A recommended reforecast configuration for the NCEP GEFS

Tom Hamill

NOAA/ESRL Physical Sciences Division

tom.hamill@noaa.gov (303) 497-3060

also: Trevor Alcott, Mark Antolik, James Brown, Mike Charles, Dan Collins, Mark Fresch, Kathryn Gilbert, Hong Guan, Hank Herr, Wallace Hogsett, David Novak, Melissa Ou, David Rudack, Philip Schafer, Michael Scheuerer, Geoff Wagner, John Wagner, Tom Workoff, Bruce Veenhuis, and Yuejian Zhu

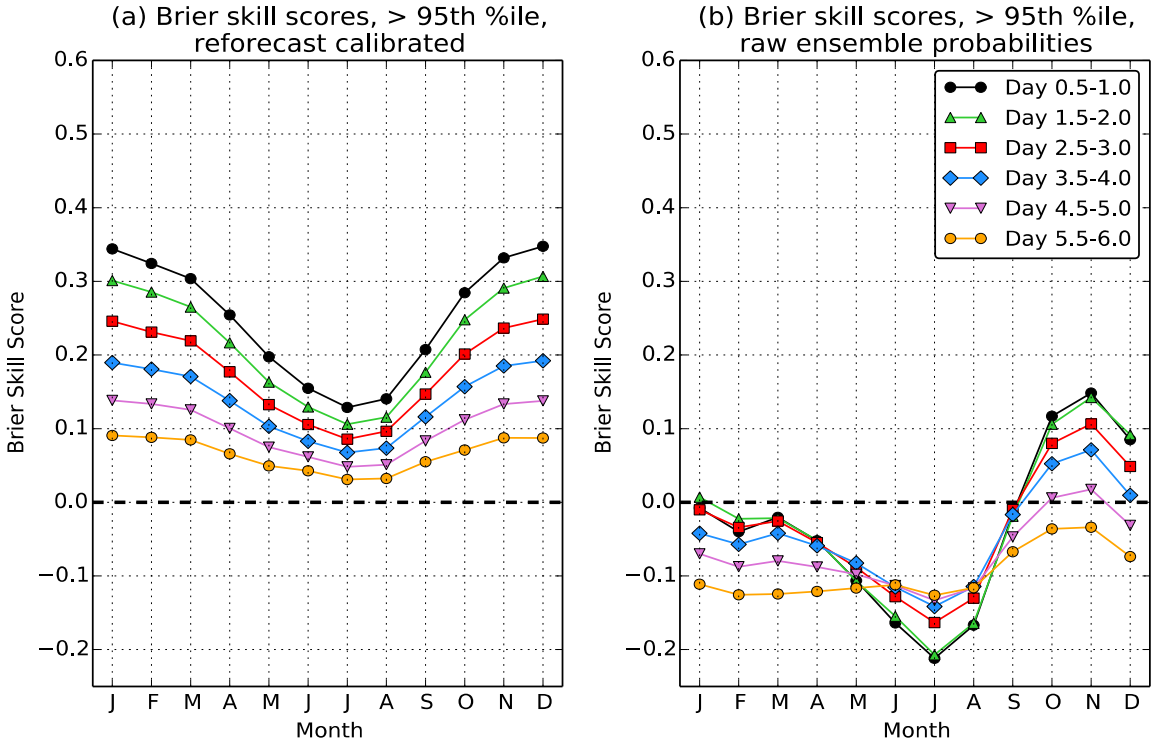
(to be presented to NCEP management)

Outline

- Introduction
  - Motivation for reforecasts
  - Users
  - Issues
- Summary of sample-size sensitivity experiments
- Recommendations
- Remaining issues and discussion.

Motivation for reforecasts

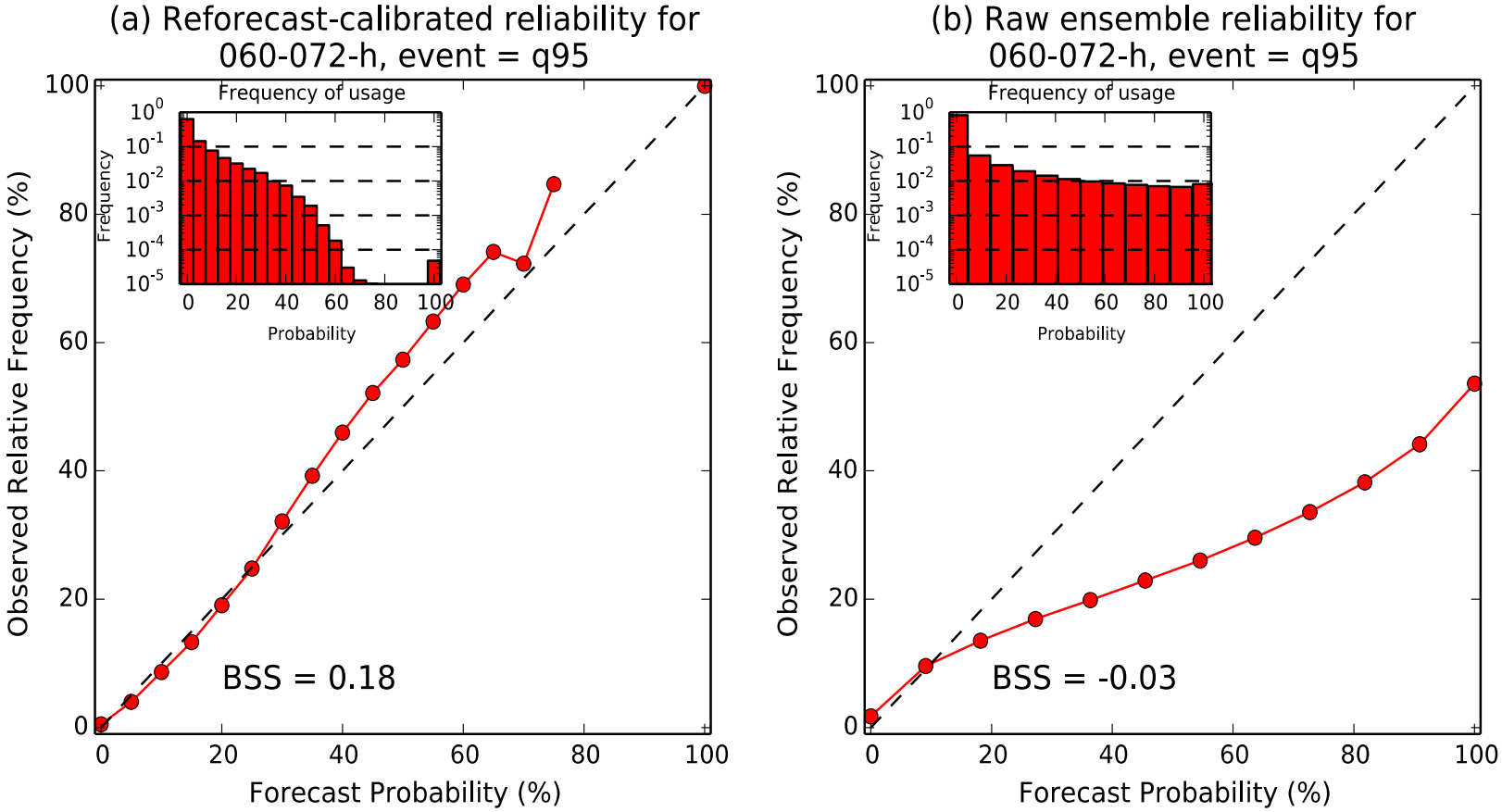
- Extra training sample size from large reforecast database (and associated observations analyses) can greatly improve skill and reliability of statistical post-processed guidance, thus improving decision support.



Raw ensemble BSS depressed perhaps ~5% because of use of 11-member reforecast ensemble rather than 21-member real-time. See Richardson, 2001 QJRMS for more on this.

