

A Quality Assessment of the Real-Time Mesoscale Analysis (RTMA) for Aviation

MATTHEW T. MORRIS

Systems Research Group, Colorado Springs, Colorado, and NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland

JACOB R. CARLEY

NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland

EDWARD COLÓN, ANNETTE GIBBS, AND MANUEL S. F. V. DE PONDECA

I.M. Systems Group, Inc., Rockville, and NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland

STEVEN LEVINE^a

NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland, and Systems Research Group, Colorado Springs, Colorado

(Manuscript received 3 October 2019, in final form 10 February 2020)

ABSTRACT

Missing observations at airports can cause delays in commercial and general aviation in the United States owing to Federal Aviation Administration (FAA) safety regulations. The Environmental Modeling Center (EMC) has provided interpolated temperature data from the National Oceanic and Atmospheric Administration's Real-Time Mesoscale Analysis (RTMA) at airport locations throughout the United States since 2015, with these data substituting for missing temperature observations and mitigating impacts on air travel. A quality assessment of the RTMA is performed to determine if the RTMA could be used in a similar fashion for other weather observations, such as 10-m wind, ceiling, and visibility. Retrospective, data-denial experiments are used to perform the quality assessment by withholding observations from FAA-specified airports. Outliers seen in the RTMA ceiling and visibility analyses during events meeting or exceeding instrument flight rules suggest the RTMA should not be substituted for missing ceiling and visibility observations at this time. The RTMA is a suitable replacement for missing temperature observations for a subset of airports throughout most of the CONUS and Alaska, but not at all stations. Likewise, the RTMA is a suitable substitute for missing surface pressure observations at a subset of airports, with notable exceptions in regions of complex terrain. The RTMA may also be a suitable substitute for missing wind speed observations, provided the wind speed is ≤ 15 kt ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$). Overall, these results suggest the potential for RTMA to substitute for additional missing observations while highlighting priority areas of future work for improving the RTMA.

1. Introduction

Adverse weather conditions are often responsible for extensive flight delays, cancellations, and diversions, which create hardships for travelers and financial losses

for the aviation industry. According to the Federal Aviation Administration (FAA) Operations Network (OPSNET), 67.4% of the “delay events” recorded in 2018 were due to weather (OPSNET 2019). In addition, there were 40 weather-related accidents in fixed-wing aircraft in 2015, of which 31 had fatalities (AOPA Air Safety Institute 2018).

To ensure the safe operation of aircraft in the National Airspace System, the FAA maintains guidelines for weather observations, which are satisfied at most airports via the Automated Surface Observing System (ASOS)

^a Current affiliation: NOAA/NWS/Meteorological Development Laboratory, Silver Spring, Maryland.

Corresponding author: Matthew T. Morris, matthew.t.morris@noaa.gov

and Automated Weather Observing System (AWOS). Failures of these automated sensors result in additional delays and cancellations, even when the ambient weather conditions would not suggest an impact on air travel. For instance, a flight from Columbus, Georgia, to Atlanta, Georgia, on 1 March 2018 was delayed for 401 min due to an automated sensor failure and the subsequent timeout of the flight crew.

According to an early 2015 letter from Airlines for America, an airline industry trade association and lobbying group, temperature was the missing element in 85% of the missing weather observations impacting aircraft operations (McGraw 2015). To minimize the impact of missing temperature observations on aircraft operations, the FAA determined that the Real-Time Mesoscale Analysis (RTMA) temperature information is a legal substitute for a missing temperature report (FAA 2016). Shortly thereafter, the National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC) began providing temperature data interpolated from the RTMA product grid at airport locations across the United States (viz., the airport weather status list) to serve in lieu of missing temperature observations. Since the airport weather status list was implemented in July 2015, there has been interest in expanding the airport weather status list beyond 2-m temperature to include additional variables of interest to the aviation community, including 2-m dewpoint temperature, surface pressure, 10-m wind, 10-m wind gust, ceiling, and visibility. The purpose of this study is to assess the quality of RTMA operational products for a subset of the aforementioned fields, which may allow for a future expansion of the airport weather status list.

The RTMA is an hourly, two-dimensional variational (2D-Var) analysis system that produces analyses of sensible weather elements (De Pondeca et al. 2011; Pondeca et al. 2015) using the NCEP Gridpoint Statistical Interpolation (GSI; Wu et al. 2002) package.¹ The 2D-Var algorithm produces an analysis by minimizing a

cost function that measures the deviation between the current solution and both the background and observations, weighted by their respective error covariance matrices. For an in-depth explanation on the analysis procedure, see De Pondeca et al. (2011). The analysis system is run over domains that encompass the contiguous United States (hereafter, CONUS), Alaska, Hawaii, Puerto Rico, and Guam (Fig. 1). The RTMA is used predominantly for situational awareness and provides analyses of 2-m temperature, 2-m specific humidity, 10-m wind, 10-m wind gust, surface pressure, ceiling, visibility, and cloud cover. The RTMA was developed in response to growing demands for the NWS to produce high-resolution meteorological analyses, with one goal being to assist NWS forecasters in populating the National Digital Forecast Database (NDFD; Glahn and Ruth 2003) grids. The first version of the RTMA was developed as an initial step toward building the Reanalysis of Record (Horel and Colman 2005) and was implemented operationally in 2006.

The RTMA is designed to closely fit observations, more so than a traditional data assimilation scheme, as the analyses are not used to initialize subsequent model forecasts. Dynamical balances, which are very important to a model's initial conditions, can therefore be somewhat relaxed in the RTMA. Knopfmeier and Stensrud (2013) compared 2-m temperature, 2-m dewpoint, and 10-m wind analyses produced by an ensemble adjustment Kalman filter (EAKF) with those generated by the RTMA in order to quantify the impact of mesonet observations on analysis quality and found lower root-mean-square (RMS) innovations² in the RTMA for temperature and dewpoint, suggesting the RTMA provides a better fit to the observations for those fields, but not winds. Ancell et al. (2014) compared surface temperature and wind analyses from the ensemble square root filter (EnSRF) and RTMA, hypothesizing that the EnSRF analyses would be superior to the RTMA analyses owing to the use of flow-dependent error covariances. The RTMA temperature analyses were slightly more skillful than EnSRF analyses, consistent with the results in Knopfmeier and Stensrud (2013) when using the EAKF. The wind analyses produced by both variants of the EnKF were superior to those produced by the RTMA. This is likely attributable to the fact that the RTMA is a univariate 2D-Var analysis system that relies principally on a static background error covariance. The EAKF and EnSRF both incorporate flow-dependent and multivariate error covariances,

¹ Analyses from the RTMA are valid at the top of the hour and are typically available 43, 38, 34, and 33 min after analysis time for the CONUS, Alaska, Hawaii, and Puerto Rico domains, respectively. A companion system, the Unrestricted Mesoscale Analysis (URMA), is run 6 h after the RTMA to assimilate late-arriving observations and is used for calibration and validation of the National Blend of Models (NBM; Gilbert et al. 2016). Along with the fields analyzed by the RTMA, URMA provides analyses of significant wave height (Hs) over oceans and minimum/maximum temperature (once daily). In addition, the Rapid Update RTMA (RTMA-RU) produces analyses every 15 min over the CONUS grid (valid at the top of the hour and 15, 30, and 45 min afterward) with a special focus on aviation applications; these analyses are available ~13 min after analysis time.

² "Innovation" refers to the difference between the model simulated observation and the observation.



FIG. 1. Boundaries of the RTMA CONUS, Alaska, Hawaii, and Puerto Rico domains are outlined in red. The boundary of the RTMA Guam domain is not shown.

which likely improved the wind analysis relative to the 2D-Var RTMA system.

While the RTMA is designed to provide a close fit to observations, all observations do not contribute equal value to the analysis. Tyndall and Horel (2013) used the adjoint of a 2D-Var surface analysis to objectively quantify the impact of nearly 20 000 surface observations using a sample of 100 analyses spanning 25 high-impact weather events. When considering individual analyses, high-impact observations coincided with regions of significant weather where the observed conditions were not accurately resolved in the background fields. Over the entire study period, observation impact was found to be a function of the observation density and local weather variability. Moreover, data denial experiments have shown that removing up to 75% of mesonet observations from EnKF analyses resulted in only nominal decreases in analysis quality as measured by RMS innovations (Knopfmeier and Stensrud 2013).

The results in Tyndall and Horel (2013) and Knopfmeier and Stensrud (2013) suggest that the RTMA may be able

to provide skillful results in the absence of a subset of METAR observations. This study performs a quality assessment of the RTMA via retrospective, data denial experiments, with a goal of determining if the RTMA could substitute for missing weather observations beyond temperature and further limit the impact of automated sensor failures on aircraft operations. Section 2 describes the data and methods. Section 3 presents results for the RTMA ceiling, visibility, temperature, surface pressure, and wind analyses. Section 4 provides the summary and discussion, including avenues for future work.

2. Data and methods

a. RTMA

The background, or first guess, for the RTMA CONUS domain is derived from the most recent forecast (typically 1 h) from the downscaled High-Resolution Rapid Refresh (HRRR; Alexander et al. 2015) for all fields, with the exception of temperature, surface pressure, and

moisture. These three fields are a blend of the downscaled HRRR and the most recent downscaled 3-km North American Mesoscale (NAM; Carley et al. 2017; Rogers et al. 2017) CONUS nest forecast valid at the analysis time. Since the HRRR and NAM nest domains do not cover the full RTMA CONUS domain, the most recent forecast from the Rapid Refresh (RAP; Benjamin et al. 2016) is used to fill gaps along the edges of the domain. Similarly, the RTMA Alaska domain leverages the most recent forecast from the downscaled RAP for all fields, with the exception of temperature, surface pressure, and moisture, which are a blend between the downscaled RAP and the latest forecast from the downscaled 3-km NAM Alaska nest. Beginning with RTMA version 2.7, first guess fields from the RAP are replaced with the new HRRR-Alaska (implemented 4 December 2018). However, the HRRR-Alaska was not available for the time periods considered in this study. The background, or first guess, for the Hawaii (Puerto Rico) domain is derived from the most recent forecast from the downscaled 3-km NAM Hawaii (Puerto Rico) nest. Finally, the most recent forecast from the HiRes Window ARW (Skamarock et al. 2008) composes the background for the Guam domain.

The RTMA uses observations from 30 min prior to 30 min after the analysis time and has a data cutoff of 30 min after the analysis time. Land-based surface observations include traditional surface observations such as METARs and synoptic observations. RTMA also takes in data from numerous mesonets, such as UrbaNet, state/university sponsored mesonets (e.g., Oklahoma Mesonet), Citizen Weather Observer Program (CWOP), state Department of Transportation (DOT) Road Weather Information Systems (RWIS), and Remote Automated Weather Stations (RAWS). Buoys, Coastal-Marine Automated Network (C-MAN) platforms, and ships provide sea-based surface and near-surface observations. Finally, satellite observations of sky cover come from the GOES Imager (Gerth 2018), while ASCAT and WindSat provide satellite observations of wind over water. In addition, the RTMA assimilates satellite low-level cloud drift winds over the oceans.

The RTMA system uses several quality control (QC) checks to filter out observations of poor quality. The gross error check compares the absolute value of the errors (observation minus background), normalized by the observation error, to a predetermined threshold.³ Values exceeding the predetermined threshold result

in a rejected observation. The gross error check is relaxed in regions of complex terrain where the background fields are often unable to resolve terrain-induced features (e.g., valley cold pools). Fixed station reject lists are also used for stations consistently reporting bad observations. These fixed station reject lists can only be updated during system upgrades and are therefore used sparingly. While mesonet wind observations are rejected by default from the CONUS analyses due to siting concerns, these observations may be used if testing reveals the observation is of sufficient quality or the observations are from known, high-quality networks (e.g., universities). Variational (nonlinear) quality control is used for temperature, pressure, specific humidity, and wind and is designed to adjust the weight assigned to a particular observation based on its fit to nearby observations and the background fields (Purser 2011, 2018). A dynamic, manually updated, station blacklist is also maintained to quickly remove problematic observations that are not captured and handled by the automated QC procedures. This list allows for any combination of observations of temperature, moisture, pressure, and wind (including gusts) to be rejected from the analysis.

b. Experiment design

We seek to quantify the impact of missing observations from airports specified by the FAA, namely those airports having “Part 139” certification (FAA 2019), on the resulting RTMA analysis to help guide the possible use of the RTMA in lieu of missing observations. Part 139 regulations require the FAA to issue operating certificates to airports serving scheduled and unscheduled air carrier aircraft with greater than 30 seats, as well as those serving scheduled air carrier aircraft with 10–30 seats. These regulations are designed to ensure safe air travel. At the time of this study, there were 530 Part 139 qualifying airports: 484 in the CONUS, 25 in Alaska, 8 in Hawaii, 3 in Puerto Rico, 1 in Guam, and 9 in the remaining U.S. territories.

