



Aggregated Verification Results for flows at 5 ABRFC test basins

**Comparison of Hydrologic and Input
Uncertainties**

&

**Comparison of streamflow from GFS
Subsystem and EPPII outputs**

Julie Demargne

Hydrologic Ensemble Prediction (HEP) Group
Office of Hydrologic Development
NOAA/National Weather Service

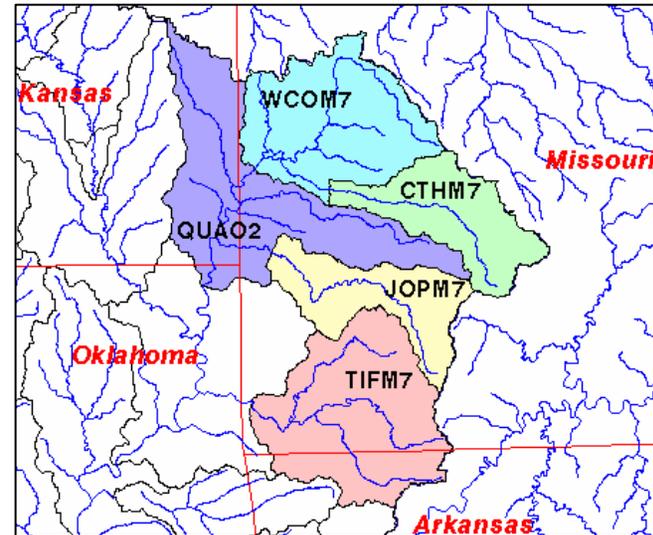
RFC Short-Term Ensemble Workshop, November 30, 2006

Computation

- Verification statistics computed for each individual basin at 3 different scales: annual, seasonal, and monthly
- Aggregate verification statistics computed as the weighted average of individual basin statistics
- Forecast flows generated for lead days 1 to 14 using
 - precipitation hindcasts generated by
 - the Ensemble Pre-Processor EPPII from RFC-QPF (days 1 & 2) and resampled climatology
 - the GFS subsystem from RFC-QPF (days 1 & 2) and GFS single-value forecasts (days 3 to 14)
 - temperature resampled climatology ensembles from EPPII
 - => Expected differences: lead days 1 & 2 from EPPII enhancements (intermittency and optimized parameters), lead days 3 to 14 from GFS single-value forecasts
- Forecast flows compared to 2 references:
 - observed flow (show all errors – hydrologic and input uncertainties)
 - simulated flow (show only errors from input – input uncertainty)
- Verification period (dependent validation):
03/06/2003 – 12/31/2004

Verification Results

- Individual test basins at ABRFC :



- Verification statistics:
 - Briers statistics (slides 4 to 12): annual scale
 - Reliability diagrams (slides 13 to 21): annual and seasonal scales
 - ROC plots (slides 22 to 30): annual and seasonal scales

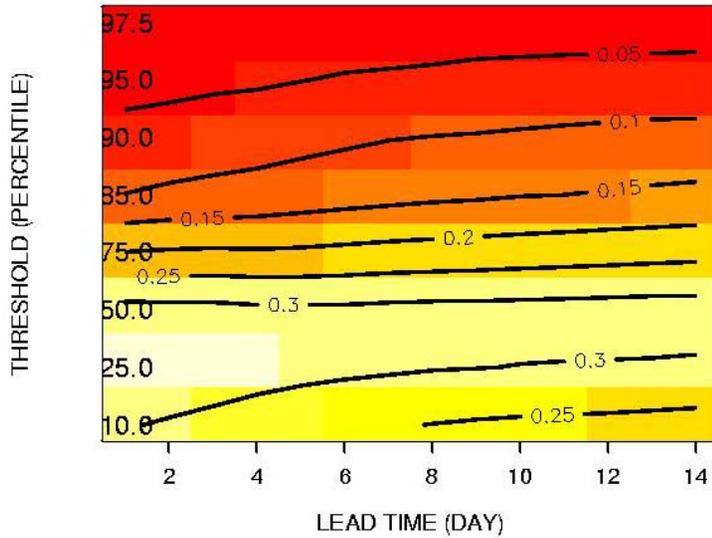
Briers statistics

- Annual results with observed and simulated flows
- Briers Skill Score (BSS) computed with 2 references: climatology and persistence
- Brier Score (BS) measures mean squared probability error
- Brier Score (BS) decomposition:
$$BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty}$$
- Brier Skill Score (BSS) measures improvement over reference forecast

