9.7

A "DEMAND PULL" APPROACH TO SHORT TERM FORECAST DEVELOPMENT AND TESTING[†]

James E. Evans*
MIT Lincoln Laboratory

1. INTRODUCTION

Product development can be motivated by "demand pull" from user needs or by "technology push" from research.

Historically, new aviation weather forecasts have been motivated by "technology push" considerations, in which better scientific validation results were the principal criterion for the forecast "goodness". For example, one might demonstrate that the forecast had a higher probability of detection (Pd) and/or lower false alarm probability (Pfa) than an alternative forecast. There were no specific numerical criteria for forecast performance that would warrant operational use of the new forecast in place of the preexisting ones, except for a generic requirement for "improved" forecasting of a given phenomenon.

However, in an era of significant government and airline budget austerity for civil aviation investments, it is becoming increasingly important to quantitatively demonstrate the benefits to the operational user community of improved aviation weather forecasts[1].

If demonstrating benefits is very important, then we propose that there should be a greater element of "demand pull" involved in the product development wherein there is:

1. a very specific set of users seeking improved performance for certain relatively well understood weather situations, and
2. reason to believe that if the forecast performance could be improved by currently exploitable scientific knowledge, measurable benefits would in fact be achieved.

The "demand pull" referred to in element (1) has been frequently addressed by "rapid prototype" development processes where a group of aviation weather forecast developers work in an iterative fashion with a group of users. The important difference in what we recommend is the addition of element (2), where one has identified users and a forecast usage situation in which end user benefits can be demonstrated.

If such a benefits demonstration is in fact going to be an integral element of the overall short term aviation forecast development process, one needs to give some thought as to how the forecast development and testing will proceed.

In the remainder of this paper, we consider the important aspects of short term forecast development and testing that are key to a successful "user benefits driven" demonstration:

First, what is the overall decision process for the effective use of the short term forecast if it is to have the desired quantifiable user benefit? As was noted by (Ballentine, 1994) "It has been said that a forecast has benefit, positive or negative, only if it changes a decision."

For example, in the case of products intended to reduce aviation weather delays, one needs to understand carefully the overall decision process that is involved in the user taking actions that will result in a reduction of delays. Also, if benefits are to be achieved, one must:

1. identify the important users, and
2. develop training for these users that is oriented towards achieving measurable benefits.

We discuss two specific short term aviation weather forecasts – convection and ceiling – to illustrate the issues that arise in thinking about the overall decision support system, key users, and training needed to generate benefits. We also consider reducing weather-related fatal accidents.

Second, what is the preexisting "baseline" of aviation forecasts/decision processes that already exists to address the user needs? In most cases, there are already various weather information sources that can be viewed as providing a short

---

[†] This work was sponsored by the Federal Aviation Administration under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

[*] *Corresponding author address*: James E. Evans, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420-9185; e-mail: jime@ll.mit.edu

[1] In some cases, it may even be important to demonstrate that benefits are likely to be achieved before the forecast development will be funded.

term forecast (e.g., a Center Weather Service Unit (CWSU) meteorologist, persistence, or animation loops of the past weather). How well do we understand how the “baseline” forecast and the associated user decision support system operate? How will the new forecast and its decision support compare? What are the training implications if the new forecast is rather different than the “baseline”?

Third, how will we measure the change in system performance? For example, if the new forecast claims to help reduce delays and/or accidents, how will one address differences in the weather between the “before” and “after” time periods? How will one determine whether the new forecast is in fact the key factor, if there was a change?

The paper concludes with some suggestions for development and testing of new aviation forecasts to improve safety and reduce delays.

2. IMPACT OF THE OVERALL DECISION MAKING PROCESS ON FORECAST DEVELOPMENT

In Figure 1, we show the major elements of decision making with short term forecasts. The left hand side of the diagram represents “classical” aviation weather forecasting, while the right hand side indicates that additional, non meteorological decision support tools (DSTs) may also be involved in the overall decision process.

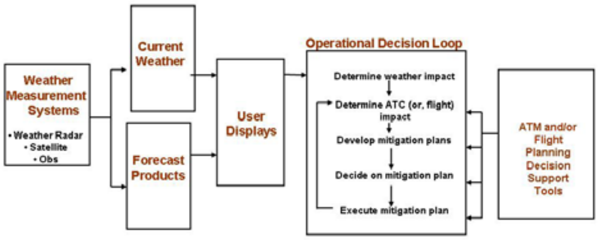

Figure 1 Decision process for use of short term forecasts for air traffic management and/or flight planning. By their nature, short term forecasts are generally most useful when the weather impacts may change fairly rapidly. In such cases, it is very important that the operational decision loop be executed in a timely and effective manner if the desired result -- mitigation of adverse aviation weather impacts -- is to be achieved.

