A Preliminary Report Out on the 2021 DTC UFS Evaluation Metrics Workshop

Tara Jensen
NCAR/RAL and DTC

Geoff Manikin
NOAA/NCEP/EMC

UFS Webinar Series - April 8, 2021

Motivation - Need for Metrics

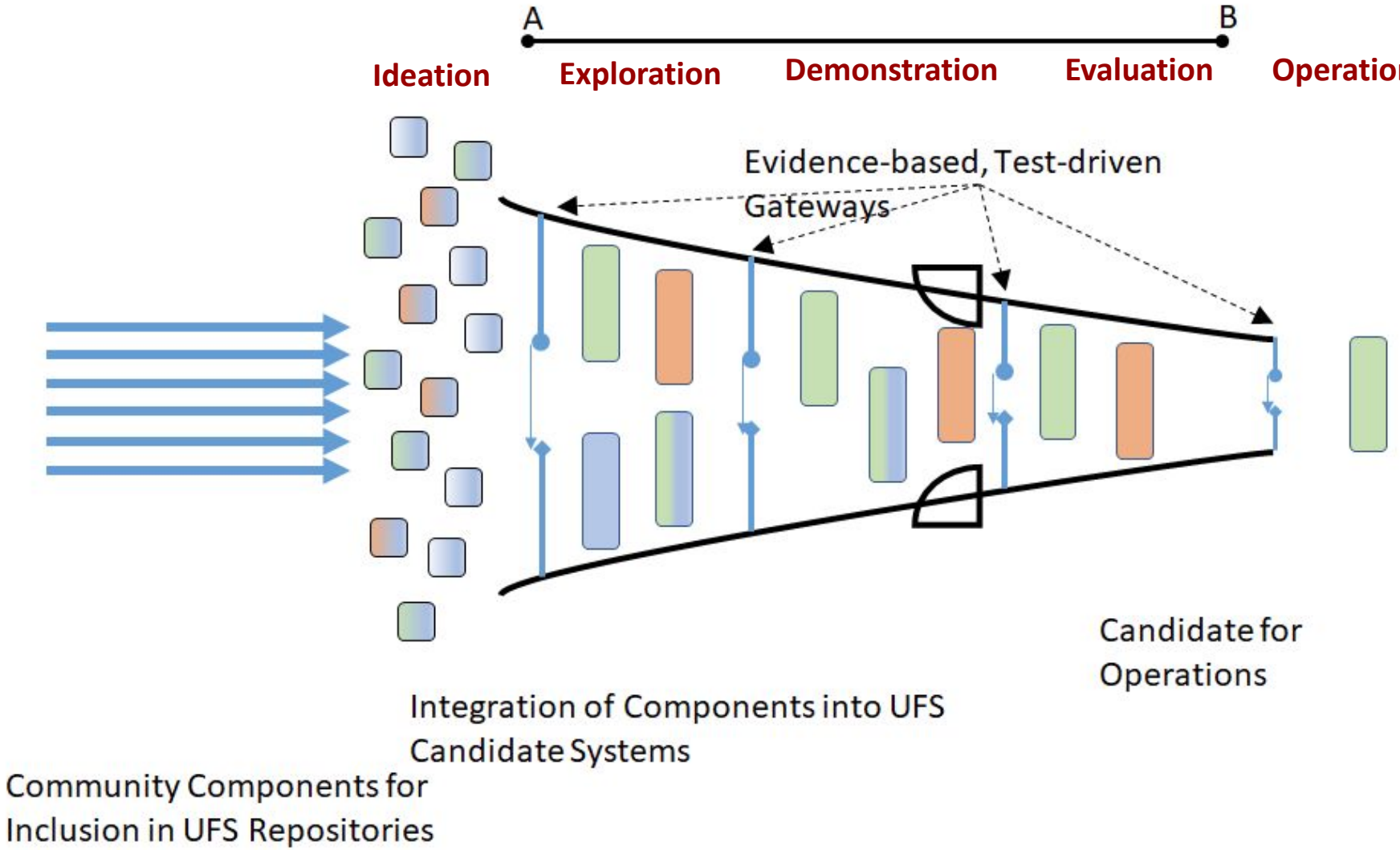
Reproducible and documented methods for evidence-based decision making

- **Identify *goals*** of verification and ***questions*** **to be answered**
- **Identify and collect *observations*** that can be used to answer the questions of interest
  - If possible, characterize ***uncertainty*** in the observations
- **Specify *type of forecasts*** **and** ***type of observations*** and how they can/should be ***matched***
- **Identify multiple *verification attributes*** that can provide answers to the questions of interest
- **Identify a *standard of comparison*** that provides a reference level of skill (e.g., persistence, climatology, reference model)

Motivation - Stages and Gates

UFS view of R2O:

As a repeated, narrowing stage and gate process



Starting Point - 2018 Metrics Workshop

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	emporal Attribut	Validation Source	Priority	Maturity	Deterministic Methodology	Determinisit
				DRAFT IN PROGRESS				
e.g. Aerosol Optical D	Compositon	Surface	1-hr, 6-hr, 24-hr T	Stage IV 1-hr, 6-hr, 24-hr Precip			Grid-to-Grid	CSI, BIAS, FSS, POD, FAR, AUR, Perform
Aerosol optical depth		total column	1hr, 24hr, monthl	AERONET L1.5, L2.0	1	1	Grid-to-observations	*CSI, BIAS, FSS, POD, FAR, AUR, Perform*
Aerosol optical depth		total column	daily	MODIS, VIIRS	1	1	Grid-to-observations, C	*CSI, BIAS, FSS, POD, FAR, AUR, Perform*
Ozone		surface	1-hr, 8-hr, daily m	EPA AIRNOW, AQS	1	1	Grid-to-observations	*CSI, BIAS, FSS, POD, FAR, AUR, Perform*
PM2.5		surface	1-hr, 24-hr, daily max,ave	EPA AIRNOW, AQS,	1	1	Grid-to-observations	*CSI, BIAS, FSS, POD, FAR, AUR, Perform*
PBL Depth	Environmental/Air Qu	Top of PBL	Instantaneous	WSR-88D, METAR Ceilometer, RAOB, ACARS, BL Profilers	1	2	Grid-to-Grid, Grid-to-Point	RMSE, BIAS, Corr
Downward Shortwave Radiation	Air Quality/Energy	Surface	Instantaneous/Av	ARM, Surfrad (Oak ridge ameriflux)	1	1	Grid-to-Point	RMSE,
Downward UV Radiation	Air Quality/Energy	Surface	Instantaneous/Av	ARM, Surfrad (Oak ridge ameriflux), AirNow UV r	1	1	Grid-to-Point	RMSE,
Aerosol optical depth		total column	15 min	GOES, Himawari, Meteosat, GEMS	2	2		
Aerosol index			daily	OMPS, OMI, TROPOMI	2	2		
smoke, ash plume height			daily	MISR, CALIPSO, MPLNET	1	1		
			1-hr, 24-hr,					

Workshop Overview

Day 1: **Opening Plenary** - Motivation, Pre-workshop Surveys, Goals
Breakout Group Session 1: Short Range Weather (SRW)

Day 2: **Breakout Group Session 2:** Medium Range Weather (MRW)
Breakout Group session 3: Subseasonal to Seasonal (S2S)
Polling to Prioritize Metrics by UFS Application

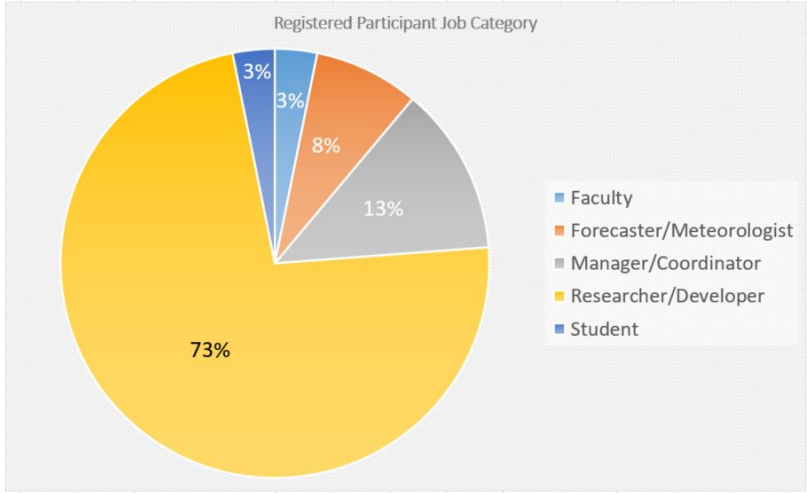
Day 3: **Report Out** on Poll Results and Discussion of How to Interpret
Breakout Group Session 4: How to Metrics to Development Gates
Final Wrap-up including Next Steps

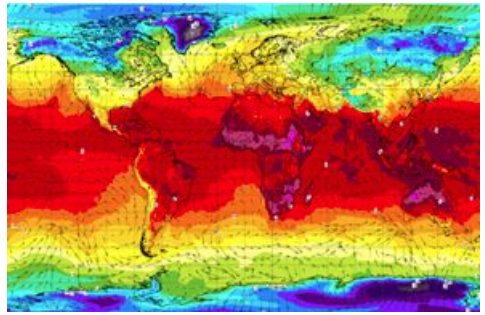
3 Pre-Workshop Surveys - 15 hours - 3 days - Live Polling - Lots of Discussion - Still not enough!!!

Participant Overview

Registered Participants: 315
Attended: Approximately 200 for the first plenary, 75 for the last
Number of Breakouts:
- SRW - 6
- MRW - 5 + DA
- S2S - 6
- Assigning to Gates - 7

Largest Breakout Group - 45
Smallest Breakout Group - 7



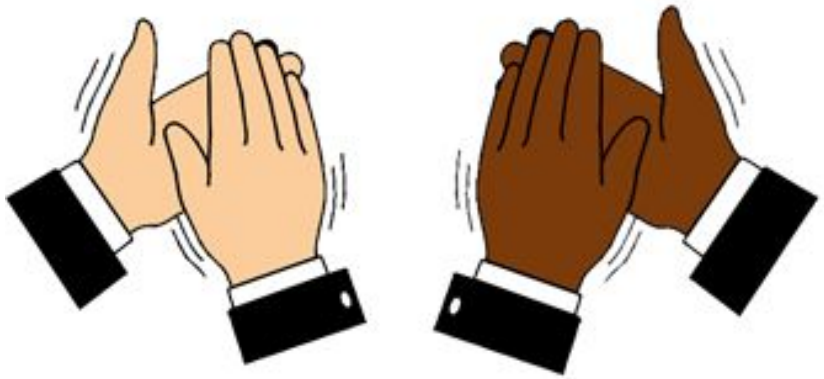


DOD: NRL, ONR, USAF
DOE: PNNL
International: Bureau of Meteorology, BC Hydro, DLR, Environment and Climate Change Canada, Int. GEWEX Project Office, Universidad Nacional Autonoma de Mexico, University of Iceland, University of Twente
NOAA: AFS, AOML, ARL, AWC, CPC, CPO, CSDL, CSL, EMC, ERT, GFDL, GLERL, GSL, MDL, NCEI, NESDIS, NHC, NOS, NSIDC, NSSL, OPC, OSTI, OWP, PMEL, PSL, RegHQs, SPC, SWPC, WFOs, WPC, WPO
WPC
Other: DTC, NASA, NCAR, USNIC, CT DOE/Env. Protection, ID DO Env. Qual., MS DO Env. Qual., PA DO Env. Protection, South Coast AQ Mgmt District
Private: Citadel LLC, FirstEnergy, Leidos/JSC Space Radiation Analysis Group, Tsunami Consultant, Systems Research Group, The Climate Corporation, The Weather Company/IBM
University: George Mason University, North Carolina State University, Purdue University, SUNY-Albany, University of Arizona, University of Colorado, University of Connecticut, University of Houston, University of Illinois at Urbana Champaign, University of Maryland, University of Miami, University of Michigan, University of Missouri, University of Oklahoma, University of Texas at Austin, University of Wisconsin-Madison, University of Wisconsin-Milwaukee, Virginia Tech

