

Khassa Chay Stream Flow Forecasting by Markove Autoregressive AR. Model.

Dr. Cheleng A. Arselan

Lecturer

Kirkuk University, College of Engineering

Jasim M . Kadir

Assis. Lecturer

Abstract

In this research a successful analysis of the monthly mean of Khassa Chay stream flow record for the period (1941-2001) was adopted, then the Markove Autoregressive AR model was fitted to the resulted stochastic component of the series, then suitable tests were done to detect the order of the model. Also another decision whether the parameter of this model should be constant or periodic was decided after testing the model. The parameters of this model were found and used to generate the monthly flow for 56 years ahead. It was concluded that Markove AR scheme is adequate to describe the structure of the monthly mean Khassa chay stream flow and the periodic type of this model is more realistic than the constant type because the high sensitivity to errors.

Keywords: Stochastic component , Autoregressive MarkoveAR, Periodic, constant, Model parameters.

التنبؤ بتصارييف نهر خاصة جاي باستخدام النموذج الرياضي Markove Autoregressive AR

جاسم، م. قادر
م. مساعد

د. جلنك، أ. أرسلان
مدرس

الخلاصة

في هذا البحث تم تحليل مقادير المعدلات الشهرية لتصارييف نهر خاصة جاي للفترة الزمنية (1941-2001). تم اجراء التحليل بنجاح ومن ثم تطبيق النموذج الرياضي المعروف Markove AR على السلسلة المستخلصة من التحليل مع اجراء مجموعة من الفحوصات للتحقق من مرتبة النموذج المستخدم بالإضافة الى بعض الفحوصات التي مكنت من تحديد طبيعة معاملات النموذج الرياضي Markov A R من حيث كونه بطبيعة ثابتة او متماوجة . تم اشتقاق المعاملات الرياضية للنموذج الرياضي المستخدم ومن ثم تم توليد سلسلة مستقبلية لفترة مقدارها 56 سنة . تم الاستنتاج في هذا البحث على قابلية النموذج Markove AR لتوصيف المركبة العشوائية للتصارييف الشهرية لنهر خاصة جاي وتم الاستنتاج كذلك من كون النموذج ذو الطبيعة المتماوجة اكثر تحسسا للاخطاء.

