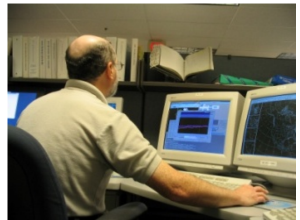
# Ensemble Post-Processor (EnsPost)

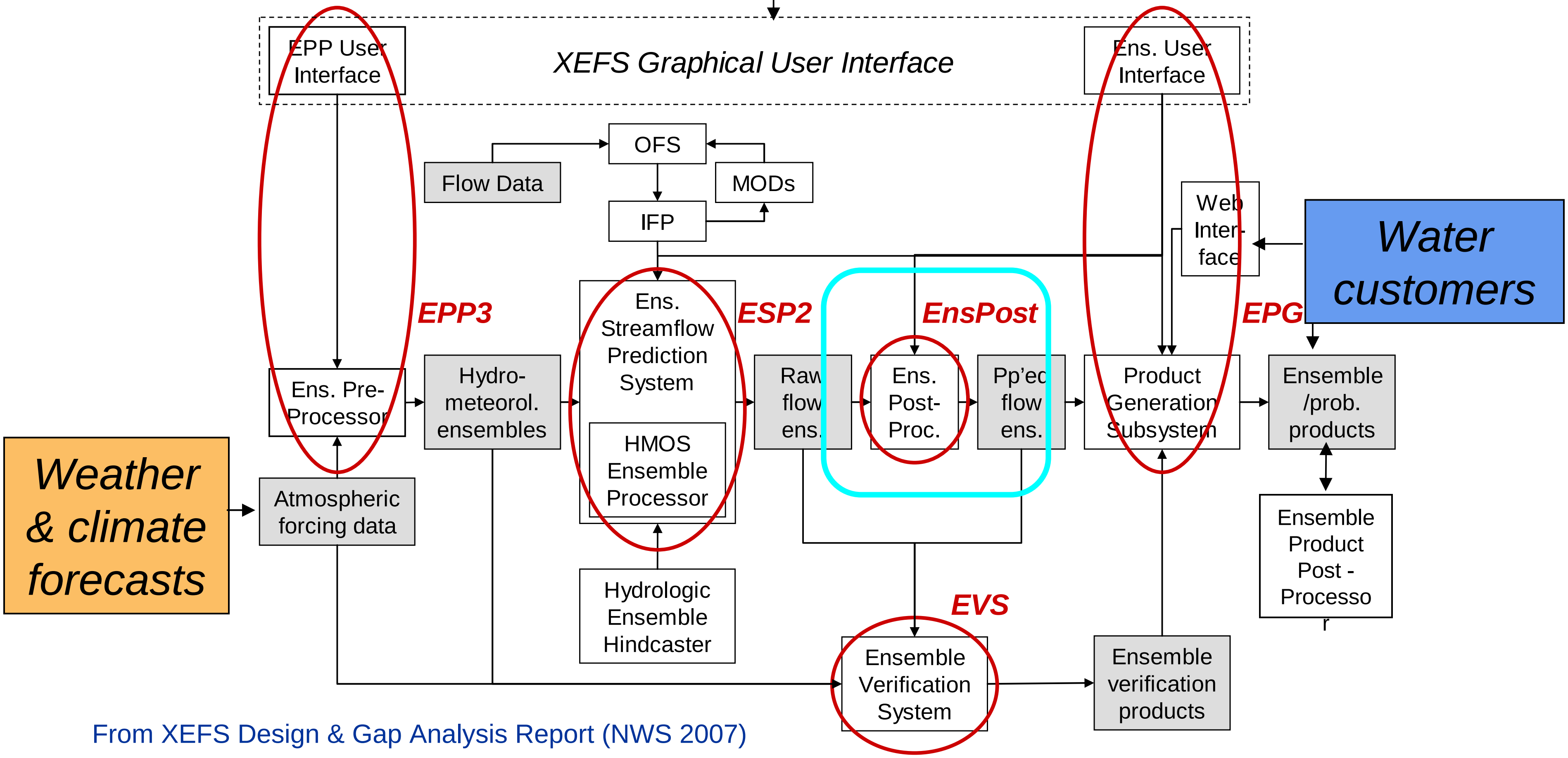
DJ Seo and Satish Regonda

Hydrologic Ensemble Prediction Group (HEP)
Hydrologic Science and Modeling Branch
Hydrology Laboratory
Office of Hydrologic Development
NOAA/National Weather Service

# EXperimental Ensemble Forecast System (XEFS)



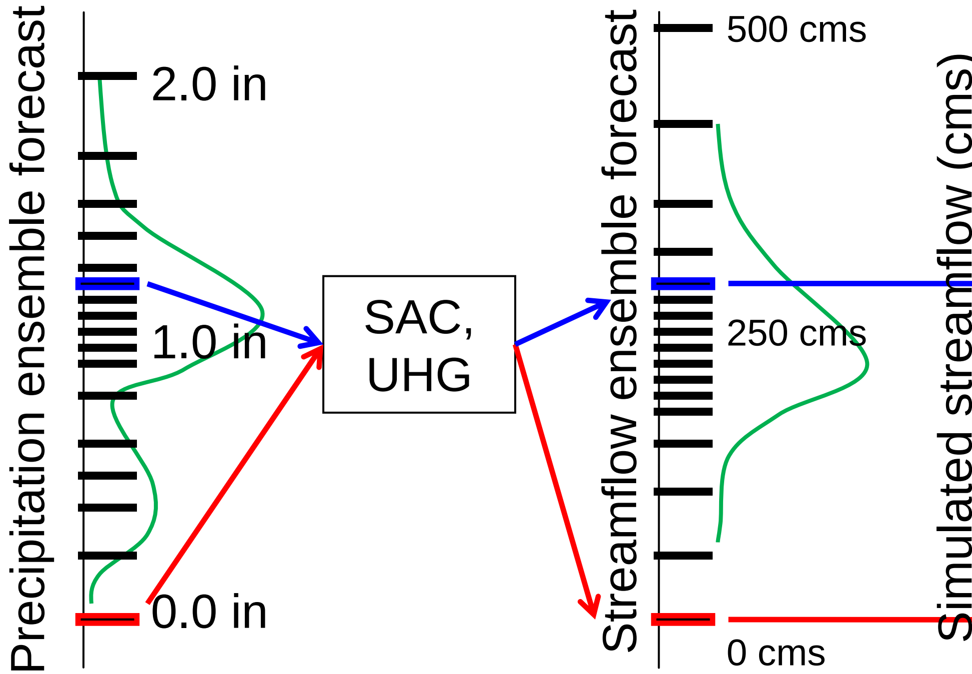
Forecasters add value

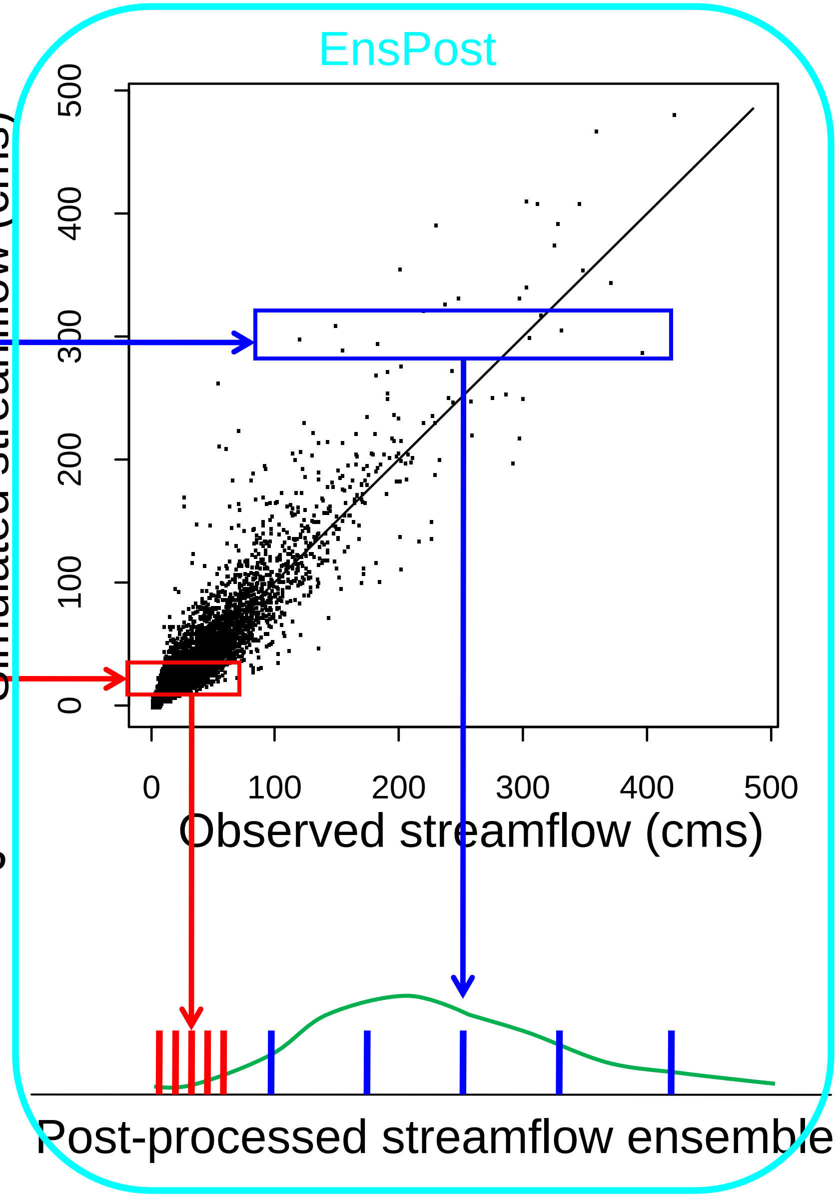



From XEFS Design & Gap Analysis Report (NWS 2007)

