

FORECAST-BASED DECISION SUPPORT FOR SAN FRANCISCO INTERNATIONAL AIRPORT

A NextGen Prototype System That Improves Operations during Summer Stratus Season

BY DAVID W. REYNOLDS, DAVID A. CLARK, F. WESLEY WILSON, AND LARA COOK

Improved forecasts of clearing time of low clouds over the approach to San Francisco International Airport reduces aircraft arrival delays and provides a substantial monetary savings to the airlines.

In an effort to streamline air traffic efficiency in the National Air Space (NAS), the Federal Aviation Administration (FAA) Aviation Weather Research Program (AWRP) funded an initiative led by the Massachusetts Institute of Technology (MIT) Lincoln Laboratory to improve the forecast of summer stratus that impacts operations into San Francisco International Airport (SFO). SFO has two pairs of closely spaced parallel runways that require visual conditions (3,000-ft ceilings and above and greater than 5 miles visibility) to perform dual approaches to maximize arrival throughput at a nominal rate of 45 to a maximum of 60 aircraft per hour. The presence of low-ceiling stratus in the approach zone precludes the

dual-parallel approach procedure, reducing the airport's arrival capacity to 30 aircraft per hour. SFO is chronically one of the highest delay airports (Fig. 1) in the NAS and is a major hub for both intra- and

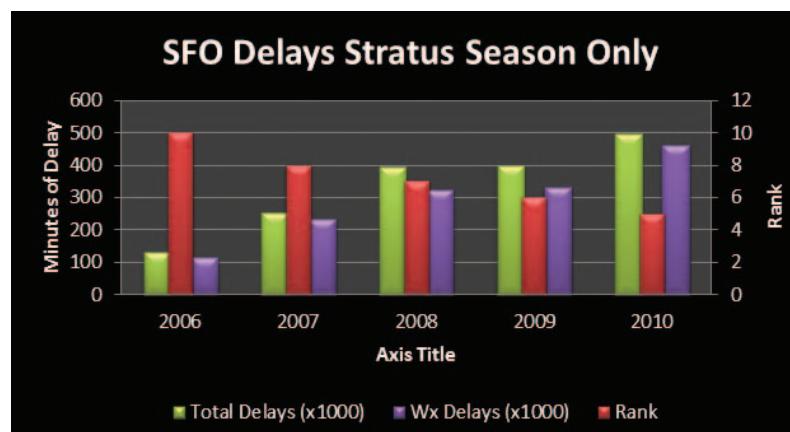


FIG. 1. SFO total minutes of delay, minutes of weather delay, and rank in weather delays for the top 35 busiest airports, during the summer stratus season 15 May–15 Oct for the period 2006–10. Note that each season has shown an increase in total minutes of delay and weather delays, and SFO's rank is in the top 5 for weather delays for 2010. (Provided by the FAA OPSNET system.)

intercontinental air traffic, so a disruption in flow has a wide-ranging impact. Roughly half of SFO's air traffic delay is attributable to summer stratus. The FAA-sponsored program led to the development of a prototype Marine Stratus Forecast System (MSFS) intended to improve the daily forecast of stratus clearing to help traffic managers more efficiently manage arrival demand with available airport capacity. The system was demonstrated operationally in 2001–04 and has since been managed by the National Weather Service Forecast Office (NWSFO) in Monterey. A 2008 report by the NWS (Delman et al. 2008) investigated the performance of the system, both in terms of forecast skill and impact on reduction of delay. It was concluded that the MSFS performed reliably to its expected skill level, yet there was a negligible reduction in aircraft delay attributable to the new forecast guidance system.

This discrepancy was reported by Clark (2009), who suggested that an improvement to the system was required that went beyond the weather component of the problem and included a solution that recognized a deficiency in the air traffic management decision-making process. Specifically, an independent effort was referenced that had already been initiated by Mosaic ATM, Inc., under a funding grant by the National Aeronautics and Space Administration (NASA) Ames Research Center. Mosaic ATM had been investigating a Monte Carlo approach to derive an optimal air traffic flow decision during SFO stratus conditions, weighing operational flow risk against potential delay reduction, using the MSFS's known error distribution characteristics with dynamic arrival demand information. They reported a proposed solution that could be implemented in real time and presented anticipated potential benefits derived analytically (Cook and Wood 2009). FAA System Operations agreed to fund development of

their methodology and integrate it into the existing forecast system, with an operational demonstration planned for the 2012 summer stratus season. Because this is a work in progress and has not been fully vetted in actual operations, the Ground Delay Parameters Selection Model (GPSM) is only briefly discussed in this article. However, this effort does serve as a significant step toward the FAA's Next Generation (NextGen) plan for future air traffic management, which will rely on translation of automated probabilistic weather information forecasts into the traffic flow strategy decisions.

One issue in this regard is the role of the forecaster in providing expert oversight in order to assure beneficial implementation. This is a recurring theme since automation is considered central to the NextGen concept, with the term "forecaster over-the-loop" coined (www.ral.ucar.edu/aap/themes/fotl.php) to describe a critical role for the forecaster in both quality control of the output of an automated system and value adding to the guidance output. Within the context of the MSFS and GPSM products, forecasters are expected to assume the role of relating to traffic managers any suspected deviation from the system forecast that would impact the integrity of the automated recommendations. This includes identification of days when the clouds are not the result of the "typical" stratus conditions, for which the automated forecasts are not expected to perform well. Forecasters have been working with the MSFS for a number of years; coupled with their expertise, they will take on the responsibility of suggesting any adjustments to the automatically generated GPSM recommendations (to either a more conservative or more aggressive approach, or to disregard them entirely). This is consistent with how they currently apply the MSFS guidance in generating their own forecasts of transitioning to dual runway operations.

The following sections outline the rationale for developing the MSFS system, the key components of the MSFS, verification of the automated guidance with a comparison to manual forecasts for all days where a ground delay program was initiated, the estimated benefits of using forecast to reduce delay, and a brief description of the GPSM.

THE IMPACT OF STRATUS ON OPERATIONS AND AIR TRAFFIC FLOW PLANNING.

A stratus cloud deck below 3,000 feet in the approach zone prevents dual approaches to SFO's closely spaced parallel runways (Fig. 2). In practice, for runways 28L/28R, side-by-side approaches are not started until the ceiling reaches at least 3,500 feet.

AFFILIATIONS: REYNOLDS*—National Weather Service, Monterey, California; CLARK—MIT Lincoln Laboratory, Lexington, Massachusetts; WILSON*—MIT Lincoln Laboratory, Lexington, Massachusetts, and National Center for Atmospheric Research, Boulder, Colorado; COOK—Mosaic ATM, Inc., Leesburg, Virginia
*Retired

CORRESPONDING AUTHOR: David W. Reynolds, 304 Mar Vista Dr., Monterey, CA 93940
E-mail: david.reynolds@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-11-00038.1

In final form 21 February 2012

This effectively cuts the airport's arrival capacity in half, from the maximum of 60 to 30 planes per hour. During the warm season (May–October), stratus forms and dissipates on a daily cycle in response to marine air advection and radiative cooling and heating, impacting operations on approximately 50–60 days each year (derived from the number of GDPs issued between 15 May and 15 October 2006–10). It should be noted that there are days during this period when low clouds impact the approach to SFO but are strongly forced by upper-level features, such as a trough just off the west coast, rather than the more typical daily thermal circulation associated with the summer season stratus.

The MSFS has a user entry option that allows the NWS Center Weather Service Unit (CWSU) forecaster located at the Oakland Air Route Traffic Control Center (ARTCC) to identify these as "not stratus" days. This is done to notify other MSFS users that the automated forecasts should be discounted, since they were not developed to accommodate this type of weather situation. There may be as many as 75–100 days per season when the MSFS issues a forecast due to clouds in the approach but, as stated above, only 50–60 qualify as days for which the system was designed.

