Post-Processing of Week 2 GEFSv12 Heat Forecasts via Neural Nets

47th Climate Diagnostics and Prediction Workshop

Session 6, Part 1: Applications of Modern Technologies to S2S Forecasting

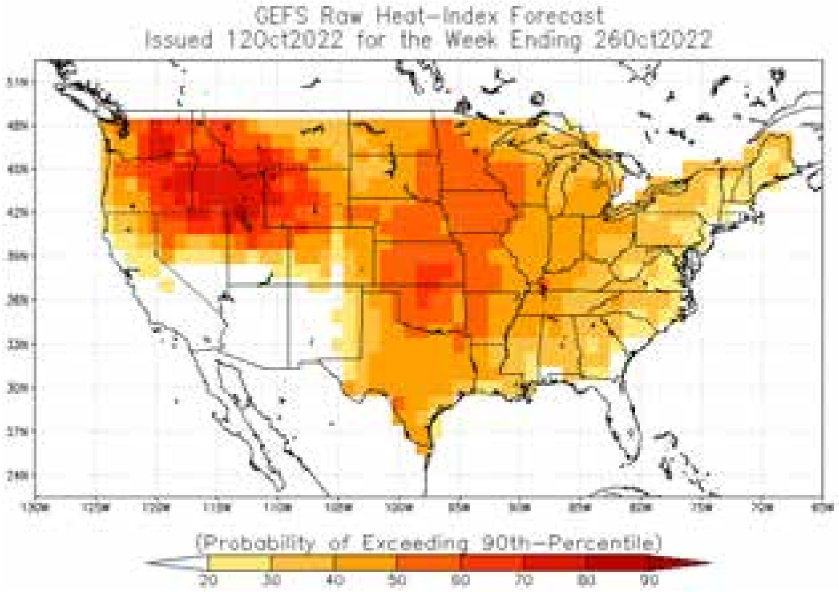
Greg Jennrich[1,2], Li Xu[1,2],
Evan Oswald[1,2], and Matthew Rosencrans[2]

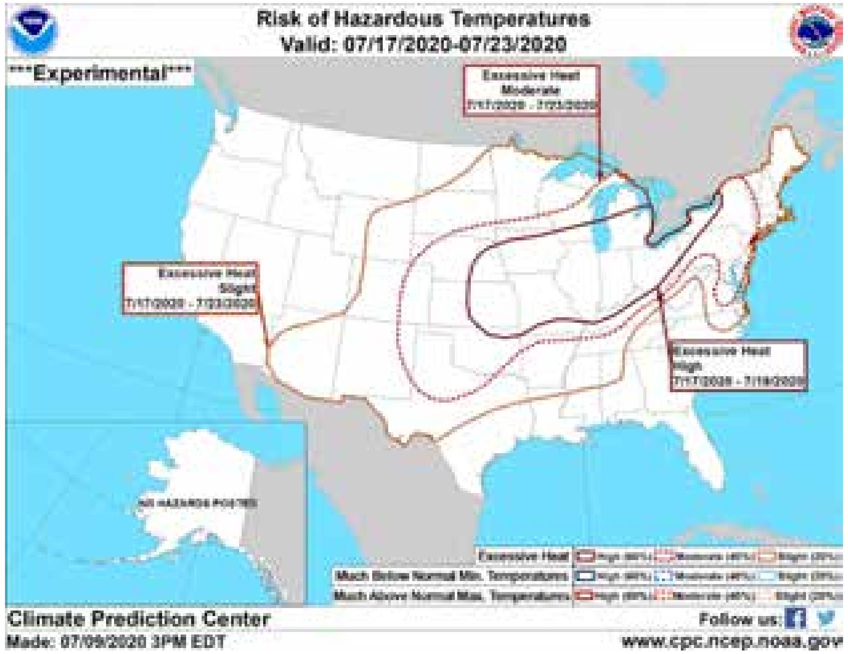
[1] ERT, Inc.

[2] NOAA Climate Prediction Center

Background

- Heat is the #1 weather related fatality in the United States
- CPC historically has focused on Week 2 anomalies, but is becoming more hazard/impacted minded.
- Like many others, CPC has started to explore the utilities of machine learning (ML) based approaches to our forecast problems.
- **Goal**: Complete initial exploration and development of a ML based extreme heat tool.

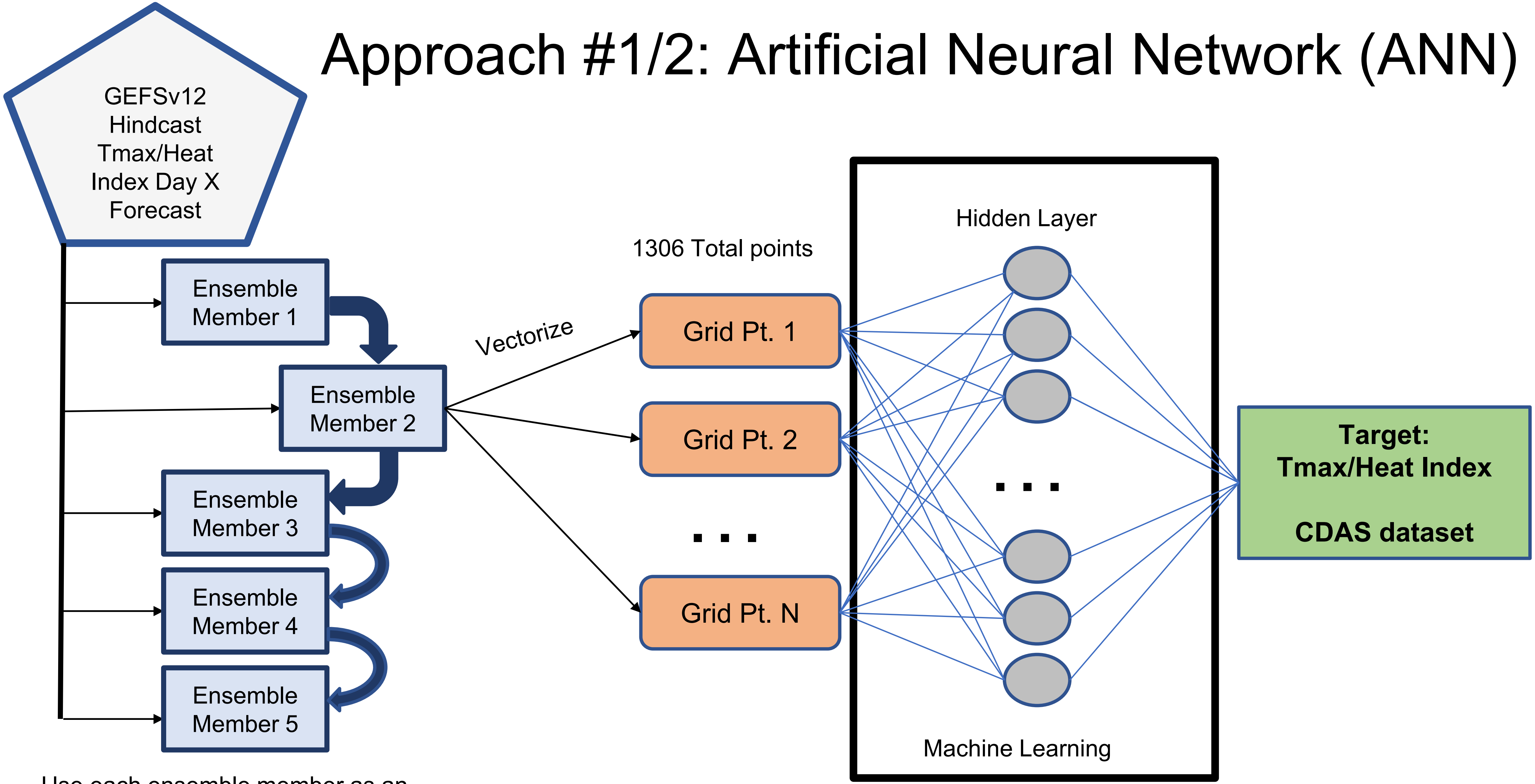




Data

- **Objective:** Improve on GEFSv12 Tmax and Heat Index forecasts via ML
- Summer period: April 9th- September 20th initialization dates (165/year)
- Initial model to be tested: Week 2 **GEFSv12** Tmax/Heat Index
  - *Hindcast*: 5 ensemble members (2000-2020)
  - *Realtime Verification*: 124 ensemble members (2021)
- Target data: **CDAS** Tmax/Heat Index Reanalysis
- 3 ML Techniques: Neural Nets (NN) via TensorFlow
  1. ANN bias correction of GEFSv12 forecasts
  2. ANN bias correction for specific heat products
  3. Deep Learning (DL) model with additional inputs.

Approach #1/2: Artificial Neural Network (ANN)



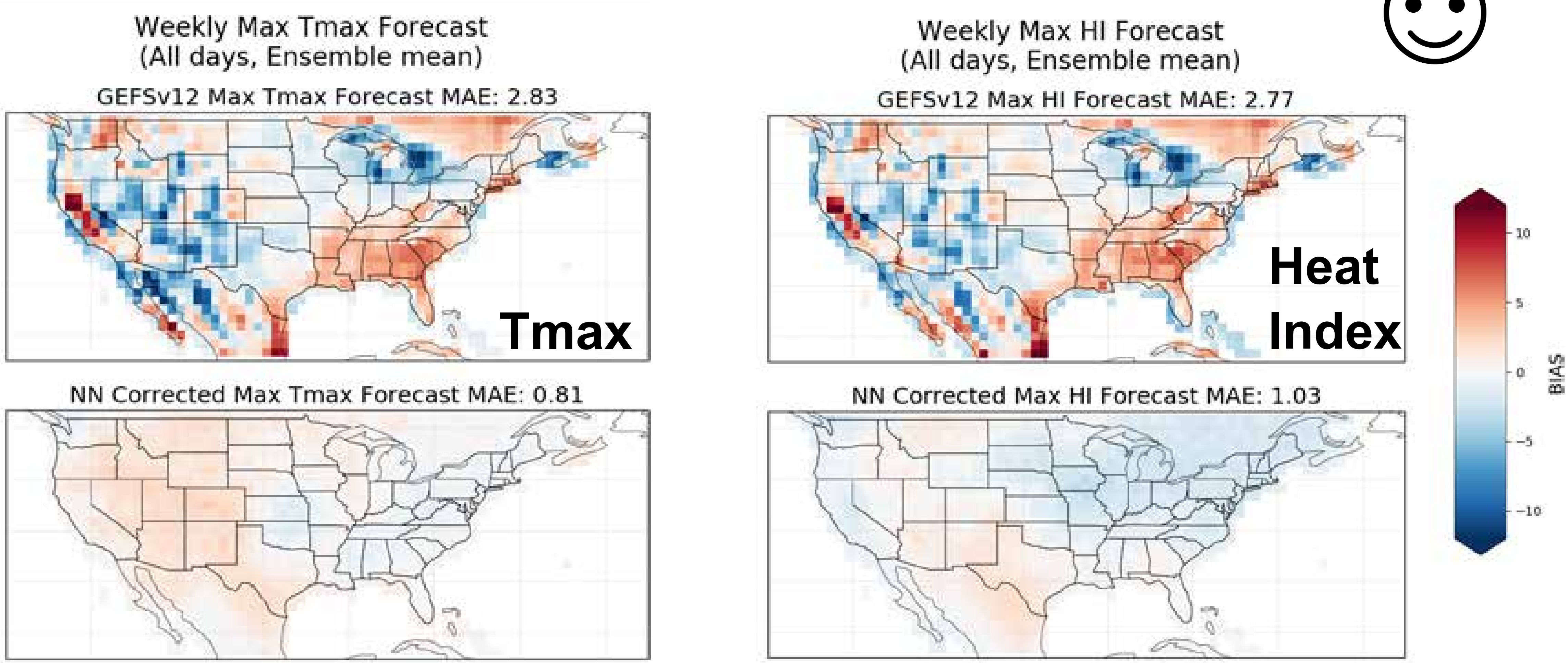
Use each ensemble member as an independent sample → Sample size= # of initialization dates * 7 forecast days * 5 ensemble members

