Overview of Hydrologic Ensemble Prediction Activities at the Office of Hydrologic Development

Julie Demargne[1,2], Limin Wu[1,3], James Brown[1,2], Satish Regonda[1,4], Yuqiong Liu[1,4], and Haksu Lee[1,2]

Acknowledgements to D.J. Seo and John Schaake

1. NOAA/NWS Office of Hydrologic Development
2. University Corporation for Atmospheric Research
3. TCoombs & Associates, LLC
4. Riverside Technology, Inc.

Riverside Technology, inc.
Water Resources Engineering and Consulting

December 2, 2010

TCoombs & Associates, LLC

Main objective

- Develop a hydrologic ensemble forecast system to account for all major sources of uncertainty and communicate uncertainty-quantified forecast and verification information to end users
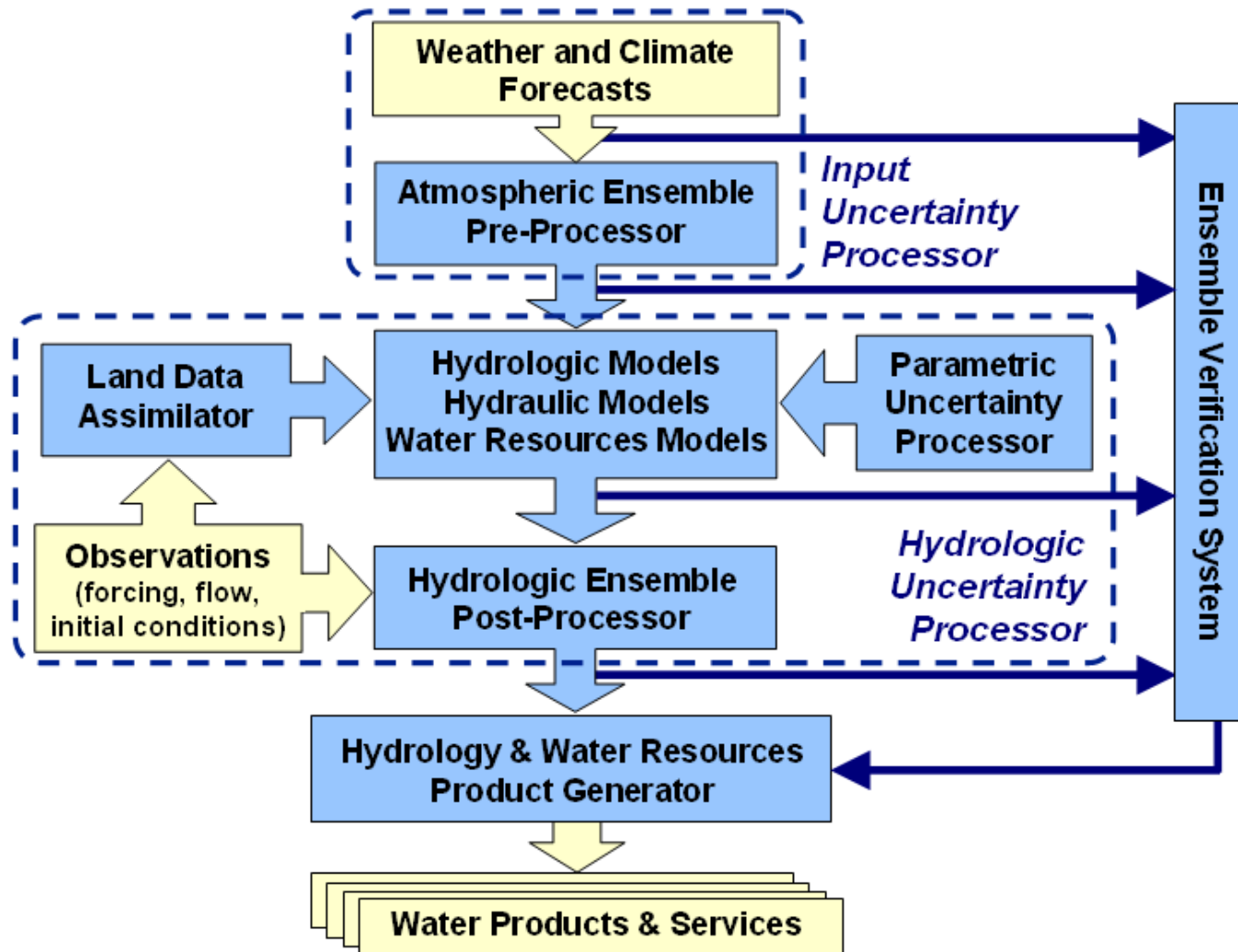
Hydrologic Ensemble Forecast Service (HEFS)
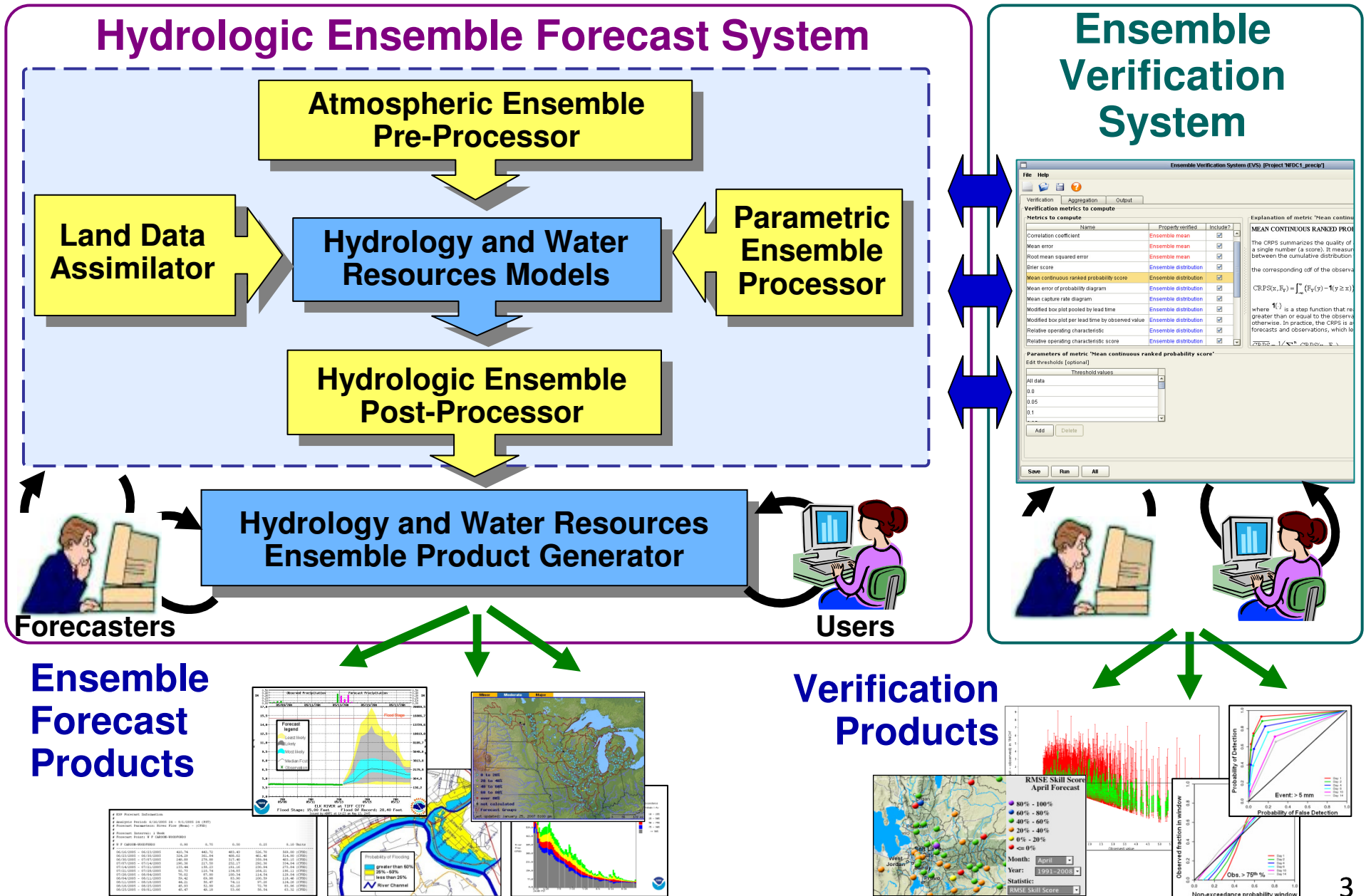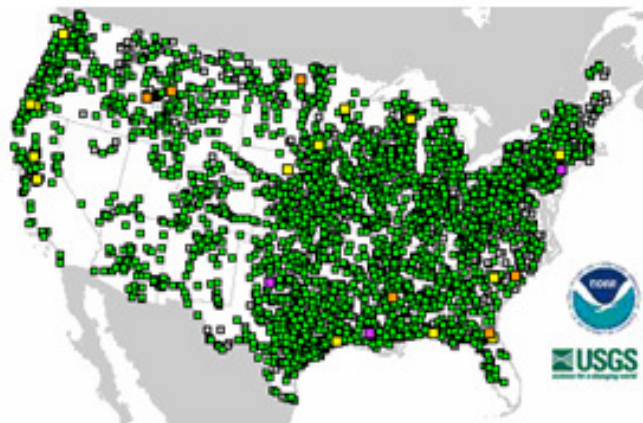
Towards Water Resources Service
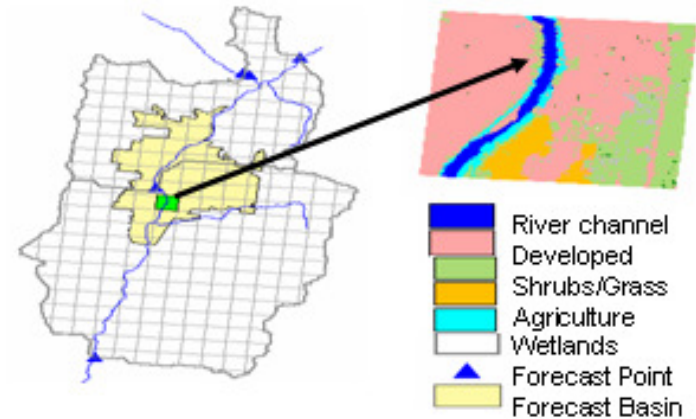## Predicting Floods to Droughts
In Your Neighborhood

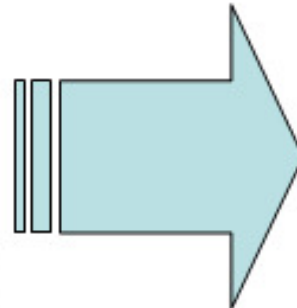**River Services**
(600 miles per forecast point)

River Conditions

Water Resource Services
(6 square mile forecast basins)

