Probabilistic Precipitation Calibration Using Two-parameter Ensemble Model Output Statistics

Xiang Su
Jiangsu Meteorological Observatory
Nanjing 210008, China

Aug 27, 2019 at NCEP, College Park

Background

EMOS

(NR)

Ensemble Model Output Statistics

(Nonhomogeneous Regression)

Need

- Reliable **probabilistic quantitative precipitation forecast (PQPF)** is essential for weather forecast centers and hydro-meteorological applications
- Due to imperfect initial condition and model configuration, ensemble forecast systems are usually subject to **biases** and **dispersion errors**

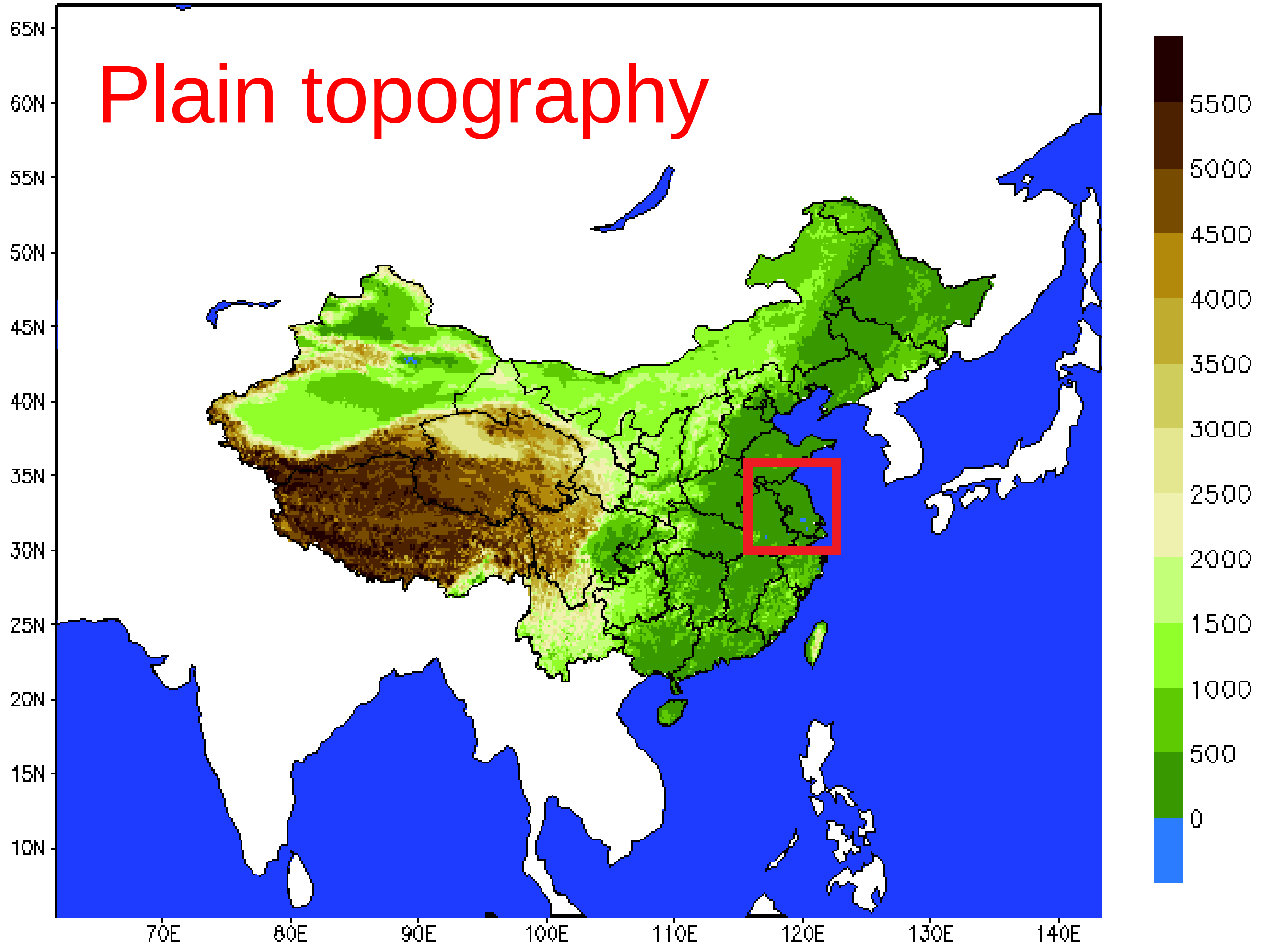
Method

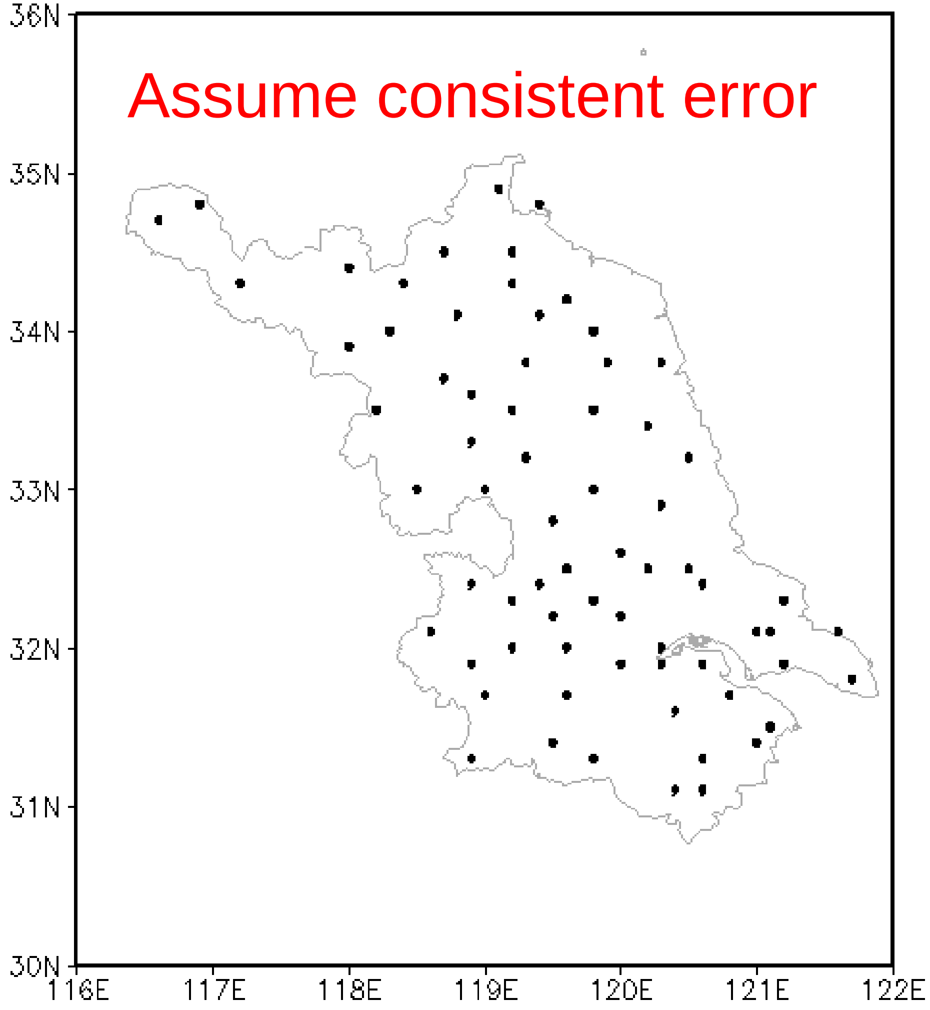
- Statistical post-processing methods are often used to calibrated the raw ensemble forecasts to generate more **reliable** and **accurate** probabilistic forecasts
- **EMOS** is one of the state of art ensemble post-processing techniques (firstly proposed by Gneiting et al. 2005 for Gaussian variables)

Use

- Scheuerer (2014) used EMOS for PQPF based on **left-censored GEV distribution**
- Scheuerer and Hamill (2015) used EMOS for PQPF based on **censored shifted gamma (CSG) distribution**

Aim: Post-process PQPFs of 70 stations in Jiangsu Province, China





Data

- Observation data: 00-00 (UTC) daily precipitation from 70 rain gauge stations in Jiangsu Province, China.
- Forecast data: ECMWF 24-h ensemble precipitation forecast initialized on 12UTC with 50 perturbed members on 0.5*0.5 degree grid.
- Forecast data is interpolated onto the 70 rain gauge stations.
- Forecast lead time: 012-036h and 036-060h.
- Validation period: June to August, 2017.
- Training method: Train each day individually using 40-day combined symmetric sliding window (20 latest days with observation and 20 days after the forecast day of previous year)

Verification methods

- Continuous Ranked Probability Score (CRPS)
- Brier Skill Score (BSS)
- Reliability Diagram

Station CRPS

$$\mathrm{crps}(\tilde{G}_i, y_i) = \int_{-\infty}^{\infty} [\tilde{G}_i(t) - \mathrm{H}(t - y_i)]^2 \, dt$$

