American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-03, Issue-06, pp-154-168
www.ajer.org

Research Paper Open Access

A Multi-Variables Multi -Sites Hydrological Forecasting Model Using Relative Correlations

Prof. Dr. Rafa H Al-Suhili[2] and Prof. Dr. Reza Khanbilvardi[2]

1 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.

2 Director , NOAA-Cooperative Remote Sensing Science and Technology Center Director International Center for Environmental , Resources and Development Prof. of Civil Engineering Dept, City College of New York, New York, USA

Abstract: **-** A multi-variables multi-sites hydrological data forecasting model was derived and checked using a case study. The philosophy of this model is to use the cross-variables correlations, cross-sites correlations and the time lag correlations simultaneously. 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. This model represents a modification of the model proposed by Al-Suhili and Mustafa(2013). The model performance was compared with four well known forecasting models developed for the same data. These models are the single-site single-variable first order auto regressive, the multi-variables single-site Matalas(1967), the single-variable multi-sites Matalas(1967), and Al-Suhili and Mustafa models. In addition to that another multi-variables multi-sites model was developed herein similar in its concept to the Matalas(1967) model considering the variables as an additional sites. The results of the six models for three forecasted series for each variable, were compared using the Akaike test which indicates that the developed model is more successful, since it gave the minimum (AIC) values for (83.33 %) of the forecasted series. This indicates that the developed model had improved the forecasting performance. Moreover the t-test for monthly means comparison between the models indicates that the developed model has the highest percentage of succeed (94.44%).

Keywords: **-** *Forecasting, Multi-sites, Multi-variables, Cross sites correlation, Serial correlation, Cross variables correlations, Hydrology,*

I. 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 changes 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 in time. This allows developing various applications linked to extreme events, such as flood analyses, and draught analysis, hence allowing 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 may 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". Matals(1967), had developed the most well known multi-sites model using cross site correlations between one variable at different sites. This model can be applied as a multi-variable model that uses multi variables cross correlation in a given site. Richardson (1981) had proposed a multi-variables stochastic weather models for daily precipitation, maximum temperature, minimum temperature, and solar radiation, as cited in Wilks (1999). This model forecast a hydrological variable at multiple sites, 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. There exist in the literature some relatively recent 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.

In real situation both cross variables and cross sites correlation may exist between different hydrological variables at different sites. Al-Suhili et al(2010) had presented a multisite multivariate model for forecasting different water demand types at different areas in the city of Karkouk, north Iraq. This model first relate 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 a multisite Matalas(1967) models for each type of demand. These models were then coupled with the regression equation to simulate the mult-isites multi-variables variation. 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 and cross variables 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. Moreover the model was applied for only eight months of the year (October to May) excluding these months of zero precipitation values.

In this research a modified multi-sites multi-variables 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 Mustafa model(2013). The modification is done such that the total weights of the weighted components 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 variable at the preceding time step. This was done by adopting a different method for estimating the parameters of the model. Moreover the model was applied for all the months of the year that includes zero values for the precipitation. The problem was overcome by adding a constant value of 0.1 to the whole precipitation data series, to investigate whether this modification can solve this problem of zero values. This model was applied to a case study of monthly data of four hydrological variables, air temperature, humidity, precipitation and evaporation at three sites located north Iraq, Sulaimania, Chwarta, and Penjwin.

II. THE MODEL DEVELOPMENT

The multivariate multisite model developed herein, utilizes single variable time lag one correlations, cross variables lag-one correlations, and cross sites lag-one correlations. In order to illustrate the model derivation consider figure 1a where the concept of the model is shown. 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, figure 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, figure 1 shows a schematic representation of the developed 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 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 time (t-1) at all sites. The expression is weighted by the 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 first 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 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 O=1, can be represented by the following equation:

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

Where :

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

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

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

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

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

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

$\rho_{1,5} = \rho$ [(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

$\rho_{1,6} = \rho$ [(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

$\rho_{6,6} = \rho$ [(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 designation ($\rho_{i,j}$) is used for simplification .

ϵ: is the stochastic dependent component.

ξ: is the stochastic independent component.

