An Overview of the Forecast Systems Laboratory (FSL) Meteorological Assimilation Data Ingest System (MADIS)

Draft

8/12/2004

NOAA

noaaresearch

FORECAST SYSTEMS LABORATORY

BOULDER, COLORADO

Introduction

FSL's MADIS Project makes integrated, quality-controlled observations available to the meteorological community. An overview of MADIS capabilities is given in this document. Comments and (or) questions should be sent to patricia.a.miller@noaa.gov. For more information, see http://www-sdd.fsl.noaa.gov/MADIS.

Ingest

MADIS provides ingest, integration, automated quality control (QC), and distribution support for both NOAA and non-NOAA observations. Observations currently supported by MADIS include:

- Meteorological Aviation Reports (METARs)
- Surface Aviation Observations (SAOs) *
- Surface maritime reports (including Buoy, Ship, and Coastal-Marine Automated Network [C-MAN])
- Modernized NWS Cooperative Observer (COOP-M) Program *
- Surface Mesonets * from
  - Aberdeen Proving Ground
  - Amateur Radio Operators APRSWXNET/CWOP program
  - AWS Convergence Technologies (discontinued on 6/1/04)
  - Anything Weather
  - Colorado Department of Transportation
  - Colorado E-470 Road Weather Information System
  - Florida Automated Weather Network
  - Ft. Collins Utilities
  - NWS Goodland forecast office miscellaneous observations
  - Gulf of Maine Oceanic Observing System
  - FSL GPSMET observing system
  - Hydrometeorological Automated Data System
  - Iowa Department of Transportation
  - NWS Boulder forecast office miscellaneous observations
  - Kansas Department of Transportation
  - Louisiana State University/Jackson State University
    - Mississippi Mesonet
    - Louisiana Agriclimatic Information System
  - MesoWest
  - Minnesota Department of Transportation
  - NOS PORTS
  - Oklahoma Mesonet
  - Interagency Fire Center RAWS network
  - Denver Urban Drainage and Flood Control District
  - Wisconsin Department of Transportation
  - Weather for You
  - West Texas Mesonet
- Radiosondes
- NOAA Profiler Network (NPN)
- Cooperating Agency Profiler (CAP) *
- Automated aircraft reports and profiles (including MDCRS, ACARS*, AMDAR*, and TAMDAR * observations)
- Microwave Radiometers *
- GOES 3-h operational winds
- GOES 1-h experimental winds *

* *Indicates datasets not currently available on NWS AWIPS via NOAAPORT*

MADIS is also currently serving as the modernized COOP (COOP-M) Central Facility by ingesting raw COOP-M observations from five different datastreams, integrating them, applying automated QC procedures, and distributing them to interested parties and organizations. Automated monitoring tools designed to locate communications outages at individual COOP-M stations are also supported by MADIS.

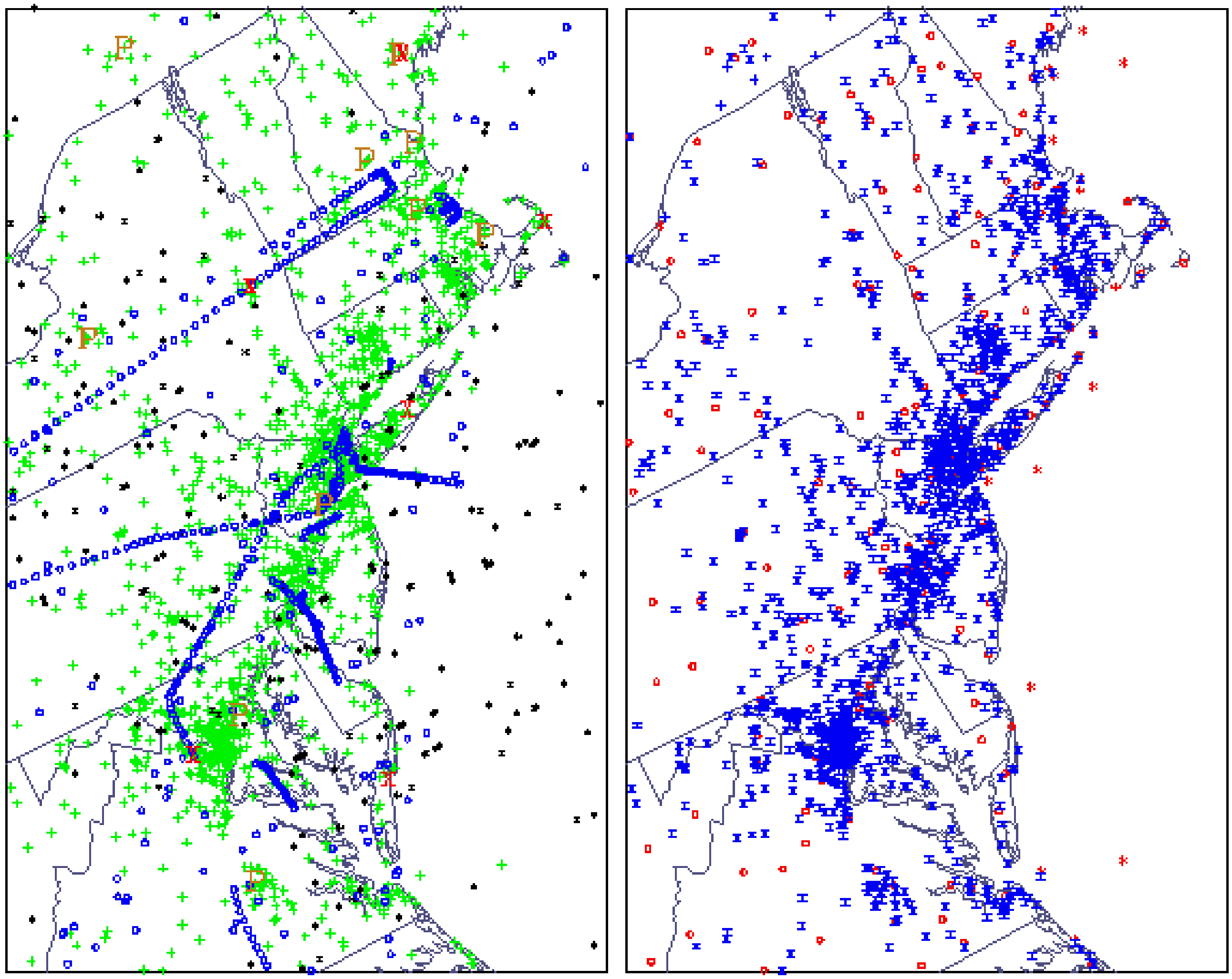

Figure 1. MADIS observations available over the east coast of the U.S. on April 29, 2004.

a - left) All MADIS observations at 00Z. Surface observations are in green, aircraft reports in blue, and satellite winds in black. Profiler locations are indicated by a "P" and radiosonde locations with "X".

b - right) MADIS surface observations at 12Z. Standard METAR and maritime data are in red. Locations of MADIS mesonet and COOP-M stations are indicated by blue "X"s.