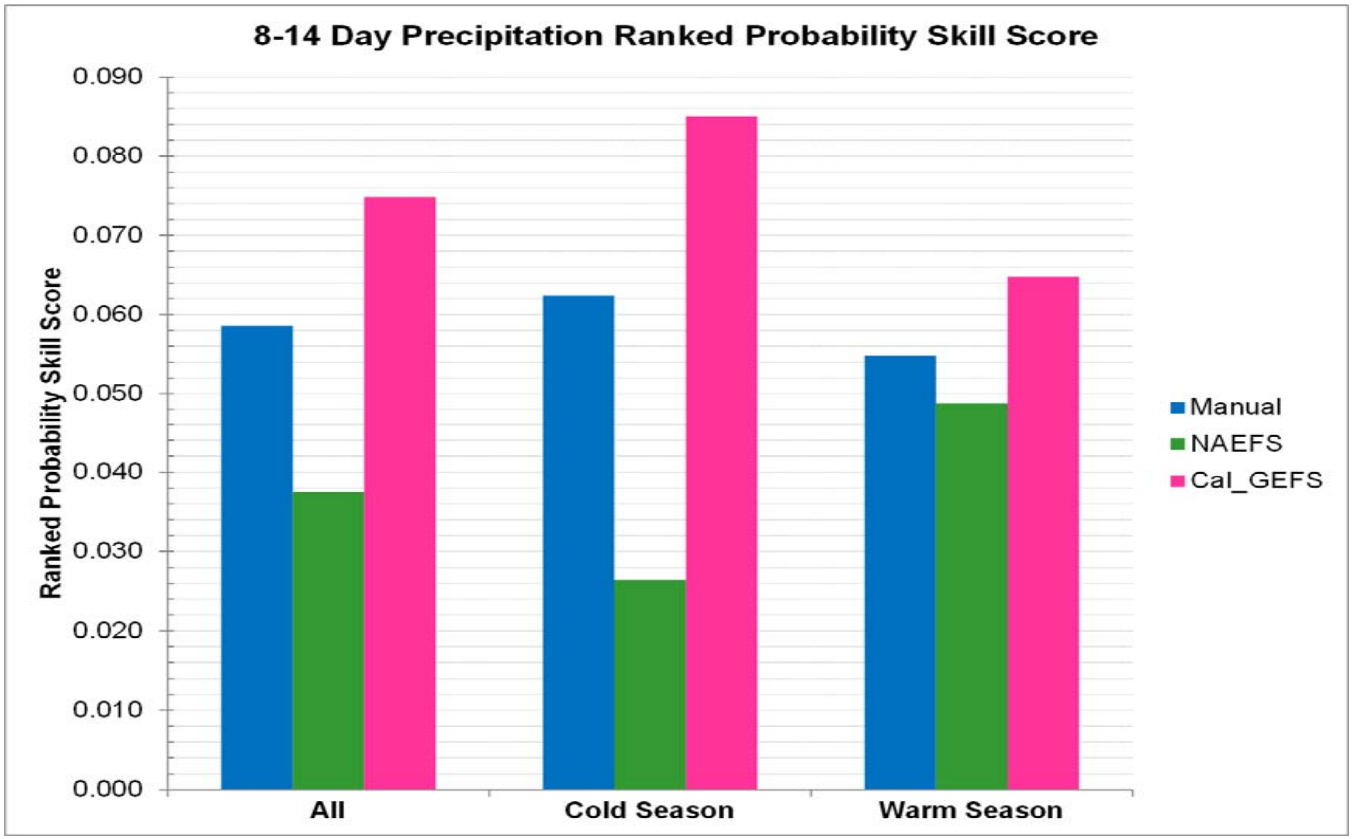
Still, note dramatic effect of post-processing, here using reforecasts, rank analog approach (Hamill and Whitaker 2006) and 2002-2013 1/8-degree CCPA over CONUS.

Reliability before and after post-processing



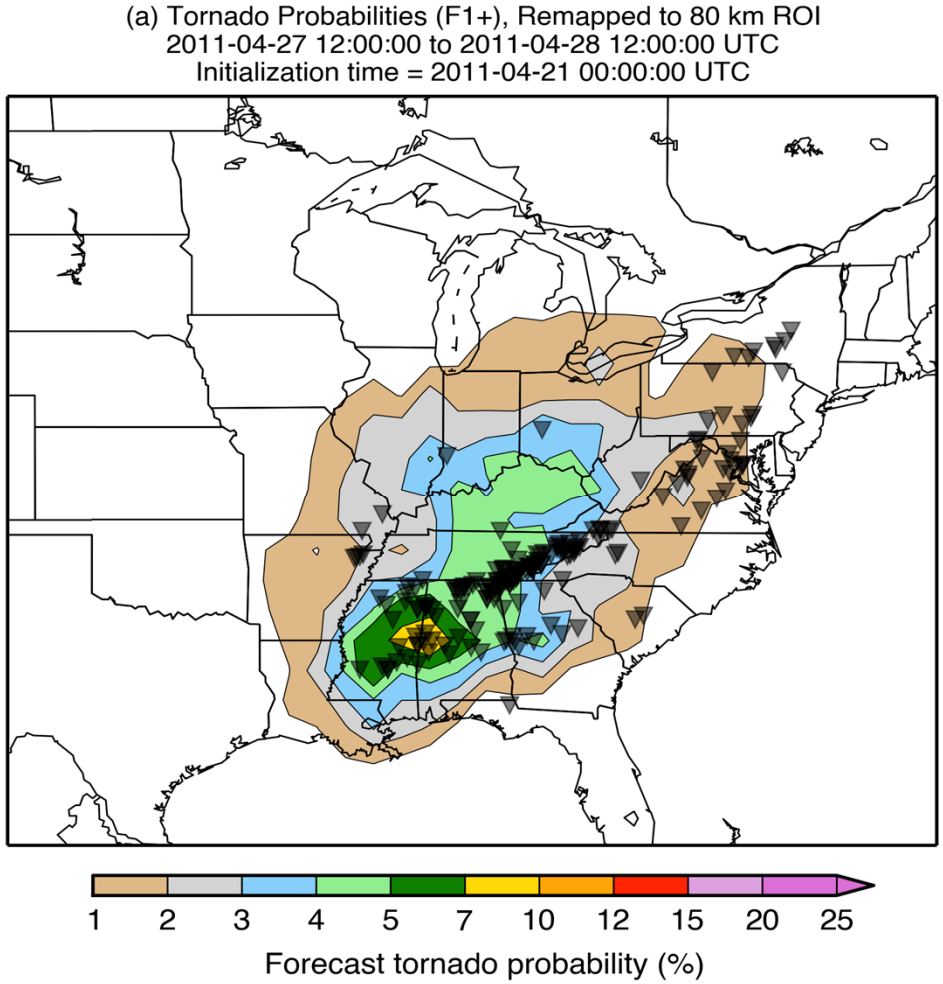
different post-processing methods than this one may have different reliability characteristics than this analog technique, which tends to under-estimate high-end probabilities.

CPC reforecast skill vs. NAEFS, manual



Skill of NCEP/CPC's forecasts of above/near/below-normal precipitation for the 8-14 day period from various methods, including manual forecasts, from NAEFS, and from the reforecast-calibrated GEFS.

