Validation and Calibration of an Observing Systems Simulation Experiment (OSSE) System using a Summary Assessment Metric (SAM) Inter-comparison of OSSE and Observing System Experiment (OSE) Results

Kayo Ide [1,2], Sid Boukabara [3], Ross Hoffman [2],
Yan Zhou [2], Narges Shahroudi [3,4], Kevin Garrett [3],
V. Krishna Kumar [3.4], Tong Zhu [5], Robert Atlas [6]

(1) University of Maryland, (2) UMD-CICS
(3) NOAA NESDIS STAR
(4) Riverside Technology Inc, (5) CSU-CIRA, (6) NOAA AOML

Background: OSSE and CGOP

- OSSE can be used to study observation impact for ”what if” scenarios, e.g.,
  - Loss of existing satellites: current vs 3-polar, 2-polar, 1-polar
  - Addition of hypothetical instruments: + small satellite instruments

  to support decision-makers with quantitative assessments of expected/proposed observing systems & readiness for new sensors

- Community Global OSSE Package (CGOP) with GMAO G5NR
  - Needs for a modular extensible framework for conducting OSSEs
  - Towards advancement of the theory and practical application of OSSEs

[Boukabara et al. 2016a, 2018a,b]

Underlying assumption:
OSSE/CGOP is realistically capable of assessing observation impact relevant to real global NWP

Objective of This Study

- OSSE can be used to study observation impact for ”what if” scenarios, e.g.,
  - Loss of existing satellites: current vs 3-polar, 2-polar, 1-polar
  - Addition of hypothetical instruments: + small satellite instruments

  to support decision-makers with quantitative assessments of expected/proposed observing systems & readiness for new sensors

- Community Global OSSE Package (CGOP) with GMAO G5NR
  - Needs for a modular extensible framework for conducting OSSEs
  - Towards advancement of the theory and practical application of OSSEs

[Boukabara et al. 2016a, 2018a,b]

Objective of this study:

- Validate and calibrate OSSE using CGOP for realistic global NWP
- using Summary Assessment Metrics (SAMs)

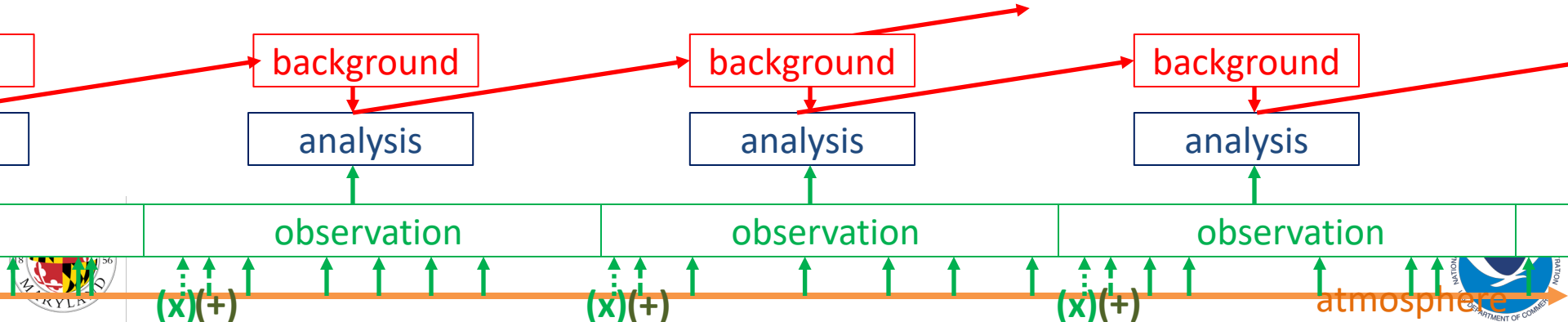
Practical Objective of Impact Assessment

- Impact assessment = comparative study w.r.t. a control

(x) Loss of existing observing systems

(+) New/hypothetical observing systems

		Real DAS	Synthetic DAS
		Obs. Sys. Exp. (OSE)	Obs. Sys. Sim. Exp. (OSSE)
NWP elements	Atmosphere	Real & unknown	Simulated NR & known
	Observations	Real (x)	Simulated (x) & (+)
	Model	Operational	(maybe simplified)
	DA method	Operational	(maybe simplified)



Approach: Inter-comparison of OSE and OSSE

- Impact assessment = comparative study w.r.t. a control

(x) Loss of existing observing systems: existing vs 3-polar, 2-polar, 1-polar

(+) New/hypothetical observing systems: + small satellite instruments

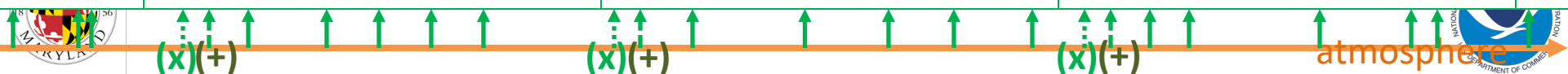
		Real DAS	Synthetic DAS
		Obs. Sys. Exp. (OSE)	Obs. Sys. Sim. Exp. (OSSE)
NWP elements	Atmosphere	Real & unknown	Simulated NR & known
	Observations	Real (x)	Simulated (x) & (+)
	Model	Operational	(maybe simplified)
	DA method	Operational	(maybe simplified)

Validation for "what if" (loss of existing observing systems) scenario

- inter-comparison of impact assessment: OSE for (x) ←→OSSE for (x)

