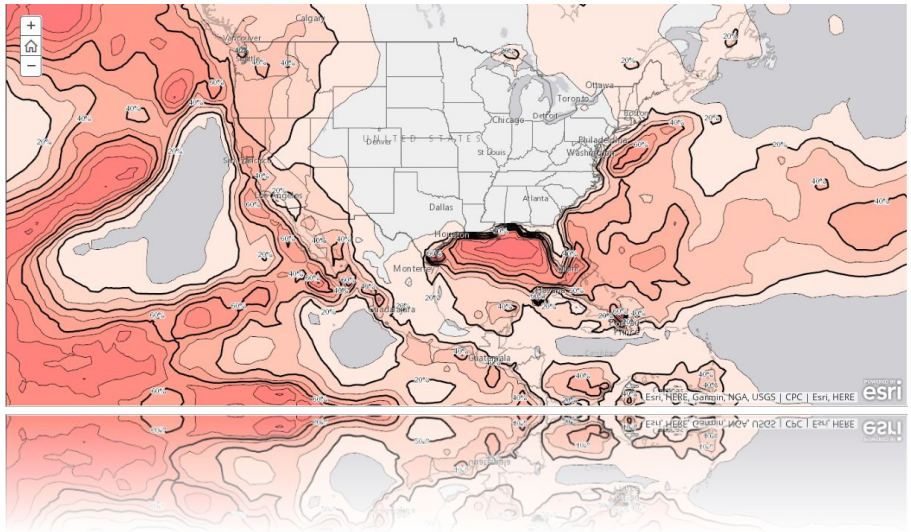
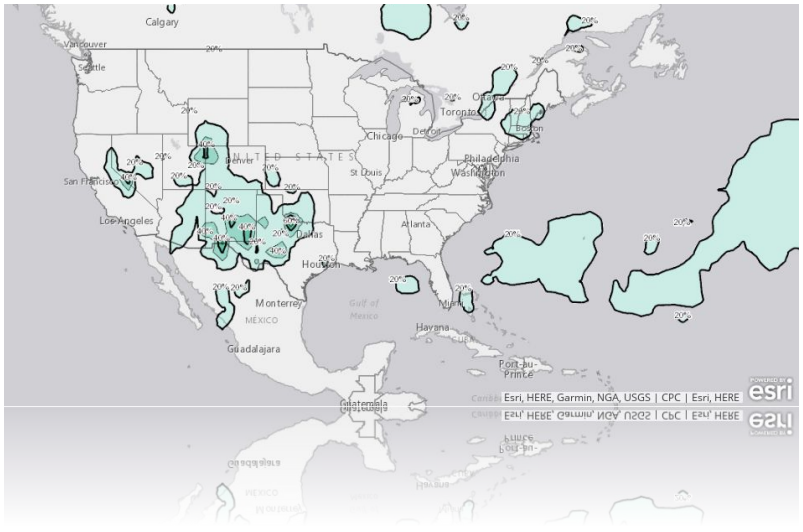
CPC's Week-2 Probabilistic Extremes Tool (PET)

NOAA / NWS / NCEP / Climate Prediction Center

Melissa Ou





Outline

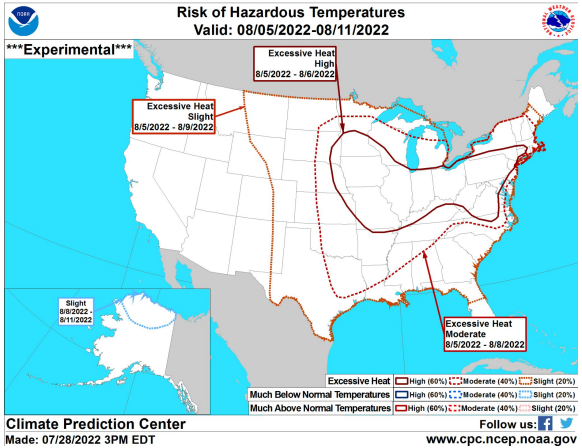
- Motivation and background
- Overview of the Week-2 Probabilistic Extremes Tool (PET)
- PET data and calibration methodology
- Verification methodology
- Cursory verification Results
- Future work / gap needs

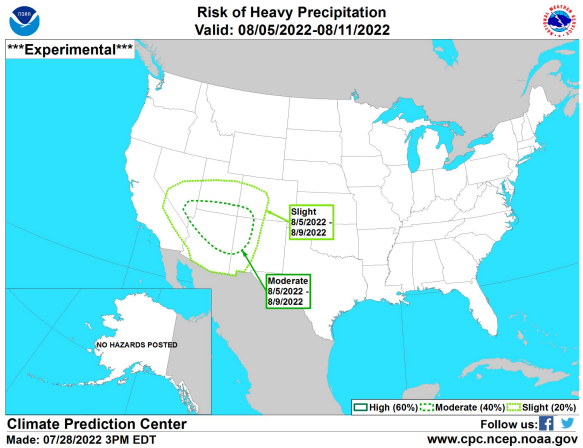
Motivation and Background

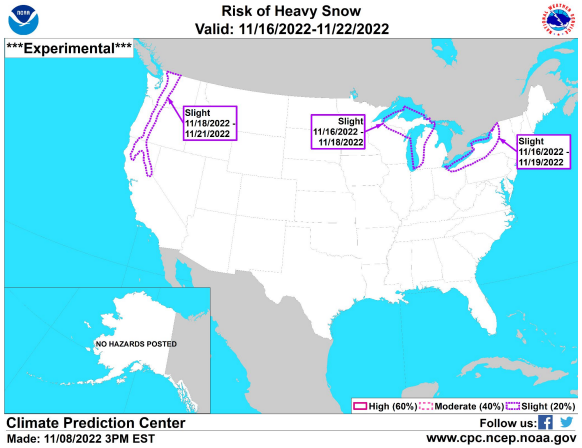
- Need for a reliable probabilistic tool for various variables to support the probabilistic Week-2 U.S. Hazards outlook, which is geared towards Impact-Based Support Services (IDSS)
- Stakeholders want more detailed regional probabilistic extremes information beyond the Week-2 U.S. Hazards Outlook
- Need for forecasts in percentile space and actual values
- Need for probabilistic snow model guidance, increasing interest in snow forecasts by users
- Need for objective verification of the PET

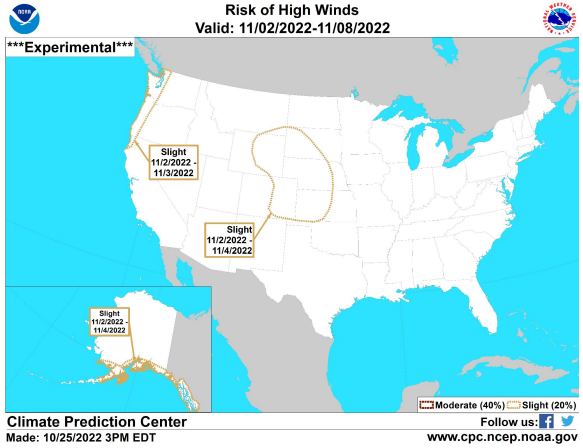
Week-2 U.S. Hazards Probabilistic Outlook

https://www.cpc.ncep.noaa.gov/products/predictions/threats/threats.php





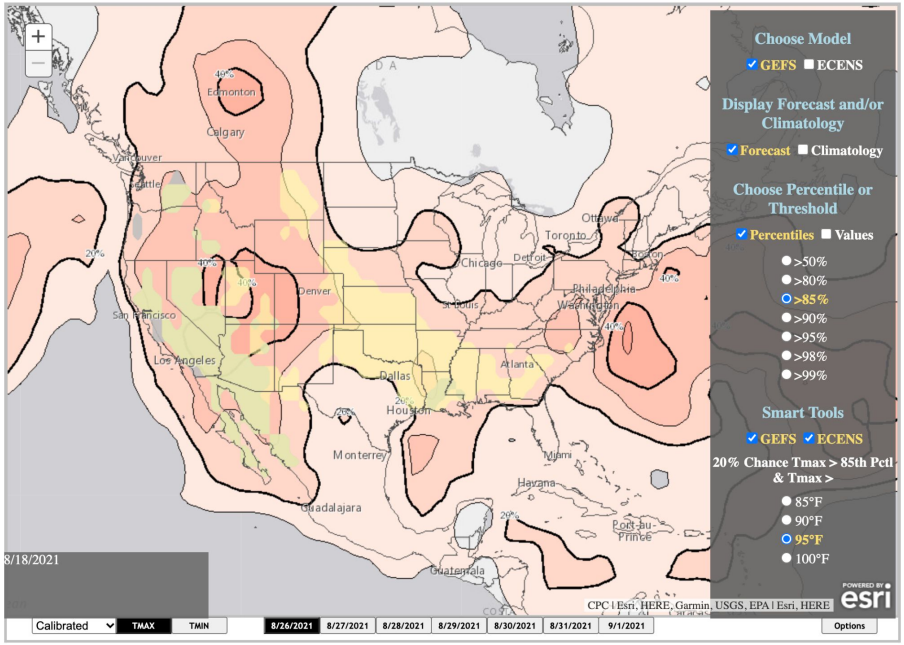




- Manually drawn forecasts of risk levels for potential hazardous events
- PET is the primary guidance tool, providing objective post-processed probabilities
- Slight, moderate and high risk of a hazardous event occurring represents at least a 20%, 40%, 60% chance of an event occurring, respectively
- Forecasters evaluate percentiles and actual values in addition to other considerations