The quality assessment is performed through the use of retrospective, data-denial experiments spanning four seasons and encompassing four domains: CONUS, Alaska, Hawaii, and Puerto Rico. Retrospective experiments are not run on the Guam domain due to the existence of only one Part 139 qualifying airport located within that region. Three experiments are performed: the CONTROL, EXP, and NODA. The CONTROL experiment assimilates all available, quality-controlled observations, as would happen in the operational RTMA. The EXP experiment rejects observations from stations on the list of Part 139 airports as a way to simulate the impact on the RTMA of missing those observations. While missing all Part 139 observations is a highly

³The predetermined threshold varies by report type (e.g., METAR or mesonet), and is sensitive to the observation error, which is not a static value for all stations in a particular network.

unlikely scenario, such an event is not unprecedented and has occurred in rare operational situations (e.g., a network failure). Finally, the NODA experiment does no data assimilation and corresponds to the background field. NODA serves as the “worst case scenario” baseline (i.e., no observations are assimilated). A comparison between the EXP and CONTROL experiments will help quantify the degradation seen at the airports for which observations are not available. If observations are missing at an airport, they will likely be missing from the RTMA as well. The retrospective periods each cover two weeks of spring, summer, fall, and winter (Table 1). For each retrospective period, the RTMA is initialized at 0000 UTC on the preceding day to allow sufficient time for the system to create an initial bias correction field for the temperature analysis and to generate analyses used in the so-called first guess at appropriate time procedure (FGAT; Lawless 2010). These additional analysis cycles are not included in the quality assessment results as they would be slightly degraded and not representative of the operational RTMA product.

The retrospective runs utilize the most recent version of the RTMA at the time of the assessment (version 2.7; Carley et al. 2018), which went into operations at 1200 UTC 4 December 2018, along with the first guess fields that were available in real time (with some exceptions). For instance, ceiling height was not analyzed over the Alaska domain prior to the December 2017 implementation, so ceiling data were appended to the downscaled RAP background for the summer 2017 and fall 2017 retrospective periods in order to accommodate the ceiling height analysis. In addition, the NAM downscaling procedure was rerun for the winter 2018 period as an obsolete terrain dataset was used to produce the operational real-time files.

3. Results

a. Ceiling

Low ceiling and visibility conditions place restrictions on the rate of air traffic into and out of airports. For instance, San Francisco, California, experiences frequent low stratus events and accompanying low cloud ceiling heights. This reduces the capacity for arrivals at San Francisco International Airport (KSFO) due to restrictions placed on parallel approaches (San Francisco International Airport 2010). Although KSFO is an airport with human augmentation (i.e., not susceptible to missing observations), the frequent low ceiling heights makes this a good location to assess the quality of RTMA ceiling analyses. Accurately predicting the clearing time is crucial for minimizing the

TABLE 1. Retrospective periods (i.e., seasons) used in the quality assessment experiments.

Period	Start	End
Summer 2017	0000 UTC 27 Jun 2017	2300 UTC 10 Jul 2017
Fall 2017	0000 UTC 1 Oct 2017	2300 UTC 14 Oct 2017
Winter 2018	0000 UTC 1 Jan 2018	2300 UTC 14 Jan 2018
Spring 2018	0000 UTC 1 Apr 2018	2300 UTC 14 Apr 2018

financial losses associated with delays, cancellations, and diversions.

The overall quality of RTMA ceiling analyses across all Part 139 airports is first assessed through the use of boxplots, which provide a visual representation of the range of analysis values stratified by METAR-observed flight category (LIFR, IFR, MVFR, and VFR; Table 2). Figure 2 shows the boxplots of ceiling for each experiment, aggregated across all retrospective periods and domains. Figure 2 reveals that the RTMA correctly characterizes the majority of LIFR conditions in both the CONTROL and NODA experiments as evidenced by the full interquartile range (IQR) for these experiments falling within the LIFR category. The EXP configuration performs similarly, but is slightly degraded for LIFR conditions as the IQR extends into the IFR range at approximately the 62nd percentile. All experiments feature a preponderance of outliers, indicating several events where the RTMA struggled to capture the LIFR event regardless of the configuration. Outliers are also observed under IFR and MVFR conditions, which may preclude the RTMA serving as a substitute for missing ceiling observations at Part 139 airports. Under all low ceiling categories (LIFR, IFR, and MVFR), EXP is similar to NODA, but features slight degradation relative to CONTROL in the form of a larger IQR, which indicates a greater variability in the range of analyzed conditions for LIFR, IFR, and MVFR flight categories. Little, if any, discernible trend is seen for the VFR flight category.

While the boxplots (Fig. 2) provide results aggregated across all stations, categorical statistics for dichotomous events (Wilks 2011), namely two-dimensional station maps of the percent change of critical success index (CSI; Gilbert 1884), are used to identify regional patterns in the analysis quality and identify stations where the RTMA struggles in the absence of Part 139 observations. In this study, a low ceiling event occurs when the ceiling value falls in a flight category at least as restrictive as the flight category being considered (Table 2). A “hit” occurs when a low ceiling event is both analyzed and observed, a “false alarm” occurs when an event is analyzed but not observed, a “miss” occurs when an event is not analyzed but is observed, and a “correct

TABLE 2. Flight categories for ceiling and visibility. Categories at the top of the table are most restrictive. In aircraft operations, the flight category would correspond to the most restrictive category obtained when considering ceiling and visibility conditions. However, in this work, the ceiling and visibility analyses are assessed individually.

Flight category	Ceiling (ft AGL)	Visibility (mi)
Low instrument flight rules (LIFR)	≤ 500	≤ 1
Instrument flight rules (IFR)	500 to ≤ 1000	1 to ≤ 3
Marginal visual flight rules (MVFR)	1000 to ≤ 3000	3 to ≤ 5
Visual flight rules (VFR)	> 3000	> 5

null” occurs when an event is neither analyzed nor observed. For instance, when considering MVFR ceiling heights, an observed ceiling height of 950 ft (IFR) and an analyzed ceiling height of 450 ft (LIFR) corresponds to a hit. Here, CSI and its associated percent change are defined as

$$\text{CSI} = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{false alarms}}, \quad (1)$$

$$\text{percent change} = \frac{\text{TEST} - \text{CONTROL}}{\text{CONTROL}} \times 100\%, \quad (2)$$

where TEST can be either the NODA (skill loss or gain associated with no assimilation) or EXP (skill loss or gain from the missing Part 139 observations) experiment. The maps are presented alongside the CSI values from CONTROL, which serve as a baseline of the RTMA performance with the full suite of observations assimilated. Note that a station will only be plotted if at least 30 events were observed in a category that is at least as restrictive as the one being considered.

Figure 3 presents station maps of ceiling statistics, aggregated across all retrospective periods, which are focused on central California and the San Francisco Bay (a region that experiences frequent low ceiling events that impact aircraft operations). There were 268 events with ceilings ≤ 3000 ft at KSFO during all periods. Overall, the CSI scores in CONTROL improve as the flight categories become less restrictive (e.g., moving from LIFR to MVFR conditions), likely a result of the observed events becoming more frequent in less restrictive categories. However, comparatively lower scores are found at KSFO for each flight category. Under LIFR

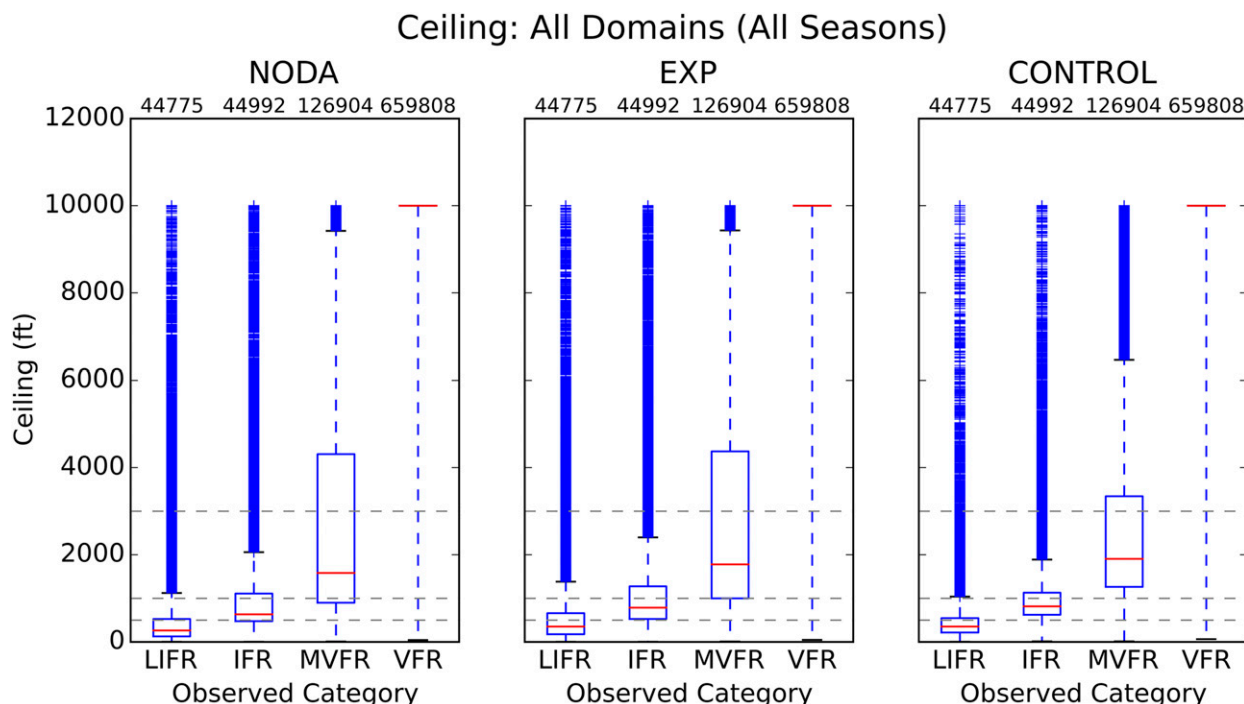


FIG. 2. Ceiling boxplots showing results aggregated across all retrospective periods and domains for each experiment. The boxplots are stratified by the METAR-observed flight category, with dashed horizontal lines delineating these categories. The rectangle shows the distribution between the first and third quartiles, while the median is shown with a red line. Note that ceiling values are capped at 10 000 ft, and the observed sample size is provided above the boxplots. In a perfect system, the LIFR boxplot would be entirely contained below the lowest dashed line at 500 ft.

