An Experiment to Measure the Value of Statistical Probability Forecasts for Airports

ROSS KEITH
Bureau of Meteorology, and James Cook University, Townsville, Queensland, Australia

STEPHEN M. LEYTON
Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 29 April 2005, in final form 5 July 2006)

ABSTRACT

The economic value of weather forecasts for airports for commercial aviation is investigated by introducing financial data into the decision-making process for fuel carriage by aircraft. Using specific operating costs for a given flight, an optimal decision probability threshold can be calculated that identifies whether that flight should carry extra fuel, in case of adverse weather conditions and subsequent diversion. Forecasts of these adverse conditions can then be applied to a critical threshold to make a real-time decision regarding the carriage of additional fuel. This study focuses on forecasts of low ceiling and/or reduced visibility and their corresponding impact on forecast value for flights arriving at three major airports in the United States. Eighteen daily flights by American Airlines were examined during a 14-month period, a total of approximately 7500 flights. Using operating cost data from this period, a critical decision threshold was derived for each daily flight. Two sets of forecasts, statistically derived probabilistic forecasts and National Weather Service terminal aerodrome forecasts (TAFs), were then applied to each flight’s fuel carriage decision-making process. The probabilistic forecasts, which utilize regional surface observations, were generated for the destination airport with a lead time appropriate to the airline’s flight planning time. If the forecast probability of adverse weather was greater than the critical decision threshold for a given flight, then additional fuel was deemed necessary for that flight. The categorical TAFs that corresponded timewise to the developed probabilistic forecasts were obtained for each location. For this study, a categorical “yes” forecast denotes the expectation that the visibility and/or cloud ceiling conditions are such that extra fuel is required, while a categorical “no” forecast does not require extra fuel. The analysis presented herein indicates that by using statistical, probabilistic forecasts rather than categorical forecasts, a significant saving is made in operating costs. This is probably because of a more optimal balance between false alarms and misses for each flight, rather than more “accurate” forecasts per se. This is the mechanism by which probabilistic forecasts create value, rather than increasing the number of hits and correct rejections and/or decreasing the number of false alarms and misses. For each of the flights investigated in this study, the total cost of using probabilistic forecasts was less than that of using TAFs. An average of $23,000 is saved per flight during this 14-month period. Projecting these figures over all American Airlines flights, a potential annual savings of approximately $50 million in operating costs would be realized by using probabilistic forecasts of adverse landing weather conditions instead of the traditional TAF.

1. Introduction

Commercial aircraft taking off and landing around the country regularly have to deal with the possible occurrences of adverse weather conditions that require decisions to be made that could be different from those made under ideal flying conditions. Perhaps visibility is so poor that airports that normally use side-by-side dual-runway approaches must reduce to single-runway approaches, cutting the flow rate by 50%. Maybe thunderstorms are expected to cross over an airport and planes may be forced into a holding pattern until the weather improves. In these instances and many others, decisions have to be made either prior to departure or en route that have significant economic impacts.

Consider a destination airport in which adverse
weather is a possibility. First, the airline dispatch must decide whether the situation warrants the carriage of additional fuel in case a deviation from the existing flight plan is necessary. Next, the plane departs but conditions may or may not be suitable for landing as the plane approaches its destination. If conditions at the destination airport are or become unfavorable for landing, depending on whether additional fuel is being carried, the pilot can decide to divert immediately to a predetermined alternate airport, go into a holding pattern for a period of time hoping that weather conditions improve, or continue to the destination airport, attempt to land one or more times and then either safely land or divert to the alternate airport. If conditions at the destination are favorable for landing, regardless of whether additional fuel is being carried, the plane continues to its destination and safely lands. Obviously, the amount of fuel being carried by the aircraft affects the range of choices that can be made in a given situation.

Because each additional gallon of fuel has an inherent weight, the choice to load additional fuel requires even more additional fuel (i.e., carriage fuel) just to carry the extra weight. If additional fuel is carried but not utilized, then the carriage fuel is burned unnecessarily, resulting in an excessive fuel expense. If additional fuel is not carried but needed, then the airline must bear the very expensive cost of a diversion. It is important to note that even with the extra fuel, weather conditions may be so poor that a holding pattern or diversion may still be warranted. Because each gallon of fuel has an inherent cost, it is in the best interest of the airline to balance the cost of carrying additional fuel when it may not be necessary with the cost of holding patterns and diversions. To optimize this decision and minimize excessive costs, it is imperative that a rational decision rule be developed to determine whether a given flight should carry the additional fuel. This decision rule must be unique to each flight and would be represented by a “critical” threshold probability of adverse weather conditions occurring.

A cost-based decision-making model is an important step in reducing costs but in order to utilize this system, weather forecasts must be expressed as a probability of occurrence for a given condition. The current system, terminal aerodrome forecasts (TAFs), contains deterministic forecasts (i.e., “yes” representing 100% probability or “no” representing 0% probability) of low ceiling and visibility. However, an alternate forecasting method, using recent research findings, would be to develop short-term probabilistic forecasts of low ceiling and visibility that range from 0% to 100%, inclusive. These probabilistic forecasts could be used to trigger the decision, with a probability less than the “critical” threshold resulting in no additional fuel added and a probability greater than the “critical” threshold resulting in additional fuel being added.

