Research Article

Assimilation of Aircraft Observations in High-Resolution Mesoscale Modeling

Brian P. Reen [1] **and Robert E. Dumais Jr.** [2]

[1]*Battlefield Environment Division, Army Research Laboratory, Adelphi, MD, USA*
[2]*Battlefield Environment Division, Army Research Laboratory, White Sands, NM, USA*

Correspondence should be addressed to Brian P. Reen; brian.p.reen.civ@mail.mil

Aircraft-based observations are a promising source of above-surface observations for assimilation into mesoscale model simulations. The Tropospheric Airborne Meteorological Data Reporting (TAMDAR) observations have potential advantages over some other aircraft observations including the presence of water vapor observations. The impact of assimilating TAMDAR observations via observation nudging in 1 km horizontal grid spacing Weather Research and Forecasting model simulations is evaluated using five cases centered over California. Overall, the impact of assimilating the observations is mixed, with the layer with the greatest benefit being above the surface in the lowest 1000 m above ground level and the variable showing the most consistent benefit being temperature. Varying the nudging configuration demonstrates the sensitivity of the results to details of the assimilation, but does not clearly demonstrate the superiority of a specific configuration.

1. Introduction

Aircraft observations can provide observations with spatiotemporal coverage not available from standard in situ measurements, and thus careful assimilation of these observations can be used to improve mesoscale model forecasts. Above-surface in situ meteorological observations are much less dense and much less frequent than surface observations. For example, radiosondes are generally available only at widely spaced locations (averaging 315 km apart over the United States [1]) and only twice a day. In contrast, weather observations taken by commercial aircraft can provide denser and more frequent observations than radiosondes.

While aircraft-based observations have been around for some time [2], referred to as aircraft meteorological data relay (AMDAR) reports or Aircraft Communications Addressing and Reporting System (ACARS), more recently a network of aircraft-based observations called Tropospheric Airborne Meteorological Data Reporting (TAMDAR) [3] has been introduced. One advantage of TAMDAR over standard AMDAR observations is the presence of TAMDAR sensors on smaller aircraft that fly into a larger number of airports and

cruise at a lower altitude; this allows for better observational coverage at lower altitudes [4]. A modified TAMDAR sensor package has been applied to an unmanned aerial system (UAS [5]). UAS weather observations could significantly enhance observing capabilities, especially in environments where observation density is very low. Another advantage of TAMDAR has been the reporting of humidity, which has not usually been included in standard AMDAR observations [4], although a subset now report moisture [6]. Additionally, while with AMDAR observations height generally must be estimated based on the pressure of the observation (e.g., [7]), TAMDAR observations include GPS-based height.

Various studies have demonstrated the value of assimilating aircraft observations (e.g., [8–11]) and more recently studies have demonstrated the value of assimilating TAMDAR aircraft observations in particular (e.g., [4, 12–14]). However, investigation of the value of assimilating TAMDAR observations in high-resolution (~1 km horizontal grid spacing) simulations has been limited.

Moninger et al. [4] used three-dimensional variational (3DVAR) analyses with TAMDAR data to initialize 20 km

horizontal grid spacing Rapid Update Cycle model simulations and found assimilation of TAMDAR data improved temperature, moisture, and wind forecasts. Gao et al. [12] also applied TAMDAR data at 20 km horizontal grid spacing with 3DVAR but used the Advanced Research version of the Weather Research and Forecasting model (WRF-ARW [15]). Gao et al. [12] demonstrated assimilating TAMDAR data improved model forecasts and also showed that TAMDAR errors are comparable to radiosondes for temperature and in the winter for moisture, smaller than radiosondes in the summer for moisture and for winds stronger than $15\,\text{ms}^{-1}$, and larger than radiosondes for winds less than $15\,\text{ms}^{-1}$. However, Jacobs et al. [16] subsequently demonstrated corrections that improve the quality of TAMDAR wind observations from a small subset of TAMDAR aircraft with larger wind direction errors. Wang and Huang [17] also used 3DVAR to assimilate TAMDAR observations in WRF-ARW at approximately the same horizontal grid spacing (18 km), but applied it to a hurricane and found improvements in track forecast.

Observation nudging has also been used to assimilate TAMDAR observations in WRF-ARW. Observation nudging uses the difference between observations and the model to create nonphysical terms which are added to the model tendency equations in order to gradually "nudge" the model towards the observations (e.g., [18]). Observation nudging generally centers the application of an observation on the time at which the observation was taken; this allows the effects of observations to be based on how the model compares to the observation at the time of the observation. Zhang et al. [14] assimilated TAMDAR observations during a 6 h preforecast to a 24 h forecast using 12 and 4 km nests. Evaluations over a 6-day period indicated that within the layer where TAMDAR shows the most impact (400–600 hPa) the greatest improvements are in temperature and moisture forecasts, whereas wind speed was improved on the 12 km domain but degraded on the 4 km domain.

While Jonassen et al. [19] did not assimilate TAMDAR observations, they assimilated soundings from a UAS using observation nudging in 9, 3, and 1 km WRF-ARW simulations for two cases. They found UAS observations improved the model's representation of the sea breeze. They also noted limited sensitivity to varying the observation nudging configuration.

We assimilate TAMDAR observations on a 1 km WRF-ARW domain using observation nudging to investigate the potential value of TAMDAR in high-resolution simulations and explore the issues associated with assimilating single-level above-surface observations. Experiments are performed that explore the sensitivity of the results to observation nudging parameters, since the high resolution of the simulations and single-level above-surface nature of the observations may limit the applicability of parameters determined from past nudging studies. We do not attempt to find universally optimal nudging parameters, but rather demonstrate the effects of limited modification of the parameters to better demonstrate the potential benefits of TAMDAR data. In Section 2 we describe the model and its configuration, and in Section 3 we describe the cases studied. The methodology

used to incorporate observations is discussed in Section 4, while the experimental design is outlined in Section 5. Sections 6 and 7 present the results and discussion of the results, respectively, and the summary and conclusions are presented in Section 8.

2. Model Description and Configuration

WRF-ARW version 3.6.1 was configured with 27, 9, 3, and 1 km horizontal grid spacing nested domains centered over San Francisco, California (Figure 1) using 56 vertical layers with the lowest prognostic level for fields such as air temperature at ≈12 m AGL. The use of a 1 km domain allows better resolution of the relatively complex terrain in this region and thus should better resolve terrain-influenced meteorological features. The model was integrated for 24 hours from 12 UTC to 12 UTC for the five case days described in Section 3.

The initial conditions for all four domains and the time-dependent lateral boundary conditions for the outermost domain were created using Global Forecast System (GFS) 0.5-degree horizontal resolution output. Sea surface temperature was specified using the Real-Time Global Sea Surface Temperature [20] from the Marine Modeling and Analysis Branch of the Environmental Modeling Center at the National Centers for Environmental Prediction, which has one-twelfth-degree horizontal grid spacing. The National Weather Service's National Operational Hydrologic Remote Sensing Center (NOHRSC) Snow Data Assimilation System (SNODAS [21]) 1 km snow fields were used where available; elsewhere GFS snow fields were used as a first guess but adjusted via National Snow and Ice Data Center (NSIDC) daily 4 km snow cover fields from the IMS (Interactive Multisensor Snow and Ice Mapping System) Daily Northern Hemisphere Snow and Ice Analysis [22]. In particular, outside of the area covered by SNODAS, the GFS snow depth was used as a first guess, but in locations where IMS indicated no snow the snow depth was set to zero, and in locations where IMS indicated snow but GFS did not indicate snow the snow depth was set to a default value.

The Mellor-Yamada-Janjić (MYJ) scheme [23] is applied to parameterize the atmospheric boundary layer, modified as in Reen et al. [24]. Microphysics are represented using the Thompson microphysics scheme [25] and on the coarser two domains (27 and 9 km) the Kain-Fritsch cumulus parameterization [26] is invoked. For longwave radiation, the Rapid Radiative Transfer Model [27] is used and for shortwave the Dudhia scheme [28] is used. The Noah land surface model [29] represents land surface processes. This combination of physics parameterizations is very similar to that used in Reen et al. [24] which applied a 9 km WRF domain centered over southern California; the only difference is the use of the Thompson microphysics scheme here. All of the physics parameterizations chosen for this study have been widely used in other studies, with the exception of the modifications to the MYJ scheme which have however been applied previously over this region [24] and in other studies [30, 31].

Observation nudging towards TAMDAR observations is applied to the 1 km domain during the first 6 h of the
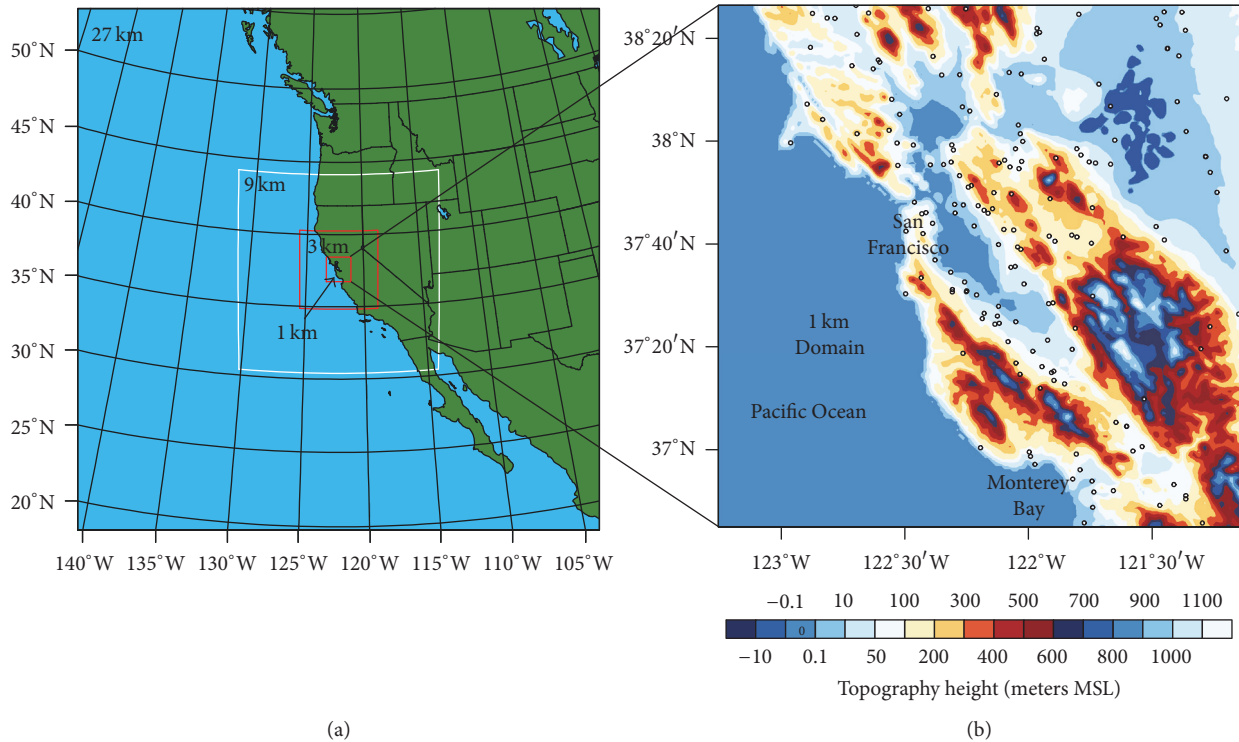
(a)

