New calibration methods for extreme precipitation probabilities in subseasonal-to-seasonal forecast models

Chiara Lepore, Michael K. Tippett

Lamont-Doherty Earth Observatory
Columbia University | Earth Institute

Michelle L'Heureux, Melissa Ou, Laura Ciasto

OWAQ -> S2S Research
NA19OAR4590159

Summary

- **Provide a model based guidance for week-2 precipitation extremes in the US (Probabilistic Extremes Forecast Tool)**
  - **3-day accumulated precipitation (8-10, 10-12, 12-14) above 85th**
- **Current calibration approach:**
  - **Ensemble regression method**
  - **Does not vary from one forecast to another**
  - **one method to calibrate the whole pdf of the process (i.e. precip is heavy tailed)**
- **Proposed calibration approach:**
  - **Focused on extremes (i.e. week-2 probability of 3-day accumulated precip above 85th)**
  - **Includes informations about the state of the forecast/process (i.e. precip, env. variables)**

Data

Probability 3-day accumulated precipitation (L= 8-10, 10-12, 12-14) above 85th

- **Forecast model:**
  - **SubX GEFS v.11 hindcast (1999-2016) [real time forecast for 2017-2019]**
    - **Hindcast are daily data (00z-00z) - 10 ens. mem. - 1 start once a week**
- **Reanalysis data (match in time and spatial grid):**
  - **NARR - 3 hourly data- aggregated to 00z - 00z**
  - **CPC Unified - daily data- 12z-12z and do not match forecast 00z-00z**
  - **PERSIANN - daily data- 00z-00z**

Roadmap

- Extract climatologies ($I_{3d,\,85th}$) for forecast and reanalysis data
- Extract occurrence of exceedance ($I_{3d} > 85^{th}$) for forecast and reanalysis for a range of lead times: L=8-10 day and NARR
- Evaluate baseline skill of uncalibrated forecasts
- Present and Assess preliminary new calibration methods.

Climatologies for extremes $I_{3d,\ 85^{th}}$

For each grid point and calendar day we extract climatologies for 85^{th}

Forecast model

- **Functional form: annual harmonics, 3 harmonics correspond to 7 params**

$$I = exp[a_0 + \sum_{k}^{3} (a_k \cdot \sin(k \cdot nu \cdot x) + b_k \cdot \sin(k \cdot nu \cdot x)]$$

nu = 2*pi/365.25 , x = day

- **For each calendar day we pool rolling window L=4-33 days**

Reanalysis

- **No Harmonics - Pooling a rolling window of data (-15 to +15 days)**

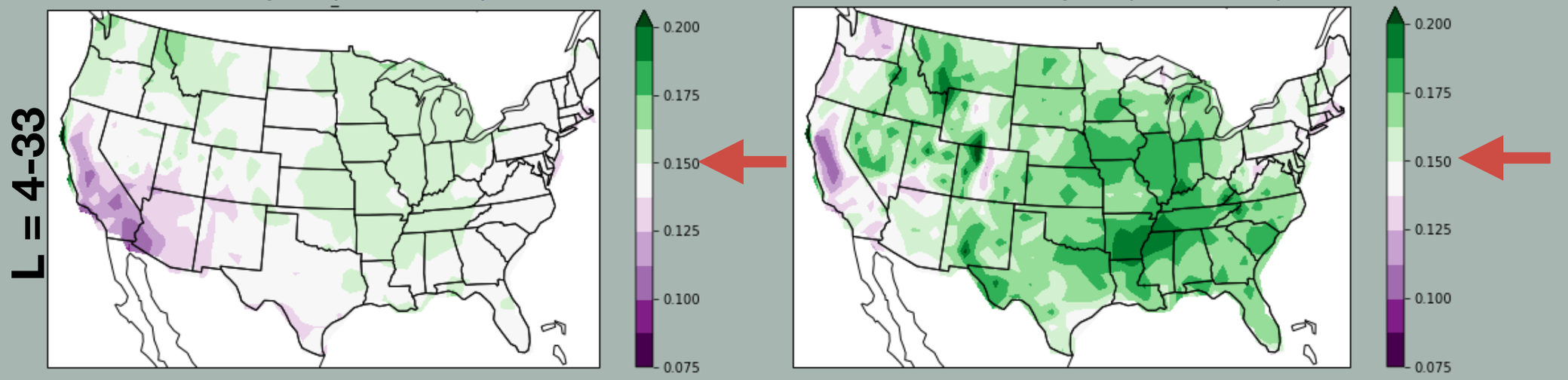
Climatologies for extremes $I_{3d,\,85^{th}}$

Annual mean of frequency of exceedence of climatology (15%)