About the Probabilistic Extremes Tool (PET)

https://www.cpc.ncep.noaa.gov/products/predictions/threats/extremesTool.php



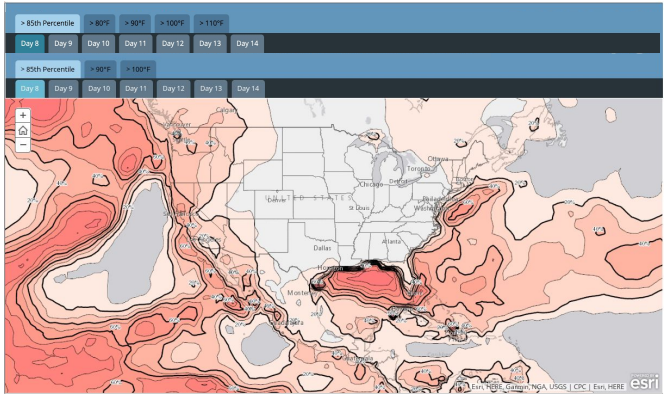
Sample screenshot of Tmax internal version

- Public and internal version (updated daily)
- Public only has GEFS, internal also has ECMWF and Canadian ensembles
- Tool publicly available as of Sept 2018
- Forecast probabilities of exceeding various thresholds
- Thresholds in percentiles and actual values
- Probability distribution built from model ensemble
- Model forecast probabilities are post-processed (bias-corrected and calibrated)
- Global

About the Public PET

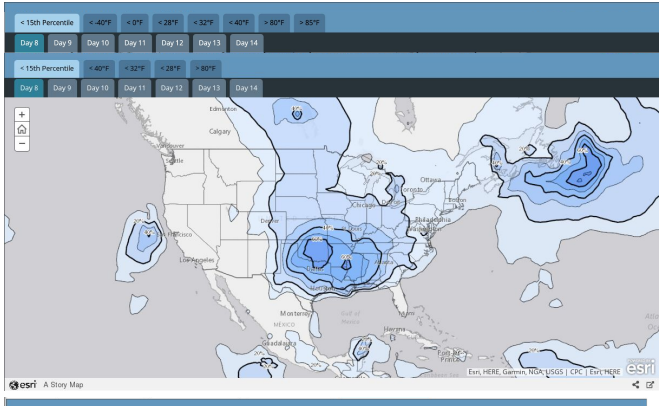
Public tool thresholds and variables

https://www.cpc.ncep.noaa.gov/products/predictions/threats/extremesTool.php



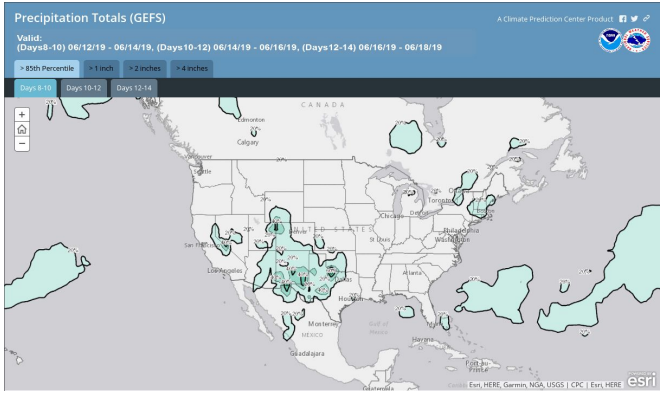
Tmax Thresholds

> 85th ptile
> 80 deg F
> 90 deg F
> 100 deg F
> 110 deg F



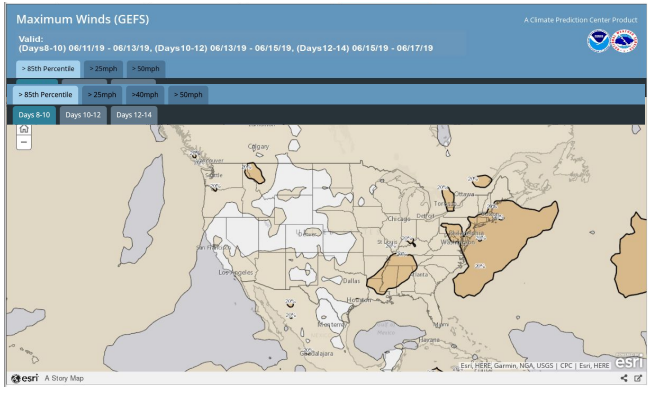
Tmin Thresholds

<15th ptile
< 40 deg F
< 32 deg F
< 28 deg F
< 0 deg F
< -40 deg F
> 80 deg F
> 80 deg F



Precipitation Thresholds

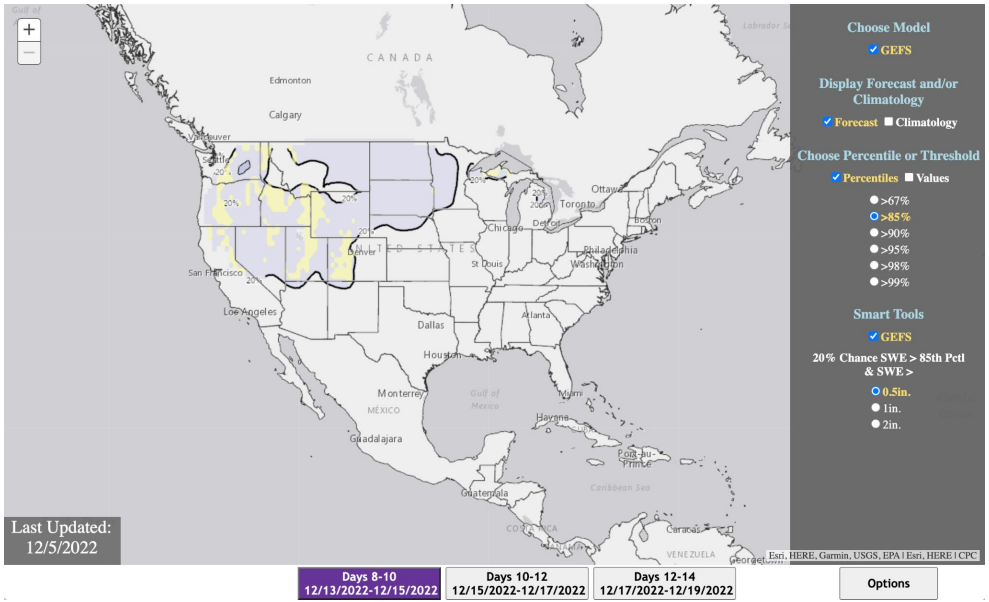
> 85th ptile
> 1”
> 2”
> 4”



Wind Thresholds

>85th ptile
> 25 mph
> 40 mph
> 50 mph

Snow Water Equivalent Forecasts



- Addition of snow water equivalent (SWE) to PET in Sep 2022 (3-day SWE change)
- Hazards forecasters found this extremely helpful for issuing probabilistic snow hazards

Sample screenshot of internal snow water equivalent (SWE) tool (currently only internal)

PET Data

- Real-time model data: GEFSv12
  - Resolution: 1 Deg (0.5 Deg for snow)
  - Number of ensemble members: GEFS 31 members
  - Based on the 0Z update cycle
- Training data:
  - Reforecasts: GEFS - 20 Years from 2000-2019
  - Observations:
    - Tmax and Tmin - CPC's gridded max and min temperature datasets based on GTS station data over land, supplemented w/ first day lead of GEFS reforecast data over water
    - Precipitation - CPC unified gauge dataset
    - Winds - Climate Data Assimilation System (CDAS)
    - Snow - Snow Data Assimilation System (SNODAS)
- Climatology data:
  - Derived from historical observations for each variable
  - 30 year base period - 1991-2020
- Temporal aggregation:
  - Tmin/Tmax is daily; Precipitation, SWE, and winds are aggregated over 3-days (lower predictability)

