

## 16B.3 MODIFICATION OF THE UNIFIED POST PROCESSOR (UPP) VERSION 3 TO ACHIEVE IMPROVED ACCURACY FOR MODEL VERIFICATION

John W. Raby<sup>1\*</sup>, Brian P. Reen<sup>2</sup>, Jeffrey A. Smith<sup>1</sup>

<sup>1</sup> US Army Research Laboratory, White Sands Missile Range, New Mexico

<sup>2</sup> US Army Research Laboratory, Adelphi, Maryland

### 1. INTRODUCTION

To support the U.S Army operations rapid-update cycling nowcast numerical weather predictions are required. The Advanced Research Weather Research and Forecasting (WRF-ARW) model (Skamarock et al., 2008) has been applied by the Army with the goal of providing gridded forecast output for use in Army mission execution, artillery and aviation. The WRF-ARW's Four-Dimensional Data Assimilation (FDDA) method (Deng et al., 2009) has been used to integrate and apply observations into the model's initialization processes. This tailored version of the WRF-ARW is called the Weather Running Estimate-Nowcast (WRE-N) modeling system. The WRE-N is designed to run up to hourly, producing 3-6 hour forecasts per cycle at horizontal grid point resolutions of several km to 500m to produce forecasts of tactically significant variables in the atmosphere with a focus on the Planetary Boundary Layer (PBL).

The WRE-N is initialized using the WRF Preprocessing System (WPS) which processes the observations for their use in the assimilation phase of the WRE-N. The Ungrib module of the WPS converts the coarse grid model (CGM) output from GRIB to an intermediate format and interpolates it to provide the field which serves as the background against which the observations are compared during the Quality-Control (QC) process which is run by the Obsgrid module. Observations from radiosonde soundings which have measurements at many levels are interpolated vertically on to the background field and thus acquire the resolution of the background field which result in a loss of vertical resolution when compared to the original raw observation. The output of Obsgrid is observations which have been QC'd using gross error checks, buddy checks and background error checks. The resulting observations are used for verification and would be used for data assimilation if it were enabled.

The assessment of model accuracy relies on the ability, or lack thereof, to generate verification statistics that accurately compare the model output to actual observations. One challenge which makes this difficult is the method by which the model forecast values are matched to the observations for verification. Common practice is to interpolate the model values in 3D space from the raw model output on the staggered native

vertical grid to a de-staggered grid with the output on pressure surfaces. A tool which performs this is the Unified Post Processor (UPP) which was developed by The National Centers for Environmental Prediction. The UPP is described in the User's Guide online at: [http://www.dtccenter.org/upp/isers/docs/user\\_guide/V3/upp\\_users\\_guide.pdf](http://www.dtccenter.org/upp/isers/docs/user_guide/V3/upp_users_guide.pdf) For this project, the ARL investigators used Version 3 of the UPP to post-process the output of the WRE-N.

For verification, ARL uses the Model Evaluation Tools (MET) which was developed by the National Center for Atmospheric Research (NCAR). MET ingests the output of the UPP in GRIB format to access the model forecast data. The Point-Stat tool is used to perform grid-to-point verification of the post-processed WRE-N output against observations. In order to extract the forecast values at the pressure level of the observations, Point-Stat interpolates the model data from the UPP pressure levels to the pressure level of the observation and then generates matched pairs of forecast and observations to be used to compute error statistics.

### 2. STANDARD PROCESS FLOWS FOR OBSERVATION QC AND UPP POST-PROCESSING OF WRE-N OUTPUT

The steps involved in the processes of post-processing using the UPP and observation QC are shown in Figure 1.

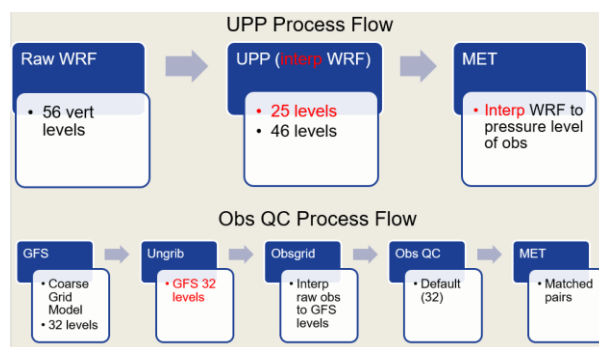


Figure 1. UPP and Observation QC Process Flows. Red text flags the parts of the process which present the challenges which impact the quality of model verification.

