Comparison of some forecasting models for air travel movements at Riyadh international airport, Saudi Arabia

Abdullah O. Ba-Fail, Seraj Y. Abed and Sajjad M. Jasimuddin
Industrial Eng. Dept., King Abdulaziz University, PO Box 80204, Jeddah 21489, Saudi Arabia

Airport is the nucleus of air transport system. Accurate forecasting of aviation demand is required for ensuring the improved air transportation facilities and systems. This paper compares three forecasting methods using the data of arrival, departure and movement at Riyadh International Airport. A number of tools are used to evaluate the adequacy of each of the models proposed. Among the forecasting models suggested, it is postulated that a polynomial model of third degree is the most suitable one. The model may be used to plan future expansion for the airport activities. Yearly forecasts are generated for a six-year planning horizon.

يعتبر المطار نواة نظام السفر بالجو، ولذا فان النظم المتطورة للنقل الجوى تتطلب التخطيط الجيد من خلال التنبؤ الدقيق باحتياجات الطيران. وتقدم هذه الدراسة ثلاثة نماذج للتنبؤ باستخدام بيانات الوصول والمغادرة في مطار الرياض الدولي، وقد استخدمت عدة وسائل لتقويم ملائمة كل من النماذج المستخدمة، وأشارت الاختبارات أن النموذج المتعدد الحدود للدرجة الثالثة هو الأكثر ملائمة لهذا التنبؤ، وقد استخدم هذا النموذج في التنبؤ المطلوب لوضع خطة سته سنوات للطيران، وكذلك يمكن استخدام هذا النموذج في تخطيط التوسعات المستقبلية لأنشطة المطار.

Keywords: Air transportation, Airport, Forecasting models, Aviation demand

1. Introduction

Air transportation has recently become the fastest and dominant form of long-haul public transportation. Transportation by air has likewise a significant impact on the economies of the world [1]. A significant increase in the number of airline passengers is expected in the next decade: 64.4% increase to 743.5 million people by the year 2001 [2]. The slow increase of airport capacity may contribute to air traffic congestion at major airports [3]. Air traffic congestion together with bad weather account for 93% of the delays in 1989 [4]. Such delays are expected to escalate unless an expansion in airport capacity, an increase in the efficiency of using existing airport capacity, and modernization of air control facilities are forthcoming.

The sudden boom in the development of air commerce and aviation has increased the size and coastlines of airport facilities and installations. The rapid developments in aviation industry produce an accelerated rate of technological obsolescence, which in turn creates a continuing need for investment to provide facilities in time to accommodate new demands. Improved air transportation facilities and systems require basic airport planning that may be achieved through accurate forecasting of aviation demand.

Riyadh is endowed with numerous opportunities for air travel by virtue of its status as the capital city of the Kingdom of Saudi Arabia, its geographical location as the center of air routes to major cities of the Kingdom, proximity to the two holy mosques of the Muslim world and its rapid multifaceted development. Consequently, Riyadh International Airport (RIA) is the busiest and the largest in the Kingdom. The primary objective of this paper is to develop suitable models for purposes of forecasting yearly arrivals, departures and movements at the RIA.

The intent of the study is to provide the concerned departments in the government and in the services industry such as the Presidency of Civil Aviation and Saudi Arabian Airlines with a planning tool. This tool may help, for example, to study the proper sizing of airports facilities such as gate requirements, apron size, terminal capacity, etc.

2. Literature review

There has been a large amount of effort devoted to the problem of analyzing and forecasting air passenger movements. Some of the studies conducted in this area will be reviewed in this section.

Poore [5] conducted a study to test the hypothesis that forecasts of the future demand for air transportation offered by airplane manufacturers and aviation regulators are reasonable and representative of the trends implicit in actual experience. The tests compared forecasts issued by Boeing, McDonnell Douglas, Airbus Industry and the International Civil Aviation Organization with actual data and the results of a baseline model of the demand for Revenue Passenger Kilometers (RPKs). The model is a combination of two equations describing the RPKs demanded by the high- and the low-income sectors respectively. Variations in the RPKs demanded by the high-income group are related to changes in income per capita. Variations in the RPKs demanded by the low-income segment are related to changes in the population size. The model conformed with the assumptions and conditions for appropriate use of regression analysis. The model also appeared to be in conformity with historical demand.

