

SAB recommendations from September 2018 meeting

Prepared by Philip Pegion and SAB board

Executive Summary

Following the meeting that occurred September 26 & 27, 2018, the SAB has learned a lot of what DTC does, and how it is constrained by its sponsors. We felt that DTC is doing a great job in several areas that are highlighted below and wish to seed sustained support in those areas. Here is an overview of the core comments from the SAB:

- First, the SAB feels that given the tremendous changes occurring within NOAA, NCEP, and NCAR, the DTC has a unique opportunity to refocus and refine its mission for the next decade or so. We encourage the DTC staff and its sponsors to consider this opportunity to help bring NOAA's next generation prediction and verification to a wider community.
- The SAB was not clear on the DTC mission from the presentations and discussion, Refining the DTC's mission (goals) and role with existing sponsors is important. If the DTC is a 'bridge' between two communities then who exactly are the customers now on each side of the bridge. Further what potential types of future customers should be considered? What will these customers want? It seems that there is a bit of a disconnect between the research community's needs/wants and the operational community's needs/wants that stretches the DTC.
- The SAB feels that while funding is year to year, there is sufficient continuity in funding that more projects should be structured on longer time scales - at least 2 years. In particular, we feel that funding "one-off" case studies which requires only 2-3 months of work is a sub-optimal use of DTC staff and resources. Longer projects will enable answering more complicated questions that the global or regional systems need answered.
- The SAB agrees that the CCPP has the tremendous potential to accelerate the testing of new physics or suites of physics in different models, but there is resistance to its widespread adoption in the operational community. The DTC could help answer the current performance and usability questions by providing some early test cases for CCPP in the global system.
- The MET project has the potential to redefine evidence-based decision making for NOAA as well as academic communities. However, there are continued perceptions as to its usability. We recognize that the MET group has put in considerable effort to educate potential users of its capabilities and advantages. As the system has been very rapidly developed, we believe that focusing on more outreach to other non-NOAA communities will help ensure its adoption across a wider range of communities.

The rest of this document is organized by project/topic that the discussions revolved around.

Planning

There is concern throughout the SAB that since DTC is tied to a one-year planning cycle, it makes it hard for them to be effective in some areas. The NGGPS SIP is a rolling 3-year plan and DTC is named in the SIP. Also, NOAA grants have a duration of more than 1 year.

Recommendations

- *We recommend that DTC sponsors allow certain aspects have a multi-year plan. Plans can be changed, but there should be a reason. I.e., plan with sponsors for multi-year projects. They can be changed every year, but the sponsor should start with the previous plan and justify changes.*
- *Sustain core funded activities (code management, CCPP, MET) that are not project based, but that support all projects.*
- *Get backing for planning for MET might be helpful.*
- *DTC needs to work with their funding agencies to define what their role is, and who is the community that they support now and in the future*

Community Common Physics Package (CCPP)

The SAB agrees that the CCPP has the potential to accelerate the implementation of testing of new physics or existing physics in different models, but this framework needs continued support to provide community support/education on what it is, and how it works.

Benefits of CCPP

- Initial heavy lift of porting a parameterization to CCPP opens up that parameterization to be tested in a wide range of configurations.
- CCPP can benefit the visitor program by lowering the barrier to entry, and visitors can do some testing and evaluation themselves.

Concerns

- The balance between immediate sponsor needs and actual quality design is a tension. CCPP efforts seem to be too reactive. DTC should ask NOAA for more breathing room to do a quality design.
- The current design still has questions. A low risk prototyping strategy could be developed. NCAR is probably ahead of NOAA in thinking through some of the more complex issues (e.g. chemistry) with the CCPP. The community is looking for clarity and a stable interface. Work with NCAR also to clearly define CCPP, and with NOAA to clearly define timelines.
- Physics suites must be brought in together.

