A Multi-variables Multi -sites Model for Forecasting Hydrological Data Series

Rafa H. Al-Suhili
Prof. ,Civil Engineering Dept., College of Engineering, University of Baghdad, Iraq.
A visiting Professor to the City College of New York, New York, USA.
Email: rafiishaker@yahoo.co.in

Nawbahar F. Mustafa
Lecturer, Dams and Water resources Eng., College of Engineering, University of Sulaimania.

ABSTRACT

A multivariate multisite hydrological data forecasting model was derived and checked using a case study. The philosophy is to use simultaneously the cross-variable correlations, cross-site correlations and the time lag correlations. The case study is of two variables, three sites, the variables are the monthly rainfall and evaporation; the sites are Sulaimania, Dokan, and Darbandikhan.. The model form is similar to the first order auto regressive model, but in matrices form. A matrix for the different relative correlations mentioned above and another for their relative residuals were derived and used as the model parameters. A mathematical filter was used for both matrices to obtain the elements. The application of this model indicates it's capability of preserving the statistical characteristics of the observed series. The preservation was checked by using (t-test) and (F-test) for the monthly means and variances which gives 98.6% success for means and 81% success for variances. Moreover for the same data two well-known models were used for the sake of comparison with the developed model. The single-site single-variable auto regressive first order and the multi-variable single-site models. The results of the three models were compared using (Akike test) which indicates that the developed model is more successful ,since it gave minimum (AIC) value for Sulaimania rainfall, Darbandikhan rainfall, and Darbandikhan evaporation, while Matalas model gave minimum (AIC) value for Sulaimania evaporation and Dokan rainfall, and Markov AR (1) model gave minimum (AIC) value for only Dokan evaporation).However, for these last cases the (AIC) given by the developed model is slightly greater than the minimum corresponding value.

Key words: forecasting, multi-sites, multi-variables, cross sites correlation, serial correlation, cross variables correlations, hydrology.

الخلاصة

تم اشتقاق نموذج تنبأ بالبيانات الهيدرولوجية لمتغيرات مختلفة وفي مواقع متعددة وتحقيقه باستخدام حالة دراسية. تعتمد فلسفة النموذج على الاستخدام الاني لمعاملات الارتباط المكانية وتلك التي توجد بين المتغيرات في الموقع الواحد بالإضافة الى الارتباط التسلسلي الزمني. الحالة الدراسية هي لمتغيرين في ثلاثة مواقع, المطر والتبخر في السليمانية, دوكان و دربندخان. ان النموذج شبيه بنموذج الارتباط التسلسلي ولكن معاملاته بصيغة المصفوفات . للنموذج مصفوفتي معاملات الاولى ذات عناصر تمثل معاملات الارتباطات النسبية والثانية تمثل معاملات بقايا الارتباط النسبية. بينت النتائج قدرة النموذج على التنبؤ بالمعلومات بصورة صحيحة حيث تم استخدام اختباري فحص الفرق بالأوساط الحسابية والتباين, وكانت نسب النجاح (81,98) على التوالي. ولغرض المقارنة بين النموذج المشتق والنماذج المعروفة في ادبيات الموضوع, تم بناء نموذج ذو المتغير الواحد لكل متغير من المتغيرات المستخدمة(ستة نماذج)و ثلاث نماذج من نوع النماذج المتعددة المتغيرات نموذج لكل

موقع . تم مقارنة نتائج هذه النماذج مع النموذج المشتق باستخدام اختبار (اكايكي) الذي يستخدم لهذا الغرض. بينت النتائج بان النموذج اعطى اقل القيم للاختبار بالنسبة للمطر في السليمانية و دربندخان والتبخر لدوكان اما فيما يخص نتائج بقية المتغيرات كانت قيم الاختبار اعلى بقليل عن القيم الصغرى المناظرة.

1. INTRODUCTION

Weather generation models have been used successfully for a wide array of applications. They became increasingly used in various research topics, including more recently, climate change studies. They can generate series of climatic data with the same statistical properties as the observed ones. Furthermore, weather generators are able to produce series for any length of time. This allows developing various applications linked to extreme events, such as flood analyses, and draught analysis, and hence putting proper long term water resources management to face the expected draught or flood events. There exist in the literature many types of stochastic models that simulate weather data required for various water resources applications in hydrology, agriculture, ecosystem, climate change studies and long term water resource management.

