Project Report
ATC-319

SFO Marine Stratus Forecast System Documentation

D.A. Clark
C.F. Ivaldi
F.M. Robasky
K. MacKenzie
R.G. Hallowell
F.W. Wilson
D.M. Sinton

17 October 2006

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Federal Aviation Administration,
Washington, D.C. 20591

This document is available to the public through
the National Technical Information Service,
Springfield, VA 22161

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. ATC-319	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle SFO Marine Stratus Forecast System		5. Report Date 17 October 2006
		6. Performing Organization Code
7. Author(s) D.A. Clark, C.F. Ivaldi, F.M. Robasky, K. MacKenzie, R.G. Hallowell, F.W.Wilson, and D.M. Sinton		8. Performing Organization Report No. ATC-319
9. Performing Organization Name and Address MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. FA8721-05-C-0002
12. Sponsoring Agency Name and Address Department of Transportation Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered Project Report
		14. Sponsoring Agency Code

15. Supplementary Notes

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, under Air Force Contract FA8721-05-C-0002.

16. Abstract

San Francisco International Airport (SFO) experiences frequent low ceiling conditions during the summer season due to marine stratus clouds. Stratus in the approach zone prevents dual approaches to the airport's closely spaced parallel runways, effectively reducing arrival capacity by half. The stratus typically behaves on a daily cycle, with dissipation occurring during the hours following sunrise. Often the low ceiling conditions persist throughout the morning hours and interfere with the high rate of air traffic scheduled into SFO from mid-morning to early afternoon. Air traffic managers require accurate forecasts of clearing time to efficiently administer Ground Delay Programs (GDPs) to match the rate of arriving aircraft with expected capacity.

The San Francisco Marine Stratus Forecast System was developed as a tool for anticipating the time of stratus clearing. The system relies on field-deployed sensors as well as routinely available regional surface observations and satellite data from the Geostationary Operational Environmental Satellite (GOES-West). Data are collected, processed, and input to a suite of forecast models to predict the time that the approach zone will be sufficiently clear to perform dual approaches. Data observations and model forecasts are delivered to users on an interactive display accessible via the Internet.

The system prototype was developed under the sponsorship of the FAA Aviation Weather Research Program (AWRP). MIT Lincoln Laboratory served as technical lead for the project, in collaboration with San Jose State University, the University of Quebec at Montreal, and the Center Weather Service Unit (CWSU) at the Oakland Air Route Traffic Control Center (ARTCC). The National Weather Service (NWS), under the direction of the NWS Forecast Office in Monterey, assumed responsibility for operation and maintenance of the system following technical transfer in 2004. This document was compiled as a resource to support continuing system operation and maintenance.

17. Key Words San Francisco, marine stratus, low ceiling, dual approaches, air traffic managers, ground delay programs, forecast models, Center Weather Service Unit, NWS Monterey	18. Distribution Statement This document is available to the public through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	466	

FORM DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

Of the dozens of individuals who participated in the system development process over the years, the authors would like to specifically acknowledge the contributions made by the Department of Meteorology at San Jose State University and the National Weather Service.

The SJSU Department of Meteorology effectively served as the field site team. We thank Peter Lester for initiating the department's role in the project. Special thanks go to Doug Sinton for his leadership role and contribution to the technical effort, particularly in the areas of data acquisition & handling, communications, and system performance monitoring. Thanks also go to John Farley, who took primary responsibility for ensuring reliable operation of the weather sensors at the two field sites. The stability and success of the project was largely dependent upon the perseverance and personal commitment exhibited by both Doug and John.

We are grateful to the National Weather Service for their partnership in the project, both in the development process and ultimately in ensuring the continued operation of the system. First, we thank Walt Strach who was the primary operational community representative on the scientific team. Second, thanks go to Curt Lutz and Wayne Bailey for their efforts during the technical transfer process. And finally, we are especially grateful to Dave Reynolds for his early support while the project was being established, and for ultimately leading the effort to ensure that the prototype continued on as an operational system.

ABSTRACT

San Francisco International Airport (SFO) experiences frequent low ceiling conditions during the summer season due to marine stratus clouds. Stratus in the approach zone prevents dual approaches to the airport's closely spaced parallel runways, effectively reducing arrival capacity by half. The stratus typically behaves on a daily cycle, with dissipation occurring during the hours following sunrise. Often the low ceiling conditions persist throughout the morning hours and interfere with the high rate of air traffic scheduled into SFO from mid-morning to early afternoon. Air traffic managers require accurate forecasts of clearing time to efficiently administer Ground Delay Programs (GDPs) to match the rate of arriving aircraft with expected capacity.

The San Francisco Marine Stratus Forecast System was developed as a tool for anticipating the time of stratus clearing. The system relies on field-deployed sensors as well as routinely available regional surface observations and satellite data from the Geostationary Operational Environmental Satellite (GOES-West). Data are collected, processed, and input to a suite of forecast models to predict the time that the approach zone will be sufficiently clear to perform dual approaches. Data observations and model forecasts are delivered to users on an interactive display accessible via the Internet.

The system prototype was developed under the sponsorship of the FAA Aviation Weather Research Program (AWRP). MIT Lincoln Laboratory served as technical lead for the project, in collaboration with San Jose State University, the University of Quebec at Montreal, and the Center Weather Service Unit (CWSU) at the Oakland Air Route Traffic Control Center (ARTCC). The National Weather Service (NWS), under the direction of the NWS Forecast Office in Monterey, assumed responsibility for operation and maintenance of the system following technical transfer in 2004. This document was compiled as a resource to support continuing system operation and maintenance.

MASTER GLOSSARY

AC	Approach Clear
AZ	Approach Zone
ACV	Arcata
ARTCC	Air Route Traffic Control Center
AWRP	Aviation Weather Research Program
AWOS	Automated Weather Observing System
BKN	Broken
CFA	Consensus Forecast Algorithm
CLR	Clear
COBEL	Couche Brouillard Eau Liquide
CWSU	Central Weather Service Unit
DCP	Data Collection Platform
DPA	Dual Parallel Approaches
FA	Free Atmosphere
FAA	Federal Aviation Administration
GMT	Greenwich Mean Time
GOES	Geostationary Operational Environmental Satellite
ITWS	Integrated Terminal Weather System
LAN	Local Area Network
LCL	Lifted Condensation Level
LSFM	Local Statistical Forecast Model
MIT/LL	Massachusetts Institute of Technology Lincoln Laboratory
MSF	Marine Stratus Forecast
NAM	North American Model
NRL	Naval Research Laboratory
NWS	National Weather Service
NWSFO	National Weather Service Forecast Office
OAK	Oakland
OVC	Overcast
PGAM	Pre-scaled Generalized Additive Modeling
PPP	Point to Point Protocol
RSFM	Regional Statistical Forecast Model
SCT	Scattered
SFM	Statistical Forecast Model
SFO	San Francisco International Airport
SMB	San Mateo Bridge
SODAR	Sonic Detection and Range
SQL	San Carlos Airport
SQLL	San Carlos High Resolution Sensor
SSFM	Satellite Statistical Forecast Model
UQAM	University of Quebec at Montreal

MASTER REFERENCES

Chernal, E., A. Bellon, B.L. Austin, and A. Kilambi, An Objective Technique for the Delineation and Extrapolation of Thunderstorms from GOES Satellite Data, *J. of Geoph. Res.*, **90**, No. D4, 6203–6210, 1995.

Clark, D.A. and F.W. Wilson, "The Marine Stratus Initiative at San Francisco International Airport," MIT Lincoln Laboratory, Project Report No. ATC-252, Lexington, MA, 41 pp., 1996.

Farley, John, personal communication, San Jose State University.

Rao, C.R.N., 2003, "Implementation of the Post-Launch Vicarisou Calibration of the GOES Imager Visible Channel," www.oso.noaa.gov/goes/goes-calibration.

Rew, R.K. and G.P. Davis, NetCDF: An Interface for Scientific Data Access, Computer Graphics and Applications, *IEEE*, pp. 76–82, 1990.

Turner, S., R. Tardif, P. Zwack, C. Page, N. Gauthier, and E. Monteiro, 2004: Stratus burnoff forecasting with the COBEL column model – product description. <Publication TBD>. University of Quebec, Montreal, Canada.

Vislocky, R.V. and J.M. Fritsch, 1995: Generalized additive models versus linear regression in generating probabilistic MOS forecasts of aviation weather parameters. *Wea. Forecasting*, **10**, 669–680.

Wilson, F.W. and D.A. Clark, "Interim Report: Products of the SFO Marine Stratus Initiative," MIT Lincoln Laboratory, Lexington, MA, Project Memorandum No. 43PM-Wx-0058, 1999.

TABLE OF CONTENTS

SFO Marine Stratus Forecast System Documentation

Section I. Overview

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

1.1 SCOPE

This document is provided as a resource for the use and maintenance of the Marine Stratus Forecast (MSF) system. The system was developed as a real-time forecasting tool for anticipating the time of stratus clearing from the runway approach zone at San Francisco International Airport. The clearing time is important to air traffic flow management, as the presence of approach zone stratus greatly reduces the airport operating capacity. Traffic managers require an accurate stratus forecast to properly meter incoming planes so that aircraft arrival demand properly matches arrival capacity.

The system relies on a network of sensors in the San Francisco Bay region, as well as satellite data from the western U.S. Geostationary Operational Environment Satellite (GOES-West). The data are collected, processed, and input to a suite of forecast models developed to predict the stratus clearing time. Both the data observations and the forecast model guidance are delivered to users via a web-browser display accessible via the Internet. The primary users of the system are forecasters responsible for providing guidance for air traffic management purposes, both in the public and private sectors.

This document is written and formatted as a reference guide to system users, and as a system maintenance guide for ensuring system availability and reliable operation. It is arranged in six main sections. In addition to this overview, there is a section covering the various functional areas within the system, plus a section which serves as a troubleshooting guide for system maintenance.

Section I.	Overview
Section II.	Sensor Descriptions and Maintenance
Section III.	Real-Time System
Section IV.	Display System
Section V.	Forecast Model Descriptions
Section VI.	Troubleshooting Guide

This overview provides background information regarding system development, including a high level description of the system configuration and operation.

Section II provides a description of the weather sensors and the data collection platform for acquiring and formatting the raw data. Actions required for monitoring sensor performance and detecting failures are identified. Included is a description of primary preventive and corrective maintenance tasks, as well as the appropriate references to vendor documentation and manufacturer contact information.

Section III describes the system for real-time data processing. It includes a complete description of the system configuration, and details of the operational system software installed on the base station computer. This software is responsible for collecting data, generating derived products, and managing the activation and timing of forecast models that utilize the data and derived products.

Section IV describes the display system. It is intended for both the audience of operational users and for maintenance of the display system. Included is an operational overview that serves as a user guide that provides a description of the display products and user interface for accessing the variety of current and historical weather observations and forecast model output. This is followed by a description of the data processing required to serve the display system, and the processing required to build the various display products. These are intended more for those responsible for maintaining the file structure and display software for ensuring proper display system operation and availability.

Section V describes the forecast models that are run in real time to provide objective guidance for estimating time of approach zone clearing. There are four "component" forecast models, and one "consensus" forecast algorithm which integrates the component forecast to provide a single estimate of clearing time. Included is a description of how each model was developed, the methodology and the predictors that are employed, and the relative strengths and weaknesses of the individual models. Also provided for each model is a description of the underlying data requirements, file structure, and processing required for the models to run in real time.

Section VI is a troubleshooting guide for isolating and solving system problems. This guide supplements the automated self-diagnostic electronic messages generated from the base station computer when problems are detected. It summarizes the symptoms of various potential system problems, and suggests explanations or recommendations for corrective action.

1.2 HISTORICAL BACKGROUND OF SYSTEM DEVELOPMENT

The local airspace surrounding San Francisco International Airport (SFO) is prone to regular occurrences of low ceiling conditions from May through October due to the intrusion of marine stratus cloudiness along the Pacific coast. The low cloud conditions prohibit dual parallel approaches of aircraft to the airport's closely spaced parallel runways, thus effectively reducing the arrival capacity by a factor of two. The behavior of marine stratus evolves on a daily cycle, filling the San Francisco Bay region overnight, and dissipating during the morning. Often the low ceiling conditions persist throughout the morning hours and interfere with the high rate of air traffic scheduled into SFO from mid-morning to early afternoon. The result is a substantial number of delayed flights into the airport. Air traffic managers face a continual challenge of anticipating available operating capacity so that the demand of incoming planes can be metered to match the availability of arrival slots.

The Central Weather Service Unit (CWSU) at the Oakland Air Route Traffic Control Center (ARTCC), the National Weather Service (NWS) Office in Monterey, and meteorologists at the Air Traffic Control System Command Center in Washington, D.C. are responsible for providing weather information to assist air traffic managers to anticipate changes in available operating capacity. Meteorologists at commercial airlines are also responsible for providing weather information for the most efficient and cost effective dispatch of airplanes.

In 1995, the FAA's Aviation Weather Research Program (AWRP) launched a "Marine Stratus Initiative" whose objective was to improve cloud forecast guidance for more reliable anticipation of available operating capacity at SFO. MIT Lincoln Laboratory served as technical lead for the project, in collaboration with San Jose State University, the University of Quebec at Montreal, and the CWSU at Oakland Center. The project led to the development of the Marine Stratus Forecast System.

An important element of the system is the generation of automated objective forecast guidance to be used by forecasters in making their recommendation to air traffic managers regarding the timing of cloud dissipation that will allow dual independent aircraft approaches. This guidance is presented to system users in the form of a "consensus forecast" derived from four independent component forecasts. In addition to providing a discrete forecast of the best estimate of transition time to dual approaches, the consensus forecast is also converted to a probabilistic representation showing the likelihood of dual approach availability at specific target forecast times during the period of highest arrival traffic demand.

During the early years of development, focus was on collection of a suitable database of relevant weather observations to be used for forecast model development. Data were collected from routinely available sources, such as hourly regional surface weather observations, twice-daily upper air soundings, satellite data, etc. Following discussion with the local aviation forecasting community, in particular the CWSU at Oakland Center, it was determined that additional specialized sensors would be required in the vicinity of the approach zone into SFO. Two sensor sites were established, one each at SFO and at San Carlos Airport (SQL) just to the south of the approach zone. A suite of sensing equipment was established at these two locations; the suite included sodars (SOnic Detection And Ranging) for estimating the height of the stratus layer top, pyranometers for measuring the amount of incoming solar radiation, and high resolution measurements of temperature, dew point, and wind that would be required to run a localized cloud prediction model.

Display products were developed to deliver the weather observations to the CWSU while the forecast models were being developed. In 2001, the first iteration of forecast models was ready for operational demonstration. The display system was expanded to include presentation of the forecast output, and deliver the observations and forecasts to a more general audience via the Internet. This initial demonstration ran through the summer months of 2001. This initial operational exposure led to a streamlining of the model development process and a modification of the models in preparation for a follow-up demonstration during the 2002 stratus season. This iterative process of test and development was repeated during the summer of 2003.

Upon reaching an acceptable level of reliability and performance, the National Weather Service (under the leadership of the Weather Service Forecast Office in Monterey) agreed to adopt responsibility for operation and maintenance of the system, as the initial research phase supported by the FAA was coming to a close. The 2004 stratus season served as a transition period for preparing the system for handoff to NWS. A major upgrade, re-configuration, and stabilization of hardware, communications, security, and software was performed, with the revitalized system put back on line by late June 2004. Operations ran throughout the remainder of the 2004 stratus season, as operational and maintenance responsibilities were gradually transitioned from the development team to the NWS.

NWS intends to continue operation and evaluation of the system's utility as part of its mission to provide accurate and timely forecasts for the aviation weather community in the Bay Area.

2. SYSTEM OVERVIEW

2.1 DATA SOURCES AND SENSOR CONFIGURATION

The SFO Marine Stratus Forecast System relies on a variety of weather observations acquired from land-based sensors throughout the Bay region, upper-air data via balloon-launched rawinsondes, and GOES-West satellite data. Data from these various sources are transferred to a base station computer located at the Oakland (ZOA) Air Route Traffic Control Center (ARTCC) in Fremont, CA at a time interval of 15 minutes. There the data are processed for display, and for input into a suite of models designed to forecast the time of Approach Zone clearing. The display of observations and the automated forecast guidance are made available to users via the World Wide Web.

The geographical configuration of sensors which supply data to the system is shown in Figure 1. There are two primary sensor sites, located at San Francisco International Airport (SFO) and San Carlos Airport (SQL) along the Approach Zone. These two sites were instrumented specifically to support the forecast system. Each site includes a Sonic Detection and Range (sodar) instrument to estimate cloud top height, a ceilometer to estimate cloud base height, a pyranometer to measure incoming solar radiation, and instrumentation to provide high resolution (in time and accuracy) measurements of temperature, humidity, and wind. Most of the remainder of observations are acquired from the standard suite of weather observations made widely available by the National Weather Service. This includes hourly surface observations of standard meteorological parameters (temperature, humidity, wind, cloud amount and heights), twice-daily vertical measurements (temperature, humidity, and wind) from the balloon rawindsonde launched from Oakland, and satellite data from GOES-West available at 15-minute intervals. The other key significant observation of note is the FAA-maintained surface weather sensor located at the San Mateo Bridge, situated directly below the primary runway approach zone into SFO airport. This sensor provides the key ceilometer measurement of cloud amount and cloud base height used for operational purposes in assessing the availability of visible approaches into SFO.

Figure 1 also shows the location of the system base station computer located at the Oakland Air Route Traffic Control Center (ARTCC) in Fremont, CA. The base station computer serves the function of collecting and processing data from the variety of sources and generating display products. In addition to local display, the base station computer acts as a web server for delivering the display products to external system users via the Internet.

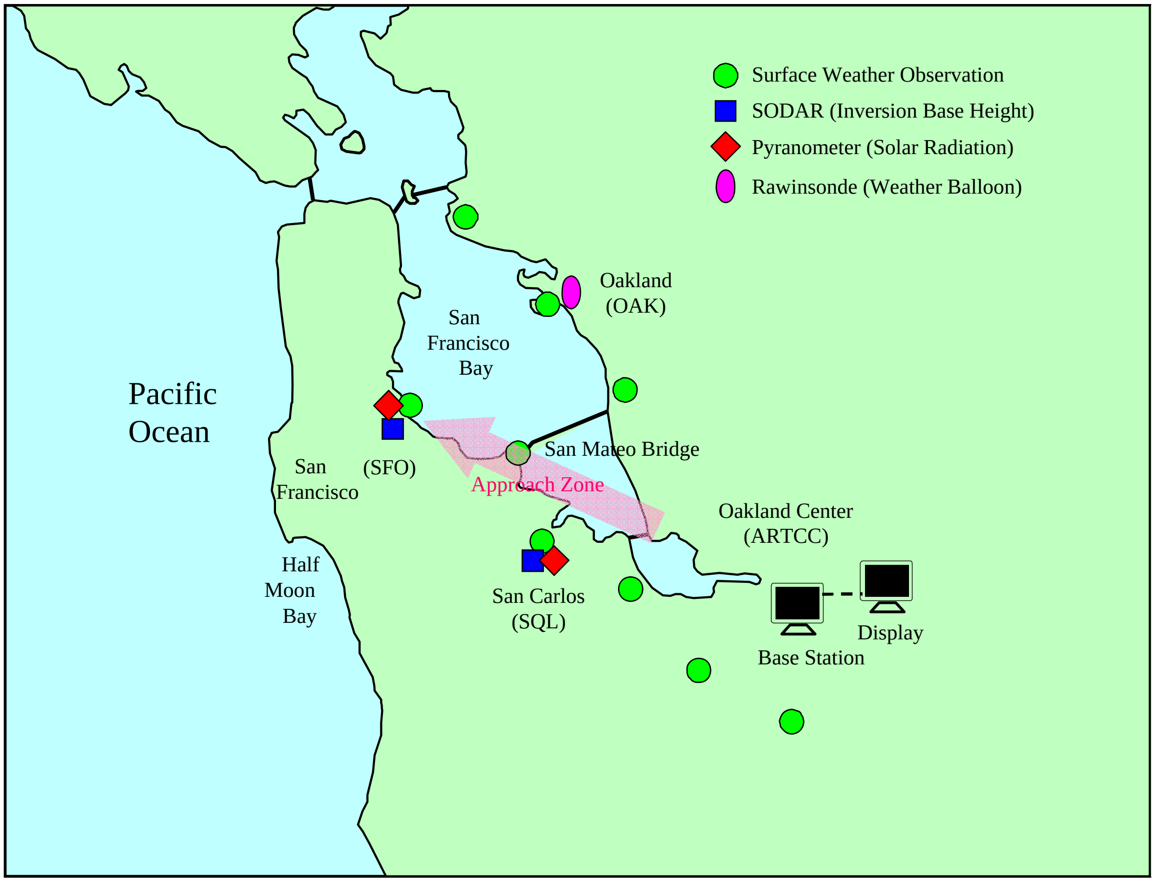


Figure 1. SF Bay area sensor locations.

2.2 DATA ACQUISITION AND REAL-TIME PROCESSING

Figure 2 shows a schematic representation of the hardware and communications configuration for data acquisition and processing. The system encompasses three locales: two field sites equipped with identical hardware and software located at San Francisco International Airport and San Carlos Airport, and a two-node local area network (LAN) housed within CWSU at the Oakland ARTCC. The LAN is composed of a base station computer used for data collection and processing, and a display computer used by forecasters to view forecasts of marine stratus burnoff within the approach zone as well as near realtime field-sensor observations and derived products. The base station acquires data from both the field sites and other sources using a commercial DSL connection to the Internet. A backup means to acquire field site data is available if needed, using a dialup Point to Point Protocol (PPP) connection over a phone line that is shared with DSL. The dialup connection is shown in Figure 2 as the dashed line connecting field site modems to the base station modem. Five-minute ASOS surface observations for San Francisco Airport are acquired through a dialup connection using the Kermit protocol. Five-minute Automated Weather Observing System (AWOS) surface observations from an FAA-managed site located at the San Mateo Bridge enter the base station via a serial RS232 line that originates within the confines of the CWSU weather desk. The Oakland upper-air sounding and hourly surface observations from stations in and around San Francisco Bay are provided by the National Weather Service Forecast Office (NWSFO) in Monterey and satellite data are retrieved from the Naval Research Laboratory (NRL), also located in Monterey. These data sources rely on the DSL connection for transfer of data to the base station.

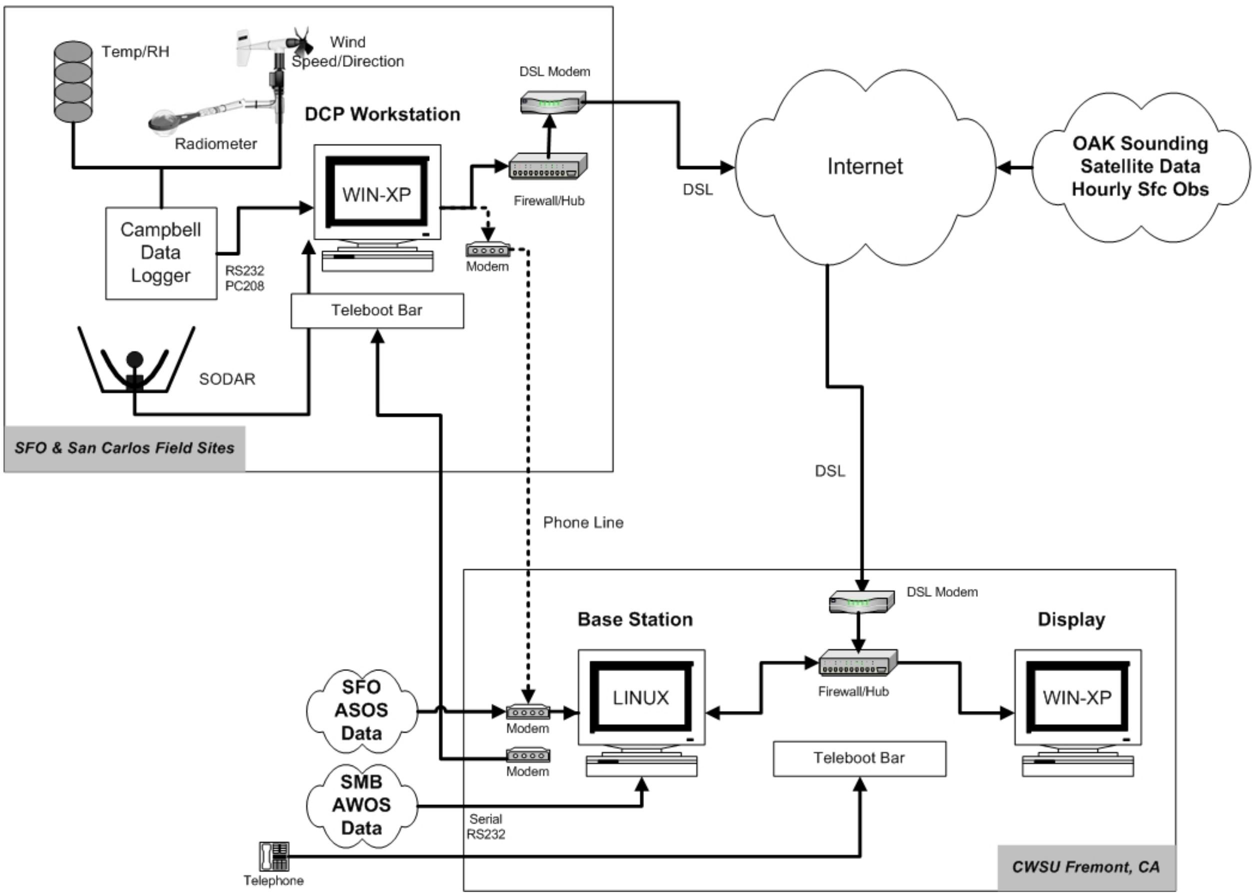


Figure 2. System diagram depicting the San Francisco Marine Stratus Forecast System.

Field site instrumentation, including temperature and humidity sensors, a pyranometer, and wind sensor, are tower mounted and connected to a data logger that prepares the sensor data for acquisition by the Data Collection Platform (DCP), a PC running the Windows XP operating system. The DCP also collects and processes data from an acoustic sounder that profiles the atmosphere just above the field site.

Both the field sites and the LAN at the CWSU are protected from Internet intrusion through inclusion of a firewall device. The firewall is configurable through a user interface to allow only permitted inbound traffic for remote login. The firewall does not limit outbound access. At the CWSU, the firewall also acts as a hub, connecting the base station and display computer in a LAN configuration.

Finally, all three locations are configured with teleboot bars. A teleboot bar allows a user or a computer to dial into the device and cycle power to any device attached to the bar. In Figure 2, a dedicated modem at the CWSU provides the base station with the ability to dial into either field site and clear a malfunctioning device without a person-in-the-loop. A teleboot bar is also located at the CWSU to provide maintenance personnel the ability to dial in to the CWSU and reboot the base station or communications equipment.

Details of the software components that run on the base station computer (data acquisition, data processing, and forecast model activation) are provided in Section III, Real-Time System.

3. FORECAST MODELS

3.1 OVERVIEW

Data collected at the base station computer are supplied to four "component" forecast models, each designed to independently provide an estimate of the time at which the approach zone is expected to be sufficiently free of clouds to allow parallel runway approaches into SFO. Individual forecasts from the four component models are combined to yield a single "Consensus" forecast. The forecast process is run on a daily cycle, beginning at 09 GMT each morning (2 AM Pacific Daylight Time). The models are initialized with new data and run at the top of the hour at 09, 11, 13, 15, 16, 17, and 18 GMT throughout the morning, as long as cloud is present in the approach zone. Once the approach zone has cleared, the models cease to run.

The four component models consist of three statistically derived forecast models, and a physics-based numerical model:

1. Local Statistical Forecast Model (LSFM)
2. Regional Statistical Forecast Model (RSFM)
3. Satellite Statistical Forecast Model (SSFM)
4. COBEL

The models are quasi-independent in that they rely on significantly different data observations as input. The benefits of this are threefold. First, it provides four different perspectives on the forecast, allowing the opportunity for either inter-model consistency or variability, which is a useful indicator to the forecaster as to the reliability of the current forecast. Second, the use of a multiple forecast/consensus approach allows the forecast system to remain available in the event that one or more data sources become unavailable. The consensus forecast algorithm is designed to effectively combine the available component forecasts using a weighting system that is based on the relative historical performance of each forecast model at each model run hour. Third, the consensus approach tends to filter the hour-to-hour variance that is sometimes observed in the individual component forecasts.

Each of the four forecast models has a specific set of input data required for execution. The models optimally run using observations from the top of the current hour; however, there is flexibility to use data observations from within 15-minutes from the top of the hour. This flexibility allows the models to run in the event that the optimal top-of-the-hour observations are unavailable. In particular, this accommodates potential missing observations in the visible satellite image which is routinely available at 15-minute intervals, and the subset of surface observations that are available at 5-minute intervals.

Beginning at the top of each forecast run hour, each model will execute provided that its required data are available. Once all of the models have received their data and have executed, the Consensus algorithm is prompted to generate a consensus forecast, followed by a posting of all forecasts to the display. The display is refreshed with new forecasts only once per hour, i.e., the individual models are not allowed to update asynchronously to the display within a given hour.

Typically, all of the top-of-the-hour data are received and processed by the forecasts models by 20-25 minutes past the top of the hour, resulting in a refresh of forecasts to the display. However, during some hours, the receipt of data is delayed, or may not occur at all. The system will wait up to 35 minutes (a parameter setting within the system) past the top of the hour for new data to arrive. Once this waiting period has expired, a consensus forecast will be generated using whatever models are available, and post the new forecasts to the display, even if one or more of the component forecasts was not able to execute.

3.2 STATISTICAL FORECAST MODEL DEVELOPMENT

Three of the four component forecast models were statistically derived. Each of the three statistical forecast models (SFMs) was developed using essentially the same methodology, with the primary difference being the nature of the input data. Since this statistical approach was central to the forecast development methodology, a brief description of the process is presented here. A more detailed description is found in Section V: Forecast Model Descriptions.

3.2.1 General Methodology

The forecast development process uses statistical regression to generate a set of forecast equations based on the selection of the appropriate set of predictors and coefficients to arrive at an effective forecast model equation for estimating a predictand. This process was performed separately for each model (Local, Regional, and Satellite) and for each model run hour (09, 11, 13, 15, 16, 17, and 18 GMT). For the purposes of the SFO Marine Stratus Forecast system, the predictand was defined to be the time (on a given day that stratus is impacting the Approach Zone) that dual independent approaches become available at SFO. This is an important distinction from defining the predictand as a meteorological parameter, as it is an attempt to directly predict the operation-impacting event that is the prime motivation for system development. In most instances, this operational event is closely tied to one or more meteorological events, most notably the time that the ceilometer at San Mateo Bridge indicates no ceiling. However, the dual independent approach event may also be influenced by other meteorological or operational circumstances. The regression analysis uses the daily record (dating back more than 15 years) of the daily transition time from single to dual approaches at SFO. This is commonly referred to as the "side-by" time, and corresponds to the time at which there is an increase in the effective aircraft arrival rate.

All of the final statistical forecast equations are of the general form:

$$T = \text{SunriseTime} + \text{Constant} + a_1X_1 + a_2X_2 + \dots a_nX_n$$

where T = Time of transition to dual approaches (i.e., side-by time)
X_n = individual predictor
a_n = predictor coefficient weighting

Note the dependence of the prediction time on the time of SunriseTime each day which reaches its minimum (earliest) near the summer solstice and increases each day for the remainder of the summer stratus season. This is consistent with solar radiation being the primary physical mechanism that forces the evaporation process.

The statistical forecast development process chooses an appropriate set of predictors and coefficients from a large set of candidate predictors, to arrive at a set of candidate forecast equations. The process is performed separately for each model, and for each model initialization hour. This five-step process is shown schematically in Figure 3.

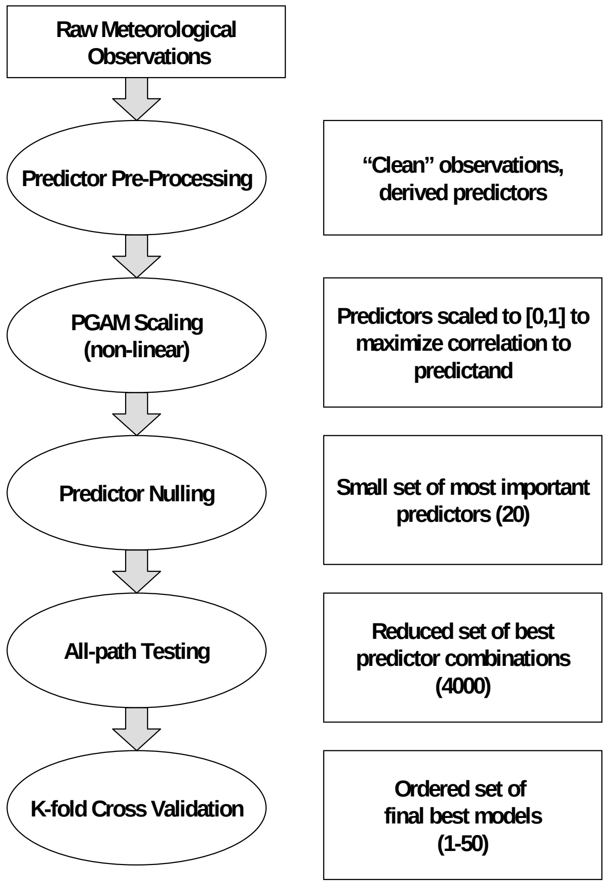


Figure 3. Primary steps of statistical forecast development process.

The process begins with a large set of potential predictors, which undergo a pre-processing step from their raw form. First, the data are quality checked to deal with erroneous or missing data. Second, the data may be processed to yield a "derived" predictor from the raw data. An example of this might be using a time-difference value of a parameter as a potential predictor.

The second step is a non-linear re-scaling of the candidate predictors using a methodology referred to as Pre-scaled Generalized Additive Modeling (PGAM). This process examines the correlation of each candidate predictor with the predictand, and re-scales its value non-linearly to a value between 0 and 1, in an attempt to expose the most sensitive range of values correlating to the predictand.

Once the candidate predictors have been re-scaled, they then undergo a process referred to as "predictor nulling," which is an attempt to reduce the large set of predictors down to a more computationally manageable number suitable for equation development (typically on the order of 20 or so predictors). Many of the predictors are redundant in that their relationship to the predictand is very similar. The nulling process examines the predictor set for these redundancies, and reduces the set to the most promising candidates.

The fourth step in the process is to examine all combinations of the 20 best candidate predictors. This process is referred to as "All-Path Testing," which applies a least squares error test on a training portion of the data set to yield the best set of candidate prediction equations containing from 1 to 20 predictors.

Finally, the candidate equations undergo a multi-fold cross validation procedure to provide a rank-order list of the predictor equations yielding the least error on quasi-independent data sets. The "quasi" independence is an artifact of the limited data availability, wherein the test data set is randomly apportioned into difference combinations of subsets to provide multiple independent assessments of performance.

Thus, this 5-step process distills a large number of predictors to produce a rank-ordered set of candidate prediction equations. This rank-order list is then used more subjectively to select the final form of the forecast equation.

The rank-order list essentially presents the quality of the candidate equation in terms of minimizing least-squares error. From the thousands of "best" possible candidates, the top group (say, a few equations to a few dozen equations) is roughly equivalent in terms of the error metric. The final selection process entails examining the components of these best candidate equations. Considerations are made with regard to the nature of the selected predictors, in terms of both their scientific integrity, and equally important, their data availability. With regard to the former, favor is given to those equations whose component predictors are most consistent with the intuitive correlations based on discussions with experienced forecasters. With regard to the latter, favor is given to those equations whose predictors involve data which are most likely to be reliably available in a real-time data acquisition setting. This was done by looking at the historical availability of all predictors in the initial raw data set. Although this final equation selection has a large subjective component, it is important to remember that the selection is being made from amongst the top objectively-derived candidate equations whose difference in the error-based metric was virtually negligible.

3.2.2 "Day Typing"

An important element of the statistical forecast development process was a logical stratification of the data set. Rather than simply include all available stratus days into the training data set for statistical development, the training database was subdivided into two or three subsets (depending on the model), based on the characteristics of meteorological parameters which are known through experience to characterize the atmospheric "regime" during which the dissipation of stratus was occurring on a given day. This distinction has come to be referred to as "day-typing," i.e., trying to distinguish the type of day under which the current forecast is being made. For example, in forecasting practice it is typical to anticipate the dissipation evolution differently based on the general direction of the geostrophic (pressure-induced) wind field, wherein an onshore wind is more likely to retard the burnoff process, while an offshore wind is more likely to accelerate the process. During the equation development process, the training dataset was therefore subdivided accordingly in order to anticipate these tendencies. The intent was to develop a different set of equations based on these generalized regimes (i.e., day types) which were known to have characteristically different behavior. Often, this subdivision resulted in equation sets for "typical" versus "less typical" conditions, that are reflected in the resultant performance of the corresponding forecast equation sets.

The final result is that each statistical forecast model set (Local, Regional, Satellite) consists of a different forecast equation for each model run hour and, in some instances, different forecast equations for a particular run hour depending on the "day type" selection for that model and run hour. The nature of the day-typing utility is more evident in the individual forecast description overviews that follow.

3.3 FORECAST MODEL SUMMARY DESCRIPTIONS

3.3.1 Local Statistical Forecast Model (LSFM)

The Local Model is considered "local" in that it relies heavily on observations from the two specially instrumented sensor sites at SFO and San Carlos Airport (SQL), in close proximity to the Approach Zone. In particular, the predictors selected by the LSFM focus heavily on cloud height/depth and the amount of solar radiation reaching the ground. Additionally, the local model also tends to include the regional geostrophic wind as an important "derived" predictor, as measured by the pressure differences across regional surface station pairs.

Due to its high dependence on predictors associated with the depth of the local cloud layer, the day-typing stratification by the LSFM is associated with the so-called "quality" of the marine inversion height estimate. The base of the inversion height is presumed to be an estimate of the cloud top height; when this height is reliably estimated, the performance of the LSFM tends to be better. Thus, the quality of the inversion height estimate (particularly as derived from the sharpness of the sodar signal) was used as a discriminator in developing the LSFM forecast equations. Separate forecast equations were developed for good inversion height estimate quality versus inferior quality days.

One of the strengths of the LSFM is its reliance on data that is measured with high precision and high temporal resolution, specifically the sodar inversion heights and the pyranometer radiation measurements.

Conversely, it is hindered by the fact that these measurements are only available at two key locations near the approach zone, rather than on a broader regional basis. As a result, the LSFM is expected to perform best on days when the stratus dissipation process is dominated by the local radiation effects, and less by the external influence of significant advective cooling or heating.

3.3.2 Regional Statistical Forecast Model (RSFM)

The Regional model derives its predictor equations from routinely available weather observations extending beyond the immediate San Francisco Bay area. The model relies heavily on standard hourly NWS surface observations from the Central California region, and the twice-daily vertical sounding from Oakland.

One aspect of the Regional Model which distinguishes it from the other models is its reliance upon multiple pairs of stations for computing pressure differences as a measure of geostrophic wind. As a result, this model is best suited to account for anticipated changes to the wind field which impact the advective component of the cloud dissipation process. Other preferred predictors include the time of first cloud ceiling observation overnight in the Bay region, the height of the cloud ceiling at SFO, and the height of the inversion as determined by the Oakland sounding. Like the other statistical models, the RSFM also takes into account the maximum solar radiation available on a given day based on the sun angle which varies throughout the season.

The “day-typing” used for the RSFM involved the geostrophic wind. Essentially, the model categorizes a given day as having an onshore, offshore, or neutral geostrophic wind, with a separate set of forecast equations available for each condition.

Since the RSFM uses only routinely available observations, it is not reliant upon data from specialized sensors. Default forecast equations, utilizing different combinations of surface observations, have been developed which makes the model less vulnerable to the lack of availability of any single surface observation.

3.3.3 Satellite Statistical Forecast Model (SSFM)

For cloud depths typical of marine stratus (usually less than 1000 feet), the cloud depth is roughly linearly proportional to the liquid water content in the cloud. The Satellite Model correlates visible satellite brightness imagery with historical times of transition to dual approaches to estimate the time of cloud dissipation in the approach zone. The SSFM is run hourly between 15 GMT and 18 GMT, when visible satellite imagery is available.

The 1-km resolution visible satellite image from central California is filtered (to eliminate small-scale structure), registered (to ensure proper geographical orientation), and normalized to eliminate the effect of sun angle on cloud image brightness. The normalization for sun angle allows for an acceptable correlation between brightness and liquid cloud water.

Once the data are pre-processed, they are divided into 52 geographical sectors, primarily based on topography. Thus, each sector may contain roughly 10 to 30 data points. Within each sector, statistics are computed to estimate the mean cloud brightness and the mean area cloud coverage within the sector. These sector statistics thus represent 52 geographical data points for input to the SSFM. Through the model development procedure, the best-correlated sectors are chosen, with the appropriate predictor coefficients derived separately for each hour.

As with the other statistical models, predictor equations are developed for various “day types.” For the satellite imagery, the discriminating factors were found to be the height of the base of the cloud, and the “roughness” of the satellite image, i.e., the variance of the cloud brightness seen in the image. More typical stratus days are associated with cloud decks showing a smooth, regular texture. Fragmented or irregular cloud imagery tends to indicate the impact of a synoptic scale feature (such as a passing trough or weather system) and/or the presence of multiple cloud layers, which may disrupt the more typical radiation-driven dissipation process.

To an even greater degree than the Regional Model, the Satellite Model has the advantage of sampling data from an extensive geographic area. However, since it relies on visible imagery, it cannot be run prior to sunrise. Also, it is susceptible to any satellite data calibration errors

3.3.4 COBEL Model

COBEL is the only model amongst the four component models that was not statistically derived. It is a very high-resolution one-dimensional (vertical column) physical model of the planetary boundary layer that simulates the evolution of the cloud dissipation process. It was adapted from a model developed for forecasting fog behavior in northern France. (COBEL is short for the French **Co**uche **B**rouillard **E**au **L**iquide.)

The model is initialized with a vertical profile of temperature, humidity, and wind at SFO. The initialization uses a hybrid of the Oakland (OAK) balloon sounding and high resolution lower atmosphere measurements from the special sensors deployed at SFO. This profile is used to characterize the stratus cloud deck in terms of total vertically integrated liquid water content. As the model steps forward in time, it simulates the physical processes associated with the cloud evaporation process. The model forecast time of cloud dissipation at SFO is declared when the cloud liquid water content reaches zero.

Since the model is designed to forecast for a single point location (SFO), the forecast must be adjusted to account for the spatial difference from the specific area of interest, i.e., the approach zone into SFO. This is done via historical correlation between the time of cloud dissipation at the point location (SFO) and the time of transition to dual approaches.

The model is re-initialized for each model run, using updated high resolution measurements of surface temperature, humidity, wind and solar radiation. In particular, the re-initialization (during daylight hours) includes an adjustment to ensure that the model’s interpretation of liquid cloud water amount is consistent with the amount of solar radiation transmitted through the cloud layer. The model is then re-run, providing a new forecast of stratus dissipation.

The operational version of COBEL for San Francisco is designed for typical stratus days when stratus is present and fully developed at or before sunrise, when there is a marked strong inversion at the top of a shallow well-mixed marine boundary layer, and when no clouds are present above the boundary layer. If a reliable estimate of the inversion base height is not available, the model will not be allowed to run within the real-time system. Also, the current version of the model adapted for the SFO application assumes no horizontal advection and a fixed value for subsidence. A horizontal pressure force, corresponding to a 2 m/s geostrophic wind, is arbitrarily applied over the entire depth of the model domain (surface to 1.5 km height) and kept constant with time. As a result, the model is not expected to perform as well on days when the influence of temperature advection is large compared to radiation effects.

3.3.5 The Consensus Forecast

The Consensus Forecast combines the component forecasts to provide a single forecast of the time that stratus is expected to clear the approach zone to allow parallel approaches of aircraft. It is computed by assigning a weight to each of the component forecasts. This weight is derived from the historical performance of each component model, evaluated separately for each of their respective day types, and for each individual initialization hour. Thus, the weights are assigned based on the current initialization hour and the current day type being employed by each component model.

The consensus forecast time 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.” Any one of the following four conditions will trigger the LOW confidence indicator:

1. The inversion base height is not clearly identifiable. This may occur when the inversion is weak or diffuse. Since the Local Model and COBEL rely heavily on the accurate inversion height, these two component models are not allowed to run within the real-time system without a reliable inversion height estimate.

2. There is an extraordinarily high cloud ceiling base. Statistically, model performance (particularly the Satellite Model) has been found to be degraded when the stratus ceiling height is above 2300 feet.

3. The cloudiness in the Bay area appears disorganized or patchy. This often occurs when there is a transient synoptic scale feature impacting the mid-California coast, such as the approach and passage of a trough or precipitating weather system. This is usually accompanied by transient cloud decks or multi-layered clouds, which the forecast models do not handle very well. Algorithmically, this condition is detected automatically in the real-time system when the variance of cloud brightness within the region exceeds an empirically derived threshold.

4. Less than three component models are available.

In addition to generating a discrete forecast time, the Consensus Forecast is also used to generate a probabilistic forecast of clearing prior to specific target times during the high traffic demand period, namely 17, 18, 19, and 20 GMT. The probabilistic forecast values are derived empirically by comparing historical consensus forecast times with actual verification times.

3.4 FORECAST MODEL PERFORMANCE SUMMARY

Development of the forecast models was an iterative process that began in 2000, using a training data set 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 and 2002, the performance of the models was monitored; during the same period, the entire model development process (particularly for the statistical models) was streamlined and automated. This allowed for more rapid analysis of model performance and modification of models to allow testing of new versions. Four different trial versions of the three statistical models were examined during that two year period. 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 a re-work of the Satellite Model, after detection of a processing error in the raw data that needed to be corrected.)

Thus, the final version of the statistical models were developed during the winter of 2002-2003, using a training data set of stratus days from 1996-2002. Once the final models version of the models were established, a two-fold cross-verification procedure was performed to provide a quasi-independent assessment of expected model performance. The assessment was not truly independent, since the two validation data sets were subsets of the training data set. This assessment was performed in the following manner: The training data set was split into two subsets. The model equation coefficients for each predictor was then re-computed (via least squares) using one of the two data sets, and then scored using the other training set. This process was repeated, swapping the training data subset with the scoring subset. The final error estimates were made by averaging the errors from the two scored subsets.

Although not a truly independent scoring exercise, this process allowed a more realistic estimate of the expected performance of each of the models at each initialization time. These results are shown in the left column of Table 1, labeled “Development.” The table presents three sets of statistics. First is the Median Absolute Error (in hours) 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 performance statistics for these two seasons combined are shown in the right column of Table 1.

For a baseline comparison, error statistics for forecasts using conditional climatology are shown in Table 2. Conditional climatology presumes a forecast for each run hour that is simply computed as the average verification time for all days that had stratus clouds present at that hour. In essence, these baseline error statistics represent the minimum acceptable level of skill that would need to be demonstrated by the forecast system.

TABLE 1
Model Performance Statistics for Development Data Set (left column) and Independent Demonstration Data Set (right column)

Development (1996-2002)

Median Absolute Error (hrs)

	Hour (GMT)						
	09	11	13	15	16	17	18
Consensus	0.77	0.77	0.71	0.59	0.49	0.42	0.57
Cobel	0.87	0.82	0.79	0.60	0.51	0.52	0.49
Local	0.67	0.71	0.61	0.57	0.47	0.45	0.53
Regional	0.80	0.86	0.77	0.77	0.72	0.66	0.64
Satellite	#	#	#	0.53	0.40	0.40	0.40

Bias (hrs)

	09	11	13	15	16	17	18
Consensus	0.05	0.05	0.02	-0.03	-0.01	-0.03	-0.03
Cobel	0.14	0.12	0.09	-0.01	0.05	0.07	0.04
Local	0.14	0.13	0.10	0.09	0.04	0.08	0.12
Regional	0.24	0.24	0.18	0.15	0.21	0.13	0.13
Satellite	#	#	#	0.12	0.08	0.07	0.22

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 (hrs)

	Hour (GMT)						
	09	11	13	15	16	17	18
Consensus	0.97	0.89	0.76	0.75	0.67	0.63	0.59
Cobel	0.94	0.78	0.73	0.71	0.62	0.65	0.90
Local	0.99	0.80	0.74	0.76	0.77	0.87	0.81
Regional	0.96	1.03	0.84	0.89	0.74	0.51	0.45
Satellite	#	#	#	0.73	0.60	0.53	0.59

Bias (hrs)

	09	11	13	15	16	17	18
Consensus	0.34	0.47	0.37	0.49	0.43	0.53	0.62
Cobel	0.00	0.14	0.20	0.22	0.31	0.47	0.69
Local	0.21	0.56	0.50	0.42	0.34	0.86	0.77
Regional	0.40	0.58	0.27	0.45	0.33	0.01	0.28
Satellite	#	#	#	0.59	0.51	0.54	0.73

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	#	#	#	123	112	91	56

TABLE 2
Median Absolute Error (hrs) of Conditional Climatology Forecasts for Each Model Initialization Hour

Hour (GMT)	09	11	13	15	16	17	18
Conditional Climatology	1.05	1.05	1.05	1.00	1.00	0.87	0.75

As would be expected, the independent data set did not score as well as that derived from the training data set. In general, however, the median absolute errors were comfortably lower than conditional climatology, particularly at the key tactical forecast hour of 16 GMT which represents the last forecast hour for which there is sufficient leadtime to cancel an active delay program prior to the late morning ramp-up of scheduled arrival traffic. The independent data set for 2003-04 also exhibited a late forecast bias. It is presumed that this was largely impacted by the 2004 season which consisted of a heavy early bias (more than 0.30 hours) in the verification distribution. This was evidenced by the consistency of the forecast bias across all four of the component forecast models, including the physics-based COBEL model.

4. PRODUCT DISPLAY

4.1 OVERVIEW

The system display is the primary interface for operational users to access observations and model forecasts as an aid toward anticipating the time of stratus dissipation in the Approach Zone into SFO. The primary display is part of a two-node Local Area Network located at the Oakland Center CWSU in Fremont, CA, which hosts the system base station computer for data acquisition and product generation. The base station also acts as a web server allowing remote users to access the display via the World Wide Web. The display is both automated and interactive. It automatically refreshes with the most recently available information, and it allows the user to investigate the details of each element as desired. In particular, it includes an on-line archive of all observations, forecasts, and verifications for a minimum of one year.

The display (Figure 4) includes both a data observation frame and a forecast frame. The observation frame provides a graphical suite of weather observations as an aid in proper diagnosis of the current stratus-related conditions. The forecast frame provides both deterministic and probabilistic forecasts of the time of stratus dissipation, based on four individual forecast models. These four “component” forecasts are combined to generate a single “Consensus” forecast. During the summer stratus season (May through October), forecasts are generated every two hours each morning from 09 GMT to 15 GMT, and then hourly until 18 GMT. The forecast frame also allows the user to review and analyze various elements of the forecasts, such as the specific parameters within each model. Additionally, the forecast frame allows the user to launch a separate web browser that allows access to view all observations and forecasts for any individual day within the past year.

The remainder of this section provides a brief summary of the various components within the observation and forecast display frames. Complete details of the display products and interaction with the display are provided in Section IV: Display System.

4.2 OBSERVATION FRAME

The left frame on the main display page is used to show current observations, namely:

1. surface weather observations
2. sodar inversion base graphs
3. visible satellite imagery
4. solar radiation plots

The time resolution of the data observations ranges from 1 to 15 minutes, depending upon data type. Each individual observation product receives new data roughly every 15 minutes. The display automatically checks every 2 minutes to see if any of the products has received new data, and automatically refreshes the entire observation frame. Due to the time required to collect and process remote data observations, the latency of the most recent data viewed on the display will typically range from 10 to 25 minutes.

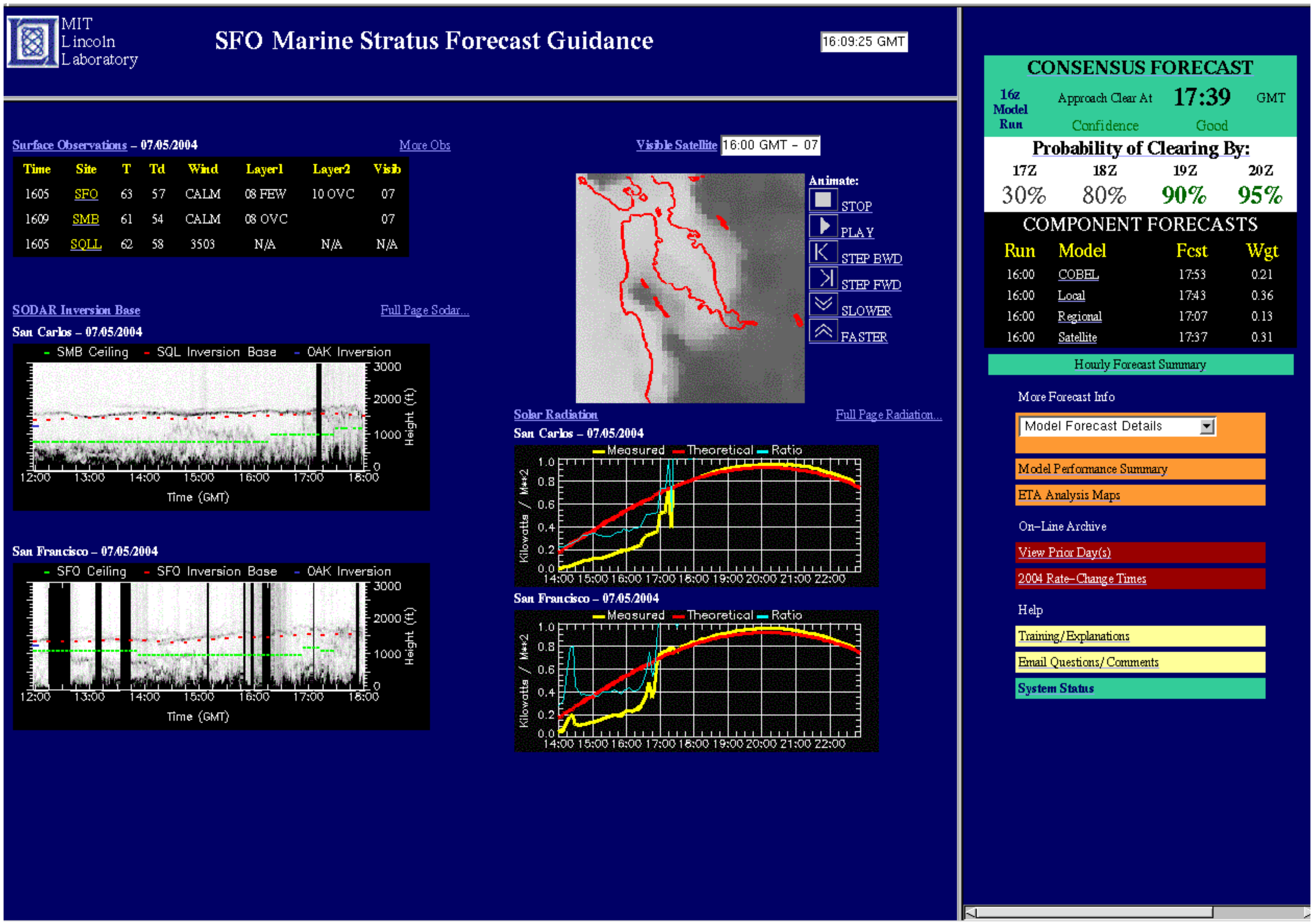


Figure 4. SFO MSF display.

4.2.1 Surface Observations

The main page shows text surface observation reports of temperature, dew point, wind, cloud amount for three key surface station locations (SFO airport, San Mateo Bridge (SMB), and SQL airport). The user can also select "More Obs" to view current and previous observations from the standard network of NWS stations from the Central California region.

4.2.2 Sodar Inversion Base Graphs

The Sodar Inversion Base product (Figure 5) provides a continuous monitoring of the local inversion base height which defines the top of the stratus deck. The acoustic sodar (SOnic Detection And Ranging) is an upwardly pointing parabolic antenna that emits an audible signal whose return signal is proportional to the vertical gradient of air density (and therefore temperature) in the atmosphere. On the sodar plot, the temperature change representing the inversion appears as a dark quasi-horizontal line above the 500-foot level of the time-height plot. The red dashed line on the sodar plot is an automated estimate of the inversion base height. This estimate is also used as input to automated forecast algorithms. The green dashed line is the ceiling height from the nearest surface reporting station to the sodar. The blue line displayed at 1200 GMT is the Oakland inversion determined from the sounding data.

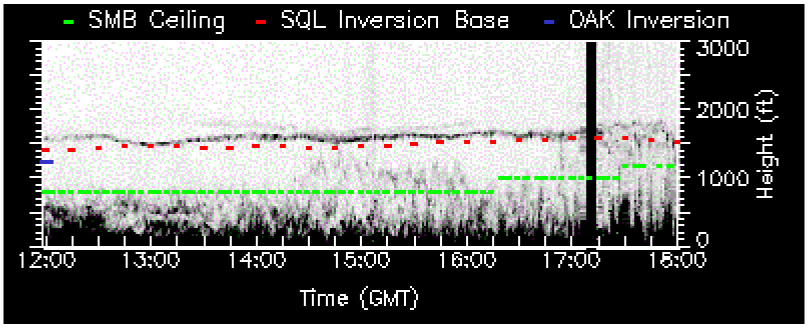


Figure 5. Example of sodar image.

4.2.3 Visible Satellite Imagery

The satellite display product is derived from the 1 km resolution western US GOES visible satellite channel. It is intended to provide a general close-up view of the cloud coverage and brightness in the San Francisco Bay Region, centered on the Approach Zone. An updated satellite image is available every 15 minutes. Unlike conventional displays of visible satellite imagery, the satellite data used for the display are "normalized" to eliminate the effects that the changing sun angle has on the brightness of the cloudiness. The result is that any changes viewed in the cloud brightness are primarily attributable to the actual physical characteristics of the cloud. For a stratus cloud deck ranging in thickness from 200 to 1200 feet (typical of the SFO Bay region), the normalized cloud albedo is roughly linearly proportional to the total liquid water content within the cloud, so that changes in the visible brightness can be directly attributable to changes in cloud liquid water. In addition to brightness normalization for sun angle, the satellite data also undergo a geographic registration process to minimize "jumpiness" in viewing a sequence of cloudiness, and a simple interpolation and filtering process to eliminate irregular high frequency changes in cloud coverage and brightness that may make it difficult to analyze trends in the cloud coverage during the burn-off process. Since this is a visible satellite image it is only available after sunrise. The satellite images can be viewed as a movie loop to observe the trend of the cloud change.

4.2.4 Solar Radiation Plots

The Solar Radiation product is an image showing the amount of shortwave radiation reaching the earth's surface (Figure 6). This value is plotted as a yellow curve. The red parabolic curve indicates the amount of radiation that would reach the ground on a cloud-free day. This varies by location, day-of-year, and time-of-day. The thin blue curve on the plot shows the ratio between the observed amount of radiation transmitted though the stratus layer and the theoretical amount that would reach the ground on a cloud-free day. This ratio is significant in that the fraction of transmitted solar radiation is roughly proportional to the liquid water content (integrated vertically) within the cloud layer.

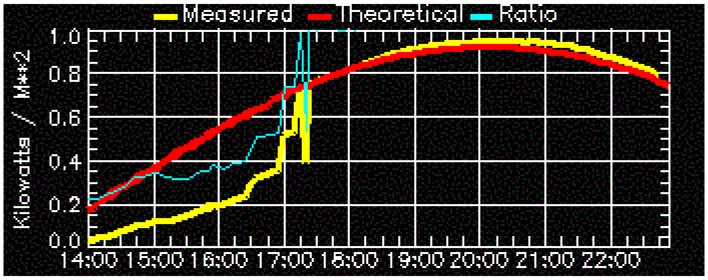


Figure 6. Solar radiation image.

4.3 FORECAST FRAME

The forecast frame is located on the right side of the main display (Figure 4). In the upper portion of this frame is the forecast box, which displays the four component model forecasts (described previously in Section 3), as well as the Consensus Forecast. The forecasts are updated automatically each morning at 09, 11, 13, 15, 16, 17, and 18 GMT. Each component model forecast clearing time is accompanied by a "weight" indicating the relative contribution of that forecast toward the consensus forecast.

In addition to the discrete forecast times, the forecast box also presents a probabilistic representation based on the consensus forecast time. The probabilities indicate the likelihood that the Approach Zone will be clear by specific target times that are relevant to the air traffic decision process, namely 17, 18, 19, and 20 GMT. These times represent the period during which the heaviest arrival traffic is due into SFO, and traffic managers must decide to implement, maintain, or cancel ground delay programs based on anticipated available capacity.

At the bottom of the forecast box is a link titled "Hourly Forecast Summary." This link displays a table of each of the model forecasts throughout the morning. On the primary display at the Center Weather

Service Unit, there is an additional button that enables the CWSU forecaster to indicate the time that the approach zone was cleared to accept dual approaches. This button is not visible on remote displays accessed via the Internet.

Once the stratus has cleared from the approach zone, the forecast models cease to run. A text message will appear above the forecast box to indicate that the approach zone is cleared.

Below the forecast box there are a number of additional links to provide the user with more information that may be used as forecast aids. These links are described briefly in the following paragraphs.

4.3.1 Model Forecast Details

A pull-down menu is provided that allows access to more detailed information regarding the current model forecasts. The user has the option to open a new browser to view a Detailed Forecast Summary, a Raw Model Output Summary, or the Key Model Parameters. These options provide additional details regarding the forecasts that have been made throughout the morning, including the specific parameters that comprise the individual component forecast models.

4.3.2 Model Performance Summary

This link provides a table indicating the historical performance of the forecast models, to provide the user with a sense of both absolute and relative reliability of the forecasts.

4.3.3 NAM (formerly ETA) Analysis Maps

This link launches a separate browser window for displaying 1 millibar North American Model (NAM, formerly ETA) model analysis maps that are accessed via the Internet. These are provided as an aid to understanding the regional dynamics, particularly in anticipating changes to the geostrophic wind.

4.3.4 Review Prior Day(s)

A key feature of the Marine Stratus Forecast System is the capability to review both observations and forecasts for previous days during the current or past stratus season. This is often useful for finding analogous forecast situations, or for investigating and understanding forecast model performance. Selection of this link launches a separate browser that provides text and graphical observations and forecast for a prior day, selectable via a pulldown menu of year/month/day.

4.3.5 Rate Change Times

This link provides an up-to-date summary table of side-by approach times for the current stratus season. Listed are the first time on each day that the Arrival Rate was increased to 45 planes per hour and 60 planes per hour. The first time during the day that the Arrival Rate increased to at least 45 planes per hour is considered the "Verification Time" for the Approach Clearing time being provided by the automated forecast guidance models.

4.3.6 Training/Explanations

This link launches a separate browser window that provides helpful information for understanding the observations and forecasts provided by the MSF System. It is organized in an outline fashion, providing a table of contents corresponding to the order in which items appear on the display. Selection of an individual item from the table of contents will bring the help page directly to that item. Additionally, within the main display page, each underlined title is also linked to the appropriate location in the help file.

4.3.7 System Status

This button changes color based on the freshness of the San Francisco sodar data. If all data are less than one hour old then the button is green. If any data are greater than an hour old but less than two hours old then the button turns yellow. If older than two hours, the button turns red.

When this link is clicked a new browser window opens a table indicating the status of the various data sources within the real-time system. The current status is determined by how recently data from that source have been received, which is also indicated in the table. At a glance, the user can determine how current or how old the data are by the color coded words. The word “Good” is in green and it means that the data are less than an hour old. The word “OK” is in yellow to indicate that the data are more than an hour old but is less than two hours old. If the data are older than two hours, the word “BAD” appears in red.

GLOSSARY

AWRP	Aviation Weather Research Program
ARTCC	Air Route Traffic Control Center
CWSU	Central Weather Service Unit
DCP	Data Collection Platform
LAN	Local Area Network
LSFM	Local Statistical Forecast Model
MSF	Marine Stratus Forecast
NAM	North American Model
NRL	Naval Research Laboratory
NWS	National Weather Service
NWSFO	National Weather Service Forecast Office
PGAM	Pre-scaled Generalized Additive Modeling
PPP	Point to Point Protocol
RSFM	Regional Statistical Forecast Model
SFO	San Francisco International Airport
SMB	San Mateo Bridge
SODAR	Sonic Detection and Range
SQL	San Carlos Airport
SSFM	Satellite Statistical Forecast Model

SFO Marine Stratus Forecast System Documentation

Section II. Sensor Maintenance

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. INTRODUCTION

1.1 SCOPE

This section provides information regarding the weather sensors and data collection equipment at the Marine Stratus Forecast system field sites. It includes a listing and description of the installed equipment, and guidelines for maintaining the equipment and identifying equipment failures and deficiencies. More detailed description and maintenance guidelines for some of the equipment are provided in the manufacturers' documentation, which is referenced within this document.

1.2 SYSTEM CONFIGURATION

The SFO Marine Stratus Forecast System relies on a variety of weather observations acquired from land-based sensors throughout the Bay region, upper-air data via balloon-launched rawinsondes, and GOES-West satellite data. Data from these various sources are transferred to a base station computer located at the Air Route Traffic Control Center (ARTCC) in Fremont, CA at a time interval of 15 minutes. There the data are processed for display, and for input into a suite of models designed to forecast the time of Approach Zone clearing. The display of observations and the automated forecast guidance are made available to users via the World Wide Web.

The geographical configuration of sensors which supply data to the system is shown in Figure 1. There are two primary sensor sites, located at San Francisco International Airport (SFO) and San Carlos Airport (SQL) along the Approach Zone. These two sites were instrumented specifically to support the forecast system. Each site includes a Sonic Detection and Range (sodar) instrument to estimate cloud top height, a ceilometer to estimate cloud base height, a pyranometer to measure incoming solar radiation, and instrumentation to provide high resolution (in time and accuracy) measurements of temperature, humidity, and wind. Most of the remainder of observations are acquired from the standard suite of weather observations made widely available by the National Weather Service. This includes hourly surface observations of standard meteorological parameters (temperature, humidity, wind, cloud amount and heights), twice-daily vertical measurements (temperature, humidity, and wind) from the balloon rawindsonde launched from Oakland, and satellite data from GOES-West available at 15-minute intervals. The other key significant observation of note is the FAA-maintained surface weather sensor located at the San Mateo Bridge, situated directly below the primary runway approach zone into SFO airport. This sensor provides the key ceilometer measurement of cloud amount and cloud base height used for operational purposes in assessing the availability of visible approaches into SFO.

Figure 1 also shows the location of the system base station computer located at the Oakland Air Route Traffic Control Center (ARTCC) in Fremont, CA. The base station computer serves the function of collecting and processing data from the variety of sources and generating display products. In additional to local display, the base station computer acts as a web server for delivering the display products to external system users via the Internet.

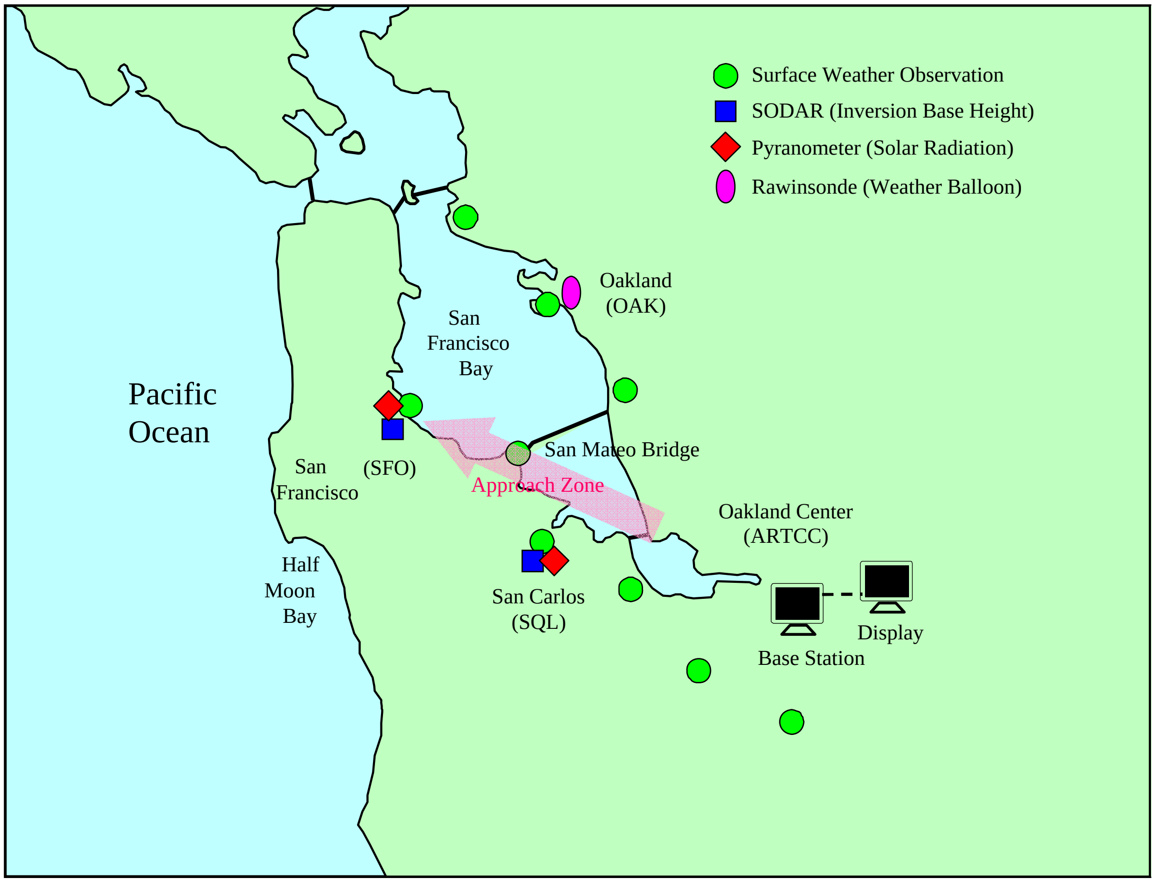


Figure 1. SF Bay area sensor locations.

1.3 DATA ACQUISITION AND REAL-TIME PROCESSING

Figure 2 shows a schematic representation of the hardware and communications configuration for data acquisition and processing. The system encompasses three locales: two field sites equipped with identical hardware and software and located at San Francisco International Airport and San Carlos Airport, and a two-node local area network (LAN) housed within the CWSU. The LAN is composed of a base station computer used for data collection and processing, and a display computer used by forecasters to view forecasts of marine stratus burnoff within the approach zone as well as near realtime field-sensor observations and derived products. The base station acquires data from both the field sites and other sources using a commercial DSL connection to the Internet. A backup means to acquire field site data is available if needed, using a dialup Point to Point Protocol (PPP) connection over a phone line that is shared with DSL. The dialup connection is shown in Figure 2 as the dashed line connecting field site modems to the base station modem. Five-minute ASOS surface observations for San Francisco Airport are acquired through a dialup connection using the Kermit protocol. Five-minute Automated Weather Observing System (AWOS) surface observations from an FAA-managed site located at the San Mateo Bridge enter the base station via a serial RS232 line that originates within the confines of the CWSU weather desk. The Oakland upper-air sounding and hourly surface observations from stations in and around San Francisco Bay are provided by the National Weather Service Forecast Office (NWSFO) in Monterey and satellite data are retrieved from the Naval Research Laboratory (NRL), also located in Monterey. These data sources rely on the DSL connection for transfer of data to the base station.

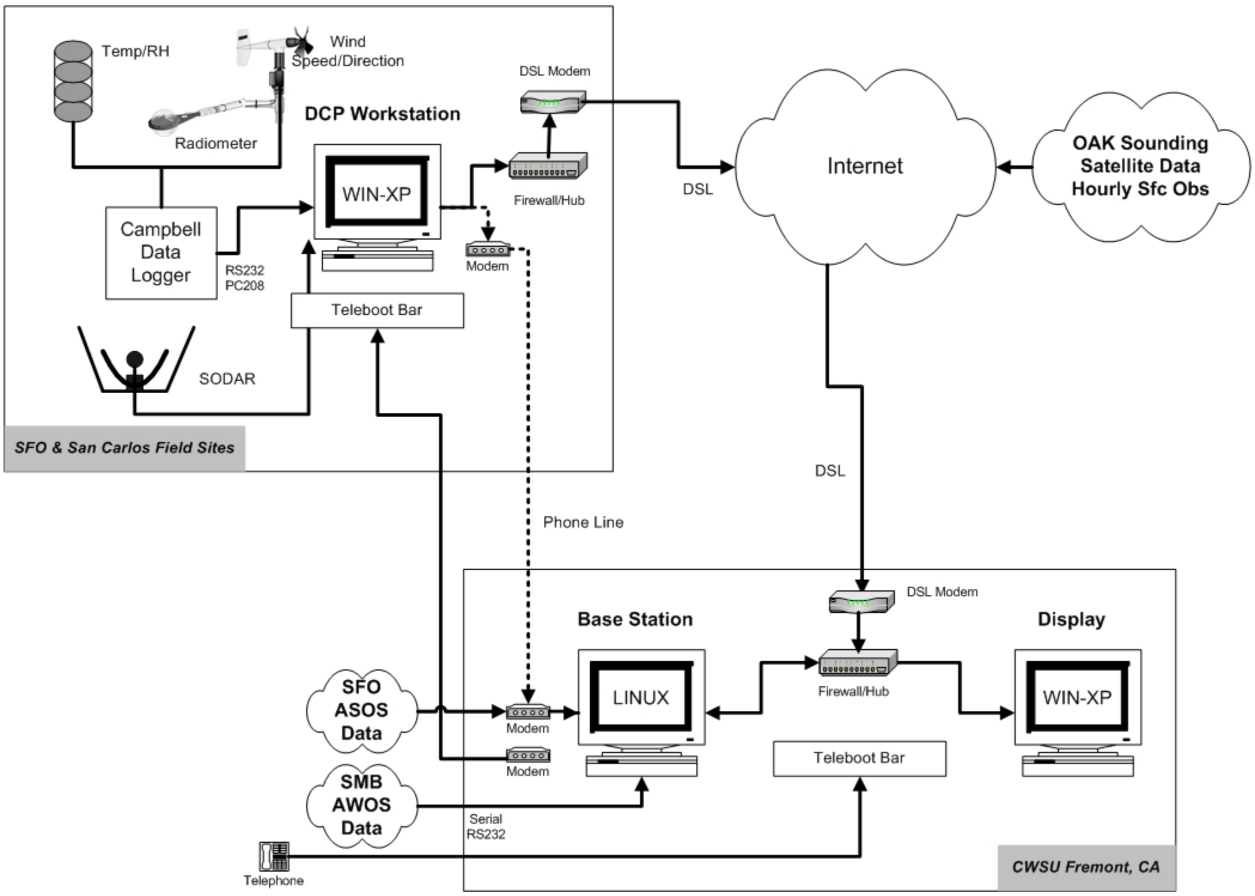


Figure 2. System diagram depicting the San Francisco Marine Stratus Forecast System.

Field site instrumentation, including temperature and humidity sensors, a pyranometer, and wind sensor, are tower mounted and connected to a data logger that prepares the sensor data for acquisition by the Data Collection Platform (DCP), a PC running the Windows XP operating system. The DCP also collects and processes data from an acoustic sounder that profiles the atmosphere just above the field site.

Both the field sites and the LAN at the CWSU are protected from Internet intrusion through inclusion of a firewall device. The firewall is configurable through a user interface to allow only permitted inbound traffic for remote login. The firewall does not limit outbound access. At the CWSU, the firewall also acts as a hub, connecting the base station and display computer in a LAN configuration.

Finally, all three locations are configured with teleboot bars. A teleboot bar allows a user or a computer to dial into the device and cycle power to any device attached to the bar. In Figure 2, a dedicated modem at the CWSU provides the base station with the ability to dial into either field site and clear a malfunctioning device without a person-in-the-loop. A teleboot bar is also located at the CWSU to provide maintenance personnel the ability to dial in to the CWSU and reboot the base station or communications equipment.

2. FIELD SITE EQUIPMENT DESCRIPTION

2.1 GENERAL

The specific field site equipment relevant to this documentation is depicted by the box in the upper left of the system diagram shown in Figure 2. The equipment at each of the two field sites is essentially equivalent. The two main subcomponents are a ground based sodar (SOnic Detection And Ranging) that is shielded by an acoustic baffle and attached to a data collection computer, and a tower-mounted weather sensor suite used for high precion measurements of temperature, humidity, wind, and solar radiation. Figure 3 is a photograph of the field site equipment.

Figure 3. Sensor field site equipment. Foreground is sodar baffle anchored to cement pad. Background shows tower with mounted sensors. Electronics box is seen at base of tower.

2.2 SODAR AND DATA COLLECTION PROCESSOR (DCP)

Under contract with the National Weather Service, this subsystem was designed, built, and installed on site by the NOAA Environmental Technology Laboratory (NOAA/ETL), as an upgrade to the prototype sodar system that was used during system testing. Detailed documentation of this subsystem was provided to NWS as part of this contract[1].

Functionally, the sodar is an upward-pointing parabolic antenna (Figure 4) that emits an audible signal in the 2 kilohertz range. The signal is sensitive to vertical changes in the refractivity of the atmosphere. This refractivity varies with air density, which in turn is related to air temperature. Thus, the return signal from the sodar is intended to provide an indication of significant vertical changes of temperature in the lower atmosphere. In the Pacic stratus environment, this temperature gradient denotes the base of the marine inversion, the point vertically where the temperature begins to increase with height. Meteorologically, this height estimates the top of the marine layer, identifying the top of the marine stratus cloud deck. The sodar has an effective vertical range for detecting this phenomenon up to approximately 4000 feet.

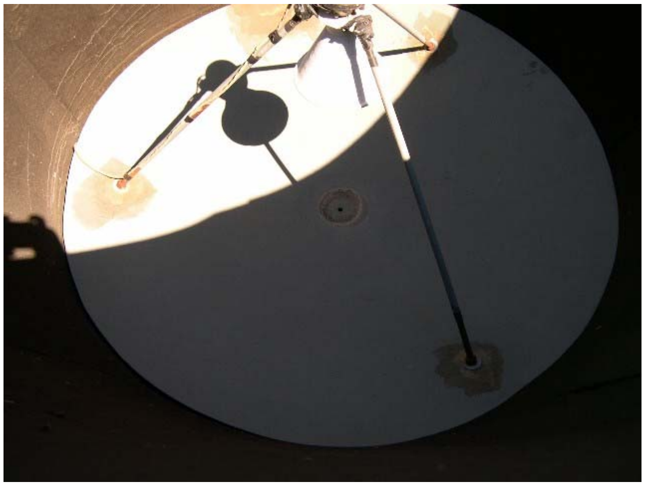


Figure 4. Sodar antenna viewed looking down into baffle.

The sodar portion of the subsystem consists of hardware and software in the DCP, a power amplifier, and a filter. These units are installed in the equipment shed. They are connected to the antenna system by a shielded cable. The Sodar system installation, operation, and service requirements are described in detail in the documentation provided by NOAA/ETL.

A personal computer (PC) running the Window XP operating system serves as the Data Collection Processor (DCP), and has three major tasks:

[1] Real-time Digital Sodar System, San Francisco and San Carlos Airport Systems, NOAA/ETL

1. data collection from the collocated Campbell data logger,
2. Sodar operation, including collection of the Sodar data, and
3. data communication to the base station for all sensors at the site.

This computer system and its maintenance are also described in detail in the documentation provided by NOAA/ETL. The ETL point of contact is Jim Jordan (303-497-6303).

2.3 WEATHER SENSORS AND DATA LOGGER

This subsystem includes the weather sensors and a data logger for collecting the weather data. It was designed by REBS, Inc. Detailed information regarding sensor installation, operation, and service requirements is available in a separate document provided by the manufacturers.

The data logger system, which prepares the sensor data for acquisition by the DCP, is contained in an electronics box (Figure 5), which houses four components:

1. a Campbell data logger,
2. a multiplexor,
3. a lightning arrestor,
4. and an optical isolator.

Figure 5. Data logger system housed inside equipment box.

The Campbell Data Logger collects data from three weather sensors:

1. REBS, Inc. pyranometer
2. REBS Inc. temperature and humidity sensor
3. R.M. Young Inc. wind sensor

These sensors are mounted on a 10-meter tower. The wind sensor is a propeller-and-vane system mounted atop the tower (Figure 6); the temperature and humidity sensors are mounted lower on the tower. The pyranometer, also mounted lower on the tower (Figure 7), is a solar radiometer which measures the amount of downward shortwave radiation reaching the earth's surface. Since shortwave radiation is reflected and absorbed by water vapor, the amount reaching the earth's surface on a cloudy day (compared to a sunny day) provides an indication of the cloud depth.

Figure 6. Propeller and vane anemometer system mounted atop tower.

Figure 7. Pyranometer system mounted on lower portion of tower. Sensor is inside glass dome on left.

Detailed documentation for the various weather sensors is available from the manufacturers, as listed below:

Item(s)	Manufacturer	Phone
Data logger system, Pyranometer, Temperature/humidity sensor	REBS, Inc (Charles Fritchen)	206-624-7221
Wind sensor system	R. M. Young Company	231-946-3980

3. MAINTENANCE

This section provides guidance for the monitoring and evaluation of system operations and for routine maintenance. Detailed troubleshooting and repairs should be conducted in accordance with the manufacturer's documentation, cited in the previous section. Field maintenance involves conducting routine maintenance, and diagnosing problems. If the field solutions described herein do not correct the problem, then unit replacements may be advised.

3.1 DATA COLLECTION PROCESSOR (DCP)

The Sodar/DCP operates from one hard drive. A second hard drive, which is identically configured, is located in the computer, but is not connected to power or to the data bus. In case of a hard drive failure, the following steps should restore the system to operational status:

1. Turn off the computer, disconnect the power cord, and open the case.
2. Remove the failed hard drive (the one that is connected to power).
3. Check that the jumpers on the replacement hard drive are set to "Master."
4. Attach the power and data cables to the replacement hard drive.
5. Close the case, turn on the modems, and boot the computer.
6. The following settings are required:
    a. Set the Clock to GMT
    b. Confirm that the computer is seeing the modem on com 4 (the modem must be connected to the system before it is booted)
    c. Check the IP addresses and masking (same at both sites)
    Path: Internet Connections – Local Area Connection – Properties – General – Internet Protocol TCP/IP – Properties – General (IP 192.168.1.2 Mask 255.255.255.0 Gateway 192.168.1.1) (DNS Server Address: 206.13.28.12, alt. 206.13.31.12)
    d. Check Sodar display and set the Sodar Noise filters
        i. – Median Factor 1.25
        ii. – Median alpha for good data 0.06
        iii. – Median alpha for bad data 0.001
7. Check system operation - Data files being created in the Xfer directory (wait for 15 minutes: they should be created and then disappear)
    a. Sodar file: sodar.dat and sodar.zip
    b. Data logger file: datalggr.dat and datalggr.zip
8. Order a replacement hard drive from NOAA/ETL

In case of any other computer failure, the computer should be replaced with the spare DCP computer, and the failed unit returned to NOAA/ETL for repair. Remember that the settings from step 6 (above) must be installed in the replacement computer.

3.2 DATA INTEGRITY AND SENSOR MAINTENANCE

If the data collection and communication system is performing properly, then the integrity of the sensor data can be checked from the user display. This section describes the process used to visually evaluate the accuracy of the data from the various sensors.

The user main display page is shown in Figure 8. It contains data displays from all of the system sensors and from some collocated meteorological sensors. This display contains all of the data that are required for visually checking the integrity of the system data collection. If there is a question regarding whether a small discrepancy is large enough to warrant attention, a forecaster should be consulted. Sensor data from prior days can be viewed from "View Prior Day(s)" panels.

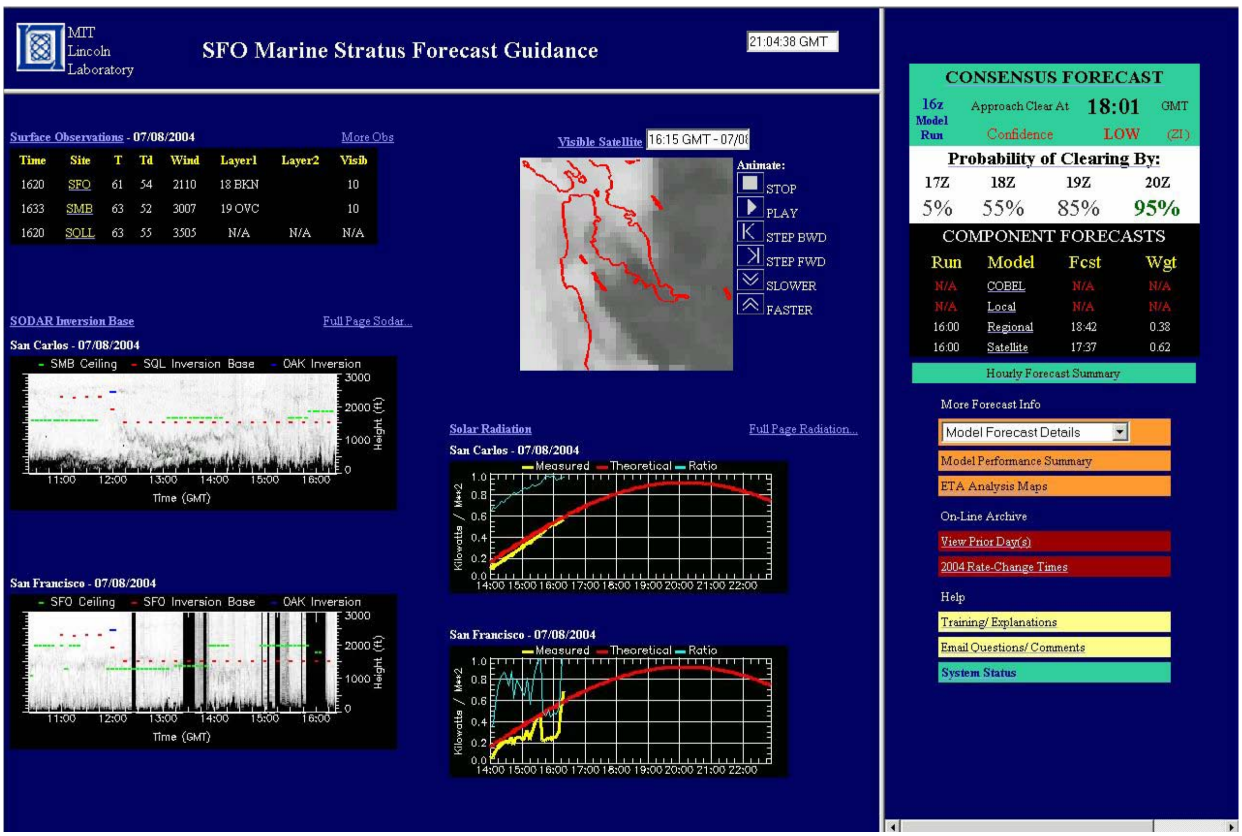


Figure 8. The Marine Stratus System Forecaster's display, available at the WFO and at the CWSU.

Sodar data are displayed in the lower left panels, *Sodar Inversion Base*. Pyranometer data are displayed in the lower central panels, *Solar Radiation*. Wind, temperature, and dew point data are displayed in the upper left panel, *Surface Observations*-More Obs, SFOL and SQLL, at the bottom of the list. These should be compared with the NWS systems, SFO and SQL.

3.3 SODAR MAINTENANCE

3.3.1 Performance Monitoring and Corrective Maintenance

There are two Sodar data displays, one for each Sodar (SFO and SQL). Examples of satisfactory displays are shown in Figure 9. Typically, the inversion base can be seen on this display as a dark, quasi-horizontal line between 1000 and 3000 feet. The red dashed line on the display product indicates an automated estimate of the inversion based height. Note that there can be a daily variation between a sharp image of a single layer and a complex image including multiple and fuzzy layers. These variations are due to normal changes in the atmosphere. There can also be days when there are no layers. Due to the frequent excessive aircraft noise at San Francisco Airport, the SFO Sodar display is typically less robust than the San Carlos display. There are also times when the automatic data quality analysis, which masks excessive airplane noise, will reject the Sodar signal for periods of time that may last for as long as ½ hour (solid vertical black bar on the display). These situations are normal, and do not require maintenance attention. Finally, when wind speeds are in excess of 25 knots, the Sodar signal may be lost totally, or may become cluttered (random speckling). The display should return to normal when the wind speed decreases.

If one Sodar has a nearly solid white or grey display, while the other has a good atmospheric signal, a site visit and antenna inspection is advised:

1. Check the acoustic foam. If it comes lose and blows over the antenna, this muffles the signal.
2. If the Sodar is beeping, but the signal is weak, check that the shield on the underground cable to the antenna is properly grounded.
3. If the Sodar fails to beep, then the Sodar transducer or the amplifier system may have failed. An ohm-meter can be used to check for an open voice coil in the transducer. A failed transducer should be replaced with a spare and returned to the shop for rebuilding.
4. If the amplifier system has failed, follow the Troubleshooting Instructions in the Sodar documentation. The amplifier has two channels; service may be restored by connecting the antenna leads to the alternate channel.

a)

b)

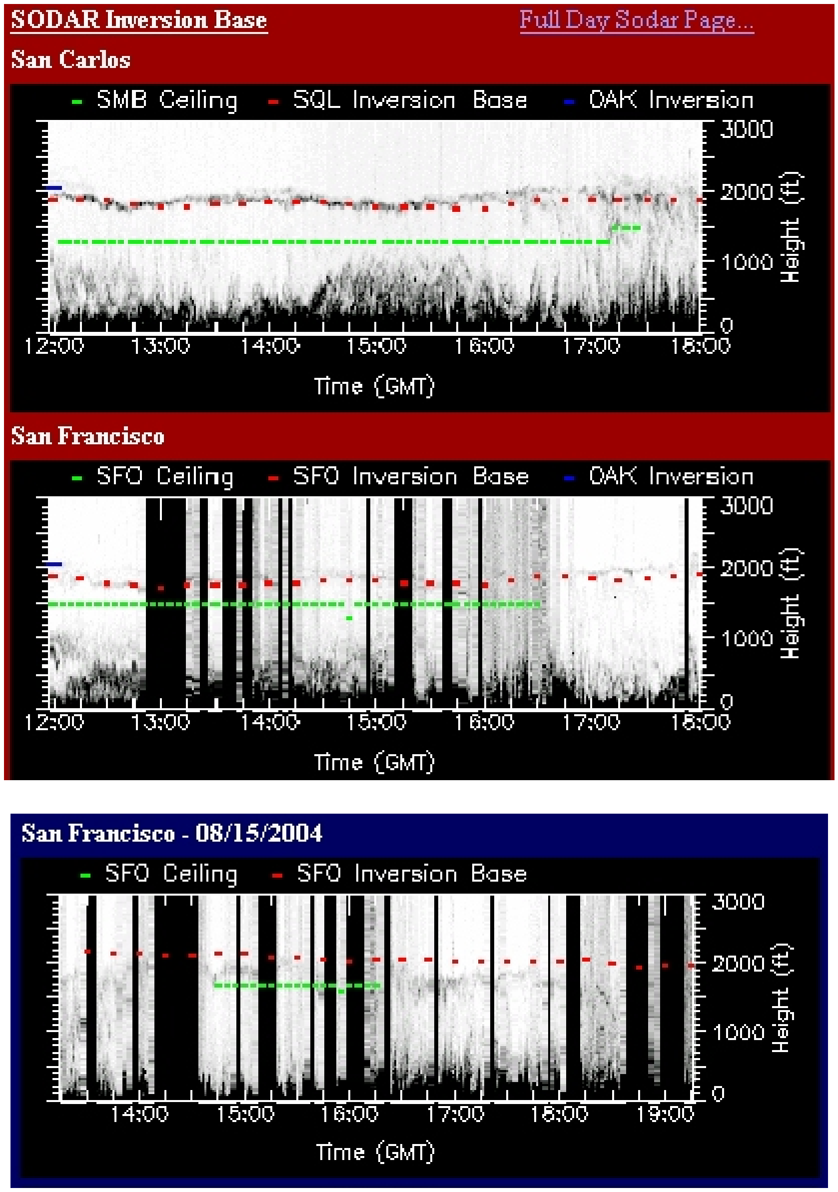


Figure 9. Typical sodar displays: a) both sodars; SQL is sharper than SFO, SFO affected by high winds at 16z, b) SFO on a day when the atmospheric signal is weak; note the substantial aircraft noise editing.

3.3.2 Scheduled Maintenance

Monthly: Conduct a visual inspection of the antenna, anchors, and foam.

Annual: Replace the Sodar transducer and rebuild the old unit.

Follow the instructions that come with the new diaphragm (voice coil). Cautions: 1) be careful to not over-tighten the screws, 2) be careful that the wires touch nothing but the contact posts, and 3) when mounted in place, don't short the transducer horn by letting it contact anything other than the mounting bracket.

Biannual: Replace the Acoustic Foam in the Antenna.

Weather and UV breaks down the foam, so replacement is needed every two years, preferably in the spring. Use 4-inch thick acoustic foam; five 2’ x 6’ sections are required. The old foam should be scraped off. The new side panels must be cut at a taper to form a solid cover for the sides of the baffle. The tapered scraps may be used to fill at the top of the sides. The floor under the antenna should also be covered. It is permissible to use old foam from the sides or scraps of new foam on the floor, so long as it is solidly covered (when the floor foam is removed, check that the drain holes in the floor are not clogged). The new foam should be secured to the sides of the baffle with contact cement (sparingly).
One Local Source: Bob's Foam Factory in Fremont, phone 510-657-2420.

3.4 DATA LOGGER SYSTEM

3.4.1 Performance Monitoring and Corrective Maintenance

Follow maintenance procedures in the REBS data logger and sensor documentation. Use REBS data logger analysis spread sheet for system checkout if problems arise. REBS will provide phone support to help isolate problems. Consider replacing the unit and returning it to REBS for refurbishing.

If the senor data are faulty, check that the Power Supply is providing sufficient voltage to the data logger. If the voltage drops below 12 v, then the data logger may appear to function, but it may collect faulty data from the sensors.

3.4.2 Scheduled Maintenance

Monthly: None

Annual: Visual Inspection, follow instructions in the REBS system documentation

Biannual: Remove Data Logger for calibration by Campbell Scientific

3.5 PYRANOMETER MAINTENANCE

3.5.1 Performance Monitoring and Corrective Maintenance

Inspect data weekly for consistency with ideal values by using the user display (Figure 10). The yellow line indicates the amount of solar radiation reaching the sensor. The red line indicates that expected radiation curve on a cloud-free day. The yellow line and the red line should be in close agreement when there are no clouds. The blue line is the ratio of these two. The blue line value should be near to 1.00 in cloudless conditions. If the blue line is consistently below 0.95 in cloudless conditions, then the instrument should be checked. Grime such as bird deposits on the dome can cause this condition. Since the red line does not account for all seasonal changes, minor disagreements are acceptable.

Follow maintenance procedures in the REBS, Inc. pyranometer documentation. Use REBS pyranometer analysis spread sheet for checkout if problems arise.

