Multi-Sites Multi-Variables Forecasting Model for Hydrological Data using Genetic Algorithm Modeling

Prof. Dr. Rafa H. Al-Suhili
The Senior Prof. of the Civil Engineering Dept., University of Baghdad, Baghdad, Iraq
A Visiting Prof. to the City College of New York, New York, USA.
Email:rafeaalsuhili@yahoo.com

ABSTRACT

A two time step stochastic multi-variables multi-sites hydrological data forecasting model was developed and verified using a case study. The philosophy of this model is to use the cross-variables correlations, cross-sites correlations and the two steps time lag correlations simultaneously, for estimating the parameters of the model which then are modified using the mutation process of the genetic algorithm optimization model. The objective function that to be minimized is the Akiake test value. The case study is of four variables and three sites. The variables are the monthly air temperature, humidity, precipitation, and evaporation; the sites are Sulaimania, Chwarta, and Penjwin, which are located north Iraq. The model performance was checked by comparing it's results with the results of six forecasting models developed for the same data by Al-Suhili and khanbilvardi, 2014.The check of the performance of the new developed model was made for three forecasted series for each variable, using the Akaike test which indicates that the developed model is more successful, since it gave the minimum (AIC) values for (91.67 %) of the forecasted series. This indicates that the developed model had improved the forecasting performance. For the rest of cases (8.33%), other models gave the lowest AIC value, however it is slightly lower than that given by the developed model. Moreover the t-test for monthly means comparison between the models indicates that the developed model has the highest percent of succeed (100%).
Keywords: forecasting, multi-sites, multi-variables, cross sites correlation, serial correlation, cross variables correlations, hydrology.

نموذج تنبأ بالمعلومات الهيدرولوجية متعدد المواقع ومتعدد المتغيرات باستخدام تقنية الجينات الوراثية

رافع هاشم السهيلي
استاذ
كلية الهندسة- جامعة بغداد

الخلاصة

تم في هذا البحث اشتقاق نموذج تنبأ بالبيانات الهيدرولوجية متعدد المواقع متعدد المتغيرات ويعتمد على خطوتين زمنيتين وتم برهنته باستخدام حالة دراسية. ان فلسفة هذا النموذج تعتمد على استخدام معاملات الأرتباط بين المتغيراتوبين المواقع ومعاملات الأرتباط الزمني لخطوتين زمنيتين سابقتين بشكل اني لأيجاد معاملات النموذج ومن ثم يتم تغيير قيمها باستخدام عملية المعايرة الخاصة بتقنبة الجينات الوراثية . كما هو معروف هن تقنية الجينات الوراثية هي تقنية تستخدم لأيجاد القيمة المثلى لدالة الهدف حيث ان الالة المستخدمة هنا والتي يتم ايجاد القيمة الصغرى لهى هي دالة اختبار اكايكي. ان الحالة الدراسية المأخوذة هس لأربع متغيرات في ثلاث مواقع . المتغيرات هي درجة حرارة الهواء والرطوبة والسقيط والتبخر الشهرية والمواقع هي السليمانية وجوارتة وبنجوين التي تقع في شمال العراق. تم مقارنة اداء النموذج مع نتائج ستة نماذج تنبأ ولنفس حالة الدراسة. هذه المقارنة تمت لثلاثة متسلسلات زمنية لكل متغير في كل موقع تم التنبأ بها بستخدام كل من النماذج الستة السابقة والنموذج الجديد وباستخدام الأختبار المشار اليه اعلاه واشارت النتائج بان النموذج الجديد اكثر نجاحا لانه اعطى اقل قيم للأختبار بنسبة 91.76 %. وهذا يدل على ان النموذج الجديد قد حسن عملية التنبأ.

الكلمات الرئيسية: التنبؤ، مواقع متعددة، متغيرات متعددة، ارتباط المواقع المتقاطعة، ارتباط متسلسل، ارتباط المتغيرات المتقاطعة، هيدرولوجي.

1. INTRODUCTION

Forecasting models of hydrological data series have been used successfully for different water resources studies. By now it is used in different research topics, including, water resources projects planning and design, climate changes studies and reservoirs operation. They can generate series of climatic data that reflects the same statistical properties as the observed ones. Moreover, weather generators are able to produce series for longer length in time than the observed ones, and hence gives much more significance on the decision making process. This allows a better identification of the consequences of extreme events, such as extreme flood, extreme draught, and hence allowing sufficient water resources management to make the required preparations for the expected draught or flood events. Different types of forecasting models are available in the literature that can be used for weather data forecasting.

Single variable single site forecasting models (SVSS) are used for forecasting a hydrological variable at a single site independent of the same variable at the near sites ignoring the spatial dependence that may exist in the observed data. This model, also ignore the cross variables relations that may physically exist between these variables. Generally correlation between variables in different sites with different variables may exist in real applications**. ,Matals, 1967.** had developed a single variable multi-sites model (SVMS) using cross site correlations between one variable at different sites. This model can be applied as a multi-variable single site (MVSS) model that uses multi variables cross correlation in a given site. ,**Richardson, 1981.** had developed a multi-variables stochastic weather models for daily precipitation, maximum temperature, minimum temperature, and solar radiation, as cited in, ,**Wilks, 1999.** 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. Recently progress had been made especially in the last twenty years to come up with theoretical frameworks for spatial analysis ,**Khalili, 2007**. **Lee, et al., 2010.** had developed a space–time model to regionalize the weather generators. In these models, the precipitation is linked to the atmospheric circulation patterns using conditional probability distributions and conditional spatial covariance functions. 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. There exist in the literature some relatively recent trials to account for the spatial variation in multi-sites, and variables correlations among different variables in the same site and in the other sites.