Ghobrial [6] conducted a study using an econometric model, and estimated the aggregate demand of an airline. The demand was expressed in terms of airline network structure, operating characteristics and firm-specific variables. A number of model formulations with different combinations of explanatory variables were estimated using the two-stage-least-square procedure. The results suggest that the airline aggregate demand is elastic with respect to yield, and inelastic with respect to network size and hub dominance. Some implications regarding airline network expansion and hubbing are discussed.

Saudi Arabian Bechtel Company [7] conducted a study to update the traffic forecasts and planning assumptions for the New Riyadh International Airport. Four economic variables related to air traffic activities were chosen for the study. These causal variables were gross domestic product, government appropriations, project appropriations and import of goods and services. Each variable was correlated with the annual domestic and international passenger value at the 6th old Riyadh airport. For international passengers, the correlation coefficient varied between 0.97 and 0.993 and the best results were obtained with the imports C.I.F. For domestic passengers, the correlation coefficient varied between 0.936 and 0.997 and the best results were obtained with government appropriations.

The literature review indicates that few studies refer to forecasting passenger movements through airport issue. There is, however, no study on air travel movements through the Saudi airports. This paper aims at developing models to forecast passengers movements from the RIA so as to increase airport capacity and reduce air traffic congestion.

3. Data analysis

Data on passenger arrivals, departures and movements at the RIA in Saudi Arabia during January 1975 through December, 1996 were obtained from the Presidency of Civil Aviation, the Saudi aviation authority [8, 9]. These yearly data were used to study the historical trends over that period and to develop suitable models so as to forecast future trends in passenger arrivals, departures and movements at this airport.

3.1. Trends and rates of growth

As in all forecasting techniques, attempts are made here first to find whether data are trended, seasonal, irregular or volatile. The data as obtained from the Presidency of Civil Aviation are plotted in figs. 1-a, 1-b and 1-c. The distributions were found to be negatively skewed. Further investigation shows that all the series are trended and non-seasonal

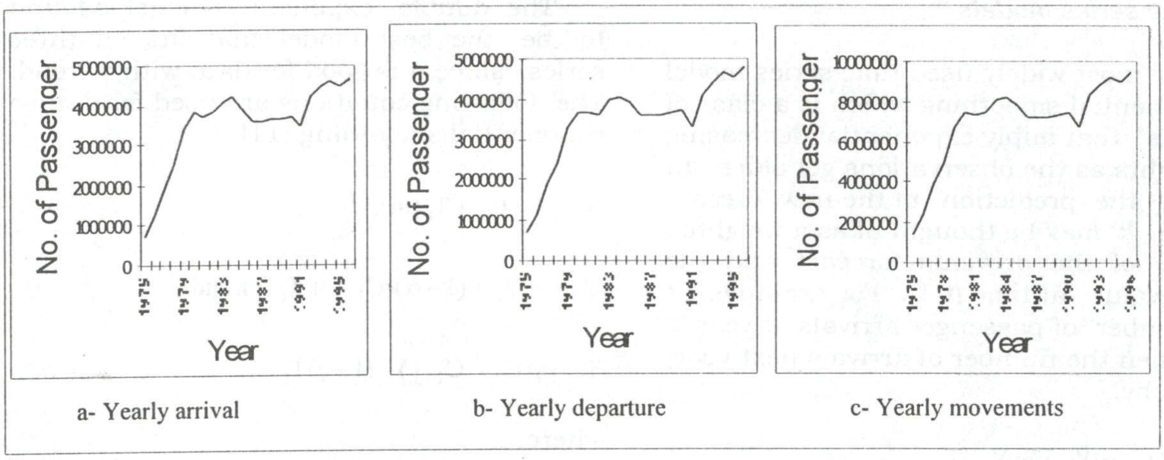


Fig. 1. Plots of original data.

