

DTC Science Advisory Board 2019 Report

Attendees:

On Site: Vincent Larson, Phil Pegion, Evan Kuchera, Rusty Benson, Vijay Tallapragada, Zhaoxia Pu, Todd Hutchinson, Alex Reinecke, Gretchen Mullendore, Yannick Tremolet, Tom Galarneau, Andrew Gettelman

Remote: Xuguang Wang, Yvette Richardson, Robert Pincus

Dates of Meeting: 19-20 September 2019

Overview

As described in the DTC [charter](#), the DTC provides a “means for the NWP community to test and evaluate new models, technologies, and techniques for use in research and operations and serves as a bridge between research and operations to facilitate the activities of both halves of the NWP community”. In practice, the board sees DTC fulfilling this charter through roles that can be grouped into three categories:

- Enabling community collaboration and communication amongst researchers, operational centers, and other interested parties
- Developing and supporting common infrastructure for both research and operations
 - software frameworks (e.g., CCpp) and tools (e.g., MET)
- Testing and Evaluation
 - Promoting development and advancement of algorithms (e.g., parameterizations)
 - Selection and migration to operations, primarily in the form of software packages

DTC’s organizational chart shows that DTC is organized with one group for each of the above three categories. This model is largely successful. Today, researchers and operational centers often use the same tools (MET, UPP, etc.) and have access to the same model components (WRF, GSI). This has enabled techniques developed in research to migrate to operations, often facilitated by DTC infrastructure.

The DTC leverages funding from many agencies, and leverages staff that report into different institutions to provide software, frameworks, and ultimately, decision-making information to stakeholders. By coordinating projects that meet the varied stakeholder needs, the DTC is able to make larger contributions to the NWP community than would be possible with more limited

funding from fewer agencies. Further, by leveraging staff from different organizations, DTC utilizes the diversity of personnel to incorporate expertise available within the different organizations. While the strengths of the funding and staffing model are evident, the board also recognizes that the varied stakeholder interests can sometimes be orthogonal, and it can be hard to recognize the overall mission of the DTC when viewing different projects side-by-side. In fact, during closed door meetings, the board did have a difficult time understanding the DTC mission. With that, **there is a suggestion that the DTC reiterate its mission throughout future SAB meetings, and specify how individual projects fit into the mission when presenting those projects during the SAB meetings.**

During the meetings, the DTC asked the SAB to provide broad forward-looking scientific advice (such as fostering more successful R2O across the entire atmospheric science community). However, there were times when the board delved into organizational and funding specifics and issues. While this certainly can't be 100% isolated from providing scientific advice, it may not be what is most needed by the DTC director from this board. **We would suggest that next year, the requests made of the board be reinforced throughout the meetings, so that the board can go into the closed door meetings and focus on the issues that the DTC needs most scientific help with.**

The rest of this report will cover specifics for each of the topics above and conclude with recommendations.

Enabling Community Collaboration

To support collaboration we suggest that DTC view itself as a hub for sharing information, code and concepts, to facilitate scientific advancement across the weather enterprise. This would be accomplished primarily through the selection of collaborative tools, guidelines for using them, and support of their use in DTC-related work. DTC may want to establish a policy that information sharing and collaboration across the community should be a part of any DTC projects. This is consistent with some current activities, for example user support for METplus.

The evolution of tools used by DTC to support collaboration has been effective. The transition to git/github is enabling more natural collaboration, as is the recent transition from email support to "forum"-style support, which fits well in today's more open collaboration. The board supports these transitions and suggests that the DTC seek other opportunities along these lines. One suggestion was to investigate additional communication tools such as Slack, Microsoft TeamWorks, but board recognizes that the DTC must carefully select tools that are expected to be widely used and lasting before simply implementing them. The DTC should see if there is opportunity to leverage tools that other organizations (e.g., NOAA, NCAR) may already have in place. **DTC might consider a user survey to determine how users would like to communicate and what tools from their experience might be most useful..**

Development and Support of Common Infrastructure

CCPP/MET

There was much enthusiasm on the board for continued support and development of infrastructure tools such as MET and CCPP, which the board sees as enabling transitions between research and operations.

The board remains unclear as to whether DTC's role in the CCPP is to enable the implementation and testing of parameterizations, or whether they are (or view themselves as) gatekeepers whose mandate is to decide whether some code is or isn't accepted into CCPP. If DTC is a gatekeeper the board identified concerns about what physics packages would be included, and even if, DTC would be hosting the packages. If DTC is hosting, the governance of what gets into the repo may need to be defined.

