**Anadolu Tarım Bilimleri Dergisi,**
Anadolu Journal of Agricultural Sciences

e-ISSN: 1308-8769, ANAJAS February 2023, 38 (1): 53-68

# Modeling of Climatic Variables Using Stochastic Approaches in Sudan

# Sudan'da Stokastik Yaklaşımlar Kullanılarak İklimsel Değişkenlerin Modellenmesi

**Mawadda. A. M. ABDALLAH[1], Bilal CEMEK[2]**

[1] Agriculutural Research Corporation (ARC), Wad Medani City, Sudan
• **17211130@stu.omu.edu.tr**, • ORCID > 0000-0002-9135-3025
[2]Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Tarımsal yapılar ve Sulama Bölümü, Samsun
•**bcemek@omu.edu.tr**• ORCID > 0000-0002-0503-6497

**Makale Bilgisi / Article Information**

**Makale Türü / Article Types:** Araştırma Makalesi / Research Article
**Geliş Tarihi / Received:** 20 Temmuz / July 2022
**Kabul Tarihi / Accepted:** 03 Kasım / November 2022
**Yıl / Year:** 2023 | **Cilt - Volume:** 38 | **Sayı - Issue:** 1 |**Sayfa / Pages:** 53-68

**Atıf/Cite as:** Abdallah, M.A.M., Cemek, B. "Modeling of Climatic Variables Using Stochastic Approaches in Sudan" Anadolu Journal of Agricultural Sciences, 38(1), February 2023: 53-68.

**Sorumlu Yazar / Corresponding Author:** Mawadda. A. M. Abdallah

https://doi.org/10.7161/omuanajas.1145

# MODELING OF CLIMATIC VARIABLES USING STOCHASTIC APPROACHES IN SUDAN


## ABSTRACT

The climatic variables play a significant role in agricultural process and irrigation management because we need to know all changes related to the climate, which will absolutely affect agricultural yield. For this purpose, the ARIMA models were suggested in this study for modeling daily average temperature, solar radiation, and relative humidity factors related to five main meteorological stations (Wad Madani, Khartoum, Al Gadaref, Al Damazin, and Dongola) in Sudan. The daily variables were obtained from the period 2013 to 2020. Time series analysis methods are used for estimating and modeling the climatic variables using Autoregressive Integrated Moving Average methods, which are called Box Jenkins models. For modeling purposes, linear stochastic models were used to estimate the future values of daily variables. The Augmented Dickey-Fuller test (ADF) was used to check the stationarity of the data at 1%, 5%, and 10% confidence levels. The time series of variables showed stationarity and no trend. The best models were selected from the autocorrelation (ACF) and partial autocorrelation (PACF) function graphs employing diagnostic testing. The adjusted R2, Standard error (S.E), Akaike information criterion (AIC), and Bayesian information criterion (BIC) values were used to assess which models were the best. The appropriate findings were observed in ARIMA (1,0,1) and (1,0,2) which can be effective for predicting future values. The ARIMA models obtained satisfactory results for temperature, relative humidity, and solar radiation variables. So, this study might be extremely helpful for agricultural engineers to achieve all the processes related to agricultural practices.


***Keywords:*** Variables, Arıma Models, Adf Test, Stochastic, Prediction, Autocorrelation.

# SUDAN'DA STOKASTİK YAKLAŞIMLAR KULLANILARAK İKLİMSEL DEĞİŞKENLERİN MODELLENMESİ


## ÖZ:

İklimsel değişkenleri, tarımsal süreç ve sulama yönetiminde önemli bir rol oynamaktadır çünkü iklimle ilgili tarımsal verimden kesinlikle tüm etkilenecek değişiklikleri bilmemiz gerekmektedir. Bu amaçla, bu çalışmada Sudan'daki beş ana meteoroloji istasyonu (Wad Madani, Hartum, Al Gadaref, Al Damazin ve Dongola) ile ilgili günlük ortalama sıcaklık, bağıl nem ve güneş radyasyonu değişkenlerinin modellenmesi için ARIMA modelleri önerilmiştir. 2013-2020 yılları arasındaki günlük iklim verileri kullanılmıştır. Box Jenkins modelleri olarak adlandırılan ARIMA (Otoregresif Entegre Hareketli Ortalama Modelleri) kullanılarak iklimsel değişkenlerin tahmin ve modellemesini için zaman serisi analiz yöntemlerini kullanılmıştır. Modelleme amacıyla, günlük değişkenlerin gelecekteki değerlerini tahmin etmek için doğrusal stokastik modeller kullanılmıştır. Durağanlığı 1%, 5% ve 10% güven düzeylerinde kontrol etmek için zaman serilerine artırılmış Dickey-Fuller testi (ADF) uygulanmıştır. Değişkenlerin zaman serileri durağanlı ve eğilimsiz olduğunu tespit edilmiştir. Modelleri kontrol etmek ve otokorelasyon (ACF) ve kısmi otokorelasyon (PACF) fonksiyon grafiklerinden en iyi modelleri seçmek için diyagnostik kontrolleri kullanılmıştır. En iyi modelleri, düzeltilmiş R2, Standart hata (S.E), Akaike bilgi kriteri (AIC) ve Bayesian bilgi kriteri (BIC) değerlerine göre seçilmiştir. En iyi sonuçları, gelecekteki değerleri tahmin etmek için etkili olabilecek ARIMA (1,0,1) ve (1,0,2) modellerinde gözlenmiştir. ARIMA modelleri sıcaklık, bağıl nem ve güneş radyasyonu değişkenleri için tatmin edici sonuçları elde etmiştir. Bu nedenle, bu çalışma ziraat mühendislerinin tarımsal uygulamalarla ilgili tüm süreçleri gerçekleştirmeleri için çok yardımcı olabilmektedir.

