© Copyright 2000 American Meteorological Society (AMS). Permission to use figures, tables, and brief excerpts from this work in scientific and educational works is hereby granted provided that the source is acknowledged. Any use of material in this work that is determined to be "fair use" under Section 107 of the U.S. Copyright Act or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC §108, as revised by P.L. 94-553) does not require the AMS's permission. Republication, systematic reproduction, posting in electronic form on servers, or other uses of this material, except as exempted by the above statement, requires written permission or a license from the AMS. Additional details are provided in the AMS CopyrightPolicy, available on the AMS Web site located at (http://www.ametsoc.org/AMS) or from the AMS at 617-227-2425 or copyright@ametsoc.org.

Permission to place a copy of this work on this server has been provided by the AMS. The AMS does not guarantee that the copy provided here is an accurate copy of the published work.

5.1 FORECAST AIDS TO LESSEN THE IMPACT OF MARINE STRATUS ON SAN FRANCISCO INTERNATIONAL AIRPORT *

F. Wesley Wilson and David A. Clark
Massachusetts Institute of Technology, Lincoln Laboratory

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. Previous 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 (Wilson and Clark, 1997).

The San Francisco Marine Stratus Initiative has provided a four-year focus on this problem and has resulted in the development of several forecast algorithms that will aid the operational forecasting of the dissipation of marine stratus in the approach to SFO (Clark and Wilson, 1997). These algorithms involve new techniques for the analysis of observational data and statistical and dynamical prognosis of the behavior of the marine stratus. This discussion of the design and the performance of these algorithms provides an overview of the status of this project.

2. PRODUCT GOALS

The goal is to provide accurate, short-term forecasts of the time that there will be sufficient clearing of marine stratus in the approach to SFO to permit side-by-side landings on the closely-spaced parallel runways, which are separated by only 700 feet. This kind of operation is possible only when the pilots are able to maintain visual contact throughout the time when they are below 3000 feet. New radar technology may allow them to reduce this requirement to 1900 feet, which will provide substantial benefits in the winter months but will have less benefit in the summer when the stratus top is usually held below 2500 feet by a strong capping inversion. The project focus has been on the summer events, where the ceiling does not lift sufficiently to permit side-by-side operations and the dissipation occurs when there has been sufficient solar heating of the marine layer to evaporate the liquid cloud water.

After 1630Z, the scheduled arrival traffic rate exceeds the single-runway capacity of 30 planes per hour, and delay programs are imposed to manage the excess demand. Many of the delayed flights originate at airports which are less than one hour away, so the short-term forecasts can be used to release these flights to take advantage of the increase of the airport arrival capacity of 55 planes per hour, which will occur when the stratus clears in the approach. Due to this flight time, there is a delay of about an hour between the time that a traffic management decision is made to increase the number of flights departing to SFO and the actual increase in landings. In the worst case, where the traffic management decision is made reactively based on observed clearing, there could be a loss of 25 landing opportunities during the first hour after the clearing. Accurate short-term forecasts would provide the information that is needed to support proactive air traffic management decisions. The automated products under development are intended to aid the operational forecasters in providing these accurate, short-term forecasts.

3. OBSERVATIONAL DATA

The basic information to support these forecasts is provided by the National Weather Service (NWS), including standard products that are widely distributed. In addition to the synoptic structure and prognosis from various forecast models, the NWS observing systems provide timely regional information, including a local sounding (Oakland), GOES-9 data (15-minute), and regional surface observations, mostly hourly, but more frequently from a few key Automated

---

* This work was sponsored by the Federal Aviation Administration under Air Force contract F19628-95-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the U.S. Government. Corresponding author address: F. Wesley Wilson, Massachusetts Institute of Technology, Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420-9185. E-mail: wesw@ll.mit.edu

Surface Observing Station (ASOS) sites. An FAA Automated Weather Observing Station (AWOS) provides an important ceiling observation at the San Mateo bridge. The local WSR-88D radar (MUX) is located 50 km to the south at an elevation of 3500 feet and is restricted to scanning above the marine layer; it provides little useful information for the forecasting of summer stratus. Of these, the satellite data usually provides the forecaster with the most effective information about changes that may affect the short-term forecast.

In addition to the NWS information infrastructure, the project has added additional observational data to the project database. The most important of these provide improved sensing at and to the south of SFO. Sensing systems at SFO and at San Carlos airport measure the solar radiation with pyranometers, the height of the inversion base with incoherent Sodar, and the 10m temperature, dewpoint, and winds. These data provide important information about the evolution and changes in the local marine layer structure, which are important for providing the short-term marine stratus forecasts. A Bay Area map showing observing station locations is shown in figure 1. Included is the location of the base station data collection computer at San Jose State University and the primary display computer at the CWSU.

