Summary Report from Strategic Implementation Plan (SIP) Workshop

Cross-Working Group (WG) Meetings

Annexes: Individual WG Reports

NB: This file is the second of a two-part workshop summary report. For part one see file: NOAA COMMUNITY-SIP X-WG WORKSHOP Summary (Part 1-of-2).pdf

Executive Summary

The NOAA Community Modeling Workshop and meetings of the Strategic Implementation Plan (SIP) Working Groups were held from April 18-20, 2017 at the National Center for Weather and Climate Prediction in College Park, Maryland.

The first half, which ran from April 18 through noon on April 19, was the NOAA Community Modeling Workshop, and was designed to interact with the broader model R&D community. As such, this portion was completely open to the public, and included a dial-on capability for the plenary sessions, which was done in order to be as open and transparent as possible. This workshop began with opening talks to set the stage, the first a summary of the approach and goals of the Next Generation Global Prediction System (NGGPS), and the second a summary of the SIP and its goals and objectives. This was followed by a panel discussion of senior leaders from the weather enterprise, including the Directors of NWS and NOAA Research, UCAR President, and senior leaders from academia, private sector, NASA, National Science Foundation, and DoD (Navy). Each were asked to provide their perspective on three items:

1. What aspects of a NOAA-led community-based next-generation unified modeling system would your organization and sector find advantageous (i.e., how they would benefit)?

2. For which parts of a community unified modeling effort would your org/sector be best able (and most likely) to contribute (i.e., what do you feel is the best role for your org/sector to play)?

3. From the perspective of your org/sector, what do you see as the greatest challenges to be overcome (or barriers to be broken down) to make this a successful community enterprise?

The remainder of the presentations were panel discussions featuring co-chairs from 12 active SIP Working Groups, each of whom provided their perspective on the ongoing activities of their WG and the overall effort to migrate the NGGPS global model, under development within NOAA, into a community-based unified modeling system.

The workshop concluded on the morning of April 19 with a series of parallel break-out groups, each of which was asked to provide their assessment based on what they saw and heard during the presentations to identify two categories of items:

1. ***Best practices:*** What are the major things that we're getting right?
2. ***Gaps:*** What are the major things that we're missing, or heading down the wrong track?

Note: A separate document will provide the reports from the break sessions.

The second portion, which ran from the afternoon of April 19 through the end of April 20, was a series of meetings between the various SIP Working Groups, which was aimed at advancing the technical planning within each WG and ensuring that they are well-coordinated across them. These meetings between WGs, also referred to as Cross-WG meetings, were also designed to identify areas of overlap vs. gaps between the WGs, and to help facilitate technical exchange.

Each WG was asked to provide (1) an overall assessment of the effectiveness of the workshop, (***Best practices:*** What are the major things that we're getting right?

3. ***Gaps:*** What are the major things that we're missing, or heading down the wrong track?

A summary of the "immediate needs" and "critical path" items are in the next sections, followed by the full reports from each WG provided in a separate Annex for each.

Immediate Needs:

Each Working Group (WG) was asked to comment on items they felt need to be addressed as soon as possible in order to inform the remaining work. Items unique to the subject of each particular WG are found in the Annexes. The following is a summary of the most commonly referenced items from all of the working groups, or items that are applicable across all the WGs:

- *Governance:*
    - *Defining governance procedures: Setting up a governance structure needs to happen ASAP for project success.*
    - *Establish governance for the overall unified modeling effort. It is important to have a group that can establish vision, prioritize the distributed development, and assist the various WGs and participating institutions in aligning their activities.*
    - *Transitional Governance needed ASAP to be able to make up-front decisions needed by other WGs in order to provide necessary guidance for them to draft initial plans.*
    - *To retain the confidence and goodwill of community participants it will be necessary for an authoritative body (e.g. steering committee) to emerge soon that can resolve conflicts and formulate and communicate integrated plans.*
    - *Introduce overall authority and coordination. There is a need for an overall authority that can resolve conflicting requirements, strategies, and goals, and develop integrated plans. This is important to the system architecture because unified modeling system design and development require communication of an overall vision and an understanding of priorities and timelines.*
    - *We need clear guidelines on the governance of the overall configuration of NWS operational applications. Perhaps a "Change Management Review Board" needs to be stood up by the Governance WG for this purpose. (Governance of the individual Earth System components should be kept separate and the communities supporting these components should maintain such primary responsibility.)*

- *Infrastructure and Architecture:*
    - *Setting up archive. Setting up a repository for code (outside NOAA firewalls), needs to happen ASAP for project success.*
    - *Access to the model and model documentation/training needs to be easy in order to enable the community to participate. In addition, funding needs to be available to allow for community participation.*
    - *Code management, CM, governance, and decision making needs to be transparent.*
    - *Need governance decision ASAP to validate consensus on standardization around code development procedures, which include type of version control system (e.g., Git), hosting service (e.g., VLab, GitHub, etc.), and guidelines for development.*
    - *Define roles and responsibilities in infrastructure and support. Understand what needs will be met with FV3's inclusion in CESM and what needs to be done by other groups (GMTB, EMC, etc.)*
    - *Workflow: Starting with the early release (planned for the end of May) the users should*

have access to a workflow that can be used with the model. This will establish good practices from the onset of the project and enable the community to focus on model development.

- *Community survey: The communication and outreach WG should work with the other working groups to help design a survey on what EMC can provide (test cases, post processing and verification tools, DA experiments etc.) and at the same time identify the major needs. This should then be broadcast to the overall research community to get a pulse of how best EMC can interact with the broader community.*
- *Coordinate system architecture definition and FV3 development. Additional interaction is needed between the System Architecture and Dynamics working groups to discuss goals, timelines, and approaches (e.g., roles and capabilities of ESMF/NUOPC and FMS) related to integrating FV3 into the unified modeling system. For example, the implementation of nesting and a standalone regional atmospheric component model employing FV3 have system architecture implications.*

- *Misc./cross-cutting items:*
    - *Community consensus emerged to proceed with a stand-alone regional FV3 model, for purposes of testing and evaluation, verification and validation, and to engage the broad R&D community that do not typically have the capability or interest for a fully-coupled global system (e.g., the WRF community).*
    - *Establish unified metrics and scorecards that address short-range convective scales to sub-seasonal and seasonal scales, as well as metrics for coupled applications. Establish testing procedures to conduct evidence-based, transparent assessment of physical parameterizations and suites.*
    - *Observations Processing: It became extremely clear that obs processing must be accounted for somewhere within the WGs. There is an increasing need for diverse data sets for verification and data assimilation activities from multiple working groups, and it seem inefficient to task the V & V or data assimilation groups with this task unless one is provided with the support to take on this effort. Perhaps an additional WG should be formed to provide subject matter expertise on observations data types and processing.*
    - *Clear Guidelines for Community Interaction: There was a resounding sentiment that clear guidelines for community interaction, including how to contribute to the community repositories is needed. Questions that need to be addressed include:*

1) How will the code be supported out to the community (O2R)?
2) Who will provide helpdesk capability (Community Support)?
3) How much testing is required prior to committing an innovation to the community repository (governance)?
4) How to lower the bar for the community to do testing (testing framework)?
5) How will computing be provided to help with development and testing (testing framework)?
6) What are the benchmarks of success, upgrading the operational model and accepting innovations into the repositories (governance)?

Critical Path items:

Each WG was asked to identify those items to be worked that are the most critical to achieving their WG goals and creating their portion of the draft SIP. Since each WG had their own perspective, there was no attempt to pull out common themes. Instead, a brief summary of the critical path items is provided below from each of the working groups.

(1) Aerosols & Atmospheric Composition

System Architecture and Post-processing: The AC-WG System would require the following capabilities to be included in the system architecture and post-processing software:

- Support for downstream modeling (mechanism, LBCs, interfaces)
- Provide for off-line coupling to chemistry & dispersion models (e.g., backtracking dispersion)

Physics: The AC-WG recommends that a small subgroup be formed to evaluate the best approach for coupling atmospheric chemistry modules with atmospheric physics (also recommended above in immediate needs section). The Atm. Comp. WG also provided a list of capabilities that a physics module would support:

- Interactions with microphysics and radiation (aerosol aware) in a modular way, allowing for aerosol models of varying complexity (e.g. bulk, modal, bin). The solution needs to account for efficiency and memory footprint.
- Incorporation of aerosol schemes that have been developed as ESMF components, this way preserving ongoing, long-term collaborations with aerosol modeling groups at NASA.

Verification and Validation: The V&V system should include at minimum surface ozone and fine particulate matter (PM2.5) observations as well as Aerosol Optical Depth (AOD) from surface and satellite instruments to meet AC-WG initial operational capability goals. Therefore, the V&V system needs to support at least WMO BUFR and NetCDF observation formats. The AC-WG also recommends that an existing or new WG be tasked to develop plans for housing key input data (obs, emissions, land surface data) for the unified model.

Data Assimilation: The AC-WG was interested in how the JEDI program would interact with this SIP process. Many issues that SIP teams will cover are also being considered as JEDI requirements. Our group emphasized the importance of the data assimilation system for constraining not only atmospheric composition species concentrations, but also the emission, which are typically based on databases that can often be several years old. Therefore, a JEDI system that supports joint state-emission estimation is of paramount importance for Atmospheric Composition.

Land/Hydro & Marine: The Land WG emphasized the importance of ensuring consistency of dust and biogenic emissions by using consistent land surface information (land use database, canopy conductance, soil moisture, surface fluxes from land model). The Land WG also recommended the formulation of terrestrial emissions on the same tiles used for the land-surface and vegetation models. Therefore, both groups recommended closer collaboration as plans and these components are developed in the unified modeling suite.

(2) Data Assimilation

- A collaboration model for JEDI with the central software repository outside NOAA is crucial for community participation and efficient collaboration in DA research.

- The success of JEDI is crucial, since it is tied to progress on key scientific issues (coupled DA, localization, etc.). In other words, JEDI is on the "critical path".
- A governance structure for JEDI that allows for rapid development is needed (current process for GSI is subject to bottlenecks, long delays in getting changes merged). Use 'agile' software practices where possible, separate high-level governance from low-level code review.
- A community-supported tangent linear and adjoint of FV3 (based on work done at GMAO) is highly desirable.
- Improved feedback between DA and physics development is desirable (by using DA diagnostics to diagnose physics errors, using JEDO/UFO to validate model in obs space).
- Issues of 'governance' remain of primary concern, not just for JEDI but for the entire unified prediction effort. This is causing a bit of fear, uncertainty and doubt among potential participants and should be resolved as quickly as possible. In other words, creating a governance policy and structure for the unified modelling effort is on the "critical path".

<u>(3) Dynamics & Nesting</u>

- Strategy for stand-alone FV3 LAM development must take into consideration global-meso unification priorities along with physics and data assimilation strategies.
- Development of moving nests for FV3 critically depend on choice of framework, feasibility in operational settings, and computational efficiency.
- 3D physics development for space weather applications might need a separate strategy than that is pursued by Physics WG
- The current data assimilation does not support nested meshes, which needs to be accounted for in the JEDI development.
- Access to the model and model documentation/training needs to be easy in order to enable the community to participate. In addition, funding needs to be available to allow for community participation.
- Code (and configuration) management, governance, and decision making process need to be transparent.

<u>(4) Ensembles</u>

- *Resources are needed for the middle of the R2O funnel*, i.e., research supporting developments that may lead to implementations in 3-5 years. It makes sense that this funding should be run through OAR, considering the recent Weather Bill.
- *Identifying resources for reanalysis/reforecast generation*. Our sense (from ensemble experts not being consulted) is that HPC procurement does not yet account for this new requirement.
- The basic business model for central computing and then widespread dissemination of data is broken. We need to figure out how to *bring computing to the data*.
- *Coming to clarity on a weekly GFS ensemble*. We need to clarify whether (i) such a system is needed to meet product requirements, and (ii) assuming so, what configuration is desirable, including the length of reanalysis/reforecast for such a system.
- *We need to define procedures and ground rules for R2O suitable to both internal-NOAA and external members*. What does it take to run the gauntlet from research to operations? See notes

below for more.
- *We need to work out procedures for eventual unification of global & regional ensembles.* These communities are largely disjoint right now, and are run out of separate WGs.

<u>(5) Governance</u>

- Governance help needed on rules and standards others will need to plan around
    - Data formats, interfaces, grids, scripting, coding standards, documentation, etc.
- Infrastructure:
    - Data availability for common tests (regression testing, science V&V, etc.)
    - Strategic decisions needed about which parts of infrastructure need to be built vs. those that can leveraged through partnering
- V&V will be an integral component for several WGs
- Comms/Outreach:
    - Common, consistent and effective messaging to community is essential

<u>(6) Infrastructure</u>

User support and training for the coupled system: While the main developers can provide user support for running the individual components of the Unified modeling system, users need support to run the coupled model. User support needs to be offered on a continuous basis. Training should be offered on a regular basis. The support and training can be provided by partnerships between NOAA testbeds and academic research centers.

Scientific advisory committee for the coupled system: While the scientific development of the individual components is guided by the developers who are also conducting testing on the individual components of the coupled system, the scientific development of the coupled system requires guidance and testing. This can be achieved by partnerships between existing testbed centers and academic research centers.

File formats: This was not an issue that the Infrastructure WG was looking at but in discussion with the Post-Processing and Verification WGs we came to realize that this is an important issue. While NetCDF has emerged as the consensus format from the research community, there are organizational and publishing requirements that dictate the need for broader discussion and consensus. The Infrastructure WG feels this should be addressed by the Post Processing and Verification WGs.

<u>(7) Land & Hydrology</u>

- An efficient model development “hierarchy” will help streamline testing various physical parameterizations, from individual components (e.g. surface drag) to fully-coupled systems (atmosphere/aerosols, ocean, land-hydrology, sea-ice, waves components), with appropriate data sets for testing and benchmarks to “pass”.

- Land DA needs will be met by other packages until GEDI is ready (e.g., SMAP). DA WG should work more closely with land/hydro WG to develop requirements and test cases.
- There should be a parallel development strategy with explicit sunsetting of "older" systems. Progress meeting stakeholder (and congressional) requirements needs to continue without stagnation while new system is being formulated and constructed.
- Unification of "climate/seasonal" and "short term NWP"--processes are the same but models are different (e.g., Noah, RUC vs. LM, CLM) Noah-MP bridges somewhat
- Unification of "land" and "hydro"--esp. Better hydrology impacts on marine and or coupled environmental prediction. Need additional management guidance on "water" in unification effort. Horizontal vs. vertical processes.

(8) Marine

- Short-term weather scale Ocean & Sea Ice based on Global HYCOM/CICE developments synchronized with US Navy efforts.
- Seasonal scale Ocean & Sea Ice based on MOM6 and SIS2 developments synchronized with GFDL efforts.
- Ocean & Sea Ice Data Assimilation based on NCODA/GODAS for providing initial Ocean and Sea Ice states. Continue exploration of ensemble-based methods (LETKF, EnKF) for both weather and seasonal scales.
- Waves model developments include coupling to other Earth System components (Atmosphere, Ocean, Sea Ice) using NEMS/NUOPC.
- Waves Data Assimilation using GSI.
- Foster involvements with community-supported modeling groups (HYCOM, MOM6, WW3, FVCOM, ADCIRC, ROMS, CICE, others?)
- Explore coupling of all Marine Modeling components using NEMS-NUOPC framework.
- Increased engagement with NOAA's Water Team, which is a cross-Line Office team (with embedded governance) developing the vision and timeline to integrate coastal and riverine observations, couple hydrologic and hydrodynamic models, engage regional stakeholders and provide better products and services to the coastal community. Notes from the recent Coastal Inundation Summit will be distributed shortly. Their work can inform the recommendations for the Marine Modeling WG (both short and long term).
- Water Quality/ Ecosystem developments - this is in the long-term plan for the NOAA Water Team, but it is resource-dependent. Again, we should engage with the Water Team to understand their timeline.
- Note: Additional critical path items for the long term (i.e., after the FY18-20 period for the SIP) were also included; see Annex 8.

(9) Mesoscale/CAM

Testing and benchmarking: The upcoming implementation of the HREFv2 at EMC will provide an operational benchmark for which other future CAM ensembles ("CAME" below) will be evaluated against. The upcoming CLUE as part of the SPC/NSSL Hazardous Weather Testbed (HWT) will provide important subjective evaluation of 10 different CAME systems and several deterministic 3-km FV3 runs. When combined with additional idealized runs and real cases, the overall strategy will be included in a FV3-CAM Test Plan.

Widespread interest to run FV3 as a standalone regional model, so development of a prototype system should be considered to be on the critical path.

From another perspective, exploring and further developing the FV3's regional-global nesting capability (such as for hurricane forecasting) should also be considered on the critical path.

(10) Physics

- *Code management and governance for FV3.* Need to identify a code management and governance system that allows the community to use and contribute to the software.
- *User and developer support.* At this time it is not clear how user and developer support for the unified model will be done. It is likely that various groups (CESM, GMTB, etc.) will play a role, and this needs to be clarified.
- *Review requirements for the physics-dynamics interface.* In recent discussions, additional requirements for the physics-dynamics interface have been suggested, which may require development and layers beyond the current Interoperable Physics Driver. It is important to review, prioritize, and set forth a community vision for this interface.
- *Create procedures for documenting physics and its interfaces*. At this moment, there is no agreed process or guidelines in place for the documentation of physical parameterizations. The result is that the documentation of the GFS 2016 suite by GMTB will age off. It is important to put in place procedures to ascertain that new code committed to trunk/master is properly documented.
- *Understand overall and individual component development milestones and schedules.* Physics suite development requires knowledge of near-term deadlines and longer term goals, and physics group needs to be aware of other software and model milestones.
- *Prioritize the physics development.* Priorities need to be established for the physics development including key physics components, testing procedures, and metrics.
- *Consider HPC requirements.* It is critical for the operational centers and community testbeds to have sufficient computational resources including storage, ease of access, and documentation to meet the demands of full testing and evaluation of the physics in uncoupled and coupled applications.

(11) Post-processing

- WGs would benefit from additional guidance on how to collaborate and make decisions. We are aware of tools out there that may help (VLab, Google Groups, COG). No established solutions to support WG work.
- Communication about overarching issues (GitHub, Licensing, possible collaboration with NOAA Big Data Project) will be important.
- NGGPS may need to focus more on data formatting and metadata.
- Post-processing NWP output is one of the most important ways that America's Weather Industry (AWI) differentiates themselves.
- Many people expressed concerns for NOAA's HPC infrastructure.
- Parallel runs that support NCEP Production Suite (NPS) are routinely shared with the entire weather enterprise. Retrospective runs that support (NPS) are not routinely shared outside of WCOSS.

- A communications plan will be a very important part of our efforts to form an NGGPS community.
- Need to de-bias model output, diagnose weather-dependent variables from model output, and combine output (model and obs) from various sources for better post-processing.
- *Develop an infrastructure to share data* (users who want to do their own post-processing -- ”private sector in the middle of the supply chain”) as well as to continue to make decision support products to support other users. A challenge will be organizing the community to move forward with both tasks.

<u>(12) System Architecture</u>

- *Hire a manager responsible for unified modeling system software development.* (Already known issue) In the absence of this function and an associated team, there is little likelihood of being able to complete the implementation of a unified modeling system architecture and operate it.
- *Identify a developer to integrate FV3 into coupled model configurations.* (New) In the absence of someone in this role, the path is unclear for implementing FV3 in the unified modeling system in a manner consistent with initial SAWG recommendations. More generally, a component lead is needed at EMC for the coupler.
- *Understand current development milestones and schedules.* (Already known issue) Evolution of the system architecture requires balancing near-term deadlines with longer term goals, but the SAWG has not yet mapped its recommendations to known development milestones.
- *Complete investigations in the SAWG’s initial recommendations* (Already known issue). Assess feasibility of replicating the GFDL coupling strategy with NEMS, and possibility of partnering with CESM to develop a community coupler that meets EMC needs.
- *Add nesting and grid refinement to topics addressed by the SAWG.* (New) Options for the implementation of nested systems should be included in SAWG considerations of the system architecture. There may also be a requirement for a standalone regional atmospheric component.
- *Add HPC to topics addressed by the SAWG, in coordination with Infrastructure WG*. (Already known issue) HPC needs to be explicitly addressed in the SIP. The SAWG, in coordination with other working groups (esp. Infrastructure, Physics, and Data Assimilation), needs to consider engineering (including best practices, standards), portability, and performance optimization and scalability of the community prediction system on current and emerging HPC systems (including processors, networks, I/O, and data management).
- *Advance a common view of the Interoperable Physics Driver (IPD) interface* (Already known issue). There is not yet a consensus about requirements, scope and design of the physics interface, especially the interface to chemistry/aerosols.
- *Document requirements for Data Assimilation applications*. (Already known issue) Recommendations are needed for a system architecture that will best support the data assimilation application. Example questions: What are the implications of weakly and strongly coupled DA? Should the architecture support Pause/Resume? Should the architecture support IAU (incremental analysis update) and Replay?
- *Address conflicting requirements for Ensembles*. (Already known issue) Existing ensemble requirements contain conflicting opinions about the level of communication needed between ensemble members – an authority outside of the SAWG needs to resolve conflicts and finalize requirements. The system architecture is impacted by the tradeoff between model integration performance (fewer simulations) and data management performance (many simulations).