***Anahtar Sözcükler:*** Değişkenler, ARIMA Modelleri, ADF Testi, Stokastik, Tahmin Etmek, Otokorelasyon.

## 1. INTRODUCTION

The climate system is known as a complex system that is characterized by major changes in climatic variables, which are very necessary for many environmental systems, especially agricultural processes, which are affected by weather conditions that can affect agricultural production. Agricultural productivity might be increased or reduced due to a longer or shorter growing season, so the most important climatic variables such as temperature, relative humidity, and solar radiation make a big sense of crop yield. However, crop yields are also expected to vary increasingly from year to year due to extreme weather events, and it varies/they(crop yields) vary? depending on different geographical locations. Therefore, many different methods have been used in many previous studies to interpret and estimate the changes in variables, such as time series analysis techniques using Autoregressive Integrated Moving Average (ARIMA) models to model future values. Time series analysis is important to analyze and predict the future values of hydrological parameters such as the evapotranspiration process which is considered a complex process in the hydrology field. In addition, time-series analysis is necessary for testing hydrologic and climatologic models with reconstructed hydrologic models from the past (Palmroth et al., 2010).

It is worth mentioning that the most evident patterns appearing in time series data are trends and seasonality (G. E. Box et al., 2015).

The Box-Jenkins model, also known as the autoregressive integrated moving average* model, is widely used in analyzing hydrologic time series and has been used to extrapolate past patterns of behavior into the future. (Mohan and Arumugam, 1995). ARIMA is the first method introduced by Box-Jenkins and has until now been the most popular model for forecasting univariate time series data (Lee, 2011). According to (Yürekli et al., 2007) the ARIMA models have been applied to simulate and estimate the climatic variables for Tokat city in Turkey and the models that obtained satisfying results can be used for predicting future values.

According to (Asteriou and Hall, 2007) time series analysis is necessary to set up a hypothetical probability model to represent data for predicting future values . It is then possible to estimate parameters, check for good data fit, and perhaps use the fitted method to improve our understanding of the processes by generating the data after an appropriate collection of models has been selected as the best. In addition, time series analysis is very important for economic statistics such as unemployment to recognize the presence of seasonal components and not being confused them with long-term trends. Time series models can be applied for controlling future values of series by adjusting parameters. For instance, time series analysis is playing the main role in prediction of future values such as testing hypothesis of global warming using recording temperature data and predicting one series from observations of another population data. Also, it is useful in simulation studies, for example, the performance of a reservoir depends heavily on the random daily inputs of water to the hydrological system. There are many studies that have been carried out in different environmental components using the time series analysis approaches with another stochastic methods to increase the performance of models as observed in the study of (Ashrafzadeh et al., 2020) which forecasted monthly ETo using different methods of time Series Models, Support Vector Machines (SVM), and Goup Method of Data Handling GMDH in Iran for Long-Term during 1993-2014 and these models were used to forecast reference evapotranspiration for the next 2-year period (2013-2014). According to (Luo et al., 2014) have proposed a method for forecasting short-term daily reference evapotranspiration using Hargreaves–Samani model and temperature forecasts. In this study, the HS model was compared with the Penman-Monteith model to calibrate and validate the Hargreaves–Samani (HS) model. The ARIMA models were applied by (Mossad and Alazba, 2016) to predict monthly reference evapotranspiration under arid climate using daily metrological parameters to develop the ET0 time series.

In the study of (Kim et al., 2011) ARIMA model has applied seasonal to evaluate and predict temporal-spatial Precipitation in Mongolia. In such an approach (Han et al., 2012) used the ARIMA models to predict drought, and the results demonstrated how the model had important short-term forecasting for the SPI index. Also, (Han et al., 2010) employed a new technique of modelling for the Spatial and temporal series the VTCI series and applied the ARIMA models to simulate and forecast the VTCI (vegetation temperature condition index) series to simulate and forecast VTCI (vegetation temperature condition index) series.

