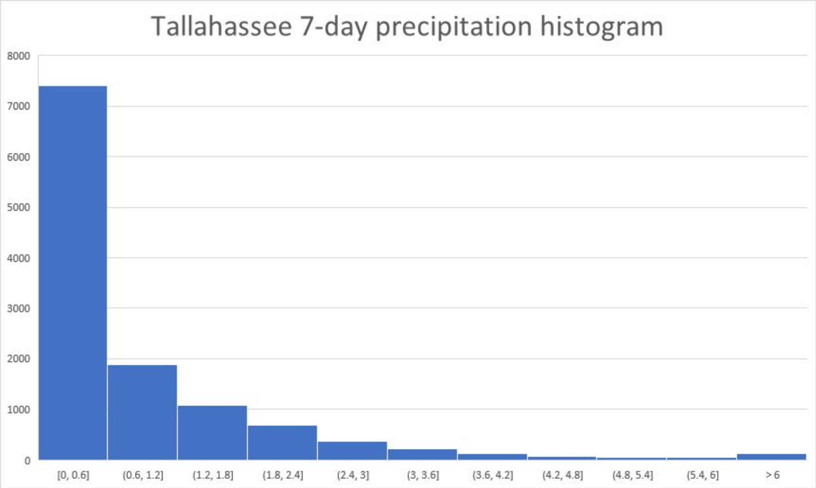
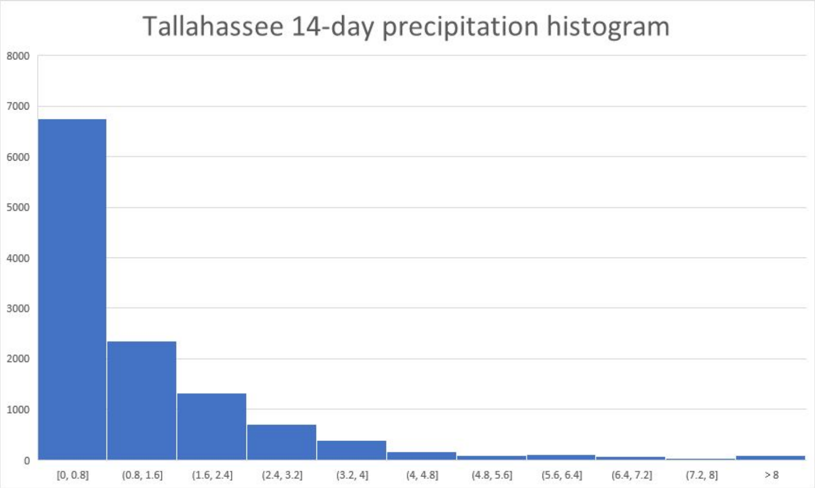
PRECIPITATION DISTRIBUTIONS ARE NOT NORMAL; CAN WE MAKE THEM LOOK THE PART?

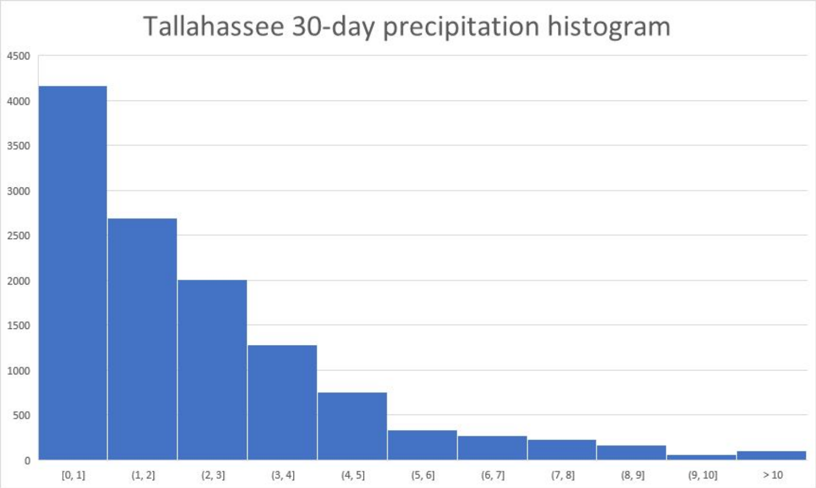
MICHAEL GOSS, EMERSON LAJOIE, CORY BAGGETT, ERICA BURROWS