Details here: https://www.cpc.ncep.noaa.gov/products/predictions/threats/extremesToolAbout.php

PET Methodology

- Bias correction first performed on the target forecasts
  - bias = historical fcst mean - historical obs mean
  - bias correction = real time fcst - bias
- Then calibration applied applied to the paired historical reforecasts and observations to create more reliable probabilities using ensemble regression (Unger, 2009)
  - A benefit is can use less members in training period to get stats (ensemble mean) and apply stats to more members in the real-time, retaining more info from individual members (Ou et al., 2016)
  - Outcome is fully calibrated-probability distribution of the ensemble forecast, with forecast probability of exceedance (POE) with 19 reference percentiles
  - Subset of the 19 percentiles are depicted in the PET outlook maps, that are deemed to be of most interest and skillful
- Actual value thresholds are obtained by interpolating to percentiles to values using climatology

Verification Methodology

- Use DTC's Model Evaluation Tools (MET); Benefit: many available skill metrics, especially for extremes
- A Python app (MET-Python-Utils) was created as a flexible framework to format data for MET, run MET, and plot results, in addition to custom reliability code for multiple models, post-processing types, variables, etc.
- **Critical Success Index (CSI, also known as the Threat Score)** and False Alarm Rate (**FAR**) skill shown for probabilities ***>20% chance of >85th percentile*** (thresholds commonly used by forecasters) and **reliability**
- Shortest week-2 leads presented (day 8 for temperature, days 8-10 for snow)
- Compare skill of **raw** vs. bias-corrected (**BC**) vs. BC+calibrated (**Rfcst-Cal**) GEFS for 3-day accumulated SWE change, Tmin, and Tmax
- Verification for 6-month cool/warm seasons for 2021-2023 (cool: 10/01/2021-03/31/2022, 10/01/2022-03312023 | warm: 04/01/2021-09/30/2021, 04/01/2022-09/30/2022)

Verification Methodology

From MET documentation: Subscripts represent forecast and observed, respectively, with 1 denoting occurrence and 0 non-occurrence

CSI is defined as

Fcst Obs

$$\text{CSI} = \frac{n_{11}}{n_{11} + n_{10} + n_{01}}.$$

It is the ratio of the number of times the event was correctly forecasted to occur to the number of times it was either forecasted or occurred. **CSI** ignores the "correct rejections" category (i.e., $\mathbf{n_{00}}$). CSI is also known as the Threat Score (TS). CSI can also be written as a nonlinear combination of POD and FAR, and is strongly related to Frequency Bias and the Base Rate.

Called "FAR" in CTS output Table 11.4

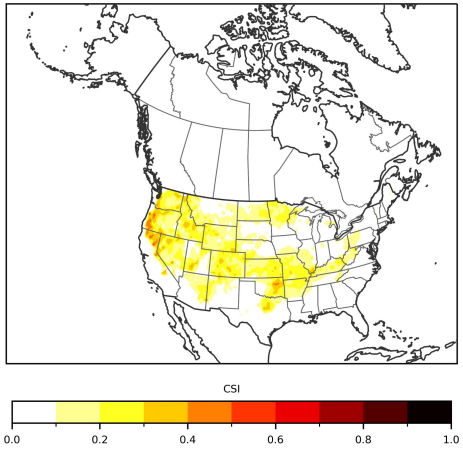
FAR is defined as

$$\text{FAR} = \frac{n_{10}}{n_{10} + n_{11}} = \frac{n_{10}}{n_{1.}}.$$

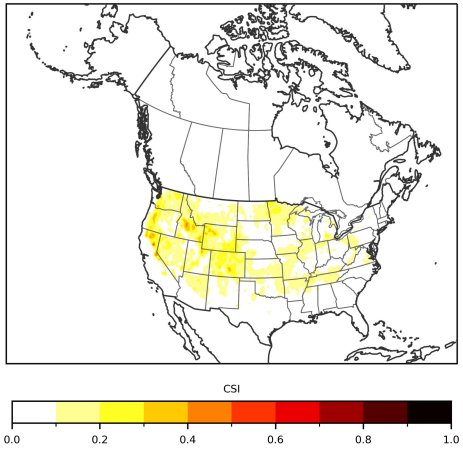
It is the proportion of forecasts of the event occurring for which the event did not occur. FAR ranges from 0 to 1; a perfect forecast would have FAR = 0.

SWE Verification Results - 2021/2022

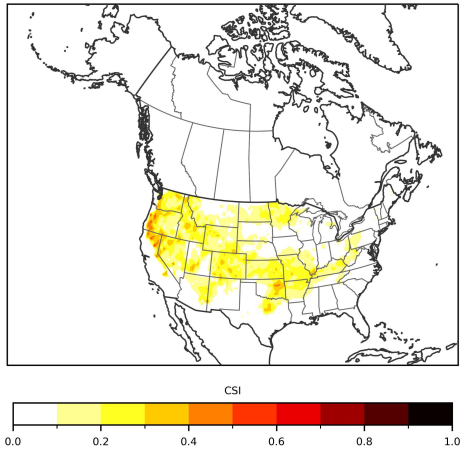
CSI of gefs 00z rfcst-cal 8-10day swe issued 20211001 to 20220331



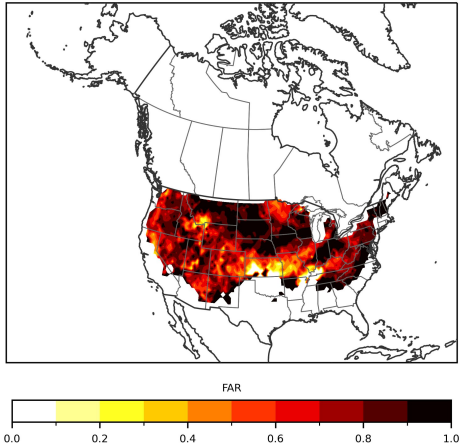
CSI of gefs 00z rfcst-bc 8-10day swe issued 20211001 to 20220331



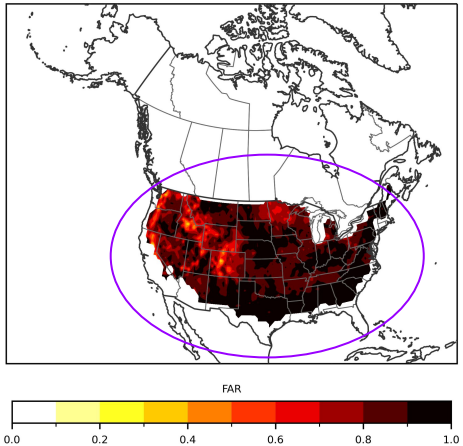
CSI of gefs 00z rfcst-cal 8-10day swe issued 20211001 to 20220331



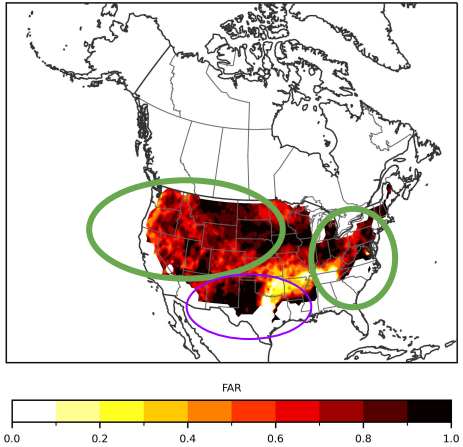
FAR of gefs 00z raw 8-10day swe issued 20211001 to 20220331



FAR of gefs 00z rfcst-bc 8-10day swe issued 20211001 to 20220331



FAR of gefs 00z rfcst-cal 8-10day swe issued 20211001 to 20220331

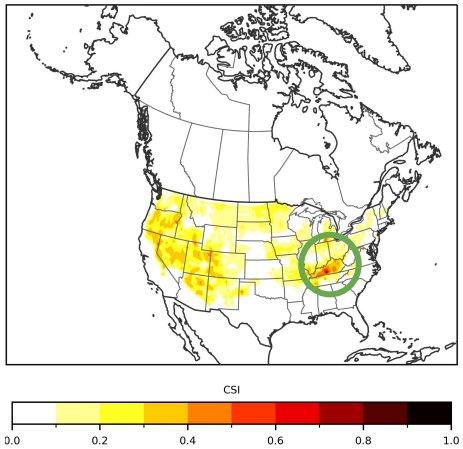


Purple: Degradation
Green: Improvements

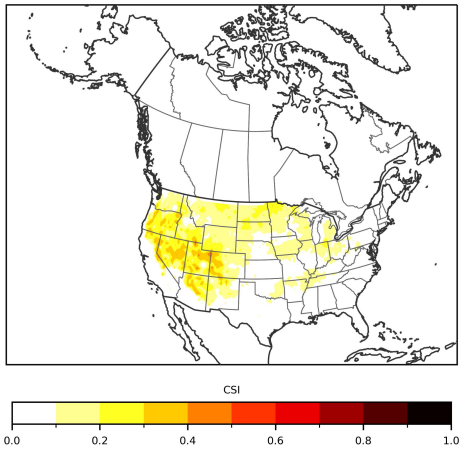
- CSI fairly similar
- Rfcst-cal brings too much snow southward to south-central but improvements for other areas
- BC much worse! Why?
- Reforecast issues?
- Climatology of e.g. 30 years ago may be much colder than the recent warmer period that could impact bias correction/calibration?
- Rcst-cal improves upon BC