Briers Score Statistics

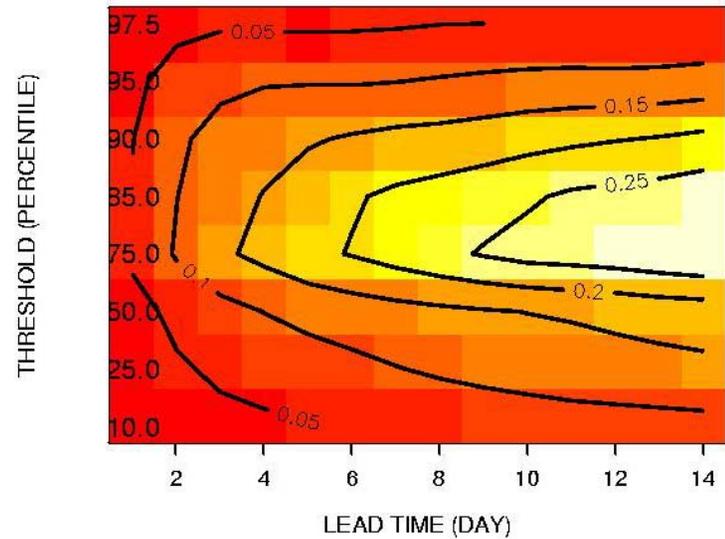
Reference flow: observed

ABRFC – BS, ENSEMBLE FCST OF 24HR FLOW

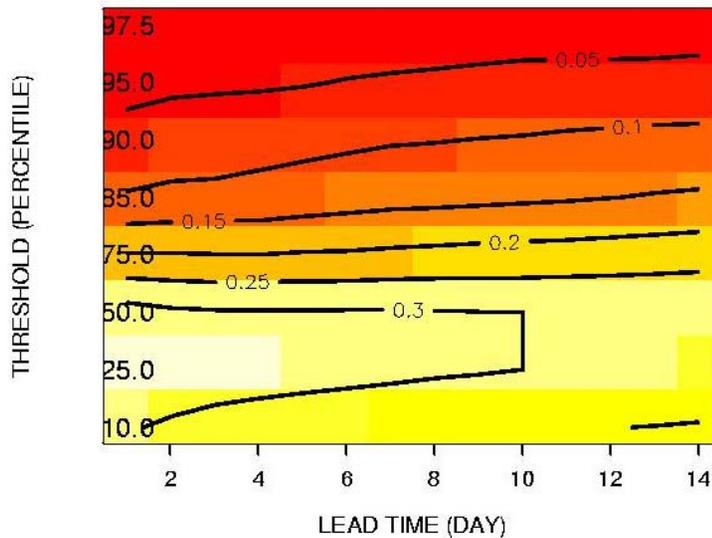


EPPII

ABRFC – BS, PERSISTENCE FCST OF 24HR FLOW



ABRFC – BS, ENSEMBLE FCST OF 24HR FLOW



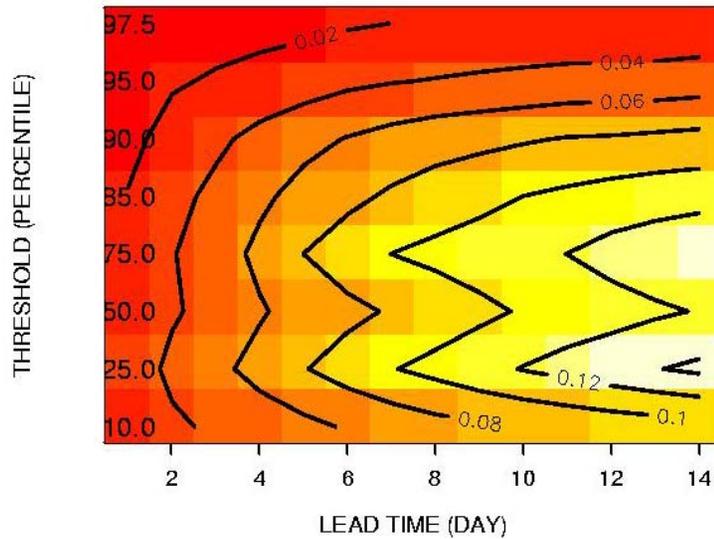
GFS

Perfect score: $BS=0$

Briers Score Statistics

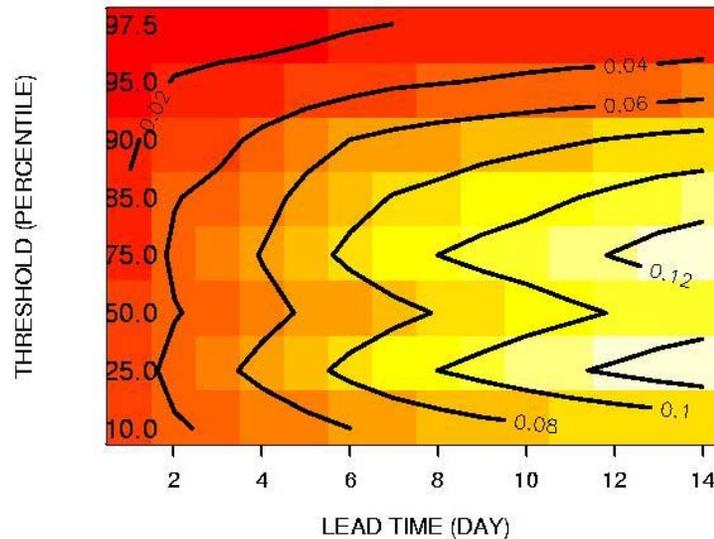
Reference flow: simulated

ABRFC – BS, ENSEMBLE FCST OF 24HR FLOW



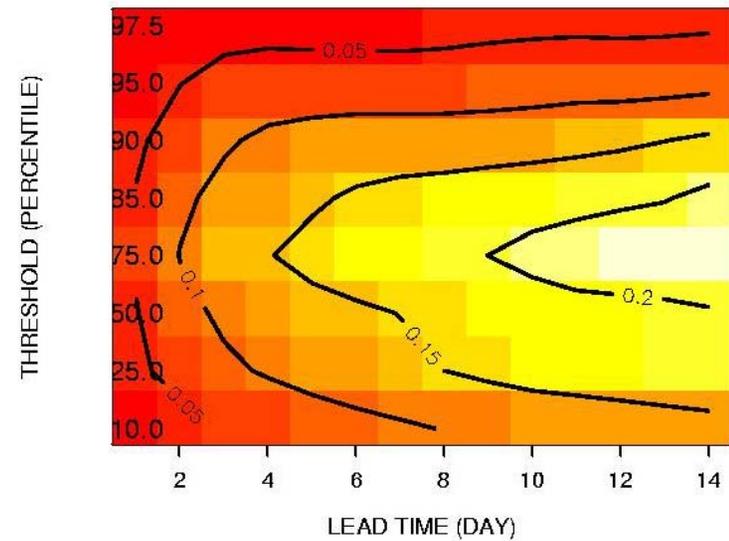
EPPII

ABRFC – BS, ENSEMBLE FCST OF 24HR FLOW



GFS

ABRFC – BS, PERSISTENCE FCST OF 24HR FLOW

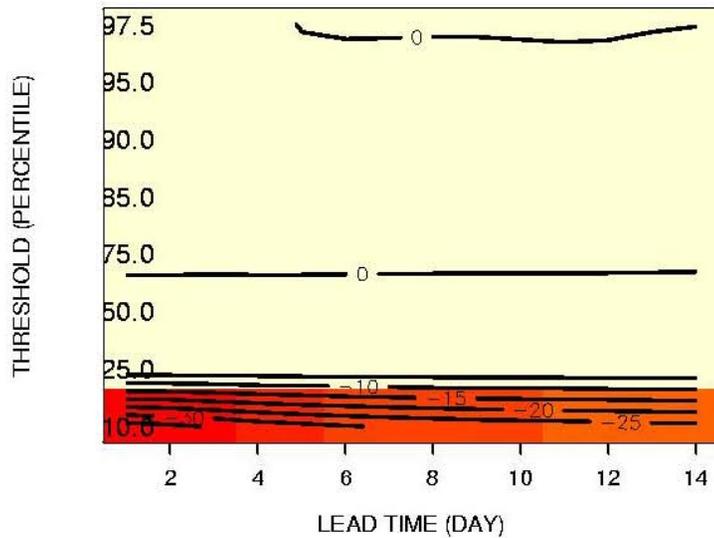


Perfect score: $BS=0$

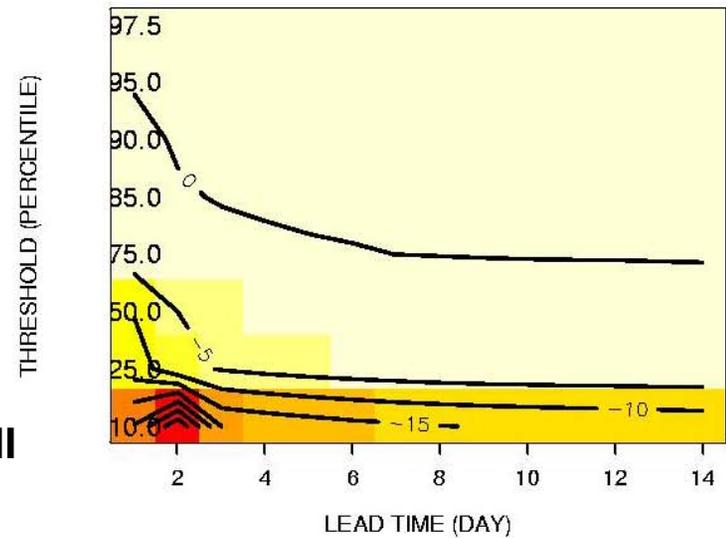
Briers Skill Score Statistics

Reference flow: observed

ABRFC – BSS, ENSEMBLE FCST VS. CLIMATOLOGY

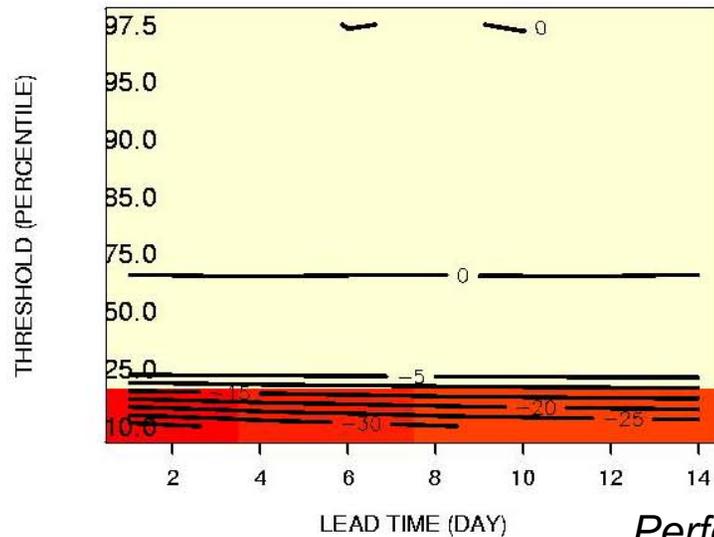


ABRFC – BSS, ENSEMBLE FCST VS. PERSISTENCE

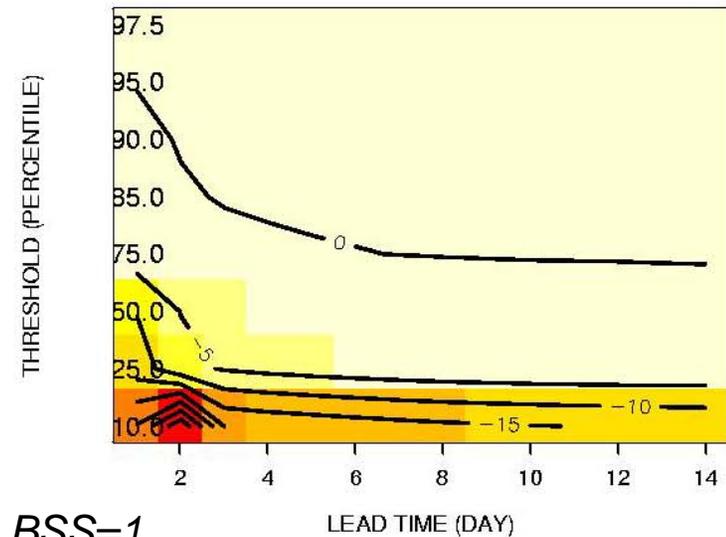


EPPII

ABRFC – BSS, ENSEMBLE FCST VS. CLIMATOLOGY



ABRFC – BSS, ENSEMBLE FCST VS. PERSISTENCE



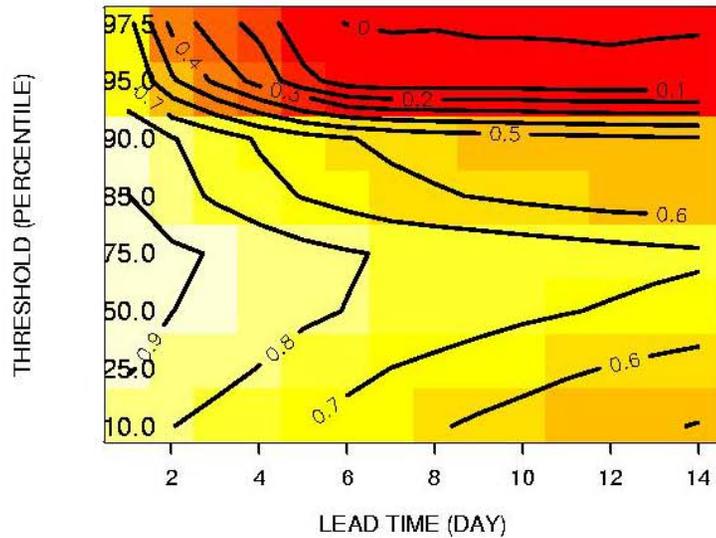
GFS

Perfect score: BSS=1

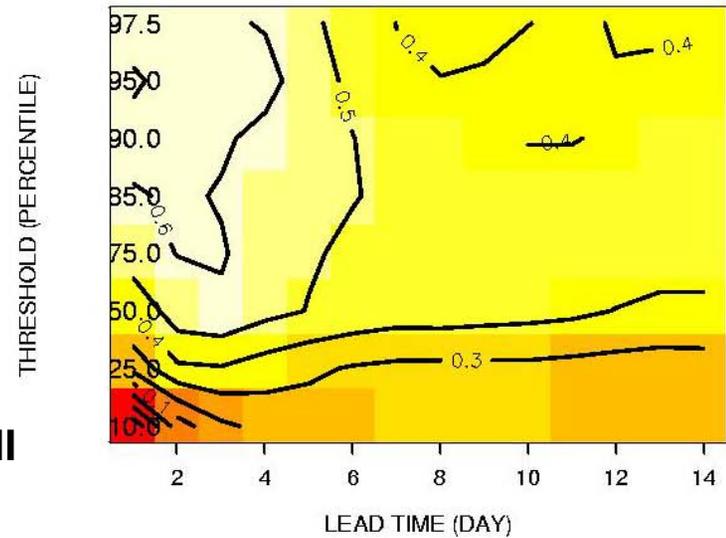
Briers Skill Score Statistics

Reference flow: simulated

ABRFC – BSS, ENSEMBLE FCST VS. CLIMATOLOGY

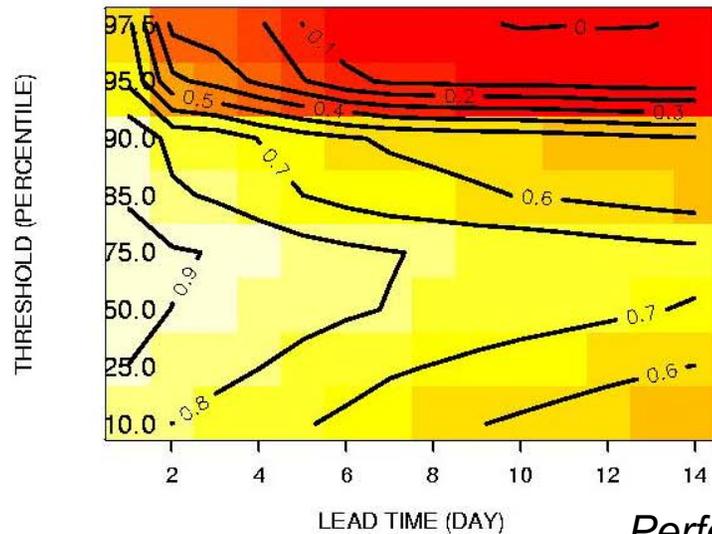


ABRFC – BSS, ENSEMBLE FCST VS. PERSISTENCE

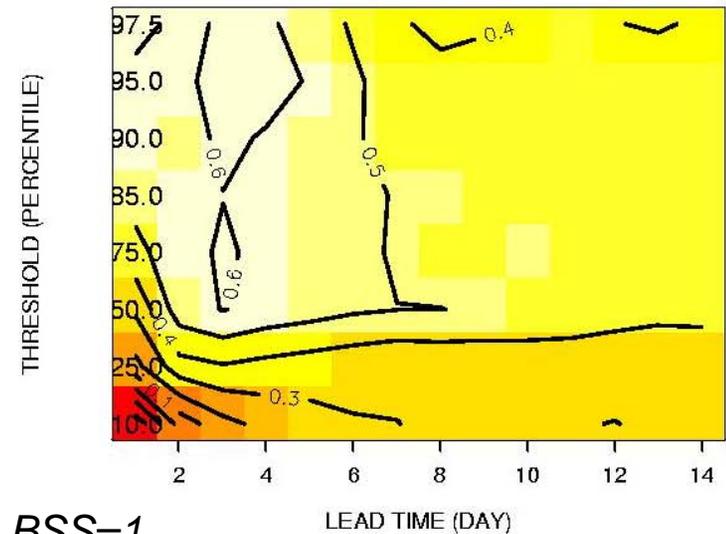


EPP11

ABRFC – BSS, ENSEMBLE FCST VS. CLIMATOLOGY



ABRFC – BSS, ENSEMBLE FCST VS. PERSISTENCE



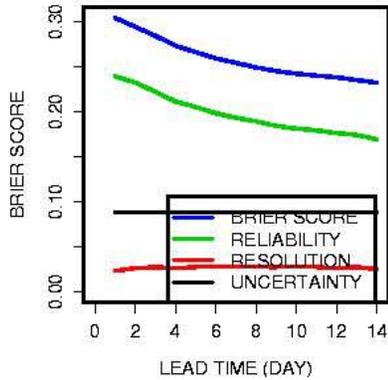
GFS

Perfect score: BSS=1

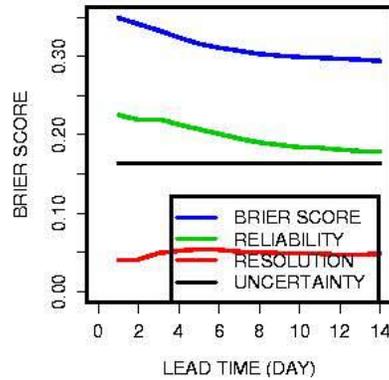
Briers Score Statistics

Reference flow: observed

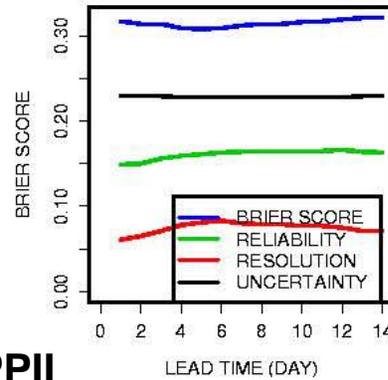
ABRFC – 24HR FLOW, 10.0TH PER.



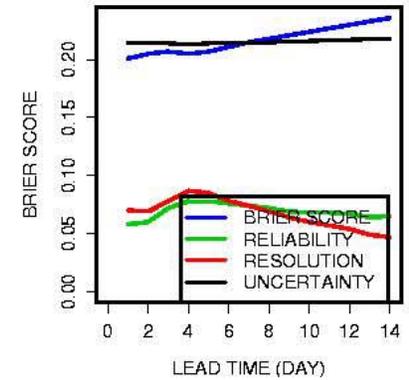
ABRFC – 24HR FLOW, 25.0TH PER.



ABRFC – 24HR FLOW, 50.0TH PER.

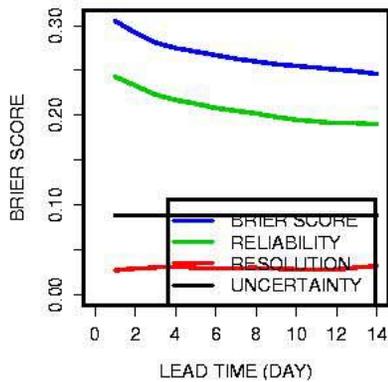


ABRFC – 24HR FLOW, 75.0TH PER.

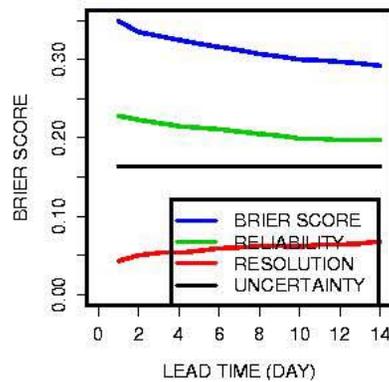


EPP11

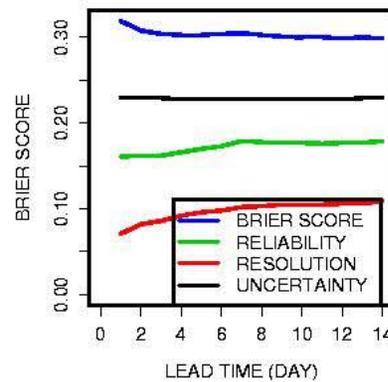
ABRFC – 24HR FLOW, 10.0TH PER.



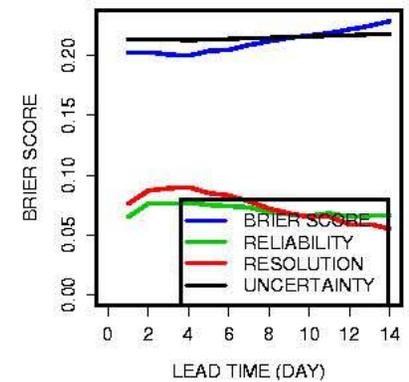
ABRFC – 24HR FLOW, 25.0TH PER.



ABRFC – 24HR FLOW, 50.0TH PER.



ABRFC – 24HR FLOW, 75.0TH PER.



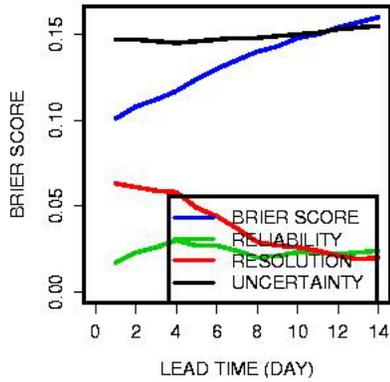
GFS

$$BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty}$$

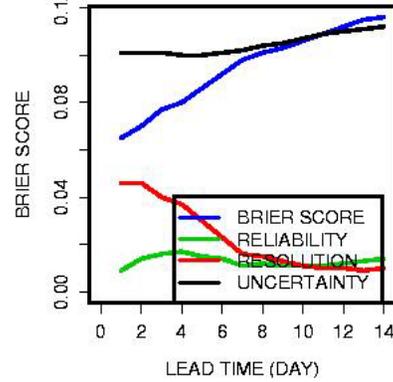
Briers Score Statistics

Reference flow: observed

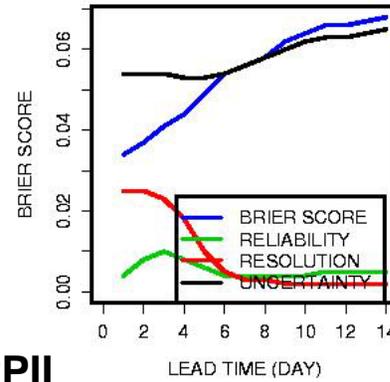
ABRFC – 24HR FLOW, 85.0TH PER.



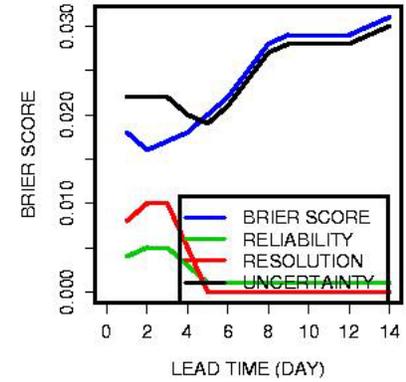
ABRFC – 24HR FLOW, 90.0TH PER.



ABRFC – 24HR FLOW, 95.0TH PER.

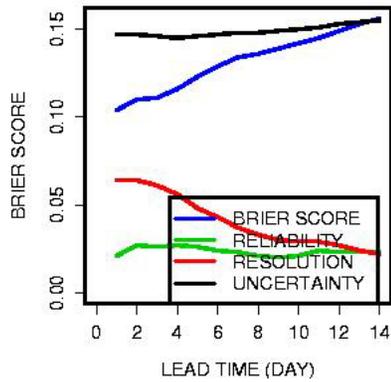


ABRFC – 24HR FLOW, 97.5TH PER.