WITH ADDITIONAL INPUT FROM: JOHNNA INFANTI, LAURA CIASTO, MATT ROSENCRANS, DAN COLLINS, LEIGH ZHANG

THE PROBLEM

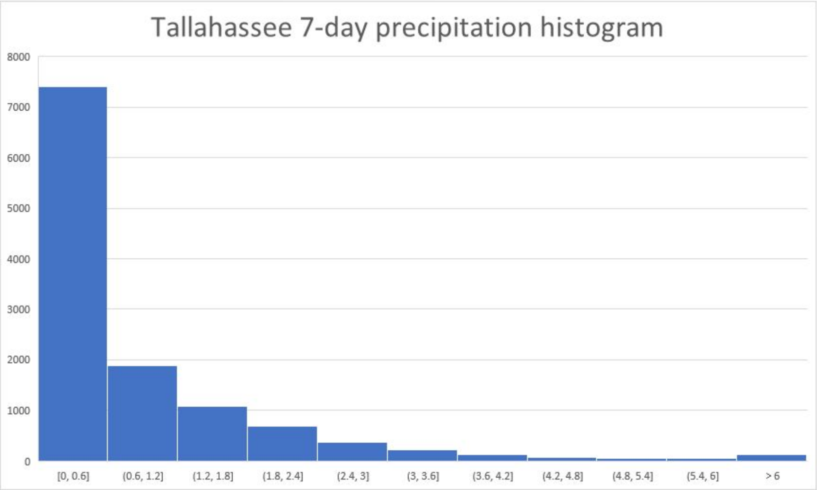


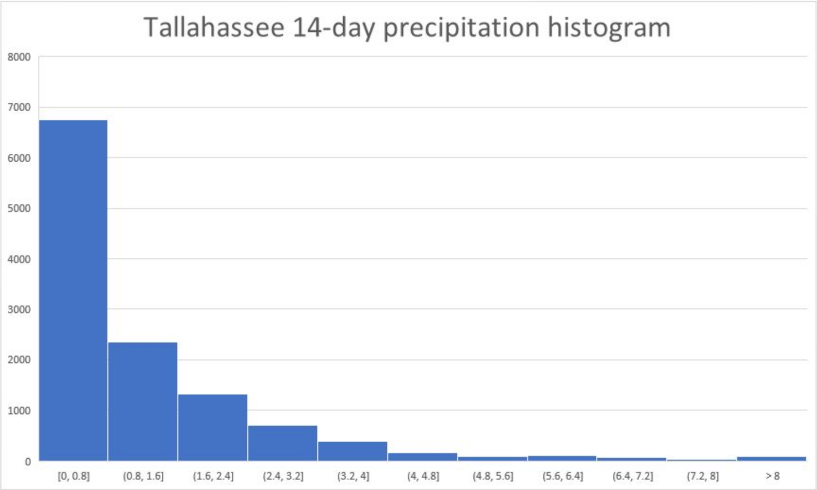


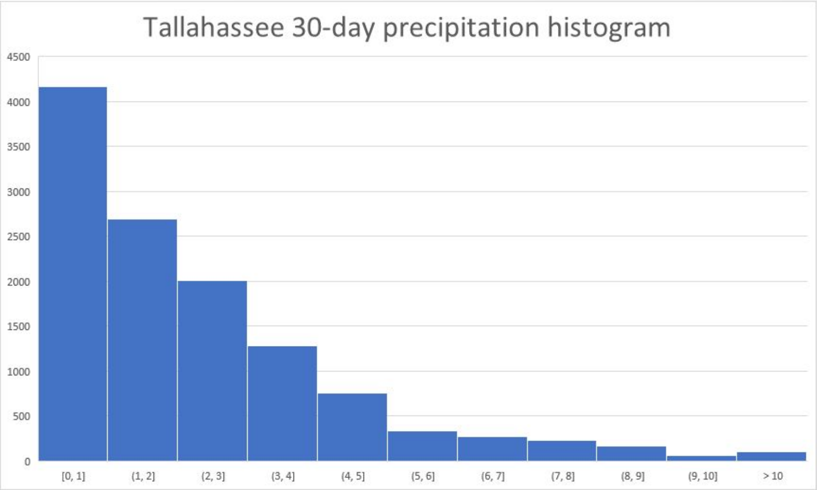


- Precipitation distributions are rarely close to normal/Gaussian, especially for:
  - Shorter integration periods
  - Drier regions