Due to ever increasing congestion at major airports and in en route airspace, achieving decisions in a timely manner such that delays are reduced is particularly difficult in the US national airspace system (NAS). As indicated in Figure 2, the NAS now must be thought of as a highly congested network in which network disturbances arising from a loss of capacity due to adverse aviation weather rapidly propagate throughout the network.

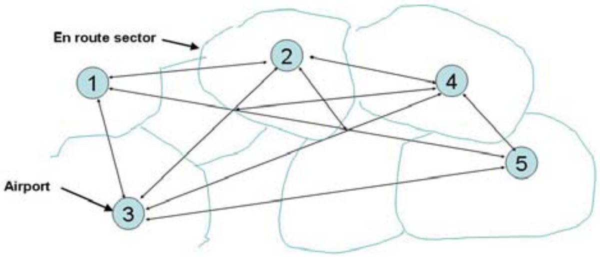

Figure 2. The NAS as a network. When capacity is lost (e.g., due to en route convective weather), rerouting is essential, but very difficult since many of the links are already near capacity in fair weather. Unanticipated losses of terminal arrival capacity can also cause widespread network disruptions if aircraft in holding patterns near the terminal block other “over flight” aircraft.

A consequence of the highly coupled nature of the NAS is the need for extensive collaboration and coordination in addressing problems with adverse weather. In Figure 3, we show the results of an analysis of inter-facility coordination issues by (Davison and Hansman, 2001).

Figure 3 highlights both the number and diversity of users that might be associated with the operational decision loop shown in Figure 1. Both sources of coordination complexity have significant implications for the design of forecasts, the design of the system to provide the forecasts to users, and the user training.

2.1 Understanding the Benefits Generating Users: the CIWS Experience

The value of identifying all of the key decision makers was highlighted in the recently conducted Corridor Integrated Weather System (CIWS) ATC workload study (Robinson and Evans, this conference). The initial CIWS deployment (see Robinson, et. al., 2004) had focused principally on providing displays to the TMCs at the ATCSCC, ARTCCs and a number of TRACONs. However, at the Washington ARTCC (ZDC), CIWS displays were also provided to the area managers who supervise the controllers in an enroute sector.

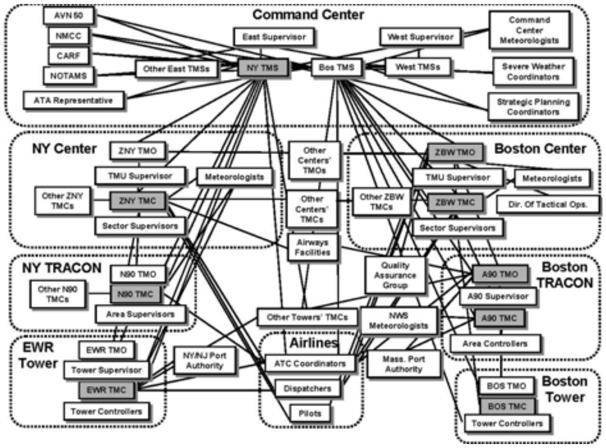

Figure 3. Interactions between various FAA facilities and airlines in addressing congestion problems related to the Newark International Airport (EWR) (from Davison and Hansman, 2001). The traffic management coordinators (TMCs) play a key role in addressing NAS network problems, but they must coordinate with many other potential aviation weather forecast users. Note that airline dispatchers are an important component of the coordination process. This is because rerouting and other adjustments to filed flight plans may be necessary to address the combination of weather and congestion problems.

One of the principal benefits metrics in the initial CIWS operational benefits study (table 7-3 in Robinson, et. al., 2004) was the number of times capacity enhancing decisions were made per convective weather impact day in an ARTCC. It was found that ZDC had a substantially higher frequency per thunderstorm day for two key decisions:

- Keeping routes open longer/reopening closed routes earlier
- Proactive reroutes of aircraft

However, it was not possible to determine whether the high frequency of capacity enhancing decisions at ZDC reflected the CIWS availability at area manager positions as opposed to some other ZDC unique factor (e.g., TMU receptiveness to the CIWS products).

In 2005, CIWS displays were installed in 4 of the 8 areas at the Cleveland ARTCC (ZOB) and real time benefits observations, similar to those conducted in 2003, were repeated. These additional observations specifically compared how well the various CIWS forecast products were used when convective weather impacts occurred in a ZOB area in which the area manager had access to the CIWS products versus a situation where the convective weather occurred in a ZOB area without access to the CIWS products.