- River channel
- Developed
- Shrubs/Grass
- Agriculture
- Wetlands
- Forecast Point
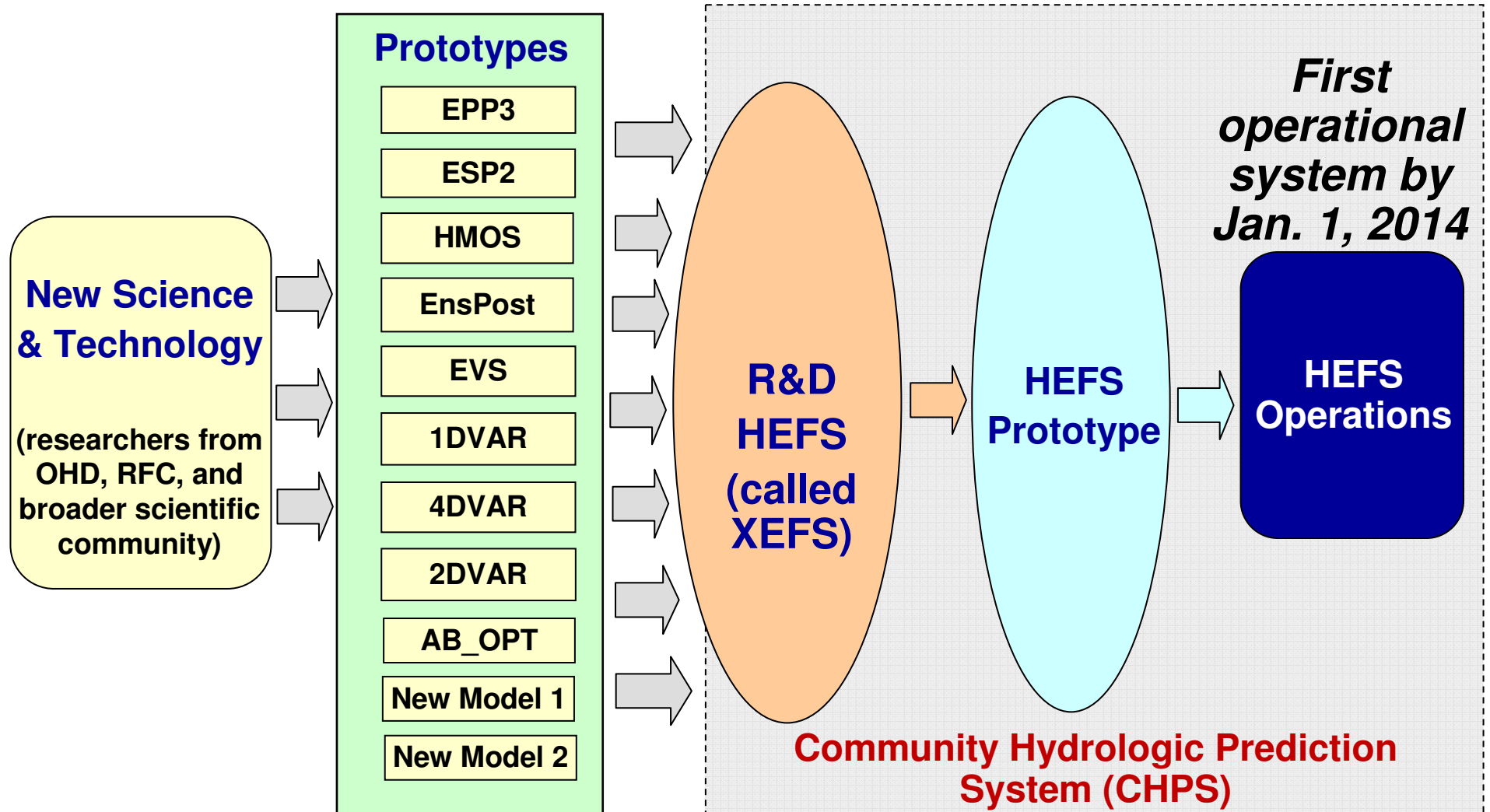- Forecast Basin
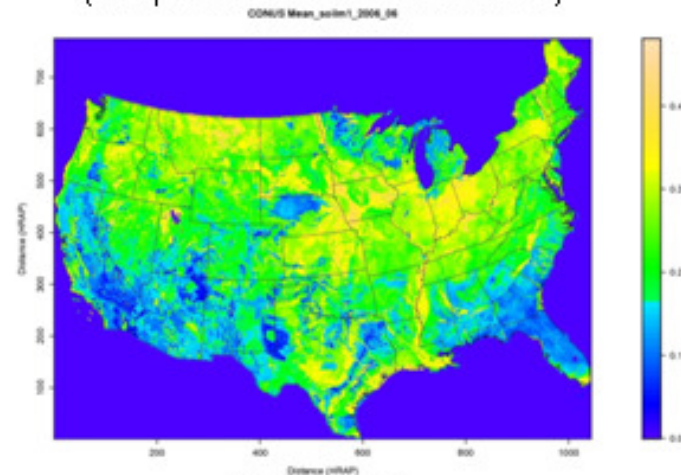
Soil Conditions

From Carter (2006)

4

Ensemble Prediction:
Current (Seasonal) Ensemble Streamflow Prediction vs. HEFS

Feature	Current	HEFS
Platform	National Weather Service River Forecast System (NWSRFS) (inflexible, outdated)	Community Hydrologic Prediction System (CHPS) (flexible, open architecture)
Forecast horizon	Weeks to seasons	Hours to years
Input forecasts	Climate outlook forecasts	Short-, medium- and long-range forecasts (HPC/RFC, GFS/GEFS, CFS/CFSv2, SREF, others)
Hydrologic uncertainty	Not addressed	Addressed (but w/ room for improvement)
Products	Limited number of graphical products	A wide array of user-tailored products via Web-enabled interactive toolbox

Collaborative R&D and RTO for HEFS

New Science & Technology

(researchers from OHD, RFC, and broader scientific community)

Prototypes
- EPP3
- ESP2
- HMOS
- EnsPost
- EVS
- 1DVAR
- 4DVAR
- 2DVAR
- AB_OPT
- New Model 1
- New Model 2

R&D HEFS (called XEFS)

HEFS Prototype

First operational system by Jan. 1, 2014

HEFS Operations

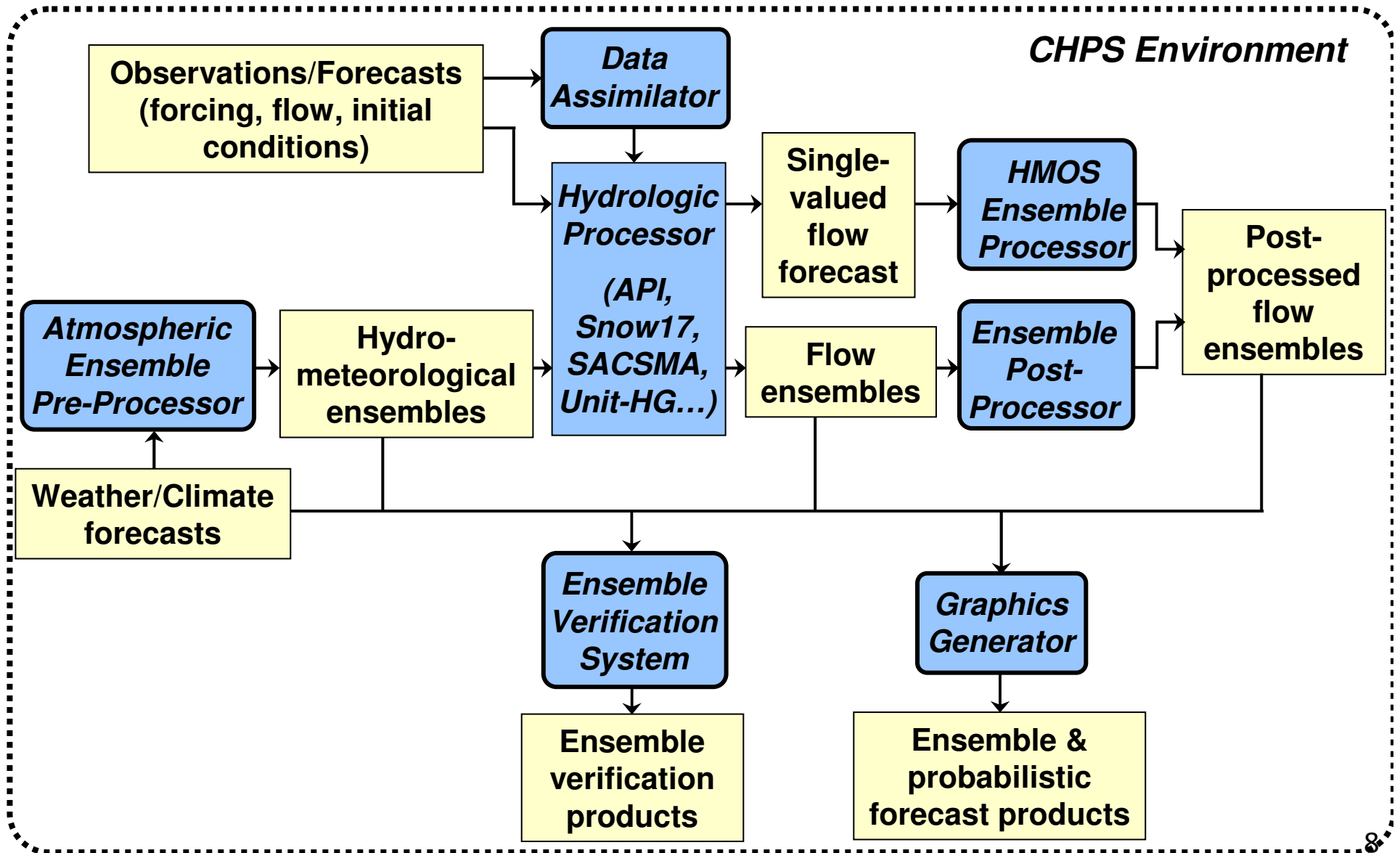**Community Hydrologic Prediction System (CHPS)**
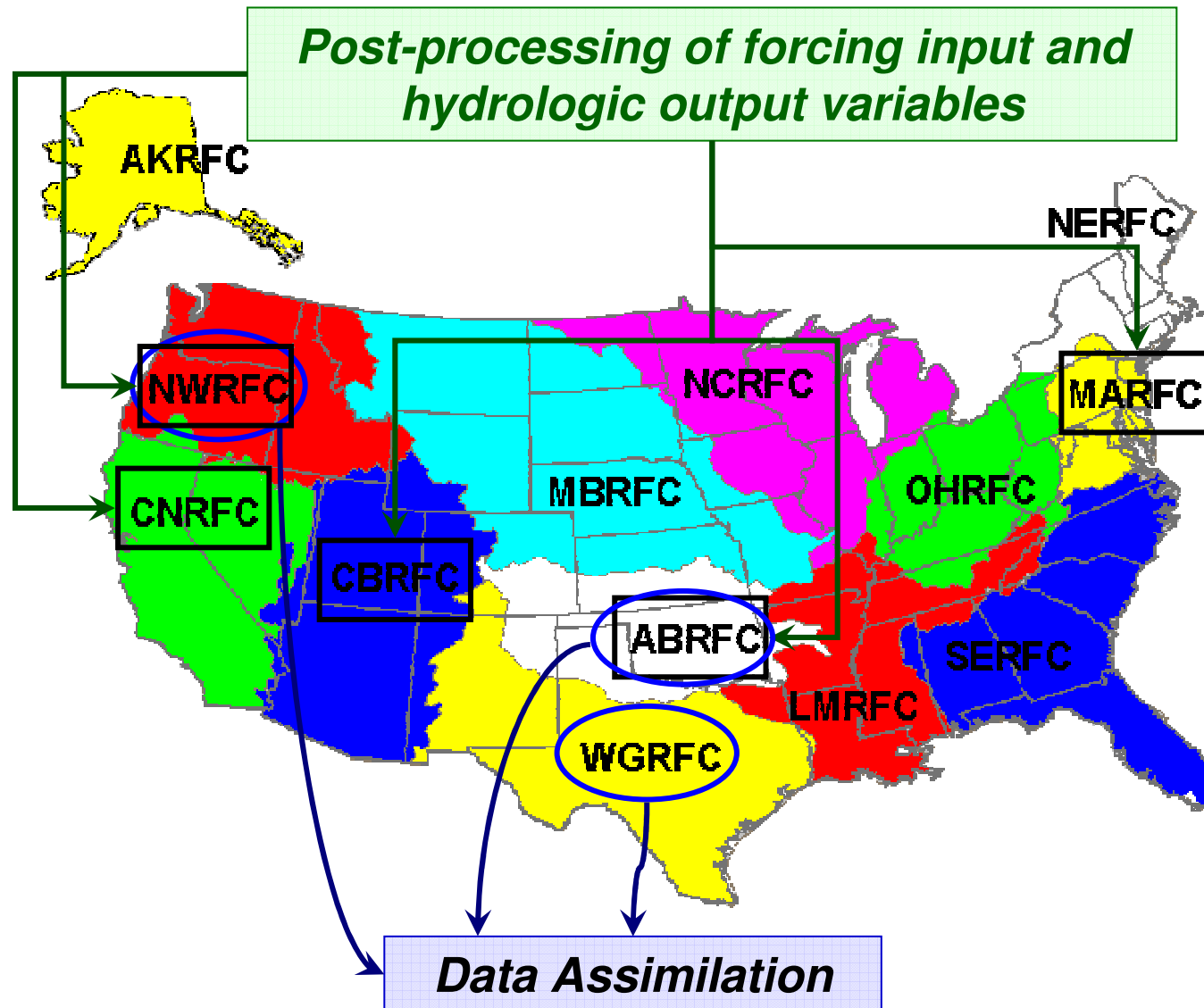
Current R&D Activities

- **Integrated eXperimental Ensemble Forecast System (XEFS)** (R&D version of HEFS) interfaced with CHPS: under evaluation

- **Atmospheric Ensemble Pre-Processing** to produce forcing input ensembles at the basin scale using available weather/climate forecasts (single-valued and probabilistic) and multiple post-processing techniques

- **Hydrologic Ensemble Post-Processing** to produce hydrologic ensembles based on multiple post-processing techniques that make use of single-valued and ensemble hydrologic forecasts at the basin scale

- **Ensemble Verification** to include verification metrics and products that support model developers, RFC forecasters, and end users for diagnostic verification and real-time verification purposes

- **Data Assimilation** to generate optimal initial states and produce improved ensemble snowmelt and flow forecasts using either deterministic or ensemble data assimilation algorithms to account for uncertainty in hydrologic models (e.g., model structure, parameters, and states) and observational data (e.g., model forcing and output)

Experimental Ensemble Forecast System (XEFS)

Interfaced with Community Hydrologic Prediction System (CHPS)