Workshop Support

Organizers: Tara Jensen (NCAR and DTC), Geoff Manikin (EMC), Burkely Gallo (SPC), Jason Levit (EMC), Jack Settlemaier (SRH), Sarah Lu (U Albany), Linden Wolf (OSTI), Deepthi Achuthavarier (OSTI), Yan Xue (OSTI), Jason Otkin (U Wisc-CIMSS), Mike Baldwin (Purdue), Dave Turner (GSL), and Cristiana Stan (GMU)

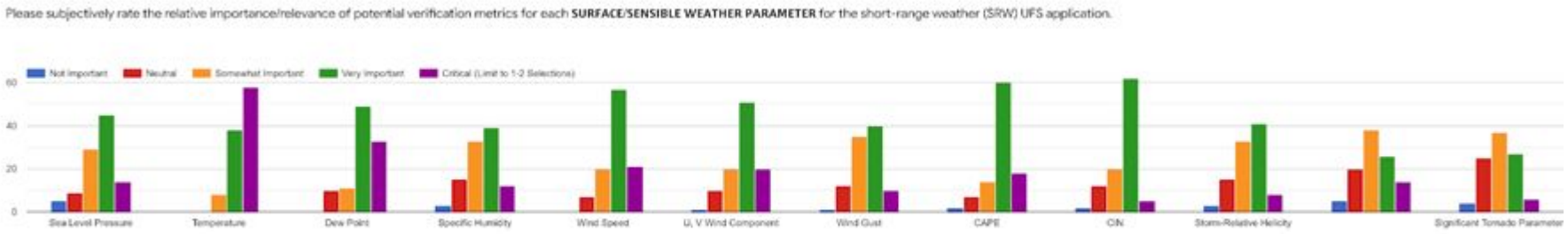
- UFS V&V Cross-Cutting Team
- EMC MEG and Verification Teams
- All of the Breakout Facilitators
- DTC administrative support and UCAR Multimedia Services



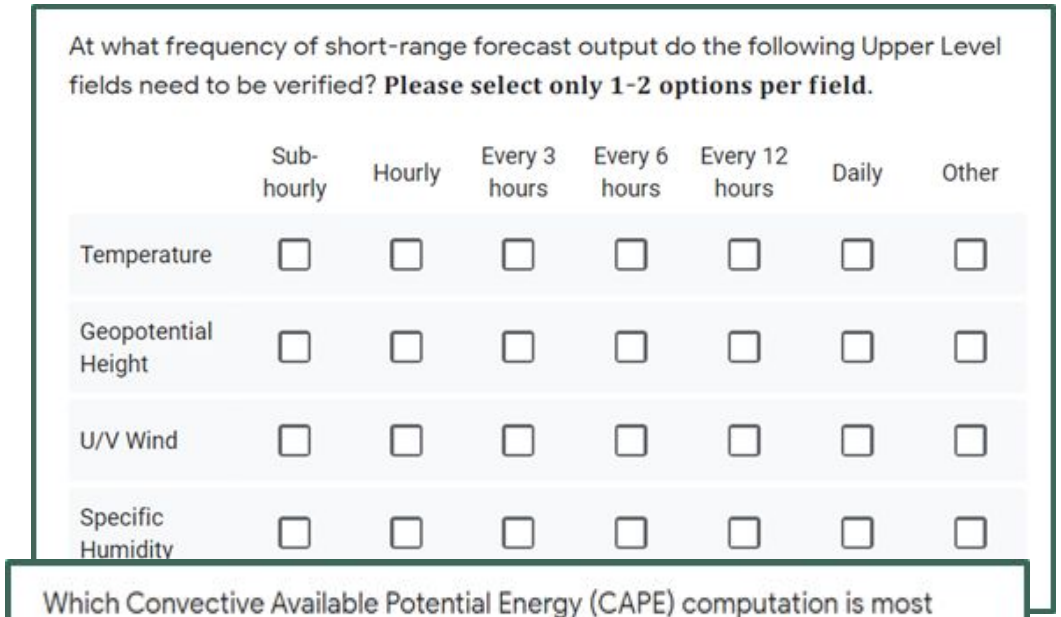
Funded by: UFS R2O Project - Hosted by: DTC

Survey 1 Methodology

- Took place in October 2020
- Focused on the Working Groups of UFS
  - Short Range Weather (SRW); Medium Range Weather (MRW); Subseasonal to Seasonal (S22)
  - Air Quality/Atmospheric Composition
  - Coastal
  - Hurricane
  - Marine and Cryosphere
  - Space Weather
- Started from 2018 Workshop findings
- Relied on Subject Matter Expertise
- Used web-based literature searches

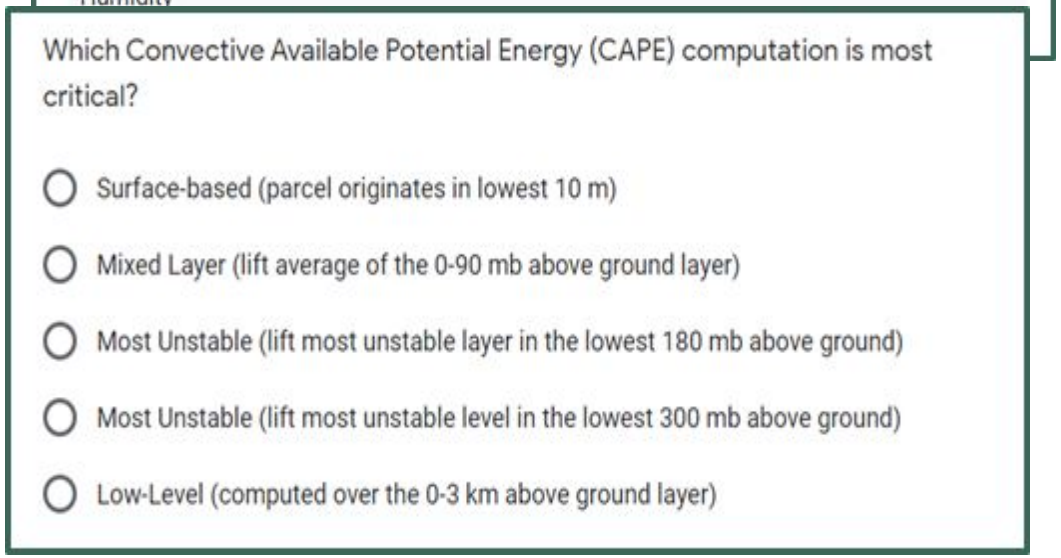


Survey 2 Methodology



At what frequency of short-range forecast output do the following Upper Level fields need to be verified? **Please select only 1-2 options per field.**

	Sub-hourly	Hourly	Every 3 hours	Every 6 hours	Every 12 hours	Daily	Other
Temperature	☐	☐	☐	☐	☐	☐	☐
Geopotential Height	☐	☐	☐	☐	☐	☐	☐
U/V Wind	☐	☐	☐	☐	☐	☐	☐
Specific Humidity	☐	☐	☐	☐	☐	☐	☐



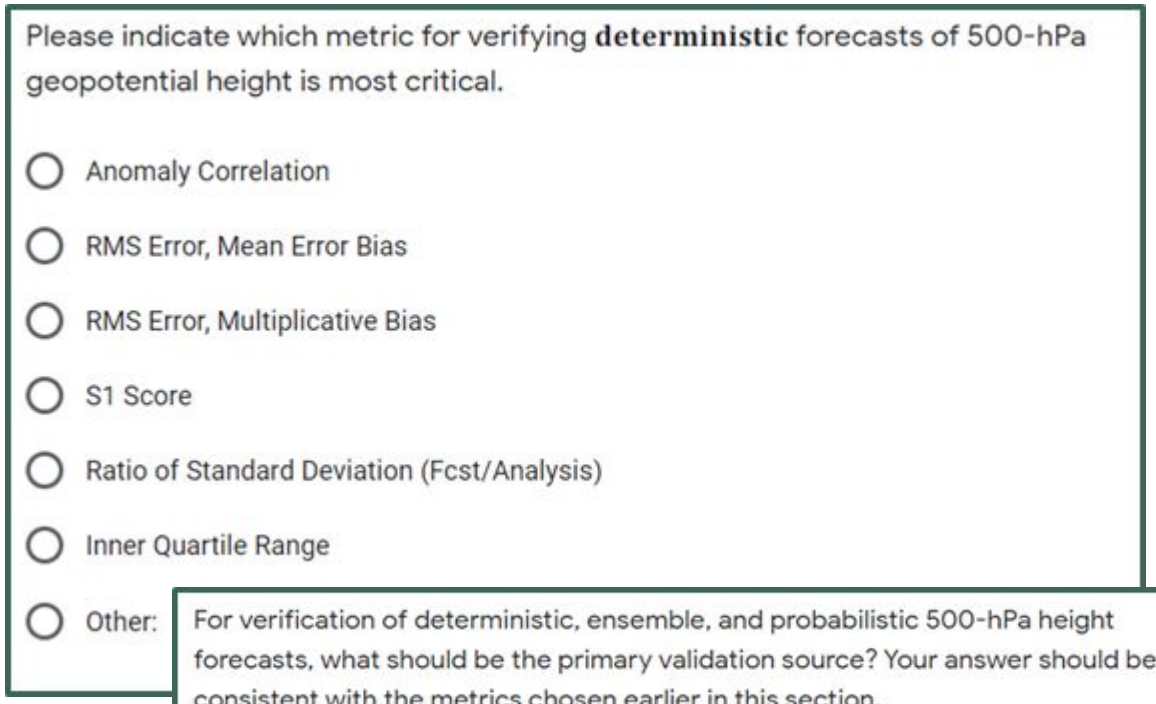
Which Convective Available Potential Energy (CAPE) computation is most critical?

- ○ Surface-based (parcel originates in lowest 10 m)
- ○ Mixed Layer (lift average of the 0-90 mb above ground layer)
- ○ Most Unstable (lift most unstable layer in the lowest 180 mb above ground)
- ○ Most Unstable (lift most unstable level in the lowest 300 mb above ground)
- ○ Low-Level (computed over the 0-3 km above ground layer)

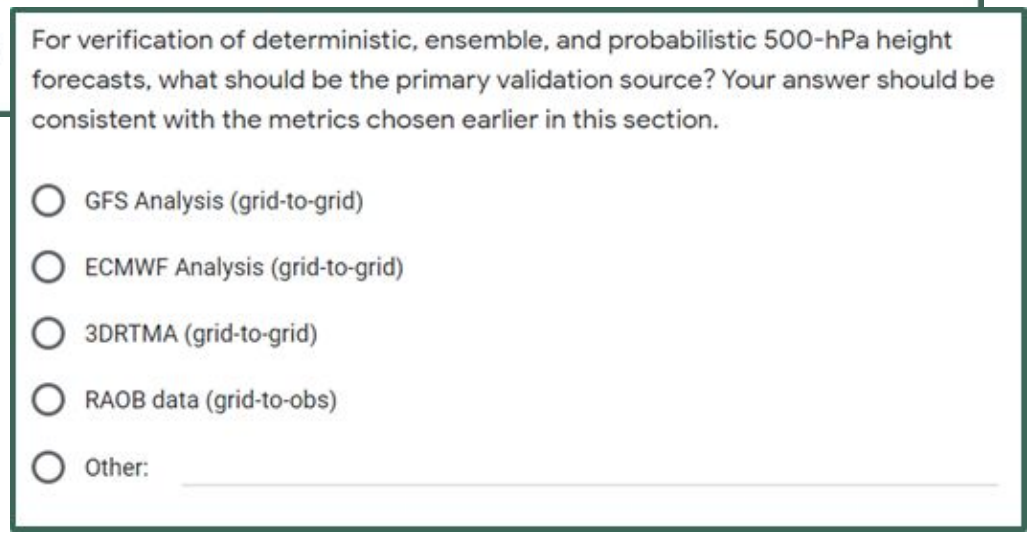
- Took place mid-December to mid-January
- Given the fields from Survey 1, was time to delve into specifics of those fields
- Organization changed from Survey 1
- Marine and Coastal were merged and Cryosphere broken out
- Subseasonal broken out from MRW
- Sections:
  - Heights and Layers – where should we be assessing this field?
  - Verification frequency – how frequently should we assess this field?
  - Temporal attributes – what accumulations or maxima should we consider?
- Also asked questions through write-in fields

