Al-Azhar University Civil Engineering Research Magazine (CERM)
Vol. (39) No. (2) April, 2017

# Prediction of Traffic Volumes on Egyptian National Highways Using Time Series Analysis

**Aya Farag[1], K. El-Araby[1], Mona Abdallah[1], and M. El Esawey[1]**

[1]Department of Public Works, Ain Shams University, Cairo, Egypt.

**ملخص البحث**

تعتبر السلاسل الزمنيه من أهم الأساليب الإحصائيه الحديثة التى يمكن من خلالها معرفة طبيعة التغيرات التى تطرأ على قيم أحجام حركه المرور مع الزمن وتحديد الأسباب والنتائج وتفسير العلاقات التى تربط بينهما والتنبؤ بما سيحدث من تغير فى قيم أحجام المرور فى المستقبل على ضوء ما تم جمعه من بيانات فى الماضى ؛فأحجام المرور على الطرق هى أحد العوامل الاساسية اللازمة لتصميم أنظمة النقل ،والتنبؤ الدقيق بأحجام المرور يوفر مؤشرات دقيقة تجعل من السهل وضع الخطط التصميمية المستقبلية المناسبة لاستيعاب هذه الزيادة , لذلك فقد اهتم الباحثون بوضع العديد من النماذج والخطط لزيادة الدقة فى العملية التنبؤية ,كان من ابرزها نماذج بوكس وجنكزالتى أثبت كفاءتها ودقتها فى مجالات تطبيقها. والهدف الاساسى من هذا البحث هو تطبيق نماذج بوكس وجنكزللتنبؤ بحجم حركة المرور اليوميه والأسبوعية على مجموعة من الطرق السريعة الاقليمية بمصر, وذلك باستخدام بيانات حجم حركة المرورالمجمعة من ١٤ محطة رصد دائمه على الشبكة في كلا الاتجاهين على مدار خمسة سنوات من ٢٠٠٨حتى عام ٢٠١٢, حيث استخدمت بيانات حجم حركة المرور اليومية والأسبوعية خلال السنوات ٢٠٠٨ حتى عام ٢٠١١ لبناء السلاسل الزمنية وتحليلها واختيار النموذج المناسب , بينما استخدمت بيانات العام ٢٠١٢ للتحقق من صحة النماذج ومطابقة نتائجها مع أحجام حركه المرور التى تم تجميعها. وقد اشارت النتائج الى إمكانية استخدام نماذج (SARIMA) الانحدار الذاتي و المتوسط المتحرك الموسمى للتنبؤ بأحجام المرور اليومية والأسبوعية المستقبلية على شبكة الطرق السريعة الاقليمية بمصر.فبشكل عام تراوحت نسبة الخطأ بين ٤,٣% و ٧,١٠% عندالتنبؤ بأحجام حركة المرور اليومية , بينما كانت نسبة الخطأ بين ٢% و٧% عند التنبؤ بأحجام حركة المرور الأسبوعية . كما أظهرت النتائج عدم وجود نموذج واحد محدد من شأنه ان يناسب جميع أنواع الطرق عند التنبؤ بأحجام الحركة الأسبوعية .

## Abstract:

Traffic volume is a fundamental measure that is used to serve different types of transportation-related applications at planning and operational levels. Traffic data collection is usually costly and hence researchers started to adopt other approaches to compensate for the scarcity of data and limited resources. In the past few decades, many effective approaches have been developed and applied to estimate/predict traffic volumes. Time series analysis is considered to be one of these powerful approaches where future values of the traffic volumes at a facility are estimated using the time-relationships with the past values. This paper investigates the application of Box-Jenkins time series models to predict daily and weekly traffic volumes for a group of Egyptian national highways. A total of 5- year of traffic volume data of 14 stations in both directions; 2008 to 2012, were used in the analysis. The daily and weekly traffic volume for the years 2008 through 2011 were used to fit the time series models. Seasonal autoregressive integrated moving – average (SARIMA) models were shown to be the most suitable modeling structure for the analysis. The developed models were used to forecast traffic volumes for the year 2012. The forecasted traffic volumes were then compared to the actual traffic volumes and the estimation error was computed. In general, estimation errors ranged between 4.3% and 7.10% for daily traffic volumes estimation, while the error was between 2% and 7% for weekly traffic volumes estimation. The results of this paper show the potential of using time series models for predicting daily and weekly traffic volumes on the national highway network of Egypt.

The results also showed that there is no single model structure that will best fit the data of all highway types.

**Key Words**: Traffic volumes, time series models, predict traffic volumes, Box & Jenkins.

## 1. INTRODUCTION

Traffic characteristics such as traffic volume, speed, and density are of interest to the users of transportation system. These variables are usually considered as performance measures of the highway network. Traffic volume is an important traffic characteristic in transportation systems. The measurement and prediction of traffic volumes are crucial in the planning, design, and operation of highway facilities. A very important requirement in nearly all transportation planning and design strategies is the knowledge of spatial and temporal distribution of traffic. Therefore, there is a growing need for time-dependent analysis techniques such as time series analysis in the fields of transportation planning and traffic engineering. Whether concerning long-range plans, transportation management strategies, or short-range decisions, time series techniques have immediate application in transportation and traffic planning. Thus a lot of active research works have been conducted during the past several years on the prediction of traffic volumes using time series models. The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. These models are then used to generate future values for the series to make forecasts.

