Decadal Trend Aware Calibration of the NMME

Dan Collins, Emerson LaJoie, and Johnna Infanti

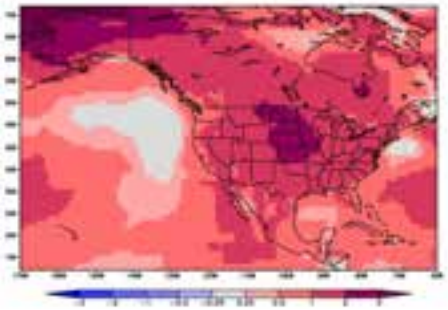
NOAA Climate Prediction Center

Outline

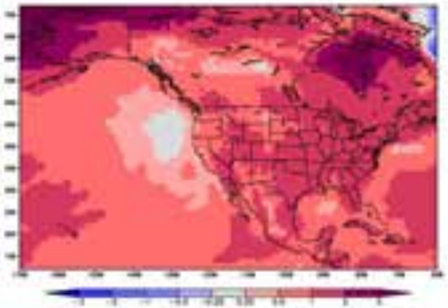
- Background
- Observed and NMME model decadal trends
- Two methods of post-processing of seasonal forecasts
  - Ensemble Regression (ER, Unger et al. 2009)
  - Regression explicitly including decadal trends (ER+T)
- Deterministic skill and decadal and other signals
- Probabilistic skill
- Frequency of above/below normal forecasts and observations
- Conclusions

Background

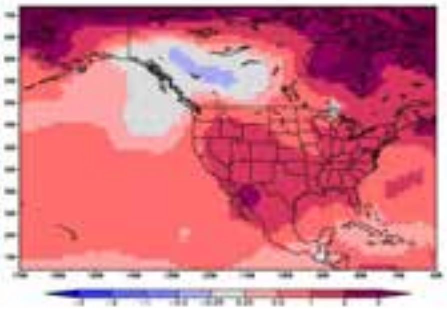
- Three-category / tercile NMME model hindcasts
  - 1991-2020
  - 1-month-lead, e.g., DJF predicted from November 1
- Linear regression based calibrated forecasts
- *Leave 3-years-following-out cross validation*
  1) *Leave out 1991, 1992 and 1993*
  2) *Train statistical model using 1994-2020*
  3) *Predict 1991 seasonal temperature*
- Forecast signals from *cross-validated correlation, R*
- Probabilistic categorical forecast skill from Heidke Skill Scores
  - Percent increase in hit rate of each category, compared to random forecasts
- Decadal trends from linear statistical model (cross-validated)