The stratus typically dissipates from the approach zone sometime between midmorning and early afternoon, roughly coinciding with the morning arrival push of aircraft into SFO (Fig. 3). When stratus is present in the approach zone during the early morning and expected to persist, traffic managers may implement a Ground Delay Program (GDP) by holding a portion of upstream aircraft on the ground to reduce the flow of incoming traffic during the period of anticipated reduced capacity. This significantly reduces the risk of excessive airborne holding and diversions that would result from an extended period of demand exceeding capacity. The operational cost of this coping mechanism is that upon stratus clearing, there is a period of wasted arrival capacity while the expected pipeline of aircraft is being filled following release of ground-held planes (Fig. 4a; MCR Federal, Inc. 2004). Figure 4b shows the annual costs associated with ground delays at SFO. You will note

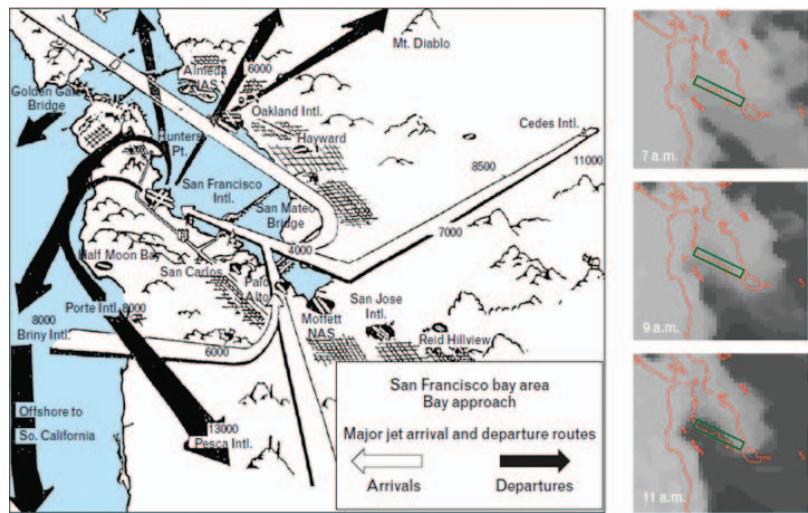


FIG. 2. (left) Primary approach and departure routes into and out of SFO [adapted from Strach (1991)]. During visible approaches, incoming aircraft arrive in parallel, merging at southern end of San Francisco Bay. (right) Typical evolution of stratus cloud dissipation in the Bay region. The green rectangle approximates the approach zone that must be free of low stratus to allow parallel approaches.

that there has been an increase in the minutes of delay due to an increase in the number of arrivals into SFO. This is mainly due to additional commercial carriers coming into SFO over the period shown, which has significantly increased the arrival demand.

Forecasting responsibility for anticipating the time of stratus clearing is shared by the CWSU, the aviation forecasting desk of the NWSFO in Monterey, and the operations centers of major commercial airlines with significant market share in SFO. Their forecasts are used by traffic managers at the Oakland ARTCC and at the FAA Air Traffic Control System Command Center (ATCSCC) to determine the duration (start time and end time), airport acceptance rate (AAR), and scope (number of planes impacted based on geographic proximity) for a proposed GDP. The GDP parameters are arrived upon as a collaborative process via a conference call facilitated by the ATCSCC with the Oakland ARTCC held each morning at ~1215 UTC (5:15 a.m. PDT). Input is provided by the CWSU forecaster, the United Airlines (UAL) forecaster, and occasionally by other commercial airlines. The conceptual process for making GDP decisions is shown in Fig. 5 and shows two-way interactions with all of the responsible parties. Development of the forecast guidance system was intended to support this decision process and subsequent modifications to the GDP as the situation evolves during the morning hours. Ultimately, the final decision is made and implemented by the FAA ATCSCC.

DEVELOPMENT OF THE MSFS MARINE STRATUS FORECAST SYSTEM. Marine stratus is advected into San Francisco Bay during the overnight hours, induced by the sea breeze circulation caused by strong heating of the interior valley during the previous afternoon. The marine stratus is trapped from the top by the marine inversion layer, and on either side of the Bay by the higher terrain associated with the Oakland East Bay Hills and the coastal hills running north-south in San Mateo County, consistent with a typical marine layer depth of 1,500 ft. Figure 6 presents the underlying physical processes impacting stratus dissipation. As the sun rises, solar radiation is transmitted through the stratus layer heating the underlying surfaces, especially the higher terrain on the eastern and western sides of the Bay, which in turn raises the potential temperature of the entire surface layer, raising the height of the zero dew point depression and thus the height of the cloud base. This can be seen in the time-height plot in Fig. 7, which shows the ceiling height lifting with time (green dashed) toward the top of the cloud deck (red dashed). Sensors and

data acquisition to support the forecast system were chosen to monitor the heat budget and to track the physical evolution of the marine stratus on both the local and regional scale (Clark 2002). The location of these key observation components in and around the approach to SFO are shown in Fig. 8. These observations feed a set of four forecast models that contribute independently to a consensus forecast of the time that the arrival rate will transition from a single runway to dual runway operations. At this point, pilots within the approach can visually see each other as they approach the parallel runways. This means the rate of landings increases to at least 45 aircraft per hour. In the context of forecast model development and performance evaluation, the time of transition to at least a 45 rate (hereafter, “45 rate”) will be used to mean a rate of 45 aircraft per hour) was used for validation. Although the acceptance rate can rise to as high as 60 aircraft per hour, a declaration of a 45 rate means the approach is clear and aircraft can land side by side, which represents an opportunity for an air traffic management decision to utilize newly available capacity.

The key new observations required to support the forecast system include the height of the marine inversion base (stratus cloud top), which is observed using two sonic detection and ranging instruments (sodars), ceilometers used to measure cloud base, pyranometers used to measure the incoming solar radiation, and high-resolution observations of surface temperature, dewpoint, and wind and their fluxes used to run a one-dimensional cloud model. These were combined with the existing twice per day upper air soundings from Oakland airport (OAK), just across the Bay from SFO, as well as existing hourly regional surface observations. Finally, after sunrise, high-resolution visible satellite imagery is used to forecast clearing time based on the anisotropic reflectance and distribution of the stratus within the Bay.

These data are transmitted in real time to a workstation running at the CWSU collocated

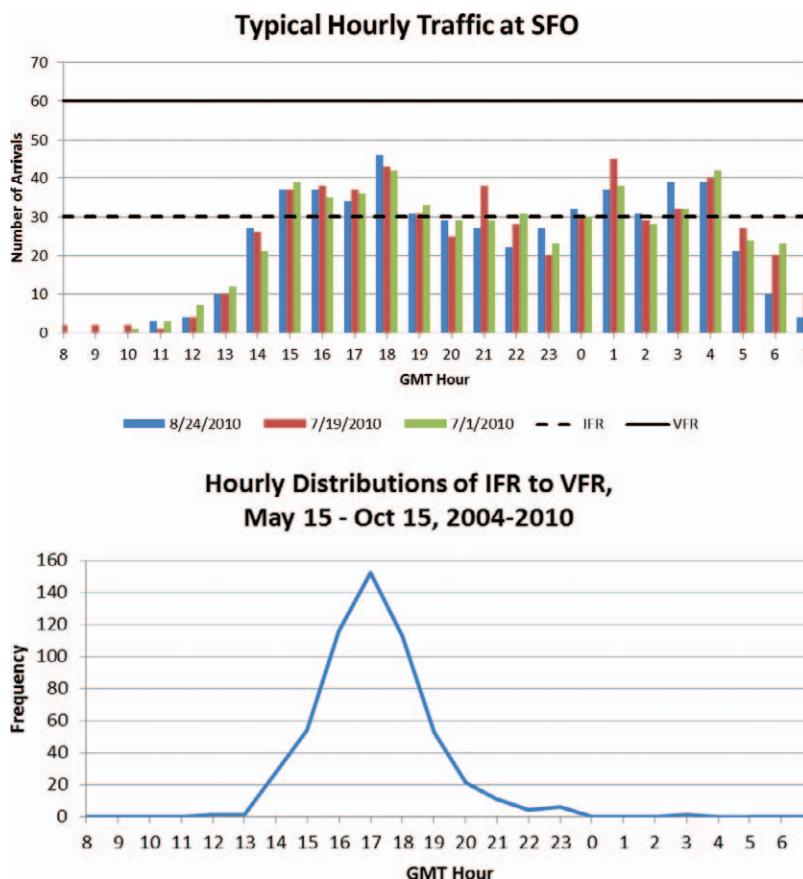


FIG. 3. Climatology of arrival rate versus transition from IFR to VFR conditions at SFO.

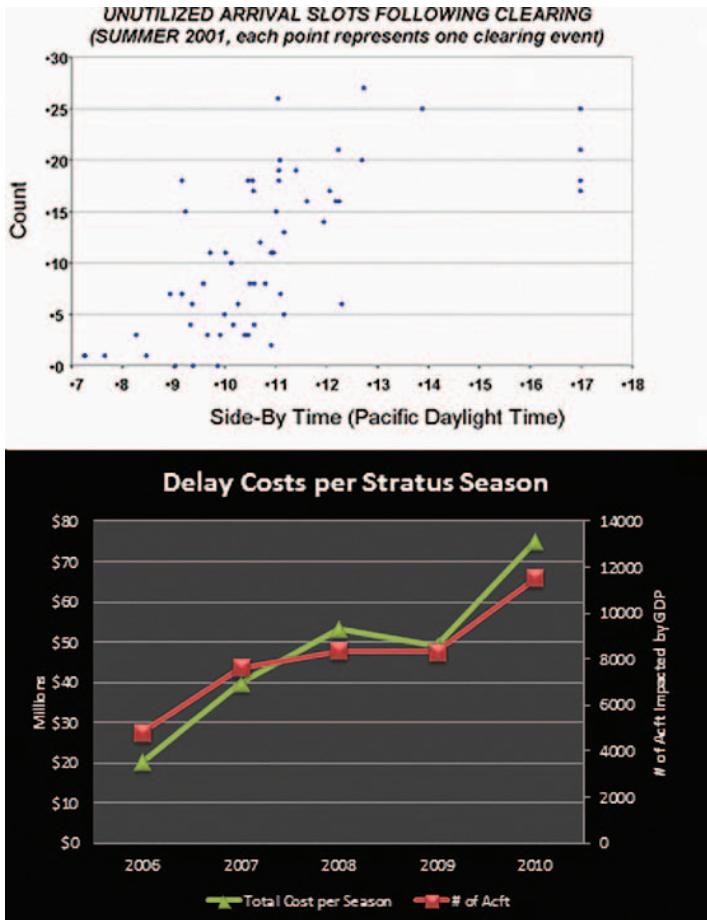


FIG. 4. (a) Unutilized arrival slots following stratus clearing, SFO summer 2001. (b) Annual cost and number of aircraft impacted during ground delay programs at SFO during the summer stratus season.

with the Oakland ARTCC (Fig. 9). A web-based situational display provides forecasters and decision makers a way to monitor the observations, individual and consensus forecasts, probabilistic forecasts, and manual forecasts and reasoning of the CWSU forecasters (Fig. 10), with all forecast information displayed in the upper right portion of the display. Forecasts from the four independent forecast models are combined to generate a single consensus forecast of the time that a 45 rate will be declared.

