Turkish J. Eng. Env. Sci.
29 (2005) , 61 – 74.


# Testing the Residuals of an ARIMA Model on the Çekerek Stream Watershed in Turkey

**Kadri YÜREKLİ, Ahmet KURUNÇ**
*Gaziosmanpaşa University, Faculty of Agriculture, Tokat-TURKEY*
*e-mail: kadriyurekli@yahoo.com*
**Fazlı ÖZTÜRK**
*Ankara University, Faculty of Agriculture, Ankara-TURKEY*

Received 15.01.2004

**Abstract**

In ARIMA modeling studies, the selection of a best model fit to historical data is directly related to whether residual analysis is performed well. Therefore, diagnostic checks including the independence, normality and homoscedasticity of residuals is the most important stage of ARIMA model building. This study is concerned with testing residuals from ARIMA models for monthly streamflow data from the Çekerek Stream watershed. Alternative tests including the Ljung-Box Q statistic, runs test and turning point test for independence analysis of the residuals; Kolmogorov-Smirnov and Anderson-Darling tests for normality of residuals; and Goldfeld-Quandt, Breusch and Pagan and Spearman's rho approaches for the homoscedasticity of residuals were used. The selected parsimony model for each data set among the ARIMA models fulfilled the diagnostic checks, considering the Schwarz Bayesian criterion.

**Key words:** ARIMA model, Monthly streamflow, Çekerek stream, Diagnostic checks.

## Introduction

Records of hydrologic variables such as rainfall and runoff are the basis of studies related to the planning and design of water resource projects. However, hydrologic data are generally unavailable, and, in conditions where they are available, these records may be too short to detect any statistically significant meaning. If the period of available data is shorter than the economic life of the project involving the planning and management of water resources, failure of the project results. In this sense, for many hydrologic studies, generating data in the form of a time series is necessary. Time series analysis and modeling, commonly used in water resources, are 2 of the major tasks in hydrologic research and development (Chatfield, 1991). The operation and management of a hydraulic structure such as a dam demand reliable information concerning monthly flow (Bayazit, 1981).

When generating synthetic data for a hydrologic variable, it is necessary to take into consideration the data that are statistically similar to the observed data (Sharma *et al.*, 1997). One of the most important and highly popularized time series models is the Box-Jenkins approach, commonly known as ARIMA (autoregressive integrated moving average). Researchers have used this approach for many different scientific and technical applications. McKerchar and Delleur (1974) used an ARIMA process to achieve stochastic modeling of monthly flows. McLeod *et al.* (1977) applied the ARIMA approach to average annual streamflows, annual sunspot number series and monthly airline passenger data and suggested a different ARIMA model for each data set. Fernando and Jayawardena (1994) used various ARIMA models in forecasting monthly rainfall records. Venama *et al.* (1996) investigated climate change in the Senegal River basin via this approach. Chaloulakou *et al.* (1999) forecasted daily maximum 1-h ozone concentrations, whereas Ahmad *et al.* (2001) analyzed water quality data using an ARIMA model.

To test the fit of Box-Jenkins ARIMA models, there are 3 steps; identification, estimation and diagnostic check (McKerchar and Delleur, 1974; Box and Jenkins, 1976; Hipel *et al.*, 1977; Enders, 1995; McLeod, 1995; Zhang, 2003). To select alternative models, the diagnostic check step is very important in ARIMA modeling. This step is focused on residuals from Box-Jenkins ARIMA models to detect whether residuals are independent, homoscedastic and normally distributed. In order to use a Box-Jenkins ARIMA model for the purpose of generating hydrologic data, the model should fulfill all the diagnostic checks.

The work in this paper is concerned with the diagnostic checking of residuals from the ARIMA models fitted to monthly streamflow data for 3 gauging stations located on the Çekerek Stream watershed by alternative methods.

## Materials and Methods

### Study area

Monthly streamflow data from 3 gauging stations (1424, 1409 and 1424) managed by the Electrical Power Resources Survey and Development Administration (EIE) at the Çekerek Stream watershed were used as materials. The Çekerek Stream watershed is bounded by 39° 30′ and 40° 45′ N latitudes, and 35° 15′ and 36° 15′ E longitudes. This area covers approximately 1,165,440 ha, which is about 1.5% of Turkey's total area. The approximate locations of the gauging stations are shown in Figure 1 and information related to these stations is presented in Table 1.

The study area is located on the north Anatolia fault line, which is one of the most important faults in the world. Therefore, tectonic movement affects the characteristics of the watershed. Çekerek Stream is formed by the joining together of small streams that originate from the Kızık, Dinar, Çalı and Kavak hills, near Çamlıbel district. Çekerek Stream is approximately 276 km in length. The stream joins the

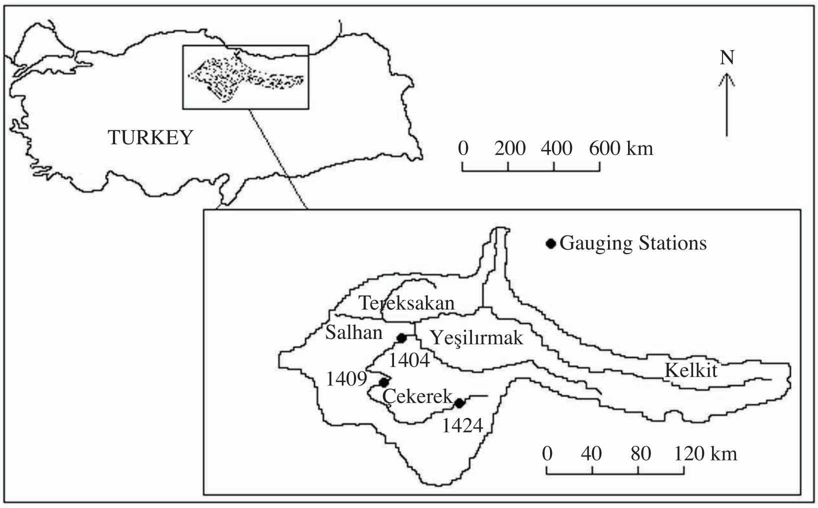


**Figure 1.** Location of the study area and the gauging stations of Çekerek Stream in the Yeşilırmak basin.

**Table 1.** Summary of information about the gauging stations.

| Station Number | Station Name | Drainage Area, $km^2$ | Number of years of data |
|---|---|---|---|
| 1404 | Çekerek-Kayabaşı | 11,724.0 | 13 |
| 1409 | Çekerek-Akçakeçili | 5,267.6 | 38 |
| 1424 | Çekerek-Çırdak Bridge | 1,032.8 | 27 |

Yeşilırmak River near Kayabaşı. The water quality of the stream is $C_2S_1$ (low salinity-low sodium) (GDSW, 1970), which can be used for irrigation purposes for plants with moderate salt tolerance in most cases without special practices for salinity control (Chhabra, 1996).

