Thorpex-Hydro meeting, NCEP Camp Springs: 12/02/10

Verification of precipitation forecasts from the SREF.

James Brown
james.d.brown@noaa.gov

Purpose and background

1. **Assess hydrologic value of SREF**

- In principle, it should be less underspread and more skilful than GEFS (caveat: 32x32 km grid not ideal for convection).

2. **Improve short-term flow ensembles**

- Merge SREF with GEFS to provide seamless forcing (temperature and precipitation) out to 14 days (see MMEFS project).
- Relies on bias-correction (not covered here).

Hydrologic value of SREF

Precip. skill and bias as a function of:

1. Location (RFC) and forecast lead time.
2. Amount.
3. Level of temporal aggregation.
4. Basin size.
5. Season.
6. Relative to GEFS.

- Different aspects of quality/metrics for each.

Caveats of the analysis

Incomplete data and results

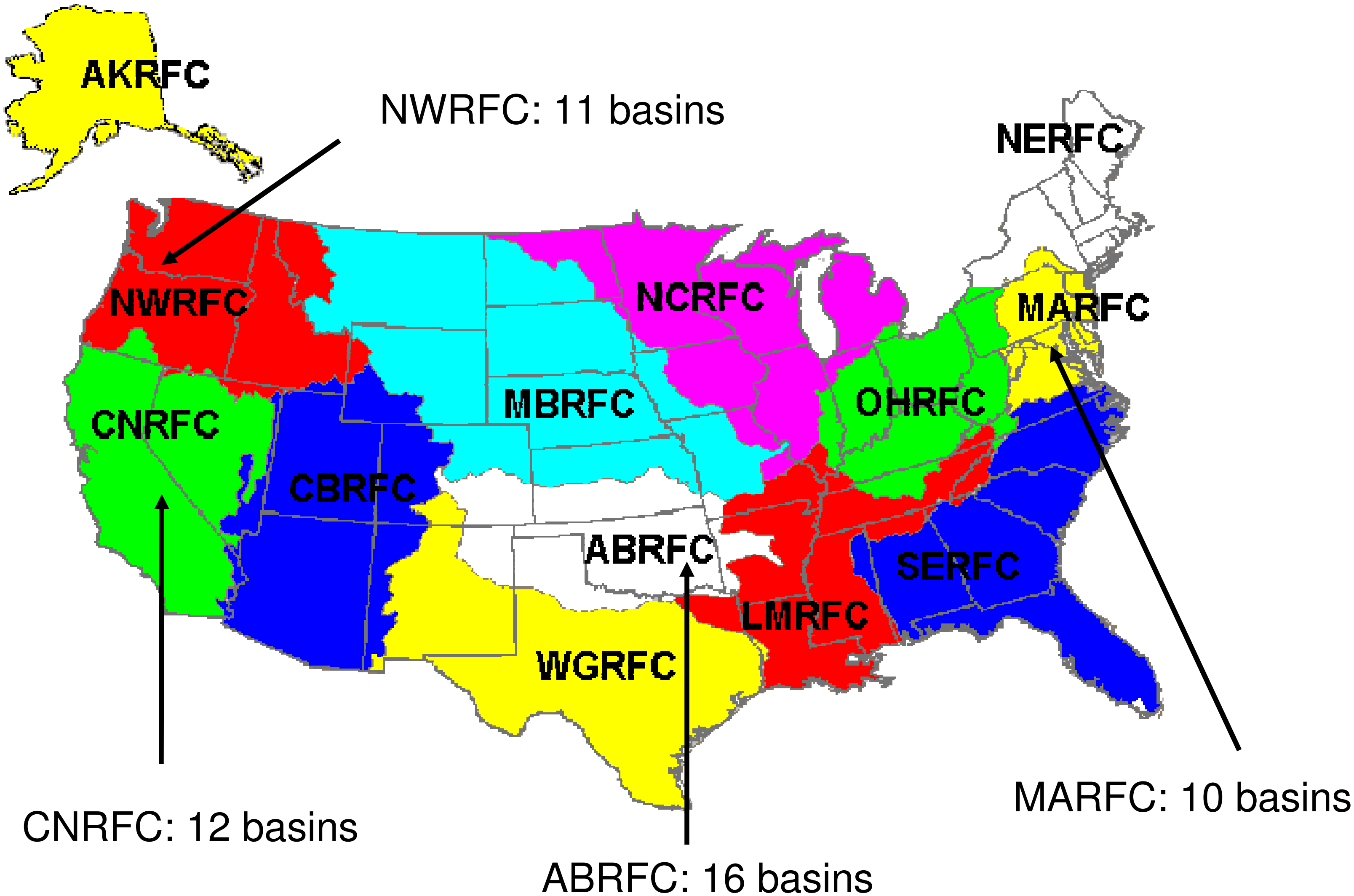
- Limited period: April 2006 – February 2010.
- Partial archive from SUNY: archive has since been updated from HPSS with all 9Z and 15Z forecasts and missing 3Z and 21Z forecasts.
- Results mainly for MARFC (three other RFCs running) and for precipitation only.
- Mismatch of results at present (work in progress)! Most w/o sampling uncertainty.

Data preparation

Derivation of basin MAP

- SREF at 32x32 km and NCEP gridded QPE (radar/gage mapped to CPC) at 0.125 degree.
- Basin MAP: average over grid cells.
- Looking at 6-, 12- 24- hourly accumulations from 6-87 hrs (same clock as observations).
- Spatial pooling across 10-15 contiguous basins to improve sample for larger events.
- Basins from four climate regions….

RFCs/basin groups



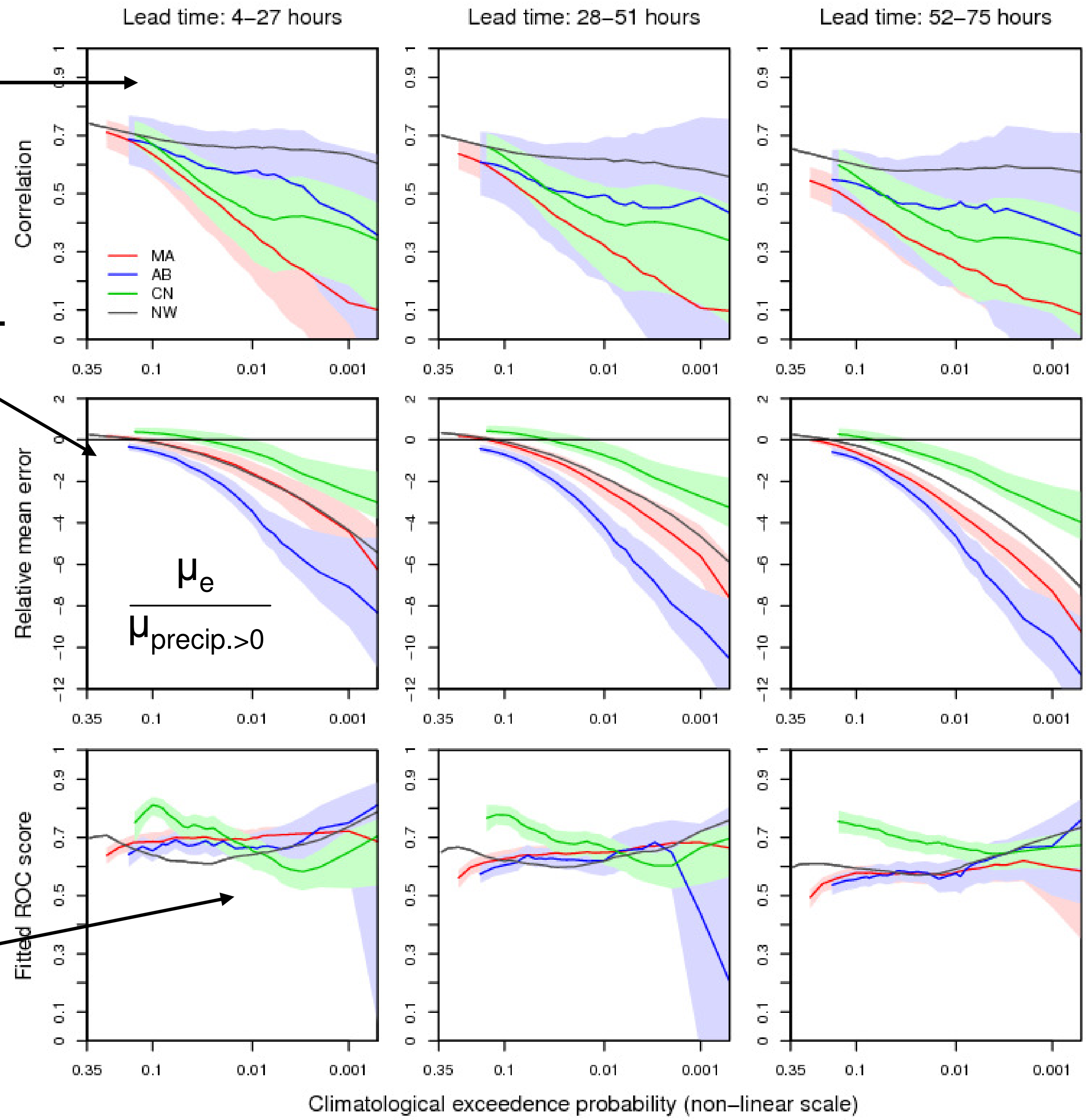
Example results I: dependence on magnitude of precipitation (across several RFCs)

