10.3 QUANTIFYING THE BENEFIT OF TEMPORAL CORRELATION IN PROBABILISTIC WEATHER FORECASTS FOR STRATEGIC AIR TRAFFIC MANAGEMENT

Christopher Provan*
Lara Shisler Cook
Mosaic ATM, Leesburg, VA

1. INTRODUCTION

Aviation operations throughout the National Airspace System (NAS) are heavily impacted by weather. According to airline-reported data compiled by the Bureau of Transportation Statistics (BTS) at the United States Department of Transportation, weather-related delays made up 39% of total aircraft delay minutes and 76% of congestion-related delay minutes in 2011.** Based on estimates of direct and indirect costs of congestion-related delay to passengers, airlines, and other industries published in a report by the Joint Economic Committee Majority Staff (2008), the total annual cost of weather delays exceeds $30 billion. A separate report by the Weather-Air Traffic Management (ATM) Integration Working Group (2007) estimated that as much as two thirds of weather delay in the NAS was potentially avoidable through improved forecasts, procedures, and decision support.

Significant reduction in avoidable weather delays is one of the nine key characteristics of the Next Generation Air Transportation System (NextGen) as laid out by the Joint Planning and Development Office (JPDO) in its NextGen Concept of Operations (2010). In particular, probabilistic weather forecasts will need to be integrated into NextGen decision support tools (DSTs) in order to minimize delays while maintaining acceptable levels of risk. In many cases, traffic managers will no longer be responsible for interpreting weather forecasts and predicting the impact on air traffic operations. Instead, DSTs will use probabilistic decision models to recommend specific ATM actions in response to weather and traffic scenarios.

Much attention has been given to the development of probabilistic forecasts for use in ATM. Approaches include simple empirical uncertainty models that build error distributions around forecast values based on historical error data and ensemble forecast models that aggregate the results of multiple component models or multiple runs of a single model with perturbed inputs to build a probabilistic forecast. However, most of the attention has focused independently on the uncertainty associated with individual forecast dimensions (e.g., wind speed) at a single forecast time or, at best, the correlated uncertainties between a small number of forecast dimensions at a single forecast time. Far less often addressed are the correlations in forecast uncertainty across forecast times that can be important in many aviation applications. At the same time, much research and development of ATM algorithms and NextGen DSTs assumes that forecasts will be presented in such a way that include this temporal correlation in uncertainties (e.g., Mukherjee, et al, 2009; Provan and Atkins, 2010).

Temporal correlation is particularly important for forecast products used in strategic ATM. Strategic ATM focuses on proactively managing traffic levels at constrained NAS resources (airports, airspace sectors, key navigational fixes, etc.) over planning horizons of 2 hours or longer. Some strategic ATM initiatives can cover planning horizons longer than 8 hours.

Strategic ATM initiatives assign delays or reroutes to flights well before they reach the constrained resource so that flights can be

* *Corresponding author address:* Christopher Provan, Mosaic ATM, Inc., cprovan@mosaicatm.com, 801 Sycolin Road SE, Leesburg, VA, 20175.
** Data taken from BTS website on December 12, 2012: http://www.rita.dot.gov/bts/help/aviation/html/understanding.html

tactically managed at the resource without risking excessive airborne holding of flights, diversions (landings at airports other than the planned arrival airport), or higher than acceptable workloads on air traffic controllers. Most often, strategic ATM initiatives are issued in anticipation of weather forcing reduced capacity at a NAS resource: e.g., fog reducing arrival capacity at an airport or thunderstorms preventing use of a large section of airspace.

Because strategic ATM manages capacity over a long period of time, cumulative effects take on a greater importance. If arrival capacity at an airport is significantly reduced over a long period of time then any excess delivery of demand builds over time. This leads to airborne holding or diversions that can cascade through time until the constraint lifts and controllers have room to recover. Thus the optimal decisions are very different when faced, for example, with a short-duration weather event that is certain to occur but whose timing is uncertain versus a long-duration weather event that has a relatively low probability of occurrence. These two scenarios may be indistinguishable in a forecast that does not explicitly address temporal correlation. While human decision makers have the benefit of being able to discuss these events with meteorologists to gather such details, an automated ATM DST would need to have this information quantified as part of any input forecast.

