Evaluation of NWS Ensemble Streamflow Prediction

Kristie Franz

Department of Hydrology and Water Resources
University of Arizona

Holly C. Hartmann, Soroosh Sorooshian, Roger Bales

NOAA CLIMAS

ISPE

Hydrology Lab

NWS

HWR

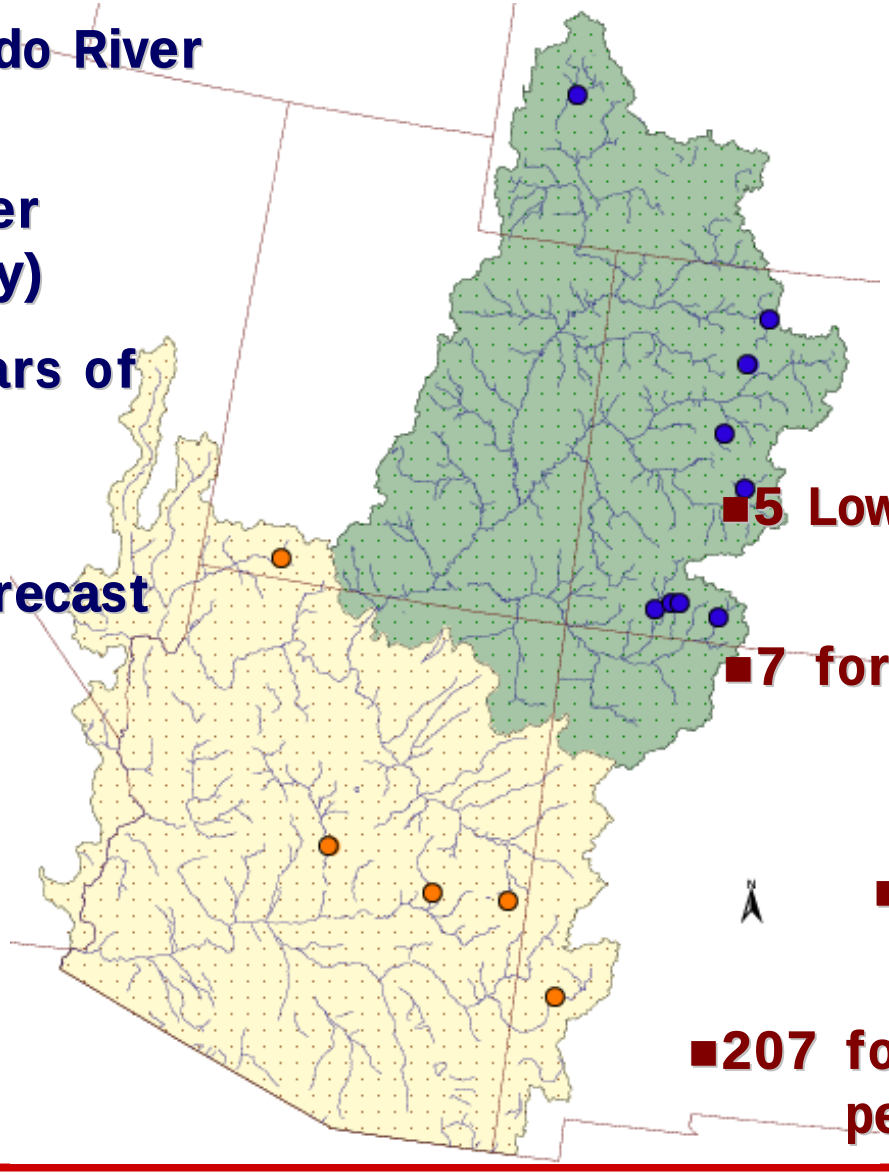
Sahra

Forecast Points

- ✓ Colorado River Basin
- ✓ Calibrated
- ✓ Data available
- ✓ Headwater location
- ✓ High Interest

Forecast Points

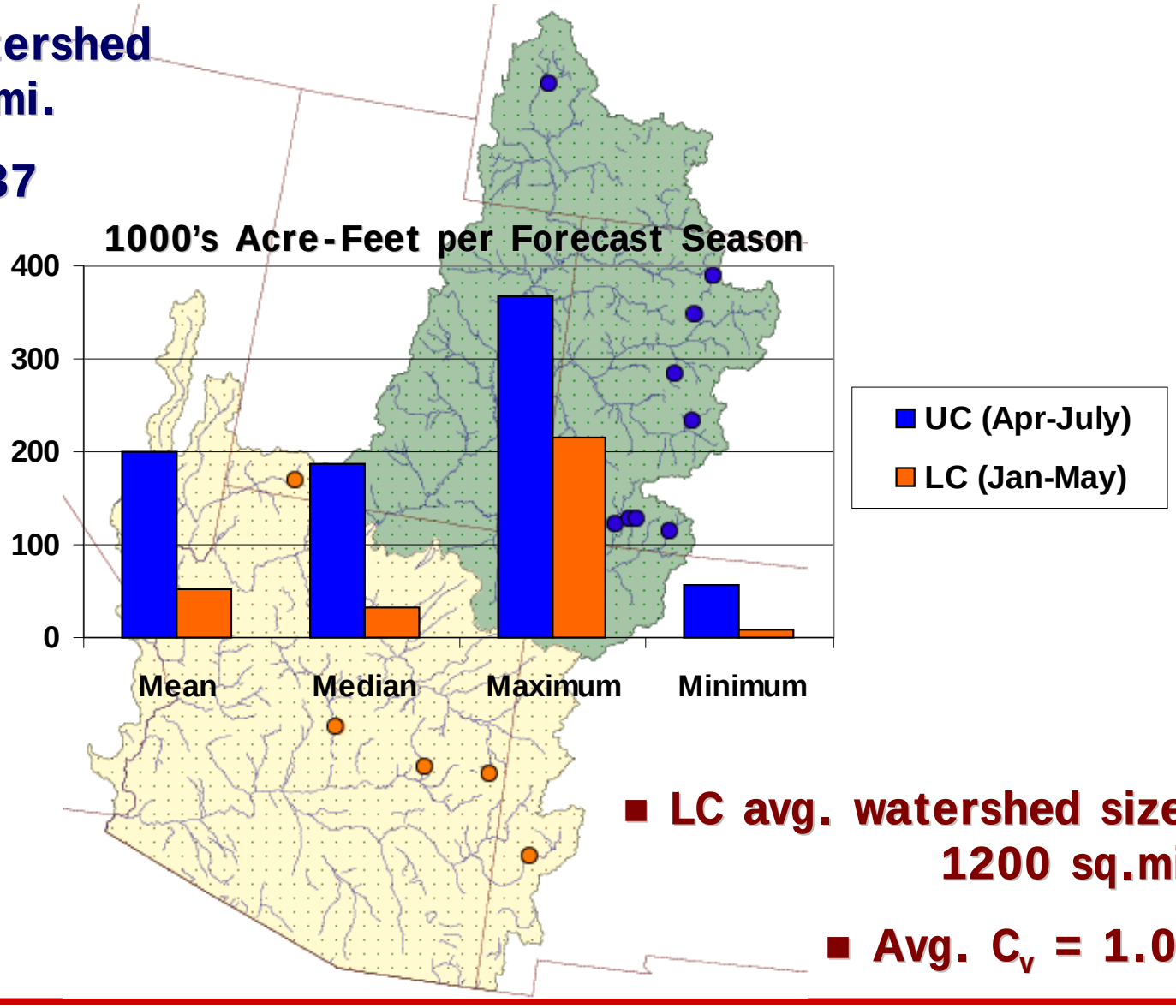
- ◆9 Upper Colorado River Basin (UC)
- ◆11 forecasts per season (April–July)
- ◆Average 31 years of observed data
- ◆337 forecasts evaluated per forecast period



- ■5 Lower Colorado River Basin Sites (LC)
- ■7 forecasts per season (Jan-May) (9 for Virgin R)
- ■Avg. 42 years of observed data
- ■207 forecasts evaluated per forecast period