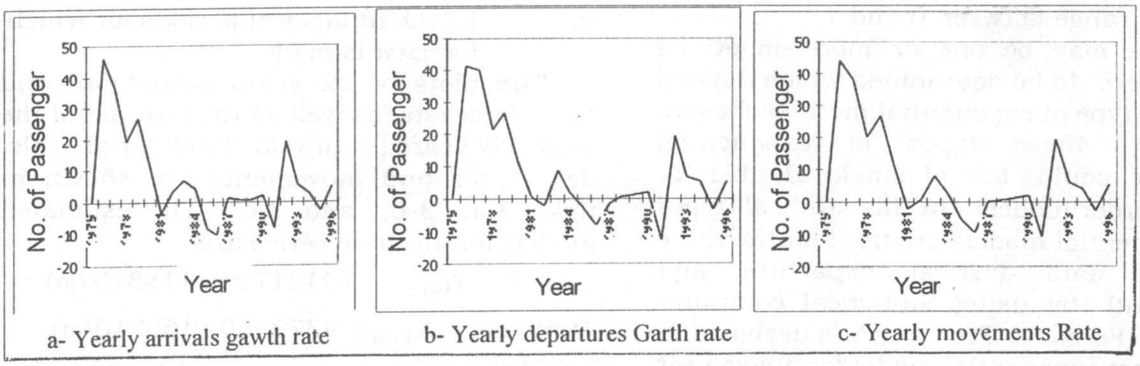


Fig. 2. Percentage growth rate.

with trend-cycles and irregular measures of (54.63% and 45.37%), (57.34% and 42.66) and (56.04% and 43.96) for arrivals, departures and movement, respectively.

The percentage growth rates of arrivals, departures and movement are shown in figs. 2-a, 2-b and 2-c. During late 70's the growth rate was very high. It starts from early 80's to mid 90's to fluctuate.

3.2. Statistical parameters

Statistical parameters are of great value for describing the characteristics of data. Table 1 gives a summary description of some important statistics for arrivals, departures, and movements of air passengers over the 22 years period from 1975 to 1996 at the RIA. It shows that, on an average, the number of annual arrivals and departures is about 2.9 million with a sample standard deviation of about 1 million. The skewness (a measure of asymmetry) with a negative sign indicates that the distributions of arrival, departure and movement are skewed to the left (i.e. the mean is smaller than the median). On the other hand, the value of the kurtosis (the measure of the heaviness of the ends of a distribution) is less than 1. It indicates that the data has a light-tailed distribution. It is quite evident that annual arrivals, departures and movement of air passengers increased, respectively, from 411097, 385201 and 796298 in 1975 to 4131782, 4372279 and 8504061 in 1996.

In this study, both the modeling methods (time series and explanatory models) are used, in addition to the Box-Jenkins model. Although a span of more than 50 years is required to use Box-Jenkin model, it is used here only to see how the model behaves in comparison to the others.

3.3. Time series models

The most widely used time series model is exponential smoothing which is a class of methods that imply exponential decreasing of weights as the observations get older. In general, the prediction of the new forecast at time 't' may be thought of as a weighted average of the previous forecast and the actual value at time (t-1). For example, let the number of passenger arrivals at year 't' be y_t, then the number of arrivals next year, is given by:

$$\hat{Y}_{t+1} = (1-\alpha)\hat{Y}_t + \alpha Y_t$$

where '∝' is the smoothing parameter and its value range between 0 and 1.

There may be one or more smoothing parameters to be determined which depend on the type of exponential method chosen. However, these types of exponential methods could be of single, double, or higher order models. In this study all types of exponential models are tried on the three sets of data (arrival, departure, and movement) by using Statistical computer package Forecast Pro which is designed to fit the best exponential model for a given set of data and determine their smoothing parameters [10].

The double exponential model is found to be the best model that fits all three series, since it is good for data with a trend. The following equations are used for double exponential smoothing [11]:

$$\hat{Y}_{t+m} = C_t + T_t(m),$$

$$C_t = \alpha Y_t + (1-\alpha)(C_{t-1} + T_{t-1}), \text{ and}$$

$$T_t = \gamma(C_t - C_{t-1}) + (1-\gamma)T_{t-1},$$

where,

C_t is the level of the series at time t,
T_t is the trend of the series at time t,
α and γ are the smoothing constants,
m is the number of periods for which forecast is made

The plots of 22 years actual data and their forecasts, as well as the forecast of the next six years (until year 2002) for arrivals, departures and movements are shown in figs. 3-a, 3-b, and 3-c. The estimated models for all three series are:

$$\text{Arrival: } \hat{Y}_{t+m} = 4131700 + 118870\,(m)$$

$$\text{Departures: } \hat{X}_{t+m} = 4372200 + 165910(m)$$

$$\text{Movements: } \hat{Z}_{t+m} = 8504000 + 294590(m)$$

Table 1
Descriptive statistics for yearly arrivals, departures and movements of air passengers