Note: When sky conditions consist of “broken” clouds, e.g., mostly cloudy conditions, the yellow line may “spike” above the red line. This is due to the fact that both direct solar radiation and indirect radiation (reflecting off the sides of the clouds) are simultaneously reaching the sensor. This is typical when the stratus cloud deck first begins to break up. Thus, this data spike is not indicative of a sensor problem. Once conditions become cloud-free, the yellow line should return to close agreement with the red line.

3.5.2 Scheduled Maintenance

Monthly:	Inspect aspirator fan (replace if it has failed) Inspect dome for excessive dust (wipe with clean, dry cloth)
Annual:	Replace aspirator fans Replace Desiccant
Biannual:	Sensor calibration and refurbish at REBS, Inc. every 2 years

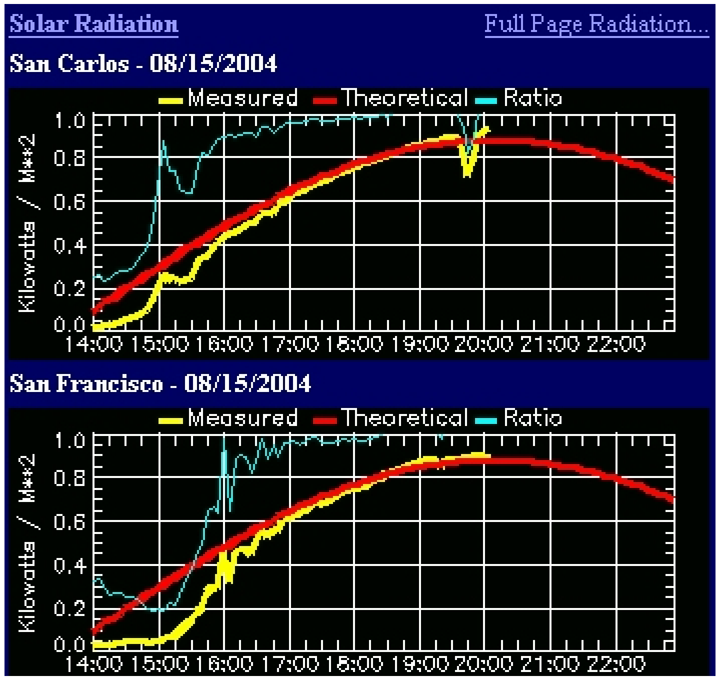


Figure 10. Pyranometer display.

3.6 TEMPERATURE & HUMIDITY SENSOR MAINTENANCE

3.6.1 Performance Monitoring and Corrective Maintenance

The temperature and dew point sensors provide high precision and high time resolution measurements at the two sensor sites, SQL and SFO. This resolution is an improvement from the standard NWS observation from these locations, which are available only once per hour. All of the surface observations are available from the main display via the panel in the upper left entitled *Surface Observations* (Figure 11). The NWS hourly observations for San Francisco and San Carlos are identified using the standard 3-character airport identifier codes SFO and SQL, respectively. Correspondingly, the field sensor identifiers employ the identifiers SFOL and SFQL. All of the regional surface observations can be viewed by selecting the *More Obs* link from the upper right corner of the *Surface Observations* box. This launches are larger table (Figure 12); SFOL and SQLL are the last two entries. SFO and SQL are found alphabetically from within the table. Temperature and dew point temperature are listed in degrees Fahrenheit under the columns labeled T and Td, respectively.

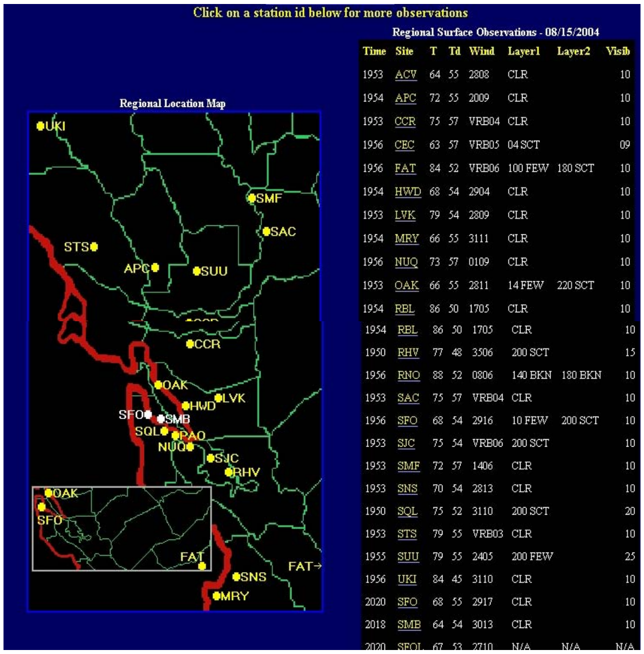
Performance of the field site temperature and humidity sensors can be monitored by comparative inspection of the field sensors (SFOL and SQLL) and the corresponding hourly NWS observations (SFO and SQL). Although they are not expected to match exactly, this will give a reasonable indication of significant inconsistencies. To view a broader sample of data, any site identifier can be selected to produce listings of sensor values for the entire day (Figure 13). Previous days can of data can be viewed by using the “View Prior Day(s) option from the links in the far right table of the main display page.

Sensor problems may also be identified and reported by operational users viewing the data displays. If corrective maintenance is necessary, follow maintenance procedures in the REBS sensor documentation. Use the REBS data logger analysis spread sheet for detailed data checkout if problems arise. If it is necessary to replace or calibrate a sensor, check the REBS documentation to see if calibration factors must be changed in the Data Logger software.

Surface Observations - 08/15/2004 More Obs

Time	Site	T	Td	Wind	Layer1	Layer2	Visib
2005	SFO	68	55	2815	CLR		10
2008	SMB	64	55	3013	CLR		09
2005	SQLL	74	52	2809	N/A	N/A	N/A

Figure 11. Surface observations on the main user display.



Regional Surface Observations - 08/15/2004

Time	Site	T	Td	Wind	Layer1	Layer2	Visib
1953	ACV	64	55	2808	CLR		10
1954	APC	72	55	2009	CLR		10
1953	CCR	75	57	VRB04	CLR		10
1956	CEC	63	57	VRB05	04 SCT		09
1956	FAT	84	52	VRB06	100 FEW	180 SCT	10
1954	HWD	68	54	2904	CLR		10
1953	LVK	79	54	2809	CLR		10
1954	MRY	66	55	3111	CLR		10
1956	NUQ	73	57	0109	CLR		10
1953	OAK	66	55	2811	14 FEW	220 SCT	10
1954	RBL	86	50	1705	CLR		10
1954	RBL	86	50	1705	CLR		10
1950	RHV	77	48	3506	200 SCT		15
1956	RNO	88	52	0806	140 BKN	180 BKN	10
1953	SAC	75	57	VRB04	CLR		10
1956	SFO	68	54	2916	10 FEW	200 SCT	10
1953	SJC	75	54	VRB06	200 SCT		10
1953	SMF	72	57	1406	CLR		10
1953	SNS	70	54	2813	CLR		10
1950	SQL	75	52	3110	200 SCT		20
1953	STS	79	55	VRB03	CLR		10
1955	SUU	79	55	2405	200 FEW		25
1956	UKI	84	45	3110	CLR		10
2020	SFO	68	55	2917	CLR		10
2018	SMB	64	54	3013	CLR		10
2020	SFOL	67	53	2710	N/A	N/A	N/A

Figure 12. Selecting "More obs" provides a complete table of observations from region.

San Francisco, CA (SFO)

Time	T	Td	Wind	Layer1	Layer2	Visib
0056	63	54	2412	11 FEW	17 OVC	10
0156	63	54	2511	19 BKN		10
0256	63	55	2412	17 BKN		10
0356	63	55	2410	17 BKN		10
0456	63	55	2309	17 BKN		10
0556	63	55	2510	15 BKN		10
0656	63	55	2513	15 BKN		10
0756	63	55	2512	13 BKN		10
0856	61	55	2612	11 OVC		10
0956	61	55	2411	15 OVC		10
1056	61	55	2510	11 SCT	15 OVC	10
1103	61	55	2610	11 SCT	15 OVC	10
1156	61	55	2609	11 SCT	15 OVC	10
1256	61	54	2510	11 SCT	17 BKN	10
1303	61	54	2407	11 FEW	17 SCT	10
1356	61	54	2406	11 FEW	15 SCT	10
1456	61	55	2510	12 SCT	17 BKN	10

Figure 13. Selecting an individual station code provides a table of data for the entire day.

3.6.2 Scheduled Maintenance

Follow maintenance procedures in the REBS sensor documentation. Use REBS data logger analysis spread sheet for checkout if problems arise.

Monthly: Visual inspection of sensor and shield (aspirator fan, deposits, grime)

Annual: Replace aspirator fan
Inspect shield tube for grime (replace tube and clean with soap and water at the shop)
Inspect T&H sensor for deposits (clean with distilled water and soft brush at site)

Biannual: Sensor calibration and refurbish at REBS

3.7 ANEMOMETER MAINTENANCE

3.7.1 Performance Monitoring and Corrective Maintenance

The anemometers provide high precision and high time resolution wind speed and wind direction measurements at the two sensor sites, SQL and SFO. This resolution is an improvement from the standard NWS observation from these locations, which are available only once per hour. All of the surface observations are available from the main display via the panel in the upper left entitled *Surface Observations* (Figure 10). The NWS hourly observations for San Francisco and San Carlos are identified using the standard 3-character airport identifier codes SFO and SQL, respectively. Correspondingly, the field sensor identifiers employ the identifiers SFOL and SFQL. All of the regional surface observations can be viewed by selecting the *More Obs* link from the upper right corner of the *Surface Observations* box. This launches are larger table (Figure 11); SFOL and SQLL are the last two entries. SFO and SQL are found alphabetically from within the table. The wind speed and direction is listed as a 4-digit set. The first two digits represent the direction from which the wind is blowing, in 10's of degrees. The second pair of digits represents the wind speed, in knots. Thus, a wind observation of 2815 would mean that the wind is blowing from 280 degrees (slightly north of due west) at 15 knots.

Performance of the field site anemometers can be monitored by comparative inspection of the field sensors (SFOL and SQLL) and the corresponding hourly NWS observations (SFO and SQL). Although they are not expected to match exactly, this will give a reasonable indication of significant inconsistencies. To view a broader sample of data, any site identifier can be selected to produce listings of sensor values for the entire day (Figure 12). Previous days can of data can be viewed by using the "View Prior Day(s) option from the links in the far right table of the main display page.

Sensor problems may also be identified and reported by operational users viewing the data displays. If corrective maintenance is necessary, follow maintenance procedures in the REBS sensor documentation. Use REBS data logger analysis spread sheet for detailed data checkout if problems arise. If it is necessary to replace or calibrate a sensor, check the REBS documentation to see if calibration factors must be changed in the Data Logger software.

3.7.2 Scheduled Maintenance

In case of data problems that cannot be corrected by realignment, the sensor unit should be replaced and the sensor returned to the manufacturer for repair.

Monthly: None

Annual: Check sensor alignment (visual inspection of the sensor and displayed wind direction).

Biannual: Bearing check by the manufacturer, R M Young, Inc.

3.8 DATA COMMUNICATION

The data communication between the remote sites and the base station is monitored by the system. A Data Status report is available on the main display page, and is accessed by the *System Status* button in the lower right corner. A typical status report is shown in Figure 14.

SFO Forecast Guidance System Data Status

Sensor Site	Status	Last Updated
San Francisco afos (afos_sfo)	GOOD	24.02 min ago (0.40 hrs ago)
San Francisco asos (asos_sfo)	GOOD	15.02 min ago (0.25 hrs ago)
San Mateo awos (awos_smo)	GOOD	7.02 min ago (0.12 hrs ago)
San Carlos sensor suite (flux_scf)	GOOD	15.02 min ago (0.25 hrs ago)
San Francisco Arpt sensor suite (radi_sfr)	GOOD	15.02 min ago (0.25 hrs ago)
San Carlos sodar (sodr_scs)	GOOD	15.02 min ago (0.25 hrs ago)
San Francisco sodar (sodr_sfs)	OK	64.02 min ago (1.07 hrs ago)
GOES 10 Visible data (vis_data)	GOOD	20.02 min ago (0.33 hrs ago)
GOES 10 Visible summary (vis_sum)	GOOD	20.02 min ago (0.33 hrs ago)

Figure 14. Data status display.

Successful data communication involves both the data source and the data receiver. In case of failure, several items must be checked: the sensor, the DCP, the site modems (optical and telephone), the base station communications, and the base station data collection system. For issues regarding the base station, refer to the Base Station documentation.

Logical clues that can help with the isolation of a communications problem:

1. If data from a single sensor are missing, check the sensor.
2. If data from all site sensors are missing, check the data logger and the communication between the data logger and the DCP. (Also check that the power supply is delivering a proper voltage to the data logger.)

3. If the data from a site are being provided over phone lines, but not over the internet, check the internet connection with the site.
4. If data from the site are unavailable by internet or phone, reboot the DCP (by teleboot). If the problem persists, check the DCP.
5. If data from both sites are unavailable, check the base station communications.

3.9 REPLACEMENT UNITS

Replacement units can be refurbished or obtained from the manufacturers. The system is designed with the following replacements intended:

NOAA/ETL (303-497-6303)

DCP computer
DCP computer hard drive (with software)
Sodar Filter
Sodar Amplifier

REBS, Inc. (206-624-7221)

Data Logger
Pyranometer
Temperature and humidity sensor

R.M. Young, Inc. (231-946-3980)

Anemometer system

GLOSSARY

ARTCC	Air Route Traffic Control Center
AWOS	Automated Weather Observing System
CWSU	Central Weather Service Unit
LAN	Local Area Network
NRL	Naval Research Laboratory
NWS	National Weather Service
NWSFO	National Weather Service Forecast Office
PPP	Point to Point Protocol
SFO	San Francisco International Airport
SODAR	Sonic Detection and Range
SQL	San Carlos Airport

SFO Marine Stratus Forecast System Documentation

Section III. Real-Time System

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS
(Continued)

LIST OF TABLES

1. INTRODUCTION

1.1 SCOPE

This document describes the Real-time System for the San Francisco Marine Stratus Forecast System. The purpose of the document is to provide a detailed description of the operational system software installed on the base station computer, which is located at the Air Route Traffic Control Center in Fremont, CA. This software is responsible for collecting conventional and field-site specific meteorological data, generating derived products, and managing the activation of forecast models that utilize the collected data and derived products.

1.2 RELATED DOCUMENTS

The SFO System documentation includes two main sections that are directly related to the real-time system information provided in this section. Section II describes field site meteorological sensors as well as the data collection platforms (computers) that collect and stage field sensor data for network-based retrieval by the base station computer. Section IV details the operational system software responsible for web-based display of meteorological data and forecast model output acquired and maintained by the base station computer.

2. OVERVIEW AND BACKGROUND INFORMATION

2.1 HISTORICAL PERSPECTIVE

The operational software installed on the base station computer is rooted in work that began as a separate effort within the FAA Integrated Terminal Weather System (ITWS) program. The effort, known as the Marine Stratus Initiative at San Francisco International Airport, began in 1995 with the objective to provide improved weather information and forecast guidance to the Oakland Center Weather Service Unit (CWSU). An early implementation of the system featured intermediate data routing computers that collected data from remote field site data collection computers and made the data available to a base station computer. The base station computer stored or "cached" hours of recent data and made the data available to remote user sites for further processing. The remote users sites in turn, acquired data from the base station using server-client software and ran analysis and display software locally to generate graphic output. Remote user sites included the CWSU at the Oakland Air Route Traffic Control Center (ARTCC) in Fremont, CA, the NWSFO/Monterey, and MIT/Lincoln Laboratory. A complete description of the early system is provided by Clark and Wilson (1996).

The operational system implementation is fairly recent, a design that dates back to 2000. This implementation greatly simplified the system, reducing the number of field sites to just the San Francisco and San Carlos field sites and utilizing the base station to communicate directly with the field site data collection computers, thereby eliminating the need for intermediate data routing computers and simplifying the overall system design. Additionally, the need to have analysis and display software installed at remote user sites was eliminated by converting the display subsystem to use a web-based client-server design. With a web-based approach, the base station would be solely responsible for generating a web page of display products and serving the page to any site with a commercial web browser.

2.2 SYSTEM OVERVIEW

Figure 1 is a schematic depicting the hardware configuration for the San Francisco Marine Stratus System. The system encompasses three locales: two field sites equipped with identical hardware and software and located at San Francisco International Airport and San Carlos Airport, and a two-node local area network (LAN) housed within the CWSU at the Oakland Center in Fremont, CA. The LAN is composed of a base station computer used for data collection and processing, and a display computer used by forecasters to view forecasts of marine stratus burn off within the approach zone as well as near realtime field-sensor observations and derived products. The base station acquires data from both the field sites and other sources using a commercial DSL connection to the Internet. A backup means to acquire field site data is available if needed, using a dialup Point to Point Protocol (PPP) connection over a phone line that is shared with DSL. The dialup connection is shown in Figure 1 as the dashed line connecting field site modems to the base station modem. Five-minute ASOS surface observations for San Francisco Airport are acquired through a dialup connection using the Kermit protocol. Five-minute Automated Weather Observing System (AWOS) surface observations from an FAA-managed site located at the San Mateo Bridge enter the base station via a serial RS232 line that originates within the confines of the CWSU weather desk. The Oakland upper-air sounding and hourly surface observations from stations in and around San Francisco Bay are provided by the National Weather Service Forecast Office (NWSFO) in Monterey and satellite data are retrieved from the Naval Research Laboratory (NRL), also located in Monterey. These data sources rely on the DSL connection for transfer of data to the base station.

Field site instrumentation, including temperature and humidity sensors, a pyranometer, and wind sensor, are tower mounted and connected to a data logger that prepares the sensor data for acquisition by the Data Collection Platform (DCP), a PC running the Windows XP operating system. The DCP also collects and processes data from an acoustic sounder that profiles the atmosphere just above the field site.

Both the field sites and the LAN at the CWSU are protected from Internet intrusion through inclusion of a firewall device. The firewall is configurable through a user interface to allow only permitted inbound traffic for remote login. The firewall does not limit outbound access. At the CWSU, the firewall also acts as a hub, connecting the base station and display computer in a LAN configuration.

Finally, all three locations are configured with teleboot bars. A teleboot bar allows a user or a computer to dial into the device and cycle power to any device attached to the bar. In Figure 1, a dedicated modem at the CWSU provides the base station with the ability to dial into either field site and clear a malfunctioning device without a person-in-the-loop. A teleboot bar is also located at the CWSU to provide maintenance personnel the ability to dial in to the CWSU and reboot the base station or communications equipment.

The remaining sections of this document focus on the software components that run on the base station computer, namely: data acquisition, data processing, and forecast model activation.

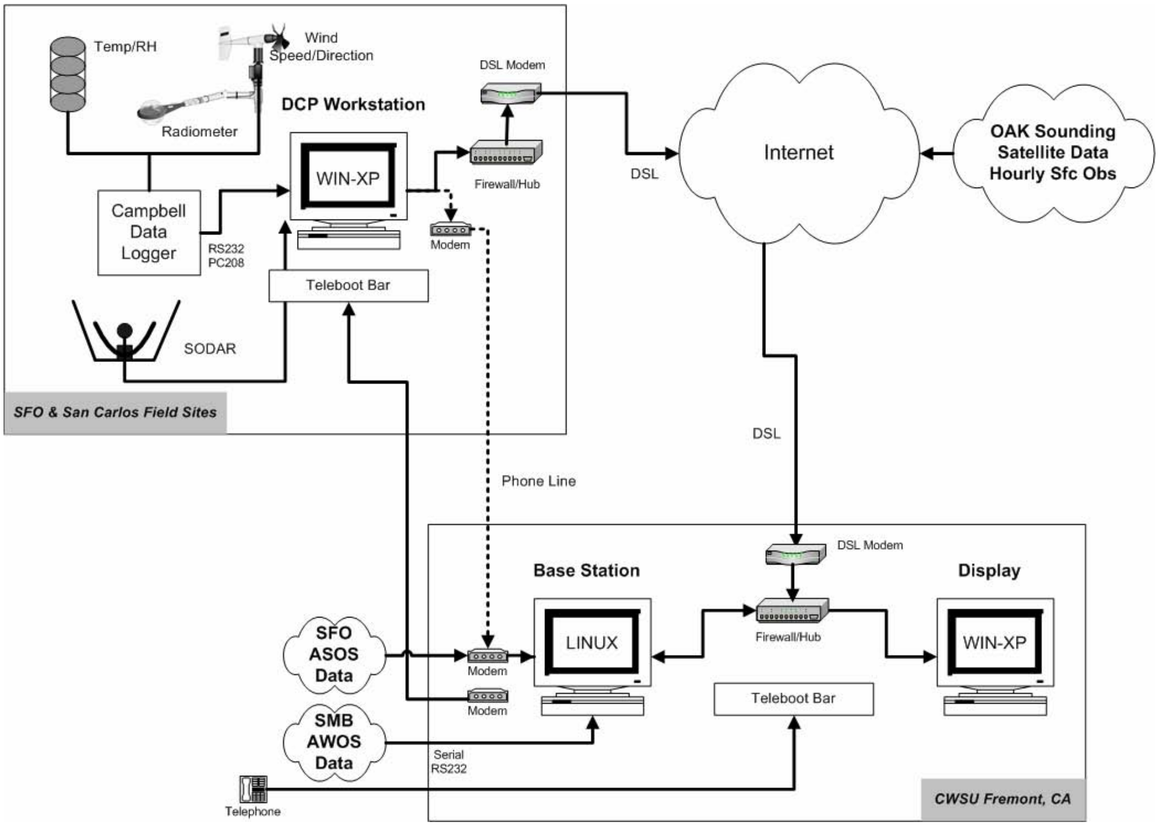


Figure 1. System diagram depicting the San Francisco Marine Stratus Forecast System.

3. DATA ACQUISITION

Data acquisition on the base station computer is the process of gathering project-specific data from field sites as well as conventional and satellite meteorological data from the NWSFO and NRL (and if necessary from other Internet sites) for the purpose of data display and as input to marine stratus forecast models. The acquisition process is composed of the following components: 1) routine acquisition, 2) acquisition failure detection and recovery, 3) data archiving, 4) data monitoring, and 5) cleanup operations. Under the UNIX operating system, these components have been implemented as *cron* jobs that run at regularly scheduled intervals. The collection of all cron jobs are defined in a master *crontab* file named */ll/cv/cv.cron*. This file is established as a working *crontab* file under the UNIX user account "cv".

During the stratus season, which runs from May through October, field-site and San Francisco Airport (SFO) 5-minute Automated Surface Observing System (ASOS) surface observations are retrieved every fifteen minutes during the hours that the Central Weather Service Unit (CWSU) is manned (5am to 9pm local time). During CWSU off-hours, calls are made once every hour. In the off-season, calls are made once an hour over the course of a day. The switch in call frequency between in-season and off-season calls is controlled by the use of different *crontab* files. During the stratus season, *cv.cron* mirrors the contents of either *cv.cron_summer_standard* or *cv.cron_summer_daylight*. In the off-season, *cv.cron* mirrors the contents of *cv.cron_winter*. Table 1 summarizes the differences between summer and winter data acquisition schedules. The file */ll/cv/ftp/pub/cv/data/SEASON* contains dates that define the beginning of the stratus and off-seasons. The default dates found in the file are May 15th for the first day of the stratus season and October 15th for the first day of the off-season.

The following subsections describe each data acquisition component (*cron* job) in greater detail.

3.1 DATA ACQUISITION VIA DSL/PHONE LINE DIALUP

The operational version of the San Francisco Marine Stratus Project employs DSL communications and the *ftp* protocol as the primary means of acquiring data from the field sites. Dialup communications using the Kermit protocol is to acquire ASOS surface observations. In the event that data acquisition fails using DSL communications during a regular data retrieval cycle, dialup communications are employed using the *ftp* protocol over a PPP connection.

The script */ll/cv/src/scatter-ftp/ftp_dial_test.tcl* is activated as a cron job to initiate two file transfers and one phone call to acquire sodar and pyronometer/surface observations from the two field sites and 5-minute ASOS surface observations from SFO. This script calls *gather.tcl* which in turn calls */ll/cv/src/kermit/calldcp.ker*. The script *calldcp.ker* invokes individual Kermit files for each phone call: *sfsodar_test.ker* for the SFO field site, *scsodar_test.ker* for the San Carlos (SQL) field site, and *asos.ker* for the SFO ASOS data. All calls are logged to the log file */ll/cv/call.log*. During the stratus season, *ftp_dial_test.tcl* is scheduled to run once every fifteen minutes starting at 0152 local time and ending at 2052 local time. Outside this time range and also during the off-season, data retrievals are made hourly at seven minutes past the hour.

Table 1

Stratus and Off-Season Data Acquisition Schedule

Data Source	Season	Acquisition Rate (min)	Hour (GMT)	Duration (hours)
Field Sites	Stratus	15	12	16
	Stratus	60	4	8
	Off-Season	60	0	24
SFO ASOS	Stratus	15	12	16
	Stratus	60	4	8
	Off-Season	60	0	24
SMB AWOS	Both	5	0	24
Hourly sfc obs from NWSFO-Monterey	Both	60	0	24
Hourly sfc obs from NWS web sites	Both	7.5	12	2
	Both	12	9, 11, 14-23	-
	Both	30	0-8, 10	-
OAK sounding from NWSFO-Monterey	Both	60	0	24
OAK sounding from NWS web sites	Both	7.5	0-5, 12-17	-
		30	6-11, 18-23	-
GOES satellite data	Both	15	0	24

Once files are successfully transferred to the base station computer, *gather.tcl* uncompresses any compressed files and initiates data format checks on the data. The program */ll/cv/src/scatter-ftp/kermit/sfdcp/sfsotest* checks the sodar file from the SFO field site for proper format while the program *datalggrsfdcp_test* performs a similar check on the pyronometer/surface observations file. In addition, the program *sodarmf_new* adjusts each one-minute sodar record so that the median value is constant. This insures a consistent gray-scale for the sodar display. Another program, *postsodrf*, removes any extraneous characters from the sodar file prior to any further processing. A similar set of data format checking programs inspect sodar and pyronometer/surface observations files from the SQL field site. The program */ll/cv/src/scatter-ftp/kermit/asos/asostest* is used to inspect the SFO ASOS file for proper file format. Results of each test produce either a file-specific success or failure file. A file acquired successfully is moved to the directory */ll/cv/data/Temp* for scattering whereas a failed file is discarded.

3.2 DATA ACQUISITION VIA THE INTERNET

Hourly surface observations, twice-daily Oakland upper-air soundings and visible satellite data from the GOES-W satellite are acquired through the Internet. The following subsection describes the acquisition of hourly surface observations and the Oakland soundings. Satellite data acquisition and preprocessing is discussed separately in Section 3.7.

Hourly surface observations and upper-air soundings from Oakland are transferred from the NWSFO in Monterey to the base station computer four times an hour starting at approximately 1 minute past the hour. The data transfer is initiated by the NWSFO using the *ftp* protocol. The data are placed in the directory */ll/ftp/pub/cv/data/nws* as 13 different files. Because of the mission critical value of these data there is a secondary independent internet-based source for these data should the primary means of acquiring the data through the NWSFO fail. Sounding data are retrieved from the web site *www.wrh.noaa.gov* using the Lynx *cron* job */ll/cv/src/scatter-ftp/getsnd.tcl*. Hourly surface observations are retrieved from the web site *weather.noaa.gov* using the Lynx *cron* job */ll/cv/src/scatter-ftp/getmetar.tcl*.

Once the hourly observations and sounding data have been acquired from either the primary source, secondary source, or a combination of both, the *cron* job */ll/cv/src/scatter-ftp/afosmaker_nws.tc*l re-formats the 13 separate files into the single file *afos_sfo.dat* and places the file in the directory */ll/cv/data/Temp* for scattering. This *cron* job runs most frequently between the hours of 12 and 13 GMT due to the need to get the Oakland sounding into the forecast cycle. For the remaining forecast hours, *afosmaker_nws.tcl* runs less frequently and only during the first half-hour of every hour.

3.3 LOCALLY-ACQUIRED DATA

An AWOS station located at the San Mateo Bridge provides 5-minute surface weather observations directly to the base station computer via a serial connection to the serial port *ttyS0*. The *cron* job *collect_awos.sh* runs once every minute to check for the presence of new text output from the binary program *readport*. If new data are found, the data are parsed (by the program *awosnew*) and relocated to the file */ll/cv/data/Temp/awos_smo.dat* for further processing.

3.4 DATA SCATTERING

Another *cron* job, */ll/cv/src/scatter-ftp/scatter_temp.tcl*, runs every minute to check for recently acquired data placed in the directory */ll/cv/data/Temp*. Files found in this directory are distributed to the main transfer directory */ll/cv/data/Xfer* for monitoring and to the directory */ll/cv/data/SFXfer* for additional processing by the cvRTCache, a program that prepares the data for the forecast models and then activates the models. A list and description of the files scattered *to /ll/cv/data/Xfer* can be found in Table 2. Documentation on the cvRTCache begins in Section 4.

Table 2

Sensor Data Files Produced by the Data Acquisition Process

Files Scattered to the Transfer Directory (*/ll/cv/data/Xfer*)	**Description**
afos_sfo.dat	Raw Surface Hourly and OAK RAOB Data
asos_sfo.dat	Raw SFO Surface ASOS Data
awos_smo.dat	Raw SMB Surface AWOS Data
radi_sfr.dat	Raw SFO Field Site Surface Data
flux_scf.dat	Raw SQL Field Site Surface Data
sodr_sfs.dat	Raw SFO Field Site SODAR Data
sodr_scs.dat	Raw SQL Field Site SODAR Data
vis_data.dat	GOES Imagery Data
vis_sum.dat	GOES Summary Data
Note: Files are also scattered to */ll/cv/data/SFXFer* for further processing by the cvRTCache. File names are as above with a unique data/time stamp appended to the end of the file: <filename>.dat.YYMMDD_HHMMSS	

3.5 DATA ARCHIVING

Files successfully acquired are also appended to archived day files in the directory */ll/cv/data/Archive* and are named according to year, Julian day and sensor type. See Table 3 for more information. Each day at 0 GMT, archived day files are compressed and relocated to */ll/cv/disk_2/data/Archive/zips/YYYY*, where *YYYY* is the appropriate year directory.

Table 3

File Naming Convention for Archived Raw Day Data Files

3-Letter File Extension	Corresponding Data File Name
afo	afos_sfo.dat
aso	asos_sfo.dat
log	gather.log
scf	flux_scf.dat
scs	sodr_scs.dat
sfr	radi_sfr.dat
sfs	sodr_sfs.dat
smw	awos_smo.dat
vsd	vis.data.dat
vsm	vis.sum.dat
File name format is: YYJJJ.Ext Where YY is the year JJ is the Julain day Ext is the 3-letter extension as above	

3.6 PRACTICAL CONSIDERATIONS

3.6.1 Programming and Resources

There are no known programming or computer resource issues with respect to data acquisition.

3.6.2 Exception Handling

The *cron* job */ll/cv/src/scatter-ftp/overseer.tcl* runs every 15 minutes during the stratus season starting at 0147 local time (just before the forecast cycle begins) through 2047 local time when CWSU operations end for the night. Outside this time range and during the off-season, the script runs every 30 minutes starting at 2 minutes past the hour. Files distributed to the main transfer directory */ll/cv/data/Xfer* are examined for age and minimum size by *overseer.tcl*. If acquired sodar and pyronometer files remain stale after 2 (see reference to DEEPLEVEL in Table 4) successive checks by *overseer.tcl*, the appropriate field site computer is telebooted. Data acquisition failures also prompt *overseer.tcl* to send out e-mail to listed recipients. Specifically, e-mail messages are issued after 2 (see reference to BEEPLEVEL in Table 4) successive checks for proper file size or age fail. Teleboot calls are made on a separate phone line to prevent interference with calls to acquire data. On the base station computer, the modem used for telebooting is connected to port *ttyS4* while the modem used for backup data ingest is connected to port *ttyS5*. A list of all e-mail messages can be found in Appendix A. Note that e-mail messages are categorized by urgency and contain instructions for what actions, if any, should be taken to resolve the problem.

In the event of a communications failure with any of the three regularly scheduled data retrievals, *ftp_dial_test.tcl* is re-run with the word "overseer" as an argument in an attempt to retry a failed retrieval or retrievals. Retried retrievals are made on the same lines as regularly scheduled retrievals but at a time sufficiently different to avoid interference. A log of retry attempts is written to */ll/cv/src/scatter-ftp/overseer.tmp*. In addition, any success files located in the directories */ll/cv/src/scatter-ftp/kermit/sfdcp*, *scdcp*, or *asos* are removed both at the beginning of each call event and also if files fail an age test. In every calling cycle a check is made in *ftp_dial_test.tcl* for the existence of success files. If success files are missing, the script */ll/cv/src/scatter-ftp/regather_test.tcl* is invoked to attempt a recall. Thus in every calling cycle, two attempts are made to retrieve data that are missing due to a communications-related failure.

A check for the existence of recent Oakland soundings is also made using *cron* jobs. The script */ll/cv/src/scatter-ftp/chksnd0.tcl* checks for a recent 0 GMT sounding at 0137 GMT each day while *chksnd.tcl* checks for a recent 12 GMT sounding at 1337 GMT. Regardless of whether a sounding is up-to-date, e-mail is sent to listed recipients that include the contents of the sounding file.

For AWOS processing, the *cron* job */ll/cv/awos/testread.sh* runs every 10 minutes to check the age of AWOS data. If the data are stale, *testread.sh* terminates the *readport* program. In turn, the cron job */ll/cv/src/awos/restart.sh* runs every minute to check if *readport* is running and restarts it if it has been stopped.

With the advent of DSL communications as the primary medium for data retrieval from the field sites, additional communications hardware has been introduced into the system. This hardware includes telephone company-supplied DSL modems as well as firewall/hub devices located at both field sites and at the CWSU. Like the field site data collection computers, these devices are also subject to malfunctions, requiring a teleboot of the device when necessary. The base station data acquisition software has been modified to monitor these additional devices and to initiate teleboot actions when necessary. Table 5 lists the teleboot phone numbers and teleboot bar plug configurations for each site. Table 5 provides the information needed for a human to perform a manual teleboot operation, should the situation arise. For information on the command language utilized to communicate with a teleboot bar, consult the NBB Series Network Boot Bars User's Guide from Western Telematic Inc., part number 13333. The manual teleboot process assumes an operator has access to a computer and modem and a communications program (e.g., Hyperterminal) to interact with the text interface of the teleboot bar.

Table 4

Input Parameters to the Raw Data Monitoring Script overseer.tcl

Input Parameter	Default Value	Description
SFBUFFER	1200 sec	Max allowable age of sodr_sfs.dat for 15 min call cycle
SCBUFFER	1200 sec	Max allowable age of sodr_scs.dat for 15 min call cycle
SFBUFFERN	4000 sec	Max allowable age of sodr_sfs.dat for 1 hr call cycle
SCBUFFERN	4000 sec	Max allowable age of sodr_scs.dat for 1 hr call cycle
SFPYROBUFFER	1200 sec	Max allowable age of radi_sfr.dat for 15 min call cycle
SCPYROBUFFER	1200 sec	Max allowable age of flux_scf.dat for 15 min call cycle
SFPYROBUFFERN	4000 sec	Max allowable age of radi_sfr.dat for 1 hr call cycle
SCPYROBUFFERN	4000 sec	Max allowable age of flux_scf.dat for 1 hr call cycle
SATBUFFER	2000 sec	Max allowable age of vis_sum.dat for daylight hours
ASOSBUFFER	1200 sec	Max allowable age of asos_sfo.dat for 15 min call cycle
ASOSBUFFERN	4000 sec	Max allowable age of asos_sfo.dat for 1 hr call cycle
BUFFER	4000 sec	Default max allowable age for data file
AFOSBUFFER	4000 sec	Max allowable age of afos_sfo.dat
DRIVBUFFER	1000 sec	Max age for sodr_*.dat driver check in 15 min call cycle
DRIVBUFFERN	2000 sec	Max age for sodr_*.dat driver check in 1 hr call cycle
SOBBUFFER	1200 sec	Max allowable age of sob.tmp showing sob link alive
AWOSBUFFER	1200 sec	Max allowable age of awos_smo.dat
WINDBUFFER	4000 sec	Max allowable age of winds.dat (WOCSS wind model)
FCSTBUFFER	2900 sec	Max allowable age of latest_* forecast files in fcst cycle
WINDSIZE	1000 byte	Min allowable size of winds.dat file (WOCSS wind model)
ASOSSIZE	100 byte	Min allowable size of asos_sfo.dat
SODSIZE	200 byte	Min allowable size of sodr_*.dat
FLXSIZE	100 byte	Min allowable size of rad_sfr.dat and flux_sct.dat
AREASIZE	4010000 byte	Min allowable size of raw satellite file (estimate)
AWOSSIZE	50 byte	Min allowable size of awos_smo.dat
AWOSBIGE	104 byte	Max allowable size of awos_smo.dat
BEEPLEVEL	2	Default minimum number of failures before action taken
WEEPLEVEL	6	Min number of WOCSS failures before action taken
CEEP LEVEL	3	Min # of low (100’) bogus ceilings from SMB ‘til action taken
DEEPLEVEL	2	Min # of driver check failures for sodr_*.dat ‘til action taken
ZEEPLEVEL	999	Min # of zero line check failures for sodr_*.dat ‘til action taken
BOOTLEVEL	10	Min # of sodar-only or datalogger-only failures before teleboot

TABLE 5

Site Phone Numbers and Teleboot Bar Plug Assignments

Location	Phone Number	Plug 1	Plug 2	Plug 3	Plug 4	Plug 5
CWSU	510 793-3623	Base Station	Cisco PIX Firewall	Data Modem	Teleboot Modem	SBC DSL Modem
SFO Field Site	650 692-2497	Computer Modem	Computer	SBC DSL Modem	Cisco PIX Firewall	Teleboot Bar Modem
SQL Field Site	650 631-0623	Computer Modem	Computer	SBC DSL Modem	Cisco PIX Firewall	Teleboot Bar Modem

Teleboot bar text interface command to cycle power to a device:

/BOOT *n*, Y

where *n* is the plug number corresponding to the device to be power cycled
and Y indicates that confirmation of the action to power cycle is to be suppressed.
BOOT can be abbreviated to BO.

3.6.3 Quality Assurance and Diagnostics

Unlike the field site data, hourly surface observations are not examined to see if the data arrive in a standard format. Instead, the scripts used to assemble the target file (*afos_sfo.dat*) from files acquired through the Internet impose a standard format. In addition, the assembly process selects only the most recent site observation data, ignoring any older data that may exist in the raw files.

3.6.4 Operator Interaction and Default Settings

The master crontab file */ll/cv/cv.cron* schedules all data collection that originates from the base station computer. It also schedules data monitoring and archiving activities. Default settings for the time-of-day activation of the various programs that comprise the data acquisition process are specified *in /ll/cv/cv/cron*. Adjustments to the schedules for these programs would only be necessary if the definition of data acquisition cycles were to change.

Table 4 lists the default threshold settings used by *overseer.tcl* to monitor file size and age and successive failure count thresholds used to govern when to take corrective action when problems occur. These settings can be found at the beginning of the *overseer.tcl* script. These settings are tuned for the 15-minute stratus season and 1-hour off-season data acquisition cycles. Adjustments would only be necessary if the duration of either cycle were to change.

3.6.5 Periodic/Potential Maintenance Tasks

The *sh* scripts */ll/cv/src/scatter-ftp/afos/delsndgud.sh* and *delsndgd0.sh* run twice each day to remove old sounding files to prepare for new sounding file checks made by *chksnd.tcl* and *chksnd0.tcl*. The script */ll/cv/src/scatter-ftp/afos/delog.sh* runs once each week to remove log files and other miscellaneous files associated with data acquisition logging.

Two minutes before the start of each 15-minute calling cycle, the *cron* job */ll/cv/src/scatter-ftp/Cleanup/cleanup* is activated. This script terminates any dialup scripts still resident from the last calling cycle to insure a clean start during the next calling cycle.

The cron job */ll/cv/src/awos/delrdprt.sh* runs once each month to delete log files associated with AWOS data acquisition.

Periodically, NWS personnel change the password required at dial in to retrieve the 5-minute SFO ASOS data. When the password changes, the new password is entered in the file */ll/cv/src/scatter-ftp/kermit/asos.ker*, at the line that starts with the word “lineout”. Older out-dated passwords will appear above this line, commented out with a leading semicolon.

3.7 SATELLITE DATA ACQUISITION AND PREPROCESSING

3.7.1 High Level Description

3.7.2 Theoretical Description

This subsystem of the SFO C&V real-time system is responsible for the ingestion and processing of visible satellite imagery. This imagery has two principal uses: to provide situational awareness via display of current and recent images, and as a basis for making statistical forecasts of stratus burn-off. The initial raw visible image is improved for these uses by correcting registration errors, normalizing to an equivalent local noon, applying a low-pass filter, and providing numerical summaries of pre-chosen sub-regions (or sectors) of the image.

3.7.3 Data Inputs

The initial input file is a Network Common Form (NetCDF)-formatted file containing data from the GOES-W visible channel. The file name is generated using the format *YYYYMMDD.HHMM.hi.gSS* with coverage of the west coast of the United States as shown in Figure 2. The input file is retrieved by the base station once every 15 minutes from the Naval Research Laboratory in Monterey, CA.

3.7.4 Data Outputs

The entire satellite process produces two output files. One contains summary statistics of the current image for various pre-defined sectors and is named *vis_sum.dat*. The other file contains albedo measures for each of the 80 x 80 points in the grid which overlays our spatial domain. This output file is named *vis_data.dat*. The contents of these two files are described in more detail in the appropriate sections below. At the end of each image's processing these two files are placed in the directory /ll/cv/data/SFXfer for further processing (see Section 4).

3.7.5 Structural Flow – Overall Steps to the Process

The input file is then subjected to the following procedures. The functional flow diagrams for the entire process are shown in Figures 3 and 4. The individual procedures are discussed more fully in the various sections below:

- Unpacking and extracting a subset of the initial image.
- Normalizing the image for sun angle and instrument degradation.
- Correcting the image for registration errors.
- Interpolating the image to the base grid.
- Filtering the image.
- Compute sector summary statistics.

These core procedures, along with others involving file transfers, are controlled by a series of scripts. These in turn are all controlled by master scripts named "StartGoes" and "StopGoes", also discussed below.

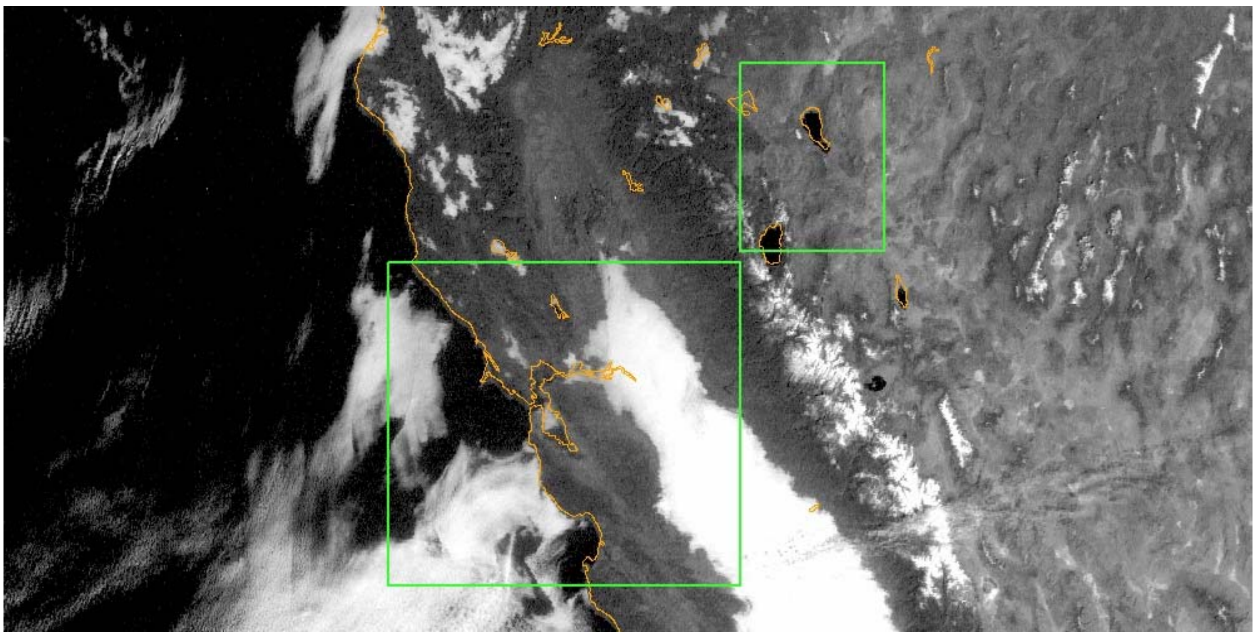


Figure 2. Sample image showing coverage of the initial GOES file. The two rectangles show the parts of this image that are extracted for subsequent processing; the larger is the main area centered on the San Francisco Bay, and the smaller is centered on Pyramid Lake, NV and is used for image registration.

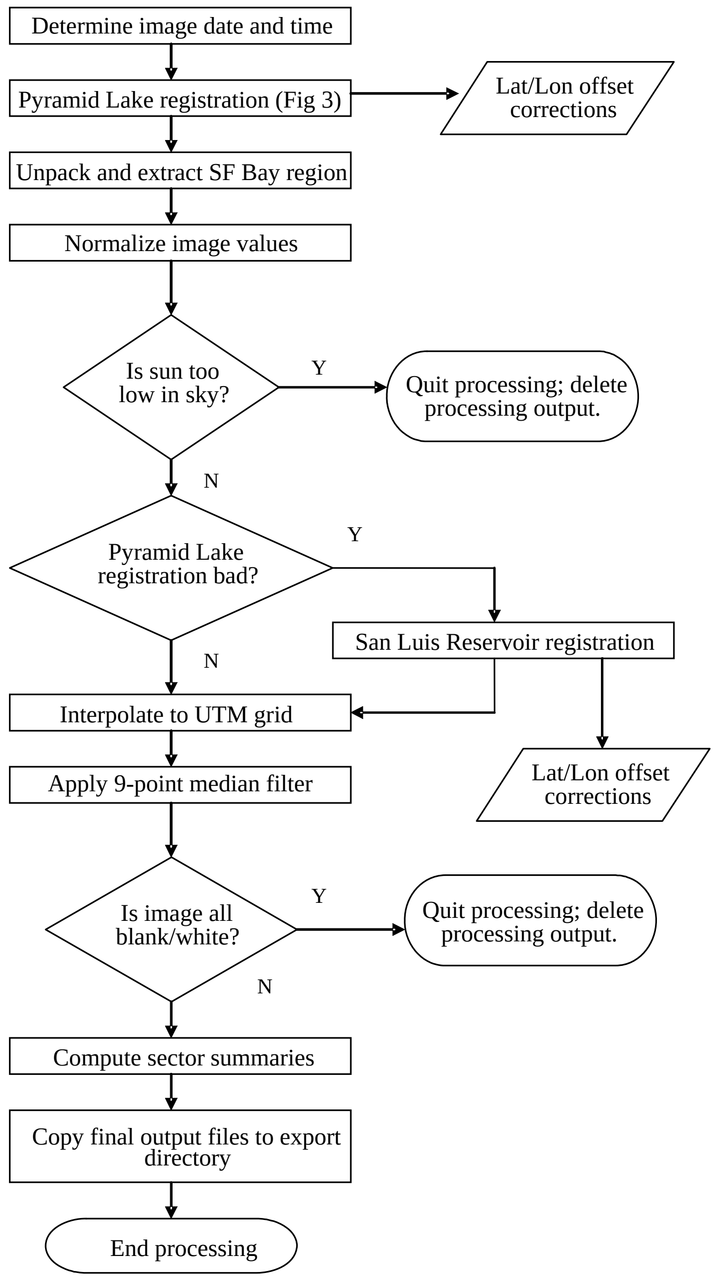


Figure 3. Functional flow diagram for satellite processing.

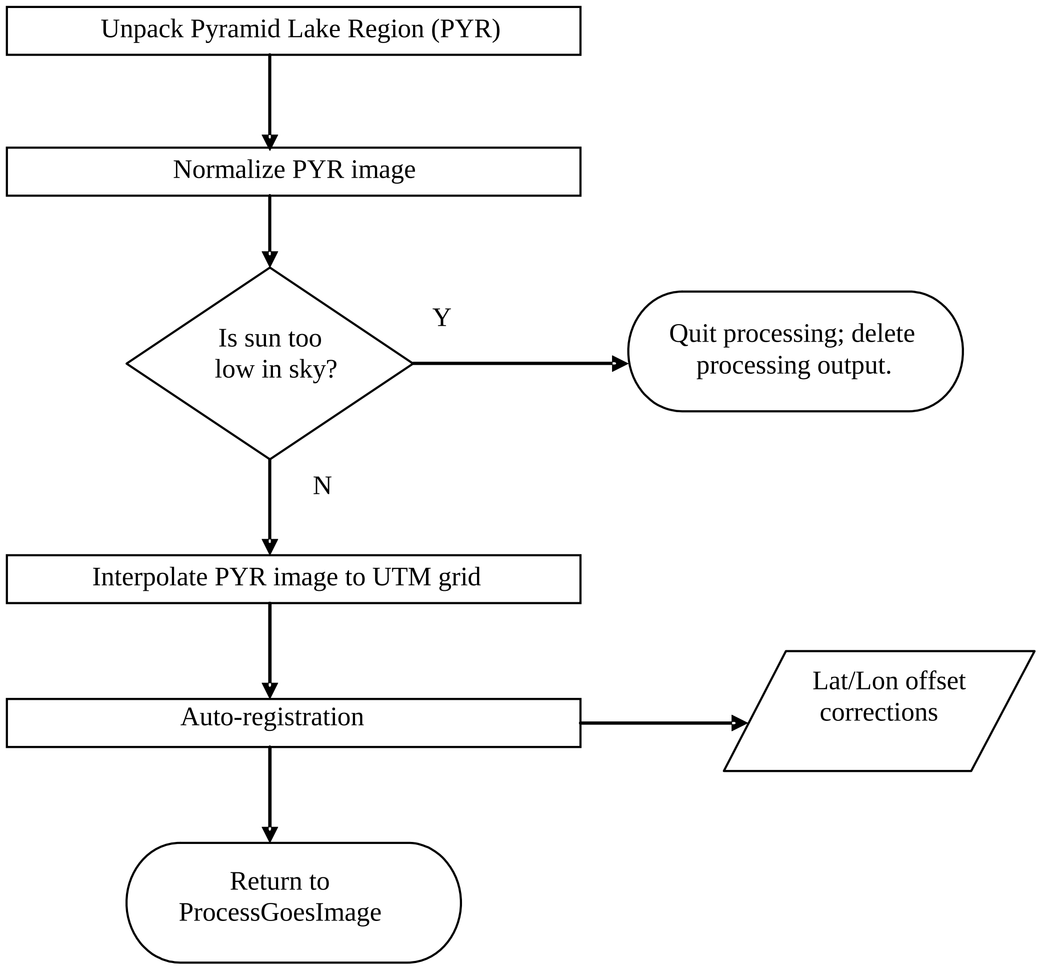


Figure 4. Functional flow diagram for Pyramid Lake registration.

3.7.6 Practical Considerations

There are two ways in which the user may have to alter the system for future conditions. The first is inevitable; the second possible. Inevitably, the GOES-W satellite in use will change from the current GOES-10 to a replacement satellite, with further changes in the far future. As is detailed in the normalization section below, imagery is corrected for progressive "graying" due to instrument degradation. The parameters that govern this degradation are empirically derived by NOAA and are unique to each satellite (in part because they are based on the launch date of the satellite). When a new satellite becomes operational, new parameters will need to be added to the appropriate parameters file. This is detailed below.

It is also possible that the raw satellite image received from the NRL could change in size. The current system expects these images to be of a certain size, and will exit otherwise (after providing the new dimensions). If this should happen, constants in the unpack, normalization, and interpolation programs would need to be updated, and the code recompiled. Details describing the changes to the constants in these programs can be found in the Sections 3.8.6.2, 3.8.7.2, and 3.8.10.4.

The system is robust with respect to images composed of partial or wholly missing (blank) values. Processing is accomplished for any valid portions of the image, with missing flags filling the rest. Should the entire image be non-valid, processing is terminated and no products are generated.

3.8 SUBSYSTEM COMPONENTS

3.8.1 StartGoes

This is the script that starts the satellite processing. It calls two scripts (rcpRaw and AcquireRaw, see below) that run continually and monitor the state of the input file (image) and begin subsequent processing with the arrival of each new image. StartGoes calls these two scripts in such a way as to enable logging of their diagnostic output.

3.8.2 RcpRaw

This script acquires each raw image file (once it has completely arrived). It copies the file to a standard import directory. As the input files are all visible images, it refuses to so copy any overnight images (which would be completely black).

3.8.3 AcquireRaw

This script completes the acquisition of the raw image file. It monitors the import directory for new files. Once one completely arrives, its filename is stamped with the current date and time, and the file is moved to the proper dated directory for raw files. The post processing script (PostProcess, see below) is then invoked.

3.8.4 PostProcess

This script accomplishes three main tasks. First, it determines whether or not the file is of sufficient size to be a "good" image. Secondly, it invokes ProcessGoesImage (see below), which accomplishes the main image processing tasks. Third, when processing of the image is completed, DistributeGoesData (see below) is invoked, which moves the final satellite products from an export directory to the transfer directory for further processing by the cvRTCache.

3.8.5 ProcessGoesImage

3.8.5.1 Theoretical Description

This is the master script which controls the details of the satellite image processing. It in turns calls various executables (from C code) and scripts that accomplish the basic tasks described in Section 3.7.5. A functional flow diagram for this script is shown in Figure 3.

3.8.5.1.1 Data Inputs

The only data input for this script is the original raw input image.

3.8.5.1.2 Data Outputs

The final outputs of the scripts, if executed successfully, are the two products of the entire satellite processing system mentioned above: *vis_data.dat* and *vis_sum.dat*.

Most of the executables called by this script produce their own unique outputs that are in turn used as inputs for subsequent executables. These are all detailed in each components description given below.

3.8.5.2 Practical Considerations

This script is written in perl.

Three command-line arguments are expected: the full path of the directory containing a copy of the raw input file, and in which the processed data is to reside; the name of the raw input file; and the type of operating system. This last argument controls where the script looks for the various executables described below.

The script contains a block of user-adjustable parameters at its head. All such parameters are designated in name by all capitals. They are given in Table 6. It is not anticipated that any of these parameters will need to be adjusted in the future.

There are several points in the script at which it can be determined that the current image is not, for various reasons, of sufficient quality to proceed with processing. If this is determined, the script deletes all of the output produced thus far for this image (but not the initial raw image itself) and exits.

3.8.6 Unpack

3.8.6.1 Theoretical Description

The original raw satellite image file is in NetCDF format, a self-describing machine-independent file format designed to represent array-oriented scientific data. The file contains discrete blocks of data known as variables that can be accessed independently of and without having to read other sections of the file. In addition to the full resolution visible channel data, each file contains coincident solar zenith angle, latitude, and longitude data stored as separate variables. This program extracts the variables using routines from the NetCDF C-interface library and formats the data into separated flat files. A detailed description of NetCDF is provided by Rew and Davis (1990).

Table 6

Parameters Used in "ProcessGoesImage" Script

Name	Current Value	Description
BIN_DIR	/ll/cv/projects/goes/bin	Location of executables, with subdirectories for each possible operating system (linux and solaris)
SUN_ANGLE_THRESH	82.5	Upper limit of solar zenith angle (deg) for allowable processing
LAT_PER_PIXEL	0.0180	Latitude (deg) size of each pixel in base UTM grid
LON_PER_PIXEL	0.0227	Longitude (deg) size of each pixel in base UTM grid
FILT	9mf	Type of filter used on image (currently 9-pt median filter)
HEADER	See source code	Header for sector summary file
EXP_DIR	/ll/cv/data/GOES/Export	Location for final satellite products

3.8.6.1.1 Data Inputs

For command-line inputs, this program expects the path of the directory containing the raw input image and the name of the raw input image. As file inputs, this program only expects the raw image file.

3.8.6.1.2 Data Outputs

This program generates a number of separate files for subsequent use. The routines determine the date and time of the image from the file name and use these to construct the base file name. This has the form of *VIS_YYMMDD_HHMM* (for example, VIS_030605_1500 would refer to the 1500 GMT image from 5 June 2003). The different files generated from this program (and also from all subsequent programs) all have this root file name, but with differing extensions. The types of files produced by this program are summarized in Table 7.

Table 7

Summary of Files Produced by Unpack

Extension	Type of File	Format	Description
.img	image values	Binary (short int)	Image values for each point in the unpacked domain. Values can range from 0-255.
.lat	latitudes	ASCII	Latitude (deg N) of each image value.
.lon	longitudes	ASCII	Longitude (deg W) of each image value.
.sun	sun angles	ASCII	Solar zenith angle (i.e., sun angle measured from the zenith, in degrees) for each image value.

3.8.6.2 Practical Considerations

This program expects the dimensions of the input image to be of a certain size. These are currently hardcoded (via a *#define* in *unpack.c*) as 1000 by 500. The dimensions of each incoming raw image are read by the program from file attributes. If these dimensions do not match those expected the current dimensions are printed to the log file and the program will exit. Should the dimensions of the incoming image change in the future, the values of *XDIM* and *YDIM* in *unpack.c* will have to be changed to the new values, and the program will need to be recompiled.

3.8.7 Normal

3.8.7.1 Theoretical Description

This program normalizes image brightness values, attempting to remove the non-meteorological changes in image brightness due to sun angle variations and satellite sensor degradation. This is accomplished as two separate functions within this code. The relevant parameters for the current satellite are kept in a separate parameter file.

The normalization can be expressed as:

Normalized_value = Original_value * sun_angle_factor / graying_factor

The "sun angle factor" is based on the cosine of the solar zenith angle (θ) along with a seasonally varying "constant" (A, basically a sinusoidal function of the day of year). It is expressed as:

$$sun\_angle\_factor = \sqrt{(1+A)/(\cos\theta + A)}$$

This relation was adapted from Chernal et al (1985).

The "graying factor" is a progressively larger number based on the number of days since the launch of the current satellite (d). It is taken from Rao (2001) (available online from NOAA at www.oso.noaa.gov/goes/goes-calibration), and has the form:

$$graying\_factor = 1/[1.049(1+0.0001022d)]$$

3.8.7.1.1 Data Inputs

For command-line input, normal expects the path of the directory containing the data, the body of the filename (i.e., the filename without the extension), and the x-y dimensions of the input data.

Two input files containing data from the satellite image under consideration are required. Both follow the naming convention for data files described above. One is the file of unpacked image values (*.img*) and the other the values of solar zenith angle (*.sun*). Both of these files were produced as output from the unpack process.

This code also requires a static file that contains four parameters governing the general sun-angle corrections, as well as specific parameters for the current GOES-W satellite in use. This file is */ll/cv/projects/goes/params/normalize.param*, and is set in the normalize code via a #define of *NORM_PARAM_FILE*.

3.8.7.1.2 Data Outputs

A single file of normalized outputs is produced, having the extension *.nml*. It is of the same form (binary) and dimensions as the input *.img* file.

3.8.7.2 Practical Considerations

It is possible that in the course of normalization that some pixels may exceed the maximum brightness (i.e., 255/255). This is more likely to happen when the image contains deep convective cloud. Such pixels are capped to a maximum of 255.

It is important to realize that this formula for the graying factor will need to be changed in the future. It may be updated based on additional analysis for GOES-10, and will certainly need to be changed when GOES-10 is replaced. The NOAA web site mentioned above can be monitored for future changes.

If this change is simply to the values of the constants used for GOES-10, these can be made via the *normalize.param* file. Simply change the relevant values in the line for GOES-10.

If, however, the change is due to a switch to a new satellite, then a line for this satellite will need to be added to *normalize.param*. This line will consist of the GOES identifier (XX, for GOES-XX), the start date for data (i.e., when the operational switch was made from the previous satellite), the launch date of the satellite, and the values of the two constants used in the graying factor formula. The GOES website (or the scientists mentioned thereon) would be a good source for this information. In addition, if a switch to a new satellite is made, the fifth line in *normalize.param* will need to be changed to the total number of satellites having degradation parameters in this file. It is currently set to 2 (for GOES-9 and –10), and should be changed to 3 when new satellite information is added.

If operations switch to the next GOES-W and no new degradation parameters are yet available for this satellite, then the old parameters may be copied and used with the new start and launch date. This should be adequate initially while the satellite is still “young”, based on the similarity of the calibration parameters for GOES-8, -9, and -10. But the difference that changes to these parameters make in the calibration are magnified the greater the number of days from launch, so it is important to implement new parameters as soon as they become available from NOAA.

If a future change to the form of the graying factor formula is required, this can be made in the function *calibration_formula.c*. Recompilation of the normalization code would have to follow. However, it is not anticipated that such a change will be necessary.

This program expects the dimensions of the input image to be of a certain size. These are currently hardcoded (via a *#define* in *normal.c)* as 1000 by 500. Should the dimensions of the incoming image change in the future, the values of *XDIM* and *YDIM* in *normal.c* will have to be changed to the new values, and the program will need to be recompiled.

3.8.8 register_pyr

3.8.8.1 Theoretical Description

Satellite images are routinely misaligned with respect to their reported navigation information. This is easily seen on coastal visible satellite images during cloud-free conditions when geographical outlines are overlaid according to the reported latitudes and longitudes; often the plotted outlines are noticeably shifted away from what should be their true locations. This is generally referred to as an error in image *registration*. The system attempts to correct these errors by aligning several easily discernable bodies of water that are usually cloud-free during the stratus season, and computing the x and y offset distances. These are then used as corrections in subsequent processing steps.

Two bodies of water are used as targets, one primary and one as backup in the event that the primary target is unavailable (usually due to cloud contamination). The primary target is Pyramid Lake, in western Nevada; the secondary is the San Luis Reservoir, southeast of San Jose on the east slope of the Diablo Range. This processing is outlined in the flowchart in Figure 4. A sample image of Pyramid Lake showing an initial mismatch in registration is shown in Figure 5.

This particular routine, *register_pyr*, computes a registration correction based on the primary target, Pyramid Lake. Parameters are declared in this routine (via #define) that are unique to this target. A separate but similar routine computes a registration correction based on the secondary target, and is described in a separate section.

The technique starts with a template that represents the shape of the target from the point of view of the base 2-km grid (shown as white dots in Figure 5). On this template, ones represent grid points within the target and zeroes points that lie outside. This template is overlaid on the image file according to the initial navigation and a dot product is calculated by multiplying the template values (zeroes and ones) by the image values. Then the template is shifted one grid point and another dot product is calculated. This is done until the template has been shifted to a maximum of five grid points in each of the four cardinal directions. As the image values in the target (representing water bodies) are darker than their surrounding over-land image values, the shifting of the template that corresponds to the minimum dot product yields the registration error of the image. These values are returned to the main processing script.

3.8.8.2 Data Inputs

As command-line arguments, this routine expects the directory path containing the image file from which the registration corrections are to be deduced, and the name of that data file.

This routine looks for a template file for Pyramid Lake. This file is located in the parameters directory (*/ll/cv/projects/goes/params*) and is named *register_pyr.template*.

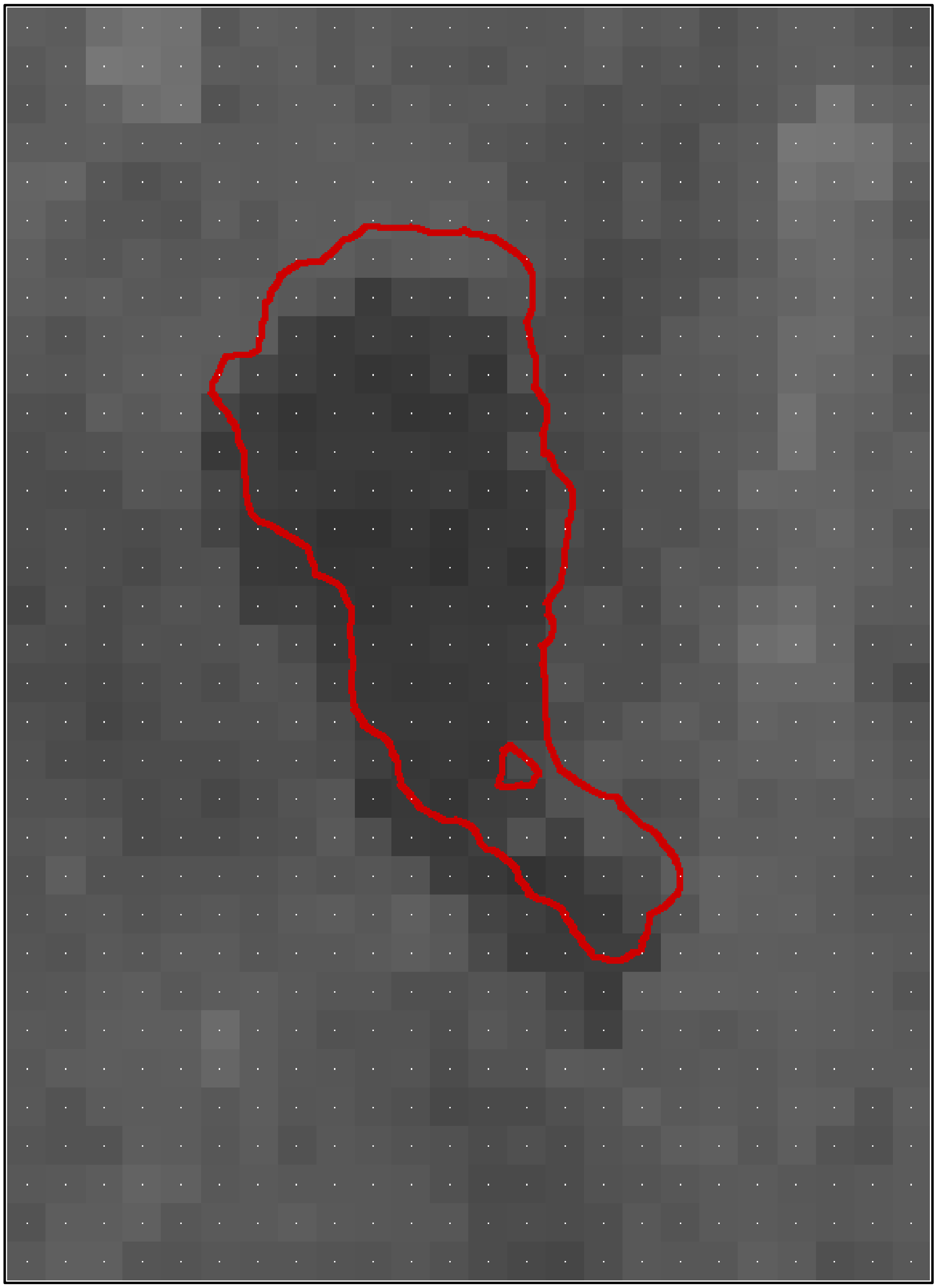


Figure 5. Pyramid Lake geography overlaid on a sample satellite image prior to registration correction.

3.8.8.3 Data Outputs

This routine produces no file output. It does print two values (on a single line) to standard output. These values are the latitude and longitude registration corrections. This output is captured in the master script and used in the subsequent interpolation routine.

3.8.8.4 Practical Considerations

The success of the registration routine assumes a clear view of the target to discern land/water brightness contrasts. If the target is obscured by cloud this routine can give misleading results. Therefore each pixel in the target image is compared to a threshold value that has been determined to effectively discriminate between cloud and no-cloud. If more than 1/3 of the target image is obscured by cloud then no registration is attempted. A “cloud contamination” message is printed to standard error, and “missing” offsets (-99’s) are produced. These missing offsets will cause the master script to attempt registration via the secondary target.

In addition, some common-sense bounds are imposed on the registration offsets that are produced, in the event that the routine runs away with itself and produces very large offsets. Currently these offsets are capped at ±5 grid points in each direction.

3.8.9 register_slr

This program is identical in function to *register_pyr* described above. The only difference is the target, which is the San Luis Reservoir. Thus a different templates file is read (*/ll/cv/projects/goes/params/register_slr.template*), and different parameters are declared in the source code specific to this target.

3.8.10 interp_main

3.8.10.1 Theoretical Description

This program interpolates the unpacked and normalized image values to a standard rectangular grid. This grid is 160 km x 160 km, with a spacing of 2 km between grid points (resulting in a grid of 80 points by 80 points). It is centered on the San Francisco Bay and is based on the Universal Transverse Mercator (UTM) projection. This grid is shown in Figure 6. The interpolation is carried out by a “nearest neighbor” approach (i.e., the value of the original pixel that is nearest the desired grid point is used).

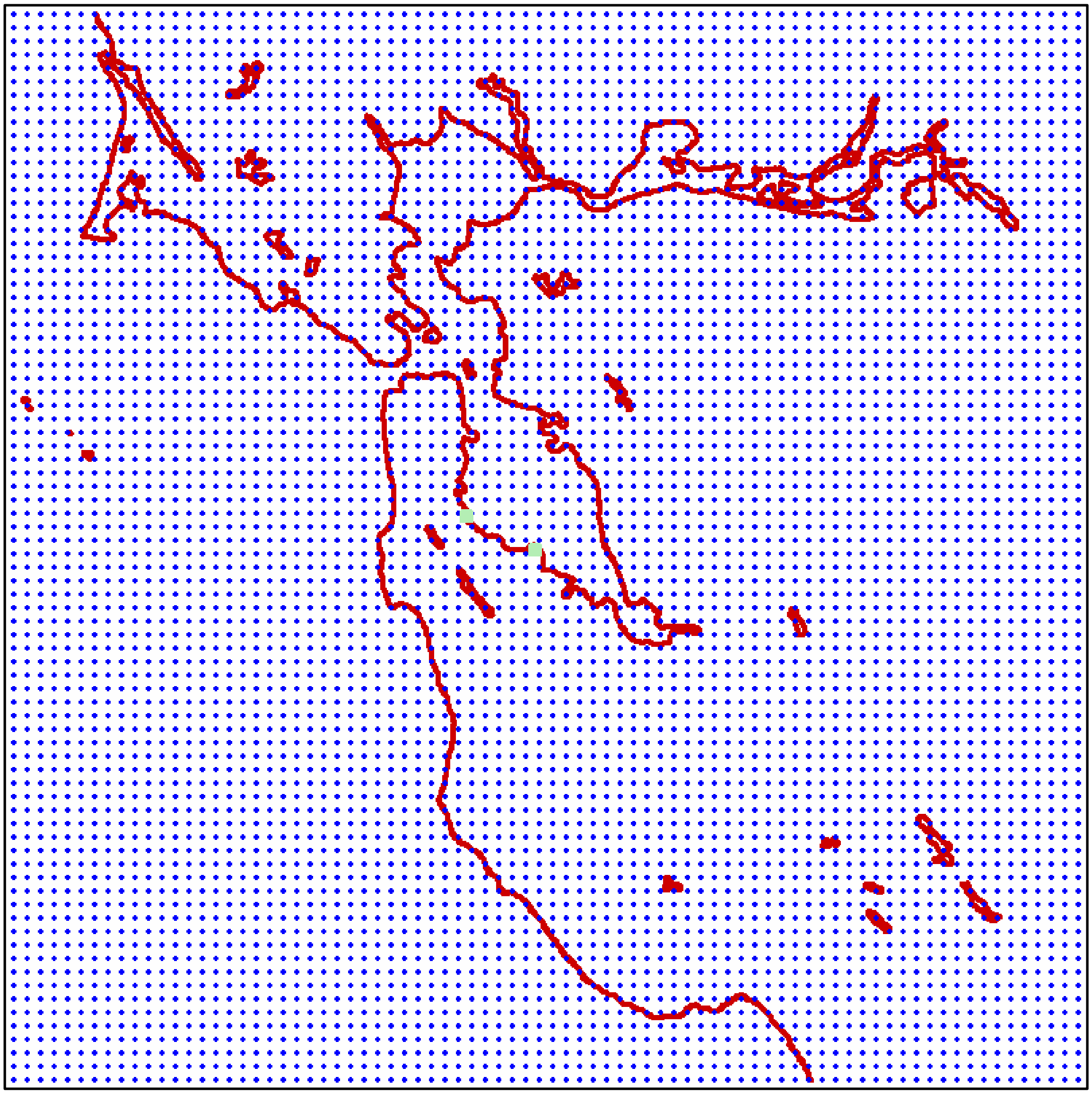


Figure 6. The main 2-km UTM grid to which satellite image values are interpolated.

This program interpolates the entire San Francisco Bay image that is used as a base for all subsequent routines. Similar interpolation programs apply to the two registration targets. These are identical as to function but differ as to several controlling parameters. They are described in subsequent sections.

3.8.10.2 Data Inputs

The numerous command-line arguments to this program are: the complete path to the input data, the body of the input file name, the input file's extension, the x-dimension of the input file, the y-dimension of the input file, the output file's extension, and the latitude and longitude registration adjustments.

The input image file is specified by the command line arguments that specify its location, body name, and extension. This program also expects the latitudes and longitudes of each of the input image's pixels. These are expected to be in the same directory as the input image, and to have the same filename body, but to have the extensions "*.lat*" and "*.lon*".

This program also expects the latitudes and longitudes of the UTM grid that the input image is to be interpolated to. These are found in the parameters directory (*/ll/cv/projects/goes/params*) and are named *utmlat_main.2km* and *utmlon_main.2km*.

3.8.10.3 Data Outputs

The output file of interpolated image values shares the input image file's location and filename body. The extension for the output file is specified in the command line; "*.int*" is the designation used in the master script. The output file is written in unsigned 1-byte binary form.

3.8.10.4 Practical Considerations

The program expects the dimensions of the input image to be of a certain size. These are currently hardcoded (via a *#define* in *interp_main.h)* as 1000 by 500. Should the dimensions of the incoming image change in the future, the values of ORIG_*XDIM* and ORIG_*YDIM* in *interp_main.c* will have to be changed to the new values, and the program will need to be recompiled.

3.8.11 interp_pyr

This program interpolates the initial image that covers the Pyramid Lake registration target to a standard 2-km UTM grid. It only differs from *interp_main* in the values of several of the parameters specified in the *interp_pyr.h* header file, and in the names of the UTM latitude and longitude files (*utmlat_pyr.2km* and *utmlon_pyr.2km*). Its output is designated with an *.pyr* extension. Like *interp_main*, the values of ORIG_XDIM and ORIG_YDIM contained in the include file (*interp_pyr.h*) for this program will need to be modified if the dimensions of the input image change.

3.8.12 interp_slr

This program interpolates the initial image that covers the San Luis Reservoir registration target to a standard 2-km UTM grid. It only differs from *interp_main* in the values of several of the parameters specified in the *interp_slr.h* header file, and in the names of the UTM latitude and longitude files (*utmlat_slr.2km* and *utmlon_slr.2km*). Its output is designated with an *.slr* extension. Like *interp_main*, the values of ORIG_XDIM and ORIG_YDIM contained in the include file (*interp_slr.h*) for this program will need to be modified if the dimensions of the input image change.

3.8.13 Filter

3.8.13.1 Theoretical Description

The interpolated satellite image, like the original image on which it is based, usually contains a great deal of high frequency, low amplitude variability that is of very little interest. It can be distracting visually and may add unwanted noise to the sector summaries (described below). This variability is removed by applying a simple low-pass filter to the image values. The filter employed is a 9-point median filter. Thus every image value is replaced by the median value of the 9 values contained in a 3x3 pixel box centered on the value in question.

3.8.3.2 Data Inputs

On the command line, this program expects the directory path to the input image file, the filename body of the input image, and the filtering method to be employed. This method is specified in the master script as "*9mf*", to indicate a 9-point median filter.

The input file that is expected by the program, and specified in the command line, is the interpolated image file and has the extension "*.int*".

3.8.3.3 Data Outputs

The filtered output image file is placed in the directory specified in the command line, and has an extension corresponding to the filtering method used. Thus in this case the output file has the extension "*.9mf*".

3.8.14 goes_bin2asc

3.8.14.1 Theoretical Description

The output image files throughout the satellite image processing are written in an unsigned 1-byte binary format. One of the final products, however, is a file of image values in ASCII format. This utility program translates an input binary file to ASCII. It also checks the image for validity; i.e. it makes sure that the image is not either entirely white or black.

3.8.14.2 Data Inputs

On the command line, this program expects the directory path of the input (and output) files, the filename body of the image to be translated, the extension of the input (binary) file, and the extension of the output (ASCII) file.

The only input file is that specified by the path, filename body, and input extension given on the command line.

3.8.14.3 Data Outputs

This program also returns the number of valid values in the image. A valid value is one that is greater than the absolute minimum (0) and less than the absolute maximum (255). This value is used by the master script to detect images that are either entirely blank/white or dark. If such an image is encountered, processing for the image is aborted and the processed files are deleted.

The output file consists of the 80x80 input image values written in ASCII format. The output filename consists of the path, filename body, and output extension given on the command line. The output values are converted from the initial 0-255 range to a 0-100 range.

3.8.14.4 Practical Considerations

This program checks that no image values lie outside the minimum and maximum bounds (values of 0 and 255, respectively). Any such values are capped to their applicable bounds. Any such values should have been dealt with in prior processing steps, so the presence of this check is redundant but retained for added safety.

3.8.15 Sector

3.8.15.1 Theoretical Description

The primary use of satellite data in the system is as potential predictors for a satellite-based statistical forecast model. The sector program converts the gridded image values into a more usable form for this purpose. Primarily this involves condensing the image from the original 80x80 values into a more manageable number. This is done by dividing the domain into approximately 50 sectors, and then producing summary statistics for each sector. These sectors are shown in Figure 7. The sector values are then used as statistical predictors.

The sectors are summarized as to mean brightness, the standard deviation of the brightness, and the fraction of the sector that is covered by cloud. The latter is determined by comparing the brightness values in the sector to an empirically determined cloud threshold. Currently this is set to 115 (out of a maximum 255). The program also generates the number of non-missing pixels within the sector for this image, and a value that allows the standard deviation of a grouping of sectors to be calculated.

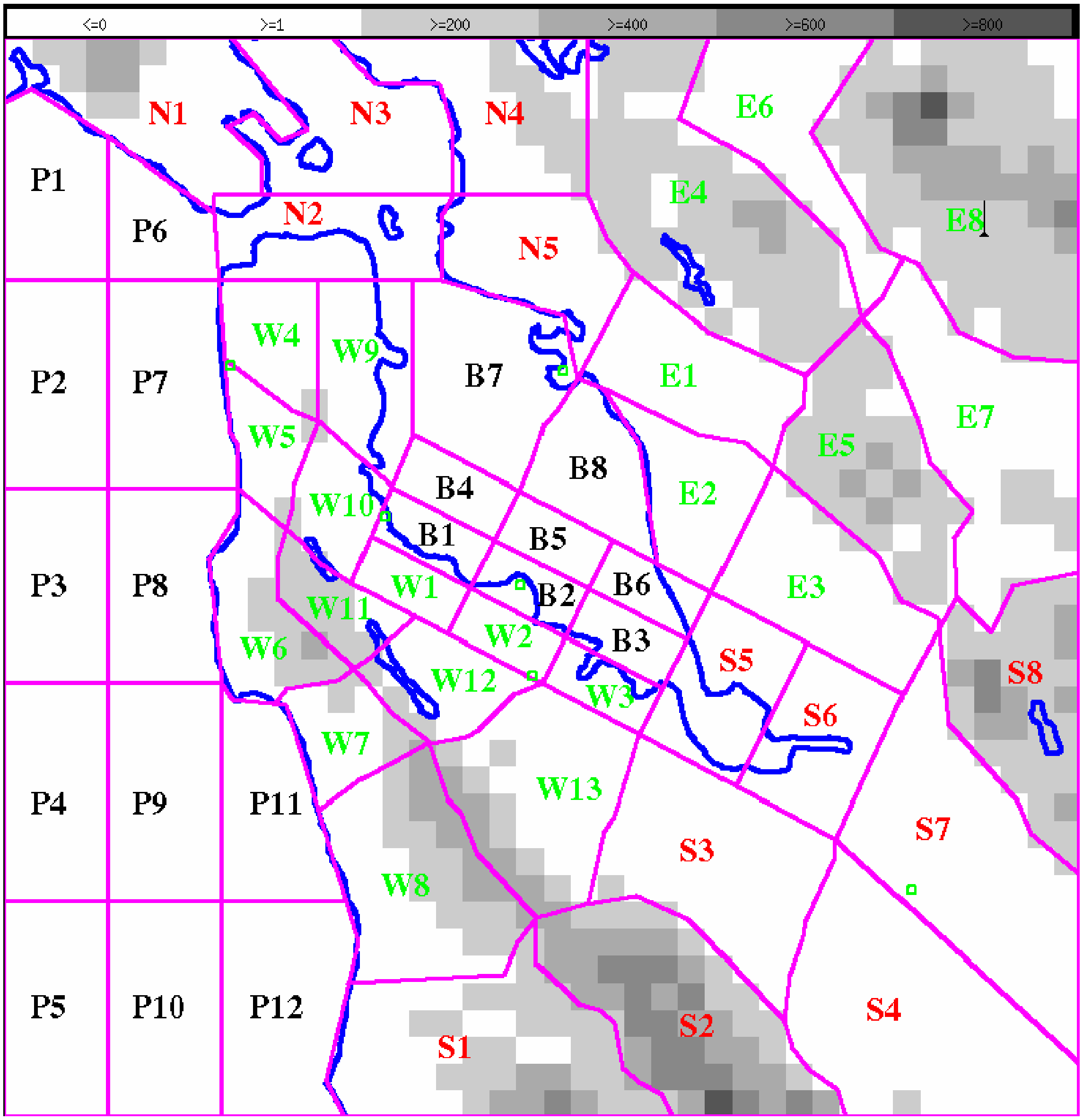


Figure 7. Satellite sector locations and labels. Gray shading indicates elevation, with the key (in meters) at the top of the figure.

3.8.15.2 Data Inputs

On the command line, this program expects the path to the input data file, that filename's body, and the extension of the input filename. The program then expects the input data in the location and under the name so specified.

This program also expects a static file specifying the configuration of the sectors. This file is fixed as "*/ll/cv/projects/goes/params/sector.dat*". This file first contains the number of sectors specified, and then for each sector its name and number of points, and the coordinates of those points within the base 80x80 image.

3.8.15.3 Data Outputs

This program sends it results to standard output, rather than a file. This output is captured into an appropriately named file within the master processing script. It is named by the standard filename body with an extension of "*.sum*", and a copy is placed in the export directory.

3.8.15.4 Practical Considerations

If the entire image is composed of missing values, the overall satellite process will never reach this step and sector summaries will not be produced. However, patches or swaths of missing values sometimes occur within images. The sector program checks each sector for its number of non-missing pixels. If the fraction of such pixels is greater than a threshold of 0.80, then a valid summary is produced for the sector based on the non-missing pixels. If the fraction of non-missing pixels is below this threshold, then the program produces missing (-0.99, in this case) flags for all the sector summary values.

3.8.16 DistributeGoesData

This script moves the final two data products from the export directory to the *cvRTCache/Xfer* directory. In the process it renames the files from their date- and time- stamped versions to the generic names of *vis_data.dat* and *vis_sum.dat* while adding a date and time header line to their contents.

4. DATA PROCESSING

4.1 TECHNICAL DESCRIPTION

Once data are acquired and placed in the transfer directory, the data are required to undergo further processing before forecast model execution can take place. Data are decoded, subjected to simple range checking, reformatted, and written to disk files in a non-competitive date ordered directory hierarchy. A single program composed of subunits that are responsible for processing individual data types manages this processing. The program, named the "cvRTCache", is implemented as a memory resident module that waits for and then processes the stream of data entering the system. The characteristics and functions of the cvRTCache are described in the following subsections.

4.1.1 Data Inputs

The input sensor data processed by the cvRTCache are one or more files placed in the transfer directory by the data acquisition process. As previously mentioned, files placed in the transfer directory are ready to be read without risk of contention. In addition, files placed in the transfer directory will queue if the rate of data acquisition exceeds the rate at which files are being processed. This is accomplished by the data acquisition process that adds a unique date/time stamp to the name of the file. Table 2 lists the names of the files that appear in the transfer directory and the format of the name once the date/time stamp is appended to the name.

The remaining input files accessed by the cvRTCache are located in a common directory. They are read once when the cvRTCache is started. Thus, changes made to these files do not take effect until the cvRTCache is restarted. These files are user-editable and control the following features:

1. How the year and month are determined for surface hourly data.
2. Which surface hourly stations' observations are processed.
3. Which raw data entering the system to process and which to ignore.
4. Parameter control file for the cvRTCache.

The first file is needed to direct the cvRTCache where to search for year and month information within the raw surface hourly/OAK RAOB file since the METAR format contains only information on the day, hour, and minute of the observation. Appendix B contains the location and name of this file along with a description and format of its contents.

The second file directs the cvRTCache as to which hourly METAR stations from those found in the raw surface hourly/OAK RAOB file to process and save. The file is composed of a simple list of stations, expressed as 3-letter station identifiers. Appendix C contains the location and name of this file along with a description and format of its contents. Section 4.2.5 provides additional details on when and how to modify this file and a discussion of the numerous applications dependent on this file.

The third file directs the cvRTCache as to which sensor data placed in the transfer directory to process and save. The file is composed of a simple list of sensor identifiers, similar to the raw files names listed in Table 2, without the file name extension. Appendix D contains the location and name of this file along with a description and format of its contents.

The fourth file directs the cvRTCache where to look for certain resources such as the location of executables and scripts that the cvRTCache activates, where to look for the raw input sensor data, where to write the formatted output sensor data, and where to place log files. This file also instructs the cvRTCache how to name the formatted output sensor data. Appendix E contains the location and name of this file along with a description and format of its contents.

4.1.2 Data Outputs

Sensor data processed by the cvRTCache are written to disk files in a non-competitive date ordered directory hierarchy. Figure 8 illustrates the directory hierarchy structure and output file naming convention. Files in this directory structure are referred to as "day summary" files because each file contains a summary of a day of data for a particular source and type of data. All processing subsequent to processing performed by the cvRTCache (e.g., forecast model execution) references day summary data files.

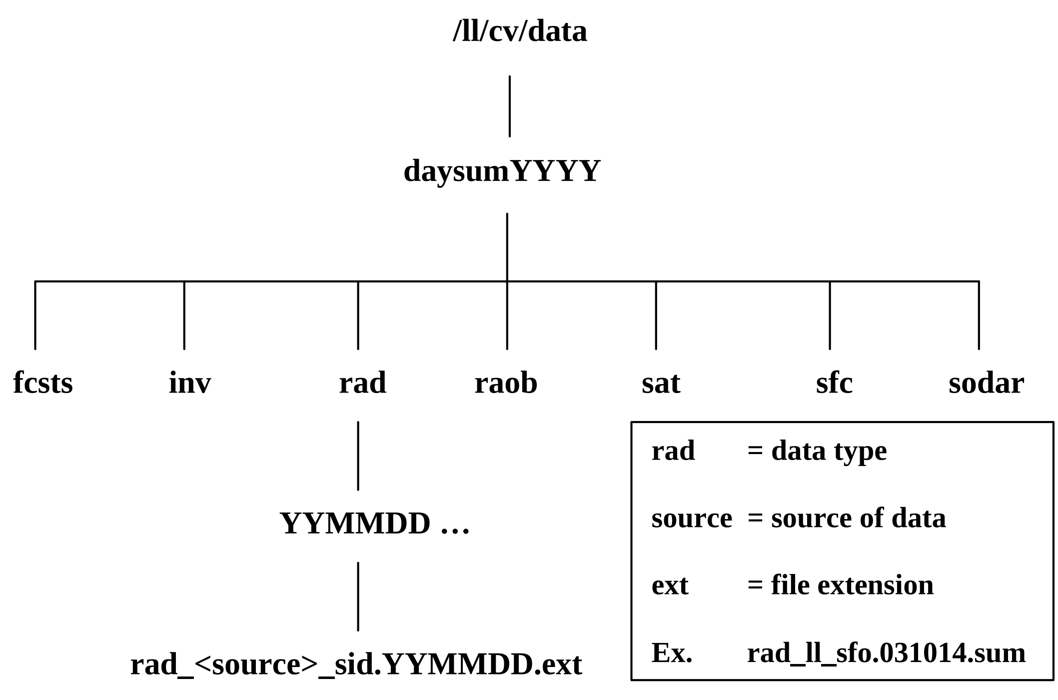


Figure 8. Schematic of date ordered directory structure for storage of processed sensor data. Also shown is file naming convention.

The cvRTCache maintains a log file each day that records its activities as well as the activities of programs and scripts that it activates. Log files are located in the date ordered directory /ll/cv/logs. If the cvRTCache is stopped and restarted, whether purposely or not, a new log file will be created upon restart. Thus, it is possible to have multiple log files for the same day. Log files are given unique file names by including the date and time that the cvRTCache was started in the file name.

Three additional files are written and maintained by the cvRTCache in the */tmp* directory. One file is used to coordinate the midnight GMT exit of the program to roll over the log file for the next day. Another file is used to log the startup and shutdown activities of the cvRTCache and a third file maintains a running log of mail messages sent to the operator giving a history of starts and restarts of the program.

4.1.3 Structural Flow

4.1.3.1 Overview

A high-level functional flow diagram of the cvRTCache is shown in Figure 9. Upon activation, the cvRTCache reads the run-time control parameters from the parameters file. It then determines the machine type of the computer it is running on so that the appropriate directory paths can be identified that contain machine-specific executable programs that the cvRTCache will be activating. Next the program establishes a log directory path and log file and writes a copy of the run-time parameters (just read) to the log file. This completes initialization tasks performed by the cvRTCache. The program then enters into a loop that begins with a timed sleep of one-minute duration. Upon expiration of the sleep event, the program examines the data transfer directory to see if any new data have arrived since it was last checked. If a new file or files are found, the cvRTCache identifies the type of data contained in the file and calls the appropriate module for further processing. Once processed, the cvRTCache deletes the file from the transfer directory. This procedure is repeated in serial for each file until all files have been processed.

After any new data files have been processed, the cvRTCache then calls a master post-processing script to allow further processing of the data without the risk of data contention. That is, tasks that require reading the data that has just been processed (written) by the cvRTCache need to be able to do so without risk of further potentially overlapped processing of newly arriving data. This is accomplished by having the cvRTCache activate the master post-processing script and wait for its completion before restarting a new cycle of data processing. The master post-processing script in turn, activates various programs needed to complete the cycle of processing for the real-time system (e.g., activate the forecast models). Note that although the web display system accesses processed sensor data, it is not part of the post-processing activities launched by the master post-processing script. Instead, the web display system has been developed as an autonomous software system that utilizes file locks to coordinate file access with the cvRTCache.

Once all post-processing tasks have completed, control is returned to the cvRTCache. If it is a new day, the cvRTCache terminates normally to allow a restart and generation of a new log file to reflect the crossover to a new day. Otherwise, the cvRTCache begins a new cycle.

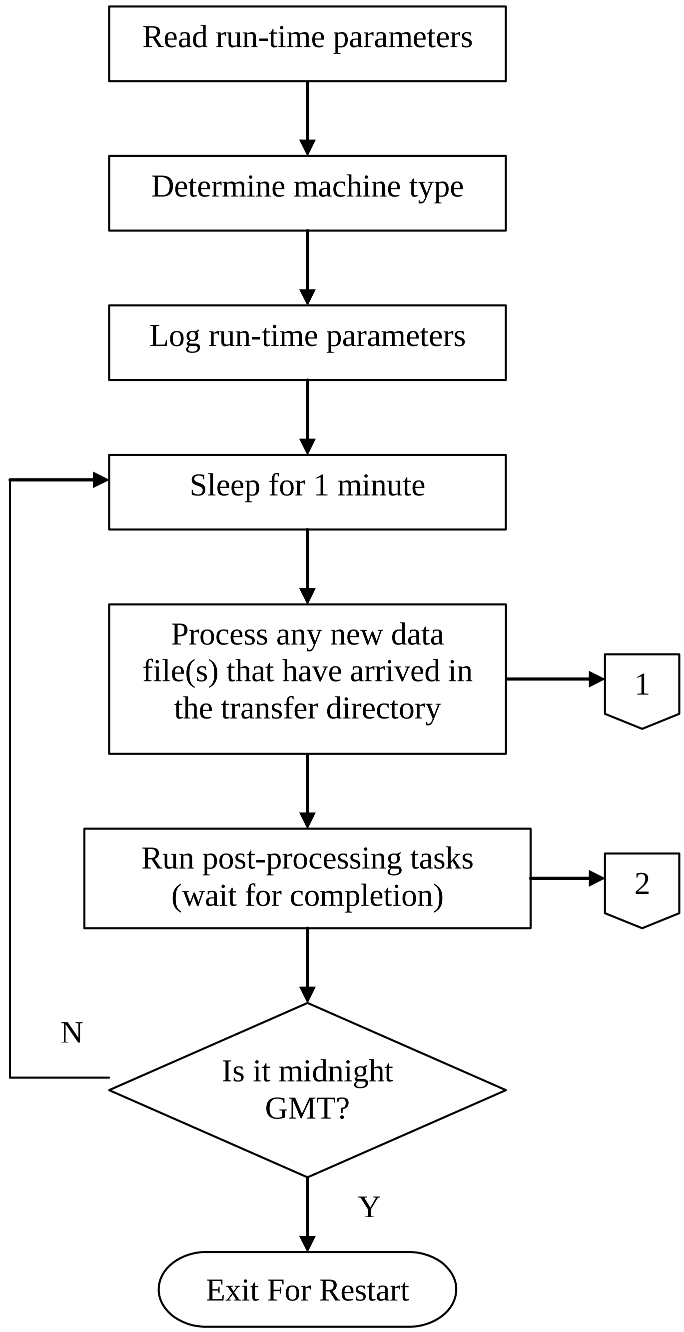


Figure 9. Functional flow diagram for the cvRTCache Program.

4.1.3.2 Data Processing Details

Further data processing details are shown in the functional flow diagram of Figure 10. This section of code is executed as the result of one or more sensor files being detected by the cvRTCache within the transfer directory. First, the sensor identifiers contained in the sensor list file (described in Section 4.1.1) are read and accessed in succession to call the sensor-specific function associated with each sensor data type. The following logic described is identical for all functions, as indicated in Figure 10. Within a function, the first step is to check if the file exists. If it does, a copy of the file is placed in a working directory and the working directory is software locked so that the cvRTCache will have exclusive write access to the day summary file. The software lock coordinates write access that the cvRTCache needs to update the file with read access that the web display software needs to post data to the web page. Next, a sensor-specific data-processing routine is called to process the data and write it to the day summary file. The function then terminates after having deleted the sensor input file from the transfer directory and releasing the software lock on the working directory. This sequence is then repeated for the remaining sensor identifiers until the list is exhausted and the end-of-file marker is detected. Control is then returned to the main body of the cvRTCache program.