Overall CRPS

$$CRPS = \frac{1}{Ns} \sum_{i=1}^{\mathrm{Ns}} \mathrm{crps}(\tilde{G}_i, y_i)$$

Station BS

$$\mathrm{bs_i} = (p_i - o_i)^2$$

Overall BS

$$BS = \frac{1}{Ns} \sum_{i=1}^{\mathrm{Ns}} \mathrm{bs_i}$$

Overall BSS

$$BSS = 1 - \frac{BS}{BS_{RAW}}$$

BS decomposition

$$BS = REL - RES + UNC$$

reliability resolution uncertainty

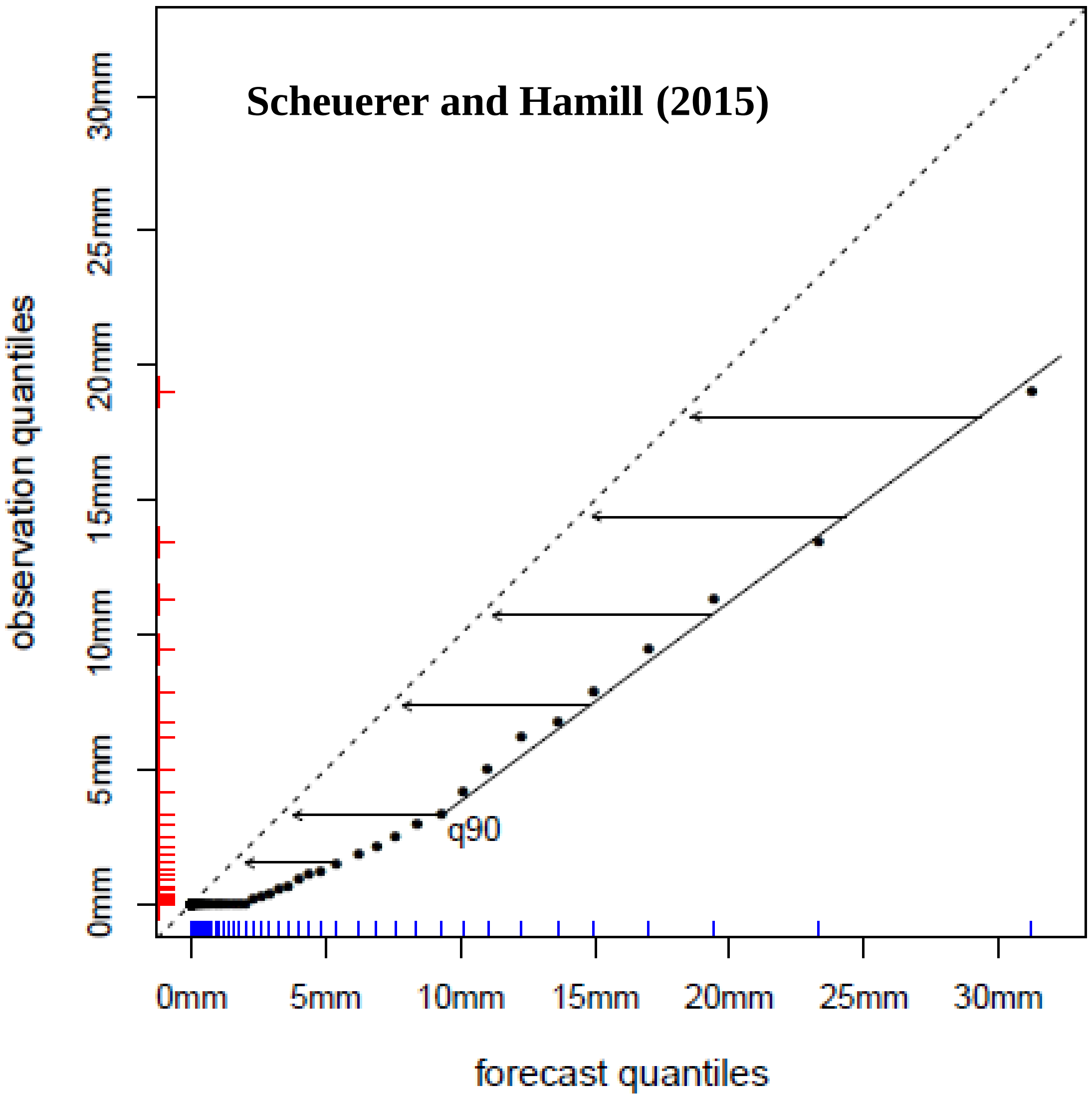
Steps of EMOS

1. First-moment bias correction of all ensemble members
2. Select appropriate distribution function for target variable
3. Link parameters to ensemble statistics
4. Minimize CRPS during training period to get optimal parameters
5. Use EMOS to post-process ensemble PQPFs

Steps of EMOS

1. First-moment bias correction of all ensemble members
2. Select appropriate distribution function for target variable
3. Link parameters to ensemble statistics
4. Minimize CRPS during training period to get optimal parameters
5. Use EMOS to post-process ensemble PQPFs

Quantile Mapping (QM)



First we calculate forecast and observation quantiles

$$q_f(k/100) \qquad q_o(k/100)$$

where k=1,2,…,99. Values between the fixed quantiles are linearly interpolated.

For forecast lower than $q_f(k_l/100)$, where $k_l = 90$ the QM corrected forecast is $q_o(k_l/100)$

As for forecast larger than $q_f(k_l/100)$, the QM corrected forecast is

$$\tilde{f}_x = \max\left\{q_o(k_l/100) + \zeta \cdot (f_x - q_f(k_l/100)), 0\right\}$$

Where

$$\zeta = \frac{\sum_{i=k_l+1}^{99}\left(q_f(i/100) - q_f(k_l/100)\right)\left(q_o(i/100) - q_o(k_l/100)\right)}{\sum_{i=k_l+1}^{99}\left(q_f(i/100) - q_f(k_l/100)\right)^2}$$

Steps of EMOS

- First-moment bias correction of all ensemble members
- Select appropriate distribution function for target variable
- Link parameters to ensemble statistics
- Minimize CRPS during training period to get optimal parameters
- Use EMOS to post-process ensemble PQPFs

Left-censored Generalized Extreme Value Distribution (GEV) distribution

The cumulative distribution function (CDF) of GEV distribution:

$$G(x)=\begin{cases}\exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right], \xi\neq 0\\ \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right], \xi=0\end{cases}$$

