ARL-TR-9115 ● Nov 2020

DEVCOM ARMY RESEARCH LABORATORY

An Evaluation of the Unrestricted Mesoscale Analysis as Gridded Observations for Spatial Model Verification

by John W Raby, Huaqing Cai, Leelinda Dawson, and Robert Dumais

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-9115 ● Nov 2020

An Evaluation of the Unrestricted Mesoscale Analysis as Gridded Observations for Spatial Model Verification

John W Raby, Huaqing Cai, Leelinda Dawson, and Robert Dumais
Computational and Information Sciences Directorate, DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
November 2020	Technical Report	1 October 2019–30 September 2020

4. TITLE AND SUBTITLE
An Evaluation of the Unrestricted Mesoscale Analysis as Gridded Observations for Spatial Model Verification

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
John W Raby, Huaqing Cai, Leelinda Dawson, and Robert Dumais

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLC-EM
White Sands Missile Range, NM 88002

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-9115

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ORCID ID(s): Leelinda Dawson, 0000-0003-4209-8459, Huaqing Cai, 0000-0003-3918-4153

14. ABSTRACT
The Army requires weather knowledge products, which include tactically significant meteorological variables and decision aid products impacting Army missions and systems. The Army's Weather Running Estimate – Nowcast (WRE-N) model provides forecasts that are ingested into such decision aids. To verify the accuracy of high-resolution WRE-N forecasts, advanced spatial methods of model verification are needed, which require high-resolution gridded observations. An evaluation was conducted to determine the accuracy of a gridded observation product called the Unrestricted Mesoscale Analysis (URMA). This novel evaluation involved applying a grid-to-point verification technique to compare the URMA observations to point ground-truth measurements, and the computation of error statistics for several meteorological variables. The evaluation used hourly URMA observations and point observations collected over a continuous 99-day period from 11 November 2016 to 17 February 2017, characterized as a "winter" season. The domain of the evaluation was over a large portion of the Western United States and Northern Mexico and included a wide range of terrain types associated with complex mountain-desert-basins, the central and southern plains, and coastal plains, as well as oceanic areas. The results of the study show that the URMA gridded observations can reliably be used for assessment of the WRE-N.

15. SUBJECT TERMS
verification, observations, nowcast, threshold, decision aid, weather impacts

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			John W Raby
Unclassified	Unclassified	Unclassified	UU	44	19b. TELEPHONE NUMBER (Include area code) (575) 678-2004

Contents

List of Figures

List of Tables

Acknowledgments

Many thanks to Ms Sandy Fletcher, Ms Lisa Lacey, and Ms Amber Bennett-Groves of DEVCOM Army Research Laboratory Technical Publishing for their attention to every detail in the formatting and editing of this technical report.

Executive Summary

An evaluation was conducted to determine the accuracy of the gridded observations generated by the National Oceanic and Atmospheric Agency (NOAA) National Center for Environmental Prediction (NCEP)/National Weather Service (NWS) for their use in the verification of Numerical Weather Prediction (NWP) models. Advanced methods of model verification require gridded observations in order to be able to verify high-resolution output spatially, as opposed to the more traditional methods, which perform point-by-point comparisons with observational ground-truth data coming from weather observations. The latter method often mischaracterizes the true skill of high-resolution forecasts. The gridded observation product evaluated was the Unrestricted Mesoscale Analysis (URMA).

Weather knowledge products for the warfighters include forecasts of tactically significant variables and decision aid products, which depict the 2-D distribution of weather phenomena that can impact Army missions and systems. The Army Weather Running Estimate – Nowcast (WRE-N) model provides the forecast grids, which are ingested into the decision aids. The decision aids apply thresholds to locate areas in time and space that exceed these limits and indicate the possibility of significant impacts. To evaluate the accuracy of the forecasts at the high resolutions of interest to the Army, high quality gridded observations are needed to perform spatial verification. The URMA gridded observations are a potential source of such observations for this use, and were used here in a study to assess their accuracy. The evaluation involved comparing URMA with point observation ground-truth data over a large domain located over the Western United States and Northern Mexico. The results of the study have determined that the URMA product is reasonably accurate and can be used for assessment of the WRE-N as forecast grids approaching 1 km grid spacing.

1. Introduction

The Army requires weather knowledge products to support mission planning and execution, to compete and defeat the adversary in a multi-domain environment. Engaging adversaries in multiple domains requires complete situational awareness of environments that may vary significantly in terms of terrain and weather conditions. The knowledge products will need to be tailored for a wide variety of terrain and meteorological conditions encountered in multiple domains. Weather knowledge products for the warfighters include forecasts of tactically significant variables and decision aid products that depict the 2-D distribution of weather phenomena that can impact Army missions and systems. The Army's Weather Running Estimate – Nowcast (WRE-N) model is a Numerical Weather Prediction (NWP) model that provides the forecast grids of meteorological variables, which are ingested into the decision aids. The decision aids apply thresholds to these forecasts to determine the spatial and temporal distribution of weather conditions that can impact the effectiveness of multi-domain formations. To evaluate the accuracy of the forecasts at the high resolutions of interest to the Army, advanced methods of model verification are needed, which require gridded observations in order to verify high-resolution output spatially, as opposed to the more traditional methods that perform point-by-point comparisons with observational ground-truth data coming from weather observations. This grid-to-point approach to verification cannot adequately assess the true skill of high-resolution forecasts.

1.1 Traditional Verification

Traditional grid-to-point model evaluation methods use point observations to verify the skill of NWP in predicting continuous meteorological variables, by computing such statistics as mean error (ME) and root-mean-square error (RMSE), which characterize model accuracy over the entire domain. When these techniques are applied to high-resolution models such as the WRE-N, the results can give misleading error estimates. When compared with error statistics for lower-resolution models, high-resolution models often score poorer when using these techniques. The issue is the inability of the verification technique to evaluate the true skill of higher-resolution forecasts, which replicate mesoscale atmospheric features in a way that is more representative of the actual phenomenon, owing to their use of a finer grid over smaller domains, higher-resolution land-surface models, and better parameterization of subgrid physical processes (Jolliffe and Stephenson 2012).

1.2 Spatial Verification

In recent years, various nontraditional verification techniques were developed to show the value of higher-resolution forecasts. In particular, spatial verification techniques were developed that overcome the limitations of grid-to-point techniques. These techniques score on the basis of the exact matching between point observations and the forecasts at those points. Fuzzy verification, also known as neighborhood verification, uses an approach that does not require exact matching; but instead focuses on how well the atmospheric feature or object is replicated by the model, even if there is a spatial displacement of the feature. Ebert (2008) reviews a number of such methods. The goal is to determine the amount of displacement by using a range of neighborhood sizes, which encompass surrounding forecast and observed grid points in the verification process. In this way, model performance, as a function of spatial scale, can be determined to allow selection of the scale required to have the desired accuracy. These spatial verification methods require gridded observations vice point observations for ground truth.

1.3 URMA Gridded Observations

Sources of gridded observations are few, particularly at the spatial scale needed for Army weather knowledge products to be tailored for multi-domain formations operating in regions with varied and complex terrain conditions. A candidate source of gridded observations is the Unrestricted Mesoscale Analysis (URMA). URMA is used by the National Oceanic and Atmospheric Agency (NOAA) National Centers for Atmospheric Research (NCEP)/National Weather Service (NWS) for verification of Numerical Weather Prediction (NWP) models. The Real Time Mesoscale Analysis (RTMA), in conjunction with URMA, provides real–time, 2-D meteorological gridded analysis products produced from NWP analyses and hourly point weather observations, from the national networks of Météorologique Aviation Régulière (METAR) and mesoscale sensors that are distributed over the continental United States (CONUS). Two-dimensional RTMA/URMA was developed by National Centers for Environmental Prediction (NCEP) in collaboration with the Earth System Research Laboratory and the National Environmental, Satellite, and Data Information Service (NESDIS) (De Pondeca et al. 2011). RTMA/URMA is produced on an hourly basis using a mesoscale analysis background field produced by a blend of the downscaled High Resolution Rapid Refresh (HRRR) model and the North American Mesoscale (NAM) model downscaled to a 3-km grid as a first guess background field (Morris et al. 2020). For the URMA products used for this study, HRRR v2 on a 3-km grid was used (Benjamin et al. 2016). To fill in gaps at the edges of the domain, the most recent

forecasts from the Rapid Refresh (RAP) were used (Morris et al. 2020). The RAP (RAP v3 for this study) provides an hourly forecast on a 13-km grid over North America (Benjamin et al. 2016). The first guess field is then adjusted through a 2-D variational data assimilation technique (2dvar) to analyze point weather observations from the national networks of METAR and mesoscale sensors, which are distributed over CONUS (De Pondeca et al. 2011). The first cycle of the analysis is the RTMA on a 2.5-km CONUS grid, which is used for weather situational awareness, calibration, and aviation safety. URMA is produced by re-running the RTMA on the same grid 6 h following the first cycle to enhance the amount of point observations used for analysis, making it a better product for model verification/validation (Pondeca et al., 2015). For example, NOAA uses URMA gridded observations for verification and bias correction of the National Blend of Models used by NWS forecasters (Ruth et al. 2017). URMA also serves as the NWS Analysis of Record (UCAR 2020). A future development anticipated for the RTMA/URMA analysis system is the 3-D RTMA, which is planned to provide 3-D analysis fields with sub-hourly updates. (Weygandt et al. 2019).

1.4 Previous RTMA Verification

A number of studies have attempted to compare the RTMA with observations. Morris et al. (2020) reviews the results from a few such studies that focused on performing an assessment of the RTMA to evaluate its value as an alternative source of weather observations for use by airports for current conditions affecting flight safety. Their study consisted of running data-denial experiments for retrospective periods of time, which involved generating RTMA output using specified ingest configurations to allow the assimilation phase to be controlled to restrict the available observational data to create three distinct cases. The three cases were 1) a CONTROL case, which assimilated all expected observations considered to be a more typical or normal scenario; 2) an EXP case, which denied access to observations coming from certain airports considered to be a rare but not unprecedented scenario; and 3) a NODA case, which denied access to all observations, which is considered to be the worst-case scenario. They determined that the RTMA could be used as a substitute for airfield weather observations under certain conditions, for only certain meteorological variables, and only at certain locations. This assessment is the most complete, compared to any others investigated. None of the previous studies have assessed the quality of the URMA product.

2. Design of the Evaluation

At 2.5-km grid spacing, the ideal method for verifying URMA would have been spatial verification since it can replicate the details of mesoscale structures similar to NWP models at this resolution. However, spatial verification was impossible to consider as there are no other sources of gridded ground-truth observations available to allow it to be used, leaving the only possible method, which is comparison with point observations. As such, URMA presents the same difficulties as NWP models due to the previously described problems encountered when using grid-to-point verification techniques to evaluate the true skill of higher-resolution forecasts. Hence NOAA's motivation for developing URMA as the first such set of gridded observations to be used for spatial model verification. The evaluation approach used for this study employed the older, traditional technique used for assessing gridded NWP model forecasts using point observations for ground truth. In this case, the URMA gridded observations over the CONUS domain served as the gridded analysis product being evaluated and the point observations for a selected domain were used as ground truth. The domain selected, shown in Fig. 1, was located over the Western United States and Northern Mexico to include portions of the Gulf of Mexico, Sea of Cortez, and the Eastern Pacific Ocean. This domain was also the outer nest region (d01) for our WRF experiments.

