Modeling Demand for Air Travel at Jeddah International Airport: an Empirical Study

SERAJ Y. ABED and ABDULLAH BAFAIL Department of Industrial Engineering, King AbdulAziz University, Jeddah, Saudi Arabia

ABSTRACT. The primary objective of the work was to suggest a suitable forecasting modeling demand for air travel at Jeddah International Airport. A large amount of data was collected to provide the basis for the conclusions. Data on traffic intensity at Jeddah International Airport for a period ranging from 1975 to 1996 was used for the analysis. These voluminous figures were fitted to most forecasting models. Results indicate that data in all series are trended but non-seasonal. The growth rate was high in the 70's and fluctuating in early 80's and mid 90's.

After the analyses, comparison and cross-checking, the authors arrived at the conclusion that the non-linear cubic model was the most suitable one for aviation planning in Jeddah city. An examination of the statistical performance measures calculated for all models conclusively points to the usefulness of the non-linear cubic model. 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 planning authority could use it for decision-making purposes as and when required.

1. Objectives

Fortunately, Jeddah is endowed with numerous natural opportunities for air travel because of its geographical location, being the "gate for the two holy mosques" and its rapid development in all spheres of life. Jeddah International airport is the busiest airport and the second largest airport at the Kingdom of Saudi Arabia.

The objective of this paper is to conduct a study on Jeddah International airport: • To discover the nature of data-historical trend and rate of growth for yearly arrivals, departures and movements.

• To statistically analyze the historical data for yearly arrivals, departures and movements in Jeddah and produce the statistical parameters of these data.

• To develop suitable forecasting models to forecast yearly arrivals, departures and movements.

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 helps, for example, to study the proper sizing of airports facilities such as gate requirements, apron size, terminal capacity, etc.

2. Literature Review

The airport, which is a major part of air transportation, performs a function of substantial economic importance to the community and to the country it serves. It provides the facility that gives business and industrial communities access to national and industrial air transport networks that link cities and nations closely together. Air transportation has become the fastest, and for long-haul passenger transportation, the dominant form of public transportation. The movement of mail, cargo, and express traffic by air has likewise had a significant impact on the economies of the world^[1].

The airport is the nucleus of this transport system. Safe and efficient air traffic flow depends upon the quality of individual airport facilities and the coordination of individual airports. The growth of air traffic, size and composition of the aircraft fleet, and technological advances in aircraft design are interdependent, and are strongly influenced by economic ability of the airport to keep pace with the expanding requirements for adequate airport facilities^[2]. The explosive development of air commerce and aviation generally has increased the size and coastlines of airport facilities and installations, the complexity of their relations with airlines, governments and the general public, and the need for coordinated planning of interrelated airport systems^[3]. The rapidly developing technology of the aviation industry produces an accelerated rate of technological obsolescence, which, in turn, creates a continuing demand for capital to provide facilities in time to accommodate new demands. The continuing demands both by the public and by the aviation industry for improved air transportation facilities and systems require that basic airport planning be programmed and conducted with thoroughness and with educated forethought^[1]. One of the primary elements of the initial planning is the forecast of aviation demand.

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

Poore (1993)^[4] 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 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 (1992)^[5] 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 (1979)^[6] 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 Riyad 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.

3. Data analysis

Data on passenger arrivals, departures and movements in major international airports in Saudi Arabia during January, 1975 through December, 1996 were obtained from the Presidency of Civil Aviation, the formal Saudi aviation au-

thority^{[7], [8]}. These monthly data were used with other sets of data about some influencing economic variables to develop an econometric model to forecast travel demand in Saudi Arabia.

The set of data about Jeddah International Airport was extracted from the original set, and was analyzed to study the historical trends over that period and to develop suitable forecasting models to forecast future trends in passenger arrivals, departures and movements at this airport.

3.1 Historic Trends and Rates of Growth

Attempts are made to find whether the data were trended, seasonal, irregular or volatile. Plots of the actual data for arrivals, departures and movements are shown below in Fig. 1(a, b, c). All of the distributions are negatively skewed. Further examinations showed that all the series were trended and non-seasonal with trend-cycles and irregular measures of (54.63%, 45.37%), (57.34%, 42.66) and (56.04%, 43.96) for arrivals, departures and movements respectively.

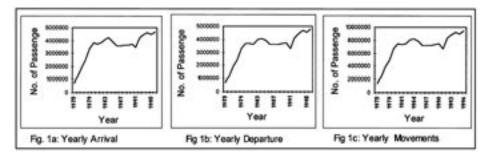


FIG. 1. Plots of original data.