Transportation researchers have developed time series traffic prediction models using Box & Jenkins techniques (1994) such as autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and seasonal autoregressive integrated moving average (SARIMA). The Box-Jenkins approach to estimate models of the general ARIMA family is currently one of the most widely implemented tools for modeling time series data. Although the three-stage approach (model identification, parameter estimation, and forecasting) was originally designed for modeling time series, the underlying strategy is applicable to a wide variety of statistical modeling approaches. This method provides best forecasts for the majority of the series tested.

## 2. APPLICATION OF TIME SERIES MODELS IN TRAFFIC FORECASTING: A LITERARTURE REVIEW

An early study on time series is a set of values of a particular variable that occur over a period of time in a certain pattern. Many statistical methods relate to data which are consisting of successive measurements or observations on quantifiable variables, made over a time interval (Bowerman, O'Connell, and Koehler, 2005). Usually the observations are chronological and taken at regular intervals (days, months, years).Many of models such as, Exponential smoothing, linear regression, Bayesian forecasting and Box – Jenkins used in Traffic engineer forecasting.

Ahmed and Cook (1979) used 166 datasets from three surveillance systems in Los Angeles, Minneapolis, and Detroit to develop a model for short-term traffic volume and occupancy forecasting. All of the datasets were best represented by an autoregressive integrated moving-average (ARIMA) model.

Another study by Nihan and Holmesland (1980) used monthly average of weekday traffic in Seattle to fit ARIMA time series models. A data set containing monthly volumes on a freeway segment for the years 1968 through 1976 is used to fit a time series model. The resulting model is used to forecast volumes for the year 1977. The forecast volumes are then compared to actual volumes in 1977. The results of this study indicate that time series techniques can be used to develop highly accurate and inexpensive short term forecasts.

Bengamin (1986) present a procedure for forecasting average daily traffic using time series model. He compared time series model with Urban Transportation Planning System (UTPS) developed demand forecast and noted that time series method work well in a stable growth situation but can't estimate sudden shifts in behavior. He also noted that adding more explanatory variables to the time series model would improve its predictive ability.

Williams et al. (1998) modeled traffic flow data by definite periodic cycles. Seasonal autoregressive integrated moving average (SARIMA) and winters exponential smoothing models were developed and tested on data sets belonging to two sites: Telegraph Road and the Woodrow Wilson Bridge on the inner and outer loops of the Capital Beltway in northern Virginia. Data were 15-min flow rates. The single-step forecasting results indicated that seasonal ARIMA models outperform the winters exponential smoothing and historical average models.

Also, Karlaftis (2003) developed multivariate time-series-state space models using 3-minute volume measurements from loop detectors near downtown Athens. The models were fed by data from upstream detectors to improve the volume prediction of downstream locations. They suggested that different model specifications are appropriate for different time periods of the day. The authors also suggested that the multivariate state-space models improve the prediction accuracy over traditional time series models.

B. Ghosh et al. (2004) modeled traffic flow at an arterial intersection in a congested urban transportation network in the city of Dublin. Three different time-series models, random walk model, exponential smoothing technique and seasonal ARIMA model are used for modeling of traffic flow in Dublin. Simulation and short-term forecasting of traffic flow data are done using these models. The data used for modeling are obtained from loop-detectors at a certain junction in the city center of Dublin. Seasonal ARIMA and exponential smoothing technique gave highly competitive forecasts and match considerably well with the observed traffic flow data during rush hours.

Sabry et al. (2007) compared regression and auto regressive integrated moving average (ARIMA) models in forecasting highway daily traffic volume. According to the analysis, the ARIMA model was shown to be the best forecasting method especially for the average monthly and average weekly daily traffic volumes. This approach is particularly appropriate when the forecast is short term and there is insufficient time and resources to build and calibrate a behavioral model.

The previous review primarily concerned with previous applications of time series models in traffic engineer, especially using of Box - Jenkins in traffic data forecasting. A number of applications of ARIMA have also been summarized in Table (1). It shows that their level of complexity models appears to have small sample size and them all neglected roads characteristics. But in this research different road facility such as Land use (agricultural & desert) ,with different Geometric ( 2 lanes two way highway roads, 4 lanes divided highway roads and 6 lanes divided highway roads ) are considered .Also, the sample size of data collected were large contain fourteen stations on Egyptian National Highways. All of the data set containing daily and weekly traffic volumes from years 2008 to 2011 to improve the predictions of traffic volumes.