- *Prioritize the scientific research agenda*. (New) Many questions with a bearing on system architecture require scientific research, with the answers relating either to the representation of Earth system processes and their interactions or to the impact on predictability and prediction skill (as a function of lead time); e.g., intra- and inter-component interactions (aerosols in 3D interface; atmospheric columns shading each other at high resolution; coupling ocean and sea ice as a “fast” process; lateral water movement at and below the land surface; etc.). All have a bearing on R2O and O2R (support). While a maximally flexible system architecture is desirable to be able to adapt to different scientific outcomes,, that must be balanced by practical considerations.

(13) Verification & Validation

- **Governance:** It became clear that governance must help establish common standards and benchmarks for verification/evaluation. And defining the standards/metrics before the evaluation of any parallel system is critical. Without this, groups are going to do their own verification on whatever they want to, and the results will be cherry-picked. A recommendation is to consider including someone from V&V on the Governance WG. Betsy Weatherhead, CU Boulder, would be a good candidate.
- **Facilities:** There are multiple components that need to be in place to make this work, including how to share data, access to observations; access computing and a funding avenue. Of these, computing is likely the most critical to start addressing immediately.
- **Communication and Outreach:** To facilitate a clear understanding of how the community can engage with the unified modeling effort, an active Communication and Outreach WG is needed. The idea of special sessions at AMS and AGU is supported by the V&V working group.
- **Verification and System Architecture:** There is a need to make sure V&V and System Architecture communicate on a routine basis to make sure the development of these two core components of the unified system are well aligned.
- **Verification and Post-Processing:** There is a critical link between post-processed fields and how to evaluate them. The use of common output formats is critical to the success of the unified system and V&V and Post-Processing need to help define that. It is recommended these to WGs develop a mechanism for meeting routinely.
- **Verification and DA:** There is a critical link between the observation database, forward operators, and diagnostic tools used by DA and what is needed by Verification and Validation. It is recommended these to WGs develop a mechanism for meeting routinely, including having an V&V member integrated into the JEDI planning. John Halley Gotway (NCAR and DTC) would be a good candidate.
- **Definition of MET+:** To take advantage of some of the pre-existing packages out in the community, especially for some of the component areas of research (land surface, hydro, etc…), it is good to start thinking of the definition of MET+ to be broader than the core MET and METViewer components embedded in a Python framework. Making sure other packages can interact with MET but also be checked out separately should be a longer range goal of the SIP.
- **Diagnostics at Every Timestep:** Dynamics and other WGs expressed a desire to be able to evaluate/diagnose some fields every timestep (e.g. fluxes and closing the energy budget). This seems to be out of the purview of a verification package that is intended to evaluate the final products. Upon discussion with the System Architecture group, they felt it might be better addressed by a combination of their group and the CCPP overseen by the Physics group.
- **Native Vertical Coordinates:** While most verification packages, including MET+, currently

operate on standard vertical levels, there is a strong need to also have capability to perform verification on native vertical levels to facilitate model development. Additionally, evaluation of some fields (e.g., brightness temperature) are best performed on native model levels rather than pressure levels.

- **Process-Oriented Evaluation:** This topic ranges from diagnosing fluxes every time-step, to evaluating physics schemes with field project data and remotely sensed observations, to diagnosing general circulation signal such as the MJO Index. Better definition of how a unified verification package needs to support these wide ranging needs, including prioritization, should be included in the SIP.

(14) Communications & Outreach

This WG had its initial meeting near the end of the workshop. Two main goals emerged:

1. To develop a *communications plan* to support *integrated decision making*. This plan should be strategic in nature with short- to long-term goals and include objectives with metrics.
2. To reach out to the community with multiple forms of communication (e.g. posting plans and keeping them updated.

Annex 1. Aerosols & Atmospheric Composition (AC) Working Group

Workshop summary report from Aerosol and Atmospheric Composition SIP WG

This document summarizes the key action items and important information obtained during the Community Modeling Workshop from April 18-20, 2017 by the Aerosol and Atmospheric Composition Working Group (AC-WG). The AC-WG met with the following other working groups during the workshop:

- System Architecture
- Verification and Validation
- Infrastructure
- Physics
- Governance
- Post-processing
- Dynamics
- Data Assimilation
- Land/Hydro and Marine (jointly with other cross-WG meetings)

Meetings with these groups helped identify immediate needs that need to be addressed and critical path items that should be ensured so that the Atmospheric Composition plans and goals can be successful.

The AC-WG was impressed with the breadth of community interest from various NOAA labs, other federal agencies, private sector and academia. Defining the scope of the SIP timeframe to a three year strategy helped the WG focus its efforts. The WG interactions during the cross-WG sessions were crucial to communication and identifying key issues to be addressed. The effort for community engagement would benefit from providing a clearer definition of the community, its role and scope. The decision making process could be better outlined as well as the role of the infrastructure WG. A final report back session after the cross-WG meetings would have helped summarize key findings and next steps.

Immediate Needs

The following items should be addressed as soon as possible so the AC-WG can proceed with planning and component development.

- Unified Coupling Strategy

Discussion with the System Architecture (SA) WG primarily emphasized how atmospheric composition component should be integrated into the unified model system architecture. Some AC routines are already built in modular form and take advantage of ESMF capabilities to couple with physics and dynamics (eg: GEOS-CHEM, NGAC GOCART). Others are applying the more direct route of putting chemistry within the physics driver. ESMF coupling would enable the atmospheric composition component to be self-contained (emissions, 1-D chemistry, deposition), allow ease of code maintenance and optimization as well as sharing of the code among users with different interests, including operations, development and research for either standalone applications or inclusion in the Earth System model with close interactions with other components. ESMF coupling may potentially result in a more complex and less efficient coupling code. A separate atmospheric composition ESMF component could enhance flexibility to more easily call chemistry at a different frequency than other physics processes. SA could permit two different approaches for coupling, but this would complicate code support and therefore, they requested that one coupling mechanism be applied for all chemistry packages. It was also recommended that ozone currently treated as a meteorological variable through the physics driver in the GFS physics be unified with the final atmospheric composition approach.

The WG recommends an analysis be performed to determine the additional expense of calling atmospheric composition processes via a separate component architecture. For example, the stratospheric ozone parameterization currently in-lined within GFS physics could be put into a separate component and computational cost be compared. Ed Hyer, NRL, pointed the group to a preliminary report on the cost of ESMF type coupling within the Community Climate System Model (CCSM): https://www.earthsystemcog.org/site_media/projects/esmf/timing_1005_ccsmoverhead.pdf

Average computational time cost for coupling to atmosphere, ocean, land or ice components was 3.15%. The AC-WG also brought this issue up to Governance WG to obtain more guidance.

- Resource Limitations for transition to operations

Regardless of the desire for greater collaboration to improve NWS global atmospheric composition models, no one has yet been identified to foster the transition from research to operations at NCEP. It would be necessary to identify a dedicated staff member who would facilitate the transition of most promising enhancements via the transition, testing and evaluation process within the NCEP computing environment.

Critical Path Items

The AC-WG identified critical path capabilities needed to support chemistry processes and meet the WG goals. The critical path items are summarized by system functionality below.

- *System Architecture and Post-processing*

The AC-WG System would require the following capabilities to be included in the system architecture and post-processing software:

- Support for downstream modeling (mechanism, LBCs, interfaces)
- Provide for off-line coupling to chemistry & dispersion models (e.g., backtracking dispersion)

These capabilities should be included in the post-processing software well in advance of off-line atmospheric composition testing. Therefore, operational regional air quality prediction and global dispersion forecast would need similar outputs from the unified global and nested weather models in order for testing to commence for ultimate implementation as downstream applications of the current global and regional systems.

- *Physics*

The AC-WG recommends that a small subgroup be formed to evaluate the best approach for coupling atmospheric chemistry modules with atmospheric physics (also recommended above in immediate needs section). The Atm. Comp. WG also provided a list of capabilities that a physics module would support:

- Interactions with microphysics and radiation (aerosol aware) in a modular way, allowing for aerosol models of varying complexity (e.g. bulk, modal, bin). The solution needs to account for efficiency and memory footprint.
  The Physics WG noted that aerosol properties are already passed to several supported physics modules including Morrison-Gettelman and Thompson microphysics. The Physics WG was committed to supporting additional double moment aerosol aware microphysics schemes with options for simplification if aerosol interactions were not desired.
- Incorporation of aerosol schemes that have been developed as ESMF components, this way preserving ongoing, long-term collaborations with aerosol modeling groups at NASA.

For operations, a decision on the final configuration of global and nested model physics would be needed at least 1 year in advance of implementation to ensure the chosen physics and chemistry are properly coupled. For example, inline convective mixing of atmospheric constituents would require mass fluxes to be generated from the convective parameterization. The AC- WG recommended that its related modules be part of the support that CCPP and/or GMTB could provide, however both groups realized there would be limitations to the number of physics and chemistry approaches that this program could maintain. A testing plan should be devised to help determine which capabilities would best meet the needs of operational and research communities.

Finally, modularity and consistency in the specification of aerosol optical properties (e.g., refractive index) in model physics, data assimilation and post-processing (e.g., ensure that CRTM is consistent with other components) should be ensured as these systems are being built.

- *Verification and Validation*

The V&V system should include at minimum surface ozone and fine particulate matter (PM2.5) observations as well as Aerosol Optical Depth (AOD) from surface and satellite instruments to meet AC-WG initial operational capability goals. Therefore, the V&V system needs to support at least WMO BUFR and NetCDF observation formats. The groups also discussed where the datasets would be stored and the possibility of taking advantage of the JEDI IODA sub-system for this critical function. The AC-WG recommends that an existing or new WG be tasked to develop plans for housing key input data (observations, emissions, land surface data) for the unified model.

- Data Assimilation

The AC-WG was interested in how the JEDI program would interact with this SIP process. Many issues that SIP teams will cover are also being considered as JEDI requirements. For example, AC-WG data assimilation key issues were taken from a similar exercise produced for JEDI. In any case, our group emphasized the importance of the data assimilation system for constraining not only atmospheric composition species concentrations, but also the emission, which are typically based on databases that can often be several years old. For example, satellite retrieved NOx and VOCs have been used to update regional anthropogenic emissions databases and satellite retrieved fire activities have been used to update global smoke emissions. Therefore, a JEDI system that supports joint state-emission estimation is of paramount importance for Atmospheric Composition.

Furthermore, it was requested that the data assimilation system supports aerosol information from IR sensors. The AC-WG did request other capabilities for data assimilation, but the most important for initial capability would be the ability to assimilate AOD, PM, and LIDAR backscatter by providing radiative properties of at least current operational species in the CRTM as well as allowing for atmospheric composition data ingest. Particular attention should be paid to surface characterization and polarization effects of UV channels.

Brad Pierce, NESDIS/CIMMS, noted that the NGGPS DA plans for coupling low resolution aerosol predictions to high resolution meteorological model assimilation. The group thought we should meet with the DA group to discuss in more detail the implications of this approach on aerosol data assimilation.

- Land/Hydro & Marine

The land WG emphasized the importance of ensuring consistency of dust and biogenic emissions by using consistent land surface information (land use database, canopy conductance, soil moisture, surface fluxes from land model). The land WG also recommended the formulation of terrestrial emissions on the same tiles used for the land-surface and vegetation models. Therefore, both groups recommended closer collaboration as plans and these components are developed in the unified modeling suite.

Annex 2. Data Assimilation Working Group

SIP planning workshop from the DA working group perspective

Thomas Auligne, Ron Gelaro, Daryl Kleist and Jeff Whitaker (DA co-chairs)

1) Assessment of workshop: The DA WG found the cross-WG meetings were very useful and informative. However, we found the presentation format for the community engagement part of the workshop less so. Much of the material was duplicative, and the format did not seem to result in much useful discussion (at least from our perspective).

2) Top Takeaways/Items on 'critical path':

1) A collaboration model for JEDI with the central software repository outside NOAA is crucial for community participation and efficient collaboration in DA research.

2) The success of JEDI is crucial, since it is tied to progress on key scientific issues (coupled DA, localization etc). In other words, JEDI is on the "critical path".

3) A governance structure for JEDI that allows for rapid development is needed (current process for GSI is subject to bottlenecks, long delays in getting changes merged). Use 'agile' software practices where possible, separate high-level governance from low-level code review.

4) A community-supported tangent linear and adjoint of FV3 (based on work done at GMAO) is highly desirable.

5) Improved feedback between data assimilation and physics development is desirable (by using DA diagnostics to diagnose physics errors, using JEDO/UFO to validate model in obs space).

6) Issues of 'governance' remain of primary concern, not just for JEDI but for the entire unified prediction effort. This is causing a bit of fear, uncertainty and doubt among potential participants and should be resolved as quickly as possible. In other words, creating a governance policy and structure for the unified modelling effort is on the "critical path".

3) Proposed DA WG membership:

Co-chairs: Thomas Auligne (JCSDA), Ron Gelaro (NASA/GMAO), Daryl Kleist (NOAA/NCEP) and Jeff Whitaker (NOAA/ESRL)

Members (proposed): Nancy Baker (NRL), Yannick Tremolet (JCSDA), John Derber (NCEP), Rahul Mahajan (NCEP), Steve Penney (Univ. of MD), Jacob Carley (NCEP), Ricardo Todling (NASA/GMAO), Clara Draper (ESRL), Chris Snyder (NCAR), Andrew Collard (NCEP)

Notes from cross-WG discussions:

DA/Infrastructure:

Arun: repository includes so many components (coupled model) hard to imagine - easier to see how JEDI can be one system

Does unified framework = unified model?

Each model component has its own repository - how do changes get back to component repos?

Need to make sure that changes from forks get merged back in regularly.

Multiple repos can be connected, using collaborative tools, issue tracking, integrated testing (e.g. Github, Atlassian,..?).

Easy to connect Github, Atlassian repositories (both based on git). What extra collaboration tools are needed?

How to avoid proliferation of options problem (WRF)? CESM process - certain combos are blessed and are supported. Different levels: runs/won't blow up, vs scientifically validated. Supported options will require resources.

Don't mix high level governance to low-level code review. Use agile software practices - define tests, standards, rules, processes - trust tests and reviewers, merges are fast and frequent. Tests are very important - both functional and scientifc quality tests.

High level governance involves which configurations are community supported. Has to be a panel for each component (DA, land model, ocean model, etc.). CESM can provide a guide. Should be an appeals process.

Can we measure community demand? Upvote feature request or pull request vote.

Datasets will need to be supported also (observations for DA).

DA/Physics

Topics of mutual interest:

Stochastic physics - test process based vs more ad hoc (physics tendency perts). Tie to observables - how to tune, test validate?

Reduce spinup - spurious physics tendencies (IAU, balance constraints in DA?)

Using DA system/increments to diagnose model errors. Time-mean increments indicate bias, correlate with physics tendencies to determine causes. Calculate time integrated model tendencies from physics to compare with mean increments.

New high-time freq obs (GOES 16, cubesats) that observe the same pixel can be used to validate model tendencies.

Synthetic sat image vs observed radiances?

Initial microphysical species and aerosols.

Linearized physics for 4DVar/TLM? Simplified scheme and linearize it, or linearize full scheme? If not consistent with full physics, may cause imbalance.

Are linearized/adjoint physics useful for parameterization development (including stochastic?).

Dan from GMAO invite to group discussion to talk about linearized physics development.

Tension between physics development (which aims to mimic noisy, nonlinear nature) and linearization/adjoint requirements (which by nature are smooth and linear).

Automatic software works for dycore. For physics it's an iterative process.

DA/Ensembles

Walter: TL/Adjoint for FV3? Ron: GMAO has already done this.

Tom: useful for FSOI (adjoint vs ensemble based)

Ensemble based DA for other components? Land (from GMAO) and ocean (Steve Penney hybrid-gain). Atmosphere is only component that uses EnVar. Wave DA is being done within GSI.

Walter: GEFS spread dropped when going from 3DEnVar to 4DEnvar. Probably due to transition from additive noise to stochastic physics, and subsequent changes to stochastic physics.

Daryl: EnKF runs in late cycle, so GEFS initialized with forecast perturbations that are recentered. Tom: so it's likely related more to treatment of model uncertainty as opposed to initial conditions. Daryl: We are also testing running EnKF in early in cycle (2:45 after cutoff vs 4:45) so analyses perturbations are available for GEFS.

Yuejian: NHC complains hurricane tracks have less spread. Near surface variables are also an issue (underspread in GEFS). Tom: we should do tests with FV3, since GFS is very diffusive.

Daryl: Stochastic physics implementation in FV3 should be a short time priority (may be suboptimal).

Tom A: there is overlap between stochastic physics random pattern generator and DA since they both need a smoothing operator on the model native grid.

Fred Carr: there was not as much of an improvement going from 3DEnVar to 4DEnVar: Daryl: as about ½ of 3DVar to 3DEnVar. How can we improve further? Tom: localization, nonlinearity/nongaussianity (outer loops), propagating of static B (ensemble hybrid 4DVar), larger ensemble size, more radiances. Met Office finds that ensemble hybrid 4DVar beats hybrid 4DEnvar handily.

Thursday 20 April

DA/Dynamics and Nesting

General topic: General DA requirements for (moving) nests? What capabilities exist for nesting/variable resolution (and stretched) grids? Current capabilities rely on use of lat/lon intermediate grids.

Tom A: Within context of JEDI, start with UFO (native grid departure calculations) and then work toward generalizing the solver.

Tom A: Can we work on optimization and compromises to make the TL/AD computationally affordable? Bill P: Some simple things to start with (memory allocation) but not much can be done in the dynamic core itself. For 10km, effective time step within time splitting is about 20s time step. C180 ~60s. Can we do something drastic (SJ: nearly impossible to move toward S-I, 5 year effort). We need Dan H.!

One possibility: Ignore the time splitting (ignore the "n split")? We should have a conversation to explore these kinds of trade-offs for use within the inner loop integrations.

Back to conversation about native grid DA and JEDI: Needs to be done in steps and native grid O-F is the most obvious starting point. One thing we'll need is an operator for B, localization, etc., that is on the native grid and extremely computationally efficient. Option for using diffusion operator, spectral operator, etc.? Can we adapt large halo diffusion operator?

Moving nest capability for FV3: Absolutely a goal we want to reach but timetable is still open/undecided. Work in the community (OU) on nested, moving grid issues, feature-based remapping, etc.

Thinking about ensemble DA: running members from single job/task to minimize I/O may be advantageous. Some of the two-way nesting is designed in such a way and this could/should be possible.

DA/System Architecture

Requirements that come from the DA application that Systems Architecture should / could be addressing?

Where does support and requirements lie for development and support for TL/AD codes? Automatic differentiation has come a long way. Physics is a bigger issue than dynamics because of linearization. We need Dan here. General consensus that we need to make a recommendation that the SA supports ability to have TL/AD software. One example would be backward clock (for adjoint).

Going through DA requirements: ability to call forecast model multiple times within same executable without large overhead (avoiding additional calls to setup). Replay with IAU at GMAO.

Important that the capabilities be available for all components in coupled context (and combinations thereof).

Cecilia: Some concern about different concerns from different groups that seem uncertain (example given by sub-seasonal project in terms of need for digital filter capability). This feeds into requirements from system architecture perspective.

Jim: Envisioning a set of "use cases" for things to support (replays, use of IAU, etc.)? Should DA impose this as a requirement? Call UFO from model or have UFO call the model, what does this mean for the architecture?

What is meant by replay mode? Running the forecast twice, all in memory, whereby second time is with IAU forecasting from offline analysis increment…

Examples of models that do not have these requirements? CESM doesn't have pause/restart but they are working on it. Requirement: component must be thread safe.

From SA perspective, ability to avoid redoing of "setup" may be very difficult or nearly impossible (see part of DA1 recommendation for SA group).

Following experience of DART: are you (DA) going to document specific requirements to add a new model or component into unified DA/JEDI? Make sure SA group has what is needed in time to react.

Jim K: "Strongly coupled da" seems to have biggest implications on SA, but in reality, it is probably a bigger requirement/implication for the DA component.

DA/CAM

Requirement for single core for effective CAM ensemble DA? HREFv2 as operationalization of SSEO this September. This has mixed physics and mixed core components including time-lagged. Are we being hurt by having this mixed system? SPC/NSSL will be looking at CLUE (community leveraged unified ensemble)?

Bottom line: Would DA be preferable using single core/model, for example? Preference for single model for ability to represent error statistics that are usable in the DA? In terms of building best-practices, single dycore would definitely be preferred. Multi-physics could work but also creates technical and scientific challenges.

Glen Romine: Single core, single physics storm scale ensemble. 80-member ensemble at 15 km to initialize a 10 member CONUS ensemble.

No plan to move SSEO into ensemble DA realm. However, this will be baseline for HRRR-E or FV3-based convective scale ensemble DA. General consensus that single core DA is probably the way to go. Multi-physics seems like a possibility but comes with its own problems. Would likely not be preferable path from a pure DA perspective. Stochastics (PBL) seem to be effective within context of HRRR-ensemble work underway. Separating spread in DA from spread in ensemble forecasts (similar issues with global)... [from Curtis]

Software engineering issue: Need to minimize communication, IO. Having ensemble members available on same task in core. IO is already an issue.

NSSL/WoF: DA every 15 minutes with forecasts at top/bottom of hour. 3hr v. 1.5hr, respectively.

Domain is size of a couple of states, 3km. HRRR DA ensemble is feeding NSSL/WoF application. EnKF is going to be running with NEWS-E but in parallel?

Thoughts on nest with DA? Covariances across nest boundaries? How much of an issue is this with full global domain with nest and rapid updating? NAM experience: for now do separate analysis for parent and 3km nest.

