsih et al. [7] proposed a hybrid model combining Long Short-Term Memory (LSTM) networks and Extreme Learning Machines (ELM) for relative humidity forecasting. The results indicated that the LSTM-ELM hybrid model outperformed standalone models in terms of accuracy. Douville and Willett [8] projected 21st-century near-surface relative humidity using gridded temperature and humidity data. They found that relying only on temperature yields weakly constrained results, stressing the need for mitigation strategies beyond global warming targets. Lenderink et al. [9] evaluated the reliability of convection-permitting climate models (CPMs) compared to regional climate models (RCMs) in simulating humidity. The study concluded that CPMs offer more reliable projections of future convective rainfall extremes and provide useful benchmarks for improving model performance. Ahn et al. [10] assessed the ability of transformer-based and RNN-based models to forecast greenhouse conditions, including relative humidity. The study highlights both the potential and current limitations of advanced deep learning architectures for time-series forecasting in greenhouse environments. Wang et al. [11] introduced an artificial neural network (ANN) approach to correct relative humidity over ice in ECMWF forecasts for the upper troposphere and lower stratosphere (UTLS). The method enhanced UTLS humidity representation, improved contrail cirrus predictions, and showed strong potential for broader applications in weather modeling, aviation, and climate research. Similarly, Shad et al. [12] used the SARIMA model for monthly relative humidity prediction in Delhi, India. The process involved model identification, parameter estimation, diagnostic checking, and forecasting, with SARIMA demonstrating good predictive performance for relative humidity. Hutapea et al. [13] employed a machine learning approach using the LSTM model to forecast relative humidity. The LSTM model was trained using two years of climatic data, and the results showed that it provided high accuracy in predicting relative humidity.
A review of the above studies highlights a clear research gap. Although AI-based models such as LSTM, ELM, and Transformer architectures have recently achieved remarkable accuracy in forecasting relative humidity, most investigations have been limited to tropical, coastal, or controlled greenhouse environments. In contrast, long-term forecasting of relative humidity in high-altitude, semi-arid regions has received little attention despite its importance for hydro-climatological studies. This underexplored context poses unique challenges such as data scarcity, high interannual variability, and strong seasonal contrasts. The novelty of the present study lies in addressing this gap by employing a classical yet robust ARIMA framework on more than three decades of historical records from the mountainous and semi-arid climate of Aligoudarz. Unlike previous works that either relied on short-term datasets or focused on humid regions, this study demonstrates how ARIMA can effectively capture the temporal variability of relative humidity under challenging climatic conditions. The findings provide not only methodological insights into the applicability of traditional time-series models in data-limited environments but also valuable implications for climate adaptation and water resource management in arid and semi-arid settings.

Materials and Methods

Study area

Aligoudarz is a city located in the southwestern part of Iran, positioned at 49°42' east longitude and 33°24' north latitude, encompassing an area of 130 hectares. The region is situated at an altitude of 2022 meters above sea level, with an average annual rainfall of 387.7 mm and an annual evaporation rate of 2048.2 mm. According to the Demartin and Amberje climate classification methods, Aligoudarz has a semi-arid and arid-cold climate, characterized by mild summers and very cold winters. Fig. 1 illustrates the geographical location of Aligoudarz within Iran. The city features a synoptic station, from which relative humidity data spanning from 1992 to 2023 were collected to predict its variations.

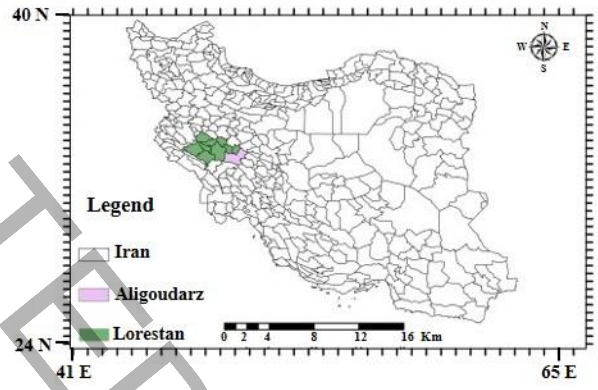


Fig. 1. Geographical location of Aligoudarz city in Iran

Mann-Kendall Test

The Mann-Kendall test is used to check the time trend for each set of data. This test is based on non-parametric linear regression logic. The results of this test show whether there is a significant increase or decrease trend at a certain level of confidence in the time series trend of the parameter under study. Using Mann-Kendall non-parametric test is not sensitive to the normality of the data. The Mann-Kendall test was first proposed by Mann, [11] and then developed by Kendall, [14]. The use of this method was recommended by the World Meteorological Organization. One of the strengths of the Mann-Kendall method is that it is suitable for time series that do not follow a specific distribution [15].

The Mann-Kendall test is used to detect trends in time series data. For the gth time step (month or year) at the kth station, the test statistic S_{gk} is computed as:

$$S_{gk} = \sum_{i}^{n-1} \sum_{j=i+1}^{n-1} \operatorname{sgn}\left(X_{jgk} - X_{igk}\right), Xigk \leq i < j \leq n \tag{1}$$

where n is the number of observations in the time series, and sgn(θ) is the sign function defined as:

$$S_{gn} = \begin{cases} if, \theta > 0, 1 \\ if, \theta = 0, 0 \\ if, \theta < 0, -1 \end{cases} \tag{2}$$

When n≤10, S_{gk} is approximately normally distributed with a mean of 0 and variance:

Kendall and I showed that when n≥ 10, the S statistic is distributed almost normally and has a mean of 0 and the following standard deviation:

$$\left(\sigma_{gg}\right)_{k}=\frac{\left[n\left(n-1\right)\left(2n+5\right)-\sum d\left(d-1\right)\left(2d+5\right)\right]}{18} \tag{3}$$

where d represents the number of tied values in the series. The standardized test statistic Z_{gk} is then calculated as:

$$Z_{gk}=\begin{cases} if, sgk>0, \dfrac{sgk-1}{\delta gk} \\ if, sgk=0, 0 \\ if, sgk<0, \dfrac{sgk+1}{\delta gk} \end{cases} \tag{4}$$

Under the null hypothesis of no trend, Z_{gk} follows a standard normal distribution. If $|Zgk|<1.96$ (95% confidence) or $|Zgk|<2.58$ (99% confidence), the trend is considered significant. Change points were visually identified using the intersection of U and U′ plots, as recommended by Kendall [14] and supported by trend visualization techniques [16].

Time Series (TS) Analysis

When analyzing a time series, the objectives can vary and are typically categorized into description, prediction, control, and pattern fitting [17]. Plotting data over time allows for the observation of changes in the statistical parameters of the series, such as mean, standard deviation, and skewness [18]. The explanation of the TS includes the detection of its stationarity and non-stationary and the investigation of the autocorrelation of the series. A time series is considered stationary when its mean and variance remain constant over time, and any strong periodic fluctuations have been eliminated. Non-stationary series can be converted into stationary series by applying transformations such as differencing to stabilize the variance [19, 20].

To investigate the stationarity of a TS, it is essential to calculate statistical properties such as mean and variance, as modeling accuracy relies on stationary data. The methods for this analysis include the autoregressive model, which is based on the principle where the value at a given time is regressed on its previous values, and the moving average model, where the value at time *t* is estimated using the random value at that moment plus a weighted sum of the random values from previous times. Some processes exhibit both autocorrelation and moving average characteristics; in such cases, combined models such as ARIMA are employed [21].

The process of autocorrelation is the dependence between consecutive values in terms of time. A function that calculates autocorrelation based on the time interval between observations is known as the Autocorrelation Function (ACF) [22]. Autocorrelation can be considered the similarity between observations. The autocorrelation function is also used to measure this similarity. The ACF function has been used to check autocorrelation using Minitab software.