Single site models of weather generators are used for forecasting a hydrological variable at a single site independent of the same variable at the near sites, and thus ignoring the spatial dependence exhibited by the observed data. On the other hand single variable forecasting models are used for forecasting a hydrological variable in a site independent of the other related variables at the same site, thus ignoring the cross variables relations that physically exist between these variables. **Tobler, 1970,** mentioned in the first law of geography that "everything is related to everything else, but near things are more related than distant things." The most commonly used multi-sites stochastic weather models are of the form proposed by **Richardson, 1981**. for daily precipitation, maximum temperature, minimum temperature, and solar radiation , **Wilks, 1999.**These models forecast a hydrological variable at multiple sites simultaneously, hence simulate the cross sites dependency between these sites. The Multi-variables models are similar to the multi-sites model but simulate the cross variables dependency that exists between some variables at a certain site. The two models forms are similar but using cross sites correlations in the first one , while the second one uses the cross variables correlations. Much progress had been made principally in the last 20 years to come up with theoretical frameworks for spatial analysis **Khalili , 2007.**Some models, such as space–time models have been developed to regionalize the weather generators. In these models, the precipitation is linked to the atmospheric circulation patterns using conditional distributions and conditional spatial covariance functions **Lee et al., 2010.** The multi-site weather generators presented above are designed using relevant statistic information. Most of these models are either complicated or some are applicable with a certain conditions. In real situation both cross variables and cross sites correlation may exist between different hydrological variables at different sites. There exist in the literature some relatively recent some trials to account for the spatial variation in multi-sites. **Calder, 2007,** had proposed a Bayesian dynamic factor process convolution model for multivariate spatial temporal processes and illustrated the utility of the approach in modeling large air quality monitoring data. The underlying latent components are constructed by convolving temporally-evolving processes defined on a grid covering the spatial domain and include both trend and cyclical components. As a result, by summarizing the factors on a regular spatial grid, the variation in information about the pollutant levels over space can be explored. **Al-Suhili et al., 2010,** had presented a multisite multivariate model for forecasting different water demand typesat different areas in the city of Karkouk, north Iraq. This model first relate the each demand type with explanatory variables that affect its type, using regression models, then obtaining the residual series of each variable at each site. These residual are then modeled using multisite Matalas models for each type of demand. These models were coupled with the regression equation to form the multisite multivariate variation. The last two cited research are those among the little work done on forecasting models of multi-sites multivariate types. However these model are rather complicated, and/or do not model the process of cross site and cross variables correlation simultaneously, which as mentioned above the real physical case that may exist. Hence researches are further required to develop a simplified multisite multivariate model. In this research a new straightforward multisite-multivariate approach is proposed to develop such a model that describe the cross variables and cross sites correlation structure in the forecasting of multi variables at multi sites simultaneously. This model was applied to a case study of monthly data of two hydrological variables, rainfall and evaporation at three sites located north Iraq, Sulaimania, Dokan, and Darbandikhan.

2. THE MODEL DEVELOPMENT

The multivariate multisite model developed herein, utilizes single variable lag correlations, cross variables lag-correlations, and cross sites correlations.

In order to illustrate the model derivation consider **Fig.1** shown. This figure illustrates the concept of two variables, two sites and first order model. This simple form is used to simplify the derivation of the model. However, then the model could be easily generalized using the same concept. For instant, **Fig. 2** is a schematic diagram for a multivariate multisite model of two variables, three sites and first order time. The concept is that if there will be two-variables, two sites, and one time step (first order), then there will exist (8) nodal points. Four of these represent the known variable, i.e. values at time (t-1); the other four are the dependent variables, i.e. the values at time (t). As mentioned before **Fig. 1** shows a schematic representation of the developed multisite multivariate model and will be abbreviated hereafter as MVMS (V, S ,O),where V: stands for number of variables in each site , S: number of sites , and O : time order, hence figure (1) can be designated as MSMV (2,2,1), while **Fig. 2** MVMS (2,3,1).

This model can be extended further to (v-variables) and / or (s-sites) and / or (o- time) orders as will be shown later .The model concept assume that each variable dependent stochastic component at time t can be expressed as a function of the independent stochastic component for all other variables at time (t), and those dependent component for all variables at time (t-1) at all sites. The expression is weighted by serial correlation coefficient, cross-site cross-correlation coefficient, cross-variable cross coefficient and cross-site, cross-variable correlation coefficient. In addition to that; the independent stochastic components are weighted by the residuals of all types of correlations. These residual correlations are expressed using the same concept of autoregressive first order model (Markov chain). Further modification of this model is to use relative correlation matrix parameters by using correlation values relative to total sum of correlation for each variable, and the total sum of residuals as a mathematical filter ,as will be shown later.

A model matrix equation for first order time lag, O=1, number of variables=V, and number of sites=S, could be put in the following form:

$$[\epsilon_t]_{v*s,1} = [\rho]_{v*s,v*s} * [\epsilon_{t-1}]_{v*s,1} + [\sigma]_{v*s,v*s} * [\xi_t]_{v*s,1} \quad (1)$$

Which for v=2,s=3,and =1

$$[\epsilon_t]_{6,1} = [\rho]_{6,6} * [\epsilon_{t-1}]_{6,1} + [\sigma]_{6,6} * [\xi_t]_{6,1} \quad (2)$$

Where :

$$\begin{pmatrix} \epsilon_{(v1,s1)} \\ \epsilon_{(v2,s1)} \\ --- \\ \epsilon_{(v1,s2)} \\ \epsilon_{(v2,s2)} \\ --- \\ \epsilon_{(v1,s3)} \\ \epsilon_{(v2,s3)} \end{pmatrix}_t = [\epsilon_t]_{6,1} \qquad (3)$$

$$\begin{pmatrix} \epsilon_{(v1,s1)} \\ \epsilon_{(v2,s1)} \\ --- \\ \epsilon_{(v1,s2)} \\ \epsilon_{(v2,s2)} \\ --- \\ \epsilon_{(v1,s3)} \\ \epsilon_{(v2,s3)} \end{pmatrix}_{t-1} = [\epsilon_{t-1}]_{6,1} \qquad (4)$$

$$\begin{pmatrix} \xi_{(v1,s1)} \\ \xi\epsilon_{(v2,s1)} \\ --- \\ \xi_{(v1,s2)} \\ \xi\epsilon_{(v2,s2)} \\ --- \\ \xi_{(v1,s3)} \\ \xi\epsilon_{(v2,s3)} \end{pmatrix}_t = [\xi_t]_{6,1} \qquad (5)$$

