Ensemble Pre-Processor (EPP) RFC Subsystem

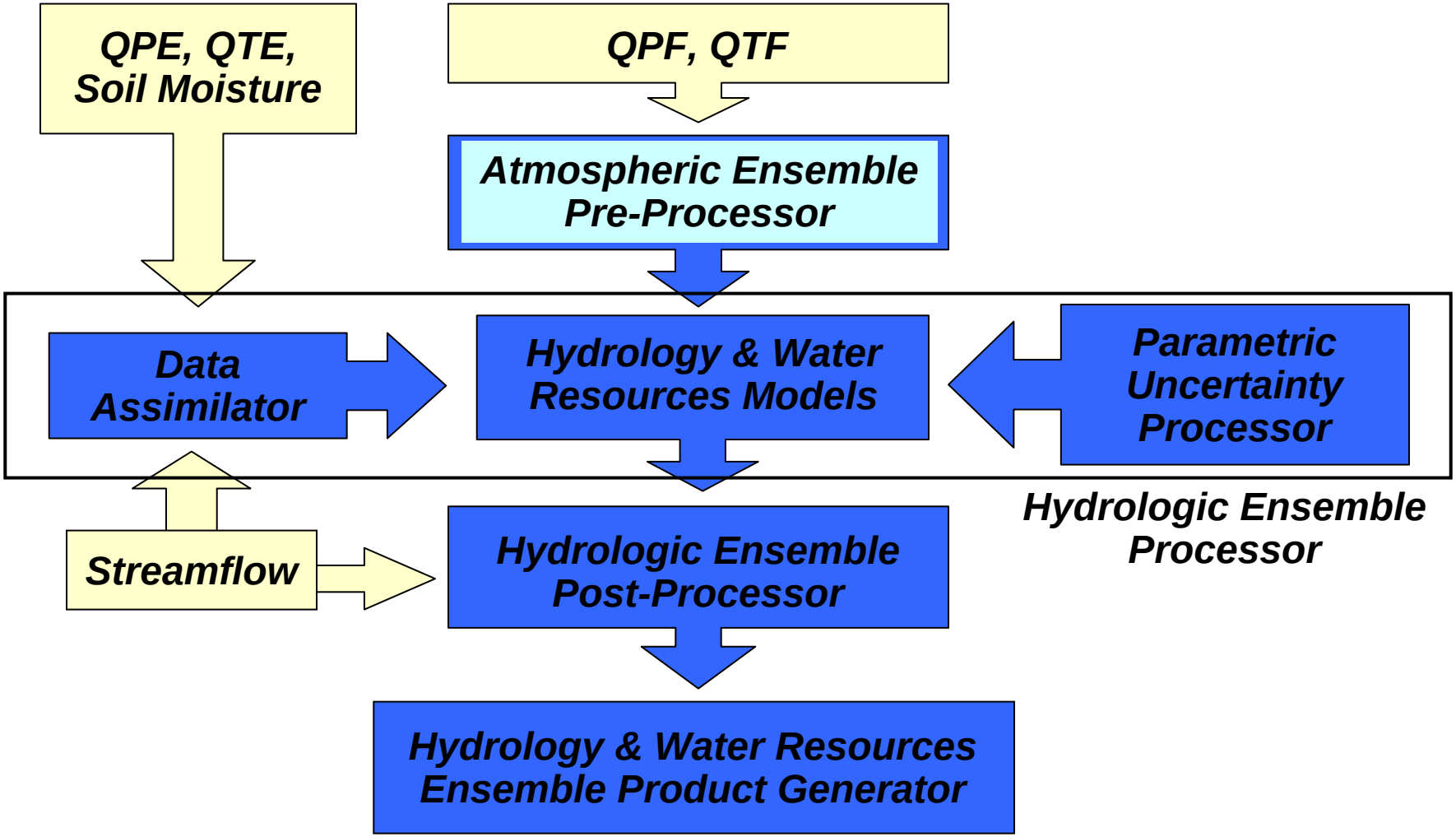
Limin Wu[1,3], Julie Demargne[1,2], John Schaake[1,4],
Henry D. Herr[1], and Dong-Jun Seo[1,2]

[1]NOAA/NWS/Office of Hydrologic Development
[2]University Corporation for Atmospheric Research
[3]RS Information Systems, Inc.
[4]Consultant

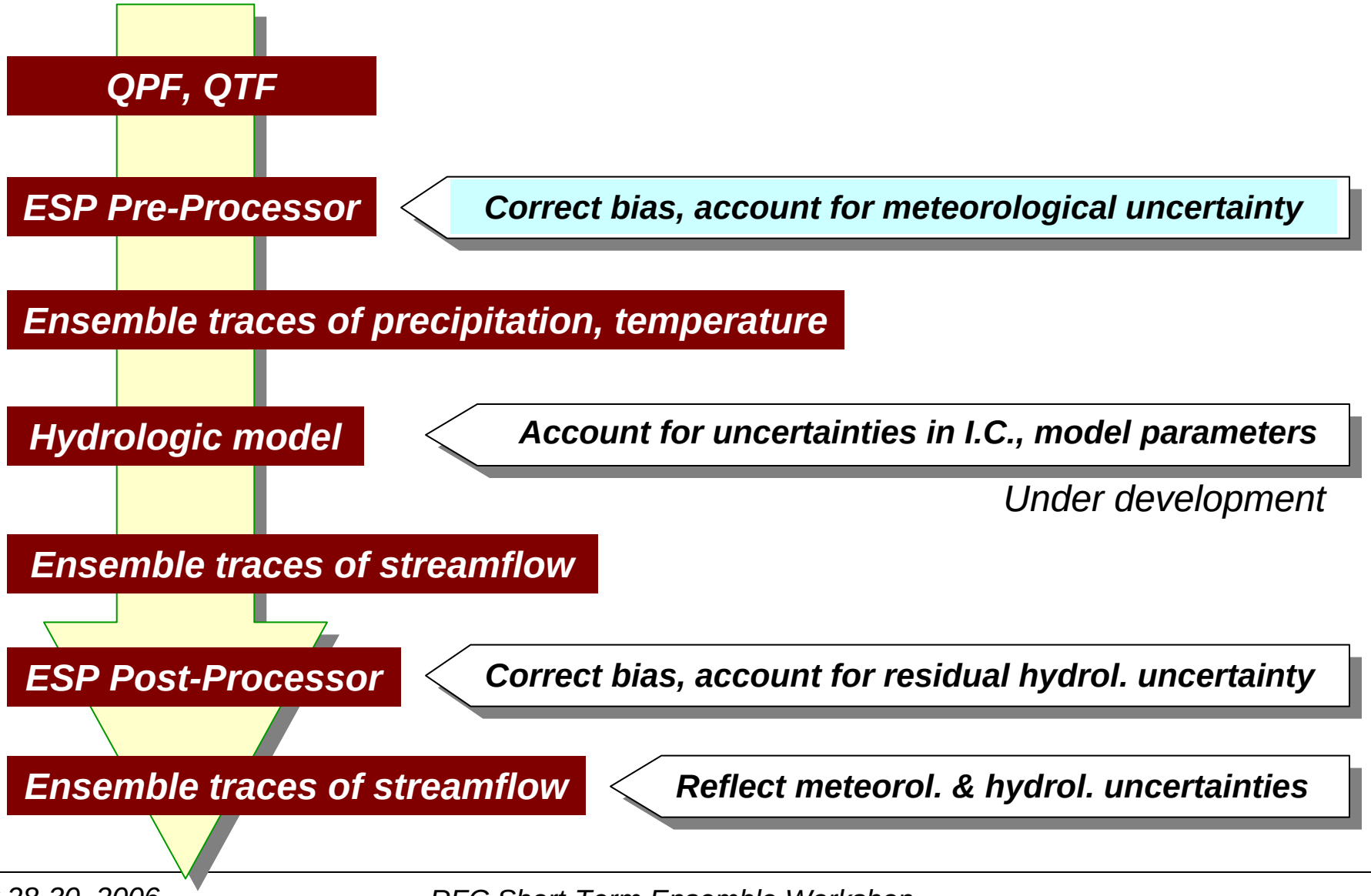
Acknowledgments

- RFC contributors: Robert Hartman, Bill Lawrence, Billy Olsen, Joseph Ostrowski, and Edwin Pryor
- Past contributors: Shuzheng Cong, Xiabiao Fan, Mary Mullusky, and Edwin Welles

