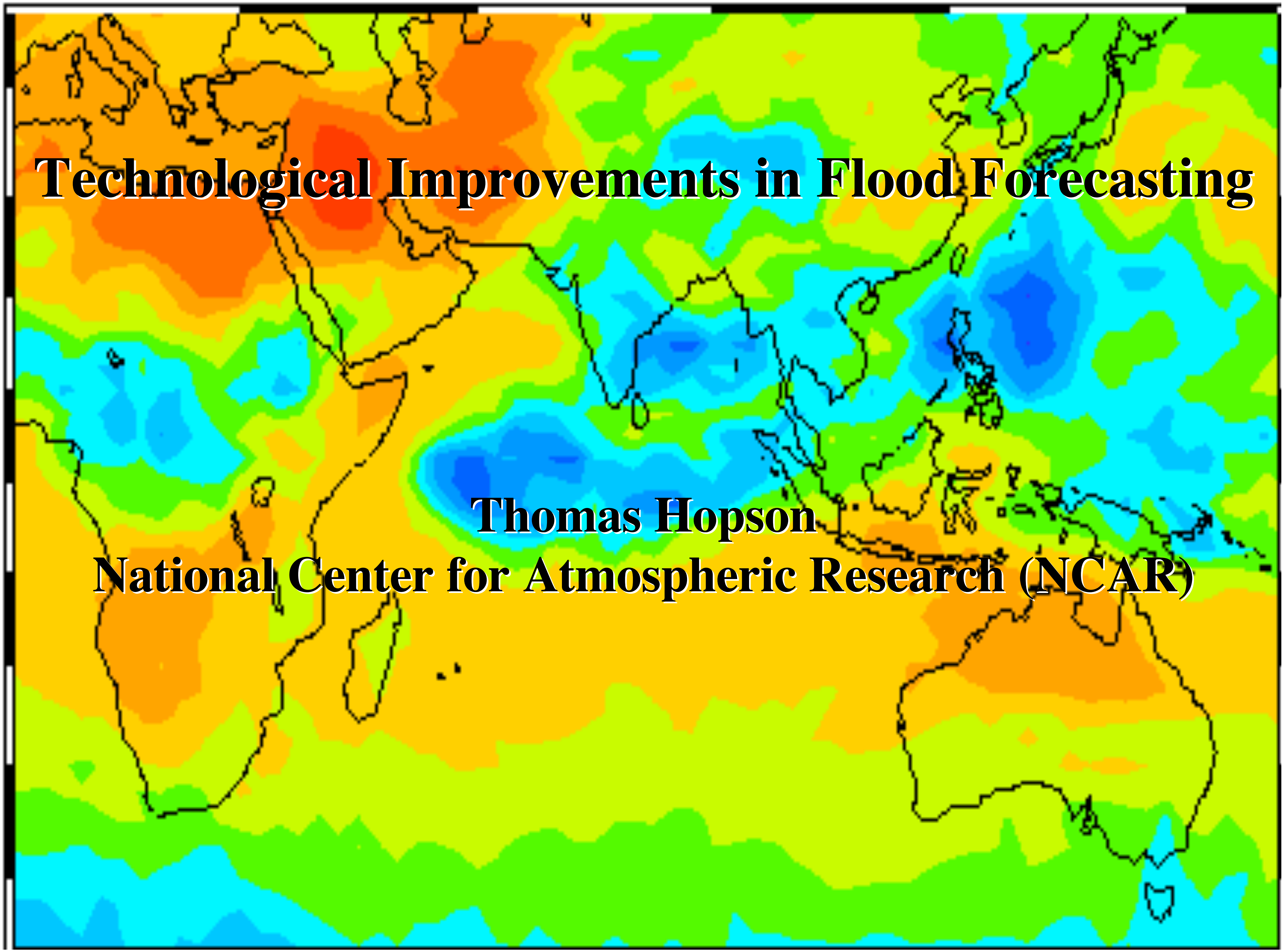
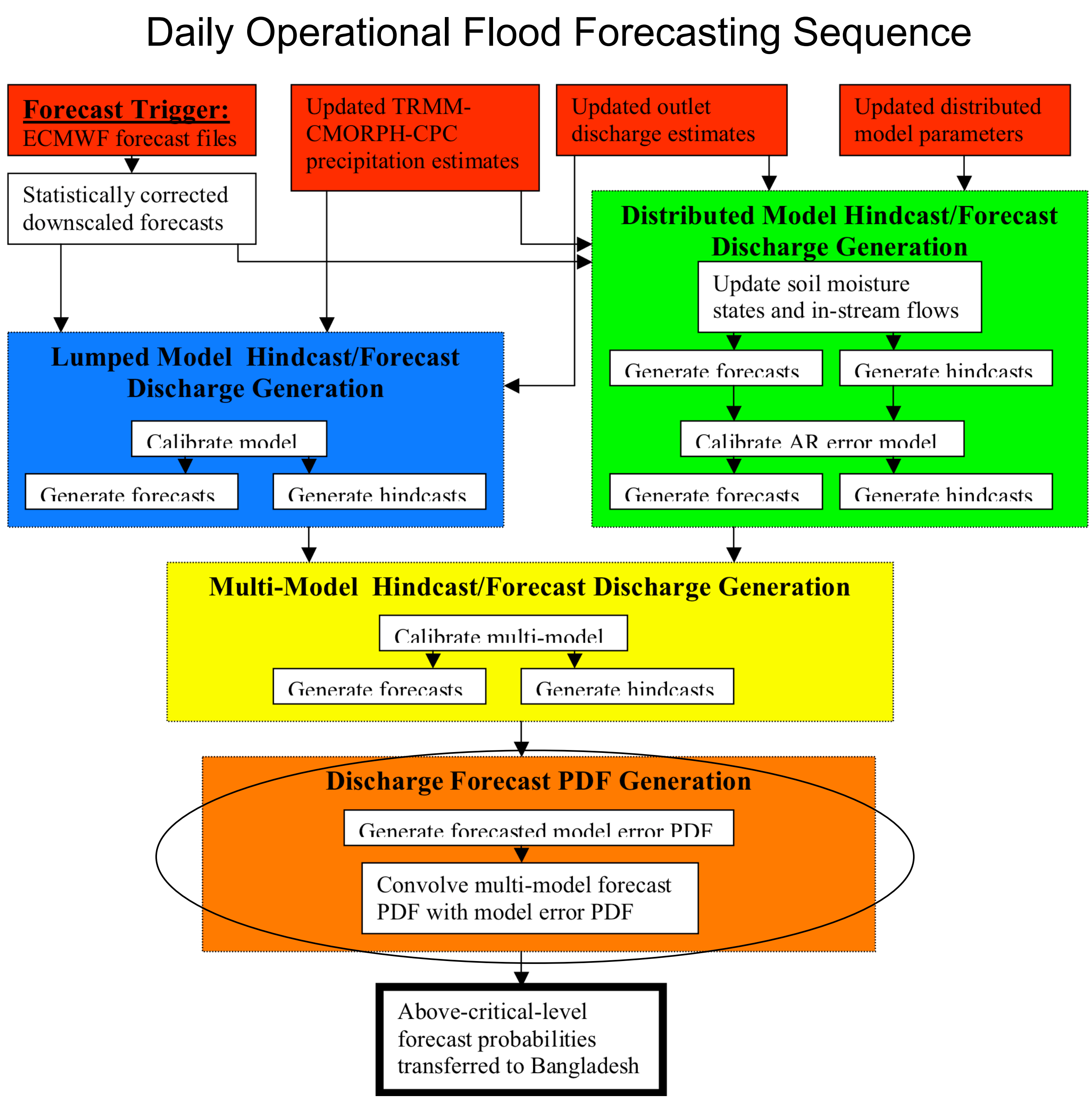


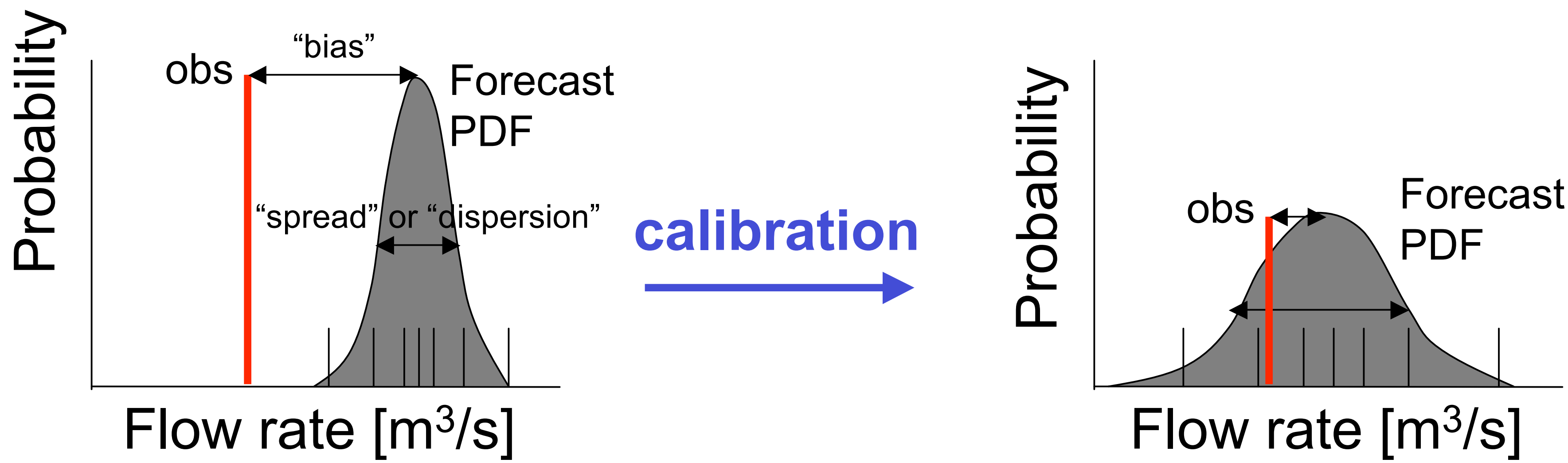
Technological Improvements in Flood Forecasting

Thomas Hopson
National Center for Atmospheric Research (NCAR)



Daily Operational Flood Forecasting Sequence
Forecast Trigger:
ECMWF forecast files
Updated TRMM-CMORPH-CPC precipitation estimates
Updated outlet discharge estimates
Updated distributed model parameters
Statistically corrected downscaled forecasts
Distributed Model Hindcast/Forecast Discharge Generation
Update soil moisture states and in-stream flows
Generate forecasts
Generate hindcasts
Calibrate AR error model
Generate forecasts
Generate hindcasts
Lumped Model Hindcast/Forecast Discharge Generation
Calibrate model
Generate forecasts
Generate hindcasts
Multi-Model Hindcast/Forecast Discharge Generation
Calibrate multi-model
Generate forecasts
Generate hindcasts
Discharge Forecast PDF Generation
Generate forecasted model error PDF
Convolve multi-model forecast PDF with model error PDF
Above-critical-level forecast probabilities transferred to Bangladesh

Final flood forecast "calibration" or "post-processing"



<u>Post-processing has corrected:</u>

• the "on average" bias

• as well as under-representation of the 2nd moment of the empirical forecast PDF (i.e. corrected its "dispersion" or "spread")

<u>Our approach:</u>

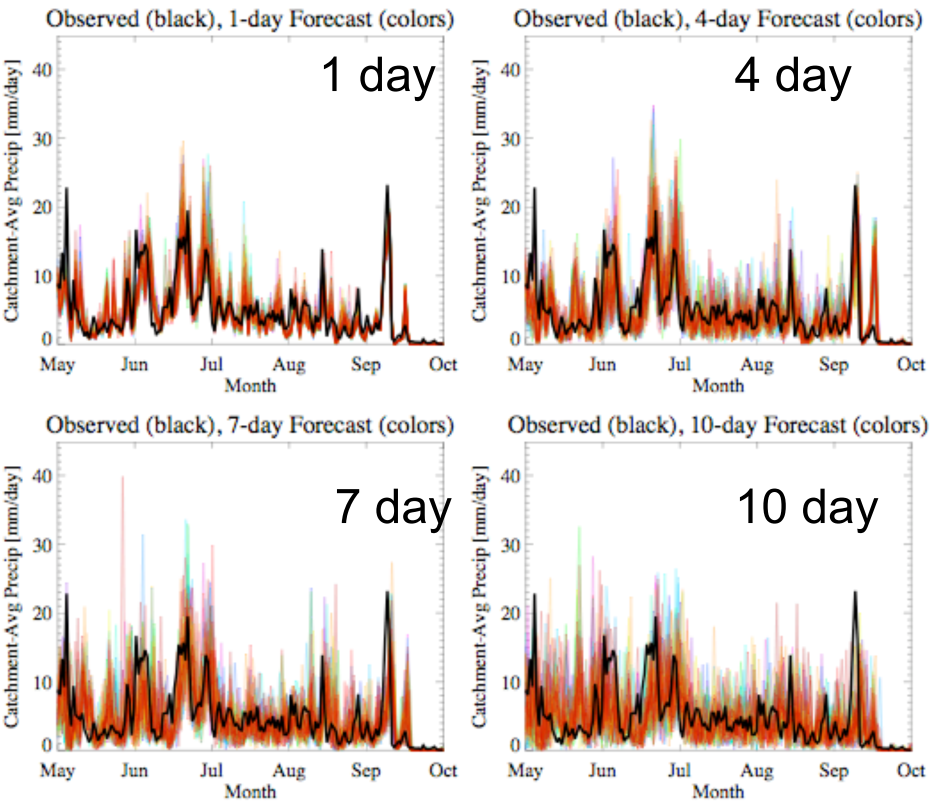
• under-utilized "quantile regression" approach

• probability distribution function "means what it says"

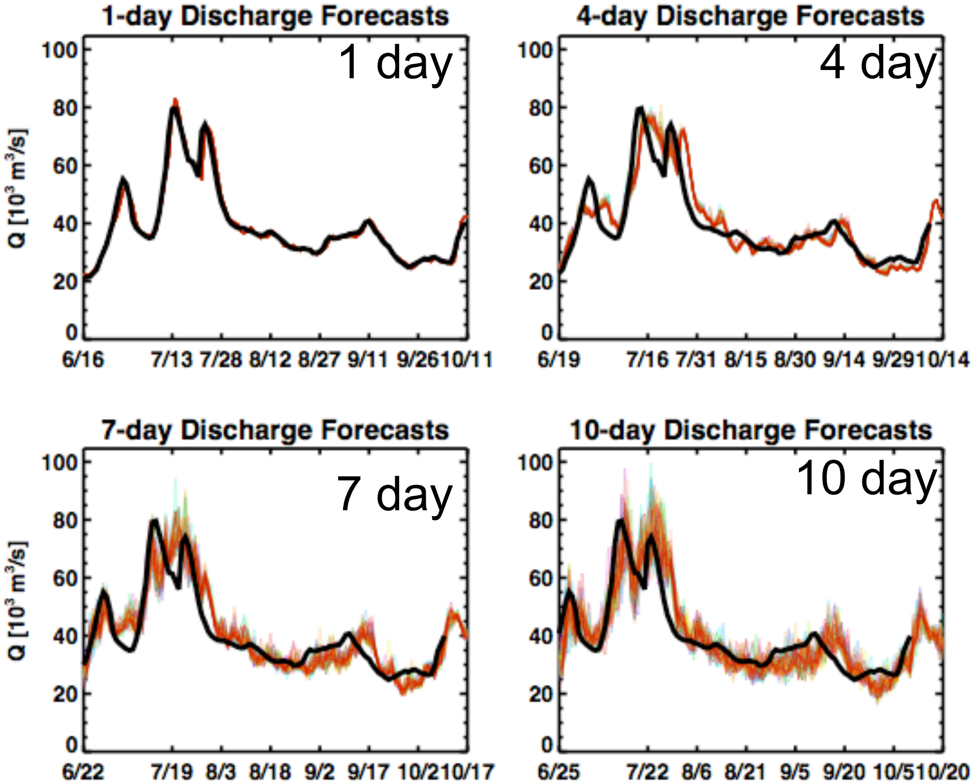
• daily variation in the ensemble dispersion directly relate to changes in forecast skill

Significance of Weather Forecast Uncertainty Discharge Forecasts

Precipitation Forecasts

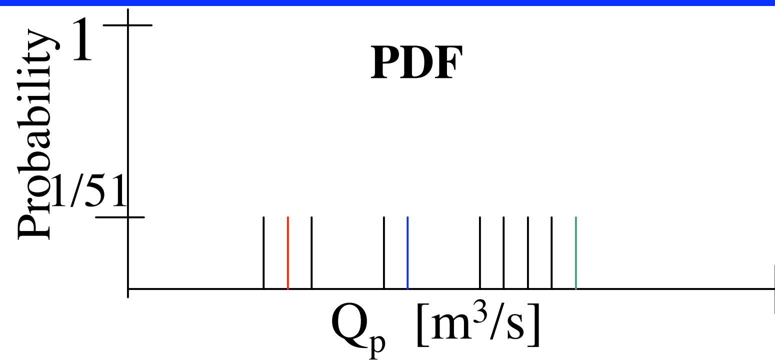


Discharge Forecasts

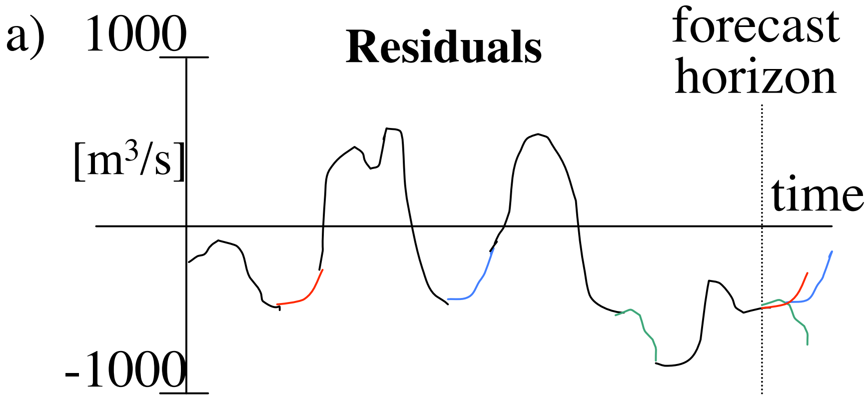


Producing a Reliable Probabilistic Discharge Forecast

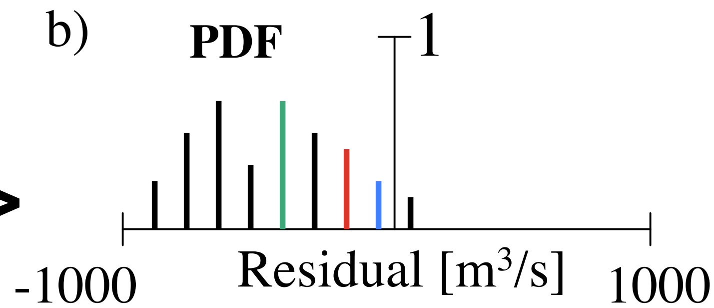
Step 1: generate discharge ensembles from precipitation forecast ensembles (Q_p):



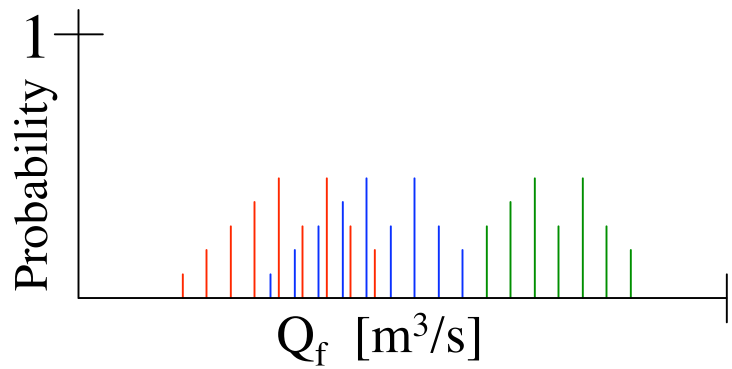
Step 2: a) generate multi-model hindcast error time-series using precip estimates; b) conditionally sample and weight to produce empirical forecasted error PDF:



=>

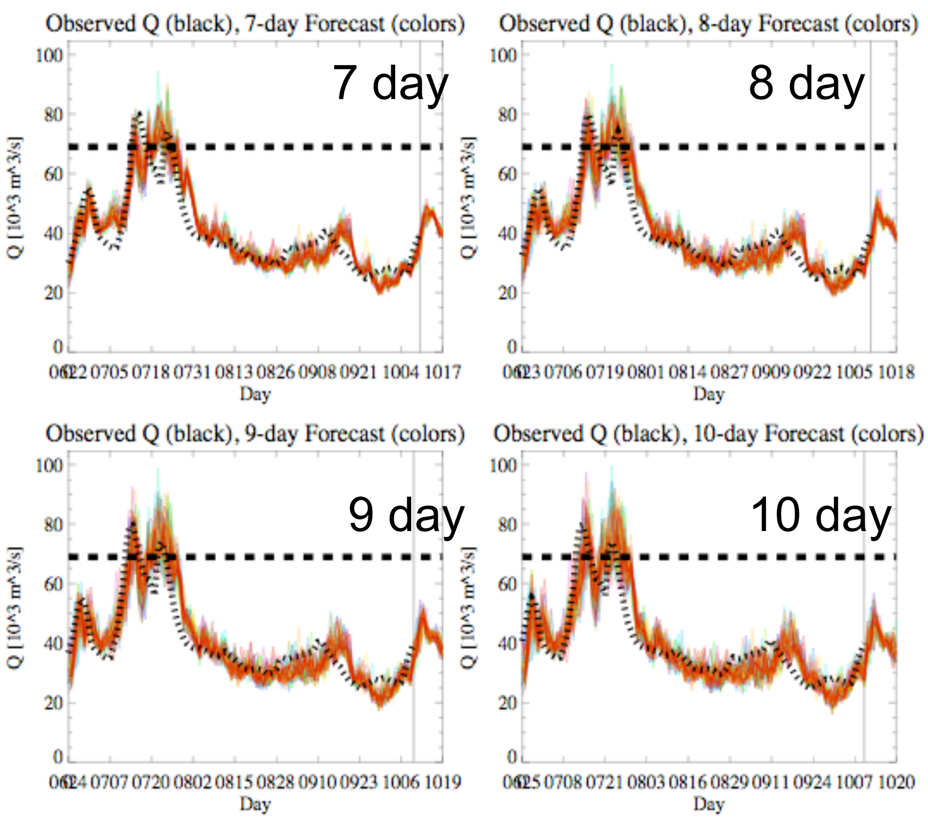


Step 3: combine both uncertainty PDF's to generate a "new-and-improved" more complete PDF for forecasting (Q_f):

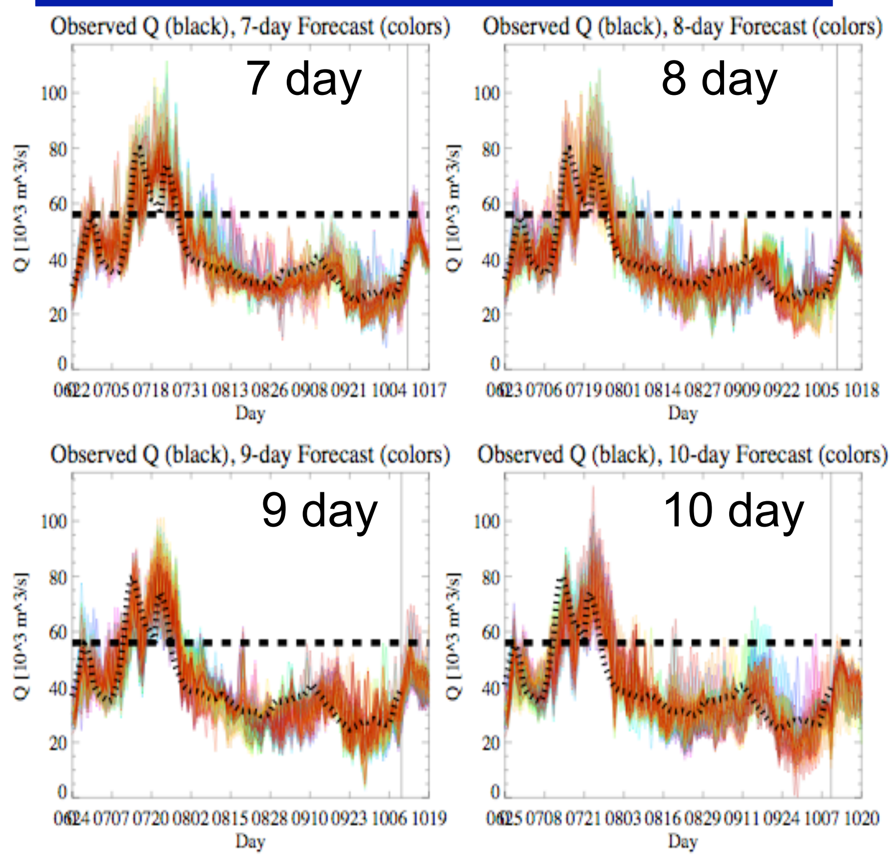


Significance of Weather Forecast Uncertainty Discharge Forecasts

2004 Brahmaputra Discharge Forecast Ensembles

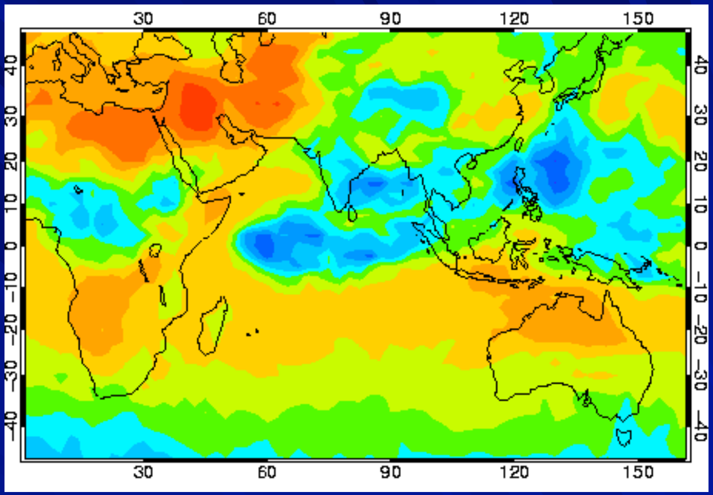


Corrected Forecast Ensembles



Overview:

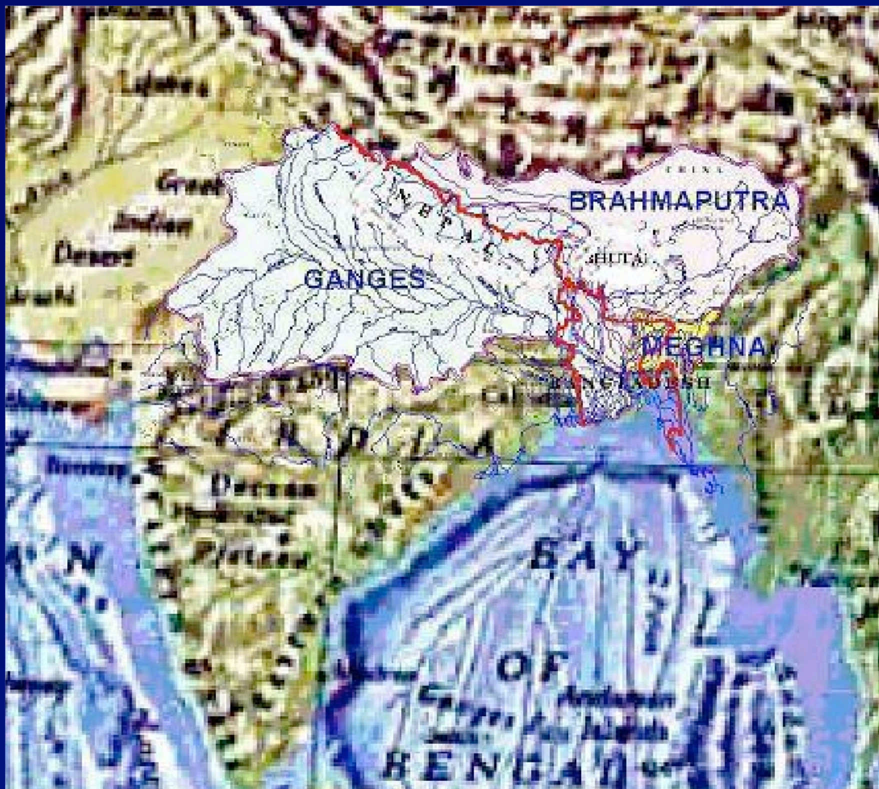
Technological improvements in flood forecasting



I. Future improvements: remotely-sensed river discharge
   - Dartmouth Flood Observatory
   - GRACE satellite system

II. Multi-Model or Post-processing: Pros and Cons

CFAB Project: Improve flood warning lead time



Problems:

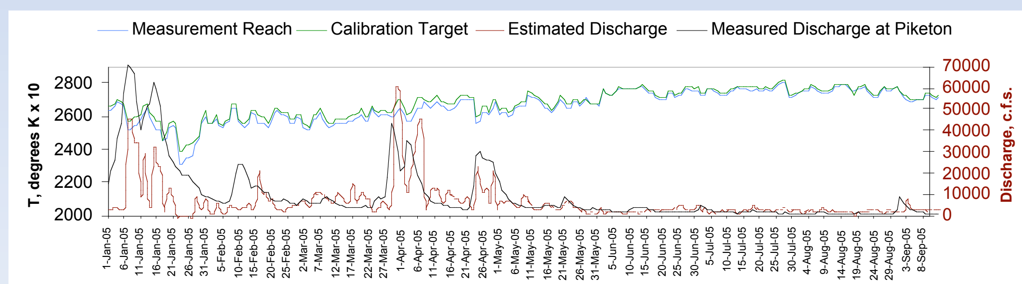
1. Limited warning of upstream river discharges

2. Precipitation forecasting in tropics difficult

Good forecasting skill derived from:
1. good data inputs: ECMWF weather forecasts, satellite rainfall
2. Large catchments => weather forecasting skill “integrates” over large spatial and temporal scales
3. Partnership with Bangladesh’s Flood Forecasting Warning Centre (FFWC) => daily border river readings used in data assimilation scheme

Satellite-based River Discharge Estimation

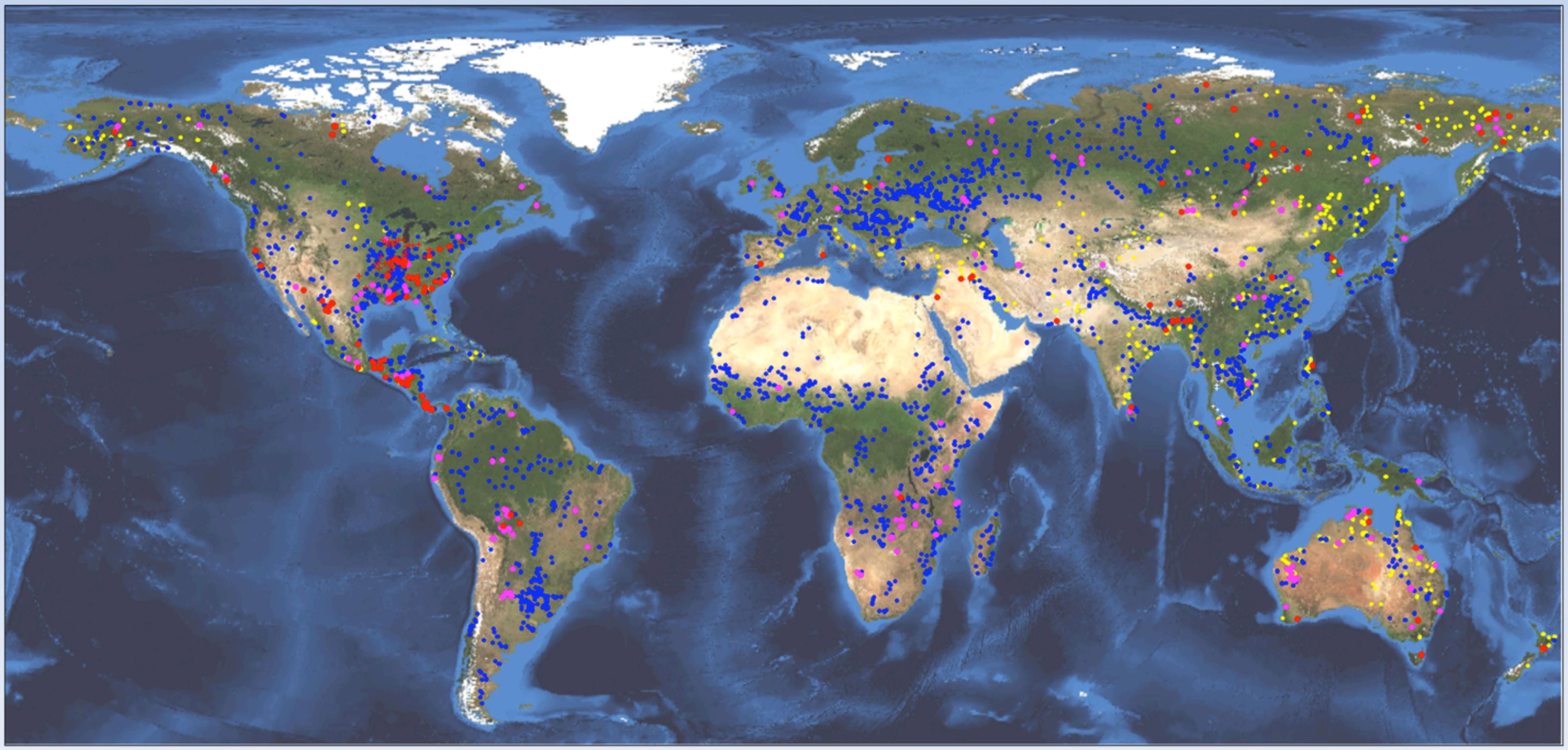
Bob Brakenridge, Dartmouth Flood Observatory, Dartmouth College



http://www.dartmouth.edu/~floods/

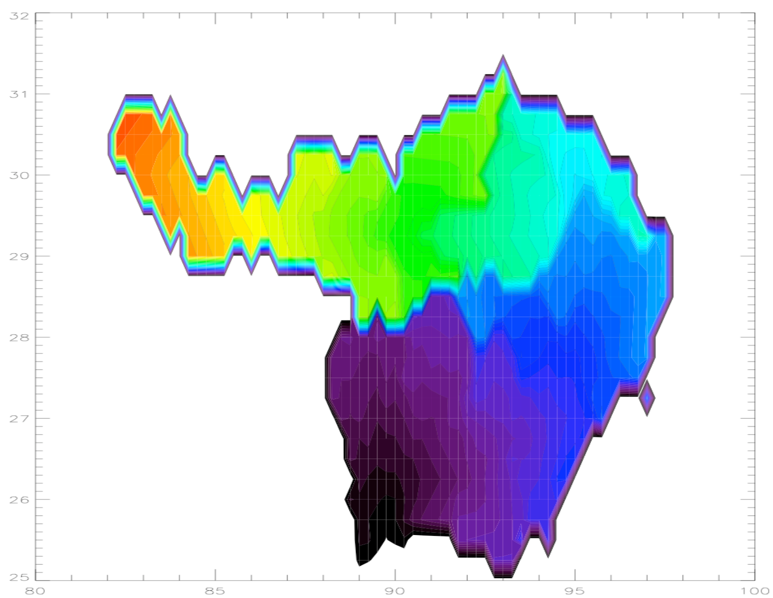
River Watch

- •Day/Night Flood detection on a near-daily basis regardless of cloud cover.
- •Measurement of river discharge changes; current flood magnitude assessments
- •Immediate map-based prediction of what is under water
- •Access to rapid response detailed mapping as new maps are made
- •Access to map data base of previous flooding and associated recurrence intervals.

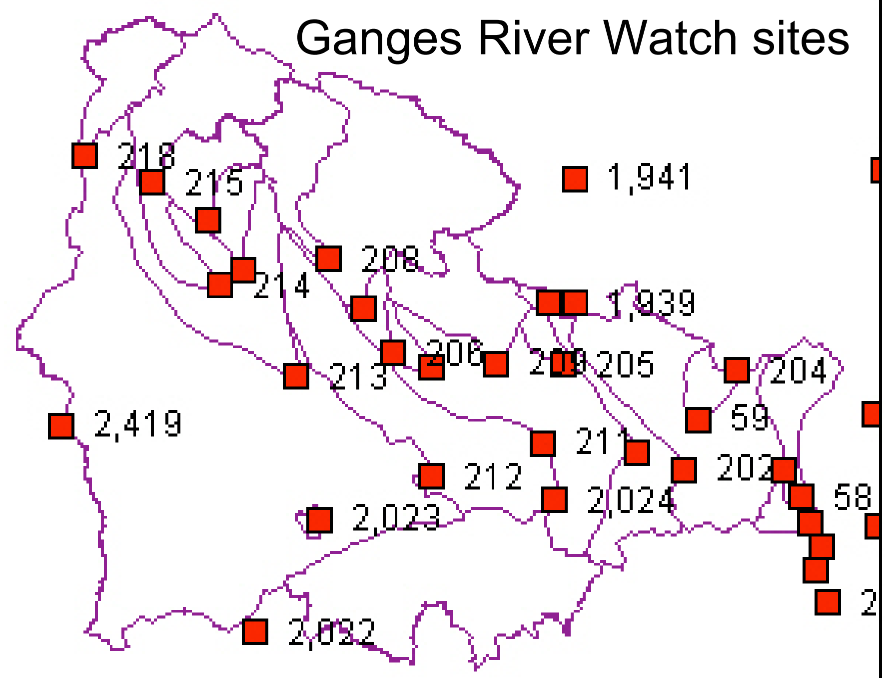


Application to the Ganges and Brahmaputra Rivers

Brahmaputra floodwave isochrons



Ganges River Watch sites



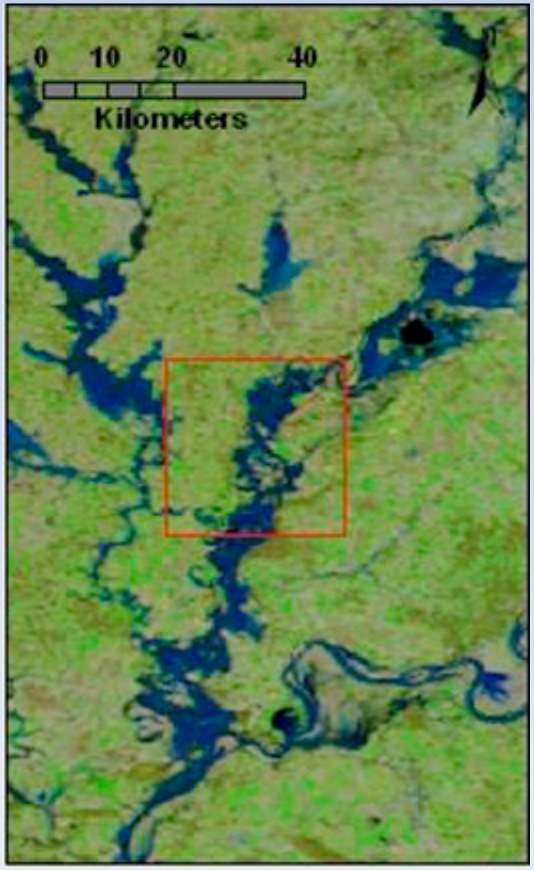
<u>Utility of River Watch discharge estimates to flood forecasting:</u>

1) Calibration of ungauged subcatchments outflow and routing
2) Operational improvements through data assimilation
   -- blending of enKF, 4DVAR, and “quantile regression”