Integration

Raw data files are received from many different data providers in many different data formats, and contain observations in various units from stations reporting from various time zones.

MADIS ingests these data files, combines the observations from non-NOAA data providers, and integrates them with NOAA datasets by encoding all of the data into a uniform format and converting all of the observations to standard observation units and time stamps.

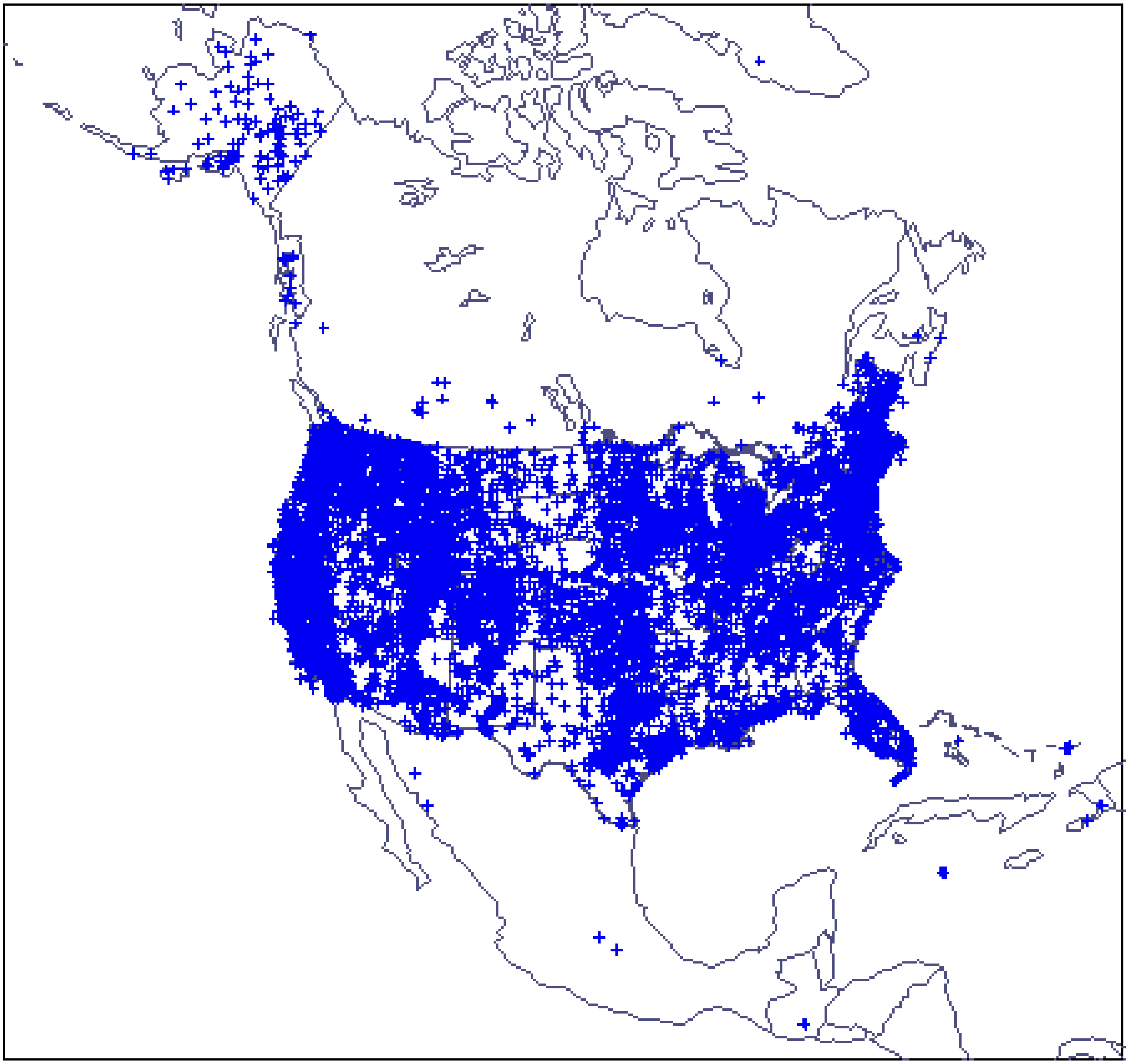

Figure 2. MADIS mesonet stations in the North-American domain on May 26, 2004.

Automated Quality Control

Observations are stored in the MADIS database with a series of flags indicating the quality of the observation from a variety of perspectives (e.g. temporal consistency and spatial consistency), or more precisely, a series of flags indicating the results of various QC checks. MADIS users and their applications can then inspect the flags and decide whether or not to use the observation.

Two categories of QC checks, static and dynamic, are implemented for each observation type. The static QC checks are single-station, single-time checks which, as such, are unaware of the previous and current meteorological or hydrologic situation described by other observations and grids. Checks falling into this category include validity, internal consistency, and vertical consistency. Although useful for locating extreme outliers in the observational database, the static checks can have difficulty with statistically reasonable, but invalid data.

To address these difficulties. MADIS also implements dynamic checks which refine the QC information by taking advantage of other available hydrometeorological information.

Examples of dynamic QC checks include position consistency, temporal consistency, and spatial consistency. QC checks are run on a subhourly basis to guarantee the timeliness of the information to MADIS users. Static QC checks, for example, are applied every five minutes to newly arrived surface observations, while the spatial consistency check is run every 15 minutes.