(Zhang, 2003) has suggested a hybrid method of ARIMA and ANNs (Artificial neural networks) models to improve the accuracy of forecasting and estimate the strength of the modeling in linear and nonlinear case. The combination of ARIMA and neural network models are especially essential to extract different patterns of data and show the ability to capture the non-linearity in the data. An Artificial neural network model is an effective tool to improve forecasting accuracy and achieve a high degree of model performance in time series data (Khashei and Bijari, 2010).

Kaur et al. (2015) used neural networks to apply a hybrid model that considered two methods of ARIMA and wavelet transform models. To forecast the hydrological drought on annual and seasonal time spans, (Bazrafshan et al., 2015) have tested the accuracy of stochastic known as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models. They also calculated the accuracy of models in terms of lead-time for forecasting. In such an approach (Mishra and Desai, 2005) and Grazhdani, (2010) have used the SARIMA model to predict droughts using standardized precipitation index series in the Kansabati river basin in India. In this study, the ARIMA models were applied for simulating and estimating the daily climatic variables which were calculated from five main meteorological stations in Sudan over a period of seven years related to the years 2013–2020.

## 2. MATERIALS AND METHODS

### 2.1. Study Area

Five major stations in the Sudan were chosen as the study experimental samples and served as the study area. Sudan is officially known as the Republic of Sudan. The Northeast of Africa is home to Sudan which is bordered by Egypt in the north, Libya in the northwest, Chad in the west, the Central African Republic in the southwest, South Sudan in the south, Ethiopia in the southeast, Eritrea in the east, and the Red Sea in the east. Sudan lies at the latitudes of 12.86280 N and a longitude of 30.217 0 E, also 405 m above mean sea level. Sudan occupies 1.886.068 km2 (728.215 square miles) making it Africa's third-largest country by area (Fig. 2.1).

## 2.2. Data Collection

The used data sets were daily climatic variables of temperature, relative humidity, and solar radiation, covering the period from 2013 to 2020 (Fig. 2.1).

## 2.3. Time Series Analysis Technique (ARIMA model)

A time series is a collection of data points that are taken at evenly spaced points or observations and are plotted based on a certain time. Time series analysis starts with plotting the observations of the time series and then applying the approach from (G. E. Box et al., 2015) using mathematical functions which were autocorrelation and partial functions that indicate how each value of the series is correlated with the previous value according to the time (Chen et al., 2009). The used mathematical functions were shown as following:

**(i) Autocorrelation Function**

A mathematical process that describes the correlation between all the of points in the time series data with separately time or lag. In this process, lags indicate the stationarity and non-stationarity of the series by using the Ljung-Box Q and p-values confidence intervals for each lag showing if the series has a statically significant or non-significant correlation.

**(ii) Partial Autocorrelation**

The partial autocorrelation function (PACF) is used to detect trends and seasonality. In this process, lags play an important role in data analysis which aims to identify the extent of the lag in an autoregressive model, by plotting this function, the appropriate lags "P" in an AR(p) model or in an extended ARIMA(p,d,q) model would be applicable to determine The partial autocorrelation function (PACF) (Chen et al., 2009).

## 2.4. Procedure of ARIMA Model

This model is fitted to time series data either to better understand the data or to predict future points in the series. In this study, the ARIMA model is applied in some cases where the data show evidence of non-stationarity features, and an initial differencing step (corresponding to the "Integrated" part of the model) can be applied to remove the non-stationarity features. The model is referred to as an ARIMA (p, d, q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model, respectively. The first parameter "p" refers to the number of autoregressive lags (not counting the unit-roots), the second parameter "d" refers to the order of integration, and the third parameter “q” gives the number of moving average lags. ARIMA model form an important part of the Box-Jenkins approach to time-series modeling (G. E. Box et al., 2015). According to the following sub-categories of ARMA model described for response series [yi], form of the ARIMA model is:

$$Xt-\phi 1\ xt-1-\ldots-\phi_p\ x_t-p = z_t+\Theta_{1zt}-1+\ldots+\Theta_{qzt}-q \qquad (1)$$

Where:

$\{Z_t\}\sim = WN(0,\sigma^2)$, Polynomials of $(1-\phi 1z-...-\phi_p z^p)$ and $(1+\Theta 1z+...+\Theta_q z^q)$ have no common factors.

The process {Xt} is released to be an ARMA (p, q) process with mean μ if {Xt – μ} is an ARMA (p,q) process. It is convenient to use the more concise form of equation (1).

$$\text{And} \quad \varphi(B)Xt = \theta\ (B)Zt, \qquad (2)$$

where: φ(B) and θ (B) are the pth and qth-degree polynomials,

$$\varphi(z) = 1 - \varphi\ 1\ z - \cdots - \varphi_{\ p}\ z_p \qquad (3)$$

$$\text{and} \quad \theta\ (z) = 1 + \theta\ 1\ z + \cdots + \theta_q\ z_q, \qquad (4)$$

where:

B is the backward shift operator $(B_{\ j}X_{\ t} = X_t{-}j,\ B_{\ j}Z_t = Z_t{-}j,\ j = 0, \pm 1\ldots)$ The time series {Xt} is said to be an autoregressive process of order p (or AR(p)) if θ (z) ≡ 1, and a moving-average process of order q (or MA(q)) if φ(z) ≡ 1.