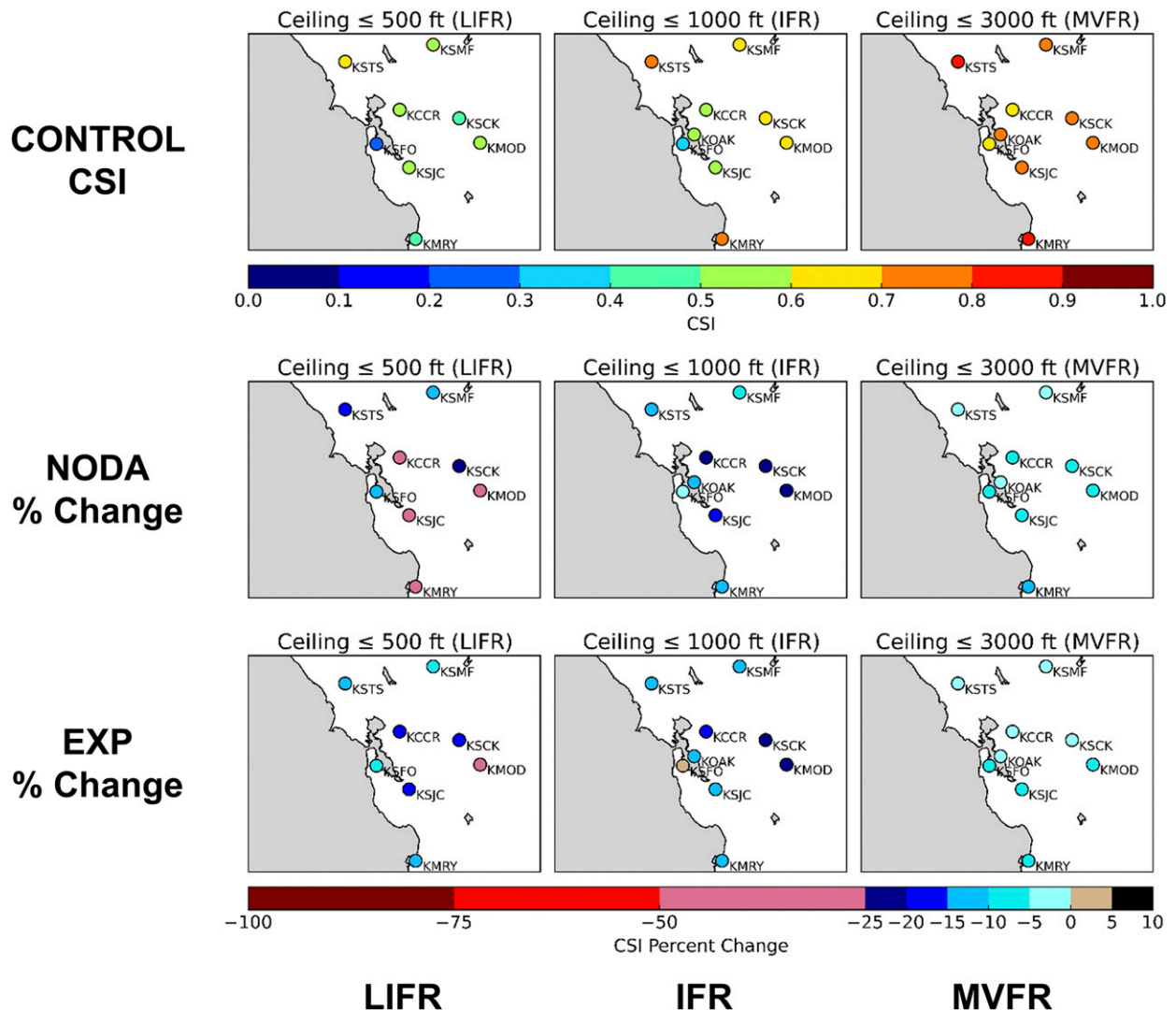


FIG. 3. Station maps of ceiling statistics aggregated across all retrospective periods, focused on the San Francisco Bay area in California; note that all stations shown are Part 139 airports. (top) The CSI from the CONTROL experiment and (middle),(bottom) the percent change of CSI relative to CONTROL for NODA and EXP, respectively.

conditions, the CSI values in CONTROL are ≤ 0.7 at all stations in the area shown, with a CSI of 0.2–0.3 at KSFO. NODA is degraded relative to CONTROL at all stations within the area shown, with degradations of $\sim 10\%$ seen at KSFO relative to the already low CSI score in CONTROL. Similar degradations are seen at KSFO in EXP, with degradations of up to 50% seen elsewhere. Under IFR or more restrictive conditions most stations exhibit a degradation of less than 25% relative to CONTROL in NODA. EXP is degraded relative to CONTROL at all stations, often by a smaller magnitude than seen under LIFR conditions. Under MVFR or more restrictive conditions, NODA is degraded relative to CONTROL by $\sim 10\%$ at all stations, with a similar pattern seen in EXP.

Figure 3 also reveals a positive percent change at KSFO for IFR conditions, indicating that CONTROL has a lower CSI score than seen in NODA; similarly EXP shows greater degradation than NODA. While initially counterintuitive, this is a result associated with the analysis of variables of a more discrete nature with non-Gaussian error statistics. In a variational analysis system employing only a climatologically specified background error, such as the RTMA, analysis increments are spread to nearby grid points with the spatial extent controlled by the decorrelation length (i.e., the background error covariance matrix). In the situations where the assimilation of Part 139 observations appears to degrade the analysis, nearby stations offer conflicting ceiling information reflective of the highly localized

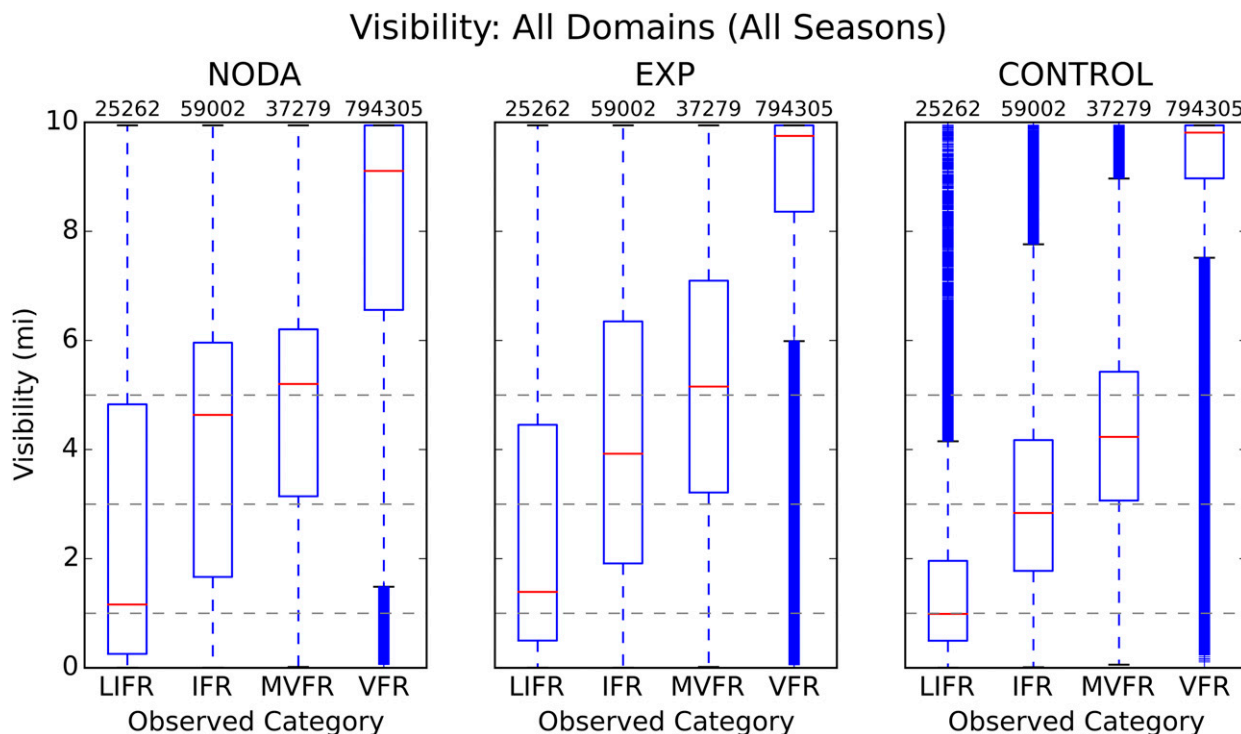


FIG. 4. Visibility boxplots showing results aggregated across all retrospective periods and domains for each experiment. The boxplots are stratified by the METAR-observed flight category, with dashed horizontal lines delineating these categories. The observed sample size is provided above the boxplots.

nature of low ceiling events. Since the analysis system assimilates all observations simultaneously, the resulting analysis will be a compromise, raising the ceiling at one location while lowering it at another. As an example, imagine two nearby stations, one reporting a ceiling height of 800 ft and the other reporting 15 000 ft. The analysis value in this vicinity, assuming no other nearby observations, could well be on the order of say 8000 ft. The exact value will be determined by the background field, background error model, observation errors, and how far the pair is from the other observations. The point here is that, in this hypothetical scenario, the ceiling height analysis would indicate VFR conditions at both stations despite IFR conditions occurring at the former.

Taken together, the results suggest that the RTMA is likely not a viable replacement for missing ceiling observations at Part 139 airports, unless future work is able to show the outliers seen in Fig. 2 are dominated by a small subset of Part 139 airports.

b. Visibility

Figure 4 shows the boxplots of visibility for each experiment and reveals that low visibility events are less frequent than low ceiling events as depicted in Fig. 2;

when considering LIFR conditions, there are 25 262 visibility events and 44 775 ceiling events, comprising 2.76% and 5.11% of the total observations, respectively. Furthermore, EXP is unable to capture most LIFR, IFR, and MVFR visibility events, with more than half of the analysis values falling outside the bounds of the respective flight category. EXP closely mirrors NODA and is degraded relative to CONTROL under LIFR, IFR, and MVFR conditions. The IQR falls within the variability limits for all experiments under VFR conditions, although the whiskers extend beyond these limits in NODA. Given the poor performance, the RTMA is likely not a suitable replacement for missing visibility observations at Part 139 airports.

As was done for ceiling in section 3a, two-dimensional station maps of the percent change of CSI are also used to assess the quality of RTMA visibility analyses in the vicinity of San Francisco, California (Fig. 5). Figure 5 reveals that, similar to ceiling, the CSI scores in CONTROL improve as the flight category becomes less restrictive. Under LIFR conditions, the CSI scores in CONTROL are lower than those seen under LIFR ceiling conditions, with the lowest score seen in Monterey, California (KMRY). KSFO experienced fewer than 30 LIFR