Lengthy reforecasts helpful for objective calibration of rare events (here, tornado forecasts)



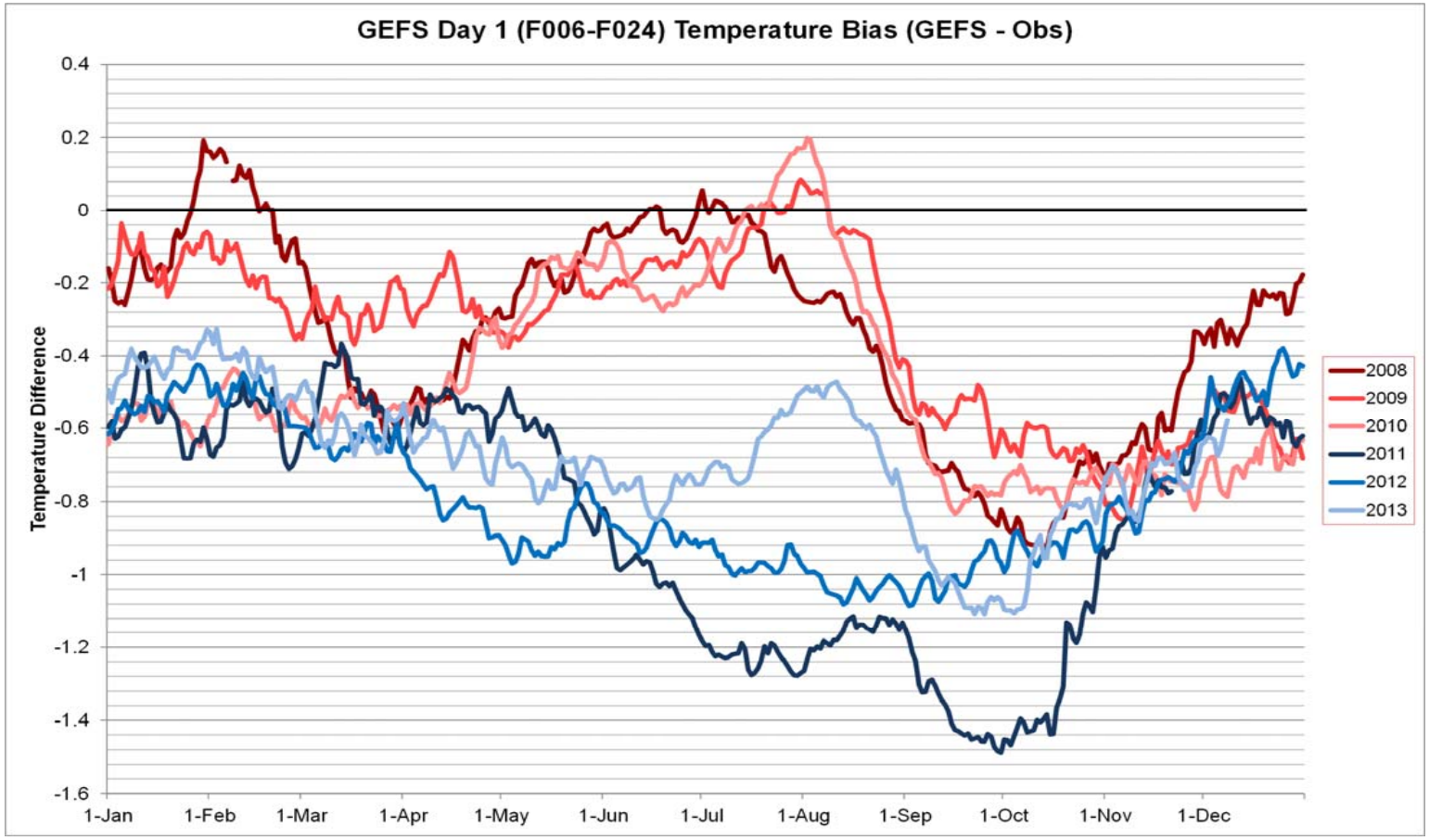
Objective tornado probabilities for the period 12 UTC 27 April 2011 to 12 UTC 28 April 2011 for forecasts initialized 6 days earlier, based on a reforecast analog procedure (Tuscaloosa outbreak). Observed tornado locations are shown with the grey inverted triangles.

c/o Francisco Alvarez, Ph.D. candidate, St. Louis University.

Issues in the computation of reforecasts

- Computational expense.
- Archival expense.
- Need coincident, high-quality verification or analysis data to get the most from reforecasts.
  - precipitation analyses.
  - surface analyses.
- Necessity of a *consistent* system for initialization and forecast.

Changing short-term forecast bias due to changes in data assimilation system



c/o CPC. In 2011, the reforecasts changed from CFSR initialization to GSI initialization, which used a slightly different version of the forecast model.

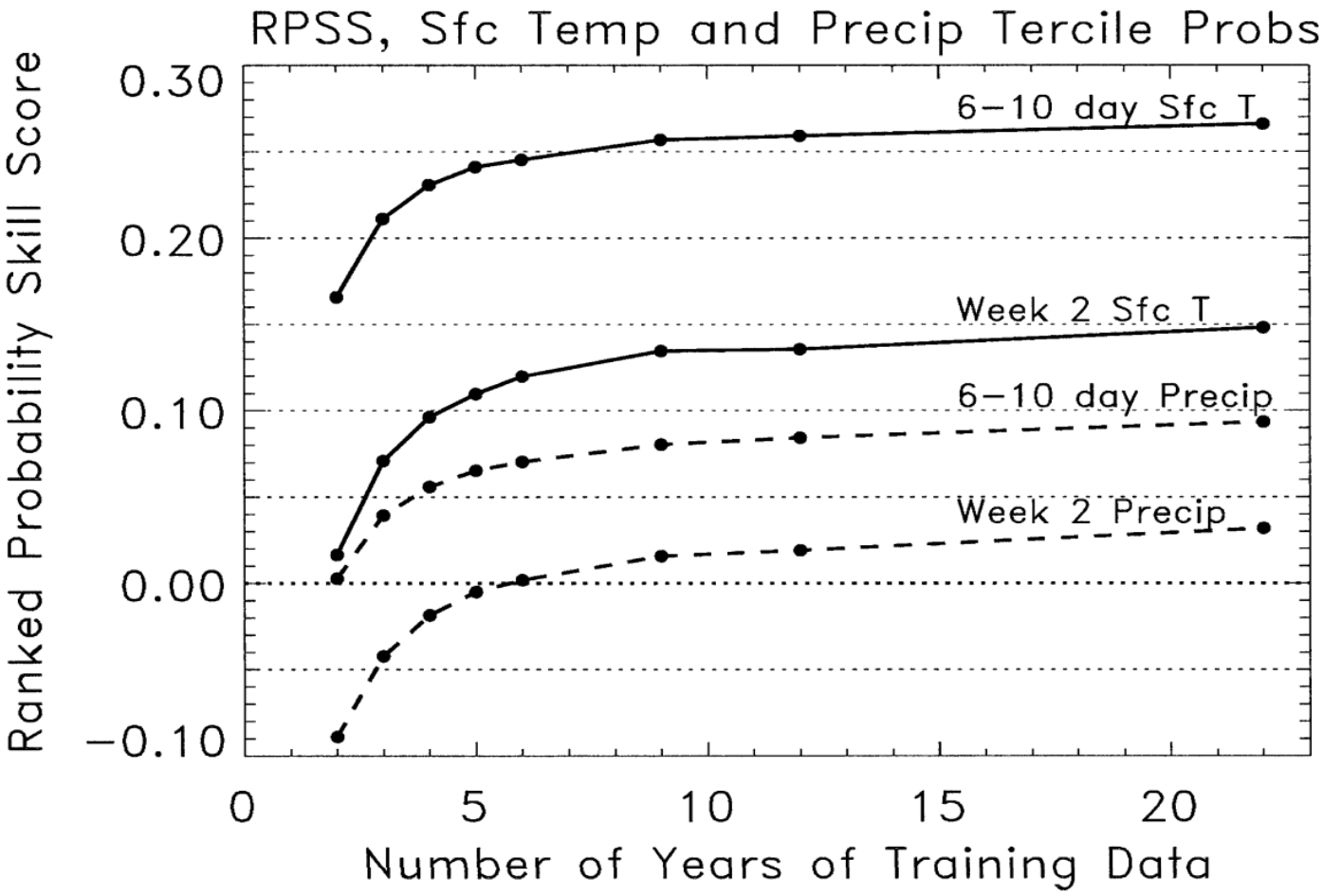
NWS users of reforecast-based products

- **CPC**: 6-10 day, 8-14 day tercile forecasts of temp and precipitation over US, week +2 temperature extremes. More products in pipeline, including global products, extended hazards outlook.
- **OHD**: for statistically post-processing of temperature and precipitation to provide input to hydrologic models. Also provides sufficient sample size for testing hydrologic models over many cases.

