BAYESIAN PROCESSOR OF ENSEMBLE (BPE)

By

Roman Krzysztofowicz
University of Virginia

Presented at the

3rd NCEP / NWS Ensemble User Workshop
Laurel, Maryland

31 October – 2 November 2006

Acknowledgments:

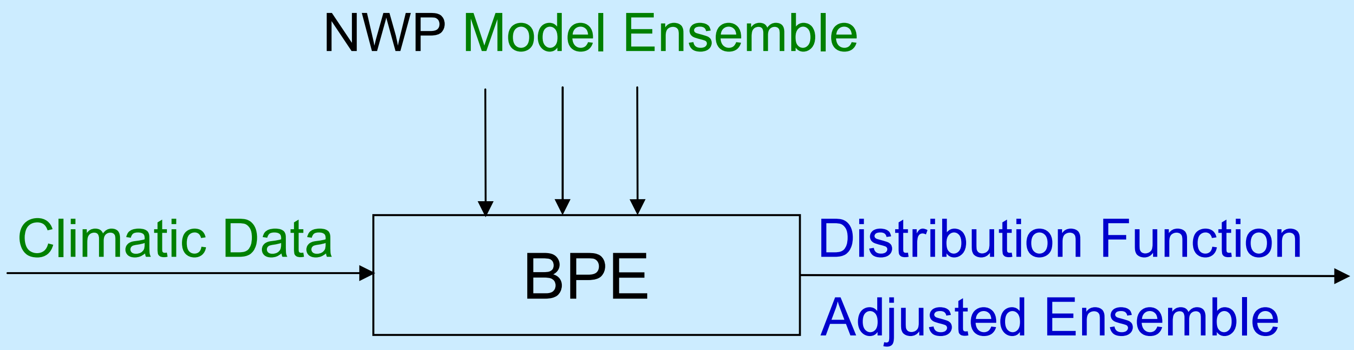
Work supported by the National Science Foundation under Grant No. ATM–0135940.

Collaboration of Zoltan Toth, Environmental Modeling Center NOAA / NWS / NCEP.

NEW STATISTICAL TECHNIQUES for PROBABILISTIC WEATHER FORECASTING

Techniques

Bayesian Processor of Output (BPO)
Bayesian Processor of Ensemble (BPE)



- extracts and fuses information
- quantifies total uncertainty
- calibrates (de-biases) ensemble

Versions for

binary predictands
multi-category predictands
continuous predictands

THEORY for CONTINUOUS PREDICTAND

Variates

W – predictand

$\mathbf{X}$ – vector of predictors, $\mathbf{X} = (X_1, \ldots, X_I)$

Bayesian Theory

$g(w)$ prior (climatic)

$f(\mathbf{x}|w)$ conditional (likelihood)

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

In real time, $\mathbf{x}$ is given; write $\phi(w)$

Fusion

- Two sources
- Asymmetric samples: climatic sample of W – long

  joint sample of $(\mathbf{X}, W)$ – short

FORECASTING EQUATIONS

- Posterior Distribution Function

$$\Phi(w) = Q\left(\frac{1}{T}\left[Q^{-1}(G(w)) - \sum_{i=1}^{I} c_i Q^{-1}(\bar{K}_i(x_i)) - c_0\right]\right)$$

- Posterior Density Function

$$\phi(w) = \frac{1}{T}\frac{\exp\left(-\frac{1}{2}[Q^{-1}(\Phi(w))]^2\right)}{\exp\left(-\frac{1}{2}[Q^{-1}(G(w))]^2\right)} g(w)$$

- Posterior Quantile

$$w_p = G^{-1}\left(Q\left(\sum_{i=1}^{I} c_i Q^{-1}(\bar{K}_i(x_i)) + c_0 + TQ^{-1}(p)\right)\right)$$

$$0 < p < 1$$

$$p = 0.1,\ 0.25,\ 0.5,\ 0.75,\ 0.9$$

EXAMPLE: Three Predictors

Quillayute, WA; cool season

W — 24-H PRECIP. AMOUNT, 12–36 h after 0000 UTC
X_1 — 24H TOTAL PRECIP. ending 36 h
X_2 — 850 REL. VORTICITY at 24 h
X_3 — 700 VERTICAL VELOCITY at 12 h