NW preserves much better corr. for large events than MA but similar conditional biases in mean: NW = better post-processing. AB = large bias.

However, post-processors estimate F(obs|fcst) not F(fcst|obs).

Skill declines fairly slowly over first three lead days.

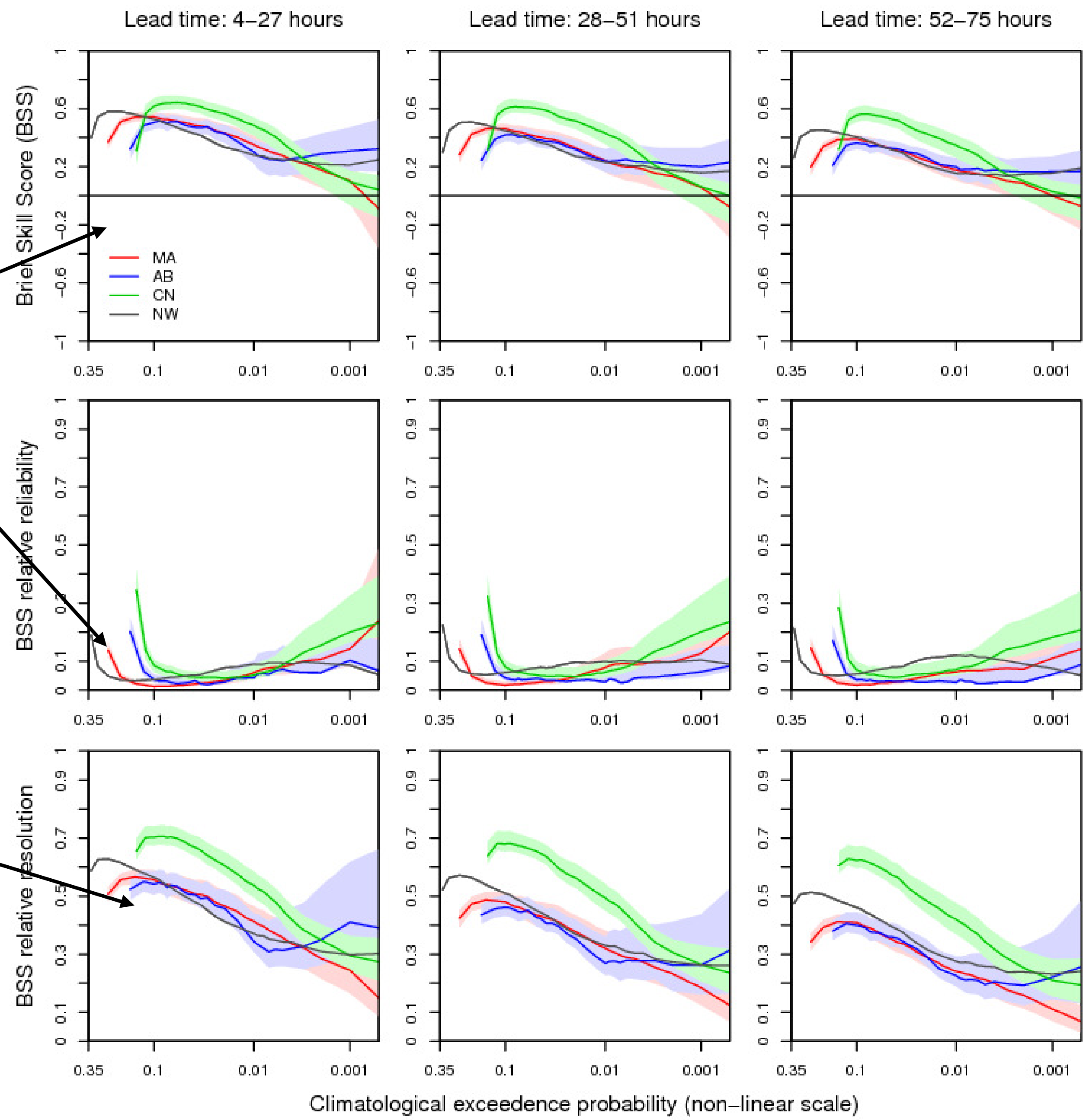
Discriminatory skill changes slowly with size of event.



BSS (CR factors)

"U-shaped" BSS due to lack of reliability for both light and heavy precipitation.