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

The matrix equation (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:

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

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

$$\boldsymbol{\sigma_{i,j} = (1 - \rho_{i,j}^2\,)^{0.5}} \quad (10)$$

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

$$\boldsymbol{\sigma r_{i,j}} = \frac{\sigma_{i,j}}{\sum_{j=1}^{n=v*s} abs(\rho_{i,j}+\sigma_{i,j})} \quad (11)$$

Then the model matrix equation is the same as that appear in eq.(2), replacing $\rho_{i,j}$ values by the corresponding relative values $\boldsymbol{\rho r_{i,j}}$ in equation (6), and $\sigma_{i,j}$ with the corresponding relative values $\sigma r_{i,j}$ in equation (7) . The differences of this model than that proposed by Al-Suhili and Mustafa(2013), are in eqs(9) and (11), where for the first equation the denominator is the sum of $\rho_{i,j}$, only , while for the second equation it is the sum of $\sigma_{i,j}$, only. The model can be generalized to any number of variables and number of sites.

III. THE CASE STUDY AND APPLICATION OF THE MODEL:

In order to apply the 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 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. Figure 2 shows a Google map of the locations of these stations. Table 2 shows the approximate distances between these stations.

The model was applied to the data of the case study described above. The 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 model building, 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. Hence the model was built for the all of the months from January to December, rather than for October to May as proposed by Al-Suhili and Mustafa(2013).

The first step of the modeling process is to check the homogeneity of the data series. 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 . The data sample was divided into two subsamples with sizes (n1=5,and n2=3) as number of years for subsample one and subsample two respectively. The split sample test estimated t-values were 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, and thus this non-homogeneity should be removed. The results of this test had showed that there are some variables exhibits non- homogeneity. Tables 3 and 4 show these results, which indicates that non-homogeneity is exist in each of Sulaimania air temperature, Penjwin humidity, Penjwin air temperature, and Penjwin evaporation series, while the series of the other variables are homogeneous. To remove this non-homogeneity the method suggested by Yevjevich (1972), was used by applying the following data transformation to the series of the non-homogeneous variables for the n1 years:

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

Where,

$H_{i,j}$: is the homogenized series at year i, month j of the first sub-sample (old n1 series).

$X_{i,j}$: 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(recent n2 series). This implies that the data is homogenized according to the second sub-sample, i.e., the most recent one which is the correct way for forecasting. Table 5 shows the values of Mean2, Sd2, A1, B1, A2, and B2, for the non-homogeneous series. Tables 6 and 7 show the results of the split sample test after the application of equation (12), which ensures that the data series are all now homogeneous.

The second step in the modeling process is to check and remove the trend component in the data if it is 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 larger than the critical t-value then trend exists, otherwise it is not. The following equation was used to estimate the t-values.

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

Where

n: is the total size of the sample.

Table 8 shows the trend test results, which indicate the absence of the trend component in all of the data series of the twelve variables.

The third step of the modeling process is the data normalization of the data 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 of the transformation.

α : is the shifting parameter.

XN : is the normalized series.

Table 9 shows the coefficients of the normalization transformation of all of the twelve series. A shifting parameter of constant value 5 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 the one that reduce the skewness to almost zero value. However it was found that for the precipitation series in the three locations the required normalization transformation is of high negative value of order less than -4. If these values were selected the series is transformed to values that numerically differ after the 6 digits beyond the point, which means a very high accuracy is needed to perform the analysis which is not assured even if the long format of the Matlab software is used. Hence, for the precipitation series the minimum power value obtained among the other variables was used (-0.55). This transformation power will not let the skewness of the precipitation reduces to nearly zero as required, but at least reduce this skewness to values that are much smaller than the skewness of the homogenized series. Table 11 shows that the skewness coefficients are reduced to almost zeros for the data series, with an exception of the precipitation , which have skewness values less than 1.

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

$$\epsilon_{i,j}=\frac{XN_{i,j}-Xb_j}{Sd_j} \tag{15}$$

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.

The existence of the periodic components is detected by drawing the corrlogram up to at least 25 lags, if the curve exhibits periodicity then the periodic components are exist, otherwise it is not. Figure 3 shows the correlograms of the normalized data, where the periodic component is clear. Figure 4 shows the correlograms of the dependent stochastic component, which indicates the removal of the periodic components.

The fifth 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 serial and cross correlation coefficients $\rho_{i,j}$, and $\sigma_{i,j}$ of matrix eqs. (6) and (7) respectively, which then used to estimate the model parameters $\rho r_{i,j}$ and $\sigma r_{i,j}$ using eqs.(9), and (11), respectively.

For the sake of comparison between the developed model and the known forecasting models in the literature, five types of forecasting models were developed for the same data of the case study. For each variable a single variable single site first order autoregressive model (8 models), for each site multi-variables single site first order Matalas model(3 models), for each variable a single variable multi-sites first order Matalas model (4 models) ,Al-Suhili and Mustafa multi-variables multi-sites model(1 model), and a multi-variables multi-sites Matalas model(1 model).

