10A.1

The Model Evaluation Tool

Lacey Holland, John Halley Gotway, Barbara Brown, Randy Bullock, Eric Gilleland, and David Ahijevych
National Center for Atmospheric Research
Boulder, CO 80307

1. INTRODUCTION

To ensure changes made to numerical weather prediction (NWP) models improve forecast quality, special efforts should be made in model evaluation. The goal of any model evaluation should be to assess the quality of forecasts and analyses produced, through the use of user-relevant metrics. The Model Evaluation Tool (MET) was developed to support the Developmental Testbed Center (DTC) at the National Center for Atmospheric Research (NCAR) in its role as liaison between the research and operational NWP communities by providing a tool to analyze forecast performance.

One of the DTC's charges is to maintain WRF (Weather and Research Forecast model) Reference Code. Unlike WRF Contributed Code, which meets minimum programming standards and has been tested by its contributors, WRF Reference Code consists of NWP code that has been rigorously tested by the DTC and meets determined quality standards. When Contributed Code is considered for elevation to Reference Code status, the DTC must assess the proposed changes by analyzing effects on forecast performance. The MET will aid in the analysis of WRF Reference code changes through application of multiple methods of evaluating NWP forecasts. This toolkit will also be available to the WRF user community and forecast developers.

The initial version of the MET focuses on datasets used by the DTC and makes use of utilities available in the WRF model. Later versions may extend capabilities to additional datasets. The MET offers grid-to-point, grid-to-grid, and advanced spatial forecast verification techniques in one unified, modular toolkit that builds on capabilities available in other verification systems. This paper summarizes the methods included in the MET, shows an example highlighting MET capabilities, and discusses capabilities to be implemented in future releases.

2. DESIGN DESCRIPTION

In addition to aiding the DTC in evaluating NWP code, the MET will also provide a freely-available verification toolkit to the WRF user community through controlled version releases. In anticipation of a variety of verification applications, the MET is highly configurable and modular in design. This design provides users with a large amount of flexibility in configuring and running the MET. Users may pick and choose which tools provided in the MET to apply to their datasets, allowing them to avoid running unnecessary programs and providing the user with greater control over their verification. One advantage of this strategy is that there are fewer interdependencies between individual programs, limiting the impact of version release changes.

Different applications used in the MET are depicted in the flowchart in Fig. 1. Tools provided in the MET can be grouped by function to describe the overall structure of the MET: data handling, statistical calculations, and data analysis. Data handling tools are provided to read in, to match precipitation rates or collection periods ("PCP Combine" in Fig.1), and/or to reformat input datasets ("PB2NC"). Another set calculates verification statistics for grid-to-point, grid-to-grid, and/or apply the spatial verification method developed at NCAR referred to as Method for Object-based Diagnostic Evaluation (MODE). Verification methods using point-based observations are hereafter referred to as grid-to-point methods ("point_stat"), while verification using gridded datasets for verification is referred to as grid-to-grid methods ("grid_stat"). The last set of tools, still under development, will aggregate and summarize statistics and provide methods for producing summary plots.

In the initial release, the input data types supported by the MET are somewhat limited. One goal of the MET was to include functionality

* *Corresponding author's address*:
Lacey Holland P.O. Box 3000 Boulder, CO 80301
e-mail: lholland@ucar.edu

provided by other forecast verification systems. For grid-to-grid or MODE verification, the forecast and gridded verification datasets must be in GRIB format. The use of the *copygb* utility provided in the WRF Post is recommended for use when the forecast and observational datasets are not on the same grid. MET assumes this has already been performed by the user and both the forecast and the gridded verification dataset are on the same grid. Currently, for grid-to-point verification, the MET supports the prepbufr observational data format.

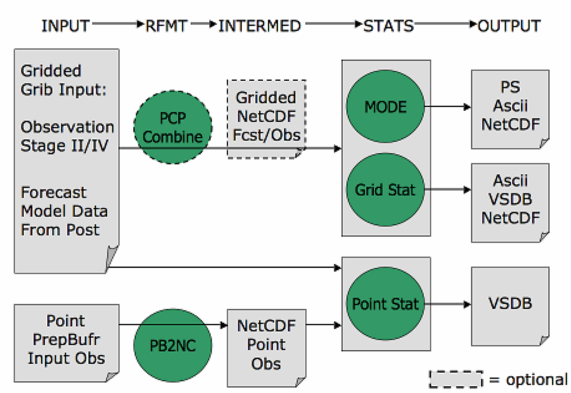

Figure 1. Flowchart depicting MET applications, input, and output.

The MET produces intermediary files in NetCDF format. Thus, if the user does not have observational datasets in prepbufr format for grid-to-point verification, he or she may reformat the desired data into NetCDF.

