

Probabilistic Solar Power Forecasts Using a Large Ensemble

Stephen Jascourt, Senior Scientist
and Christopher Cassidy, Eric Wertz, and Travis Hartman

Outline

1. Motivation for probabilistic forecasts and scenario
2. Forecast paradigm shift
3. Test evaluation set-up and scoring
4. Test evaluation results – beware of scoring system details!
5. Summary

Solar **U**ncertainty **M**anagement and **M**itigation for **E**xceptional **R**eliability in **G**rid **O**perations

Goal: Use of **probabilistic** solar power forecasts in **operational** decision systems governing electric power system in ERCOT

Accurate + reliable + sharp + resolved + discriminating probabilistic forecast

$P(\text{fcst}) = \text{freq obs}$

$P(\text{fcst}) \neq \text{clim}$

Diff fcst \rightarrow diff outcome

Diff outcome \leftarrow diff fcst

\rightarrow Risk parity economic dispatch

\rightarrow reduce operating costs

\rightarrow increase reliability

\rightarrow Dynamic adaptive reserves

\rightarrow reduce cost and reduce pollution

\rightarrow Open source solar power forecasting visualization tool and situational awareness tools

3-year project funded by DOE Solar Energy Technologies Office

Project team



Utility-scale solar farms in ERCOT

21 utility-scale PV generating units

> 1.5 GW ac capacity operating

Good history data > 1 year for 15 units

>3 years history data for some

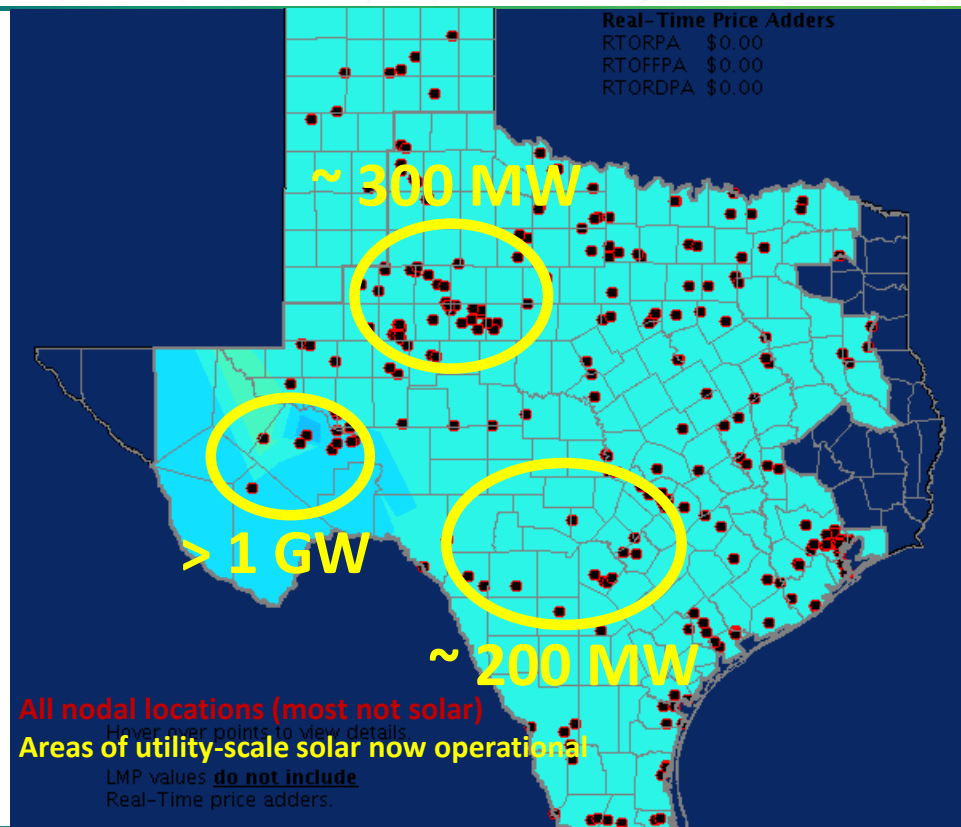
Concentration in west TX

> 1.5 GW more have full interconnection study approved

~ 30 GW in queue

20+ GW wind and 30 GW in queue

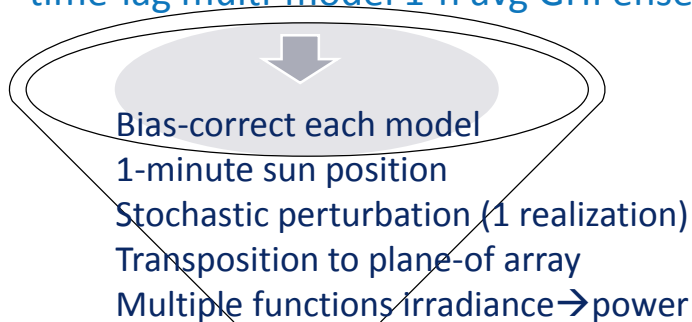