IV. FORECASTING RESULTS AND DISCUSSION

The developed models mentioned above are 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:

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) 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 using the t-test .

3. Checking the performance of the model of the hole forecasted series using Akaike test.

The Akaike test can be used also for the purpose of comparison of the forecasting performance between the new multi-variables multi-sites model developed herein and the other models. This performance comparison was made to investigate whether the new model can produce better forecasted data series. For this purpose the Akaike (AIC), test given by the following equation was used:

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

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, using the six different models mentioned above. The overall statistical characteristics are compared with those observed, for each of the generated series. It is observed that the six 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 10,11 and 12 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 four 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. However the overall succeed percent given by the Al-suhili and Mustafa model is almost similar to that given by the developed model.

As mentioned above for purpose of the comparison between the developed model performance and that of the available forecasting models and developed for the data as mentioned above, the Akaike(1974) test was used. Table 13, shows the Akaike test results for all of the forecasted variables, in each site, obtained using the developed five 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 (83.33%). Al-Suhili and Mustafa model had gave the lowest test value for the remaining cases (16.66%). However for these cases 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.

Figure 5 shows comparisons between the observed and the generated series using the developed model for the whole three years and between the monthly means of these two series. This figure indicates the capability of the model for forecasting the future variation of all of the variables.

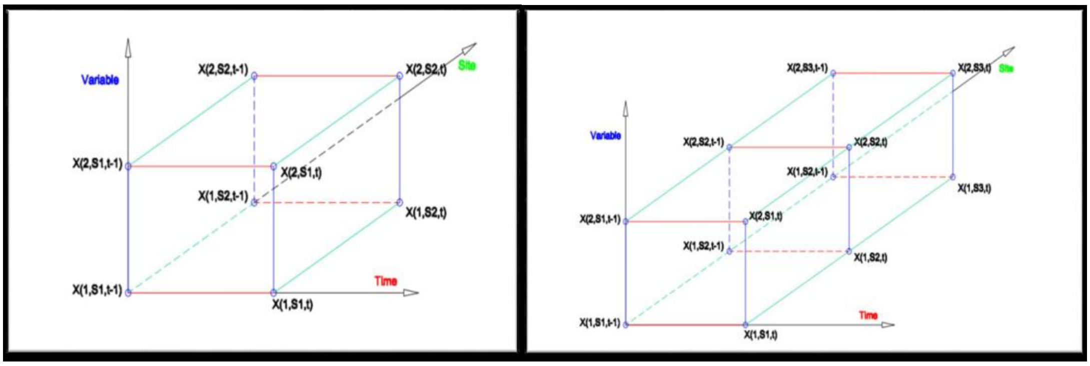
V. CONCLUSIONS

From the analysis done in this research, the following conclusion could be deduced:
The model parameters can be easily estimated and do not require any extensive mathematical manipulation.
The model can preserve the overall statistical properties of the observed series with high accuracy. However this is also observed for the other five models developed for the same variables.
The model can preserve the monthly means of the observed series with excellent accuracy, evaluated using the t-test with overall success (94.4%). This percent is almost the highest among the those obtained by the other model, except that the Al-Suhili and Mustafa model(2013), had presented a very close values.
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(83.33%). Moreover for those remaining cases where Al-suhaili and Mustafa(2013) model had the better performance(minimum AIC value); the test value of the developed model is slightly higher than this minimum value.

REFERENCES

[1] Al-Suhili, R. H. and Mustafa, N.F. " A Multi- variables Multi-sites model for
[2] Hydrological Series Data Forecasting", accepted for publication at the Journal of Eng. College of Engineering, University of Baghdad, 2003.
[3] 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.
[4] Akaike, H., 1974, “A New Look at the Statistical Model Identification”, IEEE T. Automat. Contr., 19 (6), pp716–723.

[5] 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.
[6] Box, G.E., and Jenkins, G. M. (1976),"Time Series Analysis and Control", San Francisco, California: Holden-Day, Inc.
[7] 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: pp229-247.
[8] 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
[9] 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
[10] Matalas N.C., 1967,” Mathematical Assessment of Synthetic hydrology”, journal of Water Resources Researches vol. 3: pp 937-945.
[11] 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 .
[12] Tobler W., 1970 "A computer Movie Simulating Urban Growth in the Detroit Region". Economic Geography, 46(2): pp234-240.
[13] 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.
[14] Yevjevich, V. M., 1972 "The structure of Hydrologic Time Series", Fort Collins, Colorado State University



(a) (b)

Fig. 1 Schematic representation of the developed multi-variables multi-sites model, a)MVMS(2,2,1), b) MVMS(2,3,1).