This paper provides a quantitative assessment of the potential benefit of temporal correlation models to strategic ATM planning. In particular, we focus on the benefit of temporal correlation in forecasts for a DST that helps decision makers plan ground delay programs (GDPs). In order to isolate the benefit of explicitly modeling temporal correlations in forecasts, we use a simplified model of GDP planning with a linear programming optimization formulation based on work by Ball, et al (2003). Using this optimization model as a stand-in for a GDP planning DST, we analyze a range of case studies and compare the GDPs planned with and without temporal correlation.

The next section describes current GDP planning practices and discusses how this process might differ in a NextGen environment. Section 3 describes the GDP model and optimization formulation used in the studies. Section 4 provides more detail on the study design, and section 5 presents the scenarios and analysis results. The paper closes with conclusions and implications for continuing research and development.

2. PLANNING GROUND DELAY PROGRAMS

GDPs are a tool that traffic managers often use to address weather-related shortages in airport arrival capacity. GDPs control arrival demand at a single airport by assigning delays to flights before they have departed from their origin airports. These ground delays, typically taken while the plane is at the gate with engines off, are significantly less expensive for airlines and provide lower workloads and risk levels for air traffic controllers than comparable airborne delays (Cook, et al, 2004). Passengers also benefit by knowing about delays further in advance, having cellular phone and wireless internet access so that they can modify their post-arrival schedule, and often being able to wait out delays in more comfortable terminal buildings rather than on the aircraft.

Because flights must be assigned delays prior to their departures, GDPs must be planned at least 2-3 hours before arrival demand is expected to exceed airport capacity in order to fully control arrival demand. When capacity constraints are caused by weather, this requires weather forecasts looking perhaps 6-12 hours into the future in order to predict airport arrival capacity over the full duration of the weather event. At these horizons, forecast uncertainty plays a significant role in the effectiveness of a GDP. Overestimating the impact of weather on capacity means flights will be delayed more than necessary at their departure airports, resulting in unused capacity at the arrival airport. Underestimating the impact means flights will arrive at a rate that cannot be supported by the airport, resulting in costly airborne holding, increased workload and stress for air traffic controllers, and potentially very expensive diversions to other airports as aircraft fuel reserves drop below acceptable levels.

In current practice, GDPs are planned collaboratively by FAA traffic managers, meteorologists, and airline representatives. When a threat of inclement weather is observed, the collaborative parties will typically conduct a conference call to discuss the situation. Meteorologists provide an overview of current weather forecasts, and traffic managers and airline representatives determine whether a GDP is required and, if so, discuss the parameters that will be used for the GDP. While airlines interests are given weight in the decision process, the ultimate decision lies with an FAA traffic management specialist at the Air Traffic Control System Command Center (ATCSCC). As the conditions evolve and new forecasts are issued, programs can be updated or cancelled altogether.

Uncertainty in weather forecasts and the resulting uncertainty in future arrival capacity most often force ATCSCC specialists to issue conservative GDPs that have a high probability of keeping arrival demand below capacity. Flights are assigned more ground delay than is necessary in order to hedge against the risk of airborne holding and diversions. This unnecessary delay adds to airline costs and passenger inconvenience.

NextGen DSTs will aim to help reduce these unnecessary delays while still hedging against the risks of overdelivery by applying advanced optimization techniques to turn combined traffic and weather data into GDP parameters that cause minimal delay while being robust to forecast and capacity uncertainty. Whereas present day DSTs are primarily focused on providing situational awareness to traffic managers (scheduled traffic, weather forecasts, etc.), NextGen DSTs will move toward providing recommended actions. Achieving the dual goals of risk mitigation and reduced delay require probabilistic weather forecasts that can be used to quantify the uncertainty in the corresponding capacity predictions.

Successful integration of weather forecasts into NextGen DSTs requires coordination between meteorologists developing forecast tools and aviation researchers developing NextGen DSTs. Some GDP planning DSTs are already under development, such as the GDP Parameter Selection Model (GPSM) (Cook and Wood, 2009 and 2010). For the purposes of this study, however, a simple GDP planning model provides a clearer illustration of the importance of including temporal correlation in the forecast uncertainty model. The following section describes such a model.

3. GDP LINEAR PROGRAM FORMULATION

The key decision that must be made when a GDP is issued is the arrival acceptance rate (AAR). The AAR is the number of flights that will be assigned arrival slots in each 15-minute interval during the GDP. Flights are then matched to arrival slots using previously agreed upon procedures. Each flight is assigned ground delay so that it will arrive at its assigned arrival slot without needing any airborne holding.

