Transportation Research Part A 65 (2014) 23–32

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

# A demand trend change early warning forecast model for the city of São Paulo multi-airport system

Rodrigo Arnaldo Scarpel *

*Instituto Tecnológico de Aeronáutica (ITA), Brazil*

ARTICLE INFO

*Article history:*
Received 10 December 2012
Received in revised form 2 April 2014
Accepted 8 April 2014
Available online 4 May 2014

*Keywords:*
Scenarios
Short-term fluctuations
Long-term forecasting
Classification and regression tree
Air passenger numbers

ABSTRACT

The need of accurate forecasts of air passenger numbers to assist managerial decision making for both short and long terms is well recognized and a central problem on both short and long term forecasting is how to handle future trend. The aim of this paper is to develop a demand trend change early warning forecast model (EWFM) for the city of São Paulo multi-airport system (SPMARs). For SPMARs the EWFM is based on the combination of leading indicators and alarms against possible occurrence of changes on trend component of the monthly number of domestic air passengers. A topdown induction procedure is employed to identify leading indicators to provide an interpretable prediction procedure to support the development of scenarios for future demand trend. Results show that changes on such demand trend are mostly associated to changes on the economic activity and six different scenarios were built combining the identified leading indicators. The EWFM was employed to assist managerial decision making for both short and long terms in order to evaluate different alternatives to prevent congestion delay occurrences and to support infrastructure planning.



## 1. Introduction

The need of accurate forecasts of air passenger numbers to assist managerial decision making for both short and long terms is well recognized. According to Scarpel (2013), long-term forecasting of air passenger numbers provides a key input into infrastructure planning of airports and air navigation services, and for aircraft ordering and design. Moreover, the ability to anticipate short-term fluctuations on demand trend is helpful for airlines in order to develop strategies for dealing with demand instability and for aviation policy makers in order to evaluate different alternatives to prevent congestion delay occurrences. According to Madas and Zografos (2008), the increasing imbalances between capacity and traffic has resulted in congestion and delay figures that have drawn the attention of aviation policy makers investigating alternative means of coping with the mismatch between aviation capacity and demand. Congestion during peak periods also puts a tremendous strain on airport and airline personnel and also creates additional work for air traffic controllers (Wensveen, 2011).

A central problem on both short and long term forecasting is how to handle future trend. According to Grubb and Mason (2001), the trend is the most important component to forecast on long lead-time prediction in the presence of strong growth. Moreover, trend curves will probably dominate long-term forecasts (Granger and Jeon, 2007). However, short-term shocks on the trend component frequently cause deviations of demand from its underlying growth path.

* Address: Praça Marechal Eduardo Gomes, 50 – ITA/IEM, sala 2311, São José dos Campos, SP CEP 12.228-900, Brazil. Tel.: +55 1239476973.
*E-mail address:* rodrigo@ita.br

http://dx.doi.org/10.1016/j.tra.2014.04.007
0965-8564/

Different approaches can be employed to deal with such trend component. The first and most usual approach is extrapolation using time series analysis (Armstrong, 1985). In order to give more control over trend extrapolation Gardner and McKenzie (1985, 1989) added a damping parameter to the extrapolation model. Grubb and Mason (2001) proposed damping future trend towards the historical average trend instead of with damping to zero. Such modification allowed them to vary the trend used for the predictions and to build different scenarios for future trend. According to Schnaars (1987), the term scenarios is used by many researchers to describe any set of multiple forecasts and the idea of providing multiple forecasts relies on the recognition that as a forecast is only as accurate as its underlying assumptions, it makes more sense to consider a number of plausible assumptions, rather than a single one which may later turn out to be incorrect. Ozyildirim et al. (2010) employed leading indicators for anticipating changes on future trend. According to Jones and Chu Te (1995), the usefulness of leading indicators is that it enables researchers to determine and predict turning points in the cyclical movements of an activity of interest.