How to build multi-ensemble with multiple resolutions? Follow example of Xuguang with the supplementing of the standard EnKF with time-lagged full resolution members with additional control variables in the ensemble?

From the CAM-scale: We will likely be dealing with additional control variables? What observations? What forward operators? Toward multi-scale data assimilation, general scale dependence, etc.

Fine scale DA: assumption of hydrostatic? Even now we have some flexibility depending on configuration.

From Jacob: Even though we talk about wanting to promote single core for DA perspective, we talk about the desire to incorporate global ensemble which seems to counter previous point. (Good point!).

With rapid updating and 4DEnVar, need to do so for hourly update? Sure, there are observations there, we could configure with backward looking windows, overlapping windows, etc.

Partial cycling is still a requirement for meso/CAM applications? It seems so, in particular for large domain applications? Can we come up with solutions?

Afternoon:

DA/V&V

Where do observations fit in? Does it need to be its own group? To some extent, DA spends a lot of time collecting, selecting, QCing and can probably help. Outside of real time GTS-based observations, what happens with other stuff that may not be so straightforward? Older data that may not be real-time. Non-operational observations, etc.?

Thinking about this in terms of the obs proc group in NCEP? Collecting, tanking, dumping, running some amount of QC? One key step would be at least some level of standardization. What is meant by standardized?

What are expectations on observation database side? Access should be standardized, and not necessarily what is going on under the hood?

Should be careful to separate observations for DA (thinned, super observations, etc.) from the raw data that you may want in a different form or more of for the verification?

Discussion on quality control and what types of DA information do we really want to pass to V&V?

Do we want to push back on having an observation group and/or ensure we have observation representation on various WGs?

1. Access to broader set of observations.
2. Simulation of model equivalents of observations (UFO re-use for VV).
3. Observation error estimation?

Examples of how you might do simulation of observation in DA versus VV (10 mb interpolation for raobs, reflectivity)?

What would be the most use? Hooks to the UFO is the most obviously and IODA.

To think about: How can VV help DA? Validation of ensemble cross-covariances? Inclusion of DA diagnostics as part of standard VV, scorecards, etc. Where does FSOI fit in?

DA/ Marine, Land/Hydro, Atmos. Composition

Composition: What are we talking about in terms extra variables? In terms of aerosols, 15 species but we only observe 2D (AOD). Several approaches that could be taken, GMAO choosing EnKF for cost? Navy does 2D analysis and then scales the 3D background? What kind of ensemble size are we talking about for this? For LDE, using something like 300 members but only perturbing the meteorology (from MET EnKF). How to perturb emissions, or include emissions as a control variable? Specification of emissions are probably biggest source of uncertainty in the system.

Chemistry is also uncertain. Some observations of precursors to ozone (NOx?).

How good is the model for these applications? Performs well, perhaps better than expected.

PM2.5: Forecasters are relying on this information with regularity. Known biases, particularly coming from emissions. DA could really help constrain the emissions (online bias correction, parameter estimation, etc.).

Overall Recommendation: Means for doing emission estimation either separately or jointly with state estimate.

For uv: Need a vector radiative transfer (polarized radiances).

Use of IR radiances for ozone. Currently turned off in the operational atmospheric assimilation. Aerosol content from certain channels. Other information that would be of relevance to composition DA effort.

Aerosols are important for SST retrievals.

Getting things like SO2 as a separate observer, component to CRTM, etc.

Land: Would be helpful to articulate JEDI plan for land. Land is really a different animal that does not really work well for variational approaches resulting in the exploitation of ensemble techniques.

Arlindo: For under-observed problems (aerosol, some land components), climatological B is simply not good enough. Seems then that “ensemble-B” is critical piece and not necessarily an EnKF.

For the most part, very few folks are working on radiance based assimilation for land (SMAP, L-band, exception). It is not in CRTM to our knowledge. Lots of work on assimilating retrievals. Soil moisture from SMAP (and perhaps SMOS and ASCAT) is obviously starting point perhaps as well as snow. Snow has additional issues. Fractional snow cover: how is this transformed to SWE or depth?

Interest in having soil moisture assimilation especially in coupled systems, and this is something that JEDI should support. Should definitely leverage work from people like Randy Koster on soil moisture work/operators, etc.

What are latency issues? Seems this would depend on component and application? Maybe this changes the priority, but it shouldn't stop from building/having capability.

Requirement for backward looking, catch up type work?

Ocean:

Like for other components, UFO will be an enabler. IODA will help with standardization.

Steve Penny has spoken to GFDL about modularization of remapping that could then be used independently of the model.

Avichal: Need connection to VV so that modules for QCQA and/or flags/metadata for general use. Database will carry some metadata such as QC flags, feedbacks from da, etc.

How will new applications plug into JEDI? Example given was streamflow. The whole idea is to use a structure and standardization to make this easier.

Annex 3. Dynamics & Nesting Working Group

Brief Assessment of the Workshop:

The Dynamics and Nesting Working Group has taken active participation in the Community Modeling Workshop and SIP Cross-Working Group Coordination Workshop. Overall impressions of the workshop are quite positive, with lot of information shared from across all levels of NOAA leadership and from public and private partners of community at large. The community modeling workshop had the right ingredients. It was a great opportunity for highlighting the need for unified modeling system for research and operations, and for developing a proper governance structure including concepts presented by Ming Ji for effective communications and collaborations.

The SIP Cross-WG workshop served most of its purpose through intense technical and programmatic discussions among various groups. Several D&N WG members felt that a debrief from each working group after these discussions would have been appropriate. Several of the technical discussions were cut short, and some of the cross-WG meetings were in-effective (especially with Infrastructure, Physics and CAM WG) as the difference of opinions were not successfully moderated or resolved. The follow-up procedures established by the EMC Director are expected to put these cross-WG coordination meetings in a perspective with more concrete steps that can be taken in developing a comprehensive planning document for SIP V1.0 by end of the summer of 2017. Several WG members also expressed interest in a follow-up meeting to finalize the contents of draft SIP V.0 document, and to resolve many outstanding issues related to infrastructure, framework, code management, governance, roles and responsibilities of partnering agencies, development and implementation timelines, and community support.

Top Takeaways:

- So many people are interested in NOAA unified modeling efforts is a very good sign.
- NGGPS needs to move forward with unifying the global and regional models, and in particular exploring the advantages of global-to-regional modeling to reach the convective scales.
- There was an apparent disagreement between ESRL/GSD and GFDL on unification approach, with specific interest in developing stand-alone FV3 regional Limited Area Model (LAM), and revisit idealized studies to test/confirm the suitability of FV3 dynamic core for convective allowing modeling applications.
- Engineering a standalone LAM version of FV3 could be less difficult than first imagined.
- For tropical cyclone forecast applications, moving nests are needed.
- D&N WG interactions with the DA WG were very good, especially seeing the cooperation between PSD and EMC, and the efforts at NASA.
- The Physics WG was not representative, as few if any global model physics developers (except for people at EMC) were present.
- Many demands were made of NOAA (especially GFDL) during the workshop. What do the NOAA laboratories and broader community plan to do to help improve NOAA modeling, especially our nation's forecast models?
- An excellent 3-Step plan was prepared to extend FV3 to the upper atmosphere for Space Weather applications. SPWC are very pleased with these plans and recommended moving ahead with implementing these plans.

Immediate Needs/Items on Critical Path:

- Strategy for stand-alone FV3 LAM development must take into consideration global-meso unification priorities along with physics and data assimilation strategies.
- Development of moving nests for FV3 critically depend on choice of framework, feasibility in operational settings, and computational efficiency.
- 3D physics development for space weather applications might need a separate strategy than that is pursued by Physics WG
- The current data assimilation does not support nested meshes, which needs to be accounted for in the JEDI development.
- Access to the model and model documentation/training needs to be easy in order to enable the community to participate. In addition, funding needs to be available to allow for community participation.
- Code (and configuration) management, governance, and decision making process need to be transparent.

Notes from Cross-WG Meetings:

Meeting between the Nesting+Dynamics, Meso/CAM and Ensembles working group:

NGGPS needs to move forward with unifying the global and regional models, and in particular exploring the advantages of global-to-regional modeling to reach the convective scales. Doing this will require input from a number of research communities (not just convective-scale modelers). There was an apparent disagreement between ESRL/GSD and GFDL on this approach, with specific interest in developing stand-alone FV3 regional (limited area) model (LAM), and revisit idealized studies to test/confirm the suitability of FV3 dynamic core for convective allowing modeling applications.

Many members put a high priority on creating a standalone FV3 capability anticipating that a large number of community users would have no desire to carry along a global parent for limited area studies. Others believed nests should be used for regional/mesoscale work where a global parent would always be used as the source for nested boundary data. This eliminates many inconsistencies along the boundaries that are present in Limited Area Models (LAMs). The vision was expressed that nested global models will eliminate the need for LAMs in the future. On the other hand, it was pointed out that very high resolution domains are currently run with rapid updates. Rerunning the global nested model every few minutes might not be feasible, and it was argued that a LAM configuration is needed to include, for example, new radar observations at a frequent rate.

Engineering a standalone LAM version of FV3 could be less difficult than first imagined. Currently each face of the cube is a limited area domain that continually receives boundary updates from the adjacent faces. A first major step toward the new capability would be to isolate a face of the cube and replace its boundary updates with data that instead originates from preprocessed external files. Interpolations from the GFS initial conditions to the FV3 grid are already available at GFDL. Infrastructure changes are now required to read in the boundary conditions instead of computing them in the global FV3 domain. This was viewed to be feasible by the GFDL/NASA dynamical core developers. GFDL invites community members (e.g. EMC personnel) to spend an extended amount of time (like a month) at GFDL to help build the regional model capability.

For tropical cyclone applications, moving nests are needed. This raises the question whether a potential LAM version of FV3 needs to provide this capability, or whether tropical cyclone forecasts will always use the nested global configuration (with a moving grid capability under development). A proof-of-concept design with two distinct approaches for developing moving nests for FV3 is already made available by the WG members. The first approach is the Next Generation Generalized Nesting Framework (NGGNF) being developed by AOML/HRD while the other approach is the extension of existing FV3 nesting capabilities for moving nests within the cube sphere grids. NGGNF appears to be a high-risk high-gain approach while the latter could be more efficient for TC forecast applications. It was noted that the land model uses the identical grid as FV3 at GFDL. However, FMS changes required for nesting work in the FV3 are not well defined or understood, and the choice of NEMS/ESMF approach vs. FMS approach have their own advantages and disadvantages.

There is some disagreement on the need for running many more idealized tests with FV3. Those in favor of these tests tend to be strongly in favor of building the standalone capability. However, in order to involve a larger community it is advantageous to provide a few idealized test cases with a public release (e.g. selected dynamical core tests) since they are typically easier to use and validate than complex test configurations. Idealized tests provide a training opportunity for new users and developers. A tutorial meeting 'How to use FV3' is envisioned. Since the NGGPS model is intended to be a community model, a MUCH more useful meeting would be "How to use the NGGPS model", of which FV3 is a part. This would have much broader reach than a tutorial focused just on the core, which is only a component of a complete modeling system. The need for extensive documentation of FV3 (both a scientific description and a user's guide with example namelist/parameter/runtime settings) has been highlighted.

<u>Meeting between the Nesting+Dynamics and Infrastructure working groups</u>

This meeting was initially called to discuss the FMS changes required for nesting work in the FV3. The Infrastructure group instructed the Nesting+Dynamics group to speak to the System Architecture working group. The System Architecture group, later in the day, instructed Nesting+Dynamics to go back to the Infrastructure group. This left us without an answer.

Also discussed in this meeting was the repository to use: git vs. subversion, and which server. It was decided that git was the correct path, and that EMC could be trained in git with minimal cost. The complicating consideration was about which commercial company to use. While transferring code between git repositories is easy, transferring associated material (documentation or code history) is not easy. This would almost certainly be true for any commercial service that was chosen.

<u>Meeting between the Nesting+Dynamics and Physics working groups</u>

Two major discussion points were addressed. The first focused on the physics-dynamics coupling and, in particular, on the "Interoperable Physics Driver" (IPD). The second discussion point was centered around the scale awareness of physics parameterizations.

<u>Physics coupling:</u>

IPD is an independent interface that separates the physics from the dynamical core in such a way that the physics is swappable for a given dycore and the dycore is swappable for a given physics suite. It was highlighted that IPD provides layers with additional functionality like concurrency, different processor configurations for physics and dynamics, and regridding capabilities. The latter point triggered the remark that regridding should not be part of a physics driver, but part of the dynamics component.

IPD is currently linked to the Common Community Physics Package (CCPP) that is under development by the Developmental Testbed Center (DTC). It was revealed that a release of the CCPP physics in planned for May 2017, but that only physics stubs are available at the current time. The initial vision of CCPP was to swap a complete physics suite (like the GFS physics package). It remained unclear though whether the GFS physics package (as used by the NGGPS project) will also be part of the planned release (which would be desirable). It was highlighted that it is also possible to swap individual physics schemes via an xml text file in IPD. In addition, it is possible to specify parallel-split and sequential-split physics-dynamics coupling options via the xml file. A potential future limitation is that IPD is currently not able to execute physics and dynamics on different grids, and that there are no options for 3D physics (only 1D column physics). These limitations might need to be revisited in the future.

- CCPP approach may not be suitable due to tight coupling between dynamics and physics as well as inter-connections between various physical parameterizations, and apparent need for fast and slow physics to be efficiently executed in any model integration.

Scale awareness:

Questions were raised whether the physics parameterizations need to be switched to different schemes at different resolutions (based on their scale dependencies) or whether current/future-generation physics schemes can adjust themselves. It was argued that the deep convection scheme can be made scale-aware more easily than other schemes (even on stretched grid) and that the scale-awareness of the microphysics scheme seems to be critically important. This is in particular a challenge for hurricane forecasts that need to resolve or adequately capture the hot towers and mesovortices (critically important for hurricane strength).

- The need for "scale-aware" physics in the next generation Interoperable Physics Packages was also discussed at the meeting. It was discussed that physics may also need to be "core aware," noting that a different dynamical core will likely change the spectrum of waves captured by the model.
- Any physics package will need to be core-aware. Physics packages MUST be developed in tandem with the core and the rest of the model if useful results are to be obtained. The integration of physics and dynamics is much more complex than simply the wave spectrum captured by the model; time stepping, diffusion, vertical coordinate, and especially moist thermodynamics must be taken into account.

Meeting between the Nesting+Dynamics and Verification+Validation working groups

The main take away message is on the kind of verification that may be required for improving models. I think this is still work in progress. Until we have an end to end system, it is difficult to come out with a set of developer driven needs. This is especially true for high-resolution nests. The verification group will continue working with the dynamics group. Some good progress is already made with MET and MET-TC. It was recommended that we have a registry where the required model fields for further verification and validation may be updated on the fly (before running the model). It appears that such a registry is already available for the GFS. One of the challenges for both the teams may be on mining those data sets.

The discussions addressed both technical and scientific aspects. From a technical viewpoint it was noted that the current V&V techniques are single-processor applications. Such a practice will be unsustainable for big data sets, especially for future high-resolution simulations. Concerning suitable V&V metrics, the Nesting+Dynamics group suggested evaluating process-oriented diagnostics like (as a concrete example) the warm-season propagation of Mesoscale Convective Systems (MCSs) over the Great Plains, and whether they trigger nighttime precipitation peaks. Another example is the tracking of gravity waves. Such process-oriented model diagnostics are currently not available. Other desirable metrics are wave analyses (Fourier analysis) or kinetic energy spectra. Radar data should be used for validations in combination with instrument simulators. In addition, it was noted that NASA GEOS (an FV3-powered model) has a satellite simulator that can be moved into another model. However it is not part of the FV3 core itself, but that it needs improvements to support the V&V process.

Questions were raised whether some new phenomena discovered in model simulations (e.g. a double warm core in hurricanes) can even be verified with observations? Some V&V assessments also ask for high-resolution 10m downscaled wind near the coastlines, which cannot be observed. A V&V need for vertical velocities was identified, and a need for upper atmosphere and space weather V&V. The latter should include breaking gravity waves and tides in the upper atmosphere that can be verified with satellite observations. Looking into the future, seasonal forecasts will become important. Therefore, the evaluation of the Quasi-Biennial Oscillation (QBO), the downward propagation of Sudden Stratospheric Warming (SSW) signals or, more generally speaking, troposphere-troposphere interactions should be included. The assessment of potential vorticity (PV) contours will be helpful. These are not necessarily directly available from observations. Another very important V&V focus area needs to be on 'extreme value statistics'. Future high resolution model configurations are expected to capture more extreme events and the V&V suite needs to expand in this direction.

Notes on developing FV3 Deep Atmosphere Dynamics

- Henry Juang (EMC) has prepared an excellent 3-Step plan to extend FV3 to the upper atmosphere for Space Weather applications. SPWC are very pleased with these plans and recommend moving ahead to implement the Step 1 phase, to allow for variable gas constant and Cp with transport of additional species. Step 2, to extend the grid to ~600 km (~10(-7) Pa) should follow shortly after, which will bring the dynamical aspects of the the extended FV3 to a level equivalent to the dynamical core of the current GSM-WAM, without space weather physics but with the additional option of non-hydrostatic processes in the FV3 dynamical core.
- The Dynamics-Physics break-out group discussed the need for 3D physics. The upper atmosphere requires both horizontal and vertical molecular diffusion, both of which have to be treated by implicit time integration and cannot be accommodated by column physics. In addition, high horizontal resolution planned for FV3 renders the geometry of 1D physics impractical, as horizontal resolution approach vertical scale sizes of physical processes. An effort is required to plan how 3D physics could be included in GSM-FV3 for both the restricted and extended atmosphere configurations, at minimum for horizontal molecular diffusion. GSM-WAM currently includes horizontal diffusion in the dynamical core, similar to the way horizontal filters are implemented to impose numerical diffusion. The difference is that 3D molecular diffusion relies on real physical diffusion rates (not numerical). At the same time, working with SWPC scientists, the additional space weather physical processes should be added. This would bring GSM-FV3 to a level to enable detailed comparison with GSM-WAM and validation of the Space Weather capabilities of the new dynamical core, and explore impact of additional non-

hydrostatic processes. It would be valuable to enable extend runs of the model for these comparisons and space weather validation to be possible with one year.

- The need for "scale-aware" physics in the next generation Interoperable Physics Packages was also discussed at the meeting. It was discussed that physics may also need to be "core aware," noting that a different dynamical core will likely change the spectrum of waves captured by the model, and change the gravity wave parameterization required to tune the mesosphere and lower thermosphere dynamics. Improvements in the dynamics of an extended model has potential impact on seasonal and sub-seasonal weather forecasts.
- Step 3 of Henry's plan would introduce the "deep atmosphere" changes, where the dynamics depends on radius, gravity varying with altitude, and additional Coriolis terms. This component would extend the capabilities of GSM-FV3 beyond GSM-WAM. It would be valuable in add this capability and validate within an additional year.
- SWPC welcome the Community Model concept for the deep atmosphere GSM-FV3. There is also an urgent need to simplify and provide educations of scripts, improve access to development computer system that can mimic the operational system, and address the issues of very limited computer storage.
- For verification and validation, WAM group should define variables for VV.

Annex 4. Ensembles Working Group

X-WG discussions of NGGPS/SIP Ensemble WG at NGGPS/SIP workshop, April 2017

Co-chairs: Tom Hamill (ESRL/PSD),Yuejian Zhu (NCEP/EMC), Ryan Torn (U. Albany)

Overall workshop assessment: the NGGPS/SIP workshop was a valuable exercise to bring the broader community together to try to make sure that the FV3 prediction system is a truly community-supported tool. The best part of the workshop were the cross working group meetings, as they facilitated discussions between people that often were not interacting with each other. These should happen more regularly.

Top takeaways:

1. **Immediate needs.**
    a. *Defining governance procedures, setting up archive*. Setting up a governance structure and a repository for code (outside NOAA firewalls), needs to happen ASAP for project success.
    b. *Setting up a low-barrier computational resource for community members*. As computer procurements take time, to get this in place a few seasons to years hence, the issue needs to be worked now. See summary notes below for more.
2. **Items on the critical path (related to ensembles).**
    a. *Resources are needed for the middle of the R2O funnel*, i.e., research supporting developments that may lead to implementations 3-5 years hence. It makes sense that this funding should be run through OAR, considering the recent imprimatur of the Weather Bill.
    b. *Identifying resources for reanalysis/reforecast generation*. Our sense (from ensemble experts not being consulted) is that HPC procurement does not yet account for this new requirement.
    c. Per notes below, the model for central computing and then widespread dissemination of data is broken. We need to figure out how to *bring computing to the data*.
    d. *Coming to clarity on a weekly GFS ensemble*. Hendrik Tolman has suggested this as a replacement for the SREF and to some extent for the current GEFS. Whether this happens, and what the configuration should be seem to be up in the air. We need to clarify whether (i) such a system is needed to meet product requirements, and (ii) assuming so, what configuration is desirable, including the length of reanalysis/reforecast for such a system.
    e. *We need to define procedures and ground rules for R2O suitable to both internal-NOAA and external members*. What does it take to run the gauntlet from research to operations? See notes below for more.
    f. *We need to work out procedures for the eventual unification of global and regional ensembles.* These communities are largely disjoint right now, and are run out of separate WGs.

Summary of notes and suggested recommendations stimulated by cross-WG meetings.

Community-related items:

(1) Many of the issues for ensemble system development, like other NGGPS components, are 3+ year development issues. For example, how will we treat the uncertainty of the coupled model state in ensemble systems? Right now ensemble development is occurring largely with resources for the bottom of the R2O funnel, at 1-2 year scales, with concrete operational deliverables.