In a reliable hydrological system the cross variables and cross sites correlation may exist between different hydrological variables at different sites, in addition to the time lag correlations. ,**Al-Suhili, et al., 2010,** had presented a multivariate multisite model for forecasting different water demand types at different areas in the city of Karkouk, north Iraq. This model use in advance regression analysis to relate each demand type with explanatory variables that affect its type, then obtaining the residual series of each variable at each site. These residual are then modeled using a multisite, **Matalas, 1967,** models for each type of demand. These models were then coupled with the regression equation to simulate the multi-variables multi-sites simulation. The last two cited research are those among the little work done on forecasting models of multi-sites multi-variables types. However these model are rather complicated, and/or do not model the process of cross site, cross variables correlation and time lag correlation simultaneously, which as mentioned above is the real physical case that exist. Hence researches are further required to develop a simplified multi-sites multi-variables model. ,**Al-Suhili, and Mustafa, 2013,** had proposed a multi-variables multi-sites model that uses relative correlation matrix and a residual matrix as the model parameters to relate the dependent and independent stochastic components of the data. This model represents the dependent stochastic of each variable at a time step as a weighted sum of the dependent stochastic component at the preceding time step and the present independent stochastic components. However these weights are not summed to one, while logically they should be. ,**Al-Suhili, and Khanbilvardi, 2014,** had developed a model as modification of ,**Al-Suhili, and Mustafa model, 2013,** using one degree time step model and relative weighted correlations, with one time degree, and weights that sums to 1.

In this research a modified multi-variables multi-sites approach is proposed to develop a model that describe the cross variables, cross sites correlation and lag-time correlation structure in the forecasting of multi variables at multi sites simultaneously. This model represents a modification of ,**Al-Suhili, and Khanbilvardi model, 2014,** by extending the time dependence to the second degree, and apply the mutation process of the genetic algorithm model to these parameters, such that to minimize the Akaike test value. The modification is done such that the total weights of the lag 1 and lag 2 correlations and the residual correlations are summed to 1, i.e. each variable is resulted from the weighted sum of the other variables in the same site and those in the other sites in addition to the same and other variable at the preceding two time steps. This was done by adopting a different method for estimating the parameters of the model using lag 1 and lag 2, time correlations rather than using only lag 1 time correlation, moreover these parameters were then subjected to a mutation process of the genetic algorithm model but with keeping the sum of the weights to 1 as a constraint. The mutation process continues until minimizing the Akaike test value as an objective function. This model was applied to the same case study used by ,**Al-Suhili, and Khanbilvardi, 2014,** model for the sake of comparison. The case study is for the monthly data of four hydrological variables, air temperature, humidity, precipitation and evaporation at three sites located north Iraq, Sulaimania, Chwarta, and Penjwin.

2.THE MODEL DEVELOPMENT

The multivariate multisite model developed herein, utilizes single variable time lag one and lag two correlations, cross variables lag-one and lag- two correlations, and cross sites lag-one and lag-two correlations. In order to illustrate the model derivation consider **Fig. 1a,** where the concept of these correlations is shown, ,**Al-Suhili, and Khanbilvardi, 2014**. This figure illustrates the concept for two variables, two sites and first order lag-time model. This simple form is used to simplify the derivation of the model. However, the model could be easily generalized using the same concept. For instant, **Fig. 1b**, is a schematic diagram for the multi-variables multi-sites model of two variables, three sites and first order lag-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 model and was abbreviated as MVMS (V, S, O), where V: stands for number of variables in each site, S: number of sites, and O: time order, hence the model representation in figure (1a and b) can be designated as MVMS (2, 2, 1), and MVMS (2, 3, 1), respectively.

This model can be extended further to (V-variables) and / or (S-sites) and / or (O- time) order. 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 times (t-1)and (t-2) at all sites. The expression is weighted by the first two time lags serial correlation coefficients, cross-site correlation coefficients, cross-variable coefficients and cross-site, cross-variable correlation coefficients. In addition to that; the independent stochastic components are weighted by the residuals of all types of these correlations. These residual correlations are expressed using the same concept of autoregressive second order model (Markov chain). Further modification of this model is to use relative correlation matrix parameters by using correlation values relative to the total sum of absolute lag-1 and lag-2correlations for each variable, and the total sum of the absolute residuals as a mathematical filter ,as will be shown later.

A model matrix equation for second order time lag, O=2, number of variables=V, and number of sites=S, could be put in the following form:

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

Which for V=2,S=3,and O=2, can be represented by the following equation:

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

Where:

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

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

$$\begin{pmatrix} \epsilon_{(v1,s1)} \\ \epsilon_{(v2,s1)} \\ --- \\ \epsilon_{(v1,s2)} \\ \epsilon_{(v2,s2)} \\ --- \\ \epsilon_{(v1,s3)} \\ \epsilon_{(v2,s3)} \end{pmatrix}_{t-2} = [\epsilon_{t-2}]_{6,1} \quad (5)$$