Current XEFS Test Basins
at NWS River Forecast Centers

Post-processing of forcing input and hydrologic output variables

AKRFC

NERFC

NWRFC

CNRFC

CBRFC

MBRFC

NCRFC

OHRFC

MARFC

ABRFC

SERFC

LMRFC

WGRFC

Data Assimilation

Uncertainty integration in XEFS

Uncertainty in hydrologic forecast

=

Uncertainty in future forcings **+** **Uncertainty in everything else**

precipitation,
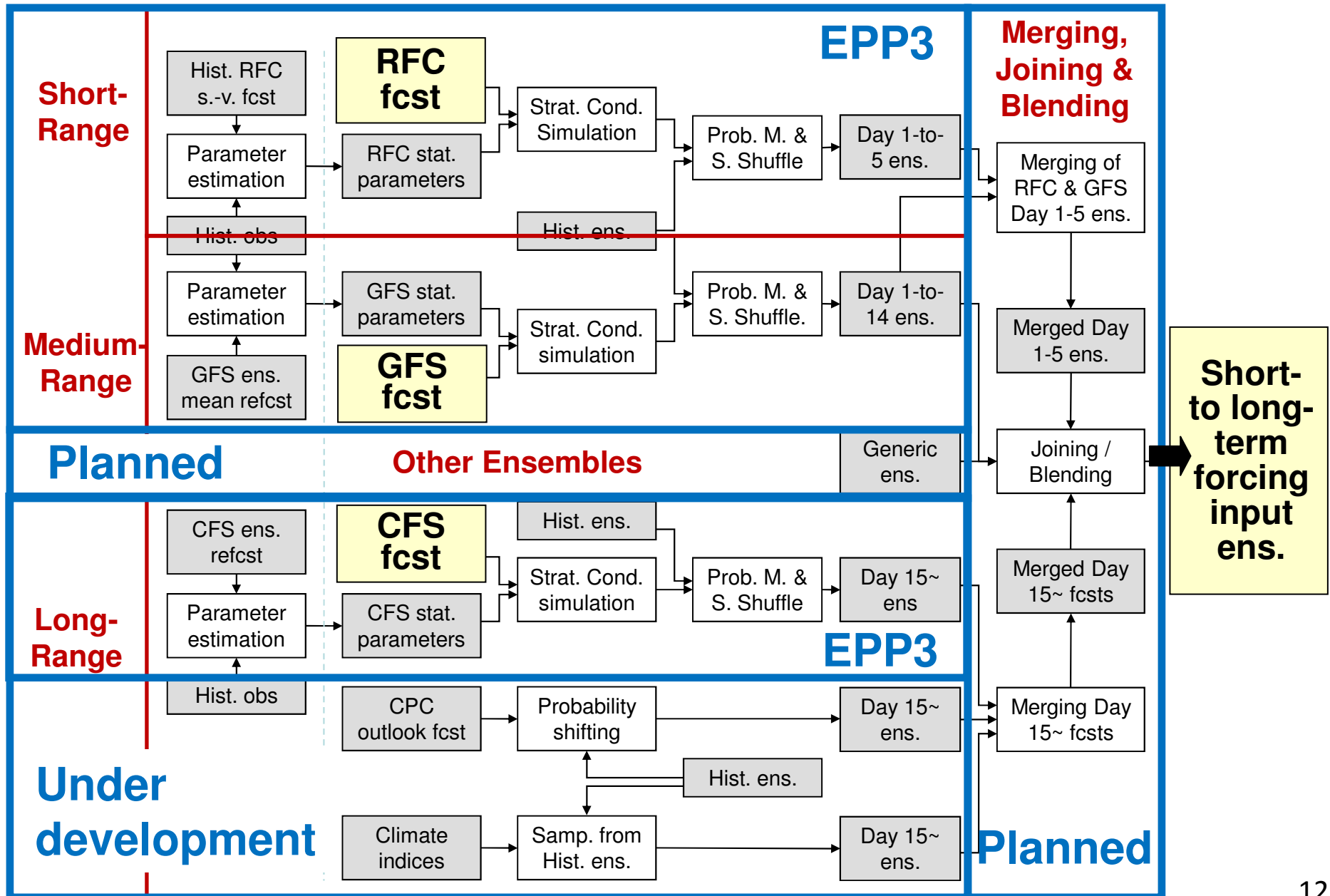temperature,
potential evaporation

initial soil moisture conditions,
model errors,
observation errors,
etc.

- Reliable and skillful (uncertainty-integrated) hydrologic ensemble forecast requires

  - Reliable and skillful ensemble forecast of forcing variables
  - Accurate modeling and accounting of hydrologic and hydraulic uncertainties

Ensemble Pre-Processor Strategy

- Current situation
  - NWP ensembles are generally biased in the mean and spread
  - For short range, HPC/RFC forecasts are generally more accurate than NWP ensemble forecasts in some mean sense
- Initial build (EPP3)
  - Given single-valued forecast (HPC/RFC single-valued QPF, GFS ensemble mean from frozen GFS, CFS ensemble mean), statistically generate ensembles based on joint probability distribution between observed and forecast precipitation
- Future builds
  - Use new CFSv2 and GEFS datasets
  - Bring in post-processed (bias-corrected and downscaled) multi-model NWP ensembles
  - Include potential evaporation
  - Include other post-processing techniques

11

Envisioned Ensemble Pre-Processor

EPP3

Merging, Joining & Blending

Short-Range

Hist. RFC s.-v. fcst → Parameter estimation → RFC stat. parameters

RFC fcst

Strat. Cond. Simulation → Prob. M. & S. Shuffle → Day 1-to-5 ens.

Hist. obs

Hist. ens.

Medium-Range

Parameter estimation → GFS stat. parameters

GFS ens. mean refcst

GFS fcst

Strat. Cond. simulation → Prob. M. & S. Shuffle. → Day 1-to-14 ens.

Merging of RFC & GFS Day 1-5 ens.

Merged Day 1-5 ens.

Planned

Other Ensembles

Generic ens.

Joining / Blending

Long-Range

CFS ens. refcst → Parameter estimation → CFS stat. parameters

CFS fcst

Hist. ens.

Strat. Cond. simulation → Prob. M. & S. Shuffle → Day 15~ ens

Merged Day 15~ fcsts

EPP3

Short-to long-term forcing input ens.

Under development

Hist. obs

CPC outlook fcst → Probability shifting → Day 15~ ens.

Hist. ens.

Climate indices → Samp. from Hist. ens. → Day 15~ ens.

Merging Day 15~ fcsts

Planned

12

Strategy for hydrologic uncertainty modeling

- Current approach

  - EnsPost: lump all hydrologic uncertainties into one and model it stochastically (Seo et al. 2007)

  - HMOS: model total uncertainty (input + hydrologic) in single-valued operational flow forecasts using QPF information

- Future approach

  - Include other statistical post-processing techniques
  - Uncertainty modeling of regulated flows
  - Initial condition uncertainty via ensemble data assimilation
  - Parametric uncertainty via parametric uncertainty processor
  - Multi-model ensembles for structural uncertainty

Example of verification results: XEFS flow ensembles compared to climatological ESP ensembles

Skill Score for Mean CRPS (CRPSS):

GFS-based flow generated w/ pre- and post-processing compared to

➢ GFS-based flow w/o post-processing

➢ climatology-based flows (operational ESP) (w/o pre- and post-processing)

▪ Very large improvement w/ pre-processing and GFS ens. means over climatological ESP

▪ Significant improvement w/ post-processing

Higher scores: better

Ref=EPP: Obs. > 95th %
Ref=Clim: Obs. > 95th %

gain from EPP using GFS ensemble means

gain from EnsPost

CRPSS

Lead time (days)

14
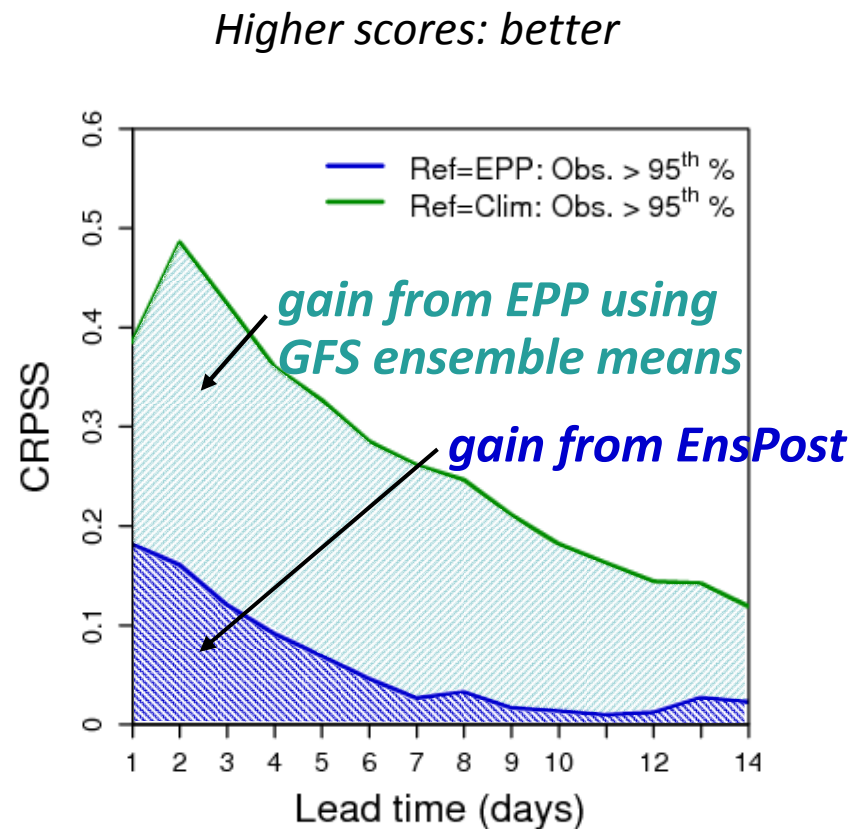
Strategy for data assimilation

- Current approach:

  - develop DA tools with single-valued forecasts and extend to ensemble DA

  - interface them into OpenDA for CHPS application

- Techniques currently tested

  - Hydrologic routing DA (1DVAR) to assimilate streamflow into 3-parameter Muskingum routing model

  - 2DVAR to assimilate streamflow, precipitation and PE for Sacramento and Unit Hydrograph models (implemented)

  - Snow and streamflow DA (including different techniques, e.g., Ensemble Kalman Filter and Maximum Likelihood Ensemble Filter)

  - 4DVAR to assimilate streamflow, soil moisture, precipitation and PE for gridded SAC and kinematic-wave routing in Research Distributed Hydrologic Model (RDHM)

Ensemble Verification System

- Currently Available Features
  - Java tool with structured GUI
  - Only verifies numerical time-series for individual points/areas
  - Supports flexible conditional verification
  - Computes several key metrics (deterministic and ensemble) for reliability, resolution, discrimination, skill

- Status
  - EVS version 3.0 released in October 2010 www.nws.noaa.gov/oh/evs.html
  - Enhancements underway (e.g., estimation of confidence intervals)
  - Fully documented and freely available

Operational hydrologic ensemble forecasting Challenges

- Appropriately model and integrate uncertainties introduced from data, model, and human sources

- Combine ensemble forcing for short, medium, and long ranges from multiple sources

- Maintain spatio-temporal relationships across different scales

- Include forecaster skill in short-term inputs (QPF, temperature, etc.)

- Include forecaster guidance of hydrologic model operation

- Maintain coherence between deterministic and ensemble forecasts

- Provide uncertainty information in a form and context that is easily understandable and useful to the customers

- Reduce the cone of uncertainty for effective decision support
  - Improve accuracy of meteorological and hydrologic models

- Improve uncertainty modeling and observations of rare and extreme events (e.g. record flooding, drought without any historical analog)

- Greatly improve computing, database and data storage capabilities

OHD-NCEP Collaborations

- Common areas of expertise

  - Bias correction, statistical downscaling, data assimilation, ensemble verification

  - Inter-comparison of different techniques and verification results

- Coordination on new forecasting systems

  - Hindcast strategy, dataset sharing (reforecasts, forecasts, and analyses), ensemble verification

  - Hydrology requirement: long-term reforecast datasets required for calibration at RFC basin scale

Thank you!