Fig. 1 Evaluation domain

The verification was conducted over the domain, which is characterized by a complex mountain-desert-basin terrain, the central and southern plains, and coastal plains, as well as the above-mentioned oceanic areas. Point observations and URMA gridded observations were collected during a 99-day period, which could be characterized as a "winter" period between 11 November 2016 and 17 February 2017. The grid-to-point verification method involved taking the differences (URMA minus observation) between the URMA value of the meteorological variable interpolated to the location of each point observation, and the value of the observed variable. The method used for interpolation was a distance-weighted mean using the four closest grid values. The differences are then used to generate data summaries called partial sums, which were accumulated as needed, and used for computing continuous error statistics that quantify the errors over the desired time period. The software used to perform these calculations was the Model Evaluation Tools (MET). MET was developed at NCAR through grants from the NSF, NOAA, the United States Air Force (USAF), and the United States

Department of Energy (DOE). NCAR is sponsored by the United States NSF. The output of MET was visualized using the METViewer tool also developed by NCAR. METViewer enabled the aggregation of the error statistics over each hour for all 99 days and then produced plots of the statistics for each hour.

3. Generation of Evaluation Data

The point observation ground-truth data collected from sensors placed at the 2-m and 10-m above ground level (AGL) were collected and postprocessed for quality assurance (QA) using the real-time implementation of the Weather Running Estimate – Nowcast (WREN_RT) (Reen and Dawson 2018). WREN_RT collects the Meteorological Assimilation Data Ingest System (MADIS) archived meteorological observations at NOAA (NOAA 2014). For this study, only near-surface observations from the MADIS database consisting of standard METAR surface observations, mesonet* surface observations, and maritime surface observations were used. The QA process first involved filtering the mesonet observations by applying the use and reject lists obtained from developers of the RTMA system (De Pondeca et al. 2011). This quality-assurance measure is especially important given the greater tendency of mesonet observations to be poorly sited compared with other, more standard, surface observations. Next, the Obsgrid component of WREN_RT was used for quality control of all observations. This component included gross-error checks, comparison of observations to a background field (here the Global Forecast System), and comparison of observations to nearby observations. The quality-controlled observations were output in hourly "little_r"-formatted text files (Raby and Smith 2016).

3.1 Data Preprocessing

As shown in Fig. 2, a few preprocessing tasks needed to be completed before both the URMA gridded observation and point observation files could be ingested into the MET Point-Stat tool to produce error statistics data. Two scripts were developed and implemented in Python to assist in making these preprocessing tasks for all 99 case-study days easier and more efficient. One script, *rename_grib2.py*, performed the renaming of each hourly URMA gridded observation file in GRIB2 for all 99 days, to a standard filename convention. Then, the renamed hourly URMA gridded observation files were used as forecast input into the MET Point-Stat tool. The second script, *run_ascii2nc.py*, performed the run and data processes associated with the MET ascii2nc tool, which converted each hourly point observation file in "little_r" format for all 99 days to NetCDF files. Then, the hourly

*A network of automated meteorological observation stations.

NetCDF point observation files were used as observation input into the MET Point-Stat tool. All of the generated output from both scripts were ingested into the MET Point-Stat tool using another script, *runPointStat.py*, developed in Python, that performed the automation of the run and data processes associated with the tool (Dawson et. al 2016).

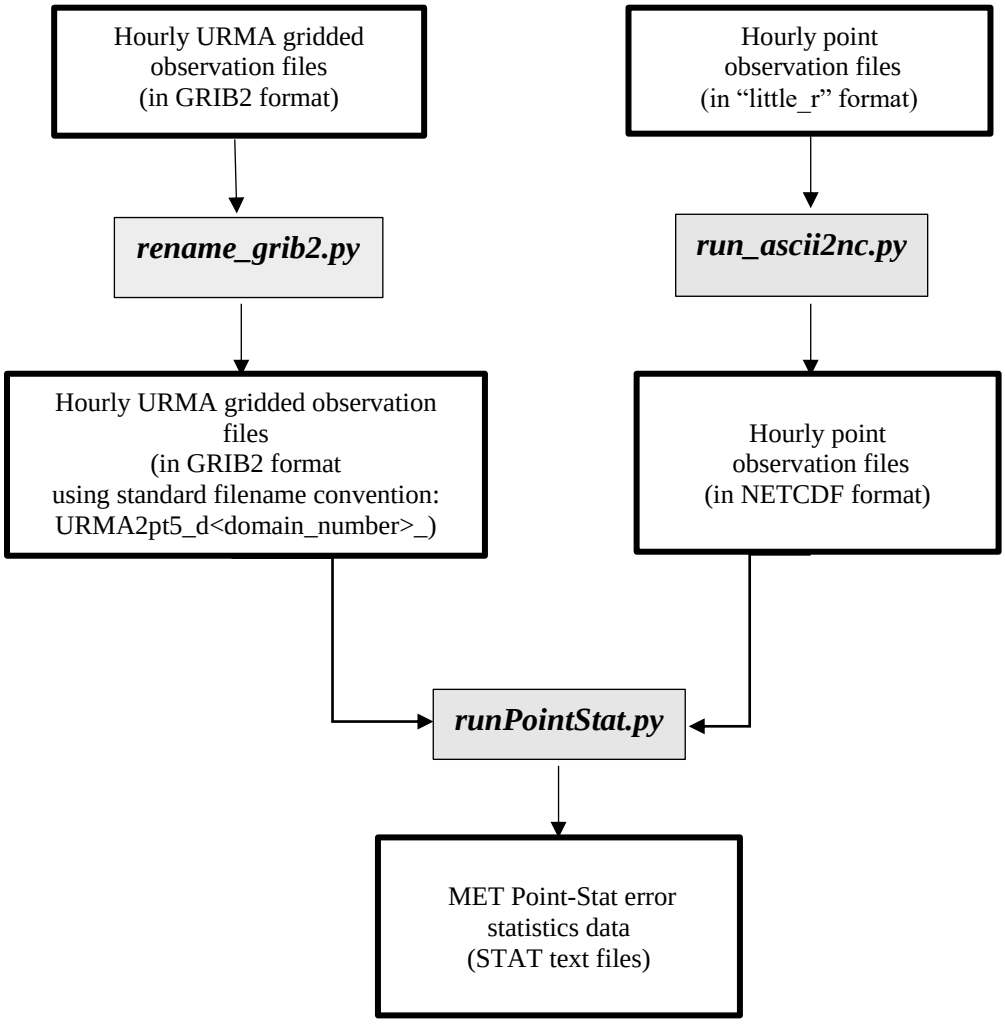

Fig. 2 Generation of evaluation data flow diagram

3.2 MET Point-Stat Processing

The MET Point-Stat tool ingests the URMA gridded observation files and the point observation files and then interpolates the value from the URMA grid to the location of the point observations to create matched pairs with all observed values. For this study, the values of the near-surface meteorological variables were used to create the matched pairs. The output of MET Point-Stat consists of a tabular ascii data formatted text file, called STAT, which is used by other MET tools to access the fundamental data from which the error statistics are calculated. The STAT file

contains 1) the matched pair data, 2) the scalar and vector partial sums generated during the calculation of the URMA-observation differences, and 3) the continuous error statistics. For this study, which involved the computation of error statistics over a 99-day period, the availability of the partial sums in the STAT file enabled the aggregation of the partial sums (i.e., the URMA values, the point observations, and their differences) to produce the error statistics (Newman et al. 2018). The METViewer software loads the STAT files, computes the error statistics according to user-specified settings by aggregating multiple partial sums, computes their weighted average where the weight is determined by the numbers of matched pairs and then generates plots of the statistics (NCAR 2020). Plots of the statistics were generated for the near-surface meteorological variables listed in Table 1.

Table 1 Near-surface meteorological variables used for the evaluation

Variable name/units	Abbreviation	Level (AGL)	Nominal matched pair count per hour
Temperature (degrees Kelvin)	TMP	2 m	6000
Dew-point temperature (degrees Kelvin)	DPT	2 m	4000
U wind component (meters/second)	UGRD	10 m	1000
V wind component (meters/second	VGRD	10 m	1000
Wind speed (meter/second)	WIND	10 m	1000
Wind direction (degrees)	WDIR	10 m	1000

3.3 Error Statistics Plots

For the 99-day period, the error statistics of all variables except for wind direction (WDIR), include the following:

- Mean value of the variable from URMA and observation with the number of matched pairs available for aggregation at each valid time.
- The ME or bias statistic aggregated over all matched pairs at each valid time.
- The RMSE statistic aggregated over all matched pairs at each valid time.

For WDIR, the following are plots which present the error statistics:

- The mean value of WDIR from URMA and the observation aggregated over the 99-day period.
- The WDIR bias statistic (WDIR_BIAS) computed by aggregating multiple vector partial sums for both wind components to compute the mean URMA and observed wind directions and their difference.
- The WDIR direction error (DIR_ERR) computed from signed angle between the directions of the averaged URMA and observed wind vectors.
- The WDIR absolute error (DIR_ABSERR) is the absolute value of the WDIR direction error (DIR_ERR).

For the formulas used for computing these statistics, see the MET User's Guide (Newman et al. 2018).

4. Analysis of Evaluation Data

The graphics showing the errors for 2-m-AGL TMP are presented in Figs. 3–5. The valid-times are shown in Coordinated Universal Time (UTC) and Mountain Standard Time (MST). MST was chosen as the most prevalent time zone across the central part of the domain, which encompasses portions of three time zones. Figure 3 shows the aggregated mean values of URMA (red points) and observed (green points) TMP for each valid time for the 99-day period.

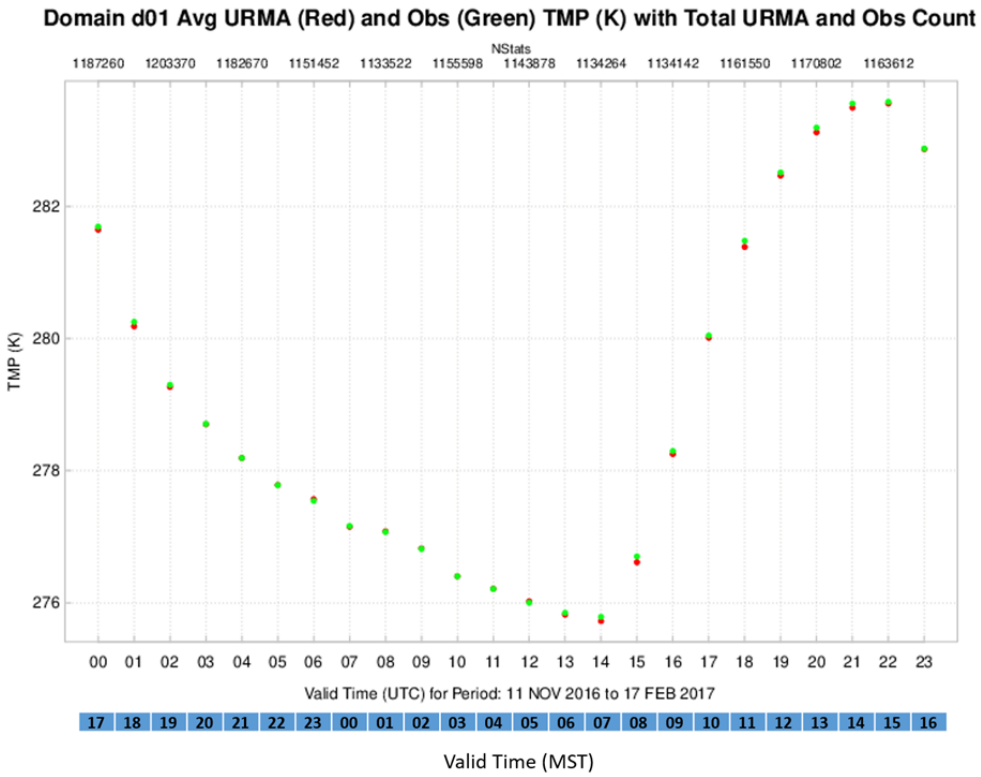

Fig. 3 Aggregated values of URMA (red points) and observed (green points) TMP for each valid time for the 99-day period. Numbers of matched pairs used for all error statistics obtained by dividing "NStats" value by 2.