This manuscript brings these ideas together in a case study of 18 daily flights over a span of 14 months by a major U.S. commercial airline in which actual flight costs are identified and a decision rule is created for each flight, which is, in turn, triggered by the probability of adverse weather conditions occurring. The estimated costs incurred by utilizing the current deterministic forecasts to make fuel carriage decisions are then compared to the estimated costs when utilizing probabilistic forecasts. It is found that for each flight route examined, the use of probabilistic forecasts results in a greater cost savings. In fact, extrapolation of these costs savings to a major carrier’s entire fleet yields an annual savings of as much as $50 million.

2. The cost-based decision rule for fuel carriage

Federal Aviation Administration (FAA) regulations require airlines to plan sufficient fuel for an aircraft to proceed to an alternate destination airport if a weather forecast predicts a cloud ceiling of \( \leq 2000 \text{ ft} \) and/or visibility \( < 3 \text{ statutory miles (SM)} \). This is referred to as the alternate minimum (sometimes referred to as “alt. min.”) threshold. The decision whether to carry additional fuel is complicated by the fact that most flights can still land at an airport even when the ceiling is at or below 2000 ft and/or the visibility is less than 3 SM. The very existence of the alternate minimum threshold is as a mandated safety buffer to account for errors in the categorical forecast.

a. Derivation of critical decision thresholds

The critical fuel carriage threshold for a given flight can be derived using a basic contingency table, as shown in Fig. 1. First, the airline dispatch must decide

<table>
<thead>
<tr>
<th>Decision</th>
<th>Weather</th>
<th>Cost (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fly w/ Additional Fuel (1)</td>
<td>Below Alt. Min. (3)</td>
<td>( C_A = \text{costs of flight diversions, holding patterns, revenue (missed connection), passenger dissatisfaction} ) → HIT</td>
</tr>
<tr>
<td>Fly w/o Additional Fuel (0)</td>
<td>Above Alt. Min. (9)</td>
<td>( C_A = \text{fuel carriage} \rightarrow \text{FALSE ALARM} )</td>
</tr>
<tr>
<td>Fly w/o Additional Fuel (0)</td>
<td>Below Alt. Min. (5)</td>
<td>( C_A = \text{lost revenue (fewer missed connections since other planes also delayed), ground service costs, passenger dissatisfaction} \rightarrow \text{MISS} )</td>
</tr>
<tr>
<td>Fly w/o Additional Fuel (0)</td>
<td>Above Alt. Min. (8)</td>
<td>( C_A = \text{zero} \rightarrow \text{CORRECT REJECTION} )</td>
</tr>
</tbody>
</table>

**FIG. 1.** Decision tree showing the possible outcomes based upon whether additional fuel is carried and whether the weather conditions at the destination are below the alternate minimum.
whether a given plane should add additional fuel to its
prescheduled fuel load. Next, the flight may or may not
encounter alternate minimum conditions at its destination. This decision tree outlines these four possible out-
comes and the costs incurred by each.

If one defines $p$ as the probability that there will be
below alternate minimum conditions, then $(1 - p)$ is
the probability of conditions being above the alternate
minimum threshold. So for the “additional fuel” option,
the expected cost, $EC_1$, is defined as

$$EC_1 = p \times C_{11} + (1 - p) \times C_{10}.$$  

Similarly, for the “no additional fuel” option, the ex-
pected cost ($EC_0$) is

$$EC_0 = p \times C_{01} + (1 - p) \times C_{00}.$$  

The logical decision then is to carry extra fuel if the
expected cost from carrying the fuel ($EC_1$) is less than
that from not carrying the fuel ($EC_0$), for example,
$EC_1 < EC_0$. Therefore, to calculate the critical prob-
ability threshold of weather below the alternate mini-
num, we set $EC_1 = EC_0$:

$$p \times C_{11} + (1 - p) \times C_{10} = p \times C_{01} + (1 - p) \times C_{00}.$$  

Rearranging,

$$p \times [C_{11} - C_{10}] + (C_{00} - C_{01}) = C_{00} - C_{10}.$$  

By definition, the false alarm cost (FAC) is the cost
of a false alarm over and above the cost of a correct
rejection; that is, $FAC = C_{10} - C_{00}$. The miss cost (MC)
is similarly defined as the cost of a miss over and above
the cost of a hit (i.e., the correct forecast of an event);
that is, $MC = C_{01} - C_{11}$. Note that both of the defini-
tions specify the avoidable cost of the weather forecast,
the cost relative to a correct forecast.

Solving for $p$ yields

$$p = \frac{FAC}{FAC + MC}.$$  

Further, if we define the ratio (FAC/MC) as the cost
ratio (CR), then

$$p = \frac{CR}{1 + CR}.$$  

This is in agreement with Keith (2003) and Mason
(2004).

As shown by the derivation above, the critical fuel
 carriage threshold is driven completely by the cost pa-
rameters of each flight. It is important to realize that
most of the cost of a weather event is still incurred even
if the event is correctly forecast. Usually, if a weather
event occurs at an airport that has an impact on incom-
ing flights, only a small portion of the impact is avoid-
able even when the forecast is correct. A perfect fore-
cast is useless to an airline if no corrective action can be
taken based on the forecast. Therefore, it is this avoid-
able component that must be studied when considering
financial impact.

b. Cost data

For each flight, the following costs are relevant for
calculating the critical fuel carriage threshold:

(a) the cost to carry the weight of extra fuel unneces-
sarily,

(b) the cost of an en route diversion, and

c) the cost of a diversion from a destination after a
missed approach.