For this study, we use an aggregate flight model as a stand-in for the more complex flight-specific planning used for actual GDPs. The model assigns scheduled arrivals to intervals of fixed length (e.g., 15 minutes) based on their scheduled time of arrival. For intervals 1 through T, let S_t be the number of scheduled arrivals in interval t. In practice, some of these flights might already be in the air or be otherwise unable to take a ground delay. For our purposes, however, we assume that all flights are eligible to be delayed.

The probabilistic weather forecast is represented by a set of N capacity vectors, each including a nonnegative integer capacity for each time interval. A single vector represents a feasible sequence of actual airport arrival capacities throughout the time horizon with a specified probability of occurrence. The arrival capacity in time period t under capacity vector n is represented by A_{nt}. The probability of occurrence for capacity vector n is p_n.

Flights can be assigned ground delay in order to reduce the arrival demand in a given time period based. Let y_t be the number of flights assigned ground delay in order to delay them from interval t to interval t+1. x_t is the number of remaining flights planned to arrive in interval t.

If the number of flights with planned arrival times in an interval exceeds the actual arrival capacity during this interval then the excess flights must be held in the air in order to delay them into the following interval. Note that the number of flights that must be assigned airborne holding is dependent on which capacity vector is observed to occur. Thus, let z_{nt} be the number of flights required to take airborne holding in order to be delayed out of time period t and into t+1 under capacity vector n.

Ground delay and airborne holding costs are both assumed to be linear. Define q to be the cost ratio between a single interval of airborne holding and a single interval of ground delay. Because airborne holding is always more expensive than ground holding, we assume q > 1. We do not model the cost of potential diversions.

Given the above notation, the GDP planning problem can be formulated mathematically as an integer linear programming (ILP) problem as first described by Ball, et al (2003). The goal is to balance the cost of assigned ground delays against the costs of airborne holding. Because airborne holding is subject to the uncertainty described by the probabilistic capacity inputs, airborne holding must be computed as an expected cost. The ILP formulation is given below.

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} y_t + q\sum_{t=1}^{T}\sum_{n=1}^{N} p_n z_{nt} && (1)\\
\text{subject to} \quad & x_t + y_t = S_t + y_{t-1} \quad \forall t && (2)\\
& x_t + z_{n,t-1} - A_{nt} \le z_{nt} \quad \forall n,t && (3)\\
& x_t, y_t, z_{nt} \in \mathbf{Z}^{+} \quad \forall n,t && (4)
\end{aligned}$$

Expression (1) is the cost function. The cost unit is the cost of assigning a one-interval ground delay to a single flight. The first term in the cost function is the sum of the number of flights delayed out of each arrival interval by a ground delay. The second term is the sum of the expected number of flights that must be given airborne holding in each interval weighted by cost ratio q.

Expressions (2-4) are the constraints that ensure that all variables maintain values that correctly model possible real-world outcomes. (2) ensures feasibility of the ground holding and scheduled arrival counts in each arrival interval by setting the sum of the flights assigned to the interval (x_t) and the flights delayed out of the interval (y_t) equal to the sum of the arrivals originally scheduled in the interval (S_t) and any flights delayed into the interval from the previous one (y_{t-1}). (3) similarly ensures feasibility of the airborne holding that would occur under each possible capacity outcome by setting the number of flights assigned airborne holding at the end of an interval (z_{nt}) to be at least as large as the number of flights assigned to arrive in the interval (x_t) plus any flights held in the air at the end of the previous interval ($z_{n,t-1}$) less the arrival capacity in the current interval (A_{nt}). If the arrival capacity is less than or equal to the number of assigned flights plus those held from the previous interval then this constraint will hold with equality. (4) ensures that all flight counts are nonnegative integer values.

A key observation by Ball, et al, is that this ILP formulation has a dual network structure, implying that when all capacity and demand parameters have integer values, the problem can be solved as a continuous linear program (LP) by relaxing the integrality constraints on the variables. LPs are a well-studied class of optimization problems that can be solved efficiently for even very large problem instances using pre-packaged LP solver software. For the analyses detailed in this paper, the GDP planning LP was formulated in Matlab and solved using lp_solve, a free, open-source LP/ILP solution software package that can be run from Matlab via an application programming interface (API).* Solution times for all scenarios presented were under 5 minutes on a laptop personal computer.