4.1.3.3 Master Post-Processing Script

Details of the master post-processing script are presented in the functional flow diagram of Figure 11. The cvRTCache launches this script, waiting for the script to complete before continuing on to the next processing cycle. The master script starts off by examining the contents of the transfer directory for any files ready to be processed. If found the oldest set of files are renamed to their generic equivalents by removing the unique date-time stamp. This action allows the cvRTCache to process the files on the next cycle. The next step is to run a script that determines the sky conditions in and around the approach zone at San Francisco Airport. The script is called once every cycle beginning one hour prior to the first pre-dawn forecast interval. The script examines ceiling reports from four surface observing stations and declares that stratus is present if at least one of the stations is reporting a ceiling below 3000 feet. By early morning, as visible satellite data becomes available, the script will relinquish the task of determining sky conditions to another similar script that analyzes conditions based solely on satellite data as long as the satellite data is timely. Next, the master script checks to see if the current time is within the forecast time interval. If so and if stratus has been detected in the area of the approach zone, the component forecast models are executed followed by invocation of the consensus model.

Although the only two criteria, forecast time interval and presence of stratus, are needed to activate the models within the master post-processing script, further logic within the forecast model software modules determines whether or not actual forecasts are generated.

4.1.4 Nominal and Degraded Performance

Under nominal conditions of data acquisition, the cvRTCache will easily process sensor data files placed in the transfer directory in a first in first out fashion. When forecast models are running, the additional processing time may introduce a minor delay in the processing of newly arrived data (because the cvRTCache waits for post-processing tasks to complete). In a case such as this, the arriving files simply queue in the transfer directory (as mentioned above by naming the files with a unique date-time stamp) until the cvRTCache can begin a new processing cycle.

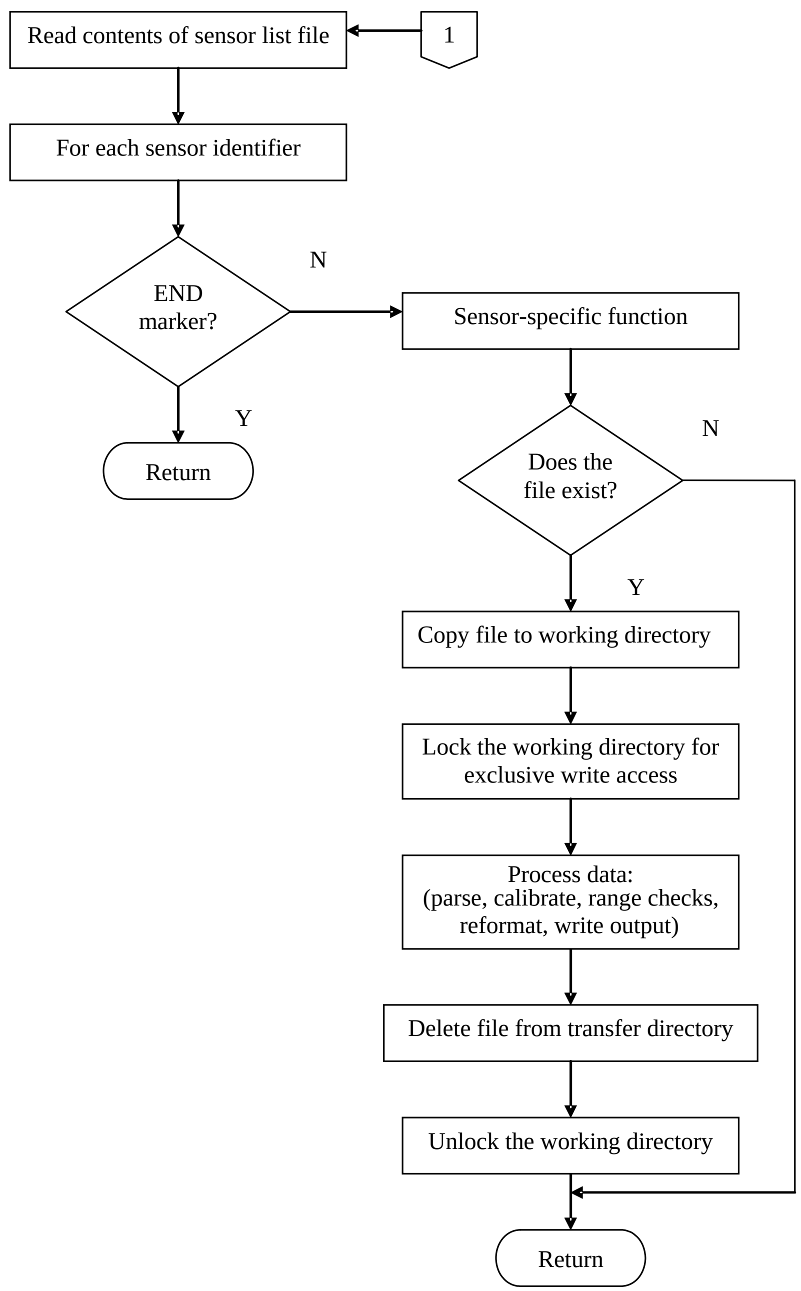


Figure 10. Functional flow diagram showing data processing details within the cvRTCache Program.

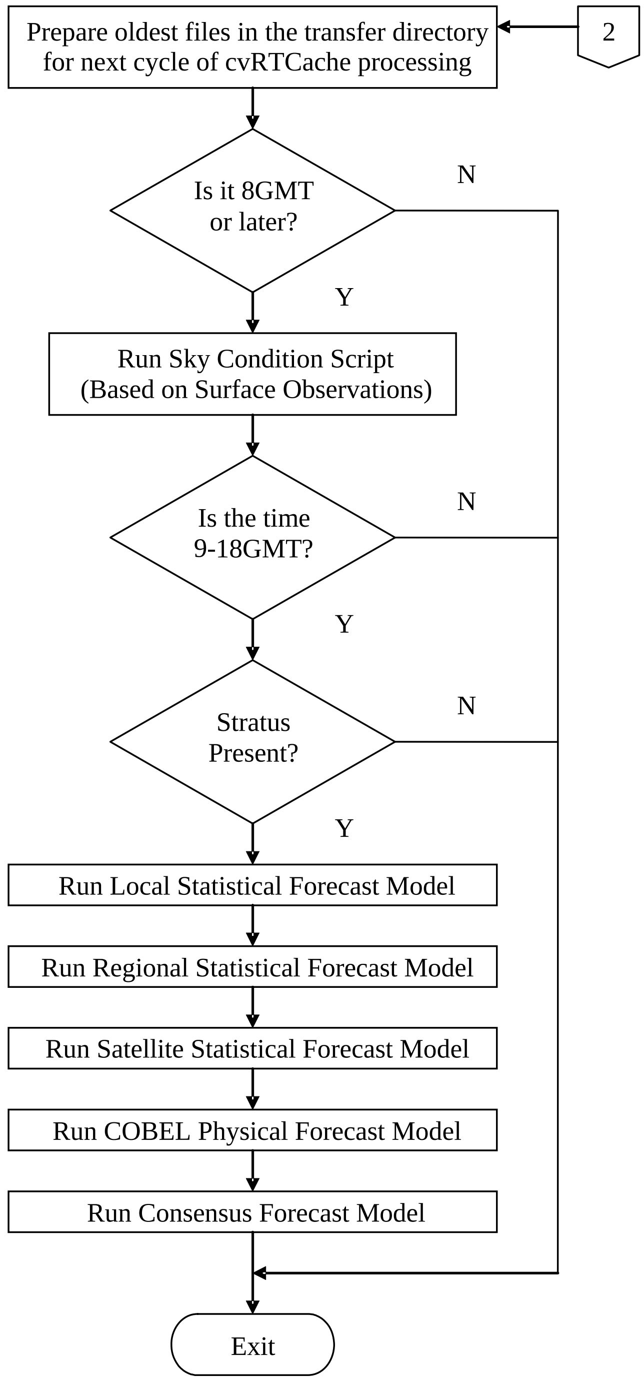


Figure 11. Functional flow diagram for the master post-processing script.

4.2 PRACTICAL CONSIDERATIONS

4.2.1 Programming and Resources

Table 8 summarizes the resources utilized by the cvRTCache.

Table 8

Computer Resources Utilized by the cvRTCache

Resource	Utilization (MB)
Memory Requirement for cvRTCache Executable	0.9
Disk Storage for a Day of Day Summary Files	3.1
Disk Storage for a cvRTCache Log File	2.8

4.2.2 Exception Handling

The cvRTCache is a data-driven program designed to process data placed in the transfer directory by the data acquisition system described in Section 3.0. As such, the cvRTCache is not sensitive or even aware of issues related to data timeliness or availability. These issues are handled by the data acquisition phase of the real-time system.

4.2.3 Quality Assurance and Diagnostics

The cvRTCache does not provide explicit data quality information in the output day summary files. Some of the sensor-specific data processing routines mentioned in Section 4.1.3.2 perform range checking on each field within a data record. A data value within a field is set to a missing value if the data value is outside the bounds of the range defined for that field. Records that contain date-time fields are discarded if the date-time within the record exceeds the computer wall-clock time.

As mentioned in Section 4.1.2, the cvRTCache maintains a log file that records the activities of the program along with the activities of programs and scripts that the cvRTCache activates. This log file is typically used as a first-level diagnostic tool in assessing problems that may be encountered with the cvRTCache or surrogate programs. Sensor-specific data processing routines also maintain individual log files, but their contents are generally limited to a record of data contents rather than software logic. More information on the use of these log files in a troubleshooting context can be found in Appendix F.

4.2.4 Operator Interaction and Default Settings

Interaction with the cvRTCache is through the four ASCII text files outlined in Section 4.1.1 and detailed in Appendices B through E. The appendices list the default settings contained in the files (as shipped with the software distribution) and document the format of each file and an explanation of each setting.

Normally the cvRTCache remains memory resident, exiting at the end of each day to roll over the log file to a new day as it restarts itself. The program runs within the confines of the script that is used to activate it, either manually or automatically through a cron job. When the program exits control returns to the calling script, which is designed to restart the program. There may be times when an operator may want to stop both the program and calling script to perform some task (e.g., install a new version of the cvRTCache). Appendix G details the procedures for manually stopping and starting the cvRTCache.

4.2.5 Periodic/Potential Maintenance Tasks

Directory trimming is performed automatically on the dated directories located under the directory */ll/cv/logs* using a cron job. Each dated directory contains the cvRTCache log file for that day as well as a subdirectory for each forecast model. Each forecast model subdirectory contains log files that archive the details of every run made by a model over the course of a day. The cron job allows the operator to adjust the number of days of log files to maintain online at any one time. There are also other cron jobs that perform directory and file trimming for other components of the data processing system. Appendix H documents each cron job within the crontab file that is used to trim directories and files. Included are the default settings (either number of files or directories to maintain online) that ship with the software distribution and documentation on how to change the default settings.

To decode and save additional hourly surface station reports that appear routinely in the file */ll/cv/data/Xfer/afos_sfo.dat*, edit the file /ll/cv/params/afos_sa_list.txt and add the three-letter station identifier(s) to the list as a separate record (see Appendix C). Make sure the end-marker “END” appears as the last record in the file. Since this file is read each time the raw file */ll/cv/data/Xfer/afos_sfo.dat* appears in the transfer directory, care should be taken not to edit this file when the raw file is being processed. When in doubt, follow the procedure in Appendix G to manually stop the cvRTCache, edit the file, and restart the cvRTCache using the restart procedure in Appendix G.

An entire calendar year of day summary files occupies, on average, about 1GB of disk space uncompressed and approximately 250MB compressed. Although this size is quite small compared to the disk capacity expected to be available on the operational base station computer, periodic backup of the data is warranted since these data are used in playback operations and statistical forecast model development.

At the conclusion of the stratus season the models can be shutdown so that they do not run on days that although cloudy, are not stratus days. The primary reason for stopping the models from running is that the models were designed to provide forecasts for marine stratus dissipation as opposed to the cloud conditions associated with synoptically-induced features typically observed in the off-season. Of secondary note is the needless waste of processing cycles and disk space that the models consume when they run. Appendix I provides the procedure to stop the models at the end of a stratus season and the procedure to allow the models to run at the start of a stratus season.

4.3 DATA PROCESSING COMPONENTS

The following sections detail the functions performed by the sensor-specific data processing modules. The modules are collectively represented in the "Process Data" function block of Figure 10.

4.3.1 Surface Data Processing

In addition to the data processing descriptions for surface-based observations given in the following sections, Appendix J provides a description of data characteristics and day summary record format.

4.3.1.1 Hourly Observations

Hourly observations in METAR format are contained in the text file *afos_sfo.dat*. Observations (including specials) with station identifiers that match those listed in the file */ll/cv/params/afos_sa_list.txt* are decoded and stored in individual station-based day summary files. Figures 12a and 12b illustrate the data processing steps for hourly observations.

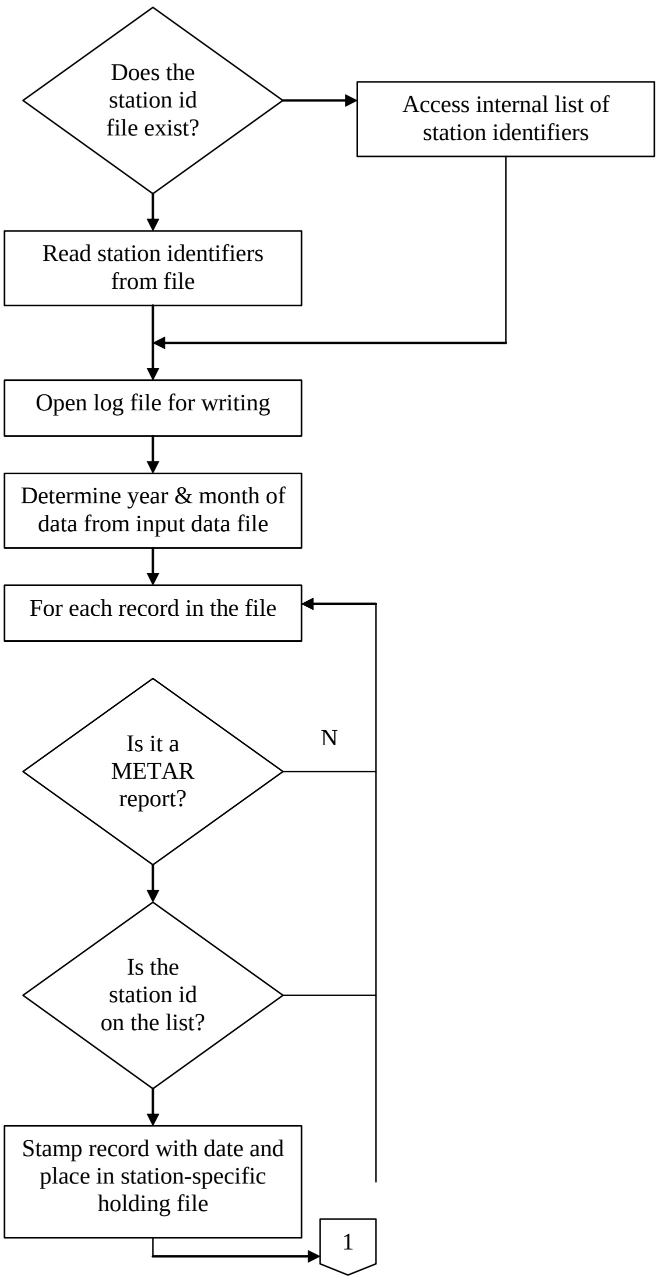


Figure 12a. Functional flow diagram for processing hourly observations.

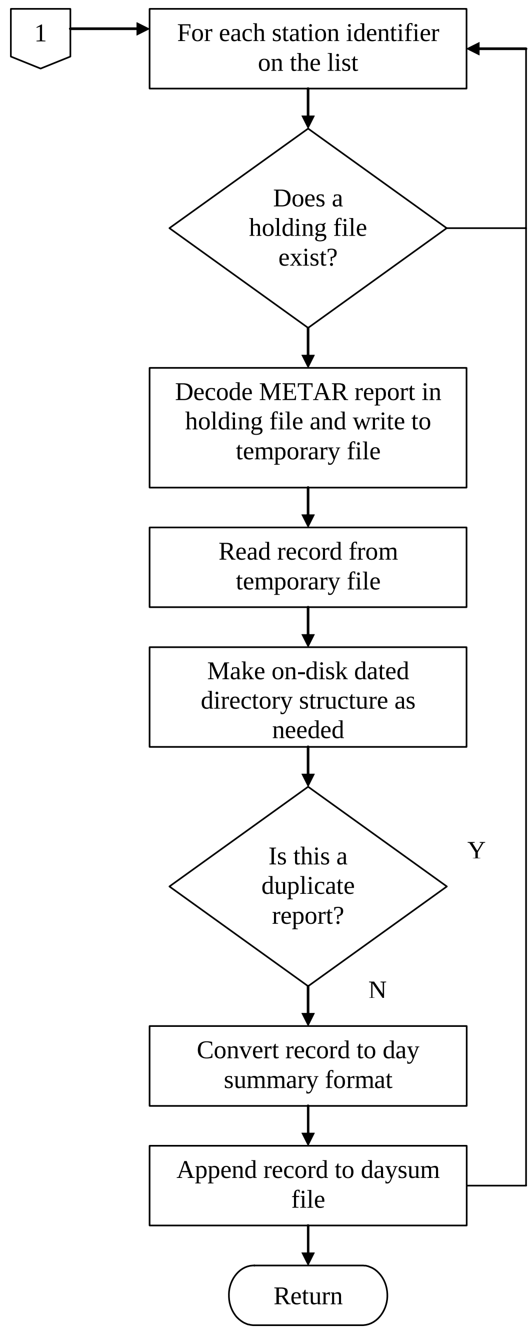


Figure 12b. Functional flow diagram for processing hourly observations (continued).

4.3.1.2 ASOS Observations

Five-minute ASOS observations for San Francisco Airport are contained in the text file *asos_sfo.dat*. To prepare ASOS observations for the METAR decoder (same decoder used to decode hourly observations), a program is called to strip leading date, time, and ASOS-specific identifying information off the report. Once decoded, reports are reformatted and written to the appropriate day summary file. Figure 13 outlines the data processing steps for ASOS observations.

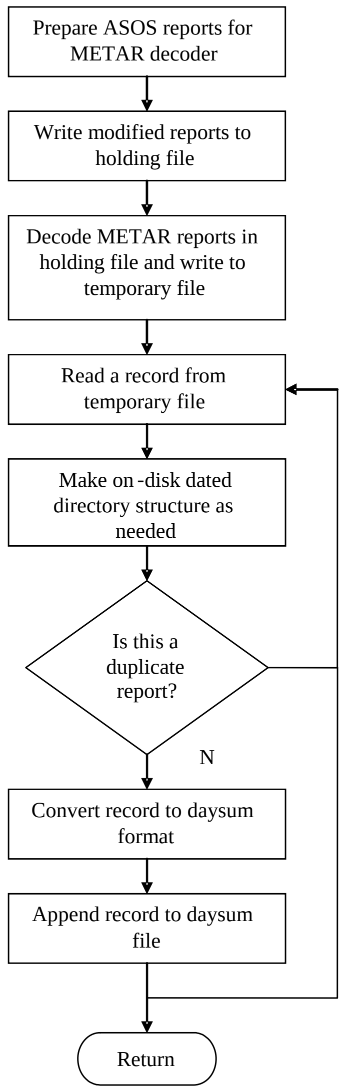


Figure 13. Functional flow diagram for processing SFO ASOS observations.

Five-minute AWOS observations for San Mateo Bridge are contained in the text file *awos_smo.dat*. The observations are decoded, reformatted, and written to the appropriate day summary file. Figure 14 shows the data processing steps involved in processing AWOS observations.

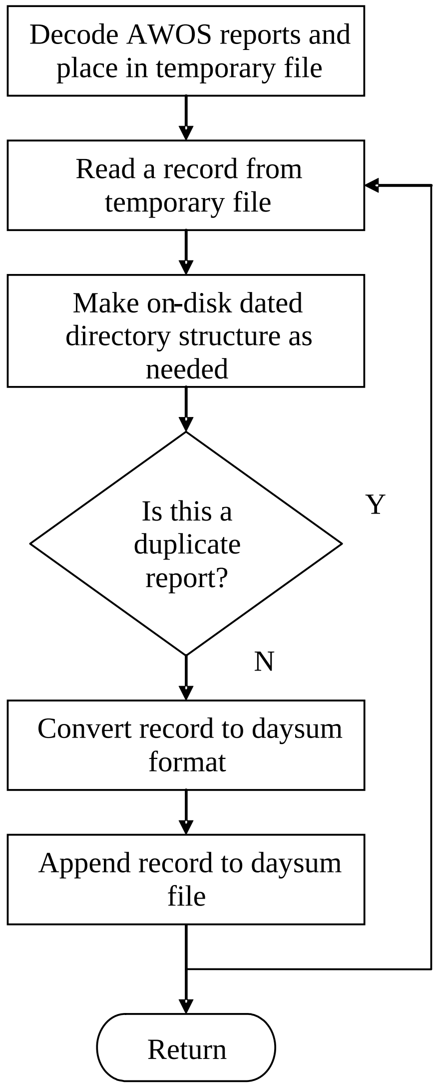


Figure 14. Functional flow diagram for processing SMB AWOS observations.

4.3.2 Upper-Air Data Processing

In addition to the data processing description for upper-air observations given in the following section, Appendix K provides a description of the NOAA/FSL upper-air data format adopted for use in storing the Oakland sounding.

4.3.2.1 Sounding Observation

The single sounding utilized in the San Francisco Marine Stratus Project is the Oakland sounding. The raw data are contained in the text file *afos_sfo.dat*. Mandatory-level pressure data and significant-level temperature and wind data are parsed, combined to form records containing both pressure and height levels, and formatted to the NOAA/FSL format. Figure 15 shows the data processing steps involved in processing the Oakland sounding.

4.3.2.2 Post-Processing

Once the Oakland sounding has been processed to its final form, a *csh* post-processing script is called to perform additional tasks that utilize the sounding. The processing steps to perform these additional tasks are shown in Figures 16a and 16b. First, copies of the sounding are placed in day summary files for all subsequent access. This is followed by activation of a program that computes an estimate of the inversion-base height based solely on the sounding. Results are written to a day summary file. Next, if the sounding just processed is different from a previously processed sounding, it is assumed that the new sounding represents either a corrected sounding or one with additional information. Under these circumstances, log files that record the progression of the component forecast models are moved aside (if they already exist) to force the models to re-run all forecast intervals up to and including the current interval. This insures that the models run with the latest sounding data and data derived from the sounding (e.g., sounding-based estimate of the inversion-base height).

4.3.3 Field-Site Data Processing

In addition to the data processing descriptions for field-site observations given in the following sections, Appendices J, L, and M provide information on data characteristics and day summary record formats.

4.3.3.1 Surface Observations

Five-minute field site surface observations from the San Francisco field site are contained in the text file *radi_sfr.dat*. Similar observations from the San Carlos field site are contained in the text file *flux_scf.dat*. For each record in the input data file, data are decoded, calibrated to engineering units (using constants read from an external text file), and written to a temporary file. Records from the temporary file are subsequently read, reformatted, and appended to day summary files; one file for the radiation data and another file for the conventional surface data. Figure 17 shows the data processing steps involved in processing field-site surface observations.

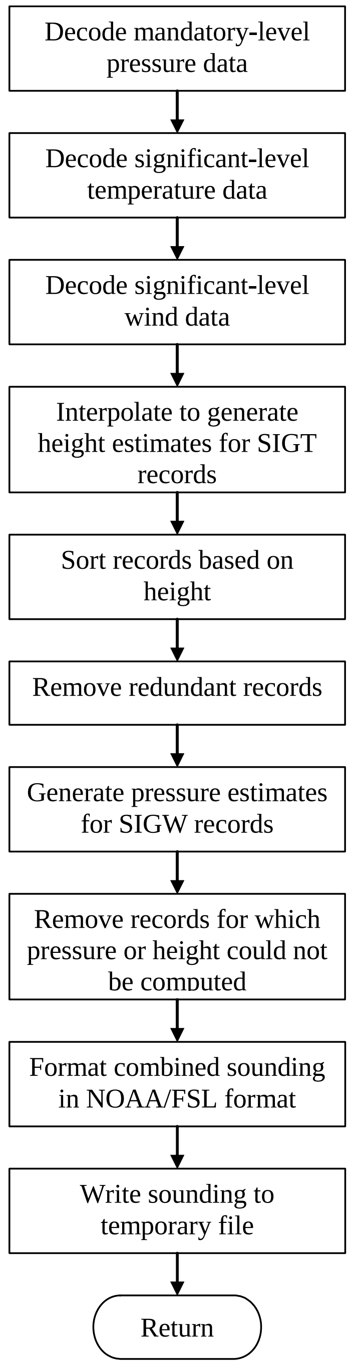


Figure 15. Functional flow diagram for processing sounding data.

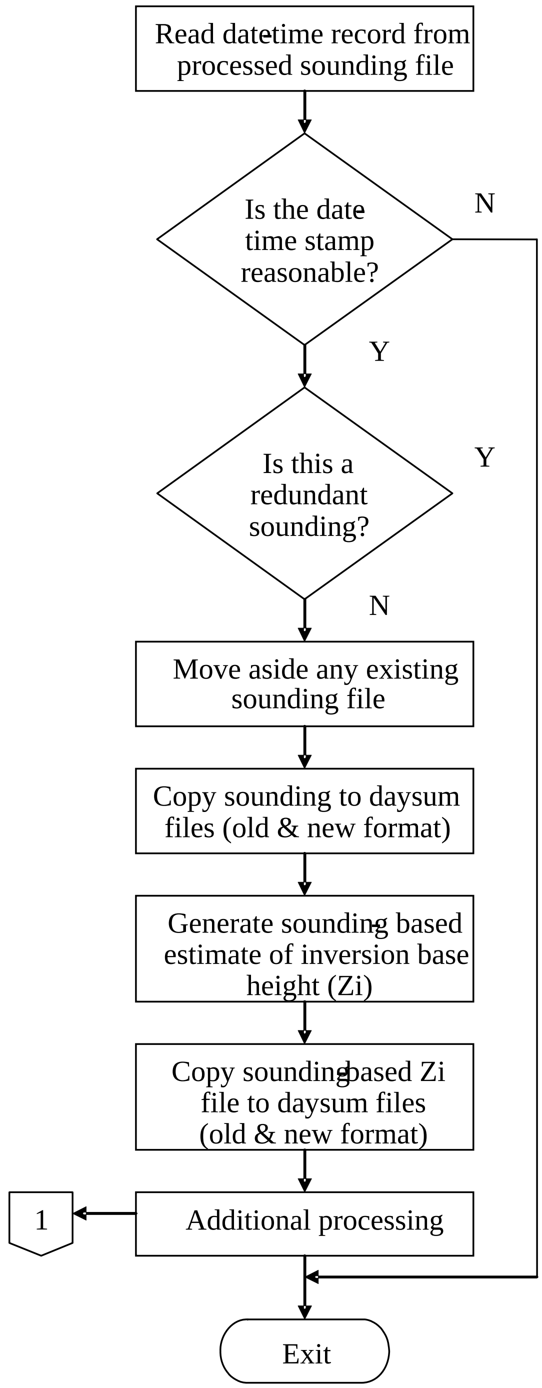


Figure 16a. Functional flow diagram for post-processing sounding data.

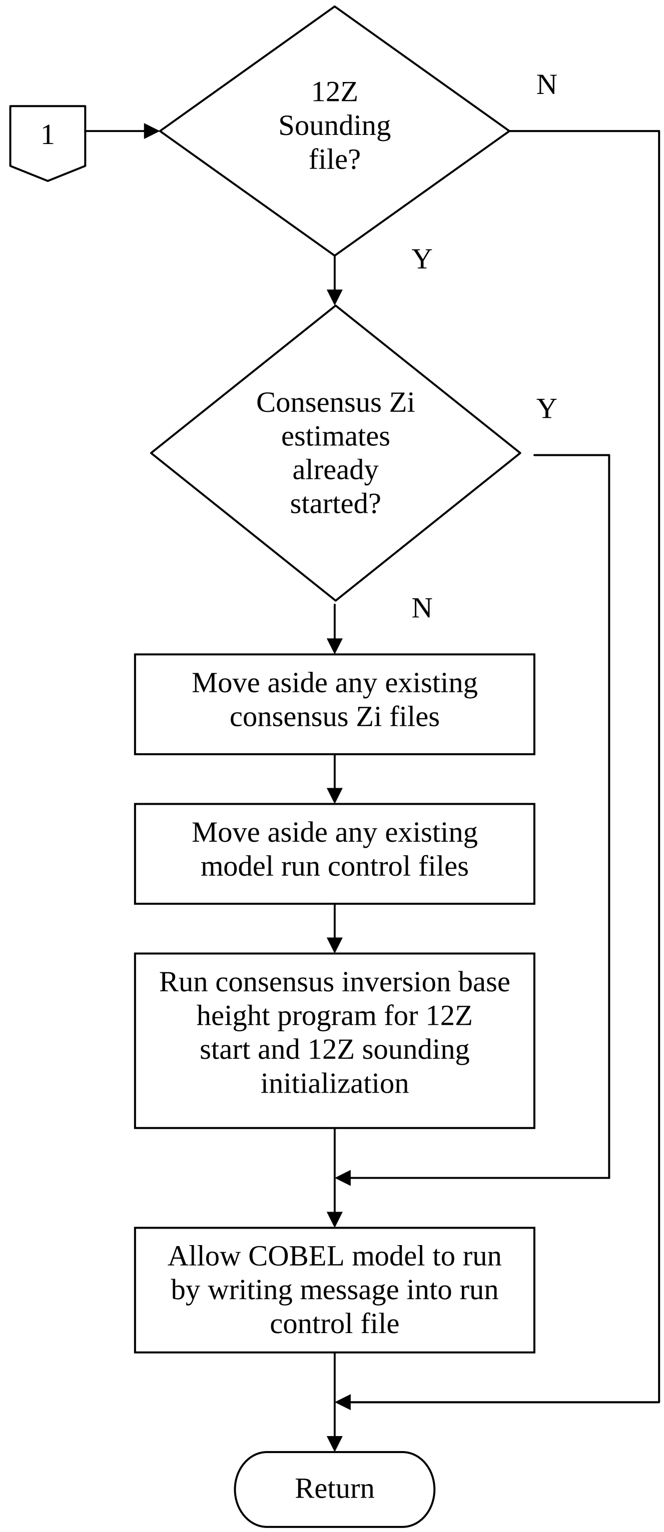


Figure 16b. Functional flow diagram for post-processing sounding data (continued).

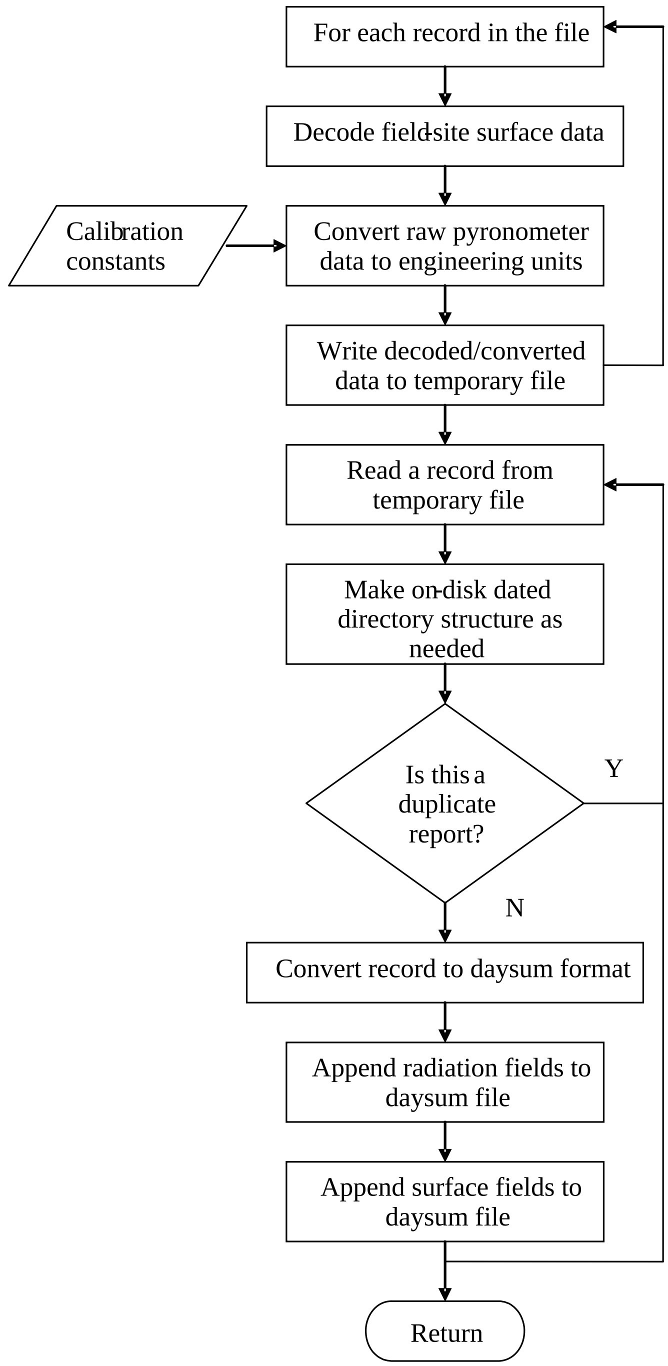


Figure 17. Functional flow diagram for processing field-site surface observations.

4.3.3.2 Sodar Data

One-minute field site sodar data from the San Francisco field site are contained in the text file *sodr_sfs.dat*. Similar data from the San Carlos field site are contained in the text file *sodr_scs.dat*. Sodar data records are appended unmodified to appropriate day summary files based on the date-time of the data contained in each record. The functional flow diagram for processing sodar data is shown in Figure 18.

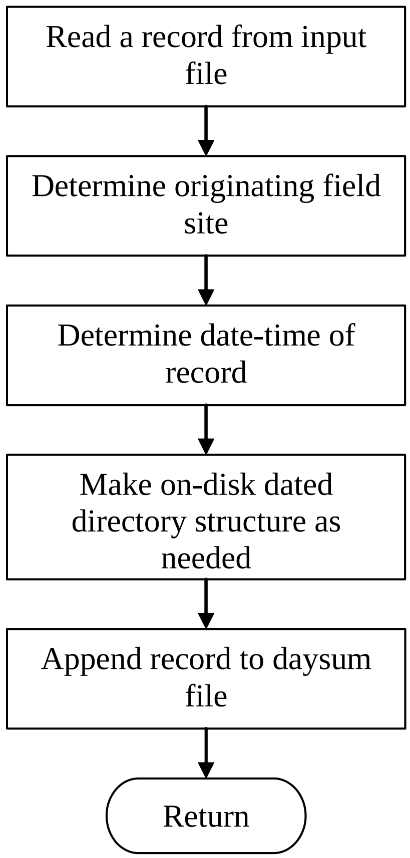


Figure 18. Functional flow diagram for processing of sodar data.

4.3.3.3 Post-Processing

Once sodar data have been placed in day summary files, a *csh* post-processing script is called to generate an estimate of the height of the inversion base using sodar data and other ancillary information. The script begins by determining if estimates of the inversion base height have already begun. If so, a new computation is generated for the current time step and the script terminates. If not and if the data are within the range of pre-dawn hours (9-11GMT), meteorological conditions are assessed and a determination is made as to whether it is feasible to start the computation of inversion base height during the pre-dawn hours. A negative determination will delay the start of the computations until either an independent estimate of the inversion base height has been made (based on the Oakland sounding) or the time exceeds 14GMT and the system has not received a morning sounding. The functional flow diagram in Figure 19 outlines these steps.

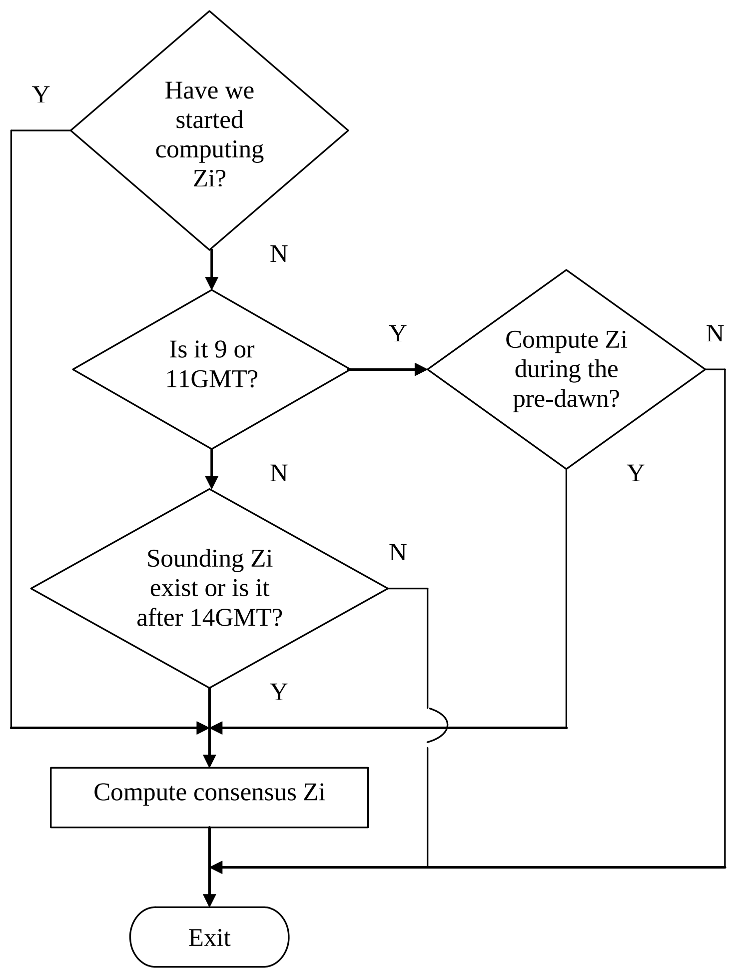


Figure 19. Functional flow diagram for post-processing of sodar data.

4.3.4 Satellite Data Processing

In addition to the data processing description for satellite data given in the following sections, Appendix N provides information on data characteristics and days summary record formats.

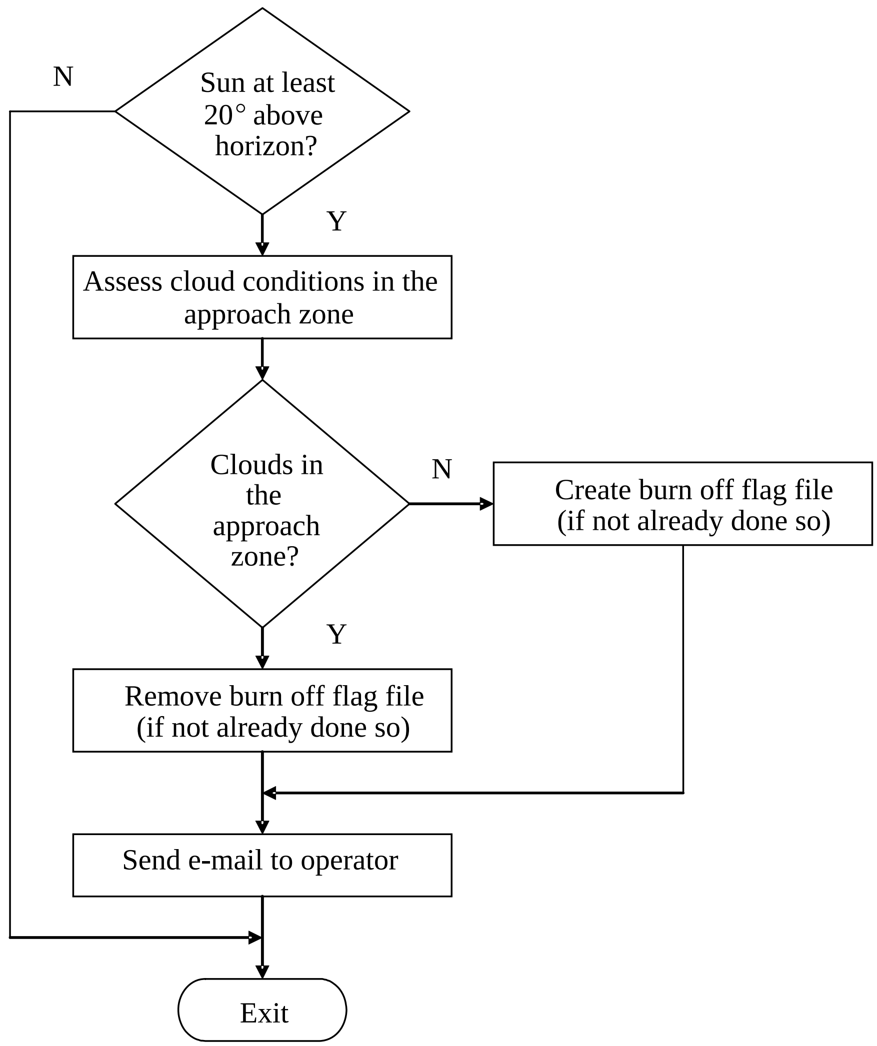


Figure 20. Functional flow diagram for post-processing of satellite data.

4.3.4.1 Satellite Data

Fifteen-minute satellite image data obtained from NRL/Monterey and preprocessed after acquisition are contained in the text file *vis_data.dat*. A companion file containing processed satellite data statistics are contained in the text file *vis_sum.dat*. The data are appended unmodified to day summary files.

4.3.4.2 Post-Processing

Once satellite data have been placed in day summary files, a csh post-processing script is called to assess cloud conditions in the approach zone. The steps carried out by this script are shown in the functional flow diagram of Figure 20. If no cloud is detected, a burn off flag file is created (if it does not already exist) and a date-time stamp is appended to the file indicating when the determination was made. If cloud is detected, the burn off flag file is deleted (if it exists). The absence of a burn off flag file is an indication that forecast models should be running if the time is within the forecast interval. The presence or absence of the burn off flag file is one piece of information the master post-processing script examines to control the activation of the forecast models (the other is forecast interval). The only other process that can create or delete the burn off flag file is a companion script that makes an assessment of cloud based solely on surface observations in and around the approach zone. This script was described in Section 4.1.3.3.

4.3.5 Derived Data Processing

4.3.5.1 Sounding-Based Inversion Base Height Product

The sounding-based inversion base height product is generated by the post-processing script (described in Section 4.3.2.2) as a result of receiving either a 0 or 12GMT Oakland sounding. The product provides a sodar-independent estimate of the height of the base of the temperature inversion (Zi) that is prevalent along the Northern California coast during the stratus season that runs from late spring through early fall. The algorithm employed to generate the product was developed with the following assumptions:

1. The objective is to determine Zi that corresponds to the top of the stratus layer.
2. There can be only one estimate per sounding.
3. Temperature gradient and moisture information are the primary sources examined in the analysis.
4. If multiple layers are present, a measure of quality is assigned to distinguish the most probable layer with a bias towards higher layers.
5. The algorithm should be able to run each day so as to provide a reference archive for stratus days that possessed no strong indication of an inversion base.
6. Focus is on layers found between 300 and 1500 meters.

The output generated is the best estimate of Zi in meters and a quality factor. The quality factor is a measure of the confidence that the chosen inversion layer is likely to be the layer associated with the inversion base capping the marine stratus. The value is derived from a cumulative analysis of layer statistics including temperature gradient, dew point depression difference, the presence or absence of cloud below the layer, and temperature difference. The range of values is 0 - 100 in increasing confidence. Values of 60 and higher indicate the analysis found cloud at the base of the inversion. Appendix O provides information on data characteristics and the day summary record format for this product.

4.3.5.2 Sodar/Sounding-Based Inversion Base Height Product

The Sodar/Sounding-based inversion base height product is generated by the post-processing script (described in Section 4.3.3.3) as a result of receiving sodar data from one or both field sites. The product provides a consensus estimate of Zi, employing both the sounding-based estimate of Zi and sodar data. If cloud has persisted in the area of the approach zone for at least one hour prior to 9 or 11GMT, the computation start time for the product is either 9 or 11GMT. Otherwise, the computation start time is 12GMT and begins once the 12GMT Oakland sounding is processed by the cvRTCache. The program begins by establishing a first-guess estimate for Zi. If the start time is 9 or 11GMT, the first-guess is chosen from one of the following sources:

1. 0 GMT Oakland sounding, if a cloud ceiling is present and Zi is above the ceiling.
2. From the last consensus Zi of the previous day if the algorithm was tracking a layer and the layer is above the cloud ceiling.
3. Computed from the surface observation of cloud ceiling and a statistically derived mean cloud thickness.

Otherwise, if the start time is 12GMT, the first-guess estimate is taken from the 12GMT Oakland sounding. If by 14GMT no 12GMT Oakland sounding has been acquired by the cvRTCache, the program will begin making consensus Zi estimates using only sodar data. Next, sodar data collected from 0GMT to the current time are analyzed to attempt to identify one or more credible layers. The first-guess estimate is then used to guide the selection of one of the candidate layers as the best estimate of Zi. For the remainder of the day as new sodar are collected, data are analyzed to track (and possibly adjust) the estimate of Zi. Refer to Appendix O for information on data characteristics and the day summary record format for the consensus Zi product.

5. COMPONENT FORECAST MODEL ACTIVATION AND PROCESSING

5.1 TECHNICAL DESCRIPTION

Each of the four component forecast models are activated by means of model-specific launching script. These scripts are called in serial fashion by the master post-processing script described in Section 4.1.3.3. Although portions of each script are model-specific, they do share some common logic described in the following sections.

5.1.1 Data Inputs

Sensor data contained in day summary files are accessed to derive day-type parameters and model predictors. The model run control file is accessed to manage the progression of forecast runs over the course of a day.

5.1.2 Data Outputs

The results of attempting to generate a forecast, whether successful or not, are logged to the model run control file. If a successful forecast is generated, the contents of the raw output file produced from the model evaluation program are copied to an area accessed by the web display software. The raw output includes the burn off time, quality factor, and statistics for each predictor. Burn off time and quality factor are also appended to a model forecast day summary file that is subsequently accessed by the consensus model to compute a consensus forecast.

In addition to the model run control file, additional log files generated during each model run are written to a model-named subdirectory under */ll/cv/logs/YYYYMMDD*. Table 9 lists the generic log file name and a description of file contents for each log file generated during a forecast interval run.

5.1.3 Structural Flow

Functions common to all model-launching scripts are depicted in the functional flow diagram of Figures 21a and 21b. For the first invocation of the day, the script begins by creating a log directory that will be used to store log files generated for each forecast interval run that is made. The directory is created under the path */ll/cv/logs/YYYYMMDD* and is named with the model name reflected in the model-launching script. Next, a list of model-specific forecast run intervals is defined. The list is referenced to guide the forecast model activation process to attempt to produce forecasts for only those run times specified in the list. The list is fixed within the script but could easily be changed in the unlikely event that forecasting requirements change. Table 10 lists the forecast run times for each component model. Note that the list represents only a guide as to the forecast intervals requested to be run. The ability to actually produce a forecast for a particular run time and model is determined by data availability.

Table 9
Log Files Used in Forecast Model Activation and Processing

Generic File Name	Description
<*Script Name*> ..log.YYYYMMDD_HHMMSS	Log file used by model launching script. Contains information on forecast interval run logic and choice of model type for run.
<*Mdl Name*> _HHMM.inputdata. <*Mdl Type*> .YYYYMMDD_HHMMSS	Input file to forecast model evaluation program. Contains name/value pair and date-time stamp for each predictor
<*Mdl Name*> _HHMM.raw.YYYYMMDD_HHMMSS	Contains data (and associated date-time stamp) read from day summary files. Used to derive predictor values.
<*Mdl Name*> _HHMM.YYYYMMDD_HHMMSS.0.log	Output file produced by the forecast model evaluation program. Identifies the model type chosen and resulting burn off time and quality factor. Also includes statistics for each predictor.
<*Mdl Name*> RunSummary.YYYYMMDD	Model run control file. Logs the success or failure of every forecast interval run made over the course of a day.

Where: <*Script Name*> is one of:

ProcessLocalModel
ProcessRegionalModel
ProcessSatelliteModel
ProcessCobelModel

<*Mdl Name*> is one of: Local, Regional, Satellite, Cobel

When <*Mdl Name*> is Local, <*Mdl Type*> is one of:

09Default, 11Default, 13ZiQGE55, 13ZiQLT55, 14ZiQGE75, 14ZiQLT75, 15ZiQGE90, 15ZiQLT90, 16ZiQGE90, 16ZiQLT90, 17Default, 18Default

When <*Mdl Name*> is Regional, <*Mdl Type*> is one of:

HHDefault, HHOnshore, HHOffshore, HHNeutral
(HH = 09, 11, 13, 15, 16, 17, 18)

When <*Mdl Name*> is Satellite, <*Mdl Type*> is one of:

14Default, 14HiCig, 14LoCig, 15Default, 15HiCig, 15LoCigRufCld, 15LoCigSmthCld, 16Default, 16HiCig, 16LoCigRufCld, 16LoCigSmthCld, 17Default, 17HiCig, 17LoCigRufCld, 17LoCigSmthCld, 18Default

All files located in:
/ll/cv/logs/YYYYMMDD/<*Mdl Name*>Model

Table 10

Daily Forecast Run Times by Model

Model	Forecast Run Times
Local	9[1], 11[1], 13, 15, 16, 17, 18
Regional	9, 11, 13, 15, 16, 17, 18
Cobel	9[1], 11[1], 13, 15, 16, 17, 18
Satellite	15, 16, 17, 18
[1] Conditional upon generation of consensus Zi product.	

Once a list of run times is established, the logic chooses the first run time from the list. The run time is then checked against the last run time on the list to see if all runs have been made. A check is also made to see if the run time actually appears on the list. If a run time does not appear on the list, it is advanced by one hour and the logic is repeated. Otherwise, the model run control file is queried to see if a run with this run time has been made. If not, a run is made if the current time is equal to or exceeds the run time. If at least one run has already been made, the current time is checked to see if it is time to move on to the next run time. If so, the current run time is advanced by one hour and the logic is repeated. Otherwise, a check is made to see if an additional run of the current run time should be made. For a run time that has not yet generated a successful forecast, the update rate is every 2 minutes until a successful forecast is made. Subsequent successful runs are then made every 4 minutes.

The additional processing referenced in Figure 21a is expanded in detail in Figure 21b. Processing enters this stage if the previous logic determines that a run should be made. Each launching script contains model-specific code that defines the suite of models to choose based on run time and the choice of one of the models from that suite based on daytype thresholding and data availability. In the event that a chosen daytype model cannot be utilized because of missing data, a default or backup model is selected. On rare occasions the default model may not run due to missing data in which case, no forecast is produced.

If a model is chosen and data availability is not an issue, data from the day summary files are gathered to derive the predictors for the model. The model is then run to produce a forecast. Output from the model is written to a number of areas:

1. Raw model output is appended to:
   /ll/cv/data/cvRTCache/LATEST/<*Mdl Name*>.KeyParams

2. Burn off time, quality factor, and model type are appended to the forecast model day summary file:
   /ll/cv/data/daysumYYYY/fcsts/YYMMDD/fcsts_<*Mdl Name*>.YYMMDD.sum

3. Burn off time, quality factor, and model type are appended to the model run control file:
   /ll/cv/logs/YYYYMMDD/<*Mdl Name*>/<*Mdl Name*>RunSummary.YYYYMMDD

5.1.4 Nominal and Degraded Performance

When data availability issues prevent a daytype model from running, a default model is substituted to run in its place.

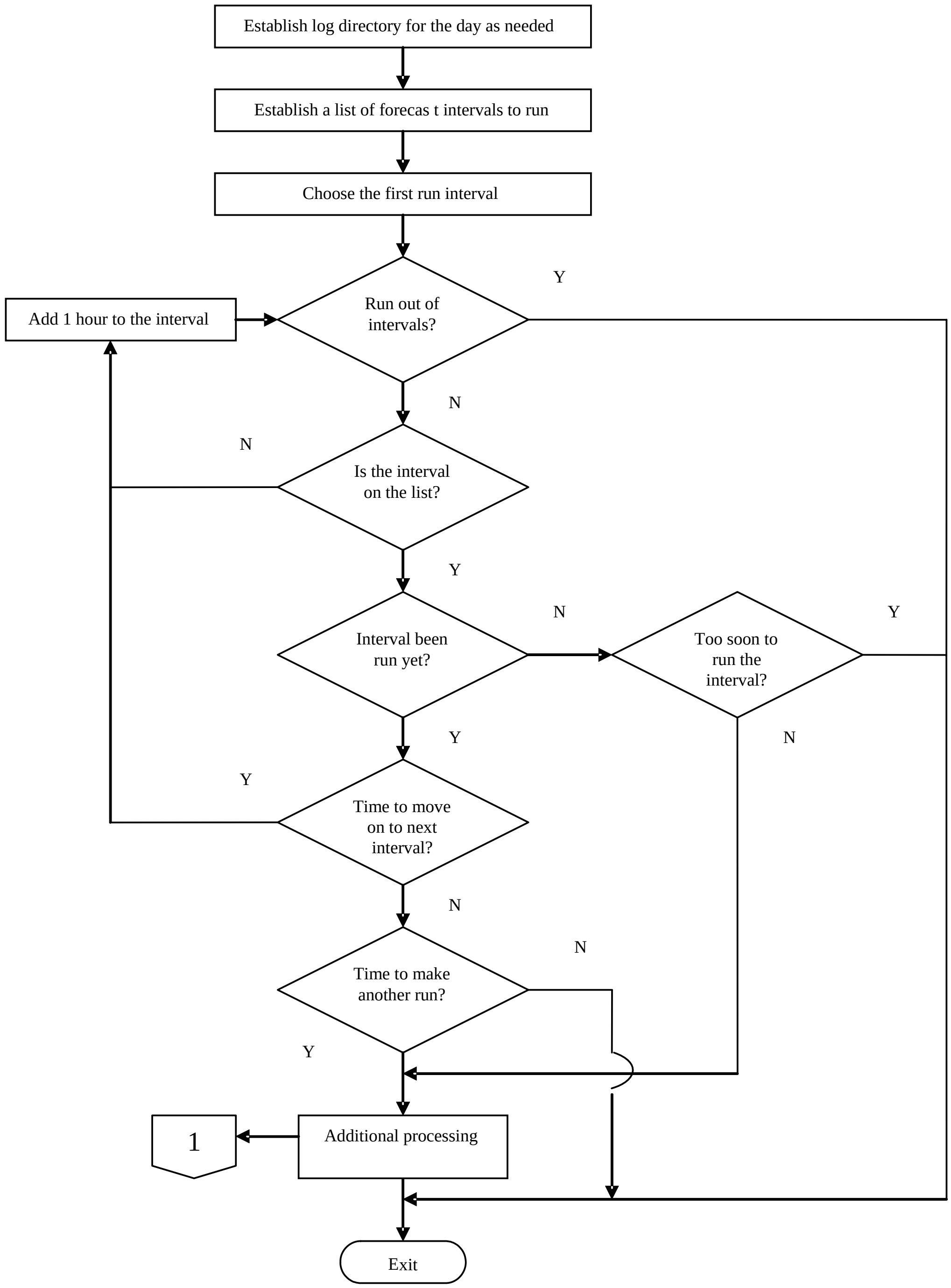


Figure 21a. Functional flow diagram for the forecast model launching scripts.

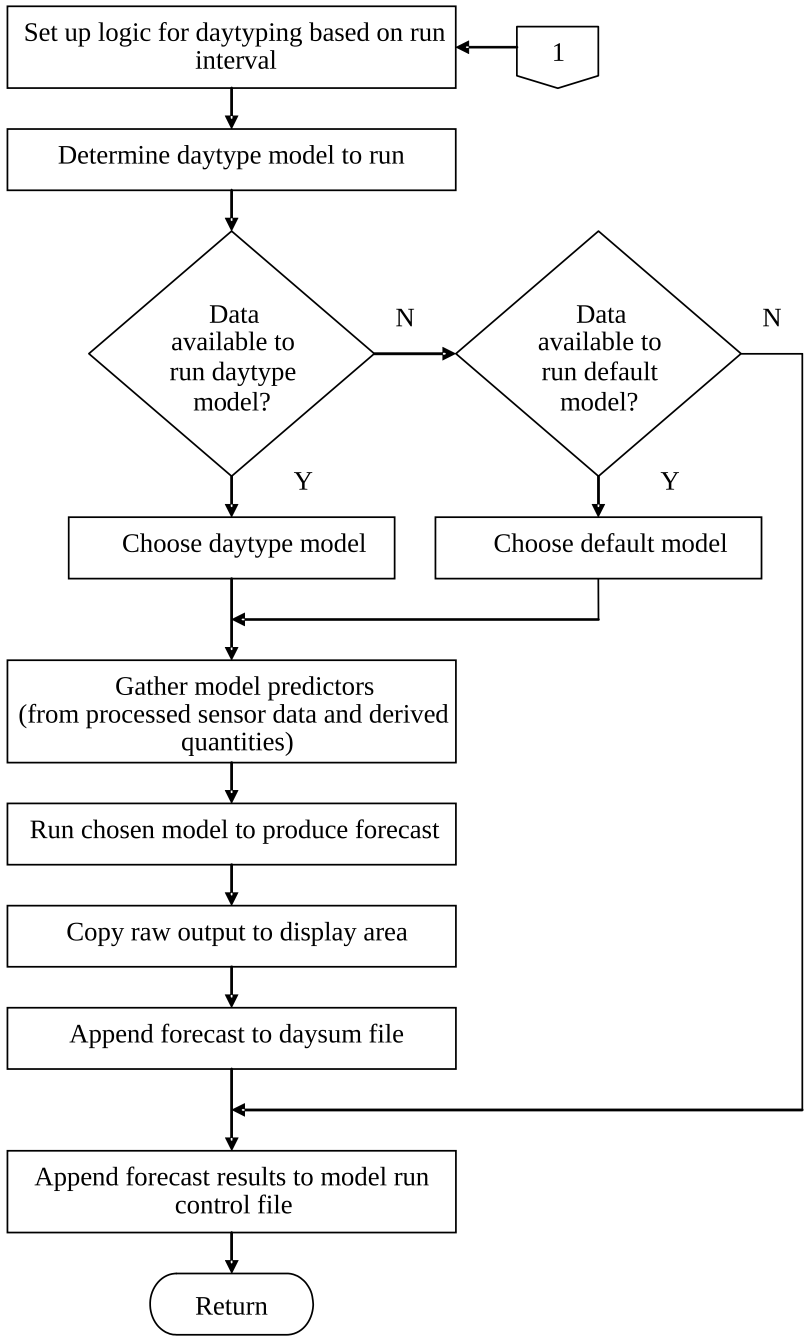


Figure 21b. Functional flow diagram for the forecast model launching scripts (continued).

The quality factor associated with a burn off time is an indication of the timeliness of the data that was used to derive the predictors for the model. A quality factor less than 5 indicates that one or more predictors was derived from data less than nominal in time.

5.2 PRACTICAL CONSIDERATIONS

5.2.1 Programming and Resources

Day type and default model definitions are defined within the launching scripts. If new models are introduced into the system, it may be necessary to modify the launching scripts.

5.2.2 Exception Handling

The launching scripts are designed to identify and run the model that can produce a forecast based on the data that are available at the time of the run. By design, the system continues to attempt to run models every few minutes with the intent that as some point within a forecast interval, the most up-to-date data will be available to produce the most relevant forecast. As such, there is no special logic to handle situations in which the most up-to-date data never arrives. Instead, a default model is chosen that can run with the data that are available.

5.2.3 Quality Assurance and Diagnostics

Forecast quality, expressed as a quality factor, is an indication of the timeliness of the data used to derive model predictors.

The log files described in Table 9 can be used as a diagnostic tool to determine why a particular model ran or didn't run or why no model ran. The model run control file posts the results of every run and will indicate if a run was unsuccessful due to missing data. To investigate further, the model evaluation program input file can be examined to determine which predictor(s) was missing and prevented the model from running.

5.2.4 Operator Interaction and Default Settings

Model launching scripts are called by the master post-processing script so no operator interaction is required. Default settings for the model forecast run intervals are shown in Table 10.

5.2.5 Periodic/Potential Maintenance Tasks

If new models were introduced into to the operational system that included day type name changes and/or selection logic, the corresponding launching script would need to be changed.

6. CONSENSUS MODEL ACTIVATION AND PROCESSING

6.1 TECHNICAL DESCRIPTION

Following the completion of the component forecast model-launching scripts; the master post-processing script completes the post-processing cycle with activation of the consensus model. The following sections provide a description of the consensus model from the perspective of its implementation within the real-time system.

6.1.1 Data Inputs

The consensus model requires the following input:

1. Run-time parameters that define resources required by the model.
2. The latest forecast burn-off time and quality factor from one or more of the component models.
3. Certain sensor information to examine meteorological conditions and compute a quality assessment based on whether the conditions are typical or atypical for summertime marine stratus.
4. Consensus weights that are applied to the component model forecasts to produce a consensus forecast.

General run-time parameters are defined in a Python script file (the consensus model is implemented as a collection of Python scripts). This file contains run-time parameters including: how often a consensus forecast is attempted within a forecast run interval, the location and generic names of the files containing the component forecasts, the location and generic names of output files, the location and generic name of the log file, the location and name of the file containing the consensus weights, and the run schedule. Appendix P contains the location and name of this file along with a description and format of its contents. Section 6.2.0 describes the conditions under which the run schedule would need to be modified.

As mentioned in Sections 5.1.2 and 5.1.3, each component model appends output to a day summary file. Thus, the latest component forecast will reside at the end of the file. The consensus model reads the latest forecast from each of the component model day summary files as input.

Select sensor information is input to assess the meteorological conditions as either typical or atypical for summertime marine stratus and to assign a low confidence to the consensus forecast if the conditions are deemed atypical. Sensor data examined include the consensus Zi height and quality factor, reported ceiling heights from stations in and around the approach zone, and satellite data.

Consensus weights are derived from the historical performance of each component forecast suite, which includes daytype and default models. The consensus model reads the consensus weights file to extract the appropriate weights based on model type and forecast run time. Appendix Q contains the location and name of this file along with a description and format of its contents.

6.1.2 Data Outputs

The consensus model writes output to a number of areas. A day summary file is created to append records for each consensus forecast made during a forecast run interval. This file resides alongside the component model day summary files used as input to the consensus model. A file is also generated for the web page display, containing forecast run interval, forecast burn off time, forecast confidence, and forecast weight for each component model and the consensus model. This file also contains results from the marine stratus quality assessment test.

6.1.3 Structural Flow

The consensus model software logic defines and references a set of structures (or dictionaries in Python) to decide when to attempt to run the consensus model. For each forecast run-time interval, there are two sub-intervals with characteristics defined in separate dictionaries (see Appendix P). The first sub-interval defines the expected time interval (start and stop times) within which component models are expected to run when marine stratus is already established. The stop time is a cut-off time that represents the latest time one is willing to wait for all relevant component models to complete before running the consensus model (the consensus model will run if at least one component model has produced a forecast). The second sub-interval covers the remaining time within a forecast run-time interval and seeks to cover instances in which marine stratus does not become established until later within an interval. In the following discussion of structural flow, references to run intervals refer to these sub-intervals.

Consensus model processing is illustrated in the functional flow diagram of Figure 22. Processing starts by generating a log file that records the activity for the current run of the consensus model. The log file is placed in the standard area for log files *(/ll/cv/logs/YYYYMMDD*) under subdirectory *ConsensusModel*. The input time is then examined to see if it falls within the current run interval. If so, a check is made to see if the run-time corresponds to a retry time. If not, the program exits. If so, a check is made to see if a consensus forecast has already been made for the current run interval. If so, the program exits. If not, forecasts generated from component models that ran are collected and examined. If all component models expected to produce a forecast for the current run interval have all produced forecasts of the highest quality, a consensus forecast is generated. Otherwise, the program exits.

Following the computation of the consensus forecast, sensor data are read and examined to make a daytype quality assessment based on the meteorological conditions as being typical or atypical for summertime marine stratus. The consensus Zi height and quality are examined to determine whether or not a well-defined inversion has been established and is being tracked. Ceiling height reports from stations in and around the approach zone are examined to see if the cloud-base heights fall within a range that is typical for marine stratus. Satellite data are processed to estimate a cloud-roughness factor to discriminate marine stratus from other cloud. For each component assessment, a quality factor is computed indicating either low quality (atypical conditions) or nominal quality (typical conditions). A low quality indication from at least one of these assessments will cause the web display software to display “low confidence” in the forecast window. Processing then concludes by writing component and consensus forecast burn of time and related statistics to a day summary file and to a file for web display. Results from the daytype quality assessment are also written to the web display file.

If the current time is not within the current run interval, the previous run interval is examined to see if a forecast was produced for that interval. If not, forecasts generated from component models that ran are collected. If at least one component model ran, a consensus forecast is computed and processing continues as described above. It is this branch of logic that executes when a run interval has ended and a consensus forecast has not been generated because one or more models failed to run (most often due to lack of data) or one or more models never produced a forecast with the highest quality factor (due to the use of less-timely data).

6.1.4 Nominal and Degraded Performance

Nominal performance is achieved when all component models expected to run for a particular run interval produce forecasts. At least one component model must produce a forecast within a forecast run-time in order for the consensus model to run.

6.1.5 Practical Considerations

6.1.6 Programming and Resources

The arrival of sensor data into the operational base station computer may vary from what has been typically found using the research version of the system. Depending on the magnitude of the difference, it may be necessary to adjust one or more sub-interval start and stop times within the dictionaries (see Appendix P to balance the attempts of the consensus model with the ability of the component models to run based on data availability.

6.1.7 Exception Handling

There is no exception processing performed by the consensus model if one or more component models doesn’t run. When a stop time for an interval is reached, the consensus model will utilize the component model forecasts that are available (assuming there is at least one) to generate a consensus forecast.

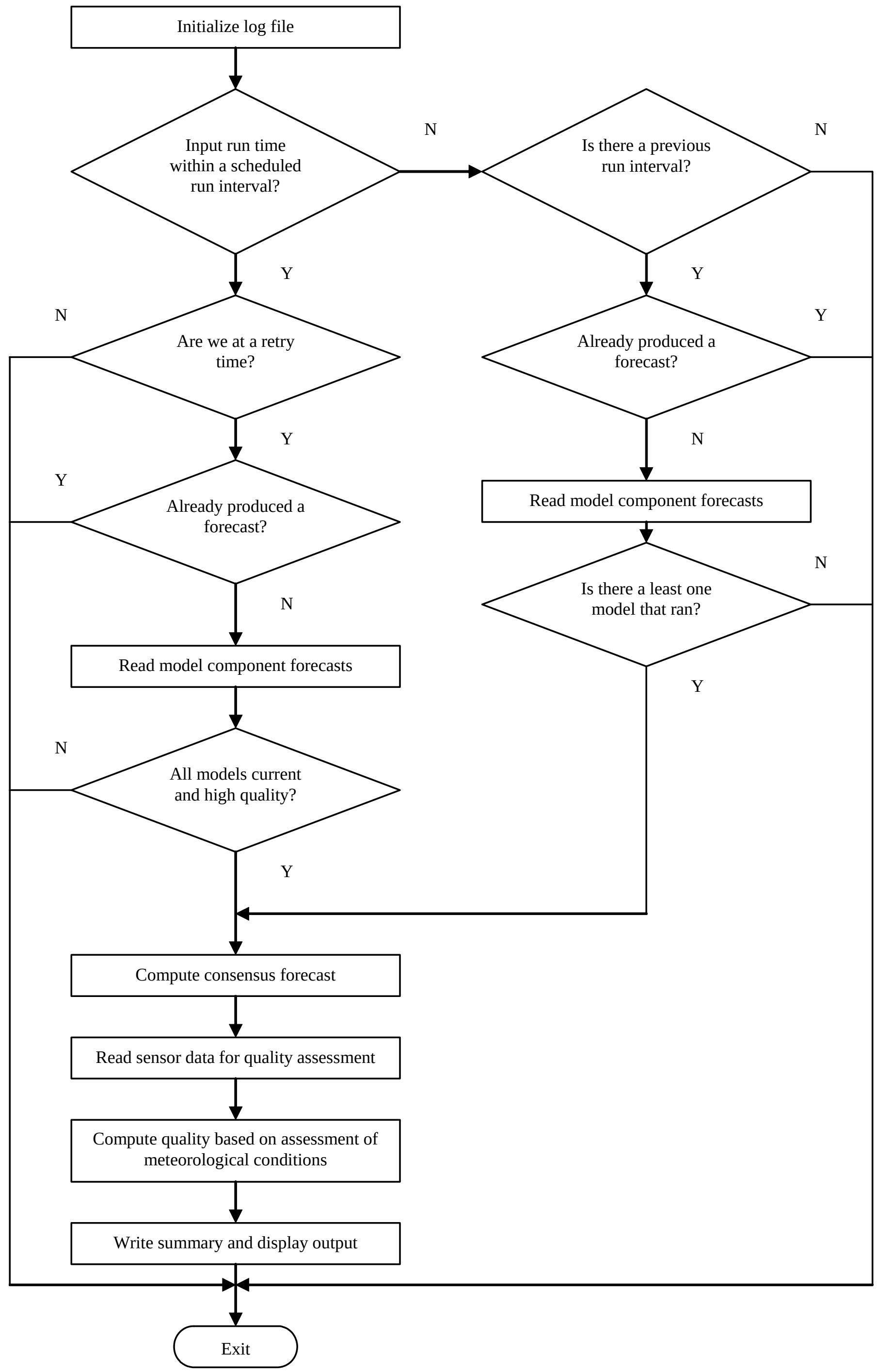


Figure 22. Functional flow diagram for the consensus model.

6.1.8 Quality Assurance and Diagnostics

The consensus model computes a confidence factor for each model. The value indicates the confidence in the forecast made by a forecast model for a particular run-time and daytype based on historical performance. The web page display system displays confidence as one of three categories (LOW, MEDIUM, or HIGH).

6.1.9 Operator Interaction and Default Settings

The consensus model is called by the master post-processing script so no operator interaction is required to activate the model. Run interval characteristics and their default settings are shown in the Python dictionaries of Appendix P.

6.1.10 Periodic/Potential Maintenance Tasks

The time that data will generally be available to allow the component models to run with the most up-to-date data may differ between the current research system and the target operational system. Once the operational system is collecting data and running models, run interval start and stop times should be established and the defaults should be adjusted if necessary.

If existing models are modified or new models are created using the automated statistical forecast development system, a file similar to the file shown in Appendix Q will need to be generated containing the new model weights.

APPENDIX A
E-MAIL MESSAGES ORIGINATING FROM THE BASE STATION

Subject: ST1 EXTREMELY-URGENT: CWSU SBC ROUTER DOWN
Date: Tue, 6 Jul 2004 20:24:54 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

CWSU SBC ROUTER NOT PINGABLE FROM CWSU BASE STATION (ST1)
If this repeats several times, contact NWS and/or SBC technical staff
Or try manually telebooting by dialing 510 793-3623 /bo 5,Y
SBC ROUTER at CWSU may be down

Subject: ST1 EXTREMELY-URGENT: SBC DSL AT CWSU DOWN
Date: Tue, 6 Jul 2004 20:24:50 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SBC INTERNET ROUTER NOT PINGABLE FROM CWSU BASE STATION (ST1) WHEN CWSU SBC ROUTER IS PINGABLE FROM ST1
If this repeats several times, contact SBC technical staff
DSL at CWSU is probably down
DSL at CWSU may be down

Subject: ST1 EXTREMELY-URGENT: CWSU PIX FIREWALL DOWN
Date: Tue, 6 Jul 2004 20:24:48 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

CWSU PIX FIREWALL NOT PINGABLE FROM CWSU BASE STATION (ST1)
If this repeats several times, contact NWS technical staff
Or try manually telebooting by dialing 510 793-3623 /bo 2,Y
Firewall at CWSU may be down

Subject: ST1 URGENT: CWSU PORT ttyS4 for TELEBOOT PHONE LINE
Date: Tue, 6 Jul 2004 20:24:49 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

CWSU TELEBOOT PHONE MODEM HAS BEEN REBOOTED but this may NOT UNLOCK the PORT
If this repeats several times, contact NWS technical staff
Kill all processes running on /dev/ttyS4 and check permissions for /dev/ttys4.
Permissions should be crw-rw-rw- and only root user can reset them.
You can also try a manual teleboot of the teleboot modem: 510 793-3623 /bo 4,Y
PORT ttyS4 and/or teleboot modem or line at CWSU may be down

Subject: ST1 URGENT: NO DATA RETRIEVAL FROM DCP COMPUTER AT SQL
Date: Tue, 6 Jul 2004 20:24:49 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

If this repeats several times, you may need to manually invoke the SQL DCP and modem teleboot script or dial
650 631-0623 /bo 1:2,Y
COMMUNICATIONS TO SQL DCP are down

Subject: ST1 URGENT: NO DATA RETRIEVAL FROM DCP COMPUTER AT SFO
Date: Tue, 6 Jul 2004 20:24:50 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

If this repeats several times, you may need to manually invoke the SFO DCP and modem teleboot script or dial
650 692-2497 /bo 1:2,Y
COMMUNICATIONS TO SFO DCP are down

Subject: ST1 URGENT: PIX FIREWALL AT SQL DOWN
Date: Tue, 6 Jul 2004 20:24:51 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

BUT SBC CAYMAN ROUTER (DSL) AT SQL IS OK
If this repeats several times, contact NWS technical staff
Or try manually telebooting 650 631-0623 /bo 4,Y
FIREWALL at SQL may be down

Subject: ST1 URGENT: PIX FIREWALL AT SFO DOWN
Date: Tue, 6 Jul 2004 20:24:52 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

BUT SBC CAYMAN ROUTER (DSL) AT SFO IS OK
If this repeats several times, contact NWS technical staff
Or try manually telebooting 650 692-2497 /bo 4,Y
FIREWALL at SFO may be down

Subject: ST1 URGENT: DSL ON DCP COMPUTER AT SQL DOWN
Date: Tue, 6 Jul 2004 20:24:52 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

BUT SBC CAYMAN ROUTER (DSL), PIX FIREWALL, AND PHONE DATA RETRIEVAL FROM SQL DCP ARE OK
If this repeats several times, you may need to manually invoke the SQL DCP teleboot script or dial
650 631-0623 /bo 2,Y
DSL on SQL DCP is down

Subject: ST1 URGENT: CWSU PORT ttyS5 for DATA PHONE LINE
Date: Tue, 6 Jul 2004 20:24:53 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

CWSU DATA PHONE MODEM HAS BEEN REBOOTED but this may NOT UNLOCK the PORT
If this repeats several times, contact NWS technical staff
Kill all processes running on /dev/ttyS5 and check permissions for /dev/ttys5.
Permissions should be crw-rw-rw- and only root user can reset them. You can also try a manual teleboot of the data modem: 510 793-3623 /bo 3,Y
PORT ttyS5 and/or data phone modem or line at CWSU may be down

Subject: ST1 URGENT: DSL ON DCP COMPUTER AT SFO DOWN
Date: Tue, 6 Jul 2004 20:24:53 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

BUT SBC CAYMAN ROUTER (DSL), PIX FIREWALL, AND PHONE DATA RETRIEVAL FROM SFO DCP ARE OK
If this repeats several times, you may need to manually invoke the SFO DCP teleboot script or dial
650 692-2497 /bo 2,Y
DSL on SFO DCP is down

Subject: ST1 URGENT: SBC DSL AT SQL DOWN
Date: Tue, 6 Jul 2004 20:24:59 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SBC CAYMAN ROUTER AT SQL NOT PINGABLE FROM CWSU BASE STATION (ST1)
BUT ST1 IS ONLINE: DSL MAY BE DOWN AT SQL
If this repeats several times, contact SBC technical staff
Try logging into 68.124.181.38 using a browser
If this does not work, try manually telebooting by dialing 650 631-0623 /bo 3,Y
DSL at SQL may be down

Subject: ST1 URGENT: SBC DSL AT SFO DOWN
Date: Tue, 6 Jul 2004 20:25:00 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SBC CAYMAN ROUTER AT SFO NOT PINGABLE FROM CWSU BASE STATION (ST1)
BUT ST1 IS ONLINE: DSL MAY BE DOWN AT SFO
If this repeats several times, contact SBC technical staff
Try logging into 67.126.122.22 using a browser
If this does not work, try manually telebooting by dialing 650 692-2497 /bo 3,Y
DSL at SFO may be down

Subject: ST1-URGENT-SFDCPOLD
Date: Mon, 19 Jul 2004 05:12:38 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_sfs.dat is old at: Mon Jul 19 05:12:38 2004
(Time Mon Jul 19 05:08:02 2004)
This indicates that there is a communication problem with the SFDCP computer.
Both DSL and phone communications are down to SFO.
There are automated scripts running that ping and teleboot the SBC and PIX routers at SFO.
These scripts generate emails with teleboot instructions.
If you receive this email repeatedly (every 15 minutes between 5 AM and 9 PM or every 30 minutes between 9 PM and 5 AM) you should manually teleboot according to the instructions in the emails with the teleboot instructions.
sfcount: 8

Subject: ST1-URGENT-SFO-FLUX-AND-SODAR-OLD
Date: Thu, 8 Jul 2004 19:33:52 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

radi_sfr.dat is old at: Thu Jul 8 19:33:51 2004
(Time Thu Jul 8 19:23:01 2004)
This indicates that neither pyranometer nor sodar data from the SFO site are available. Since no data are being received from SFO, there is probably NOT a problem with the datalogger and/or the datalogger communications link to the SFDCP computer although that can not be ruled out. As this is either accompanied by an email message called SFDCPOLD or SFDCP-HUNG, then the SFDCP computer will be automatically telebooted.
sfcount: 8

Subject: ST1-URGENT-SFDCP-HUNG
Date: Mon, 19 Jul 2004 05:12:40 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SFDCP has DSL and/or phone contact but no data at Mon Jul 19 05:12:40 2004. So it will be automatically telebooted now by calling 650 692-2497 /bo 2,Y
sfcount 8

Subject: ST1-URGENT-SCDCPOLD
Date: Mon, 19 Jul 2004 05:12:41 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_scs.dat is old at: Mon Jul 19 05:12:41 2004
(Time Mon Jul 19 05:08:04 2004)
This indicates that there is a communication problem with the SCDCP computer.
Both DSL and phone communications are down to SQL.
There are automated scripts running that ping and teleboot the SBC and PIX routers at SQL.
These scripts generate emails with teleboot instructions.
If you receive this email repeatedly (every 15 minutes between 5 AM and 9 PM or every 30 minutes between 9 PM and 5 AM) you should manually teleboot according to the instructions in the emails with the teleboot instructions.
sccount 8

Subject: ST1-URGENT-SQL-FLUX-AND-SODAR-OLD
Date: Thu, 8 Jul 2004 19:33:53 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

flux_scf.dat is old at: Thu Jul 8 19:33:53 2004
(Time Thu Jul 8 19:23:02 2004)
This indicates that neither pyranometer nor sodar data from the SQL site are available. Since no data are being received from SQL, there is probably NOT a problem with the datalogger and/or the datalogger communications link to the SCDCP computer although that can not be ruled out. As this is either accompanied by an email message called SCDCPOLD or SCDCP-HUNG, then the SCDCP computer will be automatically telebooted.
sccount: 8

Subject: ST1-URGENT-SCDCP-HUNG
Date: Mon, 19 Jul 2004 05:12:42 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SCDCP has DSL and/or phone contact but no data at Mon Jul 19 05:12:42 2004. So it will be automatically telebooted now by calling 650 631-0623 /bo 2,Y
sccount 8

Subject: ST1-URGENT-SFSOD-VERY-OLD
Date: Tue, 20 Jul 2004 15:31:12 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_sfs.dat is very old at: Tue Jul 20 15:31:11 2004
(Time Tue Jul 20 15:23:01 2004)
Automatic teleboot of SFDCP because sodar file has failed 11 successive times. If this persists, it may indicate a hardware problem with SFO sodar. Try manually telebooting if this persists: 650 692-2497 /bo 2,Y
sfcount: 11

Subject: ST1-URGENT-SCSOD-VERY-OLD
Date: Tue, 20 Jul 2004 15:12:11 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_scs.dat is very old at: Tue Jul 20 15:12:11 2004
(Time Tue Jul 20 15:08:03 2004)
Automatic teleboot of SCDCP because sodar file has failed 11 successive times. If this persists, it may indicate a hardware problem with SQL sodar. Try manually telebooting if this persists: 650 631-0623 /bo 2,Y
sccount: 11

Subject: ST1-URGENT-SCFLUX-VERY-OLD
Date: Tue, 20 Jul 2004 15:31:19 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

flux_scf.dat is very old at: Tue Jul 20 15:31:19 2004
(Time Tue Jul 20 15:23:03 2004)
Automatic teleboot of SCDCP because datalogger file has failed 11 successive times. If this persists, it may indicate a hardware problem with SQL datalogger. Try manually telebooting if this persists: 650 631-0623 /bo 2,Y
scfcount: 11

Subject: ST1-URGENT-SFFLUX-VERY-OLD
Date: Tue, 20 Jul 2004 15:12:14 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

radi_sfr.dat is very old at: Tue Jul 20 15:12:14 2004
(Time Tue Jul 20 15:08:01 2004)
Automatic teleboot of SFDCP because datalogger file has failed 11 successive times. If this persists, it may indicate a hardware problem with SFO datalogger. Try manually telebooting if this persists: 650 692-2497 /bo 2,Y
sfrcount: 11

Subject: ST1-URGENT-AWOSOLD
Date: Thu, 8 Jul 2004 19:33:57 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

awos_smo.dat is old at: Thu Jul 8 19:33:57 2004
(Time Thu Jul 8 19:31:01 2004)
This indicates a failure to get the awos_smo.dat file, which is generated at CWSU by the base (st1) computer. It may indicate that the serial port ttyS0 on st1 is locked or has improper permissions. If that is not the case, then there may be a problem with the AWOS sensor at SMB or in the communications link between the sensor and CWSU. If the CWSU forecasters are still getting SMB AWOS data on their display at CWSU, then a port problem on st1 is likely.

Subject: ST1-URGENT-MODEMLOCK5
Date: Thu, 8 Jul 2004 19:34:15 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Modem locked on ttyS5
MAY NEED TO REBOOT ST1 (Time Thu Jul 8 19:34:15 2004).
This indicates that the data modem port on st1 is locked.
Follow the instructions on the accompanying email.
If this persists and st1 is not accessible, you may have to teleboot st1: 510 793-3623 /bo 1,Y.

Subject: ST1-URGENT-MODEMLOCK4
Date: Thu, 8 Jul 2004 19:34:16 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Modem locked on ttyS4
This indicates that the teleboot modem port on st1 is locked
Follow the instructions on the accompanying email.
If this persists and st1 is not accessible, you may have to teleboot the teleboot modem on st1: 510 793-3623 /bo 4,Y.

Subject: ST1-SEMI-URGENT-SFFTP-FAIL
Date: Thu, 8 Jul 2004 19:33:50 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

ftp to sfdcp failed at: Thu Jul 8 19:33:50 2004
This indicates that DSL service to the SFO sodar site is down, however the phone call is succeeding so the sfdcp is running. You can check on the router by using a browser to access the SBC router at http://67.126.122.22/ There are automated scripts running that ping and teleboot the SBC and PIX routers at SFO as well as the SFDCP.
These scripts generate emails with teleboot instructions.
If you receive this email repeatedly (every 15 minutes between 5 AM and 9 PM or every 30 minutes between 9 PM and 5 AM) you should manually teleboot according to the instructions in the emails with the teleboot instructions.
sfftpcount: 8

Subject: ST1-SEMI-URGENT-SCFTP-FAIL
Date: Thu, 8 Jul 2004 19:33:51 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

ftp to scdcp failed at: Thu Jul 8 19:33:51 2004
This indicates that DSL service to the SQL sodar site is down, however the phone call is succeeding so the sfdcp is running. You can check on the router by using a browser to access the SBC router at http://68.124.181.38/ There are automated scripts running that ping and teleboot the SBC and PIX routers at SQL as well as the SCDCP.
These scripts generate emails with teleboot instructions.
If you receive this email repeatedly (every 15 minutes between 5 AM and 9 PM or every 30 minutes between 9 PM and 5 AM) you should manually teleboot according to the instructions in the emails with the teleboot instructions.
scftpcount: 8

Subject: ST1-SEMI-URGENT-ASOSOLD
Date: Thu, 8 Jul 2004 19:33:54 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

asos_sfo.dat is old at: Thu Jul 8 19:33:54 2004
(Time Thu Jul 8 19:23:02 2004)
This indicates that the sfo_asos data call has failed.
NWS is responsible for the modem that receives this call.
If it's not a hardware problem, it may indicate that
NWS has changed the asos password, which it does about every 9 months (see /ll/cv/src/scatter-ftp/kermit/asos.ker).
New password can be obtained from NWS Monterey from
Jeff Helms jeffrey.helms@noaa.gov 831 656-1713 ext. 262

Subject: ST1-SEMI-URGENT-SFFLUXOLD-ONLY
Date: Thu, 8 Jul 2004 19:33:55 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

radi_sfr.dat is old at: Thu Jul 8 19:33:55 2004
(Time Thu Jul 8 19:23:01 2004)
This indicates that the pyranometer data from the SFO site are not available. Since sodar data are still being received from SFO, this may indicate that there is a problem with the datalogger and/or the datalogger communications link to the SFDCP computer. As this is NOT accompanied by an email message called SFDCPOLD, then there is a problem with the datalogger at SFO. If it were accompanied by an email message called SFDCPOLD, then it would be a communication problem with the SFDCP computer
sfrcount: 8

Subject: ST1-SEMI-URGENT-SCFLUXOLD-ONLY
Date: Thu, 8 Jul 2004 19:33:55 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

flux_scf.dat is old at: Thu Jul 8 19:33:55 2004
(Time Thu Jul 8 19:23:02 2004)
This indicates that the pyranometer data from the SQL site are not available. Since sodar data are still being received from SQL, this may indicate that there is a problem with the datalogger and/or the datalogger communications link to the SCDCP computer. As this is NOT accompanied by an email message called SCDCPOLD, then there is a problem with the datalogger at SQL. If it were accompanied by an email message called SCDCPOLD, then it would be a communication problem with the SCDCP computer
scfcount: 8

Subject: ST1-SEMI-URGENT-AFOSOLD
Date: Thu, 8 Jul 2004 19:33:56 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

afos_sfo.dat is old at: Thu Jul 8 19:33:56 2004
(Time Thu Jul 8 19:31:01 2004)
This indicates that the afos_sfo.dat file is not being generated.
This can be due to a host of reasons but the most likely is that there has been some temporary format change in the SFOSWRCA file sent from NWS.
If this problem persists, you should contact Doug Sinton: sinton@met.sjsu.edu 650 326-9203

Subject: ST1-SEMI-URGENT-ASOSSIZE
Date: Thu, 8 Jul 2004 19:33:58 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

asos_sfo.dat is short (Size 18)
This indicates a short asos_sfo.dat file.
It's usually accompanied by the ASOSOLD email.
It usually indicates an NWS problem with the SFO ASOS modem or a temporary format change.
If this problem persists, contact Doug Sinton: sinton@met.sjsu.edu 650 326-9203

Subject: ST1 SEMI-URGENT: BAD 12Z SOUNDING STATUS
Date: Tue, 6 Jul 2004 20:23:33 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SOUNDING FROM ST1 Report:
12 Z SOUNDING BAD This is almost always an NWS problem with the sounding

Subject: ST1 LOW URGENCY: BAD 00Z SOUNDING STATUS
Date: Tue, 6 Jul 2004 20:23:33 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Subject: ST1-SEMI-URGENT-AWOSBIGE
Date: Thu, 8 Jul 2004 19:33:59 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

awos_smo.dat is too big (Size 105)
This indicates that the awos_smo.dat file is too big.
This is only a problem if it persists.
It indicates a change in the format of the raw awos file which is parsed on st1 by /ll/cv/src/scatter-ftp/awos/awosnew.
If it persists, contact Doug Sinton: sinton@met.sjsu.edu 650 326-9203

Subject: ST1-SEMI-URGENT-AWOSSIZE
Date: Thu, 8 Jul 2004 19:33:59 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

awos_smo.dat is short (Size 33)
This indicates that the awos_smo.dat file is short.
This is only a problem if it persists.
It indicates a change in the format of the raw awos file which is parsed on st1 by /ll/cv/src/scatter-ftp/awos/awosnew.
If it persists, contact Doug Sinton: sinton@met.sjsu.edu 650 326-9203

Subject: ST1-SEMI-URGENT-SCSIZE
Date: Thu, 8 Jul 2004 19:34:00 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_scs.dat is short (Size 43)
This indicates that the sodr_scs.dat file is short.
If this persists without the accompanying SCDCPOLD message, it can indicate a software problem on the scdcp or scsodar computer.
In this case contact Doug Sinton: sinton@met.sjsu.edu 650 326-9203
or Jim Jordan at ETL: jim.jordan@noaa.gov 303 497-6303

Subject: ST1-SEMI-URGENT-SFSSIZE
Date: Thu, 8 Jul 2004 19:34:00 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_sfs.dat is short (Size 88)
This indicates that the sodr_sfs.dat file is short.
If this persists without the accompanying SFDCPOLD message, it can indicate a software problem on the sfdcp or sfsodar computer.
In this case contact Doug Sinton: sinton@met.sjsu.edu 650 326-9203
or Jim Jordan at ETL: jim.jordan@noaa.gov 303 497-6303

Subject: ST1-SEMI-URGENT-SFRSIZE
Date: Thu, 8 Jul 2004 19:34:01 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

radi_sfr.dat is short (Size 27)
This indicates the radi_sfr.dat file is short.
If this persists without the accompanying SFFLUXOLD message, it can indicate a software problem on the sfdcp computer or a problem with the datalogger at SFO.
In this case contact NWS technical staff.

Subject: ST1-SEMI-URGENT-SCFSIZE
Date: Thu, 8 Jul 2004 19:34:02 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

flux_scf.dat is short (Size 17)
This indicates the flux_scf.dat file is short.
If this persists without the accompanying SCFLUXOLD message, it can indicate a software problem on the scdcp computer or a problem with the datalogger at SQL.
In this case contact NWS technical staff.

Subject: ST1-SEMI-URGENT-SCSDRIV
Date: Thu, 8 Jul 2004 19:34:04 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

driverbad found on sodr_scs.dat at: Thu Jul 8 19:34:04 2004 (file time Thu Jul 8 19:23:01 2004)
This indicates a potential failure of the sound driver on the SQL sodar.
It may also indicate a blown fuse on the sodar peripherals at SQL or a bad connection between the sodar peripherals and the SCDCP.
If it persists, contact NWS technical staff.

Subject: ST1-SEMI-URGENT-SFSDRIV
Date: Thu, 8 Jul 2004 19:34:04 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

driverbad found on sodr_sfs.dat at: Thu Jul 8 19:34:04 2004 (file time Thu Jul 8 19:23:01 2004)
This indicates a potential failure of the sound driver on the SFO sodar.
It may also indicate a blown fuse on the sodar peripherals at SFO or a bad connection between the sodar peripherals and the SFDCP.
If it persists, contact NWS technical staff.

Subject: ST1-SEMI-URGENT-SATOLD
Date: Sat, 24 Jul 2004 13:36:52 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Processed satellite file old at Sat Jul 24 13:36:52 2004
(/ll/cv/data/Xfer/vis_sum.dat file time Sat Jul 24 00:16:01 2004)
This indicates the processed satellite file derived from the raw satellite file sent by NWS is not current despite the presence of a current raw satellite file. If this email persists, contact LLMIT for a problem with the cvRTCache software.

Subject: ST1-SEMI-URGENT-NWS-SATOLD
Date: Sat, 24 Jul 2004 13:36:59 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>

AREA raw satellite file old at: Sat Jul 24 13:36:59 2004
(/ll/ftp/pub/cv/data/nws/AREA0702 time Sat Jul 24 13:31:53 2004)
(/ll/ftp/pub/cv/data/nws/AREA0702 size 1403136)
This indicates the raw satellite file from NRL is not current.
The raw satellite file from NRL /ll/ftp/pub/cv/data/nws/AREA0702 is old.
The problem is NOT in the cvRTCache software,
The raw satellite data are not coming in and if this persists contact:
Joe Turk: turk@nrlmry.navy.mil

Subject: ST1-SEMI-URGENT-AREASIZE
Date: Sat, 24 Jul 2004 13:37:00 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

/ll/ftp/pub/cv/data/nws/AREA0702 is short (Size 1403136)
This indicates a short /ll/ftp/pub/cv/data/nws/AREA0702 file.
It's accompanied by the SATOLD and/or the NWS-SATOLD email.
The size of this file should be 1.4 mbytes.
It usually indicates a problem with transmission of the satellite file to st1.
If this problem persists, contact: Chuck Ivaldi: Ivaldi@ll.mit.edu

Subject: ST1-SEMI-URGENT-LOWCEIL
Date: Thu, 8 Jul 2004 19:34:15 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

LOWCEIL written for awos_smo.dat at: Thu Jul 8 19:34:15 2004
This indicates a malfunction of the awos ceiling sensor at the San Mateo Bridge.
If this persists, contact CWSU at 510 745-3457 to report the problem.

Subject: ST1-LOW-URGENCY-LOCFCSTOLD
Date: Thu, 8 Jul 2004 19:34:11 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

local forecast old at: Thu Jul 8 19:34:11 2004
(file time Thu Oct 23 16:48:12 2003)
This indicates that the local forecast model did not run despite a ceiling existing.
This usually occurs because the Zi quality is low or because the Zi file(s) does (do) not exist yet or because some observation is not current. If it persists for days, contact MIT/LL.

Subject: ST1-LOW-URGENCY-COBFCSTOLD
Date: Thu, 8 Jul 2004 19:34:12 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Cobel forecast old at: Thu Jul 8 19:34:12 2004
(file time Thu Oct 23 16:14:48 2003)
This indicates that the COBEL forecast model did not run despite a ceiling existing.
This usually occurs because either the Zi quality is low or because the Zi file(s) does (do) not exist yet or because the 12Z Oakland sounding is late or missing. If it persists for days, contact MIT/LL.

Subject: ST1-LOW-URGENCY-REGFCSTOLD
Date: Thu, 8 Jul 2004 19:34:13 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

regional forecast old at: Thu Jul 8 19:34:13 2004
(file time Thu Oct 23 16:48:16 2003)
This indicates that the Regional forecast model did not run despite a ceiling existing.
This usually occurs because an observation is late or missing.
If it persists for days, contact MIT/LL.

Subject: ST1-LOW-URGENCY-CONFCSTOLD
Date: Thu, 8 Jul 2004 19:34:13 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

consensus forecast old at: Thu Jul 8 19:34:13 2004
(file time Thu Oct 23 16:36:23 2003)
This indicates that the Consensus forecast model did not run despite a ceiling existing.
If this persists for days, contact MIT/LL.

Subject: ST1-LOW-URGENCY-SATFCSTOLD
Date: Thu, 22 Jul 2004 17:40:49 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

satellite forecast old at: Thu Jul 22 17:40:49 2004
(file time Thu Jul 22 17:15:09 2004)
This indicates the Satellite forecast model did not run despite a ceiling existing and despite a current processed satellite image. If this persists, contact MIT/LL.

Subject: ST1-LOW-URGENCY-BADYEAR
Date: Thu, 8 Jul 2004 19:34:14 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

badyear written for nws.txt at: Thu Jul 8 19:34:14 2004
This indicates that a bad date is on the SFOSWRCA file received from NWS.
If this persists, contact Doug Sinton: sinton@met.sjsu.edu 650 326-9203.

