Addressing model uncertainty through statistical post-processing

Tom Hamill

NOAA Earth System Research Lab, Boulder, Colorado, USA

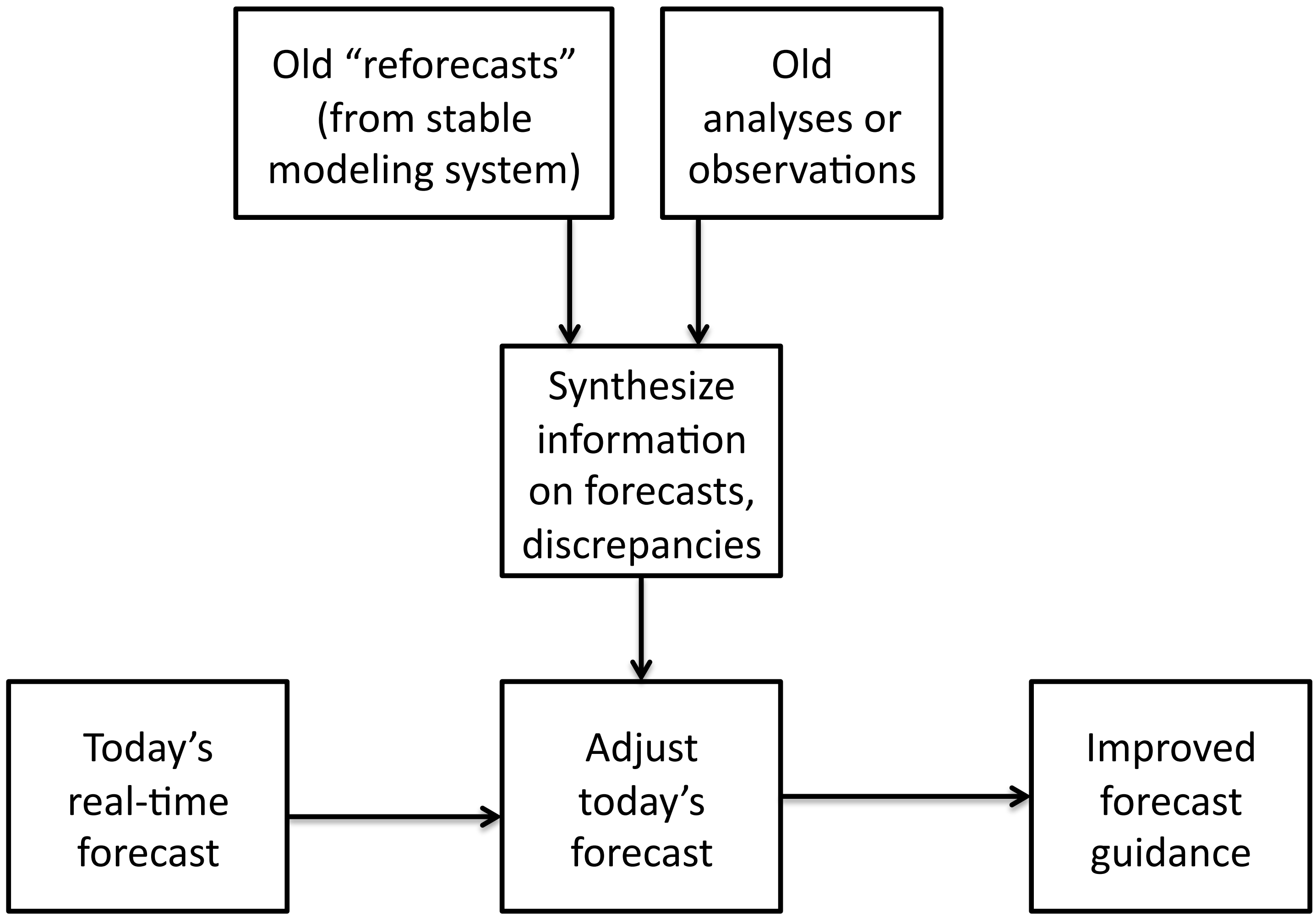
tom.hamill@noaa.gov

001 (303) 497-3060

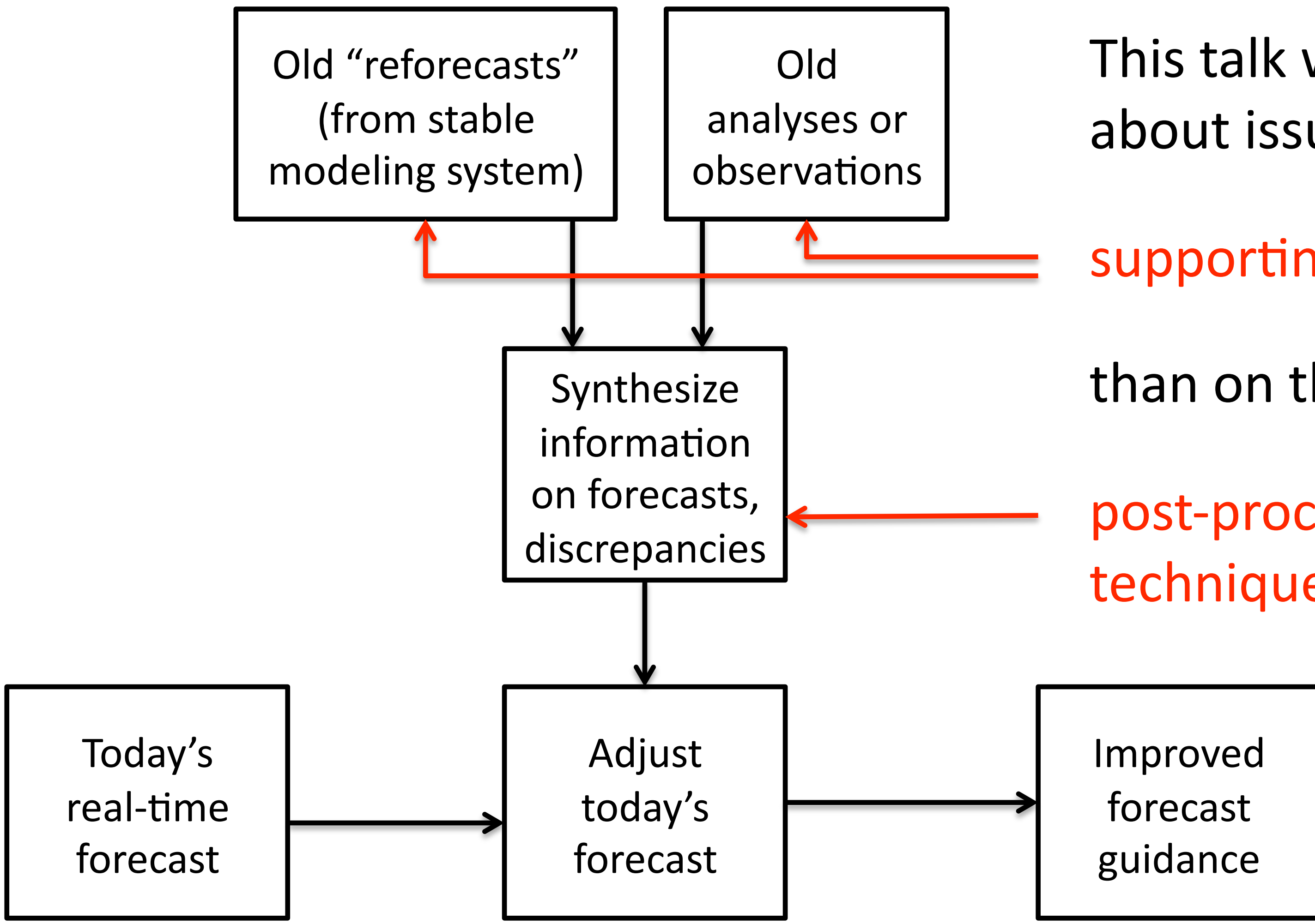
Acknowledgments: Jeff Whitaker, Renate Hagedorn, Florian Pappenberger

statistical post-processing: the janitorial service of numerical prediction

Janitors' preferred cleaning technique



Janitors’ preferred cleaning technique



This talk will be more about issues with the

supporting data sets

than on the specific

post-processing techniques

What can post-processing and reforecasts do that other model error corrections cannot?

- Provide context on how unusual today's forecast event is, relative to other *forecast* events.
- Compensate for errors due to finite ensemble size.
- Provide extra "resolution" via statistical downscaling.
- Compensate for remaining systematic model biases not addressed through stochastic techniques, thereby increase reliability, increase forecast skill.
- Provide sufficient samples to quantify forecast errors for particular locations, hydrologic basins

Disadvantages of post-processing

- Right answer perhaps, but for wrong reason? We prefer to directly improve the model in physically realistic ways.
  - Also, some errors are too complex to adjust via post-processing; for these, there is no substitute for improving the model.