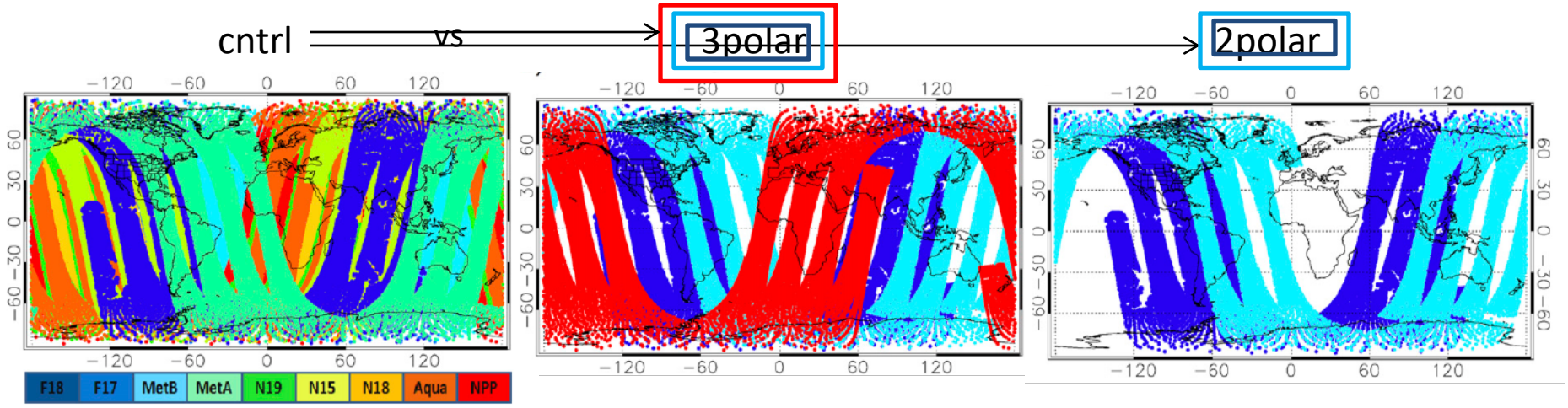
Calibration

- Expected impact for real NWP for (+) ← OSSE for (+)



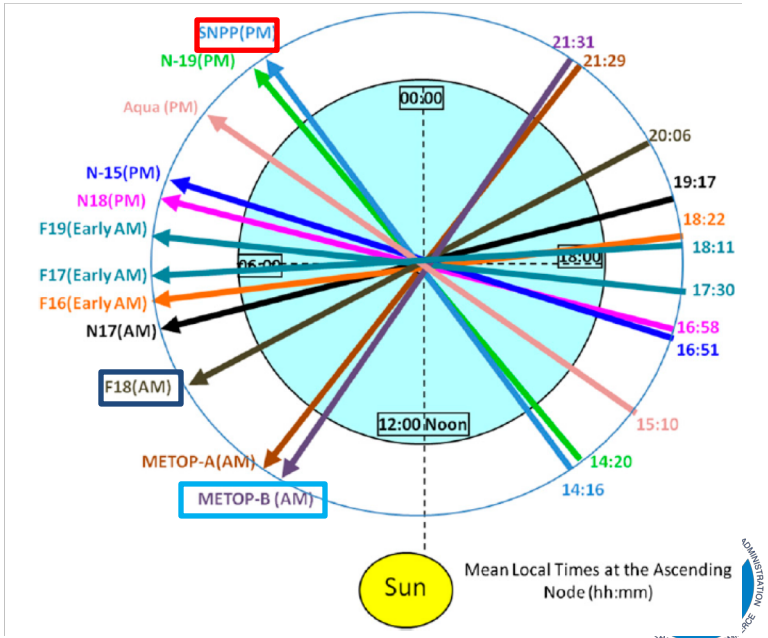
Experimental Setup: Data Gap Scenario ("What If" Scenario)

- Inter-comparison: OSSE vs OSE



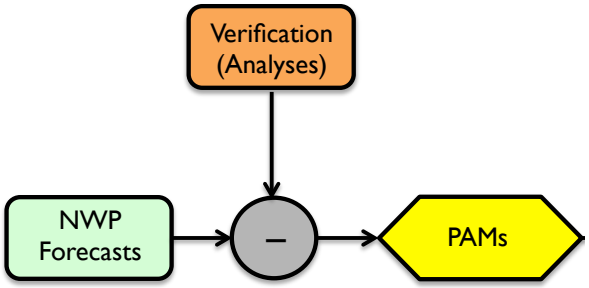
Period	7 July to 7 August, 2014
Obs system config.	2014: all conventional + satellite data gap scenarios
	OSSE only: perfect obs
DAS	3DEnVar with T670/T254 80mem
Forecast	0000UTC daily up to 168h
Verification	OSSE/OSE: own cntrl analysis
	OSSE only: G5NR

[Original OSE work by Boukabara et al (2016b)]



Impact Assessment: Conventional Approach

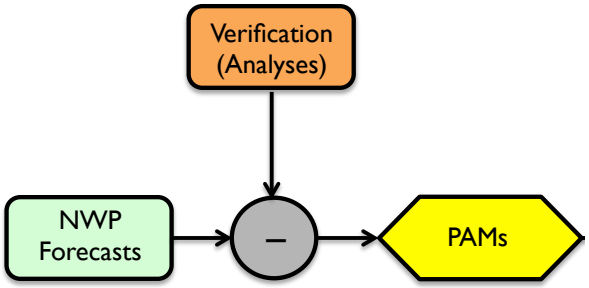
- Primary Assessment Metrics (PAMs) = conventional metrics
  - variable :: height (Z), temperature (T), wind (V);
  - statistic :: AC, RMSE, AME (absolute value of bias)
  - domain :: NHX, SHX, tropics
  - Vertical level :: 250, 500, 700, 850, 1000 hPa
  - forecast time :: 24, 48, 72, 96, 120, 144, 168 h (00h = analysis)



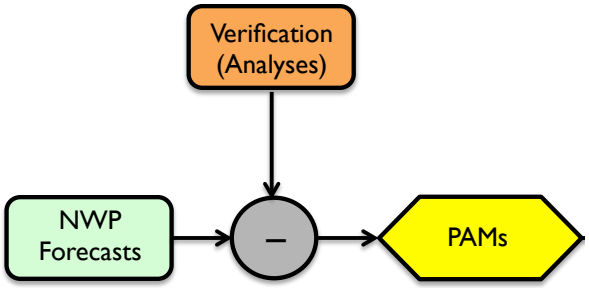
Impact Assessment: Conventional Approach

- Primary Assessment Metrics (PAMs) = conventional metrics
  - variable :: height (Z), temperature (T), wind (V);
  - statistic :: AC, RMSE, AME (absolute value of bias)
  - domain :: NHX, SHX, tropics
  - Vertical level :: 250, 500, 700, 850, 1000 hPa
  - forecast time :: 24, 48, 72, 96, 120, 144, 168 h (00h = analysis)

