18.1

Probabilistic Forecasts of Cloud Impacts at San Francisco International Airport

F. Wesley Wilson[1,2]
NCAR, Boulder, CO

1. INTRODUCTION

San Francisco International Airport (SFO) is unable to use independent parallel approaches to its closely-spaced parallel runways when Marine Stratus is present in the approach. Delay programs are imposed to regulate the flow of traffic to match the true arrival capacity of the airport. Failure to forecast accurately the times of onset and dissipation of stratus in the approach results in unnecessary delays, costly airborne holding and diversions, or in wasted capacity as the traffic management planners fail to match the arrival rate to the actual airport capacity. Theoretical studies have shown that accurate 1-2 hour forecasts of the times of clearing in the approach could provide substantial reductions in the delays and inefficiencies associated with the Marine Stratus impacts on air traffic at SFO (Clark and Wilson, 1997). This provides enough time for planes to travel to SFO from Western points of origin.

This investigation will focus on decision aids for assisting the modification or cancellation of an existing delay program. Implementation of an coordinated Forecast and Traffic Management system has been impeded by the lack of a mechanism for dealing with the uncertainties in forecast performance.

At issue is the trade-off between the delays incurred, when the strategy is overly conservative, and the airborne holding and congestion, which may occur when the strategy is overly aggressive. Requests from traffic managers and recent advances in automated traffic management models indicate that probabilistic forecasts may have a role in resolving these issues.

2. DETERMINISTIC FORECASTS FOR SFO

The SFO Marine Stratus Initiative has developed four core deterministic forecast algorithms, and a Consensus Forecast (CF) product (Wilson and Clark, 2000). The historical performance for each of the core forecasts have been studied and the error characteristics have been established for a variety of meteorological situations. There is no evidence that the errors are correlated. The consensus forecast is the weighed average of the core forecasts, using inverse variance weighting. This approach is followed, even when some of the core forecasts fail to be issued, due to missing data.

The four core forecasts are:

COBEL – a column model, which provides a detailed dynamical analysis of the heat budget for the dissipation.

RSFM – a statistical forecast model based on regional NWS surface data.

SSFM - a statistical forecast model based on NWS satellite data (visible channel).

LSFM – a statistical forecast model base on data obtained from special project sensors along the approach to SFO.

Forecasts of the expected time of clearing are issued bi-hourly from 9z to 15z, and hourly from 15z to 18z, or until the stratus has cleared in the approach. If two or more core forecasts are missing, then the CF forecast is computed from the available core forecasts, and the confidence factor is reduced. The early forecasts are used for deciding whether or not to invoke a delay program, and the later forecasts are used for deciding when to modify or cancel a delay program.

---

1. *This research was supported by the National Center of Excellence for Aviation Operations Research, under Federal Aviation Administration Research Grant 96-C-001 and contract number DTFAo3-97-D00004. This document has not been reviewed by the Federal Aviation Administration (FAA). Any opinions expressed herein do not necessarily reflect those of the FAA or the U.S. Department of Transportation.*

2. *Corresponding author address:* F. Wesley Wilson, NCAR, PO Box 3000, Boulder, CO 80307;
email: wes@ucar.edu

3. PROBABILISTIC FORECASTS FOR SFO

Probabilistic forecasts traditionally have been based on a logistic (nonlinear) regression against a set of predictors. Initial applications were directed towards Model Output Statistics (MOS) (Glahn and Lowery, 1972), where the predictors are taken from NWP model fields. Vislocky and Fritsch, 1997, have shown that using observations as predictors has merit for shorter forecast horizons, although data quality issues complicate the operational use of observational data (Allen, 2001). Another possibility is using other forecasts as the predictors (Fritsch et al, 2000). In the current application, the predictors for the logistic regression are taken from the core forecasts and the consensus forecast.