The spatial consistency check is performed using an Optimal Interpolation (OI) technique. At each observation location, the difference between the measured value and the value analyzed by OI is computed. If the magnitude of the difference is small, the observation agrees with its neighbors and is considered correct. If, however, the difference is large, either the observation being checked or one of the observations used in the analysis is bad.

To determine which is the case, a reanalysis to the observation location is performed by eliminating one neighboring observation at a time. If successively eliminating each neighbor does not produce an analysis that agrees with the target observation (the observation being checked), the observation is flagged as bad. If eliminating one of the neighboring observations produces an analysis that agrees with the target observation, then the target observation is flagged as "good" and the neighbor is flagged as "suspect."

Suspect observations are not used in subsequent OI analyses. The reanalysis step is particularly important when combining data from well-maintained and well-sited observation systems with data from less advantaged observing systems.

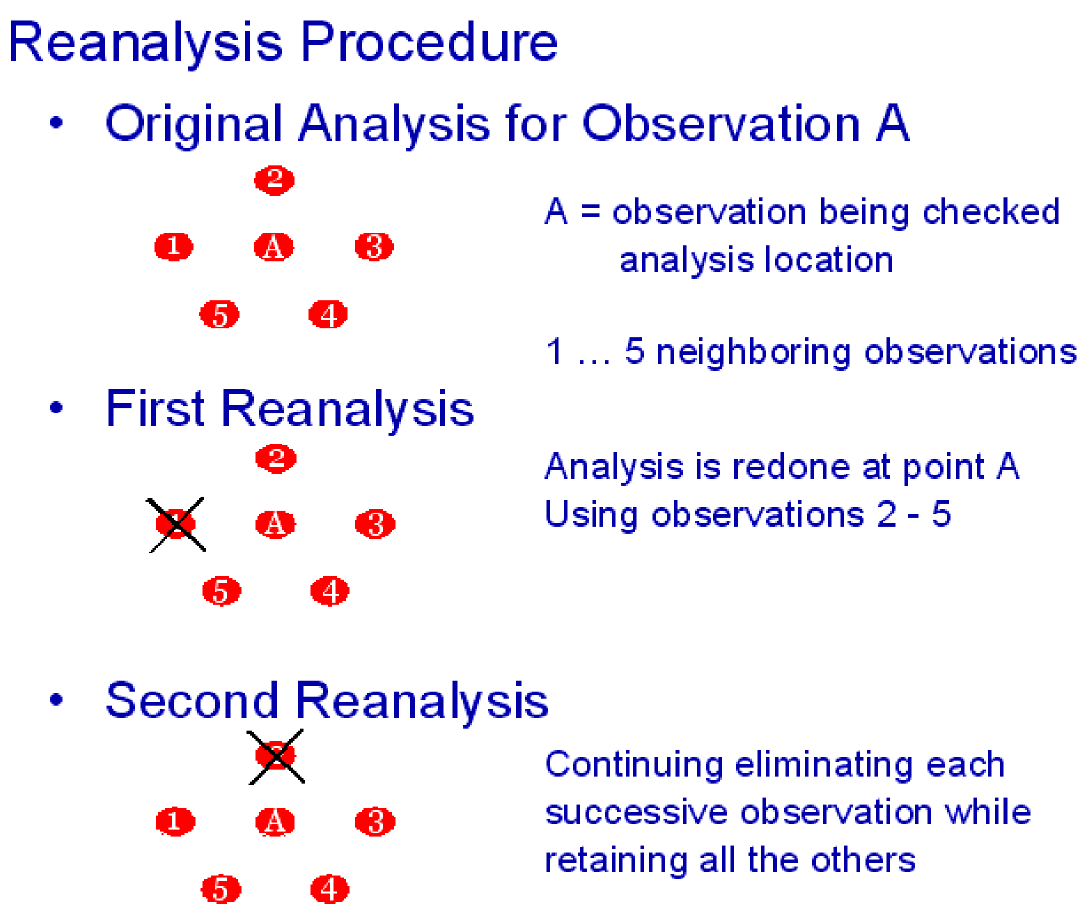

Figure 3. Graphic illustration of reanalysis procedure used in the MADIS spatial consistency check to determine if the target observation is bad or if one of the observations used in the QC analysis is bad. The reanalysis procedure is implemented only if the difference between the target observation and the analysis is greater than a specified threshold.

MADIS also provides single-character "data descriptors" for each observation, which give an overall opinion of the quality of the observation by combining the information from the various QC checks. Algorithms used to compute the data descriptor are a function of the types of QC checks applied to the observation, and the sophistication of those checks.

In addition, MADIS provides the capability for human interaction with, and override of, the automated QC results by maintaining "reject" and "accept" lists for each dataset. Observations placed on the reject list are always flagged as "bad" within the MADIS QC database, while observations placed on the accept list are flagged as "good."

Quality Control Monitoring for Surface Datasets

To assist in subjective overrides of the automated QC results, and to allow for monitoring of surface observations, MADIS keeps station monitoring statistics on the frequency and magnitude of observational errors encountered at each station location, and also provides for the visualization of the observation QC and station monitoring information through a graphical-user-interface (GUI) called the Quality Control and Monitoring System (QCMS) Browser.

The monitoring statistics are computed on an hourly, daily, weekly, and monthly basis for individual NOAA and non-NOAA surface networks (e.g. ASOS, COOP-M, MesoWest, etc.) and have proven to be very useful in locating persistent observational biases and hardware failures. Text versions of the statistics are made available to data providers as part of the MADIS data distribution and web services.

The Browser is an AWIPS-based display application that serves as the user interface to the MADIS surface QC database by implementing an interactive text and graphics display system designed to provide visualization of the QC information, and also to provider easy access to the MADIS subjective intervention capabilities. The Browser is configured at FSL for use in monitoring NOAA surface networks, as well as monitoring non-NOAA mesonets.

Users of the QCMS Browser can select all or portions of the QC information provided in the MADIS database and display the information as plan view and/or time series plots, or can choose to display the information in tabular form on the user interface. The Browser provides easy access to the MADIS database and QC information for the purposes of:

1) monitoring station performance;
2) locating persistent biases or failures in surface observations; and
3) evaluating observation/QC accuracy.