**The ARIMA model for application**

In order to analyze the monthly time series from the 3 gauging stations, an ARIMA modeling approach was used in this study. A seasonal ARIMA model denoted as ARIMA (p,d,q)*(P,D,Q) that is a combination of past values and past residuals can be written as follows (Janacek and Swift, 1993; Ahmad *et al.*, 2001; Sun and Koch, 2001):

$$Ø(B)\Phi F(B^s)(w_i - \mu) = C + \theta(B)\Theta(B^s)a_i \quad (1)$$

$$w_i = (1-B)^d(1-B^s)^D x_i \quad (2)$$

$$Ø(B) = 1 - Ø_1 B - Ø_2 B^2 - \ldots - Ø_p B^p \quad (3)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \quad (4)$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps} \quad (5)$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs} \quad (6)$$

There are 3 stages for fitting a seasonal ARIMA model to a given time series: model identification, parameter estimation and diagnostic checking (Box and Jenkins, 1976). The identification stage is intended to determine the differencing required to produce stationarity, and the order of nonseasonal and seasonal AR and MA operators for a given series. The estimation stage consists of using the data to estimate and to make inferences about the values of the parameters conditional on the tentatively identified model. The parameters are estimated such that an overall measure of residuals is minimized. This can be done with a nonlinear optimization procedure (Zhang, 2003). The diagnostic checking of model adequacy determines whether residuals are independent, homoscedastic and normally distributed.

The ARIMA model requires the use of stationary time series data (Dickey and Fuller, 1981). A stationary time series has the property that its statistical characteristics such as the mean and the autocorrelation structure are constant over time. When the observed time series presents a trend and heteroscedasticity, differencing and power transformation are often applied to the data to remove the trend and stabilize variance before an ARIMA model can be fitted. The existence or lack of stationarity in a time series can be detected by alternative methods, i.e. nonparametric tests (including Kendall's tau, Mann-Kendall and Sen tests), a unit root test and a graphical approach (Hipel and McLeod, 1994; Gibbons, 1997; Greene, 2000).

By plotting original series, stochastic trends (nonstationarity) in the mean and variance may be revealed (Box and Jenkins, 1976). Under current practice, such data requires that the observed data series be tested for unit roots. One of unit the root approaches commonly used to explain whether a time series is nonstationary is the Dickey and Fuller (DF) test (Dickey and Fuller, 1981). Residuals of the regression based on the relationship between the current value (in time t) and the last value (in time t-1) in a given time series were assumed to be independent and to have constant variance in this test. Therefore, under the conditions that residuals have serial correlation, the $\sum_{i=1}^{p} d_i \Delta y_{t-i}$ term should be augmented to DF test regression to remove serial correlation in residuals. This approach is called the augmented Dickey and Fuller (ADF) test. Typically, this test is expressed as (Enders, 1995)

$$\Delta Y_t = \alpha_0 + \beta t + \alpha_1 Y_{t-1} + \sum_{i=1}^{P} d_i \Delta Y_{t-i} + u_t \quad (7)$$

where $\Delta Y_t = Y_t - Y_{t-1}, \alpha_0$ is the drift (constant) term and t is the time trend with the null hypothesis $H_0$: $\alpha_1 \neq 0$ and alternative hypothesis $H_1$: $\alpha_1 = 0, P$ is the number of lags necessary to obtain white noise or to remove serial correlation in residuals and $u_t$ is the error term. The test involves examining the $t$ statistic, $t(\alpha_1)$, for the parameter $\alpha_1$, under the null hypothesis that $\alpha_1 = 0$. The null hypothesis is rejected if the t statistic is larger than the critical value, $\tau_1$, obtained from MacKinnon (1990). Note that the critical values are not those used in typical applications of the $t$ statistic, $t(\alpha_1)$, in linear regression models.

To determine the possible persistence structure in the data set, an autocorrelation function (ACF) and partial autocorrelation function (PACF) should be used (Janacek and Swift, 1993). The ACF and PACF give information about the nonseasonal and seasonal AR and MA operators for a time series. The ACF provides significant information about the correlation between pairs of observations that are k time units apart, called lag. The identification of the appropriate parametric time series model depends on the shape of the ACF (Brooks, 2002). In general, for an MA $(0, d, q)$ process, an autocorrelation coefficient $(r_k)$ with the order of k cuts off and is not significantly different from zero after lag q. If $r_k$ tails off and does not truncate, this suggests that an AR term is needed to model the time series. When the process is an MA $(0, d, q)$ x $(0, D, Q)$, $r_k$ truncates and is not significantly different from zero after lag $q + sQ$. If $r_k$ attenuates at lags that are multiples of s, this implies the presence of a seasonal AR component. For an AR $(p, d, 0)$ process, the partial autocorrelation coefficient $(\o_{kk})$ with the order of $k$ truncates and is not significantly different from zero after lag p. If $\o_{kk}$ tails off, this implies that an MA term is required. When the process is an AR $(0, d, q)$ x $(0, D, Q)$, $\o_{kk}$ cuts off and is not significantly different from zero after lag p+sP. If $\o_{kk}$ damps out at lags that are multiples of s, this suggests the incorporation of a seasonal MA component into the model. The ACF for seasonal series should not exceed a maximum lag of approximately $5s(5s < n/4)$. The PACF is usually calculated for 20 to about 40 lags $(40 < n/4)$. For seasonal models, higher lags of the PACF may be required for identification (Hipel *et al.*, 1977).

The diagnostic checking of model adequacy is the last stage of model building. The tests related to independence, normality and homoscedasticity should be performed at this stage as the residuals $(a_i)$ from an ARIMA model are assumed to be independent, homoscedastic, and usually normally distributed. Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit of the tentative model to the historical data. If the constant variance and normality assumptions are not true, they are often reasonably well satisfied when the observations are transformed by a Box-Cox transformation (Hipel *et al.*, 1977; Bayazıt, 1981). The Box-Cox transformation can be given as

$$z_{i=1}^{n} = \lambda^{-1}\left[(x_{i=1}^{n} + c)^{\lambda} - 1\right]\lambda \neq 0 \tag{8}$$

$$z_{i=1}^{n} = \ln\left(x_{i=1}^{n} + c\right)\lambda = 0 \tag{9}$$

If the model is inadequate, the 3-step model building process is typically repeated several times until a satisfactory model is finally obtained. The final selected model can then be used for prediction purposes (Wei, 1990).

**Alternative test approaches for diagnostic checking of independence**

Testing for independence (randomness) against serial dependence is a fundamental problem in time series analysis. To determine whether a time series, x(t), is independent, the function (ACF) of the series is examined. If the ACF is significantly different from zero, this implies that there is dependence between observations. Therefore, ACF is a powerful complementary tool for testing independence (Janacek and Swift, 1993; Ferguson *et al.*, 2000). The residual autocorrelation function (RACF) should be obtained to determine whether the residuals are white noise. There are different applications related to the RACF for the independence of residuals. The first one is the correlogram drawn by plotting $r_k$ (a) against lag $k$.

$$r_{ak} = \frac{\sum\limits_{t=k+1}^{n} a_t a_{t-k}}{\sum\limits_{t=1}^{n} a_t^2} \tag{10}$$

Under the assumption that $a_t$ follows a white noise process the standard errors of these $r_{ak}$ are approximately equal to $1\big/\sqrt{T}$. Thus, under the null hypothesis that $a_t$ follows a white noise process, roughly 95% of the $r_{ak}$ should fall within the range $\pm 1.96/\sqrt{T}$. If more than 5% of the $r_{ak}$ fall outside of this range, then most likely $a_t$ does not follow a white noise process (Lehmann and Rode, 2001).

There are many statistical tests used for diagnostic checking of randomness. In this study, the Ljung-Box Q statistic, and turning point and runs tests were used as alternative approaches for the diagnostic checking of residuals for independence.