> *Rec: OAR should develop base funding and grant mechanisms for NGGPS targeted to these 3+ year development issues. This is consistent with recommendations in the recent NOAA weather bill authorization.*

(2) The respective roles and responsibilities of the various agencies under a community paradigm are unclear. NSF / NCAR seemed to be concerned that community modeling was their responsibility, not NOAA's, and that they already have community systems. Navy and USAF had security concerns about sharing prediction code they might use.

> *Rec: NOAA should proceed with FV3 community model development despite these reservations but should try to work to address concerns of other agencies.*

(3) Aside from Brian Gross (OCIO) attending first half day, we were missing people who represent hardware procurement and utilization at the meeting. There were many questions after the first half day about that.

> *Rec: Request full participation from OCIO and NCO at next meeting.*

(4) The ensemble group supports the general Auligne-provided concept of github-type repository outside NOAA firewall.

(5) We need governance procedures drafted and agreed upon sooner rather than later.

> *Rec: development draft governance procedures quickly and circulate them for comment, as this is a necessary early step in the development of a community.*

(6) Under the community model paradigm, we envision external grant recipients and others external and internal to NOAA playing a role in development of the community system used operationally. How would this work in practice? Hamill has previously suggested some guidelines for R2O stages and gates of transition that might be a first draft of a community-wide plan. See here.

> *Rec: Solicit broader review of these guidelines, and if in agreement that the basic concepts are valuable, adapt these for more widespread use.*

(6) At the workshop, coupling was subsumed into infrastructure team.

> *Rec: consider establishing a separate coupling team, as there are significant science issues involved in coupling.*

Ensemble-development related items.

(1) Ensemble system development at the global and regional (storm scale) is mostly occurring independently. This is in part due to the differing processes of concern (mid-latitude synoptic weather vs. storm scale). Nonetheless, in the future there is the desire for system unification, with FV3-based

systems being used for both purposes.

Rec: global ensemble and Meso-CAM should develop a roadmap for eventual consolidation of ensemble systems, with shared ensemble procedures to the extent possible.

(2) There are many ensemble configuration decisions ahead of us, with the UMAC imprimatur to make these decisions in an evidence-based framework.

(a) The current concept for a GFS weekly ensemble is currently a bit ill defined. Are requirements for it clear? Is it going to happen? If not, what information is needed to make such a decision? What length of reanalysis/reforecast is needed for such a system? Will we stick with 4x daily for this weekly system?
(b) The monthly GEFS system is currently not well defined. What is the cadence of such a system, e.g., 1x/day?

Experiments to make such decisions in an evidence-based framework are expensive and are hard to accomplish rapidly with the FV3 development in progress.

Recs: (a) STI should clarify requirements for weekly system. (b) NGGPS/SIP ensemble team should develop a test plan, what tests need to be conducted to clarify system configuration.

(3) What data to be saved from reforecast? Ensemble team members asked questions at the workshop of various potential users. To support wind energy, multiple levels of near-surface winds useful. Turbulent kinetic energy information desirable. Vertical cloud and RH distribution better than downward solar for solar-radiation forecasting.

Rec: Hamill to more broadly solicit information from users for what forecast variables should be saved in real-time and reforecast ensembles.

Computing and data storage resourcing.

(1) Because of the computational expense of ensemble research, *supporting a community ensemble prediction system developers cannot rationally happen without an adequately large pool of high-performance computing and disk storage.* NGGPS grant recipients (and regular NOAA staff working with them) should have ready access to such a pool of computing, with a low barrier to access (no lengthy security clearance process).

Rec: Pass this recommendation to NOAA/OCIO and NCEP/NCO so that this is reflected in future HPC procurement.

(2) Ensemble data, including reforecasts, is too voluminous already to disseminated widely. With compute capacity increasing faster than disk storage, which in turn is increasing faster than I/O bandwidth, the problem is going to get worse. A paradigm shift is needed. One potential model is to *bring the computing to the data*. Those seeking to synthesize ensemble data (say, into event probabilities), or postprocess it, or verify model forecasts in creative ways could have a moderately sized pool of HPC that had read-only access to NOAA's data sets. They could then disseminate the (greatly reduced volume) of synthesized information.

Rec: Pass this recommendation to NOAA/OCIO and NCEP/NCO. Recommend that they explore the potential for cloud computing vendors to co-locate some of their compute resources with

NOAA data storage.

(3) The regular generation of reanalysis / reforecast is a new "requirement" for NOAA. It is not clear that it is reflected in HPC and disk purchases. Upcoming production of next-generation subseasonal ensemble reanalysis/reforecast is squeezed onto WCOSS currently. This could affect other development negatively. Further, WCOSS cycles are expensive, given the high up-time reliability.

Rec: suggest that OCIO and NCO work with NGGPS/SIP ensemble team and broader NOAA management to develop a computing procurement plant that incorporates the requirement to regularly generate reanalyses/reforecasts.

(4) There are multiple concepts for how reanalyses and reforecasts might be performed. One possibility is that reanalyses and reforecasts are generated only perhaps every fifth year or so, with the model frozen until a next-generation is ready for operational use. Another concept is to generate the reanalysis only perhaps every fifth year and then to generate reforecasts more regularly, with each new model version. White papers on reforecasts here and here provide more information on the tradeoffs involved.

Rec: Given the recommendation above for setting aside dedicated computational resources for reanalysis/reforecast, the ensemble team should develop and solicit approval for a defined concept of operations of reanalysis/reforecast, suitable for defining HPC requirements.

(5) There were differing opinions on multi-model ensembles offered by participants.

Rec: Per past UMAC recommendations, we emphasize that MMEs make sense at the international / inter-agency level, not at NOAA level. NOAA needs to consolidate resources on fewer high-quality prediction systems.

We note in recent conversations via the US-Canadian-British group on postprocessing, there is apparent buy-in from Met Office to share data. Hamill will be discussing the scientific issues behind sharing with Meteo France and ECMWF in May 2017.

(6) Under the community-development paradigm, scientists funded by multiple agencies will be developing the community system.

Rec: Appropriately leverage the deep expertise at other agencies, for example NASA with land modeling and the US Navy for coupled ocean/atmospheric modeling).

We note this follows NOAA/OAR Director Craig McLean's recommendation for defining more clear "swim lanes" for different agencies.

Annex 5. Governance Working Group

Summary Report of Governance WG for Cross-WG meetings at Strategic Implementation Plan (SIP) Workshop (April 19-20, 2017)

Immediate Needs:

- Governance:
  - Transitional Governance needed ASAP to be able to make up-front decisions needed by

other WGs in order to provide necessary guidance for them to draft initial plans (Terms of reference (transience, Membership, Community, Scope, Based on Decisions to be Made, Infrastructure, etc.)

        - What are the goals and timelines for short-term deliverables?
        - High-level decisions on infrastructure strategy
        - Strategic alliance with NCAR model governance? Infrastructure?
        - Multi-agency
        - Define short-term and long-term and bridging (This will be a persistent role of governance.)
        - Separation of operational decisions (internal) and those associated with the community. (e.g., NCEP-NCO Relationship)
    - Compile governance needs from all WGs (i.e., what immediate decisions need to be made by WGs to move forward that require governance involvement) (What are the functions that governance needs to address?)
    - Draft governance structure needed before other WGs provide their draft SIP inputs
    - What is the community?
    - Define if scope of this Governance WG is at level of the components or applications. Will this WG provide governance for, e.g., FV3, MOM6, CCPP, CICE, UPP, MET, etc. (unrealistic)? Or will it assume that those components have some governance and work above it, and governs how these components are brought together to create applications? Or perhaps the best possibility is that this WG provides guidelines and requirements to the governances of the components. As a corollary, if this WG is not responsible for FV3 Governance, who is?
    - Definition of Unified System (Unified Modeling Task Force)
- Infrastructure:
    - Need governance decision ASAP to validate consensus on standardization around code development procedures, which include type of version control system (e.g., Git), hosting service (e.g., VLab. GitHub etc.), and guidelines for development.
    - Define roles and responsibilities in infrastructure and support. Understand what needs will be met with FV3's inclusion in CESM and what needs to be done by other groups (GMTB, EMC, etc.)
- V&V:
    - Need prioritization of V&V capabilities to V&V WG can focus

Critical Path items:

- Governance help needed on rules and standards others will need to plan around
    - Data formats, interfaces, grids, scripting, coding standards, documentation, etc.
- Infrastructure:
    - Data availability for common tests (regression testing, science V&V, etc.)
    - Strategic decisions needed about which parts of infrastructure need to be built vs. those that can leveraged through partnering

- V&V will be an integral component for several WGs
    - Other points on V&V?
- Comms/Outreach
    - Common, consistent and effective messaging to community is essential

Governance WG meeting with Post Processing WG:

- Community governance tools (e.g., COG) needed to manage the community effort
    - Is there an existing or selected collaboration environment for NGGPS?
- Governance help needed ASAP on rules and standards others will need to plan around
    - Data formats, interfaces, grids, scripting, coding standards, documentation, etc.
- How to manage multiple models in a coupled system, each with own post-proc. Software
    - Need to centralize as much common post-proc. as possible (e.g., within MET), but all cannot be centralized so there will be a need to manage multiple post-proc. tools/packages.
    - Can CESM umbrella be a standard for others to follow?
- Need by other WGs for a draft governance structure soon
    - One suggested approach: create a few use cases (e.g., aviation post-proc. for FAA) and work through the draft governance structure
    - How would we share a draft governance structure with the other WGs?

Governance WG meeting with System Architecture (SA) WG:

- SA WG focus has been on coupling components and how to resolve conflicts on science, arch.
    - Implicit vs. explicit coupling
    - metrics and benchmarks
    - portability between HW platforms
- Discussion on needs for governance:
    - How do we incorporate user needs into governance and SA?
    - How will governance handle competing/conflicting requirements?
    - Governance WG needs to go back to all other WGs and solicit their functional needs for governance (i.e., what governance decisions do the WGs need to be made to inform them as they develop their strategic and plans)
- Need greater definition on what applications will be part of the unified system
    - Meso/CAM, space weather, hurricanes, etc.
- EMC product line tiered by time scale: SA WG needs the product requirements since they will affect the architecture
- Common Infrastructure for Modeling the Earth (CIME) is a collaboration between ACME (DOE) and CESM (NCAR); potential solution for NOAA unified system?
- NOAA's coupler (NEMS) is not community code; can we collaborate with CESM to potentially work towards a more open, common coupler?

Governance WG meeting with Verification and Validation (V&V) WG:

- Focus on benchmarks and standards
- Need prioritization of V&V capabilities; relying on Governance to provide guidance
- V&V tools and benchmarks needed to decide whether new code will be accepted in baseline
  - Also need standard V&V tests for initial code check-in (minimum first gate)
- Independence of V&V (separate from development group) is essential
- Availability of observations required for V&V

Governance WG meeting with Marine, Land/Hydro, and Aerosols/Atmos. Comp. WGs:

- Need for component-based metrics vs. metrics for fully-coupled system
- CESM example:
  - Each component model has its own governance Separate "panels" for ocean, land, etc.; higher-level governance for integration
  - Big challenge to manage multiple independent component models is not the lack of a common repository, but adequate/consistent regression testing (note: repository issues mitigated with use of Git by all components)
- MOM6 ocean model: Has made a commitment to coupled-system partners (like CESM) to ensure bit-wise identical matching (on multiple HW platforms, O/S) for updated versions of MOM
  - Also commitment to pre-testing within coupled system
  - Demonstration that use in the coupled system is important to the MOM6 community
  - ACTION: Need to demonstrate/communicate clear distinction between governance for specific configurations on NCEP computers and governance for code standards for the contributing systems
- Formal Change Management System (~Review Board) must be implemented
  - Model Coupling example: time needs to be allocated to test the change before it is implemented in a coordinated way across component systems
- Access to supercomputing resources from non-NOAA personnel
- Standards: realize can't be uniform among components but some minimum functional coding standards and a more common set of standards for coupler "caps" would be helpful
- Aerosols: need guidance on stand-alone vs. inline inside model physics
- Mass conservation is a must for all

Governance WG meeting with Infrastructure WG

- Top priorities for Infrastructure (need to resolve ASAP):
  - Git-based repositories
  - Data availability for common tests (regression testing, science V&V, etc.)
- How do researchers get access to data for specific cases? Note: NCAR already has archive of research-quality data for many cases (supports multiple use case scenarios).
- Training and User Support

  - Options for "summer school" or extended seminars for R&D partners: should work together with Comms/Outreach group on options like this. Also best to outsource to orgs with long track record/culture with this type of training (e.g., NCAR, COMET, CIs); can also explore corporate partnerships/sponsorship
  - Testbeds need to be linked to scientific V&V
  - Note: NCAR has a coupled model testbed (joint CESM – ACME effort)
- Need for some form of V&V panel to build common test cases
- (Governance) Ricky Rood suggested a form of an "Applications WG" to consider infrastructure issues for the end-to-end system
- Data formats:
  - Consensus to move away from GriB2 exchange format in favor of NetCDF4 and BUFR; need to work with WMO to change standards, or just work internally?
  - Also, Infrastructure WG needs to discuss data format issues with DA WG

<u>Governance WG meeting with Meso/CAM WG</u>:

- Migration of legacy Meso/CAM models (e.g., HRRR, NAM) are outside the scope of unified model
  - However some new unified model governance structures (such as a Science Board) can be leveraged to help provide decision-making guidance to the Meso/CAM group
- Scope of Meso/CAM WG related to regional models:
  - Building a prototype regional version of a FV3-based meso/CAM model is in scope
  - Regional hurricane models are not being considered now; could be added later
- NCAR briefed their new vision for unified modeling (to unite or synchronize efforts from MMM and CGD) in the weather-climate space; presents improved opportunity for NOAA collaboration
- Evolution of EMC's meso/CAM models requires a V&V plan, and oversight of a review board
  - NWS/OSTI (H. Tolman) working with Meso/CAM WG to develop V&V plan

<u>Governance WG meeting with Ensembles WG</u>:

- Who is working HPC requirements? Will the ops and R&D systems be adequate for ensembles of the new unified modeling system?
  - Disk storage and I/O needs to be properly scoped to support ensembles (more than CPU!)
- Need to understand requirements for timing/frequency of operational runs (2x/day, 4x/day, etc.) and model upgrade cycle (annual, bi-annual, etc.)
  - Need to account for needs for reforecasts needed for ensemble calibration
  - Requirements for number of years of reforecasts; can subsampling be used for cost/benefit?
  - Need to explore option for on-the-fly reforecasts in production (i.e., ECMWF method)
- Governance need: "gates" for R2O
  - See separate email from Tom Hammill
- Acknowledge that "higher moment" ensembles require larger data sizes
- Barriers between physics and ensembles need to be broken down
- Is there a desire or need to support low-res ensembles for some R&D uses?

Governance WG meeting with Physics, Dynamics and DA WGs:

- Code/repository management issues of physics, dynamics and DA software
  - How to handle NEMS?
  - GFDL (W. Anderson) suggested FV3 dynamic core be treated as its own community code, similar to the way GFDL manages MOM6
  - Consensus: while the development core code can be a living/evolving code base, we definitely need official code releases that signify what version will be supported
  - Transition to Git seen as essential; code management "ecosystem" is also important, but there is no requirement for all components to use the same one so long as it is Git-based
  - For JEDI DA development, JCSDA uses agile development approach where incremental code commits are done in conjunction with automated regression tests, V&V
  - Other?
- Dependencies within dynamics
  - Extra-high model top for space weather applications, such as Whole Atmos. Model (WAM)
  - Nesting (static vs. moving) for hurricanes
- Definition for community may be different for various components
  - R&D community for DA and dynamics will likely be much smaller than for physics
- Will need a pool of reviewers to help make science-based decisions
  - Should those that wish to commit code to the baseline be required to serve as reviewers?
- Based on precedents of other major models, these groups see need for recurring major meetings
  - Annual conference/workshop? Recurring central training courses?

Annex 6. Infrastructure Working Group

Brief assessment of the workshop: The SIP workshop was an excellent opportunity for the Working Groups that are looking at the different aspects of how EMC and the research community should proceed in developing a path that closely aligns the outside community with EMC. The open session provided a very good overview and broad participation with representatives from all aspects of the earth system community and the topics discussed covered the breadth of the needs for a community system. However, the presentations from the different Working Groups over the first day got a little overwhelming by the end. The TORs of the different Working Groups should have been shared with everyone to provide insight into how the different Working Groups were tasked. It would have been beneficial to have the slides with the key findings from the different Working Groups available as the messages got lost in the multiple slides as the day progressed. Having the TORs and slides readily available would have enhanced the one-on-one interactions during the latter half of the workshop. The infrastructure WG strongly advocates having a meeting of this type on an annual basis as it provides an excellent opportunity for the research and operations community to remain engaged. The discussions during the one-on-one meetings were frank and open, and there was general agreement that an open collaborative environment is clearly the desirable outcome. There was still some reservation on how the collaboration between EMC and the research community will play out, indicating that there is much work still to be done in trust building.

Top Takeaways

WG Role: It was clear from the meeting that there was uncertainty as to the role of the Infrastructure WG and how we were different from the SAWG. To address that in our discussions with SAWG, we came to the agreement they own all the technology and science necessary to insert the various community supported components into the Unified Global Coupling System. The Infrastructure WG is striving to handle the remainder of the needs of the Unified Global Modeling System to foster the community participation in the process of building the nation's forecast system. Addressing that was critical in having fruitful discussions.

Repositories: How the repositories for the different code bases will be structured was a point raised in all the discussions. There was wide agreement that the repositories will have to be Git-based as that provides the biggest advantages with regards to code management. Keeping the urgency of the issue on hand, the infrastructure WG will be introducing an early report on code repositories and our recommendations on the rules of engagement to the Governance WG. The choice of where the repository should sit (e.g. GitHub, VLab, BitBucket, etc) will not be a decision that the Infrastructure WG will be making but will be providing an early report on what types of support (other than the git revision control) are needed.

File formats: The Infrastructure WG did not consider file formats in advance, but in discussion with both the Post-Processing and Verification WGs we came to realize that this is an important issue. The standard formats like GRIB2 miss important metadata capabilities that are critical when dealing with large data sets from re-forecasts. (e.g. providing model revision tag as meta data). NETCDF4 was identified as a rich format for storing model data.

Portability: Code portability was a major discussion point with the Software Architecture WG. EMC has traditionally worked on a small cross section of machines that were not deemed enough for research and development. It was felt that the community codes that make up the Unified Global Modeling System should be able to work on a larger cross section of platforms. However, it is not feasible to have all the codes work across all possible platforms. In the end it was felt that the codes should be portable to a finite (yet to be defined) platforms. But it was not clear who will be taking ownership of this. We may have to rely on a distributed model where researchers port codes to the platforms they have access to (with support from the code managers of the respective repositories).

Workflow: A unified workflow was identified as a much needed resource to have the broader community be involved in the development process. The workflow needs to be able to address both case studies / one off experiments as well as cycling runs which mimic operations as close as possible. We should learn from already existing workflows at the major modeling centers to build requirements.

Testbeds: The role of testbeds still needs to be fleshed out. How they can be used to develop appropriate configurations for particular operational system needs, and how can the broader research community be involved in the process, both need to be considered. This may require either a separate WG or a smaller portion of the Infrastructure WG focused on addressing this.

User Support and Training: Regular training in the form of week-long schools was identified as a major need for developing a strong research community. This provides the opportunity for research scientists to send their students to come up to speed with the latest modelling developments in NGGPS but also provides NWS and the research community with a growing pool of future experts that they can draw on. At this moment training for different components is done on an ad hoc basis. A centralized approach would probably yield better results. Training should also be supplemented by user support to answer continual queries.

<u>Immediate needs</u>

Working Groups Organigram: A diagram showing the roles of all working groups and their relationships will help define the platform of the project and how the WGs can better serve the goals of the project.

Repositories: The list of requirements for repository should include but not be limited to a) open-source, b) documentation, c) governance, and d) regression tests

Other factors to consider when choosing a repository:

- Version control system, list management and archives
- Bug/issue tracker
- Basic web server for project/software pages
- Statistics reporting (e.g., number of commits, number of downloads)
- Forums and Wikis
- Project/release management
- How easy is it to integrate other things you run separately with the repository?
- How easy is to backup the entire repository (code, mailing lists, issue tickets)?
- How established and stable is the repository?

- Service Level Agreements for uptime, downtime, time to fix outages and bandwidth?

There are many options, from running it all yourself to paying for fully hosted services.

A comparison of various code hosting facilities is available at
https://en.wikipedia.org/wiki/Comparison_of_source_code_hosting_facilities

An important point to keep in mind when choosing a repository is that a repository serves its purpose in the present. The provisions for the repository must be regularly reviewed in case migration to another service is required in the future.

During the discussions at the SIP meetings, it became clear that a pilot project could be valuable to begin establishing and configuring the necessary repositories (including links to authoritative component model repositories). Necessary governance could also be stood up and exercised simultaneously. We are in a situation where learning from experience is the best approach, because the broad community in the U.S. weather enterprise does not currently have this experience. One immediate challenge is that no organization is currently tasked with repository maintenance and management for the next-generation modeling suite.

Workflow: Starting with the early release (planned for the end of May) the users should have access to a workflow that can be used with the model. This will establish good practices from the onset of the project and enable the community to focus on model development.

