DTC Community UFS Test Plan and Metrics Workshop
Thank you to workshop organizers:

- Tara Jensen, NCAR and DTC
- Ligia Bernardet, CIRES, NOAA/GSD and DTC
- Jason Levit, NOAA/EMC
- Ivanka Stajner, NOAA/OSTI
- Ryan Torn, SUNY Albany
- Curtis Alexander, NOAA/GSD
- James Kinter, George Mason University
- Sherrie Morris, NOAA/OSTI
- DTC administrative staff

WebEx and Telecon: Mary Hart, NOAA/EMC
Presentation organization: Perry Shafran, NOAA/EMC
Who Attended

Community

- Universities (10 Institutions)
- Commercial (Zedx/X, ENSCO, BASF, Harris)
- ECMWF and Met Office (remotely)
- NCAR
- DTC Task Leads

Federal Government

- NOAA/NWS/NCEP/EMC, WPC, SPC, CPC
- NOAA/OAR/ESRL, MDL, NSSL, AOML, ARL
- NOAA/NWS/OSTI, AFS
- NOAA/NESDIS/STAR

~100 participants
60-70 actively engaged

- ESPC
- Air Force, Navy
- NASA GMAO, CCMC
Monday
Workshop Expectations and Example Test Plans
Choosing and Using Metrics
Verifying Weather Scale Applications
3:30 - Breakout Session #1 – Drafting Test Plans
  – Physics
  – CAM/Regional I
  – CAM/Regional II
  – Marine Hycom and CICE
  – S2S Ensemble
Tuesday
Verifying Medium Range and S2S Applications
Verifying Coupled Applications
Breakout Session #2 – Defining Metrics
  – CAM/Regional Weather
  – Global Weather
  – S2S and Seasonal
  – Process Oriented for All Scales
  – Atmospheric Chemistry and Composition
Agenda

Tuesday
3:30 - Breakout session #3 – Building Guidelines for Hierarchical Model Testing
  – Atmospheric Physics
  – Tropical Cyclones
  – Coupled Systems I
  – Coupled Systems II

Wednesday
9:30 – Panel Discussion Report Out to Workshop Participants & UFS SIP Members
Break-out Groups
Test Plans

- Discussed
  - CAM: Transition from HREF, SAR, RRFS
  - Physics: Advanced physics suite for the FV3GFSv2 implementation
  - S2S Ensemble
    - Marine Hycom-CICE for Weather time scales and MOM6-CICE for S2S.
- Test plans should to align with SIP - the CAM, Physics and Marine ones were already well on their way, S2S Ensemble needs more work
- Need to determine what groups will be performing which tasks in test plan and how to incorporate group-specific test plans into larger plan.
- Encompass all scales from CAM to S2S (start preparing for scale-aware physics and unification of suite for UFS)
- Test in ensemble mode (GEFS and ensemble DA) and with components
- Tests in LES mode can support physics development/improvement/tuning
- Additional diagnostics are needed in particular tendencies from parameterizations and processes within parameterizations
## Metrics

### Categories: Important to All, Global, Regional/CAM, S2S, Process Oriented, Coupling Validation, Marine, Tropical, ACC, Land/Hydro, Upper Air and Space

- Ensembles/Probabilistic measures were embedded in each sheet
- Example Metrics Worksheet - Atmospheric Chemistry and Composition (ACC)