The results (Robinson and Evans, 2006) clearly demonstrated the significant benefits of providing the products to the area managers: the overall use of CIWS to identify and execute capacity-increasing opportunities at ZOB increased about 66 % for each of the two benefits decisions noted above, and the overall frequency of capacity enhancing decisions at ZOB was now essentially identical to that of ZDC[2].

2.2 Understanding the Benefits Generating Users: the ITWS Experience

A similar impact of “non traditional” users was also noted in testing of the Integrated Terminal Weather System (ITWS). Important elements of the ITWS (specifically, a short term convective forecast) had grown out of feedback from users of the Terminal Doppler Weather Radar (TDWR) test system in Orlando, FL. A relatively complete set of initial ITWS products, created using data fusion of real time inputs from multiple sensors, began to be used operationally in 1993 and were formally evaluated in real time by air traffic and airline users at Memphis and Orlando in 1994 (Evans and Ducot, 1994). The earlier TDWR products had gone only to towers and TRACONs. In the ITWS demonstration, it was suggested by local ATC personnel that the ITWS products be provided to the associated ARTCCs, as well as to the TRACONS and towers.

In the benefits assessment, it was found that improved decisions by the en route traffic management units were a very significant factor in reducing delays. Also, it was learned that many of the convective delay problems attributed to the airports in fact arose from weather in the transitional en route airspace surrounding the terminal area, as opposed to arising from runway impacts.

---

[2] Additionally, in 2005 CIWS provided a new echo tops forecast product that was not previously available. There was an increase in the frequency of capacity enhancing decisions at ZDC, which presumably reflected the additional benefit of the echo tops forecast. However, the % increase in the frequency of capacity enhancing decisions at ZOB was much greater than the % increase at ZDC. Overall, it appears that providing the CIWS displays to the half of the ZOB area managers alone increased the frequency of capacity enhancing decisions by about 66%.

2.3 *Optimizing the Forecasts to Meet the Operational User Needs if Benefits are to be Achieved: the CIWS Experience*

One of the very important side benefits of the 2003 CIWS operational benefits assessment (Robinson, et.al., 2004) was the identification of which CIWS products were used to make various capacity-enhancing decisions. This analysis of the relative merits of the products showed that the high resolution echo tops product (which had no corresponding forecast) was almost as important as the 0-2 hour precipitation forecast in operational user decision making.

This lead to the development and testing of an echo tops forecast, which has proven beneficial in further increasing the frequency of capacity-enhancing decisions in the CIWS domain (Dupree, et. al., 2006).

2.4 *Optimizing the Forecasts to Meet the Operational User Needs if Benefits are to be Achieved: the San Francisco Ceiling/Visibility Probability Forecast Experience*

The local airspace surrounding the San Francisco International Airport (SFO) is prone to regular occurrences of low ceiling conditions from May through October due to the intrusion of marine stratus along the Pacific coast. The low cloud conditions prohibit dual parallel landings of aircraft on the airport's closely spaced parallel runways, thus effectively reducing the arrival capacity by a factor of two. The behavior of marine stratus evolves on a daily cycle, filling the San Francisco Bay region overnight, and dissipating during the morning. Often the low ceiling conditions persist throughout the morning hours.

The FAA will put a ground delay program (GDP) into effect under these conditions, since the scheduled demand into SFO from mid morning to early afternoon typically exceeds the SFO low ceiling arrival capacity of approximately 30 aircraft per hour. These GDPs delay departures at their origin such that the arrival flow for SFO[3] matches the airport capacity. The result of the GDP is a substantial number of delayed flights into the airport and a negative impact on the National Air Space (NAS).

Such GDPs are typically put in effect for many more hours than the climatological average marine stratus dissipation time. As a consequence, when the marine stratus dissipates, the actual rate of aircraft arrivals at the airport (about 30 per hour) is much less than the effective SFO arrival capacity at that time (60 aircraft per hour).

Depending on the distance from SFO of the delayed flights, it can take one or more hours before the landing rate at SFO is comparable to the airport capacity. Hence, unnecessary delay is incurred due to the inability to match the supply of arrivals to the airport capacity.

Benefits studies (Wilson and Clark, 1997) have shown that there is a significant delay reduction benefit from proactively ending the GDPs using a forecast.

The marine stratus stratus forecast (Ivaldi, et. al., 2006; Clark, 2002) at SFO has been a success meteorologically:

1. the forecast of most likely marine stratus dissipation time outperforms climatology on average by:

   (a) about 12% for the pre-dawn forecast (e.g., 13Z), and
   (b) about 35% for forecasts issued during the morning hours (e.g., 15Z)

   based on data from 2003-2005. During 2005, high confidence morning forecasts of the dissipation time provided a 53% improvement over climatology, and

2. the automated objective probabilistic forecasts of clearing by key operational target times, namely 17, 18, 19, and 20Z, have been shown to be statistically reliable.