Summer record load ~ 70 GW



Paradigm Shift to Probabilistic Forecasts

CURRENT PARADIGM

5-minute real-time telemetry
time-lag multi-model 1-h avg GHI ensemble



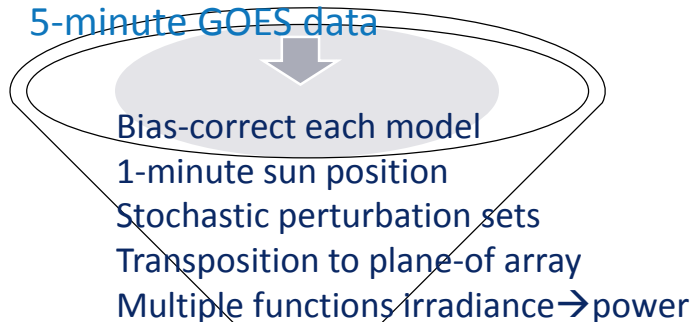
Ensemble set + short term corrections

Blend →
Deterministic 1h forecast

Unit Scheduling/
Dispatch/Reserves

NEW PARADIGM

1-minute real-time telemetry
enormous multi-model GHI ensemble
5-minute GOES data



Ensemble set **with tails well(?) sampled**

Machine learning + Statistical Methods →
Probabilistic forecast (5 min to 2 h, then 1h)

Decision Support Tools, Risk Parity Dispatch (NREL)



Paradigm Shift to Probabilistic Forecasts

Forecasts to be available:

- Statistical short-term projection/corrections using real-time solar farm obs
- Extrapolation from GOES 5-minute imagery
- HRRR – 3 km, updates every 1h to 18h, updates to 36 h, 15 min avg, time lag ensemble
- NAM – 3 km, 6 h update, 1 h time avg, small time-lag ensemble
- GFS and (coarser space/time) 20-member GFS ensemble, 6h updates
- ECMWF and 51-member ensemble now updates every 6h, 0.2 deg, 1 h avg
- Canadian global model and 20 ensemble members every 12 h
- NOAA Short-Range Ensemble 26 members every 6 h
- Experimental HRRR 9-member ensemble, updated every 12 h (subject to change)
- New WRF-Solar forecasts from the other funded projects

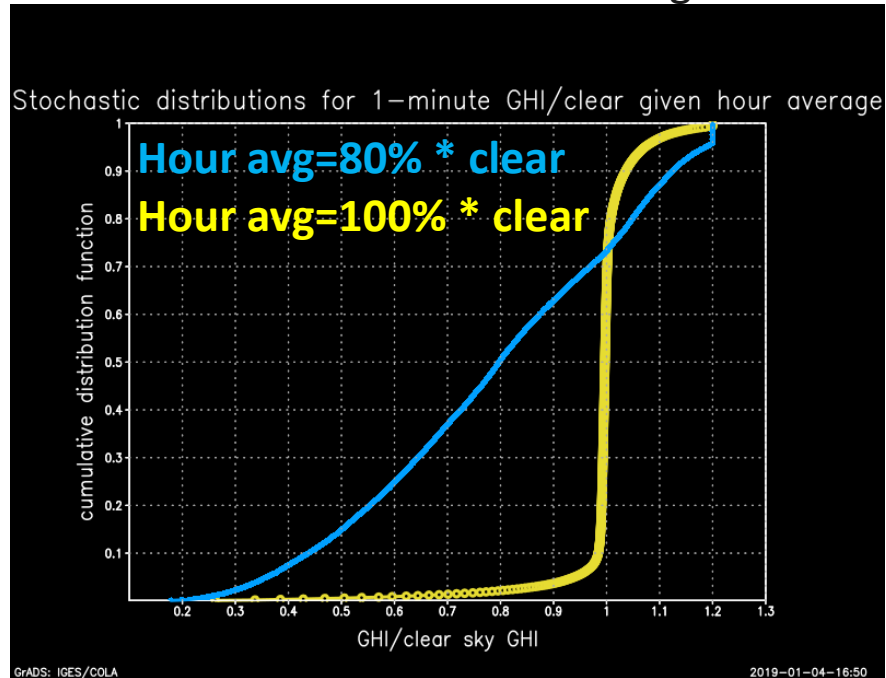
Total daily NWP model runs: >500 but only 37 high-resolution and 24 low-latency



Paradigm Shift to Probabilistic Forecasts

Forecasts to be available:

And 1000 stochastic variations for each time-average forecast!



