HIC INNOVATION MEETING
Part II – Prototype ESPVS

October 27-28, 2003
Kansas City, MO

By
Dave Brandon / CBRFC

The Current ESP Verification System (ESPVS) Was Developed in the late 90s

Software is not modular
It is hard to maintain
It is hard to make improvements/changes

It Lacks Probabilistic Statistical Metrics:
- Reliability Diagrams
- Discrimination Diagrams
- Rank Probability Score
- Rank Probability Skill Scores

It does not use the same code (ESPADP) when producing probabilistic forecasts

The current version does not work (very well).

Design of a Prototype ESPVS

RTI: Will develop the design document and write the code (in cooperation with OHD and CBRFC).
(First draft about finished, some Code written.)

U of A: Will provide the science/math for the statistical metrics.
(Currently in progress.)

CBRFC: Will provide test sets, and test and evaluate prototype code
(Expect delivery in the new year)

Prototype Verification System For NWSRFS-ESP
(Next ESPVS - 3 Components)

Component 1	Component 2	Component 3
ESPESG	ESPADP	ESPVDP
ESP Ensemble Series Generator	ESP Analysis Display Program	ESP Verification Display Program

ESPESG
ESPEnsembleSeriesGenerator

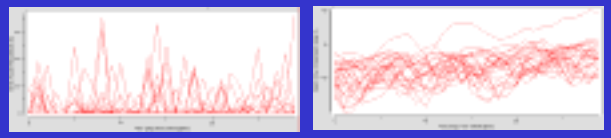
Uses These

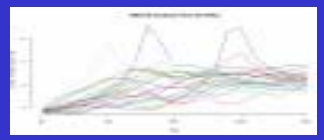
MAPS/MATs

Generates These

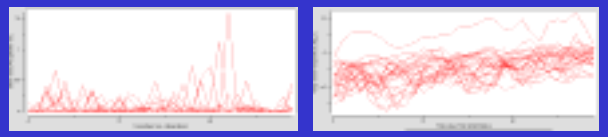
Flow Ensembles

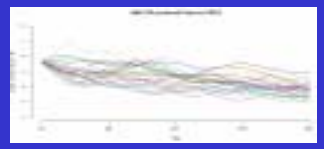
1971



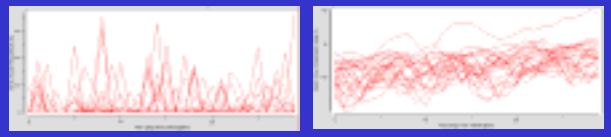


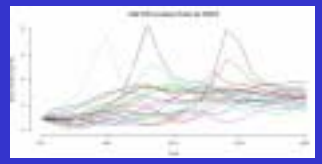
1972





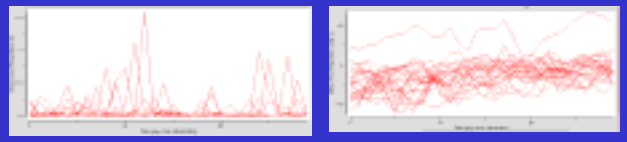
1973

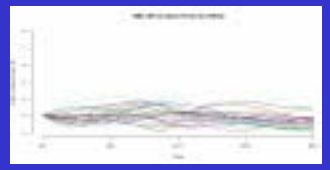




. . . . . .

n





ESPADP
ESPAnalysisDisplayProgram

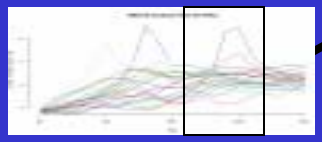
Uses These

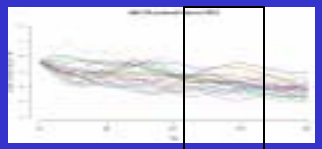
Generates These

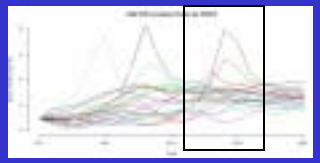
Flow Ensembles

Forecasts (e.g.)

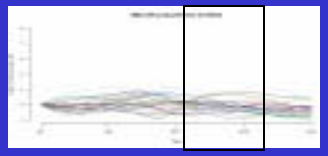
Analysis Window

1971  → APR-JUL Volume

1972  → APR-JUL Volume

1973  → APR-JUL Volume

. . . . . .

n  → APR-JUL Volume

ESPVDP
ESPVerificationDisplayProgram

Uses These
Forecasts (e.g.)