Approach #1: Daily Forecast

- Train: 17 hindcast summers
- **Validate: 4 hindcast summers**
- Test: 2021 realtime summer

- Bias in the ensemble mean (GEFS vs. ML) in the 4 Validation summers
- Mean Absolute Error (MAE) greatly reduced.

☺

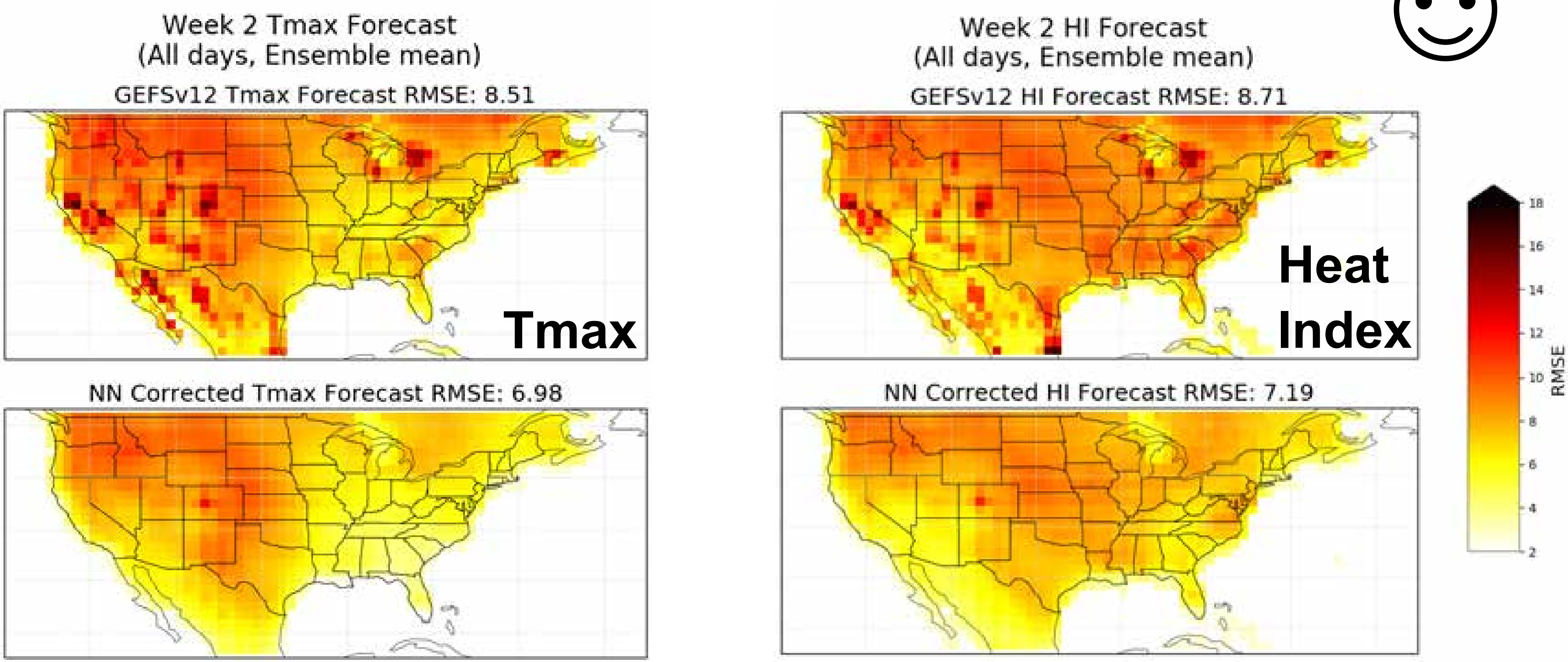


Approach #1: Daily Forecast

- Train: 17 hindcast summers
- **Validate: 4 hindcast summers**
- Test: 2021 realtime summer

- RMSE in the ensemble mean (GEFS vs. ML) in the 4 Validation summers
- Error is reduced greatly reduced