NWS users of reforecast-based products, continued

- **WPC**. For calibrated PQPF.
- **SPC**. Will be evaluating Alvarez's tornado forecast guidance, may use this in future extended-range tornado guidance.
- **NWS Western Region**. Situational awareness tool (http://ssd.wrh.noaa.gov/satable/)
- **NDFD**: Sandy Supplemental project to provide NDFD grids through multi-model blend. GEFS reforecasts are keystone input.

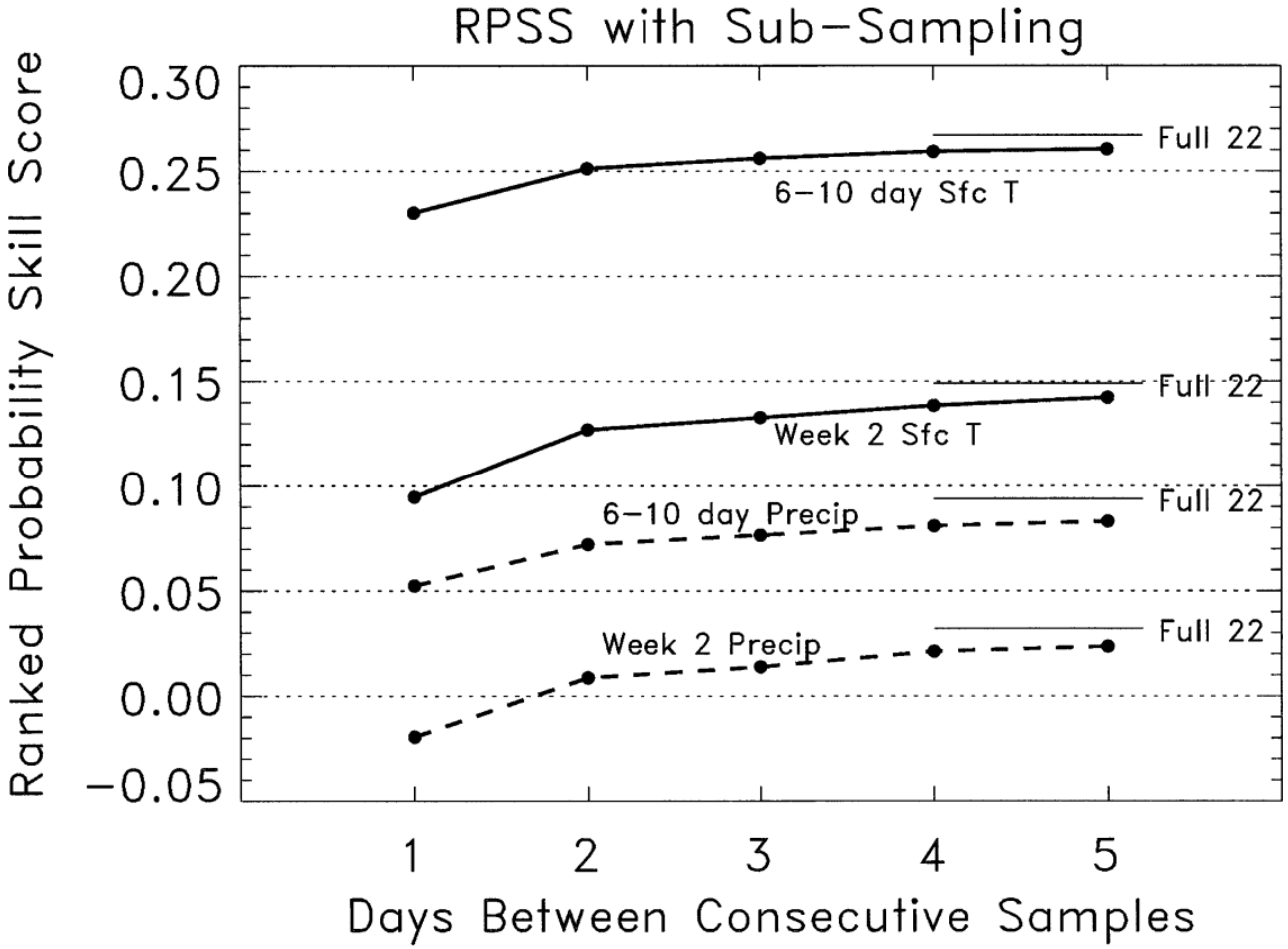
Some sample-size sensitivity results: 6-10 and 8-14 day forecasts.



Ranked probability skill score (larger is better) of post-processed surface temperature and precipitation forecasts as a function of the number of years of training data, assuming every-day samples were available.

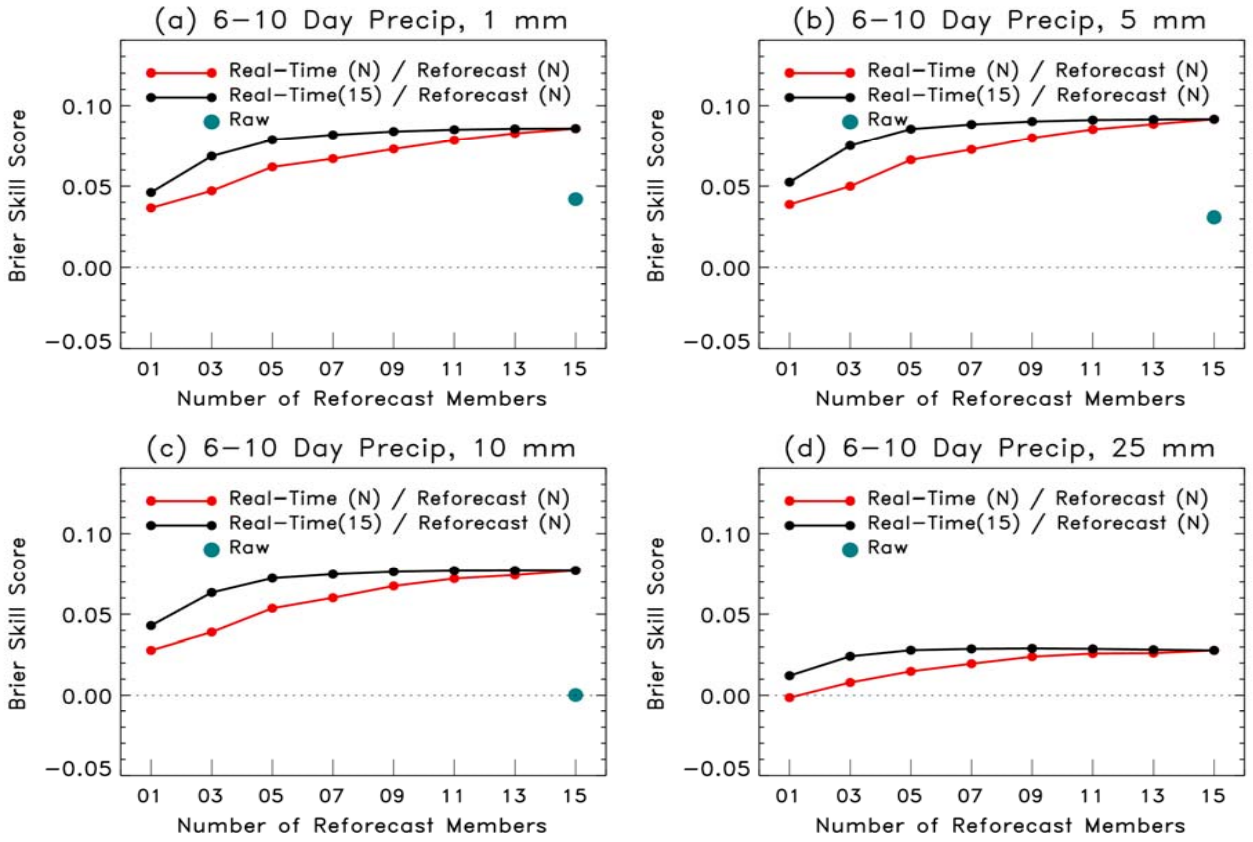
(used 1st-generation GEFS reforecasts)

Some sample-size sensitivity results: 6-10 and 8-14 day forecasts.