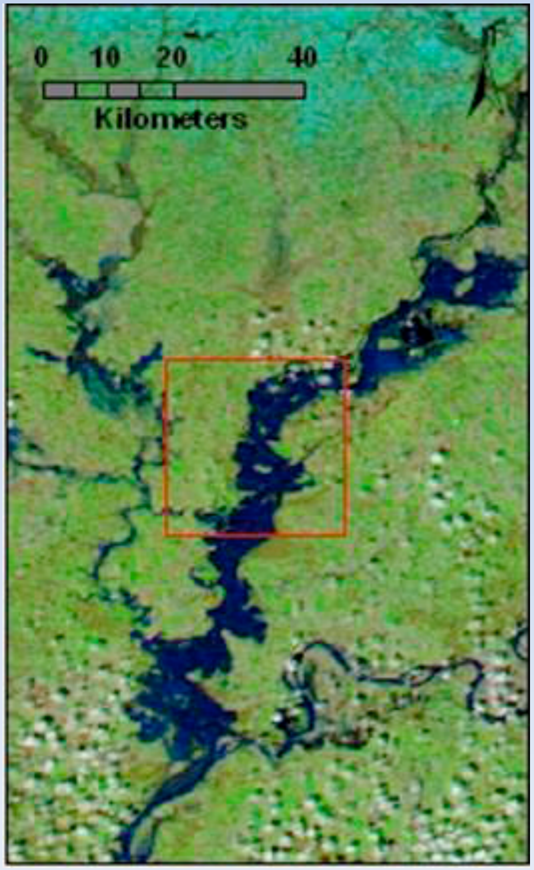
MODIS sequence of 2006 Winter Flooding

2/24/2006 C/M: 1.004

3/15/2006 C/M: 1.029

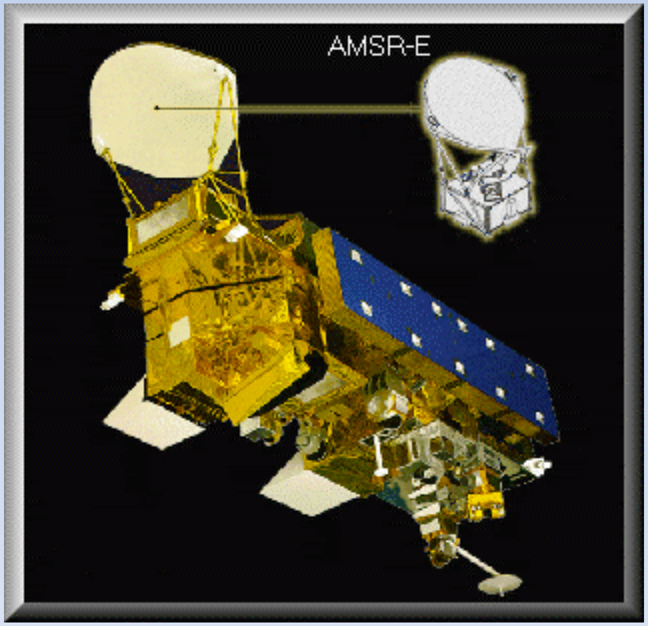


3/22/2006 C/M: 1.095



Objective Monitoring of River Status: The Microwave Solution

The Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) is a twelve-channel, six-frequency, passive-microwave radiometer system. It measures horizontally and vertically polarized brightness temperatures at 6.9 GHz, 10.7 GHz, 18.7 GHz, 23.8 GHz, 36.5 GHz, and 89.0 GHz.

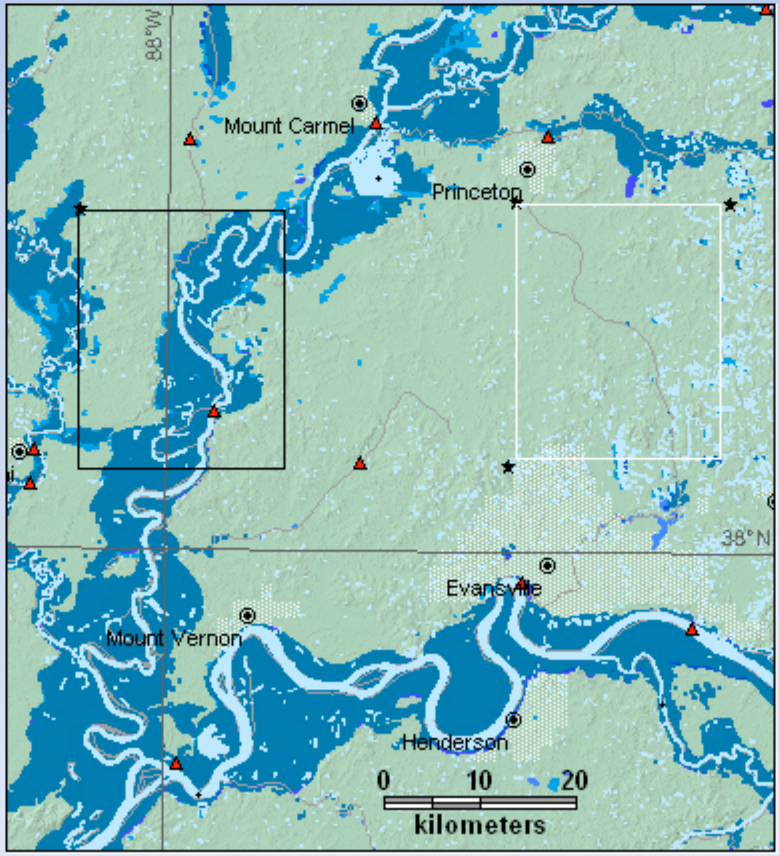


Spatial resolution of the individual measurements varies from 5.4 km at 89 GHz to 56 km at 6.9 GHz.
AMSR-E was developed by the Japan Aerospace Exploration Agency (JAXA) and launched by the U.S. aboard Aqua in mid-2002.

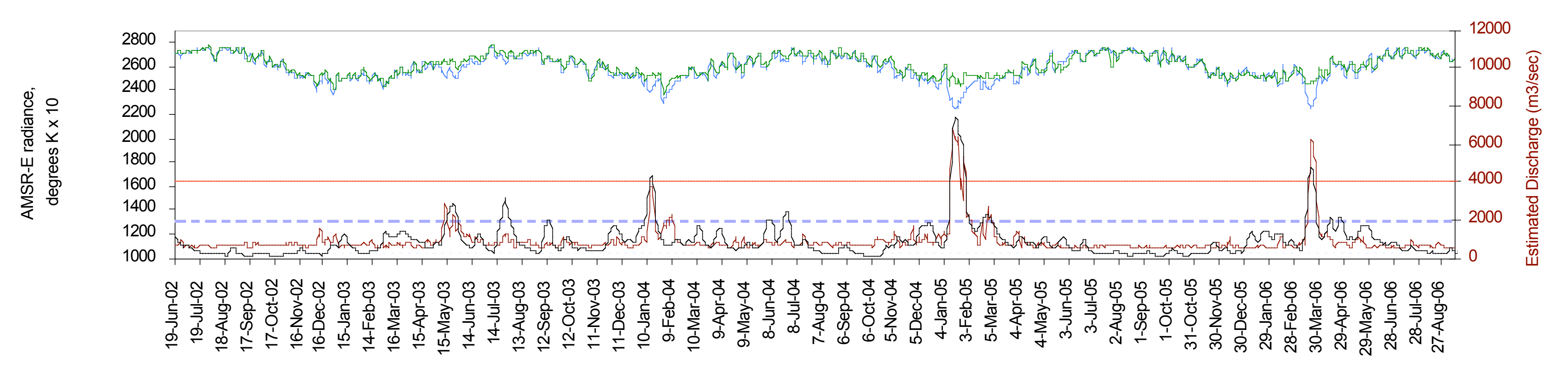
One day of data collection
(high latitudes revisited most frequently)

Example: Wabash River near Mount Carmel, Indiana, USA

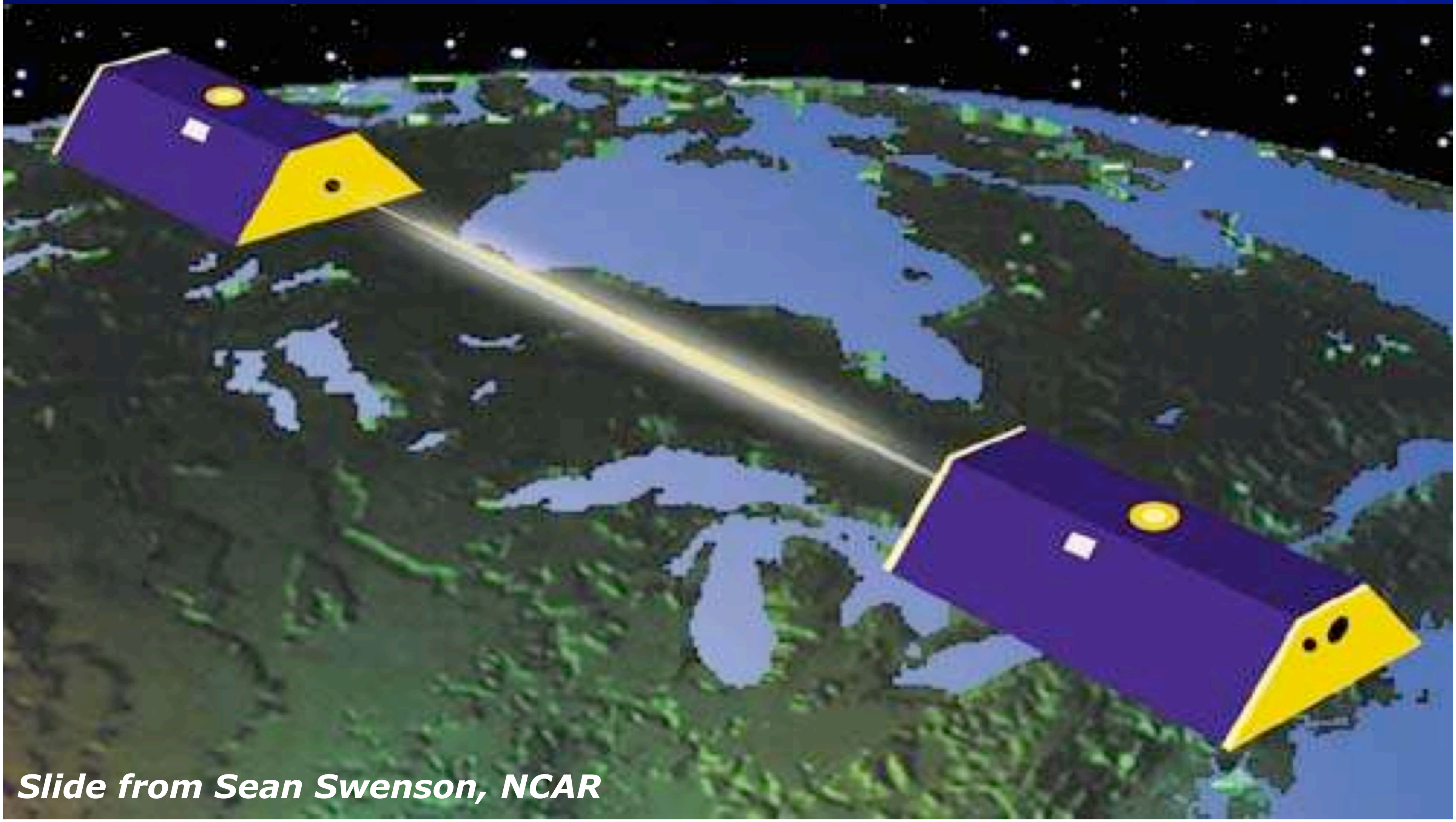
Black square shows Measurement pixel. White square is calibration pixel.



Site 98, Wabash River at New Harmony, Indiana, USA

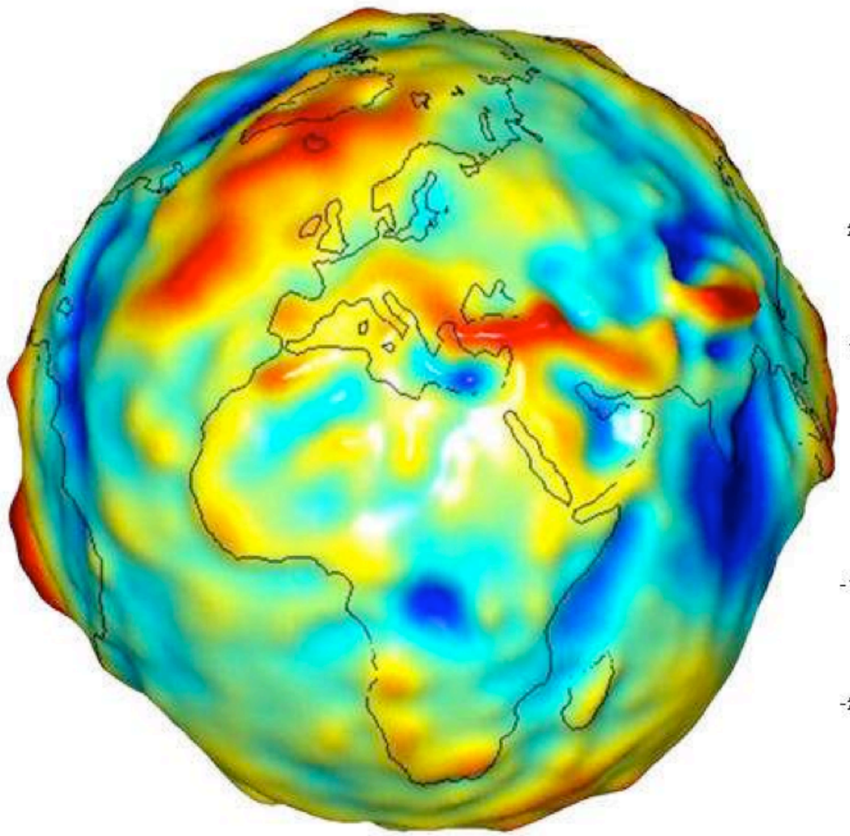


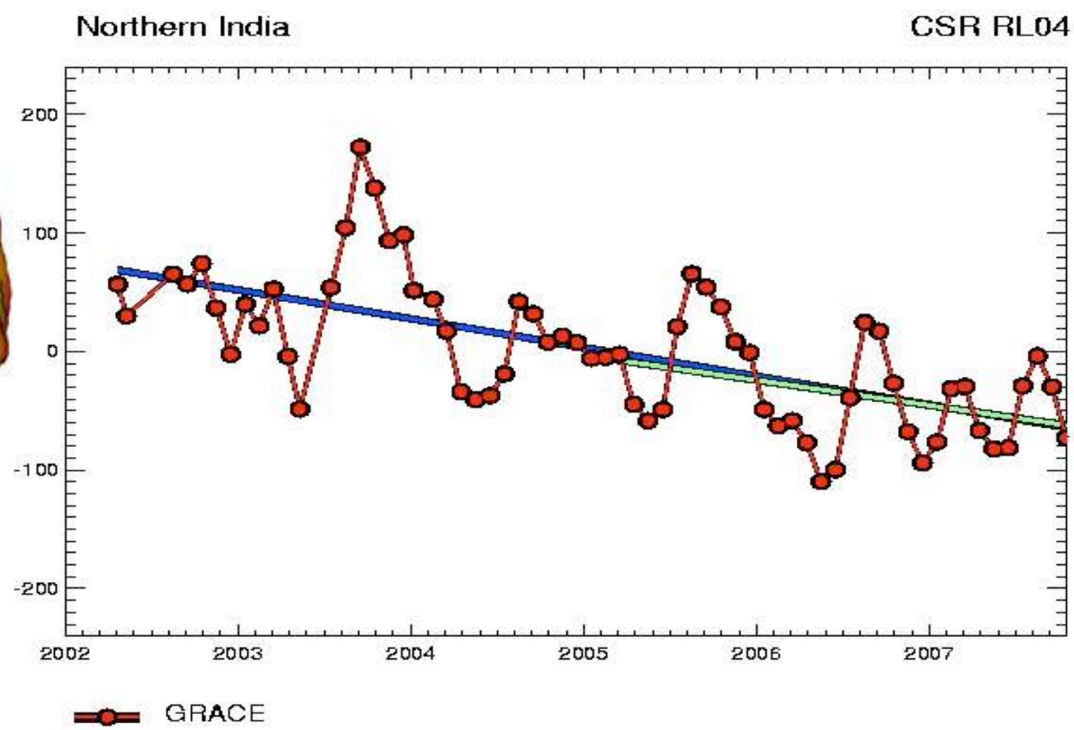
Gravity Recovery And Climate Experiment (GRACE)



Slide from Sean Swenson, NCAR

Northern India Time Series





GRACE catchment-integrated soil moisture estimates useful for:

1) Hydrologic model calibration and validation
2) Seasonal forecasting
3) Data assimilation for medium-range (1-2 week) forecasts

Slide from Sean Swenson, NCAR

Conclusions

<u>Further Advances:</u>

- Data assimilation of new satellite-derived products:
  -- Dartmouth Flood Observatory river discharge estimates
  -- GRACE integrated catchment soil moisture
  -- QSCAT and TMI soil moisture estimates (Nghiem, JPL)

- Expansion of multi-model approach (78 member multi-model)

- Daily-updated seamless weather-to-seasonal flood forecasting:
  -- utilizing short-, medium-, monthly-, and seasonal ensemble forecasts

Multi-Model or Post-processing: Pros and Cons

Tom Hopson - NCAR
Martyn Clark - NIWA
Andrew Slater - CIRES/NSIDC

Question:

How best to utilize a multi-model simulation (forecast), especially if under-dispersive?

a) Should more dynamical variability be searched for? Or

b) Is it better to balance post-processing with multi-model utilization to create a properly dispersive, informative ensemble?