A TS that has a stationary property should be analyzed. A TS is recognized as stationary if its statistical properties such as mean and variance are constant over time. In TS analysis, before modeling, the series should be made stationary. To achieve this, methods for assessing the stationarity of TS data are explored within the framework of a seasonal ARIMA model, which integrates both non-seasonal and seasonal components in a multiplicative structure. Each element of the ARIMA model functions as a parameter, following a standard notation. Specifically, ARIMA models are typically denoted as ARIMA (*p*, *d*, *q*), where the parameters *p*, *d*, and *q* are substituted with integer values to specify the particular form of ARIMA model employed. The parameters are defined as follows:

p: Represents the number of lagged observations included in the model, commonly referred to as the lag order.
d: The number of times the raw observations are differenced, also referred to as the degree of differencing.
Q: The size of the moving average window, also known as the order of the moving average.

$$\varphi(B)Z_t = \varphi(B)(1-B)Z_t = \theta(B)at \tag{5}$$

In Eq. (1), Z_t represents the experiential series, $\varphi(B)$ denotes the multinomial of order *p*, and $\theta(B)$ represents the multinomial of order *q*. For instance, a linear regression model would specify the amount and kind of terms to include. A parameter value of zero indicates that the corresponding component should be excluded from the model. The selection of parameters must satisfy the conditions of stationarity and invertibility, particularly about moving averages and autocorrelations. The significance of the parameters should be tested, which involves assessing the error terms associated with the estimates and calculating the corresponding *t*-values [5]. If θ is the point estimate of the parameter of interest and S_θ is the standard error of this estimate, the *t*-value is calculated by $t = \theta / S_\theta$. If the null hypothesis (i.e., the parameter is equal to zero) is rejected with a significance level of α = 0.05 or greater, the parameter is considered significant and is retained in the model.

Fit-adequacy tests are essential for evaluating the accuracy of the models. To assess the validity of the models fitted to the data, the residuals were examined for normality and autocorrelation using tools such as the Quantile-Quantile Plot (QQ Plot), Shapiro-Wilk test, and Kolmogorov-Smirnov test [21, 22]. In this analysis, SPSS and Minitab software were utilized to assess the normality and homogeneity of the data, along with t-statistics (T), P-values, and the Bayesian Information Criterion (BIC) to evaluate the relationship between observed and predicted data.

Prior to fitting the ARIMA model, the distribution of the observed relative humidity (RH) time series was examined using both the Shapiro-Wilk and Kolmogorov-Smirnov normality tests. The results showed no significant deviation from normality (Shapiro-Wilk p = 0.21; K-S p = 0.17), confirming that the data were appropriate for parametric time series modeling.

To ensure the appropriateness of the model, two complementary methods are employed [23]:

1. Residual analysis of the fitted model: This method assesses whether the residuals are random and uncorrelated. It includes evaluating assumptions such as the normality of residuals, constancy of variance, and independence of residuals, as well as plotting the residuals against time. These assumptions are tested using the Pert-Manto test.
2. Analysis of Models with Additional Parameters: This method examines models with more parameters to determine if they provide a better fit.

In residual analysis, it is essential to verify the assumptions of normality, constant variance, and independence. The normality assumption is considered valid if the residuals align approximately along a straight line and exhibit a uniform distribution. The Pert-Manto test, which utilizes a modified Box-Pierce statistic, offers a formal approach to testing the hypothesis that the residuals are uncorrelated. This test is represented by Eq. (6):

$$Q(LBQ) = n(n+2)\sum_{h=1}^{k}(n-k)^{-1} p_h^2 \tag{6}$$

where *n* is the number of observations, *Q* is the test statistic (a modified version of the Ljung-Box statistic). Under the null hypothesis, it follows an approximate Kido distribution (because it is for positive values and its capability for statistical significance). The first condition states that if the value of the *Q* statistic exceeds the critical value from the chi-squared distribution, the null hypothesis is rejected, indicating that the data are correlated. The second condition requires that the correction index value must also be greater than the significance level α.

Results and Discussion

Temporal Trend Analysis Using the Mann-Kendall Test

To analyze the temporal trend of the relative humidity time series in the study area, the non-parametric Mann-Kendall test was employed. Table 1 presents the Mann-Kendall statistic for the relative humidity time series in Aligoudarz. According to the table, a slight negative decreasing trend was observed in the Aligoudarz climate. However, since the Z-statistic value is -0.01, which is less than 1.96, this decrease is not statistically significant at any confidence level.

Table 1. Mann-Kendall statistic values at the study station

Mann-Kendall statistic		RH
	Z	-0.01

The following graph illustrates the temporal trend of relative humidity in Aligoudarz (Figure 2). Figure 2 depicts the changes in the variable during the statistical years under study. In these graphs, the vertical axis represents the variable values (U), while the horizontal axis shows U'. To obtain U', it is sufficient to plot the variable values in descending order of the years. If the two curves U and U' intersect in these graphs, it indicates a significant change in the variable during that particular year, referred to as a jump point. Conversely, parallel curves indicate no significant change during the study period. The results of Figure 2 confirm the Z-statistic findings, showing a slight decreasing trend at the Aligoudarz station. According to Figure 2, a noticeable decrease in relative humidity occurred at Aligoudarz station in 2009, while a significant increase was observed in 2015. These abrupt changes, or jump points, can have considerable impacts on the overall trend of drought variability in the region. Potential causes for these shifts may include regional climatic factors, such as unusual precipitation patterns, temperature extremes, or rapid snowmelt events, which directly influence atmospheric moisture. In addition, anthropogenic factors, including land-use changes, expansion of agricultural activities, or water management interventions, may also have contributed to these variations. These results highlight the importance of considering both climatic and human-induced influences when interpreting temporal changes in relative humidity and their effects on drought conditions during the studied period.