الكلمات الدالة: التصارييف الشهرية ، الطبيعة المتماوجة ، الطبيعة الثابتة

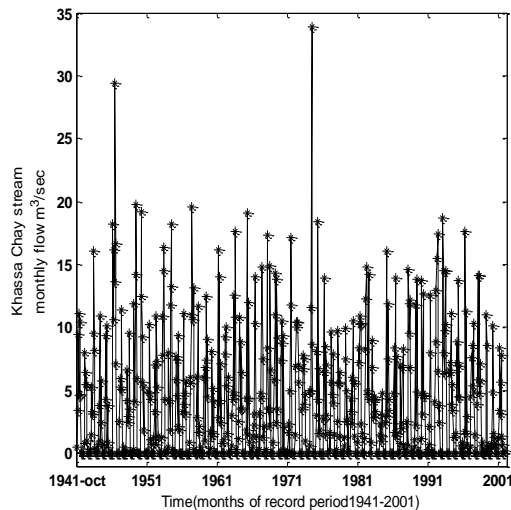
Received: 3 – 6 - 2011

Accepted: 5 – 8 - 2012

1. Introduction

Forecasting of seasonal stream flow provides many benefits' to society, by improving the ability to plan, adapt and changing water supplies. The development of stochastic models of hydrological phenomena play an important role in water resources engineering including their use to forecast river flows [3]. The choice of the right model for a given hydrological series is an important aspect of the modeling. Stochastic analysis and modeling of monthly river flows have attracted the attention of many researchers. Many researchers tend to prefer the analysis of daily stream flow rather than the monthly record. Early attempts of modeling hydrologic events were made by Young and Pisano who proposed a multisite monthly stream flow model, using Matalas algorithm in order to preserve the cross correlation between the stations [9]. Alsuhaily 1985 used stochastic analysis to the daily Tigris flow series at four stations, then the stochastic component was fitted by different models (AR, and ARIMA models) for each single site and also the multi site model (MATALAS) was applied to the mentioned four stations successfully [1]. Ten candidate models of the Auto-Regressive Moving Average (ARMA) family are investigated for representing and forecasting monthly and ten-day streamflow in three Indian rivers by P. P. Mujumdar & D. Nageshkumar 1990, the best models for forecasting and representation of data were selected by using the criteria of Minimum Mean Square Error (MMSE) and Maximum Likelihood (ML) respectively. The selected models were validated for significance of the residual mean, the periodicities in the residuals and of the correlation in the residuals [8]. Kadri Yurkli, Ahmet kuruncu, Fazli ozturk 2004 presented a methodology on modeling monthly data of Kelkit stream in which the watershed is located north Anatolia and the stream is formed together by joining small streams, Box-Jenkins or ARIMA (autoregressive-moving average) were used to simulate monthly data. Diagnostic checks were done for all the models selected from the autocorrelation function (ACF) and partial autocorrelation function (PACF) [10]. A. K. Mishra and V. R. Desai 2005 worked on Kansabati river basin in India, they concluded that due to the random nature of contributing factors, occurrence and severity of droughts can be treated as stochastic in nature. In their study, linear stochastic models known as ARIMA and Multiplicative Seasonal Autoregressive Integrated Moving Average (SARIMA) models were used to forecast droughts based on the procedure of model development [7]. In this research an attempt to model the Khassa Chy river which is a tributary of Zghitun river which is flowing into the existing Adhaim dam reservoir, (this stream is dividing city of Kirkuk into two main sectors) depending on the monthly recorded data for a period (1941-2001) which begins from October. These data were taken from a report on Khassa Chay dam project hydrological study By Engineering Consultancy Bureau of Mustansiriya University -College of Engineering, 2006 [4]. These recorded monthly Khassa Chay stream flows for the mentioned period (1941-2001) were plotted and shown in Figure(1) and the annual monthly mean values are plotted as in Figure(2).

By referring to the Figures 1 and 2 which show the monthly discharge in m^3/sec during the period (1941-2001) it is obvious that the Khassa Chy river is a seasonal stream and by referring to the discharge values for the months (Jun, July, August, and September) and for all years which are equal to zero or nearby zero, this means that this stream is suffering from a drought at these times. This characteristic of Khassa Chay stream requires from any researcher especial treatment in the analysis and modeling the stream. This treatment was done in this research by using many trials in the analysis and by rearranging the steps of the stochastic analysis till reaching the best description of the dependent stochastic component of the stream, and also by using two different trials when finding the model parameters as will be described in the following sections later.



Figure(1) The monthly discharge of Khassa Chay Stream flow . (1941-2001).

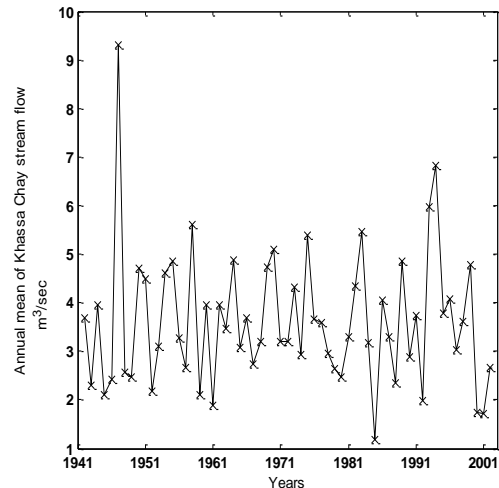


Figure (2) The Annual means of Khassa Chay Stream flow . (1941-2001).

2. Khassa Chay Streamflow Data Analysis.

As was mentioned in the previous section the data which was used in the analysis and modeling process was for the period(1941-2001). This means that the monthly khassa Chay flow was for 61 years (in m³/sec) , 56 years of this data was used for the analysis and the remained 5 years in data which were used for Calibration the model after building it. The Stochastic analysis process was done first by investigating the homogeneity and existence of trend portion in the monthly discharge of Khassa Chay stream for the 56 years, which was done successfully by method suggested by Yevjevich, depending on the T-test and by choosing a suitable level of significance [1]. After doing the process above the data was concluded to be homogenous and there was no error record in the khassa chay series. The series also was detected to be empty from trend component by using parametric and non parametric tests. The next step in the analysis was to investigate the periodicity of the series, which can be defined as a value repeats itself year after year in its seasonal position inside the year. The physical explanation of the presence of this component in seasonal hydrologic time series is that it is attributed to the astronomical events which are also periodic. The investigation of this component was done by measuring what is called as a persistency which is defined as the tendency of low flows to follow low flows or high flows to follow high flows, thus it represents the correlation between the flow at the time t with the flow at the time t+1, t+2,.....t+k which is measured by the serial correlation coefficient r_k:

$$r_k = \frac{\frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - X_{me})(X_{t+k} - X_{me})}{\frac{1}{n} \sum_{t=1}^n (X_t - X_{me})^2} \dots\dots\dots(1)$$

where X_{me} is the mean of the sample of n values of X_t and k is usually taken for values from zero up to 2*w where w represents the number of months [1],[5]. The periodic component of this series was removed using a non parametric method by standerdization.