**Ljung-Box Q (LBQ) Statistic:** The Ljung-Box Q or $Q(r)$ statistic can be employed to check independence instead of visual inspection of the sample autocorrelations. A test of this hypothesis can be done for the model adequacy by choosing a level of significance and then comparing the value of calculated $\chi^2$ with the $\chi^2$-table critical value, the present model is adequate on the basis of available data. The $Q(r)$ statistic is calculated by the following equation (Ljung and Box, 1978):

$$Q(r) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}r_k(a)^2 \tag{11}$$

**Turning Point Test:** A turning point is when the series changes from increasing to decreasing or vice versa. That is, $X_{t-1} < X_t > X_{t+1}$ or $X_{t-1} > X_t < X_{t+1}$. Equations (12) and (13) can be used to calculate the mean and variance. Then, by counting the number of turning points, the $N_T$ can be approximated well with Eq. (14). Let T = the number of turning points in an n period series. The $N_T$ statistic should be compared with the z-table critical value. If the $N_T$ statistic lies within the 5% significance interval, the null hypothesis related to the independence of the data set is accepted. The hypothesis of randomness should be rejected at $\alpha$ significance level if the absolute value of $N_T > N_{T(1-\alpha/2)}$, where $N_{T(1-\alpha/2)}$ is the 1- $\alpha/2$ quantile of standard normal distribution (Cromwell *et al.*, 1994).

$$\mu_T = (2/3)(n-2) = \tag{12}$$

$$\sigma_T^2 = (16n-29)/90 \tag{13}$$

$$N_T = |T-\mu_T|/\sigma_T^2 \approx N(0,1) \tag{14}$$

**Runs Test:** The runs test can be used to decide if a data set is from a random process. A run is defined as a series of increasing values or a series of decreasing values. The number of increasing (or decreasing) values is the length of the run. In a random data set, the probability that the $(i+1)^{th}$ value is larger or smaller than the $i^{th}$ value follows a binomial distribution, which forms the basis of the runs test. The first step in the runs test is to compute the sequential differences $(Y_i - Y_{i-1})$. Positive values indicate an increasing value, whereas negative values indicate a decreasing value. In other terms, if $Y_i > Y_{i-1}$ a 1 (one) is assigned for an observation and a 0 (zero) otherwise. The series then has an associated series of 1s and 0s. To determine if the number of runs is the correct number for a series that is random, let T be the number of observations, $T_A$ be the number above the mean, $T_B$ be the number below the mean and R be the observed number of runs. Then, using combinatorial methods, the probability P(R) can be established and the mean and variance of R can be derived (Cromwell *et al.*, 1994; Gibbons, 1997). When T is relatively large (>20) the distribution of R is approximately normal.

$$E(R) = \frac{T + 2T_AT_B}{T} \tag{15}$$

$$V(R) = \frac{2T_AT_B(2T_AT_B - T}{T^2(T-1)} \tag{16}$$

$$Z_N = \frac{R - E(R)}{\sqrt{V(R)}} \approx N(0,1) \tag{17}$$

The null hypothesis is rejected if the calculated $Z_N$ value is greater than the selected critical value obtained from the standard normal distribution table. In other words, the x(t) series is decided to be non-random.

**Alternative test approaches for diagnostic checking of normality**

If a data set is distributed according to the bell-shaped curve of the normal distribution, this set can be referred to as normal. Therefore, the histogram of a data set gives information related to normality. In addition, it is well known that a normal distribution is not skewed and is defined to have a coefficient of kurtosis of 3 (Brooks, 2002).

There are several statistical tests used for the diagnostic checking of normality. In this study, Kolmogorov-Smirnov, Anderson-Darling and skewness tests were used as alternative approaches for the diagnostic checking of residuals for normality.

**Kolmogorov-Smirnov (K-S) Test:** This is a non-parametric test of data fitting to a theoretical distribution using the maximum absolute deviation (D) between the 2 functions of cumulative distribution. The K-S test is distribution free; therefore, the critical values do not depend on the specific distribution being tested. The maximum absolute deviation is as follows (Haan, 1977):

$$D = max|Fn(x) - Fa(x)| \quad (18)$$

where Fn (x) is the cumulative density function based on n measurements, and Fa (x) is the specified theoretical cumulative distribution function. The value of the D statistic is compared with the critical value D (n, $\alpha$) obtained from Haan (1977). If D is greater than the critical value D (n, $\alpha$), the null hypothesis related to normality is rejected for the chosen level of significance.

**Anderson-Darling Test:** The Anderson-Darling (AD) test is a member of the group of goodness of fit statistics known as empirical distribution function statistics. This test is widely used in practice to test normality. It is a modification of the K-S test and gives more weight to the tails than does the K-S test. The AD test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution. The AD test for normality has the functional form (Thode, 2002)

$$AD = \sum_{i=1}^{n} \frac{1-2i}{n} \{ln(F_0[Z_i]) + ln(1 - F_0[Z_{(n+1-i)}])\} - n \quad (19)$$

where $F_o$ is the cumulative distribution function of the specified distribution and $Z_i$ is the ordered data. The null hypothesis, which has normality at the 0.05 significance level, cannot be rejected if the AD test statistic is smaller than the critical value (CV) from Eq. (20).

$$CV = \frac{0.752}{(1 + 0.75/n + 2.25/n^2)} \quad (20)$$

**Skewness Test:** This method is used for testing whether a sample comes from a normal population. If a given series has a normal distribution, the skewness measure ($\gamma$) of the series is approximately normally distributed with zero and standard deviation $(6N^{-1})^{1/2}$ (Snedecor and Cochran, 1989; Salas *et al.*, 1980). The series is decided to have a skewed distribution if the skewness statistic is greater than its standard deviation. However, when the sample size is less than 150, the table given in Snedecor and Cochran (1989) is suggested for the standard deviation of skewness statistic.

**Alternative test approaches for diagnostic checking of homosecdasticity**

For the diagnostic checking of residuals in terms of homoscedasticty, Goldfeld-Quandt, Breusch and Pagan and Spearman's rho tests are commonly used.

**Goldfeld-Quandt Test:** This test is very useful for determining whether a transformation of the data is needed. If there is a change in variance (heteroscedasticity) of residuals, a transformation is necessary for the data. The Goldfeld-Quandt test statistic ($F_{cal}$) can be obtained from Eq. (21) (Greene, 2000):

$$F_{cal} = \frac{ESS_H/(n_H - k_p)}{ESS_L/(n_L - k_p)} \approx F_{table}[(n_H - k_p), (n_L - k_p)] \quad (21)$$

If $F_{cal}$ is smaller than the F-table critical value, the residuals are assumed to be homoscedastic. The following is the method for the Goldfeld-Quandt statistic:

1. Rank or order the residuals from the model fit to the data in ascending order,
2. Omit c central residuals from the ranked residuals, where c may be approximately 1/4 or 1/5 of total residuals,
3. Obtain 2 subsets of residuals (below and above residuals of the omitted residuals),
4. Calculate sum of squares for 2 subsets of residuals.
5. Calculate $F_{cal}$ value using the equation.

**Breusch and Pagan (B-P) Test:** This test is similar to the Goldfeld-Quandt test but the main difference is that central residuals from the ranked residuals are not omitted. For the test, the residuals from the model fit to the data are divided into 2 groups. Then the residual sum of squares ($ESS_F$, $ESS_S$) for these groups is obtained. Equation (22) can be used to calculate the Breusch-Pagan test statistic ($F_{cal}$) (Breusch and Pagan, 1979):

$$F_{cal} = \frac{ESS_H/(n_H - k_p)}{ESS_L/(n_L - k_p)} \approx F_{table}[(n_H - k_p), (n_L - k_p)] \quad (22)$$

If $F_{cal}$ is smaller than the F-table critical value, the residuals are assumed to be homoscedastic.

