The Developmental Testbed Center
2007 13-km Dynamic Core Test Design

Design document written by Lígia Bernardet with input from DTC team and WRF developers

September 13, 2007

1. Introduction

One of the milestones of the Developmental Testbed Center (DTC) for the Fiscal Year 2007 is a controlled comparison between verification statistics of the forecasts generated by the two dynamic cores of the Weather Research and Forecasting (WRF) model: the Non-hydrostatic Mesoscale Model (NMM) core and the Advanced Research WRF (ARW) core. The motivation for this comparison is to ascertain whether the medium range skill of the two cores is similar enough to allow the direct transfer of research results obtained with one core to the other. Even though this test will be conducted with a particular model configuration, the results obtained will be understood to be general enough to be applicable to a broader range of configurations. Considering that the DTC is also tasked with conducting tests to create the WRF Reference Code, the results of this test will be used, together with other requirements, to elevate sections of the code from Contributed to Reference status.

This task is a follow-up to the DTC 2006 Core Test (Nance 2006) in support of the implementation of the WRF Rapid Refresh (RR) at the National Centers for Environmental Prediction (NCEP). The 2007 13-km Core Test extends the 2006 task onto a longer forecast period, 60 hours instead of 24, in order to evaluate the dependency of the forecast differences on the forecast length.

Since the 2007 13-km Core Test is an extension of the 2006 Core Test, most parameters in the end-to-end forecast system, including forecast domain and grid spacing, will not be altered from the 2006 Core Test. However, the WRF Preprocessing System (WPS) will be used instead of the WRF Standard Initialization (SI), the version of all codes will be updated, and the source of initial conditions will be altered. The need to update the model and preprocessor source code stems from the importance of using the latest developments in the physics and dynamics of the model, and from the need to use code that is in the WRF repository to maximize support from the developers.

The ARW and NMM dynamic cores will be used to forecast 60 cycles divided into the four seasons. The models will be initialized every other day, at 00 and 12 UTC. Model output will be ingested by the WRF Postprocessor (WPP) to obtain de-staggered grids interpolated in the vertical to isobaric levels and in the horizontal to a common grid. Postprocessed forecasts will be verified using the NCEP Verification System and will be used to generate images. While it would be desirable to use the Model Evaluation Tool (MET) in this test, it is not known whether the DTC will have it completed and ready for implementation. Therefore, postprocessed files from this test will be archived in order to be run through the MET at a later date.

The timeline for this test is as follows:

- 06/20: cutoff date for acquisition of new software developments
- 07/04: deadline for final software acquisition and system readiness
- 07/16 – 08/31: period to conduct model runs
- 09/30: preliminary results with confidence intervals will be released
- 12/31: deadline for final report to be released

The test described in this document will be performed at the same grid spacing used for the 2006 Core Test (13 km). An additional high-resolution test will be conducted by the DTC and its design will be addressed in a separate document.

2. Goals

The main goals of the 2007 DTC Core Test are 1) to quantify the errors in the forecasts produced by the ARW and NMM dynamic cores of the WRF model in a given configuration, and 2) to quantify the differences between forecasts produced by the two dynamic cores in a given configuration. Both analyses will be based on objective forecast verification statistics of the model.

To assess the differences in forecast verification statistics from each dynamic core, the ARW and NMM will be configured as similarly as possible in all aspects besides dynamic core. They will be run with the same source of initial and boundary conditions and with identical physics suites. Furthermore, they will undergo similar post-processing and verification treatment.

The objective verification results will be computed with the NCEP Verification System. The measures will include bias, root mean square error (RMSE), and decomposed (debiased) RMSE for surface and upper-air temperature, relative humidity and winds. For precipitation, the measures will include frequency bias and equitable threat score. The objective verification results will be stratified by season, vertical level, and regional area and, for the surface, also by forecast initialization hour (00 and 12 UTC). In addition, averages will be computed for the entire test period, for the entire verification domain (CONUS) and for all initialization hours.

An ancillary goal of the 2007 13-km Core Test is to quantify the differences in forecast verification statistics from models run on different platforms. This goal will be addressed by performing redundant forecasts, that is, running a subset of the forecasts on two platforms.

A final goal is to make all results accessible to the community. Therefore, products will be archived and results summarized in a report. It is expected that results from this test be used as a baseline for further testing.

3. Experiment Design

This section describes the end-to-end forecast system to be employed in the 2007 13-km Core Test. It is composed of the WPS, WRF model, WPP, NCEP Verification System, graphics generation, data archival and dissemination. Before this document is completed, input from the ARW and NMM developers will be sought in order to finalize the model configurations.