XEFS will enable seamless hydrologic ensemble prediction from weather to climate scales and translate weather and climate prediction into uncertainty-quantified water information

# Integration of input and hydrologic uncertainties



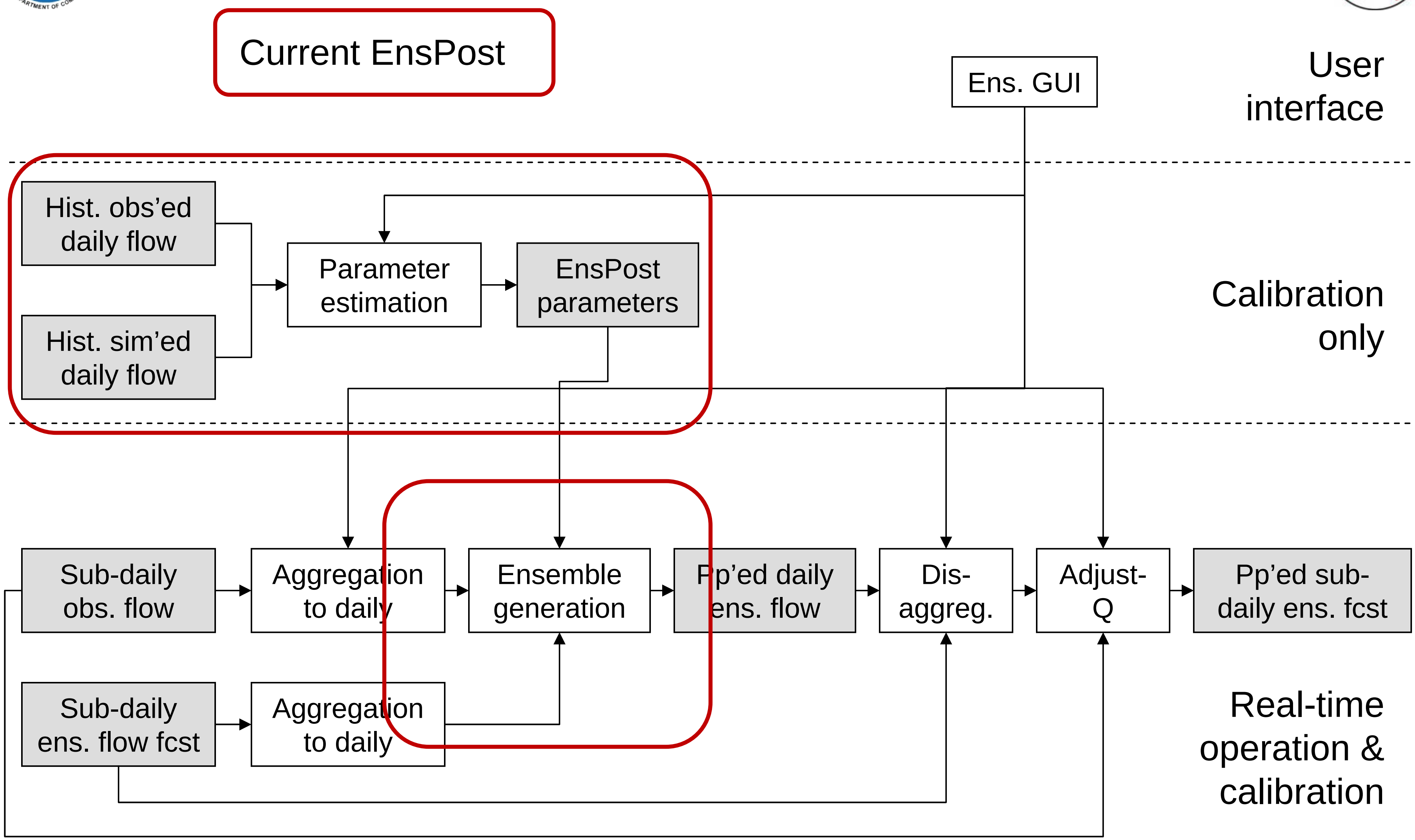

$$f_3(SQIN) = \int f_4(SQIN \mid FMAP)\ f_5(FMAP)\ dFMAP$$

$$f_1(QIN) = \int \boxed{f_2(QIN \mid SQIN)}\ f_3(SQIN)\ dSQIN$$

Krzysztofowicz (1999), Seo et al. (2006)

# Status

- A prototype ensemble post-processor exists (Seo et al. 2006)
  - An earlier version of the processor was implemented in AWIPS in 200? to support short-term (~5 days) ensemble forecasting
  - Uses daily streamflow data for statistical modeling due to general lack of availability of sub-daily historical streamflow data necessary for calibration
- Work carried out, but not completed, in 2007~2008
  - An Adjust-Q procedure for EnsPost
    - To mimic MODs in calibration
    - To mimic operational Adjust-Q in ensemble generation
  - A multiscale probability matching procedure to render ensembles of weekly, monthly, seasonal, etc., flow volume reliable




From XEFS Design & Gap Analysis Report (NWS 2007)

# Remaining work for Phase 1

- Complete the Adjust-Q procedure for Ens Post
- Complete the multiscale probability matching procedure
- Disaggregation of daily to sub-daily flow
- Assessment of data and ensemble size requirement
- User interface and display tools for parameter estimation and calibration
- Training

From EnsPost Phase 1 GAPS (XEFS Design & Gap Analysis Report, NWS 2007)

# Issues

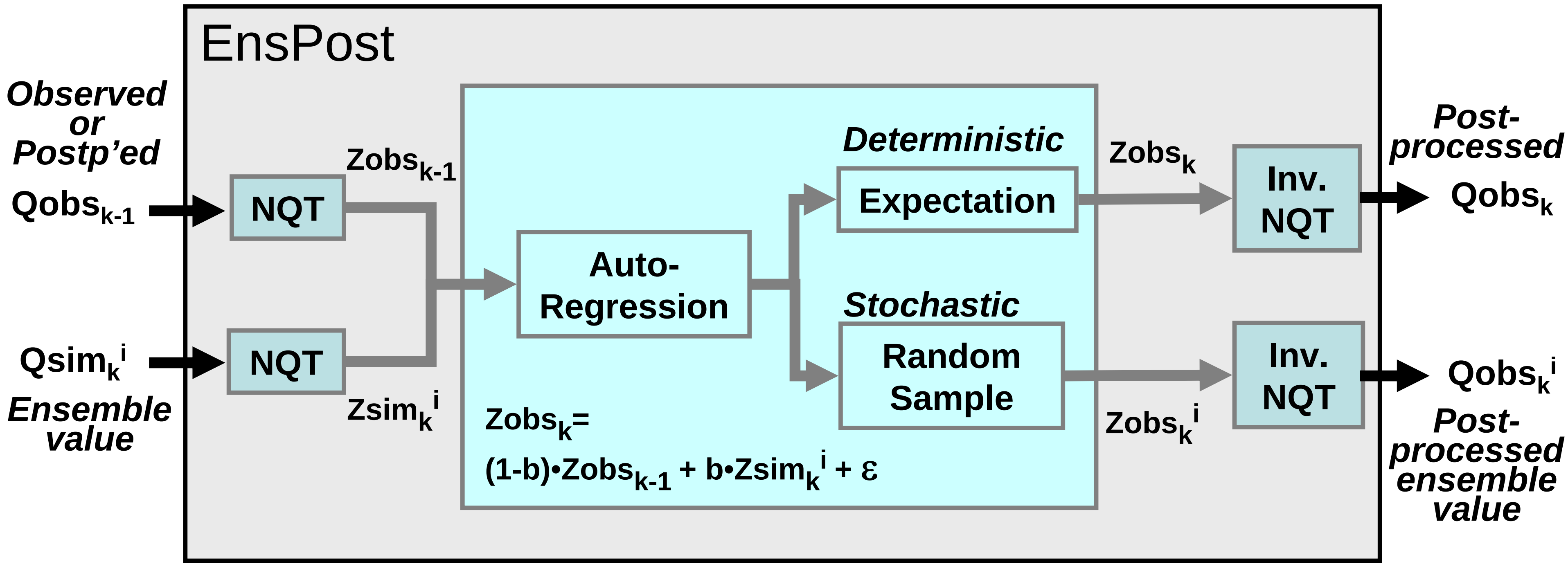
- Potency of EnsPost depends largely on:
  - The availability of long-term observed flow data
  - The degree to which stationarity in the streamflow climatology holds
- In areas where snowmelt is important, additional stratification of the data may be necessary, which would require a larger data set for parameter estimation / calibration
- Parameter estimation / calibration of EnsPost should be integrated with the general model calibration procedures