The selection of the objective function determines the characteristics of the resulting forecast system. Each of the above authors has chosen to minimize of the Brier Score (BS) (Brier, 1950). This approach produces reliable forecasts (Murphy, 1973). The probabilistic interpretation of the Pierce Skill Statistic (PSS) (Pierce, 1884 and Wilson, 2002) provides forecasts, which optimally separate correct positives from false positives, but which may not be reliable. The statistics literature suggests that Maximum Likelihood Estimation (MLE) is preferred for logistic regression (e.g. Darlington, 1990). The point is that there are choices, and we shall discover that different objective functions lead to forecasts with different strengths and weaknesses. We construct forecasts with a variety of objectives and compare the results.

To avoid over-training, it is important to have a sufficiently large set of cases in the training data. The SFO Marine Stratus Initiative provides deterministic forecasts in the summers of 1996-2002. Examination of the data reveals that the COBEL model frequently fails to issue a forecast. Since inclusion of COBEL would severely reduce the number of training cases, we have decided to build our probabilistic forecast models using the predictors RSFM, SSFM, LSFM, and CF. (CF contains COBEL information when it is available). Restricting to cases where these four forecasts are all available produces training sets with 191 cases at 15z, 181 cases at 16z, and 141 cases at 17z, sufficient for the construction of probabilistic forecast models with cross validation.

The first step is to build probability forecasts for quarter-hour categories, beginning at the time that the forecast is issued, and continuing to 24z. The forecast is the probability that the SFO approach will be cloud-free before the category end-time. The series of forecasted probabilities provides the cumulative distribution function (CDF). Forecast models are determined separately for each category. This means that we are building a succession of 5-parameter models, well within the capacity of the available training data.

To properly constrain the regression, it is necessary that there be a reasonable number of positive and negative cases for each category. For our analysis, we require at least 20 of each. Since the approach is not clear at the time that the forecast is issued, there often are not enough positives for an hour or more after the issue time. Since it usually clears before 20z, there often are not enough negatives after 19:30. The consequence of these data scarcities is that we are only able to generate probabilistic forecast models for a couple of hours in the middle of the CDF domain; we must extend the CDF to the extremes by other means. We assume the CDF is 0 at the issue time and 1 at 24z. We then extend the CDF to the full domain by logistic-linear interpolation.

This approach produces categorical probabilistic forecasts, and we can apply traditional categorical skill statistics to evaluate their performance. First of all, we evaluate the CF skill on these categories, using the PSS and the Correct Alert Ratio (CAR), the complement of the False Alert Ratio (FAR = 1 – CAR) (Wilks, 1995). Since the CAR is provides the historical performance of the deterministic forecast system, a stationarity hypothesis leads to its use as a CAR probabilistic forecast model:

$$CDF = CAR|(CF=T) \qquad (1)$$

(conditional probability).

For every CDF forecast, we can compute its probabilistic PSS and compare it to the PSS of the associated deterministic forecast. If the PSS of a CDF forecast is significantly less than the PSS of the underlying deterministic forecast, then there is a concern that we have lost skill in the transition to probabilistic forecasts.

Another measure of the accuracy of a deterministic forecast is the mean squared error (MSE). We recall that the MSE is the sum of the error variance and the squared error bias. There is an analogue for CDF forecasts. Define the Total Mean Squared Error (TMSE) to be the average error of CDF over the entire training set:

$$TMSE = (1/N)\sum_i \int (o_i - f_i(t))^2 dt \qquad (2)$$

where f_i is the probability density function of the CDF, and o_i is the observed outcome. Standard algebraic tricks lead to the decomposition:

$$TMSE = \overline{\beta}_f^2 + \overline{\sigma}_f^2 + \sigma_\beta^2 \qquad (3)$$

where β_f is the bias and σ_f^2 is the variance of the CDF, and σ_β^2 is the variance of the biases. For a deterministic forecast, the CDF is a step function (σ_f^2 =0) and so TMSE = MSE.

Minimizing the TMSE provides another possible objective for building CDF's. The practical difficulty with this approach is that it requires building the CDF's for all categories simultaneously. With approximately 10 categories in the middle of the CDF domain, this would require the simultaneous determination of 50 model parameters, well beyond the capacity of our training data. We have developed a nudging process, which allows us to adjust a candidate CDF forecast to improve its TSME, by adjusting 10 parameters. When applied to the reliable BS forecasts, this method preserves the reliability and reduces the mean variance of the CDF.