**Spearman's Rho Test:** Spearman's rho test, a nonparametric test, is another approach that can be used for determining the homoscedasticity of residuals. It is assumed that the variance of residuals is directly related to the size of the observation on x(t). If this is true, then the rank of the i$^{th}$ observation on x(t), say $d_i^x$, should correspond to the observed rank $d_i^a$. Hence, the differences in the ranks for observations and residuals would be expected to be zero on average. For this reason, observations and residuals are arranged in ascending order. Then, ranks of these variables are signed. The equations related to Spearman's rho statistic ($t_{cal}$) are as follows:

$$R_{sp} = 1 - \frac{6 * \sum D_i^2}{n * (n^2 - 1)} \tag{23}$$

$$D_i = d_i^x - d_i^a \tag{24}$$

$$t_{cal} = R_{sp} * \left[\frac{(n-2)}{(1 - R_{sp}^2)}\right]^{1/2} \tag{25}$$

where $D_i$ is the i$^{th}$ difference between rankings, $d_i^x$ is the rank of a measured variable in chronological order, and $d_i^a$ is the rank in the historical data of the i$^{th}$ observation in the ascended data. To detect whether residuals have a constant variance, the calculated $t_{cal}$ value in Eq. (5) should be compared to the t-critical value from the tables. If the $t_{cal}$ value lies within the 5% significance interval, the residuals have a constant variance.

### Schwartz bayesian criterion (SBC) for goodness-of-fit measure

Wei (1990) expressed the need to select the model that has fulfilled all the diagnostic checks and has as few parameters as possible in terms of parsimony. The SBC can measure the parsimony of model building. An ARIMA model that has the smallest SBC value among the competing models fit to time series is preferred. The mathematical formulation of the SBC can be given as follows (Wei, 1990):

$$SBC = -2L(\sum \hat{a}_t^2) + K \ln(n) \tag{26}$$

where K = p + q + P+ Q + 1, L($\sum \hat{a}_t^2$) = the log of the likelihood function of the Box-Jenkins ARIMA(p, d, q) (P, D, Q) model. The log likelihood function, $L(\sum \hat{a}_t^2)$, is a monotonically decreasing function of the sum of squared residuals, $\sum \hat{a}_t^2$.

### Results

The results of the ADF test applied to monthly data sequences to test whether monthly data sequences of 1424, 1409 and 1404 gauging stations are stationary or nonstationary are presented in Table 2. The ADF test statistics (t ($\alpha_1$)) for the data sequences of monthly streamflow from each gauging station were greater than the critical values obtained from MacKinnon (1990) at 0.01, 0.05 and 0.10 significance levels. According to these results, the null hypothesis, which has a unit root, for the data sequences should be rejected at 0.01, 0.05 and 0.10 significance levels. For each data sequence from 1424, 1409 and 1404 gauging stations, maximum lag lengths of theADF test to remove serial correlation from the residuals of the regressions based on the relationship between the current value and the last value of the data sequences from the mentioned gauging stations were selected as 10, 11 and 1, respectively. The values (V) of the parameters associated with the standard errors (SEV), t-ratios ($t_{cal}$) for drift (constant) and trend parameters in regressions of the data sequences are also presented in Table 2. The t-ratios ($t_{cal}$) related to constant and trend coefficients were compared with the critical value of 1.96 obtained from the t-distribution at the 0.05 significance level. Only t-ratios ($t_{cal}$) of constant coefficients obtained for monthly streamflow for 1409 and 1424 gauging stations were greater than the critical value (Table 2). Therefore, constant parameters for the data sequences of the mentioned gauging stations should be involved in regressions.

**Table 2.** Unit root test results for trend analysis of monthly streamflows.

| Gauge | ADF | Test Critical Value | | | Constant | | | Trend | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Station | Statistic | 0.01 | 0.05 | 0.10 | V | SEV | t-ratio | V | SEV | t-ratio |
| 1424 | -4.02 | -3.99 | -3.42 | -3.14 | 161.13 | 56.33 | 2.86 | 0.33 | 0.22 | 1.49 |
| 1409 | -4.22 | -3.98 | -3.42 | -3.13 | 7.96 | 2.69 | 2.96 | 0.01 | 0.01 | 0.93 |
| 1404 | -6.57 | -4.02 | -3.44 | -3.14 | 6.06 | 3.40 | 1.78 | 0.05 | 0.04 | 1.32 |

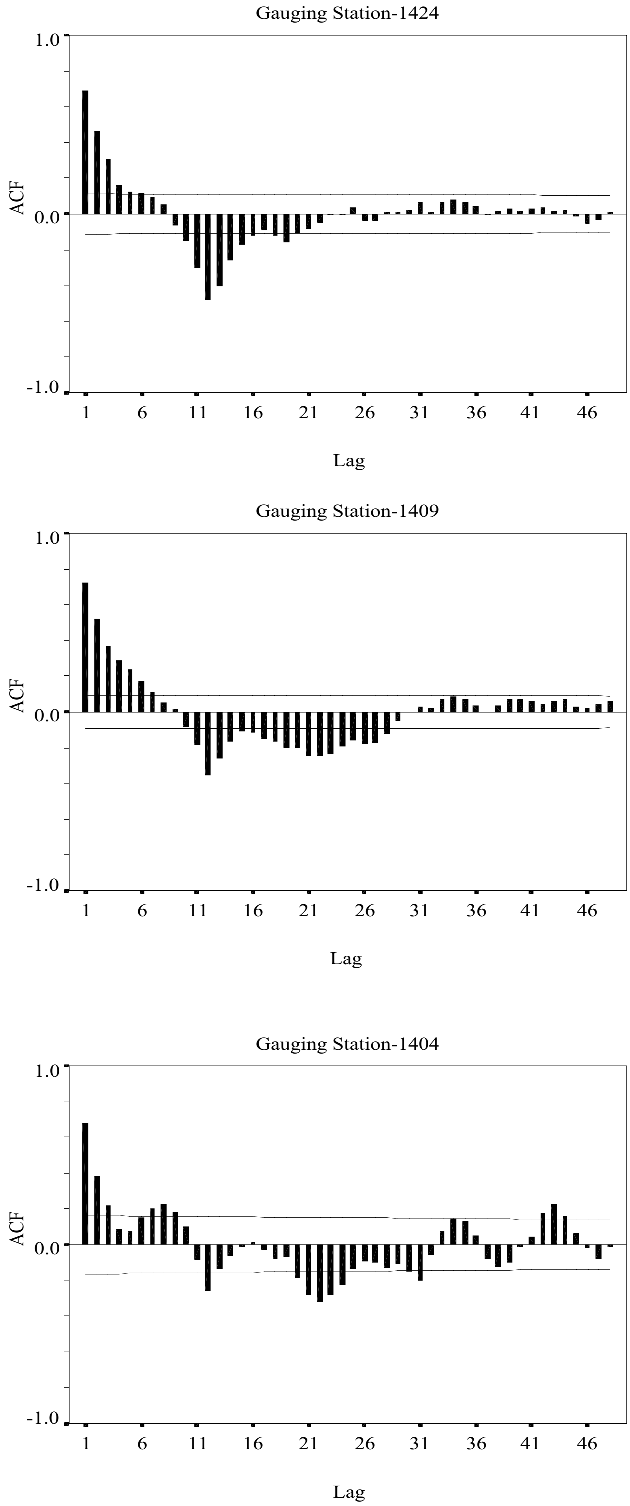


**Figure 2.** ACF seasonal differenced monthly streamflow data.

The plots of the ACF and the PACF for each monthly data sequence were drawn to gather information about the seasonal and nonseasonal AR and MA operators concerning the monthly series. The ACF graphs show an attenuating sine wave pattern that reflects the random periodicity of the data and possibly indicates the need for non-seasonal and/or seasonal AR terms in the model. For these data sequences with the cyclic seasonal component, seasonal differencing was needed. By taking the seasonal differencing operator as one (1), the seasonal wave pattern in the ACFs was removed (Figure 2).