The series whole mean after this operation completion was=-1.3*10⁻¹⁷ and the whole standard deviation = 0.9918. As it was recommended by many researchers dealing with stochastic

modeling, the analysis operation requires a transforming the series into normally distributed series, this step was done using a Box-Cox transformation method. After this transformation step the Khassa Chay series whole mean became=-6.6546 and the standard deviation = 0.3782. Another step was done to randomize the series by standardizing the series, this also was done by subtracting the whole mean of the series from the transformed data and dividing it to the whole standard deviation. The whole mean of the sires after this step was = 5.1*10⁻¹⁴ and the standard deviation=1. After the previous steps of the Stochastic analysis, the residual part in the series must be uncorrelated and random to reflect the non deterministic stochastic part of the series, this was checked by calculating the serial correlation coefficient of the series and the critical limits depending on the equation below :

$$r_c = \frac{-1 \pm 1.96\sqrt{N - K - 1}}{N - K} \dots\dots\dots(2).$$

Where rc is the critical limits of serial correlation coefficient, N: number of the data, K: the value of the lag or delay[2],[5]. Table (1) represents these values which reflect the success of the analysis that made the series nearly uncorrelated and random.

Table(1) The serial correlation coefficients with their critical limits Of the stochastic part of the khassa chay stream flow

rk	max limit	min limit
0.1247	0.0741	-0.0771
0.0774	0.0742	-0.0772
0.0025	0.0742	-0.0772
0.0379	0.0743	-0.0773
0.075	0.0743	-0.0773
0.0107	0.0744	-0.0774
-0.0617	0.0744	-0.0775
3.82*10- 4	0.0745	-0.0775
-0.0125	0.0746	-0.0776
0.0283	0.0746	-0.0776
-0.0127	0.0747	-0.0777
-0.031	0.0747	-0.0778
-0.0522	0.0748	-0.0778
-0.0496	0.0748	-0.0779
-0.0241	0.0749	-0.0779
-0.0381	0.0749	-0.078
-0.0073	0.075	-0.0781
0.0191	0.0751	-0.0781
-0.0518	0.0751	-0.0782
0.0073	0.0752	-0.0782
-0.0099	0.0752	-0.0783
-0.0052	0.0753	-0.0784
-0.0347	0.0753	-0.0784
-0.0399	0.0754	-0.0785

Now, the next step was to choose a right model to describe the series and then to find the model parameters, this will be explained in the following section.

3. Markove Autoregressive AR.

This model describes the dependent in a hydrologic stochastic series (ε_{p,t}) by assuming that each value is a combined effect of a previous value in addition to an independent stochastic component (ξ_{p,t}), which occurs at the same time of occurrence of (ε_{p,t}). The formulation of this model is given by :

$$\epsilon_{p,t} = \sum_{i=1}^{\infty} a_i \epsilon_{p,t-i} + \sigma_{\xi} \xi_{p,t} \dots\dots\dots(3).$$

Where(a_i,σ) are the parameters of the model . As the effect of the pervious values on the current values decreases with time in all series, a finite sum of the equation above is often sufficient to approximate the dependence ε_{p,t} values . So that an alternative representation of this equation is:

$$\epsilon_{p,t} = \sum_{i=1}^m a_i \epsilon_{p,t-i} + \sigma_{\xi} \xi_{p,t} \dots\dots\dots(4).$$

Where m is the order of the model. In a hydrologic time series the dependence can be well approximated by first, second or third order linear autoregressive model . Many tests are done to find the required order of this model. A modified form of autoregressive model is given by Yevjevich with general form ;

$$\varepsilon_{p,t} = \sum_{k=1}^m \alpha_{k,t} \varepsilon_{p,t-k} + \sigma_{\xi,t} \cdot \dots\dots\dots(5).$$