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

t

$$\begin{pmatrix} \phi1_{1,1} & \phi1_{1,2} & \phi1_{1,3} & \phi1_{1,4} & \phi1_{1,5} & \phi1_{1,6} \\ \phi1_{2,1} & \phi1_{2,2} & \phi1_{2,3} & \phi1_{2,4} & \phi1_{2,5} & \phi1_{2,6} \\ \phi1_{3,1} & \phi1_{3,2} & \phi1_{3,3} & \phi1_{3,4} & \phi1_{3,5} & \phi1_{3,6} \\ \phi1_{4,1} & \phi1_{4,2} & \phi1_{4,3} & \phi1_{4,4} & \phi1_{4,5} & \phi1_{4,6} \\ \phi1_{5,1} & \phi1_{5,2} & \phi1_{5,3} & \phi1_{5,4} & \phi1_{5,5} & \phi1_{5,6} \\ \phi1_{6,1} & \phi1_{6,2} & \phi1_{6,3} & \phi1_{6,4} & \phi1_{6,5} & \phi1_{6,6} \end{pmatrix} = [\phi\ 1]_{6,6} \qquad (7)$$

$$\begin{pmatrix} \phi2_{1,1} & \phi2_{1,2} & \phi2_{1,3} & \phi2_{1,4} & \phi2_{1,5} & \phi2_{1,6} \\ \phi2_{2,1} & \phi2_{2,2} & \phi2_{2,3} & \phi2_{2,4} & \phi2_{2,5} & \phi2_{2,6} \\ \phi2_{3,1} & \phi2_{3,2} & \phi2_{3,3} & \phi2_{3,4} & \phi2_{3,5} & \phi2_{3,6} \\ \phi2_{4,1} & \phi2_{4,2} & \phi2_{4,3} & \phi2_{4,4} & \phi2_{4,5} & \phi2_{4,6} \\ \phi2_{5,1} & \phi2_{5,2} & \phi2_{5,3} & \phi2_{5,4} & \phi2_{5,5} & \phi2_{5,6} \\ \phi2_{6,1} & \phi2_{6,2} & \phi2_{6,3} & \phi2_{6,4} & \phi2_{6,5} & \phi2_{6,6} \end{pmatrix} = [\phi\ 2]_{6,6} \qquad (8)$$

$$\begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} & \sigma_{1,6} \\ \sigma_{2,1} & \sigma_{2,2} & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} & \sigma_{2,6} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} & \sigma_{3,4} & \sigma_{3,5} & \sigma_{3,6} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_{4,4} & \sigma_{4,5} & \sigma_{4,6} \\ \sigma_{5,1} & \sigma_{5,2} & \sigma_{5,3} & \sigma_{5,4} & \sigma_{5,5} & \sigma_{5,6} \\ \sigma_{6,1} & \sigma_{6,2} & \sigma_{6,3} & \sigma_{6,4} & \sigma_{6,5} & \sigma_{6,6} \end{pmatrix} = [\sigma\ ]_{6,6} \qquad (9)$$

Now if we define the followings:
$\rho1_{1,1}$ = ρ1 [(x1, x1), (s1, s1), (t, t-1)]= population serial correlation coefficient of variable 1 with itself at site 1 for time lagged 1

$\rho1_{1,2}$ = ρ1 [(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

$\rho1_{1,3}$ = ρ1 [(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

$\rho 1_{1,6}$ = ρ1 [(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 for all the variables and all the sites. Similarly the lag 2 correlations are:

$\rho 2_{1,1}$ = ρ2 [(x1, x1), (s1, s1), (t, t-2)]= population serial correlation coefficient of variable 1 with itself at site 1 for time lagged 2

$\rho 2_{1,2}$ = ρ2 [(x1, x2), (s1, s1), (t, t-2)]= population cross correlation coefficient of variable 1 at site 1 with variable 2 at site 1, for time lagged 2

ϵ: is the stochastic dependent component.

ξ: is the stochastic independent component.

The coefficients of the matrices in Eqs. (7) (8) and (9) are estimated according to the correlation structure in the auto regressive models filtered by a division mathematical absolute summation filter, that makes the dependent component of each variable in each site at time t expressed as a weighted sum of the dependent components of the variables in the same site and the other sites at time steps t-1 and t-2 with additional weighted terms of the independent stochastic component of each variable and all sites. The following equations were resulted:

$$\boldsymbol{\phi 1_{i,j}} = \frac{\rho 1_{i,j}(1-\rho 2_{i,j})/(1-\rho 1_{i,j}^2)}{\sum_{j=1}^{n=v*s} abs\,(\rho 1_{i,j}) + abs(\rho 2_{i,j}) + abs(\sigma_{i,j})}. \tag{10}$$

$$\boldsymbol{\phi 2_{i,j}} = \frac{(\rho 2_{i,j}^2-\rho 1_{i,j}^2)/(1-\rho 1_{i,j}^2)}{\sum_{j=1}^{n=v*s} abs\,(\rho 1_{i,j}) + abs(\rho 2_{i,j}) + abs(\sigma_{i,j})}. \tag{11}$$

σ values are estimated using the following equation:

$$\boldsymbol{\sigma_{i,j}} = \frac{\left(1+\phi 2_{i,j}\right)(1-\phi 2_{i,j}^2-\phi 1_{i,j}^2)/\left(1-\phi 2_{i,j}\right)}{\sum_{j=1}^{n=v*s} abs\,(\rho 1_{i,j}) + abs(\rho 2_{i,j}) + abs(\sigma_{i,j})}. \tag{12}$$

3.THE GENETIC ALGORITHM MODEL

The estimated parameters in Eqs. (7),(8), and (9), are considered as a first estimation and then subjected to a mutation process of the genetic algorithm technique. The mutation process was done by adding and/or subtracting small values to the parameters such that the absolute sum of each row in matrices,1,**ϕ**2,and σ, is kept to be 1. This is the constraint of the optimization process performed by the mutation process of the genetic algorithm technique and can be represented by the following equation, for the first variable at the first site.