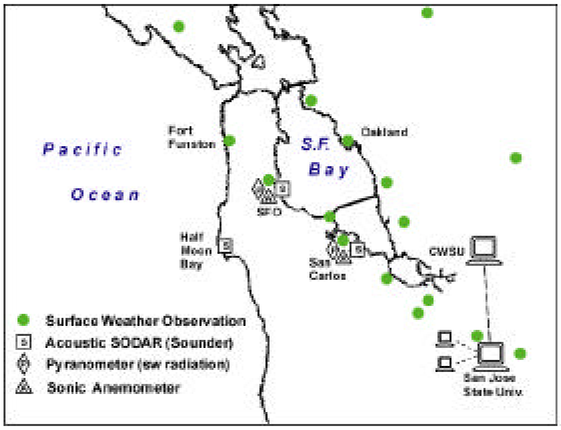

Figure 1. San Francisco Bay Area observing station locations.

4. THE FORECAST ALGORITHMS

Four forecast algorithms have been developed. A dynamical analysis of the heat budget, radiation, and cloud microphysics is provided by an extension of the COBEL (COuche Brouillard Eau Liquide) column model. Three statistical forecast techniques, which make use of the observational data, have also been developed:

1. RSFM – The Regional Statistical Forecast Model makes use of the Oakland sounding and regional hourly surface observations.
2. LSFM – The Local Statistical Forecast Model makes use of the Oakland sounding and the project sensors at SFO and San Carlos Airport.
3. SSFM – The Satellite Statistical Forecast Model makes use of the regional satellite data.

A consensus forecast product is being developed to provide a best forecast based on the results of these individual models.

COBEL was initially developed at the Laboratoire d'Aérologie, Université Paul Sabatier in Toulouse France, for the study of the nocturnal boundary layer (Estournel and Guédalia, 1987; Estournel, 1988). Its application to fog forecasting near Lille, represents an early success in automated fog forecasting (Bergot, T., and D. Guédalia, 1994; Guédalia, D. and T. Bergot, 1994). Extensions required for the marine stratus forecast include the additions of a surface parameterization, solar radiation, cloud microphysics, and a drizzle parameterization. An important innovation is the assimilation of the inversion base height, which is measured by the Sodar, to provide a response to a changing boundary layer height. These changes are known to have a substantial impact on the evolution of the marine stratus. A second innovation is the use of the measured solar radiation to estimate the cloud liquid water. COBEL is reinitialized at each hour from 14Z until clearing is observed at SFO, according to the pyranometer. The forecast of the time that the liquid cloud water evaporates is produced for each of these hourly forecasts.

The RSFM is a statistical forecast that is based on the regional hourly NWS observations. The restriction to NWS data is due to the need for a long-term data archive. RSFM equations are developed using General Additive Model (GAM) statistical techniques (Vislocky and Fritsch, 1995). The approach is a descendent of the Model Output Statistics (MOS) methodology (Dallavalle and Dagostaro, 1995), except that the predictors are taken from observational data in this application. The training data are taken from the 30-year archive of hourly surface observations and soundings in Central California. The initial predictor list includes all of the surface observation fields from the NWS surface stations and temperature and humidity information from the Oakland sounding. A sounding-based diagnosis of the height of the inversion base is included as a separate predictor. A separate forecast equation is generated for each combination of hourly forecast time and semi-hourly realization time. Since there is a large data archive, a large number of predictors can be considered. The derivation is randomized by using half of the data for training and half for evaluation. In this analysis, the optimization goal is the minimization of the Brier Score (Brier, 1950). In practice, the statistical analysis selected from 8 to 12 predictors for each forecast equation. The product is the probability of clearing before the realization time.

The LSFM is a statistical forecast model which is based on data that are local to the Bay area and mostly local to the SFO approach. The origin of this simple model is an empirical forecast aid that was developed by the operational forecasters. They have developed a scattergram forecast tool which has 14Z ceiling height and cloud thickness as its axis values and which plots the time-of-clearing in hours after sunrise for each event over several years. Subjective contours of the clearing times provide the basis for forecasts of the clearing time. The germ of this idea is that the height of the inversion base (ceiling plus cloud thickness) and the amount of cloud water (cloud thickness) are primary predictors for the time of burnoff. The LSFM has been developed using this intuition but by replacing statistical analysis for the subjective contouring, by using the inversion base height measured by the Sodar, and by allowing the possibility of additional local predictors. Due to case limitations, it is necessary to limit the number of predictors to no more than four and the forecast hours are restricted to 14Z - 17Z in the development of these equations. The derivation is randomized by using half of the data for training and half for evaluation. In this analysis, the optimization goal is the minimization of the error residual in the multi-linear regression. The product is the clearing time in hours past sunrise.

