BAYESIAN PROCESSOR OF ENSEMBLE (BPE): CONCEPT and IMPLEMENTATION

By

Roman Krzysztofowicz, *University of Virginia*

Zoltan Toth, *EMC / NCEP / NWS / NOAA*

Presented at the
4th NCEP / NWS Ensemble User Workshop
Laurel, Maryland

13–15 May 2008

Acknowledgments:

National Science Foundation, Grant No. ATM–0641572
"New Statistical Techniques for Probabilistic Weather Forecasting"

Dr. Stephen J. Lord, Environmental Modeling Center, NCEP / NWS/ NOAA
"Support for Collaboration between EMC and UVA"

PROCESSOR: Performance Requirements

1. Calibration of Forecasts (Reliability, Unbiasedness)
   - Probability distributions
   - Ensemble trajectories

2. Robustness of Estimators (Fast Convergence in Adaptive Mode)
   - Minimize the required sizes of joint samples

3. Computational Efficiency
   - Run on all model output variables

4. Structural Flexibility — Combine information:
   - Ensembles from multiple centers
   - High resolution forecasts
   - Observations (climatic data, re-analysis)

PROCESSOR: Design Considerations

1. Main Information Sources

NWP ensemble forecasts, analysis, re-analysis

- Course grid (20 – 100 km)

Observation-based analysis

- Variables relevant to users, at high resolution grid:

  Real Time Meso-scale Analysis (RTMA)
  Radar & Gage precipitation analysis (Stage IV & CPC data)

2. Two-Stage Processing

Calibrate (de-bias) ensemble forecast on NWP grid (cheap to run)

- Joint sample (forecast, analysis)
- Climatic sample (re-analysis) → standard for calibration

Downscale the calibrated ensemble onto high resolution grid

- Joint sample (NWP re-analysis, Observation-based analysis)
- Perfect prog method (no forecast sample needed)

BAYESIAN PROCESSOR OF ENSEMBLE (BPE)

"Among the various approaches
that have been proposed recently,
BPE holds considerable promise,
in that it is the only
genuinely Bayesian technique."

NSF Reviewer, 2007

BPE — Theory

w – predictand

$\mathbf{y}$ – ensemble forecast, $\mathbf{y} = (y_1, \ldots, y_J)$

$\mathbf{x}$ – sufficient statistics, $\mathbf{x} = (x_1, \ldots, x_I)$, $\mathbf{x} = T(\mathbf{y})$, $I < J$

$g(w)$ prior (climatic uncertainty) ⟵ climatic sample (long)

$f(\mathbf{x}|w)$ likelihood (stochastic dependence) ⟵ joint sample (short)

expected $$\kappa(\mathbf{x}) = \int_{-\infty}^{\infty} f(\mathbf{x}|w)\, g(w)\, dw$$

posterior $$\phi(w|\mathbf{x}) = \frac{f(\mathbf{x}|w)}{\kappa(\mathbf{x})}\, g(w)$$

Concepts

- Fusion of information (two sources, asymmetric samples)
- Revision of distribution (given ensemble)
- Calibration (against climatic prior — standard)

BPE — Outputs

Input: Model ensemble (for predictand, lead time, grid point)

Output: (1) Posterior density function

(2) Posterior distribution function

(3) Posterior ensemble (calibrated ensemble)

Each member is mapped into a posterior quantile
via the inverse of the posterior distribution function

(4) Probability of non-exceedance for each member

Usage

- Given (2), any quantile can be calculated (0.1, 0.5, 0.9)
- Given (3) and (4), the user can construct
  a discrete approximation to the posterior distribution

