Development of a New Precipitation Dataset for Bias Correction and Downscaling

Mike Charles, Zoltan Toth (NCEP)
Paul Schultz, Huiling Yuan (ESRL/GSD)

Acknowledgements:
Jess Charba, Dan Collins, Dingchen Hou, John Janowiak, Ken Mitchell, DJ Seo, Letitia Soulliard, Dave Unger, Pingping Xie (NCEP)
Roman Krzysztofowicz (Univ. VA)

Why do we need another dataset?

1. This new dataset will be our best estimate for truth on a 5x5 km (NDFD), 6-hourly grid
2. Bias correction of NAEFS precipitation:
   - Need accurate, quality controlled precipitation data for the Bayesian Processor of Ensemble (BPE) prior distribution
3. Downscaling NAEFS precipitation forecasts:
   - Need data with a high spatial resolution to downscale NAEFS precipitation forecasts to the NDFD grid
4. Verification of NAEFS precipitation forecasts:
   - Need accurate dataset to verify NAEFS forecasts

Note: This effort has limitations, as it was developed to simply combine existing datasets. Much more work will be needed for a more comprehensive approach, but this is out of the scope of this work

Using Information From 2 Datasets

1. CPC Unified Precipitation Analysis
   - Back to 2000 (eventually back to 1948)
   - ⅛° spatial resolution
   - Daily
   - Global land
2. RFC Quantitative Precipitation Estimate
   - Back to late 2000
   - ~5km spatial resolution
   - 6-hourly
   - CONUS

- RMORPH
  - Future use for global post-processing

Combining Information

CPC:

✓ More confidence in long term statistics of CPC dataset

a. Uniform QC across entire domain
b. Gauge-based

✗ Too low resolution for downscaling

RFC:

✓ High resolution nearly equal to NDFD grid → better representation of fine scale temporal and spatial variability

✗ Non-uniform QC (different RFCs have different methods)

✗ Each RFC may make their own adjustments before mosaicking

Solution: adjust RFC grids so their climatology is consistent with the CPC dataset

✓ Have the reliability of the CPC dataset, with the high spatial and temporal resolution of the RFC dataset

Summary of Methods

1. Interpolate RFC data to ⅛° match temporal and spatial resolution of CPC data
2. Establish statistical relationship between CPC and RFC datasets on the ⅛° grid
3. Adjust RFC dataset to make its climatology look like the CPC dataset
4. Downscale adjusted RFC dataset back to original temporal and spatial resolution

Step 1 - Match Resolutions

Interpolating RFC to ⅛°

1. Accumulate RFC precip over same 24-hour period
   - 12UTC - 12UTC
2. Interpolate daily RFC grid to ⅛°
   - Copygb utility
     - Budget interpolation scheme
     - Attempts to maintain area-averages

- End result:
  - ~7 years of CPC and RFC daily precipitation grids at ⅛° spatial resolution
- Relationship is established at the resolution of the coarsest dataset, otherwise we will establish relationships with imaginary data (if we interpolated CPC to 5km)

Step 2 - Establish Statistical

Regress CPC against RFC to get a & b

- Relationship established at each gridpoint, for each day of the year (a function of geography and regime)
- Use a 60-day window surrounding each day of the year
  - Max of ~420 observations per gridpoint (60 days * 7 years)
- Collect all obs. in window where RFC > 0
- Regress CPC against RFC

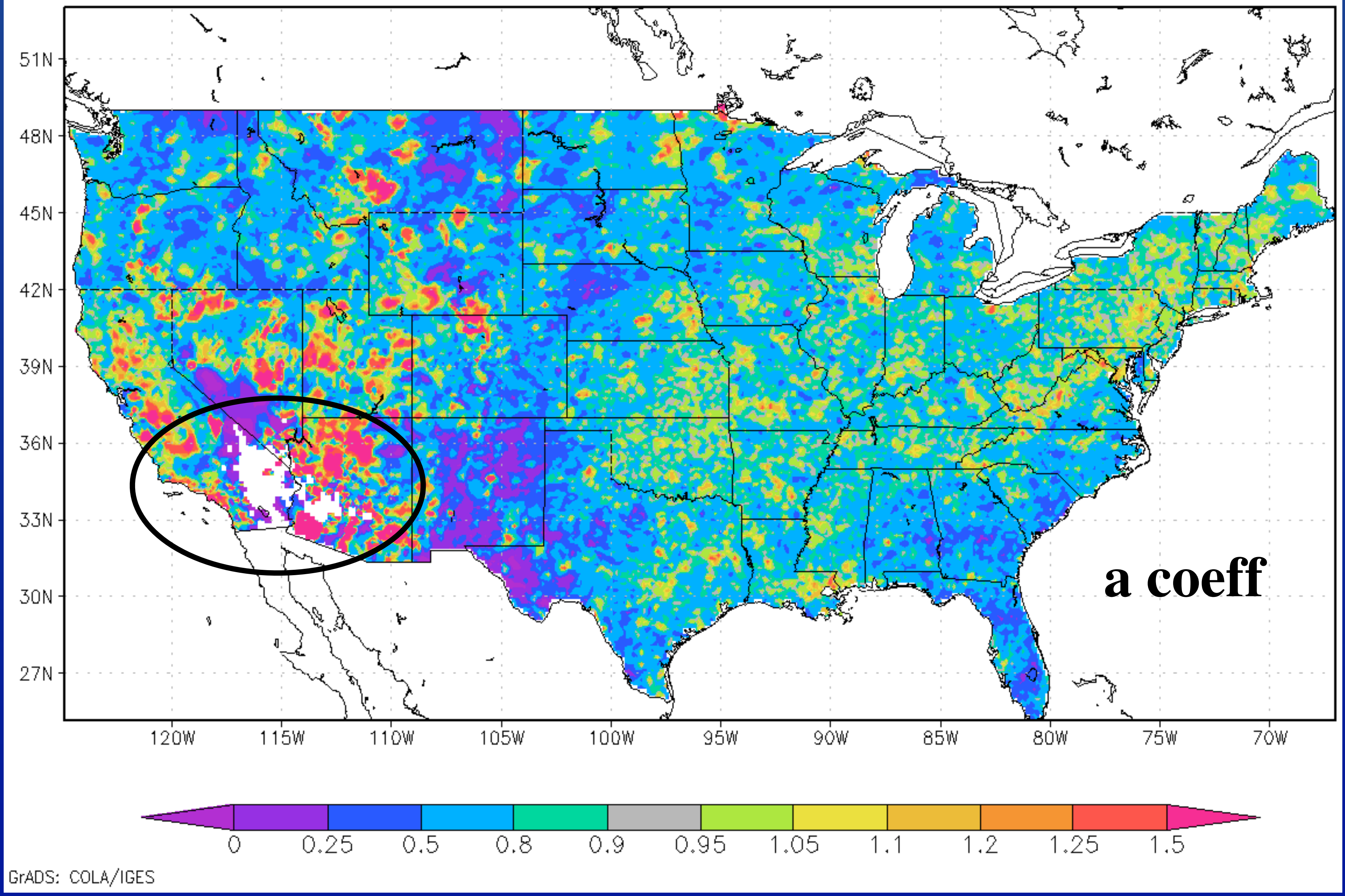
$$CPC=a*RFC+b$$

- Store an ⅛° grid of a and b for each day of the year

Step 2 - Establish Statistical Relationship

Smooth grids and fill missing data

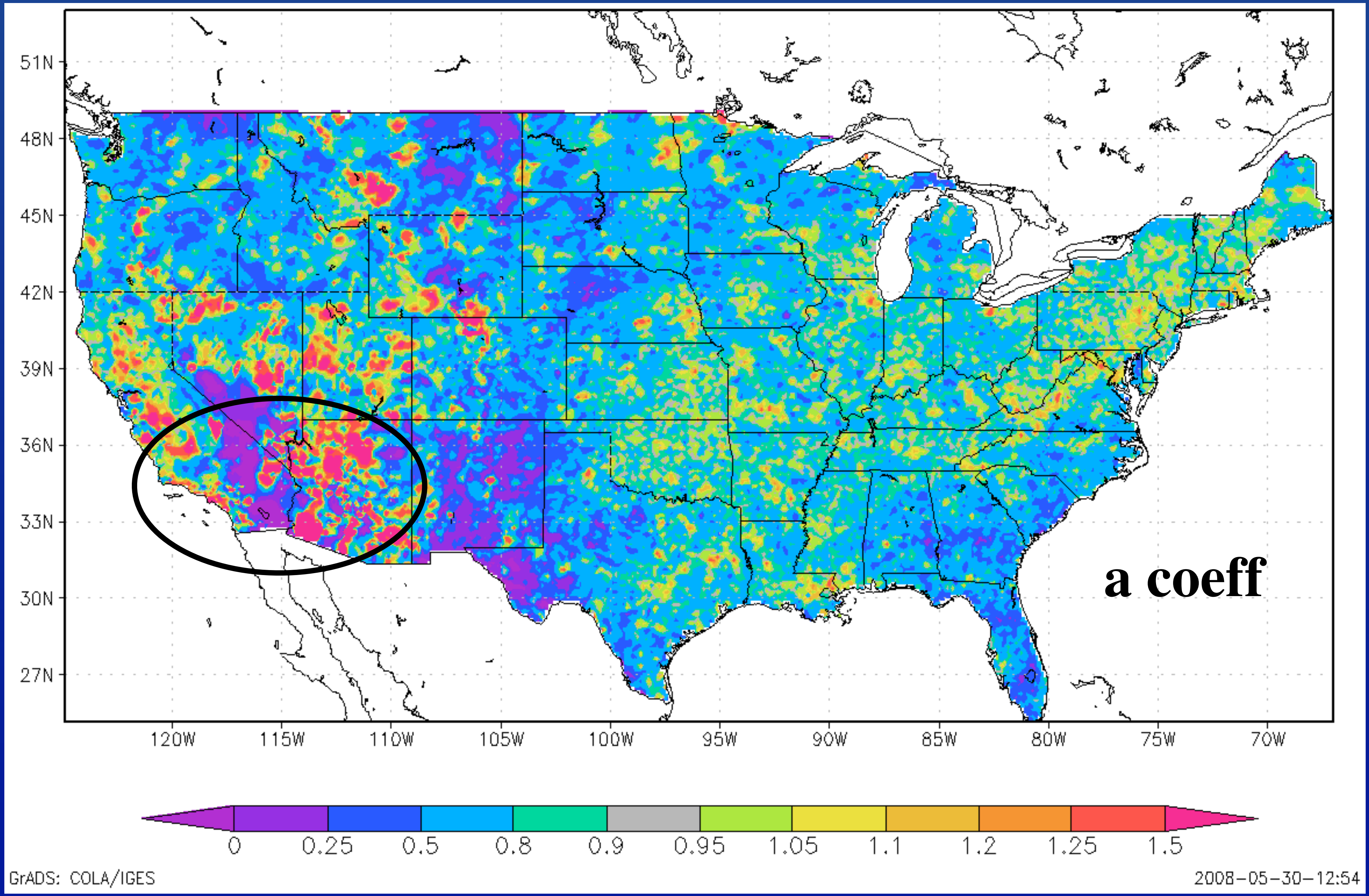
- Raw a and b grids have missing data
  - Even with 60-day window, some regions are dry