\* Corresponding author's address: John Raby, U.S. Army Research Laboratory, White Sands Missile Range, NM 88002-5501; email: john.w.raby2.civ@mail.mil

The process considered standard in this study for performing observation QC begins with ingest of the CGM which in this case is the GFS model output on 32 pressure levels as it is normally distributed. Ungrib is run to convert the GFS output to prepare the background field against which the observations are compared during the Quality-Control (QC) process which is run by the Obsgrid module. Obsgrid then interpolates the raw observations to the 32 pressure levels of the GFS data so that the observations can be QC'd against the GFS data. The resulting observations are then provided to MET to be used in the generation of matched pairs. The red text in Figure 1 flags the challenge faced when working with 32 levels which is the inability to sufficiently capture the vertical structure of the raw observations from radiosondes. Having more levels would be desirable because more of the vertical structure contained in the observations could be retained.

The process used by ARL for performing the post-processing of the WRE-N output (shown here as "WRF") begins with the generation by WRF of the model output in NetCDF format on native staggered vertical grids with 56 levels. The output files are ingested by the UPP which interpolates the model values in 3D space to a de-staggered grid with the output on pressure surfaces and GRIB formatted. The UPP produces output on 25 pressure levels based on the settings in the default configuration file, but a simple modification of the configuration files is all that is needed to enable output on all 46 levels. There is no user-friendly way to add additional levels beyond 46. The resulting output is then provided to MET. MET Point-Stat is run which interpolates the model output on pressure levels to generate the matched model-observation pairs. Point-Stat does this by interpolating the model output to the pressure level of the observation for all available observations. Both a horizontal and vertical interpolation are required to obtain the estimate of the model value at the location and pressure level of the observation. Note that the model output is interpolated twice – the first during the UPP post-processing and the second during the generation of matched pairs in MET.

### 3. CHALLENGES INHERENT IN THE USE OF THE DEFAULT UPP AND UNGRIB STEP

Figure 2 shows vertical profiles of the spacing between vertical levels for raw, 56-level model output at a high and a low elevation location and for 25 level UPP output.

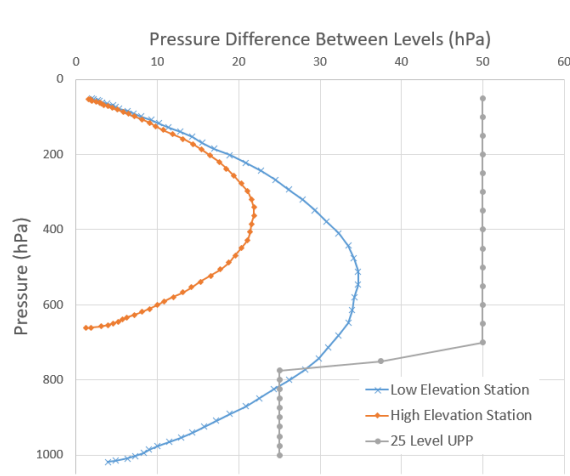


Figure 2. Vertical profiles of the spacing between vertical layers for raw WRE-N output at a high and a low elevation location and the UPP 25 level output.

Notice the high density of levels in the lowest part of the atmosphere (i.e., small pressure differences) for both the low and high elevation sites and how that compares with the density of the UPP pressure levels. The comparison is fair at the low elevation site and poor at the high elevation site. The thin layer around 800 hPa is the only place where UPP retains WRF vertical resolution and only for the low elevation station. This illustrates the deficiency of the UPP in representing the vertical structure of the model when configured to output on 25 pressure levels especially for high elevation stations.

Prior to WPSV3.9, Ungrib converted CGM data while retaining the pressure levels in the CGM data (e.g., 32 levels for GFS), which restricted the ability to retain the vertical structure of sounding data during the Obsgrid QC process. With the release of WPSV3.9 and subsequent versions, a new capability was added which enabled the user to specify interpolation of the CGM data to additional pressure levels.

### 4. MODIFICATIONS MADE TO THE UPP AND UNGRIB TO OVERCOME THE CHALLENGES

Given the above limitations imposed by the default configuration of the UPP and Ungrib software ARL decided to demonstrate methods that potentially improve the observation QC and UPP post-processing. Goals of this project were demonstrating better retention of the vertical structure of observations by using Ungrib to interpolate CGM to additional pressure levels and demonstrating better retention of the vertical structure of the model by increasing the number of pressure levels to which UPP interpolates.