1971 APR-JUL Volume

1972 APR-JUL Volume

1973 APR-JUL Volume

. . . . . .

n APR-JUL Volume

Generates These
Verification Displays

ENSEMBLE INFORMATION
Talagrand Diagram

PROBABILISTIC VERIFCATION
Ranked Probability Score RPS
Ranked Probability Skill Score RPSS
Discrimination Diagrams
Reliability Diagram

FOUR MAIN METRICS TO BE INCLUDED

(1) Ranked Probability Score RPS

(2) Ranked Probability Skill Score RPSS

(3) Discrimination Diagrams

(4) Reliability Diagram

Ranked Probability Score RPS

The Ranked Probability Score (RPS) is used to assess the overall forecast performance of the probabilistic forecasts.

A perfect forecast would result in a RPS or zero.

Gives credit for forecasts close to observation…
Penalizes forecasts further from the observation.

Looks at the entire distribution (all traces).

Good overall summary stat at all flow levels.

It is a categorical Brier Score.

Because the actual RPS value is difficult to evaluate independently, the use of the RPS in the absence of reference forecasts is limited to forecast comparison among different forecast locations.

Can be used to analyze regional consistency, i.e., possible need for recalibration.

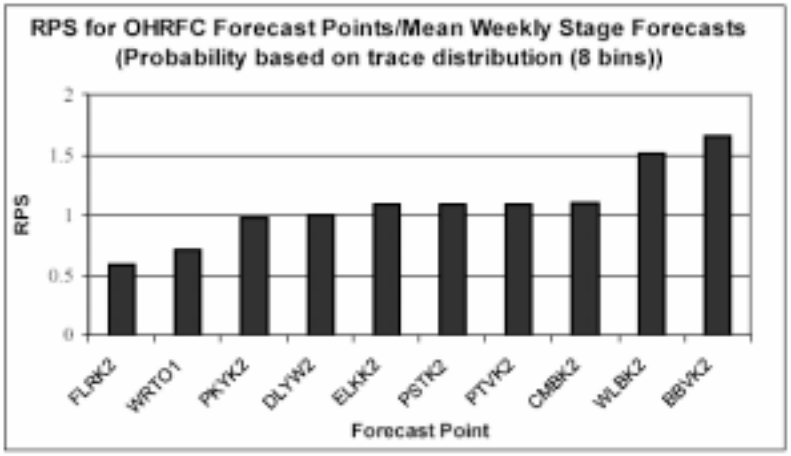


Figure 3: RPS analysis results for mean weekly stage forecasts.