The component forecast models. One of the four component forecasts is derived from a physics-based numerical weather prediction model, the Couche Brouillard Eau Liquide (COBEL) model. The three remaining models were developed via nonlinear statistical regression:

the Regional Statistical Forecast Model (RSFM), the Local Statistical Forecast Model (LSFM), and the Satellite Statistical Forecast Model (SSFM).

The COBEL model is aimed at analyzing heat budget, radiation, and cloud microphysics. It is a high-resolution, one-dimensional numerical model of the planetary boundary layer (PBL) that simulates the life cycle of stratus dissipation at a specific location. In addition, COBEL uses sodar data to track changes in the height of the inversion, and solar radiation measurements to estimate cloud liquid water content (LWC). The model requires a specific set of initial conditions in order to run. These include the existence of stratus before sunrise and a strong inversion free of clouds above it, capping a well-mixed marine boundary layer (Clark et al. 2006). Again these are the typical conditions associated with the marine stratus that impact SFO operations.

The MSFS version of the COBEL model is initialized using an adaptation of the OAK sounding through a complex procedure, which involves an assumption that the vertical profile of temperature, humidity, and wind conditions at OAK are equivalent to SFO above the boundary layer (Clark et al. 2006). The procedure interpolates the sounding at lower heights down to the surface using the high-resolu-

tion data collected at SFO. COBEL is hindered significantly when upper- and midlevel clouds are present above the boundary layer. Since the model assumes zero horizontal advection, advection of temperature

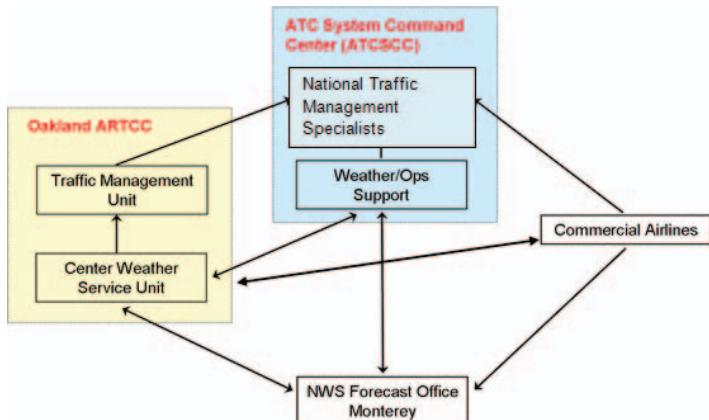


FIG. 5. Information sharing for GDP implementation strategy decision [adapted from Strach (1991)].

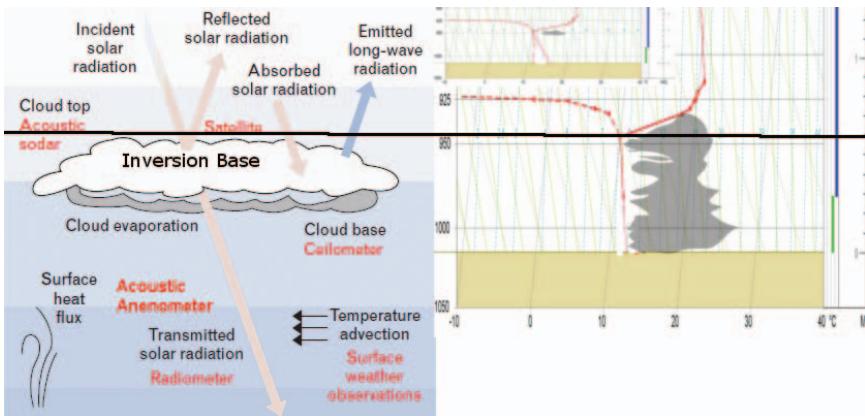


FIG. 6. (left) primary physical processes associated with stratus cloud evaporation. The key sensors associated with the various processes are shown in red. (right) vertical profile of temperature (solid) and moisture (dashed) representative of the marine stratus cloud. The inset shows how the profile change as evaporation continues from below. The base of the inversion stays constant as can be seen in Fig. 7. (Used with permission of Prof. John Monteverdi.)

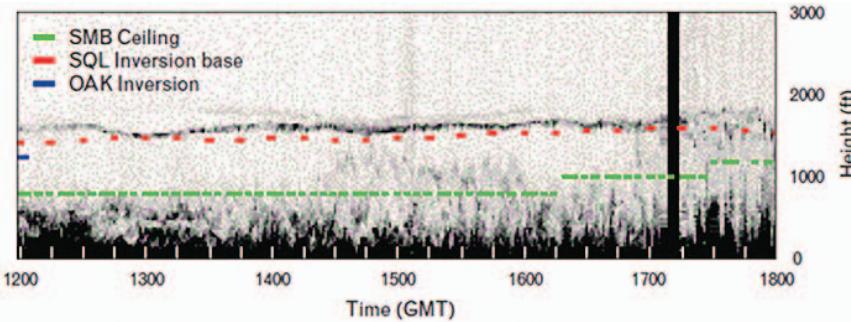
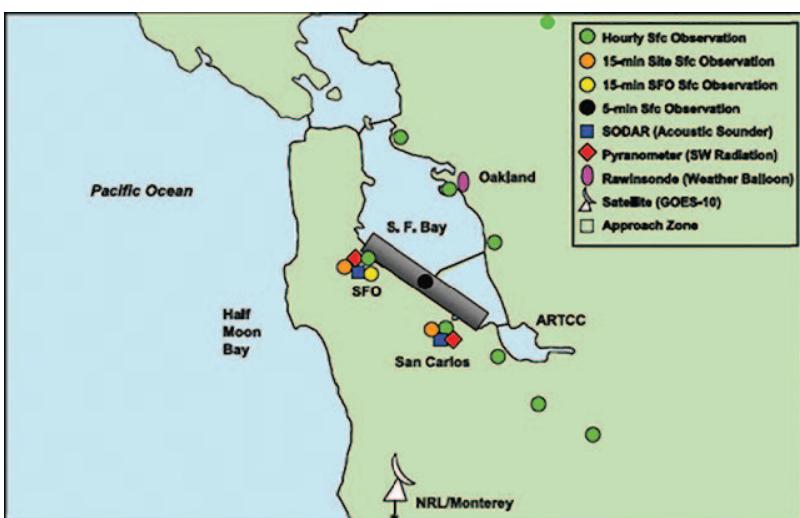


FIG. 7. Time-height profile showing decreasing depth of cloud deck with time. The quasi-horizontal black line at 1,500 ft is a signal from the sodar, indicating cloud-top height (at atmospheric inversion base). The red dashed line is the automated stable estimate of height, generated from sodar signals at 15-min intervals, for input into forecast algorithms. The dashed green line is the cloud-base height measured by a surface-based ceilometer. SMB indicates the San Mateo bridge, SQL the San Carlos airport, and OAK the Oakland airport (see Fig. 8).



and humidity during the cloud dissipation process also limits the accuracy of the model (Fidalgo et al. 2002) and would be categorized as nontypical stratus conditions. As part of the hour-to-hour sequence of COBEL model runs, model error in temperature is primarily attributed to unaccounted for advection, and this residual difference is then included in subsequent model runs. To aid the MSFS, solar radiation, cloud microphysics, and drizzle parameterizations were added to the model before its implementation for the SFO stratus application (Wilson and Clark 2000).