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} $42^{'}$ $23^{''}$

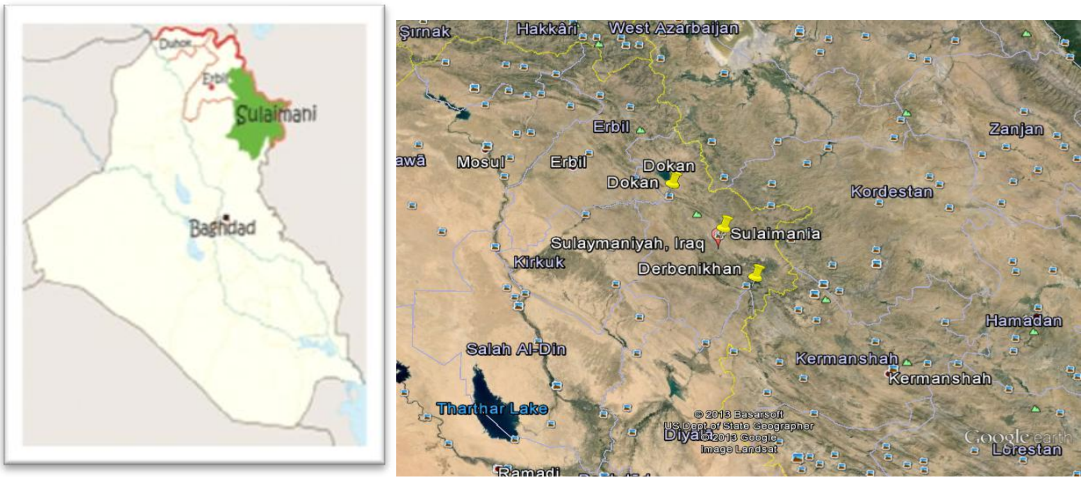


Fig. 2 Locations of the metrological stations selected for analysis.

Table 2 Approximate distances between the Sulaimania weather stations network (Km.).

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 Test of homogeneity of the original data in mean, n1=5,n2=3.

	Mean1	Mean2	s1	s2	s	t-test	Case
SulAT	19.9008	20.1485	0.4507	1.18751	0.778	-0.436	Hom.
SulHu	46.9446	45.8553	2.4575	1.46515	2.178	0.68499	Hom.
SulPr	1.69637	1.66298	0.4676	0.13639	0.39	0.11729	Hom.
SulEv	5.52551	5.23813	0.1015	0.45864	0.277	1.41827	Hom.
ChwAT	16.4929	17.1897	0.5596	1.29252	0.875	-1.0904	Hom.
ChwHu	49.6663	46.9338	3.4431	1.90116	3.018	1.23977	Hom.
ChwPr	1.9079	1.89837	0.6068	0.07698	0.497	0.02625	Hom.
ChwEv	5.60861	5.28657	0.4906	0.3626	0.452	0.97565	Hom.
PenAT	13.9755	13.5897	0.7737	1.14588	0.915	0.57744	Hom.
PenHu	63.3525	52.6272	6.9884	4.69652	6.318	2.32467	NonHom.
PenPr	2.75305	2.66979	0.9516	0.0907	0.779	0.1464	Hom.
PenEv	5.46681	4.62648	0.4845	0.25017	0.421	2.7322	NonHom.

Table 4 Test of homogeneity of the original data in standard deviation, n1=5,n2=3.

	Mean1	Mean2	s1	s2	s	t-test	case
SulAT	10.1889	9.52285	0.2866	0.47757	0.362	2.52179	NonHom.
SulHu	18.7659	17.4115	2.5053	0.99604	2.125	0.8728	Hom.
SulPr	2.18701	1.86037	0.6365	0.28427	0.545	0.82064	Hom.
SulEv	3.69204	3.52068	0.21	0.36782	0.273	0.85961	Hom.
ChwAT	10.5129	9.65657	0.5551	0.6469	0.587	1.99648	Hom.
ChwHu	16.0407	15.7313	2.1606	1.65751	2.007	0.21105	Hom.
ChwPr	2.43547	2.12283	0.8342	0.23539	0.695	0.61639	Hom.
ChwEv	3.8826	3.59904	0.2506	0.02954	0.205	1.89074	Hom.
PenAT	11.3524	9.38528	1.094	0.84784	1.019	2.64447	NonHom.
PenHu	13.033	14.0986	1.492	3.44251	2.331	-0.6259	Hom.
PenPr	3.50955	2.82328	1.3632	0.34376	1.131	0.83116	Hom.
PenEv	4.10278	3.75387	0.5069	0.2391	0.436	1.09505	Hom.