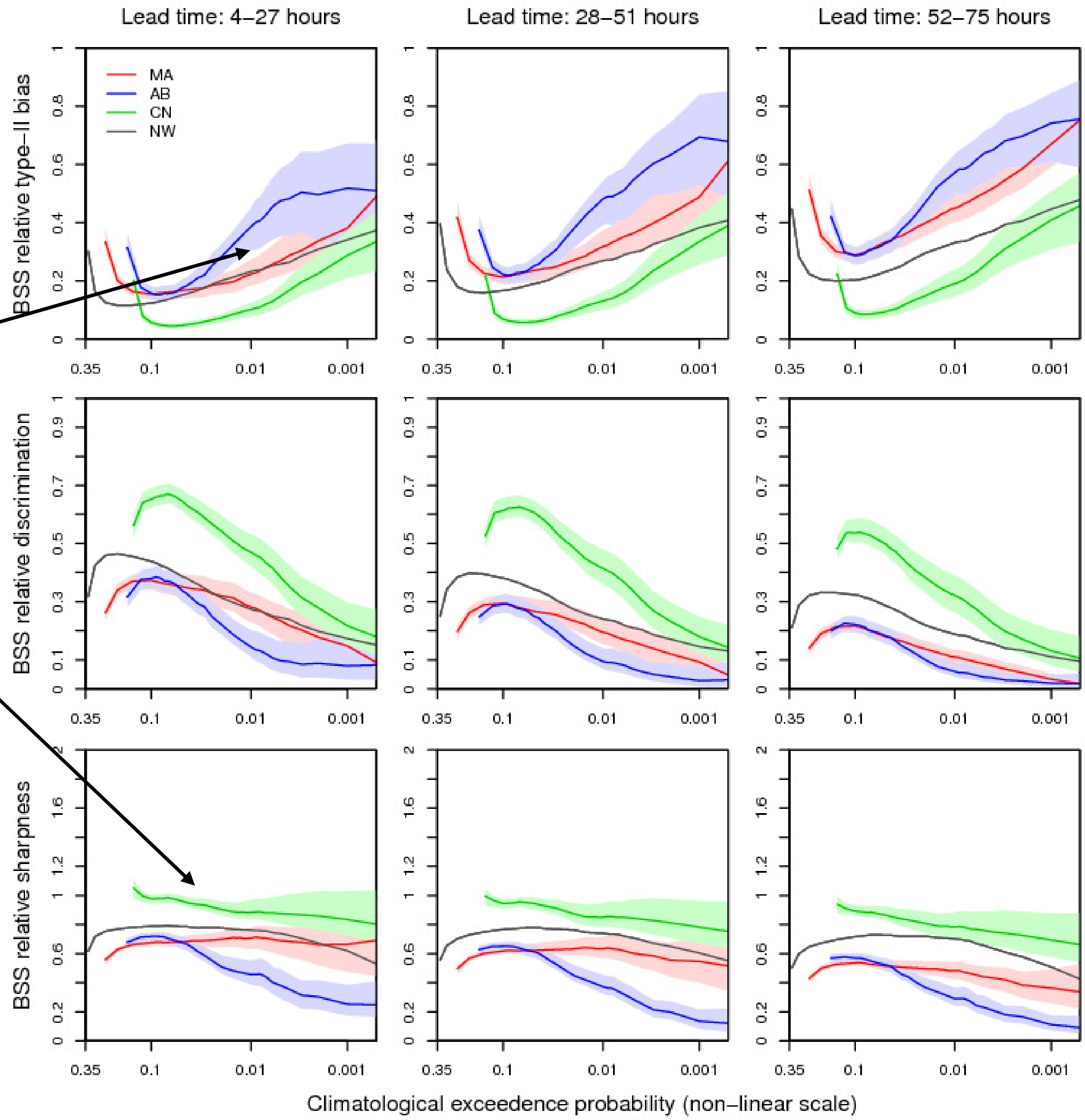
Better overall BSS for CNRFC basins, but lower reliability (Type-I bias). Due to excellent resolution.



BSS (LBR factors)

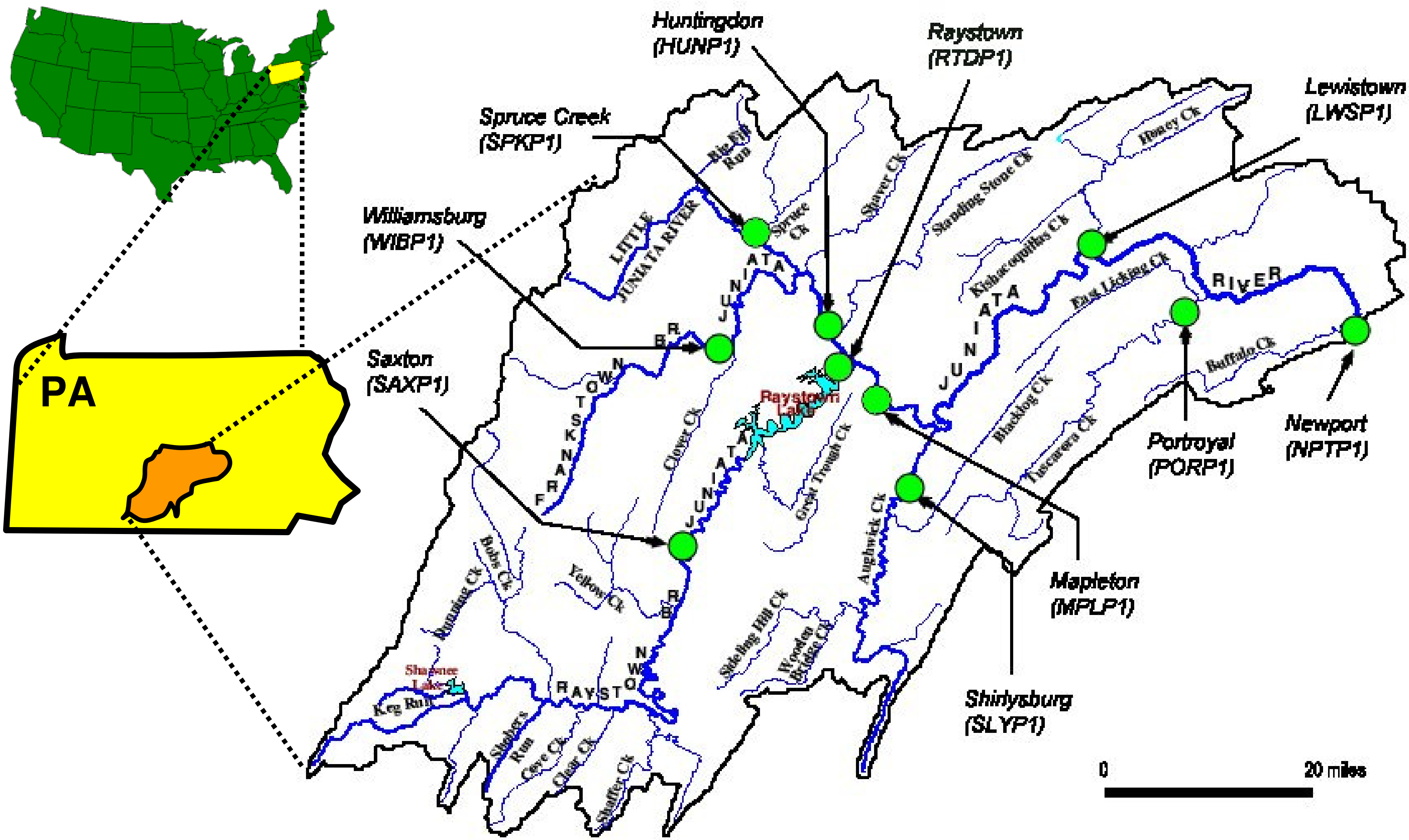
Type-II bias, i.e. F(fcst|obs) increases with size of event due to conditional bias in ensemble mean | obs. Again, worst for AB.

CNRFC has much sharper forecasts while being less reliable = under-spread. However, They have good discrimination. Can correct bias with post-processing.



