The primary agenda topic for the meeting was a discussion led by Tara Jensen (NCAR/RAL/JNT and Strategic Implementation Plan (SIP) Verification and Validation Working Group (WG) Co-Chair) regarding the upcoming 30 July - 1 August 2018 Unified Forecast System (UFS) Test Plan and Metrics Workshop and related topics. A general outline of the discussion is available at:

https://docs.google.com/document/d/1FMljNc21r7-7T8Ft0nX19YZgQMksmHxhPDNIowLw4UE/edit

Prior to the workshop discussion, there were several administrative remarks. It was mentioned that approximately 3-4 hours are expected to be allotted for UFS-SC matters during the upcoming 1-3 August 2018 SIP Coordination Meeting. Additionally, next week’s UFS-SC meeting is scheduled to be a synthesis meeting with the SIP WG Co-Chairs. A query will be sent to the co-chairs requesting information on the status of their activities and whether they have input on candidate issues for discussion with the UFS-SC.
It was noted that following the recent briefing/discussion on CCPP governance discussion with the UFS-SC, Ligia Bernardet had advanced the governance proposal to NOAA NWS OSTI. Based on input from OSTI, Ligia is coordinating discussions with partners, including NCAR, Navy NRL, and NASA, to solicit further input and potential collaboration on the proposed governance.

Tara Jensen initiated the discussion on the upcoming Test Plan and Metrics Workshop and noted that a “save the date” announcement for the workshop had been transmitted. She also indicated that workshop planning is being actively coordinated with external partners. With the workshop ending on Wednesday morning and the SIP Coordination Meeting beginning later that day, there is a hope that some of the early arriving SIP meeting attendees will be able to participate in the final sessions (report-out, etc) of the Test Plan and Metrics Workshop.

There was discussion on the definition of metrics for the UFS and whether the definition applies to different temporal/spatial modeling scales. It was noted that the Convective Allowing Modeling WG already has a list of metrics and the Verification and Validation WG is working with other SIP WG’s to advance their respective metrics preparations prior to the upcoming workshop. A list of WG liaisons is being compiled to facilitate this effort (draft list included in material at the link provided earlier in these minutes).

Interest was stated on learning more about the concrete definition and expectations for the “test plan” (scope, for example). MET (Model Evaluation Tools) was recognized to be a primary emphasis in this. There was also interest in learning more about the relationship between code verification and product validation. These issues are expected to be addressed in coming weeks, including coverage at some level during the Test Plan and Metrics Workshop.

The Verification and Validation WG expressed interest in obtaining guidance from the UFS-SC regarding what other SIP WG’s they should be coordinating with on metrics (influence from the SC is also desired to encourage collaboration from these WG’s).

Regarding the scope of the upcoming Test Plan and Metrics Workshop, the intent is to move the development of the test plan as far forward as possible. There is a desire to broaden the scope of the test plan beyond NWP for weather prediction (to consider other earth system models and ensembles, etc - at least start the conversation). The Verification and Validation WG plans to broaden and strengthen the WG membership by adding co-chairs from universities. A couple of names have been identified and, once confirmed, the SIP WG roster will be updated appropriately.

The focus of the current verification and validation effort was discussed. It was stated that the current effort is not focused on applications of verification/validation for software optimization (such as when model infrastructure changes are made). Current efforts and the workshop are focused more on verification/validation to be applied when parameterization changes are made to a model, for example. Two types of testing were noted: regression testing on
coding/software and testing for model operational improvement. Both are needed, but a comment was made that the upcoming workshop should focus on defining success up front with a goal of improving operations. The need was noted, however, to produce a written document covering the entire landscape of testing. The example of producing a coupled system was noted. In this case, there will be a need initially to ensure the coupled system is working correctly (more from a software perspective). Once working correctly, verification/validation of the scientific performance of the coupling will be needed. There is a need to capture this interface in terms of how the hand-off occurs between these types of testing/development.

A comment was made that science metrics may not be same as user based metrics and a question was asked regarding whether this will be a topic of discussion at the workshop (multiple metrics for different definitions of success). It was noted that a need exists to define what metrics are needed for different applications (big gap exists in current understanding). These issues will be captured for discussion with the liaisons identified by the Verification and Validation WG in preparing the agenda for the workshop. Based on discussion, additional liaisons may be added for this discussion. Suggestions on additional liaisons were offered such as field forecast office representatives and perhaps private sector representatives (at least polled).

In response to a query, it was stated that aspects of subseasonal to seasonal (S2S) activities will be considered in the workshop. Break-out sessions at the workshop may be based on temporal/spatial scales. It was mentioned that Climate Prediction Center representatives should be invited to the workshop given their expertise in the S2S area (have a relevant package that EMC uses).

A desire was stated to have MET working for all scales. It was noted that current plans already call for a focus on expanded applications including ensembles. The most robust current verification and validation capabilities were believed to be focused on global forecasting. There are limitations at other scales (reason for a separate ongoing Convective Allowing Modeling scale effort). A comment was made that it is imperative to have some type of agreed to verification and validation metrics across all scales and to have a way to adjudicate these (example of upgrading S2S capabilities which degrade medium scale applications). It was noted that through the MET Plus effort, work is being done on a scorecard to develop metrics (may need scorecards for different disciplines with some sort of roll-up that can apply for all). A future workshop on related teleconnections might be needed and it was recognized that multiple iterations (possible other workshops) will be required to carry this overall effort forward. Workshop organizers indicated that a strawman test plan will be sent to appropriate contacts leading up the workshop.