THE PROBLEM

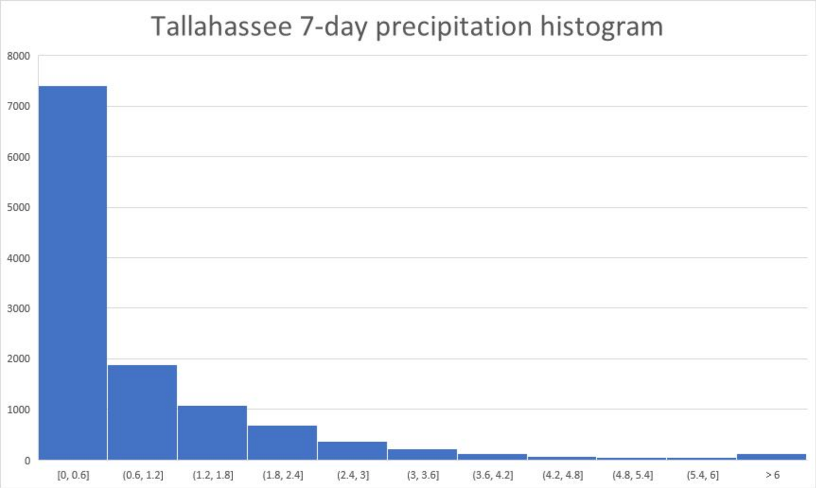


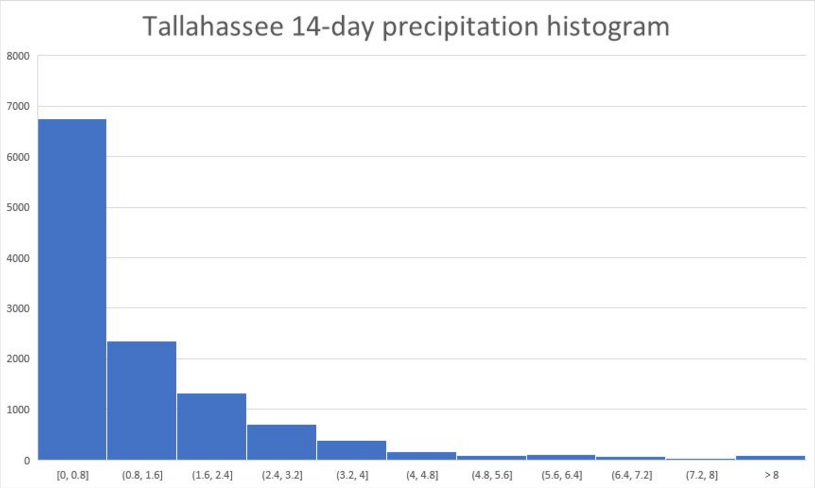


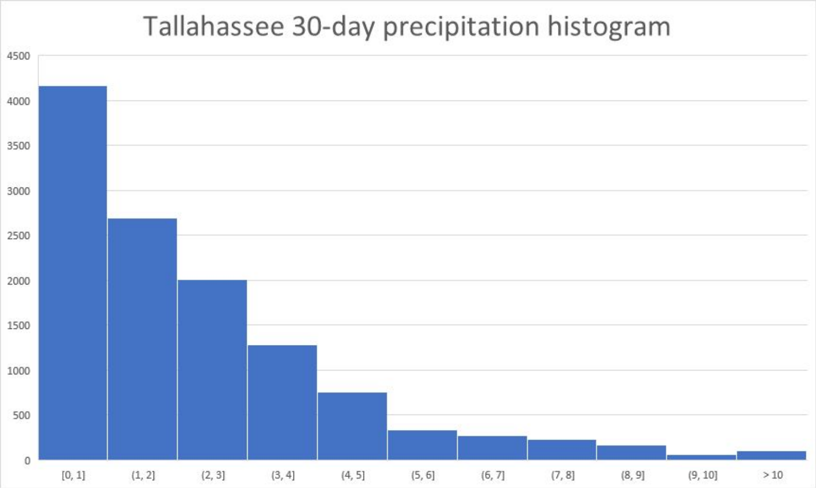


- Normally distributed data helps improve:
  - Reliability; especially in the context of two-category/three-category forecasting
  - Proper fitting in models/regressions that expect Gaussian data (i.e. some MLR code we use at CPC)

THE SOLUTION?