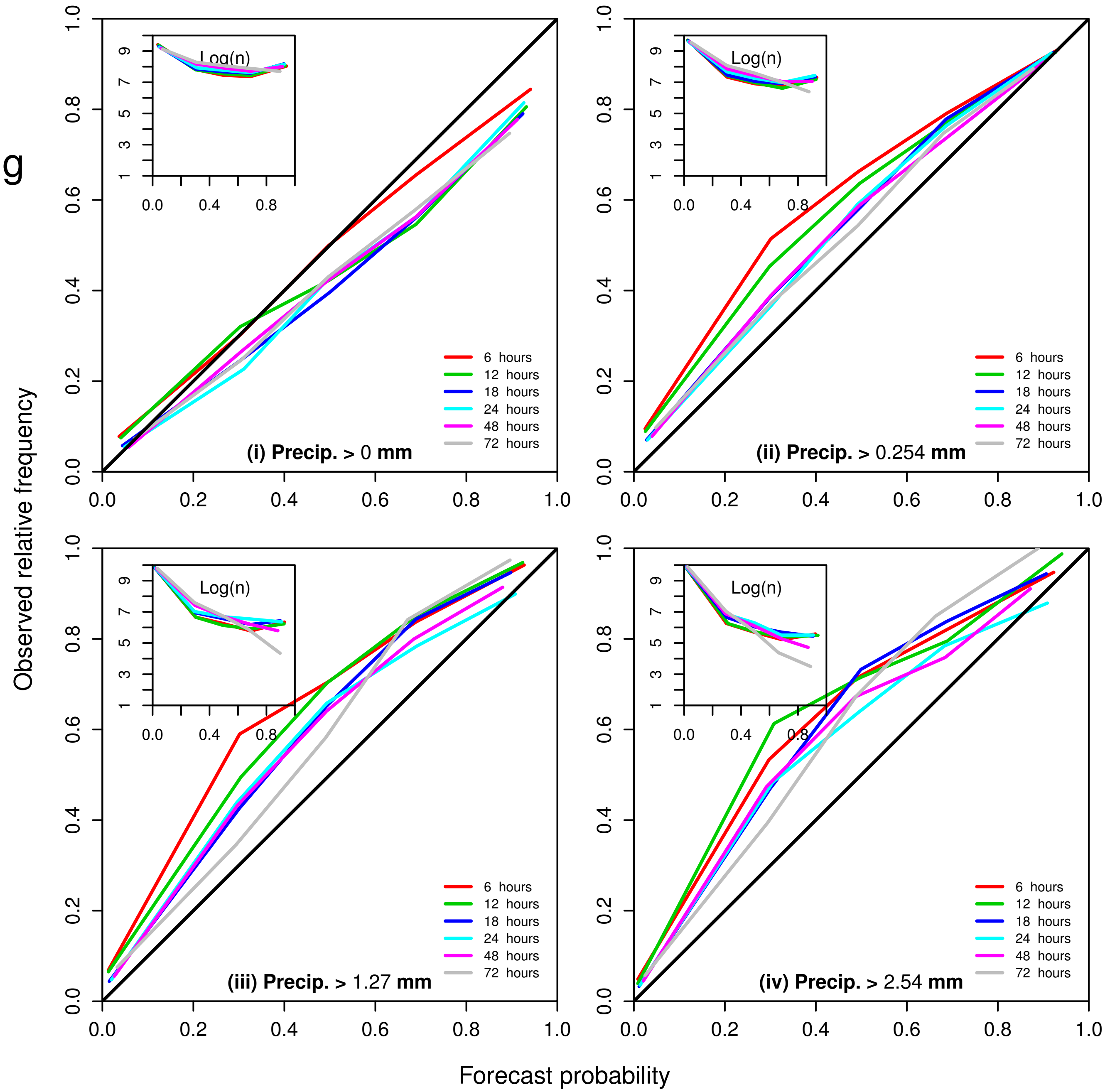
Example results II: Biases (Type-I and II) [based on SUNY archive]

Juniata River, PA



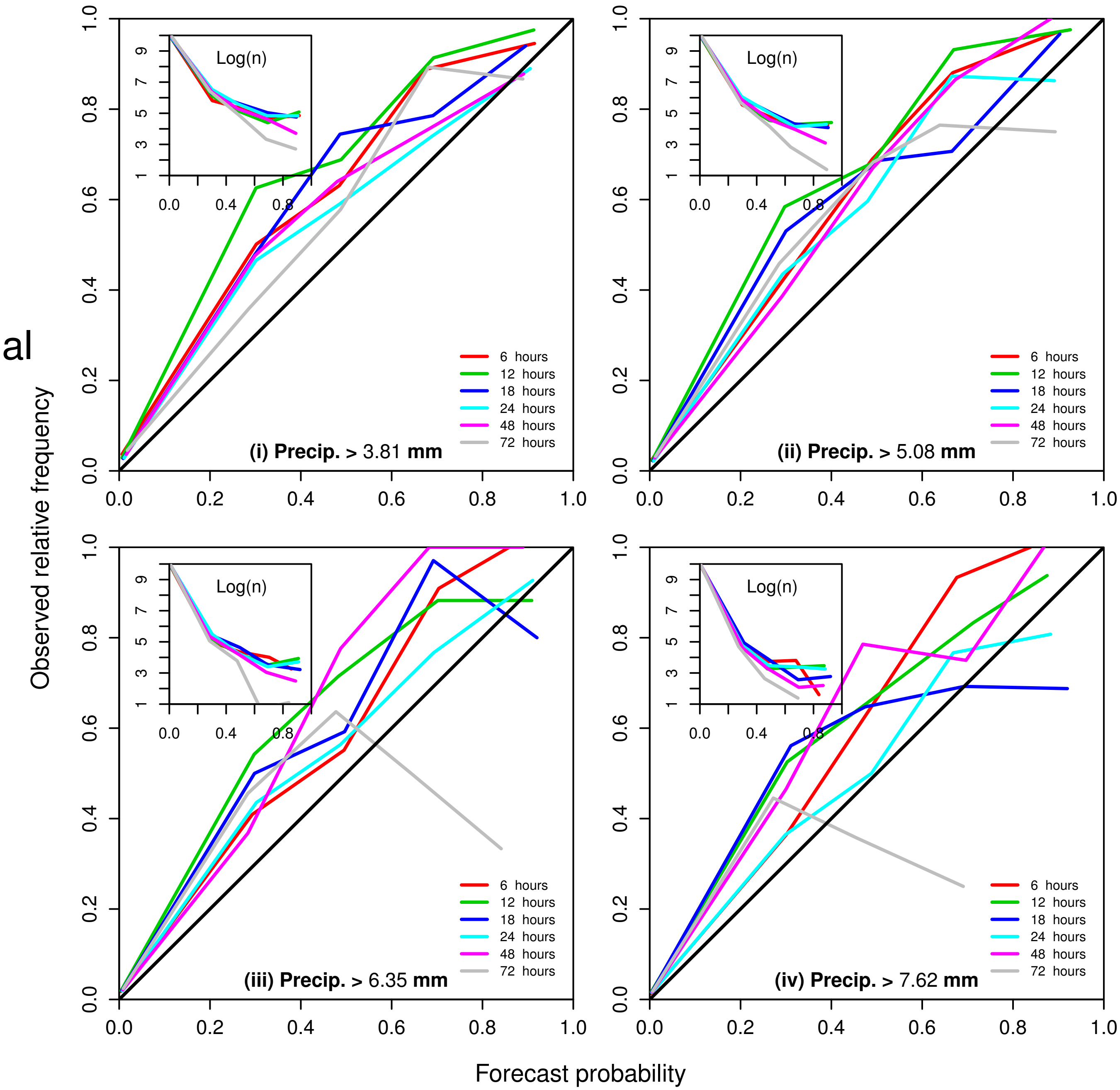
Slight over-forecasting of PoP and under-forecasting of larger events (see next slide too)

