



# The Egyptian International Journal of Engineering Sciences and Technology

Vol. 19 No. 2 (2016-Special Issue) 346–352

<http://www.eijest.zu.edu.eg>



## Forecasting of Cairo Population using ARMA Model

K. A. Abdalmalak<sup>a\*</sup>, and F. J. Gonzalez-Serrano<sup>b</sup>

<sup>a</sup>Electrical Engineering Department, Aswan University, Aswan 81542, Egypt

<sup>b</sup>Signal Processing and Communication Department, Carlos III de Madrid University, Madrid 28911, Spain

### ARTICLE INFO

#### Article history:

Received 17 May 2016  
Received in revised form  
7 July 2016  
Accepted 9 July 2016  
Available online 10 July  
2016

#### Keywords:

Population forecasting  
ARMA model  
Yule-Walker equations  
Burg equations.

### ABSTRACT

The problem of large population is one of the most important factors influencing the economy and social advancement of Egypt. Population forecasts, when carefully and intelligently made, serves a valuable purpose in helping to direct the employment of labor and capital to places or projects where they are most needed. Firstly, the paper focuses on studying the population of the capital of Egypt (Cairo). By large numbers of sampling to the population data sequence, the increasing trend is found. Then, a time series model is given which can accurately forecast the population of Cairo. Multiple Autoregressive models AR (1), AR (2) are used the forecasting of the population in the next twenty years. The parameters of the model are calculated using the famous two methods: Yule-Walker and Burg. Before using the model to make predictions, the test of model response is verified, and the MSE and MAPE are measured to verify the models. The result is a scary image of the population in this city. Full descriptions for the steps of selecting the suitable model and comprehensive MATLAB simulation are presented. Secondly, the total population density of Egypt is analyzing and forecasting with using the measured data from 1970 to 2013. The same steps of the first part are done with the population density and forecasting of the increasing of the population density of Egypt in the 20 next years is presented. The main reasons for the population problem are discussed and solution of this problem is presented.

### 1. Introduction

According to the statistic of the population in Egypt, the number of people keeps on increasing at varying rates. Egypt is the most populous country in the Middle East and the third most populous on the African continent (after Nigeria and Ethiopia) [1]. The population of Egypt represents 1.20 percent of the world's total population which arguably means that one person in every 84 people on the planet is a

resident of Egypt [2]. It grew gradually and steadily throughout the nineteenth century, beginning in the 1990s, the growth rate accelerated, and the population increased more than 600 percent. Close to 99% of all Egyptians lived along the banks of the Nile River in around 4 percent of the country's total area. Average population densities in the Nile Valley exceeded 1,500 per square kilometer, and it is one of the world's highest densities. The most crowded city in Egypt is Cairo. The population of Cairo is around 18 million people. Around 20% of the Egyptian

\* Corresponding author *E-mail address:* kerlos.atia@alumnos.uc3m.es

people live in Cairo, which has an area of 1492 km<sup>2</sup> approximately.

The AR (Auto-Regressive) model is a good model to capture the behavior of the population. Forecasting the population is a very important issue because it gives indications on the inflation of the problem. During 1970 to 2010, there was a population boom due to advancements in agricultural productivity and medical fields, which were made possible due to the Green Revolution.

There are several researchers are working on forecasting the population of different countries which faces this problem such as China [3] - [5]. Also, there are some researchers who focus on the effects of this problem for example in [6] the authors use Gray Model (GM) order (1,1) [7] "which is an approximate model has the characteristic both differential equation and difference equation" to forecast the arable land area in Sichuan province, China. Although Cairo is one of the most crowded cities in the world, there is no huge interest of the researchers to study its population's forecasting and see the size of the problem in the future.

In this paper, a study of the population of Cairo is discussed and full explanation for prediction and selecting the suitable model for prediction. The rest of the paper is organized as follows. Section 2 gives an overview of population problem in Egypt with a brief discussion of the used models which will be used to forecast the population. Section 3.1 presents the study and forecasting of Cairo's population and the steps for selecting the model, and its results are discussed. A modification in the calculation of the trend and forecasting is presented in section 3.2. Section 3.3 and 3.4 present the forecasting of population density of Egypt with the discussion of the population problem in Egypt and its possible solutions. Finally, all work is concluded in section 4.

