Sandy Supplemental Grant Recipient Final Report

Quality Control and Impact Assessment of Aircraft Observations in the GDAS/GFS

Award Number: NA13NWS4830022

The National Oceanic and Atmospheric Administration
National Environmental Satellite Data and Information Service
Center for SaTellite Applications and Research (STAR)

For the Period
1 October 2013 – 30 June 2015

On behalf of
The Cooperative Institute for Meteorological Satellite Studies (CIMSS)
Space Science and Engineering Center (SSEC)
at the University of Wisconsin-Madison
1225 West Dayton Street
Madison, Wisconsin 53706
608/262-0544

David Santek
Principal Investigator
dave.santek@ssec.wisc.edu

Ralph Peterson
Co-Investigator
ralph.petersen@ssec.wisc.edu

Sandy Supplemental Grant Recipient Final Report
Quality Control and Impact Assessment of Aircraft Observations in the GDAS/GFS

Table of Contents

Sandy Supplemental Grant Recipient Final Report
Quality Control and Impact Assessment of Aircraft Observations in the GDAS/GFS

I. Introduction

Cooperative Institute Description

The Cooperative Institute for Meteorological Satellite Studies (CIMSS) is a collaborative relationship between the National Oceanic and Atmospheric Administration (NOAA) and the University of Wisconsin-Madison (UW-Madison). This partnership has and continues to provide outstanding benefits to the atmospheric science community and to the nation through improved use of remote sensing measurements for weather forecasting, climate analysis and monitoring environmental conditions. Under the auspices of CIMSS, scientists from NOAA/NESDIS and the UW-Madison Space Science and Engineering Center (SSEC) have a formal basis for ongoing collaborative research efforts. CIMSS scientists work closely with the NOAA/NESDIS Advanced Satellite Product Branch (ASPB) stationed at the UW-Madison campus. This collaboration includes a scientist from the National Climate Data Center (NCDC), who joined the NOAA NESDIS employees stationed at CIMSS.

CIMSS conducts a broad array of research and education activities, many of which are projects funded through this Cooperative Agreement with NOAA. This Cooperative Agreement identifies four CIMSS themes:

1. Satellite Meteorology Research and Applications, to support weather analysis and forecasting through participation in NESDIS product assurance and risk reduction programs and the associated transitioning of research progress into NOAA operations,
2. Satellite Sensors and Techniques, to conduct instrument trade studies and sensor performance analysis supporting NOAA's future satellite needs as well as assisting in the long term calibration and validation of remote sensing data and derived products,
3. Environmental Models and Data Assimilation, to work with the Joint Center for Satellite Data Assimilation (JCSDA) on improving satellite data assimilation techniques in operational weather forecast models, and
4. Outreach and Education, to engage the workforce of the future in understanding and using environmental satellite observations for the benefit of an informed society.

CI Management and Organizational Structure

CIMSS resides as an integral part of the Space Science and Engineering Center (SSEC). CIMSS is led by its Director, Dr. Steven Ackerman, who is also a faculty member within the UW-Madison Department of Atmospheric and Oceanic Sciences. Executive Director Wayne Feltz provides day-to-day oversight of the CIMSS staff, science programs, and facilities. The education and outreach activities at CIMSS are coordinated by Senior Outreach Specialist Margaret Mooney. The individual science projects are led by University Principal Investigators (PIs) in conjunction with a strong and diverse support staff who provide additional expertise to the research programs. CIMSS is advised by a Board of Directors and a Science Advisory Council.

The CIMSS administrative home is within the Space Science and Engineering Center (SSEC), a research and development center within the UW–Madison's Office of the Vice Chancellor of Research. The independent CIMSS 5-year review panel for administration wrote that they were "…impressed by the people, systems and processes in place." The SSEC mission focuses on geophysical research and technology to enhance understanding of the Earth, other planets in the Solar System, and the cosmos. To conduct its science mission on the UW-Madison campus, SSEC has developed a strong administrative and programmatic infrastructure. This infrastructure serves all SSEC/CIMSS staff.

The CIMSS mission includes three goals:

- Foster collaborative research among NOAA, NASA, and the University in those aspects of atmospheric and earth system science that exploit the use of satellite technology;
- Serve as a center at which scientists and engineers working on problems of mutual interest can focus on satellite-related research in atmospheric and earth system science;
- Stimulate the training of scientists and engineers in the disciplines involved in atmospheric and earth sciences.

Executive Summary of CI Banner Research Activities

CIMSS is a collaboration between NOAA and UW–Madison that has increased the effectiveness of research and the quality of education in the environmental sciences. In a *Space Policy* article in 1986, William Bishop, former acting Director of NESDIS, noted, "Remote sensing from space can only thrive as a series of partnerships." He used CIMSS as a positive working example of the government-academia partnership, noting "The Institute pioneered the computation of wind speeds at cloud heights by tracking cloud features from image to image. These are now a stable product provided from the satellites to the global models at the National Meteorological Center." CIMSS continues to be a leader in the measurement of winds from satellite observations and leads the way in many other research endeavors as outlined above. There is great value to NOAA and UW-Madison in this long-term collaboration known as CIMSS.

II. Funded Project

Award Number: NA13NWS4830022

Project Title: Quality Control and Impact Assessment of Aircraft Observations in the GDAS/GFS

PI: Dr. David Santek

NOAA Sponsor: Andrew Collard and Stephen Lord

NOAA Sponsoring Organization: NOAA NWS/EMC

Reporting Period: 1 October 2013 – 30 June 2015

Description of Task I Activities

Primarily activity involves quarter reporting.

