The DTC Ensembles Task
A New Testing and Evaluation Facility for Mesoscale Ensembles


The National Weather Service (NWS) has identified accelerated development of mesoscale probabilistic prediction and ensemble forecast systems as a critical requirement. In particular, NWS cosponsored a workshop in 2009 (resulting in a white paper by Eckel et al.) recognizing that

the transition to probabilistic prediction and application is a vital but formidable task [with the] following goal: By 2015, the U.S. will implement a radically upgraded national capability for mesoscale probabilistic prediction to support current and future decision-making needs, helping return the U.S. to a world-leadership role in numerical weather prediction.

A National Research Council report has also expressed this need. The required critical elements of research and development to accomplish this goal are enumerated (for example) in works by Eckel et al. (2009), Buizza et al. (2005), Mass et al. (2009), Kuchera et al. (2009), and the American Meteorological Society Ad Hoc Committee on Uncertainty in Forecasts (2010). These include (but are not limited to) system optimization (to live within computing resources); treatment of model uncertainty (multimodel vs. stochastic; etc.); treatment of initial and local boundary conditions; postprocessing calibration; and display and presentation utilities (particularly those to facilitate user interpretation). For high-resolution ensemble forecasting, other critical scientific and technical challenges include best methods for creating initial perturbations, assimilation of mesoscale observations (including radar data), and verification of ensemble predictions.

In support of these scientific objectives, the Developmental Testbed Center (DTC) has established a testbed platform (the DTC Ensemble Task, or DET) intended to serve as a bridge between research and operations. In the most general sense, the goal of the DET is to provide an environment in which extensive testing and evaluation of ensemble-related techniques can be conducted such that the results are immediately relevant to the operational centers. Examples of these operational centers include the National Centers for Environmental Prediction (NCEP) and the Air Force Weather Agency (AFWA); examples of sources for new and innovative ensemble modeling techniques include universities, other testbeds that focus on specific research topics, and research and development wings of other agencies. One guiding assumption of the DET (as well as the DTC) is that as a joint venture it can function best as an independent evaluator of ensemble forecasting systems and their elements—in other

AFFILIATIONS: Tollerud, Etherton, Toth, Wharton, and McCaslin—Developmental Testbed Center, Boulder, Colorado, and NOAA Earth System Research Laboratory (ESRL), Boulder, Colorado; Jankov—Developmental Testbed Center, Boulder, Colorado; Precision Wind, Boulder, Colorado; and Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado; Jensen, Mirvis, Kuo, Brown, and Nance—Developmental Testbed Center, Boulder, Colorado, and Research Application Laboratory, National Center for Atmospheric Research (NCAR), Boulder, Colorado; Yuan—School of Atmospheric Science and Key Laboratory of Mesoscale Severe Weather/Ministry of Education, Nanjing University, Nanjing, Jiangsu, People’s Republic of China; Koch—NOAA National Severe Storms Laboratory, Norman, Oklahoma; Eckel—National Weather Service Office of Science and Technology, Washington, D.C.
CORRESPONDING AUTHOR: Edward I. Tollerud, NOAA/ESRL, 325 Broadway, Boulder, CO 80305-3328
E-mail: edward.tollerud@noaa.gov
DOI:10.1175/BAMS-D-11-00209.1
©2013 American Meteorological Society
words, as an honest broker of systems and techniques developed and/or implemented elsewhere. As a joint venture primarily between the National Center for Atmospheric Research (NCAR) and the Earth Systems Research Laboratory (ESRL), and not formally affiliated with NCEP or other operational centers, the DET is well positioned to serve that function. Another DET assumption is that in the face of operational demands, forecast centers with or without research branches will not always have the resources to compare several competing modeling components originating in the research community. How effectively the DET/DTC handles the often conflicting requirements of the research and operations communities will to a large degree determine its success in facilitating transition of research findings to operations.