The remaining three forecast models were developed based on nonlinear statistical regression. Each was developed using the time of transition from single to dual approach procedures (i.e., when the 45 rate is initiated) as the predictand, rather than a specific meteorological occurrence (e.g., cloud cover at a specific location). The reason for this choice is that it is the operationally significant issue that impacts arrival throughput, and it is verified daily by professional pilots using their judgment as to whether they can see the airport at the beginning of their approach. A structured statistical development approach was applied, which started with a large list of potential predictors that underwent a nulling process

FIG. 8. Equipment layout within the approach to SFO and data sources used from outside the approach.

to eliminate redundancies and reduce each model to a manageable set of predictors consistent with the available sample size. The historical database was then subdivided into “day types” to further isolate the effectiveness of individual predictors under various meteorological regimes (e.g., onshore versus offshore flow, etc.). The methodology employs a nonlinear, monotone rescaling of each predictor value to optimize its correlation with the predictand. These rescaled predictors are used to build forecast models using traditional multiple linear regression with cross validation.

The RSFM is a statistical model that relies primarily on NWS observational data. This model employs the use of a 30-yr archive of hourly surface observations and soundings from the County Warning Area (CWA). In addition, a history of when a 45 rate was initiated following the appearance of the first stratus cloud the night before, and the surface pressure difference between SFO and Arcata/Eureka Airport (ACV) are taken into account (Wilson 2004). The RSFM uses these factors to solve forecast equations, which predict the time that a 45 rate would be initiated. Its use of multiple grid points over a relatively large regional scale enables the model to distinguish between offshore and onshore flow and capture the regional forcing that is not captured by the other component models,

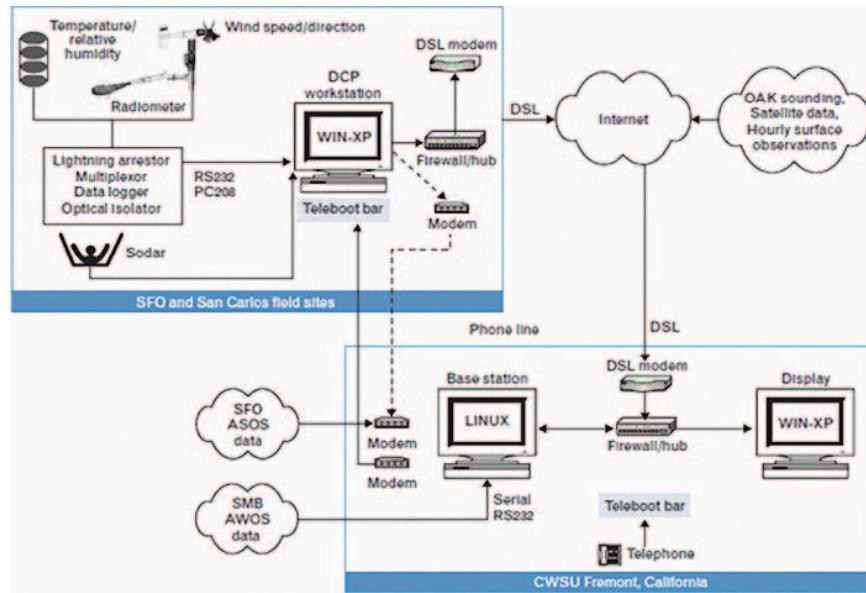


FIG. 9. MSFS data collection and data flow to the CWSU located at Oakland Center. The web page is hosted by the same machine that collects and generates the forecasts.

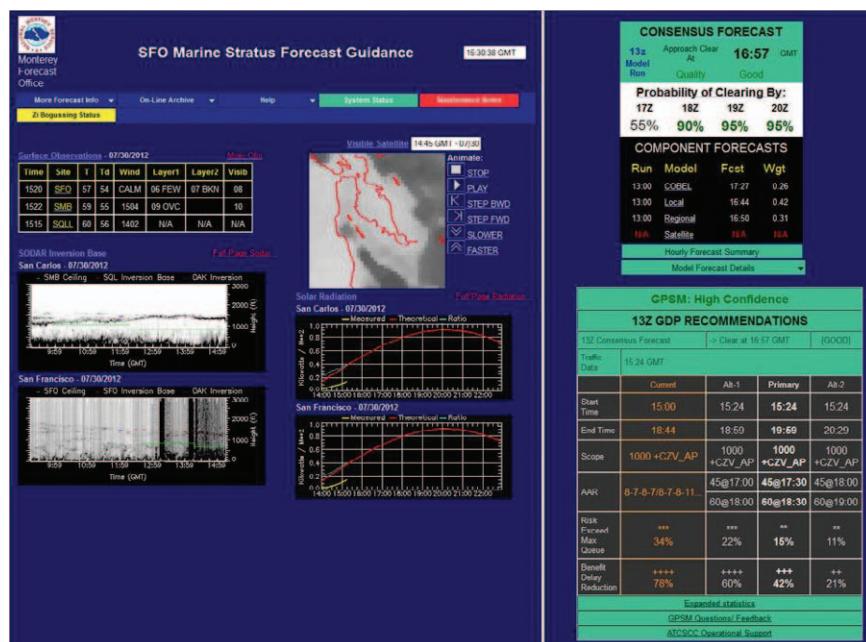


FIG. 10. Web-based MSFS observation and forecast guidance display. On the right side from top to bottom: Consensus forecast 45 rate at 1700 UTC, probability of clearing at indicated times, individual model deterministic 45 rate forecast times, and GPBM prototype display as it appeared in summer 2011. Left side of display starting from upper left is most recent 5-min Automated Surface Observing System (ASOS) and Automated Weather Observing System (AWOS) data from SFO and SMB, and temperature, dewpoint, and wind from SFO and SQL surface sites installed for this project. The sodar plots for SMB and SFO are shown under the surface data. To the right of the sodar plots are the incoming solar radiation plots (yellow trace) and the computed clear-sky amount (red) for the date. Just above these is the satellite imagery for the SFO sector used in the satellite algorithm.

and the extensive archive of NWS observations provides for more statistically stable output.

The LSFM is also a statistical forecast model. It uses observations local to the Bay Area, particularly the data from the sodars and radiometers from SFO and San Carlos Airport (SQL) that were specifically deployed for this project. This model uses historical trends in height of the inversion base, cloud layer heights, and surface wind to forecast transition to a 45 rate. As with the RSFM, this model relies heavily on historical observation data and may experience increased forecast error due to missing or incomplete records (Clark et al. 2006).

The SSFM is the third statistical forecast model, which uses 1-km visible Geostationary Operational Environmental Satellite (GOES) imagery centered on the approach zone. The data are normalized to reduce anisotropic reflectance due to the varying sun angle. This allows the satellite data to isolate brightness changes that are directly attributable to changes within the cloud layer. For statistical analysis, the region is divided into several dozen geographically homogenous “sectors,” with each sector treated as an observation point. The SSFM takes into account percent cloud coverage, mean brightness, and variance of brightness in each sector. Historical data from each sector are correlated with the time a 45 rate was initiated to forecast a probability of the time a 45 rate will begin at SFO.

The consensus forecast. There is significant statistical independence among these four forecast models, since there is limited overlap in their predictors. The consensus forecast combines these independent forecasts to provide a single deterministic forecast of the time that a 45 rate will be initiated. It is computed as a weighted average of the component forecasts. The weights are derived from the historical performance of each component model, evaluated separately for each of their respective day types, and for each individual initialization hour. The consensus forecast is accompanied by a confidence indicator. This indicator is designed to allow identification of conditions for which the consensus forecast performance is expected to be less reliable. Under these conditions, the forecast confidence is indicated as “LOW.” Otherwise, it is indicated as “GOOD.” Days classified as nontypical by the CWSU forecaster are most likely to have a LOW confidence indicator. Any one of the following four conditions will trigger the LOW confidence indicator:

- 1) The inversion base height is not clearly identifiable.
- 2) There is an extraordinarily high cloud ceiling base ($>2,300$ ft).

- 3) The cloudiness in the Bay area appears disorganized or patchy (determined automatically by examining the variance in brightness values in the sample domain), indicating a transient weather system rather than typical stratus.
- 4) Fewer than three component models are available.

Subsequent to initial system development, traffic managers at the ARTCC indicated that the deterministic forecast was difficult to translate directly to a traffic management decision without having more information about the certainty of the forecast. It was suggested that the deterministic forecast be converted to a probabilistic representation to indicate the likelihood that a 45 rate will have occurred by key target times throughout the morning arrival push. This modification was made to the system by empirically deriving the probability that a 45 rate will have been initiated by the top of each hour from 1700 through 2000 UTC (10:00 a.m. through 1:00 p.m. PDT).

From mid-May through mid-October, all of the models, with the exception of the satellite model, which requires visible satellite imagery, are run beginning at 0900 UTC (2:00 a.m. PDT) each day, then updated every 2 h at 1100, 1300, and 1500 UTC, and then run every hour until 1800 UTC (1:00 p.m. PDT). The consensus forecast provides input to all manually generated forecasts (e.g., those provided by the CWSU) and thus its accuracy is critical to these subsequent manually derived forecasts. Other tools used by forecasters include output from the Weather Research and Forecasting (WRF) model; soundings from OAK; visible, infrared, and water vapor satellite imagery; surface observations from coastal Meteorological Aerodrome Report (METARs) and offshore buoy sites throughout the CWA; web cams installed for monitoring of the stratus in the approach area; and the forecasters’ specific knowledge of Bay Area weather (Clark et al. 2006). FAA personnel rely on the accuracy of these manual forecasts when planning the scope and duration of the GDP. As stated, the GDP on a majority of stratus days should be issued by 1330 UTC or 6:30 a.m. PDT. In addition to the 1100 UTC consensus forecast, additional human input is drawn from CWSU forecasts, the 1200 UTC TAF for SFO issued by NWS forecasters in Monterey, and UAL personnel. Note that the 1100 UTC (4:00 a.m. PDT) consensus forecasts would only consist of 3 of 4 models as the SSFS model requires visible satellite imagery, which is not reliably available until 1500 UTC (8:00 a.m. PDT).