Forecast Points: Basin Streamflow Averages

- UC avg. watershed size: 400 sq. mi.
- Avg. C_v = .37



- LC avg. watershed size: 1200 sq.mi.
- Avg. C_v = 1.06

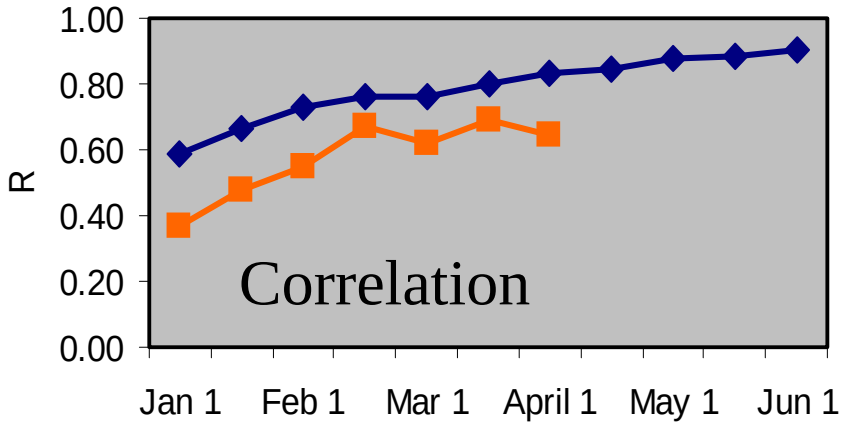
Generation & Evaluation

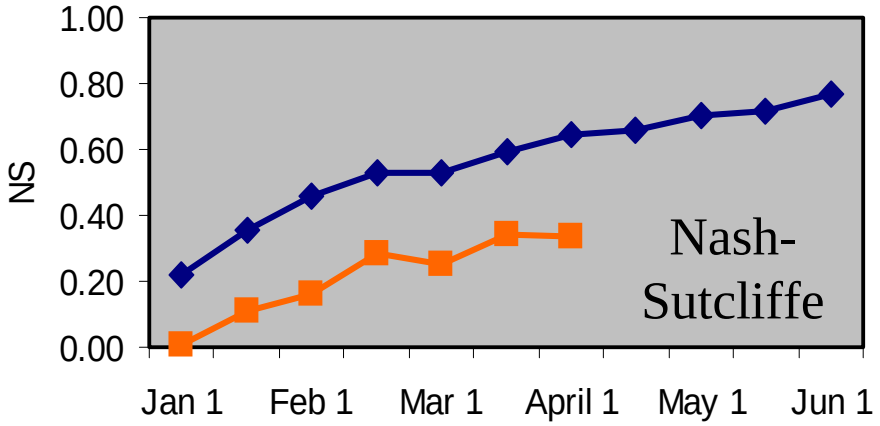
1. Generated historical ESP forecasts for all years with T, P & Q data, for the same forecast periods reported by NWS.
2. Calculated statistics for each forecast period.
   - Created single-value forecasts by taking median & "best" traces; results in deterministic forecasts
   - Disclaimer: Not Recommended, but familiar
     - MAE/Standard Deviation of observed, PBIAS, R, Nash-Sutcliffe
   - Probabilistic Forecasts
     - RPSS = % improvement over climatology
     - Discrimination & Reliability

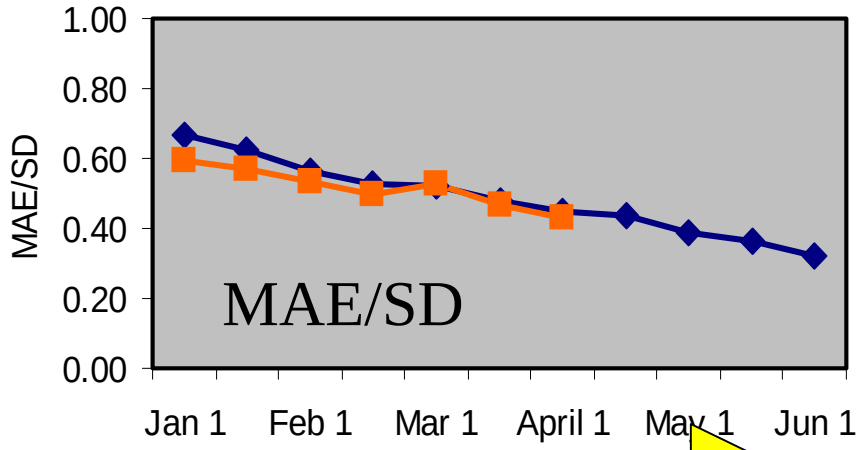
Average Basin Statistics for Median Forecast Trace