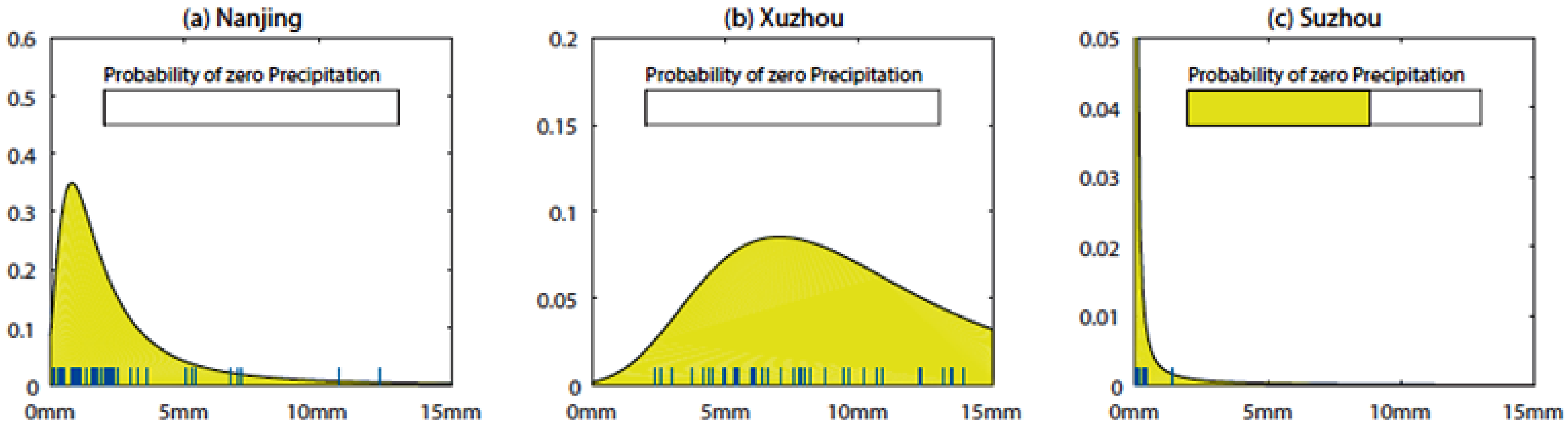
Where μ,σ,ξ are **location** parameter, **scale** parameter, and **shape** parameter, respectively.

Precipitation is a skewed non-negative variable. The left-censored GEV distribution

$$\tilde{G}(x)=\begin{cases}G(x), x\geq 0\\ 0, x<0\end{cases}$$

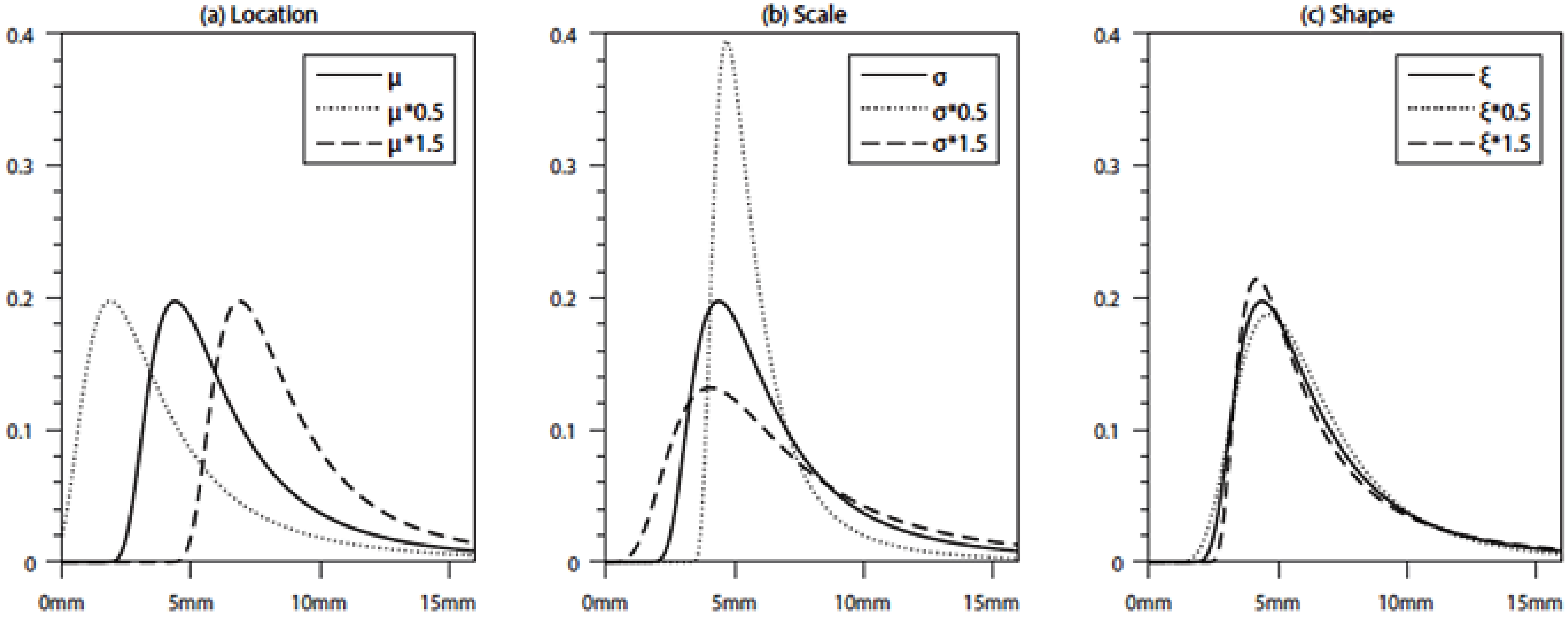
is non-negative, flexible enough and able to extrapolate precipitation extremes with a long tail

Predictive distribution of daily precipitation (00-00 UTC) on 31 July 2016 fitted by left-censored GEV distribution



Sensitivity of GEV distribution to location, scale and shape parameters by increasing and decreasing a certain parameter value while remain other parameters unchanged.

The **location parameter** mainly adjusts the **predictive mean** and the **scale parameter** mainly adjusts the **predictive variance**, while the distribution is not very sensitive to the shape parameter.