Step 2 - Establish Statistical Relationship

Smooth grids and fill missing data

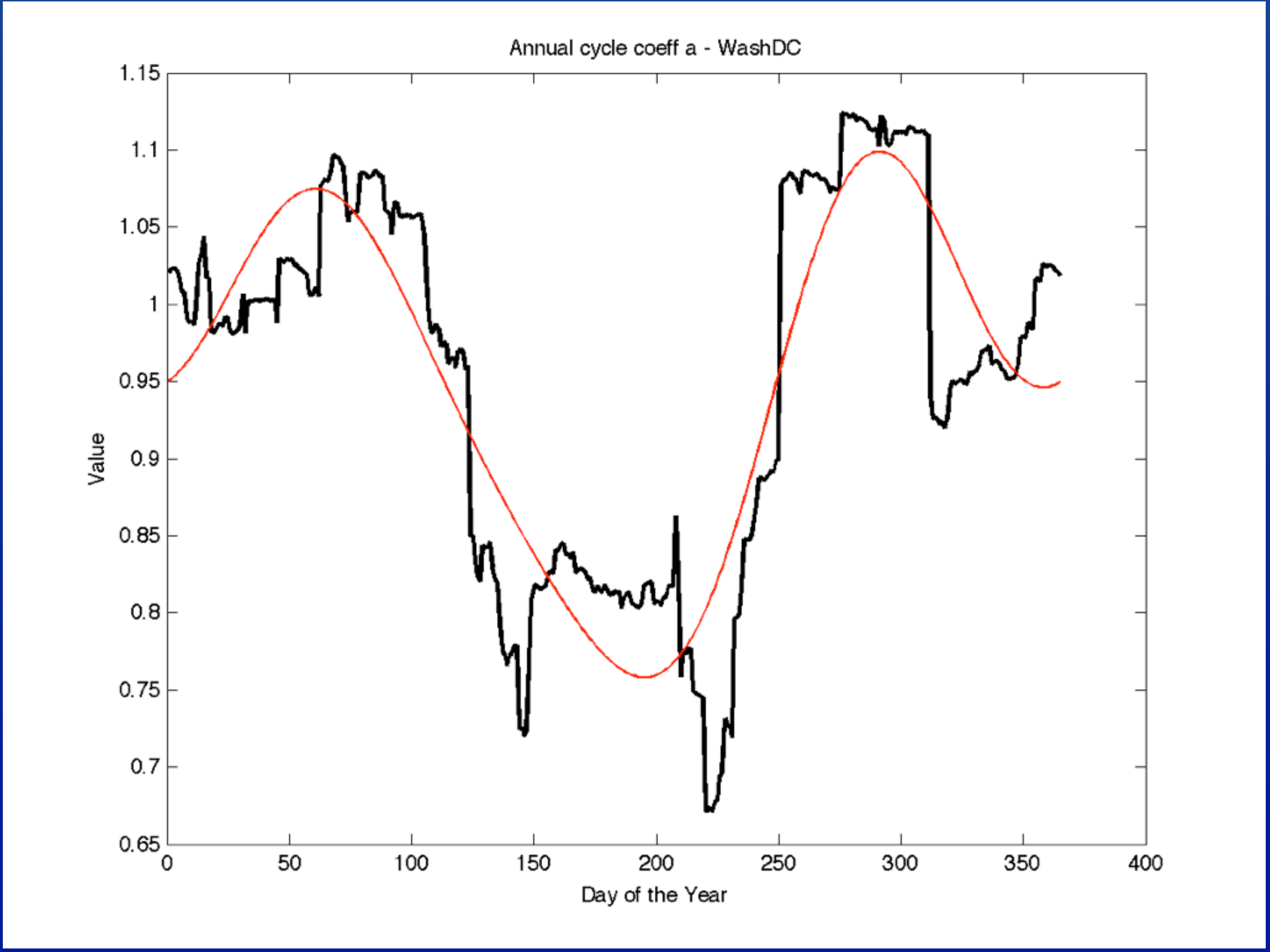
- Fill missing data with bilinear interpolation



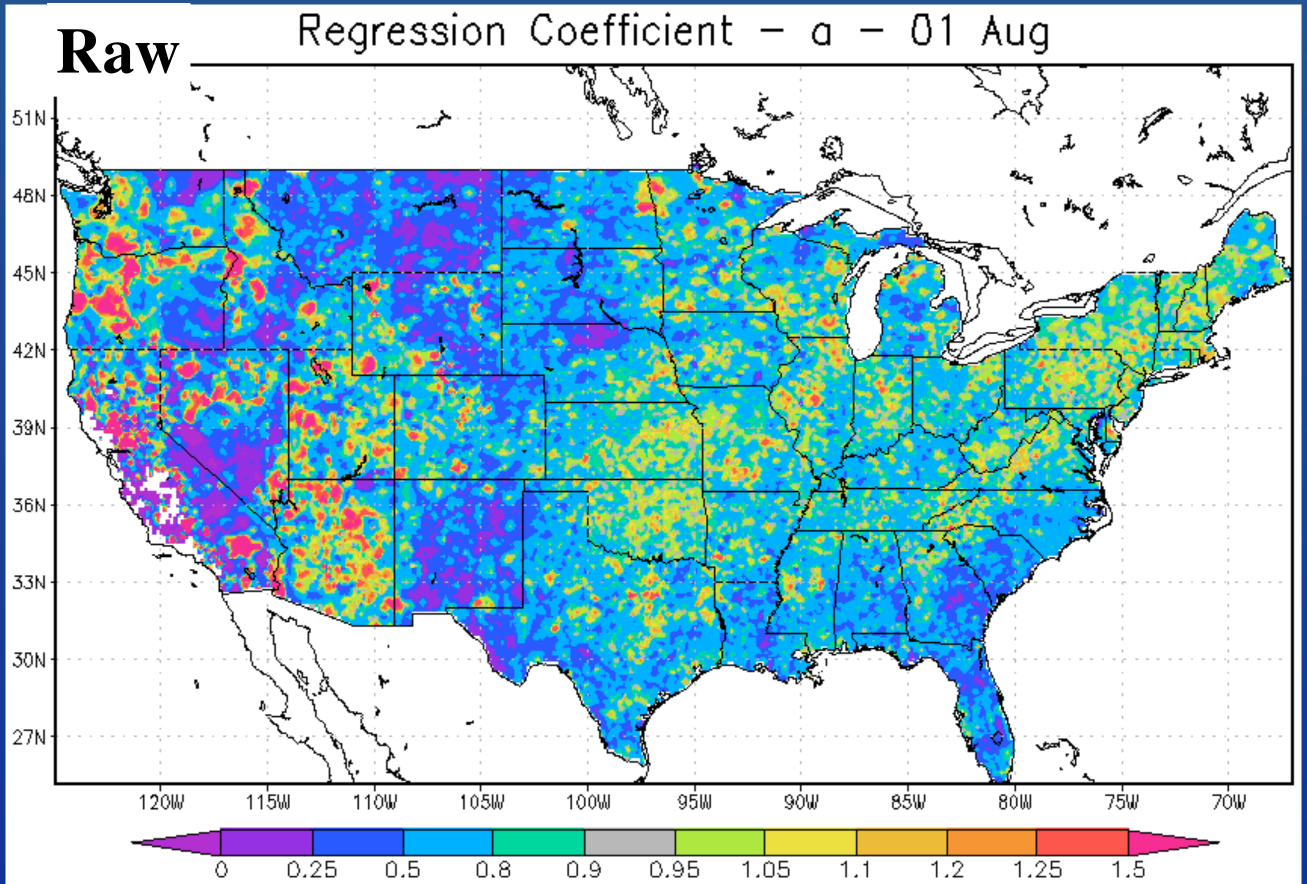
Step 2 - Establish Statistical Relationship

Smooth timeseries of a and b at each gridpoint

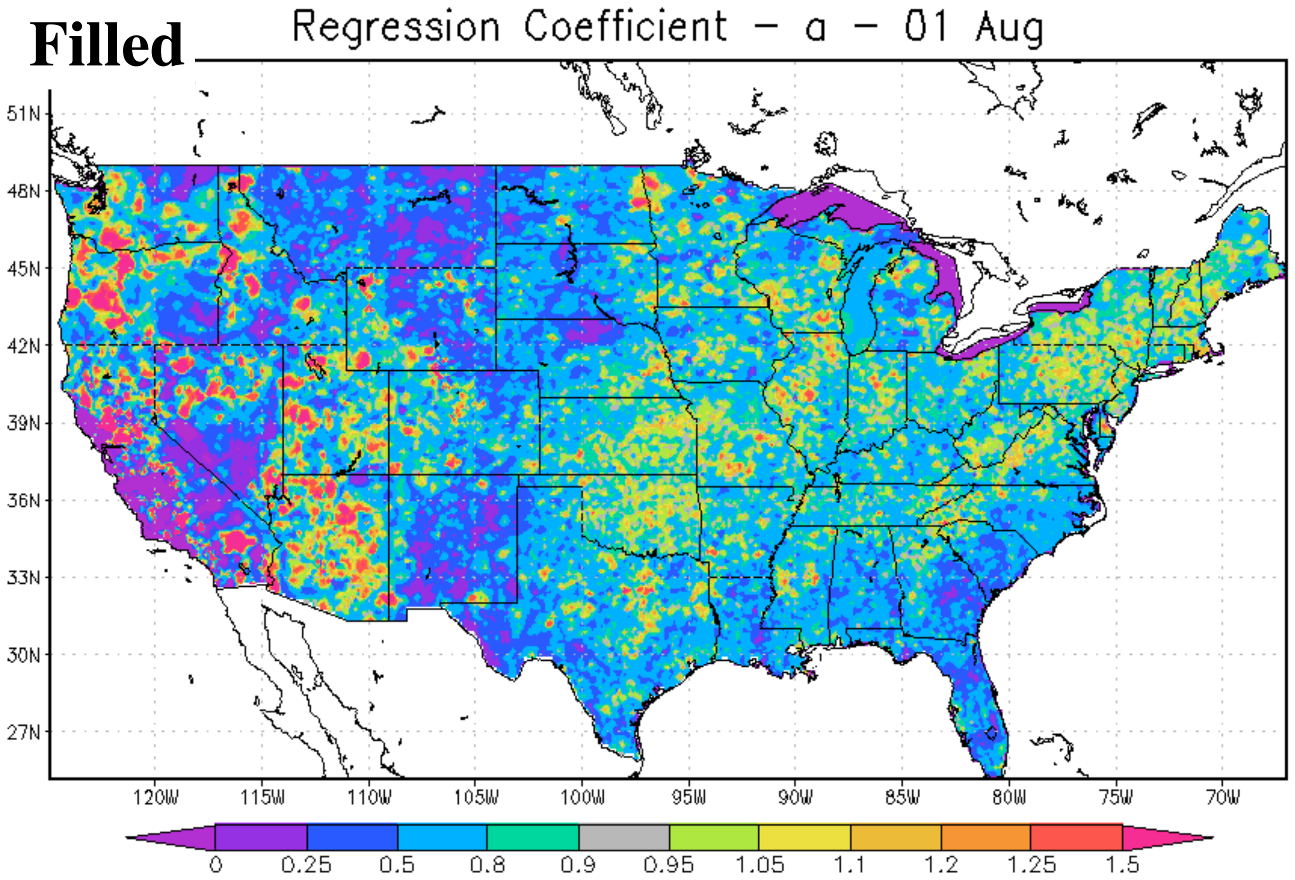
- Fit Fourier Transform (3 harmonics) to raw timeseries
- Replace grid with smoothed coefficients