☺

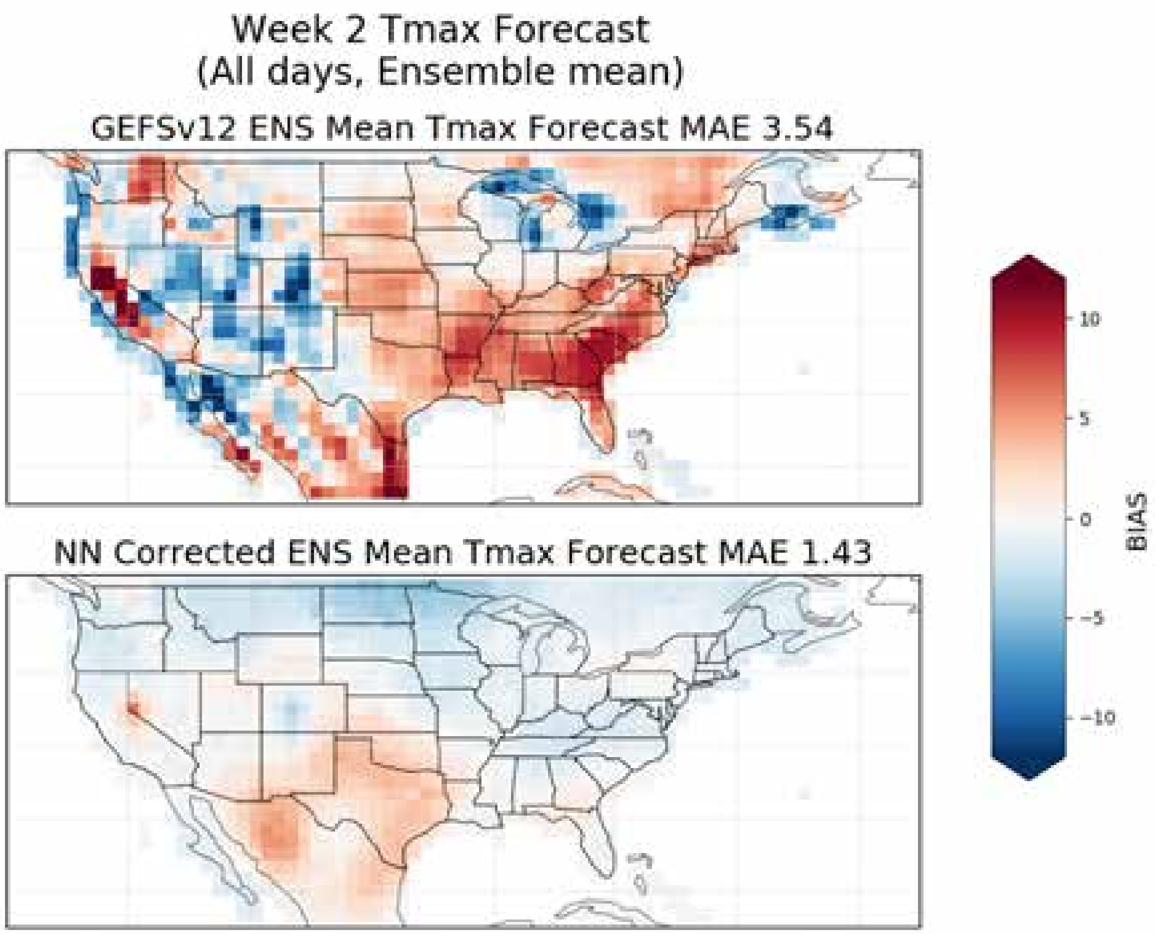


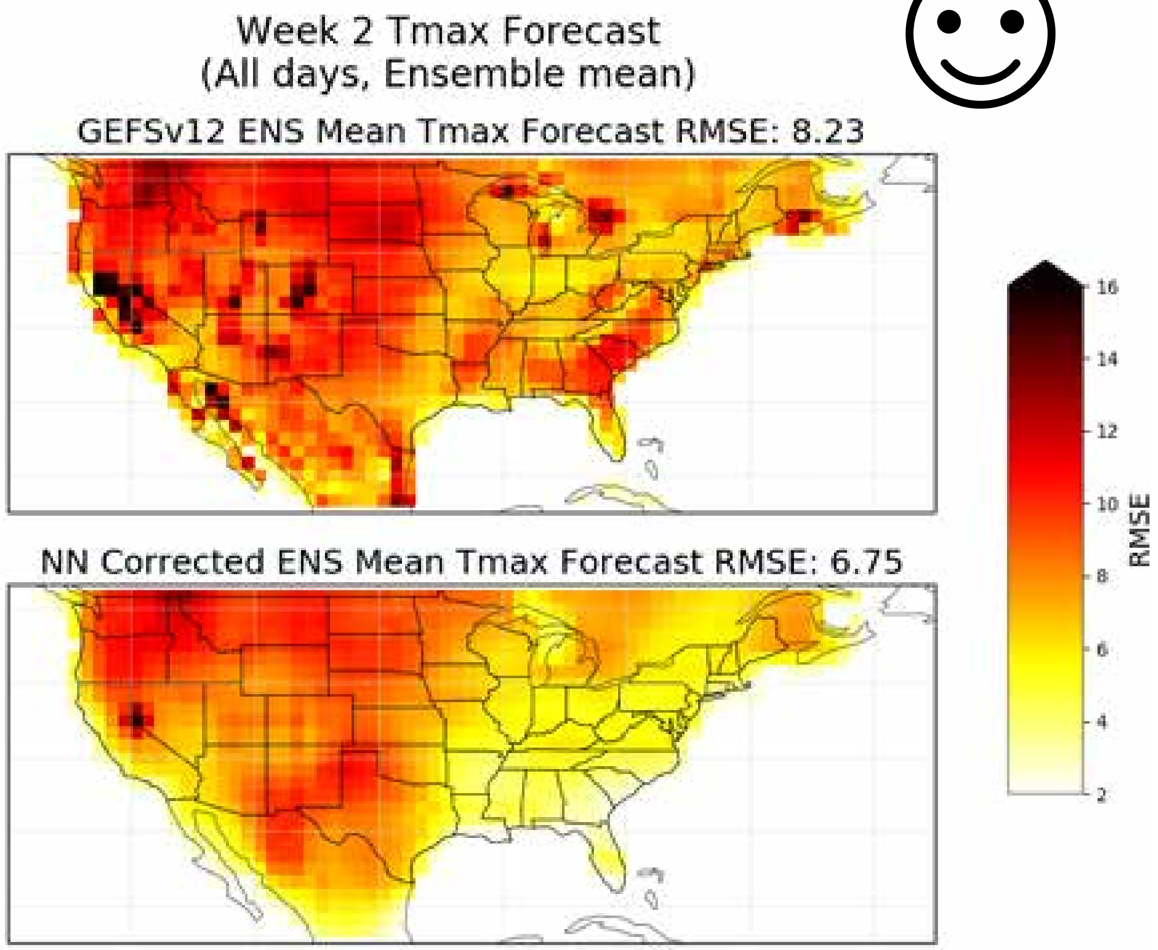
Approach #1: Daily Forecast

- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

- 2021 Realtime testing dataset for Tmax
- Both metrics show pretty good improvement.

☺





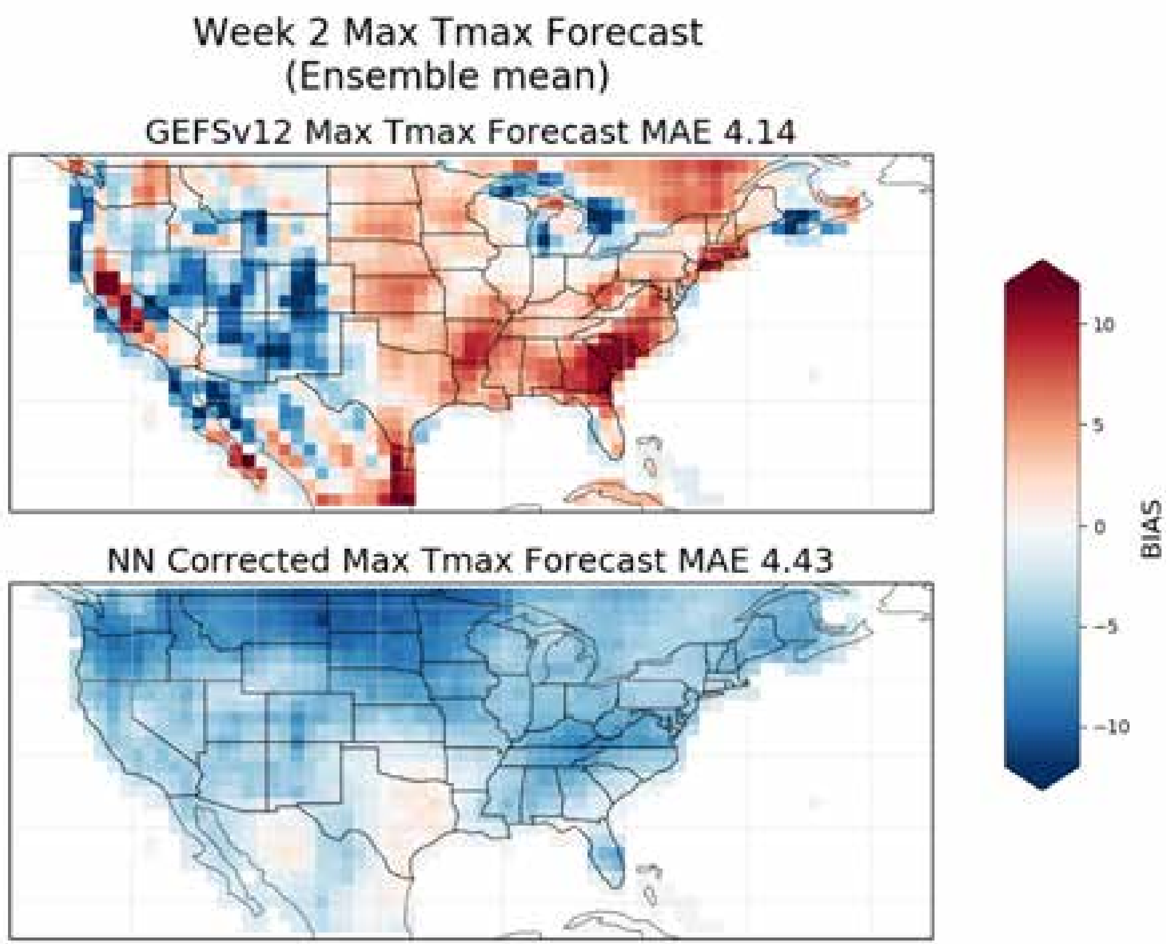
What about for Extreme Heat?

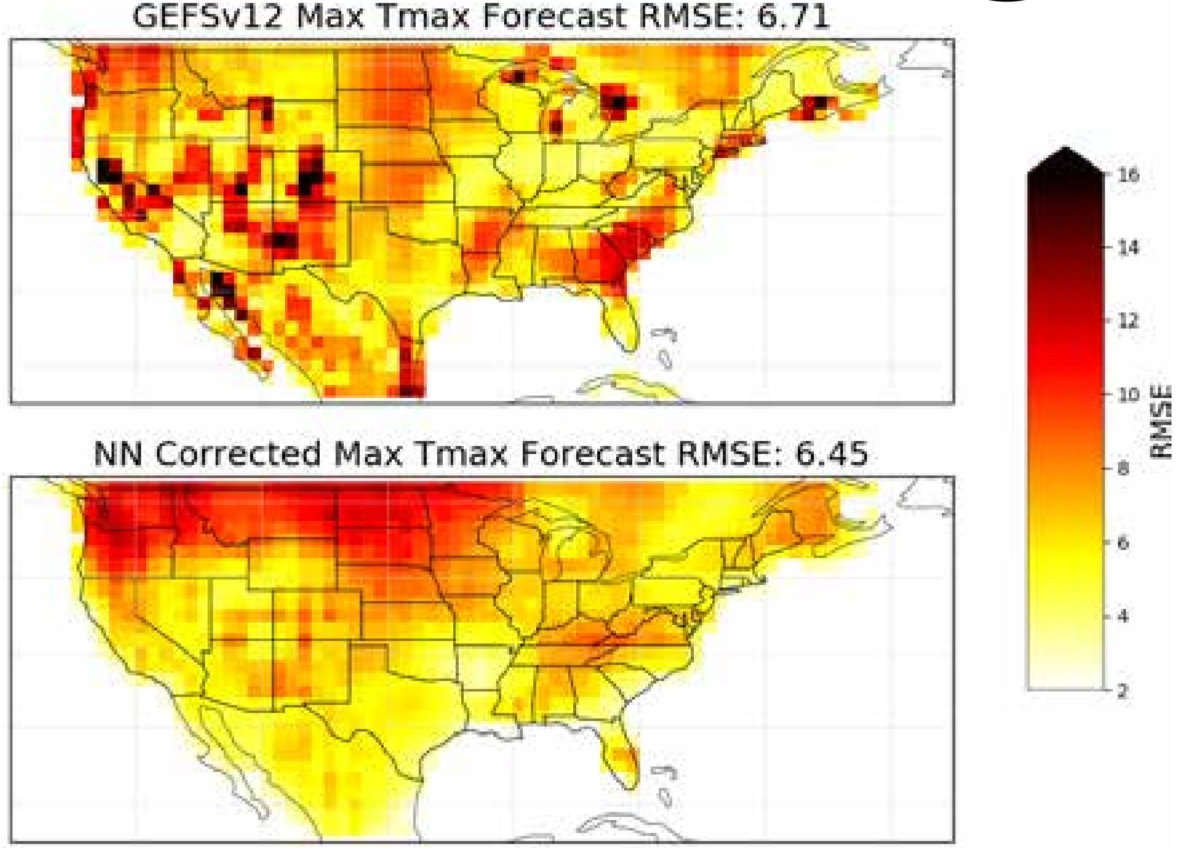
Let's look at two extreme heat products:
1. Week 2 Maximum Temperature
2. Probability of ≥ 100F during Week 2

Approach #1: Week 2 Max

- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

- The improved results do not translate to weekly maximums
- Pretty large cool bias