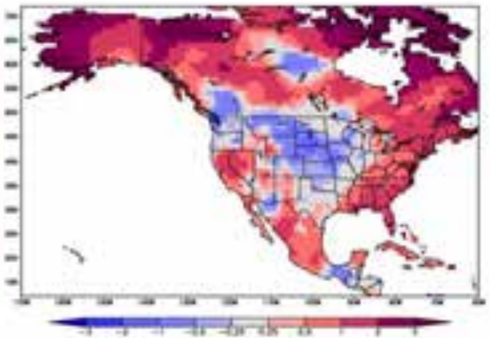
CFSv2



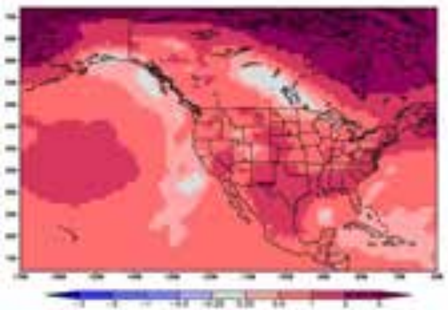
GFDL



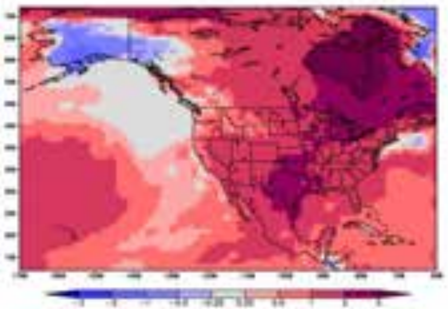
CanCM4i



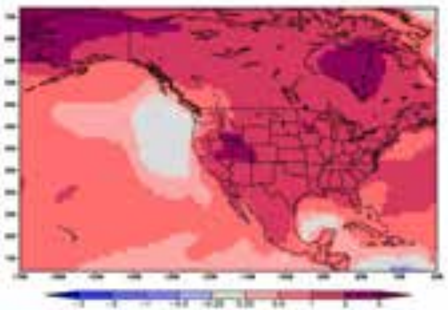
Observed



GEM5 NEMO



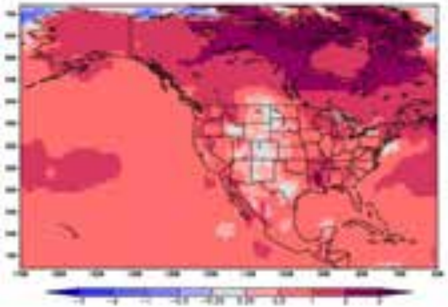
NASA



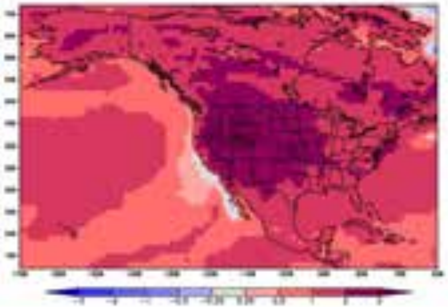
NCAR

Comparing **Linear Trend** based on **NMME model forecasts** for **lead-1 (example) DJF** forecasts of temperature from November initial conditions versus GHCN-CAMS observation trends (upper right):

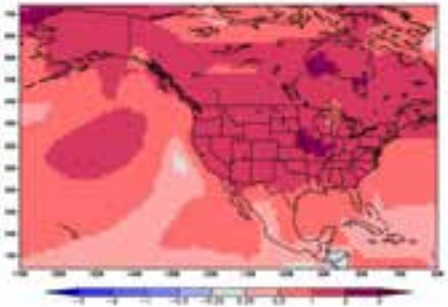
Generally dynamical model trends are more uniformly towards warmer conditions across the domain and greater than observed trends in many areas of North America.



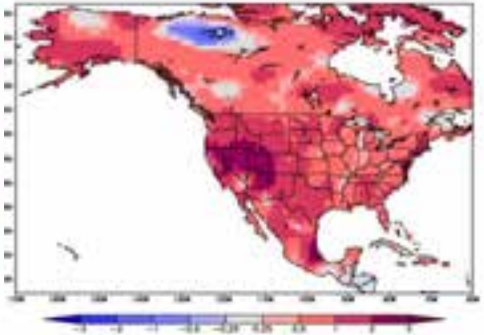
CFSv2



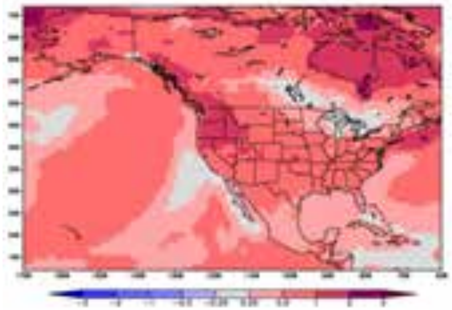
GFDL



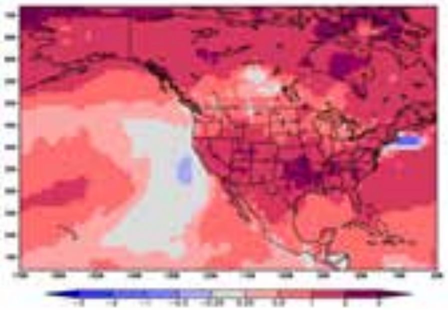
CanCM4i



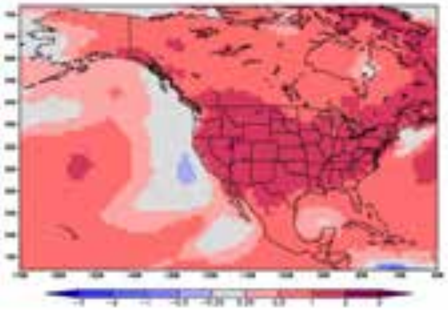
Observed



GEM5 NEMO



NASA



NCAR

Comparing **Linear Trend** based on **NMME model forecasts** versus GHCN-CAMS observation trends for **lead-1 JJA** forecasts of temperature from May initial conditions:

Warm season trends are more uniformly positive in both models and observations.