NOAA Strategic Goal(s)
NOAA Mission Goals

1. Climate Adaptation and Mitigation: An informed society anticipating and responding to climate and its impacts
2. Weather-Ready Nation: Society is prepared for and responds to weather-related events

NOAA Strategic Plan-Mission Goals

1. Serve society's needs for weather and water
2. Understand climate variability and change to enhance society's ability to plan and respond
3. Provide critical support for the NOAA mission

III. Research Progress

A particular interest in this study is to investigate the relationship between radiosonde moisture observations and AMDAR moisture observations during assimilation, when both are available in the same location. It is desirable to know, for example, how radiosonde and AMDAR moisture observations compare to the model background derived from the 6-hour forecast – a closer fit of observations to the 6 hour forecast following assimilation can indicate that the observations are high quality and improve the initial (analysis) state. Likewise, impact can be identified cross-observation; for example, an improved fit of *radiosonde* observations to the 6-hour forecast as a result of assimilating *AMDAR* observations can indicate better model performance, as this can be equivalently expressed as a closer fit of the 6-hour forecast to trusted observations.

Mean profiles of ob-minus-background (OMB) values of specific humidity are produced for radiosonde observations without AMDAR moisture assimilation, radiosonde observations with AMDAR moisture assimilation, and for AMDAR observations when they are assimilated (Fig. 1). Profiles are produced at each radiosonde launch site, averaging radiosonde OMB scores within 25 pressure-slabs between the surface and 300 hPa, which is the highest level where moisture observations are assimilated. AMDAR OMB scores are likewise averaged within these pressure-slabs, for any AMDAR moisture observation within 1 hour and 0.5 degrees of the radiosonde launch. These profiles are then averaged across all radiosonde launch sites in the continental US.

Profiles indicate that the 6-hour forecast fits to radiosonde observations more closely when AMDAR observations are assimilated during the April – May 2014 experiment (Fig. 1a), while no clear change is observed in the December 2014 – January 2015 experiment (Fig. 1b). This indicates increased model performance in the warm-season experiment. Likewise, AMDAR moisture observations fit closer to the 6-hr forecast than radiosonde observations at essentially all levels in the warm-season experiment, while this relationship does not exist in the cold-season experiment. This indicates high-quality of AMDAR moisture observations, even in comparison to radiosonde observations; in the cold-season experiment, AMDAR and radiosonde observations appear to have largely indistinguishable quality characteristics by this metric, except for perhaps the surface and near-surface levels where radiosondes are biased moist and

AMDAR observations are not. In the warm-season, the 6-hour forecast fit-to-radiosondes is improved to statistical significance within the lower troposphere just above the surface.

Precipitation skill and bias, and forecast fit-to-GPS observations

Precipitation forecast skill is measured by the Equitable Threat Score (ETS) and Bias Score for precipitation over the continental United States, binned by precipitation thresholds in amounts of millimeters per 24 hours. The mean ETS is improved to statistical significance for 12-36 hour forecasts of light precipitation in the warm-season experiment, with improvement in ETS for precipitation or below 5 mm/day (Fig. 2a). Bias is slightly reduced for these categories as well. There is statistically-significant ETS degradation for only the 10 mm/day category of the 60-84 hour forecast (not shown), while the ETS and bias are not significantly changed for any other category at any forecast lead-time. The cold-season experiment expresses no significant improvement in ETS or bias for any category or forecast lead-time (Fig. 2b), with the exception of a degradation in bias for very high precipitation (50-75 mm/day) in 60-84 hour forecasts (not shown); these categories have very few observations from which to derive statistics, and are dominated by a single event, making the statistics less reliable.

These statistics imply an improvement to short-range precipitation forecasting is achieved by assimilation of AMDAR moisture observations. An additional measure of forecast skill can be observed by computing the forecast fit-to-observations using GPS total-column precipitable water. Total-column precipitable water from the GFS forecast was compared to observations derived from Global Positioning Satellite (GPS) soundings. For each six-hourly forecast period from the analysis time to the 72-hour forecast, forecast fields are interpolated to a database of GPS observations and the mean error is computed (Fig. 3). The warm-season experiment is improved, with statistically-significant differences observed for the first 0-18 hours, after which the impact tends toward zero and is statistically insignificant. The cold-season experiment only demonstrates statistically-significant impact at analysis-time, after which the impact tends toward zero and is statistically insignificant.

AMDAR/Radiosonde redundancy experiment

i) Vertical and temporal coverage of radiosonde launch sites by aircraft observations

In this experiment, the value of radiosonde observations in regions best observed by aircraft observations was tested through a data denial experiment. The coverage of a US radiosonde site by aircraft observations was determined at each six-hourly analysis period by collecting aircraft moisture observations available within one degree in latitude/longitude space and one hour in time of the radiosonde launch. These aircraft observations are defined as 'collocated' with the radiosonde for the purposes of defining coverage of the site. The vertical profile of the radiosonde launch site is divided into 25 equally-spaced pressure slabs between the surface and 300 hPa (which is the lowest allowable pressure for assimilation of radiosonde and aircraft moisture observations). The vertical coverage of the site by aircraft observations is defined as the percentage of these slabs that contain at least one aircraft moisture observation. Likewise, the temporal coverage of the site by aircraft observations is defined as the percentage of analysis-periods where at least one collocated aircraft moisture observation is available, such that an aircraft observation profile can be produced. The total coverage score for a radiosonde launch site is the product of these two coverage statistics, varying between zero and one:

$$C_{total} = C_{Vertical} * C_{Temporal} \qquad (1)$$

Radiosonde launch sites are ranked by coverage, and the most well covered sites are used for the data-denial experiment. Table 1 shows the coverage statistics for three spatial/temporal thresholds of coverage. The ten sites chosen for the experiment include sites that appear in the top-10 for at least two thresholds, except for Las Vegas, NV, which is in the top-11 for two thresholds and in the top-3 for the strictest threshold. These sites are spread across the continental US, which allows for the assumption that they impact the forecast largely independent of one another.