In short, the DET exists to facilitate the transfer of research results to operations in cloud-scale and mesoscale ensemble prediction. In addition to testing and evaluation, the DET will also support and maintain community codes as it conducts extensive testing and evaluation of promising new capabilities and techniques that have been incorporated into these codes. Codes already supported by the DTC will serve as building blocks for the end-to-end ensemble testing and evaluation system to be assembled. Inclusion of research techniques targeted for upcoming implementation by research and development units in operational centers will ensure that DET ensemble capabilities do not lag behind existing operational capabilities. Furthermore, close collaboration with these operational centers will allow the DTC to contribute to future operational decisions.

The DET recognizes that in order to truly act as a bridge between research and operations, community input from both operational and research centers is essential. Toward that end, the DET has engaged a panel of scientists [the Weather Research and Forecasting (WRF) ensemble modeling working group (EMWG)] to provide guidance for DET planning. Other input has been provided by both the Management Board and the Science Advisory Board (SAB) of the DTC. The initial outcome of the collaboration with these entities is a plan for the structure and function of the DET. This paper describes the conceptual structure thus developed, the linkages of the DET with other DTC activities and with the larger ensemble modeling community, and some future plans and anticipated accomplishments.

**DET INFRASTRUCTURE.** Given the DET requirement to facilitate testing and evaluation of competing techniques and capabilities for specific components of the ensemble system, the DET infrastructure is designed to be as modular as possible. In order to keep the testing and evaluation of new ensemble capabilities developed by the research community relevant to operational upgrade decisions, the DET modules are configured to replicate as much as possible the algorithms used by operational centers. Modules to be included in the infrastructure are:

- *Ensemble configuration*—Encodes the characteristics that define ensemble members and their horizontal and vertical resolutions, such that different models and/or different configurations of the same model can be included.
- *Initial perturbations*—Provides the ability to represent uncertainty in initial conditions based on a variety of techniques.
- *Model perturbations*—Provides the ability to represent model-related uncertainty based on a variety of techniques.
- Statistical postprocessing—Provides the ability to specify techniques for fusing information from ensemble and high-resolution control forecasts, climatology, and other sources such as the latest set of observations; to bias-correct or calibrate forecast distributions; and to statistically downscale information to user-relevant variables.

- Product generation—Provides the ability to specify techniques for deriving information from the ensemble, generating probabilistic products, providing decision-support services, etc.

- Verification—Provides the ability to specify techniques to be used to evaluate ensemble and derived probabilistic forecasts.

Concerning the first three of these modules, we recognize that there is no single standardized code (“module”) that will work for all models or modeling activities. As a start toward standardization, however, and as part of a collaboration between NCEP’s Environmental Modeling Center (EMC) and DTC, the workflows of EMC, a related workflow for the Nonhydrostatic Multiscale Model on the B grid (NMMB), and ensemble workflow scripts have been formalized and mutually accepted. Further description of these activities is provided under “Ensemble Configuration Module.”

Implementing these modules involves two important steps:

- Establishing an initial basic capability—This step refers to establishing the capability to implement a promising new technique appropriate to the module.

- Establishing a benchmark—This step refers to establishing the capability to functionally reproduce a current operational technique appropriate to the module. Although exact replication of those products as “benchmarks” will not be the goal, the development of a “functionally similar environment” (FSE) that can credibly and closely emulate those products should be possible. A further discussion of FSEs is provided in the “Computing Resources” section.

Figure 1 presents a schematic illustration of the interrelationships between these modules. There is an overarching workflow manager for the entire ensemble configuration. This ensures that all the modules will work in sequence, and that the needed outputs are completed and compatible with subsequent modules. Once a new capability and its respective benchmark have been established, the DET has the basic infrastructure necessary to conduct testing and evaluation directed at assessing whether the new capability shows promise for improving performance over the operational benchmark.