We construct CDF forecasts using the model performance (CAR), the BS objective, the PSS objective, and the TMSE nudging of the BS result (TM). The probability densities of the CDF forecasts are plotted in Figure 1. Also shown are the conditional climatology (CC), the CF forecast (dashed line), and the observed time of clearing (solid line).

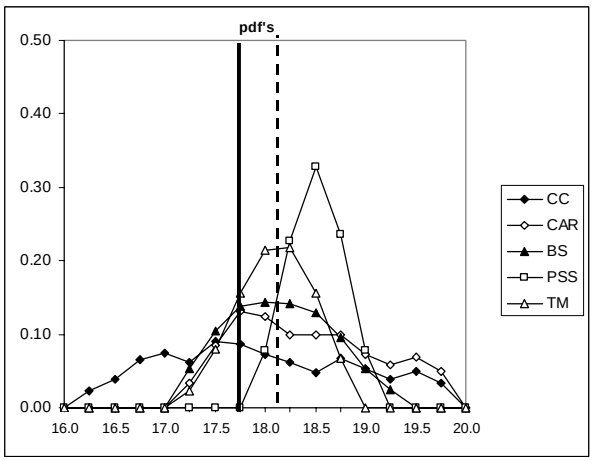

Figure 1. Probability densities of CDF forecasts.

Note that the CC forecast provides the poorest definition, that the CAR and BS forecasts have similar definition, and that the PSS and TM forecasts have the sharpest definition, but PSS has more bias. We observe this pattern throughout the training set.

4. APPLICATION TO AIR TRAFFIC

While skill metrics provide measures of technical success for developers, the practical measure of success is that the forecasts provide good service to their intended users, in this case, air traffic managers. One user, presently attempting to use these forecasts, has asked for the CAR as a measure of confidence. A proposed strategy is to make a commitment to cancel a program for the time when the CAR exceeds a threshold, e.g. 90%. Users have not discussed the possibility of using a different CDF forecast.

The current focus is the support of the Stochastic Ground Hold Plan (SGHP) model (Hoffman and Ball, 2000). This model determines the optimal strategy for holding some planes on the ground and releasing the rest to travel to the destination airport. Planes that are released prematurely will be subjected to airborne holding near the destination. Airborne holding presents additional costs to the airlines and air traffic management problems for Air Traffic Control (ATC). Planes belatedly released from the ground hold, will incur unnecessary additional delay. It is assumed that there is a cost difference between ground holding and airborne holding, with the latter being substantially greater. The objective function is the total cost, the product of cost-rate and holding times. The model determines the release strategy if clearing occurred in each category, and uses the forecasted probabilities to compute the expected associated cost. The plan for the release of ground held planes is determined by minimizing the total system cost.

Preliminary investigations indicate that smaller CDF variances are preferable, provided that the other factors are controlled. These investigations are ongoing. The initial (subjective) conclusion is that the BS-TM model provides the best guidance for use in the SGHP model.

Figure 2 provides the results when the forecasts for Figure 1 are input to the SGHP model. Three CDF forecasts are considered: CAR, BS, and TM. The vertical axis indicates the number of planes in the airborne queue at each quarter hour. More than eleven planes in the airborne queue places an excessive burden on ATC. The EAQ graphs indicate the expected airborne queue, according to the SGHP analysis. The AAQ graphs indicate the actual airborne queues, which would have been

realized under the various plans, based on the actual time of clearing. The Delay Panel lists the total aircraft delay hours that would have been realized under the various plans. The Worst and Best delay values are included to bracket the outcomes. The Worst case assumes that planes are released only after clearing is observed, and that the transit time is 1.25 hours. The Best time assumes that there was a perfect deterministic forecast, and that the traffic management decision uses this information effectively. The deterministic forecast was slightly late and this induced a slight late bias in the CDF's. Since there is some probability density prior to the time of clearing, SGHP does provide for some early releases. However, the early actual clearing provides welcome relief, and the airborne queues are not excessive. We note that the CAR plan expects some airborne queue, but none is realized, and it adds 10 a/c hours of delay. The BS plan expects the largest airborne queue, and its actual airborne queue is temporarily excessive. The BS plan adds only 4 a/c hours of delay. The TM plan adds only 5 a/c hours of delay, and has a much smaller airborne queue.