Ensemble Regression Calibration of Dynamical Model Probability Forecasts

- Derive regression equation between the dynamical model ensemble mean and the observations.
- Apply same linear equation to each dynamical ensemble member.

$$F^*_{(m)} = aF_{(m)} + b$$

- Expected residual error of the regression of the ensemble mean is related to the anomaly correlation.

$$[MSE] = \sigma_{obs}^2 (1 - R_m^2)$$

- Residual error of an ensemble is found by subtracting the average ensemble spread from the average mean square error (MSE).

$$\sigma_\varepsilon^2 = [MSE] = \sigma_{ens}^2 + \varepsilon^2 \qquad \varepsilon^2 = \sigma_\varepsilon^2 - \sigma_{ens}^2$$

Ensemble regression calibration of dynamical models uses a standard linear regression model of the form:

$$F^*=a+bF$$

F = model forecast

F^*= corrected model forecast

We propose to use a decadal trend predictor T, in addition to model forecasts F (Krakauer, 2019):

$$F^*=c+dF+eT$$

- ***Trend predictor***: Any slowly varying representation of decadal variability
    - *CPC commonly uses the Optimum Climate Normal (Huang, van den Dool, & Barnston 1996)*
    - *For this initial analysis we use the linear variation of observed temperatures with time*

Standardizing predictors and the predictand to simplify the regression equations:

The coefficients that weight the forecast according to (1) dynamical model forecast and (2) decadal trend predictors become:

$$\beta_F = \frac{(R_{FO} - R_{TF} R_{TO})}{(1 - R_{TF}{}^2)}$$

$$\beta_T = \frac{(R_{TO} - R_{TF} R_{FO})}{(1 - R_{TF}{}^2)}$$

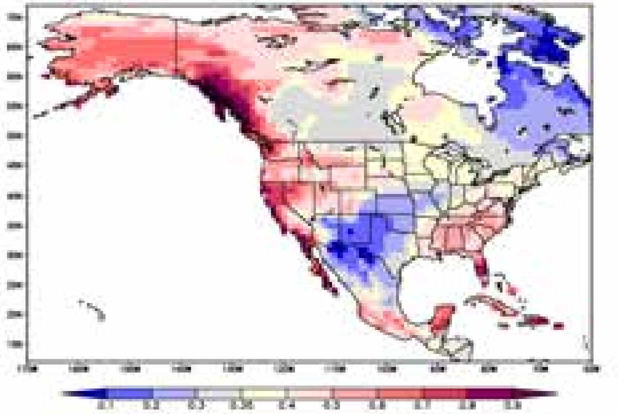
F = model predictor; T = statistical trend predictor; O = observations;

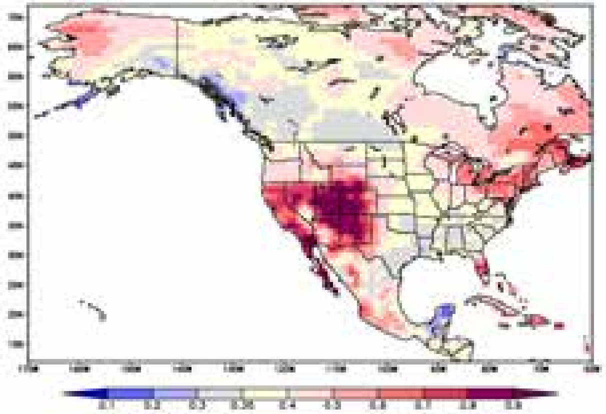
Beta F is the semi-partial correlation of the dynamical model forecasts to observations accounting for trends.

Beta T is the empirical trend forecast correlation that is not in the dynamical models.