## 2. Problem Statement and Methodology

Cairo is the capital of Egypt; it is the largest city in the Middle-East and second-largest in Africa (after Lagos). It is the third largest urban area in the Islamic world after Jakarta and Karachi, and the world's 16th largest metropolitan area [8]. Around 20% of the Egyptian people live in Cairo; its space is around 1492 km<sup>2</sup>. It has a significant increased over the years, especially in the last 20 years. This increase has many negative effects, such as lack of resources and unemployment and terrible problems in energy consumption and food. Also, it can be seen

that problem of Cairo's population as a thumbnail image of the real problem of the population of whole Egypt. The problem of the population of Egypt can be discussed with two different points of views: population only or population with space. The second point of view is related to forecast the population density instead of forecasting the population only. This point of view will answer an important question which if the problem of the population of Egypt will still exist even if the spaces of the whole country efficiently used or not. So after the complete study of the population of Cairo, the density of population of whole Egypt will be studied.

There are many models used for forecasting the time-series, Autoregressive Moving Average (ARIMA) is one of the most and famous powerful models. The ARMA (p, q) model combines two different models and has two parameters of order: p which is the order of auto-regressive (AR) model and q is the moving-average (MA) model's order models [9-10]. The two parameters are non-negative integers. Any ARMA process can be described by the following equation:

$$\Phi(z) X_t = \Theta(z) Z_t \quad (1)$$

where  $X_t$  is the time series.

$$Z_t \sim WN(0, \sigma^2).$$

$$\Phi(z) = 1 - \phi_1 z^{-1} - \dots - \phi_q z^{-q}$$

$$\Theta(z) = 1 + \theta_1 z^{-1} + \dots + \theta_q z^{-q}$$

There are two main pre-processing steps need to be performed before using the model:

1. Trend separation.
2. Seasonal component separation.

After the pre-processing of the data series, it is necessary to see if there is still important information in the residuals which results from the differences between the real data series and the trend-seasonal components. To see this importance, the Independent and Identically Distributed (IID) tests need to be performed on the residuals. The IID tests focus on studying the correlation between the residuals samples, to see this correlation two main functions need to be studied:

Autocorrelation function (ACF) and partial autocorrelation function (PACF), in IID, noises 95% of the sample autocorrelation values within  $\pm 1.96/\sqrt{n}$ . The IID can be tested by five methods: sample ACF, portmanteau test, turning point, difference-sign, and rank tests.

After proofing of existing of the correlation between the samples of the residuals (rejection of IID

hypothesis), the proper model and its order has been selected using the values of ACF and PACF. The ARMA model can be replaced by one of its two fundamental models (MA model or AR model), and this happens only in case of finite PACF or finite ACF respectively.

### 3. Experimental Results and Discussion

#### 3.1. Cairo population

The population of Cairo from 1970 to 2010 is shown in fig. 1 (dataset was taken from the World Bank [11]):

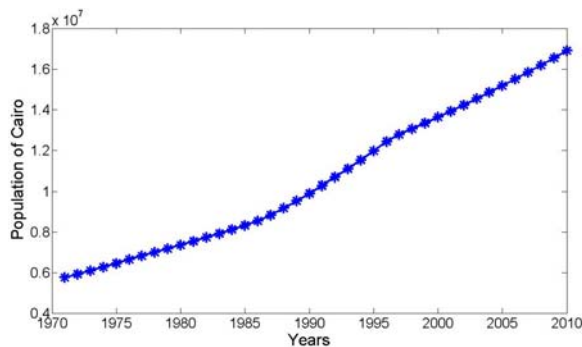


Fig. 1. The population of Cairo along the last four decades.

From the previous Figure, it can be seen that this series has not any seasonal component, and it has an increasing trend. Figure 1 also demonstrates that this data series can be fitted to a polynomial curve with the second order as in Fig. 2.