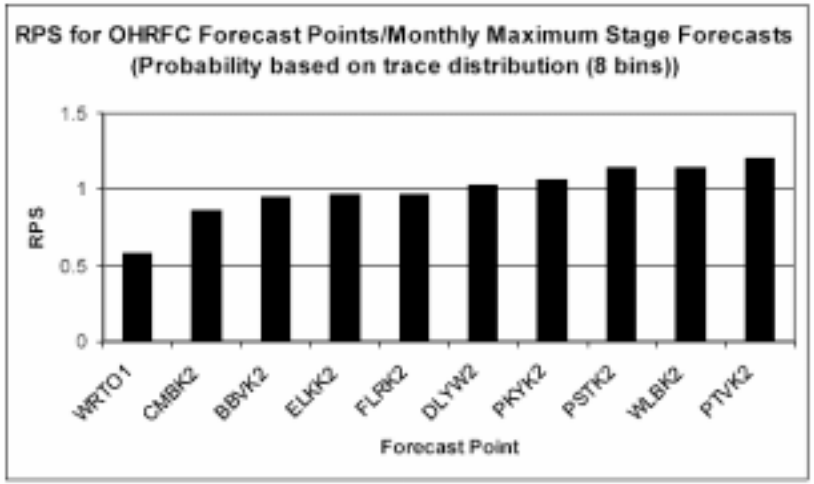
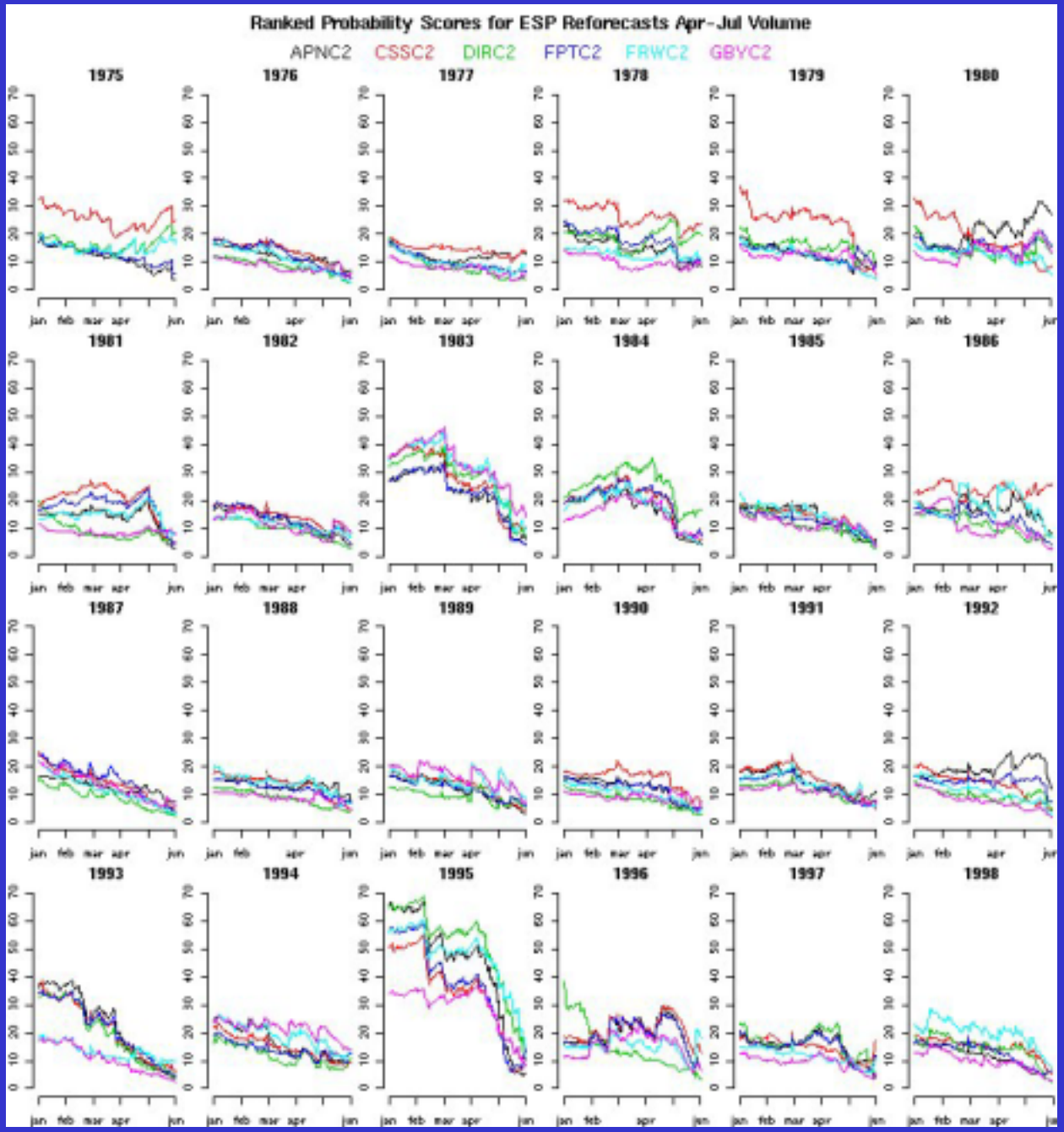


Figure 4: RPS analysis results for maximum monthly stage forecasts.

Credit: "Verification on NWS Probabilistic Hydrologic Forecasts" – Franz/Sorooshian – U. of AZ

RPS used to compare various basins. (Note RPS here was computed with 100 bins.)