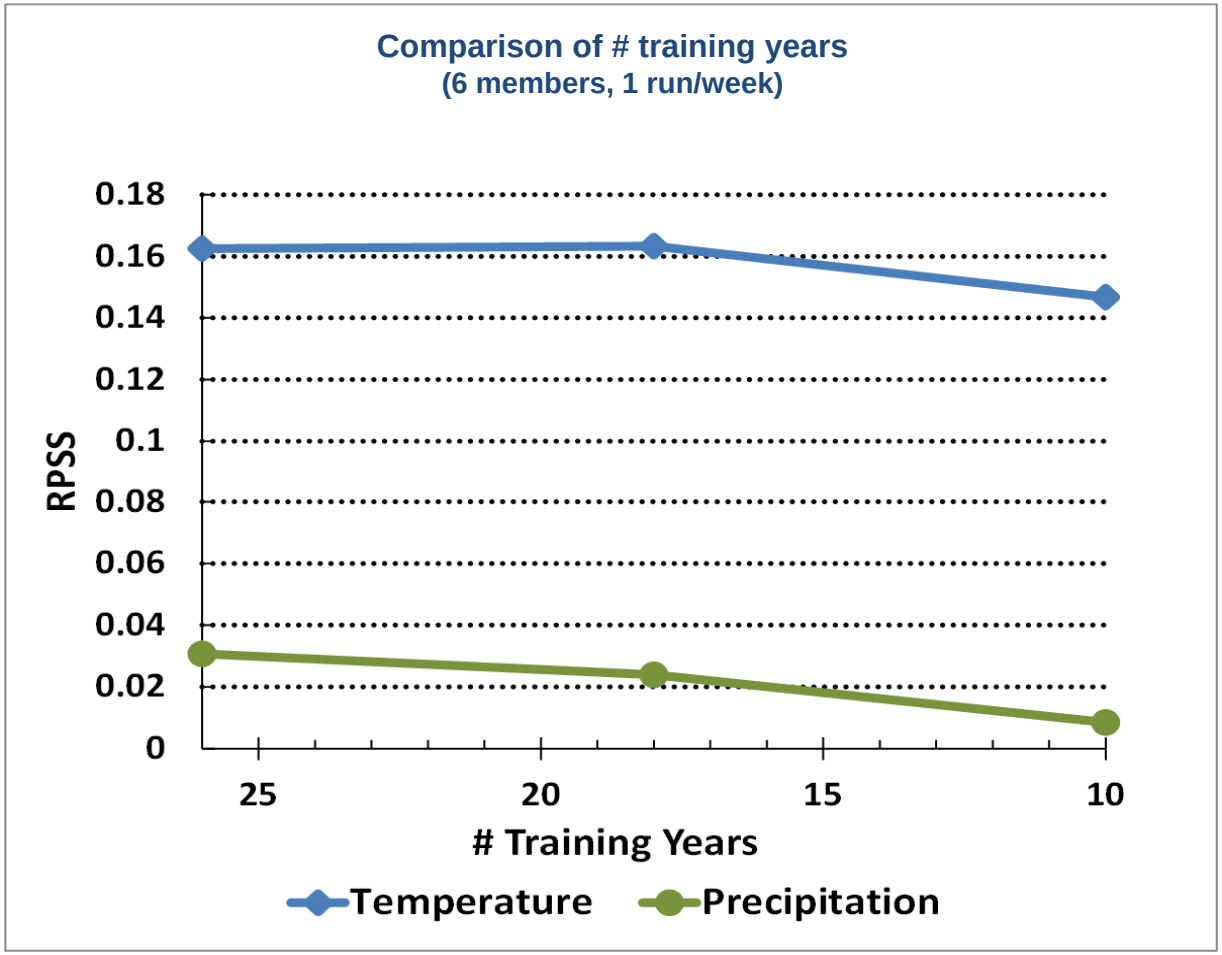
Ranked probability skill score (larger is better) of post-processed surface temperature and precipitation forecasts as a function of the number of days skipped between forecast samples. In all experiments, four total years of training data were used.

Sensitivity to number of members



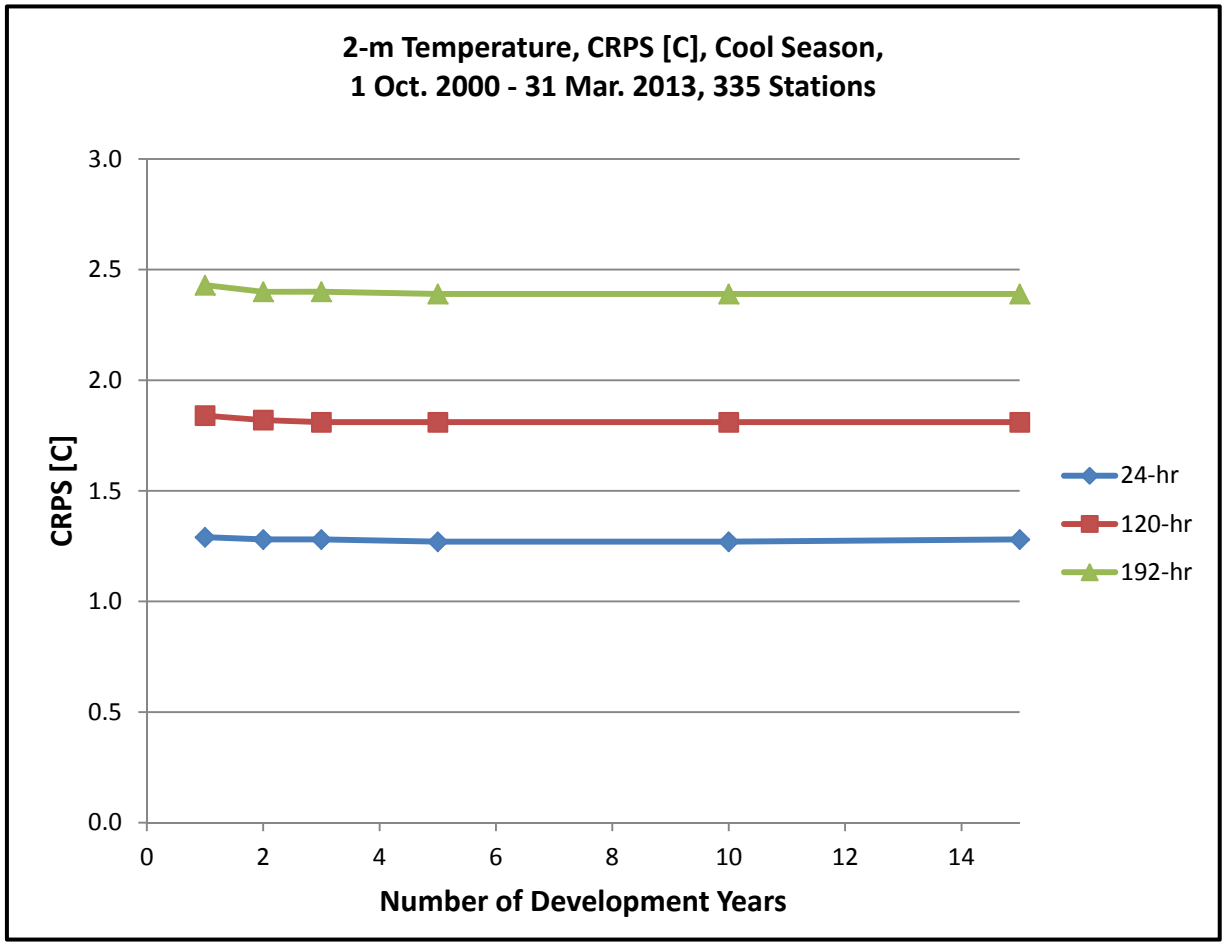
6-10 day precipitation forecast Brier Skill Score (larger is better) as a function of the number of reforecast members and the number of real-time members. The green dot shows the skill of the raw ensemble forecast guidance, here with 15 members used to set the probabilities. The red curves show the skill when n members are used both as the ensemble size for the reforecast ensemble and for the real-time ensemble. The black curve shows the skill when a n-member ensemble is used for the reforecast and a 15-member ensemble is used for the real-time forecast.