All the ACFs for monthly streamflow data from each gauging station were significantly different from zero. In addition, the Ljung-Box Q statistics calculated for ACFs were rather high compared to $\chi^2$ critical value. All of these results emphasize that the ACFs obtained for each monthly data sequence were significantly different from zero. In other words, there was a significant linear dependence between monthly observations. However, the ACFs did not cut off but rather damped out. This may suggest the presence of autoregressive (AR) terms. The PACFs possess significant values at some lags but rather tail off. This may imply the presence of moving average (MA) terms. The ACFs have significant values at lags that are multiples of 12. This may stress that seasonal AR terms are required but these values attenuate. There are peaks in the graphs of the PACFs at lags that are multiples of 12, which may suggest seasonal MA terms, but these peaks damp out.

Alternative models were selected by inspecting these diagrams and considering the principle of parsimony. Diagnostic checks were applied in order to determine whether the residuals of the alternative models were independent, homoscedastic and normally distributed. The residuals from the models given in Table 3 by taking into consideration monthly data sequences, which are not transformed, of each gauge station were tested with the K-S test for normality and the B-P test for homoscedasticity. The $D_{cal}$ values calculated based on K-S for 1424, 1409 and 1404 gauging stations, which are 0.184, 0.189 and 0.223, respectively, were greater than the $D_{Table}$ values given in Table 5. This implies that the residuals from the models do not come from a normal distribution. According to B-P, the $F_{cal}$ values were 1.809, 2.155 and 2.851, respectively. These values were greater than the $F_{Table}$ values given in Table 6. These results show a change in residual variances from the models. Normality and homoscedasticity results indicated that a Box-Cox transformation was required for the monthly data of each gauging station. The homoscedasticity and normality assumptions for residuals were achieved using Eq. (8) ($\lambda$ = -0.5 and c = 1) for gauging station 1424 and Eq. (9) (c = 0) for gauging stations 1409 and 1404. These transformations caused the residuals to be homoscedastic and approximately normally distributed. All of the ARIMA models selected from the ACF and PACF graphs did not fulfill the residual assumptions (independent, homoscedastic and normality). The models that did not fulfill at least one of the diagnostic checks were eliminated. The SBC was taken into account to obtain a parsimony model among the models fulfilling all the diagnostic checks. The model with the minimum SBC was assumed to be parsimonious. Thus, these models were expressed as the best model for time series of monthly data from each gauging station. The selected best models are presented in Table 3. The ARIMA model equations for the gauging stations (1424, 1409 and 1404, respectively) are as follows:

$$(1 - 0.70B)[(1 - B^{12})z_i] = (1 - 0.94B^{12})a_i \quad (27)$$

$$(1 - 0.76B)[(1 - B^{12})z_i] = (1 - 0.90B^{12})a_i \quad (28)$$

$$(1 - 0.86B + 0.14B^2)[(1 - B^{12})z_i] = (1 - 0.84B^{12})a_i \quad (29)$$

**Table 3.** Summary of the statistical parameters of the best fitted ARIMA models for each gauging station.

| ARIMA Model | Model Parameter | Variables in the model Value of parameter | Standard error | t-ratio | Probability < 0.05 | SBC SBC |
|---|---|---|---|---|---|---|
| (1,0,0)(0,1,1) (1424) | $\varnothing_1$ | 0.70 | 0.039 | 17.86 | 0.000 | -256.8 |
| | $\Theta_1$ | 0.94 | 0.049 | 19.12 | 0.000 | |
| (1,0,0)(0,1,1) (1409) | $\varnothing_1$ | 0.76 | 0.030 | 25.12 | 0.000 | 599.3 |
| | $\Theta_1$ | 0.90 | 0.028 | 32.10 | 0.000 | |
| (2,0,0)(0,1,1) (1404) | $\varnothing_1$ | 0.86 | 0.079 | 10.79 | 0.000 | 258.9 |
| | $\varnothing_2$ | -0.14 | 0.081 | -1.78 | 0.077 | |
| | $\Theta_1$ | 0.84 | 0.082 | 10.19 | 0.000 | |

**Table 4.** Independence test results of monthly streamflow data for each gauging station.

| Gauge Station | L-B Test k = 48 | | Decision | Runs | | Decision | Turning Test | | Decision |
|---|---|---|---|---|---|---|---|---|---|
| | Q(r) | $\chi^2_{Table}$ | | $Z_N$ | $Z_{Table}$ | | $N_T$ | $N_{Table}$ | |
| 1424 | 60.20 | 62.83 | R | -0.564 | ±1.96 | R | 0.070 | ±1.96 | R |
| 1409 | 47.08 | 62.83 | R | 0.161 | ±1.96 | R | 0.059 | ±1.96 | R |
| 1404 | 50.07 | 61.66 | R | 0.167 | ±1.96 | R | 0.220 | ±1.96 | R |
| R, residuals are independent | | | | | | | | | |

**Table 5.** Normality test results of monthly streamflow data for each gauging station.

| Gauge Station | K-S Test | | Decision | A-D Test | | Decision | Skewness Test | | Decision |
|---|---|---|---|---|---|---|---|---|---|
| | $D_{cal}$ | $D_{Table}$ | | AD | CV | | $\gamma$ | SD | |
| 1424 | 0.038 | 0.077 | ND | 0.557 | 0.750 | ND | 0.090 | 0.139 | ND |
| 1409 | 0.034 | 0.065 | ND | 0.690 | 0.751 | ND | 0.096 | 0.118 | ND |
| 1404 | 0.049 | 0.113 | ND | 0.408 | 0.748 | ND | -0.152 | 0.199 | ND |
| ND, residuals are normally distributed<br>$\gamma$, skewness statistic for the sample<br>SD, standard deviation related to skewness statistic | | | | | | | | | |

The values (V) of the parameters associated with the standard errors (SEV), t-ratios and probabilities (<5%) for the standard errors are also listed in Table 3. The standard errors calculated for the model parameters were rather small (<5%) compared to the parameter values. Furthermore, even at the 1% significance level, all of the parameters, except for $Ø_2$ for gauging station 1404, are significant; thus, these parameters should be included in the models (Table 3). The $Ø_2$parameter is significant at the 0.10 confidence level. Although the parameter is insignificant at the 5% confidence level, the selected model for gauging station 1404 is the best model according to the constraints mentioned above.

Three approaches, namely the Ljung-Box Q statistic, and runs and turning point tests, are applied for the independence assumption of residuals for the best models. The results of these tests were presented in Table 4. The selected best models were consistent with the independence assumption for all tests. The test statistics were smaller than the critical values from the tables related to the tests. This implies that residuals from the best models are independence or white noise. Furthermore, the RACF drawn for the best models (Figure 3) indicated that the residuals were not significantly different from a white noise series at the 5% significance level.