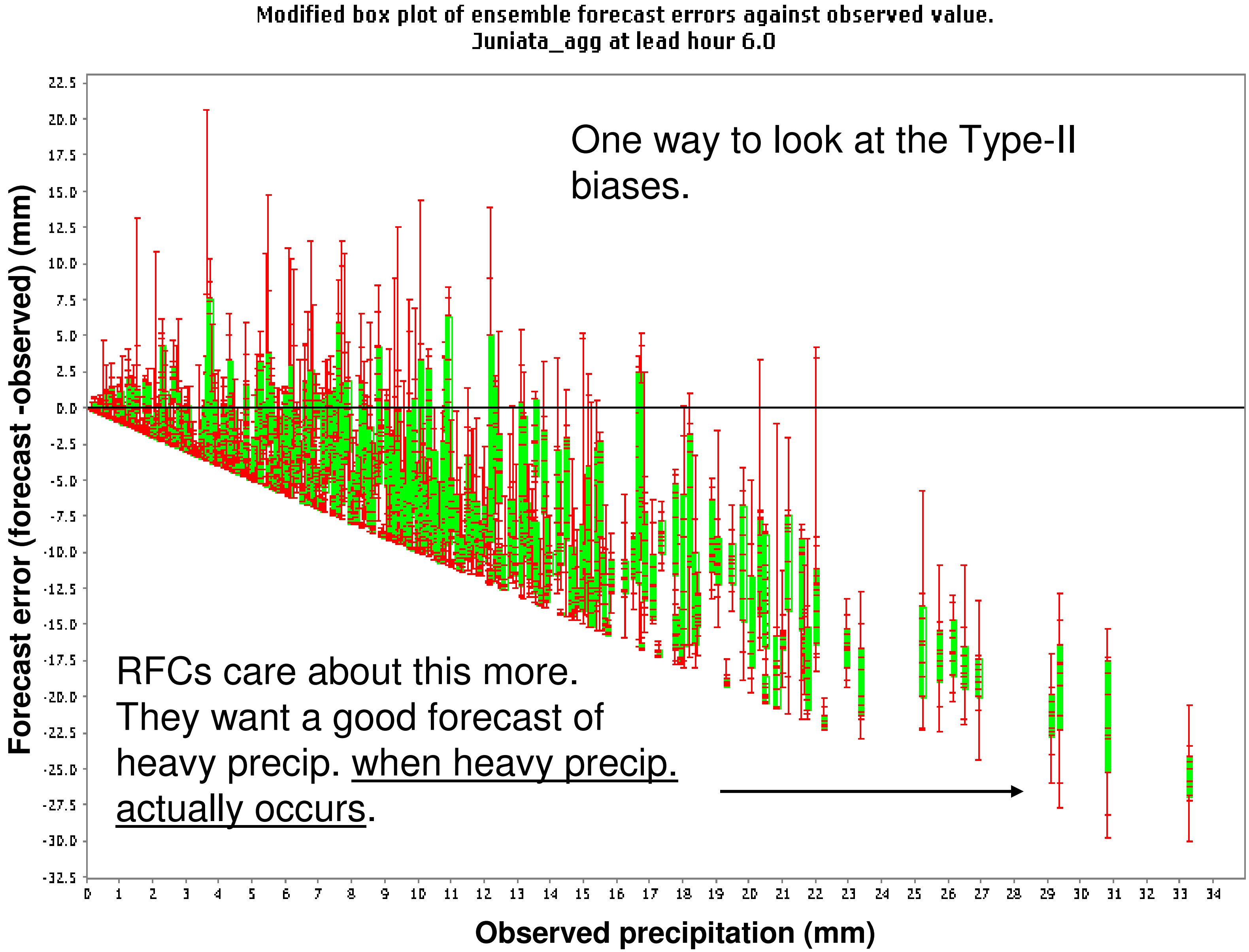
However, bias is generally small <u>in terms of</u> F(obs|fcst)=F(fcst), i.e. Type-I bias is small.



Again reliability is good for larger events.

The main problem will be the conditional bias in F(fcst|obs), i.e. the Type-II bias…..

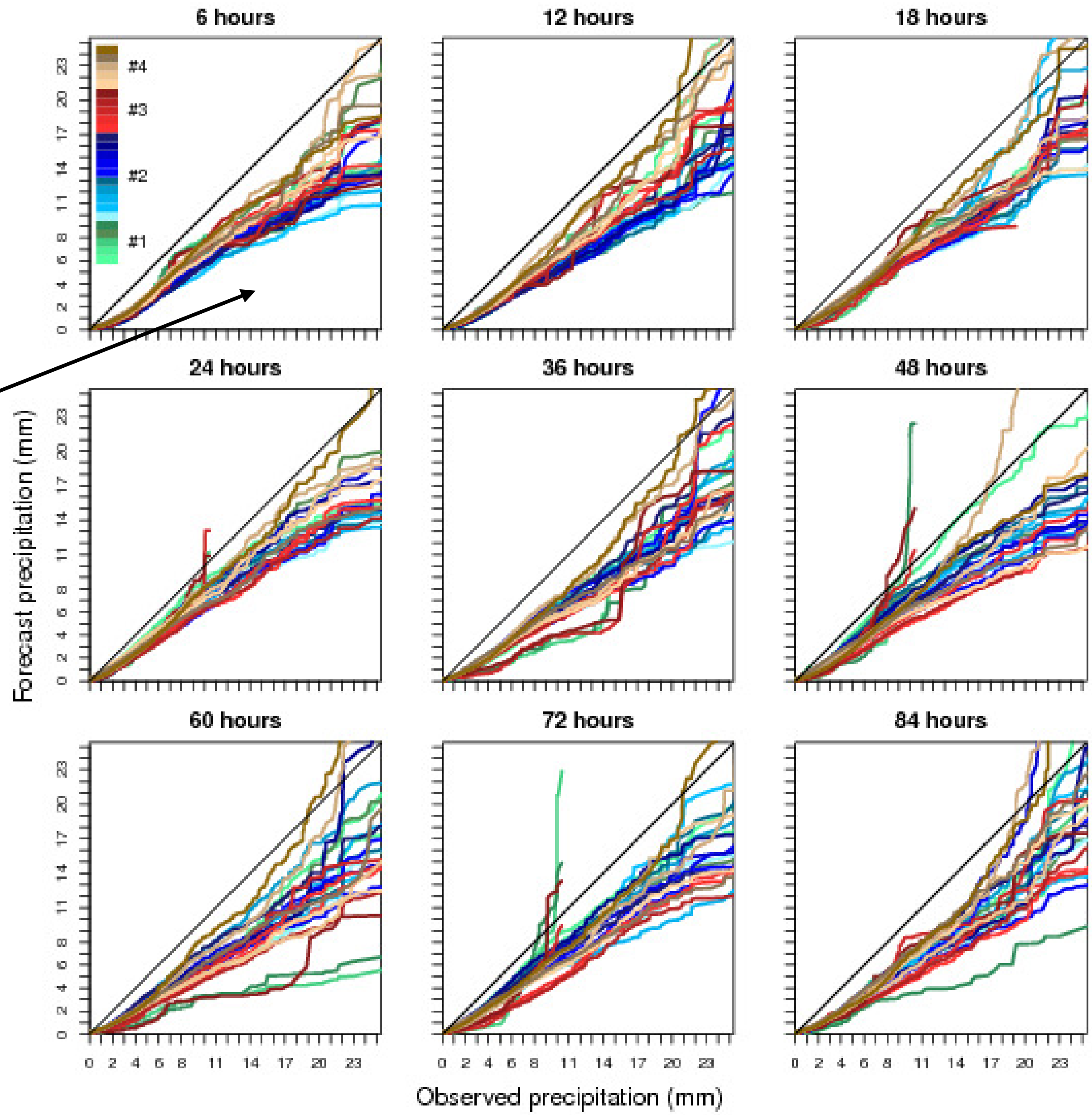



Modified box plot of ensemble forecast errors against observed value.
Juniata_agg at lead hour 6.0
Forecast error (forecast -observed) (mm)
Observed precipitation (mm)
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-17.5
-20.0
-22.5
-25.0
-27.5
-30.0
-32.5
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
One way to look at the Type-II biases.
RFCs care about this more. They want a good forecast of heavy precip. when heavy precip. actually occurs.

Member-specific Q-Q plots. Grouped by physics model and lead time.

Strong low-bias, that varies by physics model.

Even unconditional bias correction (e.g. Q-Q mapping) would improve verification statistics, but may need a model-specific approach.