Ranked Probability Skill Score RPSS

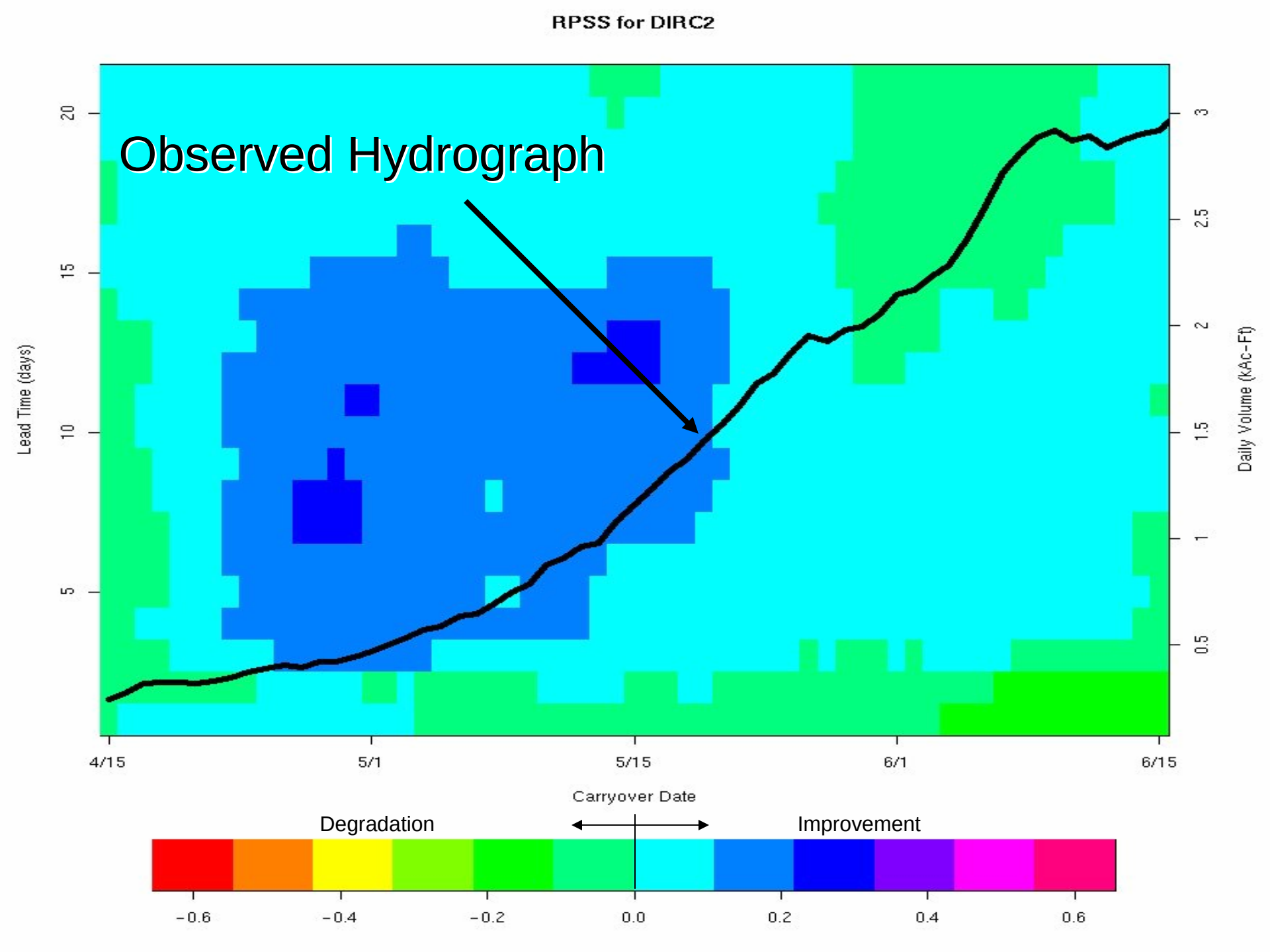
Useful to compare the forecast of interest to a reference forecast, e.g., climatology.

Is expressed as a percent improvement, e.g., over the reference forecast (e.g. how much better/worse than climatology).

Perfect score is 1.0 (100%).

Negative score indicates forecasts performed worse than reference forecast.

Credit: Presentation-"Evaluation of NWS Ensemble Sreamflow Prediction" – Kristie Franz – U. of AZ


RPSS for DIRC2
Observed Hydrograph
Lead Time (days)
20
15
10
5
Daily Volume (kAc-Ft)
3
2.5
2
1.5
1
0.5
4/15
5/1
5/15
6/1
6/15
Carryover Date
Degradation
Improvement
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6