Community survey: The communication and outreach WG should work with the other working groups to help design a survey on what EMC can provide (test cases, post processing and verification tools, DA experiments etc.) and at the same time identify the major needs. This should then be broadcast to the overall research community to get a pulse of how best EMC can interact with the broader community.

Items on the "Critical Path"

User support and training for the coupled system: While the main developers can provide user support for running the individual components of the Unified modeling system, users need support to run the coupled model. User support needs to be offered on a continuous basis. Training should be offered on a regular basis. The support and training can be provided by partnerships between existing or new NOAA testbeds and academic research centers.

Scientific advisory committee for the coupled system: While the scientific development of the individual components is guided by the developers who are also conducting testing on the individual components of the coupled system, the scientific development of the coupled system requires guidance and testing. This can be achieved by partnerships between existing testbed centers and academic research centers.

File formats: This was not an issue that the Infrastructure WG was looking at, but in discussions with the Post-Proc. and V&V WGs we came to realize that this is an important issue. While NetCDF has emerged as the consensus format in the research community, there are publishing and organizational requirements that dictate the need for broader discussion and consensus. The Infrastructure WG feels this should be addressed by the Post Proc and V&V WGs.

Summary reports from each cross-WG meetings

Meso/CAM WG

- Discussion on authoritative repositories
- role of test beds
- Recommendation on having automated tests (possibly in the repository)
- Data intensive tests / research data sets should be handled in a different space
- Need build systems across multiple platforms
- Have a code managers for all the different systems
- Need Unified workflow
- NC4 file format is a good approach
- What about MET for marine projects ?
- What will be the interaction with CESM and WRF?

System Architecture WG:

- Repo structures - TA diagram, and add how other organizations fit in.
- Recommend that ops needs to be on board (strong recommendation). Need to stop making big changes to the codes that are access the same repository as anybody else. Current EMC-NCO relationship and how codes are transferred would suggest that EMC remains in between the community repository and NCO.
- Portability - what is the goal? Portable by necessity because all the components will be external and in different organizations. Should we require some level of portability of each component?
- Pilot governance - sci and eng steering groups stood up around initial coupled capability (FV3, MOM6, what else). Initially very small with people who we bless to make some initial decisions about how codes and repositories will be managed. Initial rules expected to evolve as we learn things.
- Who gathers use cases to guide initial governance: who are we serving/supporting?
- Within that, a central location for collecting/coordinating rules, best practices, etc, may be needed before we even start making rules. Can start from Cecilia's ad hoc collection.
- Flexibility of the DA system for various coupling frequencies (e.g., daily, weekly)

Marine + Land/Hydro WGs:

- Standards for code. Recommend standard and style guide for the parts we can control. Expect other system codes (MET, MOM6, etc) to have something similar, but can't enforce it.
- Verification infrastructure. Use case is adding new capability. Need easy, tractable, and known ways to add capability and modify at workflow level. Language potentially an issue.
- From land and composition - role of IPD when existing physics coupling infrastructure exists. Ex NASA has their chemistry and physics as ESMF compliant, and would like to re-use rather than change to a new required physics cap.
- Regression tests. One thing to remember is that a regression test is part of the model. New capabilities require new regression tests. Code developers need to provide regression tests with

their new code/capability. Continuous integration - role of tests in push/pulls.
- Dedicated code manager to make sure all this works, and make sure regression tests are in place.
- Training - how will this be done? Partially a governance question. Includes documentation.
- Where did testbeds end up?

DA WG

- Everything in Git (necessary). Makes connecting repositories easy.
- List of dev tools (to support DevOps) that are necessary (issue tracking, fixes, merge, automated testing). Another list of tools that would be nice to have.
- Judgement on what makes it to the community supported level (possibly multiple levels of support) of a repository is sticky. Clear metrics need to be defined. Judgement needs to be based on what will help meet mission and goals of the community that is contributing.
- How to figure out what people want - community demand? Done through up-voting?
- Many issues related to governance - who sets the rules and what should they be? Mainly, we need the rules set somehow. Would hope that the funding organizations (NWS, etc.) see it in their interest to include external community/views when deciding what are the rules, and how they are enforced.
- Does a DA testbed exist?

Post Processing WG

- Historic data sets are important. In post-processing we often make diagnostic output and maybe post-processed in other ways (e.g. bias corrected), then save those historically for use in making predictions (e.g. via a statistical model). Those data sets are valuable. Is AWS the right vehicle here. Who is working on the big data problem in NGGPS? Someone needs to pick this up.
- Barrier to entry in post-processing is low. Community is large. Argues for a cloud solution.
- Data standards and formats are also an issue. GRIB2 has failed the PP community because of metadata restrictions. Propose an Application Profile that deals with and enforces CF + additional standards for output from the NGGPS. netCDF4.
- Provenance of experiments - ? Includes tagging data sets (model output) with unique identifier.
- Requirement that model output contains info about the implementation, e.g. tag pulled from Git repo.
- Licensing. MDL code should be attached to NGGPS code, with same license.
- New variables in the output

Governance WG

- Pilot code/system governance to get repo set up and lay initial rules in dev environment. (This is a critical need)

- Initial case is FV3 atm. Possible to put FV3 into the main community repo, rather than have it on a NOAA server, which may be in Vlab and doesn't get us around the initial access problem.
- Infrastructure WG stops at EMC (EMC relationship to NCO is out of our purview).
- Rules for repository design and engagement need to be Infrastructure priority
- Need for regular training / summer schools to continually have a talented pool of people to work with the modeling systems (can be outsourced to academia and private sector)
- Regular workshops (overall and topic specific) to engage the community
- A Liaison Officer that regularly informs the research community of EMC projects and priorities
- Role of test beds
- Recommendation for a coupled model validation panel, which will include NOAA labs, academia, and testbeds.
- Website for education of users

Ensembles WG

- How NEMS will handle the hybrid forecasts (high resolution for short lead-time and low resolution for the medium- and long-range forecast)?
- How NEMS will handle the variable ensembles configuration used to determine model uncertainty?

Verification WG

- When MET will become available for other components of the model?
- MET has a help desk, online documentation, online and in-person training
- How the evaluation metrics are setup?
- MET output (ASCII) can be used by many statistical analysis and graphics packages (R, IDL, NCL), however, the plotting capabilities are limited.
- MET can be used as a tool for validating research developments to be moved into production.

Annex 7. Land & Hydrology Working Group

Overall Assessment of Workshop

- For X-WG meetings, discussions were good and right people were there.
- A final plenary with synthesis would have been helpful, but maybe a follow-up webinar with a synthesis could be held with the SIP and/or the community to produce a more tangible summary of the workshop. Did the workshop produce the desired outcomes?
- We were almost always paired with marine and aerosol/air quality, but we didn't have any 1-1 meetings with these two groups.

Immediate Needs:

- Community Computing Infrastructure
- Flexible software design--including multiple grids/tiles for land/hydro and connections between land/pbl; land/emissions/dust; land/marine; land/microwave emission/forward RT models; creation of a "coupling" WG would be helpful here
- Different requirements for land/hydro systems need to be articulated and balanced between short-term service level needs and long-term science objectives.
- Verification group needs land/hydro expertise and a plan for assessment of land/hydro models in physics suites.

Critical Path:

- An efficient model development "hierarchy" will help streamline testing various physical parameterizations, from individual components (e.g. surface drag) to fully-coupled systems (atmosphere/aerosols, ocean, land-hydrology, sea-ice, waves components), with appropriate data sets for testing and benchmarks to "pass".
- Land DA needs will be met by other packages until GEDI is ready (e.g., SMAP). DA WG should work more closely with land/hydro WG to develop requirements and test cases.
- There should be a parallel development strategy with explicit sunsetting of "older" systems. Progress meeting stakeholder (and congressional) requirements needs to continue without stagnation while new system is being formulated and constructed.
- Unification of "climate/seasonal" and "short term NWP"--processes are the same but models are different (e.g., Noah, RUC vs. LM, CLM) Noah-MP bridges somewhat
- Unification of "land" and "hydro"--esp. Better hydrology impacts on marine and or coupled environmental prediction. Need additional management guidance on "water" in unification effort. Horizontal vs. vertical processes.

Summary of Land/Hydro Cross WG meetings

Notes from Breakout1 with SAWG + Marine, Atmos Comp

- Are land and hydro coupled via cap or not (currently both--need to test with cap to verify equivalent)
- Interface with atmospheric composition and land model: BVOC emissions, dust, depend on soil type, veg type, etc. May need dynamic veg.
- Interface with rivers, lakes, estuary, ocean: need to enable a simple “box”estuary model like CESM or explicit estuary model (2-d,3-d) linking land runoff and ocean.
- Interface with boundary layer (implicit vs. explicit coupling). Tiles and blending height.

Notes from Breakout2 with V&V + Marine, Atmos. Comp

- Water and Energy budgets (who should do this? Infrastructure?)
- Other variables? Ocean? Streamflow (would treat as point obs). Can integrate over a region (e.g., watershed) Inundated area?
- What is the V&V architecture? How could existing V&V systems be integrated? (e.g., R-WRF-Hydro package for National Water Model developed at NCAR; LVT package for Land developed at NASA/GSFC in Fortran; MOM6 user-contributed packages). Need to leverage partnerships versus reinventing the wheel. V&V will contact and work with groups like LIS and NCAR-OWP
- Need to deliver the exact data and config that EMC (and others like OWP) uses to make decisions about operational implementation so that external users can replicate operational skill metrics.

Notes from Breakout3 with Infrastructure + Marine, Atmos. Comp

- What is the charge of infrastructure?
- Who owns V&V package? Right now MET is good for ATM but what about other components? This is a governance issue.
- Coding standards? Regression testing? Part of the model. Components must have test cases.
- Workflow system will support individual components plus subsets of overall system.
- Currently recommending Git but will have additional requirements like Github. Also Vlab and Bitbucket (JEDI using Bitbucket and Atlassian suite)
- Data! Will need a “data server/preparer” for non-assimilated data (e.g., topo, land cover) required to run a model. Also sample output for testing.
- Training is a known concern. Exact implementation is pending.

Notes from Breakout4 with Physics + Marine, Atmos. Comp

- Coupling: implicit vs. explicit. Consider GFDL and/or GEOS-5 approaches.
- Consistency in land-dust-BVOC emissions. Need tile scheme for emissions as well as deposition.

Also tile land scheme and coupling to PBL. See GFDL scheme. ***This is a major action between physics, land and atmospheric composition WG.*** HEMICO (Harvard emission component is done with MEGAN). Need to work towards consistency between land cover/land use/veg type datasets for BVOC emissions vs. land surface models.

- Also need better scheme for initialization--esp. for dust forecasting.
- Wind farms and urban areas--need to better represent momentum transfers. WRF has some of this already.
- Need a test suite to quickly test physics options from single component to full system.

<u>Notes from Breakout5 with Governance + Marine, Atmos. Comp</u>

- Strategic decision making process; defining priorities, resources
- Role of UMAC, SAB's
  - UMAC will likely sunset, and some members will be invited to serve on SAB.
- How to balance governance of unified system vs. component models with own governance?
  - For V&V: will likely need a mix of component validation and system V&V.
  - E.g., some components (e.g., land, for NLDAS) are used in offline mode.
- Backward compatibility? Introduce new physics as options. Make sure new options don't break old options. Process for obsoleting old options.
- What are the values of this system? Conservation? Do we need a steering group that thinks about coupling and conservation as well as other constraints that certain components place on system?
- Might need a separate coupling WG to address issues in conservation, implicit/explicit coupling.
- Need to work with governance to figure out "bounds" of unified system….what are end user requirements for unified system and what components does it need to include (or not). Bounds will greatly impact formulation of system.

<u>Notes from Breakout6 with Post Processing + Marine, Atmos. Comp</u>

- Products
  - Bias correction is first order objective.
  - Products such as reforecasts are critical for hydrology applications.
  - Also value-added products such as hub-height (80m) winds, and insolation (direct diffuse)
  - Is time averaging available?
- Need code to map FV3 cubed sphere grid to regular grid for downstream applications.
- How will postproc change in the FV3 era? Try to maintain current capabilities. Have to be careful not to conflict with private sector value add.
- Data dissemination
- Upgraded to produce new chemistry/aerosol products (e.g., backscatter) for verification/application
- Action: Brian Cosgrove from OWP will follow up with PostProc group on NOAA big data initiative. Reforecasting a big requirement for hydrology models (NWM)

Notes from Breakout7 with Dynamics + Marine, Atmos. Comp

- Issue: how to handle moving nests? How do you get land conditions on the nest? --one possibility is to run land on finest grid and couple to other nests as sub-grid tiles.
- Action: explore existing options for lessons learned
- Problem: how to handle ocean/sea ice vs. land vs. waves on their own grids? For example: if you have a coarse land and fine ocean, how do you distribute freshwater flux to ocean?
- If you run a land on a fine grid and atmosphere on coarse grid, how do you provide products for post processing? Should be able to average or provide full resolution land data.
- Action: Need group to focus on both science and technological issues in coupling with land and marine.
- Nesting within component models is an issue for the component, but some needs overlap with infrastructure and/or architecture. Example: storm surge and waves --currently there are grids that are inactive until the water goes inland, and then it starts computing
- Action: storm surge team needs to work with NWM and land model team. (Inundation summit is next week)

Notes from Breakout8 with DA + Marine, Atmos. Comp

- Land data assimilation: Ensemble-based (not var); products vs. radiances; uncoupled (NLDAS; GLDAS) and coupled (NGGPS; HRRR). What is the plan JEDI for land?
- What are the priorities for land DA? 3, 5, 10 years?
- What are the issues with soil moisture DA? Why EnKF rather than Var?
- Soil moisture DA seems like an obvious candidate--GMAO should be able to provide working algorithm (Rolf). Why have there been no experiments run demonstrating forecast impact from SMAP DA?
- CRTM/LSMEM does not contain forward model for L-band soil moisture. Do you assimilate products or do you add and calibrate l-band forward model.
- Snow DA: need to address issues mapping from snow cover fraction to states like SWE or depth. Right now a heuristic process.
- CRTM does

Notes from Breakout10 with Meso/CAM

- Standardized land geo-database (land cover, soils, topography) sub-1km and preprocessing software (examples are geogrid for WRF and LDT for LIS). CLM/CESM and GFDL databases probably not sufficient resolution.
- Wind turbine database, irrigated area database, etc.
- Exception: WRF-Hydro/National Water Model requires data to 10s of meters for routing.
- You may need multiple levels of aggregated terrain data for the model--one can be smoothed to avoid instabilities in gravity wave drag; one can be intermediate (e.g., 1km) to be used for slope/aspect calculations for radiation in snowpack, and one can be high res (e.g., 10m) for streamflow routing.
- Unification of differences between RUC and Noah/Noah-MP? Should we add horizontal variability instead of more vertical variability?

- Groundwater modules--what is the timescale important for that? Groundwater-PBL interaction?
- Ensembles--ESRL is perturbing the soil moisture and using this in the GSI surface analysis to adjust. Leveraging stochastic physics in WRF. should work with

Annex 8. Marine Working Group

SIP Marine Modeling Working Group: Feedback on the SIP Workshop

Overall a healthy attendance at the workshop indicates that there is plenty of interest in NOAA modeling efforts. In general, the workshop was very successful. It's the first time that Marine Models Working Group (MMWG) met face to face. Two MMWG side meetings were held in addition to the cross-WG meetings. The working group members have diverse expertise and backgrounds with representation from government, academia and private industry. The working group co-chairs did an excellent job in organizing, moderating and facilitating the cross-WG breakout sessions.

Not all WG's were created equally -- some of the cross- WG discussions ran out of time to complete their interactions (e.g. with Land & Hydrology WG) while other such sessions finished well before their allocated time. Though, this allowed the MMWG members to have productive intra-WG discussions to know each other better. Members brought forth new items of interest and discussed the role of /NCEP/EMC in building Marine applications for NWS operations.

SIP Marine Modeling Working Group: Immediate Needs

There is an urgent emerging need to better define role of NCEP within NOAA's Water Initiative Vision and 5 year plan (more details can be found at: http://www.noaa.gov/explainers/noaa-water-initiative-vision-and-five-year-plan).

This impacts decisions on operational Surge modeling and future planning at NOAA for providing "Total Water" predictions. (Pat Burke will be soon sharing a report on the "Coastal Inundation Summit" held last week where this issue was also raised). The MMWG also needs to set up an engagement with the Land/Hydrology WG to ensure recommendations from the Inundation Summit are shared between the groups.

Since Marine models are planned to be integrated (coupled) with Atmospheric/Climate models using a unified modeling approach, it is extremely important to have frequent coordination and communications with other WGs, especially the System Architecture (SAWG), Infrastructure, FV3 development (Dynamics) , Model Verification and Post Processing WGs.

We need clear guidelines on the governance of the overall configuration of NWS operational applications. Perhaps a "Change Management Review Board" needs to be stood up by the Governance WG for this purpose. (Governance of the individual Earth System components should be kept separate and the communities supporting these components should maintain such primary responsibility.)

Need a decision on whether a new WG dealing specifically with science issues related to Coupling needs to be formed.

There is also an imperative need to resolve issues regarding access to NOAA's computing resources. The existing NOAA IT security rules prevent efficient collaboration setups with the research community-at-large for NOAA centers and labs. This most likely falls under the purview of the Governance WG.

Critical Path for Marine models in a Unified Modeling Framework

I.Next 3 years (FY18-FY20):

- Short-term weather scale Ocean & Sea Ice based on Global HYCOM/CICE developments synchronized with US Navy efforts.
- Seasonal scale Ocean & Sea Ice based on MOM6 and SIS2 developments synchronized with GFDL efforts.
- Ocean & Sea Ice Data Assimilation based on NCODA/GODAS for providing initial Ocean and Sea Ice states. Continue exploration of ensemble-based methods (LETKF, EnKF) for both weather and seasonal scales.
- Waves model developments include coupling to other Earth System components (Atmosphere, Ocean, Sea Ice) using NEMS/NUOPC.
- Waves Data Assimilation using GSI.
- Foster involvements with community-supported modeling groups (HYCOM, MOM6, WW3, FVCOM, ADCIRC, ROMS, CICE, others?)
- Explore coupling of all Marine Modeling components using NEMS-NUOPC framework.
- Increased engagement with NOAA's Water Team, which is a cross-Line Office team (with embedded governance) developing the vision and timeline to integrate coastal and riverine observations, couple hydrologic and hydrodynamic models, engage regional stakeholders and provide better products and services to the coastal community. Notes from the recent Coastal Inundation Summit will be distributed shortly. Their work can inform the recommendations for the Marine Modeling WG (both short and long term).
- Water Quality/ Ecosystem developments - this is in the long-term plan for the NOAA Water Team, but it is resource-dependent. Again, we should engage with the Water Team to understand their timeline.

II. Long term (FY21 and beyond)

- Engage with NextGen ALE Ocean Modeling efforts with US Navy, GFDL, NCAR, NASA-GISS, others. This model should support weather, S2S and seasonal applications.
- A unified hybrid ODA system which includes various algorithms (3DVar, 4Dvar, LETKF, EnKF). Explore methods for weakly and/or strongly coupled DA. Leverage JEDI modules. This unified system should support weather, S2S and seasonal applications.
- Build fully coupled Marine Modeling components within NEMS-NUOPC framework.
- Strong collaborations (for both R2O, O2R) with community-supported modeling groups (HYCOM, MOM6, WW3, FVCOM, ADCIRC, ROMS, CICE, others).
- Engage with community-supported "Total Water" solutions by developing linkages with hydrology (NMW), marine ecosystems and sediment transport (water quality), and the NOAA Water Team.

Annex 9. Mesoscale/CAM Working Group

Brief assessment of the workshop

EMC (Brad, Geoff): The first day of the public workshop had a lot of good information, in which I ended up taking about 12 pages of notes based on the discussions. The first panel was most the interesting, because it revealed several big-picture issues that seemed to explain the lack of cohesiveness and inherent competition within the US NWP enterprise. I was puzzled by some of the remarks from NCAR and IBM/The Weather Company, which seemed to have been at odds with what NOAA is trying to do. Although the format of the SIP coordination workshop fostered very useful and informative discussion, the number of cross WG meetings became overwhelming. Unresolved issues remain in terms of a governance structure, a clear definition of what defines the "community" of users and developers for the unified NOAA modeling effort, and the divergent and fractious views within our WG on several topics (see next section).

OAR/ESRL (Stan): First half of meeting: Without question, a lot of helpful information was presented. It was good to limit the presentations to the 3-slide format. The presentations should have been be better moderated so that there could have been more time for questions. I agree with Brad's comments above that the lack of a governance and effective repository structure is a current major gap. Obviously from Mike Farrar's talk and others, everyone at the higher management level is aware of that and working to resolve it.

OAR/GFDL (Lin): The remarks from the general community (NCAR and the private sector) were disappointing, asking essentially what NOAA should do for them instead of what they can do for the nation. If NOAA funding is involved, it should be for NOAA goals.

Second half of meeting: This was a more productive part of the meeting. The X-WG meetings were very good exposure to the different concerns of the WGs. Both the CAM WG and dynamics WG are strongly dependent on the not-yet-determined governance and repository structures.

Immediate needs

EMC: Several members within the Meso/CAM WG have identified the need to have
1. A stand-alone limited-area capability for FV3, and
2. The need to conduct additional idealized tests.