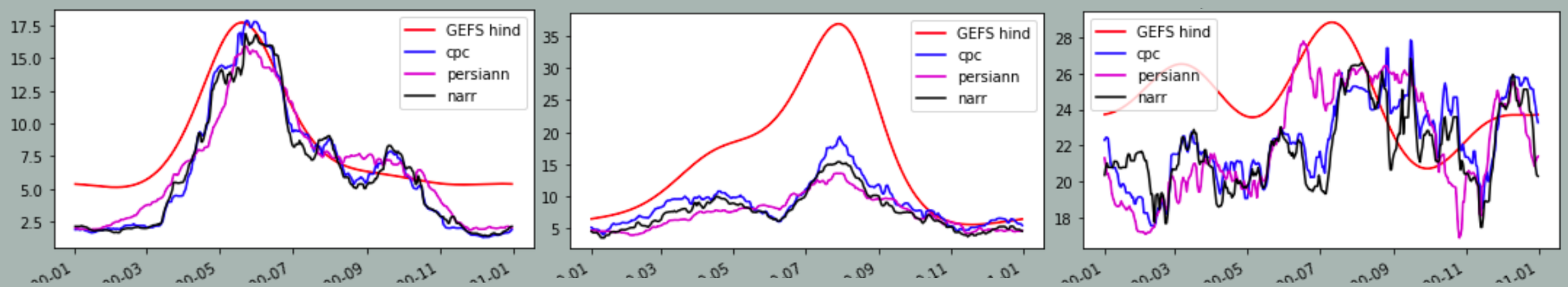
In sample (hindcast)

Out of sample (real time)

L = 4-33

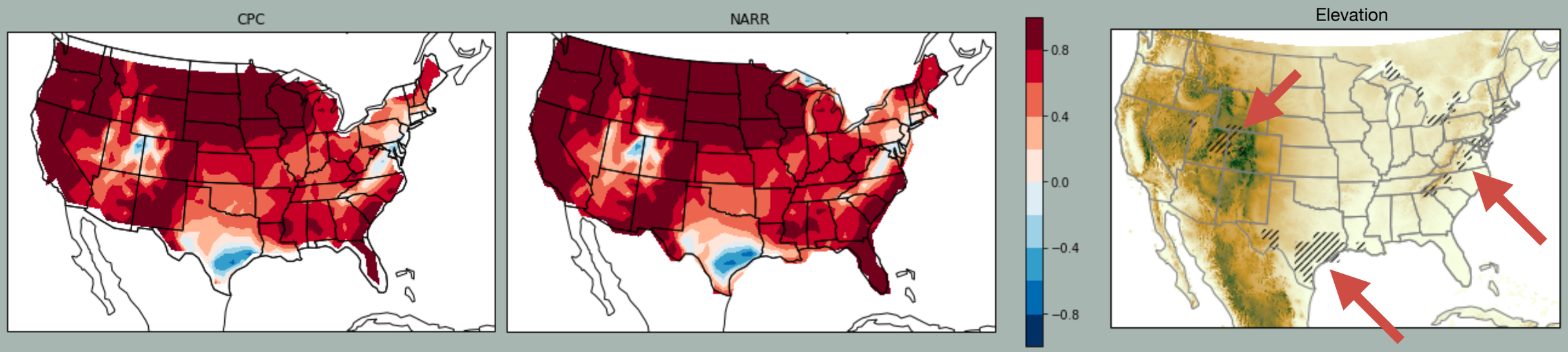


Comparing climatologies: GEFS vs. Reanalysis



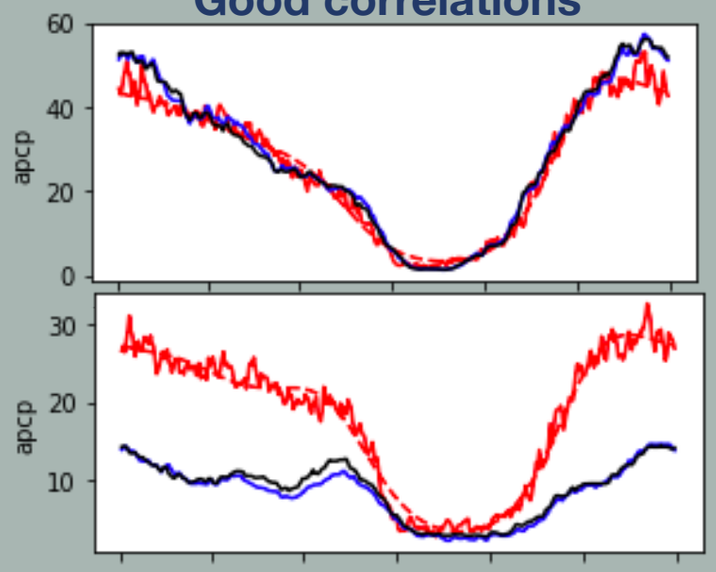
Comparing climatologies: GEFS vs. Reanalysis