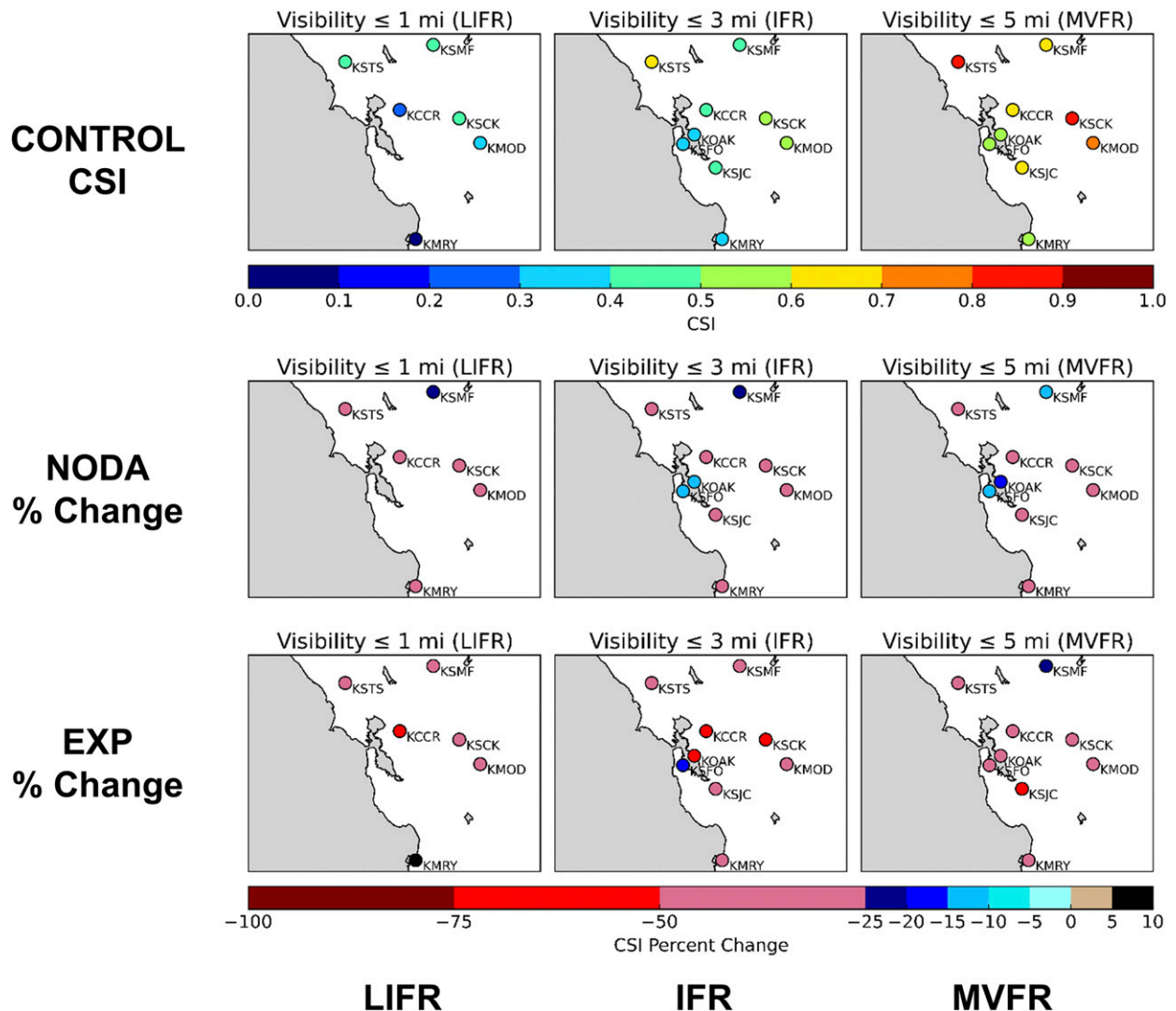


FIG. 5. Station maps of visibility statistics aggregated across all retrospective periods, focused on the San Francisco Bay area in California; note that all stations shown are Part 139 airports. (top) The CSI from the CONTROL experiment and (middle),(bottom) the percent change of CSI relative to CONTROL for NODA and EXP, respectively.

visibility events during the entirety of this study and is not plotted. NODA shows degradations between 25% and 75% at all stations except for KMRY, which exhibits higher CSI scores in NODA than in CONTROL, similar to what was seen at KSFO for ceiling. Under IFR or more restrictive conditions, all stations exhibit CSI values < 0.7 , including KSFO (0.3–0.4). Degradations of $\sim 20\%$ are seen at KSFO in NODA, with larger degradations seen elsewhere. Under MVFR conditions, degradations of 20% or greater are seen at all stations in NODA, with smaller degradations seen at several stations in EXP, most notably at KSFO.

The statistics depicted in Figs. 4 and 5 demonstrate that the RTMA system struggles to capture ongoing low

visibility events even when the observations are assimilated. This is undoubtedly exacerbated by the fact that such events are relatively rare and spatially discontinuous which leads to visibility having non-Gaussian error statistics. While recent strides have been made in the RTMA analysis algorithm to improve assimilation of non-Gaussian variables (Yang et al. 2018, 2019), these results demonstrate that visibility remains a significant challenge.

c. Temperature

Surface fields, such as 2-m temperature, are also assessed using boxplots, which provide a visual representation of the distribution of data and allow for

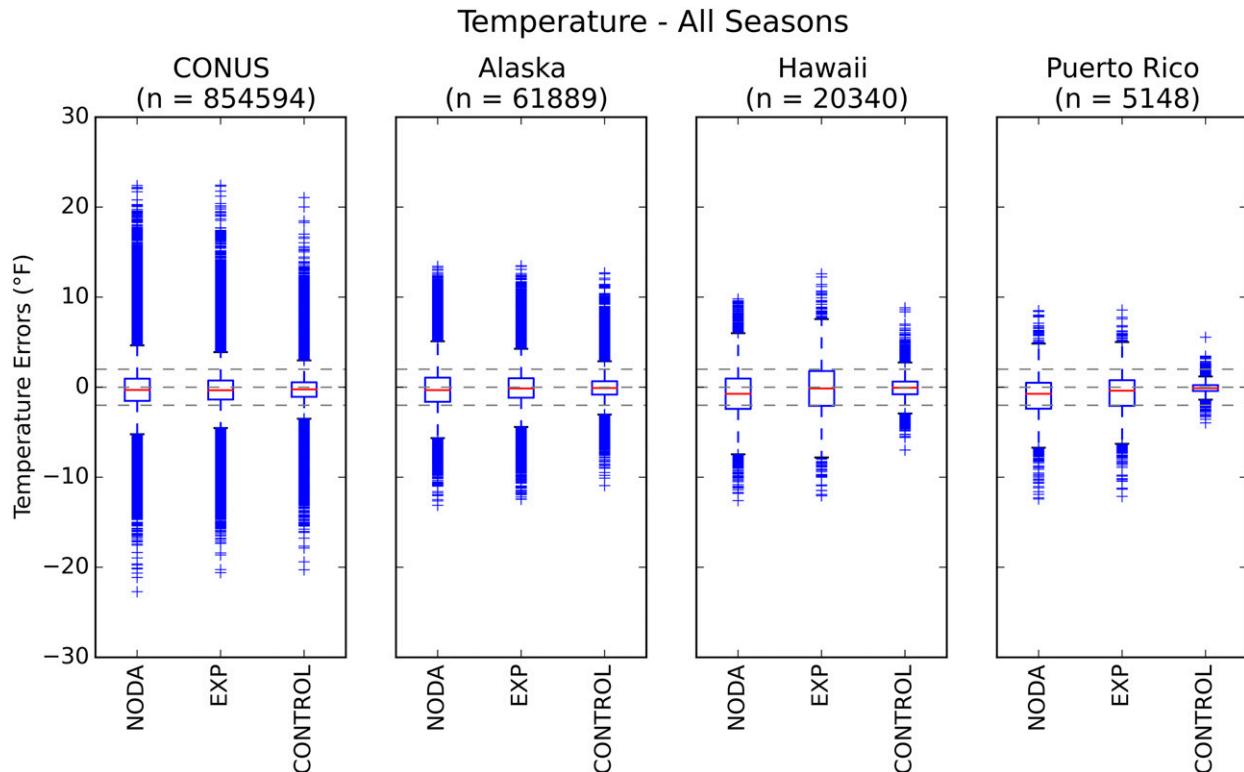


FIG. 6. Boxplots of 2-m temperature errors, showing results aggregated across all retrospective periods by domain. The observed sample size is provided for each domain. Dashed horizontal lines correspond to variability values provided by the FAA (Table 3).

comparisons among the various experiments. The surface field boxplots in Fig. 6 depict the errors⁴ (i.e., test minus observation) for a given observed value instead of the full analyzed and observed values and are aggregated across all retrospective periods over the CONUS, Alaska, Hawaii, and Puerto Rico domains. A positive (negative) value implies that the analysis or background is too warm (cold). Dashed horizontal lines correspond to guidelines for acceptable variability from the FAA (Table 3). These variability ranges are interpreted here as limits for which the RTMA may be able to serve as a substitute for missing Part 139 observations.

Figure 6 reveals that EXP offers slight improvement relative to NODA over CONUS and Alaska, as evidenced by a narrower distribution. Little difference is seen over Hawaii and Puerto Rico, though NODA (i.e., first guess fields) exhibits a slight cold bias over these areas. This bias is reduced in EXP and CONTROL over both domains. EXP is degraded relative to CONTROL over all domains, as evidenced by larger IQR values, but

these differences are more substantial over Hawaii and Puerto Rico than over CONUS and Alaska. These differences are likely due to the Part 139 observations comprising a larger portion of the overall observing system through these smaller domains. Across all domains, the whiskers⁵ and outliers extend beyond the variability bounds in all experiments. Given these results, the RTMA is likely a suitable substitute for missing temperature observations at some Part 139 airports, but future investigation of the aforementioned outliers should be performed to identify stations where this is not the case.

Diurnally aggregated time series plots of root-mean-square error (RMSE; Fig. 7) and bias (Fig. 8) correspond well with the fully aggregated boxplots (Fig. 6) by showing the closest fit to Part 139 observations (i.e., lowest RMSE values) in the CONTROL experiment over each domain. In Figs. 7 and 8, statistical significance is tested at the 95% level using bootstrap confidence intervals with replacement using 10 000 replications. EXP has a roughly 0.5°F lower RMSE value than

⁴ While the term “error” is not strictly proper in this case, as it implies the observations are truth and contain no error, we adopt the term “error” here in the interest of brevity.

⁵ The upper whisker denotes the greatest error up to $Q3 + 1.5IQR$, while the lower whisker denotes the lowest error down to $Q1 - 1.5IQR$, where $Q1$ and $Q3$ are the 25th and 75th percentiles, respectively.

TABLE 3. Quality assessment guidelines obtained from the FAA. These variability guidelines are largely based off of performance standards for nonfederal AWOS stations.

Observed element	Category	Variability
Temperature	-40° to 120°F	$\pm 2^{\circ}\text{F}$
Pressure ^a	28.5 to 31.5 in. Hg	± 0.02 in. Hg
Wind speed	Calm to 15 kt	± 3 kt
	16 to 40 kt	± 5 kt

^a Alternatively, the category limits for pressure are approximately 965.1–1066.7 hPa, while the variability limits are roughly ± 0.68 hPa.

NODA over CONUS and Alaska, which is statistically significant. This is consistent with the narrower distribution relative to NODA seen in Fig. 6 and is likely a result of the influence of nearby observations (e.g., mesonet observations) adjusting the first guess fields and improving the analysis even in the absence of the Part 139 observations. Relative to CONTROL, EXP is degraded by a comparable magnitude, with larger degradations seen over Hawaii and Puerto Rico.

Figure 8 reveals only minor differences in the temperature bias between CONTROL and EXP over CONUS and Alaska. A notable cold bias is seen in NODA

over Hawaii between roughly 1700 and 0800 UTC (up to -2.0°F) with a slight cold bias ($\sim -0.25^{\circ}\text{F}$) at the remaining analysis hours. Bias values in EXP move in the positive direction when compared against NODA for all analysis hours except between 2000 and 2200 UTC. EXP also introduces a slight warm bias (less than 0.5°F) between 0400 and 1700 UTC. Despite this pattern in bias, little, if any, discernible diurnal trends are seen in the RMSE values over Hawaii (Fig. 7). Over Puerto Rico, the RMSEs for NODA and EXP increase markedly after 1100 UTC to approximately 3.0°F . This pattern is likely in response to an increased cold bias (up to -2.5°F) in NODA beginning around 1100 UTC (Fig. 8). This bias is reduced in magnitude for the remaining experiments, with more notable improvements seen in CONTROL, even though CONTROL has a slight cold bias when the bias in NODA is greatest in magnitude.