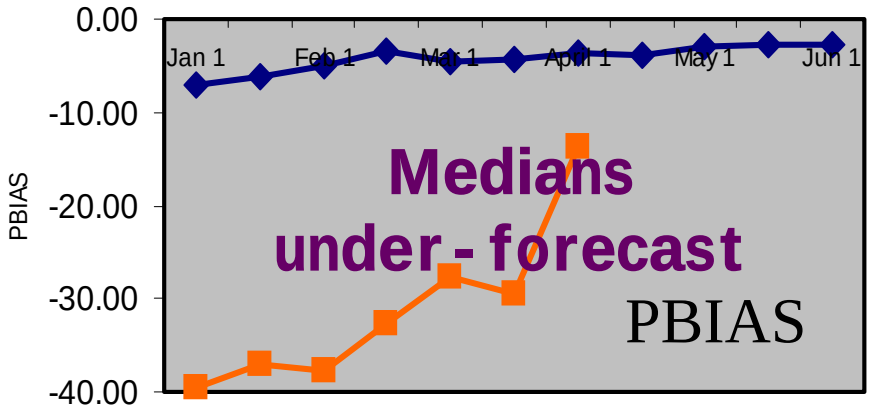
UC Med
LC Med

Median trace for UC performs better than LC on average



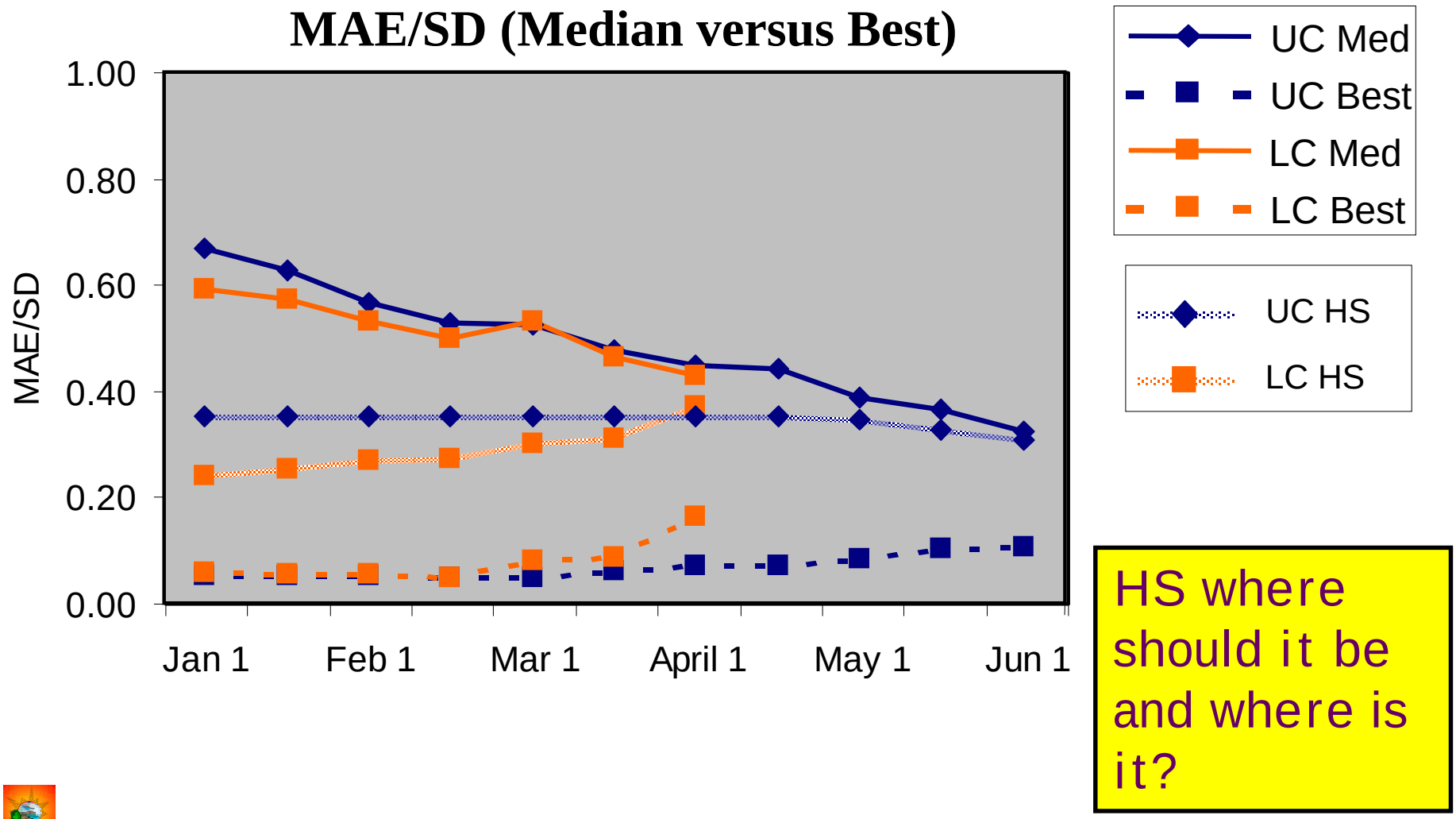






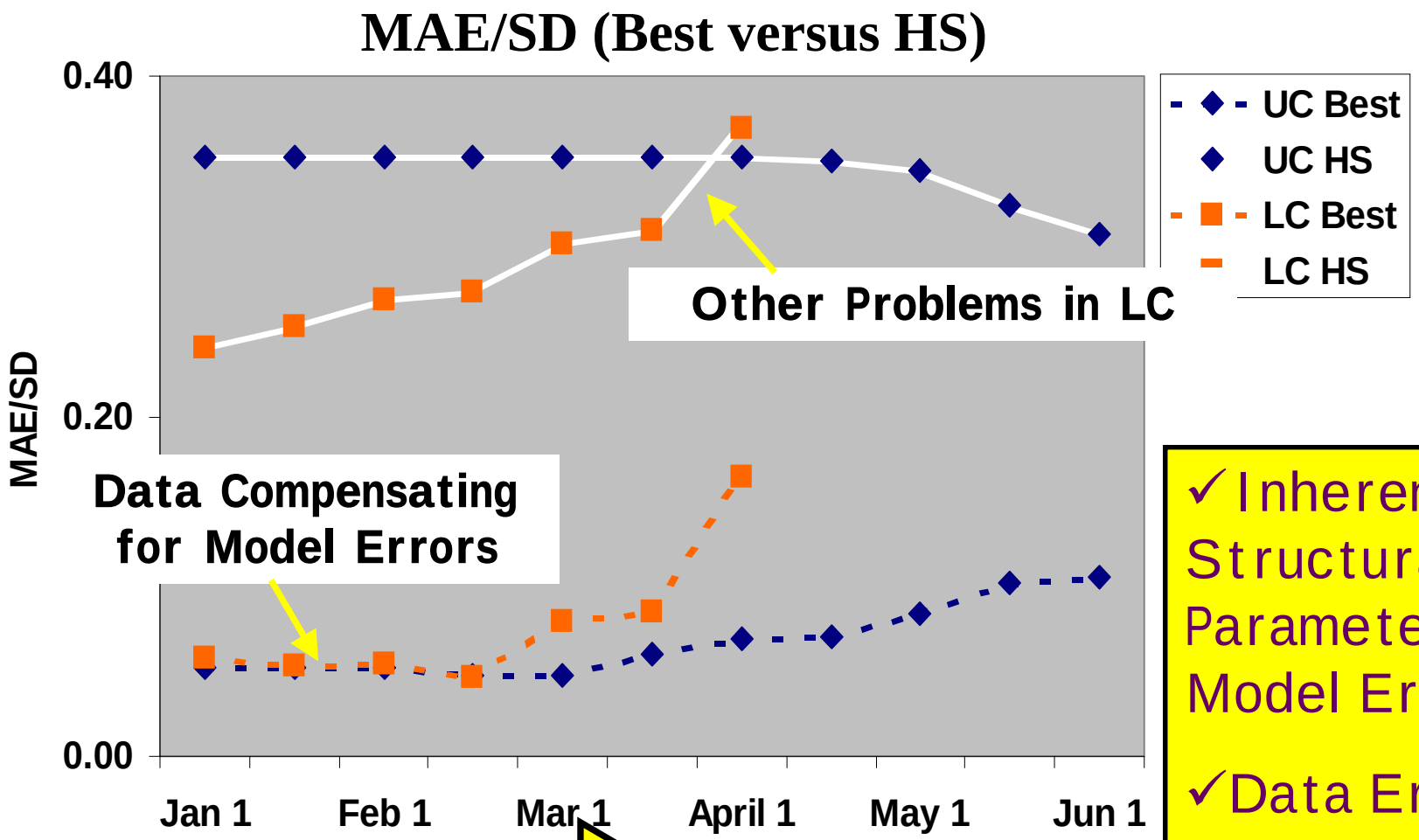
forecast improvement

The ultimate forecast trace…



HS where should it be and where is it?

Procedural Error

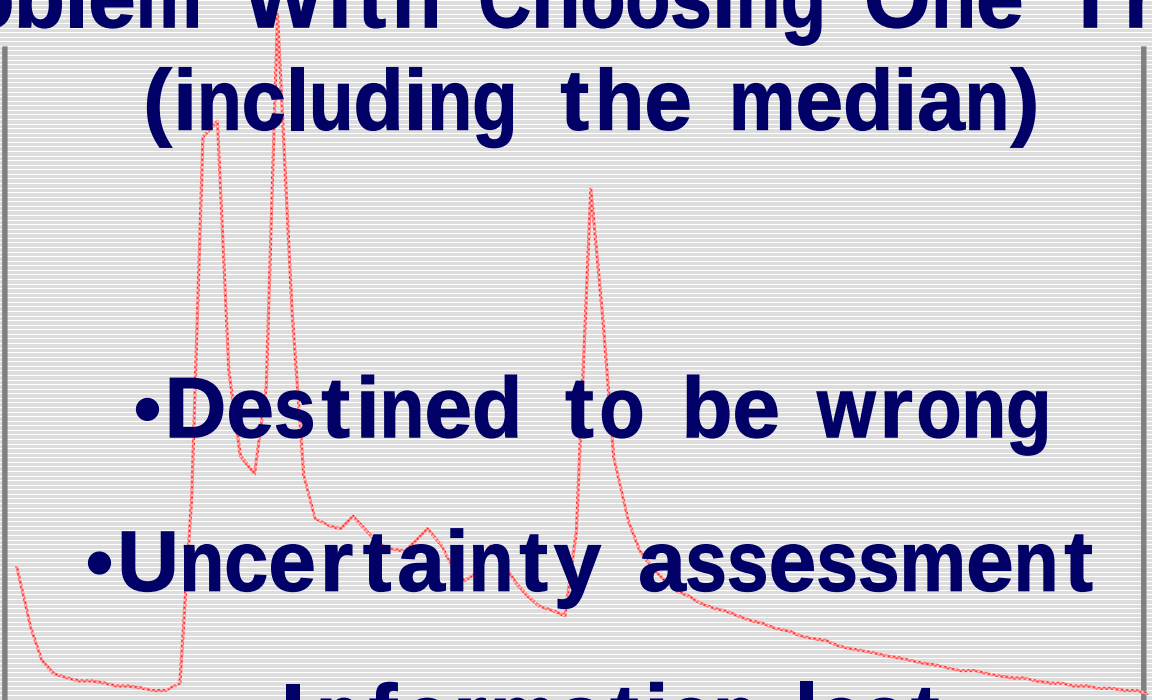


Decreased compensation time = decreased performance

- ✓Inherent Structural & Parameter Model Errors
- ✓Data Errors
- ✓Initial conditions

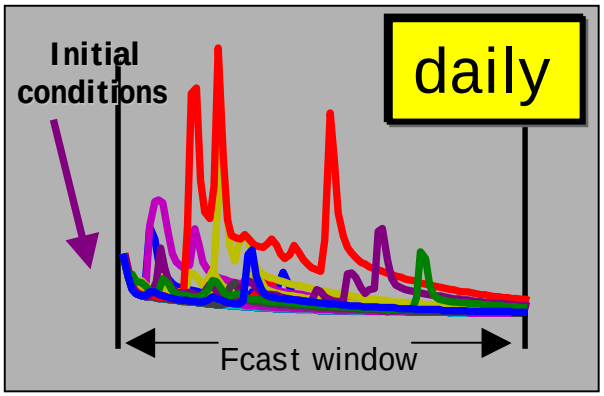
Deterministic Forecasts

Problem With Choosing One Trace (including the median)



- Destined to be wrong
- Uncertainty assessment
- Information lost
- Which do you pick (is median “good enough”?)

