Estimation the Missing Data of Meteorological Variables In Different Iraqi Cities By using ARIMA Model

Sura T. Nassir*, Ahmed B. Khamees, Wassan T. Mousa

Renewable Energy Directorate, Ministry of Science and Technology, Baghdad, Iraq.

Abstract

In this paper, ARIMA model was used for Estimating the missing data(air temperature, relative humidity, wind speed) for mean monthly variables in different time series at three stations (Sinjar, Baghdad , AL.Hai) which represented different parts of Iraq from north to south respectively.

Keywords: Meteorological Data, ARIMA Model.

تخمين البيانات المفقودة للمتغيرات الانوائية في مناطق مختلفة من العراق بأستخدام نموذج اريما

سرى ثامر ناصر*، احمد بدر خميس، وسن ثامر موسى

دائرة الطاقات المتجددة، وزارة العلوم والتكنولوجيا، بغداد، العراق.

الخلاصة

في هذا البحث استخدم نموذج اريما لتخمين البيانات المفقودة (درجة حرارة الهواء و الرطوبة النسبية و سرعة الرياح) للمتغيرات الشهرية المعدلة لسلاسل زمنية مختلفة عند ثلاث محطات (سنجار وبغداد والحي) التي تمثل اجزاء مختلفة من العراق من الشمال الى الجنوب على التوالي.

1. Introduction

The availability of meteorological data is very important in many applications, especially in renewable energy projects, where suitable locations are selected for the placement of wind turbines or solar panels. Often there are missing data in meteorological databases due to various reasons. It is therefore very important to be able to predict these data (forecasting) and to fill missing data values from databases. In almost all research fields, the procedure for handling missing values must be addressed before a detailed analysis can be made. Thus, a suitable method of imputation should be chosen to address the missing value problem. However, researchers have sometimes faced with the problem of missing data caused by equipment failure, malfunctioning equipment, terrible weather,and incorrect data recording [1]

2. Method of Calculation of Missing Data

Missing values are an unavoidable problem in many applications. In fact, the missing values are a more common and serious problem in dealing with prediction. In this work, we adopt an approach based on ARIMA model to recover the missing value of meteorological variables [2].There are many methods available to estimate missing data, most of these methods can also be applied in many different fields, such as meteorology, astronomy …etc. From these models: Autoregressive integrated moving average (ARIMA) is the most popular linear model for forecasting time series and missing data during the last three decades[3]. The linear function is based upon three parametric linear components: autoregressive (AR), integration (I), and moving average (MA)[4]. In this study, the

---

*Email: sura_thamir91@gmail.com

meteorological variables have missing data for several months within time series. The ARIMA model Figure-1.
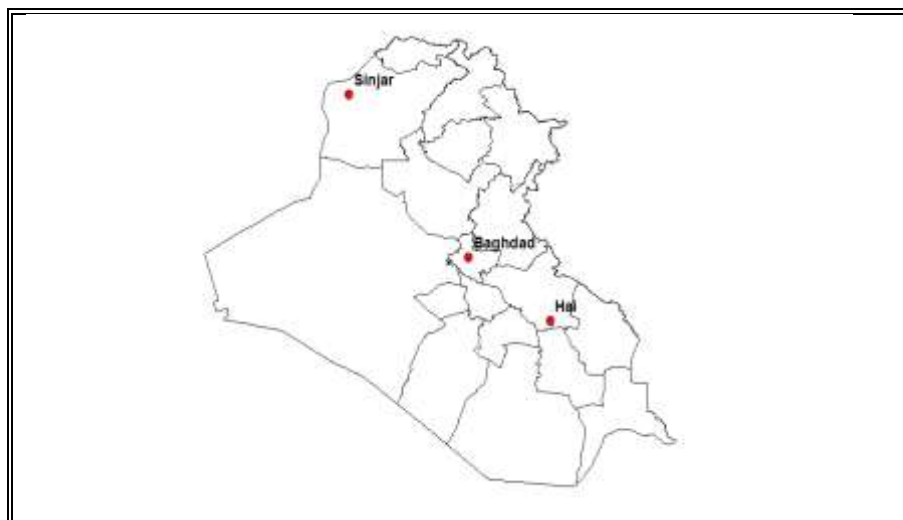
Figure 1- show map of Iraq and the location of the stations.

The missing data have been found in three cities of Iraq (Sinjar, Baghdad, and AL-HAI)   using ARIMA model for three variables (mean air temperature, mean relative humidity and mean wind speed) for different time series**.**