On the other hand, when the data series is non-stationarity ARIMA models can be applied using differencing operations and make it more suitable for estimating and forecasting the future values. In other words, The ARIMA (p,d,q) modeling approach is suitable for non-stationarity in the historic data series (Mohan and Arumugam, 1995). The following sub-categories of ARIMA model are described for response series [yi], form of the ARIMA model is:

$$\Phi\ (\beta\ )\ (\omega T-\mu) = \Theta\ (\beta\ )\ a_t \qquad (5)$$

Where:

t = is the time index

β = is the backshift operator defined as: $\beta_{yt} = y_r - 1$

$\omega r = (1\text{-}b)^d y_r$ = is the response series after differencing

μ = is the mean term

Φ (β ) and Θ (β ) are the autoregressive operator and the moving average opera-tor, respectively. They are written as:

$$\Phi(\beta) = 1 - \varphi_1 \beta - \varphi_2 \beta^2 - \ldots \Phi_p \beta^p \quad (6)$$

And

$$\Theta(\beta) = 1 - \Theta_1 \beta - \Theta_2 \beta_2 - \ldots \Theta_q \beta^q \quad (7)$$

Where:

$a_t$ are the sequence of random error, while all the other terms have already been explained.

The at values, are assumed to be independent and follow the normal distributi-on with mean zero and constant variance. The model can be written as:

$$\Phi(\beta)\omega_t = \delta + \Theta(\beta)\, a_t \quad (8)$$

Where the constant estimate δ is obtained by the formula:

$$\delta = \varphi(\beta)\, \mu = \mu - \Theta_1 \mu - \Theta_2 \mu \ldots \Theta_p \mu \quad (9)$$

The ARIMA model was developed using the approach from (T. M. Box et al., 1994), which includes four main iterative steps, as known, these main stages in building an ARIMA model based on Box-Jenkins procedure, i.e. (1) model identification, (2) model estimation, (3) model checking, and (4) model forecasting. In the identification step, differencing transformations of the data were applied to remove the seasonal and non-seasonal trends.

In this study, ARIMA models were created to evaluate the values of the climatic variables using the techniques of time series analysis. For this purpose, The statistical Eviews program was used to analyze time series and generate models for forecasting future values.

## 2.5. ARIMA (Box-Jenkins) Model Steps

The ARIMA model was applied using the time series analysis steps according to the approach of (G. E. Box et al., 2015) as following:

**(i) Identify Data**

To begin, the original data was identified to find the ARIMA patterns for the various data series by plotting both the ACF and PACF and the associated correlogram graphs, which show the statistical significance based on the lags of the graphs. A unit root test is present in a time series sample which is the null hypothesis of an Augmented Dickey-Fuller test in statistics. This test is used for checking the stationarity or trend -stationarity of the series (MacKinnon, 2010). The Augmented Dickey-Fuller test (ADF) was used with various levels of confidence to determine the stationarity of time series. This test was required to determine whether the series data was genuine or fabricated. In parliamentary law to generate valid data, both the variables and the error term must be stationary. It does signify that the error term must have a way of reverting to its average value(Saravanan, 2015). The stationarity examined using the following equation of ADF:

$$\Delta Zt = \alpha + \theta_t + \lambda Z_t - 1 + \sum k\ \varphi_i\ \Delta Z_t -1 + \varepsilon_t \qquad (10)$$

Where:

Null hypotheses, Hn = $\lambda = 0$ ($Z_t$ is not stationary, $Z_t$ contain unit root) and alternative hypotheses, $H_a = \lambda < 0$ ($Z_t$ is stationary).

It is worth mentioning that the means (μ) and the variance (σ2) of error term must be constant. The ADF test should be taken with the trend and intercept at level first (The original data), If the null hypothesis fails to decline, then the first difference should be taken. The second difference is only tested if the first difference is not important (Ibrahim and Amin, 2005).

The testing process of the ADF test is applying to the models mathematically as following

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \delta 1\ \Delta y_{t-1} + \ldots\ldots + \delta_{p-1}\ \Delta\ y_{t-p} + \varepsilon_t \qquad (11)$$

Where:

α is constant, the coefficient on a time trend, and P the lag order of the autoregressive process. By considering lags of the order *p*, the ADF formulation allows for higher-order autoregressive processes. This means that the lag length p must be determined when applying the test. One possible approach is to test down from high orders and examine the t- values on coefficients.

The unit root test is achieved by the null hypothesis $\gamma = 0$ against the alternative hypothesis of $\gamma < 0$.