(b)

FIGURE 1: (a) Placement of the four nested WRF-ARW domains, and (b) the topography of the 1 km horizontal grid spacing domain with locations of land-based surface observations during the 1–6 h forecast on 7 February indicated by black circles with a white circle inset.

simulation for certain experiments. Additional details regarding the quality control applied to the TAMDAR observations are described in Section 4.1. Observation nudging is described in more detail in Section 4.2, and how the application of observation nudging differs among experiments is described in Section 5.

3. Case Description

3.1. General Overview. Five case days were simulated over the southwestern United States; simulations started at 12 UTC on each of the five days (7 February, 9 February, 16 February, 1 March, and 5 March 2012) and ended at 12 UTC on the following day. Since these are the same case days used in Reen et al. [24], the case description from that study is used in this section with minor modifications. Within the five case days, there are days with and without strong synoptic forcing. Widespread precipitation occurred in the region on 7 February due to a trough moving onshore. More quiescent weather occurred during the 9 February case with a 500 hPa ridge centered over central California at 12 UTC. On 16 February, there was an upper-level low near the California/Arizona border with Mexico at 12 UTC, bringing precipitation to that area. The area of low pressure and the associated precipitation moved off to the south and then east as the day progressed. For 1 March, a weak shortwave trough resulted in precipitation in northern California at the beginning of the period that spread to Nevada, and then moved southward and decreased in coverage. Widespread

high-level cloudiness occurred during the 5 March case due to weak upper-level low pressure but this was accompanied by only limited precipitation.

3.2. TAMDAR Observations. The distribution of TAMDAR observations available over the 1 km WRF-ARW domain during the assimilation period is shown in Figures 2 and 3. Figure 2 shows the horizontal distribution of such observations. The observations mostly fall in a northwest-to-southeast oriented strip through the middle of the domain. The time-height distribution (Figure 3) of the observations indicates the lack of observations in the beginning of the assimilation period (12–14 UTC) for all case days; this is not surprising given that this is 04–06 local time (PST) and there are likely limited flights during this period. The height distribution indicates that all of the TAMDAR observations during the assimilation period were between 0 and 8000 m AGL, and the observations appear most concentrated in the lowest 1000 m AGL. The number of observations varies between 78 (1 March) and 125 (9 February) among the case days.

4. Methodology

4.1. Observation Quality Control. Obsgrid is part of the WRF-ARW software suite and is used here to provide quality control of both the TAMDAR observations used for observation nudging and all of the other observations used for verification except for the stage IV precipitation data. Obsgrid checks
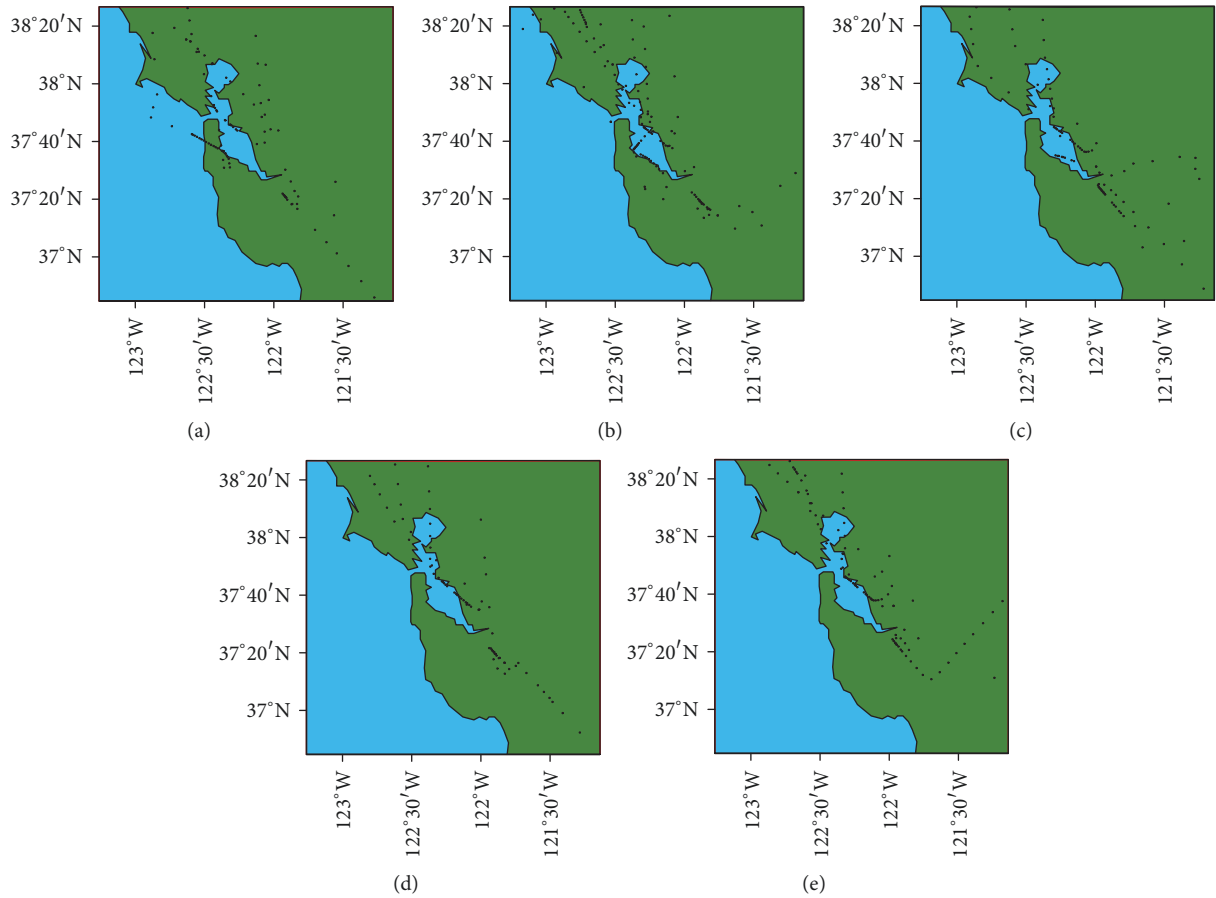
FIGURE 2: Horizontal distribution of TAMDAR observations over the 1 km WRF-ARW domain during the assimilation period (12–18 UTC) for the following dates in 2012: (a) 7 February, (b) 9 February, (c) 16 February, (d) 1 March, and (e) 5 March.

observations for gross errors, compares the observations to nearby observations (buddy check), and compares the observations to a background field (here, GFS).

Methods used by Obsgrid to quality-control surface observations and profiles of observations (e.g., radiosondes) required modification for application to single-level above-surface observations. Due to the horizontal movement of aircraft, TAMDAR observations were often processed in this study as single-level above-surface observations rather than profiles. In order to compare profiles such as radiosondes against a background field (e.g., GFS), Obsgrid vertically interpolates the observation to the available GFS pressure levels. However, since interpolation using a single observation is not possible, the standard version of Obsgrid at the time of this research employed other techniques to quality-control single-level above-surface observations that it recognized. If the pressure of the observation was near enough to the closest background field pressure level, it would modify the observation and apply it to the closest background field pressure level. In particular, it would adjust the temperature using a standard lapse rate, use the winds without modification, and discard moisture. To more effectively use TAMDAR data, we modified Obsgrid to allow single-level observations to be directly quality-controlled against the nearest pressure level

if that level is "close enough" (as determined by a user-set maximum pressure difference). Additionally, a modified version of the WRF-ARW preprocessor Ungrib (provided by Cindy Bruyere of the National Center for Atmospheric Research) was used to vertically interpolate the first-guess field to additional pressure levels prior to being ingested by Obsgrid in order to decrease the vertical separation between observations and the closest first-guess field. These changes allow TAMDAR data to be applied directly at the pressure they are observed at, remove the assumption that the atmosphere is following a standard lapse rate, remove the assumption that wind is constant with height, and permit the TAMDAR moisture fields to be assimilated. Note that the Obsgrid modifications were included in the standard releases of Obsgrid starting in V3.8, while the code to allow pre-Obsgrid interpolation to additional vertical levels is included in Ungrib starting in V3.9.