Correlations between reanalysis and GEFS climatologies

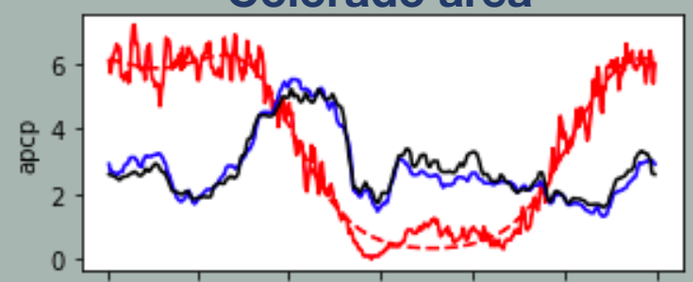


- Reanalysis and GEFS climatologies correlate well except for 3 areas where correlations < 0.
- The Colorado and Eastern blue areas could be due to some orographic effect. Not so for Texas blue area.
- Differences do not depend on time.
- tend to resolve at 50th percentiles for Colorado, not for Texas.

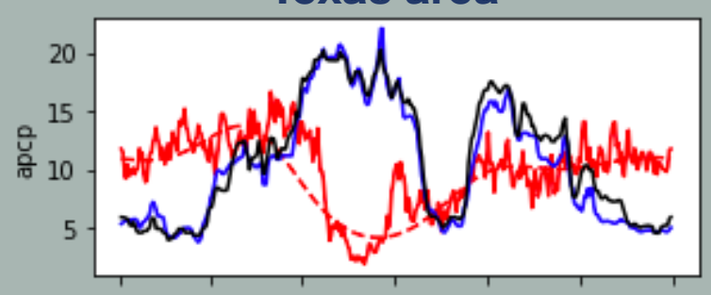
Good correlations



Colorado area



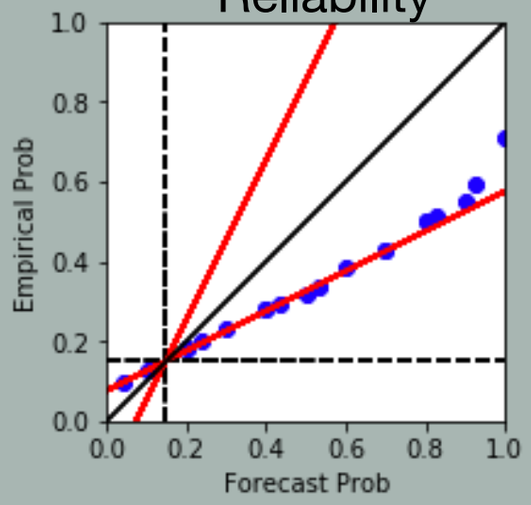
Texas area



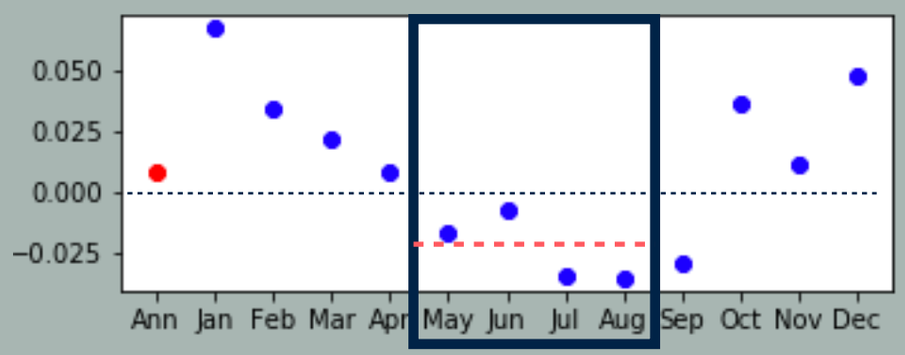
Extract occurrence of exceedance ($I_{3d}>85^{th}$)