Statistical performance measures	Arrivals	Departures	Movements
Mean	2906500.591	2848962.955	5755453.545
Standard deviation	1007955.672	1020600.890	2026639.089
Median	3137178.000	3073974.500	6198792.500
Skewness	-1.189842685	-1.044185012	-1.123205989
Kurtosis	0.976371894	0.810505680	0.891137447
Range	3720685	3987078	7707763
Minimum	411097	385201	796298
Maximum	4131782	4372279	8504061
Total	63943013	62676965	126619978
Number of years	22	22	22

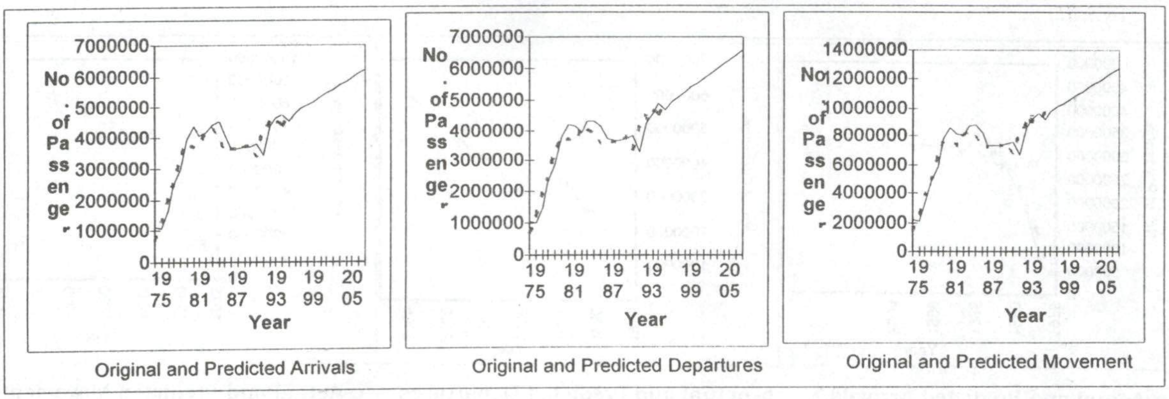


Fig. 3. Plots of 22 years actual data and their forecasts based on exponential smoothing methods.

Table 2
Exponential models for arrivals, departures and movements of air passengers

Model parameters	Arrival	Departure	Movement
Level	4131700	4372200	8504000
Trend	118870	165910	294590
Statistics			
R-square	0.9363	0.9284	0.9328
Adjusted R- square	0.9331	0.9248	0.9295
Durbin-watson	1.542	1.714	1.758
Forecast error	260600	279900	538200
MAPE	10.17	10.96	10.60
RMES	248500	266900	513200
MAD	195100	228000	402800
Forecast for six years ahead			
1997	4250600	4538096	8798639
1998	4369471	4704004	9093229
1999	4488342	4869911	9387819
2000	4607213	5035819	9682410
2001	4726084	5201727	9977000
2002	4844955	5367634	10271591

Performance measuring statistics for all the models and the forecasted values for next six years based on the estimated models are given in table 2. Error analysis does not indicate any severe problems with the models. Coefficient of determination (R^2) for all three models is above 92%. That is, around 92% of the variability in the data are explained by the double exponential models. Durbin-Watson statistic (which tests whether there is any autocorrelation exists in the residuals) for all three models are between 1.54 and 1.76. This is an indication that the errors are random. The Mean Absolute Percentage Error (MAPE) for all three models are approximately around 10% which is acceptable.