$$\sum_{j=1}^{v*s}(\boldsymbol{\phi 1_{f,j}} \pm \boldsymbol{u1_{f,j}}) + \sum_{j=1}^{v*s}(\boldsymbol{\phi 2_{f,j}} \pm \boldsymbol{u2_{f,j}}) + \sum_{j=1}^{v*s}(\boldsymbol{\sigma_{f,j}} \pm \boldsymbol{u3_{f,j}}) = 1 \tag{13}$$

With f=1, similar expression can be obtained for the other variables , by setting f=2,3, …,v*s.
And $\boldsymbol{u1_{f,j}}, \boldsymbol{u2_{f,j}}, \boldsymbol{u3_{f,j}}$: are the mutation levels of row f, and column j, of the **ϕ**1,**ϕ**2,andσ matrices, respectively.

The objective functions to be each minimized are the Akaike test value for each variable in each site as given by the following equation:

$$\boldsymbol{min\ AIC_f = 2K + n\ Ln\frac{Rss}{n}} \quad ,f=1,2,\ldots..v*s \qquad (14)$$

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. If the values of Rss/n is less than one the AIC value may be negative and the performance is better if the absolute AIC value is larger.

This means that the minimization process is done for each variable and each site , using row by row process. The number of AIC functions to be minimized is v*s.

4.THE CASE STUDY AND APPLICATION OF THE MODEL

In order to apply the developed two degree time step, MVMS(4,3,2) model explained above the Sulaimania Governorate was selected as a case study, which is the same case study used by Al-Suhili and **Khanbilvardi, 2014,** as mentioned before. This was done in order to compare the results of the new developed model with the previous models which was applied by those authors for the same case study. The following description of this case study was fully taken from this refrence. 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 are the monthly air temperature, humidity, precipitation and evaporation .These 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, Chwarta and Penjwin. These stations are part of the metrological stations network of Sulaimania governorate north Iraq. This network has eight weather stations distributed over an approximate area of (17023 km^2). **Table 1**, shows the names, latitudes, longitudes and elevations of these stations. **Fig. 2**, shows a Google map of the locations of these stations. **Table 2**, shows the approximate distances between these stations and all of the metrological stations in Sulaimania governorate.

The model was applied to the data of the case study described above. The available length of the records for the four variables and the three stations is (8) years of monthly values, (2004-2011). The data for the first (5) years, (2004-2008) were used for the estimation and the mutation of the model parameters matrices **ϕ**1,2, and σ, while the left last 3 years data, (2009-2011) were used for verification. The data includes the precipitation as a variable which has zero values for June, July, August and September, in the selected area of the case study. These months are included in the analysis, by adding a constant value to the precipitation series of 0.1 to avoid the problems that may be created by these zeros.

In all similar analysis and before applying the forecasting model shown in eq. (1), a prior analysis should be made for each variable at each site to estimate it's dependent stochastic component .These steps were done for each variable at each site of the case study by ,**Al-Suhili, and khanbilvardi, 2014.** These steps are test of homogeneity using method proposed by ,**Yevjevich,1972.** trend test and normalization transformation. The results indicate that most of the variables in all sites are homogeneous and non-homogeneity is only exist in Sulaimania air temperature, Penjwin humidity, Penjwin air temperature, and Penjwin evaporation series. These non homogeneous series were homogenized using the method proposed by Yevjevich(1972). Trend analysis indicates that all of the data variables in all of the sites are free from trend. The well known Box-Cox transformation Box and ,**Jenkins, 1976,** was used for the purpose of normalization of data and was found successful.

The estimation of the stochastic dependent component of the series, was done 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.

For more details of these analysis one can refer to **Al-Suhili and Khanbilvardi,2014**. The next step in the modeling process is to estimate the parameters of the model. The $\epsilon_{i,j}$ obtained series are used to estimate the Lag-1 and lag-2 serial and cross correlation coefficients and then estimate $\phi k_{i,j}$, k=1,and 2, and $\sigma_{i,j}$ of matrix Eqs. (7) ,(8)and (9) respectively, but with each matrix size of 12*12, using the developed Eqs.(10), (11)and (12),but with each matrix size 12*12, respectively. This step is definitely different than this used by Al-Suhili and Khanbilvardi(2014), since the model developed herein is different. Moreover these parameters are then muted using the genetic algorithm model developed above.The expected results from this development is a better performance model since the development includes the use of two time steps correlations rather than only one time step used by the previous model. The estimated and muted parameters of the model are shown in **Tables 3,4 and 5.**

5.FORECASTING RESULTS AND DISCUSSION

The developed model mentioned above was used for data forecasting, recalling that the estimated parameters above are obtained using the 5 years data series (2004-2008). The forecasted data are for the next 3- years (2009-2011), 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 which are the typical steps for forecasting:

1. Generation of an independent stochastic component (ξ) using normally distributed generator, for 3 years, i.e., (3*12) values.
2. Calculating the dependent stochastic component ($\epsilon_{i,j}$) using equation (2),with v=4,s=3 and the matrices of $\phi k_{i,j}$, k=1 and 2, and $\sigma_{i,j}$.
3. Reversing the standardization process by using the same monthly means and monthly standard deviations which were used for each variable 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.

For forecasting models, accuracy of results is considered as the overriding criterion for selecting a model. The word “accuracy” refers to the “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 typical 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 using the t-test.
3. Checking the performance of the model of the hole forecasted series using Akaike test.