Uncalibrated GEFS $I_{3d,\ 85th}$ 8-10 days skill vs. NARR

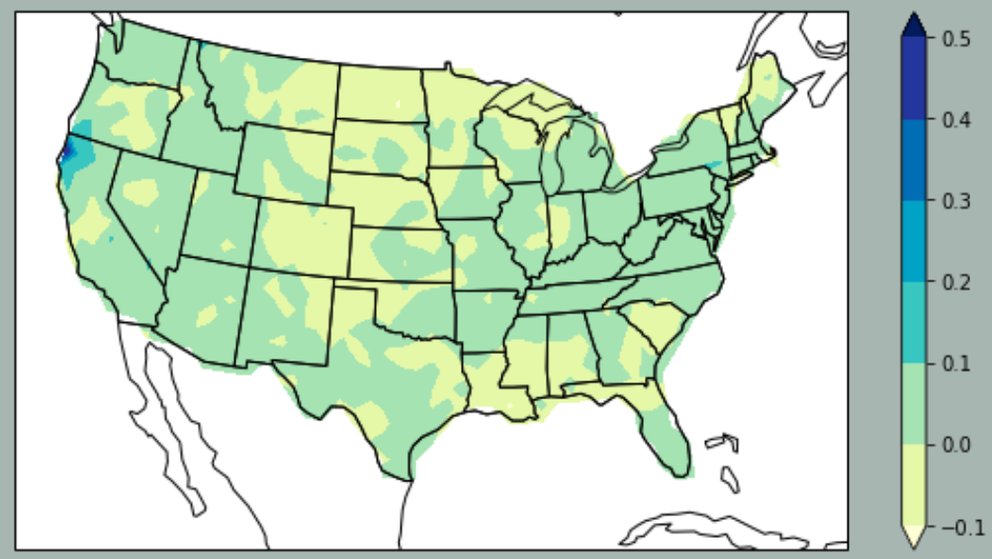
Reliability



Brier skill score (ref: Climatology)