- Additional computational and infrastructural burden to compute reforecasts and reanalyses, compile observation time series.
  - ECMWF's (relatively sparse) weekly 5-member reforecast * 20 years = 100 extra members / week to compute.
  - Generally greater benefit the more years, more days, more members in reforecast, but proportionally more expensive.
  - Without high-quality, long observation time series, many of the benefits of reforecasts + statistical post-processing are lost.
  - Need to keep computing reforecasts with current model version, else improvements are temporary.

- If real climate or model-error statistics change significantly during reforecast period, decreased accuracy of post-processed estimates

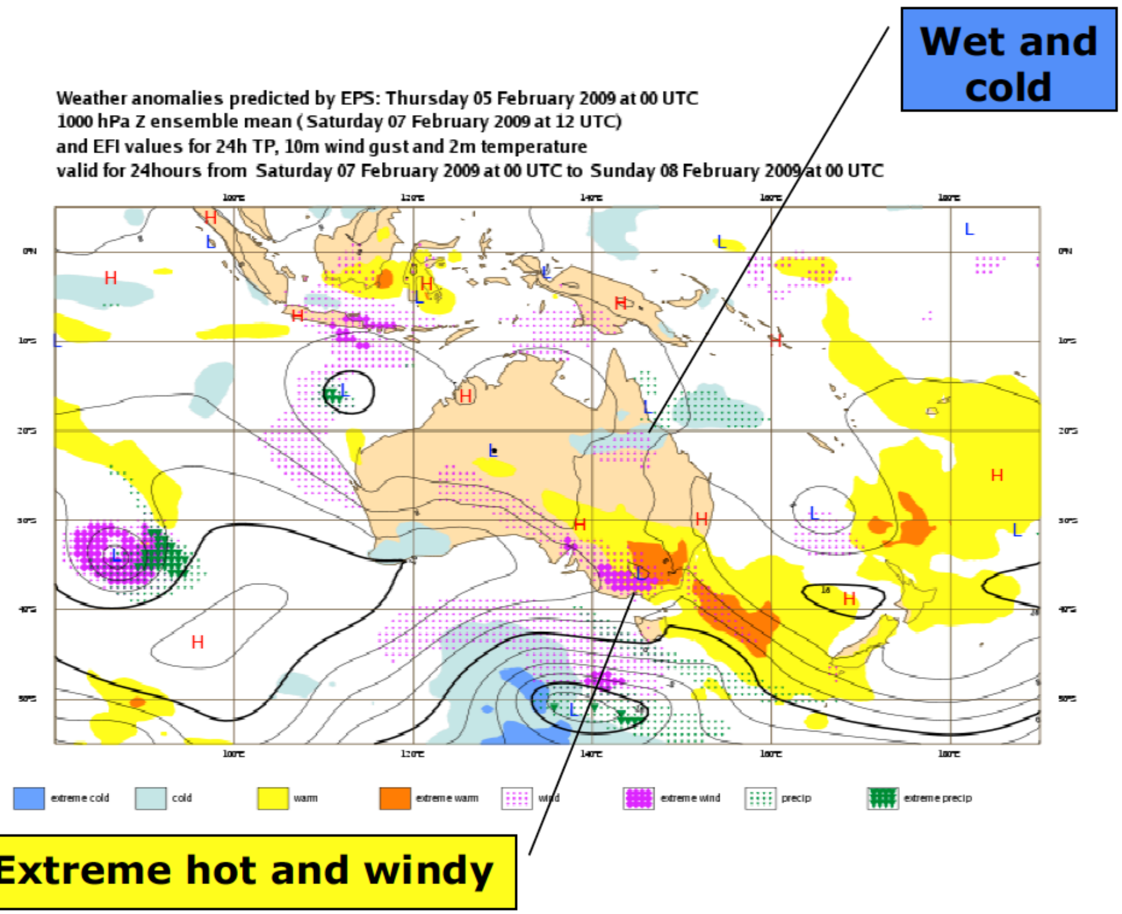
Post-processing and reforecast advantages

Reforecast advantage: facilitates quantitatively assessing how unusual an event is (EFI)

EPS I-EFI 05@00+48/72h vt 07@00-08@00

The forthcoming Interactive EFI (I-EFI) can be used to identifies areas where the ensemble forecast distribution is significantly different from the climatological distribution, and visualize the grid point distributions.

This plot shows the I-EFI +48/72h forecasts issued on 5@00UTC and valid between 7@00UTC and 8@00UTC.



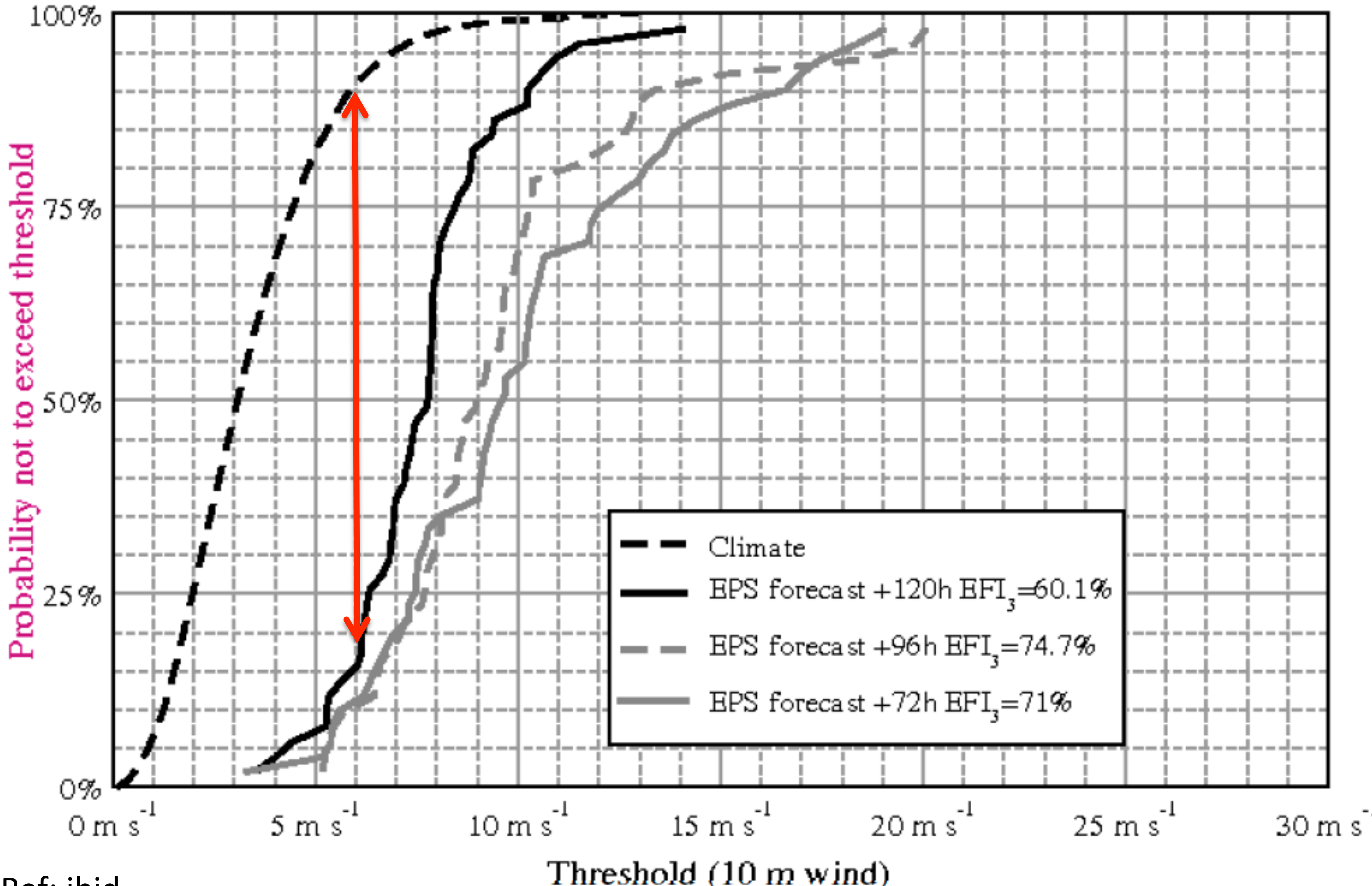
Extreme Forecast Index (needs accurate *forecast* climatology, such as provided by reforecasts)

$$EFI = \frac{2}{\pi}\int_0^1 \frac{p - F_f(p)}{\sqrt{p(1-p)}}$$

p is the percentile of the cumulative distribution estimated from the ensemble; $F_f(p)$ is how the *p*-percentile of the climate record ranks in the EPS (0 the minimum, 1 the maximum). This "Anderson-Darling" version introduces a weighted statistic that gives more power in the tails of the distribution. 2/π is normalization factor to keep -1 ≤ EFI ≤ 1.