MODEL AND FORECASTER PERFORMANCE: BENEFITS OF AN ACCURATE FORECAST.

Forecast model performance. Development of the forecast models was an iterative process that began in 2000, using a training dataset that dated back to the initial data collection phase of the project which began in 1996. Models were developed for use as real-time forecast guidance during the summer of 2001. During 2001–02, four different trial versions of the three statistical models were examined. During the winter of 2002/03, the final model versions were established, and essentially run unmodified during the two summers of 2003–04. (The only exception was the satellite model, where the detection of a processing error in the raw data needed to be corrected.) Thus, the final version of the statistical models was developed during the winter

of 2002/03, using a training dataset of stratus days from 1996–2002.

An estimate of the expected performance of each of the models at each initialization time is shown in the left column of Table 1, labeled “Development.” The table presents three sets of statistics. First is the median absolute error (MAE; in minutes) of each model for each model run hour, including the consensus forecast. The second set of statistics shows the bias of each of the errors, where a positive bias indicates that the forecast time was later than the actual verification time (i.e., a “pessimistic” forecast bias). The third set of statistics shows the number of forecasts from which the error statistics were derived. The final version of the models was run during the summer demonstrations of 2003 and 2004, representing an independent sample for model evaluation purposes. The corresponding per-

TABLE 1. Model performance statistics for (left) development dataset and (right) independent demonstration dataset.*

Development (1996–2002)

Median absolute error (min)

Hour (GMT)

	09	11	13	15	16	17	18
Consensus	46	46	43	35	29	25	34
COBEL	52	49	47	36	31	31	29
Local	40	43	37	34	28	27	32
Regional	48	52	46	46	43	40	38
Satellite	#	#	#	32	24	24	24

Bias (min)

	09	11	13	15	16	17	18
Consensus	3	3	1	-2	-1	-2	-2
COBEL	8	7	5	-1	3	4	2
Local	8	8	6	5	2	5	7
Regional	14	14	11	9	13	8	8
Satellite	#	#	#	7	5	4	13

Number of forecasts

	09	11	13	15	16	17	18
Consensus	413	440	455	450	425	329	196
COBEL	179	269	381	347	288	182	93
Local	161	237	347	305	299	226	122
Regional	381	394	386	385	366	289	168
Satellite	0	0	0	310	280	220	129

Demonstration (2003–2004)

Median absolute error (min)

Hour (GMT)

	09	11	13	15	16	17	18
Consensus	58	53	46	45	40	38	35
COBEL	56	47	44	43	37	39	54
Local	59	48	44	46	46	52	49
Regional	58	62	50	53	44	31	27
Satellite	#	#	#	44	36	32	35

Bias (min)

	09	11	13	15	16	17	18
Consensus	20	28	22	29	26	32	37
COBEL	0	8	12	13	19	28	41
Local	13	34	30	25	20	52	46
Regional	24	35	16	27	20	1	17
Satellite	#	#	#	35	31	32	44

Number of forecasts

	09	11	13	15	16	17	18
Consensus	111	121	128	127	120	92	58
COBEL	53	79	114	101	83	56	31
Local	48	74	108	99	98	69	46
Regional	107	121	124	123	116	87	55
Satellite	0	0	0	123	112	91	56

*Satellite shows no forecasts through 1300 UTC as sunrise during the summer is not until 1500 UTC.

TABLE 2. Median absolute error (min) of conditional climatology forecasts for each model initialization hour.

Hour (UTC)	0900	1100	1300	1500	1600	1700	1800
Conditional climatology	63	63	63	60	60	52	45

formance statistics for these two seasons combined are shown in the right column of Table 1.

As would be expected, the independent dataset did not score as well as that derived from the training dataset. In general, however, the median absolute errors were lower than conditional climatology (Table 2), particularly at the key tactical forecast hour of 1100 UTC, which represents the last forecast hour for which there is sufficient lead time to input to the development of the morning GDP (see next section).

The models and coefficients were held constant through the period 2004–09. In 2010 the coefficients were updated by adding the years 2004–09 to the original 1996–2002 dataset. The actual predictors remained the same, but the coefficients were modified to optimize performance with the inclusion of the new data.

Manual forecast performance. Traffic flow into SFO begins to increase dramatically around 1500–1600 UTC (8:00–9:00 a.m. PDT) due to a

combination of arriving East Coast traffic and aircraft arriving from major hubs such as Denver and Dallas and from regional airports along the West Coast. If the traffic flow is to be managed properly, the GDP needs to be issued between 1300 and 1400 UTC (6:00–7:00 a.m. PDT) on a given stratus day. Thus, the model guidance plus manual forecast need to be available prior to 1300 UTC to aid the controller in making the GDP plan. Therefore the MSFS guidance forecast available for the morning conference call typically comes from the 1100 UTC model run. This means only three of the four algorithms are available as the SSFM is not run since there is no visible satellite imagery before dawn. The MSFS guidance is available to the UAL and CWSU forecaster but not to the NWS forecaster issuing the TAF as the TAF has to be issued by 1140 UTC (4:40 a.m. PDT) and the 1100 UTC model run is not available until 1145 UTC. The ATCSCC does, however, call the

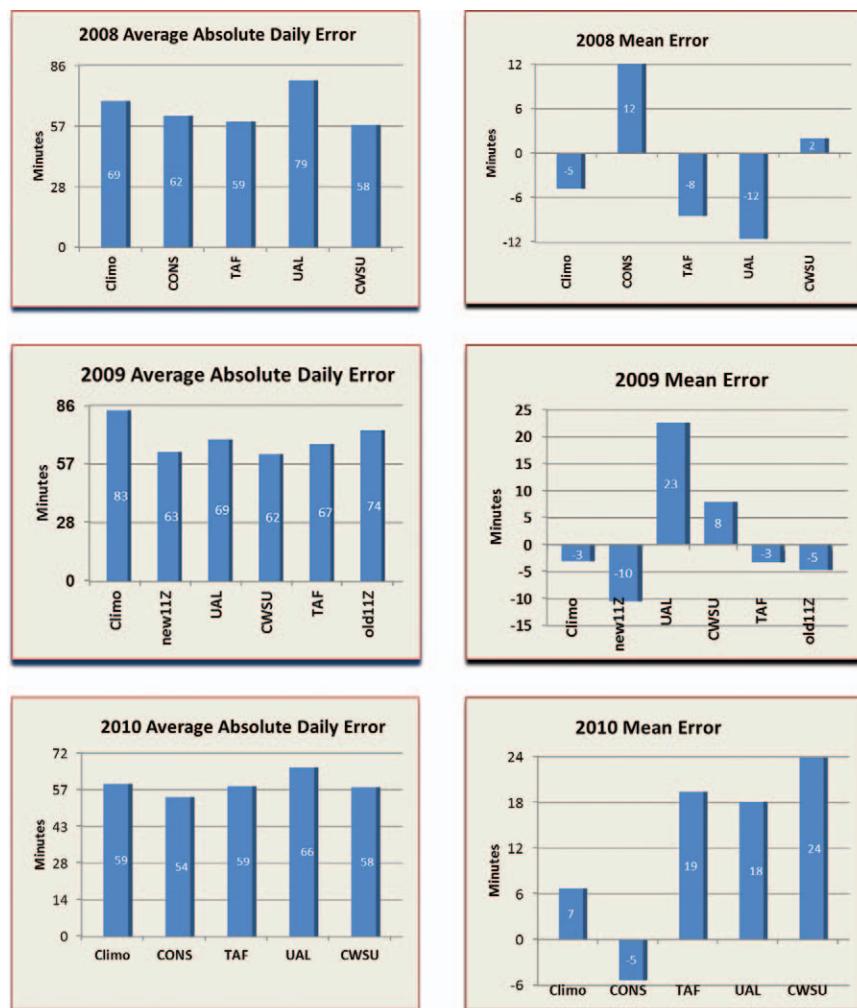


FIG. 11. Forecast verification for stratus seasons 2008–10 including climatology (1100 UTC forecast database for 2008–10), model consensus (CONS) forecast (1100 UTC), NWS 1200 UTC TAF, United Airlines 1200 UTC (UAL), and CWSU (1245 UTC). Both MAE and ME in minutes are shown. For 2009, “new11Z” refers to the CONS forecast using the revised regressions and “old11Z” to the previous CONS as per 2008. 2010 CONS uses the updated regressions.