When solved to optimality, the outputs from the model are the number of assigned arrivals (x_t) in each interval that minimize the

* lp_solve source code, installers, and all relevant documentation can be found on the internet at http://lpsolve.sourceforge.net.

defined cost function. The corresponding values of the y_t and z_{nt} variables provide the number of arrivals in each interval assigned ground holding under the optimal solution and the resulting airborne holding at the end of each interval under each capacity scenario, respectively. In addition, the cost of any feasible arrival assignment can be computed by plugging the corresponding x_t values into the constraints, computing the resulting ground delay and airborne holding values, and plugging these into the cost function. In this way, the cost of two different policies under the same set of probabilistic capacity predictions can be compared.

In the case studies presented in section 5, 15-minute intervals are used. The ratio of airborne holding cost to ground delay cost (q) is set at 3.

4. STUDY DESIGN

The analysis of the impact of temporal correlation on GDP planning examines a series of case studies. Three primary scenarios were developed, each with multiple sub-scenarios, representing a range of realistic weather events. The scenarios are not modeled on specific historical events but are instead designed to illustrate various benefit mechanisms.

In order to isolate and quantify the benefit of modeling temporal correlations in forecast uncertainty using the model presented in the previous section, we make two key assumptions. The first assumption is that the translation from a probabilistic weather forecast to a probabilistic airport capacity prediction has already occurred. In practice, this is a nontrivial step and a potential source of significant error for an actual GDP planning DST. For our purposes, however, it is enough to assume that this translation has been accurately performed.

The second assumption is that the probabilistic capacity predictions perfectly quantify the true uncertainty. Taken together with the previous assumption, this implies that the probabilistic capacity predictions can be treated as a truth model. That is, the set of capacity vectors is assumed to include all possible capacity outcomes over the planning horizon, and the associated probabilities are assumed to be the true probabilities of observing each capacity scenario given the current weather forecast.

From a given set of probabilistic capacity scenarios representing the truth model, the corresponding temporally independent capacity model is generated by computing independent probability distributions for the capacity in each interval. The distribution for an interval t is created by setting the probability of observing a given capacity k to the sum of the probabilities of all capacity scenarios in which the capacity is k in interval t. In other words, for random variable A_t representing the capacity in time period t, set:

$$P[A_t = k] = \sum_{n=1}^{N} \mathbf{1}_{[A_{nt}=k]} p_n \quad \forall t, k,$$

where $\mathbf{1}_{[]}$ is the indicator function that takes the value 1 when the expression in brackets is true and 0 otherwise.

Under the assumption of temporal independence, the probability of observing a given sequence of capacities across time periods is computed as the product of observing the individual time period capacities:

$$P[(A_1, A_2, \ldots, A_T) = (k_1, k_2, \ldots, k_T)] = \prod_{t=1}^{T} P[A_t = k_t].$$

The space of feasible capacity scenarios under the assumption of independence is then generated by enumerating all combinations of feasible capacities across time periods. Note that every capacity scenario from the original input probabilistic capacity prediction will be included in the temporally independent prediction, and thus the state space for the temporally independent version of the model will be at least as large (and typically much larger) than the original state space.

Each scenario studied is modeled twice. First, the LP instance for the original set of capacity scenarios (the truth model) is solved to optimality. The resulting ground delay and expected airborne holding costs are recorded. A second LP instance is created by generating the corresponding temporally independent scenarios. This model is solved to optimality to determine the policy, in terms of the number of flights assigned to each interval, that would be computed when the temporal correlation is ignored. This policy is then plugged into the truth model to compute ground delay, expected airborne holding, and total cost. The difference in performance between the two policies represents the impact of including or excluding the impact of temporal correlation in the capacity inputs.

5. SIMULATED SCENARIOS

The following subsections describe the design and analyze the results of each GDP scenario.

5.1 Scenario 1: Low Ceilings at San Francisco

This scenario is based on a type of low ceiling event that occurs regularly at San Francisco International Airport (SFO). Controllers at SFO typically use parallel arrival runways for simultaneous approaches at a rate of 60 arrivals per hour. However, when ceilings drop below specified thresholds, arrivals must be condensed into a single stream, which drops the arrival rate to 30 flights per hour.

In many cases, SFO begins the day operating under low ceilings. If ceilings are not expected to lift or burn off by the morning arrival push then a GDP is necessary to keep arrival rates below the 30 flights per hour threshold. This type of event is described in greater detail in Cook and Wood (2009).