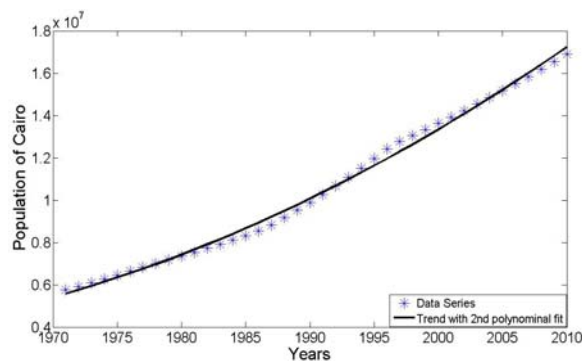


Fig. 2. The trend of the population of Cairo data series.

The trend can be expressed by the following equation:

$$P = 0.031 \times T^2 - 12.64 \times T + 11719.7 \quad (2)$$

where:  $T$  is the year for  $1970 \leq T \leq 2010$ .

$P$  is the population in million people.

To find the residuals, the calculated trend must be removed firstly. The resulted residuals are shown in Fig. 3 which has a maximum of 0.5 million instead of 18 million.

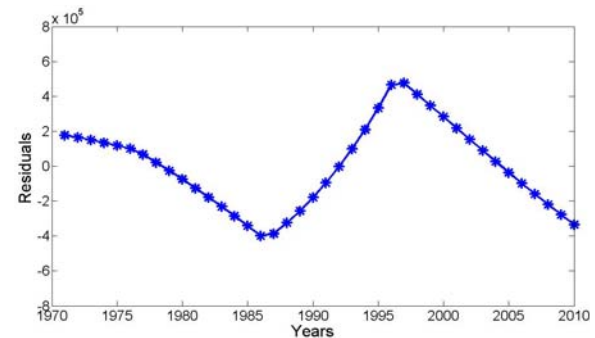


Fig. 3. The residuals after removing the trend.

The next step after calculating the residuals is to know if they are IID sequence (can be neglected as they are uncorrelated and do not contain any important information) or not. To know this, a proper test has to be performed. The famous five tests to check the IID hypothesis are: sample autocorrelation function test, portmanteau test, turning point test, difference-sign test, and rank test. These tests have a binary output true or false, which true means that these residuals are really IID sequence so they can be neglected. In the other hand, if the output of the test is false which means that the residuals are correlated (IID hypothesis rejected) a proper model has to be used to represent these residuals and extract the important information on them. The detailed explanation of these five tests can be found in [12].

After performing these five tests, the outputs for all tests are false which ensure that these residuals can not be neglected. The suitable model can be selected after calculating the autocorrelation and partial autocorrelation functions. The sample autocorrelation function of the residuals is in fig. 4(a), which has a lot of samples out the  $\pm 1.96/\sqrt{n}$  and this is clear that the sequence failed in sample ACF test, so the residuals are not an IID sequence.

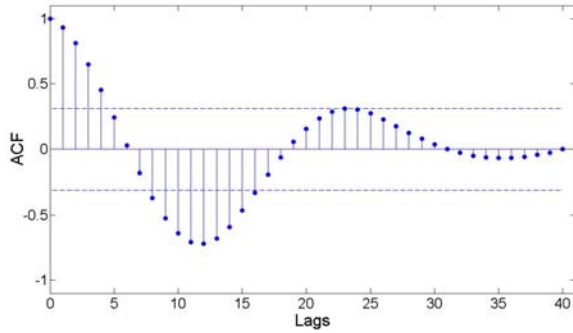


Fig. 4(a). ACF of residuals of the sequence.

Figure 4(a) shows that the ACF for these residuals is infinity, so PACF needs to be calculated.