4. Calculation and Results
   Among of all the possible tests which can be result from using ARIMA model with different p, I, and q values, only one of them have been chosen because the missing data results close to the observed data of meteorological variables. The results show in detail below:
4.1 Sinjar Station
   Time series of Sinjar city for each of the wind speed, air temperature and relative humidity began from 1993 to 2008, the missing data for air temperature were found in all months of 2003 only. The relative humidity was missed values at all the months of 2001, one month from 2002, 11 months from 2003 and 7 months from 2004 while the missing data of the wind speed were observed at one month from 1997, all the months of 2003 and 7 months of 2004 as shown below:
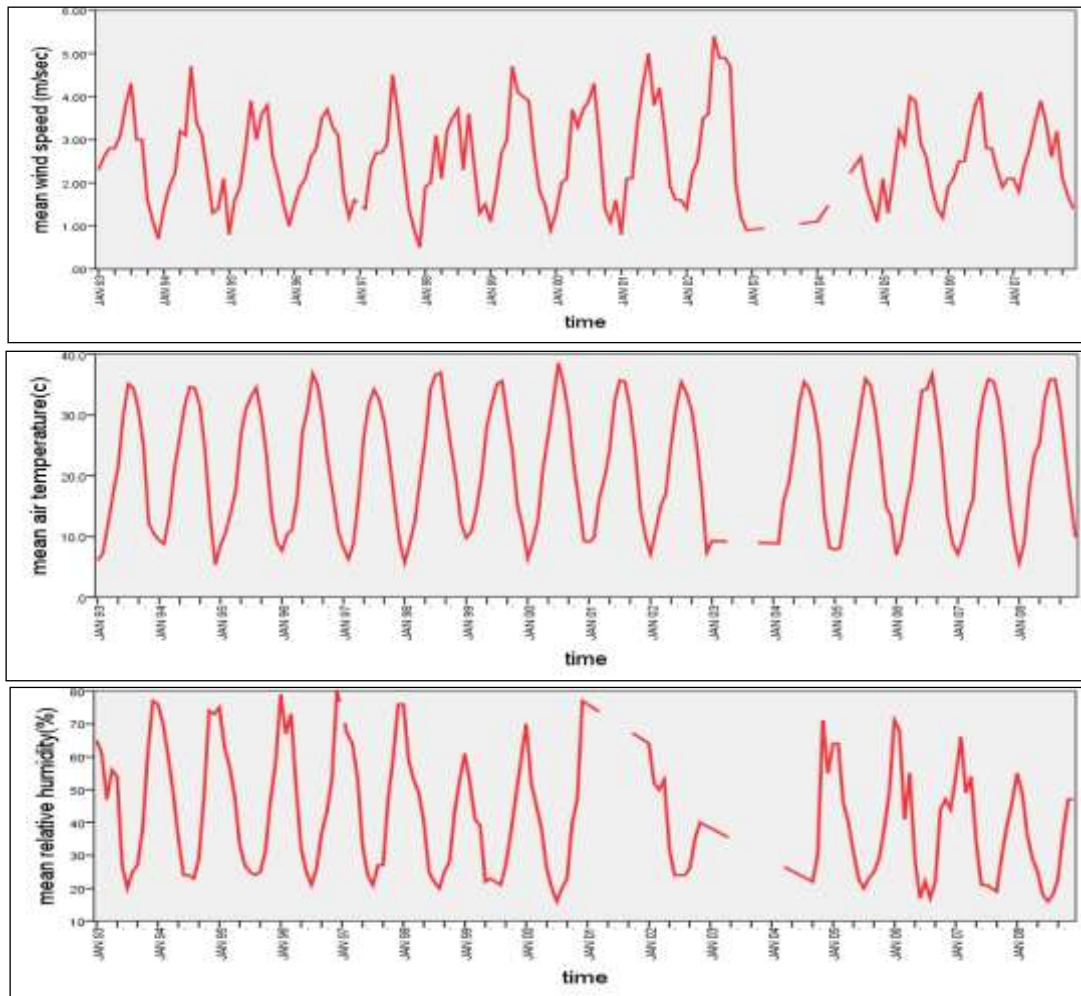
Figure 2- time series of the wind speed, air temperature and rh for sinjar station.

Choosing the correct results for the missing values, was used first the autocorrelation function (ACF) and partial autocorrelation function (PACF) and that shows there is no clear correlation within residuals of the three variables because all values within the 95% confidence limits, This means that ARIMA models which have been selected are the best models for the city of Sinjar, as shown in the following Figure
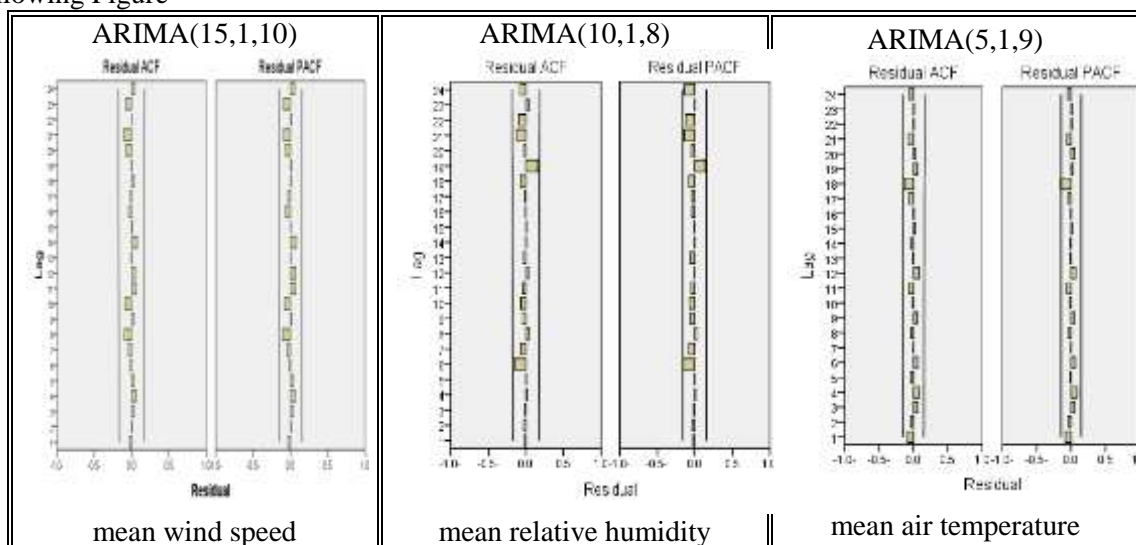