Sample-size sensitivity: CPC 8-14 day temperature & precipitation



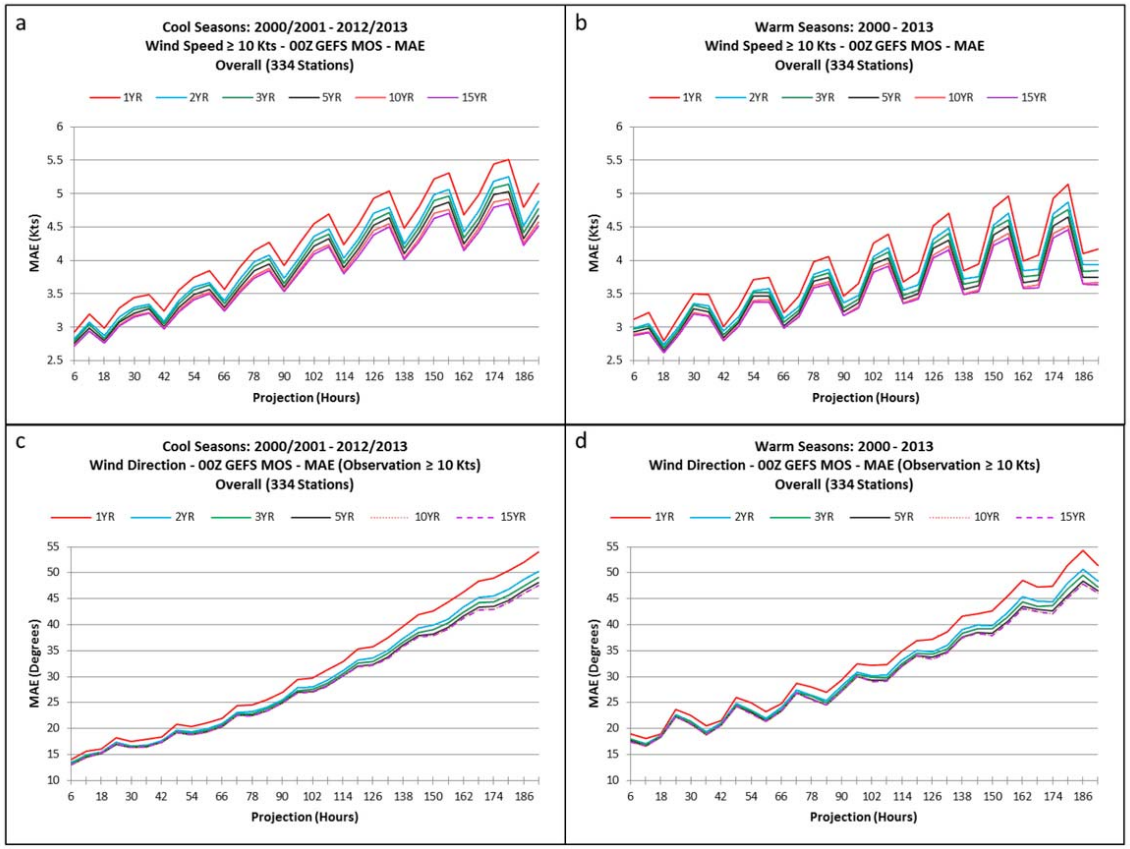
Ranked probability skill score (larger is better) for CPC 8-14 day surface temperature and precipitation skill as a function of the number of years of training data, using GEFS second-generation reforecasts and station observations over the US.

Sample-size sensitivity
MDL: 2-m temperature



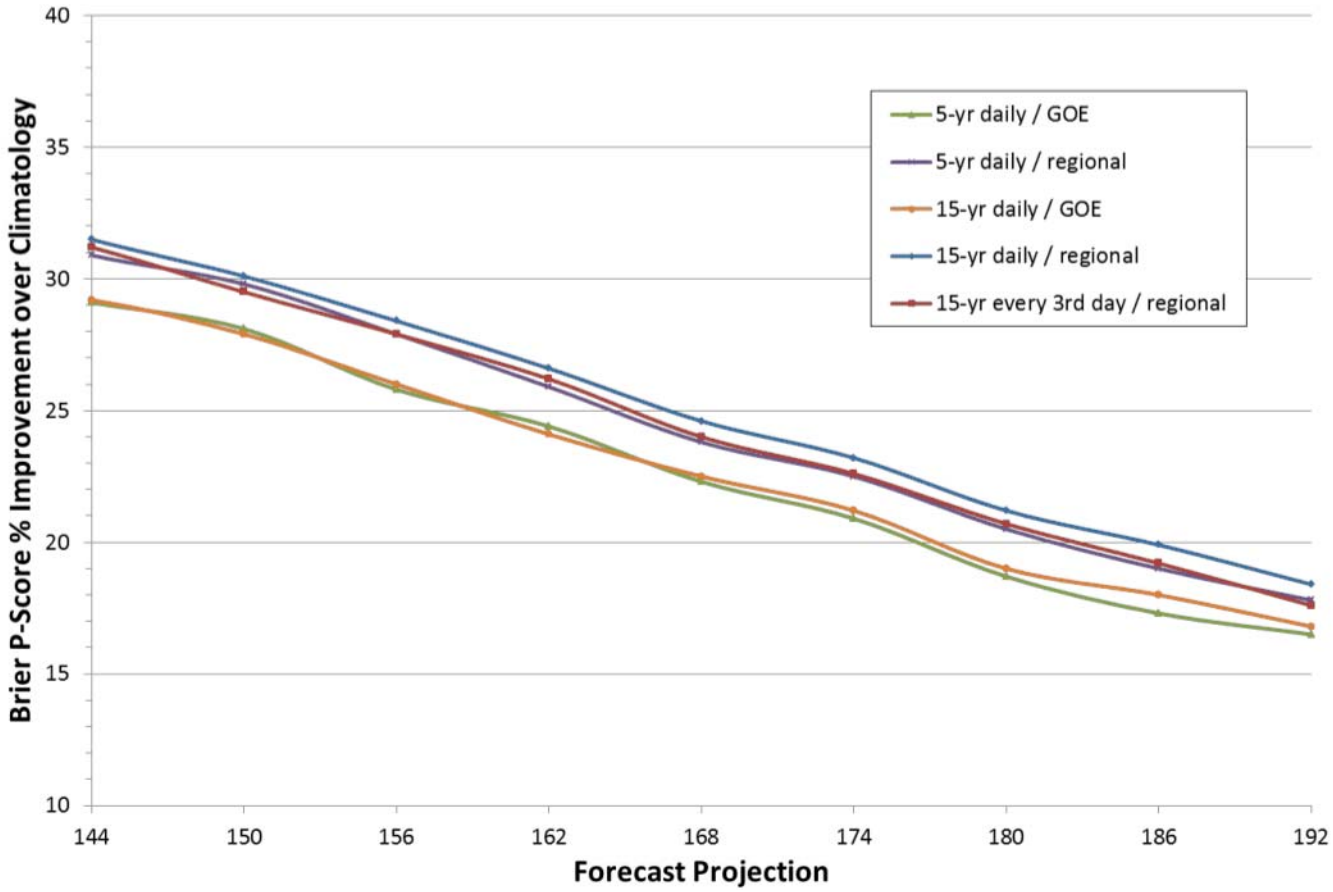
Continuous ranked probability score (CRPS; lower is better) for 2-meter surface temperature forecasts over the US during the cool season.

