Report of the Science Advisory Board (SAB)
of the Developmental Testbed Center (DTC)

November 21, 2013

Introduction

A DTC Science Advisory Board (SAB) meeting was held on 25-27 September 2013 at the NCAR Foothills Laboratory in Boulder, Colorado. According to the DTC Charter, the primary functions for the SAB are: (1) providing recommendations for strategic directions for the DTC, (2) providing recommendations for new codes or new NWP innovations for DTC to perform testing and evaluation, and (3) reviewing DTC Visitor Program proposals, and providing recommendations for selection. This report presents the SAB's recommendations to DTC regarding functions (1) and (2) above, after consolidating the presentations given during the meeting, the notes on those presentations and subsequent discussions (both those taken by SAB members themselves and those meticulously recorded by DTC staff), and additional communication among SAB task teams in the several weeks since the meeting.

Some of the highlights of the meeting itself include:

- An overview of DTC's mission, history, governance, and planning processes, as well as a recap of last year's SAB recommendations and DTC's follow-up to those recommendations, by Bill Kuo.
- High-level presentations by John Murphy (Director, NWS/OST, and Chair of the DTC EC), and Col. John Egentowich (Deputy Director of Weather, USAF, and a member of DTC EC), on the role of the DTC and metrics of its success from the perspectives of these two major sponsoring federal agencies.
- A summary of DTC's community support and outreach efforts by Louisa Nance.
- Presentations by DTC task leads on DTC efforts, accomplishments, and challenges in their particular areas.
- Presentations on strategic plans for NWP at DTC's sponsoring agencies: Geoff DiMego (NCEP/EMC) on EMC plans, Vijay Tallapragada (NCEP/EMC) on EMC Hurricane plans, and Mike Horner (AFWA) on AFWA plans. These presentations were an extremely useful addition to the meeting and should be considered essential in the future.
- Presentations by SAB members on the SAB's view of issues and strategic directions in each task area.
- A healthy discussion revisiting the role of SAB in the context of the DTC Charter and recent SAB meetings and reports.

Many of the issues brought up in the above presentations and discussions are born out in the more detailed recommendations presented in the next sections. Those recommendations are organized by DTC's Task structure. A few of those recommendations arose repeatedly across different tasks, and were considered by the SAB to be of particular importance, and so are specially listed here:

1. Since HRRR and HRRRE will play an increasingly important role in EMC's NWP suite, DTC should become more involved in HRRR T&E, O2R, and R2O.
2. The need for a software infrastructure for running and inter-comparing cycled DA and forecast experiments was strongly emphasized, as in 2012. This would be a significant undertaking that would require resources beyond DTC, but DTC could play an important role in developing and supporting such a system.
3. The emerging need for physics that operates smoothly across a range of horizontal resolution (or "scale-aware physics") was a recurring theme of discussions. DTC should facilitate testing of such parameterizations, and become familiar with international research community efforts and workshops on this topic.
4. Testing in the DTC should work toward cycled experiments, especially when evaluating new physics packages. The accumulation of physics errors in cycled mode can reveal deficiencies that are masked by continued re-initialization with a separate analysis.
5. DTC should continue active community engagement in task areas, including hosting a physics workshop and verification workshop. DTC should also facilitate ensemble configuration discussions, and remain involved in relevant NOAA testbed activities.

Task Summaries

In this section, more details of SAB meeting discussions and recommendations are presented.

a. Mesoscale Modeling

The clearest signal from the discussions about mesoscale modeling priorities for the DTC was the need to move testing and evaluation into the realm of convection permitting forecasts. The research community has been moving in this direction steadily for 10 years. Quasi-operational systems such as the High-resolution Rapid Refresh (HRRR) produce hourly forecasts across the CONUS with a grid spacing of 3 km. Experimental ensembles at 3-km, for CONUS and sub-CONUS scales, are being run for the prediction of severe convection every spring. Operationally at NCEP, NAM nests and Hi-Res Window runs have been made at convection-permitting scales (3-6km) since 2004.

