BAYESIAN PROCESSOR OF ENSEMBLE (BPE): PRIOR DISTRIBUTION FUNCTION

Parametric Models and Estimation Procedures Tested on Temperature Data

By
Roman Krzysztofowicz and Nah Youn Lee
University of Virginia

4 December 2008

Acknowledgments: Work supported by NSF, Grant No. ATM-0641572.
Data provided by Environmental Modeling Center, NCEP/NWS/NOAA.
DATA

Predictand: 2m temperature at 12 UTC
Forecast time: 00 UTC

Location: Savannah, GA
• Actual Location: 32° 8'N / 81° 13'W
• Climatic Data: 32.5° N / 80° W

Data
• Climatic Data: 1 January 1959 – 31 December 1998 (40 years)
  29 February (leap year) excluded
  5 March 1997 filled in
PRIOR DISTRIBUTION FUNCTION

\( k \) – index of the day (\( k = 1, \ldots, 365 \))
\( l \) – lead time in days (\( l = 1, 2, \ldots, 16 \))
\( W_k \) – predictand (variate)
\( w_k \) – realization of \( W_k \)

Prior Marginal D. F.

\[ G_k(w_k) = P(W_k \leq w_k) \]

Prior Markov D. F.

\[ H_{kl}(w_k | w_{k-l}) = P(W_k \leq w_k | W_{k-l} = w_{k-l}) \]

Prior D. F. \( \equiv \) Climatic Probabilistic Forecast

- Reference for calibration of ensemble
- Limit to which calibrated ensemble converges as \( LT \to \infty \)

Challenges

- Time series \( W_l, \ldots, W_{365} \) is nonstationary
- Distributions \( G_k \) and \( H_{kl} \) are non-Gaussian
STANDARDIZATION

Purpose: obtain time series that has stationary mean, variance, marginal D.F. (possibly)

Climatic sample for each day $k$ ($k = 1, \ldots, 365$)
15-day sampling window centered on day $k$

$$\{w_k(n) : n = 1, \ldots, M\} \quad M = 15 \text{ days} \times 40 \text{ years} = 600$$

Sample estimates
$ m_k$ – prior (climatic) mean
$ s_k$ – prior (climatic) standard deviation

Standardized climatic sample

$$w'_k(n) = \frac{w_k(n) - m_k}{s_k} \quad n = 1, \ldots, M$$

$$\{w'_k(n) : n = 1, \ldots, M\}$$
Sample deciles in original space (15-day window), SAV

Temperature (°K)

Day

Maximum

90
80
70
60
50
40
30
20
10
Minimum
Sample mean (15-day), Fourier series (2nd order), SAV
Sample std. dev. (15-day), Fourier series (2nd order), SAV
Sample deciles in standard space (15-day window), SAV
Two alternative hypotheses (tested empirically)

1. Nonstationary Prior (estimate 12 parametric D.F.)
   \[48 = 4 \times 12 \text{ parameters} + \text{interpolation} \implies 365 \text{ D.F.}\]
   MAD across 12 samples (1 Jan, ..., 1 Dec):
   0.012 (min), 0.024 (average), 0.039 (max)

2. Stationary Prior (estimate 1 parametric D.F.)
   4 parameters \implies 365 \text{ D.F.}
   MAD across 12 samples (1 Jan, ..., 1 Dec):
   0.038 (min), 0.053 (average), 0.065 (max)

MAD = \text{max} | \text{Empirical D. F.} - \text{Parametric D. F.} |
Prior Marginal D. F.: Analyses

0. Given: standardized climatic sample for day $k$
   \[ \{w'_k(n) : n = 1, \ldots, M\} \quad M = 600 \]
   12 days $k$ (1 Jan, ..., 1Dec)

1. Hypothesize several parametric models for $G'_k$
2. Estimate parameters of each model
3. Choose model that minimizes average (across $k$):
   \[ \text{MAD} = \max | \text{Empirical D. F.} - \text{Parametric D. F.} | \]

Model: Log-Reciprocal Type I – Log-Weibull

\[ G'(w') = 1 - \exp\left[ -\left( \frac{1}{\alpha} \ln\left( \ln \frac{\eta_U - \eta_L}{\eta_U - w'} + 1 \right) \right)^{\beta} \right] \]