HEP Group Website

http://www.weather.gov/oh/hrl/hsmb/hydrologic_ensembles/index.html

XEFS Website

http://www.weather.gov/oh/XEFS/

Extra Slides

Example: Errors in Climatological Precipitation Forecast (Day 1)
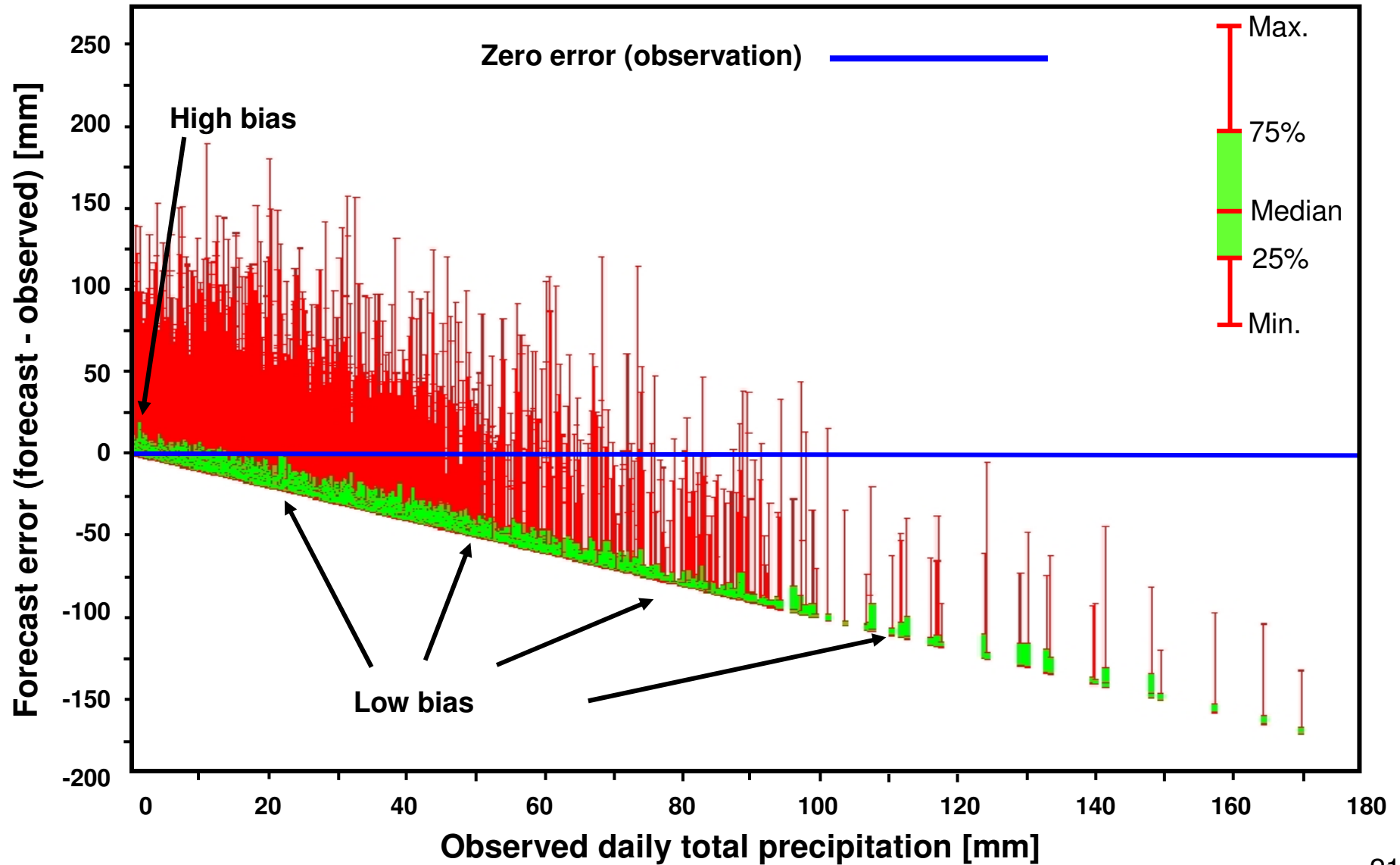
Example: Errors in GFS-based EPP Precipitation Forecast (Day 1)
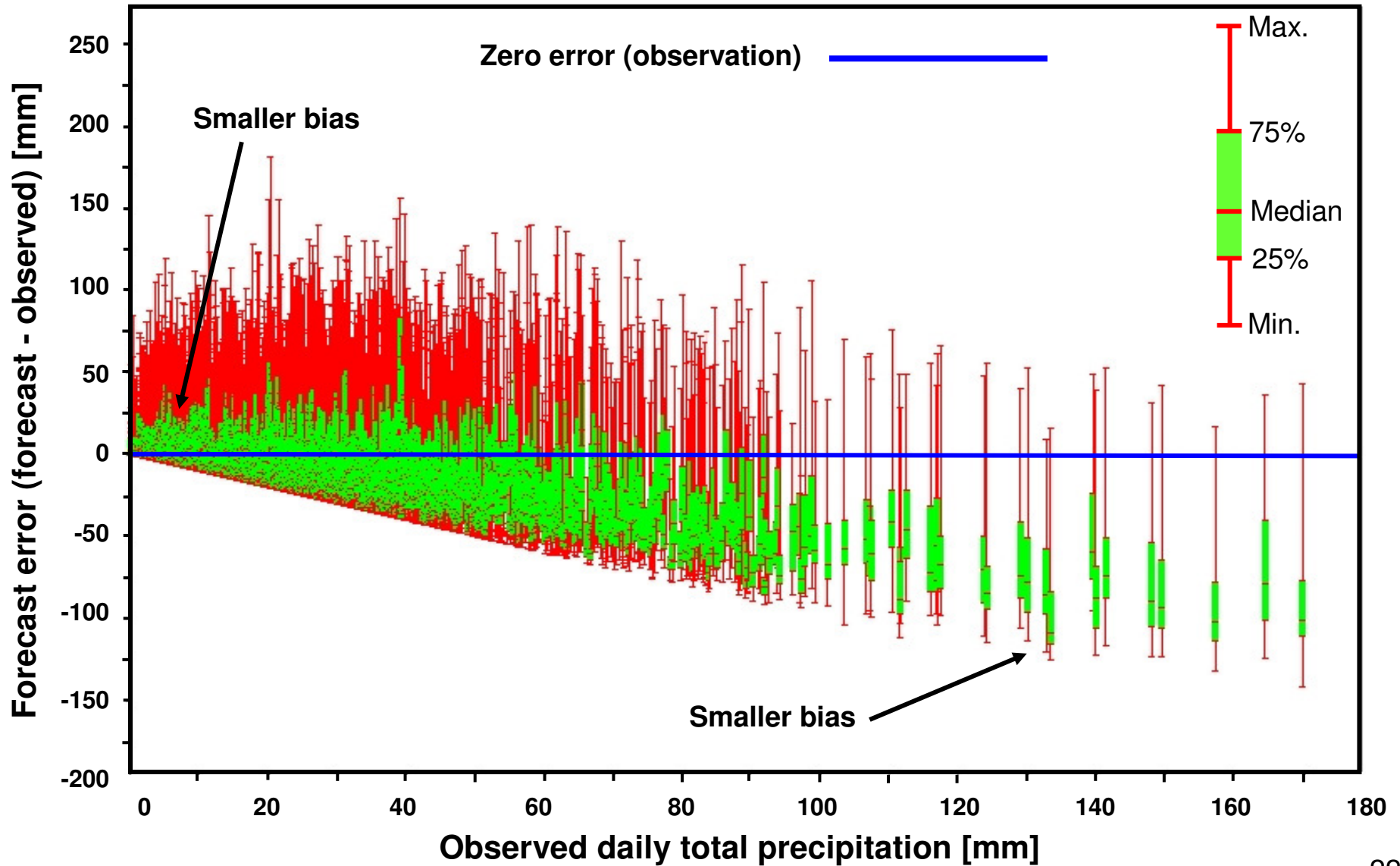
Example: Errors in Climatology-based Flow Forecast (Day 1)
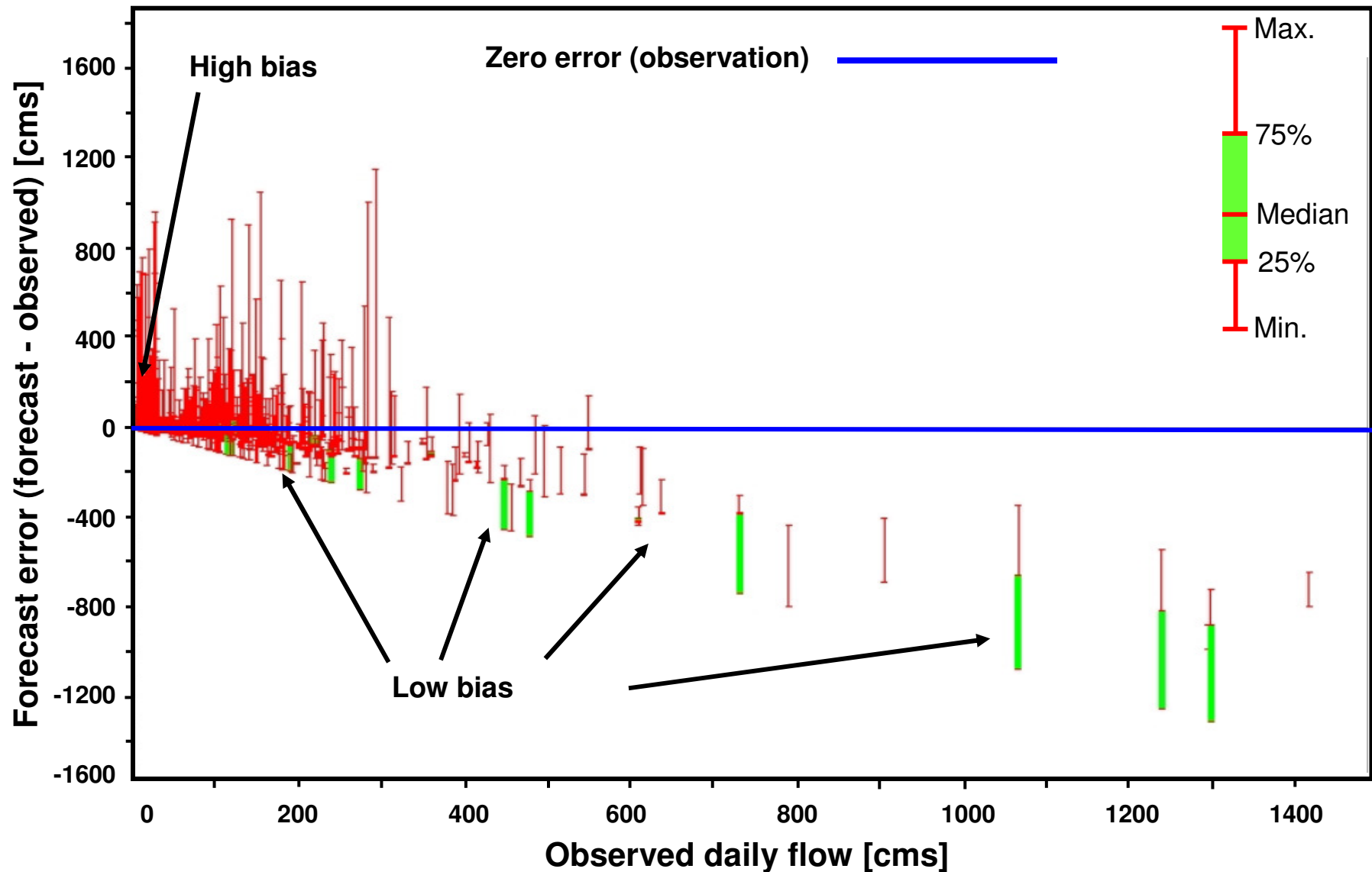
Example: Errors in GFS-based EPP-EnsPost Flow Forecast (Day 1)
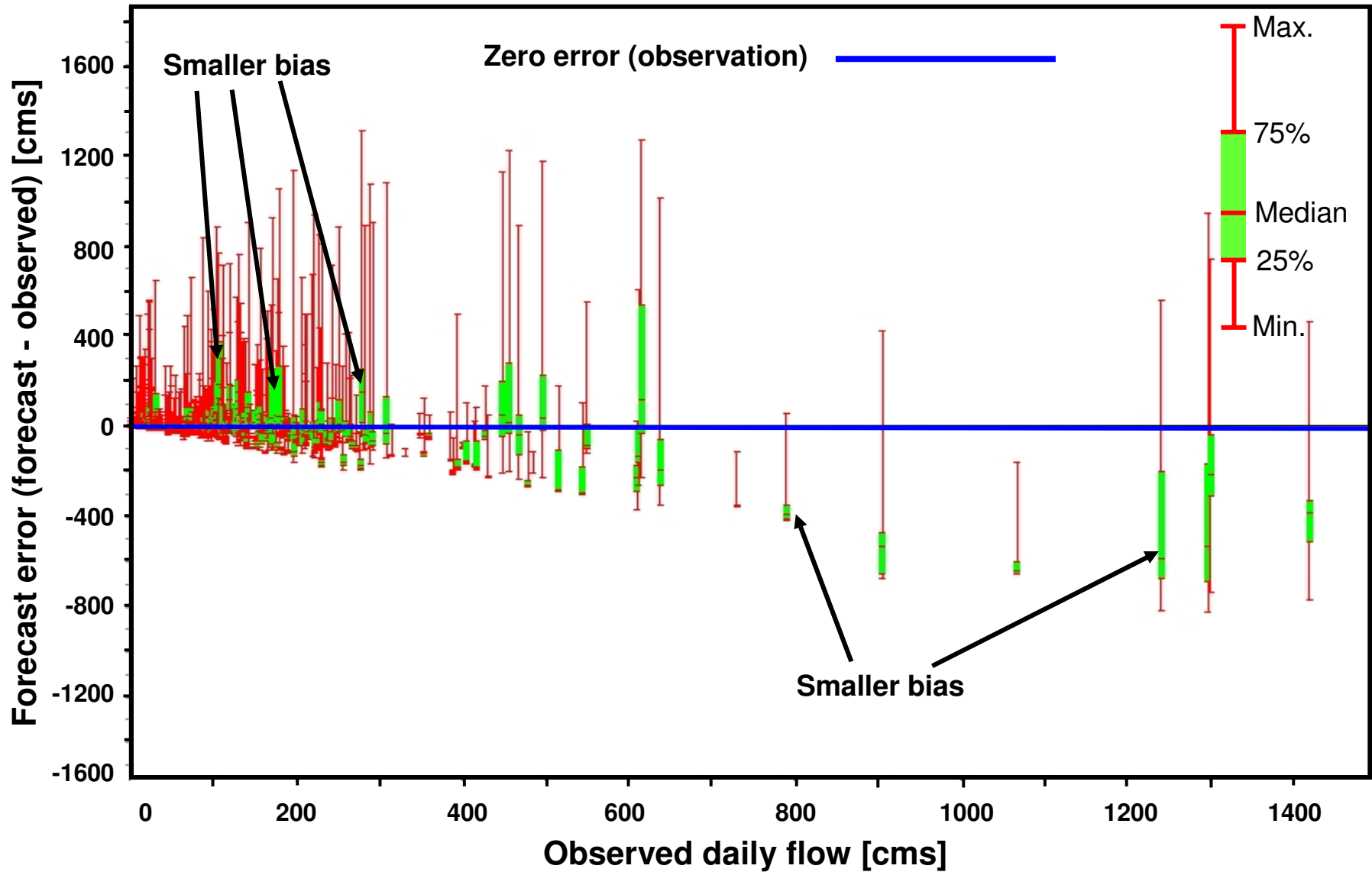
Graphic Generator