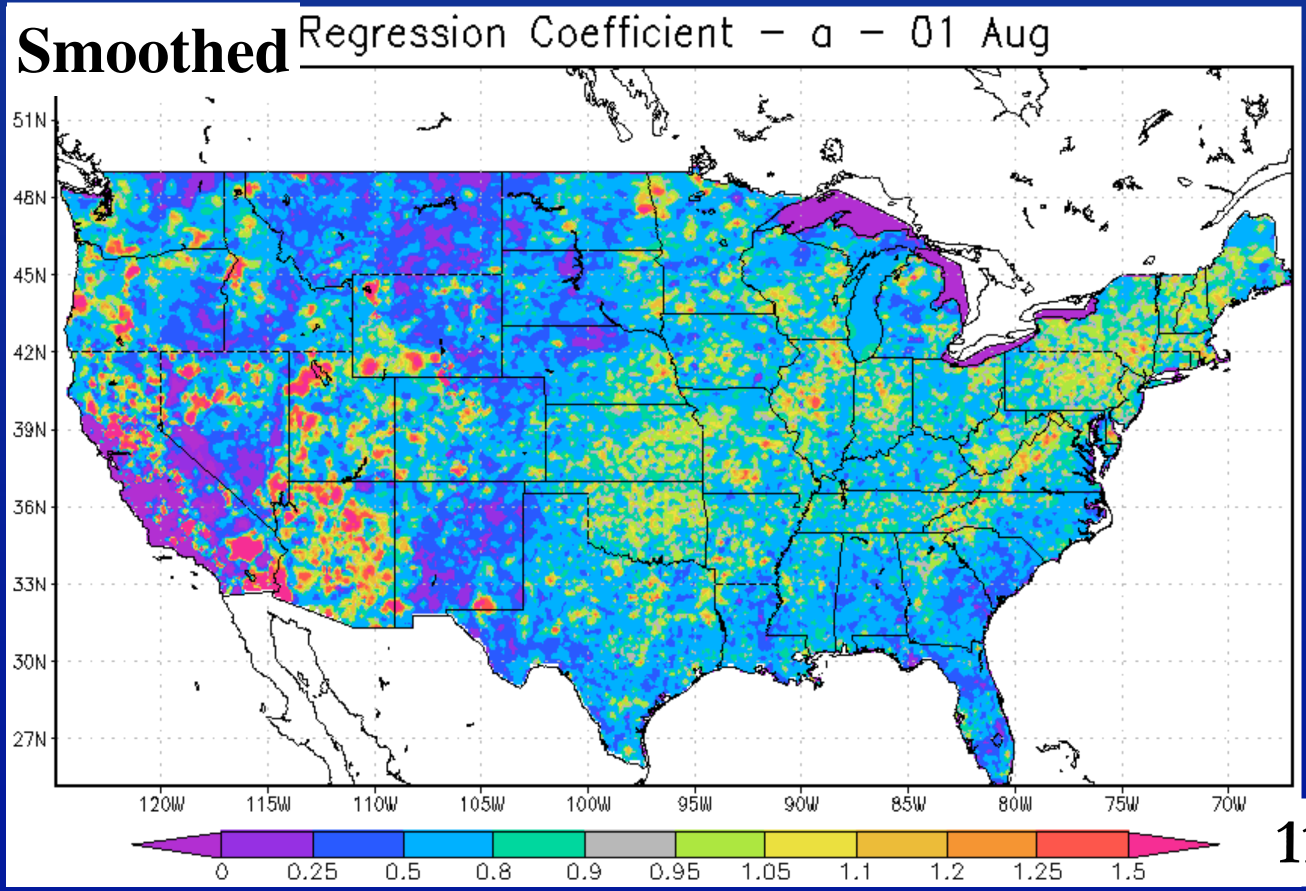


a and b Grids - Before and After



- **Raw**: before processing
- **Filled**: after filling missing data with bilinear interpolation
- **Smoothed**: after fitting Fourier Transform at each gridpoint
- See separate movie files for evolution of grids over time



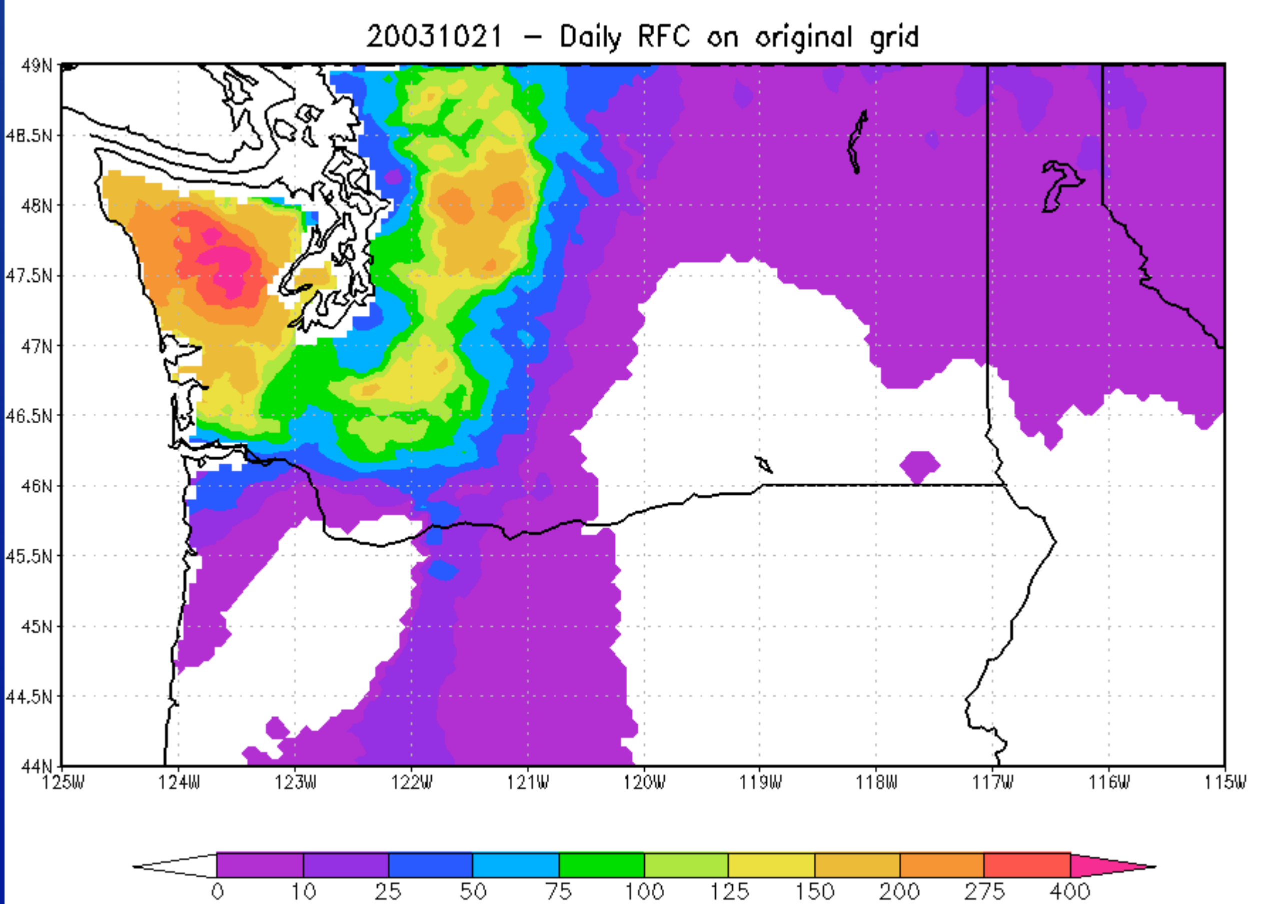


Step 3 - Adjust RFC dataset

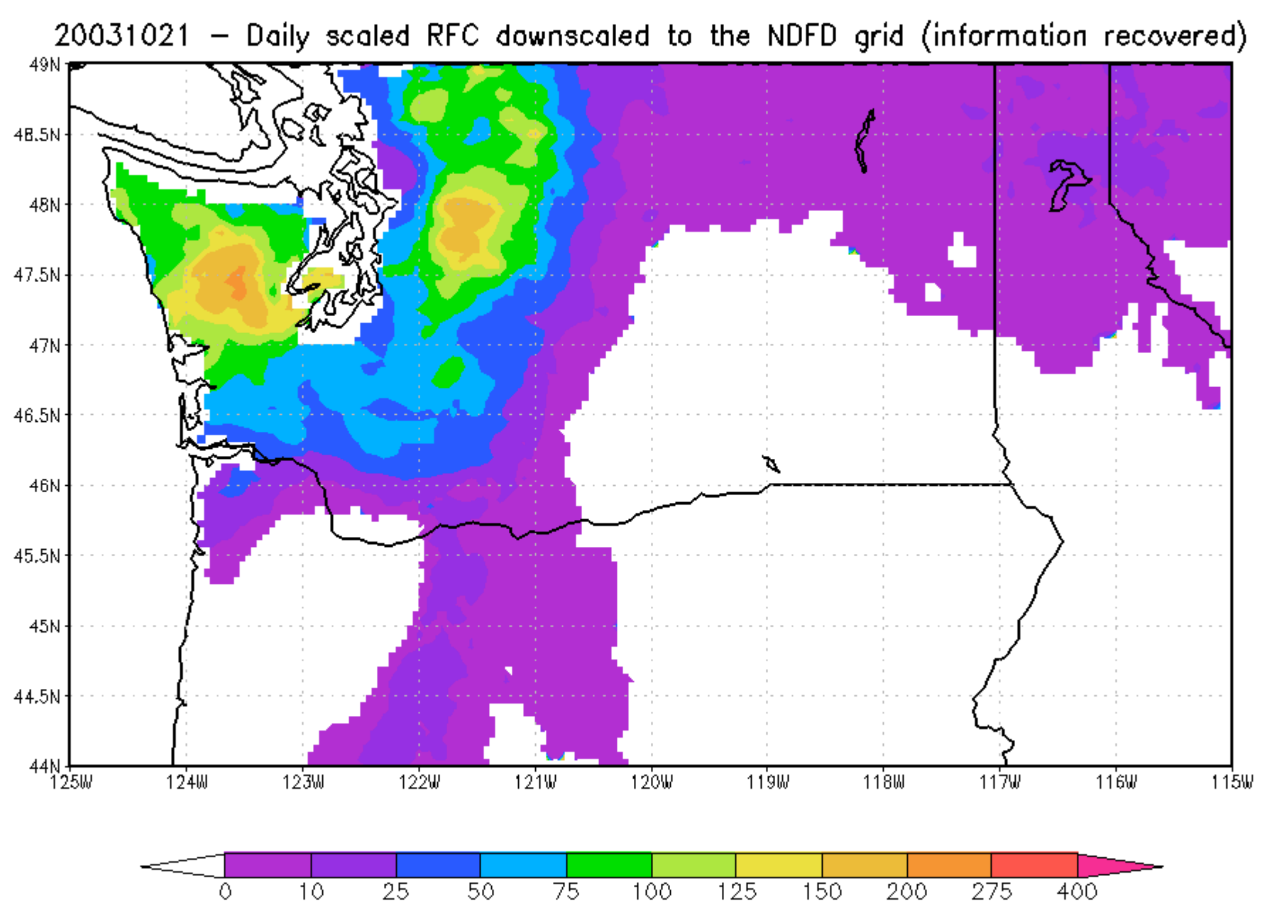
$$CPC^* = aRFC + b$$

- Transform linearly each daily ⅛° RFC grid:
  - Scale by *a* and *b* to get CPC*
- Use appropriate *a* and *b* grid based on day of the year