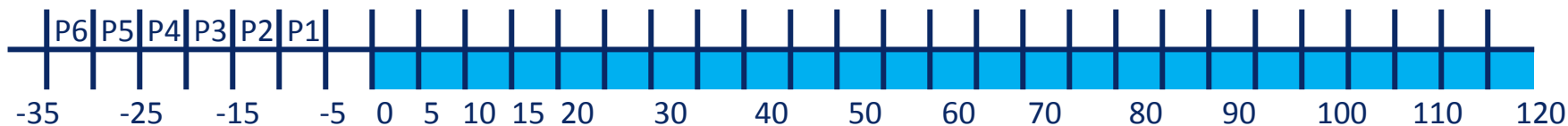
Test evaluation

- Can we evaluate ensemble skill vs. single forecast skill?
- Can we use existing set of historical forecasts to see how different sets of forecast members contribute to ensemble skill?
- Do stochastic perturbations contribute to improved ensemble skill?
- How good are 5-minute forecasts, is anything better than smart persistence?

Experiment:

Make forecasts for next 24 5-minute periods using available forecasts and smart persistence from last 6 5-minute averages.

Assume 5-min gap to collect data and make and send forecast.



Test evaluation

36 time-lag ensemble sets to test:

# smart persistence	1	2	3	4	5	6																													6	6			
# last HRRR							3	6	9	12	3		3	3	3	1		1	1	1																		12	
# last ECMWF											3	3		3	3	1	1		1	1																		3	
# last NAM											3	3	3		3	1	1	1		1																		3	
# last GFS											3	3	3	3		1	1	1	1																			3	
# last corrected HRRR																																							12
# last corrected ECM																																							3
# last corrected GFS																																							3
# last corrected NAM																																							3
# power curves/model							2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
Total # members	1	2	3	4	5	6	6	12	18	24	24	18	18	18	18	8	6	6	6	6	6	6	12	18	24	24	18	18	18	18	8	6	6	6	6	48	48		

No model only HRRR last 3 all last 1 all same but corrected model ALL

Also try using stochastic distributions for each model forecast

QC'd obs data for 1 year x 18 generating units x rolling obs, forecast every 5 minutes



Test evaluation

Evaluation metric:

Ranked probability score for 1 forecast

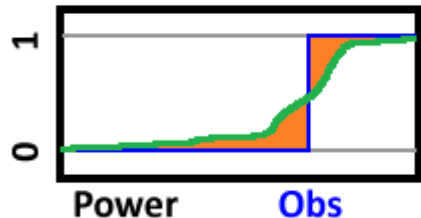
$$\sum_{\text{categories}} |prob(\text{power} > K) - obs(\text{power} > K)|^L$$

categories

L=1 for absolute error, 2=squared error
99 categories: $K=.01, .02, \dots, .99$ * capacity
prob=ensemble probability *obs*=0 or 1

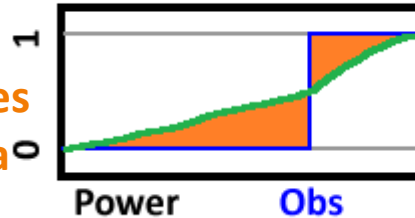
Measures difference between forecast cumulative distribution and obs (step function)

Ranked probability score for N forecasts = average score for all the forecasts

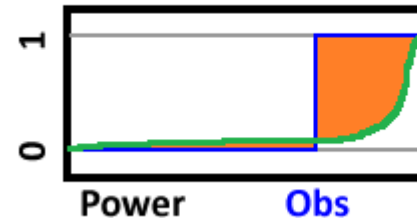


Forecast CDF good

L=1
measures
this area



too wide



sharp but too sunny

Test evaluation

Evaluation metric:

Ranked probability score for 1 forecast

$$\sum_{\text{categories}} |prob(\text{power} > K) - obs(\text{power} > K)|^L$$

L=1 for absolute error, 2=squared error
99 categories: K=.01,.02,...,.99 * capacity
prob=ensemble probability obs=0 or 1

Measures difference between forecast cumulative distribution and obs (step function)

Ranked probability score for N forecasts = average score for all the forecasts

$$\text{Ranked probability SKILL SCORE} = \frac{\text{forecast score} - \text{reference score}}{\text{Perfect} - \text{reference score}} = 1 - \frac{\text{Forecast score}}{\text{reference score}}$$

1=perfect 0=same skill as reference

Reference forecast = climatology median vs. quantile distribution (n=# ensembles)

Test evaluation

Ranked Probability Skill Score:

$L=2$ (square error) - low bias for small ensemble due to sampling of reference climatology

Simple fix *if categories equally likely* (not true here)

Instead, try using climatology quantiles

$L=1$ (absolute error) – not “proper” scoring method but no bias vs. # of members

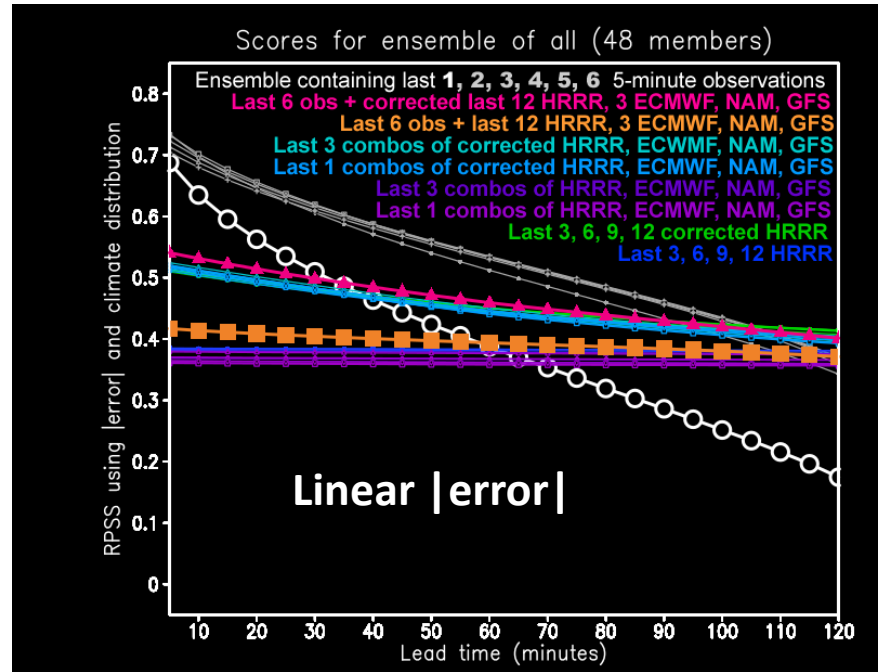
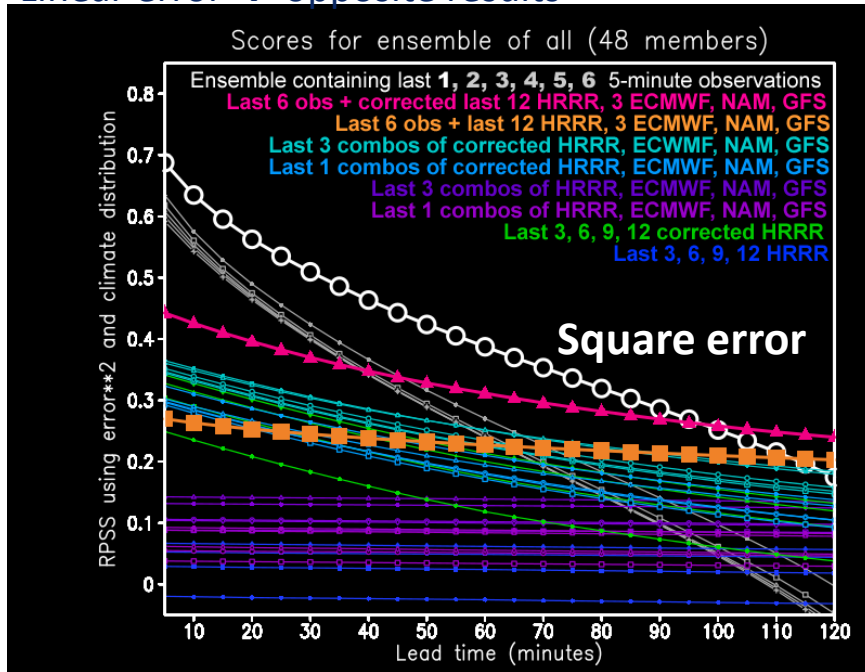
Score the ensemble sets using:

- $L=1$ vs. $L=2$ (both climo=distribution)
- $L=1$ climo=median vs. climo=distribution
- Each ensemble member as single forecast vs. stochastic perturbations
- Deterministic scores for ensemble mean: MAE and RMSE skill scores vs. climo mean



Test evaluation: square error vs. linear |error|

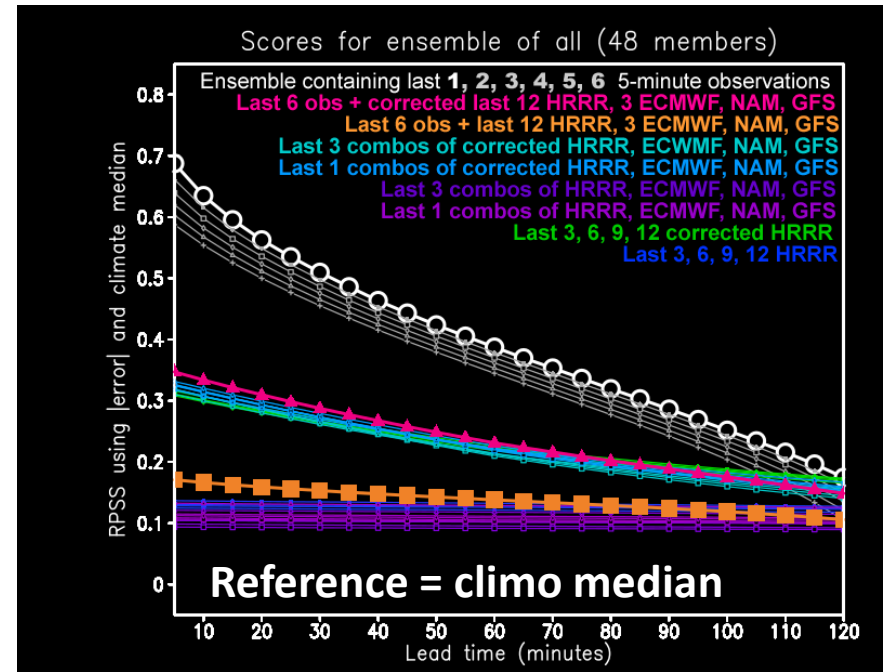
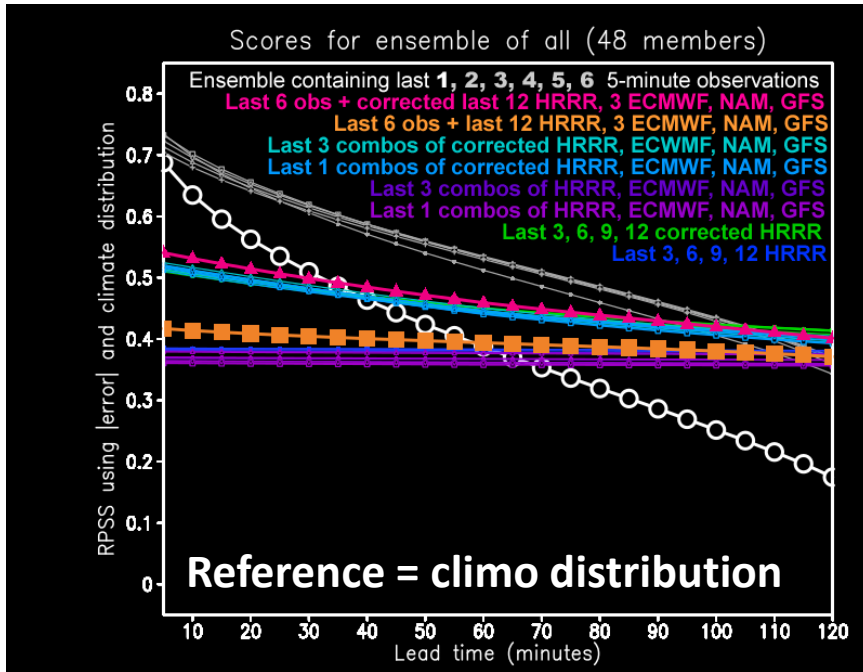
Square error → more spread among forecasts, multiple obs (smart persistence) worse than single obs, HRRR alone worse than including other models, “everything” forecast much better than model only.
Linear error → opposite results