3.4. Explanatory models

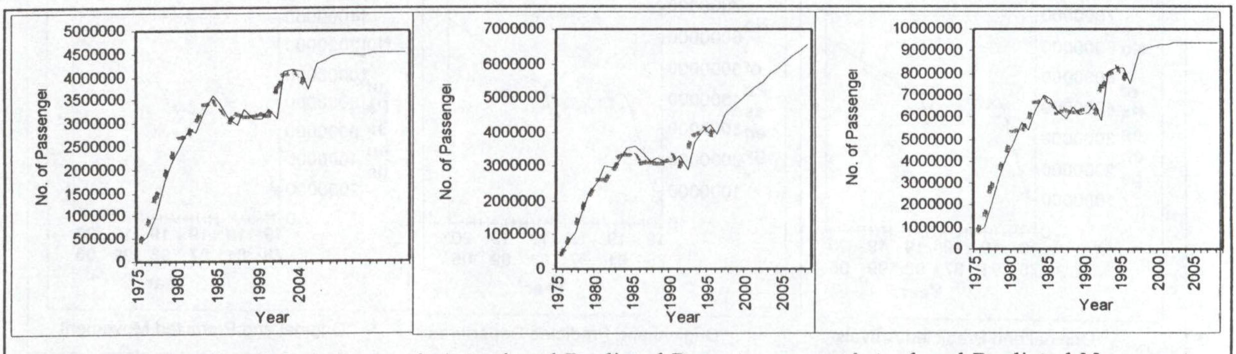
Regression analysis is the most widely used casual model for investigating and modeling relationship between two or more variables - the dependent variables (forecasted values) and the independent explanatory variables. In this study, time is the independent variable and the dependent variable is either the number of arrivals, departures or movements. Here three different models are used; the first one is linear and the other two are non-linear. The model forms are as follows:

Linear: Model I $Y_t = a + b\,t$

Non linear : Model II $Y_t = a + b\,t + c\,t^2$

Non Linear Cubic: Model III $Y_t = a + b\,t + c\,t^2 + d\,t^3$

All these models were performed separately on the number of arrivals, departures and movements. By using



a-Actual and Predicted Arrivals b-Actual and Predicted Departures c-Actual and Predicted Movement

Fig. 5. Plots for 22 years actual data and their forecasts based on Box-Jenkins model.

Table 5
Box-Jenkins models for arrivals, departures and movements of air passengers

Model parameters	Arrival	Departure	Movement
Model	ARIMA(1,1,0)	ARIMA(1,1,1)	ARIMA(1,1,0)
Coeff.	0.5919	0.9937, 0.2341	0.5267
Std. Error	0.1711	0.0196	0.1862
t-Statistics	3.4589	50.64	2.8281
Significance	0.9975	1.0	0.9896
Statistics			
R-square	0.92	0.91	0.90
Adjusted R-square	0.91	0.90	0.89
Durbin-Watson	2.123	2.064	2.123
Sjung-Box(10)	8.723 P=0.4414	9.113 P=0.5731	7.539 P=0.3262
Forecast Error	254900	277800	553600
BIC	267500	305400	580900
MAPE	8.1	6.5	8.8
RMSE	248800	264200	540200
MAD	206100	194800	436100
Forecast for six years ahead			
1997	4290456	4560972	8910879
1998	4384379	4748471	9125154
1999	4439973	4934783	9238010
2000	4472881	5119915	9297453
2001	4492359	5303874	9328762
2002	4503889	5486669	9345254

Table 6
Performance measures based on all three methods for arrivals, departures and movements of air passengers

Statistics	Arrivals			Departures			Movements		
Model	EXP	REG	B-J	EXP	REG	B-J	EXP	REG	B-J
R-square	0.88	0.94	0.83	0.89	0.95	0.84	0.89	0.94	0.84
Adjusted R-square	0.87	0.93	0.83	0.88	0.94	0.83	0.88	0.93	0.83
MAPE	9.9	6.5	8.0	10.3	7	8.4	10.1	6.6	8.1
RMSE	3.5E5	2.4E5	3.4E5	3.4E5	2.3E5	3.5E5	6.8E5	4.6E5	6.7E5
MAD	2.6E5	2.1E5	2.5E5	2.7E5	1.9E5	2.5E5	5.2E5	4.0E5	6.8E5

$$\Phi(B)Y_t = \theta(B)e_t.$$

In the above equation,

Y_t is the sequence of observation

e_t is the set of normally distributed, independent, zero-mean random variables,

p is the order of auto-regressive model, and

q is the order of the moving average model.