Based upon statistics aggregated by region and averaged over all times (Fig. 6), it appears that 2-m temperature is a suitable replacement for a subset of Part 139 observations. However, upon investigating each region as a function of the time of day, regional variations can

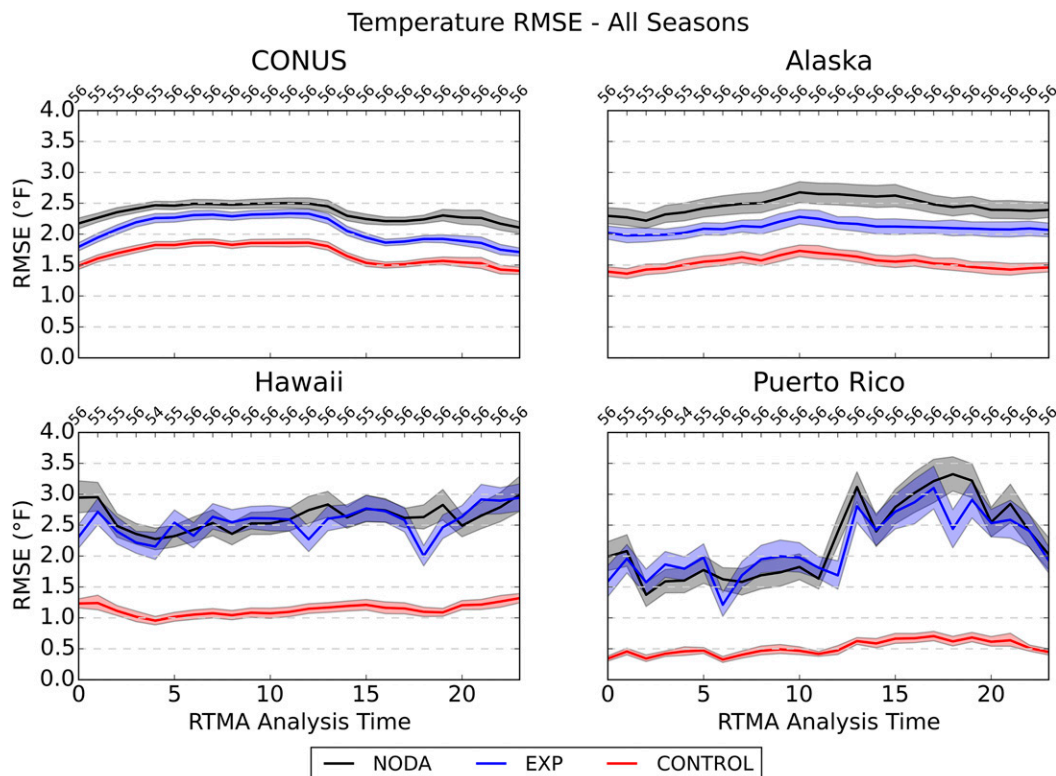


FIG. 7. Time series plots of temperature RMSE across all retrospective periods by domain. The number of analysis cycles reporting temperature observations for a given cycle hour is provided above the time series for each domain. Small discrepancies are a result of missing observations and/or input data, and thus no assessment could be performed for those cycles. The shaded area for each experiment corresponds to bootstrapped ($n = 10\,000$) 95% confidence intervals.

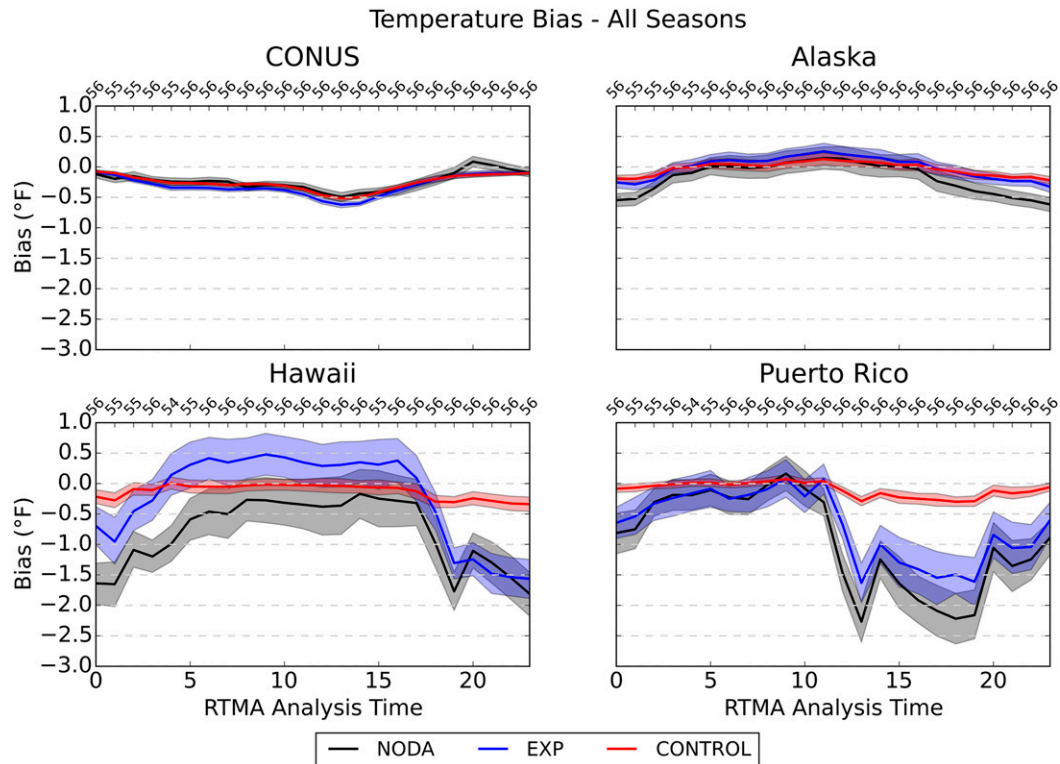


FIG. 8. As in Fig. 7, but for temperature bias.

be considerable, most notably over Puerto Rico. We hypothesize that a number of outliers seen in Fig. 6 correspond to stations from regions characterized by strong diurnal variability and/or stations that are somewhat isolated with relatively few nearby mesonet observations (e.g., Tyndall and Horel 2013). Investigation of this hypothesis is the subject of a planned follow-on project.

d. Surface pressure

Accurate observations of atmospheric pressure are critical in ensuring aircraft safety during takeoff and landing and are used to adjust the altimeter setting. Boxplots of surface pressure errors (Fig. 9) reveal a substantial low median bias in NODA over Alaska, Hawaii, and Puerto Rico with more than half of the errors in Hawaii and Puerto Rico exceeding the variability guidelines specified in Table 3. It is noted that these three regions are characterized by complex terrain. EXP offers little, if any, improvement over Hawaii and Puerto Rico relative to NODA, with more notable improvements seen over Alaska.

The bias is markedly reduced over each domain when considering the CONTROL experiment. Although the overall bias in NODA is neutral over the CONUS, biases likely exist in regions of complex terrain. EXP is clearly degraded relative to CONTROL over each

domain, as seen through wider boxplots and, in some cases, increased bias. These results suggest that the RTMA may be a suitable replacement for missing surface pressure observations in areas of noncomplex terrain.

e. Wind speed

Wind speed observations pose a unique quality control challenge in the RTMA. Most observations used in the RTMA are not located on the grounds of an airport or an otherwise flat, open environment. Many observations, especially mesonet observations near Part 139 airports, are located in urban or suburban environments. The term “mesonet” refers to a combination of multiple platforms ranging from stations provided by weather enthusiasts to high-quality university-sponsored networks. University networks notwithstanding, mesonet observations typically have little metadata and are often sited in nonstandard ways that can introduce biases in the resulting analysis. For instance, the RTMA assumes a priori that all wind observations are taken at a height of 10 m AGL. Even a casual surveying of observations used in a typical analysis shows that this is not the case as most are taken closer to the ground. In addition, wind speed observations have been shown to be more sensitive to nearby trees and buildings than

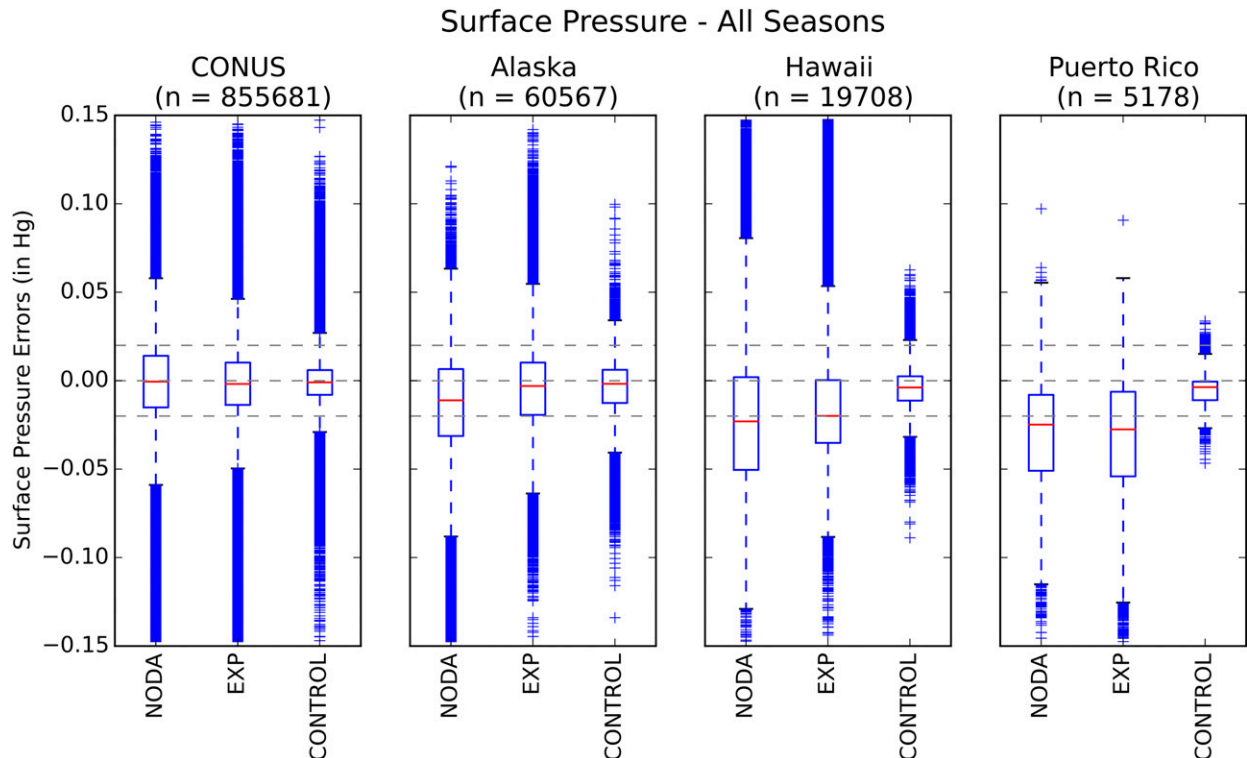


FIG. 9. Boxplots of surface pressure errors, showing results aggregated across all retrospective periods by domain. The observed sample size is provided for each domain. Dashed horizontal lines correspond to variability values provided by the FAA (Table 3).

observations of other variables (Fujita and Wakimoto 1982). Nearby obstructions and wind sensor heights of <10 m AGL combine to result in wind speed observations from these stations that are often much lower than those at airports or the background field at the relevant location. The assimilation of these observations often results in a wind speed analysis that has a low bias when compared to METAR/airport observations. Research is underway to obtain a more thorough set of metadata, advance the observation operator, and enhance quality control to address this outstanding challenge.

Boxplots of wind speed errors are stratified based on the observed wind speed, with observed wind speeds ≤ 15 kt ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$) plotted in Fig. 10 and observed wind speeds > 15 kt plotted in Fig. 11. The boxplots are stratified according to different variability values in guidelines from the FAA (Table 3): ± 3 kt for wind speeds ≤ 15 kt and ± 5 kt for wind speeds > 15 kt.

When wind speeds are ≤ 15 kt, the median overall bias in NODA is neutral over the Alaska and Hawaii domains, slightly low over CONUS (~ -1 kt) and slightly high over Puerto Rico (~ 1.5 kt; Fig. 10). An enhanced low bias is introduced in EXP relative to NODA over all domains except for Puerto Rico due to the assimilation

of mesonet observations from networks with nonstandard instrument siting, which tend to impart a low wind speed bias and reduce analyzed wind speeds in the surrounding areas. Note that there are far fewer mesonet observations over the Alaska, Hawaii, and Puerto Rico domains than over CONUS, resulting in smaller reductions in the background wind speeds over those domains. A slight high bias (less than 1 kt) is evident in EXP over Puerto Rico, perhaps due to the limited number of mesonet observations within the domain, resulting in more of a carryover of the high bias in the background (as noted in NODA). CONTROL has a narrower distribution than EXP over all domains, with a slightly reduced median bias over the CONUS and Alaska domains.