The DET’s 2011 annual operating plan (1 March 2011–29 February 2012) included detailed plans for the overall DET infrastructure with all modules, a test and evaluation protocol designed to be equivalent to that of other DTC tasks, and basic capabilities for the initial perturbations and verification modules. The infrastructure plans rely heavily on input from the ensemble community obtained through a DET Workshop held in August 2010 (www.dtcenter.org/events/workshops10/det_workshop/index.php), as well as from interactions with the reconstituted WRF EMWG. As an important initial activity, regular
meetings have been held with colleagues at EMC to ensure that DET design, software development, and testing is well coordinated to influence NCEP operations. These meetings have been particularly concerned with NCEP’s upgrade of the operational Short-Range Ensemble Forecast (SREF) system. Released to DET in February 2012 as V. 6.0.0, the upgrade consisted of new code for preprocessing, initial perturbations, model integration, ensemble postprocessing (bias correction and downscaling), clustering, and several ensemble products. The DET contribution to that upgrade is described in the section titled “Statistical Postprocessing Module.” Similar collaborations with other operational centers (e.g., AFWA) are also either anticipated or in early planning stages.

Overarching design. During AOP 2011, the DET began development and testing of the “DET Portal” (www.wrfportal.org), an interactive interface that in its advanced form can be used for setting up, running, monitoring, and evaluating DET ensemble experiments. The ensemble configuration and initial perturbation modules will be the first modules to be incorporated. With the gradual development of the portal, the use of the DET infrastructure by DET and visiting scientists will become much easier.

DET INFRASTRUCTURE MODULES. Ensemble Configuration Module. While model configuration is not a “module” in the same sense as the other modules, the definition and development parameters in the context of a particular modeling system is a set of tasks critical to ensemble forecasting. During AOP 2011, the DET established a basic capability to develop ensemble configurations. Building on the HydroMeteorological Testbed (HMT; http://hmt.noaa.gov) ensemble experience during AOP 2010, the DET infrastructure will incorporate the capability to run various models within both the NOAA Environmental Modeling System (NEMS) and the WRF model frameworks. These models will include the Advanced Research WRF and the Nonhydrostatic Mesoscale Model in the WRF infrastructure, and the NMMB in the NEMS framework. The DET will balance requirements regarding NCEP implementations (i.e., the requirement to work in the NEMS framework) and the community’s continued desire to work within the WRF framework. Specific arrangements for the testing of stochastic perturbation/physics schemes will be taken into consideration when establishing this module.

Initial Perturbations Module. During AOP 2011, the DET has contributed to the next upgrade of NCEP’s SREF system through testing and evaluation of an altitude-dependent perturbation rescaling method. Specifically, as a first attempt at estimating spatial variation of rescaling parameters, the DTC conducted a set of tests to calculate the ratio of ensemble spread to ensemble mean forecast error at four vertical levels: 250, 500, 700, and 850 hPa. Results suggested that the present method of computing initial perturbations is likely fine as is for the previous SREF implementation from 500 hPa to the top of the atmosphere, and the DET subsequently delivered modified code to NCEP/EMC. For the DET work to be fully implemented into the NCEP SREF, side-by-side comparisons of the different rescaling parameters are needed. In addition, these comparisons need to be made using current or future, not past, configurations of the NCEP SREF. During AOP 2011, such comparisons were not possible at DTC. Now, with the code repository in place, such tests can be performed.

In the future, rather than simply adjusting existing methods, entirely alternate methods for creating perturbations can also be tested. For example, starting in spring 2012, NCEP SREF intends to begin using a “blending” method inspired by the work of Wang et al. (2011). The goal of this blending is to combine large-scale features of the perturbed initial states of a global ensemble with the small-scale features provided by the perturbed initial states of a regional ensemble. The rationale for this approach is to obtain information on the mesoscale uncertainty in the analysis, which is more reliably represented by the uncertainty present in the regional ensemble than that in the global model ensemble, as these mesoscales are not resolved by the global model. Best methods for global and local perturbations could also be addressed.