OSE



OSSE

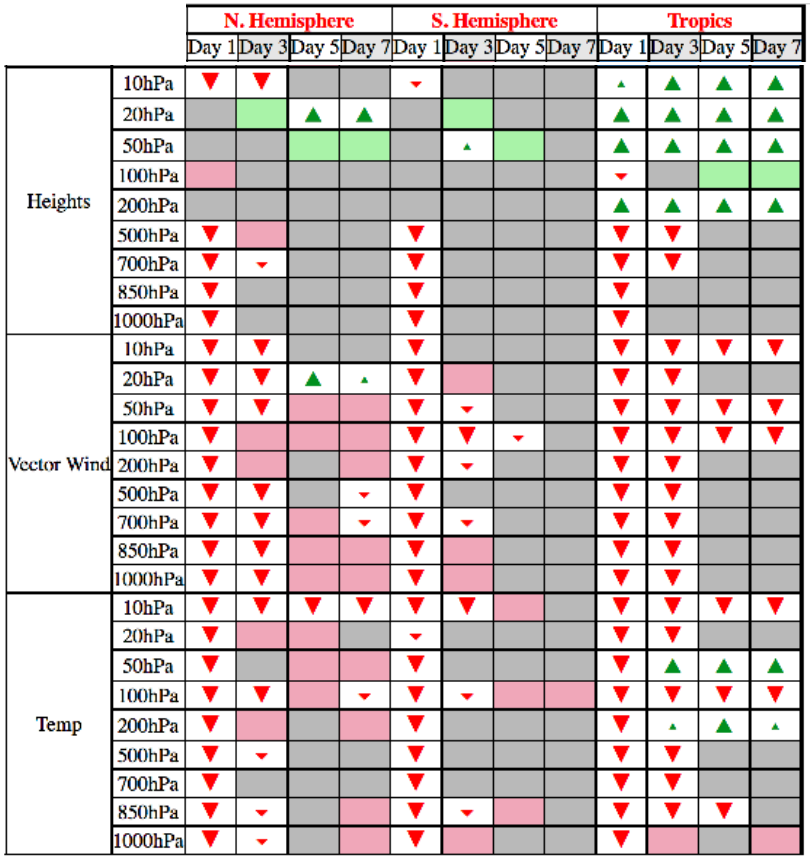


For consistency of inter-comparison, both OSE and OSSE use own cntrl-analysis for verification

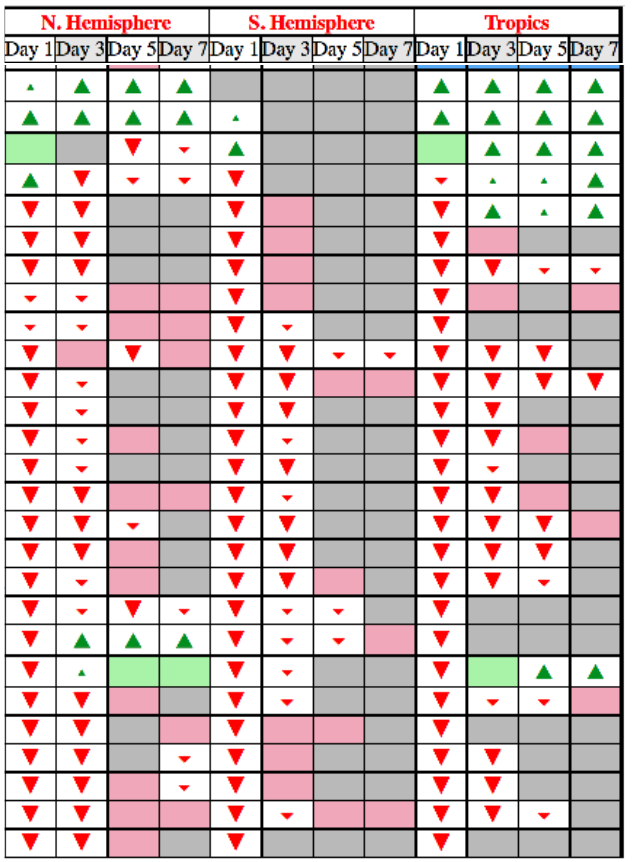
Comparison of impact assessment

Results: Scorecards (PAM - RMSE)