Figure 3-ACF and PACF with 95% confidence limits for three Models.

There are many criteria which can be used to detect the predicting results in the missing data with minimum errors. The predicting errors are related to the deference's between the actual variables and the predicted variables which obtained from the forecasting models, Such as mean average error (MAE), root mean square error (RMSE), R-squared and mean absolute percentage error (MAPE). The resulting values for these criteria for the three predicted variables are shown in the Table below:

**Table 1-**Statistical parameters of the selected ARIMA models.

Fit Statistic	Mean Wind Speeds	Mean Relative Humidity	Mean Air Temperature
R-squared	0.861	0.990	0.973
RMSE	0.506	9.156	1.805
MAPE(%)	4.401	5.378	4.295
MAE	0.015	0.034	0.043

The following Figure shows the relationship between the values of the three variables were obtained from the Iraqi meteorological organization and seismology and the values estimated by ARIMA model which explained that every increase in (x-axis) the corresponding increase in(y-axis) that   proved   by   the   simple   linear   regression   equation.
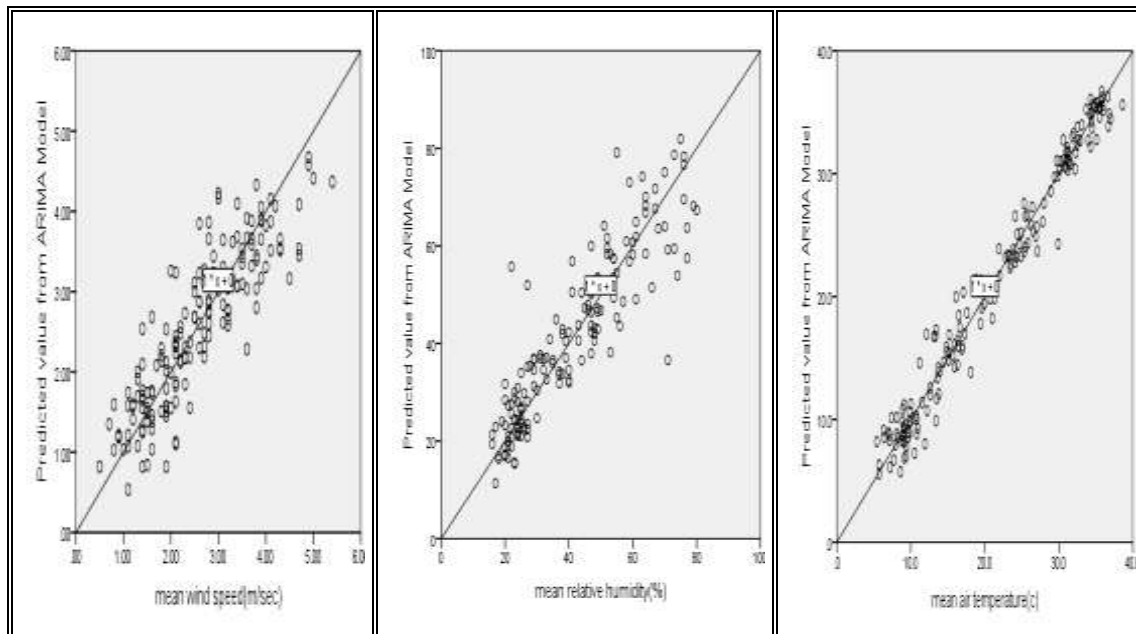
Figure 4-the comparison of actual values and predicted values for the variables.