- Sample Sizes

Prior: 818 Joint: 470

- Distribution Functions

G	is Weibull:	$\alpha = 0.592$,	$\beta = 0.880$	
$\bar{K}_1$	is Weibull:	$\alpha_1 = 9.603$,	$\beta_1 = 0.910$	
$\bar{K}_2$	is Log-logistic:	$\alpha_2 = 6.212$,	$\beta_2 = 4.863$,	$\eta_2 = -5.0$
$\bar{K}_3$	is Log-logistic (–):	$\alpha_3 = 0.539$,	$\beta_3 = 4.313$,	$\eta_3 = -0.4$

- Posterior Parameters

$c_1 = 0.505$ $\quad c_0 = -0.025$
$c_2 = 0.241$ $\quad T = 0.641$
$c_3 = -0.275$

- Informativeness Score, IS

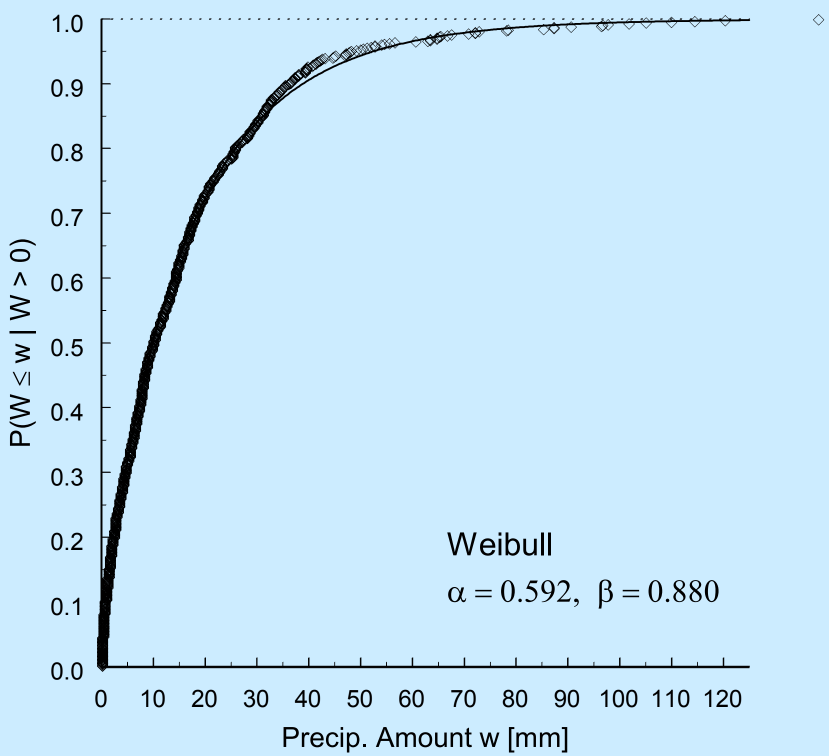
X_1	X_2	X_3	(X_1, X_2)	(X_1, X_3)	(X_1, X_2, X_3)
0.63	0.43	0.48	0.73	0.73	0.77

