

# A Model for Determining Ground Delay Program Parameters Using a Probabilistic Forecast of Stratus Clearing

Lara S. Cook, Bryan Wood

Mosaic ATM, Inc.

Leesburg, VA, USA

cook@mosaicatm.com, wood@mosaicatm.com

**Abstract**—An approach is presented for using the probabilistic forecast of stratus clearing time at San Francisco (SFO) to achieve more efficient Ground Delay Programs (GDPs) by better determining GDP end time and scope. Given a probabilistic forecast, we use a Monte-Carlo simulation approach to generate many stratus clearing times for each discrete GDP end time and scope under consideration. Various key measures are evaluated such as unnecessary ground delay if the GDP ends later than stratus clearing and the risk of airborne holding at the end of the GDP if the GDP ends earlier than stratus clearing. An objective function that includes each of the key metrics captures the cost of each scenario under consideration, and the optimal GDP parameters can then be selected. Results show reductions of 29% for unnecessary issued ground delay and reductions of 39% for unnecessarily delayed flights over the SFO GDPs during the severe weather seasons in 2006 and 2007.

**Keywords**—SFO Stratus Forecast System; Ground Delay Program; Probabilistic Forecast; Monte-Carlo Simulation

## I. INTRODUCTION

The primary challenge to air traffic management in the San Francisco area is the low altitude cloud layer that develops overnight in the San Francisco Bay area. This layer is called marine stratus and has a tremendous impact on San Francisco Airport (SFO) arrivals since it precludes simultaneous arrival operations on closely spaced parallel runways, reducing the arrival capacity from 60 to 30 flights per hour. Frequently, the stratus layer fails to burn off before the first bank of arrivals are scheduled to land in the morning, and a Ground Delay Program (GDP) must be implemented. During periods when arrival demand exceeds capacity, a GDP defers the excess demand to later periods with available capacity by assigning ground delays to flights at their departure airports. If a GDP is issued with an end time later than the actual stratus burn-off time, unnecessary delay is absorbed on the ground at the departure airports, and arrival capacity is wasted. On the other hand, if the GDP end time is earlier than the stratus burn-off time, airborne holding and diversions may occur, and safety may be compromised. The latter is a risk that air traffic controllers prefer to avoid, thus leading to current practices that tend to result in overly-conservative GDP end times. The frequency of marine stratus occurrence, combined with the tremendous aviation

impact, motivated the National Weather Service (NWS) to develop the prototype SFO Marine Stratus Forecast System [1], which is designed specifically to predict the important *stratus clearing time* in the SFO approach zone.

The model presented in this paper is motivated by research to explore the integration of probabilistic weather forecasts in traffic flow management (TFM) decision making [2]. Operational decision-making in the National Airspace System (NAS) today is primarily based on deterministic information of traffic and weather, despite the fact that there is inherent uncertainty in both of these elements. Attempts at integrating probabilistic weather forecasts into operational decision-making have proven to be challenging. Convective weather forecasts in the en route environment include uncertainty in multiple dimensions, including time, space, and severity. This requires complex models to integrate probabilistic weather forecasts with TFM decision making. In contrast, the SFO Marine Stratus Forecast System provides an opportunity to explore the integration of probabilistic weather forecasts into TFM decision-making with the reduced complexity of one dimension. This system provides a forecast of a single weather parameter (the time at which the stratus will burn off) at a fixed geographical location (the SFO approach zone).

In this paper we present a model that integrates the probabilistic forecast of stratus clearing into the current process of modeling and issuing GDPs at SFO. By utilizing the probabilistic nature of the forecast, the model can select the best GDP parameters given the uncertainty in the time of the stratus burn-off, addressing both the objectives of minimizing delay and managing risk. This model can be implemented in today's environment, with no required changes to the data, procedures, and software used to issue GDPs. By using the probabilistic forecast of stratus burn-off at SFO more effectively, GDPs in today's environment can be issued less conservatively, minimizing the overall ground delay, unused arrival slots, unnecessary delay issued, and the number of aircraft affected by the GDPs. This model is an important step toward integrating probabilistic weather forecasts with TFM decision support tools.

## II. BACKGROUND

The current FAA mechanism for building GDPs is to solve a deterministic single airport Ground Holding Problem (GHP) using a method called Ration-By-Schedule (RBS) [3] that focuses on maximizing airport throughput while assigning ground delay to flights in a manner that ensures an equitable distribution of delays across major airlines. Based on the deterministic schedule of future airport capacity and arrival demand, the algorithm first generates a fixed schedule of arrival slots and then generates slot assignments for each flight. The algorithm is implemented in a software tool called the Flight Schedule Monitor (FSM). The significance of this algorithm is that it was developed under the Collaborative Decision Making (CDM) program, and the objective is to develop “equitable” delay assignments across airlines rather than to minimize a system-wide cost tradeoff between ground and airborne holding costs. In the CDM paradigm, the FAA is responsible for forecasting capacity and determining the arrival slot schedule, whereas the airlines have flexibility in prioritizing slot assignments based on the set of slots allocated to them [4].

Given the uncertain arrival capacity created by the uncertain stratus clearing time, we need to solve a Stochastic GHP. Several formulations and solutions for the stochastic GHP have been developed since 1987, but those developed prior to 1998 all reflect the notion of minimizing a global cost metric, which deems their eventual acceptance and implementation highly unlikely, given the CDM paradigm mentioned above. The reader is directed to [5] for a more detailed literature review of stochastic ground holding research. Ball and Hoffman [6] developed the first stochastic model to consider equity. It uses a similar objective function to earlier methods that trade off ground vs. airborne holding costs, but the decision variables are the number of arrival slots to assign during each time period over the planning horizon, rather than the assignment of flights to timeslots. In addition to drastically improving the solution performance, this model formulation provides a slot availability schedule that can be post-processed by another algorithm, such as RBS, to create “equitable” assignments of flights to slots. This algorithm is classified as static since the planning decision is made once, and stochastic since the decision is made with regard to uncertain capacity.