The following Figures show the missing values which estimated by ARlMA Model for the time series of the three variables where the red curve shows observed values and the blue curve shows predicted values.
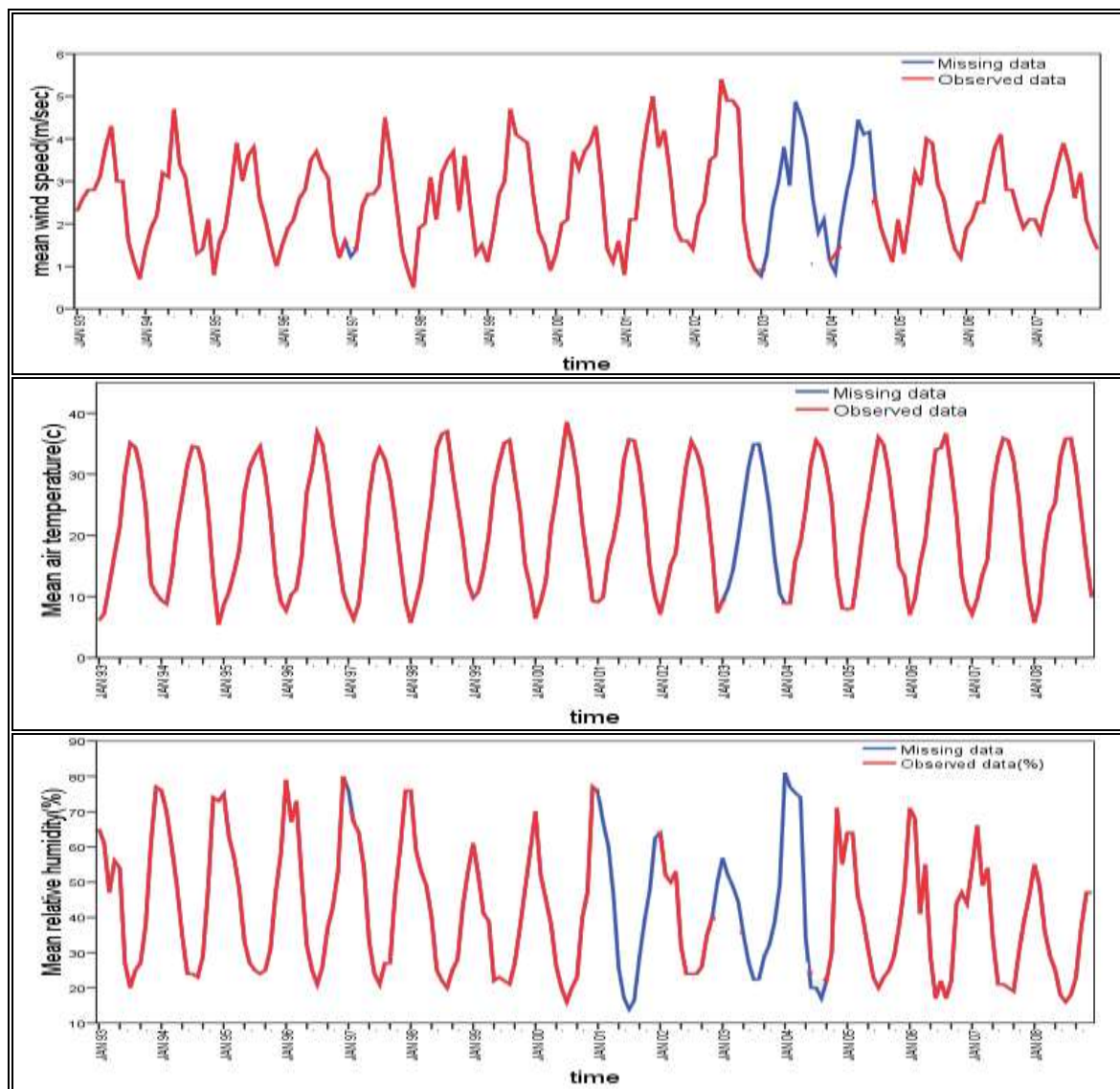
Figure 5-Time series of missing data and observed data for variable.

The following Tables will give the values of the missing data for the meteorological variables as shown below:

Table 2-The Missing Data For Air Temperature (˚C) in Sinjar Station.

year	Jan.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
RH. (2003)	9.3	11.2	14.4	19.5	25.4	31.3	34.9	35.0	30.5	24.8	15.2	9.2

Table 3-the missing data forwind speed (m/sec) in sinjar station**.**

month	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
wind speed(1997)	1.2	1.4	2.4	2.7	2.7	2.9	4.5	3.6	2.5	1.4	0.9	0.5
wind speed(2003)	1.1	2	2.8	3.6	4.5	5	5	4.5	3.6	2.5	1.7	1.1
wind speed(2004)	1.1	1.4	2.3	3.3	4.1	4.6	4.5	4.1	2.6	1.9	1.5	1.1

Table 4-the missing data for relative humidity(%) in sinjar station.

year	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
RH(2001)	76	67	60	46	26	17	14	16	29	39	48	63
RH(2000)	64	52	50	53	32	24	24	24	26	35	40	50
RH(2003)	57	52	49	45	35	27	22	29	32	39	49	70
RH(2004)	81	77	75	71	34	20	20	17	22	30	71	55

In the same way were calculated the missing data for each of the Baghdad and AL-HAI cities.