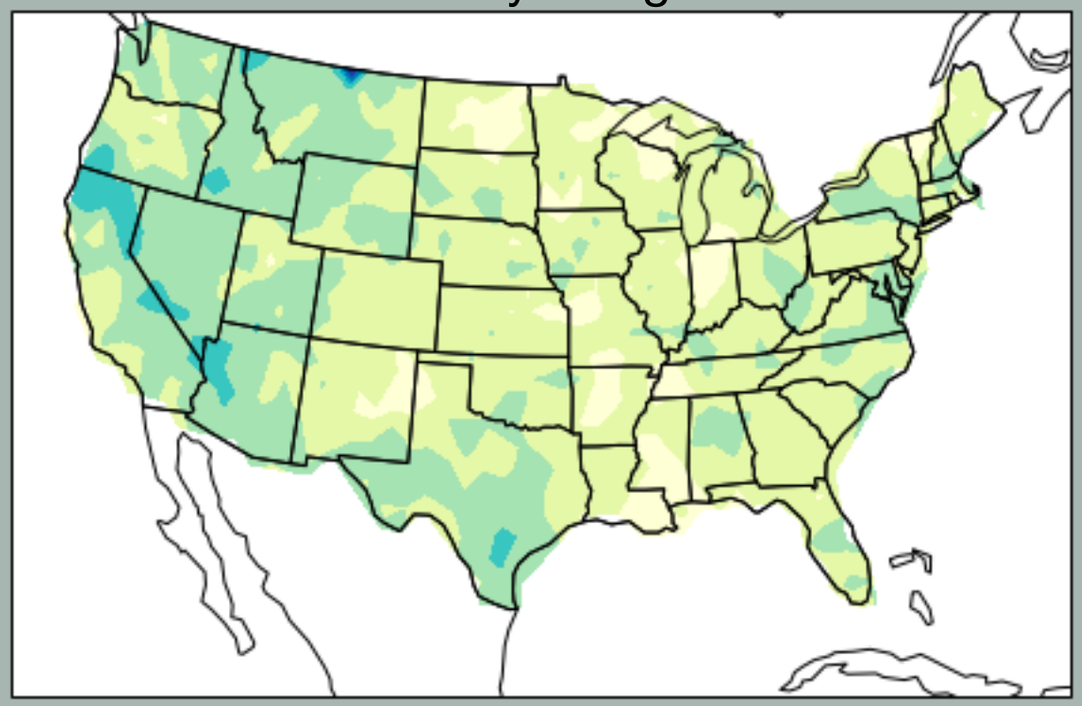


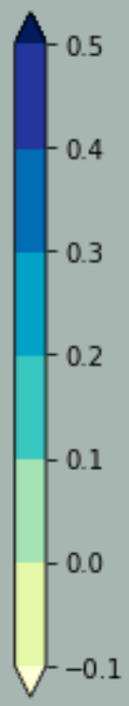
Annual Brier skill score



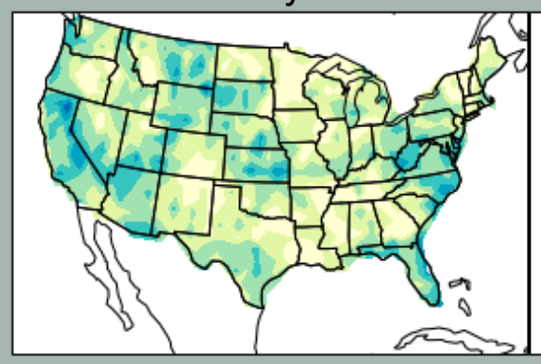
May—August Brier skill score

May—August

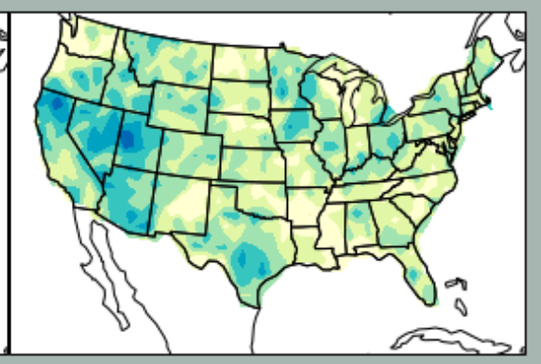




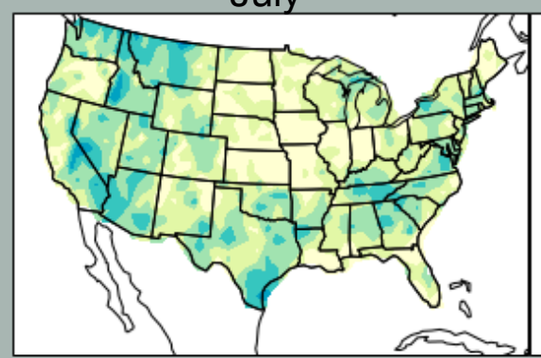
May



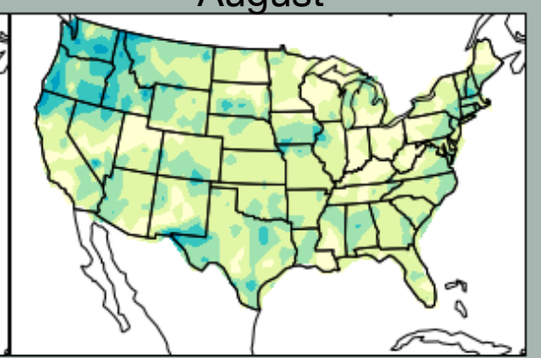
June



July



August



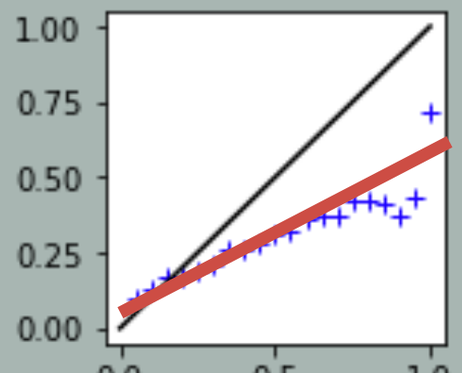
Explored calibration methods

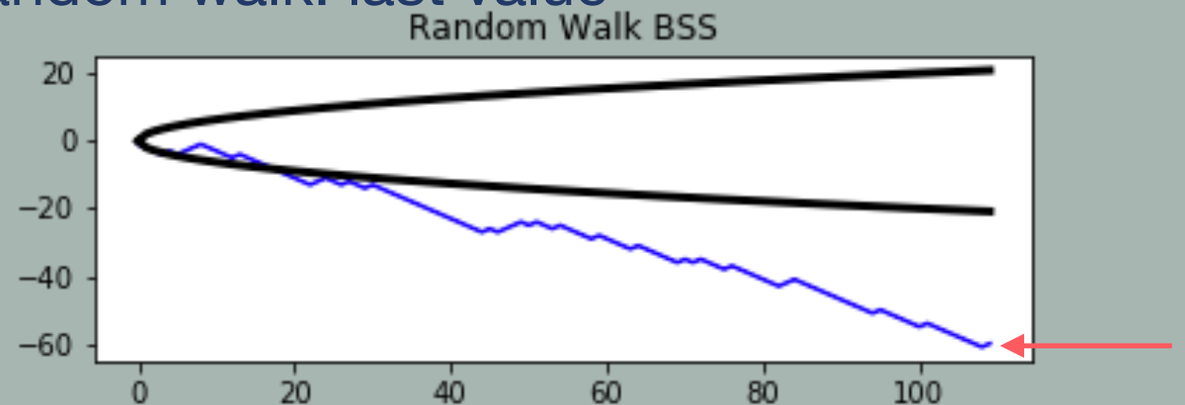
Model set up

- Extract occurrence of exceedance , $I_{3d}>85^{th}$, L=8-10 days
- Regressors: ensemble probabilities, I_{3d}, I_{3d} anomalies, and large scale environments ($CAPE, T, Td, w$): up to 3 params ~ 220 combinations
- Spatially varying multivariate Logistic Regression: $ln\frac{p}{(1-p)} = \beta_0 + \beta_1 x_1 \ldots$
- Each grid point fit pooling data from a 7x7 spatial uniform kernel
- Train on 2/3 of hindcast forecast data - Test on remaining 1/3