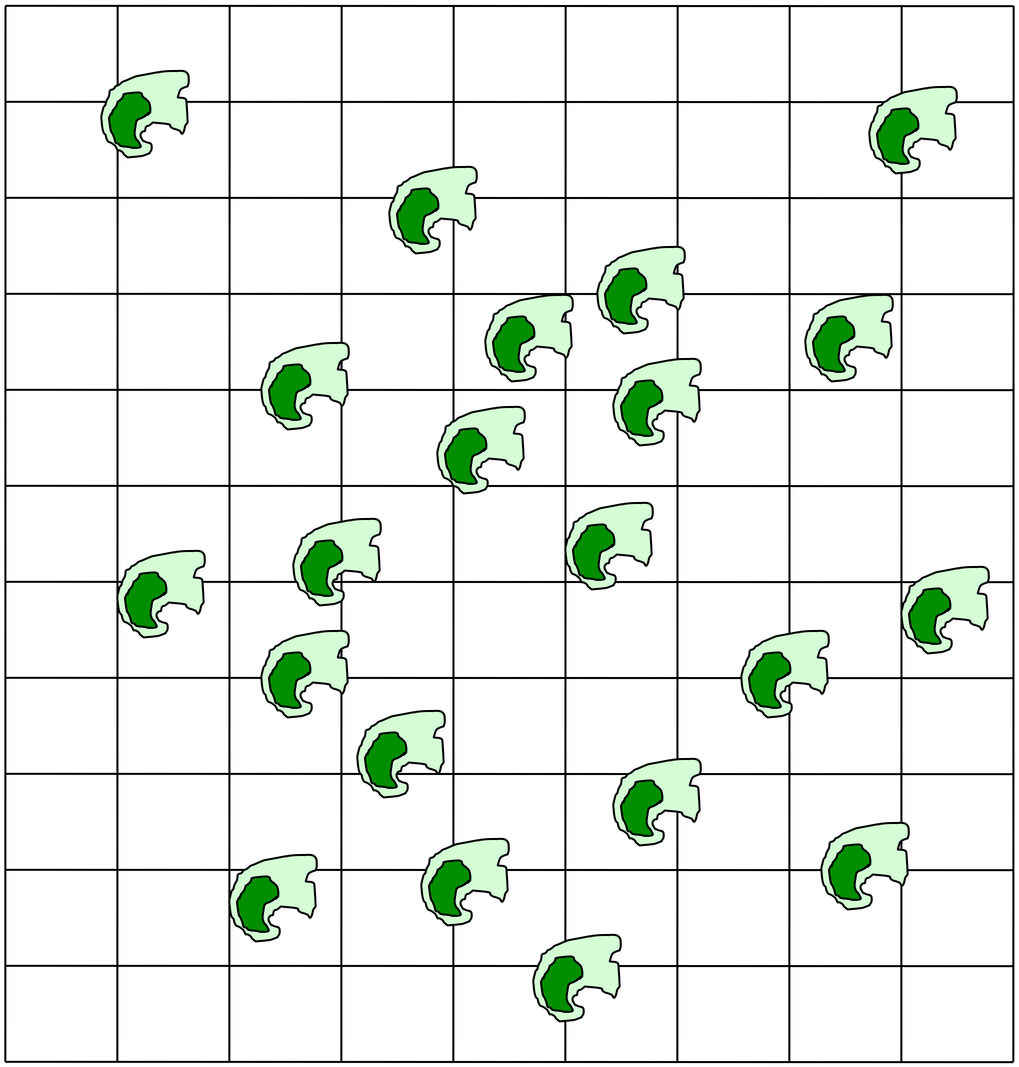
From: LaLaurette, QJRMS, 2003, and http://www.ecmwf.int/products/forecasts/efi_guide.pdf

EFI


100%
75%
50%
25%
0%
Probability not to exceed threshold
0 m s-1
5 m s-1
10 m s-1
15 m s-1
20 m s-1
25 m s-1
30 m s-1
Threshold (10 m wind)
Climate
EPS forecast +120h EFI3=60.1%
EPS forecast +96h EFI3=74.7%
EPS forecast +72h EFI3=71%

Ref: ibid

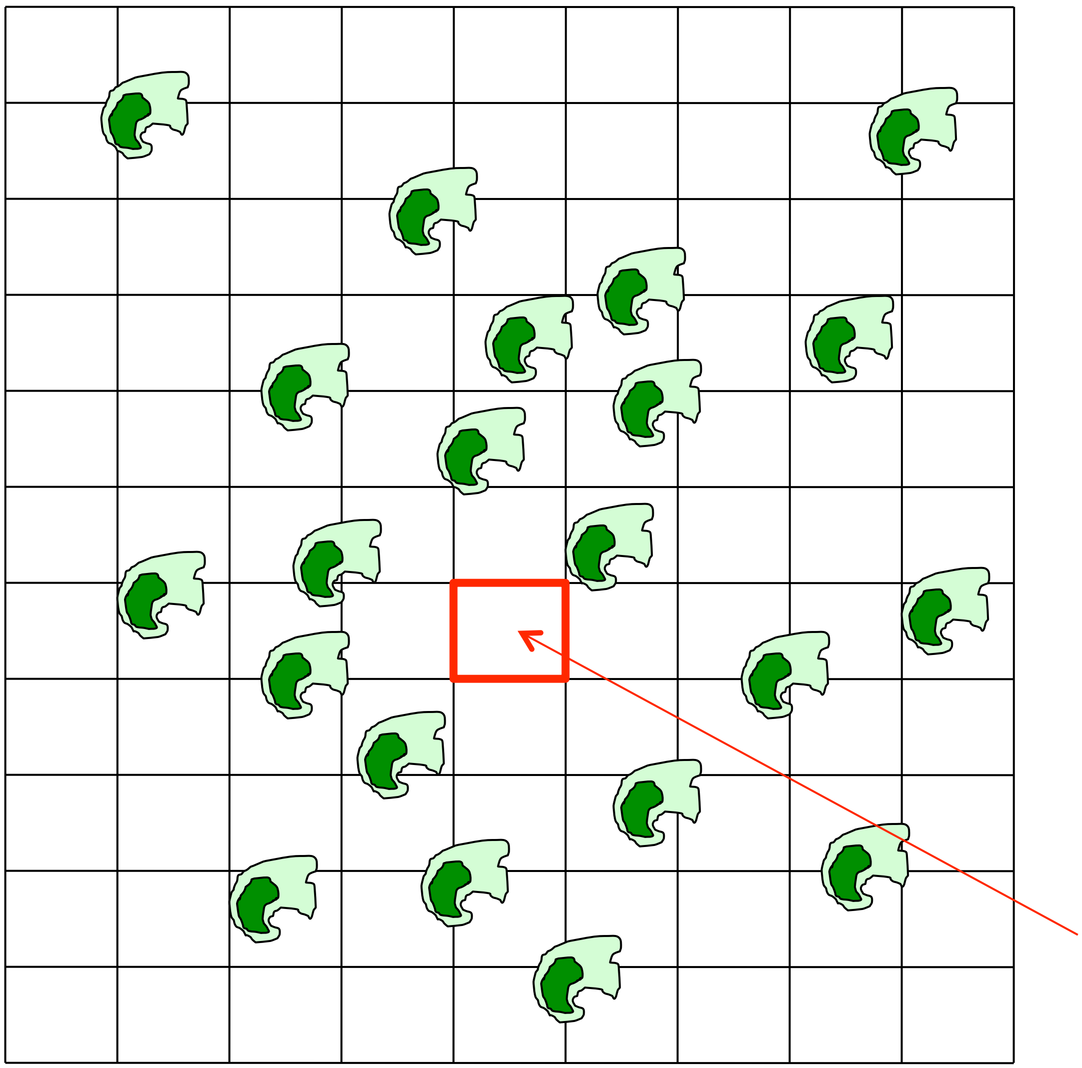
Adjusting for errors due to finite ensemble size.



Imagine: your 20-member storm-scale ensemble (which is a calibrated system, truth consistent with a random draw from ensemble)

Your job: estimate reliable probabilities on the grid.

Adjusting for errors due to finite ensemble size.