Mukherjee and Hansen [5] create a dynamic stochastic model for the single airport GHP. This model is also based on uncertain capacity but includes the ability to revise solutions dynamically based on updated capacity estimates as time evolves. The model may defer the decision of whether or not to delay a flight until it is scheduled to depart, thus, decisions for shorter length flights will be made closer to departure time, based on more accurate (hopefully) capacity estimates. Mukherjee also addresses equity considerations by extending the objective function to penalize solutions based on how much they differ from an RBS solution. The decision maker can control the importance of equity by altering the objective weight. Mukherjee also develops and tests an intra-airline substitution-optimization model that could be used by an airline to revise their slot assignments after each dynamically revised decision during the life of the program.

Although extensive efforts are made here to meet current CDM objectives of equity and flexibility, airlines may be resistant to repetitive re-planning since they are burdened with considering the downstream impact of aircraft delays on the rest of their schedule.

Both of the stochastic GHP models described above have, at their core, an economic objective of minimizing a linear combination of ground delay and airborne holding costs, both of which are the burden of the airlines. Both models also allow for the consideration of equity, another airline goal, in their ultimate application. What is missing from either model is the consideration of Air Traffic Control (ATC) risk. Early deterministic GHP models tend to focus on minimizing airline costs under the constraint that demand will be maintained at or below capacity in any time period. With probabilistic models, we must consider the likelihood that certain outcomes will result in large amounts of airborne holding. Both of these models essentially assume an unlimited arrival queue and do not consider any additional penalties such as diversions or reduced safety associated with large numbers of holding aircraft. While airlines must bear the cost of airborne holding, diversions, and (indirectly) reduced safety, they have an economic incentive for assuming some risk of those outcomes in order to realize the benefit of reduced delays. However, ATC has little incentive to assume the risks of excessive holding because they do not directly benefit from delay reduction in a GDP.

Our model is very close to [6] in function; it is static, stochastic, considers airborne and ground holding costs in its objective, and relies on RBS to fit within the CDM environment. The differences are (1) we also consider the risk of unfavorable ATC risks in our objective; (2) we solve our model by simulating hundreds of possible capacity scenarios based on empirical forecast error, rather than just a handful; and (3) we evaluate all reasonable decision alternatives with a simulation and choose the one with the lowest expected cost.

## III. THE NATURE OF SFO GDPs

The number of GDPs issued throughout the NAS from September 1998 through July 2007 totals 9,350 programs.<sup>1</sup> SFO leads all airports in the number of GDPs with a total of 1,450, 15.5% of all programs. Concentrating on methods to reduce ground delay for flights arriving in SFO can have significant positive impact on the NAS. Table 1 includes a few statistics compiled over the GDPs issued between May 15 and October 15 during 2005, 2006, and 2007; a total of 172 programs.

There are some interesting characteristics of SFO GDPs that are evident from these metrics. Only 50% of the flights arriving during a GDP timeframe are nonexempt and issued a departure delay. This is due to the number of flights that are either already airborne when the program is issued or are outside of the geographical scope of the program. Of those flights, the average delay is 44 minutes, but can be as high as 98 minutes.

---

<sup>1</sup> <http://cdm.fly.faa.gov/ad/gdp.html>

TABLE I. AVERAGE SFO GDP STATISTICS OVER 2005-2007

GDP Statistic	Value
Initial Average Total Flights	159
Initial Average Affected Flights	79
Initial Average Total Delay (Minutes)	3,642
Initial Average Maximum Delay (Minutes)	98
Initial Average Delay (Minutes)	44
Planned Average Initial Duration	4:20
Planned Average Overall Duration	4:51
Actual Average Duration (Cancel – Start Time)	2:52
Early Average Cancel Time (Planned – Actual Duration)	1:59
Average Lead Time (Start Time – Issue Time)	2:32

SFO GDPs are planned to initially run for four hours and 20 minutes. Including extensions to the programs, the average planned duration is 4:51, an increase of 31 minutes. The actual average duration of the programs is 2:52. This means that programs on average are planned to run approximately two hours longer than they are actually needed. In addition, an extension is only issued in 15% of the GDPs. These last two points illustrate the conservative nature taken by the FAA when issuing GDPs in order to mitigate the risk of excess airborne holding after the end of the program.

#### IV. SFO STRATUS FORECAST SYSTEM OVERVIEW AND USE

The SFO Stratus Forecast System [7] is based upon four different models for estimating the stratus clearing time and a model for combining those four into a consensus forecast. Three of the four models are automated to run starting at the top of the hour on the following hours: 9Z, 11Z, 13Z, 15Z, 16Z, 17Z, and 18Z. The fourth model is based on satellite imagery which requires sunlight, so it first becomes available with the 15Z run. After the models have run, the deterministic consensus forecast is generated as a weighted combination of the available component forecasts.

In addition to the consensus forecast, there are two measures of uncertainty produced by the system. The first is the confidence assigned to a forecast, either “good” or “low”. These values are based on input conditions which were found to cause marked changes in forecast accuracy. The second uncertainty measure produced is the probability of clearing before 17Z, 18Z, 19Z, and 20Z. This probability is based on empirical history of all forecasts made with a clearing time in the same 15 minute time bin. The probabilities are rounded to increments of 5%. The cumulative clearing time probabilities provided by the consensus forecast are attempting to communicate the underlying probability distribution to a human decision maker. However, they only provide a rough approximation of the actual distribution of clearing time. Additionally, the empirical error distribution used to calculate the four probabilities is based on groupings of the dependent variable (clearing time) rather than independent variable(s). In Fig. 1, we show the empirical distributions of forecast error classified by forecast confidence. The dashed stepped line is one sample of the clearing probabilities given in a consensus forecast. In comparing the values of the consensus clearing probabilities to the complete CDF, we can see that much of

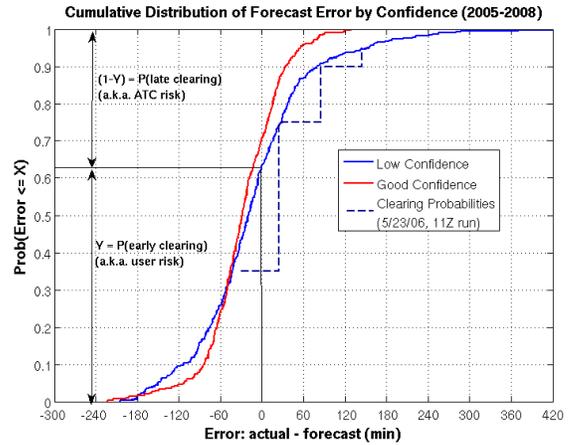


Figure 1. Cumulative Distribution Function of the Empirical Forecast Errors of the Forecasts Provided by the SFO Marine Stratus System, Classified by Forecast Confidence

the detail is omitted by using 60-minute steps in the probability distribution.