For verification purposes, in addition to TAMDAR data, observations were obtained from the Meteorological Assimilation Data Ingest System (MADIS; https://madis.ncep.noaa.gov). Specifically, standard and mesonet surface observations, ACARS observations, maritime observations, and rawinsondes were used. In addition to MADIS quality control flags and Obsgrid quality control procedures,
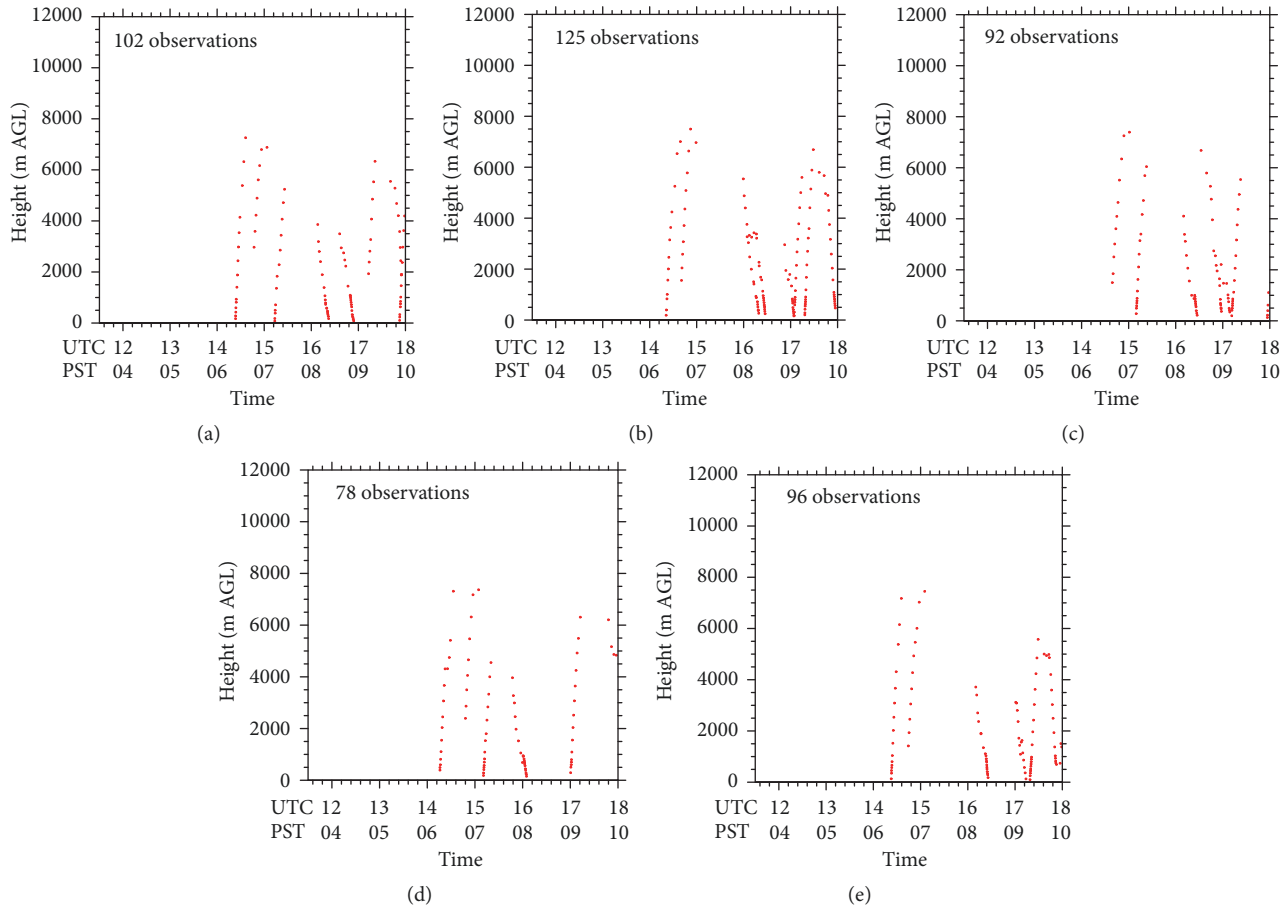
FIGURE 3: Time-height distribution of TAMDAR observations over the 1 km WRF-ARW domain during the assimilation period (12–18 UTC) for the following dates in 2012: (a) 7 February, (b) 9 February, (c) 16 February, (d) 1 March, and (e) 5 March.

MADIS mesonet observations were also quality-controlled by employing use/reject lists used in a test version of the Real-Time Mesoscale Analysis [32].

4.2. Observation Nudging.

TAMDAR observations are assimilated during a 6 h preforecast (12–18 UTC) using observation nudging [18, 24, 33]. For each TAMDAR observation of temperature, moisture, and wind, the difference between the WRF-ARW value and the TAMDAR value (the innovation) is calculated throughout the time period when the model is sufficiently close to the observation time. This time-evolving innovation is used to create a tendency term to nudge the model towards the observation over a region surrounding the observation. In addition to the innovation, the tendency term from nudging includes terms based on the difference between the location and time of the observation and the location and time at which the innovation is being applied, as well as the nudging strength (which determines the e-folding time of the innovation in the hypothetical circumstance where the only tendency term for the variable is nudging). For more details on observation nudging as implemented in WRF-ARW, see Reen [34]. During the first hour of the forecast (18–19 UTC), observations during the preforecast period continue to be assimilated, but the overall nudging weighting gradually decreases to zero during this hour.

The spreading of the innovation for above-surface single-level observations includes a horizontal and a vertical component. Vertically, in standard WRF-ARW the innovation is applied between 75 hPa below the observation and 75 hPa above the observation; the weighting decreases linearly with the difference between the current pressure and the pressure of the observation in this layer. Section 5 will detail how the depth of this layer will vary among the experiments. Horizontally, the innovation is spread in pressure-space, but if the model determines the observation to be above the atmospheric boundary layer, the innovation will not be applied to any points the model determines to be below the atmospheric boundary layer (and vice versa). The user specifies a horizontal radius of influence (HROI) which is valid at the surface, but the HROI applied increases with decreasing pressure to twice the specified HROI at 500 hPa. The user-specified HROI will be varied among the experiments, as described in Section 5.

The innovation for a given TAMDAR observation is calculated starting 1.5 h prior to the observation and continuing until 1.5 h after the observation. The temporal weighting

ramps up during the first 0.75 h and ramps down during the last 0.75 h of the 3.0 h time window for that observation. An overall observation nudging weight of $8 \times 10^{-4}\,\mathrm{s}^{-1}$ is applied here. A modification to mitigate overdrying by observation nudging developed in earlier research [24] was also applied.

5. Experimental Design

In order to test the potential impacts of assimilating TAMDAR observations and to form a preliminary understanding of the sensitivity of the impacts to the data assimilation configuration, four experiments were used for the 5 case days. The control experiment (Exp. Control) assimilated no observations. The other three experiments assimilated only TAMDAR observations and assimilated those only on the 1 km domain and for the period 12–18 UTC. The difference among these three experiments is in the horizontal and vertical radii of influence (ROIs) used to spread innovations calculated based on the difference between TAMDAR observations and the WRF prediction. Exp. H45V75 uses an HROI of 45 km and the default vertical influence of 75 hPa, and Exp. H45V200 uses an HROI of 45 km but a vertical influence of 200 hPa, while Exp. H15V75 uses an HROI of 15 km and the default vertical influence of 75 hPa. The initial conditions (12 UTC) for all four experiments are identical because the TAMDAR data is assimilated via observation nudging which applies the observations to the model while it integrates through a preforecast period (here 12–18 UTC).

6. Results

To evaluate the potential value of assimilating TAMDAR observations, model forecasts from the four experiments were compared above the surface against ACARS, TAMDAR, and rawinsonde observations and at the surface against MADIS standard and mesonet surface observations in addition to maritime observations; additionally precipitation analyses were also used for verification. Evaluation excluded the preforecast time period during which TAMDAR observations were assimilated and so no TAMDAR observations that were assimilated were used for the verification. Since the number of observations available at a given time or height can vary significantly among case days for aircraft observations, for fields other than precipitation the statistics from the five case days were combined by weighting the contribution of each case day by the number of observations it contributed to the given statistic.

The mean absolute error (MAE) of the nudging experiments was compared to the MAE of the control experiment to determine the impact of the nudging. MAE is defined here for a set of N observations as

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| \mathrm{model}_i - \mathrm{obs}_i \right|, \qquad (1)$$

where obs_i is the ith observation of some quantity (e.g., temperature) and model_i is the model prediction of this quantity interpolated to the location of the ith observation. Unlike mean error (bias), MAE is not susceptible to compensating

negative and positive errors and thus can better represent the performance of the model.

Vertical profiles comparing the performance of each experiment to Exp. Control during the first 6 h of the model forecast (19–00 UTC) are shown in Figure 4. Temperature MAE (Figure 4(a)) is improved by assimilating TAMDAR observations by 0.1–0.2 K in the lowest 1000 m AGL. Above this layer, the effect of assimilating TAMDAR observations overall has a limited positive impact. At most levels, the experiment with the smallest horizontal and vertical radii of influence (Exp. H15V75) showed the least improvement from assimilation. Note that the number of observations available for verification decreases from 1333 in the 0–1000 m AGL layer to 16 in the 11000–12000 m AGL layer. For dewpoint (Figure 4(b)), the overall impact of assimilating TAMDAR observations is mixed and small (<=0.1 K for most experiments and levels). In general, larger ROIs result in larger impacts (both positive and negative), as would be expected. The largest positive impact is in the lowest layer (0–1000 m AGL), where all three nudging experiments indicate dewpoint MAE improved > 0.1 K through assimilation of TAMDAR. This layer has more than double the observations available for verification (533) than any of the other layers. There are substantially fewer moisture observations available for above-surface verification of dewpoint (1566) compared to temperature (6253) during this period; this is not surprising given that ACARS data often do not include moisture observations. The effect of assimilating TAMDAR observations on wind forecasts varies substantially with height. All three nudging experiments show a slight improvement in the lowest layer ($\approx 0.1\,\mathrm{ms}^{-1}$) and a 0.1–0.2 ms^{-1} improvement in the 3000–5000 m AGL layer. However, a degradation of 0.1 ms^{-1} (or slightly more) is seen in one experiment (H45V75) at 2000–3000 m AGL and two experiments (H45V37 and H45V200) at 8000–9000 m AGL. Assimilating TAMDAR observations overall degrades model wind direction forecasts, although overall differences among experiments are quite small at or above 5000 m AGL. The largest degradation is in the 1000–2000 m AGL layer where H15V75 and H45V200 increase MAE by 2-3 degrees and H45V75 by 6 degrees. As with wind speed (but in contrast to temperature and dewpoint), the very largest MAE differences from Control are in H45V75 rather than H45V200, even though the latter spreads the influence of observations over a deeper layer.