# EnsPost: Methodology

- Recursive linear regression in the normal space
  - Normal transformation is based on normal quantile transformation (NQT) of observed and model-simulated flows (daily)
- $Z^{obs}_{K} = (1\text{-}b)\ Z^{obs}_{K\text{-}1} + b\ Z^{sim}_{K} + \varepsilon$
  - b, and mean and variance of $\varepsilon$ are flow range- (low and high) and seasonality-dependent parameters
- Bias correction and uncertainty accounting
  - Generates ensembles via conditional simulation of $Q^{obs}_{K}$ given $Q^{obs}_{K\text{-}L}$ and $Q^{sim}_{K}$

# Methodology (cont.)

- Process for lead day k for the $i^{th}$ ensemble member



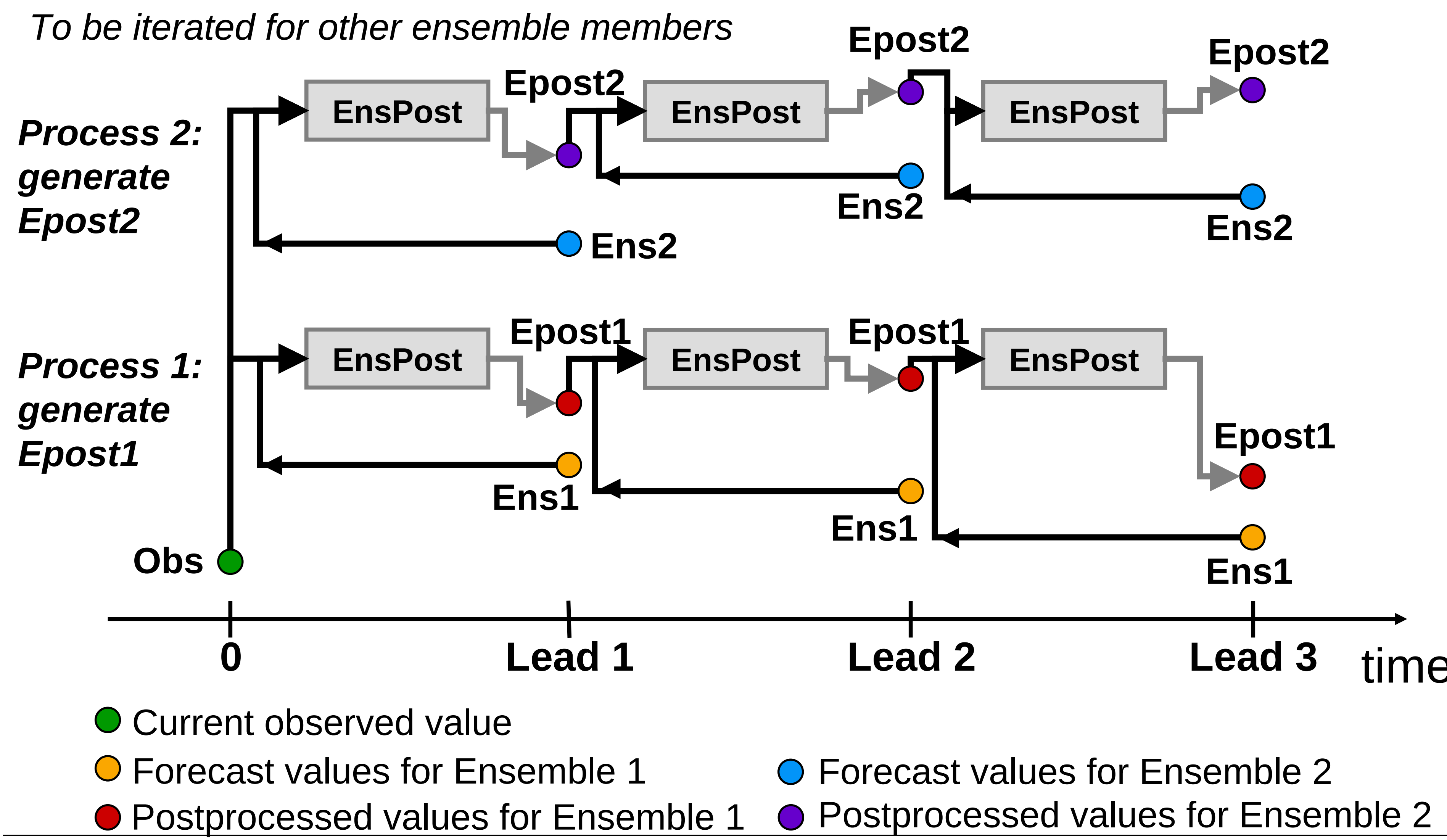

Observed or simulated value ingested by EnsPost:

Lead day k = 1: **Qobs$_0$** current observation

Lead day k > 1: **Qobs$_{k-1}$ / Qobs$_{k-1}^i$** previously post-processed value

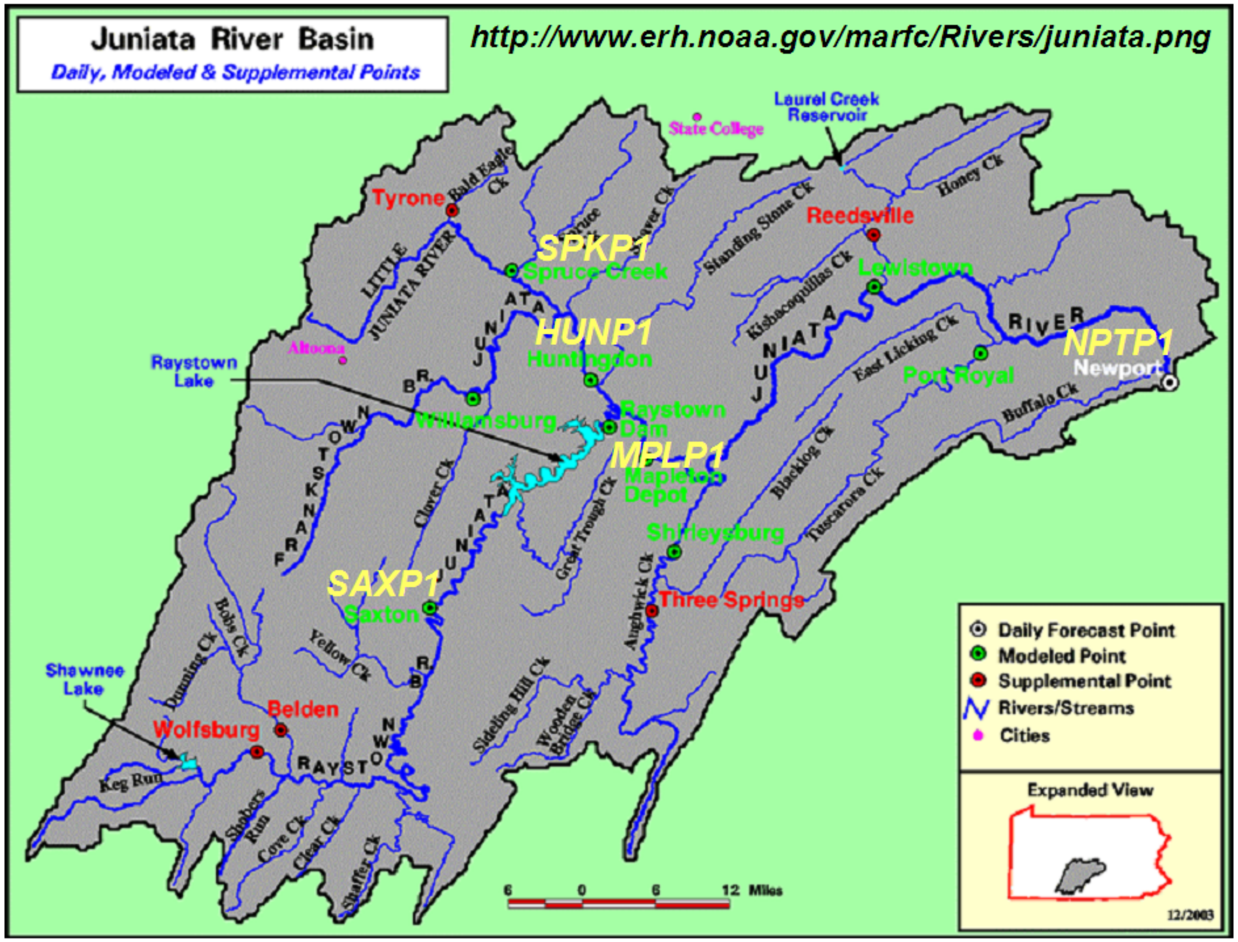
# Methodology (cont.)