Given a particular ceilings clear time, the translation to capacity is fairly straightforward in this scenario: 30 flights per hour before ceilings clear and 60 flights per hour afterward. The temporal correlation lies in the knowledge that capacities will not bounce around from time period to time period; once capacity has increased to 60 flights per hour, it will not decrease again. When temporal correlation is ignored, this will no longer be the case. Some capacity scenarios will include inconsistent capacity sequences that bounce between 30 and 60 flights per hour.

Three probability distributions for the ceilings clear time were used to generate three sets of capacity scenarios. These distributions are shown in Fig. 1. The blue distribution (forecast 1) is a unimodal distribution that might be built around a single clearing time forecast. The green distribution (forecast 2) is a bimodal distribution which might be generated from two distinct

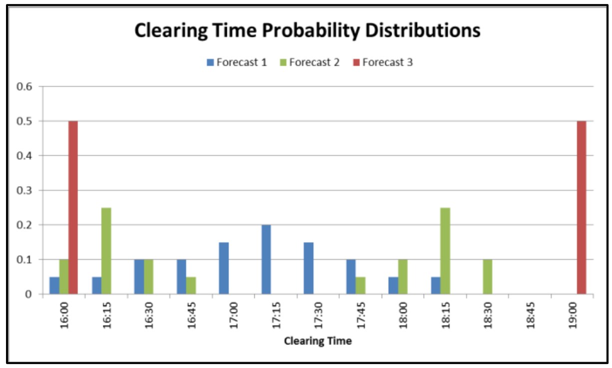

Figure 1. Clearing time probabiliy distributions with time (GMT) on the horizontal axis and probability of clearing on the vertical axis. Distributions are unimodal (blue), bimodal (green), and extreme bimodal (red).

clearing time forecasts. The red distribution (forecast 3) takes this bimodal distribution to its extreme by only allowing for two possible clearing times 3 hours apart from one another.

Two traffic scenarios were used. The low traffic scenario was based on data from October 1, 2011, and represents a typical 2011 traffic day. The high traffic scenario increases traffic in each interval by 20% and corresponds roughly with 2012 traffic levels during this same time period.

Tables 1 and 2 show the results for the low and high traffic scenarios, respectively. For each forecast, two columns are shown. The column labeled “Optimal” provides results for the policy determined by the first optimization that uses the true capacity scenarios. The column labeled “Independent” provides results for second optimization that uses temporally independent forecasts. The last row in each table provides the optimality gap, computed as the percent increase in cost under the assumption of temporal independence relative to the cost of the true optimal policy.

Table 1. Delay and holding comparions for low traffic scenario.

	Forecast 1		Forecast 2		Forecast 3	
Model	Optimal	Independent	Optimal	Independent	Optimal	Independent
Ground Delay	160	129	268	89	375	57
Expected Airborne Holding	12	25.3	4	87.5	0	159
Cost Function	**196**	**205**	**280**	**352**	**375**	**534**
Optimality Gap	**-**	**4.6%**	**-**	**26%**	**-**	**42%**

Table 2. Delay and holding comparions for high traffic scenario.

	Forecast 1		Forecast 2		Forecast 3	
Policy	Optimal	Independent	Optimal	Independent	Optimal	Independent
Ground Delay	444	344	642	346	851	374
Expected Airborne Holding	23.0	66.4	7.3	144.3	0	238.5
Cost	***513***	***543***	***664***	***779***	***851***	***1,090***
Optimality Gap	***-***	***5.8%***	***-***	***17%***	***-***	***28%***

The optimality gaps range from 4.6% to 42%. The unit of the cost function is one flight-interval of ground delay, or roughly 15 minutes of ground delay. Thus, for example, the absolute difference in cost of 72 between the optimal and independent policies under forecast 2 for the low traffic scenario corresponds to the cost of over 1000 minutes of ground delay.

In all cases, the independent policy issues a smaller amount of ground delay and, as a result, significantly higher than expected airborne holding. This implies that the programs are more aggressive, which is confirmed by the cumulative planned arrivals plotted in Fig. 2. The horizontal axis is scenario time. The blue line plots the cumulative scheduled arrivals over time. The red line plots the cumulative planned arrivals after issuing ground delays (but before computing airborne holding) under the optimal policy. The green line plots the cumulative planned arrivals when the policy is computed under the assumption of temporal independence. Note that the cumulative planned arrivals under the assumption of temporal independence is often higher (and never lower) than the optimal planned arrivals. This shows that under the assumption of independence, flights are being sent at a higher rate than in the optimal policy.

