

Modeling Weather Impact on Ground Delay Programs

2011-01-2680

Published
10/18/2011Yao Wang and Deepak Kulkarni
NASA Ames Research Center

doi:10.4271/2011-01-2680

ABSTRACT

Scheduled arriving aircraft demand may exceed airport arrival capacity when there is abnormal weather at an airport. In such situations, Federal Aviation Administration (FAA) institutes ground-delay programs (GDP) to delay flights before they depart from their originating airports. Efficient GDP planning depends on the accuracy of prediction of airport capacity and demand in the presence of uncertainties in weather forecast. This paper presents a study of the impact of dynamic airport surface weather on GDPs. Using the National Traffic Management Log, effect of weather conditions on the characteristics of GDP events at selected busy airports is investigated. Two machine learning methods are used to generate models that map the airport operational conditions and weather information to issued GDP parameters and results of validation tests are described.

I. INTRODUCTION

Air traffic congestion at the major commercial airports has been a serious problem in the National Airspace System (NAS), especially during bad weather [1]. FAA's Traffic Flow Management (TFM) manages air traffic flow to balance air traffic arrival demand against airport capacity, when the latter is reduced by bad weather or other circumstances. This results in airborne delays by holding some aircraft for landing. At major airports in the United States, when the expected demand for arrival air traffic flow exceeds the Airport Arrival Rate (AAR) for a significant period of time, GDP will be used as one of Traffic Management Initiatives (TMI) to smooth out the arrival flow and bring arrival demand in line with capacity. The most common reason for an overage of demand versus capacity is the reduction in airport acceptance rate due to adverse airport weather such as strong wind, low ceilings and visibility.

Air Traffic Control System Command Center (ATCSCC) implements GDPs after consultation with air traffic centers and airline operation centers to manage the arrival flow by holding some aircrafts at their departure airports for specified periods of time. As a result, airborne delays are reduced and less expensive and less risky ground delays are increased. If a GDP applies to a particular airport, the GDP start time and GDP stop time will be determined by the scheduled demand and forecast weather profile at the time of the GDP planning. Given the nature of weather forecast uncertainty and the inaccurate weather translation models, the initial and final planned GDP durations are often different from the actual duration. This can result in unnecessary delays.

This paper studies the use of Ensemble Bagging Decision Tree (BDT) and Neural Networks (NN) methods to predict the GDP duration time during bad weather. Our approach is to develop predictive BDT and NN models using historical GDP and weather data, and then apply these models to forecast the GDP duration when GDP is being planned. The prediction outlooks are then discussed.

The remainder of the paper is organized as follows. In the next section, statistical analysis of GDP events at several major airports is provided. Section III details BDT and NN approaches and discusses factors affecting training, performance as well as model validation. Section IV describes the data used in this study. Then section V presents computational results on the estimation of GDP duration using BDT and NN approach. Finally, section VI provides concluding remarks and future directions.

II. STATISTICS OF GROUND DELAY PROGRAMS

Traffic managers at the regional and national levels institute and modify TMIs in order to balance traffic demand with

system capacity. The FAA developed the National Traffic Management Log (NTML) to provide a single system for automated coordination, logging, and communication of TMIs throughout the National Airspace System. Some aggregate statistics obtained by the examination of the NTML data have been published [2-3] recently. We collected more than 3400 GDP events from NTML data at all US airports over the years 2007 through 2009, and generated GDP event statistics as follows:

Figure 1a displays percentage share of airport GDP counts for the top 8 airports. It shows that most GDPs are implemented at the airports in the northeast region of the United States, including the three New York-area airports (EWR, LGA, and JFK), Philadelphia (PHL), and Boston Logan International Airport (BOS). Figure 1b shows that the major cause of Ground Delay Programs is inclement weather. The diverse weather causes are shown in Figure 2. Details of weather causes for the top 8 airports are provided in Tables 1 and 2. Altogether, these data illustrate that the dominant weather causes for GDPs are different at different airports. For example, while close to 90% GDPs at SFO are caused by low ceilings due to marine stratus, wind accounts for about 50% of GDPs at the three New York-area airports, and thunder storms are the major sources of GDPs at ATL.

The GDP start time, end time, scale, and planned AAR (PAAR) are defined by TFM when a GDP is issued initially. The initial planned duration is the difference between the planned end time and start time at GDP issued time. During a GDP, these program parameters might need to be revised because of changing forecast and operation conditions. GDP revisions may lead to further GDP end time substitutions and the final planned duration is the time duration between the final revised GDP end time and start time. GDP actually ends either when the final planned GDP end time is reached or when it is cancelled. Table 3 shows average initial planned duration, average final planned duration and average actual duration for the top eight airports. The time durations for diverse weather causes are shown in Table 4. For most of these airports, the differences between averages of Final and Initial Planned Durations are much less than an hour with the exceptions of PHL and SFO, where the differences are about 1 hour. The average of the actual time duration is less than the averages of both initial planned and final planned durations. Such conservative planning of GDP events results in unnecessary delays for many GDP events. For the three New York-area airports where wind is a prominent cause of GDPs, the average time differences between the average final planned durations and the average actual durations are within one to two hours; for other airports, the time differences are more than two hours.