The SSFM is a statistical forecast model which is based on visible satellite data. The satellite data are preprocessed to remove the effects of sun slant angle and lens graying. The normalized values are clustered into geographical sectors of common meteorological or operational meaning, and sector statistics are computed (see figure 2). The statistics of interest are the percent coverage, the mean brightness, and the variance of the brightness in the ‘cloud’ pixels in the sector. This technique provides a satellite data product, which has a regional distribution similar to surface observations. Again, the data archive is limited in time, so it is necessary to use statistical techniques that are robust on smaller data sets. The number of predictors is limited to no more than four, and the forecast hours are restricted to 14Z - 17Z in the development of these equations. The derivation is randomized by using half of the data for training and half for evaluation. In this analysis, the optimization goal is the minimization of the error residual in the multi-linear regression. The product is the clearing time in hours past sunrise.

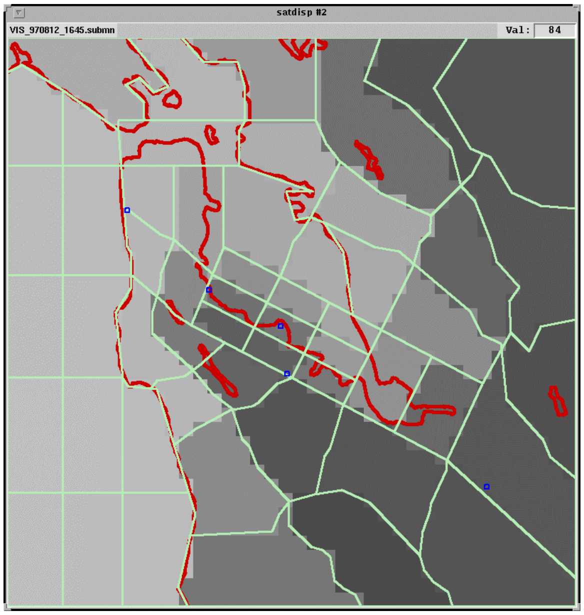

Figure 2. Map showing predefined sectors used for computation of satellite data statistics.

In the discussion, there is mention of two data analysis algorithms that deserve special attention due both to their complexity and to their potential applicability to other automation problems. The first is the Satellite Sector Product. This product converts satellite pixel data into satellite regional statistics after the normalization of the data to remove high frequency and low frequency variations. After this normalization, the data can be compared statistically over a period of several years. The second is the Sodar Z_i product, which analyzes the incoherent Sodar data to detect intensity peaks at each time step and to detect and follow atmospheric layers over time. This product shows skill far beyond that of a visual inspection. One weakness in this product is the distinction of the Z_i layer in the rare cases of multiple low layers. This technology has also been successfully applied to Profiler data for the detection of the convective layer.

5. SKILL STATISTICS

Statistical scoring metrics have been selected that have operational significance and have some meteorological tradition. These include the error distribution statistics Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Bayes Skill Statistic (BSS) (Krzysztofowicz, 1992), and the categorical skill statistics True Skill Score (TSS) (Peirce, 1884) (rediscovered by Hanssen and Kuipers, 1965, and Flueck, 1987), and a variant of the TSS which we introduce here, the Worth Ratio (WR).

Since these products are intended for operational use, it is important to measure their skill in terms of metrics that relate to their utility. The theorem of Krzysztofowicz asserts that if two forecast systems have different BSS scores, then the one with the smaller (better) score has more value to all users, provided that they make flawless use of the forecasts. And there is the hook! The perfect user is expected to know and apply all required bias and trend corrections to the forecasts. Another issue is that any linear transformation (such as bias correction) of the forecast values will leave the BSS unchanged. Thus, there is a family of forecasts providing decidedly different forecast values to the user, all of which have the same BSS. Out of the confusion comes a positive factor: there is freedom to modify the forecast output, by linear transformations, to optimize some other factor (such as removing bias) without degrading the BSS. All of this involves a posterior analysis of the forecast over some tuning period and thus has the spirit of another model output statistic.