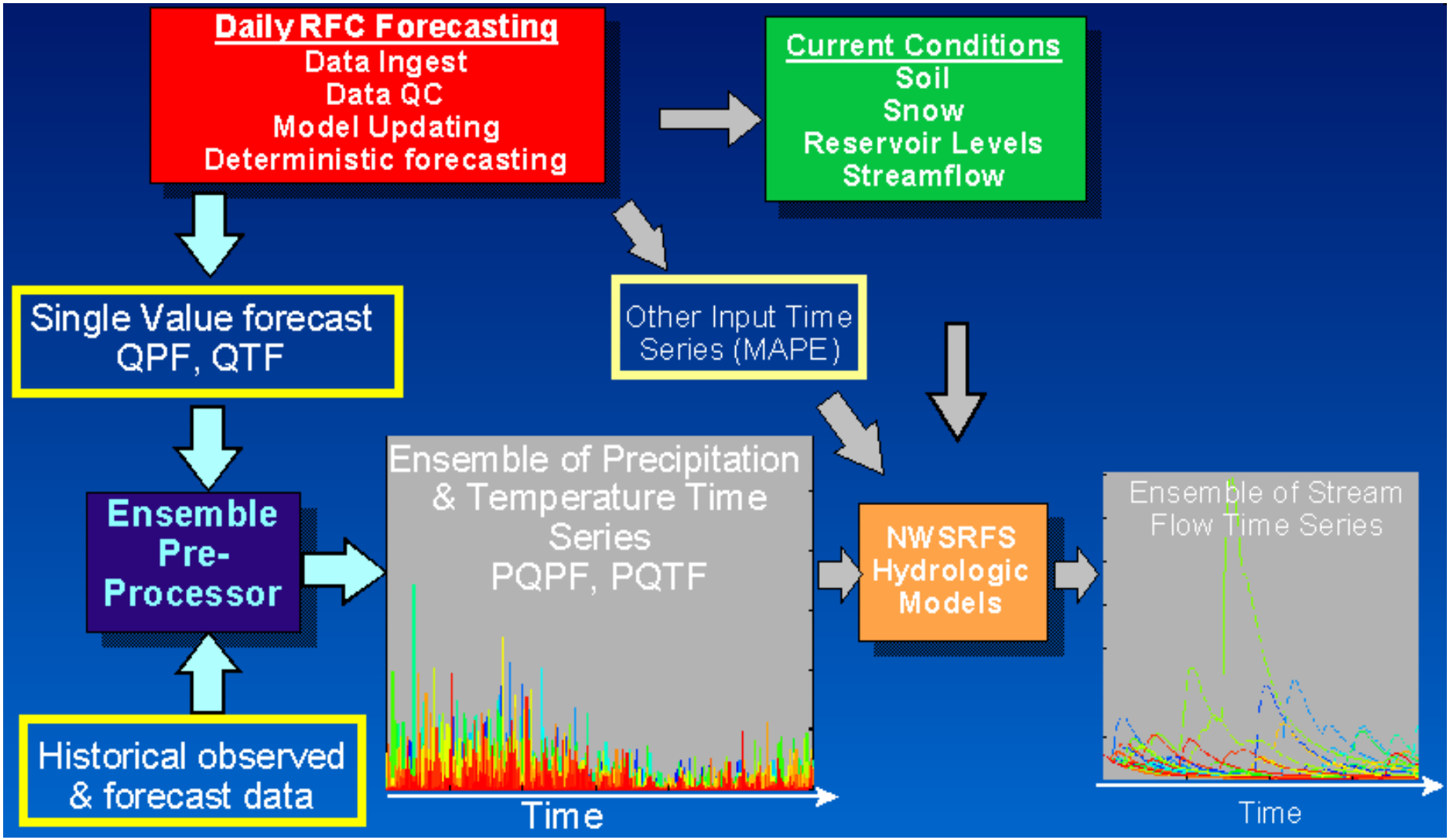
Elements of a Hydrologic Ensemble Prediction System