- Current build
  - Basic ESPADP functionality for CHPS
  - Incorporated enhanced functionality from RFC feedback
  - Tested in operating client environment

- On-going – move to operations
  - On-demand product sharing across operating clients
  - ESPADP-like default template products
  - Ability to apply a product/template to all forecast points
  - Coordinating ESPADP differences with AHPS
  - Training Manual & webinar (funded by NSTEP)

Envisioned XEFS Products

Reliable and skillful ensembles
(streamflow and inputs)
and associated
meta data
for 1 hour to 2 years

Archive of
hindcasts,
forecasts,
simulations,
observations,
and meta data

Specific data, forecasts,
ensembles, and analysis
for use in _value added_
processes (e.g. flood
inundation mapping)

User selectable attributes
period
duration
time aggregation
probability levels
thresholds
probability type
analogs based on meta data

User selectable context
mean, median, max, min
analogs based on meta data
specific years
specific forecast w/outcome

Verification
information in user
friendly form

Instructional
(annotated examples)

GRAPHICS

BINARY

010101010101
010101010101
010101

TEXT

XML

Uncertainty integration strategy in XEFS

$$f_1(q_f \mid q_o) = \int f_2(q_f \mid q_o, s_f) \ f_3(s_f \mid q_o) \ ds_f$$

Predictive uncertainty in streamflow   **Residual hydrologic uncertainty**   **Uncertainty in model-predicted streamflow**

where q_f Streamflow at some future times

   q_o Observed flow up to and including the current time

   s_f Model-predicted streamflow at the future times

Krzysztofowicz (1999)

$$f_3(s_f \mid q_o) = \iiint f_4(s_f \mid b_f, i, p, q_o) \ f_5(b_f \mid i, p, q_o) \ f_6(p \mid i, q_o) \ f_7(i \mid q_o) \ db_f \ di \ dp$$

Uncertainty in model-predicted streamflow   **Conditional hydrologic model simulation**   **Future forcing uncertainty**   **Parametric uncertainty**   **Initial condition uncertainty**

where b_f Future boundary conditions (precipitation, temperature)

   I   Initial conditions

   p   Model parameters

Seo et al. (2006)

Uncertainty integration strategy in XEFS

w/o data assimilator and parametric uncertainty processor

$$f_1(q_f \mid q_o) = \int f_2(q_f \mid q_o, s_f) \; f_3(s_f \mid q_o) \; ds_f$$

Predictive uncertainty in streamflow   **Residual hydrologic uncertainty**   **Uncertainty in model-predicted streamflow**

where q_f Streamflow at some future times
        q_o Observed flow up to and including the current time
        s_f Model-predicted streamflow at the future times

$$f_3(s_f \mid q_o) = \int f_4(s_f \mid b_f) \; f_5(b_f) \; db_f$$

Uncertainty in model-predicted streamflow   **Conditional hydrologic model simulation**   **Future forcing uncertainty**

where b_f Future boundary conditions (precipitation, temperature)

EPP Calibration Processor

Off line, model joint distribution between single-valued QPF and verifying observation for each lead time (same process used for temperature but with different distributions)
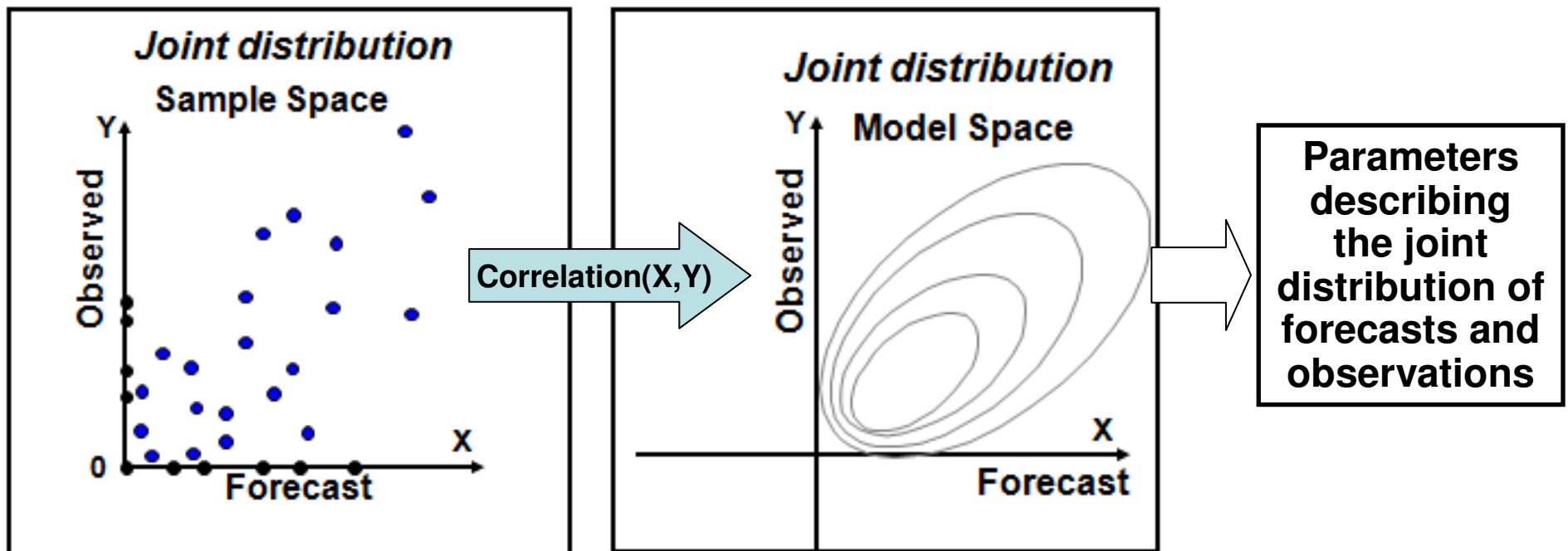
Multi-year archive of single-valued QPF necessary

Schaake et al. (2007), Wu et al. (2010)

EPP Real-Time Processor: Step 1

In real-time, given single-valued QPF, generate ensemble traces from the conditional distribution for each lead time
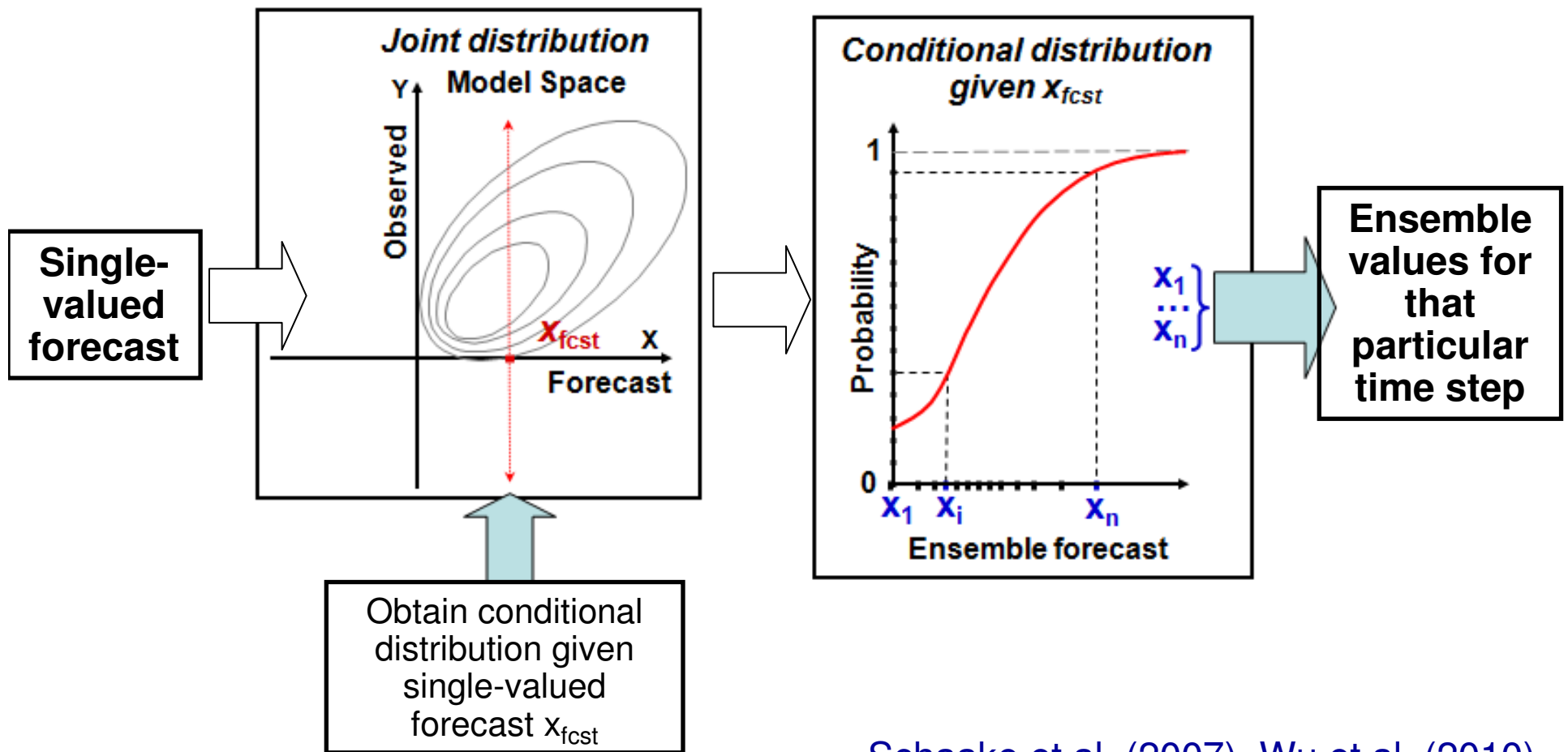
Schaake et al. (2007), Wu et al. (2010)

EPP Real-Time Processor: Step 2

In real-time, string together lead-time specific ensemble values across lead times to generate traces

Timestep 1	Timestep 2	Timestep 3	Timestep 4	
X1 (1992)	X1 (1991)	X1 (1993)	X1 (1991)	Member 1 (1991)
X2 (1991)	X2 (1995)	X2 (1994)	X2 (1995)	Member 5 (1995)
X3 (1993)	X3 (1994)	X3 (1995)	X3 (1994)	Member 4 (1994)
X4 (1994)	X4 (1992)	X4 (1991)	X4 (1992)	Member 2 (1992)
X5 (1995)	X5 (1993)	X5 (1992)	X5 (1993)	Member 3 (1993)
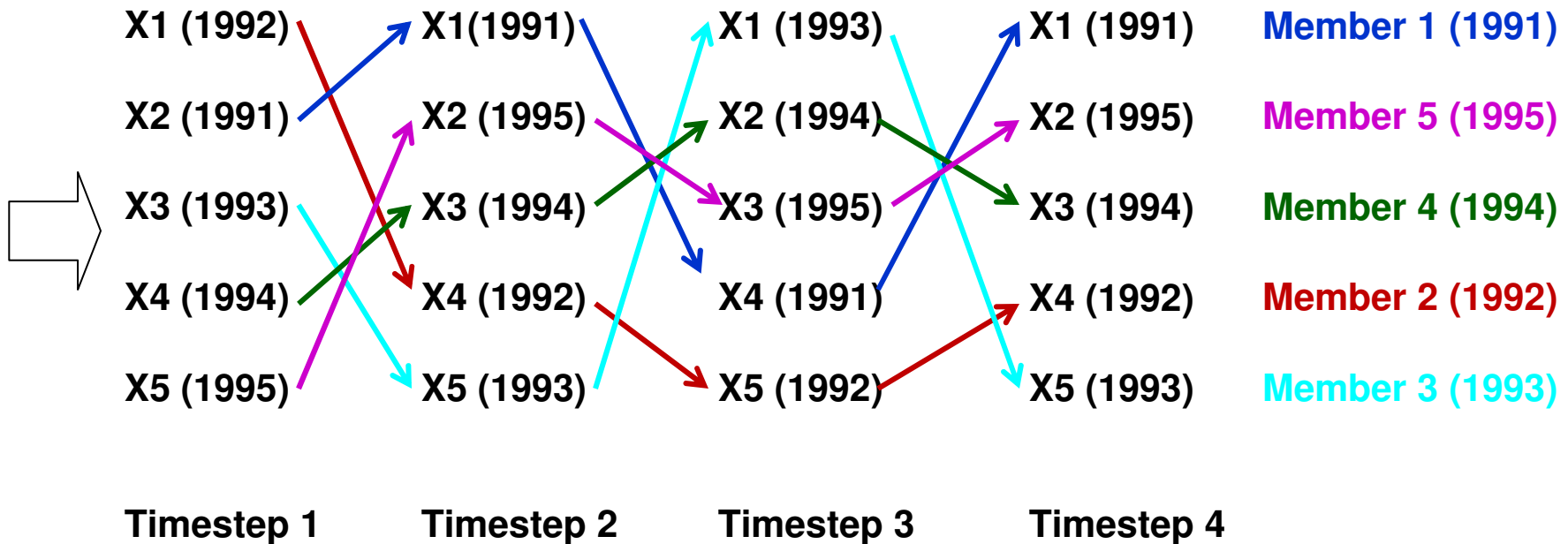
Schaake Shuffle (Clark et al. 2004)