Sample-size sensitivity
MDL's wind forecasts ≥ 10 kts.



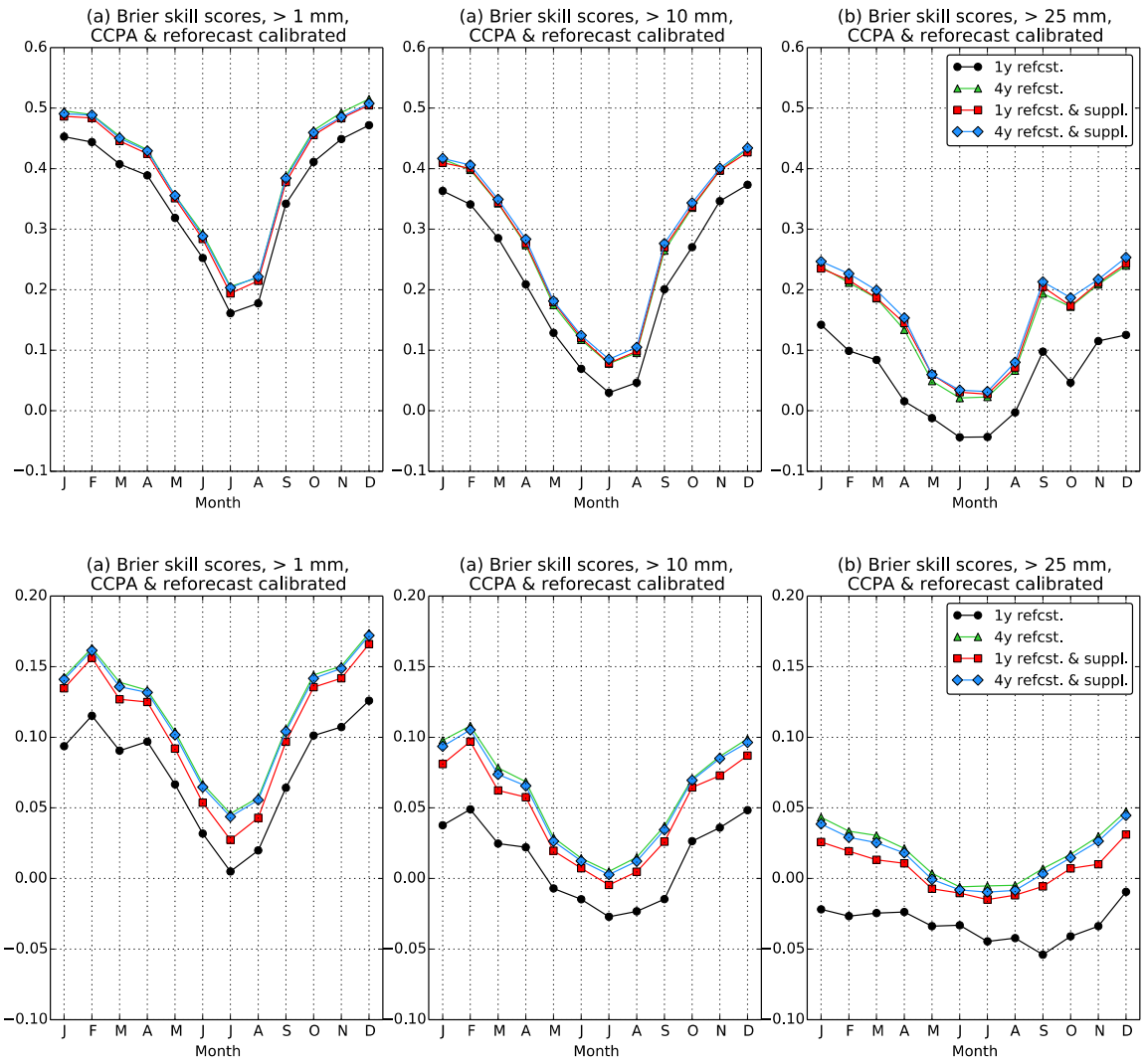
Cool-season (panels a and c) and warm season panels (b and d) mean absolute error (MAE; smaller is better) scores by projection for GEFS MOS wind speed as a function of training sample size. Top panels provide errors of wind speeds when the validation data was > 10 kts, and bottom panels provide errors of wind direction.

Sample-size sensitivity
MDL's precipitation type



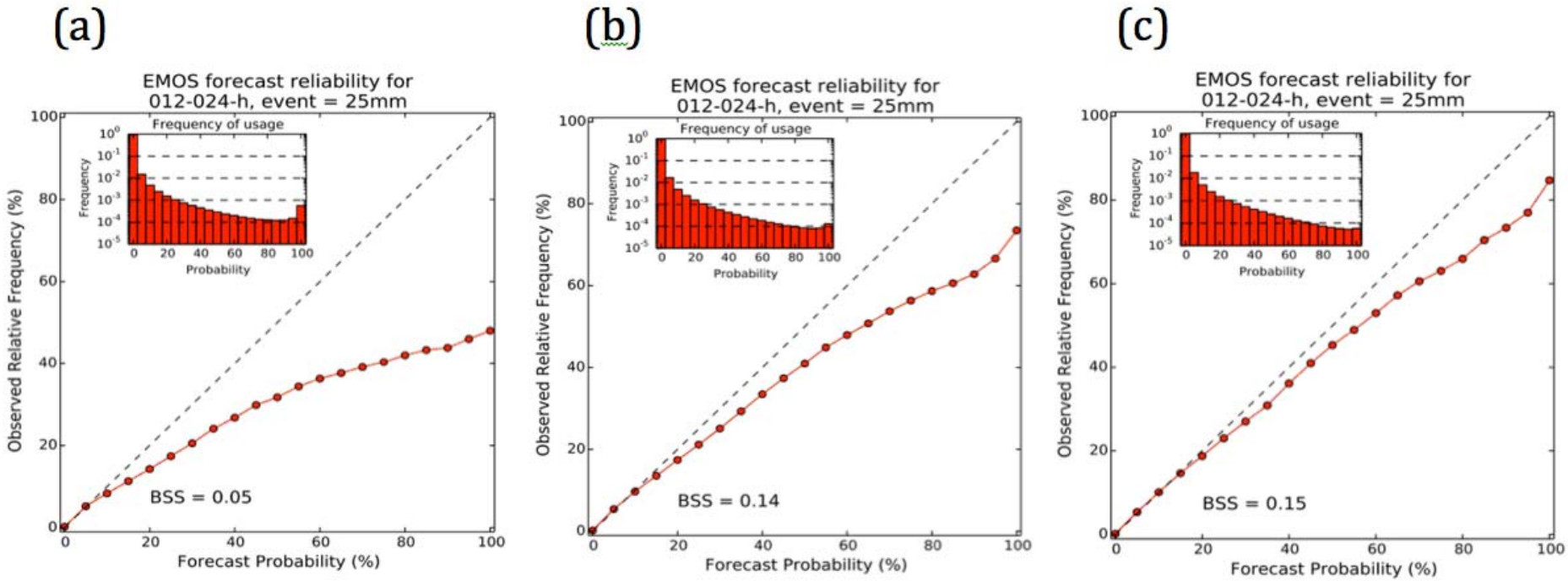
Brier skill score (larger is better) for precipitation type forecasts over the US. "Regional" and "GOE" forecasts were tested, developing the precipitation type either separately for distinct regions of the use or over the entire US, respectively. Training sample sizes included 5 years every day, 15 years every day, and 15 years every third day.