Item #1 is of interest to most members of the WG outside of GFDL and NASA. While EMC and GSD and others will be working with GFDL on developing this capability, EMC development will proceed with global nested runs and not wait for the limited-area capability to become available. Item #2 is of interest to members outside of GFDL and EMC. Both items have become wedge issues, and while this was evident before the workshop, it seems to have become more acute after the workshop (particularly item #2). There is a lack of trust between development groups, with some questioning the motives of others. Even the composition of the UMAC group seems to be split between FV3 and ARW "camps", and the concern of some of us within EMC is that the only thing that has changed is that the NMMB model has been replaced by FV3 as the source of conflict within the US mesoscale modeling “community”.

OAR/ESRL: The overall 3-slide presentation at the SIP meeting for Meso/CAM WG highlighted the need to have

- a stand-alone limited-area capability for FV3, and
- A capability to conduct additional idealized tests with FV3.

As EMC points out above, not all members of the CAM WG see these needs but there was sufficient support within the 11-member CAM WG to identify these needs within the community in the 3-slide summary.

OAR/GFDL (Lin): First, regarding the stand-alone limited-area capability, an AOML-EMC-GFDL nested grid meeting was held at GFDL late 2016. A decision was made that EMC would lead the effort to develop a stand-alone FV3 core. For unknown (to GFDL) reason, that EMC effort was not continued (see response from EMC). AOML/HRD (Gopal), on the other hand, has been working with GFDL and communicating with EMC on the development of a moving nest within the global FV3 framework for hurricane predictions. This moving nest (within a global model) development is consistent with NGGPS main goal. Second, with regards to more idealized testing with FV3, a dynamical core developer, not a casual user, is qualified to do the idealized tests for the further improvements of the core. We work with NASA and DCMIP participants in idealized tests. The CAM group is not composed of dycore experts. The CAM group should focus more on high-priority tasks at hand, implementing better and more suitable (for storm scale) physics, and subsequently, evaluating the relative performance in real-world forecasts. Our concern is all this focus on a (downgraded) stand-alone one-way model will cause a delay in the NGGPS implementation, which is GFDL’s main focus. Why would that cause a delay on NGGPS? It is because GFDL scientists would have to spend significant amount of time working on it, responding to basic questions on tests we have repeatedly performed. The level of the scrutiny we have received is stunning, and not all are constructive. Even if EMC is leading the regional effort, we will still have to spend substantial amount of time assisting them. It may not be visible to the CAM working group, but GFDL scientists (Harris, Benson, and Lin) at this point are heavily engaged with the EMC team, including exchange by emails (e.g., FV3GFS Tickets), at times on the order of 50 emails per day (for example, on May 12, 2017), for the development and testing of the NEMS based NGGPS. If we are to divert attention to the stand alone version, with focus on idealized testing and retesting, it will cause significant delay to the NGGPS implementation. NWS, not NCAR, Navy, or OAR labs, should decide what is NWS’s priority. Up to this date, GFDL has hosted visitors from OU and AOML-EMC for joint development of convection-allowing configurations (static and moving nests, respectively). GFDL will host another developer from EMC for space weather (deep atmosphere model) development during late May 2017, and more are to come (see EMC response).

EMC’s response to GFDL: Tom Black still remains the EMC POC for adding the stand-alone limited-area capability to FV3. The latest plans are for that effort to start after the 2017 HWT SFE when Lucas Harris (who has the required expertise) is available to work with Tom. With regards to idealized testing, that is a lower priority for EMC. Should a CAM test plan be established that is agreeable at least to key OAR lab stakeholders, if not the broader community as well? We are nowhere near having a consensus answer to this question.

Critical paths needed to move forward

EMC: The upcoming implementation of the HREFv2 will provide an operational benchmark for which other future CAM ensembles (“CAME” below) will be evaluated against. The upcoming CLUE as part of the SPC/NSSL HWT will provide important subjective evaluation of 10 different CAME systems and several deterministic 3-km FV3 runs. Unfortunately only a subset of these CAMEs will be evaluated for WPC’s FFaiR (flash flood) experiment, and I am not aware of any CAME evaluations for aviation forecasts. Additional metrics were provided to the V&V group for inclusion in a possible CAME-based verification scorecard. Different physics suites (or configurations) will need to be tested, as it is an issue where there is no consensus. But as noted above, the biggest issues are (1) the widespread interest to run FV3 as a standalone regional model, and (2) identifying additional testing procedures (idealized runs and real cases) to convince some in the community that FV3 is suitable for storm-scale NWP. While an ensemble-based system composed of global nested runs would satisfy multiple NWP requirements and unify across global and regional scales, that goal could only be achieved by radically changing the production suite. Because of data latency issues and the need for a longer data cutoff time, probabilistic forecasts from an ensemble of CAM nests within a global model would not be useful at the shortest ranges (<6 h), which means that a regional forecast capability with frequent, ensemble-based, storm-scale DA (HRRR and WoF) will still be needed to provide this short-range guidance used by WFOs to issue warnings and alerts.

OAR/ESRL: The upcoming FV3-CAM test plan.

OAR/GFDL: In response to EMC’s note about “widespread interest to run FV3 as a standalone regional model”, we suggest the CONUS CAM community to look forward, not backwards. Upgrade for potential gain in predictability, not downgrade to limited area model with limited predictability (a couple of days or less). The hurricane modeling community, which should be considered part of the broader “Convection-Allowing Model” community, is fully aware of the limitation of regional model (for track predictions) and the untapped potential of FV3’s regional-global nesting capability. In response to EMC’s comment “to convince some in the community that FV3 is suitable for storm-scale NWP”, GFDL believes that NHC and SPC can provide better answers because of their “core neutrality” and their qualifications to perform the evaluations. The “suitability” question is the most disturbing aspect of the CAM conversation, as it has other implications. The same suitability question can be directed to any existing CAM. Some groups may never be convinced, regardless of the evidence already available or in the near future.

In response to GSD’s FV3-CAM test plan, we believe if yet another test plan is needed, it should be developed jointly by stakeholders who have operational responsibility (in particular EMC, SPC, and NHC), not by a group that may have conflict of interests. We should not even need another test plan -- the evidence will present itself in coming experimental real-world forecasts (hurricanes, hydrology, severe weather, etc.).

Notes from cross-working-group (WG) discussions between the Mesoscale/Convection Allowing Models (Meso/CAM) WG

Breakout #1 with the Verification and Validation (V&V) and Ensembles WGs.

The need for a verification score card was identified that has common elements between the global and regional ensemble systems, as well as additional metrics for Meso/CAM (for brevity will hereafter be referred to as *CAM*). A list of metrics from this CAM WG summary document was subsequently forwarded to the V&V WG.

Discussion also touched on the need for CAM ensembles for several reasons:

1. Suboptimal error growth rates from the GEFS during days 1-3.
2. Only higher resolution guidance can resolve high-impact severe weather events, which are transient and small in scale, are primarily driven by deep convection, and are not adequately captured in coarser resolution runs that parameterize deep convection.
3. Suboptimal guidance of thermodynamic conditions from the current GFS & GEFS for forecasts of pre-convective environment and severe weather (SPC complaint). However, this situation may be addressed as future model improvements are made in FV3.

It was agreed by all that it is desirable to have a single dynamical core as the basis for an ensemble system, including development focus on a single model. Introduction of stochastic physics and ensemble DA is now successful with the global application for providing suitable spread and skill. Similar application at CAM-scale is now starting (e.g., NCAR Ensemble, HRRR Ensemble). While various stochastic methods have improved (including perturbing land states), it remains to be seen whether they can provide superior forecasts compared to ensembles of opportunity composed of multiple dynamic cores with mixed physics. It was agreed that the HREFv2 (a mixed-model ensemble) would serve as the benchmark for evaluating other CAM ensembles, and that new single-dycore CAM ensembles should at least match skill/spread of HREFv2. The CLUE will provide an excellent starting point for evaluating candidate systems at this year's Hazardous Weather Testbed (HWT), but it was recognized that eventually the evaluations need to span all four seasons. Initial sets of variables to evaluate for all seasons, including aviation and precipitation, came out of the CAM WG metrics mentioned above. No formal plan has yet been established yet for evaluating CAM ensembles for QPF/flash floods and for aviation applications.

Breakout #2 with the Dynamics/Nesting and Ensembles WGs.

EMC's observations: Much of the discussion focused on whether development should be made on improving global nested runs for improving CAM ensembles or on continuing development in limited-area domains. There is no consensus within the CAM WG. Until now only members from GFDL, NASA, and OU/CAPS have run FV3 at 3 km resolution, and GFDL views development of a stand-alone limited-area capability as a lower priority item. Many within the storm-scale modeling community, some forecast offices, private industry, and the wider academic community run the limited-area WRF ARW.

Storm-scale ensemble data assimilation (DA) applications, which are driven primarily by using data from a network of weather radars at frequent (<15 min) intervals over CONUS, makes it impractical to do from global nested runs. Being able to make runs over a small, limited area domain is also helpful in early phases of physics testing and in debugging. In response, GFDL said that the FV3 can be run optionally in a "*solo_core*" mode over a domain defined simply by a set of four points on a cubed face using periodic boundary conditions in both directions. While this may be potentially useful for some applications, it is not sufficient to be able to emulate the limited-area capabilities of the WRF ARW or NMMB, in which boundary conditions are provided from a forecast model (e.g. from the GFS, NAM, or RAP). That additional capability would require additional development work (<1 FTE) to be able to achieve. GFDL offered to host 1-3 developers to work alongside FV3 experts to accomplish extra nesting capabilities.

The additional topic of idealized testing was raised by development groups outside of GFDL, NASA, and EMC. This was a contentious topic, in which GFDL felt that this would lead to another bake-off and re-litigation of the NGGPS dynamic core tests, while ESRL, NSSL, NCAR, and select representatives from the university community in the CAM WG argued that it was the best way to engage the ARW user community. EMC seems to be caught in the middle on this issue.

GFDL's observations (Lin and Harris): The statement "*Only higher resolution guidance can resolve high-impact severe weather events, which are transient and small in scale, are primarily driven by deep convection, and are not adequately captured in coarser resolution runs that parameterize deep convection*". This statement is only partially true unless the CAM resolution is actually cloud-resolving (1-km or perhaps finer). The current CAM at 3-km is not near cloud-resolving, and would still benefit from a suitable convective parameterization, particularly the shallow-convection. A unified "scale aware" physics that performs at 1-km and up to 50-km for S2S applications will be very beneficial.

In response to the statement above "*It was agreed that the HREFv2 (a mixed-model ensemble) would serve as the benchmark for evaluating other CAM ensembles*", rather than running numerous CAM ensembles (with different cores, physics, and even different DA), can the HPC resource be better utilized for improving the resolution of a unified regional-global modeling system (for both deterministic and ensemble) to establish it as the flagship US model that can truly be "second to none"? We have seen, based on hindcasts for a full two-year period, evidence that a 13-km prototype NGGPS can already provide nearly as good a guidance on hurricane intensity as the best regional operational model (Bender et al 2017 presentation). The storm-scale prediction is also achievable with such a unified system. We spent the huge amount of human and HPC resources in the past few years to develop the non-hydrostatic core for NGGPS, and we must put that into its intended purpose -- beating the ECMWF at its weakest spot at the non-hydrostatic resolution (<< 10 km) that the IFS is currently ill-suited for. Furthermore, a much improved global system will provide better boundary conditions to the existing CAMs. We should first prioritize the limited resource on the key upstream product (GFS) so that it will benefit all the downstream products.

There are demands for more "idealized testing" from the competing groups, but there has been little justification as to what is inadequate about the existing idealized testing from NGGPS and DCMIP, nor has there been an explanation as to what is inadequate about the real-world evaluation performed by the ongoing 2017 Spring Experiment, and GFDL's presentations at the Intergovernmental Hurricane

Conference on the success of the near cloud-resolving 2-km demonstration using FV3GFS for hurricane predictions. Without such a justification we cannot assume that there is scientific rationale for endless tests and retests that will divert GFDL developers' attention and limited human resources, which will then delay the NGGPS implementation. How many "idealized tests" do we need to repeatedly do to satisfy everyone's desire? Idealized tests are useful mostly for dynamical core developers, not a necessary component for the general community users.

ESRL's observations: The discussion started with the CAM WG asking the dynamics WG how could we together move forward on the CAM WG needs expressed in the 3-slide summary in the SIP meeting on a stand-alone regional version of FV3 for wider-community testing and most importantly, for operational NOAA application for high-frequency (hourly or higher) for which a global domain would exceed NCEP computing for the next several years. Also identified was a need from the wider NOAA community (but not from EMC or GFDL) for an FV3 configuration/namelist to allow idealized testing by the wider FV3 community building from previous GFDL idealized testing for the DCMIP (https://www.earthsystemcog.org/projects/dcmip-2016/). Such a broad set of testing environments for any community participant is commonplace for community models.

The discussion then went to whether development should be made to solely use global nested runs for improving CAM ensembles (advocated initially by some in the dynamics WG and also some in the CAM WG) or to also develop a limited-area FV3 capability. EMC, ESRL, and NSSL expressed a need within the CAM WG for the stand-alone regional FV3 capability. Storm-scale ensemble data assimilation (DA) applications assimilating radar data at frequent (<=15 min) intervals over CONUS already running at NCEP would be too expensive currently to do so with global nested runs. ESRL (and some within EMC) view that hourly-subhourly storm-scale assimilation with a global-nest and later, a fully global storm-scale grid, will eventually happen at NCEP, but assessed to be far enough in the future to require the regional stand-alone capability for the interim period.

Some possibilities were identified to move towards a stand-alone regional FV3-based model. The easiest solution was to use the capability in FV3 to use an arbitrary quadrilateral grid and then adding code for an alternative specification for lateral boundary conditions from an external model. Staff from EMC, ESRL, and NSSL are now being identified to work with assistance from GFDL to develop the stand-alone FV3 configuration. GFDL invited CAM-related modeling staff from any of these organizations for a visit to aid this development. ESRL emphasized the need for carefully controlled deterministic experiments with common physics and initial conditions and resolution for this community testing of the FV3-CAM in comparison with current CAMs.

GFDL's response to ESRL's observations: The regional-global nested FV3 will likely take less HPC resources than the proposed HRRRv3 at the same resolution covering CONUS. The NGGPS prototype can finish the global and the 3-km CONUS forecasts at the same time within a reasonable cost -- a game-changing capability not currently available to NCEP. The overall resource requirement to NCEP will then be significantly reduced, that will allow increased resolution for both regional and global configurations. In case of the ensemble, it will be even more efficient -- because the cost of the global grid will become diminishingly small. The standalone FV3 can and will be built, but we consider that as a downgrade, and it should not be at the expense of the more advanced two-way nested FV3 based NGGPS, which can't be duplicated by existing CAMs. The two-way regional-global CAM should clearly be the higher priority -- because that capability is nearly ready for operation. In fact, it has already been used successfully in the

2017 Spring Experiment in operational situation.

EMC's response to ESRL's observations: EMC will be running controlled experiments that will compare regional FV3 forecasts against what is currently running in operations. The wording was changed slightly in ESRL's discussion to reflect differing views within EMC about subhourly storm-scale DA within global nests versus limited-area domains. Some within EMC may agree with ESRL's statements, while others are not yet convinced. More testing and development is needed to answer a list of questions during the next several years. EMC is open to all reasonable ideas, but decisions need to be based on evidence and available computer resources.

Breakout #3 with the Post Processing WG.

It was recognized that the community needs to work towards a truly unified post-processing code including global coarser-scale and storm-scale applications. The UPP (Unified Post-Processor, code used now at NCEP for all models from GFS to HRRR) needs to undergo substantial redesign (a responsibility that EMC currently plans to address themselves). Sometimes different algorithms are used in post processing that vary depending on the model source, which leads to conflicting information being presented to forecasters. There was some discussion about neighborhood methods used for CAM ensembles.

There was some discussion on the relatively recent trend for CAMs to perform diagnostics within the model for some new variables and passing those diagnostic values (e.g., maximum values, reflectivity) to post-processing instead of diagnosing later within the post-processing code. Obviously, improved optimization over the current UPP would be highly desirable for the rapidly-updating high-volume CAMs.
Regarding ensemble-based post-processing, there was agreement on development of improved capability for new probabilistic fields for variables relevant to severe weather, aviation, hydrology, and many other applications.

Breakout #4 - no meeting.

The Meso/CAM WG was free and did not meet with another WG.

Breakout #5 with the Physics WG.

In general, CAMs have incorporated more advanced cloud microphysics than coarser-scale global models. Enabled by more recent scale-aware physics parameterizations, it may be possible to reduce these differences, although there is still a large computing price to pay for horizontal transport of an increased number of prognostic variables. GFDL and EMC are experimenting with more advanced schemes for global modeling.

Many regional models are now using microphysical parameterizations and even cumulus parameterizations that can be configured to be aerosol-aware, and some in the CAM WG regard introduction of simple prognostic aerosols into upcoming operational CAMs as necessary for improved cloud, radiation, and precipitation fields for aviation and energy forecasts with rapid updating.

Other CAM members prefer to continue to work on improving cloud forecasts, since clouds have a much larger impact on radiation than aerosols. One of the larger sources of uncertainty in forecast models is associated with errors in incoming shortwave radiation at the surface in cloudy areas. The complexity of the problem is challenging. Uncertainties in fractional coverage, hydrometeor type (phase), and particle sizes associated with clouds lead to much larger uncertainties in incoming surface shortwave than for aerosols, and these uncertainties are further compounded by additional uncertainties in the treatment of the land surface (see breakout #9). Since incoming shortwave is the largest energy source at the surface during the day, these error sources can grow upscale because they lead to uncertainties in surface energy fluxes, PBL evolution, and whether convection (of some form) occurs. Spectacular forecast busts still occur when models struggle to forecast clouds properly; e.g., failure to predict the clearing of overcast conditions earlier in the day that led to the outbreak of convection and vice versa. Upcoming FV3-CAM plans will allow intercomparison of candidate physics suites at 3km, including the NAM physics (currently used with the operational 3km NAM-nest), the RAP/HRRR suite (currently used with the operational 3km HRRR and also being evaluated for global NGGPS application), the GFDL suite (currently in spring 2017 3km testing), and the GFS global suites (current GFS and advanced NGGPS).

An EMC decision was made after the workshop that EMC would first evaluate the advanced NGGPS physics for suitability in globally nested 3-km FV3 runs, but that other physics options could be considered depending on what is found from that testing.   It is not yet clear if a single scale-aware physics suite could be applied in all NGGPS applications in all global subdomains but agreed this would be desirable.

Breakout #6 with the Data Assimilation (DA) WG.

Discussion was centered on information exchange on the effective development efforts for global data assimilation using GSI and FV3 along with the storm-scale ensemble DA efforts by GSD, NSSL and NCAR. Some consideration was expressed from both WGs to consider multi-scale DA approaches to merge larger-scale and convective-scale treatment.

Effective storm-scale DA will require development with respect to a single FV3-based system. The Canadians apparently have used different physics in their DA (details not specified), but the preference among the experts was to also unify around a single physics package. The desire to unify around a single core and single set of physics is to avoid violating some fundamental assumptions made in the DA algorithms (e.g. Gaussianity, projections of biases from one core/scheme onto another, etc.). For systems that would involve frequent storm-scale DA (e.g. from a network of radars and [eventually] GOES radiances), ongoing runs always resident in memory are considered to be necessary (also discussed in the early April JEDI workshop at NCEP). At the same time, future DA design must ensure fault tolerance for this continuous DA process to be viable in operations. Further discussion emphasized the importance of addressing outstanding research questions regarding multiscale data assimilation. Simultaneous assimilation of broadly distributed observations, such as rawinsonde data, alongside comparatively dense observations, such as radar data, presents practical challenges owing to these networks' abilities to resolve different scales. Future DA algorithms will need to address this challenge.

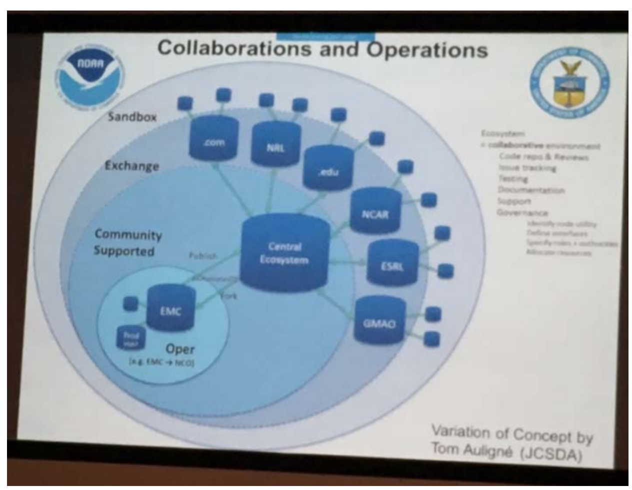
<u>Breakout #7 in with the Governance WG.</u>

One possibility for governance being considered is that used for the US Community Earth System Model (CESM). CESM already shares codes with DOE's Accelerated Climate Modeling for Energy (ACME). Of course, codes and scripts must be fully documented, adhere to minimum coding standards and performance requirements, and follow "best practices" (all TBD, possibly borrowing from CESM). Multiple configuration options (including dynamic core) exist in CESM as do WGs that determine evolution of the official version. Discussion also centered (again) around the issues of running FV3 in a standalone limited-area domain and for allowing some in the Meso/CAM WG to conduct additional idealized testing.

GFDL's comments: The idealized tests are best done by the dynamical core developers for the purpose of validating/improving the core algorithms. There are certainly educational purposes for the university community. But that will be lower priority at this early stage. The working group that is suitable to do idealized tests is the dynamics and the nesting group, not the CAM group. To evaluate the overall performance, a model is best evaluated in real-world situations. For this purpose, EMC, SPC, and NHC are the true and more neutral stakeholders/evaluators.