Outline

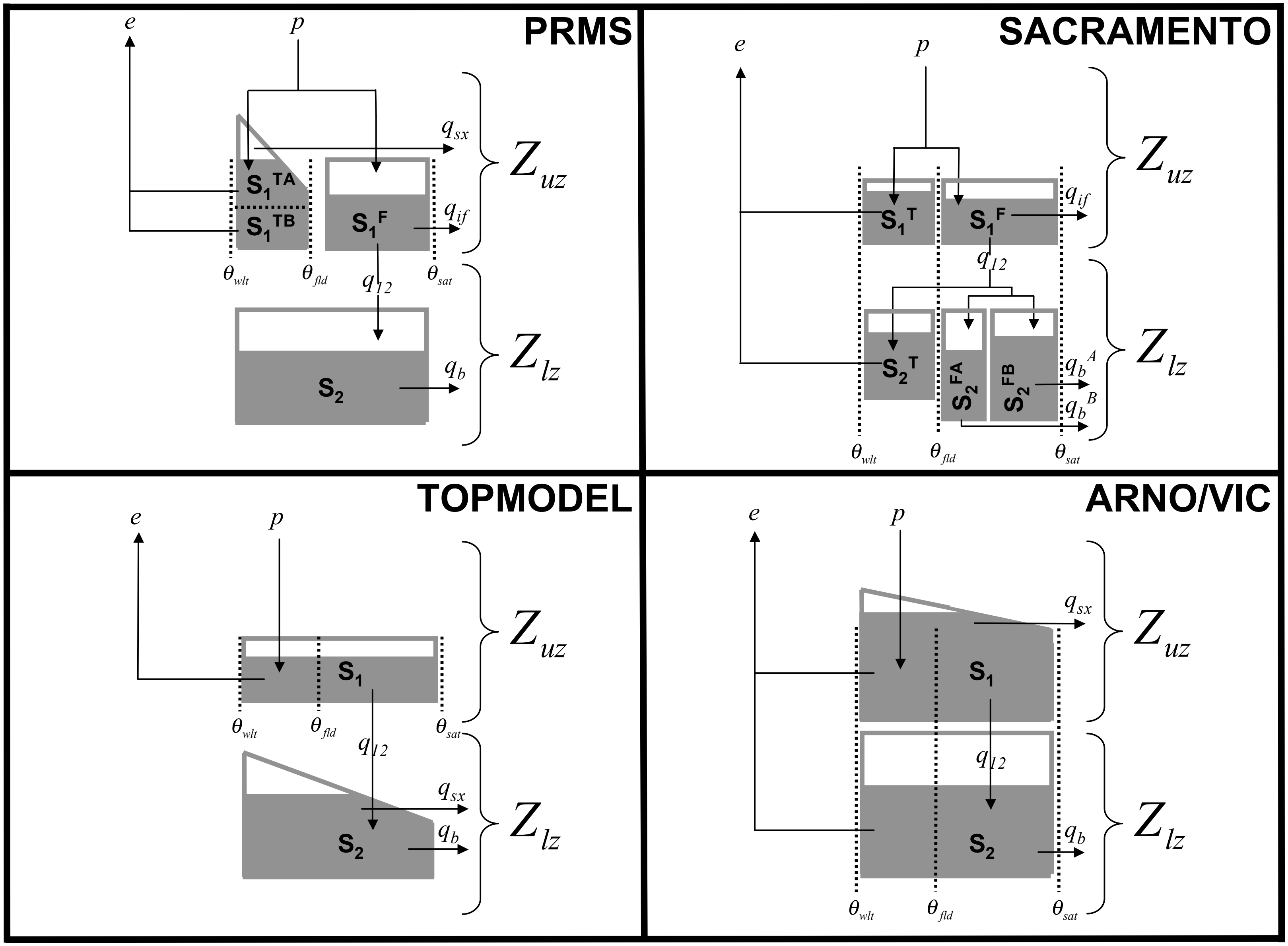
Explore this question using multi-model simulations for the French Broad River, NC of MOPEX

I. Multi-model: Framework for Understanding Structural Errors (FUSE)
   - Pre-calibration results => under-dispersive

II. Calibration procedure
   - Introduce Quantile Regression (“QR”; Kroenker and Bassett, 1978)

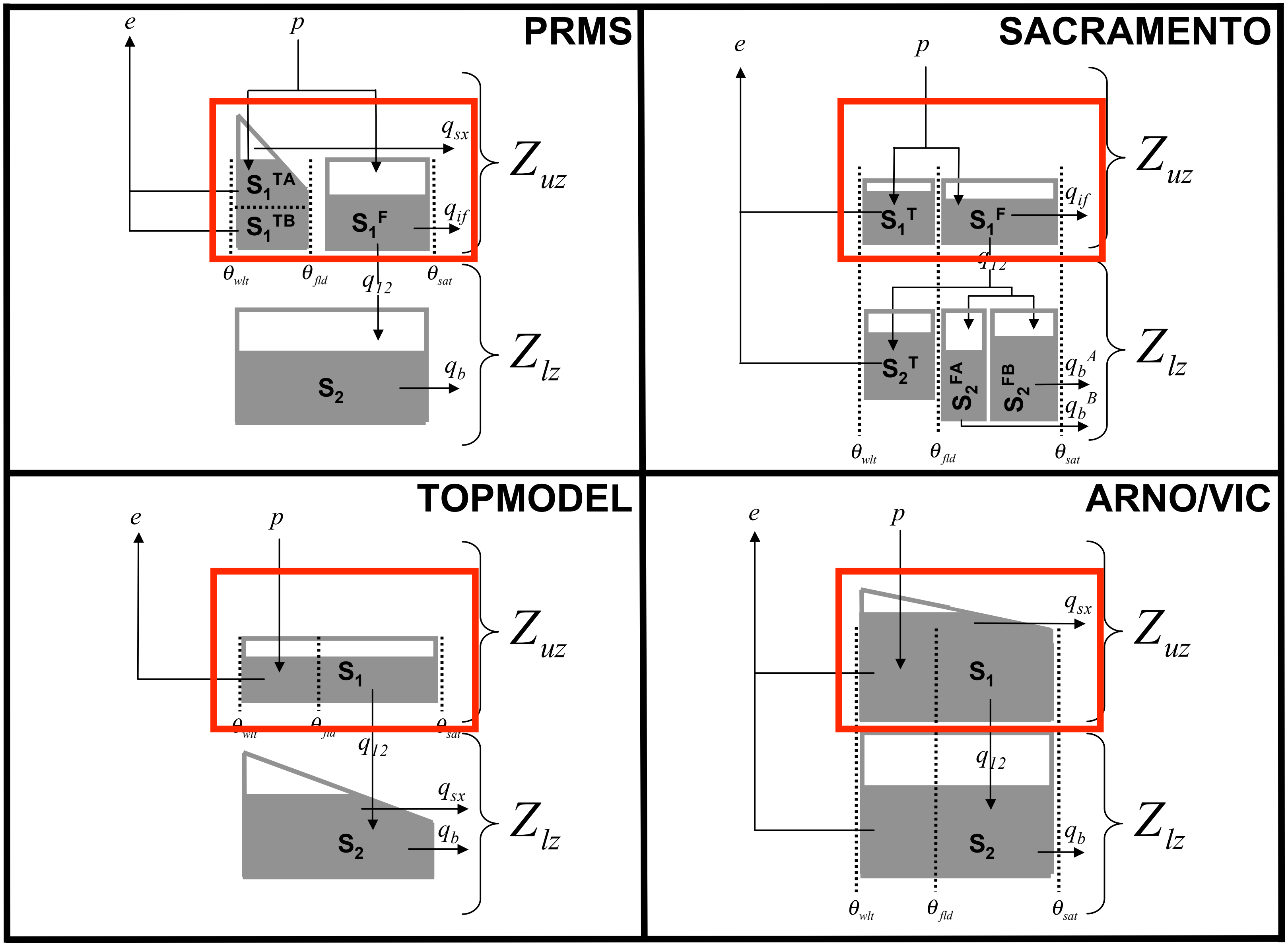
III. Discussing of Question -- how best to utilize multi-model

FUSE: Framework for Understanding Structural Errors



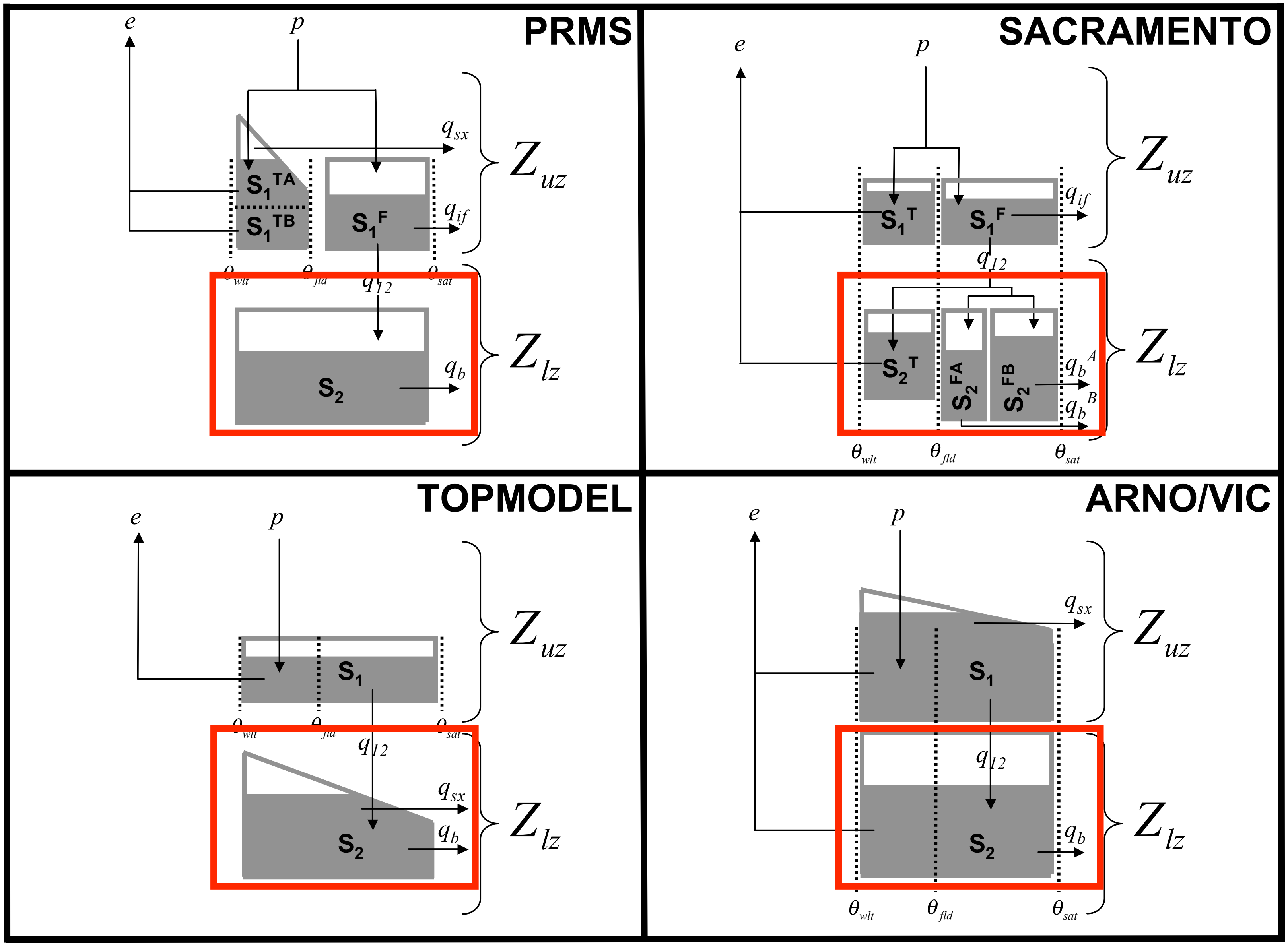
Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

Define development decisions: upper layer architecture



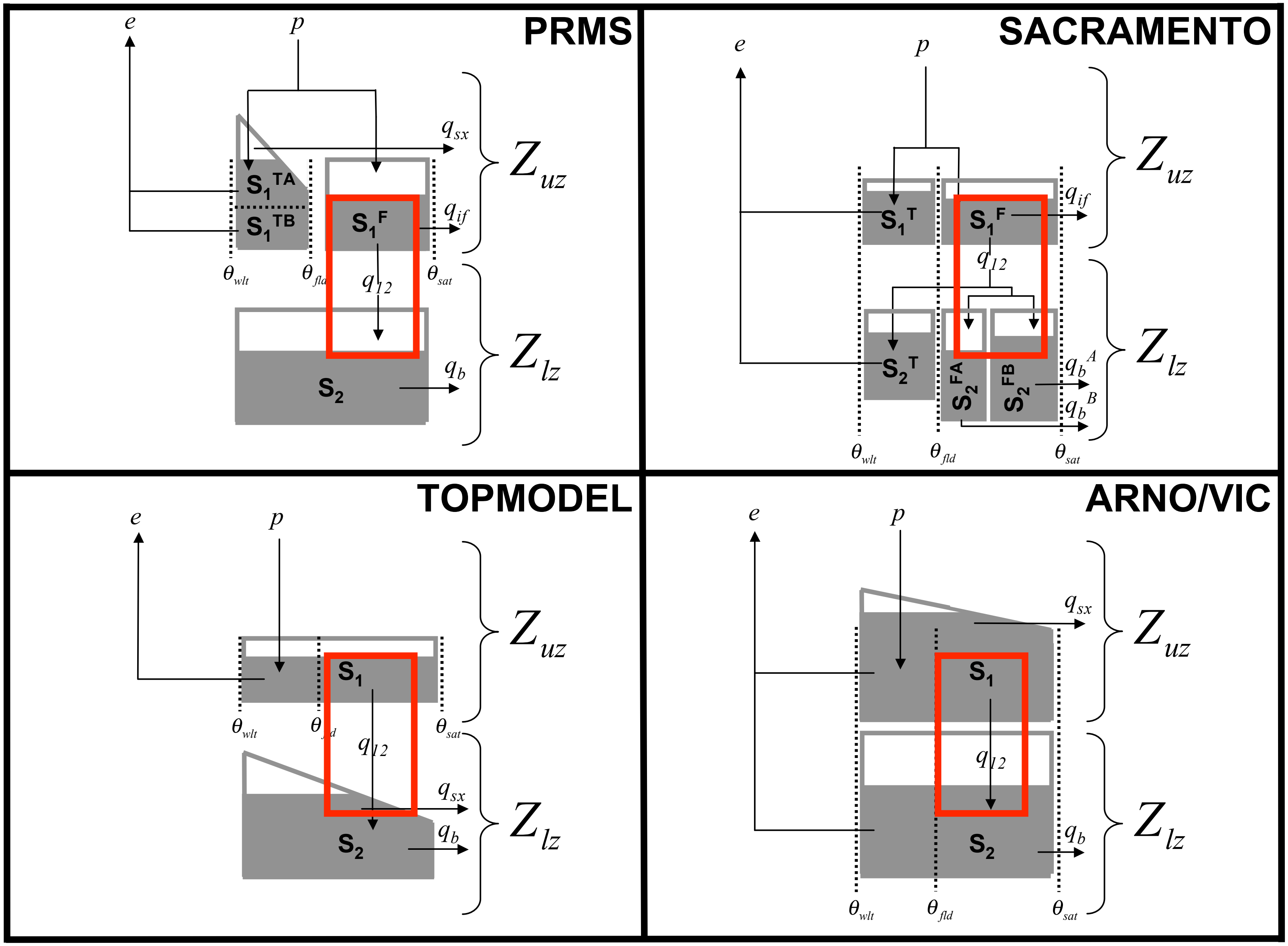
Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

Define development decisions: lower layer / baseflow



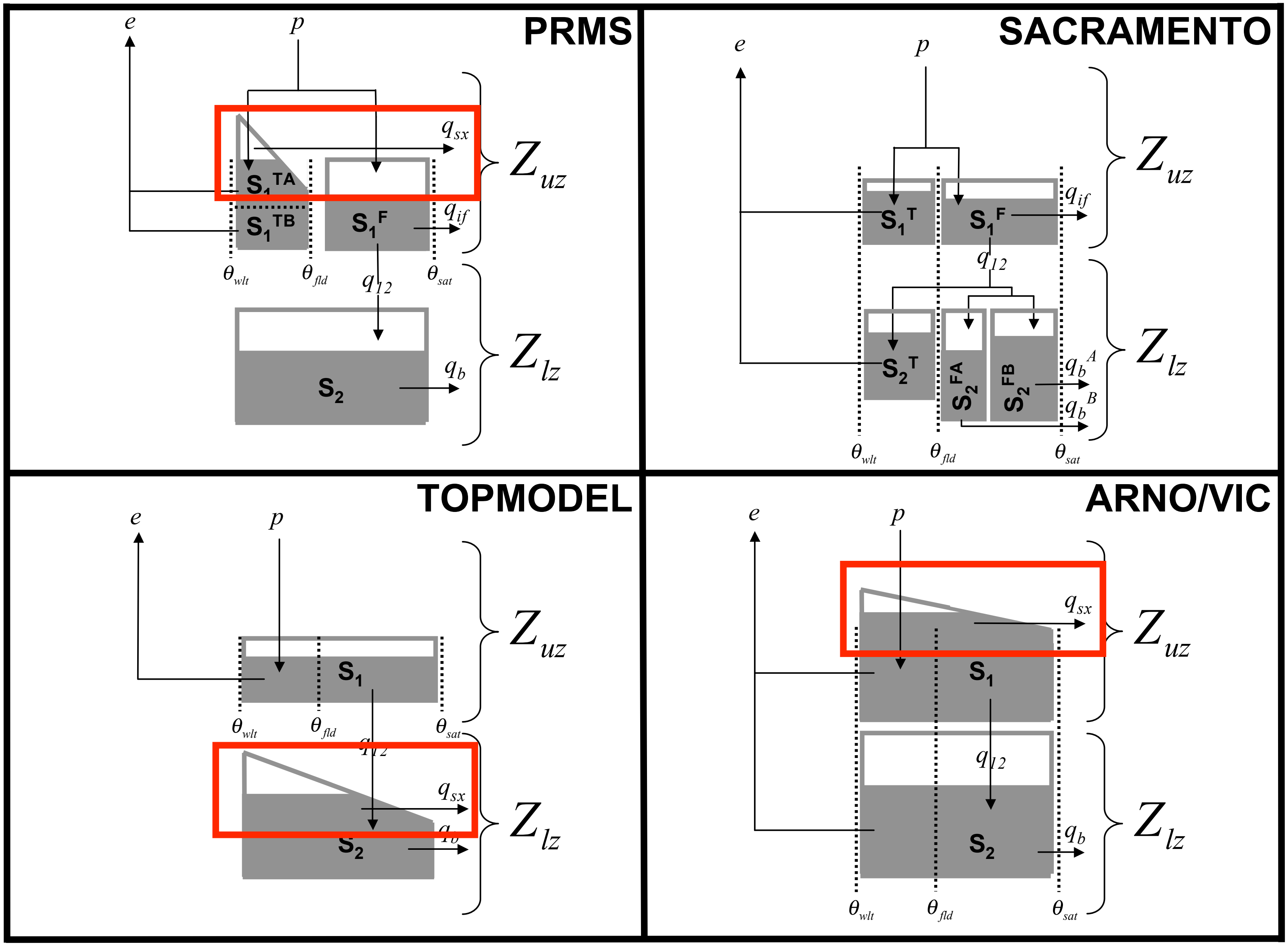
Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

Define development decisions: percolation



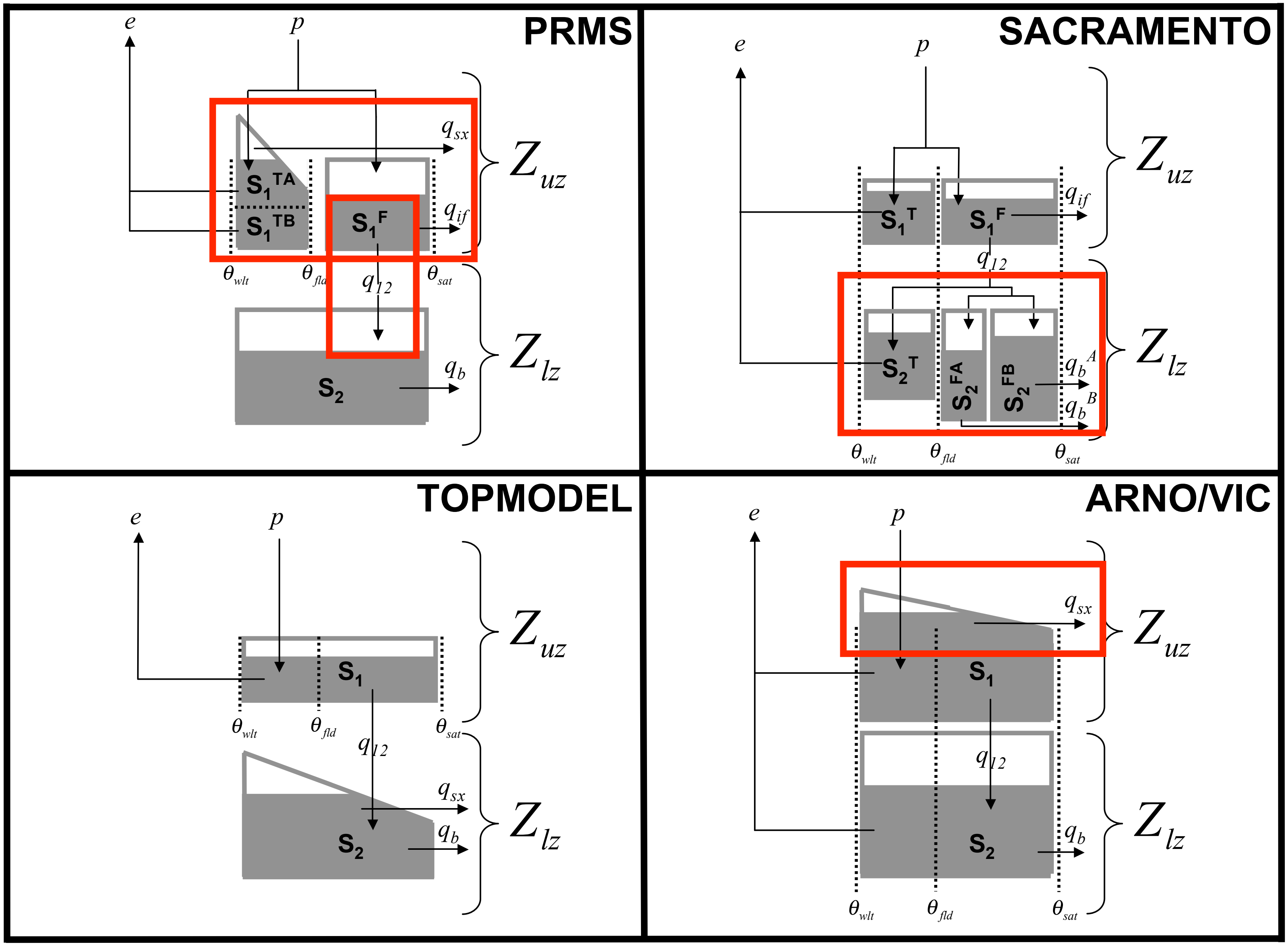
Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