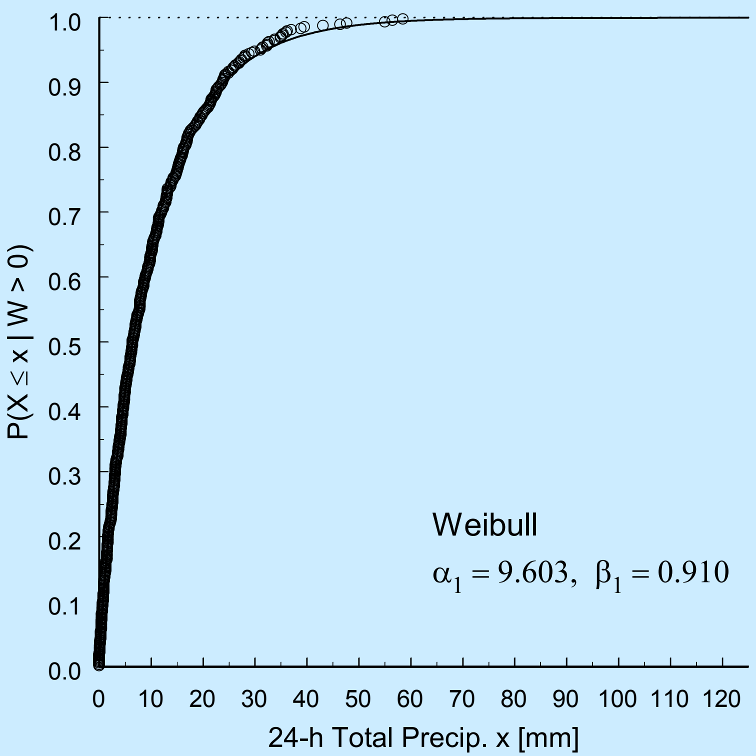
Distribution Functions

KUIL 12-36h Cond. Precip. Amount Cool

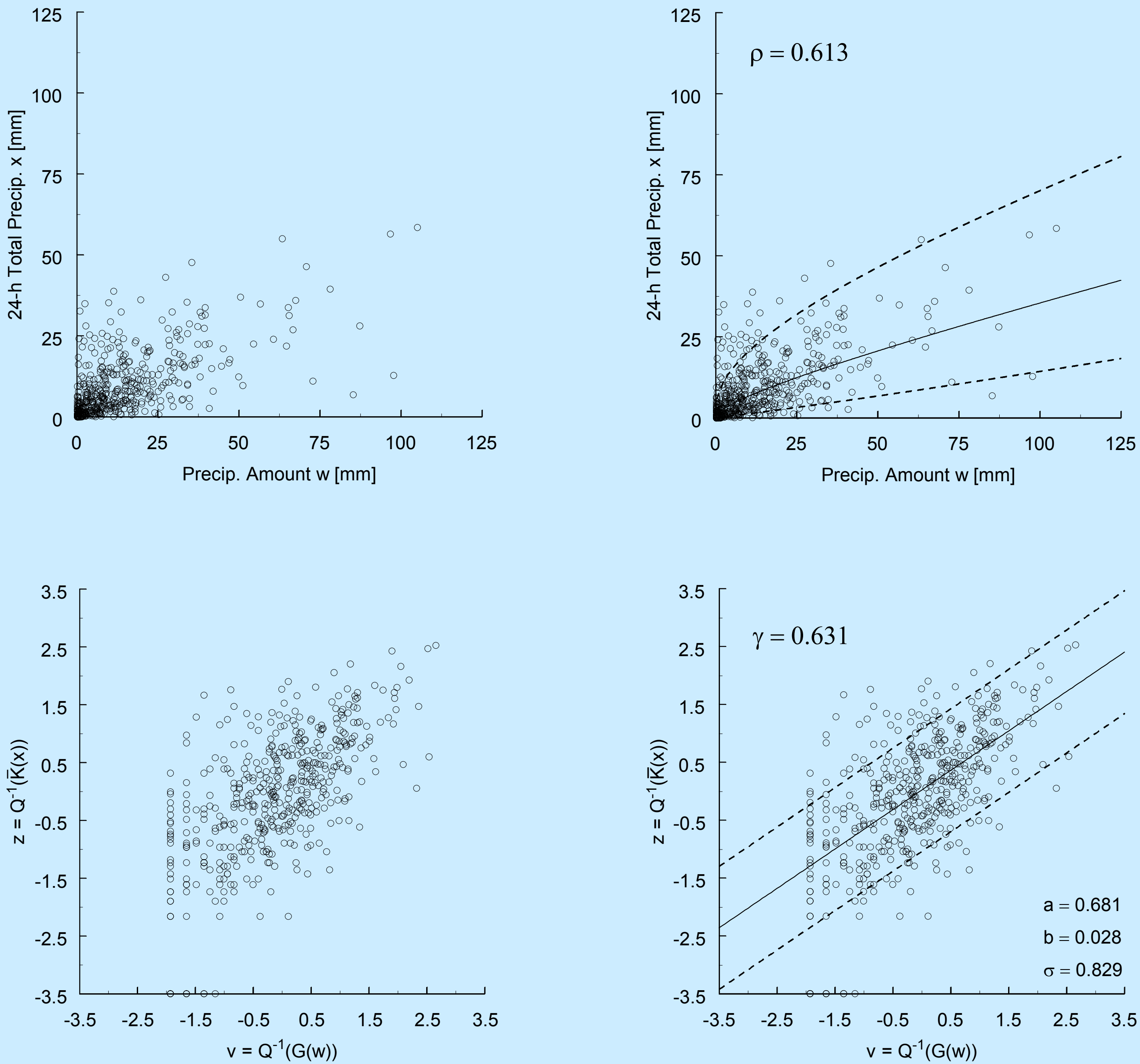
PRIOR (CLIMATIC) OF *W* N = 818

MARGINAL OF *X* N = 470





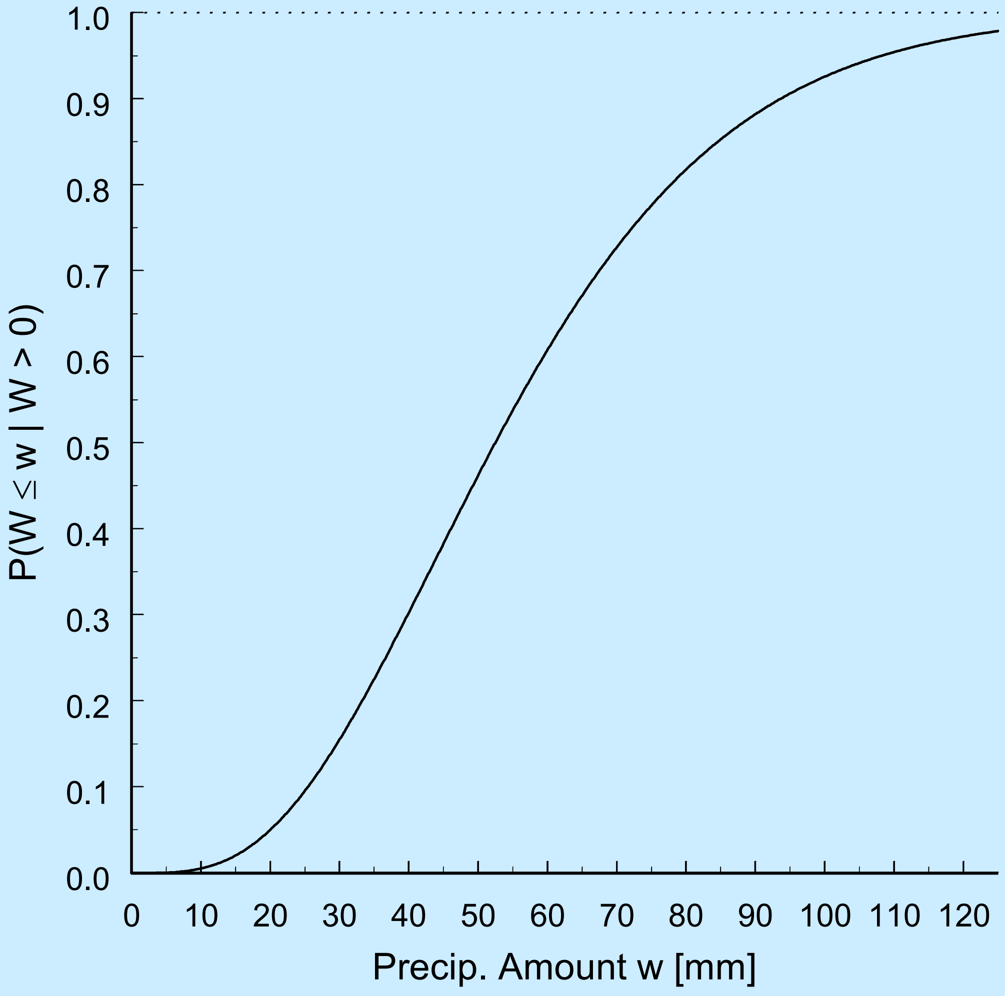
Likelihood Dependence Structure



KUIL 12-36h Cond. Precip. Amount

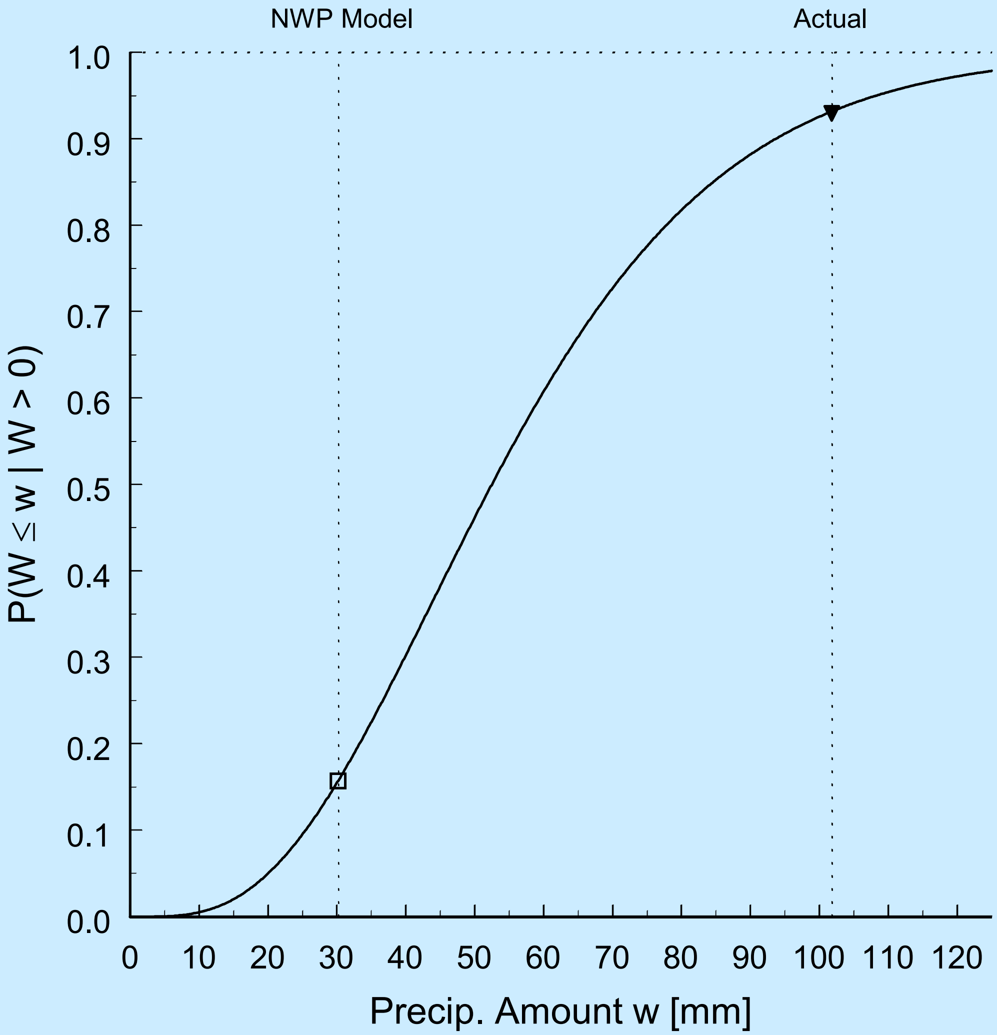
Cool

BPO forecast

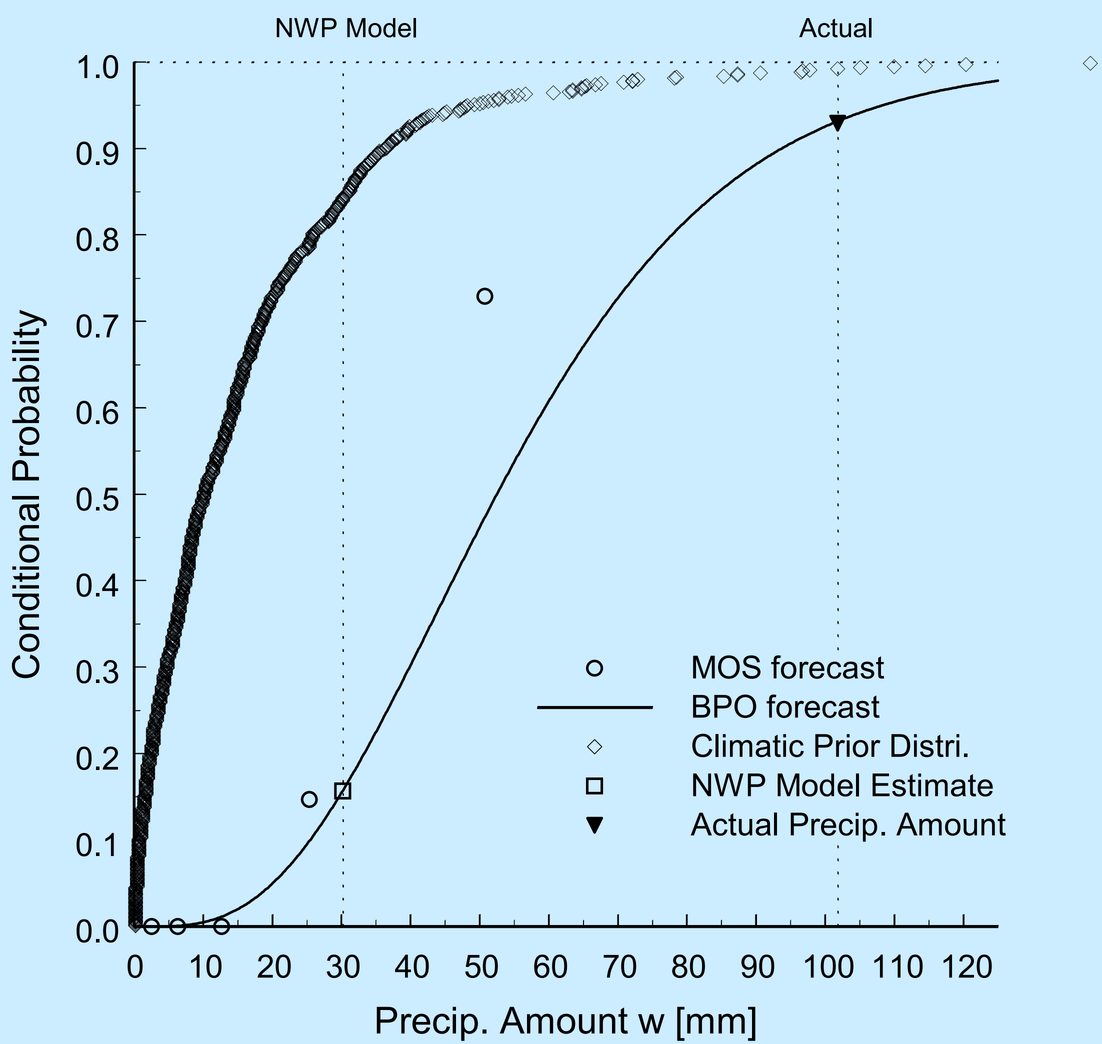


KUIL 12-36h Cond. Precip. Amount Cool

BPO forecast

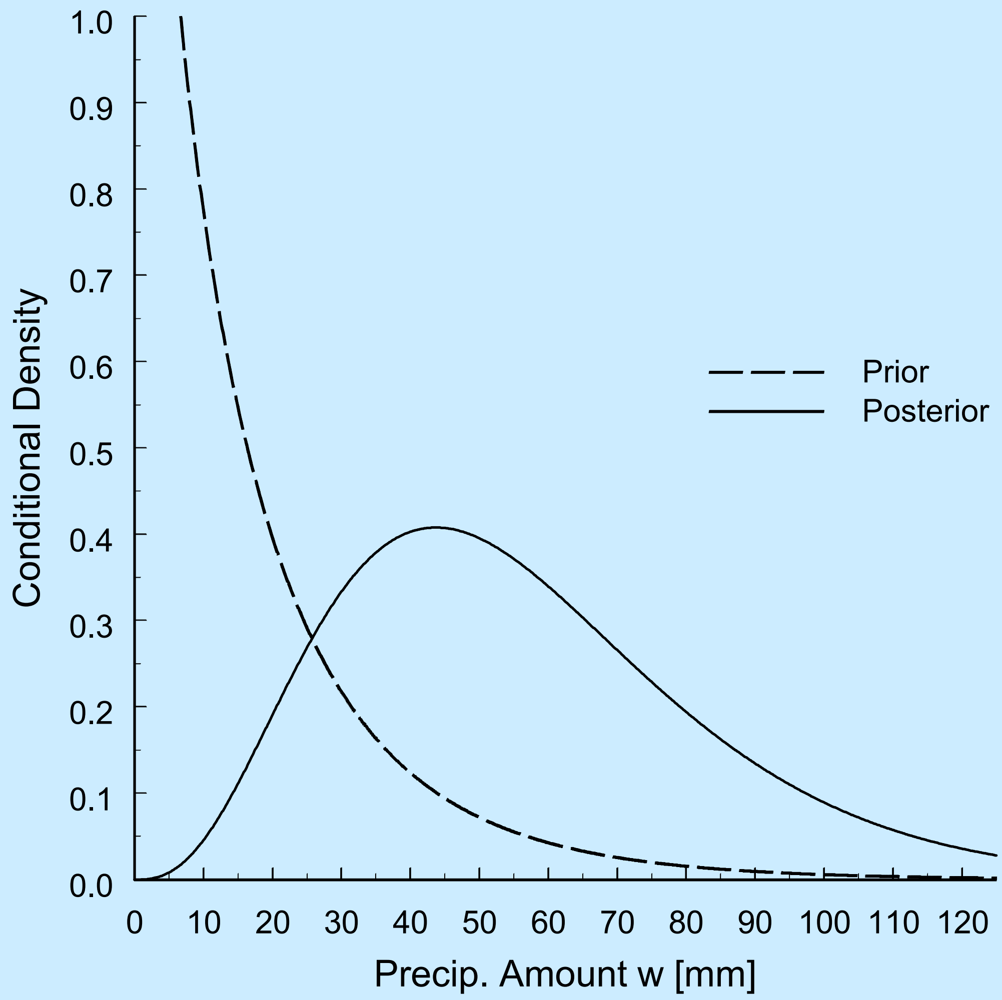


KUIL 12-36h Cond. Precip. Amount Cool



KUIL 12-36h Cond. Precip. Amount Cool

BPO forecast



BPE — Outputs

Input: Ensemble forecast of a predictand

Output: (1) Posterior distribution function (continuous cdf)