Survey 3 Methodology

- Time to dig into metrics – surveys out in early February
- This survey was to answer remaining questions about the specifics of verification for each field
  - Metrics for deterministic, probabilistic, and ensemble usage
  - Thresholds and Verification domains
  - Observation sources
- One complication – subject matter experts in different fields may not be as familiar with verification and validation specifics as they are with their application
- For some surveys, questions about nature of the fields, the spatial and temporal errors, and the variability of peak values across the globe were asked to allow statisticians to help suggest metrics to use

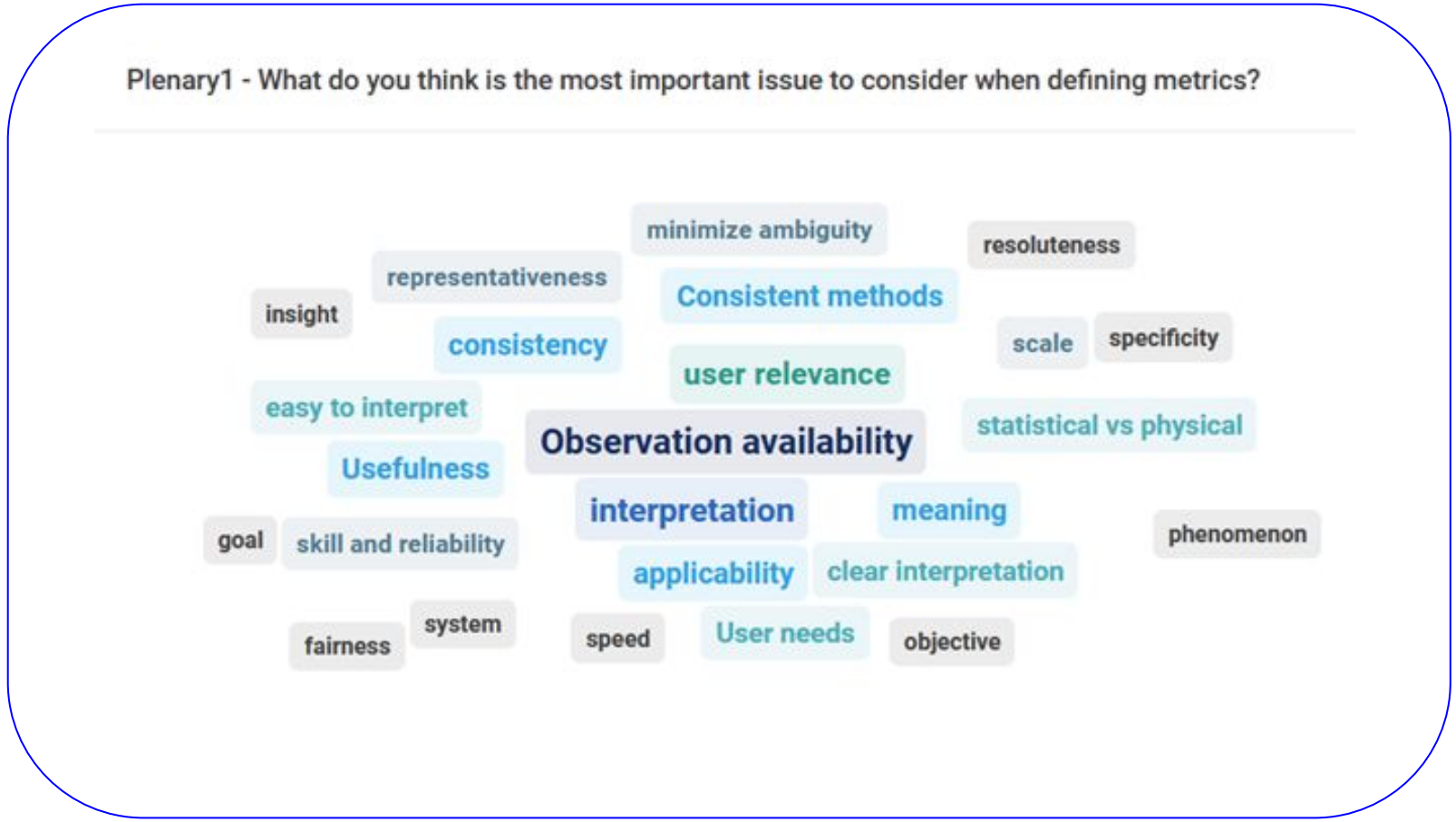

Please indicate which metric for verifying **deterministic** forecasts of 500-hPa geopotential height is most critical.

- ○ Anomaly Correlation
- ○ RMS Error, Mean Error Bias
- ○ RMS Error, Multiplicative Bias
- ○ S1 Score
- ○ Ratio of Standard Deviation (Fcst/Analysis)
- ○ Inner Quartile Range
- ○ Other:


For verification of deterministic, ensemble, and probabilistic 500-hPa height forecasts, what should be the primary validation source? Your answer should be consistent with the metrics chosen earlier in this section.

- ○ GFS Analysis (grid-to-grid)
- ○ ECMWF Analysis (grid-to-grid)
- ○ 3DRTMA (grid-to-grid)
- ○ RAOB data (grid-to-obs)
- ○ Other:

Plenary Poll



Breakout Groups

MONDAY - 2/22 Breakout Groups 1 - SRW
1.1 AQ/Comp - SRW
1.2 Hurricane - SRW
1.3 Marine, Coastal, and Cryosphere - SRW
1.4 Land and Hydro - SRW
1.5 Space and Aviation - SRW
1.6 High Impact Weather (Precip/Severe) - SRW
1.0 Support

TUESDAY - 2/23 Breakout Groups 2 - MRW
2.1 AQ/Comp - MRW
2.2 Hurricane - MRW
2.3 Marine, Coastal, and Cryosphere - MRW
2.4 Land and Hydro - MRW
2.5 Data Assimilation
2.6 High Impact Weather (Precip/Severe) - MRW
2.0 Support

Monday and Tuesday Breakout Groups:

Each group:

- Prioritized Upper Level and Sensible Weather from "temporal parent survey" plus their own application metrics
- Identified which metrics / validation sources / regions that still need clarification and resolve uncertainty

Wednesday Breakout Groups:

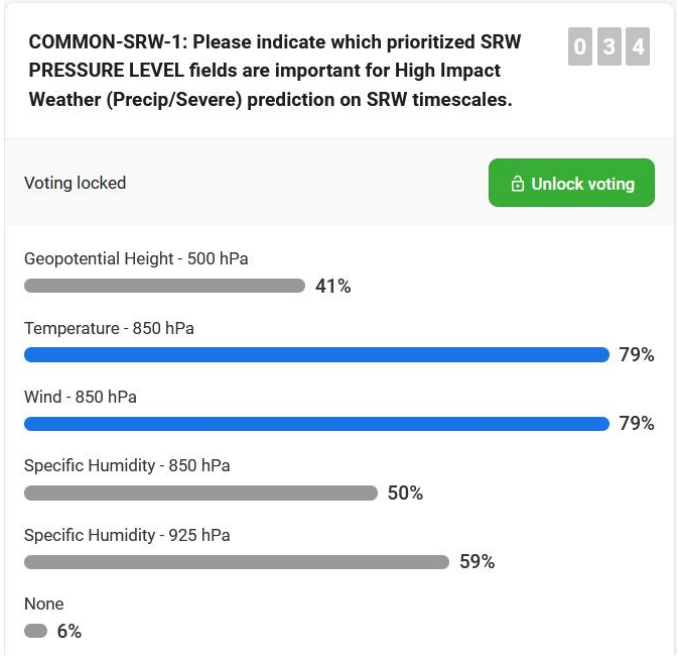
Each group discussed the stages and gates identified by UFS and how to assign prioritized metrics to them

TUESDAY - 2/23 Breakout Groups 3 - Subseasonal and Seasonal
3.1 AQ/Comp - S2S
3.2 Hurricane - S2S
3.3 Marine, Coastal, and Cryosphere - S2S
3.4 Land and Hydro - S2S
3.5 General Circulation Phenomena (MJO, ENSO, N
3.6 General Circulation Phenomena (MJO, ENSO, N
3.7 Weather Extremes (Drought, Fire, Temp Extrem
3.8 Weather Extremes (Drought, Fire, Temp Extrem
3.0 Support

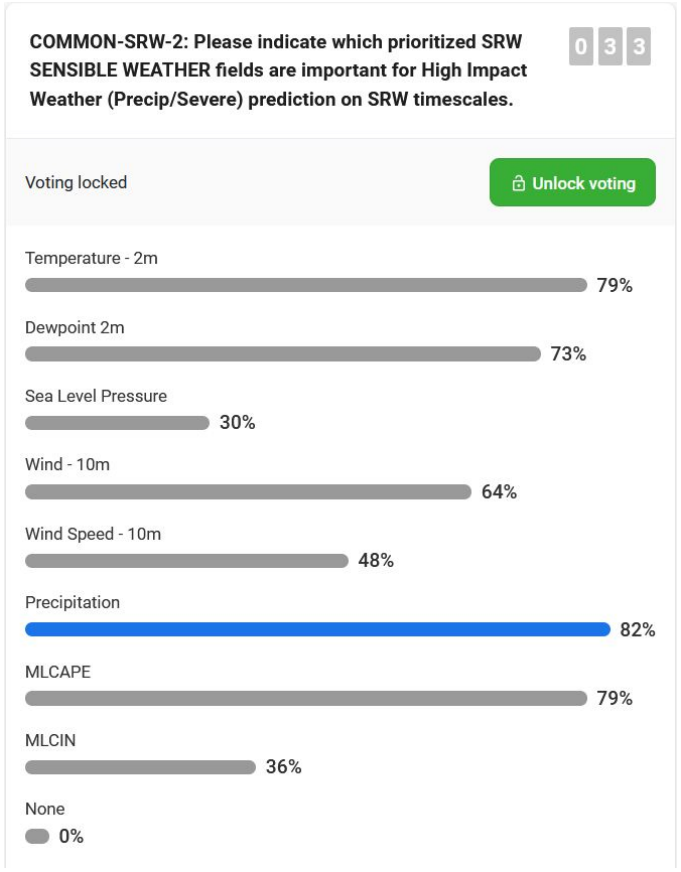
WEDNESDAY - 2/24 Breakout Groups 4 - How to Assign Metrics to Ga
4.1 SRW Gates - Blank
4.2 SRW Gates - Wolff
4.3 SRW Gates - Dawson
4.4 SRW Gates - Harrold
4.5 MRW Gates - Gottschalck
4.6 MRW Gates - Rood
4.7 MRW Gates - Tolman
4.8 Seasonal Gates - Janiga
4.9 Seasonal Gates - Kim
4.0 Support

Breakout Online Polling

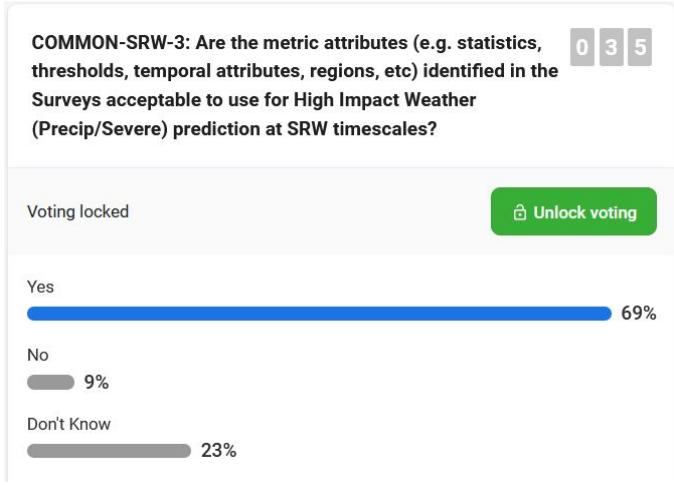
COMMON-SRW-1: Please indicate which prioritized SRW PRESSURE LEVEL fields are important for High Impact Weather (Precip/Severe) prediction on SRW timescales.