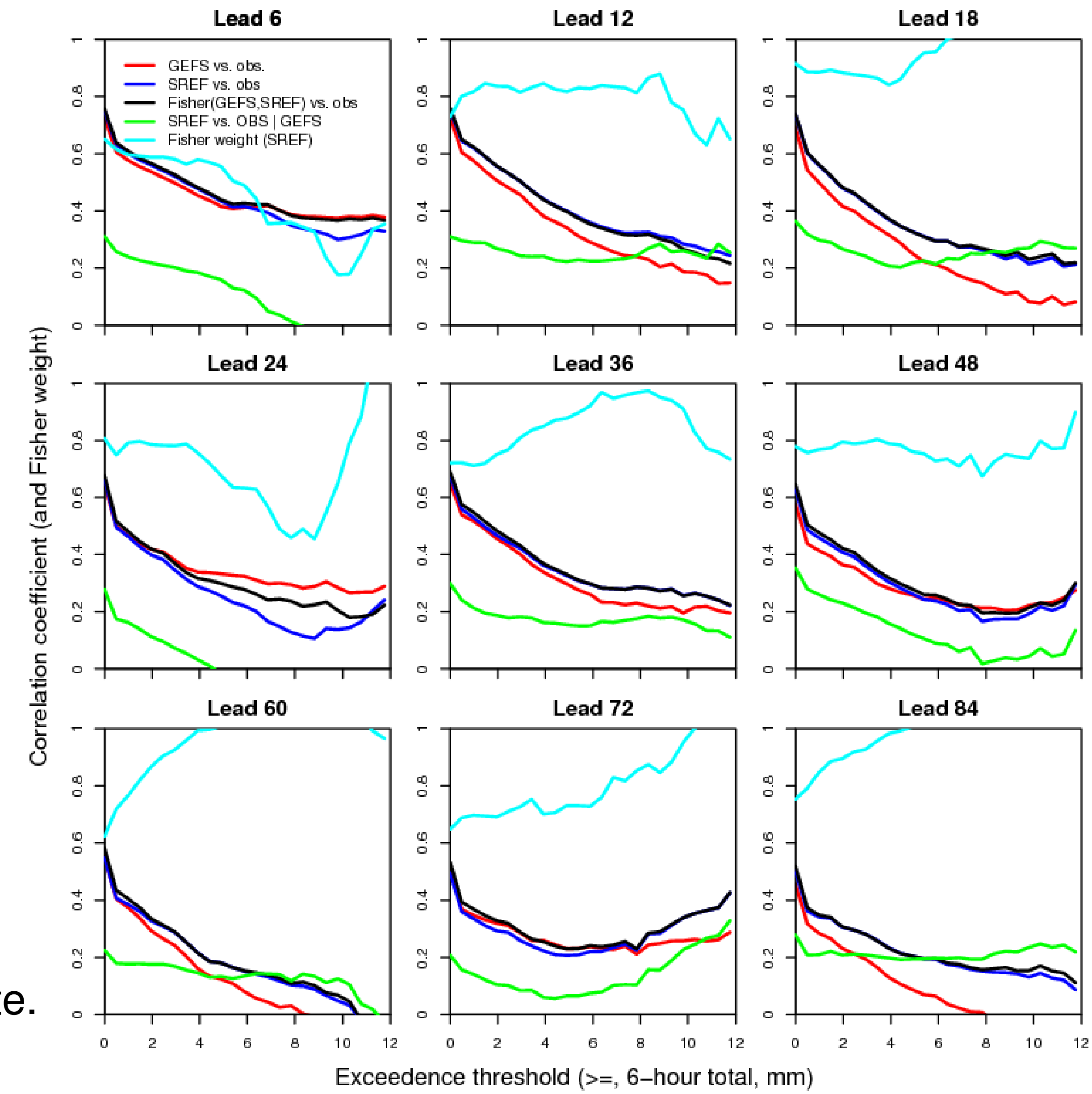


Example results III: any benefit from merging (operational) SREF and (operational) GEFS? [based on SUNY archive]

Used Fisher optimal estimation to blend ensemble means of GEFS and SREF based on how well they predict observed (basically akin to mean of BMA result).

Fisher weight (cyan) is [0,1] and shows contribution from each. A value of 1 = all from SREF, 0 = all GEFS.

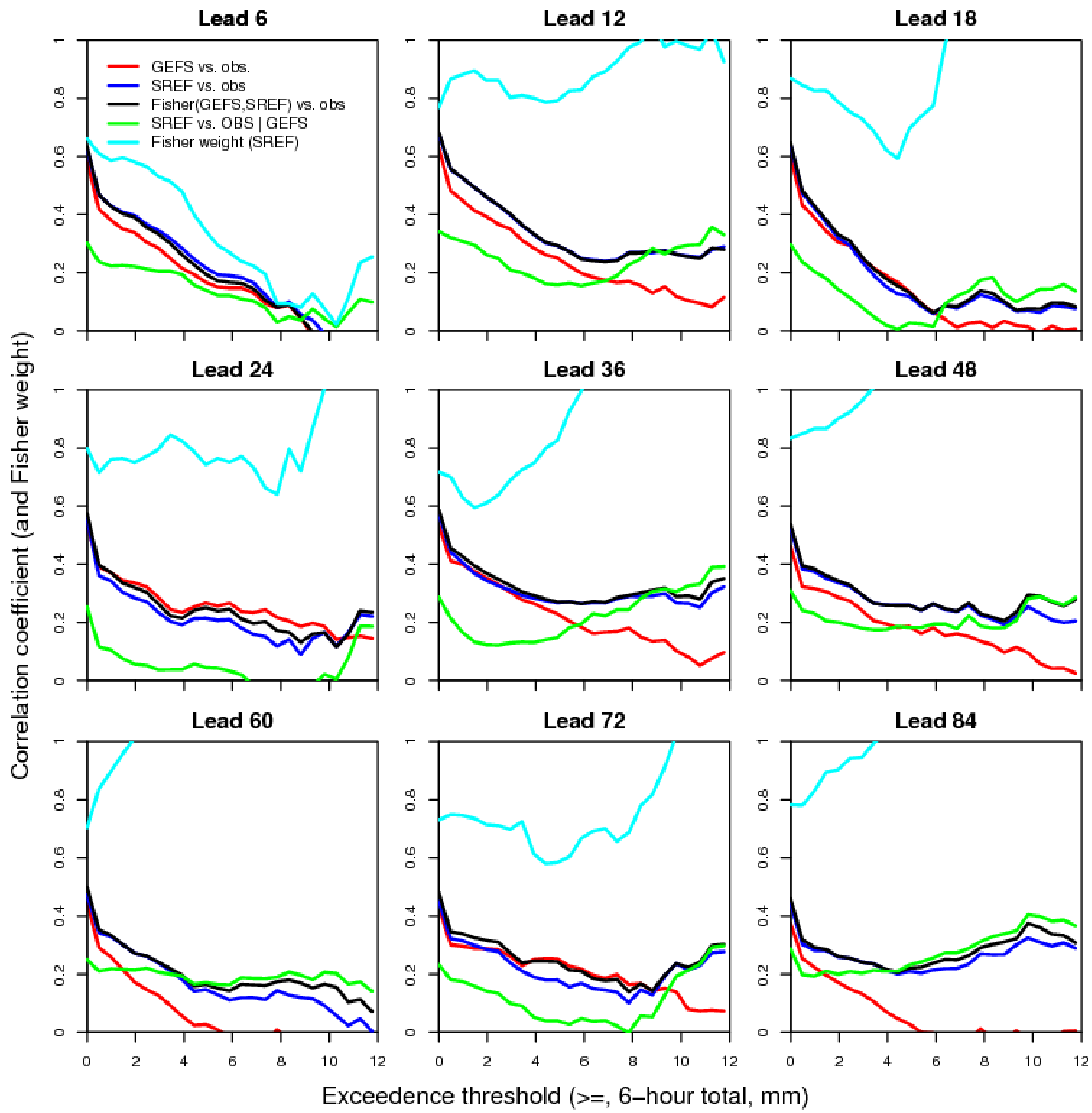
Most information comes from SREF for high-valued precipitation and for longer lead times but GEFS does contribute.