The previous equation can be used to model stationary processes. One can also model some types of non-stationary processes by differentiating the original process, Y_t, to obtain a stationary process, W_t in the form,

$$W_t = \nabla^d Y_t,$$

where,

$$\nabla^d = (Y_t - Y_{t-1})^d$$

This results in an Auto-Regressive Integrated Moving Average model (ARIMA) (p,d,q)

$$\Phi(B)\nabla^d Y_t = \theta(B)e_t.$$

Techniques for preliminary identification of the model order depend on the analysis of the auto-correlation and partial auto-correlation functions. The auto-correlation function describes inherent correlation between observations of a time series which separated in time by some lag, k. On the other hand, partial auto-correlation function measures the degree of association between Y_t and Y_{t-k}, when the effects of other time lags (1,2,3,..., k-1) are somewhat partitioned out. Although the three series are reasonably short for utilizing with Box-Jenkins method, attempt has been made here to forecast the required values. By using Forecast Pro, the best ARIMA model for arrivals, departures, and movements are:

Arrivals : $Y_t = 1.5919\ Y_{t-1} - 0.5919\ Y_{t-2} + e_t$

Departures: $X_t = 1.9937\ X_{t-1} - 0.9937\ X_{t-2} - 0.2341\ e_{t-1} + e_t$

Movements: $Z_t = 1.5267\ Z_{t-1} - 0.5267\ Z_{t-2} + e_t$

The plots of the 22 years actual data and their forecast, as well as the forecast of the next six years for arrivals, departures, and movements are shown in figs. 5-a, 5-b, and 5-c. The performance measuring statistics for all the models and the forecasted values for six years ahead based on the estimated models are given in table 5. The correlogram demonstrates that the auto-correlation function dies out quickly. Coefficients of determination (R^2) for all three models are more than 90%. Durbin-Watson values are around 2. MAPE for all three models are around 8 %.

3.6. Comparison of the forecasting techniques based on models statistics

Several performance measures were used to compare all these models with each other in order to select the best one. These measures are Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), and Mean Squares Error (MSE). Table 6 summarizes performance measuring statistics for all different models. Out of the three methods attempted for forecasting arrivals, departures and movements at the RIA, the Non-linear Regression Model (Model III) yields better results. The Box-Jenkins model is the next choice.

Coefficient of determination (R^2) for all the three models range between 88 percent and 94 percent. About 94 percent of the variability in the data are explained by Non-linear regression model. The MAPE for the model is approximately around 10 percent that are acceptable. The non-linear regression model can be marked with "Best thus far."

4. Conclusions

Air passenger movement is an exogenous variable in which the aviation authority has no control. However, it is possible to handle the passengers at airport effectively if the authority has the forecasted data of air passengers arrival, departure and movement. Since there is a

Forecast Pro, the following regression models were calibrated:

Performance measuring statistics for all the models are given in table 3. It is obvious from the table 3 that the third model (non-linear in the cubic form) is the best for arrivals, departures, and movements. Based on these criteria, the non-linear cubic form model will be used for forecasting the arrivals, departures, and movements of air passengers. Using model III, plots of actual data for 22 years, their forecast values and are shown in figure 4 and the forecast for the next six years in table 4.

Arrivals
$Y_t = 1334524 + 136693\ t$
$Y_t = 541329 + 334992\ t - 8621\ t^2$
$Y_t = -485091 + 817707\ t - 59942\ t^2 + 1487\ t^3$

Departures
$X_t = 1239720 + 139933\ t$
$X_t = 539506 + 314986\ t - 7611\ t^2$
$X_t = -577397 + 840255\ t - 63456\ t^2 + 1618\ t^3$

Movements
$Z_t = 2574245 + 276626\ t$
$Z_t = 1080835 + 649978\ t - 16232\ t^2$
$Z_t = -1062488 + 1657963\ t - 123398 t^2 + 3106\ t^3$

* All Parameters are statistically significant.

Table 3
Performance measures for linear and non-linear regression models for arrivals, departures and movements

Statistics	Arrivals			Departures			Movements		
Models	I	II	III	I	II	III	I	II	III
R-square	0.77	0.87	0.96	0.79	0.87	0.97	0.78	0.87	0.97
Adjusted R-square	0.76	0.86	0.95	0.78	0.85	0.96	0.77	0.85	0.96
MAPE	26	16	6.7	26	17	7	26	16.5	6.8
RMSE	4.6E5	3.5E5	1.8E5	4.5E5	3.6E5	1.7E5	9.2E5	7.1E5	3.5E5
MAD	3.7E5	3.1E5	1.6E5	3.5E5	3.1E5	1.4E5	7.2E5	6.2E5	2.9E5

Table 4
Forecast for the next six years using the non linear regression cubic model for arrivals, departures and movements