Steps of EMOS

1. First-moment bias correction of all ensemble members
2. Select appropriate distribution function for target variable
3. Link parameters to ensemble statistics
4. Minimize CRPS during training period to get optimal parameters
5. Use EMOS to post-process ensemble PQPFs

Left-censored Generalized Extreme Value Distribution (GEV) distribution

Take a **linear** relationship between the **location** parameter and **ensemble mean**, and between the **scale** parameter and **ensemble variance**

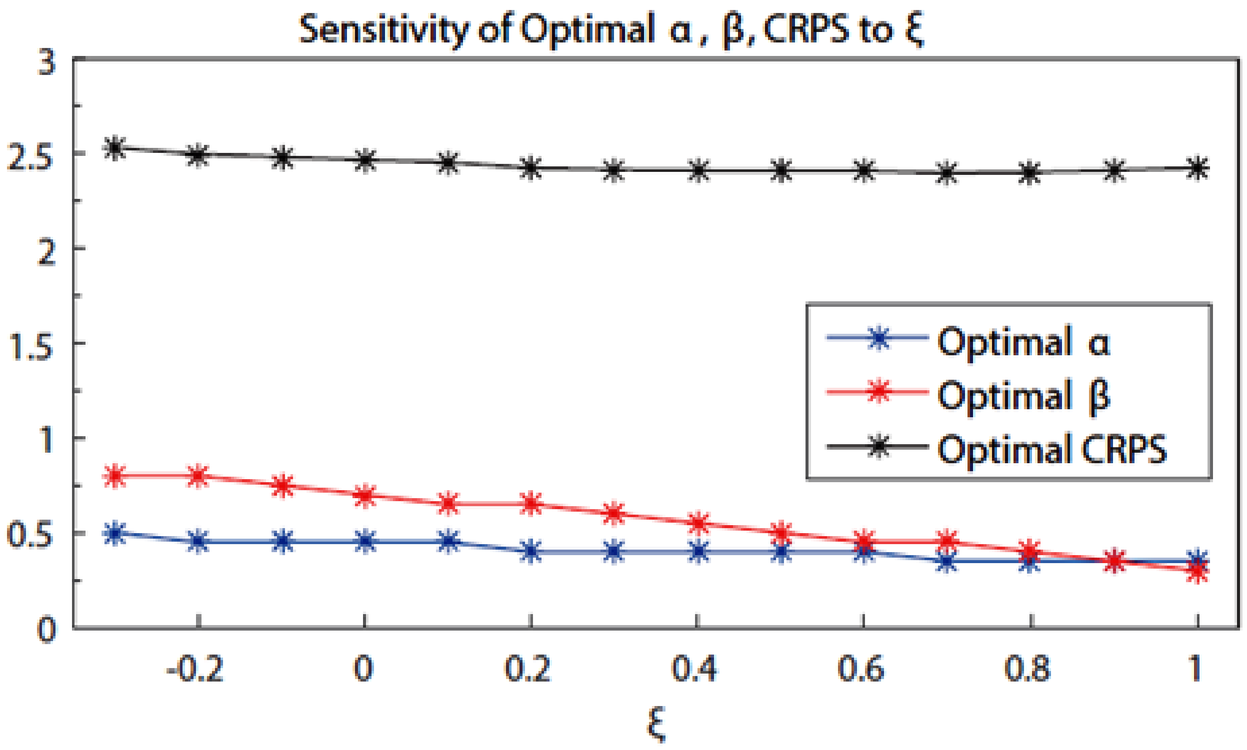
$$
\begin{cases}
\mu = \alpha \cdot \left(\frac{1}{M} \sum_{i=1}^{M} X_i \right) \\
\sigma = \beta \cdot \sqrt{\frac{1}{M-1} \sum_{i=1}^{M} \left(X_i - \overline{X} \right)^2} \\
\xi = const.
\end{cases}
$$

The shape parameter is set to constant.

(Lerch and Thorarinsdottir, 2013)

In Scheuerer (2014), the shape parameter is always assumed to be between -0.278 and 1.

We tested the different values of shape parameter.



Steps of EMOS

1. First-moment bias correction of all ensemble members
2. Select appropriate distribution function for target variable
3. Link parameters to ensemble statistics
4. Minimize CRPS during training period to get optimal parameters
5. Use EMOS to post-process ensemble PQPFs

Multi-parameter optimization problem (minimize CRPS)

◆ The quasi-newton method, for example, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, is usually used to obtain the approximate solution.

Problem No. 1

◆ These algorithms cannot guarantee convergence to a global optimal solution, i.e., they can only get locally optimal solution and better solution could exist somewhere.

Obtain a global optimal estimation of the parameters with brute-force method

Problem No. 2

◆ The solution could be sensitive to multi-parameter initial values, which are required to start these algorithms.