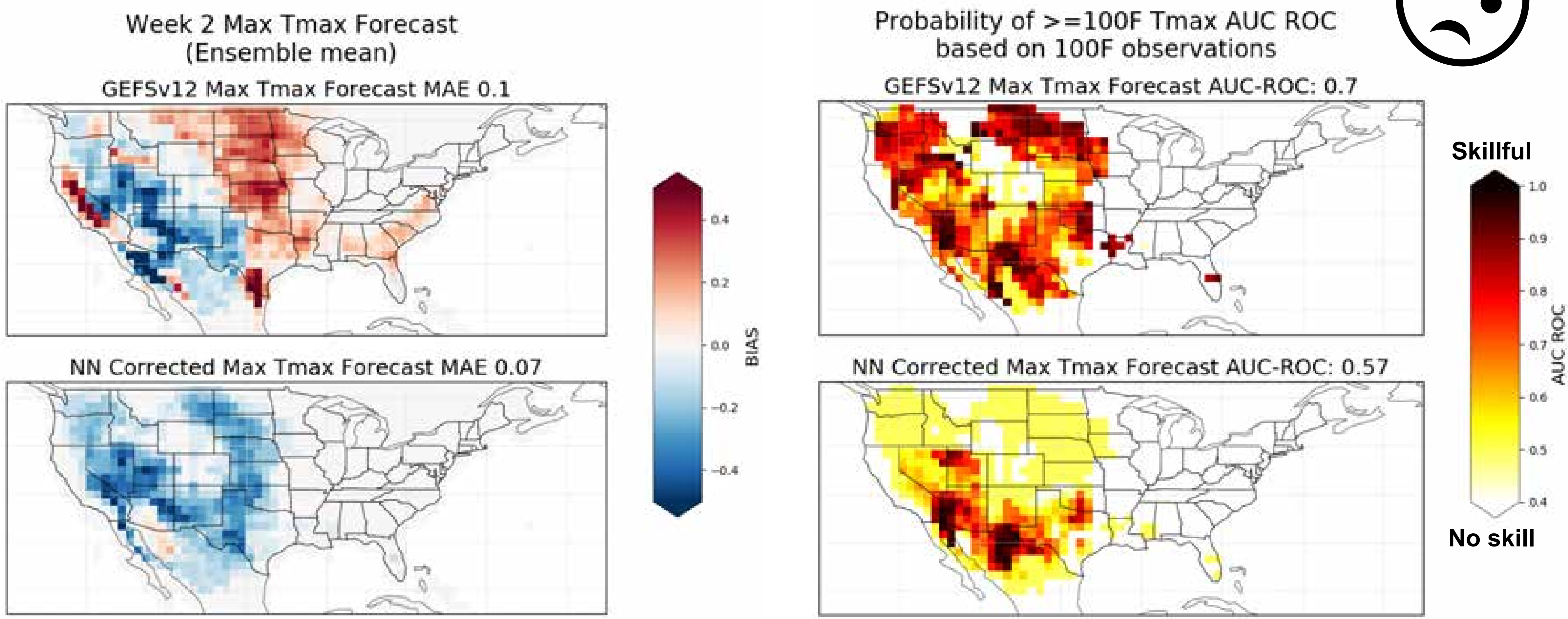




Approach #1: Probability ≥100F

- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

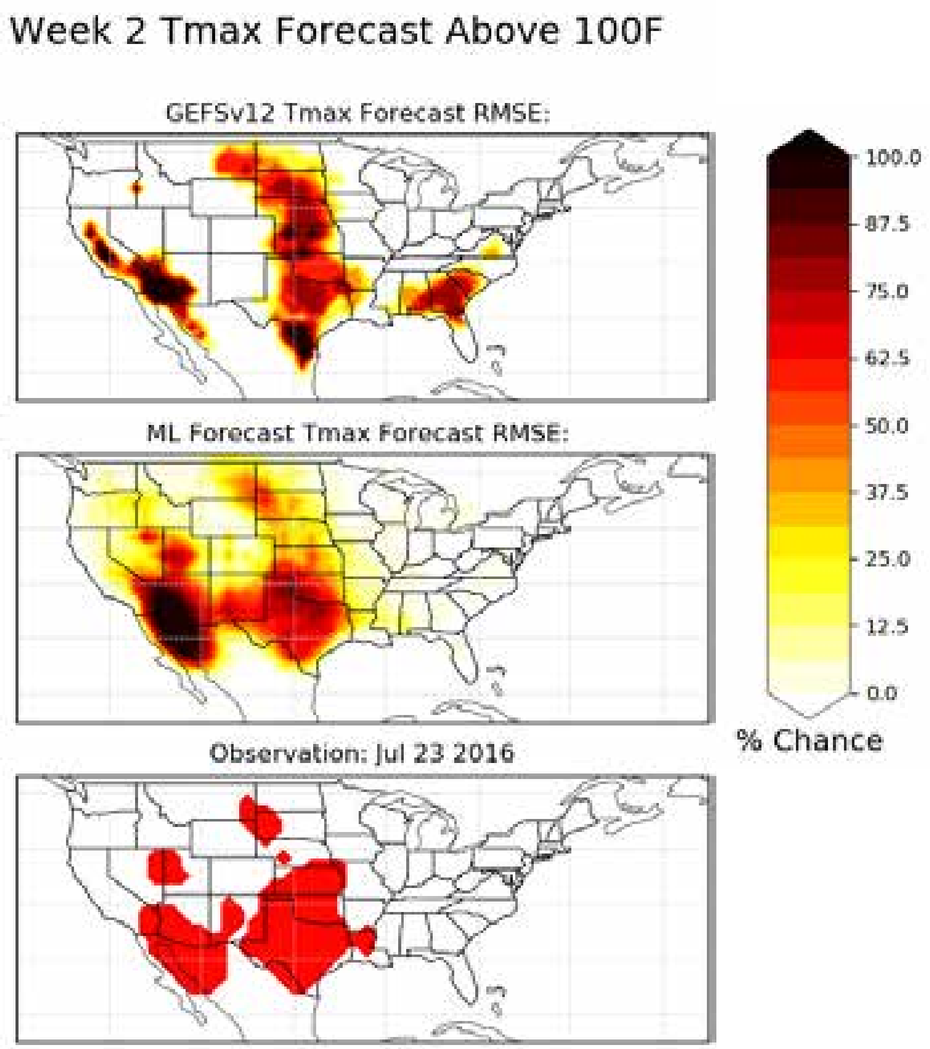
- This model does not predict enough 'hot' temperatures
- AUC-ROC = Area Under the ROC Curve → Measure of classification skill (>=100F or not)



Approach #2: Product Specific Models

- Instead of using one ML model and then calculating heat products, can we use ML models to directly improve these products

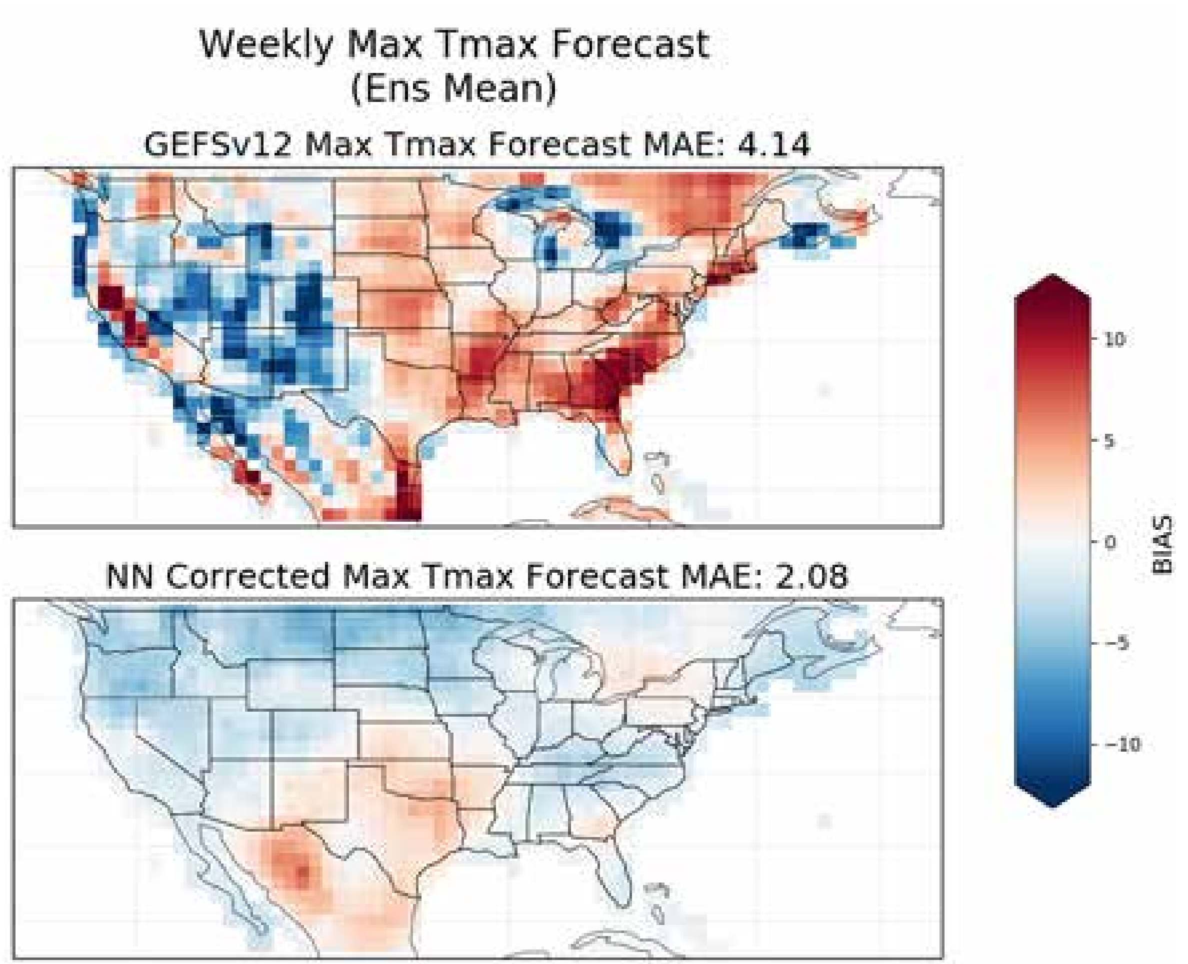
- Let's test these two heat products:
  1. Week 2 Maximum Temperature ML model
  2. Probability of ≥ 100F during Week 2 ML model