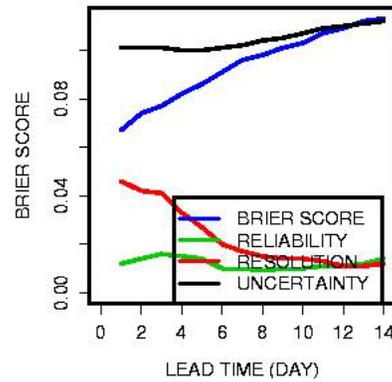


EPP11

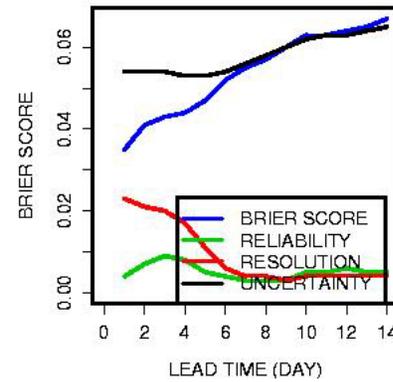
ABRFC – 24HR FLOW, 85.0TH PER.



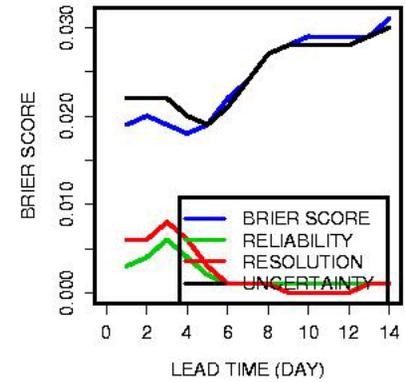
ABRFC – 24HR FLOW, 90.0TH PER.



ABRFC – 24HR FLOW, 95.0TH PER.



ABRFC – 24HR FLOW, 97.5TH PER.

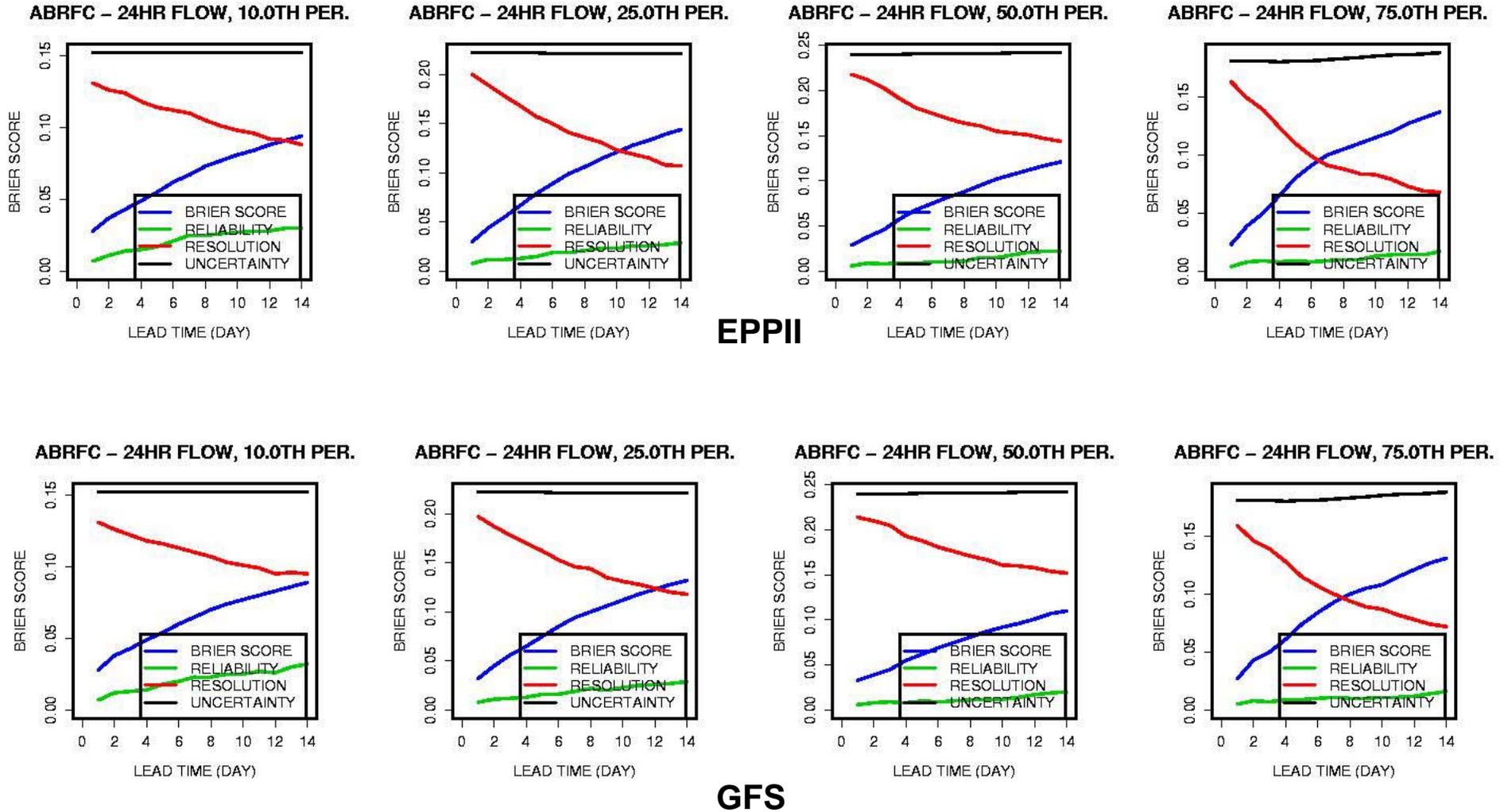


GFS

$$BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty}$$

Briers Score Statistics

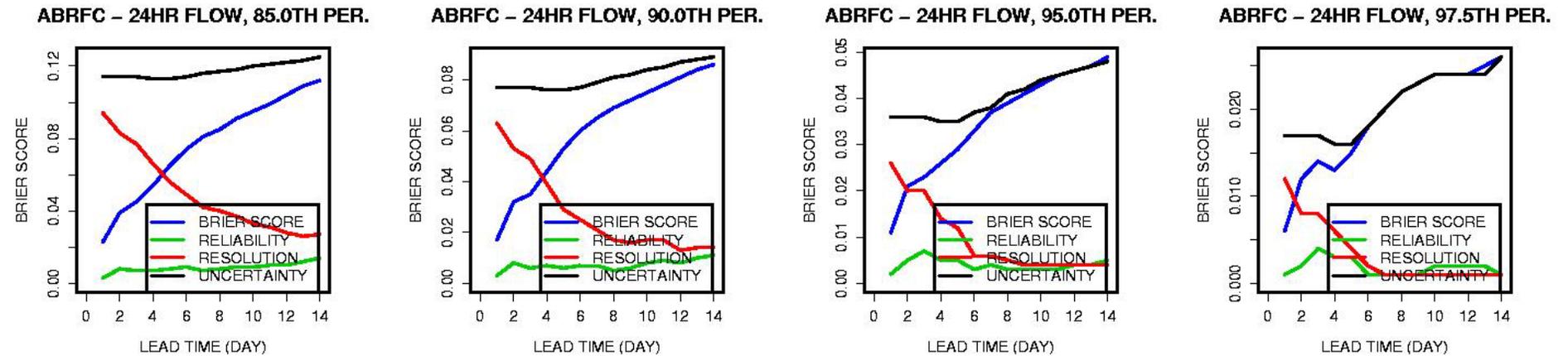
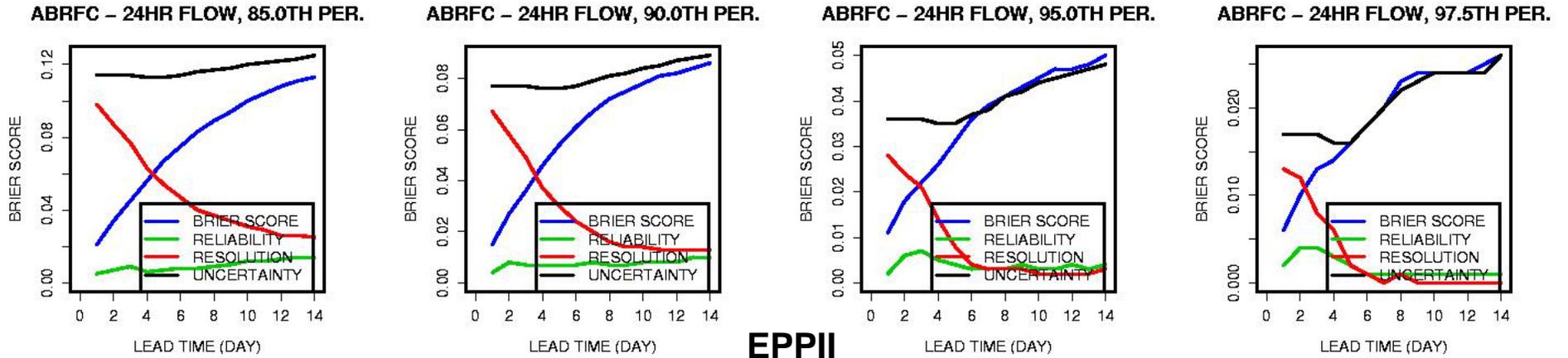
Reference flow: simulated



$$BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty}$$

Briers Score Statistics

Reference flow: simulated



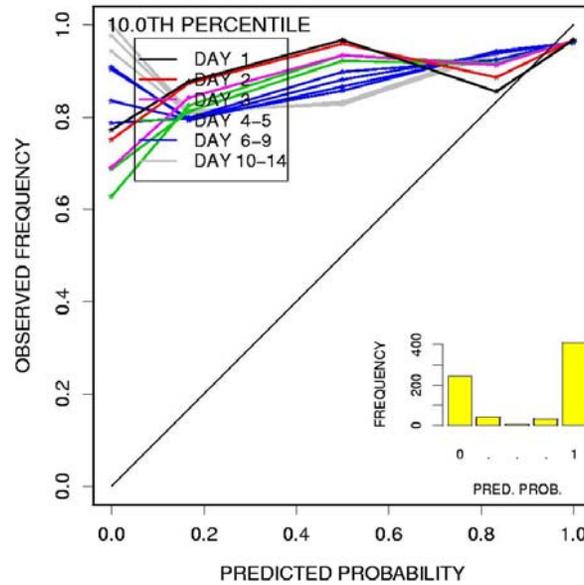
$$BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty}$$

Reliability statistics

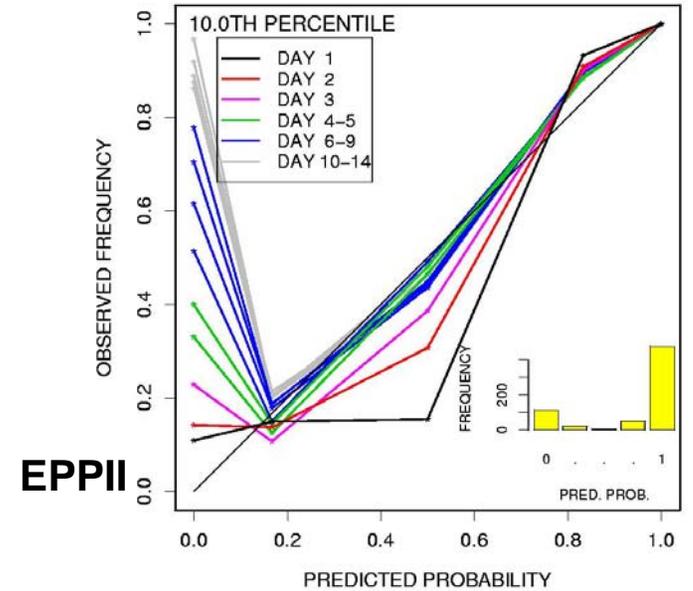
- Annual results with observed and simulated flows
- Threshold values:
10%, 25%, 50%, 75%, 85%, 90%, 95%, 97.5%
- Reliability plot with 5 probability bins (including bins for 0 and 1).
- Histogram gives sample size in each bin for lead day 1.
- Reliability Diagram measures agreement between forecast probability and mean observed frequency
- Deviation from diagonal gives conditional bias
 - Below diagonal: over-forecasting
 - Above diagonal: under-forecasting