The null hypothesis assumes that $Z_t$ is not stationary or $Z_t$ contain unit root) and the alternative hypotheses $\lambda < 0$ means (Zt is stationary).

$$DF\tau = \frac{\grave{Y}}{S.E(\grave{Y})} \qquad (12)$$

Where: DFτ is the Dickey -Fuller test value.

Once the value of the test statistic is computed, it can be compared to the relevant critical value for the Dickey–Fuller test. As this test is asymmetrical, we are only concerned with negative values of our test statistic DFγ. If the calculated test statistic is less (more negative) than the critical value, then the null hypothesis of $\gamma = 0$ is rejected and no unit root is present in the time series.

**(ii) Estimation of ARIMA Model**

The ARIMA and ARMA patterns were estimated to be the most sensitive models by considering the ACF and PACF correlogram graphs with the lowest values of the determination criteria and the most sensitive model. The most comparison appropriate model has been chosen according to the criteria of accuracy in time series analysis which are the highest significant coefficient, the least volatility, the highest adjusted R2, P-values (should be greater than 0.05), the least of AIC (Akaika info criterion) and SIC (Schwarz info criterion) values that all make sense for choosing the best model. In addition, the smallest values of RMSE are considered as criteria for choosing the best model.

**(iii) Diagnostic of Model**

In this stage, The Lunj-box test is performed for testing autocorrelation, and it is also called the autocorrelation test. This test was used to check the residuals correlogram to figure out if there is any information yet to be captured in the model whereas the flat correlogram in the ACF and PACF graphs referred to the most ideal model. For instance, it is very important to avoid over-fitting when checking the residuals of the correlogram.

**(iv) Forecasting of the Model**

The forecasting process is the essence of fitting an ARIMA model to forecast future values of the series using the past values of the original dataset. The forecasting is based on the selected model. In this step the forecasting graphs are plotted and verified as successful, and the forecasting has predicted the future values of the series.

https://doi.org/10.7161/omuanajas.1145094

## 2.6. Statistical Criteria (Information Criteria)

As known, there are many statistical criteria for selecting the most accurate models, some of them were used in this study to select the best model and assess the ARIMA models in forecasting long-term data of the climatic variables. The statistical criteria are defined mathematically according to (G. E. Box et al., 2008) approach as follows:

(1) The root squared error, RMSE is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Oi-Pi)^2}{n}} \quad (13)$$

(2) The coefficient of determination (R2) is:

$$R2 = \frac{\sum(Pi - \bar{P})(Oi - \bar{O})}{\sqrt{\sum(Pi - \bar{P})2}\sqrt{\sum(Oi - O)2}} \quad (14)$$

Where: Pi is the predicted value of the sample.

Oİ is the observed or the calculated value of the sample.

i is the forecast sample sequence number; i=1,2, . . . n.

P is the average predicted value of the sample sequence number.

O is the average observed value of the sample sequence number; and

n is the sample number of the predicted value.

(3) Akaike information criterion (AIC): Akaike's entropy-based Information Criterion (AIC) has had a fundamental impact in statistical model evaluation problems, in addition it is important to select the optimal model for forecasting the values of series data (Ozaki and Oda, 1977):

$$\hat{A} = n \, Ln \, R + 2 \, Kn \quad (15)$$

where:

n is the number of experimental points or observations,

R is the residual sum of squares and

Ln is the log likelihood function (assuming normally distributed errors), which is computed by the equation:

$$\ell = -n2 \cdot (1+\ln(2\cdot\pi) + \ln(1n \cdot \sum i=1n(yi-y\hat{}i)2)) \quad (16)$$

p is the number of parameters in the model to choose is the one for which Â is least (Webster and McBratney, 1989). When comparing many alternative models, the one with the minimum AIC value assures a good balance of goodness of fit and complexity.

(4) The Bayesian information criterion (BIC) is also known as Schwarz Criterion, it is another statistical measurement for the comparative evaluation among time series models, it is closely related to the AIC (Profillidis and Botzoris, 2018).

(5) The difference between BIC and AIC is manifested when we add number of K parameters (regressors or/and intercept), to increase the goodness of fit of the model. The BIC deals more with increasing parameters compared to the AIC. The BIC is calculated by the equation:

$$\text{SIC or BIC} = -2\cdot\ell/n + k\cdot\ln n/n \quad (17)$$

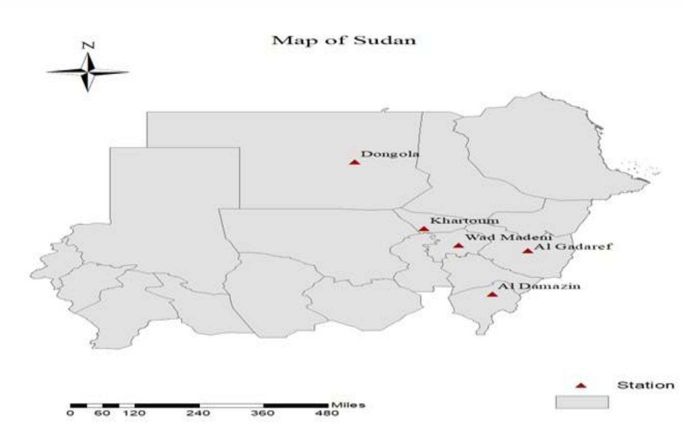



**Fig. 2.1**. Distribution Map of Main Meteorological Station.

## 3. RESULT AND DISCUSSION

According to the Table 3.1. the Dikey Fuller *values were obtained based on the original data and were most likely used to check the series' stationarity. As shown in the table 3.1. the values of calculated DF values of data series were less (more negative) than T-test critical values at 0.01, 0.05, and 0.10 significance levels for all climatic variables which means the null hypothesis is rejected and there is no unit root is present in the series. On other words, the data series was stationarity at all significance levels except solar radiation parameter in Khartoum station the DF value only at 1% was greater than T-test critical value which was -3.01> -3.432 then in this case the null hypothesis cannot be reject at 1% (there is unit root=not stationary) but it is rejected at 5% and 10% significance levels (there is no unit root) (Dickey, 2015).

It is important to mention that the daily data was investigated by plotting the ACF and PACF correlogram graphs which show the statistical significance based on associated lags with the data using the Box -Jenkins technique. For instance, before deciding on the ideal model, the residuals of the correlogram should be checked to indicate all information has been captured then the forecasting will be based on this model.

As shown in Table 3.2. the best models were chosen according to Statistical criteria for Selection models which were the adjusted R2, standard error SE, AIC, and SIC parameters. As presented in Table 3.2. the models with the highest values of R2 are considered as the best models for each daily variable related to the stations. The models with fewer errors outperform the others in terms of predicting future values.

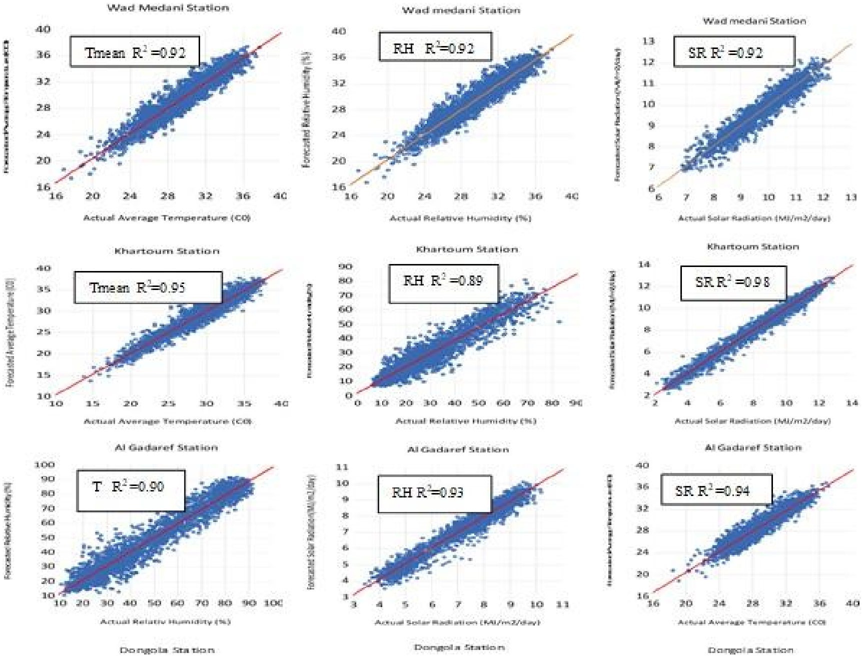
The daily climatic variables related to 2013-2020 were forecasted based on ARIMA model and compared with the original data series for each station as given in Figure 3.1.

**Table 3.1.** The Unit Root Test for Variables Depending on Df Aand T-Test Values at Different Level of Confidence.

| Station | Variable | DF value | T-test critical values | | |
|---|---|---|---|---|---|
| | | | 1% | 5% | 10% |
| Medani | Temperature | -6.056 | -3.432 | -2.862 | -2.567 |
| | Humidity | -4.068 | -3.432 | -2.862 | -2.567 |
| | Radiation | -5.436 | -3.432 | -2.862 | -2.567 |
| Khartoum | Temperature | -4.810 | -3.432 | -2.862 | -2.567 |
| | Humidity | -6.06 | -3.432 | -2.862 | -2.567 |
| | Radiation | -3.01 | -3.432 | -2.862 | -2.567 |
| AlGadaref | Temperature | -6.658 | -3.432 | -2.862 | -2.567 |
| | Humidity | -3.675 | -3.432 | -2.862 | -2.567 |
| | Radiation | -3.848 | -3.432 | -2.862 | -2.567 |
| AlDamazin | Temperature | -5.498 | -3.432 | -2.862 | -2.567 |
| | Humidity | -3.457 | -3.432 | -2.862 | -2.567 |
| | Radiation | -5.180 | -3.432 | -2.862 | -2.567 |
| Dongola | Temperature | -3.780 | -3.432 | -2.862 | -2.567 |
| | Humidity | -6.09 | -3.432 | -2.862 | -2.567 |
| | Radiation | -3.561 | -3.432 | -2.862 | -2.567 |