Figure 3 shows the UPP and Observation QC Process Flows as modified to address the challenges and achieve the goals.

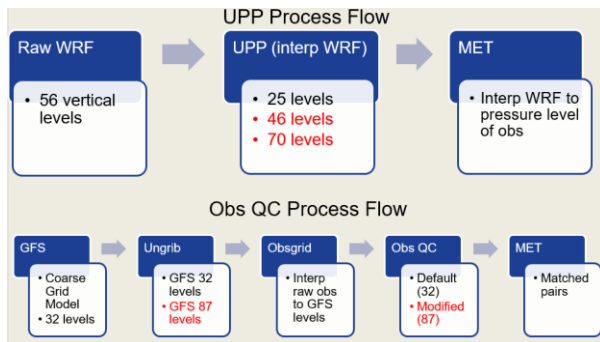


Figure 3. Modified UPP and Observation QC Process Flows. Red text flags the parts of the processes which were modified.

To achieve the goal of obtaining a better estimate of the model value to be used for verification, the UPP software was investigated to see if there was a way to increase the number of pressure levels it can output to achieve a better match in the overall density of levels to be closer to those present in the raw WRE-N output. A simple edit change of the UPP configuration file enabled output to be produced on all 46 default levels. Since there is no user-friendly way to specify the numbers and values of the pressure levels, the module of the UPP which specifies the values and number of pressure levels which are output was modified to replace the default values with a new set of values to better preserve the vertical resolution of the model for a range of locations to account for differing terrain heights within the domain. The new set of pressure levels increased the number of levels from 46 to 70, but the implementation of the new set of levels required a recompilation of the UPP software. Subsequently, the UPP was modified to enable ingest of a user-generated file specifying the desired pressure levels which can be implemented without recompiling.

Figure 4 shows the results of modifications made to the UPP to increase the number of pressure levels from 25 to 46 and 70.

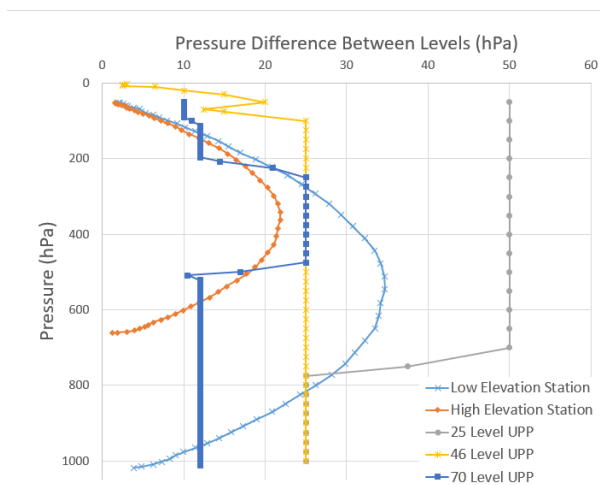


Figure 4. Vertical profiles of the spacing between vertical layers for raw WRE-N output at a high and a low elevation location and the UPP output at 25, 46, and 70 pressure levels.

Notice the improvement in the matching of the vertical density of UPP output with 70 levels with the raw model output especially at the low elevation station. The layers where 70-level UPP output has coarser vertical resolution than the raw WRE-N output at the low elevation stations are the layer below  $\approx 960$  hPa, a thin layer at  $\approx 250$  hPa and the layer above  $\approx 150$  hPa.

To achieve the goal of generating radiosonde observations for verification which better retain the original vertical structure of the raw observations, a modified version of Ungrib provided by NCAR was used which allows the user to specify pressure levels to which the CGM should be interpolated (the capability is now standard in WPS V3.9 and later). This capability was used to increase the number of vertical level being used for interpolation from 32 to 87 levels.

## 5. EXPERIMENTATION TO QUANTIFY IMPROVEMENTS

ARL conducted an experiment to quantify any improvements from the use of the modified UPP process flow and observation QC process flow. The experiment entailed using the modified UPP configuration to generate output to evaluate the quality of representation of model vertical structure and arriving at a better estimate of the model value to be used for verification. The experiment also involved using the modified Ungrib process to provide a GFS background field with additional pressure levels; this GFS field is used in Obsgrid quality control processing of observations to evaluate any improvement in retention of vertical structure which would result in better quality observations for use in verification. Note that the modified Ungrib process does not add vertical structure to the GFS data but merely interpolates it to additional pressure levels so that more of the vertical structure of observations can be retained.

