HEPEX workshop
9th June 2011

Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions

James D. Brown[1,2], Dong-Jun Seo[3]

[1] NOAA/NWS/OHD, Silver Spring, MD, USA

[2] UCAR, Boulder, CO, USA

[3] University of Texas at Arlington, TX, USA

Problem definition

Difficulties:

- **Parametric assumptions don’t always apply**
- **Marginal transforms (NQT) often problematic**
- **Bias/skill are strongly amount dependent**
- **Type-II conditional bias (CB) often overlooked**

Motivation (for operational use):

1. **Broadly applicable w/ limited supervision**
2. **Capture amount-dependent bias and skill**
3. **Address Type-I CB (reliability) *and* Type-II CB**

Problem definition

The random variables:

X = observed (assumed unbiased)

$\boldsymbol{Y} = \{Y_1, \ldots, Y_m\}$ = vector of (ensemble) predictors (biased)

The required conditional distribution:

$$F(c_f \mid y_{1,\ldots,} y_m) = Pr[\, X \le c_f \mid Y_1 = y_1, \ldots, Y_m = y_m \,] \quad \forall c_f$$

Which can be written as:

$$F(c_f \mid y_{1,\ldots,} y_m) = E[\, I(x; c_f) \mid Y_1 = y_1, \ldots, Y_m = y_m \,] \quad \forall c_f$$

$$\text{where} \quad I(x; c_f) = \left\{ 1, x \le c_f ; 0, \text{otherwise} \right\}.$$

Proposed solution: indicator co-kriging (ICK)

Proposed solution

Brown & Seo (2010): JHM, 11(3), 642-665:

Probability of not exceeding a single threshold, c_f, given the predictors

$$E\left[I(x;c_f)\,|\,Y_1=y_1,\ldots,Y_m=y_m\right]\approx$$

$$E\left[I(x;c_f)\,|\,I(y_j;c_k)=i(y_j;c_k);\ j=1,\ldots,m;\ k=1,\ldots,v\right]\approx$$

$$\underbrace{E\left[I(x;c_f)\right]}_{\text{Climatology}}+\underbrace{\sum_{j=1}^{m}\sum_{k=1}^{v}\boxed{\lambda_{c_f}(y_j;c_k)}\times\left\{i(y_j;c_k)-E\left[I(y_j;c_k)\right]\right\}}_{\text{Conditional adjustment (expectation of 0)}}.$$

Weights to estimate

Climatology + Conditional adjustment (expectation of 0)

How to estimate weights?

Mixed objective function (papers in prep.)

= predicted

= observed

$$J = J_1 + J_2$$

$$= E\left[\left(\left\{ \boxed{ICK(c_f)} - \boxed{i(x;c_f)} \right\}^2 + E\left[\left\{ E\left[\boxed{ICK(c_f)} \mid I(x;c_f) = \boxed{i(x;c_f)} \right] - \boxed{i(x;c_f)} \right\}^2 \right] \right) \middle| \; I(y_j;c_k) = i(y_j;c_k);\; j = 1,\ldots,m;\; k = 1,\ldots,v \right]$$

$$= (\text{Brier score} + \text{Type-II CB}) \mid \text{predictors}.$$

This has a closed form solution.

Application of (CBP)-ICK to multi-model ensemble of hydrologic simulations (MOPEX basins)

Application to MOPEX data

Multi-model ensemble of simulated flow

- Daily flow; 7 models (no gr4j); 9 basins.
- No tuning for specific basins!
- Used one prior observation as aux. predictor

