## A Machine Learning Approach to Improve Air Traffic Management Initiatives

Farzan Masrour Shalmani<sup>\*1</sup>, Milad Memarzadeh<sup>1</sup>, Aida Sharif Rohani<sup>1</sup>, and Krishna Kalyanam<sup>2</sup> <sup>1</sup>Universities Space Research Association (USRA), NASA Ames Research Center, Moffett Field, CA 94035, USA <sup>2</sup>NASA Ames Research Center, Moffett Field, CA 94035, USA

## I. Extended Abstract

Collaborating closely with commercial air carriers and related organizations, the Federal Aviation Administration (FAA) regulates air traffic and ensures the safety and efficiency of air operations. Air traffic controllers make strategic decisions, such as delaying, rerouting, or canceling flights, partly based on guidance provided by the FAA's Air Traffic Control System Command Center (ATCSCC). The guidance includes, among other things, control measures known as Traffic Management Initiatives (TMIs) designed to enhance safety and improve operational efficiency. TMIs play a crucial role in managing the demand and capacity within the U.S. National Airspace System (NAS). Two major TMIs that are routinely used (primarily to mitigate the adverse effects of bad weather) are Ground Delay Programs (GDPs) and Ground Stops (GSs). In a GDP, flights destined for airports facing thunderstorm activity experience delays at their origin airports. This proactive approach minimizes the risk of routing aircraft through hazardous weather conditions and also replaces (fuel burning) airborne delays with ground delays. In a GS, a temporary restriction is imposed on the departure or arrival of aircraft at a specific airport or within a designated airspace. Although other TMIs (e.g., miles-in-trail) are also implemented as part of (air) traffic flow management in the NAS, the focus of this work is on GDPs and GSs.

Since TMIs, by design, lead to flight delays or cancellations, it is crucial to put in place the right set of parameters (e.g., scope and duration of the GDP). For example, when the end time of a GDP extends beyond what is necessary, it imposes unnecessary delays on departing flights. This situation could occur as a result of inaccurate prediction of the (required) duration of the GDP based on the weather forecast. On the other hand, if a GDP ends prematurely before the underlying capacity constraints are resolved at the destination airport, it may result in airborne holding. The delicate balance lies in matching the termination of the GDP precisely with the resolution of capacity constraints, avoiding both the imposition of unnecessary ground delays and the need for airborne holding due to premature program termination. Failing to specify the right parameters for TMIs also leads to flight delays, creating a significant obstacle in managing the increasing traffic volumes causing increased work load for the controllers.

To address this issue, we propose the integration of Machine Learning (ML) models in the traffic flow management (TFM) pipeline. In current operations, decisions are made by human experts based on extensive training, historical patterns, available traffic and weather data. Since we have an abundance of data from past events that tell us the likely impact of various TMIs, by ingesting historical data, properly trained ML models can offer valuable insights and aid human decision-making. With the FAA increasingly exploring advanced analytics, ML emerges as a focal point for enhancing TFM within the National Airspace System (NAS). As a first step, this study aims to provide traffic controllers with decision-making support for the issuance and adjustment of TMIs.

Data analytics and machine learning have been previously employed to address some of the challenges associated with TMIs. Numerous studies have concentrated on various facets of TMI issuance, exploring factors influencing TMI parameters, including arrival rate, airport capacity, and delay prediction. For example, using weather forecasts, several statistical methods were used to produce probabilistic capacity profiles which in conjunction with deterministic models provided insights into the GDP planning process [1–4]. The downside of using deterministic models is that they rely on fixed inputs and predetermined rules, which lack the ability to account for the inherent uncertainty and variability present in real-world scenarios.

In a separate series of studies, researchers aimed to predict the occurrences of GDPs and GSs. The majority of these studies utilized various supervised learning methods, including Decision Trees, Naive Bayes, Support Vector Machines, and Random Forests to analyze the influence of weather conditions and arrival demand on TMI incidents [5–8]. However, these studies primarily focused on predicting the incidence of TMIs without explicitly addressing the scope of TMIs, including their duration and their geographical coverage. Furthermore, the emphasis of these studies was largely on GDPs, given their higher frequency and longer duration when compared to GSs.