The DTC should enhance their collaborations with the HRRR group. This would be a useful way to establish testing for CONUS-scale convection-permitting forecasts within the DTC during the coming year. Evaluation of high-resolution forecasts was also discussed. Developments such as MODE and neighborhood methods should be integrated into the testing of convection-permitting forecasts, as should convective-scale observational data beyond rainfall rates and radar reflectivity. Such efforts would also be useful as a collaboration with the HRRR and EMC developers.

Because it is clear that EMC will increasingly emphasize ensembles of ARW and NMM-B, the interoperability of physics and the transferability of testing between the two cores should be addressed by the DTC. DTC should work with EMC to discuss priorities on which physics packages are transitioned from the WRF infrastructure to NMMB and then work with physics developers to achieve the integration. Questions still exist regarding the utility of preliminary testing in the WRF infrastructure prior to integrating new packages into NMM-B. At a minimum, DTC staff members need to be conversant in WRF and NEMS frameworks. Related to this issue is the extent to which the DTC supports O2R for NMM-B in the NEMS framework. The DTC should consider holding a workshop for friendly users of NMM-B. Given the level of interest, the DTC can assess the importance of broader support of NMM-B to the research community.

Continued effort needs to be made to promote the growth and use of MMET (Mesoscale Model Evaluation Testbed) and make it an effective facility for the community-at-large to use. DTC should continue to solicit current, relevant cases. New cases should include longer-term evaluations (weeks to months) to elicit model mean behaviors and errors. This could be done most simply with sequential cold-start runs, but most effectively with a DA cycling capability to simulate the operational forecast environment. Additionally, the testing of new releases of WRF ARW with a set of commonly used physics options that DTC has performed recently has been extremely valuable. These tests have revealed important new biases that have resulted from changes in model physics.

More generally, testing in the DTC should work toward cycled experiments, especially when evaluating new physics packages. The accumulation of physics errors in cycled mode can reveal deficiencies that are masked by continued re-initialization with a separate analysis. Furthermore, there can be significant adjustments in the first several hours because of the incompatibility of the physics of the forecast model compared to the physics of the model that was used to produce the initial conditions. However, the best path forward was not immediately determined. MMET is perhaps a logical place to first implement cycling for the examination of retrospective test periods of interest, as suggested above. In the long run, however, cycling needs to be an integral part of all MM testing. Since HRRR & NMMB both use GSI and there is already a GSI support group within DTC, this should be feasible.

The emerging need for physics that operates smoothly across a range of horizontal resolution (or "scale-aware physics") was a recurring theme of discussions. AFWAs intention of moving toward a unified model as early as 2016, with a variable-resolution global model being one possible solution, may also bring attention to this topic. But scale-aware physics is also critical for models with uniform resolution that increases with increasing computer power. There should not be abrupt changes in the performance of physics as resolution changes. Areas for emphasis will be the interaction of boundary layer schemes and shallow convection. These schemes take on added importance as parameterizations of deep convection become inactive at convection-permitting grid lengths. Furthermore, continued testing on microphysics, especially the interaction of clouds and hydrometeors with radiation and aerosols, will be important.

Because of the move to convection-permitting grid lengths, it was felt that another community workshop on physics would be desirable. The feedback obtained during the last workshop (June 2011) was very valuable and the conversation should be continued/sustained, if possible. This report has outlined numerous areas of physics that can be emphasized [e.g. interoperable, smoothly scalable]. It was also suggested that the DTC should connect to other workshops already being planned elsewhere to establish connections and opportunities for leveraging.

b. Ensembles

A growing number of groups in the meteorological community are researching convection-allowing (1-4 km horizontal resolution) ensembles. Various systems are in place operationally or pseudo-operationally already, but the design of those ensemble systems so far has been relatively ad hoc, primarily due to the large costs associated with such fine horizontal resolution. Because maturation of operational convection-allowing ensembles is possibly still several years off, the DTC may provide greater service by targeting system components expected to be in place two or more years in the future. More specifically, DTC could play a key role in answering many of the outstanding questions on operational convection allowing ensemble systems in the following (non-inclusive) areas:

- Physics/model core combination testing
- Stochastic physics to improve ensemble reliability and probabilistic forecasting
- Design tradeoffs (number of members, vertical levels, forecast hours, etc.)
- Benefits from time lagging
- Ensemble data assimilation methodologies

DTC should be cognizant of operational goals and constraints when performing any tests of the above. For example, traditionally the Ensemble task has focused on ARW cores with LAPS initialization. Such a configuration is not a component of the future High Resolution Rapid Refresh Ensemble (HRRRE) whose initialization will be GSI and hybrid-EnKF based. Testing of a 3 km configuration with ARW core and HRRR physics would be a step towards greater operational relevance. A further step would be testing of NMMB with NAM physics. At a minimum, the DTC could play a role in bringing the HRRRE to fruition by coordinating a joint EMC, GSD, and PSD meeting to discuss options for future generation system. Additionally, it should be noted that some of the other task areas, such as mesoscale modeling and data assimilation, will likely be looking at capabilities that would benefit operational ensemble systems, so that perhaps some of these testing efforts could be completed jointly, to the benefit of multiple user communities.

In recognition of limited resources and expertise available, the ensemble configuration, perturbation, and physics tasks (DTC Ensemble Task Modules 1-3) appear to be the most fruitful at this time. However, other areas of ensemble research are also reaching maturity and could be considered for DTC evaluation as resources and opportunity allow. For example, operational forecasting and decision-making tools for ensemble data are needed to ensure users get the most benefit from the data, and this is a key and often neglected part of the weather forecast process. These include statistical correction methodologies, probabilistic algorithms, clustering approaches, techniques to identify high-impact forecasts, and visualization/data mining software. A DTC effort to build a common software framework for these sorts of tools would be quite beneficial to facilitate both research and a rapid transition to operations for these techniques.

One of the goals of the DTC is “facilitating operational and research collaborations” and this is an area where DTC could reach out to university or research labs that have resources for ensemble testing, and determining if partnerships can be built to perform more extensive tests than DTC resources allow. This would also hopefully tie research organizations in more closely with the nation’s operational modeling goals.

c. Hurricanes

The main role of the DTC should remain as the connection between the operational and research communities. DTC staff should work side-by-side with visitors and experts in the community to identify model deficiencies, implement innovations, and conduct tests. The DTC staff does not have the resources to carry out development, create the model capabilities needed for testing, and perform diagnostics.

As also discussed under the DA Task, DTC should strive to create an ECMWF-like infrastructure that supports researchers in code management and experimental design and configuration. This infrastructure should make it relatively easy to reproduce others’ experiments and understand exactly what codes, observations etc. were used to make the runs. Also, the Hurricane task needs to work synergistically with the Data Assimilation and Ensemble tasks, since all these aspects are critical for hurricanes. For example, the Hurricane Task could work with both the DA Task and Ensemble Task (as well as the EMC Hurricane Team) in efforts to test the field alignment technique for improving background error covariances in ensemble-based data assimilation for hurricane forecasting (as discussed in the DA Task section below).

With regard to modeling systems, while it is expected that HWRF will eventually become an NMM-B-based system in the NEMS framework, the transition is not expected in less than 3 years. Therefore, it is sensible for DTC to continue working with HWRF in the NMM-E framework for now. EMC’s planned configuration of HWRF in the NMM-B framework involves a global model with multiple moving nests. To prepare for this implementation, DTC should provide support to the basin-scale effort in the WRF-NMM-E framework. The basin-scale configuration consists of a large parent domain with multiple moving nests. DTC should assist AOML/HRD in integrating the basin-scale development onto the centralized HWRF code, and prepare it for becoming a supported, optional configuration in the HWRF public release. It is important that the DTC continue to increase the flexibility of HWRF so that it can be configured in alternative ways for research. The basin-scale is an example of that, and additional flexibility should be sought in other HWRF aspects as well.