OSSE: 3polar vs cntrl



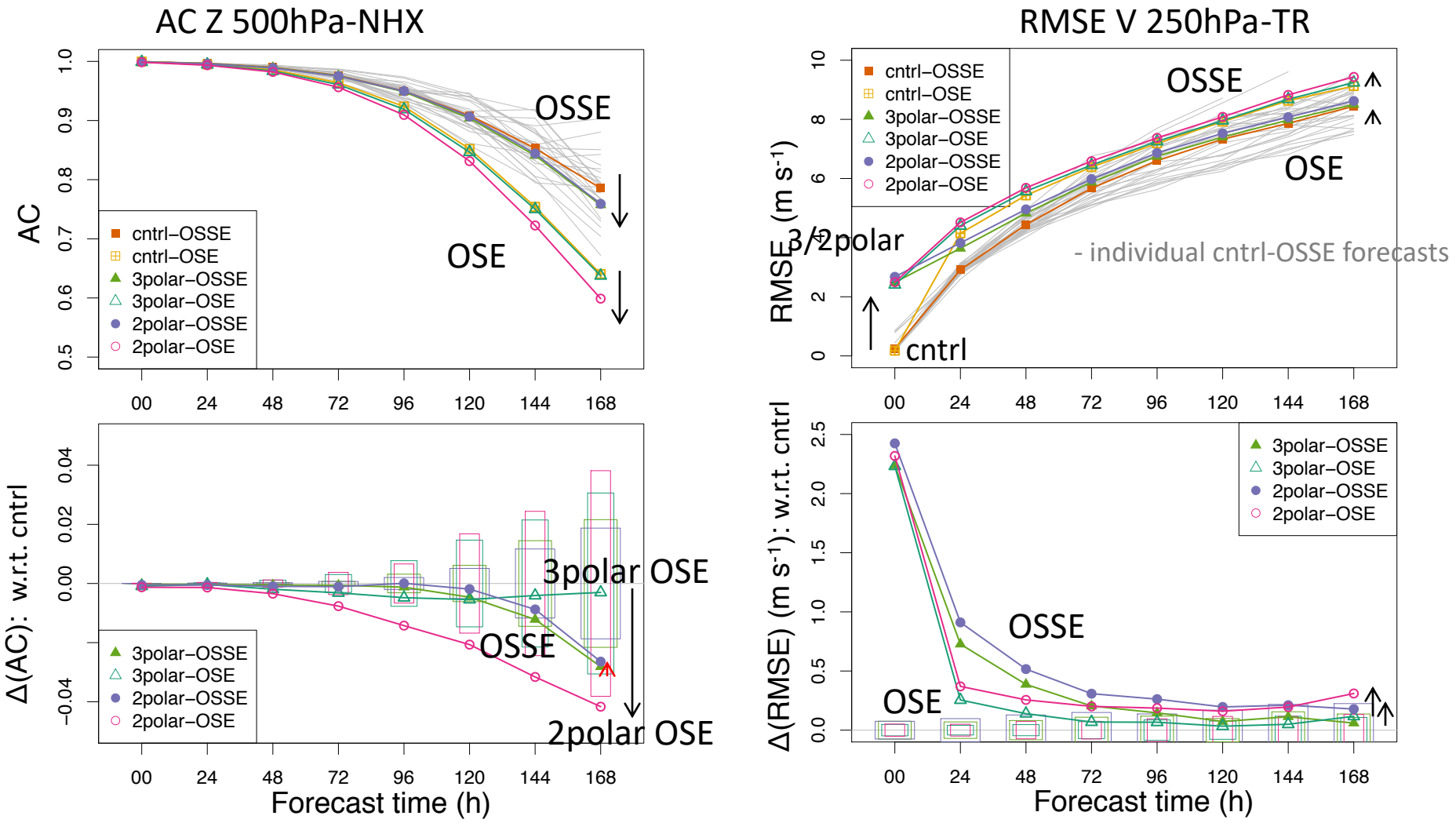
OSE: 3polar vs cntrl



3polar worse than cntrl: $0, \blacktriangledown, 0.001, \blacktriangledown, 0.01,$ ▁ $, 0.05,$ ▒ $, 0.95,$ ▇ $, 0.99, \blacktriangle, 0.999, \blacktriangle, 1$: 3polar better than cntrl

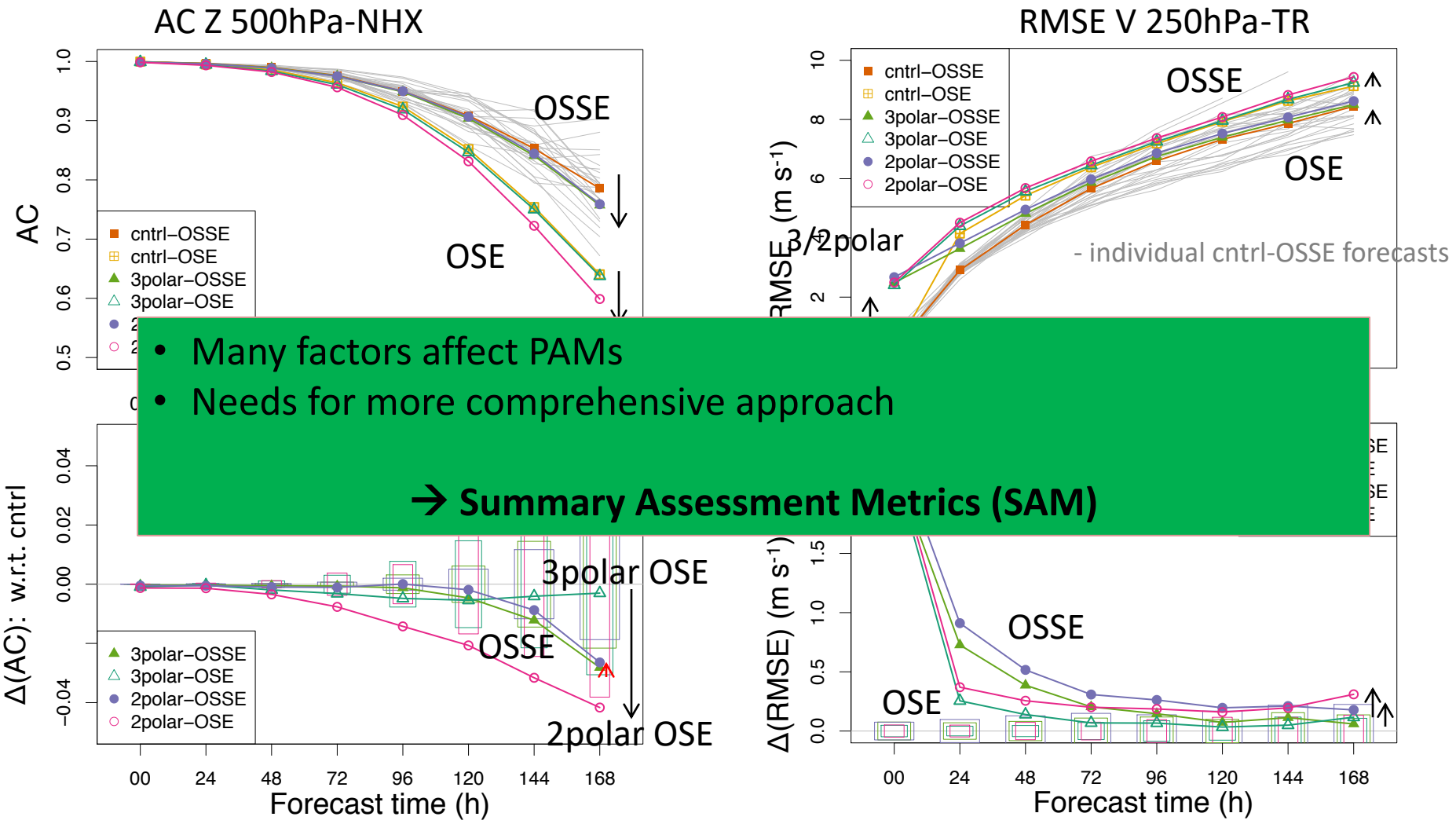
- OSSE and OSE qualitatively show similar impact
- The 3polar scenario degrades the forecast skill, except for Z in high atmosphere
- The 2polar scenario shows consistent results (not shown)

Results: Forecast Skill (PAM - AC & RMSE)



- All data gap scenarios result in poor forecast skills
- Tendency of impact mostly as expected [Δ(AC) for OSSE is statistically insignificant] although there are bit of variabilities in OSSE vs OSE inter-comparison results

Impact Assessment: Forecast Skill (PAMs: AC & RMSE)



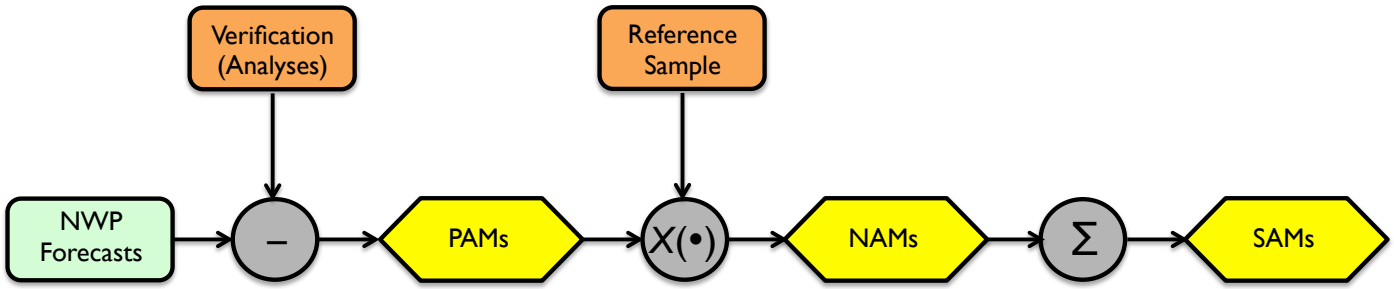
- Many factors affect PAMs
- Needs for more comprehensive approach