<sup>\*</sup>Corresponding author: farzan.masrourshalmani@nasa.gov

A limited number of studies focused on predicting the parameters of TMIs, specifically addressing their duration and extent. In one such study focusing on optimizing the TMI parameters at San Francisco International Airport (SFO), the authors utilized a probabilistic forecast of fog [9]. They simulated various capacity scenarios based on the (fog) burn-off forecasts, selecting GDP parameters that minimized airborne and overall ground delays. However, this approach exclusively emphasizes stratus (fog) burn-off as the primary determinant of GDP and GS, neglecting other influential factors like severe weather events, runway closures, lower capacity than traffic demand, and other important variables.

Given the complexity of predicting the TMI and determining its scope, we seek a more holistic approach. We aim to consider all significant factors that could impact TMIs and their parameters. What sets this research apart is the fusion of all data sources relevant to the issuance and adjustment of TMIs and it represents the first comprehensive attempt to optimize TMIs in this manner. Since this comprehensive solution involves various aspects, we break down the problem into smaller components and input all parameters into a unified model called the "TMI Adjuster". Figure 1 shows the overall framework and the list of datasets used in each model.



Fig. 1 TMI adjuster framework and ML modeling pipeline.

The objective of the TMI Adjuster module is to deliver reliable, consistent and expedited recommendations for the progression, adjustment, and termination of TMIs. The ML solution entails developing a pipeline capable of predicting the necessity of a TMI (e.g., GS or GDP) along with its various parameters. For example, in the case of a GS, this includes the scope of the GS either in terms of distance from the destination airport or based on pre-defined airspace sectors. Here, scope refers to those regions and departing airports that are subject to the GS.

In this paper, we concentrate on the issuance of GSs in the three major airports in the New York area — LaGuardia (LGA), John F. Kennedy International (JFK), and Newark Liberty International (EWR). We fuse traffic, weather and other relevant aviation data from years 2017 to 2019 to train and validate the ML models. In particular, we use the following datasets:

- Terminal Aerodrome Forecast (TAF): meteorological forecasts specific to each airport, issued four times a day, covering predefined time periods.
- TMI data: includes all GSs and GDPs along with their respective parameters.
- Aviation System Performance Metrics (ASPM): includes traffic related data such as aircraft delays, arrival, and departure rates.
- Notices to Airmen (NOTAMs): utilized to extract runway closure data and manage interdependencies between terminals in close proximity.
- Flight cancellation data
- Airspace Flow Programs (AFP): includes information on flight airborne holdings caused by TMIs.

The data preprocessing entails transforming ASPM, TMI, AFP, NOTAMs, and weather data into an hourly format and consolidating all datasets by merging them based on date and time as the primary key.

The TMI Adjuster framework comprises two parallel models: one dedicated to GS and a second model focused on GDP. As previously mentioned, our specific focus is on the GS model as a multi-classification problem. In this framework, each data point of the GS model input summarizes ten hours of data. Specifically, the data loader for the GS model generates the input and output of the model as follows: at a given time step, the input includes the actual traffic, weather, and TMI data from the two-hour window before the time step, alongside the weather forecast and scheduled traffic for the next 8 hours starting from the time step. Based on this information, the output of the GS model for each time interval consists of three dimensions. The first dimension represents a binary decision on whether there should be a GS in place for the next hour or not. The second dimension is related to the scope of the GS in the United States, and the third dimension is related to the scope of the GS in Canada (i.e., to determine if the GS impacts airports in Canada).

One of the challenges with TMI modeling is the sparsity of TMI events, particularly regarding its scope. To address this challenge in the scope of the GS model output, we implement grouping. The GS scope for the US region is defined based on a list of centers that should be included when the GS is in place. With 20 centers in the US, we utilized historical data to group them into 4 categories. In particular, we summarized our historical data in a graph format where nodes represent centers, and link weights are defined based on the co-occurrence of centers in the scope parameter of TMIs. By identified strongly connected components in this graph, we were able to partition the centers into four groups.



Fig. 2 The US centers, alongside the color-coded grouping.

Figure 2 illustrates the US centers, with the color coding representing the four groups we defined for these centers. Consequently, the model output for the US dimension will indicate the groups impacted by the TMI in the next hour. For Canada, we further simplified the process by setting the model to simply decide *whether Canada should be included or not in the TMI*. Our study concentrates only on determining TMI parameters when their implementation is confirmed. In particular, we limited the training, validation, and testing data to the time steps when there was actually a TMI in place, i.e., a GS or a GDP, or when a TMI had just terminated. Figure 3 illustrates a snapshot of this process. For training, we considered data from 2017 and 2018, the first six months for validation and hyperparameter tuning, and the second six months of 2019 for testing.