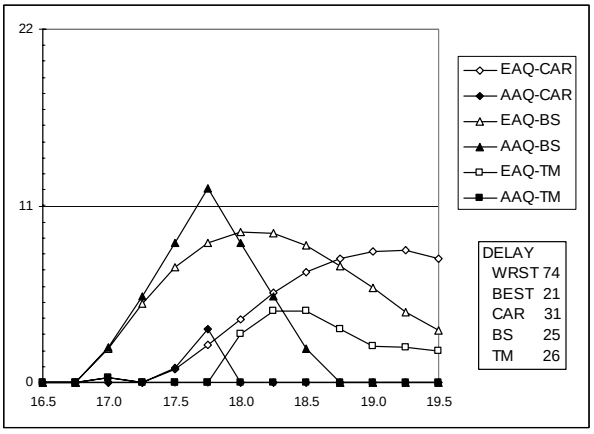

Figure 2. An example of the expected and actual airborne queues using different CDF forecasts as input for the SGHP model.

5. CONCLUSIONS

We have introduced a methodology for using the deterministic forecasts from the SFO Marine Stratus Initiative as predictors for developing probabilistic forecast models. There are options in the optimization objective, which leads to models with different forecast characteristics. When forecast skill is measured by PSS, the probabilistic forecasts have similar, but lesser skill than the underlying deterministic forecasts. In the cases under investigation, there seems to be a trade-off between reliability and skill in the sense of Pierce.

The TMSE provides an additional measure of skill. Practical considerations preclude direct TMSE optimization, but a nudging method has been developed, which permits TMSE improvement of BS optimized forecasts. These TM forecasts have qualities similar to the PSS optimized forecasts. Preliminary investigations indicate that the TM forecasts have a good synergy with the needs of the SGHP traffic management model.

6. REFERENCES

Allen, Rebecca L., 2001: Observational Data and MOS: The Challenges in Creating High-Quality Guidance. Conference on Weather Analysis and Forecasting, Ft Lauderdale, FL.

Brier, G.W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev*. **78**, 1-3.

Clark, D.A. and F.W. Wilson, 1997: The Marine Stratus Initiative at San Francisco International Airport. Seventh Conf. on Aviation, Range & Aerospace Meteorology, Long Beach, January 1997.

Darlington, R.B., 1990: *Regression and Linear Models,* McGraw Hill, NY.

Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky, 2000: Model Consensus. *Wea. and Forecasting* **15**, 571-582.

Glahn, H.R. and D.A. Lowery, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Met*. **16**, 672-682.

Hoffman, R. and M. O. Ball, 2000: A comparison of formulations for the single-airport, ground-holding problem using banking constraints, *Operations Research* **48**, 578-590.

Murphy, A.H., 1973: A new vector partition of the probability score. *J. Appl. Met.* **12**, 595-600.

Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science* **4**, 453-454.

Vislocky, R. and J. M. Fritsch, 1997: An automated, observations-based system for short-term prediction of ceiling and visibility. *Wea. and Forecasting*, **12**, 31-43.

Wilks, D. 1995: *Statistical Methods in the Atmospheric Sciences*, Academic Press, NY.

Wilson, F. W., and D.A. Clark, 2000: Forecast aids to lessen the impact of Marine Stratus on San Francisco International Airport. 9th Conference on Aviation, Range & Aerospace Meteorology, Orlando, FL.

Wilson, F. W., 2002: Additional measures of skill for probability forecasts. 10th Conference on Aviation, Range & Aerospace Meteorology, Portland, OR.