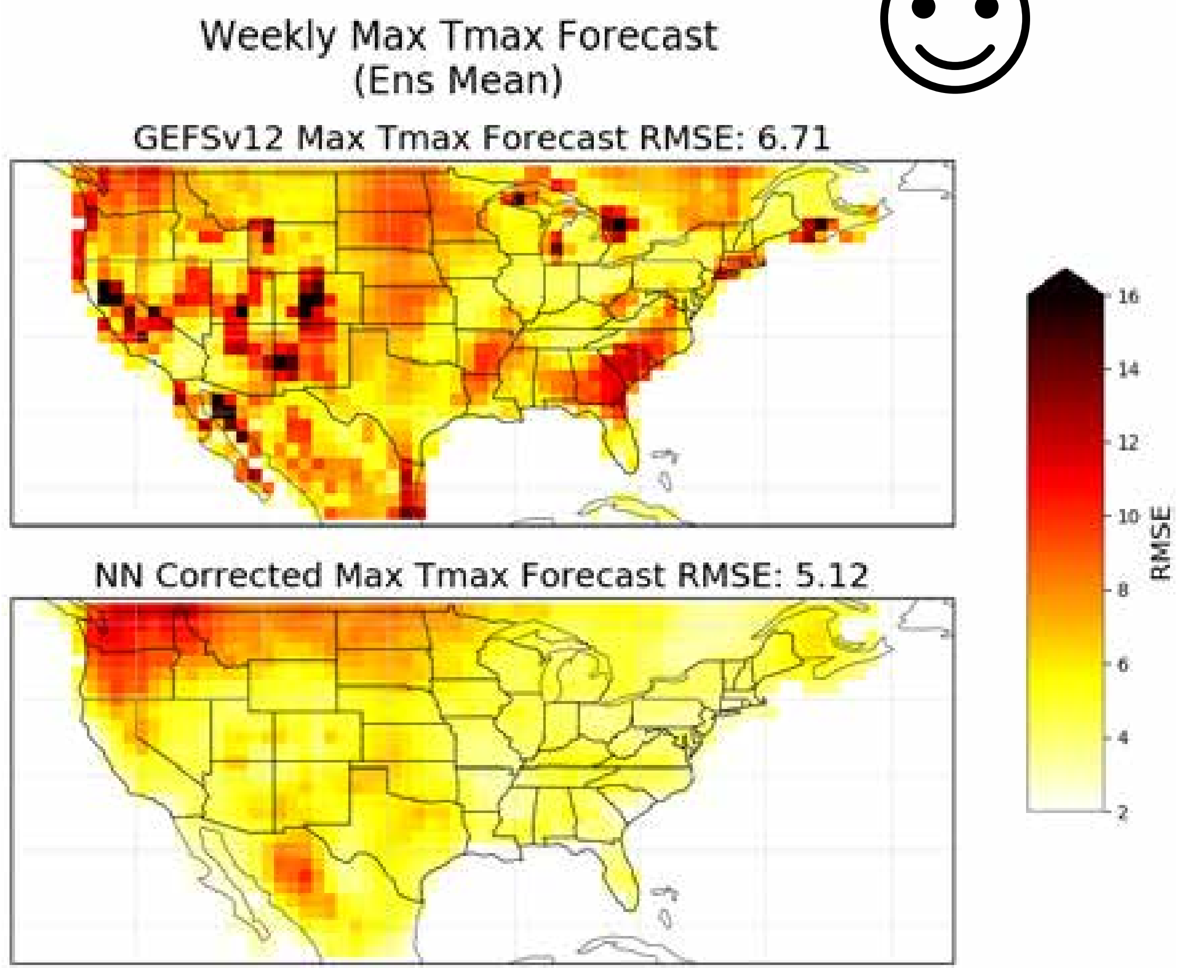


Approach #2: Week 2 Max

- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

- Much better than the corresponding results on slide 9





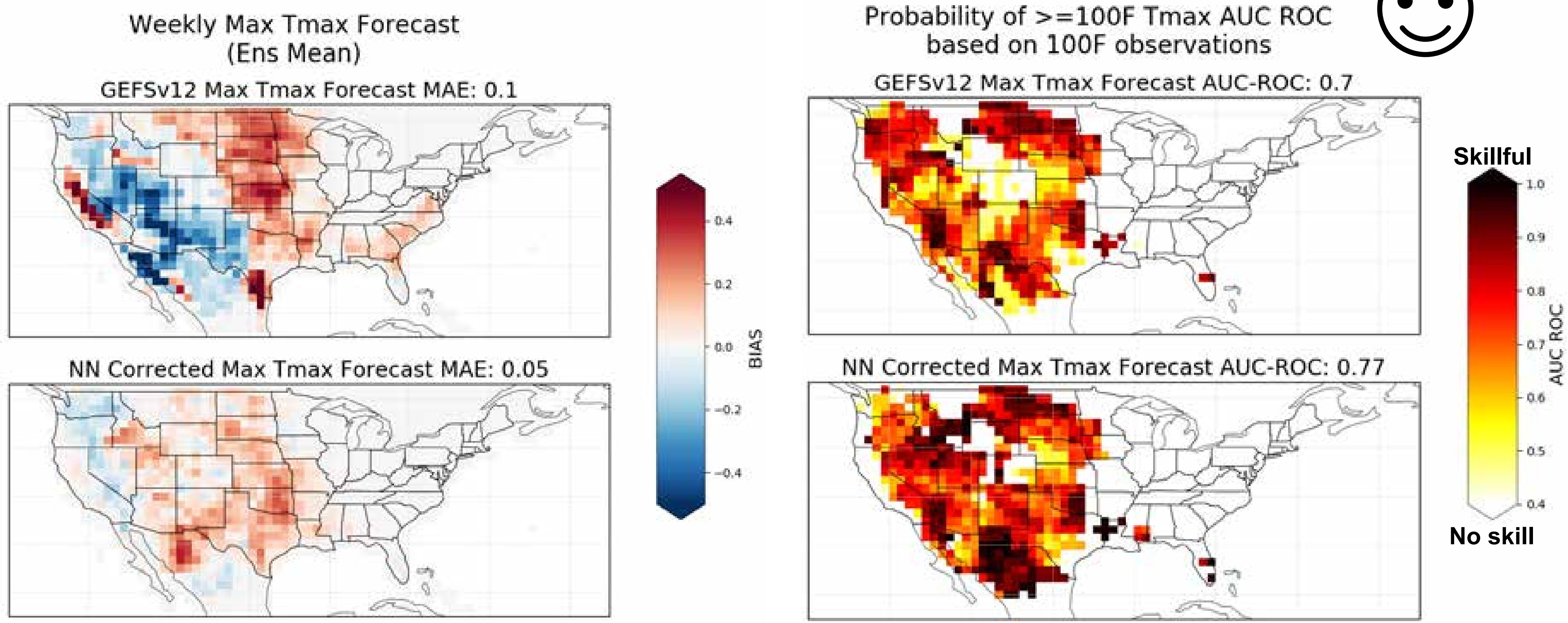
Results: Approach 2

Approach #2: Probability ≥100F

- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

- Much better than the corresponding results on slide 10
- AUC-ROC = Area Under the ROC Curve → Measure of classification skill (>=100F or not)

☺

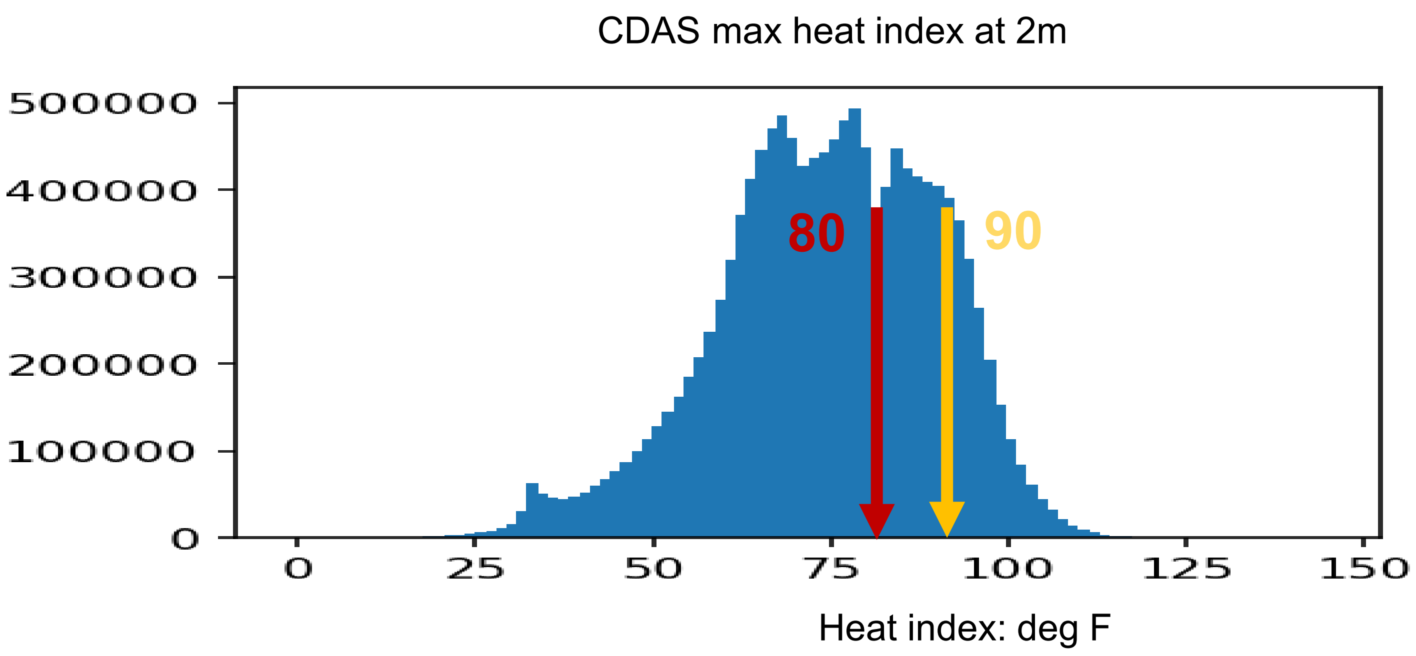


Results: Approach 2

Approach #3: The Long tail paradox

Work by Li Xu

- Since we are focused on impactful heat, we focused on only adjusting the tails of the distribution
- Tmax/HImax > 80F
  - Focus on Right tail (~28.4% total samples)
  - Total 23.1M grid-point sample forecasts during 2000-2019 hindcast period
- Tmax/HImax >90F
  - ~9.0% total sample forecasts
  - 7.3M grid-point sample forecasts
- Train points with a Deep Learning Net (DLN), one for each variable.