Test evaluation: linear error climo median vs. climo distribution

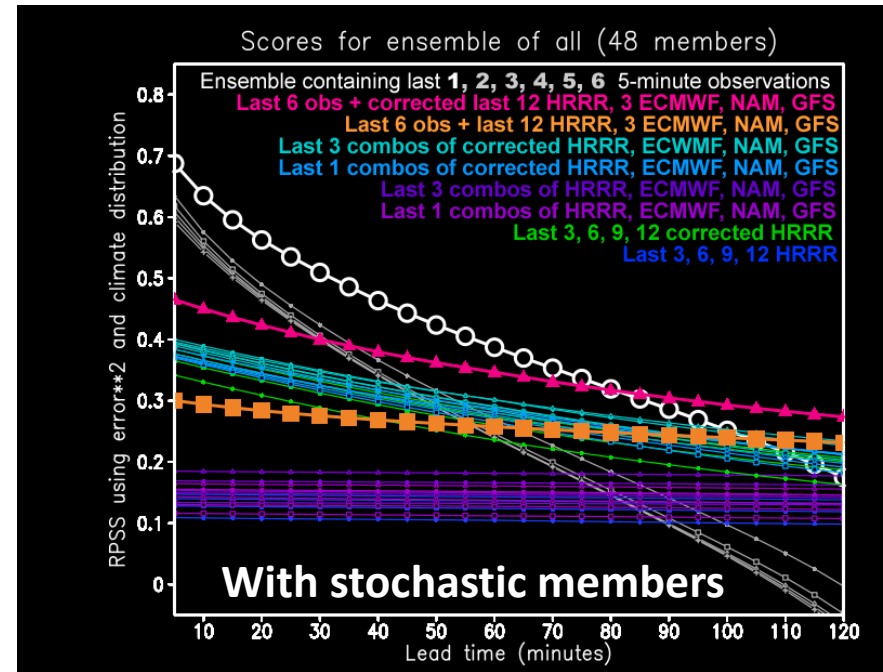
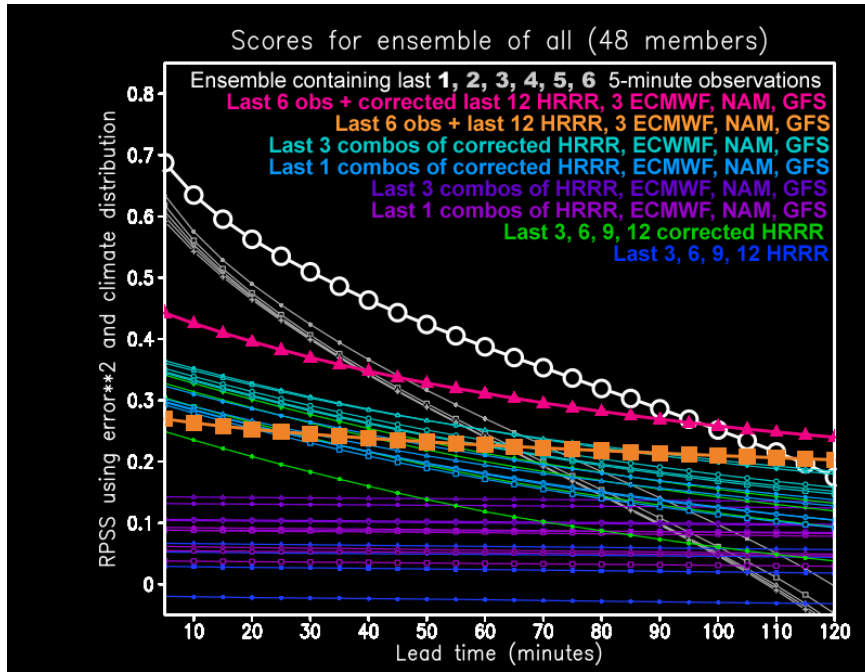
Climate median → Single obs better than multiple obs (for smart persistence)

Climate distribution → model-based scores higher, implying less skill for climate distribution than median



