SAB recommendations from September 2016 meeting

***General**: Compiled and synthesized by Russ Schumacher, 2016 SAB meeting chair*

1) Continue the work toward putting DTC-supported codes in containers for community use
Many of the challenges associated with setting up and running are associated with compilation on different platforms, building the appropriate libraries that the codes depend on, and so forth. We appreciate that DTC has taken the initiative to put supported codes into Docker containers, which can significantly reduce these challenges, and we recommend that these efforts continue.

2) Streamlining the O2R and R2O process, which includes streamlining the code repositories for community development and use, etc.
Related to the previous recommendation that is specific to containers, we recognize that there are also more general ways in which these processes can be streamlined. DTC can be a leader in lowering the bar for external users to be able to contribute improvements to community codes, but this requires several things. In particular, to test potentially operational codes, users need access to operational (or operational-equivalent) machines, and the complexity of the process to commit code changes needs to be reduced.

3) Sustain continuity in T&E effort from one year to next
We have noticed that for some of the tasks, the testing and evaluation (T&E) activities appear to shift substantially from year to year (either through rapid changes in focus, or through ramping-up or ramping-down of activity). We encourage the DTC to have more year-to-year continuity in these efforts. This allows the staff to have more spin-up time on the model or code they are testing and to gain in-depth knowledge of that system, which in turn should increase productivity overall, and yield more fruitful outcomes of the T&E work.

4) DTC needs to continue thoughtfully balancing T&E activities and user support.
In particular, this balance may need to be substantially different between the different tasks. For example, the data assimilation group suggests that greater attention is required in user support for the GSI and related DTC-supported codes. On the other hand, other tasks may require greater resources for T&E. In other words, this balance should be carefully considered on a task-by-task basis.

5) We propose that DTC embrace as an overall metric of success: to increase R2O activity and productivity.
This metric is at least somewhat measureable, and reflects the core goal that DTC is designed to achieve. Of course, to achieve this goal, robust O2R (operations-to-research) activity is also required.

6) Recommendation regarding the transition to supporting NEMS and FV3.
At present, NEMS is unfamiliar to the academic community, but with the transition to FV3 eventually the academic community will need to be able to work with NEMS. The

community will need DTC's leadership in supporting FV3 and NEMS. We're not convinced that FV3 will be ready for widespread community use in 2017, so we hesitate to recommend DTC spend significant resources in supporting FV3 (to the larger community) immediately, but DTC should prepare to provide this support in future years. This involves building up internal DTC expertise in advance of these transitions.

7) Expanding the DTC interactive research community
[A detailed version of this recommendation can be found under the hurricanes working group, but we believe it is a good thing to consider across all of DTC.]

8) Graduate student activity in DTC community support
[A detailed version of this recommendation can be found under the hurricanes working group, but we believe it is a good thing to consider across all of DTC.]

9) Verification is central to the work of all of the other DTC task areas, and connections between the verification task and other tasks could be strengthened.
All of the DTC task areas use verification as a key tool in their work, and they require robust verification tools to achieve their objectives. Although there are clear collaborations that already exist between the verification task group and the other task groups, we recommend that these collaborations be strengthened and supported. This may involve increasing the flexibility of MET (and related tools such as METviewer) to support needs of both DTC task and the community. Furthermore, there should be clear mechanisms for those who develop new verification tools or methods to have those tools incorporated into MET.

10) Clarify the roles of model/code developers, DTC, and the user community, in accessing, supporting, adding new innovations, etc.
For different models and codes, there are different procedures for users to access the code, to obtain user support, and to make their own contributions to the code. For example, some tasks fall on the model developers, some on centralized staff, some on the user community, and so forth, but the distribution of these tasks can vary substantially for different codes, which can be confusing. We don't have a specific recommendation about how these roles should be distributed, but we do think that DTC can be a leader in clarifying these roles and reducing possible confusion and redundancy.