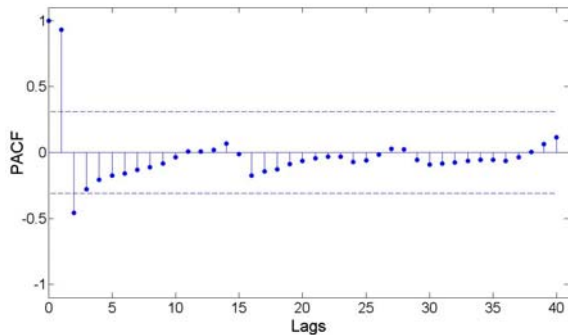


Fig. 4(b). PACF of residuals of the sequence.

Figure 4(b) shows the partial autocorrelation function of which only two samples are out of the predefined range. From the two previous figures, the suitable model is autoregressive with the order (p) of two because ACF decays slowly and PACF cuts off at lag 3 (P+1). The coefficients of AR model can be calculated either with Yule-Walker or Burg equations.

Table 1. Comparison between the results of Yule-Walker and Burg techniques for forecasting the population of Cairo.

Parameters	Yule-Walker	Burg
Coefficients	$\Phi_1=1.66$ $\Phi_2=0.73$	$\Phi_1=1.87$ $\Phi_2=0.94$
Mean Square Error *10 <sup>-9</sup> (MSE)	50.4	30.5
Mean Absolute Percentage Error (MAPE)	0.65%	0.49%

Depend on the results of error in Table 1; the Burg algorithm is preferred for forecasting the population of Cairo in next three decades as in fig. 5.

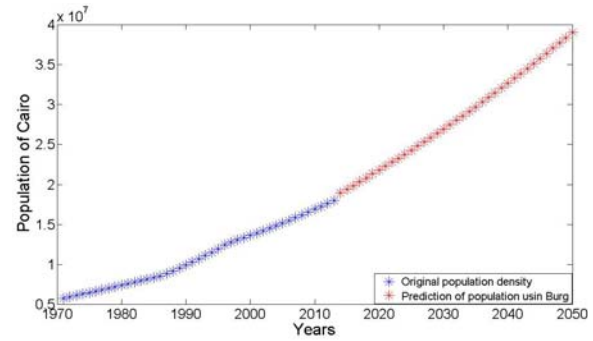


Fig. 5. Forecasting the population of Cairo in next three decades.

It is seen from the forecasting that in 2050 Cairo city will have a population about 40 million people.

### 3.2. Modification in Cairo population forecasting

By focusing on the data series, it can be seen that from 1990 the series has a slightly different behavior. So we can see the series as it consists of two linear parts with two different slopes. So it is better to calculate the trend using data from 1990 to 2010 (20 samples).

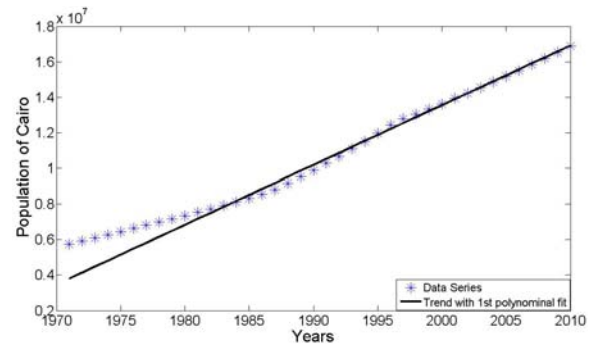


Fig. 6. The modified trend of the population of Cairo data series.

The trend can be expressed by the following equation:

$$P = 0.299 \times T - 584.661 \quad (3)$$

where: T is the year for  $1990 \leq T \leq 2010$ .

P is the population in million people.

To find the residuals, this trend needs to be removed firstly. Applying IID tests give the same results as the previous one (rejecting the IID hypothesis in all five tests). To choose the best model, the autocorrelation, and the partial autocorrelation functions are calculated and show in

Fig.7 (a), Fig.7 (b) respectively.