Table 5 Linear fitting equations for removal of non-homogeneity.

	A1	B1	R1	A2	B2	R2
SulAT	19.534	0.102	0.342	10.22	-0.062	0.318
SulHu	48.102	-0.348	0.407	20.859	-0.578	0.678
SulPr	1.904	-0.049	0.333	2.481	-0.092	0.425
SulEv	5.601	-0.041	0.336	3.817	-0.042	0.385
ChwAT	15.886	0.193	0.533	10.522	-0.073	0.256
ChwHu	51.877	-0.719	0.562	17.164	-0.275	0.362
ChwPr	2.178	-0.061	0.325	2.834	-0.114	0.422
ChwEv	5.529	-0.009	0.05	3.843	-0.015	0.151
PenAT	14.562	-0.162	0.457	12.504	-0.42	0.741
PenHu	72.717	-2.975	0.904	13.267	0.037	0.04
PenPr	3.213	-0.109	0.371	4.238	-0.219	0.485
PenEv	5.63	-0.106	0.446	3.993	-0.005	0.026

Table 6 Test of homogeneity of the homogenized data in mean, n1=5,n2=3.

	Mean1	Mean2	s1	s2	s	t-test	Case
SulAT	20.207998	20.1485	0.2963	1.18751	0.727	0.112	Hom.
SulHu	46.944628	45.8553	2.4575	1.46515	2.1775	0.68499	Hom.
SulPr	1.6963688	1.66298	0.4676	0.13639	0.3899	0.11729	Hom.
SulEv	5.5255085	5.23813	0.1015	0.45864	0.2775	1.41827	Hom.
ChwAT	16.492926	17.1897	0.5596	1.29252	0.875	-1.0904	Hom.
ChwHu	49.666252	46.9338	3.4431	1.90116	3.018	1.23977	Hom.
ChwPr	1.9079033	1.89837	0.6068	0.07698	0.4974	0.02625	Hom.
ChwEv	5.6086085	5.28657	0.4906	0.3626	0.452	0.97565	Hom.
PenAT	13.492285	13.5897	0.5189	1.14588	0.7856	-0.1698	Hom.
PenHu	52.170046	52.6272	3.5289	4.69652	3.9566	-0.1582	Hom.
PenPr	2.7530521	2.66979	0.9516	0.0907	0.7787	0.1464	Hom.
PenEv	4.7733789	4.62648	0.584	0.25017	0.4982	0.40373	Hom.

Table 7 Test of homogeneity of the homogenized data in standard deviation, n1=5,n2=3.

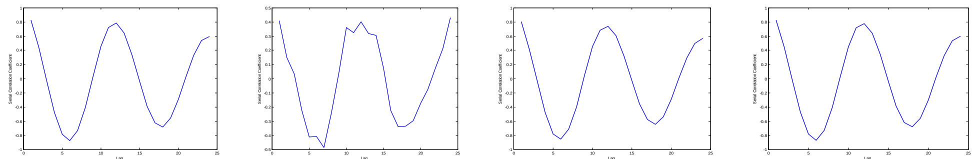
	Mean1	Mean2	s1	s2	s	t-test	case
SulAT	9.6723865	9.52285	0.3562	0.47757	0.4008	0.51093	Hom.
SulHu	18.76593	17.4115	2.5053	0.99604	2.1249	0.8728	Hom.
SulPr	2.1870135	1.86037	0.6365	0.28427	0.545	0.82064	Hom.
SulEv	3.6920379	3.52068	0.21	0.36782	0.273	0.85961	Hom.
ChwAT	10.512893	9.65657	0.5551	0.6469	0.5873	1.99648	Hom.
ChwHu	16.040672	15.7313	2.1606	1.65751	2.007	0.21105	Hom.
ChwPr	2.4354703	2.12283	0.8342	0.23539	0.6945	0.61639	Hom.
ChwEv	3.882598	3.59904	0.2506	0.02954	0.2054	1.89074	Hom.
PenAT	9.4755896	9.38528	0.7297	0.84784	0.7711	0.16037	Hom.
PenHu	13.737973	14.0986	1.6012	3.44251	2.379	-0.2075	Hom.
PenPr	3.5095495	2.82328	1.3632	0.34376	1.1306	0.83116	Hom.
PenEv	3.8723131	3.75387	0.4844	0.2391	0.4189	0.38718	Hom.