→ **Summary Assessment Metrics (SAM)**

- All data gap scenarios result in poor forecast skills
- Tendency of impact mostly as expected [Δ(AC) for OSSE is statistically insignificant] although there are bit of variabilities in OSSE vs OSE inter-comparison results

Results: Impact Assessment Method

- Primary Assessment Metrics (PAM): individually evaluated for each



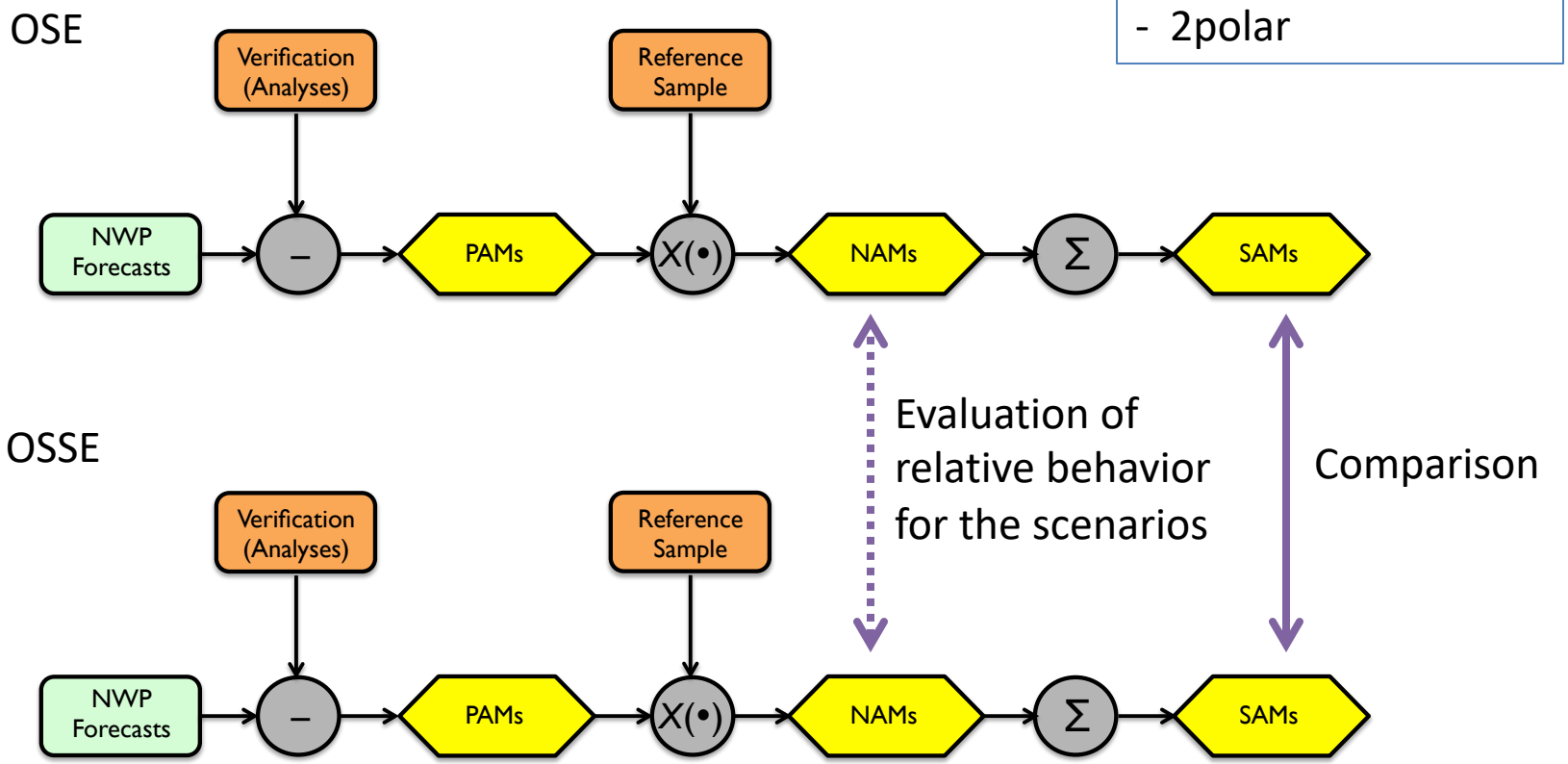
- Normalized Assessment Metrics (NAM) $\subset [0,1]$ = [worst, best]
  - Empirical Cumulative Density Function (normalized by its rank)
  - Min-Max (normalized by its min and max)

- Summary Assessment Metrics (SAM): summative assessment
  - Total SAM
  - Stratified SAMs
    - Categories
    - Forecast skills