1. Codes to be employed

The baseline codes for WRF and WPS will be the tarfiles used for the DTC Winter 2007 NMM Tutorial. New developments may be incorporated if they are in the code repository by the cut off date for acquisition of new developments (June 20, 2007). Relevant bug fixes may be incorporated if they are in the repository by the final software acquisition deadline (July 04, 2007). The requirement of only using code originating from the WRF repository guarantees that the code tested by the DTC is an integral part of the WRF system and eligible to become Reference code, and guarantees that the results are meaningful for future WRF users.

The WPP code will be based on the one used for the DTC Winter 2007 WRF-NMM Tutorial. An updated version of the NCEP Verification System code will be obtained from NCEP. The WPP and the NCEP Verification System may have new developments added by the cut off date for acquisition of new developments (June 20, 2007). Relevant bug fixes may be incorporated by the final software acquisition deadline (July 04, 2007). In case platform-porting problems prevent the use of the updated NCEP Verification System, the version used for the 2006 Core Test will serve as a fall back.

Forecast images will be generated with NCAR Command Language (NCL). The scripts to generate the images will be based on the ones used in the 2006 Core Test, with revisions to add or improve images and to make the scripts consistent with the latest WPP code.

The execution of the end-to-end system will be managed by the Workflow manager developed by the National Oceanic and Atmospheric Administration (NOAA) Earth Systems Research Laboratory. This workflow controls the system execution, starting tasks as soon as their dependencies have been met, monitoring task execution, and re-starting tasks when needed.

2. Domains Configuration

Since the ARW and NMM use different grid projections, it is not possible to exactly match the forecast grids of the two models. The grids used in the 2006 Core Test will be employed in this test, since extensive efforts were used to minimize the differences in area coverage and number of grid points between the dynamic cores. The ARW and NMM domains along with the common grid used for the Post-processing are shown below. Both dynamic cores use 58 vertical levels (59 sigma entries).

The specifications of each domain are:

- NMM: 280 x 435 gridpoints; total 121,800 gridpoints.
- ARW: 400 x 304 gridpoints; total 121,600 gridpoints.
- Post-processing (RUC 13 Grid): 451 x 337 gridpoints; total 151,987 gridpoints.

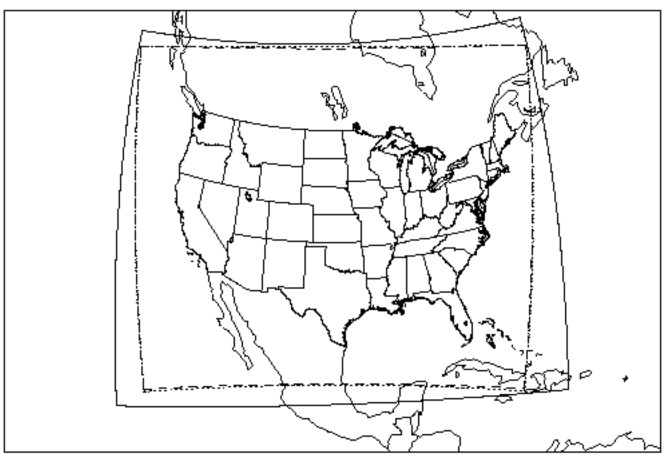

Figure 1. Map showing the boundaries of the computational domains used for the ARW (dashed line) and the NMM (dotted). The solid line shows the boundaries of the domain for the RUC 13, which was used to initialize the ARW and NMM forecasts.

Several additional domains will be used as sub-regions for verification. The results of the NCEP Verification System will be area-averaged over the entire CONUS (G164), and over the Eastern and Western sectors of the CONUS (Grids165 and 166) (Figure 2). Additionally, the surface and upper air component of the NCEP Verification System will be configured to use the 14 regional domains shown in Figure 3 to compute regional averages of statistics. If the DTC finds expertise to do this task, the NCEP Precipitation Verification System will also be configured to use the 14 regional domains.

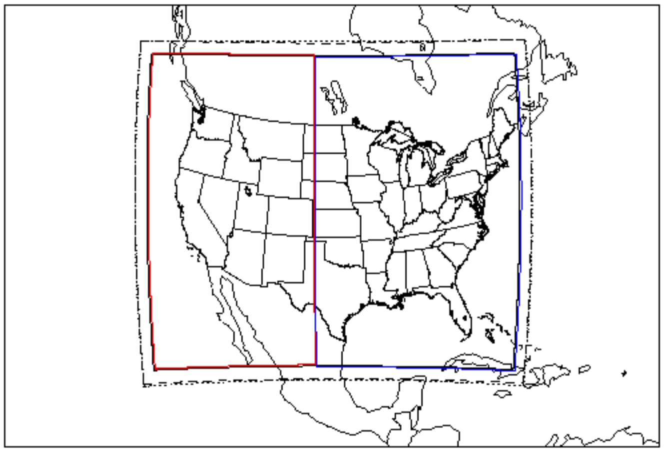

Figure 2. Map showing the boundaries of the verification domains: CONUS (solid black), West (solid red), and East (solid blue). The computational domains used for the ARW (dashed line) and the NMM (dotted) are shown for reference.

