

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

Also

Thank you to the UFS Working Groups that contributed, the Session Chairs, and the Break-out Group Facilitators

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
- ESPC
- Air Force, Navy
- NASA GMAO, CCMC

**~100
participants
60-70 actively
engaged**

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

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

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)

Forecast Field	Specialty	Vertical Attribute	Temporal Attribute	Validation Source	Priority	Maturity	Deterministic Methodology	Deterministic Metrics
DRAFT IN PROGRESS								
e.g. Aerosol Optical Depth	Composition	Surface	1-hr, 6-hr, 24-hr	Stage IV 1-hr, 6-hr, 24-hr Precip			Grid-to-Grid	CSI, BIAS, FSS, POD, FAR, AUR, Performance
Aerosol optical depth		total column	1hr, 24hr, monthly	AERONET L1.5, L2.0	1	1	Grid-to-observations	CSI, BIAS, FSS, POD, FAR, AUR, Performance
Aerosol optical depth		total column	daily	MODIS, VIIRS	1	1	Grid-to-observations	CSI, BIAS, FSS, POD, FAR, AUR, Performance
Ozone		surface	1-hr, 8-hr, daily max	EPA AIRNOW, AQS	1	1	Grid-to-observations	CSI, BIAS, FSS, POD, FAR, AUR, Performance
PM2.5		surface	1-hr, 24-hr, daily max, ave	EPA AIRNOW, AQS, WSR-88D, METAR Ceilometer, RAOB, ACARS, BL Profilers	1	1	Grid-to-observations	CSI, BIAS, FSS, POD, FAR, AUR, Performance
PBL Depth	Environmental/Air Quality	Top of PBL	Instantaneous	ARM, Surfrad (Oak ridge ameriflux)	1	1	Grid-to-Grid, Grid-to-Point	RMSE, BIAS, Correlation
Downward Shortwave Radiation	Air Quality/Energy	Surface	Instantaneous/Average	ARM, Surfrad (Oak ridge ameriflux), AirNow UV radiation	1	1	Grid-to-Point	RMSE, BIAS, Correlation
Downward UV Radiation	Air Quality/Energy	Surface	Instantaneous/Average	ARM, Surfrad (Oak ridge ameriflux), AirNow UV radiation	1	1	Grid-to-Point	RMSE, BIAS, Correlation
Aerosol optical depth		total column	15 min	GOES, Himawari, Meteosat, GEMS	2	2		
Aerosol index smoke, ash plume			daily	OMPS, OMI, TROPOMI	2	2		

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 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 Hierarchical 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