**Table 3.2.** The Best Models The Determination Criteria for Each Variable.

| Station | Variable | Model | Adjusted $R^2$ | S.E | AIC | BIC |
|---|---|---|---|---|---|---|
| Wad medani | Temperature | ARIMA(1,0,2) | 0.92 | 0.0005 | 2.844 | 2.851 |
| | Humidity | ARIMA(1,0,1) | 0.92 | 0.0007 | 2.857 | 2.863 |
| | Radiation | ARIMA(1,0,1) | 0.92 | 0.0005 | 0.350 | 0.356 |
| Khartoum | Temperature | ARIMA(1,0,1) | 0.95 | 0.0008 | 3.073 | 3.079 |
| | Humidity | ARIMA(1,0,2) | 0.89 | 0.0017 | 6.059 | 6.006 |
| | Radiation | ARIMA(1,0,1) | 0.98 | 0.0007 | 0.675 | 0.681 |
| ALGadaref | Temperature | ARIMA(1,0,2) | 0.90 | 0.0005 | 2.711 | 2.718 |
| | Humidity | ARIMA(1,0,2) | 0.93 | 0.0014 | 6.214 | 6.220 |
| | Radiation | ARIMA(1,0,2) | 0.94 | 0.0006 | 0.661 | 0.667 |
| Al Damazin | Temperature | ARIMA(1,0,2) | 0.90 | 0.0005 | 2.671 | 2.678 |
| | Humidity | ARIMA(1,0,2) | 0.95 | 0.0001 | 5.871 | 5.877 |
| | Radiation | ARIMA(1,0,1) | 0.87 | 0.0119 | 4.603 | 4.609 |
| | Temperature | ARIMA(1,0,1) | 0.98 | 0.0001 | 3.150 | 3.156 |
| Dongola | Humidity | ARIMA(1,0,2) | 0.82 | 0.002 | 5.415 | 5.421 |
| | Radiation | ARIMA(1,0,1) | 0.97 | 0.0002 | 0.883 | 0.889 |



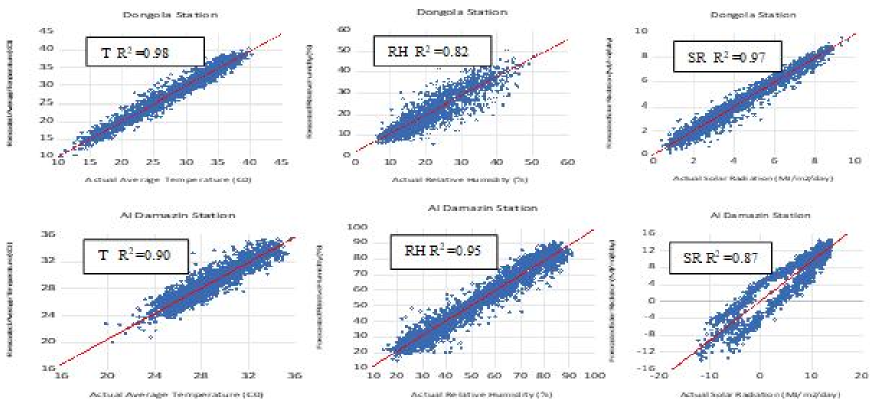



**Fig.3.1.** Scatter Graphs for Each Station of The Actual and Predicted Values of The Variables Temperature, Relative. Humidity, and Solar Radiation.

## 4. CONCLUSION

The climatic variables play a significant role in agricultural process and irrigation management because we need to know all changes related to the climate which absolutely will effect on agricultural yield. Also, the climatic variables are very necessary for the hydrological process that can indicate to the quantity of water and how much water will be lost by evapotranspiration process. These parameters are affected by the climatic conditions and geographical locations; therefore, many different methods have been used in many previous studies to interpret and estimate the complexity of variables such as time series analysis using models' techniques to predict future values. In this study, the stochastic methods were applied to develop and increase the agricultural yield of different regions in Sudan using the ARIMA models which obtained satisfactory results of temperature, relative humidity, and solar radiation parameters. So, this study might be extremely helpful for agricultural engineers to achieve all the process related to the agricultural practices.

For instance, it is important to note that the performance of the models is reduced by the ARIMA models' poor forecasting, which has an impact on the model's goodness of fit. As a result, less-parameterized models yield better outcomes than over-parameterized ones.

