

# A Non-Parametric Definition of Summary NWP Forecast Assessment Metrics

## Application to Polar Data Gap Impact Assessment and NWP Centers Skills Inter-Comparison

Ross N Hoffman<sup>1,2</sup>, Sid-Ahmed Boukabara<sup>3</sup>, V Krishna Kumar<sup>4</sup>, Kevin Garrett<sup>4</sup>, Sean PF Casey<sup>1,2</sup>, Robert Atlas<sup>1</sup>

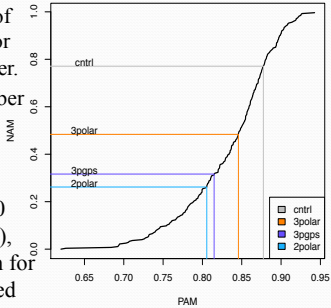
<sup>1</sup>NOAA Atlantic Oceanographic and Meteorological Laboratory, Miami FL; <sup>2</sup>Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL

<sup>3</sup>NOAA/NESDIS/Center for Satellite Applications and Research (STAR), College Park, MD

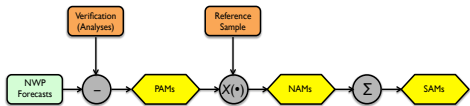
<sup>4</sup>Riverside Technology Inc., Joint Center for Satellite Data Assimilation and NOAA/ NESDIS/Center for Satellite Applications and Research (STAR), College Park, MD

### Empirical cumulative density function (ecdf) normalization

- The ecdf score for a metric is the fraction of cases in the sample, for which this case is better.
- The average of a number of ecdf's has a normal distribution.
- The ecdf for 5-day forecasts of NHX 500 hPa height AC (BGK), and the normalization for the forecasts initialized 00 UTC 18 July 2014.



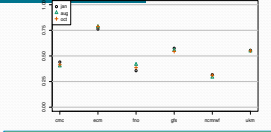
### Flow chart



- PAMs** :: primary assessment metrics
  - (ACC, RMSE) × (NHX, Tropics, SHX) × Level × Variable × Forecast length × Verification time
- NAMs** :: normalized assessment metrics
  - Normalized using the reference sample
- SAMs** :: summary assessment metrics
  - Average of all or some NAMs

### 2015 SAMs for NWP centers

- The reference samples are all initial times for all centers, month by month.
- Each center's analysis is used for verification.
- PAM dimensions** :: levels
  - variables :: geopotential height (HGT), temperature (T), and vector wind (WIND)
  - levels :: 250, 500, 700, 850, 1000 hPa
  - forecast times :: every 24 hours from 1 to 6 days
  - geographic domains :: NHX, SHX, tropics (TRO)
  - valid times :: 00 UTC from 01 until 31 of each month in 2015
  - centers: [cmc](#), [ecm](#), [fno](#), [gfs](#), [ncmrwf](#), [ukm](#)



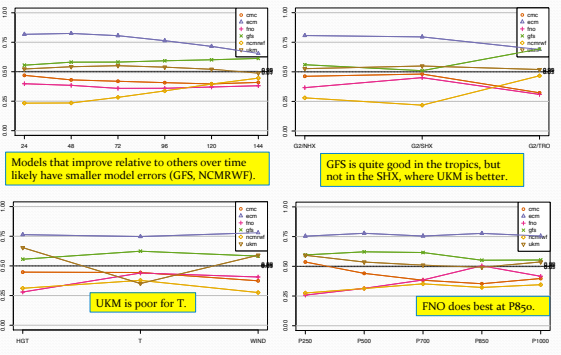
### 3-step calculation procedure

- Sample.** Define the reference sample. Example: all experiments, all initial times, for NHX AC for 2-day forecasts of 500 hPa height. Under  $H_0$ , the null hypothesis, all the members of a subset are from the same distribution.
- Normalize.** Each PAM is converted to a NAM that ranges from 0 (poor) to 1 (excellent). The normalization depends on the subset. Empirical *c.d.f.* normalization is proportional to rank in the reference sample. Under  $H_0$ , the NAMs are uniform on [0,1].
- Average.** Since the NAMs are comparable, we may average them over dimensions and values. Under  $H_0$ , the averages (SAMs) are approximately Gaussian with mean 0.5, and variance  $1/(12n)$ .

### SAMs for the BGK OSE

- Hoffman et al (2017) apply ecdf SAM to the data gap OSE experiments of Boukabara et al (2016, BGK) using the January 2015 NOAA operational system
  - cntrl: All observing systems used in operations.
  - 3polar: Retains only one satellite in each primary orbit.
  - 3pgps: Like 3polar, but with few RO obs poleward of 24°.
  - 2polar: Like 3polar but without the PM satellite.
- PAM dimensions** :: levels
  - variables :: geopotential height (HGT), temperature (T), and vector wind (WIND)
  - levels :: 250, 500, 700, 850 hPa
  - forecast times :: every 24 hours from 1 to 7 days
  - geographic domains :: NHX, SHX, tropics (TRO)
  - initial times :: 00 UTC from 25 May until 31 July 2014
  - experiments :: 2polar, 3pgps, 3polar, cntrl

### NWP centers Jan 2015 RMSE SAMs



Models that improve relative to others over time likely have smaller model errors (GFS, NCMRWF).

GFS is quite good in the tropics, but not in the SHX, where UKM is better.

UKM is poor for T.

FNO does best at P850.

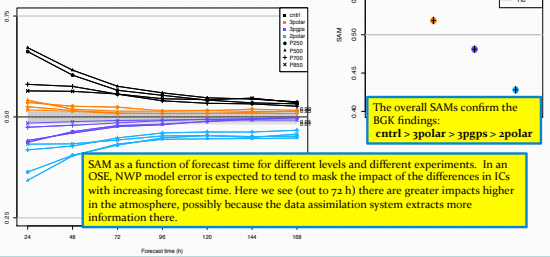
### References

- Boukabara, S.-A., 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 (7), 2547–2563, doi:10.1175/MWR-D-16-0013.1.
- R. N. Hoffman, S.-A. Boukabara, V. K. Kumar, K. Garrett, S. P. F. Casey, and R. Atlas. An empirical cumulative density function approach to defining summary NWP forecast assessment metrics. *Mon. Weather Rev.*, 15 July 2016. In press. doi:10.1175/MWR-D-16-0271.1.
- Hölm, E., R. Forbes, S. Lang, L. Magnusson, and S. Malardel, 2016: New model cycle brings higher resolution. *ECMWF Newsletter*, (147), 14–19.

**Acknowledgments.**

- Fanglin Yang explained the intricacies of the VSDB file structure.
- ECMWF gave permission to use the scorecard image.
- Financial support for this work is gratefully acknowledged, including funding provided by the Disaster Relief Appropriations Act of 2013 (H.R. 152).

### OSE SAM results



### Summary and Conclusions

- SAMs are defined as the average of a collection of NAMs.
  - The main advantages of the ecdf approach are that it is amenable to statistical significance testing.
- The ecdf normalization is tested for two cases.
  - The ecdf SAMs are relatively easy to interpret since the metrics for various subsets vary relatively consistently.
  - The OSE results are consistent with BGK's conclusions.
  - The NWP centers results agree with our prior assessment of relative forecast skill and show some interesting details.