<u>Breakout #8 CC#3 with the Infrastructure WG.</u>

As expressed by the Infrastructure WG, a central repository would be established that is linked to other repositories and coordinated with Git revision software and located via vLab or GitHub or some other venue. Candidate repositories must be open, include clear documentation, and undergo regular regression tests. Operational repository(ies) would be kept separate from the community repositories and maintained by EMC as suggested earlier in the unified modeling meeting (see figure). There was interest in a unified workflow that is core-independent and modular with switches to run with varying levels of sophistication (e.g. atmospheric runs with or without ocean coupling).



Breakout #9 with the Land Surface/Hydrology WG.

Some of the discussion focused on establishing a more unified land-surface modeling effort between the CESM LSM, the Noah (GFS, NAM), RUC-Smirnova (RAP/HRRR), and NoahMP (National Water Model, NWM) land-surface models (LSM). Discussion was held about the NASA GMAO LIS (Land Information System), now using an ensemble LSM approach including the Noah and RUC LSMs and others. While the more sophisticated treatment within the NoahMP is used in the NWM, there was general agreement that development efforts in Noah and RUC-Smirnova should merge. The 9-layer RUC-Smirnova LSM was designed to do a better job capturing the diurnal cycle, and it was agreed that the coarser 4-layer Noah model should be expanded to include more layers. Including more soil layers has a minor impact on run times because none of the fields are advected horizontally. The biggest limitation on the number of soil layers in LSMs without an implicit solution is to prevent temperature oscillations from developing when the time step is too large and/or the soil layers near the surface are too thin. Development of an implicit solution LSM will be important in the future merger. Land-surface data assimilation methods were discussed, including that used at NOAA GMAO and the atmosphere-soil relationship in GSI used by RAP/HRRR.

Annex 10. Physics Working Group

SIP Working Group (WG) on Physics: Workshop Feedback

1. <u>Brief assessment of the workshop</u>: for both the first half (community engagement) and second half (SIP X-WG meetings) of the workshop, what worked best, what didn't, what could be improved, and most important…what it might mean for any follow-up meetings we may have in the future.

Overall the workshop was very successful in bringing a diverse cross-section of the community to discuss issues related to research and operational NWP. Over the first 1.5 days of the workshop, it was very encouraging to hear of NCEP's commitment to a community-based modeling system, as well as to an evidence-based decision making system. The breakout groups generally arrived at similar conclusions, which helped to identify the key next steps and gaps for this community effort. However, the discussion was mostly at a high level, and perhaps at the next meeting the discussion can be centered more of the key substantive issues.

The second part of the workshop, in which the WGs met with each other, was very useful, but at times overwhelming and long. While it would be valuable to repeat this type of workshop in a year, it is recommended that the topics for discussion be more guided. Most of the group discussions at times lacked a leader to better organize and moderate the discussion. Nonetheless, it was valuable to hear various perspectives on the topics discussed. In particular, the discussions about modernization of the software infrastructure and accessibility to code repositories was welcomed.

A remaining challenge is to distill the discussion into priorities, decisions, and actions.

2. <u>Top takeaways</u>: What were the most important things your WG took away from the workshop. This is open to whatever you want to flag, but there are a couple of categories I'd like to make sure you address:

a. <u>Immediate needs</u>: Coming into the workshop, I had hoped that for the most part that all WGs could work along the same schedule, and deliver your draft SIP inputs at roughly the same time. However it became abundantly clear that there are some things that need to get done much sooner in order for the rest of the group to proceed. Please flag any of those "immediate needs" that your WG feels are important, either from another WG(s), or from your own WG.

- *Solidify a code management and governance process for advancing the physics suite.* NCEP is working toward an advanced physics suite for implementation with FV3. The suite and its parameterizations will be continuously upgraded over the years. It is important to establish a transparent and evidence-based process for proposing, testing, and reviewing physics that can be followed by all developers (from EMC, GFDL, and the general community). GMTB has proposed the CCPP concept and roadmap as a mechanism for advancing physics. This concept needs to be reviewed, modified as needed, and adopted. Failure to do so will perpetuate the perception that physics development at NCEP is not an open process.
- *Establish governance for the overall unified modeling effort.* It is important to have a group that

can establish vision, prioritize the distributed development, and assist the various WGs and participating institutions in aligning their activities.

- *Coordinate the physics development and testing with the overall unified model development*. Interactions are needed between the physics group and many of the other working groups to establish timelines, objectives, and possible impacts.
- *Establish unified metrics and scorecards* that address short-range convective scales to sub-seasonal and seasonal scales, as well as metrics for coupled applications. Establish testing procedures to conduct evidence-based, transparent assessment of physical parameterizations and suites.

b. Items on the "critical path": Those items to be worked that are the most critical to achieving your WG goals and creating your portion of the draft SIP. It would also be nice to know which ones were identified as a consequence of the meeting vs. those you already knew beforehand, but that is not required…just a 'nice to have' if you are able to do so.

- *Code management and governance for FV3.* We welcome the FV3 release planned for mid-May. Going forward, it is necessary to identify a code management and governance system that allows the community to use and contribute to the software.
- *User and developer support.* At this time it is not clear how user and developer support for the unified model will be done. It is likely that various groups (CESM, GMTB, etc.) will play a role, and this needs to be clarified.
- *Review requirements for the physics-dynamics interface.* In recent discussions, additional requirements for the physics-dynamics interface have been suggested, which may require development and layers beyond the current Interoperable Physics Driver. It is important to review, prioritize, and set forth a community vision for this interface.
- *Create procedures for documenting physics and its interfaces*. At this moment, there is no agreed process or guidelines in place for the documentation of physical parameterizations. The result is that the documentation of the GFS 2016 suite by GMTB will age off. It is important to put in place procedures to ascertain that new code committed to trunk/master is properly documented.
- *Understand overall and individual component development milestones and schedules.* Physics suite development requires knowledge of near-term deadlines and longer term goals, and physics group needs to be aware of other software and model development-related milestones.
- *Prioritize the physics development.* Priorities need to be established for the physics development including key physics components, testing procedures, and metrics.
- *Consider HPC requirements.* It is critical for the operational centers and community testbeds to have sufficient computational resources including storage, ease of access, and documentation to meet the demands of full testing and evaluation of the physics in uncoupled and coupled applications.

Annex 11. Post-processing Working Group

Shared Post-processing Takeaways from NGGPS Workshop 201704
Contributing: Matt Peroutka, Melissa Ou, Zoltan Toth

Assessment of the first half (community workshop)

- The organizers tried to avoid "Death by Powerpoint." It's not clear that they succeeded. Strict enforcement of a time limit for presentations would have helped.
- With the exception of one memorable incident, audio/visual support for meeting seemed to work well.
- Breakout sessions suffered from NOAA/non-NOAA segregation.

Assessment of the second half (X-WG meetings)

- Very interesting discussions with other WGs. All involved seemed to learn a great deal.
- Some disappointment that remote WG members could not participate in second half.

Top takeaways; immediate needs

- Are we forming one NGGPS community served by 13 Working Groups? Or are we forming 13 communities?
- WGs would benefit from additional guidance on how to collaborate and make decisions. We are aware of tools out there that may help (VLab, Google Groups, COG). No established solutions to support WG work.
- Communication about overarching issues (GitHub, Licensing, possible collaboration with NOAA Big Data Project) will be important.
- NGGPS may need to focus more on data formatting and metadata.

Top takeaways; not-so-immediate

- Post-processing NWP output is one of the most important ways that America's Weather Industry (AWI) differentiates themselves.
- Many people expressed concerns for NOAA's HPC infrastructure.
- Parallel runs that support NCEP Production Suite (NPS) are routinely shared with the entire weather enterprise. Retrospective runs that support (NPS) are not routinely shared outside of WCOSS.
- Interesting input from Communications and Outreach WG meeting
    - A communications plan will be a very important part of our efforts to form an NGGPS community.
        - It needs to support integrated decision making.
        - If we have a community, we need a way to communicate decisions to the community.
        - If we have a community, we need a way to listen to communications from the community.
    - Advocacy and outreach are important.
    - Training and education are also important.
    - It's important for NOAA to get our planning into the open.

Didn't come up at the workshop, but WG raised them

- At what point do we infuse additional science and post-processing techniques?
- De-bias model output
- Diagnose weather-dependent variables from model output
- Combine output (model and observations) from various sources for better post-processing
- Develop an infrastructure to share data (users who want to do their own post-processing -- ”private sector in the middle of the supply chain”) as well as to continue to make decision support products to support other users. A challenge will be organizing the community to move forward with both tasks.

Summary Reports from X-WG meetings

Post-processing and Governance WG

- COG is an option for Community Governance, communication, and non-software collaboration.
- No established solution to support governance, communication, and non-software collaboration for all of NGGPS.
- Might be useful if Governance WG addressed a handful of use cases where governance will be challenged.
- Boundaries of current WGs should be same as old NGGPS groups.

Post-processing and Infrastructure WG

- Infrastructure has not considered data storage and metadata standards to date. This may be something for them to consider.
- “Rules of Engagement” for GitHub are under development.
- Licensing statement for software is under development.
- Unclear how governance processes for NGGPS community can or should interact with NWS governance processes.

Post-processing and Meso/CAM WG

- Interesting discussion, but no notes.

Post-processing and Ensembles WG

- Substantial, broad-based concern for effective storage of EPS output. Common examples are temporal and spatial resolution of output (e.g., hourly and 0.25 degree output from GFS).
- Cloud computing and NOAA Big Data Project are obvious options that seem to be worth pursuing.
- Some discussion of lossy compression without much support.
- Some discussion of NOAA-hosted computing that might be leased to non-NOAA members of the weather enterprise.

Post-processing and Verification/Validation WG

- Overlaps between two WGs seem obvious. We will plan to conduct regular meetings.
- Going forward, we will need to define boundaries between and integration points for two groups.

Annex 12. System Architecture Working Group

Strategic Implementation Plan Workshop Comments (18-20 April 2017)
System Architecture Working Group

Following the Strategic Implementation Plan (SIP) Workshop hosted by NCEP on 18-20 April 2017, the various SIP Working Groups were asked to provide feedback, responding to two questions. The questions are below, with responses from the System Architecture Working Group (SAWG), based on a group discussion on 28 April 2017.

1. Brief assessment of the workshop : for both the first half (community engagement) and second half (SIP X-WG meetings) of the workshop, what worked best, what didn't, what could be improved, and most important…what it might mean for any follow-up meetings we may have in the future.

The workshop – both halves – was generally very successful. The moderated discussions served the purpose of getting all the WGs up to speed on each other's charge and progress and set the stage for the later cross-WG discussions. Everybody is pleased with the new openness and transparency, commitment to community engagement, evidence-based decision-making, and commitment to modern software engineering practice. The breakouts in the first half all came to very similar conclusions, which might mean that they were unnecessary but more likely means that there was value in giving people an opportunity to discover their common views in separate venues.

The cross-WG discussions were effective in some cases but not in others. That was a weak function of the timing – people got burned out toward the end – but it was a stronger function of the maturity of the groups. Those groups that had been working on their issues for a while were better prepared to interact with the other groups. For the groups that formed only recently, the discussions were useful in providing them with context and an understanding of the relationship among the working groups, but the level of engagement was limited. There was ambiguity about who was in charge of each of the cross-WG meetings, which led to awkwardness in how to organize the discussion. It would have been better to have a more concrete set of questions defined in advance with more detail than the "headline" questions that were circulated, and a moderator or chair for each session so that the discussions could be more focused. The result was that there was more discussion of conceptual aspects and less discussion of specifics and practical considerations.

Despite the general agreement on the high level issues from the first breakouts, there are conflicting opinions and diverging plans within individual working groups and among the different working groups. "How is this going to work?" was a common refrain among participants. To retain the confidence and goodwill of community participants it will be necessary for an authoritative body (e.g. steering committee) to emerge soon that can resolve conflicts and formulate and communicate integrated plans. There is a need for an annual meeting schedule, including more frequent subject-specific WG meetings. In particular, another workshop to discuss concrete ways forward is needed in less than one year. Virtual means of more frequent, regular coordination can also be explored through the new Communication WG.

2. Top takeaways : What were the most important things your WG took away from the workshop. This is open to whatever you want to flag, but there are a couple of

categories I'd like to make sure you address:
a. **Immediate needs**: Coming into the workshop, I had hoped that for the most part that all WGs could work along the same schedule, and deliver your draft SIP inputs at roughly the same time. However it became abundantly clear that there are some things that need to get done much sooner in order for the rest of the group to proceed. Please flag any of those "immediate needs" that your WG feels are important, either from another WG(s), or from your own WG.

- *Introduce overall authority and coordination*. There is a need for an overall authority that can resolve conflicting requirements, strategies, and goals, and develop integrated plans. This is important to the system architecture because unified modeling system design and development require communication of an overall vision and an understanding of priorities and timelines.
- *Coordinate system architecture definition and FV3 development*. Additional interaction is needed between the System Architecture and Dynamics working groups to discuss goals, timelines, and approaches (e.g. roles and capabilities of ESMF/NUOPC and FMS) related to integrating FV3 into the unified modeling system. For example, the implementation of nesting and a standalone regional atmospheric component model employing FV3 have system architecture implications.
- *Define land surface model disposition*. There is a need to supplement existing land surface requirements in areas such as implicit/explicit coupling and interface with PBL, and in the longer term, river/estuary/lake models, aerosols/atmospheric composition, dust and GHG emissions.

b. **Items on the "critical path"**: Those items to be worked that are the most critical to achieving your WG goals and creating your portion of the draft SIP. It would also be nice to know which ones were identified as a consequence of the meeting vs. those you already knew beforehand, but that is not required…just a 'nice to have' if you are able to do so.
The items under "immediate needs," plus (not necessarily in priority order):

- *Hire a manager responsible for unified modeling system software development.* (Already known issue) In the absence of this function and an associated team, there is little likelihood of being able to complete the implementation of a unified modeling system architecture and operate it.
- *Identify a developer to integrate FV3 into coupled model configurations.* (New) In the absence of someone in this role, the path is unclear for implementing FV3 in the unified modeling system in a manner consistent with initial SAWG recommendations. More generally, a component lead is needed at EMC for the coupler.
- *Understand current development milestones and schedules.* (Already known issue) Evolution of the system architecture requires balancing near-term deadlines with longer term goals, but the SAWG has not yet mapped its recommendations to known development milestones.
- *Complete investigations in the SAWG's initial recommendations* (Already known issue). Assess feasibility of replicating the GFDL coupling strategy with NEMS, and possibility of partnering with CESM to develop a community coupler that meets EMC needs.
- *Add nesting and grid refinement to topics addressed by the SAWG.* (New) Options for the implementation of nested systems should be included in SAWG considerations of the system architecture. There may also be a requirement for a standalone regional atmospheric component.
- *Add HPC to topics addressed by the SAWG, in coordination with Infrastructure WG*. (Already known issue) HPC needs to be explicitly addressed in the SIP. The SAWG, in coordination with other working groups (esp. Infrastructure, Physics, and Data Assimilation), needs to consider engineering (including best practices, standards), portability, and performance optimization and

scalability of the community prediction system on current and emerging HPC systems (including processors, networks, I/O, and data management).
- *Advance a common view of the IPD physics interface* (Already known issue). There is not yet a consensus about requirements, scope and design of the physics interface, especially the interface to chemistry/aerosols.
- *Document requirements for Data Assimilation applications*. (Already known issue) Recommendations are needed for a system architecture that will best support the data assimilation application. Example questions: What are the implications of weakly and strongly coupled DA? Should the architecture support Pause/Resume? Should the architecture support IAU (incremental analysis update) and Replay?
- *Address conflicting requirements for Ensembles*. (Already known issue) Existing ensemble requirements contain conflicting opinions about the level of communication needed between ensemble members – an authority outside of the SAWG needs to resolve conflicts and finalize requirements. The system architecture is impacted by the tradeoff between model integration performance (fewer simulations) and data management performance (many simulations).
- *Prioritize the scientific research agenda*. (New) Many questions with a bearing on system architecture require scientific research, with the answers relating either to the representation of Earth system processes and their interactions or to the impact on predictability and prediction skill (as a function of lead time); e.g., intra- and inter-component interactions (aerosols in 3D interface; atmospheric columns shading each other at high resolution; coupling ocean and sea ice as a "fast" process; lateral water movement at and below the land surface; etc.). All have a bearing on R2O and O2R (support). While a maximally flexible system architecture is desirable to be able to adapt to different scientific outcomes,, that must be balanced by practical considerations.

<u>Gaps that go beyond System Architecture:</u>
- *Articulation of vision*. What are the expected outcomes and criteria for success at the 2-yr, 5-yr and 10-yr marks? N.B.: This may be an immediate need since SIP can't be developed completely in isolation from long range strategic plan; decisions made in the short term may be hard to reverse, even if they are antithetical to longer range vision.
- *Refinement of "community" definition*. Better define and articulate how different parts of the community interact with EMC and each other, and how decisions will be made on behalf of the community. Establish a scientific steering committee. acknowledge and resolve the tension between inclusiveness and manageability.
- *Better delineation of community responsibilities and plans for community support*. Plan for governance: goals, priorities, timelines (including accumulation points for changes). How is support for community developers going to happen?
- *High - level framework for code pathway*. Diagrams showing code repositories are good, but need details about how code moves between the shells, rules for submitting, releasing, documenting, testing and evaluating (hierarchy of testing and criteria for advancement) plus plan for training and support.
- *Definition of resource challenges*. Consolidation of systems, inter-agency and inter-institutional exchanges, access to critically important resources.
- *Involvement of NCO*. Decisions involving code repositories and R20 transition require willing involvement and participation by NCO.

Involvement of CPC. Decisions about design and implementation of extended prediction codes would benefit from participation by CPC leads.

•

Annex 13: Verification and Validation (V&V) Working Group

Community Modeling Workshop and SIP Working Group Meeting

1. Assessment of workshop:

The workshop was overall effective, with much useful collaboration and information.

Community Engagement:

The fact that the community was brought into the discussion was important and critical to the process. The entire community, including NOAA staff, seemed to be supportive of the new openness being exhibited by EMC. There were two sentiments expressed within the V&V working group. The external community felt that the overview of each working group provided an excellent introduction to help all better understand each other's perspectives and how to move forward. Additionally, the community found the NOAA partners to be refreshingly open and honest and as a result felt they were actually engaged.

However, some participants closer to EMC and NOAA labs remarked that the first day was somewhat repetitive, and while the WG report out schedule looked like it was formatted in a way to allow for a reasonable amount of discussion, it didn't achieve its mission. Several felt that this session with reports from the different working groups were nowhere near as interactive as they needed to be and that was likely due to the reports often took most of the session, leaving little time for follow-up.

The Breakout Session was very useful, with some common themes clearly emerging. Meeting room limitations dictated that NOAA staff clustered into break-out sessions separate from the external community. Initially, this seemed like it would be counter-productive, but after the session, it seemed like it allowed the external community to share their ideas independently. The fact that common themes seemed to prevail throughout the Breakout Group reports was very encouraging.

Ultimately, the community was left at the end of the first day and a half looking for more feedback from NOAA on how to proceed and how to engage. Outstanding questions include:

1. What are the roles of different community participants to make this interaction productive?
2. What are the metrics of success for the community to determine if efforts are being constructive (e.g. Is the community trying to help NGGPS meet schedule? Or to help EMC grow and be successful? What does NOAA want to get out of this?)?
3. How can the community contribute immediately?
4. How can the community participate in model development and the community repository?
5. What compute and monetary resources will be available to allow the community to participate?
6. What follow-on interactions should the community expect?

SIP X-WG Meetings:

The cross-working group meetings were overall very useful. Some of the meetings followed expected lines of discussion and others were unexpectedly very useful. Many of the items listed in the Immediate Needs and Critical Issues section were ideas that came forward from the X-WG meetings. The meetings seemed to be small enough for all to participate but large enough to draw in several perspectives and much of the groups were committed to the end. It was somewhat surprising how different the groups were both in how much prep-work had been done as well as their responses to our questions about what services we can provide. Some felt they didn't need any help; others wanted to make sure V&V wasn't re-inventing the wheel; and others were very appreciative and interested in partnering.

Constructive comments about the X-WG meetings include:

- Meeting with each working group was very helpful to V&V
- Breaks needed to be built into the schedule. Serious burnout was occurring by Thursday afternoon for groups that never got to sit out one session.
- Putting more than two groups in a room together was not a great idea. Discussions necessarily would go into the topics that only two of the groups want to discuss, and the third would be left out for a while.
- The auditorium was not very conducive for these small group sessions. The set-up made it feel more like an interrogation than a conversation. Also, the three groups many times had different needs and points to make
- The group felt that a plenary session to end the meeting would have been helps wrap up the workshop. It felt like there was no closure

2A. Top Takeaways from the Workshop - Immediate Needs:

Clear Guidelines for Community Interaction: There was a resounding sentiment that clear guidelines for community interaction, including how to contribute to the community repositories is needed. Questions that need to be addressed include:

- How will the code be supported out to the community (O2R)?
- Who will provide helpdesk capability (Community Support)?
- How much testing is required prior to committing an innovation to the community repository (governance)?
- How to lower the bar for the community to do testing (testing framework)?
- How will computing be provided to help with development and testing (testing framework)?
- What are the benchmarks of success, upgrading the operational model and accepting innovations into the repositories (governance)?