(2) Posterior density function (continuous pdf)

(3) Adjusted 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

This probability is identical for all predictands

USAGE

- Given (3) and (4), the user can construct
  a discrete approximation to the posterior cdf
- Given (1), any quantile can be calculated (10, 50, 90 for NDGD)

ENSEMBLE PROCESSING: Challenges & Solutions

W – predictand

$\mathbf{Y}$ – ensemble (vector of estimators) $\mathbf{Y} = (Y_0, Y_1, \ldots, Y_J)$

1. Samples are asymmetric
   - Climatic sample of W – long

     NCEP / NCAR re-analysis ~50 years, 2.5 x 2.5 grid
   - Joint sample of $(\mathbf{Y}, W)$ – short

     Recent ensemble forecasts and observations ~ 90 days

   * BPE: prior distribution, likelihood function → Bayesian fusion

2. Time series of $(\mathbf{Y}, W)$ are non-stationary (seasonality)

   * "Standardize" using climatic statistics for the day

     → Stationary

     → Ergodic

   * "Homogenize" across variates using ensemble statistics

ENSEMBLE PROCESSING: Challenges & Solutions

3. Ensemble members are:
   - not independent

$$0.43 < Rank\ Cor(Y_i', Y_j') < 0.82 \qquad i \neq j, \qquad j = 0, 1, \ldots, 10$$

   - not conditionally independent

$$f(y_i', y_j' | w) \neq f_i(y_i' | w) f_j(y_j' | w)$$

* BPE: models dependence (meta-Gaussian likelihood function)

4. Ensemble members have
   - very different informativeness