11) Keep close attention on where future DTC funding might be anticipated, and fill openings on the SAB with expertise in these areas.
This is, of course, already done by the DTC leadership, and we strongly encourage it to continue, especially as the NGGPS paradigm emerges. For example, if coupled models (e.g., ocean/atmosphere, or sea ice/atmosphere) are becoming more important to the NGGPS development, it would be beneficial to have someone with expertise in this area on the SAB in the future when current members end their terms. (i.e., we aren't recommending an expansion of the SAB, but that these areas be considered when openings arise).

12) Not lose sight of the unique function that DTC serves: to objectively evaluate model performance in a variety of applications and variables.
As outlined above, DTC serves the community in important ways by maintaining and supporting community codes and developing new tools. However, there is no other organization in the US that has, as a core responsibility, the mission to be an unbiased evaluator of NWP models, and some SAB members want to ensure that this is not lost amid other activities.

Regional Ensemble Working Group: *Rob Fovell (lead), Russ Schumacher, David Gochis, David Vollmer*

(1) **We recommend that ensemble (statistical) post processing should be a priority, especially in the transition to convection-allowing models**.
This was recommended last year, was not funded, and should be considered again, as the community still needs it. The precise role that DTC should play in this effort isn't completely clear, but DTC can play a lead role through the visitor program and connections with MDL (NOAA's Meteorological Development Lab) that need to be established. In this way, community needs and ideas can be brought to MDL. Can methods for deriving CAM-based ensemble probabilities be standardized (neighborhoods, probability matching, etc.)? Perhaps DTC should organize another workshop specifically on this issue? Engaging EMC, NSSL, and other potential partners in this effort is encouraged.

(2) **We recommend involvement in the CLUE (Community Leveraged Unified Ensemble) collaboration, an HWT initiative**.
This initiative coordinates forecasts from different groups, to permit direct comparisons. The SAB previously recommended leveraging ensembles being produced by others, to mitigate limited resources. That particular effort was not funded, but the CLUE initiative is emerging as an opportunity, as it is developing a track record and visibility. Accordingly, we recommend getting engaged in CLUE, using output in post-analysis mode (as they don't have enough "eyes on the outputs"), help design standardized MET configurations so statistics can be easily compared, and be involved in design of next year's ensemble. Also, DTC should attract more visitor projects relating to this work.

(3) **Continue and enhance current efforts relating to stochastic physics for both DA and forecasting**.
This is needed to provide an alternative to mixed physics ensembles, which are difficult to maintain in an operational environment, and to help EMC in its move towards the unified modeling system, and it is a cutting edge scientific opportunity. Additionally, in the short term, this effort will more effectively inform 3-km HRRR ensemble design and (importantly) insight gained will be transferrable to other models and systems. Along those lines, we also suggest initiating an early involvement with FV3 to facilitate ensemble design for the new system. Specifically, collaborations with the data assimilation (DA) team are encouraged.

(4) **Encourage further research on the role of land surface states and**

parameter uncertainty in ensemble prediction systems, particularly with regards to convective forecasting.
The motivations are that the land surface state and fluxes have not been handled well in models, and are not well-constrained by observations. These have important but not fully explored influences, such as on warm season convection, surface hydrological responses, surface air temperature, snow cover, etc.. The opportunity is develop an uncertainty quantification framework for ensemble construction and to explore role of different initializations, e.g., NLDAS, National Water Model analyses, etc.. This could be dealt with using stochastic perturbations. The goal is to enhance the characterization of uncertainly owing to surface and subsurface processes.

(5) **Encourage engagement with global ensemble community**, especially a collaboration with GMTB, and specifically revisit recommendation from last year to explore stochastic approaches at resolutions relevant to global modeling.

(6) **Regarding MMET, we see previously raised issues regarding visibility as being addressed**.
We recognize that team is working to provide MMET cases, UPP, and MET packaged in Docker containers, to facilitate adoption in academic curricula. We encourage further movement towards incorporating ensemble aspects, as reflected by the merging of the mesoscale model and ensemble tasks into the regional ensemble working group. Graduate students and instructional staff are particular targets of this effort. We still think that the acronym is not optimal, particularly with regards to its present expansion beyond just models, and the confusion the name causes with the MET (Model Evaluation Tools) is unhelpful.