The TMP varies diurnally as expected and the errors are very small as shown by the proximity of the two data points at each valid time. The numbers of TMP matched pairs, which were used to compute the aggregated error statistics, are obtained by dividing the values located along the x-axis at the top of the plot, by two. Figure 4 shows the ME or bias statistic for TMP.

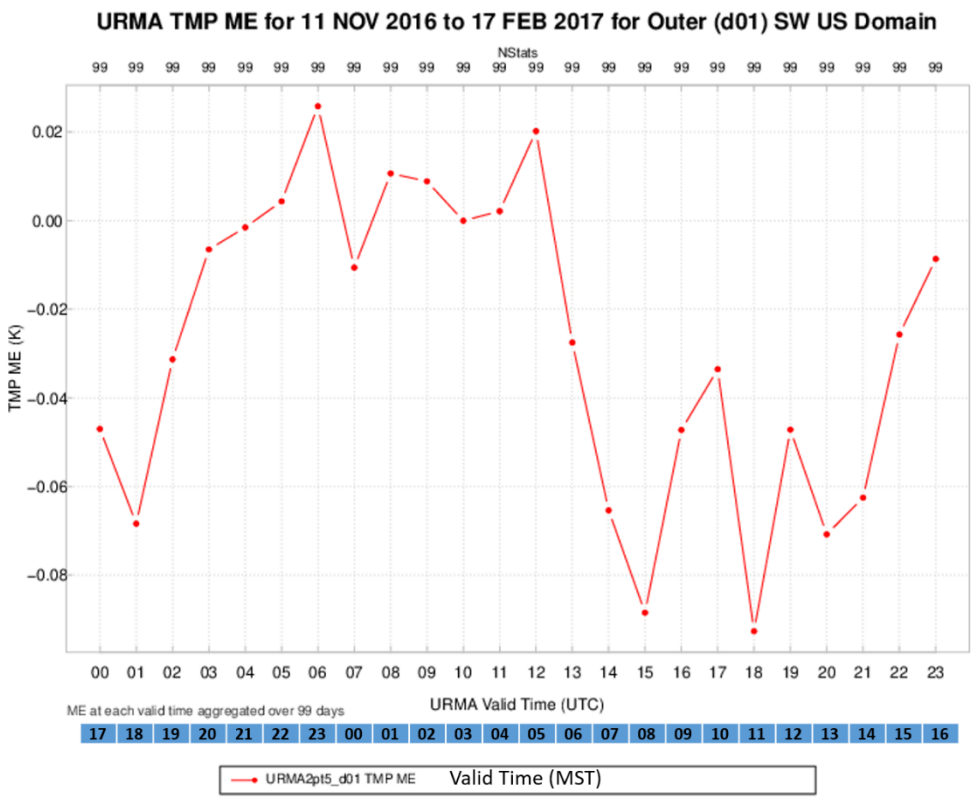

Fig. 4 Aggregated values of URMA TMP ME for each valid time for the 99-day period

The ME varies from mostly positive from 2200 MST to 0500 MST, indicating a tendency for URMA to slightly overestimate the TMP during nighttime to negative at all other hours indicating a slight underestimate of TMP during daytime. The period of best accuracy (ME close to zero) is during the period from 0300 UTC to 1300 UTC. The magnitude of the bias over the course of the day is very low, showing good agreement with the point observations. Figure 5 shows the RMSE statistic for TMP.

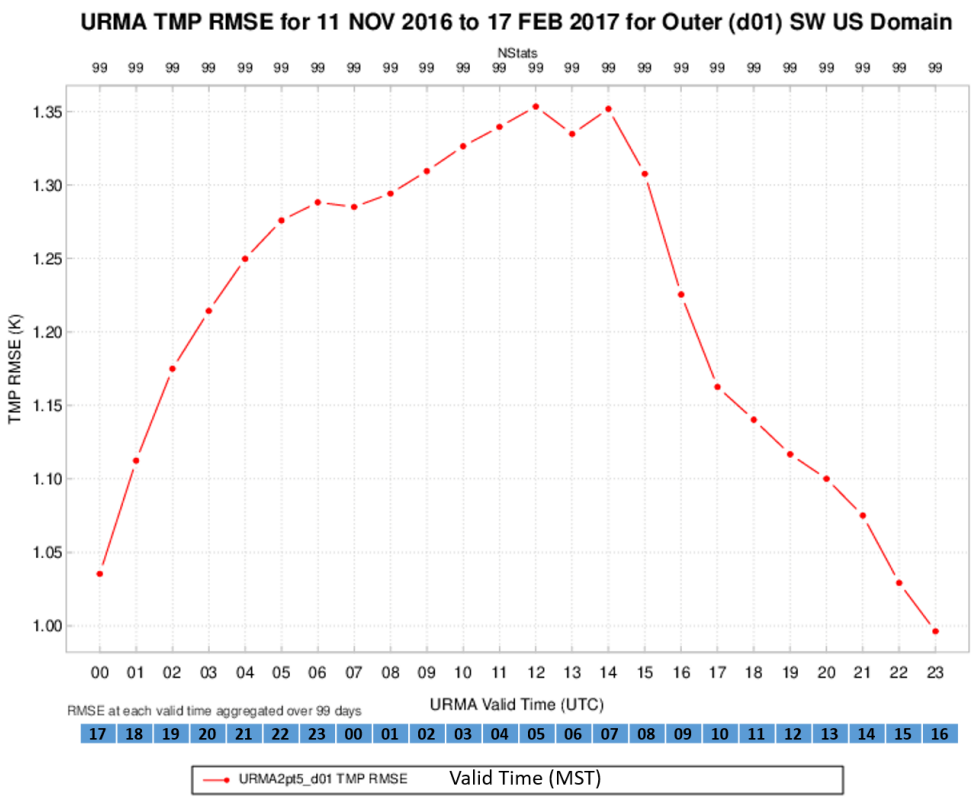

Fig. 5 Aggregated values of URMA TMP RMSE for each valid time for the 99-day period

The RMSE varies from slightly above 1.0 K to 1.35 K then back to 1.0 K, at the end of the diurnal period. The near-minimum values are in the late afternoon between 1500 and 1700 MST and the near-maximum values occur in the early morning between 0200 and 0800 MST. Overall, these RMSE values show good agreement between URMA and observed TMP.

The graphics showing the errors for 2-m-AGL DPT are presented in Figs. 6–8. Figure 6 shows the aggregated mean values of URMA (red points) and observed (green points) DPT for each valid time of the 99-day period.

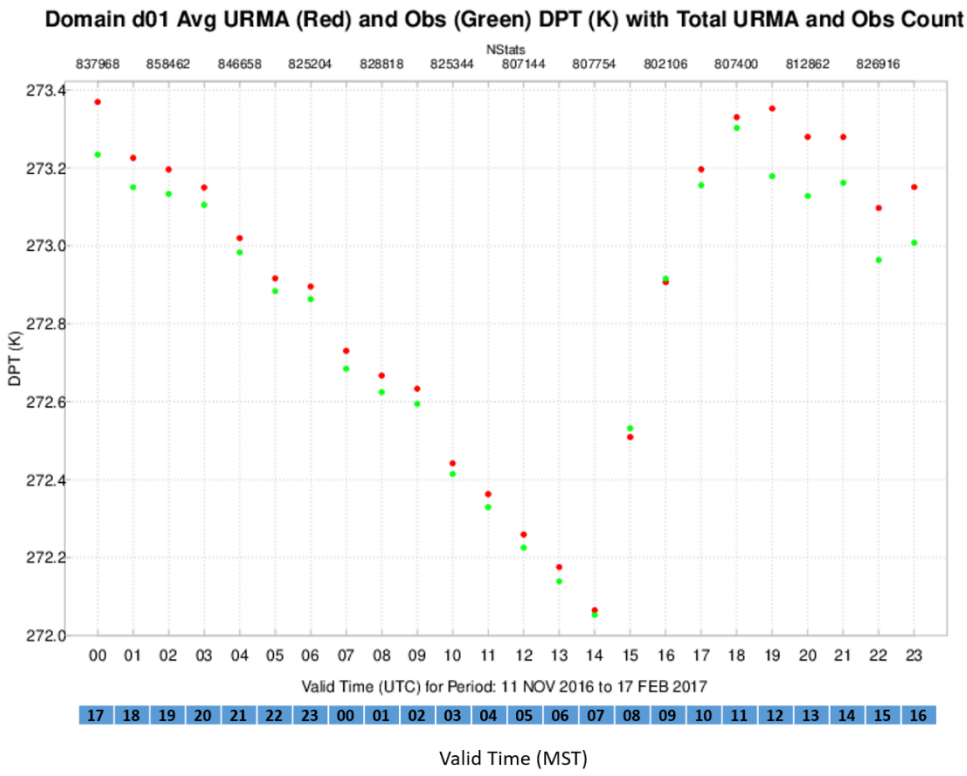

Fig. 6 Aggregated values of URMA (red points) and observed (green points) DPT for each valid time for the 99-day period. Numbers of matched pairs used for all error statistics obtained by dividing "NStats" value by 2.

The DPT varies diurnally like TMP and the errors are very small as shown by the proximity of the two data points at each valid time, but not quite as small as those for TMP. The numbers of DPT matched pairs, which were used to compute the aggregated error statistics, were obtained by dividing the values located along the x-axis at the top of the plot, by two. Figure 7 shows the ME or bias statistic for DPT.