- Generation of ensemble forecasts at daily time step



# Examples – MARFC/Juniata

The following 5 slides are excerpted from
http://www.copernicus.org/EGU/hess/hessd/3/1987/hessd-3-1987.htm



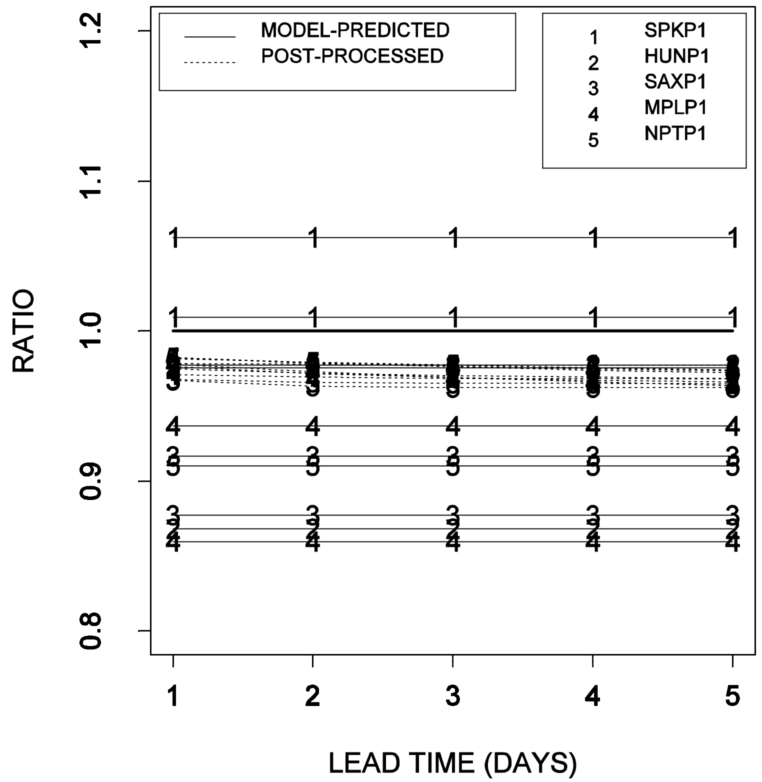



Fig 4. Ratio of the sum of the observed flow to that of the model-predicted (in solid line) or the post-processed (in dotted line) in the parameter estimation period versus lead time (days).

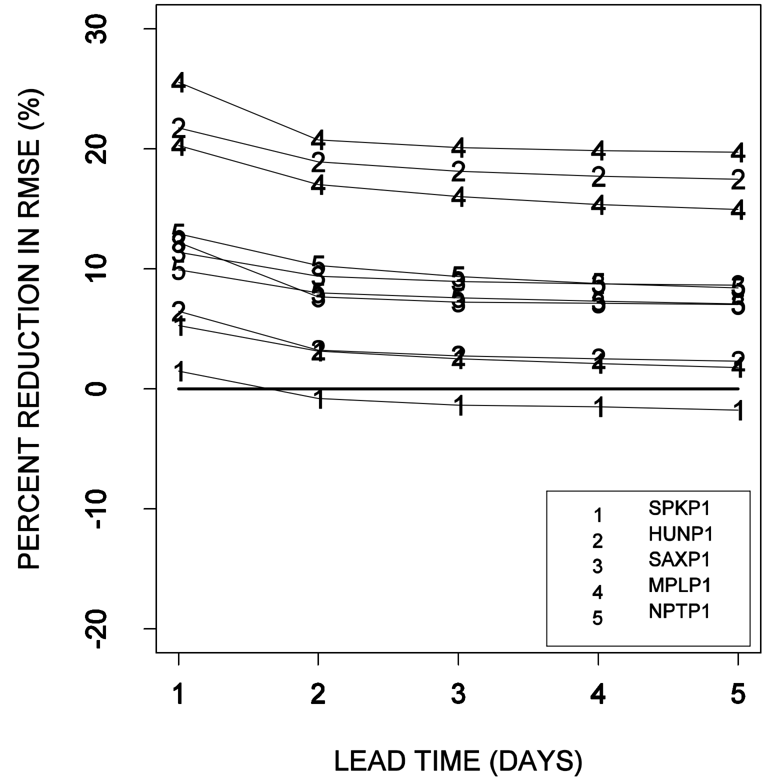



Fig 5. Percent reduction in root mean square error (RMSE) by the post-processed flow over the model-predicted in the parameter estimation periods versus lead time (days).

From Seo et al. (2006)

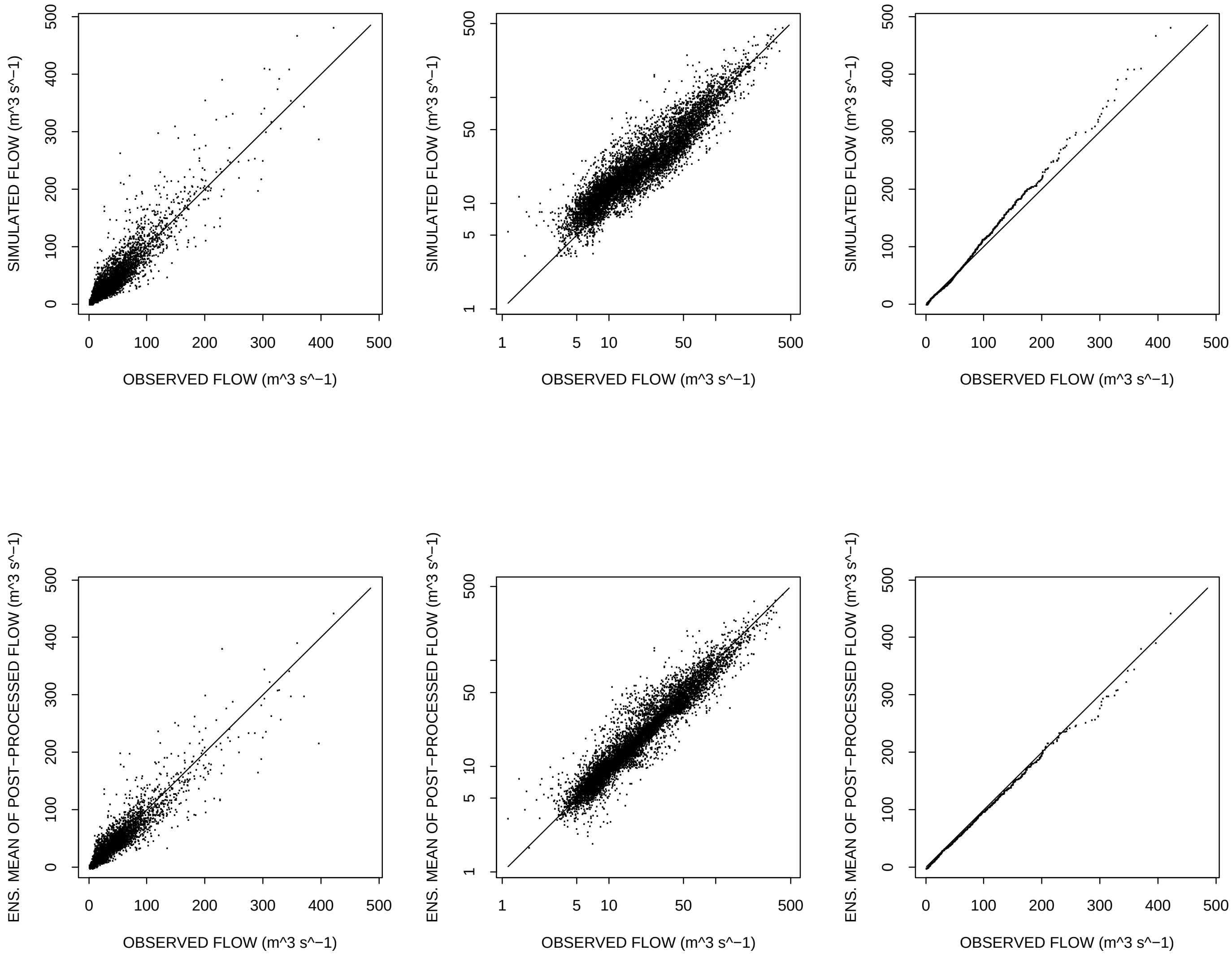



Fig 6. Scatter-plots in linear scale (left panels) and in log scale (middle panels), and quantile-quantile plots (right panels) of daily flow between the observed and the model-simulated flows (upper panels) and between the observed and the post-processed flows (lower panels) at HUNP1 for a parameter estimation period. The lead time is 1 day.