Recommendations

- *More support is needed so that CCPP is available quickly in the Global Modeling Testbed.*
- *EMC also needs support during the transition from IPD to CCPP.*
- *Success stories of CCPP from outside users would solidify the utility of CCPP as a community asset. This could include NCAR. Test cases (sample parameterizations and use cases) to determine limitations of architecture.*
- *Need to continue to listen EMC's concerns: adding diagnostics, issues such as PBL is closely tied to surface physics.*
- *See if the DTC can agree on a realistic timeline with NOAA that allows DTC time to do quality design work, while still doing 'prototyping' (not operational). This could be based on an independent review of the CCPP and its status and cost/time to completion.*
- *CCPP needs continued funding support for operational implementation and community support.*
- *Facilitate a process so that the NCAR and NOAA versions of the CCPP do not diverge.*
- *The CCPP and physics under the CCPP need to have an authoritative code repository. There are community efforts to port code into the CCPP, a repository for such code should be made available soon.*
- *Suggest an independent review of its current state in light of the original design goals with the aim of determining ongoing costs and whether it can be fixed to achieve originally stated goals*

Model Evaluation Toolkit (MET)

MET provides a comprehensive verification capability that has great value across the NWP community. NCEP/EMC, along with other centers, trust MET as a verification tool.

There seem to be many strong opinions on MET - often contradictory. There has been tremendous progress on MET, both as a software system and as a verification tool. Its adoption by EMC as its verification system should help improve its standing within the operational community. The challenge will be to figure out how to promote its use in the academic community.

Recommendations

- *Survey Users to determine what metrics are most important and the pain points of MET usage*
- *Continue to streamline the source code, required packages, and installation*
- *Make it clear to users that interfacing with existing databases is not trivial, and, often requires development*
- *Work with data assimilation development groups to share knowledge and code about interfacing with newer models and observations sources*
- *Develop a baseline standard set of metrics for forecast evaluation, and, drive a path to adoption through documentation, training and simple tools*
- *Redevelop MET to be more modular, e.g. what GSD is doing with MET express.*
- *Need clarity of what metrics are more important for forecast evaluation.*
- *Attempt to remove legacy code*

Concerns

Comments below are culled from individuals on the board.

Individual comment 1:

There are concerns that MET can be difficult to install. The need for several 3rd party libraries (BUFR, GSL, g2clib, etc., often with compiled with different compilers) can make installation tricky and portability difficult. This is not unique to MET, as many other meteorological packages have similar dependencies and can be hard to install from source code. A very informal survey of MET users by one of the board members, found that the Docker container for MET has been extremely helpful in solving the installation issues for general users. Further, even outside of containers, installation is much simpler now than it was a couple of years ago, perhaps due to refined source code and lessons learned that have been incorporated into documentation. So, with that, MET is certainly moving forward, though DTC should continue to streamline the code (perhaps continuing to minimize prerequisite libraries) in order to further simplify the usability and encourage further adoption.

Individual comment 2:

MET has become a large Swiss Army Knife that does everything. It has so many statistics, it is not clear how to focus on higher priority statistics. MET is straining the usability limit for many people (see modularity below). In addition, MET has a problem querying observation database and it has a lot of legacy code. This makes for a technical limit at NCEP. Testing MET on different platforms is difficult (hard to port due to legacy code). Sponsors are pulling MET in different directions and contributing to these problems.

General:

There is a sentiment amongst the board that interfacing MET to existing databases, can be challenging. This was reported as an issue with observational databases at NCEP, and other board members mentioned similar experiences in their home institutions. With varied databases at different institutions, it is understood that it would be very difficult for MET to support all interfaces. As such, for users looking to interface with their own databases and generate simple statistics (e.g., site-based MAE, RMSE), MET may not be the appropriate tool, as it may be simpler for users to calculate their own statistics. However, the value of MET becomes clear when one looks to build a verification system that will have the capability of generating verifications that are comprehensive and sometimes sophisticated. For example, MET shines when users are looking to build gridded verifications (e.g., Fractional Skill Score, object-based verification, etc.). As such, it should be made clear on the MET web pages that the value of MET is a tool for comprehensive and sometimes sophisticated verifications, though MET may not be the best tool to replace existing simple database-driven verifications.

With the wealth of verification metrics that MET can provide, it can be difficult to understand how to make best use of the system for a given application. The MET documentation and tutorial pages are comprehensive and thorough, and with this information, one can make full use of MET. However, what is often not clear is how one would generate meaningful verification for their application. While MET can be used for many applications, the focus of MET usage has generally been for operational weather forecasting. Walking a user through the generation of a verification that has been used before to verify operational forecasts would be very helpful in order to teach a user both how to use MET and how to verify forecasts. Further, this could lead to a set of baseline statistics that could become a standard verification. Ideally that standard could be used to enable verification comparisons between forecasts from different NWP organizations. The scorecards that DTC has presented at many conferences may provide a basis for this. Perhaps having a set of 'standard' statistics that drive a standard scorecard would be very helpful here.