SWE Verification Results - 2022/2023

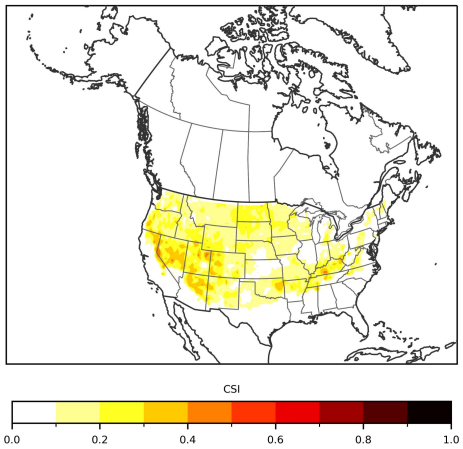
CSI of gefs 00z raw 8-10day swe issued 20221001 to 20230331



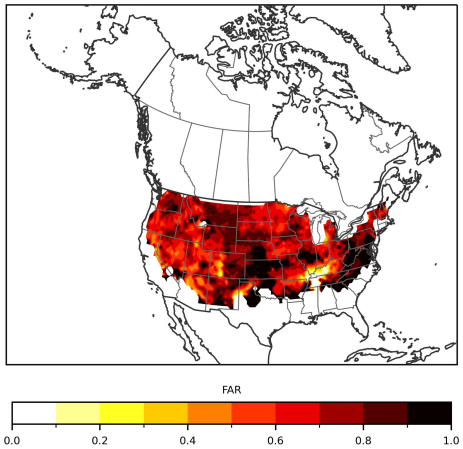
CSI of gefs 00z rfcst-bc 8-10day swe issued 20221001 to 20230331



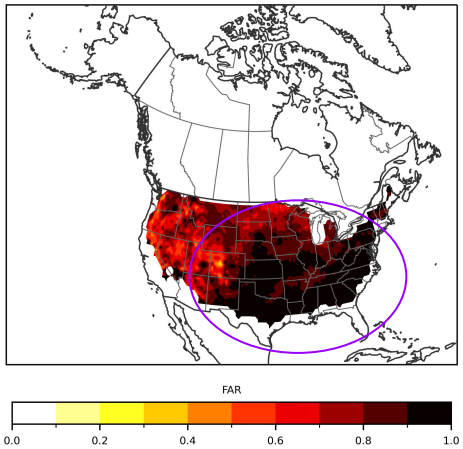
CSI of gefs 00z rfcst-cal 8-10day swe issued 20221001 to 20230331



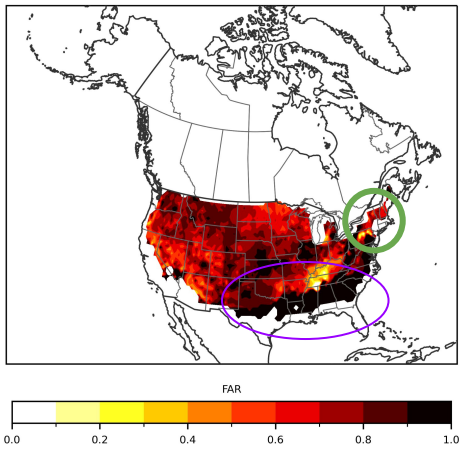
FAR of gefs 00z raw 8-10day swe issued 20221001 to 20230331



FAR of gefs 00z rfcst-bc 8-10day swe issued 20221001 to 20230331

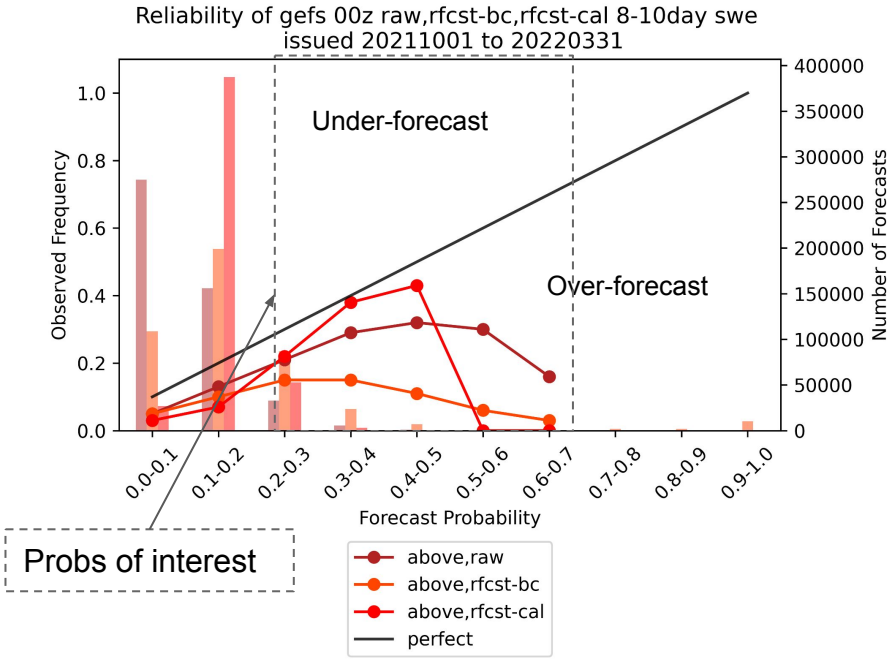


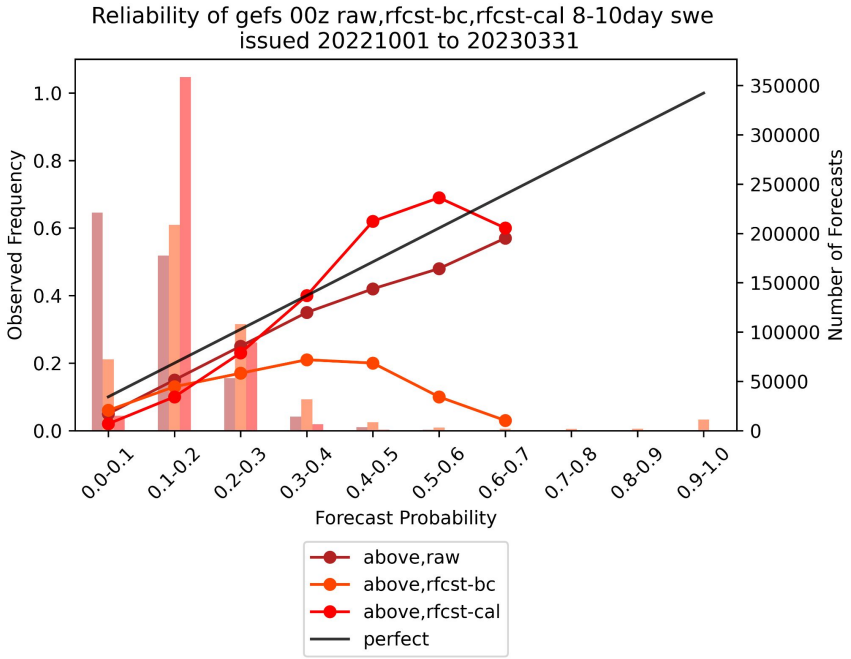
FAR of gefs 00z rfcst-cal 8-10day swe issued 20221001 to 20230331



- Higher CSI across the West than previous year, possibly just more heavy snow events

SWE Verification Results

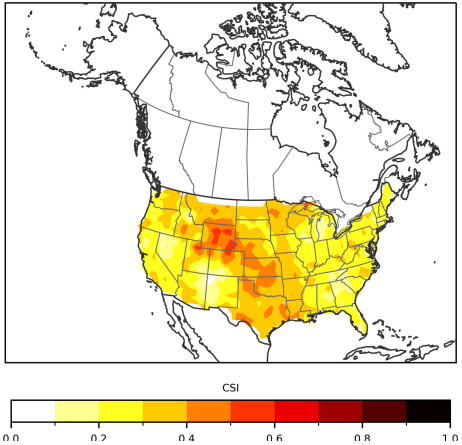




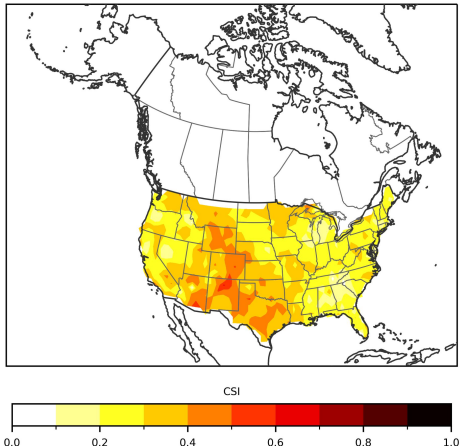
- Significantly improved reliability of Rfcst-cal over raw forecasts at probabilities (probs) above 30%
- Over-forecasting of snow likely due to known bug in PTYPE
- 2022-2023 season calibration resulted in under-forecasting for probs > 40%, but not for the previous season
- Similar to CSI and FAR, BC degrades reliability but calibrating significantly improves it, this is noticed for multiple variables - systematic issues in reforecast?