These costs summarize the economic impacts of air-
port forecasts. Scenario a is the cost of a false alarm,
whereby the forecast causes extra fuel to be carried
unnecessarily. The cost is the extra fuel used to carry
that extra weight, usually a fairly modest figure of the
order of $100. Scenario b occurs if no extra fuel is car-
rried, and during the flight the weather at the destination
is such that a landing may not be possible, and a deci-
sion is made en route to divert. Scenario c happens if
extra fuel is carried and the weather at the destination
is bad. According to American Airlines, a pilot will
almost always make an attempt to land given that he or
she has adequate fuel to return to cruising altitude and
proceed to another airport. This leads to the seemingly
incongruous fact that, on a small number of occasions,
the biggest cost actually occurs following a correct fore-
cast. There are undoubtedly more subscenarios than
these three, but in order to simplify the issue of opera-
tional forecast use for this study, we have settled on
these three scenarios. Discussion with American Air-
lines staff suggested that these handle the great major-
ity of situations.

c. Calculation of critical thresholds

To calculate a critical threshold for each flight, one
must accurately calculate the cost of a “hit” above stan-
dard operating costs as well as the cost of a “miss”
above standard operating costs. In addition, one must
be able to account for the impact of forecast errors. The
only viable way of making allowances for this is to fol-
low the same methodology as Keith (2003). Keith cal-
culated the climatological rate of low ceiling and re-
duced visibility events in which landing was impossible
and expressed this as a ratio of the climatological rate
that conditions were below the alternate minimum.
This ratio is calculated for each month and different
times of the day.

Not only did we have to consider the FAA-mandated
alternate minimum ceiling and visibility conditions for
fuel carriage, but we had also to consider the landing minimum, which represents the values of ceiling and visibility that would preclude an aircraft from landing. As opposed to the fixed values of ceiling and visibility that represent alternate minimum conditions, the landing minima vary according to a number of factors, including aircraft type, standard of airport landing instrumentation, and training level of the pilot. That said, we learned that American Airlines most commonly implements landing minima of ceiling $\leq 200$ ft and visibility $< \frac{3}{4}$ mi (W. Qualley 2005, personal communication). Therefore, we decided it was reasonable to apply these landing minima throughout our study. While we were interested in utilizing the landing minima, an investigation of climatology at the arrival airports revealed that these conditions occur quite rarely, with an occurrence rate of $\leq 1\%$. Because of this, the ratio of “below landing minimum” events to “below alternate minimum” events would be too low to develop stable probabilistic forecast equations. Therefore, we had to focus solely on alternate minimum conditions in this study, other than for the purpose of calculating flight operating costs.

Some difficulty arises in calculating the cost of a “hit,” or correct forecast of weather conditions being below the alternate minimum. If the TAF correctly forecast conditions below the alternate minimum, advice from American Airlines is that the pilot will usually attempt a landing, but if conditions are such as to preclude landing, he or she will then climb out and proceed to another airport. If the climatological ratio of the number of occasions of below landing minimum conditions to the number of occasions of below alternate minimum conditions is $r$, then

$$\text{Hit Cost} = (r \times \text{Cost of Diversion from Destination}) + [(1 - r) \times \text{False Alarm Cost}].$$

Recall that for a hit, alternate fuel is being carried. Therefore, $r$ is usually a small value, but is multiplied by a large dollar value (i.e., the diversion cost), whereas $(1 - r)$ is typically close to a value of 1 but is multiplied by a smaller number (i.e., the cost of carrying additional fuel). In practice both terms of the equation are significant in the calculation of the hit cost.

The cost of a missed event is also complex. Consider the situation where an airline dispatcher plans minimum fuel due to a forecast of above alternate minimum conditions. Occasionally, after the flight departs, the weather at the destination airport deteriorates to below alternate minimum. This is almost always followed by an amended TAF, and on some occasions an amended observations-based probabilistic forecast. Sometimes the dispatcher, in consultation with the pilot, can agree upon an alternate airport that is nearby the actual destination while en route. This allows the flight to continue along its original flight path and then make a decision about whether to proceed when the flight is closer to the planned destination. However, this is not always possible and the flight diverts to another airport, in order to land and refuel. We were not able to access detailed information for each flight showing how often in-flight deterioration of weather caused en route diversions. Therefore, for the purpose of the study, we considered three scenarios: that en route weather deteriorations caused diversions 20%, 60%, and 100% of the times that extra fuel was not carried.

3. The deterministic aviation weather forecast

TAFs, created by the National Weather Service (NWS), are airport-specific forecasts of visibility and ceiling at a microscale level and have lead times of up to 24 h (though only the first 12 h are typically used operationally). Numerical models are unable to adequately define the fine scales required to analyze and then predict the detailed evolution of these parameters, which is required by the aviation industry. So, forecasters are required to produce these forecasts by primarily reverting to climatology and their own experiences. Thus, it can be said with a fair degree of validity that TAFs are still at a stage where synoptic-scale forecasting was before the advancement of numerical weather prediction. Various verification studies, for example, Dal-lavalle and Dagostaro (1995), Keith (2003), and Harvey et al. (1992) have shown that, in the first 3–6 h of the TAF, raw persistence usually demonstrates greater skill.