In the data-denial experiment, the GDAS was run on a six-hourly cycle from 01 April – 19 May 2014, following a spin-up period of one week. The aircraft moisture observations were assimilated, but the entire radiosonde (wind, temperature, and moisture observations) at each of the ten chosen sites was excluded. The purpose of this experiment is to determine if the forecast is significantly impacted by the missing radiosonde data, given that there is an abundance of high-quality aircraft observations present at or near the radiosonde's location.

ii) Precipitation equitable threat score (ETS) and bias score

The change in ETS and bias scores in the data-denial experiment is similar in form to the impact from the first assimilation experiment which included both the aircraft moisture observations and the ten selected radiosondes (Fig. 4). The 12-36 hour ETS score is improved to statistical significance for low precipitation amounts 0.2-2.0 mm/day) and bias is improved for precipitation amounts less than 10 mm/day. When compared to a portion of the first assimilation experiment over the same time period, there is a notably larger positive impact on both ETS and bias scores when the ten selected radiosondes have been removed, achieving statistical significance over these same precipitation thresholds (Fig. 5). Impacts on longer-range forecasts do not reach statistical significance (not shown). This result implies that the ten selected radiosondes may be reducing precipitation skill rather than improving it; as shown in Fig. 1, the 6-hour forecast fit-to-observations is better for aircraft moisture observations than for radiosondes on average. Thus it is possible that removing radiosondes in regions of dense aircraft observational coverage could yield a positive impact.

iii) Forecast fit-to-observations: GPS total-column precipitable water

The forecast fit-to-observations was calculated for the data-denial experiment in the same manner that was applied to the assimilation experiment for comparison (Fig. 6). While short-range (0-36 hour) forecasts are improved by the assimilation of aircraft moisture data regardless of whether the selected radiosondes are denied, the improvement is greater when the radiosondes are assimilated. Moreover, when the radiosondes are denied the mid-range forecast (72 hours) degrades to statistical significance.

iv) Reconciliation of results

The results of the forecast fit-to-observation test is in contrast to the short-range precipitation skill scores. The precipitation skill scores suggest that removal of radiosondes (and thus greater reliance on aircraft observations near the denied radiosondes) yields a forecast improvement in the short range. Meanwhile, the forecast fit-to-observations test tells a different story; the forecast system is most improved when both the aircraft and radiosonde observations are assimilated, and the forecast degrades at 72 hrs when the radiosondes are denied.

Reconciliation of these results requires both an examination of the differences in the tests, as well as the role of assimilating aircraft temperature observations versus moisture observations.

Two differences between these tests must be considered. First, GPS observational coverage is more comprehensive spatially and temporally than precipitation data, which allows every forecast to be tested for accuracy at more locations than with the more sparse precipitation data. For example, the 12-36 hour forecast period over which the precipitation skill scores are presented is binned by precipitation amount, with the largest number of precipitation observations in the lowest-value bin. For the warm-season experiment, there are 42,057 data points used to determine ETS and bias skill scores when comparing the assimilation experiment and the data-denial experiment. By contrast, in the forecast fit-to-observations test, 69,071 observations were tested over the same forecast period, a 64% increase in available observations, even when compared to the *largest* precipitation bin. Second, the forecast fit-to-observations shows statistically significant degradation in the medium-range, while statistically significant impact on precipitation scores only extends to the 12-36 hour forecast. For these reasons, one could argue that the forecast fit-to-observations test is more comprehensive.

Another consideration is in the relative impact of moisture versus temperature observations from aircraft. As shown previously, aircraft moisture observations near radiosondes exhibit a lower OMB than radiosonde observations, which implies that aircraft moisture observations may be of higher quality. However, temperature observations from aircraft have been shown to suffer biases that can vary by individual aircraft as well as by whether the aircraft is ascending or descending (e.g. Ballish and Kumar 2008). While efforts to address these biases are currently being investigated (Isaksen et al. 2012, Zhu et al. 2015), NCEP does not currently employ a bias correction mechanism for these observations. It is possible that higher-quality moisture observations from aircraft improve the short-range precipitation forecast, while biased temperature observations from aircraft degrade the mid-range forecast, explaining the results from both the precipitation skill score test and the forecast fit-to-observations test.

v) Obsevation-minus-analysis statistics

As a final test of the impact of the selected radiosondes, the ob-minus-analysis (OMA) statistics of aircraft moisture observations assimilated near the missing radiosonde sites was compared with and without the radiosondes present, and the difference was plotted against the density of aircraft observations present (Fig. 7). While there is no strong correlation (i.e. linear relationship) between these two statistics ($r =$ -0.0144), a relationship becomes clear when the points are plotted on a phase-space. The more aircraft observations nearby (higher values along the abscissa), the less the OMA statistic for aircraft moisture observations is capable of changing when the radiosonde is denied (lower values along the ordinate). Thus the relationship between these two statistics is represented by an upper-bound on the ordinate as a function of the abscissa, which appears to obey an exponential-decay-like form.

vi) Evaluation of radiosonde data-denial

The tests presented do not reach a clear conclusion about the importance of radiosondes in the presence of dense aircraft observational coverage. While precipitation skill can be improved in the short range by their exclusion, forecast fit-to-obs against GPS total-column precipitable water suggests that denying the radiosondes increases error, even to statistically significant degradation against a control that contains no aircraft moisture observations for 72-hour forecasts.

The impact of the missing radiosonde, measured as the change in OMA of the aircraft observations when the radiosonde is excluded, demonstrates a relationship to the number of aircraft observations present; the more aircraft observations present, the smaller the upper-bound on the expected impact of denying the radiosonde. Based on the best-fit curve describing the upper-bound, the expected OMA-impact on a single, lone aircraft moisture observation is 1.47×10^{-3} g/g. To reduce this upper-bound by 50%, roughly 22 aircraft observations need to be present to reduce the impact of the radiosonde. To reduce the upper-bound by another 50%, roughly 43 aircraft observations must be present. Given a threshold maximum allowable impact from a denied radiosonde, a minimum number of aircraft observations must be present.