Where ($\alpha_{k,t}$ and $\sigma_{\xi,t}$) are the parameters of the model, both which may be periodic or non periodic depending on the estimated $r_{k,t}$ values using the equation below :

$$r_{k,t} = \frac{\sum_{p=1}^N (X_{p,t} - \frac{1}{n} \sum_{p=1}^N X_{p,t})(X_{p,t+k} - \frac{1}{n} \sum_{p=1}^N X_{p,t+k})}{\left[\sum_{p=1}^N (X_{p,t} - \frac{1}{n} \sum_{p=1}^N X_{p,t})^2 \right]^{1/2} \left[\sum_{p=1}^N (X_{p,t+k} - \frac{1}{n} \sum_{p=1}^N X_{p,t+k})^2 \right]^{1/2}} \dots\dots\dots(6).$$

where
 rk is lag k serial correlation coefficient
 N : number of annual data or years,
 K lag in units $X_{p,t}$ = value at time t, P is the year .
 If $t = w - k + 1$ then n will be replaced by n-1 and $X_{p,t+k}$ with $X_{p+1,k}$,
 W : number of seasonal positions in year[1],[5],[3].

If these values exhibit periodicity then $\alpha_{k,t}$ and $\sigma_{\xi,t}$ should be periodic, otherwise they should be constant. The estimation of the first and second order Autoregressive model parameters is shown below :

3. 1. Constant Parameters Model.

3. 1. 1. First Order Model.

$$\varepsilon_{p,t} = a_1 \varepsilon_{p,t-1} + \sqrt{1 - a_1^2} \xi_{p,t} \dots\dots\dots(7)$$

Where $a_1 = r_1'$ or $a_1 = r_1$

3. 1. 2. Second Order Model.

$$\varepsilon_{p,t} = a_1 \varepsilon_{p,t-1} + a_2 \varepsilon_{p,t-2} + \sigma_{\xi}^2 \dots\dots\dots(8)$$

Where

$$a_1 = \frac{r_1 - r_1 r_2}{1 - r_1^2} \dots\dots\dots(9).$$

$$a_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \dots\dots\dots(10).$$

$$\sigma = 1 - \sqrt{(a_1^2 + a_2^2 + 2a_1 a_2 r_1)} \dots\dots\dots(11)$$

3. 2. Preriodic Parameters Model

3. 2. 1. First Order Model.

$$\varepsilon_{p,t} = \alpha_{1,t} \varepsilon_{p,t-1} + \sigma_t \xi_{p,t} \dots\dots\dots(12).$$

Where:

$$\alpha_{1,t} = r_{1,t-1}$$

For t=1 , t-1 =w and w=12 in this case.

$$\sigma_t = (1 - \alpha_{1,t})^{1/2} \dots\dots\dots(13).$$

3. 2. 2. Second Order Model.

$$\varepsilon_{p,t} = \alpha_{1,t} \varepsilon_{p,t-1} + \alpha_{2,t} \varepsilon_{p,t-2} + \sigma_t \xi_{p,t} \dots\dots\dots(14)$$

$$\alpha_{1,t} = \frac{r_{1,t-1} - r_{1,t-2} - 2r_{2,t-2}}{1 - r_{1,t-2}^2} \dots\dots\dots(15)$$

$$\alpha_{2,t} = \frac{r_{2,t-1} - r_{1,t-2} - 2r_{2,t-2}}{1 - r_{1,t-2}^2} \dots\dots\dots(16).$$

$$\sigma_t = (1 + \alpha_{1,t}^2 + \alpha_{2,t}^2 - 2\alpha_{1,t}r_{1,t-1} - 2\alpha_{2,t}r_{2,t-2} + 2\alpha_{1,t}r_{1,t-1})^{(1/2)} \dots\dots\dots(17).[1]$$

By using equation(6) to find the lag one , two, and three months serial correlation coefficient , the values of $r_{1,t}$, $r_{2,t}$, $r_{3,t}$ where calculated for the khassa chay stream flow time series after the analysis steps and found as shown in Table (2).

Table(2) The lag 1, 2, 3 months serial correlation coefficients values

$r_{1,t}$	$r_{2,t}$	$r_{3,t}$
0.1553	-0.0441	0.0231
0.0762	0.0817	- .0048
0.175	0.1826	0.1643
-0.2491	-0.0711	-0.0619
0.2057	0.1311	0.1661
0.1932	0.1052	0.068
0.3954	0.0778	-0.0109
0.5057	0.0393	-0.098
-0.0507	-0.0355	-0.0703
-0.0259	0.4325	-0.051
-0.0359	0.2228	-0.0605
0.2919	-0.036	-0.4072
mean values		
0.1364	0.0905	-0.0286