EPP Real-Time Processor: Step 2

In real-time, string together lead-time specific ensemble values across lead times to generate traces

For each time step of the forecast period, arrange precip/temp ensemble members such that they have the same ordering as historical observations
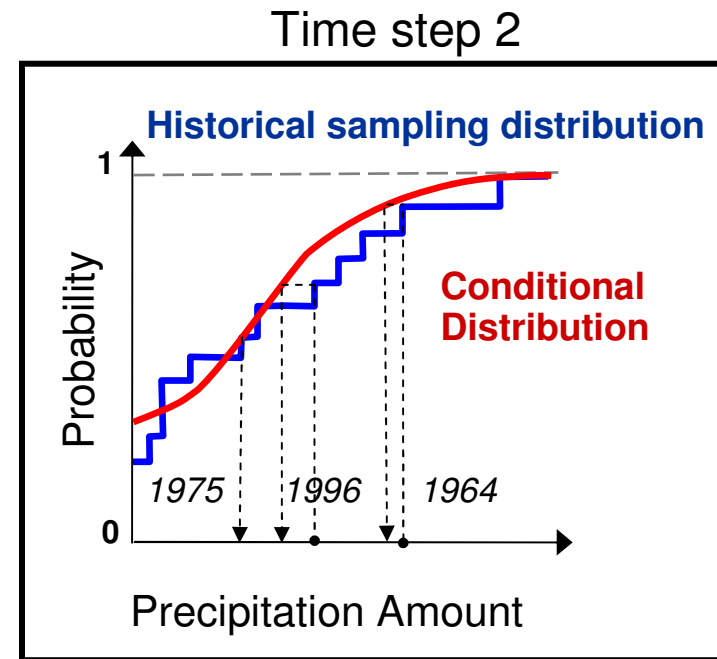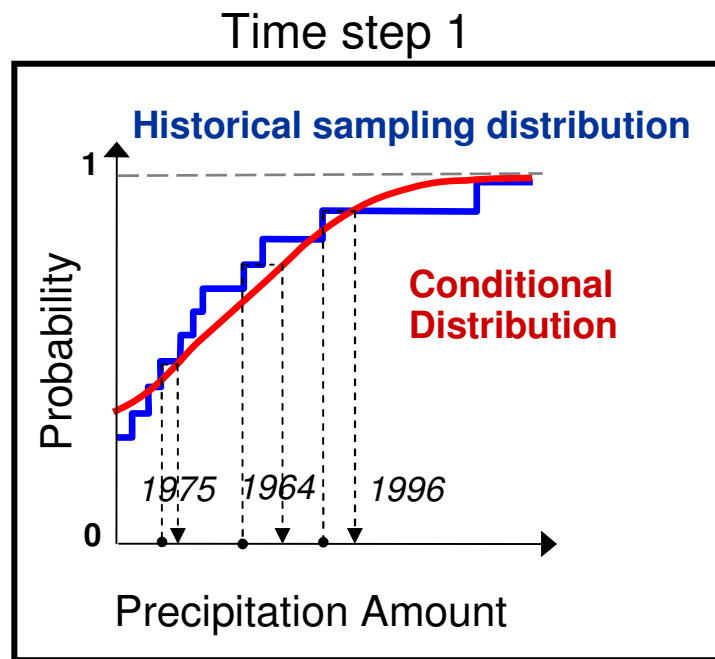
Different illustration of Schaake Shuffle (Clark et al. 2004)

EPP Real-Time Processor: Step 2

In real-time, string together lead-time specific ensemble values across lead times to generate traces

Stratified sampling: n+1 intervals of probability to define n ensemble values

Schaake Shuffle: re-order ensemble values based on historical observations
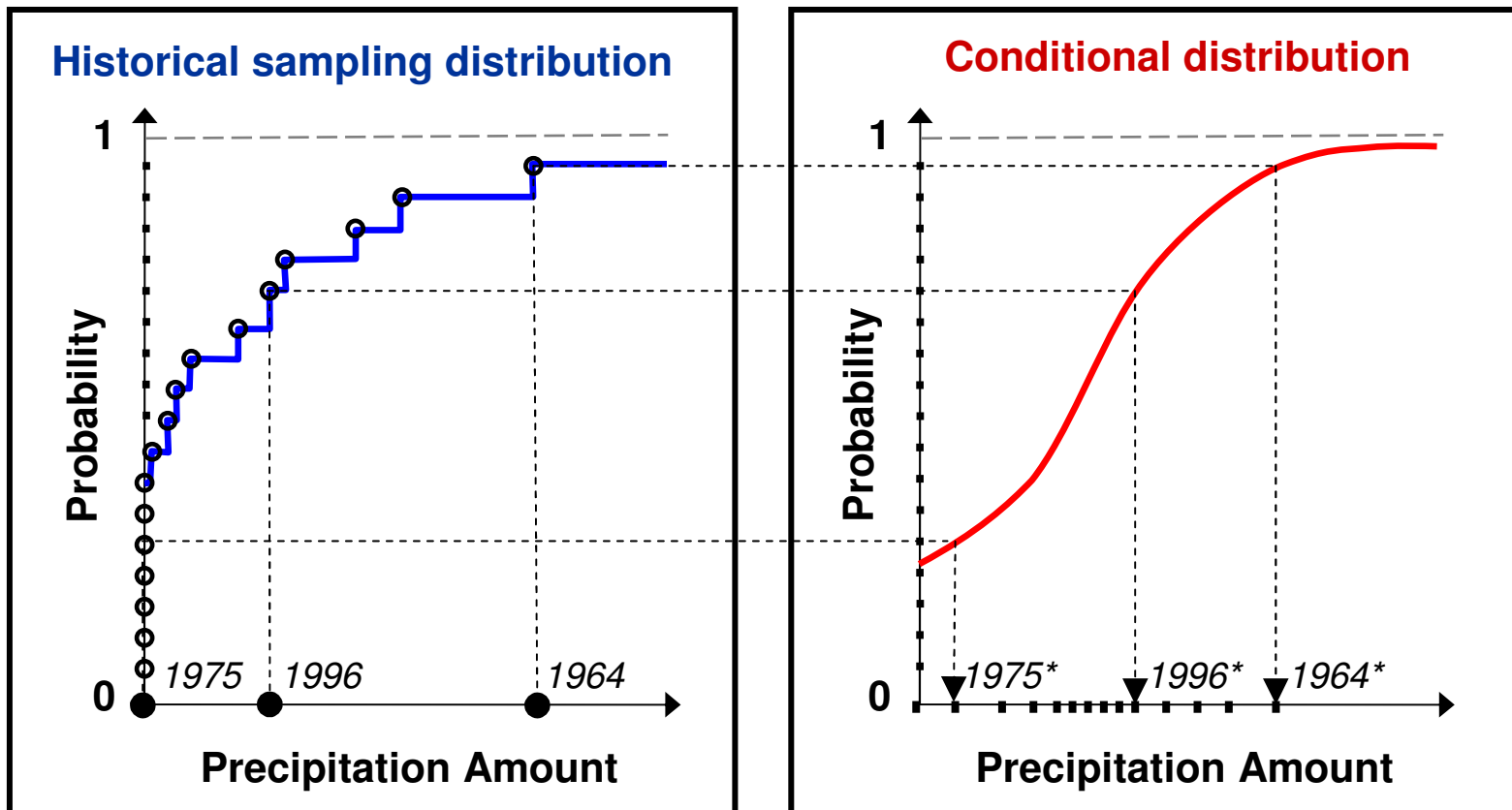
Different illustration of Schaake Shuffle (Clark et al. 2004)

Ensemble Preprocessor: Schaake Shuffle

Climatology used to identify properties of precipitation and temperature in space and time

Example of historical events for a given date

1975: clear + hot

cloudy + warm

clear + warm

cloudy + warm

clear + hot

Low flow

1996: rain + cold

rain + cold

rain + cold

snow
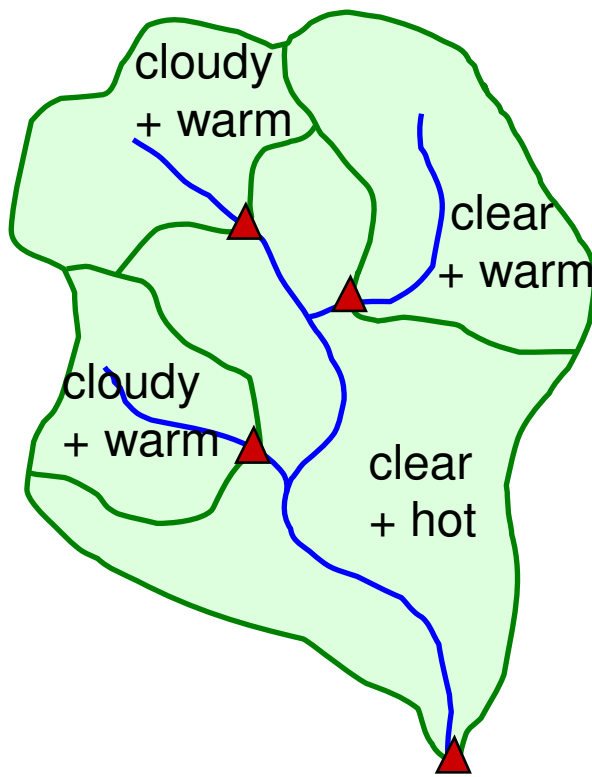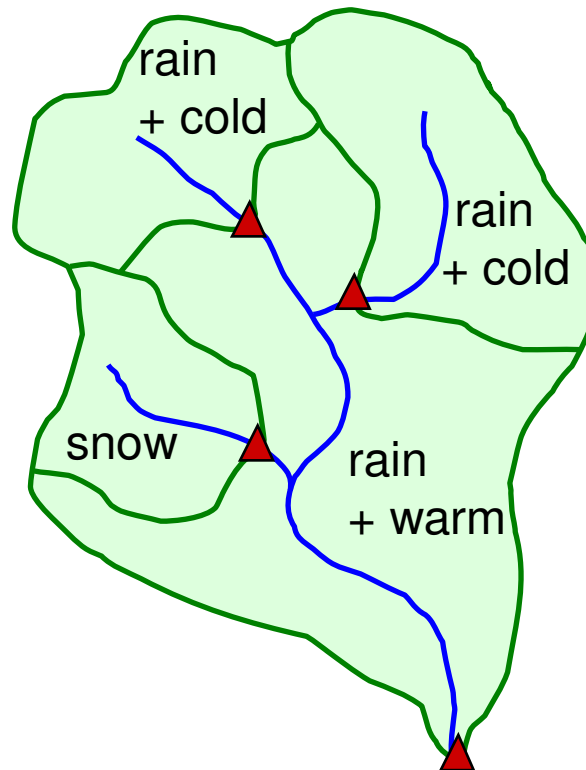
rain + warm

Medium flow

1964: very heavy rain + warm

heavy rain + warm

rain + warm

extreme rain + cold

heavy rain + warm
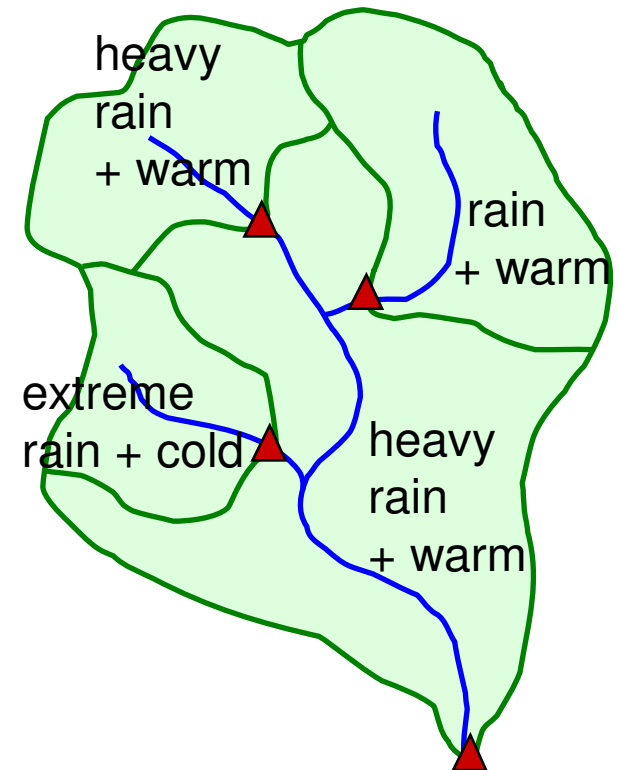
High flow

Ensemble Preprocessor: Schaake Shuffle

Conditional forecast ensembles constructed to maintain properties of precipitation and temperature in space and time