Since the amount of aircraft observational coverage is highly variable, even for the most well-covered radiosonde sites, permanent exclusion of these sondes in favor of aircraft observations does not seem plausible. However, the opposite case could be considered: "Where would an *additional* radiosonde observation provide the *most* impact, based on aircraft observational coverage?" This scenario occurs during off-time radiosonde launches, which have become part of the adaptive observation network, especially during the Atlantic hurricane season when a significant hurricane threatens to make landfall on the east coast of the United States. Under extreme scenarios radiosondes can be launched at 0600 and 1800 UTC from all operating sites in the continental US, as was the case with the days leading up to landfall of Hurricane Sandy (2012).

In scenarios such as these, the goal may be to deploy a limited number of off-time radiosondes with a goal to maximize the impact on the analysis and the forecast of an extreme weather event. One could then expect that radiosondes deployed where there is an expectation of dense aircraft observations would have less impact than radiosondes deployed where there is an expectation of sparse aircraft observational coverage. The decision to launch an off-time radiosonde at a particular site could be aided by statistics on the aircraft observational coverage at existing radiosonde sites for these times.

Operational implementation

Based on the presented research, the decision was made to implement assimilation of aircraft moisture observations in the GDAS, as part of the next upgrade. Implementation is currently slated for February 2016.

References

Ballish, B. A., and K. Kumar, 2008: Systematic differences in aircraft and radiosonde temperatures. *Bull. Amer. Meteor. Soc.*, **89**, 1689-1708.

Isaksen, L., D. Vasiljevic, D. Dee, and S. Healy, 2012: Bias correction of aircraft data implemented in November 2011. *ECMWF Newsletter*, No. 131, ECMWF, Reading, United Kingdom, 6-6.

Zhu, Y., J. C. Derber, R. J. Purser, B. A. Ballish, and J. Whiting, 2015: Variational correction of aircraft temperature bias in the NCEP's GSI analysis system. *Mon. Wea. Rev.*, **143**, 3774-3803.

Site	$\Delta T = 0.75, \quad \Delta D = 0.25$	$\Delta T = 1.00, \quad \Delta D = 0.50$	$\Delta T = 1.25, \quad \Delta D = 0.75$
Miami, FL	(25th) 0.046	(5th) 0.446	(5th) 0.653
Tampa, FL	(7th) 0.166	(2nd) 0.569	(2nd) 0.861
Atlanta, GA	(21st) 0.076	(8th) 0.376	(7th) 0.517
Fort Worth, TX	(12th) 0.126	(1st) 0.603	(3rd) 0.739
Nashville, TN	(1st) 0.231	(4th) 0.524	(4th) 0.717
Las Vegas, NV	(3rd) 0.213	(11th) 0.318	(11th) 0.410
Sterling, VA	(2nd) 0.222	(3rd) 0.540	(1st) 0.864
Denver, CO	(5th) 0.199	(9th) 0.368	(10th) 0.446
Oakland, CA	(4th) 0.209	(6th) 0.394	(8th) 0.496
Upton, NY	(10th) 0.132	(7th) 0.379	(9th) 0.478

Table 1. Ten radiosonde launch sites chosen for data denial experiment. Rankings of each site by coverage are provided for three thresholds defining collocation of aircraft observations to the radiosonde: (left) observations within 0.75 hours and 0.25 degrees of the site, (middle) observations within one hour and 0.5 degrees of the site, and (right) observations within 1.25 hours and 0.75 degrees of the site. Rankings in the top-10 are highlighted in red.