To determine a suitable order of this model a simplified practical method suggested by Yevjevich is used . This method expresses the goodness of fit the model by determination coefficients (D_i , $i=1, 2, 3, 4, \dots$) which represents the percentage of the total variance of $(\varepsilon_{p,t})$ explained by i th order term of an autoregressive equation, while the remaining portion of this variance is explained by the $(\sigma_t \xi_{p,t})$ term. The criteria used is that if a higher order model explains a percent of the variance less than a chosen quantity (ΔD) then this higher order model is neglected, i.e, if the difference between the percentage of the explained variance by j th and i th order terms of the model ($D_j - D_i$) with $j=i+1$ is smaller than (ΔD) then the model order is taken to be i and (ΔD) is usually taken =0.01 , i.e, 1percent of the total variance of $(\varepsilon_{p,t})$ series. It is expected

that the degree of the model does not exceed the third degree, therefore the determination coefficients D_1, D_2, D_3 only were calculated by using the following equations :

$$D_1 = r_1^2 \dots\dots\dots(18).$$

$$D_2 = (r_1^2 + r_2^2 - 2r_1r_2)/(1 - r_1^2) \dots\dots\dots(19).$$

$$D_3 = \frac{(r_1^2 + r_2^2 + r_3^2 + 2r_1^3r_3 + 2r_1^2r_2^2 + 2r_1r_2^2r_3 - 2r_1^2r_2 - 4r_1r_2r_3 - r_1^4 - r_2^4 - r_1^2r_3^2)}{(1 - 2r_1^2 - r_2^2 + 2r_1^2r_2)} \dots\dots\dots(20).$$

And the order of the model was found by using the following steps:

- 1-if $(D_2 - D_1) \leq \Delta D$ and $(D_3 - D_1) \leq 2\Delta D$ then $m=1$
- 2-if $(D_2 - D_1) > \Delta D$ and $(D_3 - D_2) \leq \Delta D$ then $m=2$
- 3-if $(D_2 - D_1) > \Delta D$ and $(D_3 - D_2) > \Delta D$ then $m=3$

Where the values of r_1, r_2, r_3 , are the mean values of $r_{1,t}, r_{2,t}, r_{3,t}$ respectively [1]. The results of the application of this test are shown below:

$r_1 = 0.1364, r_2 = 0.0905, r_3 = -0.0286, D_1 = 0.0186, D_2 = 0.0021, D_3 = 0.0264, (\Delta D) = 0.0100, D_2 - D_1 = -0.0165, D_3 - D_1 = 0.0078, D_3 - D_2 = 0.0243$, then the model order is 1. The results indicated that the model is 1st order and that the first degree autoregressive model fits the series, since higher degree models do not account for an increase of the explained variance of 1 percent or more over that explained by the first order model. Another decision, whether model parameters should be constant or periodic over the year, should be made. Since those parameters depend on $r_{1,t}$ then they should be periodic if $r_{1,t}$ found to be periodic. By comparing the found values of lag1 serial correlation coefficients with the critical limits which can be taken as was recommended by many researchers ,according to Anderson test :

$$S_{u,1} = \frac{-1 \pm 1.96\sqrt{N-3}}{N-2} \dots\dots\dots(21)$$

Where N is the number of the data [2]. See Table(2). The minimum value is -0.077 and the maximum one is 0.0602. Due to above limits the values of $r_{1,t}$ reflect a periodicity therefor the model should be considered as a periodic. The Auto regressive model now is considered as a periodic and from 1st degree.

4. Model Parameters.

The Autoregressive periodic 1st degree model parameters were found from the equations (12) and shown in Table(3).

By using the model parameters, the independent deterministic stochastic part was extracted and the resulted series was analyzed to know its properties. The whole mean of this series was =0.0063, and the standard deviation=1.0535. After this operation a series of normally distributed numbers were generated with the same mean and same standard deviation with length covering the same 56 future years to represent the monthly independent deterministic part of khassa chay stream flow for the future years. The reason of selecting this length was to ensure future forecasting years reaches nearly 50 years ahead after 2001 since the 5 years of the data period (1941-2001) will

Table(3) The markove model parameters of khassa chay stream flow

α	σ
0.2919	0.8415
0.1553	0.9191
0.0762	0.9611
0.175	0.9083
-0.2491	1.1176
0.2057	0.8912
0.1932	0.8982
0.3954	0.7776
0.5057	0.7031
-0.0507	1.025
-0.0259	1.0129
-0.0359	1.0178

be used for verification or calibration of the model. The generated series was used with the resulted parameters above to find the stochastic dependent

component of monthly Khassa Chay flow series for the future years(1997-2052)years, this was done by reversing the operation which were used in extracting the independent part of the series the following steps later were also reversing all the operations before the step of extracting the independent part with taking into consider the sequence of the steps which mean de-standardization the resulted series by using the same values of the mean and standard deviation which were used in the standardization then the power transformation method was reversed to return the series un transformed, the periodicity was also returned by using the same values of monthly means and monthly standard deviations. Many statistical tests were done to the generated series using the 1st 5 years of the generated series as calibration period. Table(4) shows the comparison between the forecasted khassa chay stream