For the purpose of comparison of the forecasting performance between the new multi-variables multi-sites model developed herein and the other models the Akaike test can be used. This performance comparison was made to investigate whether the new model can produce better forecasted data series. For this purpose the Akaike (AIC), given by the equation (14), without minimization for the six models and with minimization for the present developed model.

For each site and variable three sets of data are generated, using the seven different models mentioned above. The overall statistical characteristics are compared with those observed, for each of the generated series. It is observed that the seven models can all give good resemblances for these general statistical properties. 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. No distinguishable performance of any of the model can be identified in this comparison of the general statistical properties. **Tables 6,7. and 8**. show the t-test percent of succeed comparison summary for all of the variables and sites, for the three generated series. As it is obvious from the results of these tables, the generated series for the first six model succeed in (t-test) with high percentages except for the Penjwin station where sometimes low percentage is observed. It is also clear that the developed model had increased the percent of succeed. The developed model had the highest overall percent of succeed among the other models (100%). However the differences are small.

For purpose of the comparison between the developed model performance and that of the available forecasting models and the developed model for the data as mentioned above, the Akaike(1974) test was used. **Table 9**, shows the Akaike test results for all of the forecasted variables, in each site, obtained using the six models 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. These cases represent (91.67%). However for these cases where the lowest AIC value is given by a model other than the developed model, the developed model had gave the next lowest AIC values. Moreover for these cases it is observed that very small differences are exist between these test values of the new model and the minimum obtained one.

6.CONCLUSIONS

From the analysis done in this research, the following conclusion could be deduced:
The model parameters estimation and mutation processes are simple and the sum of each row of the **ϕ**1, 2, and σ matrices is equal to 1, which reflects the weighted sum of the variables.The model can preserve the monthly means of the observed series with excellent accuracy, evaluated using the t-test with overall success (100%). However, the differences between the percent success is not so high between the developed model and the other models.

The comparison of the model performance with the other models performances using the Akaike test had proved that the developed model had a better performance for the most cases (91.67%). Moreover for those remaining cases where other model had the better performance (minimum AIC value); the test value of the developed model is slightly higher than this minimum value.

REFERENCES

- Al-Suhili, R. H. and Mustafa, N.F, 2003, *A Multi- variables Multi-sites Model for Hydrological Series Data Forecasting,* accepted for publication at the Journal of Eng. College of Engineering, University of Baghdad.

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

- Al-Suhili R.H., and Khanbilvardi R., 2014, *A Multi-Sites Hydrological Forecasting Model Using Relative Correlations,* American Journal of Engineering Research Vol.3, Issue 6, pp154-168.

- Akaike, H., 1974, *A New Look at the Statistical Model Identification,* IEEE T. Automat. Contr., 19 (6), pp716–723.

- 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 Sulaimania.

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

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

- *Lee S. J. and Wents E. A., 2010. Space-Time Forecasting Using Soft Geo statistics: A Case Study in Forecasting Municipal Water Demand for Phonex, Arizona,* Journal of Stochastic Environ Risk Assess vol. 24: pp 283- 295.

- *Matalas N.C, 1967. Mathematical Assessment of Synthetic hydrology,* Journal of Water Resources Researches vol. 3: pp 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.

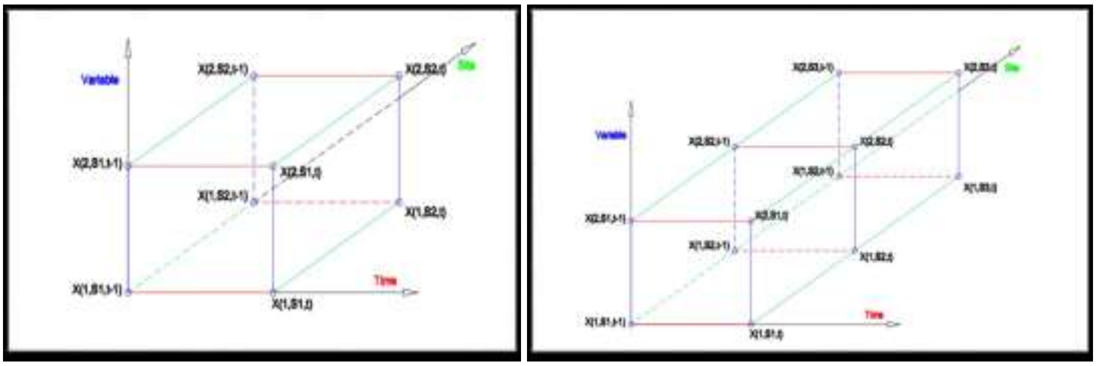
- *Wilks D. S., 1999. Simultaneous Stochastic Simulation of Daily Precipitation, Temperature and Solar Radiation at Multiple Sites in Complex Terrain,* Journal of Elsevier, agricultural and forest meteorology 96:pp 85-101,.

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

ABBREVIATIONS

SulAt: Air temperature at Sulaimania.
SulHu: Humidity at Sulaimania.
SulPr: Precipitation at Sulaimania.
SulEv: Evaporation at Sulaimania.
ChwAt: Air temperature at Chwarta.
ChwHu: Humidity at Chwarta.

ChwPr: Precipitation at Chwarta.
ChwEv: Evaporation at Chwarta.
PenAt: Air temperature at Penjwin.
PenHu: Humidity at Penjwin.
PenPr: Precipitation at Penjwin.
PenEv: Evaporation at Penjwin
SVSS: Single variable, single site model.
SVMS: Single variable, multi-sites model.
MVSS: Multi variables, single site model.