While the published probabilities are intended to help operational decision-makers, it is unlikely that they are actually being used in decision-making. The model presented in this paper addresses this issue by integrating the probabilistic forecast of clearing time automatically into the GDP modeling process and eliminating the need for human interpretation of probabilistic information.

#### V. MODEL OVERVIEW

In order to include the uncertainty in a forecast in a decision support model and to evaluate the risk associated with any given decision, we need to know what the possible outcomes are and how likely they are to occur. We derive the likelihood information by adding the consensus forecast clearing time to the appropriate error distribution in Fig. 1 based on forecast confidence. We then define performance metrics which allow us to measure the outcomes of a given decision over the entire distribution of possible clearing time outcomes. This essentially turns probabilistic inputs into probabilistic outcomes.

We have developed a model to find the “optimal” GDP parameters based on criteria such as minimizing unnecessary delay and managing the risk of airborne holding. The parameters that are “optimized” as part of our model include start time, end time, and scope. Our actual approach is to find the optimal parameters one-by-one instead of trying to evaluate all possible combinations of these parameters at once. Different decision criteria are relevant for different parameters, further justifying this approach. Once the optimal value of a parameter is found, it is set as a constant before evaluating the next parameter. We evaluate the parameters in the following order, using the following rules to set the other parameters:

1) *Start Time*—This is easily set by finding the start time of the earliest 15-minute increment in which the demand exceeds the capacity.

2) *End Time*—The start time is set based on the previous step, the scope is set to “All,” capturing all domestic departures, and each possible end time (15-minute increments only) is evaluated.

3) *Scope*—The start time and end time are set based on the previous steps, and each possible scope selection is evaluated.

The details of the models that determine the GDP end time and GDP scope are included in the following two sections.

#### A. *GDP End Time*

A Monte Carlo simulation approach is used to find the best GDP end time. This requires the full distribution of the probabilistic forecast of stratus clearing provided by the SFO Stratus Forecast System.

The steps of the model for selecting the GDP end time are as follows:

1) Simulate the distribution of clearing times as the deterministic forecast time plus the distribution of errors from the empirical CDF of forecast errors.

2) Based on the GDP start time and the deterministic forecast of stratus clearing time, generate the list of GDP end times that reasonably can be considered. Note that a GDP can only end by 15-minute period at 14, 29, 44, or 59 minutes past the hour. For each possible GDP end time, model the GDP in the FSM and generate the resulting flight delays. Note that all GDP parameters should stay constant except the GDP end time and the rate, which should stay consistent with the GDP end time. The rate is set to 30/hour prior to the GDP end time and 60/hour after the GDP end time. The scope is set to “ALL,” including the three west coast Canadian airports.

3) Generate metrics, such as unnecessary delay and airborne holding, which are used to evaluate GDP end time. Some metrics are dependent not only on the GDP end time, but on the GDP cancel time/stratus clearing time. This illustrates the importance of using a Monte Carlo simulation approach to simulate the range of likely clearing times. For each possible GDP end time scenario (result of Step 2), we generate the metrics for each simulated GDP cancel time in the set of clearing times generated in Step 1. We calculate the mean metrics for each GDP scenario as the average of each metric in the set of metrics. This results in the expected value for each metric given the probabilistic forecast of stratus clearing.

4) Calculate the cost of each scenario using the objective function, and find the minimum value over all GDP end time scenarios. This is the optimal GDP end time.

The goal of the model is to quantify the impact of uncertain clearing time on the outcome metrics. We wish to choose an end time which results in good performance over a wide variety of clearing time scenarios. In this problem, some of the scenarios happen to result in a “perfect” outcome where the decision value may actually provide the best possible answer (GDP end time = actual clearing time).

However, that is not the goal of this model. Instead, it is conceptually more important to imagine that the model will choose end times that perform well, on average, and will guard against the risk of undesirable outcomes, such as a late clearing time.

Risk is expressed in terms of likelihood and severity. For example, any reasonable decision for GDP end time has some probability of being too late. In this case, the severity is defined by the amount of financial loss to users who experience unnecessary delay while some available capacity goes unused. The user risk, therefore, is based on economic loss and is realized when GDP end times are too late and have to be canceled early. ATC risk, however, occurs when the GDP ends before the stratus clears. When this occurs, an airborne holding queue forms and quickly becomes large enough to have a negative impact on safety and controller workload. In the interest of maintaining safety, drastic remedial actions such as Ground Stops or diversions could result. ATC risk is very critical and must be explicitly guarded against by the model. We include two terms in the objective function to mitigate ATC risk.

The objective function includes the expected unnecessary ground delay and expected total airborne holding. The terms have the same units, similar scales, and increase/decrease monotonically with increasing GDP end time. This is also consistent with earlier stochastic GHP models, which use ground holding and airborne holding in their objectives. We include weights for both metrics in order to quantify the FAA's willingness to trade ground delay for air delay. This captures the tradeoff between the FAA's desire to serve the industry efficiently and its operational and safety concerns [6]. We apply a weight of 2 to the mean total airborne holding after end time, giving twice the weight to airborne vs. ground delay, consistent with previous research [5].