Short-Term Ensemble Streamflow Prediction (ESP): Current Process

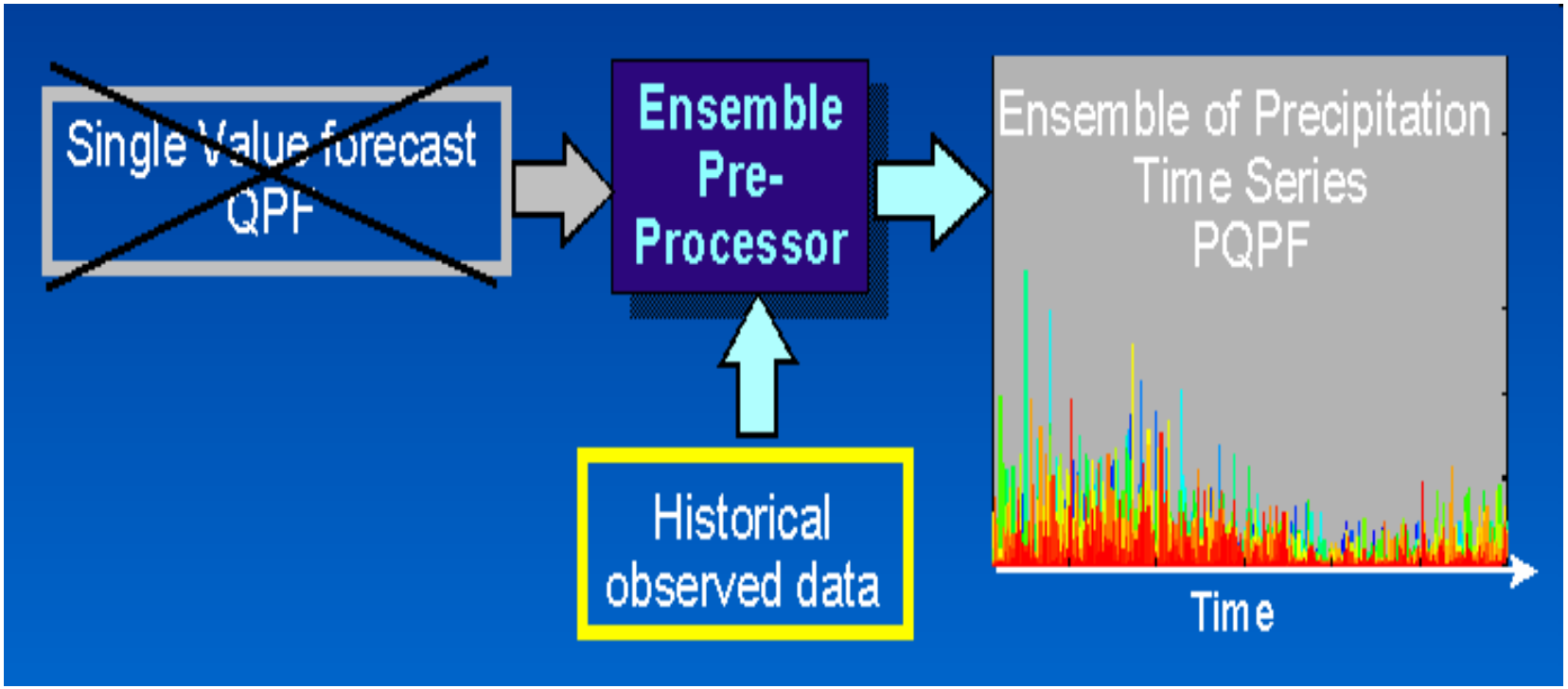


ESP Using EPP

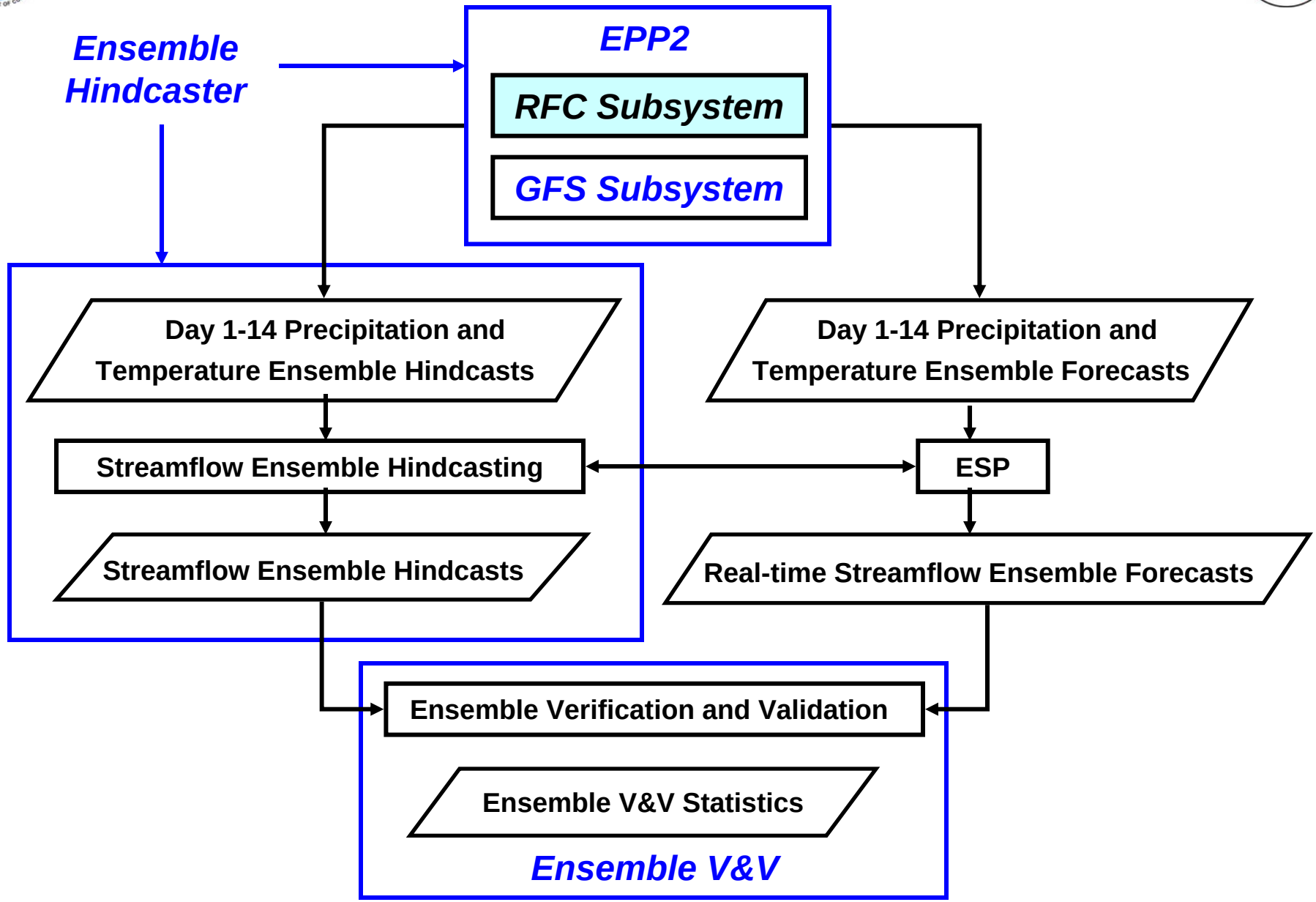


What if single-value forecast is not available?

- Climatology fills in:
  - Estimate and smooth statistical parameters for climatological probability distributions of precipitation and temperature
  - Sample from the climatological probability distributions



FY06 Ensemble Projects and their Relationships



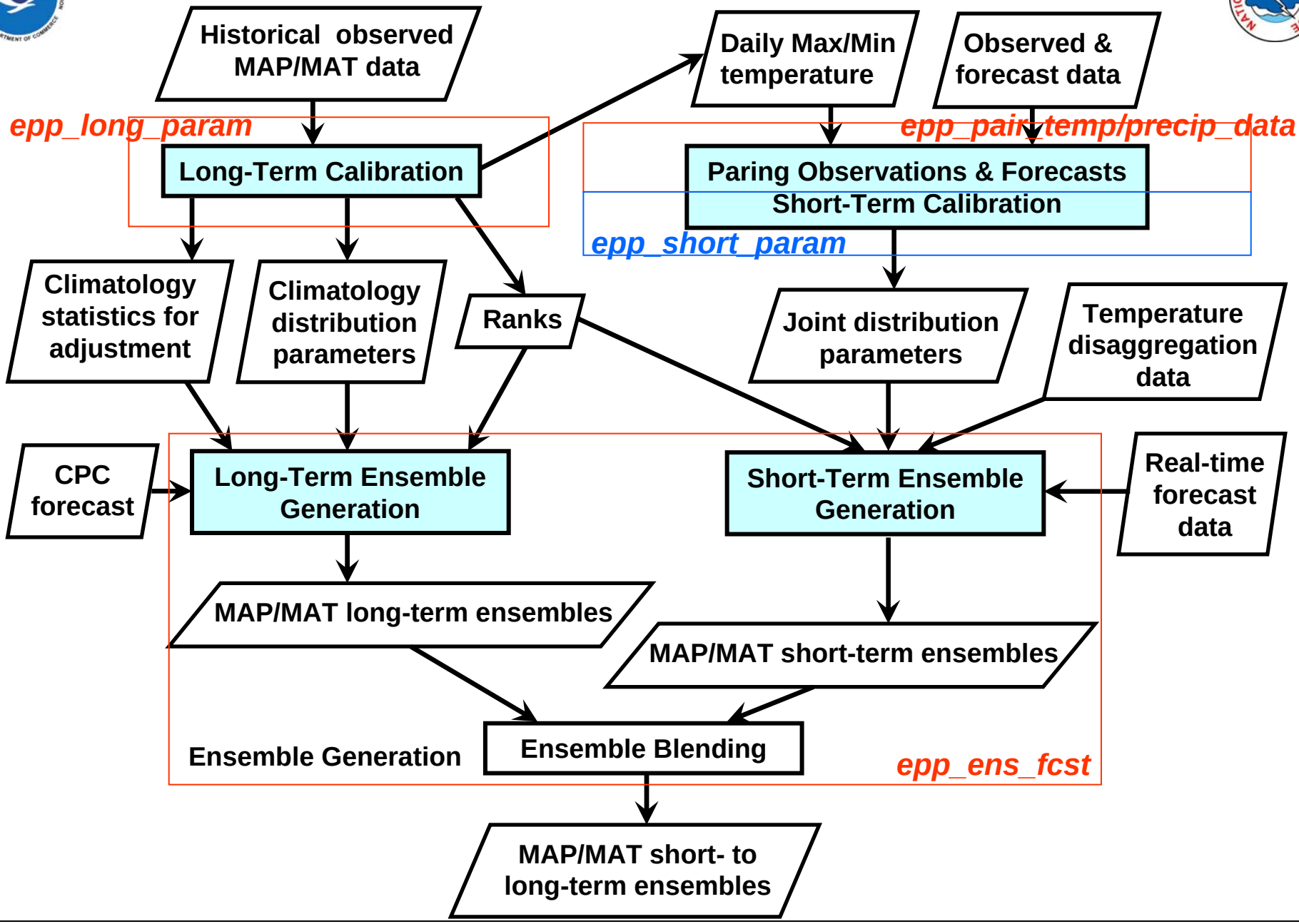
Why ensembles from single-value forecast?