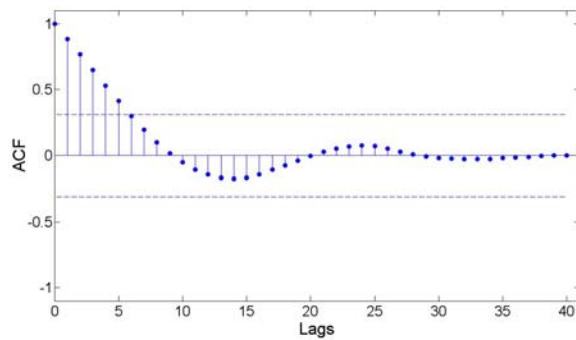


Fig. 7(a). ACF of modified residuals of the sequence.

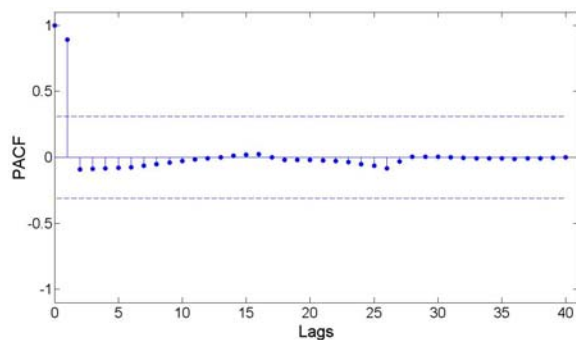


Fig. 7(b). PACF of modified residuals of the sequence.

From the results of ACF and PACF figures, the suitable model is AR (1) because ACF samples decay slowly, and PACF has only one sample outside the bands. The coefficients of AR model can be calculated either with Yule-Walker or Burg algorithms.

Table 2. Comparison between the results of Yule-Walker and Burg techniques for modification of forecasting the population of Cairo.

Parameters	Yule-Walker	Burg
Coefficients	$\Phi_1=0.91$	$\Phi_1=0.99$
MSE *10 <sup>-9</sup>	19.3	20.3
MAPE	0.412%	0.423%

This modified method decreases the MAPE (using Yule-Walker technique) from 0.646 % to 0.412 %. Depend on the results of error in Table 2; the Yule-Walker algorithm is used for forecasting the population of Cairo in next three decades as in Fig. 8.

It is seen from the forecasting that in 2050 Cairo will have a population about 32 million people (instead of 40 million people in the first method). One of the best methods for a fair comparison between the two methods is Akaike's Information Corrected Criterion (AICC) which gives an

indication not only to prediction error but also to the complexity of the used method. The modified method is given AICC=32101 instead of 55492 in the first one.

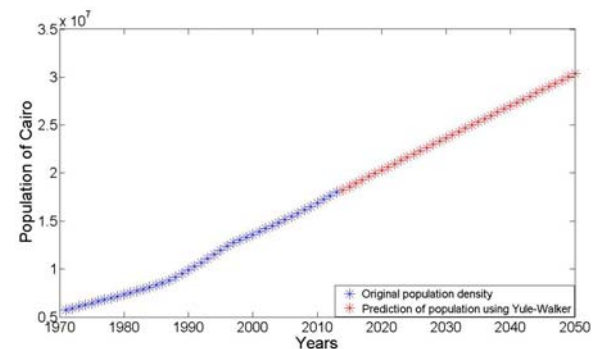


Fig. 8. Forecasting the population of Cairo in next three decades.

It is seen from the results of forecasting the population of Cairo in each of the previous two methods that population explosion will occur in the near future. The main problem is: Cairo now faces a lot of problems related to the current population (19 million only), so it can not imagine the size of this problems when the population reaches to the higher predicted values. City's resources will not bear this enormous population, and the problem must be solved nowadays before converting to a terrible disaster.

### 3.3. Egypt population density

From any tries of forecasting population of large cities in Egypt such Cairo (as in the previous two sections) or forecasting the population of whole Egypt, the results are very scary, and all confirm the occurrence of a population explosion very soon. The majority of these studies have looked at the problem of narrow outlook because they studied only the population. In this section, the problem will be studied in a different point of view as the focus will be not only in the number of people but the overall space will be taken into account also. So the density of population will be used instead of population only.