To illustrate the accuracies achieved: the SFO system produced 136 forecasts in the three-year period 2003-2005 with a 90% or greater probability of clearing before 17Z or 18Z that verified 94% of the time. Of the 8 forecasts that did not verify, 7 had an offset time of less than 30 minutes.

(Ivaldi, et. al., 2006) summarize the possible operational user actions based on the forecasts and operational consequences as follows:

“There are several ways in which the SFO forecast could influence traffic flow decision making for SFO. The first is to avoid a GDP if ceilings and visibilities are forecast to improve prior to arrival rates exceeding acceptance rates. Second would be to cancel a GDP proactively,

---

[3] For example, the weekday arrival demand at SFO in the summer of 2005 was typically 55 airplanes per hour between 9 am and 10 am local time (16Z to 17Z).

once initiated, if confidence was high that clearing would occur prior to the arrival rate exceeding the acceptance rate. A third possibility is to maintain the GDP, but gradually increase the acceptance rate at some agreed upon time prior to clearing, based on the confidence of the forecast.

Each of these decisions carries with it a level of risk. Obviously the first option carries the greatest risk but also the greatest potential benefit to the NAS and the traveler. However if the forecast is wrong, the Oakland Center will be dealing with vectoring many aircraft into a holding pattern and most likely invoking a ground stop. The second option carries with it a reduced benefit, as well as a reduced risk, as fewer aircraft would be in the air to manage. The third option carries with it even less risk, but also reduced benefit, as it is dependent on the rate at which the acceptance rate is increased prior to clearing."

There have been very few events in which a GDP was cancelled proactively The current FAA policy is to add two hours to the burn off time to arrive at a GDP cancellation time[4]. Since the vast majority of stratus events dissipate well before 2 hours after the projected burn off time, most of the projected benefit from the forecast is not being achieved.

In cases where there is a (subjective) "high confidence" that burn off will occur at a given time (based on discussions between the Oakland CWSU, the Monterey NWS and United Airlines meteorology[5]), an intermediate (e.g., 45 per hour) arrival rate is used for the last two hours of the GDP. This partially reduces the number of landing slots that were not utilized, but still leaves a significant "avoidable" delay.

We have identified three key problems in the operational utilization of what appears to be a technically very successful probabilistic forecast:

1. the ARTCC operational users are very concerned about the possibility of too many aircraft holding in the Oakland ARTCC airspace,
2. the traffic flow management unit personnel do not have academic training or practical experience at using probabilities for decision making, and
3. important forecast information that would be needed to apply standard techniques for decision making under uncertainty were not being provided to the users in the current forecast.

Making decisions using well-defined probability forecasts (that is probabilities that can be manipulated by the standard rules for probability use) involves the application of statistical decision theory which is a relatively well understood area conceptually.

The essential components (see, e.g., Chernoff and Moses, 1959) are:

- the available actions (e.g., GDP parameters),
- the possible states of nature (the marine stratus dissipation times),
- the consequence of actions for a given action that is taken when nature has some state (e.g., amount of delay, the number of aircraft in a holding pattern, etc),
- the probability of the various possible states of nature, given some measurements (these probabilities for various states would be generated by the SFO forecast algorithm), and
- the strategy used to choose between the actions, given the forecast probabilities.

It should be noted that there is extensive literature on optimizing GDP parameters, given a probabilistic forecast of the future capacity [(Mukherjee and Hansen, 2005) show contemporary results as well as providing references to the past literature)]. These studies did not explicitly consider the cost to air traffic personnel from too many aircraft in a holding pattern (e.g., if the GDP was ended proactively in error). In addition, they generally assume that the costs and benefits could be expressed by a combined metric, such that one could optimize the GDP parameters using an expected loss criteria.

If one sets about to convert the SFO marine stratus forecast information into information more directly tailored to the consequence of a given action that is taken when nature has some state, one finds quickly that a key factor, the probability distribution of extremely late dissipation times (e.g., the cases where the dissipation time was after the 90% forecast time), was not being provided to the decision makers.

This, coupled with the difficulties in relating the forecast probabilities to trading off possible

---

[4] The general guidelines for how low ceiling/visibility traffic flow operations are conducted at SFO are provided on the following web site: http://www.fly.faa.gov/ois/west/zoa/sfo/sfo_tm.htm

[5] All of these participants have access to the automated forecast.

outcomes for various strategies, lead us to consider instead providing a decision theory-based presentation for the forecasts.