ESP Forecasts :Creating Probability



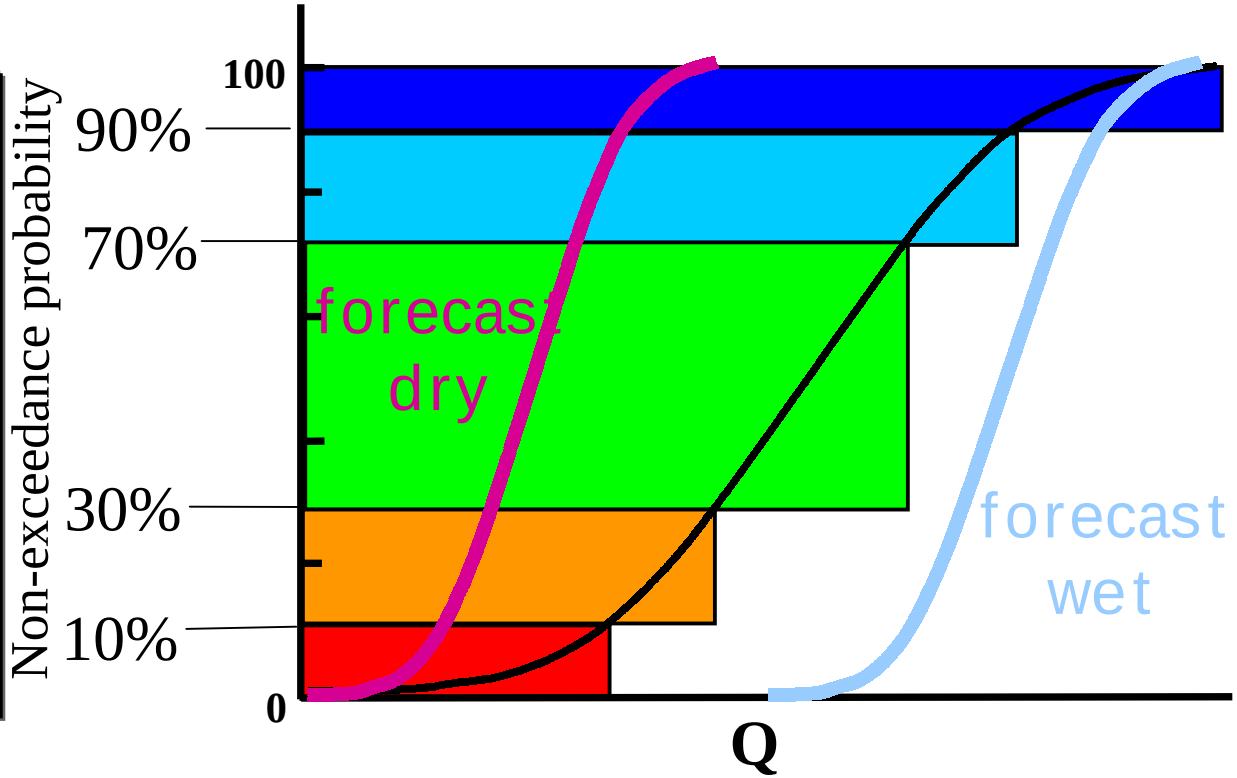
Seasonal totals

Rank and create probability distribution

Historical observation; dist. = climatology

Percentiles based on WS forecasts

Relate ESP dist. to historical



RPS for J = 5 Percentiles

RPS calculation

Probability Interval	Forecast	Observation	if	Cum.Sum F	Cum.Sum O	$(F-O)^2$
0-10%	0.1	0	1	0.1	0	0.01
>10-30%	0.2	0	0	0.3	0	0.09
>30-70%	0.4	1	0	0.7	1	0.09
>70-90%	0.2	0	0	0.9	1	0.01
>90-100%	0.1	0	0	1	1	0
					Sum Equals RPS	0.2

RPS = 1.4

$$\overline{RPS} = \frac{1}{n}\sum_{k=1}^{n} RPS_k$$

For particular forecast window
(k = # Forecasts)

RPS and SS

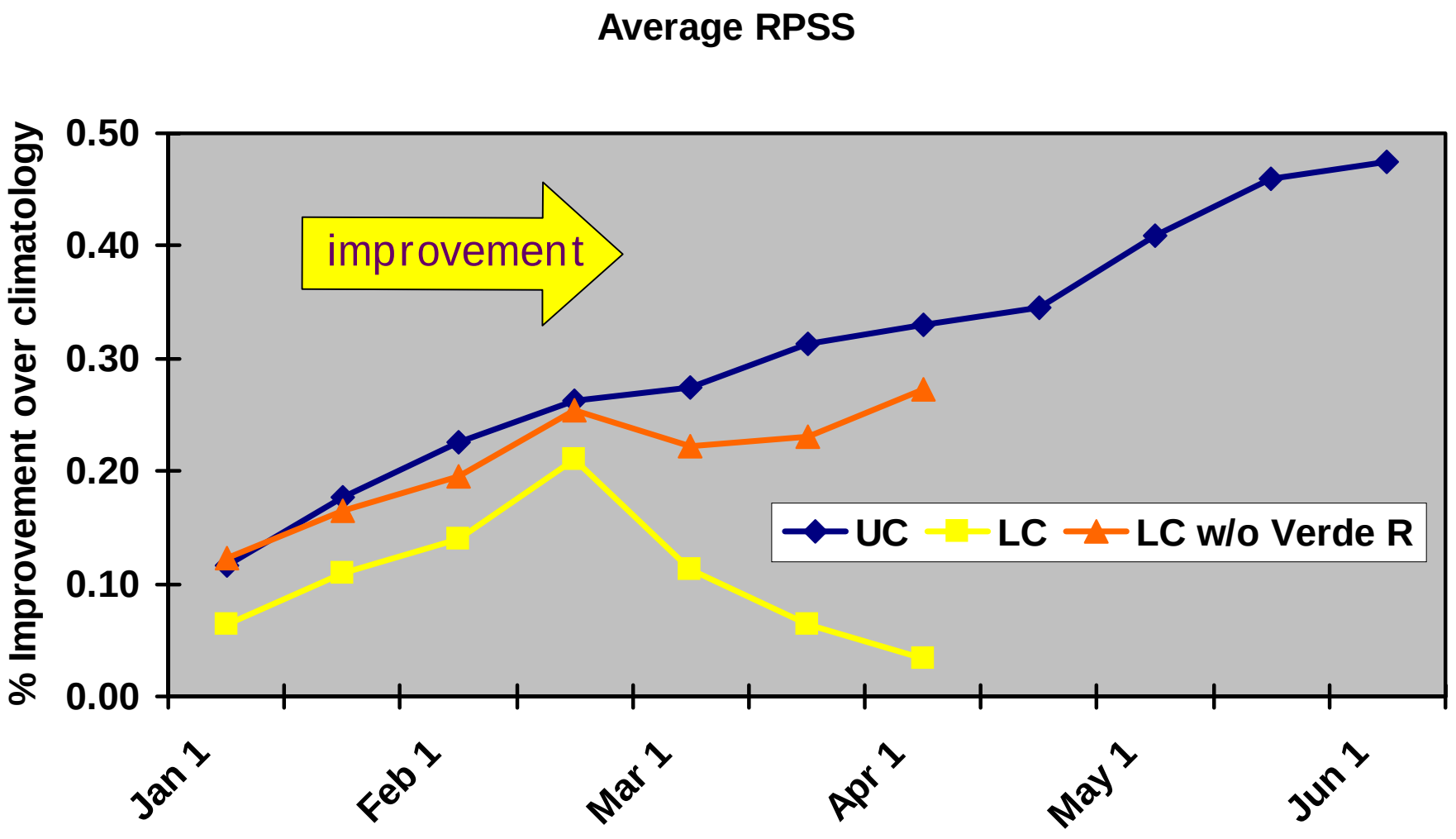
Why the RPS?

- looks at entire distribution (all traces)
- gives credit for probability close to observed
- penalizes for probability far from observed
- overall summary stat for those concerned with all flow levels (water managers)

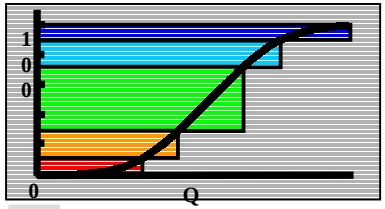
$$RPSS = \frac{RPS_f - RPS_{cl}}{1 - RPS_{cl}} x100\%$$

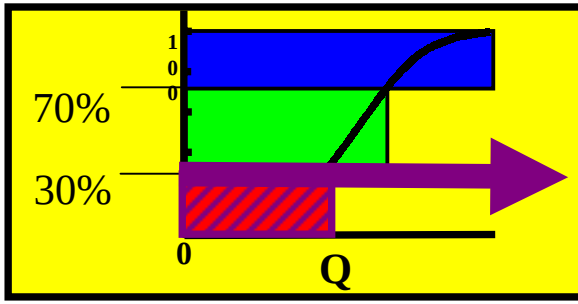
Calculate percent improvement over climatology as next best

Forecast Statistics:RPSS



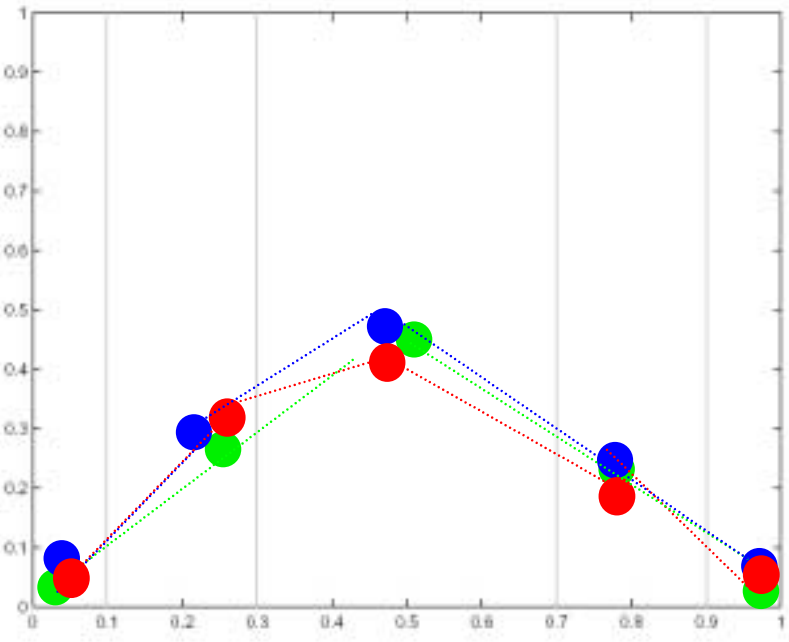
Example Discrimination Diagram





How often was a particular forecast probability given when the observation occurred in the lowest 30%