This question seems to arise every year and deserves clarification. There was no consensus within the board on what the DTC's role should be.

Visitor Program

The board views the visitor program as successful and supports DTC's desire to improve the quantity and quality of visitor proposals. The board suggests that DTC does more advertising of the program and directly solicits requests for people to write a proposal. The board also continues to support approval from a DTC staff member on a project summary before accepting a full proposal for committee review.

Retirement of Software

The board recognizes that there is a cost to maintain and continue to support software over time and fully supports the DTC's push to reduce support for and ultimately retire older software (e.g., parts of GSI, WRF, etc.) that has a limited user base. Following this line of thinking the board suggests that the current UPP and the entire GSI system be considered for reduced support and ultimately retirement. Continuous review of when to retire packages should be continued.

Management of NWP Workflows

General NWP workflow management was discussed several times by the DTC staff and board members, especially pertaining to the need for UFS workflow management. **The board did not come to a consensus on how the workflow should be supported, but did agree that workflows are often organization-centric and can be difficult to standardize.** One operational center is likely to have different requirements from another operational center.

Often, the workflow design and implementation, is what enables different centers to meet their requirements. Designing a workflow that meets the needs of multiple centers could lead to a complex system that is hard to maintain. Further, implementing the same workflow management software in research and test systems, could further complicate the software and it ultimately may not effectively meet the needs of the researchers. **Before committing to building a common set of workflow management software, the DTC should understand the needs of all users that it intends to support, and, determine if it is possible and reasonable to expect varied users to use common software.** If requested and supported, the DTC could perhaps consider a library of different workflows, or information sharing for different workflow code and practices, perhaps hosting a repository and discussion forum on workflows. **Understanding user base and needs is critical before developing any workflow systems or libraries. If unclear, the primary user base should be clarified with the funding agency.**

Testing And Evaluation

Presentations during the meeting demonstrated DTC's expertise in testing and evaluation. In many cases the exercises support better understanding of NWP model behavior and provide quantitative advice as to what to implement in operations. There is clearly interest and enthusiasm within DTC, along with tools and staffing, to undertake more activities along these lines and to provide great interpretation of the results.

The board notes, though, that testing and evaluation is often more limited than is ideal. Sample sizes are often quite small and/or limited in scope, perhaps due to limited computing resources. The board was concerned that the scope of current activities is often not sufficient to thoroughly inform decisions.

To avoid the risk of doing the job half-well, the board recommends that DTC make a conscious decision as to whether to consider testing and evaluation as a central part of its mission, or to support this work with infrastructure and leave the running and analysis of forecasts to the organizations making the decisions. If this line of activity is embraced DTC should seek opportunities for funding for dedicated (or sufficiently reserved) computational resources, perhaps through new programs such as EPIC. **While the board agreed that the amount of testing and evaluation done by DTC is less than ideal, it did not agree on whether the program should be expanded or reduced.**

The board did note that a role for the DTC in evaluating candidates for operational implementation runs the risk of being a conflict of interests. Many DTC staff wear multiple hats within and across organizations. Many of these joint affiliations are with organizations whose parameterizations are being assessed for implementation. This is especially notable for employees at NOAA/GSD: many staff affiliations are listed as GSD/CIRES/DTC/GMTB, and DTC accomplishments were highlighted in a recent GSD newsletter as if they were part of that division's work. Since GSD is actively championing their own physics suites for operations, this

overlap in staff and missions clearly threatens DTC's credibility as an independent assessment body.

Recommendations

A summary of recommendations is here:

- Review and enable modern tools for communication amongst DTC collaborators
- Continue retirement of legacy applications
- Continue software transitions to git/github
- Determine DTC path for T&E, expand or contract the T&E program
- Promote the visitor program and solicit proposals
- Fully evaluate whether workflow management is appropriate for DTC before embarking

Addendum

The following addendum provides specific recommendations from individual board members. While some of the comments were incorporated into the text above, they are also left here verbatim

Suggestion for Physics Evaluation Process from a Board Member

DTC's main role is R2O, specifically helping to translate new algorithms and software from the academic community into operational forecasts. This requires "bake-offs" between the default parameterizations in UFS and new parameterizations from the community. (Without bake-offs of some sort, the R2O process would be too subjective and politicized.) A challenge of bake-offs is to incentivize parameterization developers to spend the considerable time needed to develop parameterizations and tune them until they can perform better than the default, i.e., overcoming the "valley of death." A discouraging factor is that in any bake-off, there can be only one winner. If the losers are simply informed that they have lost, without any promise of a re-competition, then the losers will become discouraged. Then, unfortunately, a parameterization suite that is close to winning won't receive the final push it needs in order to replace the default.