Assessment

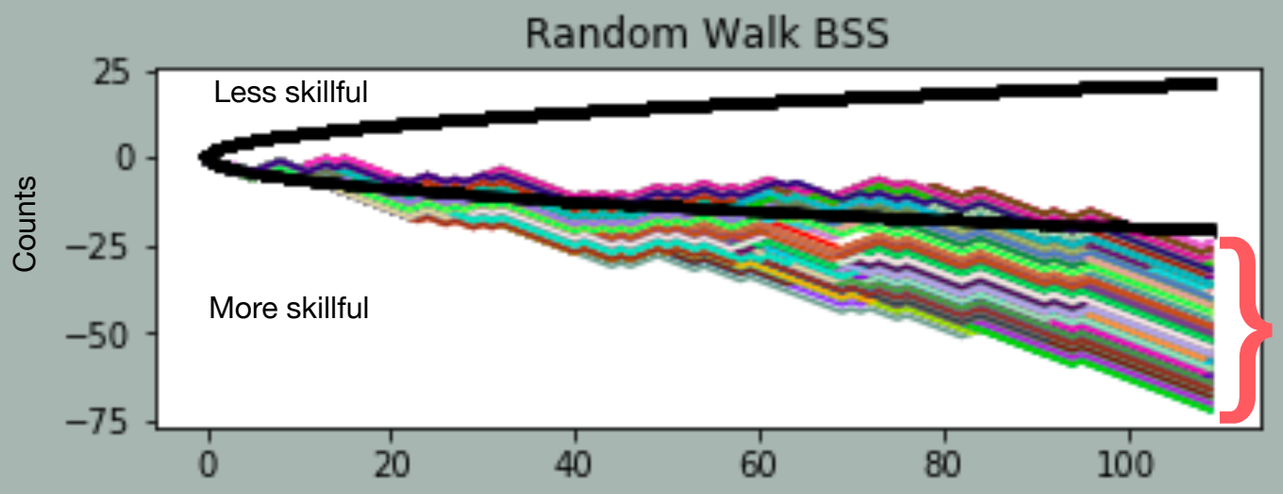
- Look at Reliability, BSS, time varying BSS (random walk).
- Overall Reliability: WLS fit; BSS random walk: last value