NWS TAF forecaster prior to the issuance of the TAF to get a “heads up” as to the likelihood of a GDP for the day and an estimate of the time of clearing. Thus, the ATCSCC utilizes these four inputs in preparing the scope and duration of the GDP. The mean absolute error and mean error (bias) of forecast transition to a 45 rate for the last three stratus seasons, 2008–10, are shown in Fig. 11 for the consensus and the manually prepared forecasts. The number of days in the sample by year was 52 (46 GOOD and 6 LOW confidence) for 2008, 58 (37

GOOD and 21 LOW confidence) for 2009, and 53 (47 GOOD and 6 LOW confidence) for 2010. The days included in these calculations required the consensus and manual forecasts to be available, a GDP to have been issued, and that the forecaster did not categorize the day as “not typical stratus.” Again, this implies that the normal diurnal cycle of stratus dissipation as described in Fig. 6 may not take place in that larger-scale forcing, such as a weak upper trough off the California coast, may interfere with this process. For the three years analyzed, the days in which a GDP was issued but the day was classified as not typical stratus was 20 in 2008, 16 in 2009, and 13 in 2010.

Utilizing the same sample days that went into the statistics for each of the 3 years used in Fig. 11, the mean of the 45 rate initiation time was calculated and considered the climatological mean for the period 15 May to 15 October. Using this time, 1816 UTC, the error and bias were calculated if one were just to use climatology each day. This allows a basis for comparison to the manual and automated forecasts errors. For 2009 the consensus forecast error and bias are shown using the original coefficients (old 1100 UTC) from the 1996–2002 dataset and the updated 1996–2009 dataset (new 1100 UTC). In general the CWSU forecast has shown the least error on average. This is somewhat expected as CWSU forecasters have the benefit of seeing all other forecasts plus the consensus forecast prior to their issuance. Based on the statistics in Fig. 11, the consensus forecast provides very useful guidance, which provides the human forecaster with a good starting point.

- **Old Guidelines:**
 - Forecast provided by 8am EST
 - Proposed GDP sent to users no earlier than 8am. Users request 10West and 12West programs with minimum 4 hour lead time.
 - **Rate of 30 per hour; variable rates not recommended.**
 - **Scope recommendations:**
 - 6West – High confidence by 1800Z /1900Z
 - 10 or 12 WEST if burn-off after 1800Z/1900Z and by 2000Z/2100Z. Users expressed a strong preference for 10 West or greater program.
 - 12 or 15 WEST if burn-off after 2000Z/2100Z and by 0000Z/0100Z
 - 15 WEST or ALL if all day event
 - **Length of program = stratus burn off + 2 hrs**
 - GS into a 6West GDP if burn-off forecasted by 1700Z/1800Z, but does not occur
- **New Guidelines for 2008 Season:**
 - Distance-based programs preferred
 - **Multi-rate ground delay programs are requested with the last hour or two at a higher acceptance rate. (ex 30/30/45/45)**
 - **Scope Recommendations:**
 - 800 to 1000 miles, if high confidence or consensus that the weather condition will clear by 1800Z (daylight savings time)
 - 1200 to 1625 miles, if the weather condition is forecast to clear after 1800/1900Z and by 2000/2100Z.
 - 1900 miles, if the weather condition is forecast to clear after 2000/2100Z and by 0000/0100Z
 - ALL if the weather condition is forecast to be an all day event. SFO volume could dictate two distinct programs in this case.

The recommended length of the program will be collaboratively discussed during its development.

- GS and/or GS into a GDP if burn-off forecasted by 1700/1800Z, but does not occur.

FIG. 12. Procedures for issuing a ground delay program for SFO during summer stratus season since 2008 and with prior rules noted for comparison purposes.

Benefits of accurate manual deterministic forecast. Much attention has been paid in recent years to the question of derived benefits from developmental systems, as the FAA wants to ensure a return on their research investment. As Delman et al. (2008) and Clark (2009) showed in their review of the SFO MSFS, there were many opportunities for increasing arrival rates prior to clearing taking place, yet few were realized. A recent cost/benefit analysis performed in conjunction with the GPSM product development currently in progress showed a notable improvement in throughput efficiency at SFO since 2008 (Cook and Wood 2009). This important observation has been attributed to a significant change in the guidelines used by traffic managers in establishing delay parameters (Fig. 12). The most significant change is associated with the ability to “ramp up” planned arrival rates prior to the end of the GDP; that is, the planned rate was increased by incremental steps from 30 to 45, rather than a single abrupt change. Furthermore, prior to 2008, a very conservative approach was practiced that called for establishing the baseline program end time to be 2 h after the forecast 45 rate time and that the planned arrival rate should be held to 30 until the end time. This conservatism neutralized any real benefits from the forecast provided. This information was brought to the attention of senior FAA personnel who were instrumental in having the GDP implementation procedures modified (Fig. 12). With the new guidelines, analysis of GDPs from 2008 on showed many more days with planned arrival rates above 30 and possibly as high as a 45 rate during the last several hours of

the GDP. As such, benefits observed and quantified subsequent to 2007 are largely attributed to the change in operational procedures.

A simple methodology was applied to provide an estimate of these benefits. For each day used in the verification in the section titled "Manual forecast performance," a comparison was made between the early morning forecast of the 45 rate and the actual observed 45 rate declaration. GDP days were separated into what we will call pessimistic, with a 45 rate declared an hour or more earlier than forecast, and optimistic if the 45 rate was declared an hour or more after the forecast 45 rate, with forecasts that verified within an hour of the observed 45 rate being declared.

Pessimistic forecasts provide little benefit in that arrival rates would have still been at the planned 30 rate and the available slots for landing (45–60 rate) would not be filled. To determine what we will call the nominal recovery rate (i.e. how many slots can be filled when the rate increases earlier than forecast), we examined actual arrival rates during the 2-h period after a 45 rate was declared. Hourly rates above 30 within these 2 h suggest a nominal recovery rate without the benefit of a quality forecast. For the three years 2008–10, the nominal recovery rate in this 2-h period was from 17 to 20 additional aircraft (above the 30 rate) or 77–80 aircraft landing in this 2-h window. This nominal rate would need to be exceeded in order for a forecast to be considered to have provided a benefit. For the sample of days when the forecast was within 1 h of the declared 45 rate, a benefit is produced only when the number of arrivals exceeds 77–80 aircraft in the 2 h following a 45 rate declaration. For the seasons 2008–10, the days meeting these criteria were identified. Figure 13 summarizes the estimated benefits for the three stratus

seasons 2008–10. It shows the percentage of days in which a GDP was issued and additional aircraft were able to land due to an accurate forecast and an aggressive planned arrival rate in the GDP, the number of additional aircraft, and the dollar valuation based on the number of minutes saved. The number of minutes of savings was calculated by taking the average delay per aircraft, multiplied times the number of additional aircraft able to land, and assumes approximately 125 passengers per aircraft. Since 95% of all delays into SFO are ground holds (Cook and Wood 2009), the value of 1 min of delay for ground holds was used based on the University of Westminster analysis (Cook et al. 2004) shown in Fig. 14. The results show that during the summer stratus season, between 40% and 50% of the GDP days benefit from an accurate forecast and an aggressive planned arrival rate in the GDP. This translates to between 300 and 350 additional aircraft able to land that would otherwise have still been in a ground hold when clearing took place. Using the \$100 per minute value as derived from Fig. 14, assuming 125 passengers per aircraft, leads to an average benefit per year of \$1.67 million. However this must be adjusted taking into account the days categorized as optimistic.

For those days in which the forecast was optimistic, meaning it did not clear for more than an hour after the forecast clearing time, and the planned arrival rate was greater than 30, a GDP revision, a ground stop, or in-flight delays would have been necessary, meaning aircraft inbound to SFO were slowed so arrivals would not exceed 30. The numbers of minutes of airborne delay were calculated using actual flight data. These minutes were multiplied by \$119.28 using the long-term airborne holding line in Fig. 14 for 125 passenger aircraft. This number was subtracted from

the benefit calculated for each year to arrive at the true benefit. This is what is shown in Fig. 13. This drops the benefit to an average of \$0.85 million per stratus season. This can be compared to the average operating costs of the MSFS system of ~\$50,000 per year.