The root mean square error (RMSE) provides similar information to MAE, but because RMSE squares the error it is more sensitive to large errors. RMSE is defined here for a set of N observations as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\mathrm{model}_i - \mathrm{obs}_i \right)^2}. \qquad (2)$$

Vertical profiles of RMSE (not shown) parallel to those of MAE in Figure 4 are generally very similar to the MAE profiles in terms of the relationship between errors among experiments and vertical levels. One area where differences are seen is in low-level dewpoint. At 0–1000 m AGL, while
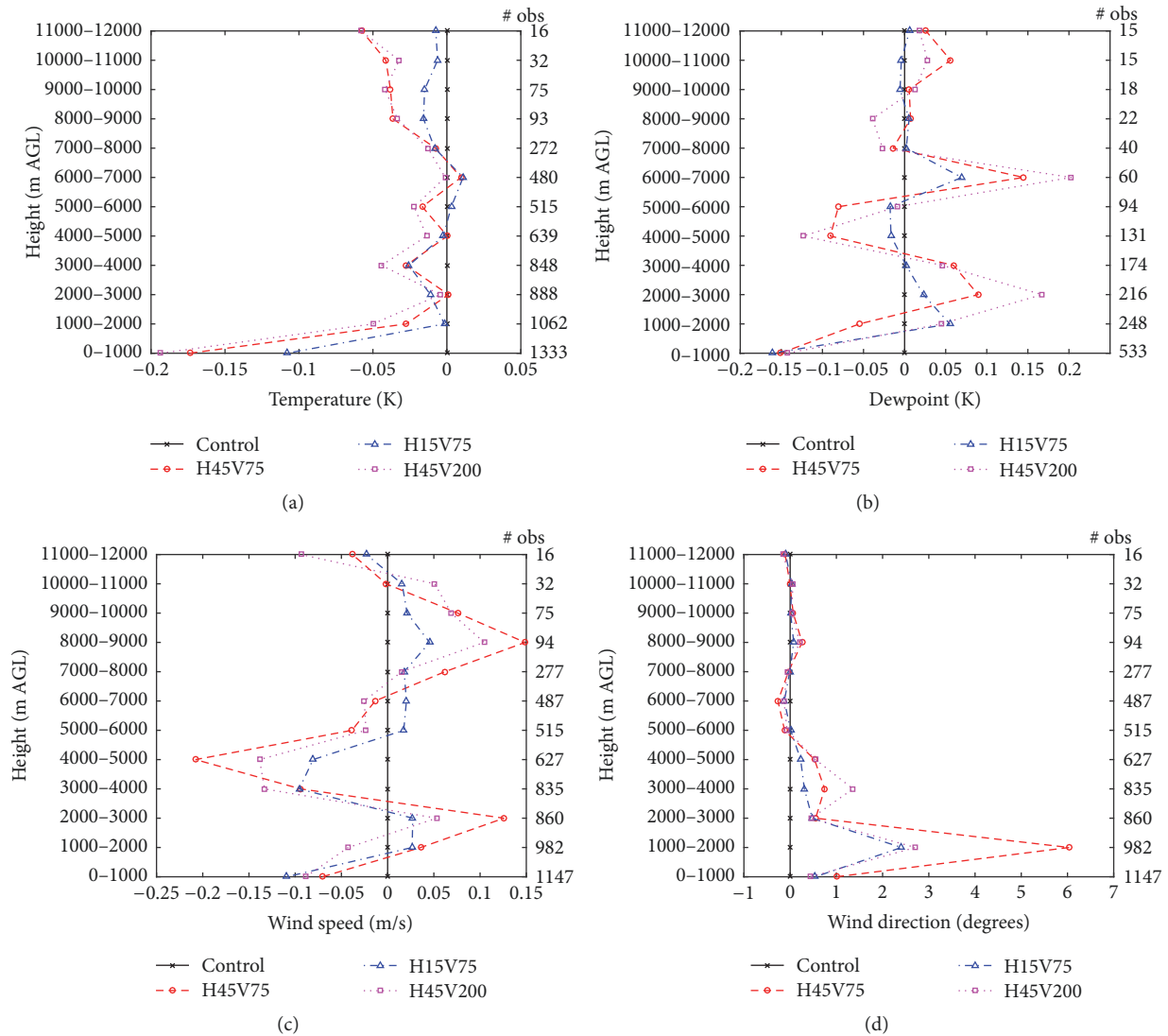
FIGURE 4: Vertical profiles of the difference between MAE for each experiment compared to the MAE for Exp. Control for the 1–6 h forecast (19 UTC) as compared to above-surface observations and calculated over observations from all five case days for (a) temperature, (b) dewpoint, (c) wind speed, and (d) wind direction. Values to the left of the vertical line indicate an experiment performed better than the Exp. Control. The number of observations used in the verification to create each vertical layer is noted along the right edge of each plot.

all three experiments perform very similarly in MAE (Figure 4(b)), in RMSE the ranking among the experiments remains the same but the difference between Exp. H15V75 and the other nudging experiments increases. This may indicate that one or more observations where Exp. H45V75 and H45V200 performed particularly poorly were improved by Exp. H15V75. At 1000–2000 m AGL, the slight degradation in MAE due to nudging (Figure 4(b)) for Exp. H45V200 and H15V75 is in RMSE a slight improvement for Exp. H45V200 and no change for Exp. H15V75. This may indicate that some observations that Exp. Control predicted particularly poorly were notably improved by nudging. For temperature 1000–2000 m AGL in Exp. H45V200 the MAE indicates slight improvement (Figure 4(c)) while RMSE indicates no change. This suggests that the difference between predictions of Exp.

H45V200 and certain observations are relatively large and dominate the RMSE more than the MAE.

The 0–1000 m AGL layer has the largest number of observations, and for temperature, dewpoint, and wind speed this is a layer where assimilating TAMDAR observations improves the 1–6 h forecast. The distribution among the model-observation pairs used in verification of the sizes of improvements and degradations from the assimilation of TAMDAR observations is shown for Exp. H45V75 in Figure 5. Note that each plot in Figure 5 shows the distribution that created a single point in Figure 4 (the 0–1000 m AGL value for Exp. H45V75). While overall temperature in this layer improves ≈0.2 K with the assimilation of TAMDAR (comparing Exp. H45V75 with Exp. Control), the distribution of the improvements and degradations (Figure 5(a))
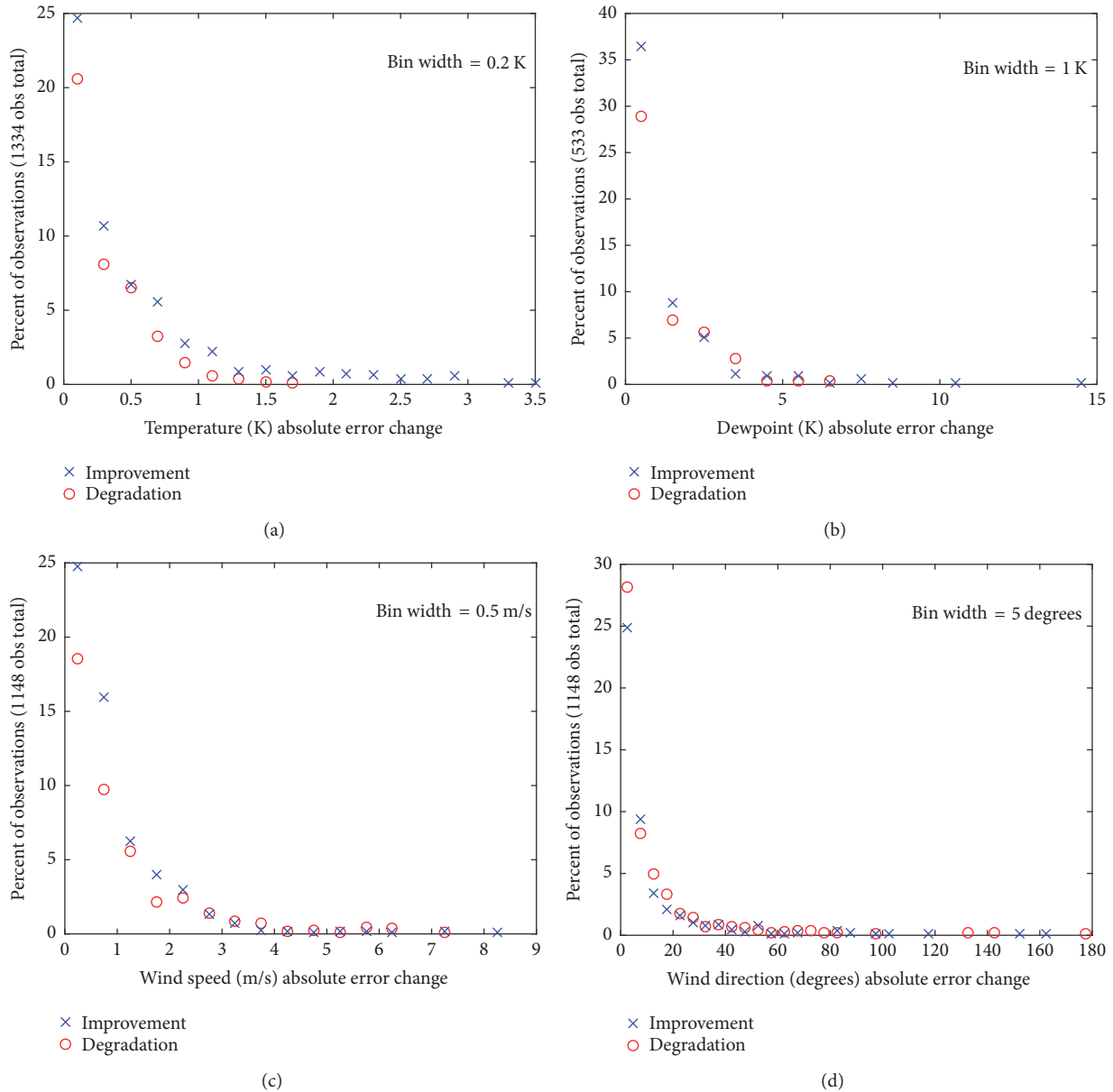
(a)

(b)

(c)

(d)

FIGURE 5: Distribution of the differences between the absolute value of model error in Exp. H45V75 and the absolute value of model error of Exp. Control for (a) temperature, (b) dewpoint, (c) wind speed, and (d) wind direction as verified against above-surface observations during the 1–6 h forecast (19 UTC). Bins representing model-observation pairs where the absolute value of model error in Exp. H45V75 is smaller than that in Exp. Control are denoted by "Improvement," whereas pairs where Exp. H45V75 has the larger error are denoted by "Degradation" but are plotted on the positive x-axis to simplify comparison against mirror "Improvement" bins.