Reliability Statistics

Reliability Diagram
(agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow

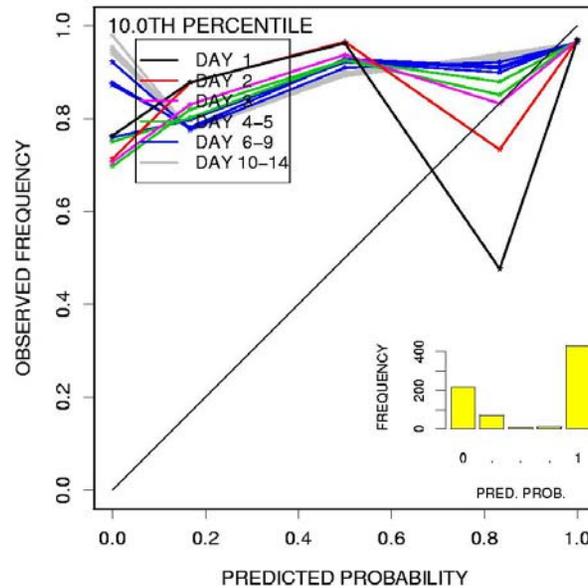


Reference flow: observed

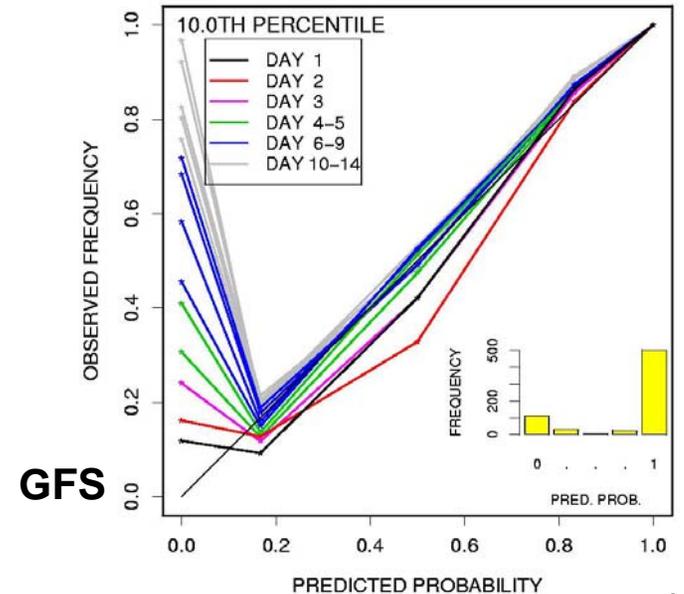


Reference flow: simulated

With 5 bins

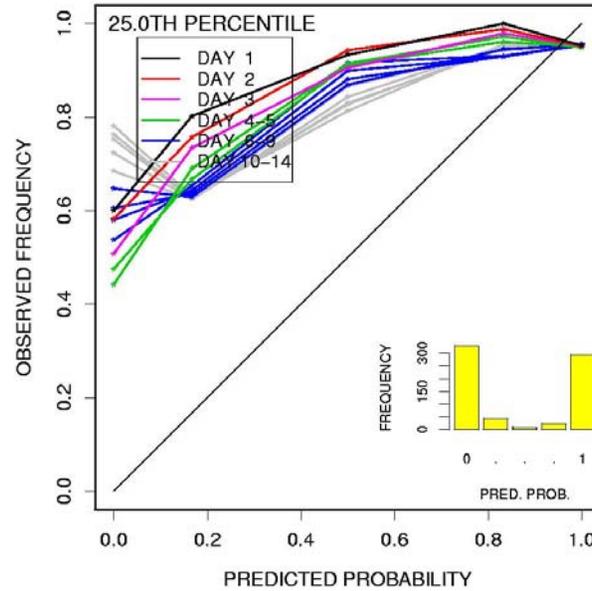


Deviation from diagonal gives conditional bias

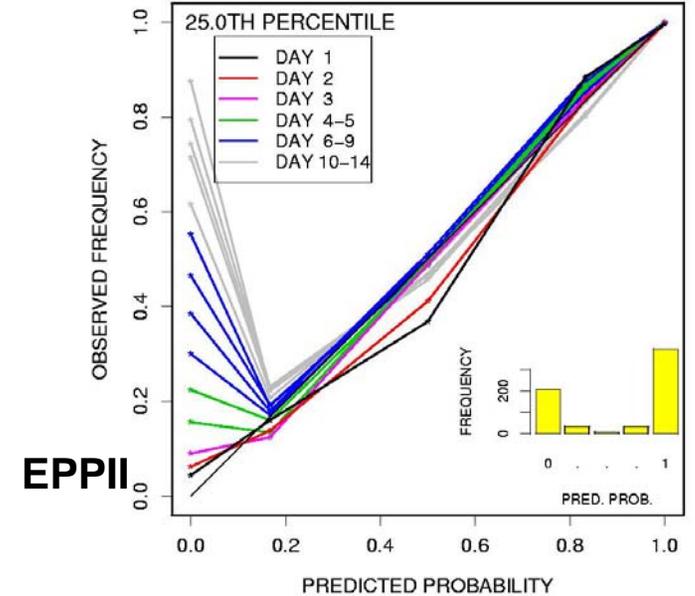


Reliability Statistics

Reliability Diagram
(agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow

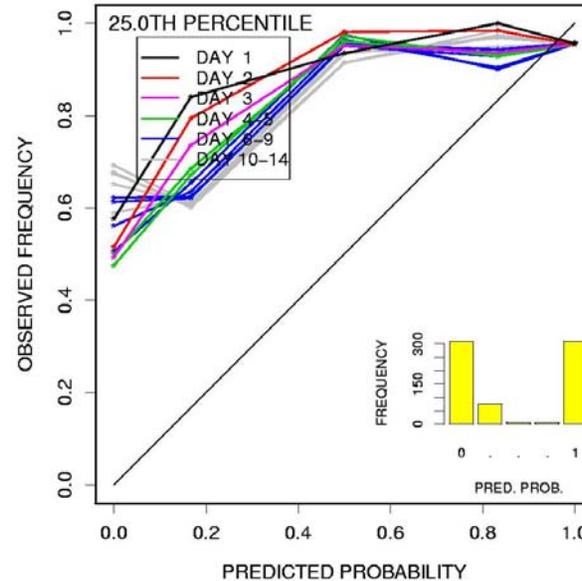


Reference flow: observed

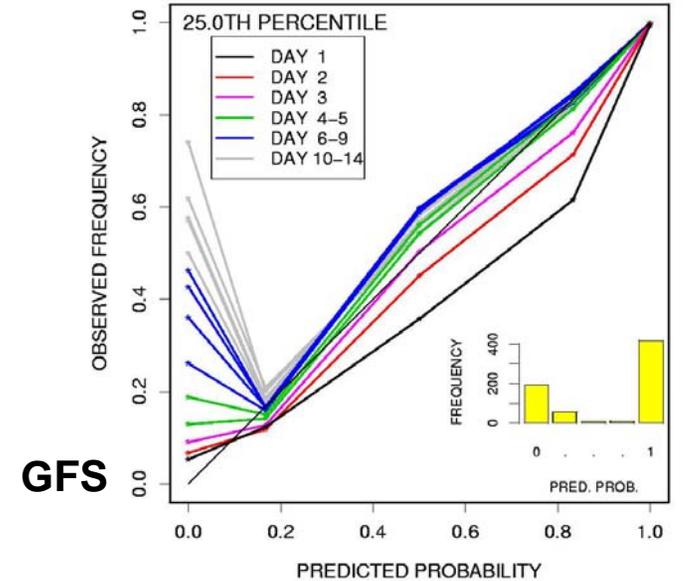


Reference flow: simulated

With 5 bins



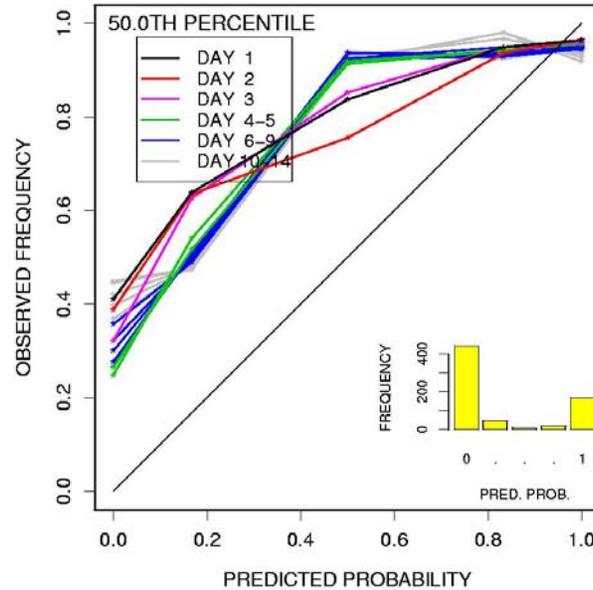
Deviation from diagonal gives conditional bias



GFS

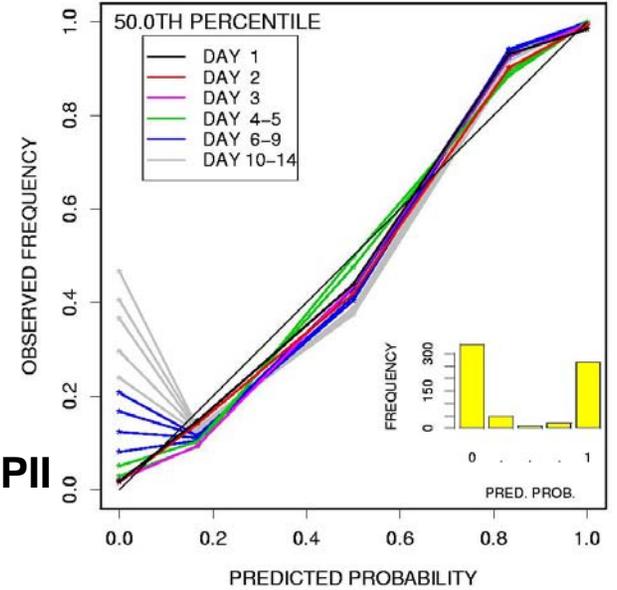
Reliability Statistics

Reliability Diagram
(agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow



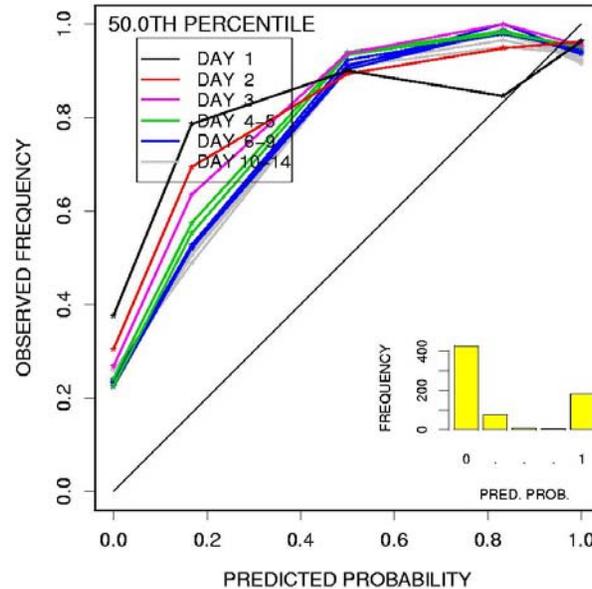
Reference flow: observed

EPPII



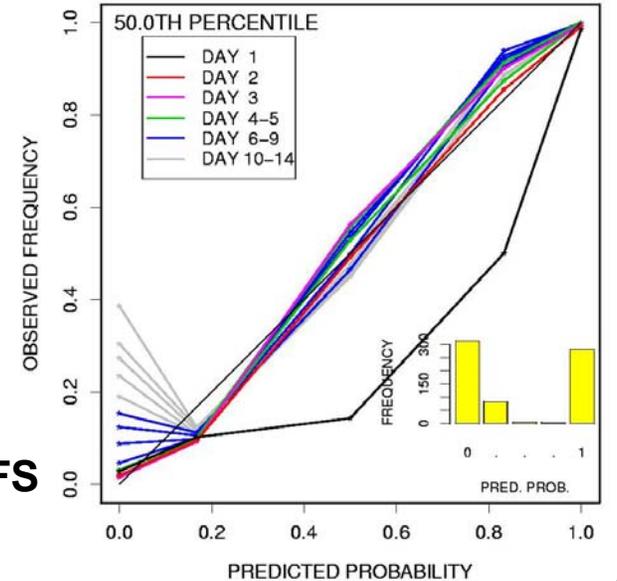
Reference flow: simulated

With 5 bins



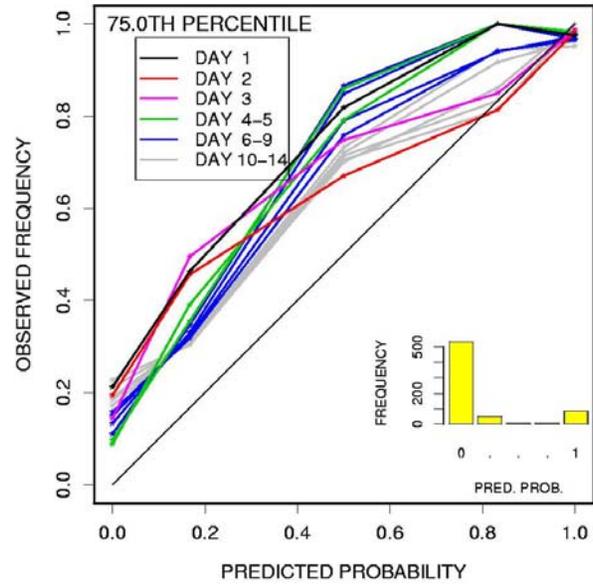
Deviation from diagonal gives conditional bias