Cross-validation/verification

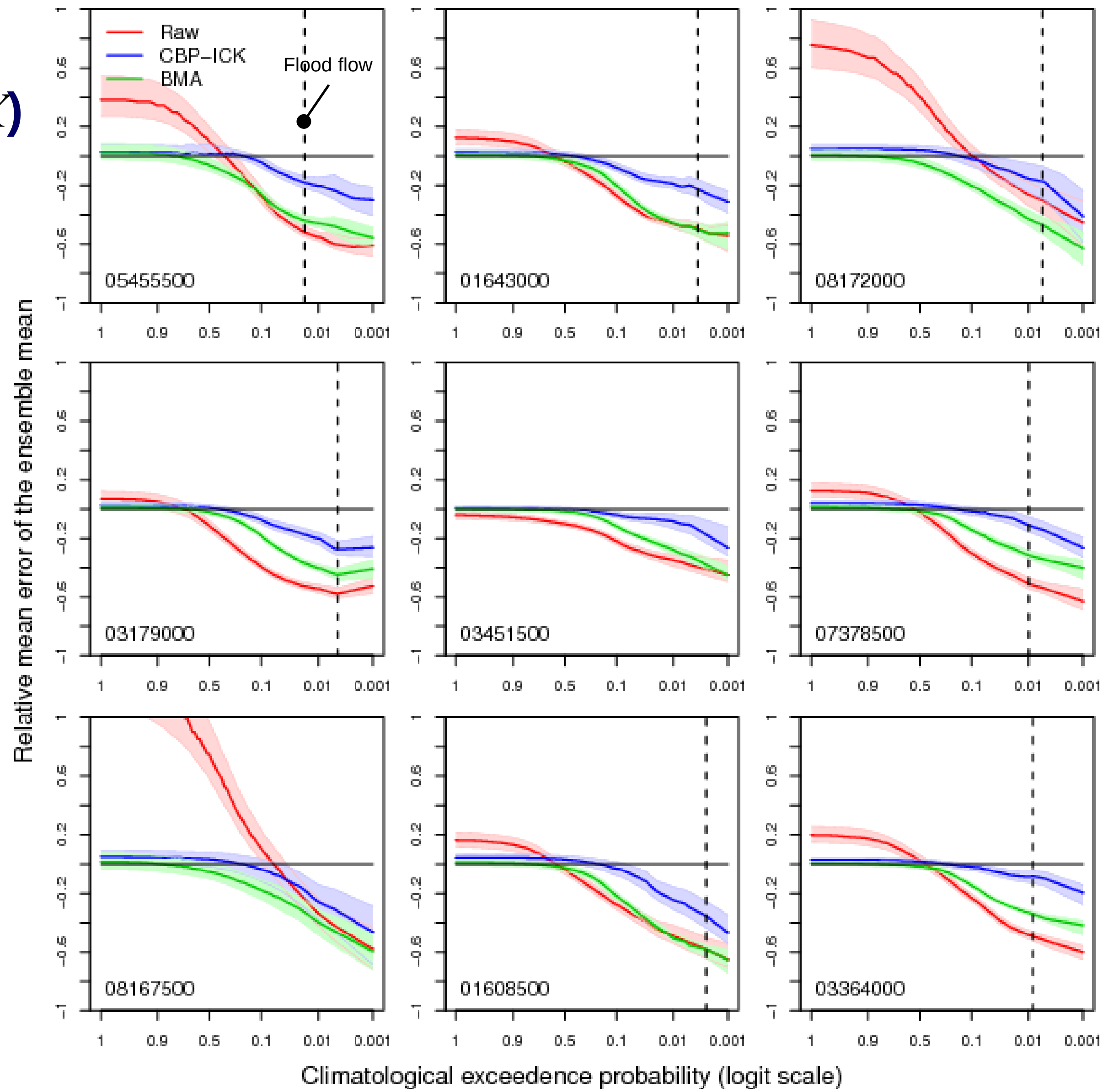
- "Leave-one-year-out" cross-validation
- Verified ensemble mean & probabilities (EVS)
- Skill: raw ensemble & Gaussian mixture BMA
- Block bootstrap to assess sampling unc.

Type-II CB of ensemble mean

Relative error of ens. mean (μ_Y) given obs. (X) > threshold (c_p), which has clim. prob, p:

$$RE = \frac{\sum_{x>c_p} \mu_Y - x}{\sum_{x>c_p} x}$$

- CB in raw ensemble: over-predict low flow, under-predict high.
- Much reduced by CBP-ICK (note that CB is only part of objective function).
- BMA fairs less well, except at low thresh.