EMC plans to run HWRF coupled with downstream models in the near future (wave, storm surge, and inundation). Due to resource limitations, it was not recommended that DTC start working with those additional components at the moment. Instead, DTC can contribute in the areas of land surface modeling and precipitation forecast, as detailed in the next section.

While DTC's main focus should remain the improvement of HWRF, DTC should not restrict itself to working only with HWRF. Important insight can be gained through diagnostics of HFIP Stream 1.5 and 2.0 models. One area that would be of interest is rapid intensification performance in the various models.

The SAB responded positively to DTC involvement in addressing a concern from HFIP management about a decrease of track skill in GFS model output on days 4-5 of the forecast. However, the SAB felt that the first step was to confirm that there is indeed a problem, since the reported indications of the problem are based on a small number of cases. This first step is one the DTC would be ideally suited to undertake.

DTC should invest more in diagnostics and evaluation to complement its testing activities. Such activities should provide feedback for developers about model deficiencies and avenues for improvement. Priorities for verification and diagnostics are in precipitation, large-scale fields, radiative fluxes, sensible and latent heat fluxes, and model energetics.

Two areas that DTC should engage in improvement of forecasting skill are rapid intensification / rapid weakening (RI/RW) of tropical cyclones, an important HFIP goal for which little progress has been made; and precipitation performance for land-falling storms in HWRF, which has yet to be evaluated.

Testing of promising new or improved physical parameterizations should remain a priority for DTC. Such testing should be done in conjunction with the scientists who have developed or improved the parameterization in question. Successful parameterizations are the ones that improve forecasts for generic areas and weather systems, as well as for tropical cyclones. Additionally, scale-aware physics, which can work in a model running both low- and high-resolutions, is a high-priority.

It is recommended that DTC interact with the HFIP physics (strategic and tiger) teams to define priorities for testing. Also, DTC should be abreast of research efforts and workshops in the international community, particularly in the area of scale-aware physics. It is important to consider physics suites, not isolated packages. The broad areas for prioritized testing are

- Aerosol-aware microphysics and radiation parameterizations
- Land surface model
- Planetary boundary layer parameterizations
- Stochastic parameterizations

d. Data Assimilation

The data assimilation (DA) subgroup of the SAB presented a list of current directions in data assimilation research. The ones that were suggested to be the most relevant for the DTC at this time are:

1. Mesoscale/convective scale ensemble-variational data assimilation for NCEP's NAM, RAP/HRRR, and HWRF systems, and AFWA's current regional forecast system.
2. Methods for better representation of model uncertainty (stochastic physics).

3. Field alignment for tropical cyclones (as an upgrade to TC relocation for HWRF).
4. Improving the use of radiances (clear and cloudy) in regional data assimilation.

It was noted that there is significant overlap with the ensemble task for many of these research areas, especially (1) and (2), and perhaps (3). During the discussion it was suggested that item (4) would best be accomplished in collaboration with the hurricane task and the HFIP data assimilation team, since this is a top priority for HFIP.

Although RAP, HWRF and NAM are (or soon will be) using the GSI ensemble-variational hybrid data assimilation system, they currently use the global GFS EnKF ensemble to define the background-error covariance. For hurricane and storm-scale DA, the covariances from the relatively low-resolution global ensemble are probably highly sub-optimal. For this reason, it will be important to test the use of high-resolution EnKF ensembles cycled with the regional models (WRF-ARW, HWRF and NMM-B) in the GSI hybrid DA system. DTC will need to become familiar with and support the current operational NOAA EnKF code in order to contribute to this.

Regarding item (2), there are a number of schemes currently used operationally for global prediction that could be tested in the regional models. Some of these (stochastic kinetic energy backscatter and stochastically perturbed physics tendencies) are already implemented in WRF-ARW, but have not been extensively tested in the context of ensemble data assimilation. Others have been implemented in the GFS and are currently being tested within the GSI ensemble-variation hybrid, and could be ported to the regional systems.