GFS

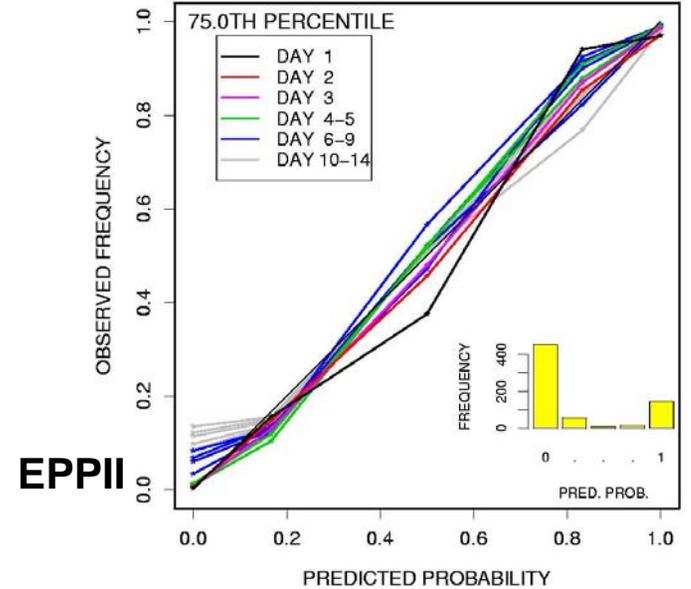


Reliability Statistics

Reliability Diagram
 (agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow



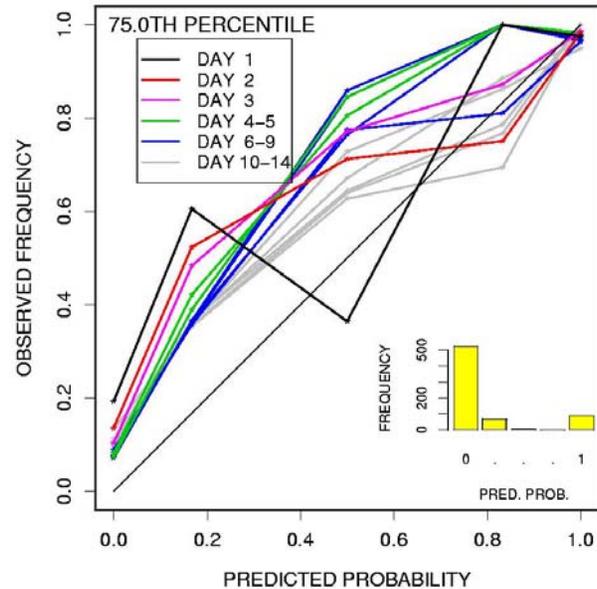
Reference flow: observed



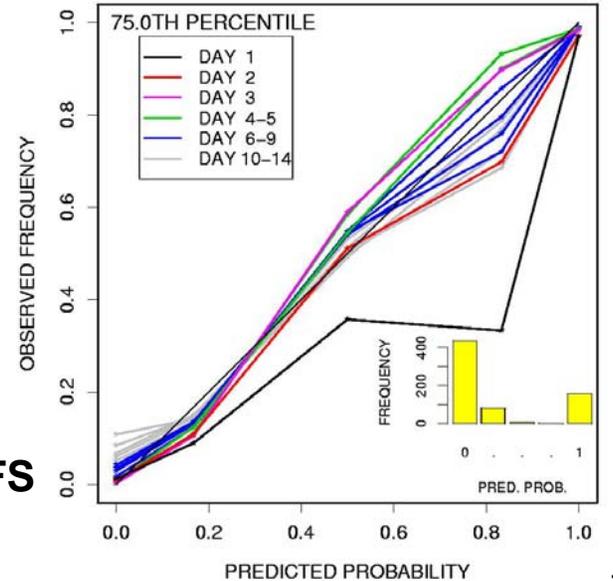
Reference flow: simulated

With 5 bins

Deviation from diagonal gives conditional bias

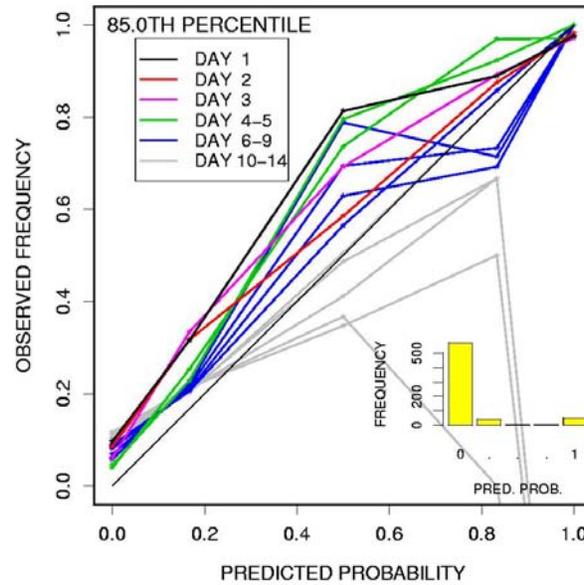


GFS

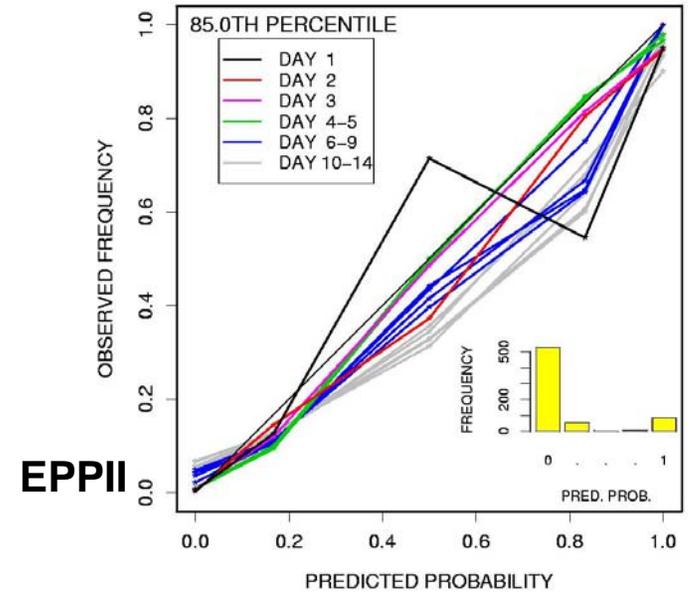


Reliability Statistics

Reliability Diagram
(agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow

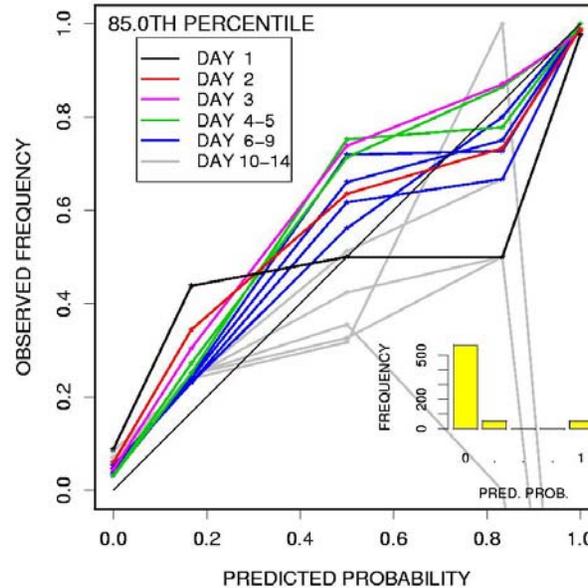


Reference flow: observed

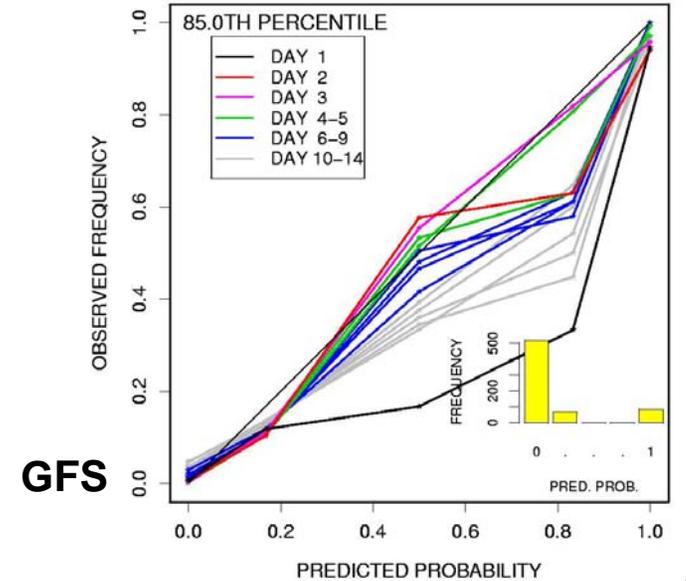


Reference flow: simulated

With 5 bins



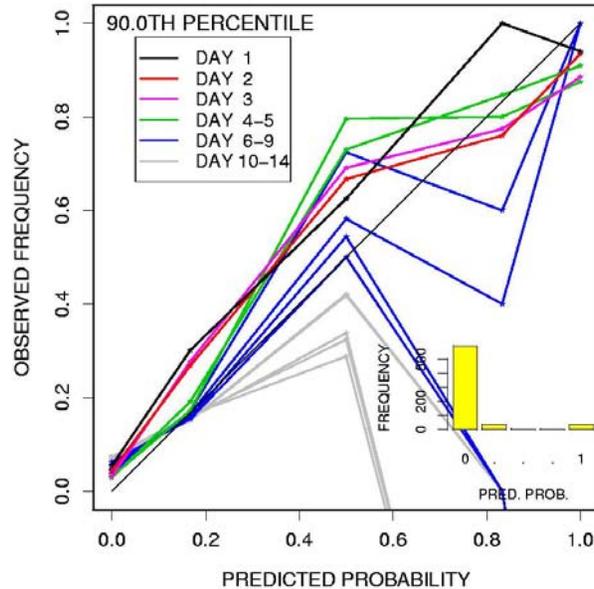
Deviation from diagonal gives conditional bias



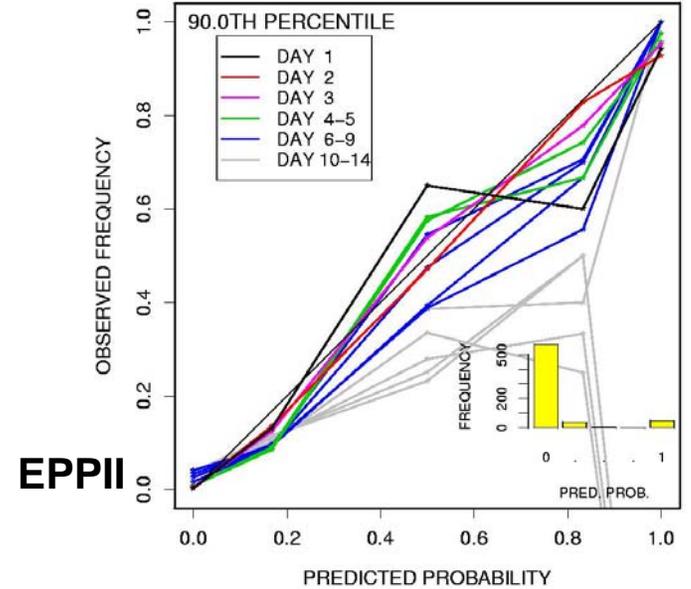
GFS

Reliability Statistics

Reliability Diagram
(agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow



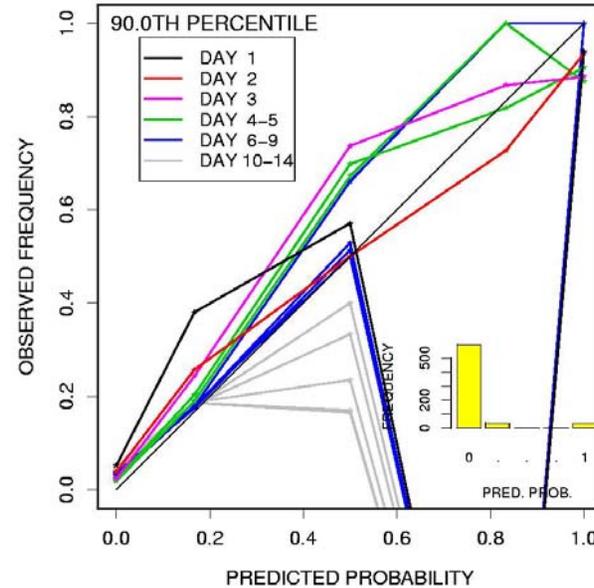
Reference flow: observed



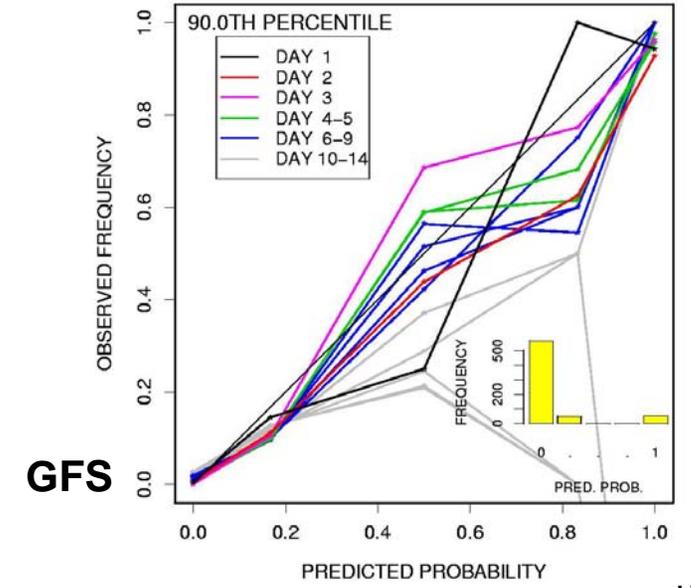
EPPII

Reference flow: simulated

With 5 bins



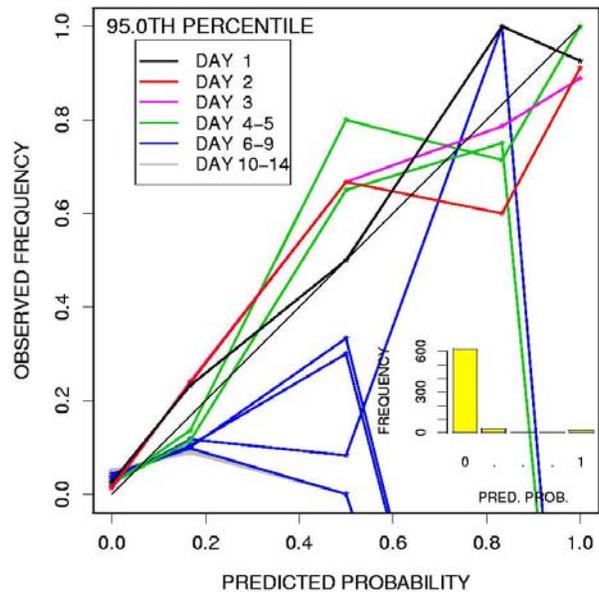
Deviation from diagonal gives conditional bias



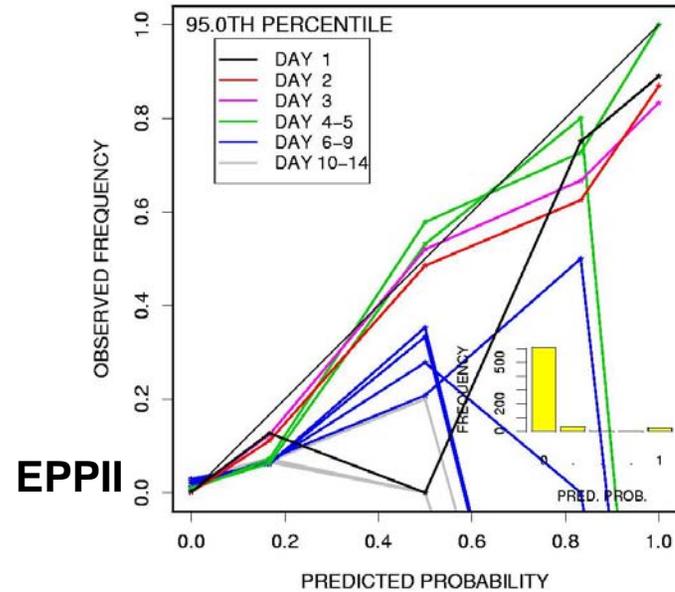
GFS

Reliability Statistics

Reliability Diagram
(agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow



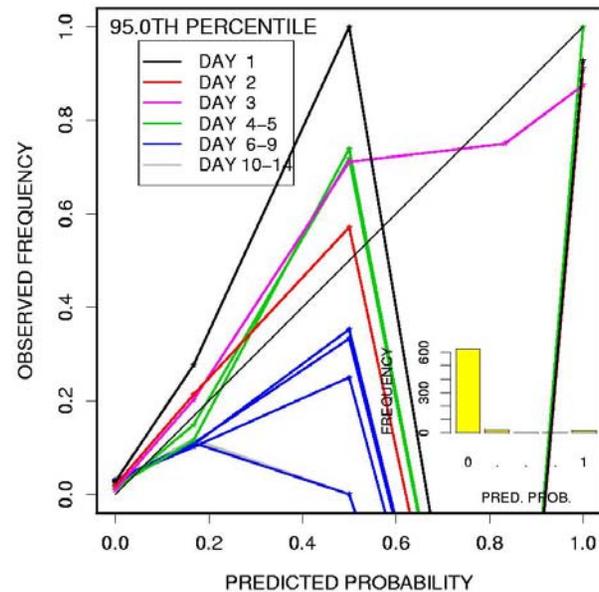
Reference flow: observed



EPPII

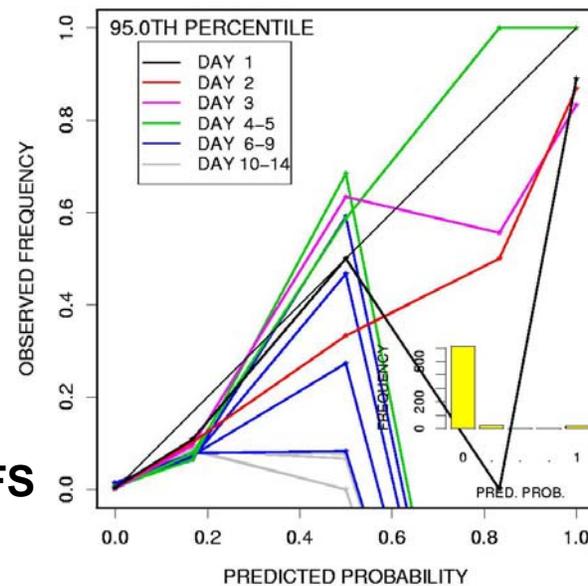
Reference flow: simulated

With 5 bins



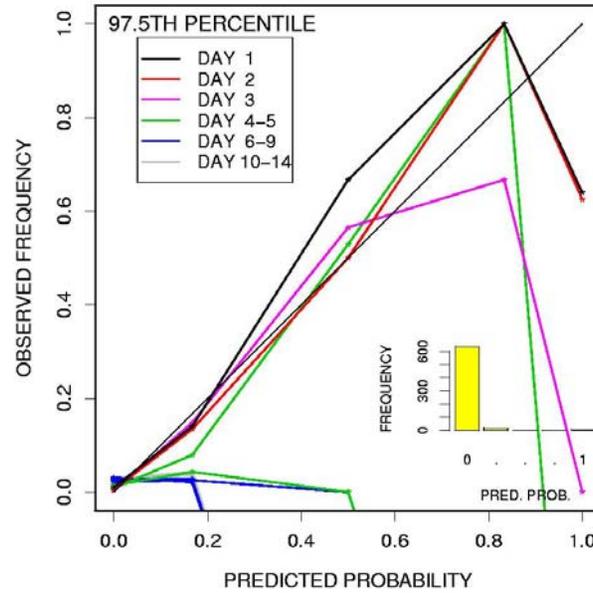
Deviation from diagonal gives conditional bias