(a) **(b)**

Figure 1. Schematic representation of the developed multi-variables multi-sites model, a)MVMS(2,2,1), b) MVMS(2,3,1),**Al-Suhili and Khanbilvardi,2014**.

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''

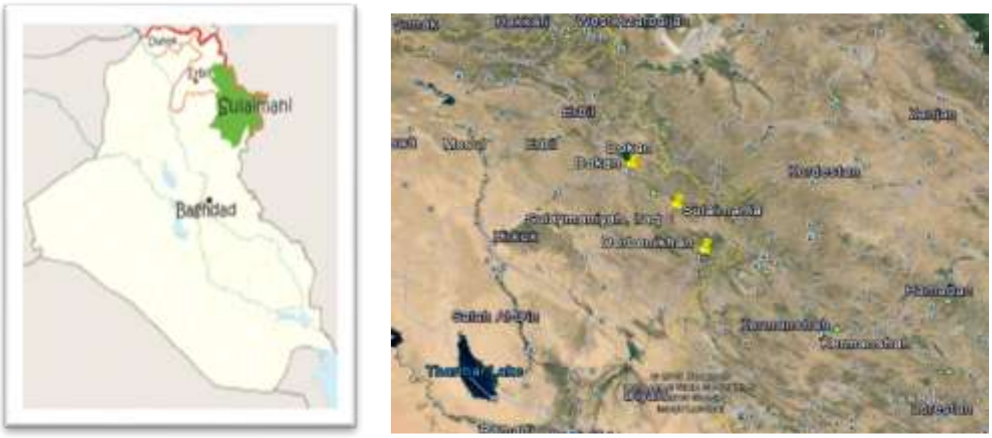


Figure 2. Locations of the metrological stations selected for analysis.

Table 2. Approximate distances between the Sulaimania weather stations network (km.), **Barzingi, 2003.**

Name of Weather Station	Sulaimani	Dukan	Darbandikhan	Penjwin	Chwarta	Halabjah	Bazian	Chamchamal
Sulaimani	0	62.76	54.00	45.88	20.85	63.36	29.17	56.10
Dukan	62.76	0	114.73	97.10	61.20	125.85	42.00	47.90
Darbandikhan	54.00	114.73	0	61.40	68.68	28.36	73.98	90.57
Penjwin	45.88	97.10	61.40	0	36.53	48.22	74.15	102.12
Chwarta	20.85	61.20	68.68	36.53	0	69.73	41.30	69.90
Halabjah	63.36	125.85	28.36	48.22	69.73	0	89.50	111.05
Bazian	29.17	42.00	73.98	74.15	41.30	89.50	0	28.41
Chamchamal	56.10	47.90	90.57	102.12	69.90	111.05	28.41	0

Table 3. Model Coefficients Matrix **ϕ**1.

	SulAt	SulHu	SulPr	SulEv	ChwAt	ChwHu	ChwPr	ChwEv	PenAT	PenHu	PenPr	PenEV
SulAT	9E-05	0.00083	0.001216	-0.00089	0.001039	0.003111	0.00092	0.00302	0.0004902	0.00264	0.00306	0.00334
SulHu	-0.002	-7E-05	0.00179	0.001458	0.0009656	0.001133	0.00162	0.00496	0.0033658	0.00154	0.00154	0.00349
SulPr	-7E-04	-0.0005	4.68E-05	0.004946	-0.000473	-0.00016	0.00014	0.00396	-0.000569	0.00029	-0.0005	0.00166
SulEv	-5E-04	-0.0004	0.005277	-3.8E-06	0.0002179	0.002646	0.00466	0.00075	-3.75E-06	0.00187	0.00471	0.00024
ChwAT	0.0009	0.00258	0.001425	-2.3E-05	-8.26E-05	0.006609	0.00122	0.00363	0.0007436	0.00244	0.00387	0.0041
ChwHu	-5E-04	0.00087	0.001276	0.001938	0.0018334	0.000174	0.00074	0.00164	0.0040532	0.00201	0.00097	0.0069
ChwPr	-8E-04	-0.0007	0.00026	0.005164	-0.000885	-0.00038	1.1E-05	0.00429	-0.000984	0.00032	-0.0003	0.00085
ChwEv	0.0003	0.00237	0.002408	-0.00049	0.0010564	0.001358	0.0025	- 0.00077	-0.002271	0.0058	0.00344	0.00269
PenAT	0.0015	0.00069	0.000239	0.000859	0.0022123	0.003494	3.6E-05	0.00097	0.0001551	-9E-05	0.00223	0.00031
PenHu	0.0006	0.00113	0.002238	-8.4E-05	0.0012027	0.000747	0.00159	0.00713	-0.000475	-0.0003	0.00182	0.01189
PenPr	-0.001	8E-05	0.000863	0.002284	-0.00036	-0.0001	0.00066	0.00315	-0.000664	0.00112	-9E-07	0.00138
PenEv	0.0035	0.00207	0.000333	-0.00015	0.0037207	0.00343	0.00039	0.00388	-0.00017	0.01135	0.002	0.00089

Table 4. Model Coefficients Matrix **ϕ**2.