The Browser also interacts with the MADIS subjective intervention capabilities to allow individual variables at select stations to easily be either removed from or placed onto the reject or accept lists. MADIS automated QC and station monitoring procedures are not affected by subjective intervention lists, with the sole exception that observations on the reject list will be labeled as "suspect" and not used to check the spatial consistency of neighboring observations. This allows for the continued monitoring of the stations contained in the lists. For example, a station with wind observations that fail the QC checks a large percentage of the time may be added to the reject list. However, once the observation failure rate at the station falls back to near zero (possibly due to an anemometer that has been repaired), the station will likely be deleted from the list.

Current resources at FSL do not allow for continuous human monitoring of surface meteorological observations. Instead the subjective intervention lists are maintained in response to data issues raised by users or noticed by the MADIS developers. In any case, the QCMS Browser has been found to be a useful tool for monitoring surface observations and QC, and should be considered for use in NOAA operational monitoring centers. Plans are also in the works to install the Browser at each NWS forecast office as part of their operational AWIPS systems.

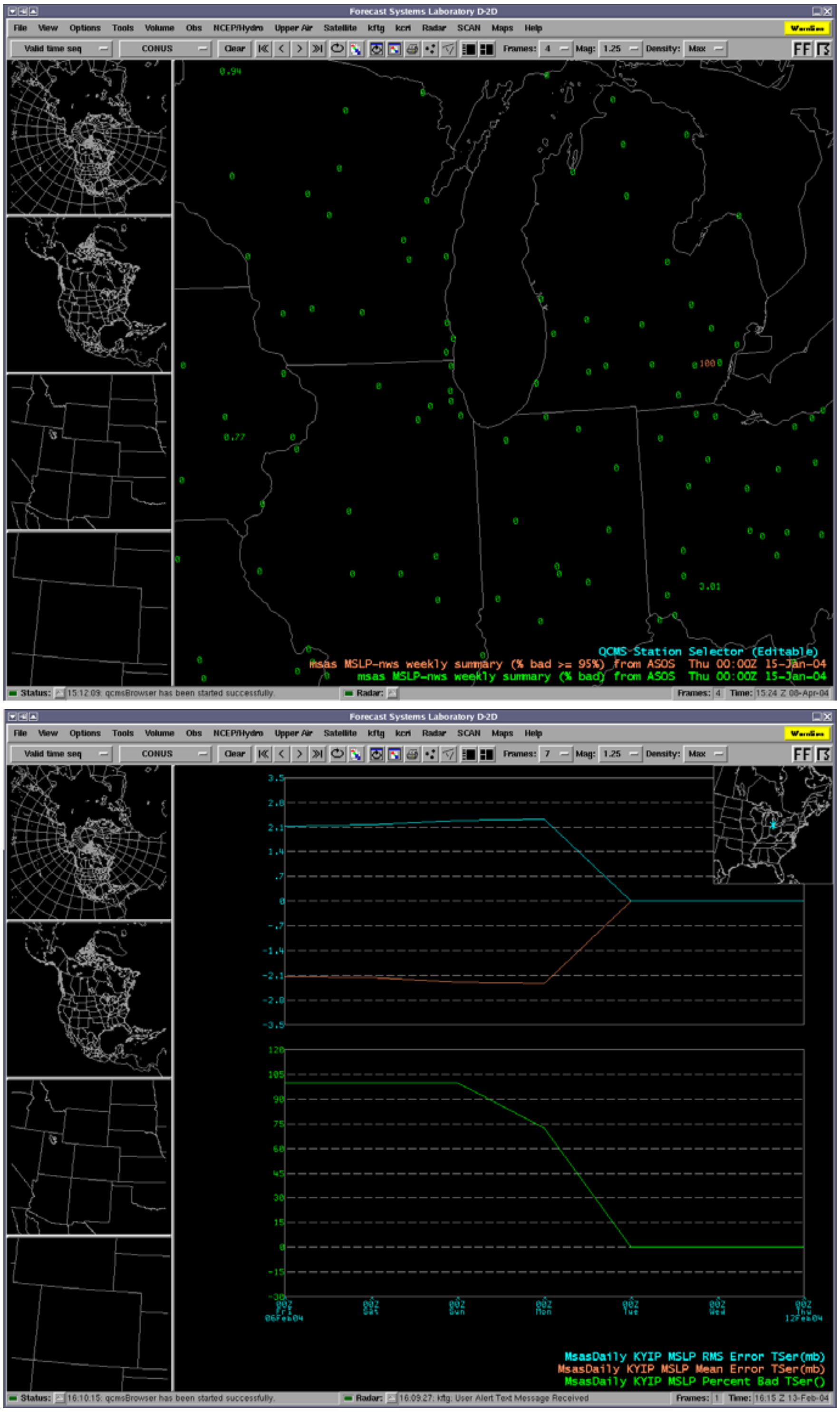

Figure 4. QCMS Browser-produced AWIPS displays detailing the detection and correction of a persistent bias in the sea-level pressure observations reported by an ASOS station in Ypsilanti, MI. a) top, Plan view display of the weekly percentage of QC failures indicating a 100% failure for the Ypsilanti, MI, sea-level pressure observations before NWS personnel were alerted to the problem. Percentage failures above 95% are highlighted in red. b) below, Time series display of the daily RMS (blue) and mean (red) errors for the observations, and the percentage failure (green) over the days both before and after NWS personnel corrected the problem. The QC information before the correction indicated persistent RMS and mean errors of approximately 2.1 mb and a continuous failure rate near 100%. After the observations were corrected, both the errors and the percentage failure fell to zero.

Output Data Files and Software Support

MADIS output files are compatible with AWIPS and AWIPS-like display systems (e.g. FSL's FX-Connect and FX-Net) and the analysis software provided by the Local Analysis and Prediction System (LAPS) and the Weather Research and Forecasting (WRF) Model 3D-Variational (3DVAR) Data Assimilation System. They have also been used to initialize the Advanced Regional Prediction System (ARPS), MM5 , and Coupled Ocean/ Atmosphere Mesoscale Prediction System (COAMPS™) forecast models.