$$0.28 < IS_j < 0.59 \qquad j = 0, 1, \ldots, 10$$

   - diminishing marginal informativeness

$$IS(Y_0) = 0.586, \quad IS(Y_0, Y_2) = 0.605, \quad IS(Y_0, Y_2, Y_1) = 0.614$$

* BPE: Sufficient Statistic: $\mathbf{X} = T(\mathbf{Y}')$

ensemble	$\mathbf{Y}'$	dimension	22	(NCEP, 2008)
statistic	$\mathbf{X}$	dimension	2–5	

- $\mathbf{X}$ is as informative as $\mathbf{Y}'$
- $f(\mathbf{y}' | w)$ is replaced with $f(\mathbf{x} | w)$

ENSEMBLE PROCESSING: Challenges & Solutions

5. Distributions of W and X_i have many forms (non-Gaussian)

   * BPE: allows any form of the distribution (meta-Gaussian model)

   - Parametric distribution of each variate (2–3 parameters)
   - Library of 43 parametric distributions
   - Automatic estimation and selection

6. Dependence structure between $\mathbf{X}$ and W is

   - non-linear (in mean)
   - heteroscedastic (in variance)

   * BPE: models structure (meta-Gaussian likelihood function)

   - Normal Quantile Transform (NQT)
   - Each variate transformed into a standard normal
   - Multiple linear regression

BPE — Basic Properties

Theoretically-based optimal fusion of ensemble forecast with climatic data

Updates prior (climatic) distribution with 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 regime
   - Parametric (easy to estimate and manipulate)
   - Robust when joint sample is small

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