relative frequency of forecasts

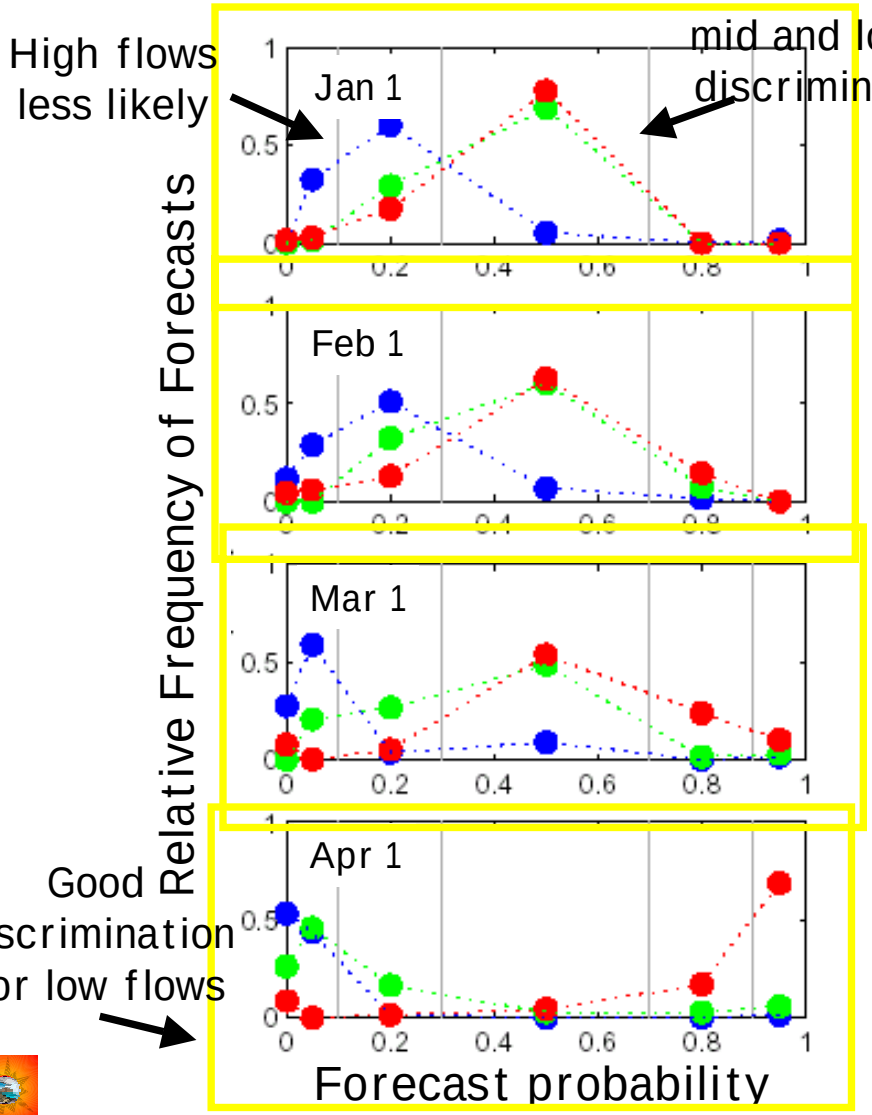


this was the forecast probability given to a particular flow level

When observed occurred in lowest 30%, this was the flow level forecasted

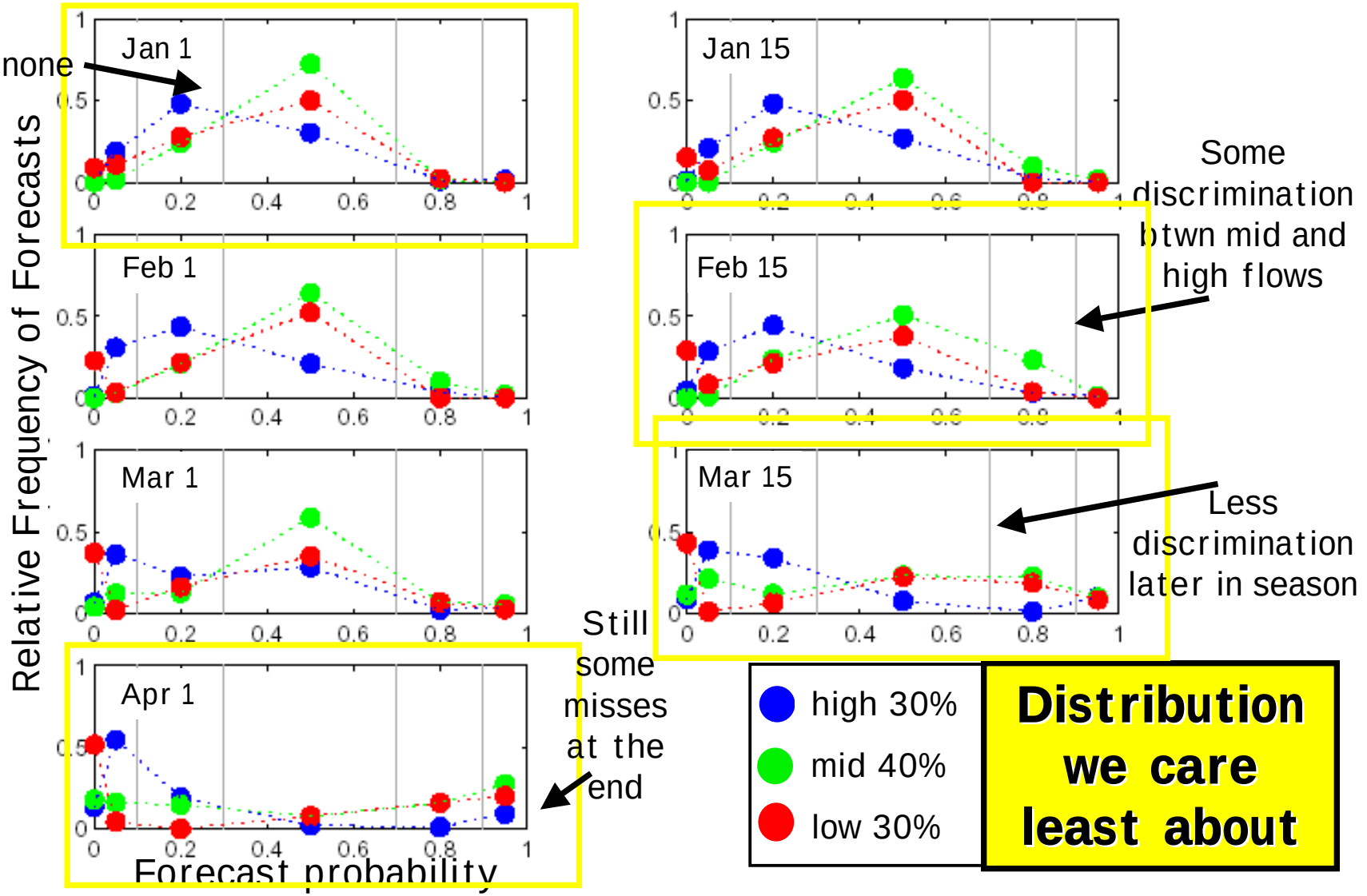
- high 30%
- mid 40%
- low 30%

Discrimination Diagrams for Observation in Lowest 30% for LC