Define development decisions: surface runoff



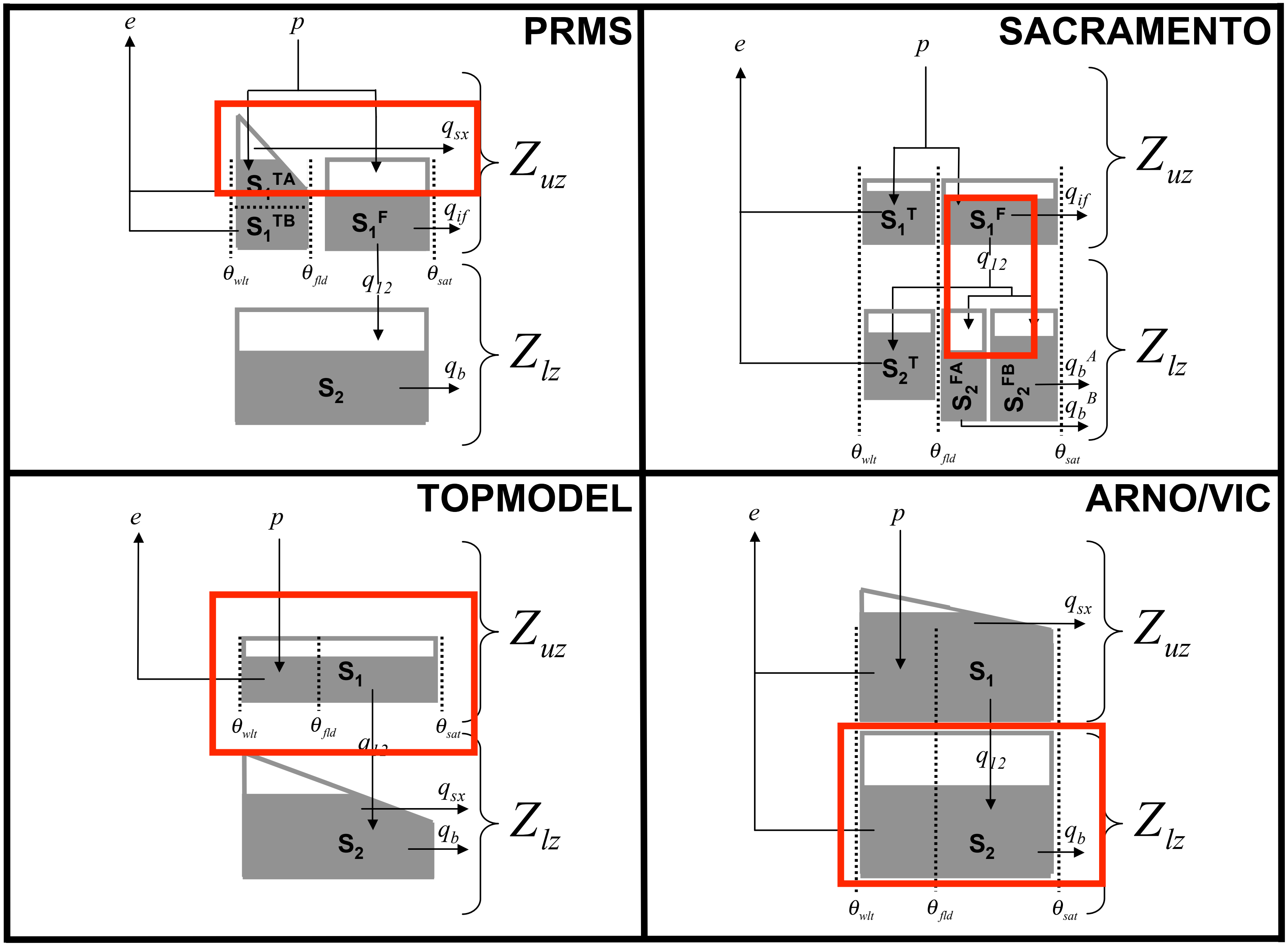
Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

Build unique models: combination 1



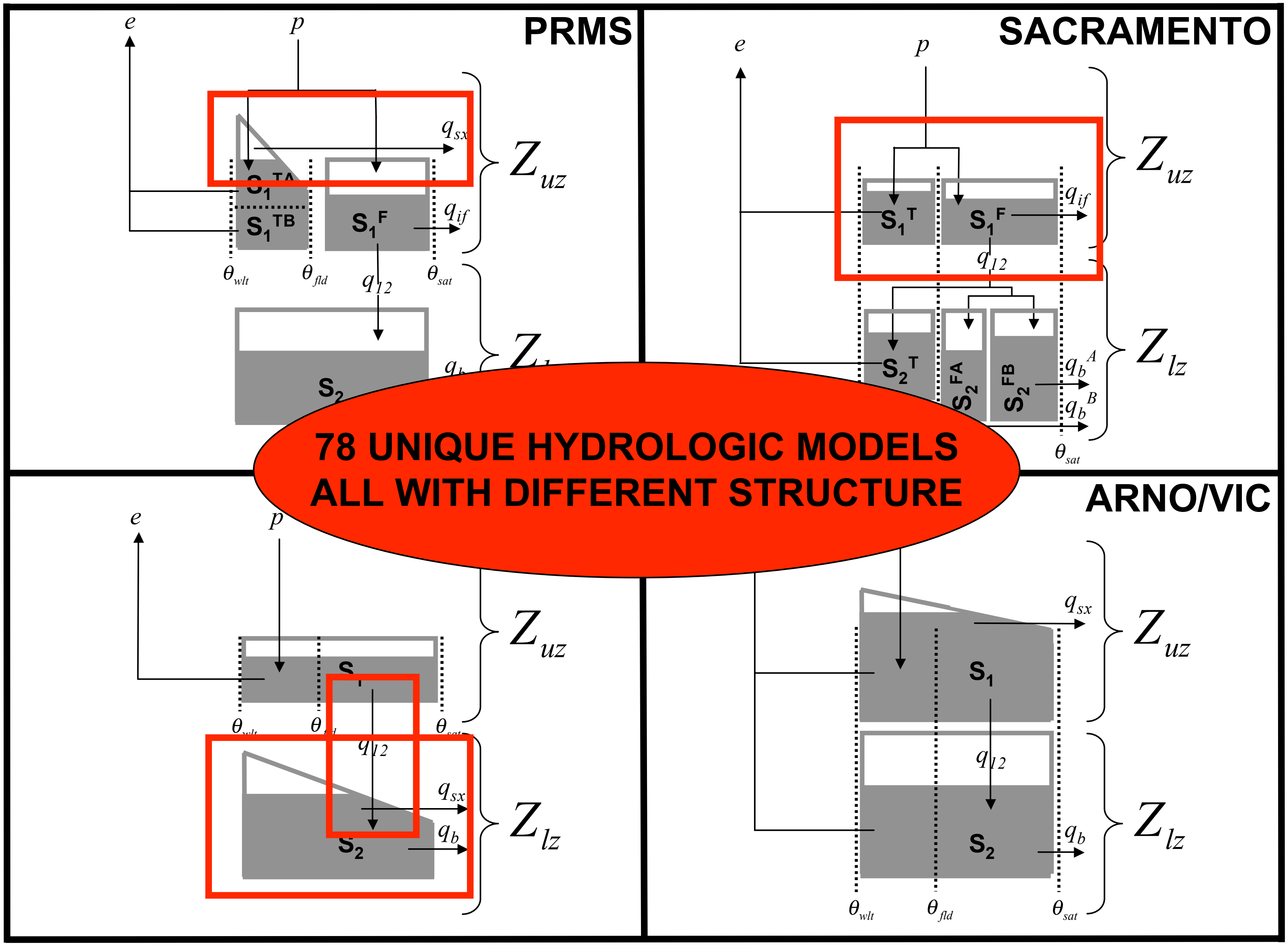
Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

Build unique models: combination 2



Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

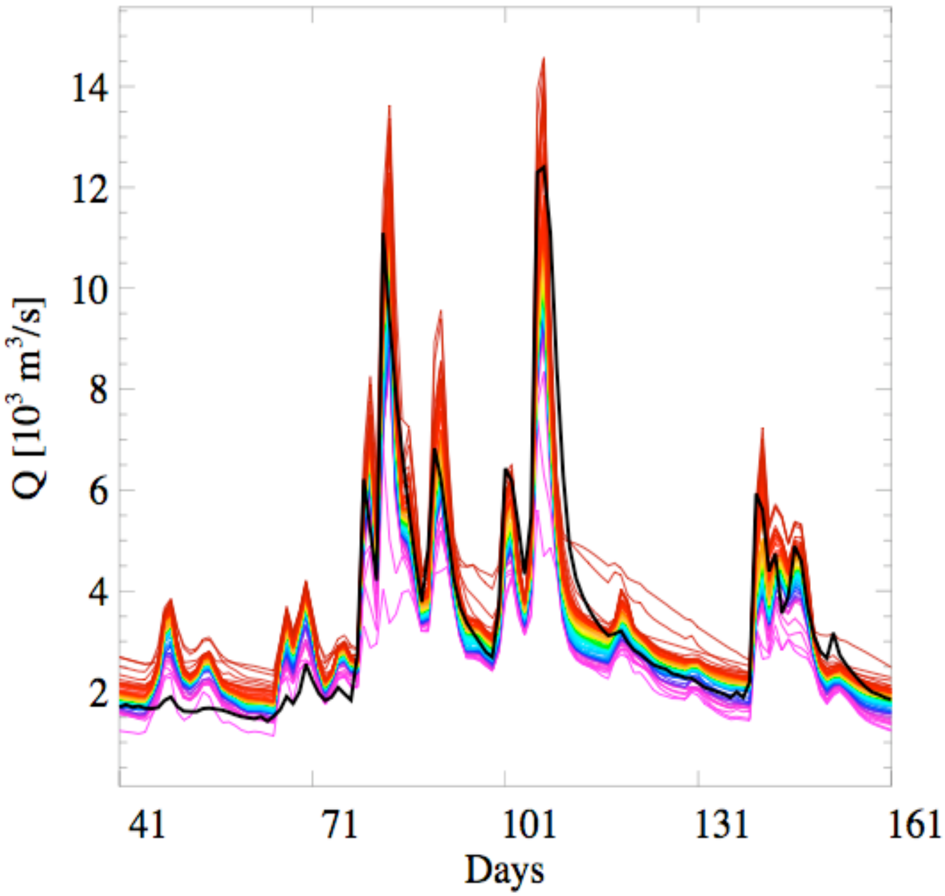
Build unique models: combination 3

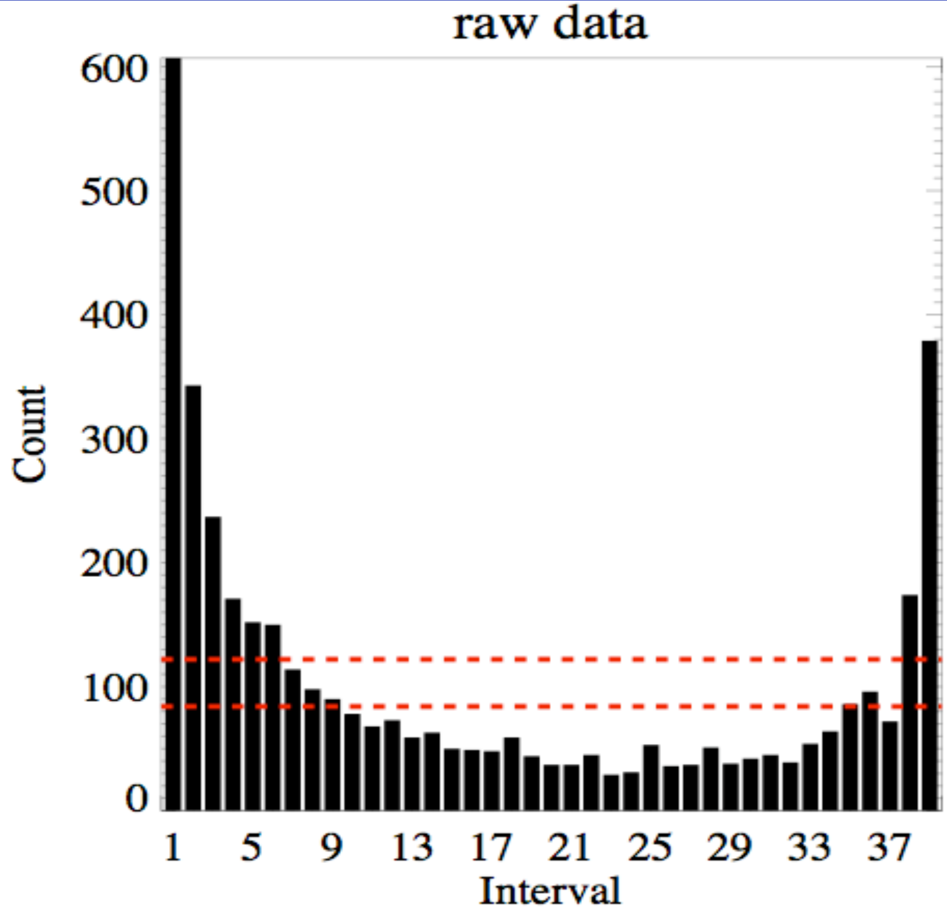


Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay (2008) Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research,* 44, W00B02, doi:10.1029/2007WR006735.

Example: French Broad River

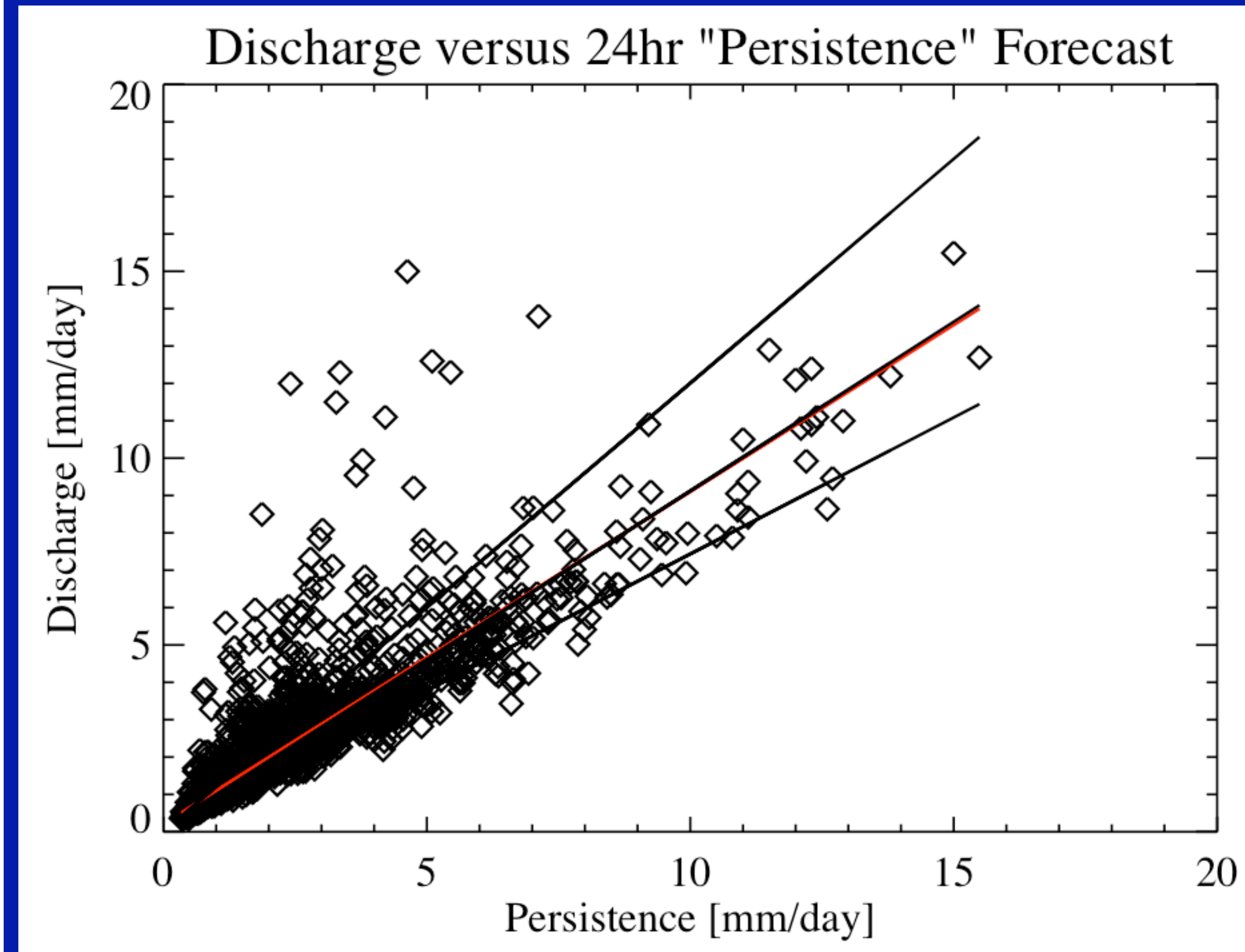
Before Calibration => underdispersive





Black curve shows observations; colors are ensemble

Our approach: Quantile Regression (QR)



Benefits

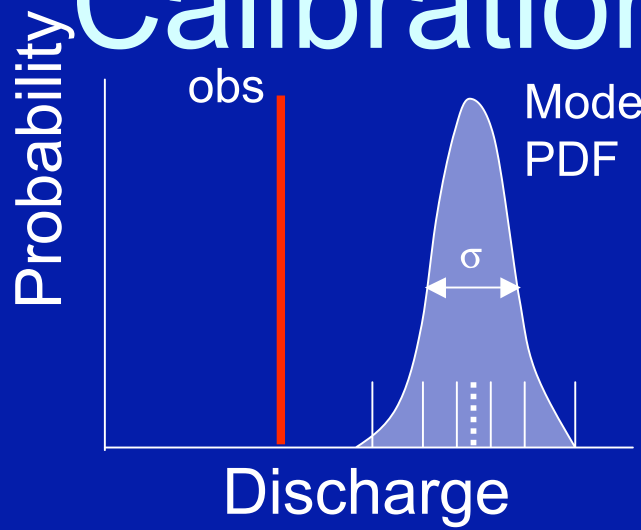
1) Less sensitivity to outliers

2) Works with heteroscedastic errors

3) Optimally fit for each part of a (non-gaussian) PDF

4) "flat" rank histograms

Calibration Procedure



Use QR to perform a fit on 78 quantiles individually (recall: 78 FUSE models simulations).