Hurricanes Working Group: *Jenni Evans (lead), Kristen Corbosiero, S. G. Gopalakrishnan*

1) Building Community
<u>Expanding the DTC interactive research community</u>
Build an interactive community of active users through monthly online research events: informal themed discussion with one user taking the lead to begin the discussion. The goal is to hear from active users with fresh ideas. These could be exciting new incremental results, problems they've solved, or sophisticated problems [not something that can be solved through the help desk or straightforward consultation] they want to bring to the community. There can be a disconnect between research and operations, so this provides a great link. Since DTC is meant to be looking 3 years out, this provides a conduit between research now and possible operations in the future via DTC assessment of feasibility.

This would be a good forum for interacting with active scientists in other areas/models. This is also a promising mechanism for getting users involved in the DTC visitors program.

In the research community, this group may be dominated by graduate students and postdocs. In this way, it could become a defacto young scientists in modeling mentoring group. This group could also request presentations from senior scientists now and then.

Graduate student activity in DTC community support

In terms of DTC concerns with having adequate staffing for support of WRF (or other DTC-supported code), we suggest a graduate student model: identify senior graduate students whose expertise as at least a first-line consultant, DTC can evaluate. For students who pass the evaluation, DTC pays their stipend and tuition and the students work in the email support system, only referring major questions to DTC staff. Questions dealt with should be geared towards operational model/problems relevant to the DTC; other questions should be referred to the WRF user support or other relevant support communities.

In addition to addressing the DTC staffing problem, this builds/enhances capabilities of upcoming researchers relevant to the needs of operational models. It would benefit the university community by supporting students with a gap in funding (a win-win situation for both DTC and the university.)

2) Hurricane Modelling

DTC support during evolution of the model dynamical cores

The WRF -> HNMMB -> FV3 transition is going to be sticky. Since the NWS have decided to go with the FV3 dynamical core and are pushing to implement it as soon as their processes allow, we suggest that the DTC should provide academic/industry users with a roadmap between WRF/HWRF and FV3. This probably means not engaging them in development with the HNMMB model. However, we acknowledge that NHC is extremely careful when adopting new models (by in large, rightly so); this means that operational use of the FV3 core for hurricanes is likely farther off in time than we are being led to believe. Does this change our recommendation to the DTC?

Huge overhead to switch to HNMMB [replaces GFDL] based on experience with NMMB. Don't want the community to be restricted to HNMMB infrastructure. Should DTC be supporting HNMMB? No, because funds for that support are not likely from HFIP.

A roadmap would be a good start: DTC helps guide academic researchers in efforts that would ultimately contribute to the FV3.

Since hurricane modeling is a global problem, so make very high resolution global model available to students. DTC is already working on physics-related core for the FV3 global model.

At present, NEMS is unfamiliar to the academic community, but with the transition to FV3 eventually the academic community will need to be able to work with NEMS. The community will need DTC's leadership in supporting FV3 and NEMS. We're not convinced that FV3 will be ready for widespread community use in 2017, so we hesitate to recommend DTC spend significant resources in supporting FV3 (to the larger community) immediately, but DTC should prepare to provide this support in future years. This involves building up internal DTC expertise in advance of these transitions.

3) Science questions relevant to the DTC mission

Scale-aware physics. How can this be tested and how can test models be made available to the user community? Don't wait for "perfect" model: make models available so that academic community can start applying it to case studies and providing feedback on relative skill in different atmospheric flow structures/phenomena.

Common community physics package in DTC global task. DTC should work to move the hurricane model physics into the global model framework.

Hurricane-relevant physics. Test global model for multiple hurricanes in a basin.