	SulAt	SulHu	SulPr	SulEv	ChwAt	ChwHu	ChwPr	ChwEv	PenAT	PenHu	PenPr	PenEV
SulAT	0.0406	0.00348	-0.00121	0.014164	0.0326137	0.001676	-0.0008	0.01678	0.0185219	0.00054	-0.002	0.00431
SulHu	0.0027	0.04271	0.011687	0.005042	0.005783	0.024246	0.01223	0.00617	0.0033607	0.0054	0.01746	-0.0021
SulPr	-3E-05	0.01266	0.036896	0.004796	0.001517	0.013677	0.03391	0.00397	0.0005742	0.00959	0.02605	-0.0008
SulEv	0.0147	0.00584	0.005347	0.04445	0.012055	0.004235	0.00797	0.01529	0.0110532	0.00074	0.00392	0.00246
ChwAT	0.0302	0.00539	0.000757	0.010837	0.0406819	0.003852	0.00041	0.01726	0.0194488	0.00117	0.00087	0.00519
ChwHu	0.0031	0.02582	0.013814	0.004269	0.0066439	0.039592	0.01164	0.00467	0.0033397	0.0083	0.01733	-1E-04
ChwPr	-7E-04	0.01256	0.032747	0.006544	0.0007594	0.010903	0.03839	0.00448	-0.000249	0.00835	0.02783	-3E-05
ChwEv	0.0163	0.00665	0.004293	0.012837	0.0188531	0.004149	0.00504	0.04542	0.001972	0.00373	0.00526	0.01221
PenAT	0.0199	0.00542	0.000964	0.011641	0.0221966	0.004502	0.00089	0.00505	0.0455255	0.00065	0.00148	0.0022
PenHu	0.0024	0.00686	0.009741	0.001986	0.002636	0.010262	0.00924	0.00478	0.0003667	0.04882	0.01738	-0.0004
PenPr	0.0006	0.01691	0.023759	0.00423	0.003011	0.015529	0.02632	0.00541	0.0019096	0.0142	0.03727	0.00115
PenEv	0.0064	0.00135	0.001288	0.002988	0.0098815	0.004767	0.00086	0.01719	0.0025273	0.00219	0.00166	0.04873

Table 5. Model Coefficients Matrix σ.

	SulAt	SulHu	SulPr	SulEv	ChwAt	ChwHu	ChwPr	ChwEv	PenAT	PenHu	PenPr	PenEV
SulAT	0.1622	0.04904	0.039409	0.074906	0.1322718	0.044986	0.04027	0.08161	0.0870505	0.04273	0.03765	0.05055
SulHu	0.0467	0.17091	0.067669	0.051751	0.0534633	0.103785	0.06908	0.05358	0.0477855	0.05256	0.08332	0.03671
SulPr	0.0384	0.068	0.14765	0.047889	0.0416035	0.070726	0.13617	0.04636	0.0396642	0.06007	0.10822	0.03682
SulEv	0.0787	0.05682	0.054907	0.177799	0.0717521	0.053052	0.06104	0.08019	0.0692286	0.04579	0.05195	0.04943
ChwAT	0.1231	0.05148	0.041675	0.064794	0.1627617	0.046955	0.04099	0.08132	0.0884664	0.04243	0.04157	0.05071
ChwHu	0.0473	0.10868	0.073023	0.04967	0.0550543	0.158355	0.06737	0.05059	0.0472773	0.05893	0.08276	0.03944
ChwPr	0.0372	0.06779	0.131862	0.051848	0.0401359	0.0635	0.15355	0.0475	0.0381013	0.05711	0.1143	0.03853
ChwEv	0.08	0.05497	0.049646	0.070487	0.0871308	0.049449	0.05128	0.18249	0.0446837	0.04759	0.0516	0.06853
PenAT	0.0949	0.058	0.048481	0.072703	0.1012188	0.055655	0.04833	0.05715	0.1820766	0.04785	0.04942	0.05105
PenHu	0.0522	0.062	0.068644	0.051373	0.0527015	0.070048	0.06753	0.05605	0.0480483	0.19546	0.0883	0.04359
PenPr	0.0382	0.07864	0.099759	0.045894	0.0433918	0.07468	0.10837	0.04838	0.0410341	0.07089	0.14909	0.03942
PenEv	0.0648	0.05418	0.054126	0.057668	0.0727816	0.06122	0.05325	0.09105	0.0566975	0.05327	0.0548	0.19451

Table 6. Comparison between the percent of succeed in t-test for differences in monthly means of the generated and observed data for set 1 generated series, by each model.

	SVSS	SVMS	MVSS	**Matalas, 1967**	**Al-Suhili and Mustafa, 2013**	**Al-Suhili and Khanbilvardi, 2014**	The Developed Model
SulAT	100	91.667	100	91.66667	100	100	100
SulHu	100	100	100	91.66667	83.3333333	100	100
SulPr	83.33	100	100	100	91.6666667	91.667	100
SulEv	100	100	100	100	100	100	100
ChwAT	100	91.667	91.667	91.66667	91.6666667	91.667	100
ChwHu	100	100	91.667	91.66667	100	100	100
ChwPr	91.67	91.667	83.333	100	91.6666667	91.667	100
ChwEv	91.67	91.667	91.667	91.66667	91.6666667	91.667	100
PenAT	83.33	100	91.667	91.66667	100	91.667	100
PenHu	66.67	66.667	83.333	75	83.3333333	83.333	100
PenPr	100	91.667	91.667	91.66667	100	100	100
PenEv	66.67	83.333	100	91.66667	100	91.667	100
Overall	90.28	92.361	93.75	92.36111	94.4444444	94.444	100

Table 7. Comparison between the percent of succeed in t-test for differences in monthly means of the generated and observed data for set 2 generated series, by each model.