<u>For each of quantile:</u>

1) Perform a "climatological" fit to the data
   => simulation always as good as "climatology"

2) Starting with full regressor set, iteratively select best subset using "forward step-wise cross-validation"

   – Fitting done using QR
   – Selection done by:
     a) Minimizing QR cost function
     b) Satisfying the binomial distribution
     => Verification measures directly inform the model selection

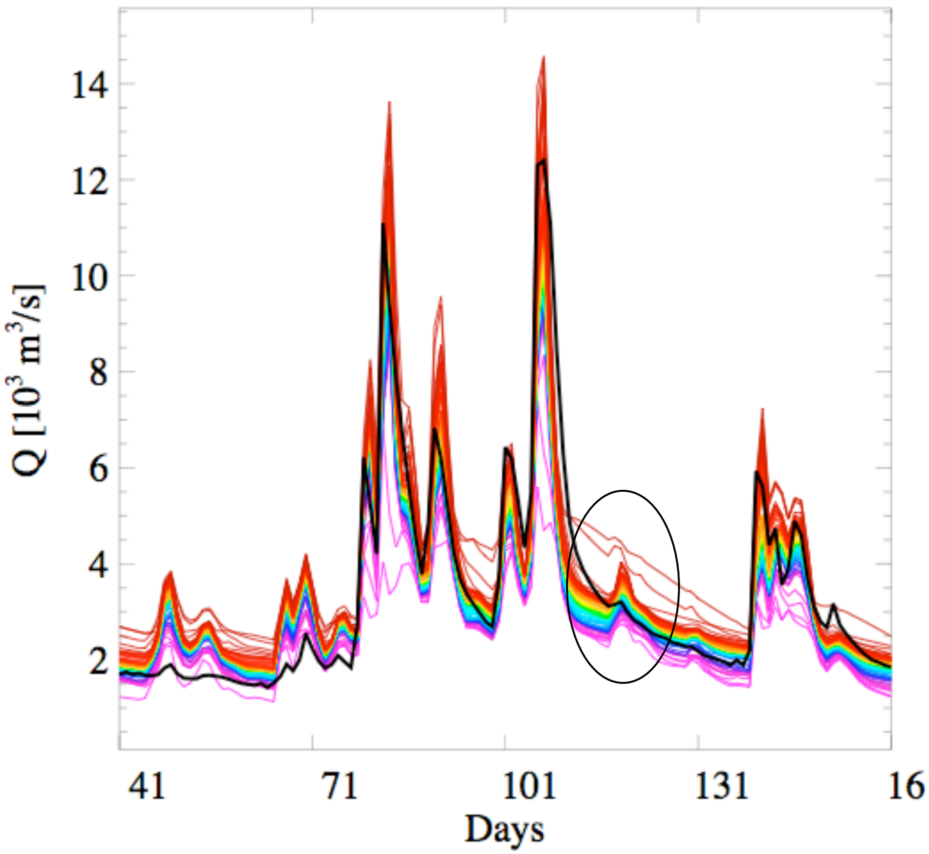
3) 2nd pass: segregate forecasts into differing ranges of ensemble dispersion, and refit models
   => forcing skill-spread utility

<u>Regressor set for each quantile:</u>

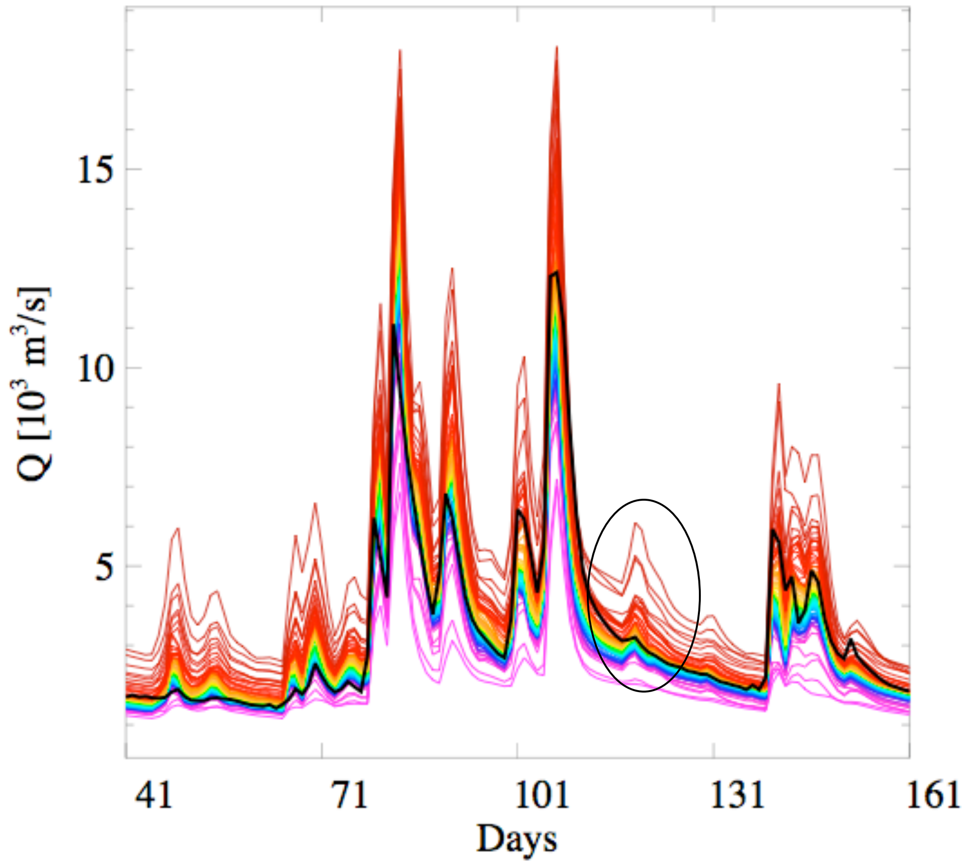
1) - 78) All individual 78 model simulations
79) ensemble mean
80) ensemble standard deviation
81) ranked ensemble member (sorted ensemble that corresponds to quantile being fit)

Example: French Broad River

Before Calibration



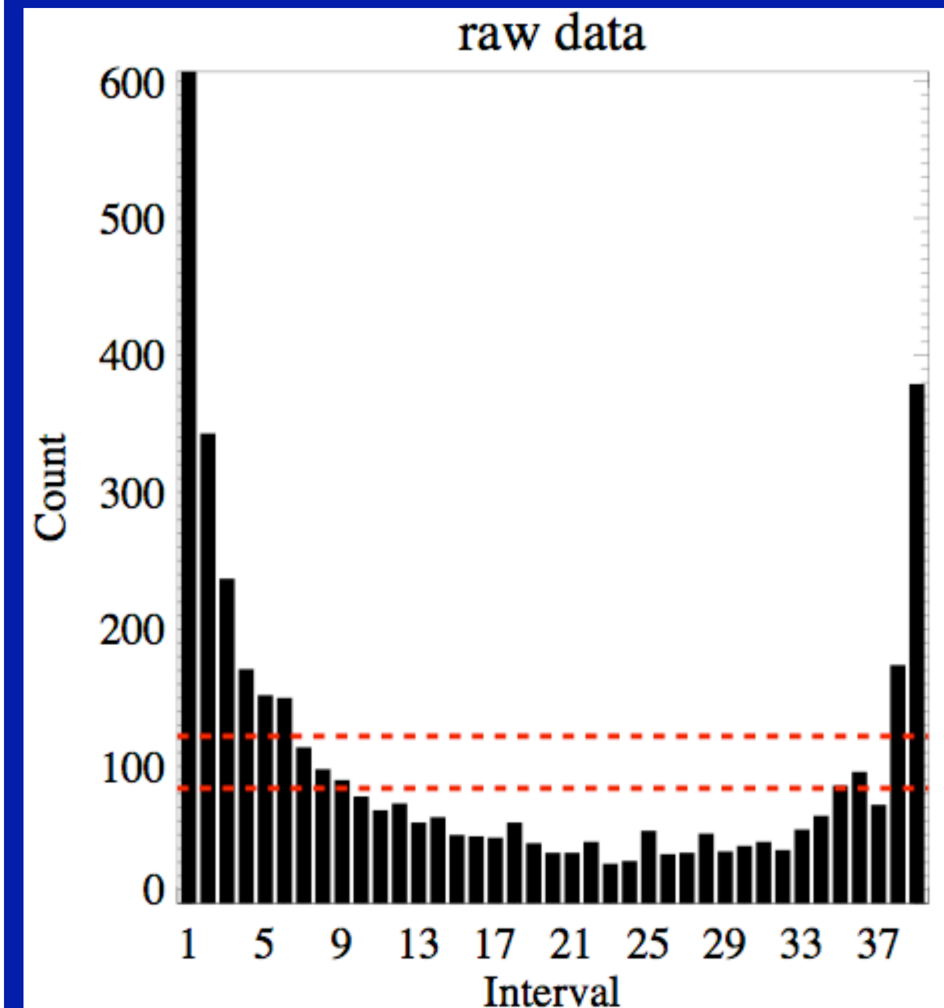
After Calibration



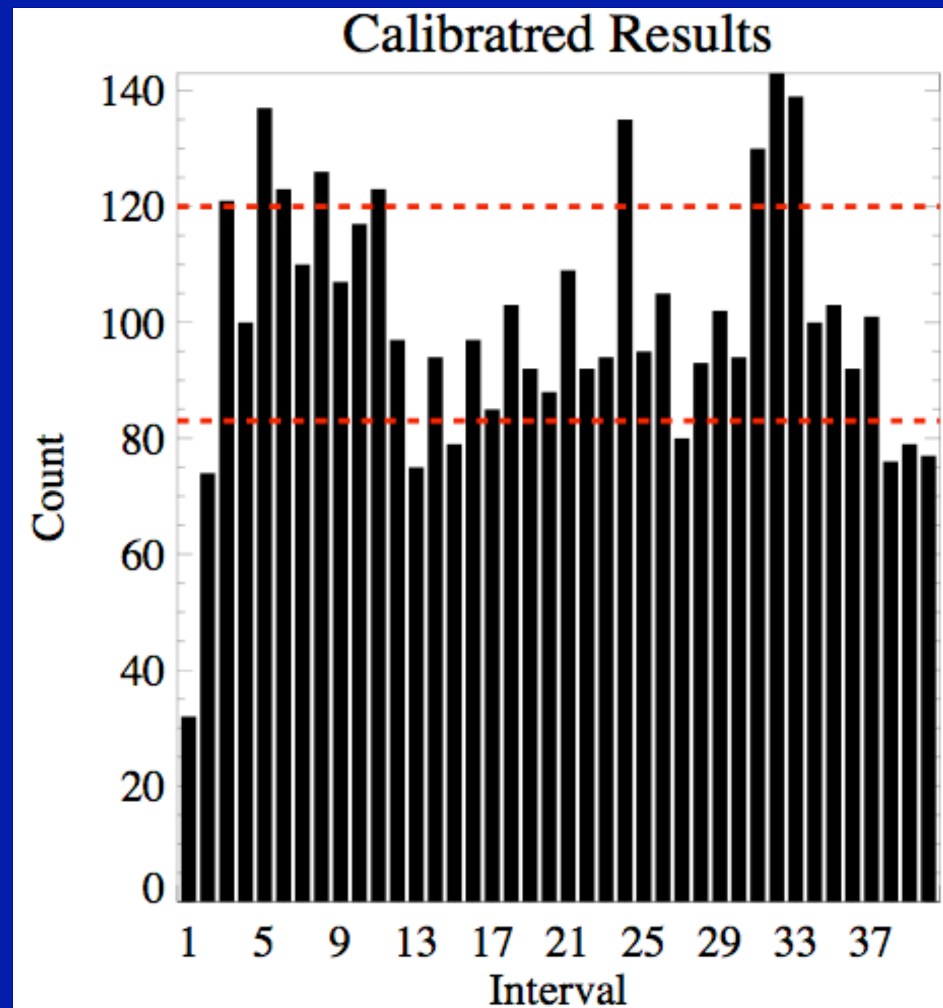
Black curve shows observations; colors are ensemble

Rank Histogram Comparisons

Raw full ensemble



After calibration



After quantile regression, rank histogram more uniform

(although now slightly over-dispersive)

What Nash-Sutcliffe implies about Utility

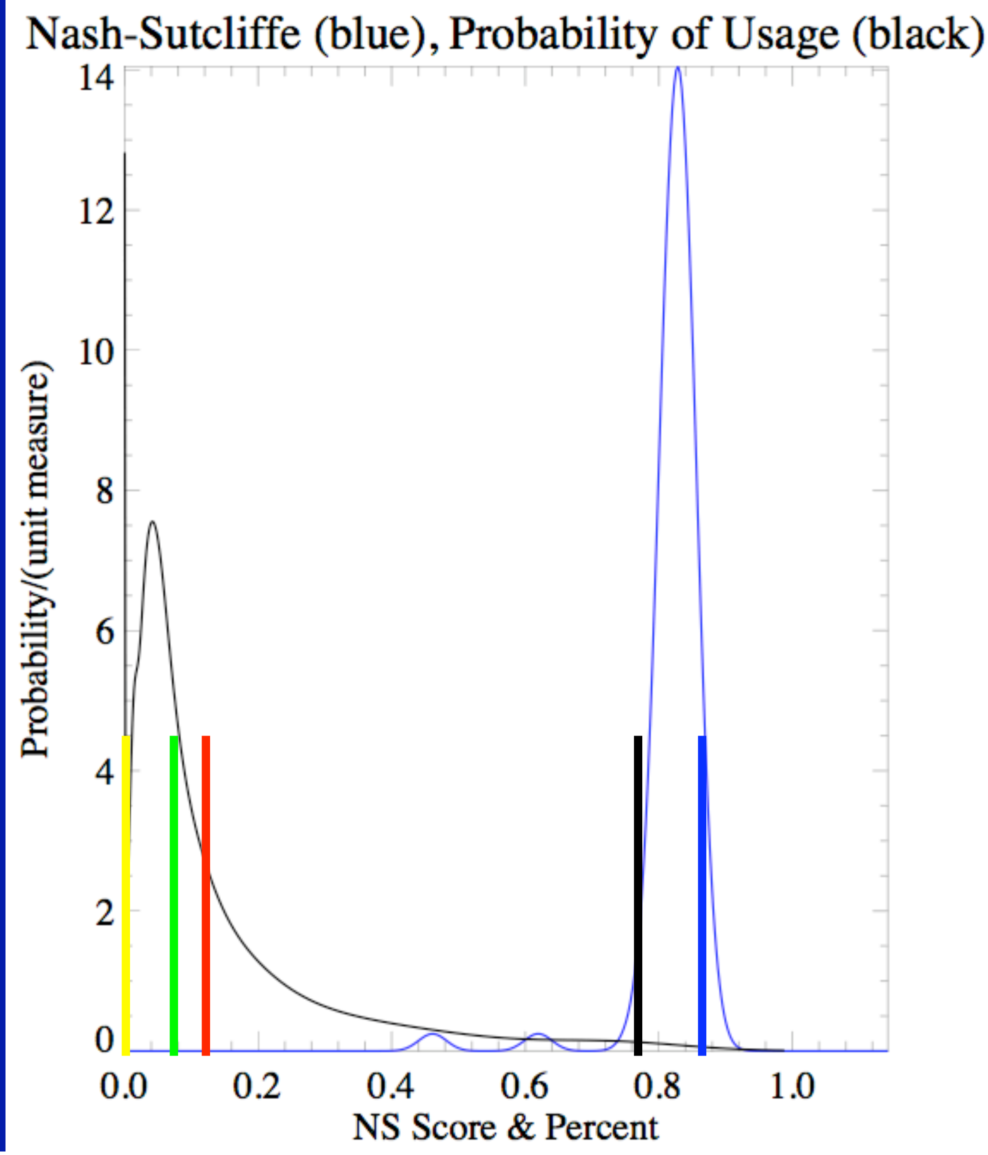
Frequency Used for Quantile Fitting of Method I:

Best Model=76%

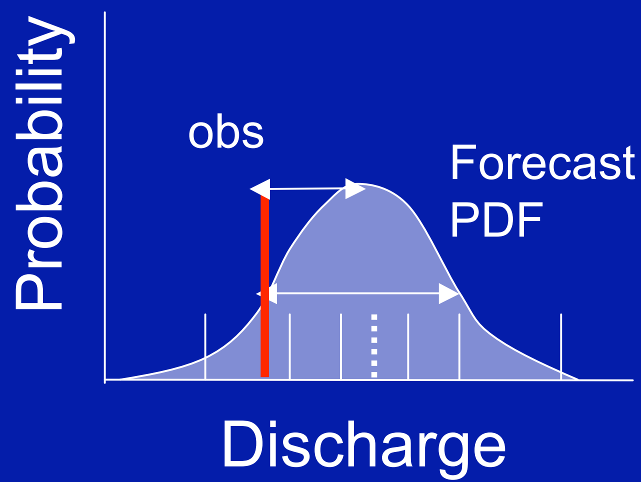
Ensemble StDev=13%

Ensemble Mean=0%

Ranked Ensemble=6%



Note:



$$\langle\ \rangle_i = ensemble\ average$$

$$\left\langle (f_i - o)^2 \right\rangle_i \ versus \left\langle (\bar{f} - o)^2 \right\rangle_i$$

Simplifying

$$eq1: \ \left\langle f_i^2 \right\rangle - 2o\bar{f} + o^2$$

$$eq2: \ \bar{f}^2 - 2o\bar{f} + o^2$$

$$o \sim f_j \quad \Rightarrow \quad \langle\ \rangle_j$$

$$eq1: \ 2\left(\left\langle f^2 \right\rangle - \bar{f}^2\right)$$

$$eq2: \ \left\langle f^2 \right\rangle - \bar{f}^2$$

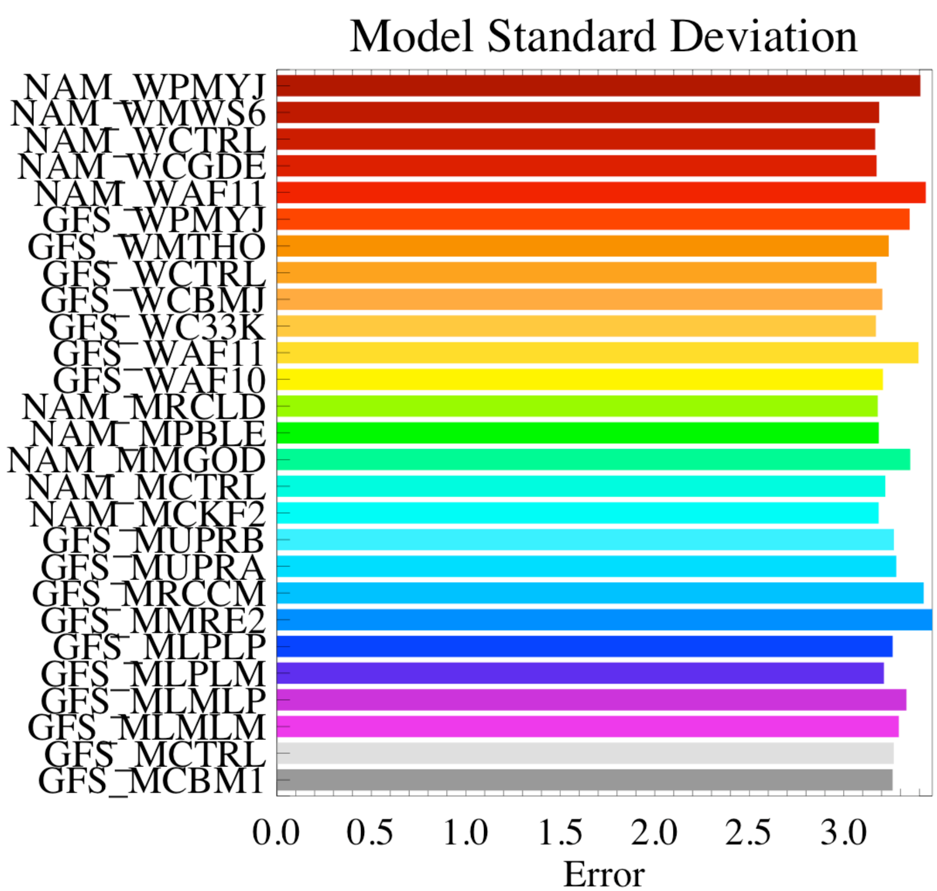
$$\Rightarrow \quad eq1 = 2\ eq2$$

<u>Take home message:</u>

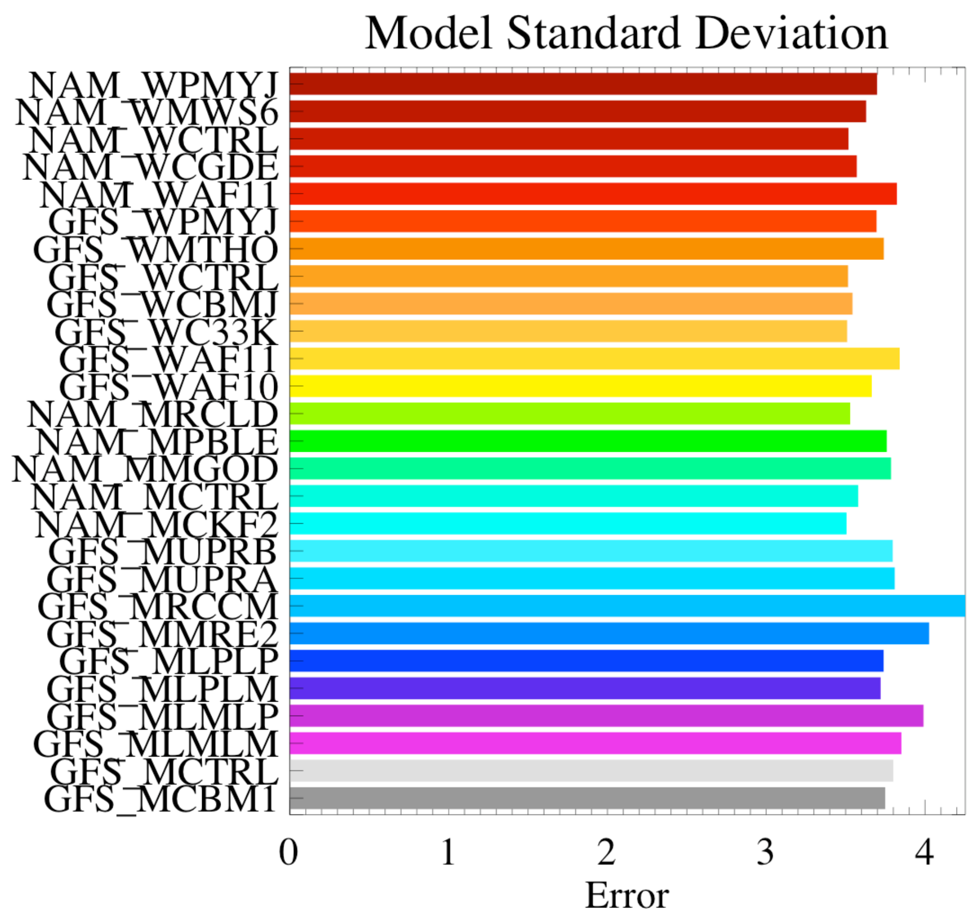
For a "calibrated ensemble", error variance of the ensemble mean is 1/2 the error variance of any ensemble member (on average), independent of the distribution being sampled