COMMON-SRW-2: Please indicate which prioritized SRW SENSIBLE WEATHER fields are important for High Impact Weather (Precip/Severe) prediction on SRW timescales.



COMMON-SRW-3: Are the metric attributes (e.g. statistics, thresholds, temporal attributes, regions, etc) identified in the Surveys acceptable to use for High Impact Weather (Precip/Severe) prediction at SRW timescales?



Candidate Metrics Polling

TUESDAY - 2/23 Application Metrics Polling
ALL POLLS are at: https://app.sli.do/event/vccfxvzl

- Polls were generated from the Candidate Metrics Spreadsheets used in the Breakout Groups
- Generally, one poll was generated for each application
- For MRW and S2S, the poll was split to make it easier to take
- Also added polls for Aviation (at the request of AWC), Data Assimilation, and Land

Example of Results shown in later slides

- FP.1 SRW - RRFS
- FP.2a MRW - GFS/GEFS
- FP.2b S2S - GFS/GEFS
- FP.3 Seasonal - CFS
- FP.4 AQ

- FP.5 Coastal - NWPS
- FP.6 Hydro - NWM
- FP.7 Hurricane - HAFS
- FP.8 Lake - GLWM
- FP.9 Marine/Cryosphere - RTOFS

- FP.10 Space - WAM-IPE
- FP. 11 Aviation Prediction
- FP.12 Data Assimilation Component
- FP.13 Land Component

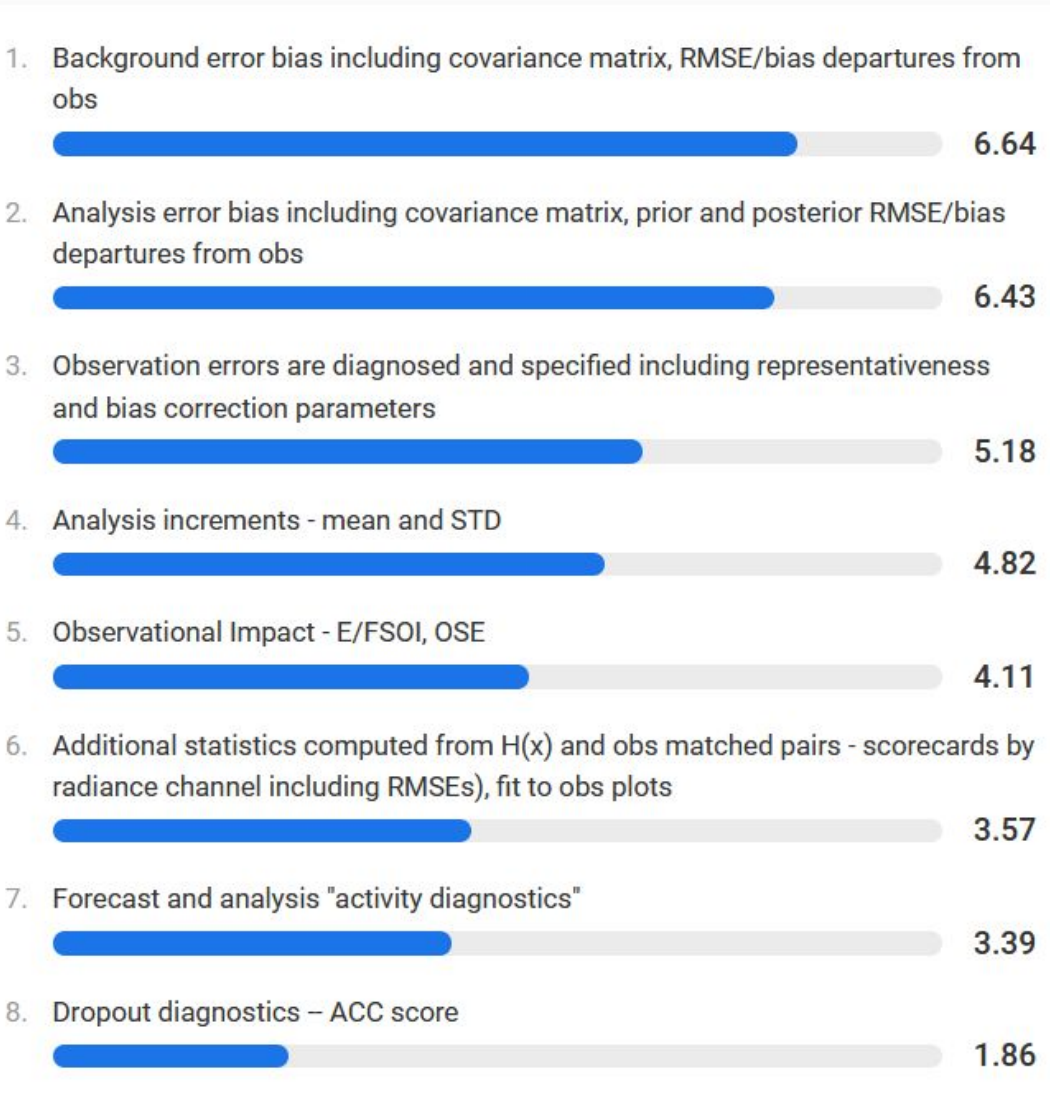
Candidate Metrics Polling

Please Rank the following fields for the Data Assimilation Application. You MUST GIVE A RANKING TO ALL fields before you can submit.

Select options from the list below.

- (●) Analysis error bias including covariance matrix, prior and posterior RMSE/bias departures from obs
- ( ) Background error bias including covariance matrix, RMSE/bias departures from obs
- ( ) Observation errors are diagnosed and specified including representativeness and bias correction parameters
- ( ) Observational Impact - E/FSOI, OSE
- ( ) Additional statistics computed from H(x) and obs matched pairs - scorecards by radiance channel including RMSEs), fit to obs plots
- ( ) Forecast and analysis "activity diagnostics"
- ( ) Dropout diagnostics -- ACC score
- ( ) Analysis increments - mean and STD

Send



The Approach to Data Synthesis

- Following the three rounds of surveys, we had data for each UFS application covering fields, metrics, temporal attributes, key thresholds, regions, validation source, and verification approach
- **Phase 1** of the synthesis (**pre-workshop**) involved the creation of worksheets with all of the data aggregated and gaps identified, to serve as a starting point for the breakout sessions
- **Phase 2** of the synthesis (**during workshop**) involved updating the aggregated data to account for changes made and new fields added during the workshop breakout sessions to prepare for the workshop voting
- **Phase 3** of the synthesis (**post workshop**) involved reorganizing the aggregated data to match the voting results
- Phase 3 also involved efforts to provide alternate rankings of prioritized metrics to represent the needs of smaller but important groups of customers and developers

Data Synthesis

- A worksheet for each UFS application was built to collect information
- The initial tab contained a compilation of data from each round of surveys

Hurricane Metrics
Aggregate

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	FIELD	LEVEL	DETERMINISTIC METRIC	ENSEMBLE METRIC	PROBABILISTIC METRIC	TEMPORAL ATTRIBUTE	NOTABLE THRESHOLDS	REGION	VERIFICATION APPROACH	VALIDATION SOURCE	CLIMATOLOGY SOURCE	SPATIAL MATCHING	STATISTIC available in METplus?	FILE FORMAT supported in METplus?	COMMENTS/DISCUSSION TOPICS
10	TC CHARACTERISTICS (TCC)														
11	Intense Precipitation	Surface	ETS,FreqBias	ETS of Ens Mean	Reliability Diagram	Every 6h or Every 24h	0.25 and 0.5" for 6h 1 and 2" for 24h	All basins separate	grid-to-grid, grid-to-obs	tie MRMS QPF, Stage IV					break tie between MRMS QPF and Stage IV
12	Rapid Intensification (Decay) Rate (MSLP)	Surface	RMSE, Bias	RMS of Ens Mean + Spread		Mostly every 12hrs, some 6hr	MSLP breaks (1000,980)	All basins separate	grid-to-grid, grid-to-obs	ECMWF Analysis, NHC Best Track					need to identify if RI should be evaluated both by MSLP and Wind Speed; also clarify if evaluation is every 6 or 12 hours
13	Rapid Intensification (Decay) Rate (Wind Speed)	Surface	RMSE, Bias	RMS of Ens Mean + Spread		Mostly every 12hrs, some 6hr	SS Cats (TS, Hurr, Major)	All basins separate	grid-to-grid, grid-to-obs	ECMWF Analysis, NHC Best Track					Settle Columns F,G
14	Radius of Maximum Winds (RMW)	10 m	RMSE, Bias	RMS of Ensemble Mean + Spread		Every 6 h		All basins separate (Atlantic is the basin with consistent recon data)	grid-to-obs	NHC operational estimates, Recon Vortex messages					
15	Maximum extent of winds (wind radii)	10 m	Bias	RMS of Ensemble Mean + Spread		Every 3h	34,50,64 kt	All basins separate (or aggregate)	grid-to-obs grid-to-grid	NHC Best Track, METARS, ECMWF Analysis					
16	Wave Height	surface													identify if wave height is critical and if so, get all meta-data
17	TC INTENSITY (TCI)														
18	Max Wind Speed	Sfc/10-m	RMS, Bias, Perf Diagram	RMS of Ensemble Mean + Spread		Every 6h		All basins separate	grid-to-grid, grid-to-obs	ECMWF, NHC Best Track data					ask if thresholds are the same as for radius of max wind criteria
19	TC TRACK (TCT)														
20	Absolute Track Error	Surface	Average NM	Avg Error of Ensemble Mean + Spread		Every 6h	65kt storms		grid-to-obs	Best Track		SRW - 50 or 100km suggested; MRW/S2S - greater distance			ask if evaluation should be performed for each basin separately, determine if 50-100km spatial matching is appropriate for MRW
21	Along Track Error	Surface	Average NM	Avg Error of Ensemble Mean + Spread		Every 6h	65kt storms		grid-to-obs	Best Track					ask if evaluation should be performed for each basin separately, determine if 50-100km spatial matching is appropriate for MRW
22	Cross Track Error	Surface	Average NM	Avg Error of Ensemble Mean + Spread		Every 6h	65kt storms		grid-to-obs	Best Track					ask if evaluation should be performed for each basin separately, determine if 50-100km spatial matching is appropriate for MRW

Pre-Workshop Data

- Orange and gray shading was used to identify necessary breakout group discussion; orange indicated that the survey voting was extremely close, while gray indicated that the necessary information was not obtained from the surveys