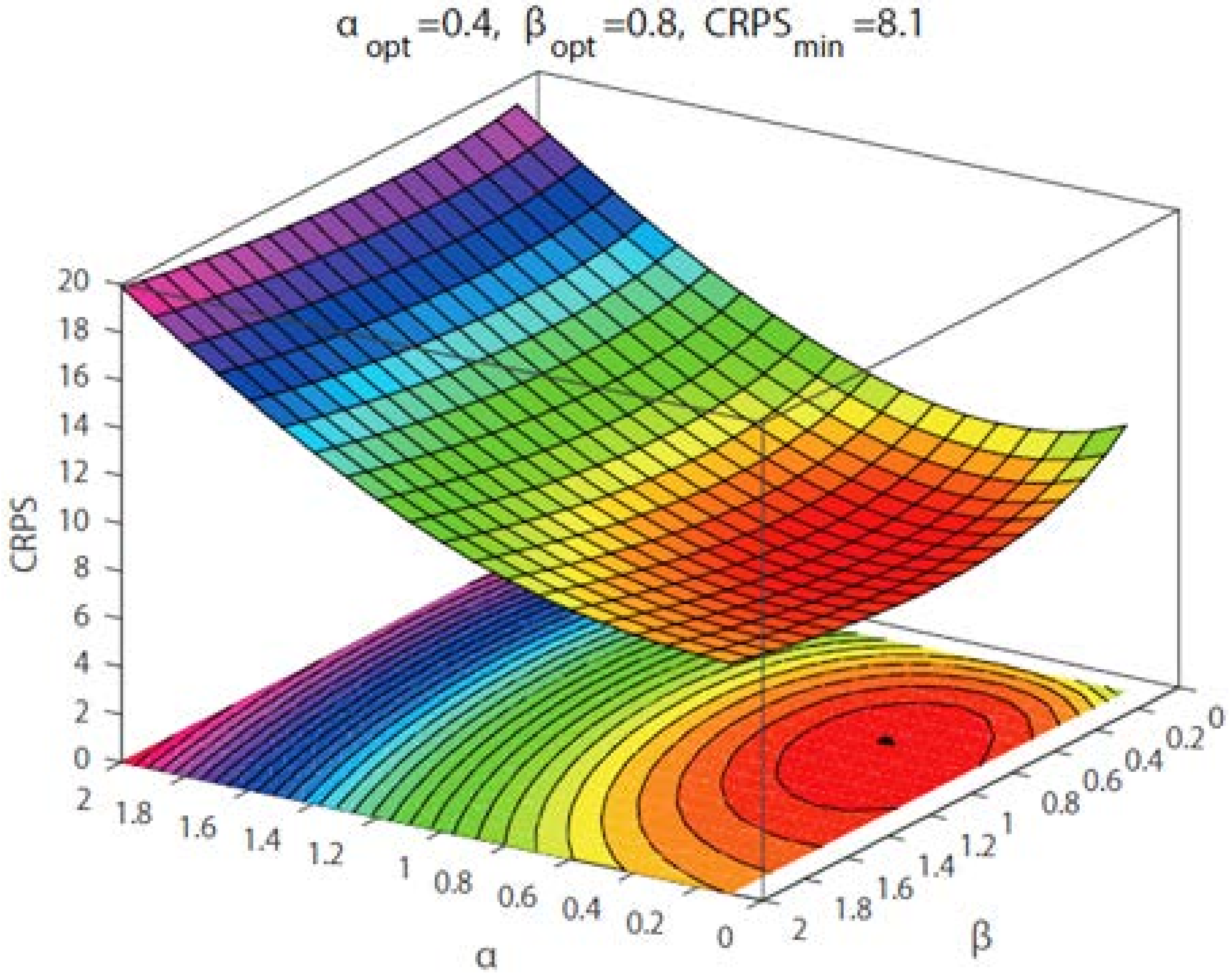
No initial values are required any more

Problem No. 3

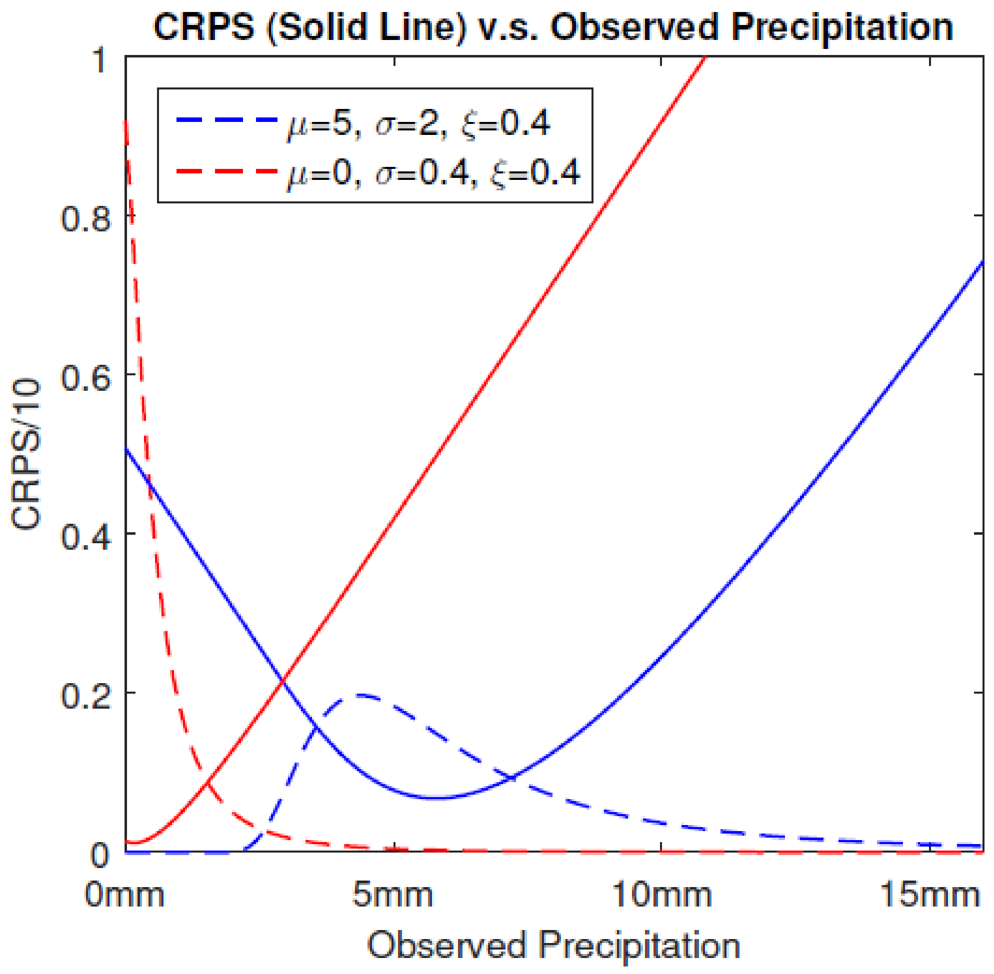
◆ Too many parameters often result in overfitting.