We then add two terms to mitigate the ATC risk. The first is based on the ability to extend a program. If the expected number of exempt flights by 1600Z in the first hour after the GDP end time exceeds a threshold, we deem that the program is not extendable and eliminate it as a feasible decision. This term ensures that the ATC risk of ending the program too early can be mitigated by extending the program at 1600 after the 15Z forecast becomes available. We use a conservative threshold of 20 flights, despite a capacity of 30. In several instances, this precluded end time decisions that were economically advantageous.

The last term in the model is probably the most important for risk aversion. From our empirical data and the TFM guidelines that recommend selecting GDP end time by adding two hours to forecast clearing time [8], we can show that traffic managers are currently willing to accept only a single digit percentage chance of GDP end time being too early. The evidence is in Fig. 1, where you can see that the forecast clearing time plus 120 minutes results in about a 94% chance that the clearing time is before GDP end time, which corresponds to a 6% chance of late Stratus clearing time (ATC risk) when forecast confidence is low. Similarly, there is only a 1% chance when confidence is good. What we desire is an objective term that permits low probabilities of

ATC risk and quickly increases to penalize the objective heavily for risky end time decisions. The Upper Partial Moment (UPM) of the maximum flights in holding at the end of the GDP is the correct type of term for this situation, but it is not naturally sensitive enough to moderate levels of risk to ensure that only small amounts of risk will be tolerated. To improve the level of risk aversion, we first use the exponential function to create the rapidly nonlinearly increasing penalty values for moderate increases in ATC risk. We then add an optional multiplier to the exponent to help tune the shape. This multiplier is essentially a risk aversion dial in the objective function. High values lead to a response that is closer to an “L” shape, where lower values lead to a more linear response. We keep this multiplier at 1 in our objective function but recommend its inclusion in the model as a way for users to adjust their risk tolerance.

The resulting objective function is

$$Z(t, \xi) = E_{\xi} \{w_1 U(t, \xi) + w_2 H(t, \xi) + e^{k(upm(F,t,\xi))} + w_3 T(t, \xi)\}, \quad (1)$$

where:

$t$  is the chosen GDP end time, and

$\xi$  is the probability distribution of clearing time, and

$E_{\xi}$  is the expectation over the distribution of clearing times, and

$U$  is the total unnecessary ground delay in minutes issued to the non-exempt flights in the GDP, and

$H$  is the total airborne holding after the GDP end time, and

$F$  is the max flights in holding after the GDP end time, and

$w_1$  is the weighting coefficient for unnecessary ground delay, and

$w_2$  is the weighting coefficient for airborne holding after the GDP end time, and

$w_3$  is an arbitrarily large weighting coefficient, and

$k$  is the risk controlling coefficient, and

$T$  is equal to 1 if the number of exempt flights in the first new hour if the GDP extended at 1600Z exceeds a threshold, 0 otherwise.

After evaluating  $Z$  for all reasonable values of GDP end time ( $t$ ) and simulating  $\xi$  we choose the end time that yields the lowest objective.

Fig. 2 illustrates the objective function on a sample GDP on May 30, 2007.

The competing objectives of minimizing airborne holding after the GDP end time and unnecessary ground delay are plotted in green and red, respectively. The risk component, the exponential function of the UPM of the number of flights in holding after the GDP end time, is plotted in teal. Note the sharp increase in this component at earlier GDP end times.

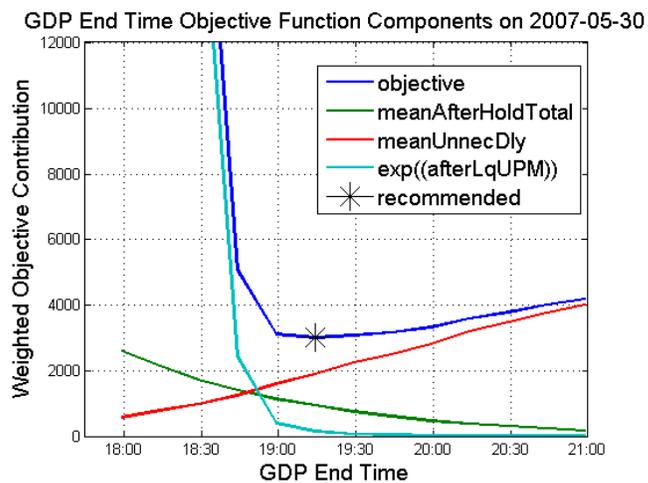


Figure 2. Illustration of the GDP End Time Objective Function for the GDP on May 30, 2007

This component in the objective function drives solutions away from earlier end times that hold too much risk of a scenario where the stratus clears later than the end time and there is not enough time to extend the program to manage the demand. The resulting objective function is plotted in blue, with the recommended GDP time noted at the minimum point.

### B. GDP Scope

Given the GDP start and end times, the model now selects the best scope for the program. The scope sets the geographical region for the program. All flights departing from airports within that region are nonexempt and can be issued departure delays. The scope can be a critical decision when modeling a GDP; the broader the scope, the more flights captured, which reduces average delays and usually results in more equitable programs. But larger scopes can increase unnecessary delay, including flights with longer flying times that must start accumulating ground delay earlier than short haul flights. Scope can be set based on center tiers or distance from the airport. We use center tiers to illustrate the model, but our algorithm supports either method.