The percentage growth rates of arrivals, departures and movements are shown in Fig. 2(a,b,c). During the late 70's, the rate was very high, and starting from the early 80's to mid 90's, it was found to be fluctuating.

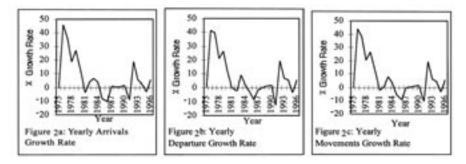


FIG. 2. Percentage growth rate.

3.2 Statistical Parameters

Numerical measures 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 passengers over the 22 year period. It shows that, on an average, the number of annual arrivals and departures was about 3.5 million with a sample standard deviation of about 1 million. The skewness or the measure of asymmetry with the negative sign indicates that the distributions of arrival, departure and movement are skewed to the left (i.e., this is why the mean is smaller than the median). Since the value of the kurtosis (the measure of the heaviness of the tails of a distribution of data) is less than 3, it indicates that the data has a light-tailed distribution. It is evident that the annual arrivals, departures and movement of passengers decreased from 695073, 666916 and 1361989 in 1975 to 4693223, 4803331 and 9496554 respectively in 1996.

Statistical performance measures	Arrivals Departures		Movements
Mean	3523075	3474332	6997407
Standard deviation	1034816	1065550	2098587
Median	3706582	3675534	7413778
Skewness	- 1.559640697	- 1.438160124	- 1.503516023
Kurtosis	2.105399167	1.801334167	1.956829317
Range	3998150	4136415	8134565
Minimum	695073	666916	1361989
Maximum	4693223	4803331	9496554
Total	77507649	76435299	153942948
Number of years	22	22	22

TABLE 1. Descriptive statistics for arrivals, departures, and movements.

3.3 Developing Suitable Techniques for Forecasting

Quantitative forecasting techniques are methods based either on intuition or statistical principle. They are divided into two types. The first type is the time series model. A time series model assumes that some pattern or combination of patterns is recurring over time. The objective of time forecasting methods is to discover the pattern in the historical data series and extrapolate that pattern into future. The second type is explanatory model or causal model. The explanatory model assumes that the value of the forecasted dependent variable is a function of one or more other independent variables. The objective of the explanatory model is to find the relationship that exists between the forecasted variable and other independent variables^[9]. In this study, both modeling methods and Box-Jenkins model (also a time-series) were used. Although the number of years is less than 50, Box-Jenkin model is used only to see how the model behaves in comparison to the others.

3.3.1 Time series models

The most widely used time series forecasting model is exponential smoothing. Exponential smoothing technique is a class of methods that implies exponential decreasing of weights as the observations get older^[9]. In general, the prediction of the new forecast at time 't+1' may be thought of as a weighted average of the previous forecast and the actual value at time 't'. For example, let the number of passenger arrivals at year 't' be y_t , then, to forecast the number of arrivals next year (i.e. \hat{Y}_{t+1}) the following equation is to be used:

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

where \propto is the smoothing parameter and its value ranges between 0 and 1. There are one or more smoothing parameters to be determined which depend on the type of exponential method selected. These types could be single, double, or higher order models. In this study, all types of exponential methods were tried on the three series (arrival, departure, and movement) by using Forecast Pro software^[10]. *Forecast Pro* is designed to fit the best exponential model for a given set of data and determine their smoothing parameters. The best estimated model that fits all three series was found to be the double exponential model. The double exponential model is good for data with a trend. The following equations are used for double exponential smoothing^[9]:

$$\begin{split} \hat{Y}_{t+m} &= C_t + T_t (m) \\ C_t &= \alpha Y_t + (1 - \alpha) (C_{t-1} + T_{t-1}) \\ T_t &= \gamma (C_t - C_{t-1}) + (1 - \gamma) T_{t-1} \end{split}$$

where,

 C_t = the level of the series at time t,

 T_t = the trend of the series at time t,

 \propto and γ are the smoothing constants,

m = the number of periods for which forecast is made.

The plots of 22 years of actual data and their forecasts, as well as the forecast of the next six years (until 2002) for arrivals, departures and movements are shown in Fig 3(a-c). From the figures, it is clear that double exponential models does fit the data well. The estimated models for all three series are:

$$\begin{array}{lll} Arrival & : & \hat{Y}_{t+m} = 4693000 + 125810 \ (m) \\ Departure & : & \hat{X}_{t+m} = 4803300 + 14735 \ (m) \\ Movements & : & \hat{Z}_{t+m} = 946400 + 274790 \ (m) \end{array}$$

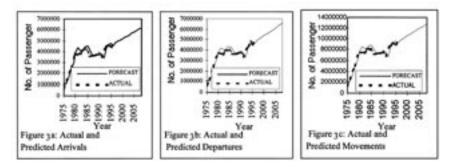


Fig. 3. Plots of 22 years actual data and their forecasts based on exponential smoothing methods.