indicates model-observations pairs with improvements as large as ≈3.5 K and degradations as large as ≈1.7 K. The forecast error changes little (<0.2 K) for ≈45% of observations, but in every bin there are more model-observation pairs that are improved than are degraded. For dewpoint (Figure 5(b)), ≈65% of model-observation pairs have forecast errors that change by no more than 1.0 K. While there are more improvements than degradations in most bins, the 2-3 K and 3-4 K bins have more degradations than improvements. The small number of changes > 7 K are improvements. In wind speed changes (Figure 5(c)), improvements

outnumber degradations for all bins representing changes up to 2.5 ms^{-1}, but for the limited number of model-observation pairs with larger changes, the results are more mixed. In regard to changes in the model error in wind direction caused by assimilating TAMDAR observations (Figure 5(d)), degradations outnumber improvements in three out of four of the bins < 20 degrees. There are a limited number of model-observation pairs with very large differences in wind direction caused by assimilating TAMDAR. Removing the 5.4% of model-observation pairs with improvements or degradations greater than 45 degrees decreases the mean degradation
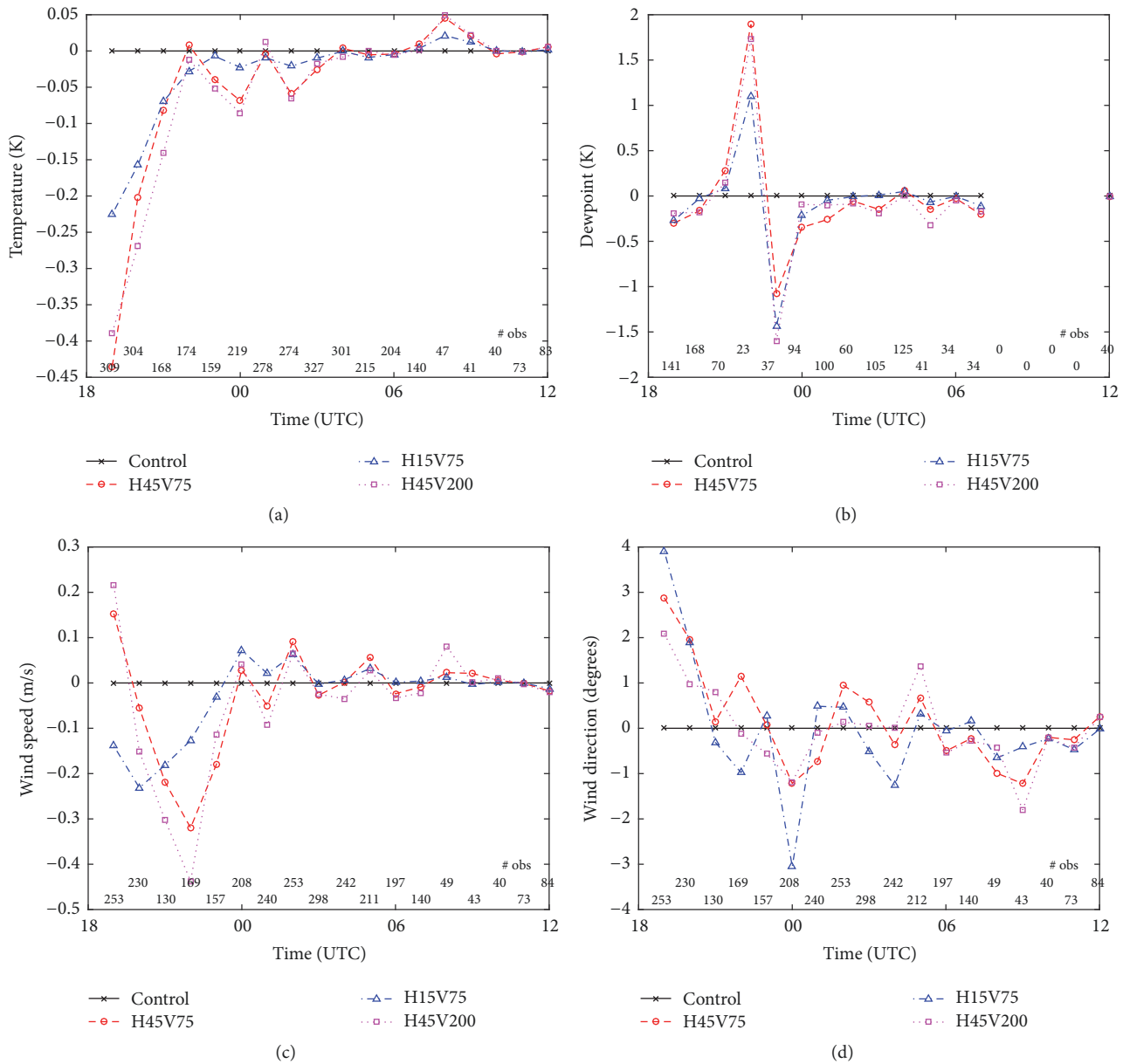
FIGURE 6: Time series of the difference between MAE for each experiment compared to the MAE for Exp. Control for the 1–18 h forecast (19–12 UTC) for the 0–1000 m AGL layer for (a) temperature, (b) dewpoint, (c) wind speed, and (d) wind direction as compared to above-surface observations and calculated over observations from all five case days. Values below the horizontal line indicate an experiment performed better than Exp. Control. The number of observations used to create the verification at each time is noted just above the horizontal axis at the bottom of the plot.

from 1.0 to 0.5 degrees. Similarly, for the apparent outlier in Figure 4(d) where for wind direction one layer higher (1000–2000 m AGL) TAMDAR degrades the forecast by 6.0 degrees, removing the 4.1% of observations with improvements or degradations greater than 45 degrees decreases the degradation to 2.7 degrees. This illustrates how significant degradations in a small subset of wind observations can significantly affect the mean effect.

The temporal evolution of the 0–1000 m AGL error in this layer demonstrates the length of time over which TAMDAR observations continue to influence the forecast (Figure 6).

For temperature (Figure 6(a)), assimilation of TAMDAR observations improves the 1 h forecast by 0.2–0.4 K, with the improvement of 0.4 K for both experiments using the larger HROI (45 km). However, this improvement rapidly decays with time, such that the improvement appears mainly limited to the 1–3 h forecast (19–21 UTC). For dewpoint (Figure 6(b)), a small improvement due to TAMDAR is seen in the 1 h forecast for all three experiments (0.2–0.3 K), as well as a smaller improvement for the nudging experiments with larger ROIs at 2 h (0.2 K in Exp. H45V75 and Exp. H45V200), but then a small degradation at 3 h. For the 4 h and 5 h forecasts

(22 and 23 UTC), a large improvement followed by a large degradation is noted. However, the number of observations available for verification greatly decreases for this period, and thus it is difficult to determine the significance of this temporal variation. In particular, the hourly progression of the number of observations available for dewpoint verification in the 0–1000 m AGL layer 1–6 h forecast is 141, 168, 70, 23, 37, and 94. For wind speed (Figure 6(c)), the 1 h forecast shows TAMDAR causing a small improvement for the smallest ROIs (0.1 ms^{-1} for Exp. H15V75) but a small degradation for the larger ROIs (0.1–0.2 ms^{-1}). However, all nudging experiments show improvements after this for the 2–5 h forecast (20–23 UTC) (except at 5 h Exp. H15V75 shows little effect). For wind direction (Figure 6(d)), all nudging experiments show degradation for the first 2 h of the forecast (2–4 degrees at 19 UTC and 1-2 degrees at 20 UTC), but after that the results are mixed. For 19 UTC in Exp. H45V75, excluding the 9.1% of model-observation pairs with large degradations (>45 degrees) changes the mean from a degradation of 2.9 degrees to an improvement of 0.6 degrees. This again highlights the substantial impact that model-observation pairs with large wind direction degradations can have on the mean statistics.

The relative performance among experiments and times for the 0–1000 m AGL layer as measured by RMSE (not shown) is very similar to that indicated by MAE except for wind direction. Wind direction forecasts appear to be slightly degraded by nudging during the first couple hours of the forecast (19–20 UTC) according to both MAE (by 1–4 degrees; Figure 6(d)) and RMSE (by 1-2 degrees) and after 20 UTC both MAE and RMSE differences due to nudging are generally less than 2 degrees. However, in general at 0–1000 m AGL the RMSE differs much more notably from the MAE for wind direction than for the other variables; this is not surprising given that the nature of wind direction lends itself to a distribution of wind direction errors with a small number of relatively very large errors (e.g., Figure 5(d)).

The time evolution of bias (mean error) near the surface (0–1000 m AGL) generally indicates differences among the experiments that decrease with forecast lead time (Figure 7). For temperature (Figure 7(a)), at 19 UTC (1 h forecast) a cool bias in Exp. Control (−0.9 K) is improved somewhat by nudging to TAMDAR data with a horizontal radius of influence of 15 km (Exp. H15V75; bias of −0.7 K) but improved most by a horizontal radius of influence of 45 km (Exp. H45V75 and Exp. H45V200; bias of −0.4–−0.5 K). The improvement disappears by the 4 h forecast, consistent with the MAE (Figure 6(a)). Dewpoint bias (Figure 6(b)) is improved by assimilation of the TAMDAR data at the 1 h forecast (19 UTC) by >0.5 K, and for the experiments with a 45 km radius of influence is improved at the 2 h forecast (20 UTC) and degraded at the 3 h forecast (21 UTC). The drastic decreases in the number of dewpoint observations to verify against starting at 21 UTC (70 observations compared to 168 at 20 UTC) and especially so at 22–23 UTC (23 and 37 observations) make it harder to attach importance to these differences. Assimilating TAMDAR data increased the forecast wind speed (Figure 6(c)), which improved the bias for the 1 h forecast (19 UTC), but degraded the bias for

subsequent times. For wind direction (Figure 6(d)), while it appears that nudging with a 15 km horizontal radius of influence (Exp. H15V75) degraded wind direction bias for the 1 h forecast, other than that it is difficult to ascertain a clear signal regarding the effects of TAMDAR data assimilation on wind direction bias. While mean error (bias) can provide helpful information regarding the nature of model error, due to the possibility of compensating negative and positive errors, metrics such as MAE and RMSE can provide a more robust evaluation of model performance.

In addition to evaluating the model output against above-surface observations, verification was also performed against surface observations (Figure 8; the spatial distribution of land-based surface observations is shown in Figure 1(b)). In general, overall effects on the MAE at the surface are small. Surface temperature is slightly improved at the 1 h forecast (Figure 8(a); 19 UTC), but the improvement rapidly dissipates. For surface dewpoint (Figure 8(b)), while MAE is generally slightly decreased via assimilation of TAMDAR observations, the magnitude of the changes is quite small (<0.1 K). For surface wind, both speed (Figure 8(c)) and direction (Figure 8(d)) show slight degradation with the assimilation of TAMDAR data.