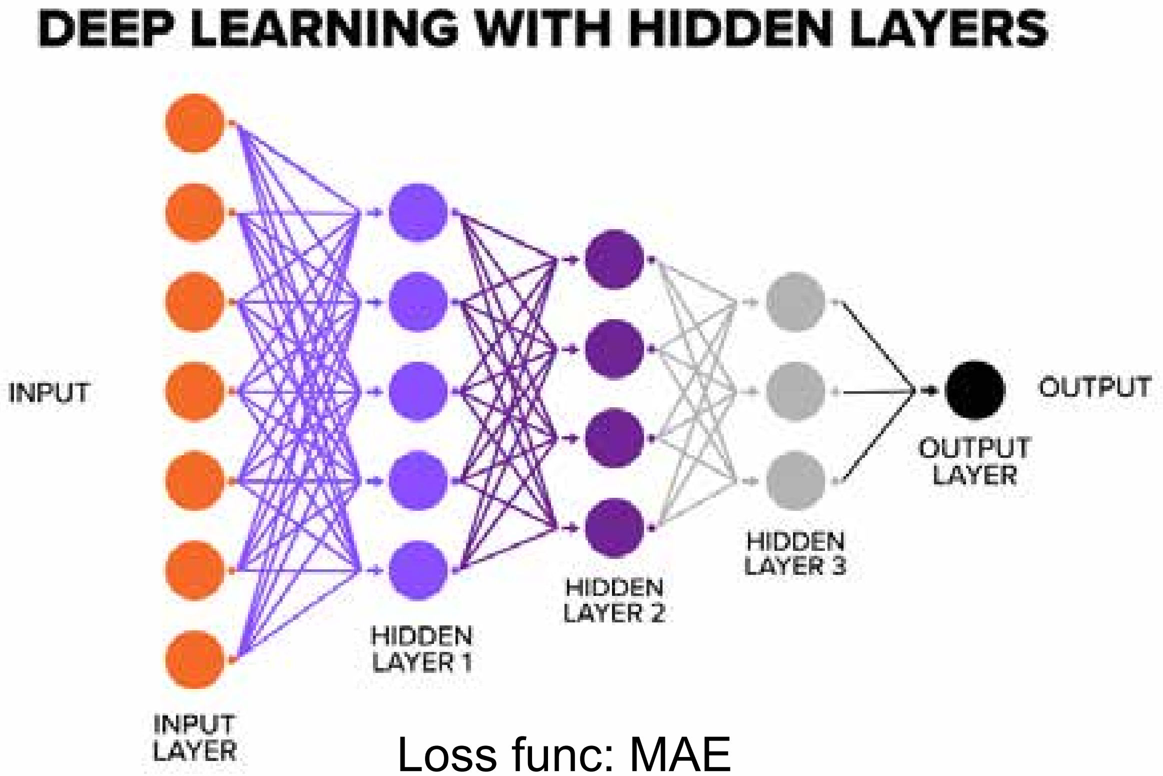
Histogram of Daily max **Heat Index** over the CONUS, data source: CDAS

Deep learning Input (Predictor) for Tmax/HImax

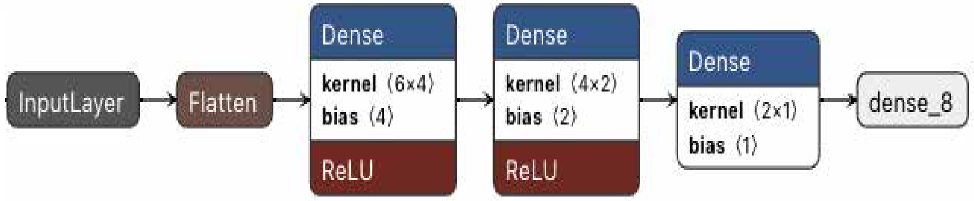
- Model guidance
  1. GEFSv12 Heat Index
  2. GEFSv12 Tmax
  3. Relative humidity (implicit)
  4. ~~blocking index (Tibaldi and Molteni 1990)~~
- Geographic and seasonal information
  1. Latitude / Longitude
  2. Elevation (DEM)
  3. ~~Vegetation type (fraction)~~
  4. The Day of year (normalize with distance to Aug 1st)
  5. Forecast leading (8-14)

Tried, but didn't work out

- Low Boundary forcing(physics driver)
  1. Soil Moisture (Standardized SM Index from LB model)
  2. ENSO (nino3.4 index)



Loss func: MAE

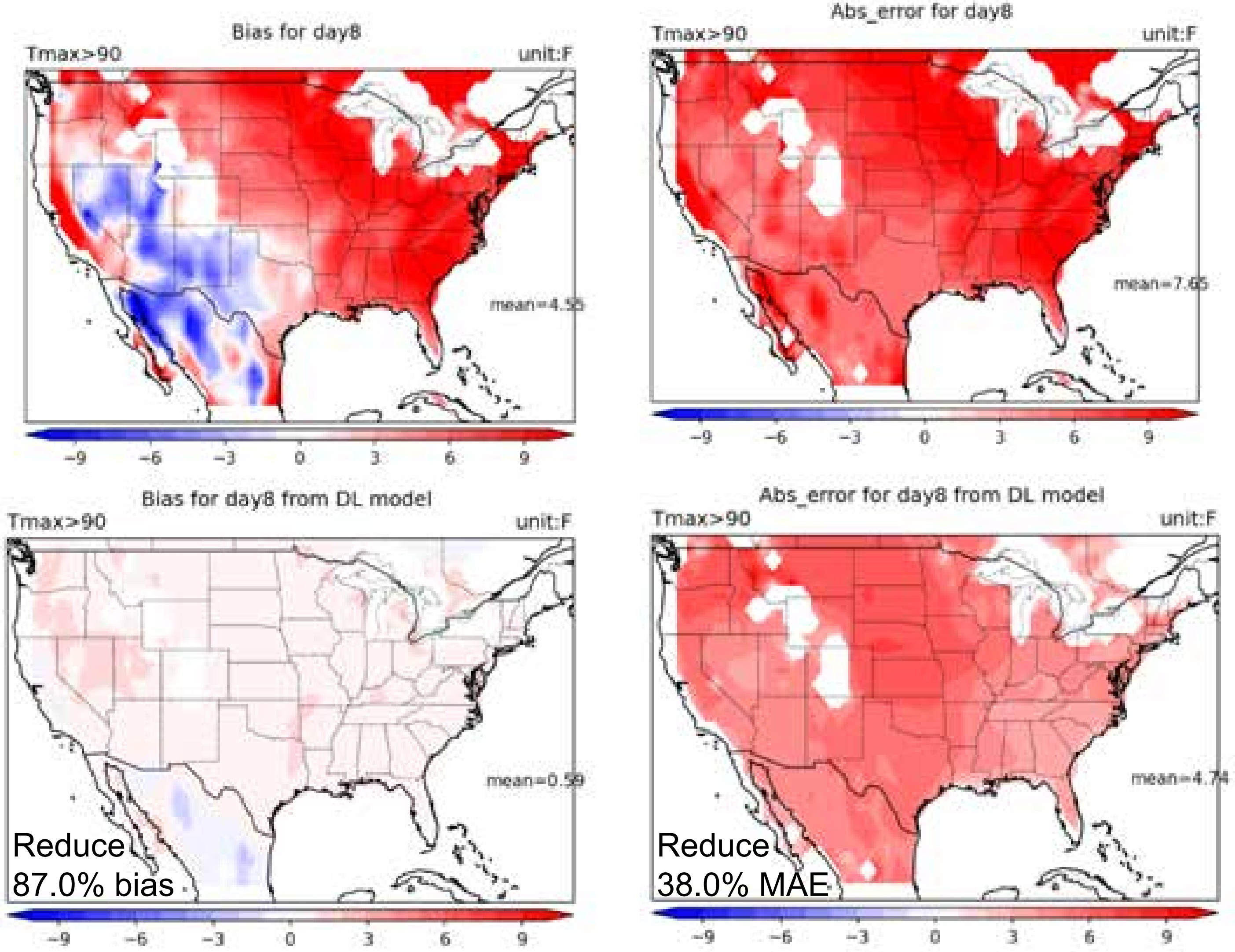


Results: Approach 3

Approach #3: Daily Forecast

- Train: 17 hindcast summers
- **Validate: 4 hindcast summers**
- Test: 2021 realtime summer

- Day 8 Forecast ensemble bias and mean absolute error (MAE) of the GEFS and the >90F Deep Layer model.
- Only scoring the points above 90F, thus the sample size is not equal for each point
- The DL model show improvements for Bias and MAE

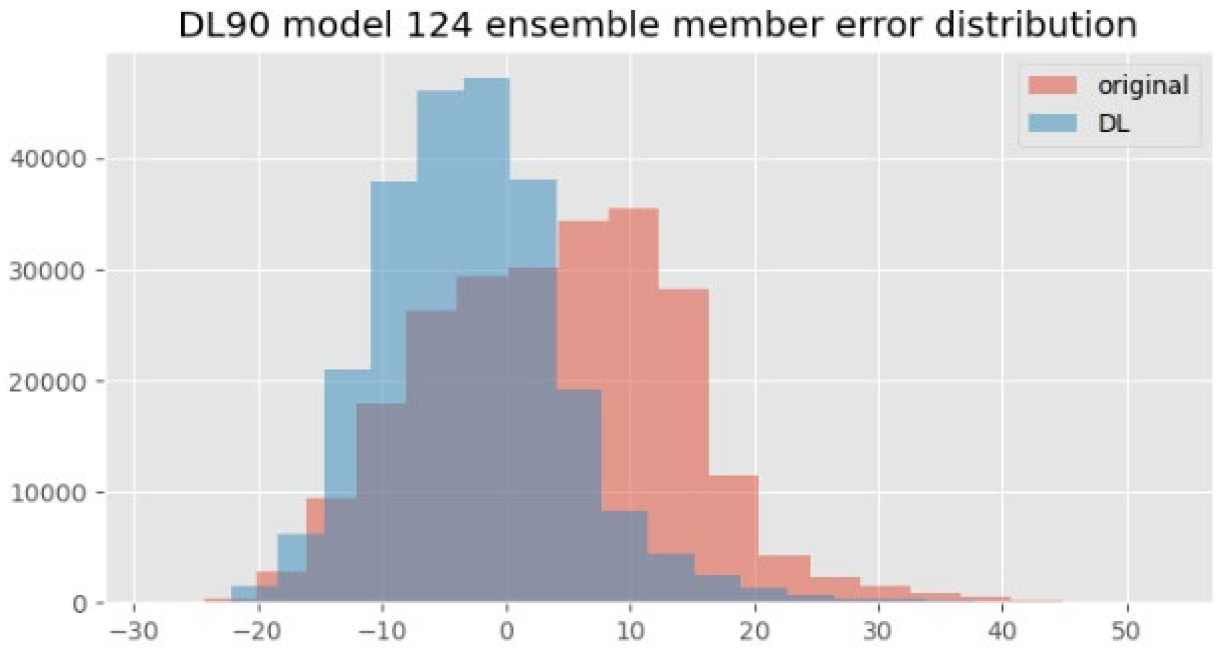


Results: Approach 3

Approach #3: Daily Forecast

- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

- We have just begun to test this model for the 2021 period
- Looking at each DL point, the GEFS bias distribution skews warm
- The DL forecast bias distribution was narrowed and shifted to center near zero, but suggests a small cool bias
- More 2021 testing in progress