Specifically, we suggest that there needs to be a substantially different, risk management based, approach to presentation and use of the SFO probabilistic weather forecasts:

1. The operational decision makers (e.g., the FAA traffic flow managers in consultation with key airlines) need to be provided the expected consequences of various actions (i.e., GDPs), given the probability distribution of expected dissipation times. This operational consequences-oriented presentation would include key factors such as expected average delays, expected unnecessary "avoidable" delay, average holding time, and the probabilities of various numbers of aircraft (e.g., 10, 20 or 30) in airborne holding within the Oakland ARTCC, for various GDP options.
2. Much more attention needs to be paid to how to mitigate the risk of rare late stratus dissipation events that would cause an excessive number of holding aircraft. There are at least two options for such risk mitigation: improved use of the daytime forecasts (e.g., 15Z) to modify a GDP that was put into effect in the predawn period (e.g., 13Z), and developing a fair and equitable system by which SFO-bound planes in a holding pattern would be diverted to an alternative airport in the event that the number of holding aircraft exceeds an agreed upon threshold. It should be noted that the diversion option would have to be developed collaboratively with the airlines.

The hope is that the above alternative approach to decision making with the SFO probabilistic forecast can be developed and implemented under the auspices of the FAA/industry Collaborative Decision Making (CDM) program (see http://cdm.metronaviation.com/).

In summary, we have identified a need for a dramatically different presentation of the forecast results at SFO, as well as a need for explicit, collaborative risk management of some possible consequences of actions taken by users. The risk management procedure itself is likely to generate additional requirements for marine stratus forecast information (and development). Training will clearly be very important if such a major paradigm shift in the GDP decision making approach is to be successful.

The above experience in achieving operational benefits with what we would regard as a meteorologically successful, probabilistic forecast, for a situation where the consequences of various actions are fairly well understood, highlights the challenges ahead for the practical application of statistical decision theory to the use of probabilistic convective weather forecasts.

2.5 Forecasts to Improve Safety

The FAA Flight Plan (FAA, 2005) discusses reducing the air carrier and general aviation accident rates in the near term. The air carrier fatal accident performance target description does not have any "strategic activities" that are specifically related to weather. Given that fatal air carrier accidents are very rare, there is relatively little guidance that we have as to how short term forecasts should be tailored to reduce air carrier fatal accidents.

Weather is a significant factor in many of the general aviation (GA) fatal accidents. It has been shown (Office of Federal Coordinator for Meteorology, 2003) that the fatal accident rate (per flight hour) for Part 91 GA aircraft has been significantly reduced from 1995 to 2001; however, there was no analysis to determine what the roles of various factors (on board sensing, pilot training, dissemination of products, forecasts, etc.) were in reducing the accident rate.

None of the FAA Flight Plan "strategic initiatives" to reduce GA fatal accidents has an aviation weather component. However, the performance target to reduce Alaska accidents does include a strategic initiative involving the use of weather cameras and alternative techniques to provide improved information to air carriers and general aviation accidents.

We have not been able to find papers that discuss how past analyses of GA weather accidents and pilot decision making have been conducted to determine how forecasts might be explicitly tailored to change the decision making of the subset of GA users that have accidents.

The 2005 Flight Plan has a strategic initiative in human factors to "identify human factors that may cause accidents and develop strategies, methods and technologies that will reduce those accidents". However, none of the activity targets identified in the 2005 Flight Plan specifically addresses optimizing weather forecasts to improve decision making for the pilots that are most likely to have weather-related fatal accidents.

2.6 The Role of Training

If measurable benefits are to be achieved, then it is very important that the training explain to the end users how the weather forecast can be used to do a better job of making the decisions that lead to the benefits.

This is a very different orientation than the usual training, which emphasizes the theory underlying the forecasts and how the various display features work. This is not to say that one should not explain the basis for the forecasts and it is clear that the users need to be able to interact with the displays[6]. However, far too often, the training is left to a training group that may have little first hand experience with the forecast products and their operational utilization, or, training is only provided by either a computer-based instruction CD or a help page on a WWW site.

Our experience is that benefits-oriented training for the use of aviation weather forecasts is best accomplished by a two step process:

1. small group training in a classroom setting, with a display depicting past recorded weather data sets so that the trainer can show practical situations in which the forecast is used, as well as answer questions about the basis for the forecast. The availability of a display that behaves like the operational one helps develop the ability to manipulate the display (e.g., overlays, pan, zoom, stored configurations, etc) to meet the specific user needs for various operational situations. An important element of the classroom approach is the use of trainers who have personal experience in observing the real time use of the forecasts for decision making at operational facilities [that is, Subject Matter Experts (SMEs)].
2. in situ training whereby experienced trainers are present during actual weather events so that operational users can ask questions about the forecasts in the context of solving an actual problem. This latter form of training, which resembles in some respects the medical training of interns and residents in hospital wards, has probably been the most effective training for air traffic personnel who do not have an academic orientation.