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	FIELD	LEVEL	DETERMINISTIC METRIC	ENSEMBLE METRIC	PROBABILISTIC METRIC	TEMPORAL ATTRIBUTE	NOTABLE THRESHOLDS	REGION	VERIFICATION APPROACH	VALIDATION SOURCE	CLIMATOLOGY SOURCE	SPATIAL MATCHING	STATISTIC available in METplus?	FILE FORMAT supported in METplus?	COMMENTS/DISCUSSION TOPICS
10	TC CHARACTERISTICS (TCC)														
11	Intense Precipitation	Surface	ETS,FreqBias	ETS of Ens Mean	Reliability Diagram	Every 6h or Every 24h	0.25 and 0.5" for 6h 1 and 2" for 24h	All basins separate	grid-to-grid, grid-to-obs	tie MRMS QPF, Stage IV					break tie between MRMS QPF and Stage IV
12	Rapid Intensification (Decay) Rate (MSLP)	Surface	RMSE, Bias	RMS of Ens Mean + Spread		Mostly every 12hrs, some 6hr	MSLP breaks (1000,980)	All basins separate	grid-to-grid, grid-to-obs	ECMWF Analysis, NHC Best Track					need to identify if RI should be evaluated both by MSLP and Wind Speed; also clarify if evaluation is every 6 or 12 hours
13	Rapid Intensification (Decay) Rate (Wind Speed)	Surface	RMSE, Bias	RMS of Ens Mean + Spread		Mostly every 12hrs, some 6hr	SS Cats (TS, Hurr, Major)	All basins separate	grid-to-grid, grid-to-obs	ECMWF Analysis, NHC Best Track					Settle Columns F,G
14	Radius of Maximum Winds (RMW)	10 m	RMSE, Bias	RMS of Ensemble Mean + Spread		Every 6 h		All basins separate (Atlantic is the basin with consistent recon data)	grid-to-obs	NHC operational estimates, Recon Vortex messages					
15	Maximum extent of winds (wind radii)	10 m	Bias	RMS of Ensemble Mean + Spread		Every 3h	34,50,64 kt	All basins separate (or aggregate)	grid-to-obs grid-to-grid	NHC Best Track, METARS, ECMWF Analysis					
16	Wave Height	surface													identify if wave height is critical and if so, get all meta-data
17	TC INTENSITY (TCI)														
18	Max Wind Speed	Sfc/10-m	RMS, Bias, Perf Diagram	RMS of Ensemble Mean + Spread		Every 6h		All basins separate	grid-to-grid, grid-to-obs	ECMWF, NHC Best Track data					ask if thresholds are the same as for radius of max wind criteria
19	TC TRACK (TCT)														
20	Absolute Track Error	Surface	Average NM	Avg Error of Ensemble Mean + Spread		Every 6h	65kt storms		grid-to-obs	Best Track		SRW - 50 or 100km suggested; MRW/S2S - greater distance			ask if evaluation should be performed for each basin separately, determine if 50-100km spatial matching is appropriate for MRW
21	Along Track Error	Surface	Average NM	Avg Error of Ensemble Mean + Spread		Every 6h	65kt storms		grid-to-obs	Best Track					ask if evaluation should be performed for each basin separately, determine if 50-100km spatial matching is appropriate for MRW
22	Cross Track Error	Surface	Average NM	Avg Error of Ensemble Mean + Spread		Every 6h	65kt storms		grid-to-obs	Best Track					ask if evaluation should be performed for each basin separately, determine if 50-100km spatial matching is appropriate for MRW

Post-Breakout Group Data

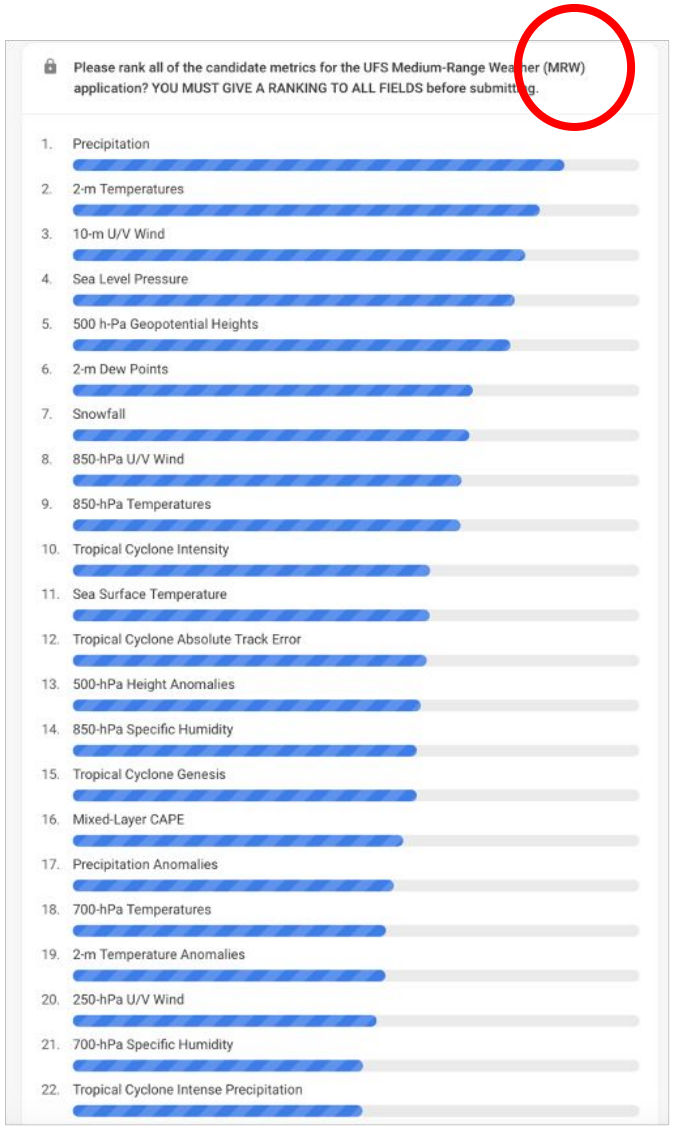
- Red was used to denote any resolutions of orange and gray boxes as well as any new fields added during the breakout sessions

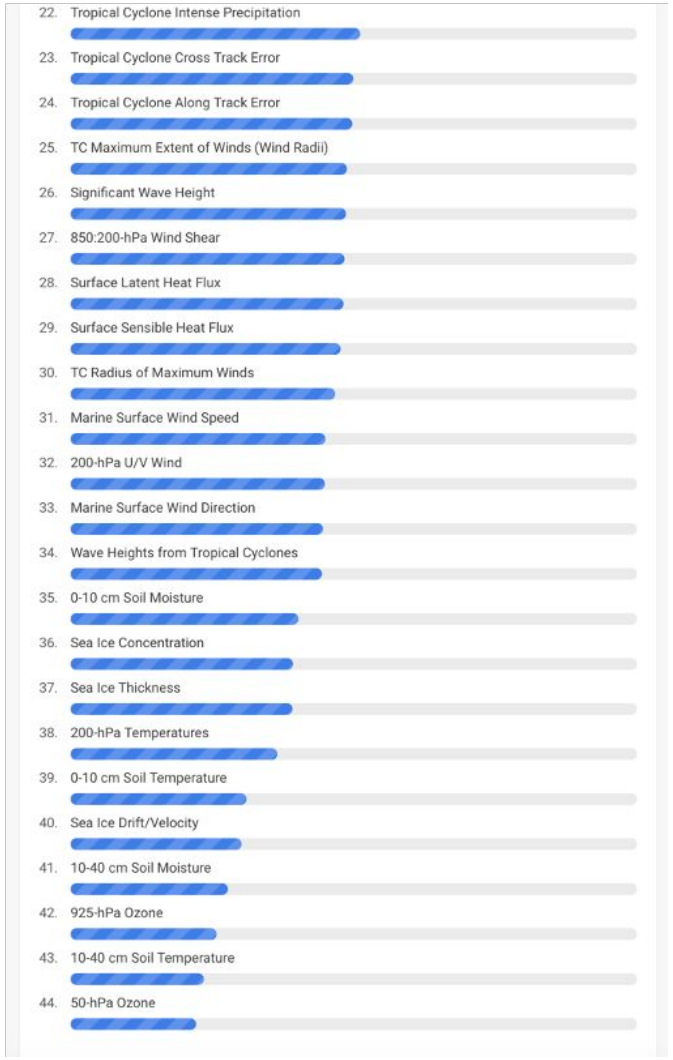
FIELD	LEVEL (S2)	DETERMNISTIC METRIC (S3)	ENSEMBLE METRIC (S3)	PROBABILISTIC METRIC (S3)	TEMPORAL ATTRIBUTE (S2)	NOTABLE THRESHOLDS (S3)	REGION (S3)	VERIFICATION APPROACH (S3)
SENSIBLE WEATHER								
Temperature	Sfc/2-m	RMS + Bias	RMSE of Ens. Mean + Ensemble Spread		Hourly	0 C, 60 F (when paired with high Td)?	CONUS divided into thirds + Alaska	Grid-to-obs
Dew Point	Sfc/2-m	RMS + Threshold Bias (do not compute stats for low values)	RMSE of Ens. Mean + Ensemble Spread		Hourly	50, 60, 70 F (possibly 40 and 50 in the west?); need lower threshold for fire wx	CONUS divided into thirds + Alaska	Grid-to-obs
Sea Level Pressure	Surface				Hourly			
Wind	Sfc/10-m	RMSE + Mean Error Bias	RMSE of Ens. Mean + Ensemble Spread		Hourly		CONUS divided into thirds + Alaska	Grid-to-obs
Wind Speed	Sfc/10-m	RMS Error + Bias	RMSE of Ens. Mean + Ensemble Spread		Hourly	25, 34, 48 kt (marine) 30 kt (blizzard) 20 (fire wx)	CONUS divided into thirds + Alaska	Grid-to-obs
Precipitation	Surface	Total Interest (MODE), FSS, and Contingency Table Elements	Performance diagram	Reliability Diagram	Hourly	6h: 0.25", 0.5", 1" (include 0.1" in winter) and 24h: 1" and 2" (include 0.5" in winter)	CONUS divided into thirds + Alaska	Grid-to-grid, grid-to-obs
Rainfall Rate	Surface				Hourly			
MLCAPE	Average of 0-90 mb AGL	RMS Error	RMSE of Ens. Mean + Ensemble Spread		Hourly	500, 1000, 2000 (and 250 for low CAPE SVR)	CONUS divided into thirds + Alaska/North America/SPC Convective Outlook Areas	Grid-to-grid, grid-to-obs
~~MLCIN~~	~~Average of 0-90 mb AGL~~	~~RMS Error~~	~~RMSE of Ens. Mean + Ensemble Spread~~		~~Hourly~~		~~CONUS divided into thirds + Alaska/North America/SPC Convective Outlook Areas~~	~~Grid-to-grid, grid-to-obs~~

New field

Items resolved

Live Polling at the Workshop





- Every field listed on the aggregate spreadsheet (following the breakout sessions) was added to a Slido poll for that particular application
- A full ranking was generated by the workshop participants for each survey

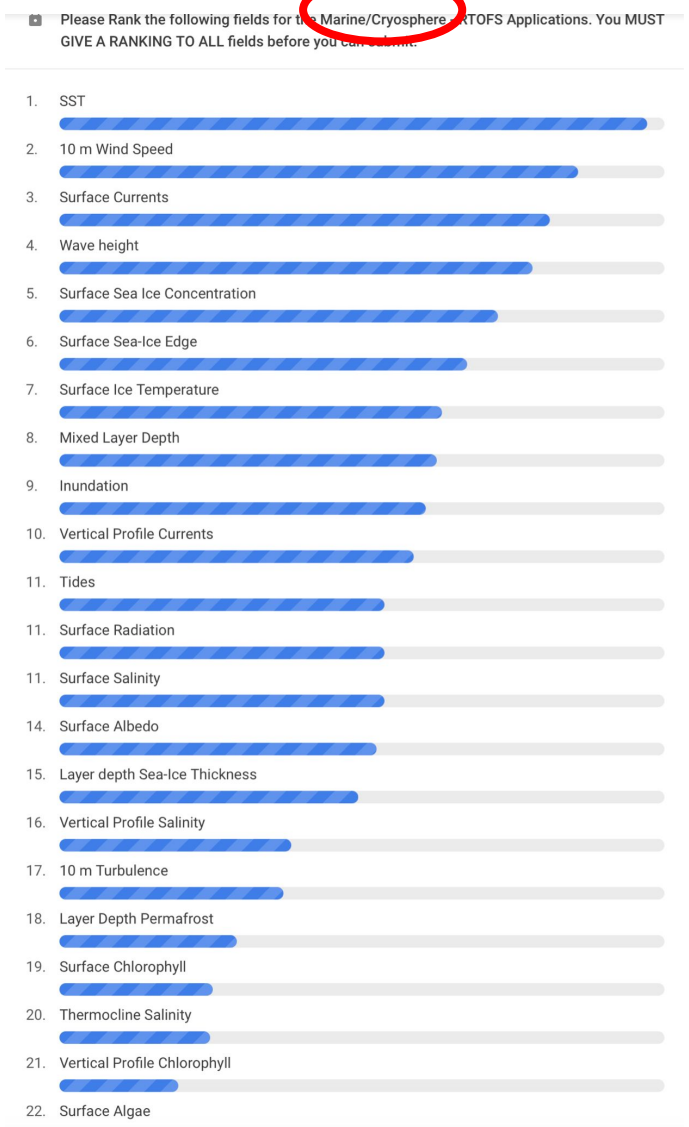
Live Polling at the Workshop

Please Rank the following fields for the Space - WAM-IPE application. You MUST GIVE A RANKING TO ALL fields before you can submit.