The Worth Ratio is introduced as a candidate for this optimization measure. For categorical forecasts, it is necessary to make a distinction in the meaning of the benefit of a forecast. For an event space parsed into cases where an event occurred (w=T) and did not occur (w=F), the benefits formula has the form

$$E = P(w = T) * G - P(w = F) * L \qquad (1)$$

where G is the anticipated gain when the forecast verifies and L is the loss when the forecast fails to verify. In this application, the gain is reduced delay and the loss is additional airborne holding and diversions. Taking no action results in the status quo, E = 0. The scenario is that there are many days when the forecast might be applied, and on some days the forecast is positive and on other days it is not. The distinction is whether the benefit is averaged over all days, expected benefits per trial (E_T), or if it is just averaged over the days when an action is taken (E_A, the forecast is positive). Investigating an analogue to Krzysztofowicz results for categorical forecasts, the first author has been able to establish two results:

1. There is no skill statistic for categorical forecasts that has a monotone relation to E_T for all users.
2. WR has a monotone relation to E_A, i.e., a higher WR score corresponds to a higher E_A for all users.

Since linear transformations of the forecast do modify the WR, one can envision a program where posterior analysis is used to define the linear transformation of the forecast, which optimizes WR on the trial set. The optimizing transformation will maintain the BSS and will improve categorical skill in this sense.

The definition of the WR is straightforward. Details must be left to another presentation. Doswell, et al., 1990, show that

$$TSS = P(f=T \mid w = T) - P(f = T \mid w = F) \qquad (2)$$

We note that the forecast system has skill when TSS > 0. A mathematical alternative is to define

$$WR = P(f=T \mid w = T) / P(f = T \mid w = F) \qquad (3)$$

and to use the skill test WR > 1. While this is a simple mathematical reformulation, there are important consequences. First, WR is related to benefits for all users. Second, WR carries independent information in that all probability information from the forecast contingency table can be derived algebraically from the pair (TSS, WR). A typical practice is to use the pair (POD, FAR) (Wilks, 1995). Finally, if the event ratio is defined by $ER = P(w=T) / P(w=F)$ and the forecast confidence ratio is defined by $CR = P(w=T \mid f=T) / P(w=F \mid f = F)$, then an application of Bayes Formula (e.g., Wilks) shows that WR is related to the confidence that the user can have in a positive forecast by $CR = WR * ER$.

6. PERFORMANCE RESULTS

A principal scientific consideration in the selection of San Francisco for the product development site is the abundance of cases and repeatability of the phenomena. These considerations are paramount for the development of statistical models that depend on data collection within the project. In addition, they provide a database that can be used for substantive statistical evaluation of the product performance.

The results presented here are based on 141 summer marine stratus cases from 1996-1997. The results presented are typical of the general class that have been obtained in the project. Only a few examples can be presented here, as shown in figures 3 through 6.

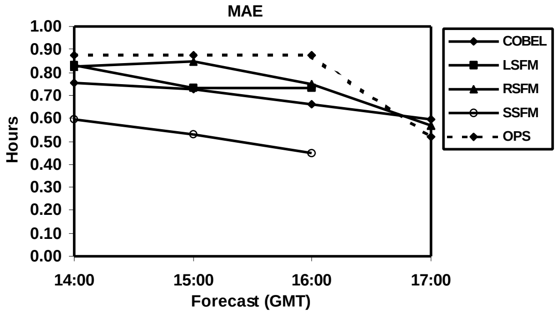

Figure 3. Comparison of the MAE for all forecast systems (based on '96-'97 data), plotted against the time that the forecast is issued. The MAE for operational forecasts issued at 14Z and 17Z are plotted for comparison. Lower MAE values reflect better forecast skill.

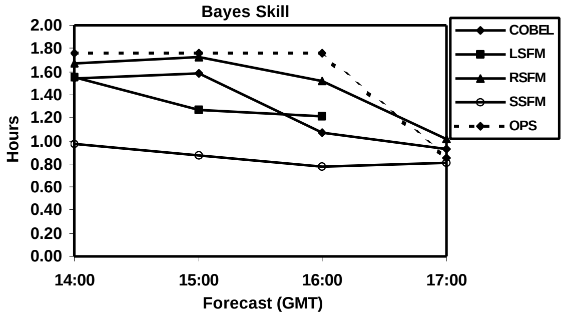

Figure 4. Comparison of the BSS for all forecast systems (based on '96-'97 data), plotted against the time that the forecast is issued. The BSS for operational forecasts issued at 14Z and 17Z are plotted for comparison. Lower BSS values reflect better forecast skill.