Regardless of the way the operational users obtain a forecast [e.g., via Web browser displays or a full capability situation display], we have found it essential to make at least a yearly visit to the operational user facility to conduct refresher training and answer questions.

We should note that this is a very different approach that usually provided for FAA production systems (e.g. ITWS) where typically there is only a computer-based instruction course with no local SME who understands the weather forecast and its operational use.

3. ESTABLISHING A BASELINE FOR BENEFITS

One of the important elements in benefits assessment is the baseline system for comparison. To date, such baseline assessments have very rarely been carried out before an aviation weather forecast was deployed.

In the case of the ITWS, there had been experience with the use of TDWR prior to the beginning of ITWS forecast development and testing. However, because most of the products of ITWS were new products developed based on “rapid prototyping” feedback from the users, it was not clear that a baseline assessment of TDWR usage would have looked at the operational decisions that were addressed by the ITWS products.

In the case of CIWS, there was a very rich, complicated user environment with many other convective weather products and forecasts available to the users (e.g., WARP products including time lapse animation of the past weather, precipitation and the National Convective Weather Forecast (NCWF) on the ETMS display, and the information from the CWSU). No effort was made to establish a pre-CIWS baseline. In retrospect, it would have been very helpful to have made observations of operational decision making before the CIWS forecasts were deployed, both to identify the key decision makers and to anticipate important forecast features.

In the initial CIWS benefits usage assessment (Robinson, et. al., 2004), users were asked in real time what the differences would have been in making a specific decision had they not had CIWS. This was viewed as a reasonable approach at that time, especially since the viewing of the CIWS products by an operational user could

---

[6] An important element of the Lincoln Laboratory ITWS and CIWS training is laminated Quick Reference Cards (QRCs) with a compressed version of the training material that can be left near the user displays.

be confirmed by the benefits data gatherer. Additionally, this "in situ" assessment approach made it fairly straightforward to understand which CIWS products had been used to make the different decisions.

However, during presentations of the 2003 operational benefits results (Robinson, et. al., 2004) to the FAA investment analysis group, concerns were expressed about the repeatability of the benefits observations (especially, the estimation of decision making with the baseline system).

One option for estimating the baseline decision making performance would have been to turn off CIWS for a time period and observe the operations of the NAS during convective weather. This was not viewed as practically possible.

Hence, in 2005, observations of TMU decision making in convective weather were conducted at two ARTCCs [Atlanta (ZTL) and Jacksonville (ZJX)] that did not have CIWS, but which were adjacent to an ARTCC that did have CIWS (Washington). The focus of the TMU observations was on events that were close to the boundary between ARTCCs with and without CIWS. We observed that ZTL and ZJX have very different congestion constraints and styles of traffic flow management than does ZDC, so it is not clear to what extent a quantitative baseline for ZDC can be developed from the ZTL and ZJX observations. However, when CIWS is deployed to Atlanta and Jacksonville, there will be a prior baseline for comparison, albeit the problem of different convective weather events will still exist.

Determining a baseline for the forecasts used to determine SFO GDPs to address summer marine stratus events is very difficult. The forecasts generated by the various participants (the CWSU, the Monterey forecast office, and United Airlines meteorologists) prior to the operation of the automated forecast system were generally not recorded and probably were not quantitative. There was a consensus forecast discussion with the TMU that was not recorded and then a GDP decision was made in which the perceived likelihood of various outcomes was also not recorded.

Additionally, the nature of the SFO human generated forecasts and the decision process for use of those forecasts is quite different from the statistical decision theory-based approach that we have suggested above. Hence, one is not just comparing forecasts, but rather a completely different approach to making decisions.

One option for evaluation of the SFO forecast would be to compare the "unavoidable" delay at the end of a stratus event before and after the forecast was deployed. However, there are three significant baseline problems.

First, the determination of when the stratus has dissipated in terms of operational procedures is based on pilot reports as opposed to measurements from an archived weather sensor (e.g., a METAR for SFO). Unfortunately, the side-by-side times were not recorded before the automated SFO forecast was deployed.

Additionally, there have been important procedural changes made since the forecast was deployed (e.g., the multi-rate GDP). Hence, it is not clear how much of the difference in "unavoidable" delay could be attributed to the forecast versus the procedure change.

Finally, the arrival demand time profile at SFO has changed significantly during the period in which the SFO forecast was in test.