<table>
<thead>
<tr>
<th>Forecast Field</th>
<th>Specialty</th>
<th>Vertical Attribute</th>
<th>Temporal Attribute</th>
<th>Validation Source</th>
<th>Priority</th>
<th>Maturity</th>
<th>Deterministic Methodology</th>
<th>Deterministic Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. Aerosol Optical D</td>
<td>Compositon</td>
<td>Surface</td>
<td>1-hr, 6-hr, 24-hr T</td>
<td>Stage IV 1-hr, 6-hr, 24-hr Precip</td>
<td>Grid-to-Grid</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerosol optical depth</td>
<td>total column</td>
<td>1hr, 24hr, monthly</td>
<td>AERONET L1.5, L2.0</td>
<td>1</td>
<td>Grid-to-observations</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerosol optical depth</td>
<td>total column</td>
<td>daily</td>
<td>MODIS, VIIRS</td>
<td>1</td>
<td>Grid-to-observations</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ozone</td>
<td>surface</td>
<td>1-hr, 8-hr, daily m</td>
<td>EPA AIRNOW, AQS</td>
<td>1</td>
<td>Grid-to-observations</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM2.5</td>
<td>surface</td>
<td>1-hr, 24-hr, daily max, ave</td>
<td>EPA AIRNOW, AQS, WSR-88D, METAR Ceiometer, RAOB, ACARS, BL Profilers</td>
<td>1</td>
<td>2</td>
<td>Grid-to-Grid, Grid-to-Point</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
<td></td>
</tr>
<tr>
<td>PBL Depth</td>
<td>Environmental/Air Qu Top of PBL</td>
<td>Instantaneous</td>
<td>1</td>
<td>Grid-to-Grid, Grid-to-Point</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Downward Shortwave</td>
<td>Radiation</td>
<td>Surface</td>
<td>Instantaneous/Av</td>
<td>ARM, Surfrad (Oak ridge)</td>
<td>1</td>
<td>1</td>
<td>Grid-to-Point</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
</tr>
<tr>
<td>Downward UV</td>
<td>Radiation</td>
<td>Surface</td>
<td>Instantaneous/Av</td>
<td>ARM, Surfrad (Oak ridge)</td>
<td>1</td>
<td>1</td>
<td>Grid-to-Point</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
</tr>
<tr>
<td>Aerosol optical depth</td>
<td>total column</td>
<td>15 min</td>
<td>GOES, Himawari, Meteosat, GEMS</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>Grid-to-Point</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
</tr>
<tr>
<td>Aerosol index</td>
<td>smoke, ash plume</td>
<td>daily</td>
<td>OMPS, OMI, TROPOMI</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>Grid-to-Point</td>
<td>CSI, BIAS, FSS, POD, FAR, AUR, Performance</td>
</tr>
</tbody>
</table>
Metrics

- Teleconnections: MJO, QBO, etc ...
- Data from every time step to inform time steps and to see cumulative impact from parameterization schemes.
- Observations: Do we have what we need? If not, could use an LES (Large Eddy Simulation) in some cases. Could use an analysis but not necessarily for small-scale processes.
- Need to focus on areas beyond hemispheric and CONUS regions and deterministic metrics (e.g., diurnal, precip., tropics, polar, etc …)
- Need to be multi-disciplinary and pushed out to the research area. Funding!
- Reduction of number of Scorecards. Need to prioritize metrics.
  - Summary scorecard of to 10 changed metrics or …
  - Have stakeholder specific scorecards with some performance graphics (synthesis measures).
Hierarchical Testing

- Hierarchical coupled model testing is needed for R&D at universities and partner agencies so that testing of innovations is relevant for potential future R2O. Good for process studies; exploring parameter space; sanity checks; idealized testing.
- In addition to a single column version of the atmosphere, reduced-complexity versions of other components are needed:
  - Ocean: mixed layer model (no dynamics); Land surface: bucket (or leaky, dirty bucket);
  - Sea ice: thermodynamics (no motion); Atmosphere: small radius aquaplanet
- A statistical expert should be involved in experimental design to ensure that the hypothesis being tested is well formulated
- Diverse cases and optimal case number (initial conditions, physical processes, forecast difficulty etc)
- DA, ensemble, and in-situ data impacts should be included but cold-starts can also add knowledge
- HTF needs to be model agnostic, very user-friendly, well-documented, with some support to be adopted by the community.
Pre-Implementation Test Plans
- Should align with SIP - CAM, Physics and Marine already well on their way, S2S Ensemble needs a more massaging
- There’s a slightly murky path for how to move forward with utilizing the test plans but we need to keep working on how to clarify the water

Metrics
- S2S, Tropics and Polar do not have SIP WG; Upper Atmosphere / Space was not filled in
- Needed more discussion specifically focused on Ensemble and DA metrics
- Need a list of observations necessary to facilitate meaningful evaluation

Hierarchical Testing
- We now have some idea on how to discuss a DA HTF
- Mapping of Hierarchichal Testing Plans to Readiness Levels (RL) (e.g. there could be templates for RL 6-8 that are different than for R-L 1-5)
- Need a testing framework to allow for turning on/off components for coupled UFS

SIP Action
- UFS SIP WGs should map to test plans and metrics organization and vice-versa