For the selected best models, the results related to the normality of residuals using K-S, AD and skewness tests are given in Table 5. The test statistics ($D_{cal}$ and AD) based on these 2 methods were smaller than the critical values ($D_{Table}$ and CV) at the 5% level of significance. As can be seen in Table 5, skewness measures ($\gamma$) related to residuals from each model were less than its standard deviation (SD). These results suggest that the residuals of the best models are normally distributed. In addition to these tests, Figure 4 shows the frequency histograms of the residuals. As expected, the normal curves significantly reflect a normal distribution.

Test statistics from the G-Q, B-P and Spearman's rho approaches for the homoscedasticity of the residuals are presented in Table 6. Test statistics ($F_{cal}$) from G-Q and B-P methods were smaller than the critical values obtained from $F_{Table}$ at the 5% significance level. These results concerning the G-Q and B-P tests imply that the residual variances are constant. Similarly, the homoscedasticity test statistic ($T_{cal}$) of residuals based on Spearman's rho approach was also smaller than the critical value ($T_{Table}$) from the t-distribution at the 5% significance level. Spear-

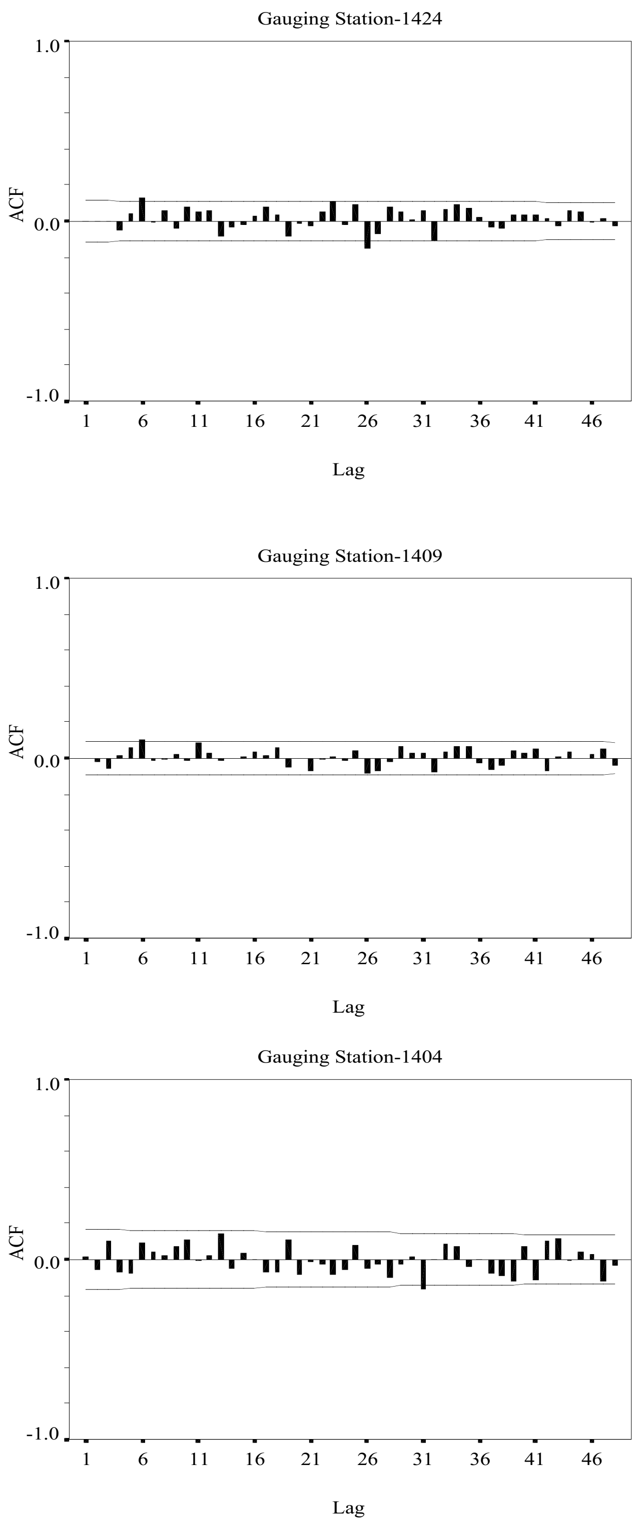


**Figure 3.** Residual ACF monthly streamflow data for each gauging station.

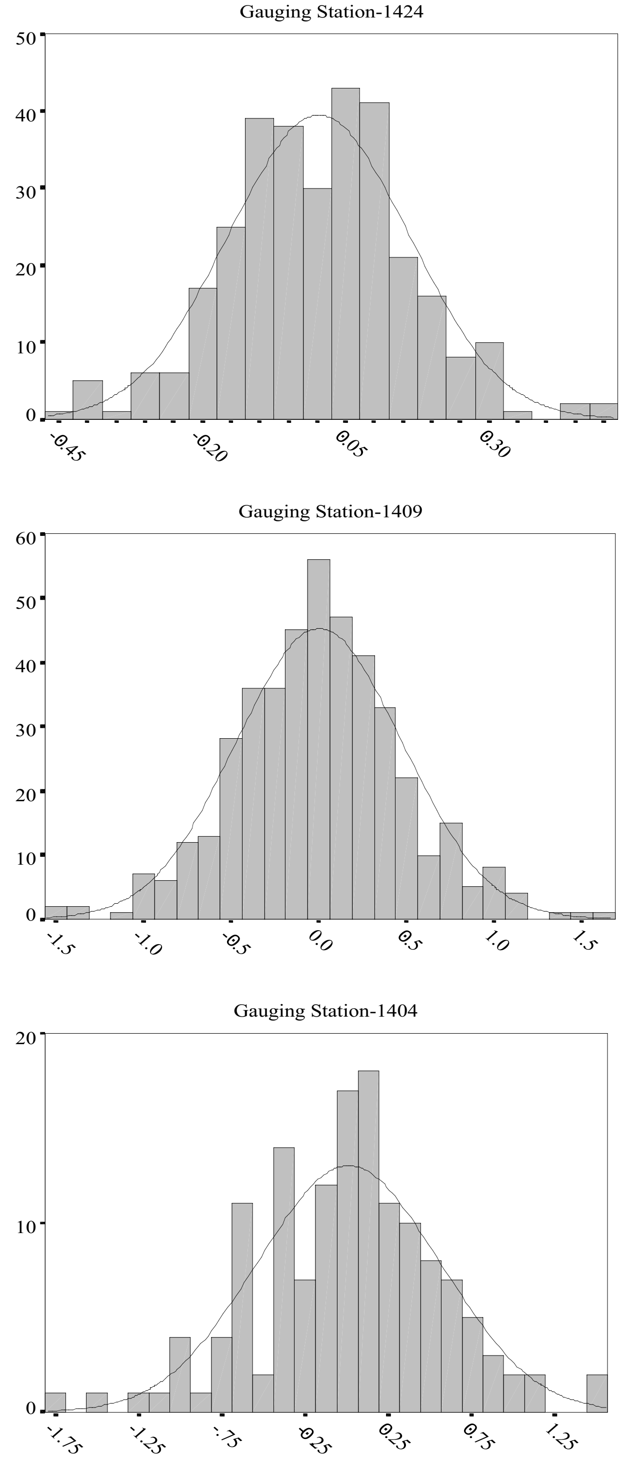


**Figure 4.** Frequencies of residuals from the best model for each gauging station.

man's rho approach also causes no change in the residual variances. Thus, all 3 methods satisfy condition the related to homoscedasticity.