$$\begin{pmatrix} \rho_{1,1} & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} & \rho_{1,5} & \rho_{1,6} \\ \rho_{2,1} & \rho_{2,2} & \rho_{2,3} & \rho_{2,4} & \rho_{2,5} & \rho_{2,6} \\ \rho_{3,1} & \rho_{3,2} & \rho_{3,3} & \rho_{3,4} & \rho_{3,5} & \rho_{3,6} \\ \rho_{4,1} & \rho_{4,2} & \rho_{4,3} & \rho_{4,4} & \rho_{4,5} & \rho_{4,6} \\ \rho_{5,1} & \rho_{5,2} & \rho_{5,3} & \rho_{5,4} & \rho_{5,5} & \rho_{5,6} \\ \rho_{6,1} & \rho_{6,2} & \rho_{6,3} & \rho_{6,4} & \rho_{6,5} & \rho_{6,6} \end{pmatrix} = [\rho]_{6,6} \qquad (6)$$

$$\begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} & \sigma_{1,6} \\ \sigma_{2,1} & \sigma_{2,2} & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} & \sigma_{2,6} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} & \sigma_{3,4} & \sigma_{3,5} & \sigma_{3,6} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_{4,4} & \sigma_{4,5} & \sigma_{4,6} \\ \sigma_{5,1} & \sigma_{5,2} & \sigma_{5,3} & \sigma_{5,4} & \sigma_{5,5} & \sigma_{5,6} \\ \sigma_{6,1} & \sigma_{6,2} & \sigma_{6,3} & \sigma_{6,4} & \sigma_{6,5} & \sigma_{6,6} \end{pmatrix} = [\sigma]_{6,6} \qquad (7)$$

where:
$\rho_{1,1}$ = ρ [(x1, x1), (s1, s1), (t, t-1)]= population serial correlation coefficient of variable 1 with itself at site 1 at site 1, for time lagged 1

$\rho_{1,2}$= ρ [(x1, x2), (s1, s1), (t, t-1)]= population cross correlation coefficient of variable 1 at site 1 with variable 2 at site 1, for time lagged 1

$\rho_{1,3}$= ρ [(x1, x1), (s1, s2), (t, t-1)]= population cross correlation coefficient of variable 1 at site 1 with variable 1 at site 2, for time lagged 1

$\rho_{1,4}$= ρ [(x1, x2), (s1, s2), (t, t-1)]= population cross correlation coefficient of variable 1 at site 1 with variable 2 at site 2, for time lagged 1

$\rho_{1,5}$= ρ [(x1, x1), (s1, s3), (t, t-1)]= population cross correlation coefficient of variable 1 at site 1 with variable 1 at site 3, for time lagged 1

$\rho_{1,6}$= ρ [(x1, x2), (s1, s3),(t,t-1)]= population cross correlation coefficient of variable 1 at site 1 with variable 2 at site 3, for time lagged 1,the definition continues… , finally

$\rho_{6,6}$= ρ [(x2, x2), (s3, s3), (t, t-1)]= population serial correlation coefficient of variable 2 at site 3 with variable 2 at site 3, for time lagged 1.

The designated (ρ i,j) is used for simplifying .That is variables at site 1 ,as 1, and 2,for this model (in general to 1,2,…v),then for variables at site 2,as 3 ,and 4 (in general from v+1 to 2v and so on) hence (r1,v+1) stands for the correlation between variable 1 at site 1,and variable 1 at site 2 and so on.
ϵ: is the Stochastic dependent component.

ξ: is the Stochastic independent component.

$\sigma_{i,j}$: are the residual of the correlation coefficient $\rho_{i,j}$.

The matrix, Eq. (2) can be written for each term, for example, for the first term:

$$\epsilon_{(1,s1,t)} = \rho_{1,1} * \epsilon_{(1,s1,t-1)} + \rho_{1,2} * \epsilon_{(2,s1,t-1)} + \rho_{1,3} * \epsilon_{(1,s2,t-1)} + \rho_{1,4} * \epsilon_{(2,s2,t-1)} + \rho_{1,5} * \epsilon_{(1,s3,t-1)} + \rho_{1,6} * \epsilon_{(2,s3,t-1)} + \sigma_{1,1} * \xi_{(1,s1,t)} + \sigma_{1,2} * \xi_{(2,s1,t)} + \sigma_{1,3} * \xi_{(1,s2,t)} + \sigma_{1,4} * \xi_{(2,s2,t)} + \sigma_{1,5} * \xi_{(1,s3,t)} + \sigma_{1,6} * \xi_{(2,s3,t)} \quad (8)$$

Similar equations could be written for the other variables. The correlation coefficient in each equation is filtered by a division summation filter, as in the following equation:

$$\rho r_{\mathbf{i,j}} = \frac{\rho_{i,j}}{\sum_{j=1}^{n=v*s} abs\rho_{i,j}}. \quad (9)$$

Where $\rho r_{\mathbf{i,j}}$ is the relative correlation coefficient of row i and column j of the matrix in eq.(6). The corresponding σ values are estimated using the following equation:

$$\sigma_{i,j} = \sqrt{1 - \rho_{i,j}^2} \tag{10}$$

Then these $\sigma_{i,j}$ are also filtered using equation similar to eq.(9) as follows:

$$\sigma r_{\mathbf{i,j}} = \frac{\sigma_{i,j}}{\sum_{j=1}^{n=v*s} abs\sigma_{i,j}} \tag{11}$$

Then the model matrix equation is the same as that appear in Eq.(2), replacing $\rho_{\mathbf{i,j}}$ values by the corresponding relative values $\rho r_{\mathbf{i,j}}$in the matrix Eq.(6), and $\sigma_{i,j}$ with the corresponding relative values $\sigma r_{i,j}$ in the matrix Eq.(7) . The model can be generalized to any number of variables and number of sites.