In the case of SFO, scope is not as critical a factor as it can be at other airports. The step from one scope to next largest scope tends to not capture many additional flights, especially at the larger scopes. This is due to the fact that the SFO GDPs tend to have shorter durations in the early morning when the traffic is primarily short-haul arrivals. This is shown in Fig. 3. Each of the colored series plots the number of flights added by increasing the scope to the next largest tier. For example, the purple line shows the additional flights captured when moving from a 6 West (the smallest scope) to a 10 West program (the next largest scope). This first step in scope usually has the largest impact in terms of number of flights added, six on average. The remainder of the scope steps, 10 to 12 West, 12 to 15 West, and 15 West to All, tend to add very few flights, if any. The average additional flights added are two, two, and one, respectively. The dashed black line, scaled to the right axis, shows the

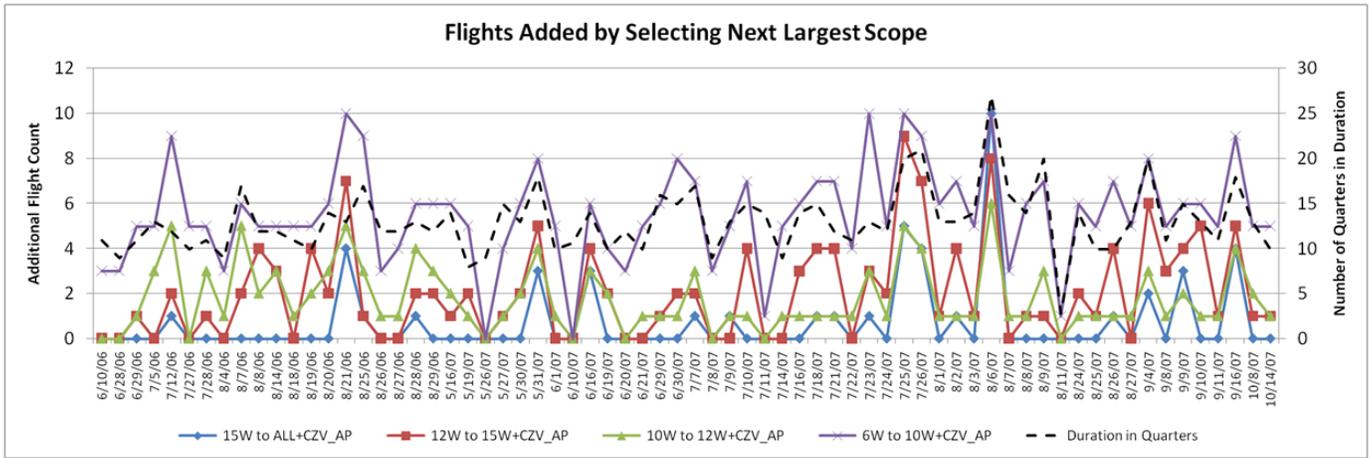


Figure 3. The Additional Number of Affected Flights by Increasing the Scope Plotted Against the GDP Duration

number of 15-minute periods included in the duration of each of the GDPs. There is a strong correlation (as high as 0.67) between the GDP duration and the number of additional flights included in each step of scope. This implies that the longer the duration of the GDP, the more relevant the scope decision becomes.

Though scope selection may not have a big impact on the majority of SFO GDPs, it still can be important for those programs that will run longer due to later forecasted stratus clearing. For that reason, scope selection is included in the model. This stage of the model uses the same approach as the previous stage, which selected the GDP end time using a Monte Carlo simulation of the distribution of stratus clearing times. The only difference in this stage of the model is the set of relevant metrics generated and the objective function used. Each scope scenario for evaluation is generated using the “optimal” GDP start and end time already identified by the previous stages of the model.

The key metrics identified as relevant to the decision of scope are as follows:

- Average Delay—the average delay over all nonexempt flights in the GDP
- Mean Unnecessary Delay—the delay that will be already unnecessarily absorbed if the GDP cancels earlier than planned
- Ability to Extend at 1600Z—measures the number of nonexempt controllable flights in the hour following the current GDP end time if extended at 1600Z
- Delay Variability—a measure of carrier equity; the standard deviation in the carrier average delay
- Equity Metric for Airlines (EMA)—a measure of equity determined by comparing the delay assigned in a proposed program to that which results by using an airborne holding model. Any deviation from the airborne holding model is decreased equity. The metric assesses the variance in delay by carrier.

- Equity Metric for Flights (EMF)—Similar to EMA, but assesses the variance by flight instead of carrier

The last three equity-related metrics are included in the FSM software and are used operationally to assess the equity of different GDP scenarios [9].

The objective function is

$$Z(s, \xi) = E_p\{w_1A(s, \xi) + w_2U(s, \xi) + w_3V(s, \xi) + w_4T_1(s, \xi) + w_4T_2(s, \xi) + w_4T_3(s, \xi)\}, \quad (2)$$

where:

$s$  is the chosen GDP scope, and

$\xi$  is the probability distribution of clearing time, and

$E_p$  is the expectation over the distribution of clearing times, and

$A$  is the average delay per flight, and

$U$  is the total unnecessary ground delay in minutes issued to the nonexempt flights in the GDP, and

$V$  is the delay variability across carriers, and

$T_1$  is equal to 1 if the number of exempt flights in the first new hour if the GDP extended at 1500Z exceeds a threshold, 0 otherwise, and

$T_2$  is equal to 1 if the EMA threshold is violated, 0 otherwise, and

$T_3$  is equal to 1 if the EMF threshold is violated, 0 otherwise, and

$w_1$  is the weighting coefficient for the average delay, and

$w_2$  is the weighting coefficient for unnecessary ground delay, and

$w_3$  is the weighting coefficient for delay variability, and

$w_4$  is a very large weighting coefficient.

After evaluating  $Z$  for all reasonable values of GDP scope ( $s$ ) and simulating  $\xi$  we choose the scope which yields the lowest objective.

A total of six adjustable parameters are used in the objective function; there are three weights and three thresholds. An analysis was conducted to understand the sensitivity of the objective function to different combinations of weights and thresholds. Fig. 4 plots the number of programs in which a particular scope was selected by the objective function. Each series represents a different set of weights and thresholds, and a wide range of scenarios are represented. The overall shape of each trend line is fairly consistent, especially for the broader scopes. This implies that the weights and thresholds are not very sensitive. More variability exists in the smaller scopes, where there is more likely additional flights included by moving to the next largest scope, as was explained earlier with Fig. 3.

The weighting and threshold parameters selected to test our model are illustrated in the black dashed line scenario in Fig. 4. This scenario weights average delay and delay variability 30 times greater than unnecessary delay. This is because unnecessary delay is a sum over all flights, where average delay is on the scale of an average value per flight. The exempt flights in extension threshold value selected is consistent with that used in the selection of GDP end time (20 flights), and the values used for EMA and EMF are the upper bound to those values considered to reflect “good” equity, which are eight in the case of both metrics [9].