**Table (1): Summary of the literature reviews on ARIMA Models**

| Authors | Year | Model technique | Sample size | Main Outcome |
|---|---|---|---|---|
| Ahmed, M.S. & Cook, | 1979 | Autoregressive integrated moving average (ARIMA) model, moving-average, and double-exponential smoothing | A total of 166 data sets from three surveillance systems in Los Angeles, Minneapolis, and Detroit to provide short-term forecasts of traffic data. | ARIMA models were found to be more accurate in representing freeway time-series data, than moving-average, double-exponential smoothing |
| Nihan and Holmes-land | 1980 | Autoregressive integrated moving average (ARIMA) model | monthly average of week day volumes on a freeway segment in Seattle Washington for the years 1968 through 1977 | time series can be used to develop highly accurate of short-term traffic volumes forecasting |
| Williams, B.M., P. Durvasula and D. Brown. | 1998 | Autoregressive integrated moving average (ARIMA) and winters exponential smoothing models | Telegraph Road and the Woodrow Wilson Bridge on the inner and outer loops of the Capital Beltway in northern Virginia. Data were 15-min flow rates | Seasonal ARIMA models outperform the winters exponential smoothing and historical average models. |
| A & Karlaftis, M.G. | 2003 | Autoregressive integrated moving average (ARIMA) model | Core urban area loop detector data, 3-min volume measurements from urban arterial streets near downtown Athens, models were developed that feed on data from upstream detectors to improve on the predictions of downstream locations. | different model specifications are appropriate for different time periods of the day and multivariate state space models improves on the prediction accuracy over analysis time series. |
| Bidisha Ghosh, Biswajit Basu, and Margaret O'Mahony | 2004 | Random walk model, exponential smoothing technique and seasonal ARIMA model are used for modeling of traffic flow. | The data used for modeling are obtained from loop-detectors at a certain junction in the city center of Dublin. | Seasonal ARIMA and exponential smoothing technique give highly competitive forecasts and match considerably well with the observed traffic flow data during rush hours. |
| Sabry, Abd-El-Latif and Badra | 2007 | logistic regression model and an autoregressive integrated moving – average (ARIMA) model | A total of 14-year traffic data for a Tanta – Mansoura road in both directions (from 1990 to 2003) were used to predict traffic volume. | ARIMA model seems to be the best forecasting method especially for short term forecasts of both average monthly and average weekly daily traffic volume. |

## 3. DATA DESCRIPTION

This paper investigates the application of time series techniques to predict daily and weekly traffic volumes for a sample of Egyptian national highways. These roads are described in Table 1.

**Table 2: Selected Egyptian Highways for Traffic Volumes Data collection**

| Road Class | Station No. | Direction | SECTION |
|---|---|---|---|
| Two-lane Two Way Agriculture | 1 | T* | Sharkia – Ismailia (AHM – ISM) |
| | 2 | T | Minya – Asyut (MNY - ASU) |
| Four-lane Divided - Agriculture | 3 | 1,2** | Gharbia – Beheira (TNT - DMH) |
| | 4 | 1,2 | Qaliubia – Sharkia (ORR - BLB) |
| | 5 | 1,2 | Gharbia – Dakahlia (TLK - SMN) |
| | 6 | 1,2 | Mit Ghamr – Aga (MTG – AGA) |
| | 7 | 1,2 | Beheira – Alexandria (DMH - ALX) |
| Six-lane Divided - Agriculture | 8 | 1,2 | Defra-KafrElziat (DEF-ZIT) |
| | 9 | 1,2 | Menoufia – Gharbia (QSN - TNT) |
| | 10 | 1,2 | Cairo – Qaliubia (SHB - BNH) |
| Four-lane Divided - Desert | 11 | 1,2 | Cairo – Ismailia (10R - ISM) |
| | 12 | 1,2 | Cairo – Suez (ORR -TNLj) |
| | 13 | 1,2 | Sadat City - Wadil Natrun (SDT - AMT) |
| | 14 | 1,2 | Giza – Faiyum (ALX - FYM) |

Source: Information Center of General Authority for Roads, Bridges & Land Transport (GARBLT).
*T: Total for both directions
**1, 2: Direction1 and opposite direction 2

As Table 2 shows, the available data include daily traffic volumes at 14 different count stations in both directions for a period of 5 years (2008 to 2012). The available dataset covered different highway types including agricultural and desert roads with different lane configurations (two-way-two-lane, four lanes, and six lanes). The analysis was carried out for the sum of traffic volumes of the two travel directions for Two-lane-two-way highways. On the other hand, the analysis was undertaken for each direction (direction 1 and direction 2) for divided highways. Daily and weekly traffic volumes for the years 2008 to 2011 were used to fit the time series models. The resulting models were used to forecast traffic volumes for year 2012. The forecasted traffic volumes for the models were then compared with the actual traffic volumes for the same period.

Statistical Package for the Social Sciences SPSS was used to develop the models used in this analysis. As well, several Excel Macros were created on the sample data.

## 4. THE MODELING FRAMEWORK

Box-Jenkins models consist of one or more of the following elements:

- Autoregressive (AR);
- Integrated (I);
- Moving Average (MA); and
- Combinations: ARMA, ARIMA, SARIMA or IMA.