1. Electron Density (F-Region
2. TEC (Integrated)
3. Plasma Drift
4. Electron Density
5. Scintillation (F-Region)
6. Neutral Density
7. Proton Density
8. Solar Wind Vp (ground, F-region)
8. O/N2 Ratio (column integrated)
10. U/V Winds (250 km)
11. IMF-BZ (ground, F-region)
11. Kp Index (ground, F-region)
13. f0F2 (peak plasma density - coupled with height)
14. F10.7 (ground, F-region)
15. Electron Energy Flux
16. U/V Winds (80 km)
16. O (F-region)
18. SYM-H and Dst (ground, F-region)
19. Aurora Boundary
20. O+ (F-region)
21. Turbulence (0-2 km lower level)
22. Ceiling (N/A)
23. Wind Shear (0-2 km)
24. Turbulence (FL340 upper level)
25. Visibility (Sfc)

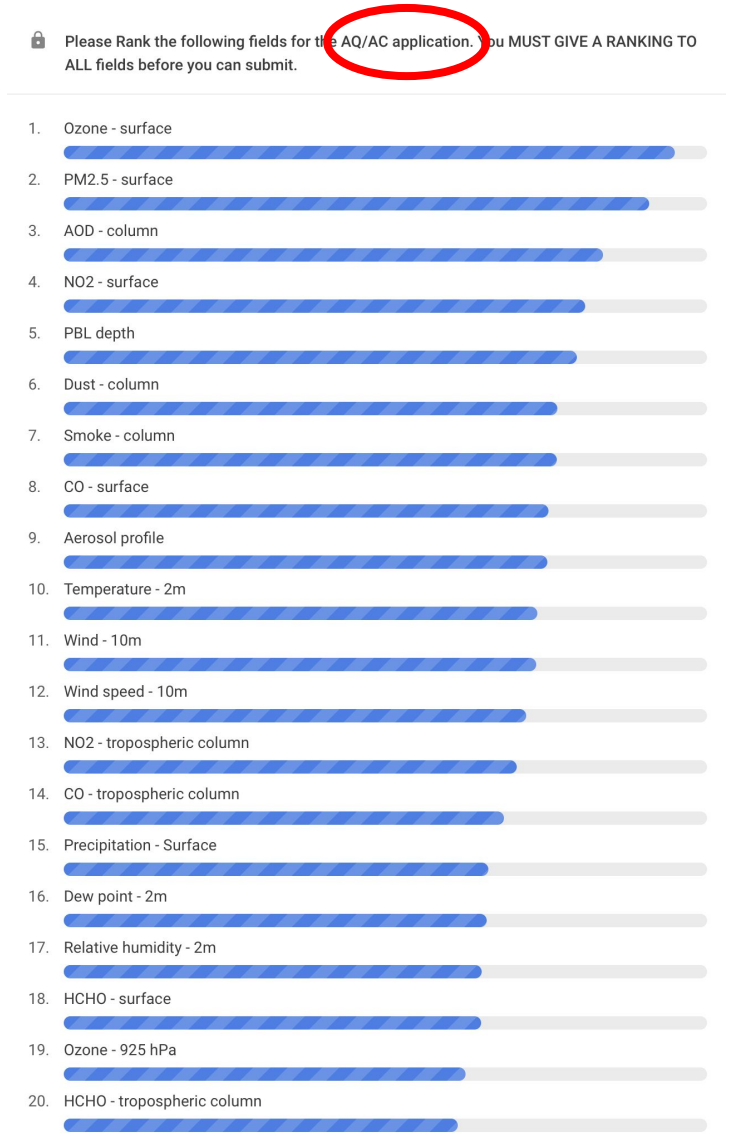
- Every field listed on the aggregate spreadsheet was added to a Slido poll for that particular application
- A full ranking was generated by the workshop participants for each survey

Live Polling at the Workshop



- Every field listed on the aggregate spreadsheet was added to a Slido poll for that particular application
- A full ranking was generated by the workshop participants for each survey
- Marine/Cryosphere was tricky, as the longer range UFS plans have this application absorbed by the MRW application; we kept this survey separate but added in some additional marine/cryosphere elements to the MRW survey (this will need future revisitiation)

Live Polling at the Workshop



- Every field listed on the aggregate spreadsheet was added to a Slido poll for that particular application
- A full ranking was generated by the workshop participants for each survey

Live Polling at the Workshop

Please Rank the following fields for Seasonal Application-UFS. You MUST GIVE A RANKING TO ALL fields before you can submit

1. Oceanic Nino Indices (ENSO)
2. Precipitation Anomaly at Surface
3. Temperature Anomaly at 2-meter
4. Geopotential Height Anomalies at 500-hPa
5. Precipitation at surface
6. Temperature at Sea Surface
7. Temperature at 2-meter
8. Geopotential Height at 500-hPa
9. Jet Stream Position & Intensity at 250-hPa
10. AO/AAO Index
11. NAO/PNA Index
12. RMM1 and RMM2 (MJO)
13. Sea Ice Concentration at Surface
14. Snow Accumulation at Surface
15. QBO Equatorial Zonal Winds at 100-10 hPa
16. Temperature Anomaly at 850-hPa
17. U/V Wind at 10-meter
18. Upper Ocean Heat Content (Temp Avg) at 0-300 meter
19. Outgoing Longwave Radiation at Top of Atmos
20. U/V Wind at 250-hPa
21. Soil Moisture at Surface
22. Standardized Precip Index
23. Basin-Wide TC Counts
24. U/V Wind at 850-hPa
25. Temperature at 850-hPa
26. Blocking Tibaldi-Molteni Idex
27. Palmer Drought Severity Index
28. Atmos River Integrated Water Vapor Transport in Column
29. Heat Wave # of Defined Based on Percentile
30. Soil Temperature at Surface
31. Monsoon Onset and End Dates
32. Temperature at 500-hPa
33. Zonal Winds at 60 N at 10 mb
34. Fire Danger Index
35. East Asian Summer Monsoon Index
36. Heat Wave # of Defined Based on Threshold
37. Streamflow at Surface
38. Vegetation at Surface
39. Energy-Helicity Index (Severe Weather)
40. Land-atmosphere Coupling (CTP-Hi)
41. Supercell Composite Parameter
42. Groundwater at Sub-surface

- Every field listed on the aggregate spreadsheet was added to a Slido poll for that particular application
- A full ranking was generated by the workshop participants for each survey

Prioritized Metrics

- Worksheets were updated to order the entries based on the polling results

FIELD	LEVEL	DETERMNISTIC METRIC	ENSEMBLE METRIC	PROBABILISTIC METRIC	TEMPORAL ATTRIBUTE	NOTABLE THRESHOLDS	REGION	VERIFICATION APPROACH	VALIDATION SOURCE	CLIMATOLOGY SOURCE	SPATIAL MATCHING	STATISTIC available in METplus?	FILE FORMAT supported in METplus?	COMMENTS	COMMENTS	RANKING
Precipitation	Surface	ETS/Bias FSS, Contingency table elements		Reliability Diagram	Every 24h	0.5", 1", 2" ?	CONUS: East/West/Central	grid-to-grid	Stage IV Analysis			YES				1
Temperature	Sfc/2-m	RMS/Bias	RMS of Ensemble Mean + Spread		Every 3h	0C	CONUS:East/West /Central/Alaska	grid-to-obs	METARS			YES				2
U/V Wind	Sfc/10-m	RMS/Bias	RMS of Ensemble Mean + Spread		Every 3h	10, 15 m/s	CONUS:East/West /Central/Alaska	grid-to-obs	METARS			YES				3
Sea Level Pressure	Surface	ACC	RMS of Ensemble Mean + Spread		Every 6h		NH+SH+Tropics	grid-to-grid	GFS Analysis			YES				4
Geopotential Height	500-hPa	ACC	RMS of Ensemble Mean + Spread	Reliability	Every 6h	540 dam, 546 dam, 588 dam	NH+SH+Tropics	grid-to-grid	GFS Analysis	ERA-5 Climo		YES				5
Dew Point	Sfc/2-m	RMS/Bias	RMS of Ensemble Mean + Spread		Every 3h	50, 60, 70 F	CONUS: East/West/Central/ Alaska	grid-to-obs	METARS			YES				6
Snowfall	Surface	ETS/Bias FSS, Contingency table elements	Performance Diagram	Reliability Diagran	Every 24h	2",6", 12"	CONUS: East/West/ Central	grid-to-grid	NOHRSC Analysis							7
U/V Wind	850-hPa	RMS, Bias	RMS of Ensemble Mean + Spread	Reliability	Every 6h	15 m/s, 25 m/s, 35 m/s (Anomalies, ideal	NH+SH+Tropics	grid-to-obs	RAOBS + Aircraft Data			YES				8
Temperature	850-hPa	RMS, Bias	RMS of Ensemble Mean + Spread	Reliability	Every 6h	0C, 10C, 20C, 30C	NH+SH+Tropics	grid-to-obs	RAOBS + Aircraft Data			YES				9
TC Intensity	Surface	Average kt	Avg Error of Ensemble Mean + Spread		Every 6h	no distinction needed between hurricanes and weaker storms	All Basins (Atlantic, East Pac, Central Pac, West Pac, North Indian Ocean, SH)	grid-to-obs	Best Track					Need to figure out how to verify tracks and intensity of storms in longer-range forecasts that haven't yet formed; it won't get verified if it doesn't exist at the start		10
Temperature	Sea Surface	RMS + Mean Error Bias	RMS of Ensemble Mean + Spread	BSS, CRPSS	Every 3h and Daily	0C, 1C, 26.5C	NH+SH+Tropics	grid-to-grid	GHRSST							11
Absolute Track Error	Surface	Average NM	Avg Error of Ensemble Mean + Spread		Every 6h	no distinction needed between hurricanes and weaker storms	All Basins (Atlantic, East Pac, Central Pac, West Pac, North Indian Ocean, SH)	grid-to-obs	Best Track					Dissipation - Must give a model penalty for continuing a TC after it has dissipated or dissipated even if it continues on nature		12
Geopotential Height Anomalies	500-hPa	ACC	RMSE of Ensemble Mean + Ensemble Spread		Daily (more important at extended ranges)		NH + SH + Tropics	grid-to-grid	GFS Analysis	ERA-5 Climo				added based on late request		13
Specific Humidity	850-hPa	RMS, Bias	RMS of Ensemble Mean + Spread	Reliability	Every 6h	5 g/kg, 10 g/kg	NH+SH+Tropics	grid-to-obs	RAOBS + Aircraft Data			YES				14
											suggested spatial matching of 50 or					

MRW Application shown here

Tiered Metrics

	FIELD	LEVEL
2	TIER 1	
3	Precipitation	Surface
4	Temperature	Sfc/2-m
5	U/V WInd	Sfc/10-m
6	Sea Level Pressure	Surface
7	Geopotential Height	500-hPa
8	Dew Point	Sfc/2-m
9	Snowfall	Surface
10	U/V Wind	850-hPa
11	Temperature	850-hPa
12	TC Intensity	Surface
13	Temperature	Sea Surface
14	Absolute Track Error	Surface
15	Geopotential Height Anomalies	500-hPa
16	Specific Humidity	850-hPa