## VI. RESULTS

The end time and scope selection models are evaluated over 68 SFO GDPs during the severe weather season in 2006

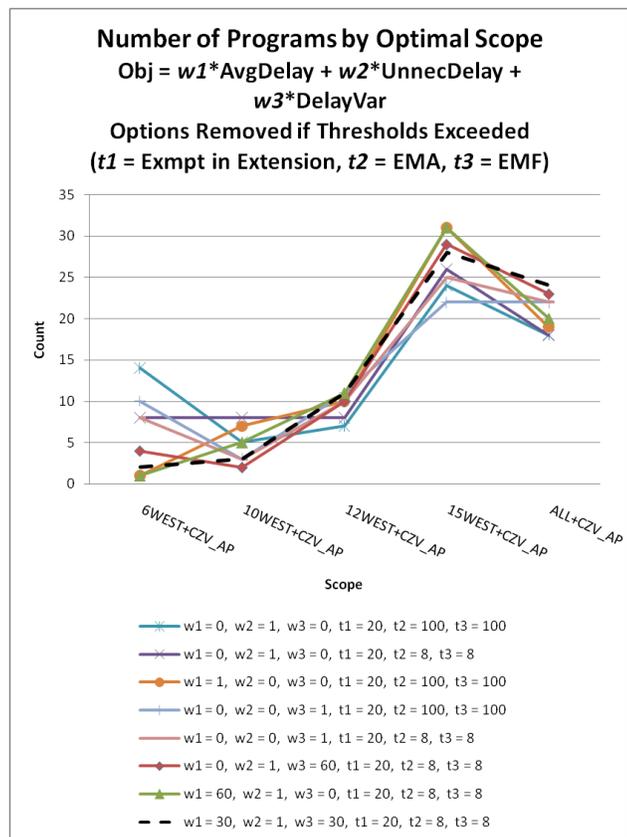


Figure 5. Count of Programs by Optimal Scope for Various Objective Function Weights and Thresholds

and 2007. Both the end time and scope selections chosen by the model are compared to those that were selected operationally.

In the case of end time, an end time was chosen operationally that was on average 128 minutes later than the clearing time. The model’s selection of end time was on average only 56 minutes later than the clearing time, a 56% reduction in unnecessary GDP duration. The model end time selected in 10 of the 68 programs (15% of the programs) was more than five minutes earlier than the stratus clearing time, and an extension would have been required for these programs. But there would have been plenty of time to issue these extensions; the time between the initial issue time and end time was on average for these programs over five hours. In 59 of the programs, the model matched or out-performed operations, making the same or better selection for the end time. In only 13% of the programs did the end time selected operationally result in a better end time than that selected by the model.

In the case of scope selection, the majority of the time the model selected a broader scope than was used operationally. We compare the scopes by assigning a value of 1, 2, 3, 4, or 5 to each program based on whether the scope selected was 6 West, 10 West, 12 West, 15 West, or All respectively. On average the operational scope selection is 1.7 (which is closer to 10 West than 6 West), while the average model selection is 4.0, a 15 West. The difference in scope from the model selection vs. the operational selection is 2.3 scope tiers, meaning that the model on average selects scopes that are over two step scope selections larger than was selected operationally. In only three cases did the model select a scope smaller than the one used operationally. These results are due to the importance placed on equity in the objective function and the smaller variability in unnecessary delay across the scopes.

## VII. BENEFITS ASSESSMENT

To capture the benefits of using our proposed model, we measure specific key metrics over three possible scenarios as illustrated in Fig. 5. First, using the actual GDP parameters used historically, we remodel those GDPs as they were actually run. This provides the value of the key metric for the

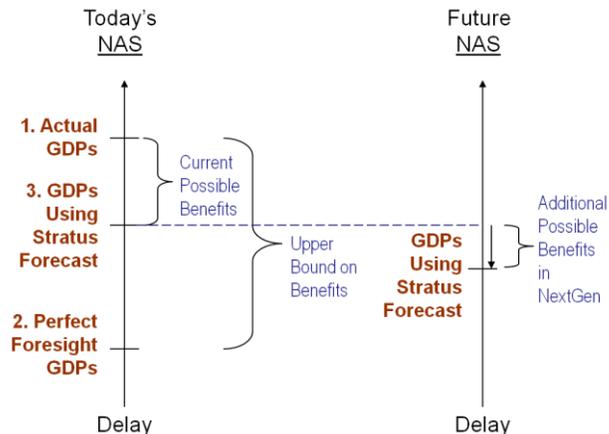


Figure 4. Approach to Assessing Benefits of the Model Using the Probabilistic Forecast of Stratus Clearing

actual GDPs, illustrated as number 1 on the figure. Second, we modify the GDP parameters to the ideal end time that should have been used to run that program, the actual clearing time. This captures how the GDP would have been run if the users had perfect foresight of the stratus clearing time. This provides the value of the key metric for the perfect foresight GDPs, illustrated as number 2 on the figure. Lastly, we use our model, which takes advantage of the probabilistic forecast to generate the recommended GDP end time and scope. This provides the value of the key metric for the GDPs using the stratus forecast, illustrated as number 3 on the figure. The difference between the actual GDPs and the perfect foresight GDPs provides an upper bound on the possible benefits that can be gained in improving the modeling of GDPs. The difference between the actual GDPs and the GDPs generated using our model provides the benefits that could have been achieved using our model in today's NAS environment.

We recognize that additional benefits can be achieved in the NextGen environment. There are various papers that have been published that propose mechanisms for managing the traffic flow into SFO when ceilings are below minimums that replace the current mechanism of issuing GDPs, specifically Ration-by-Schedule. Some of these possible approaches include a dynamic stochastic model [5], a stochastic integer program with dual network structure [10], optimization based on distance [11], ration-by-distance with equity guarantees [12], banking constraints [13], and optimization and mediated bartering models [14]. We hypothesize that probabilistic forecasts of stratus clearing times in conjunction with more efficient traffic flow management models can result in even further benefits than can be realized in today's system, which is shown on the right-hand side of Fig. 5.