The output files for all MADIS datasets are available in uniform formats with uniform quality control structures within the data files. Software support is also provided for the datasets through the use of an Application Program Interface (API) that provides users with easy access to the data and QC information. The API allows each user to specify station and observation types, as well as QC choices, and domain and time boundaries.

Many of the implementation details that arise in data ingest programs are automatically performed by the API, greatly simplifying user access to the disparate datasets, and effectively integrating the database by allowing, for example, users to access many different types of surface observations (e.g. ASOS, COOP-M, maritime, and non-NOAA mesonets) through a single interface. Users of the MADIS API, can also, for example, choose to have their wind data automatically rotated to a specified grid projection, and/or choose to have mandatory and significant levels from radiosonde data interleaved, sorted by descending pressure, and corrected for hydrostatic consistency.

For users who do not want to write their own programs to use the API, but would like easy access to the data, utility programs for each dataset are included in the API package. These programs can be used to read station information, observations, and QC information for a single time, and dump these to an output text file.

The operation of each program is controlled by a text parameter file that allows the user to exercise all of the options included in the MADIS API. The programs can be run as needed to access any MADIS data stored on the user's system, or can be run as time-scheduled tasks to get data keyed to the current time.

AWIPS Support

To assist NWS AWIPS users, MADIS provides documentation on how to customize AWIPS systems to ingest and display MADIS datasets. For surface mesonet data, NWS forecast offices can choose to select only those mesonets that exist in their local area, and can optionally specify a latitude/longitude box to filter national-scale mesonets in order to reduce the data volume. Instructions are provided on how to bring the data in through the AWIPS local data ingest subsystem (LDAD) along with the necessary metadata files, and any necessary scripts and preprocessors. The station tables are updated weekly.

Also, where applicable (such as for GPSMET integrated precipitable water vapor), the documentation will explain how to customize AWIPS to display additional variables that are not part of the baseline AWIPS system.

In addition to the LDAD data, several of the MADIS datasets are supplementary to what's currently available to AWIPS via NOAAPORT. Customization packages (including instructions and the necessary files to be used in the AWIPS localization tasks) are provided.

Metadata Updates

Station tables are updated for all of the datasets as needed, including weekly automated updates for the mesonet data. In addition to the station table maintenance, additional metadata such as site photos, are collected when available.

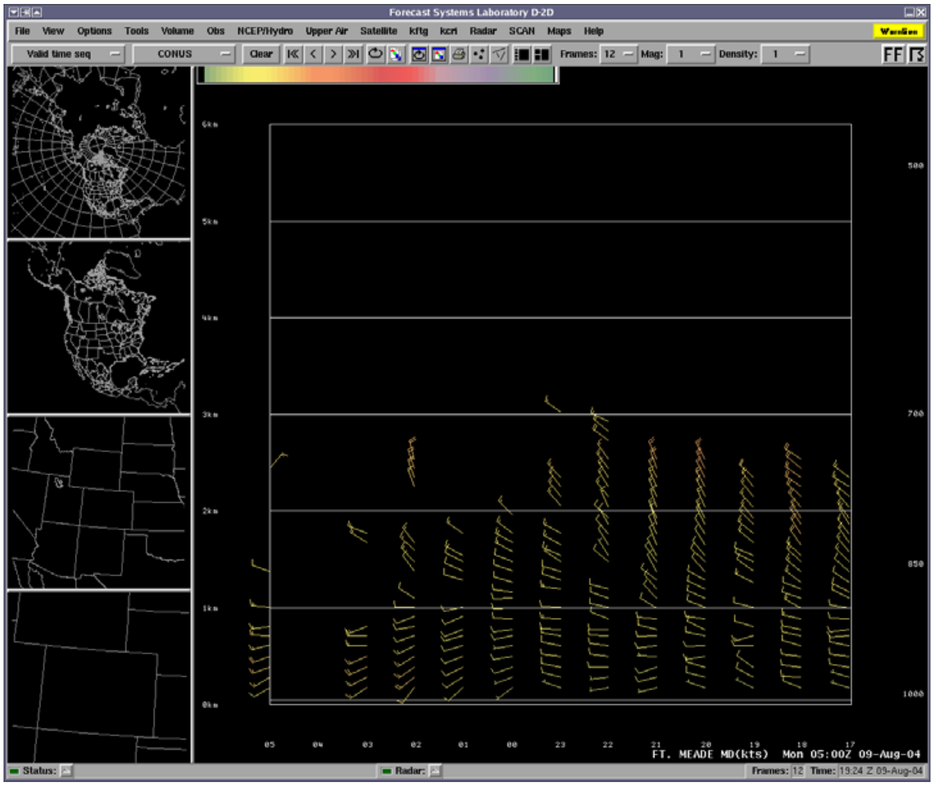

Figure 5. AWIPS Cooperating Agency Profiler (CAP) time-height plot for the Ft. Meade profiler provided by the Maryland Department of the Environment. MADIS supports NWS AWIPS users by providing both real-time access to CAP data and instructions for how to ingest and display the data on AWIPS.

Figure 6. Photograph indicating the northern view from the Alton, IA station in the Iowa Department of Transportation mesonet. Photographs of the northern, eastern, southern, and western views for each station in the mesonet were provided by the Iowa Department of Transportation and are kept on the MADIS web server for reference by users of the mesonet data.

User Support

MADIS developers and support staff are available via telephone and/or e-mail to assist users of the MADIS datasets and API. FSL staff, for example, frequently assist NWS forecast office staff with setting up AWIPS ingest and display of MADIS datasets.

FSL Central Facility Support

The MADIS development team works closely with FSL Information and Technology Services (ITS) staff to operate and monitor MADIS ingest, processing, and distribution functions. ITS consists of three functional groups:

- the Data Systems Group which acquires and processes meteorological datasets;
- the High-Performance Computing Group which operates and maintains one of the world's fastest supercomputers; and
- the Systems Support Group which provides real-time computer operator support for the FSL Central Facility.

In addition to the functional groups, ITS staff also provide the network management, computer systems administration, and computer security services required by the MADIS project.