Once these TAFs are created, FAA guidelines dictate to the airlines whether additional fuel must be carried. If the forecasted ceiling or visibility level at the destination airport is less than a specified limit (e.g., the alternate minimum), then the aircraft is required to carry extra fuel. This FAA rule is not obviously related to costs, and is certainly not sensitive to differences in costs between different flights. Thus, the airline experiences significantly excessive fuel costs simply by adhering to FAA requirements.

For the purpose of the case study described below, an archive of TAFs, spanning from April 2002 to May 2003, was obtained in order to examine how these forecasts would compare to the overall costs resulting from the cost-based decision rule system. These TAFs were produced by the National Weather Service office that is responsible for each of the three airports listed in the case study discussion. Of interest to this study is wheth-
er the TAF predicts adverse weather such that flight dispatchers must specify that a given flight should add extra fuel.

4. The probabilistic aviation weather forecast

There have been some attempts to investigate alternatives to traditional numerical model prediction of those parameters that are crucial for landing aircraft. In particular, observations-based statistical techniques have shown great promise for improving short lead time forecasts of these surface weather conditions. Vislocky and Fritsch (1997) demonstrated that such a statistical forecasting system has superior skill compared to numerical models for forecasting ceiling and visibility at these short lead times. This system considered a network of surface observations surrounding an observing site (often an airport) to produce (via regression techniques) probabilistic forecasts of ceiling and visibility for that observing site. The results indicated that such a system has greater skill than model output statistics (MOS) derived forecasts, as well as persistence climatology, out to a lead time of 6 h. Leyton and Fritsch (2003, 2004) extended this work by demonstrating a further increase in skill when utilizing higher-density and higher-frequency surface observations for these short lead time forecasts.

While observations-based statistical systems have shown greater forecast skill over traditional techniques, little has been done to investigate whether the economic value of the forecast has been improved as well. Keith (2003) demonstrated how human forecasters’ subjective airport forecasts, when produced in a probabilistic format, should increase the value of that forecast to an airline compared to the traditional categorical airport forecast. This study looks to further examine the impact of probabilistic forecasts on forecast value by utilizing actual financial data.

a. Data

Two datasets were utilized to develop the statistical forecasts equations. The first contained hourly surface weather observations [i.e., Surface Airways Observations (SAOs)] for the period of January 1982–December 1996. The second dataset contained hourly observations [in the METAR (routine aviation weather report) format] spanning from January 1997 to July 2003. Each dataset included hourly observations of temperature, dewpoint, wind speed and direction, cloud cover (from which ceiling can be derived), visibility, and present weather at over 1500 automated weather-observing sites around the United States. To calculate climatological values of various weather parameters, the datasets were combined to produce values that represented weather conditions over 22 yr. In addition, the latter dataset is utilized for statistical forecast development. Leyton and Fritsch (2003, 2004) provide a detailed explanation of the data processing procedures.

b. Development of probabilistic forecasts

Statistical forecast equations were developed in the same manner as described by Leyton and Fritsch (2003, 2004). To replicate the decision-making process involved in fuel carriage for a flight, careful consideration was given to the initialization and valid times for the probabilistic forecasts. This is mainly because decisions regarding the loading of jet fuel typically are made during the dispatcher’s flight planning, which usually occurs 1–2 h prior to the flight’s departure. Another complicating factor is due to the nature of observations-based probabilistic forecasts. As described by Leyton and Fritsch (2003, 2004), these forecasts can only be initialized and verified at the top of the hour. However, scheduled flights arrive and depart throughout a given hour. Therefore, to create the most accurate representation of the actual decision-making process, all of these factors were carefully considered.

It is important to note that in this study, adverse weather conditions were limited to low ceiling and/OR reduced visibility. Thus, no specific attempt was made to statistically forecast the occurrence of thunderstorms at a given location. Operationally, the forecast of a thunderstorm is considered as being below alternate minimum. Of the total number of occasions of below alternate minimum conditions, less than 10% were as a result of thunderstorms, and of this number over half were captured by the 2000-ft and/or 3-mi criteria. While thunderstorms are the major consideration for in-flight weather, low ceiling and visibility at terminal airports are by far the more important issue for initial fuel allocation. The omission of thunderstorm forecasts is therefore not expected to significantly affect the validity of the result of the comparison.

5. A case study of cost-based decision rule system

A case study was performed in which the cost-related impact of low-ceiling and/or visibility forecasts was examined for 18 daily flights arriving at three American Airlines “hub” airports: Dallas–Fort Worth International Airport (DFW), John F. Kennedy International Airport in New York City (JFK), and Lambert–St. Louis International Airport (STL). A database of op-
Critical fuel carriage decision probabilities for each flight, for each month.