Although overall the impact of assimilating the TAMDAR observations on the surface verifications was minimal, on 7 February the 1 h forecast shows a noticeable impact on surface temperature. The model 1 h forecast (19 UTC) 2 m temperature on 7 February is shown for Exp. Control (Figure 9(a)) and Exp. H45V75 (Figure 9(b)) with the observed 2 m temperatures overlaid as circles centered at the observation locations. One notable impact at 19 UTC is that near the center of the domain (approximately 37°20′N, 122°00′W) assimilating the TAMDAR data creates a warmer 2 m forecast at this time which more closely matches the observations (as seen by the closer match in color between the overlaid circles and the underlying shading in Figure 9(b) in this region as compared to that in Figure 9(a)). In the vicinity of airports, TAMDAR observations can be available close to the surface that originate from aircraft taking off or landing. The presence of these near-surface observations provides one method for TAMDAR observations to correct low-level model biases; for this case there were approximately 38 TAMDAR observations available for assimilation in the lowest 1000 m. This case demonstrates the ability of aircraft-based observations to correct a low-level temperature bias.

Model precipitation forecasts were verified for the two cases with precipitation in the 1 km domain during the 0–6 h forecast (7 February and 1 March). The 0–6 h precipitation forecast was verified against the 4 km stage IV precipitation analyses [35] using the Model Evaluation Tools (MET [36]). The Fractions Skill Score (FSS [36, 37]) is a neighborhood-based method which means that it has the possibility of distinguishing between forecasts of precipitation that are near misses compared to forecasts that are much farther from observations. Neighborhood methods can be especially advantageous when verifying high-resolution forecasts since traditional point-to-point verification methods can indicate poor model performance for cases where the model forecast closely matched observations but was spatially or temporally
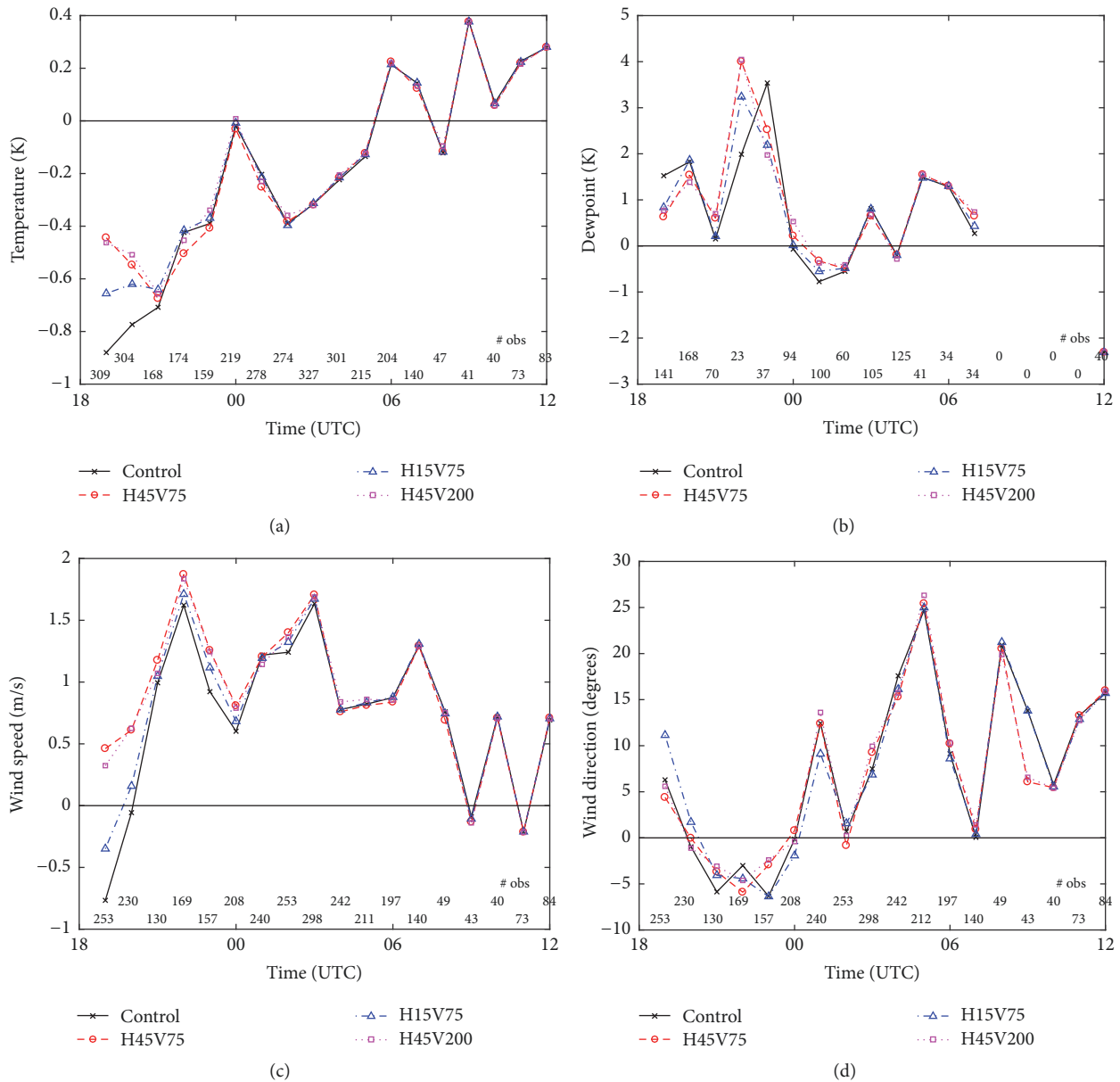
FIGURE 7: Time series of the ME for each experiment for the 1–18 h forecast (19–12 UTC) for the 0–1000 m AGL layer for (a) temperature, (b) dewpoint, (c) wind speed, and (d) wind direction as compared to above-surface observations and calculated over observations from all five case days. The number of observations used to create the verification at each time is noted just above the horizontal axis at the bottom of the plot.

offset by a small amount. However, neighborhood methods are most straightforward to apply for fields where a spatially continuous verification field (e.g., radar) is available and categorical verification (e.g., >1 mm) provides the desired information.

As applied here, calculating FSS involves comparing the fraction of a "neighborhood" (a collection of one or more grid points having an equal number of grid points in both horizontal directions) over which the WRF precipitation forecast exceeds a threshold to the fraction of that neighborhood that was observed to exceed that threshold; the results are then

combined for all possible neighborhoods of that size. Because the observed precipitation is on a 4 km grid, converting the WRF forecast to the stage IV precipitation grid results in each grid cell used in verification representing approximately 16 grid cells of the model forecast. Therefore, verifying a single grid cell of this 4 km grid is actually representing information from a 16-grid cell neighborhood of the WRF model forecast. The FSS was calculated for neighborhoods of 1, 9, and 25 4 km grid cells. The 1.0 mm threshold was used for both days, but the 2.5 mm threshold was used only for 1 March because very few grid cells in the observed precipitation reached that
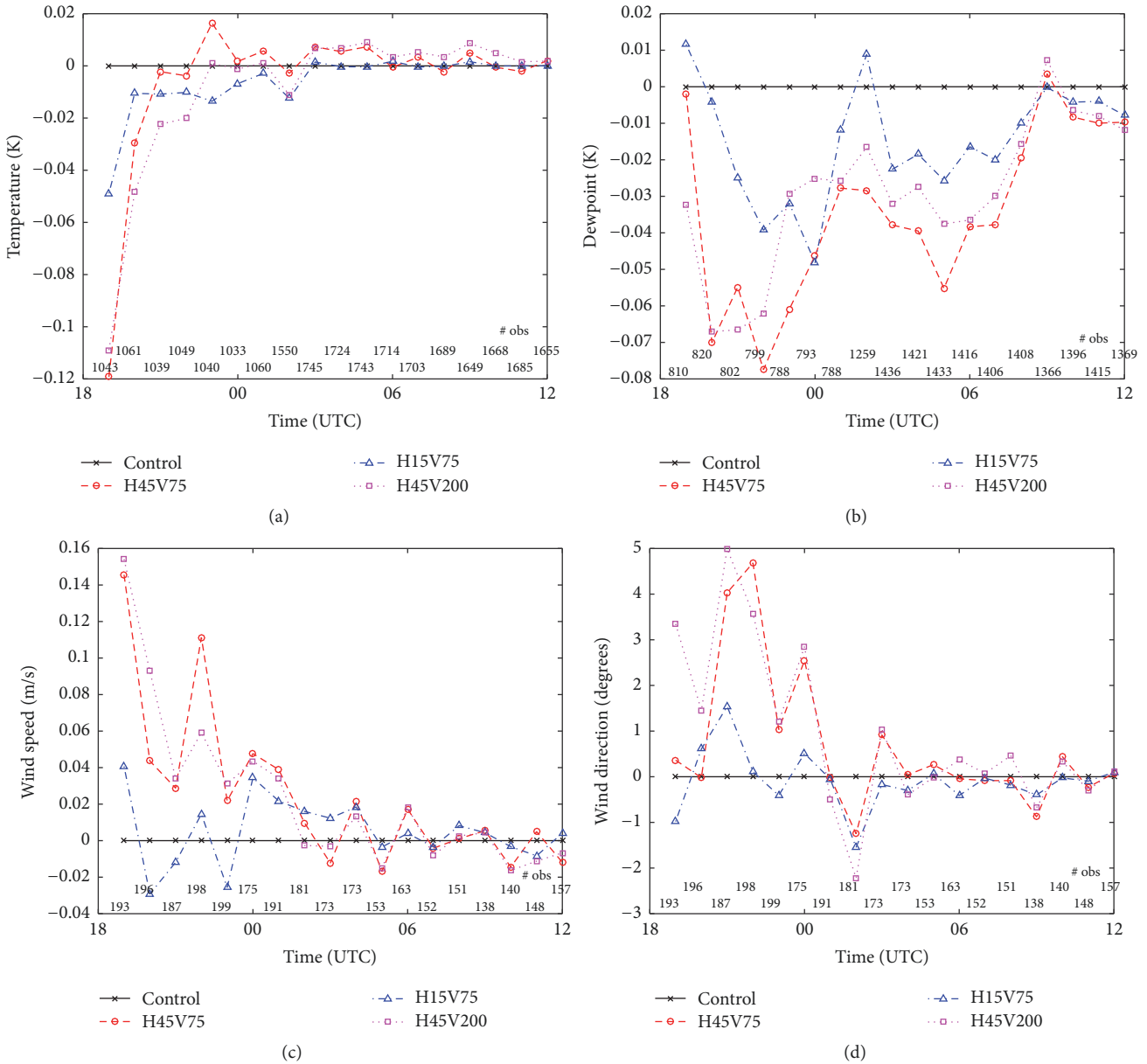
FIGURE 8: Time series of the difference between MAE for each experiment compared to the MAE for Exp. Control for the 1–18 h forecast (19–12 UTC) for the surface calculated over observations from all five case days for (a) temperature, (b) dewpoint, (c) wind speed, and (d) wind direction. Values below the horizontal line indicate an experiment performed better than Exp. Control. The number of observations used to create the verification at each time is noted just above the horizontal axis at the bottom of the plot; the use of use-reject lists in the quality control contributes to significant variations in the number of observations available among variables and between night and day.

threshold on 7 February during that 0–6 h forecast. Note that higher values of FSS indicate better forecasts, with 1.0 being the highest possible FSS.