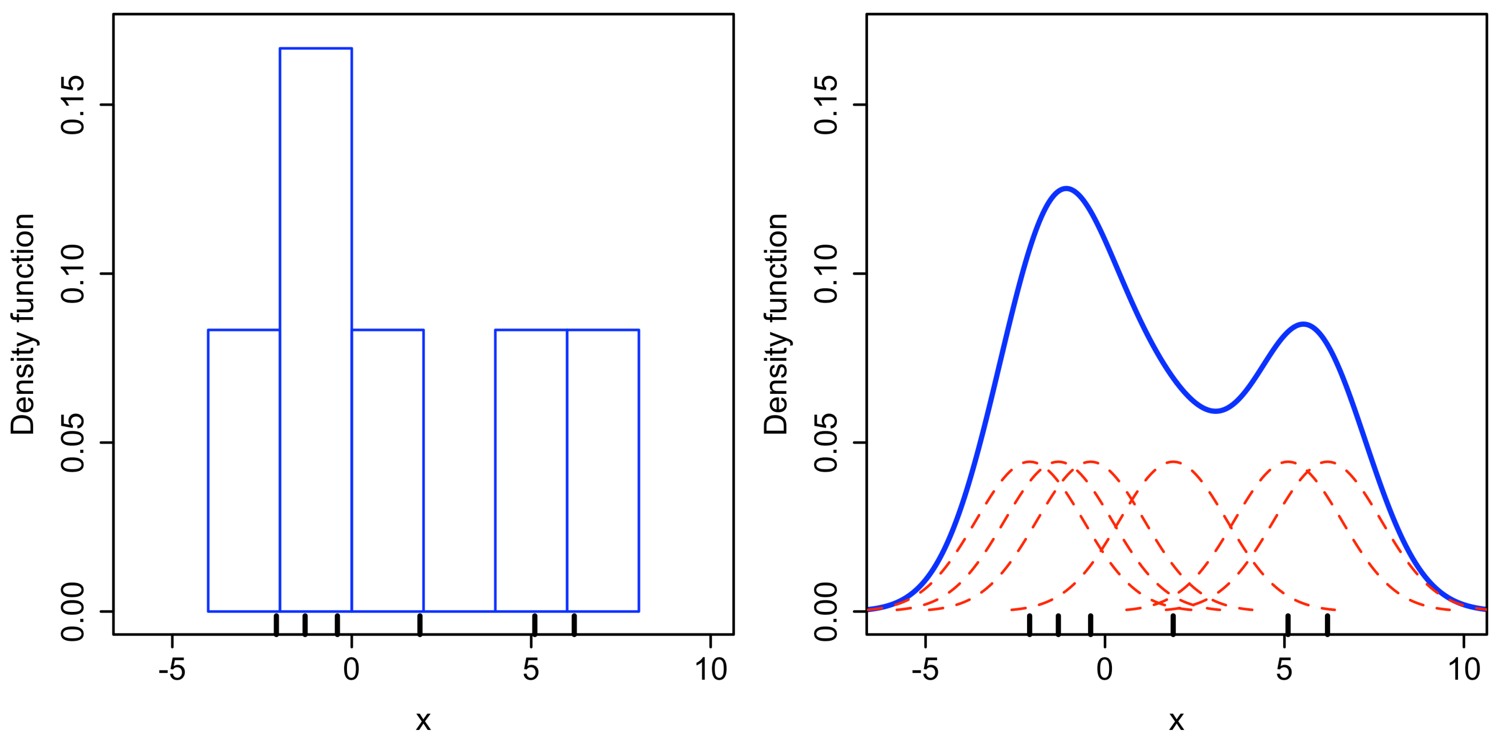


Imagine: your 20-member storm-scale ensemble (which is a calibrated system, truth consistent with a random draw from ensemble)

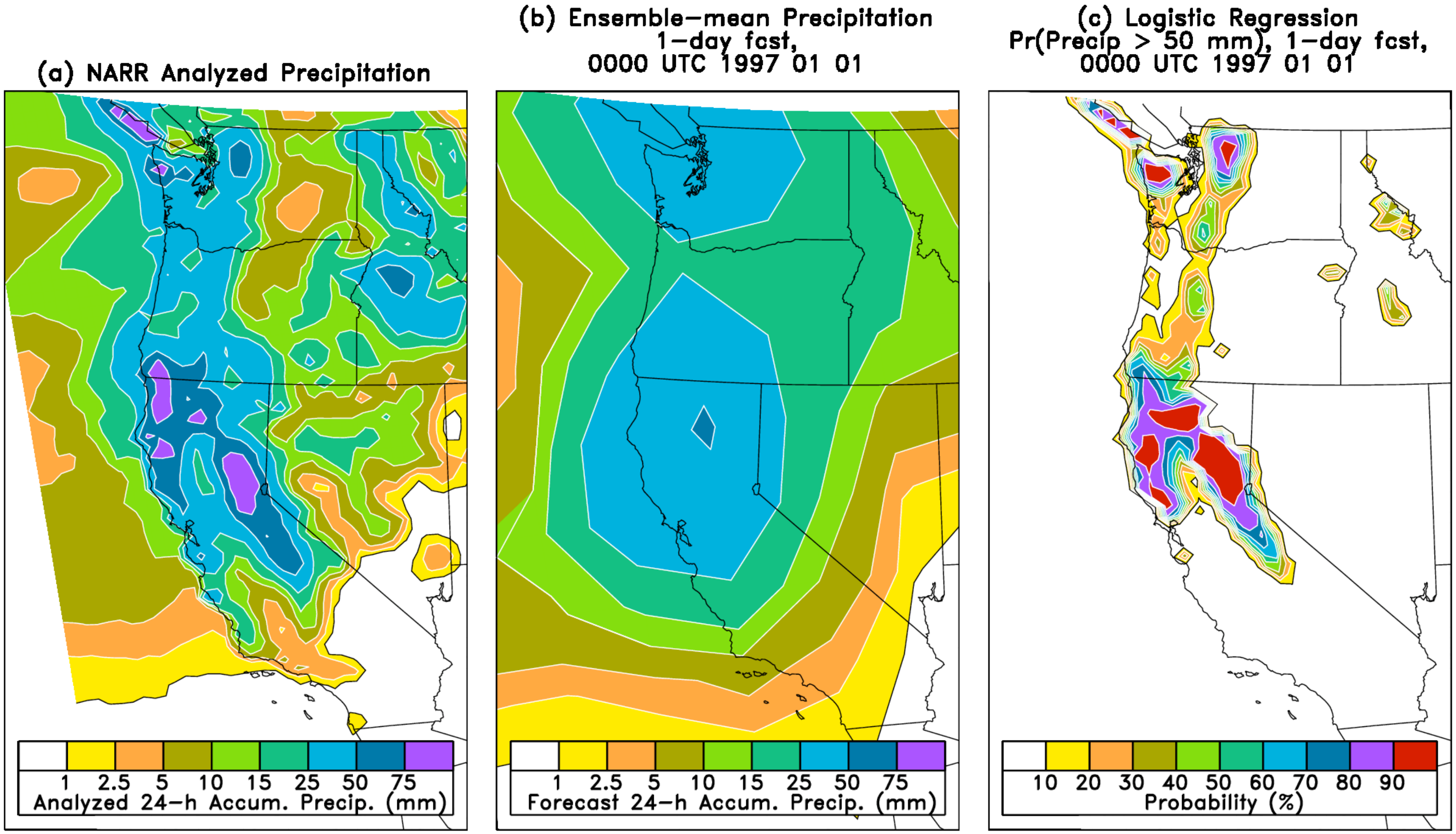
Your job: estimate reliable probabilities on the grid.

Zero probability for this cell? Yes if you use ensemble relative frequency

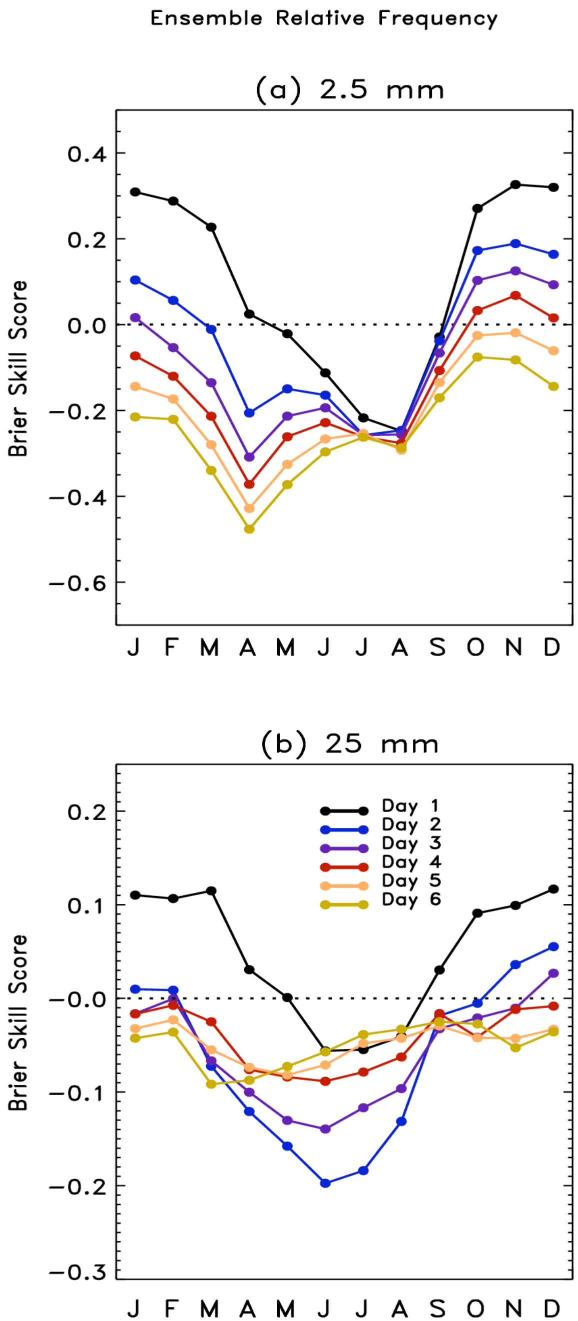
Kernel density estimation to produce smooth pdf from limited-size ensemble