The above discussion only covers the deterministic portion of the MSFS system and how this guidance is used by forecasters and controllers to produce

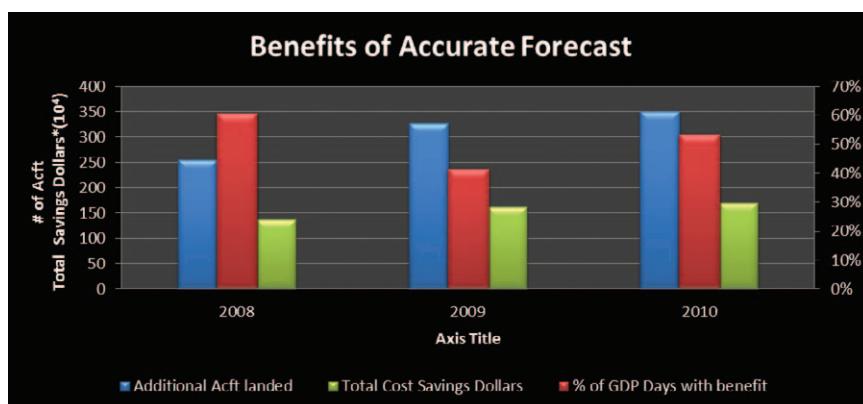


FIG. 13. Benefits of accurate forecast showing number of additional aircraft landed, percent of GDP days with benefit, and total value of minutes saved (in USD).

an improved GDP on a given stratus day. The probabilities provided by the MSFS have been extensively used in the development of the GPSM. In essence, the GPSM uses the known historical error distribution of the MSFS consensus forecasts and a frequently updated (in near-real time) aircraft arrival traffic demand profile to generate recommendations to air traffic managers for establishing GDP parameters. The GPSM presents a baseline recommendation, plus an aggressive and a conservative alternative for consideration by air traffic managers. These recommendations are presented in a separate table that has been integrated into the lower right portion of the main MSFS display (refer back to Fig. 10). Air traffic managers can then set GDP parameters using the table information as guidance, combined with their expertise, and input from the aviation forecasters familiar with the MSFS system. Details of the GPSM historical forecast performance data, which derives from the probabilities provided by the MSFS, are discussed in the next section.

GROUND DELAY PROGRAM PARAMETERS SELECTION MODEL. The GPSM integrates the probabilistic forecast of transitioning to a 45 rate into the current process of modeling and issuing GDPs at SFO. By utilizing the probabilistic nature of the forecast, the model can select the best GDP parameters given the uncertainty in the forecast, addressing the objectives of both minimizing delay and managing risk. By using the probabilistic forecasts at SFO more effectively, GDPs in today's environment can be issued less conservatively, minimizing the overall ground delay, unused arrival slots, unnecessary delay issued, and the number of aircraft affected by the GDPs. This model is an important step toward integrating probabilistic weather forecasts with traffic flow management decision support tools.

Any deterministic forecast can be translated into a probabilistic forecast by utilizing the historical validation data of that forecast product. The MSFS did exactly this, and includes in its display the

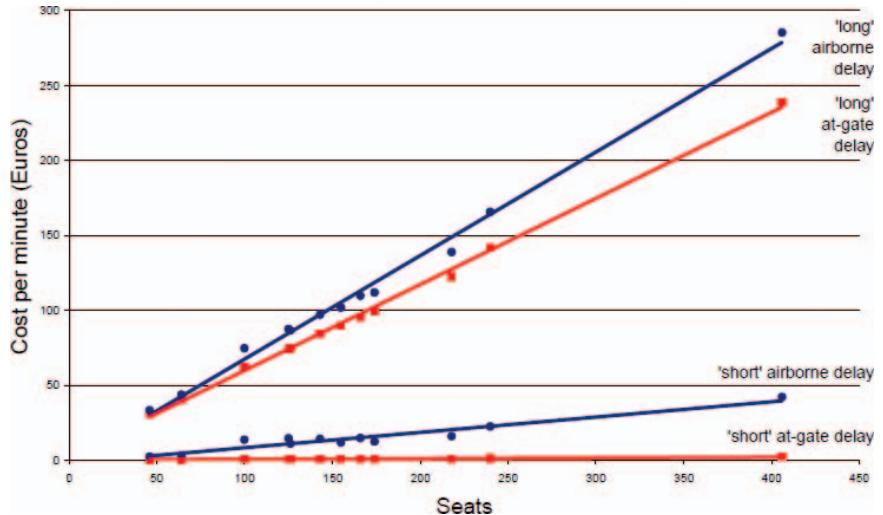


FIG. 14. University of Westminster results of 1 min of ground or airborne delay. For the MSFS benefits study we assumed 95% of the delays were ground based and an average of 100 passengers were on each additional aircraft that landed above the nominal rate (see text) on days with accurate forecast and aggressive planned arrival rate. From Cook et al. (2004).

probability of clearing at four key points in time. While the published probabilities are intended to help operational decision makers, it is difficult to integrate these probabilities into the decision making process. In fact, analysis indicated little correlation between probabilities and aggressive GDPs issued by controllers (Delman et al. 2008). GPSM addresses this issue by integrating the probabilistic forecast of clearing time mathematically into the GDP modeling process and eliminating the need for human interpretation of probabilistic information. By using 14 yr of data comparing the MSFS forecast for a 45 rate to be issued versus the actual declared 45 rate, GPSM is able to construct an error distribution around any given deterministic forecast provided by MSFS. GPSM then uses this distribution to evaluate any given set of GDP parameters (start time, end time, arrival rate, and scope) by generating expected outcomes of key metrics, such as unnecessary ground delay in the case the GDP is too conservative, and airborne holding in the case the GDP is too aggressive. These key metrics are combined into a cost function, and a GDP scenario is selected that minimizes the cost of the expected outcomes. A detailed explanation of the model can be found in Cook and Wood (2009).

The GPSM model was initially tested against GDP data from the 2006–07 stratus seasons. Based on those very promising results, the FAA funded the development of a real-time GPSM prototype to use in a field evaluation. As subsequent stratus seasons were evaluated to update the GPSM benefits assessment,

it was found that the gap between delay resulting from actual GDPs and delay resulting from GDPs recommended by GPSM started to decrease in 2008. As addressed earlier, due to a growing awareness of the amount of unnecessary delay issued during SFO stratus events, operational procedures were improved and unnecessary delay was decreased. The question remained whether GPSM could still provide benefits over those being achieved by utilizing the deterministic forecast and improved operational procedures.

Figure 15 summarizes the benefits from the GPSM system that can be achieved in terms of delay reduction based on the 2010 stratus season. These data comprise 59 GDPs that were issued during typical stratus events. Since a GDP can be modified at any point via a GDP revision, we show a comparison of delay for both the initial GDPs and the final values after all subsequent GDP revisions and possible airborne holding. The actual ground delay issued over those 59 programs is shown on the blue bars. The purple bars represent the amount of delay that would have been required given a perfect forecast (in other words, the “ideal” GDP). The red bar in between represents the resulting delays if the GPSM recommendations had been implemented. After all revisions, GPSM would have reduced delay by 29% from the delays caused by the GDP issued operationally,

which equates to an 81% reduction in the unnecessary delay that was caused by the actual programs (some portion of the actual delay is necessary, as illustrated by the ideal GDP delay—the remainder of the actual delay is unnecessary).

It is important to note that this estimate of significant savings can be achieved based on a recent stratus season, where the improved operational procedures have been implemented and refined. Using an estimate of ~\$100 cost per minute of ground delay and \$119.28 per minute of airborne delay (Cook et al. 2004), this equates to a savings of over \$11.5 million per stratus season. Figure 16 summarizes the delay reductions that could have been achieved by utilizing GPSM grouped by the past five years of stratus seasons. The decrease in the GPSM delay reduction in 2008, 2009, and 2010 illustrates the improved operational procedures and better utilization of the MSFS, but there still remains an estimated 29% of the delay that can be reduced by the use of GPSM. A full operational evaluation of GPSM is planned for the stratus season in 2012, and a future publication is planned to discuss the approach and use of GPSM in an operational real-time environment.

SUMMARY AND DISCUSSION. This paper describes the development of the Marine Stratus

Forecast System, a forecaster decision aid designed to improve the forecast 45 rate time for dual approaches into San Francisco International Airport during summer stratus events. On approximately 50–60 days between mid-May and mid-October, low-ceiling marine stratus over the approach zone precludes usage of the parallel runways at SFO, reducing arrival rates by a factor of 2. To accommodate this impact in reduced arrival capacity, the FAA ATCSCC must implement a GDP that impacts the traffic flow across the NAS. This paper has shown that ground holds can be reduced if FAA flight controllers utilize forecaster input that are based on the MSFS guidance and build in an aggressive arrival rate prior to the termination of the GDP. Prior to 2008, FAA procedures precluded or discouraged controllers from building in an aggressive arrival rate or reducing the length of the GDP based on forecast of an increased capacity. Beginning with the 2008 stratus season, procedures were modified to allow increasing the planned arrival rate above the nominal 30 per

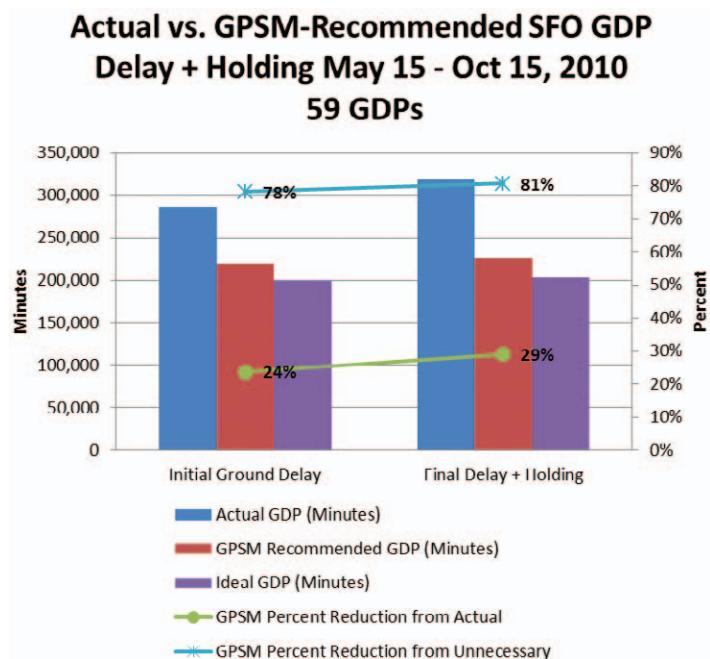


FIG. 15. Actual versus GPSM total minutes of delay along with the minutes of delay from what would have been an ideal GDP implementation. The GPSM percent reduction in delay minutes against actual GDP issued and against unnecessary minutes of delay. See text for details.