Figure 9 presents the data series of the population's density of Egypt and the increasing trend which can be expressed by following first order equation:

$$p_d = 1.858 \times T - 2104.7 \quad (4)$$

where:  $T$  is the year for  $1990 \leq T \leq 2010$ .



$P_d$  is the density of population in million people per  $\text{km}^2$ .

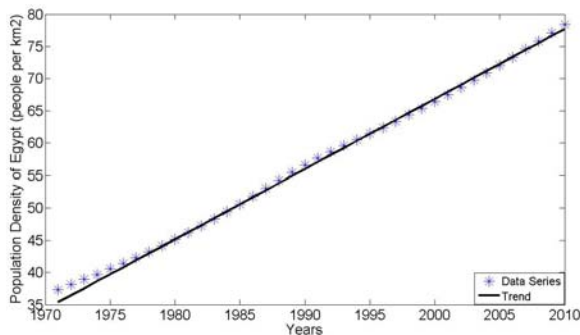


Fig. 9. The data of density of population of Egypt and its trend.

After subtracting the trend from the data, it is found that the remaining residuals have a maximum of 2 million/ $\text{km}^2$  as in fig. 10. But these residuals are somewhat correlated, and ACF and PACF functions need to be calculated to select the proper model.

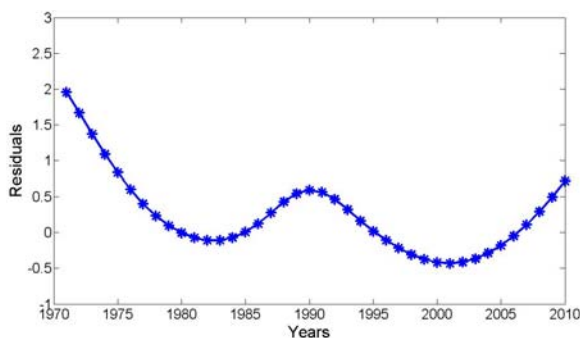


Fig. 10. The residuals of the density of population after removing the trend.

After calculating these two functions, it can be concluded that the first order of autoregressive is the suitable model. The coefficient of the model can be calculated either with Yule-Walker or Burg algorithms. The better algorithm can be selected by comparing the predicted sample (in each case) with the tested data and calculate the mean square error and mean absolute percentage error as in Table 3.

Table 3. Comparison between the results of Yule-Walker and Burg techniques for forecasting the density of population of Egypt.

Parameters	Yule-Walker	Burg
Coefficients	$\Phi_1=0.86$	$\Phi_1=0.96$
MSE $\times 10^{-9}$	1.59	1.02
MAPE	0.84%	0.67%

From the coefficient of AR (1) and the equation of trend, linear prediction can be applying for forecasting the future 30 years as in Fig. 11.

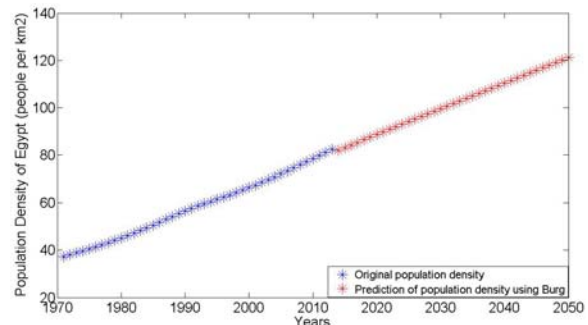


Fig. 11. Forecasting the density of population of Egypt in next three decades.

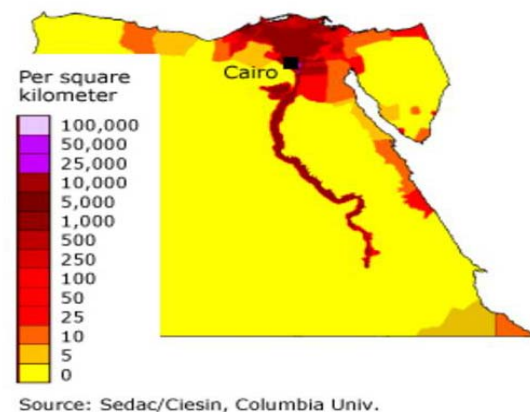
It is seen from the forecasting that in 2050 Egypt will have total population density below 130 people per  $\text{km}^2$  which is considered a low-density value.