Sample-size sensitivity: PQPF



Brier Skill Scores for US precipitation forecasts at a lead time 012–024 h as a function of month of the year. (a) event of > 1mm/12h, (b) > 10mm/12h, and (c) > 25 mm/12h. Different training sample sizes are shown with the different colored curves. Red and blue curves display skill when the training data at each point was supplemented with training data from 20 other points chosen to have similar climatologies and forecast errors.

Sample-size sensitivity: Reliability for ≥ 25 mm / 12h

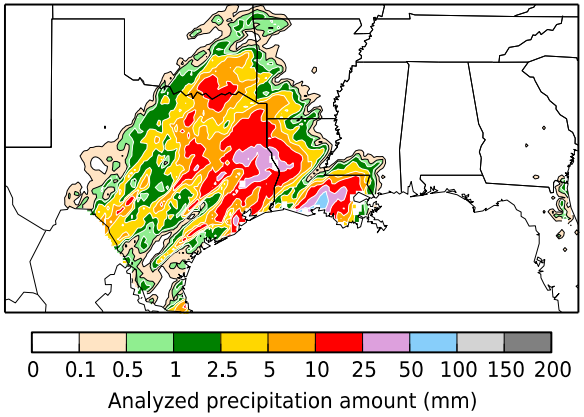


Reliability diagrams for the > 25 mm/12-h forecast event and 12-24 hour forecast lead times. (a) with 1 year of training data, (b) 4 years of training data, and (c) 4 years and 20 supplemental locations for each grid point. Inset histograms provide the frequency of usage for each probability category, in 5-percent intervals.

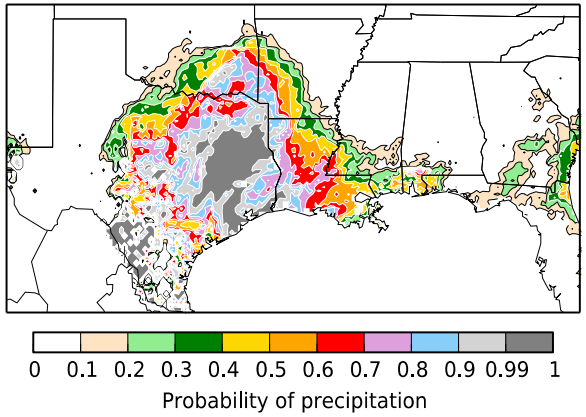
Sample-size sensitivity

improved spatial consistency with larger training sample size

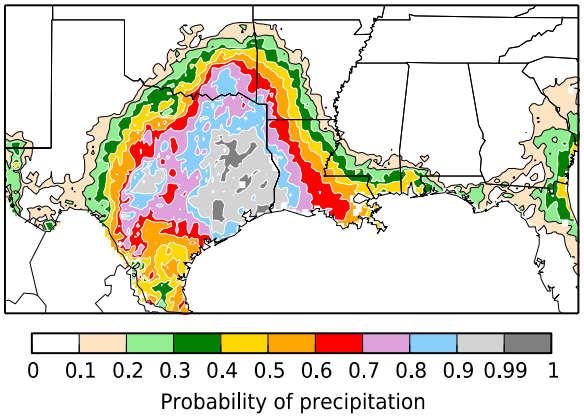
(a) 12-h accumulated precipitation analysis, 2013010812 to 2013010900



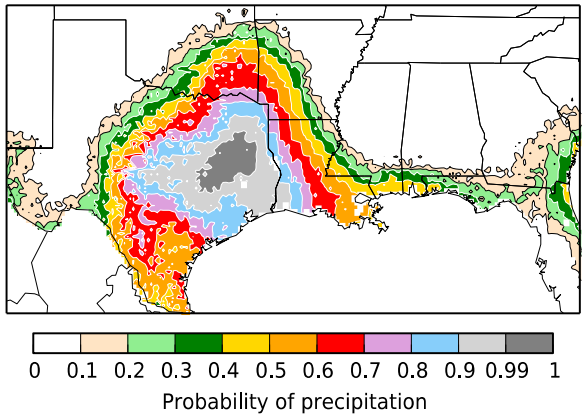
(b) Probabilistic forecast P(>1 mm): 1 year



(c) Probabilistic forecast P(>1 mm): 4 years



(d) Probabilistic forecast P(>1 mm): 4 years + suppl.



(a) 12-h analyzed precipitation amount ending on 00 UTC 09 January 2013, and post-processed forecasts of the probability of precipitation accumulations > 1 mm/12 h obtained with parameter estimates based on training sets corresponding to (b) 1 year, (c), 4 years, and (d) and 4 years plus supplemental locations.

Recommendations

Recommendation 1

- Until a next-generation reanalysis and reforecast in place and ready for utilization, NCEP/EMC should continue the production of an 11-member GEFS ensemble for the 00 UTC cycle in its current (circa 2012) configuration. These real-time forecasts will be approximately consistent with the GEFS reforecast, so existing products can continue to be generated from them. Given the next-generation GEFS will be higher in resolution, this will be a minor computational expense.

Recommendation 2

- NOAA should immediately begin preparations for the production of a next-generation reanalysis to support the reforecast generation process, as well as to facilitate other applications inside and outside of NOAA. The reanalysis configuration should match the operational data assimilation configuration as much as possible. The necessary preparations include determining the computational, storage, and personnel resources needed, as well as organizing the observational data that will be assimilated. The configuration details of the data assimilation system to be used in the reforecast should be decided in consultations between relevant NWS and OAR scientists. We assume that a future reanalysis will be ensemble-based, providing a number of initial analyses suitable for ensemble reforecast initialization.

Recommendation 3

- NOAA should prepare to conduct a reforecast using the anticipated operational configuration of the GEFS. We recommend the following configuration for a GEFS reforecast: 20 years, once every 5 days, with 5 members, and twice daily, from the 00 and 12 UTC cycle. This would be an extra 200 members computed every 5 days, compared with the operational 21x4x5 = 420 members computed in those 5 days, i.e., an extra ~50% computational expense. The discussion below will include possible ways to deal with this computational burden.

Recommendation 4

- The GFS should have at least 2 years of retrospective forecasts computed for it prior to implementation, given that it is also expected to be post-processed by MDL and used for a variety of applications such as in the blender project. Since typically 6 months of prior forecasts are already computed for quality assurance, this request is rather modest. Skipping days between retrospective samples is acceptable.

Recommendation 5

- Given the requirements for NDFD guidance of surface weather elements at high (2.5-km) resolution, NCEP should devote the necessary resources to generate a high-quality retrospective analysis of surface weather with its Real-Time Mesoscale Analysis System.

Discussion of recommendations

- Computational expense of retaining 11-member GEFS after next implementation?
  - approximately 4 percent of the cost of the anticipated 84-member (21 x 4 cycles) T574 GEFS.

Ideas: how to fit new reforecasts with constrained computational resources

- Allocate 50% more time to GEFS, with expanded WCOSS computing in coming years.
- Back off model resolution increases. For example, instead of next-next-generation GEFS going to T800, perhaps going to T700 will allow reforecasts to fit into time allowed.
- Free up cycles by eliminating or shortening the 06Z, 18Z cycles of GEFS.
- Compute reforecasts during slack time in production schedule (they don't need to be real-time).

Conclusions

- Recommendations for reforecasts delivered here.
- Philosophically speaking:
  - *Post-processing needs to be thought of as an integral part of the NWP process*.
  - NCEP and NWS should give the same attention to detail to post-processing that they give to development of dynamical cores, assimilation methods, parameterizations, ensemble systems.
  - Expect stumbles, but we'll get there if we work together.
- Associated white paper (DRAFT) available at http://www.esrl.noaa.gov/psd/people/tom.hamill/White-paper-reforecast-configuration.pdf

Discussion

- Are participants ok with these tests and with the recommendations?
  - Any suggested modification?
  - Any absolutely key tests missing?
  - New content for associated white paper, figures you think are essential?
- We need participants to finish their sample-size test reports and provide me with hyperlinks.