Table 8 Trend detection test after removing non-homogeneity.

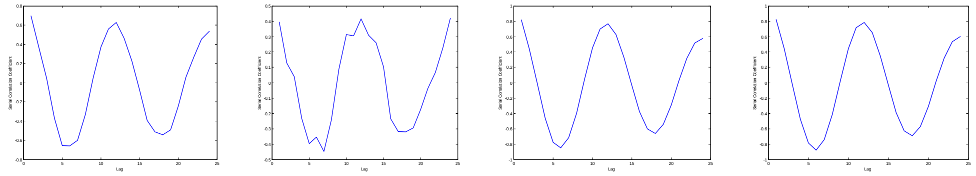
	r	t
SulAT	0.0766045	0.1882
SulHu	-0.407047	-1.0916
SulPr	-0.333184	-0.8656
SulEv	-0.336426	-0.8751
ChwAT	0.5332421	1.54401
ChwHu	-0.562345	-1.6658
ChwPr	-0.325523	-0.8433
ChwEv	-0.049619	-0.1217
PenAT	-0.109125	-0.2689
PenHu	-0.187953	-0.4687
PenPr	-0.370671	-0.9776
PenEv	0.2005191	0.50135

Table 9 Normalization transformation power, and skewness for data (2004-2008).

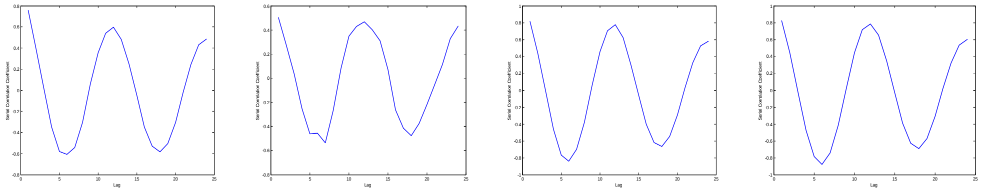
	Power	Skewness
SulAt	1.1	0.007658458
SulHu	0.9	0.008000455
SulPr	-0.55	0.756794614
SulEv	-0.55	0.001047691
ChwAt	1	0.000888238
ChwHu	1	0.001705557
ChwPr	-0.55	0.749611239
ChwEv	-0.35	-0.005834051
PenAT	1.05	-0.003692929
PenHu	1.1	-6.97121E-05
PenPr	-0.55	0.62709928



a1 **a2** **a3** **a4**

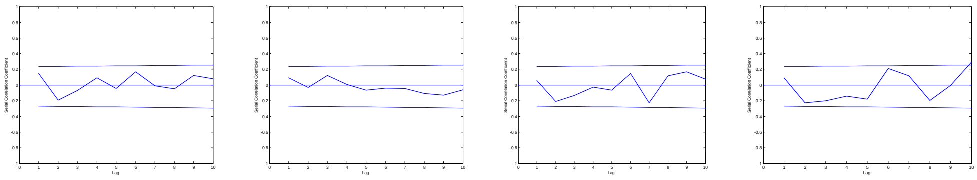


b1 b2 b3 b4

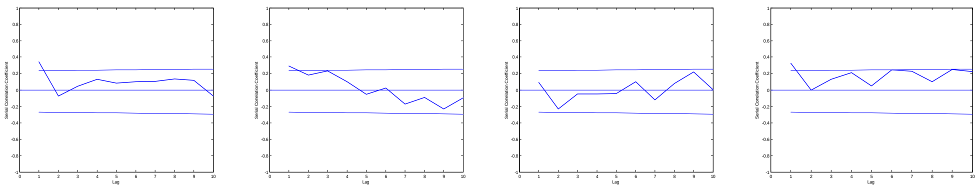


c1 c2 c3 c4

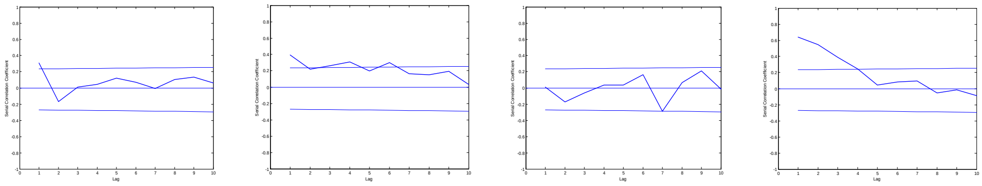
Fig. 3 Correlograms of the normalized data series,a) Sulaimania, b) Chwarta, c) Penjwin, 1) Air temperature,2) Humidity,3) Precipitation,4) Evaporation.