RFC



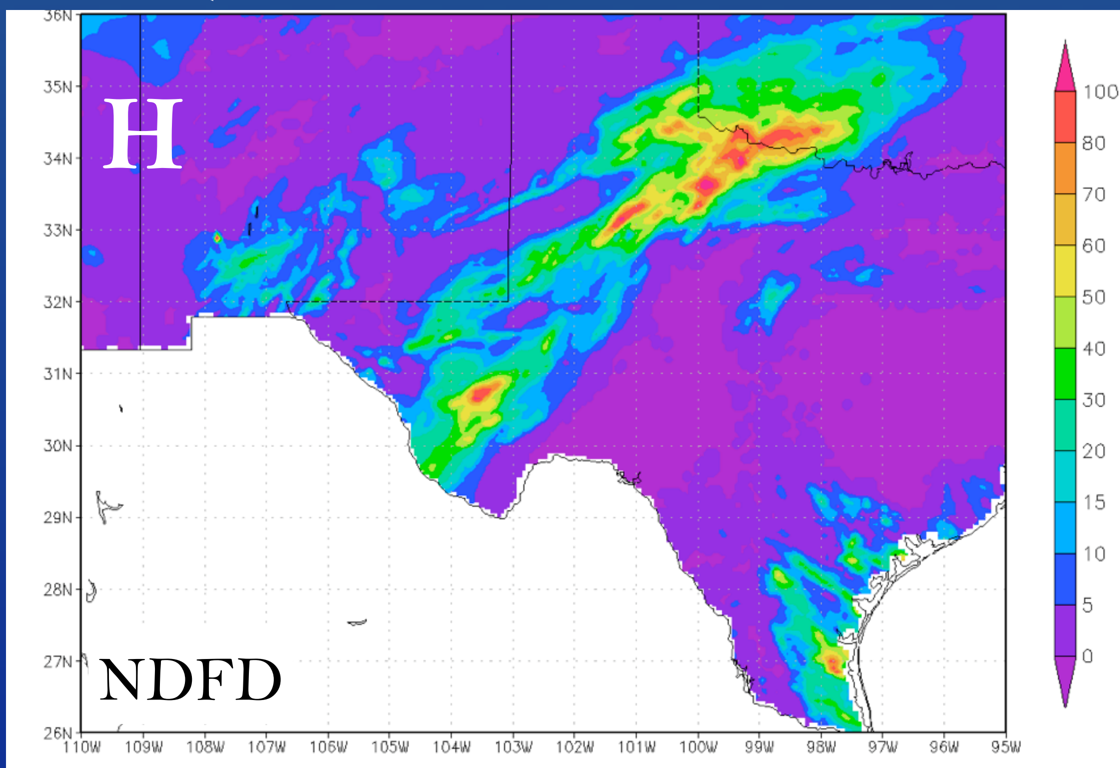
CPC*

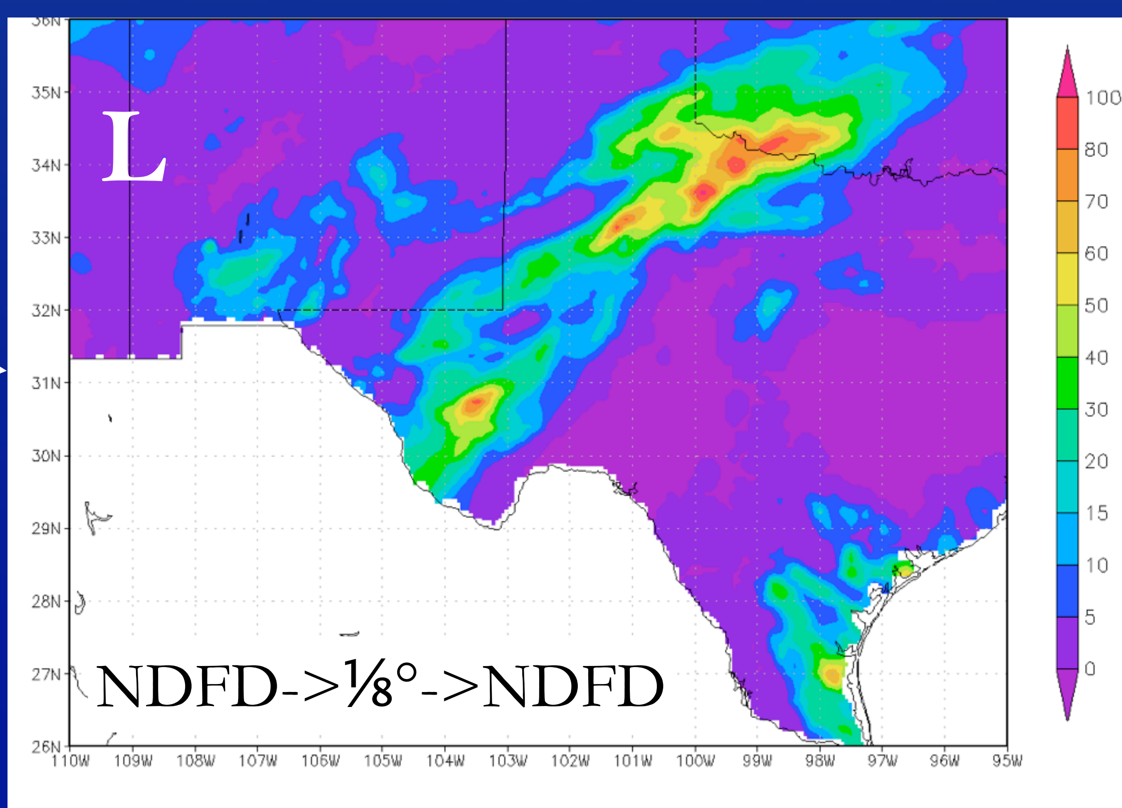


Recovering Original RFC Resolutions

Spatial Disaggregation

Daily RFC precip

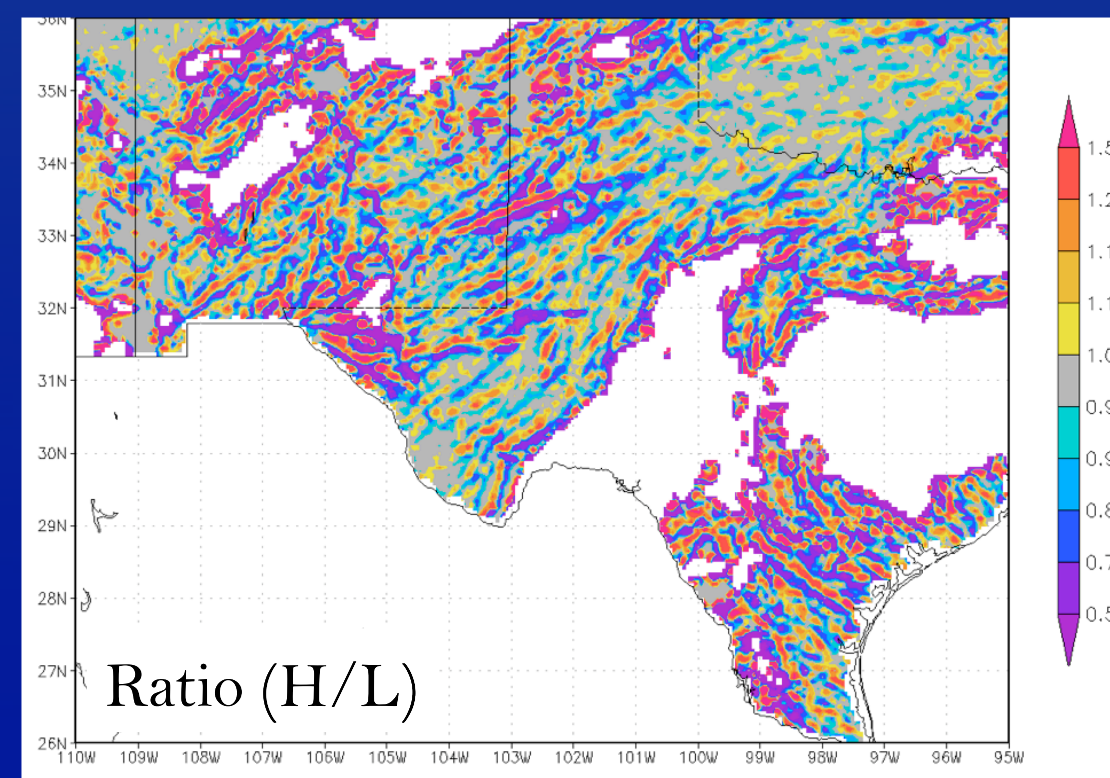




- Information is lost between H and L
- How much information?
  - Take ratio H/L (below)
  - This ratio can be used to put high resolution information back into RFC*

1. Interpolate RFC* to NDFD
2. Multiply by H/L

- End with RFC* at NDFD resolution.
- Spatial information recovered from RFC_{orig}