- Single-value (i.e. deterministic) forecast has additional skill from human forecasters, particularly in Day 1
- Precipitation ensembles from NWP have significant biases in the mean and in the spread
- The goal of the ESP Preprocessor is to produce precipitation ensembles that are unbiased, and that reflect the additional skill
  - A practical, “observation-driven” approach for generation of precipitation ensembles

EPP RFC Subsystem

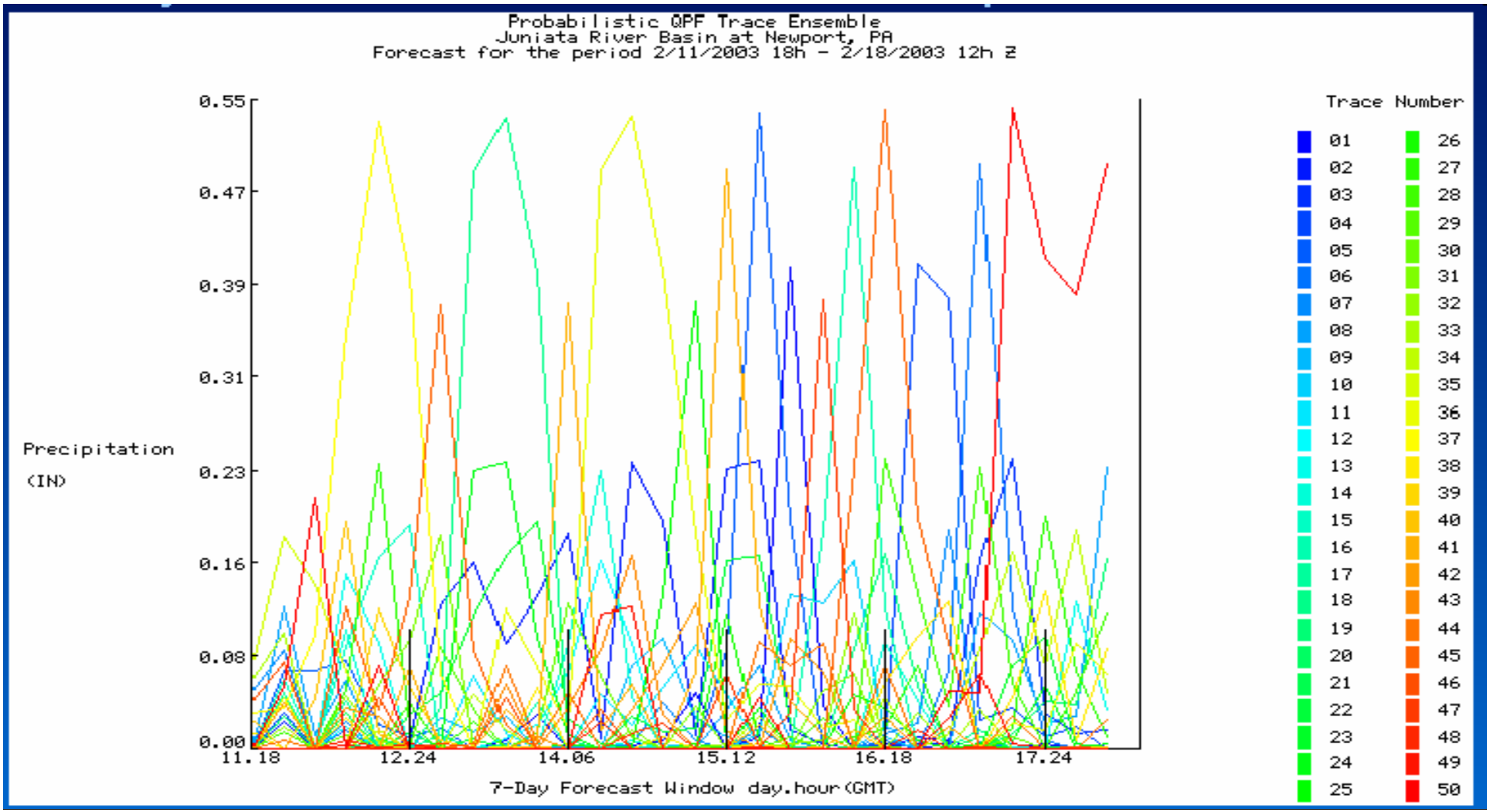
- Built on previous versions
- Has been in experimental operation at AB- and MARFCs
- Models joint probability distributions between 6-hr MAP/MAT and single-value forecast of precipitation/temperature
- Requires multi-year archive of 6-hr MAP, 6-hr QPF, 6-hr MAT, and forecast of 24-hr Tmin and Tmax
- Generates short-term precipitation/temperature ensembles via:
  - Stratified sampling from conditional distribution of observed precipitation/temperature given single-value forecast of precipitation/temperature
  - Probability matching of historical ensembles via Schaake Shuffle (Clark et al. 2004)
- Generates long-term ensembles from historical traces using resampled climatology
- Operates in real-time and hindcasting modes
- Written in Fortran and C++; Korn shell script for hindcasting mode

EPP RFC Subsystem



A Sample Output Displayed by ESPADP

7-day Probabilistic Quantitative Precipitation Forecasts



Ensemble generation from single-value forecast

- Step 1) Model joint probability distribution between observed and single-value forecast precip/temp from multi-year archive (go to the schematic)
  - For each segment, forecast lead time, and time of the year (centered over some user-specified time window)
- Step 2) construct via stratified sampling the conditional probability distribution of observed precip/temp given single-value forecast precip/temp (go to the schematic)
- Step 3) Arrange the output forecast ensemble members such that they have the same ordering as the historical ensemble members (i.e. Schaake Shuffle, Clark et al. 2004) (go to the schematic)

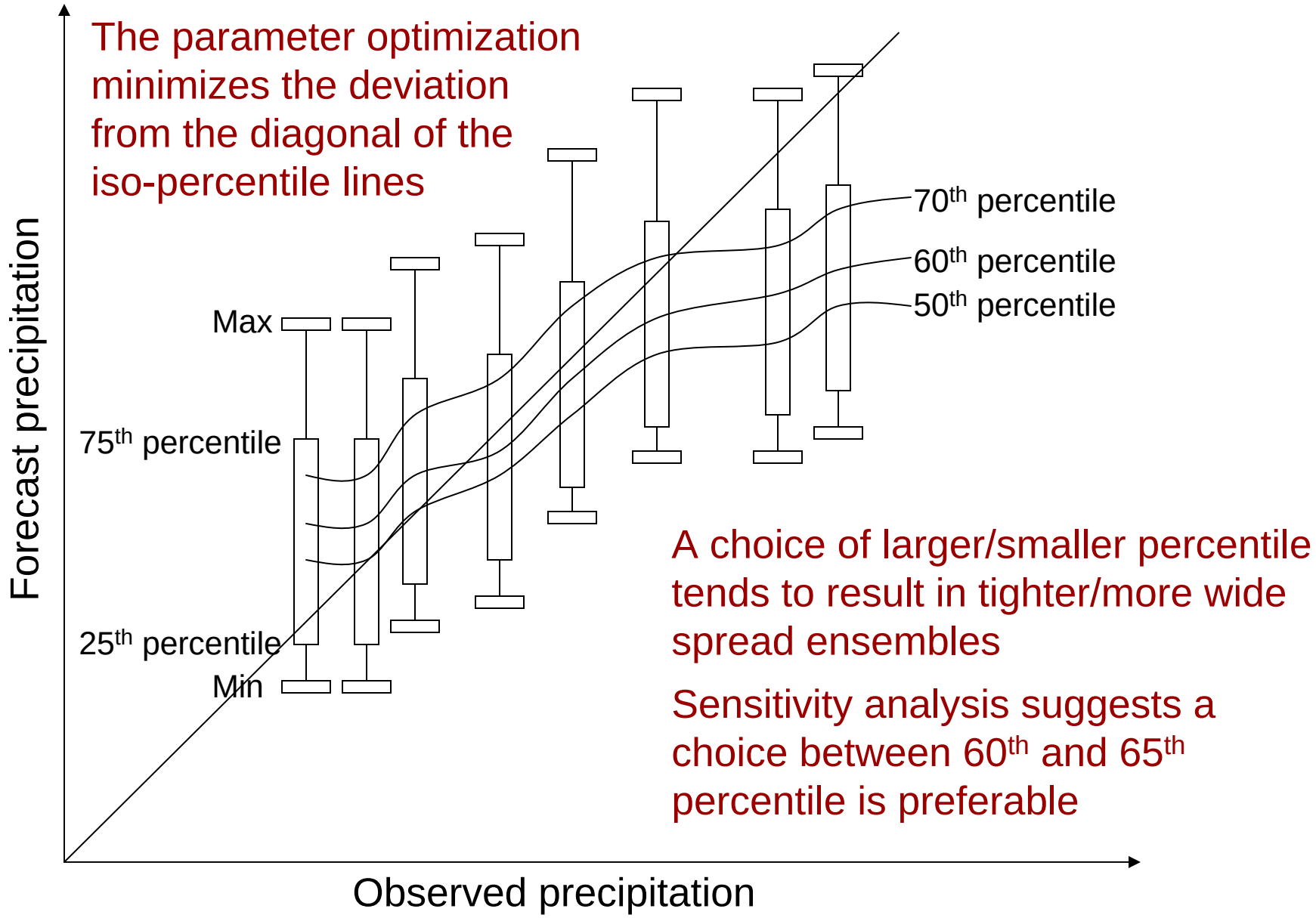
Explicit modeling of precipitation intermittency