Tmin Verification Results - 2021/2022

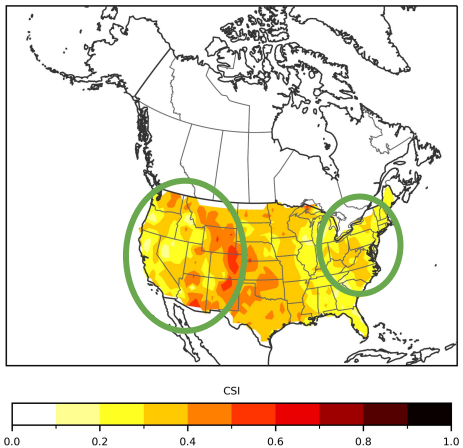
CSI of gefs 00z raw day8 tmin
issued 20211001 to 20220331



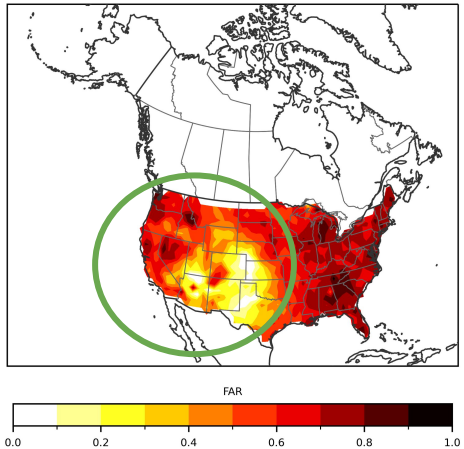
CSI of gefs 00z rfcst-bc day8 tmin
issued 20211001 to 20220331



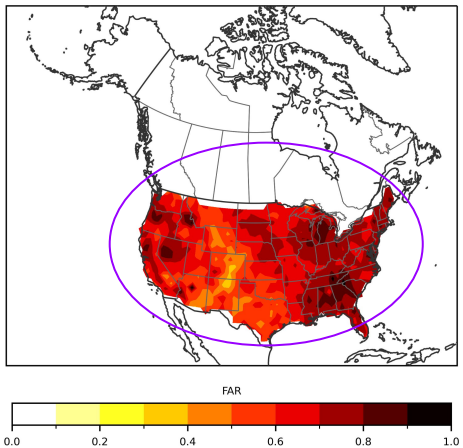
CSI of gefs 00z rfcst-cal day8 tmin
issued 20211001 to 20220331



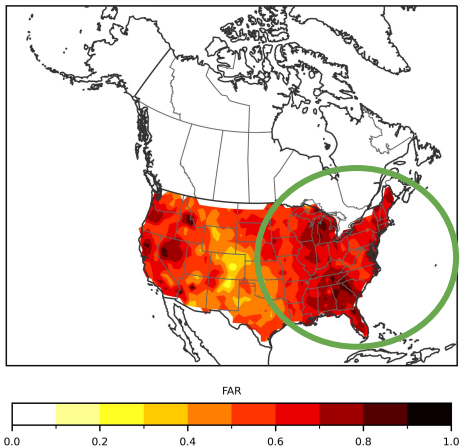
FAR of gefs 00z raw day8 tmin
issued 20211001 to 20220331



FAR of gefs 00z rfcst-bc day8 tmin
issued 20211001 to 20220331



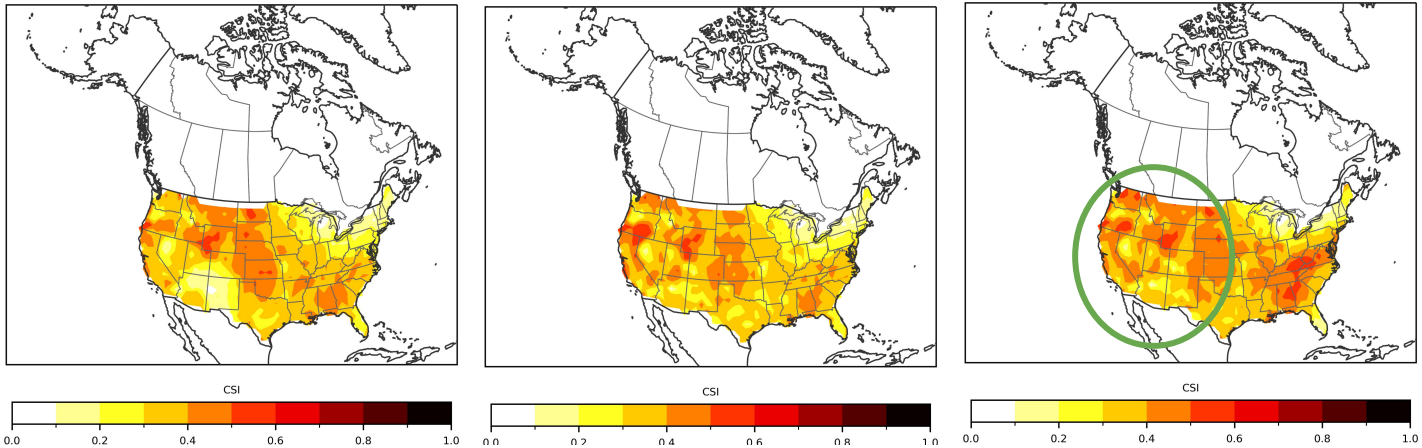
FAR of gefs 00z rfcst-cal day8 tmin
issued 20211001 to 20220331



- Rfcst-cal shows widespread improvement in CSI compared to raw across much of the CONUS, especially for the West
- Raw: lowest FAR over West, Rfcst-cal lower over East (BC worst)

Tmin Verification Results - 2022/2023

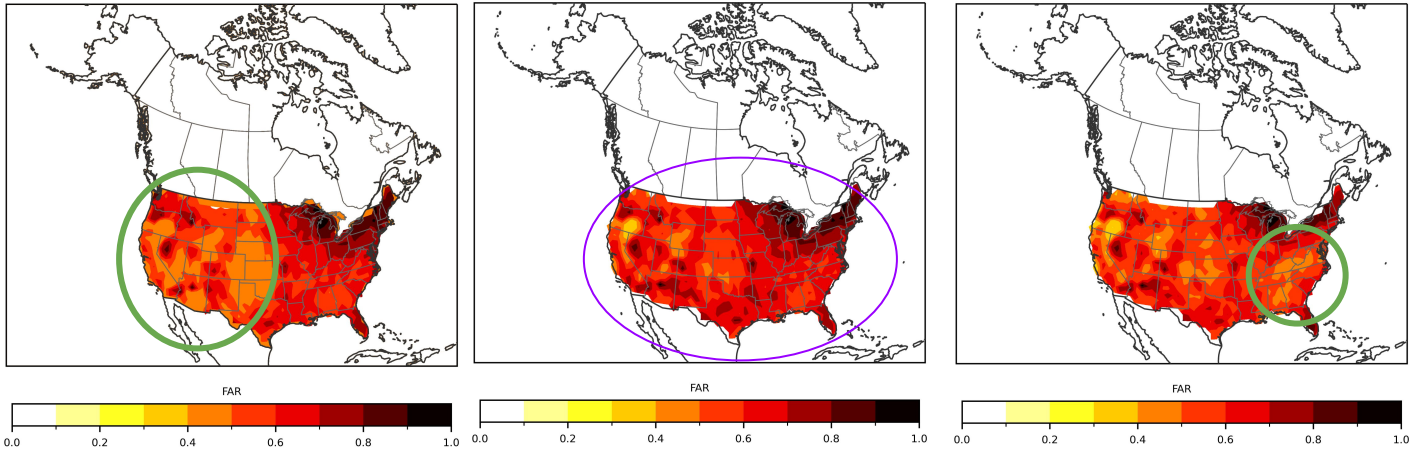
CSI of gefs 00z raw day8 tmin
issued 20221001 to 20230331