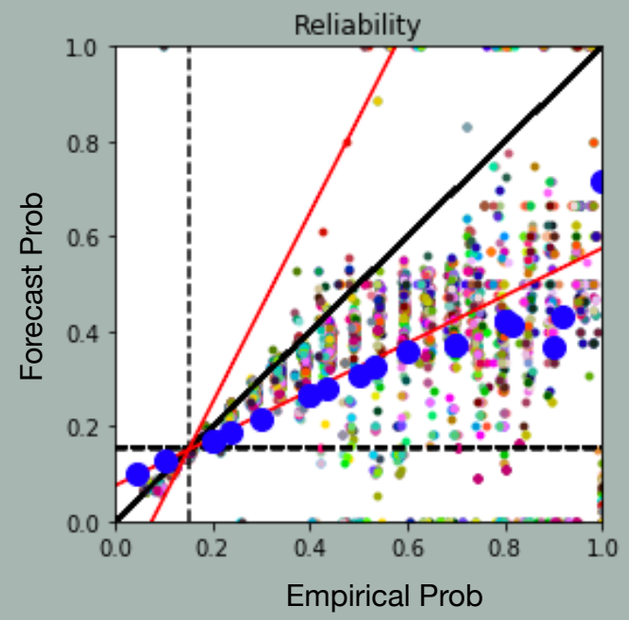




Results

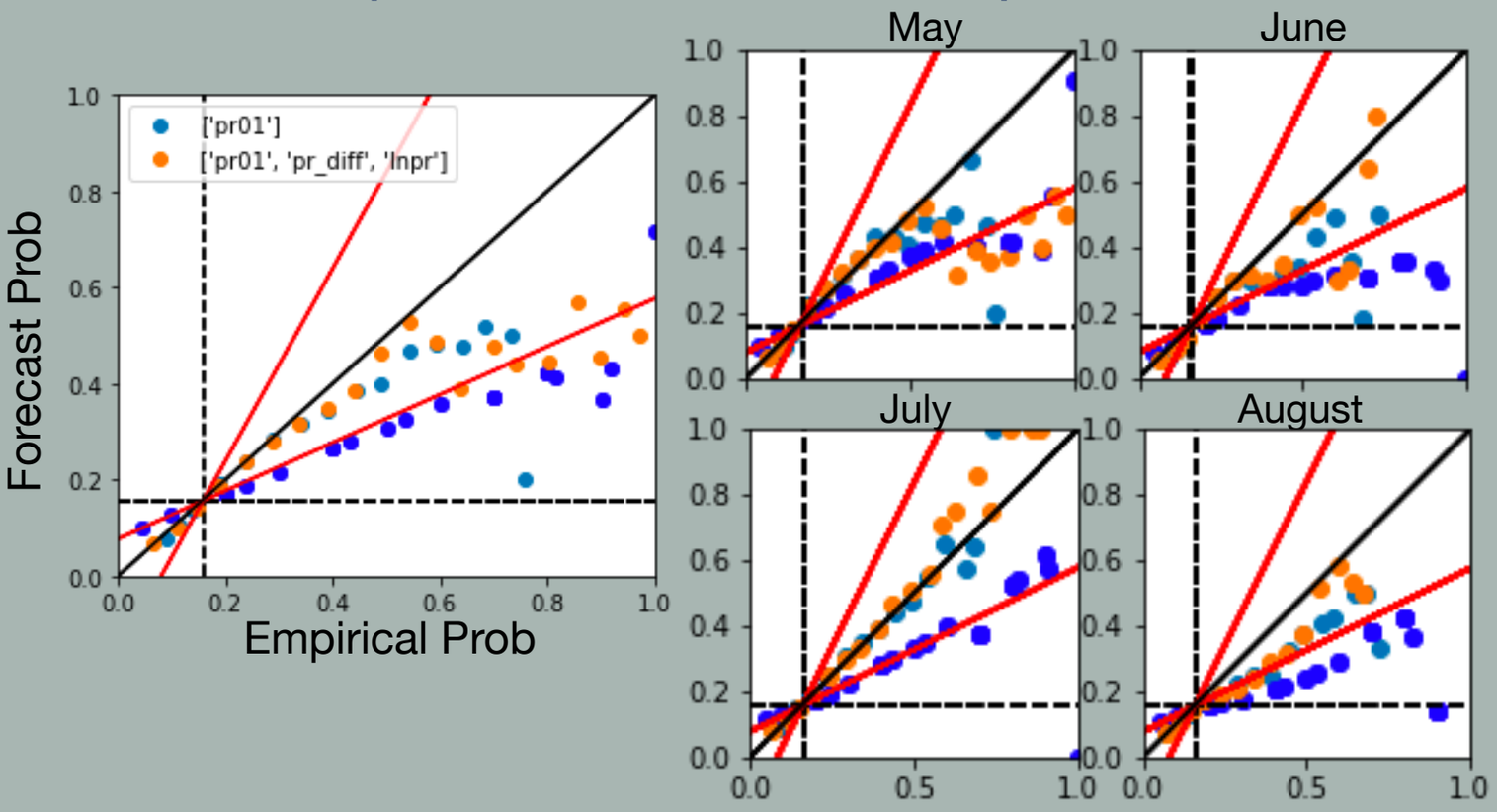
- All combinations (1 to 3 params) result in significantly better model than the raw forecast wrt BSS (all negative total counts in BSS random walk below black line).
- Not all models improve reliability.
- Looking at different metrics (Reliability vs. BSS) identify a small pool of best performing models.