Example SON forecasts: Ensemble Regression Calibration of Dynamical Model Probability Forecasts (top) compared to Ensemble Regression plus Trend (bottom)

ER calibrated

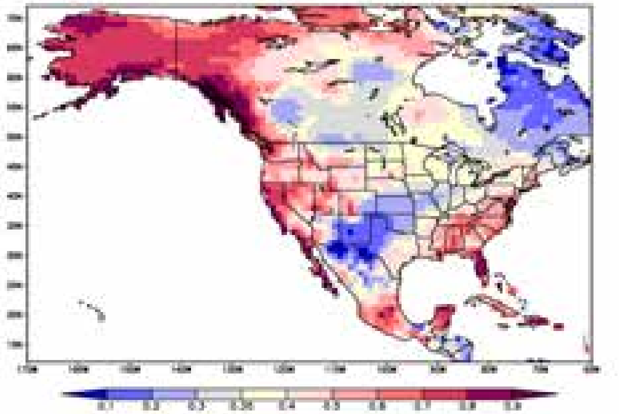


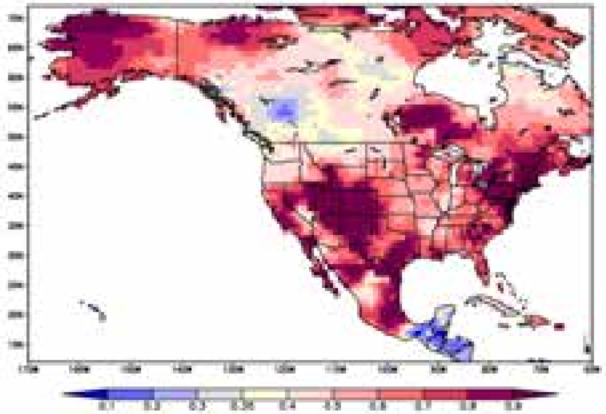


2015