a1 a2 a3 a4



b1 b2 b3 b4



c1 c2 c3 c4

Fig. 4 Correlograms of the dependent stochastic series, a) Sulaimania, b) Chwarta, c) Penjwin, 1) Air temperature,2) Humidity,3) Precipitation,4) Evaporation.

Table 10 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.

	SS	MSSV	MVSS	Al-Suhili and Mustafa	Matalas MVMS	MVMS
SulAT	100	91.667	100	100	91.66666667	100
SulHu	100	100	100	83.3333333	91.66666667	100
SulPr	83.33	100	100	91.6666667	100	91.667
SulEv	100	100	100	100	100	100
ChwAT	100	91.667	91.667	91.6666667	91.66666667	91.667
ChwHu	100	100	91.667	100	91.66666667	100

ChwPr	91.67	91.667	83.333	91.6666667	100	91.667
ChwEv	91.67	91.667	91.667	91.6666667	91.66666667	91.667
PenAT	83.33	100	91.667	100	91.66666667	91.667
PenHu	66.67	66.667	83.333	83.3333333	75	83.333
PenPr	100	91.667	91.667	100	91.66666667	100
PenEv	66.67	83.333	100	100	91.66666667	91.667
Overall	90.28	92.361	93.75	94.4444444	92.36111111	94.444

Table 11 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.

	SS	MSSV	MVSS	Al-Suhili and Mustafa	Matalas MVMS	MVMS
SulAT	100	100	100	100	100	100
SulHu	91.67	91.667	100	91.6666667	100	91.667
SulPr	100	100	100	100	100	91.667
SulEv	100	100	100	100	100	100
ChwAT	83.33	100	91.667	91.7	75	91.7
ChwHu	100	91.667	91.67	100	100	100
ChwPr	91.67	91.667	91.667	91.6666667	91.66666667	91.667
ChwEv	91.67	91.667	91.667	91.6666667	83.33333333	91.667
PenAT	100	100	100	100	91.66666667	91.667
PenHu	66.67	66.667	75	75	91.66666667	83.3
PenPr	100	100	100	91.6666667	91.66666667	100
PenEv	100	91.667	91.667	100	91.66666667	100
Overall	93.75	93.75	94.445	94.4472222	93.05555556	94.444

Table 12 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.

	SS	MSSV	MVSS	Al-Suhili and Mustafa	Matalas MVMS	MVMS
SulAT	83.33	91.667	100	100	100	100
SulHu	100	91.667	83.333	83.3333333	100	91.667
SulPr	100	100	91.667	91.6666667	100	100
SulEv	100	100	100	100	100	100
ChwAT	100	100	91.667	91.6666667	91.66666667	100
ChwHu	91.67	100	100	91.6666667	91.66666667	91.667
ChwPr	91.67	100	91.667	100	91.66666667	100
ChwEv	100	83.333	91.667	91.6666667	91.66666667	91.667
PenAT	91.67	100	91.667	100	100	91.667
PenHu	75	66.667	66.667	75	75	75
PenPr	91.67	91.667	91.667	91.6666667	91.66666667	91.667
PenEv	100	100	83.333	100	83.33333333	100
Overall	93.75	93.75	90.278	93.0555556	93.05555556	94.444