Approach: Inter-comparison

- Comparison of summative impact assessment

"What if" scenarios:
- cntrl
- 3polar
- 2polar

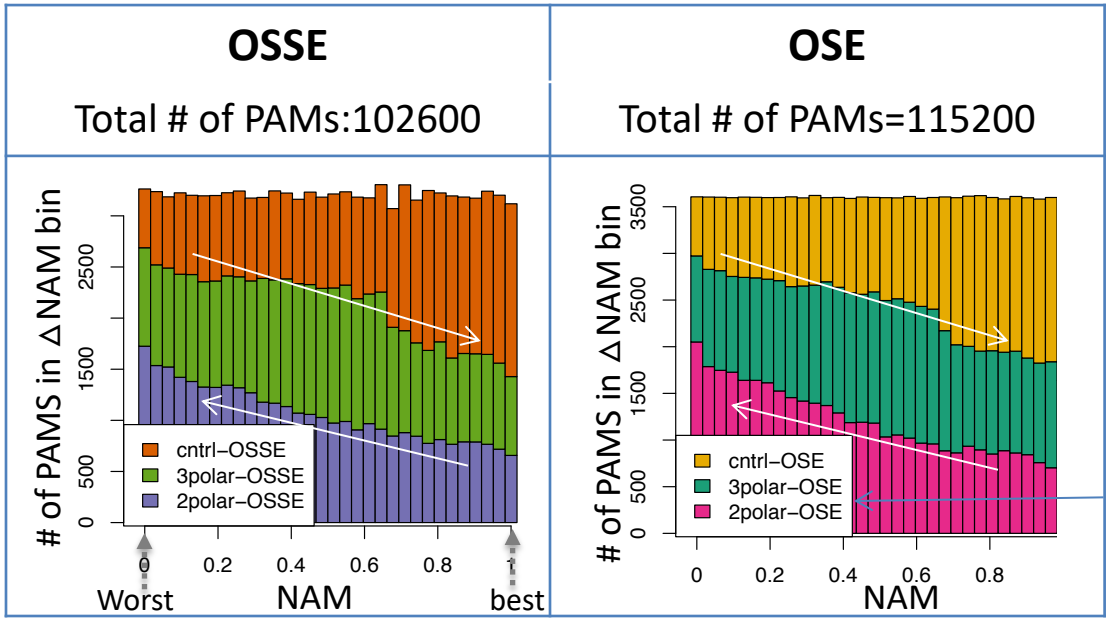


Evaluation of NAM behavior for OSSE vs OSE

- NAMs: evaluation of relative behavior (among scenarios) for OSSE and OSE
  - ECDF

$$\text{NAM} = \frac{\text{rank(PAM in } R) - 1}{\text{size}(R)}$$

R : reference sample of PAM



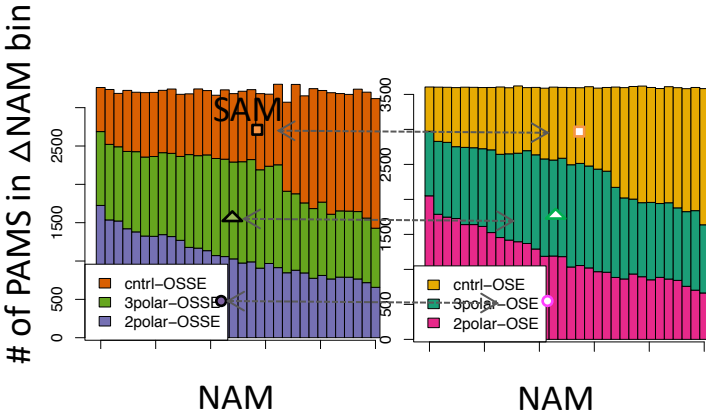
"What if" scenarios

NAMs of 3 scenarios behave in a similar manner for OSSE and OSE

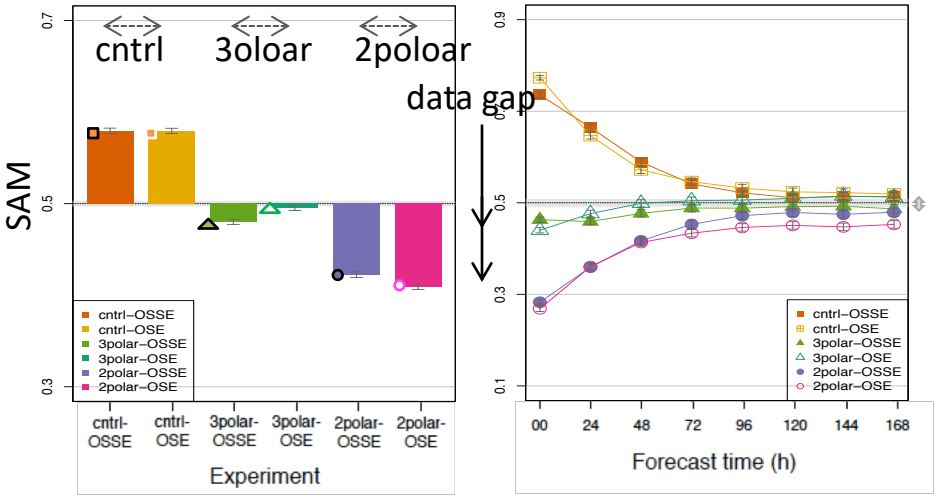
OSE vs OSSE Inter-comparison (ECDF) : Total SAMs