Warm season:

GEFS adds less info. during the warm season, particularly for high-valued precipitation at longer lead times.

Potential problem: how to deal with transition from SREF to GEFS at 3.5 days?



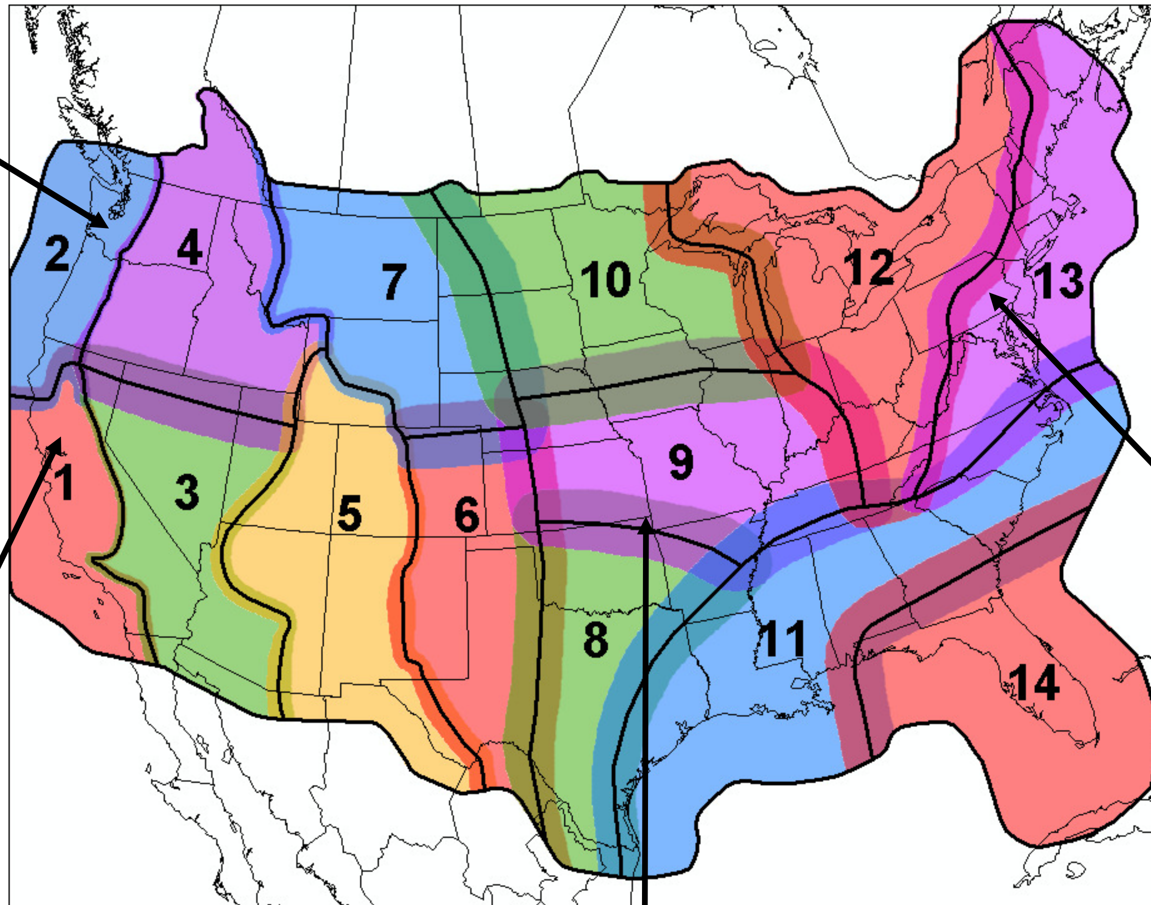
Brief conclusions

- SREF forecasts are surprisingly skilful and reliable even before bias correction.
- Obviously this varies with location. For example, most skill for large events in PNW basins. Much less in Mid-Atlantic basins
- Type-I biases: some over-forecasting of light precipitation and under-forecasting of heavy.
- Type-II biases: extreme under-forecasting of heavy precipitation (i.e. when heavy actually occurs). This will be problematic for RFCs.