	SVSS	SVMS	MVSS	**Matalas, 1967**	**Al-Suhili and Mustafa, 2013**	**Al-Suhili and Khanbilvardi, 2014**	The Developed Model
SulAT	100	100	100	100	100	100	100
SulHu	91.67	91.667	100	100	91.6666667	91.667	100
SulPr	100	100	100	100	100	91.667	100
SulEv	100	100	100	100	100	100	100
ChwAT	83.33	100	91.667	75	91.7	91.7	100
ChwHu	100	91.667	91.67	100	100	100	100
ChwPr	91.67	91.667	91.667	91.66667	91.6666667	91.667	100
ChwEv	91.67	91.667	91.667	83.33333	91.6666667	91.667	100
PenAT	100	100	100	91.66667	100	91.667	100
PenHu	66.67	66.667	75	91.66667	75	83.3	100
PenPr	100	100	100	91.66667	91.6666667	100	100
PenEv	100	91.667	91.667	91.66667	100	100	100
Overall	93.75	93.75	94.445	93.05556	94.4472222	94.444	100

Table 8. Comparison between the percent of succeed in t-test for differences in monthly means of the generated and observed data for set 3 generated series, by each model.

	SVSS	SVMS	MVSS	**Matalas, 1967**	**Al-Suhili and Mustafa, 2013**	**Al-Suhili and Khanbilvardi, 2014**	The Developed Model
SulAT	83.33	91.667	100	100	100	100	100
SulHu	100	91.667	83.333	100	83.3333333	91.667	100
SulPr	100	100	91.667	100	91.6666667	100	100
SulEv	100	100	100	100	100	100	100
ChwAT	100	100	91.667	91.66667	91.6666667	100	100
ChwHu	91.67	100	100	91.66667	91.6666667	91.667	100
ChwPr	91.67	100	91.667	91.66667	100	100	100
ChwEv	100	83.333	91.667	91.66667	91.6666667	91.667	100
PenAT	91.67	100	91.667	100	100	91.667	100
PenHu	75	66.667	66.667	75	75	75	100
PenPr	91.67	91.667	91.667	91.66667	91.6666667	91.667	100
PenEv	100	100	83.333	83.33333	100	100	100
Overall	93.75	93.75	90.278	93.05556	93.0555556	94.444	100

Table 9. Comparison between the AIC test for the three generated series by each model.

	SulAT	SulHu	SulPr	SulEv	ChwAT	ChwHu	ChwPr	ChwEv	PenAT	PenHu	PenPr	PenEv
-SVSS	60.04	183	54.89	2.104	58.31	165.55	65.53	14.39	63.32	143.64	95.86	6.103
-SVMS	42.13	185	57.73	14.42	70.01	186.98	99.93	4.1	57.76	152.49	91.31	13.22
-MVSS	66.97	158	58.37	7.095	71.31	169.69	51.35	-1.246	60.95	157.9	130.7	-8.323
Matals, 1967	47.22	162	52.12	11.61	59.34	155.12	50.15	16.89	59.67	148.91	82.56	-10.56
Al-Suhili and Mustafa, 2013	45.02	158	39.68	-1.38	55.91	153.2	47.48	-13.86	53.88	143.23	62.2	-11.74
Al-Suhili and Khanbilvardi, 2014	36.43	144	36.58	-5.35	48.84	144.92	48.62	-15.76	46.57	126.39	61.89	-20.55
The Developed Model	26.34	136	**38.14**	-7.34	42.67	137.56	35.67	-19.87	39.67	111.14	55.67	-21.69
-SSSV	54.38	163	50.84	10.68	79.51	187.49	55.09	-19.04	61.29	153.73	91.79	19.75
-MSSV	73.38	168	72.78	13.97	62.9	173.26	63.21	6.373	67.8	173.56	101.3	15.52
-MVSS	48.31	190	62.51	6.358	81.06	169.01	66.25	-3.124	61.9	142.18	95.18	29.9
Matals, 1967	42.67	157.8	47.45	4.56	59.67	153.67	49.15	11.69	54.15	139.45	78.34	26.56
Al-Suhili and Mustafa, 2013	40.6	153	33.71	-3.7	53.7	148.72	46.15	-12.47	50.31	137.7	59.68	-19.29
Al-Suhili and Khanbilvardi, 2014	35.09	150	36.09	-3.21	51.71	146.4	46.3	-9.43	45.59	127.45	63.46	-14.18
The Developed Model	33.21	143.16	29.87	**-1.1**	45.67	122.56	36.89	-14.78	41.34	121.78	58.99	-22.45
SSSV	46.43	145	63.35	17.81	64.45	171.17	110.7	22.29	56.22	164.29	107.7	-5.484
MSSV	48.88	165	67.72	11.97	69.84	177.03	86.33	29.87	70.4	149.45	63.05	20.53
MVSS	48.96	167	45.03	25.9	81.77	160.22	87.54	-1.396	69.7	153.59	98.92	28.11
Matals, 1967	45.91	143	44.54	19.87	58.67	157.78	76.57	-10.12	56.12	143.56	81.23	22.13
Al-Suhili and Mustafa, 2013	43.53	150	42.85	-6.46	57.58	147.49	55.41	-16.01	55.76	132.87	67.45	-14.84
Al-Suhili and Khanbilvardi, 2014	41.61	144	35.67	-2.82	57.4	142.15	45.95	-11.54	51.3	117.14	62.19	-15.31
The Developed Model	33.54	121.5	32.67	**-4.32**	48.78	139.98	36.78	-18.92	46.78	113.56	58.75	-17.23