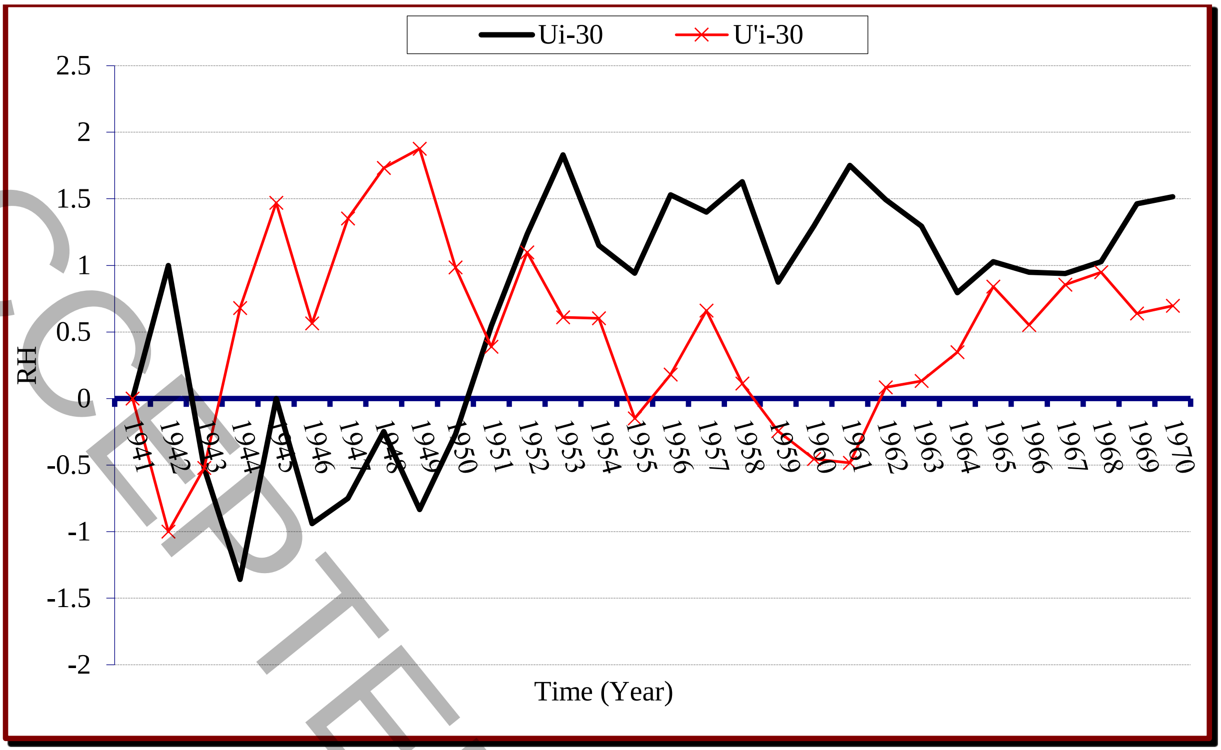


Fig. 2. Mann-Kendall graph of Relative Humidity

Future Forecasting with Time Series

This section presents the results of ARIMA model analysis for predicting the relative humidity trend over the next ten years. To simulate and predict future values using the ARIMA model, an optimized ARIMA model must first be developed. Several essential factors should be evaluated to establish an optimal model, including identifying the presence of a trend in the data, examining autocorrelation among the data, and assessing stationarity in both variance and mean. The first step in the process involves checking for the presence of a trend in the data using a Time Series Plot (Figure 3). This visual analysis helps determine whether the data exhibit a trend that needs to be addressed before fitting the ARIMA model.

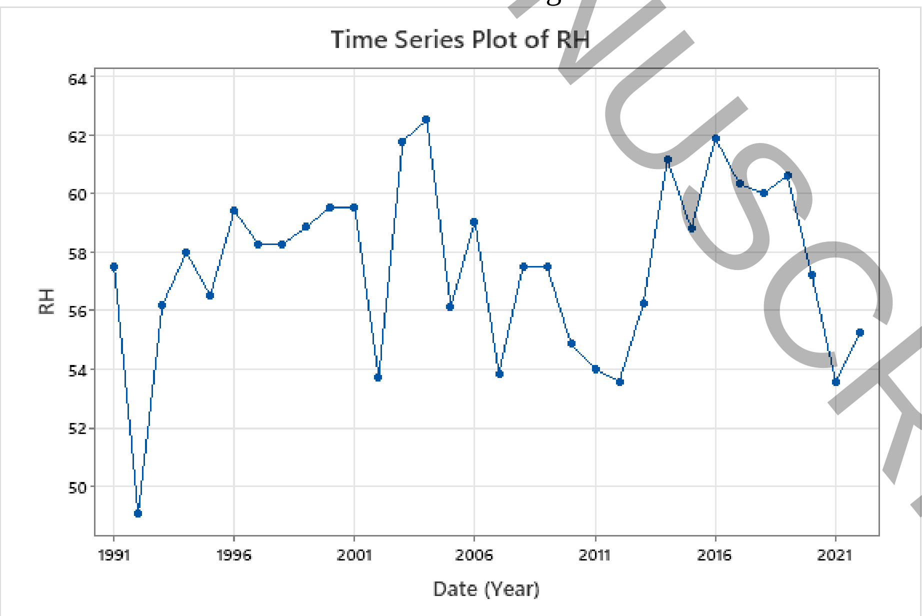


Fig. 3. Time Series Plot of Observational Data

To better analyze the data trend, a Smoother was applied to the Time Series Plot (Figure 4). The red line represents the overall trend, indicating fluctuations across the years. Initially, the trend shows a slight increase, followed by a decline during the mid-period (2001–2007), and finally continues to decrease toward the end of the study period. The original data points (blue dots) reveal significant annual variations, which may suggest the influence of factors such as climate change or short-term phenomena affecting relative humidity. In this trend, Critical Periods include: Lowest Values: The lowest relative humidity levels were observed around 1992, and 2019 and Highest Values: Relative humidity peaked during the mid-period in 2002 and 2011. Given the arid and semi-arid climate of the study area, these fluctuations in relative humidity can have significant implications for water resource management and agriculture. Monitoring periods of reduced humidity is essential for developing effective drought management strategies.

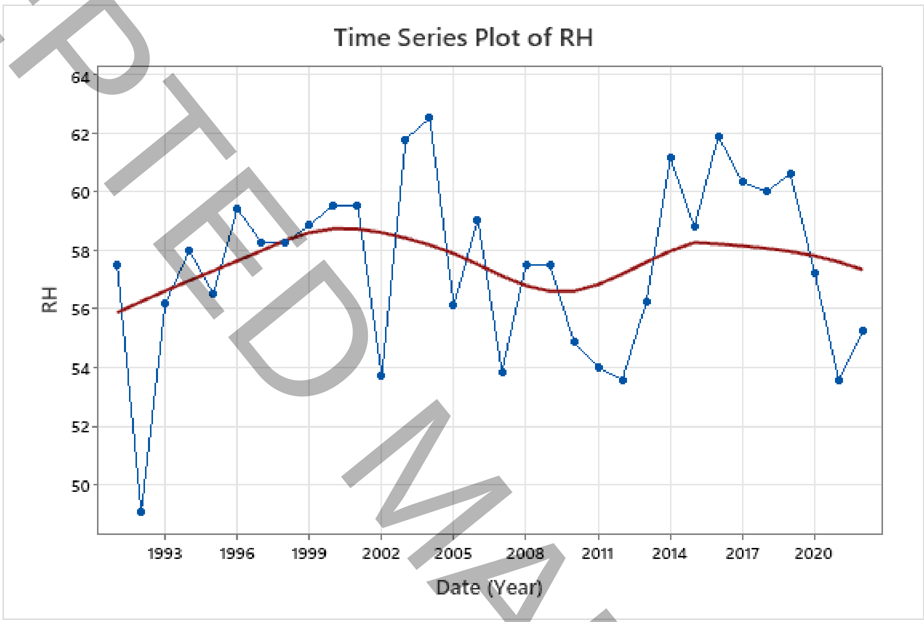


Fig. 4. Smoother of Observational Data

To assess the autocorrelation within the time series data during the observation period, ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots were generated. Figure 5 and Table 2 show the ACF plot results for the relative humidity time series data. Autocorrelation evaluates the relationship between time series data at different lags. In these plots, if the autocorrelation bars extend beyond the critical bounds (red lines), it indicates significant correlation at that specific lag. At Lag 1, a strong positive autocorrelation is observed. In some other lags, negative values appear, but most of these values remain within the non-significant range. This indicates that the time series data likely exhibit a weak seasonal or oscillatory pattern, though this pattern does not consistently repeat across all lags. If high autocorrelation is observed in the initial lags, it may suggest the presence of non-stationarity in the time series.