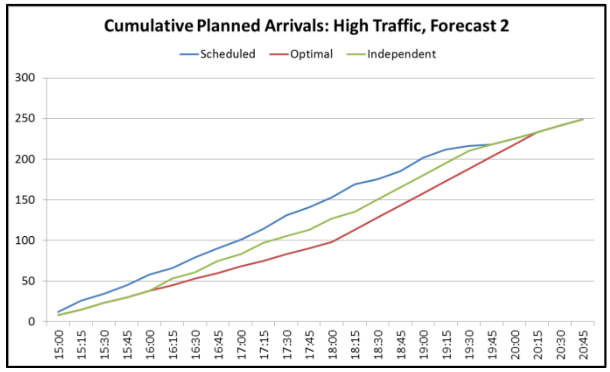

Figure 2. Cumulative planned arrivals for high traffic scenario under forecast 2: scheduled (blue), optimal (red), and independent (green).

The decrease in ground delay in each case is outweighed by the corresponding increase in expected airborne holding under the assumed cost ratio of 1:3.

This bias can be intuitively explained. Under the true capacity scenarios, capacity is always low up until the clearing time, and thus there are high-probability capacity scenarios in which the cumulative capacity is very low relative to the maximum. When temporal correlations are ignored, there will be many scenarios with capacities fluctuating between low and high values. These scenarios will have cumulative capacities that are neither among the highest nor the lowest scenarios and will have a high aggregate probability of occurrence. As a result, an automated DST will see lower risk of capacity staying low for a long period of time and will recommend releasing flights earlier than under the optimal policy. The result is the high amount of airborne holding observed.

5.2 Scenario 2: Short-duration Event of Uncertain Timing

The second case study uses the example of a frontal passage that results in a short-term drop in arrival capacity. The drop may, for example, be due to a line of thunderstorms that cause a temporary arrival stoppage or to a significant change in wind direction that requires a runway configuration change that results in a temporary arrival rate of zero. Regardless, the pattern is one of operating at maximum capacity (60 flights per hour) followed by a 30-minute time

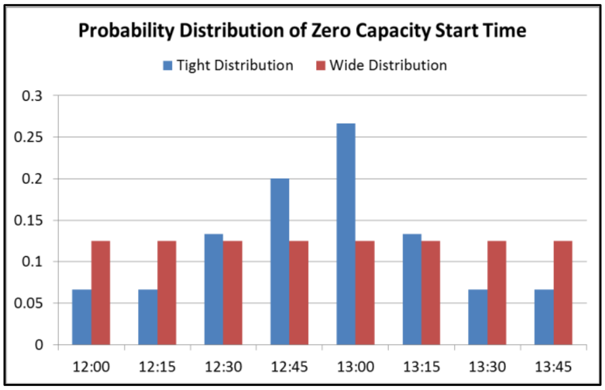

Figure 3. Probability distribution for zero-capacity start time: tight distribution (blue) and wide distribution (red).

period with no arrivals and a return to maximum capacity. The timing of this arrival stoppage, however, is uncertain.

Two probability distributions for the timing of the drop in capacity were investigated. Fig. 3 shows the two distributions. The tight distribution (blue) is predicts with high probability that the event will occur between 12:30 and 13:15 with lesser probabilities going 30 minutes further in each direction. The wide distribution (red) has equal weight on the drop occurring at each 15-minute interval between 12:00 and 13:45. Scheduled traffic was assumed to be at near capacity, implying that recovery from the drop in capacity would take a relatively long period of time.

The results for these two distributions are summarized in Table 3. The optimality gaps for the tight and wide forecast distributions are 9.7% and 14%, respectively, corresponding to absolute cost differences equivalent to 480 and 945 minutes of ground delay. In contrast to the SFO scenarios in the previous subsection, the independent policies are more conservative – issuing more ground delay and incurring less expected holding – than the optimal policies. And in this case the savings in airborne holding are not enough to make up for the large increase in ground delay.

Table 3. Delay and holding comparisons for short-duration event scenarios.