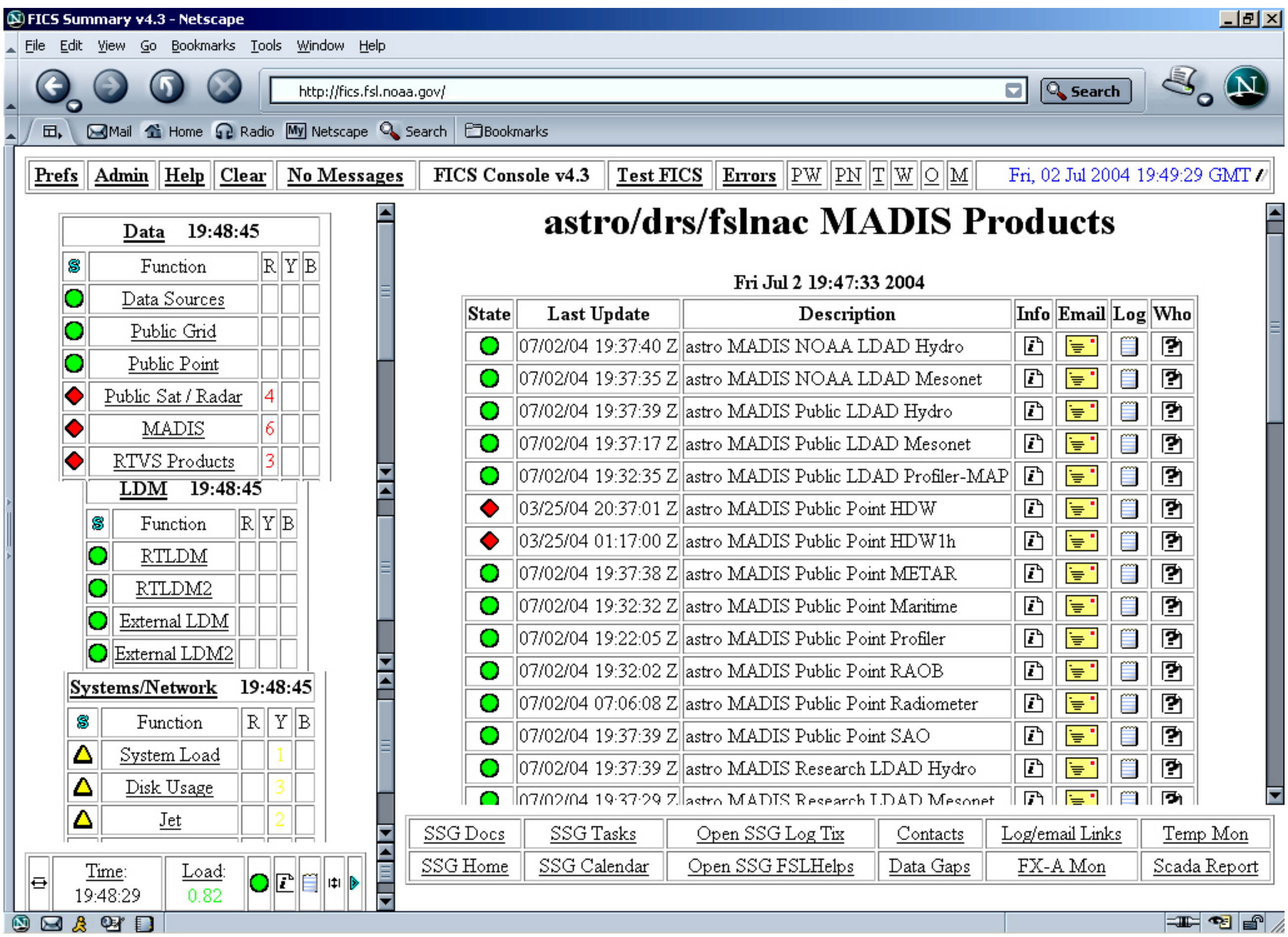

Figure 7. Web-based monitoring system used by the FSL operators to assess the availability of the MADIS datasets. Green "state" lights indicate that datasets are available; red indicate missing data. In this case, GOES High-Density Winds (HDW) were missing due to a server failure at NESDIS.

Processing Environment

The MADIS data processing is performed on a system of 21 computers using Intel processors and the Linux operating system, many of which are configured into "high-availability" (HA) pairs. The HA pairs are used to increase reliability by providing redundant hot backups for the key processing components. A failure of the primary computer will be immediately detected by the secondary computer, which will then automatically start distributing its outputs to the downstream computers.