CSI of gefs 00z rfcst-bc day8 tmin
issued 20221001 to 20230331

CSI of gefs 00z rfcst-cal day8 tmin
issued 20221001 to 20230331

FAR of gefs 00z raw day8 tmin
issued 20221001 to 20230331

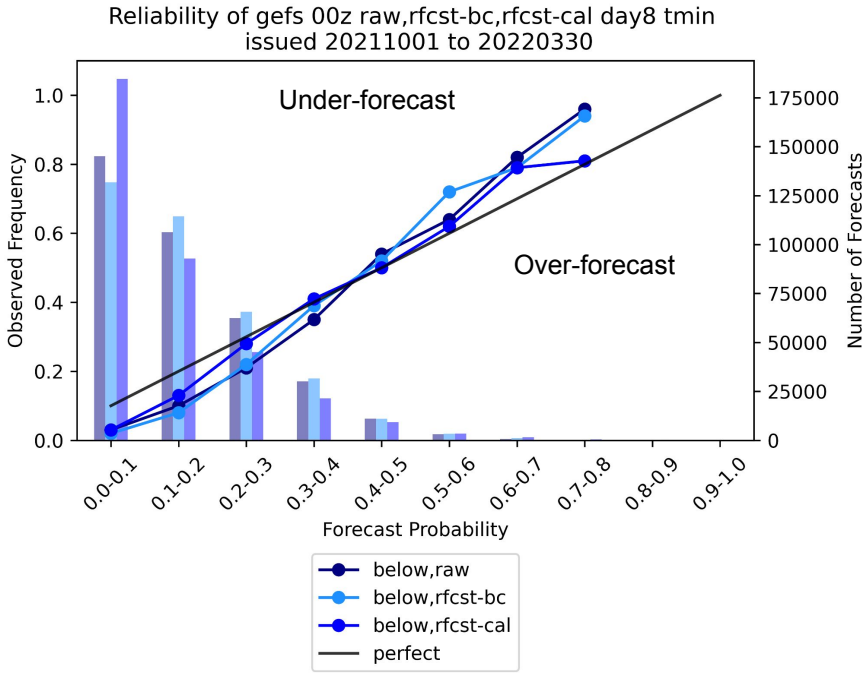


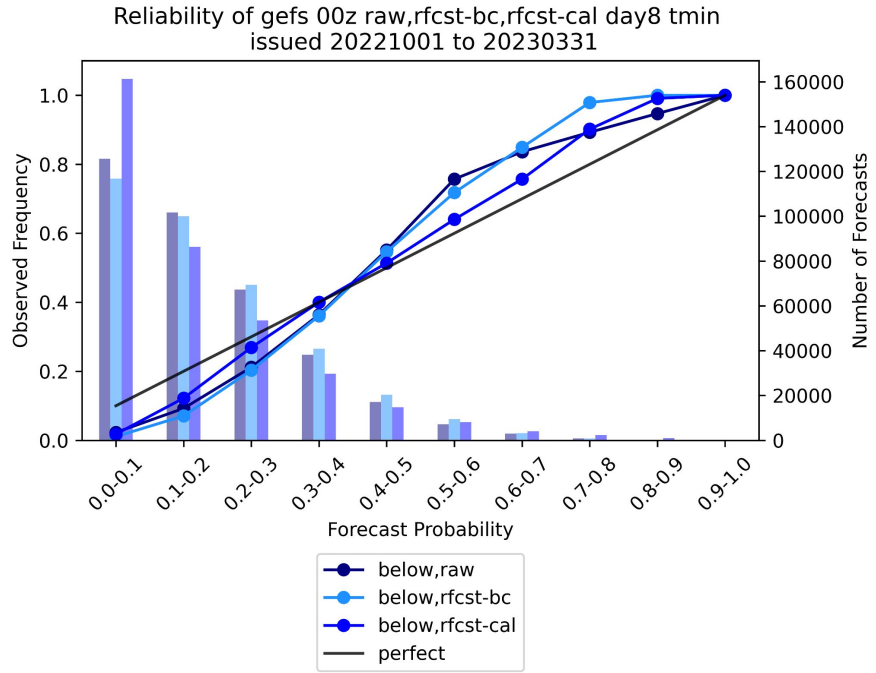
FAR of gefs 00z rfcst-bc day8 tmin
issued 20221001 to 20230331

FAR of gefs 00z rfcst-cal day8 tmin
issued 20221001 to 20230331

- Similar results to previous year
- Higher CSI across the West compared to previous year

Tmin Verification Results

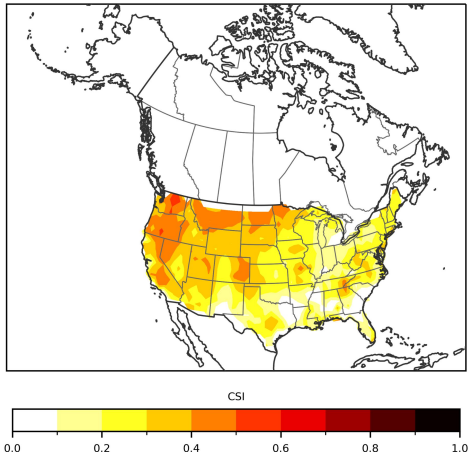




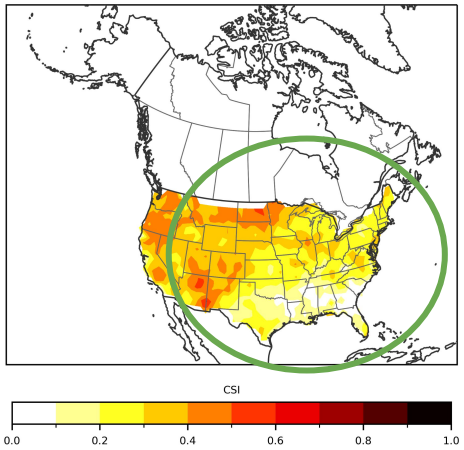
- Rfcst-cal has noticeable improvement in reliability
- Generally Tmin is underforecast at probabilities >50%

Tmax Verification Results- 2021/2022

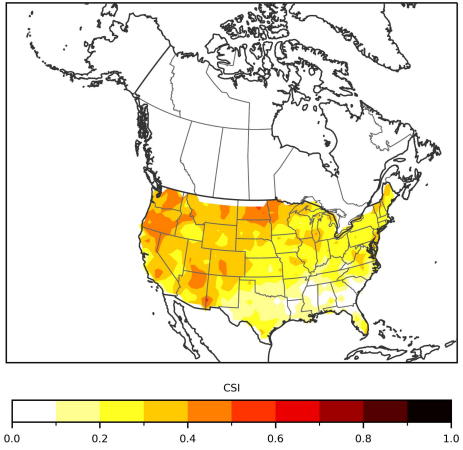
CSI of gefs 00z raw day8 tmax
issued 20210401 to 20210930



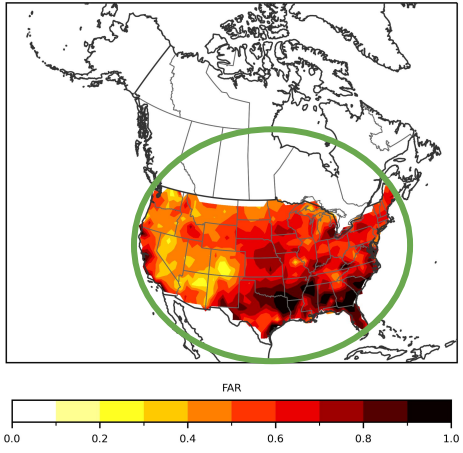
CSI of gefs 00z rfcst-bc day8 tmax
issued 20210401 to 20210930



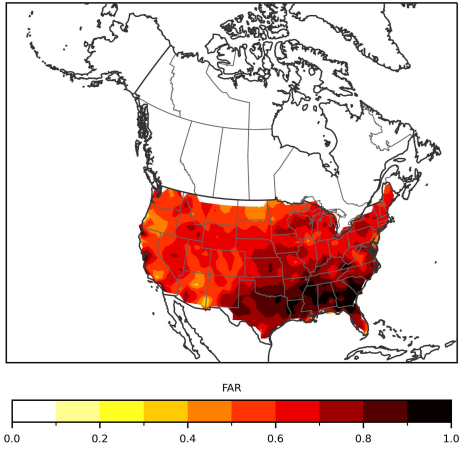
CSI of gefs 00z rfcst-cal day8 tmax
issued 20210401 to 20210930