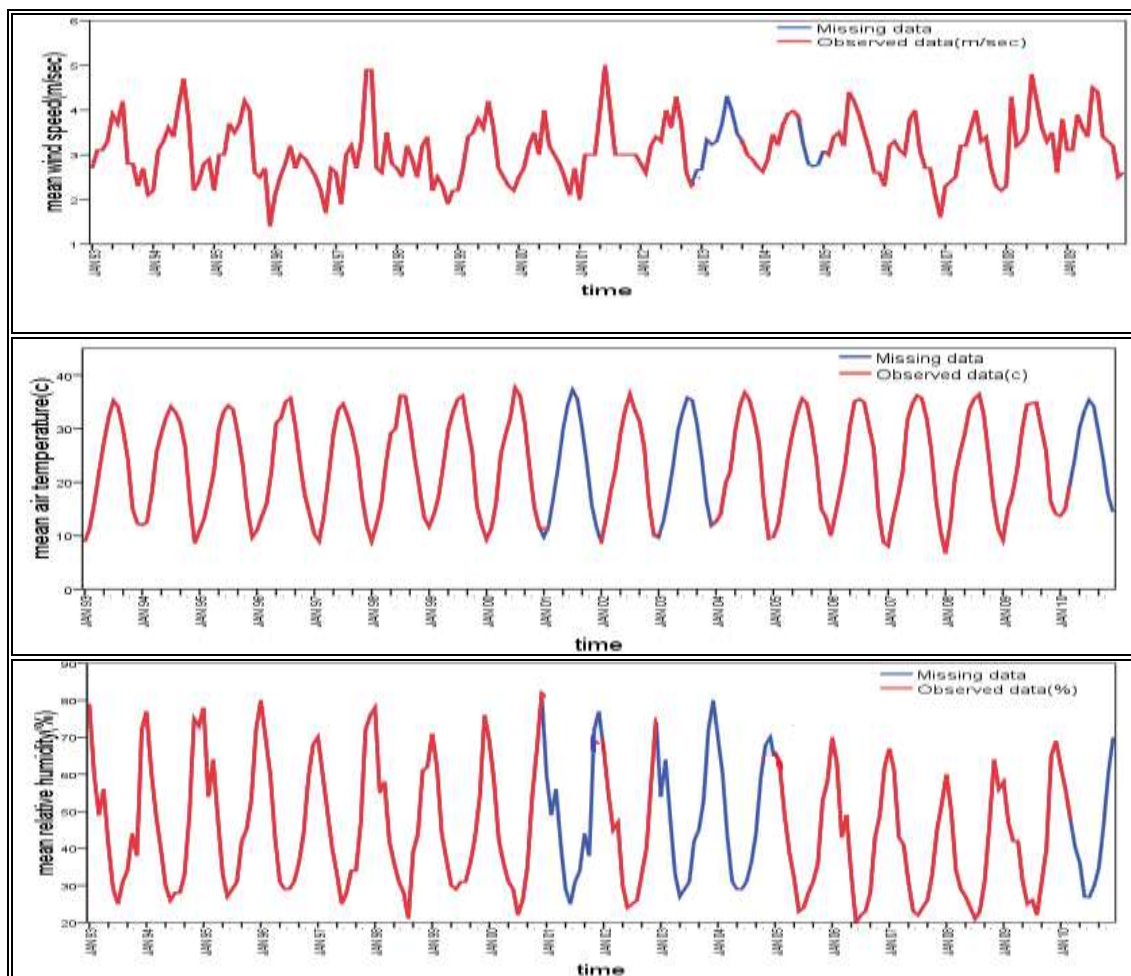## 4.2. Baghdad station

Time series for the Baghdad city for each air temperature and relative humidity began from 1993 to 2010 but wind speed began from 1993 to 2009. the missing data for air temperature was found in all months of 2001 and 9 month from 2010 while the missing relative humidity was found missing at all the months of 2001,2003, 2004 and 9 months from 2010. but, the wind speed was found missing in one month from 1999, 2001, 2002, 2006 and 4 months from 2003 and 2 months from 2004 as shown below:
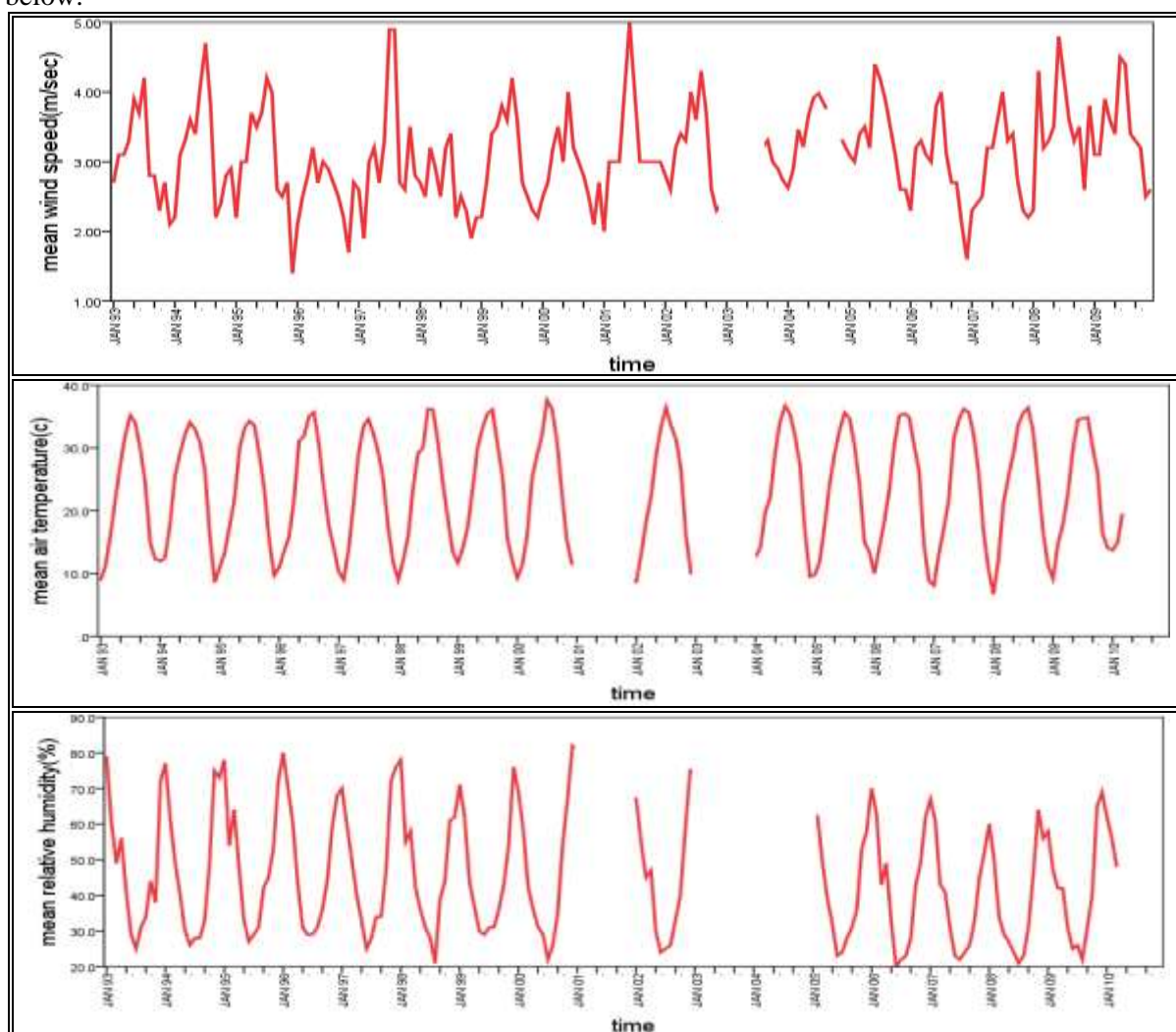
Figure 6- Time series of the mean wind speed, mean air temperature and mean RH.