<table>
<thead>
<tr>
<th>Flight</th>
<th>Departure time (UTC)</th>
<th>Flight duration (h:min)</th>
<th>Arrival time (UTC)</th>
<th>Critical probability values for each flight</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIA–JFK</td>
<td>0025</td>
<td>2:47</td>
<td>0312</td>
<td>Jan: 0.05 Apr: 0.04 Mar: 0.04 Jun: 0.04 Aug: 0.04 Sep: 0.04 Oct: 0.04 Nov: 0.04 Dec: 0.04 Annual avg: 0.04</td>
</tr>
<tr>
<td>DFW–JFK</td>
<td>2336</td>
<td>3:20</td>
<td>0256</td>
<td>Jan: 0.06 Apr: 0.05 May: 0.05 Jun: 0.05 Jul: 0.05 Aug: 0.05 Sep: 0.05 Oct: 0.05 Nov: 0.05 Dec: 0.05 Annual avg: 0.05</td>
</tr>
<tr>
<td>SJU–JFK</td>
<td>2215</td>
<td>3:55</td>
<td>0210</td>
<td>Jan: 0.03 Apr: 0.03 May: 0.03 Jun: 0.03 Jul: 0.03 Aug: 0.03 Sep: 0.03 Oct: 0.03 Nov: 0.03 Dec: 0.03 Annual avg: 0.03</td>
</tr>
<tr>
<td>LAX–JFK</td>
<td>2200</td>
<td>5:10</td>
<td>0310</td>
<td>Jan: 0.10 Apr: 0.08 May: 0.08 Jun: 0.11 Jul: 0.08 Aug: 0.08 Sep: 0.07 Oct: 0.08 Nov: 0.07 Dec: 0.08 Annual avg: 0.08</td>
</tr>
<tr>
<td>AUS–DFW</td>
<td>1856</td>
<td>1:00</td>
<td>1956</td>
<td>Jan: 0.05 Apr: 0.05 May: 0.05 Jun: 0.05 Jul: 0.05 Aug: 0.05 Sep: 0.05 Oct: 0.05 Nov: 0.05 Dec: 0.05 Annual avg: 0.05</td>
</tr>
<tr>
<td>SEA–DFW</td>
<td>1920</td>
<td>3:47</td>
<td>2307</td>
<td>Jan: 0.04 Apr: 0.04 May: 0.04 Jun: 0.04 Jul: 0.03 Aug: 0.03 Sep: 0.06 Oct: 0.04 Nov: 0.04 Dec: 0.04 Annual avg: 0.04</td>
</tr>
<tr>
<td>SAT–DFW</td>
<td>2206</td>
<td>1:08</td>
<td>2314</td>
<td>Jan: 0.06 Apr: 0.06 May: 0.05 Jun: 0.06 Jul: 0.05 Aug: 0.10 Sep: 0.06 Oct: 0.06 Nov: 0.06 Dec: 0.06 Annual avg: 0.06</td>
</tr>
<tr>
<td>ELP–DFW</td>
<td>2114</td>
<td>2:36</td>
<td>2350</td>
<td>Jan: 0.02 Apr: 0.02 May: 0.02 Jun: 0.02 Jul: 0.02 Aug: 0.02 Sep: 0.02 Oct: 0.02 Nov: 0.02 Dec: 0.02 Annual avg: 0.02</td>
</tr>
<tr>
<td>SFO–DFW</td>
<td>0115</td>
<td>3:30</td>
<td>0445</td>
<td>Jan: 0.03 Apr: 0.03 May: 0.03 Jun: 0.03 Jul: 0.03 Aug: 0.03 Sep: 0.03 Oct: 0.03 Nov: 0.03 Dec: 0.03 Annual avg: 0.03</td>
</tr>
<tr>
<td>STL–DFW</td>
<td>1458</td>
<td>2:00</td>
<td>1658</td>
<td>Jan: 0.05 Apr: 0.04 May: 0.04 Jun: 0.04 Jul: 0.04 Aug: 0.04 Sep: 0.05 Oct: 0.04 Nov: 0.04 Dec: 0.04 Annual avg: 0.04</td>
</tr>
<tr>
<td>LAS–STL</td>
<td>2135</td>
<td>3:00</td>
<td>0035</td>
<td>Jan: 0.02 Apr: 0.02 May: 0.02 Jun: 0.02 Jul: 0.02 Aug: 0.02 Sep: 0.02 Oct: 0.02 Nov: 0.02 Dec: 0.02 Annual avg: 0.02</td>
</tr>
<tr>
<td>LAX–STL</td>
<td>1830</td>
<td>3:33</td>
<td>2203</td>
<td>Jan: 0.02 Apr: 0.02 May: 0.02 Jun: 0.02 Jul: 0.02 Aug: 0.02 Sep: 0.02 Oct: 0.02 Nov: 0.02 Dec: 0.02 Annual avg: 0.02</td>
</tr>
<tr>
<td>ATL–STL [1]</td>
<td>1222</td>
<td>1:52</td>
<td>1414</td>
<td>Jan: 0.08 Apr: 0.06 May: 0.06 Jun: 0.05 Jul: 0.05 Aug: 0.05 Sep: 0.06 Oct: 0.06 Nov: 0.06 Dec: 0.06 Annual avg: 0.06</td>
</tr>
<tr>
<td>MSP–STL [1]</td>
<td>1356</td>
<td>1:35</td>
<td>1531</td>
<td>Jan: 0.07 Apr: 0.05 May: 0.05 Jun: 0.05 Jul: 0.05 Aug: 0.05 Sep: 0.05 Oct: 0.05 Nov: 0.05 Dec: 0.05 Annual avg: 0.05</td>
</tr>
<tr>
<td>ORD–STL</td>
<td>1702</td>
<td>1:15</td>
<td>1817</td>
<td>Jan: 0.06 Apr: 0.05 May: 0.05 Jun: 0.05 Jul: 0.05 Aug: 0.05 Sep: 0.05 Oct: 0.05 Nov: 0.05 Dec: 0.05 Annual avg: 0.05</td>
</tr>
<tr>
<td>MIA–STL</td>
<td>1723</td>
<td>3:00</td>
<td>2023</td>
<td>Jan: 0.08 Apr: 0.07 May: 0.07 Jun: 0.07 Jul: 0.07 Aug: 0.07 Sep: 0.07 Oct: 0.07 Nov: 0.07 Dec: 0.07 Annual avg: 0.07</td>
</tr>
<tr>
<td>ATL–STL [2]</td>
<td>1840</td>
<td>1:52</td>
<td>2032</td>
<td>Jan: 0.04 Apr: 0.03 May: 0.03 Jun: 0.03 Jul: 0.03 Aug: 0.07 Sep: 0.07 Oct: 0.07 Nov: 0.07 Dec: 0.07 Annual avg: 0.07</td>
</tr>
<tr>
<td>MSP–STL [2]</td>
<td>1720</td>
<td>1:52</td>
<td>1912</td>
<td>Jan: 0.04 Apr: 0.03 May: 0.03 Jun: 0.03 Jul: 0.04 Aug: 0.02 Sep: 0.02 Oct: 0.02 Nov: 0.02 Dec: 0.02 Annual avg: 0.02</td>
</tr>
</tbody>
</table>