Another important metric would be the frequency and severity of situations where drastic traffic flow management actions had to be taken to avoid overloading the ARTCC controllers. Unfortunately, some of the TMU mechanisms used in such situations (e.g., miles-in-trail spacing) may not be archived in the national logs for the time periods of interest. Additionally, it might not be possible to determine how many aircraft were holding in the Oakland ARTCC airspace prior to 2001.

Another baseline-related topic is determining how much of the existing weather delay is "avoidable" and to what extent the "avoidable" delay is being (or will be) reduced by other systems. (Allan and Evans, 2005) provide estimates of the weather delay in 2005. Determining how much of that delay is "avoidable" requires an NAS model that considers the capacities of terminals, en route sectors, and the optimal allocation of flights to the available capacity. Research is underway to develop such a model; its overall structure is discussed in (Weber, et. al., 2005).

4. HOW DO WE MEASURE A CHANGE IN PERFORMANCE

We have attempted to use delays in convective events before and after CIWS and ITWS were deployed as a measure of a change in NAS performance. The most experience in studying this connection was gained through the ITWS testing at Atlanta (Allan and Evans, 2005). Our experience has been that this analysis is very difficult to accomplish for a number of reasons:

- Difficulty in finding comparable convective weather events (e.g., spatial patterns and time history of significant convective weather)
- Changes in the NAS (e.g., demand, traffic mix, operational procedures, airline scheduling and automation capabilities)
- Other forecasts (e.g., CCFP) and the decisions made using those forecasts

Our current view is that the use of quantitative delay comparisons as the basis for forecast benefit assessment will probably require a NAS model that considers the complexities of the NAS as a network that were discussed above.

Alternatively, one can instead focus on measuring the differences in beneficial decisions made by operational users before and after a system was deployed. The archives of NEXRAD level 2 products at NCDC and ETMS flight track data since about 2002 are a major help in this respect for convective weather assessment.

We are in the process of comparing time animations of flight tracks and the CIWS products before and after CIWS was deployed to see if the capacity enhancing benefits decisions discussed above were in fact being accomplished more frequently after CIWS was deployed. Initial experience with such analyses has shown that one needs automated tools to examine the time animations since:

1. the convective weather events are generally not repeatable[7], and
2. in congested airspace, there are a myriad of different traffic management decisions going on at one time.

Similar comparisons of the capacity enhancing decisions in convective weather before and after ITWS was deployed will commence shortly.

For ceiling/visibility forecasts such as at SFO, one would really like to compare the “unavoidable” delay at the end of a stratus event before and after the forecast was deployed, along with metrics that indicated the frequency and severity of cases where the supply of arrivals exceeded the effective SFO arrival capacity. Given that:

- the forecast at SFO has evolved significantly since 2001, and that
- a very different operational decision strategy might be deployed in the near future,

it might be possible consider the initial test years (e.g., 2001) of the SFO stratus dissipation forecast as a baseline for assessing the change in user decisions.

Demonstrating that a change in performance in reducing air carrier or GA fatal accidents can be attributed to a forecast appears to be a neglected area. For example, although there have been a number of safety-oriented weather forecast products developed by aviation weather researchers that are available via Internet Web sites, there is a paucity of papers describing the extent to which fatal weather accidents have been prevented through use of the forecasts.

5. SUMMARY

In this paper, we have discussed the implications of a shift in paradigm from “technology driven” aviation weather forecast development to a paradigm where being able to demonstrate operational user benefits is a principal focus.

In the preceding sections, we have discussed the identification of the operational users that are most likely to generate the benefits and the tailoring of both the forecasts and training to facilitate improved end user decision making. The experience with both ITWS and CIWS showed the importance of identifying as many of the key decision makers as possible. Additionally, we noted that the CIWS operational benefits assessment identified additional forecast capabilities that would enhance the operational utility of the system.

We view the San Francisco experience with a probability forecast that was very successful meteorologically, but has yet to demonstrate operational benefits, as very important to the aviation weather community given that a major thrust of the multi-agency Joint Development Program Office (JPDO) is to utilize probabilistic forecasts. Clearly much more applied research into the operational use of such forecasts is necessary.

A common denominator for all of the short term forecast applications discussed is that the training needs to advise the end users on how to properly apply the forecast to the operational decisions that they make on a daily basis, as well as providing information on the basis of the

---

[7] See section 6.3 in (Allan and Evans, 2005).

forecast and use of the display. When possible, this training should include face-to-face training by an SME on the operational use of the forecast so that the user can ask questions.

Establishing a baseline for the aviation forecast/operational user decision system that will be used as a comparison was shown to be challenging for both convective weather and the SFO forecasts.