Table 2. Autocorrelation results of the Relative Humidity results

Lag	ACF	T	LBQ
1	0.253728	1.44	2.26
2	0.067493	0.36	2.42
3	0.061971	0.33	2.57
4	-0.197665	-1.04	4.09
5	-0.060510	-0.31	4.23
6	-0.274779	-1.40	7.39

7	-0.314697	-1.51	11.70
8	-0.225172	-1.01	14.00

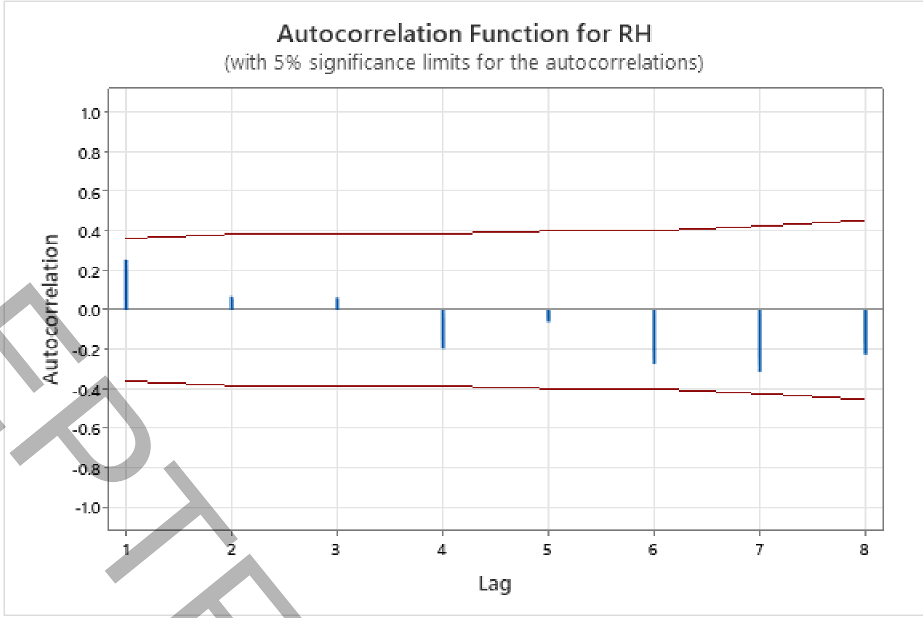


Fig. 5. Autocorrelation of the Relative Humidity time series

Figure 6 and Table 3 show the Partial Autocorrelation Function (PACF) results for the time series data of relative humidity. Significant correlations are observed at Lag 1 and Lag 4. At other lags, the partial autocorrelation is close to zero. This pattern may indicate the influence of an autoregressive (AR) process of order 1 or 4. Models such as AR(1) or AR(4) could be suitable for modeling. On the other hand, the blue lines representing the upper and lower significance thresholds (red lines) are not crossed. Therefore, the assumption of the parameter P being equal to zero can be considered. Although autocorrelation exists between the data, it is not significant at any of the lags. As shown in the figure, T for all lags falls within the range of -1.96 to +1.96, indicating non-significance at the 95% confidence level. The partial autocorrelation is high at the initial lags but decreases rapidly, which may suggest relative stationarity or the need for first-order differencing to achieve full stationarity.

Table 3. Partial Autocorrelation results of Relative Humidity results

Lag	PACF	T
1	0.253728	1.44
2	0.003329	0.02
3	0.047090	0.27
4	-0.240693	-1.36
5	0.052456	0.30
6	-0.301142	-1.70
7	-0.161419	-0.91
8	-0.203563	-1.15

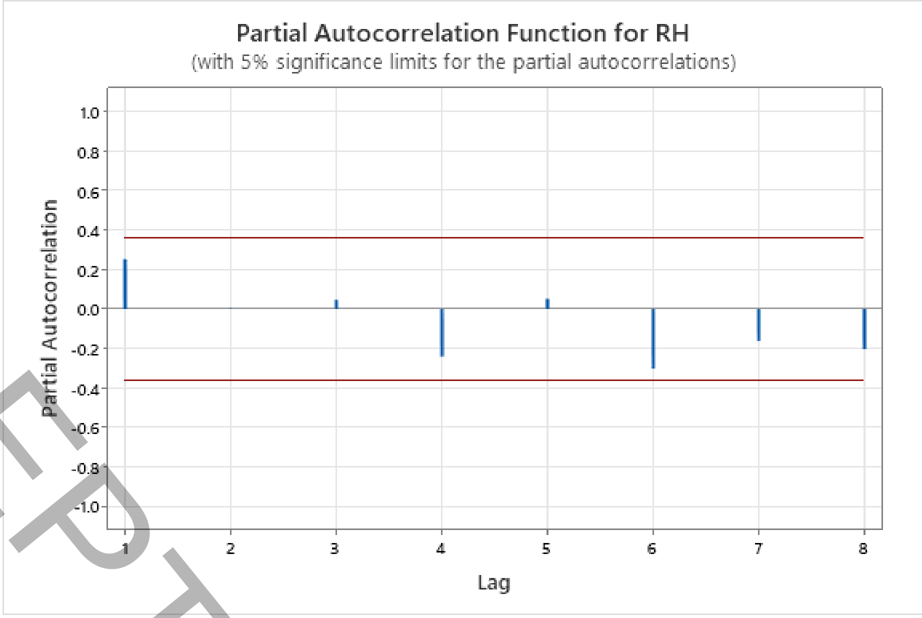


Fig. 6. Partial Autocorrelation of Relative Humidity time series

In the third step, the stationarity condition of the observational data was examined using the Box-Cox plot (Figure 7). In this plot, the horizontal axis (λ - lambda) shows different values of λ used in the Box-Cox transformation. The vertical axis (StDev) represents the standard deviation of the data after applying the Box-Cox transformation. The vertical lines (Lower CL and Upper CL) represent the 95% confidence interval for the optimal value of λ. The horizontal line (Limit) shows the value set as the standard deviation limit. The estimate (λ) is the estimated optimal value of λ (in this case, 0.29). Lower CL and Upper CL represent the lower and upper bounds for λ (respectively -2.61 and 4.09). The Rounded Value is the rounded optimal λ value (in this case, 0.50). The curve shows how the standard deviation of the data changes with variations in λ. At λ close to 0.29, the standard deviation reaches its minimum value, indicating the best transformation for normalizing the data. The value of 0.29 falls within the 95% confidence interval (-2.61 to 4.09), suggesting that this value is valid. The rounded λ value is 0.50, which may be used for simplicity. The optimal value of λ for the Box-Cox transformation is approximately 0.29. If a simpler approximation is required, the value 0.50 can also be used. This transformation can help normalize the data and improve statistical analysis. As shown in Figure 7, the upper and lower limits are not significant, indicating that the stationarity condition holds for the time series analysis.