Operating costs for several American Airlines flights arriving into each airport was obtained; this dataset contains the cost per flight information for each of the 18 daily flights, in 2003 dollars. It is important to note that the flights used in this study were predetermined in consultation with the airline and represent varying distances and departure–arrival times during the day. Thus, this study is not focused simply on a specific type of flight but rather a broad representation of an airline’s typical daily flights. These operating costs represent typical costs of various scenarios for each flight during the April 2002–May 2003 period; approximately 7500 flights in all.

For the purpose of calculating the hit cost, values of $r$ were calculated for DFW, JFK, and STL for each month during the 14-month period, in 3-h segments, and then applied to each flight as appropriate. Thus, the expected hit cost was calculated and subtracted from the cost of a missed event to arrive at the miss cost. Using the cost-based decision rule system derived above, the critical fuel carriage threshold is driven completely by the cost parameters of each flight. Critical thresholds have been identified as high as 0.27 for some long international flights, and as low as 0.02 for shorter domestic flights, as shown in Table 1.

Once the static critical fuel carriage threshold was calculated for each daily flight, forecasts of below alternate minimum conditions could then be utilized to assess whether a given flight should carry additional fuel. Two sets of independent forecasts were used in this trial study: statistical probabilistic forecasts of low ceiling and/or reduced visibility (ranging from 0% to 100%, inclusive) and TAFs from the airport’s local NWS office (either 0% or 100%). The purpose of utilizing two forecasts was to not only identify whether this technique worked but also whether one forecast type delivers superior value compared to the other. NWS TAFs are the current source of aviation weather forecasts. However, as noted previously, recent studies have demonstrated the superior skill of probabilistic forecasts, especially when considering short lead times.

Table 2 shows the hit rate and false alarm rate for each flight, with 95% confidence intervals. The 95% confidence intervals have been calculated using the formula quoted by Stephenson (2000). It can be seen that the probabilistic forecasts are driven by the cost calculations to a much more conservative tactic (higher hit rate and false alarm rate) than the TAFs. Perusal of the data showed that the TAFs sometimes missed expensive events, while the probabilistic forecasts had more false alarms than the TAFs. The tight confidence intervals give one some faith in the robustness of the value calculations. The dollar values are calculated directly (i.e., no inference is involved) from the probabilistic forecasts and the TAFs. So the significant differences in value are likely to be caused by the different tactic or decision threshold.

As mentioned above, we were not able to access information that showed how often in-flight deterioration of weather caused en route diversions. Thus, the analysis has been performed using three different assump-
tions: that en route weather deteriorations caused diversions 20%, 60%, and 100% of the times that extra fuel was not carried. Table 3 shows a comparison of total costs per flight using the traditional TAF and the observations-based probabilistic forecasts, with consideration given to the different assumptions regarding weather deterioration. These costs, in 2003 dollars, represent the total costs from April 2002 to May 2003. It is important to note that for each of the 18 flights, the use of observations-based probabilistic forecasts resulted in lower total costs than the traditional TAFs. Extrapolating these results for 18 flights across the whole American Airlines domestic fleet (approximately 2200 flights per day), the airline would have saved between $15 million and $50 million annually if the initial flight planning had been done using the observations-based probabilistic airport forecasts rather than the traditional TAF. By comparison, in 2003, the annual fuel expense of a major U.S. commercial airline was $1 billion or more. Thus, by utilizing this simple decision rule system along with the probabilistic forecasts, total annual fuel expenses could be reduced by as much as 2.5%. (Note: in 2005, the total annual fuel expense of American Airlines was near $4 billion, while the total annual operating expense was over $18 billion.)

6. Summary and discussion

The sole purpose of this study was to examine whether significant savings in fuel expenses could be realized by switching from the current decision-making process to one that utilizes cost-based decision rules and probabilistic weather forecasts of adverse weather conditions. The current system utilizes TAFs in conjunction with FAA policy that allows airline dispatchers to make fuel carriage decisions arbitrary of the costs involved. We explore the development of a cost-based decision system that identifies a critical probability of adverse weather conditions for each flight. Finally, we examine two different types of probabilistic forecasts of low ceiling and/or visibility to apply to the cost-based decision rule: converting the TAF deterministic forecasts into probabilistic forecasts (100% to represent “yes” and 0% to represent “no”) and creating statistically derived probabilistic forecasts (0%-100%, inclusive). We feel that a cost-based decision system is preferable in order to minimize fuel costs and we find that utilizing statistically derived probabilistic forecasts to determine fuel carriage results in a significant cost savings compared to the deterministic TAF forecasts. The savings found in this study suggest an annual fuel cost reduction of $15–$50 million for a major commercial airline.