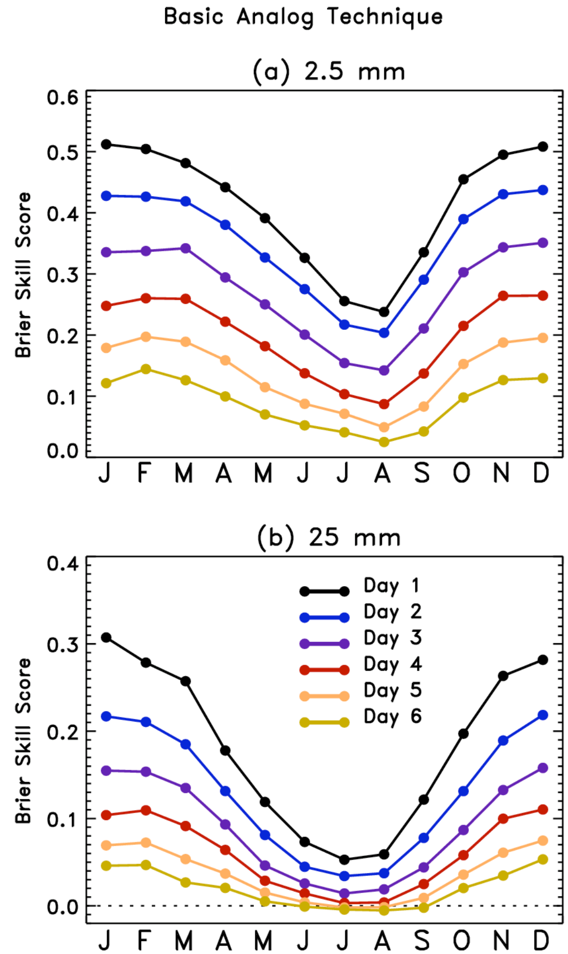


Provide extra "resolution" via statistical downscaling



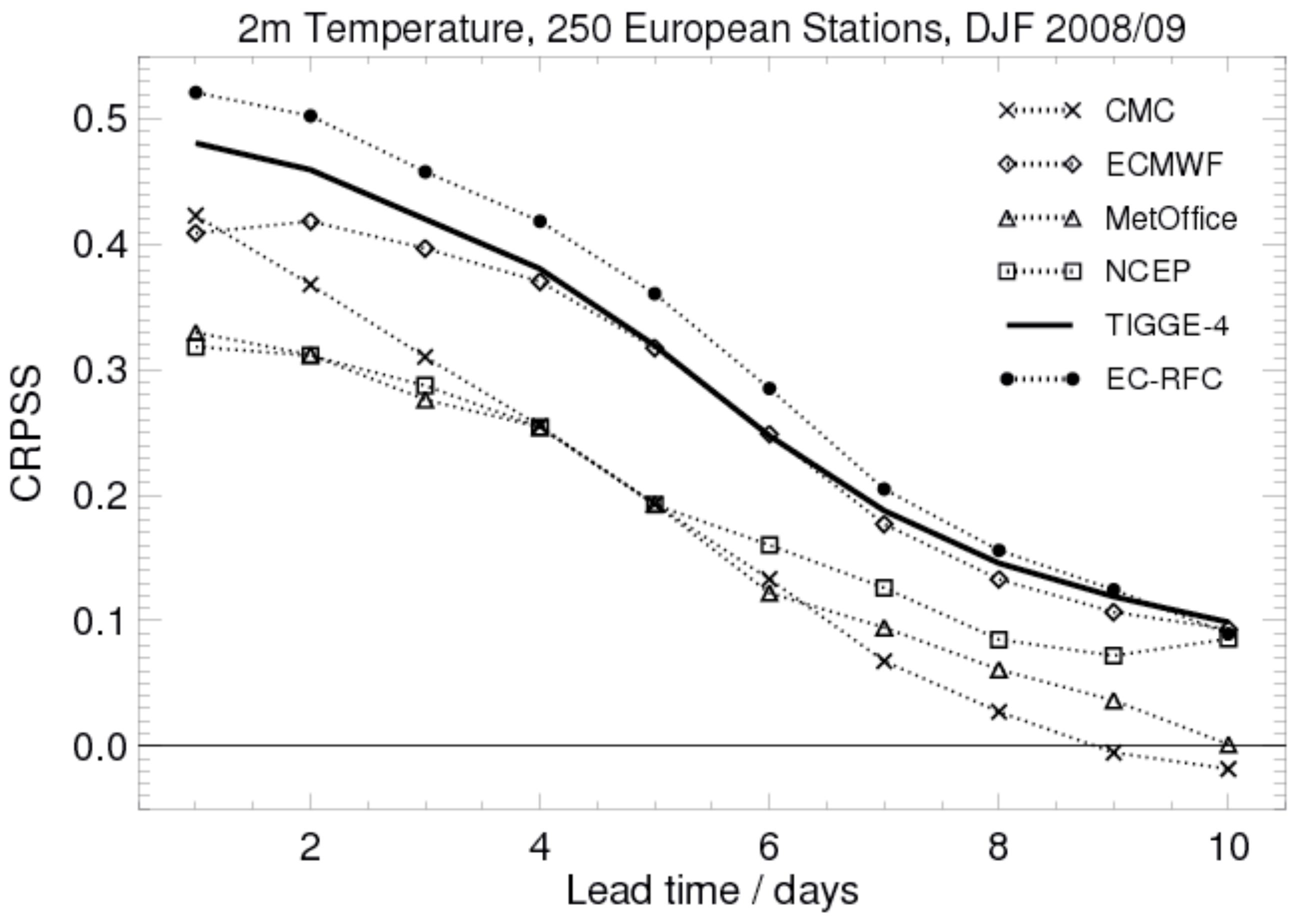
"resolution" here is used as in its definition in the Brier Score decomposition, the ability of a forecast model to successfully forecast deviations from the overall climatological probability.





Verified over 25 years of forecasts; skill scores use conventional method of calculation which may overestimate skill (Hamill and Juras, QJRMS, Oct 2006).

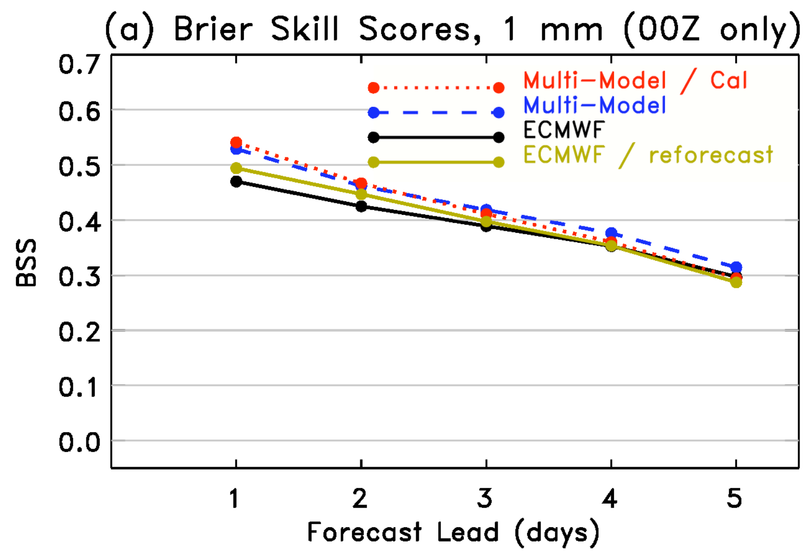
Reforecast vs. multi-model, T_{sfc}

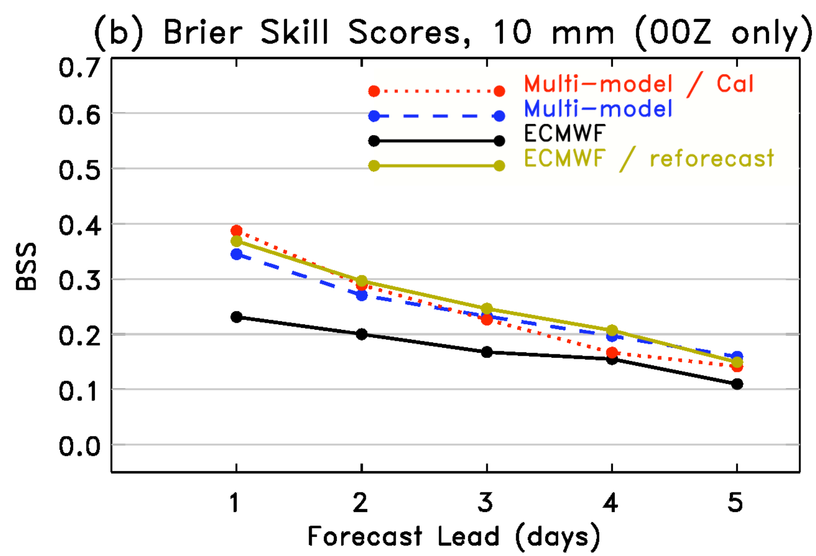


ECMWF's forecasts were corrected here using a blend of bias correction from the past 30 days of forecasts and a more sophisticated regression approach using reforecasts.