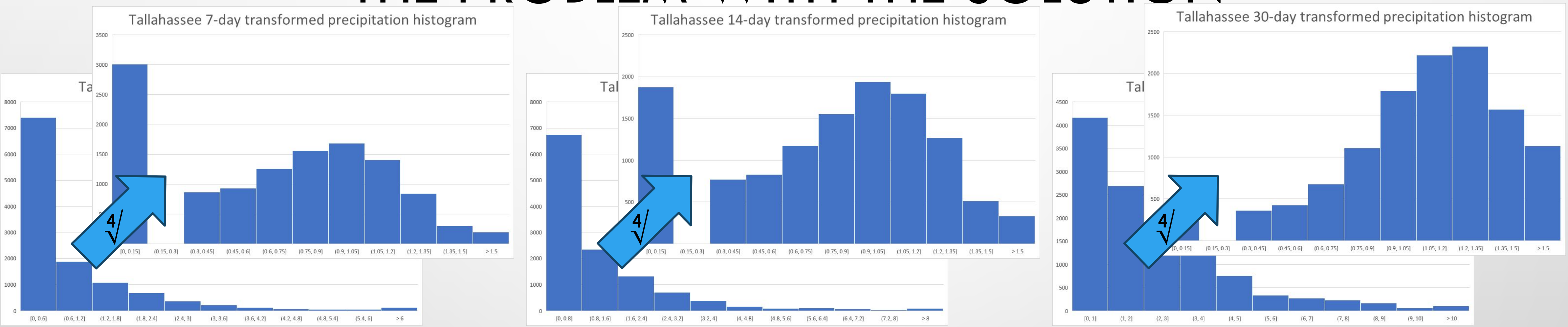






- One method of addressing this issue is to use one of many standard functions to transform the distribution
- Commonly, CPC uses 4th-root, square-root, or logarithmic transformations with precipitation data
  - Truly, probably a gamma distribution is “best”

THE PROBLEM WITH THE SOLUTION



- The "best" transform function for one situation may be quite poor for another
  - Transformations can be quite sensitive to integration period, grid point, and season

WHAT WE WANT

- Transformations that give more normally distributed precipitation data as a function of:
  - Integration period
  - Grid point
  - Season
- “Best” transformation parameters that don’t take too long to derive or calculate

WHAT WE DID

- Box-Cox transformation
  - One parameter
  - Parameter (λ) values of 0, ¼, ½ correspond with already-used log, 4th-root, and square-root transforms

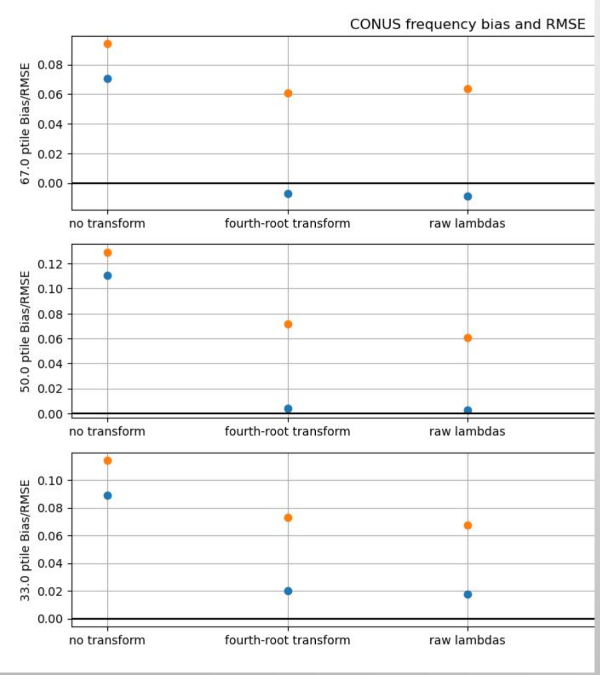
$$\psi(y,\lambda) = \begin{cases} \dfrac{(y+1)^{\lambda}-1}{\lambda} & y \geq 0 \text{ and } \lambda \neq 0, \\ \log(y+1) & y \geq 0 \text{ and } \lambda = 0, \\ -\dfrac{(-y+1)^{2-\lambda}-1}{2-\lambda} & y < 0 \text{ and } \lambda \neq 2, \\ -\log(-y+1) & y < 0, \lambda = 2. \end{cases}$$

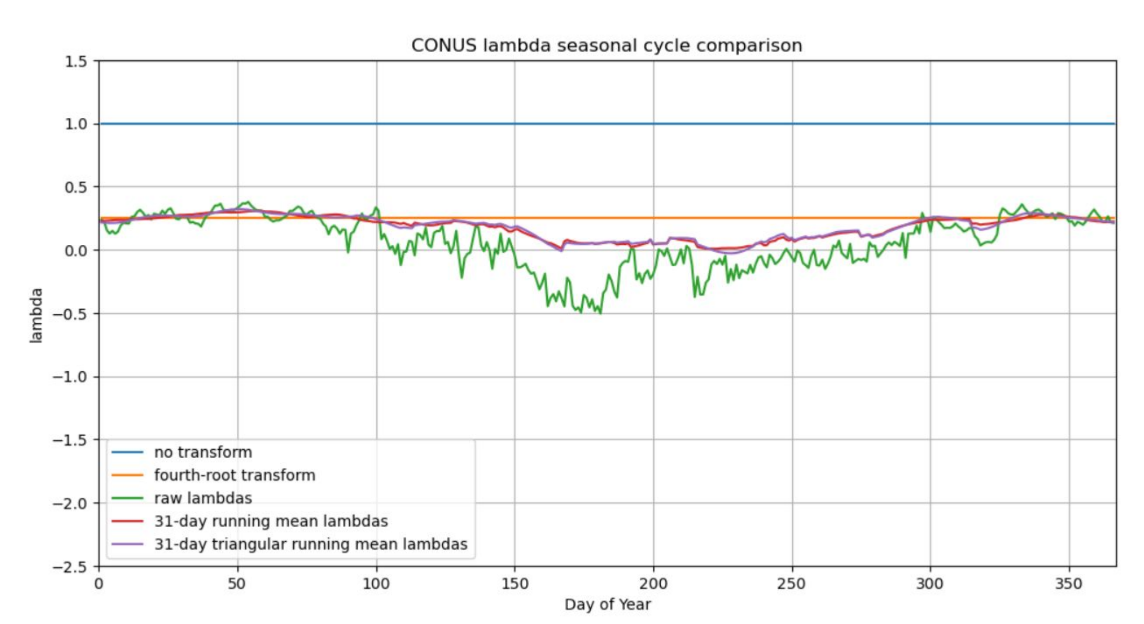
Yeo and Johnson (2000); Also Wilks' textbook