Other slides

HR-MOS climate regions

NWRFC

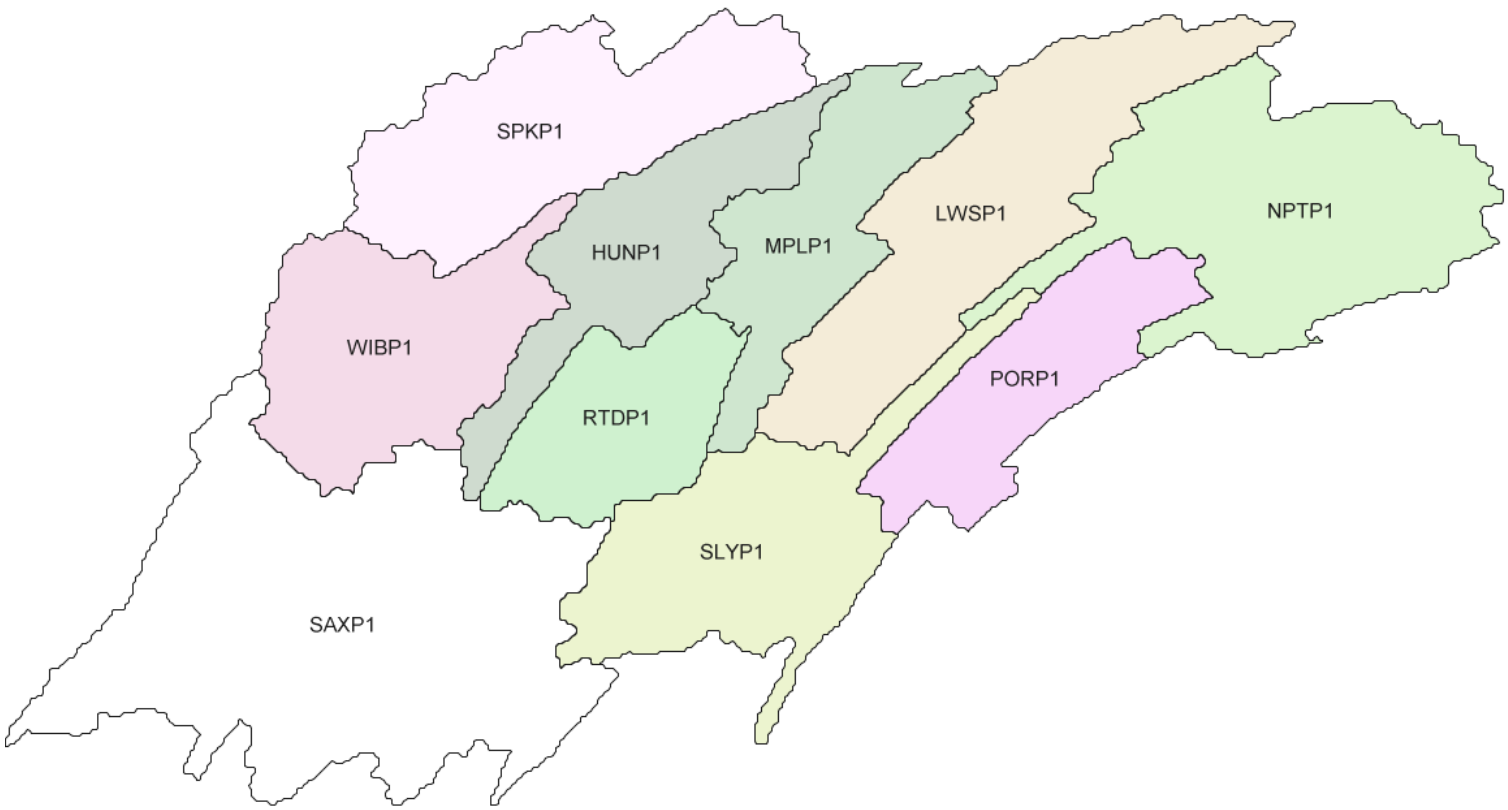


MARFC

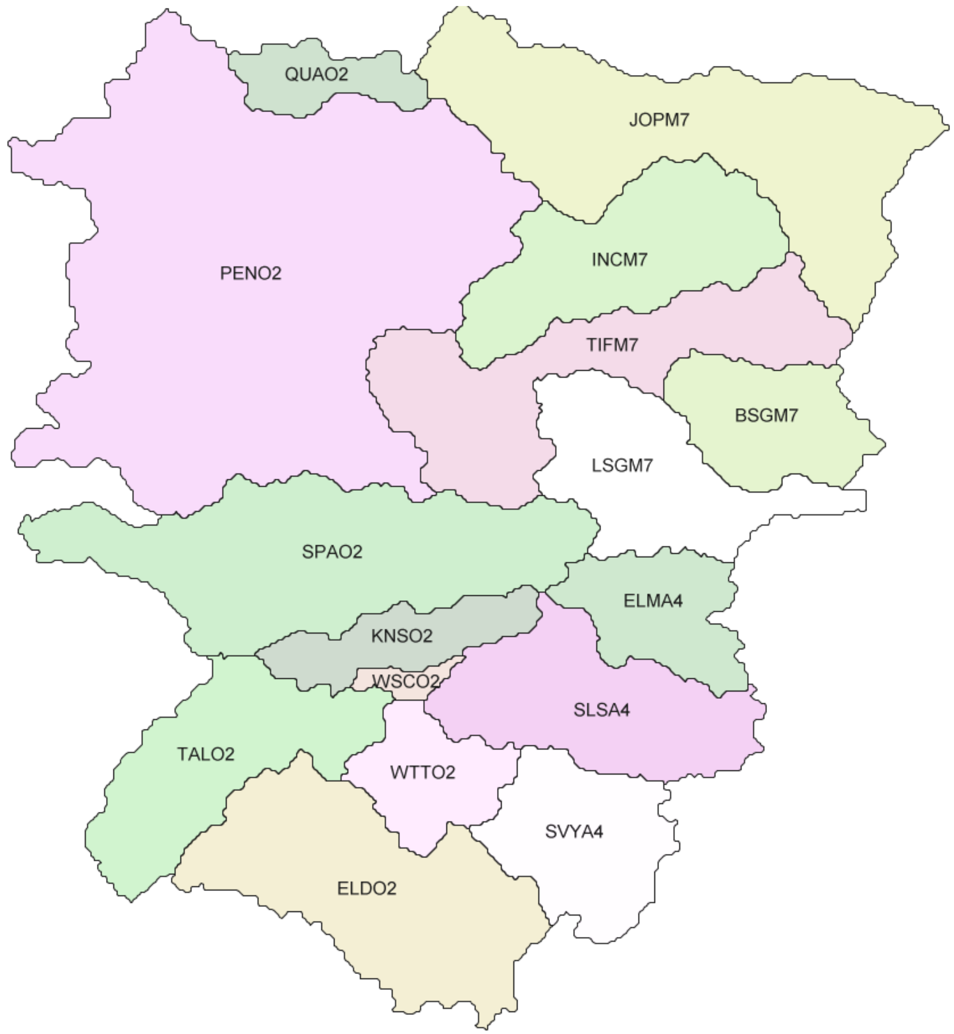
CNRFC

ABRFC: 16 basins

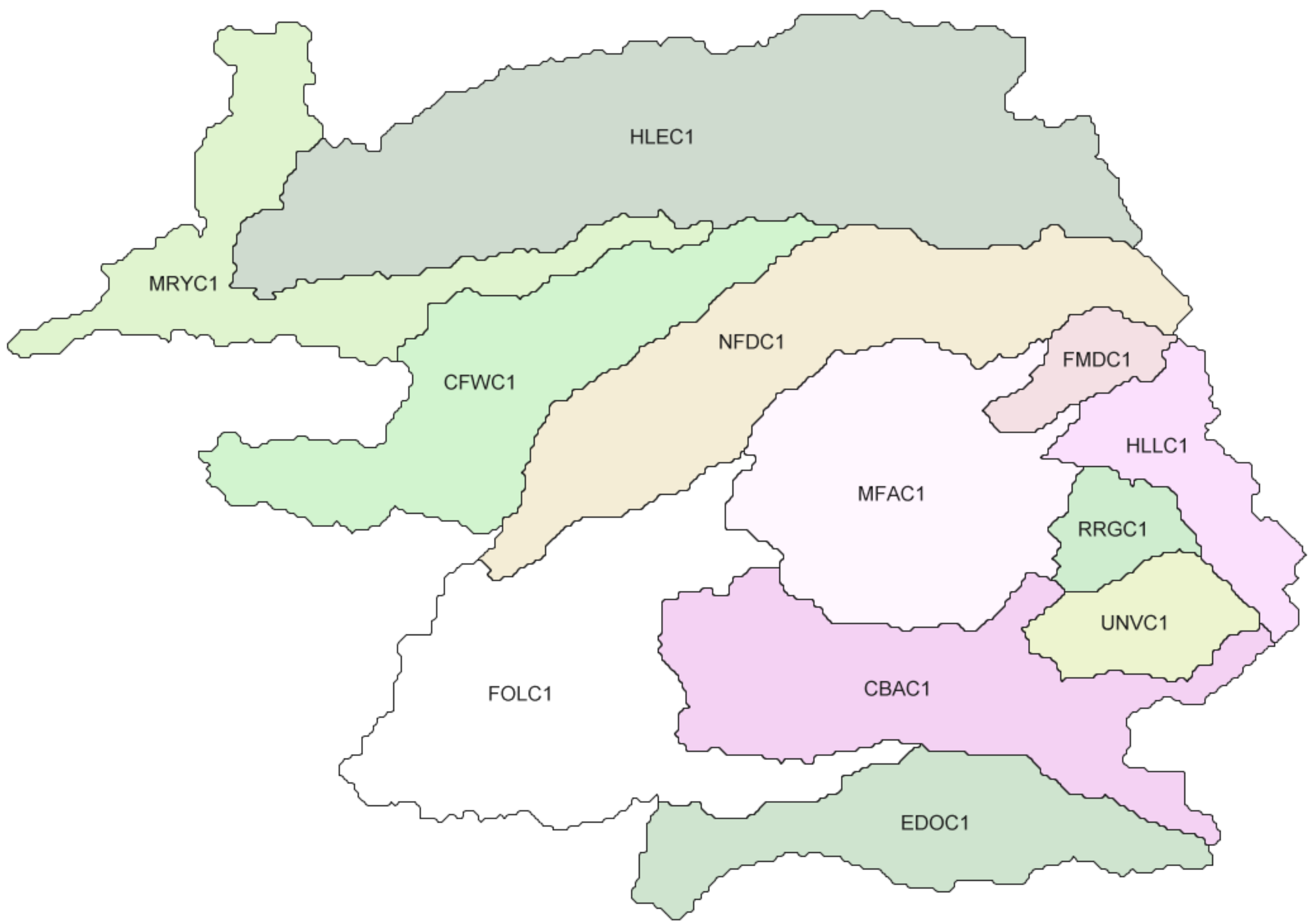
MARFC basins



ABRFC basins



CNRFC basins



NWRFC basins