Parameters: $\alpha, \beta, \eta_L, \eta_U$
Empirical and Parametric Distribution Functions on 1 Jan ($k = 1$) ($M = 600$), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0277
$\alpha = 0.7971$
$\beta = 3.5186$
$\eta_L = -4.0396$
$\eta_U = 2.3533$
Empirical and Parametric Distribution Functions on 1 Feb ($k = 32$) ($M = 600$), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0201
$\alpha = 0.7467$
$\beta = 3.0885$
$\eta_L = -3.4939$
$\eta_U = 2.3624$
Empirical and Parametric Distribution Functions on 1 Mar ($k = 60$) ($M = 600$), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0236
$\alpha = 0.8139$
$\beta = 3.0467$
$\eta_L = -3.6945$
$\eta_U = 2.1512$
Empirical and Parametric Distribution Functions on 1 Apr ($k = 91$) ($M = 600$), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0234
$\alpha = 0.7974$
$\beta = 3.3631$
$\eta_L = -3.7663$
$\eta_U = 2.2657$
Empirical and Parametric Distribution Functions on 1 May ($k = 121$) ($M = 600$), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0387
$\alpha = 0.7703$
$\beta = 3.3851$
$\eta_L = -3.4261$
$\eta_U = 2.2486$
Empirical and Parametric Distribution Functions on 1 Jun ($k = 152$) ($M = 600$), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0199
$\alpha = 0.9259$
$\beta = 5.2807$
$\eta_L = -5.7415$
$\eta_U = 2.1794$
Empirical and Parametric Distribution Functions on 1 Jul \((k = 182) (M = 600)\), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0289
\(\alpha = 0.7362\)
\(\beta = 4.2803\)
\(\eta_L = -4.5610\)
\(\eta_U = 2.9318\)
Empirical and Parametric Distribution Functions on 1 Aug ($k = 213$) ($M = 600$), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0185
$\alpha = 0.6738$
$\beta = 4.2969$
$\eta_L = -4.1051$
$\eta_U = 3.2253$
Empirical and Parametric Distribution Functions on 1 Sep \((k = 244) \ (M = 600), \ SAV\)

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0115
\(\alpha = 0.8015\)
\(\beta = 4.7956\)
\(\eta_L = -5.0532\)
\(\eta_U = 2.7160\)
Empirical and Parametric Distribution Functions on 1 Oct \((k = 274) (M = 600)\), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0249
\(\alpha = 0.8507\)
\(\beta = 3.7870\)
\(\eta_L = -4.0384\)
\(\eta_U = 2.1044\)
Empirical and Parametric Distribution Functions on 1 Nov ($k = 305$) ($M = 600$), SAV

- Log-Reciprocal Type I
- Log-Weibull
- MAD: 0.0242
- $\alpha = 0.8238$
- $\beta = 3.2029$
- $\eta_L = -3.8365$
- $\eta_U = 2.1703$
Empirical and Parametric Distribution Functions on 1 Dec (k = 335) (M = 600), SAV

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0262
α = 0.7528
β = 2.8569
η_L = -3.3537
η_U = 2.2668
Prior Marginal D. F.: Conclusions

- Small diversity of empirical D. F.
- Large similarity of parametric D. F.
- Near-stationarity of parameters $(\alpha, \beta, \eta_L, \eta_U)$
Empirical Distribution Functions: the two most diverse on 1 Jun ($k = 152$) and 1 Dec ($k = 335$) ($M = 600$), SAV
Parametric Distribution Functions on First Day of Each Month \((M = 600)\), SAV

Log-Reciprocal Type I – Log-Weibull Distribution
Time-series distribution parameters: First of Each Month ($M = 600$), SAV

Log-Reciprocal Type I – Log-Weibull Distribution
Stationary Prior Marginal D.F. in Standard Space

1. Form a *pooled* standardized climatic sample from several days

2. Determine *bounds* from the grand sample (365 days, 40 years)
   \[
   \eta_L = \min_{k,n} \{ w'_k(n) \} - 0.5 \\
   \eta_U = \max_{k,n} \{ w'_k(n) \} + 0.5
   \]

3. Estimate stationary D.F. from the pooled sample
   \[
   G'(w'), \quad \eta_L < w' < \eta_U
   \]
   stationary parameters: \( \alpha, \beta, \eta_L, \eta_U \)
Stationary Prior Marginal Distribution Functions: Empirical and Parametric, SAV

Pooled Sample from 4 Days: 1 Jan, 1 Apr, 1 Jul, 1 Oct ($M = 2400$)

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0139
$\alpha = 0.7659$
$\beta = 5.4736$
$\eta_L = -5.8833$
$\eta_U = 3.3777$
Prior Marginal D.F. for day $k$ in Original Space