A variety of options are provided for to allow the user to specify the type of output. Most statistics are output in ASCII format. The grid-to-point and grid-to-grid verification tools can output statistics in the VSDB (Verification System DataBase) format used by the NCEP Forecast Verification System (Brill 2007), as well as an enhanced VSDB format that includes additional types of measures. Contingency table counts and statistics for continuous variables may also be output in ASCII format. Output from the MODE tool is also available in ASCII. Graphics produced by MODE are in Postscript format.

3. VERIFICATION METHODS

The MET is designed in a modular fashion such that methods for the verification of forecasts using both point observations and gridded datasets are available. Spatial verification information is provided by the MODE tool. The MODE tool is discussed further in section 3.3.

3.1 Grid-to-point methods

Grid-to-point methods are used when point-based observational datasets are available as the verification dataset. Grid-to-point methods are applied using datasets that are ungridded. The user may select the following options: data types are used in grid-to-point verification, the GRIB forecast to be verified, the interpolation methods used, a mask over which to perform the verification, the type of output desired, thresholds used to discretize continuous variables, the statistics desired, and a host of other options through the use of a commented configuration file. A summary of parameters available to be specified by the user is defined in Table 1.

Table 1. User-specified parameters in the MET point_stat verification tool.

Parameter	Description
vx_grib_code	Grib code and level corresponding to field to be verified
mask_grids	Grid indicating mask to apply, if any
mask_polys	Latitude/longitude polyline identifying verification area, if any
thresholds	Thresholds used to discretize forecast/observation fields to produce contingency table scores, if any
ci_alpha	Alpha value used in producing confidence intervals
output_flag	Flag indicating the type of output produced
message_type	Prepbufr message type used
interp_flag	Flag indicating method of interpolation used
interp_width	Indicates how many surrounding grid point to use in the interpolation
interp_threshold	Number of surrounding points that must be available to perform the interpoloation

The input prepbufr file is reformatted to NetCDF format in *pb2nc*, and user-specified parameters regarding observational data types and minimum requirements for quality marks are applied. Any derived quantities to be verified (e.g. CAPE, RH, potential temperature) are stored in NetCDF. Verification statistics are then computed and provided in ASCII format. Verification statistics for both continuous and discrete variables are provided. Statistics for continuous variables available are as follows:

- forecast/observation mean
- forecast/observation standard deviation
- Pearson correlation coefficient
- Spearman rank correlation coefficient
- Kendall tau rank correlation coefficient
- mean error
- standard deviation of the error
- frequency bias
- mean absolute error
- mean squared error
- bias-corrected mean squared error
- root-mean squared error
- percentiles of the error

Confidence intervals on the mean, standard deviation, Pearson correlation coefficient, Spearman rank correlation coefficient, Kendall tau rank correlation coefficient, mean error, and standard deviation of the error are also available.

Statistics can be computed for discrete variables, and for continuous variables that have been discretized. A list of statistics for discrete variables is given below:

- Total observations
- FHO statistics (forecast rate, hit rate, observation rate)
- Contingency table counts
- Contingency table proportions
- Accuracy
- Bias
- Probability of Detecting Yes
- Probability of Detecting No
- False Alarm Ratio
- Critical Success Index
- Gilbert Skill Score
- Hanssen and Kuipers Discriminant
- Heidke Skill Score
- Odds Ratio

As discussed earlier, the output of this file is in ASCII format. Specifically, the output may be in VSDB format, or in a similar ASCII-based format that describes statistics for continuous variables, continuous variables with a threshold applied, and contingency table counts.

3.2 Grid-to-grid methods

Many of the same methods for performing grid-to-point verification can be applied to gridded forecast and verification datasets. These methods may be employed using a variety of datasets such as gridded analyses. Because the MET assumes any interpolation needed has already been performed (i.e. the forecast and verification datasets are on the same grid), no interpolation routine is provided. As in the MET's grid-to-point forecast verification, statistics for both continuous and discrete variables are available. Other user-specified parameters available in the configuration file are shown in Table 2. As for grid-to-point methods, the output is available in VSDB format, and an ASCII format that provides information similar to the grid-to-point methods.

Table 2. User-specified parameters available in the grid_stat tool.

Parameter	Description
vx_grib_code	Grib code and level corresponding to field to be verified
mask_grids	Grid indicating mask to apply, if any
mask_polys	Latitude/longitude polyline identifying verification area, if any
thresholds	Thresholds used to discretize forecast/observation fields to produce contingency table scores, if any
ci_alpha	alpha value used in producing confidence intervals
output_flag	flag indicating the type of output produced

3.3 The MODE Tool

The MODE Tool is provided as an initial object-based method of forecast verification for spatial forecasts. There are several other objected orientated methods available in the user community, which may be included in future releases. Although initially developed for the special problem of verifying radar reflectivity and precipitation forecasts, MODE has been generalized for use on a wider variety of datasets.