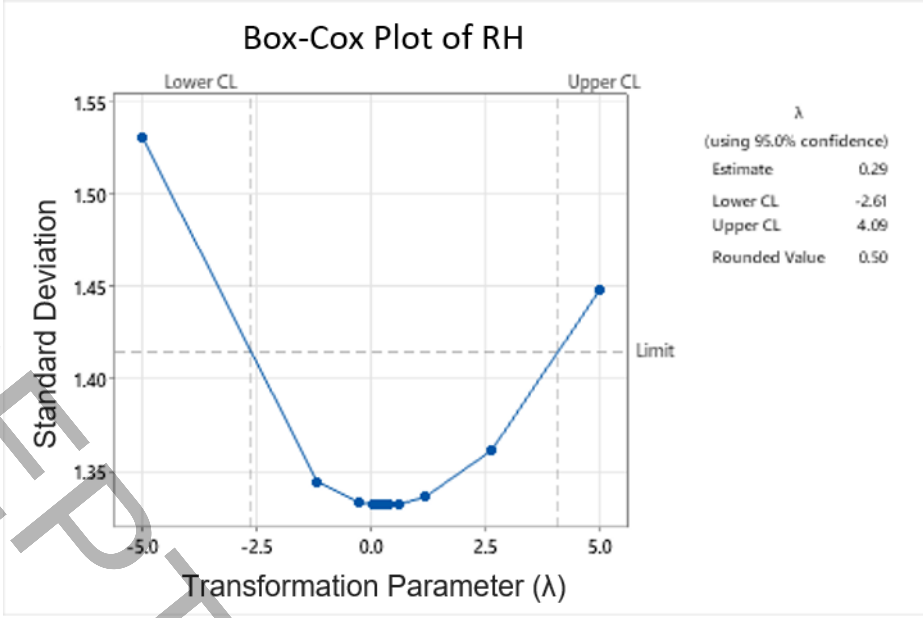
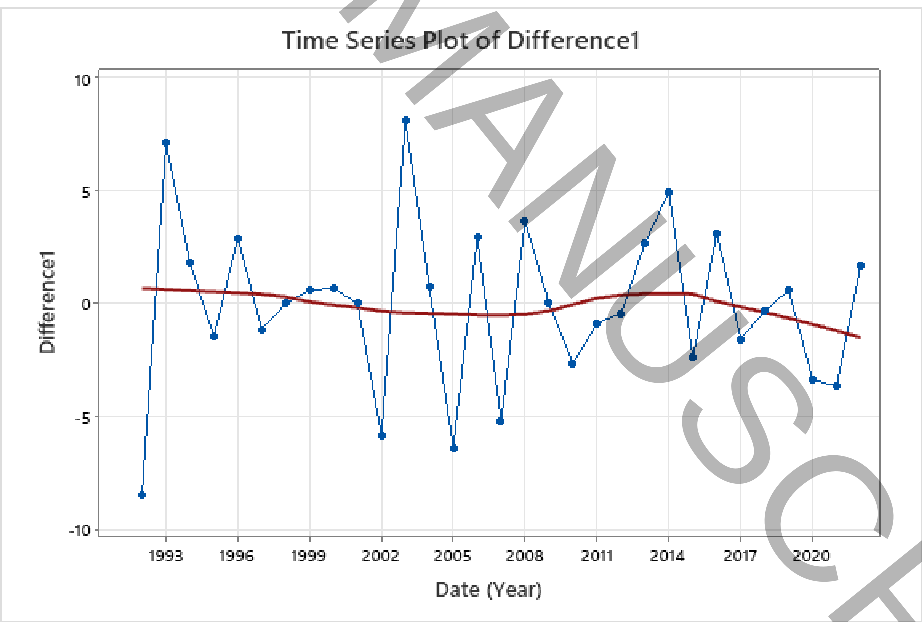


Fig. 7. Box-Cox plot of the Relative Humidity time series

After analyzing the characteristics of the time series under study, the trend of the relative humidity time series was removed based on the plot in Figure 8. For this purpose, the differencing method was used to remove the trend. This method was applied in two stages, LAG1 and LAG2. The results of these two stages are shown in Figure 8. As observed, the series becomes more stationary after the first differencing, and therefore, the differencing parameter in the ARIMA model is considered to be 1.



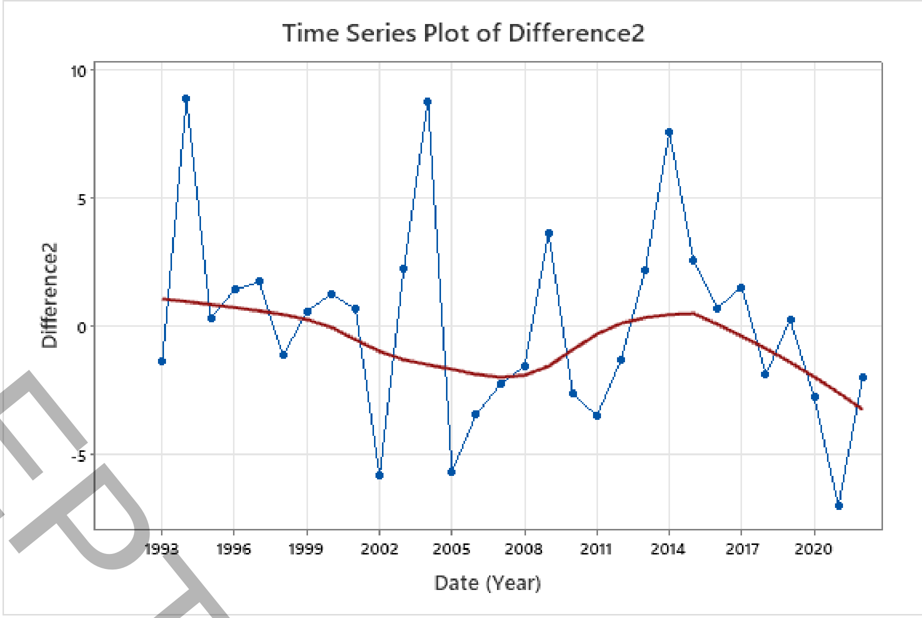


Fig. 8. The left plot shows the trend of the data at LAG1, and the right plot displays the trend of the data at LAG2

Based on the results obtained in Figure 8, the second and third assumptions, namely d and q, were considered to be equal to 1. However, these coefficient values will be determined through trial and error. To examine the ARIMA model, four important assumptions need to be considered: (1) the residuals must be uncorrelated, (2) the residuals must be normally distributed, (3) the residuals must be random, and (4) the variance of the residuals must be constant. First, the ARIMA(p,d,q) model with the assumption (0,1,1) was executed. According to Table 4, the Constant value has a P-value greater than 0.05. Therefore, it is not statistically significant and should be removed from the table.

Table 4. Final Estimates of Parameters, including Constant

Type	Coef	SE Coef	T-Value	P-Value
MA 1	0.937	0.111	8.41	0.000
Constant	0.0402	0.0735	0.55	0.589

Based on Table 5, the P-value for the MA (Moving Average) is 0, which is less than 0.05. Therefore, the significance condition is satisfied.

Table 5. Final Estimates of Parameters, without Constant

Type	Coef	SE Coef	T-Value	P-Value
MA 1	0.737	0.130	5.65	0.000

To ensure the condition of uncorrelated residuals, the Modified Box-Pierce test was evaluated using the statistical K-Square test. As shown in Table 6, for LAG 12 and 24, the P-value is greater than 0.05. This indicates that the condition of uncorrelated residuals is satisfied. Since the relative humidity data spans 32 years, the software displayed up to LAG 24.

Table 6. Modified Box Pierce

Lag	12	24	36	48
Chi-Square	17.42	31.59	*	*
DF	11	23	*	*
P-Value	0.096	0.109	*	*

To ensure the normality of the residuals, the ACF and PACF plots were examined. As shown in Figure 9, none of the residual values (represented by the blue plot) cross the upper and lower significance thresholds (shown in red) in either the ACF or PACF plots.