For the experiment, we looked for a weather situation which would have strong, near-surface gradients where vertical structure is a significant factor. We selected a case study day 20/21 FEB 2018 which had a strong cold front with freezing rain in Missouri. We ran the default configuration of Ungrib which uses the GFS on 32 pressure levels and the modified configuration of Ungrib on 87 pressure levels and compared the results with the raw observations. We ran the WRE-N from 12 UTC 20 February 2018 until 12 UTC 21 February 2018 which included an initial preforecast period of 3 hours for observation nudging for a domain which included soundings with strong vertical temperature gradients. We used UPP to generate output on 25, 46, and 70 pressure levels and then ran MET Point-Stat to generate the matched pairs and error statistics.

## 6. ANALYSIS OF RESULTS

We used the experimental data to do the following comparisons:

- 1) Error statistics for UPP output on 25 and 70 pressure levels.
- 2) Forecast temperature difference distributions for UPP output on 25, 46, and 70 levels.
- 3) Vertical profiles of model output derived from use of the three sets of UPP output compared to raw model output.
- 4) Differences in the vertical temperature profile between the interpolated radiosonde observation using the default and modified Ungrib and the original raw observation data.

The error statistics comparison is given in Table 1 below.

FCST_LVL	OBS_LVL	Pairs	Levels ->		25		70	
			ME	MAE	RMSE			
Z2	Z2	812	1.35	1.35	1.65	1.65	2.58	2.58
P1200-0	P900-1200	840	1.13	1.14	1.39	1.40	2.56	2.55
	P750-899	31	-0.32	-0.28	1.18	1.14	1.37	1.37
	P500-749	51	-0.13	-0.14	0.88	0.70	0.90	0.92
	P300-499	52	-0.21	-0.18	0.40	0.38	0.56	0.56
	P050-299	244	0.81	0.16	1.91	1.27	2.42	1.76
P000-049	6	-12.07	-12.07	12.07	12.07	12.71	12.71	

Table 1. Temperature error statistics generated using UPP output at 25 and 70 pressure levels for the forecast period from 15 UTC 20 February 2018 to 12 UTC 21 February 2018. The FCST\_LVL column indicates whether the statistic was computed using 2-m AGL model-diagnosed temperature (Z2) or the full prognostic temperature field (P1200-0 indicating any model prognostic levels with pressures between 1200 and 0 hPa). The OBS\_LVL column indicates whether the statistic was computed using only surface temperature observations (Z2) or using any temperature observations with pressures within the specified ranges (hPa). The pairs column indicates the number of model-observation pairs used to calculate the statistic, while ME=mean error, MAE=mean absolute error, and RMSE=root mean square error.

The error statistics were computed using observations which had been processed and QC'd using output from Ungrib with 87 levels. No statistically significant differences were noted on the basis of applying the 95% confidence limits to the error statistics with the noted exception at 50-299 hPa where the errors for 70 level UPP were smaller than those of the 25 level UPP. The difficulty with this type of analysis is that the errors which are based on model-observation differences tend to average out. For this reason we also present an analysis of the distribution of the forecast temperature differences which focuses on the differences themselves without averaging.

A histogram summarizing the forecast temperature differences is shown in Figure 5.

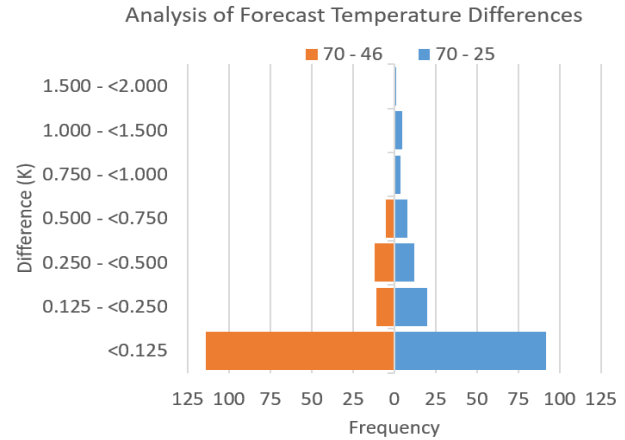


Figure 5. Histogram of the absolute value of forecast temperature differences between 70- and 25-level UPP output (70 – 25) and between 70- and 46-level UPP output (70 – 46) for 12 UTC 21 February 2018.