Comparisons of monthly mean flow and standard deviation values for observed and predicted data from the ARIMA model are given in Figure 5. To determine whether there is significant difference for the mean flow and standard deviation values of the observed and predicted data for each month, a z-test (for means) and F-test (for standard deviations) were applied (Haan, 1977; Devore and Peck, 1993). Since monthly mean values from observed and predicted data for each gauging station were between z-critical table values (± 1.96 for 2 tailed at the 5% significance level), the data support the claim that there is no difference between the mean values of observed and predicted data. Similarly, monthly standard deviation values from observed and predicted data for each gauging station were smaller than F-critical table values at the 5% significance level (1.96 for 1424, 1.76 for 1409 and 2.82 for 1404 gauging station). Furthermore, these results show that the predicted data preserve the basic statistical properties of the observed series.

The results of the relationship between the observed and predicted monthly data sequences in terms of regression are presented in Table 7. The coefficient of correlation (R), which measures the strength of the association between 2 variables, and the significance level ($R_{sig}$) related to the R of regression for each gauge station shows that there is a statistically significant linear relationship between the observed and predicted data. On the other hand, the coefficient of determination (R-square), which is interpreted as the proportionate reduction of total variation associated with the use of the predictor variable (the observed data in this study), and adjusted R-square measure, which presents the sample response of the population for each regression, were close to one. In addition, the results (F-value and $F_{Sig}$) concerning tests applied for determining whether the estimated regression functions adequately fit the data emphasize that the association between the observed and predicted monthly data sequences is linear. Based on these results, it is concluded that the selected best ARIMA model for each station can make accurate estimates.

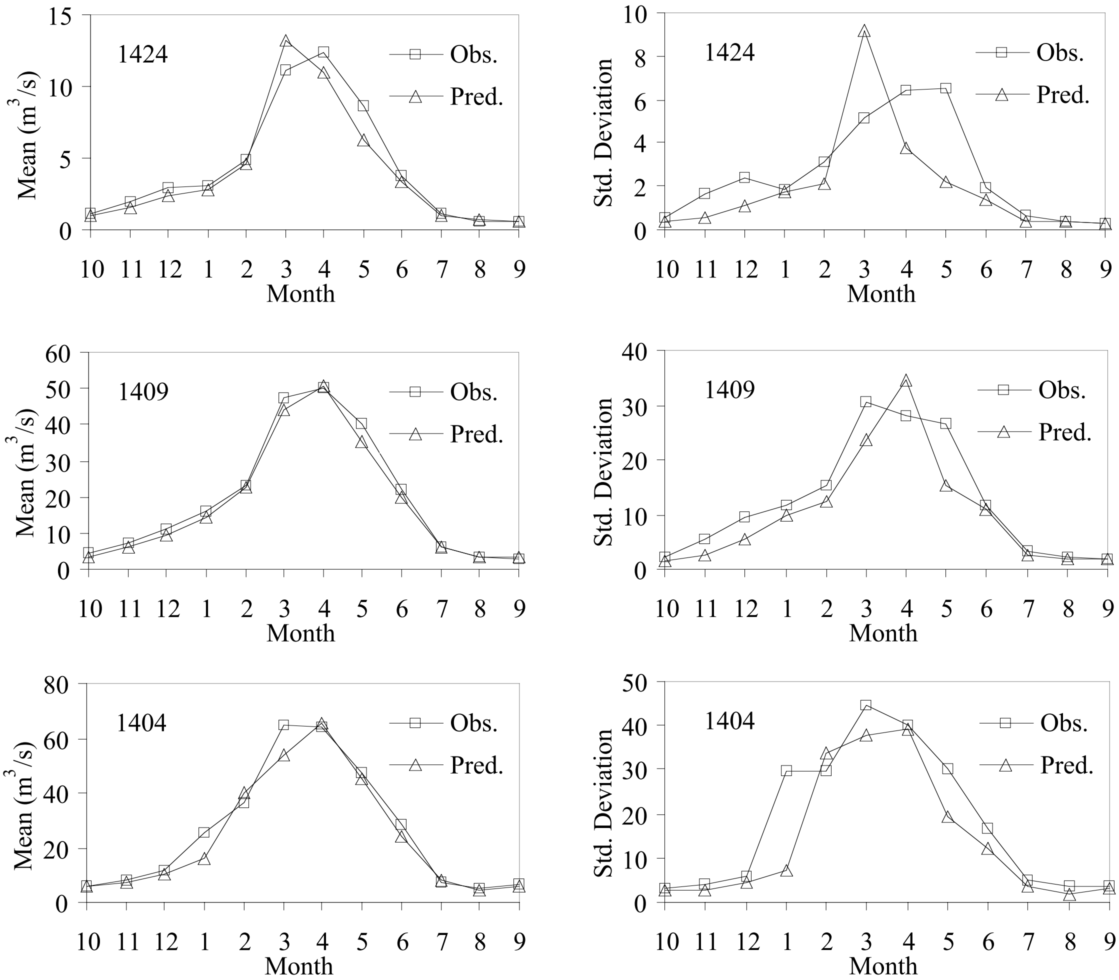


**Figure 5.** Comparison of some statistical properties from observed and predicted data.

**Table 6.** Homoscedasticity test results of monthly streamflow data for each gauging station.

| Gauge Station | G-Q Test | | Decision | B-P Test | | Decision | Spearman's Rho Test | | Decision |
|---|---|---|---|---|---|---|---|---|---|
| | $F_{cal}$ | $F_{Table}$ | | $F_{cal}$ | $F_{Table}$ | | $T_{cal}$ | $T_{Table}$ | |
| 1424 | 1.11 | 1.35 | NC | 1.11 | 1.30 | NC | 1.39 | ±1.96 | NC |
| 1409 | 1.07 | 1.29 | NC | 1.07 | 1.25 | NC | 0.76 | ±1.96 | NC |
| 1404 | 0.95 | 1.57 | NC | 0.96 | 1.49 | NC | 1.74 | ±1.96 | NC |
| NC, no change in variance of residuals. | | | | | | | | | |

**Table 7.** Regression summaries of the relationship between the observed and predicted data.

| Gauge station | Correlation | | R-square | Adjusted R-square | | |
|---|---|---|---|---|---|---|
| | R | $R_{Sig}$ | | | F-value | $F_{Sig}$ |
| 1424 | 0.920 | 0.000 | 0.847 | 0.846 | 1709.8 | 0.000 |
| 1409 | 0.921 | 0.000 | 0.848 | 0.848 | 2402.8 | 0.000 |
| 1404 | 0.895 | 0.000 | 0.801 | 0.800 | 572.7 | 0.000 |
| $R_{Sig}$, significance level concerning coefficient of correlation<br>$F_{Sig}$, significance level for F-value of regression model | | | | | | |

## Conclusion

The estimation of monthly flow is, in practice, often based on small samples of data, which may cause severe uncertainty. Therefore, the main priority is to obtain reliable estimates of monthly flows. Simulation on the basis of samples of historical flow records is needed for reliable information as for many water resources studies the available streamflow records are scarce. The selection of a best model fit to historical data is directly related to whether residual analysis is performed well.