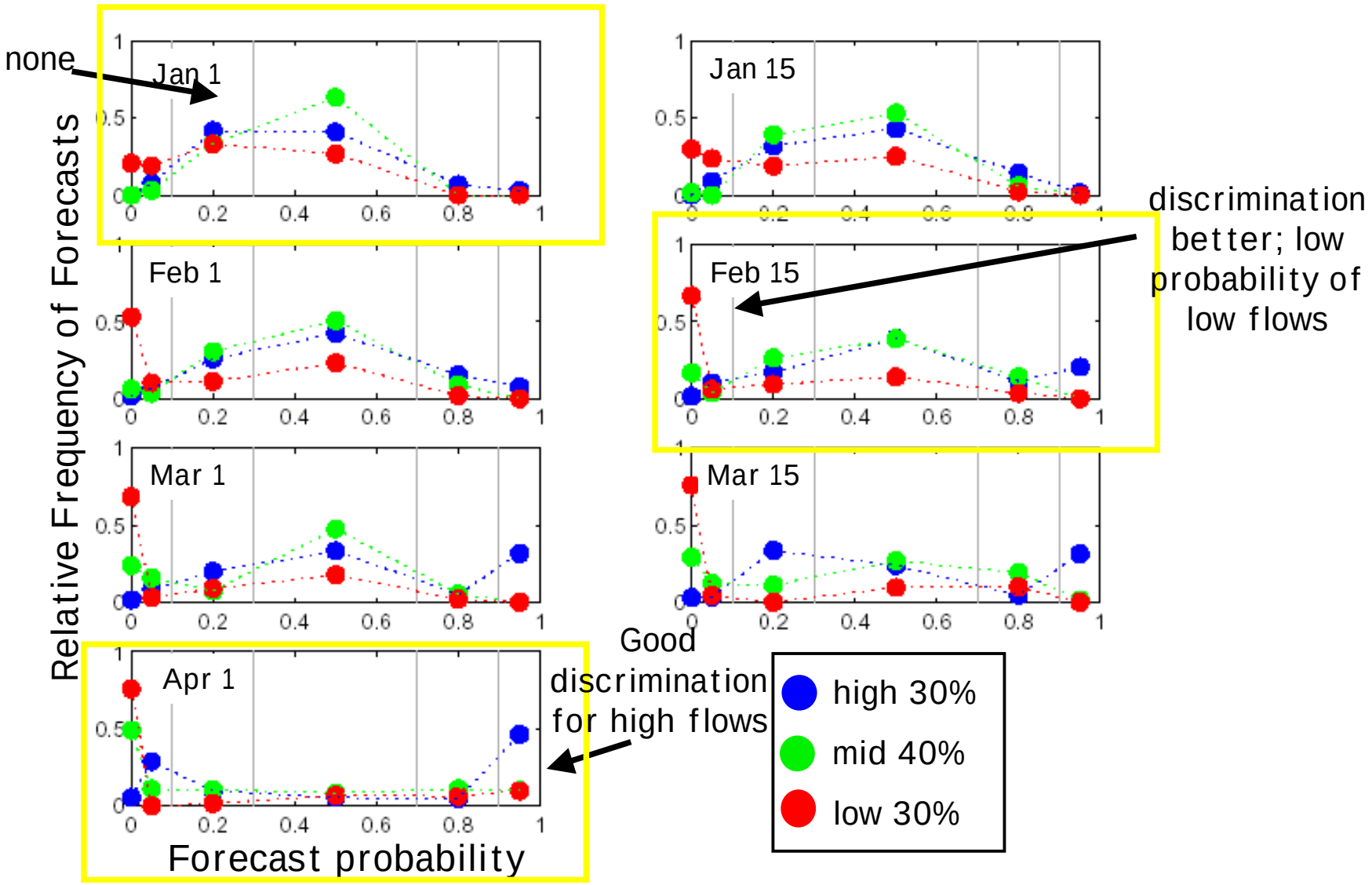


Important info for junior water rights holders; low probability for high flows indicates that high flows will not occur

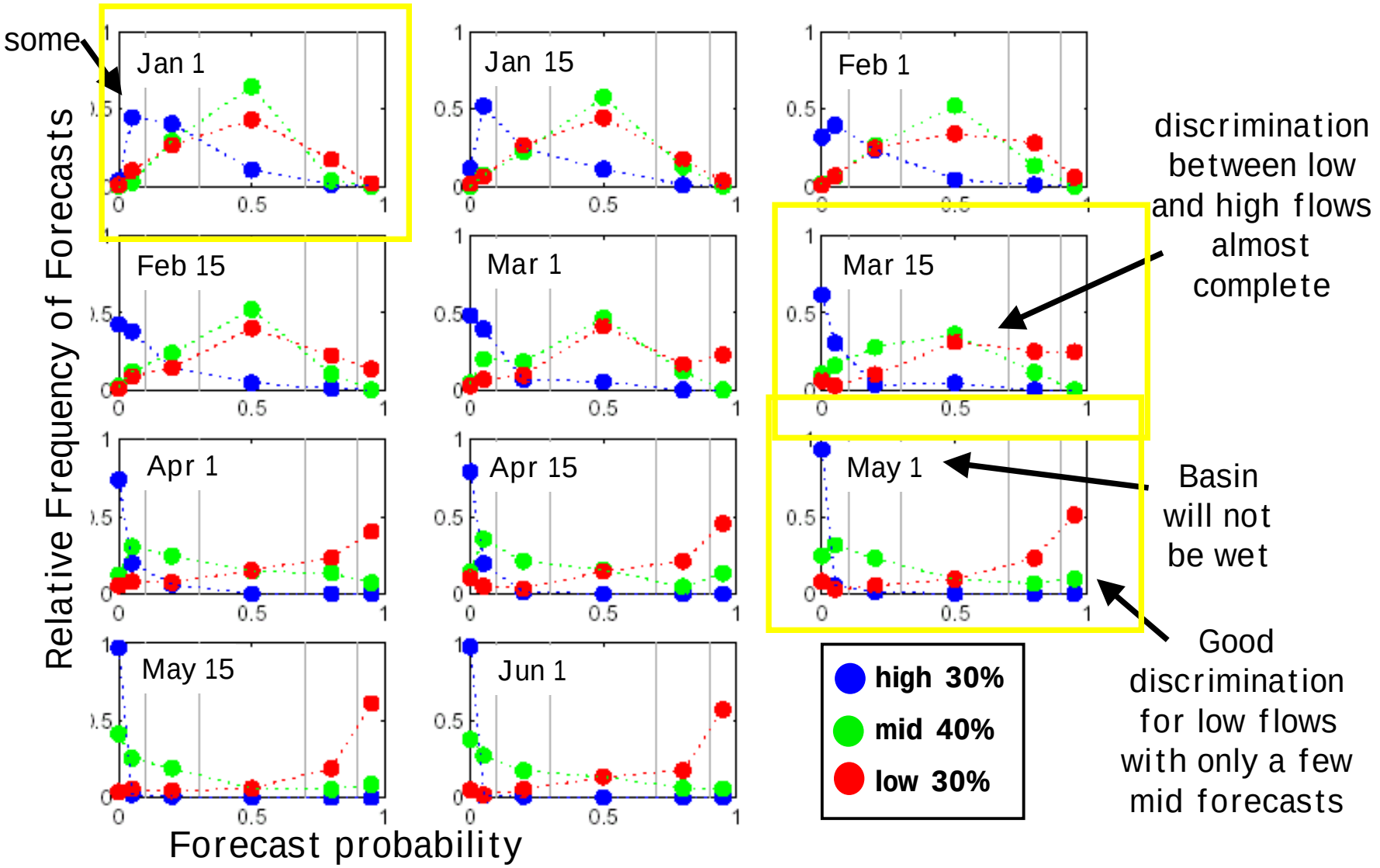
Discrimination Diagrams for Observation in Middle 40% for LC



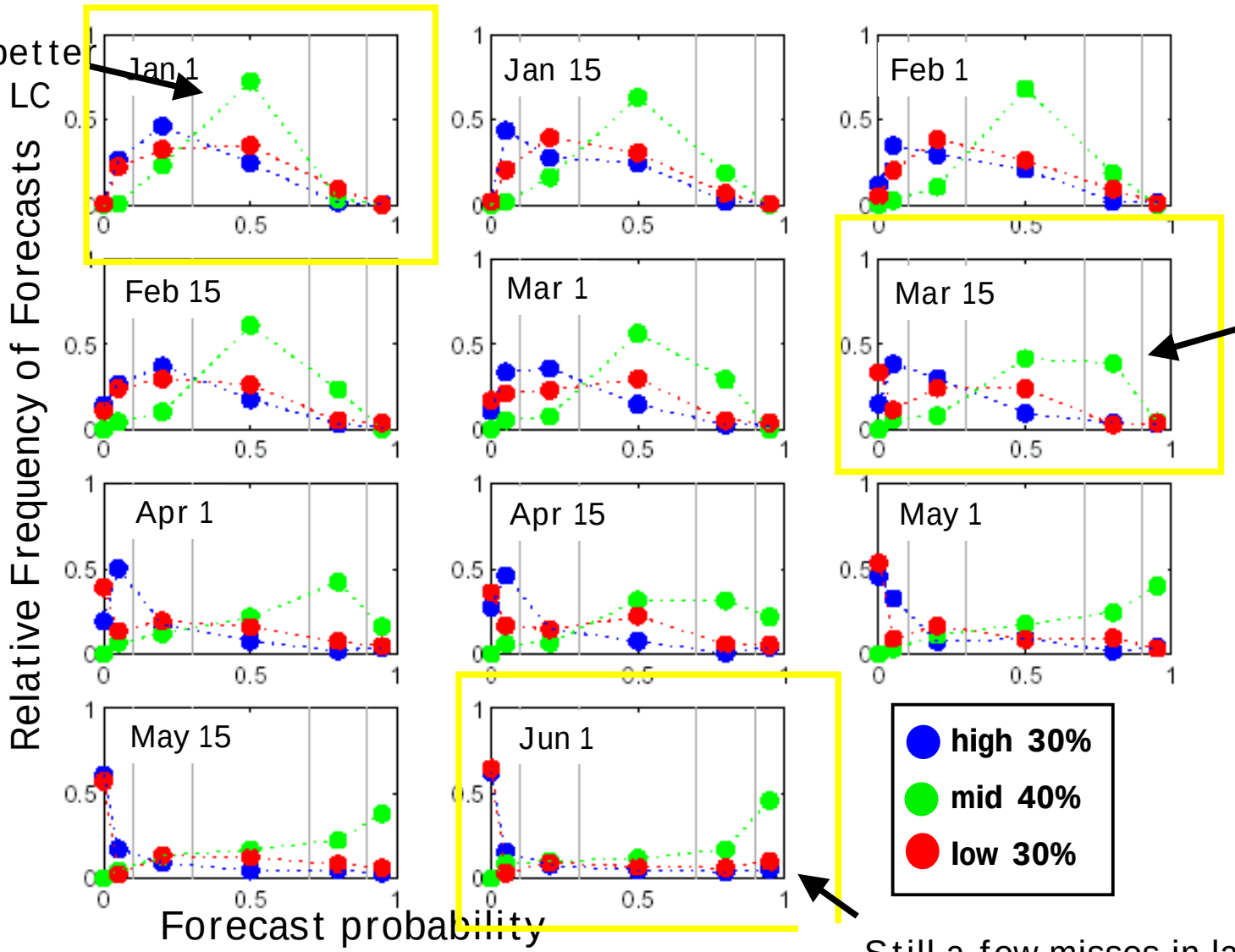
Discrimination Diagrams for Observation in Highest 30% for LC