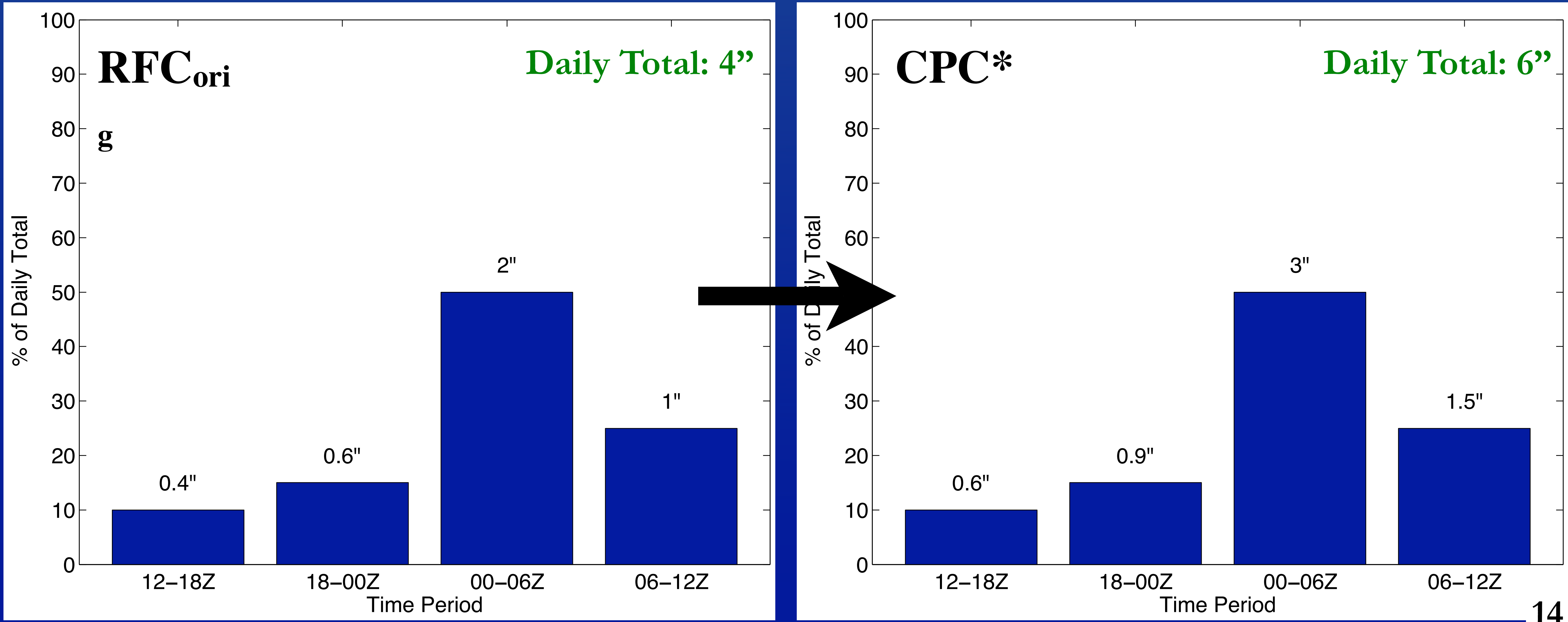


Recovering Original RFC Resolutions

Temporal Disaggregation

1. Determine percentage of daily total precipitation in each 6-hour period in RFC_{orig}
2. Divide 24 hour RFC* into four 6-hour precip amounts using the percentages from RFC_{orig}

Percent of daily total in each 6-hourly period

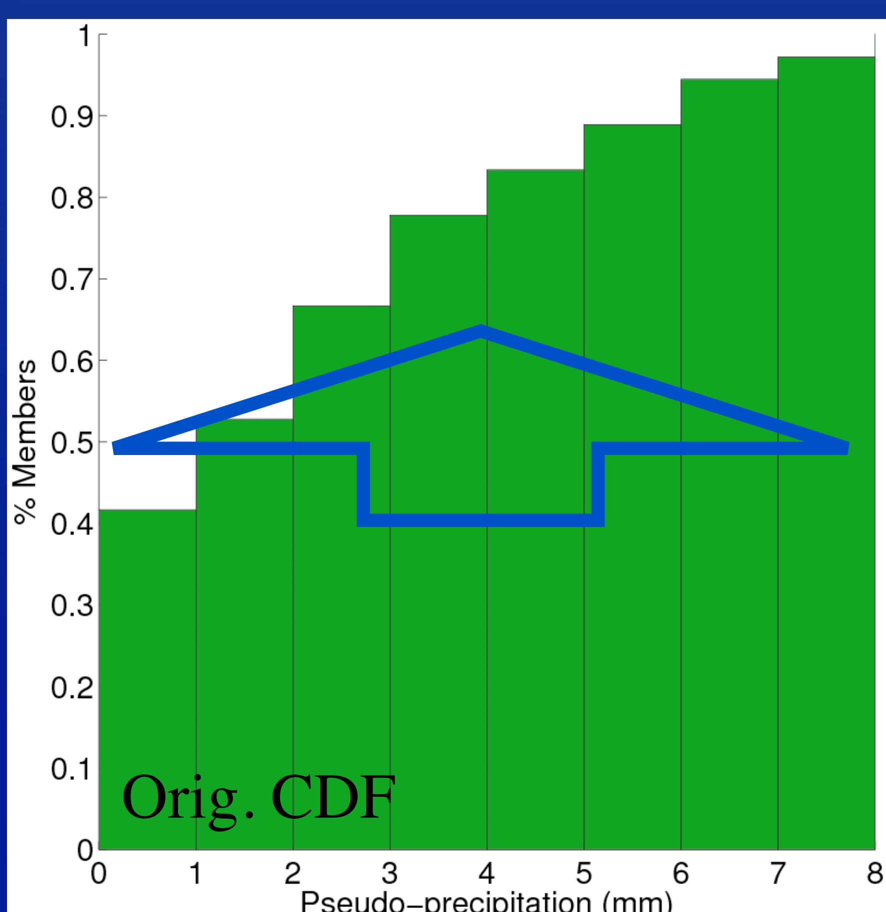


Pseudo-precipitation

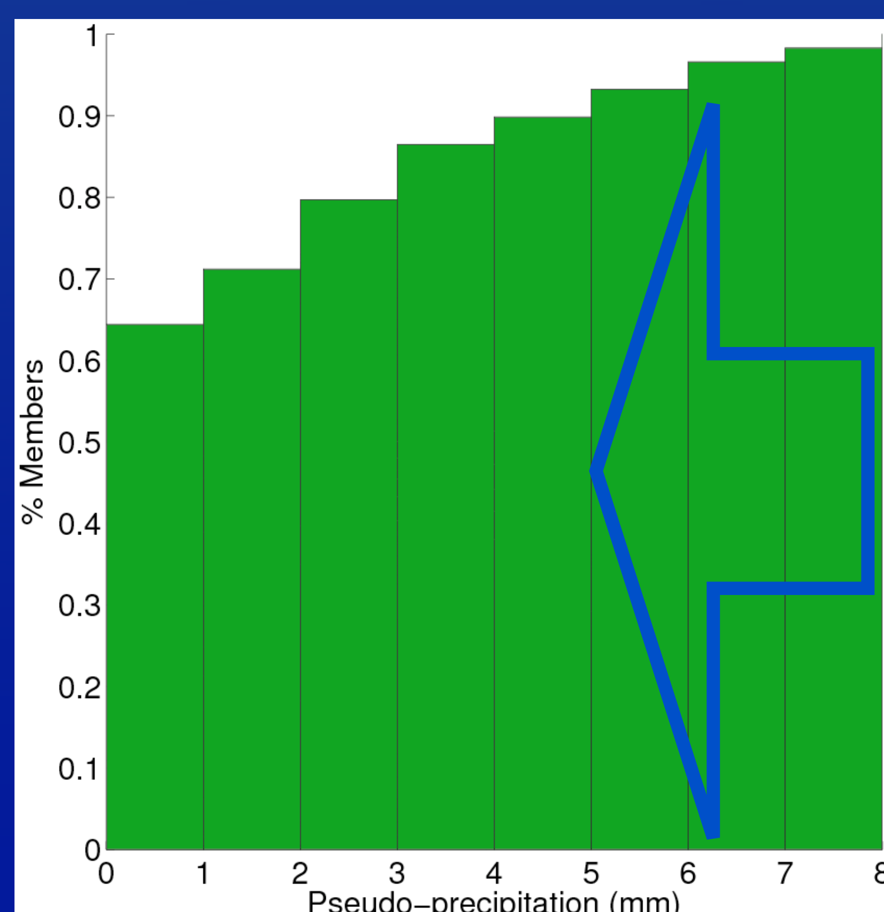
Problems with Precipitation

- Unlike most other variables, precipitation is discontinuous in magnitude
  - Requires different processing than continuous variables:
    1. Adjust P.O.P.
    2. Adjust mean/spread