From Seo et al. (2006)

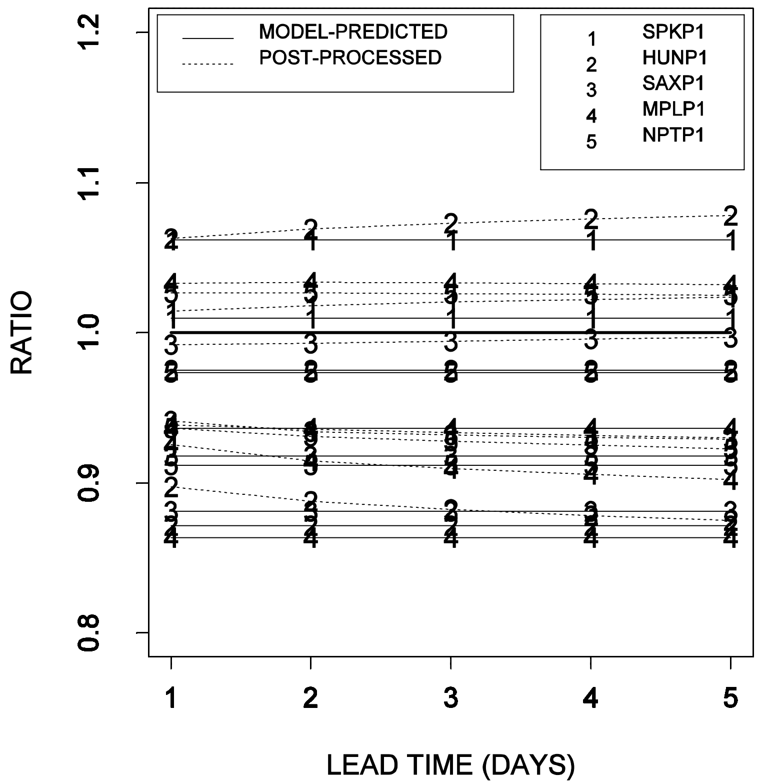



Fig 7. Same as Fig 4, but for the validation periods.

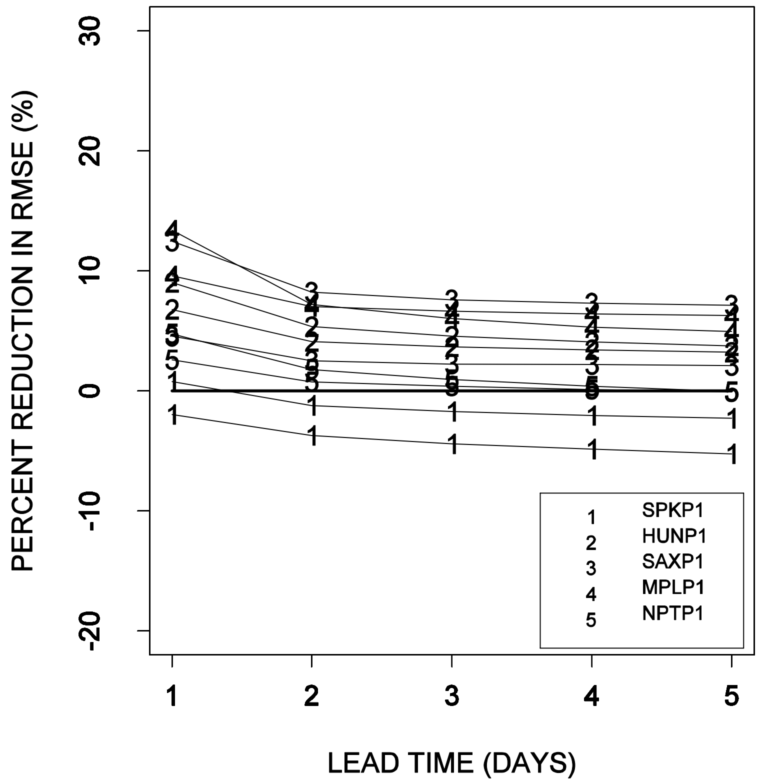
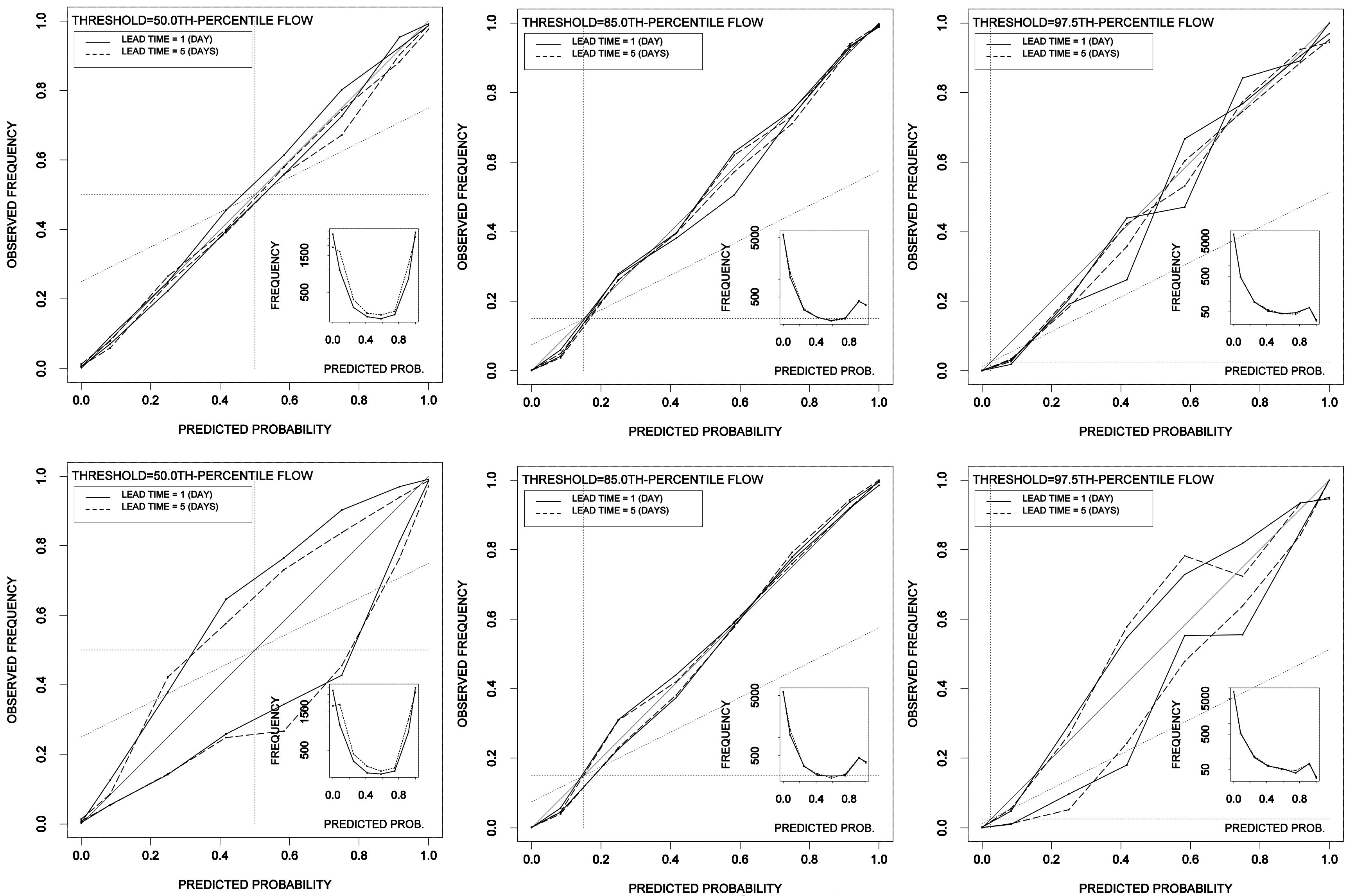