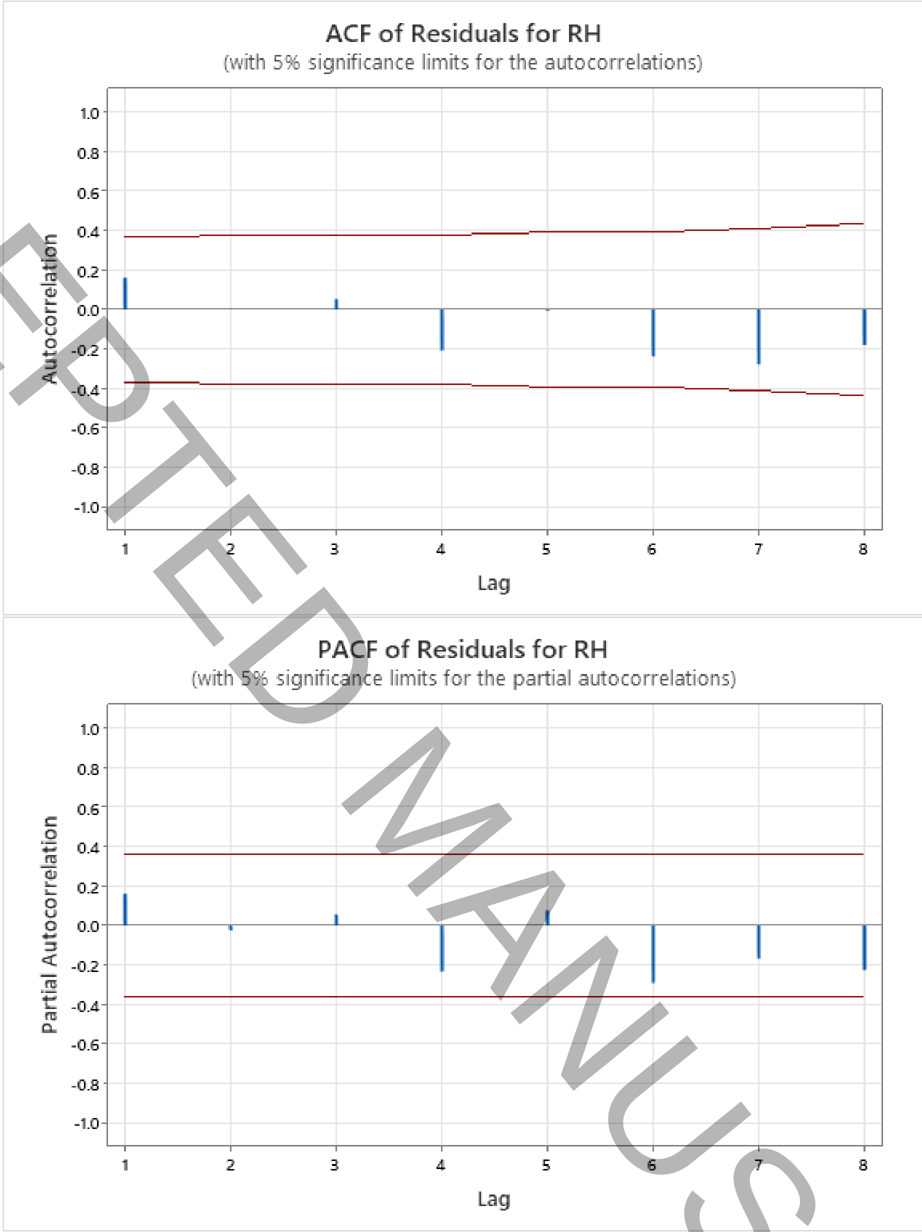


Fig. 9. ACF and PACF plots of residuals

Figure 10 shows 4 in 1 graph of residuals in the ARIMA results of Relative Humidity. The residuals in the Normal Probability plot are centered around the first quartile. The histogram also exhibits a bell-shaped curve, indicating a normal distribution. All these checks indicate that the condition of normality of the residuals is satisfied. To ensure the condition of randomness of the residuals, the "Versus Fits" plot from Figure 10 was used. If an upper and lower limit is set on this plot, and all residuals fall within these limits, the condition of randomness is satisfied. In this plot, the selected range is between +6 and -6. To ensure the condition of constant variance of the residuals, the "Versus Order" plot from Figure 5 was used. No specific trend was observed in this plot, thus confirming that the condition of constant variance of the residuals holds.

Furthermore, for additional assurance, the Kolmogorov-Smirnov plot was also examined, where the P-value was greater than 0.05. Therefore, the normality condition of the residuals holds (Figure 11).

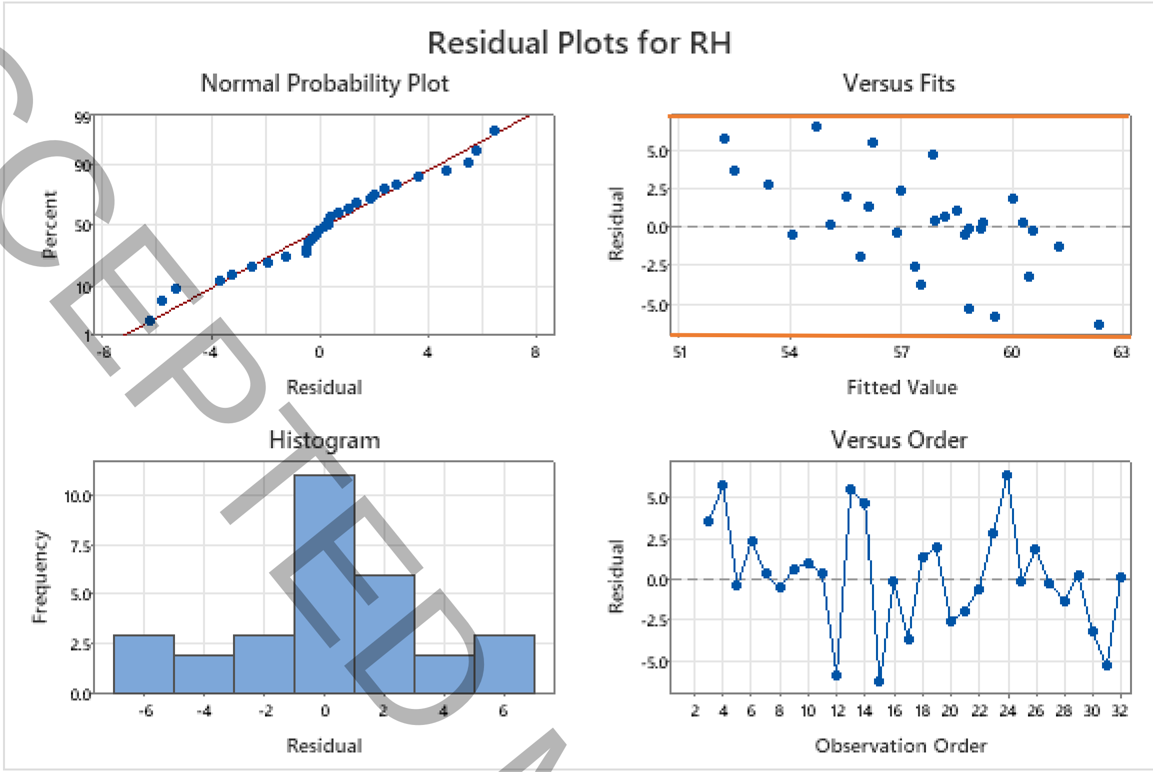


Fig. 10. 4 in 1 plots of residuals include: Histograms, Normal Probability, Versus Fits, and Versus Order

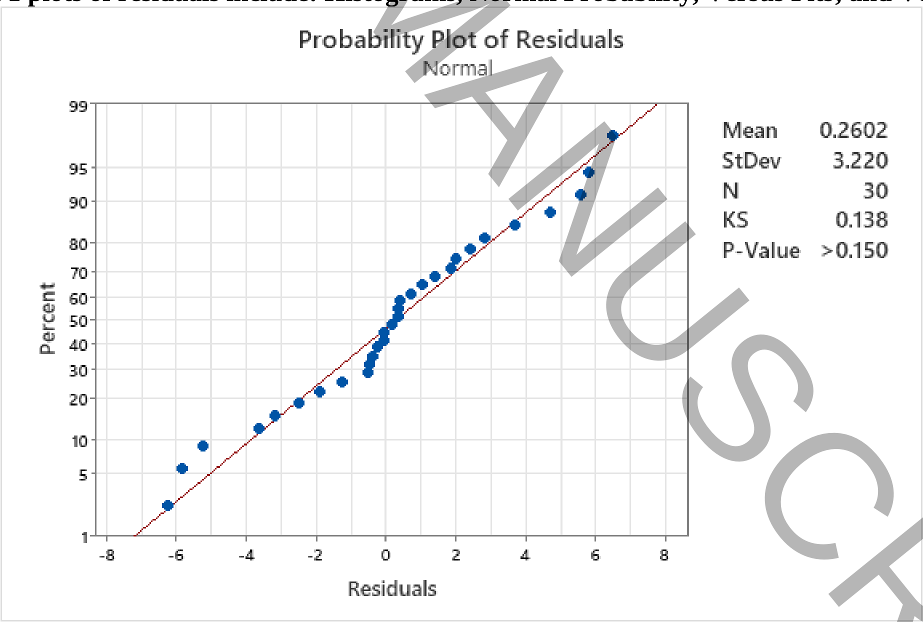


Fig. 11. Kolmogorov-Smirnov plot of residuals

Based on the necessary prerequisites for time series analysis, the ARIMA(0,1,1) model is a suitable choice. However, other configurations such as (1,1,1), (2,1,2), (0,0,1), and (1,0,1) evaluated through trial and error. By repeating this process and based on AIC and BIC, it was concluded that increasing p and q did not result in any change in the significance of AR and