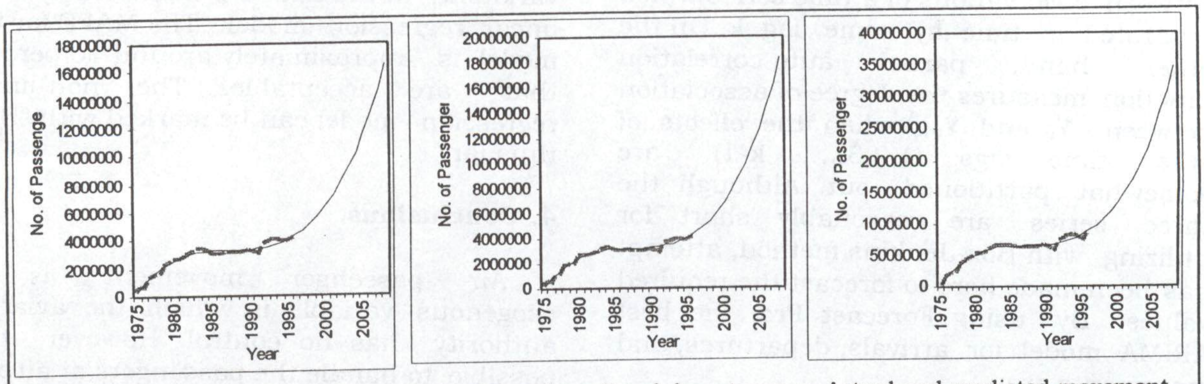
Model parameters	Arrival	Departure	Movement
1997	4711703	4874876	9586578
1998	5177001	5414876	10591878
1999	5736622	6061059	11797681
2000	6399493	6823134	13222628
2001	7174537	7711815	14885352
2002	8070685	8733810	16804498

3.5. Box-Jenkins model

The basic auto-regressive moving average ARMA (p, q) model is of the form

$$Y_t - \Phi_1 Y_{t-1} - \ldots - \Phi_p Y_{t-p} = e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q},$$

which can be written in the short notation form as:



a-Actual and predicted arrivals b-Actual and predicted departures c- Actual and predicted movement

Fig. 4. Plots for 22 years actual and their forecasts based on the cubic model non- linear regression.

number of forecasting models available for such purposes, choice of the appropriate model is crucial and decisive for aviation planning.

This paper compares three forecasting methods using the data of arrival, departure and movement at the RIA in Saudi Arabia. Some forecasting models are proposed for arrivals, departures, and movements at the RIA. A number of tools are used to evaluate the adequacy of each of the models proposed. An examination of the statistical performance measures calculated for all models conclusively points to the usefulness of the non-linear cubic model. Among the forecasting models suggested, it is postulated that a non-linear cubic model is the most suitable one.

Although the proposed model may help project movement of air passengers at the RIA and reduce air traffic congestion, these findings have to be interpreted with some cautions since they are based on yearly traffic intensity data for a period ranging from 1975 to 1996. An attempt can be made to use monthly data for future study. A further extension of the work is possible with the development of a Decision Support System (DSS) based on these models so that the aviation authority could use it for decision making purposes as and when required.

References

[1] A. Siddiqui., Operating levels and airline economics, Saudia World, (1994)

[2] Federal Aviation Authority (FAA), Airman's Information Manual, Federal Aviation Authority, Washington DC. (1990).

[3] F. Sokkar et al., Examination of air traffic flow at a major airport, in: Proceedings of the 1990 Winter Simulation Conference (1990).

[4] J. Thomassia., Flight delays blamed on traffic, Weather, USA Today, Jannet Publication Company (1990).

[5] J. W. Poore., Forecasting the demand for air transportation services. Journal of Transportation Engineering (1993).

[6] A. Ghobrial., Aggregate demand model for domestic airlines. Journal of Advanced Transportation (1992).

[7] Saudi Arabian Bechtel Company, Update Forecast and Planning Assumptions, New Riyadh International Airport (1979).

[8] IMF Statistics Department. International Financial Statistics Yearbook.Vol. Xlvii. (1994).

[9] Kingdom of Saudi Arabia-Ministry of Planning. Achievements of the Development Plans (1992).

[10] Forecast Pro for Windows. Business Forecast Systems, Inc. (1994).

[11] W. Makridkis and W. McGee. Forecasting methods and applications. John Wiley and Sons, Inc., New York (1978).

Received Augusts 10, 2000
Accepted December 6, 2001