courtesy of Renate Hagedorn, ECMWF & DWD

Reforecast vs. multi-model precipitation over US, Jul-Oct 2010





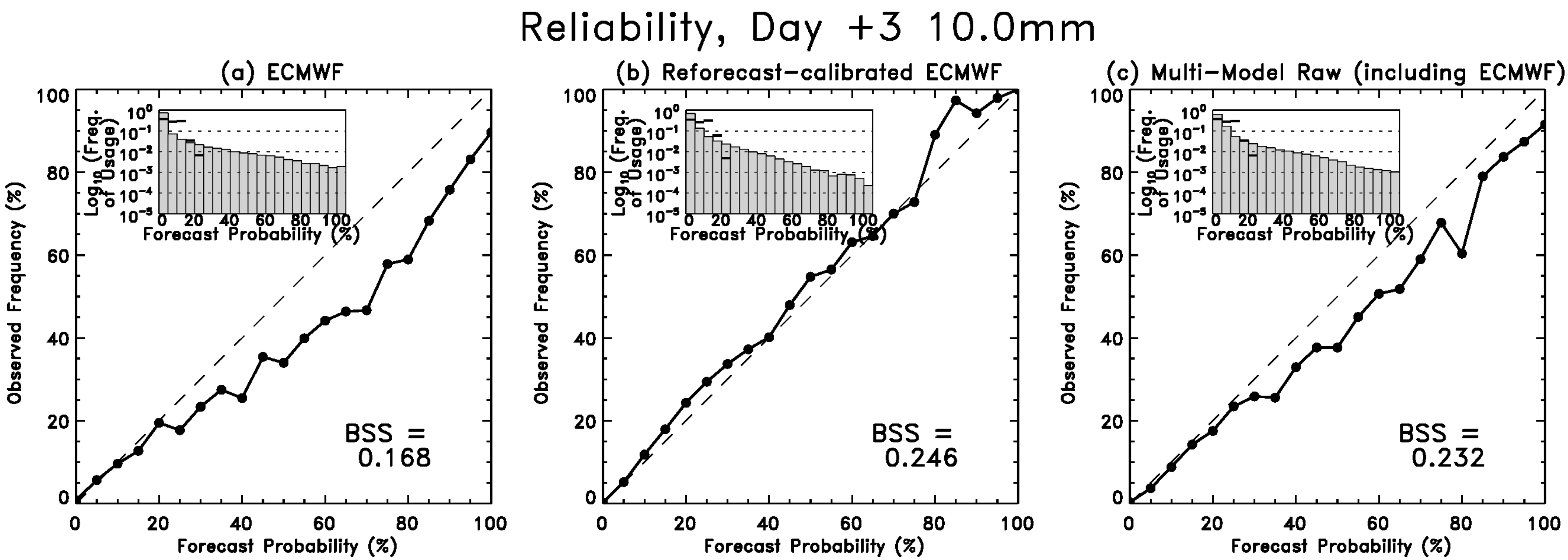
Verification of 1-degree resolution forecasts against 1-degree precipitation analyses over CONUS.

The following forecasts are plotted: 20-member ECMWF forecasts (black); ECMWF, calibrated via logistic regression using 9 years of ECMWF 4-member weekly reforecasts (green); multi-model (blue) and multi-model, calibrated using the last 30 days of forecasts/analyses.

Reforecasts appear to provide most improvement at heavy precipitation thresholds, consistent with other previous results.

Sample reliability diagrams

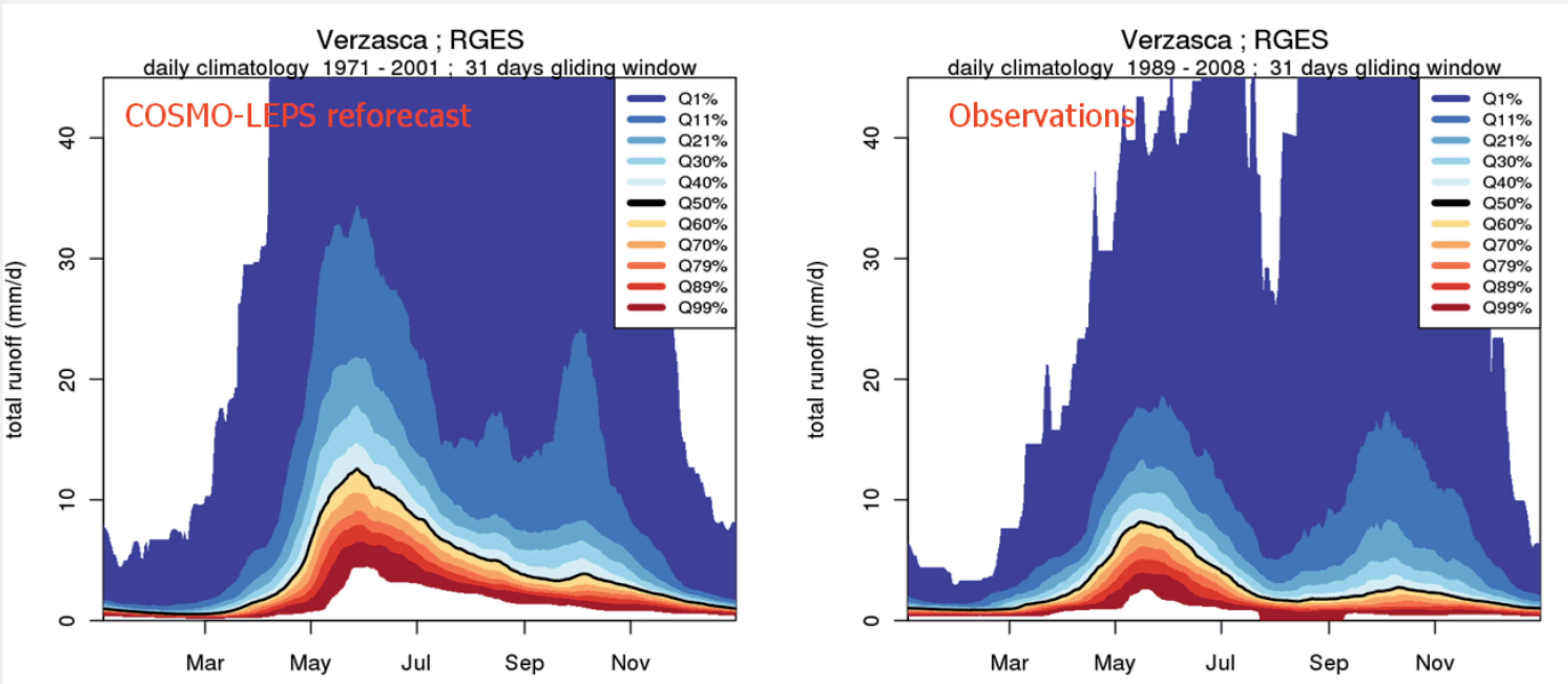
ECMWF, reforecast-calibrated, multi-model



(more reliable)

(sharper)

Reforecast advantage: permits quantifying forecast errors for particular locations, basins