When wind speeds > 15 kt are observed (Fig. 11) a larger low bias (i.e., more negative) is seen in NODA across all domains relative to cases when winds are ≤ 15 kt. This bias is exacerbated in EXP due to the impact of surrounding nonstandard mesonet observations. For instance, over CONUS, the median value for all experiments is below the minimum variability value of -5 kt (i.e., more than half of the errors are greater than 5 kt in magnitude). Overall, the IQR for CONTROL is similar to that seen in NODA over all domains except over

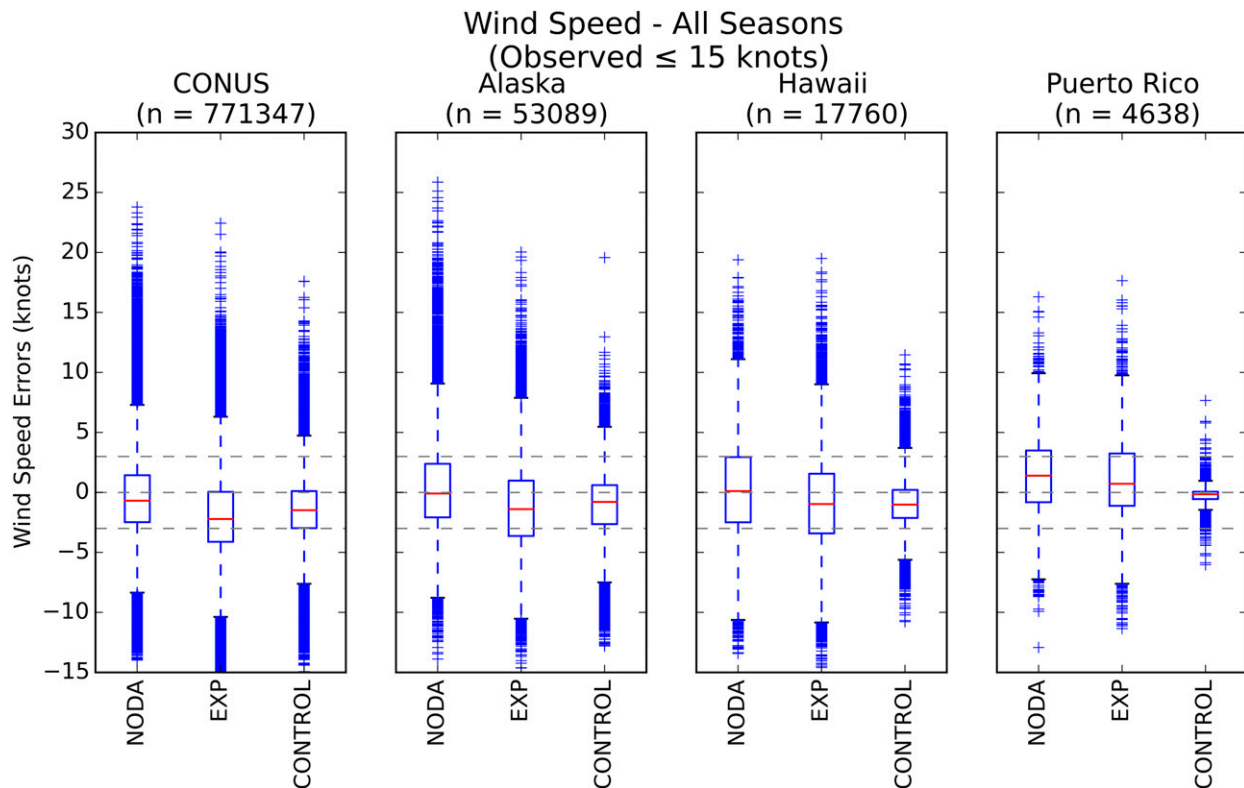


FIG. 10. Boxplots of wind speed errors, showing results aggregated across all retrospective periods by domain, when wind speeds ≤ 15 kt are observed. The observed sample size is provided for each domain. Dashed horizontal lines correspond to variability values provided by the FAA (Table 3).

Puerto Rico, where notable improvements are seen in CONTROL relative to NODA. CONTROL has a narrower distribution than EXP over all domains, most notably over Puerto Rico where EXP shows more negative bias than CONTROL. Overall, the boxplots (Figs. 10, 11) suggest that the RTMA may be a suitable replacement for missing wind speed observations of ≤ 15 kt, but not for stronger wind speeds.

Time series plots of wind speed RMSE over the CONUS, Alaska, Hawaii, and Puerto Rico domains, aggregated across all retrospective periods, are shown in Fig. 12. Over CONUS, the RMSE in NODA is no more than 4 kt for all analysis hours. EXP has a roughly 1 kt higher RMSE than NODA that is statistically significant, showing the wind speed analysis in EXP is degraded relative to NODA in the absence of Part 139 observations while also highlighting the negative influence of nearby mesonet observations. CONTROL offers minor improvements relative to NODA between 0000 and 1500 UTC, with little difference for the remaining hours. Over Alaska, the RMSE in NODA is roughly 4 kt for all analysis hours. CONTROL offers more substantial improvements in RMSE relative to NODA over Alaska of roughly 1 kt and is statistically

significant. Consistent with CONUS, EXP has a somewhat higher RMSE than NODA at most hours, although this difference is smaller in magnitude than seen over CONUS.

Over Hawaii, the RMSE in NODA is slightly greater than that seen over CONUS, with these values generally around or slightly above 4 kt. Consistent with patterns seen over CONUS and Alaska, the RMSE in EXP is somewhat higher than in NODA, or nearly equivalent, at all analysis hours. The RMSE in CONTROL is statistically significantly lower than in NODA and EXP at all analysis hours. Over Puerto Rico, the RMSEs associated with NODA and EXP are less distinguishable and generally not statistically significantly different. The RMSE values seen in CONTROL are lower than in both NODA and EXP by about 2 kt, which is statistically significant.

Time series plots of the wind speed bias shown in Fig. 13 reveal a slight low wind speed bias (0.5–1.5 kt) in NODA over the CONUS domain, with this bias worsened in EXP by roughly 2 kt when all non-Part 139 observations are assimilated. When Part 139 observations are included in the CONTROL the bias is improved relative to EXP by 0.5–1 kt, though still more

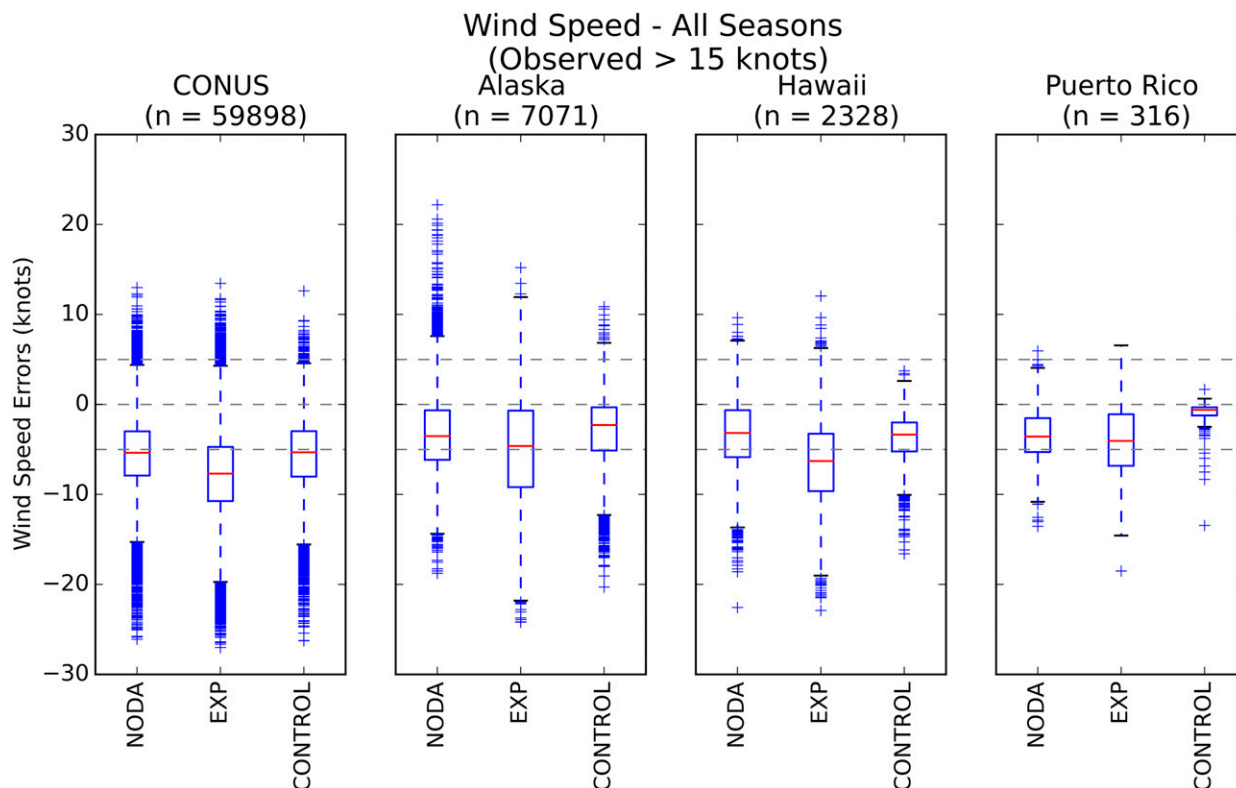


FIG. 11. As in Fig. 10, but when wind speeds > 15 kt are observed.

negative than NODA. This behavior is consistent with that seen in Figs. 11 and 12, which suggests that the assimilation of mesonet observations introduces a negative wind speed bias relative to Part 139 observations. The assimilation of the Part 139 observations clearly decreases this bias, but it is still present when compared to NODA. Over Alaska, the pattern in bias and the differences across experiments mirrors that seen in CONUS. The only exceptions are that the bias is slightly closer to zero and there is less separation between EXP and CONTROL.

The pattern seen in the bias over Hawaii follows the pattern of the RMSE seen in Fig. 12; in other words, the hours with the least (most) negative bias also have the lowest (highest) RMSE values. Over Puerto Rico, the bias in NODA and EXP is negative (up to 1.5 kt) from roughly 1400 to 2200 UTC, while the bias is positive (up to 3.5 kt) outside these times. Despite the diurnal pattern in the bias, little discernible pattern is seen in the RMSE over Puerto Rico (Fig. 12). The transient convective nature of the winds in tropical domains (Hawaii and Puerto Rico) likely contributes to the diurnal variability seen in the RMSE and bias scores on those domains. The time series (Figs. 12, 13) reinforce the conclusions drawn from the boxplots (Figs. 10, 11),

namely that the RTMA may be a suitable replacement for missing wind speed observations of ≤ 15 kt.

While the boxplots in Figs. 10 and 11 and time series in Figs. 12 and 13 provide results aggregated over the entire domain, the station maps of wind speed statistics, such as those shown in Fig. 14 over the southern Plains, facilitate a detailed analysis of the spatial variability across stations. CONTROL has a low wind speed bias at all stations within the area shown in Fig. 14 due to the assimilation of mesonet observations, with this bias in excess of -1 kt at many stations (more negative), consistent with prior analysis seen in the boxplots and time series. A larger low bias (greater than 3 kt) is seen in major cities, such as Dallas (boxed region, Fig. 14), San Antonio, Texas (KSAT), and Tulsa, Oklahoma (KTUL).