Example of ensemble forecasts for a given date

1975: mostly clear + hot*   *1996*: snow/rain + cold*   *1964*: heavy rain + warm*

cloudy + warm

clear + **hot**

light rain + warm

clear + hot

Low flow

snow

rain + cold

snow

rain + **cold**

Medium flow

heavy rain + **cold**

rain + warm

heavy rain + **warm**

rain + warm

High flow

35
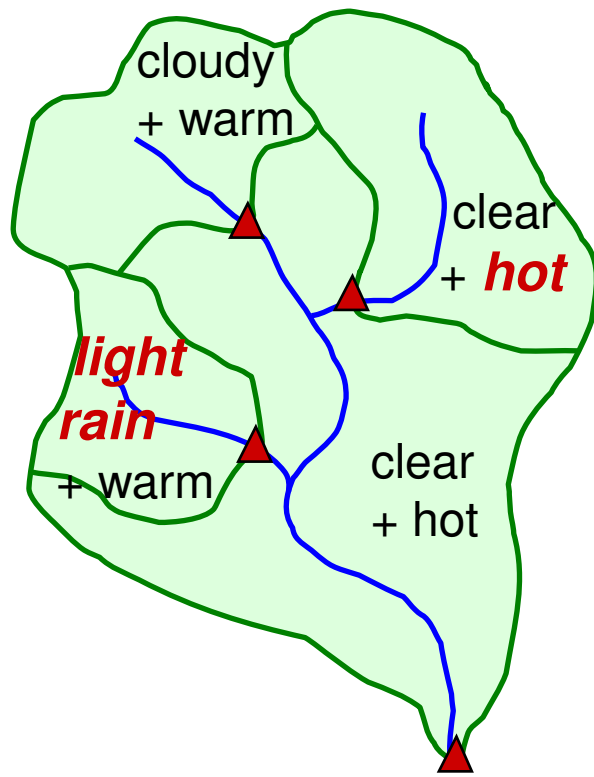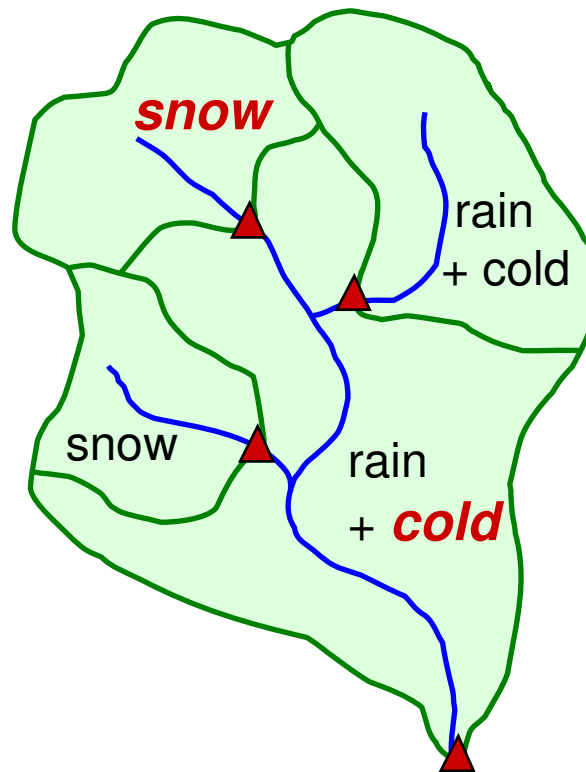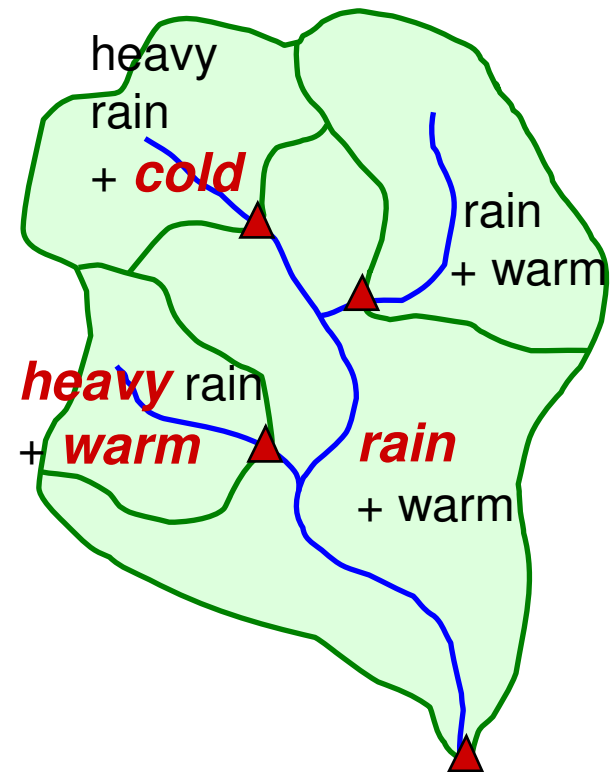
Ensemble Postprocessor: Methodology

- Process for lead day k for the ith ensemble member
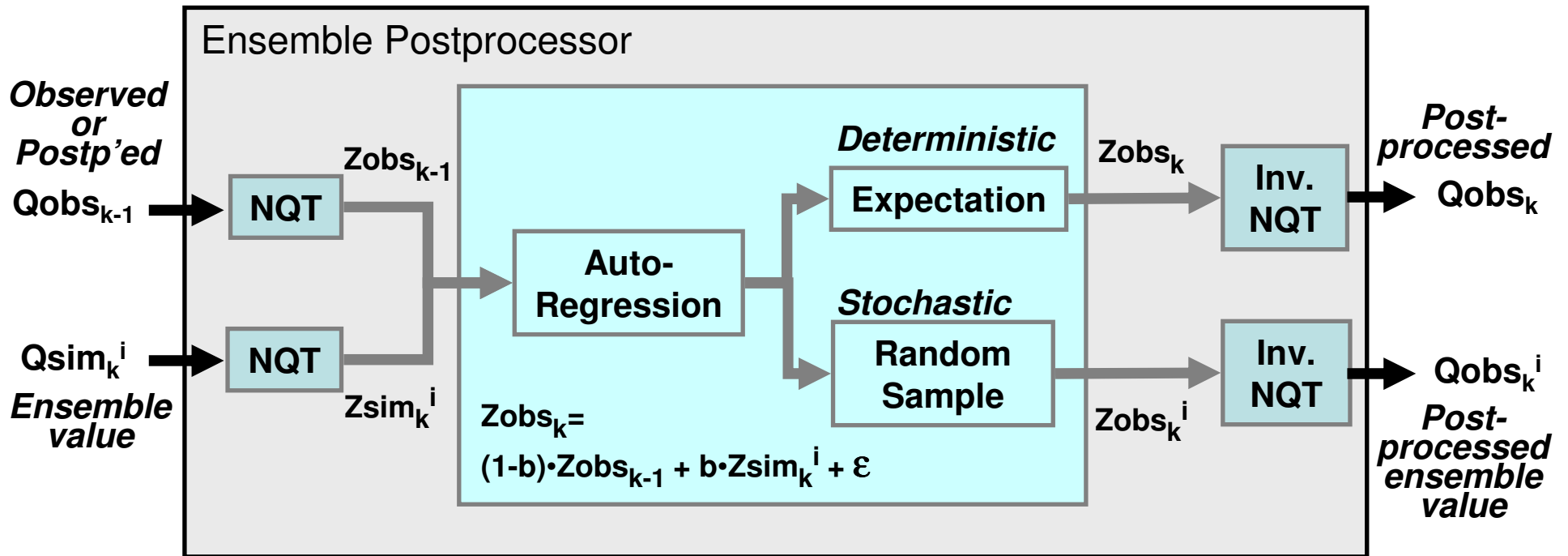
Observed or simulated value ingested by Ensemble Postprocessor:

Lead day k = 1: $\mathbf{Qobs_0}$ current observation

Lead day k > 1: $\mathbf{Qobs_{k-1}}$ / $\mathbf{Qobs_{k-1}^{i}}$ previously postprocessed value

Ensemble Postprocessor: Future Research

- Disaggregating technique to generate 6-hr ensembles

- Performance of the ensemble postprocessor is sensitive to data availability

- If long-duration data is available, the postprocessor performs as expected (correct model biases, produce reliable ensemble traces)

- Ensemble postprocessor does not handle regulated flows very well

- Storm typing/stratification/conditioning is necessary to handle disparate events (e.g. rain-on-snow)

- Perturbation and hybrid approaches should also be considered to deal with data paucity and to capture longer-memory model dynamics better

Hydrologic Model Output Statistics (HMOS)

- HMOS Approach:
  - Accounts for total uncertainty via establishing statistical relationships between single-valued forecasts and observed flows, using QPF info

- Current prototype:
  - Simpler approach for short-term flow ensemble generator
  - Combination of Model Output Statistics (MOS) and statistical adjust-Q technique
  - Statistical relationships between forecast and observed flow for each forecast lead time (ensembles generated at 6-hr time step) based on flow stratification (multiple categories defined from bias-adjusted flow forecast and accumulated QPF amount)
  - Uses forecast flows, recent observed flows, and QPF info
  - Needs long records of model stage/forecast flows, stage/forecast observations, and QPF

HMOS: Methodology

- Process for lead time k to generate the ith ensemble member from single-value forecast: similar to Ensemble Postprocessor
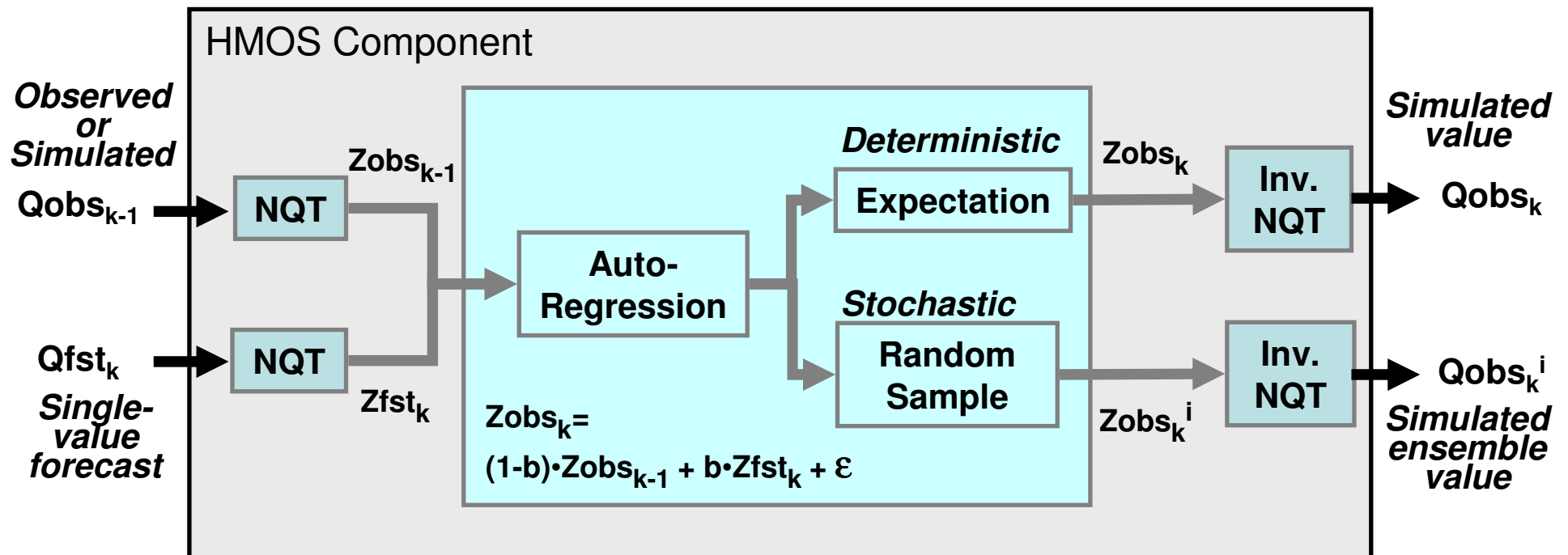
Observed or simulated value ingested by HMOS:

Lead time k = 1: $Qobs_0$ current observation

Lead time k > 1: $Qobs_{k-1}$ / $Qobs_{k-1}^i$ previously simulated value

Case Study

- North Fork of the American River (875 km^2) near Sacramento, California
- Daily products, 14 lead days, 45 members, 1979-2005
- GFS-based Ens. Pre-Processor (**EPP**) and Ens. Post-Processor (**EnsPost**) against climatology, evaluated via Ensemble Verification System (**EVS)**