One capability provided by the MET with respect to the verification of precipitation forecasts is the ability to sum accumulated precipitation over user-specified periods of time. This capability may be used on either the gridded verification dataset or the forecast dataset.

The most common application of the MODE is with precipitation forecasts. The MODE has been most thoroughly tested as a tool to verify WRF accumulated precipitation forecasts using Stage II or Stage IV precipitation analyses as verifying observations.

For the user interested in only precipitation verification, he/she may skip the grid-to-grid module provided in the MET and use only the MODE, as the MODE can output the results of both an object-based method of precipitation forecast verification and many of the traditional verification statistics produced in the grid-to-grid module. A variety of parameters controlling the output the results of and flow of the MODE tool are available. Table 3 shows some of these parameters. Additionally, a large number of parameters pertaining to the way objects are defined and how matching and merging are performed are also available for user specification, but for the sake of brevity these are not shown.

Table 3. User-specified parameters available in the MODE tool.

Parameter	Description
grid_res	nominal grid spacing
vx_grib_code	Grib code and level or accumulation period corresponding to field to be verified
mask_missing_flag	indicates whether/how to mask out missing data
mask_grid	grid indicating mask to apply, if any
mask_grid_flag	flag indicating how to apply grid mask
mask_poly	Latitude/longitude polyline identifying verification area, if any
mask_poly_flag	flag indication how to apply the lat/lon polyline mask

The MODE outputs a summary Postscript file that includes both graphics and summaries of the object interest fields (Fig. 2). This tool can also be used to produce an ASCII file of traditional verification statistics. The MODE also outputs an ASCII file showing many attributes of the objects identified, as well as comparisons between observed and forecast objects.

4. ANALYSIS TOOLS

The MET Analysis tool, although still under development, will provide ways of aggregating and producing graphics of the statistics produced by the grid-to-point, grid-to-grid, and MODE Tools. These tools will allow the production of time series, spatial maps and height series of statistics, and aggregations of statistics over user-defined regions. It will also allow the calculation of indices such as the GO Index and others that are of special interest to specific user groups. Graphics tools will be provided but may not be fully integrated.

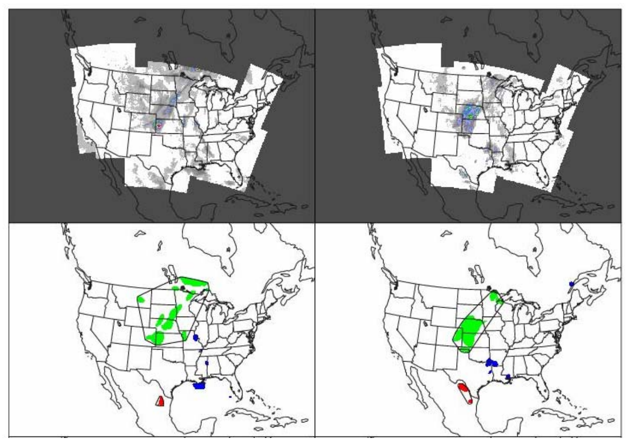

Figure 2. Example of MODE-identified objects in the forecast (bottom left) and observation (bottom right) fields. Corresponding raw fields are shown in the top panels.

5. SUMMARY

The MET is a highly configurable toolkit freely available to the model developer, verification, and WRF-user communities. The initial release of MET will provide verification methods using both grid-to-point and grid-to-grid methods. The MODE tool also will provide an object-based method for spatial forecast verification. Initial support for graphics will also be provided. Additional functionality will be made available through controlled version releases.

6. FUTURE WORK

Controlled version releases in the future will provide additional capabilities. These capabilities will include support for additional observational formats, a database, and additional advanced spatial verification techniques.

User input from a townhall meeting at the 87th American Meteorological Society Annual meeting in San Antonio, several ideas for both initial capabilities and capabilities to be included in future releases were identified. This was followed up by a WRF verification workshop in which members of the NWP and verification communities met to identify and discuss which features and capabilities would be important to include in future releases. Some of the needs identified for future implementation include extensions to support the verification of probabilitistic and ensemble forecasts, and the inclusion of additional spatial verification methods. The results of an intercomparison of spatial verification techniques (Gilleland 2007) will identify additional spatial techniques considered for inclusion in the MET.

7. ACKNOWLEDGMENTS

We thank the U.S. Air Force Weather Agency and the National Oceanic and Atmospheric Administration for their support. Our gratitude also goes to the members of the WRF community who participated in the AMS town hall meeting and the verification scientists who participated in the WRF Verification Workshop in February 2007.

8. REFERENCES

Brill, K. cited 2007: EMC verification database. [Available online at: http://www.emc.ncep.noaa.gov/mmb/papers/brill/VSDBformat.txt]

Gilleland, E. cited 2007: ICP. [Available online at: http://www.ral.ucar.edu/projects/icp/]