The following Figures show the missing values which estimated by ARlMA Model for time series of the three variables where the red curve shows observed values and the blue curve shows predicted values.

Figure 7-time series of missing data and observed data for variables.

The following tables will give the values of the missing data for the meteorological variables as shown below:

Table 5-The missing data for Air temperature (˚C) in Baghdad station.

year	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
air temp(2001)	9.7	11.9	17.1	22.9	29.8	34.4	37.2	35.5	30.1	24.4	15.8	11.6
air temp(2010)	13.7	14.9	19.5	24.2	30	33.2	35.3	34.1	29.2	24.0	17.5	14.4

Table 6-The Missing Data For Wind Speed(m/sec) in Baghdad Station.

year	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
wind speed(**2002**)	2.8	2.6	3.2	3.4	3.3	4.0	3.6	4.3	3.7	2.6	2.3	2.7
wind speed(**2003**)	2.7	3.3	3.2	3.3	3.7	4.3	4.0	3.5	3.3	3.0	2.9	2.7
wind speed(**2004**)	2.6	2.9	3.5	3.2	3.7	3.9	4.0	3.9	3.2	3.8	2.7	2.8

Table 7-the missing data for relative humidity(%)in Baghdad station.

| year	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
RH (2001)	70	61	50	43	34	29	24	29	40	54	67	75
RH (2003)	68	57	48	40	31	26	24	26	34	46	58	67
RH (2004)	67	60	51	40	30	25	24	27	37	48	59	66
RH (2010)	62	56	48	40	36	26	26	29	34	46	60	70