Most of Box-Jenkins models combine the first three elements (ARIMA). The model is generally referred to as an ARIMA (p,d,q) model where p, d, and q are integers greater than or equal to zero, (p) refers to the order of the autoregressive part of the model, (d) integrated (it means the time-series are differenced), and (q) refers to the order of moving average part of the model. The AR, MA, and ARMA models can be expressed mathematically as:

AR model: $Z_t = \phi_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots\ldots + \phi_p Z_{t-p}$

MA Model: $Z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots\ldots - \theta_q a_{t-q}$

ARMA:

$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$

Where,

$Z_t$: Variable that follows time series

$\delta$ : Constant

$a_t$ : white noise error

$\mu$: the mean

$\phi_1 \ldots \phi_p$ , $\theta_1 \ldots \theta_q$ : model parameters

The multiplicative seasonal ARIMA model generalization (p, d, q) $(P,D,Q)_s$ is considered as an extension of the method to series in which a pattern repeats seasonally over time . Where:

p : the order of AR model

q : the order of MA model

d: the order of difference

P: the order of seasonal Process AR

Q: the order of seasonal Process MA

D: the order of seasonal difference

s: the length of seasonal period

By considering the patterns of the autocorrelations and the partial autocorrelations, a reasonable model for the data can be estimated.

An Autocorrelation Function (ACF) and partial autocorrelation (PACF) are measures of the correlation between current and past series values and indicate which past series values are most useful in predicting future values.

The x axis of the ACF plot indicates the lag at which the autocorrelation is computed; the y axis indicates the value of the correlation (between −1 and 1). A positive correlation indicates that large current values correspond with large values at the specified lag; a negative correlation indicates that large current values correspond with small values at the specified lag. The absolute value of a correlation is a measure of the strength of the correlation, with larger absolute values indicating stronger relationships.

There are three primary stages in building a Box-Jenkins time series model:

- **Model Identification**: determining the order of the model required (p,d,and q).

- **Model Estimation**: estimating the parameters of different model components.
- **Model Validation:** Checking for inadequacies by considering the autocorrelations of the residual series (the series of residual, or error, values).

Once a suitable model is selected, the software program that fits Box-Jenkins models used to generate forecasts and associated probability limits.

The following steps describe the overall calibration and validation procedure (Figure 1):

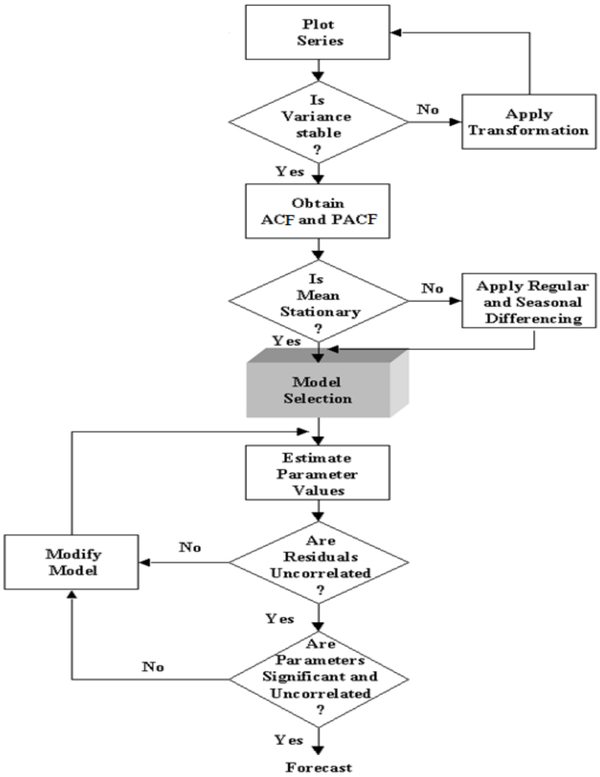


**Figure 1: Box-Jenkins modeling approach**

1. Set of values observed sequentially through time (Plotting the series) to find out whether the series shows any upward or downward trend and whether some sort of data transformation might simplify analysis

- If series exhibits a trend, remove a deterministic trend by difference the series. If the differenced series still does not appear stationary, we would have to difference it again.
- If the probability distribution is the same for all starting values of t. A statistical process is stationary .This implies that the mean and variance are constant for all values of t.
- If mean and variance are not constant, transform a non-stationary series into a stationary one by a log transformation of the series.

2. Obtain the Autocorrelation function (ACF) and partial Autocorrelation function (PACF) to find out whether any kind of seasonal pattern is apparent and guess a reasonable model for the data.

- If the (ACF) show high values at fixed interval, series including seasonal autoregressive term.
- If the (ACF) no decay to zero, series is not stationary. Apply differencing.

3. Select a suitable model; ARMA or ARIMA or SARIMA.

4. Estimating the parameters for the Box-Jenkins models is a quite complicated non-linear least squares and maximum likelihood estimation problem. For this reason, the parameter estimation should be left to a high quality software program that fits Box-Jenkins models.