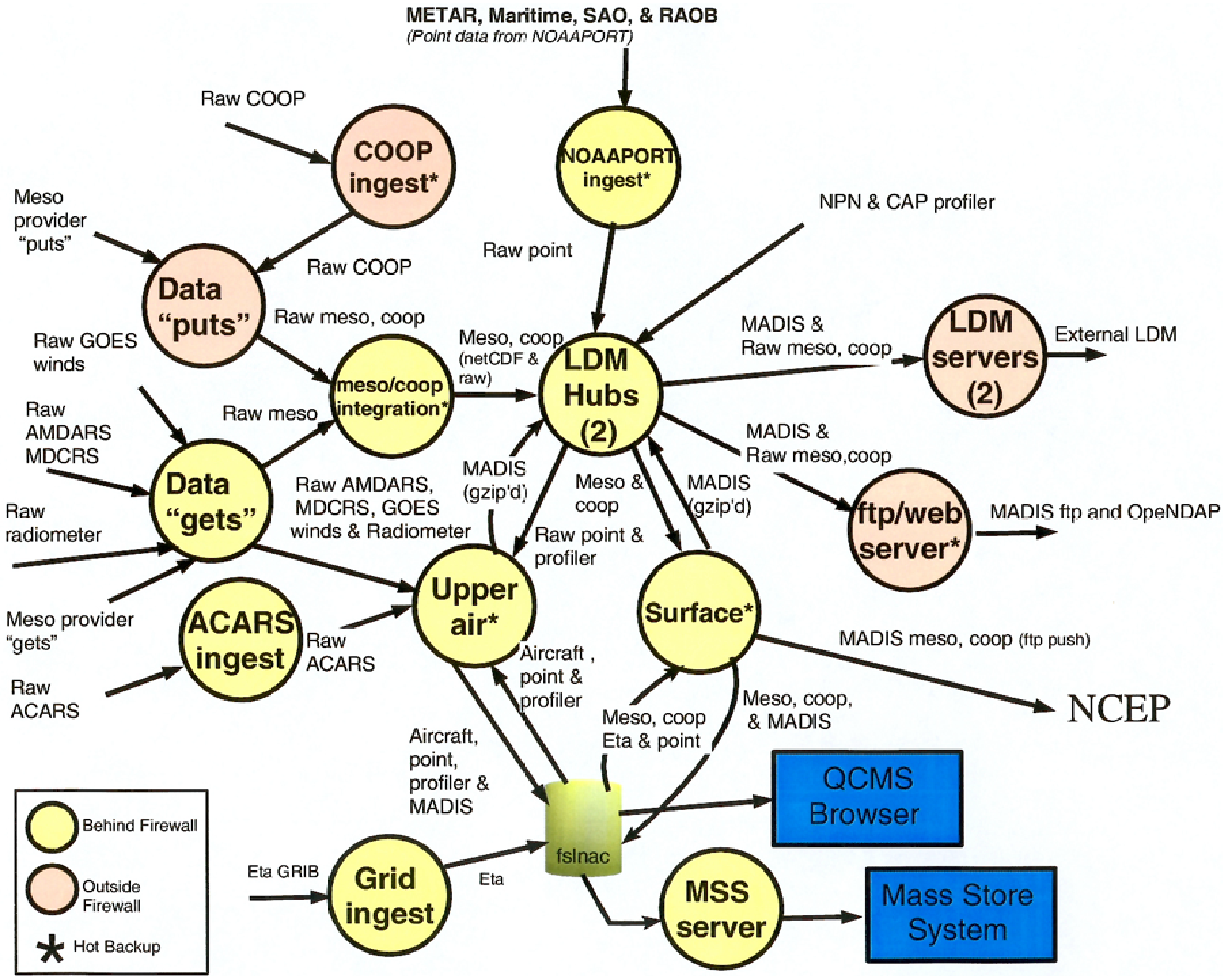

Figure 8. Schematic depiction of the MADIS processing environment. FSL is currently in the process of installing "high-availability" pairs for each MADIS processor to help increase the reliability of the MADIS datasets.

Distribution

MADIS subscribers can receive real-time data or obtain access to on-line storage of saved real-time data by requesting an ftp or OPeNDAP (OPen source project for Network Data Access Protocol) account. For real-time data, LDM (Local Data Manager) access is also available.

Since some of the data are proprietary, different distribution categories have been set up to handle restricting these datasets, which include some of the mesonets, and the automated aircraft data. In general, no restrictions apply to government agencies supporting forecasting operations, or to research and educational organizations or institutions.

To accommodate the different distribution categories, all distribution mechanisms include authentication, with different levels of authentication required for the more restricted datasets (e.g., the Internet address of the user's computer is validated in addition to requiring a password-based ftp account).

Additional services are provided so that all MADIS data files are directly compatible with AWIPS displays and applications without the use of the MADIS API software, and without the need to change AWIPS software to handle new datasets.

Data Storage

In addition to the real-time data available from the MADIS servers, on-line access is available to saved real-time data going back, for most datasets, to July 1, 2001. These data can be accessed either via ftp or by OPeNDAP clients. The MADIS API package includes OPeNDAP clients that can be used to access each dataset. To provide a backup in case of disk failure on the servers, the FSL Central Facility mass store system also contains a complete set of MADIS data files.

Data Outage Monitoring and E-Mail Notifications

Real-time monitoring procedures have been established for each MADIS dataset, that report on the status of both input and output datastreams with a combination of automated and human operator procedures. The status of each individual mesonet input datastream, for example, is automatically monitored for data outages. After a sustained outage of six hours, automated e-mail messages are sent to personnel designated by the data provider. Output datastream outages and computer failures are tracked by ITS operators via a trouble ticket reporting system.

Operators also maintain logs of data outages for internal communication and tracking purposes, and take corrective action in response to failed real-time systems based on documentation created for each system or subsystem. The operators are also responsible for maintaining the troubleshooting documentation and other pertinent informational and training materials.

User Response

MADIS currently supports hundreds of users, including the majority of NWS forecast offices, NCEP, NCDC, NASA, and many other federal, state, academic, and commercial organizations. In addition to the individual observation and QC information provided to general users, organizations contributing observations to the MADIS integrated surface observation database are provided with access to the QC statistics as previously described. Here are some comments on the value of the MADIS data and QC:

1. From Ron Sznaider, Vice President for Business Development at Meteorlogix Corporation:

> **"About a year ago, Meteorlogix became aware of the MADIS project. We found friendly, knowledgeable, and professional researchers who had built a very robust system designed to collect and merge weather information from a wide variety of disparate sources. Furthermore, they had applied state-of-the-art quality control to this expanded set of weather observations, and made access to the data available more frequently. It seemed too good to be true! In fact, through the**

continued efforts of Patty Miller and Mike Barth, the MADIS data has exceeded all expectations and has made a substantial difference in our efforts to provide decision makers throughout the country with more timely, accurate, high-quality weather information...The MADIS Project is an exemplary model for a successful public-private partnership that ultimately benefits businesses as well as providing potential widespread societal benefits."

2. From the U.S. Weather Research Program's Advanced Operational Nowcasting System Working Group (who acknowledged that the establishment of a national surface station mesonetwork was their highest priority for new or enhanced observing systems):

   "The NOAA/FSL 13,000 surface stations that are utilized for the Meteorological Assimilation Data Ingest System (MADIS) should form the starting point for a [more extensive] national network. NOAA should take the lead in quality assurance and setting standards for surface mesonet data. It should be noted that FSL is already doing this. NOAA should formalize this FSL activity and fully support the effort..."

3. From Jeffrey Medlin, Science and Operations Officer, NWS Mobile, Alabama Weather Forecast Office:

   "Next to the WSR-88D, AWIPS, and WES, this [MADIS] is one of the greatest gifts associated with modern day weather forecasting."