Reduce parameters to two

Brute-force method to minimize CRPS

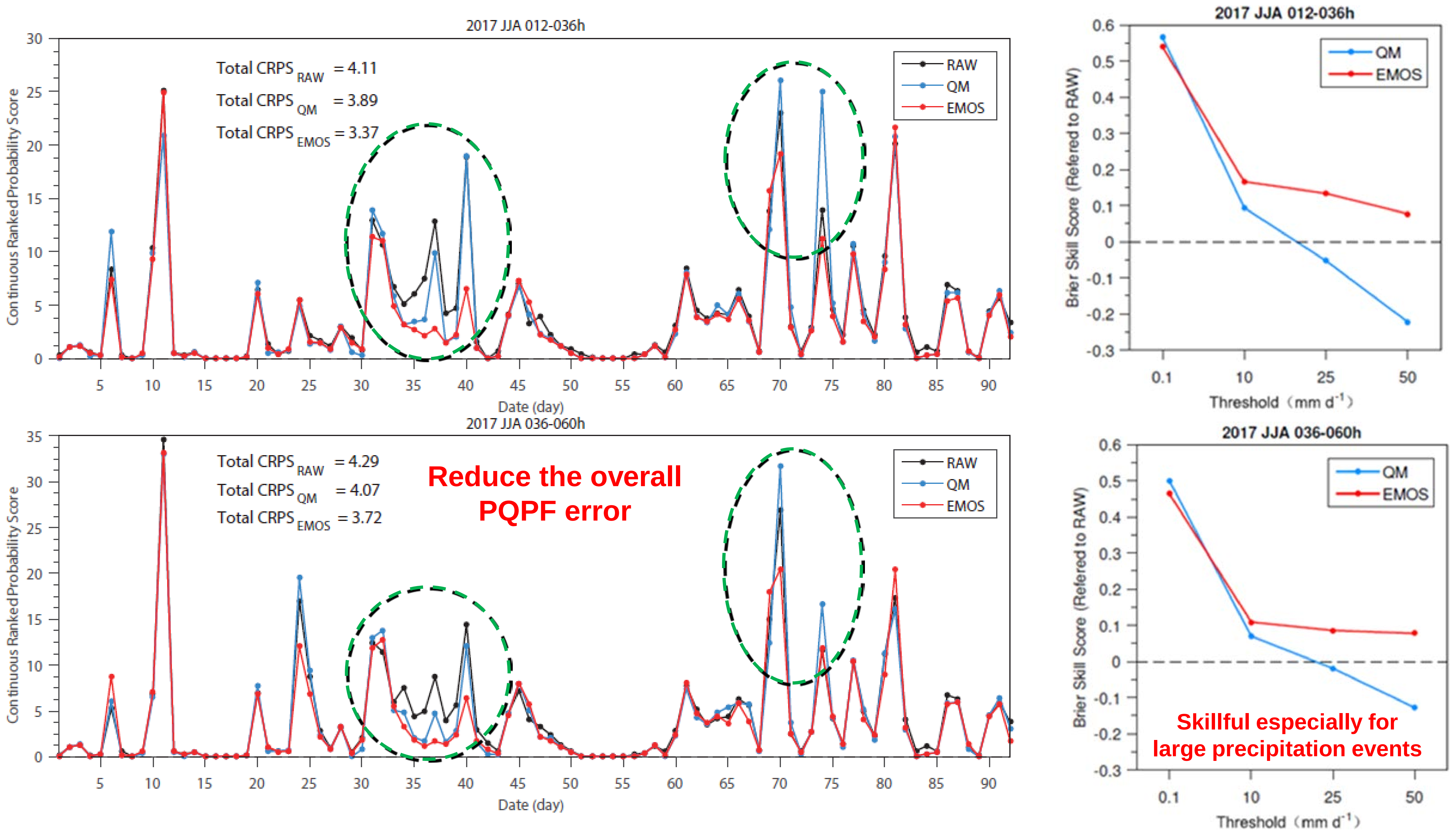


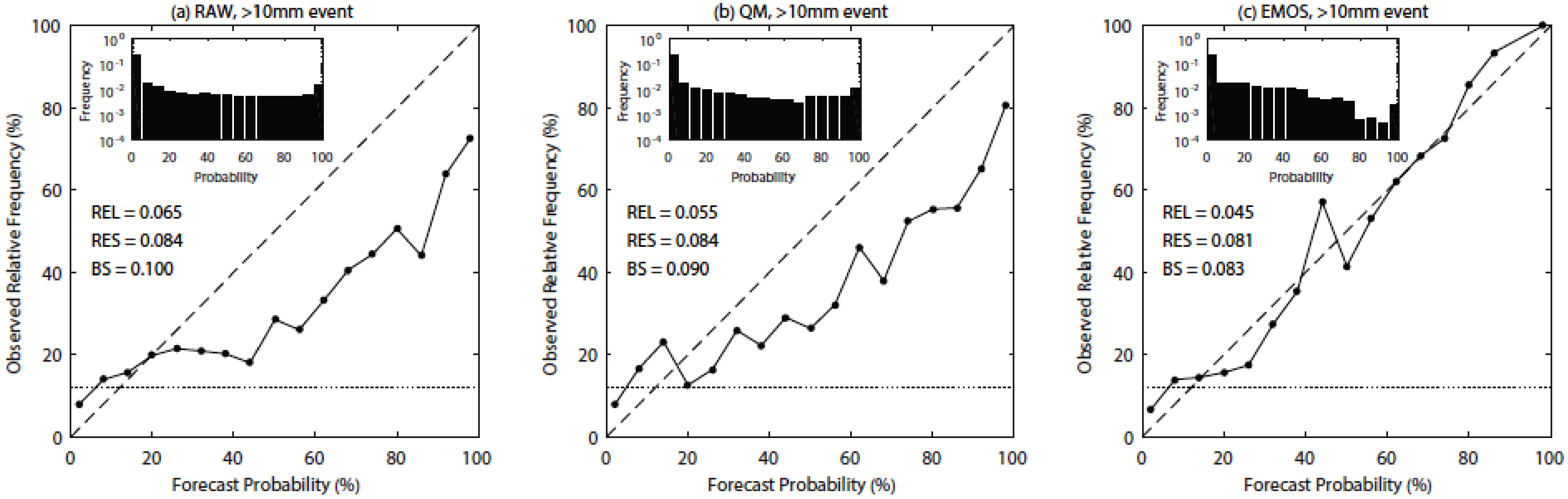
Given the predictive GEV distribution,
how CRPS changes with
different values of observation.