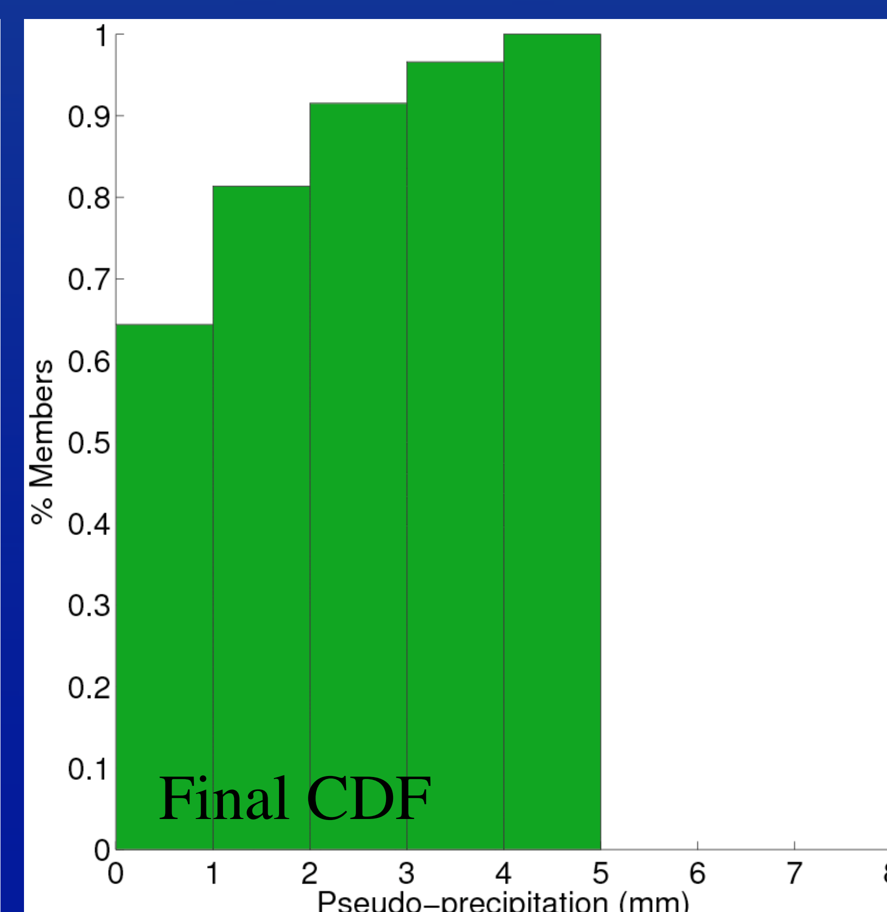
P.O.P. is adjusted first



Then mean/spread are adjusted

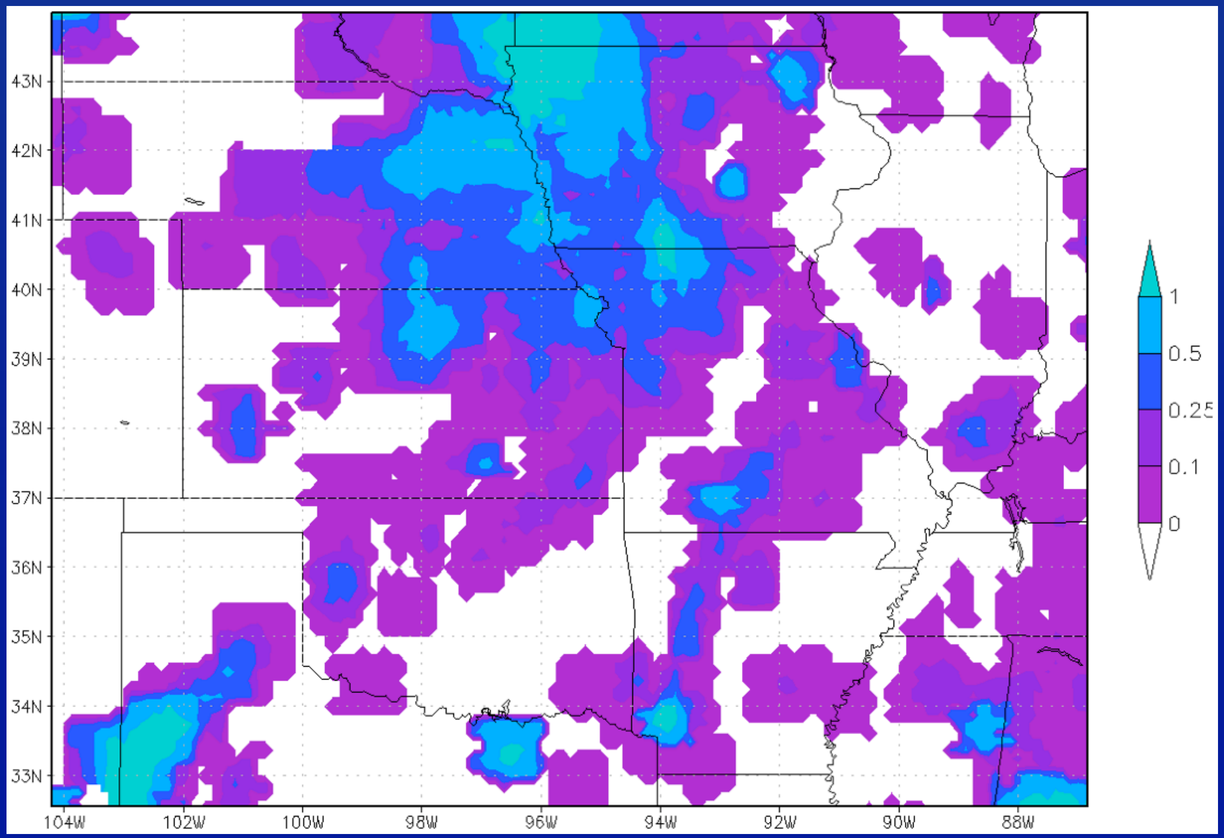


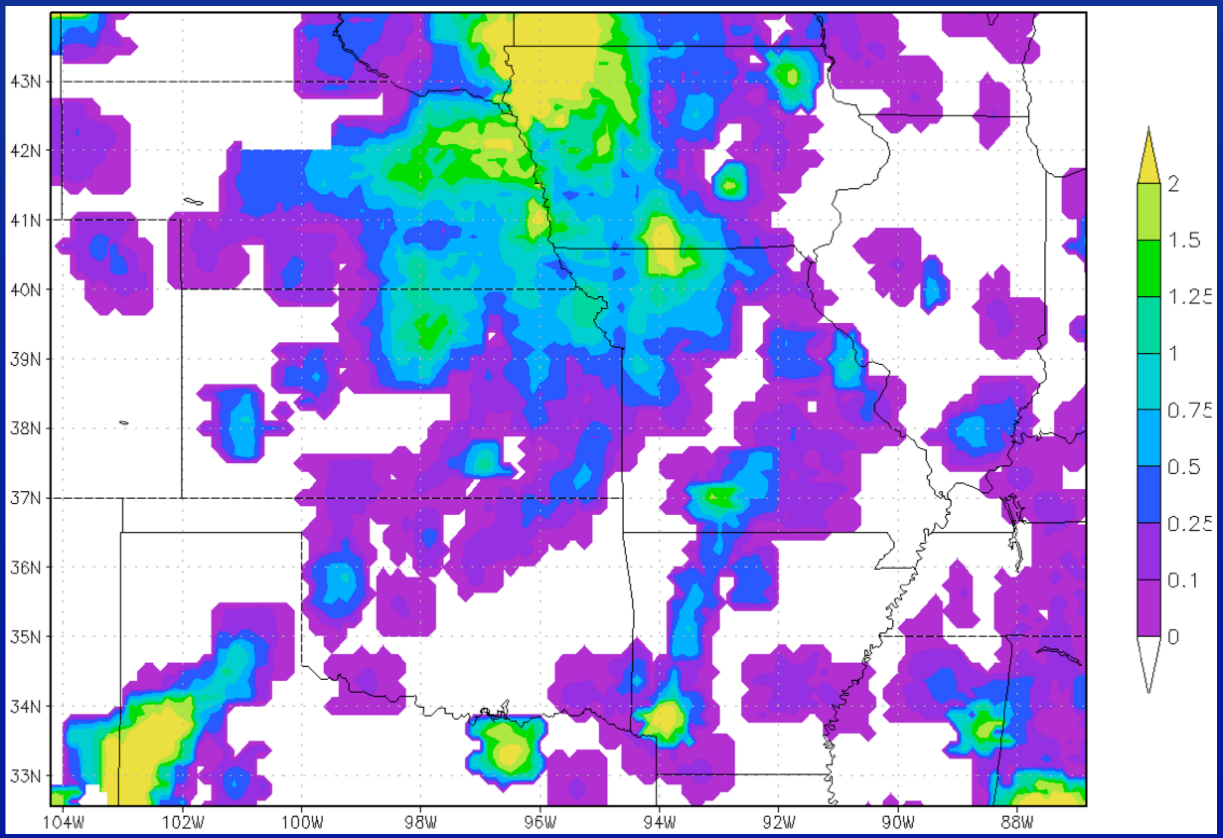
New POP and amounts



Problems with Precipitation

- Unlike most other variables, precipitation is also discontinuous in space
  - Have to treat regions with minimal precip. independently of regions of no precip. (even though they are closely related)
  - For example, large region of light drizzle (below). Model generates too little precip over too little area (left). With normal precip., cannot force members to precipitate, can just increase precip. in already wet members (right)