Subject: ST1-LOW-URGENCY-SA2-LATE
Date: Thu, 8 Jul 2004 19:34:45 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sa2.out LATE data time: file time: (Time Thu Jul 8 19:34:43 2004) at: Thu Jul 8 19:34:45 2004
This indicates that the SFOWRKSA2 metar file is late.
It is sent from slug at NWSFO Monterey.
If this goes on for days, you should contact David Rosenberg at NWSFO.

Subject: ST1-LOW-URGENCY-SFOWRKSA2-LATE
Date: Thu, 8 Jul 2004 19:34:46 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SFOWRKSA2 LATE data time: 06 1900 file time: (Time Thu Jul 8 19:34:43 2004) at: Thu Jul 8 19:34:46 2004
If this goes on for days, contact David Rosenberg at NWSFO Monterey.

Subject: ST1-LOW-URGENCY-SFOWRKSA2-IRREGULAR
Date: Thu, 8 Jul 2004 19:34:47 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SFOWRKSA2 irregular file time: (Time Thu Jul 8 19:34:43 2004) at: Thu Jul 8 19:34:47 2004
If this goes on for days, contact David Rosenberg at NWSFO Monterey.

Subject: ST1-LOW-URGENCY-METARS
Date: Thu, 8 Jul 2004 19:34:47 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

metars.txt is short (Size 8) at: Thu Jul 8 19:34:47 2004
This means that the metars UNIDATA file pushed to st1 from methost8 at SJSU is late.
If it goes on for days, contact Mike Voss at SJSU: voss@met.sjsu.edu

Subject: ST1-NOT-URGENT-SFSODOLD
Date: Thu, 8 Jul 2004 19:33:52 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_sfs.dat is old at: Thu Jul 8 19:33:52 2004
(Time Thu Jul 8 19:23:01 2004)
This usually indicates that excessive aircraft noise at SFO has caused the sodar algorithm to drop at least 15 consecutive minutes of noise contaminated data resulting in no new sodar file.
This is a normal occurrence at SFO and is most likely to occur between 13Z and 04Z.
sfcount: 8

Subject: ST1-NOT-URGENT-SCSODOLD
Date: Thu, 8 Jul 2004 19:33:54 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

sodr_scs.dat is old at: Thu Jul 8 19:33:54 2004
(Time Thu Jul 8 19:23:01 2004)
This usually indicates that excessive aircraft noise at SQL has caused the sodar algorithm to drop at least 15 consecutive minutes of noise contaminated data resulting in no new sodar file. This is a sometime occurrence at SQL and is most likely to occur during daylight hours.
sccount: 8

Subject: ST1-NOT-URGENT-WINDSOLD
Date: Thu, 8 Jul 2004 19:33:57 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

winds.dat is old at: Thu Jul 8 19:33:57 2004
(Time Thu Jul 8 19:30:19 2004)
This indicates that the wocss wind data file is not being generated.
Normally the files winds.dat and temps.dat are sent to USGS once an hour for the SFBAY wind model.
A failure in generating the winds.dat file can indicate a problem in the afos_sfo.dat file.
Usually this is due to a temporary format change in the SFOSWRCA file sent by NWS.
If this persists, you should contact Doug Sinton: sinton@met.sjsu.edu or 650 326-9203

Subject: NOT-URGENT-WOXOLD
Date: Thu, 8 Jul 2004 19:33:56 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Wocss wind model at CWSU old: Thu Jul 8 19:33:56 2004 Time Thu Jul 8 19:33:16 2004.
This indicates the Wocss wind pc, ALFALFA, at CWSU has hung and may need to be rebooted by CWSU staff.
Call 510 745-3457. You can also try pinging ALFALFA by entering the wing command.

Subject: ST1-NOT-URGENT-WINDSIZE
Date: Thu, 8 Jul 2004 19:33:58 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

winds.dat is short (Size 36)
This indicates that the winds.dat file used by USGS is small.
It usually means that the afos_sfo.dat file does not have enough current data.
This problem usually resolves itself

Subject: ST1-NOT-URGENT-SCSZERO
Date: Thu, 8 Jul 2004 19:34:03 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

zeroline found on sodr_scs.dat at: Thu Jul 8 19:34:03 2004
This indicates that zeroes show up on the sodr_scs.dat file.
Do not worry about this.

Subject: ST1-NOT-URGENT-SFSZERO
Date: Thu, 8 Jul 2004 19:34:03 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

zeroline found on sodr_sfs.dat at: Thu Jul 8 19:34:03 2004
This indicates that zeroes show up on the sodr_sfs.dat file.
Do not worry about this

Subject: ST1-NOT-URGENT-WETARS
Date: Thu, 8 Jul 2004 19:34:48 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

wetars.txt is short (Size 0) at: Thu Jul 8 19:34:48 2004
This means that the metars file retrieved from the NWS website (see /ll/cv/src/scatter-ftp/getmetar.tcl) is late.
Not much you can do about this.

Subject: ST1 GOOD 00Z SOUNDING STATUS
Date: Tue, 6 Jul 2004 20:23:32 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SOUNDING FROM ST1 Report:
00 Z SOUNDING OK No worries

Subject: ST1 GOOD 12Z SOUNDING STATUS
Date: Tue, 6 Jul 2004 20:23:32 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

SOUNDING FROM ST1 Report:
12 Z SOUNDING OK No worries

SOUNDING FROM ST1 Report:
00 Z SOUNDING BAD This is almost always an NWS problem with the sounding

Subject: ST1 12Z SOUNDING
Date: Wed, 7 Jul 2004 21:15:07 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Wed Jul 7 21:15:07 2004
USUS46 KSTO 071215
MANOAK
72493 TTAA 57121 72493 99012 14827 30002 00109 14018 28004
92764 15866 23510 85500 23679 26508 70159 11268 19505 50587
07974 19005 40756 22370 23005 30960 37966 14503 25084 46564
17005 20228 56361 16501 15408 60960 26512 10659 62160 26512
88164 60960 23510 77999 51515 10164 00008 10194 25009 24506=
UMUS46 KSTO 071216
SGLOAK
72493 TTBB 57120 72493 00012 14827 11966 11607 22944 11608
33936 12061 44925 15866 55909 24478 66894 25680 77831 23479
88750 16876 99674 08066 11666 08278 22656 08677 33602 03878
44437 17564 55400 22370 66222 52562 77174 60161 88122 63360
99100 62160 31313 05102 81101 41414 863//=

PPBB 57120 72493 90012 30002 24007 23009 90346 25512 25511
27006 90789 25006 23506 22505 91246 19006 20511 23008 9205/
20509 22505 9305/ 15501 16506 947// 27014 9504/ 26516 27012=

Subject: ST1 00Z SOUNDING
Date: Wed, 7 Jul 2004 21:18:09 GMT
From: CCvV Account <cv@adsl-64-175-253-249.dsl.sntc01.pacbell.net>
To: mtr.eltechs@noaa.gov

Wed Jul 7 21:18:09 2004
USUS46 KSTO 070015
MANOAK
72493 TTAA 57001 72493 99012 14827 30002 00109 14018 28004
92764 15866 23510 85500 23679 26508 70159 11268 19505 50587
07974 19005 40756 22370 23005 30960 37966 14503 25084 46564
17005 20228 56361 16501 15408 60960 26512 10659 62160 26512
88164 60960 23510 77999 51515 10164 00008 10194 25009 24506=
UMUS46 KSTO 070016
SGLOAK
72493 TTBB 57000 72493 00012 14827 11966 11607 22944 11608
33936 12061 44925 15866 55909 24478 66894 25680 77831 23479
88750 16876 99674 08066 11666 08278 22656 08677 33602 03878
44437 17564 55400 22370 66222 52562 77174 60161 88122 63360
99100 62160 31313 05102 81101 41414 863//=

PPBB 57000 72493 90012 30002 24007 23009 90346 25512 25511
27006 90789 25006 23506 22505 91246 19006 20511 23008 9205/
20509 22505 9305/ 15501 16506 947// 27014 9504/ 26516 27012=

APPENDIX B
DATA CONTROL FILE FOR HOURLY METAR DATA

File Name: AFOS_DD.txt

File Location: /ll/cv/params

File Description: File used to direct the cvRTCache to date information within the input data file *afos_sfo.dat*. The valid date-time of the data is necessary in order to store the data in an on-disk dated directory.

File Contents:

```
SFOWRCA      6
SFOOSOSFO    5
NMCFRHT73    3
NMCFRH73     3
SFOCGRSFO    5
```

Contents Description:

One or more name-value pairs (one pair per line). Each pair consists of a header name that is routinely present in the input data file *afos_sfo.dat* and a count that represents the maximum number of records to search for date information following the detection of the header name. The cvRTCache reads this file each time the input data file *afos_sfo.dat* is processed from the transfer directory.

APPENDIX C
STATION LIST CONTROL FILE FOR HOURLY METAR DATA

File Name: afos_sa_list.txt

File Location: /ll/cv/params

File Description: File containing a list of 3-letter station identifiers. Determines which station reports contained in the input data file *afos_sfo.dat* are to be processed by the cvRTCache.

File Contents:

```
ACV
APC
CEC
CCR
FAT
HWD
LVK
MCC
MRY
NUQ
OAK
RBL
RHV
RNO
SAC
SFO
SJC
SMF
SNS
SQL
STS
SUU
UKI
END
```

Contents Description:

One 3-letter station identifier per line. Note that the list must be terminated with the marker “END”. This cvRTCache reads this file each time the input data file *afos_sfo.dat* is processed from the transfer directory.

APPENDIX D
SENSOR DATA LIST CONTROL FILE FOR CVRTCACHE PROCESSING

File Name: sensor_list.txt

File Location: /ll/cv/params

File Description: File containing a list of sensor data identifiers. Determines which incoming data files are to be processed by the cvRTCache.

File Contents:

```
sodr_sfs
sodr_scs
awos_smo
asos_sfo
afos_sfo
radi_sfr
flux_scf
vis_data
vis_sum
END
```

Contents Description:

One sensor data identifier per line. Note that the list must be terminated with the marker "END". The cvRTCache reads this file each time an input data file is processed from the transfer directory.

APPENDIX E
CVRTCACHE PARAMETERS FILE

File Name: cvcvRTCache.prm

File Location: /ll/cv/params

File Description: File containing runtime resource information needed by the cvRTCache.

File Contents:

```
XDIR                          "/ll/cv/data/Xfer"
HOURDIR                       "/ll/cv/data/cvcvRTCache/HOUR"
DISPDIR                       "/ll/cv/data/cvcvRTCache/LATEST"
LOG_HOURDIR                   "/ll/cv/data/cvcvRTCache/HOUR/LOG"
BINDIR                        "/ll/cv/cvcvRTCache/bin/"
SCRIPTDIR                     "/ll/cv/cvcvRTCache/scripts"
DAYSUM_L1DIR                  "/ll/cv/data"
DAYSUM_L2DIRPREFIX            "daysum"
AFOS_DAYSUM_DATATYPE          "sfc"
AFOS_DAYSUM_DATASOURCE        "afos"
AFOS_DAYSUM_FILESUFFIX        "sum"
AFOS_DATEDETECT_FILE          "/ll/cv/params/AFOS_DD.txt"
ASOS_DAYSUM_DATATYPE          "sfc"
ASOS_DAYSUM_DATASOURCE        "asos"
ASOS_DAYSUM_FILESUFFIX        "sum"
AWOS_DAYSUM_DATATYPE          "sfc"
AWOS_DAYSUM_DATASOURCE        "awos"
AWOS_DAYSUM_FILESUFFIX        "sum"
LLSFO_DAYSUM_DATATYPE_SFC     "sfc"
```

LLSFO_DAYSUM_DATATYPE_RAD	"rad"
LLSFO_DAYSUM_DATASOURCE	"ll"
LLSFO_DAYSUM_FILESUFFIX	"sum"
LLSQL_DAYSUM_DATATYPE_SFC	"sfc"
LLSQL_DAYSUM_DATATYPE_RAD	"rad"
LLSQL_DAYSUM_DATASOURCE	"ll"
LLSQL_DAYSUM_FILESUFFIX	"sum"
GOES_DAYSUM_DATATYPE	"sat"
GOES_DAYSUM_DATASOURCE	"vissec"
GOES_DAYSUM_STATIONID	"goes10"
GOESI_DAYSUM_FILESUFFIX	"dat"
GOESS_DAYSUM_FILESUFFIX	"sum"
SODAR_DAYSUM_DATATYPE	"sodar"
SODAR_DAYSUM_FILESUFFIX	"sum"

Contents Description:

One name-value pair per line. The cvRTCache reads this file only once at startup.

APPENDIX F
MANUALLY STOPPING AND STARTING THE CVRTCACHE

How to stop the cvRTCache:

1. Edit the crontab file (crontab –e) and comment out (with # in column 1) the line that reads:

    a. 10,20,30,40,50 * * * * . /ll/cv/.profile && /ll/cv/scripts/Rts/SonOfcvChecker.sh 1>>/tmp/SonOfcvChecker.log 2>&1

2. Run the following command at the unix prompt:

    a. /ll/cv/scripts/Rts/stop_cvcvRTCache_shell

How to start the cvRTCache:

1. Edit the crontab file (crontab –e) and uncomment out the line that was commented out in step 1 of "How to stop the cvRTCache".

2. Optionally run the following command at the unix prompt if it desirable to start the program sooner than if left to the cron job reactivated in step 1:

    a. /ll/cv/scripts/Rts/SonOfcvChecker.sh

APPENDIX G
CRON JOBS USED TO CONTROL DIRECTORY AND FILE TRIMMING

Scripts utilized within *cron* jobs for directory and file trimming:

/ll/cv/scripts/Misc/trimDirs: *csh* script that recursively trims dated directories. The script takes two arguments; the name of the root directory that contains the dated directories and the number of dated directories to keep.

/ll/cv/scripts/Misc/trimFiles.pl: *perl* script that trims files from a target directory. The two options used are:

1. –d <root_directory_name>
2. –a <count>

where <root_directory_name> is the name of the directory containing the files to trim and <count> is the number of files to keep.

***Cron* jobs:**

1. <u>Dated directories containing raw satellite data</u>

*0 4 * * * /bin/csh -c '/ll/cv/scripts/Misc/trimDirs /ll/cv/data/GOES/Raw 2 >>&! /dev/null'*

2. <u>Dated directories containing processed satellite data</u>

*5 4 * * * /bin/csh -c '/ll/cv/scripts/Misc/trimDirs /ll/cv/data/GOES/Processed 2 >>&! /dev/null'*

3. <u>Directory containing a copy of file pairs (satellite image and summary files) routinely sent to the cvRTCache for processing:</u>

*10 4 * * * /ll/cv/scripts/Misc/trimFiles.pl -d /ll/cv/data/GOES/Export/Moved -a 1 > /dev/null 2>&1*

4. Directory containing GOES satellite data processing log files:

*15 4 * * * /ll/cv/scripts/Misc/trimFiles.pl -d /ll/cv/data/GOES/Logs -a 10 > /dev/null 2>&1*

5. Dated directories containing log files for the cvRTCache and forecast models:

*20 4 * * * /ll/cv/scripts/Misc/trimDirs /ll/cv/logs 10 > /dev/null 2>&1*

APPENDIX H
SURFACE DATA CHARACTERISTICS AND DAY SUMMARY FILE FORMAT

```
SURFACE DATA
============

Filename convention:
   SAO        sfc_afos_xxx.yymmdd.sum
   ASOS       sfc_asos_sfo.yymmdd.sum
   AWOS       sfc_awos_smb.yymmdd.sum
   PG&E       sfc_baaq_xxx.yymmdd.sum
   MIT/LL     sfc_ll_xxx.yymmdd.sum


Description of included files
=============================
```

Filename	Site	Update Rate	On-Line Date	Reports Wx/Clds	Lat	Lon
sfc_ll_sfo.yymmdd.sum	SFO Airport -MIT/LL Sensor	5 min	Jul 19 1996	no	37.62	-122.38
sfc_ll_sql.yymmdd.sum	SQL Airport -MIT/LL Sensor	5 min	Jul 19 1996	no	37.75	-122.50
sfc_asos_sfo.yymmdd.sum	SFO Airport –ASOS	5 min	May 1 1996	yes	37.61	-122.38
sfc_awos_smb.yymmdd.sum	San Mateo AWOS	5 min	May 1 1996	yes	37.56	-122.27
sfc_afos_acv.yymmdd.sum	SAO ACV – Arcata	Hourly	May 1 1996	yes	40.99	-124.10
sfc_afos_apc.yymmdd.sum	SAO APC – Napa	Hourly	May 1 1996	yes	38.22	-122.28
sfc_afos_ccr.yymmdd.sum	SAO CCR – Concord	Hourly	May 1 1996	yes	37.98	-122.05
sfc_afos_fat.yymmdd.sum	SAO FAT – Fresno	Hourly	May 1 1996	yes	36.77	-119.72
sfc_afos_hwd.yymmdd.sum	SAO HWD – Hayward	Hourly	May 1 1996	yes	37.65	-122.11
sfc_afos_lvk.yymmdd.sum	SAO LVK – Livermore	Hourly	May 1 1996	yes	37.70	-121.82
sfc_afos_mry.yymmdd.sum	SAO MRY – Monterey	Hourly	May 1 1996	yes	36.58	-121.85
sfc_afos_nuq.yymmdd.sum	SAO NUQ – Moffet NAS	Hourly	May 1 1996	yes	37.41	-122.05
sfc_afos_oak.yymmdd.sum	SAO OAK – Oakland	Hourly	May 1 1996	yes	37.73	-122.20
sfc_afos_rbl.yymmdd.sum	SAO RBL - Red Bluff	Hourly	May 1 1996	yes	40.15	-122.25
sfc_afos_rno.yymmdd.sum	SAO RNO - Reno, NV	Hourly	May 1 1996	yes	39.50	-119.78
sfc_afos_sac.yymmdd.sum	SAO SAC - Sacramento Exec	Hourly	May 1 1996	yes	38.51	-121.50
sfc_afos_sfo.yymmdd.sum	SAO SFO - San Francisco	Hourly	May 1 1996	yes	37.61	-122.38
sfc_afos_sjc.yymmdd.sum	SAO SJC - San Jose	Hourly	May 1 1996	yes	37.36	-121.91
sfc_afos_sns.yymmdd.sum	SAO SNS - Salinas	Hourly	May 1 1996	yes	36.67	-121.60
sfc_afos_sql.yymmdd.sum	SAO SQL - San Carlos	Hourly	May 1 1996	yes	37.75	-122.50
sfc_afos_sts.yymmdd.sum	SAO STS - Santa Rosa	Hourly	May 1 1996	yes	38.52	-122.82
sfc_afos_suu.yymmdd.sum	SAO SUU – Fairfield	Hourly	May 1 1996	yes	38.27	-121.93
sfc_afos_uki.yymmdd.sum	SAO UKI – Ukiah	Hourly	May 1 1996	yes	39.13	-123.20

Matrix of reported data fields for each site (by site id)

Site	data field reported (yes/no) TEMP	DEWP	DIR	SPD	PRES	VIS	weather/clouds
ll_sfo	yes	yes	yes	yes	no	no	no
ll_sql	yes	yes	yes	yes	no	no	no
asos_sfo	yes	yes	yes	yes	yes	yes	yes
awos_smb	yes	yes	yes	yes	yes	yes	yes
SAO sites	yes	yes	yes	yes	yes	yes	yes

Description of included data

(data line is 122 characters long)

APPENDIX I

SURFACE DATA CHARACTERISTICS AND DAY SUMMARY FILE FORMAT (CONTINUED)

```
SAMPLE FILE HEADERS
-------------------

YY/MM/DD HH:MM HH.MM   TEMP  DEWP   DIR   SPD   PRES    VIS   WEATHER   CEIL   CLD1 T HGT1     CLD2 T HGT2     CLD3 T HGT3
96/07/31 00:45  0.75   36.0  14.0   310   5.1  1010.9  32.0             -99 -99.00 0  -99   -99.00 0  -99   -99.00 0  -99

YY/MM/DD HH:MM HH.MM   TEMP  DEWP   DIR   SPD   PRES    VIS   WEATHER   CEIL   CLD1 T HGT1     CLD2 T HGT2     CLD3 T HGT3
96/09/08 03:00  3.00   22.0  21.0   270   6.2  1010.9   3.2        BR   304    0.19 0  213     0.75 0  304     0.75 0  457


Description of each field
-------------------------

YY/MM/DD     year/month/day
HH:MM        hour:minute
HH.MM        decimal hour and minute
TEMP         temperature in degrees Celcius
DEWP         dewpoint temperature in degrees Celcius
DIR          wind direction in whole degrees
SPD          wind speed in m/s
PRES         barametric pressure in millibars
VIS          visibility in km
WEATHER      weather condition (NWS codes)
CEIL         ceiling height in meters, defined as lowest cloud of broken or overcast that is not
             a thin layer
CLD1 T HGT1 first cloud layer;  decimal cloud amount, thin indicator flag, height in meters
CLD2 T HGT2 second cloud layer; decimal cloud amount, thin indicator flag, height in meters
CLD3 T HGT3 third cloud layer;  decimal cloud amount, thin indicator flag, height in meters

Cloud amount is reported as a decimal value, created as follows:
NWS cloud report  Decimal equivalant
----------------  -----------------
   OVC            1.00
  -OVC            1.00
   BKN            0.75
```

-BKN	0.75
SCT	0.38
-SCT	0.38
FEW	0.19
-FEW	0.19
CLR	0.05
SKC	0.00

The thin indicator flag is currently not used.
If data for a specific time period is missing, it will simply not appear in the data file.
However, if data within the observation is missing (i.e. missing dewpoint or cloud cover, etc) the field will
be filled with a -99. The exception for this is the weather field, which if the station does not report weather this field is left blank.

APPENDIX J
STOPPING AND STARTING THE FORECAST MODELS

Stopping the Models:

Normally, a cron job will stop the models, at 1Z on October 15. To change the stop date, perform the following:

1. Login to st1 using the cv account

2. % crontab –e (edit the active crontab file)

3. Look for the entry: 0 1 15 10 * /bin/csh -c '/ll/cv/scripts/Rts/stopmodels.csh >>&! /dev/null'

where	0	1	15	10	*
	min.	hr.	day	month	day of week

4. Edit the line changing the date indicators as needed.

5. Exit out of crontab as if you were saving changes with the vi editor (COMMAND MODE w q).

6. The above procedure updates the running crontab. You must also make the same changes to the permanent copies of the crontab file. The files are:

/ll/cv/cv.cron and /ll/cv/cv.cron_summer_daylight.

Starting the Models:

Normally, a cron job will start the models, at 1Z on May 15. To change the start date, perform the following:

1. Login to st1 using the cv account

2. % crontab –e (edit the active crontab file)

3. Look for the entry: 0 1 15 5 * /bin/csh -c '/ll/cv/scripts/Rts/startmodels.csh >>&! /dev/null'

where	0	1	15	5	*
	min.	hr.	day	month	day of week

4. Edit the line changing the date indicators as needed.

5. Exit out of crontab as if you were saving changes with the vi editor (COMMAND MODE w q).

6. The above procedure updates the running crontab. You must also make the same changes to the permanent copies of the crontab file. The files are:

/ll/cv/cv.cron and /ll/cv/cv.cron_summer_daylight

APPENDIX K
NOAA/FSL SOUNDING FILE FORMAT

UPPER AIR DATA
==============

Filename convention:
   Oakland sounding   raob_oak.yymmdd.fsl (contains 0 and 12Z data)
   Oakland sounding   OAK_raobs_yymmdd_hh.fsl (old namimg convention, 1 sounding per file)

FSL Output Format Description

(FSL Radiosonde Database)

FSL Rawinsonde data format

The official FSL data format is similar to the format used by the National Severe Storms Forecast Center (NSSFC) in Kansas City. When a user of the database retrieves data using the direct access method discussed in Section 7, data appears in this ASCII "card image" format. The first 4 lines of the sounding are identification and information lines. All additional lines are data lines. An entry of 32767 indicates that the information is missing, not reported, or not applicable.

---COLUMN NUMBER---

1 LINTYP	2	3	4	5	6	7
			header lines			
254	HOUR	DAY	MONTH	YEAR	(blank)	(blank)
1	WBAN#	WMO#	LAT	LON	ELEV	RTIME
2	HYDRO	MXWD	TROPL	LINES	TINDEX	SOURCE
3	(blank)	STAID	(blank)	(blank)	SONDE	WSUNITS
			data lines			
9	PRESSURE	HGT	TEMP	DEWPT	WIND DIR	WIND SPD
4						
5						
6						
7						
8						

LEGEND

```
LINTYP: type of identification line
        254 = indicates a new sounding in the output file
          1 = station identification line
          2 = sounding checks line
          3 = station identifier and other indicators line
          4 = mandatory level
          5 = significant level
          6 = wind level (PPBB) (GTS or merged data)
          7 = tropopause level (GTS or merged data)
          8 = maximum wind level (GTS or merged data)
          9 = surface level

HOUR:       time of report in UTC
LAT:        latitude in degrees and hundredths
LON:        longitude in degrees and hundredths
ELEV:       elevation from station history in meters
RTIME:      is the actual release time of radiosonde from TTBB. Appears in GTS
            data only.
HYDRO:      the pressure of the level to where the sounding passes the
            hydrostatic check (see Section 4.3).**
MXWD:       the pressure of the level having the maximum wind in the sounding.
            If within the body of the sounding there is no "8" level then MXWN
            is estimated (see Section 3.2).
TROPL:      the pressure of the level containing the tropopause. If within the
            body of the sounding there is no "7" level, then TROPL is
            estimated (see Section 3.3)**
LINES:      number of levels in the sounding, including the 4
            identification lines.
TINDEX:     indicator for estimated tropopause. A "7" indicates that
            sufficient data was available to attempt the estimation; 11
            indicates that data terminated and that tropopause is a
            "suspected" tropopause.
SOURCE:     0 = National Climatic Data Center (NCDC)
            1 = Atmospheric Environment Service (AES), Canad
            2 = National Severe Storms Forecast Center (NSSFC)
            3 = GTS or FSL GTS data only
            4 = merge of NCDC and GTS data (sources 2,3 merged into sources
            0,1)
SONDE:      type of radiosonde code from TTBB. Only reported with GTS data
            10 = VIZ "A" type radiosonde
            11 = VIZ "B" type radiosonde
            12 = Space data corp.(SDC) radiosonde.
WSUNITS:    wind speed units (selected upon output)
            ms = tenths of meters per second
            kt = knots

PRESSURE:   in whole millibars (mb)
HGT:        height in meters (m)
```

```
TEMP:        temperature in tenths of degrees Celsius
DEWPT:       dew point temperature in tenths of a degree Celsius


WIND DIR:    wind direction in degrees
WIND SPD:    wind speed in either knots or tenths of a meter per second
             (selected by user upon output)
```

An example of FORTRAN format statements necessary to read output rawinsonde data, according to LINTYP, is as follows:

```
                    LINTYP
                       254    (3i7,6x,a4,i7)
                         1    (3i7,2f7.2,2i7)
                         2    (7i7)
                         3    (i7,10x,a4,14x,i7,5x,a2)
               4,5,6,7,8,9    (7i7)
```

**- section of NOAA tech memo on the data base (in print)

TECHNICAL INFORMATION: Schwartz, B.E., and M. Govett, 1992: "A hydrostatically consistent North American Radiosonde Data Base at the forecast Systems Laboratory, 1946-present." NOAA Technical Memorandum ERL FSL-4. Available from NOAA/ERL/FSL 325 Broadway, Boulder, CO 80303.

Begin RAOB database request

Prepared by Mark Govett, govett@fsl.noaa.gov, ; and Bill Moninger Moninger@fsl.noaa.gov

APPENDIX L
RADIATION DATA CHARACTERISTICS AND DAY SUMMARY FILE FORMAT

Radiation data available from shortwave radiometers (pyranometers).

RADIATION DATA
==============

Filename	Site	Update Rate	On-line Date	Lat	Lon
rad_sfo.YYMMDD.sum	SFO Airport -MIT/LL Sensor	5 min	Jul 19 1996	37.62	-122.38
rad_sql.YYMMDD.sum	SQL Airport -MIT/LL Sensor	5 min	Jul 19 1996	37.75	-122.50

SAMPLE FILE HEADER
==================

```
YY/MM/DD HH:MM HH.MM   Kd      Kn     Max      Ratio
96/07/31 00:05  0.08  553.0   441.2   541.4    1.02
```

```
Field       Value shown
-----       -----------
YY/MM/DD    year/month/day
HH:MM       hour:minute
HH.MM       decimal hour and minute
Kd:         downward shortwave insolation (watts/m^2)
Kn:         net shortwave insolation (watts/m^2)
Max:        maximum downward insolation (cloudless value) (watts/m^2)
Ratio:      Kd/Max : value represents the percent of total possible
      downward shortwave insolation
```

APPENDIX M
SODAR DATA CHARACTERISTICS AND DAY SUMMARY FILE FORMAT

```
SODAR DATA
============

Filename               Site                          Update         On-line        Lat     Lon
                                                     Rate           Date

sodar_sfo.YYMMDD.sum   SFO Airport -MIT/LL Sensor    5 min          Jul 19 1996    37.62   -122.38
sodar_sql.YYMMDD.sum   SQL Airport -MIT/LL Sensor    5 min          Jul 19 1996    37.75   -122.50


SAMPLE FILE
===========

-9999
V0.3 300C SITENAME02
8
6 07012002 000100
4 40 50 3 160 200 25
25348 11368 9794 6894 1840 1930 1200 1596 1045 967 807 816 692 264 258 130 192 396 87 410
141 95 70 93 27 98 117 78 91 211 204 132 226 304 263 289 298 469 218 472
497 393 234 675 466 219 140 46 14 7 0 0 0 15 5 5 0 10 17 0
0 3 0 0 36 26 11 14 33 0 2 18 10 40 16 0 3 26 9 0
0 4 44 17 32 34 2 30 42 49 50 40 35 39 15 20 21 40 35 21
2 84 48 24 30 39 26 42 29 19 10 33 25 14 14 3 0 10 9 8
33 21 22 6 8 23 4 15 10 8 0 15 20 12 23 10 5 16 10 19
32 12 9 3 17 26 52 11 13 17 14 7 24 6 15 15 18 32 33 15

Field                   Value shown
-----                   -----------
-9999                   Start of record marke
                        (Multiple records can appear in one file)
V0.3 300C SITENAME02    (Version number) (Model number) (Site id)
8                       No. of data sub-records contained in the record

----- Sub-record #1 -----

6 07012002 000100       (Version number) (MMDDYYYY) (HHMMSS)
4 40 50 3 160 200 25    Pulses Bwidth Delta Start Points Pwidth Temp
( Data )                160 data values ordered as 8 rows, with 20 data values
                        per row

----- Sub-record #2 -----
      .
      .
      .
      .
      .
----- Sub-record #8 -----
```

APPENDIX N
SATELLITE DATA CHARACTERISTICS AND DAY SUMMARY FILE FORMAT

```
Data filename conventions:
   sector statistics data   sat_vissec_goesX.YYMMDD.sum
   gridpoint VIS data       sat_visdat_goesX.YYMMDD.txt
   gif image file           sat_visdat_goesX.YYMMDD_HHMM.gif

Locator files:
   gridpoint latitudes      sat_lat.dat
   gridpoint longitudes     sat_lon.dat
   sector id image          sat_sect.gif
   grid point image         sat_grid.gif

Description of Included Files and Data:
=======================================

Each visible image is:
      1.  Normalized to eliminate sun-angle variations.
      2.  Interpolated to a 2km grid, 160 km on a side, centered on the San
          Mateo Bridge observing site (an 80 x 80 point grid).  The locator
          file "sat_grid.gif" shows this grid.
      3.  Passed over by a 9-pt median filter.
      4.  Divided into sectors (chosen based on topographical and operational
          considerations), and summarized accordingly.  These sectors cover an
          inner 80km by 80km area, centered on SMB.  These sectors can be seen
          in the locator files "sat_grid.gif" and "sat_sect.gif".

All base reflectance values are expressed as a fraction of maximum
reflectance; i.e., between 0 and 1.

Nominally, satellite image times are 15 minutes apart starting at the top of
each hour.  Frequent departures occur, however.  All image times are adjusted
to the nearest 5-minute multiple.

1) Sector statistics file (sat_vissec_goesX.YYMMDD.sum):
--------------------------------------------------------

Sample header and data:

YY/MM/DD HH:MM HH.MM  SEC    MEAN     SD   COV    N    MOM2
97/08/17 16:15 16.25  B1    0.672  0.016  1.00    8   3.615
97/08/17 16:15 16.25  B2    0.607  0.028  1.00    8   2.952
97/08/17 16:15 16.25  B3    0.501  0.088  0.75    8   2.062
```

Description of each field:

```
YY/MM/DD year/month/day
HH:MM    hour:minute
HH.MM    decimal hour
SEC      sector identifier
MEAN     mean value for sector
SD       standard deviation of sector values
COV      fraction of sector covered by cloud
N        number of grid points used
MOM2     second moment of sector values
```

Additional comments:

There is one file for each day. Each contains all the available summaries for each sector and time, in chronological order.

Sector identifiers (SEC) have a maximum of 3 characters. The location of each sector can be found in the sat_sect.gif locator image.
(This image also shows background elevations, in meters, at the 2-km resolution of the satellite data.) Thhe first character is a lette designating the general region in which the sector lies: (B)ay,
(W)est, (P)acific, (N)orth, (E)ast, or (S)outh. The sectors within each region are then numbered in a general sequence from west to east.

The fraction of a sector covered in cloud (COV) expresses the fraction of non-missing grid points above an empirically derived threshold.
This threshold was found via statistical comparisons of SFO and SMB surface cloud observations with the corresponding visible satellite pixel values. More specifically, the median of all satellite values corresponding to "broken" cloud surface observations was chosen. This yielded a reflectance value of 0.451 as the delineator between operationally significant cloud and no-cloud.

The number of grid points used (N) is the number of non-missing values in the sector for a particular time. If this number falls below 80% of the total possible points in the sector, then all other sector fields are given a value of -0.99, which designates "missing". These missing sectors are still reported, however, to preserve a constant number of lines for each data time.

The second moment (MOM2) is the sum of the squared values in the sector. It is included to enable computations of standard deviations for combinations of sectors.

2) Grid point VIS data (sat_visdat_goesX.YYMMDD.txt):

There is one file for each day. Each contains all the available images for each time, in chronological order. Each ASCII image contains the normalized and filtered reflectance values. Each image contains 80 lines, each containing 80 values. These images have a top-down orientation; i.e., the values are arranged in the file just as they would appear on a map. The first line of the image is the northern-most row of data, with the first value being the western-most value and the last value being eastern-most. The values are scaled by 100 to give integers between 0 and 100. Missing values are designated by -99. The corresponding latitude and longitude values for each of the grid points can be found in the sat_lat.dat and sat_lon.dat locator files.

3) Gif image files (sat_visdat_goesX.YYMMDD_HHMM.gif):

These images portray the values contained in the grid point data files described above. Overlaid are geographical outlines and the locations of FUN, SFO, SMB, SQL, and SJC.

Note on 1998 Data Quality:

The operational GOES-WEST satellite, GOES-9, began to fail in 1998, and was replaced by GOES-10 on 26 July. Various analyses have given reason to doubt the quantitative reliability of the GOES-9 values for the 1998 season, especially for purposes of developing statistical forecast models. Please use caution with these data.

APPENDIX O
INVERSION BASE HEIGHT PRODUCT CHARACTERISTICS AND DAY SUMMARY FILE FORMAT

```
Filename convention:
   OAK sounding inversion base product    inv_raob_oak.yymmdd.sum
   OAK sounding inversion base product    soundingLayer_yyyymmddhh
                                    (old naming convention)
   Hybrid inversion base product          inv_hyb_xxx.yymmdd.sum

                                          where xxx is either sfo or sql.

Description of included files
=============================

Filename                     Site                   Update period
--------                     ----                   -------------

inv_raob_oak.yymmdd.sum      Oakland                Twice Daily
inv_hyb_sfo.yymmdd.sum       SFO Arpt. -LL Sensor   15 min starting at 9,11 or 12Z
inv_hyb_sql.yymmdd.sum       SQL Arpt. -LL Sensor   15 min starting at 9,11 or 12Z


Description of included data: Oakland sounding inversion base product
=====================================================================

SAMPLE FILE HEADER
------------------

The Oakland sounding inversion product file does not have a header.

SAMPLE FILE DATA
----------------

2001082112
586 53

Description of each field
-------------------------

Line 1:
2001082112      year,month,day, and hour in yyyymmddhh format.

Line 2, field 1:
586             height of the inversion base in meters.

Line 2, field 2:
53              quality factor assigned to the derived height.
```

The quality factor is a measure of the confidence that a chosen inversion layer is likely to be the layer associated with the inversion base capping the marine stratus. The value is derived from a cumulative analysis of layer statistics including temperature gradient, dew point depression difference, the presence or absence of cloud below the layer, and temperature difference. Range is 0 - 100 in increasing confidence.

Description of included data: Hybrid inversion base product
===

(data line is 55 characters long)

SAMPLE FILE HEADER AND DATA

```
YY/MM/DD HH:MM HH.MM   inv_height(feet)  (meters)  QFactor
01/08/21 12:00 12.00   1929.13           588.00      73
```

Description of each field

```
YY/MM/DD                year/month/day
HH:MM                   hour:minute
HH.MM                   decimal hour and minute
inv_height(feet)        height of the inversion base in feet
(meters)                height of the inversion base in meters
QFactor                 quality factor assigned to the derived height
```

The quality factor is a measure of the confidence that a chosen sodar layer is likely to be the layer associated with the inversion base capping the marine stratus. The value is derived from a cumulative analysis of layer statistics including proximity to the previously derived height, layer persistence, matching strength between the two sodars, and whether the layer is above or below ceiling height when present. Range is 0 - 100, in increasing confidence.

APPENDIX P
CONSENSUS MODEL PARAMETERS FILE

File Name: consparams.py

File Location: /ll/cv/consensus/scripts

File Description: File containing runtime resource information needed by the consensus model.

File Contents:

```
#!/usr/local/bin/python

import string
import fileinput

# Choose method for computing consensus forecast and confidence
conf_method = 0  # +-delta t method
#conf_method = 1  # Probability method

# This parameter is only used when conf_method = 1
interval_prob = 0.5  # Hours

# Run parameters

# How often we try to compute a consensus forecast
# while we're in a run interval (in minutes)
retry_interval = 1

# List of all component models
models = ['LSFM', 'RSFM', 'SSFM', 'COBEL']

# Bad or missing value flags
bad_value = -999.00

# Input files
fcst_file = {'LSFM': '/ll/cv/data/daysum%s/fcsts/%s%s%s/fcsts_Local.%s%s%s.sum',
             'RSFM': '/ll/cv/data/daysum%s/fcsts/%s%s%s/fcsts_Regional.%s%s%s.sum',
             'SSFM': '/ll/cv/data/daysum%s/fcsts/%s%s%s/fcsts_Satellite.%s%s%s.sum',
             'COBEL': '/ll/cv/data/daysum%s/fcsts/%s%s%s/fcsts_Cobel.%s%s%s.sum'
             }

# Output files
display_out_file = '/ll/cv/data/cvcvRTCache/LATEST/latest_consensus.dat'
summ_out_file = '/ll/cv/data/daysum%s/fcsts/%s%s%s/fcsts_Consensus.%s%s%s.sum'
# Archive directory
```

```
archive_dir = '/ll/cv/data/cvcvRTCache/LATEST'

# File to log diagnostic output
logfile = '/ll/cv/logs/%s/ConsensusModel/%s.log.%s'

# Lock file
lockfile = '/ll/cv/Locks/ForecastUpdate.lock'

# Run schedule. This determines how the consensus forecast is computed # for a
set of given hours.  Each entry is a dictionary with the
# following elements:
#
# forecast_hour = hour from which to grab data for the forecast
# start_hours   = start hours of the interval during the current
#                 runtime hour that we will attempt to compute the
#                 consensus forecast
# start_minutes = start minutes
# end_hours     = end hours
# end_minutes   = end minutes

sched = [
         {'forecast_hour': '0900',
        'models': ['LSFM', 'RSFM', 'COBEL'],
        'start_hours': 9,
        'start_minutes': 18,
        'end_hours': 9,
        'end_minutes': 35,
        'data_start_hours': 8,
        'data_start_minutes': 45,
        'data_end_hours': 9,
        'data_end_minutes': 5
        },

         {'forecast_hour': '0900',
        'models': ['LSFM', 'RSFM', 'COBEL'],
        'start_hours': 9,
        'start_minutes': 37,
        'end_hours': 10,
        'end_minutes': 0,
        'data_start_hours': 8,
        'data_start_minutes': 45,
        'data_end_hours': 9,
        'data_end_minutes': 5
        },

         {'forecast_hour': '1100',
        'models': ['LSFM', 'RSFM', 'COBEL'],
        'start_hours': 10,
        'start_minutes': 2,
        'end_hours': 11,
```

```
'end_minutes': 35,
'data_start_hours': 10,
'data_start_minutes': 45,
'data_end_hours': 11,
'data_end_minutes': 5
},

 {'forecast_hour': '1100',
'models': ['LSFM', 'RSFM', 'COBEL'],
'start_hours': 11,
'start_minutes': 37,
'end_hours': 12,
'end_minutes': 0,
'data_start_hours': 10,
'data_start_minutes': 45,
'data_end_hours': 11,
'data_end_minutes': 5
},

 {'forecast_hour': '1300',
'models': ['LSFM', 'RSFM', 'COBEL'],
'start_hours': 12,
'start_minutes': 2,
'end_hours': 13,
'end_minutes': 35,
'data_start_hours': 12,
'data_start_minutes': 45,
'data_end_hours': 13,
'data_end_minutes': 5
},

 {'forecast_hour': '1300',
'models': ['LSFM', 'RSFM', 'COBEL'],
'start_hours': 13,
'start_minutes': 37,
'end_hours': 14,
'end_minutes': 0,
'data_start_hours': 12,
'data_start_minutes': 45,
'data_end_hours': 13,
'data_end_minutes': 5
},

{'forecast_hour': '1500',
'models': ['LSFM', 'RSFM', 'SSFM', 'COBEL'],
'start_hours': 14,
'start_minutes': 2,
'end_hours': 15,
'end_minutes': 35,
'data_start_hours': 14,
```

```
 'data_start_minutes': 45,
 'data_end_hours': 15,
 'data_end_minutes': 5
 },

{'forecast_hour': '1500',
 'models': ['LSFM', 'RSFM', 'SSFM', 'COBEL'],
 'start_hours': 15,
 'start_minutes': 37,
 'end_hours': 16,
 'end_minutes': 0,
 'data_start_hours': 14,
 'data_start_minutes': 45,
 'data_end_hours': 15,
 'data_end_minutes': 5
 },

{'forecast_hour': '1600',
 'models': ['LSFM', 'RSFM', 'SSFM', 'COBEL'],
 'start_hours': 16,
 'start_minutes': 2,
 'end_hours': 16,
 'end_minutes': 35,
 'data_start_hours': 15,
 'data_start_minutes': 45,
 'data_end_hours': 16,
 'data_end_minutes': 5
 },

{'forecast_hour': '1600',
 'models': ['LSFM', 'RSFM', 'SSFM', 'COBEL'],
 'start_hours': 16,
 'start_minutes': 37,
 'end_hours': 17,
 'end_minutes': 0,
 'data_start_hours': 15,
 'data_start_minutes': 45,
 'data_end_hours': 16,
 'data_end_minutes': 5
 },

{'forecast_hour': '1700',
 'models': ['LSFM', 'RSFM', 'SSFM', 'COBEL'],
 'start_hours': 17,
 'start_minutes': 2,
 'end_hours': 17,
 'end_minutes': 35,
 'data_start_hours': 16,
 'data_start_minutes': 45,
 'data_end_hours': 17,
```

```
         'data_end_minutes': 5
         },

        {'forecast_hour': '1700',
         'models': ['LSFM', 'RSFM', 'SSFM', 'COBEL'],
         'start_hours': 17,
         'start_minutes': 37,
         'end_hours': 18,
         'end_minutes': 0,
         'data_start_hours': 16,
         'data_start_minutes': 45,
         'data_end_hours': 17,
         'data_end_minutes': 5
         },

        {'forecast_hour': '1800',
         'models': ['LSFM', 'RSFM', 'SSFM', 'COBEL'],
         'start_hours': 18,
         'start_minutes': 2,
         'end_hours': 18,
         'end_minutes':35,
         'data_start_hours': 17,
         'data_start_minutes': 45,
         'data_end_hours': 18,
         'data_end_minutes': 5
         }
        ]

# Performance score file
perf_score_file = '/ll/cv/consensus/ascii/perf_scores.txt'

# Read performance scores into a dictionary
def read_perf_scores(filename):
    perfsc = {}
    for line in fileinput.input(filename):
      if fileinput.filelineno() > 1:
          parts = string.split(line)
          assert len(parts) == 3, ('read_perf_scores: line %d of scores
                                    table only contains %d fields.'
                                    %(fileinput.filelineno(),
                                    len(parts)))
          model_id = parts[0]
          sigma = float(parts[2])
          perfsc[model_id] = sigma

    assert len(perfsc) > 1, 'read_perf_scores: scores table contains
                             only one line.'
    return perfsc
# Read in performance scores
perf_scores = read_perf_scores(perf_score_file)
```

APPENDIX Q
CONSENSUS MODEL WEIGHTS FILE

File Name: perf_scores.txt

File Location: /ll/cv/consensus/ascii

File Description: File containing component model weights needed by the consensus model.

File Contents:

```
MODEL                   NCASES      SIGMA
Cobel09                 179         1.764
Cobel11                 269         1.664
Cobel13                 381         1.597
Cobel14                 372         1.470
Cobel15                 347         1.268
Cobel16                 288         1.157
Cobel17                 182         1.248
Cobel18                 93          1.360
Local09Default          161         1.491
Local11Default          237         1.430
Local13ZiQLT55          44          2.195
Local13ZiQGE55          303         1.208
Local14ZiQLT75          52          1.410
Local14ZiQGE75          296         1.193
Local15ZiQLT90          66          1.423
Local15ZiQGE90          239         1.056
Local16ZiQGE90          231         0.884
Local16ZiQLT90          68          1.070
Local17Default          226         1.005
Local18Default          122         1.201
Regional09Default       928         1.699
Regional09Neutral       588         1.544
Regional09Onshore       58          1.458
Regional09Offshore      203         1.541
Regional11Default       952         1.687
Regional11Neutral       583         1.622
Regional11Onshore       57          1.882
Regional11Offshore      194         1.418
Regional13Default       951         1.650
Regional13Neutral       659         1.590
```

Regional13Onshore	83	1.613
Regional13Offshore	165	1.494
Regional14Default	944	1.515
Regional14Neutral	642	1.449
Regional14Onshore	90	1.555
Regional14Offshore	173	1.412
Regional15Default	918	1.486
Regional15Neutral	664	1.497
Regional15Onshore	99	1.461
Regional15Offshore	131	1.193
Regional16Default	883	1.452
Regional16Neutral	640	1.487
Regional16Onshore	79	1.817
Regional16Offshore	105	0.809
Regional17Default	697	1.343
Regional17Neutral	531	1.384
Regional17Onshore	73	1.292
Regional17Offshore	70	1.107
Regional18Default	445	1.299
Regional18Neutral	334	1.402
Regional18Onshore	54	1.615
Regional18Offshore	24	0.494
Satellite15Default	323	1.168
Satellite15LoCigSmthCld	217	1.057
Satellite15LoCigRufCld	63	1.178
Satellite15HiCig	34	1.923
Satellite16Default	287	1.031
Satellite16LoCigSmthCld	206	0.931
Satellite16LoCigRufCld	41	1.728
Satellite16HiCig	33	1.143
Satellite17Default	228	0.986
Satellite17LoCigSmthCld	134	0.995
Satellite17LoCigRufCld	42	1.208
Satellite17HiCig	44	0.914
Satellite18Default	134	1.182

Contents Description:

One model-runtime-daytype per line.

GLOSSARY

ARTCC	Air Route Traffic Control Center
AWOS	Autmoated Weather Observing System
CWSU	Central Weather Service Unit
DCP	Data Collection Platform
ITWS	Integrated Terminal Weather System
LAN	Local Area Network
NRL	Naval Research Laboratory
NWSFO	National Weather Service Forecast Office
PPP	Point to Point Protocol
SFO	San Francisco International Airport
SODAR	Sonic Detection and Range
SQL	San Carlos Airport

REFERENCES

Chernal, E., A. Bellon, B.L. Austin, and A. Kilambi, "An Objective Technique for the Delineation and Extrapolation of Thunderstorms from GOES Satellite Data," *J. of Geoph. Res.*, **90**, No. D4, 6203–6210, 1985.

Clark, D.A. and F.W. Wilson, "The Marine Stratus Initiative at San Francisco International Airport," Massachusetts Institute of Technology, Lincoln Laboratory, Project Report ATC-252, Lexington, MA, 1996.

Rao, C.R.N., "Implementation of the Post-Launch Vicarisou Calibration of the GOES Imager Visible Channel," www.oso.noaa.gov/goes/goes-calibration, 2003.

Rew, R.K. and G.P. Davis, "NetCDF: An Interface for Scientific Data Access, Computer Graphics and Applications," *IEEE*, pp. 76–82, 1990.

SFO Marine Stratus Forecast System Documentation

Section IV. Display System

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (CONTINUED)

LIST OF TABLES

1. INTRODUCTION

1.1 SCOPE

This document serves as both an operational user manual and a display maintenance & data processing manual for the SFO Marine Stratus Forecast (MSF) System. The system is designed to provide forecast guidance for the dissipation of aviation-impacting stratus from the approach zone into San Francisco International Airport. Display products are generated from data collected on a base station computer located at the Center Weather Service Unit (CWSU) at the Oakland Air Route Traffic Control Center (ARTCC) in Fremont, CA, and delivered to remote users via the World Wide Web. In addition to CWSU forecasters, other users include air traffic managers and commercial airline weather forecasters and dispatchers.

Three sections follow this Introduction. Section 2 is primary intended as an operational user manual. It provides a description of the products and details regarding interaction with the display. Sections 3 and 4 are intended more for maintenance of the display system. Section 3 provides a description of the data handling processes required for serving the display. Section 4 describes the processes required to generate the various display products. Many of the key display maintenance elements are documented and summarized in the appendices.

1.2 BACKGROUND

1.2.1 Project History

1.2.1.1 Stratus Phenomenology and Operational Impact

The local airspace surrounding the San Francisco International Airport (SFO) is prone to regular occurrences of low ceiling conditions from May through October due to the intrusion of marine stratus cloudiness along the Pacific coast. The low cloud conditions prohibit dual parallel approaches (DPA) of aircraft to the airport's closely spaced parallel runways, thus effectively reducing the arrival capacity by 50%. The behavior of marine stratus evolves on a daily cycle, filling the San Francisco Bay region overnight, and dissipating during the morning. Often the low ceiling conditions persist throughout the morning hours and interfere with the high rate of air traffic scheduled into SFO from mid-morning to early afternoon. The result is a substantial number of delayed flights into the airport. Air traffic managers face a continual challenge of anticipating available operating capacity so that the demand of incoming planes can be metered to match the availability of arrival slots. When the approach clearing occurs earlier than anticipated, available arrival capacity goes unused as upstream planes are not properly positioned to exploit the newly available arrival slots. The number of unutilized arrival slots can range from three to as many as twenty per stratus event, resulting in a substantial amount of unnecessary aircraft delay. Conversely, if decisions are made to release upstream aircraft in anticipation of additional capacity that is unrealized, the result is an unacceptably high number of airborne holds and possible diversions.

The Central Weather Service Unit (CWSU) at the Oakland Air Route Traffic Control Center (ARTCC), the National Weather Service (NWS) Office in Monterey, and meteorologists at the national System

Command and Control Center in Washington, D.C. are responsible for providing weather information to assist air traffic managers anticipate changes in available operating capacity. Meteorologists at private airlines are also responsible for providing weather information for the most efficient and cost effective dispatch of airplanes. These are considered the primary users of the stratus forecast guidance system.

1.2.1.2 Historical Overview

In 1995, the FAA's Aviation Weather Research Program (AWRP) launched a "Marine Stratus Initiative" whose objective was to improve cloud forecast guidance for more reliable anticipation of available operating capacity at SFO. MIT Lincoln Laboratory served as technical lead for the project, in collaboration with San Jose State University, the University of Quebec at Montreal, and the CWSU at Oakland Center. The project led to the development of the MSF System. This system includes observations of the current weather and forecast guidance predicting the burn off of marine stratus clouds that form in the San Francisco Bay. Early in the project it was determined that a Display and Analysis (D&A) system was necessary to serve up new products to the primary user, the Central Weather Service Unit (CWSU) forecaster at the Oakland Center. The first system used a Server-Client Library architecture designed by MIT Lincoln Laboratory's Weather Sensing Group. This system has been used by various data services and display systems within the group. The design used a single base station computer for product generation and to serve a D&A computer at the CWSU. There was a need to supply this display data to multiple users without the need of a dedicated D&A computer. It was determined that the best way to accomplish this was by creating a publicly available internet based display, the MSF Display. The software underwent an extensive redesign during the winter of 1995 in order to create a file-based architecture. This legacy software has formed the backbone of the current cache design.

1.2.2 System Overview

The MSF Display System is a web based program that accesses a variety of data from the base station computer and sends that data over the internet. The data are accessed using a file-based cache structure. The data are read from the multiple directory trees and served to the MSF Display. These data include generated GIF images, and text files that include the model forecasts.

The model forecast guidance products were developed using data from the summer season when the stratus dissipates in a predictable manner. Since the model guidance is only tuned for summer stratus prediction, the MSF Display will run in one of two modes. In summer mode the forecast table will be shown and displays the latest forecast guidance when the models are operating. In winter mode the forecast table is replaced with a text table that will include information about the MSF operations or sensor maintenance that might take place. While in winter mode none of the guidance products will be displayed. Appendix H contains details about how to set the MSF Display into winter mode.

1.2.3 MSF System Products

The MSF System products that appear on the display include two main categories. These two main categories are observation products and automated forecast guidance products. The Observation Products described below are of four types: satellite imagery, sodar images, solar radiation images, and hourly surface data. The Forecast Products contain four component models and one Consensus model that weigh the component models to provide a single forecast. The four component models are COBEL, Regional, Local and Satellite forecast models.

1.2.3.1 Observation Products

The Observation Products are listed below and is discussed in Section 2.4.
- Satellite Imagery: 1 km resolution data depicting the San Francisco Bay
- sodar images: six hour, 1 minute interval vertical scan of the atmosphere
- Solar Radiation: ~12 hours of measured incoming solar radiation
- Surface Observations: measure of the current conditions at the station

1.2.3.2 Forecast Products

The Forecast Products are listed below and is discussed in Section 2.5.
- COBEL: one-dimensional physical model
- LSFM: Local Statistical Forecast Model
- RSFM: Regional Statistical Forecast Model
- SSFM: Satellite Statistical Forecast Model
- Consensus: Weighted combination of the four component models

1.2.4 MSF Real-Time Data Processing

Since the MSF Display System is a web based program that delivers data to multiple systems simultaneously, the data must be processed and stored in real time so that the latest data are always available. The observation data processes run continuously and produce both graphical images and text based hourly surface observations. The images are created and stored in graphics interchange format (GIF). The GIF images that are created include solar radiation data, sodar data, and satellite data. Other text data include CWSU year-to-date rate change times, model performance scores, and system data status. Additional data are downloaded from the Internet FTP and processed to produce ETA analysis maps.

1.2.4.1 Observation Data

The observation data listed below requires processing and is discussed in Section 3.2.
- Sodar
- Solar Radiation
- Satellite
- Surface Observations

1.2.4.2 Forecast Data

An overview of the model forecast guidance products is discussed in Section 3.3.

1.2.4.3 Text Data

The text data that requires processing is listed below and is discussed in Section 3.4.

Rate Change
Model Performance
System Status
Apache Web Access Logs

1.2.4.4 FTP Data

These data are downloaded by a cron and processed to produce analysis maps. This is discussed in Section 4.5.

1.2.5 MSF Display Processing and Performance

The MSF Display is a dynamic web page that uses JavaScript and Common Gateway Interface (CGI) scripts written in Perl. Every piece of the display uses Perl scripts to set up the layout and determines the data to be shown on the display. The maintenance required of the MSF Display System is small compared to the real-time data that are created. Occasionally it may be necessary to investigate why data are not being displayed or why the data are not outputting correctly. It may also be necessary to confirm that changes in internet browser technology have not significantly affected the functionality of the display.

1.2.5.1 MSF Display

An introduction to the three processing scripts that produce the three frames is discussed in Section 5.1.
Title Frame
Observation Frame
Forecast Frame

1.2.5.2 Observation Processes

The clickable links in the observation frame are considered observation processes and each script associated with the link is discussed in Section 4.2.
Surface Observations Display Processes
Sodar Display Processes
Satellite Display Processes
Solar Radiation Display Processes

1.2.5.3 Forecast Processes

The many clickable links found in the forecast frame are considered forecast processes and each script associated with the link is discussed in Section 4.3.
Forecast Summary
Raw Model Output Summary
Key Model Parameters
CWSU Rate Change Form
Model Performance Summary
ETA Analysis Maps
Review Prior Days Display

Rate Change Times Summary
System Status

1.3 MSF DISPLAY MAINTENANCE

In general, the MSF display is designed to run continuously throughout the year with no software or file maintenance. However, there are some maintenance issues that may need to be addressed over the longer term, i.e. over a period of one or more years. These associated tasks fall into two general categories: 1) the need to remove files (observation files, log files, etc.) from the system disk that continuously grow in size, and 2) the need to edit files or scripts in response to external changes that may influence data formats or processing requirements. Within the various functional sub-sections of this document, these possible long term software and file maintenance tasks are identified and described. Furthermore, in order to provide a quick reference to support system maintenance, each of these maintenance item descriptions are also presented sequentially in Appendix A, including their corresponding paragraph references from the main body of this document.

1.4 MSF DISPLAY CONFIGURATION

The SFO MSF Display relies on weather observations from a network of sensors surrounding the Bay region. Data are collected from these sensors at 15-minute intervals, and transferred to a database computer located at the ARTCC, also referred to as Oakland Center, in Fremont, CA. Once there the data are processed for display, and for input into a suite of algorithms designed to forecast the time of Approach Zone (AZ) clearing. The display of observations and the automated forecast guidance are made available to users via the World Wide Web. This distribution medium allows all users to access the display without the need of a separate dedicated system display. The display is password protected to prevent unauthorized access to the system although a selected group of users are allowed access without the password. Any user not included in the selected users group must use a user name and password to access the display system.

1.5 RELATED DOCUMENTS

For complete details of real-time data acquisition processing, see Section III: Real-time System. For a more in-depth explanation of the forecast models and their prediction parameters, see Section V.

2. MSF DISPLAY OPERATIONAL OVERVIEW

2.1 INTRODUCTION

The MSF display is the primary interface for operational users to access observations and model forecasts as an aid toward anticipating the time of stratus dissipation in the Approach Zone into SFO. The primary display is located at the Oakland Center CWSU in Fremont, CA, which hosts the system base station computer for data acquisition and local product generation. Remote users are able to access the display via the World Wide Web. The display is both automated and interactive. It automatically refreshes with the most recently available information, and it allows the user to investigate the details of each element as desired. In particular, it includes an on-line archive of all observations, forecasts, and verifications for a minimum of one year.

The display includes both a data observation frame and a forecast frame. The observation frame provides a graphical suite of weather observations as an aide in proper diagnosis of the current stratus-related conditions. Most importantly, these observations include daily time plots of the sodar-derived inversion base heights and incoming solar radiation measurements. The forecast frame provides both deterministic and probabilistic forecasts of the time of stratus dissipation, based on four individual forecast models. These four “component” forecasts are combined to generate a single “Consensus” forecast. Forecasts are generated every two hours every morning from 09 GMT to 15 GMT, and then hourly until 18 GMT. The forecast frame also allows the user to review and analyze various elements of the forecasts, such as the specific parameters within each model. Additionally, the forecast frame allows the user to launch a separate browser that allows access to view all observations and forecasts for any individual day within the past year.

More details regarding the various observation and forecast elements within the display, as well as display interaction, are provided in the remainder of this operational overview.

2.1.1 Mouse Pointing Device

2.1.1.1 Click the Mouse

In the document when the phrase “click on xxx” is found it means to move the mouse cursor to the specified location (xxx) and then push and release the left mouse button.

2.1.1.2 Mouse Over Event

There are references to a “mouse over” event in this document. This means that when the mouse cursor is moved to the specified location that an event, such as opening a pop-up window, will take place.

2.2 GENERAL DISPLAY INFORMATION

2.2.1 Introduction

The SFO MSF Display is a web based display that is divided into three frames. These three frames each contain different data types.

2.2.2 Display Sizing

The SFO MSF Display was designed to be viewed on a screen size no less than 1152 pixels x 864 pixels. A smaller screen size causes scroll bars to appear in both the left and right frames in order to view all the data contained on the display. It is possible to eliminate the scroll bars but only if the graphics card and the display can handle the higher resolution size required.

2.2.3 Frame Sizing

The three frames on the SFO MSF Display appear at a default size when the display is initially opened. The frames can be resized by moving the mouse cursor to the frame border edge, pushing and holding the left mouse button, and dragging the frame border to a new location. Each time the frame border is resized the frames will automatically refresh.

2.2.4 Refresh Rate and Response Time

While the display is running the browser will automatically update every five minutes for the left frame and every minute for the right frame. The amount of time for the display to respond during the update cycle or after clicking a link depends on both the speed of the user's internet connection and the traffic load on the MSF System's internet connection. Every time a link is clicked that opens new browser window the parent frame from which the link was selected automatically refreshes.

2.2.5 Display Not Updating

If it appears that the display is not updating and the data has become old or is unavailable then the user should try to manually refresh the browser by clicking on the browser's reload or refresh button. If the display still does not refresh with newer data then click the System Status (Section 2.5.11) button in the lower right corner of the display to see if the data is still current on the MSF System. If the data appears to be current in the System Status window then there could be a problem with the MSF Display System.

2.3 DISPLAY LAYOUT

An example of the display layout is shown in Figure 1.

2.3.1 Title Frame

The upper left frame contains the page title and the current GMT time to the right of the title. The time is determined by the computer system clock.

2.3.2 Observation Frame

2.3.2.1 Description

The lower left frame on the display is used to display current observations that include surface weather observations (Section 2.4.1), Sodar Inversion Base graphs (Section 2.4.2), visible satellite imagery (Section 2.4.3), and solar radiation plots (Section 2.4.4). The time resolution of the data observations ranges from 1 to 15 minutes, depending upon data type. Each individual product receives new data roughly every 15 minutes. The display automatically checks every 2 minutes to see if any of the products has received new data, and automatically refreshes the entire observation frame. Due to the time required to collect and process remote data observations, the latency of the most recent data viewed on the display will typically range from 10 to 25 minutes. When the most recent data appears more than 30 minutes old, it is usually indicative of some problem with the remote sensor, data communications, or possibly the web browser. In these instances, the viewer may click the System Status (Section 2.5.11) button in the Forecast Frame (Section 2.3.3) for more information. If the data appears to be current then there may be a problem with the web browser. In this case review the information in Section 2.2.5.

2.3.2.2 User Interaction

The underlined titles located above each of the Observation Products (Section 2.4) are links to an online help document that describes the observation data. When these titles are clicked, a new browser window opens the online help at the selected section. There are other links located on this page that either link to more detailed data or full page data. The satellite image is a clickable link that brings up a more extensive image animation controller.

2.3.3 Forecast Frame

2.3.3.1 Description

The right frame displays the latest model forecasts and links to more detailed forecast data as well as archived data from prior days. The page is automatically updated every minute in order to verify that the latest forecasts are always displayed. The model forecasts are only updated once per hour typically 20-30 minutes past the top of the hour. Directly above the forecast table is where the Approach Clear Message (Section 2.3.3.1.1) appears. Below the forecast table is a drop down selector that has links to a list of detailed model data and other links to open up other archived data. The Status Button (Section 2.5.11) at the bottom of the page will open up a new window indicating latency of the data.

2.3.3.1.1 Approach Clear Message

Directly above the forecast table, space is reserved to display an “Approach Clear” (AC) message to indicate that the AZ is sufficiently free of clouds to allow dual approaches (DPA). The automated forecast algorithm stops running and the model forecasts are no longer displayed when the message appears. The AC message can be generated either automatically based on surface ceilometer observations or satellite data or manually via input from the on-duty meteorologist at the CWSU.

In the pre-dawn hours prior to the availability of visible satellite imagery, the surface ceilometer reports are examined for the presence of a cloud deck in the proximity of the AZ. If none of the surface stations are reporting a cloud deck, the AC message will appear on the display. Following sunrise, the visible satellite imagery is examined for cloudiness in the AZ. If there is no visible cloud and the surface stations are indicating no cloud deck, then the AC message appears.

Once the AZ is formally declared by air traffic managers to be sufficiently clear for DPA, the on-duty CWSU meteorologist may enter the official clearing time into the system, which will appear within the AC message. This acts as the verification time used to access performance of the model forecasts.

2.3.3.2 User Interaction

Clicking on any one of the links below the forecast table will open a new browser window. The model names and the underlined titles in the forecast table are all clickable links that will open up the help pages and will give an overview of the selected model.

The confidence indicator in the forecast table has a mouse over event. When the user points the mouse cursor over the confidence level or the two letter confidence level code a pop-up window appears that explains why the model ran with a low confidence.

2.4 OBSERVATION PRODUCTS

2.4.1 Surface Observations

2.4.1.1 Description

The most recent observations from the surface observing stations at San Francisco International Airport (SFO) and San Mateo Bridge (SMB) are shown in the upper left corner of the Observation Frame. Surface observations are also reported from a high resolution sensor suite deployed at San Carlos (SQLL). The SMB observation is particularly significant in that it lies directly beneath the flight path of the AZ into SFO. The SQLL observation does not have an instrument to detect ceiling height or visibility so these fields are marked with an “N/A.” These observations are updated at 15-minute intervals; the observations may appear 10 to 25 minutes old due to latencies in data collection, transmission, and processing. If the observations become more than forty minutes old then all the observation fields become “N/A.” An example of the surface observation table is shown in Figure 2.

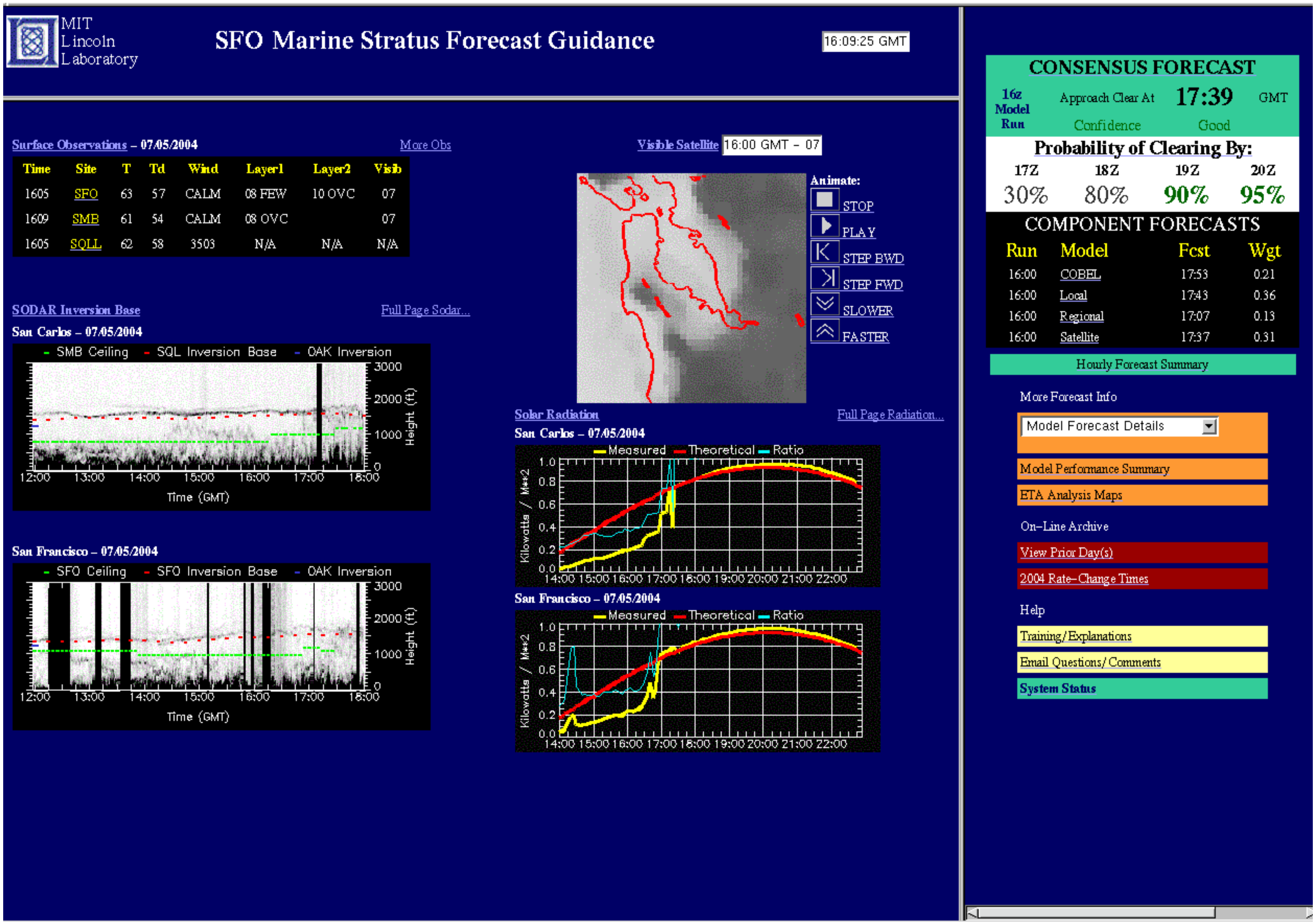


Figure 1. SFO MSF display layout.

The columns in the data table from right to left are:

1. Time of observation (GMT)
2. 3-character site identifier
3. Temperature, degrees F
4. Dew Point Temperature, degrees F
5. Wind Direction and Speed (ddff), where dd is direction in tens of degrees, and ff is wind speed in knots
6. Cloud Layer 1 Height (hundreds of feet) and Coverage
7. Cloud Layer 2 Height (hundreds of feet) and Coverage
8. Visibility (miles)

Time	Site	T	Td	Wind	Layer1	Layer2	Visib
1605	SFO	63	57	CALM	08 FEW	10 OVC	07
1609	SMB	61	54	CALM	08 OVC		07
1605	SQLL	62	58	3503	N/A	N/A	N/A

Figure 2. Surface observation table.

2.4.1.2 User Interaction

The surface observation table includes data for three stations: San Francisco Airport (SFO), San Mateo Bridge (SMB), and a special sensor at San Carlos Airport (SQLL). The three letter station identifier is a clickable link that opens up a new window that shows a full days worth of observations for that station.

Additional regional surface observations are available by clicking the "More Obs" link that appears above the observation table in the Observation Frame. A map of regional weather observing locations is shown. Standard hourly National Weather Service stations are shown in yellow, and more frequently updated (5-15 minutes) stations are shown in white. To the right of the map is a table showing the most recent observations for each location. Clicking one of the 3-character station identifiers, either on the map or in the table, will provide a table of recent observations for that station. To return to the Observation Frame click the underlined link at the top or bottom of the page or click the back button on the browser.

2.4.2 Sodar

2.4.2.1 Description

The Sodar Inversion Base product is intended to provide a continuous monitoring of the local inversion base height which defines the top of the stratus deck. The acoustic sodar (SOnic Detection And Ranging) is an upwardly pointing parabolic antenna that emits an audible signal whose return signal is proportional to the vertical gradient of air density (and therefore temperature) in the atmosphere. On the sodar plot, the temperature change representing the inversion appears as a dark quasi-horizontal line above the 500-foot level of the time-height plot. Darkness in the plot below 500 feet is usually attributed to ambient noise. The red dashed line on the sodar plot is an automated estimate of the inversion base height. This estimate is also used as input to automated forecast algorithms. The green dashed line is the ceiling height from the nearest surface reporting station to the sodar. The blue line displayed at 1200 GMT is the Oakland inversion determined from the sounding data. For a dark line of significant thickness, the inversion base height is presumed to be at the *bottom* of the dark layer, where the temperature change associated with the inversion begins. During the cloud burn-off process, a slowly rising inversion base height may be indicative of a later than anticipated clearing time. In addition, it has been observed that the appearance of the inversion signal on the plot tends to become more diffuse during the hour before clearing due to vertical mixing of the atmosphere. A black vertical line or bar on the sodar image is caused by significant noise at the ground level or a missing sodar scan at that time period. The noise is most likely an aircraft passing close to the sensor. An example of a sodar image is shown in Figure 3.

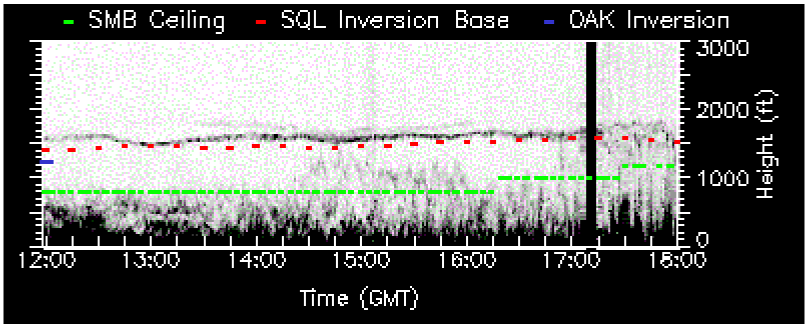


Figure 3. Example of sodar image.

2.4.2.2 User Interaction

By clicking the "Full Page Sodar..." link the two sodar images will be enlarged to fill the entire Observation Frame. This may distort the images slightly depending on the browser.

In order to return to the Observation Frame click the underlined link at the top or bottom of the page or click the back button on the browser.

2.4.3 Satellite Imagery

2.4.3.1 Description

The satellite display product is derived from the 1 km resolution western US GOES visible satellite channel. It is intended to provide a general close-up view of the cloud coverage and brightness in the San Francisco Bay Region, centered on the AZ into SFO. An updated satellite image is available every 15 minutes. Unlike conventional displays of visible satellite imagery, the satellite data used for the display are "normalized" to eliminate the effects that the changing sun angle has on the brightness of the cloudiness. The result is that any changes viewed in the cloud brightness are primarily attributable to the actual physical characteristics of the cloud. For a stratus cloud deck ranging in thickness from 200 to 1200 feet (typical of the SFO Bay region), the normalized cloud albedo is roughly linearly proportional to the total liquid water content within the cloud, so that changes in the visible brightness can be directly attributable to changes in cloud liquid water. In addition to brightness normalization for sun angle, the satellite data also undergo a geographic registration process to minimize “jumpiness” in viewing a sequence of cloudiness, and a simple interpolation and filtering process to eliminate irregular high frequency changes in cloud coverage and brightness that may make it difficult to analyze trends in the cloud coverage during the burn-off process. Since this is a visible satellite image it is only available after sunrise. An example of the visible satellite image is shown in Figure 4.

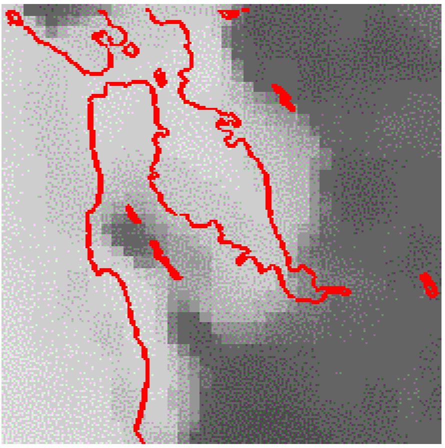


Figure 4. Visible satellite image.

2.4.3.2 User Interaction

Directly to the right of the satellite image there are six buttons. These buttons control the animation of the satellite image in the Observation Frame. From top to bottom, the control buttons allow the standard function of Stop/Pause, Play, Step Backward, Step Forward, Increase Loop Speed and Decrease Loop Speed. Directly above the satellite image there is a frame containing the time at which the satellite image was recorded.

2.4.3.2.1 Image Animation Controller

If the satellite image is clicked in the Observation Frame then a new window opens with a larger satellite image and more extensive movie looping controls. The satellite animation tool is divided into a left and right frame. The right frame contains a slightly larger satellite image above and the satellite data time below. The left frame has a row of buttons on top that control the satellite loop. From left to right the buttons are: Step Backward, Play Backward, Stop, Play Forward and Step Forward. The next row of buttons controls the playback loop mode. From left to right the buttons are: Play Normal, Play Forward/Backward. The next two rows of buttons controls the loop speed and the amount of dwell time either before or after the image loop. The frame number of the satellite image is included in a box below the buttons. If any of the satellite images are corrupt or appear to have poor registration these can be excluded from the loop by clicking the image frame number at the bottom of the left frame. The window can be closed by clicking the "Close" button below the animation tool. An example of the image animation controller is shown in Figure 5.

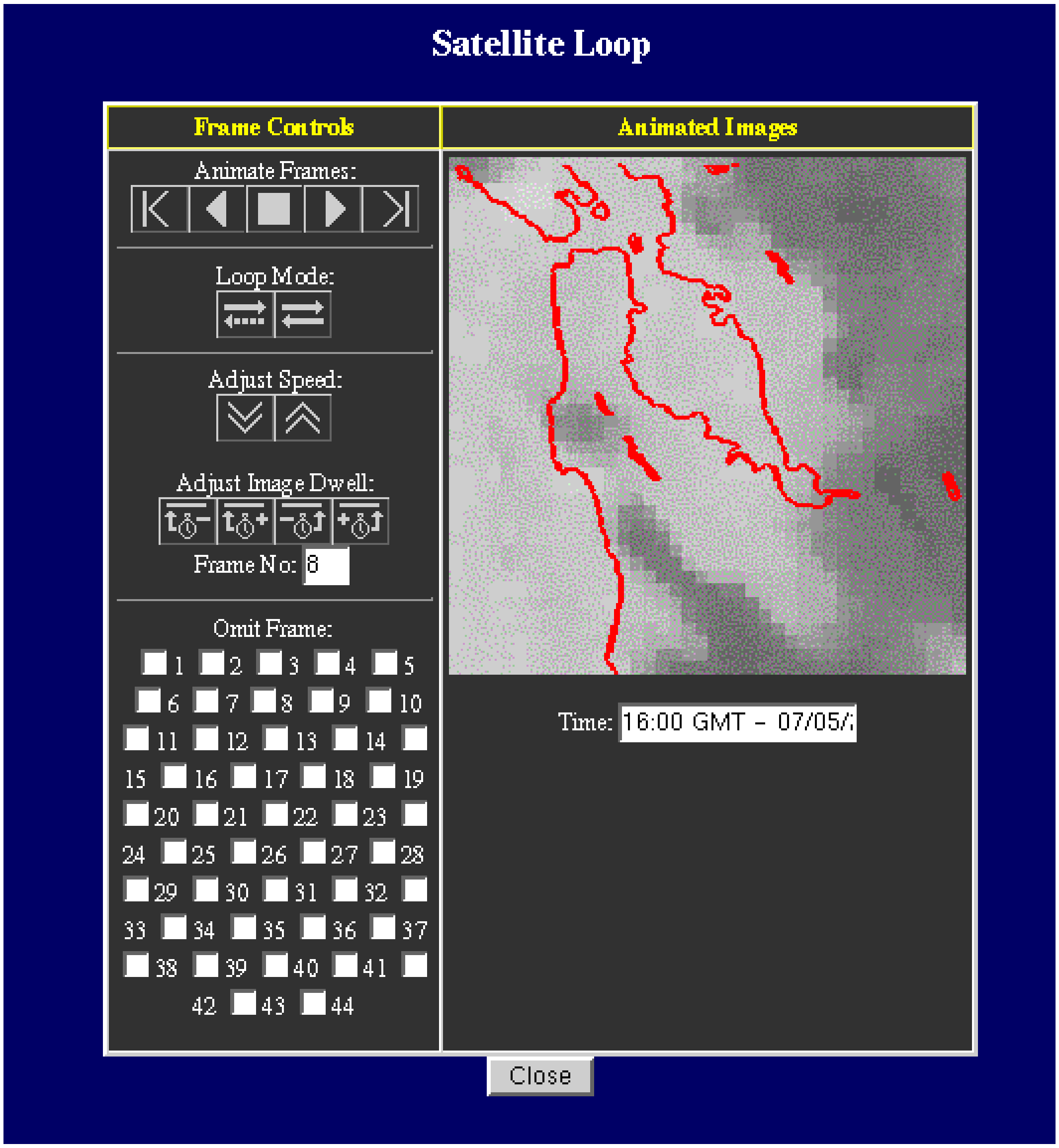


Figure 5. Image animation controller.

2.4.4 Solar Radiation

2.4.4.1 Description

The Solar Radiation product is an image showing the amount of shortwave radiation reaching the earth's surface. This value is plotted as a yellow curve. The red parabolic curve indicates the amount of radiation that would reach the ground on a cloud-free day. This varies by location, day-of-year, and time-of-day. The thin blue curve on the plot shows the ratio between the observed amount of radiation transmitted though the stratus layer and the theoretical amount that would reach the ground on a cloud-free day. This ratio is significant in that the fraction of transmitted solar radiation is roughly proportional to the liquid water content (integrated vertically) within the cloud layer. This image is usually created after 14 GMT or when the sun begins to rise. If data has not been collected by the MSF System by 15 GMT, the solar radiation data is assumed to be unavailable and the display will reflect that on the screen. An example of the solar radiation image is shown in Figure 6.

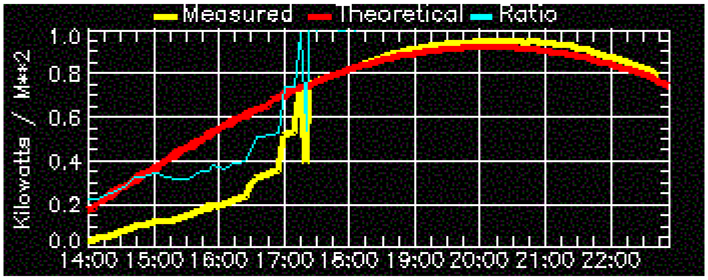


Figure 6. Solar radiation image.

2.4.4.2 User Interaction

By clicking the "Full Page Radiation..." link the two solar radiation images will be enlarged to fill the entire Observation Frame. This may distort the images slightly depending on the browser.

In order to return to the Observation Frame click the underlined link at the top or bottom of the page or click the back button on the browser.

2.5 FORECAST PRODUCTS

2.5.1 Forecast Guidance

2.5.1.1 Description

The right side of the display features automated forecast guidance for predicting the time of stratus clearing in the AZ into SFO. (The forecasts are not designed to handle the behavior of clouds associated with extra-tropical cyclones.) There are four "Component" forecasts models, the COBEL model, the Local SFM, the Regional SFM, and the Satellite SFM, designed to predict the clearing time (also referred to as the cloud “burn-off”), and one “Consensus” forecast that weighs the component forecast to provide a single forecast. The forecast model strengths and weaknesses are shown on Table 1.

TABLE 1

Model Strengths & Weaknesses

MODEL	Strengths	Weaknesses
COBEL	• Only physical model • Adjusts to feedback • Direct estimate of cloud liquid	• Remote point forecast • Does not handle non-local forcing
Local	• High resolution measurements • Emulates existing techniques	• Single point forecast • Limited historical database
Regional	• Multiple data point locations • Handles larger scale forcing	• High forecast variance • Vulnerable to incomplete data
Satellite	• Direct observation of predictor • Area input	• Sensitive to calibration errors • Limited to visible imagery

The forecast algorithms were developed using a history of the times at which the airport transitioned from a 30 plane per hour Acceptance Rate to a higher rate (either 45 or 60 planes per hour). This is often referred to as the “side-by” time, as it represents the time at which arriving aircraft may begin performing DPA into the airport. This transition time was used to verify the time of “Approach Clearing” during the automated algorithm development process. In a meteorological sense, it is most closely associated to the first time that the automated observing station at the San Mateo Bridge (SMB) transitions from a cloud ceilometer report overcast (OVC) or broken (BKN), to a report of scattered (SCT) or clear (CLR). However, since there are other meteorological conditions which could preclude the performance of DPA such as lingering clouds elsewhere in the AZ, or the presence of haze, the Approach Clearing time is not precisely equivalent to the SMB clearing time. In addition, the start time of DPA may also be affected by Air Traffic Management decisions which include other non-weather-related factors. Therefore, the prediction times provided by the algorithms are intended as the best estimate of the time that DPA will be allowed for aircraft on approach to SFO.

The COBEL, Local, and Regional models generate forecasts each day at 09, 11, and 13 GMT. The Satellite Model requires visible satellite imagery which is not available until after sunrise. Starting at 15 GMT, all four of the component models generate a forecast every hour through 18 GMT, or until the stratus has cleared from the AZ.

In addition to providing a forecast time, a forecast "weight" is also presented for each of the four component models. This weight indicates the relative contribution of each model forecast in generating a Consensus Forecast. The weights are derived from the relative historical performance of the models at each forecast initialization time, and additionally taking into account the relative performance during pre-defined categorical meteorological conditions including onshore versus offshore flow or high versus low ceiling base heights.

2.5.1.1.1 Consensus

The Consensus Forecast provides a single unified forecast of the time that stratus is expected to clear the AZ to allow DPA of aircraft. It is derived from the four component forecasts.

The Consensus Forecast is computed by assigning a weight to each of the component forecasts. From historical data, performance of the four component models has been evaluated separately for each of their respective day types, and for each model initialization time. In generating the Consensus Forecast, each component model's contribution is weighted accordingly based on the current model run time and the current day types.

The Consensus Forecast time of clearing is accompanied by a confidence indicator. There are particular model day types (based on meteorological conditions) for which the model performance is expected to be substandard. Under these conditions, the forecast confidence is indicated as "LOW;" otherwise it is indicated as "Good."

2.5.1.1.2 Probability of Clearing

These probabilities indicate the likelihood that the Approach Zone will be clear by specific times that are relevant to the air traffic decision process, namely 17, 18, 19, and 20 GMT. These times represent the period during which the heaviest arrival traffic is due into SFO, and traffic managers must decide to implement, maintain, or cancel ground delay programs based on anticipated available capacity.

The probability numbers are derived from a historical analysis of the correlation of Consensus Forecast clearing time and actual clearing time. Earlier Consensus Forecast times correspond to higher likelihood of clearing, and later times correspond to lower percentages. Lower percentages are displayed in black; when the probability of clearing by a particular time is sufficiently high (90% or greater) to confidently recommend proactive release of ground-held aircraft, the percentage is displayed in green.

2.5.1.1.3 COBEL

COBEL is a very high-resolution 1-dimensional (column) numerical model of the planetary boundary layer that simulates the evolution and burn-off of marine stratus. It is unique to the MSF System in that it is the only component forecast derived from a physics-based model. It was adapted from a model developed for forecasting fog behavior in northern France. (COBEL is short for the French **Co**uche **B**rouillard **E**au **L**iquide.)

The model is initialized with a vertical profile of temperature, humidity, and wind at SFO. The initialization uses a hybrid of the Oakland (OAK) balloon sounding and high resolution lower atmosphere measurements at SFO. This profile is used to characterize the stratus cloud deck in terms of total vertically integrated liquid water content. As the model steps forward, physical processes are applied within the model to anticipate the evolving decrease of the cloud liquid water with time. The model forecast time of cloud dissipation is declared when the liquid water content reaches zero.

Since the model is adapted for a single point location, the forecast for that location (SFO) is then adjusted to account for the spatial difference from the area of interest, i.e. the AZ into SFO. This is done via historical analysis of the relationship between cloud coverage at the model point location and the recorded transition to DPA (the so-called *side-by* time.) Thus, the COBEL forecast time shown on the display represents the *expected time of clearing in the AZ.*

The model is re-initialized for each model run, using updated high resolution measurements of surface temperature, humidity, wind, and solar radiation. In particular, the re-initialization includes an adjustment to ensure that the model's interpretation of liquid cloud water amount is consistent with the amount of solar radiation transmitted through the cloud layer. The model is then re-run, providing a new forecast of stratus burn-off.

The operational version of COBEL for San Francisco is designed for typical stratus days when stratus is present and fully developed at or before sunrise, when there is a marked strong inversion at the top of a shallow well-mixed marine boundary layer, and when no clouds are present above the boundary layer. The model includes (via measurement or parameterization) all of the physical processes important for the dissipation of marine stratus including solar and terrestrial radiation, turbulence, as well as cloud and rain processes. Although designed to input horizontal advection of temperature and humidity and vertical motion from meso-scale models, the current operational version assumes no horizontal advection and a fixed value for subsidence of (0.7 cm/s). A horizontal pressure force, corresponding to a 2 m/s geostrophic wind, is arbitrarily applied over the entire depth of the model domain (surface to 1.5 km height) and kept constant in time.

2.5.1.1.4 Local SFM

The LSFM uses statistical regression to forecast the time of stratus burn-off in the AZ. The model is considered "local" in that its predictions are derived from weather observations that are very local to the AZ, particularly the airports at SFO and SQL. Important predictors from these two locations are the height of the inversion base, cloud layer heights, the surface wind, and trends (in space or time) in these parameters. Regression analysis was performed to correlate the behavior of these parameters to the historical times of DPA. Since solar radiation is the primary forcing mechanism during the burn-off process, model statistics were developed in terms of the elapsed time after sunrise. The model is run each day at 09, 11, and 13 GMT, and then hourly from 15 to 18 GMT.

2.5.1.1.5 Regional SFM

The RSFM uses statistical regression to forecast the time of stratus burn-off in the AZ. It is designed to use routinely available surface and upper air observations as input. As such, its advantage over the other two statistical models is its broader view of synoptic conditions, particularly its ability to discriminate onshore versus offshore geostrophic flow. As with the other two statistical models the forecast is produced at 09, 11, and 13 GMT, and then hourly from 15 to 18 GMT.

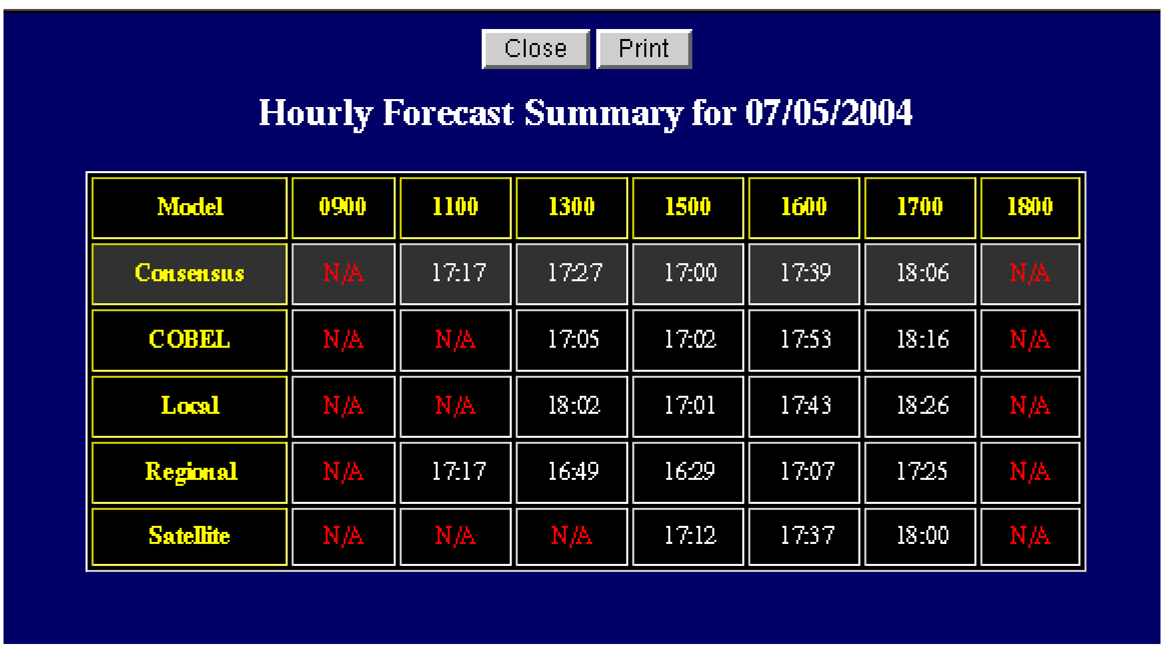
2.5.1.1.6 Satellite SFM

The SSFM is based on statistical correlation of the evolution of cloudiness in the Bay Region, as characterized by the western US GOES visible satellite imagery, and the ultimate cloud burn-off time in the AZ. Forecasts are issued hourly from 15 to 18 GMT.

There are three major steps involved in the generation of the SSFM forecasts. The first is a data pre-processing step to optimize the quality of the data being used for analysis. The second is a grouping of the satellite data grid points into "sectors" that are geographically homogeneous. In the final step, the evolution of the cloudiness within each sector is correlated with the ultimate time of burn-off in the AZ.

2.5.2 Hourly Forecast Summary

When this link is clicked a new window opens up providing a table of each model forecast time (GMT) throughout the current day. This allows the user to view forecast trends. It shows the forecast time of burn-off provided by each of the Component Forecast models for each forecast run time, as well as the Consensus Forecast. If a model forecast has not been produced yet the window will open up indicating that the models are unavailable. If the "Print" button at the top of the page is clicked, the page will print to the default printer. Clicking the "Close" button above the table will close this window. An example of the hourly forecast table is shown in Figure 7.



Hourly Forecast Summary for 07/05/2004

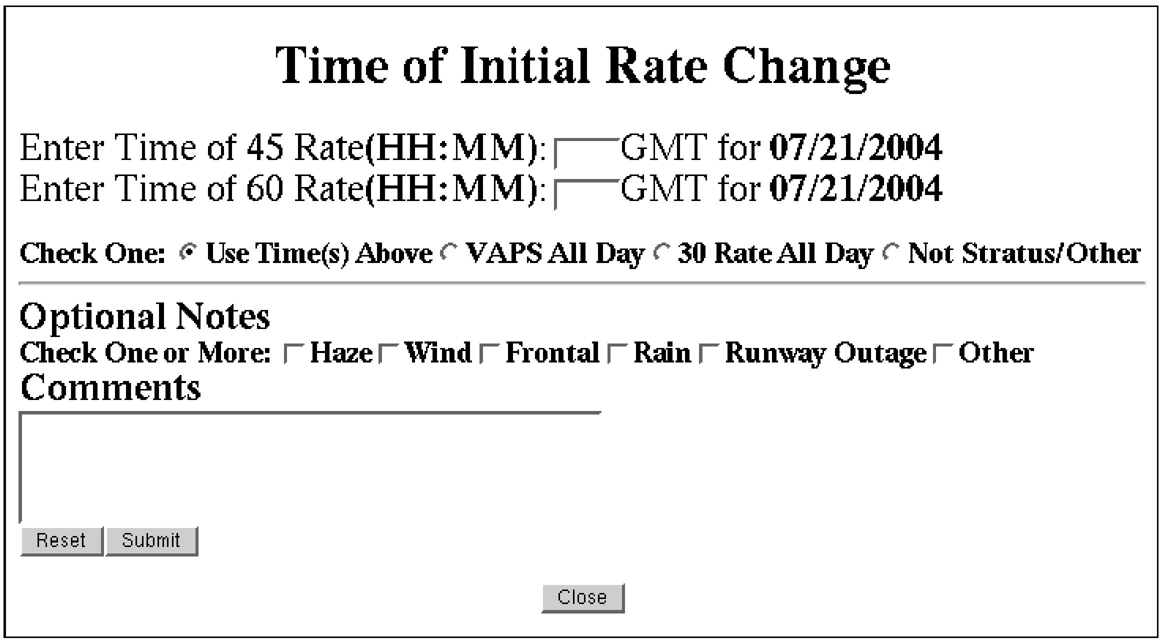
Model	0900	1100	1300	1500	1600	1700	1800
Consensus	N/A	17:17	17:27	17:00	17:39	18:06	N/A
COBEL	N/A	N/A	17:05	17:02	17:53	18:16	N/A
Local	N/A	N/A	18:02	17:01	17:43	18:26	N/A
Regional	N/A	17:17	16:49	16:29	17:07	17:25	N/A
Satellite	N/A	N/A	N/A	17:12	17:37	18:00	N/A

Figure 7. Example of hourly forecast table.