	Tight Forecast		Wide Forecast	
Policy	Optimal	Independent	Optimal	Independent
Ground Delay	203	289	203	277
Expected Airborne Holding	77.1	62.4	81.6	78
Cost	*434*	*476*	*448*	*511*
Optimality Gap	-	*9.7%*	-	*14%*

As in the previous case, this can be explained intuitively. Under the actual capacities, there is guaranteed to be only a 30-minute period at a zero capacity. Thus the cumulative capacity available cannot be less than the maximum capacity less 30 minutes of arrival capacity. When temporal dependence is ignored, however, there is the possibility of being at zero capacity for more than 30 minutes. Since each interval is evaluated independently, scenarios exist with a zero rate for as long as 135 minutes. Thus there is some probability that the cumulative capacity is very low. The result is that the policy computed under the assumption of independence releases arrivals more slowly in order to hedge against the risk of the low-capacity scenarios – even though such scenarios do not exist in the true capacity state space.

5.3 Scenario 3: Long-Duration Event of Uncertain Occurrence

The final scenario investigated is one of a long-duration event that may or may not occur. This could represent, for example, a blizzard or large thunderstorm that may or may not develop or that might hit or miss the airport of interest. In this case, if the event occurs, it is expected to reduce the airport capacity by 50% for a period of 2 hours. If the event does not occur then airport capacity is not impacted. Three probabilities of occurrence are modeled: 10%, 50%, and 75%. Traffic identical to that in the previous case study was used.

The results for these three scenarios are shown in Table 4. Under a 10% probability of occurrence, neither policy issues a significant amount of ground delay, but the optimal policy achieves nearly the same level of expected airborne holding with half of the issued ground delay. The optimality gap is 5.6%, representing the cost equivalent of approximately 160 minutes of ground delay.

Table 4. Delay and holding comparisons for long-duration event scenarios. Column header is probability of occurrence.

	10%		50%		75%	
Model	Optimal	Independent	Optimal	Independent	Optimal	Independent
Ground Delay	17	32	582	248	582	453
Expected Airborne Holding	56.5	55	0	167	0	96.75
Cost Function	**186.5**	**197**	**582**	**749**	**582**	**743.25**
Optimality Gap	**-**	**5.6%**	**-**	**29%**	**-**	**28%**

The cases with 50% and 75% probability of occurrence are instructive. In both cases, the optimal policy is to hedge completely against the possibility that the event occurs by issuing enough ground delay to ensure that no airborne holding is required. When temporal dependence is ignored, however, there is a high probability that there will be mixed high- and low-capacity intervals, so the cumulative capacity is likely to be somewhere between the maximum and the minimum. In fact, under the assumption of temporal independence, it is impossible to distinguish between a long-duration event of uncertain occurrence and a short-duration event of uncertain timing (as described in the previous section), even though the optimal policies in each case are significantly different. The result is that the policy under temporal independence releases some arrivals early, which creates a high likelihood of airborne holding. The optimality gap in both the 50% and 75% case is nearly 30%.

6. DISCUSSION AND CONCLUSIONS

The case studies presented in the previous section show that, under a wide range of conditions, assuming temporal independence in a weather forecast – or, more accurately, in the translation of a weather forecast into a probabilistic capacity prediction – can significantly reduce the efficiency of a strategic ATM planning tool. This is due to the cumulative nature of strategic ATM initiatives. Capacity reductions over a long period of time must be handled differently than those over a short period of time. Particularly in high-traffic environments, it often takes a long time to recover from lost capacity, and unexpected airborne holding often can cascade through time and impact flights many hours after they first occur.

In present day practice, this is typically not an issue. Human decision makers are looking at weather forecast tools and consulting with meteorologists to build a comprehensive understanding of the weather situation before they take action. Meteorologists in particular are able to describe the risks to these decision makers in ways that make clear what is likely to happen over time. Is this a long-duration or short-duration event? Are there meaningful thresholds – e.g., a change in wind direction – that can be planned against? How likely are various scenarios? As a result, human ATM experts are able to make decisions that appropriately account for risks.

In the NextGen operational environment, however, DSTs will be required to aggregate traffic data and automated weather forecast or capacity prediction data in order to recommend ATM actions that can improve system efficiency. An automated DST requires forecast uncertainty to be quantified in a meaningful way. This includes, in particular, a quantification of temporal correlation in the uncertainty around capacity predictions. Thus there is significant need for coordination between designers of ATM DSTs and researchers that are developing new forecast products or improving existing products in order to determine how this requirements gap can be addressed.