5. Check the model by study the autocorrelation plots of the residuals to see if further large correlation values can be found.

- If all the autocorrelations and partial autocorrelations are small, the model is considered adequate and forecasts are generated.
- If some of the autocorrelations are large, the values of p and/or q should be adjusted and the model is re-estimated.

6. Estimate the significance and uncorrelated of the selected model parameter.

- If model unsignifical, the values of p and/or q should be adjusted and the model is re-estimated.
- If model significal, generating forecasting in a validation period, for which you have data, or forecasting into the future.

# 5. MODEL DEVELOPMENT AND EVALUATION

## 5.1 Daily Traffic Volumes

Several multiplicative ARIMA (p, d, q) models were examined. The daily traffic volumes were calculated for the period of 2008 to 2012 for 14 count locations where one count station is installed for each location and each direction. Daily volume data of four years (from January 2008 to December 2011) were used for models development. While, the data of the period between January 2012 and December 2012 was used for models testing and evaluation. For the two-lane-two-way highways, the total traffic volume of both directions was used as the dependent variable. On the other hand, for divided highways, the traffic volume of each direction was handled separately. A total of 26 models were developed using SARIMA models as in Table 3.

**Table 3: Results of Daily traffic volumes models using SARIMA (p, d,q) (P, D, Q)$_S$**

| Highway Class | Road name | DIRECTION | SARIMA Model(p,d,q) (P,D,Q)$_S$ | *$R^2$ | ** MAPE | ***Ljung-box Sig. |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| Two-lane-2 Way Agric | Sharkia-Ismailia | T | $(7,1,7)(1,0,0)_{365}$ | 0.79 | 4.9 | 0.81 |
| | Minya-Asyut | T | $(7,1,7)(1,0,0)_{365}$ | 0.71 | 7.1 | 0.76 |
| Four-lane Divided - Agric | Gharbia-Beheira | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.80 | 6.4 | 0.09 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.88 | 6.9 | 0.07 |
| | Qaliubia-Sharkia | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.77 | 5.4 | 0.68 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.75 | 5.3 | 0.73 |
| | Gharbia-Dakahlia | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.80 | 4.4 | 0.18 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.79 | 5.6 | 0.25 |
| | Mit Ghamr - Aga | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.77 | 6.3 | 0. 27 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.84 | 5.8 | 0.38 |
| | Beheira-Alexandria | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.78 | 4.6 | 0.14 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.85 | 4.9 | 0.08 |
| Six-lane Divided - Agric | Defra-KafrElziat | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.88 | 6.0 | 0.18 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.90 | 6.4 | 0.23 |
| | Menoufia-Gharbia | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.75 | 5.5 | 0.14 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.72 | 6.6 | 0.27 |
| | Cairo- Qaliubia | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.77 | 4.9 | 0.21 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.83 | 4.3 | 0.12 |
| Four-lane Divided - Desert | Cairo- Ismailia | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.83 | 4.7 | 0.20 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.79 | 5.6 | 0.17 |
| | Cairo - Suez | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.62 | 4.9 | 0.21 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.58 | 4.8 | 0.09 |
| | Sadat City - Wadil Natrun | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.86 | 5.4 | 0.46 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.83 | 6.1 | 0.37 |
| | Giza- Faiyum | 1 | $(7,1,7)(1,0,0)_{365}$ | 0.76 | 5.6 | 0.37 |
| | | 2 | $(7,1,7)(1,0,0)_{365}$ | 0.77 | 4.7 | 0.49 |

*$R^2$: model's coefficient of determination

**MAPE: mean absolute percentage error

***Ljung-box sig.: is a type of statistical test of whether any of a group of autocorrelations of a time series are different from zero.

A general Seasonal ARIMA model (7,1,7) $(1,0,0)_{365}$ was found appropriate for all locations as shows in Table (3). The model consists of two parts; a non-seasonal part (7,1,7) and a seasonal part $(1,0,0)_{365}$. The non-seasonal part of the model is not stationary (trend is present), so differencing was necessary d=1 (to transform the data to be stationary; the series varies randomly around its mean level). The order of autoregressive and moving average (p=7& q=7), means that the pattern of daily traffic volumes data is repeated weekly (from Saturday to Friday). The order of the AR model defines the number of past days that should be included in predicting the next day, where the order of moving average is a linear regression of the current value of the series against current and previous white noise error terms. For the seasonal part $(1,0,0)_{365}$, it was found that seasonality in daily traffic volumes data which has high values always tends to occur in some particular days and low values tend always to occur in other particular days. In this case, S = 365 (days per year) is the span of the periodic seasonal behavior.

The last three column on Table (3) show the model statistics ($R^2$, MAPE, Ljung-box test). Diagnostic checking and accuracy test proved that this model is adequate and fits the data well and hence can be used for forecasting daily traffic volumes. In addition, the models have $R^2$ values that range from 0.58 to 0.90 which implying that about 58% to 90% of the variations in traffic volumes data have been accounted for by the model. Also, the value of mean absolute percentage error (MAPE) ranges from 4.3 to 7.1%. According to Lewis (1982), the level of accuracy for the MAPE test is divided into four stages as shown in Table 4.