Performance measuring statistics for all the models and the forecasted values for six years ahead based on the estimated models are are shown in Table 2. Error analysis does not indicate any severe problems with the models. Coefficient of determination (\mathbb{R}^2) for all three models above 88% which means that 88% of the variability in the data are explained by the double exponential models. Durbin-Watson statistic (the statistic tests the hypothesis that there is no autocorrelation present in the residuals) for all three models are around 1.6 which indicate that the errors are random. The mean absolute percentage error (MAPE) for all three models around 10% which is acceptable.

3.3.2 Causal or explanatory models

The most widely used technique for the development of causal models is regression analysis. Regression analysis is a technique for modeling and investigating the relationship between two or more variables. This relationship could be linear or non-linear, and defined in the form of an equation that expresses the relationship between the dependent (forecasted variable) and the independent (explanatory variables). Regression analysis could be simple or multiple. Simple regression is defined by one independent variable, while multiple regression is defined by more than one independent variable^{[11], [12]}. In this study, the dependent variables were the number of arrivals, departures, or movements, while the independent variable was time. Three different models were used. The first one was linear and the rest were non-linear. The form of the three models is as follows:

 $\begin{aligned} \text{Model I} & \text{Y}_t = a + b \ t \\ \text{Model II} & \text{Y}_t = a + b \ t + c \ t^2 \\ \text{Model III} & \text{Y}_t = a + b \ t + c \ t^2 + d \ t^3 \end{aligned}$

Model parameters	Arrival	Departure	Movement
Level	4693100	4803300	9496400
Trend	125810	147350	274790
Statistics			
R-square	0.8815	0.8914	0.8888
Adjusted R-square	0.8756	0.8859	0.8832
Durbin-Watson	1.6290	1.6640	1.6430
Forecast error	365000	359900	717100
МАРЕ	9.96	10.33	10.05
RMES	348000	343100	683700
MAD	257300	271200	517600
Forecast for six years ahead			
1997	4818961	4950634	9771200
1998	4944772	5097985	10045992
1999	5070584	5245337	10320784
2000	5196395	5392689	10595576
2001	5322207	5540040	10870368
2002	5448018	5687392	11145160

TABLE 2. Exponential models for arrivals, departures and movements.

All these models were performed separately on the number of arrivals, departures and movements. By utilizing Interactive Statistical Programs (ISP)^[13], the following regression models were calibrated:

Arrivals	$Y_t = 2098815 + 123848 t$ $Y_t = 1157860 + 359087 t - 10227 t^2$ $Y_t = -439313 + 1110221 t - 90086 t^2 + 2314 t^3$
Departures	$\begin{split} X_t &= 1956249 + 132007 \text{ t} \\ X_t &= 1073966 + 352577 \text{ t} - 9590 \text{ t}^2 \\ X_t &= -528189 + 1106055 \text{ t} - 89697 \text{ t}^2 + 2322 \text{ t}^3 \end{split}$
Movements	$Z_{t} = 4055064 + 255856 t$ $Z_{t} = 2231827 + 711665 t - 19817 t^{2}$ $Z_{t} = -967502 + 2216277 t - 179784 t^{2} + 4636 t^{3}$

*All parameters are statistically significant.

Performance measuring statistics for all models are shown in Table 3. From the Table, it is clear that the third model (non-linear in the cubic form) is the best for arrivals, departures, and movements. Based on these criteria, the nonlinear cubic form model will be used for forecasting the future. Plots of the 22 years of actual data and their forecasts, as well as the six years of forecasts for arrivals, departures, and movements using model III are shown in Figure 4. Forecasts for the next six years are shown in Table 4.

TABLE 3. Performance measures for linear and non-linear regression model for arrivals, departures and movements.