The differences in the model temperature values were calculated for all matched pairs for the two UPP options using the 70 level option as the reference since it's assumed to be the closest to the raw model output. This is equivalent to the difference in the model errors using the two UPP options. The difference calculation is described as follows:

For each matched pair, the absolute value is taken of the difference between the model values for two UPP configurations. Thus, for each matched pair we calculated:

$$| M_A - M_B |$$

Where:

$M_A$  = WRF model output interpolated to a set of pressure levels per UPP Option "A" and then interpolated to the pressure of the observation in MET.

$M_B$  = WRF model output interpolated to a set of pressure levels per UPP Option "B" and then interpolated to the pressure of the observation in MET.

The differences 70 minus 25 are generally larger than the differences 70 minus 46. So changing from 25 to 46 levels is an improvement as indicated by the removal of the largest differences (bins  $\geq 0.750$  K) and decreases in the  $0.125 - <0.250$  K and  $0.500 - <0.750$  K bins when going from the 70 – 25 bars (right side of the chart) to the 70 – 46 bars (left side of the chart), and that changing from 46 to 70 levels gives additional improvement as indicated by the non-zero values in the  $0.125 - <0.250$  K,  $0.250 - <0.500$  K, and  $0.500 - <0.750$  K bins for the 70 – 46 bars.

Figure 6 shows vertical temperature profiles generated using the modified UPP and the raw model output.

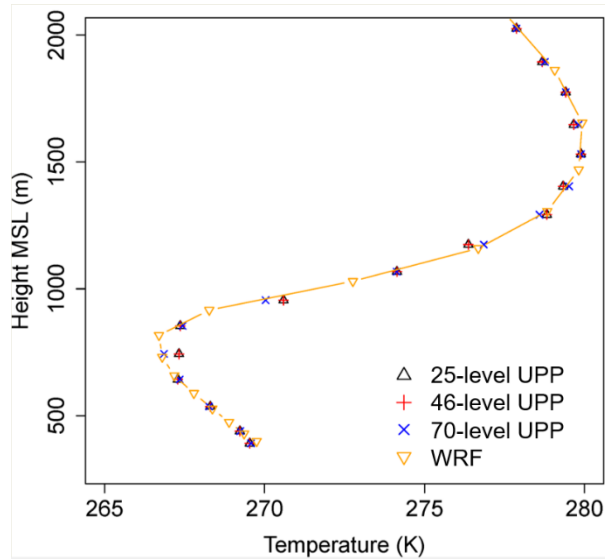


Figure 6. Vertical temperature profiles for Springfield, MO, generated using the UPP and from the raw model output at 12 UTC 21 February 2018.

The profiles in Figure 6 compare the temperature profile from the raw model output with those from the 3 types of UPP output. The 25 level output data points in black triangles are covered by the red, “+” symbol data points of the 46 level output because these two configurations use the same pressure levels in this vertical range. Both the 25- and 46-level output differ from the raw model output at some levels as seen by the horizontal separation on the plot between the 25- and 46-level points and the line representing the interpolation between the vertical levels of the raw WRF output. The 70 levels data points are blue, “x” symbols. The 70 level output tracks the model output quite closely.

Figure 7 also shows vertical temperature profiles generated using the modified UPP and the raw model output, but at a higher altitude where there is a change in the temperature lapse rate.

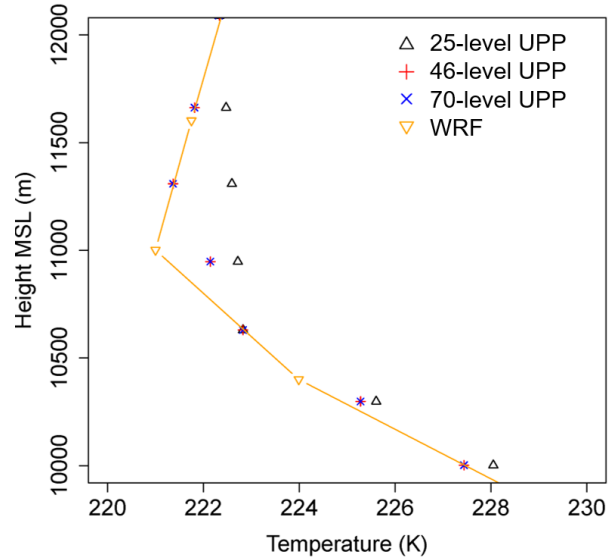


Figure 7. Vertical temperature profiles for Springfield, MO, generated using the modified UPP and from the raw model output at a higher altitude than Figure 6 at 12 UTC 21 February 2018.