Table 4: level of accuracy for the MAPE test

| MAPE value | Level of accuracy |
|---|---|
| MAPE < 10% | very Accurate |
| 10% < MAPE < 20% | Accurate |
| MAPE < 50% 20% < | Medium |
| 50% < MAPE | Less accurate |

Source: Lewis (1982, p. 40)

Consequantly ,the models can be considered as "Accurate very ".The p-value of Ljung-Box's is shown in the last column of Table 3. Its value is greater than 5% which indicated that the chosen model is adequate . According to Box-Ljung test (1978) ,The test is applied to the residuals of a time series after fitting an ARIMA model to the data. The test examines autocorrelations of the residuals. If the autocorrelations are very small, the model does not exhibit significant lack of fit. So, it can be concluded that SARIMA (7,1,7) $(1,0,0)_{365}$ is adequate enough to be used for daily traffic volumes time series analysis.

## 5.2 Weekly Traffic Volumes

Several models were examined to fit weekly traffic volumes including a multiplicative SARIMA (p, d, q) $(P, D, Q)_s$ model. Similar to the analysis of daily traffic volumes, the data for the period between 2008 and 2012 are used in models development. The weekly traffic volumes prediction models are important for comparing different highway classes in terms of seasonality, autoregressive, order of seasonal and non-seasonal differencing. Results of estimating weekly traffic volumes prediction models are presented in Table 5.

**Table 5: Results of weekly traffic volumes models using SARIMA (p,d,q)$(P, D, Q)_s$**

| Highway Class | Road name | DIRECTION | SARIMA Model(p,d,q) $(P,D,Q)_S$ | $*R^2$ | **MAPE | ***Ljung-box Sig. |
|---|---|---|---|---|---|---|
| Two-lane Two Way Agric. | Sharkia- Ismailia | T | $(1,0,0)(0,1,0)_{52}$ | 0.83 | 3.4 | 0.06 |
| | Minya-Asyut | T | $(1,0,0)(0,1,0)_{52}$ | 0.85 | 2.8 | 0.62 |
| Four-lane | Gharbia- Beheira | 1 | $(1,1,0)(0,1,0)_{52}$ | 0.85 | 4.1 | 0.08 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Divided Agric. | | 2 | $(1,1,0)(0,1,0)_{52}$ | 0.98 | 2.7 | 0.30 |
| | Qaliubia- Sharkia | 1 | $(1,1,0)(0,1,0)_{52}$ | 0.85 | 2.0 | 0.14 |
| | | 2 | $(1,1,0)(0,1,0)_{52}$ | 0.83 | 2.3 | 0.30 |
| | Gharbia- Dakahlia | 1 | $(1,1,0)(0,1,0)_{52}$ | 0.89 | 4.0 | 0.28 |
| | | 2 | $(1,1,0)(0,1,0)_{52}$ | 0.79 | 4.4 | 0.30 |
| | Mit Ghamr - Aga | 1 | $(1,1,0)(0,1,0)_{52}$ | 0.69 | 4.7 | 0.07 |
| | | 2 | $(1,1,0)(0,1,0)_{52}$ | 0.78 | 4.4 | 0.07 |
| | Beheira- Alexandria | 1 | $(1,1,0)(0,1,0)_{52}$ | 0.63 | 5.2 | 0.07 |
| | | 2 | $(1,1,0)(0,1,0)_{52}$ | 0.73 | 5.4 | 0.09 |
| Six-lane Divided Agric. | Defra-KafrElziat | 1 | $(1,1,0)(1,0,0)_{52}$ | 0.78 | 7.0 | 0.14 |
| | | 2 | $(1,1,0)(1,0,0)_{52}$ | 0.75 | 8.0 | 0.06 |
| | Menoufia- Gharbia | 1 | $(1,1,0)(1,0,0)_{52}$ | 0.69 | 4.0 | 0.06 |
| | | 2 | $(1,1,0)(1,0,0)_{52}$ | 0.52 | 5.0 | 0.09 |
| | Cairo- Qaliubia | 1 | $(1,1,0)(1,0,0)_{52}$ | 0.68 | 5.2 | 0.08 |
| | | 2 | $(1,1,0)(1,0,0)_{52}$ | 0.67 | 4.5 | 0.09 |
| Four-lane Divided Desert | Cairo- Ismailia | 1 | $(1,1,0)(1,0,0)_{52}$ | 0.64 | 4.1 | 0.71 |
| | | 2 | $(1,1,0)(1,0,0)_{52}$ | 0.63 | 5.2 | 0.40 |
| | Cairo – Suez | 1 | $(1,1,0)(1,0,0)_{52}$ | 0.64 | 3.9 | 0.57 |
| | | 2 | $(1,1,0)(1,0,0)_{52}$ | 0.78 | 3.9 | 0.21 |
| | Sadat City - Wadil Natrun | 1 | $(1,1,0)(1,0,0)_{52}$ | 0.87 | 3.3 | 0.09 |
| | | 2 | $(1,1,0)(1,0,0)_{52}$ | 0.91 | 2.7 | 0.79 |
| | Giza- Faiyum | 1 | $(1,1,0)(1,0,0)_{52}$ | 0.77 | 3.5 | 0.12 |
| | | 2 | $(1,1,0)(1,0,0)_{52}$ | 0.79 | 3.0 | 0.14 |