- Probability distribution of precipitation is of mixed type (mass at zero, continuous elsewhere; go to the schematic)
- Until recently, this mixed nature of probability distribution of precipitation was modeled only approximately in EPP
- In FY06, new modules were developed, and implemented in the RFC Subsystem for testing and evaluation, that explicitly treat probability distribution of precipitation as mixed type (Herr and Krzysztofowicz 2005; Cong et al., manuscript under preparation)
- The effects are subtle and improve various performance measures for light precipitation
- See “Ensemble verification II” presentation this afternoon for examples

Improved modeling of conditional probability distribution, parameter estimation & user control

- In FY06, new modules were developed, and implemented in the EPP RFC Subsystem for testing and evaluation, to allow:
  - A more flexible regression model to better-fit the data in modeling of conditional probability distribution
  - Explicit optimization of the regression parameter using a multi-objective function
  - Some control by the user over certain properties of the ensembles being generated
- See “Ensemble verification II” presentation this afternoon for examples

User control of ensemble behavior



EPP RFC Subsystem Data Requirements

- For precipitation ensemble generation
  - Historical 6-hour MAP time series in the OHD DATACARD format
  - Historical 6-hour QPF time series
  - Real time 6 or 1 hour QPF values
- For temperature ensemble generation
  - Historical 6-hour MAT time series in the OHD DATACARD format
  - MOS daily max and min temperature data files
  - MOS station list
  - Real time 6 QTF values

EPP Internal File Format for Storing Historical Observed and Forecast Values

A sample segment from a temperature file for CHTM7 in ABRFC

```
CHTM7 1843 1215199300Z 1231199800Z DF
DATE       OBS  FCST
12151993 38.87 44.00
12161993 43.10 45.00
12171993 46.92 39.00
12181993 42.21 39.00
12191993 48.59 42.00
12201993 41.44 32.00
12211993 47.41 35.00
12221993 33.69 24.00
...
```

EPP RFC Subsystem
Data Availability Survey Results for 5 RFCs

RFC	Format of 6-hr MAP Archive (Period of Record)	Format of 6-hr QPF Archive (Period of Record)	Format of 6-hr MAT Archive (Period of Record)
APRFC	Archive Database (02/2006?)	Archive Database (02/2006 w/ winter gaps)	Archive Database (02/2006)
LMRFC	Database (1996-2006)	Database (2004-2006) Files (1996-2006)	None
NCRFC	SHEF .B (2000-2006 w/ gaps)	OFS mod or xmrg (2000-2006 w/ gaps)	None
SERFC	Shef-coded Text Product (05/1997 - present)	Shef-coded Text Product (05/1997 - present)	None
WGRFC	OHD DATACARD (1/1996 – 10/2005)	QPS – SHEF (12/2003 – present)	None

Operation

- The ensemble generation program in the EPP RFC Subsystem operates in two modes:
  - Real time (forecast) mode
  - Retrospective (hindcast) mode
- The execution of the programs in the EPP RFC Subsystem involves:
  - Placing input files in the appropriate directories
  - Setting Apps-Defaults tokens
  - Preparing the input control file (input deck)
  - Running the executable at the command line prompt For details, see the user's manual available as Appendix B in Folder "dj" at http://www.weather.gov/ohd_files/quickpost/index.php

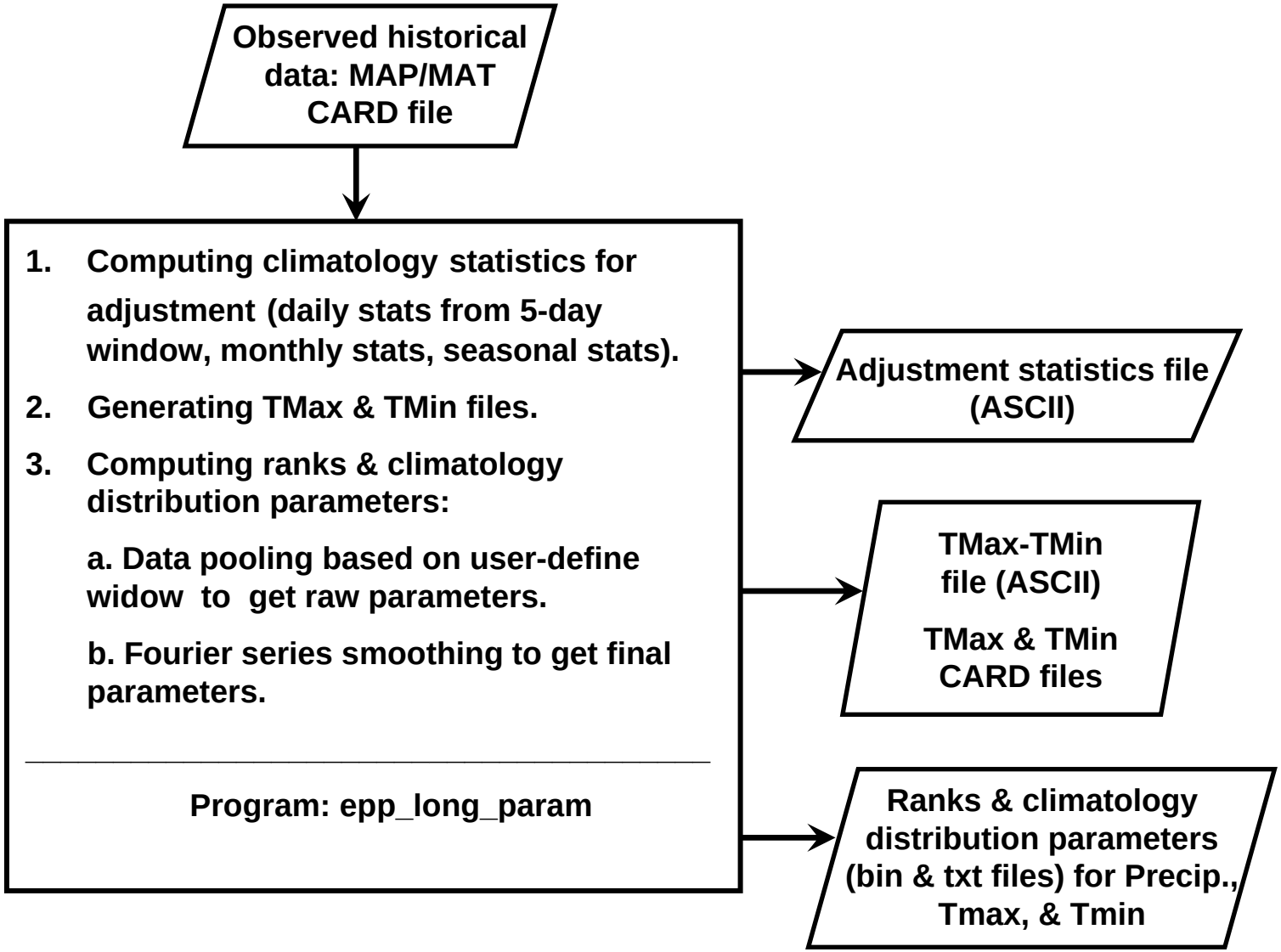
Connection to the Hydrologic Ensemble Hindcaster and Ensemble Verification System (EVS)