Airport	# Train	#Validation	#Test	GS Positive Class%	GS Negative Class%	No TMI %
JFK	2435	335	390	40%	60%	13%
LGA	3635	1022	1100	42%	58%	12%
EWR	5324	1349	1317	28%	72%	11%

Table 1The summary of statistics for GS Data.



Fig. 3 Our study focuses solely on determining TMI parameters when their implementation is confirmed. The yellow boxes indicate the time windows we consider in the training data, while data points outside of these time windows are ignored.

We consider two model structures for the GS Model. Firstly, a hierarchical classification model [10], where the human decision-making for a GS is of hierarchical nature. The decision-maker first decides whether there is a need for a GS, and if the answer is yes, determines the scope. A hierarchical classification model organizes the problem into a class hierarchy, typically a tree or a Directed Acyclic Graph (DAG) structure, and considers the dependency of the decision in the previous step to the next component [10]. Here, we employ the local classifier per level approach, which involves training one multi-class classifier for each level of the class hierarchy. The second structure is the independent structure. In this setting, as the name suggests, we do not consider the dependency of the decisions in the different dimensions of the output of the model. Instead, for each dimension, we train a multi-class classifier independently.

Table 1 summarizes GS model statistics for training, validation and testing. The table documents the effect of limiting data to the time steps when there was actually a TMI in place or when a TMI had just terminated. This resulted in a more balanced distribution of the GS class(GS positive class)versus "No GS"(GS negative class), which might help the training process. While JFK and LGA follow very similar distributions, with 40% and 42% GS positive class respectively, EWR has proportionally fewer GS incidents at 28%.

Our subsequent phase involves evaluating the performance of both hierarchical structure and independent structure using different state-of-the-art multi-class classifier models such as Random Forest, Decision Trees, K-nearest Neighbors, and Logistic Regression and forecast the duration and scope of the GSs.

## References

- Buxi, G., and Hansen, M., "Generating day-of-operation probabilistic capacity scenarios from weather forecasts," *Transportation Research Part C: Emerging Technologies*, Vol. 33, 2013, pp. 153–166.
- [2] Provan, C. A., Cook, L., and Cunningham, J., "A probabilistic airport capacity model for improved ground delay program planning," 2011 IEEE/AIAA 30th Digital Avionics Systems Conference, IEEE, 2011, pp. 2B6–1.
- [3] Kicinger, R., Krozel, J., Steiner, M., and Pinto, J., "Airport capacity prediction integrating ensemble weather forecasts," *Infotech@ Aerospace 2012*, 2012, p. 2493.

- [4] Kicinger, R., Cross, C., Myers, T., Krozel, J., Mauro, C., and Kierstead, D., "Probabilistic airport capacity prediction incorporating the impact of terminal weather," *AIAA Guidance, Navigation, and Control Conference*, 2011, p. 6691.
- [5] Liu, Y., Hansen, M., Zhang, D., Liu, Y., and Pozdnukhov, A., "Modeling Ground Delay Program Incidence using Convective and Local Weather Information," *Proceedings of the Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017), Seattle, WA, USA*, 2020, pp. 1–8.
- [6] Mukherjee, A., Grabbe, S. R., and Sridhar, B., "Predicting ground delay program at an airport based on meteorological conditions," *14th AIAA aviation technology, integration, and operations conference*, 2014, p. 2713.
- [7] Mangortey, E., Pinon-Fischer, O. J., Puranik, T. G., and Mavris, D. N., "Predicting The Occurrence of Weather And Volume Related Ground Delay Programs," AIAA Aviation 2019 Forum, 2019, p. 3188.
- [8] Mangortey, E., Bleu-Laine, M.-H., Puranik, T. G., Pinon Fischer, O., and Mavris, D. N., "Machine learning approach to the analysis of traffic management initiatives," *Journal of Air Transportation*, Vol. 29, No. 2, 2021, pp. 56–68.
- [9] Cook, L. S., and Wood, B., "A model for determining ground delay program parameters using a probabilistic forecast of stratus clearing," *Air traffic control quarterly*, Vol. 18, No. 1, 2010, pp. 85–108.
- [10] Silla, C. N., and Freitas, A. A., "A survey of hierarchical classification across different application domains," *Data mining and knowledge discovery*, Vol. 22, 2011, pp. 31–72.