Statistics	Arrivals			Departures			Movements		
Model	Ι	II	III	Ι	II	III	Ι	II	III
R-square	0.60	0.73	0.94	0.64	0.76	0.95	0.63	0.74	0.94
Adjusted R-square	0.58	0.71	0.93	0.63	0.73	0.94	0.61	0.72	0.93
MAPE	24	18	6.5	24	18	7	24	18	6.6
RMSE	6.3E5	5.2E5	2.4E5	6.2E5	5.1E5	2.3E5	12.5E5	10.3E5	4.E5
MAD	4.8E5	4.6E5	2.1E5	4.8E5	4.5E5	1.9E5	9.6E5	9.1E5	4.E5

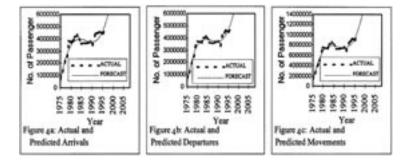


FIG. 4. Plots for 22 years actual and their forecasts based on the cubic model non-linear regression.

Model parameters	Arrival	Departure	Movement
1997	5603552	5712287	11315842
1998	6315243	6450041	12765282
1999	7180083	7342763	14522849
2000	8211961	8404385	16616345
2001	9424767	9648834	19073602
2002	10832390	11090051	21922442

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

3.3.3 Box-Jenkins Model

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

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

which can be written in the short notation form as:

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

where Y_t is the observation sequence and e_t is a sequence 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^{[9], [14]}. The previous equation can be used to model stationary processes. One can model some types of non-stationary processes by differencing the original process, Y_t , to obtain a stationary process, W_t ,

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

This results in an Auto-regressive Integrated Moving Average model ARI-MA (p,d,q).

Techniques for preliminary identification of the model order rely on the analysis of the auto-correlation and partial auto-correlation function. The auto-correlation function describes inherent correlation between observations of a time series which are separated in time by some lag, k. While the 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 partialled out^[9]. Although the three series are reasonably short for utilizing Box-Jenkins method, forecasting was attempted. Based on the SPSS^[15] software analysis, the best ARIMA model for arrivals, departures, and movements was:

Arrivals :
$$Y_t = 0.5255 Y_{t-1} + Y_{t-1} - 0.5255 Y_{t-2} + e_t$$

Departures : $X_t = 0.4889 X_{t-1} + X_{t-1} - 0.4889 X_{t-2} + e_t$
Movements : $Z_t = 0.5195 Z_{t-1} + Z_{t-1} - 0.5195 Z_{t-2} + e_t$

The plots of the 22 years of actual data and their forecast, as well as the forecast of the next six years for arrivals, departures, and movements are shown in Figure 5(a-c). Performance measuring statistics for all models and the forecasted values for six years ahead based on the estimated models are shown in Table 5. The correlogram shows that auto-correlation function dies out quickly. Coefficients of determination for all three models are above 82%. Durbin-Watson statistics are around 2. MAPE for all three models are around 8%.

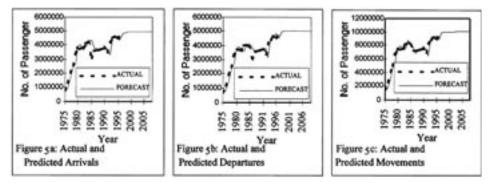


FIG. 5. Plots for 22 years actual data and their forecasts based on Box-Jenkins model.

Model parameters	Arrival	Departure	Movement			
Model	ARIMA(1,1,0)	ARIMA(1,1,0)	ARIMA(1,1,0)			
Coeff.	0.5255	0.4889	0.5195			
Std. error	0.1757	0.1848	0.1786			
t-statistics	2.9905	2.6456	2.9086			
Significance	0.9928	0.9845	0.9913			
Statistics						
R-square	0.8270	0.8382	0.8373			
Adjusted R-square	0.8270	0.8382	0.8373			
Durbin-Watson	2.1250	2.2040	2.1690			
Sjung-Box(10)	4.538 P = 0.08018	5.258 p = 0.1267	3.675 p = 0.03916			
Forecast error	349400	355100	367500			
BIC	366600	372600	728200			

TABLE 5. Box-Jenkins models for arrivals, departures and movement.

Model parameters	Arrival	Departure	Movement	
MAPE	0.08072	0.08383	0.08084	
RMSE	341000	346600	677300	
MAD	251100	252200	677300	
Forecast for six years ahe	ad	\$		
1997	4817533.50	4929894.00	9753934.00	
1998	4882854.00	4991776.00	9887647.00	
1999	4917177.50	5022033.00	9957113.00	
2000	4935213.50	5036827.00	9993202.00	
2001	4944590.50	5044060.50	10011951.00	
2002	4949670.50	5447597.00	10021691.00	

TABLE 5. Contd.