We do not claim to have precisely represented all the operational intricacies of forecast use by airlines. To quantify this more exactly it would be necessary to work directly with an airline, to more adequately specify the details of the financial impact of weather and weather forecasts on airline operations. This impact will undoubtedly be complex some of the time, and will require extension of our current algorithm. However, the gross differences in the dollar results from use of the two different types of forecasts certainly justify

---

### Table 2. Hit rate and false alarm rate for each flight, for both the probabilistic forecasts and the NWS TAFs, with 95% confidence intervals, assuming that en route weather deteriorations caused diversions 20%, 60%, and 100% of the time.

<table>
<thead>
<tr>
<th>Probability forecasts</th>
<th>Hit rate</th>
<th>False alarm rate</th>
<th>TAFs</th>
<th>Hit rate</th>
<th>False alarm rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIA–JFK</td>
<td>0.983 (0.966, 0.992)</td>
<td>0.391 (0.346, 0.438)</td>
<td>0.816 (0.776, 0.850)</td>
<td>0.081 (0.059, 0.111)</td>
<td></td>
</tr>
<tr>
<td>DFW–JFK</td>
<td>0.992 (0.978, 0.997)</td>
<td>0.352 (0.308, 0.399)</td>
<td>0.810 (0.770, 0.845)</td>
<td>0.085 (0.062, 0.115)</td>
<td></td>
</tr>
<tr>
<td>SJU–JFK</td>
<td>0.993 (0.979, 0.998)</td>
<td>0.802 (0.761, 0.837)</td>
<td>0.818 (0.778, 0.852)</td>
<td>0.082 (0.059, 0.112)</td>
<td></td>
</tr>
<tr>
<td>LAX–JFK</td>
<td>0.983 (0.966, 0.992)</td>
<td>0.262 (0.222, 0.306)</td>
<td>0.838 (0.800, 0.870)</td>
<td>0.087 (0.064, 0.118)</td>
<td></td>
</tr>
<tr>
<td>AUS–DFW</td>
<td>0.961 (0.938, 0.976)</td>
<td>0.116 (0.089, 0.150)</td>
<td>0.765 (0.722, 0.803)</td>
<td>0.056 (0.038, 0.082)</td>
<td></td>
</tr>
<tr>
<td>SEA–DFW</td>
<td>0.980 (0.962, 0.990)</td>
<td>0.163 (0.131, 0.201)</td>
<td>0.768 (0.725, 0.806)</td>
<td>0.047 (0.031, 0.072)</td>
<td></td>
</tr>
<tr>
<td>SAT–DFW</td>
<td>0.968 (0.947, 0.981)</td>
<td>0.109 (0.083, 0.142)</td>
<td>0.722 (0.677, 0.763)</td>
<td>0.044 (0.028, 0.068)</td>
<td></td>
</tr>
<tr>
<td>ELP–DFW</td>
<td>0.972 (0.951, 0.984)</td>
<td>0.124 (0.096, 0.156)</td>
<td>0.306 (0.264, 0.351)</td>
<td>0.080 (0.058, 0.110)</td>
<td></td>
</tr>
<tr>
<td>SFO–DFW</td>
<td>0.982 (0.964, 0.991)</td>
<td>0.291 (0.250, 0.336)</td>
<td>0.628 (0.581, 0.673)</td>
<td>0.048 (0.031, 0.073)</td>
<td></td>
</tr>
<tr>
<td>STL–DFW</td>
<td>0.985 (0.968, 0.993)</td>
<td>0.643 (0.596, 0.687)</td>
<td>0.597 (0.550, 0.643)</td>
<td>0.045 (0.029, 0.069)</td>
<td></td>
</tr>
<tr>
<td>LAS–STL</td>
<td>0.946 (0.920, 0.964)</td>
<td>0.868 (0.832, 0.897)</td>
<td>0.594 (0.547, 0.640)</td>
<td>0.109 (0.083, 0.142)</td>
<td></td>
</tr>
<tr>
<td>LAX–STL</td>
<td>0.980 (0.966, 0.992)</td>
<td>0.514 (0.470, 0.561)</td>
<td>0.631 (0.582, 0.675)</td>
<td>0.078 (0.057, 0.108)</td>
<td></td>
</tr>
<tr>
<td>ATL–STL [1]</td>
<td>0.971 (0.950, 0.983)</td>
<td>0.477 (0.430, 0.525)</td>
<td>0.657 (0.611, 0.701)</td>
<td>0.062 (0.043, 0.089)</td>
<td></td>
</tr>
<tr>
<td>ATL–STL [2]</td>
<td>0.976 (0.957, 0.987)</td>
<td>0.294 (0.253, 0.339)</td>
<td>0.667 (0.621, 0.710)</td>
<td>0.087 (0.064, 0.118)</td>
<td></td>
</tr>
<tr>
<td>MSP–STL [1]</td>
<td>0.992 (0.978, 0.997)</td>
<td>0.534 (0.486, 0.581)</td>
<td>0.634 (0.587, 0.679)</td>
<td>0.051 (0.034, 0.076)</td>
<td></td>
</tr>
<tr>
<td>MSP–STL [2]</td>
<td>0.994 (0.981, 0.998)</td>
<td>0.669 (0.623, 0.712)</td>
<td>0.596 (0.549, 0.642)</td>
<td>0.071 (0.050, 0.099)</td>
<td></td>
</tr>
<tr>
<td>ORD–STL</td>
<td>0.996 (0.984, 0.999)</td>
<td>0.239 (0.201, 0.282)</td>
<td>0.688 (0.642, 0.730)</td>
<td>0.074 (0.053, 0.103)</td>
<td></td>
</tr>
<tr>
<td>MIA–STL</td>
<td>0.952 (0.927, 0.969)</td>
<td>0.215 (0.179, 0.257)</td>
<td>0.548 (0.500, 0.595)</td>
<td>0.081 (0.059, 0.111)</td>
<td></td>
</tr>
</tbody>
</table>
further investigation of the use of observations-based probabilistic airport forecasts as a more valuable product than categorical TAFs.