The red “+” symbols in Figure 7 are the data points for the 46 level output and are covered by those of the 70 level output which are blue “x” symbols. The 25 level output in black triangles shows the impact of using fewer levels for interpolation of the model value used in verification. The 46- and 70-level output provide a better estimate of the actual model output.

Figure 8 shows the results when comparing the default and modified Ungrib during the QC process.

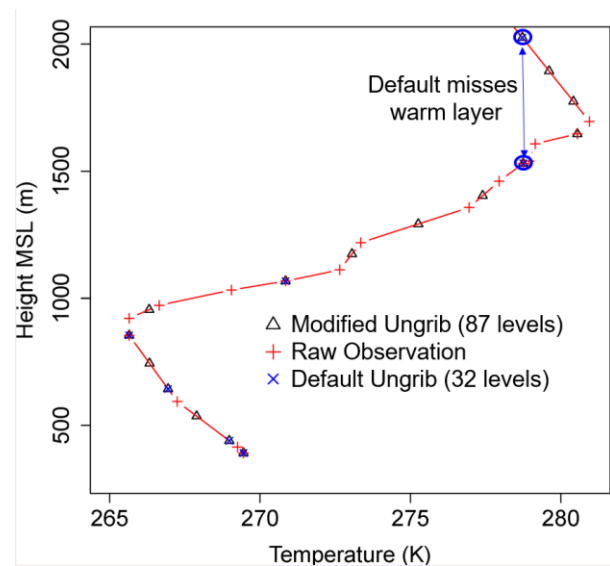


Figure 8. Vertical temperature profiles for Springfield, MO, generated using the default and modified Ungrib and the raw observation data at 12 UTC 21 February 2018.

Figure 8 shows the profile of temperature from the raw observation shown by the solid red line with "+" symbols compared to the profiles obtained through interpolation of the observation to the 32 levels of the GFS background field and to the 87 levels in the GFS background field obtained by interpolating the 32-level GFS field to additional levels via Ungrib. The default data points indicated by blue "x" symbols are covered by the modified points shown by black triangles where they coincide. Note the warm layer around 1600 m where you can see the impact of interpolating to fewer levels. The large blue circles mark the two default data points available at this level. The 87 level profile more closely captured the structure in the raw observation. If the model had forecast isothermal conditions between 1500 m and 2000 m, it would not have been penalized for missing the warm layer if the default method for quality controlling the observations had been used for verification.

## 7. SUMMARY AND NEXT STEP

Increasing the vertical resolution of the pressure levels in the UPP increases the accuracy with which MET measures model skill. Increasing the retention of the observation vertical structure through the quality control process allows the model to be more fully evaluated. The best solution would be to use model data on native vertical grids for the interpolation to the observations for verification thus avoiding double interpolation. The next step is to develop an approach for using model output on native vertical grids for verification.

## 8. REFERENCES

1. Deng A., D. Stauffer, B. Gaudet, J. Dudhia, C. Bruyere, W. Wu, F. Vandenberghe, Y. Liu, and A. Bourgeois, 2009: Update on WRF-ARW end-to-end multi-scale FDDA system. 10th WRF Users' Workshop, NCAR, 23-26 June, Boulder, CO. 1.9. 14 pp.
2. National Center for Atmospheric Research, 2016: Model Evaluation Tools Version 5.2 (METv5.2). User's Guide 5.2. [Available online at <http://www.dtcenter.org/met/users/>.]
3. Skamarock W. C., J. B. Klemp, 2008: A time-split non-hydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, 227, 3465-3485.
4. UPP User's Guide V3.0, 34 pp. [available online at [http://www.dtcenter.org/upp/users/docs/user\\_guide/V3/upp\\_users\\_guide.pdf](http://www.dtcenter.org/upp/users/docs/user_guide/V3/upp_users_guide.pdf)]

## 9. ACKNOWLEDGEMENTS

Model Evaluation Tools (MET) was developed by the National Center for Atmospheric Research (NCAR). MET was developed through grants from the United States Air Force Weather Agency (AFWA) and the National Oceanic and Atmospheric Administration (NOAA). NCAR is sponsored by the United States National Science Foundation.