Contact:
Greg Jennrich
Gregory.Jennrich@noaa.gov

Conclusions

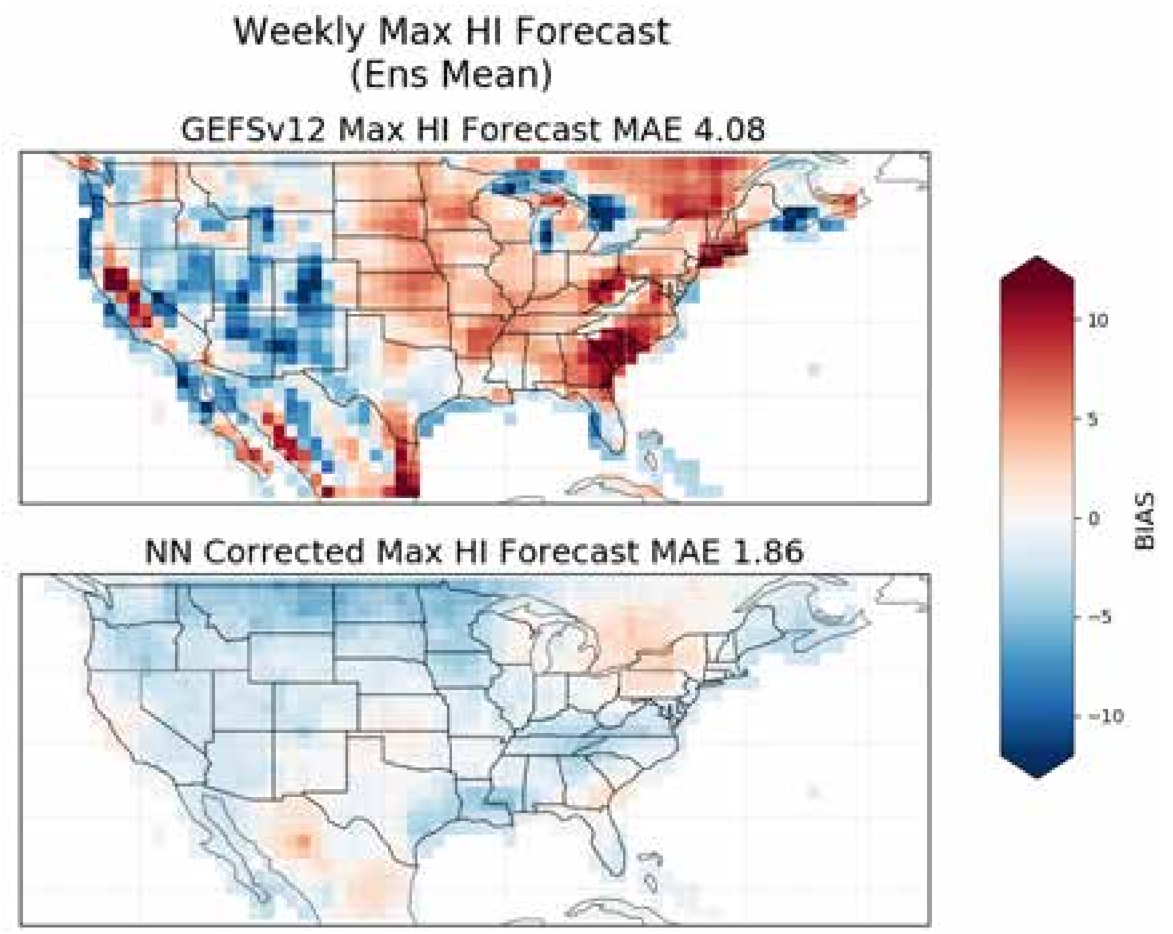
- For extreme events, the general ML/AI model that fit for the whole distribution does not work well (approach 1)
  - Works well for the 2 Week average, but not on extreme heat
  - Can't expect to use a ML model on a different product than it was designed for
- We can overcome this issue in two ways:
  - Use heat product specific models (approach 2)
  - We can trim/adjust the model to only fit for the long tail (approach 3)

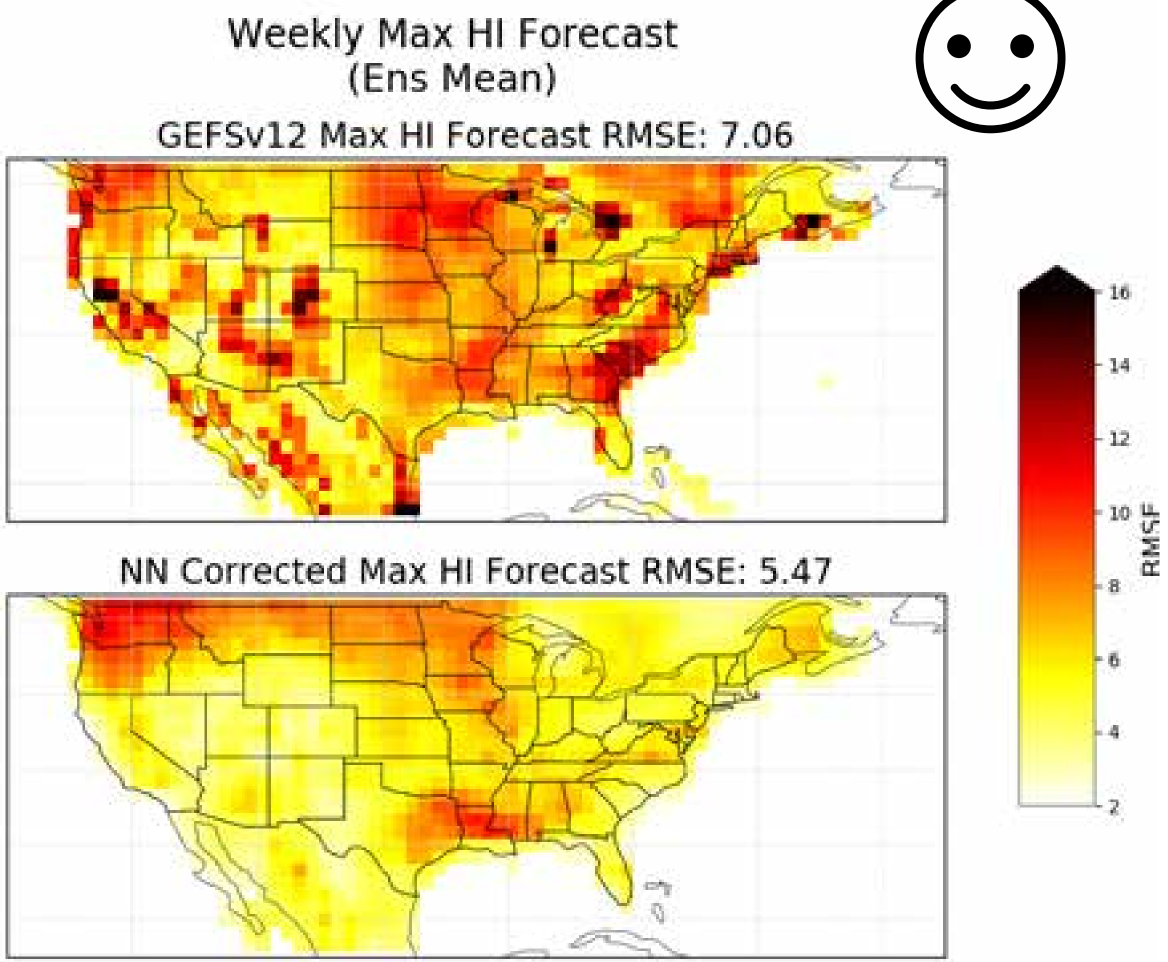
Future Work

- Further analysis with Approach #3
- Apply and test with other models (ECMWF) and with 2022
- Possible realtime tool in 2023

Approach #2: Week 2 Max

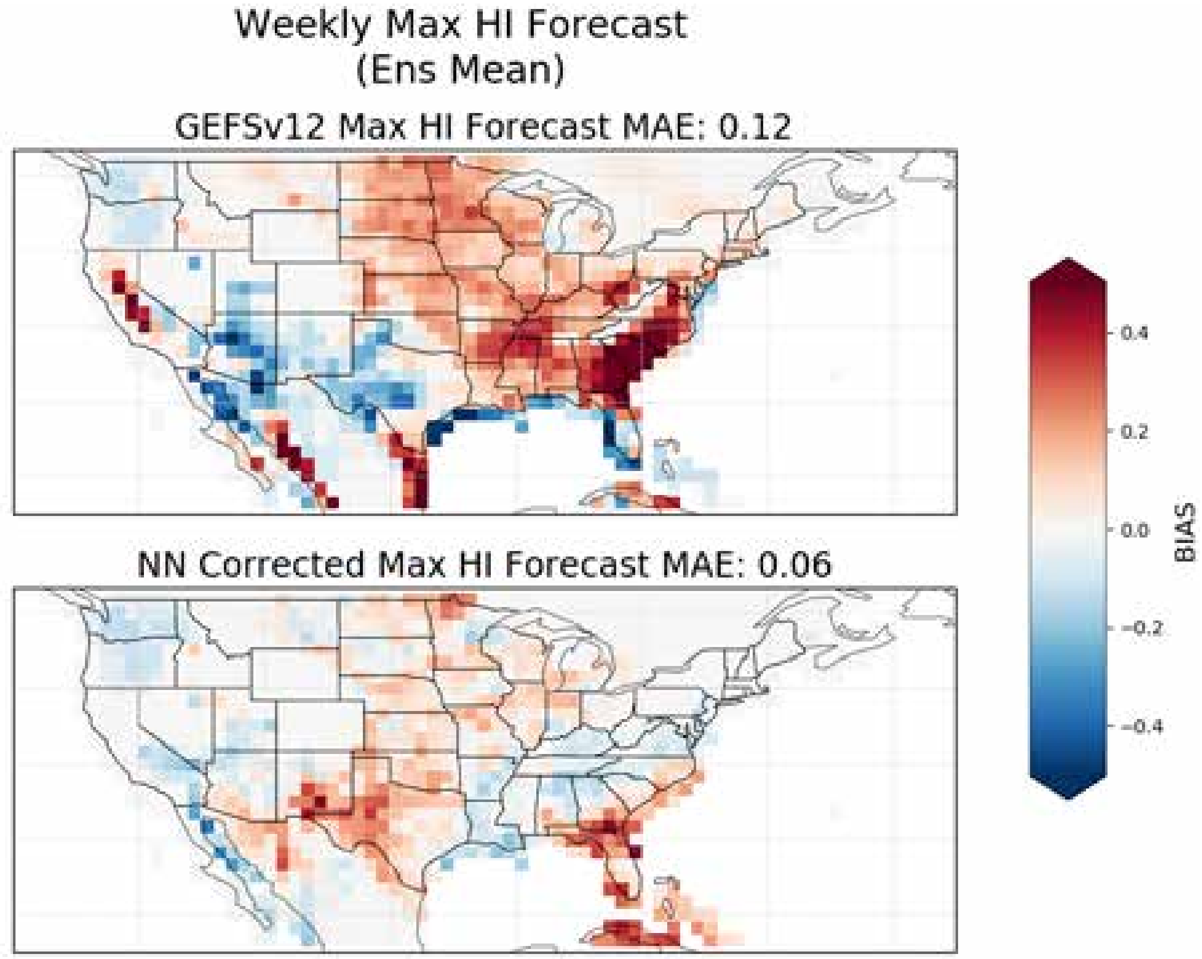
- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

