Developmental Testbed Center (DTC)

Testing protocol for transitioning mesoscale modeling code from research to operations (R2O)

Point of Contact: Jamie Wolff

1. Overview

The Developmental Testbed Center (DTC) is a distributed facility with components at the National Center for Atmospheric Research's (NCAR) Research Applications Laboratory (RAL) and the National Oceanic and Atmospheric Administration's (NOAA) Earth Systems Research Laboratory (ESRL). The fundamental purpose of the DTC is to facilitate the interaction and transition of numerical weather prediction (NWP) technology between research and operations. In order to accomplish this mission, focus is placed on two key aspects: 1) advancing science research by providing access to and community support for functionally similar components of operational NWP systems to the research community, and 2) extensively testing and evaluating promising codes emerging from the research community in a common framework to demonstrate the potential benefits for operational systems. The ultimate goal of these specific activities is to accelerate the rate at which new science innovations are infused into operational weather forecasting, promoting an overall improvement in forecast skill.

A wide range of NWP science innovations are under development in the research community. With this extensive number of new techniques, a testing protocol detailing the procedures necessary to advance new innovations through the research to operations (R2O) process is imperative. Ultimately, the effort should be streamlined to make it not only efficient, but also effective, providing a path to potential operational use. This document details a three stage process of testing conducted by the research community, DTC and, ultimately, operational centers. The current focus for this effort is to assess mid-latitude mesoscale modeling research techniques with a focus on model numerics and physical parameterizations.

2. Testing Procedure

Stage One – Proving ground for research community

During development of new science innovations in NWP systems, researchers generally perform case studies in order to further delve into the complex interactions of non-linear systems. Detailed case studies also allow the researcher to tune the new technique for a particular event in an optimal way. The natural progression of testing may then lead to longer test periods on the order of a month or season in order to assess the new technique across a more diverse set of situtations. In order to assist the research community with conducting this initial stage of testing, the DTC has established and is maintaining the Mesoscale Model Evaluation Testbed (MMET; http://www.dtcenter.org/eval/mmet). The motivation of MMET is to assist the research community in efficiently demonstrating the merits of a new development that could positively impact an operational configuration in the future.

MMET provides a variety of initialization and observation data sets for a number of routine, high-impact and field campaign cases[1]. Baseline results for select operational configurations are also produced by the DTC in a functionally similar environment to operations and made available through MMET. The baselines established by the DTC utilize the MMET data sets and, thus, do not exactly emulate those used in operations, which include restricted data not available to the research community. The so-called "data sets of opportunity" available in MMET are housed at the DTC and accessible to the community-at-large via a web interface (http://dtcenter.org/repository). RAMADDA[2], a **R**epository for **A**rchiving, **M**anaging and **A**ccessing **D**iverse **DA**ta, is used as the foundation for hosting and distributing the data. The data sets can be utilized by the research community to test and evaluate newly developed innovations that address forecast needs in areas that may include, but are not limited to: hydrology, severe weather, aviation, energy, ground transportation, air quality and fire weather. In addition to single case events, longer time periods to further investigate a new technique are also included. Through the common framework provided by MMET, researchers have the ability to perform direct comparisons between multiple innovations tested by the research community and/or against the baseline operational configurations established by the DTC. It is recommended to run several case studies spanning multiple weather regimes to illustrate the versatility of the new innovation for operational use.

The DTC advocates the communication of objective verification results produced by the research community in a way that is concise and allows for the value of the new technique to be readily assessed. The research community is encouraged to use the DTC-supported verification software system Model Evaluation Tools (MET) to assist with performing consistent verification among the research community. For ease of use, a template configuration of the verification system is provided for each case included in the MMET. This template provides efficient generation of objective verification scores for both grid-to-point and grid-to-grid comparisons. Verification metrics in the template configuration include standard verification statistics in wide use by the NWP community (e.g., RMSE, bias, GSS). In addition, MET provides the opportunity to assess new innovations with state-of-the art verification techniques (e.g., object-based, neighborhood methods). Depending on the application, a range of variables and properties may be examined. For general, mid-latitude forecasts, the evaluation should include an assessment of surface and upper-air predictions for temperature, humidity and wind, as well as precipitation. If MET is not used to conduct the evaluation, a detailed description of how the verification scores were computed is vital for comparing different verification approaches and resulting statistical values. Providing the DTC with objective verification output is necessary to conduct a direct comparison with other established results. A preliminary estimate of computation resources required for running a new innovation is also critical information for operational assessment.

Nomination Process – Transitioning code from the research community to the DTC

If improved forecast accuracy is demonstrated through objective verification results during the first stage of testing, a recommendation may be made for the innovation to be considered for the second

[1] Recommendations for additional cases to be included in MMET can be submitted from the MMET website: http://www.dtcenter.org/eval/mmet/cases/form_submission.php.