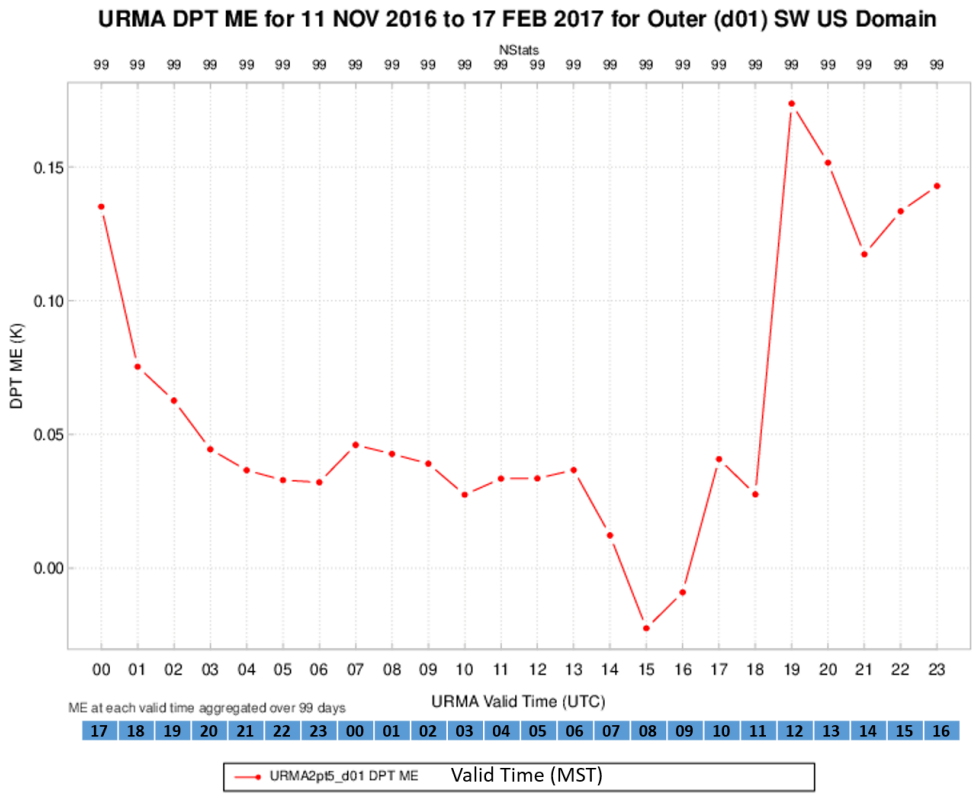

Fig. 7 Aggregated values of URMA DPT ME for each valid time for the 99-day period

The ME varies diurnally in a pattern distinct from that of TMP. There is a tendency to slightly overestimate the DPT for most of the day except for a brief period between 0800 and 0900 MST, where there is a very small underestimate of DPT. The period of best accuracy with ME close to zero is during the morning period from 0200 UTC to 1000 MST, but the magnitude of the bias over the course of the day is very low, showing good agreement with the point observations. Figure 8 shows the RMSE statistic for DPT.

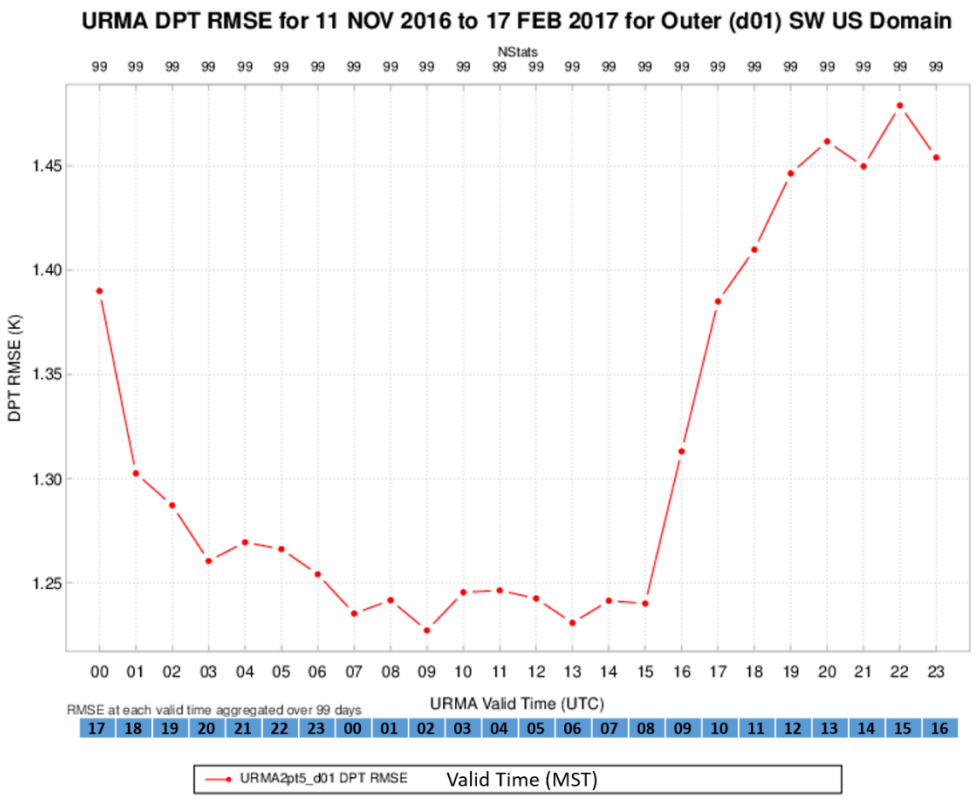

Fig. 8 Aggregated values of URMA DPT RMSE for each valid time for the 99-day period

The DPT RMSE varies diurnally in a pattern, which is the opposite of that for TMP. The near-minimum RMSE values of 1.20 K to 1.25 K occur in the early morning 0000 to 0800 MST and the near-maximum values of 1.39 K to 1.49 K occur in the afternoon 1200 to 1700 MST. Overall, these RMSE values show good agreement between URMA and the observed DPT.

The graphics showing the errors for 10-m-AGL UGRD are presented in Figs. 9–11. Figure 9 shows the aggregated mean values of URMA (red points) and observed (green points) UGRD for each valid time for the 99-day period.

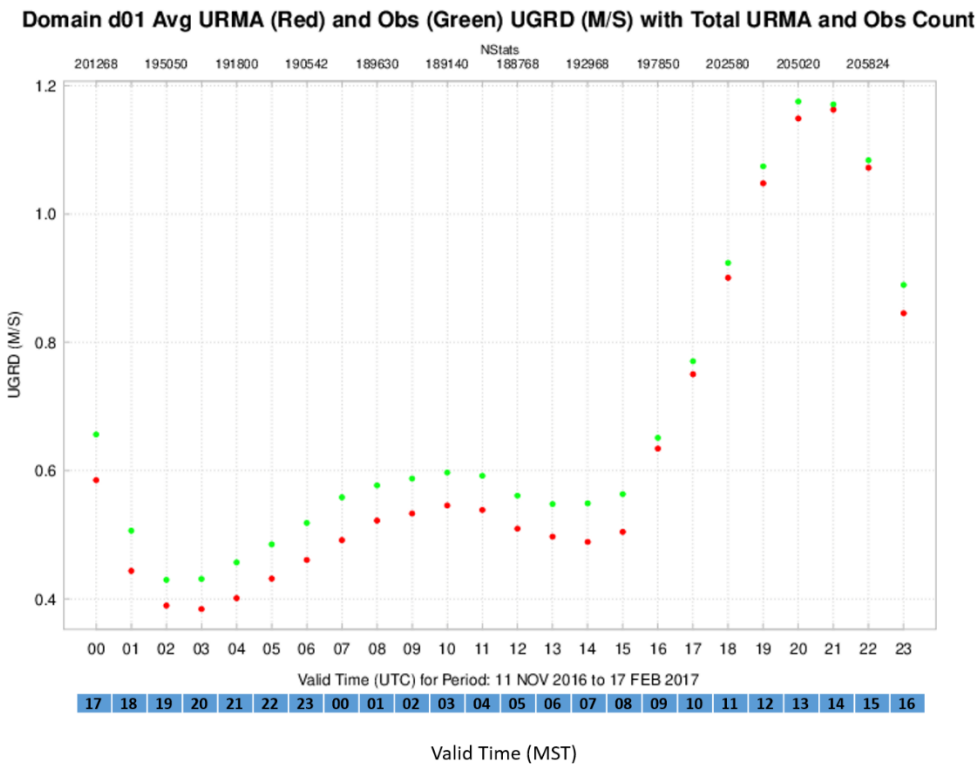

Fig. 9 Aggregated values of URMA (red points) and observed (green points) UGRD for each valid time for the 99-day period. Numbers of matched pairs used for all error statistics obtained by dividing "NStats" value by 2.

The UGRD varies diurnally in a pattern unlike TMP and DPT with the minimum values for the day being between 1700 and 0900 MST with a decided peak in magnitude during the daytime between 1000 and 1600 MST associated with the maximum in surface heating and planetary boundary layer (PBL) turbulent mixing. The errors are quite small, as shown by the proximity of the two data points at each valid time, with a tendency for slightly smaller errors during the peak in the magnitude of UGRD. The numbers of UGRD matched pairs, which were used to compute the aggregated error statistics, were obtained by dividing the values located along the x-axis at the top of the plot, by two. Figure 10 shows the ME or bias statistic for UGRD.

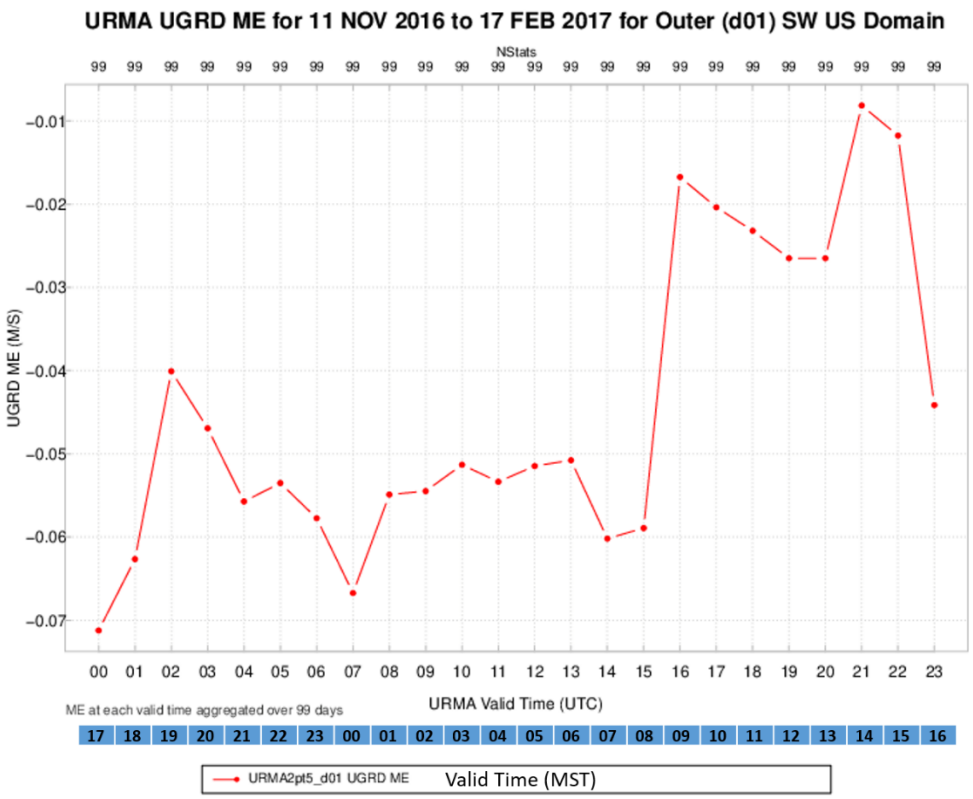

Fig. 10 Aggregated values of URMA UGRD ME for each valid time for the 99-day period

The UGRD ME varies diurnally in a pattern with the largest magnitude occurring between 1700 and 0800 MST with a decided trend toward less magnitude for the remainder of the day. The general tendency is for underestimating the magnitude of UGRD, but the accuracy compared to the observations is very good overall. Figure 11 shows the RMSE statistic for UGRD.