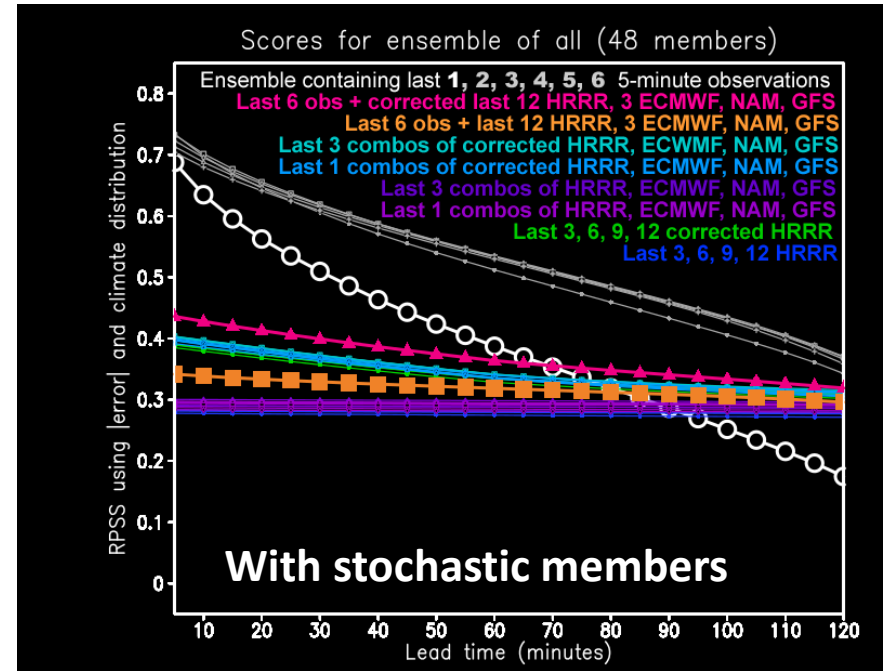
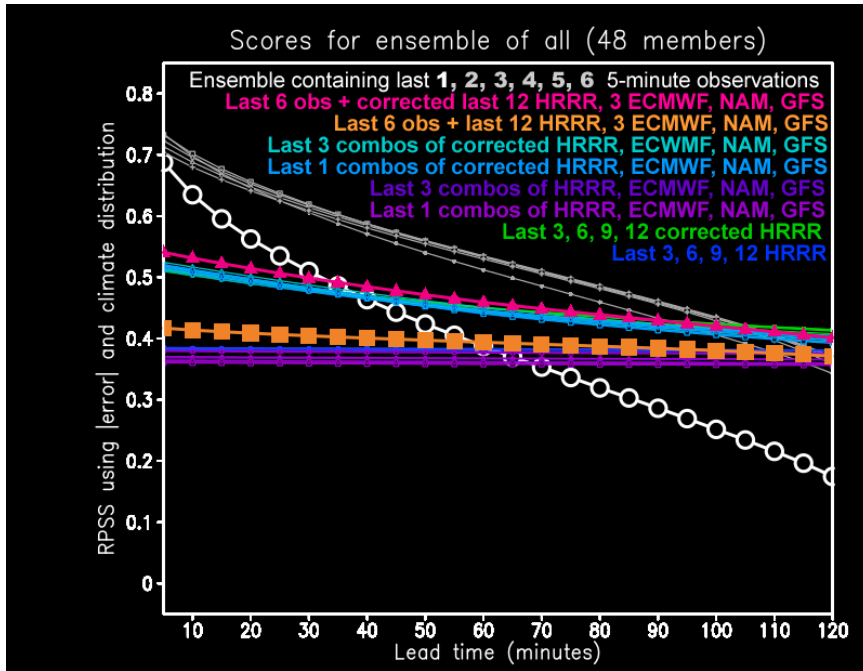
Test evaluation: Stochastic vs. single value per member

Square error comparison: using 1000 stochastic members per original member reduces spread between model ensemble sets and increases scores