The primary metric of importance when evaluating the benefits of improved GDPs is issued ground delay. Every minute of ground delay is a cost to the NAS users, and thus any reduction in ground delay can provide significant benefits. Fig. 6 summarizes the ground delay issued for each

GDP under the three different scenarios: what actually occurred, what our model generated, and what ideally should have been done (GDP end time = stratus clearing time). Our model reduced the total ground delay by 17% of the actual delay as shown by the red line. But some of the ground delay, 43% to be exact, was necessary since the stratus did not clear prior to the GDP start time in 66 out of the 68 GDPs. With the ideal GDPs acting as a lower bound on ground delay, our model achieved a 29% reduction of the ground delay that was unnecessary.

It is important to note that no model can successfully eliminate all unnecessary ground delay and be able to consistently find the ideal GDP end times. The main reason for this is the need to manage ATC risk. The cost of running a GDP that ends too early can be high, resulting in excessive airborne holding, diversions, and ground stops. Our model is not only focused on the goals of the NAS users (smaller amounts of ground delay and less affected flights), but is also focused on the goals of the FAA, ensuring that the available capacity can adequately accommodate the traffic demand. Since our model considers the risks in selecting a GDP end time that is earlier than the clearing time, we do not expect to reduce ground delay close to the ideal number. Thus, a 29% reduction is significant given our conservative approach to managing ATC risk.

The same approach was used to calculate the benefits in terms of the number of flights affected by the GDPs, and is captured in Fig. 7. It is interesting to compare the benefits in the number of flights controlled vs. the benefits in ground delay. The reduction in affected flights is 18%, as opposed to 17% in ground delay. The reduction in unnecessarily affected flights is 39% compared to 29% for unnecessary ground delay. Percent reductions are higher for affected flights due to the arrival rates used in some of the actual GDPs. In some cases, a 30-rate was not used all the way to the GDP end time; the rate was gradually increased in the last hour or two, decreasing the ground delay but providing no decrease in the number of flights affected. A 39% reduction in unnecessarily delayed flights can have significant impact on the airlines,

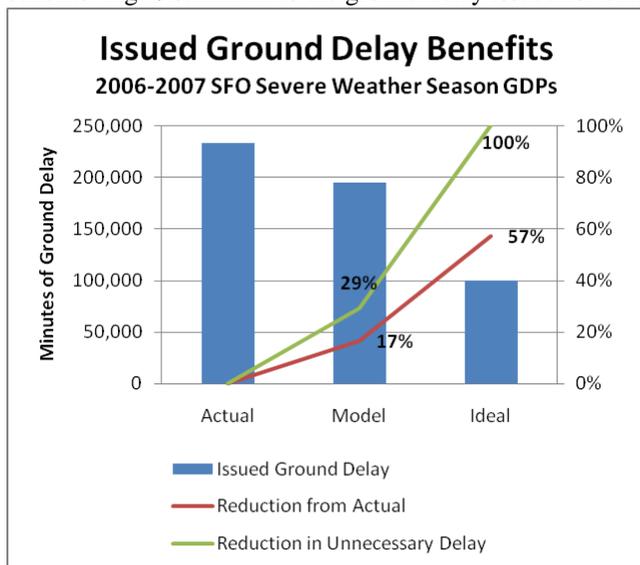


Figure 6. Model Benefits From Reductions in Ground Delay

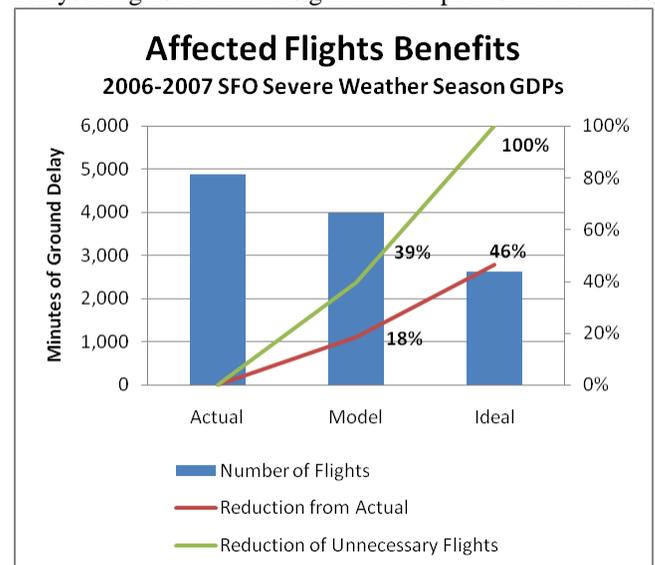


Figure 7. Model Benefits From Reductions in Affected Flights

greatly reducing the impact on their aircraft, crew, and passenger networks by maintaining a much higher number of connections.

We further estimate what the reduction in ground delay could mean in terms of reduced costs to the NAS users. Using data from 2005–2007, we calculate the average total ground delay issued per SFO GDP during the severe weather season when the SFO Stratus Forecast System is in use (May 15–Oct. 15). We also calculate the average number of annual SFO GDPs during those same time periods. This provides the average severe weather season number of ground delay minutes issued for SFO. Since our model results in a reduction of issued ground delay by 17%, we can calculate the number of ground delay minutes that can be eliminated with the use of our model. Using the ATA-provided value of \$60.46<sup>2</sup> as the cost per minute of delay, we can then calculate the expected annual savings to NAS users from using our model, \$2.83M. The results are summarized in Table 2.