GFS

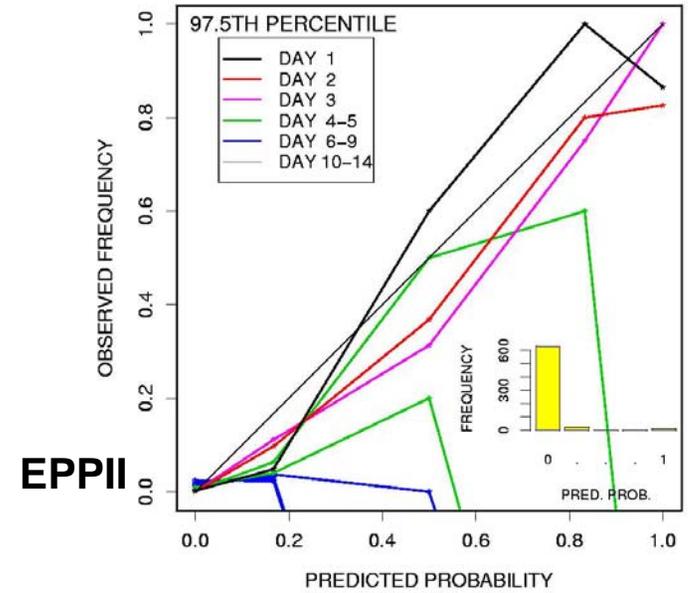


Reliability Statistics

Reliability Diagram
(agreement between forecast probability and mean observed frequency) for a range of threshold percentiles for the 24-hr annual flow



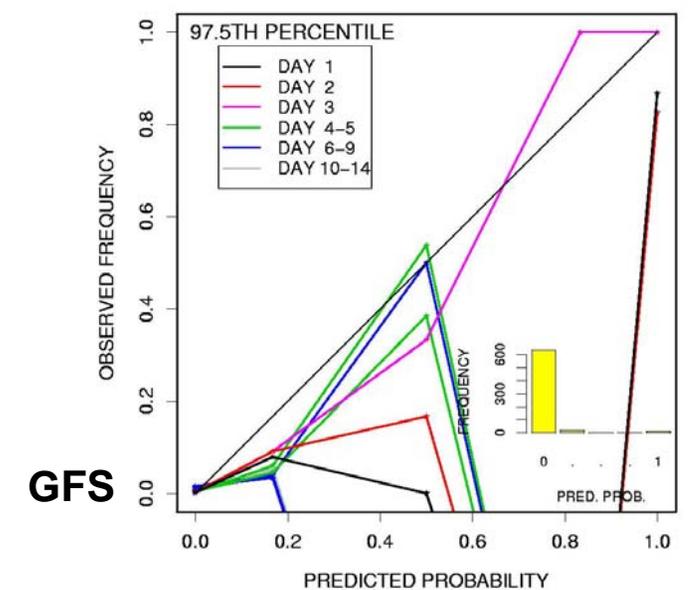
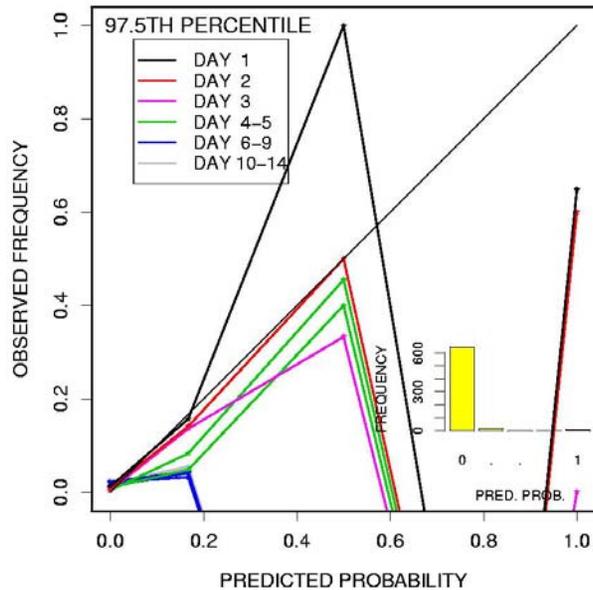
Reference flow: observed



Reference flow: simulated

With 5 bins

Deviation from diagonal gives conditional bias



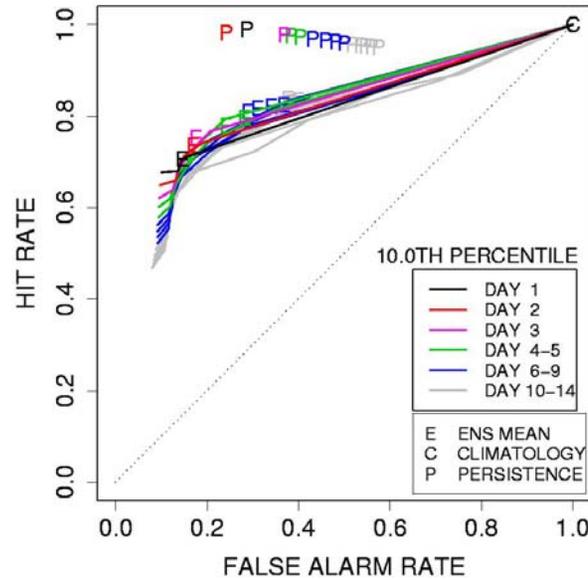
GFS

ROC statistics

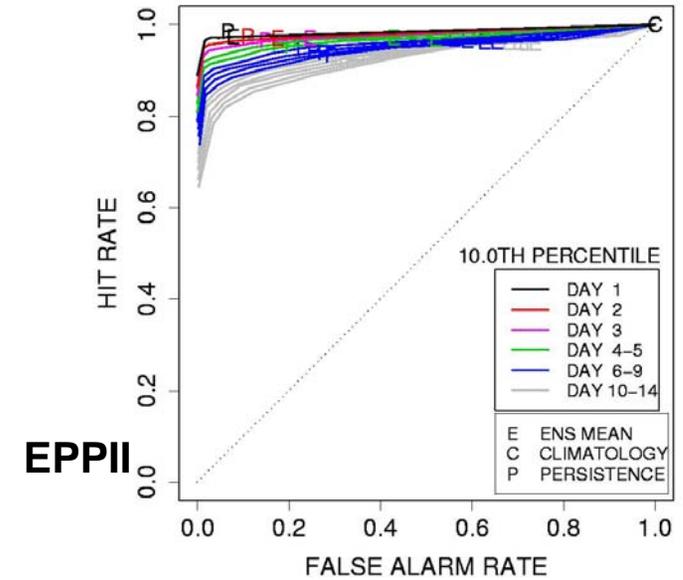
- Annual results with observed and simulated flows
- Threshold values:
10%, 25%, 50%, 75%, 85%, 90%, 95%, 97.5%
- ROC diagram with 10 points
- ROC points for Ensemble Mean, Monthly Climatology Mean, and Persistence forecasts
- Relative Operating Characteristic measures resolution (ability of forecast to discriminate between events & non-events)

ROC Statistics

ROC
 (ability of forecast to discriminate between events & non-events)
 for a range of threshold percentiles
 for the 24-hr annual flow

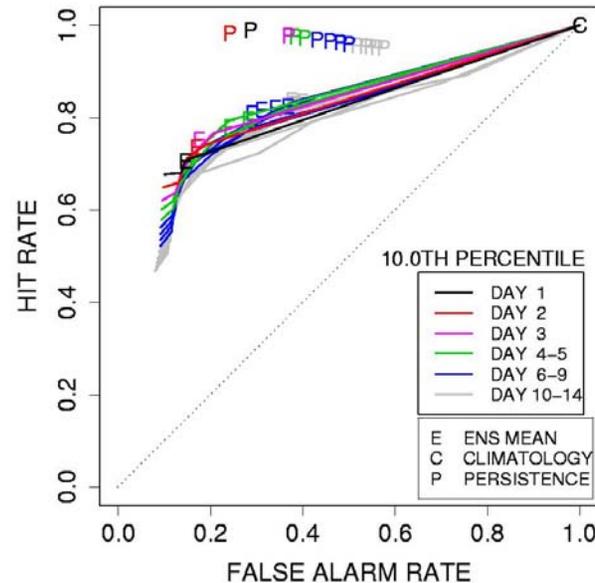


Reference flow: observed

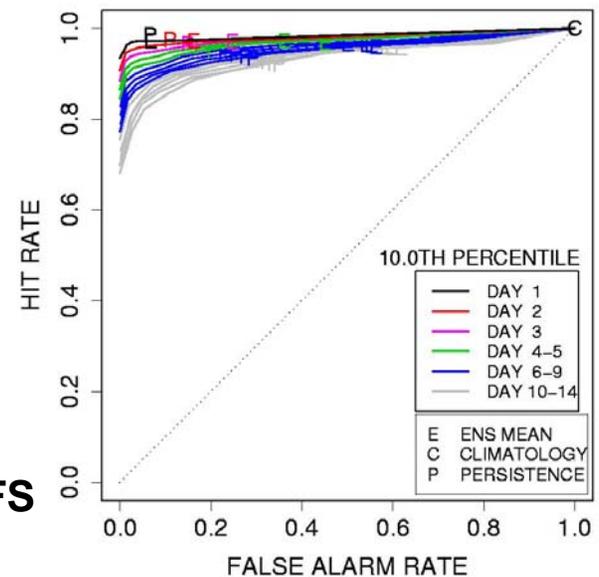


EPPII

Reference flow: simulated



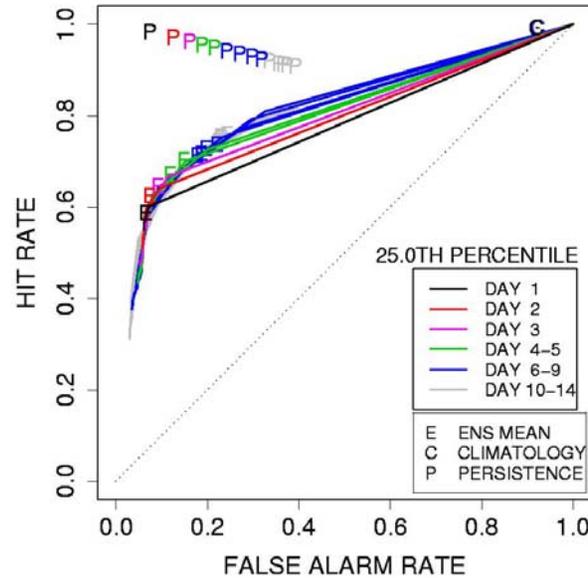
GFS



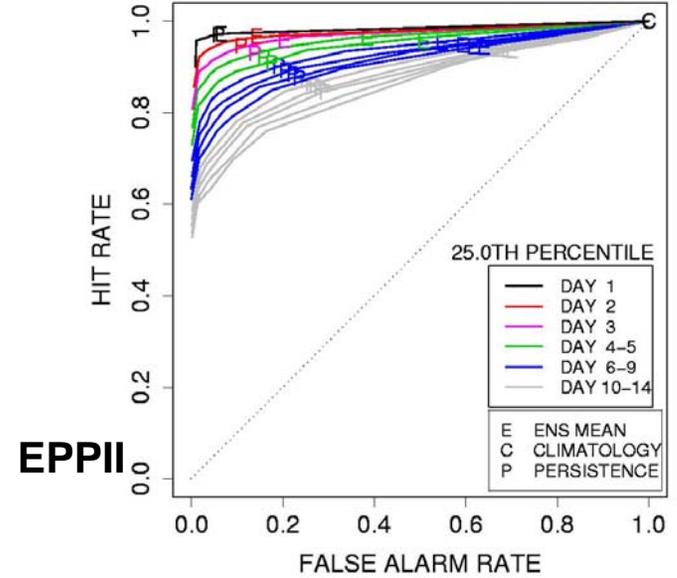
Perfect scores:
 $HR = 1$ and $FAR = 0$

ROC Statistics

ROC
 (ability of forecast to discriminate between events & non-events)
 for a range of threshold percentiles for the 24-hr annual flow