3. THE CASE STUDY AND APPLICATION OF THE MODEL.

In order to apply the new developed (MVMS) model explained above the Sulaimania Governorate was selected as a case study. Sulaimania Governorate is located north of Iraq with total area of (17,023 km2) and **population, 2009.** 1,350,000. The city of Sulaimania is located (198) km north east from Kurdistan Regional capital (Erbil) and (385) km north from the Federal Iraqi capital (Baghdad). It is located between (33/43- 20/46) longitudinal parallels, eastwards and 31/36-32/44 latitudinal parallels, westwards. Sulaimania is surrounded by the Azmar Range, Goizja Range and the Qaiwan Range from the north east, Baranan Mountain from the south and the Tasluje Hills from the west. The area has a semi-arid climate with very hot and dry summers and very cold winters. **Barzanji, 2003**.

The variables used in the model among other meteorological recoded data are (rainfall and evaporation) for monthly model as a two main variables that are expected to be useful for catchment management and runoff calculation. Data were taken from three meteorological stations (sites) inside and around Sulaimania city, which are Sulimania, Dokan dam, and Darbandikhan dam meteorological stations. Dokan dam metrological station is located (61 km) north east, and Darbandikhan dam metrological station is located (55 km) south east of Sulaimania city. While Dokan dam meteorological station is located (114 km) north east of Darbandikhan dam metrological station .The sites coordinates are given in **Table 1**, **Barzinji ,2003**.The Satellite image of the locations of the three stations showed in **Fig.3.**

The model was applied to the data of the case study described above. The length of record for the two variables and the three stations is (27) years, (1984-2010). The data for the first (22), (1984-2005) years were used for model building, while the left last 5 years data were used for verification,(2006-2010). It is worth to mention that the data are on monthly basis. Moreover since the analysis includes the rainfall as a variable which has zero values for June, July, August and September, in the selected area of the case study, these months are excluded from the analysis. Hence the model was built for the continuous period from October to May.

In order to give a general view for the data used the descriptive statistics (Mean, Standard deviation Sd, Coefficient of Skewness Cs, Coefficient of kurtosis Ck, Maximum Max, Minimum Min) were calculated for rainfall and evaporation of Sulaimania, Dokan dam, and Darbandikhan dam meteorological stations and are shown in **Table 2.**

Before proceeding with the modeling process the data series should be checked for their homogeneity . The split sample test suggested by **Yevjevich, 1972,** was applied for this purpose for each data series to test the homogeneity both in mean and standard deviation values. Different sizes of the subsamples were used for dividing the data sample into two subsamples with (n1,and n2) as number of years for subsample one and subsample 2 respectively. That is (n1:n2) as (1,26),(2,25),(3,24), and so on. The split sample test result on estimated t-values that was compared with the critical t-value. If the t-value estimated is greater than the critical t-value then the data series is considered as non-homogeneous, **Yeijevich, 1972 ,** and thus this non-homogeneity should be removed. The results of this test had showed that there are some different subsamples splitting (n1:n2) values that exhibit non- homogeneity exist, however these cases that gives the maximum t-test values were considered for each of the 6- data series.**Table.3** shows these results, which indicates that non-homogeneity is exist in Sulaimania evaporation, Dokan rainfall, and Derbendikhan evaporation data series, while the series of the other variables are homogeneous. To remove this non-homogeneity the method suggested by **Yeijevich,1972** was used that using the following equation:

$$H_{i,j} = Mean2 + \frac{X_{i,j} - (A1 - B1 * i)}{A2 - B2 * i} * Sd2 \qquad (12)$$

Where,
$H_{i,j}$: is the homogenized series at year i,month j of the first sub-sample (old).
$X_{i,j}$: is the original series at year i, month j, of the first sub-sample .
A1, B1: are the linear regression coefficients of the annual means.
A2,B2 : are the linear regression coefficients of the annual standard Deviations.
Mean2,Sd2 : are the overall mean and standard deviation of the second sub-sample.
This implies that the data is normalized according to the second sub-sample, i.e., the most recent one which is the correct way for forecasting. **Table.4** shows the values of the of Mean2,Sd2,A1,B1,A2,and B2, for the three non-homogeneous series.
The homogenized data were then retested to make sure that the transformation applied in Eq.(12), had removed the non-homogeneity. **Table.5** shows these results which ensure that the data series are all now homogeneous.
The next step in the modeling process is to check and remove the trend component in the data if exist. This was done by finding the linear correlation coefficient(r) of the annual means of the homogenized series, and the t-value related to it. If the t-value estimated is located in the r=0 hypothesis rejecting area t> + or - critical t-value of 2.83 then trend exist otherwise it is not. The following equation is used to estimate the T-values.