4. From Dr. Pat Welsh, Science and Operations Officer, NWS Jacksonville, Florida Weather Forecast Office:

   "What you [the MADIS team] have done is technology transfer at its finest... You have filled a large need for mesoscale data while encouraging a rare and truly collaborative effort among the public, private, and academic sectors...The FSL MADIS Project has been crucially important in Florida where mesoscale weather is the rule, but until MADIS, only synoptic scale surface data was available. Several commercial firms with data in Florida provide data to MADIS that is now available to the NWS offices... [Additionally] Florida State University, Embry-Riddle Aeronautical University, the University of North Florida and the University of South Florida all have data projects that include FSL MADIS."

5. From Dr. Jan F. Dutton, Director of Weather Services, AWS Convergence Technologies, Inc.

   "The MADIS system has been very valuable in AWS's efforts to not only monitor the real-time data quality of its network of weather stations but also to improve the quality the data. AWS operates a nationwide network of approximately 6000 weather stations located primarily in urban areas. Prior to inclusion [in MADIS], we had a range of relatively simple data quality control procedures but nothing matching the sophistication or scope of the MADIS system. The AWS data became input to the MADIS system in 2002 and since then we have used the resulting quality control information to great benefit in the operation of our weather network... Our ability to improve the AWS network results in better products for our end customers across the board whether it be for our 30 Million WeatherBug users, our 100 broadcast television partners reaching 80 million people per day, short term electricity traders who make trading decisions based on AWS real-time data, and emergency managers and first responders who use the data during emergencies."

Short-Term Plans

FSL plans to support the Great Lakes Flight Experiment, scheduled for fall 2004, through its MADIS system by ingesting, integrating, quality controlling, and distributing Tropospheric Airborne Meteorological Data Reporting System (TAMDAR) sensor data to numerous experiment participants, including NCEP and NWS forecast offices. Support for additional datasets, such Polar orbiting satellite observations and observations from up to 25 new surface mesonets is also planned for the 2004 - 2005 time frame, as well as support for new distribution capabilities, such as XML-based output files. A "data recovery system" to process surface observations that were not originally available in real time due to communications outages is also under development.

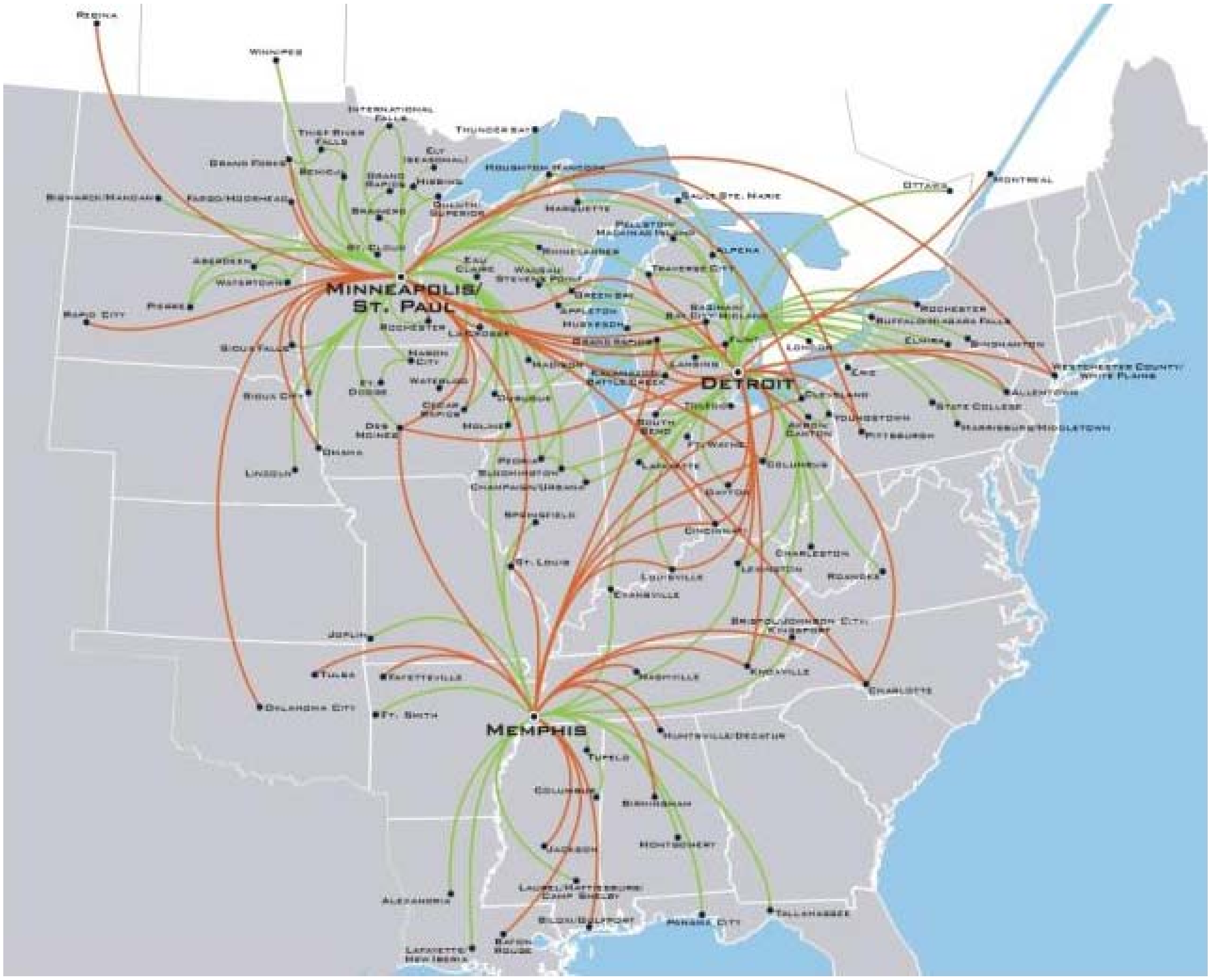

Figure 9. Mesaba Airline SAAB-340 Flight routes planned for the Great Lakes Flight Experiment. Mesaba plans to install TAMDAR units on 64 SAAB-340 aircraft (routes in green) to support the Experiment. Raw data from the units will be ingested, integrated, quality controlled, and distributed by the MADIS system at FSL.