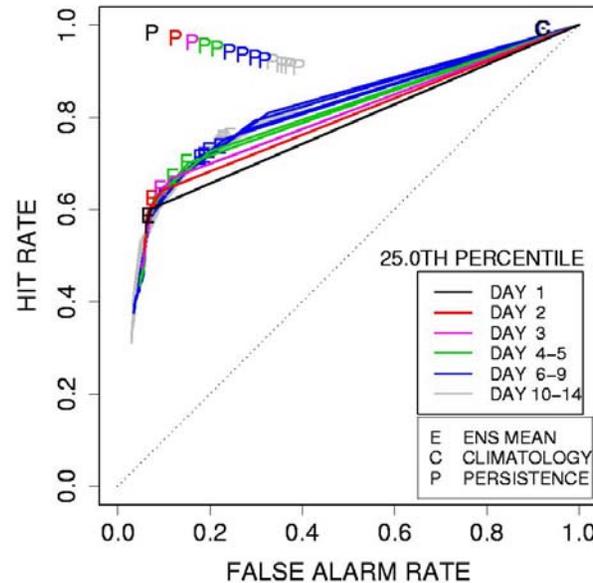


Reference flow: observed

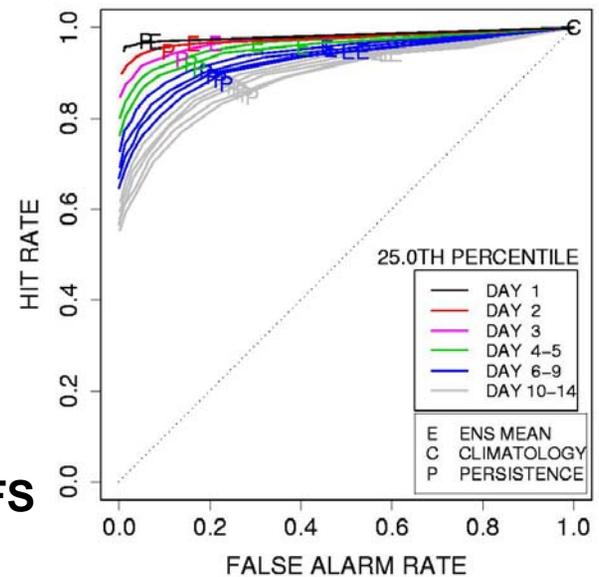


EPPII

Reference flow: simulated



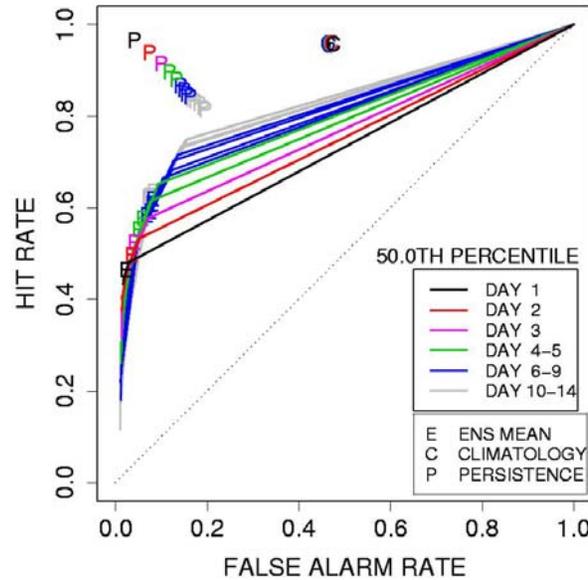
GFS



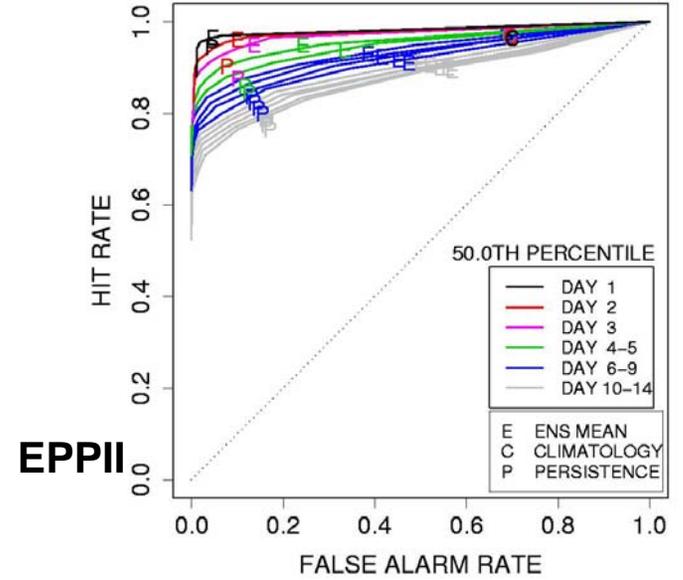
Perfect scores:
 $HR = 1$ and $FAR = 0$

ROC Statistics

ROC
(ability of forecast to discriminate between events & non-events) for a range of threshold percentiles for the 24-hr annual flow

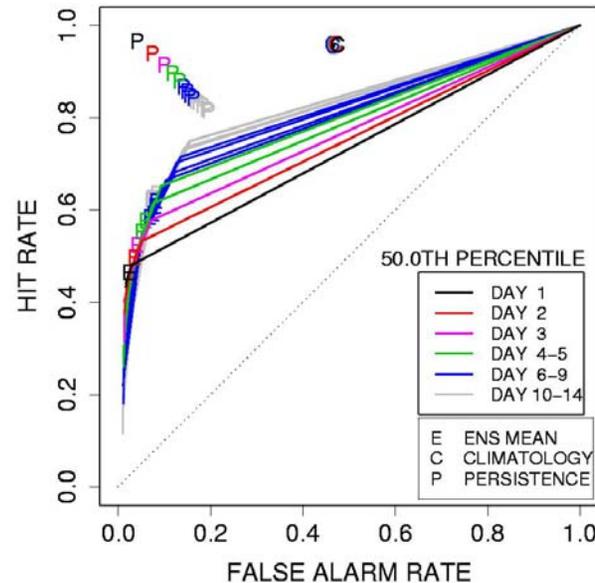


Reference flow: observed

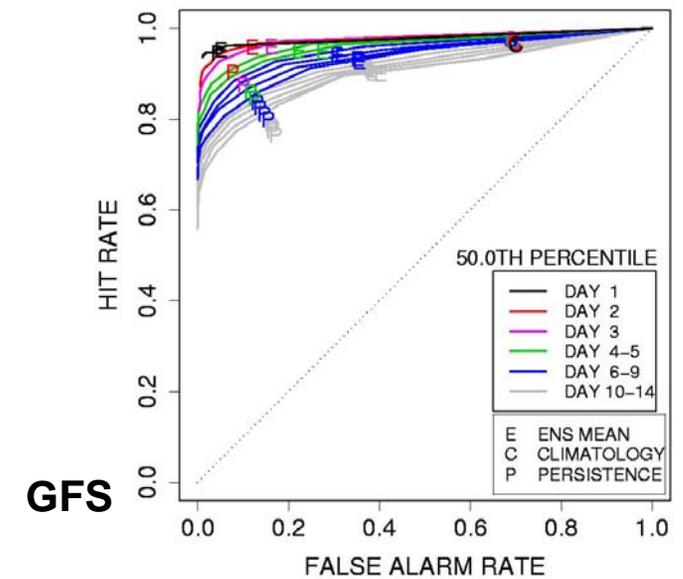


EPPII

Reference flow: simulated



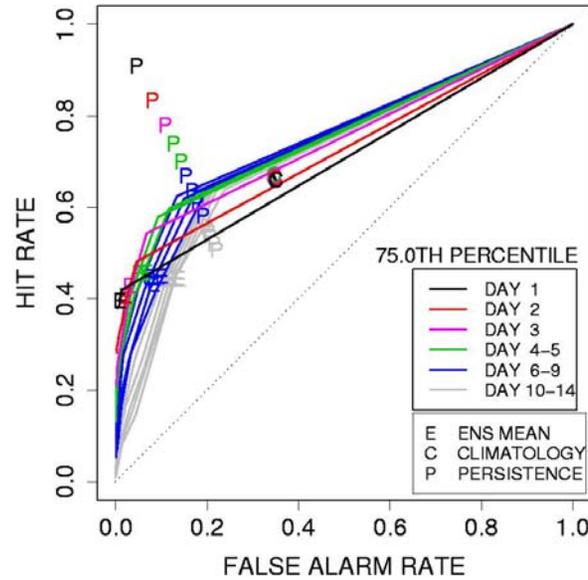
Perfect scores:
 $HR = 1$ and $FAR = 0$



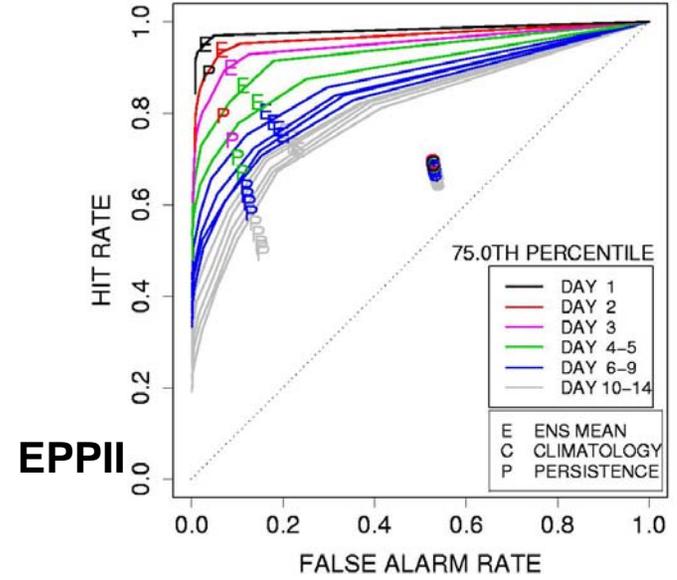
GFS

ROC Statistics

ROC
 (ability of forecast to discriminate between events & non-events)
 for a range of threshold percentiles
 for the 24-hr annual flow

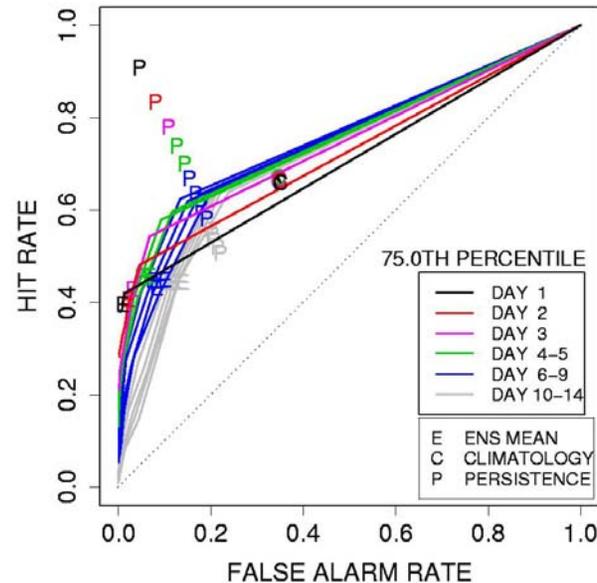


Reference flow: observed

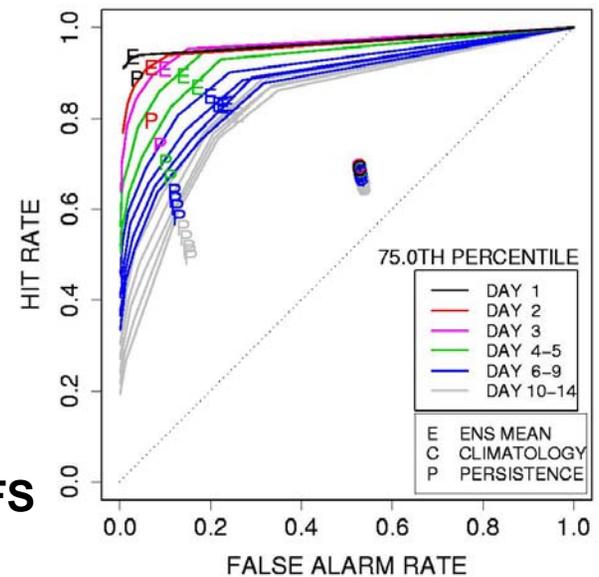


EPPII

Reference flow: simulated



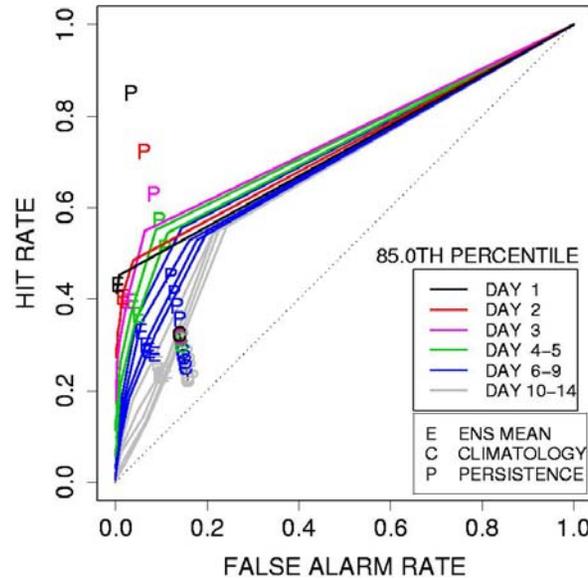
GFS



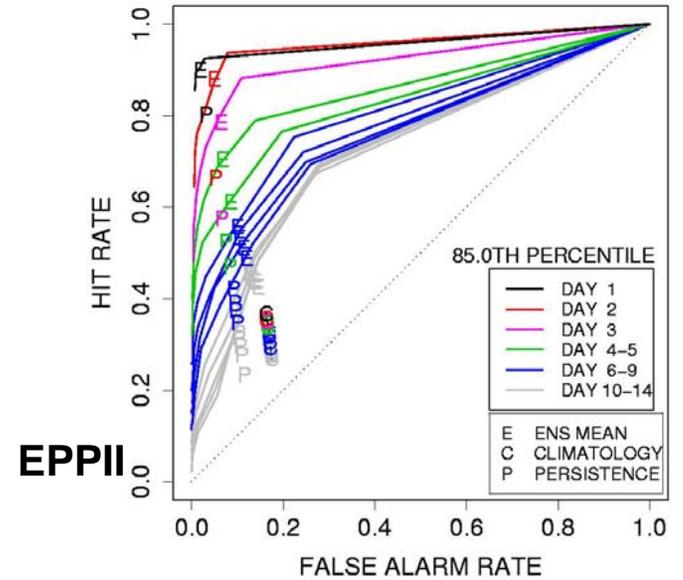
Perfect scores:
 $HR = 1$ and $FAR = 0$

ROC Statistics

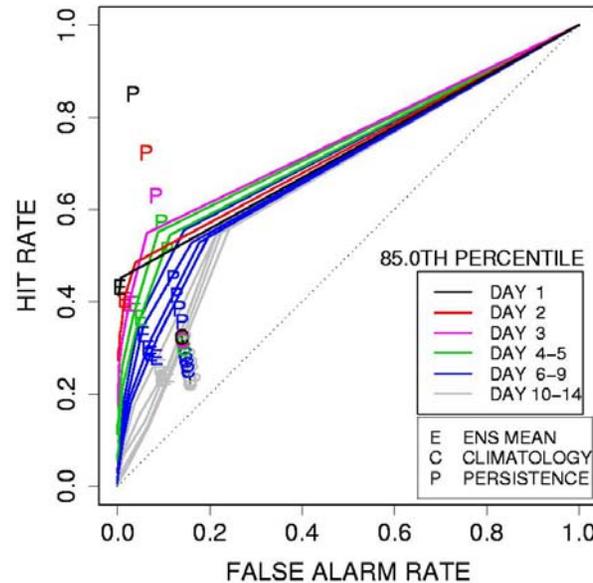
ROC
 (ability of forecast to discriminate between events & non-events)
 for a range of threshold percentiles
 for the 24-hr annual flow