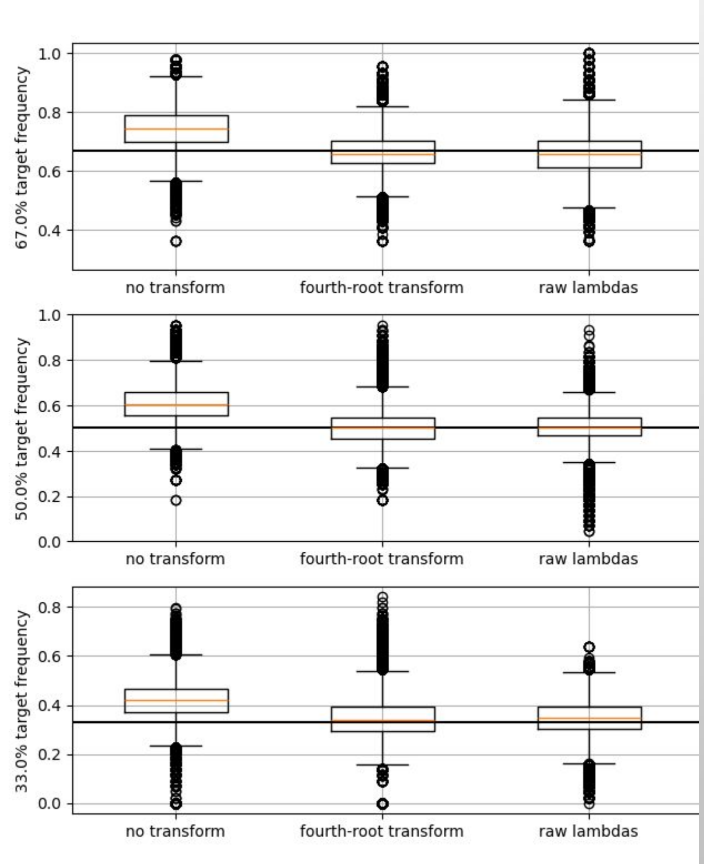
- λ values that minimize the d_{λ} statistic (Hinkley, 1977)
  - Essentially, the normalized difference between the mean and the median

WHAT WE FOUND

- For CONUS as a whole, the 4th-root transformation is indeed a pretty good approximation for 14-day precipitation