Table(4) the statistical properties comparison between the forecasted and recorded khassa chay stream flow

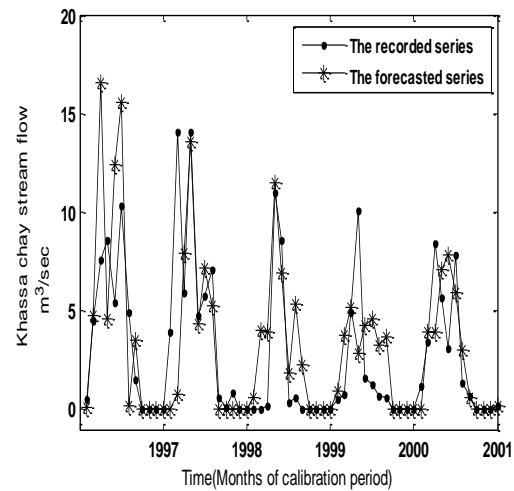
Statistical Results	Oct	Nov	Dec	Jan	Feb	Mar
Mean(for)	1.1071	4.2481	6.4685	6.8121	7.0941	9.1046
Mean(rec)	1.2200	4.5600	5.4000	7.3639	7.480	8.0180
std(for)	1.1593	2.9657	2.8796	4.2454	3.8964	6.4299
std(rec)	1.5579	5.6412	3.8802	4.503	2.6401	6.4534
min(for)	0	0	2.3321	1.5038	1.3331	0.0888
min(rec)	0	0	0.2	5.7	1.6	0.4
max(for)	5.0689	12.0573	16.5908	19.4354	22.4978	33.8140
max(rec)	3.9	14.1	8.4	14.1000	8.6000	10.3000
Cs(for)	1.3238	0.5325	1.3379	1.1221	1.6124	1.5812
Cs(rec)	1.2335	1.1190	0.6569	0.5534	0.3097	1.8436
Ck(for)	4.5020	2.7765	4.7672	3.7851	6.4895	6.0986
Ck(rec)	2.8829	2.7517	3.2274	2.1450	2.007	7.242
StatisticalResults	Apr	May	Jun	Jul	Aug	Sep
Mean(for)	5.9193	2.6879	0.0103	0.0085	0	0.0584
Mean(rec)	5.5197	1.9951	0.0148	0.0180	0	0.0410
std(for)	4.5421	3.2107	0.0246	0.0122	0.0116	0.0944
std(rec)	4.44	3.382	0.0447	0.0115	0	0.0894
min(for)	0.2067	0	0	0	0	0
min(rec)	0.6	0	0	0	0	0
max(for)	21.8424	16.1183	0.1134	0.0575	0.0756	0.4030
max(rec)	7.1000	1.5000	0.1	0.9	0	0.2
Cs(for)	1.4664	1.8537	2.6994	1.7271	4.2597	2.0315
Cs(rec)	0.8568	2.7016	4.3053	7.3560	7.6169	4.6635
Ck(for)	4.8679	7.2889	9.8365	6.1840	25.0226	6.5862
Ck(rec)	3.0677	10.806	21.1245	56.2423	59.0167	24.006

*(for) means the statistical property for the forecasted series(just for calibration years)..

** (rec) means the statistical property for the recorded data (just for calibration years).

flow and the pervious recorded data by comparing the (mean, standard deviation, coefficient of sekweness and coefficient of kurtosis) while the results of the statistical tests on the generated series and the recorded data for the 5 years which are the calibration years were as: t-value for the mean value=-0.4866;t-value for the standard deviation value =-0.0356, while the critical limit =2.306 at 95% significance level,and the found F-value=1.125 while the critical value is=6.39.

In addition to these tests the same t, f tests were done to the monthly statistical parameters (monthly mean and monthly standard deviation, monthly coefficient of skwness, monthly coefficient of kurtosis)which were found. Most of these tests indicated to the success of the fitting the Markove autoregressive model to Khassa chay stream flow series, seeTable(8). Figure (3)shows the approximate conformity between the forecasted and the recorded data of Khassa chay stream flow for 1st 5 years in most positions .