GPSM Percent Reduction in Total Ground Delay + Airborne Holding from Actual GDPs

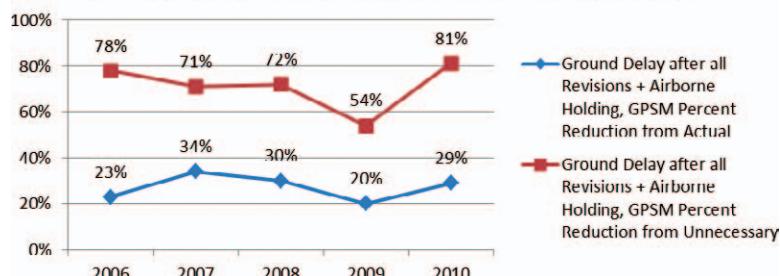


FIG. 16. GPSM percent reduction in total delay minutes compared to the actual GDP implemented taking into account ground and airborne holding as well as unnecessary holding. Again, see text for details.

hour within the last 2 h of the program. As shown in Fig. 13, utilizing the current operational MSFS from 2008–2010, approximately \$0.85 million has been saved annually by utilizing skillful forecasts. This has been accomplished by allowing between 250–350 additional arrival slots to be filled during the stratus season, thus taking advantage of what otherwise would have been wasted arrival capacity. As discussed, there is a penalty for being too aggressive, leading to airborne holding and additional cost to the airlines. However, as arrival demand is anticipated to continue to increase at SFO, it is highly beneficial that skillful forecasts be incorporated into the issuance of GDPs.

Although no funding exists to improve the MSFS at this time, some potential improvements would consist of the following: 1) Utilizing the geostationary satellite low cloud detection algorithm (Lee et al. 1997) prior to sunrise to improve the 1100 and 1300 UTC forecasts; 2) incorporating real-time boundary layer winds (boundary layer wind profiler) into a regression for predicting a 45 rate onset, knowing southerly to southwest winds in the lowest 3,000–5,000 feet can signal an early clearing over the approach to SFO; and 3) better identifying days in which the stratus fails to clear from the approach leading to all-day GDPs. These days typically lead to multiple extensions of the GDP and most likely additional delays to passengers that might have simply had flights cancelled early and been able to book flights into nearby airports such as OAK or San Jose (SJC).

The GPSM appears to offer substantial improvement in the reduction of arrival delays over the current MSFS by providing guidance to traffic managers for applying probabilistic forecast information. GPSM utilizes the historical error distribution of the MSFS's probabilistic forecasts determining

transitioning to dual runway operations to address both the objectives of minimizing delay and managing risk. Based on retrospective application of the GPSM on probabilities available for the 2008–10 season, an additional 25%–30% reduction in delays could have been realized over those reductions that FAA controllers were able to obtain utilizing the MSFS deterministic forecasts. This translates to over \$11 million in savings potential per stratus season. In cooperation with the FAA ATCSCC, a preliminary test of the GPSM was conducted during

the summer 2011 stratus season to refine the code and display. A more formal evaluation will be conducted under full operational conditions during the 2012 stratus season to see if it can produce the types of efficiencies it has shown in the retrospective analyses. This would become the first systematic attempt to integrate objective probabilistic weather information into the air traffic flow decision process, which is a cornerstone element of the FAA's visionary NextGen program for U.S. flight operations.

ACKNOWLEDGMENTS. The MSFS 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. Key contributing organizations were MIT Lincoln Laboratory, San Jose State University, the University of Quebec at Montreal, and the NWS CWSU at Oakland Center.

Since 2004, the Aviation Services Branch of the National Weather Service has funded the operation of the MSFS. The GPSM development work was performed by Mosaic ATM, Inc., funded by NASA Ames Research Center and the FAA System Operations Programs. The Marine Meteorology Division of the Naval Research Laboratory in Monterey, California, provides the visible satellite imagery for use in MSFS. Special thanks go to Chris Stumpf, with the Student Cooperative Employment Program (SCEP) at NWS Monterey, for preparing several of the figures dealing with forecast verification. We would also like to thank three anonymous reviewers for their careful review and thoughtful comments. This has greatly improved the paper.

We dedicate this paper to the memory of Peter Zwack, who participated enthusiastically in the developments of this project, who provided the COBEL technology, and who enriched our scientific understandings in many ways.

REFERENCES

Clark, D. A., 2002: The 2001 demonstration of automated cloud forecast guidance products for San Francisco International Airport. *Extended Abstracts, 10th Conf. on Aviation, Range, and Aerospace Meteorology (ARAM)*, Portland, OR, Amer. Meteor. Soc., JP1.26. [Available online at http://ams.confex.com/ams/13ac10av/techprogram/paper_38862.htm.]

—, 2009: Investigating a new ground delay program strategy for coping with SFO stratus. *Extended Abstracts, Aviation, Range and Aerospace Meteorology Special Symp. on Weather-Air Traffic Management Integration*, Phoenix, AZ, Amer. Meteor. Soc., 4.6. [Available online at http://ams.confex.com/ams/89annual/techprogram/paper_149997.htm.]

—, C. V. Ivaldi, F. M. Robasky, K. MacKenzie, R. G. Hallowell, F. W. Wilson, and D. M. Sinton, 2006: SFO marine stratus forecast system documentation. MIT Lincoln Laboratory Rep. ATC-319, 482 pp. [Available online at www.ll.mit.edu/mission/aviation/publications/publication-files/atc-reports/Clark_2006_ATC-319_WW-15318.pdf.]

Cook, A., G. Tanner, and S. Anderson, 2004: Evaluating the true cost to airlines of one minute of ground delay. University of Westminster on behalf of EUROCONTROL, 260 pp. [Available online at http://westminsterresearch.wmin.ac.uk/171/Cook%2CTanner%2CAnderson_2004_final.pdf.]

Cook, L., and B. Wood, 2009: A model for determining ground delay program parameters using a probabilistic forecast of stratus clearing. *Proc. Eighth USA/Europe Air Traffic Management Research and Development Seminar*, Napa, CA, FAA, Paper 125. [Available online at www.atmseminar.org/seminarContent/seminar8/papers/p_125_W.pdf.]

Delman, M., D. W. Reynolds, and N. M. Ray, 2008: Marine stratus initiative and effect of stratus on arrival rates at SFO. *Extended Abstracts, 13th Conf. on Aviation Range and Aerospace Meteorology*, New Orleans, LA, Amer. Meteor. Soc. [Available online at http://ams.confex.com/ams/88Annual/techprogram/programexpanded_474.htm.]

Fidalgo, C., D. Sims, S. McGettigan, and J. Weinrich, 2002: Demonstration of the marine stratus forecast product for the San Francisco International Airport. *Extended Abstracts, 10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 8.11. [Available online at http://ams.confex.com/ams/13ac10av/techprogram/paper_38896.htm.]

Lee, T., J. Turk, and K. Richardson, 1997: Stratus and fog products using GOES-8–9 3.9-mm data. *Wea. Forecasting*, **12**, 664–677.

MCR Federal, Inc. 2004: Ceiling and visibility at SFO—Marine stratus forecast benefits estimate. MCR Briefing BR-M021/080–4.

Strach, W. J., 1991: The effects of summer time stratus at San Francisco International Airport on the nationwide flow of commercial airline traffic. *Proc. First NWS Aviation Weather Workshop*, Kansas City, MO, NWS.

Wilson, F. W., 2004: Probabilistic forecast of cloud impacts at San Francisco International Airport. *Extended Abstracts, 20th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Seattle, WA, Amer. Meteor. Soc., 18.1. [Available online at http://ams.confex.com/ams/84Annual/techprogram/paper_71372.htm.]

—, and D. A. Clark, 2000: Forecast aids to lessen the impact of marine stratus on San Francisco International Airport. *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 5.1. [Available online at www.ll.mit.edu/mission/aviation/publications/publications-by-name/wilson-fw.html.]