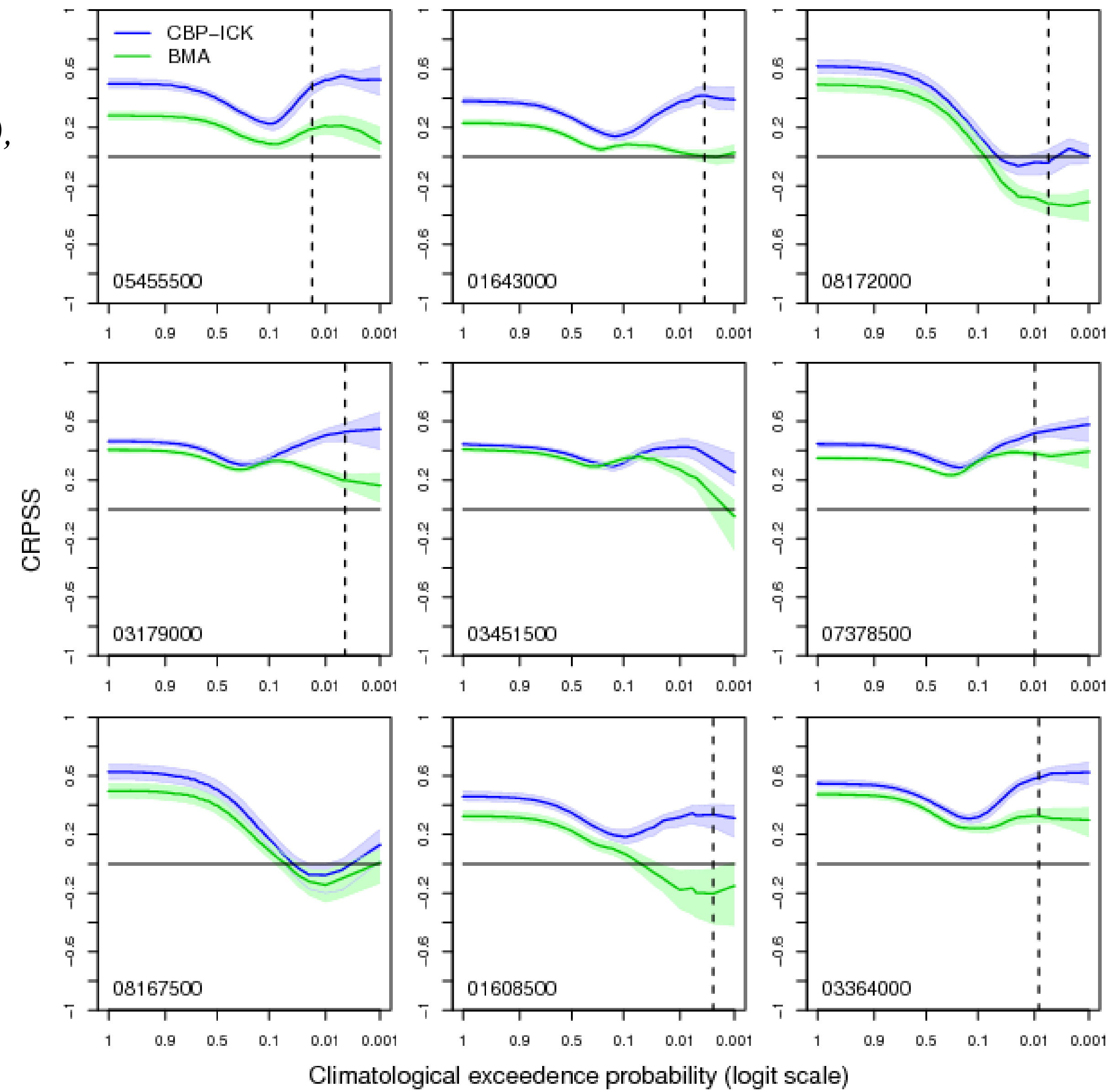
Conditional skill

Conditional CRPSS:

$$\overline{CRPSS} = 1 - (\overline{CRPS} \,/\, \overline{CRPS}_{raw}),$$

$$\text{where} \quad \overline{CRPS} = \frac{1}{|x > c_p|} \times \sum_{x > c_p} \int \left[F_Y(q) - I_X(q)\right]^2 dq.$$

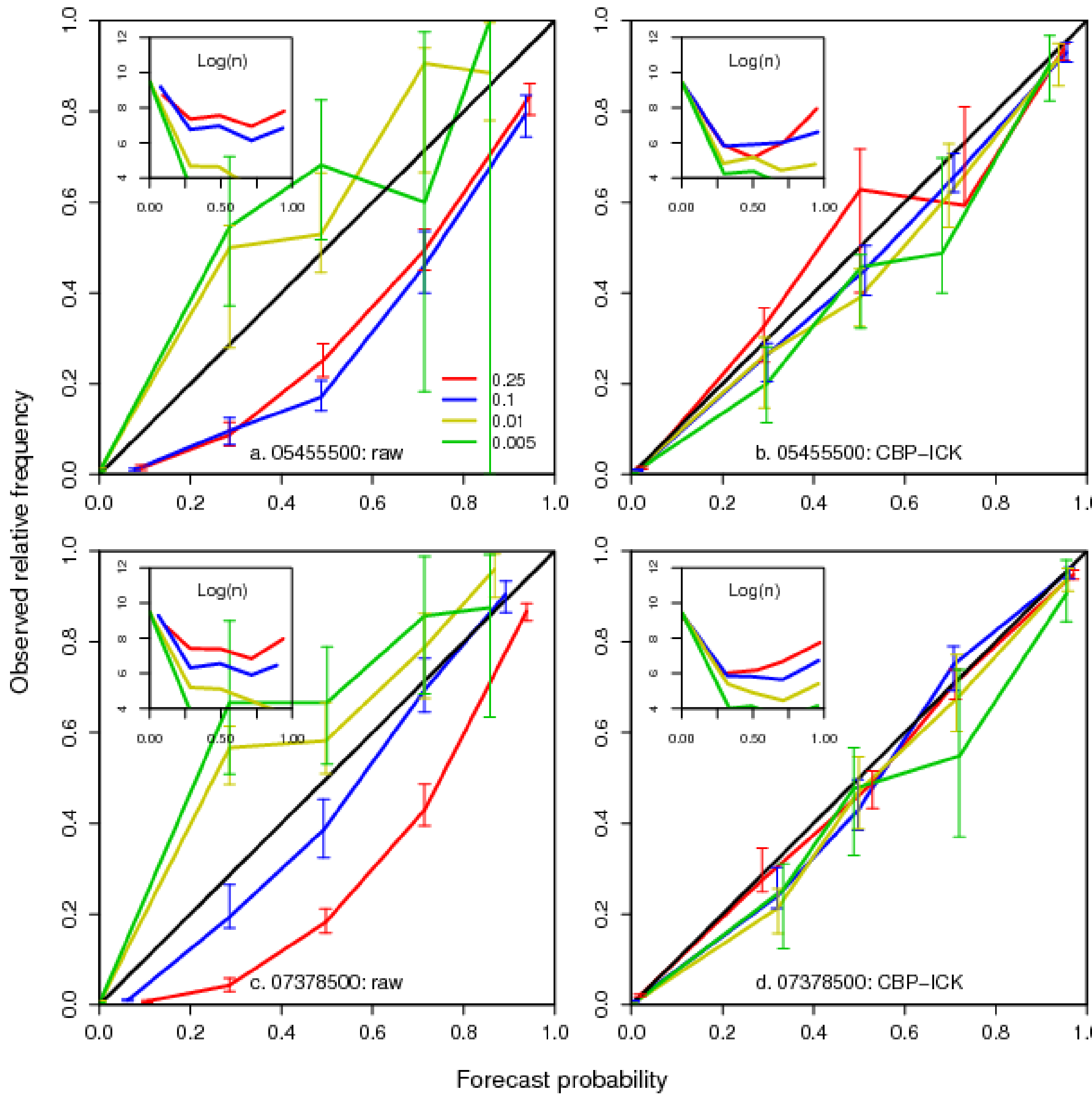
- 40-60% gain in skill for most locations and flow thresholds.
- Greatest gain where conditional bias was large and predictors remain powerful.
- 0817/0816 (in TX): convective precip.?