Verification and Evaluation. So much hurricane verification is based on track and intensity, but DTC should expand verification to a variety of structure measures, such as cloud structures (cloud bands, asymmetries), and much more. Can also leverage existing tools in MET by collecting them in METTC, as well as the track and intensity tools available now. This only need be a website that points to the relevant MET tools, so should not add a lot of overhead to DTC staff. Community users can also suggest MET tools that could be added to METTC. Tools developed during DTC visitor projects need to be incorporated into the overall code package.

One problem relevant to hurricanes now is rainfall distribution. This is all relevant to the operational forecasts in regional NWS offices and *would be of strong interest in HFIP*. Utilize rainfall relocation to evaluate rainfall against analyses even with track errors; incorporate the same methodology using satellite/radar data as the analysis; incorporate rainfall shape evaluation (object-based verification) to capture skill at simulating rainband structures.

Verification Working Group: *Adam Clark (lead), Geoff Dimego, Kelly Mahoney, Kathy Gilbert*

1) Continue to popularize and support verification in general and MET in particular through community outreach, such as regional workshops and AMS short courses and developing super users at key organizations.
There are still many in the research and forecasting community that are not familiar with MET, and/or are not aware of all the new capabilities built into the new version of MET (e.g., MODE time domain), so continuing outreach efforts should be a priority. Also, the concept of "super users" was discussed – these are users that have developed enough expertise with MET that they can train others at their organizations, act as advocates for MET, and serve as a first layer of support when questions/problems arise. For DTC, the super users can serve two important purposes: (1) outreach, and (2) lessening the load on MET help staff. Examples of places where super users could be targeted are national research laboratories like NSSL and ESRL/GSD, national forecasting centers like SPC, WPC, and AWC, and academic institutions with strong meteorology programs doing verification work (e.g., U. of Oklahoma, Purdue, U. of Washington, Colorado State, etc.).

2) Continue to develop a set of best practices for using MET, especially including visualization and reporting of verification statistics using METViewer and an effort to build a user-friendly, multi-platform MET+.
We strongly support the development and concept of MET+. Providing a unified and comprehensive verification tool will make R2O more efficient. Also, there appears to be a strong need and interest in visualization, thus, integrating METViewer capabilities into the official MET release should be a priority. This could involve developing something with some functionality of METViewer, but not requiring all the underlying software packages.

3) Continue to invest a portion of NGGPS funding to support multi-scale applications of MET and "process-oriented" approaches to verification.
The consensus from the meeting was that the definition of "process oriented" was not clear, thus, some energy should be devoted to clarification via outreach with partners and/or funding agencies. Also, the wording "multi-scale applications" was chosen over "global applications" given that the vision for a unified operational modeling system may start with global scale through NGGPS, but will eventually have to consider multi-scale applications.

4) Develop connections to testbed (e.g., HMT, HWT, AWT, DTB) and national centers. CLUE experiment with HWT is an obvious opportunity.
Many of the testbeds are conducting experiments in which collaboration with DTC through either real-time or post-experiment analysis/verification would be mutually beneficial. One very obvious opportunity is collaboration with the NOAA/HWT's annual Spring Forecasting Experiment (SFE). During 2016, a new paradigm for the ensemble guidance used in the experiment was introduced known as the Community Leveraged Unified Ensemble (CLUE). The CLUE involved an unprecedented effort to leverage several academic and government research institutions to help guide NOAA's near-term operational environmental modeling efforts. Specifically, during the 2016 SFE, convection-allowing ensemble configurations were coordinated much more closely than in previous years, with all groups agreeing on a set of model specifications so that simulations contributed by each group could be used in carefully designed controlled experiments geared toward identifying optimal configuration strategies for the first generation of operational convection-allowing ensembles. The CLUE contained 65 members designed for 8 unique, controlled experiments. DTC's role in the CLUE could include post-experiment verification, along with collaborative design of future CLUE experiments based on results from past experiments.