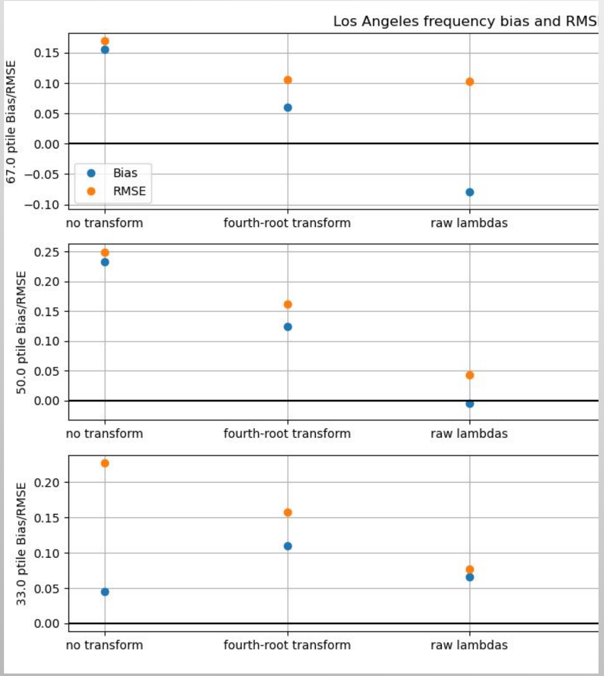


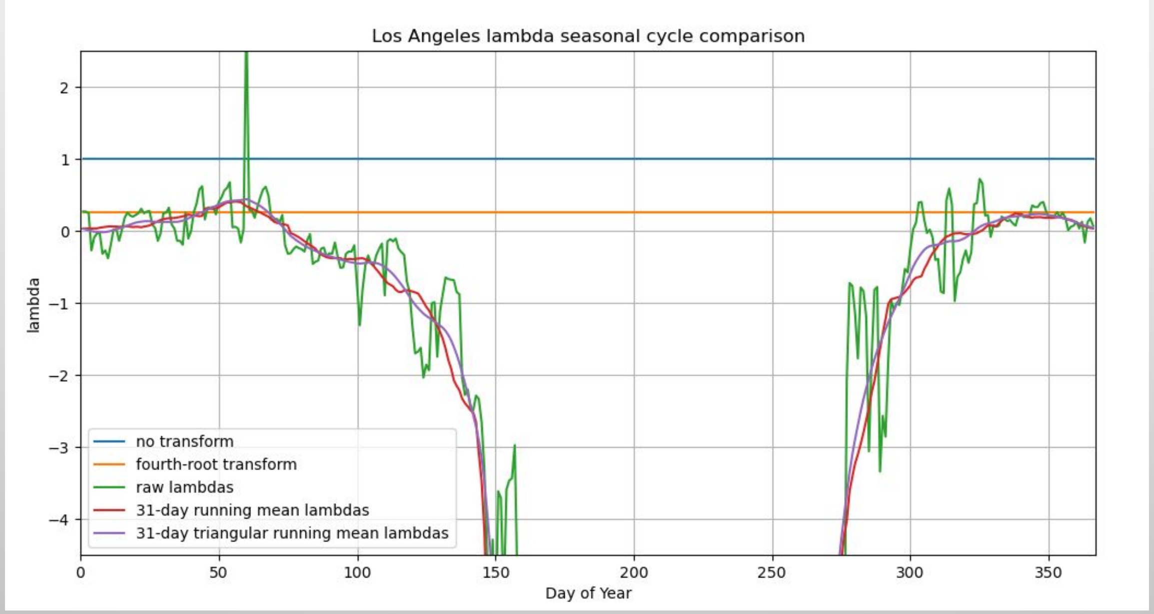


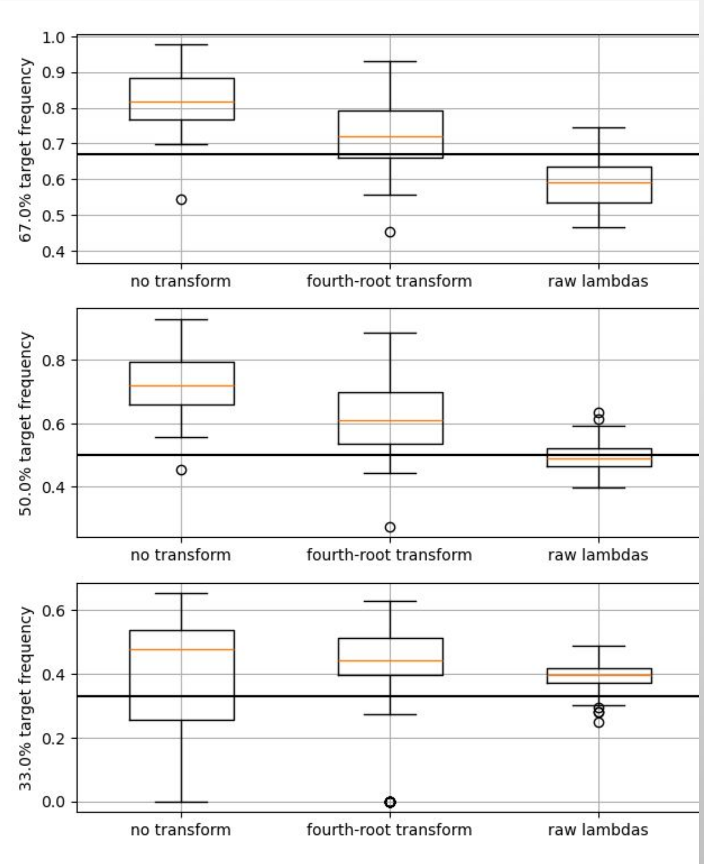


WHAT WE FOUND

- However, the best λ varies substantially by integration period (not shown), grid point, and season