To characterize the portion of the RMSE not attributable to the aforementioned low wind speed bias, RMSE statistics are recomputed by removing the influence of the bias at each station through the use of a bias-corrected RMSE (BCRMSE):

$$\text{BCRMSE} = \sqrt{\text{RMSE}^2 - \text{bias}^2}. \quad (3)$$

The BCRMSE in CONTROL is >1.5 kt at all stations within the area depicted in Fig. 14, with values > 3 kt

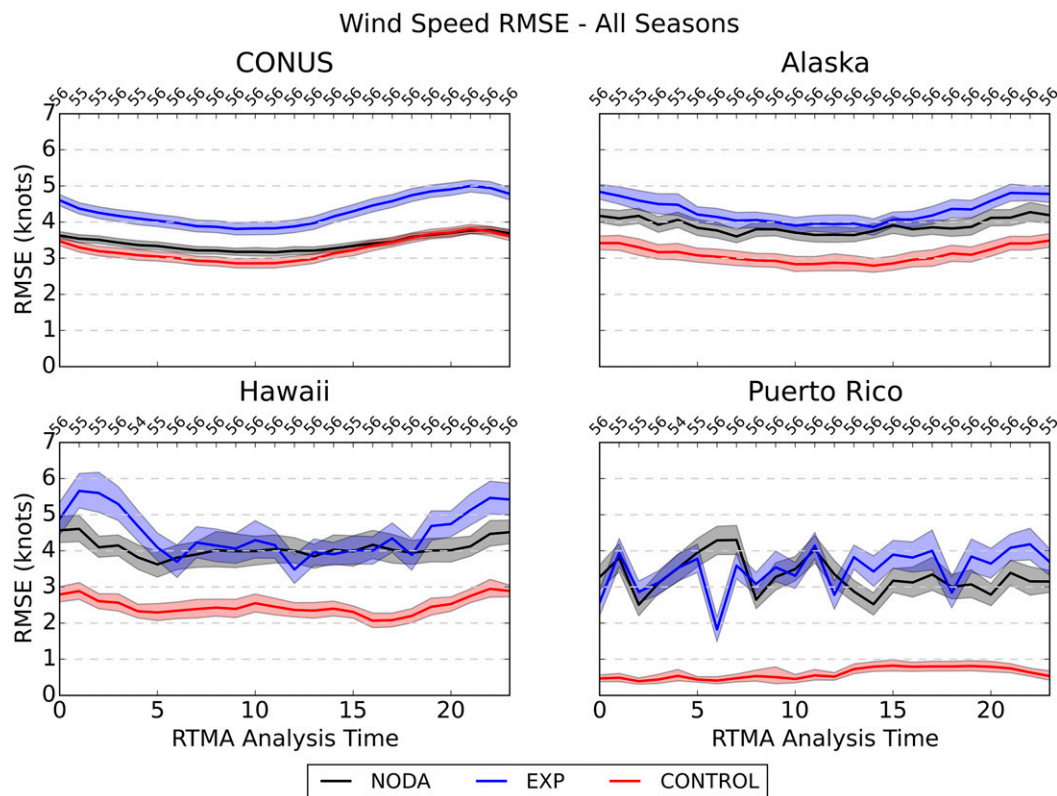


FIG. 12. Time series plots of wind speed RMSE across all retrospective periods by domain. The number of analysis cycles reporting wind speed observations for a given cycle hour is provided above the time series for each domain. The shaded area for each experiment corresponds to bootstrapped ($n = 10\,000$) 95% confidence intervals.

seen in the Dallas, Texas, metro area as well as at KSAT and KTUL. The percent change plots for BCRMSE will generally be positive at most stations since CONTROL will tend to have a lower BCRMSE compared to NODA and EXP [Eq. (2)]. This holds true in NODA at the majority of stations in Fig. 14, with many stations having degradations of greater than 25% (i.e., increased BCRMSE in NODA versus CONTROL). However, there are instances for which NODA has lower BCRMSE scores than CONTROL, such as over the Dallas–Fort Worth metroplex (boxed area), where Dallas Love Field Airport (KDAL), Dallas–Fort Worth International Airport (KDFW), Fort Worth Alliance Airport (KAFW), and Fort Worth Meacham International Airport (KFTW) all exhibit negative percent changes. Similar patterns are seen in other metropolitan areas, such as Austin, Texas (KAUS), KSAT, and KTUL. EXP is degraded relative to CONTROL at all stations in the area shown, with smaller degradations of less than 25% (blue shading) seen in areas where there is likely more widespread coverage of mesonet observations (e.g., Dallas–Fort Worth, Oklahoma City, and Tulsa).

The degradation in wind speed seen in EXP relative to NODA (Figs. 12, 13) is in response to the overwhelming influence of nonstandard mesonet observations. This impact is also seen in Fig. 14 where NODA features lower BCRMSE in metropolitan areas compared to CONTROL owing to the influence of nearby mesonet stations having nonstandard instrument siting. Despite these limitations, the results suggest that the RTMA may be a suitable replacement for missing Part 139 observations when wind speeds ≤ 15 kt are observed, but caution should be exercised for stronger winds > 15 kt.

4. Summary and discussion

Since 2015, EMC has provided temperature pseudo-observations derived from the Real-Time Mesoscale Analysis (RTMA) at airport locations throughout the United States in order to minimize the impacts of missing temperature observations on the nation's aviation industry. This work presents a quality assessment of the RTMA to evaluate the current quality of the 2-m temperature analysis while also providing a basis for

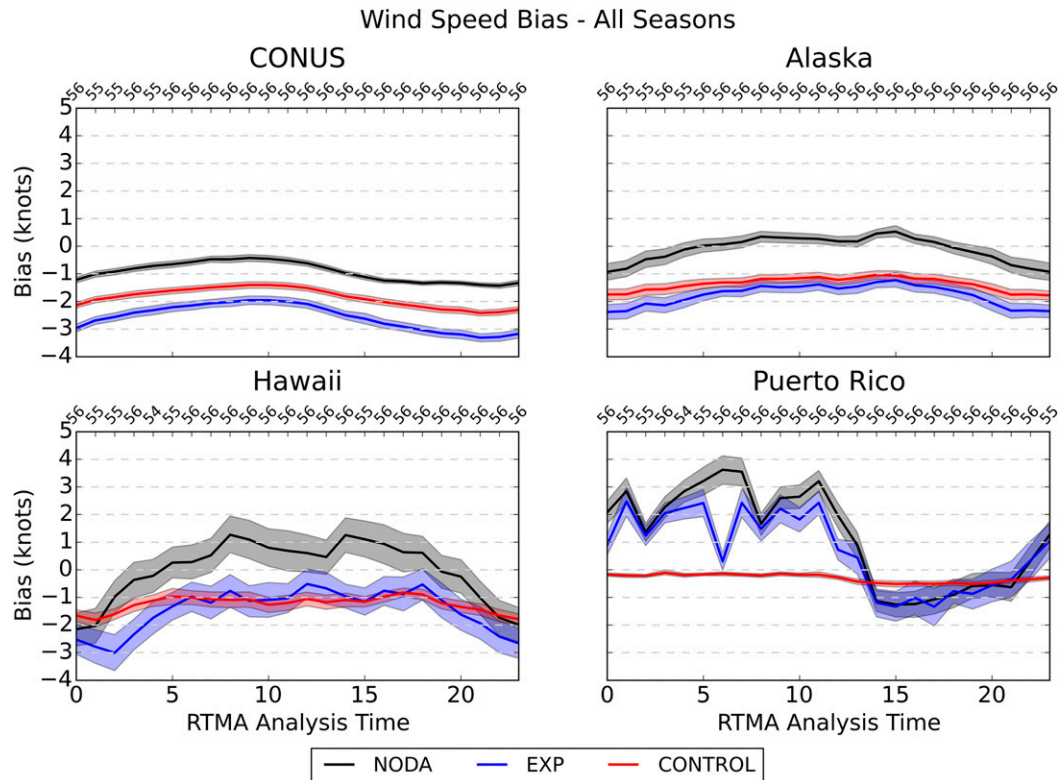


FIG. 13. As in Fig. 12, but for wind speed bias.

extending the pseudo-observation capability for other weather observations, including surface pressure, 10-m wind, ceiling, and visibility.

The quality assessment is performed via retrospective, data-denial experiments that span two weeks for each season (Table 1) and are run over the CONUS, Alaska, Hawaii, and Puerto Rico domains. Three experiments are performed: 1) the CONTROL experiment assimilates all available, quality-controlled observations as in operations, 2) the EXP experiment rejects observations from Part 139 airports, but is otherwise configured similarly to CONTROL, and 3) NODA represents the first guess, or background, fields that are used in the analysis procedure.

Key findings from the quality assessment are summarized in Table 4. When considering ceiling, high analysis outliers are seen in the boxplots aggregated across all domains. For example, there are instances in which the RTMA reports VFR conditions when in actuality LIFR conditions are observed. These outliers will likely preclude the RTMA serving in lieu of missing ceiling observations on all domains, unless future work is able to show that the aforementioned outliers are dominated by a small subset of stations which may allow station-specific refinement.

The results of this study reveal that RTMA struggles to capture low visibility events more so than for ceiling, with EXP failing to capture more than half of the observed LIFR, IFR, and MVFR visibility events. The poor performance noted above suggests that the RTMA should not be considered a viable substitute for missing visibility observations across all domains.

In terms of 2-m temperature, outliers are seen across all domains (Fig. 6), as well as degradations in EXP relative to CONTROL over all domains, most notably over Hawaii and Puerto Rico. Diurnal variations are seen in the RMSE and bias over Puerto Rico, with bias magnitudes higher overall than seen over CONUS and Alaska. The degradations seen in EXP relative to CONTROL are expected and the errors generally remain within the FAA threshold. Taken together, these results suggest that the RTMA is likely a suitable substitute for missing 2-m temperature observations for at least a subset of airports throughout CONUS and Alaska, although future investigation of the aforementioned outliers should be performed to determine if particular stations should be excluded from the airport status list. Furthermore, caution should be exercised when using the RTMA in lieu of missing temperature observations in Hawaii and Puerto Rico if RMSEs approaching 3°F are intolerable.

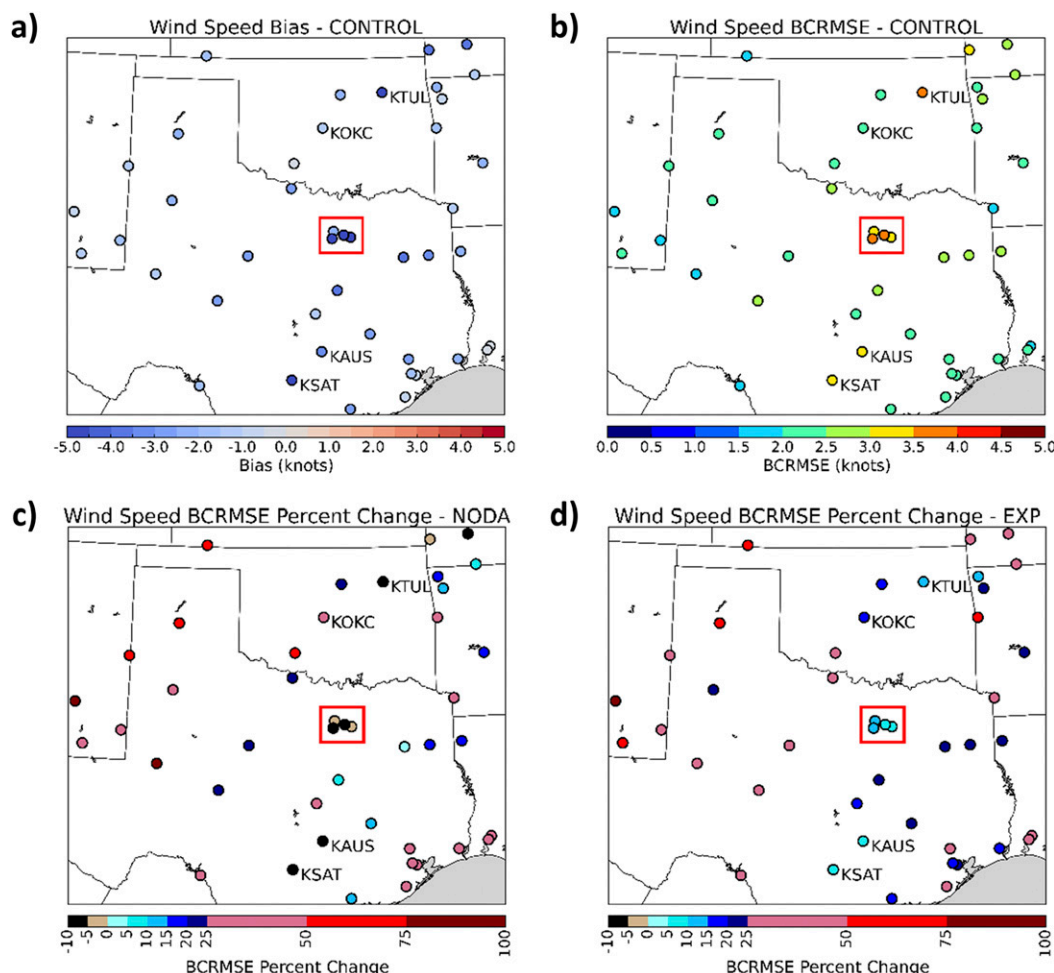


FIG. 14. Station maps of (a) wind speed bias and (b) bias-corrected RMSE for CONTROL, along with bias-corrected RMSE percent change for (c) NODA and (d) EXP. The red box shows the location of four airports in the Dallas-Fort Worth metroplex, namely KDAL, KDFW, KAFW, and KFTW. Note that all stations shown are Part 139 airports.