During the discussion of item (3), Vijay Tallapragada mentioned that preliminary efforts were made to integrate the MIT field alignment technique into the HWRF system, but that work was halted due to software licensing and/or patent issues. It was suggested that this could be revisited, since these restrictions may no longer exist. Field alignment of tropical cyclones (or any other coherent feature) is a potentially important issue in data assimilation, since position errors can result in highly non-Gaussian background-error covariances that cause Kalman-filter based systems to perform sub-optimally.

There was some discussion of whether the DTC could be involved in re-factoring the GSI code to be more easily extendable. One example is the current restriction to only rectangular grids. However, it was suggested that efforts involving large-scale code changes should originate at the operational center, to avoid the "boarding the moving bus" problem at the stage when those modifications need to be merged into the operational code repository.

The need for a software infrastructure for running and inter-comparing cycled DA and forecast experiments was emphasized again. This was first discussed in the 2012 SAB report, but DTC sponsors elected not to fund such an activity in 2013. The SAB emphasized again that such a system (inspired by the PrepIFS/ODB system used at ECMWF) would be crucial to fulfill DTC's mission of improving R2O transitions. This is particularly true for DA related development, since the fully cycled DA forecast system is so complex and difficult for visiting scientists to master in a short time period. It was also emphasized that such a system must also be used at the operational centers in order to be effective at enhancing R2O. Therefore, the operational centers will have to be intimately involved in the development of the software.

e. Verification

The discussion covered the big issues facing the verification community. The DTC verification group is a clear leader in the field and the challenge is how to capitalize on the expertise of the relatively small group. Many questions were raised and remain unanswered by the research community. Thus, as solutions present themselves, they should be incorporated into the efforts of the DTC verification group. The questions include:

- How to estimate and incorporate uncertainty into verification, i.e., probabilistic forecast verification?
- How do we estimate and incorporate observation uncertainty into verification?
- How to verify high-resolution forecasts? What about multi-scale models?
- With what methods / observations? What do we have for non-precip fields? How do we handle "negative objects"?
- How can we better incorporate temporal and spatial information into verification?
- How can we make verification relevant for more users?
- How can verification connect better with DA? They have similar issues in uncertainty, incorporating new observations, multiple scales.
- Is there value to doing DA-style pre-verification?

At many points during the SAB meeting, the issue of differing user needs came up. A great challenge exists in making relevant metrics as models become higher resolution and users expect more from the models. The DTC Verification team could help novice verification users by having a problem of the month or supplying predefined metrics for certain verification problems, emphasizing the benefits of using the Model Evaluation Toolkit (MET) package. Users tend to be conservative in using and understanding new techniques until they can clearly see the benefits.

Two ideas that were raised that involve somewhat more novel approaches to model evaluation, and are areas that DTC could help facilitate for the research community:

- It would be helpful to have a capability to access internal model diagnostics, such as tendencies and intermediate quantities, that are not part of standard output. Although this capability is not strictly "verification" because there is typically no data against which to verify such quantities, such a capability would allow developers and testers who are exploring model error to address more than just the question of "how big?" but also "why?" DTC, in collaboration with NCAR's WRF development team, should take the lead in developing code modifications and/or documentation in the WRF-ARW package to enable easier user access to internal model diagnostics.

- Cycled DA provides useful intermediate fields: *innovation* (observation-minus-background), *analysis residual* (observation-minus-analysis) and *analysis increment* (analysis-minus-background). When examined statistically, these fields can shed light on systematic model errors. As DTC ramps up its support for DA cycling, it should ensure that these fields are accessible to users, with possible integration into DTC verification tools (MET).