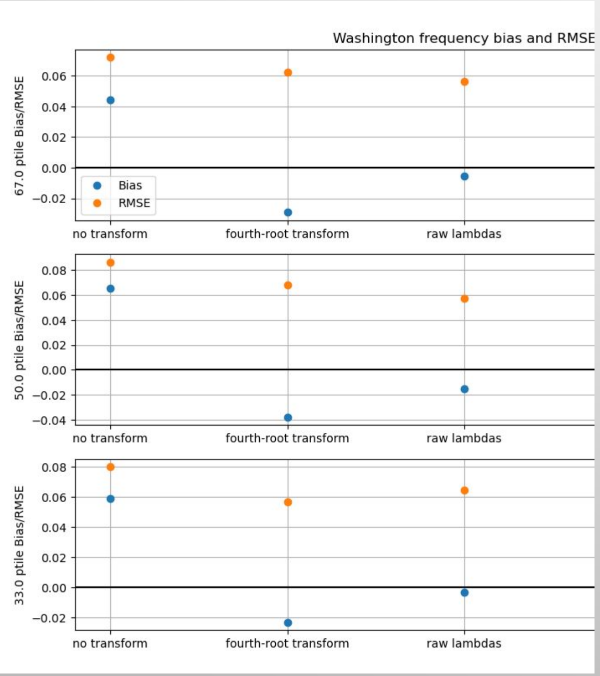


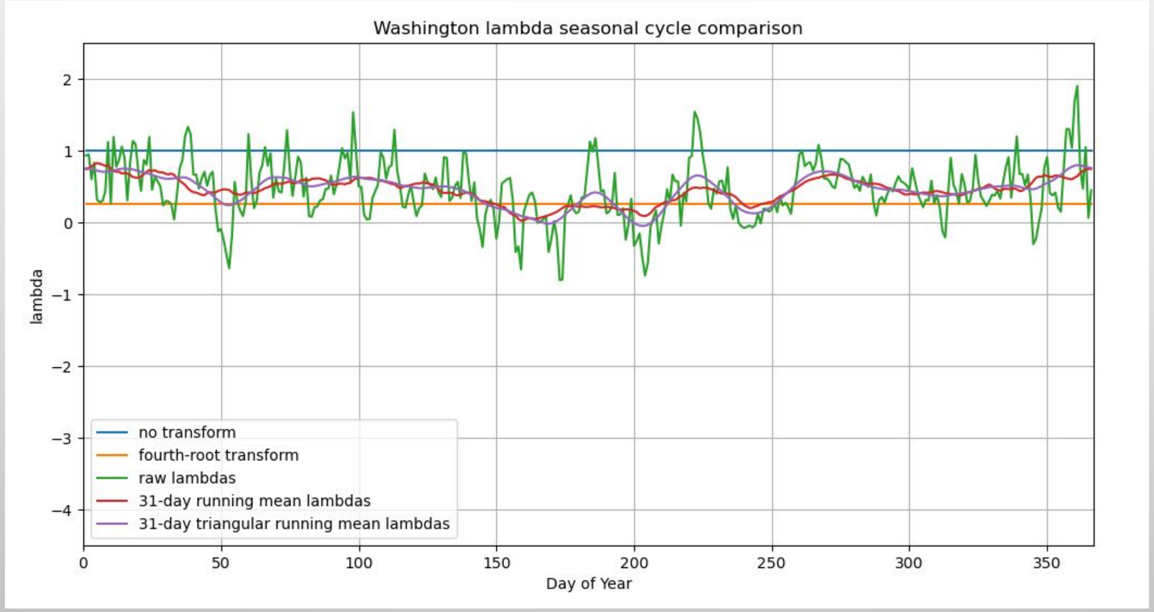


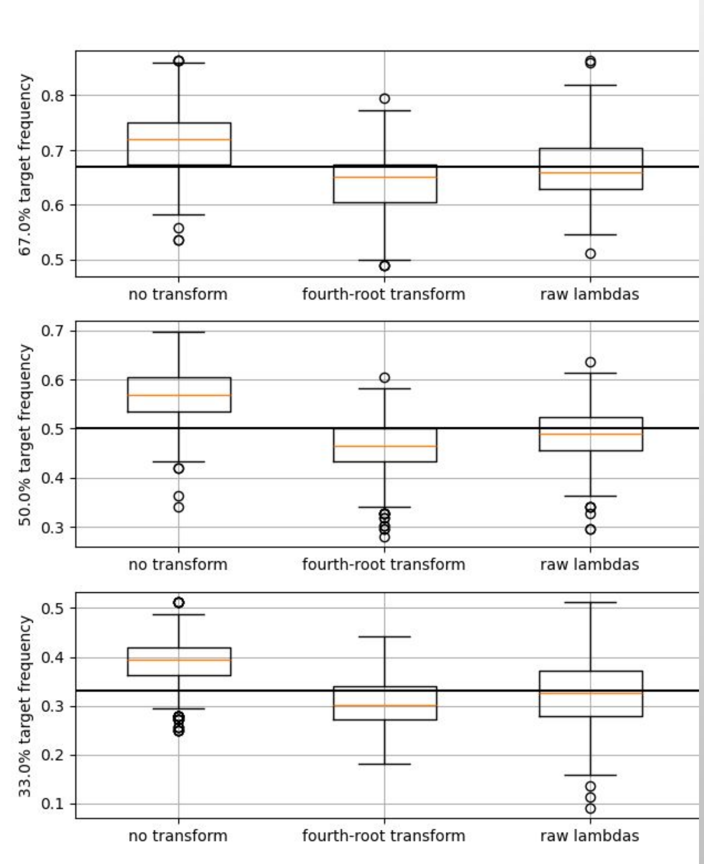


WHAT WE FOUND

- However, the best λ varies substantially by integration period (not shown), grid point, and season

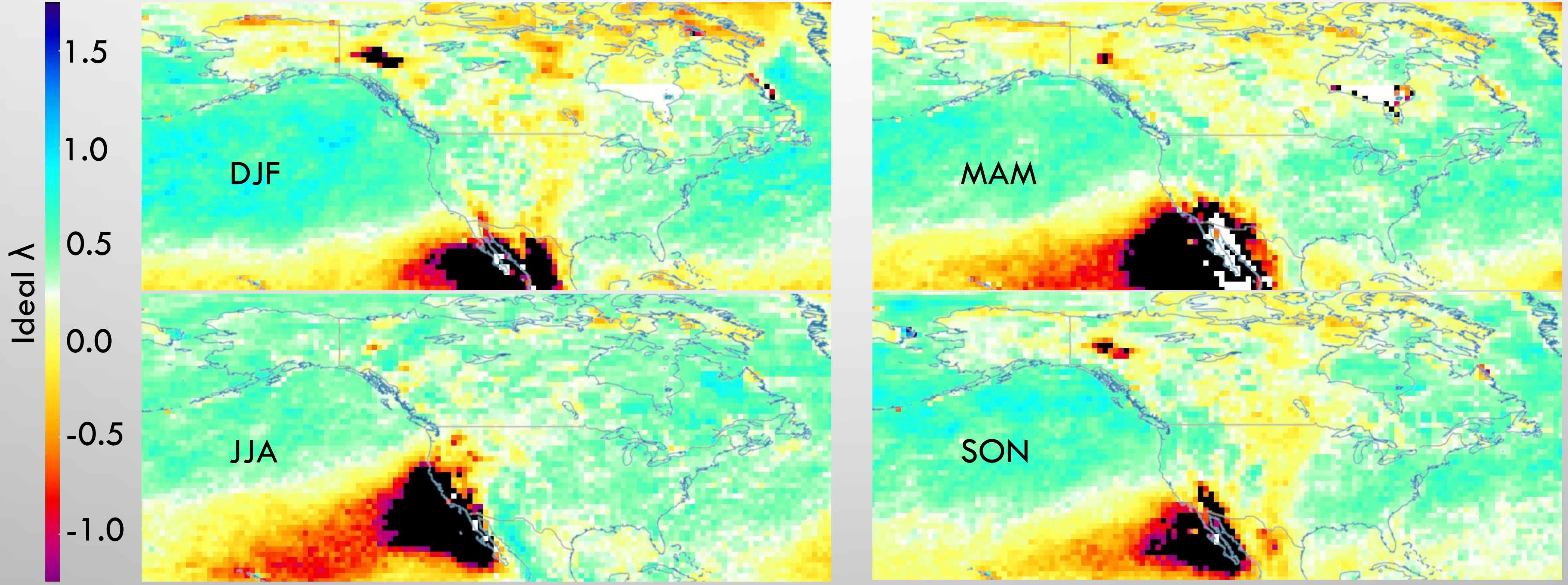






WHAT WE FOUND

- However, the best λ varies substantially by integration period (not shown), grid point, and season



WHAT WE'RE DOING

- Now that we have λ as a function of integration period, grid point, and day of year (and python functions to transform and inverse-transform using a map of lambdas!)…
  - Emerson is currently testing some MLR code with the newly transformed precipitation data to test whether or not categorical forecast skill can be improved
  - Preliminary results are equivocal, but more testing needed!

THANK YOU!

- MICHAEL.GOSS@NOAA.GOV