Event reliability (Type-I CB)

Type-I CB (reliability):

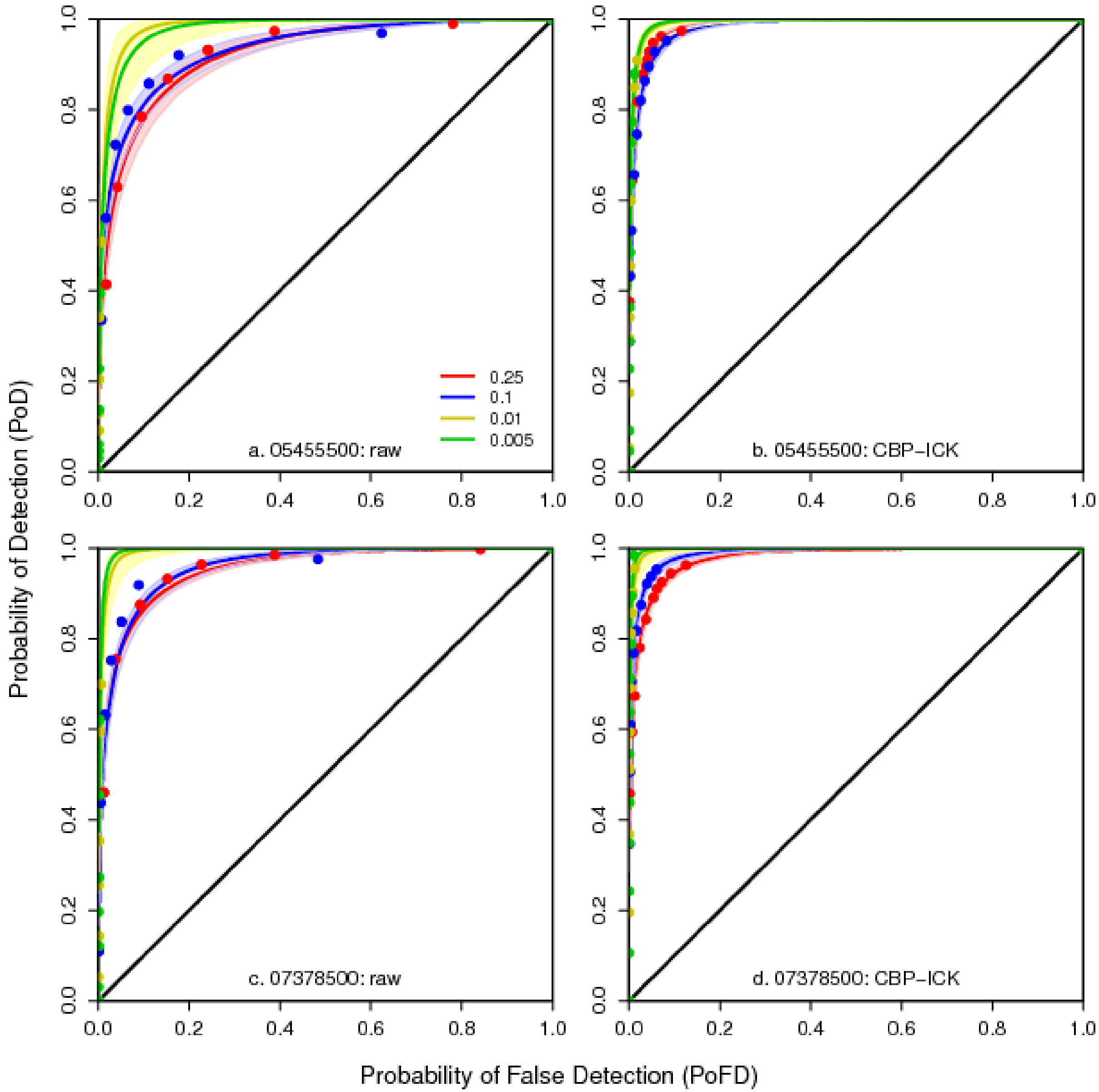
- Raw ensembles are reasonably reliable (models reasonably calibrated overall).
- Some improvement following bias-correction, both in reliability and sharpness.
- However, greatest gains in skill stem from reducing Type-II CB, not Type-I.



Event discrimination

ROC curves

- Empirical PoD/PoFD fitted with bivariate normal model.
- Raw ensembles do well at high thresholds.
- Significant gains in event discrimination at most thresholds following bias-correction.



Conclusions

- Goal of “maximizing sharpness subject to reliability” should be reframed.
- Type-II CB <u>can</u> be addressed when estimating post-p. statistical parameters.
- We propose a flexible non-parametric technique with Type-II CB minimization.
- Appropriate for multi-year datasets of single/multi-model ensembles.
- Need to evaluate for regulated flows.

Questions???