MA. For all configurations, the Constant parameter was not found to be significant. Since d in LAG1 had been smoothed, d was initially set to 1, and the case for d equal to zero was also tested. Setting d to zero resulted in the insignificance of AR and MA. Finally, after executing the optimal model, all conditions were met. Therefore, forecasting for the next 10 years was performed with this model, and the results are shown below. Table 7, shows the forecasted values and Figure 12 shows trend on time series resulted ARIMA(0,1,1). As shown in Table 8, ARIMA(0,1,1) yielded the lowest AIC and BIC. Therefore, it was selected as the most suitable model for forecasting RH in this study.

Table 7. Forecasting results of ARIMA(0,1,1)

		95% Limits	
Period	**Forecast (%)**	**Lower**	**Upper**
33	54.1529	47.7070	60.5988
34	54.3986	47.0021	61.7951
35	53.9941	44.8813	63.1069
36	53.9064	43.5977	64.2151
37	53.6643	42.0329	65.2957
38	53.4975	40.6599	66.3351
39	53.2940	39.2345	67.3535
40	53.1083	37.8576	68.3591
41	52.9140	36.4716	69.3564
42	52.7239	35.0970	70.3508

Table 8. Comparison of different ARIMA tested models

Model	AIC	BIC
ARIMA(0,1,1)	129.33	131.02
ARIMA(1,1,1)	131.90	134.02
ARIMA(2,1,2)	134.45	137.88
ARIMA(0,0,1)	140.77	142.44
ARIMA(1,0,1)	138.62	141.04

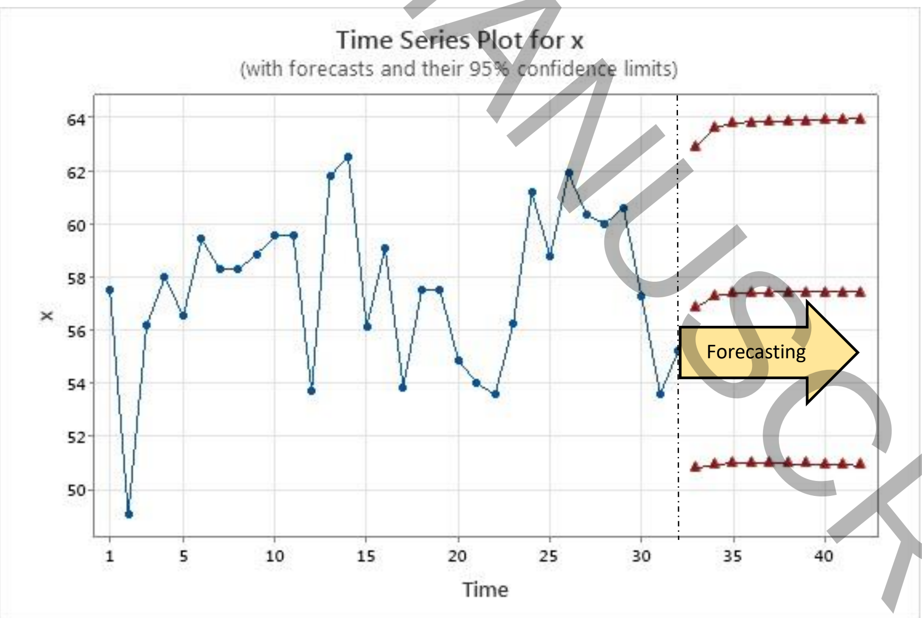


Fig. 12. Future trend plot of ARIMA(0,1,1)

Model Accuracy in Historical Forecasting

To assess the forecasting accuracy of the selected ARIMA(0,1,1) model, we performed one-step-ahead back-casting on the historical dataset from 1993 to 2022. The predicted values were compared to the actual observed RH values for each year. The model's performance was evaluated using standard error metrics, including: Root Mean Square Error (RMSE): 2.45, Mean Absolute Error (MAE): 1.89, and Coefficient of Determination (R^2): 0.72. These results confirm that the ARIMA(0,1,1) model provides an acceptable level of predictive accuracy, especially considering the climatic complexity and limited resolution of the dataset. Figure 13 illustrates the comparison between the actual and predicted RH values over the historical period.

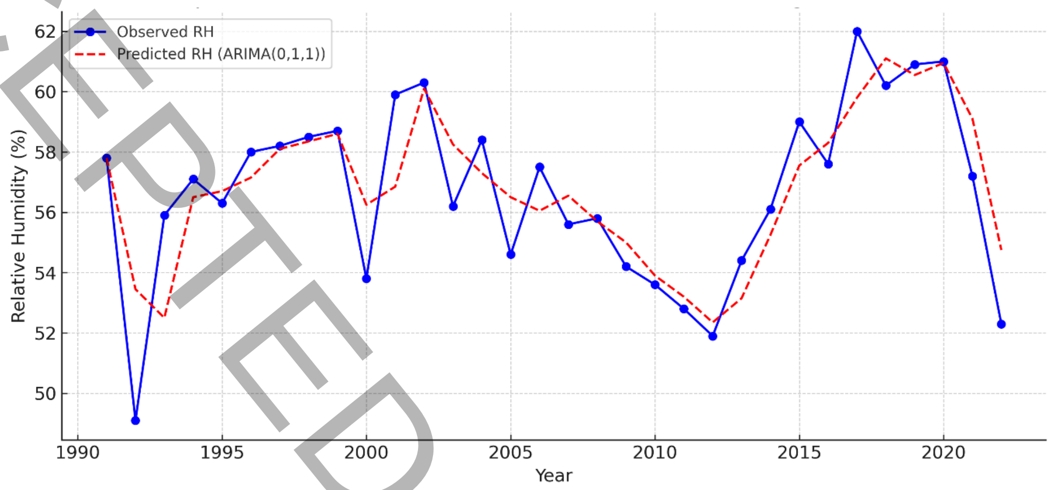


Fig. 13. comparison between the actual and predicted RH values over the historical period using ARIMA(0,1,1)

Interpretation of Model Selection and Implications

Although recent studies demonstrate the superior predictive performance of advanced AI-based models such as LSTM, ELM, and Transformer architectures [24, 25], their implementation typically requires extensive datasets, high-quality observations, and substantial computational resources. In contrast, the present study focuses on a long-term historical dataset from a semi-arid mountainous region, where data scarcity and strong seasonal variability make the application of complex AI models less practical [26]. The ARIMA model was therefore selected for its robustness, interpretability, and proven performance in forecasting hydro-climatic variables under limited-data conditions [27, 28].

Nevertheless, future work could extend this research by incorporating hybrid approaches or comparing ARIMA with state-of-the-art AI techniques to assess potential improvements in forecasting accuracy under similar climatic conditions.