- Korn shell scripts developed to run the ensemble generation program repeatedly for the user-specified verification period to produce precipitation and temperature hindcasts
  - Hindcast files are in DATACARD format
  - Date-Time stamp is attached to the filename of the hindcast files
- The hydrometeorological hindcasts may be ingested:
  - by the Hydrologic Ensemble Hindcaster to generate streamflow ensemble hindcasts
  - by the Ensemble Verification System (EVS) for verification

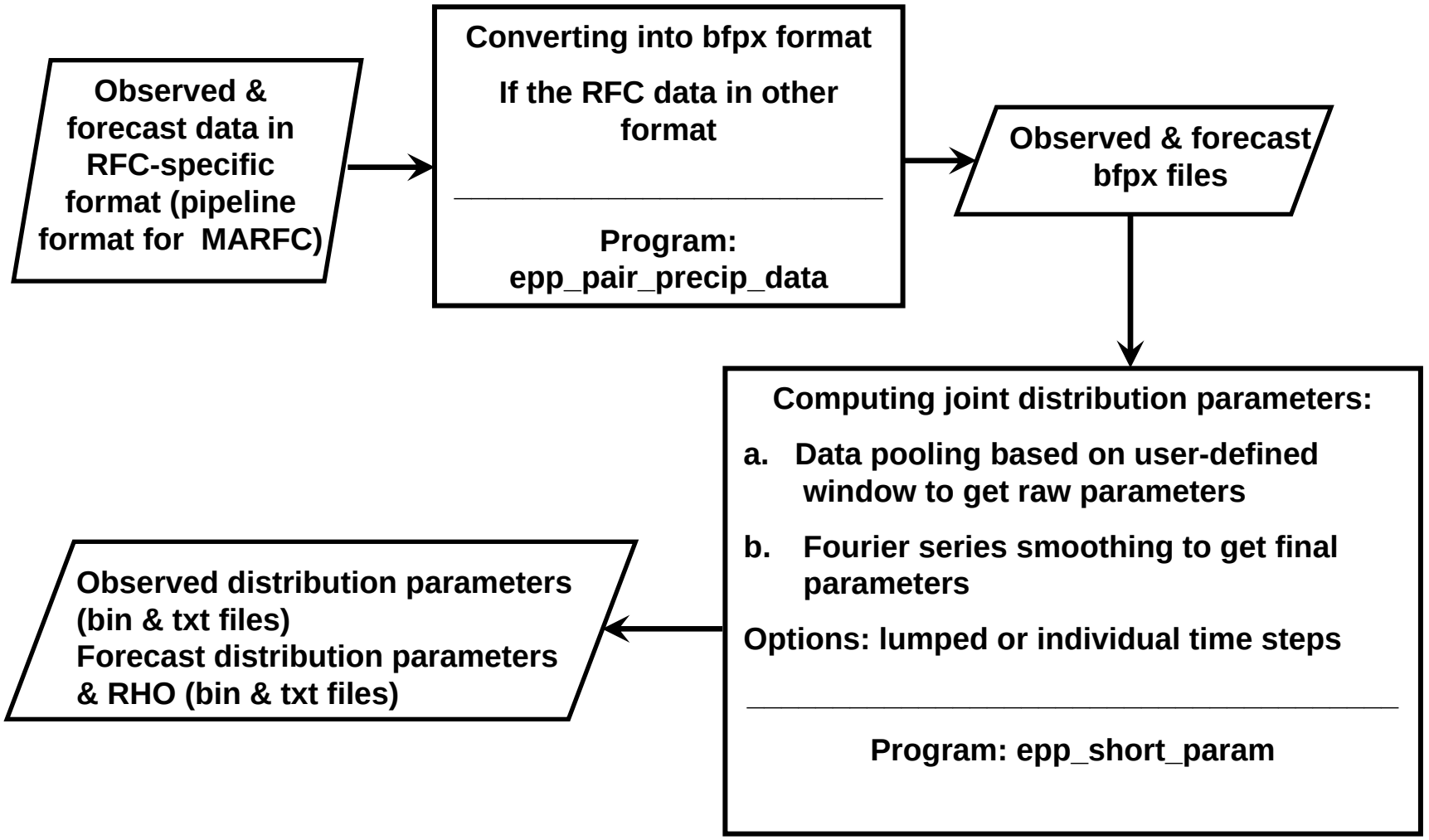
Thank you

Linked slides follow

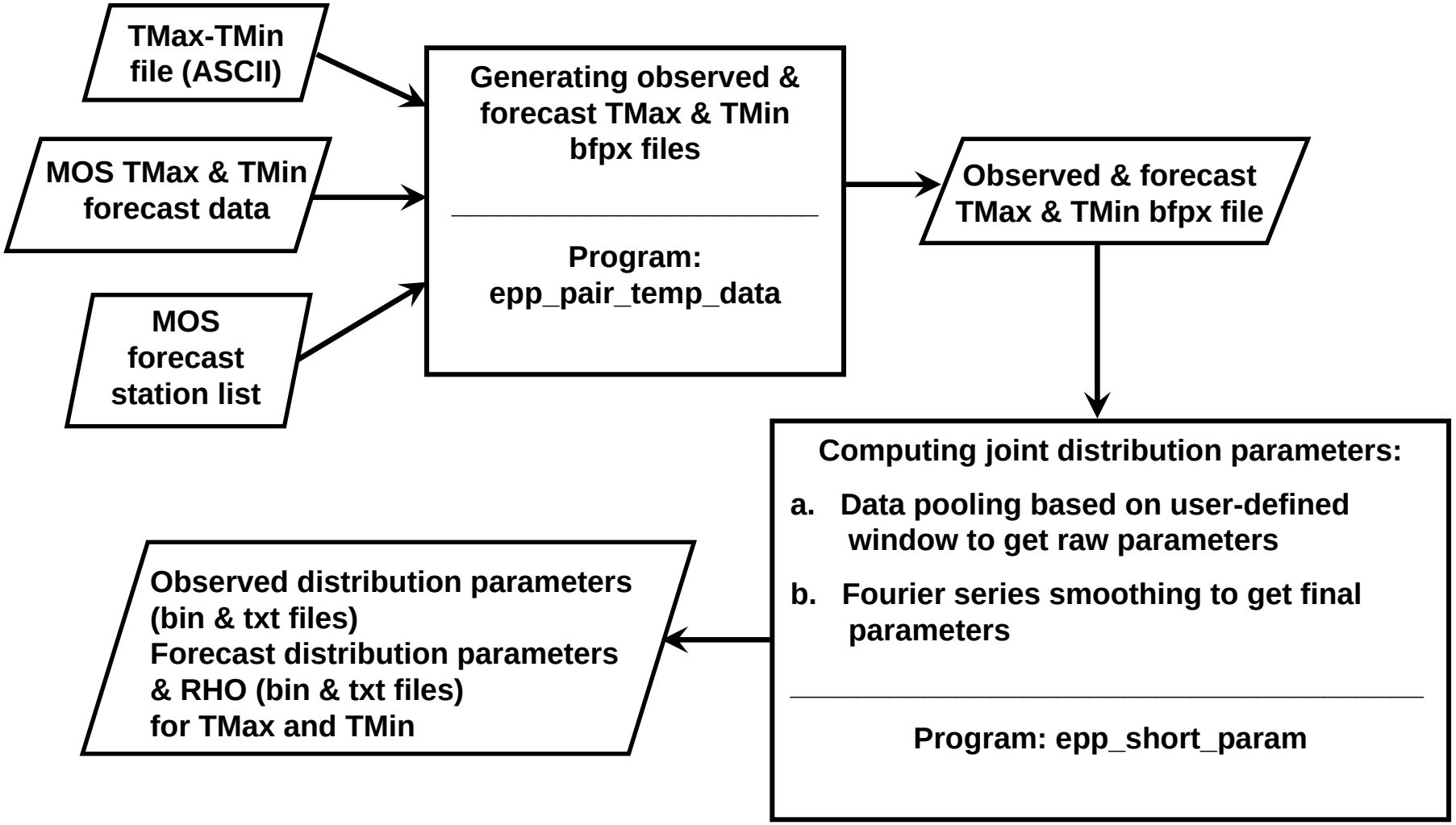
Long-Term Calibration



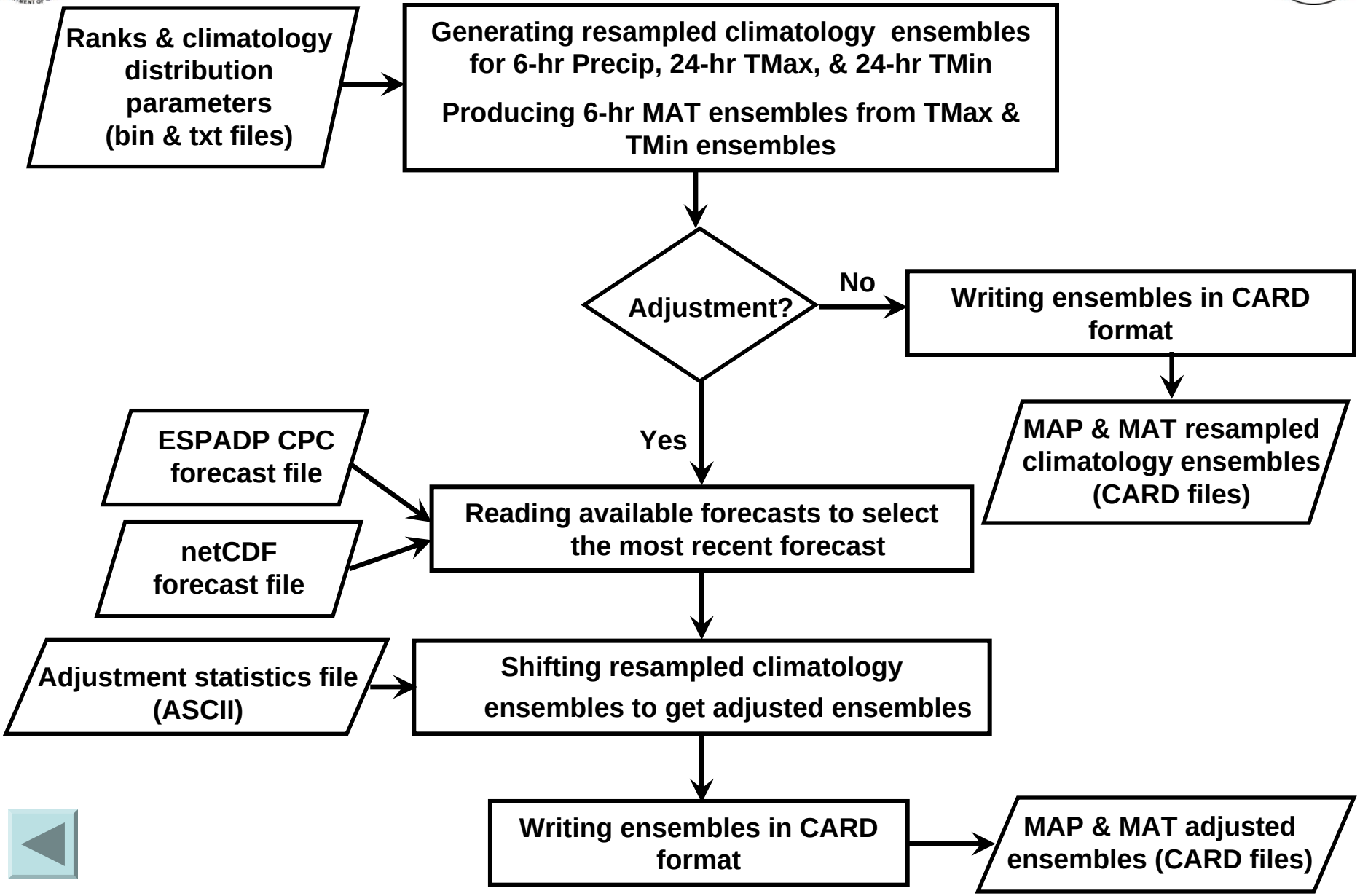
Precipitation Short-Term Calibration for each lead day



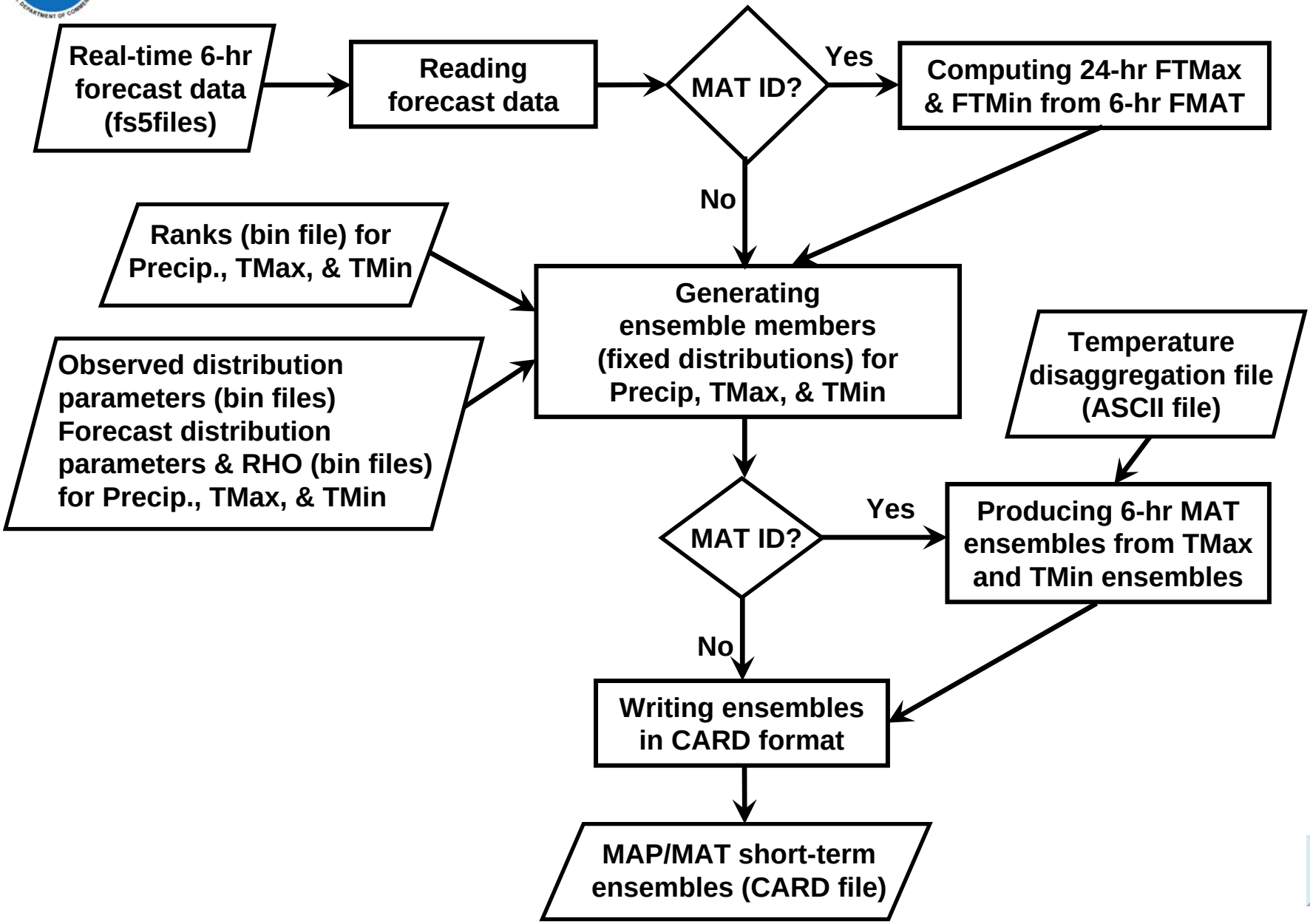
Temperature Short-Term Calibration for each lead day