	TIER 2	
18	TC Genesis	Surface
19	CAPE	Mixed Layer
20	Precipitation Anomaly	
21	Temperature	700-hPa
22	Temperature Anomaly	2-meter
23	U/V Wind	250-hPa
24	Specific Humidity	700-hPa
25	TC Intense Precipitation	Surface
26	Cross Track Error	Surface
27	Along Track Error	Surface
28	Maximum extent of winds (wind radii)	10 m
29	Significant Wave Height	Surface
30	Wind Shear	850-200 hPa
31	Latent Heat Flux	Surface

	TIER 3	
33	Sensible Heat Flux	Surface
34	Radius of maximum winds (RMW)	10 m
35	Wind Speed	Marine Surfac
36	U/V Wind	200-hPa
37	Wind Direction	Marine Surfac
38	Wave Height from TCs	Surface
39	Soil Moisture	0-10 cm
40	Sea-Ice Concentration	Surface
41	Sea-Ice Thickness	Layer depth
42	Temperature	200-hPa
43	Soil Temperature	0-10 cm
44	Sea-Ice Drift / Velocity	
45	Soil Moisture	10-40 cm
46	Ozone	925-hPa

MRW Application results shown here

- The metrics were divided into different tiers (roughly by thirds) to loosely represent potential different gateways in the R2O development process
- The initial generation of tiers was based on absolute survey ranking

Tiered Metrics

- Inspection of the data, however, revealed that the fields from certain “disciplines” (categories) were ranked much higher than those from others, leaving out representation from all user and developmental communities in the tiered approach

MRW Application	# of Metrics Pre-Workshop	# of Metrics Post-Breakout	# of Metrics in “Tier 1"	# of Metrics in “Tier 2”	Mean Rating (44 total)	Median
Upper Level	7	12	5	3	21.2	20
Sfc-Sensible Wx	6	9	6	3	8.3	6
Marine / Cryosphere	4	7	1	1	30.6	33
Land-Sfc / Hydro	6	11	0	1	35.8	37
Tropical	10	10	2	6	22.2	24

Alternate Approach to Tiered Metrics

	TIER 1	
3	Precipitation	Surface
4	Temperature	Sfc/2-m
5	U/V WInd	Sfc/10-m
6	Sea Level Pressure	Surface
7	Geopotential Height	500-hPa
8	Dew Point	Sfc/2-m
9	Snowfall	Surface
10	U/V Wind	850-hPa
11	Temperature	850-hPa
12	TC Intensity	Surface
13	Temperature	Sea Surface
14	Absolute TC Track Error	Surface
15	Geopotential Height Anomalies	500-hPa
16	Specific Humidity	850-hPa

need to add representation for land-sfc and marine/cryosphere disciplines in tier 1

two fields get "bumped" to tier 2 to make room in tier 1 for the moved fields

MRW Application

	TIER 2	
18	Geopotential Height Anomalies	500-hPa
19	Specific Humidity	850-hPa
20	TC Genesis	Surface
21	CAPE	Mixed Layer
22	Precipitation Anomaly	
23	Temperature	700-hPa
24	Temperature Anomaly	2-meter
25	U/V Wind	250-hPa
26	Specific Humidity	700-hPa
27	TC Intense Precipitation	Surface
28	TC Cross Track Error	Surface
29	Significant Wave Height	Surface
30	Sensible Heat Flux	Surface

- This approach requires that each discipline be represented in each tier; if a discipline was not represented in a tier, the highest-ranked field from that category was moved up in the rankings
- This approach to generating a list of tiered prioritized metrics was the best-received during V&V Working Group meetings
- Of course, not all of the UFS applications have clear-cut categories like SRW, MRW, and Seasonal do

SRW Synthesis

SRW Application	# of Metrics Pre-Workshop	# of Metrics Post-Breakout	# of Metrics in "Tier 1"	# of Metrics in "Tier 2"	Mean Rating (47 total)	Median
Synoptic	5	9	3	6	18.9	20
Sfc-Sensible Wx	8	8	6	1	6.9	5
Severe / Winter	5	8	4	3	16.4	16
Land-Sfc / Hydro	6	9	0	0	40.1	42
Aviation	6	5	0	4	27.4	28
Air Quality	4	9	1	1	34.3	35

- Sfc/Sensible Weather fields were ranked very high
- Land-Sfc, Aviation, and AQ fields were ranked lower; generating tiers with required representation from all disciplines was extremely useful to properly account for verification needs in these areas
- New fields added during the breakout sessions had a mean rating of 22; two of them ended up in the top 10
- Some of the severe weather fields (like updraft helicity) were ranked surprisingly low, likely reflecting the growing diversity of the hi-res modeling customers
- Still room to parse the disciplines further (make land-sfc and hydro separate?)

Further Steps in the Data Synthesis

- The breakout groups did a remarkable job resolving many issues, but some gaps on the spreadsheets remain (especially for new fields added), and those need to be addressed
- The plan is to bring in small teams of subject matter experts to fill in those items
- The experts will also review the spreadsheets to identify any potential mistakes (i.e. selection of a verification data set that is inconsistent with the region chosen for verification)
- Potential addition parsing of the “disciplines” to make sure that all components of model development and customer need are properly represented in the lists of prioritized metrics

Verification, Validation, and Evaluation at EMC

Performance
- Monitor real-time model performance
- Report statistics to Congress and HQ

Evaluation
- Evaluate both parallel and real-time models
- Contribute to evaluation reports

Research
- Develop new metrics and indexes
- Create new metrics software and technology with METplus

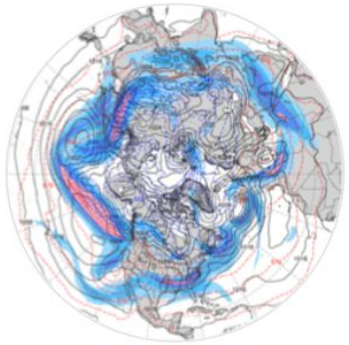
Community
- Support the Unified Forecast System community
- Involve partners and stakeholders in evaluations

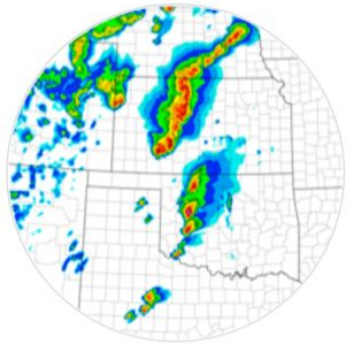
Courtesy:
Jason Levit
Chief, VPPPGB

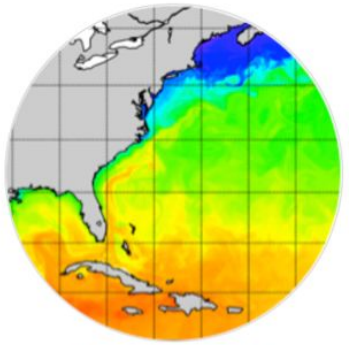
EMC's New(ish) Verification Page

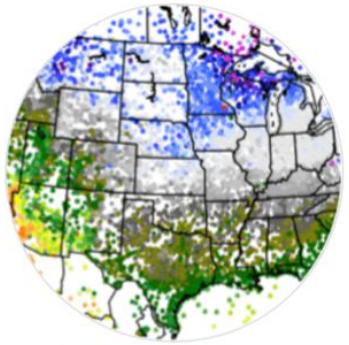
www.emc.ncep.noaa.gov/users/verification

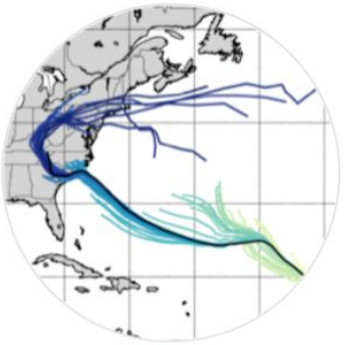
EMC Home / EMC Verification

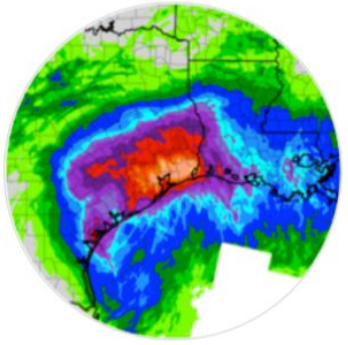

Global Models

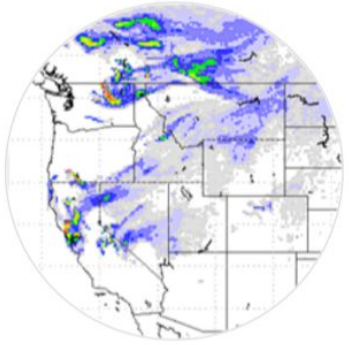

Regional/Hi-Res Models


Ocean/Wave Models


Real-time Analyses


Cyclones


Precipitation/Clouds

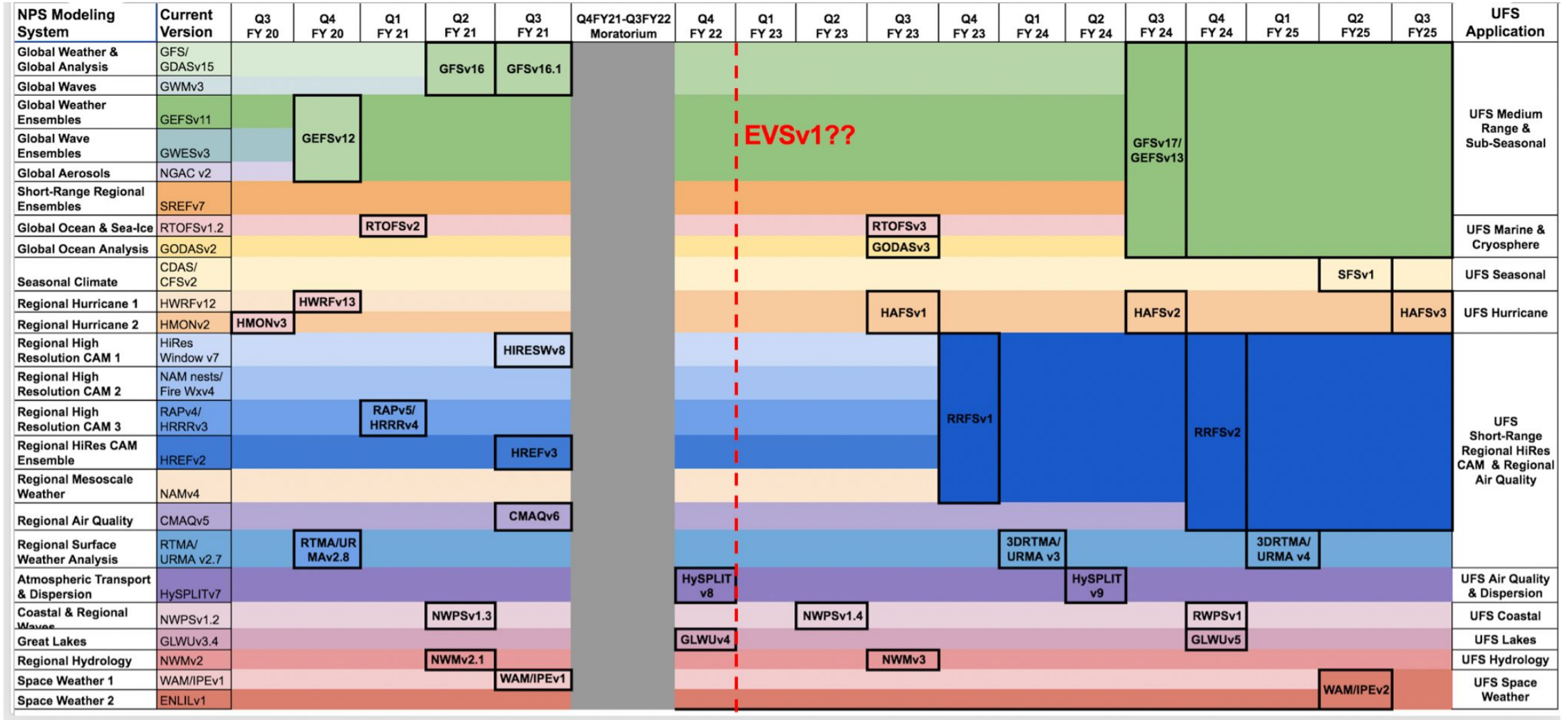

Air Quality

- A one-stop shopping site for all EMC verification images
- No tables!
- Common plotting tools and standards
- More images and organization to come!