3.4 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 these measures. Based on the performance measuring statistics for all different models, it seems that out of the three methods attempted for forecasting arrivals, departures and movements in Jed-dah international airport, the Non-linear Regression Model (Model III) yields better results. The Box-Jenkins model is the next choice. The Softwares used in analyzing these data were SPSS^[15], Microsoft Excel (Version 5)^[16], Forecast Pro for windows^[10] and ISP (Interactive Statistical Programs)^[13].

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

TABLE 6. Performance measures based on all three methods for arrivals, departures and movements.

4. Conclusion

The primary objective of the work was to suggest a suitable forecasting modeling demand for air travel at Jeddah International Airport. Since there are a number of forecasting models available for such purposes, choice of the appropriate model is crucial and decisive for aviation planning. The work was carried out with this aspect in mind and a large amount of data was collected to provide the basis for the conclusions. Data on traffic intensity at Jeddah International Airport for a period ranging from 1975 to 1996 was used for the analysis. These voluminous figures were fitted to most forecasting models. After the analyses, comparison and cross-checking, the authors arrived at the conclusion that the non-linear cubic model was the most suitable one for aviation planning in the Jeddah city. The validity of this conclusion and reasons thereof are discussed elsewhere in the paper. The cross-checking, done using Box-Jenkins model which is one of the most advanced techniques of forecasting, proves this decisively. An examination of the statistical performance measures calculated for all models conclusively points to the usefulness of the non-linear cubic model. 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 planning authority could use it for decision-making purposes as and when required.

References

- [1] Siddiqui, Aziz (1994) Operating Levels and Airline Economics. Saudia World, July.
- [2] Boeing Commercial Company (1987) Air Freight, Air Travel, and Strategic Planning.
- [3] Boeing Commercial Airplane Company (1993) Proceedings of Airline Evaluation Seminar.
- [4] **Poore, James Walter** (1993) Forecasting the Demand for Air Transportation Services. *Journal of Transportation Engineering,* September.
- [5] **Ghobrial, Atef** (1992) Aggregate Demand Model For Domestic Airlines. *Journal of Advanced Transportation.*
- [6] Saudi Arabian Bechtel Company (1979) Update Forecast and Planning Assumptions, New Riyadh International Airport, July.
- [7] IMF Statistics Department (1994) International Financial Statistics Yearbook. vol. xlvii.
- [8] Kingdom of Saudi Arabia, Ministry of Planning (1992) Achievements of the Development Plans.
- [9] Makridkis, Wheel Wright, McGee (1978) Forecasting Methods and Applications. John Wiley & Sons, Inc., New York.
- [10] Forecast Pro for Windows (1994) Business Forecast Systems, Inc.
- [11] Walpole, Ronald E. and Raymond, H. Myers (1985) Probability and Statistics for Engineers and Scientists. 3rd ed., Macmillan Publishing Company, New York.
- [12] Montgomery, Douglas C. and Elizabeth A. Pech (1982) Introduction to Linear Regression Analysis. John Wiley & Sons Inc., New York.
- [13] Interactive Statistical Programs (ISP) (1992) Lincoln Systems Corporation.
- [14] Koutsoyiannis, A. (1977) Theory of Econometrics, 2nd ed., The Macmillan Press Ltd., Hong Kong.
- [15] SPSS User Manual, Release 6 (1993) SPSS Inc., U.S.A., June.
- [16] Microsoft Corporation (1994) Building Applications Using Microsoft Excel.

المستخلص. الهدف الأساسي من العمل هو بناء نموذج مقترح للتنبؤ الملائم لطلب السفر الجوي في مطار جدة الدولي حيث جمعت كمية كبيرة من البيانات لدعم قواعد الاستنتاجات ، كما استخدمت بيانات كثافة السير أو الحركة في مطار جدة الدولي خلال الفترة ١٩٧٥ – ١٩٩٦م لبناء عدة نماذج مقترحة . وقد أشارت الرسومات البيانية إلى أن البيانات في كل السلاسل الزمنية ذات اتجاه ولكنها ليست موسمية ، كما أن معدل النمو كان عاليًا في السبعينيات ومتذبذبًا في بداية الثمانينيات ومنتصف التسعينيات .

بعد التحليل والمقارنة والتعديلات توصل الباحثان إلى أن النموذج المكعب غير الخطي كان الأكثر ملائمة بناءً على اختيار مقاييس الأداء الإحصائية لكل النماذج . وإن أردنا التوسع في المشروع فإننا نستطيع أن نطور نظام القرار المساند اعتماداً على هذه النماذج بحيث تستطيع سلطات الملاحة الجوية استخدامها للمساعدة في اتخاذ القرار عند الحاجة إليها .