On 7 February, all experiments notably overforecast the coverage of precipitation, and the resultant FSS (Table 1) were low and varied little among the experiments. All of the experiments FSS fell well within the 95% confidence intervals of the other experiments and thus there does not appear to be any notable difference among the experiments precipitation forecast skill for this case. For the other case with precipitation during the 0–6 h forecast, 1 March, the

FSS were much higher than on 7 February (Table 1). For the 1.0 mm threshold, Exp. H45V200 has a lower FSS compared to Exp. Control, while Exp. H45V75 and Exp. H15V75 have slightly higher FSS compared to Exp. Control. However, the 95% confidence intervals of the four experiments overlap for all neighborhood sizes at the 1.0 mm threshold. For the 2.5 mm threshold, all of the experiments have higher FSS than the control experiment for all three neighborhood sizes; furthermore, Exp. H45V75 has higher FSS than any of the other experiments at all three neighborhood sizes. However, the only values with nonoverlapping confidence intervals are
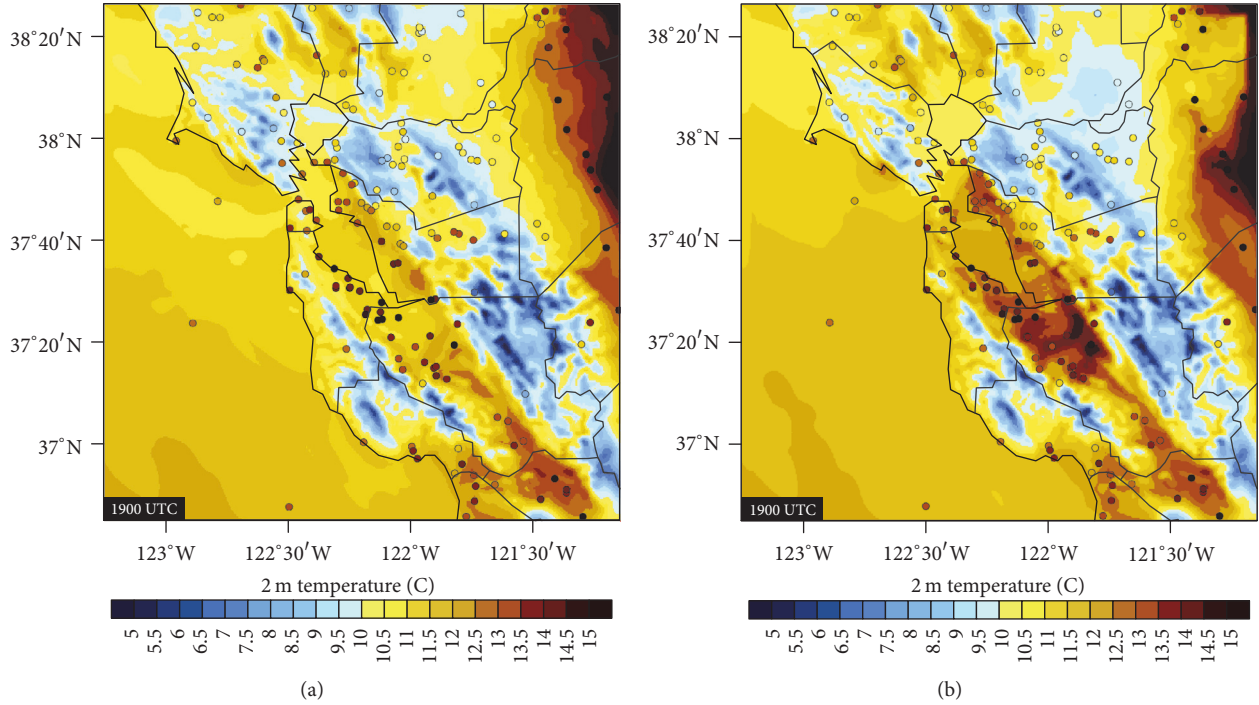
Figure 9: Model 1 km horizontal grid spacing domain 1 h forecast (19 UTC 7 February) 2 m temperature for (a) Exp. Control and (b) Exp. H45V75 with the observed 2 m temperature overlaid as colored circles centered over the observation locations.

Table 1: FSS for the 0–6 h (18–00 UTC) precipitation forecast for each experiment on the two case days with precipitation during the 0–6 h forecast.

Case	Threshold (mm)	Neighborhood size (4 km grid cells)	Experiment name			
			Control	H45V75	H15V75	H45V200
7 Feb	>=1.0 mm	1	0.084	0.086	0.085	0.087
		9	0.078	0.075	0.073	0.075
		25	0.063	0.065	0.063	0.065
1 Mar	>=1.0 mm	1	0.571	0.581	0.577	0.562
		9	0.673	0.676	0.677	0.653
		25	0.723	0.719	0.725	0.694
	>=2.5 mm	1	0.296	0.317	0.307	0.316
		9	0.400	0.452	0.425	0.432
		25	0.437	0.517	0.482	0.492

the 25-grid cell neighborhood FSS for the control experiment and Exp. H45V75.

7. Discussion

Assimilation of TAMDAR observations showed the clearest benefit in terms of temperature, while results were mixed for dewpoint and wind speed, and TAMDAR observations degraded the wind direction forecast. For the two case days with precipitation during the 0–6 h forecast, the assimilation of TAMDAR observations had a limited effect when measured by a 1.0 mm verification threshold; however, for the single case warranting application of a 2.5 mm threshold,

assimilation of TAMDAR observations appears to improve the model precipitation forecast. The mixed results for dewpoint may partially be a result of a more limited observational dataset to verify dewpoint against compared to temperature. In regard to wind, Jacobs et al. (2014) note that a subset of TAMDAR observations contain notably larger wind errors due to the method used to obtain wind direction on some aircraft and demonstrate a methodology to correct for this error. It may be that some of the observations in this study are from aircraft with this issue and thus this issue may have contributed to the wind direction degradation and also suppressed the capability of the TAMDAR observations to improve wind speed forecasts.

Temperature, dewpoint, and wind speed were all improved in the lowest 1000 m AGL averaged over the 1–6 hour forecast (Figures 4(a)–4(c)). This is the layer with both the most TAMDAR observations to assimilate and the most observations to verify against, and thus it is encouraging to see the most improvement here. While the mean improvements were small, the distribution of changes indicated some model-observation pairs with much larger improvements (Figure 5).

The benefit of assimilating TAMDAR observations, even for temperature in the 0–1000 m AGL layer, decreases rapidly with forecast time. In this layer, temperature improvements are limited mostly to the first three hours of the forecast (Figure 6(a)), dewpoint improvements disappear (at least temporarily) after the first two hours of the forecast (Figure 6(b)), and, excluding the mixed results in the 1 h forecast, wind speed improvements in this layer are seen in the 2–5 h forecast.

Most of the observations being assimilated are in a northwest-to-southeast strip through the center of the domain (Figure 2), and since the entire 1 km domain is 188 km across, the time for air affected by the TAMDAR observations to be advected out of the domain will be somewhat limited. This will be especially the case above the atmospheric boundary layer, where winds may be stronger. This advection may contribute to the rapid degradation with time of the benefits of assimilating TAMDAR observations and also provide further reason for the generally decreased benefits above the 0–1000 m AGL layer.

In general, the experiment with the smallest HROI and vertical spreading (Exp. H15V75) showed the smallest improvements (and smallest degradations) consistent with the decreased influence TAMDAR observations have in this experiment. This suggests that use of a larger HROI (45 km instead of 15 km) allows the TAMDAR observations to improve a larger portion of the domain when the TAMDAR observation improves the forecast, but also can magnify the area degraded by a TAMDAR observation. Among the two experiments using an HROI of 45 km (Exp. H45V75 and Exp. H45V200), the results did not indicate whether 75 hPa or 200 hPa vertical spreading was better.

It is not certain why assimilating TAMDAR observations degrades the model forecast for certain variables and heights. Probable causes include weaknesses in the configuration of the nudging data assimilation and errors in the TAMDAR observations. Stricter quality control may need to be imposed on the TAMDAR observations to limit the possibility of TAMDAR observations with large errors being ingested by WRF. The horizontal and vertical spreading of the influence of observations may need to differ among temperature, moisture, and dewpoint or may need to vary across the domain based on factors not yet fully accounted for in the nudging data assimilation system. For example, it may be that near-surface TAMDAR observations over land that are near the coast should have their influence spread less far over the ocean than over the land (based on the assumption that the model error at the location of the TAMDAR observation is better correlated with the model error at a point over land than with the model error at an equidistant point that is over water).

It can be challenging to evaluate the forecast skill of short-term high-resolution model forecasts above the surface due to a relative lack of observations, especially in cases without precipitation. Aircraft data can provide an important source of above-surface observations for verification, as leveraged in this study. However, there is significant spatial and temporal variability in the availability of aircraft observations and even in this study over an area with a major airport, additional observations would have allowed for a more robust evaluation of the potential value of TAMDAR observations.