EPP: retain skill in single-valued input fcst & generate unbiased ENS

EnsPost: account for all hydrologic uncertainties
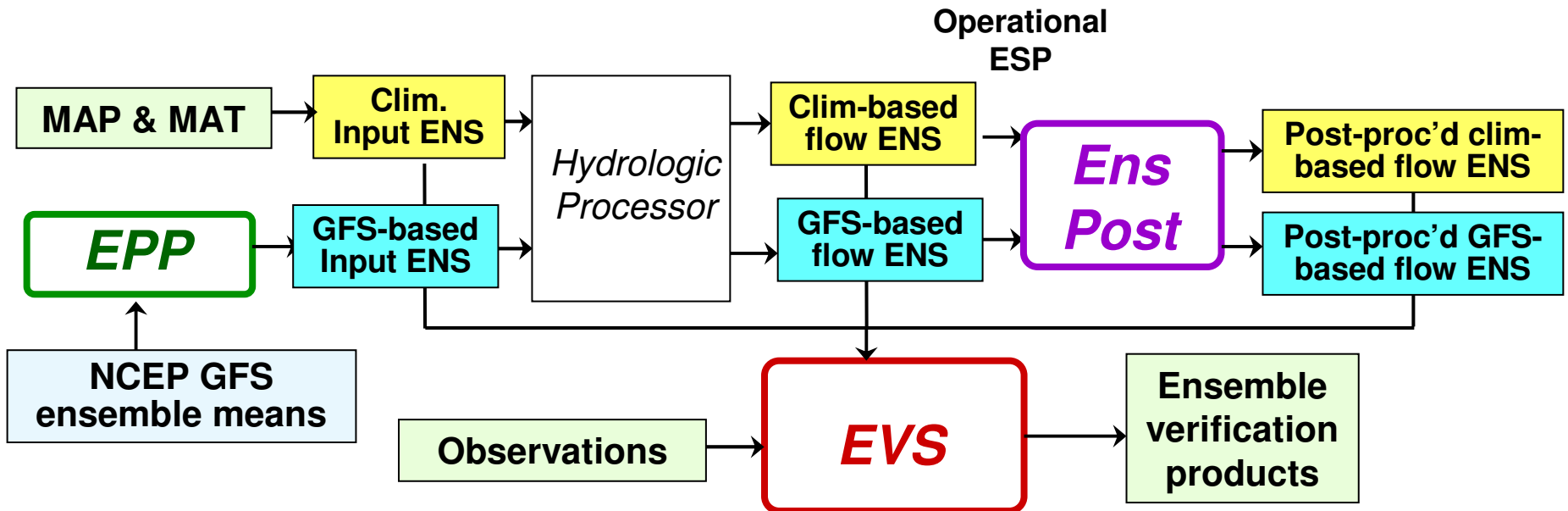
Verification Results: Pre-Processor

- GFS-based 24-hr **precipitation ENS** from EPP vs. Climatology:

  - Mean Continuous Ranked Probability Score (CRPS)

  - Mean CRPS decomposition
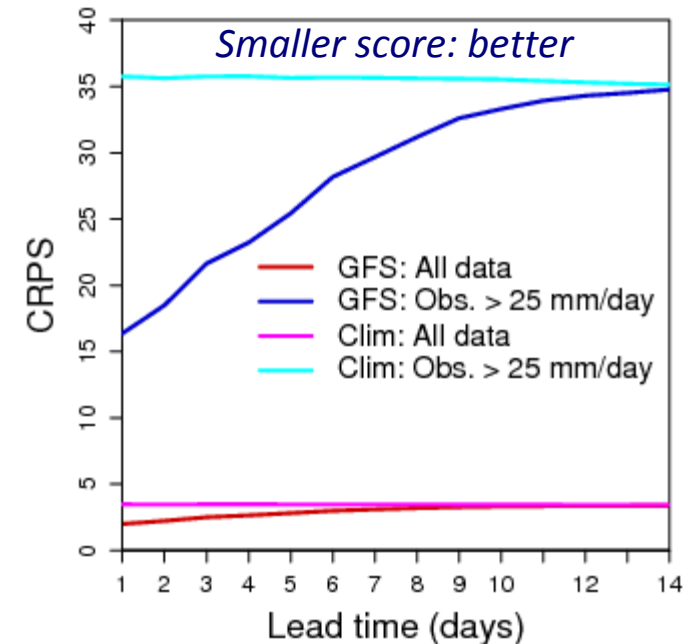
Mean CRPS= Reliability
  + Potential CRPS
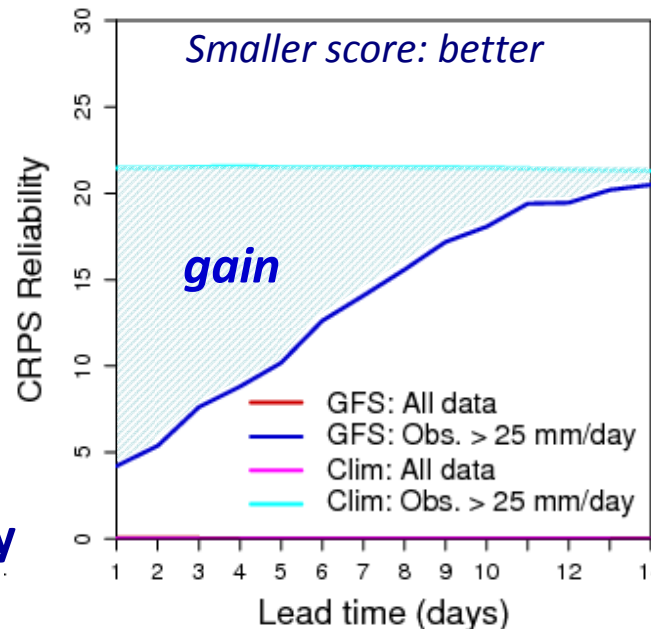
Gain is mostly in reliability
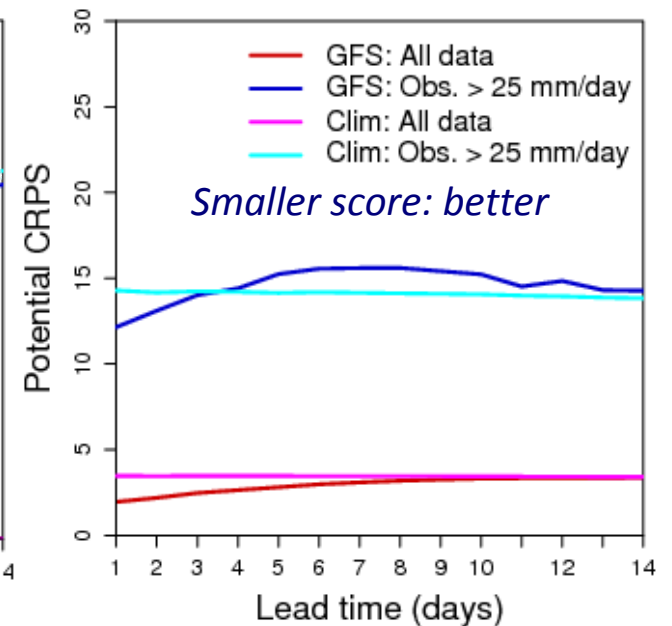
Verification Results: Post-Processor

- 24-hr **flow ENS** from EPP GFS-based forcing w/ vs. w/o EnsPost

  - **Mean CRPS**: improvement from EnsPost is most significant at short lead time, from reducing uncertainty in initial conditions
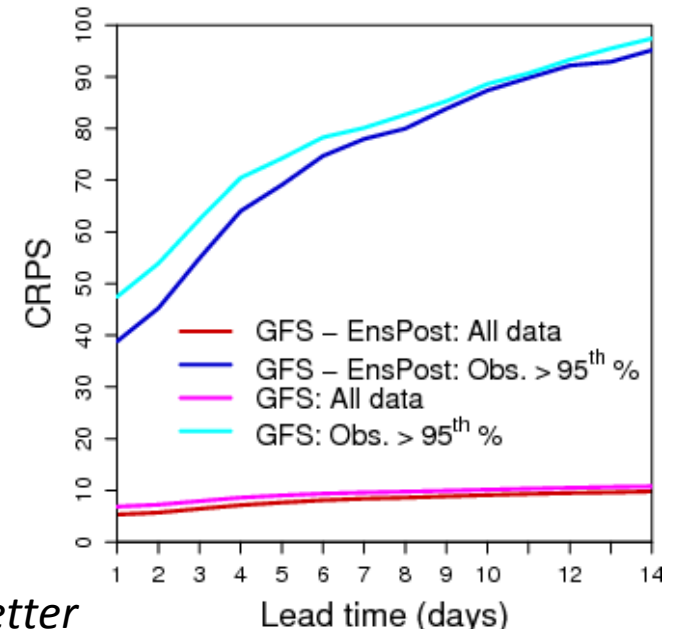
  - **Mean CRPS decomposition**

  Mean CRPS=
    Reliability
    + Potential CRPS

  Significant gain in reliability w/ EnsPost at all lead times

Smaller scores: better
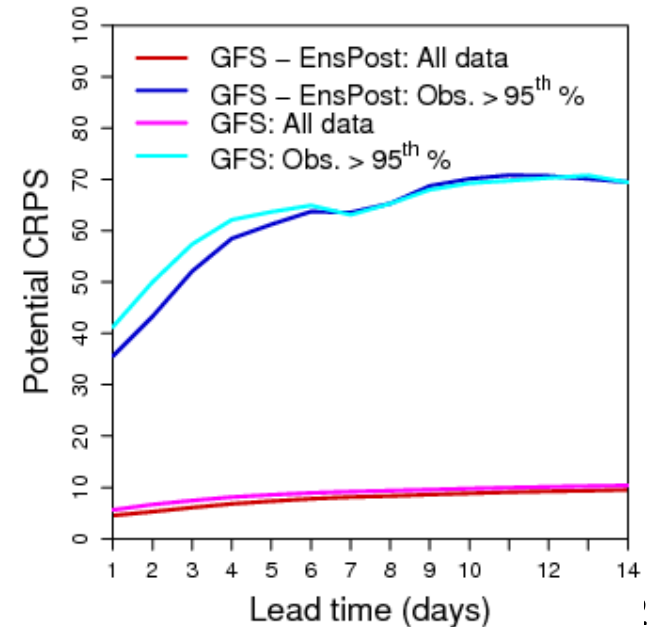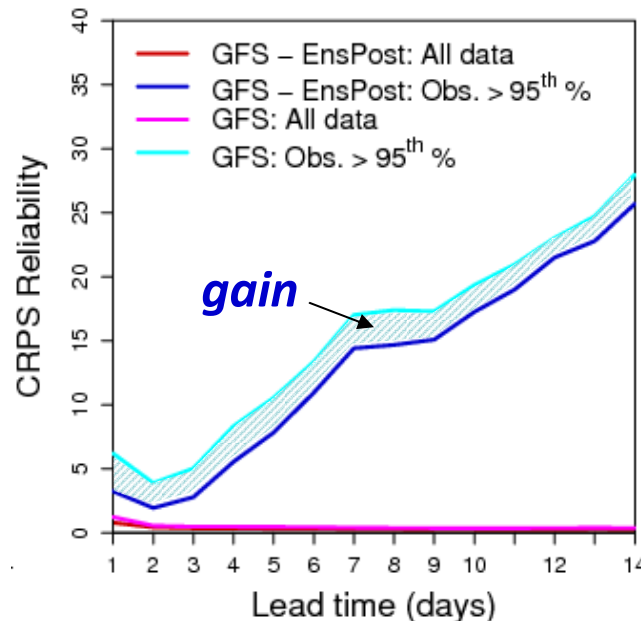
Why ensemble forecasting?

- Provide estimate of forecast uncertainty
  - Forecasters get objective guidance for level of confidence in forecasts
  - End users decide whether to take action based on their risk tolerance
- Help extend forecast lead time or ascertain its limit
  - Weather and climate forecasts are highly uncertain and noisy; they cannot practically be conveyed as single-valued
- Improve forecast accuracy
  - Averaging two good (or bad) forecasts in some way is usually better than either of the two
- Help improve forecast quality cost-effectively
  - Assessment of relative importance of major sources of uncertainty helps define targeted improvements

Uncertainties in hydrologic forecast

Quantify meteorological/Input uncertainty

Reduced uncertainty due to pre-processing

Ensemble pre-processor

Quantify parametric uncertainty

Reduced uncertainty due to calibration

Parametric uncertainty processor

Quantify uncertainty in initial conditions

Reduced uncertainty due to data assimilation

Data assimilator

Uncertainty

Lead Time

Structural uncertainty, residual uncertainty

Ensemble post-processor, multimodel ensemble

Flow regulations - A large challenge

Integration of Input and Hydrologic Uncertainties
An illustration of EPP-EnsPost approach

Precipitation ensemble forecast

2.0 in

1.0 in

0.0 in

Hydrologic, hydraulic models

Streamflow ensemble forecast

500 cms

250 cms

Simulated streamflow (cms)

Observed streamflow (cms)

Uncertainty-integrated streamflow ensemble forecast

Repeat for all individual streamflow ensemble members

Post-processed streamflow ensemble member

EVS interface, graphics and free download