Statistical Post-Processing Module. A wide range of possible postprocessing techniques can be candidates for testing within this module. As examples of activities undertaken to date, the DET was provided with bias correction codes and scripts (from NCEP SREF) and downscaling codes and scripts (from the North American Ensemble Forecast System). These codes and scripts were modified and applied to archived forecasts from the NCEP SREF from June and July 2011 to assess the forecast improvements resulting from these two statistical postprocessing techniques. Preliminary evaluation of results (using both analyses from the Real Time Mesoscale Analysis system and METAR observations as verifying datasets) indicated
that both techniques resulted in improved forecasts. The code and scripts were then ported back to EMC and implemented in a parallel version of the NCEP SREF. Results from this parallel run also showed improved forecasts, and will be part of the new implementation of SREF coming online in spring 2012.

**Products and Services Module.** At the request of Hazardous Weather Testbed (HWT) participants, many ensemble products were generated during the 2010 Spring Experiment. Some of these products proved useful subjectively, while others did not. The DET evaluated a small subset of these. Objective evaluation of single-value products (i.e., ensemble mean and probability matched mean), as well as the probabilistic products (probability and neighborhood probability) for composite reflectivity, quantitative precipitation forecasts, and radar echo top height, showed that each product provided a different level of skill. This result suggests that an ensemble and its products cannot be evaluated by one method and variable alone.

In support of the ensemble task area, the DTC plans to accelerate its implementation of the DET Products and Services module through its collaboration with HWT as well as other NOAA testbeds such as HMT. The approach will be twofold:

- **Identification of new products**—The DTC will evaluate in real time and prepare a report on an evaluation of all testbed research ensemble products (both single-value and probabilistic) for select variables (i.e., composite reflectivity, 1-km reflectivity, and accumulated precipitation). Ensemble mean, spread, and probabilistic output will be compared to operational guidance from AFWA, NCEP, and other operational centers. After the initial evaluation is complete, the DET will work with these operational centers and testbeds to determine if the technique is promising enough to be included in the products and services module.

- **Begin incorporation of new products into Products and Services Module**—The DET has obtained product generation code from two operational centers (NOAA/NCEP/EMC and AFWA) and is working with the testbed research organizations [CAPS and NOAA/ESRL/Global Systems Division (GSD)] to acquire some of their ensemble product algorithms for inclusion in the DET product generation module. An initial basic capability for this module has thus been established, and can undergo further testing at the DET.

**Verification Module.** The DET will apply its verification module—which is largely composed of ensemble verification capabilities developed during AOP 2010 through the HMT collaboration—to its testing and evaluation activities. In addition, verification metric packages consisting of a set of forecast variables, forecast levels, and statistical metrics to be used for several forecast and evaluation challenges (e.g., operational model evaluation, aviation weather forecasts) will be established. These verification metric packages will be structured to be similar to DTC reference configurations (www.dtcenter.org/config) and will be defined through discussions with major operational users (EMC, the Federal Aviation Administration, the Office of Hydrologic Development, the Aviation Weather Center, etc.). Experimental DET runs for the initial perturbation module, for example, will be evaluated using such a defined EMC verification metric package. Additionally, as represented by the sidebar component on Fig. 1, HWT and HMT ensembles may be evaluated beyond the testbed collaboration efforts to improve understanding and interactions with these organizations. Resources permitting, the DET verification module will also be used for the evaluation of some Hurricane Forecast Improvement Project (HFIP) regional ensemble experiments.

**COMPUTING RESOURCES.** From the start, the DET has been faced with a major challenge identifying computational resources to construct and test its infrastructure and to evaluate new methods developed by the community. Fortunately, over the last year the DET has been successful in identifying and gaining access to a much-improved array of computing resources. During AOP 2010, the DET prepared and was granted a request for a startup allocation on the Teragrid computational network supported by the National Science Foundation (NSF). This allocation allowed the DET to test the portability of its infrastructure and run a limited amount of tests in support of its AOP 2011 work. With AOP 2012 in mind, DTC personnel are now operating on the NOAA Zeus machine and will work with the NCAR Yellowstone machine coming online later in 2012. Between these resources, it is anticipated that there will be sufficient computational resources to carry out most of the planned DET activities. Potential additional requests for computing resources are also pending or planned, including renewal and expansion of the initial Teragrid request.