Figure(3) The conformity between the recorded stream flow for 5 years & forecasted

5. Seperating Khassa Chay Stream Series (The 2nd Approach) :

Because of the sensitivity of khassa chay stream due to the drought periods there was another trial which was done to model the khassa chay stream by the same autoregressive model Markove AR. But after separating the Khassa Chay stream data into wet periods and dry periods then the same analysis was done to the wet months in each year (oct, nov, dec, jan, feb, mar, apr, may) for the period of 56 years. The same stochastic analysis steps were done to the selected wet months as was resulted in the wet khassa chay months flow series was composed from a periodic component in addition to the stochastic non deterministic part The periodic part was removed by the same manner (non parametric), and the stochastic non deterministic part (dependent component) was analyzed after the operation of the normalization and the standardization, The resulted dependent stochastic residual was analyzed to find the Markov model parameter. The results which were depended on the same equations above to find the order of the model were: $r_1 = 0.1415$, $r_2 = 0.0246$, $r_3 = 0.0255$, $D_1 = 0.0200$, $D_2 = 0.0139$, $D_3 = 0.0205$, $(\Delta D) = 0.0100$, $D_2 - D_1 = -0.0061$, $D_3 - D_1 = 4.8759 \times 10^{-4}$, $D_3 - D_2 = 0.0066$.The model according to the above parameters was 1st order and due to results of $r_{1,t}$ was periodic see Table(5) .

Table(5) The lag 1,2,3 months serial correlation Coefficients (wet months)

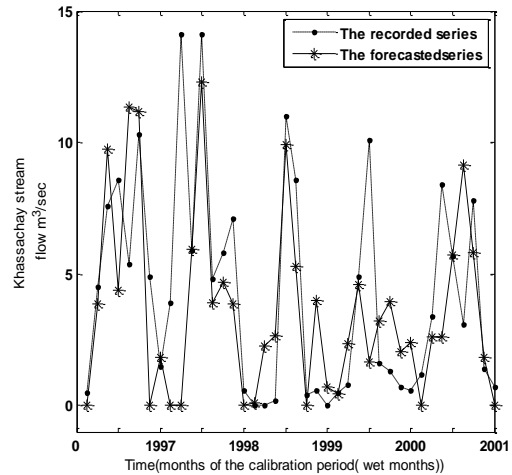
	$r_{1,t}$	$r_{2,t}$	$r_{3,t}$
1	0.1507	-0.048	0.0227
2	0.0724	0.0914	-0.0028
3	0.1769	0.1752	0.1633
4	-0.2522	-0.0644	-0.0521
5	0.2157	0.1228	0.1645
6	0.2028	0.1093	-0.0197
7	0.4015	-0.2347	-0.0293
8	0.1641	0.0454	-0.0426
mean	0.1415	0.0246	0.0255

The parameters of the model was found as shown in Table(6).

Table(6) Markove model parameters (for wet months)

α	0.1641	0.1507	0.0724	0.1769	-0.2522	0.2157	0.2028	0.4015
σ	0.9143	0.9216	0.9631	0.9072	1.119	0.8856	0.8929	0.7736

The same forecasting operation steps were done to forecast the khassaChay stream flow just for the wet months. As was done in the pervious trial many statistical tests were done to check the wet months model, the results were as explained below: t-value=0.6733 for mean, t-value for standard deviation =0.1293, F-value=1.3419, which were a good indications of the success of the model and also the tests were done to the monthly parameters values and indicated to the same conclusion of model success, see Table (8). Figure (4) is also an other indication of the conformity of the forecasted and recorded two series.



Figure(4) The conformity between the forecasted & recorded stream flow for 5 year (just wet months)

Table(7) shows a comparison of the statistical parameters between the forecasted and recorded two series for calibration periods.

Table(7) the statistical properties comparison between the forecasted and recorded khassa chay stream flow for wet months .

Statistical Results	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Mean(for)	0.1146	2.2125	5.1111	6.7973	6.5836	5.1244	2.3437	0.9865
Mean(rec)	1.2200	4.5600	5.4000	7.3639	7.480	8.0180	5.5197	1.9951
std(for)	0.1991	1.3939	2.9553	4.2809	3.5218	4.0305	1.6524	1.0863
std(rec)	1.5579	5.6412	3.8802	4.503	2.6401	6.4534	4.44	3.382
min(for)	0	0	2.5933	1.6605	3.2149	0	0	0
min(rec)	0	0	0.2	5.7	1.6	0.4	0.6	0
max(for)	0.4599	3.8560	9.7643	12.3006	11.3617	11.1770	3.9826	2.4013
max(rec)	3.9	14.1	8.4	14.1000	8.6000	10.3000	7.1000	1.5000
Cs(for)	1.2976	-0.6530	0.7651	0.1652	0.4239	0.3659	-0.3100	0.3209
Cs(rec)	2.9056	1.3329	0.6569	0.5534	0.3097	1.8436	0.8568	2.7016
Ck(for)	2.9280	2.5897	2.2335	1.6210	1.5190	2.4085	1.8062	1.4343
Ck(rec)	2.8829	2.7517	3.2274	2.1450	2.007	7.242	3.0677	10.806