As shown in Tables 3 and 4, these GDP programs usually last for several hours and the average durations depend on weather conditions as well as the airports. Table 5 lists the

average total EDCT (Expect Departure Clearance Time) delay minutes (in 1000 minutes) and the total numbers of aircraft assigned with EDCT delays for the top 8 airports due to different weather causes. As shown in Table 5, GDPs generated more delays for ORD and ATL airports than for other airports. At any one airport, the delays can be quite different due to different weather causes.

Reducing the number of GDP revisions and the time difference between GDP initial planned duration and actual GDP duration will reduce air traffic management work load as well as unnecessary aircraft delays. With more accurate air traffic demand prediction, less uncertainty in weather forecast, and better airport capacity models, the initial issued GDP can be improved [4-5]. In this paper, we present machine learning methods to reduce the number of GDP revisions and to improve initial planned GDP duration in order to support TFM.

III. APPROACH AND MODELING METHODOLOGY

We propose to use two machine learning techniques to predict characteristics of a GDP. Ensemble learning with Bagging Decision Tree (BDT) model was used here to classify the difference between GDP initial planned duration and final planned duration. Furthermore, Neural Networks (NN) regression model was applied to predict GDP duration. Supervised machine learning was used to train models, i.e., BDT and NN, and the models were validated using data cross validation methods.

ENSEMBLE BAGGING DECISION TREE

Ensemble methods use multiple machine learning models to obtain better predictive performance than what any of its individual constituent members can produce. Bagging is an ensemble method that uses random resampling of a dataset to construct models [6]. In classification scenarios, the random resampling procedure in bagging induces some classification margin over the dataset. Additionally, when bagging is performed in different feature subspaces, resulting classification margins are likely to be diverse, which is essential for an ensemble to be accurate. This method takes into account the diversity of classification margins in feature subspaces to improve the performance of bagging. First, it studies the average error rate of bagging, converts the task into an optimization problem for determining some weights for feature subspaces. Then, it assigns the weights to the subspaces via a randomized technique in classifier construction. Experimental results demonstrate that the ensemble method is robust to classification noise and often generates improved predictions than any single classifier [7-8].

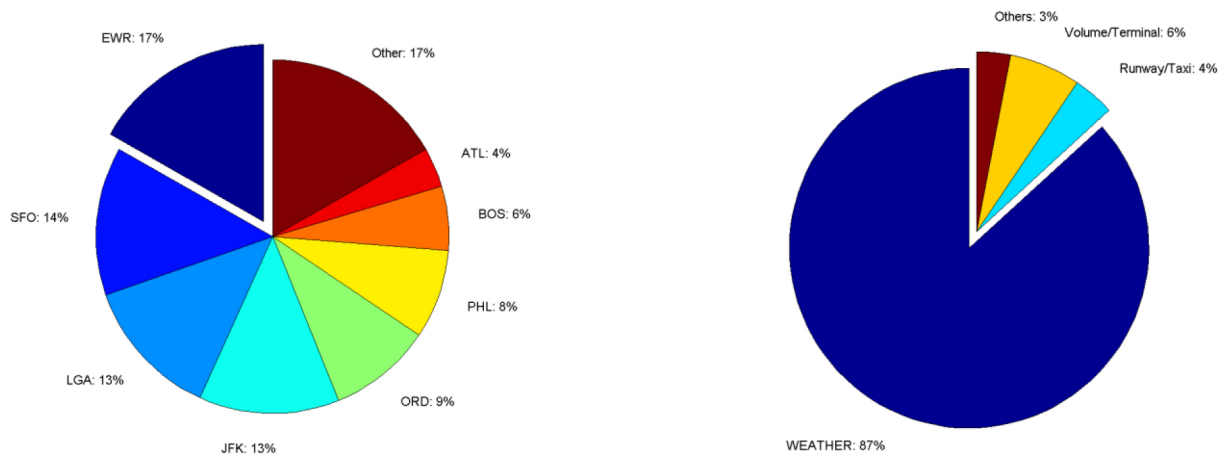


Figure 1. (a) GDP Airport Counts and (b) GDP Causes

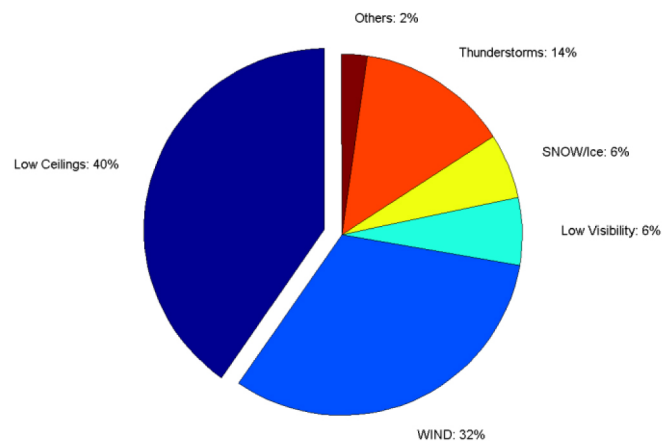


Figure 2. Distribution of Weather GDPs among Different Weather Subcategories

Table 1. Category Percentage Ratio for the Top 8 airports

Airport	Weather	Equipment	Center Volume	Terminal Volume	Runway Taxi	Others
EWR	92%			4%	3%	1%
SFO	96%				3%	1%
LGA	88%	1%		9%	2%	
JFK	78%		1%	17%	2%	2%
ORD	98%	2%				
PHL	91%			1%	8%	
BOS	95%	2%			2%	1%
ATL	96%	4%				