Obs Processing: It became extremely clear that observation processing must be accounted for somewhere within the working groups. There is an increasing need for diverse data sets for verification and data assimilation activities from multiple working groups, and it seem inefficient to task the V & V or data assimilation groups with this task unless one is provided with the support to take on this effort. An additional working group should be formed to provide subject matter expertise on observations data types and processing.

2B. Top Takeaways from the Workshop - Critical Items:

Governance: It became clear that governance must help establish common standards and benchmarks for verification/evaluation. And defining the standards/metrics before the evaluation of any parallel system is critical. Without this, groups are going to do their own verification on whatever they want to, and the results will be cherry-picked. A recommendation is to consider including someone from V&V on the Governance WG. Betsy Weatherhead, CU Boulder, would be a good candidate.

Facilities: There are multiple components that need to be in place to make this work, including how to share data, access to observations; access computing and a funding avenue. Of these, computing is likely the most critical to start addressing immediately.

Communication and Outreach: To facilitate a clear understanding of how the community can engage with the unified modeling effort, an active Communication and Outreach WG is needed. The idea of special sessions at AMS and AGU is supported by the V&V working group. Hui Shao attended the meeting at the workshop. Other V&V members who are interested in participating include Jason Otkin (U. Wisc Madison) and Betsy Weatherhead (CU Boulder).

A summary of what Hui took from the meeting includes:

- More clear information is needed about SIP scope, who is the community, what the roles are, who is the decision maker
- Requirements need to be clearly communicated - e.g. right now there is no clear signal on what's the path forward
- Set up a mailing list and web-page to keep people aware of the progress and provide community feedback in a regular way
- Sustained and centralized support and effort needed – e.g. training and education
- Communication not only needs to be outward but also inward into EMC so current internal staff know what is going on – the actual developers need to be more involved – not just managers
- Advocating this effort beyond the current community of professional scientist is very important. This group should also reach out to students to help change the culture

Verification and System Architecture: There is a need to make sure V&V and System Architecture communicate on a routine basis to make sure the development of these two core components of the unified system are well aligned.

Verification and Post-Processing: There is a critical link between post-processed fields and how to evaluate them. The use of common output formats is critical to the success of the unified system and V&V and Post-Processing need to help define that. It is recommended these to WGs develop a mechanism for meeting routinely.

Verification and DA: There is a critical link between the observation database, forward operators, and diagnostic tools used by DA and what is needed by Verification and Validation. It is recommended these to WGs develop a mechanism for meeting routinely, including having an V&V member integrated into the JEDI planning. John Halley Gotway (NCAR and DTC) would be a good candidate.

Definition of MET+: To take advantage of some of the pre-existing packages out in the community, especially for some of the component areas of research (land surface, hydro, etc…), it is good to start thinking of the definition of MET+ to be broader than the core MET and METViewer components embedded in a Python framework. Making sure other packages can interact with MET but also be checked out separately should be a longer range goal of the SIP.

Diagnostics at Every Timestep: Dynamics and other WGs expressed a desire to be able to evaluate/diagnose some fields every timestep (e.g. fluxes and closing the energy budget). This seems to be out of the purview of a verification package that is intended to evaluate the final products. Upon discussion with the System Architecture group, they felt it might be better addressed by a combination of their group and the CCPP overseen by the Physics group.

Native Vertical Coordinates: While most verification packages, including MET+, currently operate on standard vertical levels, there is a strong need to also have capability to perform verification on native vertical levels to facilitate model development. Additionally, evaluation of some fields (e.g., brightness temperature) are best performed on native model levels rather than pressure levels.

Process-Oriented Evaluation: This topic ranges from diagnosing fluxes every time-step, to evaluating physics schemes with field project data and remotely sensed observations, to diagnosing general circulation signal such as the MJO Index. Better definition of how a unified verification package needs to support these wide ranging needs, including prioritization, should be included in the SIP.

3. Report

The Environmental Modeling Center (EMC) Strategic Implementation Plan (SIP) Verification and Validation (V&V) Working Group was expanded from the Next Generation Global Prediction System (NGGPS) V&V Team. The SIP V&V team includes thirty members, all of which participated in at least one teleconference. Seven members are from the university community and several members from the Naval Research Lab (NRL), National Center for Atmospheric Research (NCAR) and Developmental Testbed Center. There are at least one member from each of the National Center for Environmental Prediction (NCEP) Centers and several members from many of the NOAA Labs. The goal of the V&V team is to provide guidance on how to provide efficient and robust community verification package that addresses the following needs:

- Works on varying computer architectures and complexity
- Supports evaluation of the coupled system
- Available via GitHub or VLab with community access
- Provides friendly interface and extensive user training, outreach
- Provides access to allow community contribution
- Must interface seamlessly with JEDI Obs Database and Forward Operator but also be able to support stand-alone obs files.

By providing a unified community verification package, a goal of the SIP will be to make Research to Operations (R2O) and Operations to Research (O2R) easily accessible to external community to perform experiments in an efficient manner. The statistics and scores computed using a unified verification package will be transferrable across institutions and allow future evidence based decisions about the Unified Modeling Package to be reliable and reproducible.

Key Issues identified by the working groups include communication, research, new methods and technical needs, barriers and risks.

- Communication and Research
    - Requirements of each community (i.e. operational centers vs. academia) should be expanded beyond what was done by the NGGPS V&V team
    - More research on “optimal methods” to evaluate unified model at different time and spatial scales is needed
    - Greater emphasis is required on easily using remotely-sensed observations and incorporating obs/analysis/hindcast uncertainty
- Tool should incorporate methods to help
    - Developers better understand their models and ensembles (process-oriented)
    - Forecasters gain confidence in new products, especially probabilistic (regime dependent, extremes)
    - Managers decide what to implement (scorecards and NWP indices) \

  - Support determining the impact of errors on end-users (societal impacts)
- Technical Needs, Barriers, Risks
  - Must run with or without a workflow manager and on single or multiple processors
  - Computational efficiency to run operationally and on large datasets
  - Flexibility to also run in research labs, academia and other orgs
  - Breadth of scope to serve the needs of many groups may make tools unwieldy

In the cross-working group meetings, many ideas were brought forward for the V&V working group to consider. A common theme amongst the meetings was a need to define benchmarks and determine a mechanism to make it easy for the community to test against them. Additionally, it became clear through these meetings that the diversity of observations needed for verification and the need for expertise beyond the V&V and DA groups. The meetings served as an avenue for the V&V working group to identify requirements not already covered by the documentation produced by the NGGPS V&V team (https://drive.google.com/a/noaa.gov/file/d/0BwjqxMjULlDVU0wMkJtSnl5SFE/view?usp=sharing). The following is a sampling of salient points:

- Some scores and metrics need to be focused on Convection Allowing Model (CAM) resolution. The CAM working group provided their initial ideas as a starting point
- How does one determine what is a meaningful spread and assess bimodal solutions that occur from multi-model ensembles
- There should be guidelines on the discrimination between statistical significance and practical significance
- Closing the water or energy budget is considered to be verification that the model is working properly
- V&V needs a broader scope of observations that are used in DA therefore JEDI may need to extend beyond their requirements
- Evaluation of hurricane tracks and intensity, precipitation, conservation properties, field decomposition, and energy spectra all lead to validating the dynamics are working properly
- Verification of processes such as sudden stratospheric warmings, Quasi Biennieal Oscillation (QBO); Madden Julien Oscillation (MJO) and electromagnetic fields are also critical to evaluating if the dynamics are handled properly.
- Satellite fields may be used to evaluation stratospheric gravity waves and tides as well as fields pertinent to space weather (i.e. upper atmosphere temperature errors and ionosphere fields)
- The verification package also needs to be able to provide statistics for extreme events

Through the discussions the V&V working group identified areas that are likely not within the purview of the unified verification package at this time, including replicating the forward operators to convert model output resemble observations (likely better handled in DA); performing the checks required to close the water and energy budget in a coupled system (likely better handled in Physics and System Architecture); and replication of fully developed verification packages such as that for the Land Surface Modeling, Hydrology, and Marine groups.

Most of the group discussions included some mention of benchmarks and the need for them. Have a suite of benchmarks would allow evidence based decisions to be made without cherry-picking cases or examples. The benchmarks should be formulated with guidance of the other working groups and the Governance group and should include the current operational capability. It should include a suite of tests including all four seasons but also select case studies to demonstrate potential improvement. Special attention should be payed to seasonal and sub-season benchmarks as well as those for weather. There was also a recommendation that forecasters should recommend some of the metrics. The benchmarks should be bundled with observation and analysis fields so the comparisons will be universal. There was also mention of the need for verification of input variables prior to post-processing.

Several groups include discussion on Community Code aspects of a unified verification package. There was a request to consider it as a place to share tools. Toward this end, the concept of broadening the definition of MET+ to be broader than the core MET and METViewer components embedded in a Python framework. Making sure other packages can interact with MET but also be checked out separately was an idea that was discussed several time. In discussions with the Infrastructure group, the requirements to be considered an authoritative repository for the unified system were stated as: Code, Governance, Open to community, Documentation, and Regression testing. MET+ already has all of these except governance. Setting that up will need to be a focus over the next year or two. There was also discussions with several WGs regarding conforming to data standard throughout the system, specifically NetCDF4 and Grib2.

X-WG meeting notes (Tara Jensen)

Ensembles and Meso/CAM

Impact of ensembles and V&V on final product

What is needed for global ensembles and what

How will global ensemble merge with meso/regional

When V&V metrics for ocean; da products – using insitu obs

CAM VX and ensembles

Score of V&V is very broad – **what type of metrics**

How can forecasters can use diag and scores in V&V

Best ways to verify ensembles

Getting good spread

CAMs transitions to R2O

V&V of probabilistic and ensembles

Verify CAM ensembles – first and second moment

Commonalities and differences – global, regional and CAM ensembles

- Stochastic parameterizations – commonality
- Error growth characteristics – SKEBS and SPPT don’t work as well on CAM
- GFS – how much spread can get for aviation purposes
- I.e. if you’re tuning for day 1 spread will likely result in too much in 6day and if you can get good

6-7day spread… not enough in 1 day

- Best practices:
- Land surface perturbation is key
- What does a global system with a regional/cam look like from probabilistic – is it something we could consider

Verification

Overlap on verification – scorecard; upper air VX; precip on shorter time scale; heavy p

Updraft helicity; impact based metrics

Clouds are a common verification need

Pre-convective environment – GFS currently doing poorly on the pre-convective environment

How do you determining what is meaningful spread

How do they compare to a compare new models – baselines and

CAM working group –

Application point of view – move toward probabilistic – solar irradiance and low-level winds – probabilistic distribution/dispersion

Unified post-processing system – establish a baseline is standardized – what we're running in operation should be a benchmark;

Uncertainty vs. spread; when there's a 1-1 spread skill – doesn't tell us if there's reliability and discrimintation; more sophisticated tools, bin spread – what was the expected error of the mean – Walter K.s

Sometimes bimodal solutions – needs tools to diagnose and determine which way to go…

Clusters of clear and cloud; depends on how binary the fields are

Value form of validation, forecaster point of view;

Global Experiment; have model developers sit with forecasters

Subjective eval is also important

What is statistically significant

What role should forecasters play in validation

Processing of obs/analysis for comparing against –

What do you verify against – best practices, Obs/analysis uncertainty

How do you treat extreme values and evaluate them.

Components Models

Land – focus on where people live – not enough emphasis on near surface metrics; land outputs

Water and energy "budgets" – take the variables and perform the calculation; where does this overlap with post-processing

Bias – looking for model drift

Closing the budget is a "verification" – infrastructure and architecture

Diagnosing the fluxes every time step -

Sources of aerosol

Change over time period – what is coming in is going out

When DA is used – track increments and include them in the budgets

Obs that are needed: More ocean drifters; sst; vertical cross sections; profiles; derived products; transports

Whatever metrics are important to them – these are the metrics we want to deliver on

Need to verify

Controlling the drift of the mean state;

Wheeler – kolotis diagrams; MJO – modes of variability

Benchmarks – MME of season and sub-seasonal forecasts… is NOAA better or the same

ECMWF – weather to S2S is benchmark; ensemble of score;

Aerosols and Atmos. Comp/Air quality – Point obs at surface – Fraction Correct vs. thresholds;

Comparisons with satellite and aeronet obs;

VIIRS - BUFR

Validation – set of forecasters that provides feedback

How do collaboration work;

How do we pull these together other packages already exist – why rewrite them;

Dynamics/Nesting

Hurricane track and intensity and intensity change– driven by dynamics – 34,50, 64

Not validating fields;

ACC - if good;

Hurricanes; precip; **process-oriented metrics – conservation properties; energy spectrum –**

Moving nests –

R20 –

V and V products – for general users

Night-time precip in OK – due to MCS – propogation

Rossby Waves

Fourier transform

Benchmarks – forecaster metrics?

Storm total rainfall swath; some say more physics

Combo of bulk statistics and case studies;

Find something in the model and search the observations to see if it exists

How to handle these big grids – need to handle 3km – diagnostic is part of model output

Separation of noise vs. signal – not just the uncertainty – partially predictable cases

How do we move in observational space; how do you analyze the biases from the model

- Snow depth;
- Who does the research about verification?
- 10 m downscaled winds – how do you evaluate this
- WMO standards
- Uncertainty estimates –
- Mean temperatures
- What is not there today – high resolution structure;
- Vertical Velocity
- Space weather – dynamics play a role – temperature errors; ionosphere and plasma-sphere

Process question: How do you evaluate – sudden stratospheric warmings. QBO; electromagnetic fields

PCA used for sudden stratospheric warming – use an AMIP type run

Process questions: Satellite fields to provide stratospheric gravity waves and tides;

How to “process” observations to match.

Extreme value statistics – longer record

<u>Governance</u>

- Need to define priorities
- What V&V needs to be done for R2Repo
- What test can be run to make sure users –
- Look at the scorecard they use – understand what they're using and make recommomendation
- Community version – model intercomparison; indiv researchers to create a metric
- **Provide a place to share tools**
- **Benchmarks – most things over all 4 seasons; episodic over at 10 cases – need to make recommendations**
    - Who sets benchmarks
    - What are they
    - Validation Test Report
    - Goes into V&V after the initial test report
- Need to avoid cherry-picking
- Validation is the heart and soul that makes it work
- Need to define terms – glossary
- V&V needs to guide Governance on tests that need to be completed
- How to test for reliability
- How do we handle sharing of data with community
- How much testing, how many "cases" are needed to move down the funnel
- Validating in a fully-coupled vs. validating stand-alone – when do you move something in and do a fully coupled

Tiers of testing – Governance wants input on what each looks like

Model development – a few cases – needs to be lightweight – how do others do this; how does CESM

R2repo – standardized tests

Operational system – multi-year testing

What algorithm to use for computing AC (WMO, EMC, ESRL, methods)

Obs needed for verification – get list of obs needed to Fred Carr

Meteorological realism – subjective eval

Look at work-breakdown for developing the testplan;

Strategic guidance and goals

<u>Post-processing</u>

NOAA's big data initiative – one success is radar data – the data are parked there

Post processing – barriers to entry are trivial; if sufficient model data got loaded onto cloud –

Amazon will capture the data for free if they can sell computing

AWC turbulence – icing

Renewables – taking raw data and making products

Get data in the format that people can use

Output fields that need to be verified before post-processing (inputs) and the verified after (outputs)

There's a risk that if we pick only a few

Three areas – ensembles are embedded in there

- **Model post – translate model variables into sensible**
- **Diagnostics – doesn't use training**
- **Statistical - Uses training**
    - **Blends**
- **Decision support and derived variables**

Morphing the idea of MET+ to be the package we're developing but also the way HWF

Need verification of input variables prior to post-processing – relevant to post-processing

Eddy dissipation rate – EDR – magnitude of turbulence

Base Verification – standard outputs of NWP

Application Verification – done by end-users

Need to separate operational capabilities from

Repository and **data standards**

- How popular it is; space saving and information intensive
- Metadata needs to be standard

Satellite data – where does this observation data sit between VX, PostProc, UFO

ICAO – wants WMO to define a more descriptive netcdf

System Architecture

User specified diagnostic list – performance and resource

As for the closed budgets – there's 1st and 2nd order closures; may want to see how

Certain fields need to – non-conservation of energy 0.1 of a watt – has to be computed for the entire coupled system

- What are the outputs – basic table of numbers – globally averaged
- Minimal set of numbers, coupler and whole system

Expose the info to the top – this seems like a system-architecture

because – output of prognostic variables and

pull out intermediate variables to compare to prognostic variables written out

There's other requirements to print out info at every time step – per time step wind – and fields for the water model

The concept of Diagnostic manager for all components but this is still also not the V&V group

Exemplar use cases

List of diagnostics

Data Assimilation

Reanalysis exists

DA goal is produce initial conditions – state estimates

Standardizing the datasets –

IODA is more an interface

- Obs that get into interface
- Obs that make it into data assimilation
- Obs that don't make it into the interface

Types of QC

- Sanity check
- Based on Model biases

Need both

QC can be an estimate of the error – can we have a larger dataset – VARQC builds a probability model for the error, through the minimization

V&V needs:

- **A broader scope of obs than used in DA**
- **Simulate model equivalent of obs**
- **Obs error estimation – QC**

We use where we can likely estimate the obs as accurately as possible.

RTMA/URMA currently using GSI – it will leverage JEDI

Will need to still be able to read in URMA, etc., and likely do some additional processing of grids and points to handle neighborhood probability

Something that's not part of verification but a useful for DA is interesting and vice versa

Assess confidence in observations – one single metric for fit-to-obs – what is JO

DA scorecard – super index

Forecast horizon, broken down by horizon, bias in NH, sign of analysis increments, spread-error

Verify the ensemble forecast – verify covariance

DA: Can provide uncertainty for gridded data

Explore FSOI with DA further

Physics

Verify parameters

Temperature tendency, moisture tendency, cloud moisture

Optical thickness, irradiance, increments from DA

It would be nice look at point and spatial averages and compare to increments

Soundings of species

What observations are available for these types of evaluations

In the physics group – there's an effort to produce an inventory of physics

Framework that we can provide to tendencies

Will CCPP test cases, benchmarks,

If you do an average over increments, and there's a pattern where there tends to be convection… then there's work to be done

Observationally based heating profiles

TRMM and GPM – independent source with heating profile data

With satellite analysis – will need to composite the forecast fields to compare

Need to compare to native grid – there's a intermediary product that is a lat/lon

MJO task force – in Fortran; Chidong will connect us with this tool

Infrastructure

Needed to be authoritative repository:

- Code
- Governance
- Open
- Documentation
- Regression

Make a recommendation - storage NETCDF v4 and GriB2 for delivery

Consider a way of plotting MET output with respect to baseline… need to update regularly

When 1 version is released, it needs to be released with scorecard/report card

Annex 13: Communications and Outreach Working Group

Communications & Outreach Working Group, Initial Meeting, 20 April 2017

This was the initial meeting of this Working Group, with largely ad hoc participants (group membership not yet finalized). As such, there were no X-WG meetings with other WGs.

Participants (other WGs):

- Ligia Bernardet, GSD/GMTB (Governance, Physics, Systems Architecture, Infrastructure)
- Bill Bua, UCAR/COMET (Floater; tent. Infrastructure)
- Susan Jasko, Social Scientist (Floater)
- Daryl Kleist, NWS-EMC (DA)
- Sarah Lu, SUNY-Albany (Aerosol/Atmos. Comp.)
- Matt Peroutka, NWS-STI (Post Processing)
- Ricky Rood, Univ. of Michigan (Governance)
- Tim Schneider, NWS-STI (Co-Chair, Communications/Outreach)
- Hui Shao, NCAR (DA, Verification)
- Hendrik Tolman, NWS-STI (Governance, Infrastructure, CAM)

A couple of goals for this working group were discussed:

3. To develop a *communications plan* to support *integrated decision making*. This plan should be strategic in nature with short- to long-term goals and include objectives with metrics.
4. To reach out to the community with multiple forms of communication (e.g. posting plans and keeping them updated.

As a foundation for moving forward with this group, issues of scope, and various characteristics and properties of communications and outreach were discussed:

- It should be a cultural vehicle for younger and more diverse inputs
- Training and education are within the scope of this activity
- It should communicate outbound decisions; receive inbound collegial inputs; and foster internal cross-communications
- Relationships are constructed and maintained through communication → TRUST
    - Consider mechanisms by which we communicate
- It was observed that NOAA's communication processes are "arcane and archaic" – the world is change, so we must change as well
- Sustained support from NOAA is needed to deal with security and access issues; facilitating R2O; repository training; and two-way follow-through
    - Connect/engage with John Ogren (in KC) on these issues
- Communication = power, therefore we need transparency and communication at multiple levels (these levels need to be identified)
    - ⇒ We're asking for cultural change
- We need to identify who the "opinion leaders" are in this community, and target them first and win them over

- We also have to help people to see how they fit into the big picture
- Recognize that EMC has to balance tight deadlines and mission focus with looking outward
- A question: do we have 1 community or 14 (WGs)?
- Lessons learned:
  - Adults want to learn quickly and apply immediately
  - MEG has been excellent (mix of developers and users)
  - Field staff (users) have strong opinions – so we need to work on *user buy-in* of new models
- Document motivation and rationale for the communication plan
  - Rules of order
  - Enforce meeting practices (e.g. publish minutes
  - Establish email lists
  - Membership lists
  - All info is public – all group-related (any WG) communications should be shared with all group members (no splinter/side conversations)
- Look at various models of communication