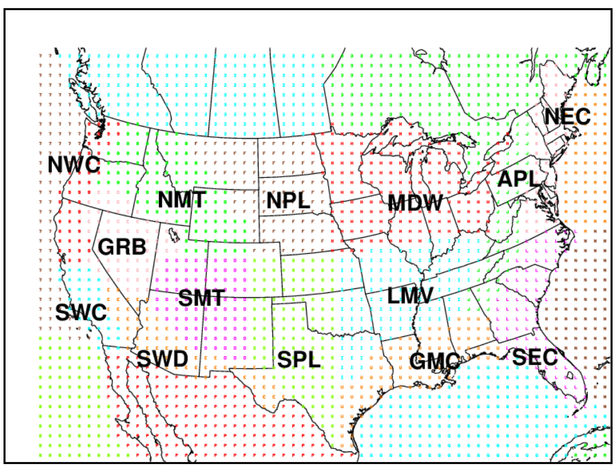

Figure 3. Map showing the location of the regional verification domains.

3. Initial and Boundary Conditions

The Core Test will employ initial and boundary conditions from the North American Mesoscale Model (NAM). For the retrospective periods to be used, the NAM corresponded to the Eta model.

The time-invariant component of the lower boundary conditions (topography, soil and vegetation type etc.) will be generated through the geogrid program of WPS. The use of a unified pre-processor for the two dynamic cores should minimize the differences in lower boundary conditions. Currently, the topography smoothing differs between the cores. The DTC is working with WPS developers to seek unification.

4. Forecast Periods

Forecasts will be computed for 15 days in each season. Forecasts will be initialized at 00 and 12 UTC every other day and run out to 60 hours.

5. Physics Suites

The physics suite is described in the Table below.

Microphysics	Ferrier
Surface Layer	Janjic
Planetary Boundary Layer	Mellor-Yamada-Janjic
Convection	Betts-Miller-Janjic
Land-Surface Model	Noah
Radiation SW and LW	GFDL

6. Other aspects of model configuration

The timesteps used in the two dynamic cores are different due to numerical stability requirements of each dynamic core. The NMM will use a timestep of 30 s, and the ARW a long timestep of 72 s with an acoustic timestep of 18 s. To make the frequency of physics calls in each dynamic core more uniform, calls to the boundary layer, microphysics and cumulus parameterization in the ARW will be done every time step, while in the NMM they will be called every other timestep. Calls to radiation will be done every 30 minutes for both dynamic cores. History files will be written every three hours.

7. Post-processing

The WPP will be used to destagger the ARW and NMM forecasts and to interpolate them to a common Lambert-Conformal grid (Figure 1). Additionally, the WPP will be used to generate derived meteorological variables including mean sea level pressure, and to interpolate the forecasts to isobaric surfaces, as required by the plotting and verification programs. Postprocessed fields in model native vertical coordinates will not be generated.

8. Model verification

Model verification statistics will be generated with the NCEP Verification System, comprised of the Surface and Upper Air Verification System and the Quantitative Precipitation Forecast (QPF) Verification System.

Through the NCEP Surface and Upper Air Verification System, forecasts will be interpolated to the location of the observations (METARs, RAOBS, and ACARS). The observational dataset will be the Rapid Update Cycle Prepbufr files. The verification statistics produced will include the bias, the RMSE and the decomposed (debiased) RMSE, computed separately for each observational type.

For the precipitation verification, a grid-to-grid comparison will be made in which the forecasts and the precipitation analyses will first be interpolated to the 12-km grid spacing Grid 218 and then compared. Accumulation periods will be 24 h and 3 h. The observational datasets will be the NCEP Stage II analysis for the 3-h accumulations and the River Forecast Center Analysis for the 24-h accumulation (valid at 12 UTC). The statistics computed will include the bias and the equitable threat score.

Verification statistics will be stratified by season, vertical level, regional area and, for the surface, also by forecast initialization hour (00 and 12 UTC). In addition, averages will be computed for the entire test period (all seasons), for the entire verification domain (CONUS) and for all initialization hours. The areas for aggregation of verification statistics are: CONUS, Eastern-CONUS, and Western-CONUS (Figure 2). Additionally, the surface and upper air component of the NCEP Verification System will be configured to use the 14 regional domains shown in Figure 3 to compute regional averages of statistics. If the DTC finds expertise to do this task, the NCEP Precipitation Verification System will also be configured to use the 14 regional domains.