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (13)$$

Where
Table 6 shows these results, which indicate the absence of the trend component in all of the data series of the six variables.
Before proceeding into the modeling process the data should be normalized to reduce the skewness coefficient to zero. The well-known Box-Cox transformation Box and **Jenkin , 1976** was used for this purpose as presented in the following equation:

$$XN = \frac{(H+\alpha)^{\mu} - 1}{\mu} \qquad (14)$$

Where:
μ : is the power
α : is the shifting parameter.
XN : is the normalized series.
Table.7 shows the coefficients of the normalization transformation of all of the six series. The shifting parameter is selected to ensure avoiding any mathematical problem that may occur due

to the fraction value of the power µ. The power value is found by trial and error so as to select its value that reduce the skewness to almost zero value. **Table 8** shows the statistical properties of the series before and after normalization, which indicate that the skewness coefficients are reduced to almost zero a property of the normal data.

The next step in the modeling process is to remove the periodic component to obtain the stochastic dependent component of the series, which is done by using Eq.(15), as follows:

$$\epsilon_{i,j}=\frac{XN_{i,j}-Xb_j}{Sd_j} \qquad (15)$$

Where:

$\epsilon_{i,j}$: is the obtained dependent stochastic component for year i, month j.

Xb_j : is the monthly mean of month j of the normalized series XN.

Sd_j : is the monthly standard Deviation of month j of the normalized series XN.

Table 9 shows the monthly means and monthly standard deviations of the normalized data series XN. The $\epsilon_{i,j}$ obtained series are then used to estimate the Lag-1 serial and cross correlation coefficients $\rho_{i,j}$, and $\sigma_{i,j}$of matrix Eqs.(6) and (7) respectively, which then used to estimate $\rho r_{i,j}$ and $\sigma r_{i,j}$ using Eqs.(9), and (11), respectively..

4. RESULTS AND DISCUSSION

The developed model above is used for data forecasting, recalling that the estimated parameters above are observed using the 22 years data series (1984-2005). This model will be used to forecast data for the next 5- years (2006-2010) since the data available are up to 2010, that could be compared with the observed series available for these years, for the purpose of model validation.

The forecasting process was conducted using the following steps:

1. Generation of an independent stochastic component (ξ) using normally distributed generator, for 5 years,i.e., (5*12) values.

2. Calculating the dependent stochastic component ($\epsilon_{i,j}$) using Eq. (2) and the matrices of $\rho r_{i,j}$ and $\sigma r_{i,j}$ as shown in Eqs. (9) and (11), respectively.

3. Reversing the standardization process by using the same monthly means and monthly standard deviations which were used for each variable to remove periodicity using Eq. (15) after rearranging.

4. Applying the inverse power normalization transformation (Box and Cox) for calculating un-normalized variables using normalization parameters for each variable and Eq.(14).

In most forecasting situation, accuracy is treated as the overriding criterion for selecting a model. In many instance the word “accuracy” refers to “goodness of fit,” which in turn refers to how well the forecasting model is able to reproduce the data that are already known. The model validation is done by using the following steps:

1. Checking if the developed monthly model resembles the general overall statistical characteristics of the observed series.

2. Checking if the developed monthly model resembles monthly means, monthly standard deviations using t-test for the means and F-test for the standard deviation.

Furthermore the performance of the new multi-variables multi-sites model developed herein was compared with the well-known single variable single site model, and multi-variables single site model (MATALS model). This performance was made to investigate whether the new model can produce better forecasted data series. For purpose of comparison of different forecasting models performance, the Akaike (AIC), test given by the following equation:

$$AIC = 2K + nLn\frac{Rss}{n} \tag{16}$$

Where:
n: is the number of the total forecasted values .
K: number of parameters of the model plus 1.
Rss: is the sum of square error between the forecasted value and the corresponding observed value.
For each site and variable three sets of data are generated. The overall statistical characteristics are compared with those observed, for each of the generated series. **Table 10** shows these comparisons. For all variables and sites the generated sets resemble the statistical characteristics not exactly with the same values of the observed series but sometimes larger or smaller but within an acceptable range. **Table 11** shows the t-test and F-test summary for all of the variables and sites. As it is obvious from the results of these tables, that the generated series succeed in (t-test) for all of the monthly means, except for two months for Sulaimania rainfall, i.e. overall succeed percent of (98.6%). This indicates that the model is successfully resembled the monthly means values, with excellent accuracy.
Based on (F-test) which seek the variance differences between the observed and generated series; the success percentage ranking of the generated series was: the best being for Sulaimania rainfall (96%), followed by Darbandikhan evaporation (88%), Darbandikhan rainfall (83%), Dokan evaporation (83%), Dokan rainfall (71%), and finally Sulaimania evaporation (67%). The overall success percentage was (81%). These results of the F-test indicate that the model was successfully resembled the monthly standard deviations, with a very good accuracy. As mentioned above for purpose of the comparison of the model performance with the available forecasting models, the **Akaike , 1974** test was used. Before that six single variable single site models were developed, one for each variable, and three single variable multi-site models, **Matalas ,1967** one for each site. These models were then used for forecasting monthly data for the same period (2006-2010), forecasted by the developed model.
Table.12 shows the Akaike test results for all of the forecasted variables, in each sites, obtained using these model and those obtained by the developed model. It is obvious that the developed model had produced for most of the cases the lowest test value, i.e, the better performance. Even though for some cases it has higher test value than the other models, but for these cases it is observed that a very little differences are exist between these test values and the minimum obtained one.

5. CONCLUSIONS

From the analysis done in this research, the following conclusion could be deduced:

1- The model parameters can be easily estimated and do not require any extensive mathematical manipulation.
2- The model can preserve the overall statistical properties of the observed series with high accuracy.