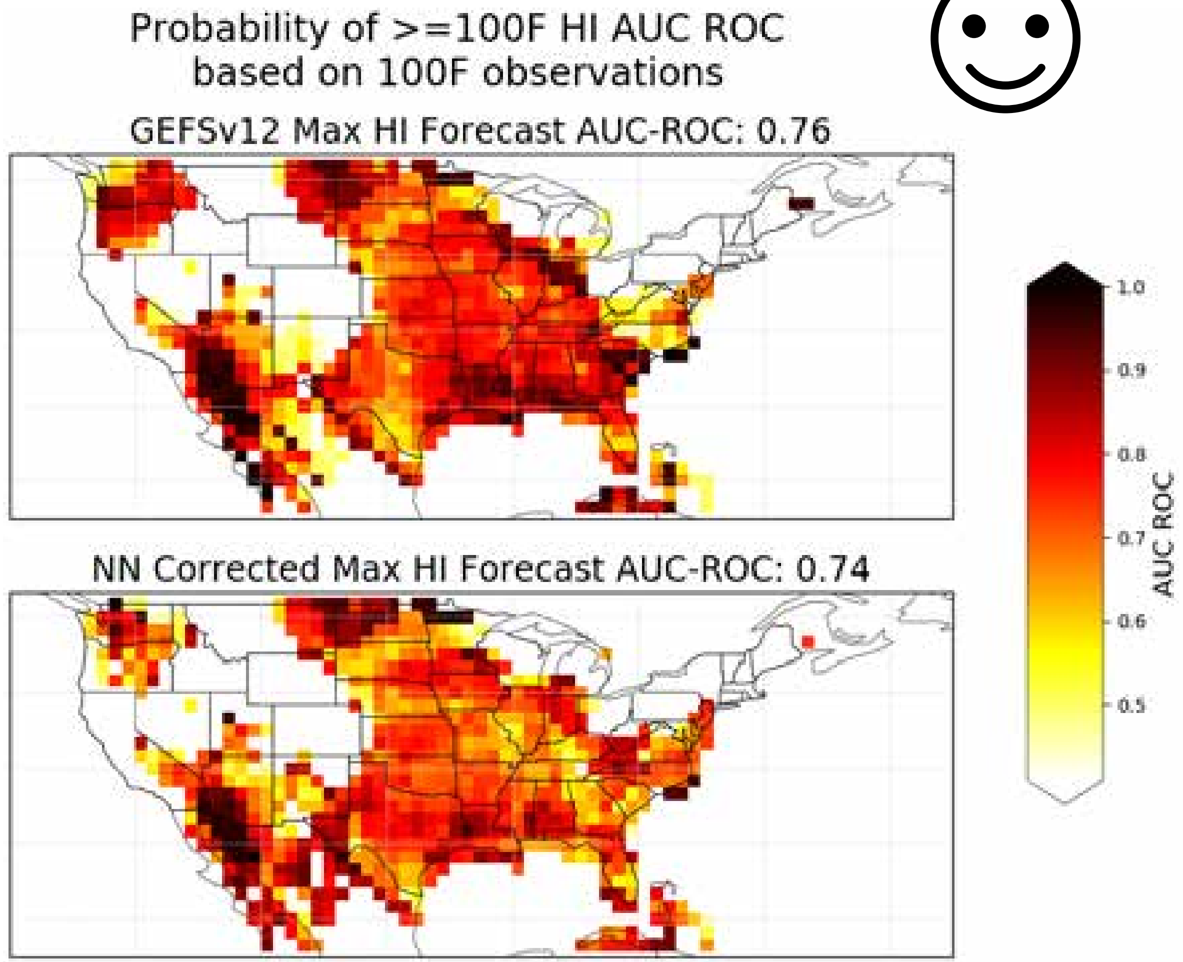




Approach #2: Probability ≥100F

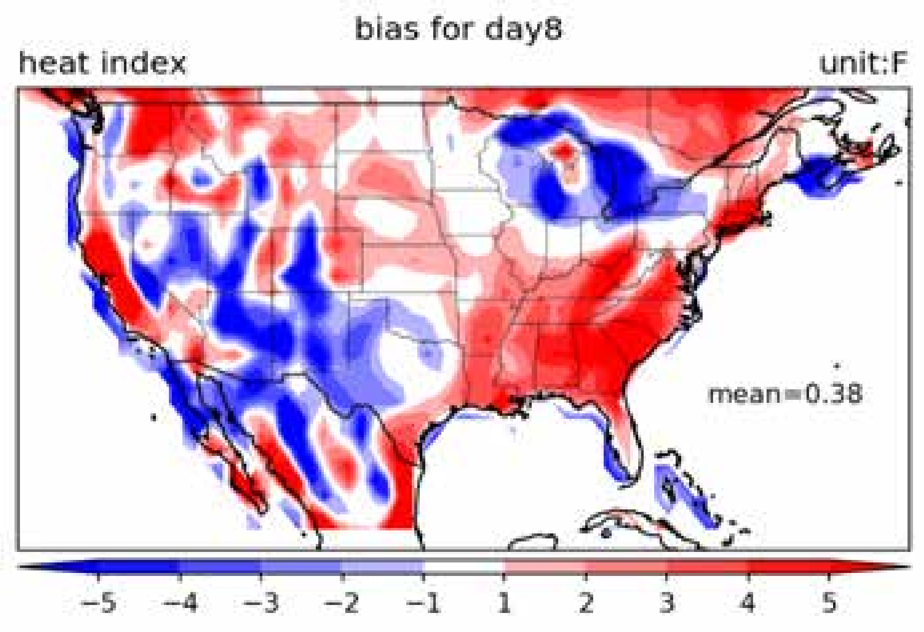
- Train: 17 hindcast summers
- Validate: 4 hindcast summers
- **Test: 2021 realtime summer**

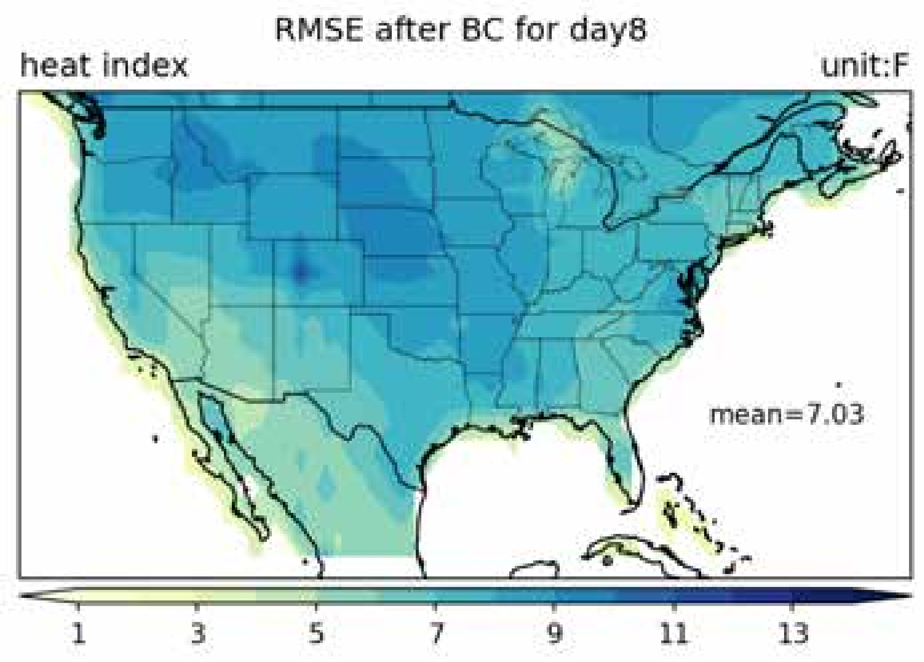


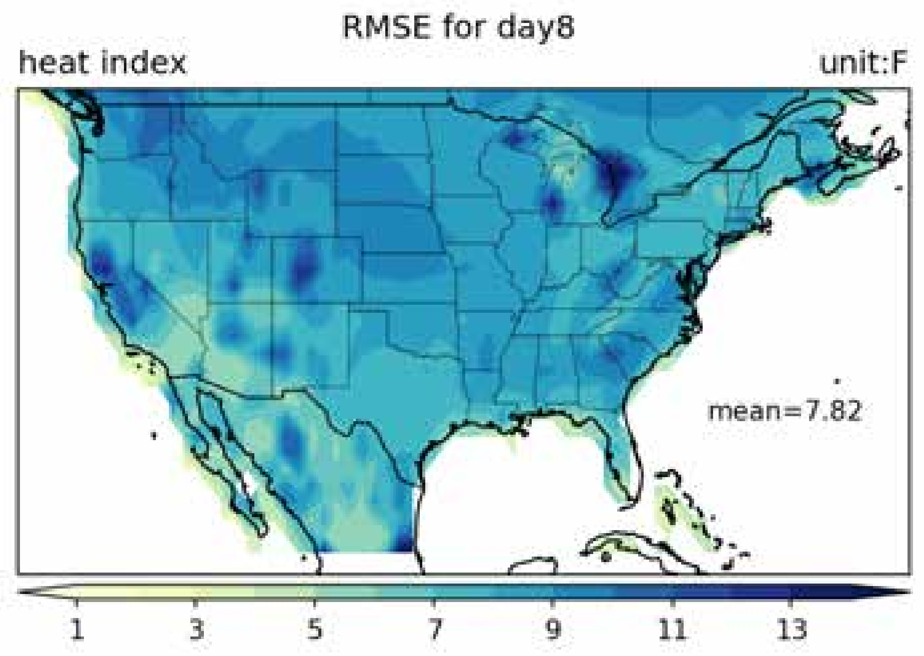


Simple Bias Correction

- Why go through all the trouble with ML if we can just slap on a simple Bias correction?
- It doesn't provide much improvement in the Root Mean Square Error (RMSE)
- We would need a separate bias correction for each forecast product







Extras