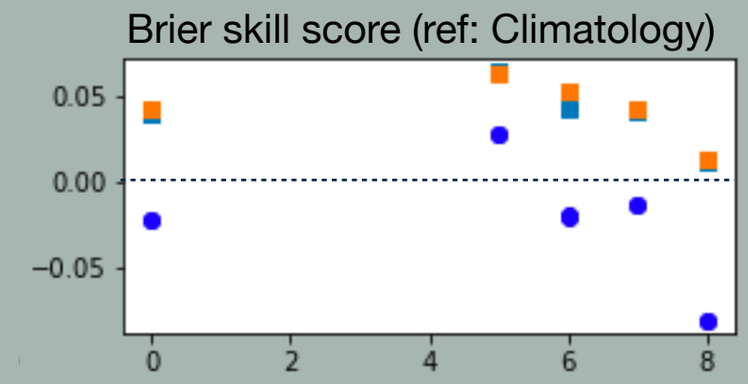




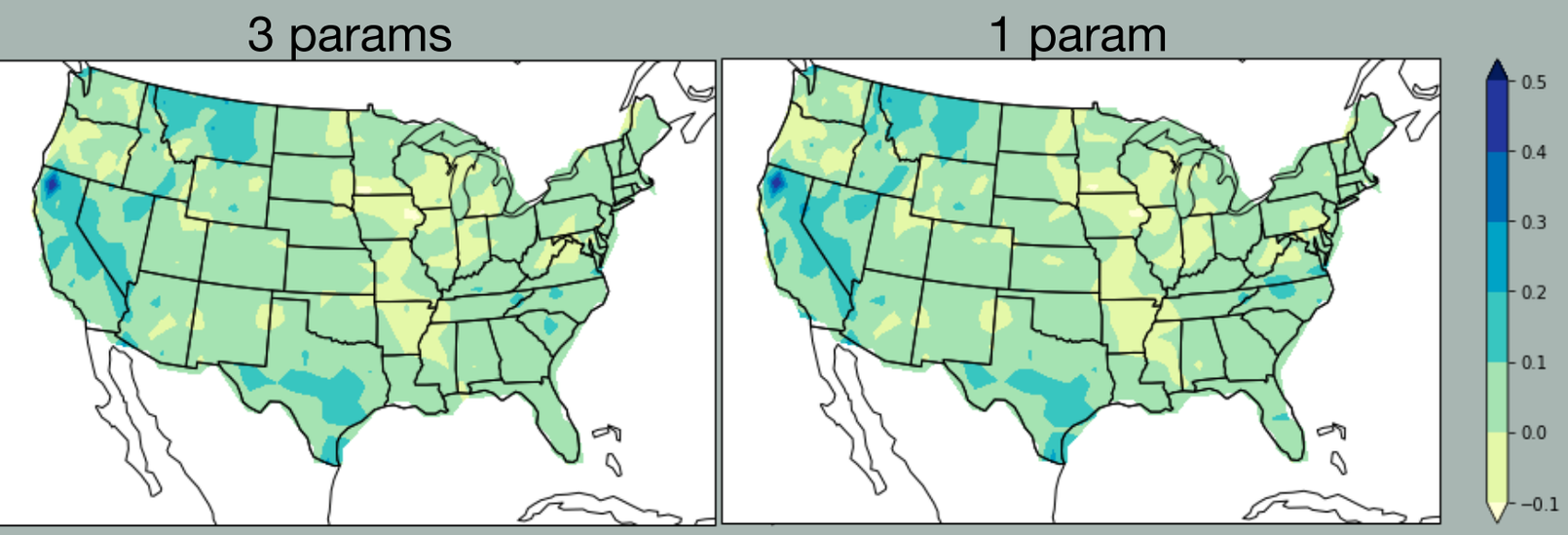
Results

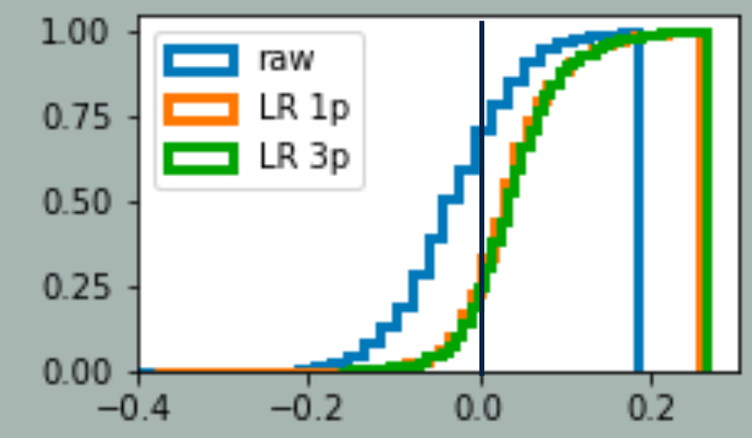
- Best of 3 parameter: ensemble probabilities, anomalies, log-precip
- Best of 1 parameter: ensemble probabilities





May—August Brier skill score





Conclusions

- **Provide a model based guidance for week-2 precipitation extremes in the US (Probabilistic Extremes Forecast Tool)**
  - **3-day accumulated precipitation (8-10, 10-12, 12-14) above 85th**

- **Proposed calibration approach:**
  - **Focused on extremes (i.e. week-2 probability of 3-day accumulated precip above 85th)**
  - **Includes informations about the state of the forecast/process (i.e. precip, env. variables)**

- **Future work:**
  - **Extend to all year, all SubX Models**
  - **Detailed comparison of differences among various models (including ER)**
  - **Extend to GEFS v.12**