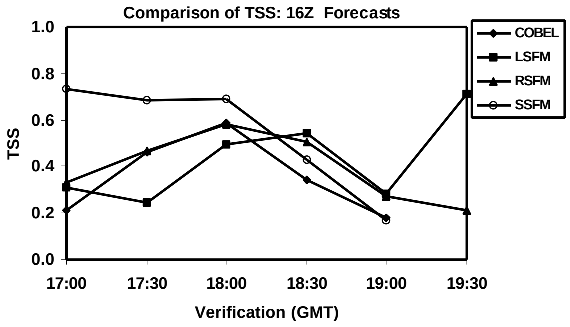

Figure 5. Comparison of the TSS for 16Z forecasts (based on '96-'97 data), plotted against the realization time for the forecast. Higher TSS values reflect better forecast skill.

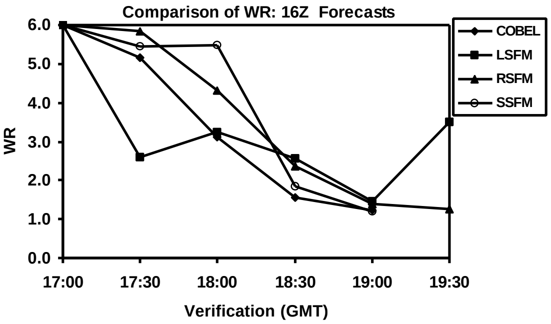

Figure 6. Comparison of the WR for 16Z forecasts (based on '96-'97 data), plotted against the realization time for the forecast. Higher WR values reflect better forecast skill.

7. CONCLUSIONS

Several automated systems have been developed for forecasts of the clearing of marine stratus in the approach to SFO. Each of these systems shows evidence of skill when compared with the performance of operational forecasts. Skill measures that reflect the utility of the forecasts have been introduced, and these provide evidence that these forecasts will have value in an operational setting.

8. REFERENCES

Bergot, T., and D. Guédalia, 1994: "Numerical Forecasting of Radiation Fog. Part I: Numerical Model and Sensitivity Tests." *Mon. Wea. Rev.*, **122**, pp.1218-1230.

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 San Francisco Marine Stratus Initiative." Amer. Meteor. Soc., *7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, CA, pp. 384-389.

Dallavalle, J.P., and V.J. Dagostaro, 1995: "The Accuracy of Ceiling and Visibility Forecasts Produced by the National Weather Service." Preprints, Amer. Meteor. Soc., *Sixth Conf. on Aviation Weather Systems,* Dallas, TX, pp. 213-218.

Doswell, C.A., R. Davies-Jones, and D.L. Keller, 1990: "On Summary Measures of Skill in Rare Event Forecasting Based on Contingency Tables." *Weather and Forecasting*, **5**, pp. 576-585.

Estournel C., and D. Guédalia, 1987: "A New Parameterization of Eddy Diffusivities for Nocturnal Boundary Layer Modeling." *Bound. Lay. Meteor.*, **39**. pp.191-203.

Estournel, C., 1988: *Étude de la Phase Nocturne de la Couche Limite Atmosphérique*. Thèse de Doctorat d'État, Université Paul Sabatier, Toulouse, France, No. d'ordre 1361, 176 pp.

Flueck, J. A., 1987: "A Study of Some Measures of Forecast Verification." Amer. Meteor. Soc., *Preprints, Tenth conference on Probability and Statistics in the Atmospheric Sciences*, pp. 69-73.

Guédalia, D. and T. Bergot, 1994: "Numerical Forecasting of Radiation Fog. Part II: A Comparison of Model Simulations with Several Observed Fog Events." *Mon. Wea. Rev.,* **122**, pp.1231-1246.

Hanssen, A. W. and W. J. A. Kuipers, 1965: "On the Relationship between the Frequency of Rain and Various Meteorological Parameters." De Bilt, Royal Netherlands Meteorological Institute, Mededeelingen en Verhandelingen, No. 81, 65 pp.

Krzysztofowicz, R., 1992: "Bayesian Correlation Score: a Utilitarian Measure of Forecast Skill." *Mon. Wea. Rev.,* **120,** pp. 208-219.

Peirce, C. S., 1884: "The Numerical Measure of the Success of Predictions." *Science*, **4**, pp. 453-454.

Vislocky, R.V., and J.M. Fritsch, 1995: "Generalized Additive Models versus Linear Regression in Generating Probabilistic MOS Forecasts of Aviation Weather Parameters." *Weather and Forecasting*, **10**, pp. 669-680.

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

Wilson, F. W. and D. A. Clark, 1997: "Estimation of the Cost of Ceiling and Visibility Events at Major Airports." Amer. Meteor. Soc., *7th Conference on Aviation, Range, and Aerospace Meteorology,* Long Beach, CA, pp. 480-485.