Reference papers:

Brown, J.D. and Seo, D-J. (2010) A nonparametric post-processor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *Journal of Hydrometeorology*, 11(3), .

Brown, J.D., Demargne, J., Seo, D-J., and Liu, Y. (2010) The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling and Software*, 25(7), 854-872.

Brown, J.D. and Seo, D-J (in preparation) Evaluation of a non-parametric post-processor for hydrologic uncertainty estimation and bias-correction, with application to a multi-model ensemble of simulated streamflows from test basins in the southeast U.S. TBD.

Extra slides

How to estimate weights?

Min. square error (Brown & Seo, 2010):

$$J = J_1$$

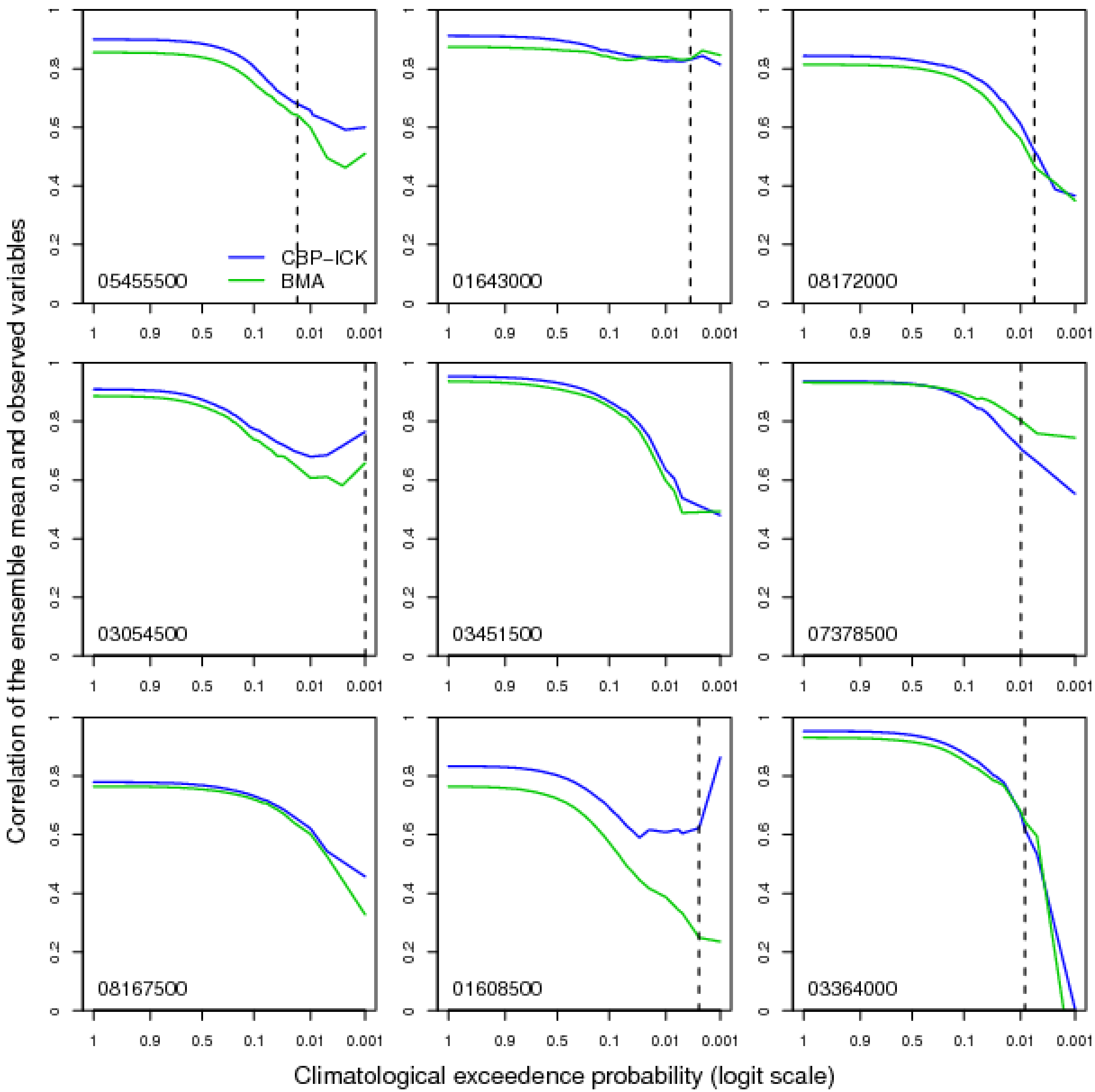
$$= E\left[\begin{array}{l}\left\{\underline{ICK(c_f)} - \underline{i(x;c_f)}\right\}^2 \mid \\ \underline{I(y_i;c_k) = i(y_i;c_k); i = 1,\ldots,m; k = 1,\ldots,v}\end{array}\right]$$