2.5.3 Report Rate Changes

2.5.3.1 Description

This button appears only on the CWSU display at the Oakland Center. It provides an interface for the CWSU to enter verification time of the aircraft arrival rate change each day. The CWSU forecaster on duty will typically enter the verification value prior to the end of the current work shift. If the verification is not entered by 20 GMT, the text on the button will begin to scroll as a reminder to enter the appropriate information. If the CWSU forecaster fails to enter the verification by 00 GMT then an automated program fills in missing report information and the rate report can be changed via the button on the Review page at a later date. An example of the rate change form is shown in Figure 8.


Time of Initial Rate Change

Enter Time of 45 Rate(**HH:MM**): ____ GMT for **07/21/2004**
Enter Time of 60 Rate(**HH:MM**): ____ GMT for **07/21/2004**

Check One: ◉ **Use Time(s) Above** ○ **VAPS All Day** ○ **30 Rate All Day** ○ **Not Stratus/Other**

Optional Notes
Check One or More: ☐ **Haze** ☐ **Wind** ☐ **Frontal** ☐ **Rain** ☐ **Runway Outage** ☐ **Other**
Comments

Reset | Submit

Close

Figure 8. Rate change form.

2.5.3.2 User Interaction

Selection of this button opens a new browser window. The window is formatted to allow entry of a specific verification time at which the aircraft acceptance rate increased to 45 planes per hour and 60 planes per hour on that day. It also provides selectable buttons to indicate that the approach procedure was visible all day (VAPS All Day), if the rate never increased from 30 (30 Rate All Day), or if the day was not an actual stratus day (Not Stratus/Other). The window also includes a section for optional notes where the CWSU forecaster can include comments about the conditions of the day, i.e., Haze, Rain, Wind, Runway Outage, or the forecaster can type in some general comments of their own. The official side-by verification time for that day is the earliest time at which the rate increased to at least 45 planes per hour.

When the “Reset” button is clicked the entire rate change form will be cleared. After the CWSU forecaster enters the rate change, the “Submit” button must be clicked. A message will be displayed informing the CWSU forecaster of the forecast that was submitted. The “Close” button must then be clicked to close the window. If the “Close” button is clicked before the forecast is submitted, an alert message is display indicating that the rate changes were not entered and asks for confirmation of the action taken. If the CWSU forecaster wants to close the window without entering the rate change then click the “OK” button otherwise click the “CANCEL” button to return to the rate change form.

Once the CWSU forecaster has entered a rate change time, the label of the button in the Forecast Frame (Section 2.3.3) will change to “Edit Rate Changes.” If selected the forecaster can edit the previously entered times for that day.

2.5.4 Model Forecast Details

A pull-down menu is provided that allows access to more detailed information regarding the current model forecasts. When one of the menu selections is clicked a new browser window will open to display the Detailed Forecast Summary (Section 2.5.4.1), the Raw Model Output Summary (Section 2.5.4.2) or the Key Model Parameters (Section 2.5.4.3), (Section 2.5.4.4), (Section 3.5.4.5). Much of this information was targeted for the system developers in monitoring model and system performance, but the information may also be useful to operational forecasters.

2.5.4.1 Detailed Forecast Summary

This table (see Figure 9) similar to the Hourly Forecast Summary, with additional details regarding each forecast model. The first column indicates the time of day (GMT) that the Consensus model was generated for each run cycle. This is followed by the Run ID, which simply indicates the initialization hour for each model run. The rest of the information is the Consensus model forecast and the four component model forecasts that were used to create the Consensus model.

The Consensus forecast for each run time is shown, followed by the Confidence indicator (Good or LOW). The next four columns are labeled Pr17 through Pr20. These show the probability that the AZ will be clear prior to each future target hour (17 through 20 GMT) based on the current forecast.

The next set of columns show the forecast for each of the four component forecast models, plus the fractional weight that each forecast was assigned toward generating the Consensus Forecast for that hour. The four weights will always sum to 1.00. If a model forecast has not been produced yet the window will open up indicating that the models are unavailable. Clicking the “Close” button above the table will close this window.

Close

Detailed Forecast Summary for 07/05/2004

Time of Day	Run ID	Consensus						COBEL		Local		Regional		Satellite	
		Fcst	Conf	Pr17	Pr18	Pr19	Pr20	Fcst	Weight	Fcst	Weight	Fcst	Weight	Fcst	Weight
N/A	0900	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
11:36	1100	17:17	LOW	40%	75%	90%	95%	N/A	N/A	N/A	N/A	17:17	1.00	N/A	N/A
13:24	1300	17:27	Good	40%	80%	95%	95%	17:05	0.27	18:02	0.47	16:49	0.27	N/A	N/A
15:23	1500	17:00	Good	60%	95%	95%	95%	17:02	0.22	17:01	0.32	16:29	0.16	17:12	0.31
16:19	1600	17:39	Good	30%	80%	90%	95%	17:53	0.21	17:43	0.36	17:07	0.13	17:37	0.31
17:20	1700	18:06	Good	N/A	60%	85%	95%	18:16	0.22	18:26	0.33	17:25	0.18	18:00	0.28
N/A	1800	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figure 9. Detailed forecast table.

2.5.4.2 Raw Model Output Summary

This table provides more specific detail regarding the sequence of asynchronous model runs each hour. If a model forecast has not been produced yet the browser window will open up indicating that the models are unavailable. For each model run hour, the models will attempt to run repeatedly based on data availability. Each row of the table depicts a change in either the model forecast run time, the model forecast time or the quality number in any one of the models so it is possible that each column could contain identical multiple entries. The first column indicates the sequential time (GMT) that any of the four models attempted to run. The most recent forecast results for the four models are then shown. For each component model, the Run ID, Forecast, Quality, and DayType are shown, as described below. An example of the raw model output is shown in Figure 10.

The Run ID shows the model initialization time (GMT) for each forecast cycle, followed by that model's forecast of clearing time (GMT). This is followed by a "Quality" indicator. If all of the data required by the model were available at run time, the quality of the run is assigned a "5." If all of the required data was not available, or was not sufficiently close to the initialization time (top of the hour), the model quality is reduced to a "4." Since data is received asynchronously, the first model run during a given run cycle hour may initially have a quality of 4, but then upgraded to a 5 upon a subsequent run if new data arrives. The Consensus Forecast generator will allow sufficient time for all models to run with a quality of 5, at which time the Consensus Forecast will be generated. If, however, all of the models do not reach a quality of 5 by 0:35 minutes past the top of the hour, the Consensus Forecast will be generated using the most recent forecast provided by each component model, regardless of the Quality indicator.

The three statistically-based component forecasts are also accompanied by a "DayType" indicator. Each model has multiple choices of forecast equations to use at each initialization time, based on the meteorological conditions at run time. The first two characters of the DayType indicator show the model run hour (GMT); the remaining characters indicate the meteorological conditions which determined the specific forecast equations to use. These are described as follows.

Close

Raw Model Output Summary for 07/05/2004

Time of Day	COBEL			Local				Regional				Satellite			
	Run ID	Fcst	Quality	Run ID	Fcst	Quality	DayType	Run ID	Fcst	Quality	DayType	Run ID	Fcst	Quality	DayType
10:07	----	-----	-	----	-----	-	-----	0900	17:12	5	09Neutral	----	-----	-	-----
11:21	----	-----	-	----	-----	-	-----	1100	17:17	5	11Neutral	----	-----	-	-----
13:21	----	-----	-	----	-----	-	-----	0900	17:12	5	09Neutral	----	-----	-	-----
13:21	1200	17:07	5	----	-----	-	-----	0900	17:12	5	09Neutral	----	-----	-	-----
13:23	1200	17:07	5	----	-----	-	-----	1100	17:17	5	11Neutral	----	-----	-	-----
13:23	1300	17:05	5	----	-----	-	-----	1100	17:17	5	11Neutral	----	-----	-	-----
13:24	1300	17:05	5	1300	18:02	5	13ZiQGE55	1100	17:17	5	11Neutral	----	-----	-	-----
13:24	1300	17:05	5	1300	18:02	5	13ZiQGE55	1300	16:49	5	13Neutral	----	-----	-	-----
15:11	1300	17:05	5	1300	18:02	5	13ZiQGE55	1300	16:49	5	13Neutral	1500	17:22	4	15LoCigSmthCld
15:13	1300	17:05	5	1500	17:01	5	15ZiQGE90	1300	16:49	5	13Neutral	1500	17:22	4	15LoCigSmthCld
15:13	1300	17:05	5	1500	17:01	5	15ZiQGE90	1500	16:29	5	15Neutral	1500	17:22	4	15LoCigSmthCld
15:13	1500	17:02	5	1500	17:01	5	15ZiQGE90	1500	16:29	5	15Neutral	1500	17:22	4	15LoCigSmthCld
15:23	1500	17:02	5	1500	17:01	5	15ZiQGE90	1500	16:29	5	15Neutral	1500	17:12	5	15LoCigSmthCld
16:00	1500	17:02	5	1600	17:25	4	16ZiQLT90	1500	16:29	5	15Neutral	1500	17:12	5	15LoCigSmthCld
16:12	1500	17:02	5	1600	17:21	5	16ZiQLT90	1500	16:29	5	15Neutral	1500	17:12	5	15LoCigSmthCld
16:13	1500	17:02	5	1600	17:21	5	16ZiQLT90	1600	17:07	5	16Neutral	1500	17:12	5	15LoCigSmthCld
16:13	1600	17:53	5	1600	17:21	5	16ZiQLT90	1600	17:07	5	16Neutral	1500	17:12	5	15LoCigSmthCld

Figure 10. Raw model output table.

Local SFM: The Local model makes its forecast equation discrimination based on the quality of the inversion height, which is the key predictor in the model. Each time the inversion height is automatically determined, it is assigned a quality number (0 to 100) based primarily on the strength of the inversion signal. For any given hour, the Local model can run under one of three DayTypes. These are indicated in Table 2.

TABLE 2

Local SFM Day Types

Day Type	Day Type Explanation
ZiQGExx	Inversion quality greater than/equal to threshold xx
ZiQLTxx	Inversion quality less than threshold xx
Default	Default, when quality is indeterminate

Regional SFM: The Regional model makes its forecast equation discrimination based on the direction (onshore, offshore) of the geostrophic flow, as determined by pressure comparisons from San Francisco (SFO) to Arcata (ACV). The DayType indicator shows model run hour (GMT), followed by one of the identifiers in Table 3.

TABLE 3

Regional SFM Day Types

Day Type	Day Type Explanation
Onshore	Onshore geostrophic wind
Offshore	Offshore geostrophic wind
Neutral	No significant onshore or offshore wind
Default	Geostrophic wind is indeterminate

Satellite SFM: The Satellite model makes its forecast equation discrimination based on two characteristics in the cloudiness: 1) The height of the cloud base, and/or 2) the "roughness" of the satellite cloud image in the Bay area. The model performance was found to differ when the cloud base height exceeded 2400 feet due to either the depth of the stratus layer, or the presence of a higher cloud deck. The cloud "roughness" in the visible satellite image (i.e., the variance of the brightness) is also an indication of either a multiple cloud layer situation, or the approach and/or the passage of a synoptic feature. The DayType indicator shows the current conditions under which the Satellite model is operating. These can be seen in Table 4.

TABLE 4

Satellite SFM Day Types

Day Type	Day Type Explanation
LoCig	Low Ceiling
HiCig	High Ceiling
Smooth	Smooth texture (low variance) in Bay area visible cloud brightness
Rough	Rough texture (high variance) in Bay area visible cloud brightness

2.5.4.3 Key Model Parameters

When this new browser window opens it provides detailed information regarding the parameters contributing to the current day's model forecast. Essentially it provides a series of tables, each one representing a sequential run of the model. Each table is accompanied by a header, followed by a list of parameter information. Note that the model run may be repeated several times within a single runtime hour, showing a repetition of tables for the same runtime hour.

The first line of the header indicates the runtime hour for which the model was initialized, followed by the specific model runtime (yyyymmdd_hhmmss), the resulting forecast of the model run (hh:mm hhmm), and the run quality index (5=optimal, 4=degraded). The second line of the header indicates the DayType conditions for which the model was run.

This is followed by a list of parameters that are the terms in the forecast equation, plus their values and contribution to the current forecast. The parameter name is a coded string, indicating the hour from which the data is taken (hh), followed by the location of the data (three-letter airport ID) and a brief description of the parameter.

The specific values of each parameter are then provided in column format. The weight of the parameter indicates its coefficient in the forecast equation, showing the relative importance of its contribution. The current value is then given, accompanied by a "scaled" value, which essentially converts the raw value to a fixed scale ranging from 0.00 to 1.00. Finally, the "partial sum" is provided. This is the product of the weight and the scaled value, and represents the absolute contribution (in hours) that the parameter is contributing to the forecast time.

The last two parameters shown are the Constant Term (in hours) and the Sunrise Time (in GMT). The resulting forecast is computed by starting with the Sunrise Time, adding the Constant, and then adding all of the partial sums. Thus, the table provides an indication of how each parameter contributed to the current forecast. Clicking the "Close" button above at the top of the page will close this window. An example of the key model parameters is shown in Figure 11.

Close

Key Model Parameters For the Local Model on 07/05/2004

Results of Local Model for Run: 1300 Time: 20040705_132400 Result: 18:02 1804 5

Model/Daytype: Local13ZiQGE55

Model Structure Summary

Term	Name	Weight	Value	Scaled_Value	Partial_Sum
1	13sfoTh1	1.664	85.0000	0.2935	0.4885
2	13smbTh1	1.149	211.0000	0.4852	0.5575
3	13sfoTh1dt1	0.656	0.0000	0.4926	0.3233
4	13rblPdx	0.871	1.4000	0.0000	0.0000
5	13sfoZi	0.863	420.0000	0.1445	0.1248
6	13sfoDir	−0.638	0.0000	0.0000	−0.0000
7	13smbCigdt1	0.831	0.0000	0.9629	0.7997
8	13acvPdx	0.884	−0.7000	0.8935	0.7901
9	Constant_Term	1.000	2.0475	2.0475	2.0475
10	GMT_Sunrise	1.000	12.9101	12.9101	12.9101

Figure 11. Key model parameters.

2.5.4.3.1 Local SFM

The Local Key Model Parameters are shown in Table 5.

TABLE 5

Local SFM Parameters

Parameter	Parameter Description
Cig	ceiling height
Cigdt	ceiling height change (1-hour)
Dew	dew point temperature
Dir	wind direction
Mxrad	maximum (clear-day) radiation
Pdx	pressure difference
Rad	solar radiation
Raddt	solar radiation change (1-hour)
Spd	wind speed
Th1	cloud thickness
Th1dt	cloud thickness change (1-hour)
Tmp	temperature
U925	925-mb u wind component
V	v wind component change
Vdt	v wind component change (1-hour)
Zi	Inversion base height

2.5.4.4 Regional SFM

The Regional Key Model Parameters are shown in Table 6.

TABLE 6

Regional SFM Parameters

Parameter	Parameter Description
12InvOak	Inversion height from 12Z Oakland sounding
Cig	ceiling height
CldMax	Maximum fractional cloud cover
Dew	dew point temperature
DwptDprs	dew point depression
First Cloud	Hour at which first pre-dawn cloud ceiling was observed
Mxrad	maximum (clear-day) radiation
Prs	pressure at the surface
PrsDX sfo	pressure difference from SFO
Spd	wind speed
Tmp	temperature
U925	925-mb u wind component
Ucomp	u wind component
Vcomp	v wind component
Vis	visibility

2.5.4.5 Satellite SFM

The parameter is usually a specific satellite sector or group of sectors. The sector ID is then followed by an indicator of the particular statistical parameter used for that sector, where M represents the median cloud brightness value, C represents the fractional cloud coverage of the sector, and “dt” represents an hourly change in the parameter. See Table 7 for an example of the parameters.

TABLE 7

Satellite SFM Parameters

Parameter	Parameter Description
15s2M	15Z satellite image; Sector s2; Median brightness value
16e1e2e3Mdt	16Z image; Sectors e1, e2, and e3; change in Median brightness

One other parameter is sometimes used, coded as sfoMxrad. This represents the maximum solar radiation available at SFO for a given day. Essentially, this is a seasonal indicator of the amount of solar heating available to evaporate the cloud layer.

2.5.5 Model Performance Summary

2.5.5.1 Description

When the Model Performance Summary is selected a new browser window opens that provides an indication of the skill of the various forecast models. If the model performance summary file is not found a message will be displayed indicating that the data is unavailable. The summary is shown as two tables (see Figure 12). The first is a static table showing the performance of the various models during the 2003 and 2004 summer demonstration periods. Performance is presented as Median Absolute Error (hh:mm), for each model, at each runtime hour. Essentially, this represents the “typical” forecast errors observed for each model during the 2003–04 demonstration periods.

The second table is dynamic, updating each day, beginning with the start of the 2004 operational season. A minimum number (adaptable parameter) of rate change verifications need to be entered in order to see the model performance. As with the first table, the second table shows the Median Absolute Error (MAE) for each model, except that the table is updated every day using the previous day's forecast and verification information. The number in parentheses following the MAE shows the number of forecasts that were used in computing the median error. Additionally, this table breaks down the performance of the Consensus Forecast into three different categories: ALL, Good, and LOW. ALL includes forecasts from all days, Good includes only forecasts for which the Confidence was indicated as “Good,” and LOW includes only forecasts for which the Confidence was indicated as “LOW.”

Close

Model Performance Summary

Historical Performance: 2002 – 2003
Median Absolute Error hh:mm

Model	0900	1100	1300	1500	1600	1700	1800
Consensus	1:01	0:54	0:50	0:41	0:38	0:29	0:24
COBEL	1:18	1:06	0:55	0:47	0:42	0:41	0:54
Local	1:20	0:52	0:46	0:43	0:37	0:40	0:34
Regional	1:04	1:02	0:58	0:53	0:57	0:37	0:21
Satellite	N/A	N/A	N/A	0:39	0:32	0:23	0:17

Running Performance: 29 Jun 2004 – 20 Jul 2004
Median Absolute Error hh:mm (Number of forecasts)

Model		0900	1100	1300	1500	1600	1700	1800
Consensus	All	0:44(11)	0:52(15)	0:24(16)	0:45(15)	0:48(13)	0:52(9)	1:14(3)
	Good	0:40(3)	0:52(9)	0:25(11)	0:25(8)	0:27(8)	0:33(3)	1:19(1)
	LOW	0:51(8)	0:41(6)	0:15(5)	0:59(7)	0:55(5)	0:54(6)	1:08(2)
COBEL		0:45(3)	0:43(9)	0:56(12)	0:22(8)	0:08(3)	0:26(3)	0:51(2)
Local		1:04(3)	0:55(9)	0:18(12)	0:38(10)	0:29(8)	0:53(3)	1:54(1)
Regional		0:44(11)	0:57(15)	0:39(15)	1:00(15)	0:26(13)	0:14(9)	0:30(3)
Satellite		N/A	N/A	N/A	0:45(15)	0:48(13)	1:02(9)	1:49(3)

Figure 12. Model performance tables.

2.5.5.2 User Interaction

When the mouse cursor is positioned over any of the performance table titles a mouse over event opens a small pop-up window that briefly explains the scoring technique. Clicking the "Close" button above the tables will close the window.

2.5.6 ETA Analysis Maps

2.5.6.1 Description

This link will launch a separate browser window for displaying 1mb ETA model analysis maps that are stored online. These are provided as an aid to understanding the regional dynamics, particularly in anticipating changes to the geostrophic wind. The default setting is for these maps to be generated only during the summer season.

2.5.6.1.1 Upper Frame

This frame includes the title on the left side and the right side is a date selector to view a different day's worth of data. The four pull-down menus on the right allow the user to select a Year, Month, Day, and Hour of the data the user wants to view. The dates that appear in this list are only added if there is data available for that day so it is possible that the list could skip days if there was no data collected for that day.

2.5.6.1.2 Lower Frame

The lower frame contains an analysis map and links to other types of analysis maps. When the page is first opened the default data that appears is from yesterday.

2.5.6.2 User Interaction

After a year, month, day and hour have been selected from the pull-down menu, clicking the "Submit" button will bring up the data from that time frame. When one of the links to the left of the image is click it will open up the selected image if it is available. If that image is not available, it will state "Image Unavailable." If the image being displayed is clicked, another window will appear that allows a loop of seven day of images. The image animation controls are the same as what is used to loop the satellite images (Section 2.4.3.2.1). Since this page is a new browser window it can be closed the same as any other application.

2.5.7 View Prior Days

2.5.7.1 Description

This display function allows the user to review observations and forecasts from prior days within recent years. Selection of this link will launch a separate browser window for displaying review information from another day.

2.5.7.1.1 Upper Frame

This frame includes the title on the left side and the right side is a date selector to view a different day's worth of data. In the upper right hand corner is a pull-down menu that allows the user to select the Year, Month, and Day of the day of interest.

2.5.7.1.2 Lower Frame

When the page is first opened the display defaults to yesterday's data but any other day for the current season can be reviewed. The layout for this frame is broken down into three columns. The upper left column includes a table of forecasts that is laid out similar to the Hourly Forecast Summary (Section

2.5.2). The actual rate-change verification times for that day are indicated in the green box at the top of the forecast table. Below the forecast table there are two sodar images (Section 2.4.2) that show the 12 to 18 GMT scan from both the San Carlos and San Francisco sites. The middle column contains the solar radiation image data of the entire day (Section 2.4.4) for San Carlos and San Francisco. The right column contains the visible satellite image (Section 2.4.3) on top and the CWSU will see the Report Rate Changes button (Section 2.5.3) below the satellite image. The other links in this column are exactly like what appears on the MSF Display.

2.5.7.2 User Interaction

After a year, month, and day have been selected from the pull-down menu, clicking the "Submit" button will bring up the data from that time frame. Interaction with the Review pages is similar to that of the main display page, with the exceptions noted in the following sections.

2.5.7.2.1 Rate Change Verification

This link opens a table showing the time of transition to parallel approaches on each day of the current stratus season. If the word "More" appears in the green box above the forecast table, dragging the mouse cursor over the word will activate a mouse over event that will open a grey pop-up window containing the CWSU comments.

2.5.7.2.2 Full Day Sodar Page

By clicking this link a new browser window opens up that contains eight sodar images. These images represent four six hour sodar scans for both San Carlos and San Francisco. When finished viewing this page, clicking the "Close" button at the top of the page will close the window.

2.5.7.2.3 Full Page Radiation

Similar to clicking the link on the display page (Section 2.4.4.2); this opens a new browser window showing full-page images of the solar radiation data. Clicking the "Close" button on the top of the page will close the window.

2.5.7.2.4 Satellite Image

When the satellite image is clicked it will open a new browser window similar to what occurs on the display page (Section 2.4.3.2.1). This satellite image can also be animated from this review page. The looping controls are located above the satellite image.

2.5.7.2.5 Report Rate Change

See Section 2.5.3.

2.5.7.2.6 Model Forecast Details

See Section 2.5.4.

2.5.7.2.7 Rate-Change Times

See Section 2.5.8.

2.5.7.2.8 Regional Surface Obs

See Section 2.4.1.2.

2.5.7.2.9 Training/Explanations

See Section 2.5.9.

2.5.7.2.10 Email Question/Comments

See Section 2.5.10.

2.5.8 Rate-Change Times

2.5.8.1 Description

This link provides an up-to-date summary table of side-by approach times for the current stratus season. Listed are the first time on each day that the Arrival Rate was increased to 45 planes per hour and 60 planes per hour. The first time during the day that the Arrival Rate increased to at least 45 planes per hour is considered the "Verification Time" for the Approach Clearing time being provided by the automated forecast guidance models. An example of the rate change table is shown in Figure 13.

2.5.8.2 User Interaction

Clicking on any one of the dates on this window will bring up the review page for that date. If the word "More" appears in yellow in the last column, dragging the mouse cursor over the word will activate a mouse over event that will open a grey pop-up window containing the CWSU comments.

2.5.9 Training/Explanations

This link will launch a separate browser window that provides helpful information for understanding the observations and forecasts provided by the MSF System. It is organized in an outline fashion, providing a Table of Contents corresponding to the order in which items appear on the display. Selection of an individual item from the Table of Contents will bring the Help page directly to that item. Within the display page, each underlined title is also linked to the appropriate location in the help file.

Close

Previous Sideby Times Entered At the CWSU

Date	Verification	45 rate@	60 rate@	Comments
07/20/2004	17:42	17:42	18:12	
07/19/2004	18:11	18:11	18:11	
07/18/2004	N/A	N/A	N/A	No Verification Reported
07/17/2004	N/A	N/A	N/A	No Verification Reported
07/16/2004	15:45	15:45	17:09	
07/15/2004	None	None	None	VAPS All Day
07/14/2004	16:37	16:37	18:35	
07/13/2004	18:35	18:35	N/A	
07/12/2004	15:20	N/A	15:20	
07/11/2004	None	None	None	VAPS All Day
07/10/2004	15:32	15:32	16:43	
07/09/2004	17:30	17:30	N/A	
07/08/2004	17:14	17:14	N/A	
07/07/2004	16:15	16:15	N/A	
07/06/2004	16:51	N/A	16:51	
07/05/2004	17:45	17:45	18:01	
07/04/2004	16:31	N/A	16:31	
07/03/2004	None	None	None	VAPS All Day
07/02/2004	None	None	None	VAPS All Day
07/01/2004	18:55	18:55	19:58	
06/30/2004	17:43	17:43	N/A	
06/29/2004	18:00	N/A	18:00	

Figure 13. Rate change table.

2.5.10 Email Question/Comments

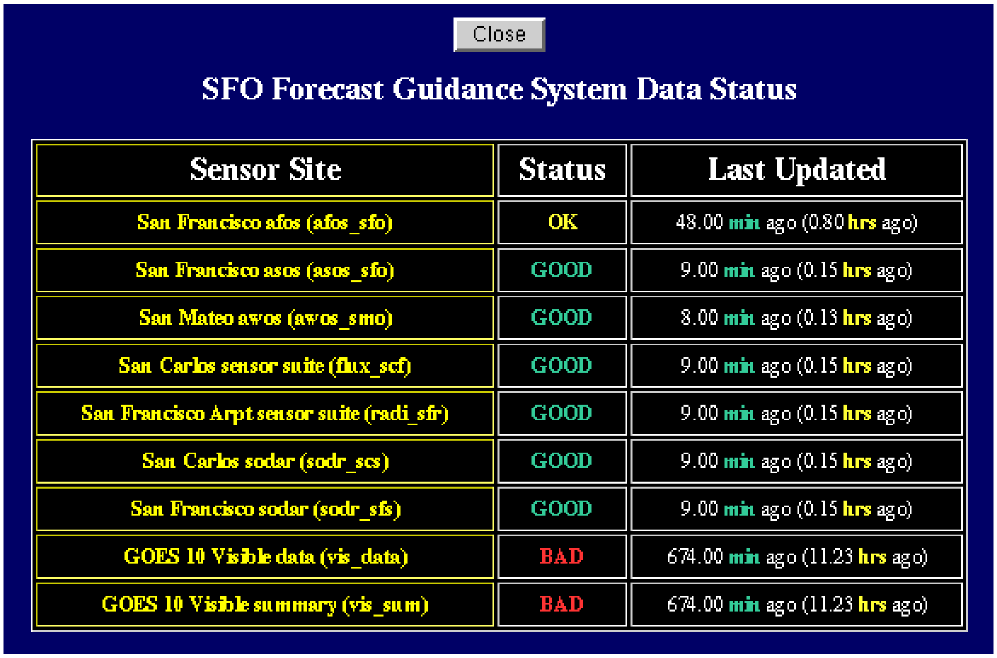
External users may send email to ask questions or provide feedback on the MSF System. When the email link is selected the computer will open up the default mail tool. The email recipients will already be inserted in the email as well as a subject of the email.

2.5.11 System Status

2.5.11.1 Description

This button changes color based on the freshness of the San Francisco sodar data. If these data are less than an hour old then the button is green. If the data are greater than an hour old but less than two hours old then the button turns yellow. If it is older than that the button turns red.

When this link is clicked a new browser window opens that indicates the status of the various data sources within the real-time system. The current status is determined by how recently data from that source have been received, which is also indicated in the table. At a glance the user can determine how current or how old the data are by the color coded words. The word “Good” is in green and it means that the data are less than an hour old. The word “OK” is in yellow to indicate that the data are more than an hour old but less than two hours old. If the data are older than that the word “BAD” appears in red. An example of the system status table is shown in Figure 14.


Close

SFO Forecast Guidance System Data Status

Sensor Site	Status	Last Updated
San Francisco afos (afos_sfo)	OK	48.00 min ago (0.80 hrs ago)
San Francisco asos (asos_sfo)	GOOD	9.00 min ago (0.15 hrs ago)
San Mateo awos (awos_smo)	GOOD	8.00 min ago (0.13 hrs ago)
San Carlos sensor suite (flux_scf)	GOOD	9.00 min ago (0.15 hrs ago)
San Francisco Arpt sensor suite (radi_sfr)	GOOD	9.00 min ago (0.15 hrs ago)
San Carlos sodar (sodr_scs)	GOOD	9.00 min ago (0.15 hrs ago)
San Francisco sodar (sodr_sfs)	GOOD	9.00 min ago (0.15 hrs ago)
GOES 10 Visible data (vis_data)	BAD	674.00 min ago (11.23 hrs ago)
GOES 10 Visible summary (vis_sum)	BAD	674.00 min ago (11.23 hrs ago)

Figure 14. System status table.

2.5.11.2 User Interaction

Clicking the “Close” button at the top of the page will close the window.

3. MSF DATA PROCESSING OVERVIEW

3.1 REAL-TIME SYSTEM INFORMATION

3.1.1 Introduction

The MSF System uses a file-based cache design. A program called the "cvRTCache" handles all real-time data processing from the remote sensors and places these data into a directory structure referred to as day summary (daysum) files. This system design contains a legacy directory structure called the hour file (Hfile) directory. Since the MSF Display System is completely separate from the real-time data processing it still relies on the old hour file (Hfile) directory structure. It is necessary to use lock files to coordinate file access with the cvRTCache. These lock files are written into the Hfile directory. The MSF Display also reads status files written to this directory by the cvRTCache to determine the arrival time of new data.

3.2 OBSERVATION DATA

3.2.1 Overview

The processes that produce these data are controlled by a cron file under the web account. The master cron file is called web.cron and is stored in directory /ll/cv/web/crontab. A master script called webProcCheck located at /ll/cv/web/scripts/ runs from the cron every five minutes to verify that the processing scripts are still running. If it becomes necessary to stop the processing scripts, view Appendix D for details. When any one of the four observation processor or the Interactive Data Language (IDL) Remote Procedure Call (RPC) is not running then all the processes are restarted. After the processes have been restarted an email is sent to the system maintainer to indicate the processes that were not running and that everything was restarted. The email address list parameter file is discussed in Appendix J. IDL is called from a C program and produces the GIF images that are viewed on the MSF Display.

3.2.1.1 Input Data

The observation data processors read data and control files from two different directories. All data are read directly from the daysum directory /ll/cv/data/daysumYYYY, where YYYY is the current four digit year. This directory structure is shown in Table 8. When the three control files for each data type exist they can be read from the Hfile directory /ll/cv/data/cvRTCache/HOUR. These files are the lock file, LOCK.FILE, the status file, STATUS.FILE and the update file, LAST.UPDATE. When a lock file exists it contains the process name and the numerical process ID that produced the lock. When new data arrive, the data's status file is updated to include the data time in epoch seconds. The update file created by the observation data processors contains a copy of the status file data time in epoch seconds. The daysum and Hfile directory does not provide a data quality assessment. Therefore, any data received by the processes described below are assumed to be valid data.

TABLE 8

Daysum Directory Structure

Daysum Directory (/ll/cv/data/daysumYYYY/)		**Description**
inv/	**YYMMDD/**	Inversion height data
rad/	**YYMMDD/**	Solar Radiation data
raob/	**YYMMDD/**	Oakland sounding data
sat/	**YYMMDD/**	Satellite Imagery data
sfc/	**YYMMDD/**	Hourly surface observation data
sodar	**YYMMDD/**	Sodar data

3.2.1.2 Output Data

The processed image data and text from surface observations are placed into a dated directory structure at /ll/cv/web/data. Each different piece of observation data is placed into its own directory. See Table 9 for details. The observation generators maintain a separate log file for each of the processes during the day. These logs are also stored in a dated directory structure at /ll/cv/web/logs. The log files are time stamped to the start time of the processes. At the end of each day a program automatically rolls over the log file to a new date and continues to produce logs as long as the process is running. If any one process is stopped and restarted during the day, a new time stamped log file is created. Therefore, there could be many log files for one or all processes in a single dated directory.

TABLE 9

Display Data Directory

Directory (/ll/cv/web/data/YYMMDD/)	**Description**
obs/	Hourly surface observation data
pyro/	Solar Radiation data
satellite/	Satellite Imagery data
sodar	Sodar data

3.2.1.3 Processing Details

The observation generators run an infinite loop using a two minute sleep time between each data check to see if new data have arrived. This ensures that any new data that are placed in the daysum directory will be processed no more than two minutes later.

Before the data are read the process searches the Hfile directory for a lock file to see if the data are locked by the cvRTCache. If the data are not locked, the observation generator locks the directory for reading and runs through its process to compare the time in the status file to the time in the update file. If the two times are the same the process releases the lock and enters sleep mode again but if the status file time is greater than the update file time then new data are processed by the observation generator. If any of the processes have not produced data for an extended period of time but it appears that data are getting into the daysum directory then the log files should be viewed to verify if the cause of the problem is invalid data or if the status file time indicates no new data. The status file when compared to the update file is used to control the time to update any new data. The process concludes that no new data is available and continues to wait if the status file is deleted or is not updated.

3.2.1.3.1 Handling Lock Files

The observation generators use lock files to coordinate reading and writing of new data with the cvRTCache. After data have been processed the lock must be removed in order to allow the cvRTCache to write new data to the directory. If the process is killed or ends prematurely before the lock is removed it will prevent both the observation generator and the cvRTCache from processing data. The cvRTCache will force a removal of the lock if it finds it to be significantly old.

3.2.1.3.2 Handling Update Files

The update files are written by the observation generators after processing new data. If the update file has not been created or was deleted accidentally or deliberately the observation generator assumes that all data were not processed and the data are processed as if it were new. The only reason for deleting the update file is if the data changed in the daysum directory. After the update file is deleted and the two minute sleep time has elapsed all of the data associated with that update file will be re-generated.

3.2.1.4 Possible Maintenance Tasks

All the observation data are stored in dated directories and they will accumulate over the course of the year. It will be necessary to purge any older data after a certain period of time in order to maintain at least one year's worth of data that can be displayed on the review page. The directory trimming can be performed manually or automatically through a script or a cron. Appendix E describes the directory trimming script in detail.

3.2.2 Sodar Image Generator

3.2.2.1 Description

The script that starts up the sodar image generator is located at /ll/cv/web/scripts/start_generate_sodars. The data required to produce the sodar GIF images are in the daysum directory under the sub-directory called "sodar." The San Francisco sodar image is created from the data file sodar_sfo.YYMMDD.sum and the San Carlos sodar image is created from the data file sodar_sql.YYMMDD.sum. If the sodar data scans an inversion layer it will be depicted on the display as a dark quasi-horizontal line above the 500-foot level. Image generation varies slightly during the "summer operations" mode and the "winter mode."

During summer operations, the Zi inversion algorithm output (red dash line) and the inversion height determined from the OAK balloon sounding (blue tick at 1200GMT) are plotted over the sodar data image. These data are located in the daysum directory in the sub-directory “inv.” The ceiling height data from the surface station closest to the associated sodar (green dash line) is plotted over the sodar data image in both the summer and winter modes. These data are found in the daysum sub-directory “sfc.” The sodar overlay control parameter file is located at /ll/cv/web/params/sodar_ctl.pm and is discussed in Appendix L. This file initializes when the summer operations begin and end and sets flags indicating if the overlays should be drawn. Since each sodar image only contains six hours of data it is necessary to create four six hour sodar images for each sensor site to use them for review at a later date. The images are time stamped according to the last available data that was processed. Figures 15a, b, and c illustrate the data processing details that produce the sodar images.

3.2.2.2 Data Quality and Diagnostics

Occasionally there could be a data outage for which a single one minute scan is not available. In this case the image will contain a thin vertical black line. If several consecutive one minute scans are not available then a thicker vertical black line will appear on the sodar image. There has been a known case where the sodar thought it produced a signal scan but the transmitter/receiver was not working. This occurrence was a field sensor error that produced completely noiseless data. It should be noted that all sodar scans receive ambient noise at ground level and this appears on the sodar image as a horizontal black line at the lowest level on the image. If a sodar image does not display some ground level noise and is completely free of any inversion layers at upper levels, then the sensor may need to be checked.

3.2.3 Solar Radiation Image Generator

3.2.3.1 Description

The script that starts up the sodar image generator is located at /ll/cv/web/scripts/start_generate_sodars. The data required to produce the sodar GIF images are in the daysum directory under the sub-directory called “sodar.” The San Francisco sodar image is created from the data file sodar_sfo.YYMMDD.sum and the San Carlos sodar image is created from the data file sodar_sql.YYMMDD.sum. If the sodar data scans an inversion layer it will be depicted on the display as a dark quasi-horizontal line above the 500-foot level. Image generation varies slightly during the “summer operations” mode and the “winter mode.” During summer operations, the Zi inversion algorithm output (red dash line) and the inversion height determined from the OAK balloon sounding (blue tick at 1200GMT) are plotted over the sodar data image. These data are located in the daysum directory in the sub-directory “inv.” The ceiling height data from the surface station closest to the associated sodar (green dash line) is plotted over the sodar data image in both the summer and winter modes. These data are found in the daysum sub-directory “sfc.” The sodar overlay control parameter file is located at /ll/cv/web/params/sodar_ctl.pm and is discussed in Appendix M. This file initializes when the summer operations begin and end and sets flags indicating if the overlays should be drawn. Since each sodar image only contains six hours of data it is necessary to create four six hour sodar images for each sensor site to use them for review at a later date. The images are time stamped according to the last available data that was processed. Figures 16a, b, and c illustrate the data processing details that produce the sodar images.

3.2.3.2 Data Quality and Diagnostics

The theoretical incoming solar radiation is plotted on the image all day even if the sun angle is still low. The measured solar radiation will sometimes spike above the theoretical solar radiation because the sensor will occasionally measure solar radiation reflected from the cloud as well as the incoming radiation. This typically occurs as the clouds begin to break up. Most of the time the measured radiation curve follows the theoretical radiation curve fairly closely during cloud-free conditions.

If the measures radiation curve varies significantly (> 5%) from the theoretical curve during cloud-free conditions, it is an indication that the sensor likely needs to be recalibrated.

3.2.4 Satellite Image Generator

3.2.4.1 Description

The script that starts up the satellite image generator is located at /ll/cv/web/scripts/start_generate_ satellite. The data required to produce the Satellite images are located in the daysum directory under the sub-directory called "sat." The satellite image is created from a data file called sat_vissec_goes10.YYMMDD.dat. The data is a color map of values from 1-254. The other colors are reserved for the overlay of the San Francisco Bay region. Before the data are handed to the IDL process the data are checked to be sure the color values are within that range. If the color value is 0 or 255, it is set to 1. Figures 17a, b, and c illustrate the data processing details that produce the satellite images.

3.2.4.2 Data Quality and Diagnostics

The satellite image is produced from processed visible satellite data whose brightness values are normalized to account for the sun angle. If the image looks distorted or the registration appears to be off, this is usually attributed to the processed satellite data and not from the image generating software.

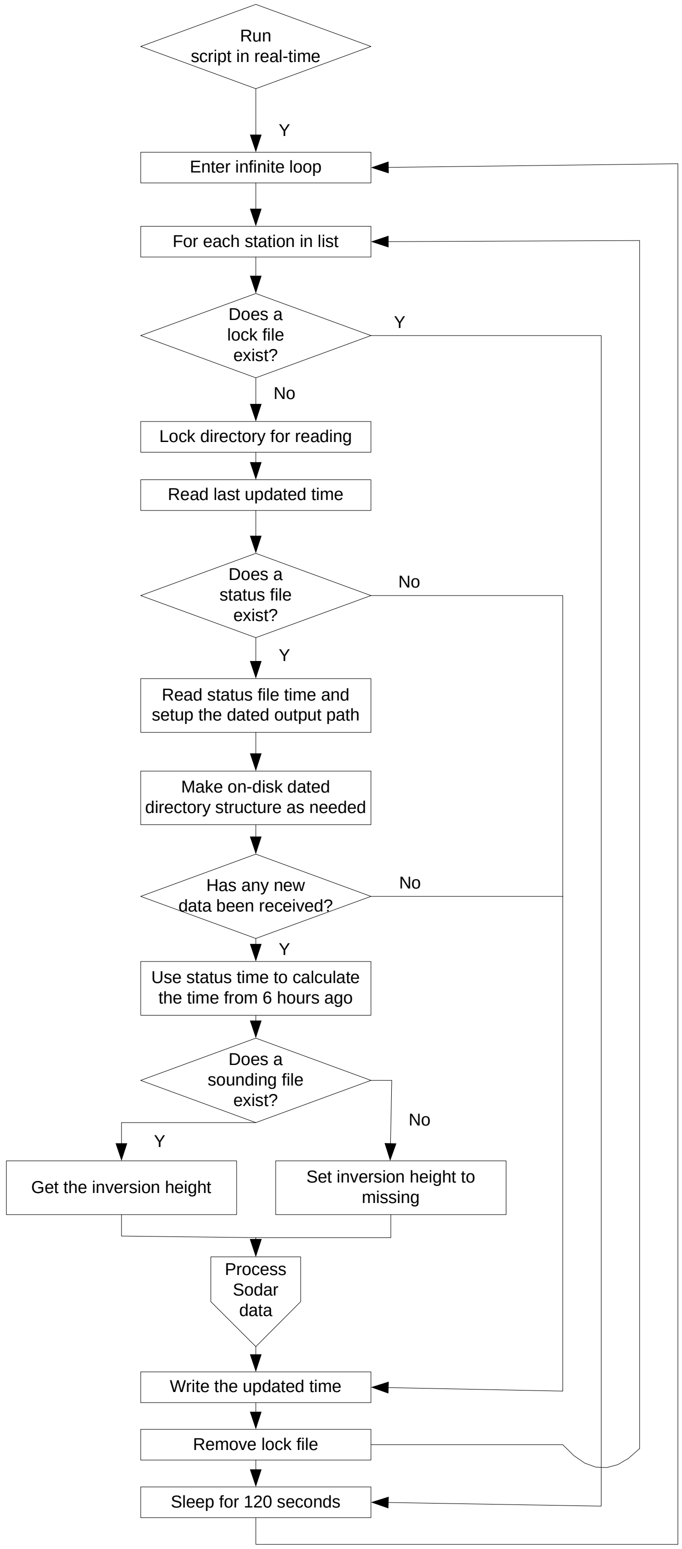


Figure 15a. Generate sodar flow diagram.

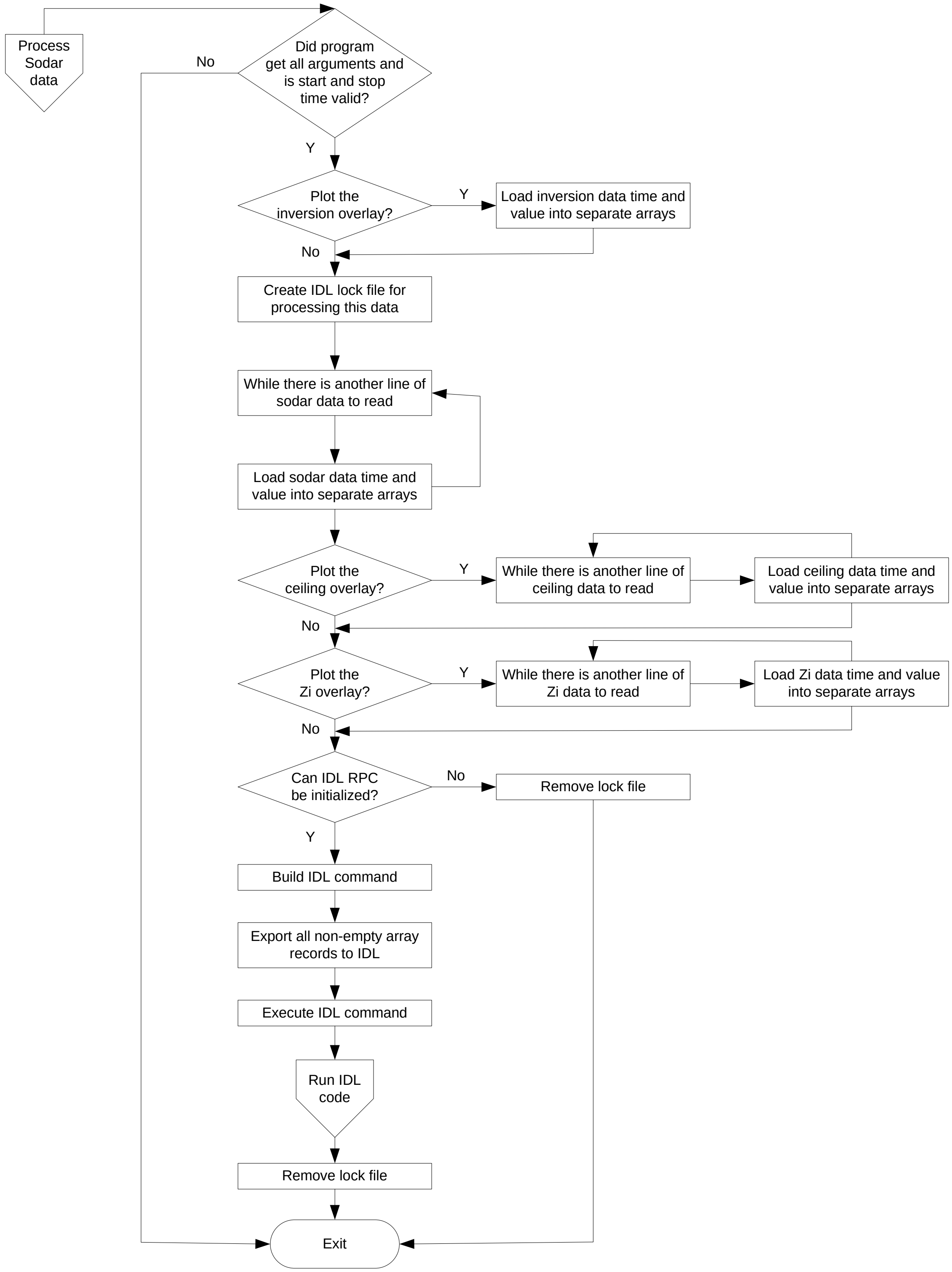


Figure 15b. Generate sodar flow diagram (continued).

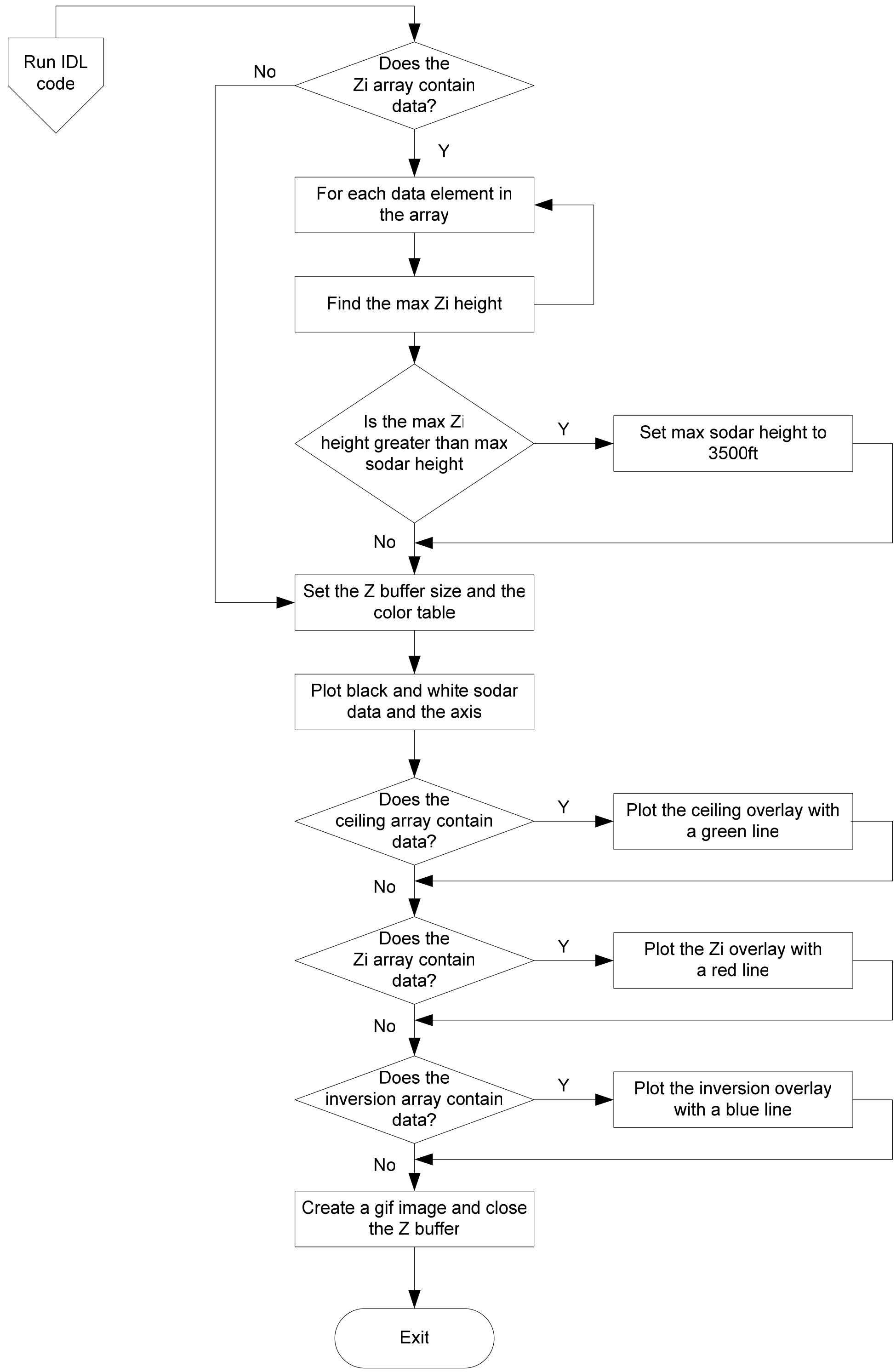


Figure 15c. Generate sodar flow diagram (continued).

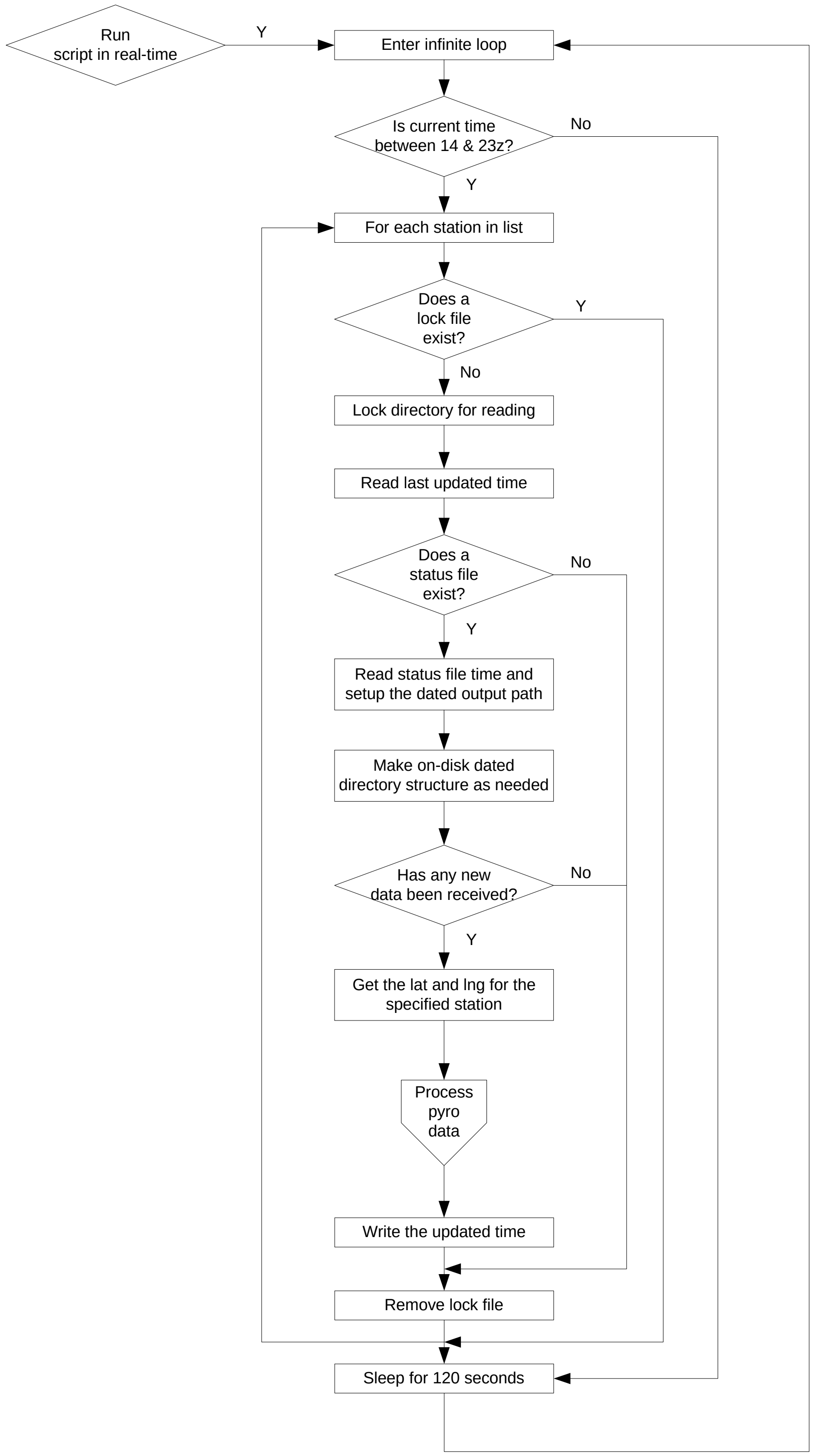


Figure 16a. Generate pyro flow diagram.

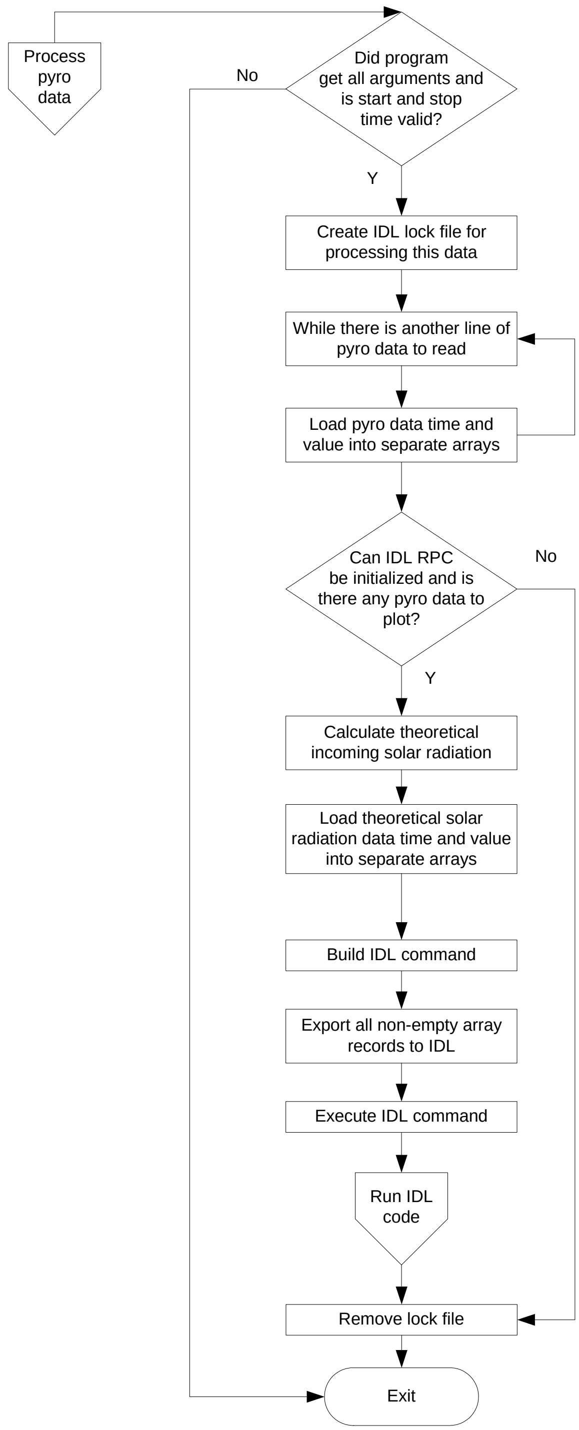


Figure 16b. Generate pyro flow diagram (continued).

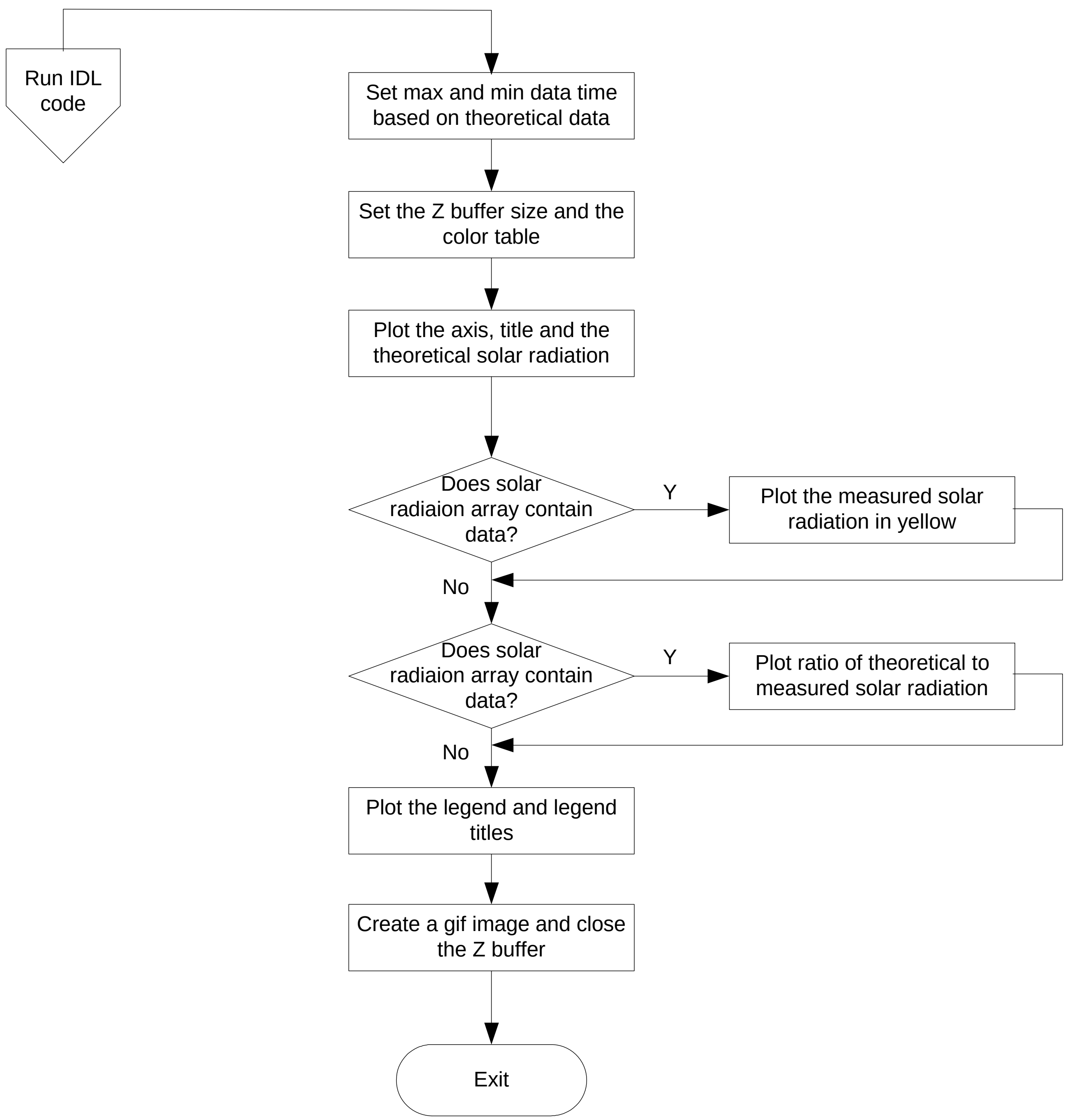


Figure 16c. Generate pyro flow diagram (continued).

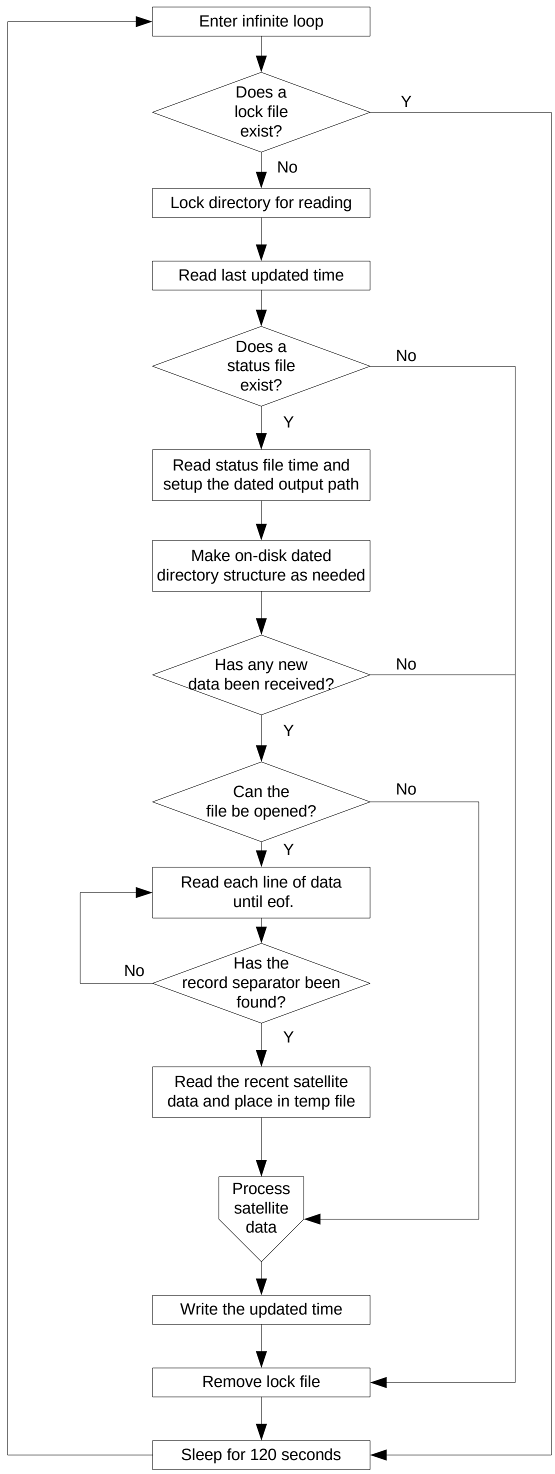


Figure 17a. Generate satellite flow diagram.

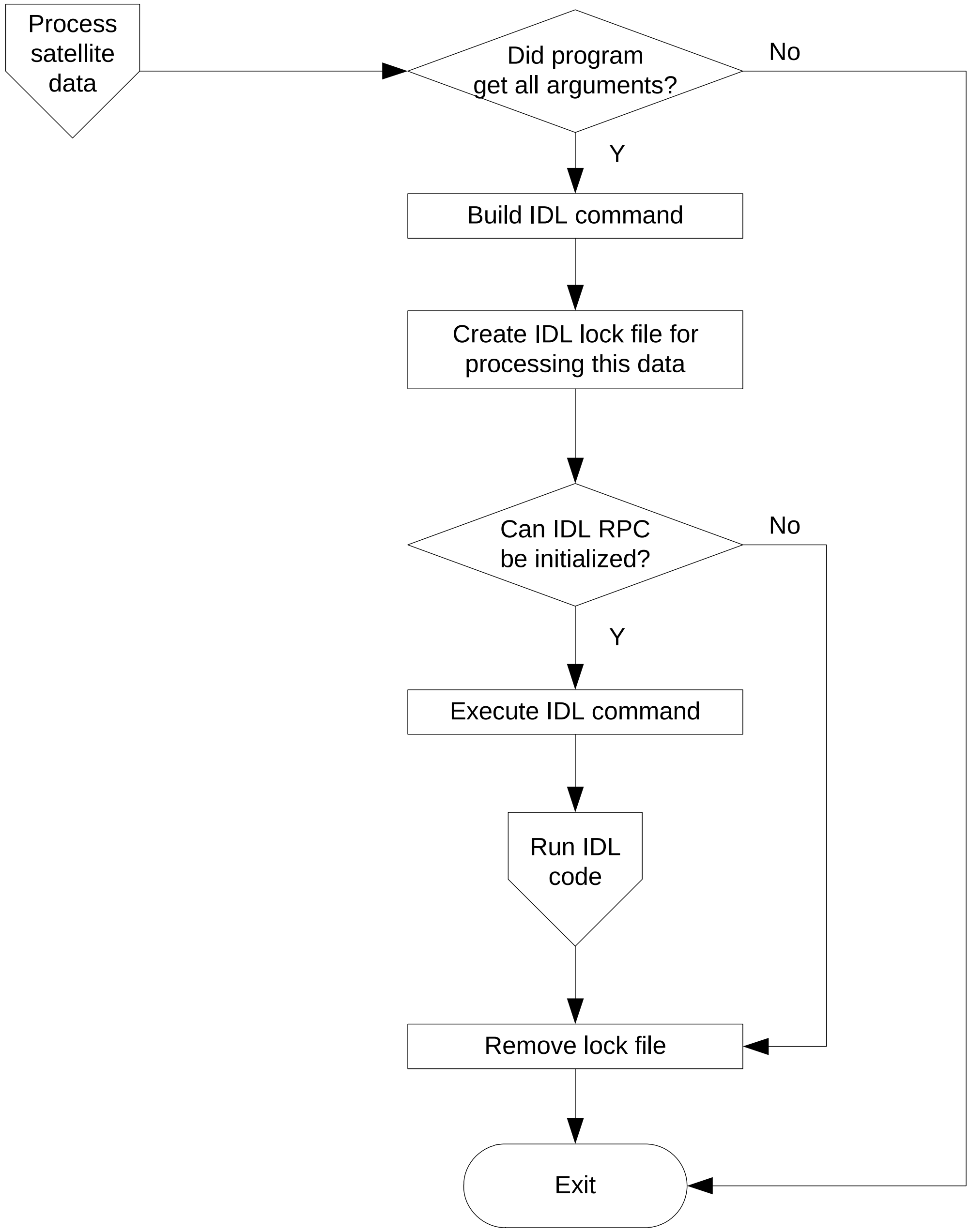


Figure-17b. Generate satellite flow diagram (continued).

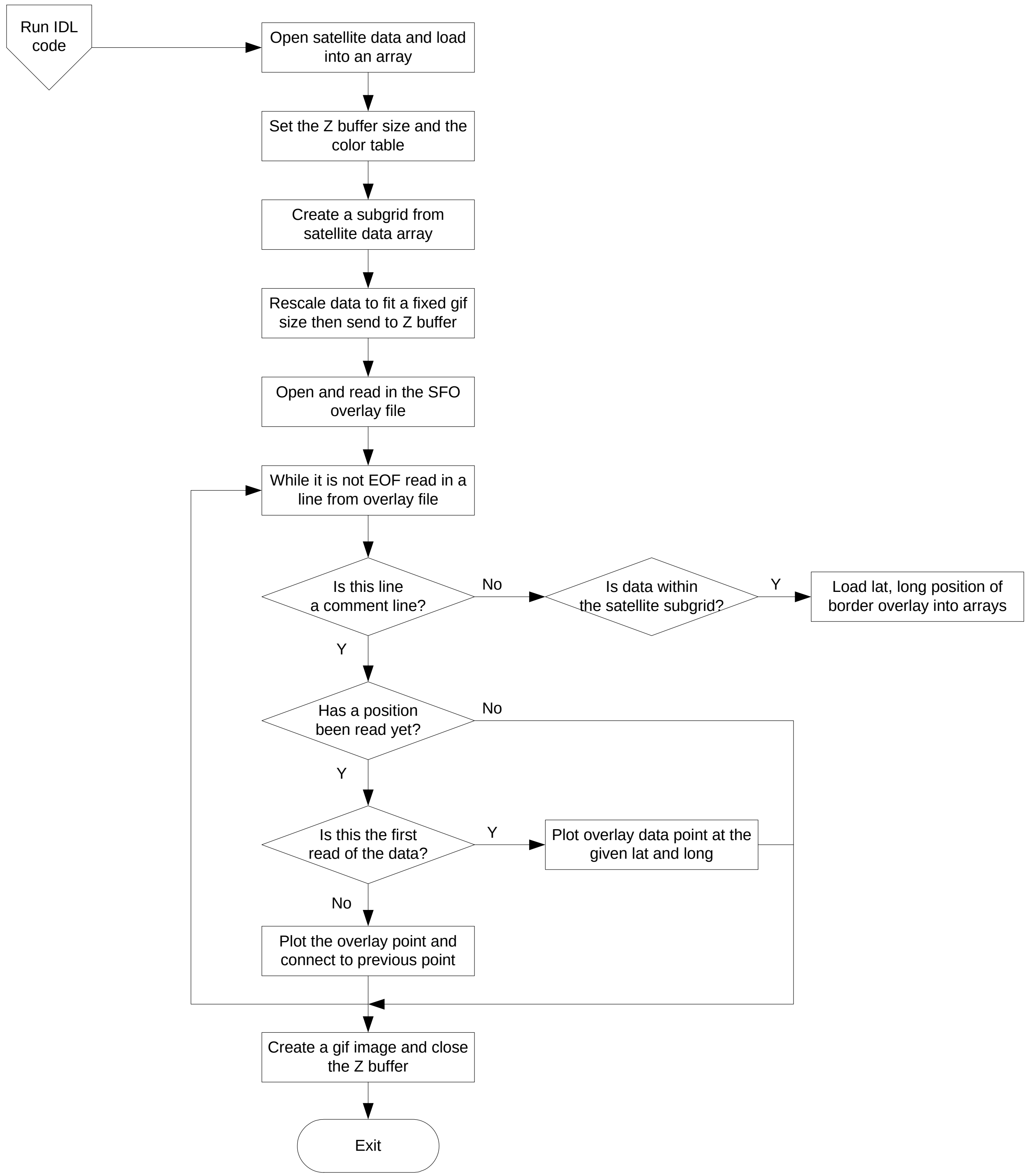


Figure 17c. Generate satellite flow diagram (continued)

3.2.5 Surface Observation Generator

3.2.5.1 Description

The script that starts up the surface observation generator is located at /ll/cv/web/scripts/start_generate_obs. The data required to produce the surface observations are located in the daysum directory under the sub-directory called “sfc.” The surface observation data are created from data files that have the following format (sfc_“type”_“station”.YYMMDD.sum, where type is afos, asos, or awos and station is the three letter surface station abbreviation). Figure 18 illustrates the data processing details that produce the surface observations images.

3.2.5.2 Data Quality and Diagnostics

Most of the surface observations are collected hourly but a few stations have five minute updates. Some of the hourly surface stations only collect data during the day and then shut down for a few hours at night. This is a normal occurrence and can be observed on a daily basis. The SMB surface observation has been know to produce an erroneous low ceiling condition at or below 100 feet. This will appear in the SMB surface observations and the ceiling overlay in the San Carlos sodar image. There might be an occasion where there could be a significant data outage from one of the surface stations. If this occurs it would be best to look in the daysum directory to verify that the data were actually there to be processed. The observation generator will process any data contained in the daysum directory.

3.3 FORECAST DATA

3.3.1 Overview

A detailed discussion of the model forecast guidance products are contained within a companion document. This section will give a general overview of the component and consensus model guidance and the logic that creates the models. The model guidance will only run during the summer stratus season when clouds and stratus are in and around the San Francisco Bay Area.

3.3.1.1 Component Models

During the summer stratus season, the four component models COBEL, Local SFM, Regional SFM, and Satellite SFM are run from 09:00 GMT to 18:00 GMT. The component models run every five minutes around the top of the hour using the most recent observation data available to each model.

3.3.1.2 Consensus Model

The consensus model is run only once per hour. It will run as soon as four component models have run, or at 35 minutes past the hour, whichever comes first.

3.4 TEXT DATA

3.4.1 Overview

The processes that produce the text data are controlled by a cron file under the web account. The master cron file is called web.cron and is stored in directory /ll/cv/web/crontab. Two processes run all year long the other processes run only during the summer stratus season. The processes that run all year long are the system data status and the Apache web access logs. All the files produced are used to determine the performance of the system and the access load on the system during the day.

3.4.2 Rate Changes

3.4.2.1 Description

The CWSU has the ability to enter the aircraft acceptance rate, AAR, during the day. During a stratus event, the AAR drops to 30 planes per hour. If the AAR increases to 45 and/or 60 planes per hour the CWSU enters the report into the base station computer. If the AAR never dropped below 45 planes per hour, that is considered a VAPS day, Visual Approaches all day. If the AAR never increased above 30 planes per hour then that is considered IFR, Instrument Flight Rules, all day. This file is a summary of all the rate reports entered by the CWSU during the summer stratus season.

3.4.2.2 Input Data

The "sideby" times are stored in a dated directory corresponding to the date for which the report was intended. The location of the data is /ll/cv/web/data/YYYYMMDD/sideby_times.YYYYMMDD.

3.4.2.3 Output Data

A tab delimited text file of the yearly sideby times entered by the CWSU is stored at /ll/cv/web/data/daily/burnoff_times.YYYY. This file is overwritten every time the CWSU enters a new sideby time. This file includes the time that AAR went to 45 planes per hour and/or to 60 planes per hour if available. If the CWSU entered a comment about the day's events it is also included in the file. If this file output is ever changed or altered in the processing script, then the display page process that reads this file must also be changed to read the new output style.

3.4.2.4 Processing Details

This file is written in one of two ways. The first method is when the CWSU enters a rate change report. If a report was not issued, a cron job, /ll/cv/web/scripts/check_yesterdays_burnoff.pl, will set the rate change report to state that the CWSU did not enter the rate change and an email is sent to the system maintenance staff to inform them that the rate change was not entered. The script that actually produces the output file is /ll/cv/web/scripts/generate_burnoff_file.pl. This script scans all the dated directories in /ll/cv/web/data for a sideby time report. When a report is found it is parsed and formatted into a single tab delimited line. The entire list of rate change reports are re-written with the most recent being at the top of the list. The reason that the entire list is re-written is in the event that the CWSU changes a rate report at a later date. This change could be to modify a missing report to an actual rate report or the CWSU may add comments to an existing report.

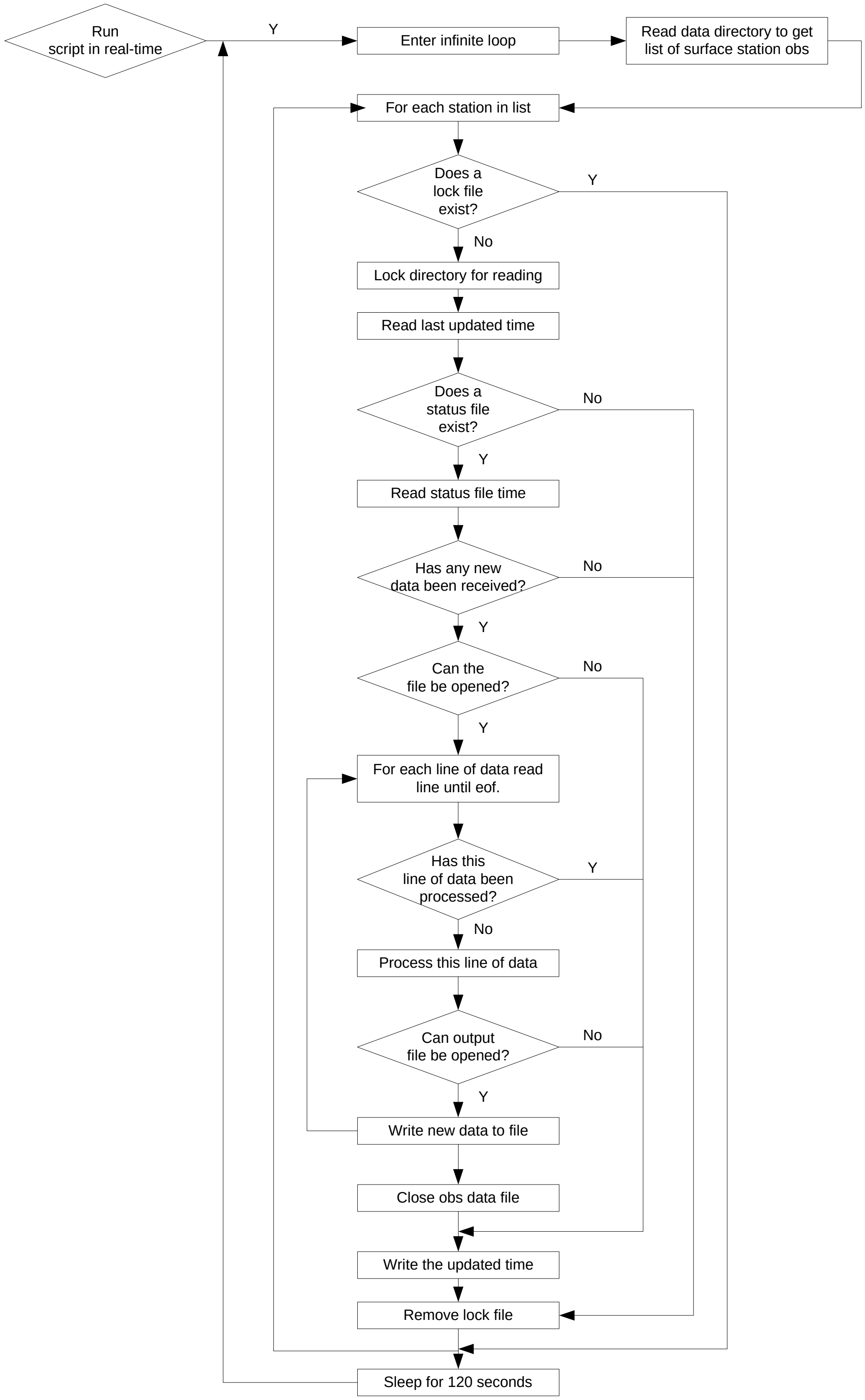


Figure 18. Generate obs flow diagram.

3.4.2.5 Possible Maintenance Tasks

Every year a new file is created that contains the full year's worth of CWSU rate changes. The sideby time report is date stamped and is placed in /ll/cv/web/data/daily. This file is only about 6.5KB but the old files are never removed from the directory. Therefore, these files will continue to build up in this directory and will eventually need to be removed.

3.4.3 Model Performance

3.4.3.1 Description

During the summer stratus season the forecast models are scored against the sideby time that is entered by the CWSU. The CWSU rate change time is considered the verification of when the stratus cleared. All the forecast models were developed using this verification time. This file contains two forecast scores, the median absolute error, and the percentage of time that forecast was within 45 minutes of the verification.

3.4.3.2 Input Data

The data to score the model performances are gathered from two different places. The first file that is read is the sideby time report which is the year to date summary of the time the AAR changed. The other data is stored by the cvRTCache at /ll/cv/data/ProductsArchive/YYYYMMDD/ConsensusForecast/. The files named latest_consensus.dat are archived copies of the consensus forecasts that were displayed on the MSF Display's forecast frame.

3.4.3.3 Output Data

After the scoring has been performed the file is placed at /ll/cv/web/data/daily/model_ performance.YYYY. This file contains two types of scoring data, a median absolute forecast error and percentage of time that the forecast error was no more than 45 minutes. The beginning and ending dates of this performance score are at the top of the file that is produced. The rest of the data is tab delimited text that is formatted to look similar to a table to allow for easy reading of the data. This file is overwritten every time the model performance script is run. If this file output is ever changed or altered in the processing script then the display page process that reads this file must also be changed to read the new output style.

3.4.3.4 Processing Details

Before the model data is scored with the script /ll/cv/web/scripts/ModelPerf.pl a few parameter files must be read in. One parameter file /ll/cv/web/params/model_perf.pm controls the beginning and ending date that will be scored, the minimum number of verification reports that must be issued before the score is displayed, and the allowable forecast error, forecasted time + or - 45 minutes. The parameter file also controls the maximum verification and forecast times. If the forecast or verification time is greater than this value it is set to this maximum time.

3.4.3.5 Possible Maintenance Tasks

Every year a new file is created that contains the model performance scores for the entire stratus season. The model performance file is stamped with the current year. This file is very small and will only occupy about 1.2KB. These model performance files will accumulate in this directory year after year and will eventually need to be removed.

3.4.4 System Status

3.4.4.1 Description

In order to keep track of how current the data is on the display the System Status script located at /ll/cv/web/scripts/SystemStatus.pl runs every minute off a cron. When new data arrive its status file is updated to include the data time in epoch seconds. All the status files for every piece of incoming data are checked and assessed to determine its age.

3.4.4.2 Input Data

This process only requires the status files, STATUS.FILE, which are stored in the Hfile directory /ll/cv/data/cvRTCache/HOUR.

3.4.4.3 Output Data

The tab delimited text file is stored at /ll/cv/web/data/daily/system_status. This file includes the data type, the age of the data, “good,” “ok,” “bad,” and the last time the data were updated. If this file output is ever change or altered in the processing script then the display page process that reads this file must also be changed to read the new output style.

3.4.4.4 Processing Details

The first line of the file is a determination of the overall age of the data. This is based only on the San Francisco sodar data. If the data are less than or equal to one hour old then the data is considered to be “good,” less than or equal to two hours old is “ok,” and greater than two hours old is “bad.” Every other line in the rest of the file includes the data type followed by the age of the data and the last time the data were updated. The calculation to find the last update time is done by comparing the status file to the current system time.

3.4.5 Apache Web Access Logs

3.4.5.1 Description

The MSF Display is a web based display that serves it pages through an Apache server. Each time the page is accessed or is refreshed the Apache server receives a request from the web browser to send new data to the display. Every data request is logged in the Apache server log files.

3.4.5.2 Input Data

The Apache server log files are located at /var/log/httpd/YYYYMM. In this directory there are two files per day. The access logs contain a request from a web browser to send data. The error logs contain problems that occurred during the request processes. The information in these files contains the IP address or DNS host that accessed data from the Apache server, the user name that accessed the base station computer if the user did not have trusted access and the document that was accessed. In order to create a new trusted user, see Appendix B for details. If the user does not know their IP address or DNS host name and they have accessed the web page then the access or error logs can be checked to find the IP address or the DNS host name. See Appendix C for details.

3.4.5.3 Output Data

The daily log files are read in and are created as viewable web pages containing a summary of access requests that the Apache server logged. The daily statistics files are located at /var/www/html/Admin/daily/ and the monthly statistics file is located at var/www/html/Admin/monthly/. These files can be viewed through a web browser by entering the current MSF Display IP address along with /Admin, ie http://64.175.253.249/Admin/. This page is password protected and can be set or changed by the system's root account.

3.4.5.4 Processing Details

Since the log files that are being processed are produced by the Apache web server, there is a utility that creates the HTML summary files. This utility is free and can be downloaded from the internet. The utility program is called wwwstat and is located at /usr/local/bin/. The script that initializes all the variables and file names that are to be processed is /ll/cv/web/scripts/web_logger/process_weblogs.pl. The HTML daily summary files are only stored for the last 10 days but all the monthly summary files are stored. After the access logs are created an email is sent to the system maintainer to indicate that the latest web access logs are ready to be viewed. The email address list parameter file is discussed in Appendix J.

3.4.5.5 Possible Maintenance Tasks

Both the Apache log files and the HTML monthly summary files are never deleted from the base station computer. Each HTML monthly summary file occupies only 165KB on average but the Apache access log files occupy about 2 to 3 MB for each day.

3.5 FTP DATA

3.5.1 Overview

The processes that collect the FTP data for generating the ETA analysis maps are controlled by a cron file under the web account. The default setting is to collect these data only during the summer season. The master cron file is called web.cron and is stored in directory /ll/cv/web/crontab. The data are collected in GRIB format from NCEP four times a day and include the ETA model initialization at 00, 06, 12, and 18 GMT. Each file is approximately 14 to 15 MB in size. After the data are downloaded and used for creating GIF images, the GRIB file is deleted.

3.5.2 Analysis Maps

3.5.2.1 Description

A regional 1-mb resolution surface pressure analysis is generated from GRIB-formatted ETA model data. The analysis maps are generated from the ETA model initializations at 00, 06, 12, and 18 GMT.

3.5.2.2 Input Data

The data required to produce the analysis maps must first be downloaded via FTP from the NCEP forecast computer. The data are downloaded into /ll/cv/web/tmp.

3.5.2.3 Output Data

The data that are produced from this process are placed into a dated directory structure at /ll/cv/web/archive/YYYYMMDD/grads/. The log files are placed into the /ll/cv/web/tmp/ directory.

3.5.2.4 Processing Details

A free processing package called the Grid Analysis and Display System (GrADS) produces all the one mB analysis maps in GIF format. In order to download and process the data the script /ll/cv/web/arc_scripts/start_eta or start_gdas must be set to run. Six different maps are produced every six hours at five different levels of the atmosphere. Two maps depict the geopotential height and the vorticity at the 500mB level. One map shows the geopotential height and relative humidity at the 700mB level. The 850mb level map depicts the geopotential height and the temperature. At the 925mB level the map includes the geopotential height, the temperature and wind vectors. The last map shows the surface level which includes the mean sea level pressure and the 1000mb to the 500mb thickness.

3.5.2.5 Data Quality and Diagnostics

Since all this data are collected via an FTP web address, these scripts could potentially require multiple changes over the years in order to collect the data in the event that the web site address changes. It is possible that when the script runs that it never actually gets the FTP data. When this happens it does not crash and it does not affect the system’s performance but any maps that are produced from this data will appear completely white. If users continue to want to see this data then it should probably be set up so that the data are pushed into the system instead of having to be pulled through FTP.

3.5.2.6 Possible Maintenance Tasks

All the GIF images that are produced by this process are stored in a dated directory. The data will eventually need to be pruned to prevent filling the disk. Directory trimming can be performed manually or automatically through a script or a cron. Appendix E describes the directory trimming script in detail.

4. MSF DISPLAY PROCESSING/PERFORMANCE OVERVIEW

4.1 MSF DISPLAY

4.1.1 Overview

The MSF Display is a web page with multiple frames created with Common Gateway Interface (CGI) scripts written in Perl and includes JavaScript in order to make the display more dynamic. All of the CGI files located on the base station computer reside in the directory /var/www/cgi-bin/. The display is launched when the main script is requested by a web browser. The main script is in the CGI directory /cv_main/CvTop.pl. When CvTop.pl is called it creates the three frames on the web browser and opens the Title frame, Observation frame, and Forecast frame. The display can be set into one of three modes from the browser window. The default setting is to display the MSF Display into real-time mode. The MSF Display can be set into a review or snapshot mode and this is discussed in Appendix E. When the MSF Display is set into test mode, see Appendix G, it behaves as a real-time display but the CWSU rate change button will appear in the forecast frame. This allows testing of the CWSU rate change form but it will not issue an official rate report. Appendix K details the host computer parameter file the controls which computer the official rate change report can be entered from.

4.1.2 Title Frame

4.1.2.1 Description

The title frame, /var/www/html/cv/title.html, is only an HTML text file that contains the MSF Display title and JavaScript that creates the clock to the right of the title.

4.1.2.2 Processing Details

Since this frame is an HTML text file the only processing occurs within the JavaScript clock. The clock on the frame is based on the computer's system clock where the web browser is running. A correction factor is added to account for the different time zones across the nation. If the system clock on the computer is not set to the correct time zone and/or the time is not correct then the title page will show the wrong time.

4.1.3 Observation Frame

4.1.3.1 Description

The observation frame, /var/www/cgi-bin/cv_main/observations.pl, is a script that controls the layout and flow of the observation data around SFO. The lower left hand frame of Figure 1 shows an example of the observation frame. Several links are located within the frame in order to view Observation Products. The processes to view these products are discussed in Section 4.2. The processing of the GIF images and the hourly surface data was discussed in Section 3.2.

4.1.3.2 Input Data

All the data that are displayed in the observation frame are pre-processed and stored for retrieval later. This pre-processing step was discussed in the previous section. See Table 4-2 for the directory structure of the stored observation data that serve as input to the display. Since the display page can not access the image data directly from the file location, symbolic links are placed in the directory /var/www/html/cv/data/. If these links are ever removed then the image data would become inaccessible.

4.1.3.3 Processing Details

The observation frame is produced by two processing scripts, observations.pl and observations.pm. The processing that takes place in the Perl module, /ll/cv/web/params/observations.pm sets up a variety of properties for the frame as well as the data images and links. File names and directory locations are initialized in /ll/cv/web/params/script_files.pm. The refresh rate of the frame is set in this file. Also a data time out is included that places an image unavailable flag onto the display for any of the data that have not arrived within a set time limit. Titles and sizes of the images are all set in this file as well. Clickable links are set up to direct a user to the data for the current day, to the online help pages, and to the additional observation products.

The Perl script, /var/www/cgi-bin/cv_main/observations.pl, sets up the format and location of the images and the table data. The background color of the frame as well as the color of the text on the frame is set using an HTML formatting standard called Cascading Style Sheets (CSS). Inline JavaScript code is used to open up separate browser windows to display either help pages or selected observations. JavaScript code that is linked to the frame controls the GIF image animation tool that controls the animation of the satellite image.

4.1.3.4 Data Quality and Diagnostics

The observation frame refreshes every five minutes to ensure that the most recent data are displayed. The frame will not distinguish between current data and old data although there is one exception. Any data that are place in the web data directory /ll/cv/web/data/ will be read and assumed to be current, but the surface observation data will be checked for timeliness. If the surface observations are older than forty minutes then they are timed out and “N/A” is shown across all the fields.

If sodar data do not come in at all during the day then the unavailable image is shown. If data were coming in but stopped during the day then the last available sodar image will be displayed until a new one replaces it.

The solar radiation data are only available after sunrise. If after 15:00 GMT the data have still not arrived then the unavailable image is placed into the solar radiation location on the frame. If the data were coming in but stopped, then the last available solar radiation image is displayed until a new image replaces it.

The satellite data will always show the latest image available on the observation frame even if the satellite image is old relative to the other data available. If no satellite data are sent then the unavailable image is placed into the satellite image location on the frame.

4.1.3.5 Possible Maintenance Tasks

The observation frame is mostly maintenance free with a few minor exceptions. Changes in web browser technology and web browser standards have altered the interaction of the display over the last few years. Minor adjustments were required in the past for older HTML tags that became obsolete and where no longer used or supported. The GIF animation JavaScript software works differently on older versions of Internet Explorer and Netscape than it does on the newer versions. This required the use of different coding standards based on the type of browser software being used.

4.1.4 Forecast Frame

4.1.4.1 Description

The forecast frame, /var/www/cgi-bin/cv_main/forecasts.pl, is a script that controls the layout and flow of the model guidance data and the links for additional forecast details. The right hand frame of Figure 3-1 shows an example of the forecast frame. The frame will run in one of two modes, summer and winter modes. These different modes were discussed in Section 1.2.1.3. This frame is one of the most dynamic frames on the MSF Display with a majority of the interaction occurring in this frame. All the links launch Perl processing scripts that read and format data from various files in various different formats. The forecast frame format parameters are discussed in Appendix I. These processing scripts are discussed in Section 4.3, Forecast Processes.

4.1.4.2 Input Data

This frame accesses several pieces of data that control the forecast guidance update. The forecast guidance data is located at /ll/cv/data/cvRTCache/LATEST/latest_consensus.dat. This data is placed into a table at the top of the forecast frame. Above the forecast table there will occasionally appear a report stating the approach zone is clear. This section reads data from /ll/cv/data/cvRTCache/LATEST/ burnoff_occurred or from the CWSU rate change file /ll/cv/web/data/YYYYMMDD/sideby_ times.YYYYMMDD. The Probability of Clearing table reads data from the directory /ll/cv/web/params/prob_tables. The table that is read for input to the display is based on the initialization time of the forecast guidance.

The status button at the bottom of the frame reads a file to determine what color the button should be based on the timeliness of the data. This file is /ll/cv/web/data/daily/system_status.

4.1.4.3 Processing Details

The forecast frame is produced by two processing scripts, forecasts.pl and forecasts.pm. The Perl module, /ll/cv/web/params/forecast.pm, initializes a variety of properties, images and files as well as controls the layout and format of the frame. All the JavaScript links in the forecast frame are created in this file to open up any one of the forecast products.

The Perl script, /var/www/cgi-bin/cv_main/forecasts.pl, sets up the format and layout of the forecast table and probability data. The background, table, and font colors are all created by a CSS. Inline JavaScript code is used to open up separate browser windows to display either help pages or forecast summary data. Before displaying any data two files are checked to see if any clouds are in the approach zone. If the CWSU rate report or the automated stratus burn off file has been produced then the files are read to find the time of the burn off.

The MSF Display is setup and reformatted every time the frame is refreshed or reloaded. The approach clear message is formatted first by verifying that either the automated algorithm detected clearing, or the CWSU entered the time of clearing. If the summer stratus season has ended according to the parameter file /ll/cv/web/params/operation_period.pm then the forecast table is replaced with a Winter Weather Mode table. This displays user editable notes about the system. The editable file is located at /ll/cv/web/params/webdisplay_mode.pm. While in the winter weather mode the forecast is no longer running so the forecast table and all the links are not necessary. The CWSU rate change button is also removed since it is not be necessary to enter the rate reports during the winter season. If it is determined that the CWSU users want to continue seeing the rate change button during the winter this can be changed using a parameter setting in the file /ll/cv/web/params/display_format.pm. During the summer season, if the CWSU have not entered the rate change by a certain time of day then the words on the button begin to scroll. The default time is set to 19:30 GMT but can be changed by editing a variable in the file /ll/cv/web/params/forecasts.pm. This is done as a reminder to the CWSU users to enter the rate change that occurred during the day.