The impact of such a DST when this modeling gap is successfully bridged can be significant. The GPSM tool discussed previously focuses on the problem of planning GDPs at SFO under low ceilings – similar to the case studies presented in section 5.1. The forecast product used to generate capacity estimates focused on the threshold event of low ceiling clearing time. Thus capacity predictions could be easily generated that included the temporal dependence of low capacity before clearing and high capacity afterward. An operational evaluation of the tool over 5 months in 2012 along with a larger set of simulated results suggested that ground delays could be decreased by approximately 20% per GDP compared with present day operations with negligible increase in airborne holding if the GDPs recommended by the tool were issued (Cook and Provan, 2013).

The question of how to provide forecasts with temporal correlation is nontrivial. Two possible approaches are threshold events and ensemble capacity scenarios. Threshold events are changes in weather conditions that can be directly translated into changes in capacity or operations. GPSM operates using ceiling thresholds that allow SFO to move from a rate of 30 arrivals per hour to 60 arrivals per hour. Other thresholds might be changes in wind direction or wind speed that alter the viability of a runway for arrival operations. By forecasting the timing of threshold events instead of using a block forecast for each hour, the temporal correlation can be modeled by predicting capacity before or after the event. There may still be uncertainty in the capacity of the constrained resource before or after the event, but these uncertainties may be primarily operational instead of meteorological. Thresholds can be difficult to identify precisely, however, and they often involve more than just a single weather dimension. This can make forecasting the timing of these events more difficult.

Ensemble capacity prediction is an approach that would use the individual component forecasts that are typically aggregated into a single component forecast. For physical forecast models, the component forecasts often consist of many possible evolutions of weather conditions based on a range of algorithms or starting conditions. The range of outcomes represents a reasonable model of uncertainty in the forecast. Currently, the component models are most often aggregated before being sent to operational users. If instead a capacity prediction model could work directly with each component forecast then the range of predicted capacity scenarios associated with the components should represent a reasonable model of the range of possible capacity outcomes. This would appear to be the most promising approach. One key difficulty, however, would be the need either to distribute the component model outputs, which would be a very large amount of data, to a large number of users or to define a common set of capacity models that can be applied to the component forecasts as they are computed. In the latter case, it would then be these capacity predictions, along with the aggregate forecast data, that would be distributed to users.

Regardless of the approach taken, meteorologists and DST designers will need to collaborate with each other in order to better understand the requirements on both sides and to define responsibilities for the various steps involved in translating advanced weather forecasts into capacity predictions with accurate models of uncertainty. Achieving the NextGen ATM goals and efficiencies will require greater integration of strategic planning, research, and development efforts between these two communities.

ACKNOWLEDGEMENTS

This work was funded by NASA Ames Research Center (ARC) under contract number NNA11AA18C. The authors would like to thank Shon Grabbe and William Chan at NASA ARC for their support of this effort.

REFERENCES

Ball, M. O., R. Hoffman, A. R. Odoni, and R. Rifkin, "A stochastic integer program with dual network structure and its application to the ground-holding problem," *Operations Research*, 2003, vol. 51, 167-171.

Cook, A. J., G. Tanner, and S. Anderson, "Evaluating the true cost to airlines of one minute of airborne or ground delay: final report," Eurocontrol Project Report, Brussels, Belgium, 2004.

Cook, L., and B. Wood, "A model for determining ground delay program parameters using a probabilistic forecast of stratus clearing," *8th USA/Europe Air Traffic Management (ATM) Research and Development Seminar*, Napa, CA, June 2009.

Cook, L., and B. Wood, "A model for determining ground delay program parameters using a probabilistic forecast of stratus clearing," *Air Traffic Quarterly*, 2010, vol. 18, no. 1, 85-108.

Cook, L. S., and C. Provan, "An operational evaluation of a ground delay program (GDP) decision support tool using an integrated forecast of stratus clearing at San Francisco Airport," *AMS 16th Aviation, Range, and Aerospace Meteorology (ARAM) Conference*, Austin, TX, January 2013.

Joint Economic Committee Majority Staff Report, "Your flight has been delayed: flight delays cost passengers, airlines, and the US economy billions," May 2008.

Joint Planning and Development Office, " Concept of operations for the Next Generation Air Transportation System, version 3.2," September 2010.

Mukherjee, A., M. Hansen, and S. Grabbe, "Ground delay program planning under uncertainty in airport capacity," *AIAA Guidance, Navigation, and Control (GNC) Conference*, Chicago, IL, August 2009.

Provan, C. and S. Atkins, "Optimization models for strategic runway configuration management under weather uncertainty," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Dallas, TX, September 2010.

Weather-ATM Integration Working Group Report, "Research, engineering, and development advisory committee," October 2007.