Conclusion

Time series modeling for predicting relative humidity, as one of the key variables in hydro-climatological studies, is of significant importance. In this study, using historical relative humidity data and Mann-Kendall and time series analysis methods, the trend of changes in this variable in the past and over the future period was examined. The results indicated that in period of 1993-2022 RH had decreasing trend that wasn't significant. Based on time series analysis, relative humidity in the coming years will experience a slight increase. Although this increase is not substantial, it could have important implications for water resource management, rainfall forecasting, and agricultural planning. Other outcomes of this modeling include identifying the general trend in historical data and confirming the influence of regional climatic factors on relative humidity. In the statistical analyses performed, the ARIMA model was selected as an appropriate model for forecasting, and the predictions made using this model showed satisfactory accuracy. For the 10 years ahead, the projections indicate a gradual increase in relative humidity. This increase may lead to a reduction in the risk of drought and improvement in soil moisture conditions, which is crucial for natural resource management and climatic planning. Even a modest increase of 1–2% in relative humidity may have practical implications in semi-arid environments, including improved soil moisture retention and reduced evapotranspiration. These effects can benefit local agriculture and drought resilience strategies. However, it is recommended that similar studies consider other influencing variables such as temperature, rainfall patterns, and wind speed to provide more comprehensive and accurate forecasts. Lastly, it is suggested that more advanced hybrid models, such as Artificial Neural Networks (ANN) or ARIMA-ANN hybrid models, be used to improve the accuracy of predictions. The impact of changes in relative humidity on other climatic indicators, such as drought, soil water balance, and evapotranspiration, should also be analyzed. Expanding the period of input data to the model could enhance the stability of the prediction results.

References

[1] M. Hulme, E.M. Barrow, N.W. Arnell, P.A. Harrison, T.C. Johns, T.E. Downing, Relative impacts of human-induced climate change and natural climate variability, Nature, 397(6721) (1999) 688–691.

[2] J. Smith Jr, L.R. Johnson, R.Q. Thomas, Assessing Ecosystem State Space Models: Identifiability and Estimation: JW Smith Jr. et al, Journal of Agricultural, Biological and Environmental Statistics, 28(3) (2023) 442–465.

[3] A. Agrawal, V. Kumar, A. Pandey, I. Khan, An application of time series analysis for weather forecasting, International Journal of Engineering Research and Applications, 2(2) (2012) 974–980.

[4] Z. Zhang, K. Zhang, A. Khelifi, Multivariate time series analysis in climate and environmental research, Springer International Publishing Cham, 2018.

[5] G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[6] R.J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts Melbourne, Australia, 2014.

[7] K. Kurnianingsih, A. Wirasatriya, L. Lazuardi, A. Wibowo, I.K.A. Enriko, W.H. Chin, N. Kubota, Big data analytics for relative humidity time series forecasting based on the LSTM network and ELM, International Journal of Advances in Intelligent Informatics, 9(3) (2023) 537–550.

[8] H. Douville, K.M. Willett, A drier than expected future, supported by near-surface relative humidity observations, Science Advances, 9(30) (2023) eade6253.

[9] G. Lenderink, N. Ban, E. Brisson, S. Berthou, V.E. Cortés-Hernández, E. Kendon, H. Fowler, H. de Vries, Are dependencies of extreme rainfall on humidity more reliable in convection-permitting climate models?, Hydrology and Earth System Sciences Discussions, 2024 (2024) 1–31.

[10] J.Y. Ahn, Y. Kim, H. Park, S.H. Park, H.K. Suh, Evaluating time-series prediction of temperature, relative humidity, and CO2 in the greenhouse with transformer-based and RNN-based models, Agronomy, 14(3) (2024) 417.

[11] Z. Wang, L. Bugliaro, K. Gierens, M.I. Hegglin, S. Rohs, A. Petzold, S. Kaufmann, C. Voigt, Machine learning for improvement of upper-tropospheric relative humidity in ERA5 weather model data, Atmospheric Chemistry and Physics, 25(5) (2025) 2845–2861.
[12] M. Shad, Y. Sharma, A. Singh, Forecasting of monthly relative humidity in Delhi, India, using SARIMA and ANN models, Modeling earth systems and environment, 8(4) (2022) 4843–4851.
[13] M.I. Hutapea, Y.Y. Pratiwi, I.M. Sarkis, I.K. Jaya, M. Sinambela, Prediction of relative humidity based on long short-term memory network, in: Aip conference proceedings, AIP Publishing LLC, 2020, pp. 060003.
[14] W. Kendall, R. Sherwood, Palatability of Leaves of Tall Fescue and Reed Canarygrass and of Some of Their Alkaloids to Meadow Voles 1, Agronomy Journal, 67(5) (1975) 667–671.
[15] S. Eslamian, Y. Sabzevari, A regional case study of investigation of the meteorological drought trend using the SPI index and the Mann–Kendall test, in: Drought Dynamics, Elsevier, 2026, pp. 375–393.
[16] Y. Sabzevari, S. Eslamian, Predicting the effect of temperature changes on reference evapotranspiration by means of time series modeling (case study: Khorramabad Basin), علوم و مهندسی آبیاری, ۴۵(۲) (۲۰۲۲) ۱۲۵–۱۳۸.
[17] C. Chatfield, H. Xing, The analysis of time series: an introduction with R, CRC press, 2019.
[18] Y. Sabzevari, S. Eslameian, K. Moradalivand, Monitoring annual precipitation changes in Dezful plain with statistical analysis and time series, (2022).
[19] P.J. Brockwell, R.A. Davis, Introduction to time series and forecasting, Springer, 2002.
[20] Y. Sabzevari, S. Eslamian, Forecasting the State of Climate Warming Using Time Series Analysis (Case Study: Aligudarz), Applied Research in Water Engineering, 2(1) (2026) 93–107.
[21] O. Ryan, J.M. Haslbeck, L.J. Waldorp, Non-stationarity in time-series analysis: Modeling stochastic and deterministic trends, Multivariate Behavioral Research, 60(3) (2025) 556–588.
[22] H. Smit, Autocorrelation and time series analysis, in: Chemometrics: Mathematics and Statistics in Chemistry, Springer, 1984, pp. 157–176.
[23] J. Martin, D.D.R. De Adana, A.G. Asuero, Fitting models to data: Residual analysis, a primer, Uncertainty quantification and model calibration, 133 (2017).
[24] W. Li, B. Sun, Y. Sun, Y. Huang, Y.-m. Cheung, F. Gu, DC-SHADE-IF: An infeasible–feasible regions constrained optimization approach with diversity controller, Expert Systems with Applications, 224 (2023) 119999.
[25] T. Liu, S. You, Analysis and forecast of Beijing's air quality index based on ARIMA model and neural network model, Atmosphere, 13(4) (2022) 512.
[26] V. Vasylkivska, R. Dilmore, G. Lackey, Y. Zhang, S. King, D. Bacon, B. Chen, K. Mansoor, D. Harp, NRAP-open-IAM: A flexible open-source integrated-assessment-model for geologic carbon storage risk assessment and management, Environmental Modelling & Software, 143 (2021) 105114.
[27] L. Yan, X. Lihua, L. Qinghua, C. Jiang, Y. Kunxia, X. Chong-Yu, Correction to: On the Applicability of the Expected Waiting Time Method in Nonstationary Flood Design, Water Resources Management, 34(9) (2020) 3057–3058.
[28] M. Khazaeiathar, B. Schmalz, Addressing volatility and nonlinearity in discharge modeling: ARIMA-iGARCH for short-term hydrological time series simulation, Hydrology, 12(8) (2025) 197.