TABLE II. ESTIMATION OF MODEL SAVINGS

Ave Actual Delay per Severe Wx Season GDP	3,642.0
GDPs 5/15 - 10/15 2005	120
GDPs 5/15 - 10/15 2006	47
GDPs 5/15 - 10/15 2007	65
Ave Annual Severe Wx Season GDPs	77
Ave Delay Minutes per Severe Wx Season	281,648
Ave Percent Reduction in Ground Delay	17%
Ave Annual Savings in Minutes	46,822
Cost per Minute Delay	\$60.46
Total Annual Savings	\$2,830,843

### VIII. CONCLUSION

The forecast generated by the SFO Stratus Forecast System provides an ideal opportunity for exploring the integration of probabilistic forecasts with TFM decision-making. The model presented in this paper can provide the link between the probabilistic forecast of stratus clearing and the FSM software used to model and issue GDPs, automating the interpretation of the uncertainty associated with the forecast. This model can be implemented in today’s environment and provides an opportunity for operationally testing a TFM decision support tool integrated with probabilistic weather forecasts. Based on the initial benefits assessment, this model has the potential to reduce unnecessary ground delay issued for SFO arrivals by 29% and to provide \$2.8 million in annual savings to the air carriers.

In fact, what we have also done is demonstrate how empirical error information can be used to create probabilistic forecasts from a deterministic product. The initial assessment of the model presented in this paper focused on its implementation at SFO due to the one dimensional nature of the problem and the availability of the SFO Stratus Forecast System product, which provides an abundance of forecast and validation data. However, this

approach could be applied at other airports in the future using different forecast products as long as historical forecast and verification data is available. Future decision support systems could then ingest forecasts as probability distributions, and base decision on distributions of outcomes. As the weather forecast products mature to include measurements of uncertainty, Traffic Management Initiatives (TMIs) such as GDPs can be issued such that ATC risk is specifically managed while still achieving benefits in terms of less delay for the air carriers. As with SFO, we expect that TMIs across the NAS can be issued less conservatively, resulting in reductions in delays and more efficient use of available capacity.

### ACKNOWLEDGMENT

The authors appreciate the support and guidance at NASA Ames Research Center from William Chan, Shon Grabbe, and Cedric Walker. The authors also thank David Clark of MIT Lincoln Laboratory and David Reynolds and Ken Venzke of the National Weather Service for their assistance in accessing and understanding the SFO Stratus Forecast System. We also appreciate the assistance of Midori Tanino of the FAA in accessing the Flight Schedule Monitor software to support our benefits analysis.

### REFERENCES

- [1] Clark, D. A. and Wilson, F. W., “The San Francisco marine stratus initiative”, 7<sup>th</sup> Conference on Aviation, Range and Aerospace Meteorology, Long Beach, CA, pp. 384-389, 2003.
- [2] Clark, D., “Investigating a new ground delay program strategy for coping with SFO stratus”, Aviation, Range, and Aerospace Meteorology Special Symposium on Weather - Air Traffic Management Integration, 89th AMS Annual Meeting, Phoenix, AZ, 11-15 January, 2009.
- [3] Hoffman, R., Hall, W. D., Ball, M. O., Odoni, A. R., and Wambsganss, M., “Collaborative decision-making in air transportation”, Technical Report #RR-99-2, NEXTOR, UC Berkeley, 1999.
- [4] Chang, K., Howard, K., Oiesen, R., Shisler, L., Tanino, M., and Wambsganss, M., “Enhancements to the FAA ground-delay program under collaborative decision making”, Interfaces Vol. 31, pp. 57-76, Feb. 2001.
- [5] Mukherjee, A. and Hansen, M., “A dynamic stochastic model for the single airport ground holding problem”, Transportation Science vol. 41 no. 4, pp.444-456, Nov 2007.
- [6] Ball, M. O., Hoffman, R., Odoni, A., and Rifkin, R., “The stochastic ground holding problem with aggregate demands”, Technical Report #RR-99-1, NEXTOR, 1999.
- [7] MIT Lincoln Laboratory, “SFO marine stratus forecast system documentation”, April, 2005.
- [8] Artist, M., “SFO Ground Delay Program Guidelines”, FAA ATCSCC Memorandum, October, 2007.
- [9] Metron Aviation, Inc., “The Flight Schedule Monitor Users Guide, Version 8.5”, March 2007.
- [10] Ball, M. O., Hoffman, R., Odoni, A. R., and Rifkin, R., “A stochastic integer program with dual network structure and its application to the ground-holding problem”, Operations Research 51, pp. 167-171, 2003.
- [11] Ball, M. O. and Lulli, G., “Ground delay programs: optimizing over an included flight set based on distance”, Air Traffic Control Quarterly 12, pp. 1-25, 2004.
- [12] Hoffman, R., Ball, M. O., and Mukherjee, A., “Ration-by-distance with equity guarantees: A new approach to ground delay program planning and control”, Proceedings of the 7th USA/Europe Air Traffic Management R&D Seminar, Barcelona, Spain, 2007.

<sup>2</sup> <http://www.airlines.org/economics/specialtopics/ATC+Delay+Cost.htm>

- [13] Hoffman, R. and Ball, M. O., "A comparison of formulations for the single-airport ground-holding problem with banking constraints", *Operations Research* 48, pp. 578-590, 200.
- [14] Vossen, T. and Ball, M. O., "Optimization and mediated bartering models for ground delay programs", *Naval Research Logistics* 51, pp. 75-90, 2006.

#### AUTHOR BIOGRAPHIES

**Lara Shisler Cook** is a Principal Analyst with Mosaic ATM. She has over 11 years experience supporting research and development activities for the FAA and NASA in Air Traffic Management. Prior to that, Ms. Cook worked as an Operations Research Analyst for two major carriers, both passenger and cargo. She received an M.S. in 1991 in operations research and management science from George Mason University in Fairfax, Virginia. Her B.S. in mathematics was received in 1989 from The College of William and Mary in Williamsburg, Virginia.

**Bryan Wood** is a Senior Analyst with Mosaic ATM with over 6 years of experience as an analyst in the aviation industry. He is experienced with Geographical Information Systems (GIS) and the spatial analysis of air traffic and airspace data. Mr. Wood is also very familiar with aviation weather products and their use for decision support in air traffic management. Most recently, he has been involved with the analysis of airport surface operations and decision support systems. Mr. Wood received a M.S. in 2002 in industrial and management engineering from Rensselaer Polytechnic Institute in Troy, New York. He also received a B.S. in 2000 in industrial and management engineering from Rensselaer Polytechnic Institute.