3- The model can preserve the monthly means of the observed series with excellent accuracy, evaluated using the t-test with overall success (98.6%).
4- The model can preserve the monthly standard deviations of the observed series with a very good accuracy, evaluated using the F-test with overall success (81%).
5- The comparison of the model performance with the single variable single site and the multi-site single variable models, using the Akaike tset had proved that the developed model had proved better performance in the most cases. Moreover for those less cases where other models had the better performance; the test value of the developed model is slightly higher than the minimum value.

REFERENCES

Al-Suhili R.H., Al-Kazwini, M. J., and Arselan, C. A., *Multivariate Multisite Model MV.MS. Reg. for Water Demand Forecasting* , Eng. and Tech. Journal Vol. 28, No. 13 ,2010, pp 2516-2529.

Akaike, H., 1974, *A New Look at the Statistical Model Identification* , IEEE T. Automat. Contr., 19(6), 716–723.

Barzinji K. T., 2003, *Hydrogic Studies for Goizha Dabashan and Other Watersheds in Sulimani Governorate* , M.Sc. thesis submitted to the college of Agriculture, University of Sulaimani

Box, G.E., and Jenkins, G. M. (1976), *Time Series Analysis and Control*, San Francisco, California: Holden-Day,Inc.

Calder C.A., 2007, *Dynamic Factor Process Convolution Models for Multivariate Space-Time Data with Application to Air Quality Assessment*, J. Environ.Ecol. Stat. Vol.14: 229-247.

Khalili M, Leconte R. and Brissette F., 2007, *Stochastic Multisite Generation of Daily Precipitation Data Using Spatial Autocorrelation*, J Hydrometeorology, Vol.8, P 396-412

Lee Seung-Jae and Wents E. A., 2010, *Space-Time Forecasting Using Soft Geostatitiscs: A Case Study in Forecasting Municipal Water Demand for Phhonex, Arizona* , Stoch Environ Risk Assess 24: pp 283- 295
.
Matalas N.C., 1967, *Mathematical Assessment of Synthetic Hydrology*, Water Resoures 3: 937-945.

Richardson C. W. and Wright D. A., 1984, *WGEN: A Model for Generating Daily Weather Variables*, United States Department of Agriculture, Agriculture Research Service ARS-8

Tobler W., (1970) , *A Computer Movie Simulating Urban Growth in the Detroit Region*, Economic Geography, 46(2): 234-240.

Wilks D. S., 1999, *Simultaneous Stochastic Simulation of Daily Precipitation, Temperature and Solar Radiation at Multiple Sites in Complex Terrain*, Elsevier, agricultural and forest meteorology 96:85-101.

Yevjevich, V. M., *The Structure of Hydrologic Time Series*, Fort Collins, Colorado State University, 1972.

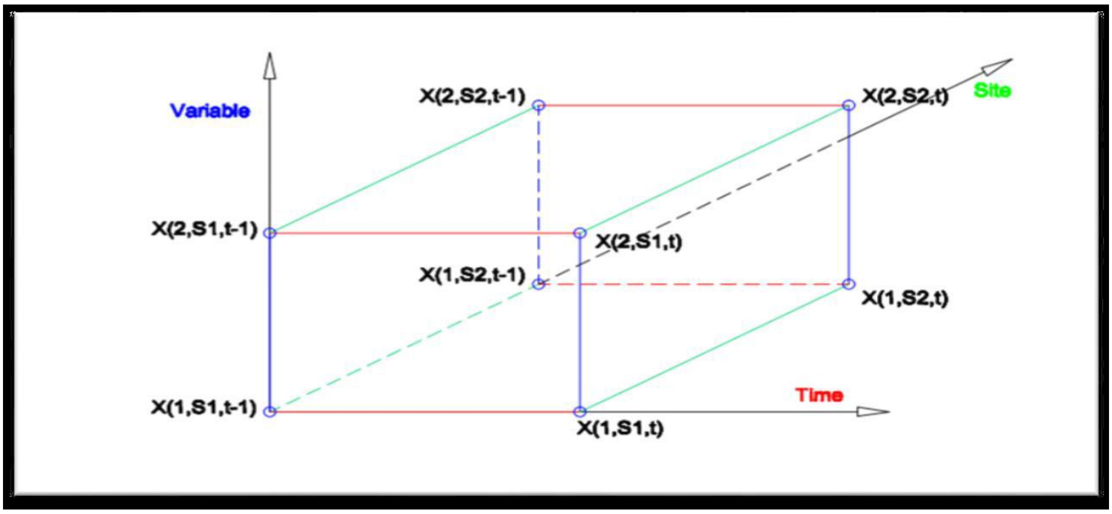

Figure 1. Schematic representation of the two variables two sites multi variables multisite model.