In order to fulfill its stated mission and become a useful testing facility, the DET plans to develop an
FSE. The FSE will support diverse potential research applications on a number of targeted operating platforms. The idea behind this development is to create a testbed computing foundation where candidate research models, data, and workflows can be tested, ported, verified, and modified in computing environments similar to an operational environment. A combination of High-End Computing (HEC) resources, networking, system and third-party libraries, code repositories, data, etc. will help minimize the time required to transfer DTC/DET product deployment into real operations.

Obviously, requirements for similarity may vary between models, testing goals, applications, resources, and targets. However, we foresee that after succeeding with several models and applications utilized at AFWA and EMC (NEMS, SREF, etc.) that involve merging with ESRL/GSD and NCAR/Research Application Laboratory workflows, the DET’s FSE will be able to provide a flexible, scalable, and heterogeneous HEC environment design.

DTC FSE flexibility is approached by building an application-driven “Gnu’s Not Unix” modules package; scalability and an HEC environment design are achieved by porting and integrating a model test-based benchmark suite with currently available hardware. While an exact replication of operational HEC resources is not an objective, ongoing DET requirements identify IBM Power 6 as a targeted architecture, and FSE can utilize the NOAA ESRL Integrated Linux cluster, research IBM Power 575, and the NSF TeraGrid’s Common User Environment.

COMMUNITY INTERACTIONS. The DET established additional critical community connections with ensemble developers and ensemble users by collaborating with NCEP and the National Unified Operational Prediction Capability (NUOPC) on the organization of the 5th Ensemble User Workshop held in the Washington, D.C., area in the spring of 2011 (www.dtcenter.org/events/workshops11/det_11) and with the NUOPC on plans for a workshop in 2012. The 2011 workshop provided important input into the planning and requirements for the statistical postprocessing, product generation, and verification modules of DET infrastructure, and helped strengthen the DET’s interactions with NCEP. It also identified numerous directions for research and development. For ensemble forecast generation, for instance, methods for ensemble configuration, probabilistic forecasting, statistical postprocessing, reforecasting, and product generation were proposed.

The DET has also benefited from ensemble verification capabilities added to the model evaluation tools through work reflected under the verification task and capabilities implemented through the DTC’s collaborations with HWT and HMT. We note that the NOAA/ESRL/GSD ensemble modeling effort for the 2011–2012 HMT winter field exercise made use of the DET infrastructure for its real-time forecast system.

A natural first step in the development of the DET has been to emphasize collaborations with EMC, a research and development branch of NCEP. AFWA has also had a substantial involvement with the DTC over the years, and these joint projects will continue as well. Beyond that, it will be necessary to establish a review system (a protocol of sorts) to make decisions about other potential collaborations. This will be particularly important as the DET is faced with decisions about promising ensemble techniques and critical needs that emerge from other research organizations and from universities with strong ensemble-modeling programs. The DET visitor program represents one kind of interaction that can support and encourage community involvement with the DET, and it is also possible that a more formal proposal review system may be instigated. The DET will also solicit input from the DTC SAB and the EMWG about the best approach to take to make these decisions.

FUTURE PLANS. As the FSE of the NCEP SREF and other modeling systems and the full DET Portal come into being in the years ahead, effective involvement of the broader community with the DET will become progressively easier. As indicated in previous sections, to ease this involvement, a protocol for evaluation and decisions about potential projects, techniques, and collaborators is planned. In addition, DET efforts will extend to other regional-scale modeling work—in particular, to HWRF ensembles. Building upon this ensemble work will entail a greater involvement in the application of ensembles to data assimilation. The DTC has in place an active Gridpoint Statistical Interpolation (GSI) effort, full code that allows it to interact with the community, with the potential to lead to the transition of research results to operations. Developing a code repository for a GSI-hybrid data assimilation scheme will expand this ability, hopefully leading to the transition of numerous HFIP-funded projects onto the pathway to operations.