4.3. AL-HAI Station

Time series for AL-HAI city for each of air temperature and relative humidity began from 1993 to 2009 but wind speed began from 1993 to 2007. the missing data for air temperature was found in all months of 2001 and 0ne month from 2002, 3 months from 2003, one month from 2004, 2 month from 2006 and 7 months from 2009 while the missing relative humidity was found missing at all the months of 2001, 4 months from 2003, 1 month for 2004, 2 months from 2006 and 7 months of 2009. But, the wind speed was found missing in all months of 2001, 1 month for 2002, 3months from 2003 and 1 month from 2004
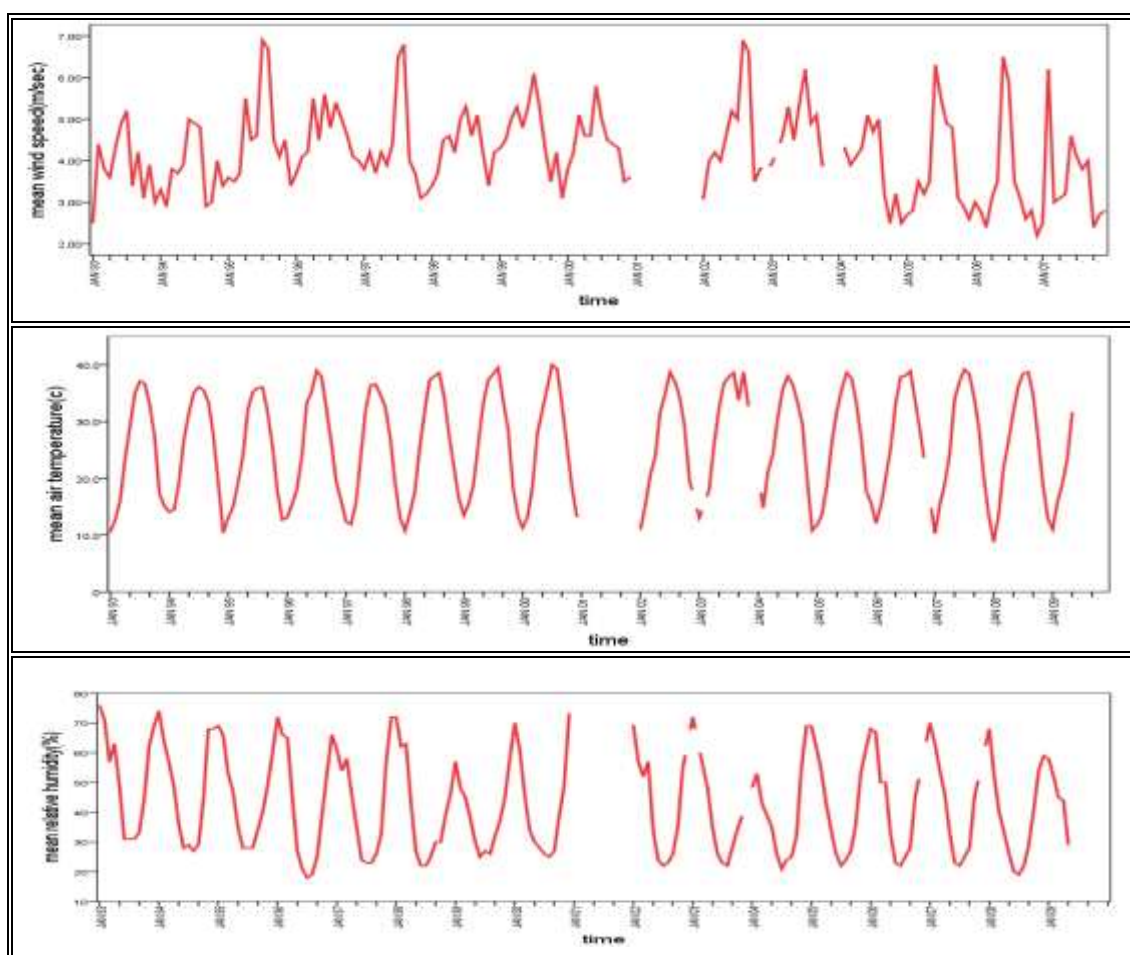
Figure 8-Time series of the wind speed, air temperature and rh.

The following Figures shows the missing values which estimated by ARlMA Model for time series of the three variables where red curve shows observed values and the blue curve shows predicted values.
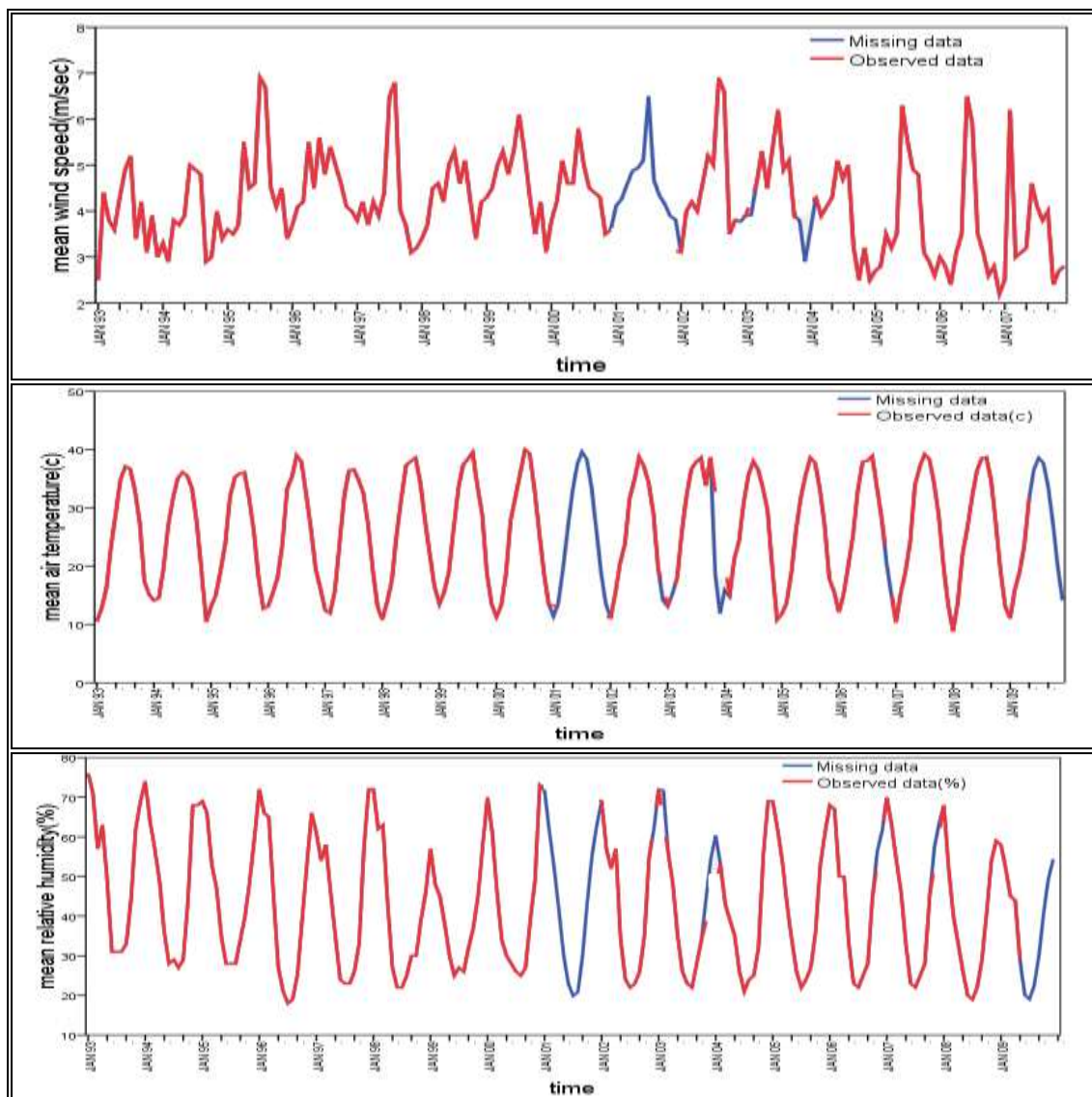
Figure 9-time series of the missing data and observed data in variables.