Acknowledgments. We must express our gratitude to Warren Qualley for sharing his expertise in commercial aviation operations and weather forecasting. Kelvin Droegemeier is to be thanked for his continued support and insight during this effort. We also thank American Airlines for providing the operating cost data for the flights as well as NCAR for the weather data utilized in this study.

REFERENCES


### Table 3. A comparison of cumulative costs per flight using the traditional TAF and the observations-based probabilistic forecasts, with consideration given to the different assumptions regarding weather deterioration.

<table>
<thead>
<tr>
<th>Flight (distance in mi)</th>
<th>Costs using TAFs ($)</th>
<th>Costs using probability forecasts ($)</th>
<th>TAF costs – probability costs ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20% 60% 100%</td>
<td>20% 60% 100%</td>
<td>20% 60% 100%</td>
</tr>
<tr>
<td>MIA–JFK (1100)</td>
<td>29,167 38,762 54,406</td>
<td>23,320 26,910 31,387</td>
<td>5,847 11,852 23,019</td>
</tr>
<tr>
<td>DFW–JFK (1400)</td>
<td>59,020 61,674 71,388</td>
<td>52,810 57,345 59,914</td>
<td>6,210 4,329 11,474</td>
</tr>
<tr>
<td>SJU–JFK (1600)</td>
<td>81,072 96,300 113,996</td>
<td>73,592 87,694 97,688</td>
<td>7,480 8,606 16,308</td>
</tr>
<tr>
<td>LAX–JFK (2500)</td>
<td>72,274 79,870 88,743</td>
<td>68,743 73,904 78,272</td>
<td>3,531 5,966 10,471</td>
</tr>
<tr>
<td>AUS–DFW (200)</td>
<td>17,434 22,866 28,848</td>
<td>9,786 11,086 12,444</td>
<td>7,648 11,780 16,404</td>
</tr>
<tr>
<td>SEA–DFW (1650)</td>
<td>14,578 22,432 30,738</td>
<td>9,868 11,946 14,721</td>
<td>4,710 10,486 16,017</td>
</tr>
<tr>
<td>SAT–DFW (250)</td>
<td>17,460 19,875 21,706</td>
<td>4,986 6,875 8,859</td>
<td>12,474 13,000 12,847</td>
</tr>
<tr>
<td>ELP–DFW (550)</td>
<td>16,324 31,786 50,136</td>
<td>5,086 5,965 8,370</td>
<td>11,238 25,821 41,766</td>
</tr>
<tr>
<td>SFO–DFW (1450)</td>
<td>21,340 37,754 54,636</td>
<td>10,460 12,375 14,076</td>
<td>10,880 25,379 40,560</td>
</tr>
<tr>
<td>STL–DFW (550)</td>
<td>39,670 55,904 77,768</td>
<td>29,134 32,470 35,688</td>
<td>10,536 23,434 42,080</td>
</tr>
<tr>
<td>LAS–STL (1400)</td>
<td>15,320 25,575 37,106</td>
<td>10,667 8,946 10,704</td>
<td>5,953 16,629 26,402</td>
</tr>
<tr>
<td>LAX–STL (1600)</td>
<td>30,042 36,761 50,983</td>
<td>20,877 24,661 26,776</td>
<td>9,165 12,100 24,207</td>
</tr>
<tr>
<td>ATL–STL (500) [1]</td>
<td>39,328 51,785 65,430</td>
<td>32,087 35,620 38,730</td>
<td>6,521 16,165 26,700</td>
</tr>
<tr>
<td>MSP–STL (450) [1]</td>
<td>45,407 61,231 82,723</td>
<td>33,260 36,593 42,067</td>
<td>12,147 24,638 40,656</td>
</tr>
<tr>
<td>MSP–STL (450) [2]</td>
<td>30,945 43,139 57,849</td>
<td>26,431 29,923 34,048</td>
<td>4,514 13,216 23,801</td>
</tr>
<tr>
<td>ORD–STL (250)</td>
<td>17,672 29,877 43,049</td>
<td>11,756 15,676 19,543</td>
<td>5,916 14,201 23,506</td>
</tr>
<tr>
<td>MIA–STL (1050)</td>
<td>21,137 33,843 44,441</td>
<td>14,549 18,282 21,846</td>
<td>6,588 15,561 22,595</td>
</tr>
<tr>
<td>Total for all flights</td>
<td>590,959 778,842 1,010,889</td>
<td>454,404 520,055 583,928</td>
<td>136,555 258,787 426,961</td>
</tr>
</tbody>
</table>