5) Incorporate recommendations from Ensemble Design Workshop into MET and METViewer.
This August 29-30 Workshop held at EMC focused on convective scale verification. Recommendations from the workshop will focus around best practices for convective scale verification (e.g., metrics and scorecards) and ensemble configuration strategies. DTC already has staff that attended and is connected to the workshop, so they should be sure to follow-up on the final set of recommendations and work to incorporate anything that is relevant into MET and METViewer.

6) Continue to make ensemble verification a high priority.
This is obvious given the strong push towards probabilistic forecast guidance across the weather community.

7) Engage the convection-allowing model community: What are the unique verification limitations for using MET to verify CAMs?
Application of MET to high-resolution models is a unique challenge simply because of the large amounts of data. As one example, simply not being able to directly read wrfout

files is a serious impediment to many researchers working at convective scales. In fact, there are some (many?) that won't even consider using MET unless it reads directly from wrfout files because that is how they do their own research using netcdf-friendly packages like python. So, it would be helpful to at least build in a capability to read 2D fields. Finally, there simply are not many researchers that have used MET for convection-allowing models. If a "super-user" could be identified for CAM applications, that could be very helpful for identifying limitations and additional applications.

8) Partnering with DA community in terms of observation-based verification (e.g., directly simulating observations with model output, etc.)

9) Consider reach out to other communities that have their own verification systems (or would benefit from NWP-based verification techniques), such as high-resolution climate models, land surface, marine, and atmospheric chemistry.

Global Model Test Bed Working Group: *Tim Whitcomb (lead), Zhuo Wang, Brad Colman*

1) DTC should focus on developing and maintaining the technical and social infrastructure for effective community engagement, testing, and transition.
If NOAA is to move toward community modeling, this requires maintenance of both the technical infrastructure (common datasets, evaluation tools, predefined workflows, scripts, etc.) for the various components in the community modeling system (e.g. the GMTB). Another part of doing community modeling isn't just code release – it's ways for developers and users in the public, academic, and private sectors to contact NOAA and contact each other for support, questions, and to pave the way to transition new modifications from the community.

We enumerated a sample of properties of a "good" community model, including (but not limited to): user-friendly documentation, responsive support, standard datasets with well-documented results, code designed for extensibility and portability, and sponsored workshops. We feel that DTC's expertise projects strongly onto many of these areas and DTC can effectively serve as the interface between NWS operations and the wider community. In addition, DTC can continue development of the testing hierarchy and support materials (such as documented baseline results).

2) DTC should work with EMC and others within NWS to clearly define transition paths for community contributions and rules for contribution.
Community modeling is not just taking code developed within an organization and making it available to the outside world. There must be a clearly defined procedure that describes what a developer must do, who they must contact, and the coding standard that they need to follow in order to take their idea and see it through to running in operations.

3) DTC should begin testing FV3 at high resolution with a focus on informing NWS during their migration to a single dynamical core for unified modeling, in collaboration with EMC and for future DTC support decisions.

While the decision has been made that FV3 will be the dynamical core for the next NWS operational atmospheric global forecast model, there are still several dynamical cores used for limited-area modeling within EMC. While there is desire to consolidate dynamical cores, that will be driven by the performance of the FV3 core at high resolution compared to the existing cores – these tests, performed by DTC in collaboration with EMC, should help EMC decide how to consolidate dynamical cores and which cores should be supported by DTC.

4) DTC should work with MEG to identify prioritized areas for collaboration and community engagement, with bundled tools and appropriate datasets.
One downside of a community modeling approach is that there is sometimes little control over the directions taken by the community for the development. This can be fruitful, but we feel that DTC can better engage community resources by working with EMC's Model Evaluation Group to identify significant model shortcomings (along with all the data required to reproduce and demonstrate that the problem has been addressed) as challenges to the community interested in improving NWS operational forecasts. A similar effort (MMET) was not perhaps as successful as was originally hoped, so it may be useful to identify why engagement in that project was not as strong as expected.