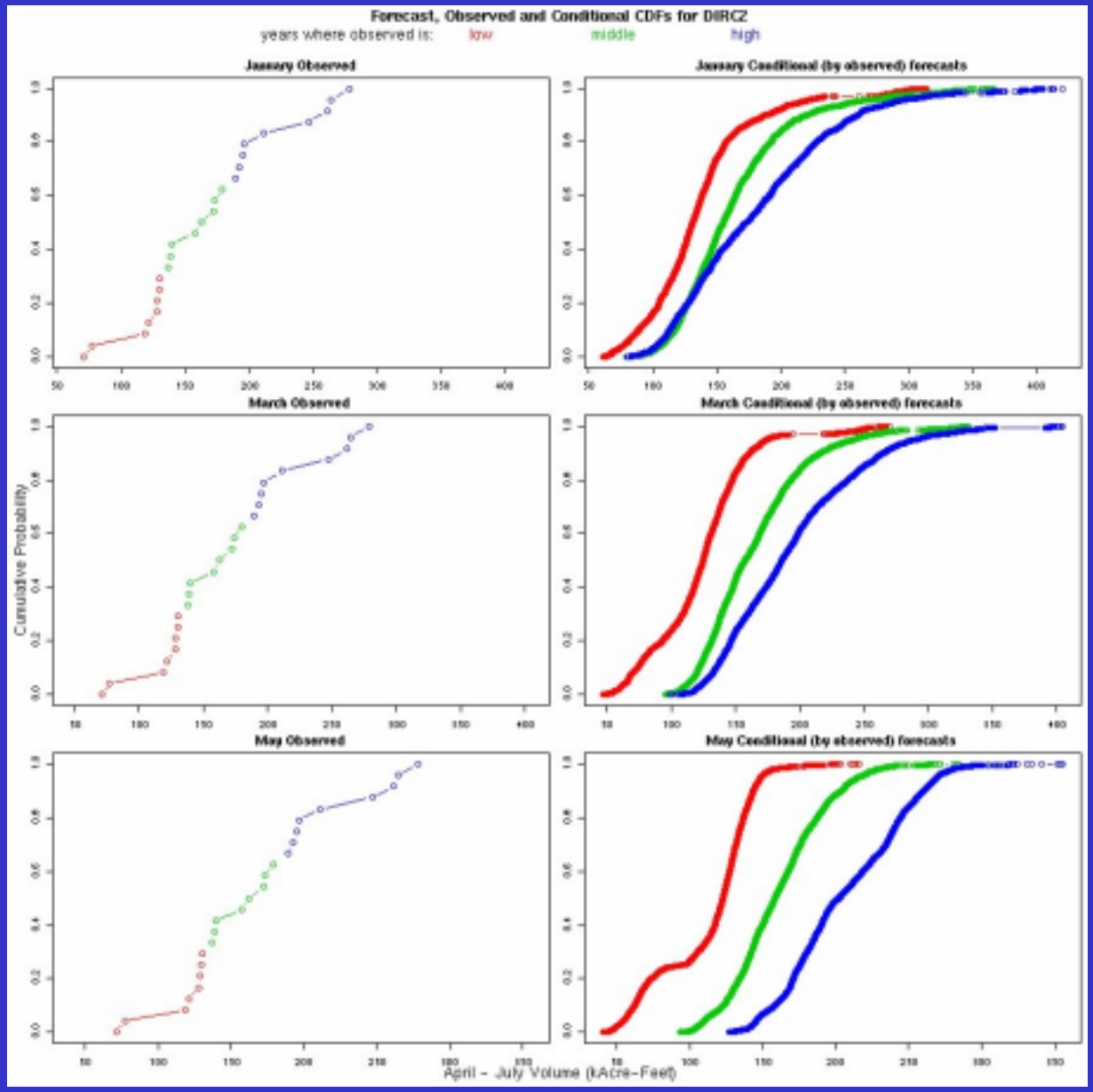
Discrimination Diagram

A discrimination diagram displays forecast probability distribution(s) conditioned on observation(s).

For example, given that the observations were high flows, did the forecasts prior to the observations reflect high flows.

Discrimination seen through CDFs conditioned on observation:

CDFs are sorted and averaged according to the observed volume. Good discrimination indicated by seperation between conditional CDFs



Observed Forecast

Reliability Diagram

A reliability diagram is used to display forecast reliability for a particular ‘type’ of forecast (e.g., flows above flood stage, flows in the lower quartile).

It displays how often an observation occurred given a particular forecast probability.

For a set of forecasts where a forecast probability value was given to a particular observation, the forecasts are considered perfectly reliable if the relative frequency of the observations equals the forecast probability.

Reliability Diagram