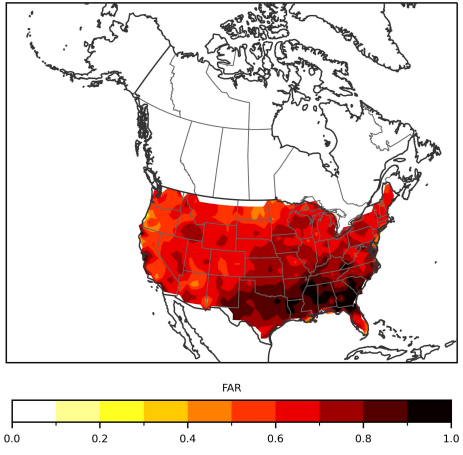
FAR of gefs 00z raw day8 tmax
issued 20210401 to 20210930



FAR of gefs 00z rfcst-bc day8 tmax
issued 20210401 to 20210930

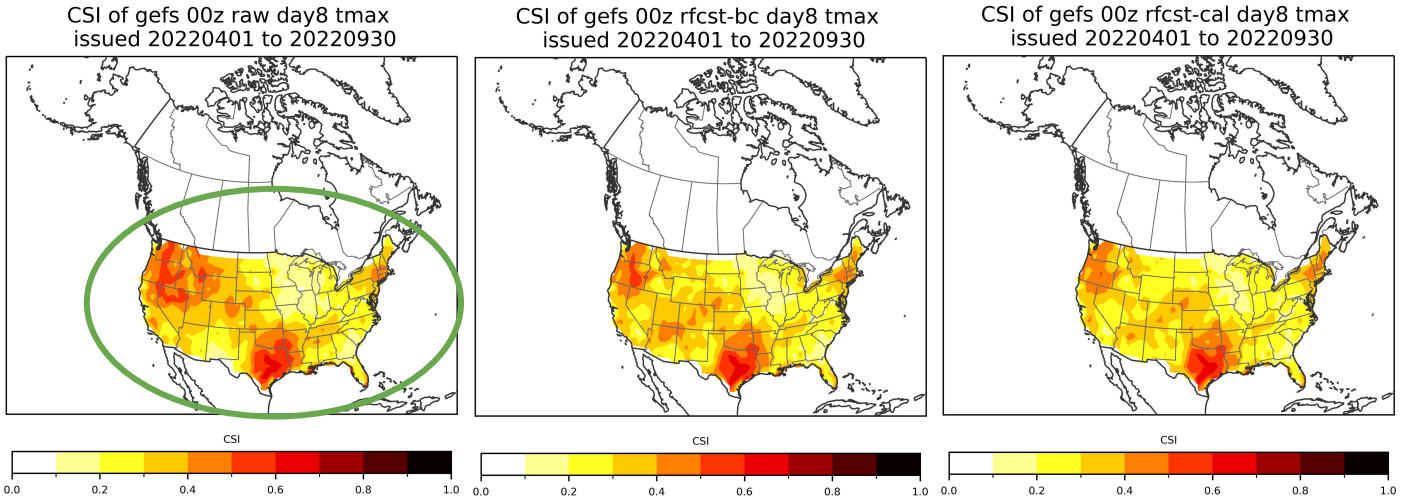


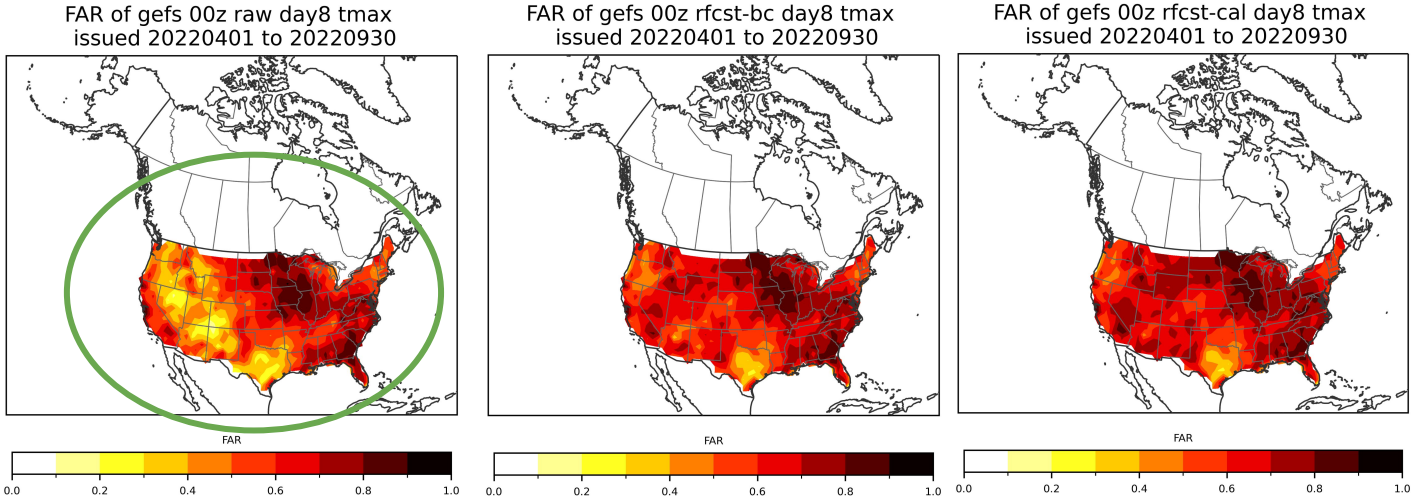
FAR of gefs 00z rfcst-cal day8 tmax
issued 20210401 to 20210930



- Interestingly, BC has the greatest CSI but raw has the lowest FAR
- Greatest FAR over south-central and Southeast

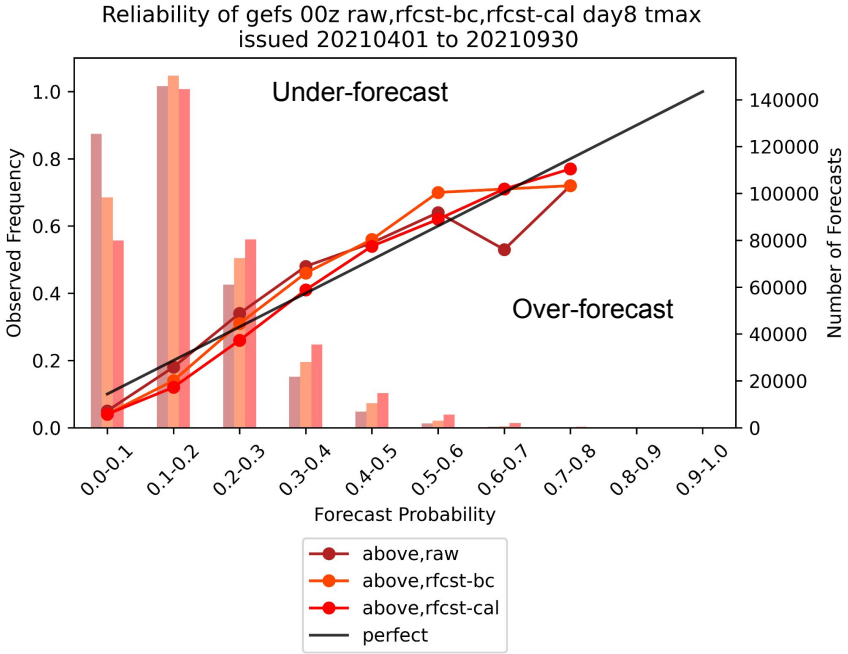
Tmax Verification Results - 2022/2023

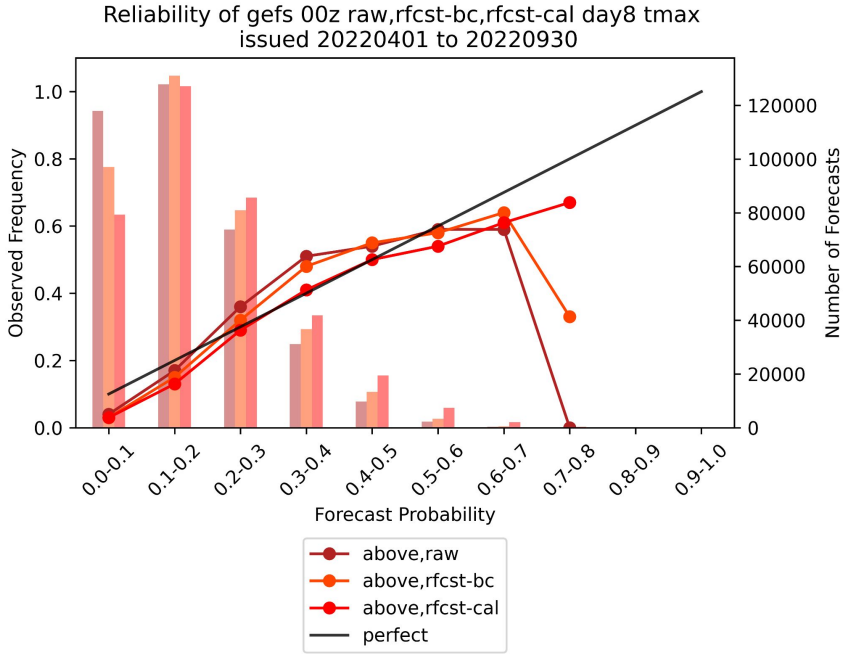




- CSI and FAR better for raw than BC and Rfcst-cal
- climatology too cold further back in reforecast compared to current conditions?