5) We recommend that DTC, not EMC, be the primary location for the interoperability layer (e.g., for physics) and infrastructure to facilitate community engagement.
The original version of the NUOPC physics interoperability layer was written by EMC, and some development has now moved to DTC as part of the Global Model Testbed. We feel that EMC is too close to operations, and that an interoperability layer designed to allow for effective community engagement should reside within an organization focused much more on community engagement (DTC) than direct operational support with less of a public interface (EMC).

Comments (not specific recommendations):
We like the multi-tiered community approach raised by EMC management in his presentation. There are three tiers: operational forecasters, super users (or core partners), and general users. DTC can provide different services for different levels of users.

For the operational centers, model verification is one of the things DTC can offer. It is important to collaborate with EMC and the NGGPS model developers at the very early stage of NGGPS implementation. In the past, there was some resistance to use the DTC MET products because EMC already had something in house. DTC can work with EMC and make the MET the unified verification package for NGGPS.

For the super users, DTC can take the lead to develop the infrastructure for testing and evaluation. Testing can start with something simple, like the single column model, and then move to a more advance level. The key is to have clearly defined testing standards along with carefully chosen metrics. In addition to performance-based metrics, physics-based metrics should also be developed. The design of the infrastructure requires input from forecasters and developers. Such an infrastructure will be critical for the R2O transition.

For the general users, DTC can help to provide clear documentation of the FV3, user's guide, and organize some users workshop. When distributing the model, it may be a good idea to bundle the MET with the model as a post processing and verification package. This can help potential R2O transition from the general user, and can also help to increase the visibility of DTC.

DTC needs to work with partners to develop clearer guidance and definition on governance and transition procedures. This community includes the private sector in addition to universities and government.

For the general users, DTC can help to provide clear documentation of the FV3, user's guide, and organize some users workshop. When distributing the model, it may be a good idea to bundle the MET with the model as a post processing and verification package. This can help potential R2O transition from the general user, and can also help to increase the visibility of DTC.

DTC needs to work with partners to develop clearer guidance and definition on governance and transition procedures. This community includes the private sector in addition to universities and government.

Data Assimilation Working Group: *Kayo Ide (lead), Sharanya Majumdar, Tom Auligne*

1) Streamline O2R and R2O activities; define community and metrics of success

(a) The continuation of efforts to bridge the "valley of death" between research and operations (both O2R and R2O) is recommended. In order to streamline efforts, the areas where DTC will provide added value (as opposed to single projects that may end up being dead ends) need to be prioritized with metrics of success defined. It is expected that these clearly defined activities will increase productivity in both R2O and O2R.

(b) The community of R2O developers needs to be defined. While this largely involves EMC/NESDIS/DTC/OAR/NASA/JCSDA, it will also need to include partners from other organizations including the commercial community.

(c) There is presently overlap and duplication across multiple groups. We recommend that the DTC works with the community of R2O developers to streamline code repositories for community development and use. This includes compilers, software and libraries. To accomplish this, developers need access to the appropriate computing platforms. The preparation of a common suite of canned regression tests would be helpful. Specific tasks can be identified for individual developers that also map on to their own priorities as PIs and within their agencies. Interactions can be accelerated via the DTC Visitor Program. One example is to collaborate with JCSDA on the development of JEDI, including community support. Ultimately, though these efforts, a streamlined repository is expected to yield greater involvement and support from the user community.

2) Establish a balance between Testing & Evaluation and User Support

(a) Continuity in Testing & Evaluation efforts needs to be sustained each year. This involves allowing time for these efforts to spin up, which will ultimately yield greater productivity and an in-depth knowledge of the system. The connection between DA and verification needs to be maintained. Specific examples of ongoing efforts that need to be maintained include
    - Verification of ensemble perturbation methods
    - Inflation: stochastic parameterization approach
    - Working with verification and ensemble groups.

(b) The User Support effort needs to be streamlined.

For each of these recommendations, we recommend continuing to work on current issues, and also establishing a dialog with EMC on the NGGPS with FV3, with priority given to

- Community support and a user guide tutorial
- Portability and a code repository
- High-performance computing access for the community.