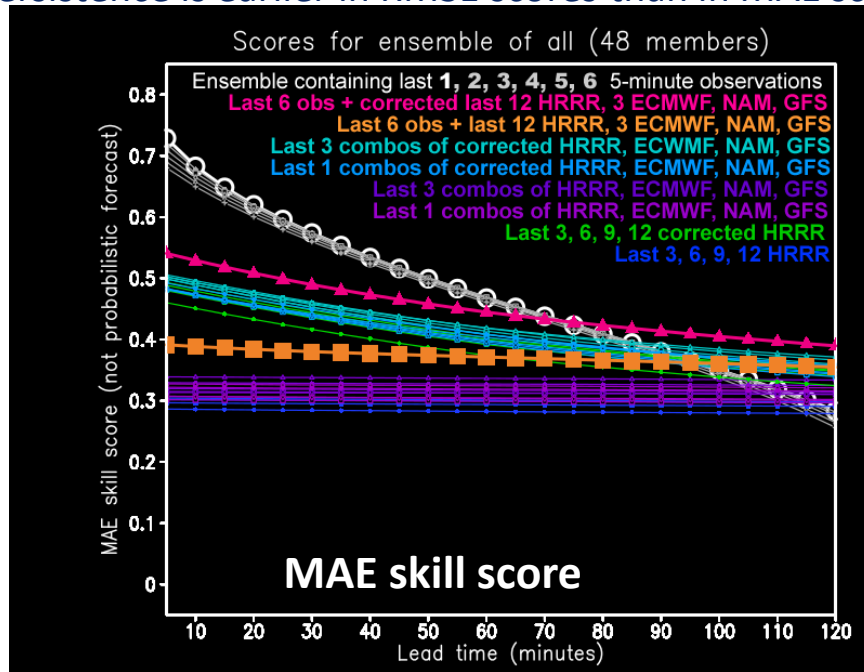
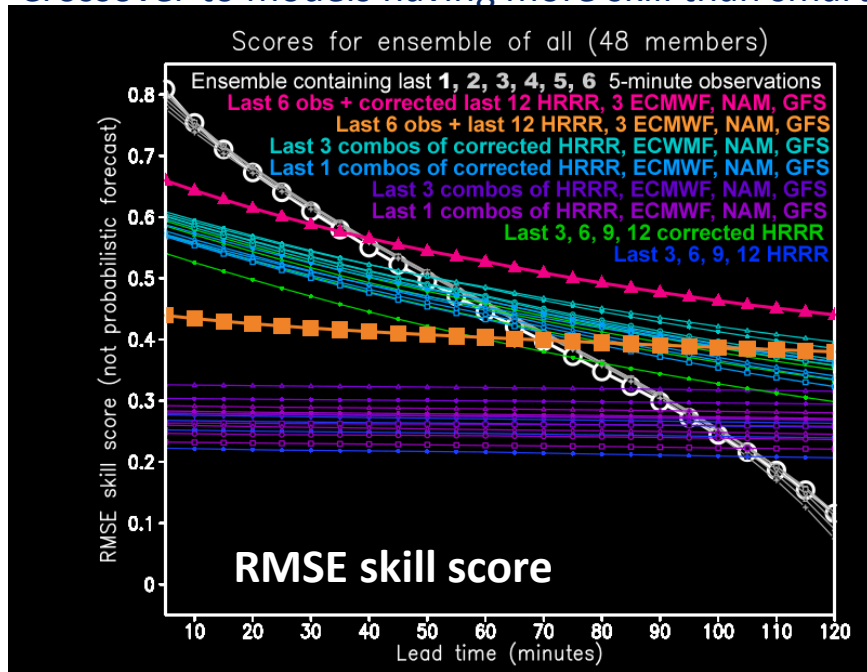
Test evaluation: Stochastic vs. single value per member

Linear |error| comparison: using 1000 stochastic members per original member reduces all scores of model-based forecasts and flips skill of HRRR-only from better than other model combinations to worse



Test evaluation: deterministic results

Skill score for ensemble mean vs. climo mean shows less skill distinction in either direction for different ensemble sizes of smart persistence. RMSE scores show more spread between ensemble sets. Crossover to models having more skill than smart persistence is earlier in RMSE scores than in MAE scores.



Summary

- SUMMER-GO is going! Benefits of probabilistic forecasts appear promising!
- More high-resolution ensembles with small latency are needed
- Evaluating forecast improvements in ensemble sets is difficult using a single measure even as comprehensive and powerful as the ranked probability skill score
 - Evaluation is sensitive to details of the reference or climatology forecast
 - Evaluation is sensitive to |linear| vs. square error in probability space
 - Whether a careful stochastic enhancement of the forecast distribution appears to improve skill depends on details of the scoring system
 - The cross-over time when model-based forecasts are superior to smart persistence depends on the scoring system details
 - Whether the last observation or a collection of recent observations produces a higher skill score depends on the scoring system details

Thank You

Stephen Jascourt, Senior Scientist

Stephen.Jascourt@radiantsolutions.com

(240)-833-8233