Tmax Verification Results





- Improvement of Rfcst-cal best at probs > 30%

Future Work / Gap Needs

- Continue ongoing work performing forecast evaluation across multiple PET vars, leads, models, post-processing types (e.g. skill diff plots) and more detailed analysis to understand skill differences
- Try to figure out what is causing skill degradation in bias correction - reforecast issues? Too long of a training period where the beginning does not reflect current climate well? Test different periods of BC, remove BC just do calibration

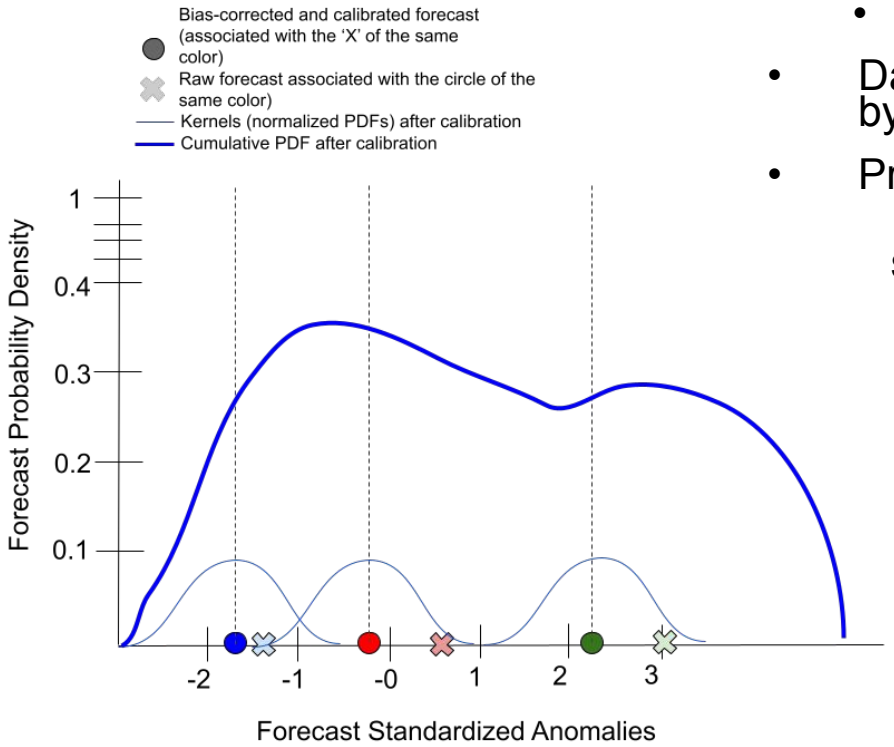
GEFS real-time/reforecast needs:

- Improved reforecasts - reanalysis with consistent initialized conditions (currently we cannot use before 2000 due to this difference)
- 6-hr max wind speed values
- Output of explicit snowfall accumulations in addition to SWE
- Fix for known GEFS PTYPE bug -> over-forecast of snow

Thank You!

Contact melissa.ou@noaa.gov or any comments or questions!

Ensemble Regression

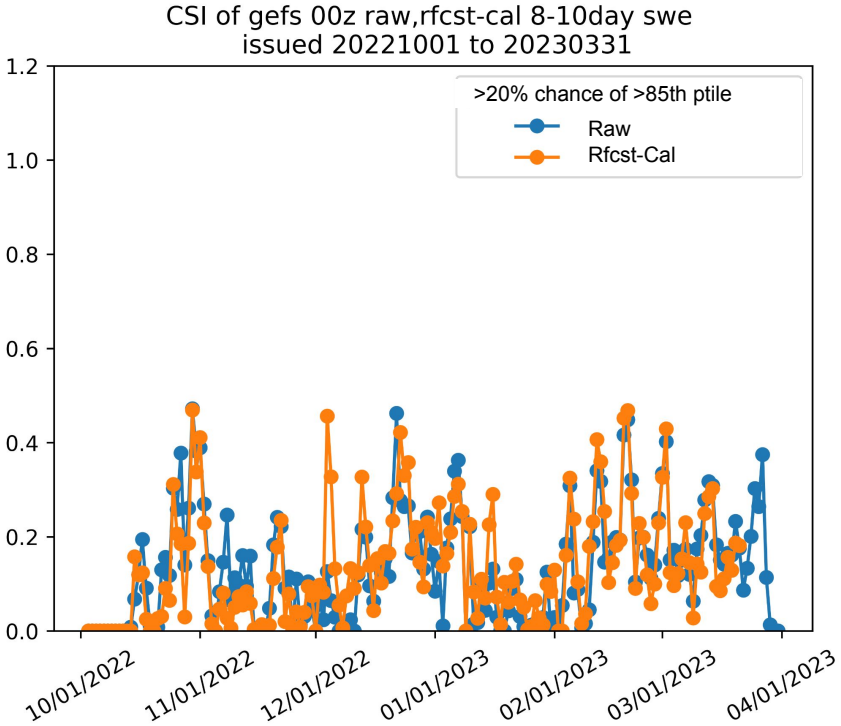


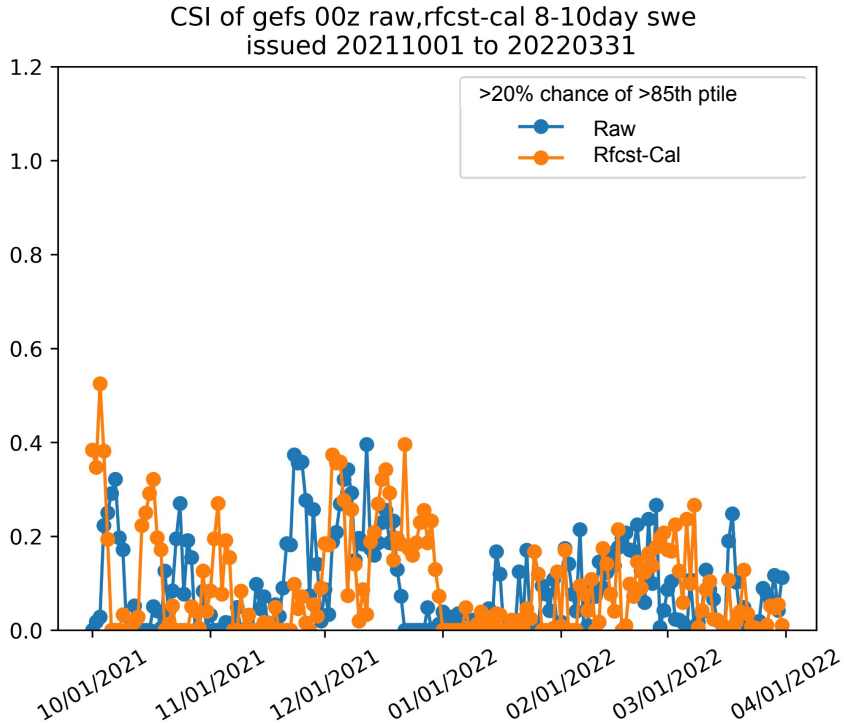
- Remove mean bias by removing model climatology
- Corrects the variance of the ensemble mean
  - $\sigma^2_{\text{Ensemble members}} = Es\text{^}2 + \sigma^2_{\text{Ensemble Mean}}$
- Damps forecasts towards observed climatology by skill
- Produces reliable probability forecasts

Steps (see diagram)

1) Raw individual ens members ("X"s) get bias-corrected and calibrated using precalculated stats based on historical reforecasts and observations, via linear regression. Results in adjusted value (Circles).
2) Kernels (normalized PDFs) are created based on the new adjusted ens values.
3) A cumulated PDF is created over all the kernels, which is converted to CDF -> probability of exceedance data at 19 reference percentiles

SWE Verification Results





- ** Put mean line through here