4.1.4.4 Data Quality and Diagnostics

The MSF Display reads in the consensus forecast file every minute to verify that the latest data will always be displayed. All the data is checked for valid dates, times and forecast times. If the date of the data does not match today’s date then the forecast values are set to “N/A.” If an invalid forecast time is created and placed in the file then it would be checked and corrected before being displayed. An invalid time is a time value that is not within the 00:00 GMT to 23:59 GMT period.

The probability of stratus burn off file, /ll/cv/web/params/prob_tables/prob_HHHH.txt, is opened based on the hour of the day and the four probability values are read from the file based on the forecast time. This file is not read if valid forecast was not found.

4.1.4.5 Possible Maintenance Tasks

There are no major maintenance tasks required for the forecast frame. The operational period parameter can be edited to set the start and stop dates of the summer and winter operational modes. During the winter weather mode there is a file that can be edited to include notes about the status of the sensors and the MSF Display.

4.2 OBSERVATION PROCESSES

4.2.1 Overview

All of the products can be viewed on the MSF Display observation frame. The section that follows will explain the function of each of the links that appear on the observation frame. Also these observation products can be selected from the Review Page. An overview of the processing of the observation data was discussed in Section 3.2.

4.2.2 Surface Observations

4.2.2.1 Description

All the surface observations that are received and processed by the base station computer can be viewed from the observation frame by clicking on the link "More Obs" or by clicking the "Regional Surface Obs" link on the Review Page. The link launches a script, /var/www/cgi-bin/cv_support/regionalobs.pl, which will format the latest observations from all the stations into a table. When the page opens, links are created so that if an individual station name within the table is clicked, it will bring up a list of all the observations of that site for the entire day.

4.2.2.2 Input Data

In Table 3-2 the directory structure listed as "obs" is the location of all the processed surface observation. If data are unavailable for a selected station then every field in the surface observation is filled with "N/A" by the script.

4.2.2.3 Processing Details

This script reads a file which contains pre-processed surface observations. Changing this format within the Surface Observation Generator (Section 3.2.5) will cause the data to be displayed in the wrong columns in the table. When this script is launched the file and image names are read in and stored. If a station id was not read in from the obs directory in /ll/cv/web/data/YYYYMMDD/obs then it will not appear in the table on the display.

JavaScript commands are created that will open a new browser window that contains a list of all the observation of individually selected stations for the entire day. This script, /var/www/cgi-bin/cv_support/obsserv.pl, formats the observation list into a table. The station name that appears above the observation table is written within this script. In addition to the raw observations, the script also computes and displays the pressure difference from SFO, shown in the far right column of the table.

4.2.2.4 Data Quality and Diagnostics

The observation data are checked every five minutes and the current time is compared to the observation time in order to display only the most current data. If observation data for a selected site have become older than forty minutes then all the observation fields for that station will timeout and display "N/A." If the data do time out, it is most likely that the sensor stopped sending data. Some sensors do not send data during the overnight hours.

4.2.2.5 Possible Maintenance Tasks

There are no major maintenance tasks for these scripts. Of course, if the Surface Observation Generator is ever changed and that changes the output of the data then a change will be required for the scripts that read and format the data for the display.

4.2.3 Sodar

4.2.3.1 Description

The sodar data are processed and a six hour GIF image is created for all the field sites. These images are viewed on both the MSF Display and the review page. Clicking the link “Full Page Sodar” on the display page launches the script /var/www/cgi-bin/cv_support/sodars.pl. This script switches the observation frame to only view larger sodar images. On the review page the script is launched when the link “Full Day Sodar Page” is clicked. A new browser window will appear that displays four six-hour sodar images for each of the field sites. These images represent the sodar scans for the full day.

4.2.3.2 Input Data

In Table 3-2 the directory structure listed as “sodar” is the location of all the processed sodar images. Since the display page cannot access the image data directly from the file location, symbolic links are placed in the directory /var/www/html/cv/data/. If these links are removed then the image data would become inaccessible.

4.2.3.3 Processing Details

The script initializes the file and image names that will be required upon being launched. When the script is launched from the MSF Display the two current sodar images are requested to be displayed. When the script is launched from the review page the six sodar images for each field site are displayed in a new browser window.

4.2.3.4 Data Quality and Diagnostics

Every five minutes the page is refreshed in order to be sure the most recent data is displayed. The sodar.pl script does not timeout the data like the regional_obs.pl script does. If the requested image is not available a default image is displayed that states the “Sodar Image is Unavailable.” If data were coming in but stopped during the day then the last available sodar image will be displayed until a new one replaces it. If the image is unavailable it could be due to no new data being sent by the field sites or the sodar image generator is no longer running.

4.2.3.5 Possible Maintenance Tasks

Unless a change of file names is performed in the sodar image generator then no maintenance is required.

4.2.4 Satellite Imagery

4.2.4.1 Description

After the satellite data are processed they are stored for access by the MSF Display. The satellite image appears on both the display and review page. When the satellite image is clicked on the display or review page a GIF Image Animation Controller appears in a new web browser.

4.2.4.2 Input Data

In Table 3-2 the directory structure listed as “sat” is the location of all the processed Satellite images. Since the display page can not access the image data directly from the file location, symbolic links are placed in the directory /var/www/html/cv/data/. If these links are removed then the image data would become inaccessible.

4.2.4.3 Processing Details

When the animation controller is launched, the script /var/www/cgi-bin/cv_support/jsmovie.pl initializes file and image names. The parameter file /ll/cv/web/params/jsparam/ll_sat.pm searches the dated directory for image files that fit the satellite regular expression. When the script is launched a new web browser displays the GIF animation controller containing the satellite image.

4.2.4.4 Data Quality and Diagnostics

Only the most recent data will be displayed because the parameter file searches the specified dated directories for the satellite data. The satellite images are produced from data that arrive every 15 minutes. Occasionally not all 15 minute satellite images are available. This occurs if the satellite data stop coming in or if two 15 minute satellite scans come in simultaneously. If it is before sunrise or if the satellite data feed stops sending data for the entire day then the unavailable image will appear stating “Satellite Image available after sunrise.”

4.2.4.5 Possible Maintenance Tasks

If the names of the satellite image files are changed in the Satellite image generator then the parameter file would need to be changed to search for the images. Otherwise, no maintenance is required.

4.2.5 Solar Radiation

4.2.5.1 Description

The solar radiation data are processed and a GIF image is created that illustrates the diurnal cycle of the incoming solar radiation measured at the field site. The images can be viewed on the MSF Display and the review page. When the script /var/www/cgi-bin/cv_support/radiation.pl is launched by clicking the link “Full Page Radiation” on the display page, the observation frame switches the view to only see larger solar radiation images. When the script is launched from the review page by clicking the link “Full Page Radiation,” a new browser window will appear with larger solar radiation images.

4.2.5.2 Input Data

In Table 3-2 the directory structure listed as "pyro" is the location of all the processed solar radiation images. Since the display page can not access the image data directly from the file location, symbolic links are placed in the directory /var/www/html/cv/data/. If these links are removed then the image data would become inaccessible.

4.2.5.3 Processing Details

When this script is launched the file /ll/cv/web/params/observation.pm is read to initialize the file and image names. When the script is launched from either the MSF Display or the review page the image is requested to be displayed.

4.2.5.4 Data Quality and Diagnostics

Every five minutes the page is refreshed in order to ensure that the most recent data are displayed. Unlike the regional_obs.pl script, the radiation.pl script does not time out the data. If the requested image is not available before 15:00GMT, then the message will state, "Solar Radiation Data Available After Sunrise." If it is after sunrise and the image is still unavailable then the default image will display, "Solar Radiation Data Unavailable." If the data were coming in but stopped, then the last available solar radiation image is displayed until a new image replaces it. In the event that solar radiation data are not available, it could be due to either a lack of new data coming from the field sites or the solar radiation image generator no longer running.

4.2.5.5 Possible Maintenance Tasks

Unless a change of file names is performed in the solar radiation image generator then no maintenance is required.

4.3 FORECAST PROCESSES

4.3.1 Overview

All of the products found on the forecast frame are links to data that when clicked will open a new browser window to display the requested data. Each of these scripts read text data from various files without requiring locks. All the data read in by these scripts are parsed and re-formatted into an easy-to-read table format. This section gives a brief description and processing details of each of these scripts. Appendix L details the model runtime parameters used when displaying model data in a table view. An overview of the forecast frame was discussed in Section 5.1.4.

4.3.2 Forecast Summary

4.3.2.1 Description

The automated forecast algorithm produces a guidance forecast throughout the stratus season when it is determined that there is stratus in the SFO bay area. Every forecast that is produced is saved in a dated directory. The link "Hourly Forecast Summary" launches the script /var/www/cgi-bin/cv_support/fcstsummary.pl which opens a new browser window and displays the hourly forecast summary in a table format. The link in the drop-down list below the forecast table called "Detailed Forecast Summary" launches the script /var/www/cgi-bin/cv_support/fcstsummary.pl which opens a new browser window and displays the detailed forecast summary if the argument, "?sumtype=Detailed" is included. The detailed forecast contains a variety of data formatted into a table.

4.3.2.2 Input Data

The forecast summary tables read data from several locations to produce its output. The forecasts are stored at /ll/cv/data/ProductsArchive/YYYYMMDD/ConsensusForecast and the naming format appears as YYYYMMDD_HHMMSS__latest_consensus.dat. The preceding date format is the year, month, day, hour, minute and second of the forecast run. If they exist, the CWSU rate change report (/ll/cv/web/data/YYYYMMDD/sideby_times.YYYYMMDD) and the automated burnoff algorithm output (/ll/cv/data/cvRTCache/LATEST/burnoff_occurred) are read and stored. The Probability of Clearing table reads data from the directory /ll/cv/web/params/prob_tables.

4.3.2.3 Processing Details

After all the data are read in and stored, a determination is made as to whether the forecasts should be shown. If there was a VAPS day, which is determined by the CWSU, then the forecast data are ignored. When the CWSU does enter a rate change report or when the automated algorithm presumes that a stratus burnoff has occurred, all the forecasts up to the time of the rate change or the presumed time of burnoff are shown. If this script is run to review a previous day of forecast data the "burnoff occurred" flag is ignored but the CWSU rate change report time is used to show all forecasts generated up to that time. If an all-day IFR (i.e. no burnoff) was reported then all the available forecasts for the entire day are shown.

There are two forecast summary tables that can be displayed. The hourly forecast summary displays a table of the forecast time for each model for each hour. The detailed forecast summary displays a table showing the run time of the model, the model confidence, the probability of clearing, the model forecast time and the model weight used to determine the consensus forecast.

4.3.2.4 Data Quality and Diagnostics

Occasionally a model forecast may not be available or might not be shown. If a model forecast time was not available for a certain model run time then an "N/A" will be shown. If the VAPS report is issued then all the model fields are filled with the "N/A." If the CWSU issued a rate change report then all the models after the time of the reported burnoff will be shown as "N/A." When the models do not run at all and the script is executed then the browser window will open stating that "Forecast Guidance is Unavailable."

4.3.2.5 Possible Maintenance Tasks

The script that produces the forecast summaries does not require any maintenance unless the format of the consensus forecast file, the "burnoff occurred" file, or the CWSU rate change file is modified.

4.3.3 Raw Model Output Summary

4.3.3.1 Description

The forecast model guidance for each model computes a likely stratus burnoff time every four to five minutes as described in Section 3.3. The forecast model output will change slightly for a single hour due to the arrival of more current data. The script /var/www/cgi-bin/cv_support/model_detail.pl will launch a new web browser window that will display each of the model's guidance forecast times for every time that the model forecast output changed.

4.3.3.2 Input Data

The raw model summary table reads data from several sources. All the model forecast guidance results are stored in the directory /ll/cv/data/daysumYYYY/fcsts/YYMMDD/. The file format appears as "fcsts_modelName.YYMMDD.sum" where modelName is the name of the models, such as "Local," "Regional," etc. If it exists the CWSU rate change report, /ll/cv/web/data/YYYYMMDD/sideby_ times.YYYYMMDD, and the automated burnoff algorithm, /ll/cv/data/cvRTCache/LATEST/burnoff_ occurred, are read and stored.

4.3.3.3 Processing Details

After the file's existence above have been determined then the data is stored. All the forecast data are temporarily saved in a file after being concatenated and sorted by the run time of the forecast model. Then the data in the temporary file are stored and compared to determine if the previously read data element is not exactly the same as the current data element being read. If the data element was not read before then the element is stored in an array. This ensures that only the data from each model that has changed will be displayed.

If the rate change report or burnoff file exists then the raw model output will be suppressed based on the report type or the time of reported burnoff. A "VAPS" report will suppress the entire list of raw model data. A rate report indicating the time of reported stratus burnoff would suppress any raw model data created after this time.

4.3.3.4 Data Quality and Diagnostics

The raw model forecast guidance may not always be available either due to a lack of data to process the models or the lack of stratus clouds in the approach zone. When this occurs the table created when the script is implemented will show red dashed in the location of the missing forecast guidance data. If it was a VAPS day then all the models get suppressed regardless of whether models processed data. If none of the models ran for the day then the window will open stating "Forecast Unavailable."

4.3.3.5 Possible Maintenance Tasks

This script requires no maintenance but if the format of the forecast data in the daysum directory is modified, the script will produce unexpected results.

4.3.4 Key Model Parameters

4.3.4.1 Description

Every time the model guidance is processed, the forecast model parameters are written to a file. These files are stored and dated for every day the models run. The files can be viewed using the MSF Display or the review page during the summer stratus season. When the link "Local SFM," "Regional SFM," or "Satellite SFM" found in the drop-down list below the forecast table is clicked the script /var/www/cgi-bin/cv_support/ModelKey.pl will be launched in a new browser window.

4.3.4.2 Input Data

The key parameters are stored in the directory /ll/cv/data/cvRTCache/LATEST. Each model has a file name formatted as "modelname.KeyParams.YYYYMMDD." There is one file for each model for each day that the models run.

4.3.4.3 Processing Details

This script reads the file and reformats the data into a table.

4.3.4.4 Data Quality and Diagnostics

It is possible that the input file could be read and written simultaneously. If data are still being written to the input data file by the model forecast processes, the MSF Display will wait until the file write is completely before it reads the file. If the model parameter file has not been created the browser window will open stating "Model Key Parameter Unavailable." If the output of this file is changed, the script will be affected and could require modification in order to read the new file format.

4.3.4.5 Possible Maintenance Tasks

Every time the model forecast guidance process runs the model parameters for each model will be stored in the directory. These model parameters will begin to accumulate over the course of the summer stratus season. It is important that this directory be trimmed occasionally to prevent filling the disk. If these files are removed the review page will no longer be able to display the model parameters. See Appendix E about file trimming.

4.3.5 CWSU Rate Change Form

4.3.5.1 Description

The CWSU can enter the time change of the AAR each day. The button to open the link is only visible on the CWSU display computer. When the button is pressed the script, /var/www/cgi-bin/cv_support/verify_form.pl, will be launched in a new web browser window. If the link is launched from any other computer, a message appears stating that only the CWSU has the authorization to enter the rate change report.

4.3.5.2 Input Data

If the CWSU has entered a rate change report for the day then this rate report serves as input data to this form. The file /ll/cv/web/data/YYYYMMDD/sideby_times.YYYYMMDD is opened and all the data in the file is parsed to refill the form data so if the CWSU wants to edit the report, the entire form will not have to be filled in again. If the rate report was not entered before then a blank form appears.

4.3.5.3 Output Data

When the CWSU enters a rate change report a new file is created. This new file, /ll/cv/web/data/YYYYMMDD/sideby_times.YYYYMMDD, contains the rate change for the day as well as any comments the CWSU forecaster wants to enter.

4.3.5.4 Processing Details

When the script is launched it verifies that it was the CWSU that is trying to enter the rate report. If someone other than the CWSU opens the rate change report a message is displayed stating that only the CWSU is authorized to enter the rate change report. This access limitation is controlled by the parameter file /ll/cv/web/params/host_computer.pm discussed in Appendix K. The IP address of the CWSU display computer is entered into the parameter file in order to confirm that only this computer will be allowed to enter the rate report. When a member of the CWSU opens the rate change form the script checks to see if the rate change was previously entered. If the rate change was entered then the file is read and the required information is parsed from the file. This data is used to fill in the JavaScript form and a notice will appear below that states the rate change was entered for the selected date. When the CWSU submits the rate change report, a summary of the report is emailed to the display and system maintainers to verify that the report was entered from the correct IP address. The email address list parameter file is discussed in Appendix J. Any unauthorized rate change report that was entered should be corrected and investigated to see who entered the report. This can be done by checking the access logs or by searching for the owner of the IP address that was listed in the email.

4.3.5.5 Data Quality and Diagnostics

When the “Submit” button is clicked, the form checks to see if a logical rate report was entered by the CWSU. For example, if the CWSU entered a time that was not in the stated format or was not within a valid time range then the report is not accepted and a request is made to fix the error. If the CWSU selected the radio button VAPS all day and then entered a rate change time, the form will again refuse to enter the report since a rate change does not happen if there was a declaration of VAPS all day. The same logic is applied when the radio button IFR all day is selected. If the “Close” button is clicked before the rate report was submitted then a pop-up window appears stating that the rate report was not entered and asks if the CWSU forecaster wishes to close the form. If the CWSU forecaster clicks the button “OK” in the pop-up window the form will close but if the button “Cancel” is clicked the pop-up window will disappear and the form will continue being displayed.

4.3.5.6 Possible Maintenance Tasks

There is no major maintenance required for this script. If the output of this script is ever modified then it will affect the rate change summary file and the model performance algorithm. Maintenance will be necessary if the IP address of the CWSU display computer is changed because the rate change button will only appear on a computer with the selected IP address during the summer stratus season. This can be corrected by editing the parameter file /ll/cv/web/params/host_computer.pm and changing the IP address listed in the file to the new IP address.

4.3.6 Model Performance Summary

4.3.6.1 Description

Models are scored daily for the current stratus season in order to determine the accuracy of the forecast guidance. The model performance scoring process was discussed in Section 4.4.3. The model scores are written to a file and can be viewed by clicking the link “Model Performance Summary.” This link will launch a script, /var/www/cgi-bin/cv_support/modelperf.pl, which will open a new web browser that formats the model scores into a table format.

4.3.6.2 Input Data

The model performance file computed every day is located at /ll/cv/web/data/daily/model_ performance.YYYY. The historical performance file containing the previous year’s performance can be found at /ll/cv/web/params/model_historical_perf.txt

4.3.6.3 Processing Details

The text file contains a median absolute forecast error and a percentage success rate. When the script is launched it stores both forecasts performance scores in different arrays and then outputs these arrays into a table format.

4.3.6.4 Data Quality and Diagnostics

This data is only written at the end of the day from a cron. A working file is created temporarily containing the model performance scores and it is moved to the model performance file. The reason for the creation of a temporary file that is moved into the permanent file format is to avoid using file locks that prevent reading of the file while it is being written. If the file has not been created for the current year or has been removed when the file is requested, the browser window will display “Model Performance Summary Unavailable.” Any major changes to file output format produced by the processor mentioned in Section 4.4.3 will require a change in the file reader.

4.3.6.5 Possible Maintenance Tasks

The script is maintenance free but if the model performance file output format is changed then this script will need to be modified to read the new format.

4.3.7 ETA Analysis Maps

4.3.7.1 Overview

The ETA Analysis Maps is a web page utilizing two frames created with CGI scripts written in Perl and includes Javascript to make the display more dynamic. All of the CGI files located on the base station computer reside in the directory, /var/www/cgi-bin/. The main script, /cv_archive/ArcTop.pl, produces an upper and lower frame in a new web browser when the link, "ETA Analysis Maps," is clicked. This display will show 1 mB ETA analysis maps of the western US centered over California. The default setting is to generate these maps only during the summer stratus season, and not during winter mode.

4.3.7.2 Upper Frame

4.3.7.2.1 Description

The design of the upper frame /var/www/cgi-bin/cv_archive/arc_title.pl was based on the review page title interface that was created. It contains the title text of the display and has drop-down selectors to choose the year, month, day, and hour of the archive data that are available.

4.3.7.2.2 Input Data

In order to assemble a collection of dates, the directory, /ll/cv/web/archive/, is searched to find the dated directories.

4.3.7.2.3 Processing Details

The dates that are collected for the archive directory are broken up into year, month, and day and are stored as arrays. The drop-down selectors use the arrays to determine what months are available for the selected year and what days are available for the selected month. All the processing for the drop-down selection changes occur in JavaScript. The collection of the dated directories occurs in Perl.

4.3.7.2.4 Data Quality and Diagnostics

If a dated directory is found it is assumed that valid data are available in the directory and the date is stored. The dated directory is created only when archive data for the specified date is produced.

4.3.7.3 Lower Frame

4.3.7.3.1 Description

The lower frame, /var/www/cgi-bin/cv_archive/arc_observations.pl, displays images created from the FTP data, discussed in Section 3.5.2, that were collected. When the image is clicked it will open up a 7 day animation of the selected image in a new browser window.

4.3.7.3.2 Input Data

The image data that are displayed on this frame is stored at /ll/cv/web/archive/YYYYMMDD/grads/. The images can only be accessed through symbolic links pointing at the data directory. The symbolic links can be found at /var/www/html/cv/archive/. If these symbolic links are ever removed then the image data would no longer be accessible.

4.3.7.3.3 Processing Details

When this display is first opened the lower frame will display the first available image data from the current date. When any of the image links are clicked the selected image will appear to the right of the table of links.

4.3.7.3.4 Data Quality and Diagnostics

If the display opens using the current date and time and no data is available for that time frame then the header in the lower frame states “Archive data is unavailable.” If no data were available and no images were produced then an unavailable image will appear in the lower frame.

4.3.8 Review Prior Days

4.3.8.1 Overview

The MSF Display was designed with the ability to review previous days of past weather. The review page uses two frames created with CGI scripts written in Perl and includes JavaScript to make the display more dynamic. All of the CGI scripts are located in the directory, /var/www/cgi-bin/. When the link “Review Prior Days” is clicked on the main display page a new web browser will launch the review page script /cv_main/RvTop.pl. The upper frame will include the page title and the date interface. The lower frame will show all the image data and includes links to the text data for the selected date.

4.3.8.2 Upper Frame

4.3.8.2.1 Description

The frame located at the top, /var/www/cgi-bin/cv_main/rv_title.pl contains the title text of the display and includes drop-down menu selectors to choose the year, month, and day of any available review data.

4.3.8.2.2 Input Data

The drop-down selector menus include a selection of dates all of which are created based on available data. The directory, /ll/cv/web/data/, is searched to find all the dated directories and these dates are added to the drop-down selectors.

4.3.8.2.3 Processing Details

The dates that are collected for the data directory are broken up into year, month, and day and are stored as arrays. The drop-down selectors use the arrays to determine what months are available for the selected year and what days are available for the selected month. All the processing for the drop-down selection changes occur in JavaScript. The collection of the dated directories occurs in Perl.

4.3.8.2.4 Data Quality and Diagnostics

If a dated directory is found it is assumed that valid data is available in the directory and the date is stored. The dated directory is created only when past data for the specified date is produced.

4.3.8.3 Lower Frame

4.3.8.3.1 Description

The lower frame, /var/www/cgi-bin/cv_main/rv_observations.pl, displays several images depicting the weather conditions around the SFO Bay area for the selected time frame. If the model forecast guidance was produced then it will also be displayed in the upper left corner of this frame in a table along with the CWSU rate change report. There are links to other data in this frame as well.

4.3.8.3.2 Input Data

Since this display is used for reviewing previous stratus days, all the data is collected from the same locations as where the MSF Display received its data. The review page gathers image data from the "pyro," "satellite" and "sodar" directories listed in Table 3. The forecast summary table uses data from the same locations discussed in Section 4.3.2. Since the review page can not access the image data directly from the file location, symbolic links are placed in the directory /var/www/html/cv/data/. If these links are ever removed then the image data would become inaccessible.

4.3.8.3.3 Processing Details

The review page uses the same data processing techniques as the MSF Display frames and the Forecast Products. The lower frame uses a parameter file located at /ll/cv/web/params/rv_observations.pl. This parameter file initializes file names, and image sizes that will appear on the display. The forecast summary table located in the upper left corner of the lower frame implements similar logic from both the "Rate Change Times" in Section 4.3.8 and the "Forecast Summary" in Section 4.3.2. The sodar images that appear below the forecast summary table will usually be the 12:00 to 18:00 GMT scan if available. If that scan is not available then the last sodar image produced for the day is shown, otherwise, "Image Unavailable" is shown. The last solar radiation images created for the day are placed to the right of the sodar images but if these are unavailable "Image Unavailable" is shown in that location. The first satellite image created for the day is also shown to the right of the solar radiation images.

All the links on the page open a new web browser that will show more detailed data. The "Full Day Sodar Page" link was discussed in the Sodar Product section, Section 4.2.3. The link "Full Page Radiation" was discussed in the Solar Radiation Product section, Section 4.2.5. The satellite image can be animated from this frame and the animation controls were discussed in the Satellite Imagery section, Section 4.2.4. The forecast links below the satellite image are similar to the forecast links on the forecast frame. Each of these forecast links were discussed in the Forecast Products section, Section 4.3.

4.3.8.3.4 Data Quality and Diagnostics

If the display opens using the current date and time and no data is available for that time frame then the header in the lower frame states “Review data is unavailable.” If any of the image products were not produced then “Image Unavailable” will appear in place of the missing image.

4.3.9 Rate Change Times

4.3.9.1 Description

The rate changes entered by the CWSU are placed into a summary file to easily review the entire year of sideby times. The process that produces the file was discussed in Section 3.4.2. Clicking on the link “YYYY Rate-Change Times” will launch the script, /var/www/cgi-bin/cv_support/sideby.pl, which will open a new browser window that formats the rate changes into a table format.

4.3.9.2 Input Data

The rate change file which is re-written every time the CWSU enters a new AAR is located at /ll/cv/web/data/daily/burnoff_times.YYYY.

4.3.9.3 Processing Details

This script stores the rate change information into arrays and outputs the data into a table format. The word “More” appears in the table for each day that the CWSU entered text comments. When the mouse cursor is positioned over the word “More,” the written comments hidden in an invisible layer become visible. The date in the date column is a clickable link that when clicked will open the review page for the selected date.

4.3.9.4 Data Quality and Diagnostics

The rate change report is created during the stratus season after the CWSU issues a rate change time. If the table is requested before the stratus season begins or the file being requested has not been created yet the web browser opens stating that “Rate change report not available.”

4.3.9.5 Possible Maintenance Tasks

The script is maintenance free but if the content of the burn off file is changed then this script will need to be modified to read the new format.

4.3.10 System Status

4.3.10.1 Description

It is important to know the timeliness of the data observed on the MSF Display. All the data that feeds the display system is checked to verify its current age. The process that produces this status file was discussed in Section 3.4.4. Clicking on the link "System Status" will launch the script, /var/www/cgi-bin/cv_support/status.pl, which will open a new browser window that formats the age of the data into a table format.

4.3.10.2 Input Data

The tab delimited text file is stored at /ll/cv/web/data/daily/system_status.

4.3.10.3 Processing Details

This script will store the status of the observation data into arrays and outputs the data in a table format. The title names of all the data are stored in this file.

4.3.10.4 Data Quality and Diagnostics

If the system status file is removed or has not been created then the web browser opens stating "System Status not available."

4.3.10.5 Possible Maintenance Tasks

The script is maintenance free but if the content of the system status file is changed then this script will need to be modified to read the new format.

APPENDIX A
LONG TERM MAINTENANCE TASKS

This appendix is a compilation of all the possible maintenance tasks identified in this document, presented sequentially as they appear in the main body of the text. Typically, these tasks will only need to be performed infrequently over a period of one or more years. They fall into two general categories: 1) the need to remove files (observation files, log files, etc.) from the system disk that continuously grow in size, and 2) the need to edit files or scripts in response to external changes that may influence data formats or processing requirements. Each subsection is labeled and numbered as it appears in the main body of this document.

MSF Data Processing Overview

Section 3.2 Observation Data

All the observation data are stored in dated directories and they will accumulate over the course of the year. It will be necessary to purge any older data after a certain period of time in order to maintain at least one year's worth of data that can be displayed on the review page. The directory trimming can be performed manually or automatically through a script or a cron. Appendix E describes the directory trimming script in detail.

Section 3.4.2. Rate Changes

Every year a new file is created that contains the full year's worth of CWSU rate changes. The sideby time report is date stamped and is placed in /ll/cv/web/data/daily. This file is only about 6.5KB but the old files are never removed from the directory. Therefore, these files will continue to build up in this directory and will eventually need to be removed.

Section 3.4.3. Model Performance

Every year a new file is created that contains the model performance scores for the entire stratus season. The model performance file is stamped with the current year. This file is very small and will only occupy about 1.2KB. These model performance files will accumulate in this directory year after year and will eventually need to be removed.

Section 3.4.5. Apache Web Access Logs

Both the Apache log files and the HTML monthly summary files are never deleted from the base station computer. Each HTML monthly summary file occupies only 165KB on average but the Apache access log files occupy about 2 to 3 MB for each day.

Section 3.5.2. Analysis Maps

All the GIF images that are produced by this process are stored in a dated directory. The data will eventually need to be pruned to prevent filling the disk. Directory trimming can be performed manually or automatically through a script or a cron. Appendix E describes the directory trimming script in detail.

MSF Display Processing/Performance Overview

Section 4.1.3. Observation Frame

The observation frame is mostly maintenance free with a few minor exceptions. Changes in web browser technology and web browser standards have altered the interaction of the display over the last few years. Minor adjustments were required in the past for older HTML tags that became obsolete and where no longer used or supported. The GIF animation JavaScript software works differently on older versions of Internet Explorer and Netscape than it does on the newer versions. This required the use of different coding standards based on the type of browser software being used.

Section 4.1.4. Forecast Frame

There are no major maintenance tasks required for the forecast frame. The operational period parameter can be edited to set the start and stop dates of the summer and winter operational modes. During the winter weather mode there is a file that can be edited to include notes about the status of the sensors and the MSF Display.

Section 4.2.2. Surface Observation

There are no major maintenance tasks for these scripts. Of course, if the Surface Observation Generator is ever changed and that changes the output of the data then a change will be required for the scripts that read and format the data for the display.

Section 4.2.3. Sodar

Unless a change of file names is performed in the sodar image generator then no maintenance is required.

Section 4.2.4. Satellite Imagery

If the names of the satellite image files are changed in the Satellite image generator then the parameter file would need to be changed to search for the images. Otherwise, no maintenance is required.

Section 4.2.5. Solar Radiation

Unless a change of file names is performed in the solar radiation image generator then no maintenance is required.

Section 4.3.2. Forecast Summary

The script that produces the forecast summaries does not require any maintenance unless the format of the consensus forecast file, the “burnoff occurred” file, or the CWSU rate change file is modified.

Section 4.3.3. Raw Model Output Summary

This script requires no maintenance but if the format of the forecast data in the daysum directory is modified, the script will produce unexpected results.

Section 4.3.4. Key Model Parameters

Every time the model forecast guidance process runs the model parameters for each model will be stored in the directory. These model parameters will begin to accumulate over the course of the summer stratus season. It is important that this directory be trimmed occasionally to prevent filling the disk. If these files are removed the review page will no longer be able to display the model parameters. See Appendix E about file trimming.

Section 4.3.5. CWSU Rate Change Form

There is no major maintenance required for this script. If the output of this script is ever modified then it will affect the rate change summary file and the model performance algorithm. Maintenance will be necessary if the IP address of the CWSU display computer is changed because the rate change button will only appear on a computer with the selected IP address during the summer stratus season. This can be corrected by editing the parameter file /ll/cv/web/params/host_computer.pm and changing the IP address listed in the file to the new IP address.

Section 4.3.6. Model Performance Summary

The script is maintenance free but if the model performance file output format is changed then this script will need to be modified to read the new format.

Section 4.3.9. Rate Change Times

The script is maintenance free but if the content of the burn off file is changed then this script will need to be modified to read the new format.

Section 4.3.10. System Status

The script is maintenance free but if the content of the system status file is changed then this script will need to be modified to read the new format.

APPENDIX B
APACE PASSWORD ACCESS

This section will not cover specifics about how the Apache server works or about specific configuration changes. It will only discuss how to set up the user name and password access. If you do have specific questions about the Apache server see the documentation that came with the system or go to an online web site. Apache is free and the internet has many references about configuring and enhancing the Apache server.

Included in this appendix will be how to add a new user name and password into the system and how to change the password of the default user name, SFO. There is also a section that describes how to add user DNS names and IP addresses so that a user name and password will not be required when viewing the display. You must have the root access of the system to do any of these changes so if you do not have that access please see the system administrator.

Change password or add new user name:

1. First login as the root user of the system.
2. Change to directory /etc/httpd/users/
3. If changing the password for the default user
   Run this script and follow the directions to change the password.
      enter_new_user SFO
4. If adding a new user name
   Run this script and follow the directions to add a new password.
      enter_new_user “new_user_name” “password_filename”

Add DNS names and IP addresses to allow password free access:

1. First login as the root user of the system.
2. Change to directory /etc/httpd/conf.d/
3. Edit the file access.conf
4. Now you need to go to the section that looks like this
   - <Directory /var/www/html/cv>

   then move to the sub section that looks like this
   - deny from all
   - allow from ...

   The allow from sub section contains multiple DNS names as well as IP addresses. This is where you would need to add, change or remove a new IP address.
5. Now move to the section that looks like this
   - <Directory /var/www/cgi-bin>

   then move to the same sub section described above

- deny from all
- allow from ...

and then add the same IP address or DNS name to the allow from section that you did above.

6. Save this file and exit the editor.
7. Now change to directory /usr/sbin/
8. Type this command (command between the quotation marks)
   - “httpd -k restart”
9. This restarts the Apache server and reads in the file that was just edited.

APPENDIX C
SEARCHING FOR IP ADDRESSES IN THE ACCESS LOGS

It may be necessary to occasionally update or add a new user IP address or DNS name to the access configuration file to allow access to the display without entering a user name and password, trusted access. These Apache access logs can be used to create MSF Display access statistics. The processing of the access logs was discussed in Section 4.4.5. Below is a list of instructions of where to find the access log files, how to search through the file and what to search for in the file.

<u>Searching the access log files:</u>

1. Start by going to the directory containing the Apache access log files, /etc/httpd/logs/YYYYMM.

2. Use the more command to search through the Apache access file, YYMMDD_access.log.

3. A line of the access log file starts with the IP address or DNS name that requested data from the web server. A dash follows that address and the user name, SFO, appears after the dash. A trusted user will not need to enter the user name and password and in this case a second dash appears after the first dash.

4. If the user is having a problem accessing the site without entering a user name and password then search the log files with this command:

   cat YYMMDD_access.log | grep " – SFO"

5. This command will display only the users that have needed to access the MSF Display with the user name and password.

6. Using this information it is possible to see what IP addresses and DNS names might be necessary to add to the trusted access list.

APPENDIX D
MANUALLY STOPPING AND STARTING WEBPROCCHECK

The web processes run throughout the day processing incoming data acquired by the cvRTCache. These processes can be stopped manually but the crontab will restart them again. If the processes are stopped as listed below the observation data on the MSF Display will become unavailable but the cvRTCache will continue to collect data from the field sites. Stopping these processes will not prevent the forecast data from being displayed. It will only stop the observation data from being processed.

<u>How to stop webProcCheck:</u>

1. Login as the web root account.

2. Edit the crontab file (crontab –e) and comment out (with # in column 1) the line that reads:

   0,5,10,15,20,25,30,35,40,45,50,55 * * * * /ll/cv/web/scripts/webProcCheck

3. Run this command from the UNIX prompt to stop the web processes.

   /ll/cv/web/scripts/stop_webpage_support

<u>How to start webProcCheck:</u>

1. Login as the web root account.

2. Edit the crontab file that was mentioned above and uncomment that line.

3. It will take a maximum of five minutes for webProcCheck to start but the program can be started immediately by running the script below:

   /ll/cv/web/scripts/webProcCheck

APPENDIX E
DIRECTORY AND FILE TRIMMING

Many of the data directories are allowed to build up over a period of time. This data are created from the "MSF Data Processing" algorithms and are used by both the "MSF Review Page" and the "Analysis Maps." Before performing directory or file trimming, the data should be backed up on an available storage medium, such as a disk or tape. Once any of the directories or files has been trimmed, the data will no longer be retrievable.

/ll/cv/scripts/Misc/trimDirs: the csh script will recursively trim dated directories. The script takes two arguments. The name of the root directory that contains the dated directories and the number of dated directories to keep.

/ll/cv/scripts/Misc/trimFiles.pl: the perl script that trims files from a target directory. The two options used are:

1. –d <root_directory_name>
2. –a <count>

where <root_directory_name> is the name of the directory containing the files to trim and <count> is the number of files to keep.

<u>Trimming Dated Directory</u>

1. Locate the dated directory and determine the number of directories that will be kept. In the case of review data, only the last year's worth of dated directories are normally retained.

2. On the command line type the following command assuming the dated observation data directory requires pruning but keep the latest 100 directories:

   /ll/cv/scripts/Misc/trimDirs /ll/cv/web/data 100

APPENDIX F
MSF DISPLAY IN SNAPSHOT MODE

It is possible to place the MSF Display into a snapshot mode. When the display is in this mode it will act like a review page. The date and time are entered as two separate arguments and if the data is available for that date then it will be displayed on the MSF Display.

<u>How to enter Snapshot Mode:</u>

1. OPEN THE MSF DISPLAY USING THE COMMAND BELOW (ASSUMING THE IP ADDRESS IS CURRENT.)

   http://64.175.253.249/cgi-bin/cv/CvTop.pl

2. ENTER BOTH A DATE AND HOUR ARGUMENT AFTER THE COMMAND SHOWN ABOVE.

   http://64.175.253.249/cgi-bin/cv/CvTop.pl?date=YYYYMMDD&hour=HHMM

3. AFTER ENTERING THE ARGUMENTS THE DISPLAY ACCESSES THE OBSERVATION AND FORECAST DATA FOR THE DATE REQUESTED THAT IS CLOSEST TO THE HOUR THAT WAS SELECTED.

APPENDIX G
MSF DISPLAY IN TEST MODE

The MSF Display was designed with a test mode. The test mode option was created to allow the display developer to see the CWSU rate change button. This button is only seen on the CWSU MSF Display computer, so in order to perform a test of the button and the Rate Change form, the MSF Display must see set to test mode. When the CWSU button is clicked the CWSU rate change form will be displayed exactly as the CWSU forecaster would see it. The difference is that if the maintainer enters a rate report the file, /ll/cv/web/tmp/sideby_times.YYYYMMDD, is written. The official CWSU rate report can be entered or modified by this script if the correct arguments are entered. These features are discussed below:

How to enter Test Mode:

1. Open the MSF Display in test mode by using the command below (assuming the IP address is current.)

   http://64.175.253.249/cgi-bin/cv/CvTop.pl?test=yes

Testing the CWSU Rate change form:

1. After the MSF Display is in test mode click the CWSU rate change button in the forecast frame.

2. When the new browser window opens a blank form will appear if the CWSU has not yet entered a report and a completed form will appear if they did enter the report. The rate report can be edited but when the form is submitted it will not overwrite the original CWSU rate change report and it will not create a new official rate change report.

3. If it is necessary to submit the rate change report from this mode then another argument, “write_access,” must appear as shown below on a single line. Using this argument will allow write access to the official rate change report.

   http://64.175.253.249/cgi-bin/cv_support/verify_form.pl?test=yes&write_access=yes

APPENDIX H
SETTING THE MSF DISPLAY INTO WINTER MODE

Overview:

The forecast guidance products were not designed or trained to predict stratus burn-off during the winter. Therefore, it was decided that the forecast guidance table would not be displayed during the winter season and instead a text table would be shown. This table replaces the forecast table at a set time determined by a start and stop date listed in a parameter file, /ll/cv/web/params/operation_period.pm. When the current date is between the start and stop times of the winter mode, the MSF Display automatically enters the winter mode. This parameter does not control the forecast guidance products from running. It only controls the visibility of the forecast guidance products on the MSF Display and the review display. When the display enters the winter mode the parameter file, /ll/cv/web/params/webdisplay_mode.pm, is used to display information about the MSF Display operational period or any sensor maintenance that may take place.

How to set the winter mode manually:

4. Open the file /ll/cv/web/params/operation_period.pm for editing.

5. Once opened move to the section of the file listed below and edit the month and day that the display should start and end on:

   start_wintr = MMDD
   stop_wintr = MMDD
      MM is the two digit month
      DD is the two digit day

How to disable the winter mode manually:

1. Open the file /ll/cv/web/params/operation_period.pm for editing.

2. Move to the section of the file listed below and set the winter mode to no:

   $winter = no; #yes;

How to change content on Winter Mode text table:

1. Open the file /ll/cv/web/params/webdisplay_mode.pm for editing

2. Move to the section that looks like the item listed below and add the text content that is require. If the text exceeds the width of the table on the MSF Display then it will automatically scroll to the next line.

```
$system_notes[0] = "…";
$system_notes[1] = "…";
```

APPENDIX I
DISPLAY FORMAT PARAMETERS

File Name: display_format.pm

File Location: /ll/cv/web/params

File Description: Parameter file containing the MSF Display formatting for the forecast frame.

File Contents:

```
require '/ll/cv/web/params/operation_period.pm';

#==================================================#
# If the user types in test=yes after the display
# page address as an argument then this parameter
# sets up the display in a test mode.
#==================================================#
  if( param('test') ){ $test_page = 1; }
  else             { $test_page = 0; }

#==================================================#
# This link will allow the CWSU to enter the current rate
# change verification during the winter mode.
#==================================================#
  $viewInwinter = “yes”;
  $start_view   = $stop_wintr;

#==================================================#
# Do you want to show the forecasts in red until
# the limit_fcst time has been reached?
#==================================================#
  if( $test_page ){ $showInred = “no”;  }
  else           { $showInred = "no";  }

#==================================================#
# Do you want to suppress the forecasts after
# burnoff or verification is entered?
#==================================================#
  if( $test_page ){ $suppress = “no”;  }
  else           { $suppress = “yes”; }
```

```
#==================================================#
# Do you want to view the verification
# button on the page?
#==================================================#
 if( $test_page ){ $verif_btn = “yes”; }
 else         { $verif_btn = “no”;  }

#==================================================#
# If we are creating a snapshot of a particular
# day then set snapshot to yes
#==================================================#
 if( param(‘date’) && param(‘hour’) ){
   $snapshot  = “yes”;
   $winter    = “no”;
   $showInred = “no”;
   $suppress  = “no”;
 }
 else{
   $snapshot = “no”;
 }
```

APPENDIX J
E-MAIL ADDRESS PARAMETERS

File Name:	e-mailList.pm
File Location:	/ll/cv/web/params
File Description:	Parameter file containing the e-mail addresses of the MSF Display maintainers and users.

File Contents:

```
#==============================================================#
# The person maintaining the system will receive regularly
# scheduled emails.
#==============================================================#
$maintenance = “kirkm\@ll.mit.edu”;

#==============================================================#
# The people in this list will receive email confirmation
# about the daily airport arrival rate changes and any
# questions or comments made about the display pages.
#
# Keep the recipients’ e-mail addresses within double quotes
# ex.  “davec\@ll.mit.edu kirkm\@ll.mit.edu”
#==============================================================#
$recipients = “davec\@ll.mit.edu kirkm\@ll.mit.edu ivaldi\@ll.mit.edu”;
```

APPENDIX K
HOST COMPUTER PARAMETERS

File Name: host_computer.pm

File Location: /ll/cv/web/params

File Description: Parameter file containing the host computer name and the host computer IP address.

File Note: The IDL processes need to know name of the host computer in order to access license information and the host IP address is necessary in order to show the rate change button on only the CWSU display computer.

File Contents:

```
# This is the CWSU display computer in Freemont, CA
  $host     = "192.168.1.3";

# This is the main computer that the IDL process run on.
  $idl_host = "st1";
```

APPENDIX L
MODEL RUNTIME PARAMETERS

File Name: model_runtime.pm

File Location: /ll/cv/web/params

File Description: Parameter file containing all the forecast model guidance run times and model names. These parameter are used in the MSF Display's forecast frame, in the model performance scoring algorithm and in all the model summary tables.

File Note: These parameters only controls the ability to view these models and model run times on the MSF Display.

File Contents:

```
#---------------------------------------------------
# The array of model run time hours to display
#---------------------------------------------------
  @model_runtime = qw/0900 1100 1300 1500 1600 1700 1800/;

  @derv_models = qw/Consensus/;
  @base_models = qw/COBEL Local Regional Satellite/;

  @model_names = (@derv_models, @base_models);
```

APPENDIX M
SODAR PARAMETER FILE

File Name: sodar_ctl.pm

File Location: /ll/cv/web/params

File Description: The sodar images contain three overlays and these overlays are plotted on the image based on these control flags.

File Note: The operational period parameter file is required in order to only show the overlays during the summer stratus season.

File Contents:

```
require '/ll/cv/web/params/operation_period.pm';

#-------------------------------------------------
# The dates below look like mmdd. Where mm is
# the 2 digit month # and dd is the 2 digit day
#-------------------------------------------------
$show_overlays = $stop_wintr;
$hide_overlays = $start_wintr;

#================================================#
# The choses for what the ovrly should be equal
# to is...
# $inv_cntrl = “inv_yes”; / “inv_no”;
# $zi_cntrl  = “zi_yes”;  / “zi_no”;
# $cig_cntrl = “cig_yes”; / “cig_no”;
#================================================#
$cig_cntrl = “cig_yes”;

if( $mon_day < $show_overlays || $mon_day >= $hide_overlays ){
  $inv_cntrl = “inv_no”;
  $zi_cntrl  = “zi_no”;
}
else{
  $inv_cntrl = “inv_yes”;
  $zi_cntrl  = “zi_yes”;
}
```

GLOSSARY

AC	Approach Clear
AZ	Approach Zone
ACV	Arcata
AWRP	Aviation Weather Research Program
ARTCC	Air Route Traffic Control Center
BKN	Broken
CLR	Clear
CWSU	Central Weather Service Unit
COBEL	Couche Brouillard Eau Liquide
DPA	Dual Parallel Approaches
FAA	Federal Aviation Administration
GMT	Greenwich Mean Time
GOES	Geostationary Operational Environmental Satellite
LSFM	Local Statistical Forecast Model
MIT	Massachusetts Institute of Technology
MSF	Marine Stratus Forecast
NWS	National Weather Service
OAK	Oakland
OVC	Overcast
RSFM	Regional Statistical Forecast Model
SCT	Scattered
SFM	Statistical Forecast Model
SFO	San Francisco International Airport
SMB	San Mateo Bridge
SODAR	Sonic Detection And Ranging
SQLL	San Carlos High Resolution Sensor
SSFM	Satellite Statistical Forecast Model

SFO Marine Stratus Forecast System Documentation

Section V. Forecast Model Descriptions

TABLE OF CONTENTS

Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Page

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS
(Continued)

LIST OF TABLES

LIST OF TABLES

(Continued)

1. INTRODUCTION

1.1 SCOPE

This document provides a description of the forecast algorithms developed for the San Francisco Marine Stratus Forecast System. This includes four component forecast models (Local, Regional, Satellite, and COBEL models), and a Consensus forecast algorithm which integrates the four component model forecasts to produce a single forecast. The document includes both a scientific description of the forecast development approach, as well as details regarding data and file management associated with the real-time model processing.

This introductory section is intended to provide an overview of model development, providing the context in which the models were developed, as well as important considerations that impacted the forecast model approach. A quantitative assessment of model performance is also presented to provide both an absolute and relative measure of expected forecast accuracy of the models, for consideration when using the forecasts as an operational guidance tool. Specific detail associated with each of the individual models is provided in their respective sections within this document (Sections 2 through 6).

1.2 RELATED DOCUMENTS

A companion document entitled "Real-time System" details the operational system software responsible for collecting required forecast model data and activating the component models. Appendix A, entitled "Design and Automation of Statistical Forecast Model Development Procedures," details the automated system developed at Massachusetts Institute of Technology/Lincoln Laboratory (MIT/LL) that was used to generate the three multiple predictor statistical models.

1.3 BACKGROUND

The local airspace surrounding the San Francisco International Airport (SFO) is prone to regular occurrences of low ceiling conditions from May through October due to the intrusion of marine stratus cloudiness along the Pacific coast. The low cloud conditions prohibit dual parallel approaches of aircraft to the airport's closely spaced parallel runways, thus effectively reducing the arrival capacity by a factor of two. The behavior of marine stratus evolves on a daily cycle, filling the San Francisco Bay region overnight, and dissipating during the morning. Often the low ceiling conditions persist throughout the morning hours and interfere with the high rate of air traffic scheduled into SFO from mid-morning to early afternoon. The result is a substantial number of delayed flights into the airport. Air traffic managers face a continual challenge of anticipating available operating capacity so that the demand of incoming planes can be metered to match the availability of arrival slots.

The Central Weather Service Unit (CWSU) at the Oakland Air Route Traffic Control Center (ARTCC), the National Weather Service (NWS) Office in Monterey, and meteorologists at the national System Command and Control Center in Washington, D.C. are responsible for providing weather information to assist air traffic managers anticipate changes in available operating capacity. Meteorologists at commercial airlines are also responsible for providing weather information for the most efficient and cost effective dispatch of airplanes.

In 1995, the FAA's Aviation Weather Research Program (AWRP) launched a "Marine Stratus Initiative" whose objective was to improve cloud forecast guidance for more reliable anticipation of available operating capacity at SFO. MIT Lincoln Laboratory served as technical lead for the project, in collaboration with San Jose State University, the University of Quebec at Montreal, and the CWSU at Oakland Center. The project led to the development of the Marine Stratus Forecast System.

An important element of the system is the generation of automated objective forecast guidance to be used by forecasters in making their recommendation to air traffic managers regarding the timing of cloud dissipation that will allow dual independent aircraft approaches. This guidance is presented to system users in the form of a "consensus forecast" derived from four independent component forecasts. In addition to providing a discrete forecast of the best estimate of transition time to dual approaches, the consensus forecast is also converted to a probabilistic representation showing the likelihood of dual approach availability at specific target forecasts times during the period of highest arrival traffic demand.

1.4 SYSTEM OVERVIEW

1.4.1 Sensor Configuration and Data

The system relies on weather observations from a network of sensors surrounding the Bay region. Data are collected from these sensors at 15-minute intervals, and transferred to a base station computer located at the Air Route Traffic Control Center (ARTCC) in Fremont, CA. There the data are processed for display, and for input into a suite of models designed to forecast the time of Approach Zone clearing. The display of observations and the automated forecast guidance are made available to users via the World Wide Web.

The configuration of sensors which supply data to the forecast models is shown in Figure 1. There are two primary sensor sites located at San Francisco International Airport and San Carlos Airport, along the Approach Zone. These two primary sensor sites each include a Sonic Detection and Range (sodar) instrument to estimate cloud top height, a ceilometer to estimate cloud base height, a pyranometer to measure incoming solar radiation, and instrumentation to provide high resolution measurements of temperature, humidity, and wind.

In addition to data from the primary sites, data supplied to the forecast algorithms also include standard surface weather reports from the central California region, and the vertical profile of temperature, humidity, and wind provided by the 00 GMT and 12 GMT daily balloon soundings from Oakland.

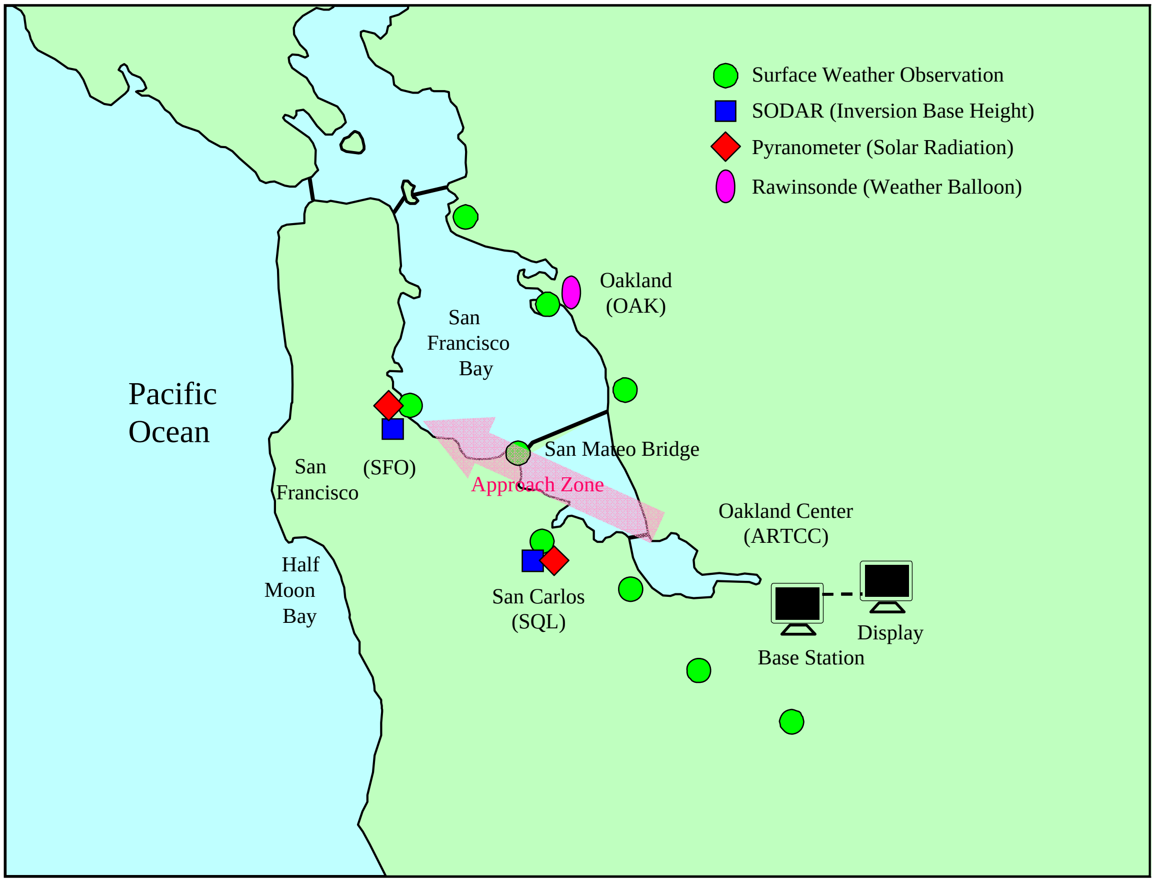


Figure 1. SF Bay area sensor locations.

1.4.2 Forecast Processing

Data collected at the base station computer are supplied to four "component" forecast models, each designed to independently provide an estimate of the time at which the approach zone is expected to be sufficiently free of clouds to allow parallel runway approaches into SFO. Individual forecasts from the four component models are combined to yield a single "Consensus" forecast. The forecast process is run on a daily cycle, beginning at 09 GMT each morning (2 AM Pacific Daylight Time). The models are initialized with new data and run at the top of the hour at 09, 11, 13, 15, 16, 17, and 18 GMT throughout the morning, as long as cloud in present in the approach zone. Once the approach zone has cleared, the models cease to run.

Each of the four forecast models has a specific set of input data required for execution. The models will optimally run using observations from the top of the current hour; however, there is flexibility to use data recorded within 15-minutes from the top of the hour. This flexibility allows the models to run in the event that the optimal top-of-the-hour observations are unavailable. In particular, this accommodates potential missing observations in the visible satellite image which is routinely available at 15-minute intervals, and the subset of surface observations that are available at 5-minute intervals.

Beginning at the top of each forecast run hour, each model will execute provided that its required data are available. Once all of the models have received their data and have executed, the Consensus algorithm is prompted to generate a consensus forecast, followed by a posting of all forecasts to the display. The display is refreshed with new forecasts only once per hour, i.e., the individual models are not allowed to update asynchronously to the display within a given hour.

Typically, all of the top-of-the-hour data are received and processed by the forecasts models by 20-25 minutes past the top of the hour, resulting in a refresh of forecasts to the display. However, during some hours, the receipt of data is delayed, or may not occur at all. The system will wait up to 35 minutes (a parameter setting within the system) past the top of the hour for new data to arrive. Once this waiting period has expired, a consensus forecast will be generated using whatever models are available, and post the new forecasts to the display, even if one or more of the component forecasts was not able to execute.

1.5 FORECAST MODEL OVERVIEW

The four component models consist of three statistically derived forecast models, and a physics-based numerical model:

1. Local Statistical Forecast Model (LSFM)
2. Regional Statistical Forecast Model (RSFM)
3. Satellite Statistical Forecast Model (SSFM)
4. COBEL

The models are quasi-independent in that they rely on significantly different data observations as input. The benefits of this are threefold. First, it provides four different perspectives on the forecast, allowing the opportunity for either inter-model consistency or variability, which is a useful indicator to the forecaster as to the reliability of the current forecast. Second, the use of a multiple forecast/consensus approach allows the forecast system to remain available in the event that one or more data sources become unavailable. The consensus forecast algorithm is designed to effectively combine the available component forecasts using a weighting system that is based on the relative historical performance of each forecast model at each model run hour. Third, the consensus approach tends to limit the hour-to-hour variance that is sometimes observed in the individual component forecasts.

1.5.1 The Statistical Forecast Models

Each of the three statistical forecast models (SFMs) was developed using essentially the same methodology, with the primary difference being the nature of the input data. A detailed description of this process is provided in Appendix A. A brief description of the process is presented here.

In general, the forecast development process uses statistical regression to generate a set of forecast equations based on the selection of the appropriate set of predictors and coefficients to arrive at an effective forecast model equation for estimating a predictand. This process was performed separately for each model (Local, Regional, and Satellite) and for each model run hour (09, 11, 13, 15, 16, 17, and 18 GMT). For the purposes of the SFO Marine Stratus Forecast system, the predictand was defined to be the time (on a given day that stratus is impacting the Approach Zone) that dual independent approaches become available at SFO. This is an important distinction from defining the predictand as a meteorological parameter; as it is an attempt to directly predict the operation-impacting event that is the prime motivation for system development. In most instances, this operational event is closely tied to one or more meteorological events, most notably the time that the ceilometer at San Mateo Bridge indicates no ceiling. However, the dual independent approach event may also be influenced by other meteorological or operational circumstances. The regression analysis uses the daily record (dating back more than 15 years) of the daily transition time from single to dual approaches at SFO. This is commonly referred to as the “side-by” time, and corresponds to the time at which there is an increase in the effective aircraft arrival rate.

All of the statistical forecast equations are of the general form:

$$T = \text{SunriseTime} + \text{Constant} + a_1X_1 + a_2X_2 + \ldots a_nX_n$$

where T = Time of transition to dual approaches (i.e., side-by time)
X = predictor
a = predictor coefficient weighting

Note the dependence of the prediction time on the time of sunrise time each day which reaches its minimum (earliest) at the summer solstice and increases each day for the remainder of the summer stratus season.

1.5.1.1 Determining Candidate Predictor Equations

The statistical forecast development process chooses an appropriate set of predictors and coefficient from a large set of candidate predictors. Initially, this entire process was performed manually. Once the process was refined and stabilized, then it was automated to allow handling of large data sets and many iterations of test runs. The essence of this process can be viewed as the 5-step sequence shown schematically in Figure 2.

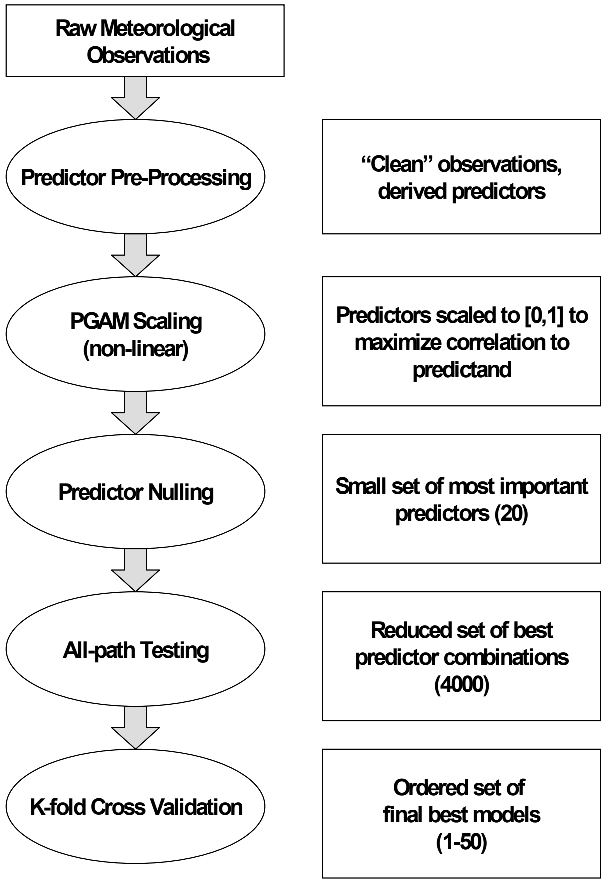


Figure 2. Primary steps of Statistical Forecast Development Process.

The development process begins with a large set of potential predictors, which undergo a pre-processing step from their raw form. First, the data are quality checked to deal with erroneous or missing data. Second, the data may be processed to yield a "derived" predictor from the raw data. An example of this might be using a time-difference value of a parameter as a potential predictor.

The second step is a non-linear re-scaling of the candidate predictors using a methodology referred to as Pre-scaled Generalized Additive Modeling (PGAM). Details are provided in Appendix A. Basically, this process examines the correlation of each candidate predictor with the predictand, and re-scales its value non-linearly to a value between 0 and 1, in an attempt to expose the most sensitive range of values correlating to the predictand. For example, the height of a cloud ceiling may be an important predictor; however, any height above 3000 feet may have very similar "predictive" meaning, whereas values between 0 and 3000 feet may be very sensitive with regard to their correlation with the predictand. The values would correspondingly be scaled to allow more sensitivity in the lower range of values, but treat all of the higher range values as virtually equivalent.

Once the candidate predictors have been re-scaled, they then undergo a process referred to as “predictor nulling,” which is an attempt to reduce the large set of predictors down to a more computationally manageable number suitable for equation development (typically on the order of 20 or so predictors). Many of the predictors are redundant in that their relationship to the predictand is very similar. The nulling process examines the predictor set for these redundancies, and reduces the set to the most promising candidates.

The fourth step in the process is to examine all combinations of the 20 best candidate predictors. This process is referred to as “All-Path Testing,” which applies a least squares error test on a training portion of the data set to yield the best set of candidate prediction equations containing from 1 to 20 predictors.

Finally, the candidate equations undergo a multi-fold cross validation procedure to provide a rank-order list of the predictor equations yielding the least error on quasi-independent data sets. The “quasi” independence is an artifact of the limited data availability, wherein the test data set is randomly apportioned into difference combinations of subsets to provide multiple independent assessments of performance.

1.5.1.2 Selecting Final Predictor Equations

The process described in the previous section describes the automated process for distilling a large number of predictors into a smaller, rank-ordered set of candidate prediction equations. This rank-order list is then used more subjectively to select the final form of the forecast equation.

The rank-order list essentially views the quality of the candidate equation in terms of minimizing least-squares error. From the thousands of “best” possible candidates, the top group (say a few equations to a few dozen equations) is roughly equivalent in terms of the error metric. The final selection process entails examining the components of these best candidate equations. Considerations are made with regard to the nature of the selected predictors, in terms of both their scientific integrity, and equally important, their data availability. With regard to the former, favor is given to those equations whose component predictors are most consistent with the intuitive correlations based on discussions with experienced forecasters. With regard to the latter, favor is given to those equations whose predictors involve data which is most likely to be reliably available in a real-time data acquisition setting. This was done by looking at the historical available of all predictors in the initial raw data set. Although this final equation selection has a large subjective component, it is important to remember that the selection is being made from amongst the top objectively-derived candidate equations whose difference in the error-based metric was virtually negligible.

1.5.1.3 "Day Typing"

One final element to the model development process must be included in order to complete the summary description of the statistical forecast model development process. Rather than simply include all available stratus days into the training data set for statistical development, the training database was subdivided into two or three subsets (depending on the model), based on the characteristics of meteorological parameters which are known through experience to characterize the atmospheric "regime" during which the dissipation of stratus was occurring on a given day. This distinction has come to be referred to as "day-typing," i.e., trying to distinguish the type of day under which the current forecast is being made. For example, in forecasting practice it is typical to anticipate the dissipation evolution differently based on the general direction of the geostrophic (pressure-induced) wind field, wherein an onshore wind is more likely to retard the burnoff process, while an offshore wind is more likely to accelerate the process. During the equation development process, the training dataset was therefore subdivided accordingly in order to anticipate these tendencies. The intent was to develop a different set of equations based on these generalized regimes (i.e., day types) which were known to have characteristically different behavior. Often, this subdivision resulted in equation sets for "typical" versus "less typical" conditions, that are reflected in the resultant performance of the corresponding forecast equation sets.

The final result is that each statistical forecast model set (Local, Regional, Satellite) consists of a different forecast equation for each model run hour and, in some instances, different forecast equations for a particular run hour depending on the "day type" selection for that model and run hour. The nature of the day-typing utility is more evident in the individual forecast description overviews that follow.

1.5.1.4 Statistical Model Summaries

This section provides a brief overview of each of the three statistical models. More detailed descriptions are provided in the corresponding sections which appear later in this document.

1.5.1.4.1 Local Statistical Forecast Model (LSFM)

The Local Model is considered "local" in that it relies heavily on observations from the two specially instrumented sensor sites at SFO and San Carlos Airport (SQL), in close proximity to the Approach Zone. In particular, the predictors selected by the LSFM focus heavily on cloud height/depth and the amount of solar radiation reaching the ground. Additionally, the local model also tends to include the regional geostrophic wind as an important "derived" predictor, as measured by the pressure differences across surface station pairs.

Due to its dependence on predictors associated with the depth of the local cloud layer, the day-typing stratification by the LSFM is associated with the so-called "quality" of the inversion height estimate. The base of the inversion height is presumed to be an estimate of the cloud top height; when this height is reliably estimated, the performance of the LSFM tends to be better. Thus, the quality of the inversion height estimate (particularly as derived from the sharpness of the sodar signal) was used as a discriminator in developing the LSFM forecast equations. Separate forecast equations were developed for good inversion height estimate quality versus inferior quality days.

One of the strengths of the LSFM is its reliance on data that is measured with high precision and high temporal resolution, specifically the sodar inversion heights and the pyranometer radiation measurements. Conversely, it is hindered by the fact that these measurements are only available at two key locations near the approach zone, rather than on a broader regional basis. As a result, the LSFM is expected to perform best on days when the stratus dissipation process is dominated by the local radiation effects, and less by the external influence of significant advective cooling or heating.

1.5.1.4.2 Regional Statistical Forecast Model (RSFM)

The Regional Model derives its predictor equations from routinely available weather observations extending beyond the immediate San Francisco Bay. The model relies heavily on standard hourly NWS surface observations from the Central California region, and the twice-daily vertical sounding from Oakland.

One aspect of the Regional Model which distinguishes it from the other models is its reliance upon multiple pairs of stations for computing pressure differences as a measure of geostrophic wind. As a result, this model is best positioned to account for anticipated changes to the wind field which impact the advective component of the cloud dissipation process. Other preferred predictors include the time of first cloud ceiling observation overnight in the Bay region, the height of the cloud ceiling at SFO, and the height of the inversion as determined by the Oakland sounding. Like the other statistical models, the RSFM also takes into account the maximum solar radiation available on a given day based on the sun angle which varies throughout the season.

The “day-typing” used for the RSFM involved the geostrophic wind. Essentially, the model categorizes a given day as having an onshore, offshore, or neutral geostrophic wind, with a separate set of forecast equations available for each condition.

Since the RSFM uses only routinely available observations, it is not reliant upon data from specialized sensors. Default forecast equations, utilizing different combinations of surface observations, have been developed which makes the model less vulnerable to the lack of availability of any single surface observation.

1.5.1.4.3 Satellite Statistical Forecast Model (SSFM)

For cloud depths typical of marine stratus (usually less than 1000 feet), the cloud depth is roughly linearly proportional to the liquid water content in the cloud. The Satellite Model correlates visible satellite brightness imagery with historical times of transition to dual approaches to estimate the time of cloud dissipation in the approach zone. The SSFM is run hourly between 15 GMT and 18 GMT, when visible satellite imagery is available.

The 1-km resolution visible satellite image from central California is filtered (to eliminate small-scale structure), registered (to ensure proper geographical orientation), and normalized to eliminate the effect of sun angle on cloud image brightness. The normalization for sun angle allows for an acceptable correlation between brightness and liquid cloud water.

Once the data are pre-processed, they are divided into 52 geographical sectors, primarily based on topography. Thus, each sector may contain roughly 10 to 30 data points. Within each sector, statistics are computed to estimate the mean cloud brightness and the mean area cloud coverage within each sector. These sector statistics thus represent 52 geographical data points for input to the SSFM. Through the model development procedure, the best-correlated sectors are chosen, with the appropriate predictor coefficients derived separately for each hour.

As with the other statistical models, predictor equations are developed for various “day types.” For the satellite imagery, the discriminating factors were found to be the height of the base of the cloud, and the “roughness” of the satellite image, i.e., the variance of the cloud brightness seen in the image. More typical stratus days are associated with cloud decks showing a smooth, regular texture. Fragmented or irregular cloud imagery tends to indicate the impact of a synoptic scale feature (such as a passing trough or weather system) and/or the presence of multiple cloud layers, which may disrupt the more typical radiation-driven dissipation process.

To an even greater degree than the Regional Model, the Satellite Model has the advantage of sampling data from an extensive geographic area. However, since it relies on visible imagery, it cannot be run prior to sunrise. Also, it is susceptible to any satellite data calibration errors

1.5.2 The COBEL Model

COBEL is a very high-resolution one-dimensional (vertical column) model of the planetary boundary layer that simulates the evolution of the cloud dissipation process. It is the only component forecast model in the system that is physics-based. It was adapted from a model developed for forecasting fog behavior in northern France. (COBEL is short for the French **Co**uche **B**rouillard **E**au **L**iquide.)

The model is initialized with a vertical profile of temperature, humidity, and wind at SFO. The initialization uses a hybrid of the Oakland (OAK) balloon sounding and high resolution lower atmosphere measurements from the special sensors deployed at SFO. This profile is used to characterize the stratus cloud deck in terms of total vertically integrated liquid water content. As the model steps forward in time, it simulates the physical processes associated with the cloud evaporation process. The model forecast time of cloud dissipation at SFO is declared when the cloud liquid water content reaches zero.

Since the model is designed to forecast for a single point location (SFO), the forecast must be adjusted to account for the spatial difference from the specific area of interest, i.e., the approach zone into SFO. This is done via historical correlation between the time of cloud dissipation at the point location (SFO) and the time of transition to dual approaches.

The model is re-initialized for each model run, using updated high resolution measurements of surface temperature, humidity, wind and solar radiation. In particular, the re-initialization (during daylight hours) includes an adjustment to ensure that the model’s interpretation of liquid cloud water amount is consistent with the amount of solar radiation transmitted through the cloud layer. The mode is then re-run, providing a new forecast of stratus dissipation.

The operational version of COBEL for San Francisco is designed for typical stratus days when stratus is present and fully developed at or before sunrise, when there is a marked strong inversion at the top of a shallow well-mixed marine boundary layer, and when no clouds are present above the boundary layer. If a reliable estimate of the inversion base height is not available, the model will not be allowed to run within the real-time system. Also, the current version of the model adapted for the SFO application assumes no horizontal advection and a fixed value for subsidence. A horizontal pressure force, corresponding to a 2 m/s geostrophic wind, is arbitrarily applied over the entire depth of the model domain (surface to 1.5 km height) and kept constant with time. As a result, the model is not expected to perform as well on days when the influence of temperature advection is large compared to radiation effects.

1.5.3 The Consensus Forecast

The consensus forecast combines the component forecasts to provide a single forecast of the time that stratus is expected to clear the approach zone to allow parallel approaches of aircraft. It is computed by assigning a weight to each of the component forecasts. This weight is derived from the historical performance of each component model, evaluated separately for each of their respective day types, and for each individual initialization hour. Thus, the weights are assigned based on the current initialization hour and the current day type being employed by each component model.

The consensus forecast time 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.” Any one of the following four conditions will trigger the LOW confidence indicator:

1. The inversion base height is not clearly identifiable. This may occur when the inversion is weak or diffuse. Since the Local SFM and COBEL rely heavily on the accurate inversion height, these two component models are not allowed to run within the real-time system without a reliable inversion height estimate.

2. There is a high cloud ceiling base. Statistically, model performance (particularly the Satellite Model) has been found to be degraded when the stratus ceiling height is above 2300 feet.

3. The cloudiness in the Bay area appears disorganized or patchy. This often occurs when there is a transient synoptic scale feature impacting the mid-California coast, such as the approach and passage of a trough or precipitating weather system. This is usually accompanied by transient cloud decks or multi-layered clouds, which the forecast models do not handle very well. Algorithmically, this condition is detected automatically in the real-time system when the variance of cloud brightness within the region exceeds an empirically derived threshold.

4. Less than three component models are available.

In addition to generating a discrete forecast time, the Consensus Forecast is also used to generate a probabilistic forecast of clearing prior to specific target times during the high traffic demand period, namely 17, 18, 19, and 20 GMT. The probabilistic forecast values are derived empirically by comparing historical consensus forecast times with actual verification times.

1.5.4 Forecast Model Performance Summary

Development of the forecast models was an iterative process that began in 2000, using a training data set 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 and 2002, the performance of the models was monitored; during the same period, the entire model development process (particularly for the statistical models) was streamlined and automated. This allowed for more rapid analysis of model performance and modification of models to allow testing of new versions. Four different trial versions of the three statistical models were examined during that two year period. 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 a re-work of the Satellite Model, after detection of a processing error in the raw data that needed to be corrected.)

Thus, the final version of the statistical models was developed during the winter of 2002-2003, using a training data set of stratus days from 1996-2002. Once the final models version of the models was established, a two-fold cross-verification procedure was performed to provide a quasi-independent assessment of expected model performance. The assessment was not truly independent, since the two validation data sets were subsets of the training data set. This assessment was performed in the following manner: The training data set was split into two subsets. The model equation coefficients for each predictor were then re-computed (via least squares) using one of the two data sets, and then scored using the other training set. This process was repeated, swapping the training data subset with the scoring subset. The final error estimates were made by averaging the errors from the two scored subsets.

Although not a truly independent scoring exercise, this process allowed a more realistic estimate of the expected performance of each of the models at each initialization time. These results are shown in the left column of Table 1, labeled “Development.” The table presents three sets of statistics. First is the Median Absolute Error (in hours) 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 performance statistics for these two seasons combined are shown in the right column of Table 1.

For a baseline perspective, error statistics for forecasts using conditional climatology are shown in Table 2. Conditional climatology presumes a forecast for each run hour that is simply computed as the average verification time for all days that had stratus clouds present at that hour. In essence, these baseline error statistics represent the minimum acceptable level of skill that would need to be demonstrated by the forecast system.

TABLE 1

Model Performance Statistics for Development Data Set (left column) and Independent Demonstration Data Set (right column)

Development (1996-2002)

Median Absolute Error (hrs)

	Hour (GMT)						
	09	**11**	**13**	**15**	**16**	**17**	**18**
Consensus	0.77	0.77	0.71	0.59	0.49	0.42	0.57
Cobel	0.87	0.82	0.79	0.60	0.51	0.52	0.49
Local	0.67	0.71	0.61	0.57	0.47	0.45	0.53
Regional	0.80	0.86	0.77	0.77	0.72	0.66	0.64
Satellite	#	#	#	0.53	0.40	0.40	0.40

Bias (hrs)

	09	**11**	**13**	**15**	**16**	**17**	**18**
Consensus	0.05	0.05	0.02	-0.03	-0.01	-0.03	-0.03
Cobel	0.14	0.12	0.09	-0.01	0.05	0.07	0.04
Local	0.14	0.13	0.10	0.09	0.04	0.08	0.12
Regional	0.24	0.24	0.18	0.15	0.21	0.13	0.13
Satellite	#	#	#	0.12	0.08	0.07	0.22

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 (hrs)

	Hour (GMT)						
	09	**11**	**13**	**15**	**16**	**17**	**18**
Consensus	0.97	0.89	0.76	0.75	0.67	0.63	0.59
Cobel	0.94	0.78	0.73	0.71	0.62	0.65	0.90
Local	0.99	0.80	0.74	0.76	0.77	0.87	0.81
Regional	0.96	1.03	0.84	0.89	0.74	0.51	0.45
Satellite	#	#	#	0.73	0.60	0.53	0.59

Bias (hrs)

	09	**11**	**13**	**15**	**16**	**17**	**18**
Consensus	0.34	0.47	0.37	0.49	0.43	0.53	0.62
Cobel	0.00	0.14	0.20	0.22	0.31	0.47	0.69
Local	0.21	0.56	0.50	0.42	0.34	0.86	0.77
Regional	0.40	0.58	0.27	0.45	0.33	0.01	0.28
Satellite	#	#	#	0.59	0.51	0.54	0.73

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	#	#	#	123	112	91	56

TABLE 2

Median Absolute Error (hours) of Conditional Climatology Forecasts for Each Model Initialization

Hour (GMT)	**09**	**11**	**13**	**15**	**16**	**17**	**18**
Conditional Climatology	1.05	1.05	1.05	1.00	1.00	0.87	0.75

As would be expected, the independent data set did not score as well as the errors derived from the training data set. In general, however, the median absolute errors were comfortably lower than conditional climatology, particularly at the key tactical forecast hour of 16 GMT which represents the last forecast hour for which there is sufficient leadtime to cancel an active delay program. The independent data set for 2003-04 also exhibited a late forecast bias. It is presumed that this was largely impacted by the 2004 season which consisted of a heavy early bias (more than 0.30 hours) in the verification distribution. This was evidenced by the consistency of the forecast bias across all four of the component forecast models, including the physics-based COBEL model.

The following sections provided a more detailed description of the four component forecast models, and the consensus forecast algorithm.

2. LOCAL STATISTICAL FORECAST MODEL

2.1 INTRODUCTION

This section describes the Local Statistical Forecast Model (LSFM) for the San Francisco Marine Stratus Forecast System. The purpose is to provide an overview of the model and how the model is run within the framework of operational system software installed on the base station computer, which is located at the Air Route Traffic Control Center in Fremont, CA.

2.2 HISTORICAL PERSPECTIVE

The LSFM is rooted in work that first began with a scattergram forecast tool that was developed and utilized at the CWSU in Fremont. CA. The scattergram provided guidance in estimating the time of burn off of marine stratus in the approach zone to San Francisco International Airport. The scattergram was derived from 3 years of data, plotting cloud base against cloud thickness and noting the time of burn off. A subjective hand-drawn contour analysis completed the scattergram. The scattergram reinforced the forecasters experiences; that 1) in general, thick stratus will take longer to burn off than thin stratus, 2) burn off is delayed on days with a high inversion base, and 3) marine stratus burnoff appeared to be influenced by whether the dominant synoptic feature was a ridge or trough.

The work first begun by the forecasters at the CWSU was extended by developing a least-squares fit to a simple physical model that used limited historical data as a training set. The cloud parameters deemed important by forecasters were included in the model and partitioned by wind direction in the training set to simulate the partitioning of ridge and trough cases as the forecasters had done for the scattergram. This approach reinforced using a statistics-based methodology with the cloud parameters identified by CWSU forecasters as important to the dissipation of marine stratus in the boundary layer. This work was later extended to include manual procedures for model development based on nonlinear regression of nonlinearly scaled predictors. These manual procedures in turn, formed the basis of the automation effort that is described in Appendix A.

2.3 TECHNICAL DESCRIPTION

2.3.1 Overview

The LSFM is one of three statistically based forecast models that use statistical regression to forecast the time of stratus burn off in the approach zone to San Francisco International Airport. The model is considered “local” in that predictors are derived from weather observations taken from stations close to the San Francisco Bay area, including data from the project-specific field sites located at San Francisco Airport and San Carlos Airport.

The two field sites are unique among other reporting sites in that they host monostatic acoustic sodar (SOund Detection And Ranging) sensors that provide thermal soundings of the marine boundary layer. The sodar data are routinely input to an algorithm that provides an estimate of the height of the inversion base, which is also a reasonable estimate of cloud-top height for marine stratus. The inversion base height and other quantities derived from it (e.g., cloud thickness) are utilized as predictors and are unique to the LSFM. The field sites also host pyranometers that provide measurements of down-welling and net short-wave radiation. The LSFM utilizes the ratio of down-welling radiation to the maximum clear-sky radiation (location and time dependent) as a predictor.

The LSFM is run during days of marine stratus, providing forecasts at 13, 15, 16, 17, and 18GMT and during the pre-dawn hours of 9 and 11GMT, contingent upon established cloud at those hours. Like the Regional and Satellite Statistical Forecast Models, the LSFM was developed using the automated statistical forecast model development system described in Appendix A.

2.3.2 Model Day Type

In 2002, MIT/Lincoln Laboratory developed a statistical forecast development system to automate the model development process (see Appendix A). The system made feasible the ability to develop models based on stratifying cases with similar recurring meteorological conditions. These conditions may favor a different set of predictors over the more general case of developing models for all cases and all meteorological conditions. The process of stratifying cases in this manner has come to be known as day typing to distinguish all stratus days from a subset of days exhibiting similar meteorological conditions.

The conditions chosen to stratify cases for the LSFM was based on the suggestion that exhibiting a stable and well-established marine boundary layer should provide increased confidence in probability of burn off because local-scale processes important to stratus dissipation dominate the situation (Farley, personal communication) [1]. That is, local-scale processes are not cloaked by larger-scale events such as a synoptic-scale system disturbing the marine boundary layer.

To distinguish days in which the marine boundary layer is considered to be stable and well established from days in which it is not, day type modeling for the LSFM examines the quality factor produced by the consensus inversion base height product algorithm. The quality factor provides an indication of the confidence of the algorithm in identifying and tracking the height of the inversion base that caps the marine boundary layer. In a typical marine stratus situation, the inversion base is well defined and varies little over the course of the day. Under these conditions, the consensus algorithm tracks the inversion base with a high degree of confidence and corresponding high quality factor. The LSFM day type models are models trained on such days, whereas “default” models are trained on all days. Table 3 shows the quality factor thresholds that must be matched or exceeded in order for a day type model to be selected.

TABLE 3

Consensus Inversion Base Quality Factor Thresholds for the LSFM

Forecast Run Time (GMT)	Consensus Inversion Base Quality Factor Threshold
9	Default Model Only
11	Default Model Only
13	55
15	90
16	90
17	Default Model Only
18	Default Model Only

2.3.3 Model Characteristics and Expected Performance

The LSFM is composed of multiple predictor sets, with day type and default sets having been developed for the 13, 15, and 16GMT forecast runs and default sets developed for 9, 11, 17 and 18GMT runs. Day type models were not developed for the predawn hours due to the limited number of marine stratus cases available during those hours. Day type models developed for 17 and 18GMT showed insignificant improvement over default models developed for the same time periods and were therefore not included in the delivered model. The LSFM predictor sets delivered to the operational system were manually selected from a field of sets generated through an automated statistical forecast model development process. That process began with a field of predictors that were limited to observations and derived quantities from the two field sites, the NWS field site located at the San Mateo Bridge, 5-minute surface observation data from San Francisco International Airport (SFO), and hourly observation data from San Carlos Airport (SQL).

Table 4 provides a summary of predictors that persist throughout the model sets, for both default and day type models. Of particular note is the persistence of sodar-derived cloud top height (Zi), an observation that serves to strengthen the same observation made by forecasters at the CWSU who noted the dependence of cloud top height with burn off time. In work conducted by UQAM (personal communication) to develop a predawn version of the COBEL model, researchers noted similar results with sensitivity studies that showed an almost linear relationship between cloud top height and burn off time. Their analysis showed a mean delay of 7 minutes in burn off for every 30-meter rise in cloud top height for heights above 200 meters.

As seen in Table 4, during the predawn and early morning forecasts, the predictor sets favor San Francisco-Arcata and Red Bluff pressure gradient predictors, gradients that are routinely used by forecasters to examine the geostrophic flow for synoptic-scale pressure-induced forcing along the coast of California. The Red Bluff pressure gradient reappears during the early afternoon intervals, suggesting that these models may possibly be reexamining strong onshore flow as one factor for persistent late day stratus conditions. By mid morning, cloud base and thickness-related predictors as well as the amount of down-welling short-wave radiation reaching the sensor sites dominate the predictor sets. The presence of these predictors is expected because their rates of change over the course of a stratus event are an indirect measure of the physical processes involved in the burn off of marine stratus. For ceiling, the expectation is that ceiling heights will rise as cloud is evaporated from below, resulting in a concomitant decrease in cloud thickness and increase in measured down-welling short-wave radiation. Finally, early afternoon forecasts at 17 and 18GMT include surface temperature and dew point temperature as predictors. Both are important in characterizing boundary layer heating and moisture content. Sensitivity studies conducted at UQAM (personal communication) have shown a strong dependence in the variation of surface temperature and burn off under conditions of low humidity and a weak dependence with temperature under moist surface conditions.

The LSFM is unique among the statistical models in that it has available to it: 1) high temporal and spatial sodar measurements used to derive cloud top height, 2) high temporal resolution measurements of down-welling short-wave radiation reaching the field sites, and 3) high temporal resolution observations of surface temperature, dew-point, and wind within the approach zone to San Francisco Airport. These observations provide the LSFM with information that helps to characterize the marine boundary layer within the approach zone in terms of cloud vertical extent and surface characteristics below the cloud including surface heating, moisture, and wind information. Consequently, the LSFM generally performs best under conditions in which radiative processes are unencumbered by advection. Under these conditions, day type models will generally outperform their default counterparts when the marine boundary layer is stable and well established. Conversely, the LSFM is not expected to perform well when advective processes dominate a marine stratus event. Although some predictor sets do contain wind information, the LSFM is at a disadvantage because it is essentially providing forecasts based on limited spatial information within the Bay area (i.e., the approach zone).

TABLE 4

Preferred Predictors as a Function of Forecast Interval

Forecast Interval (GMT) →	09	11	13	15	16	17	18
Predictor ↓							
Zi[1]	X	X	X	X	X	X	X
RBLPdx[2]	X	X	X			X	X
ACVPdx[2]	X	X	X				
Cig[3]**, Δcig**			X	X	X	X	X
CldThk[4]**, ΔCldThk**		X	X	X	X	X	
Rad[5]**, ΔRad**				X	X	X	X
T, T_d[6]						X	X

[1] Cloud top height as determined by sodar.
[2] Pressure gradient, determined as $SFO_{pressure}$ - <station>$_{pressure}$
[3] Cloud base ceiling and one-hour rate of change.
[4] Cloud thickness (cloud top height – cloud base height) and one-hour rate of change.
[5] Ratio (down-welling radiation / maximum clear sky radiation) and one-hour rate of change.
[6] Surface temperature, dew-point temperature.

2.4 REAL-TIME PROCESSING

2.4.1 Data Inputs

There is a set of three input files required for each model or models developed for each run-time. Files are stored in subdirectories named by model type under the directory */ll/cv/StatisticalModels/Local/ascii*. Table 5 identifies the subdirectory names for each run-time/model type combination.

The first file with the *params* extension is a file that specifies the raw data required to derive the predictors of the model. For day type models, this file will also contain an additional reference to the consensus inversion base height quality factor, the day type threshold parameter used by the LSFM. The contents of this file do not change unless the predictors of the model change.

The second file with the *lbls_out* extension lists the symbolic name of each model predictor. For day type models, this file will also contain an additional reference to the consensus inversion base height quality factor. The contents of this file also do not change unless the predictors of the model change. This file is used together with the first file to derive the values of the predictors listed.

The third file with the *mdl* extension contains model evaluation parameters for each predictor including predictor weights, parameters that define piece-wise linear interpolation functions and constants. The contents of this file do not change unless the predictors change or the model is retuned with additional data.

2.4.2 Intermediate Outputs/Inputs

Two intermediate output log files are generated in the course of running the LSFM. They are located in the directory */ll/cv/logs/YYYYMMDD/LocalModel*. The first file (*raw* file) contains observations extracted from day summary files according to specifications listed in the input *params* file. This output subsequently becomes input to a program that derives the model predictors listed in the *lbls_out* input file, producing another intermediate output file (*inputdata* file) of predictor name/value pairs. This file along with the *mdl* input file is then input to the model evaluation program to generate a forecast burn off time.

2.4.3 Data Outputs

If a successful forecast is produced, the model evaluation program generates an output file containing the model used, burn off time, quality factor, and statistics for each predictor. The output file can be identified with the file suffix *.0.log* and is also located in the same directory as the intermediate output log files described above. An example of the contents of the output file is shown in Figure 3.

TABLE 5

Subdirectories Containing Input Files for Each Run-Time/Model Type Combination of the LSFM

Directory Name	Description
Local09Default	Default model used for pre-dawn forecast at 9GMT.
Local11Default	Default model used for pre-dawn forecast at 11GMT.
Local13ZiQGE55	Day type model used at 13GMT when marine boundary layer is well established.
Local13ZiQLT55	Default model used at 13GMT when marine boundary layer is not well-established or missing data prevents the day type model from running.
Local15ZiQGE90	Day type model used at 15GMT when marine boundary layer is well established.
Local15ZiQLT90	Default model used at 15GMT when marine boundary layer is not well-established or missing data prevents the day type model from running.
Local16ZiQGE90	Day type model used at 16GMT when marine boundary layer is well established.
Local16ZiQLT90	Default model used at 16GMT when marine boundary layer is not well-established or missing data prevents the day type model from running.
Local17Default	Default model used at 17GMT.
Local18Default	Default model used at 18GMT.

2.4.4 Structural Flow

LSFM processing is illustrated in the functional flow diagram of Figure 4. Processing begins by establishing model parameters based on the current run-time. Parameters include day type and default model names and day type threshold. The logic then queries to see if a day type model exists for the current run-time. If no day type model exists (i.e., 9, 11, 17 and 18Z models), the default model is used. If the day type model exists, the consensus inversion base height quality factor for San Carlos (valid at the current run-time) is compared to the threshold. If the quality factor equals or exceeds the threshold, the day type model is chosen, otherwise the default model is chosen. Once a model is chosen, the final step before running the model is to insure that data are available to generate the predictors of the model.

The second-half of the processing shown in Figure 4 illustrates the data preparation steps that precede model execution. The first step is to read data from the day summary files sufficient to generate the model predictors. The data to read is governed by the input *params* file that includes references to data that may be used as a substitute for missing data. Table 6 lists the primary and substitute data sources used by the LSFM. The output from the data collection program is a *raw* file. The second step is to generate the model predictors in a name-value format acceptable to the model evaluation program. Model predictor names are read from the input *lbls_out* file together with the (now) input *raw* file to derive predictor values, using substituted values as needed. The resultant output *inputdata* file, containing name-value pairs for each predictor, together with the *mdl* file, containing predictor weights and function definitions, are inputs to the model evaluation program. The output produced by the model evaluation program is written to a *0.log* file and replicated in another directory accessed by the display software. The forecast burn off time and quality factor are also written to an LSFM day summary file to facilitate access by the consensus model.

2.4.5 Nominal and Degraded Performance

When data availability issues prevent one of the LSFM day type models from running, a default model is substituted to run in its place.

In order for any model to run, all predictors that comprise the model must have valid values. In this respect, there is no configuration in which a model can run in a degraded mode in which one or more predictors are missing.

```
#   -
#   ! Current Time  02-oct-03 15:29:10
#   -

Evaluate for Time  <20031002> => 1065106800 => <02-oct-03 15:00:00> (
Jday 275 )
 Use Time window of   720 seconds



ModelSSum Model Structure Summary
ModelSSum Model/Daytype: Local15ZiQGE90

ModelSSum   Term  Name           Weight   Value   Scaled_Value
      Partial_Sum
ModelSSum   1       15sfoV           0.0471   1.5730  0.6662
      0.0314
ModelSSum   2       15sfoCig         -0.3971  182.0000     0.0000
      -0.0000
ModelSSum   3       15sqlTh1         1.4310   0.0000  0.0000
      0.0000
ModelSSum   4       15sqlRad         -0.4869  0.5000  1.0000
      -0.4869
ModelSSum   5       15sfoZi          0.8287   556.0000     0.6297
      0.5218
ModelSSum   6       15smbTh1         1.3080   190.0000     0.3317
      0.4339
ModelSSum   7       15sfoRad         -1.3960  0.2000  0.2376
      -0.3317
ModelSSum   8       Constant_Term    1.0000   5.2908  5.2908
      5.2908
ModelSSum   9       GMT_Sunrise 1.0000   14.1359 14.1359    14.1359

 RESULT
/ll/cv/StatisticalModels/Local/ascii/Local15ZiQGE90/Local15ZiQGE90.mdl
=>   19.595 with result of 0 = < - OK - >  and 0 Substituted values

#<EndOfProcessing>
```

Figure 3. Sample contents of the output file generated by the Model Evaluation Program for Statistically Based Models.

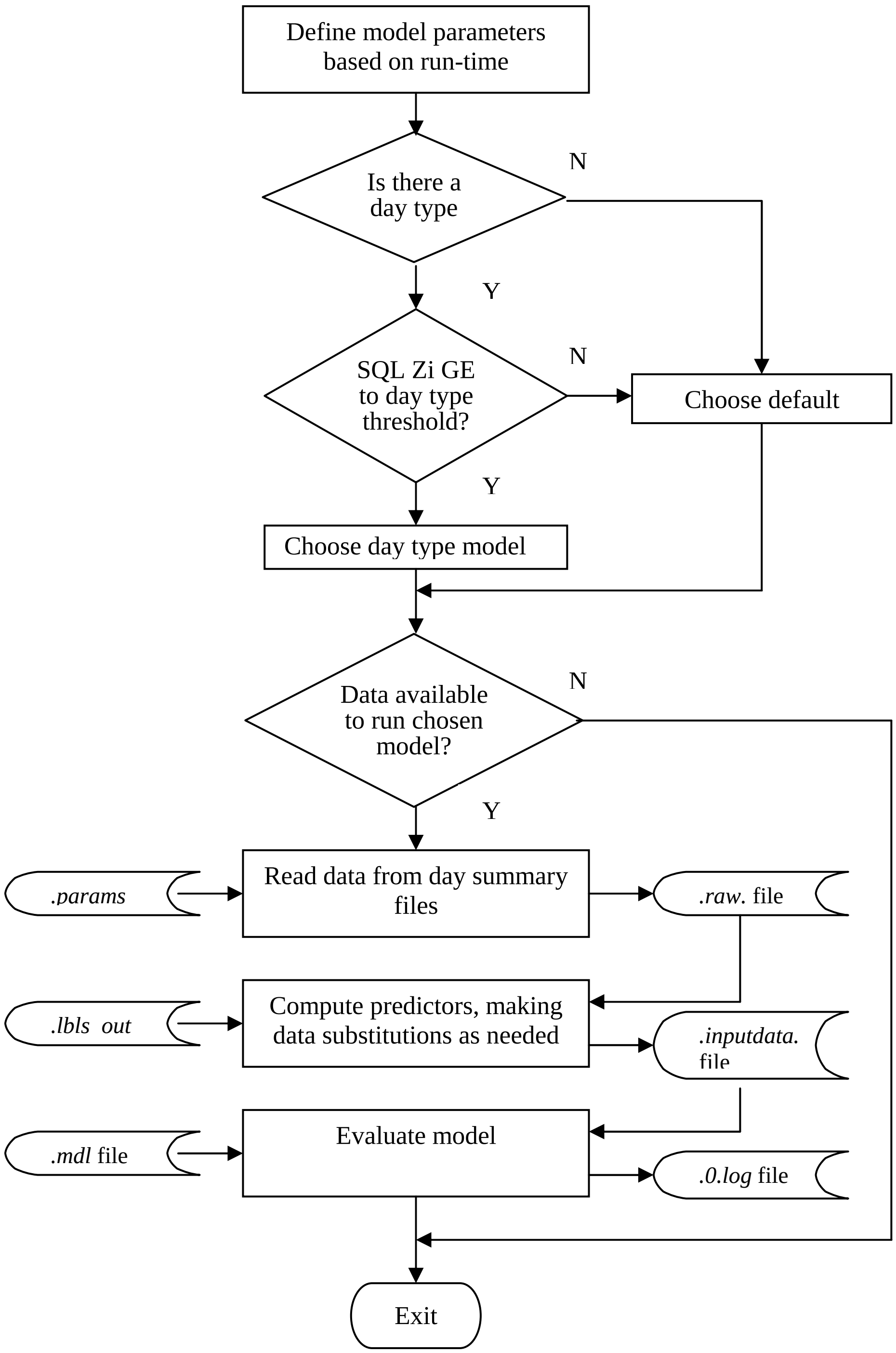


Figure 4. Functional flow diagram for LSFM processing.