[2] RAMADDA is developed and maintained by Unidata: http://www.unidata.ucar.edu/software/ramadda.

stage of testing. Community users can nominate innovations for more extensive DTC testing by going to the MMET website and filling out the nomination form (http://www.dtcenter.org/eval/mmet/candidates/form_submission.php).

In an effort to ensure that the testing conducted in the second stage of the R2O process is relevant to the broad user community and readily accessible within DTC-supported software systems, developers are encouraged to check the new technique into the WRF code repository as a community contribution prior to any extensive testing and evaluation being conducted by the DTC. Working through this process also ensures the code changes have been vetted through a developers committee and run through an initial level of testing. Researchers may work with a member of the WRF Developers Committee (WDC) to contribute well-integrated techniques in to the WRF repository. To be assigned a WDC liaison, please contact wrfhelp@ucar.edu for more information. Detailed information on the management of the WRF repository can be found in the WRF Code Repository and Release Administration document (http://www.mmm.ucar.edu/wrf/docs/code_admin.pdf). Briefly, the WDC provides an interface to outside developers and facilitates incorporation of new or enhanced features and functionalities into WRF. These additions/modifications are subject to the constraints of appropriateness, correctness, impact on existing features and functionalities and available resources for implementation and testing.

If more extensive assistance is required to finish developing a new innovation and ensure it meets the minimum level of testing required for inclusion in the WRF repository, researchers are encouraged to consider the DTC Visitor Program (http://www.dtcenter.org/visitors/). This program supports visitors to work with the DTC to test new forecasting and verification techniques, models and model components for NWP, as well as offering an opportunity for visitors to introduce new techniques into the DTC-supported community software systems. Financial and computational resources are awarded to successful applicants.

Stage Two – Comprehensive testing and evaluation performed by the DTC

Due to the large number of potential candidates, a prioritized list of eligible community innovations will be established in consultation with the DTC's Science Advisory Board, Management Board, and, ultimately, Executive Committee. Operational Centers also provide insight and guidance on whether the new technique addresses a specific need in operations and whether the demonstrated compute resource requirements are reasonable for consideration. Innovations that are available in the WRF repository and have demonstrated benefits through the first stage of testing with MMET cases will be given initial priority.

The second stage of testing is conducted by the DTC or a member of the research community through support of the DTC Visitor Program. It is more extensive in nature and may include data assimilation cycling depending on the target application. With a large number of cases tested across a broad range of weather regimes, the overall flexibility and appropriateness of a new research technique implemented in a 24x7 operational system is assessed and the strengths and weaknesses are determined.

It is critically important to ensure that forecast results produced during the second stage of testing and evaluation are immediately relevant to the operational centers. In this context, a functionally similar environment to operations is utilized. In general, the end-to-end forecast system utilized for the extensive testing and evaluation of community contributed mesoscale model innovations will be comprised of a pre-processor (potentially including data assimilation), model, and post processor.

Upon completion of the comprehensive testing for the specified set of cases, extensive verification statistics are computed using MET as described at the end of stage one. It is widely accepted that better evaluation techniques are needed to show improvements in parameterization schemes, especially as the resolution moves to finer scales. To address this, in addition to applying standard verification statistics (e.g., RMSE, bias, GSS), new spatial verification techniques (e.g., object-based, neighborhood methods) are applied in the evaluation when appropriate. For this extensive testing, verification statistics generated for each retrospective case are used to compute and plot specified aggregated statistics, and each type of verification metric will be accompanied by confidence intervals (CIs) where appropriate, computed using the proper statistical method. The overall impact on the resulting forecast when including a new research innovation compared to an operational baseline is assessed by applying a pair-wise difference methodology and highlighting differences that are statistically and practically significant.

Computational resources associated with a new technique will be assessed in terms of total compute time, calculated as the number of processors multiplied by the central processing unit (CPU) time, as well as by maximum memory usage.

Stage Three – Pre-implementation testing at Operational Centers

Upon completion of DTC testing and evaluation during the second stage of testing, the extensive results are shared with interested operational entities, along with the user community-at-large. The ultimate decision to proceed to the pre-implementation testing phase is made by operational centers and is based on a variety of factors, including forecast performance and computational resource requirements. For each target application, operational centers have well-defined, rigorous pre-implementation testing procedures that must be followed; the specifics of the testing performed in the third stage are left to the discretion of the operational center in order to meet their stringent criteria.

3. Summary

By defining a testing protocol for transitioning new innovations from R2O, the DTC is working to ensure that efforts are streamlined to make it both efficient and effective. Ultimately, the testing process is a shared responsibility between the researchers, DTC and operational centers. It is believed that, with better coordination among the NWP community as a whole, major benefits towards improving model physics can be realized, resulting in more accurate and reliable operational NWP forecasts in the end.