As an example, a FAQ entry might be:

How can I verify operational weather forecasts using MET?

This could lead to a description of how DTC used MET to verify AFWA's forecast system and then show a user how to take their own data to compare against archived NCEP GFS forecast data in order to generate a scorecard.

As NWP models (e.g., FV3, MPAS) are moving towards irregular and unstructured grids, MET will need to adopt to a framework that can support spatial verification on native model coordinate systems. Additionally, data assimilation systems face a similar challenge of migrating to new native model coordinates, and there may be opportunities to share code and knowledge here.

Community support

DTC should clearly define its 'community'. The SAB identifies the DTC 'Community' as those outside of NOAA that are interested in improving NOAA Forecast models. Are NOAA research staff outside the operational centers considered party of the community?

There are several challenges here. First is identifying community beyond those that NOAA is directly funding (NOAA could help with a list of those funded). Another challenge is that it is hard to support the community when there is no HPC platform that can be identified. Although this is a NOAA issue, it restricts the amount of support DTC can offer the community. NOAA does appear to have time on a university supercomputer that could serve as the HPC platform.

Recommendations

- *CCPP: Community support aspect. Help educate developers.*
- *SAB recommends that supporting the FV3 based global model should be a priority. The stand-alone regional convective allowing model (SAR/CAM) is still not well-defined, and the SAB feel that DTC should not be forced to provide community support for this yet until the sponsor actually has a model*
- *The SAB recommends that if/when DTC begins supporting a new system (e.g. JEDI) that they stop supporting older systems (e.g. GSI).*

Testing and Evaluation

Some projects may span multiple years, lay groundwork for that

Could follow up with PIs of NOAA projects to offer DTC services directly

Recommendations

- *Seek longer term projects that span multiple years, rather than current system of testing X-number of improvements in model Y for current year.*
- *DTC create tiger teams for global models, with a focus of improving an identified bias rather than one off T&E. Combine process-oriented problems for T&E.*

Visiting scientists program

The visitors program is working well in getting people from the academic community but less well on the operational side.

Recommendation

- *DTC should seek visitors from operational centers to work with visiting scientists at DTC.*

Hierarchical Testing Framework (HTF)

The SAB agree that the HTF is a good paradigm for model development. The questions centered around how the DTC should be involved in the framework

Recommendations

- *Let individual developers create their own parameterization simulator, and DTC gets involved at the Single Column Model (SCM) tests phase, which they are well suited to do with CCPP.*
- *Create a framework and workflow to test innovations in models with baseline, and to test high impact cases/events.*

Also recommend a set of case studies for FV3GFS.

Website

The first recommendation from the SAB is to refine information on the DTC website to further that will benefit from a mission statement and publicize it on the DTC website. The new NCAR website (<https://ncar.ucar.edu/>) has a great front page with tabs: Who we are, What we offer, and Where we Focus., which could serve as a model of defining what DTC does.

The DTC webpage (<https://dtcenter.org>) has a statement on the frontpage that states:

“DTC is a distributed facility where the NWP community can test and evaluate new models and techniques for use in research and operations.”

Further, the DTC webpage has additional information of purpose under the “About” (<https://dtcenter.org/about>).

While the general statements are very informative, what is not clear from descriptions is what users and organizations are the target audience of DTC. On the operations side, the research-to-operations (R2O) has primarily flowed toward operational US centers (NCEP, AFWA). There is a question as to whether other national and/or commercial sector organizations should be encouraged to fund DTC to enable R2O. If this is indeed encouraged, it should be made clear on the webpage.

Besides clearly defining DTC’s mission on the webpage, we feel there could be other improvements. For example, the GMTB page has a list of 3 projects. To an outsider, it looks like GMTB testing is only concerned with the performance of the Grell-Frietas convection scheme and the CICE5 model. Are these just examples, or all that GMTB does?