$$= (\textit{predicted} - \textit{observed})^2 \mid \textit{predictors}$$

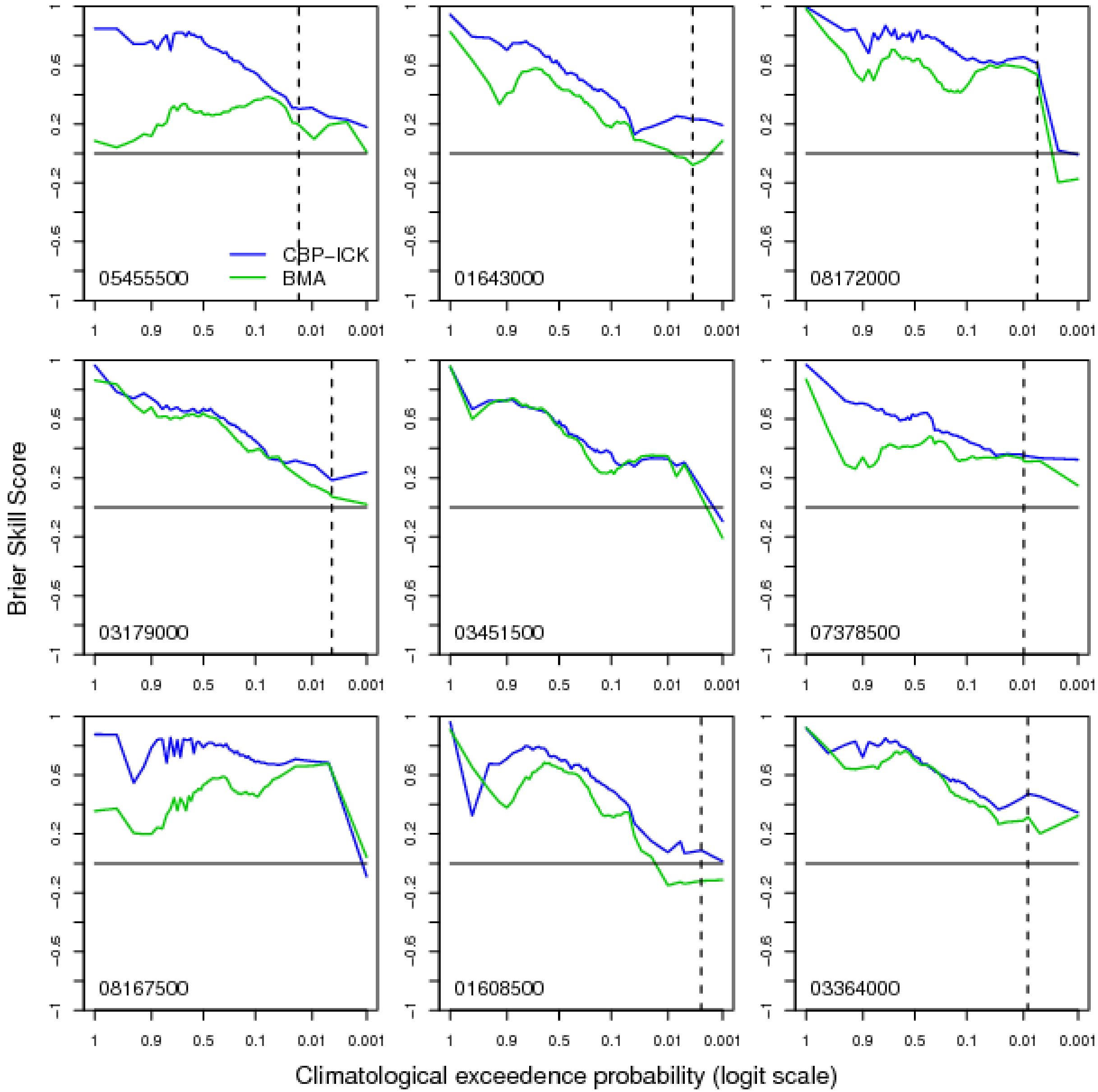
$$= \textit{conditional error variance or conditional "Brier Score"}.$$

Problem: ICK is unconditionally unbiased, but no mention of conditional bias in *J*. "Errors in variables" effect exaggerates conditional bias at high thresholds.