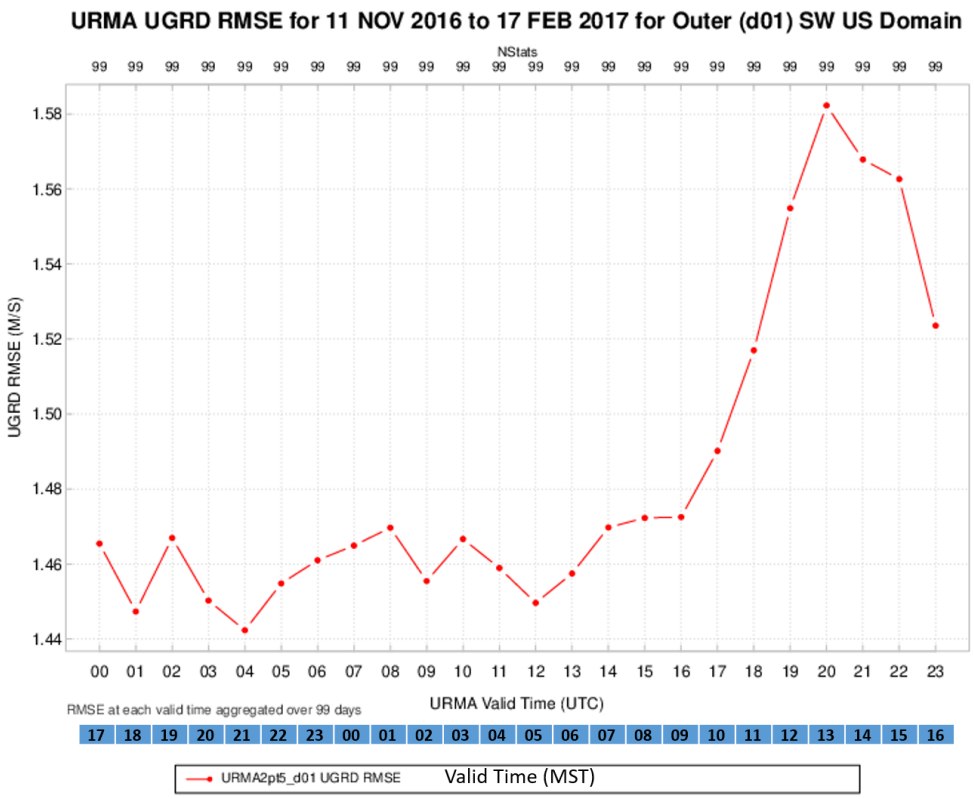

Fig. 11 Aggregated values of URMA UGRD RMSE for each valid time for the 99-day period

The UGRD RMSE varies diurnally in a pattern similar to DPT, with the near-minimum values of approximately 1.44 M/S to nearly 1.48 M/S occurring between 1700 to 0900 MST, and the near-maximum values of 1.50 M/S to 1.59 M/S occurring during the daytime 1000 to 1600 MST. These values for UGRD RMSE are quite good and show reasonable agreement with the observations.

The graphics showing the errors for 10-m-AGL VGRD are presented in Figs. 12–14. Figure 12 shows the aggregated mean values of URMA (red points) and observed (green points) VGRD for each valid time for the 99-day period.

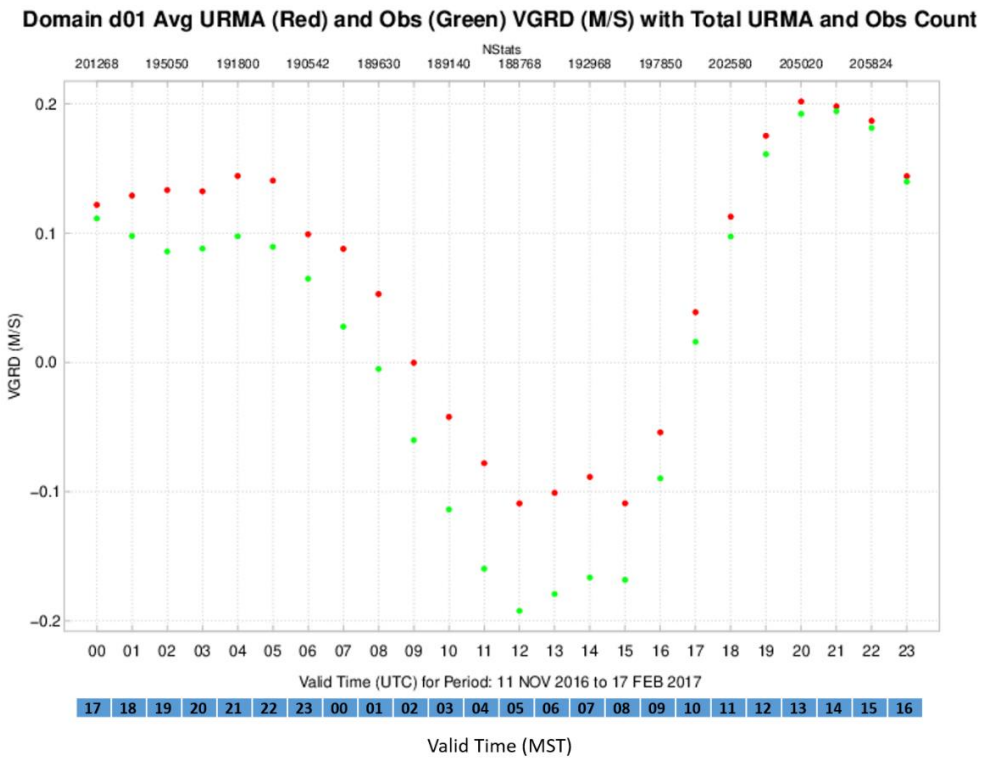

Fig. 12 Aggregated values of URMA (red points) and observed (green points) VGRD for each valid time for the 99-day period. Numbers of matched pairs used for all error statistics obtained by dividing "NStats" value by 2.

The VGRD varies diurnally in a pattern quite different than UGRD. Southerly component winds dominate during the period 1000 and 0000 MST and switch to northerly component winds during the period 0100 to 0900 MST. It is also noteworthy that the overall magnitude of VGRD is significantly smaller than that of UGRD and ranges between ±0.2 M/S. The errors are quite small, as shown by the proximity of the two data points at each valid time, with a tendency for slightly smaller errors during the period 1000 to 1700 MST. However, the errors for period 1800 to 0900 MST are slightly larger than the UGRD. This may be attributable to the relatively low VGRD magnitude as well as UGRD, which puts into question the accuracy of the observations themselves and, consequentially, the URMA values. The numbers of VGRD matched pairs, which were used to compute the aggregated error statistics, are obtained by dividing the values located along the x-axis at the top of the plot, by two. Figure 13 shows the ME or bias statistic for VGRD.

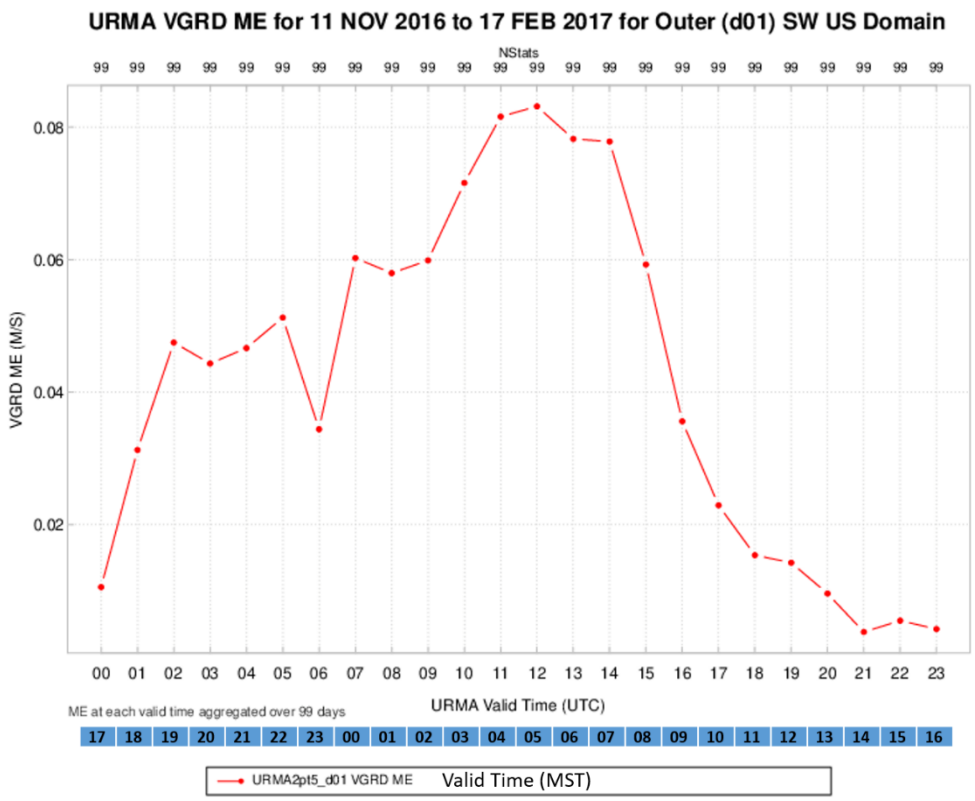

Fig. 13 Aggregated values of URMA VGRD ME for each valid time for the 99-day period

The VGRD ME varies diurnally in a pattern with the peak magnitude occurring between 0000 and 0800 MST. A decided trend toward less magnitude is observed during the daytime period 1000 to 1700 MST. The general tendency is for overestimating the magnitude of VGRD. However, the overall accuracy compared to the observations is very good. Figure 14 shows the RMSE statistic for VGRD.