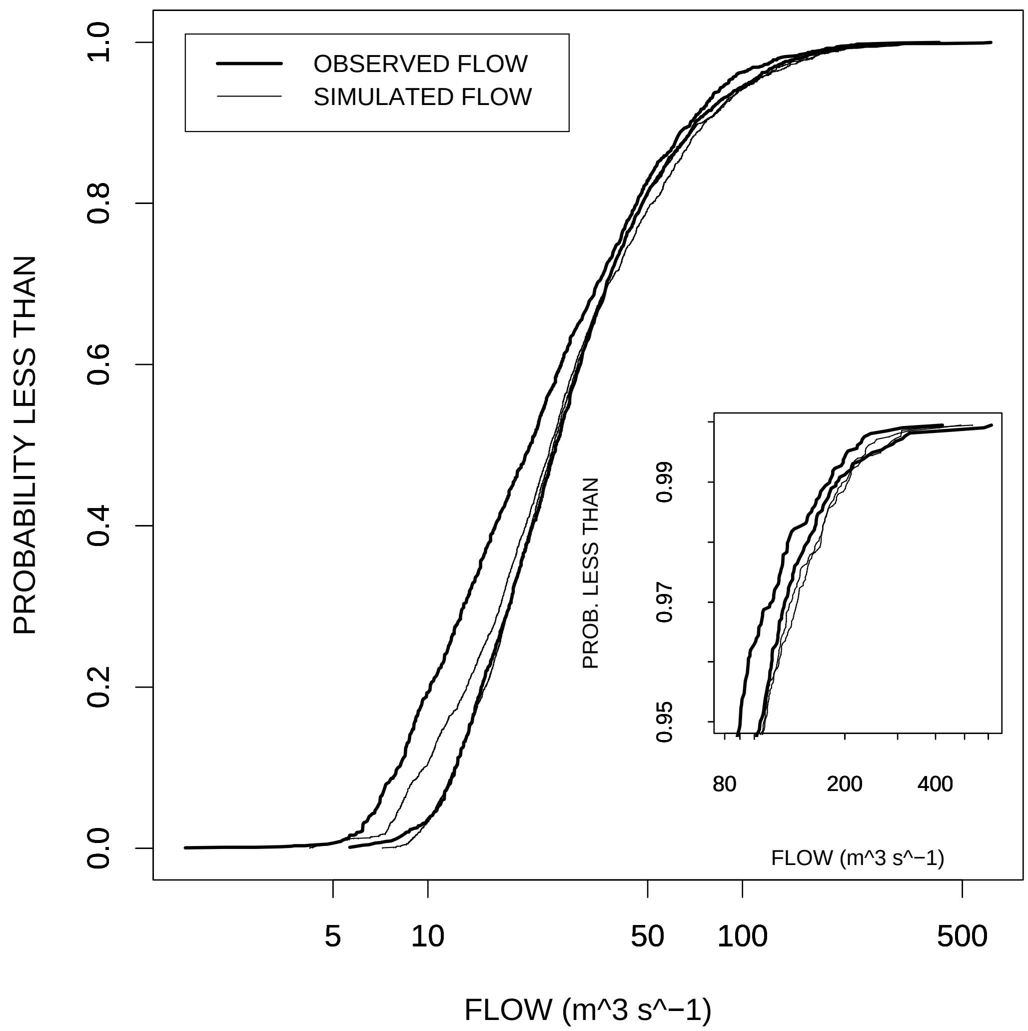

Fig 8. Same as Fig 5, but for the validation periods.

From Seo et al. (2006)

# Reliability diagrams of post-processed ensemble flows (daily) at 50th (left), 85th (middle) and 97.5th (right) percentiles for calibration (upper) and validation (lower) periods




From Seo et al. (2006)




Empirical cumulative distribution functions (ECDF) of observed (thick solid line) and simulated (thin solid line) flows at HUNP1. The upper-tail portion is magnified in the lower-right corner.

From Seo et al. (2006)

# EnsPost: Findings

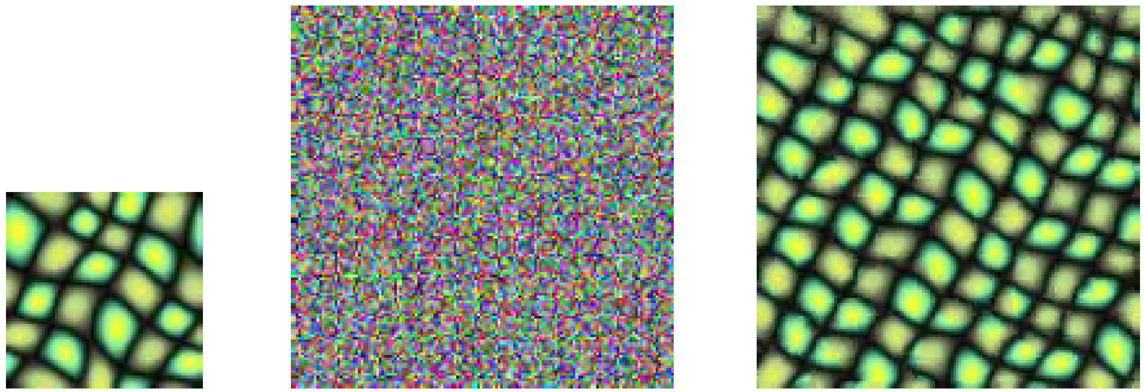
- Performance of the ESP post-processor is sensitive to data availability
- If long-duration data is available, the post-processor performs as expected (correct model biases, produce reliable ensemble traces)
- The ESP post-processor does not handle regulated flows very well
- Storm typing/stratification/conditioning is necessary to handle disparate events (e.g. rain-on-snow)

# Post-processing streamflow ensembles: Attribution

- Resolution
- Reliability (i.e. bias in probability)
  - At the model-native time step/scale (e.g. hourly)
    - Univariate error modeling
- Scaling
  - Reliability at other time scales of aggregation (e.g. daily, weekly, monthly, seasonal, annual)
    - Not an issue if scaling holds $f(cx) = a(cx)^k = c^k f(x) \propto f(x)$.
    - Multivariate (in time) error modeling

# Some ideas from image processing

## 2.1 Wei-Levoy algorithm



(a) (b) (c)

Figure 2.1: Illustration of the synthesis input and output. The texture sample (a) is used to transform the noise (b) in the image (c).

From ECOLE POLYTECHNIQUE, PROMOTION X-98, RAPPORT DE STAGE D'OPTION SCIENTIFIQUE, Paul BILLAULT (2001)

In: Proceedings of the First IEEE Workshop on Internet Vision (held in conjunction with CVPR), 2008.

# Super-Resolution Texturing for Online Virtual Globes


Diego Rother
University of Minnesota
diroth@umn.edu

Lance Williams
Google, Inc.
lancew@google.com

Guillermo Sapiro
University of Minnesota
guille@umn.edu

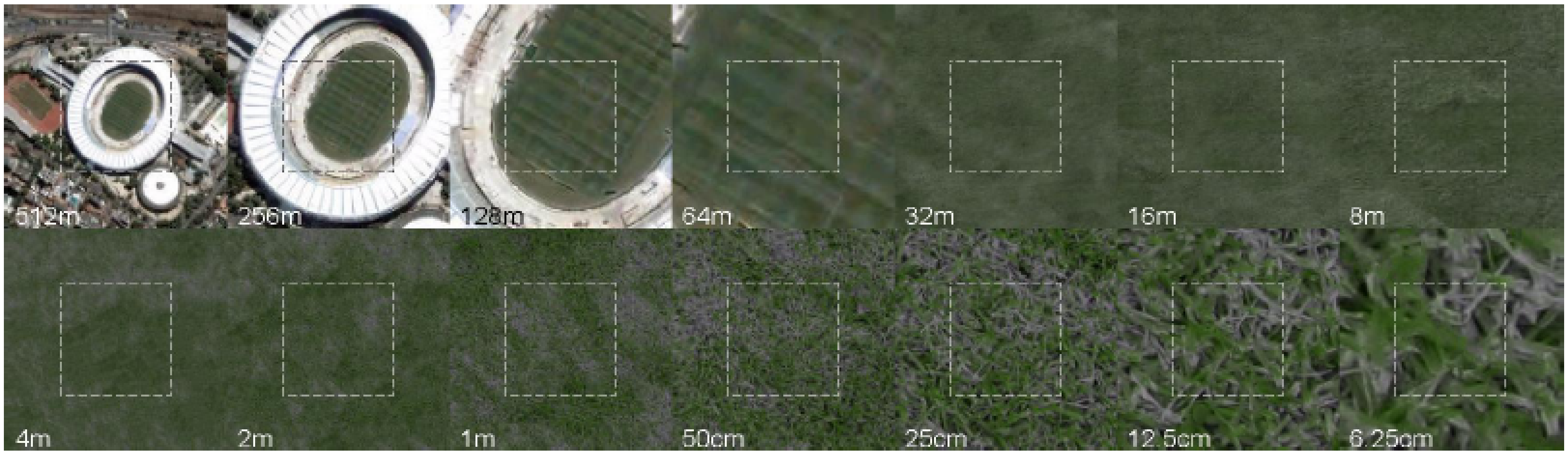




**Figure 1**: "Powers of two." This sequence of images illustrates the framework proposed in this article to supplement and enhance the imagery of virtual globe applications (e.g., *Google Earth*). The first four images were extracted from *Google Earth*, the rest were synthesized with the proposed framework following the user's zoom-in request. See the corresponding video at [28].