- ECDF SAMs

SAM = (weighted) average of NAM



Total no of PAMs:10^5

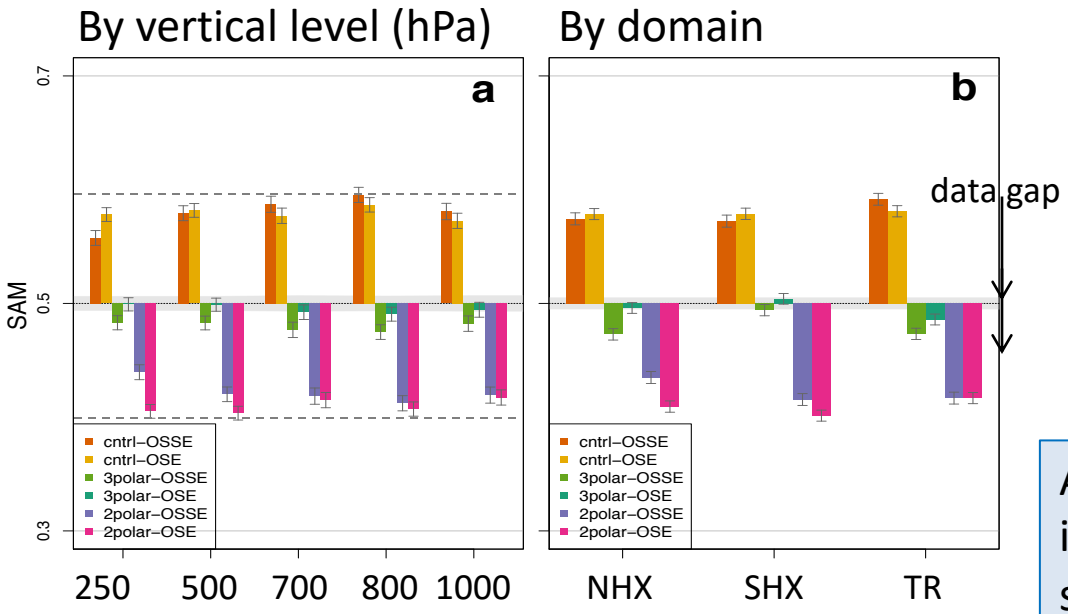


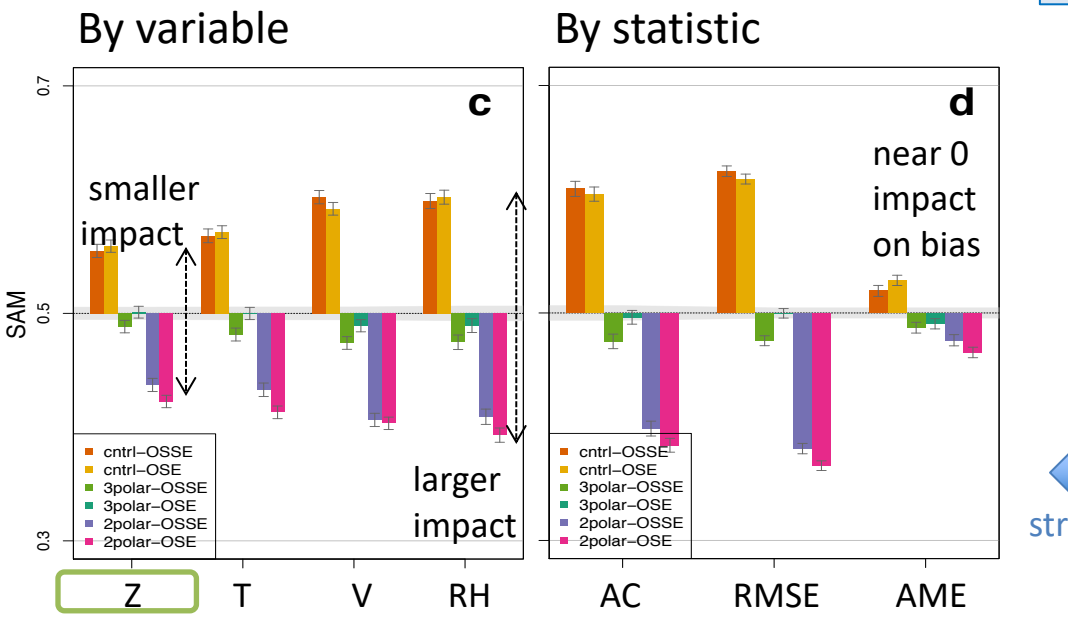
Inter-comparison shows the summative impact for OSSE and OSE is quite similar & statsitically consistent.

Properties of SAMs

- The impact is defined as the difference between the calculated SAM and its expected value under the null hypothesis (H_0), i.e., 0.5 for ECDF.
- ECDF SAM values of 0.75/0.25: very large positive/negative impacts.

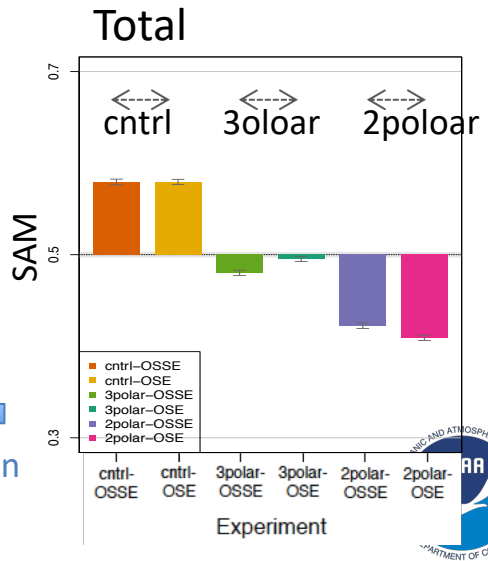
OSE vs OSSE Inter-comparison: Stratified SAMs (ECDF)





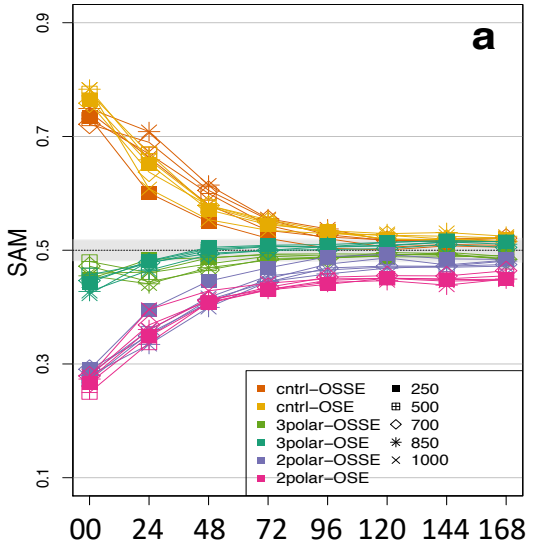
Across all categories, data gap impact on OSSE and OSE are very similar

stratification

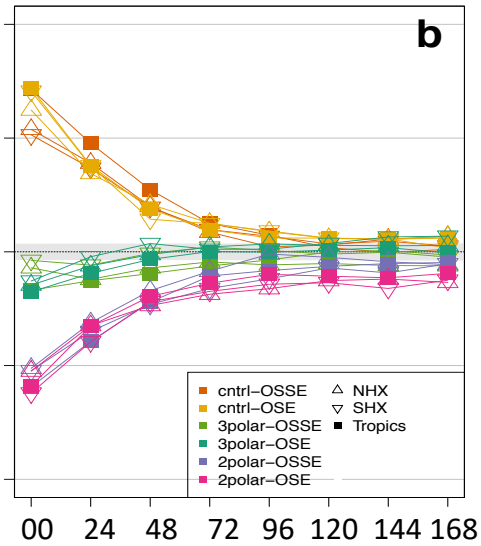


OSE vs OSSE Inter-comparison on Forecast: Stratified SAMs (ECDF)

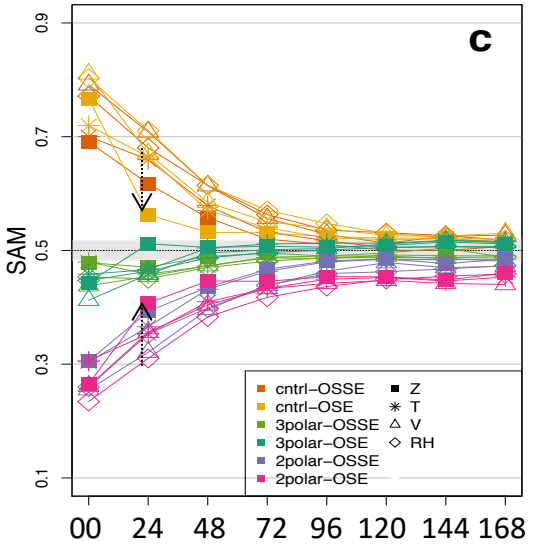
By vertical level (hPa)