RMSE of Models

6hr Lead-time

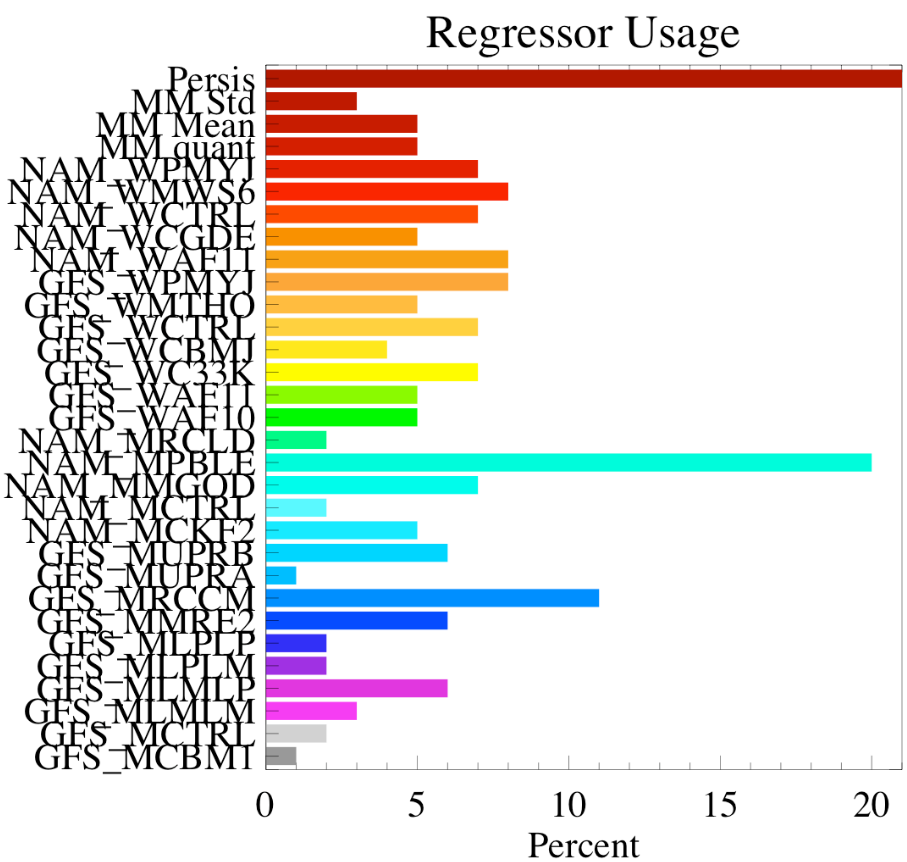


36hr Lead-time

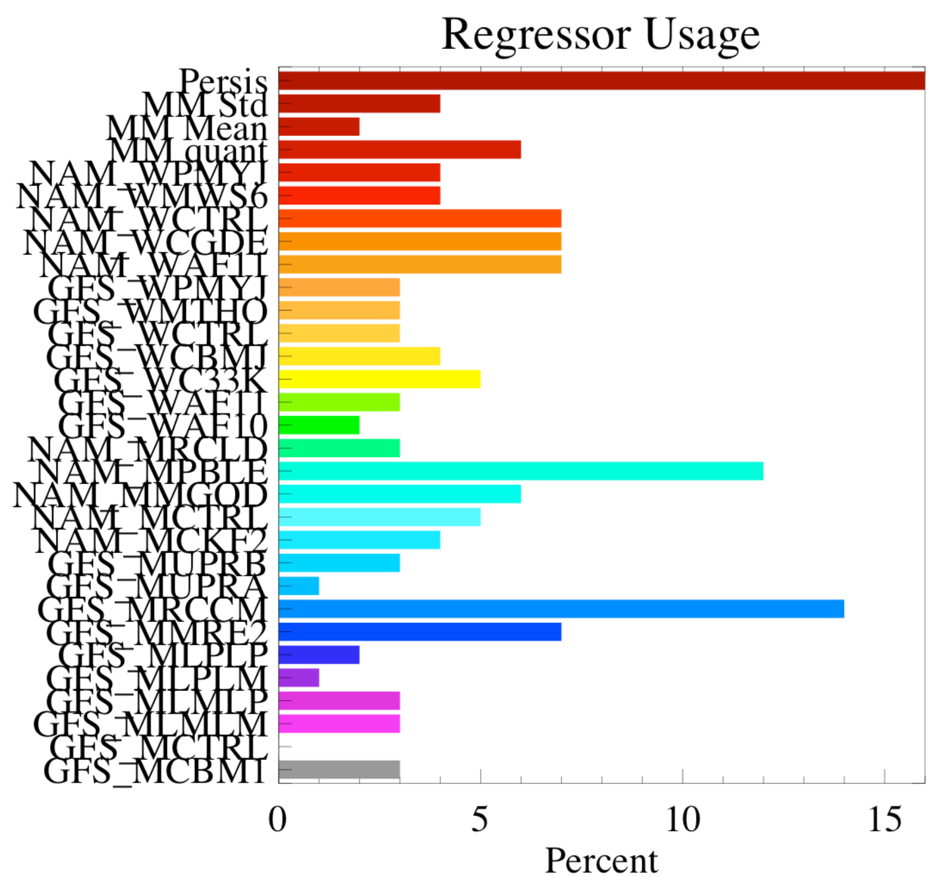


Significant Calibration Regressors

6hr Lead-time



36hr Lead-time



What Nash-Sutcliffe implies about Utility (cont)

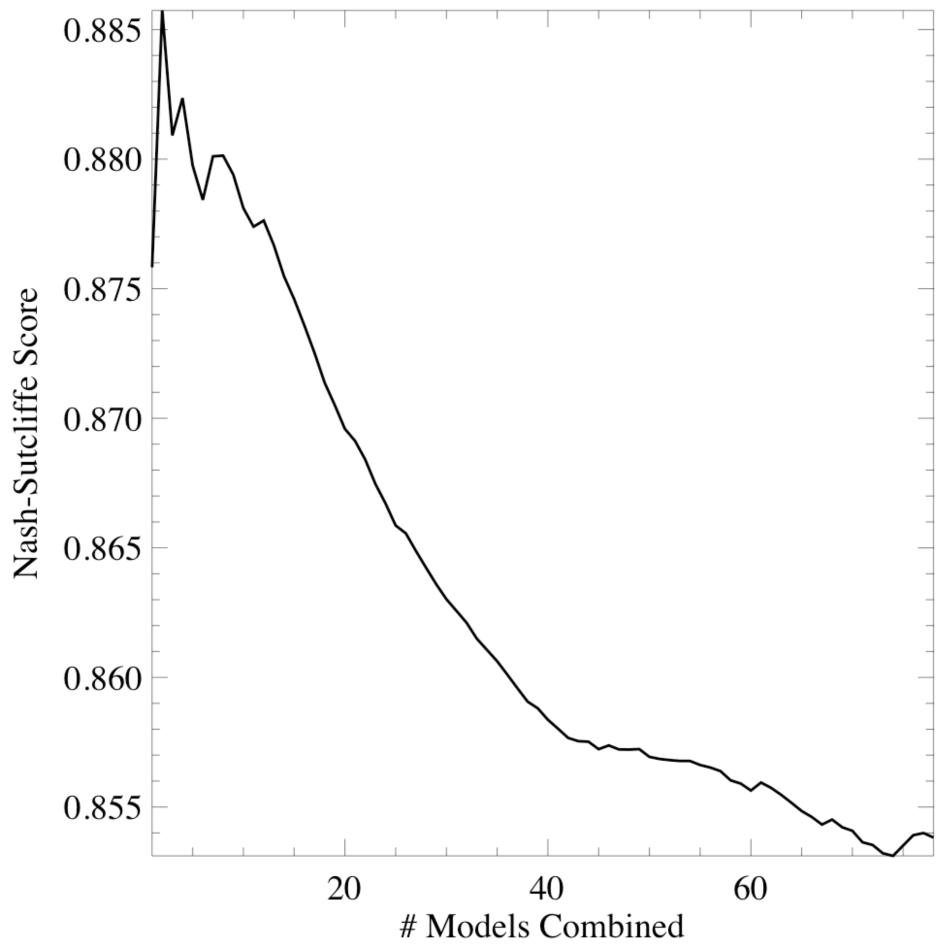
-- degradation with increased ensemble size

Sequentially-averaged FUSE models (ranked based on NS Score) and their resultant NS Score

⇒ Notice the degradation of NS with increasing # (with a peak at 2 models)

⇒ For an equitable multi-model, NS should rise monotonically

⇒ Maybe a smaller subset of models would have more utility? (A contradiction for an under-dispersive ensemble?)



What Nash-Sutcliffe implies about Utility (cont)

… earlier results …

…using only top 1/3 of models
To rank and form ensemble mean …

<u>Initial Frequency Used for Quantile Fitting:</u>

Best Model=76%
Ensemble StDev=13%
Ensemble Mean=0%
Ranked Ensemble=6%

<u>Reduced Set Frequency Used for Quantile Fitting:</u>

Best Model=73%
Ensemble StDev=3%
Ensemble Mean=32%
Ranked Ensemble=29%

⇒Appears to be significant gains in the utility of the ensemble after "filtering" (except for drop in StDev) … however "proof is in the pudding" …
⇒Examine verification skill measures …

Skill Scores

$$SS = \frac{A_{forc} - A_{ref}}{A_{perf} - A_{ref}}$$

- Single value to summarize performance.
- Reference forecast - best naive guess; persistence, climatology
- A perfect forecast implies that the object can be perfectly observed
- Positively oriented – Positive is good

Skill Score Comparisons between full- and “filtered” FUSE ensemble sets

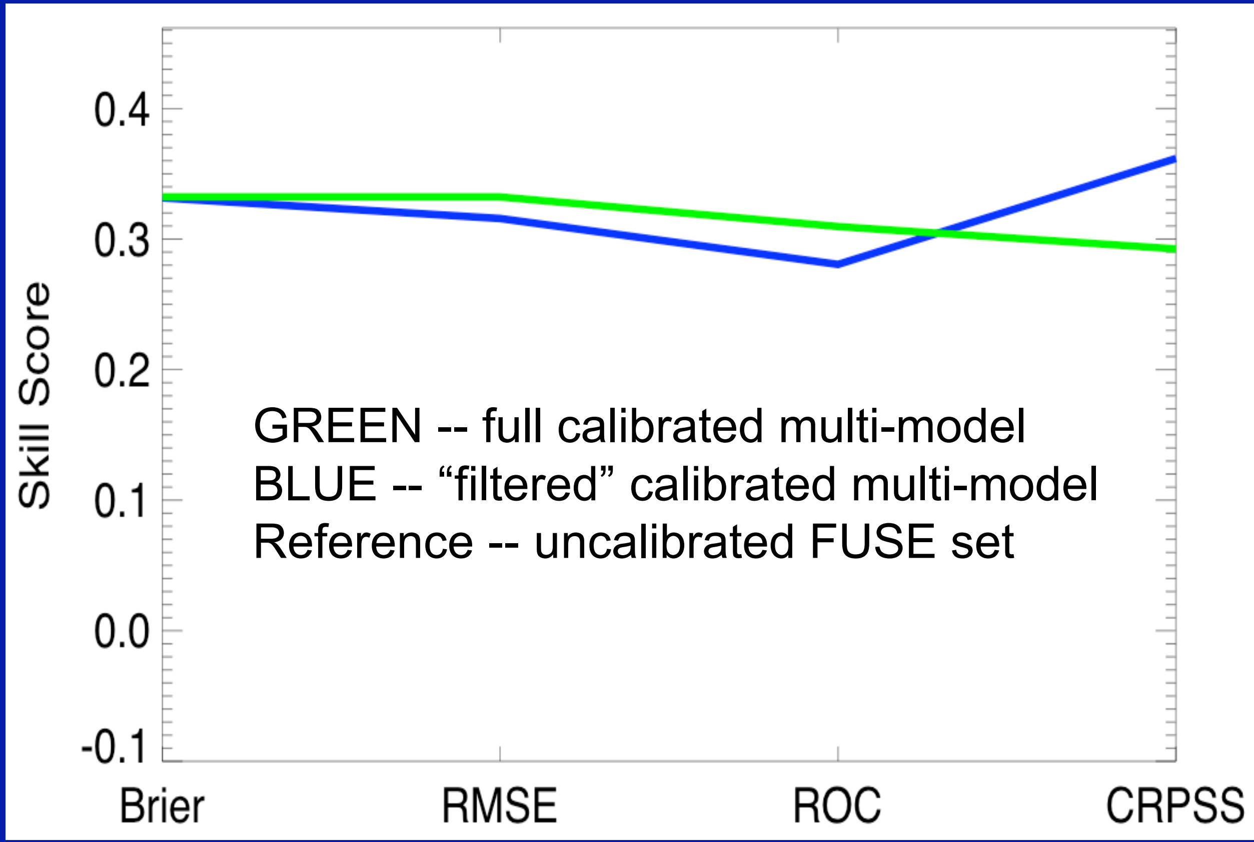
Points:

-- quite similar results for a variety of skill scores

-- both approaches give appreciable benefit over the original raw multi-model output

-- however, only in the CRPSS is there improvement of the “filtered” ensemble set over the full set

⇒post-processing method fairly robust
⇒More work (more filtering?)!



Question revisited:

How best to utilize a multi-model simulations (forecast), especially if under-dispersive?

a) Should more dynamical variability be searched for? Or
b) Is it better to balance post-processing with multi-model utilization to create a properly dispersive, informative ensemble?

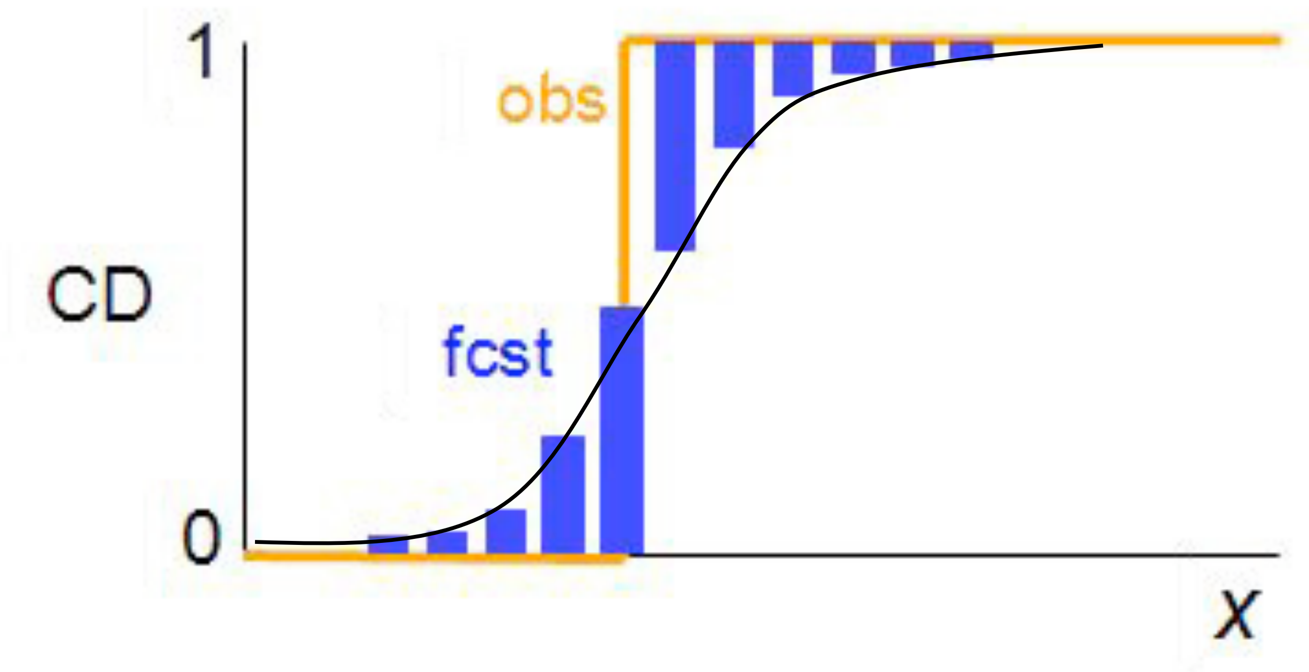
"Answer": adding more models can lead to decreasing skill of the ensemble mean (even if the ensemble is under-dispersive)

Further, quantile-regression-based calibration is fairly robust and can do a lot with just a single model (not shown), especially if a variety of approaches are utililized.