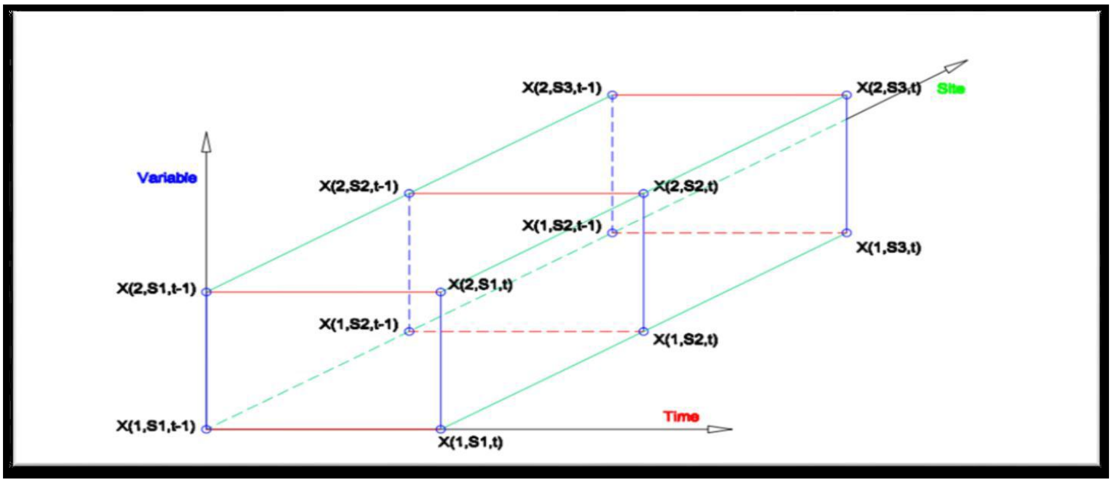

Figure 2. Schematic representation of the two variables three sites multi variables multisite model.

Table 1. North and east coordinates of the metrological stations selected for analysis.

Metrological station	N	E
Sulaimania	$35^{o}\,33'\ 18''$	$45^{o}\ \ 27'\ 06''$
Dokan	35^{o} 57’ 15”	$44^{o}\ \ 57'\ 10''$
Derbenikhan	$35^{o}\,06'\ 46''$	$45^{o}\,42'\ 23''$

Table 2. Descriptive statistics of the original data series.

Variable	1984-2005						2006-2010					
	Mean	S.D.	Skewne ss	Kurtosi s	Max	Min	Mean	S.D.	Skewne ssCs	Kurtosi s	Max	Min
Sulaimania Rainfall(mm)	91	69	0.9	0.7	354	0.1	73	63	1	3	310	0.1
Sulaimania Evapor.(mm)	120	70	1	1	415	36	106	51	1	-1	220	40
Dokan Rainfall(mm)	93	80	1	0.7	416	0.1	64	56	1	2	262	1
Dokan Evapor.(mm)	116	68	0.9	-0.1	322	24.9	101	60	1	1	284	35
Derbendkan Rainfall(mm)	78	67	1.2	1.3	326	0.4	69	59	1	1	247	0
Derbendkan Evapor.(mm)	134	69	0.7	-0.3	341	34.1	111	62	1	0.1	276	37

Table 3. Test of homogeneity results, with tc=2.38, at 95% significant level.

Variable	Test for Means				Test for S. D.			
	t-statistic	N1	N2	Homo.	t-statistic	N1	N2	Homo.
Sulaimania Rainfall(mm)	2.08	15	12	Yes	1.7	23	5	Yes
Sulaimania Evapor.(mm)	2.56	21	6	No	2.91	22	5	No
Dokan Rainfall(mm)	2.76	15	12	No	2.77	24	3	No
Dokan Evapor.(mm)	2.31	25	2	Yes	1.77	23	4	Yes
Derbendkan Rainfall(mm)	2.3	5	22	Yes	2.05	6	21	Yes
Derbendkan Evapor.mm)	3.25	18	9	No	0.93	17	10	Yes

Table 4. Coefficients of non-homogeneity removal.

Variable	Mean2	S.D.2	A1	B1	A2	B2
Sulaimania Evapor.(mm)	106.3	50.62	127.297	-0.602	62.958	0.7065
Dokan Rainfall(mm)	55.96	41.33	101.91	-0.815	79.314	-0.319
Derbendkan Evapor.mm)	117.5	70.45	131.18	0.4	69.93	-0.229

Table 5. Re- test of homogeneity, with tc=2.38, at 95% significant level.

Variable	Test for Means				Test for S. D.			
	t-statistic	N1	N2	Homo.	t-statistic	N1	N2	Homo.
Sulaimania Rainfall(mm)	2.08	15	12	Yes	1.7	23	5	Yes
Sulaimania Evapor.(mm)	1.69	26	1	Yes	1.17	9	18	Yes
Dokan Rainfall(mm)	1.92	1	26	Yes	1.61	2	25	Yes
Dokan Evapor.(mm)	2.31	25	2	Yes	1.77	23	4	Yes
Derbendkan Rainfall(mm)	2.3	5	22	Yes	2.05	6	21	Yes
Derbendkan Evapor.mm)	1.08	26	1	Yes	1.33	8	19	Yes

Table 6. Test of trend results, with tc=2.38, at 95% significant level.

Variable	T for means	T for S.D.
Sulaimania Rainfall(mm)	0.16	0.023
Sulaimania Evapor.(mm)	0.21	1.06
Dokan Rainfall(mm)	0.2	0.4
Dokan Evapor.(mm)	0.04	0.13
Derbendkan Rainfall(mm)	1.04	0.04
Derbendkan Evapor.mm)	0.28	0.41

Table 7. Coefficients of the normalization transformation.

Variable	Power μ	Shifting α
Sulaimania Rainfall(mm)	0.47	1
Sulaimania Evapor.(mm)	-0.52	0
Dokan Rainfall(mm)	0.27	0
Dokan Evapor.(mm)	-0.054	0
Derbendkan Rainfall(mm)	0.359	0
Derbendkan Evapor.mm)	0.232	0

Table 8. Statistical properties before and after the normalization transformation.