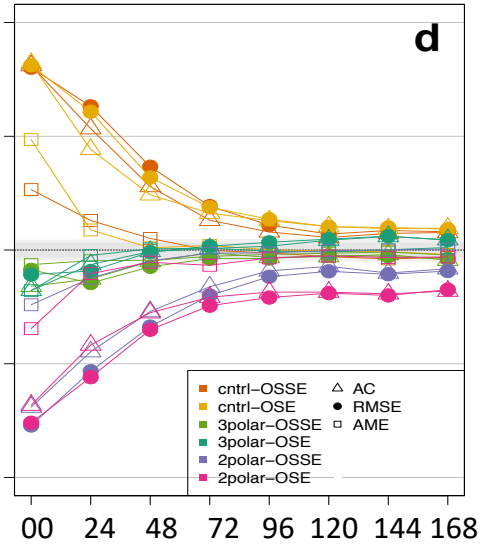
By domain



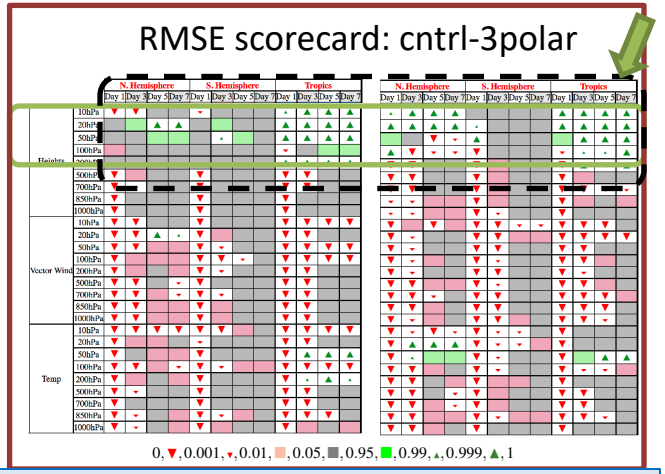
By variable



By statistic

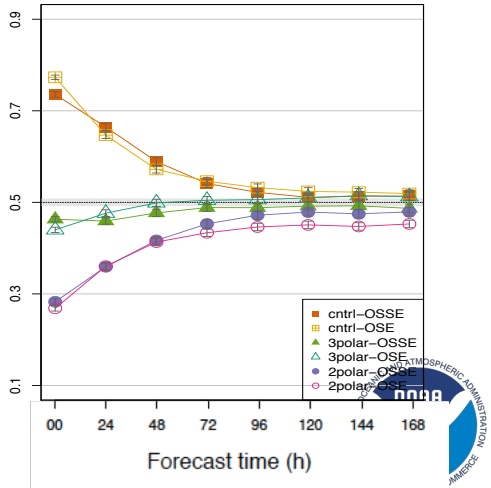


RMSE scorecard: cntrl-3polar



- Data gap impact are very similar
- cntrl-OSE Z impact are reduced in magnitude at 24h (→ behavior in RMSE scorecard)

stratification



Concluding Remarks

- An approach to validate and calibrate the OSSEs for real NWP is proposed
  - Based on the OSSE–OSE inter-comparison.
  - Using the Summary Assessment Metrics (SAMs)

- Consistent inter-comparison offers a way to suppress the apparent shortcomings of OSSE, i.e., unknown observation error characteristics (the same inter-comparison can be applied to any consistent obs. config)

- A posteriori calibration of the results
  - determines adjustments that make the parallel OSSEs and OSEs similar
  - then applies the same adjustments to the other OSSEs for real NWP

- SAMs are effective assessment metrics for
  - Validation of OSSE to OSE by inter-comparison
  - Calibration of OSSE results to real NWP impact assessment

References

- CGOP
  - Boukabara, S., I. Moradi, R. Atlas, S.P. Casey, L. Cucurull, R.N. Hoffman, K. Ide, V. Krishna Kumar, R. Li, Z. Li, M. Masutani, N. Shahroudi, J. Woollen, and Y. Zhou, 2016: Community Global Observing System Simulation Experiment (OSSE) Package (CGOP): Description and Usage. *J. Atmos. Oceanic Technol.,* **33**, 1759–1777, https://doi.org/10.1175/JTECH-D-16-0012.1
  - Boukabara, S., K. Ide, N. Shahroudi, Y. Zhou, T. Zhu, R. Li, L. Cucurull, R. Atlas, S.P. Casey, and R.N. Hoffman, 2018: Community Global Observing System Simulation Experiment (OSSE) Package (CGOP): Perfect Observations Simulation Validation. *J. Atmos. Oceanic Technol.,* **35**, 207–226, https://doi.org/10.1175/JTECH-D-17-0077.1
  - Boukabara, S., K. Ide, Y. Zhou, N. Shahroudi, R.N. Hoffman, K. Garrett, V.K. Kumar, T. Zhu, and R. Atlas, 2018: Community Global Observing System Simulation Experiment (OSSE) Package (CGOP): Assessment and Validation of the OSSE System Using an OSSE–OSE Intercomparison of Summary Assessment Metrics. *J. Atmos. Oceanic Technol.,* **35**, 2061–2078, https://doi.org/10.1175/JTECH-D-18-0061.1
- NR
  - W. Putman, A.M. da Silva, L.E. Ott and A. Darmenov, 2014: Model Configuration for the 7-km GEOS-5Nature Run,. GMAO Office Note No. 5 (Version 1.0), 18pp
  - Gelaro, R., and Coauthors, 2015: Evaluation of the 7-km GEOS-5 nature run. NASA Goddard Space Flight Center, NASA/TM-2014-104606/Vol.36, 285 pp.

References

- SAMS
  - Hoffman, R.N., S. Boukabara, V.K. Kumar, K. Garrett, S.P. Casey, and R. Atlas, 2017: An Empirical Cumulative Density Function Approach to Defining Summary NWP Forecast Assessment Metrics. *Mon. Wea. Rev.,* **145**, 1427–1435, https://doi.org/10.1175/MWR-D-16-0271.1
  - Hoffman, R.N., V.K. Kumar, S. Boukabara, K. Ide, F. Yang, and R. Atlas,2018: Progress in Forecast Skill at Three Leading Global Operational NWP Centers during 2015-2017 as seen in Summary Assessment Metrics (SAMs). *Wea. Forecasting,* early release, https://doi.org/10.1175/WAF-D-18-0117.1
- OSE
  - Boukabara, S., K. Garrett, and V.K. Kumar, 2016: Potential Gaps in the Satellite Observing System Coverage: Assessment of Impact on NOAA's Numerical Weather Prediction Overall Skills. *Mon. Wea. Rev.,* **144**, 2547–2563, https://doi.org/10.1175/MWR-D-16-0013.1