Steps of EMOS

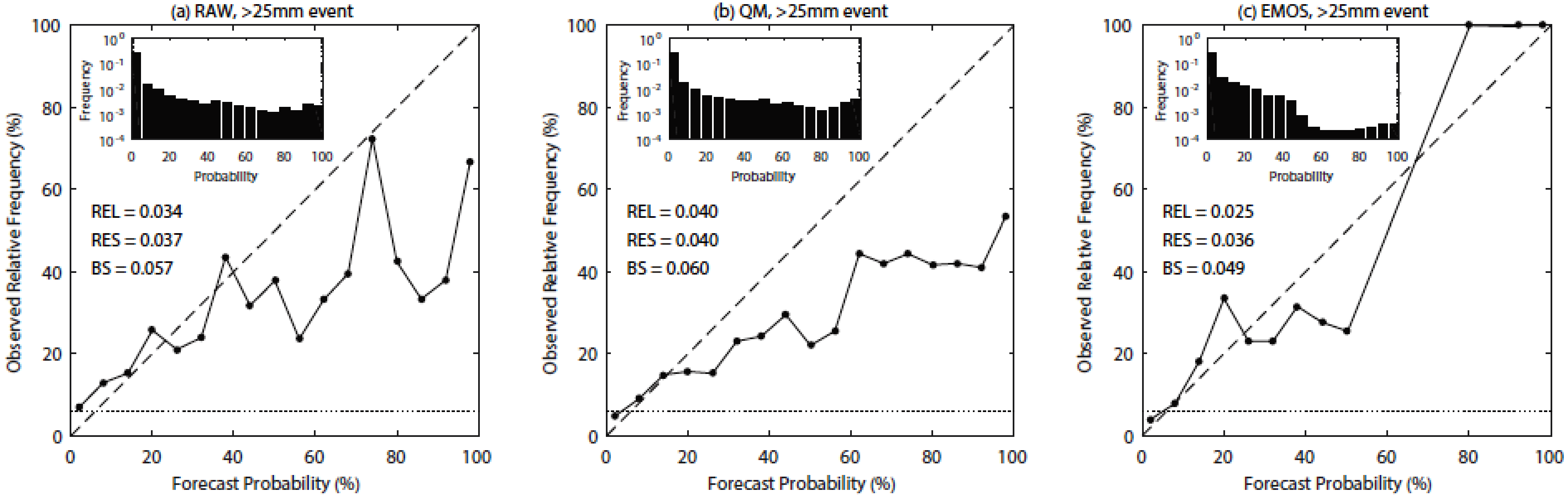
1. First-moment bias correction of all ensemble members
2. Select appropriate distribution function for target variable
3. Link parameters to ensemble statistics
4. Minimize CRPS during training period to get optimal parameters
5. Use EMOS to post-process ensemble PQPFs


2017 JJA 012-036h
Continuous Ranked Probability Score
Date (day)
Total CRPS RAW = 4.11
Total CRPS QM = 3.89
Total CRPS EMOS = 3.37
RAW
QM
EMOS
2017 JJA 012-036h
Brier Skill Score (Refered to RAW)
Threshold (mm d-1)
QM
EMOS
2017 JJA 036-060h
Continuous Ranked Probability Score
Date (day)
Total CRPS RAW = 4.29
Total CRPS QM = 4.07
Total CRPS EMOS = 3.72
Reduce the overall PQPF error
RAW
QM
EMOS
2017 JJA 036-060h
Brier Skill Score (Refered to RAW)
Threshold (mm d-1)
QM
EMOS
Skillful especially for large precipitation events



The conditional biases are well calibrated

The better BS of EMOS is mainly due to the well calibrated reliability



Conclusion

- A two-parameter EMOS post-processing model based on the left-censored GEV distribution for the short-term ECMWF ensemble precipitation forecast is proposed.
- The purpose is to avoid overfitting and obtain global optimal solution of model parameters.
- The predictive mean and variance of ensemble precipitation forecast are mainly adjusted by the location and scale parameters respectively, while the predictive distribution is not sensitive to the shape parameter.
- The two-parameter EMOS can reduce overall probabilistic forecast error and improve the probabilistic precipitation forecast skill of different precipitation thresholds during summer time, especially for large precipitation events.
- The good forecast skill of the two-parameter EMOS post-processing model is mainly due to the better calibrated reliability.

Thank you!