*(for) means the statistical property for the forecasted series . (just for calibration years).

** (rec) means the statistical property for the recorded data (just for calibration years).

6. Conclusions.

Many conclusions were found during the completion of this study, some of them are listed below :

1. The treatment of a seasonal streams that are suffering from drought periods require especial arrangement in the stochastic analysis steps which were done with Khassa Chay stream flow many times till reaching the most successful analysis .
2. The Markov AR scheme is found to be adequate to describe the structure of the stochastic component of the monthly mean Khassa chay stream flow and it was too suitable model to describe the Khassa Chay stream flow behavior.
3. The periodic type of this model is more realistic than the constant type because the model parameters are highly sensitive to errors in the estimates of the serial correlation coefficients, which was also clear from the results of lag one month serial correlation coefficients values for both models .
4. Dealing with wet times separately can be more adequate, this is clear by referring to the figures (3,4), since it is obvious from the conformity of the wet months model more than the conformity in the comprehensive model, although the monthly mean values were described in the comprehensive model better than the wet model , see Table(8).

Table(8) The Statistical tests results on the two models

Model(1)	t-value	t-value	f-value
Comprehensive	mean	std	
monthly mean	-0.1164	-0.094	1.0798
Monthly std	0.2868	0.1917	1.1749
Monthly Cs	1.2108	2.09	7.8479
monthly Ck	1.4244	2.3932	12.533
Model(2) separated			
monthly mean	1.1287	-0.0375	-0.0375
Monthly std	2.0302	0.0863	0.0863
Monthly Cs	2.5012	0.9562	0.9562
monthly Ck	1.8385	2.2925	2.2925

Std: the standard deviation

Cs: coefficient of skewness

Ck :coefficient of kurtosis

References.

- [1] Al-suhaily, A. H, (1985), "Stochastic Analysis of Daily Stream Flow of Tigris River" M Sc Thesis submitted to College of Engineering , University of Baghdad.
- [2] Anderson, O. F ,(1976), "Time Series Analysis and Forecasting", The Box-Jenkins approach London, Butterworths, 311-327.
- [3] Bread, L. R ., (1967), "Simiulation of Hydrologic in Water Yeilds Analysis" The international Hydrology Symposium, Fortcollins, Colrado U. S. A, ASCE, Vol 19, N. 6 , 33-42.
- [4] Engineering consultancy bureau, (2006)"Khassa Chi Dam "Preliminary Report Final – Mustansiriya University-College of Engineering Baghdad. Iraq.
- [5] Green , N. M. D, 1973, "A Synthetic Model for Daily Stream Flow". Journal of Hydrology, Vol. 20, No. 8, 538-542.

- [6] Kurunc, A. ,Yurekli, K, Cevek (2005) "Performance of Two Stochastic Approaches for Forecasting Water Quality and Stream Flow Data From Yisil Irmak, River, Turkey, "Enviromental modeling and software 20(2005), No.7 1195-1200 available /www.sciencedirect.com.
- [7] Mishra, A. K and Desai, V. A (2005), "Drought Forecasting Using Stochastic Models", Journal of Stochastic Environmental Research and Risk Assessment Vol .19, No.5, 57-72.
- [8] Mujumdar, P. P, Nageshkumar, D. N(1990), "Stochastic Models of Stream Flow, Some Case Studies", Hydrological sciences –journal-des sciences Hydrologiques vol 35, No.4, 395-410.
- [9] Young, G. KandPisano. G., (1968), "Operational Hydrology Using Residuals" , Journal of Hyd., Div., ASCE, Vol. 94, No. HY4,
- [10] Yurekli, K, Kurunc, A, Ozturk. F, (2005) "Testing The Residuals of ARIMA Models on The Cekerik Stream watershed In Turkey". Turkish .j. Eng. Env. Sin 29(2005), Tubikat, No. HF3, 61-74

The work was carried out at the college of Engineering. University of Kirkuk