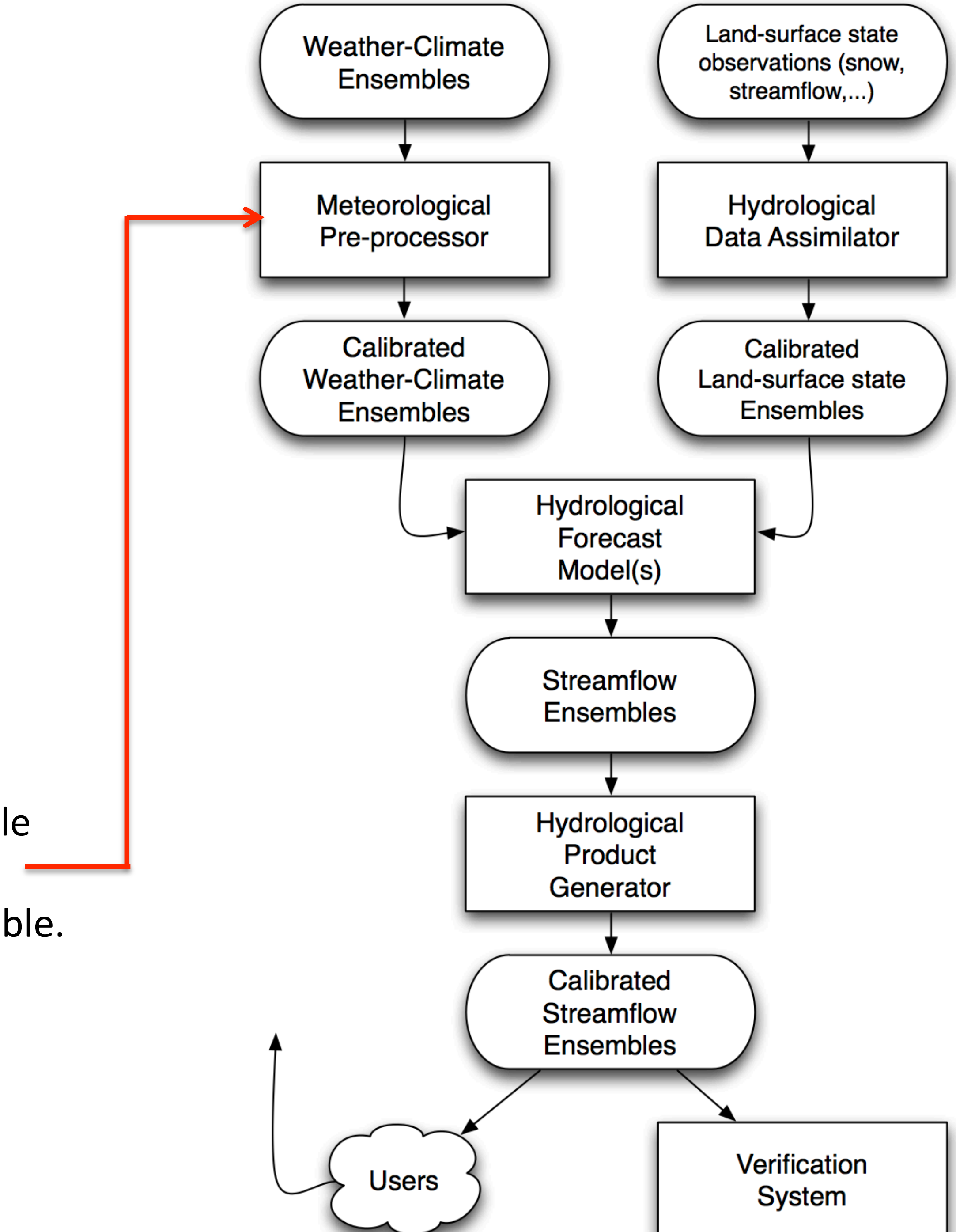


TopLeft: Discharge Climatology Quantiles (30 day gliding mean) for the Verzasca basins obtained forcing the hydrological model PREVAH with COSMO-LEPS reforecasts (1971-2000). TopRight: Observed daily discharge climatology (1989-2008)

from Felix Fundel et al. poster, *ftp://ftp.wsl.ch/pub/zappa/imprints/del3_1/cleps_bcn.pdf*

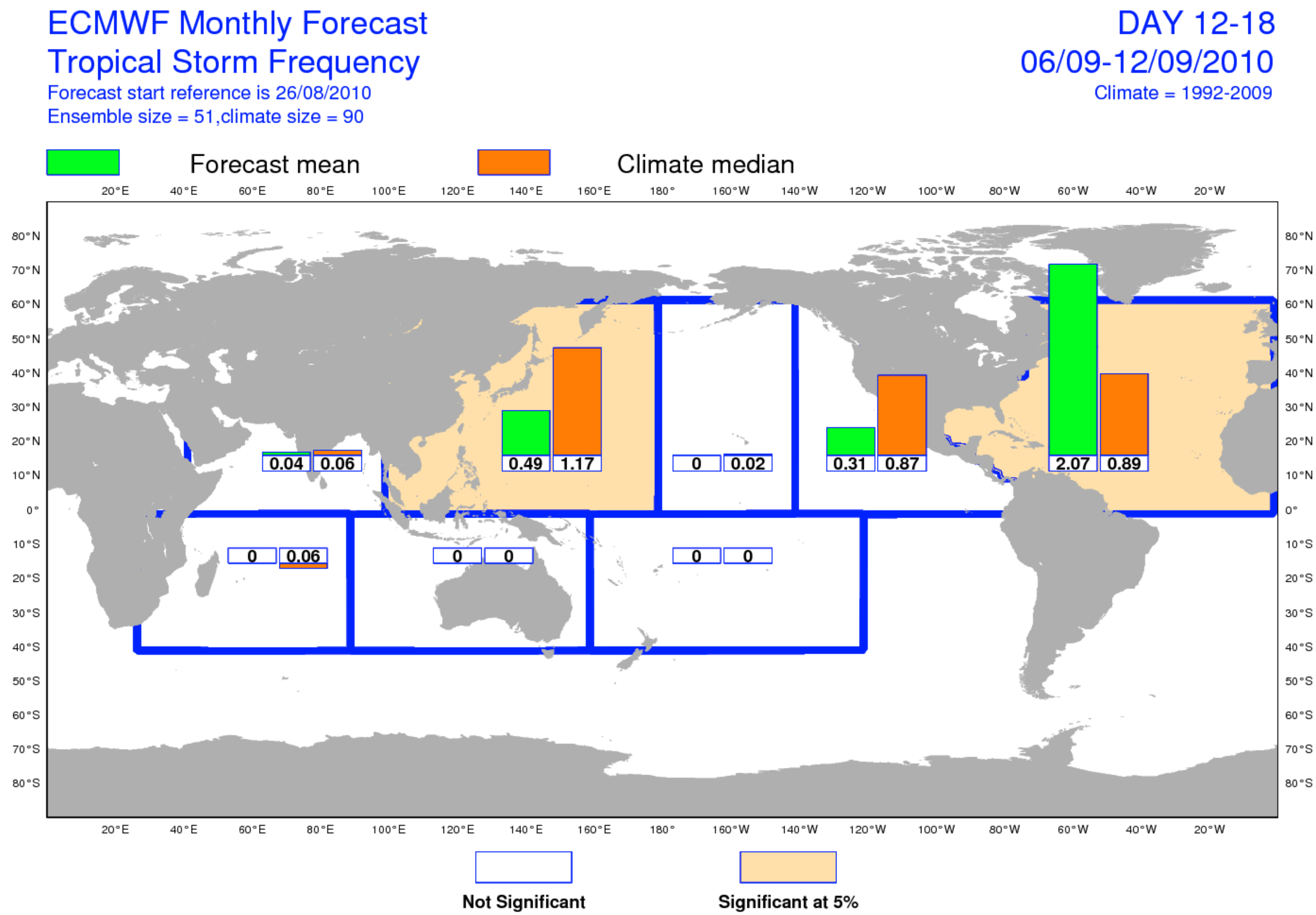
Hydrologic Ensemble Prediction Experiment

Note that hydrologists envision a step to make sure that ensemble inputs to their hydrologic system are as reliable and sharp as possible.



from Schaake et al. 2007 *BAMS* article

Reforecast use: tropical cyclogenesis



Many forecast models over-forecast tropical cyclogenesis. This ECMWF product uses TCgenesis from reforecasts to provide some calibration for possible biases.

Ref: D. Richardson, personal communication, ECMWF.

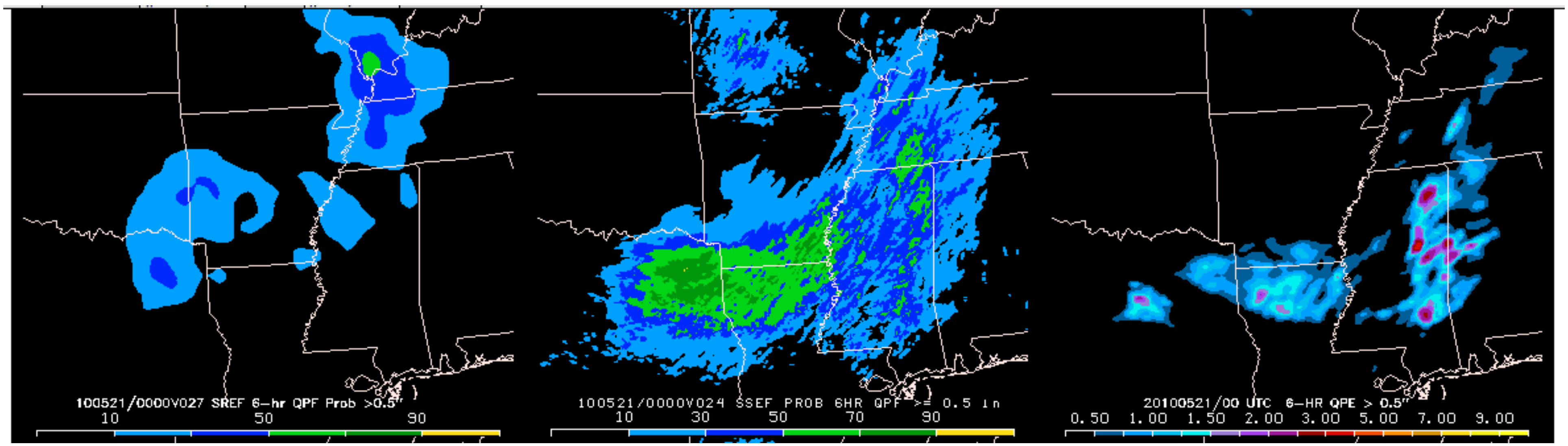
Reforecast / calibration disadvantages

Not all model deficiencies can be addressed easily through post-processing

An example from NSSL-SPC Hazardous Weather Test Bed, forecast initialized 20 May 2010

http://tinyurl.com/2ftbvgs

30-km SREF P > 0.5”　　4-km SSEF P > 0.5 “　　Verification



With warm-season QPF, comparatively coarse resolution and parameterized convection in operational SREF system produces forecast that is clearly inferior to the 4-km, resolved convection in SSEF. Calibration isn’t likely to provide nearly the improvement that the extra resolution will provide.