When considering surface pressure, the boxplots (Fig. 9) revealed a notable low bias in NODA over the Alaska, Hawaii, and Puerto Rico domains, with little to no improvement seen in EXP over Hawaii and Puerto Rico. Little overall bias was seen over CONUS, although local biases likely exist in regions of complex terrain. These results suggest that the RTMA may be a suitable substitute for missing surface pressure observations, provided the terrain is locally flat.

In terms of wind speed, Fig. 10 shows that NODA has a neutral bias over Alaska and Hawaii, a negative bias over CONUS, and a positive bias over Puerto Rico when considering observed wind speeds ≤ 15 kt. The negative bias is exacerbated in EXP on all domains except over Puerto Rico, which is in response to the assimilation of mesonet observations, which, with the exception of university-sponsored mesonets, are noted

to have no uniform standards for instrument siting, are typically most dense in metropolitan areas, and subsequently introduce a low wind speed bias in the analysis (e.g., Fig. 14). Despite this, the median errors fall within the FAA guidelines (Table 3) for wind speeds ≤ 15 kt. For stronger wind speeds (Fig. 11), there is a larger (more negative) low bias in NODA across all domains, with this bias worsened in EXP (owing to the assimilation of mesonet data); CONTROL has a narrower distribution relative to EXP over all domains.

Aggregating wind statistics by time of day shows higher RMSE values in EXP relative to NODA over CONUS and Alaska at all times of the day (Fig. 12). The maximum RMSE scores in EXP are similar across all domains, although a diurnal pattern is seen on both the Hawaii and Puerto Rico domains. Taken together, these results imply that the RTMA may be a suitable

TABLE 4. Summary of key findings for the weather observations considered in this study.

Weather observation	Key findings and recommendations
Ceiling	High analysis outliers seen for all low ceiling categories (LIFR, IFR, and MVFR) Caution should be taken across all domains, unless future work shows the outliers are attributed to a small subset of stations
Visibility	The RTMA struggles to capture low visibility events and should not be used in place of missing visibility observations
2-m temperature	Boxplots reveal outliers on all domains (analysis minus observations $> 2^{\circ}\text{F}$); future investigation may be warranted EXP is degraded relative to CONTROL on all domains, most notably over Hawaii and Puerto Rico RMSEs over Hawaii and Puerto Rico approach $\sim 3^{\circ}\text{F}$ The RTMA is likely a suitable replacement for missing temperature observations at a subset of airports, with expansion possible following examination of outliers
Surface pressure	EXP is degraded relative to CONTROL on all domains, most notably over Hawaii and Puerto Rico The RTMA may be able to serve in lieu of missing surface pressure observations, but caution should be exercised in regions of complex terrain Future analysis could be useful in identifying regional patterns (stations) where the RTMA should not be used
10-m wind speed	NODA tends to have a low bias across all domains for wind speeds > 15 kt, which is worsened in EXP The RTMA may be a viable replacement for missing wind speed observations for wind speeds ≤ 15 kt Stakeholders should exercise caution when using the RTMA for stronger wind speeds

substitute for missing wind speed observations when wind speeds ≤ 15 kt are considered, but stakeholders should be cautious when using the RTMA in cases of stronger wind speeds.

The results of this study represent a highly unlikely scenario, in which all observations from Part 139 airports are denied from the RTMA analysis. In all likelihood, missing observations would be confined to a much smaller subset of airports. Future work could be performed to withhold observations from a small, representative subset of hypothetically missing airports (e.g., coastal, mountainous, urban, and rural regions). The quality of the RTMA could also be assessed on “fair” weather days using the existing retrospective data, which would involve computing sensible weather statistics only for the observations in which the observed ceiling and visibility are within the VFR flight category. Another area of focus is the moisture analysis, which is performed by considering specific humidity rather than dewpoint temperature; thus, future work could leverage these retrospective experiments to evaluate the quality of the derived dewpoint analysis relative to criteria specified by the FAA. Last, as the current 2D version of the RTMA (2DRTMA) will soon be replaced with a 3D version (3DRTMA), it would be worthwhile to perform a similar quality assessment on a prototype version of the 3DRTMA as was performed for the 2DRTMA, and compare the results between the two systems.

Acknowledgments. This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those

of the authors and do not necessarily represent the official policy or position of the FAA. The authors thank Danny Sims, Dr. Alicia Bentley, Dr. Wan-Shu Wu, and three anonymous reviewers for their reviews of an earlier version of this manuscript.

REFERENCES

- Alexander, C. R., G. Manikin, S. Benjamin, S. S. Weygandt, G. DiMego, M. Hu, and T. G. Smirnova, 2015: The High-Resolution Rapid Refresh (HRRR): The operational implementation. *31st Conf. on Environmental Information Processing Technologies*, Phoenix, AZ, Amer. Meteor. Soc., 6A.2, <https://ams.confex.com/ams/95Annual/webprogram/Paper267782.html>.
- Ancell, B. C., C. F. Mass, K. Cook, and B. Colman, 2014: Comparison of surface wind and temperature analyses from an ensemble Kalman filter and the NWS real-time mesoscale analysis system. *Wea. Forecasting*, **29**, 1058–1075, <https://doi.org/10.1175/WAF-D-13-00139.1>.
- AOPA Air Safety Institute, 2018: 27th Joseph T. Nall report: General aviation accidents in 2015. Accessed 30 September 2019, https://www.aopa.org/training-and-safety/air-safety-institute/accident-analysis/joseph-t-nall-report?_ga=2.201873046.645886912.1566841245-261580673.1561470215.
- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Carley, J. R., and Coauthors, 2017: The nested domains of version 4 of the NAM forecast system. *28th Conf. on Weather Analysis and Forecasting/24th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 1204, <https://ams.confex.com/ams/97Annual/webprogram/Paper311148.html>.
- , and Coauthors, 2018: Ongoing upgrades to NOAA’s real time mesoscale analysis system. *29th Conf. on Weather Analysis and Forecasting/25th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 3B.1, <https://ams.confex.com/ams/29WAF25NWP/meetingapp.cgi/Paper/345007>.

- De Pondeva, M. S. F. V., and Coauthors, 2011: The real-time meso-scale analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. Forecasting*, **26**, 593–612, <https://doi.org/10.1175/WAF-D-10-05037.1>.
- FAA, 2016: FAA Order 8900.1, Volume 3 General technical administration, Chapter 26, 3-2073 A.5 Regulatory sources of weather reports—Parts 91k, 121, and 135. Accessed 27 September 2019, http://fsims.faa.gov/WDocs/8900.1/V03%20Tech%20Admin/Chapter%2026/03_026_002.htm.
- , 2019: Part 139 airport certification. Accessed 2 July 2019, https://www.faa.gov/airports/airport_safety/part139_cert/.
- Fujita, T. T., and R. M. Wakimoto, 1982: Effects of miso- and mesoscale obstructions on PAM winds obtained during project NIMROD. *J. Appl. Meteor.*, **21**, 840–858, [https://doi.org/10.1175/1520-0450\(1982\)021<0840:EOMAMO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1982)021<0840:EOMAMO>2.0.CO;2).
- Gerth, J. J., 2018: Shining light on sky cover during a total solar eclipse. *J. Appl. Remote Sens.*, **12**, 020501, <https://doi.org/10.1117/1.JRS.12.020501>.
- Gilbert, G. K., 1884: Finley's tornado predictions. *Amer. Meteor. J.*, **1**, 166–172.
- Gilbert, K. K., and Coauthors, 2016: The National Blend of Global Models, version one. *23rd Conf. on Probability and Statistics in the Atmospheric Sciences*, New Orleans, LA, Amer. Meteor. Soc., 1.3, <https://ams.confex.com/ams/96Annual/webprogram/Paper285973.html>.
- Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195–202, <https://doi.org/10.1175/BAMS-84-2-195>.
- Horel, J., and B. Colman, 2005: Real-time and retrospective mesoscale objective analyses. *Bull. Amer. Meteor. Soc.*, **86**, 1477–1480, <https://doi.org/10.1175/BAMS-86-10-1477>.
- Knopfmeier, K. H., and D. J. Stensrud, 2013: Influence of mesonet observations on the accuracy of surface analyses generated by an ensemble Kalman filter. *Wea. Forecasting*, **28**, 815–841, <https://doi.org/10.1175/WAF-D-12-00078.1>.
- Lawless, A. S., 2010: A note on the analysis error associated with 3D-FGAT. *Quart. J. Roy. Meteor. Soc.*, **136**, 1094–1098, <https://doi.org/10.1002/qj.619>.
- McGraw, P. J., 2015: Letter from Airlines for America to Margaret Gilligan (FAA). 2 pp.
- OPSNET, 2019: Monthly report for December 2018. OPSNET, 10 pp.
- Pondeca, M., and Coauthors, 2015: The 2014/2015 projected expansion of NCEP's RTMA and URMA. *19th Conf. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Phoenix, AZ, Amer. Meteor. Soc., 4.3, <https://ams.confex.com/ams/95Annual/webprogram/Paper265321.html>.
- Purser, R. J., 2011: Mathematical principles of the construction and characterization of a parameterized family of Gaussian mixture distributions suitable to serve as models for the probability distributions of measurement errors in nonlinear quality control. NOAA/NCEP Office Note 468, 42 pp.
- , 2018: Convenient parameterizations of super-logistic probability models of effective observation error. NOAA/NCEP Office Note 495, 8 pp, <https://doi.org/10.25923/kvmz-vf34>.
- Rogers, E., and Coauthors, 2017: Mesoscale modeling development at the National Centers for Environmental Prediction: Version 4 of the NAM forecast system and scenarios for the evolution to a high-resolution ensemble forecast system. *28th Conf. on Weather Analysis and Forecasting/24th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 3B.4, <https://ams.confex.com/ams/97Annual/webprogram/Paper311212.html>.
- San Francisco International Airport, 2010: Weather and operations at SFO: A primer for the media. San Francisco International Airport, 4 pp., https://media.flysfo.com/media/sfo/media/weather-operations-primer_0.pdf.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Tyndall, D. P., and J. D. Horel, 2013: Impacts of mesonet observations on meteorological surface analyses. *Wea. Forecasting*, **28**, 254–269, <https://doi.org/10.1175/WAF-D-12-00027.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, [https://doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).
- Yang, R., R. J. Purser, J. R. Carley, M. Pondeva, Y. Zhu, and W. S. Wu, 2018: Applying a general nonlinear transformation to the analysis of surface visibility and cloud ceiling height. *25th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 51, <https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345402.html>.
- , —, —, —, —, S. Levine, and W. S. Wu, 2019: Applying a nonlinear transformation to the analysis of surface visibility and cloud ceiling height. 2 pp., http://bluebook.meteoinfo.ru/uploads/2019/docs/01_Yang_Runhua_RTMA_sensitivity.pdf.