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
Workshop on representing model uncertainty and error in numerical weather and climate prediction models

20-24 June 2011, at ECMWF
Sponsored by ECMWF, WGNE, THORPEX, WCRP

There are a number of techniques to represent model uncertainty in ensemble forecasts. These range from the multi-model techniques which feature prominently in IPCC assessment reports, to the stochastic parametrisations pioneered at ECMWF and widely used at weather forecast centres around the world. A key outcome of the meeting was that the stochastic parametrisation paradigm needs further development at the process level, and hence needs to be incorporated as part of general parametrisation development. Key tools will include sophisticated analyses of observational datasets, output from cloud resolving models, and analyses from objective data assimilation. Data assimilation techniques themselves will benefit from better representations of model uncertainty.
Statistical post-processing and its niche in addressing model uncertainty

Statistical post-processing is the janitorial service of ensemble prediction.

If ensemble forecasts were consistently unbiased, reliable, and sharp, with physically based model uncertainty parameterizations, there would be little need for this activity.

For the foreseeable future, we there will be enough “mess” to clean up from the raw model guidance that you’ll need our services.
Statistical post-processing and its niche in addressing model uncertainty

Reforecasts, the tool that helps us NWP janitors do our job.
Basic idea

Old “reforecasts” (from stable modeling system)

Old analyses or observations

Synthesize information on forecasts, discrepancies

Today’s real-time forecast

Adjust today’s forecast

Improved forecast guidance
Basic idea

This talk will be more about issues with the supporting data sets than on the specific post-processing techniques.

- Today’s real-time forecast
- Adjust today’s forecast
- Improved forecast guidance

Old “reforecasts” (from stable modeling system) → Old analyses or observations → Synthesize information on forecasts, discrepancies → Adjust today’s forecast → Improved forecast guidance
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 identify 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.

Wet and cold

Extreme hot and windy
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. \( 2/\pi \) is normalization factor to keep \(-1 \leq EFI \leq 1.\)

Ref: ibid
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.
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.

Adjusting for errors due to finite ensemble size.
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

Reliability, Day +3 10.0mm

(a) ECMWF

(b) Reforecast-calibrated ECMWF

(c) Multi-Model Raw (including ECMWF)

BSS = 0.168
BSS = 0.246
BSS = 0.232

(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)

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 / 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

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

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.

From Dee et al., QJRMS, 2011 article on ERA-Interim
Changing climate: today’s forecasts warmer than those in training data set?

![Temperature Anomaly Chart](chart.png)

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.