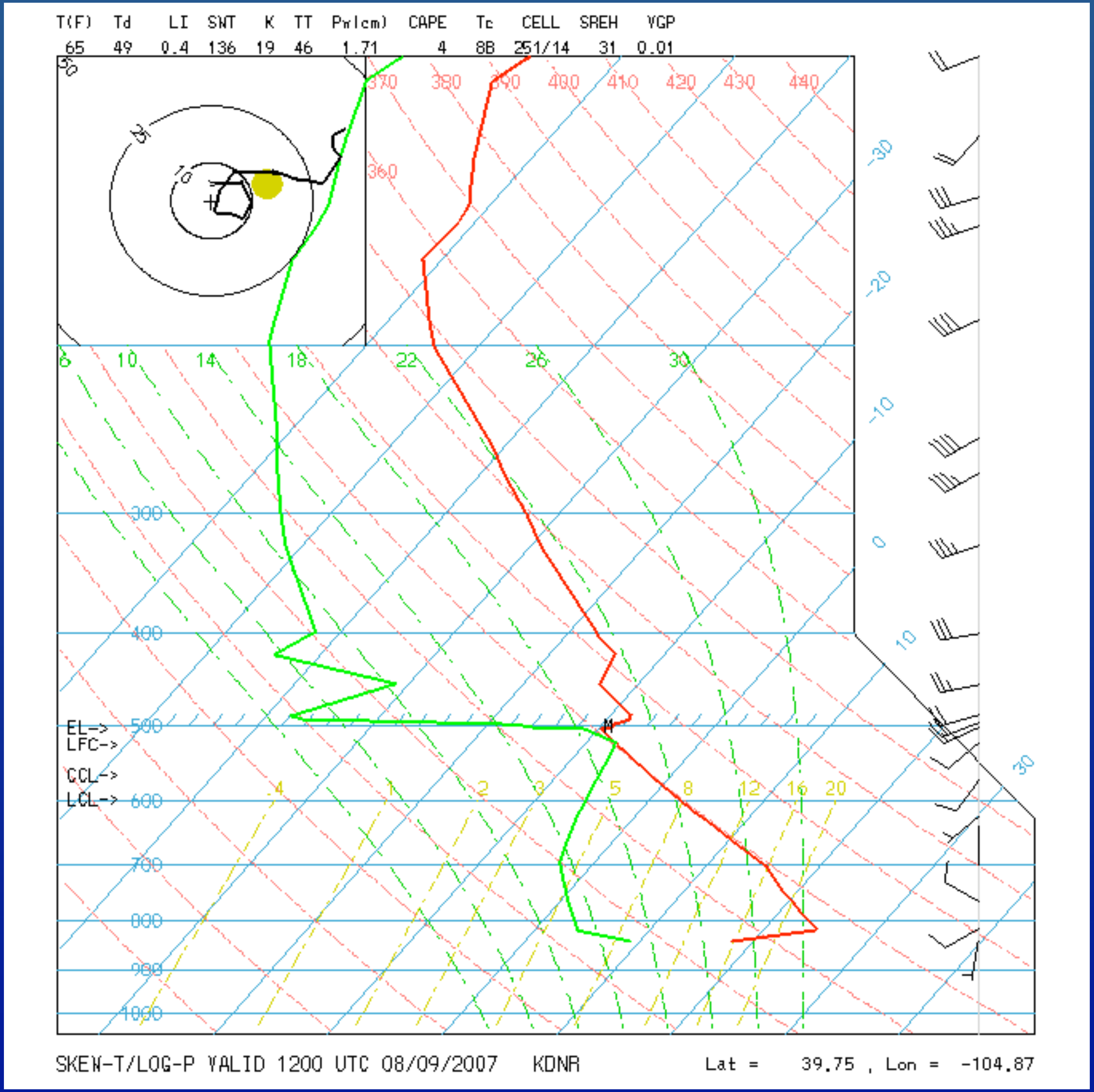
Introducing pseudo precipitation

$$PP(i,j) = P \qquad \text{if } P > 0$$

$$= \frac{1}{g}\int (q_v - q_{vsat}(T))\,dp \qquad \text{if } P = 0$$

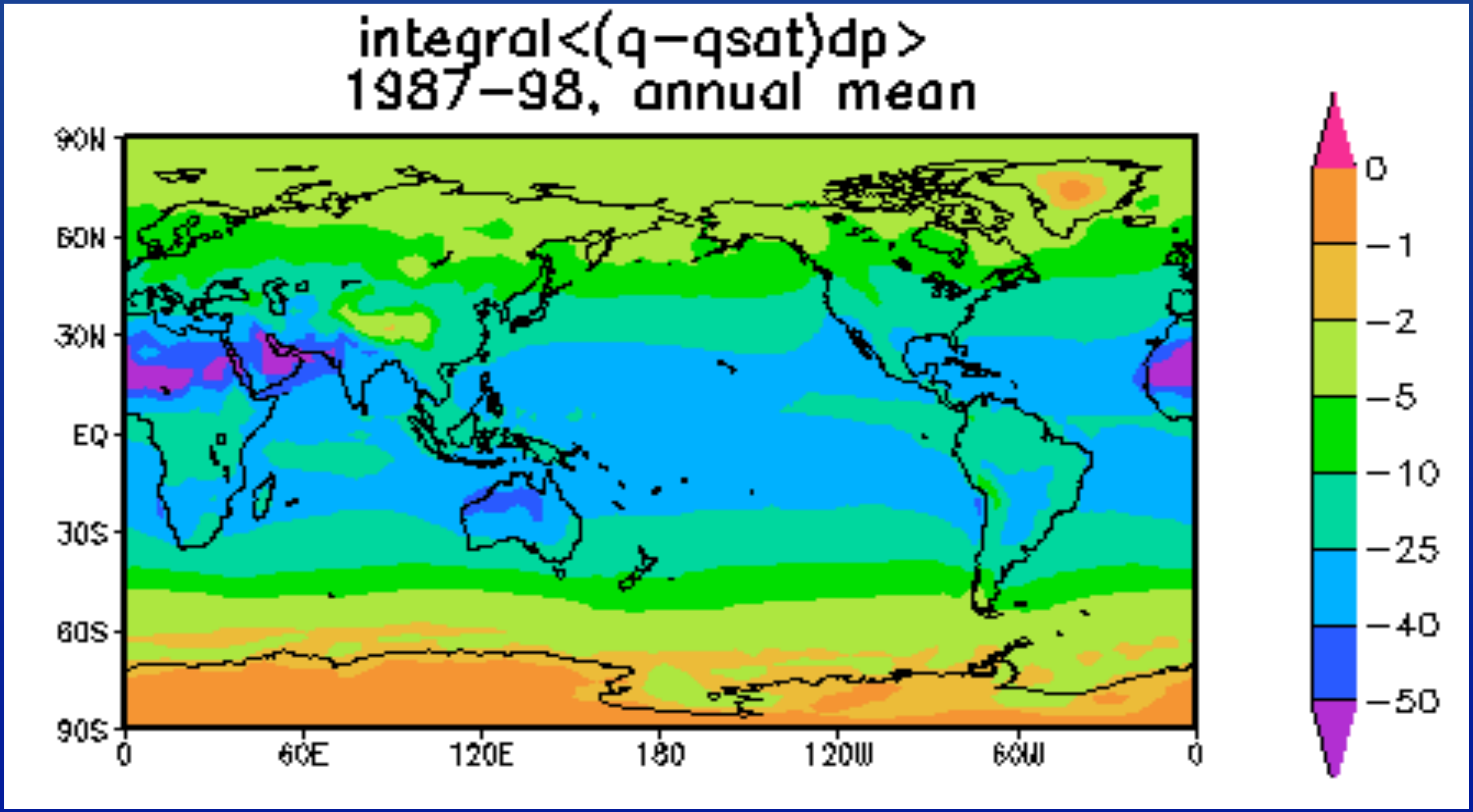
- "Dry side" is the vertical integral of saturation deficit
  - Negative-definite
  - Highest humidity represented by values approaching zero
  - Units of depth, like precipitation
  - Dry end is asymptotic
- Integrated quantity could also be $-(1-RH)*q_{vsat}$
- Benefits of PP:
  - Continuous in space and time (therefore can be processed like any other variable)
  - Normal precip. is preserved (only 'dry side' is modified)

Dry side of PP is the area between the red and green traces, multiplied by -1

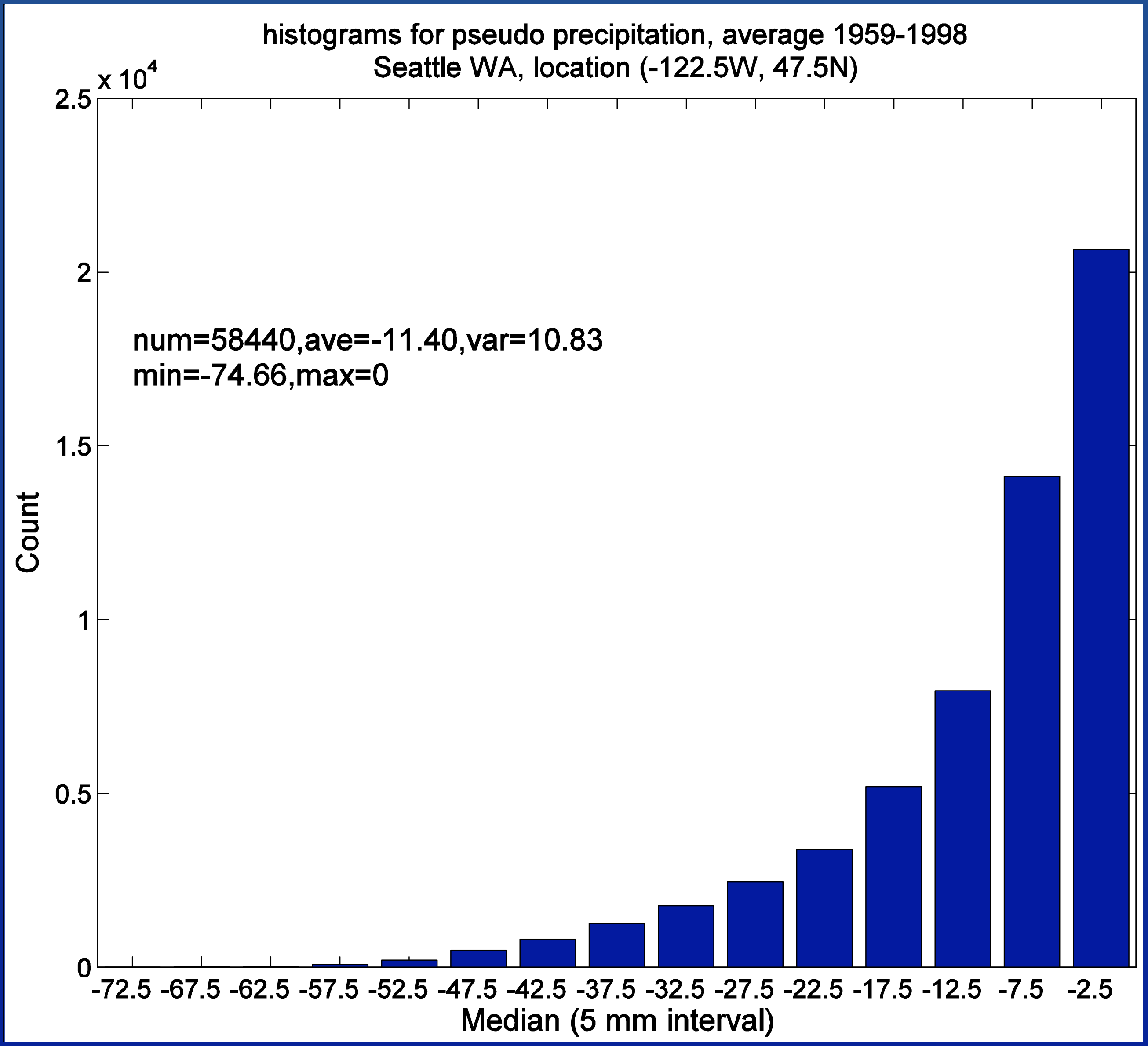


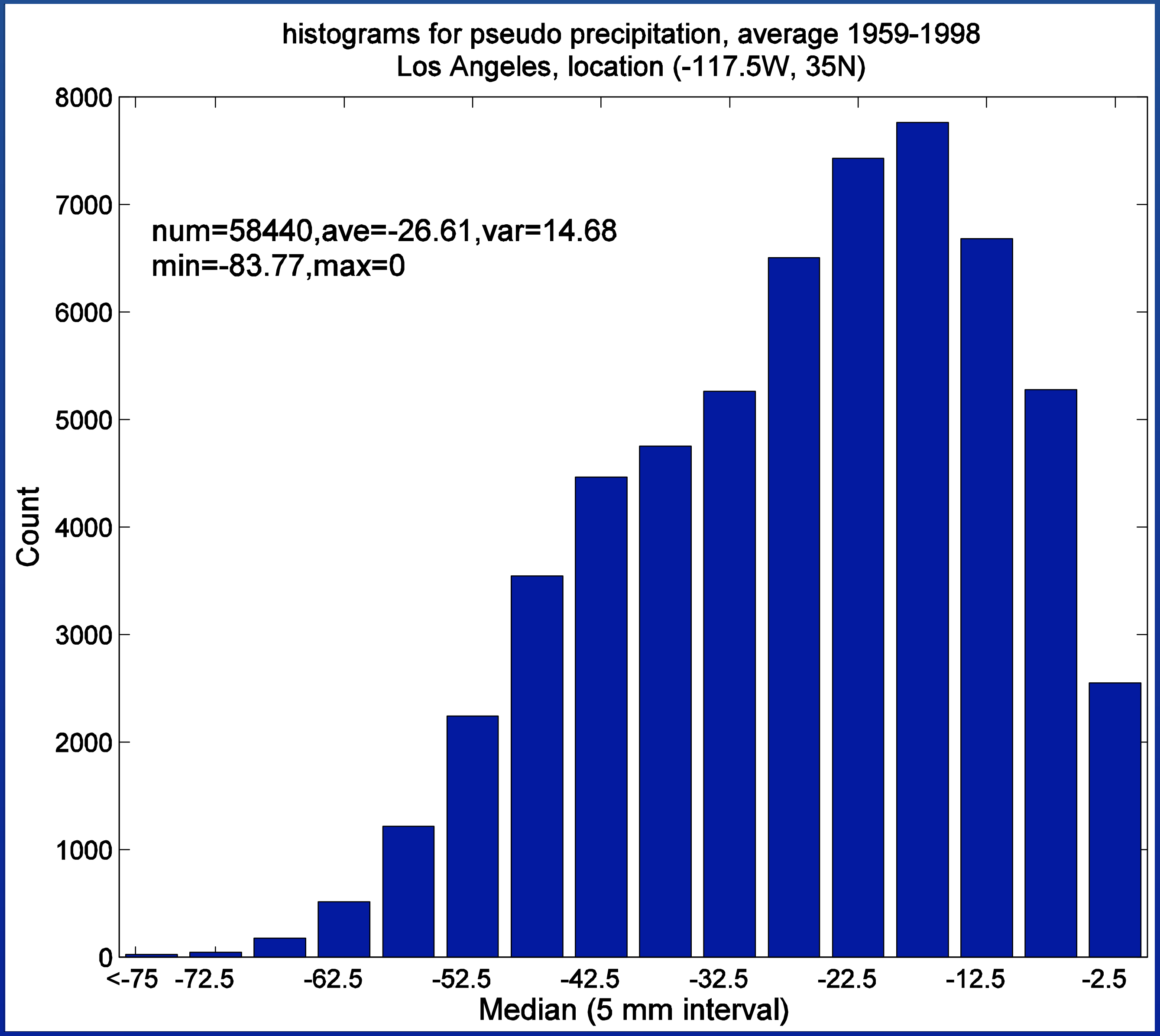
Example of dry part of PP

- Based on NCAR-NCEP reanalysis
- 1460 or 1465 samples per year
- Each pixel is an average of 17532 samples over 12 years