Different model specifications were found appropriate for different types of rural highways data. For agriculture roads; seasonal ARIMA models $(1, 0, 0)(0, 1, 0)_{52}$ appeared appropriate for two-lane two way highways. Meanwhile, seasonal ARIMA models $(1,1,0)(0,1,0)_{52}$ best fitted Four-lane highways, and finally seasonal ARIMA models $(1,1,0)(1,0,0)_{52}$ showed the best fit to Six-lane highways.

For desert roads, seasonal ARIMA models $(1,1,0)(1,0,0)_{52}$ were found to be appropriate for Four lane highways. Again, all models showed very good to the data where the coefficient of determination ($R^2$) ranged between 0.52 and 0.91 and the MAPE was between 2.0 and 7.0%.

For the two-way-two-lane highways, the model consists of a non-seasonal part (1,0,0) and a seasonal part $(0,1,0)_{52}$. The non-seasonal part of weekly traffic volumes data does not show any trend (d=0). The order of the autoregressive AR model (P=1) shows that only the weekly volume of the previous week should be included to predict next week's volume. For the seasonal part, it was found that a seasonality difference (D=1) from the previous year may be about the same for each week of the year giving us a stationary series (trend isn't present and series varies randomly around its mean level). In this case, S = 52 (weeks per year) is the span of the periodic seasonal behavior.

For Four-lane divided agricultural road, the model SARIMA model consists of two part non seasonal part (1, 1, 0) and seasonal part $(0, 1, 0)_{52}$. The non-seasonal part of weekly traffic volumes data are not stationary (trend is present), so differencing appear necessary d=1, the order of the autoregressive AR model (p =1) shows the number of past weeks should be included to predect next week traffic volumes. For seasonal part it is found that there is seasonality differences because, removing trend doesn't mean that we have removed the seasonality. In this case, S = 52 (weeks per year) is the span of the periodic seasonal behavior.

For Six-lane divided Agriculture road, and Four-lane divided Desert road, the model consists of two part non seasonal part (1,1,0) and seasonal part $(1,0,0)_{52}$. The non-seasonal part of weekly traffic volumes data are not stationary (trend is present), so differencing appear necessary (d=1), the order of autoregressive AR model (p =1) shows the number of past weeks should be included to predect next week traffic volumes. For seasonal part it is found that the sreies become stationary after non seasonal differencing, so seasonal differencing does not appear necessary. The order of the seasonal autoregressive SAR model(p =1) shows there is seasonality in weekly traffic volumes data for which high values tend always to occur in some particular weeks and low values tend always to occur in other particular weeks traffic volumes. In this case, S = 52 (weeks per year) is the span of the periodic seasonal behavior.

For Agriculture road, It can be concluded that the three types of lane groups (2 lane,4 lane, 6 lane) have autoregressive (AR) term with order (p =1),which means that the data of past weeks should be included to predect a next week traffic volumes.The weekly traffic volumes data for two way two lane agriculture road are stationary (d=0),but it shows trend for four and six lane road group class.While the seasonal differencing equal zero (D=0) for six lane divided Agriculture road, it is equal one (D=1) for two lane and four lane divided agriculture road,which mean that non seasonal differencing was not enough for series to be stationary in case of four lane divided Agriculture road. The result are summarized in Table 6.

Table 6: Comparison of weekly traffic volume models for different lane group of agriculture highways

| | **Two-lane Two Way** | **Four-lane Divided** | **Six-lane Divided** |
|---|---|---|---|
| Order of autoregressive AR(p) | 1 | 1 | 1 |
| Order of non seasonal differencing (d) | 0 | 1 | 1 |
| Order of moving average MA(q) | 0 | 0 | 0 |
| Order of seasonal autoregressive SAR (P) | 0 | 0 | 1 |
| Order of seasonal differencing (D) | 1 | 1 | 0 |
| Order of seasonal moving average SMA(Q) | 0 | 0 | 0 |

For four-lane divided Agriculture and desert roads, the non-seasonal part of the model have the same order of autoregressive AR term (p=1), the same order of moving average MA term (q=0),and the same order of non- seasonal differencing (d=0).The various between two models appear in term of seasonal autoregressive (SAR) and differencing (D). Four-lane divided agriculture roads have a seasonal differencing (D=1), which mean that seasonality differences from the previous year may be about the same for each week of the year. Four-lane divided desert roads have a seasonal autoregressive (SAR) term (P=1), which mean that there is seasonality in weekly traffic volumes data for which high values tend always to occur in some particular weeks and low values tend always to occur in other particular weeks traffic volumes. In this case, S = 52 (weeks per year) is the span of the periodic seasonal behavior. The results are summarized in Table 7.