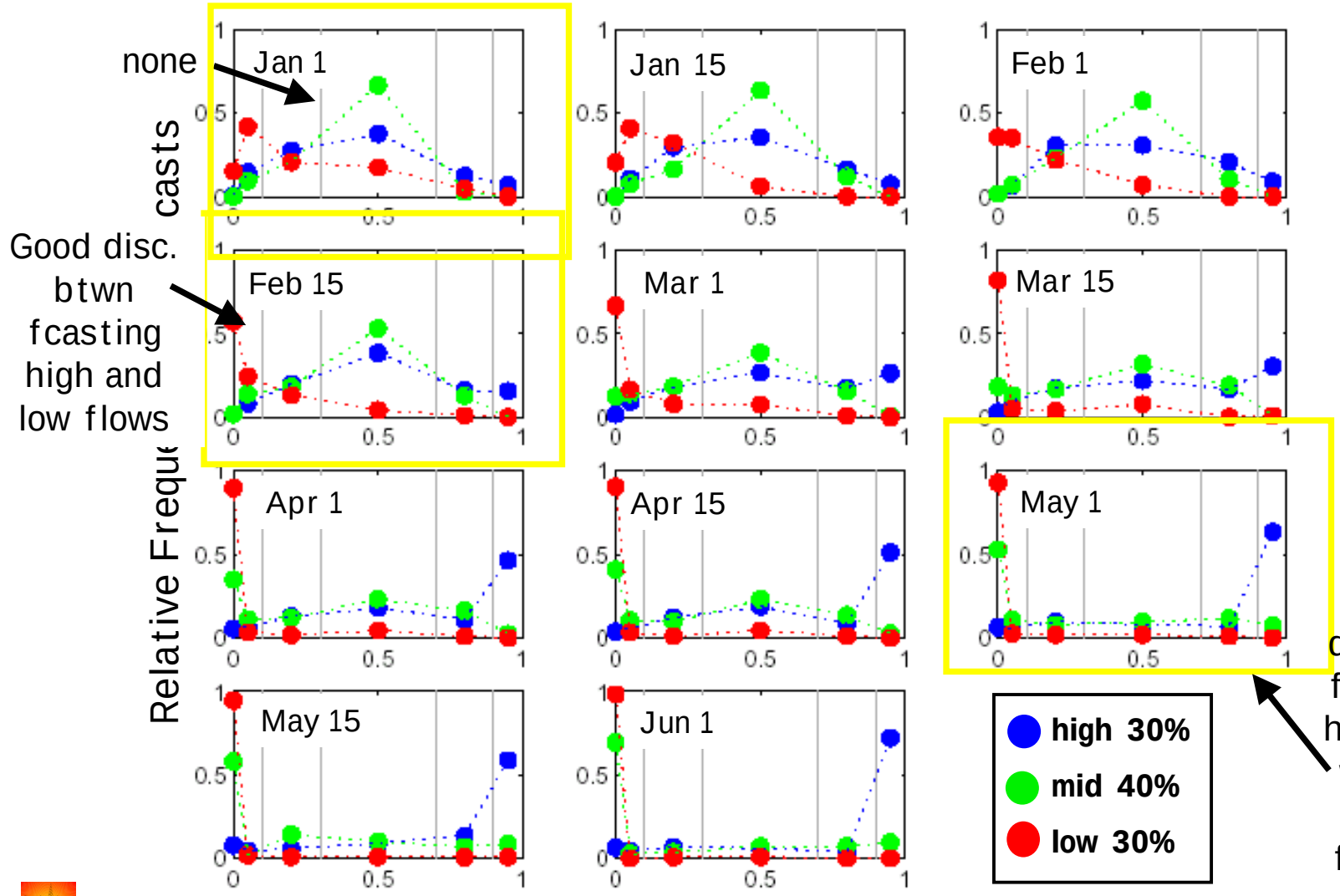
Discrimination Diagrams for Observation in Lowest 30% for UC



Discrimination Diagrams for Observation in Middle 40% for UC

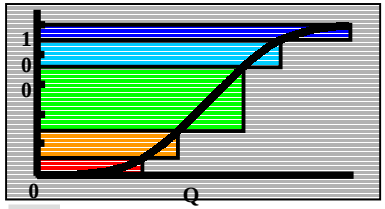


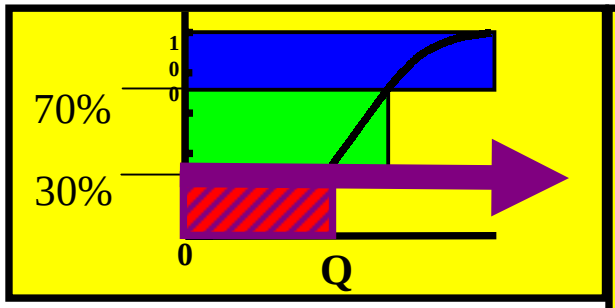
Discrimination Diagrams for Observation in Highest 30% for UC



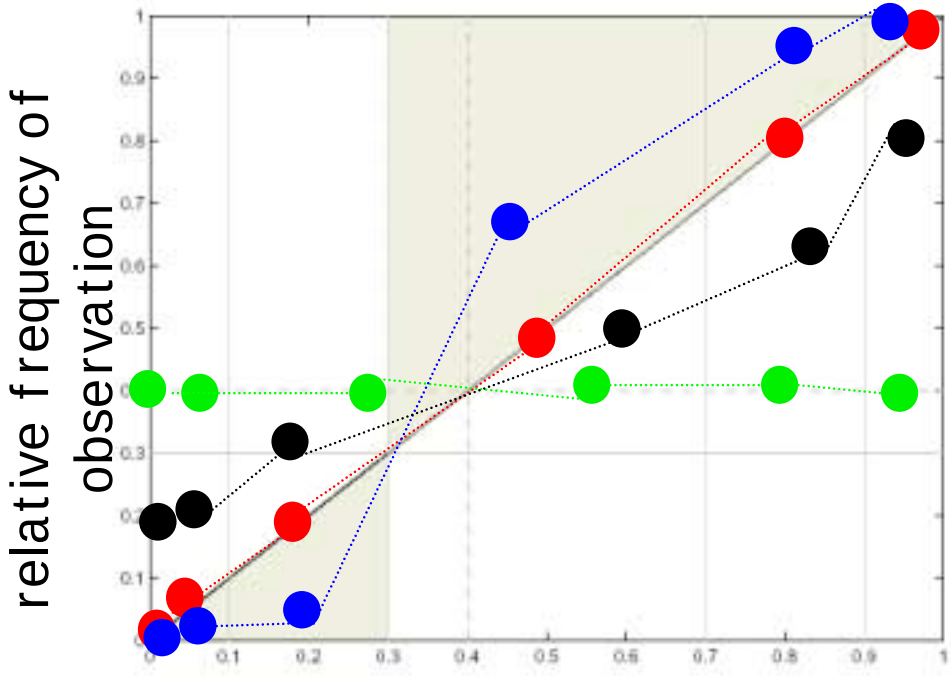
Good discrimination for high flows has developed, w/occasional high prob. forecasts for mid flows

Example Reliability Diagram





Does an observation occur at a frequency equal to the forecast probability for a particular flow level.
ex. When low flows forecasted with 80% prob., they occur 80% of the time.