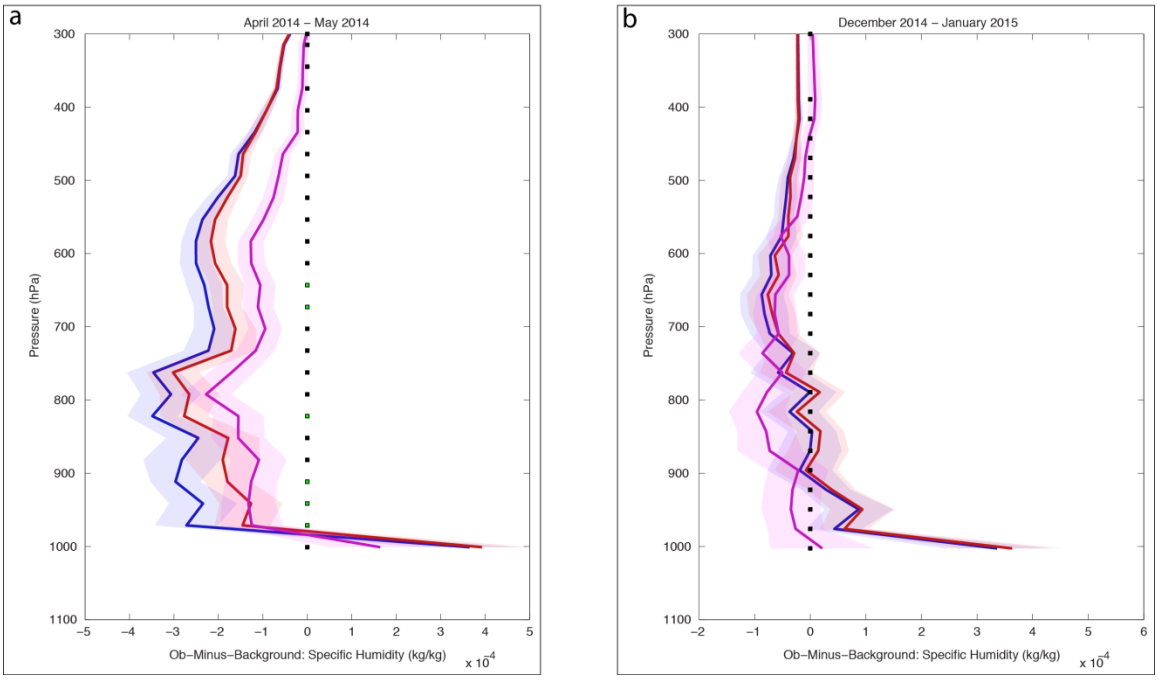

Figure 1. Mean profiles of specific humidity ob-minus-background (OMB) for the warm-season experiment (left) and cold-season experiment (right) at radiosonde launch sites. The blue profile is the mean radiosonde moisture OMB when AMDAR moisture observations are not assimilated. The red profile is the mean radiosonde moisture OMB when AMDAR moisture observations are assimilated. The magenta profile is the mean AMDAR moisture OMB. The shading around each profile represents the 5% and 95% confidence limits around the mean, and pressure-levels where the radiosonde OMB changes to statistical significance are highlighted with green squares along the zero-line.

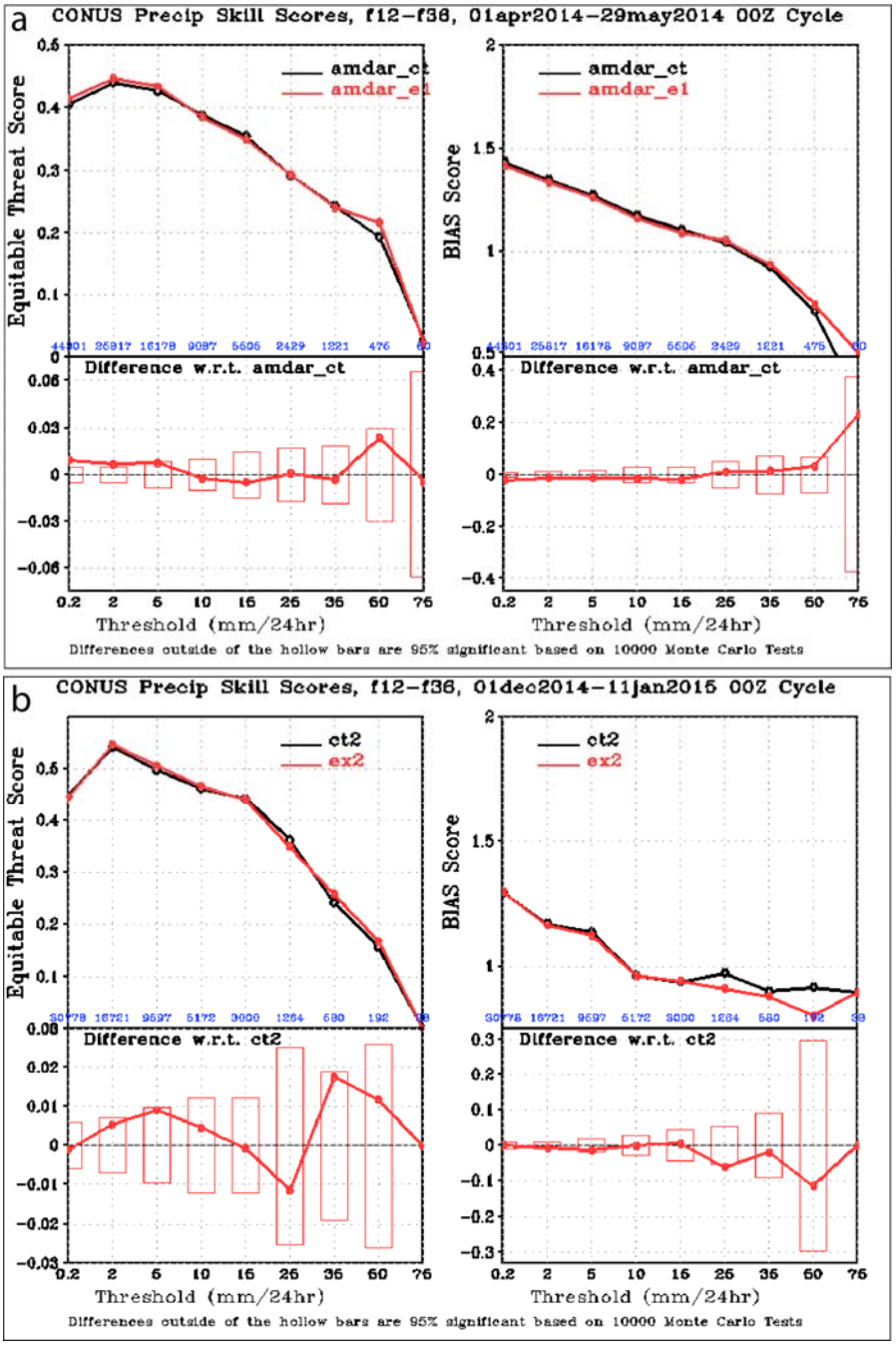

Figure 2. Precipitation skill and bias scores of 12-36 hour forecast over the continental United States for (a) warm-season experiment, and (b) cold-season experiment. The left panel of each plot shows the Equitable Threat Score (ETS) for precipitation binned by precipitation amounts in mm/24 hrs. The right panel of each plot shows the precipitation bias score in the same bins. The black curve is for the control simulation, and the red curve is for the experiment. Below each panel is a plot of the difference (experiment minus control), with bars indicating the minimum value necessary for statistical significance.