Table 7: Comparison of weekly traffic volumes models for Four-lane divided agriculture and desert roads.

| | Four-lane Divided agriculture roads | Four-lane Divided desert roads |
|---|---|---|
| Order of autoregressive AR(p) | 1 | 1 |
| Order of non seasonal differencing (d) | 1 | 1 |
| Order of moving average MA(q) | 0 | 0 |
| Order of seasonal autoregressive SAR (P) | 0 | 1 |
| Order of seasonal differencing (D) | 1 | 0 |
| Order of seasonal moving average SMA(Q) | 0 | 0 |

## 6. SUMMARY AND CONCLUSIONS

In this paper, ARIMA time series models were fitted to predict both daily and weekly traffic volumes on Egyptian National Highways. Desert and agricultural national highways models are compared for more understanding of traffic patterns. The main outcomes of our results can be summarized as follows:

- Seasonal ARIMA model (7, 1, 7) × $(1, 0, 0)_{365}$ is found to be appropriate to predict the daily traffic volumes data for Egyptian national highways.
- The length of seasonal period for daily traffic volumes are 365 (days per year) which is the span of the periodic seasonal behavior.
- As for weekly traffic volumes, different model specifications are appropriate for different types of rural highways. Seasonal ARIMA models (1, 0, 0) $(0, 1, 0)_{52}$ found to be appropriate to predict weekly traffic volumes for Two-lane Two Way agricultural roads. However, Seasonal ARIMA models $(1,1,0)(0,1,0)_{52}$ were found to be better fit the weekly traffic volumes for Four-lane agricultural roads, while, seasonal ARIMA models $(1,1,0)(1,0,0)_{52}$ appeared more appropriate to predict the weekly traffic volumes for six-lane agriculture roads data and four lane desert roads data.
- The length of seasonal period for weekly traffic volumes are 52 (weeks per year) which is the span of the periodic seasonal behavior.
- For agricultural road, the three types of lane groups (2 lane,4 lane, 6 lane) have autoregressive (AR) term with order (p =1),which means that the data of past weeks should be included to predect a next week traffic volumes..
- The weekly traffic volumes data for two way two lane Agriculture road are stationary -traffic volumes data does not show any trend- (d=0), but it shows trend for four and six lane road group class.
- The seasonal differencing equal one (D=1) for Two lane and four lane divided Agriculture road, which mean that seasonality differences from the previous year may be about the same for each week of the year.
- seasonal differencing equal zero (D=0) for six lane divided Agriculture road, which mean that non seasonal differencing was enough for series to be stationary.
- For Four-lane divided agriculture and desert roads, the non seasonal part of the model have the same order of autoregressive AR term (p=1), the same order of moving average MA term (q=0), and the same order of non- seasonal differencing (d=0).
- The various between two models appear in term of seasonal autoregressive (SAR) and differencing (D).
- Four-lane divided agriculture roads have a seasonal differencing (D=1).
- Four-lane divided desert roads have a seasonal autoregressive (SAR) term (P=1).

## 7. REFERENCES

A & Karlaftis, M.G. (2003) , “A multivariate state space approach for urban traffic flow modeling and prediction engineering”, Transportation Research Part C 11, pp. 121–135.

Benjamin, Julian (1986), “A Tine-Series Forecast of Average Daily Traffic Volumes,” Transportation Research, 20A(1), 51-60.

Bidisha Ghosh, Biswajit Basu, and Margaret O’Mahony. (2003) , “Time-series modelling for fore- 15 casting vehicular traffic flow in dublin”. In 84th Transportation Research Board Conference, 16 National Academies, Washington D.C, pages 1–22.

Box, G., G. Jenkins and G. Reinsel, (1994), “Time Series Analysis: Forecasting and Control”, Prentice Hall.

G. M. Ljung; G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". Biometrika(1978)65(2):297/303. doi: 10.1093 /biomet/65.2.297.

Lewis, C.D. (1982). “International and Business Forecasting Methods”. Butterworths, London.

M. S. Ahmed and A. R. Cook(1979), “Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques,” Transportation Research Record, No. 722, pp. 1-9.

Nihan N. and K. Holmesland (1980) , “Use of the Box and Jenkins Time Series Techniques in Traffic Forecasting ”,Transportation, 9: 125-143.

Sabry, Abd-El-Latif and Badra (2007), “Comparison between Regression and Arima Models in Forecasting Traffic Volume”. Australian Journal of Basic and Applied Sciences, 1(2): pp.126-136.

Williams, B.M., P. Durvasula and D. Brown. “Urban Freeway Traffic Flow Prediction: Application of Seasonal ARIMA and Exponential Smoothing Models”, Transportation Research Record: Journal of the Transportation Research Board, No. 1644, TRB, National Research Council, Washington D. C., 1998.