Long-Term Ensemble Generation for a Segment ID



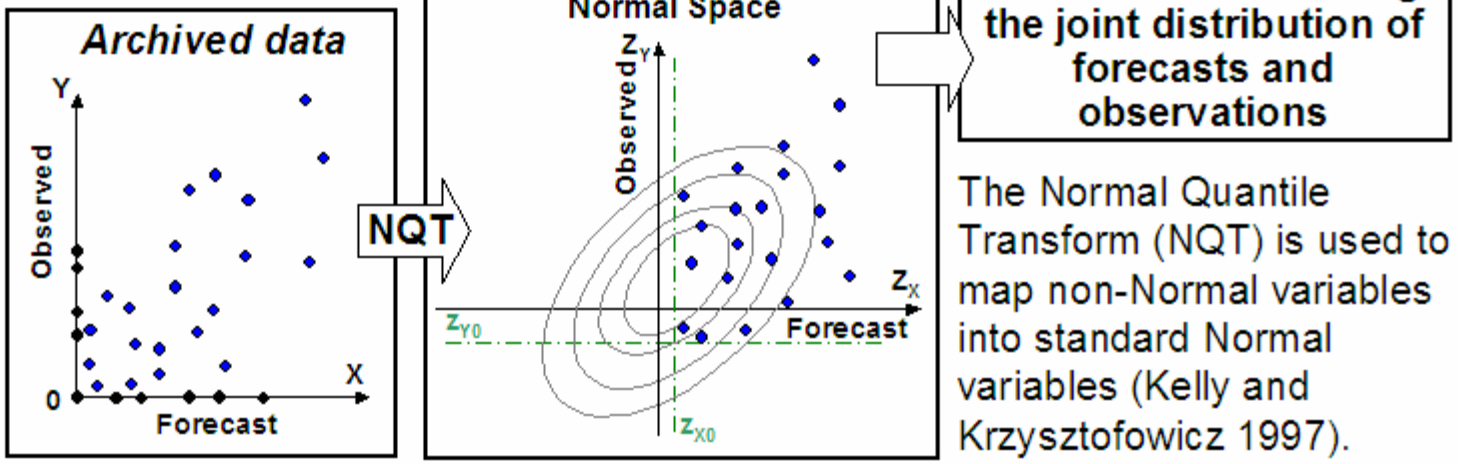
Short-Term Ensemble Generation for a Segment ID



The Ensemble Preprocessor generates short-term precipitation and temperature ensembles for each lead time and each location based on the 3 following steps.

Step 1: Joint Distribution Calibration

At each time step of the whole year and for each location, estimate the parameters of the joint distribution of observed and forecast values from archived data.



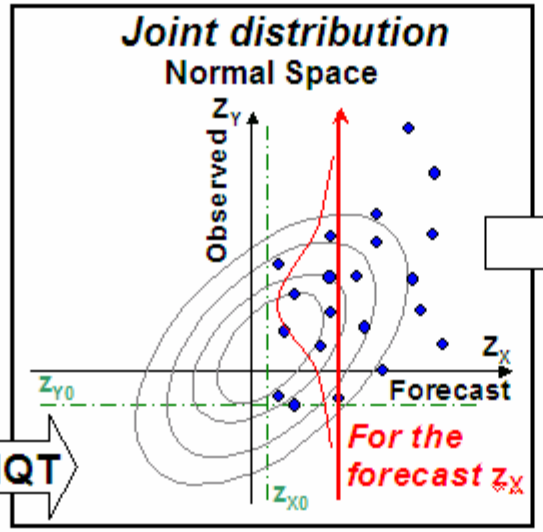
The Normal Quantile Transform (NQT) is used to map non-Normal variables into standard Normal variables (Kelly and Krzysztofowicz 1997).

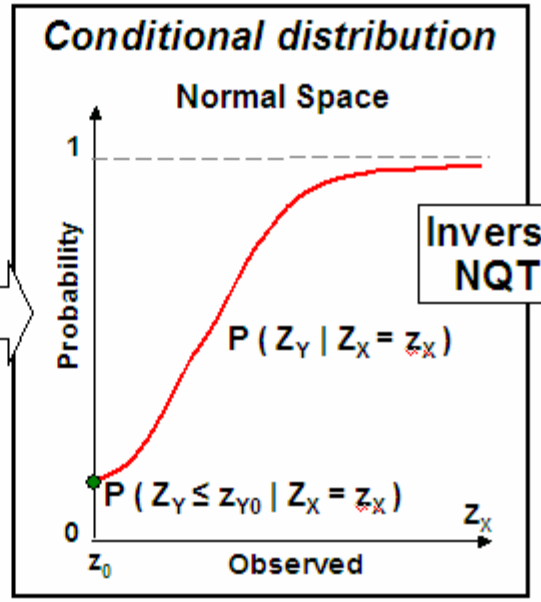
Back to Intermittency Back to Methodology

Step 2: Conditional Probability Distribution

At each time step of the forecast period and for each location, compute the parameters of the conditional distribution of future values given the single-value forecast.

Single-value forecast → **NQT**





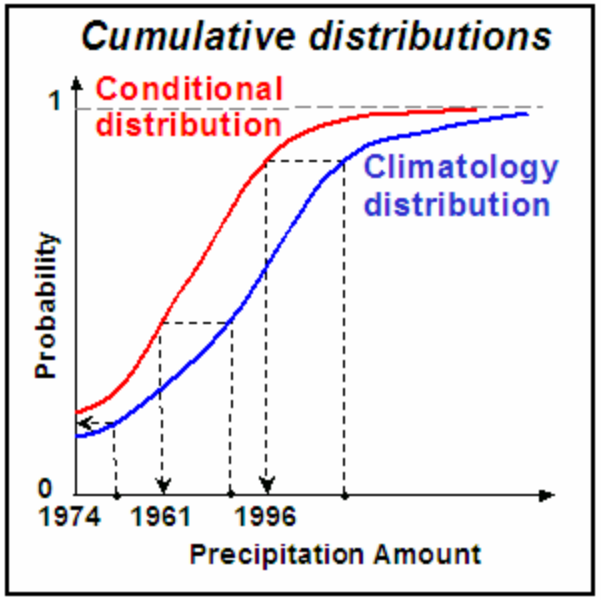
Inverse NQT → **Probabilistic forecast given the single-value forecast**

Schaake Shuffle (hopefully) explained!

Step 3: Distribution Mapping

At each time step of the forecast period and for each location, generate ensemble points given the conditional distribution of future events by sorting and then re-scaling climatological observations.

Resulting ensemble forecasts capture the skill of the single-value forecasts. By re-scaling climatological observations, this method actually maintains the space-time patterns between any two hydrometeorological variables (e.g., precipitation and temperature).



This distribution mapping technique, called the Schaake Shuffle method, is described in Clark et al. (2004).

End of all slides