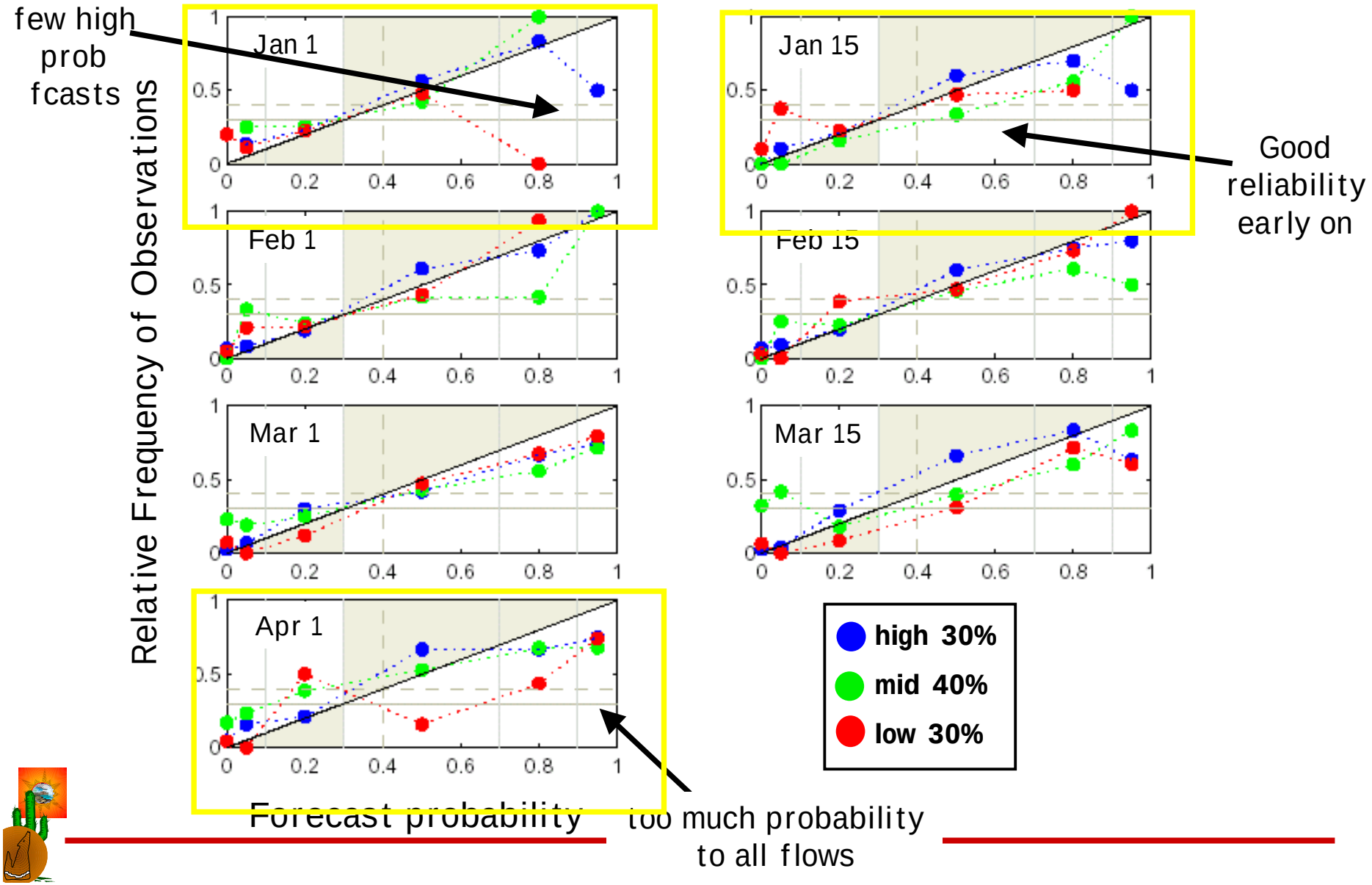


this was the forecast probability given to a particular flow level

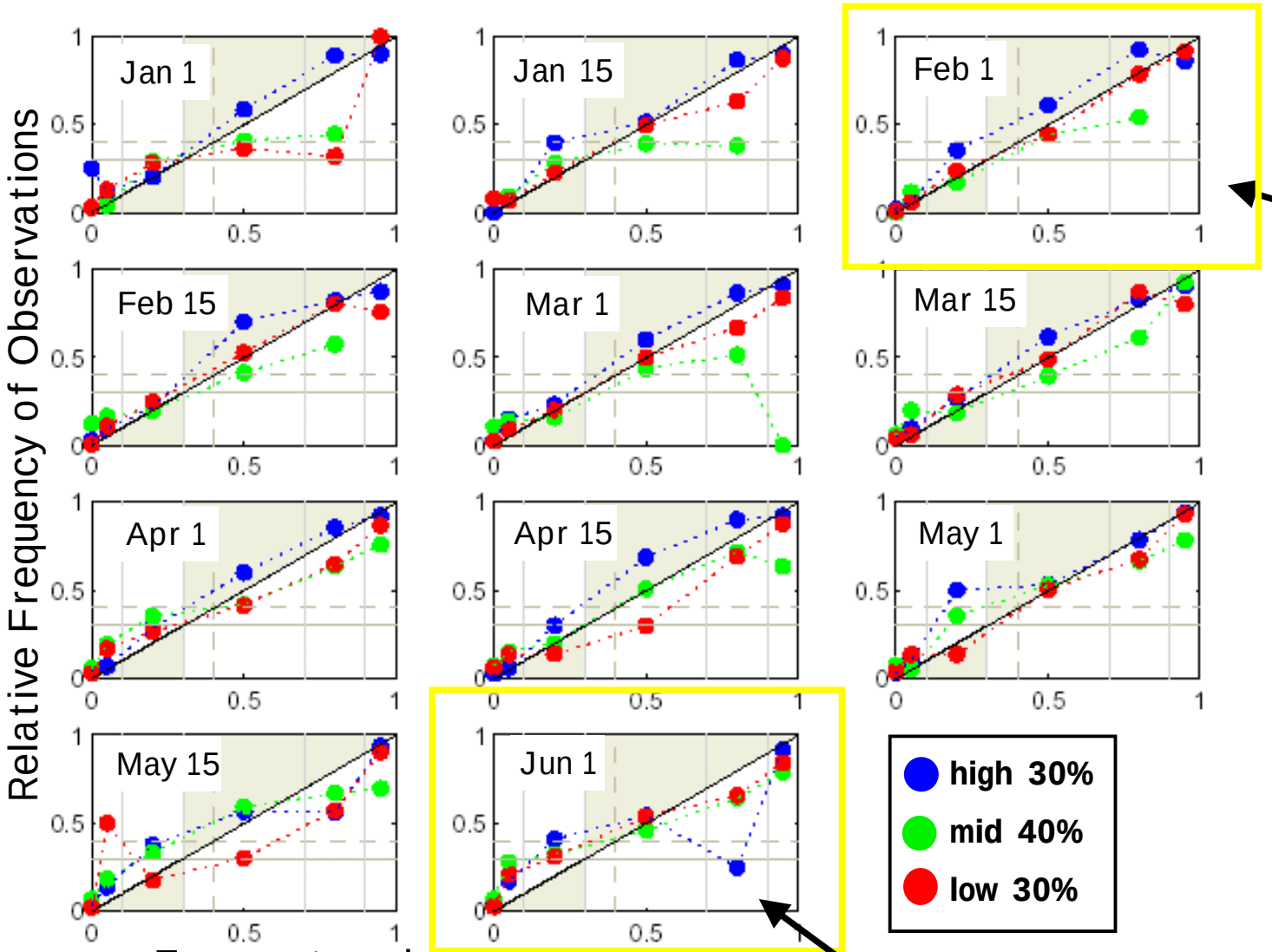
flow level forecasted

- high 30%
- mid 40%
- low 30%

Reliability Diagrams for LC



Reliability Diagrams for UC



Good reliability early on, better than LC

Generally due to low sample pop.

Summary & Conclusions

•Median
- transforms the probabilistic forecast into a deterministic forecast
- results in a loss of valuable information
- forecast is destined to be wrong

•ESP
- provides probabilistic forecasts
- more explicit statement of total uncertainty (model + future)
- ESP (probability) forecasts are “never wrong”
- show improvement over climatology forecasts for seasonal water supply prediction
- accurately predicts the correct flow quantile in which the observed will occur for some basins with lead times of 2-3 months
- provides water supply information from start of the forecast season; ex. low probability for high flows often associated with low observed streamflow volumes

Still needed

In addition, operational forecast can be expected to perform better than hindcast forecasts:
-initial conditions updates
-incorporation of climate information

Next Steps
•compare ESP forecasts to historical regression forecasts by transforming deterministic to probabilistic
•explore conditional probabilities based on climate information
•assess usefulness of ESP forecast information for forecast users
•snow conditions

Acknowledgements

- Office of Hydrology
  - Edwin Welles
  - Hank Herr
- CBRFC
  - Cara McCarthy
  - Dave Brandon

ISPE

NOAA CLIMAS

Office of Hydrology

NWS

HWR