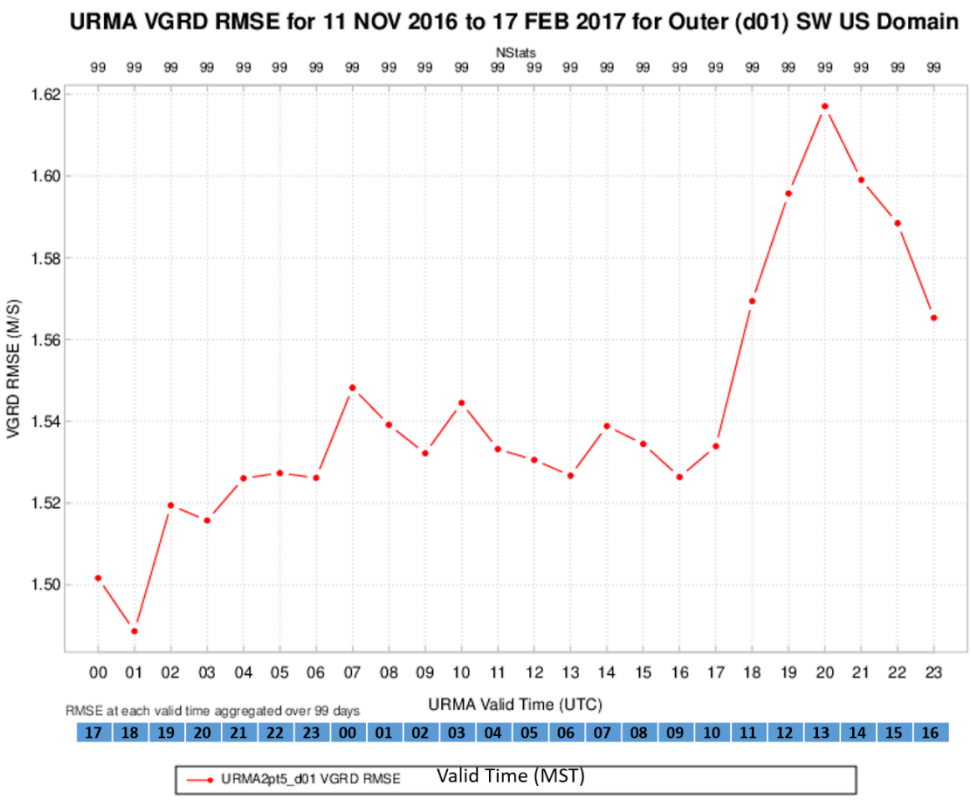

Fig. 14 Aggregated values of URMA VGRD RMSE for each valid time for the 99-day period

The VGRD RMSE varies diurnally in a pattern similar to UGRD with the near-minimum values of approximately 1.48 M/S to nearly 1.55 M/S occurring between 1700 to 1000 MST; and the near-maximum values of 1.57 M/S to 1.61 M/S occurring between 1100 to 1600 MST. These values for RMSE appear quite good, show reasonable agreement with the observations, and are consistent with the RMSE values of UGRD, except for being slightly larger than those for UGRD. This is probably attributable to the very low URMA and observed-VGRD magnitudes during the study period.

The graphics showing the errors for 10-m-AGL WIND are presented in Figs. 15–17. Figure 15 shows the aggregated mean values of URMA (red points) and observed (green points) WIND for each valid time for the 99-day period.

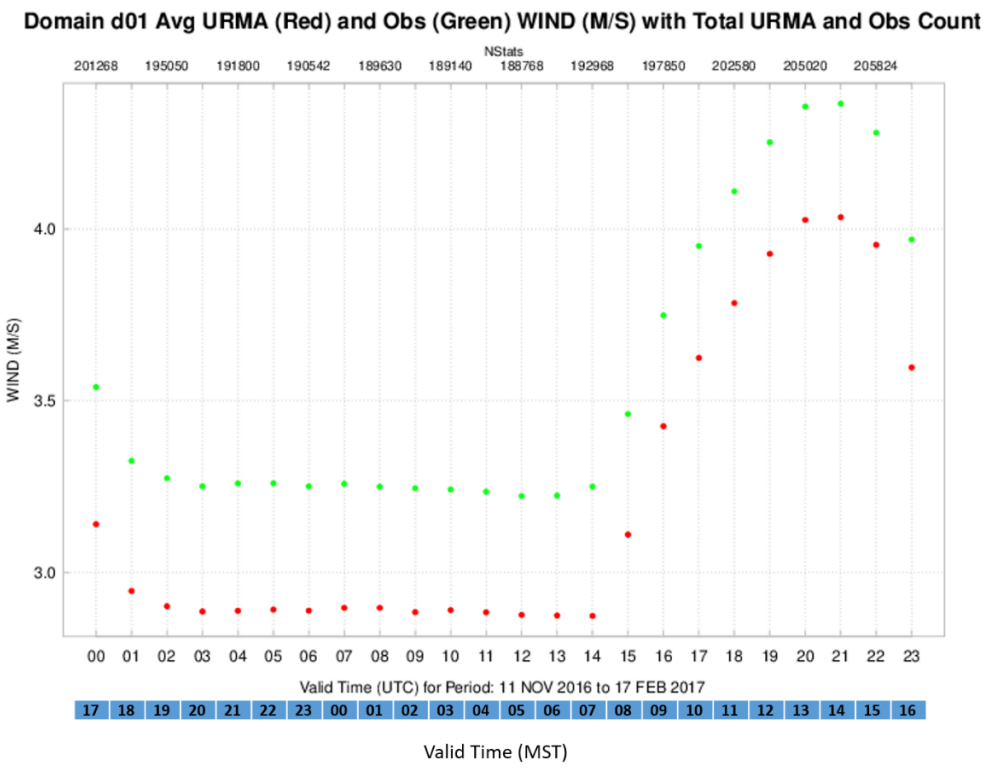

Fig. 15 Aggregated values of URMA (red points) and observed (green points) WIND for each valid time for the 99-day period. Numbers of matched pairs used for all error statistics obtained by dividing "NStats" value by 2.

The WIND varies diurnally in a pattern very similar to UGRD. This similarity is likely due to the dominant magnitude of the UGRD component, as compared to that of VGRD, during the study period. The minimum values for the day occurred between 1700 and 0800 MST, with a decided peak in magnitude during the daytime between 0900 and 1600 UTC associated with the maximum in surface heating and PBL turbulent mixing. The errors, as indicated by the spacing between the two data points at each valid time, appear larger than was the case for UGRD and VGRD, owing to the additive affect when the error from both components is combined. The numbers of WIND matched pairs, which were used to compute the aggregated error statistics, were obtained by dividing the values located along the x-axis at the top of the plot, by two. Figure 16 shows the ME or bias statistic for WIND.

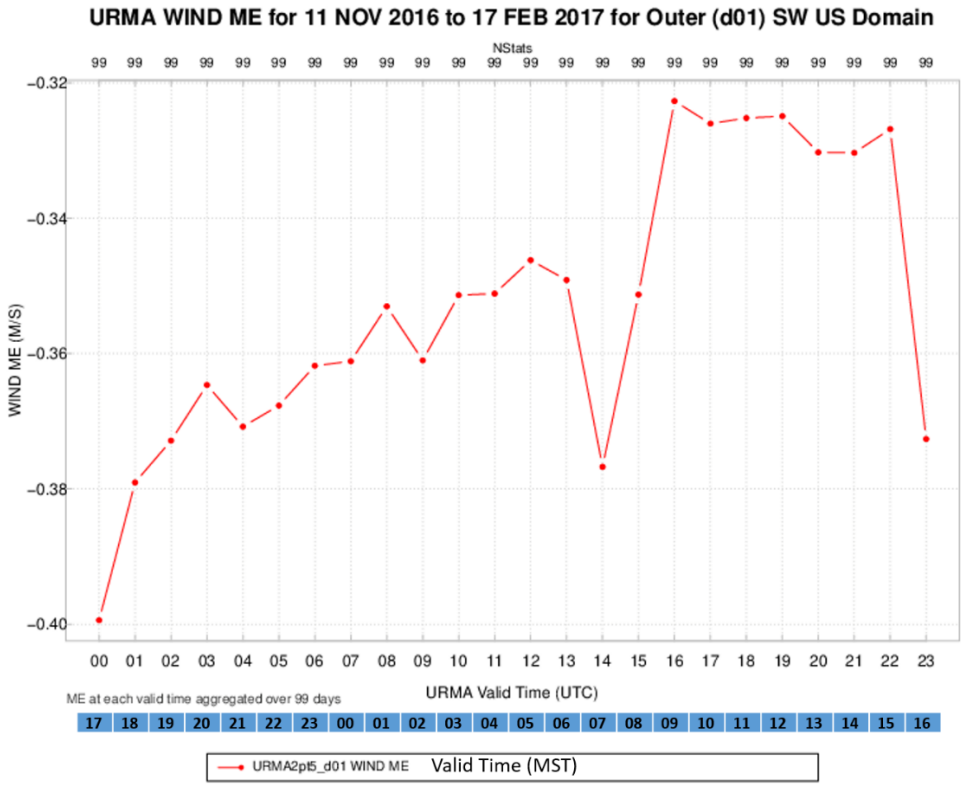

Fig. 16 Aggregated values of URMA WIND ME for each valid time for the 99-day period

The WIND ME varies diurnally in a pattern resembling that of UGRD with the largest magnitude occurring between 1600 and 0800 MST, with a decided trend toward less magnitude for the remainder of the day. The general tendency is for the URMA to underestimate the magnitude of WIND, but the accuracy compared to the observations is very good overall. Figure 17 shows the RMSE statistic for WIND.

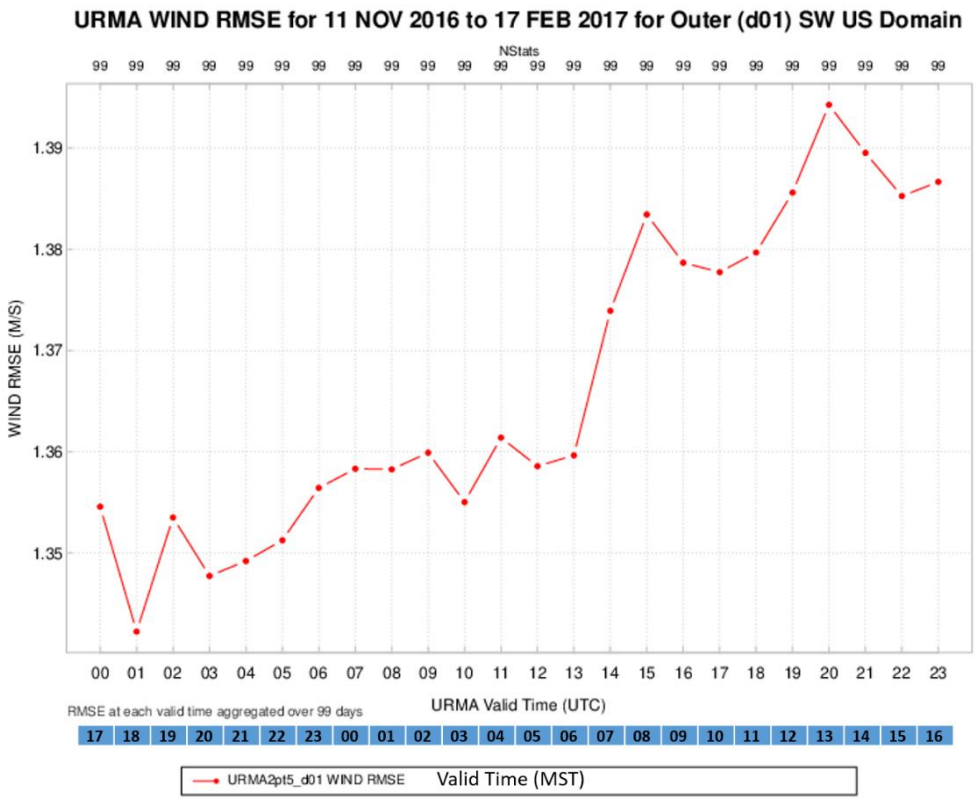

Fig. 17 Aggregated values of URMA WIND RMSE for each valid time for the 99-day period

The WIND RMSE varies diurnally in a pattern similar to UGRD; with the near-minimum values of approximately 1.34 M/S to nearly 1.36 M/S occurring during the evening and nighttime between 1700 to 0600 MST, and the near-maximum values of 1.37 M/S to 1.40 M/S occurring during the daytime from 0700 to 1600 MST. These values for RMSE are quite good and show reasonable agreement with the observations.