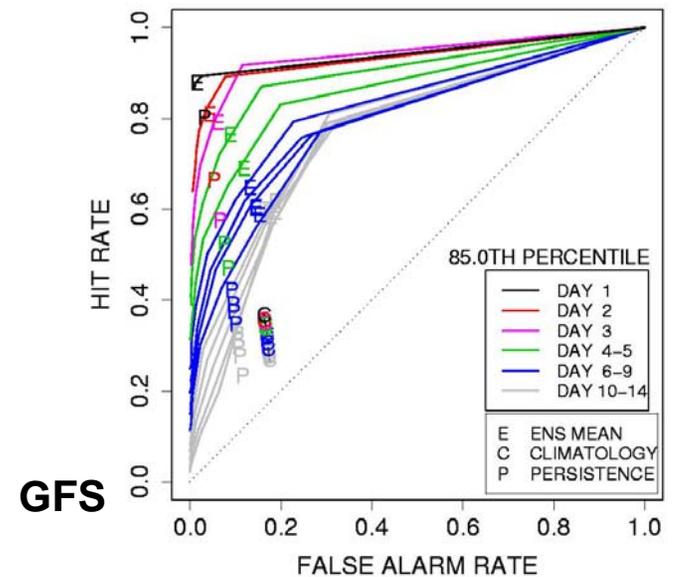
Reference flow: observed



Reference flow: simulated



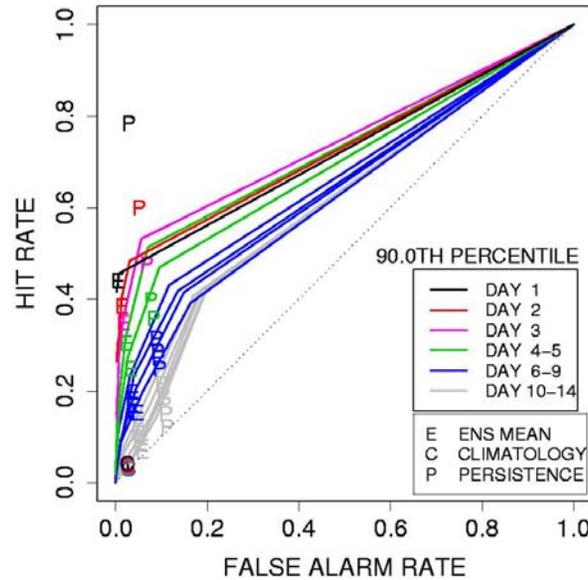
Perfect scores:
 $HR = 1$ and $FAR = 0$



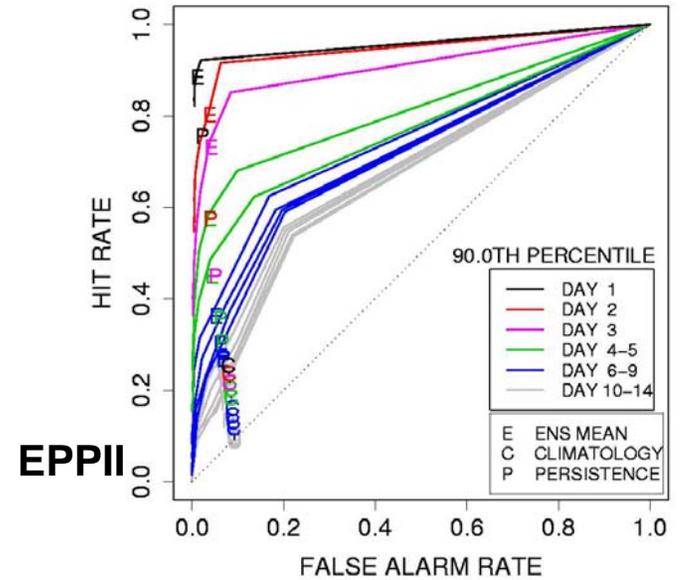
GFS

ROC Statistics

ROC
(ability of forecast to discriminate between events & non-events) for a range of threshold percentiles for the 24-hr annual flow

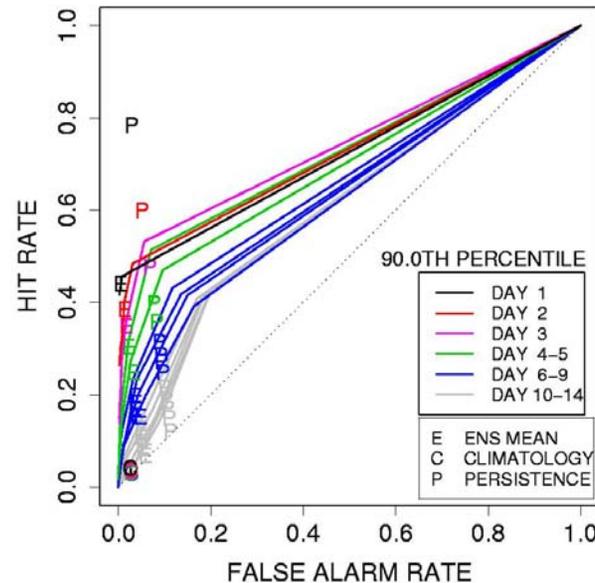


Reference flow: observed

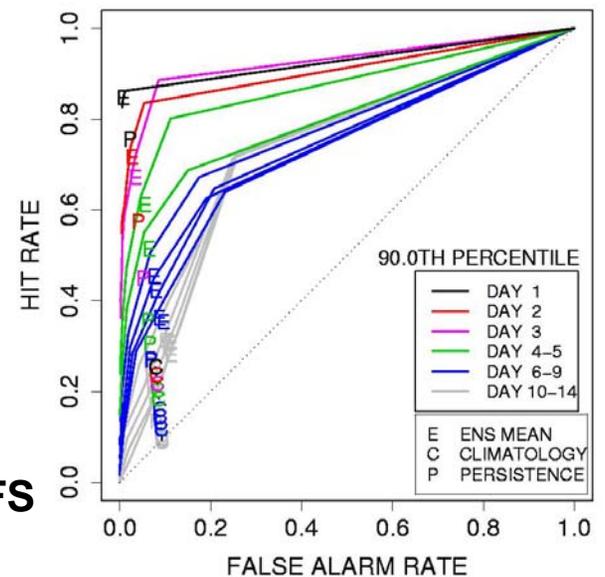


EPPII

Reference flow: simulated



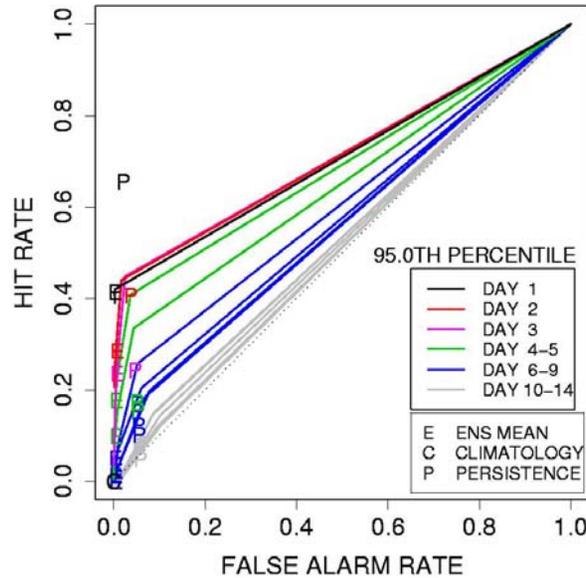
GFS



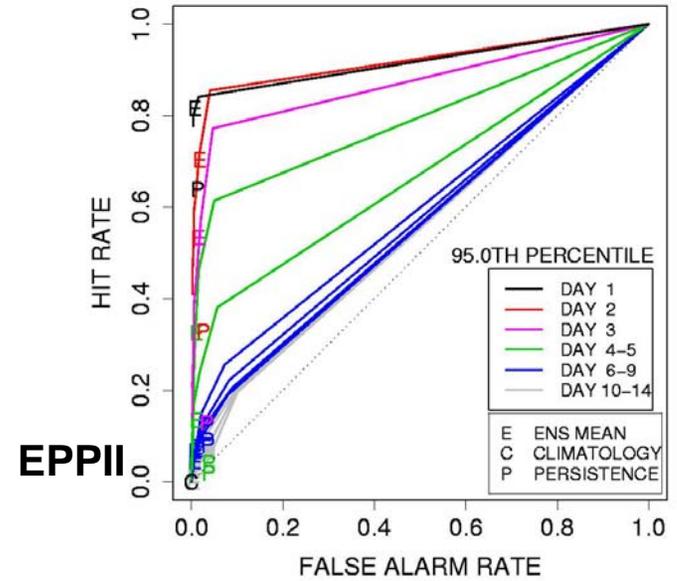
Perfect scores:
 $HR = 1$ and $FAR = 0$

ROC Statistics

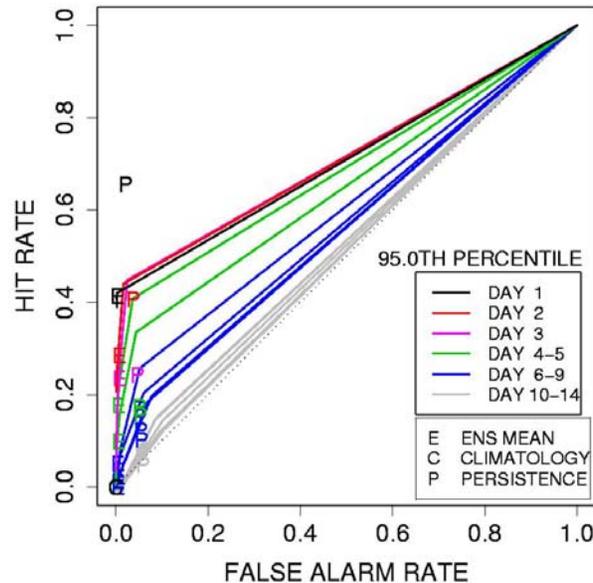
ROC
(ability of forecast to discriminate between events & non-events) for a range of threshold percentiles for the 24-hr annual flow



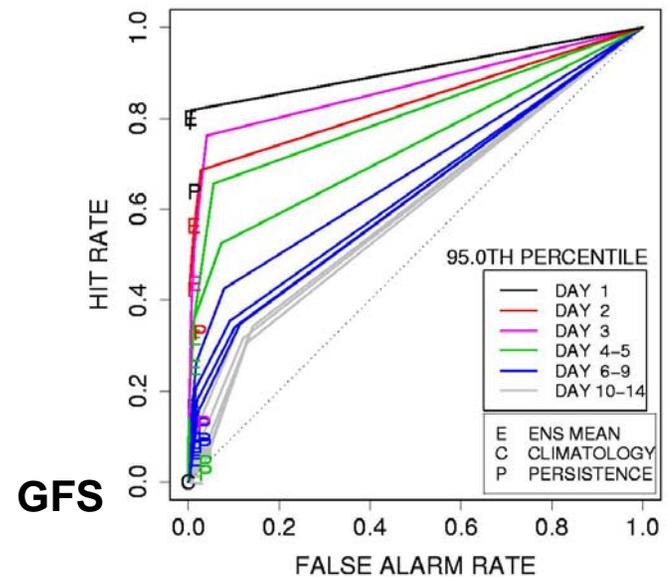
Reference flow: observed



Reference flow: simulated

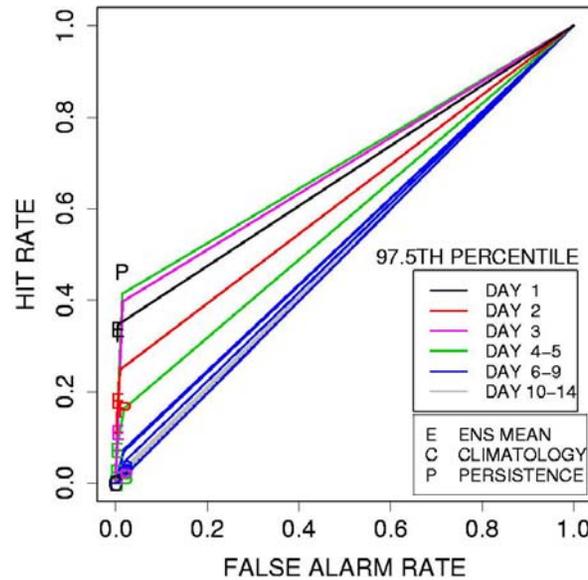


Perfect scores:
 $HR = 1$ and $FAR = 0$

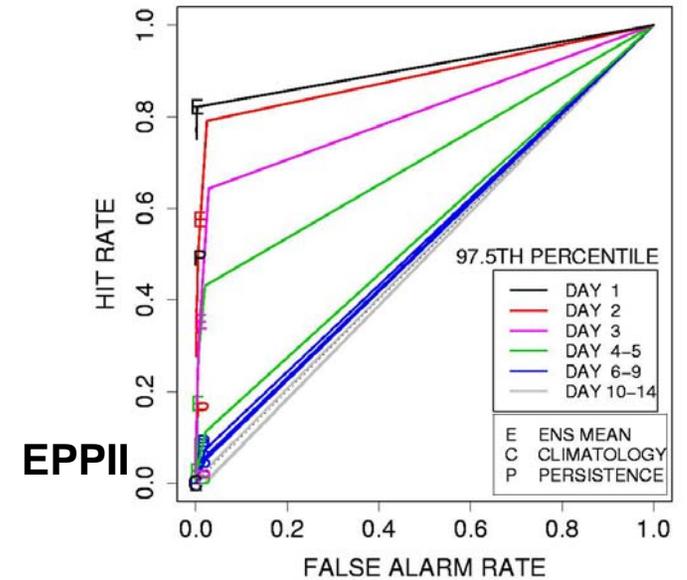


ROC Statistics

ROC
 (ability of forecast to discriminate between events & non-events)
 for a range of threshold percentiles
 for the 24-hr annual flow

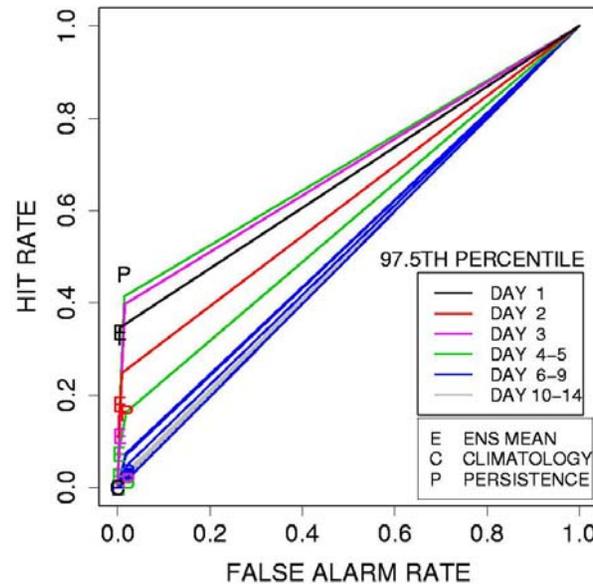


Reference flow: observed

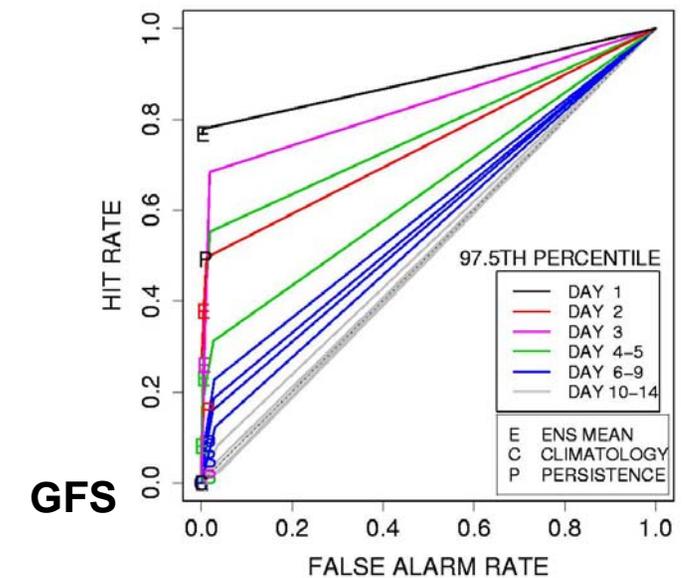


EPPII

Reference flow: simulated



Perfect scores:
 $HR = 1$ and $FAR = 0$



GFS