This study is concerned with testing residuals from an ARIMA model, which are the most popular for generating stochastically synthetic data, applied to monthly streamflow data from the Çekerek Stream watershed. Independence analysis of the residuals was examined by using the Ljung-Box Q statistic, and runs and turning point tests. To determine whether the residuals are normally distributed, Kolmogorov-Smirnov, Anderson-Darling and skewness tests were used. The homoscedasticity of residuals was tested by Goldfeld-Quandt, Breusch and Pagan and Spearman's rho approaches. The selected parsimony model for each data set among the ARIMA models fulfilled the diagnostic checks considering the Schwarz Bayesian criterion. All of these tests related to assumptions including independence, normality and homoscedasticity showed that the assumptions concerning the residuals from the selected best ARIMA model held. Furthermore, comparisons of monthly mean and standard deviation values for observed and predicted data from the ARIMA model showed that the predicted data preserved the basic statistical properties of the observed series. The simple linear regression approach was applied to explain the association between the observed and predicted monthly data sequences. The results from the regression analysis support the existence of a statistically significant linear relationship between the observed and predicted data.

## Nomenclature

| | |
|---|---|
| $a_i$ | white noise time series value at time i |
| B | backward shift operator |
| C | constant term in ARIMA model |
| c | constant for Box-Cox transformation |
| d | order of the nonseasonal differencing operator |
| $ESS_L$ | the residual sum of square for the low group |
| $ESS_H$ | the residual sum of square for the high group |
| $k_p$ | degree of freedom |
| n | the number of observation |
| $n_L$ | the number of residuals in the low group |
| $n_H$ | the number of residuals in the high group |
| Q(r) | Ljung-Box statistic at lag m |
| $r_k$(a) | ACF of $a_i$ at lag k |
| s | seasonal length |
| $x_i$ | discrete time series value at time i |
| $w_i$ | stationary series formed by differencing the $x_i$ |
| $z_i$ | transformation of $x_i$series |

## Greek Symbols

| | |
|---|---|
| $\lambda$ | exponent for Box-Cox transformation |
| $\mu$ | mean level of the $w_i$ series (if D + d > 0 often $\mu \approx 0$) |
| Ø(B) | nonseasonal AR operator of order p |
| $\theta$ (B) | nonseasonal MA operator of order q |
| $\Phi$ (B) | seasonal AR parameter of order P |
| $\Theta$ (B) | seasonal MA parameter of order Q |
| $\sigma_a^2$ | residual variance |

## References


Ahmad, S., Khan, I.H. and Parıda, B.P., "Performance of Stochastic Approaches for Forecasting River Water Quality", Water Resources, 35, 4261-4266, 2001.

Bayazıt, M., Statistical Methods in Hydrology, İstanbul Technical University Press, No: 1197, İstanbul, 1981.

Box, G.E.P. and Jenkins, G.M., Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, USA, 1976.

Breusch, T. and Pagan, A., "A Simple Test of Heteroscedasticity and Random Coefficient Variation". Econometrica, 47, 1287-1294, 1979.

Brooks, C., Introductory Econometrics for Finance, Cambridge University Press, UK, 2002.

Chatfield, E., The Analysis of Time Series: An Introduction, Chapman and Hall Ltd., London, 1991.

Chaloulakou, A., Assimacopoulos, D. and Lekkas, T., "Forecasting Daily Maximum Ozone Concentrations in the Athens Basin", Environmental Monitoring and Assessment, 56, 97-112, 1999.

Chhabra, R., Soil Salinity and Water Quality, A.A. Balkema, Inc., Brookfield, VT, USA, 1996.

Cromwell, J.B., Labys, W.C. and Terraza, M., Univariate Tests for Time Series Models, A Sage Publications, Series/number: 07-99, London, 1994.

Devore, J. and Peck, R., Statistics: The Exploration and Analysis of Data, Second Edition, Duxbury Press, California, 1993.

Dickey D.A. and Fuller, W.A., "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", Econometrica, 49, 1057-1072, 1981.

Enders W., Applied Econometric Time Series, John Wiley and Sons, Inc., New York, 1995.

Ferguson T.S., Genest, C. and Hallin, M., "Kendal's Tau for Serial Dependence", The Canadian Journal of Statistics, 28, 587-604, 2000.

Fernando, D.A.K. and Jayawardena, A.W., "Generation and Forecasting of Monsoon Rainfall Data", $20^{th}$ WEDC Conference, Colombo, Sri Lanka, 310-313, 1994.

GDSW, Soils of Yeşilırmak Basin, General Directorate of Soil and Water Publications, No: 241, Ankara, 1970.

Gibbons, J.D., Nonparametric Methods for Quantitative Analysis. American Sciences Press, Inc., Ohio, USA, 1997.

Greene, W.H., Econometric Analysis, Prentice Hall International, Inc., New Jersey, USA, 2000.

Haan, C.T., Statistics Methods in Hydrology, Iowa State Press, Iowa, USA, 1977.

Hipel, K.W., McLeod, A.I. and Lennox, W.C., "Advances in Box-Jenkins Modeling 1. Model Construction", Water Resources Research, 13, 567-575, 1977.

Hipel, K.W. and McLeod, A.I., Time Series Modelling of Water Resources and Environmental Systems. Development in Water Science, vol. 45, Elsevier Scientific Publishing Company, Amsterdam, The Netherlands, 1994.

Janacek, G. and Swift, L., Time Series Forecasting, Simulation, Application, Ellis Horwood, New York, USA, 1993.

Lehmann, A. and Rode, M., "Long-Term Behaviour and Cross-Correlation Water Quality Analysis of the River Elbe, Germany", Water Resources, 35, 2153-2160, 2001.

Ljung, G.M. and Box, G.E.P., "On a Measure of Lack of Fit in Time Series Models", Biometrika, 65, 297-303, 1978.

McKerchar, A.I. and Delleur, J.W., "Application of Seasonal Parametric Linear Stochastic Models to Monthly Flow Data", Water Resources Research, 10, 246-255, 1974.

McKinnon, J.G., Critical Values for Co-integration Tests. UC San Diego Discussion Paper, 90-4, 1990.

McLeod, A.I., Hipel, K.W. and Lennox, W.C., "Advances in Box-Jenkins Modeling 2. Applications", Water Resources Research, 13, 577-586, 1977.

McLeod, A.I., "Diagnostic Checking Periodic Autoregression Models with Application", The Journal of Time Series Analysis, 15, 221-233, 1995.

Salas, J.D., Delleur, J.W., Yevjevich, V. and Lane, W.L., Applied Modeling of Hydrologic Time Series, Water Resources Publications, Michigan, USA, 1980.

Sharma A., Tarboton D.G. and Lall U., "Streamflow Simulation: A Nonparametric Approach", Water Resources Research, 33, 291-308, 1997.

Snedecor, G.W. and Cochran, W.G., Statistical Methods, Iowa State University Press, Iowa, USA, 1989.

Sun, H. and Koch, M., "Case Study: Analysis and Forecasting of Salinity in Apalachicola Bay, Florida, Using Box-Jenkins Arima Models", Journal of Hydraulic Engineering, 127, 718-727, 2001.

Thode, H., Testing for Normality Marcel Dekker, Inc., vol. 164, USA, 2002.

Venema, H.D., Schiller, E.J. and Adamowski, K., "Evidence of Climate Change in the Senegal River Basin", Water Resources Development, 12, 531-546, 1996.

Wei, W.W.S., Time Series Analysis, Addison-Wesley Publishing Company, Inc., New York, USA, 1990.

Zhang, G.P., "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model", Neurocomputing, 50, 159-175, 2003.