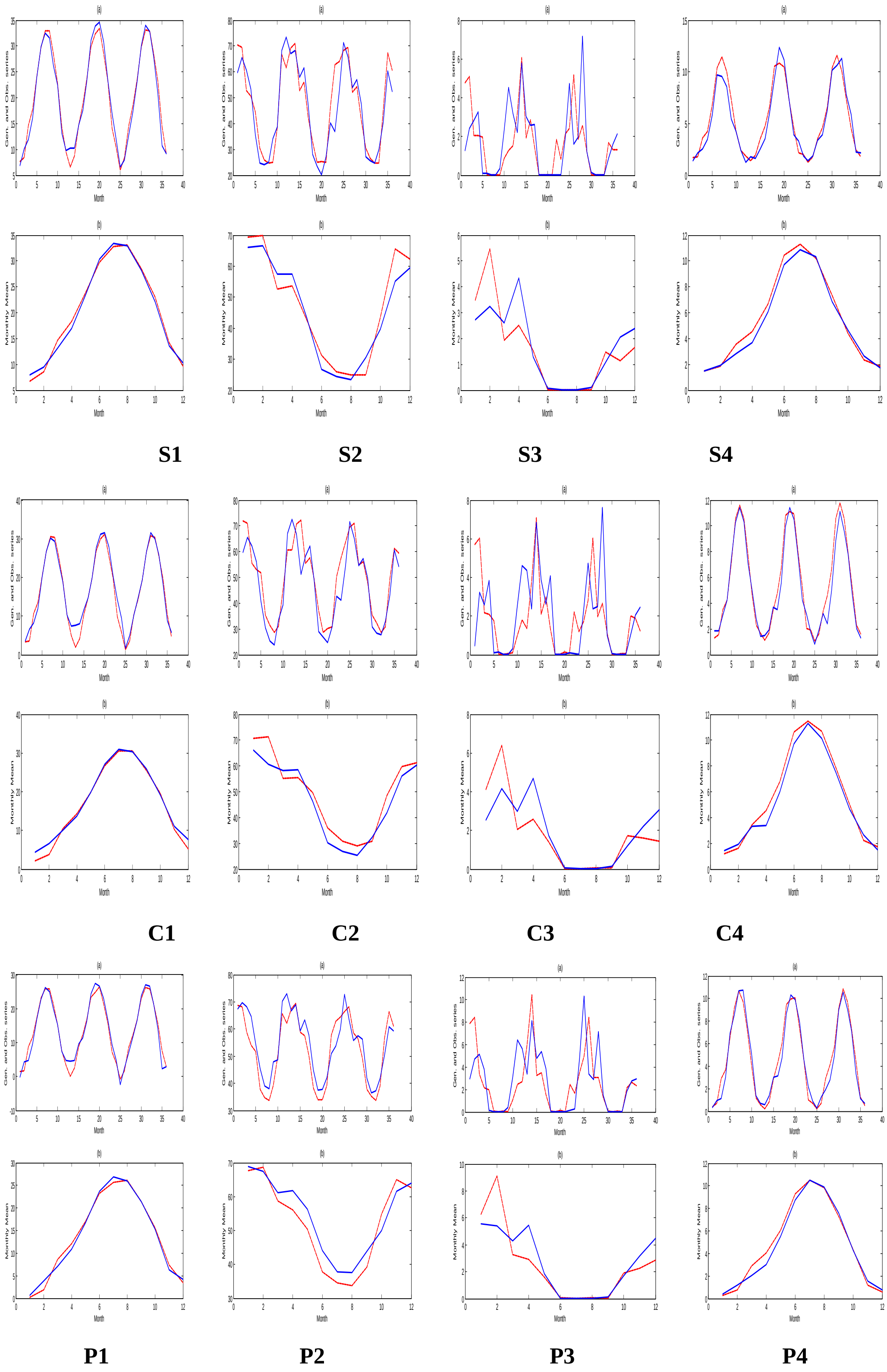


Fig. 5 Comparison between observed and forecasted series(2009-2011), S:Sulaimania,C:Chwarta,P:Penjwin,1:Airtemperature, 2:Humidity, 3:Pecipitatio-n Evaporation, a:Three years series, b:Monthly means.

Table 13. 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
Set 1	SS	60.04	183	54.89	2.104	58.31	165.55	65.53	14.39	63.32	143.64	95.86	6.103
	MSSV	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
	Al-Suhili and Mustafa	45.02	158	39.68	-1.38	55.91	153.2	47.48	-13.86	53.88	143.23	62.2	-11.74
	MVMS	36.43	144	36.58	-5.35	48.84	144.92	48.62	-15.76	46.57	126.39	61.89	-20.55
Set 2	SS	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
	Al-Suhili and Mustafa	40.6	153	33.71	-3.7	53.7	148.72	46.15	-12.47	50.31	137.7	59.68	-19.29
	MVMS	35.09	150	36.09	-3.21	51.71	146.4	46.3	-9.43	45.59	127.45	63.46	-14.18
Set 3	SS	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
	Al-Suhili and Mustafa	43.53	150	42.85	-6.46	57.58	147.49	55.41	-16.01	55.76	132.87	67.45	-14.84
	MVMS	41.61	144	35.67	-2.82	57.4	142.15	45.95	-11.54	51.3	117.14	62.19	-15.31