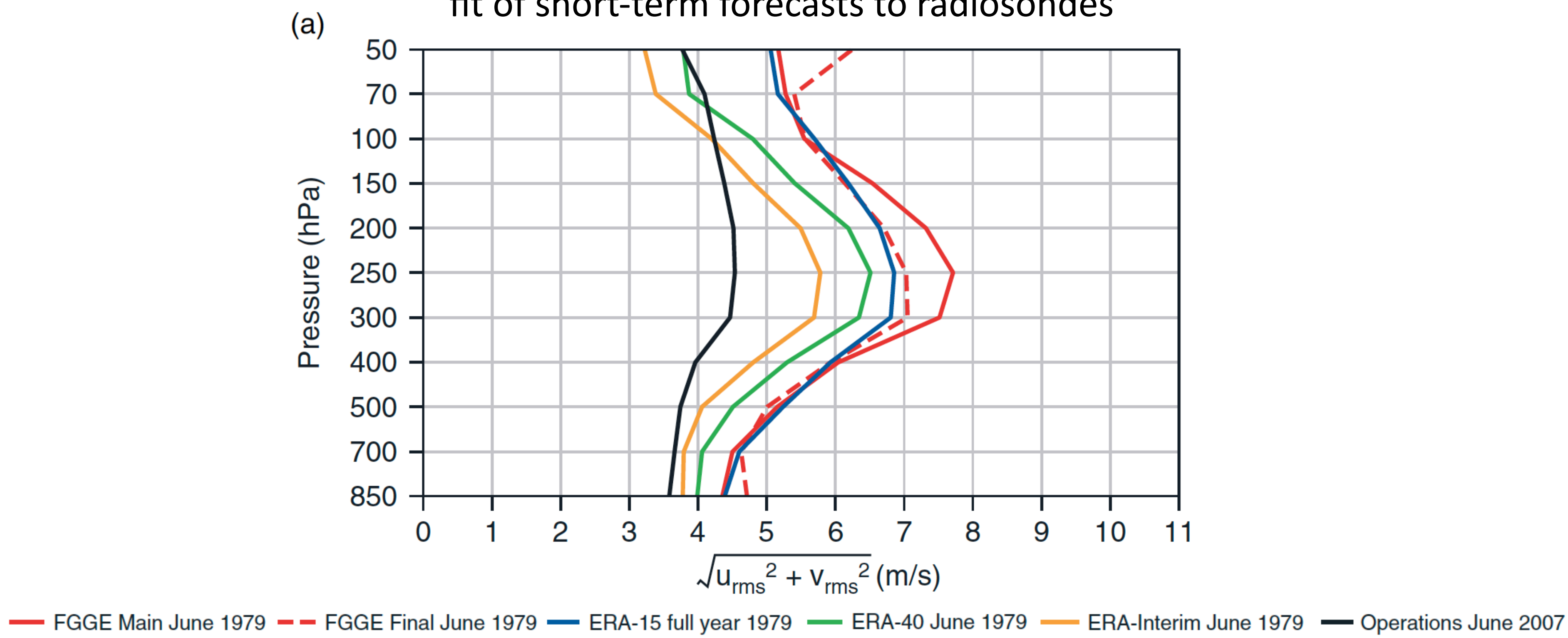
Computational burden

- Real-time ensemble: assume 50 members, 2x daily = 100/day = 700/week
- Minimal reforecast: 5 members, 20 years, 1x weekly = 100/week : 1/7 extra
- Moderate reforecast: 10 members, 30 years, 1x daily = 2100/week : 3x extra.
- Full reforecast: 50 members, 30 years, 2x daily = 21000/week : 30x extra!

Disadvantage: non-stationary forecast errors in reforecasts?

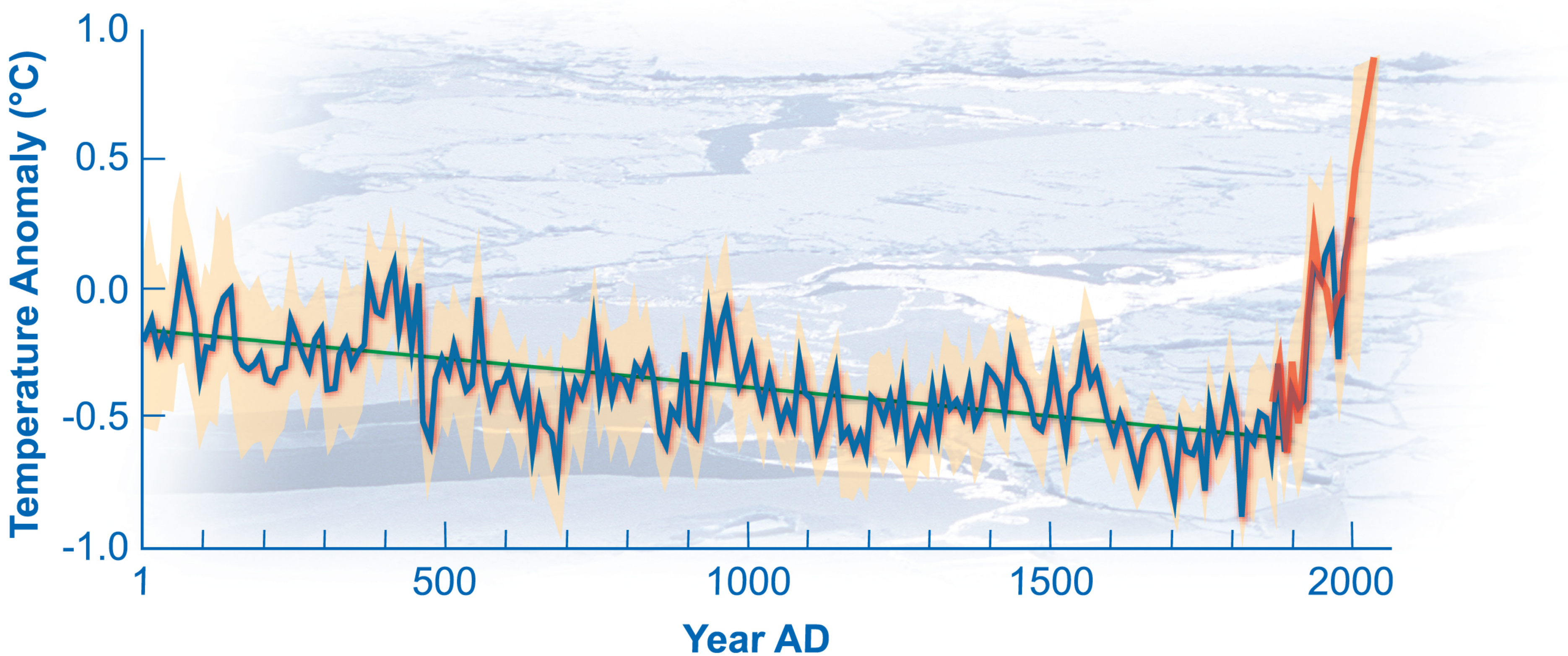
- If real climate or model-error statistics change significantly during reforecast period, decreased accuracy of post-processed estimates.

fit of short-term forecasts to radiosondes



From Dee et al., QJRMS, 2011 article on ERA-Interim

Changing climate: today's forecasts warmer than those in training data set?



If forecast today is warmer than any in the reforecast training data set, we'll be "extrapolating the regression" when we apply statistical corrections.

Conclusions

- Statistical post-processing (using reforecasts) may complement other methods of addressing model uncertainty.
  - correct for model bias.
  - generally large improvements in forecasts of rare events.

- It's not a solution for all problems, and it does increase system complexity, computational burden.

- Worth considering how these data sets may be leveraged to facilitate model uncertainty research.