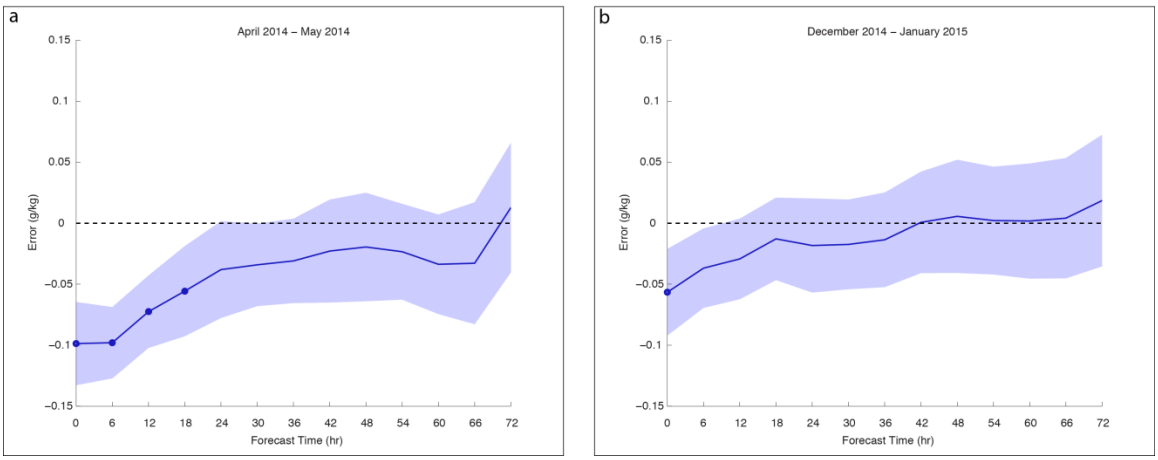

Figure 3. Difference in error between experiment and control, verifying forecast of total column precipitable water against GPS observations. Negative values represent forecast improvement relative to the control forecast. (a) Mean difference in error between experiment and control (exp – ctl) for the warm-season experiment. (b) Same as panel-a, but for the cold-season experiment. The shading around the mean represents the 95% confidence in the mean, and dots along the mean represent times when difference between experiment and control is statistically significant to 95% confidence.

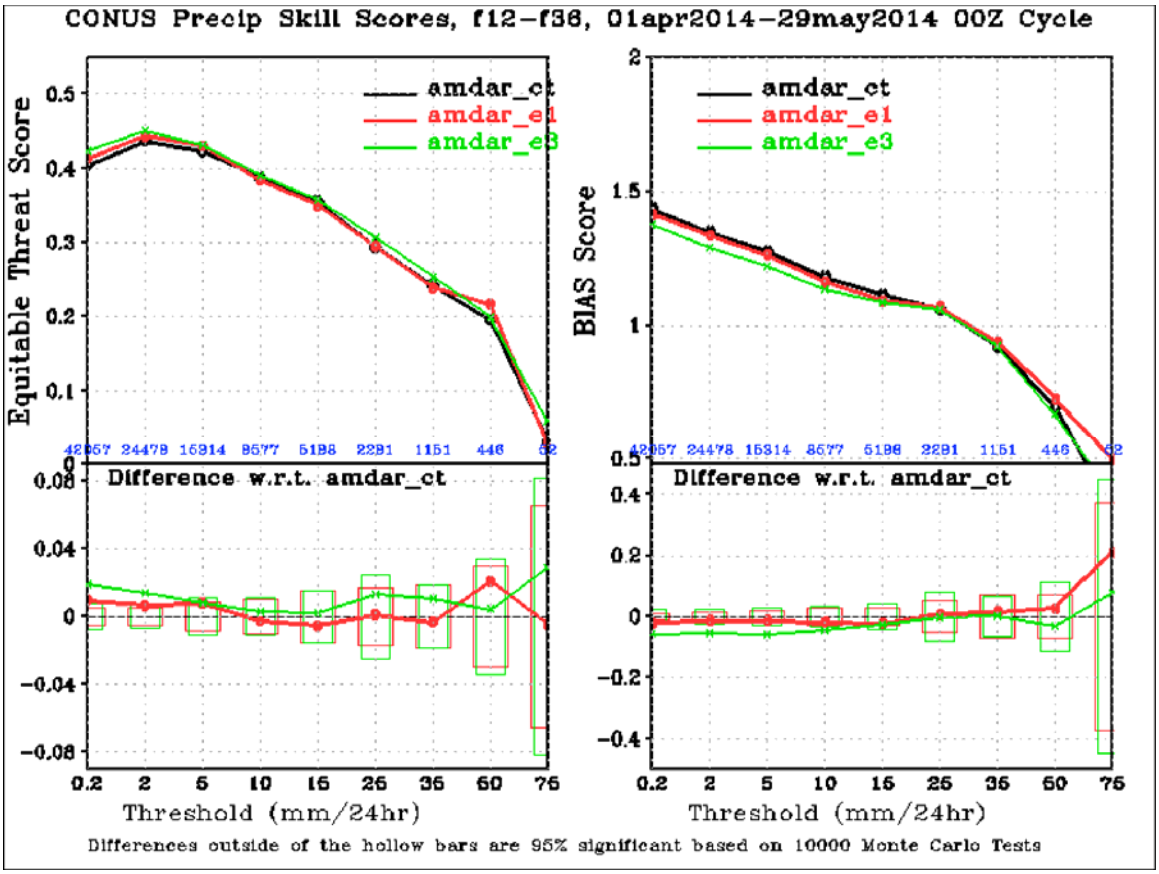

Figure 4. Precipitation skill score of the 12-36 hour forecast over the time-period of the data-denial experiment. (Left) equitable threat score (ETS) and (right) bias score by precipitation threshold, measured in mm/day. The control simulation (radiosondes, no aircraft moisture observations) is plotted in black. The assimilation experiment (radiosondes, aircraft moisture observations) is plotted in red. The data-denial experiment (selected radiosondes removed, aircraft moisture observations) is plotted in green. Below each plot is a plot of the difference with respect to the control. Bars indicate the necessary deviation from the control to achieve statistical significance.