The aim of this paper is to develop a demand trend change early warning forecast model (EWFM) for the city of São Paulo multi-airport system (SPMARs) to assist managerial decision making for both short and long terms. SPMARs encompasses the São Paulo International Airport (SBGR), Congonhas Airport (SBSP) and Viracopos Airport (SBVR). An EWFM is based on the combination of leading indicators and alarms against possible occurrence of changes on a variable of interest. For SPMARs the variable of interest is the trend component of the monthly number of domestic air passengers. Two of the most critical tasks for a EAFM success are the leading indicators identification and the scenarios building. Therefore, in this work a top-down induction procedure is employed to identify leading indicators that could be used for the purpose of anticipating changes on the trend component of time series data and to provide an interpretable prediction procedure to support the development of scenarios for future demand trend. In order to assist managerial decision making the developed EWFM is employed to anticipate short-term fluctuations on demand trend and to perform long-term forecasting of domestic air passenger numbers.

Air transport in Brazil has been recently liberalized and one of the consequences of this process was the concentration of flights in a few hubs (Costa et al., 2010). According to Wensveen (2011), although hubbing seems to benefit airlines and offers some advantages to travelers, the extent of excessive concentration at a hub can result in some negative economic impacts, namely, congestion delay which increases passenger's total travel time and airlines' operating costs. SPMARs, that up to date, is the largest multi-airport system in Brazil in terms of total number of air passengers is the place that most suffers with such hub concentration and congestion delays caused by excess demand. Thus, a demand trend change EWFM is helpful for SPMARs not only for long-term forecasting to support infrastructure planning but also to anticipate short-term fluctuations and help minimizing congestion delay occurrences.

The rest of the paper is organized as follows. Section 2 outlines the employed top-down induction procedure. Section 3 has two sub-sections, the first focuses on the data selection, pre-processing and transformation steps and the second focuses on model building and on reporting and discussing the obtained results. In Section 4 the EWFM is employed to assist managerial decision making for both short and long terms. Conclusions are presented in the final section.

## 2. Decision trees and CART

Different statistical methods can be employed to extract information from a data set and transform it into an understandable structure. Such statistical methods most common functions include attribute selection, classification, regression and clustering. In this work two of those functions are performed: attribute selection to identify leading indicators and regression to provide an interpretable prediction procedure to support the development scenarios for future trend. According to Olafsson et al. (2008), attribute selection involves a process for determining which attributes are relevant in that they predict or explain the data, and conversely which attributes are redundant or provide little information. In regression, the goal is to map the relationship between a response variable and a set of explanatory variables.

The top-down induction of decision trees is employed in this work due its transparency and hence relative advantage in terms of interpretability. Decision tree methods are attractive when the interpretability is an important issue since they are designed to detect the important predictor variables and to generate a tree structure to represent the identified recursive partition. Fig. 1 presents the equivalent partition and tree obtained for a hypothetical example.

According to Banks (2010), such approach has minimal statistical or model assumptions and its most general solution is as follows: Suppose that we have a sample of $n$ observations, a response variable $Y_1,\ldots,Y_n$ and each observation has a $r$-dimensional vector of covariates $x$. Most decision tree induction algorithms construct a tree in a top-down manner by selecting attributes one at a time and splitting the data according to the values of those attributes. The most important attribute is selected as the top split node, and so forth. In summary, in such method an algorithm is employed to induce a binary tree on the given data, which in turn results in a set of 'if–then' rules.

Partitioning acts as a smart bin-smoother that performs automatic variable selection. Formally, it fits the model

$$Y = \sum_{j=1}^{r} \beta_j I(x \in R_j) + \varepsilon \quad (1)$$

where the regions $R_j$ and the coefficients $\beta_j$ are estimated from data and $\varepsilon$ is commonly assumed to be a random error. Usually the $R_j$ are disjoint and the $\beta_j$ is the average of the $Y$ values in $R_j$. A recursive partitioning algorithm has three parts: (1) a