Measuring the change in the frequency (and impact) of capacity-enhancing decisions seems relatively feasible for the ITWS and CIWS convective weather forecasts in the near future. However, there are many challenges to carrying out such an assessment for the SFO ceiling forecasts.

Reducing GA fatal accidents is a major objective of the FAA and weather is a factor in a large fraction of such accidents. However, our limited investigation has indicated a paucity of published papers on the operational effectiveness of the various aviation weather forecasts that have been deployed in the past few years to reduce the GA fatal accident rate.

We have suggested that there needs to be focused investigations of the GA pilots that are most like to have fatal weather related accidents to determine how the appropriate weather forecasts and training should be tailored to better improve the decision making by these pilots. Similarly, there will need to be considerable research work on identifying an appropriate baseline for comparing and normalizing the various key factors (e.g., location and severity of the weather that causes fatal GA accidents) involved in fatal GA accidents.

6. REFERENCES

Allan, S. and J. Evans, 2005: Operational Benefits of the Integrated Terminal Weather System (ITWS) at Atlanta, Massachusetts Institute of Technology, Lincoln Laboratory, Project Report ATC-320.

Ballentine, V., 1994: The use of marketing principles to maximize the economic benefits of weather, Met Apps 1, 165-172.

Chernoff, H. and J. Moses, 1959: Elementary decision theory, Dover.

Clark, David, 2002: The 2001 Demonstration of automated cloud forecast guidance products for San Francisco International Airport, Portland, OR, 10th Conference on Aviation, Range, and Aerospace Meteorology, American Meteorological Society.

Davison, H and R.J. Hansman, 2001: Identification of inter-facility communication and coordination issues in the U. S. Air Traffic Control System, MIT International Center for Air Transportation Paper 2001-11-21 (available at http://icat-server.mit.edu/Library/).

Dupree, W, M. Robinson, R. DeLaura and P. Bieringer, 2006: Echo top forecast generation and evaluation of air traffic flow management needs in the national airspace system, 12th Conference on Aviation, Range, and Aerospace Meteorology, Atlanta, GA. Amer. Meteor. Soc.

Evans, J.E., and E.R. Ducot, 1994: The Integrated Terminal Weather System (ITWS) Massachusetts Institute of Technology, Lincoln Laboratory Journal, Vol. 7, No. 2, p. 449.

Evans, J., M. Robinson and S. Allan, 2005: Quantifying convective delay reduction benefits for Weather/ATM systems, 6th USA/Europe Seminar on Air Traffic Management Research and Development, ATM-2005, http://atmseminar.eurocontrol.fr/, Baltimore, MD.

Federal Aviation Administration, 2005: Air traffic organization fiscal year 2005 business plan.

Ivaldi, C., D.A. Clark, and D. Reynolds, 2006: Upgrade and technology transfer of the San Francisco Marine Stratus Forecast System to the National Weather Service, 12th Conference on Aviation, Range, and Aerospace Meteorology, Atlanta, GA. Amer. Meteor. Soc.

Mukherjee, Avijit and Mark Hansen 2005: Dynamic stochastic optimization model for air traffic flow management with en route and airport capacity constraints, 6th USA/Europe Seminar on Air Traffic Management Research and Development ,ATM-2005, http://atmseminar.eurocontrol.fr/, Baltimore, MD,

Office of Federal coordinator for meteorological services, 2003: National Aviation Weather Program Mid-Course Assessment, Report FCM-R20-2003, Aug. 2003

Robinson, M., J. Evans, B. Crowe, D. Klingle-Wilson and S. Allan, 2004: CIWS operational benefits 2002-3: Initial estimates of convective weather delay reduction, MIT Lincoln Laboratory Project Report ATC-313 (9 April 2004)

Rhoda, D. and M. Pawlak, 1999: An assessment of thunderstorm penetrations and deviations by commercial aircraft in the terminal area, Massachusetts Institute of Technology, Lincoln Laboratory, Project Report NASA/A-2

Robinson, M and J. Evans, 2006: Quantifying ATC productivity enhancement for aviation convective weather decision support systems, 12th Conference on Aviation, Range, and Aerospace Meteorology, Atlanta, GA. Amer. Meteor. Soc.

Weber, M., J. Evans, M. Wolfson, R. DeLaura, W. Moser, B. Martin, J. Welch , J. Andrews, and D. Bertsimas, 2005: Improving Air Traffic Management during Thunderstorms, Digital Avionics System Conference (DASC), Washington, DC.

Wilson, F.W. and D.A. Clark, 1997: Estimation of the avoidable cost of ceiling and visibility events at major airports. 7th Conference on Aviation, Range and Aerospace Meteorology, Long Beach, CA.