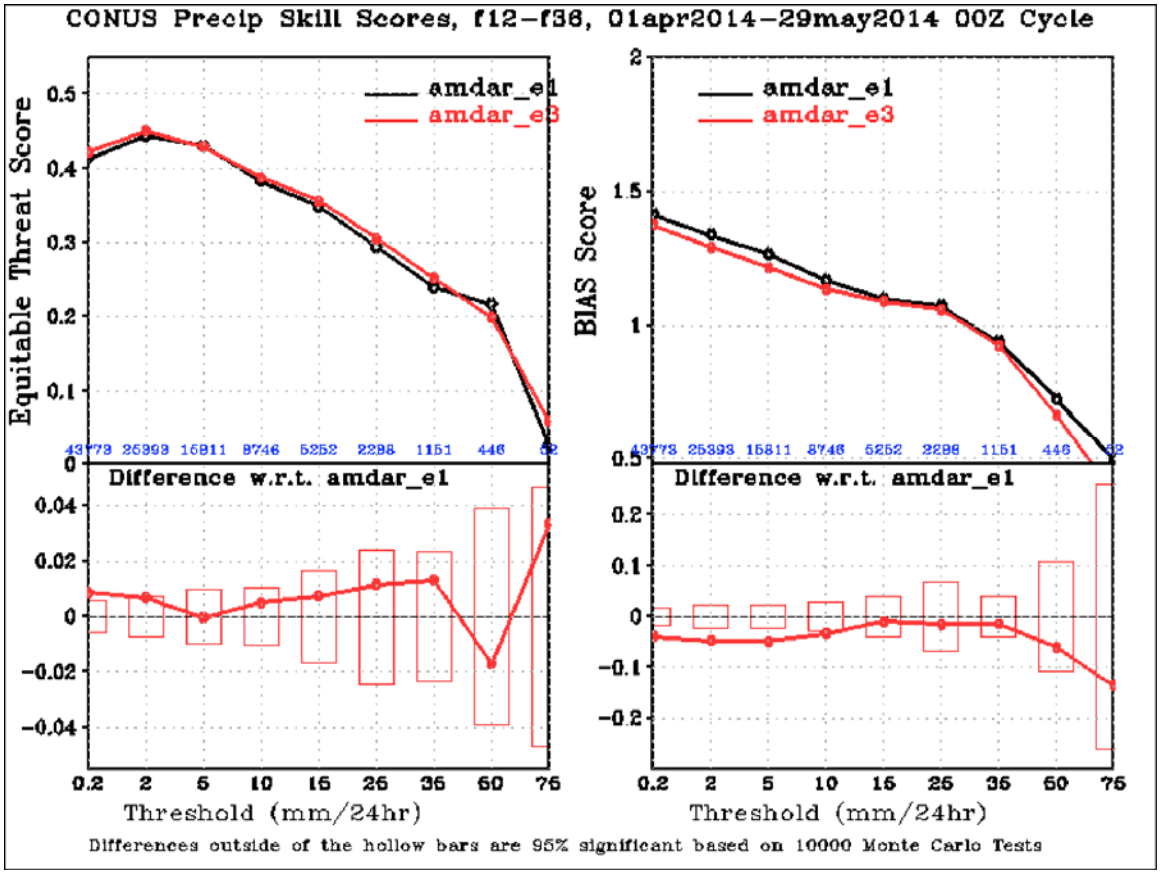

Figure 5. Precipitation skill score of the 12-36 hour forecast over the time-period of the data-denial experiment. (Left) equitable threat score (ETS) and (right) bias score by precipitation threshold, measured in mm/day. The assimilation experiment (radiosondes, aircraft moisture observations) is plotted in red. The data-denial experiment (selected radiosondes removed, aircraft moisture observations) is plotted in green. Below each plot is a plot of the difference with respect to the assimilation experiment. Bars indicate the necessary deviation from the control to achieve statistical significance.