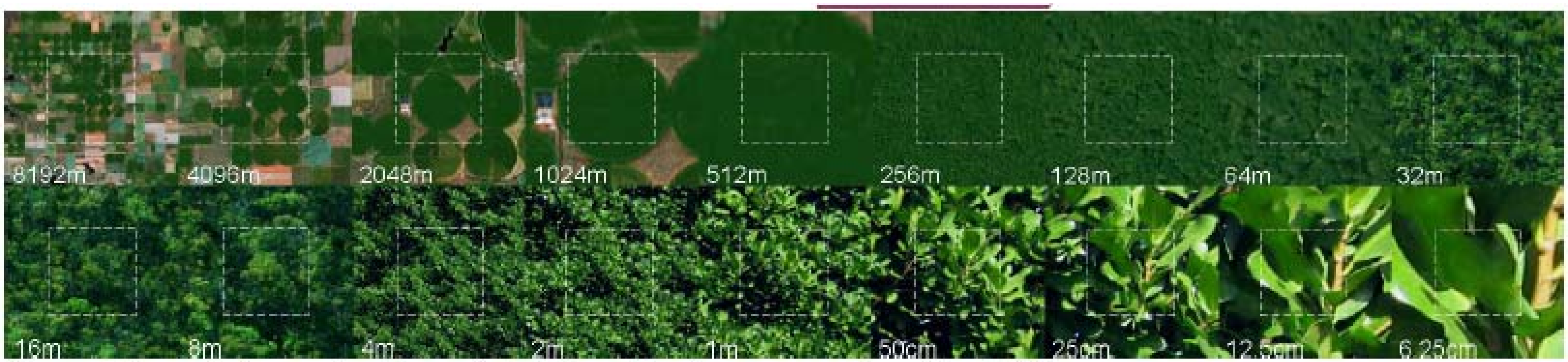
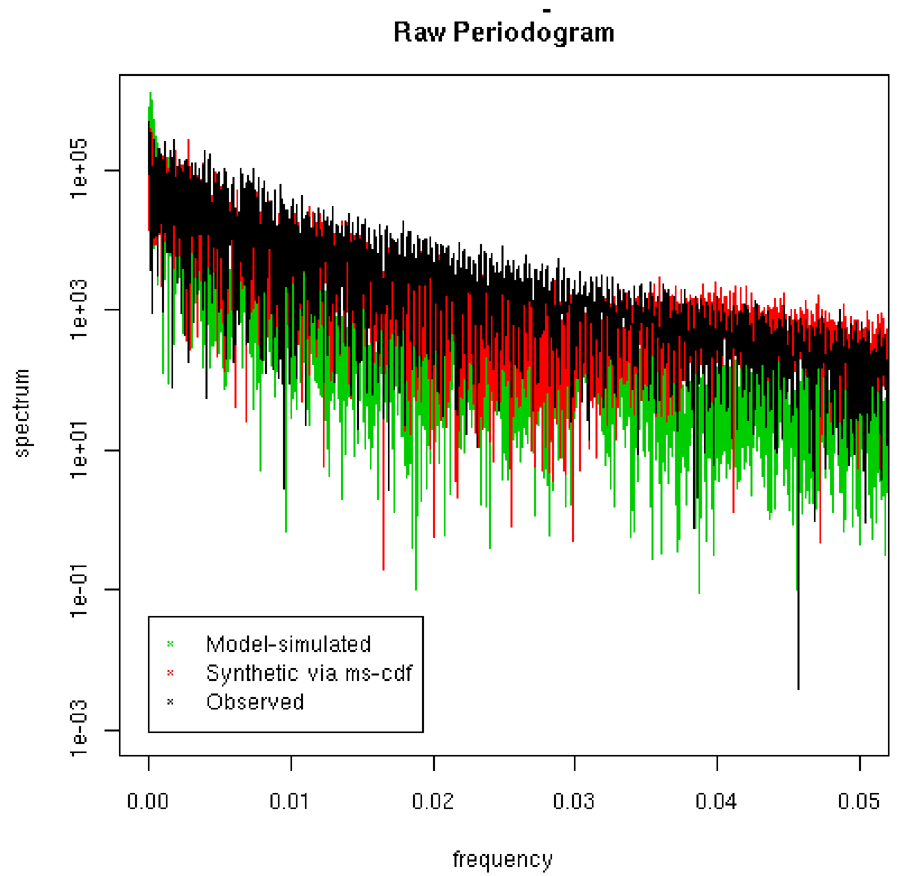
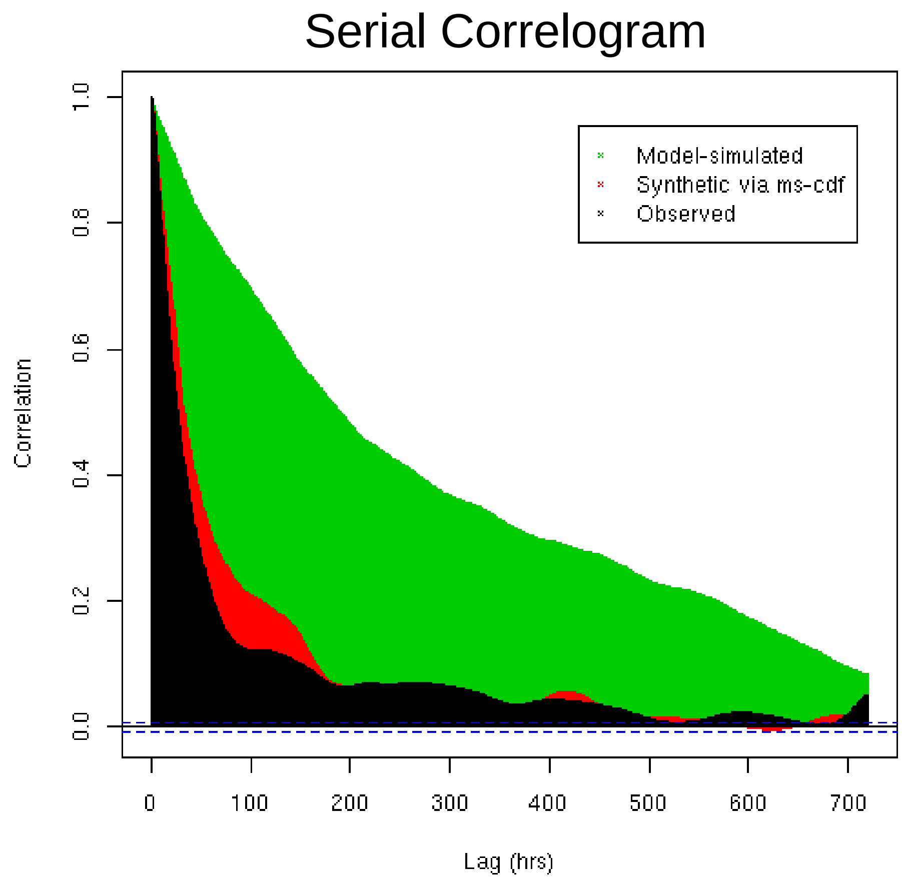