Variable	Before Norm.				After Norm.			
	Mean	S.D.	Skewness	Kurtosis	Mean	S.D.	Skewness	Kurtosis
Sulaimania Rainfall(mm)	91	69	0.9	0.7	14.36	6.84	-0.15	-0.49
Sulaimania Evapor.(mm)	106	49	1	0.7	1.8	1.32	0.1	-1.1
Dokan Rainfall(mm)	54.8	42.9	0.9	0.5	6.5	2.5	-0.1	-1.0
Dokan Evapor.(mm)	116	68	0.9	-0.1	4.1	0.5	0.0	-0.9
Derbendkan Rainfall(mm)	93	80	1	0.7	9.2	4.5	-0.12	-0.61
Derbendkan Evapor.(mm)	118.3	71.3	0.75	-0.27	8.3	1.8	-0.04	-0.79

Table 9. Monthly means and standard deviations for the dependent stochastic component.

Variable	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May
Sulaimania Rainfall Means	6.923738	15.95873	18.27623	17.95162	16.6502	17.02956	15.15592	8.927809
Sulaimania Rainfall S.D	5.851398	7.895653	6.532181	5.975568	5.221715	5.869138	5.521562	5.782993
Sulaimania Evapor. Mean	1.790015	1.750803	1.715869	1.707902	1.715362	1.756555	1.779253	1.810782
Sulaimania Evapor. S.D.	0.010111	0.016859	0.019150	0.022393	0.013306	0.017357	0.016122	0.010803
Dokan Rainfall Mean	3.908560	1.816542	3.908560	1.816542	3.908560	1.816542	3.908560	1.816542
Dokan Rainfall S.D.	6.718586	2.294015	6.718586	2.294015	6.718586	2.294015	6.718586	2.294015
Dokan Evapor Mean	4.56241	0.091936	4.56241	0.091936	4.56241	0.091936	4.56241	0.091936
Dokan Evapor. S.D.	4.006939	0.176918	4.006939	0.176918	4.006939	0.176918	4.006939	0.176918
Derbendkan Rainfall Mean	4.932498	9.747093	11.70001	11.72048	11.67984	11.31086	7.90416	4.803818
Derbendkan Rainfall S.D.	3.35148	4.540729	4.442926	3.669128	2.898333	3.798071	2.871659	2.83866
Derbendkan Evapor. Mean	10.02914	7.414099	6.914041	6.468882	6.859431	8.805194	9.050721	10.86809
Derbendkan Evapor. S.D.	0.599518	0.585723	2.012912	1.68868	0.775737	1.093621	0.677337	0.609332

Table 10. Statistical properties of observed and forecasted rainfall series (2006-2010).

Variable	Min	Max	Kurtosis	Skewness	S.D.	Mean
Sul. Obs. R	0.1	310	3	1	63	73
Sul. Gen 1 R	2.3	360.1	-0.6	0.4	66.1	75.1
Sul. Gen 2 R	2.5	282.1	0.1	0.7	64.2	72.7
Sul. Gen 3 R	6.8	248.3	-0.2	0.5	65.5	75.3
Dok. Obs. R	1	262	2	1	56	64
Dok. Gen 1 R	3	261	0.5	0.7	54	68
Dok. Gen 2 R	3	246	-0.1	0.7	52	58
Dok. Gen 3 R	6	266	0.2	0.7	49	65
Der. Obs. R	0.1	247	1	1	59	69
Der. Gen 1 R	0.9	375	-0.4	0.7	55	71
Der. Gen 1 R	1.4	243	0.1	0.9	51	66
Der. Gen 1 R	6.0	241	-0.1	0.7	53	78
Sul. Obs. E	40	220	-1	1	50	105
Sul. Gen 1 E	43	220	0.1	1.0	52	94
Sul. Gen 2 E	49	255	-0.1	1.0	53	101
Sul. Gen 3 E	53	227	0.0	0.9	48	109
Dok. Obs. E	35	262	1	1	60	101
Dok. Gen 1 E	27	276	-0.2	0.9	61	100
Dok. Gen 2 E	35	318	-0.1	0.9	68	105
Dok. Gen 3 E	37	279	-0.3	0.9	62	109
Der. Obs. E	37	276	0	1	62	111
Der. Gen 1 E	10	270	-0.4	0.7	62	95
Der. Gen 1 E	19	303	-0.5	0.6	67	123
Der. Gen 1 E	16	262	-0.5	0.7	60	108

R: Rainfall, E: Evaporation.

Table 11. Percentage success of T-test for monthly means and F-test for monthly standard deviations for three generated series for each variable, for years (2006-2010).

Variable	% Success in T-test	% Success in F-test
Sulaimania Rainfall(mm)	97.22	96
Sulaimania Evapor.(mm)	100	67
Dokan Rainfall(mm)	100	71
Dokan Evapor.(mm)	100	83
Derbendkan Rainfall(mm)	100	83
Derbendkan Evapor.mm)	100	88
Over all	98.60	81

Table 12. Comparison between the minimum Akaike test values obtained by the developed model, the multi-variable single site model, and the single variable single site model, for three generated series for each variable, by each model, for years (2006-2010).

variable	Sulaim. Rainfall	Sulaim. Evap.	Dokan Rainfall	Dokan Evap.	Derbend. Rainfall	Derbend. Evap.
The Developed Model	1,544	1,189	1,438	1,264	1,535	1,404
Multi Variable Multi Sites Model	1,600	1,179	1,421	1,253	1,628	1,423
Single Variable Single Site Model	1,606	1,191	1,463	1,214	1,550	1,420