To mitigate this problem, DTC might consider forming and maintaining an ensemble of suites of parameterizations that are under development, plus the default suite. Then DTC could run a global UFS forecast that uses each suite on a regular schedule. Developers would be given free rein to tweak their suites in order to beat the default. If a suite begins to consistently beat the default over some period of time, say, 6 months, then it will be considered by EMC for replacement of the default in the next development cycle. The concept is in some respects similar to a very scaled-back version of NOAA's Hazardous Weather Testbed, except that the DTC version would be targeted toward global forecasts. (However, DTC has much more limited resources and cannot expect to convene a meeting of scientists to analyze the forecasts.)

Regular scheduled re-competitions would avoid the need to tell a developer that he has lost. Instead, the developer can simply be told that he hasn't won *yet*. Each developer would be allowed to keep tweaking his results until he beats the default or simply gives up. But the key point is that it would be the developer who makes the decision to quit, not a bake-off judge who terminates the losers. (If a developer does not update his parameterization suite after a bake-off, the forecast would not need to be re-run for that suite.) This would mitigate the contentious nature of bake-offs.

There are many details that would need to be settled before such a "DTC ensemble" could be implemented. Resources are scarce, and so the ensemble size would have to be limited to, say, 10 suites. (The DTC could start with the 4 suites that are already implemented.) A developer who cannot break into the top ten, or who falls out of the top ten, would have to be told to quit. Furthermore, somebody would have to merge the developers' code into new versions of the UFS host model. (The 4 suites are already on the master branch at EMC, obviating the need to merge them.) The metrics for model comparison should be based on EMC's standard metric for the inclusion of a new parameterization, but perhaps scaled back. Computer time would need to be found in order to run the ensemble. Hence the computational cost of the global runs would need to be limited.

The scientific possibilities of a DTC ensemble are exciting. One could envisage EMC calling a developer and saying "Your suite produced excellent results for metric A but poor results for metric B. Can you modify your suite so that it improves metric B? If so, commit the code change before date X and your suite will be re-run." This would motivate developers. In addition, one could imagine outsiders analyzing the results from the ensemble and comparing what parameterization strategies seem to work best for particular metrics. For instance, if a suite abruptly improves for a particular metric, outsiders would be able to see the code change in the repository that led to the improvement. This could be illuminating. Over time, the DTC ensemble might evolve into a highly skillful one. Each ensemble member would be based on a different parameterization approach, and some members would be based on a different parameterization philosophy altogether, and yet each ensemble member would be well-tuned and skillful because each would undergo regular development by incentivized developers. The resulting ensemble would be both diverse and high in quality.

Secondary Suggestion

Recommendation for a T&E framework that would be coordinated with NCEP's testing.

DTC Modeling Testbed analogous to the Development Phase section 3.1.1 (*Project Plan and Charter for Global Forecast System (GFS) V16.0.0* - previously shared with DTC):

1. Determine whether a proposed physics scheme should be accepted in CCPP
 - a. The developer must implement scheme in CCPP and run SCM cases and verify, done by the developer and possibly during a visit
 - b. DTC enables a scaled-down version of EMC's tests. Either run by DTC staff or developer. However, there is disagreement among the group on whether DTC should simply enable T&E for others, or be the organization that does T&E and in fact, do more T&E.
 - c. DTC would likely need to determine some criteria to get into the game, e.g., compile, run to completion and look realistic
2. DTC may need to limit the total number of parameterizations to incorporate over time and have a way to eliminate schemes
3. The stages should map to the CCPP levels of acceptance (developer, community-supported repo, oper)

Other items

CCPP

DTC should figure out with NOAA what is desired for CCPP support and ongoing maintenance/testing of physics parameterization suites. It was not clear from the presentations what the role of the DTC is in the operation of CCPP and the transition to CCPP physical parameterizations. Should the DTC maintain a repository of tested CCPP parameterizations and/or 'suites'? If so, and assuming this should be open source, then who manages it, and who 'vets' parameterizations going in, as well as ensuring that reference 'suites' match operational code in the UFS. The DTC should work with sponsors to figure this out.

DTC has helped developed CCPP, but is it going to maintain a repository? Has that been agreed/decided/funded?