### 3.4. Discussion on population problem in Egypt

From the results of the previous section, it can be demonstrated that the problem is mainly in the distribution of people in Egypt not only in population. This because of nearly 98% of the country (which is around 82.5 million people in 2012 [1]) live in three major regions of the country:

- 1- Cairo (the capital of Egypt).
- 2- Alexandria.
- 3- Along the banks of the Nile and the Suez Canal.

The distribution of people with the space of Egypt is presented in the following fig. 12.



Source: Sedac/Ciesin, Columbia Univ.

Fig. 12. The map of the density of population of Egypt.

These three regions are among the world's most density populated, containing an average of over 1,540 per km<sup>2</sup>, as compared to 80 persons per km<sup>2</sup> for the country as a whole. The space of these three regions is lower than 4% on the total space of Egypt because the large regions of the desert, which constitute most of Egypt's territory, are sparsely inhabited.

There are some trends to solve the problem of the population by trying to reclaim these deserts and construction of new cities, but now a lot of these trends are stopped due to the circumstances of the country's political. But now it is necessary for the government to focus on building new cities on deserts because it is the faster solution to the problem of the population before the population explosion takes place.

#### 4. Conclusion

In this paper, the population of Cairo was studied, forecasting of the population in the next twenty years was done using multiple Autoregressive models AR (1), AR (2). It has been seen from the results of forecasting the population of Cairo that population explosion will occur in the near future. After that the density of Egypt population studied and forecasting in the next twenty years was done using multiple AR (1) models. The results show that there will not be any problems with density shortly and the main problem is in people distribution. The desert is significantly affected by this problem; all the results show that the only effective and rapid solution is in focusing on correcting people distribution in Egypt.

#### Acknowledgements

Thanks to Erasmus Mundus “Green-IT” program for its grant for providing the funding for this work.

#### References

- [1] [https://en.wikipedia.org/wiki/Demographics\\_of\\_Egypt](https://en.wikipedia.org/wiki/Demographics_of_Egypt), last accessed: January 25, 2015.
- [2] <http://www.tradingeconomics.com/egypt/population>, last accessed: January 25, 2015.
- [3] K. P. Meng, and W. Zeng, “China's Population Development and Prediction in the Next 50 Years,” *Research on Number Economic and Technologic Economic*, vol.3, pp.12-17, 2004.
- [4] Y. Wan, L. Xiao, and C. Wu, “An Optimum Intelligent Algorithm and its Application in Population Statistic and Forecast”, *WRI Global Congress on Intelligent Systems*, Xiamen, pp. 40 – 44, 19-21 May 2009.
- [5] L. Juan, and W. Liu, “Population Forecasting in China Based on the Grey-Markov Model”, *International Conference on Information Management, Innovation Management and Industrial Engineering (ICIIM)*, Shenzhen, pp.133 -136, 26-27 Nov. 2011.
- [6] G. He, “Arableland Area Forecast Based on GM (1, 1) Model: A Case of Sichuan Province, China”, *International Journal of Advances in Management Science*, vol.1, pp. 7-12, July 2012.
- [7] L. Dang, L. Sifeng, and D. Yaoguo, “The Optimization of Grey Model GM (1, 1)”, *Journal of Engineering Science*, Vol. 8, pp.50-53, 2003.
- [8] <http://en.wikipedia.org/wiki/Cairo>, last accessed: January 25, 2015.
- [9] W. W. S. Wei, “Time series analysis”, Addison-Wesley, 1994.
- [10] G. Box, G. Jenkins, and C. Reinsel, “Time Series Analysis: Forecasting and Control”, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [11] <http://data.worldbank.org/country/egypt-arab-republic>, last accessed: January 25, 2015.
- [12] P. J. Brockwell, and R. A. Davis, “Introduction to time series and forecasting”, Springer Science & Business Media, 2006.