BPE — Probabilistic Forecast

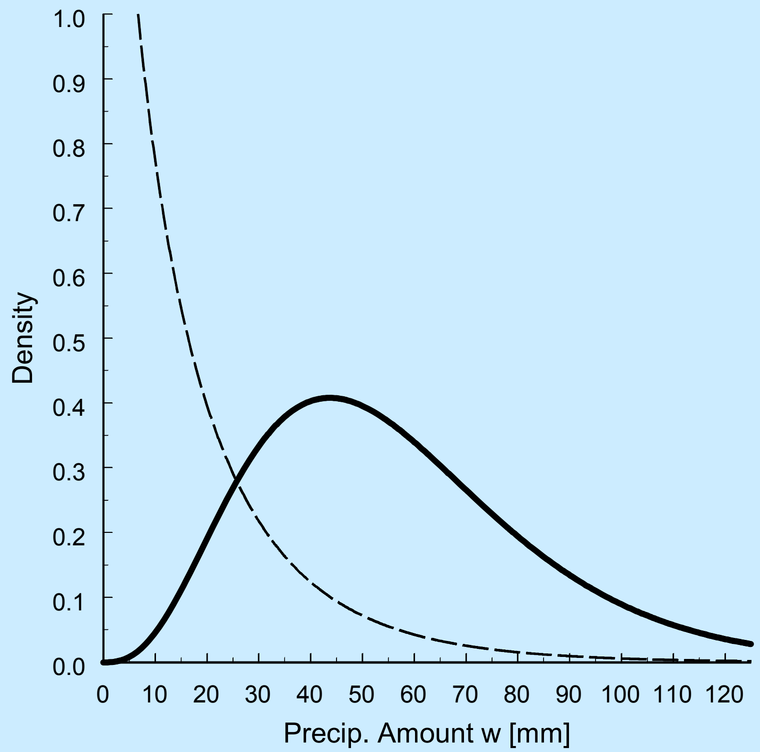
Prior d.f. $g(w)$

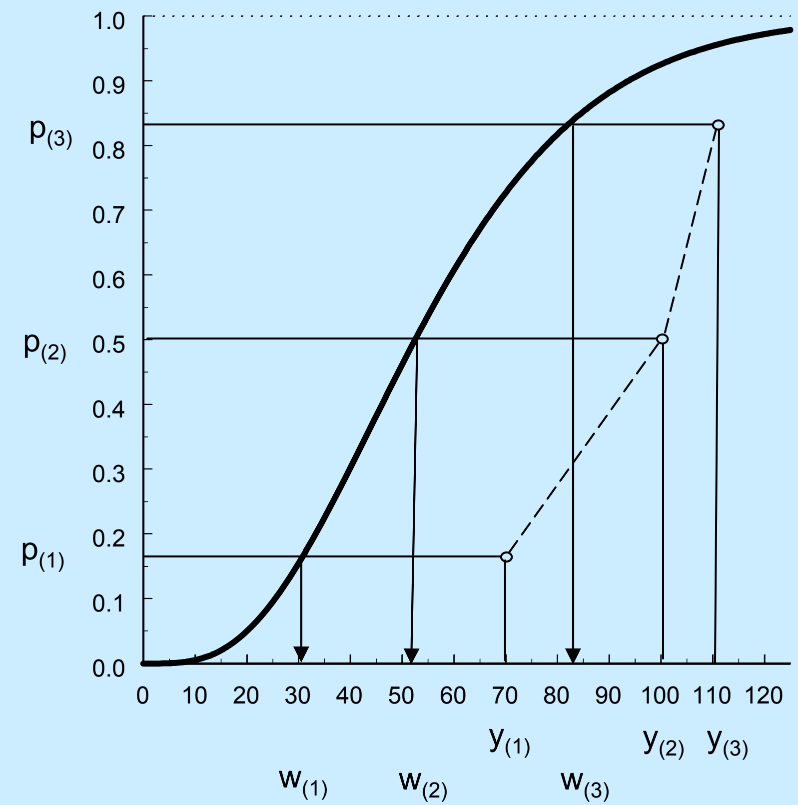
Posterior d.f. $\phi(w|\mathbf{x})$

Posterior D.F. $\Phi(w|\mathbf{x})$

Model ensemble

Posterior ensemble





BPE — Calibration: Some Results

Predictand: Daily max temperature., Savannah, GA

Sample sizes — Prior: 600 per day (120 years x 5 days)

— Joint: Estimation 110 (Cool), 60 (Warm)

Validation 220 (Cool and Warm)

Element	Distribution	Lead Time [days] 1	4	7
Median	Climatic	.44	.42	.43
	BPE	.47	.42	.39
50% Credible Interval	Climatic	.49	.49	.49
	BPE	.52	.56	.56
Quantiles:	Average Difference	.03	.01	.04

BPE — Basic Properties

Theoretically-based optimal fusion of ensemble forecast with climatic data

Revises prior (climatic) distribution given ensemble forecast based on comparison of past forecasts with observations

1. CORRECT THEORETIC STRUCTURE
   - Always valid
   - Modular: Framework for different – modeling assumptions
     – estimation procedures

2. FLEXIBLE ANALYTIC MODELS
   - Handle distributions of any form (not only normal)
   - Handle non-linear, heteroscedastic dependence structures
   - Parametric (easy to estimate and manipulate)
   - Robust when joint sample is small (no need for voluminous re-forecasting)

3. UNIQUE PERFORMANCE ATTRIBUTES
   - Removes bias in distribution
   - Guarantees calibration of the adjusted ensemble
     - Stable calibration (against climatic distribution | antecedent, regime)
     - Stationary calibration (equally good for all lead times)
     - User-specific calibration (point-specific, time-specific)
     - When predictability vanishes:
       adjusted ensemble → climatic ensemble
   - Preserves spatial / temporal / cross-variate rank correlations in ensemble