**Figure 8**: Example in a rural setting, a field in Iowa. See the corresponding video at [28].

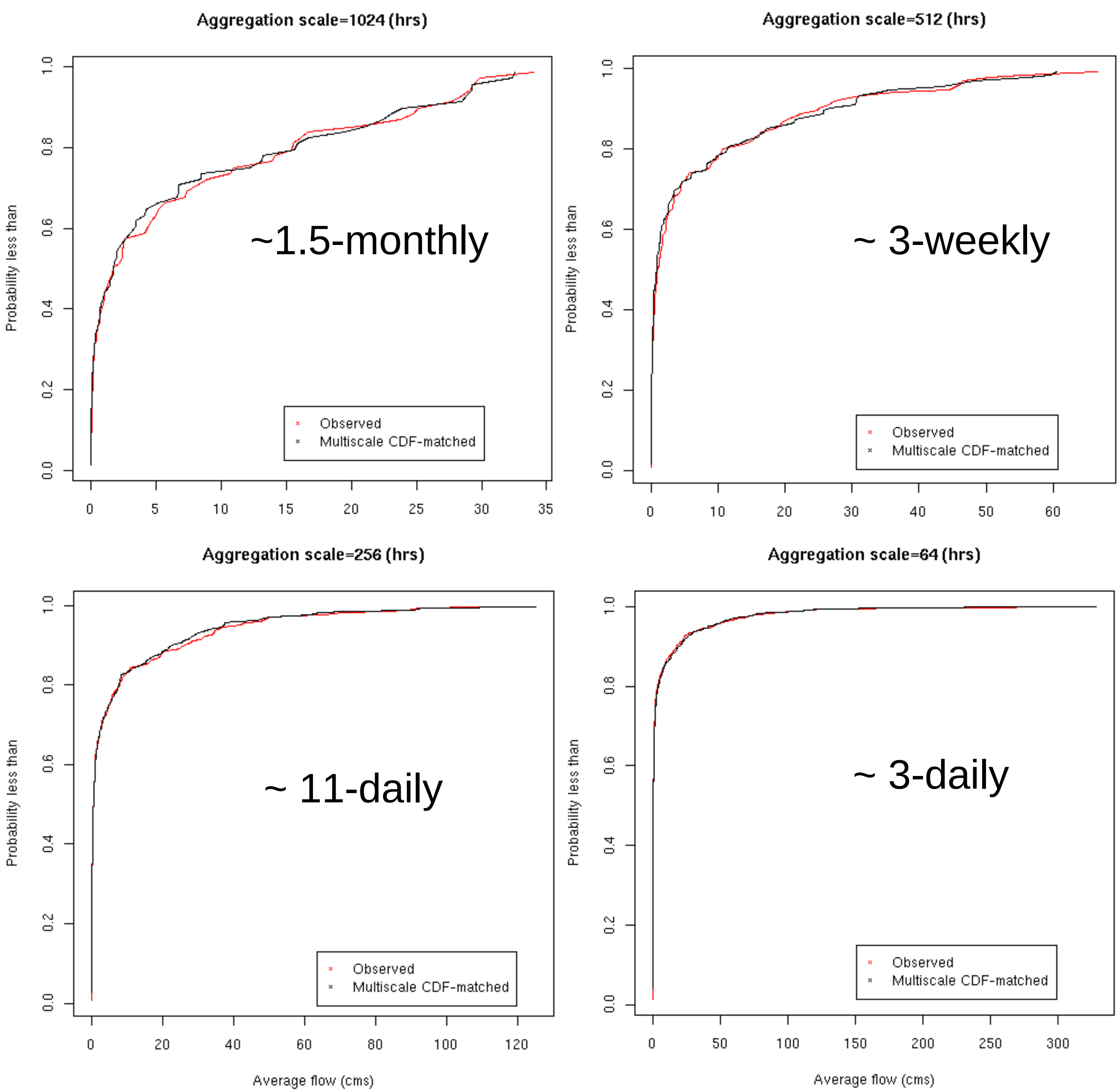
# A proposed procedure for scaling (to make weekly, monthly, seasonal, etc., flows reliable)

- Generate a temporally correlated streamflow ensemble (e.g. in the normal space, then back-transform)
- Apply probability matching from large to small temporal scales of aggregation
  - 1, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64, 96, 128, 192, 256, 384, 512, 768, 1024, 1536, 2048 (hrs)
- Do kernel smoothing

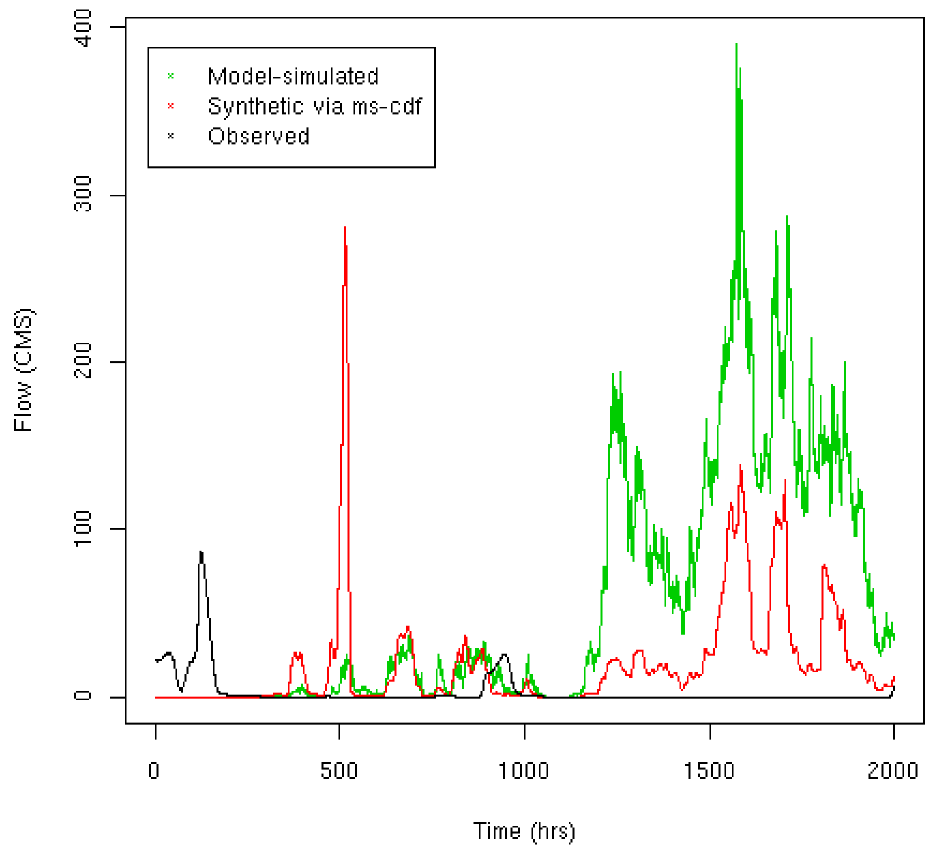


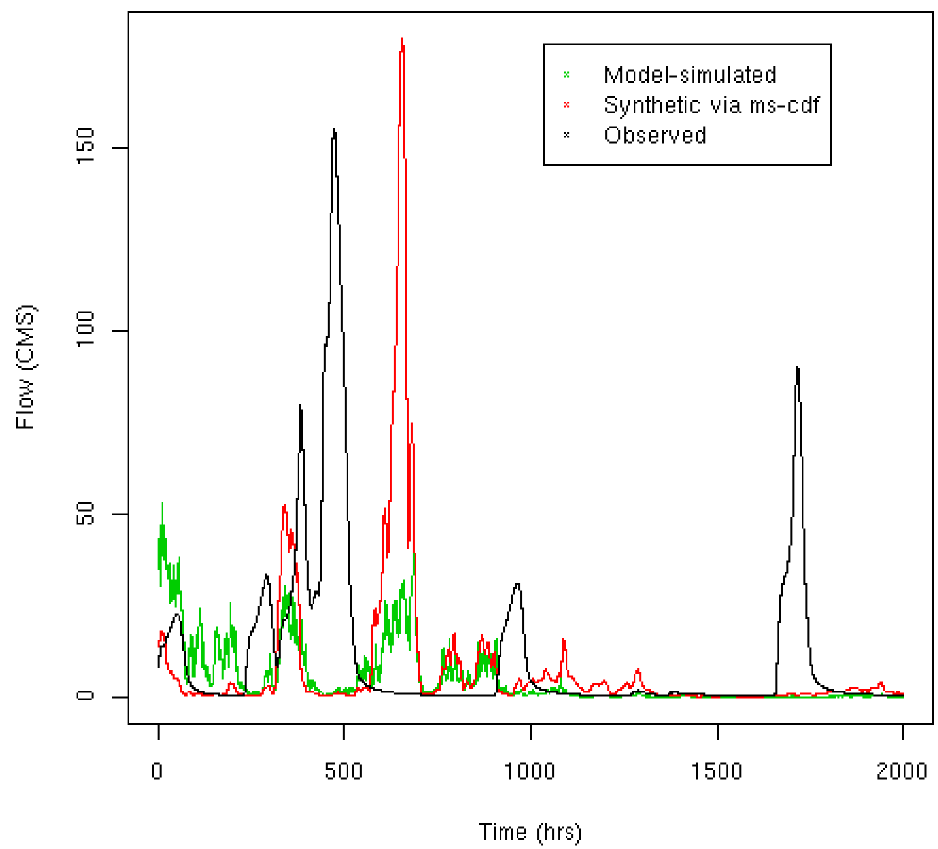




Streamflow ensembles post-processed via multiscale probability matching closely reproduce statistics of the observed flow




Streamflow ensembles post-processed via multiscale probability matching closely reproduces (marginal) probability distributions of the observed flow over a range of temporal scales of aggregation




Streamflow ensembles post-processed via multiscale probability matching resemble the observed flow reasonably closely

# Thank you

## Q/A, Discussion