8. Summary and Conclusions

TAMDAR observations of temperature, moisture, and wind were assimilated into 1 km WRF-ARW forecasts for five case days over the San Francisco, California, region. Evaluation focused on above-surface observations during the forecast period. Overall, temperature was improved via assimilation of TAMDAR observations, while the impact on dewpoint and wind speed was mixed, and the impact on wind direction was negative. The impact of the assimilation decreased rapidly with time, which is not surprising given the relatively narrow area within which TAMDAR observations were available. Using a 15 km HROI resulted in TAMDAR having less effect on the model simulations, although overall it was unclear which TAMDAR experiment performed best.

As limited research has been reported on assimilating TAMDAR observations in high-resolution ($\approx$1 km horizontal grid spacing) models, the experiments reported here explore the potential value of TAMDAR observations at high resolutions. High-resolution domains may often have limited traditional above-surface observational data to assimilate due to their limited size, and aircraft-based observations provide a promising potential source to better initialize such forecasts. However, further work is needed to explore how to best utilize these observations at these resolutions, including work with various data assimilation techniques. This work should include application to additional cases and locations to expand the robustness of this evaluation.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

precipitation data was provided by the National Center for Atmospheric Research (NCAR) Earth Observing Laboratory which is under the sponsorship of the United States National Science Foundation (NSF). Precipitation verification was completed using the Model Evaluation Tools (MET) which was developed at NCAR through grants from the United States Air Force Weather Agency (AFWA) and NOAA. NCAR is sponsored by the NSF.

References

[1] Office of the Federal Coordinator for Meteorological Services and Supporting Research (US), "Federal Meteorological Handbook No. 3: Rawinsonde and Pibal Observations," FCM-H3-1997, 1997.

[2] W. R. Moninger, R. D. Mamrosh, and P. M. Pauley, "Automated meteorological reports from commercial aircraft," *Bulletin of the American Meteorological Society*, vol. 84, no. 2, pp. 203–216, 2003.

[3] T. S. Daniels, J. J. Murray, C. A. Grainger et al., "Validation of tropospheric airborne meteorological data reporting (TAMDAR) temperature, relative humidity, and wind sensors during the 2003 Atlantic THORPEX regional campaign and the Alliance Icing Research Study (AIRS II)," in *Proceedings of the 11th AMS Conference on Aviation, Range, and Aerospace*, American Meteorological Society, Hyannis, MA, USA, 2003.

[4] W. R. Moninger, S. G. Benjamin, B. D. Jamison, T. W. Schlatter, T. L. Smith, and E. J. Szoke, "Evaluation of regional aircraft observations using TAMDAR," *Weather and Forecasting*, vol. 25, no. 2, pp. 627–645, 2010.

[5] J. E. Passner, S. Kirby, and T. Jameson, "Using Real-Time Weather Data from an Unmanned Aircraft System to Support the Advanced Research Version of the Weather Research and Forecast Model," United States Army Research Laboratory Technical Report ARL-TR-5950, 2012.

[6] R. A. Petersen, L. Cronce, R. Mamrosh, R. Baker, and P. Pauley, "On the impact and future benefits of AMDAR observations in operational forecasting part II: Water vapor observations," *Bulletin of the American Meteorological Society*, vol. 97, no. 11, pp. 2117–2133, 2016.

[7] D. A. Rahn and C. J. Mitchell, "Diurnal climatology of the boundary layer in Southern California using AMDAR temperature and wind profiles," *Journal of Applied Meteorology and Climatology*, vol. 55, no. 5, pp. 1123–1137, 2016.

[8] B. E. Schwartz, S. G. Benjamin, S. M. Green, and M. R. Jardin, "Accuracy of RUC-1 and RUC-2 wind and aircraft trajectory forecasts by comparison with ACARS observations," *Weather and Forecasting*, vol. 15, no. 3, pp. 313–326, 2000.

[9] T. H. Zapotocny, S. J. Nieman, W. Paul Menzel et al., "A case study of the sensitivity of the Eta data assimilation system," *Weather and Forecasting*, vol. 15, no. 5, pp. 603–621, 2000.

[10] R. J. Graham, S. R. Anderson, and M. J. Bader, "The relative utility of current observation systems to global-scale NWP forecasts," *Quarterly Journal of the Royal Meteorological Society*, vol. 126, no. 568, pp. 2435–2460, 2000.

[11] R. A. Petersen, "On the impact and benefits of AMDAR observations in operational forecasting Part I: A review of the impact of automated aircraft wind and temperature reports," *Bulletin of the American Meteorological Society*, vol. 97, no. 4, pp. 585–602, 2016.

[12] F. Gao, X. Zhang, N. A. Jacobs, X.-Y. Huang, and P. P. Childs, "Estimation of TAMDAR observational error and assimilation experiments," *Weather and Forecasting*, vol. 27, no. 4, pp. 856–877, 2012.

[13] X. Zhang, H. Wang, X.-Y. Huang, F. Gao, and N. A. Jacobs, "Using adjoint-based forecast sensitivity method to evaluate TAMDAR data impacts on regional forecasts," *Advances in Meteorology*, vol. 2015, Article ID 427616, 13 pages, 2015.

[14] Y. Zhang, Y. Liu, and T. Nipen, "Evaluation of the Impacts of Assimilating the TAMDAR Data on 12/4 km Grid WRF-Based RTFDDA Simulations over the CONUS," *Advances in Meteorology*, vol. 2016, Article ID 3282064, 13 pages, 2016.

[15] W. C. Skamarock, J. B. Klemp, J. Dudhia et al., "A description of the Advanced Research WRF version 3," NCAR Tech. Note NCAR/TN-475+STR, National Center for Atmospheric Sciences, Boulder, Colo, USA, 2008.

[16] N. A. Jacobs, D. J. Mulally, and A. K. Anderson, "Correction of flux valve-based heading for improvement of aircraft wind observations," *Journal of Atmospheric and Oceanic Technology*, vol. 31, no. 8, pp. 1733–1747, 2014.

[17] H. L. Wang and X.-Y. Huang, "TAMDAR observation assimilation in WRF 3D-Var and its impact on hurricane Ike (2008) forecast," *Atmospheric and Oceanic Science Letters*, vol. 5, no. 3, pp. 206–211, 2015.

[18] D. R. Stauffer and N. L. Seaman, "Multiscale four-dimensional data assimilation," *Journal of Applied Meteorology and Climatology*, vol. 33, no. 3, pp. 416–434, 1994.

[19] M. O. Jonassen, H. Ólafsson, H. Ágústsson, Ó. Rögnvaldsson, and J. Reuder, "Improving high-resolution numerical weather simulations by assimilating data from an unmanned aerial system," *Monthly Weather Review*, vol. 140, no. 11, pp. 3734–3756, 2012.

[20] W. Gemmill, B. Katz, and X. Li, "Daily real-time, global sea surface temperature—High-resolution analysis: RTG_SST_HR," NOAA/NWS/NCEP/MMAB Office Note 260, 2007.

[21] National Operational Hydrologic Remote Sensing Center, *Snow Data Assimilation System (SNODAS) Data Products at NSIDC*, National Snow and Ice Data Center, Boulder, Colo, USA, 2004.

[22] National Ice Center, *IMS Daily Northern Hemisphere Snow and Ice Analysis at 4 km Resolution, Version 1*, National Snow and Ice Data Center, Boulder, Colo, USA, 2008.

[23] Z. I. Janjić, "Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model," NCEP Office Note 437, 2002.

[24] B. P. Reen, R. E. Dumais, and J. E. Passner, "Mitigating excessive drying from the use of observations in mesoscale modeling," *Journal of Applied Meteorology and Climatology*, vol. 55, no. 2, pp. 365–388, 2016.

[25] G. Thompson, P. R. Field, R. M. Rasmussen, and W. D. Hall, "Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II. Implementation of a new snow parameterization," *Monthly Weather Review*, vol. 136, no. 12, pp. 5095–5115, 2008.

[26] J. S. Kain, "The Kain-Fritsch convective parameterization: an update," *Journal of Applied Meteorology and Climatology*, vol. 43, no. 1, pp. 170–181, 2004.

[27] E. J. Mlawer, S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, "Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave," *Journal of Geophysical Research: Atmospheres*, vol. 102, no. 14, pp. 16663–16682, 1997.

[28] J. Dudhia, "Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model," *Journal of the Atmospheric Sciences*, vol. 46, no. 20, pp. 3077–3107, 1989.

[29] F. Chen and J. Dudhia, "Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity," *Monthly Weather Review*, vol. 129, no. 4, pp. 569–585, 2001.

[30] J. A. Lee, W. C. Kolczynski, T. C. McCandless, and S. E. Haupt, "An objective methodology for configuring and down-selecting an NWP ensemble for low-level wind prediction," *Monthly Weather Review*, vol. 140, no. 7, pp. 2270–2286, 2012.

[31] B. P. Reen, K. J. Schmehl, G. S. Young, J. A. Lee, S. E. Haupt, and D. R. Stauffer, "Uncertainty in contaminant concentration fields resulting from atmospheric boundary layer depth uncertainty," *Journal of Applied Meteorology and Climatology*, vol. 53, no. 11, pp. 2610–2626, 2014.

[32] M. S. F. V. De Pondeca, G. S. Manikin, G. DiMego et al., "The Real-Time Mesoscale Analysis at NOAA's National Centers for Environmental Prediction: Current status and development," *Weather and Forecasting*, vol. 26, no. 5, pp. 593–612, 2011.

[33] A. Deng, D. R. Stauffer, B. J. Gaudet et al., "Update on WRF-ARW End-to-End Multi-Scale FDDA System," in *Proceedings of the 10th WRF Users' Workshop*, vol. 14, National Center for Atmospheric Research, Boulder, Colo, USA, 2009.

[34] B. P. Reen, "A Brief Guide to Observation Nudging in WRF," WRF-ARW Documentation, 2016, http://www2.mmm.ucar.edu/wrf/users/docs/ObsNudgingGuide.pdf.

[35] Y. Lin, "GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0," UCAR/NCAR – Earth Observing Laboratory.

[36] B. Brown, R. Bullock, T. Fowler, J. Halley Gotway, K. Newman, and T. Jensen, "Model Evaluation Tools Version 6.0 (METv6.0) User's Guide," Developmental Testbed Center Report, 2017.

[37] I. T. Jolliffe and D. B. Stephenson, Eds., *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, John Wiley & Sons, 2nd edition, 2012.