The following tables will give the values of the missing data for the meteorological variables as shown below:

Table 8-the missing data for mean air temperature(˚c) in AL-HAI station.

year	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
air temp.(2001)	11.4	13.4	19.5	26.6	32.5	37.5	39.6	38.3	33.5	26.3	18.9	13.5
air temp.(2002)	11.1	15.3	20.6	23.7	31.6	34.8	38.7	37.1	34.3	29.1	19.6	14.1
air temp.(2003)	13.1	15.3	18.0	26.0	32.3	36.7	37.9	38.6	33.8	38.7	18.7	11.9
air temp.(2004)	15.9	14.8	21.2	24.3	30.9	35.7	38.0	36.6	33.4	29.6	19.6	10.8
air temp.(2006)	12.1	15.2	20.2	25.4	33.2	37.9	38.1	38.9	33.5	28.1	20.6	15.2
air temp.(2009)	11.0	16.0	19.1	23.6	31.7	36.6	38.6	37.7	33.6	27.4	20	14.1

Table 9-the missing data for wind speed(m/sec) in AL-HAI station.

year	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
wind speed(**2001**)	4.1	4.3	4.6	4.9	4.9	5.1	6.5	4.7	4.3	4.2	3.9	3.8
wind speed(**2002**)	3.1	4.0	4.2	4.0	4.6	5.2	5.0	6.9	6.6	3.5	3.8	3.8
wind speed(**2003**)	3.9	3.9	4.6	5.3	4.5	5.4	6.2	4.9	5.1	3.9	3.8	3.5
wind speed(**2004**)	3.6	4.3	3.9	4.1	4.3	5.1	4.7	5.0	3.2	2.5	3.2	2.5

Table 10-the missing data for relative humidity in AL-HAI station.

year	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.
RH (2001)	72	62	52	42	30	23	20	21	30	44	55	63
RH (2002)	69	57	52	57	34	24	22	23	26	35	55	62
RH (2003)	72	72	56	48	35	26	23	22	28	34	45	55
RH (2004)	60	53	43	39	35	26	21	24	25	32	55	69
RH (2006)	68	67	50	50	33	23	22	25	28	45	57	63
RH (2009)	58	52	45	44	29	20	19	22	30	41	50	57

Conclusion

The results which obtained in this study explain that the ARIMA models provide a useful value of the mean monthly variables. ARIMA model is accurate but requires a lot of computational skill. The data should be available in sufficient large numbers to estimate the missing data which found within time series. The time series should be at least 10 years successive until achieve the appropriate model.

References
1. Nurul, k., Ahmad, M., Kamarulzaman, I. and Azami, Z. **2013**. Application of the Single Imputation Method to Estimate Missing Wind Speed Data in Malaysia. *Research Journal of Applied Sciences, Engineering and Technology* **6**(10): 1780-1784.
2. Haibo, H. Yuan, C. Yi, C. and Jinyu, W. **2013**.Ensemble learning for wind profile prediction with missing values. *Neural Computing & Applications*. **22**(2): 287-294.
3. Shamshad A. I., Hussin, W. M. A., Bawadi, M. A. and Sanusi, S. A. **2003.** Autoregressive Integrated Moving Average Model (ARIMA) For Forecasted Wind Speed. *School of Civil Engineering, University of Science Malaysia*, **125**: 591–613.
4. Kumar, A. and Goyal, P. **2011.** Forecasting of daily air quality index in Delhi. *Science of the Total Environment*, **409**: 5517-5523.