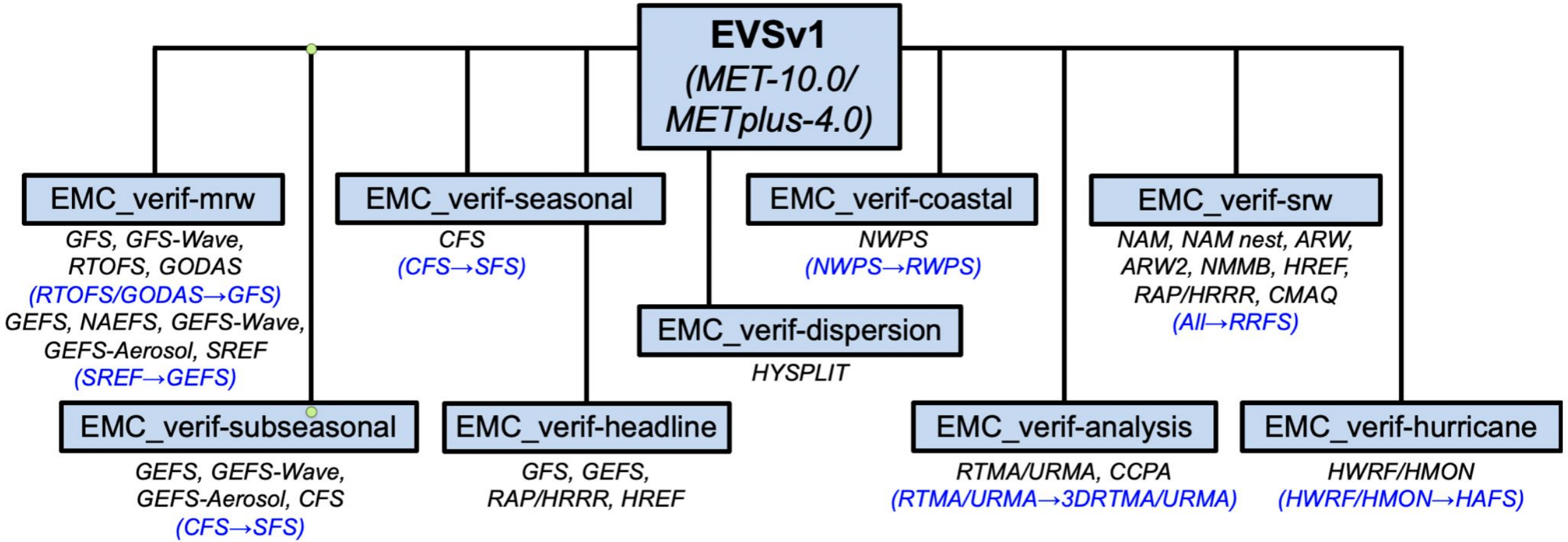
EMC Verification System

- EMC is planning to build an operational EMC verification system (EVS)
- This is part of a transition from in-house, custom software to all verification being performed using the community-based METplus code
- The EVS will produce real-time stats and graphics for the EMC verification website
- The EVS will allow EMC to share and use community code
- The EVS will be organized around model (UFS application)

UFS Rainbow Diagram



Initial EVS Planning Schematic



- Organized by UFS applications which gives EMC the critical flexibility to easily swap models in and out with the contraction / evolution of the NCEP Production Suite

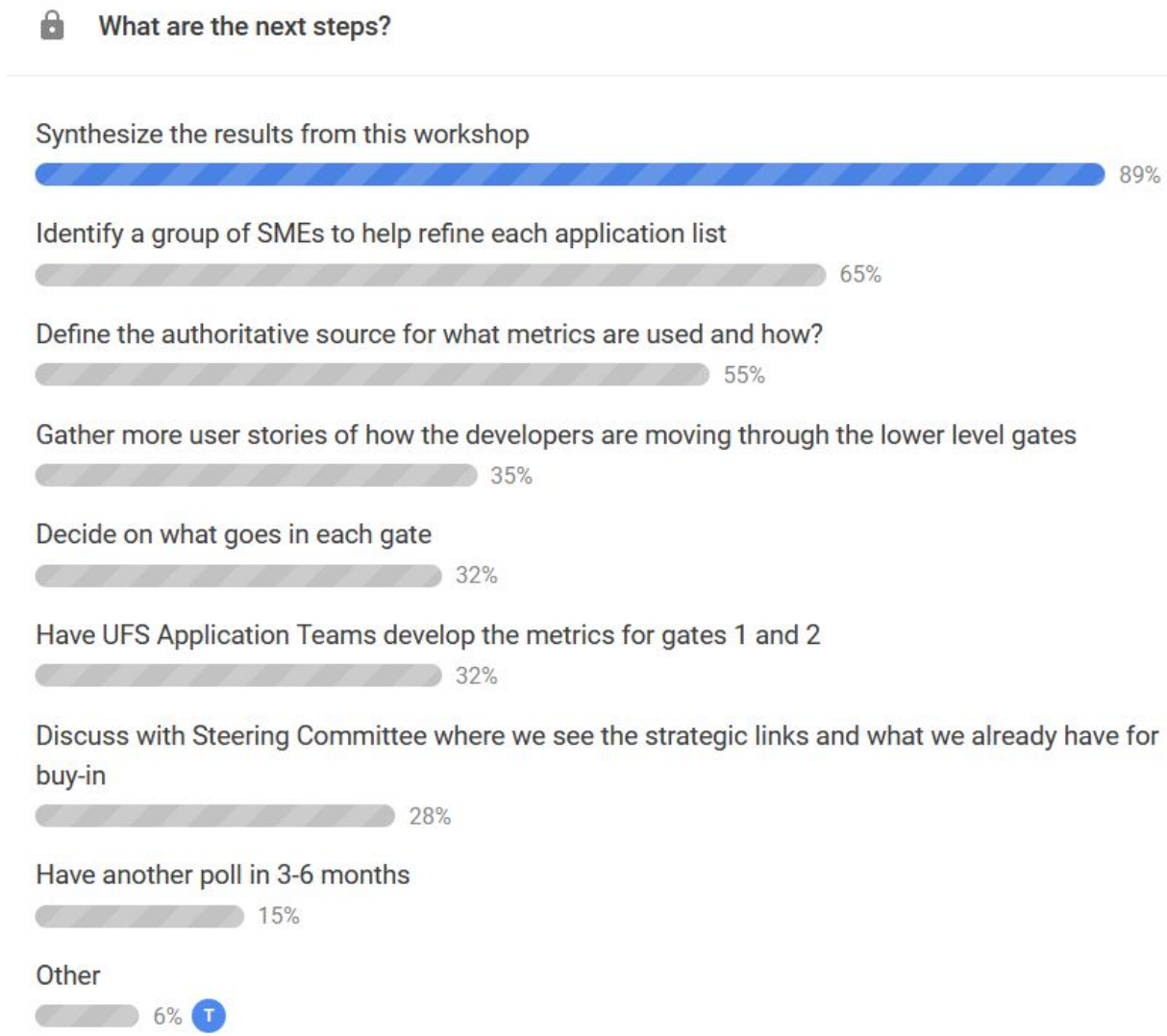
EMC Plans for the Workshop Data

- The lists of prioritized metrics will be used to build Version 1 of the EVS
- Will work with the MET team to identify gaps in the code so that any missing metrics or issues with the ability of the MET code to read and process an observational data set can be resolved
- The prioritized metrics lists will be used by EMC to generate test plans for formal evaluations of future upgrades of UFS components

Next Steps

- Formulate SME teams to help fill in the candidate lists
- Review Metrics to Identify Gaps
- Publish up to DTC and UFS websites
- Implement currently unsupported capability (R20 Year 2-4)

Next Steps - Slido



Status	
Reporting progress here	
Looking for volunteers	
Requested guidance from UFS SC	
To be done	
First attempt reported here	
Requested via UFS SC	
Requested guidance from UFS SC	
TBD after prioritized lists developed	

Next Steps

- METplus Release
  - Next METplus release - version 4.0.0 end of April 2021
  - Much of the additional development work identified by this workshop will be available in following release - version 4.1 slated for the end of December 2021 and version 5.0 in September 2022

- METplus AMI to explore prototypes
  - AWS Machine Image has beed developed
  - Intended use: NOAA Big Data Project where UFS Prototypes are available to the community
  - Status: Beta-testing, documentation, and training development
  - Expected availability to community: Mid-May
  - Want to get a jump on that? Volunteer to be a beta-tester

Thanks Again

- UFS V&V Cross-Cutting Team - Chairs and Contributors:
  - Geoff Manikin (EMC), Tara Jensen (NCAR/RAL and DTC), Jason Otkin (UW-Madison CIMSS), Dave Turner (GSL), Mike Baldwin (Purdue), Matt Janiga (NRL)
  - **Burkely Gallo (SPC),** Jason Levit (EMC), Jack Settlemaier (SRH), Sarah Lu (U Albany), Linden Wolf (OSTI), Deepthi A (OSTI), Yan Xue (OSTI), and Cristiana Stan (GMU)
- EMC MEG and Verification Teams:
  - Perry Shafran, Mallory Row, Alicia Bentley, Logan Dawson, Shannon Shields, Chris MacIntosh, Marcel Caron
- All of the Breakout Facilitators:

Ivanka Stajner, EMC, Perry Shafran, EMC, Logan Dawson, EMC, John Opatz, DTC, Malaquias Peña, UConn, Lindsay Blank, DTC, Mike Ek, DTC Dan D'Amico, DTC, Barb Brown, DTC Lisa Goodrich, DTC, Michelle Harrold, DTC, Mike Baldwin, Purdue, Ligia Bernardet, DTC, Kathryn Newman, DTC, Mark Klein, WPC, Paul Dirmeyer, GMU, Dan D'Amico, DTC, Curtis Alexander, GSL, Alicia Bentley, EMC, Mike Baldwin, Purdue, Julie Prestopnik, DTC, Tatiana Gonzalez, NWS/AFS, Mrinal Biswas, DTC, Xia Sun, DTC, Matt Rosencrans, CPC, Weiwei Li, DTC, YJ Kim, NWS/AFS, Matthew Janiga, NRL, Ben Albright, WPC, Jamie Wolff, DTC, Logan Dawson, EMC, Will Mayfield, DTC, Jon Gottschalck, CPC, Ricky Rood, U Michigan, Hendrik Tolman, NOAA

- DTC administrative support and UCAR Multimedia Services

For More Information

- **DTC UFS Evaluation Metrics Website:** https://dtcenter.org/events/2021/2021-dtc-ufs-evaluation-metrics-workshop
- **Contact:**
  - jensen@ucar.edu
  - geoffrey.manikin@noaa.gov
  -
- **Look for updates:**
  - Webinar for Attendees (May-June timeperiod
  - 2022 AMS Probability/Stats and R2O Meetings

- **DTC Visitor Program:** https://dtcenter.org/visitor-program
- **DTC Workshop - June 7-9, 2021:** http://dtcenter.org/events/workshop/2219

INTEGRATING CLOUD AND CONTAINER TECHNOLOGIES INTO UNIVERSITY NUMERICAL WEATHER PREDICTION (NWP) CURRICULUM