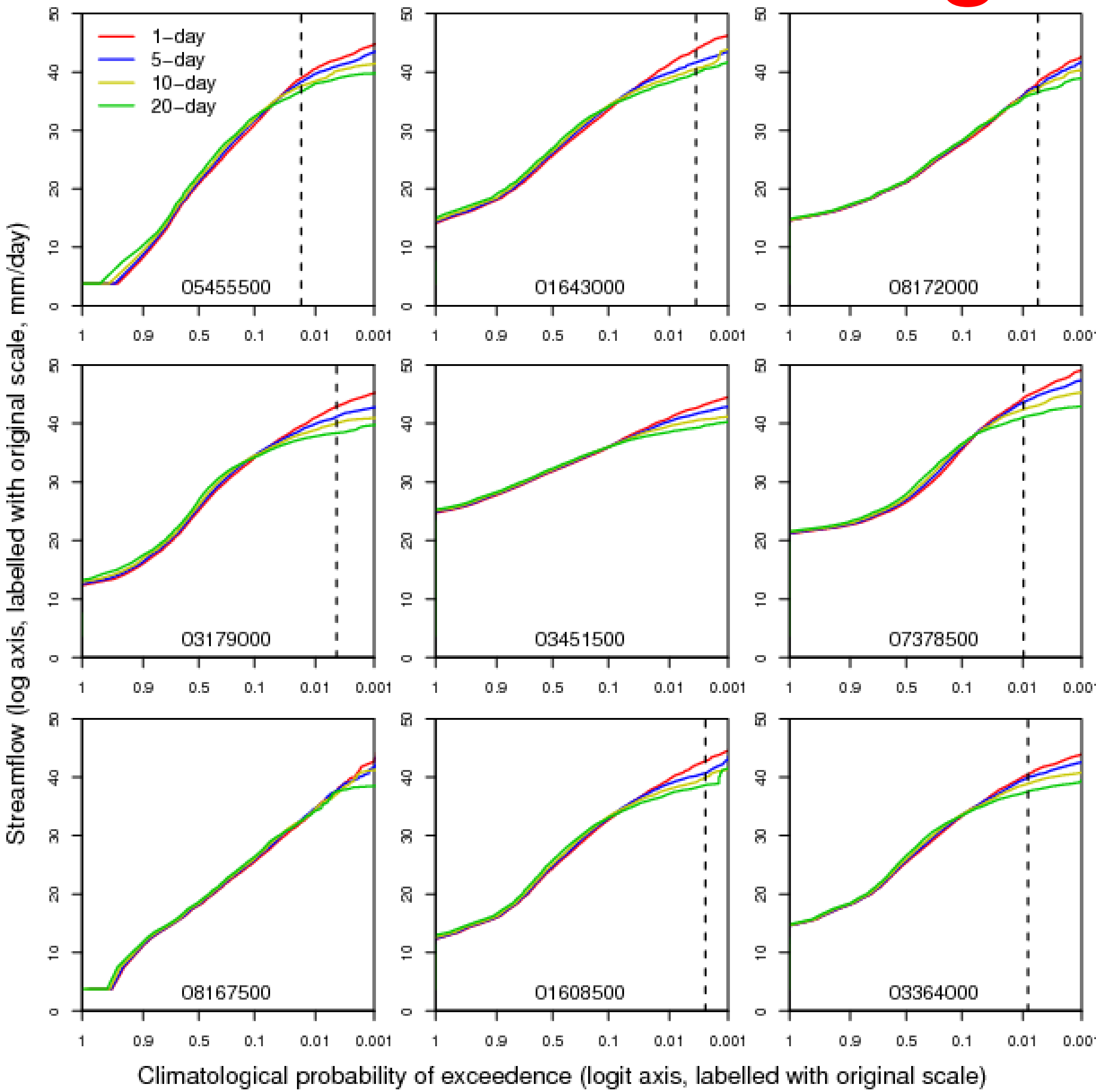
Correlation

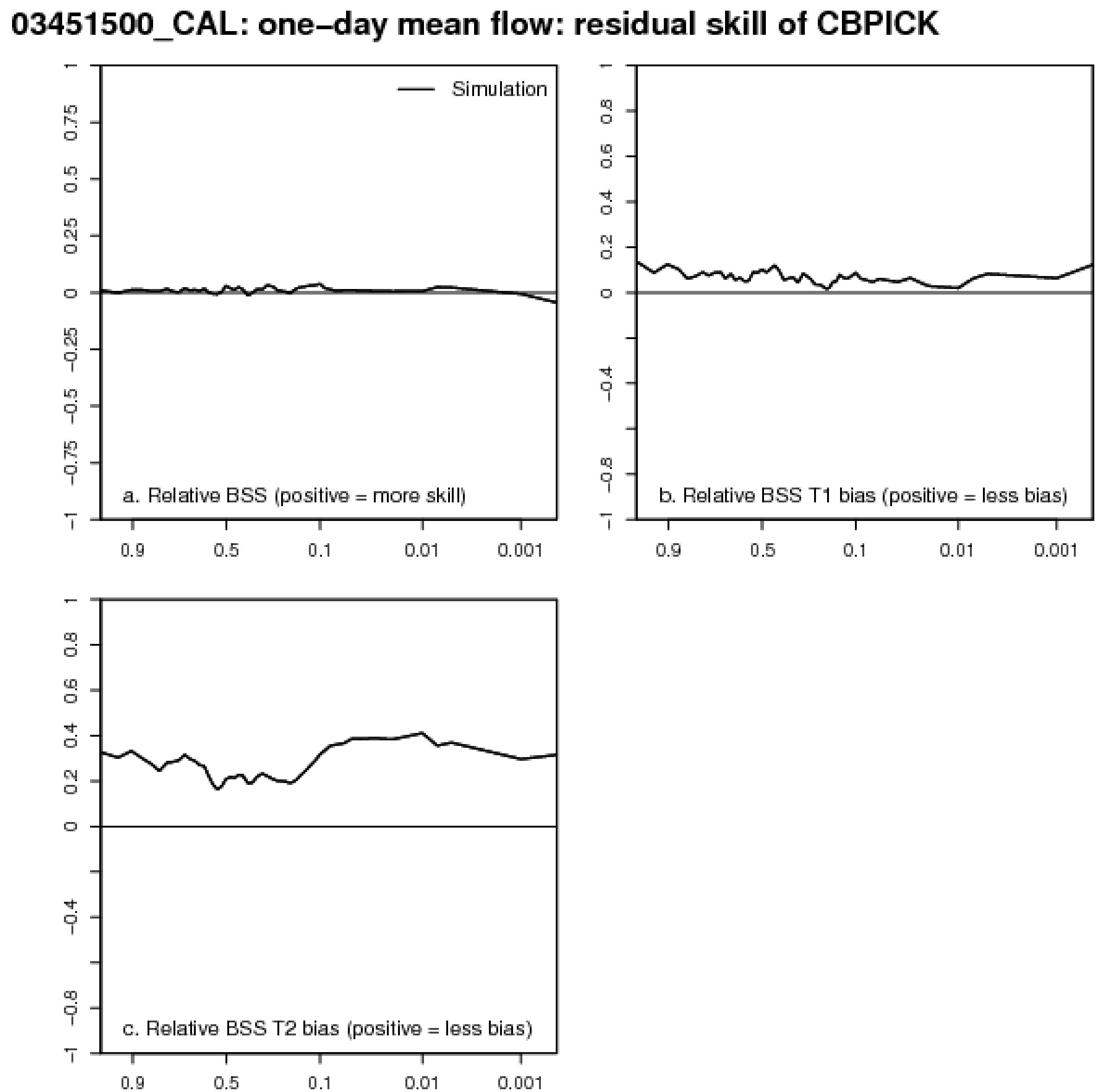


Brier Skill Score

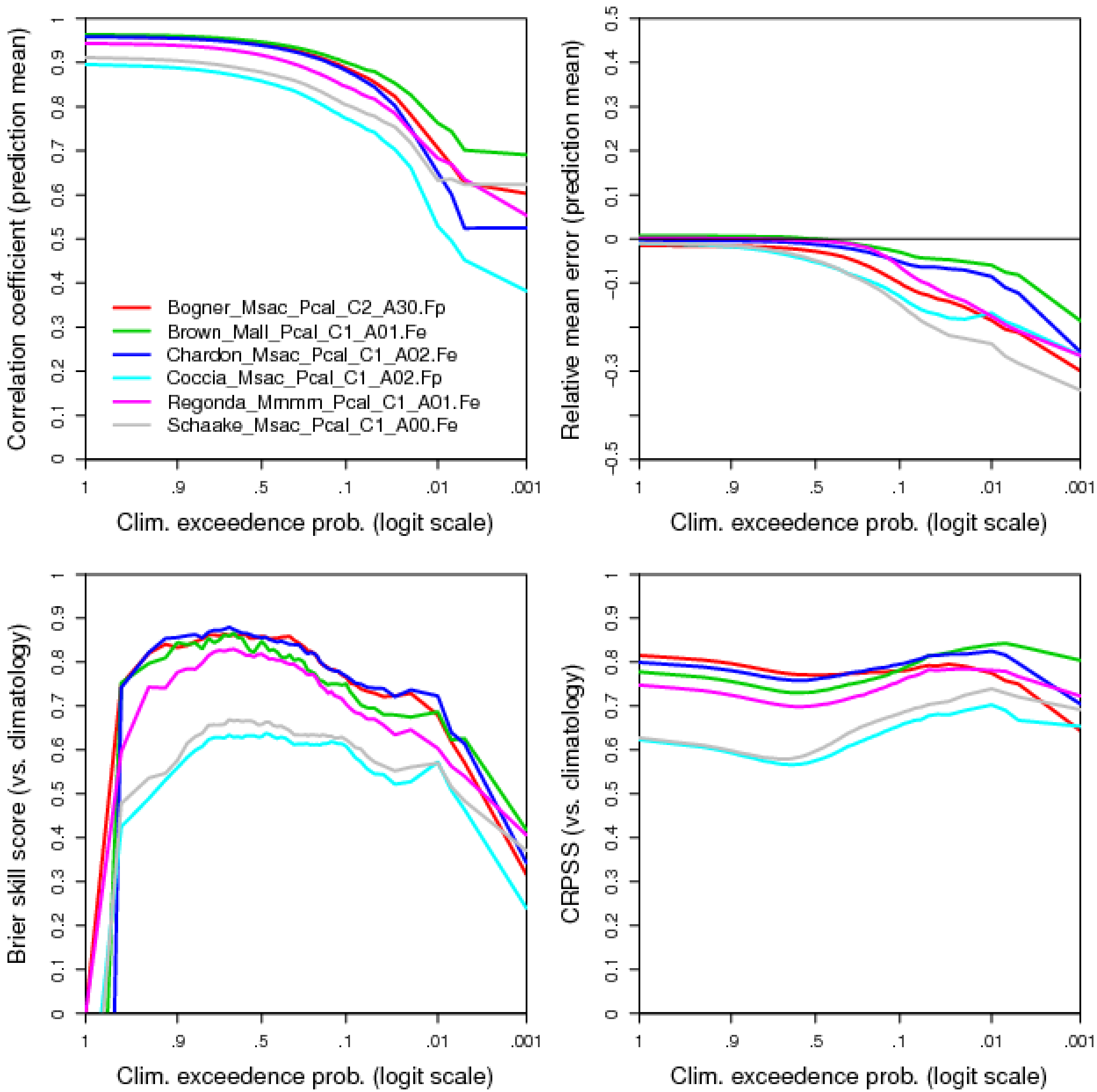


Observed climatologies




03451500_CAL: one−day mean flow: residual skill of CBPICK
Simulation
a. Relative BSS (positive = more skill)
b. Relative BSS T1 bias (positive = less bias)
c. Relative BSS T2 bias (positive = less bias)
0.9
0.5
0.1
0.01
0.001

03451500: Scenario 1



01643000: Scenario 1