### Conflict of interest:

The authors declare that there is no conflict of interest.

### Ethics:

This study does not require ethics committee approval.

### Author Contribution Rates:

Design of Study: MAMA (%50), BC (%50)

Data Acquisition: MAMA (%70), BC (%30)

Data Analysis: MAMA (%50), BC (%50)

Writing up: MAMA (%50), BC (%50)

Submission and Revision: MAMA (%70), BC (%30)

## REFERENCES


Ashrafzadeh, A., Kişi, O., Aghelpour, P., Biazar, S. M., & Masouleh, M. A. (2020). Comparative study of time series models, support vector machines, and GMDH in forecasting long-term evapotranspiration rates in northern Iran. *Journal of Irrigation and Drainage Engineering*, 146(6), 04020010.

Asteriou, D., & Hall, S. G. (2007). Applied Econometrics: a modern approach, revised edition. *Hampshire: Palgrave Macmillan*, 46(2), 117-155.

Bazrafshan, O., Salajegheh, A., Bazrafshan, J., Mahdavi, M., & Fatehi Maraj, A. (2015). Hydrological drought forecasting using ARIMA models (case study: Karkheh Basin). *Ecopersia*, 3(3), 1099-1117.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2008). Time series analysis: forecasting and control John Wiley & Sons. *Hoboken, NJ*.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control: John Wiley & Sons*.

Box, T. M., White, M. A., & Barr, S. H. (1994). A contingency model of new manufacturing firm performance. *Entrepreneurship theory and practice*, 18(2), 31-45.

Chen, C.-F., Chang, Y.-H., & Chang, Y.-W. (2009). Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. *Transportmetrica*, 5(2), 125-140.

Dickey, D. A. (2015). Stationarity issues in time series models. *SAS Users Group International*, 30.

Han, P., Wang, P., Tian, M., Zhang, S., Liu, J., & Zhu, D. (2012). *Application of the ARIMA models in drought forecasting using the standardized precipitation index*. Paper presented at the International Conference on Computer and Computing Technologies in Agriculture.

Han, P., Wang, P. X., & Zhang, S. Y. (2010). Drought forecasting based on the remote sensing data using ARIMA models. *Mathematical and computer modelling*, 51(11-12), 1398-1403.

Ibrahim, M. H., & Amin, R. M. (2005). Exchange Rate, monetary policy and manufacturing output in Malaysia.. *Journal of Economic Cooperation among Islamic Countries*, 26(3).

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, 37(1), 479-489.

Kim, B. S., Hossein, S. Z., & Choi, G. (2011). Evaluation of temporal-spatial precipitation variability and prediction using seasonal ARIMA model in Mongolia. *KSCE Journal of Civil Engineering*, 15(5), 917-925.

Lee, M. H. (2011). Forecasting of tourist arrivals using subset, multiplicative or additive seasonal Arima Model. MATEMATIKA: *Malaysian Journal of Industrial and Applied Mathematics*, 27, 169-182.

Luo, Y., Chang, X., Peng, S., Khan, S., Wang, W., Zheng, Q., & Cai, X. (2014). Short-term forecasting of daily reference evapotranspiration using the Hargreaves-Samani model and temperature forecasts. *Agricultural Water Management*, 136, 42-51.

MacKinnon, J. (1991). Critical values for cointegration tests. *Long-run economic relationships*, *13*.Mishra, A., & Desai, V. (2005). Drought forecasting using stochastic models. *Stochastic environmental researchand risk assessment*, 19(5), 326-339.

Mohan, S., & Arumugam, N. (1995). Forecasting weekly reference crop evapotranspiration series. *Hydrological sciences journal*, 40(6), 689-702.

Mossad, A., & Alazba, A. (2016). Simulation of temporal variation for reference evapotranspiration under arid climate. *Arabian Journal of Geosciences*, 9(5), 386.

Ozaki, T., & Oda, H. (1977). Non-linear time series model identification by Akaike's information criterion. *IFAC Proceedings Volumes*, 10(12), 83-91.

Palmroth, S., Katul, G. G., Hui, D., McCarthy, H. R., Jackson, R. B., & Oren, R. (2010). Estimation of long term basin scale evapotranspiration from streamflow time series. *Water Resources Research*, 46(10).

Profillidis, V. A., & Botzoris, G. N. (2018). *Modeling of transport demand: Analyzing, calculating, and forecasting transport demand*.

Saravanan, V. (2015). The Determinant of Consumer Price Index in Malaysia. *Journal of Economics, Business and Management*, 3(12).

Webster, R., & McBratney, A. (1989). On the Akaike information criterion for choosing models for variograms of soil properties. *Journal of Soil Science*, 40(3), 493-496.

Yürekli, K., Simsek, H., Cemek, B., & Karaman, S. (2007). Simulating climatic variables by using stochastic approach. *Building and environment*, 42(10), 3493-3499.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.