2.5 PRACTICAL CONSIDERATIONS

2.5.1 Programming and Resources

There are no issues related to programming and resources.

2.5.2 Exception Handling

The inability of a default model to run because of the lack of data does not trigger any exception processing. The LSFM simply does not generate a forecast.

2.5.3 Quality Assurance and Diagnostics

Forecast quality, expressed as a quality factor, is an indication of the timeliness of the data used to derive model predictors. A quality factor less than 5 indicates that one or more predictors were derived from data less than nominal in time.

Table 3 of the Real-time System Document lists the location and naming structure of log files that are produced or updated for each run of the LSFM. These files can be used as a diagnostic tool to determine why an LSFM day type or default model ran or didn't run or why no model ran. The model run control file posts the results of every run and will indicate if a run was unsuccessful due to missing data. To investigate further, the model evaluation program input file (*inputdata* file) can be examined to determine which predictor or predictors were missing and prevented the model from running. Refer to the trouble-shooting guide (Appendix A) for additional information on how to identify missing predictor values in this file.

2.5.4 Operator Interaction and Default Settings

The LSFM is activated by a model launching script, so no operator interaction is required. Default settings for the LSFM run intervals are shown in Table 4 of the Real-time System Document.

2.5.5 Periodic/Potential Maintenance Tasks

No periodic maintenance is required of the LSFM. If a new model is introduced into the system, the input files described in Section 2.4.1 will need to be updated to reflect the new predictor set. If however, an existing model is simply retuned, only the *mdl* file requires updating. Programming changes would only be necessary if a new model introduced predictors new to the system.

TABLE 6

Substitute Data Sources Available to the LSFM

Primary Data Source	Substitute Data Source
Hourly observations from SFO	ASOS observations from SFO
Hourly wind observations from SQL	Wind observations from the SQL field site
Hourly pressure observation at SAC	Hourly pressure observation at MCC
Maximum radiation computation for SFO	Maximum radiation computation for SQL

3. REGIONAL STATISTICAL FORECAST MODEL

3.1 INTRODUCTION

This section describes the Regional Statistical Forecast Model (RSFM) for the San Francisco Marine Stratus Forecast System. The purpose is to provide an overview of the model and how the model is run within the framework of operational system software installed on the base station computer, which is located at the Air Route Traffic Control Center ARTCC in Fremont, CA.

3.2 HISTORICAL PERSPECTIVE

In 1998, Pennsylvania State University under funding from the FAA developed a regional stratus cloud forecast model for SFO. The Penn State regional model utilized standard hourly weather observations and was developed using statistical regression of historical observation data. The original Penn State model, however, was designed to predict the ceiling height at SFO, and not the time of approach zone clearing. A supplemental algorithm was used to correlate the predicted ceiling height to SFO clearing time. Subsequently, MIT/LL completely redeveloped this regional model utilizing standard observations and some derived observation, using on a database of actual verification times of dual approach availability into SFO, and employing an advanced statistical model development tool (see Appendix A). Additionally, the CWSU was independently using several rules-of-thumb for determining the general bias of the stratus based on the pressure difference from ACV to SFO. This technique formed the basis of the day typing for the RSFM that will be discussed in detail.

3.3 TECHNICAL DESCRIPTION

3.3.1 Overview

The RSFM is one of three statistically based forecast models that uses statistical regression to forecast the time of stratus burn off in the approach zone to San Francisco International Airport. The model is considered “regional” in that predictors are derived from weather observations taken from regional stations surrounding the San Francisco Bay area. It is designed to require as input only routinely available National Weather Service surface and upper air observations.

In 2002, MIT/Lincoln Laboratory developed a statistical forecast development system to automate the model development process. The system made feasible the ability to develop models based on stratifying cases with similar recurring meteorological conditions. These conditions may favor a different set of predictors over the more general case of developing models for all cases and all meteorological conditions. The process of stratifying cases in this manner has come to be known as “day typing” and is used to distinguish broad categories of stratus days from one another. In the case of the regional model, the day type is based on the strength of the on/off shore flow as determined by the geostrophic pressure difference.

A total of 19 NWS surface stations and 1 balloon sounding station were selected for inclusion as potential predictors in the SFO regional modeling (see Table 7). All of the potential input predictor parameters are detailed in Table 8. The basic input predictors consisted of all the standard temperature, pressure wind and visual observations along with the ceiling height. In addition, the pressure and temperature difference from the target station to SFO were also used as predictors. Temperature, dewpoint and winds were used from the balloon sounding (at 1000, 925, 850, 700, and 500 mb) in addition to a derived calculation of dewpoint depression. The inversion height estimate from the 00Z and 12Z Oakland sounding, and an objective measure of its quality of the inversion height (i.e., the strength of the inversion signal observed in the data) were also used as initial input predictors. Finally, a series of special predictors were also added:

MaxRad: The maximum potential solar radiation available at the forecast hour. This is not a measured value but a climatological potential value, based on the sun angle for a specific location. Models were allowed to use any hour’s measurement. For example, the 09Z model was allowed to use the 14ZMaxRad value since sunrise always occurred after 09Z. The statistical model typically chose a solar radiation value that was nearest in time to the model forecast period.

FirstCloud: The first hour (after 00Z) that a ceiling below 3500ft is present at any of the four bay-bordering NWS stations (SFO, OAK, NUQ, SQL). There is a reverse correlation between the time of onset and the time of burn-off (i.e., late onset indicates an early burn-off).

Wxx: An index representing the level/type of precipitation that is falling. 0-No weather, 1-Haze or Fog, 2-Drizzle, 3-Shower, 4-Light Rain, 5-Rain(or higher). Winter precipitation was not considered for the bay area stations.

TABLE 7

Input Predictor Station Set for the SFO RSFM Model

Station	Name (in CA unless noted)	Station	Name (in CA unless noted)
ACV	Arcata	RNO	Reno, NV
APC	Napa	SAC	Sacramento
CCR	Concord	SFO	San Francisco International Airport
FAT	Fresno	SJC	San Jose
HWD	Hayward	SMF	Sacramento
LVK	Livermore	SNS	Salinas
MRY	Monterey	SQL	San Carlos
NUQ	Moffett Field	STS	Santa Rosa
OAK	Oakland	SUU	Fairfield
RBL	Red Bluff	UKI	Ukiah

TABLE 8

Description of All Available Input Predictors

Predictor	Description
Standard ASOS station predictors	
Tmp	Ambient temperature (F)
Dew	Dewpoint temperature (F)
Dir	Wind direction (deg, 0-360)
Spd	Wind speed (knots)
Prs	Station pressure (mb)
Cig	Ceiling height (??)
Vis	Visibility (statute miles)
Derived ASOS station predictors	
Ucomp	U-component of wind speed, direction (m/s)
Vcomp	V-component of wind speed, direction (m/s)
DwptDprs	Dewpoint depression (Tmp-Dew) (F)
xxxTmpDXsfo	Temperature change from station xxx to SFO at current hour (F)
xxxPrsDXsfo	Pressure change from station xxx to SFO at current hour (F)
CldMax	The maximum cloud coverage at any reported cloud level (0-1)
FirstCloud	The first hour (00Z or later) at which the cloud ceiling reported at any bay area station (OAK, SFO, NUQ, SQL) was below 3500 feet. This value is calculated for each forecast and only considers the ceilings at or prior to the forecast hour. For example, if the ceiling is below 3500 feet starting at 12Z, the 09Z model will report a 10Z onset time (because at 9Z all we know is that the onset, if it occurs, will happen after 9Z).
Standard upper level air predictors (at 1000, 925, 850, 700, and 500 mb)	
Tmp	Ambient temperature (F)
Dew	Dewpoint temperature (F)
Dir	Wind direction (deg, 0-360)
Spd	Wind speed (knots)
Derived upper level air station predictors	
Ucomp	U-component of wind speed, direction (m/s)
Vcomp	V-component of wind speed, direction (m/s)
DwptDprs	Dewpoint depression (Tmp-Dew) (F)
InvOak	The inversion height (??)
InvQfl	The “quality” of the inversion height based on the Zi algorithm
Solar radiation	
14-18sfoMaxrad	The theoretical solar radiation influx for SFO based on the date and hour. (units??)

3.3.2 Day Types

Day types are used to subset the stratus dataset when a specific predictor identifies a common grouping of case days. For example, the CWSU has routinely used the atmospheric pressure difference from ACV (Arcata, CA) to SFO as a measurement of the geostrophic (synoptic-scale pressure induced) forcing along the coast of California. This measure of synoptic scale forcing allows each day to be categorized as a function of the burn-off bias. The SFO RSFM uses three mutually exclusive day types that are based on subsets of the ACV-SFO pressure difference measurement. The day type criteria are shown in Table 9.

TABLE 9

Day Type Breakdown for SFO RSFM Model

ACV-SFO Pressure Delta (mb)	Type of day	Expected burn-off bias
$\Delta P < -2.0$	Offshore	Early
$-2.0 >= \Delta P < 1.0$	Neutral	None
$\Delta P > 1.0$	Onshore	Late

These values were determined by balancing model performance with a data set that was large enough to perform a valid statistical regression. A default model based on all the available case days was also developed and is used as a backup when predictors from the appropriate day type model are unavailable.

3.3.3 Forced Predictors

Certain predictors are forced into the final set of 20 potential predictors in an attempt to make the models more consistent from hour to hour. The following 5 predictors are forced into the final 20 predictors if they are not already there after the initial statistical processing (<HH> – forecast hour) (<stn> 3letter station name):

<HH>sfoMxrad The maximum theoretical cloud-free solar radiation at SFO for the current hour.

<HH>sfoCig The ceiling height at SFO for the current hour.

<HH>xxxFirstCloud The first hour (after 00Z) that a ceiling lower than 3500ft was reported at either SFO, SQL, or NUQ.

<00/12>InvOak The inversion height at Oakland for either 12Z or 00Z.

<HH><stn>PrsDXsfo The pressure difference from some station to SFO at the current hour.

The maximum solar radiation is used as a surrogate for radiometer data that is excluded from use in the regional model (since radiation information is not a routinely available NWS observation). The ceiling height at SFO and the inversion height in the Oakland balloon sounding are important because they represent the bottom and top of the stratus cloud deck respectively. The pressure difference, as discussed under day types, represents the onshore/offshore influence of the particular synoptic flow. Including a pressure difference as a potential predictor allows the model to weight the magnitude of the pressure difference in addition to using the pressure difference as a day type for sample clustering. Finally, FirstCloud is included because the time of onset of the first cloud cover has been observed to be highly negatively correlated to the time of ultimate cloud dissipation.

The model optimization process was typically set to require that at least three of the forced predictors are predictors in the final model selected. When there are a limited number of test case days for a particular hour and day type as few as 1 of these predictors may be required in the final model. In addition, the final model must contain no more than 1 predictor per 20 test case days. The forced predictors and maximum predictor size were a restriction unique to the Regional model processing due to the extremely large set of initial potential predictors.

3.3.4 RSFM Forecast Equations

The RSFM model runs at 9Z, 11Z, and 13Z, and hourly from 15Z through 18Z, generating a new side-by time prediction each model hour. There are two levels of models: day type and default. The day type model generally performs better than the default model, but the statistical models require that all observations be present in order for the model to make a valid prediction. Sometimes (~10-15% of the time) an observation will be missing in the day type model and the processing will run the default model instead. Infrequently (~5% of the time) both models will be missing observations, and in that case no RSFM prediction will be made.

As mentioned above, the conditions chosen to stratify cases for the RSFM are based on the pressure difference between ACV and SFO (ACV minus SFO, in mb). A separate model equation was developed for each forecast hour and day type. In the event that the pressure reading from either ACV or SFO is missing, a default model based on statistically processing all day types at that forecast hour is used as a backup.

Each forecast model has a set of 1 to N predictors that are assigned weights. These weights are associated with the relative historical performance of each model equation. Table 10 lists each model and its' corresponding predictor set. It is difficult to break down each statistical model and explain the technical reasons why the statistical processing chose specific predictors, let alone combinations of predictors. However, we can say some general physical reasons why groups of predictors might be important to the burn-off process. Preferred predictors are shown in gray on the left-hand side; the rationale for each of these predictors was discussed previously. Other predictors include: Ceiling heights from stations other than SFO, the maximum cloud coverage (CldMax), wind measurements, temperature/dewpoint measurements, visibility/weather and pressure. The SFO stratus problem tends to be a regional phenomenon along the Central California coast; therefore, using the ceiling heights and maximum cloud coverage of the stratus from neighboring stations makes intuitive sense. Wind measurements (speed, direction, u-, v-component) may be indicative of on- or offshore flow, while strong wind magnitudes may indicate both the strength of the moisture flow and a measure of the surface mixing. Winds aloft (sounding) would be indicative of upper level flow and mixing. Temperature and dewpoint measurements represent the potential for cold air and/or moisture advection. Visibility and weather measurements help the models distinguish days where precipitation is actually falling in the region (a distinct minority of the summer stratus cases for SFO). Finally, pressure measurements provide a broad metric of the synoptic conditions (low and high pressure areas) that are important to stratus development and longevity.

TABLE 10

Forecast Model Predictor Sets

Model		Preferred Predictors					Other Predictors					
Name	#Preds	MaxRad	sfoCig	PrsDXsfo	1stCloud	InvOak	Ceiling	CldMax	Winds	Temp/Dew	Vis/Wxx	Other
DEFAULT MODELS												
09Default	6	14sfoMxrad	09sfoCig				09snsCig 09suuCig		09oakUcomp 09fatDwptDprs			
11Default	6	14sfoMxrad	11sfoCig		11xxx		11snsCig	11suuCldMax		11sfoDwptDprs		
13Default	7		13sfoCig	13rnoPrsDXsfo	13xxx			13sacCldMax	13oakUcomp	13oakTmpDXsfo		00InvQfi
15Default	7	15sfoMxrad	15sfoCig		15xxx			15lvkCldMax	15oakSpd 15lvkSpd	12RRRTmp700		
16Default	6	16sfoMxrad	16sfoCig		16xxx			16sjcCldMax	16oakSpd	12RRRTmp700		
17Default	6		17sfoCig		17xxx	12InvOak			17lvkSpd	17sfoTmp 12RRRTmp700		
18Default	8	18sfoMxrad	18sfoCig		18xxx			18ccrCldMax 18sqlCldMax	18ccrSpd 18oakSpd	18sfoDwptDprs		
NEUTRAL DAY TYPE MODELS												
09Neutral	6		09sfoCig	09acvPrsDXsfo	09xxx		09snsCig		09oakUcomp 09smfSpd			
11Neutral	5		11sfoCig	11acvPrsDXsfo	11xxx		11snsCig		11oakUcomp			
13Neutral	6	14sfoMxrad	13sfoCig		13xxx				13oakUcomp 13smfSpd	13fatDwptDprs		
15Neutral	6		15sfoCig	15acvPrsDXsfo	15xxx			15lvkCldMax 15sfoCldMax	15oakSpd			
16Neutral	7	16sfoMxrad	16sfoCig	16acvPrsDXsfo	16xxx	12InvOak			16oakSpd	16acvTmp		
17Neutral	7	17sfoMxrad	17sfoCig		17xxx			17sjcCldMax		17fatTmp 17sfoTmp		17rnoPrs
18Neutral	8	18sfoMxrad		18acvPrsDXsfo	18xxx			18sqlCldMax 18ccrCldMax	18oakSpd	18acvTmpDXsfo	18oakVis	18rnoPrs
OFFSHORE DAY TYPE MODELS												
09Offshore	6	14sfoMxrad		09snsPrsDXsfo	09xxx				09smfSpd	09suuDew	09oakVis	
11Offshore	8			11acvPrsDXsfo 11rnoPrsDXsfo	11xxx			11acvCldMax	11oakSpd 11smfSpd	11oakDew	11sfoVis	
13Offshore	8	15sfoMxrad		13acvPrsDXsfo			13acvCig	13sacCldMax	12RRRSpd500	13oakTmpDXsfo 13sfoDwptDprs	13oakVis	
15Offshore	8	15sfoMxrad		15acvPrsDXsfo					15hwdSpd 15hwdDir 15smfSpd 15acvSpd	12RRRDew850	15sacVis	
16Offshore	7	16sfoMxrad	16sfoCig			12InvOak	16sqlCig		16lvkSpd 16acvSpd	16sjcDew		
17Offshore	4		17sfoCig			12InvOak	17snsCig		17smfSpd			
18Offshore	1	18sfoMxrad										
ONSHORE DAY TYPE MODELS												
09Onshore	4		09sfoCig	09rnoPrsDXsfo					09snsVcomp			09snsPrs
11Onshore	4	14sfoMxrad		11rnoPrsDXsfo				11sacCldMax	11acvUcomp			
13Onshore	5		13sfoCig			12InvOak			13sacUcomp 13oakSpd		13oakVis	
15Onshore	8			15rnoPrsDXsfo	15xxx		15oakCig	15hwdCldMax 15ukiCldMax	15sfoUcomp 15sacDir		15oakVis	
16Onshore	7		16sfoCig	16rnoPrsDXsfo					16smfSpd 16snsDir 16snsVcomp	16sfoDwptDprs	16smfWxx	
17Onshore	5	17sfoMxrad			17xxx			17fatCldMax 17lvkCldMax	17lvkVcomp			
18Onshore	4	18sfoMxrad		18rnoPrsDXsfo					18rnoSpd	18sacDwptDprs		

3.4 REAL-TIME PROCESSING

3.4.1 Data Inputs

There is a set of three input files required for each model or models developed for each run-time. Files are stored in subdirectories named by model type under the directory */ll/cv/StatisticalModels/Regional/ascii*. Table 11 identifies the subdirectory names for each run-time/model type combination.

The first file with the *params* extension is a file that specifies the raw data required to derive the predictors of the model. For day type models, this file will also contain an additional reference to the ACV-SFO pressure difference, the day type threshold parameter used by the RSFM. The contents of this file do not change unless the predictors of the model change.

The second file with the *lbls_out* extension lists the symbolic name of each model predictor. For day type models, this file will also contain an additional reference to the ACV-SFO pressure difference. The contents of this file also do not change unless the predictors of the model change. This file is used together with the first file to derive the values of the predictors listed.

The third file with the *md*l extension contains model evaluation parameters for each predictor including predictor weights, parameters that define piece-wise linear interpolation functions and constants. The contents of this file do not change unless the predictors change or the model is retuned with additional data.

3.4.2 Intermediate Outputs/Inputs

Two intermediate output log files are generated in the course of running the RSFM. The first file (*raw* file) contains observations extracted from day summary files according to specifications listed in the input *params* file. This output subsequently becomes input to a program that derives the model predictors listed in the *lbls_out* input file, producing another intermediate output file (*inputdata* file) of predictor name/value pairs. This file along with the *mdl* input file is then input to the model evaluation program to generate a forecast burn off time. Further information on these two intermediate files can be found in Table 3 of the Real-time System and Display Document

3.4.3 Data Outputs

If a successful forecast is produced, the model evaluation program generates an output file containing the model used, burn off time, quality factor, and statistics for each predictor. The output file can be identified with the file suffix *.0.log*. An example of the contents of the output file is shown in Figure 5. Further information on the output file can be found in Table 3 of the Real-time System and Display Document.

TABLE 11

Subdirectories Containing Input Files for Each Run-Time/Model Type Combination of the RSFM

Directory Name	Description
Regional09Onshore	Onshore model (ΔP > 1.0) for 9 GMT
Regional09Neutral	Neutral model (-2.0 >= ΔP < 1.0) for 9 GMT
Regional09Offshore	Offshore model (ΔP < -2.0) for 9 GMT
Regional09Default	Default model for 9 GMT
Regional11Onshore	Onshore model (ΔP > 1.0) for 11 GMT
Regional11Neutral	Neutral model (-2.0 >= ΔP < 1.0) for 11 GMT
Regional11Offshore	Offshore model (ΔP < -2.0) for 11 GMT
Regional11Default	Default model for 11 GMT
Regional13Onshore	Onshore model (ΔP > 1.0) for 13 GMT
Regional13Neutral	Neutral model (-2.0 >= ΔP < 1.0) for 13 GMT
Regional13Offshore	Offshore model (ΔP < -2.0) for 13 GMT
Regional13Default	Default model for 13 GMT
Regional15Onshore	Onshore model (ΔP > 1.0) for 15 GMT
Regional15Neutral	Neutral model (-2.0 >= ΔP < 1.0) for 15 GMT
Regional15Offshore	Offshore model (ΔP < -2.0) for 15 GMT
Regional15Default	Default model for 15 GMT
Regional16Onshore	Onshore model (ΔP > 1.0) for 16 GMT
Regional16Neutral	Neutral model (-2.0 >= ΔP < 1.0) for 16 GMT
Regional16Offshore	Offshore model (ΔP < -2.0) for 16 GMT
Regional16Default	Default model for 16 GMT
Regional17Onshore	Onshore model (ΔP > 1.0) for 17 GMT
Regional17Neutral	Neutral model (-2.0 >= ΔP < 1.0) for 17 GMT
Regional17Offshore	Offshore model (ΔP < -2.0) for 17 GMT
Regional17Default	Default model for 17 GMT
Regional18Onshore	Onshore model (ΔP > 1.0) for 18 GMT
Regional18Neutral	Neutral model (-2.0 >= ΔP < 1.0) for 18 GMT
Regional18Offshore	Offshore model (ΔP < -2.0) for 18 GMT
Regional18Default	Default model for 18 GMT

```
#  -
#  ! Current Time  02-oct-03 15:29:10
#  -

Evaluate for  Time <20031002> => 1065106800 => <02-oct-03 15:00:00> ( Jday 275 )
 Use Time window of   720 seconds



ModelSSum Model Structure Summary
ModelSSum Model/Daytype: Regional15Default

ModelSSum   Term  Name           Weight   Value      Scaled_Value Partial_Sum
ModelSSum   1      15sfoV        0.0471   1.5730     0.6662       0.0314
ModelSSum   2      15sfoCig      -0.3971  182.0000   0.0000       -0.0000
ModelSSum   3      15sqlTh1      1.4310   0.0000     0.0000       0.0000
ModelSSum   4      15sqlRad      -0.4869  0.5000     1.0000       -0.4869
ModelSSum   5      15sfoZi       0.8287   556.0000   0.6297       0.5218
ModelSSum   6      15smbTh1      1.3080   190.0000   0.3317       0.4339
ModelSSum   7      15sfoRad      -1.3960  0.2000     0.2376       -0.3317
ModelSSum   8      Constant_Term 1.0000   5.2908     5.2908       5.2908
ModelSSum   9      GMT_Sunrise   1.0000   14.1359    14.1359      14.1359

 RESULT  /ll/cv/StatisticalModels/Local/ascii/Regional15Default/Regional15Default.mdl
=>   19.595 with result of 0 = < - OK - >  and 0 Substituted values

#<EndOfProcessing>
```

Figure 5. Sample contents of the output file generated by the Model Evaluation Program for Statistically Based Models.

3.4.4 Structural Flow

RSFM processing is illustrated in the functional flow diagram of Figure 6. Processing begins by establishing model parameters based on the current run-time. Parameters include day type and default model names and day type threshold. The logic then queries to see if a day type model exists for the current run-time. If no day type model exists then the default model is used. If the day type model exists, the ACV-SFO pressure difference is used to define the appropriate day-type model that should be invoked. The day-type threshold values are shown in Table 9. If the ACV-SFO pressure difference is not known (either ACV or SFO pressure is missing) then the default model is invoked. Once a model is chosen, the final step before running the model is to insure that data are available to generate the predictors of the model.

The second-half of the processing shown in Figure 6 illustrates the data preparation steps that precede model execution. The first step is to read data from the day summary files sufficient to generate the model predictors. The data to read is governed by an input *params* file that includes references to data that may be used as a substitute for missing data. Table 6 lists the primary and substitute data sources used by the RSFM. The output from the data collection program is a *raw* file. The second step is to generate the model predictors in a name-value format acceptable to the model evaluation program. Model predictor names are read from the input *lbls_out* file together with the (now) input *raw* file to derive predictor values, using substituted values as needed. The resultant output *inputdata* file, containing name-value pairs for each predictor, together with the *mdl* file, containing predictor weights and function definitions, are inputs to the model evaluation program. The output produced by the model evaluation program is written to a *0.log* file and replicated in another directory accessed by the display software. The forecast burn off time and the ACV-SFO pressure difference are also written to an RSFM day summary file to facilitate access by the consensus model.

3.4.5 Nominal and Degraded Performance

When data availability issues prevent one of the RSFM day type models from running, either because the day type readings are unavailable or any of the predictors within the model are missing, a default model is substituted to run in its place. In the event that a predictor is missing that is needed for the default model, the RSFM will not provide a forecast for that hour.

In order for any model to run, all predictors that comprise the model must have valid values. In this respect, there is no configuration in which a model can run in a degraded mode in which one or more predictors are missing.

3.5 PRACTICAL CONSIDERATIONS

3.5.1 Programming and Resources

There are no issues related to programming and resources.

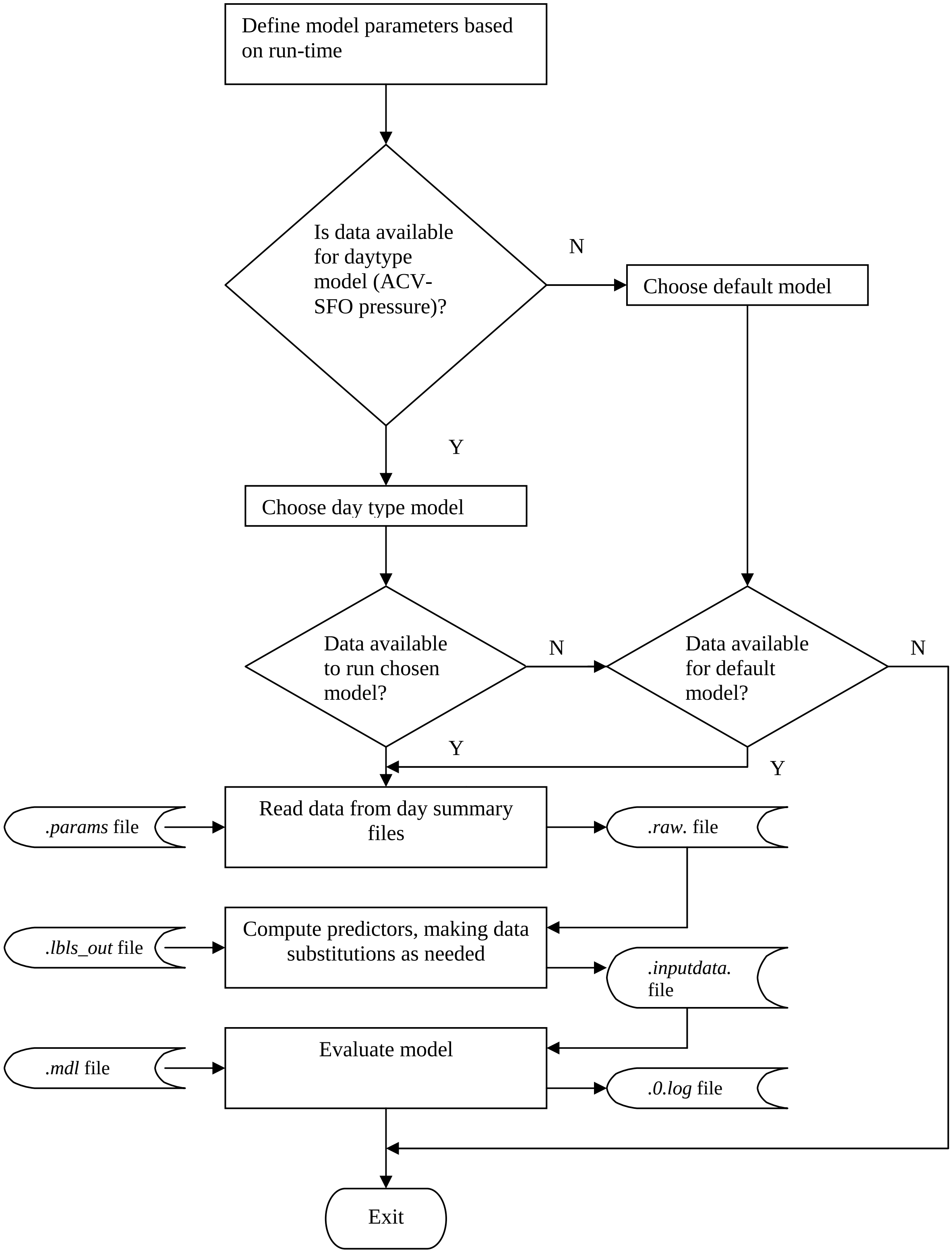


Figure 6. Functional flow diagram for RSFM processing.

3.5.2 Exception Handling

The inability of a default model to run because of the lack of data does not trigger any exception processing. The RSFM simply does not generate a forecast.

3.5.3 Quality Assurance and Diagnostics

Forecast quality, expressed as a quality factor, is an indication of the timeliness of the data used to derive model predictors. A quality factor less than 5 indicates that one or more predictors were derived from data less than nominal in time.

The log files described in Table 3 of the Real-time System and Display Document can be used as a diagnostic tool to determine why an RSFM day type or default model ran or didn't run or why no model ran. The model run control file posts the results of every run and will indicate if a run was unsuccessful due to missing data. To investigate further, the model evaluation program input file (*inputdata* file) can be examined to determine which predictor or predictors were missing and prevented the model from running.

3.5.4 Operator Interaction and Default Settings

The RSFM model is activated by a model-launching script; therefore, no operator interaction is required. Default settings for the RSFM run intervals are shown in Table 4 of the Real-time System and Display Document.

3.5.5 Periodic/Potential Maintenance Tasks

No periodic maintenance is required of the RSFM. If a new model is introduced into the system, the input files described in Section 3.5.1 will need to be updated to reflect the new predictor set. If however, an existing model is simply retuned, only the mdl file requires updating. Programming changes would only be necessary if a new model introduced predictors that are new to the system.

4. SATELLITE STATISTICAL FORECAST MODEL

4.1 INTRODUCTION

This section describes the Satellite Statistical Forecast Model (SSFM) for the San Francisco Marine Stratus Forecast System. The purpose is to provide an overview of the model and how the model is run within the framework of operational system software installed on the base station computer, which is located at the Air Route Traffic Control Center in Fremont, CA.

4.2 TECHNICAL DESCRIPTION

4.2.1 Overview

The SSFM is one of three statistical forecast models that provide real-time forecasts of stratus burn-off in the approach zone to San Francisco International Airport. This model consists of a weighted sum of select observations (or “predictors”) and a constant term. These model equations are created by multiple linear regression.

Visible satellite data is unique within the SFO forecasting system in that it is the only source providing information for every point over the entire San Francisco Bay area. However, this benefit comes at a cost. Given the parameters governing the construction of the satellite image, this amounts to 1600 data points, far too many to be used in statistical model development. A method of reducing the dimensionality of the input data needed to be found. This was accomplished by dividing the area into a set of roughly fifty sectors (Figure 7) and providing summary statistics of their image values. These sector summaries form the pool of candidate statistical predictors. The sectors were chosen based on topographical and operational considerations. The smallest sectors correspond to the SFO approach zone (B1-B6 and W1-W3) to provide detailed characteristics for this important area. To the east of the Bay high elevation sectors are distinguished from those with low elevations. To the west of the Bay, sectors are divided into those east and west of the ridge line, with the sector pairs W5/W10 and W7/W12 designed to capture the gaps in the ridge.

This processing of the raw satellite data includes registration corrections, filtering, and a normalization of values to account for both sun-angle and image degradation effects. The principal types of sector summaries are the mean and standard deviation of sector brightness values, and the fraction of the sector that is covered by cloud. These summaries may also be computed for combinations of sectors if a predictor that covers a larger spatial area is desired. A detailed description of the satellite processing system can be found in the Real-time System document under “Satellite Data Acquisition and Preprocessing.”

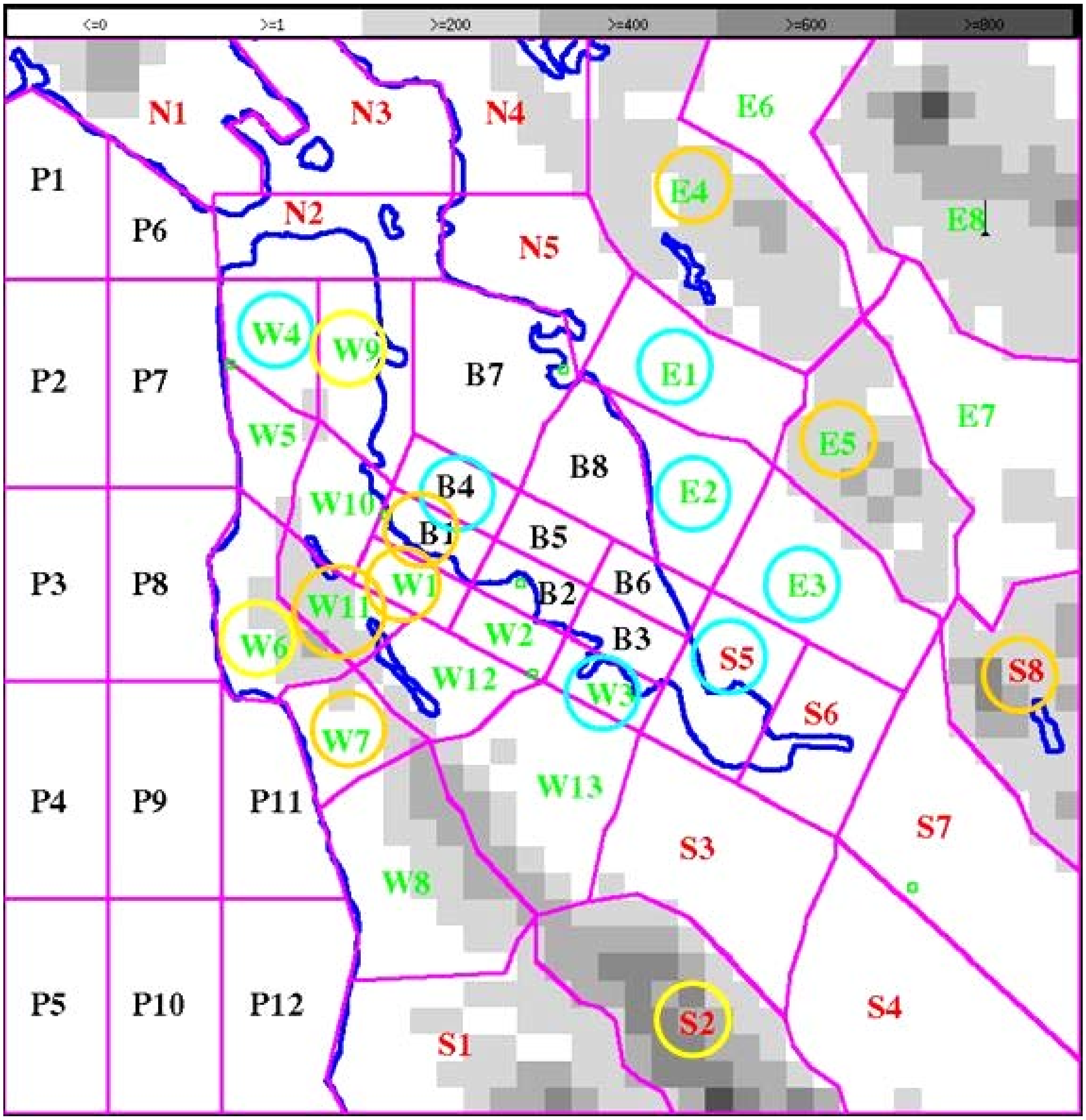


Figure 7. Sector labels and locations used to construct candidate predictors for the SSFM. Circled sectors are those chosen for use in the real-time models. Orange circles indicate those used in the 15 Z primary models, blue circles those used for 16Z – 18Z, and yellow circles those common to all primary models.

The SSFM has one predictor that does not originate in satellite imagery. It is a theoretical maximum radiation term for SFO. It functions as a seasonal term, representing the varying strength of incoming solar radiation from a maximum near the summer solstice to a much smaller value at the end of the stratus season.

The SSFM makes forecasts at 15Z, 16Z, 17Z and 18Z when marine stratus is present. (The other component models make forecasts at 9Z, 11Z, and 13Z as well, but no visible satellite data are available at these pre-dawn times.) The predictors for a given hour consist primarily of data from the same hour, though occasionally a trend predictor (the difference between the data from this hour and the hour previous) is used.

4.2.2 Day Types

At each hour, one of several different SSFMs may be used depending on the current "day type." In developing the SSFM, the cases used to determine the predictor weights were subdivided in an effort to distinguish between the majority of "typical" stratus cases and that minority exhibiting some non-typical complicating factor. This was found to provide more accurate forecasts for the typical cases, and allows for the identification of those other cases where the forecasts are likely to be less accurate. Two parameters are used to determine the day type. One is the standard deviation of all cloud-covered Pacific Ocean sectors (see Figure 7). This measures the visible texture of the clouds apart from any topographical influences. Typical stratus cases exhibit a "smooth" cloud deck. A "rough" cloud deck, with standard deviations ≥ 0.05, often indicates that some non-local dynamic, possibly synoptic in scale, is at work. The other parameter used is the height of the cloud ceiling in and near the SFO approach zone. (More specifically, the minimum of the SFO, SQL, and SMB ceilings is used.) It was found that the SSFM provided poorer forecasts for high-ceiling cases (> 500 m), possibly because for these cases the cloud deck is completely above the ridge of the San Francisco Peninsula to the west. These parameters are used to create three distinct day types (and thus three distinct model equations for each hour). The typical type has a low ceiling and exhibits a smooth cloud deck (labeled "LoCigSmthCld"). Two non-typical types are created: one with high ceiling ("HiCig"), the other with low ceiling but a rough cloud deck ("LoCigRufCld"). If the day type cannot be determined, the system attempts to use the "default" model (i.e., one that was developed using all of the cases). Also, for 18Z, there were too few stratus cases to allow the development of reliable day type models. For this hour, only a default model is used.

4.2.3 Predictor Selection

The mean, standard deviation, and ratio of cloud coverage are computed for each sector. Much experience and experimentation in developing the SSFM has led us to only consider the sector mean values as predictors (without thereby sacrificing forecast model performance). Standard deviations are susceptible to ground effects for partially cloud covered sectors and do not always purely describe the cloud texture. The ratio of cloud coverage is very sensitive to image values close to the cloud / no-cloud threshold and was found to be an unreliable predictor. Changes in sector mean values from the prior hour are also calculated. Though intuitively important, the automated process that chooses the most important predictors almost always failed to select these trend predictors. Attempts to force more of these trend predictors into the models resulted in a marked drop in performance.

Repeated experience in developing the SSFM pointed to several areas of sectors that were consistently chosen as predictors. These are indicated in Figure 7. These were the approach zone, areas east of the Bay, the highest terrain in the domain at its southernmost edge, and central and northern portions of the San Francisco Peninsula. The approach zone was important due to its proximity to what is being predicted. Areas to the east likely signal the erosion of stratus during typical cases. Areas to the west possibly give some reflection of advection of cloud and on shore flow. The high terrain in the south provides some indication of the depth of the stratus, with this sector (S2) being cloud covered only when the stratus is especially deep.

To ensure a measure of stability in the forecasts from hour to hour, the primary models ("LoCigSmthCld" and "Default") were forced as far as possible to have identical predictors that represented these general areas. As shown in Figure 7, a somewhat different set of predictors was used for 15Z as opposed to 16-18Z as the latter include a trend predictor (for a grouping of sectors E1, E2, and E3) that could not be used for 15Z due to the absence of reliable 14Z imagery. These predictors were retuned resulting in different weights for each forecast hour.

The non-typical day types ("HiCig" and "LoCigRufCld") could not be forced to have these same predictor sets either because their relatively small number of development cases would not allow the fitting of the 7 to 8 standard predictors, or because forcing the same predictor set resulted in markedly poorer models. This is not surprising, as the evolution of stratus for a non-typical case would likely be indicated by different areas of the domain than for a typical case.

4.3 REAL-TIME PROCESSING

4.3.1 Data Inputs

For each run-time a set of input files is required to specify day type parameters, model predictors, and the statistical model itself. These files are stored in subdirectories under */ll/cv/StatisticalModels/Satellite/ascii.* Table 12 identifies and describes the various subdirectories that are used.

Each model requires three files, which are contained in the model's subdirectory. The first file, with the *params* extension, specifies the raw data required to derive the predictors of the model. Most of them are based on the current satellite image, with a few used to determine the maximum radiation predictor. The contents of this file do not change unless the predictors of the model change.

TABLE 12

Subdirectories Containing Input Files for Each Run-Time/Model Type Combination of the SSFM

Directory Name	Description
Satellite15DayType	Used to determine 15GMT day type parameters
Satellite15LoCigSmthCld	Day type 15GMT model used for typical cases of approach zone cloud ceiling < 500 m and Pacific cloud standard deviation < 0.05
Satellite15LoCigRufCld	Day type 15GMT model used for non-typical cases of approach zone cloud ceiling < 500 m but Pacific cloud standard deviation ≥ 0.05
Satellite15HiCig	Day type 15GMT model used for non-typical cases of approach zone cloud ceiling ≥500 m (regardless of Pacific cloud standard deviation).
Satellite15Default	Default 15GMT model used when day type cannot be determined or missing data prevents the day type model from running.
Satellite16DayType	Used to determine 16GMT day type parameters
Satellite16LoCigSmthCld	Day type 16GMT model used for typical cases of approach zone cloud ceiling < 500 m and Pacific cloud standard deviation < 0.05
Satellite16LoCigRufCld	Day type 16GMT model used for non-typical cases of approach zone cloud ceiling < 500 m but Pacific cloud standard deviation ≥ 0.05
Satellite16HiCig	Day type 16GMT model used for non-typical cases of approach zone cloud ceiling ≥500 m (regardless of Pacific cloud standard deviation).
Satellite16Default	Default 16GMT model used when day type cannot be determined or missing data prevents the day type model from running.
Satellite17DayType	Used to determine 17GMT day type parameters
Satellite17LoCigSmthCld	Day type 17GMT model used for typical cases of approach zone cloud ceiling < 500 m and Pacific cloud standard deviation < 0.05
Satellite17LoCigRufCld	Day type 17GMT model used for non-typical cases of approach zone cloud ceiling < 500 m but Pacific cloud standard deviation ≥ 0.05
Satellite17HiCig	Day type 17GMT model used for non-typical cases of approach zone cloud ceiling ≥500 m (regardless of Pacific cloud standard deviation).
Satellite17Default	Default 17GMT model used when day type cannot be determined or missing data prevents the day type model from running.
Satellite18Default	Default 18GMT model

The second file, with the *lbls_out* extension, lists the symbolic name of each model predictor. The contents of this file also do not change unless the predictors of the model change. This file is used together with the first file to derive the values of the predictors listed.

The third file, with the *md*l extension, contains model evaluation parameters for each predictor including predictor weights, parameters that define piece-wise linear scaling functions, and constants. The contents of this file do not change unless the predictors change or the model is retuned with additional data.

In addition, for 15Z, 16Z, and 17Z a "Day Type" directory is present. These contain two files that allow the computation of parameters that are used to determine the day type. These files are the *params* and *lbls_out* files described above.

4.3.2 Intermediate Outputs/Inputs

Two intermediate output log files are generated in the course of running the SSFM. The first file (*raw* extension) contains observations extracted from day summary files according to specifications listed in the input *params* file. This output subsequently becomes input to a program that derives the model predictors listed in the *lbls_out* input file, producing another intermediate output file (*inputdata* extension) of predictor name/value pairs. This file along with the *mdl* input file is then input to the model evaluation program to generate a forecast burn off time. Further information on these two intermediate files can be found in Table 3 of the Real-time System document.

4.3.3 Data Outputs

If a successful forecast is produced, the model evaluation program generates an output file containing the model used, burn off time, quality factor, and statistics for each predictor. The output file can be identified with the file suffix *.0.log*. An example of the contents of the output file is shown in Figure 8. Further information on the output file can be found in Table 3 of the Real-time System and document.

The predictor names shown in the file are constructed according to the following convention. The first two digits refer to the GMT hour of the observation. Then follows a set of characters up to the next capital character. These refer to the location of the predictor. For satellite observations these are the sector label, or labels if the predictor is formed by a combination of sectors (Figure 8). The remaining group of characters indicates the content of the predictor. For satellite predictors this can be M (mean brightness), SD (standard deviation of brightness), or CV (fraction of sector, or sectors, covered by cloud). A trailing set of characters consisting of "dtN" may also be present, indicating that the predictor is a time difference over the past N hours.

4.3.4 Structural Flow

The SSFM process is summarized in Figure 9 as a functional flow diagram. The process can roughly be divided into two parts. The first consists of deciding, based on day type parameters and available models, if a day type or default model is to be run. If the former, then the proper day type model is chosen by comparing the day type parameters (cloud ceiling, visible satellite standard deviation) with their respective thresholds (also detailed in Table 12). If the run time is 18Z, then the default model is chosen, as it is the only one available at this hour.

```
#  -
#  ! Current Time  30-dec-03 21:41:43
#  -

Evaluate for  Time <20031003> => 1065193200 => <03-oct-03 15:00:00> ( Jday 276 )
 Use Time window of   180000 seconds



ModelSSum Model Structure Summary
ModelSSum Model/Daytype:  Satellite15LoCigSmthCld
ModelSSum   Term          Name        Weight        Value    Scaled_Value   Partial_Sum
ModelSSum     1       15sfoMxrad     0.8027      95.3000       0.1919        0.1540
ModelSSum     2       15e4e5s8M      1.9256       0.6267       0.1962        0.3778
ModelSSum     3           15s2M      1.9675       0.4760       0.1696        0.3337
ModelSSum     4         15w1b1M      1.3537       0.5395       0.1508        0.2042
ModelSSum     5          15w11M      0.5763       0.6490       0.2371        0.1366
ModelSSum     6         15w6w7M     -0.4822       0.6944       0.1103       -0.0532
ModelSSum     7           15w9M     -0.2123       0.6690       0.3196       -0.0678
ModelSSum     8   Constant_Term      1.0000       2.7342       2.7342        2.7342
ModelSSum     9     GMT_Sunrise      1.0000      14.1500      14.1500       14.1500
 RESULT
/11/cv/StatisticalModels/Satellite/ascii/Satellite15LoCigSmthCld/Satellite15LoCigSmthCld.mdl
=>   17.969 with result of 0 = < - OK - >  and 0 Substituted values

#<EndOfProcessing>
```

Figure 8. Sample output file generated by the Model Evaluation Program for Statistically Based Models.

The second half of the processing shown in Figure 9 illustrates the data preparation steps that precede model execution. The first step is to read data from the day summary files sufficient to generate the model predictors. The data to read is governed by an input *params* file, which may include references to data that may be used as a substitute for missing predictors. (For instance, in the SSFM the maximum solar radiation at SQL is used as a substitute when that observation is missing for SFO.) The output from the data collection program is a *raw* file. The second step is to generate the model predictors in a name-value format acceptable to the model evaluation program. Model predictor names are read from the input *lbls_out* file together with the (now) input *raw* file to derive predictor values, using substituted values as needed. The resultant output *inputdata* file, containing name-value pairs for each predictor, together with the *mdl* file, containing predictor weights and function definitions, are inputs to the model evaluation program. The output produced by the model evaluation program is written to a *0.log* file and replicated in another directory accessed by the display software. The forecast burn off time and quality factor are also written to an SSFM day summary file to facilitate access by the consensus model.

4.3.5 Nominal and Degraded Performance

When data availability issues prevent one of the SSFM day type models from running, a default model is substituted to run in its place.

In order for any model to run, all predictors that comprise the model must have valid values. In this respect, there is no configuration in which a model can run in a degraded mode in which one or more predictors are missing.

4.4 PRACTICAL CONSIDERATIONS

4.4.1 Programming and Resources

There are no issues related to programming and resources.

4.4.2 Exception Handling

The inability of a default model to run because of the lack of data does not trigger any exception processing. The SSFM simply does not generate a forecast.

4.4.3 Quality Assurance and Diagnostics

Forecast quality, expressed as a quality factor, is an indication of the timeliness of the data used to derive model predictors. A quality factor less than 5 indicates that one or more predictors were derived from data less than nominal in time.

The log files described in Table 3 of the Real-time System document can be used as a diagnostic tool to determine why an SSFM day type or default model ran or didn't run or why no model ran. The model run control file posts the results of every run and will indicate if a run was unsuccessful due to missing data. To investigate further, the model evaluation program input file (*inputdata* file) can be examined to determine which predictor or predictors were missing and prevented the model from running.

4.4.4 Operator Interaction and Default Settings

The SSFM is activated by a model launching script so no operator interaction is required. Default settings for the SSFM run intervals are shown in Table 4 of the Real-time System document.

4.4.5 Periodic/Potential Maintenance Tasks

No periodic maintenance is required of the SSFM. If a new model is introduced into the system, the input files described in Section 4.3.1 will need to be updated to reflect the new predictor set. If however, an existing model is simply retuned, only the *mdl* file requires updating. Programming changes would only be necessary if a new model introduced predictors new to the system.

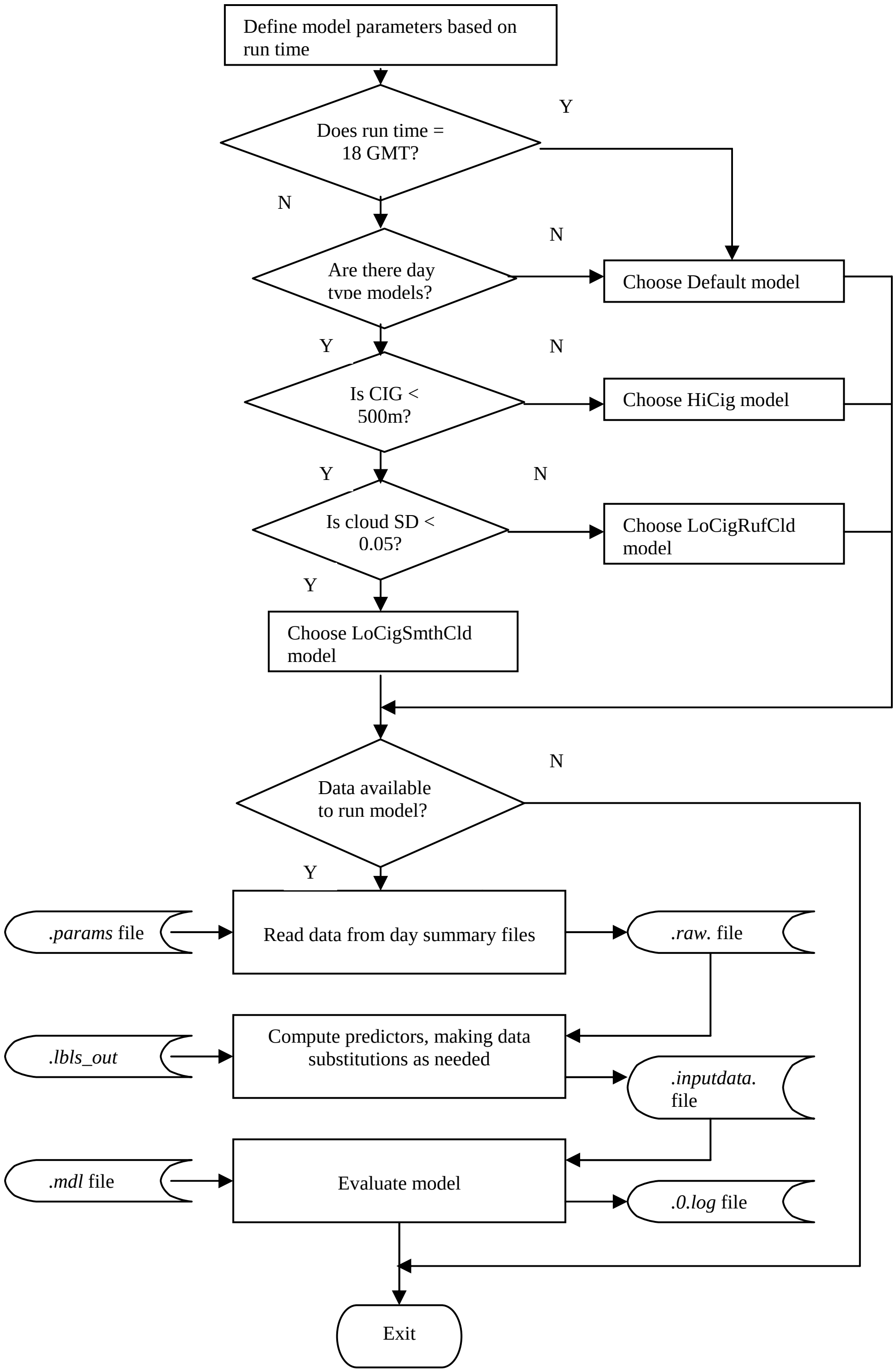


Figure 9. Functional flow diagram for SSFM processing.

5. COBEL FORECAST MODEL DESCRIPTION DOCUMENT

5.1 INTRODUCTION

This document describes the version of the COBEL model adapted by the University of Quebec at Montreal (UQAM) for the San Francisco Marine Stratus Forecast System. The purpose of the document is to provide an overview of the model and how the model is run within the framework of operational system software installed on the base station computer, which is located at the Air Route Traffic Control Center in Fremont, CA.

5.1.1 Related Documents

A companion document entitled "Real-time System" details the operational system software responsible for collecting required forecast model data and activating the component models, including the COBEL model. An in-depth scientific description of the COBEL model is provided by Turner et al. [2].

5.2 HISTORICAL PERSPECTIVE

The COBEL (French for Couche Brouillard Eau Liquide) model is a very-high resolution one-dimensional numerical model of the planetary boundary layer. The original model was developed in the late-1980's at the Laboratoire d'Aérologie, Paul Sabatier University of Toulouse, France to study the nocturnal boundary layer. Later in 1993, COBEL was adapted as a radiation fog-forecasting tool in the Nord-Pas de Calais region of France. In 1999, UQAM extended the 1993 adaptation of COBEL to simulate the evolution and burn off marine stratus. The UQAM version of COBEL was first introduced to the San Francisco Marine Stratus System in 2000.

5.3 TECHNICAL DESCRIPTION

5.3.1 Overview

The COBEL model is unique to the San Francisco Marine Stratus System in that it is the only component forecast derived from a physics-based model. The model provides high-resolution forecasts of the structure of the marine boundary layer, including cloud water content of the marine stratus. A forecast of the burn off time for the stratus is obtained when the simulated total water content of the cloud goes to zero.

The model is initialized with a vertical profile of temperature, humidity, and wind at San Francisco International Airport (SFO). The initialization uses a hybrid of the Oakland (OAK) balloon sounding and high-resolution lower atmospheric measurements at SFO. This profile is used to characterize the stratus cloud deck in terms of total vertically integrated liquid water content. As the model steps forward, physical processes are applied within the model to anticipate the evolving decrease of the cloud liquid water with time. The model forecast time of cloud dissipation is declared when the liquid water content reaches zero.

Since the model is adapted for a single point location, the forecast for that location (SFO) is then adjusted to account for the spatial difference from the area of interest (i.e., the approach zone into SFO). This is done via historical analysis of the relationship between cloud coverage at the model point location and the recorded transition to dual approaches (the so-called side-by time). Thus, the COBEL forecast time shown on the display represents the expected time of clearing in the approach zone.

The model is re-initialized for each model run; using updated high-resolution measurements of surface temperature, humidity, wind, and solar radiation. In particular, the re-initialization includes an adjustment to ensure that the model's interpretation of liquid cloud water amount is consistent with the amount of solar radiation transmitted through the cloud layer. The model is then re-run, providing a new forecast of stratus burn off.

The operational version of COBEL for San Francisco is designed for typical stratus days when stratus is present and fully developed at or before sunrise, when there is a marked strong inversion at the top of a shallow well-mixed marine boundary layer, and when no clouds are present above the boundary layer. The model includes (via measurement or parameterization) all of the physical processes important for the dissipation of marine stratus including solar and terrestrial radiation, turbulence, as well as cloud and rain processes. Although designed to input horizontal advection of temperature and humidity and vertical motion from meso-scale models, the current operational version assumes no horizontal advection and a fixed value for subsidence of (0.7 cm/s). A horizontal pressure force, corresponding to a 2 m/s geostrophic wind, is arbitrarily applied over the entire depth of the model domain (surface to 1.5 km height) and kept constant in time.

5.3.2 Predawn Forecasts

In 2002, two additional forecast run intervals were added to the COBEL model as one component in a suite of models to support a predawn forecast outlook product. This product is designed to generate forecasts prior to 12GMT and be made available to CWSU forecasters who brief air traffic control personnel at 12GMT. The two run intervals added were 9 and 11GMT; however, the actual ability of the COBEL model to produce forecasts at these hours depends on the availability of a reliable estimate of cloud top as determined from sodar data as well as surface observations at SFO.

The predawn initialization procedure in the COBEL model differs from the 12GMT procedure in that there is no coincident Oakland sounding. The steps of the initialization procedure are outlined below:

Determine the cloud top. Cloud top is taken from the estimate of inversion base height (Zi) as derived from the consensus Zi algorithm if the value is within the range 100 – 1200m.

Compute the initial liquid water content profile in the cloud. The cloud top and the maximum cloud water content (set to 0.5 g/kg, the threshold for the production of drizzle) are assumed to coincide with the boundary layer inversion. The vertical profile is assumed to be adiabatic.

Determine the temperature profile from the cloud top to the surface. The temperature profile must match the observed temperature at 10 m. This is accomplished by an iterative process that changes the temperature at the top of the cloud, under the assumption that the temperature profile is pseudo-adiabatic in the cloud and dry-adiabatic under the cloud. The cloud water content is iteratively calculated to match the modified temperature profile.

Compute the specific humidity profile from cloud base to the surface. At cloud base saturation is assumed and the specific humidity profile is determined by interpolating between the saturation value at cloud base and the observed value at 10 m (computed from the observed dew point temperature).

Free atmosphere (FA) thermodynamic state. The free atmosphere is assumed to be the same at Oakland and SFO. A default sounding (taken from a historical Oakland sounding representing a "typical" stratus day) is used to initialize the temperature and specific humidity profiles above the cloud top (the inversion base). The FA humidity and temperature profiles are matched with the boundary layer profiles determined at prior steps. The matched region is defined by the lowest level in the Oakland sounding above the boundary layer inversion (first vertical level in the free atmosphere) and the boundary layer at SFO. The match is achieved by linear interpolation between the values at this point and the values at the SFO cloud top.

Once initialized, the model steps forward in time; the solar radiation adjustment step is not applied, of course, prior to sunrise. Since the model is considered to be operating at a somewhat degraded mode prior to sunrise, it is only run when a high quality estimate of inversion base height (cloud top) is available from the consensus inversion algorithm.

5.3.3 Forecasts After Sunrise

After sunrise each day, COBEL is initialized using the available Oakland 12GMT sounding and local data from project sensors. First the Oakland sounding is adapted for SFO using the following procedure:

The initialization uses the temperature and humidity structure above the marine layer from the sounding and places it above the cloud top, which is assumed to be the inversion base height (Zi) computed from the SFO sodar.

The cloud maximum liquid water content is again assumed to be 0.5 g/kg at cloud top and decreases downward according to the adiabatic liquid water content rate. Cloud base is determined as the altitude at which the liquid water content becomes zero.

The entire temperature profile is fixed using the 10m SFO temperature and assuming that the temperature is dry adiabatic from the surface to cloud base, moist adiabatic in the cloud layer, and assuming the lapse rates from the Oakland sounding above the inversion base.
Sub-cloud humidity is set by linear interpolation between saturation at cloud base and the observed dew point temperature near the surface (10m).

The wind is initialized as being geostrophic and decreases linearly from approximately the top of the surface layer to zero at the surface. The effective cloud droplet radius is fixed for the entire forecast at 5.5 microns.

Every hour from 13Z on, COBEL is reinitialized using filtered local measurements (to remove high-frequency fluctuations) of temperature, dew point, wind, and solar radiation to generate a new forecast.

For the 0 to 2 hour interval following sunrise, the COBEL hourly reinitialization assumes that the cloud layer is mature during the interval. Consequently, the same procedure as the one applied at 12GMT is used, but instead of using the sounding as a first guess, model profiles produced from a previous short-forecast simulation are used. Steps 1 through 4 above are performed using updated observations with the exception of wind speed, which retains its first-guess value.

Beyond the 2-hour interval following sunrise, the procedure used to determine the cloud water content (step 2 above) is modified to take into account that the cloud layer may be in its dissipation stage (thinner than a mature cloud layer). The Lifted Condensation Level (LCL) is determined from near-surface (10m) measurements of temperature and dew point temperature. The LCL is taken as cloud base. The cloud water content is then set to 0 at cloud base and increased, following the adiabatic liquid water content rate, up to a maximum value at a height corresponding to cloud top, as determined by sodar (Zi). Once thermodynamic profiles have been set by this procedure, the cloud droplet effective radius (representing the cloud drop size distribution) is modified so that the model computed solar flux matches (+/- 1 W/m2) the observed surface down welling short-wave radiation at SFO. This effective radius is then used during the forecast run.

5.4 REAL-TIME PROCESSING

5.4.1 Data Inputs

The COBEL model requires three sources of observation data for the predawn and 13-18GMT runs, valid at the forecast run time: the SFO ASOS observation report, the SFO field site observation report, and an estimate of the inversion base height (cloud top) derived from sodar data. Runs beginning at 14GMT also use the surface observation of ceiling at San Carlos Airport (SQL) to select the SFO-to-approach zone burn off adjustment when it is timely, but will not prevent COBEL from running if it is not available. Predawn forecasts and the 13Z forecast have fixed burn off adjustments that do not require the SQL surface observation. The 12GMT run of COBEL requires the 12GMT Oakland sounding in addition to the three data sources mentioned above. If the 12GMT Oakland sounding is not available, COBEL will not run beyond any predawn runs that may have taken place.

COBEL also utilizes static input files that are part of the delivered COBEL software package. Table 13 lists these files.

5.4.2 Intermediate Inputs

The COBEL model generates intermediate input files derived from the observations mentioned above. SFO ASOS and field site data are filtered to remove high-frequency fluctuations and are stored in a private day summary directory hierarchy under */ll/cv/cobel/obs*. This includes temperature, dew point temperature, and pressure for the ASOS data; and temperature, dew point temperature, down welling short wave radiation, and net short wave radiation for field site data.

TABLE 13

COBEL Static Input Files Located in the Root Directory *//ll/cv/cobel*

File Name	Description
cobel_files/cobel_grille_30m	COBEL vertical grid.
cobel_files/default_sounding	Sounding used for predawn forecasts.
cobel_files/weight	Weights used by the filtering program.
radcoeff/moska1	Coefficients used in infrared radiation calculation.
radcoeff/pade.sal	Coefficients used in solar radiation calculation.

Various intermediate input files are also generated and placed in dated directories within */ll/cv/cobel/input*. These files are summarized in Table 14.

5.4.3 Data Outputs

The COBEL model places output in dated directories within the root directories */ll/cv/cobel/output* and */ll/cv/logs/YYYYMMDD/CobelModel.* Tables 15 and 16 list the critical files written to these two directories along with a description of their contents. Additional output files written to these directories are used strictly by UQAM staff for diagnostic study and are not documented.

TABLE 14

COBEL Dynamic Input Files Located in the Root Directory

/ll/cv/cobel/input/YYYYMMDD

File Name	Description
cobel_demarr.dat	COBEL configuration file.
cobel_grille_30m_YYYYMMDD	COBEL vertical grid.
inieau_HH00z	Cloud water profile.
iniq_HH00z	Water vapor mixing ratio profile.
init_HH00z	Temperature profile.
iniz_HH00z	COBEL grid heights.
ts_HH00z	Soil temperature profile.
info/cldtop_HH00z	Value of consensus Zi product.
info/inidewp_HH00z	Initial dew point temperature profile.
info/pres_ceil_HH00z	Surface pressure.
info/sounding_HH00z	Initial OAK sounding.
info/t_td_HH00z	T and Td from the SFO field site.

5.4.4 Structural Flow

COBEL model processing begins with the activation of the model launching script illustrated in the functional flow diagram of Figure 10a. For the first invocation of the day, the script begins by creating a log directory that will be used to store log files generated for each forecast interval run that is made. The directory is created under the path */ll/cv/logs/YYYYMMDD* and is named with the model name reflected in the model-launching script (in this case, Cobel). Next, a list of model-specific forecast run intervals is defined. The list is referenced to guide the forecast model activation process to attempt to produce forecasts for only those run times specified in the list. The list is fixed within the script but could easily be changed in the unlikely event that forecasting requirements change. Table 17 lists the forecast run times for COBEL, including a breakdown of the simulations required for each forecast run interval and the prerequisite conditions necessary for a simulation to run. Note that the list represents only a guide as to the forecast intervals requested to be run. The ability to actually produce a forecast for a particular run time and model is determined by data availability.

Once a list of run times is established, the logic chooses the first run time from the list. The run time is then checked against the last run time on the list to see if all runs have been made. A check is also made to see if the run time actually appears on the list. If a run time does not appear on the list, it is advanced by one hour and the logic is repeated. Otherwise, the model run control file is queried to see if a run with this run time has been made. If not, a run is made if the current time is equal to or exceeds the run time. If at least one run has already been made, the current time is checked to see if it is time to move on to the next run time. If so, the current run time is advanced by one hour and the logic is repeated. Processing continues for predawn forecast intervals, for the 12GMT run interval so long as the 12GMT Oakland sounding is available, and for successive forecast intervals beyond 12GMT as long as the previous run interval ran successfully.

The additional processing referenced in Figure 10a is expanded in detail in Figure 10b. Processing enters this stage if the previous logic determines that a run should be made. Checks are performed to insure that COBEL has the necessary input data to run. This includes the SFO ASOS observation report, SFO field site observation report, and consensus inversion base height with non-zero quality. The absence of any of these data sources will cause processing to exit with the intent that missing sources will become available at a later time. Processing may also be delayed in running the COBEL model to wait for the arrival of the hourly observation of ceiling from San Carlos (SQL) (up to 30 minutes past the hour) to accommodate the burn off adjustment algorithm as outlined in Section 3.3 above. Once all data sources are present and either the SQL observation has arrived or the wall clock time exceeds 30 minutes past the hour, the COBEL model is run. Following a run, a burn off adjustment may be applied based on the forecast run interval. Table 18 shows the linear adjustment formula used and the offset that varies with forecast run time.

TABLE 15

Output Files Generated by the COBEL Model and Placed in the Directory *//ll/cv/logs/YYYYMMDD/CobelModel*

Generic File Name	Description
out_prepare_input_YYYYMMDD	Contains output from the script */ll/cv/cobel/scripts/prepare_input.* Generated once during the 12GMT run.
err_prepare_input_YYYYMMDD	Contains error information from the script */ll/cv/cobel/scripts/prepare_input* if the script encounters errors. Otherwise, contains limited non-error-related information.
out_prevision_YYYYMMDD_12Z_radclr	Contains output from running the COBEL model in prevision mode for 12 hours for the 12GMT run.
err_prevision_YYYYMMDD_12Z_radclr	Contains error information for the COBEL model run if errors were encountered. Otherwise contains 2 blank lines.
out_prevision_YYYYMMDD_12Z_1Hr	Contains output from running the COBEL model in prevision mode for 1 hour for the 12GMT run.
err_prevision_YYYYMMDD_12Z_1Hr	Contains error information for the COBEL model run if errors were encountered. Otherwise contains 2 blank lines.
out_prevision_YYYYMMDD_12Z_TBo	Contains output from running the COBEL model in prevision mode from 12GMT until burn off.
err_prevision_YYYYMMDD_12Z_TBo	Contains error information for the COBEL model run if errors were encountered. Otherwise contains 2 blank lines.
out_prevision_YYYYMMDD_HHZ	Contains output from running the COBEL model for 1) 1 hour and 2) until burn off, beginning at time HH.
err_prevision_YYYYMMDD_HHZ	Contains error information for the COBEL model run if errors were encountered. Otherwise contains 4 blank lines.

TABLE 16

Output Files Generated by the COBEL Model and Placed in the Directory */ll/cv/cobel/output/YYYYMMDD*

File Name	Description
Burntime	Contains the unadjusted burn off time from the latest long run (until burn off) of the COBEL model or 9999.000 if COBEL failed to produce a forecast.
Burntimep	Contains the unadjusted burn off time from the latest short run (1 hour) of the COBEL model or 9999.000 if COBEL failed to produce a forecast.
burntotfile_prevision_YYYYMMDD	Contains a list of unadjusted burn off times updated throughout the day,
init_restart_prevision_YYYYMMDD_HHZ	Restart initialization file. Created during the one-hour short run and read during the follow-on burn off run. Also read during the next forecast run.

TABLE 17

Summary of COBEL Run Sequence

Forecast Run Interval (GMT)	Time Interval of Simulation	Condition(s) to Run
09	12 – Burn Off	Consensus Zi at 09GMT
11	12 – Burn Off	Consensus Zi at 11GMT
12	12 hrs.	No Conditions
12	12 – 13	No Conditions
12	12 – Burn Off	12-13 Run OK
13	13 – 14	12-13 Run OK
13	13 – Burn Off	13-14 Run OK
14	14 – 15	13-14 Run OK
14	14 – Burn Off	14-15 Run OK, Burn Off > 1hr. in the future
15	15 – 16	14-15 Run OK
15	15 – Burn Off	15-16 Run OK, Burn Off > 1hr. in the future
16	16 – 17	15-16 Run OK
16	16 – Burn Off	16-17 Run OK, Burn Off > 1hr. in the future
17	17 –18	16-17 Run OK
17	17 – Burn Off	17-18 Run OK, Burn Off > 1hr. in the future
18	18 – 19	17-18 Run OK
18	18 – Burn Off	18-19 Run OK, Burn Off > 1hr. in the future

TABLE 18
Values of the Offset (b) for the Linear Adjustment of the COBEL Forecast to Burn Off Time in the Approach Zone

Forecast Time (GMT)	No Ceiling at SQL	Ceiling at SQL	All Conditions
09			0.17
11			0.11
13			-0.13
14	-0.63	0.10	0.01
15	-0.49	0.48	0.38
16	-0.09	0.63	0.47
17	-0.18	0.76	0.54
18	-0.24	0.63	0.35

Adjustment Equation: $F_{az} = F_c * a + b$

Where:

F_{az} Adjusted forecast of time of burn off for the approach zone
F_c COBEL forecast of time of burn off for SFO
a Derived slope (constant = 1.0)
b Derived offset

Output from the model is written to a number of areas:

Burn off time, quality factor, and model type are written to:
/ll/cv/data/cvRTCache/LATEST/latest_Cobel_forecast

Burn off time, quality factor, and model type are appended to the forecast model day summary file:
/ll/cv/data/daysumYYYY/fcsts/YYMMDD/fcsts_Cobel.YYMMDD.sum

Forecast run results are appended to the model run control file:
/ll/cv/logs/YYYYMMDD/CobelModel/CobelRunSummary.YYYYMMDD

Figures 11 through 15 show expanded detail of COBEL model processing referenced from Figure 10a and 10b. The figures distinguish predawn, 12GMT, 13-14GMT, and 15-18GMT runs because of differences related to input data preparation and sounding adaptation processing (predawn and 12GMT), the inclusion of a 1-hour short-forecast simulation to produce model-derived temperature and moisture profiles (post 12GMT runs), and runs that modify the initialization restart file with data from current observations (15-18GMT runs). Figure 12 shows an expanded view of input data preparation steps performed for the predawn and 12GMT runs.

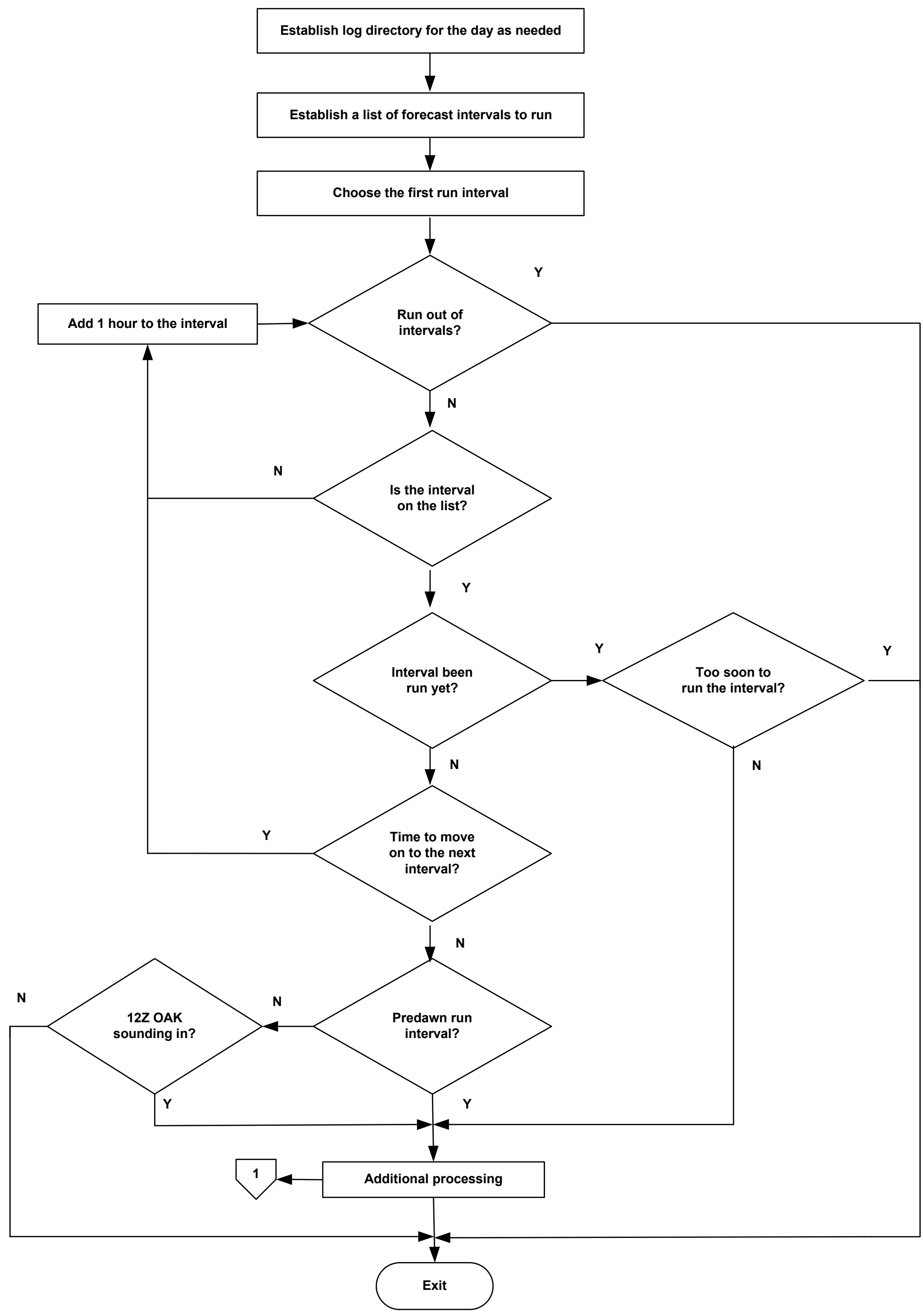


Figure 10a. Functional flow diagram of the COBEL model launching script.

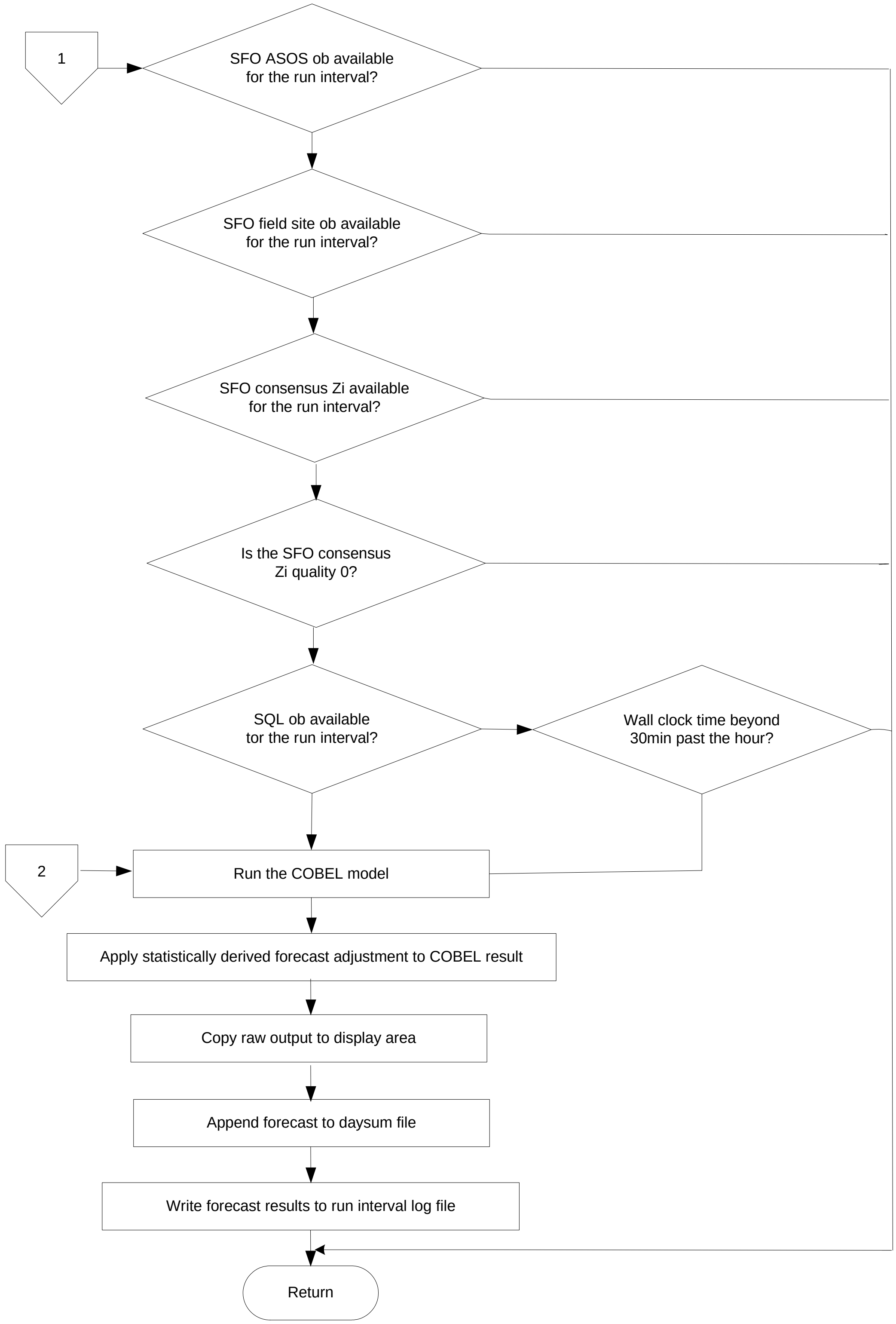


Figure 10b. Functional flow diagram of the COBEL model launching script (continued).

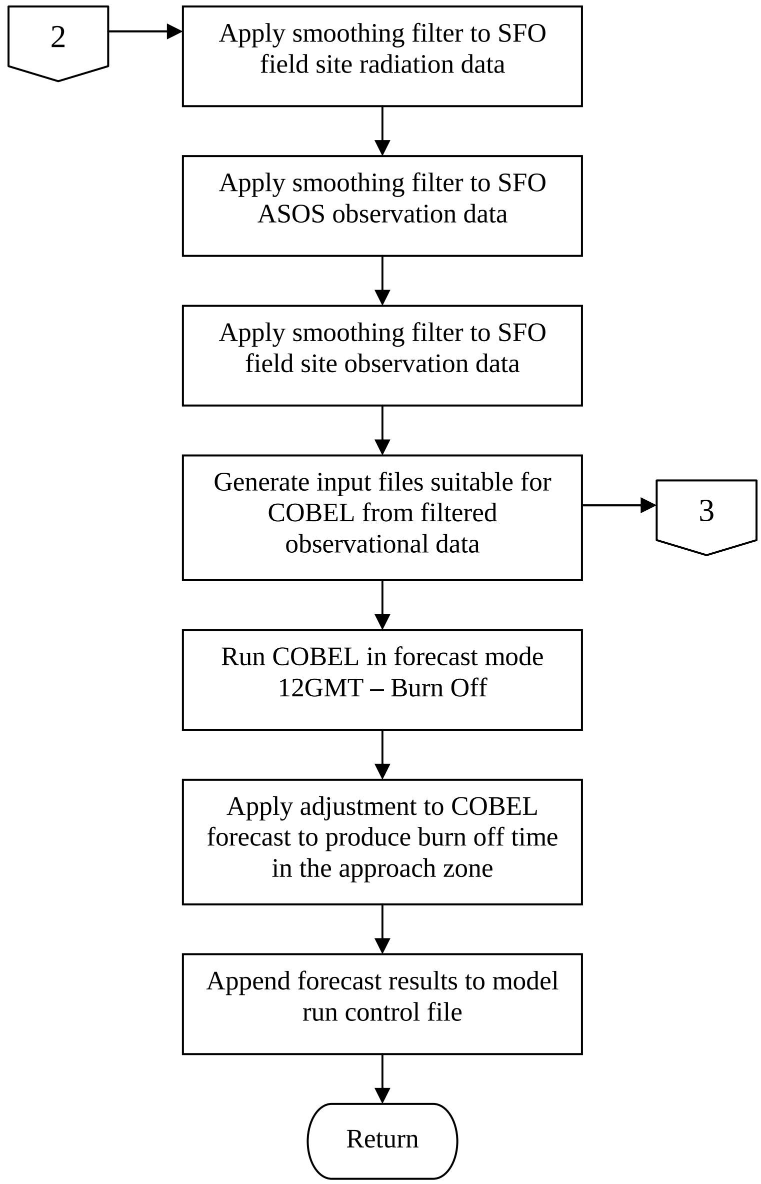


Figure 11. Functional flow diagram of the COBEL model for the 09 and 11GMT forecast runs.

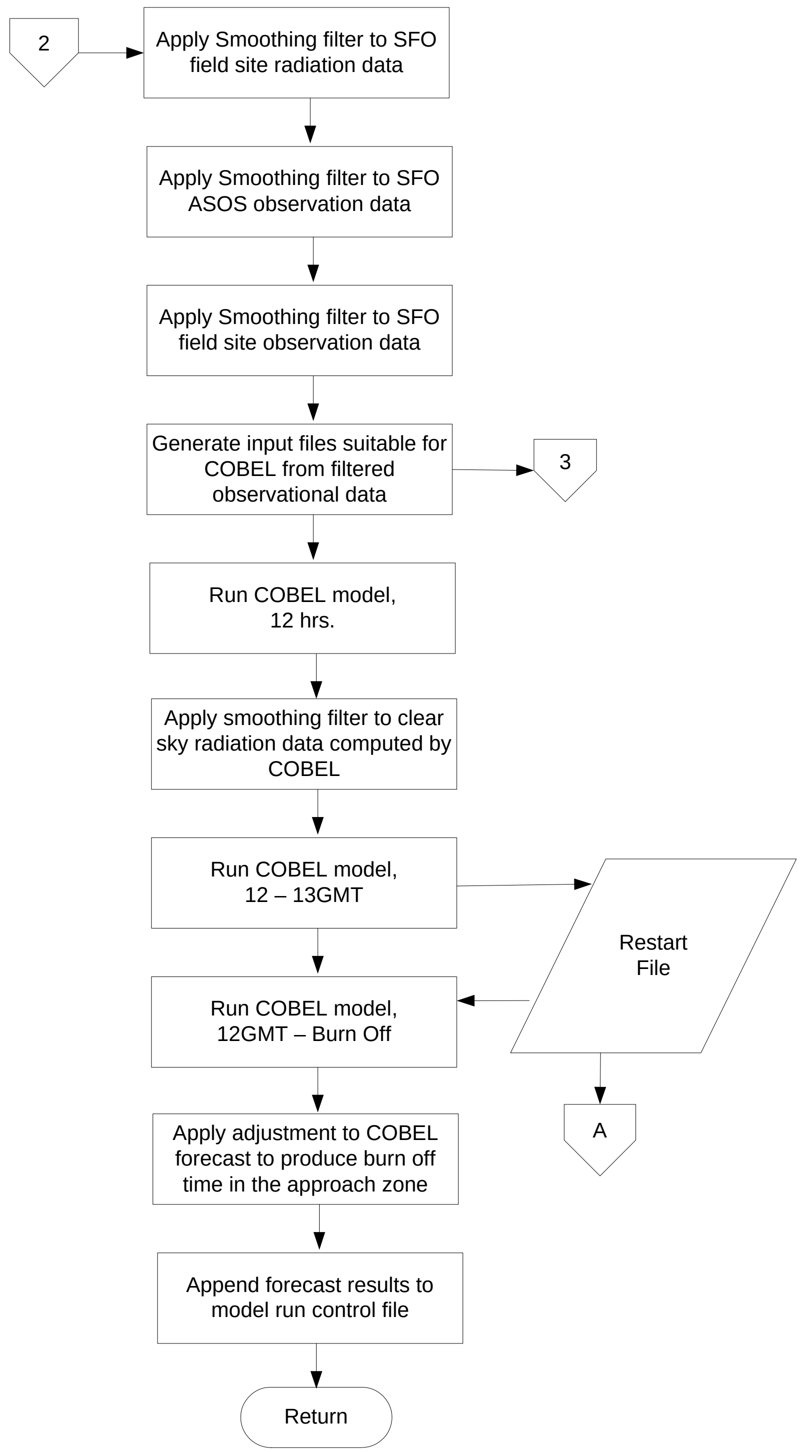


Figure 12. Functional flow diagram of the COBEL model for the 12GMT run.

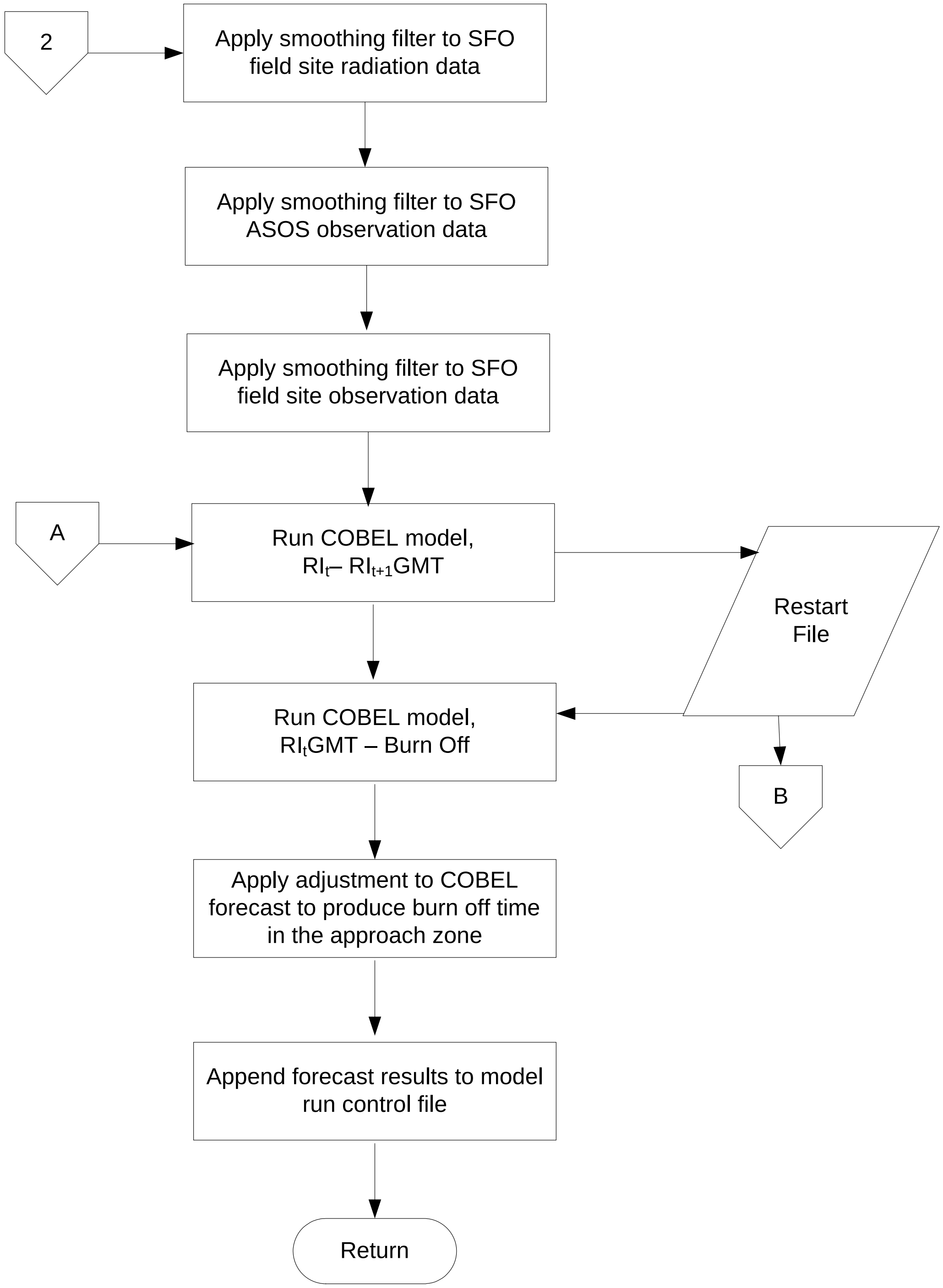


Figure 13. Functional flow diagram of the COBEL model for the 13 – 14GMT forecast runs.

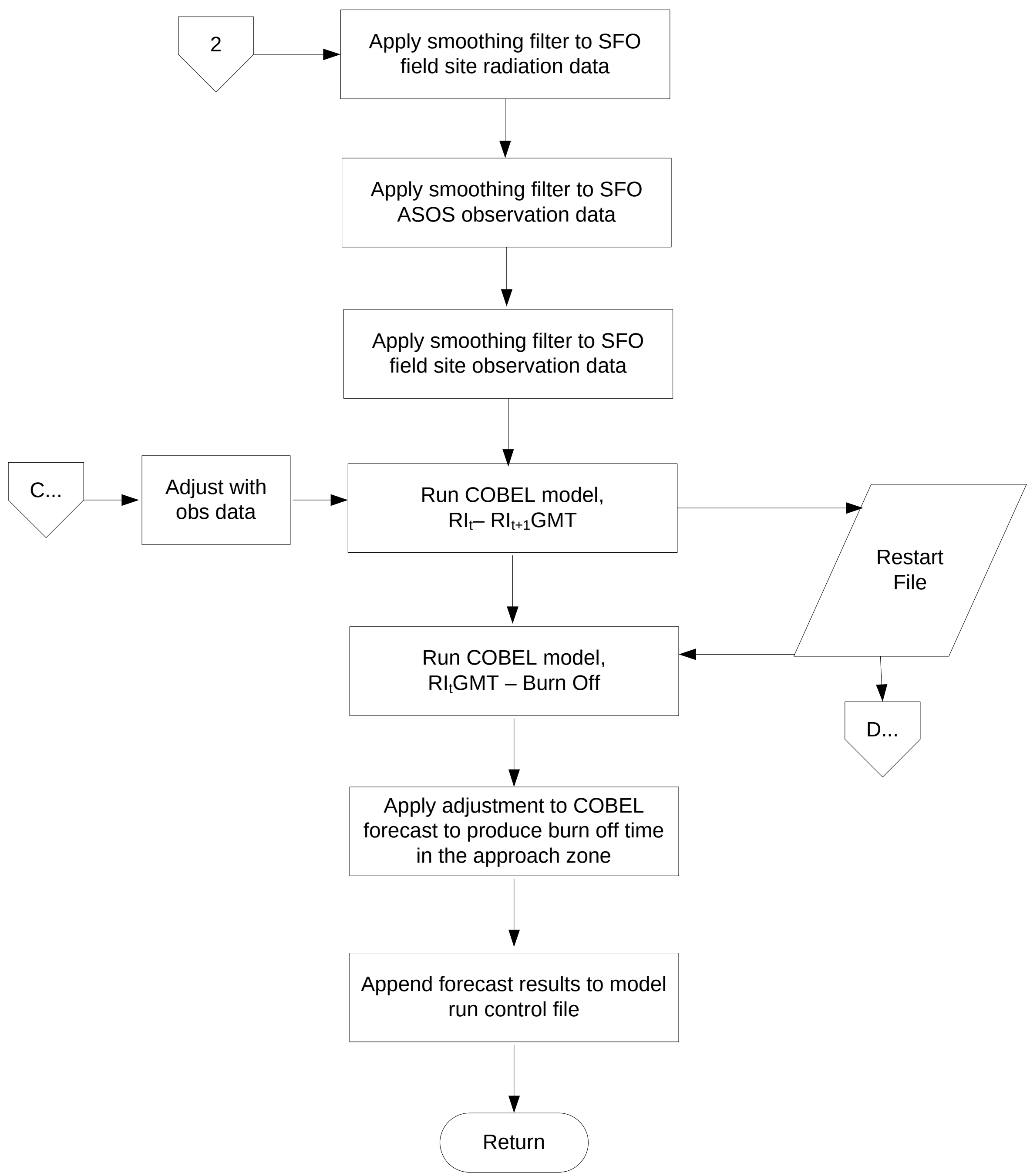


Figure 14. Functional flow diagram of the COBEL Model for the 15 – 18GMT forecast runs.