The topic of hosting a verification workshop was raised. In the past, DTC held these workshops nearly every year. It has now been three years since the last one, and the DTC Verification Task believes another is long overdue. Growth in expertise in the verification community has typically come from such workshops. Such growth also makes incorporation of new concepts into the library of available techniques much easier. Suggestions for locations included the new NCEP building and the usual NCAR Boulder location. Topics mentioned were hurricane verification, uncertainty quantification and use, and parallel developments with data assimilation. In particular, a great number of data assimilation experts reside at NCEP, so if the workshop were to be held at that location, data assimilation should be a primary topic.

Finally, the SAB suggests that DTC stay involved with the NOAA Testbeds (e.g. HWT, HMT) and use data and lessons learned during NOAA Testbed experiments to contribute to design of evaluated systems.

f. Global modeling

Although there is currently not a defined DTC Task regarding global modeling, both DTC and the SAB have over the past two SAB meetings recognized the importance of this topic and its growing overlap with DTC activities. With the trend of operational global NWP moving from low-resolution spectral to high-resolution gridpoint representations, it won't be long before operational global models are nearly indistinguishable from the regional mesoscale models we are now familiar with. Global models being run at mesoscale resolution will face the same challenges of mesoscale modeling, including testing and tuning of physical parameterization at high resolution, optimal use of mesoscale observations, and the verification of high-resolution model forecasts. The work that DTC has been performing related to regional and mesoscale modeling will be relevant to future global NWP.

There was considerable discussion on whether DTC should engage in global modeling. The SAB felt that these models are not sufficiently mature to warrant DTC activities at this time, but that DTC should continue to monitor progress in global model development. In the event that DTC becomes more active in testing and evaluation of global models in future years, three possible areas of participation might be: (i) testing of scale-aware physics for variable resolution global models, (ii) verification of global models being performed at high-resolution and (iii) conducting workshop in collaboration with operational centers on the planning and development of next-generation global models. A significant challenge would be re-tooling of DTC capabilities for non-Cartesian grids, since many of the emerging gridpoint global models (FIM/NIM, MPAS, GME, NICAM) are adopting icosahedral or similar grid structures.

SAB Structure and Role

After 2012, Professor Cliff Mass (U. Washington) stepped down as chair of the SAB after serving for multiple years. The DTC Charter is silent on the term of the SAB Chair. It only states "*the SAB may elect one of its members as chairperson to facilitate its* *business*". SAB decided to make this a rotating chair, with a one-year term. Mark Stoelinga serves as the current SAB Chair, and Josh Hacker has been elected as the next Chair.

There was also considerable discussion among the SAB and DTC MB/EC members present about the role of SAB, and attempting to realign this role with the specifications in the Charter. The SAB recognized that it should avoid making recommendations that are purely programmatic or budgetary, as those are the purview of the MB and EC. The SAB should not attempt to provide what is essentially a program review. Instead the SAB should be forward-looking and focused on science-based strategic recommendations. These recommendations should be made in a context of awareness of DTC's capabilities and constraints. The increased role and responsibilities of the DTC Task Leads in the SAB meeting helped in this regard, focusing the discussions on actionable recommendations that are useful and relevant to the crafting of DTC's Annual Operational Plan (AOP). However, at the same time, the SAB members felt strongly that SAB must maintain its independence, write its own recommendation report, and not be constrained in its discussions and recommendations.

It was pointed out that the DTC SAB is not compliant with the Federal Advisory Committee Act (FACA). This does not affect SAB's role with DTC because DTC is not a government agency, but DTC and SAB should be aware that SAB cannot directly advise DTC's sponsoring federal agencies. Nevertheless, the SAB will maintain a broad scope in its recommendations, even if it includes issues or topics that significantly overlap with the strategic plans of the sponsoring agencies.

The participants generally felt that this was a more productive meeting (compared with previous SAB meetings). Inviting operational centers to discuss their future plans was well received, as it provided excellent dialog between operational centers and the SAB. This was very helpful in formulating future direction for the DTC, and is something that DTC SAB should continue to do.

The DTC will conduct a SAB teleconference meeting in approximately six months. The focus will discuss on how SAB recommendations were incorporated into DTC AOP and the reaction of DTC Management Board and Executive Committee to the SAB recommendations.