Since every cycle will be run for both dynamic cores, the presentation of the results will take advantage of the pair wise nature of the Core Test. With this methodology, also used in the 2006 Core Test, in addition to verification statistics for each individual configuration, differences between the verification statistics of two configurations will be computed. For example, a difference could be (CONUS 500 hPa temperature bias from ARW run at NCAR) -(CONUS 500 hPa temperature bias from NMM run at NCAR). Pairs will include:

1. ARW (run at NCAR) x NMM (run at NCAR)
2. ARW (run at NOAA) x NMM (run at NOAA)

In order to evaluate the differences between runs of a core processed at two different platforms (more details in Section 3.10), two additional pairs will be constructed.

3. ARW (run at NCAR) x ARW (run at NOAA)
4. NMM (run at NCAR) x NMM (run at NOAA)

All verification statistics will be accompanied by confidence intervals (CIs) computed using parametric tests for the surface and upper-air statistics and a bootstrapping method for the precipitation statistics. The CIs on the differences between statistics for two configurations will assist in determining whether the members of the pair are significantly different. Since uncertainty measures are not computed by the NCEP Verification System, they will be generated separately. The confidence intervals will be in a spreadsheet format and will be incorporated by the DTC into the final written report. To gain further insight into the results, practical significance will be used in addition to statistical significance.

9. Graphics generation and display

Graphics will be generated using NCL. The suite of images will comprise: 1) 2-m temperature and 10-m wind vectors; 2) 2-m dewpoint and 10-m wind vectors, 3) 10-m wind speed and vectors, 4) runtotal accumulated precipitation and mean-sea-level pressure (MSLP), 5, 6, 7) 3-h accumulated convective, resolved and total precipitation, with MSLP and 1000-500 hPa thickness, 8) precipitation type, 9) simulated RADAR composite reflectivity, 10) snow cover water equivalent, 11) convective available potential energy, 12) convective inhibition, 13) precipitable water, 14) 850 hPa temperature, geopotential height and wind vectors, 15) 850 hPa geopotential height, wind vectors and wind speed, 16) 850 hPa RH, 17) 850-500 hPa mean RH and 700 hPa wind vectors, 18) 700 hPa vertical velocity and geopotential height, 19) 500 hPa absolute vorticity and geopotential height, 20) 250 hPa wind vectors and geopotential height, and 21) Skew-T soundings.

The scripts to generate the images will be based on the ones used in the 2006 Core Test, with possible revisions to add or improve images and to make the scripts consistent with the latest WPP code. All graphics will be displayed on a website so they can be accessible to the staff conducting the test and to researchers afterwards.

10. Platform intercomparison

While it is expected that identical configurations of a cycle will produce somewhat different forecasts on two different computational platforms, it is not known whether forecast verification measures averaged over several forecast cycles will differ between platforms.

To determine if time-averaged verification statistics are independent of the computer platform on which the models are processed, a subset of the forecasts will be computed both at the National Center for Atmospheric Research (NCAR) and NOAA supercomputers. This subset will initially be comprised of a few cycles from each of the four seasons. Statistical results will be used as a guideline to determine whether additional cycles need to be run in order to obtain statistically significant results regarding the platform differences. The statistics will be considered independent of computer platform if differences between the platforms are not found to be statistically significant.

4. Computer Resources

- Processing resources

  The NCAR and NOAA Boulder supercomputers will be used for all calculations. All forecast cycles will be computed on the NCAR supercomputer, and some cycles will be repeated on the NOAA supercomputer. This redundancy will address the issue of how different the forecasts errors are for identical configurations on different platforms.
- Storage resources

  All storage will be done on the NCAR Mass Storage System. Forecasts generated at NOAA ESRL will be transferred to NCAR for storage.
- Web resources

  Model forecast and verification graphics will be accessed through a web interface.

5. Deliverables

The NCAR Mass Store will be used to store the files produced by the forecast system. The following files will be stored:

- NAM files used for initial and boundary conditions.
- Datasets used for forecast verification (RUC Prepbufr and Stage II and RFC precipitation analyses).
- Static files produced by the geogrid component of WPS.
- Final output of WPS.
- Input and boundary condition files produced by real.
- History files produced by WRF.
- Output of the copygb component of the WPP.
- Prepfits and vsdb files output by the NCEP verification system.

Additionally, all source codes and executables used will be stored.

These files will be made available to the DTC and to the community for further studies.

The DTC will produce a report outlining the results and conclusions from this Core Test.

References

Nance L., 2006. Weather Research and Forecasting Core Test – DTC Report (available from http://ruc.fsl.noaa.gov/coretest2/DTC_report.pdf).