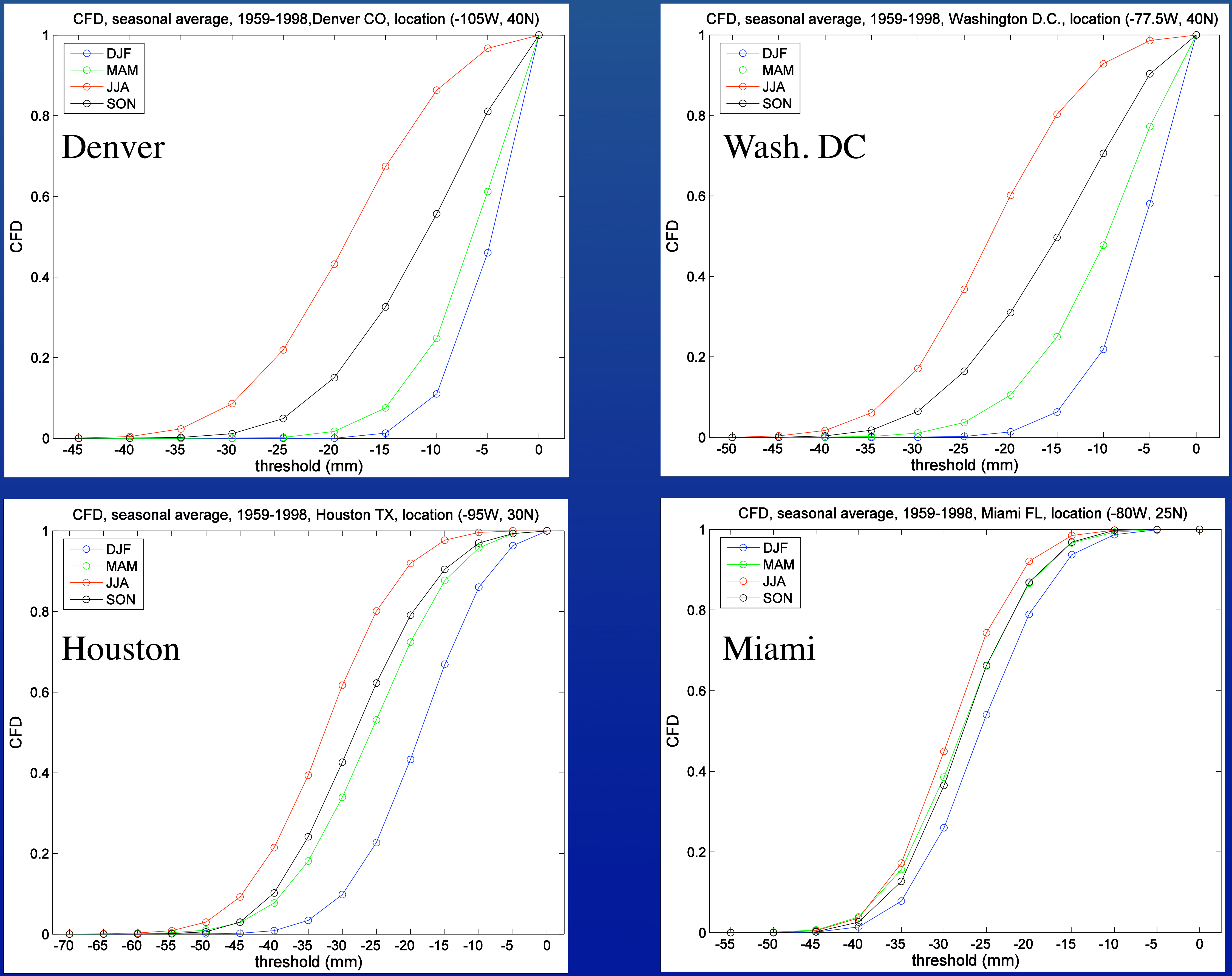


Histograms – Seattle and LA





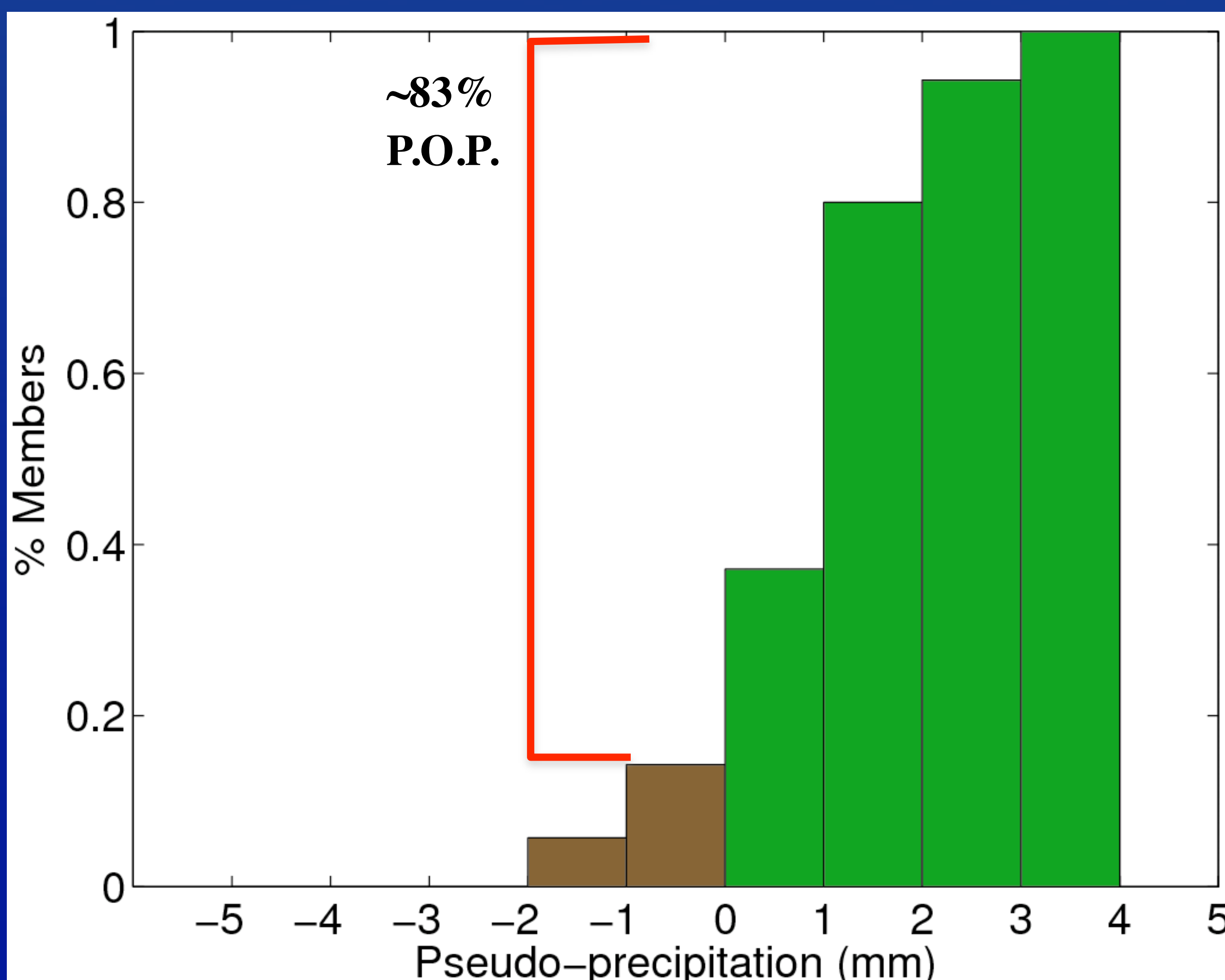
CDFs of Dry PP - Seasonal Means



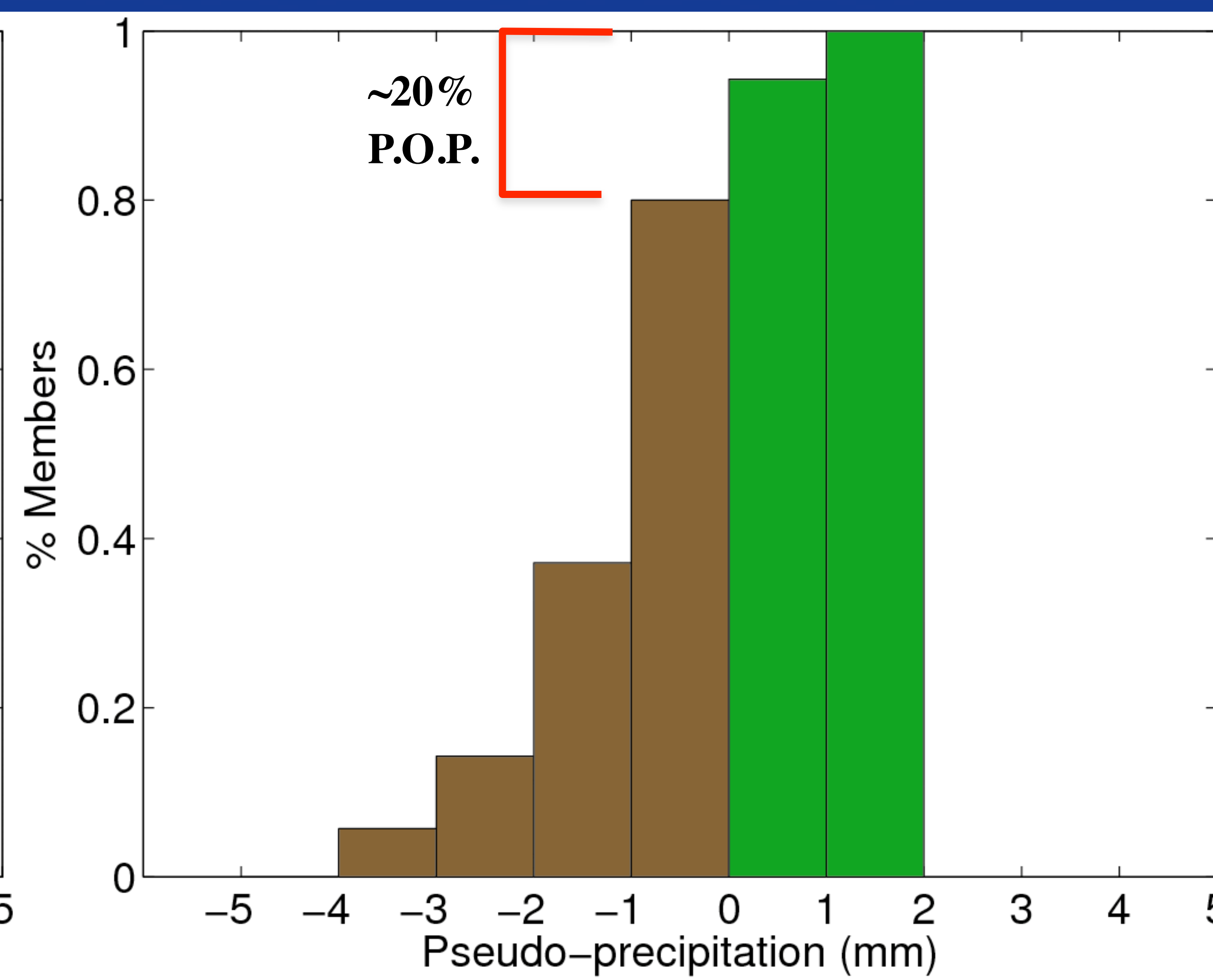
Application of Pseudo-precipitation

- With a single continuous distribution, can bias correct in one step
- Shift entire ensemble distribution to remove bias in PP
- For example, some members constantly produce drizzle (doesn't verify)… Bias correction shifts the PP distribution to the left, forcing some of those members into the negative PP (stopping the members from precipitating)

Raw Ensemble CDF



Bias Corrected Ensemble CDF



Application of Pseudo-precipitation

- Another benefit of PP is that the distribution naturally implies P.O.P.
- If there are many members with larger negative values of PP, this implies a greater probability that the atmosphere is not supportive of precipitation (negative portion of PP indicates moisture deficit)
- Figure: even though case a & b have the same % members > 0, case b has most dry members at a smaller negative value, implying higher P.O.P.