Thank You!
hopson@ucar.edu

Rank Probability Score

for multi-categorical or continuous variables



$$RPS = \frac{1}{n-1}\sum_{i=1}^{n}\left(CDF_{fc,i} - CDF_{obs,i}\right)^2$$

Continuous scores: MSE

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2$$

Attribute: measures accuracy

Average of the squares of the errors: it measures the magnitude of the error, weighted on the squares of the errors

it does not indicate the direction of the error

Quadratic rule, therefore large weight on large errors:

→ good if you wish to penalize large error

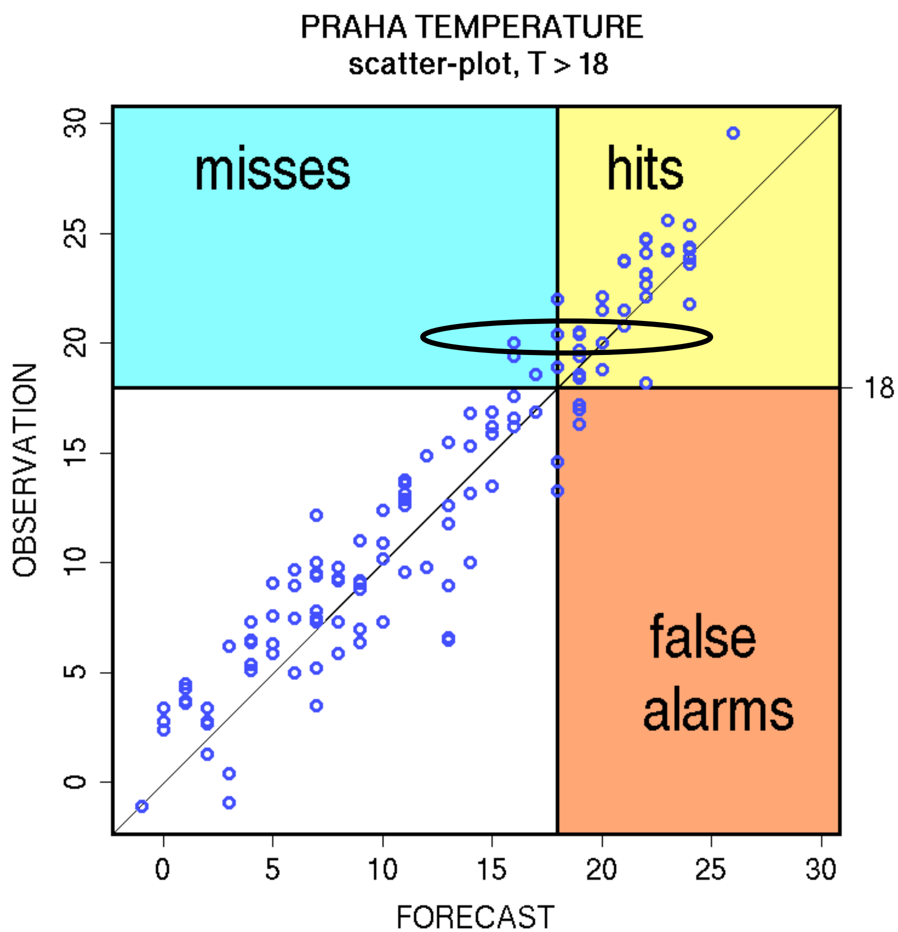
→ sensitive to large values (e.g. precipitation) and outliers; sensitive to large variance (high resolution models); encourage conservative forecasts (e.g. climatology)

=> For ensemble forecast, use ensemble mean

Slide from Barbara Casati

Scatter-plot and Contingency Table

Does the forecast detect correctly temperatures above 18 degrees ?



Slide from Barbara Casati

Brier Score

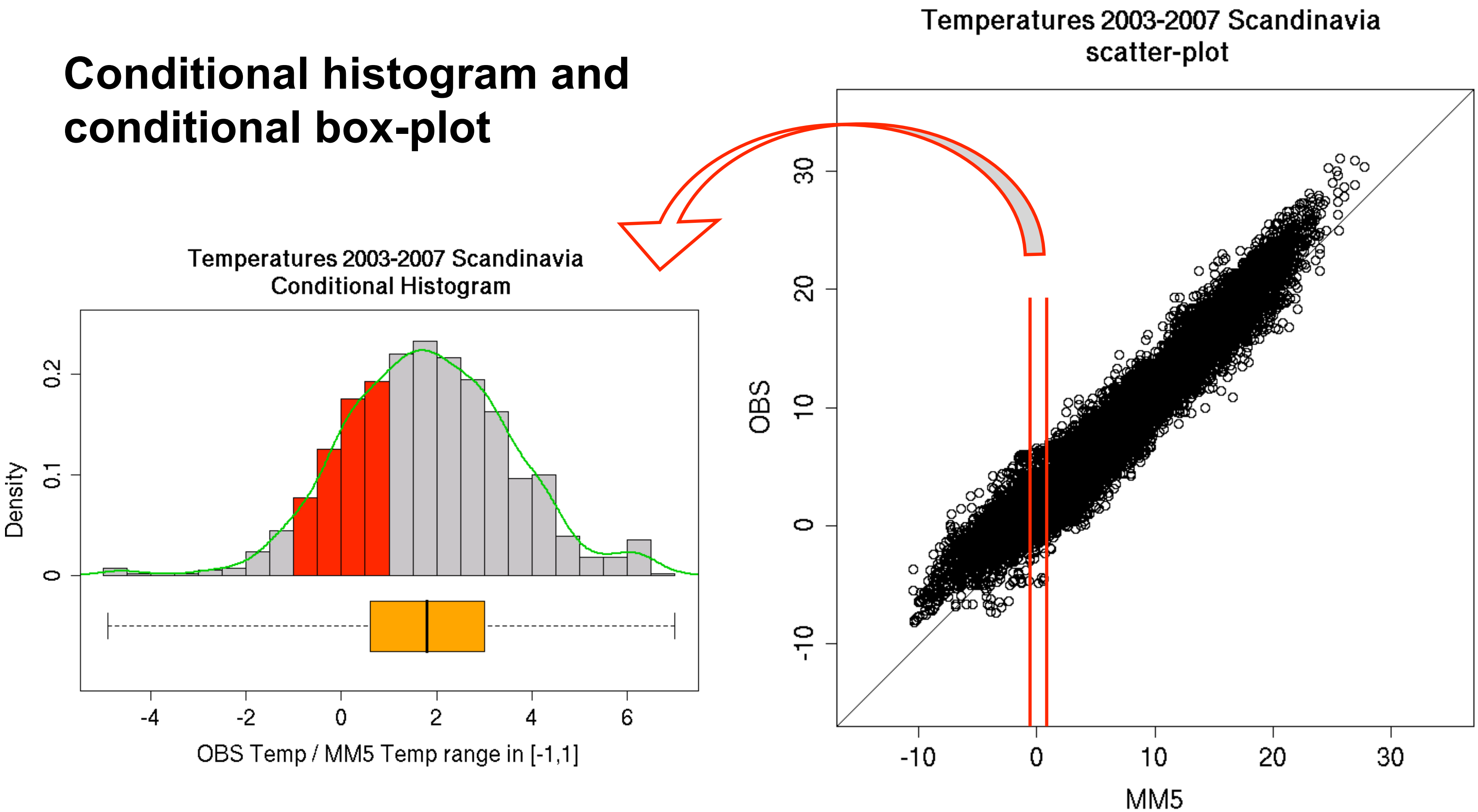
$$BS = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - o_i\right)^2$$

y = forecasted event occurence
o = observed occurrence (0 or 1)
i = sample # of total n samples

=> Note similarity to MSE

Conditional Distributions

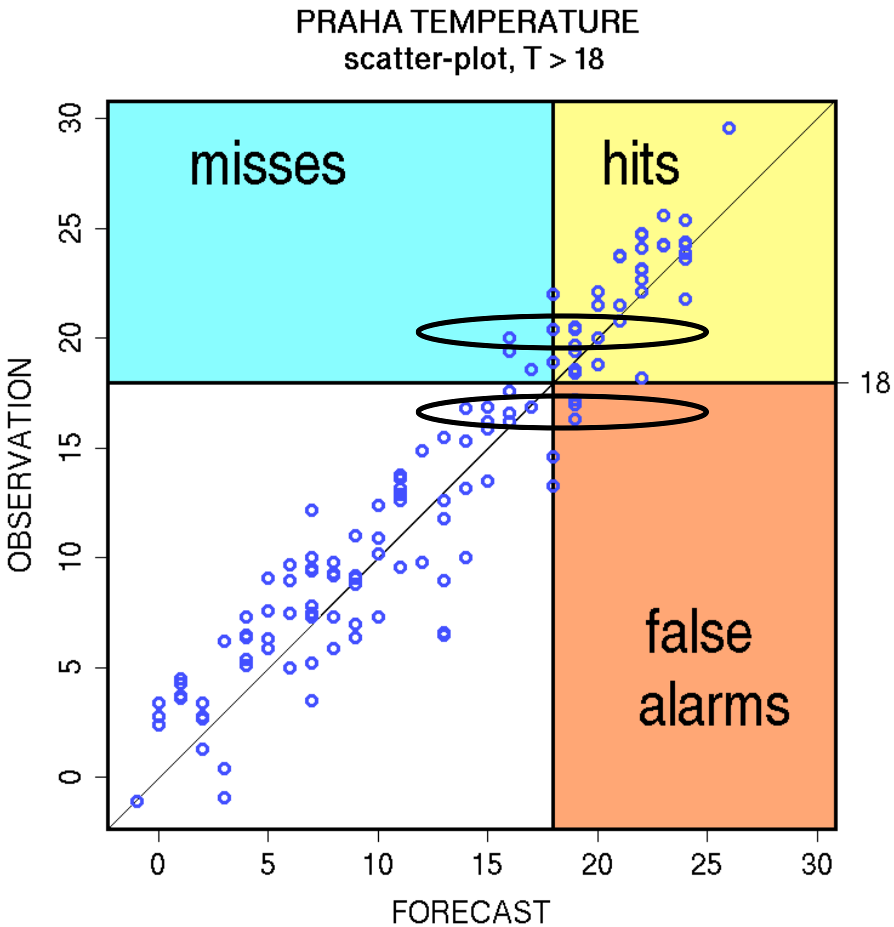
Conditional histogram and conditional box-plot



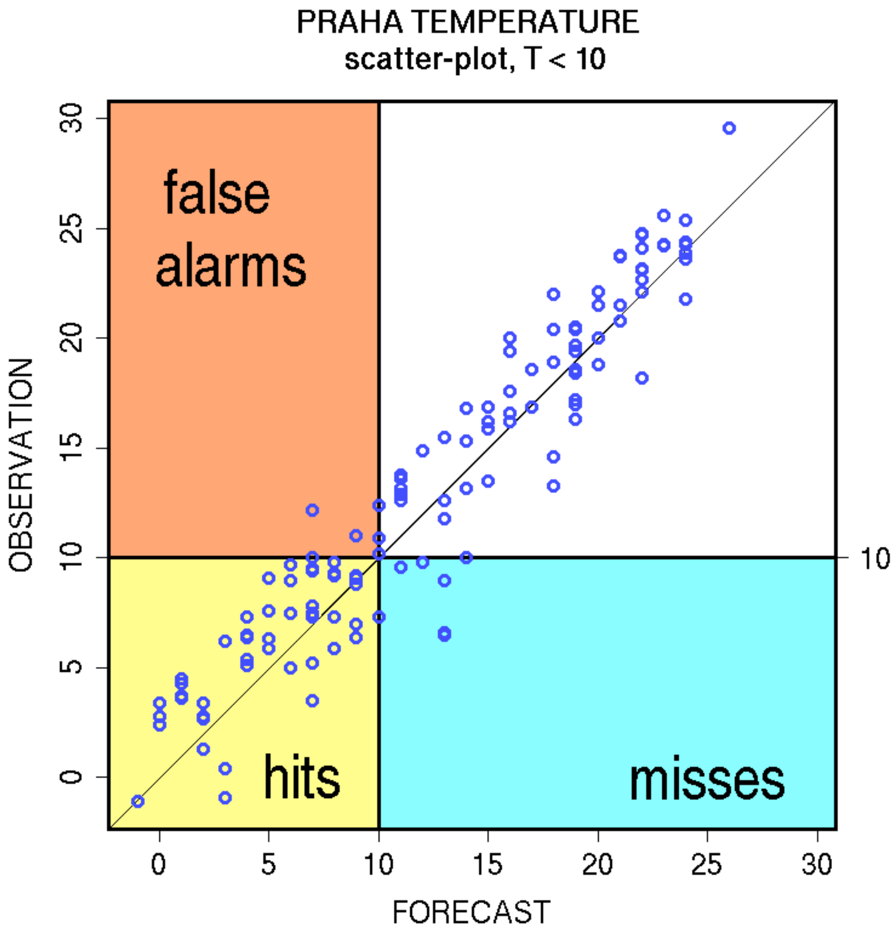
Slide from Barbara Casati

Scatter-plot and Contingency Table

Does the forecast detect correctly temperatures above 18 degrees ?

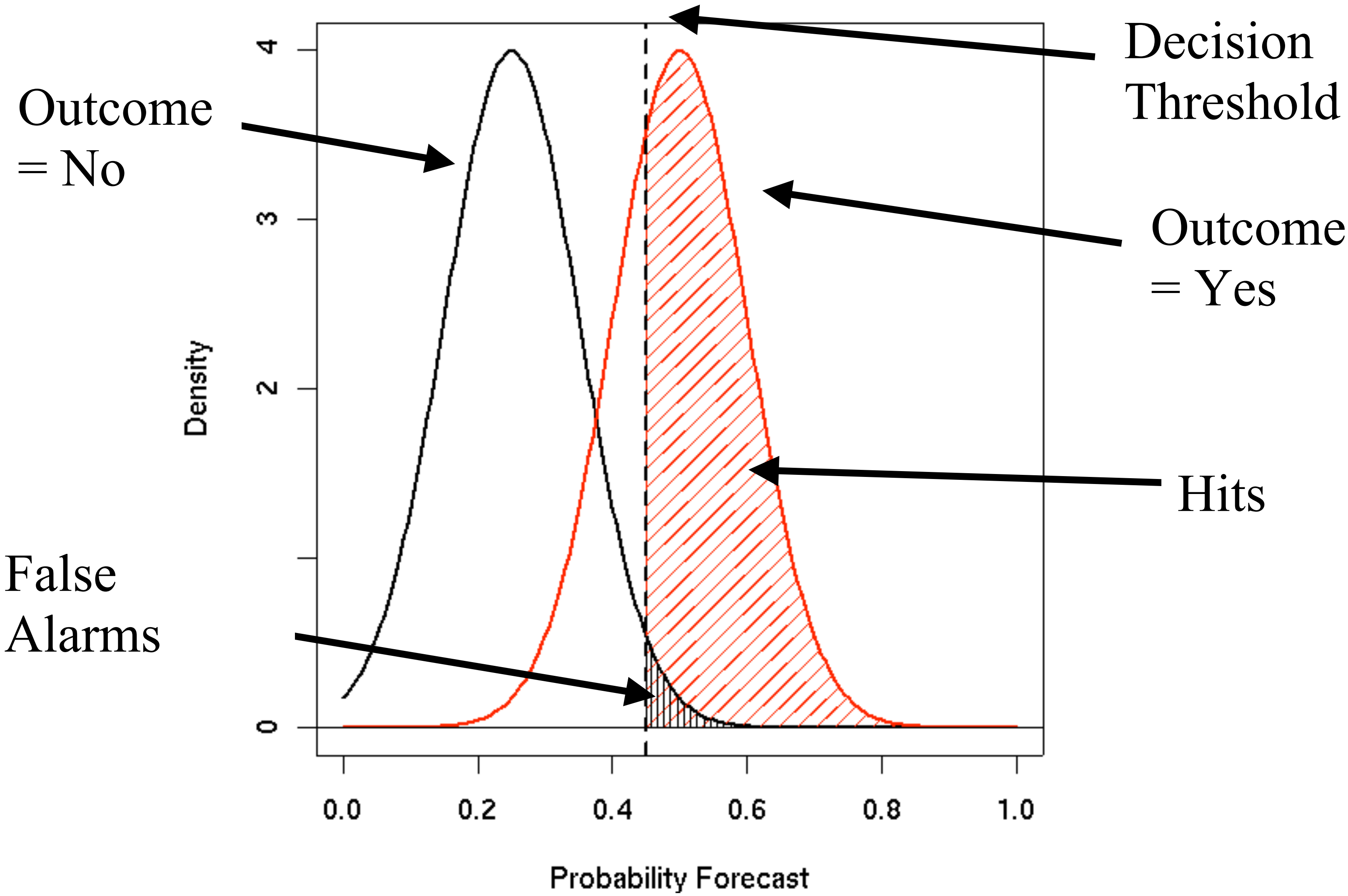


Does the forecast detect correctly temperatures below 10 degrees ?



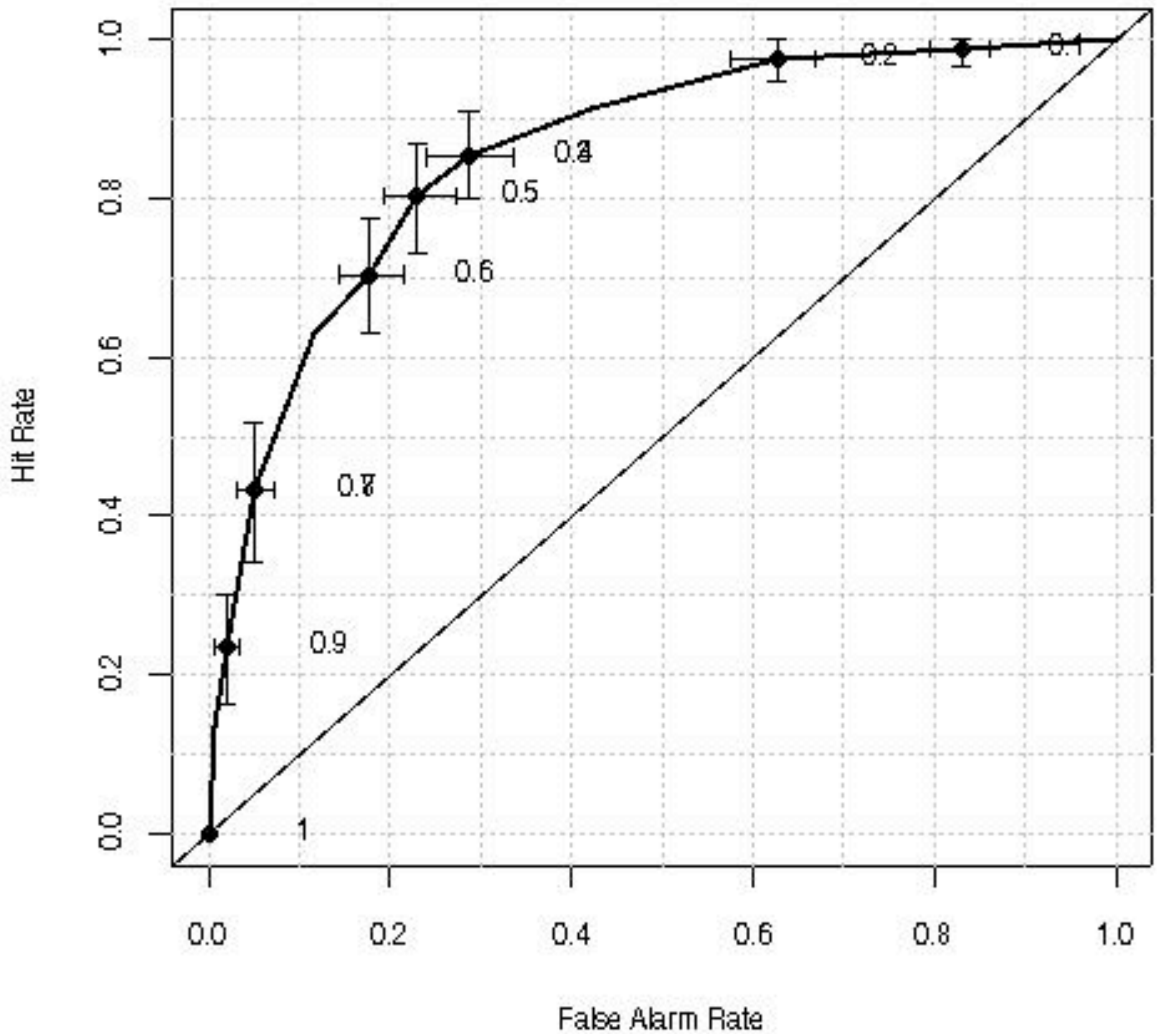
Slide from Barbara Casati

Discrimination Plot



Slide from Matt Pocernic

Receiver Operating Characteristic (ROC) Curve



Slide from Matt Pocernic