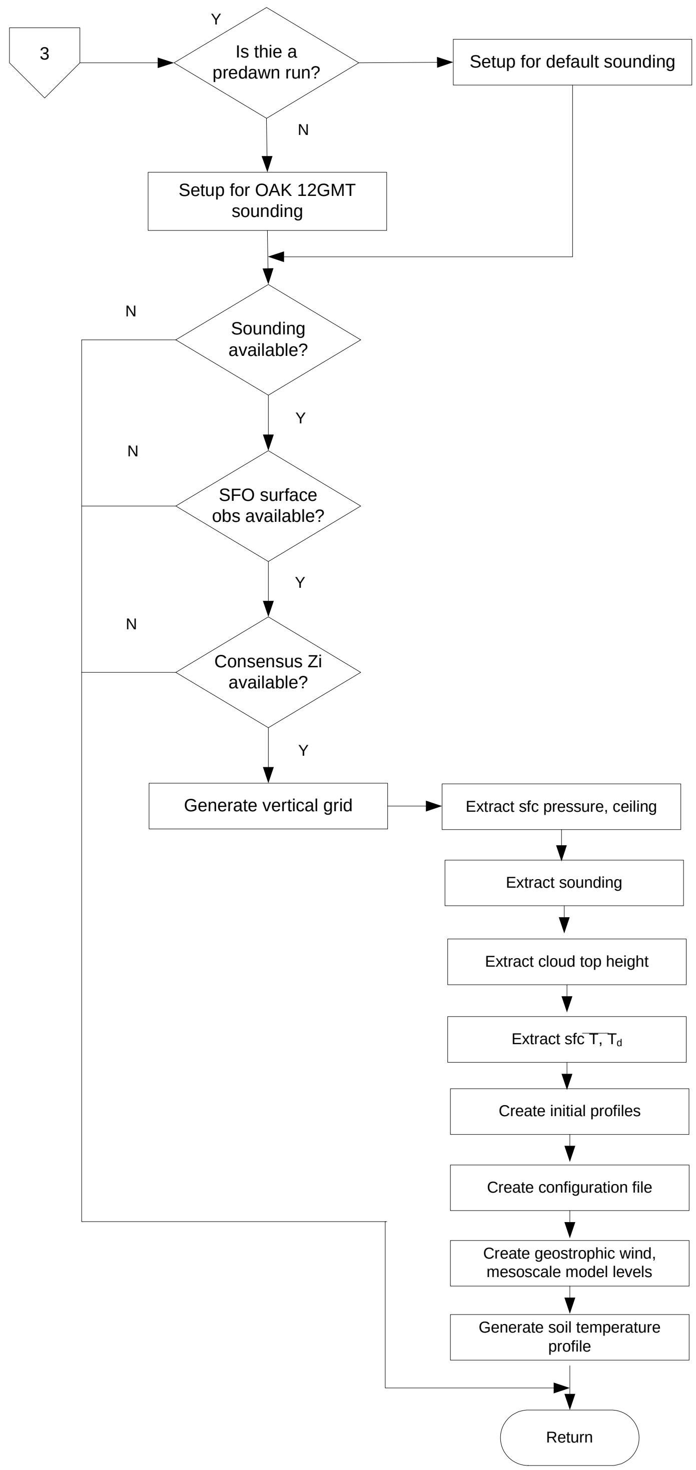


Figure 15. Functional flow diagram of the input preparation step for predawn and 12GMT COBEL runs.

6. CONSENSUS FORECAST ALGORITHM

6.1 INTRODUCTION

The consensus forecast algorithm is designed to combine the four component model forecasts to provide a single forecast of approach zone clearing. This is done by weighting each of the component forecasts based on relative historical performance of each model at each run hour. Accompanying the consensus forecast is an indicator categorizing the forecast confidence as either "Good" or "LOW." Additionally, the consensus forecast is used to provide an empirically derived probabilistic estimate of the likelihood of clearing prior to key operational target times, namely the top of each hour from 17 GMT through 20 GMT.

6.2 TECHNICAL DESCRIPTION

6.2.1 Overview

Four largely independent models produce forecasts of stratus clearing in the approach zone to San Francisco International Airport. (These models are each described in separate documents.) One of these is a dynamic primitive equation model (COBEL), with the other three being statistical models based on varying predictor sets (the Local, Regional, and Satellite Statistical Forecast Models). Because of this independence as to method and data sources, the models provide differing estimates of stratus burn off time. Certain classes of stratus cases will be forecast more accurately by some of the models, and less accurately by others. The Consensus Forecast Algorithm (CFA) seeks to blend the four component forecasts into a single one in such a way as to capitalize on their individual strengths yet mitigate their individual weaknesses, thus providing the single most reliable forecast.

6.2.2 Model Weighting

The CFA blends the component forecasts by a weighted arithmetic sum. The challenge lies in finding the optimal set of weights for any given case. The first step towards determining these weights is to estimate the expected performance of each model at each run time. This is accomplished by quantifying the error of each model/run-time based on an historical archive of cases. This archive consists of each "stratus day" (i.e., a day with morning stratus in the approach zone) for the months of May through October for the period 1996 through 2002. The archive contains 400 to 450 cases at each hour, except for 17 GMT (329 cases) and 18 GMT (196 cases), which have less due to stratus having burned off before these hours for a large number of cases. Forecasts from each model were generated for each case and compared to the actual burn-off time. The resulting errors are represented as

$$\sigma = \sqrt{\sum_i (x_i - y_i)^2 \Big/ (N-1)}$$

where x is the forecast, y is the truth (actual burn off time) for each case, and N is the total number of cases for which a forecast and verification exist (or, equivalently, the total number of i's). This error, which can be considered the standard deviation of the individual errors (if they are unbiased), is computed for each model at each run-time hour.

This can be applied in a straightforward way to COBEL, as it is a primitive equation numerical model and is not subjected to day typing. But the three statistical models require a somewhat more sophisticated approach. One reason for this is that, being statistical in nature (through multiple linear regression) the historical archive set used to evaluate their performance is also the data set used to develop the model coefficients. Using this set to measure the performance of the statistical models results in an overestimation the models' quality. These errors will be substantially lower than they would be from a sample of entirely independent cases. To bring the errors more in line with expected performance on future independent cases, the archive of forecasts from the statistical models are filled not by the models that are used in real-time, but with a degraded set. This is accomplished by splitting the available cases randomly in half (designated A and B). The same predictors used in the full model are tuned using only the A cases, and this A model is used to make forecasts for the B cases. Another model with the same predictors is tuned using the B cases, and this B model then makes forecasts for the A cases. Thus a set of forecasts for all cases can be assembled, with each forecast coming from a somewhat independent model. This technique was found to degrade the forecasts to levels approaching what was observed on entirely independent cases.

The use of day types also sets the statistical models apart from COBEL. For each of the statistical models, one of a number of equations may be used on a given day depending on a set of meteorological conditions unique to that model. The cases in the historical archive are classified as to each model's day types. The expected performance for the statistical models is made more precise by computing and using the errors that are based on these day-type subsets.

Any forecast from any model now has a measure of expected quality associated with it. These are used to construct the weights with which the models are combined to yield the consensus forecast. The errors are inversely proportional to model quality, so they themselves cannot be used as weights. Furthermore, on a theoretical basis it is found that the relative errors are best handled in terms of σ^2 rather than σ. These two considerations lead to the following equation for the weighs:

$$W_k = \frac{1}{\sigma_k^2} \Bigg/ \sum_k \frac{1}{\sigma_k^2}$$

where W is the model's weight and k refers to each individual model. This relation hold if less than all four of the component models are present, and reduces to 1 if only one component model is present. The consensus forecast is then computed as:

$$F_{consensus} = \sum_k W_k F_k$$

where F is a forecast and k again refers to the individual models.

6.2.3 Consensus Forecast Quality Indicator

In addition to providing a consensus forecast time, it was also deemed desirable to provide an estimate of the expected reliability of the forecast. In the development of the project several methods have been used to compute and express this. One recent way was to compute a consensus error (or σ) by using the same weights and method employed to compute the consensus forecast. Assuming normally distributed errors, this consensus error could then be expressed in any number of ways, such as an expected median absolute error, or as a probability that the error would be less than a certain value. Experience showed these estimates to be unreliable due to a violation of the assumption of normally distributed errors. A more basic method of estimating consensus forecast quality was settled upon, though this prior method and its results still exist (but is ignored) in the code and logs of the algorithm.

The quality of the consensus forecast is now expressed in binary fashion (“Good” vs “LOW”). The technique to determine this is similar to that used to determine the day types of the statistical forecast models. By examining the historical archive of consensus forecasts, the meteorological indicators (with their respective thresholds) were found that were best associated with markedly poor model performance. The indicators and thresholds for each forecast hour are shown in Table 19. If any of the three criteria are met, then the forecast is designated as “low quality.” (If any of the indicators are missing, such as visible satellite for the pre-dawn model times, that indicator is considered indeterminate and does not affect the overall quality decision.) In addition, if for some reason (e.g., missing predictor data) only one of the component models is able to make a forecast, then the consensus forecast is also labeled as “low quality,” regardless of the status of the indicators in Table 19. Forecasts from any single model tend to be less reliable than those from the consensus algorithm.

TABLE 19

Indicators and Thresholds Used To Determine the “Good/Low” Quality Flag in the Consensus Forecast Algorithm

Indicator	Applicable Hours (GMT)	Location(s)	Low Quality if...
Sodar Inversion Quality	9,11,13,15,16,17,18	SFO	= 0
Cloud Ceiling Height	9,11,13,15,16,17,18	SFO,SQL,SMB,OAK,HWD	Maximum > 750 m
Visible Satellite Standard Deviation	15,16,17,18	All completely cloud covered satellite sectors	> 0.096

6.2.4 Probability of Clearing Forecasts

As the forecast system development evolved, it became apparent that it would be operationally useful to have the stratus dissipation forecast expressed as a probability, in addition to a discrete forecast time. In particular, air traffic decisions hinge largely on traffic demand and available capacity; key decision points for SFO hinge on the heavy traffic volume scheduled to arrive between 17 GMT and 20 GMT. As such, it was decided to express the probability that dual approaches would be available during this period. Air traffic managers have indicated that there would be an opportunity for more proactive delay program cancellation for instances when there was a very high likelihood (i.e., at least 90%) that dual approaches would be available prior to the large wave of arrival traffic. Toward this end, an approach was developed to convert the discrete forecast time generated by the Consensus Forecast Algorithm into a probabilistic representation. Specifically, the probabilistic representation indicates the likelihood that dual approaches will be available prior to 17, 18, 19, and 20 GMT.

The conversion from the consensus forecast time to a probabilistic expression was developed empirically. Following forecast model development, the historical data base consisted of hundreds of consensus forecasts, and hundreds of clearing verification times. The consensus forecast times were group into 15-minute increments, and then compared to the verification times for those forecasts. For example, for all forecasts made at 13Z, we examined all the consensus forecasts that predicted a time between 16:00Z and 16:15Z. For that particular subset of forecasts, all of the corresponding verification times were examined to determine what fraction of time the actual clearing occurred prior to the target times of 17Z, 18Z, 19Z, and 20Z. This fraction yielded the corresponding probabilities for those target times. In this example, it was found that for all 13Z consensus forecasts for times ranging from 16:00Z to 16:15Z, the actually clearing occurred prior to 17Z on 85% of those cases; thus, for the 17Z “target” time, this is the probability assigned to all 13Z forecasts of clearing between 16:00Z and 16:15Z.

This procedure was repeated using consensus forecasts made at all model run hours. The result is an empirically derived set of look-up tables that resides within the real-time system that assign a probability of clearing based on the time that the actual consensus forecast was made, and the discrete clearing time that was forecast. As an example, the look-up table for consensus forecasts made at 15Z is shown in Table 20. For a consensus forecast made at 15Z for the time shown in the leftmost column (represented in fraction hours), the probability of clearing *any time prior to* the four target times (17Z through 20Z) is shown in the corresponding four column labeled 17-20.

TABLE 20
Sample Look-up Table Indicating Probability of Clearing By Each Target Hour of 17Z through 20Z, Based on 15Z Consensus Forecast Time. Leftmost Column Indicates Forecast Time From 15Z Forecast.

15z	17	18	19	20
16.00	0.90	0.95	0.95	0.95
16.25	0.85	0.95	0.95	0.95
16.50	0.75	0.95	0.95	0.95
16.75	0.70	0.95	0.95	0.95
17.00	0.60	0.95	0.95	0.95
17.25	0.50	0.90	0.95	0.95
17.50	0.40	0.90	0.95	0.95
17.75	0.25	0.75	0.90	0.95
18.00	0.20	0.60	0.90	0.95
18.25	0.10	0.45	0.80	0.95
18.50	0.05	0.35	0.75	0.95
18.75	0.05	0.30	0.65	0.90
19.00	0.05	0.20	0.65	0.85
19.25	0.05	0.15	0.55	0.85
19.50	0.05	0.10	0.50	0.85
19.75	0.05	0.10	0.50	0.80
20.00	0.05	0.05	0.35	0.60
20.25	0.05	0.05	0.20	0.40
20.50	0.05	0.05	0.10	0.30
20.75	0.05	0.05	0.10	0.25
21.00	0.05	0.05	0.10	0.20
24.00	0.05	0.05	0.10	0.10

6.3 REAL-TIME PROCESSING

6.3.1 Data Inputs

The four component forecasts are found in four separate files. The files, for a given day *YYMMDD*, are found in */ll/cv/data/daysumYYYY/fcsts/YYMMDD* and are labeled *fcsts_TYPE.YYMMDD.sum* where *TYPE* can be *Cobel*, *Local*, *Regional*, or *Satellite*. Each of these files contains all the forecasts for the day thus far, in chronological order. Logic within the code determines the latest forecast and decides whether it is timely enough to be included in the current consensus calculations. These files also include the day types of the models responsible for each forecast.

The errors (σ) for each model and day type, which are used to compute the weights given to the component forecasts, are read from a static table. It is found in */ll/cv/consensus/ascii/perf_scores.txt* and is shown in Figure 16. The number of cases used to compute each σ is included in this table for reference but is not used in the algorithm.

The algorithm also reads the current observations needed to determine the overall consensus quality flag. These files are found in the daysum directory structure. The names and locations of these files are given in Table 21.

TABLE 21
Input Files Used to Determine "Good/Low" Quality Flag in the Consensus Forecast Algorithm

Indicator	Input File (/ll/cv/daysumYYYY/)
SFO Sodar Inversion Quality	inv/YYMMDD/inv_hyb_sfo.YYMMDD.sum
SFO Ceiling	sfc/YYMMDD/sfc_afos_sfo.YYMMDD.sum
SQL Ceiling	sfc/YYMMDD/sfc_afos_sql.YYMMDD.sum
SMB Ceiling	sfc/YYMMDD/sfc_awos_smb.YYMMDD.sum
OAK Ceiling	sfc/YYMMDD/sfc_afos_oak.YYMMDD.sum
HWD Ceiling	sfc/YYMMDD/sfc_afos_hwd.YYMMDD.sum
Visible satellite coverage and standard deviation values	sat/YYMMDD/sat_vissec_goes10.YYMMDD.sum

6.3.2 Intermediate Input/Outputs

No intermediate input is required or intermediate output generated by this algorithm.

MODEL	NCASES	SIGMA
Cobel09	179	1.764
Cobel11	269	1.664
Cobel13	381	1.597
Cobel14	372	1.470
Cobel15	347	1.268
Cobel16	288	1.157
Cobel17	182	1.248
Cobel18	93	1.360
Local09Default	161	1.491
Local11Default	237	1.430
Local13ZiQLT55	44	2.195
Local13ZiQGE55	303	1.208
Local14ZiQLT75	52	1.410
Local14ZiQGE75	296	1.193
Local15ZiQLT90	66	1.423
Local15ZiQGE90	239	1.056
Local16ZiQGE90	231	0.884
Local16ZiQLT90	68	1.070
Local17Default	226	1.005
Local18Default	122	1.201
Regional09Default	928	1.699
Regional09Neutral	588	1.544
Regional09Onshore	58	1.458
Regional09Offshore	203	1.541
Regional11Default	952	1.687
Regional11Neutral	583	1.622
Regional11Onshore	57	1.882
Regional11Offshore	194	1.418
Regional13Default	951	1.650
Regional13Neutral	659	1.590
Regional13Onshore	83	1.613
Regional13Offshore	165	1.494
Regional14Default	944	1.515
Regional14Neutral	642	1.449
Regional14Onshore	90	1.555
Regional14Offshore	173	1.412
Regional15Default	918	1.486
Regional15Neutral	664	1.497
Regional15Onshore	99	1.461
Regional15Offshore	131	1.193
Regional16Default	883	1.452
Regional16Neutral	640	1.487
Regional16Onshore	79	1.817
Regional16Offshore	105	0.809
Regional17Default	697	1.343
Regional17Neutral	531	1.384
Regional17Onshore	73	1.292
Regional17Offshore	70	1.107
Regional18Default	445	1.299
Regional18Neutral	334	1.402
Regional18Onshore	54	1.615
Regional18Offshore	24	0.494
Satellite15Default	310	1.265
Satellite15LoCigSmthCld	216	1.073
Satellite15LoCigRufCld	51	1.663
Satellite15HiCig	33	2.021
Satellite16Default	280	1.079
Satellite16LoCigSmthCld	206	0.958
Satellite16LoCigRufCld	34	1.904
Satellite16HiCig	33	1.160
Satellite17Default	220	1.032
Satellite17LoCigSmthCld	136	1.100
Satellite17LoCigRufCld	28	1.405
Satellite17HiCig	45	0.835
Satellite18Default	129	1.293

Figure 16. Static file perf_scores.txt, which gives error estimates for each model/day type.

6.3.3 Data Outputs

Two files are produced or augmented by the consensus forecast algorithm. The first is in the same daysum forecasts directory as the component model forecasts. This file is located in */ll/cv/data/daysumYYYY/fcsts/YYMMDD*, and is named *fcsts_Consensus.YYMMDD.sum*. An example of such a file is shown in Figure 17. There is only one of these files per day, so upon successful computation of the current consensus forecast this information is appended to the end of the day's file. Each line of the file contains the forecasts for each component model as well as for consensus. Associated with each forecast is an estimate of confidence. As described above, these are not currently displayed by the system.

The other output file is placed in */ll/cv/data/ProductsArchive/YYYYMMDD/ConsensusForecast* and is named *YYYYMMDD_HHMMSS__latest_consensus.dat* where *HHMMSS* is the clock time at which the consensus forecast was made. Unlike the former output, a separate file is produced for each consensus forecast. An example of this file is shown in Figure 18. Here there is one line for each of the component forecasts and consensus, containing the forecast, confidence estimate, and weight. In addition, the consensus line of this file contains "Y/N" indications of the state of the three consensus quality indicators, with the last column being the boolean "and" of these and indicating the overall quality flag.

6.3.4 Structural Flow

The consensus forecast algorithm as described in this document is called by the master post-processing script, which is described in the Real-time System Document. The details regarding the timing of the consensus executions can be found there as well. Figure 19 shows the consensus model functional flow diagram from that document. The portion of the functional flow corresponding to the content of this current document is highlighted by the red box and expanded in detail as Figure 20.

There are three basic parts of the process. First the component models and their corresponding error estimates are acquired. Then the model weights are computed and used to blend the forecasts into a single estimate of stratus burn off. Finally, select observations are acquired and an evaluation of the overall quality ("good" or "low") of the consensus forecast is made.

6.3.5 Nominal and Degraded Performance

Nominal performance of the consensus algorithm occurs when timely forecasts from all four component models are available. Such forecasts will also be the most reliable. However, if only two or three component forecasts are available, the weights are adjusted accordingly and a forecast is still produced. If only one component forecast is available this is in effect simply passed along as the consensus forecast. However, the "low" quality flag will be set, regardless of the status of the meteorological indicators. If no component forecasts are available, a consensus forecast will not be produced.

YY/MM/DD	HH:MM	HH.MM	RunId	ConsFcst	ConsConf	LFcst	LConf	RFcst	RConf	SFcst	SConf	CFcst	CConf
04/07/05	11:36	11.6	1100	17.29	1.09	N/A	N/A	17.29	1.09	N/A	N/A	N/A	N/A
04/07/05	13:24	13.4	1300	17.46	0.95	18.04	0.81	16.82	1.07	N/A	N/A	17.09	1.08
04/07/05	15:23	15.38	1500	17.00	0.79	17.03	0.71	16.49	1.01	17.21	0.72	17.04	0.86
04/07/05	16:19	16.31	1600	17.65	0.70	17.73	0.60	17.13	1.00	17.62	0.65	17.89	0.78
04/07/05	17:20	17.33	1700	18.11	0.78	18.44	0.68	17.43	0.93	18.01	0.74	18.27	0.84

Figure 17. Sample fcsts_Consensus.YYMMDD.sum Output File.

YY/MM/DD	HH:MM	Model	RunId	Fcst	Conf	Weight	LowZiQ	HiCig	RufCld	LowConf
04/07/05	17:20	CONS	1700	18.11	0.78	1.00	N	N	N	N
04/07/05	17:20	LSFM	1700	18.44	0.68	0.33	-	-	-	-
04/07/05	17:20	RSFM	1700	17.43	0.93	0.18	-	-	-	-
04/07/05	17:20	SSFM	1700	18.01	0.74	0.28	-	-	-	-
04/07/05	17:20	COBEL	1700	18.27	0.84	0.22	-	-	-	-

Figure 18. Sample YYYYMMDD_HHMMSS_latest_consensus.dat output file.

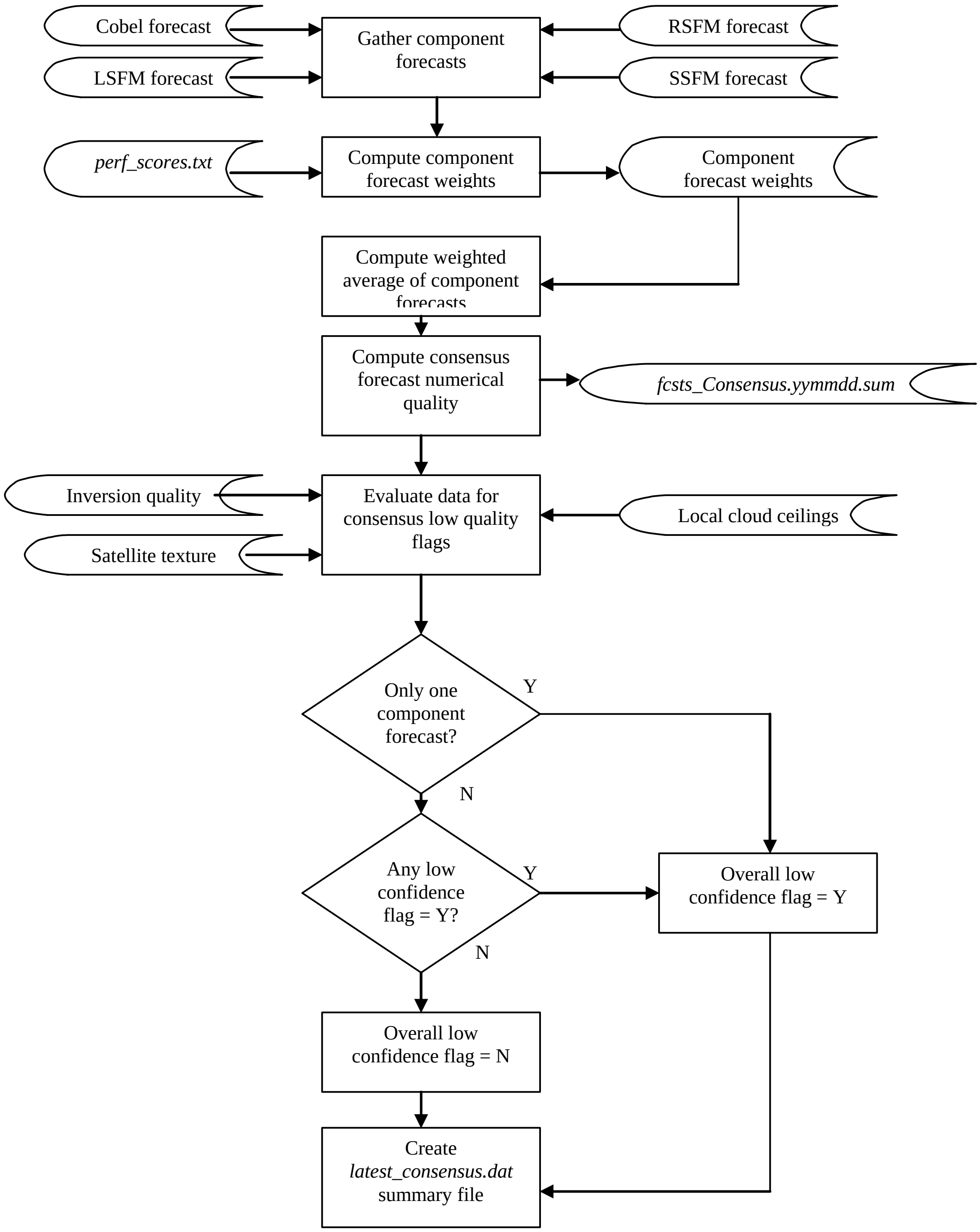


Figure 19. Structural flow diagram for the consensus forecast algorithm, corresponding to the highlighted area of Figure 20.

Figure 20. Functional flow diagram for the consensus model, from the Real-Time System and Display Document.

6.4 PRACTICAL CONSIDERATIONS

6.4.1 Programming and Resources

There are no issues related to programming and resources.

6.4.2 Exception Handling

The inability of the consensus algorithm to run because of the lack of input data does not trigger any exception processing. The consensus algorithm simply does not generate a forecast.

6.4.3 Quality Assurance and Diagnostics

Real-time execution of the consensus algorithm can be tracked in the log file. These can be found in */ll/cv/logs/YYYYMMDD/ConsensusModel* and are named *UpdateConsesusForecast.py.log.YYYYMMDD_ HHMMSS*. One of these is generated each time the consensus algorithm is called (currently every minute). Most of these log files indicate either that the consensus algorithm could not yet be run (due to missing forecasts), or that the consensus forecast has already been computed for the present cycle. The log files of relatively large size correspond to times when the consensus algorithm computed a forecast. It is these files that can be used as a diagnostic tool to investigate the operation of the consensus algorithm.

6.4.4 Operator Interaction and Default Settings

The Consensus Forecast Algorithm is activated by a model launching script, as described in the Real-time System and Display Document. No operator interaction is required.

6.4.5 Periodic/Potential Maintenance Tasks

No periodic maintenance is required for the Consensus Forecast Algorithm. If any of the component models or their day types changes, however, the error estimates for these models will need to be computed and placed in the *perf_scores.txt* table (see Table 1).

APPENDIX A
DESIGN AND AUTOMATION OF STATISTICAL FORECAST MODEL DEVELOPMENT PROCEDURES

1.0 INTRODUCTION

A methodology for generating statistical forecast equations for cloud ceiling & visibility prediction has been developed in support of the FAA's Aviation Weather Research Program. The initial forecast application was the prediction of dissipation of stratus cloud which restricts the flow of air traffic into San Francisco International Airport (SFO). Since ceiling and visibility impacts aviation operations at a number of major airport terminals throughout the National Airspace System, the statistical forecast development method designed for San Francisco has been generalized and automated to expedite similar application at other airport terminals. This report describes the design and automation of these model development procedures.

Statistical models are often a desirable way of generating meteorological forecasts, from short-term in time to climatic. They provide a simpler and less costly alternative to dynamic primitive-equation models, and can parameterize aspects of the phenomena that may be below the resolution of numerical models. The statistical approach also has particular appeal for supporting development for a large number of airport terminals, for which a substantial historical archive of routine meteorological observations is available. However, a statistical forecast approach presents a myriad of possible techniques and predictors in arrival at an optimal set of forecast solutions. For example, the efforts in SFO involved several years of experience in developing and applying statistical techniques, with early methodologies requiring a substantial amount of manual analysis for proper selection of candidate predictors and forecast equation selection. Through this experience, a standardized process was established, including the development of mechanisms for automation of steps to expedite much of the analysis and processing.

Since forecast model development for the SFO Marine Stratus Initiative (hereafter referred to as MSI) is cited often within this document to exemplify the considerations for techniques and automation, a brief overview of the MSI forecast mission is warranted here. The primary objective of the MSI was to predict the time that stratus would be sufficiently clear of the approach zone to allow dual-parallel (side-by) approaches into the airport. Four separate forecast guidance models were developed for this purpose, with each model attempting to predict the time of side-by operations (i.e., the predictand) that correspond with stratus burn-off. One is a physical model, while the other three are statistical models. Development of the three statistical models serves as the prototype for the design and automation of the methodology presented in this report. Each of the three statistical models developed for SFO require different types of input observations: 1) visible satellite imagery, 2) regional observations from standard surface and upper air sensors operated by the National Weather Service, and 3) local observations from sensors deployed specifically for the SFO project. These statistical models are referred to as the Satellite Statistical Forecast Model (SSFM), the Regional Statistical Forecast Model (RSFM), and the Local Statistical Forecast

Model (LSFM). For real-time operations, each model is run hourly throughout the morning to provide regular updates to the estimated time of stratus burnoff.

The remainder of this document is presented in two main sections. The first describes the model development approach, citing examples from the SFO experience, and extending the important considerations for standardizing the approach for application at other airports. The second describes the automation procedures that were developed to expedite the development process. This is particularly important for extending these techniques at other airports, each of which are likely to have a large number of potential predictor parameters to consider for statistical forecast equation development.

2.0 MODEL DEVELOPMENT CONSIDERATIONS

2.1 Overview

Model development in the context of this report refers to selection of the appropriate set of predictors and coefficients to arrive at an effective forecast model equation for estimating a predictand. For the applications addressed herein, the forecast development process assumes a large number of meteorological observations as potential predictors for the occurrence of some meteorologically significant event (predictand), such as the crossing of some parametric threshold (ceiling height, horizontal visibility, etc) that impacts aviation operations. The experience for the SFO MSI led to three important considerations for building statistical models:

- To better represent the required information, observations may need to be transformed into other “derived” observations or rescaled.
- In general, the full set of possible predictors leads to far more permutations of potential models than can be thoroughly examined, so the full list of predictors must be pared down into a more practically manageable list of “most important” predictors.
- Once the full list of possible predictors is limited to a set of important predictors, a “best” model must be selected from all the models that can be built from the limited set of predictors.

The flow of the model-building processes developed to accommodate these considerations is shown schematically in Figure 1. The processing stages are shown on the left, with the results of each stage represented on the right. The process begins with any raw meteorological observations that the analyst deems relevant for predicting the phenomena. These observations are subjected to quality control, possible transformations into new predictors, and a non-linear scaling that maximizes their correlation to the variable being predicted. The heart of the process then accomplishes a series of reductions. First, the number of original variables is reduced to a set of most important predictors small enough to allow testing of all of their combinations. These many combinations are tested via linear regression to find a moderately large set with the best error fits to the observations. This set in turn is subjected to a cross-validation procedure that determines the small set of models likely to perform the best on independent data. The analyst reviews this set to determine the single model to be implemented.

The following sections provide more details regarding the practical implementation of each of these steps, including specific examples from the SFO MSI, as well as considerations for adapting these steps for forecasting applications at other airports.

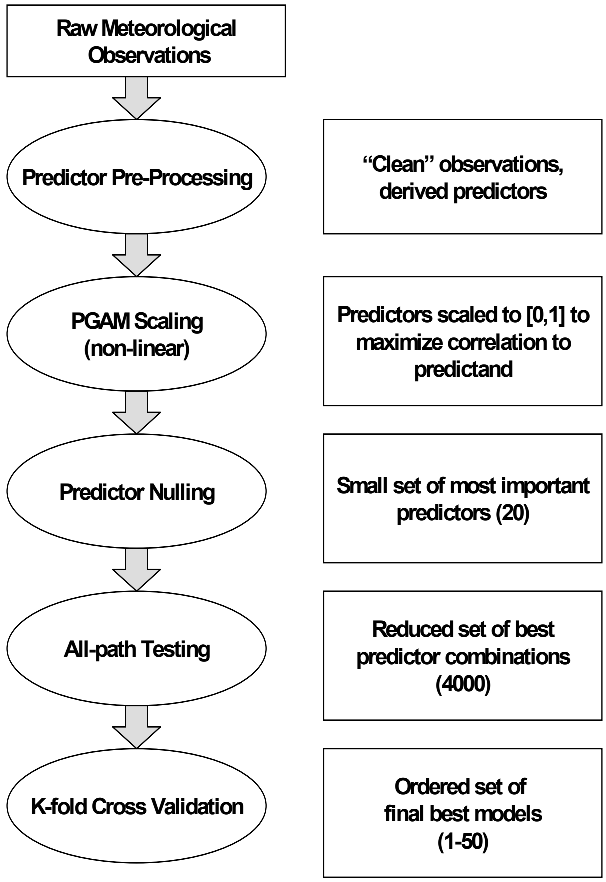


Figure A-1. Schematic overview of SFM-building process.

2.2 Data Preprocessing

The process starts with raw data, collected in the field, and subjected to some form of rudimentary quality control. Beyond this, the data can be put into a form that makes them amenable for statistical model building. For example, visible satellite imagery in the MSI was processed in such as way as to provide basic summary statistics (mean, standard deviation, and fractional cloud coverage of pixel brightnesses) for roughly 50 geographical sectors which covered the local spatial domain. This was done after standardization efforts strove to insure pixel brightness variations were due only to changing cloud characteristics, and not to causes such as solar angle variations of the illumination, or to sensor decay.

Similarly, height of the crucial base of the thermal inversion can be seen in sodar displays. Software was written to objectively determine the height of this base from such imagery. Details of these types of data processing efforts of the MSI are documented in Wilson and Clark (1999).

Whatever the source of the data, any model-building effort must deal with the challenges of routine data drop-outs. Missing observations, especially for projects with a relatively small amount of development data, quickly curtail the number of usable cases for which all predictors are present. A predictor that is seen to be important, but that is often missing, presents the developer with the uncomfortable choice between a good model which can not be used often versus a more dependable but degraded model. At the most basic level, missing data can be substituted with reasonable neighboring observations.

2.3 Predictor Rescaling: PGAM

In this step the data are rescaled to better represent what is important about each observation. An example (from the MSI) is the u wind component (east/west wind component) in the vicinity of the SFO Approach Zone. Any wind from the east indicates that there is no onshore wind component, and thus the wind is not advecting cold air into the bay. The magnitude of the wind is not important. That is, as a wind from the east increases in strength the forecast should not change; all that matters is that there is a wind from the east. In contrast, a wind from the west corresponds to cold air advection from the Pacific Ocean, which will tend to retard the stratus burn-off. Furthermore, when the wind is from the west, the stronger the wind the greater the advective effect. Unlike a wind from the east, a forecast model should respond to an increasing wind from the west. In order to achieve the proper response in a linear model, the data should be rescaled so that any u value that is negative (wind from the east) is given a constant value, while the u values that are positive should be rescaled so that increasing u values result in increasing rescaled values.

While in some cases, such as the u wind component, it is easy to guess a reasonable rescaling, in most cases it is less clear how to achieve a reasonable rescaling. In all cases it is not possible to simply know a priori what is the best rescaling. To automate this step we choose the monotonic piecewise linear rescaling that maximizes the correlation between the input observation and the predictand, and that rescales the data into the range of [0,1].

Rescaling into [0,1] aids human understanding of the resulting linear models. In the MSI example, in a linear model some predictor has a weight of *W*, then that predictor has at most an effect of *W* hours on the predicted burn-off. Thus by simply looking at the weight given any predictor, an analyst sees the maximum effect of a predictor. A small weight indicates an unimportant predictor. Rescaling into [0,1] also aids the numerical stability of the regression process.

With a rescaling into [0,1] a large predictor weight does not automatically correlate with a predictor's importance, since a predictor may be rescaled so that it is almost always very small, but only occasionally equal to one. In this case the predictor on average has a very small effect on the forecasts, even though it gets a large weight. Some predictors rescale so that they are almost always one, in which case they tend to just trade off against the constant term in the forecast equation. (Such predictors would likely not be among the set of most important predictors. See the discussion of predictor nulling below.) A third consideration is that some predictors rescale so they nearly always take on only the two values zero and one, in which case they tend to perform the job of a switch between two models. When this occurs it may be better to develop separate models for each of these two classes, starting the process over. These behaviors are not necessarily bad, but the analyst must keep them in mind.

If an analyst wished to know the average effect of a predictor, in a Root Mean Square (RMS) sense, the rescaling could be done so that the RMS value of the rescaled variable was one. In this case the weight given a predictor is just the RMS effect of that predictor on the forecast. While different choices for the range for the rescaling can be chosen, the choice does not ultimately affect the choice of model or its performance.

This rescaling is similar in some respects to what is done in Generalized Additive Models (GAM). In GAM, the predictors that are being used have already been selected, and all predictors are rescaled together in one step. If the predictor list is changed, the entire rescaling process must be repeated for each predictor. For our application, on the other hand, it is desirable to rescale once so that the rescaled predictors can be used in the predictor selection process. For an overview of GAM with a meteorological application see Vislocky and Fritsch (1995). There are some other technical differences between the rescaling presented here and GAM. Therefore we refer to this rescaling as Pre-scaled (or Pseudo-) GAM, or PGAM.

2.4 Preliminary Predictor Selection: Predictor Nulling

Given fast enough computers one would simply take the set of potential predictors and build and evaluate every possible model. However, it is typical to have far too many predictors to build all possible models. For example, each satellite image for the MSI project is turned into a three observations for each of 50 sectors, for a total of 150 potential predictors, which corresponds to 1.4×10^{45} possible forecast model equations! Often current observations and observations from the previous hour are used together, giving rise to some 300 predictors, or 2.0×10^{90} possible models. Since we cannot build and test all possible models we wish to find a subset of the predictors that is small enough to test all possible models and that would likely include many nearly optimal models. We call this subset the set of important predictors. A set with, for example, 20 predictors is small enough to allow for all model testing, and in practice seems large enough to contain many nearly optimal models.

In the early stages of the model development effort for the SFO MSI, selecting a small set of important predictors from the large set of potential predictors was accomplished by selecting only those predictors with a high correlation to the time of burn-off. Because the process of model evaluation was not fully automated the number of predictors was limited to approximately 12. Once the process was started and certain predictors were eliminated, the analyst could try inserting other predictors, but the number of predictors that could be examined was very limited. Even with 12 predictors, the process had to be ad-hoc as only a relatively small number of models could be tested manually.

The downside of limiting the predictor set by only using the direct correlation of a predictor to the predictand is best illustrated by looking at a simple example from the MSI. Suppose cloud thickness is an important predictor, and we only have observations of the height of the cloud base and the height of cloud top. Since cloud thickness is just the difference in cloud top and cloud base, it may be desirable to keep both observations in the set of important predictors. However, if one of these values is not highly correlated to the predictand (time of burn-off) it is eliminated before there is any chance to evaluate its importance in conjunction with the other predictors. In this example, an analyst may well have enough knowledge to keep both cloud base and cloud top in play, but in general these interrelationships are not as straightforward and may involve several predictors. A model building process should be able to find such important groups of predictors.

Another consideration can be illustrated by expanding this example. Suppose in addition to cloud top and cloud base, there is a direct measurement of cloud thickness. Now the process should keep cloud thickness, but not both of cloud top and cloud base, as the three together represent highly correlated information. Building models using highly correlated predictors is not only inefficient; it leads to highly unstable calculations of the predictor weights.

In order to find a small set of important predictors in a way that deals with the considerations illustrated by the above examples, the process must consider all potential predictors and all of their interrelationships. Approaches that simply look at correlations to the predictand, or work by adding or subtracting one predictor at a time will not necessary yield an optimal solution. Since a direct consideration of all possible interrelationships is not feasible, an indirect method is needed. *Predictor nulling* was developed to provide this indirect method of considering all possible interactions of predictors.

Predictor nulling is conceptually simple. The process starts with all potential predictors and computes a solution to the predictor weights that gives a nearly optimal solution (optimal in the usual least squares sense) while at the same time striving to drive small weights to zero. Minimizing a cost function formed from two terms does this. One is the square error fit of the model to the training data, which tends to drive the solution to the least squares solution; the other is a penalty function that tends to drive small weights further towards zero. If a model can be built that fits the data nearly as well as the true least squares solution, and with the weights for some predictors being nearly zero, then the predictors with nearly zero weight do not play an important roll and they can be discarded from further consideration. In practice the process does not find all unimportant predictors in one step. Instead several unimportant predictors are found and deleted from the set of predictors, and the process is repeated. This process is repeated until the desired number of predictors is left in play. Because the predictors we use are highly cross-correlated, and due to the nature of non-linear optimization, the final list of important predictors is not unique. Changing the parameters that control the nulling will result in a different list of important predictors. This appears to have little or no effect on the performance of the final model, since as we discuss later, there are many models with nearly optimal performance.

If we return to the conceptual examples used above we can see on a practical level how nulling works. We again assume that cloud thickness is important to the problem. If the predictor set has cloud base and cloud top, but does not contain cloud thickness, then the least squares solution will make strong use of cloud top and cloud base, giving a positive weight to cloud top and a negative weight to cloud base, thus forming cloud thickness. If the weight given either cloud top or cloud base is pushed towards zero, the fit to the data will suffer significantly. The nulling process requires a solution nearly as good as the least squares solution, so the weights for cloud top and cloud base are not pushed towards zero. Thus both cloud top and cloud base are left with large weights, and so remain in play. On the other hand if the predictor set contains cloud top, cloud base, and cloud thickness, it is possible to form optimal models using all three predictors, cloud top and cloud base only, or cloud thickness only. Driving two weights to zero, supposing cloud top and cloud base are not important in their own right, gives a smaller value to the nulling cost function than driving one weight to zero. Thus the nulling process chooses near zero weights for cloud top and cloud base, leaving only cloud thickness in play.

2.5 Initial Model Selection: All-path Testing

This step starts with the subset of important predictors, as determined by nulling. Generally, a subset of 20 predictors is used. To judge the performance of a model, ideally the available data set would be split into a set for training and a set for evaluation, so that the evaluation would be independent of the training. However, our data sets are too small, with ramifications for model selection that are discussed in the next section.

For this reason we choose to evaluate models based on a 10-fold split of the training data. This process is somewhat time consuming, so we first go through all possible models, training them on all the data, and computing their fit to the training data. The fit to the training data is an optimistic measure of model skill, so models that do not fit the training data are presumed to be poor models. However, since the fit to the training data may not be a good indication of how a model will perform on independent data care must be taken to keep as many models in play as possible at this stage. The fit to the training data also does not provide any useful information on the optimal number of predictors (models with more predictors tend to be preferred), so models of each size should be kept. For each model size, from one predictor to 20 predictors, the 200[1] models with the best fit to the training data are kept in play.

2.6 Quasi-independent Evaluation: K-fold Cross Validation

After the initial model selection process is finished, the number of models left is small enough for a more comprehensive evaluation of each model. If there were enough data cases the data could split into two sets, one for calculating the predictor weights and one for an independent evaluation of the resulting model. Unfortunately the data sets are too small for this to work well. If our data sets are split we are faced with two problems. The number of predictors that can be used without over fitting the training data (causing numerical instability in the calculation of the predictor weights) is a function of the number of training cases. If half the data are used for training, for example, the evaluation will tend to select models with half as many predictors as are justified by the full data set, leading to models with less than optimal performance. The other problem is that there are too few cases for a meaningful evaluation, further contributing to the selection of a less than optimal model.

In order to address these issues a k-fold cross validation is employed. We have chosen to use k=10, although other choices can be made. The available data are split into 10 subsets. One subset is removed, and the model is trained on the remaining 90% of the data. The resulting model is used to forecast on the 10% of the cases that was held out of the training. The forecast errors on those days are then collected. A different 10% is then held out, the model retrained, and another set of forecast errors generated. This process is repeated 10 times until each subset has been withheld from the training. In the end there is an error estimate for each training case. These errors are independent of the model training, and are used to

[1] Note that for some model sizes there are less than 200 possible models. For example there are only 20 models with one predictor, and only one model with 20 predictors.

compute a root mean square error for the model. Because each time the model was only trained on 90% of the data, the error estimates tend to be slightly pessimistic.

Like all cross validation methods, when there are too few data the resulting estimate of the model error may not faithfully represent the true performance. One problem is that the split into 10 subsets may accidentally result in some sets that do not contain a representative distribution of cases. Since the training uses 90% of the data at each step, if the full set has a good distribution of cases it is likely that the training sets will too. But we have found that with less than about 60 cases, this process does not always behave well. The solution is to run with many different k-fold splits. The current implementation constructs 10 random 10-fold splits of the set of training set, which are then used as a basis for 10 separate cross-validation analyses.

2.7 Final Model Selection

The k-fold cross validation does not give the one optimal model that should be fielded. There are many models, sometimes hundreds, with an estimated RMS error essentially equal to the best model. It is extremely unlikely that the optimal model will be the one with the lowest estimated error. This is due to noise in the data, the distribution of cases of various characteristics across the cross-validation splits, and the fact that the models are not trained on 100% of the data. Even if the optimal model (in a least squares error sense) could be found with certainty, there are factors other than RMS error that are important. Once a set of models with essentially the optimal RMS error is assembled, these other factors are used to determine which model to field.

Over fitting the training data is an important consideration. If a model has too many predictors for the amount of training data, the model starts to fit the noise in the data. This makes the calculation of the predictor weights unstable, and leads to poor model performance. One clue that a model has too many predictors is that small changes in the training data give rise to large changes in the predictor weights. For this reason the stability of the weights is computed as part of the cross validation. If the cross validation trains the model 10 times, 10 estimates of each predictor weight are produced. The standard deviation of each weight is calculated from the 10 estimates. The largest standard deviation is the “weight stability” for that model. Models with more stable weights are preferable to models with less stable weights. The issue of over-fitting is also dealt with directly by giving preference to models with fewer predictors.

Another important consideration is the forecast range of a model. It is important to correctly forecast across the largest range possible of predictand values. In the MSI, for example, it is desirable to forecast both early and late stratus burn-off situations well. A model that sometimes forecasts early and late burn-off is preferable to one that only forecasts burn-off in the middle of the day. This behavior is captured by computing the difference between some high and low percentile values (to avoid using any obviously incorrect outliers) of the forecast values.

Data availability is usually an important consideration; some sensors provide data more reliably than other sensors. Preference is given to models that more frequently have their required input data, the assumption being that the relative absence of data in the development set has a bearing on how reliable the data will be in the (independent) future.

Each of these factors is mapped into a value in [0,1]. The factors are then merged by using a weighted sum. The lowest score provides the model that has essentially the lowest possible RMS forecast error and the best performance by these additional criteria. The models are ordered by this score and presented to the developer for final selection.

3.0 AUTOMATION OF MODEL DEVELOPMENT PROCEDURES

3.1 Initial Considerations

This section describes the actual automation procedures used in the implementation of the ideas presented in the previous section. It will of necessity contain repetitions of certain aspects already mentioned, but the focus now is practical rather than descriptive. Illustrations of the various steps will be drawn from our experience in applying the process in the MSI, but the automation has been implemented for easy applicability to future meteorological forecasting tasks.

The codes that implement these procedures are of various types and exist on various platforms, due to a diversity of authorship in a research environment. All of the principal components save one (PGAM) were written in MATLAB. Scripts for pre-processing, formatting, and other such tasks that serve as "glue" between the various main components were written in Perl. Also, when hard numbers are discussed which describe the workings of these stages, they almost always are implemented as parameters which can be easily altered to suit future conditions.

3.2 Base Data

The building of SFMs must be based on historical records of the phenomenon that is to be forecasted (the predictand) and the various independent observations from corresponding times/dates from which predictors are to be drawn.

In the MSI, the data from which potential predictors are drawn were collected as an ongoing project archive. It consists of conventional NWS observations, various quantitative measures extracted from GOES-W visible satellite imagery, and observations from special project-specific sensors (e.g., sodars, radiometers). The satellite and project sensors are available since 1996; NWS observations extend back to 1987. All these data are kept in a project database. Each SFM-building experiment draws its potential predictors from this common data source.

In terms of the MSI, SFMs are only meant to apply to days for which warm-season stratus cloud is present in the San Francisco Bay region. This has led to a formal declaration of which days during the project's years of record are "stratus days." Obviously, days of clear sky would have no need of a forecasted low cloud burn-off. But such a step also removes days from the cold season (i.e., November through April), as well as warm-season days with a low cloud ceiling (and possibly precipitation) due primarily to synoptic-scale systems. Only predictors from these "stratus days" are used for SFM generation and evaluation.

The MSI project database also contains records of the time that side-by-side landing could be initiated corresponding to cloud burn-off for these stratus days. These are the predictands used to optimize and evaluate the various models. They are expressed and used in terms of decimal hours after sunrise.

Future projects may require different concepts of what constitutes a predictand and thus a "case" for development. The MSI assigns one deterministic predictand per date, but future applications may be more complicated. If the need is for probability forecasts, predictands may be simply a (0,1) designation of the absence/presence of the phenomena in question for a certain time, with a single day possibly containing many predictands/cases.

3.3 Organization of Experiments

In the automated process, the building of statistical models is organized according to "experiments." A single experiment encompasses models built from common data – data of the same *type* from the same *cases* using the same pool of possible *predictors*. Each experiment is assigned a unique name, which by convention consists of a system-defined label designating the data type and initialization time, followed by any user-defined text string further describing the experiment. Each experiment is given its own subdirectory of this name, which contains all of its relevant files. Each file within this subdirectory will be identified by the directory name followed by a unique extension. In all the following discussions, files of certain type will be referred to only by their extension.

In the MSI, data type is one of [lsfm, rsfm, ssfm] corresponding to data for the Local, Regional, and Satellite models. The cases within an experiment would be analyst-chosen stratus days, sometimes using a subset of the entire data set that corresponds to a particularly type of meteorological regime. Examples of experiment labels are "lsfm13all" for the Local model for 1300 GMT based on all possible stratus days, and "rsfm15onshore" for the Regional model for 1500 GMT based only on stratus days with onshore low-level winds.

To start an experiment, a user constructs its name, then decides on the cases to base the models on, and what the pool of predictors will be. The former is shown as a list of stratus dates and put into a *.dates* file. The latter is described by a list of predictor labels and put into an *.lbls_out* file. The predictor labels are also constructed by convention. They consist of the time of the observation (2 digits as we are only using top of the hour information), followed by its location, followed by a description. For example, "16sfoCig" indicates the 1600 GMT observation of cloud ceiling at SFO. These two files are all that is needed to begin an experiment. Scripts are then invoked that pull "raw" observations and predictands for these dates from the common project archives.

3.4 Predictor Pre-processing

Model development does not proceed from the raw predictors. They are first subject to quality control. Also, the user has the opportunity to create any "derived" predictors that may be useful. All of these tasks are bundled in a single pre-processing script. This script executes a different set of tasks depending on the type of data presented to it.

When project data is generated in real-time it is subjected to some level of quality control. The quality control performed here has to do with missing values and is of two main types: the standardization of missing value flags, and the possible filling of missing values via intelligent substitution. Missing values can arise at this stage from two causes. The desired predictor for a certain day may be missing in the original file in the archive, or the file itself may be missing for that day. The script which pulls data from the archives provides different missing flags for these two situations. The pre-processing script here converts them both to the same value.

Missing data are the bane of statistical model building, with the problem magnified when the number of cases available for development is limited. Depending on the stage of the process, one missing observation can disqualify an entire case from further use; conversely for comparisons of all possible models that can be built from the same N predictors, all N predictors must be present for each case. Thus we employ any step we can to reasonably fill missing values. Only after some experience working with the predictors will strategies suggest themselves. One simple example from the MSI is to fill missing values from SFO AFOS surface observations (the default) with values from ASOS (which the project database contains for SFO). A slightly more complicated example concerns inland surface pressures. It was found that intermittent drop-outs were common for Sacramento (SAC) surface pressures, and that by re-defining the current pressure as the average of the current and previous hour's observation, if both were present, or the previous or current observation if only one were present, the missing values decreased substantially.

The pre-processor also computes derived predictors. They can consist of several raw observations being combined to create a new predictor or of a raw observation being cast into a new form. An MSI example of this was described earlier, where cloud top (which corresponds to the base of the typical temperature inversion) and cloud base (which corresponds to the surface observation of ceiling) are combined to create cloud thickness. A simpler example is the computation of east-west (u) and north-south (v) wind components from the given wind speed and direction. Another type of derived predictor consists of temporal or spatial differences of the same raw observation. Examples of temporal trends that have been used in the MSI include 1-hour differences in inversion heights and 1-hour differences in cloud coverage in visible satellite imagery for certain sectors of the approach zone. An example of spatial differences employed are the differences in surface pressure between SFO and SAC to the east and ACV (Arcata, CA) to the north, which are commonly used by NWS forecasters as an indication of whether the surface pressure gradient favors on- or off- shore flow.

The result of pre-processing is a file which contains clean values for each candidate predictor, both raw and derived, for each date/case of the experiment. This file is designated by *.init*.

3.5 PGAM

This step accomplishes the rescaling of all potential predictors to a [0,1] range in such a way as to maximize each predictor's correlation with the predictand. This rescaling is expressed as a piece-wise linear function that maps the raw values to their rescaled values. Inputs to this process are the *.init* and predictand files.

The function that describes the non-linear rescaling can be of varying degrees of complexity, as can be seen by the number of *nodes* between the maximum and minimum values. The number of nodes is one less than the number of internal line segments in the function. (Each rescaling includes two line segments that represent the maximum and minimum cutoffs.) Thus the simplest linear scaling would consist of a single line segment and no nodes. A single node would result in two line segments. (An example, with one node, is shown in Figure A-2.) The PGAM software determines the number and location of nodes that maximizes the correlation with the predictand, up to a maximum of two nodes. (The higher complexity and increased computation time of additional nodes was not seen to be justified by the relatively small increases in correlation.) An internal parameter determines whether an increase in correlation is sufficiently greater than one node less to justify the additional node. If the correlation of original predictor is less than 0.2 no increase in correlation is attempted, and only a simple rescaling to [0,1] is done.

Any scaling is described in terms of a *structure function*. This enables the scaling to be applied to new data. The structure function consists of triplets of numbers that represent the endpoints of the linear segments. The three numbers are 1) the original, unscaled value and 2) the rescaled value of the point, along with 3) the slope of the following segment. (See the caption of Figure A-2.) The PGAM software writes a file for each predictor consisting of the correlations and structure functions for 0, 1, and 2 nodes, its determination of the best scaling, and a list of all the resulting scaled values. These files are placed in a *pgam* subdirectory of the experiment directory.

The PGAM software collects the rescaled values of all the predictors into a single file. This rescaled counterpart to *.init* is labeled *.pgam*. All subsequent model building is based on this file.

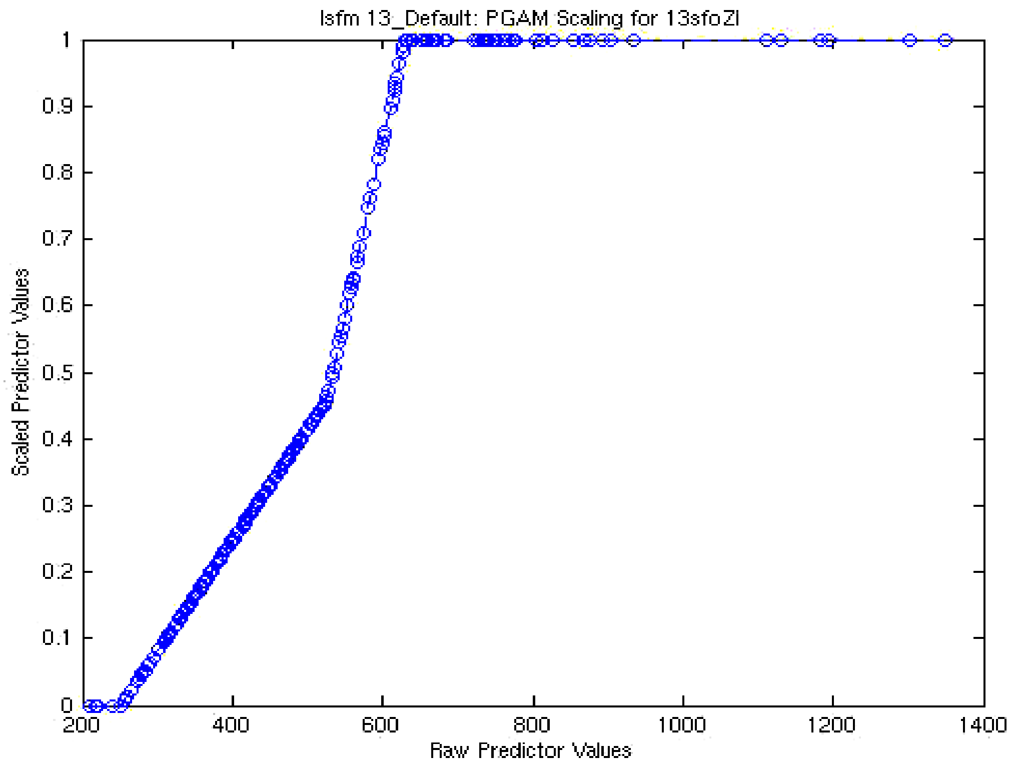


Figure A-2. Piece-wise linear PGAM structure that scales the 13Z SFO inversion height to [0,1]. The structure function can be expressed by the following triplets: (200,0,0), (252,0,0.00167), (525,0.457,.00518), and (630,1,0).

3.6 Predictor Nulling

The predictor nulling procedure attempts to select the most important predictors out of a large set of candidates. Its inputs are the PGAM-scaled values of the predictors, the predictands, and a list of the predictor names. The output consists of a reordered list of these names composed as the process discards the predictors, with those names at the bottom being the most significant.

The problem of missing data is most critical in this step. Initially, the process is considering a large number of predictors, often several hundred. Meaningful decisions about the relative importance of these predictors can only be made on cases for which all predictors are available, a very severe requirement for observational data. The process requires that all predictors are present for at least 25% of the total number of cases, hoping that this will cover enough of the phenomenology to serve as a basis for initial decisions about which predictors to discard. If all predictors are not present for this fraction of cases, predictors with the poorest coverage are discarded, one by one, until common coverage reaches this threshold. Cases can be recovered as predictors are discarded. A command-line argument controls the maximum number of predictors that can be discarded at the end of each nulling cycle. After each cycle, the process looks afresh at the number of cases having full data coverage for the remaining predictors. Since the number of remaining predictors is continually decreasing, the number of cases with full coverage usually increases. To take full advantage of this, the maximum number of predictors discarded per cycle is usually set to the lowest value possible, trading this off against the increase in computation time. In practice, 2-5 nulled predictors per cycle were used.

The only other parameter that can vary with an experiment is an internally set "bias" vector. This vector contains one element per predictor and is used to give preference to certain predictors. In the current implementation the bias vector corresponds to data coverage of the predictors. The nulling process uses this vector to slightly prefer those predictors that exhibit better data coverage, under the assumption that the coverage characteristics of historical data indicate the future reliability of these data. All other parameters are fixed in the current implementation of nulling, but some of these may be made external in future implementations to allow multiple nulling runs with slightly different controlling parameters.

The output is written to a file with extension *.nullout*. Predictors correlated at the level of noise with the predictand (< 0.1) are discarded at the very start. Then the predictors with poor coverage are discarded, as discussed above. Each nulling cycle concludes with the process writing the names of the predictors discarded to *.nullout*, preceded by how many common cases this nulling was based on. Nulling only needs to proceed to the point where the best N (20 for the MSI) predictors are determined. Attempting to null beyond this point can increase processing time substantially, as it is often difficult for the nulling to choose among the best predictors. Consequently, nulling terminates when the number of remaining predictors first falls below an internal parameter, currently set to 20. This is marked in *.nullout* at the location of the constant "predictor" name (*const)*. An example of such a file illustrating these features is shown in Figure A-3.

In practice, the order in which predictors are nulled was found to be sensitive to the ordering of the input predictors, to the point of often making a non-trivial difference in the composition of the final set of important predictors. This input ordering in no way affects the numerical content of the input, but does affect the optimization starting point for each nulling cycle. This, and the high correlation between many of the input predictors, leads us to attribute this behavior to numerical instability and round-off issues. Experiments that took these differing sets of most important predictors all the way through the process to their respective final models found no significant differences in model performance or quality. However, to make the nulling results repeatable a step was added to the procedure that initially sorts all predictors by an internal "importance" measure.

```
---Corr<0.10---          13smbPrs                ---0282cases---
13sfoZidt1               13smbPrsdt1        const
13sqlZidt1               12sfoDir           13sfoZi
13sfoCigdt1              12sfoU             13sqlZi
13sfoTh1dt1              12sqllDir          13sfoTh1
13sfoDir                 12sqllV            13sqllSpd
13sfoDirdt1              12Dir925           13smbCigdt1
13sfoSpddt1              12Spd925           13smbTh1
13sfoU                   ---0278cases---    13acvPdx
13sfoUdt1                13sfoSpd           13sacPdx
13sfoVdt1                13sfoV             13rblPdx
13sfoPrs                 ---0279cases---    12sfoZi
13sfoPrsdt1              13sqllU            12sqlZi
13sqllDir                12sfoTh1           12sfoSpd
13sqllDirdt1             ---0280cases---    12sfoV
13sqllSpddt1             13avgZi            12sqllSpd
13sqllUdt1               13rnoPdx           12sqllU
13sqllV                  ---0282cases---    12smbTh1
13sqllVdt1               12avgZi            12U925
13smbTh1dt1              13suuHCld          12V925
```

Figure A-3. Sample .nullout file.

3.7 Selection of Initial Candidate Models: All-paths Testing

This stage determines a (relatively) small subset of all possible combinations of the best N predictors, which are the most promising candidates to subject to thorough quasi-independent cross-validation in the next stage. Just over one million models result from all possible combinations of 20 predictors. This stage determines the 200 (maximum) models of each size (1-20) with the best fit to the entire development set, based on simple linear regression.

To construct inputs for this stage, the last N predictor names are taken from the end of *.nullout*. The PGAM scaled values for these names are extracted to form a smaller matrix of predictors. Any cases, which contain missing predictor values, are discarded. A column of 1's is added to this matrix to enable the fitting of the constant term.

The generic combinations themselves are created separately and kept in a static file, which also serves as input for this stage. Independent code exists that determines all lexicographic combinations of integers from 1 to N up to maximum number of integers M. (In the MSI both N and M are 20.) This stage reads from this file line by line and uses each combination of integers to index directly into the predictor matrix.

The size of model considered is capped based on the number of cases in the development data set. A general rule of thumb is to have at least 10 cases to support each predictor in a final model. Consequently, no models with sizes greater than the one-tenth the number of cases are fit or saved. This is a guard against eventually settling on an over-fit model for real-time implementation, and also saves considerably on processing time for this stage.

As each model/combination is fit to the predictand, its characteristics are saved in a structure. These characteristics include the indices and predictor names of the model, the predictor weights, the fit of the model to all of the development cases (i.e., sigma, or the square root of the mean squared model error adjusted for the number of parameters in the model), and r-squared (the fraction of the total variance of the predictand explained by this model). An example of such a structure is shown in Figure A-4. An output matrix of these structures is created, with N columns (one for each model size) and 200 rows to hold the best fits. As each model is created and evaluated, its sigma is compared to the 200th largest sigma thus far for models of that size. If the current model has a lower sigma, it replaces the model with the 200th largest value. This matrix of candidate models is the output for this stage of the process, and is saved to a file for further use downstream.

```
>> MDLS(1,5)

       comb: [10 12 13 16 18]
      names: {'13rblPdx 13acvPdx 13smbTh1 13sfoTh1 13sfoZi CONST '}
    weights: [1.8940 1.2123 1.0134 1.3102 1.0358 2.1577]
      sigma: 1.0284
         r2: 0.4589
     ncases: 311
```

Figure A-4. Example descriptive model structure after fitting to all development data, showing the 5-predictor model with best fit for this experiment.

Computationally, this is the most time-consuming part of the process. At the beginning of development for the MSI , 400 MHz Sun Ultra 5s were used. Fitting just over a million models for 300 cases took roughly four hours. Later this stage was transferred to a 2.4 GHz dual-processor PC running MATLAB under LINUX. Processing time dropped by a factor of 8 to roughly one-half hour.

3.8 Cross-validation of Candidate Models

The cross-validation stage subjects the initial candidate models to a k-fold cross-validation. For reasons explained earlier, this is an attempt to determine which models will perform the best on independent data in the absence of a true independent evaluation set. The input consists of the same predictor matrix and predictand vector used in the previous stage (with no cases containing missing values), the matrix of candidate models, and a list of "preferred predictors" to be described below.

This process first takes the matrix of candidate models and deletes those elements that are empty or that have the poorest fits to all the development cases (i.e., > 1 standard deviation from the mean fit). These obviously poor models result from the need to present models of each size to the cross-validation. Aside from experiments with a very small number of development cases, it is doubtful that any of the one-parameter models, for example, would be among the final candidates. After this trimming of models, the matrix of candidates is reformatted as a vector.

All of the remaining models are then subjected to a k-fold cross-validation scheme. In this scheme as currently implemented, 10 random 10-fold splits of the training cases are made (to guard against a single 10-fold split possibly being grossly inequitable). Cross-validation is applied 10 times, with the result being a vector of quasi-independent forecasts, 10 for each case. This vector is the basis for additional evaluative measures that are computed and added to the structure element for the model. Among these measures, as described earlier, are:

- The cross-validation fit, or sigma, computed as in the fit to all the development data.
- The "confidence," being the fraction of the forecasts that fall within some desired error bound, taken to be +/- 0.5 hours for the MSI.
- The "range," being the difference between the latest and earlier forecasts, after trimming the 5^{th} percentile values at each end to guard against considering outliers.
- The "stability" of the predictor weights, being the maximum standard deviation of the K weights for each predictor that result from the K model fittings.
- The number of cases in the full development set that this model would be able to provide forecasts for. Actual fitting is only done for cases in which all N predictors are present, but this measure expresses how many cases with non-missing predictors for this particular model (which consists of some subset of the N predictors) are present in the development data.
- The total number of predictors (not counting the constant term) in the model.
- The number of preferred predictors (described below) that this model contains.

Future projects may lead to the construction of additional measures, which may be useful aid in selecting the best models. All of these measures are computed in the same function of the MATLAB code, which can be easily modified to suit future needs.

Selections of these measures are averaged together to arrive at an overall score. To facilitate this, scaled versions of the above measures are computed. The scaling is from the range 0 to the largest value of the measure among all the models to [0,1], and is done in such a way that smaller values of the scaled measure always mean a better model. For example, for model fit (sigma) smaller values are better, but for unscaled confidence values, higher values correspond to better models. The scaling standardizes the sense across all measures to enable them to be blended into a single number. These scaled measures are added to the model structure. A sample resulting structure is shown in Figure A-5.

```
>> MDLS(1)

        comb: [10 12 13 15 16 18]
       names: '13rblPdx 13acvPdx 13smbTh1 13sqllSpd 13sfoTh1
               13sfoZi CONST'
     weights: [1.6866 1.2070 1.0237 0.5236 1.1072 1.0169 2.1679]
       sigma: 1.0242
          r2: 0.4646
      ncases: 305
     ind_fit: 1.2556
 weight_stab: 0.1319
        conf: 0.4468
       range: 5.5192
      npreds: 6
       npref: [1 2 0]
  scale_ifit: 0.8833
  scale_stab: 0.0679
  scale_conf: 0.0526
 scale_range: 0.0492
scale_ncases: 0.1528
scale_npreds: 0.3000
 scale_npref: 0.7692
  eval_score: 0.1433
```

Figure A-5. Example of descriptive model structure after cross-validation. This is the model with best "eval score" for this experiment.

The user has the option to define a set of "preferred predictors" that can be used to filter the models considered, or merely to track how many such predictors each model contains. This allows user experience and intuition with regards to important predictors to enter the process. The list of preferred predictors can be completely specified by name (e.g., "sfoDewpt") or merely indicated by a string to match (e.g., "sfo" for any predictor from SFO). A threshold can be set which only passes models with the required number of such predictors on for cross-validation. Additionally, the preferred predictors can be organized into any number of groups, each with its own threshold.

Once the cross-validation of each model has been accomplished and any filtering by preferred predictors applied, a final set of candidate models is determined. All of the "equivalently best" models are separated by considering those whose cross-validation fits are within 1 minute of the best fit, discarding the rest. The other evaluative measures are then averaged to compute overall evaluation scores (named "eval_score" in the model structure, as seen in Figure A-5). This is a weighted average, in which a parameterized vector contains the weights for each measure and is able to ignore any measure by assigning it a weight of 0. The standard deviation of all measures with non-zero weights is also computed and included in the weighted average. This is done so that preference is given to models with uniformly good measures across all the chosen measures, as opposed to models which may have one excellent measure with the rest being mediocre.

After much experimentation, the MSI used an equal weighting of the stability, confidence, number of potential cases, and number of predictors to compute the evaluation scores. Future projects can determine different weightings.

The user has several command-line arguments that give further control over the finalist models that are output. On rare occasions all of the finalist models from an experiment contain an unacceptably large number of predictors (e.g., 15 or greater), a number perhaps supportable given the number of development cases but too large to be comfortably implemented. Therefore the user can specify an upper bound to the size of model to be subjected to cross-validation. Also, again on rare occasions, an experiment can produce a set of finalist models that all have unacceptably poor case coverage; i.e., they include predictors whose data availability has been historically unreliable such that they would be able to forecast for an unacceptably small fraction of cases. This fact may be disguised because the models are evaluated relative to one another in this regard. Other models may exist in the experiment that have degraded performance but acceptable case coverage. The user has recourse to an additional command-line argument that screens the models subjected to cross-validation by some minimal coverage threshold.

The finalist models are ordered by their overall scores and saved to a file for subsequent evaluation and use.

3.9 Final Model Selection

At this stage the user has a set of models, all nearly equivalent in terms of their fit to quasi-independent data, and ordered by some combination of other measures. The choice of a final model for real-time implementation is now left to the analyst. Several summaries can be created to help in this task. One is a text file summary of the top models sorted by their overall evaluation score, with one line summarizing each model. An example of portions of such a file is shown in Figure A-6 predictors. The summary lines contain several of the evaluation measures along with the indices of the comprising the model. The end of this file contains several summaries of overall predictor use and coverage for all of these finalist models. A graphical summary of aspects of this file is also created. An example is shown in Figure A-7.

```
Cross-validation summary: max model size 20, 20 possible predictors
Number of final models: 70

Top 10 models:
RNK scor ifit(scal) conf(scal) stab(scal) rnge(scal)  n  prfp    Combination
 1  0.14 1.26(0.88) 0.45(0.05) 0.13(0.07) 5.52(0.05) 305   3   10 12 13 15 16 18
 2  0.15 1.25(0.88) 0.45(0.05) 0.17(0.09) 5.52(0.05) 307   4   3 10 12 16 17 20
 3  0.15 1.26(0.88) 0.44(0.06) 0.17(0.09) 5.51(0.05) 307   4   3 9 10 12 16 17
 4  0.16 1.25(0.88) 0.46(0.03) 0.16(0.08) 5.61(0.03) 302   4   3 10 12 15 16 17 20
 5  0.16 1.25(0.88) 0.43(0.09) 0.20(0.10) 5.54(0.04) 309   4   9 10 12 13 16 18
 6  0.16 1.25(0.88) 0.46(0.03) 0.20(0.10) 5.55(0.04) 304   4   10 12 13 15 16 18 20
 7  0.16 1.25(0.88) 0.45(0.05) 0.17(0.09) 5.58(0.04) 302   5   3 4 10 12 16 17 20
 8  0.16 1.25(0.88) 0.45(0.05) 0.17(0.09) 5.61(0.03) 302   4   3 9 10 12 15 16 17
 9  0.16 1.24(0.87) 0.42(0.11) 0.18(0.09) 5.52(0.05) 307   4   3 9 10 12 16 18
10  0.16 1.25(0.88) 0.45(0.05) 0.20(0.10) 5.66(0.02) 304   4   9 10 12 13 15 16 18

Predictors Use in Final Models
Pnum           Pname   NModels     NCov
  3         12smbTh1    45(0.64)   348(0.87)
  4          12sqllU    11(0.16)   376(0.94)
  5        12sqllSpd    11(0.16)   376(0.94)
  6           12sfoV    11(0.16)   398(1.00)
  7         12sfoSpd     5(0.07)   398(1.00)
  8          12sqlZi    15(0.21)   387(0.97)
  9          12sfoZi    55(0.79)   387(0.97)
 10         13rblPdx    70(1.00)   374(0.94)
 11         13sacPdx     5(0.07)   370(0.93)
 12         13acvPdx    70(1.00)   363(0.91)
 13         13smbTh1    36(0.51)   351(0.88)
 14      13smbCigdt1     4(0.06)   358(0.90)
 15        13sqllSpd    28(0.40)   376(0.94)
 16         13sfoTh1    70(1.00)   381(0.96)
 17          13sqlZi    10(0.14)   389(0.98)
 18          13sfoZi    65(0.93)   389(0.98)
 20          12avgZi    14(0.20)   387(0.97)

Distribution of Sizes of Final Models:
Size N_Mdls
   6     5
   7    28
   8    34
   9     3
```

Figure A-6. Sample summary of cross-validation results.

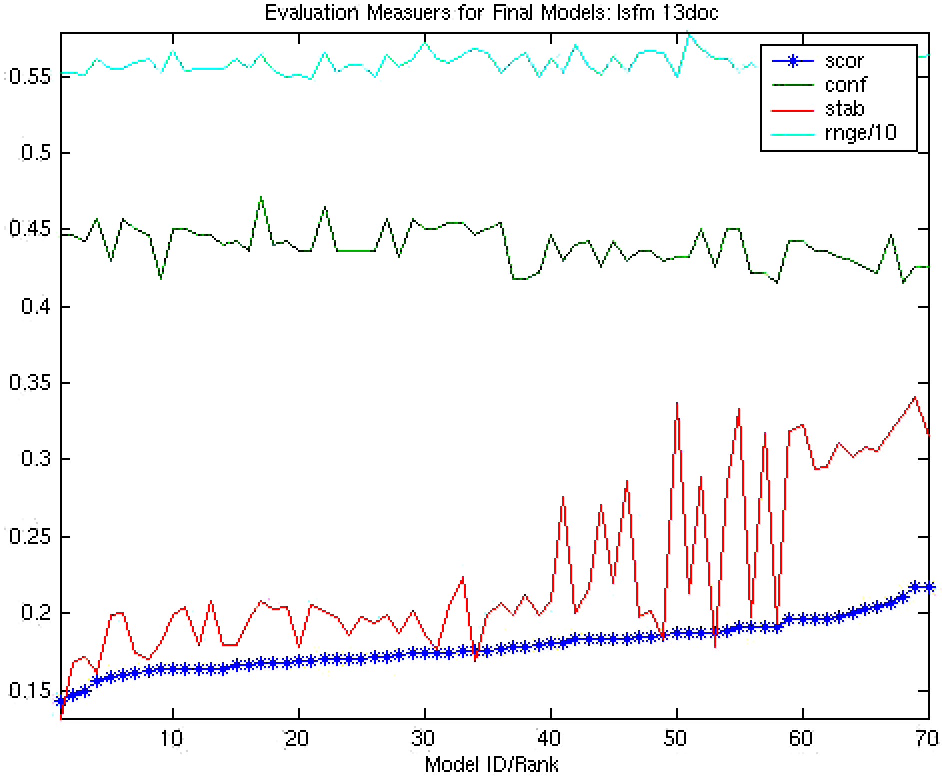


Figure A-7. Graphical depiction of evaluation score (scor), confidence (conf), stability (stab), and range (rnge/10) for the finalist models of a sample experiment. The models are sorted by evaluation score.

3.10 Generation of Forecasts

It was found desirable to know the forecasts that would result from selected candidate models when applied to any arbitrary set of cases. For instance, early stages of the development of this process considered how models developed on MSI 1996-2000 data would have performed on 2001 cases. To accomplish this, utilities were developed that pulled predictor and verification values for any set of cases, dependent or independent of the ones comprising the experiment, and placed these in a subdirectory named *fcst* within the experiment directory. The predictor values are PGAM-scaled using the structure functions already determined for the development data. Forecasts are made using any of the finalist models identified by their index, and the results stored in a file. The forecasts can then be evaluated and compared with those from other models. A graphical summary of the forecasts produced are shown in Figure A-8.

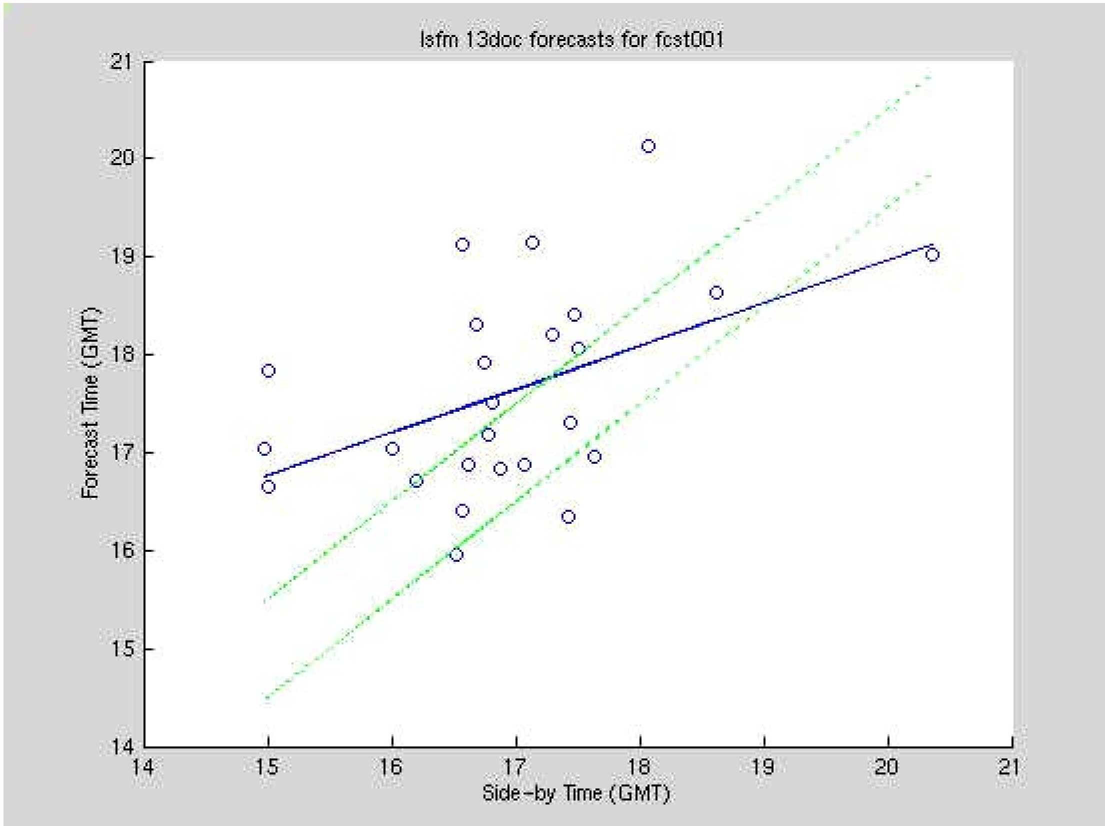


Figure A-8. Scatter of forecasts versus verifications produced by the "forecast" subsystem for a sample model. The blue line is the linear fit of this relationship. The parallel green lines bound the +/- 0.5 hour "confidence" window.

3.11 Construction of Real-Time Control Files

Eventually from an experiment a single model may be chosen for real-time forecasting. Various tools were developed that creates the specific file that the MSI's real-time system needs to implement the model. This file contained three lines for each predictor, specifying the predictor's name, weight, and structure function that accomplishes the PGAM scaling. These tools can be altered to suit the real-time needs of any future project.

3.12 Unresolved Automation Issues

The automated SFM-building process has been used with success by several analysts for the MSI. Extensive use of the process, however, has uncovered certain issues and ambiguities that need to be addressed. Several are collected in this section. None of these issues preclude successful use of the process, but addressing them may lead to improvements in the final models produced. Each of the following items represents possible areas for future research and experimentation.

Currently PGAM scaling of the raw predictors is performed from scratch at the beginning of each experiment, and is based on only its cases. The same predictor may thus have significantly different structure functions for different experiments. It has been suggested that a single pre-experiment PGAM scaling could be accomplished for each predictor once, based on all relevant data. This would save computation time, as the predictor data would already be scaled at the initiation of each experiment. A single scaling would more fully capture the relationship between the predictor and predictand, minimizing the constitution of the specific experiment. A single scaling may be preferable for experiments with few cases, where the current method which performs the scaling based on these few cases may lead to peculiar structure functions. It is unknown what the difference in final model goodness is for scaling performed at these two different points of the process.

Sometimes the PGAM scaling for a certain predictor yields a structure function that transforms the original predictor to essentially a binary switch. All raw values are turned into either 0s or 1s. It is not clear how best to use such predictors. In models with relatively few predictors, the presence of a single binary predictor, depending on its weight, can lead to large swings in the forecasted value. They may indicate the need to split the experiment in two, developing separate models for each binary value. For example (from the MSI), a ceiling predictor in the regional model may yield a scaling such that all values less than 3000 feet map to 0 and all values greater than 3000 feet map to 1. Separate models could be created , if enough cases are present, for all cases with the lower and higher ceilings. Currently the nulling process can indicate which input predictors are binary, with no automated special treatment of such predictors beyond this notification.

As mentioned earlier, the set of important predictors determined by the predictor nulling routine was found in practice to be somewhat unstable with regards to its input predictors. Sensitivity to the ordering of input predictors has been addressed. However, the presence/absence of as little as a single input predictor also results in non-trivial differences in the order in which these predictors are nulled and the content of the set of important predictors. This behavior is magnified with greater numbers of initial predictors. This is likely due to the varying distributions of missing cases among these predictors, as the available cases depend on full data coverage among all the available predictors. Adding a predictor changes the number and constitution of available cases, which can have pronounced effects. In practice this has led to a more conservative approach to constructing the initial pool of possible predictors that is presented to the nulling. (At beginning stages of using the automated process, a “kitchen sink” approach was often used, especially with regards to the construction of derived predictors.) Constructing the set of important predictors from several nulling runs with slightly tweaked parameters may yield a more stable result.

The desire has arisen to force an experiment to contain a certain set of predictors. In the MSI this was pursued as a way to achieve forecast consistency within a set of models. Thus if each regional model, for example, from 13Z through 18Z was forced to use several common predictors, greater consistency of these forecasts may result. Accomplishing this is possible in the current process but involves excessive manual interventions, and does not extend to the nulling or the initial candidate model stages. The process could be improved to facilitate this goal.

GLOSSARY

ARTCC	Air Route Traffic Control Center
AWRP	Aviation Weather Research Program
CFA	Consensus Forecast Algorithm
CWSU	Central Weather Service Unit
FA	Free Atmosphere
LCL	Lifted Condensation Level
LSFM	Local Statistical Forecast Model
MIT/LL	Massachusetts Institute of Technology/Lincoln Laboratory
NWS	National Weather Service
OAK	Oakland
PGAM	Pre-scaled Generalized Additive Modeling
RSFM	Regional Statistical Forecast Model
SFO	San Francisco International Airport
SODAR	Sonic Detection and Range
SQL	San Carlos Airport
SSFM	Satellite Statistical Forecast Model
UQAM	University of Quebec at Montreal

REFERENCES

1. Farley, John, personal communication, San Jose State University.

2. Turner, S., R. Tardif, P. Zwack, C. Page, N. Gauthier, and E. Monteiro, "Stratus burnoff forecasting with the COBEL column model – product description," <Publication TBD>, University of Quebec, Montreal, Canada, 2004.

3. Vislocky, R.V. and J.M. Fritsch, "Generalized additive models versus linear regression in generating probabilistic MOS forecasts of aviation weather parameters," *Wea. Forecasting*, **10**, 669-680, 1995

4. Wilson, F.W. and D.A. Clark, "Interim Report: Products of the SFO Marine Stratus Initiative," Wx Project Memorandum No. 43PM Wx-0058, MIT Lincoln Laboratory, Lexington, MA, 1999.

SFO Marine Stratus Forecast System Documentation

Section VI. Troubleshooting Guide

TABLE OF CONTENTS

1. INTRODUCTION

1.1 SCOPE

This document provides a troubleshooting guide for the San Francisco Marine Stratus Forecast System. Under normal operating conditions (in which the base station and networking hardware are working properly and electronic mail relay service provided by the phone company is working properly) designated NWS recipients receive electronic mail from the base station when problems detected by the base station arise. The mail message or messages sent out may identify a problem directly and provide a solution to correct the problem. They may also provide a procedure to follow (to identify the problem from among multiple problems that produce similar effects) before suggesting a course of action to correct the problem. The purpose of this document is to provide guidance in identifying and solving problems when the base station is not providing electronic mail messages either because the mail system is not functioning properly, one or more networking components at the CWSU has malfunctioned, or the base station has malfunctioned.

1.2 RELATED DOCUMENTS

Appendix A to the document “Real-Time System” contains the list of mail messages that collectively describe all the system problems (and actions to take to correct the problems) detectable by the base station.

2. OVERVIEW

Guide A provides a starting point to identify whether a problem is related to web display, data acquisition, or data processing and also provides a FAQ for model forecast generation. Guide B should be referenced when the web display page is not being served to the local client. Guides C, D and E address problems related to the acquisition of data from the field sites, the AWOS feed, and the ASOS feed, respectively. Guide F should be referenced for problems related to the satellite feed from NRL/Monterey. Finally, Guide G documents files on the base station that may be helpful to examine both for troubleshooting and to check the status of the communication links between the base station and the field sites and the health of the field site computers.

Troubleshooting Guide A. Quick Reference for Web Display, Data Acquisition and Processing and Forecast Model Activation

Problem	Explanation/Action
One or more data displays on the web page is not updating with current data.	1. Manually refresh the web page frame. 2. Refer to the relevant troubleshooting guide (B through F). 3. Check to see the cvRTCache is running: % ps -ef | grep Cache should return 2 entries: /bin/csh /ll/cv/scripts/Rts/start_cvRTCache /ll/cv/bin/i586/cvRTCache
cvRTCache is not running.	Manually restart the cvRTCache (See Appendix F, Section III).
Data acquisition appears ok and the cvRTCache is running but the web page is not updating with current data.	1. Check the raw file that corresponds to the data in question (See Table 1, Section III). 2. For field-site data, if the file modification time is current but the data are old, suspect a problem at one of the field sites. Refer to guide C, D, or E. 3. For internet-based data, if the file modification time is current but the data are old, suspect a problem with the data provider. Refer to guide F.
One or more component models didn't make a forecast.	1. Has the web page removed "Approach Clear"? 2. Does the component model normally forecast for the run-time in question? (See Table 10, Section III) 3. Was the default model missing required predictors? (See Table 9, Section III) 4. Is there a lock file for the model launching script in the lock directory */ll/cv/Locks*? If so, contact MIT/LL.
Why did the default model run instead of the daytype model?	Was the daytype model missing required predictors? (See Table 9, Section III) If not, contact MIT/LL.
Why did it take so long for the models to post their forecasts?	1. Did the models finally post at the end interval time? If so, check to see if one or more models didn't run. Consensus waits until the end interval time before posting an incomplete set of forecasts. 2. If all the models ran, access the pull down menu labeled "Forecast Model Details" and choose "Raw Model Output Summary". Check to see when the quality of all the models was sufficient to allow posting.

Troubleshooting Guide B. No Web Display Page

Web page does not display. A message similar to “Could not connect to remote machine: connection timed out” is received.

↓

Ping the CWSU SBC router:
64.175.253.254

↓

If the CWSU SBC router does not respond, try telebooting the router:
510 793-3623 /bo 5,Y

↓

Otherwise, ping the CWSU PIX firewall:
64.175.253.251

↓

If the CWSU PIX firewall does not respond, try telebooting the firewall:
510 793-3623 /bo 2,Y

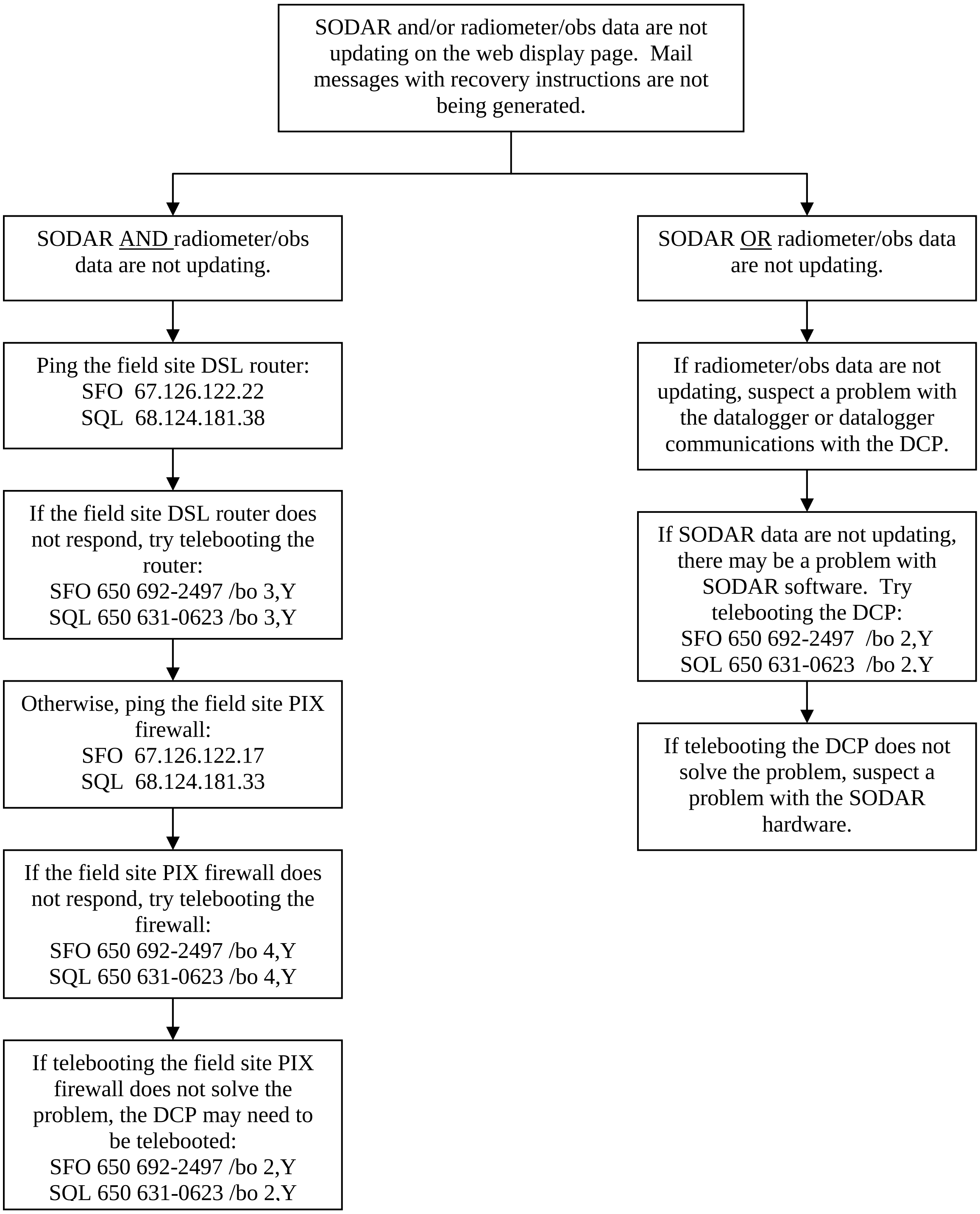
↓

Otherwise, ping the CWSU base station:
64.175.253.249

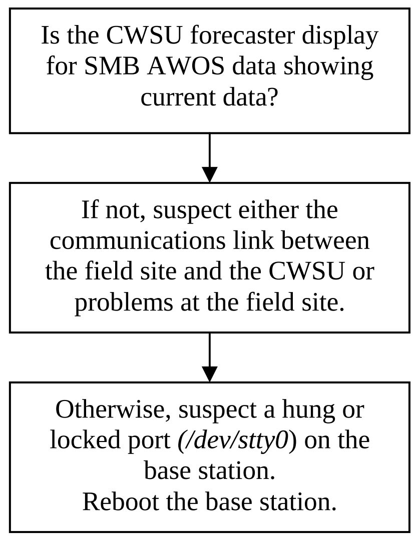
↓

If the CWSU base station does not respond, try telebooting the base station:
510 793-3623 /bo 1,Y

Troubleshooting Guide C. Field Site Data From One or More Sensors are Not Updating on the Web Display Page



Troubleshooting Guide D. Data From the AWOS Feed (San Mateo Bridge) are Not Updating on the Web Display Page



Troubleshooting Guide E. Data from the NWS ASOS Feed are Not Updating on the Web Display Page

On the base station, check the log file */ll/cv/call.log* to see if there are successive failures to acquire data.

↓

If the logs show a modem connection is being made but no data are being transferred, suspect the ASOS password has changed. Contact Jeff Helms
jeffrey.helms@noaa.gov
831 656-1713 ext. 262

↓

Otherwise, manually dial the ASOS phone number from a separate phone on a known working phone line:
1 650 872-0129

↓

If a modem connect signal is not heard, suspect a problem with either the remote modem at the NWS or the phone line. Contact either Jeff Helms or SBC.

↓

Otherwise, check the log file */var/messages* to see if the system is complaining about the data modem */dev/ttyS5*.

↓

If so, cycle power to the modem manually or try a manual teleboot of the data modem:
510 793-3623 /bo 3,Y

↓

Monitor the ASOS phone calls to see if the calls go through. If they continue to fail, suspect the port */dev/ttyS5* is hung. Reboot the base station.

Troubleshooting Guide F. Satellite Data From NRL/Monterey are Not Updating on the Web Display Page

From the base station, run the UNIX command “ls –ltr” on the directory */ll/cv/data/GOES/Raw/YYMMDD* where *YYMMDD* is the current year, month, and day.

↓

If the UNIX modification date is recent but the time stamp embedded in the file name is old, NRL is not providing updated imagery on the NRL FTP site.
Contact Joe Turk
turk@nrlmry.navy.mil
831 656-4888

↓

If the UNIX modification date is old, suspect either DSL is down or the NRL FTP site is down. Check independent sources to determine if DSL is down. If DSL is ok, suspect the NRL FTP site.
Contact Joe Turk

Troubleshooting Guide G. Resources on the Base Station to Monitor System Status and to Assist in Troubleshooting

1. Log files located in */ll/cv*:

   a. sfodsl.log — Health of DSL and phone line to SFO field site.
   b. sqldsl.log — Health of DSL and phone line to SQL field site.
   c. sfoping.log — Ping status between base station and SFO field site.
   d. sqlping.log — Ping status between base station and SQL field site.
   e. call.log — Status of data transfers from field sites and ASOS feed.

2. Commands that can be run from the base station (cv account) to check DSL service to the field sites:

   a. ftpsfdcp — If ftp service does not respond, DSL at SFO is down.
   b. ftpscdcp — If ftp service does not respond, DSL at SQL is down.

3. Files located in */ll/cv/src/scatter-ftp/kermit/sfdcp[scdcp]*:

   a. count.0 — UNIX modification time is time of last DSL transfer. A missing file indicates the last DSL contact failed.

   b. phone — UNIX modification time is time of last phone transfer. A missing phone file indicates the last phone contact failed.

4. Files located in */var*:

   a. messages — Messages related to */dev/ttyS4* (teleboot modem) and */dev/ttyS5* (data modem)