The graphics showing the errors for 10-m-AGL WDIR are presented in Figs. 18–21. Since WDIR is a circular variable with the range of values between 0 and 360, different methods were used to present the results of the evaluation. The aggregations were performed on vector partial sums instead of scaler partial sums to compute the error statistics, which involve angular differences rather than simple scalar differences.

Figure 18 shows the aggregated mean values of URMA (red points) and observed (green points) WDIR for each valid time of the 99-day period.

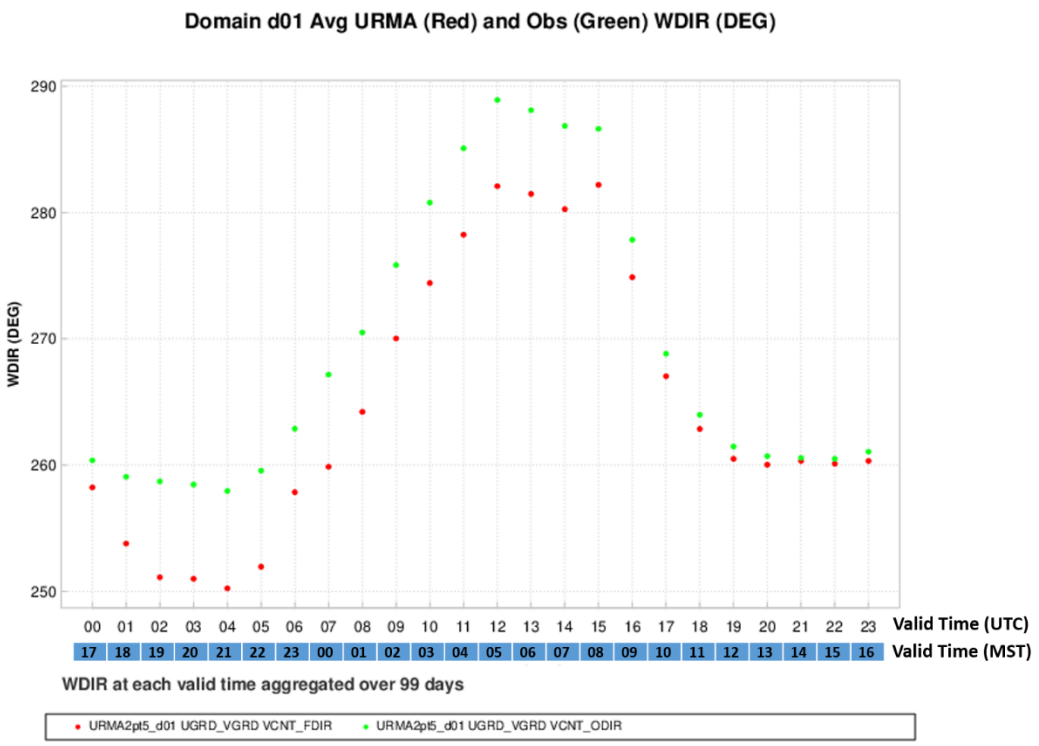

Fig. 18 Aggregated values of URMA (red points) and observed (green points) WDIR for each valid time for the 99-day period. Numbers of matched pairs used for all error statistics not shown, but are the same as for WIND.

The WDIR varies diurnally with southwest winds during the daytime and evening between 1000 and 0100 MST, becoming northwest winds during the early morning hours between 0200 and 0900 MST. This pattern may coincide with the evolution of terrain-induced nighttime and perhaps more northerly drainage flows prevailing over those from other directions. These flows are typical of the inter-mountain basins and valleys, which onset soon after sunset followed by a shift to more southwesterly winds, which develop with increased daytime heating and resultant PBL mixing down of stronger southwesterly winds from aloft. This pattern in combination with the increased and seasonal presence of mid-latitude wintertime storm systems has the net effect of bringing cooler air from north to south over the domain favored during nighttime decoupling from southwesterly flow aloft. The numbers of WDIR matched pairs, which were used to compute the aggregated error statistics, are not shown in this plot, but they were the same as those for the other wind variables. Figure 19 shows the bias for WDIR.

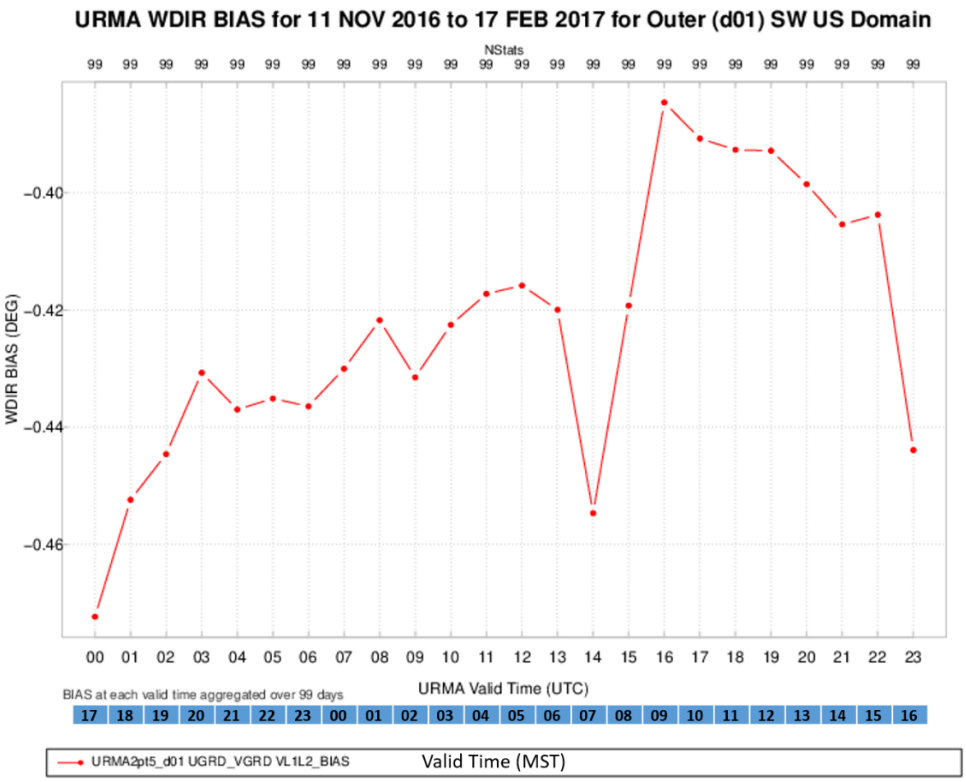

Fig. 19 Aggregated values of URMA WDIR_BIAS for each valid time for the 99-day period

The bias statistic (WDIR_BIAS) is computed by aggregating multiple vector partial sums for both wind components, and computing their weighted average where the weight is determined by the numbers of matched pairs. The mean URMA and observed wind directions are computed, as well as their difference, which is the bias error.

The WDIR_BIAS shows an overall tendency for the URMA to analyze a slightly more southwesterly wind, as compared to the observed wind. This tendency is slightly less prevalent during the hours of 0900 to 1500 MST, but the magnitude of this difference is minimal. The accuracy compared to the observations is very good overall. Figure 20 shows the DIR_ERR for WDIR.

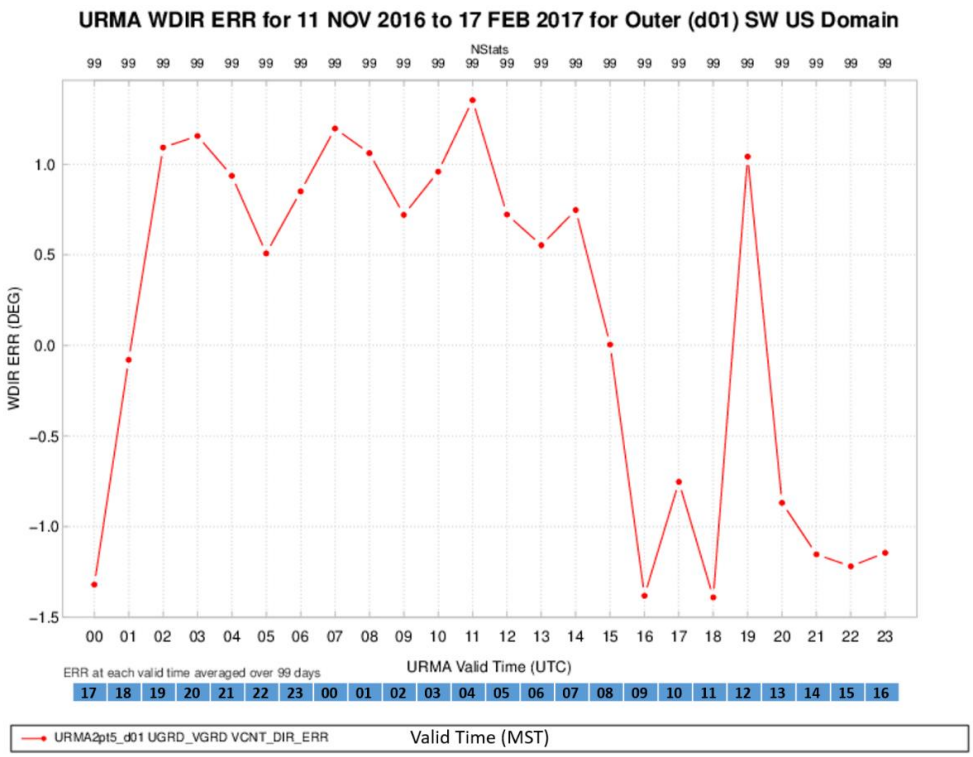

Fig. 20 Aggregated values of URMA WDIR_ERR for each valid time for the 99-day period

DIR_ERR is a vector continuous error statistic computed by taking the signed angle between the directions of the averaged forecast and observed wind vectors. The value is positive when the URMA wind vector is counterclockwise from the observed wind vector. Because this error involves assigning positive and negative values for the angle between the URMA direction and the observed direction in the counterclockwise direction, there are positive and negative error values depending on whether the URMA direction is counterclockwise or clockwise, respectively.

Figure 20 shows that the URMA direction is predominantly counterclockwise (positive) relative to the observed direction during the nighttime and early morning period 1900 to 0800 MST, and mostly clockwise (negative) during the daytime between 0900 and 1800 MST. One notable exception occurs at 1900, where there is a reversal in this pattern. Overall, the accuracy compared with the observations is very good. Figure 21 shows the DIR_ABSERR for WDIR.