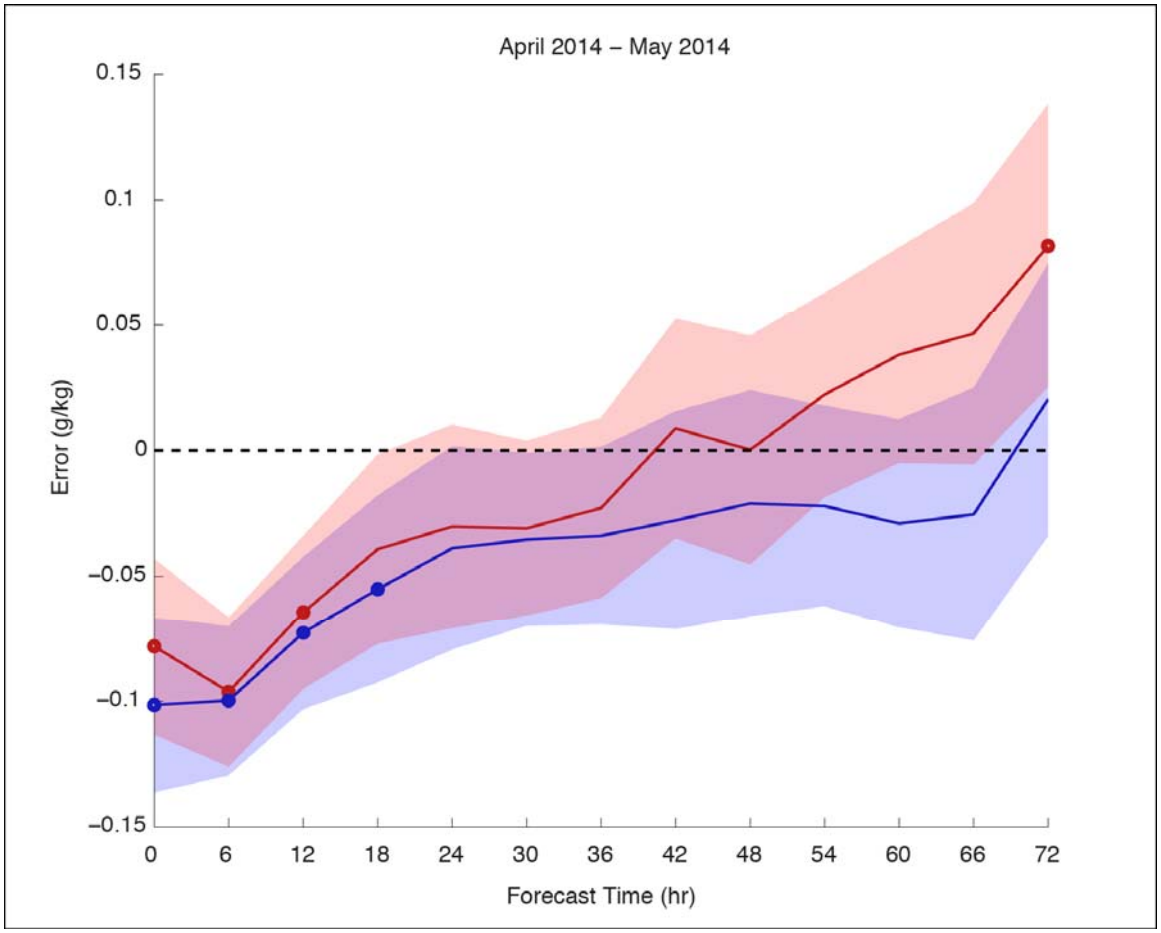

Figure 6. Difference in error between experiment and control, verifying the total atmospheric column precipitable water against GPS observations. The blue line represents the difference in the mean error over all forecasts at each forecast time between the assimilation experiment and the control. The red line represents the difference in the mean error over all forecasts at each forecast time between the assimilation + radiosonde denial experiment and the control. The shading around each mean represents the 95% confidence, and dots along the mean show where difference between experiment and control is statistically significant to 95% confidence.

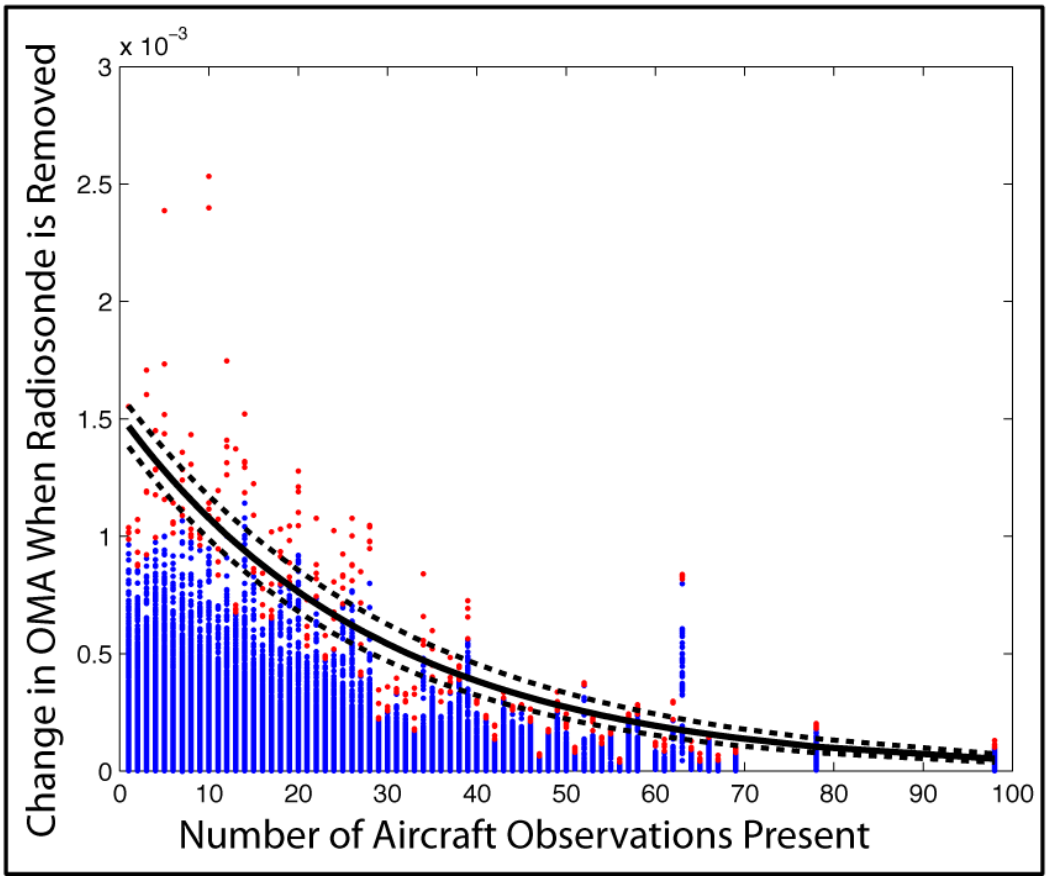

Figure 7. Phase-space plot of impact of denied radiosonde observations versus aircraft observational coverage of radiosonde launch site. Each point on the phase-space represents an aircraft moisture observation collocated to a denied radiosonde, with a position on the ordinate equal to the absolute change in ob-minus-analysis (OMA) when the radiosonde is denied, and a position on the abscissa equal to the number of aircraft moisture observations collocated to the radiosonde within the observation's pressure-slab. The red points are the five largest values on the ordinate for each unique value on the abscissa, and the black (solid) curve is an exponential curve fitted to the red points using a nonlinear least-squares approach. The 95% confidence limits on the parameters defining the best-fit curve are plotted as dashed curves.