0. Given: prior mean and standard deviation $m_k, s_k$

stationary parameters $\alpha, \beta, \eta_L, \eta_U$

1. Construct

$$G_k(w) = G\left(\frac{w - m_k}{s_k}\right)$$

$$\eta_{Lk} < w < \eta_{Uk}$$

$$\eta_{Lk} = \eta_L s_k + m_k$$

$$\eta_{Uk} = \eta_U s_k + m_k$$
Empirical Distribution Function on 1 Jan \((k = 1) (M = 600)\), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0455
\(\alpha = 0.7659\)
\(\beta = 5.4736\)
\(\eta_{Lk} = 256.18\)
\(\eta_{Uk} = 302.89\)
Empirical Distribution Function on 1 Feb ($k = 32$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0464
$\alpha = 0.7659$
$\beta = 5.4736$
$\eta_{Lk} = 254.23$
$\eta_{Uk} = 301.90$
Empirical Distribution Function on 1 Mar \((k = 60)\) \((M = 600)\), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0588
\(\alpha = 0.7659\)
\(\beta = 5.4736\)
\(\eta_{Lk} = 258.07\)
\(\eta_{Uk} = 302.60\)
Empirical Distribution Function on 1 Apr ($k = 91$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0421
$\alpha = 0.7659$
$\beta = 5.4736$
$\eta_{Lk} = 269.31$
$\eta_{Uk} = 300.41$
Empirical Distribution Function on 1 May ($k = 121$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0604
$\alpha = 0.7659$
$\beta = 5.4736$
$\eta_{Lk} = 277.76$
$\eta_{Uk} = 300.74$
Empirical Distribution Function on 1 Jun ($k = 152$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0592
$\alpha = 0.7659$
$\beta = 5.4736$
$\eta_{Lk} = 286.01$
$\eta_{Uk} = 301.80$
Empirical Distribution Function on 1 Jul \((k = 182)\) \((M = 600)\), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0652
\(\alpha = 0.7659\)
\(\beta = 5.4736\)
\(\eta_{Lk} = 291.76\)
\(\eta_{Uk} = 301.61\)
Empirical Distribution Function on 1 Aug ($k = 213$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

- Log-Reciprocal Type I
- Log-Weibull
- MAD: 0.0583
- $\alpha = 0.7659$
- $\beta = 5.4736$
- $\eta_{Lk} = 294.14$
- $\eta_{Uk} = 301.30$
Empirical Distribution Function on 1 Sep \((k = 244) \ (M = 600)\), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0382
\(\alpha = 0.7659\)
\(\beta = 5.4736\)
\(\eta_{Lk} = 291.45\)
\(\eta_{Uk} = 301.57\)
Empirical Distribution Function on 1 Oct ($k = 274$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

Log-Reciprocal Type I
Log-Weibull
MAD: 0.0580
$\alpha = 0.7659$
$\beta = 5.4736$
$\eta_L = 280.95$
$\eta_U = 303.28$
Empirical Distribution Function on 1 Nov ($k = 305$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

- Log-Reciprocal Type I
- Log-Weibull
- MAD: 0.0432
- $\alpha = 0.7659$
- $\beta = 5.4736$
- $\eta_{L} = 270.65$
- $\eta_{U} = 302.83$
Empirical Distribution Function on 1 Dec ($k = 335$) ($M = 600$), SAV

Parametric Distribution Function from Pooled Sample from 4 Days

- Log-Reciprocal Type I
- Log-Weibull
- MAD: 0.0620
- $\alpha = 0.7659$
- $\beta = 5.4736$
- $\eta_{Lk} = 260.07$
- $\eta_{Uk} = 303.57$

Temperature $w$ (°K)

$P(W_k \leq w)$
Stationary Prior Marginal D.F.: Conclusions

- Sampling Window: 5 – 15 days (200 – 600 realizations)
  - 15-day => empirical D. F. smoother, tails better delineated

- Pooled Sample Size
  - $M = 2400$, 4 days, MAD = 0.014
  - $M = 3600$, 6 days, MAD = 0.014
  - $M = 7200$, 12 days, MAD = 0.011

- Families of Good Parametric D.F.
  - Bounded: Log-Reciprocal Type I – Log-Weibull (Weibull, Log-Logistic)
  - Bounded Above: Log-Weibull (Weibull, Log-Logistic)
CLIMATIC AUTOCORRELATION

\( W_k' \) – standardized predictand
\( G' \) – prior marginal D. F. (stationary)
\( V_k \) – normalized predictand
\( Q \) – standard normal D. F.

Normal Quantile Transform (NQT)
\[
V_k = Q^{-1}(G'(W_k'))
\]

Autocorrelation coefficient (Pearson’s product-moment)
\[
c_k = Cor(V_{k-1}, V_k) \quad k = 1, \ldots, 365
\]

*Note: Standardization does not make \( c_k \) stationary
Autocorrelation in normal space (15-day), Fourier series (8th order), SAV
Prior Markov D. F. for day $k$ in Original Space

$k$ – day for which forecast is made ($k = 1, \ldots, 365$)

$l$ – lead time in days ($l = 1, 2, \ldots, 14$)

$k - l$ – day of the last observation (forecast day)

0. Given: prior marginal D. F. $G_k, G_{k-l}$

   autocorrelation coefficient $c_k$

1. Construct

$$H_{kl}(w_k|w_{k-l}) = Q\left(\frac{Q^{-1}(G_k(w_k)) - c_k^l Q^{-1}(G_{k-l}(w_{k-l}))}{(1 - c_k^{2l})^{1/2}}\right)$$