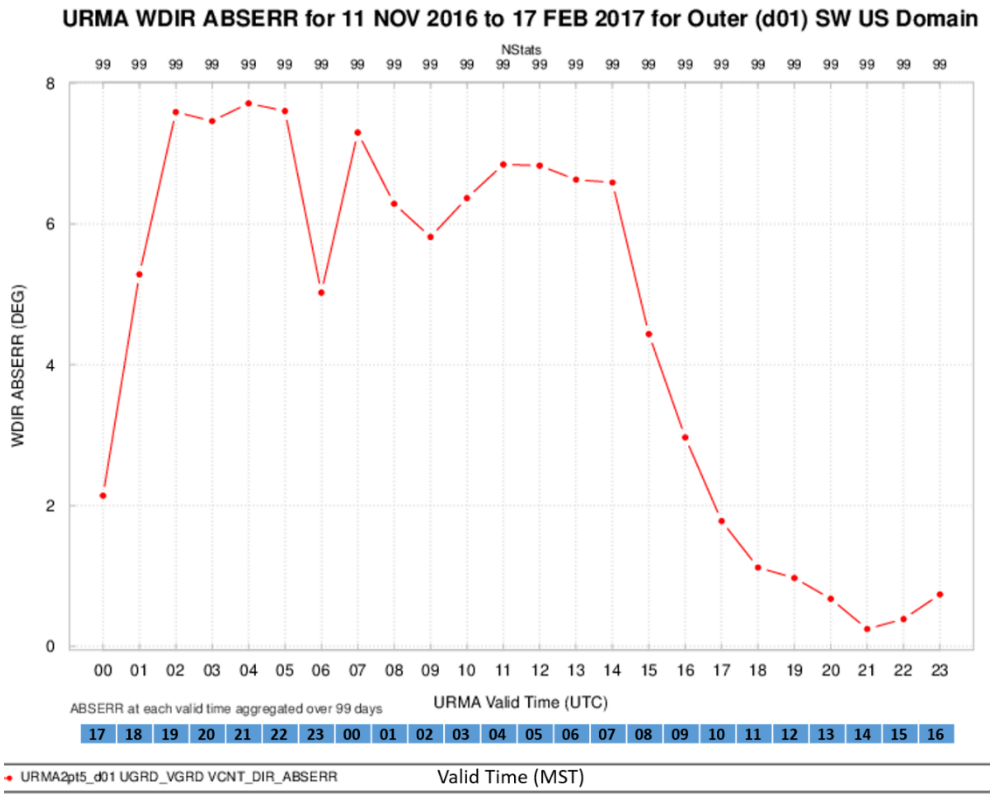

Fig. 21 Aggregated values of URMA WDIR_ABSERR for each valid time for the 99-day period

DIR_ABSERR is the absolute value of the signed angle between the directions of the averaged forecast and observed wind vectors (DIR_ERR). It is a statistic that quantifies the magnitude of the error regardless of the sign. The error is largest during the nighttime and early morning period, namely, between 1800 and 0800 MST. The error is also smallest during the daytime period 0900 to 1700 MST. The period of the smallest errors correlates well with the period of the maximum WIND (speed) values. Likewise, the period of largest errors corresponds with the period of minimum WIND values. Some of this error is probably attributable to the tendency for increased wind measurement errors during the nighttime period, with associated lighter wind speeds that degrades the ability to compute WDIR accurately. Overall errors of this magnitude are very reasonable and show good agreement with observations.

5. Summary and Conclusion

The goal of this evaluation was to establish the validity of URMA gridded observations for use as ground truth in conducting spatial verification of the Army's WRE-N Nowcast model; as well as to determine the accuracy of the forecast input

to decision aids, which predict the impacts of weather on military missions and systems. The evaluation used the grid-to-point verification method. This strategy employs point weather observations as ground truth for assessing the accuracy of the URMA analysis fields of meteorological variables. The data used for the evaluation consisted of URMA gridded observations and point weather observations, which were collected for a 99-day "winter" period from 11 November 2016 to 17 February 2017. The domain for which these data were collected was a large area over the southwestern United States, and Northern Mexico, including portions of the Gulf of Mexico, Sea of Cortez, and the Eastern Pacific Ocean. This region is characterized by terrain types ranging from complex mountain-desert-basin topography to oceanic areas. The MET Point-Stat tool was used to perform the verification, which involved the ingest of hourly URMA analyses and point observations and the paring of URMA and measured values of several near-surface meteorological variables for each observing station. The Point-Stat tool computed the differences between the URMA and observed variables at each location and generated partial sums data from which error statistics were calculated. The METViewer tool was used to ingest the output of Point-Stat, aggregate the partial sums data, and generate the error statistics and graphics, which display them as 24-h time series. The error statistic plots were analyzed to quantify the statistical error attributable to URMA. The accuracy of the URMA analyses varied diurnally with the expected patterns depending on the particular variable. Based on the analysis of error statistics, in general, the accuracy of the URMA gridded observations was considered very good for use as ground truth for spatial verification. It should be noted here that the error statistics presented may suffer from the same systematic issues previously described. These issues center on the inherent difficulties when using grid-to-point verification of gridded fields at high resolution. Two caveats about these results are 1) the point observations used as ground truth were to a great extent the same as those which were assimilated into URMA and 2) the evaluation was conducted only at the locations where point observations were available. The first caveat restricts the scoring to values from the URMA grid, which had the benefit of a nearby point observation, and thus would tend to be significantly influenced by the value of the observation during assimilation. Regarding the second caveat, although the numbers of matched pairs used in this evaluation was large, their density over the entire domain is not sufficient to adequately characterize the variable at every location within the URMA grid. As such, for verification of high-resolution NWP such as the WRE-N at 1-km or less grid spacing, there may be limitations on the quality of the URMA analyses in areas where there are no point observations available for assimilation. To overcome this limitation, it would be advantageous to exploit mesonets, remotely sensed measurements such as those available from satellites, and

specialized sources of meteorological measurements such as the Army's Meteorological Sensor Array (MSA) located at White Sands Missile Range, New Mexico, whose observations were not included in the URMA analyses, for the purpose of conducting cross-validation studies to characterize the accuracy of URMA in areas where no observations were assimilated. Although the calculated evaluation errors would be propagated into any NWP model assessment where URMA is used as ground truth, and given the above-mentioned caveats regarding the computed URMA error quality, the value of using URMA to enable spatial verification for high-resolution forecasts is judged to be reasonably good and justifiable in view of the absence of any other alternative sources of gridded observations, pending the results of future cross-validation studies.

6. References

Benjamin S, et al. A North American hourly assimilation and model forecast cycle: The Rapid Refresh. Mon Wea Rev. 2016;144:1669–1694.

Dawson L, Raby J, Smith J. The automation of nowcast model assessment processes. Adelphi (MD): Army Research Laboratory (US); 2016 Sep. Report No.: ARL-MR-0940.

De Pondeca Manuel SFV, Manikin G, DiMego G, Benjamin S, Parrish D, Purser RJ, Wu WS, Horel J, Myrick D, Lin Y, et al. The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: current status and development. Weather Forecast. 2011;26:593–612.

Ebert E. Fuzzy verification of high resolution gridded forecasts: a review and proposed framework. Meteorol Appl. 2008;15:51–64.

Jolliffe IT, Stephenson DB. Forecast verification: a practitioner's guide in atmospheric science. 2nd ed. Hoboken (NJ): John Wiley and Sons; 2012.

Newman K, Jensen T, Brown B, Bullock R, Fowler T, Halley Gotway J. The model evaluation tools v8.1 (METv8.1) user's guide. Developmental Testbed Center; 2018 [accessed 2020 Nov 9]. https://dtcenter.org/sites/default/files/community-code/met/docs/user-guide/MET_Users_Guide_v8.1.1.pdf.

Morris M, Carley J, Colon E, Gibbs A, De Pondeca M, Levine S. A quality assessment of the real-time mesoscale analysis (RTMA) for aviation. Weather Forecast. 2020;35:977–996

Pondeca M, Levine S, Carley J, Lin Y, Zhu Y, Purser J, McQueen J, Yang R, Gibbs A, Parrish D, DiMego G. Ongoing improvements to the real time mesoscale analysis (RTMA) and unrestricted mesoscale analysis (URMA) and NCEP/EMC. World Climate Research Organization; 2015 [accessed 2020 Nov 6]. https://www.wcrp-climate.org/WGNE/BlueBook/2015/individual-articles/01_Pondeca_Manuel_etal_RTMA.pdf.

[NCAR] METViewer. Boulder (CO): National Centers for Atmospheric Research [accessed 2020 Sep 27]. https://dtcenter.org/sites/default/files/community-code/met/docs/presentations/met-tutorial-20180131/19_METViewer_Jan18.pdf.

[NOAA] Meteorological assimilation data ingest system (MADIS). College Park (MD): National Oceanic and Atmospheric Administration; 2014 [accessed 2016 Jul 27]. http://madis.noaa.gov.

Raby J, Smith J. Domain-level assessment of the weather running estimate-nowcast (WRE-N) model. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Nov. Report No.: ARL-TR-7881.

Reen B, Dawson L. The weather running estimate – nowcast realtime (WREN_RT) system, version 1.03. Adelphi (MD): Army Research Laboratory (US); 2018 Sept. Report No.: ARL-TR-8533.

Ruth D, Huntemann T, Plumb D. Verification of the national blend of models. 28th Conference on Weather Analysis and Forecasting/24th Conference on Numerical Weather Prediction; 2017; Seattle, WA. Amer Meteor Soc. 7B.3. https://ams.confex.com/ams/97Annual/webprogram/Paper305573.html.

[UCAR] Operational models encyclopedia. Boulder (CO): University Corporation for Atmospheric Research; 2015 [accessed 2020 Sep 27]. https://sites.google.com/a/ucar.edu/model-encyclo-determ/deterministic/analyses/rtma-urma.

Weygandt S, Alexander C, Ge G, Hu M, Ladwig T, Hartsough C, Carley J, Zhao G, Pondeca M, Yang R. Evaluation of a Prototype Version of the 3D-Real-Time Mesoscale Analysis (3D-RTMA) for Situational Awareness and Nowcast Applications. 35th Conference on Environmental Information Processing Technologies; 2019 Jan 8; Phoenix, AZ [accessed 2020 Oct 12]. https://ams.confex.com/ams/2019Annual/webprogram/Paper353171.html.

List of Symbols, Abbreviations, and Acronyms

2-D	two-dimensional
2dvar	two-dimensional variational data assimilation
3-D	three-dimensional
AGL	above ground level
CONUS	continental United States
DEG	degrees
DEVCOM	US Army Combat Capabilities Development Command
DIR_ERR	wind direction error
DIR_ABSERR	wind direction absolute error
DOE	United States Department of Energy
DPT	dew-point temperature
HRRR	High Resolution Rapid Refresh
K	Kelvin
M/S	meters per second
MADIS	Meteorological Assimilation Data Ingest System
ME	mean error
MET	Model Evaluation Tools
METAR	Météorologique Aviation Régulière
MSA	Meteorological Sensor Array
NAM	North American Mesoscale
NCAR	National Centers for Atmospheric Research
NCEP	National Center for Environmental Prediction
NESDIS	National Environmental, Satellite, and Data Information Service
NOAA	National Oceanic and Atmospheric Agency
NSF	National Science Foundation
NWP	Numerical Weather Prediction
NWS	National Weather Service
PBL	planetary boundary layer

QA	quality assurance
RAP	Rapid Refresh
RMSE	root-mean-square error
RTMA	Real Time Mesoscale Analysis
STAT	tabular ascii data format
TMP	temperature
UGRD	U wind component
URMA	Unrestricted Mesoscale Analysis
USAF	United States Air Force
UTC	Coordinated Universal Time
VGRD	V wind component
WDIR	wind direction
WDIR_BIAS	wind direction bias
WIND	wind speed
WRE-N	Weather Running Estimate – Nowcast
WREN_RT	Real-Time implementation of the Weather Running Estimate – Nowcast

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLD DCI TECH LIB
5 (PDF)	DEVCOM ARL FCDD RLC EM R DUMAIS H CAI J RABY L DAWSON B REEN