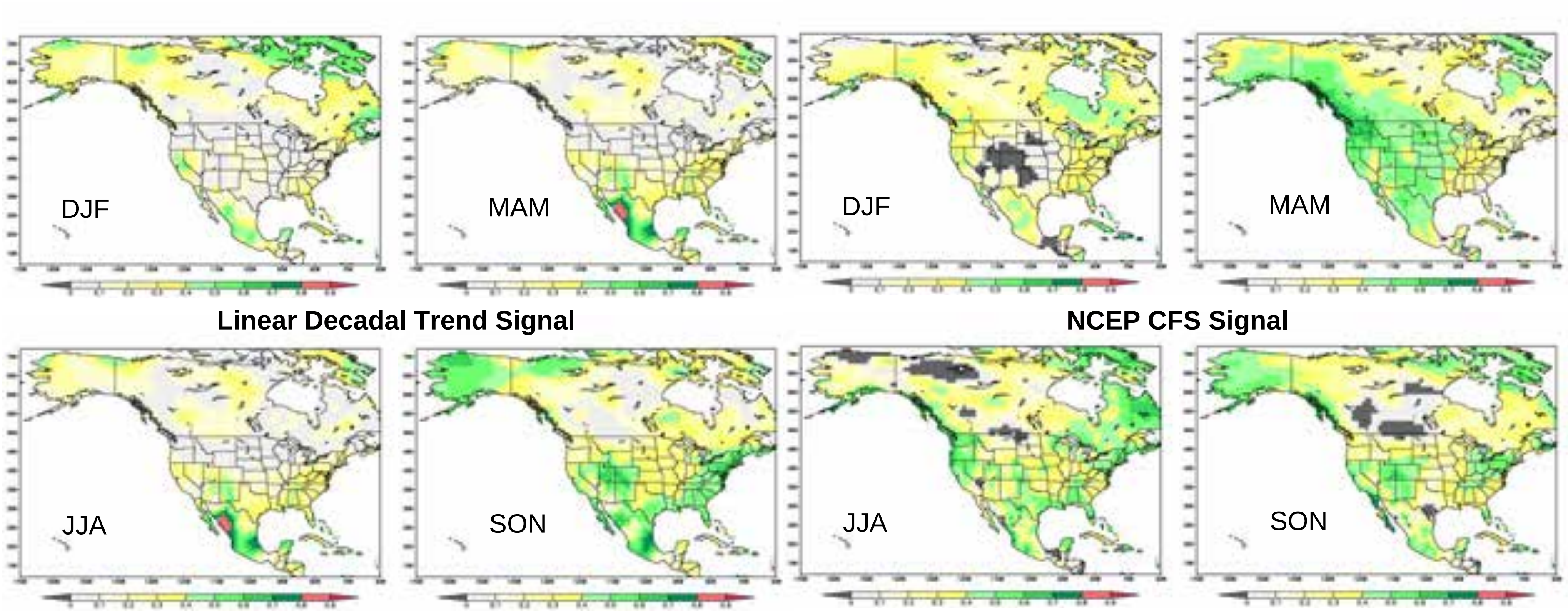
2020

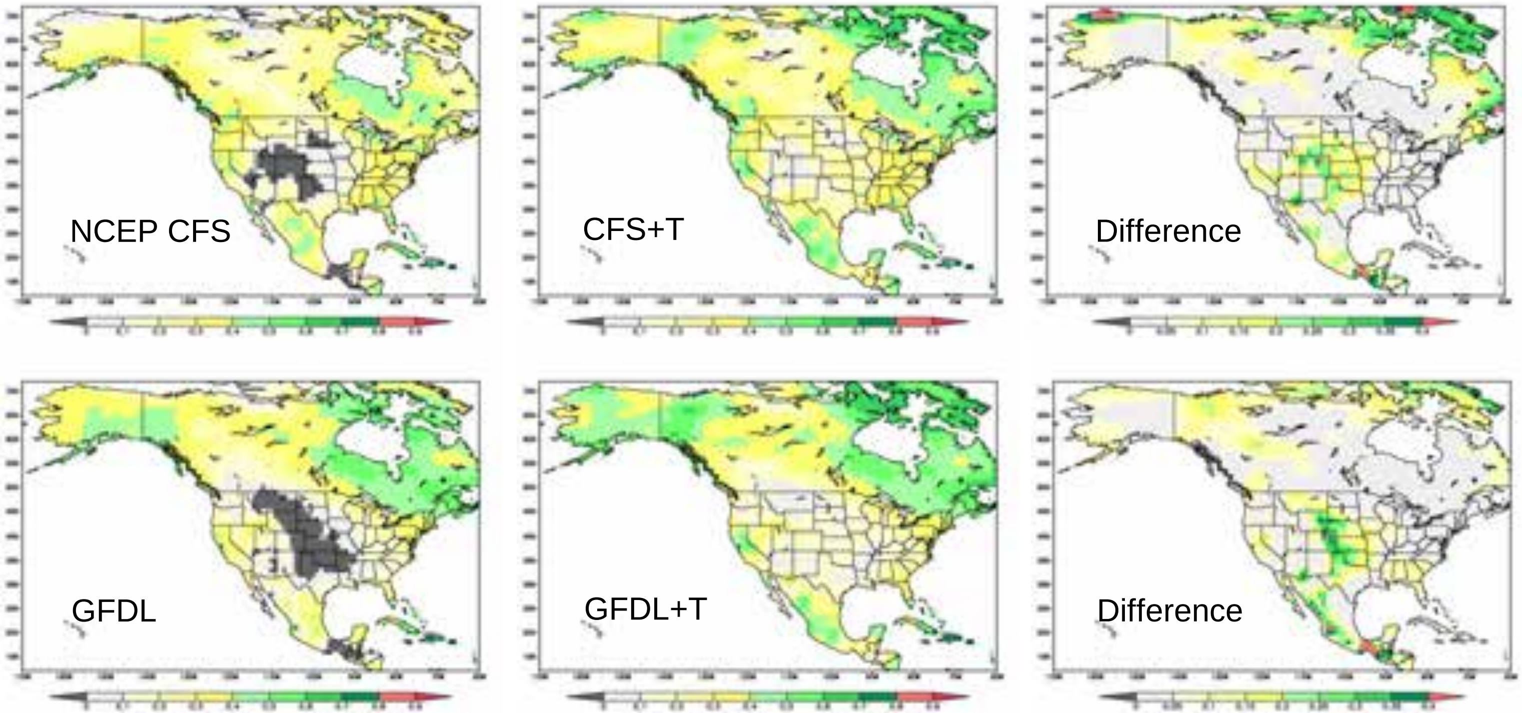
ER + T calibrated





Decadal Trend Signal (Correlation, Leave next 3 years out Cross-validation) (left) compared to
Dynamical Model Signal (right, e.g., NCEP CFSv2)

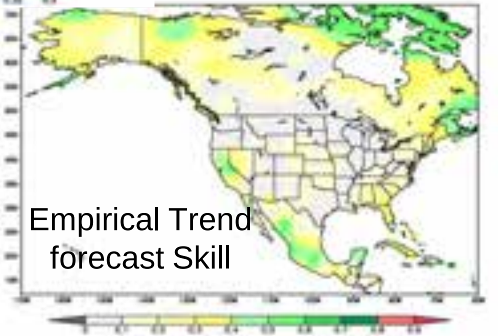


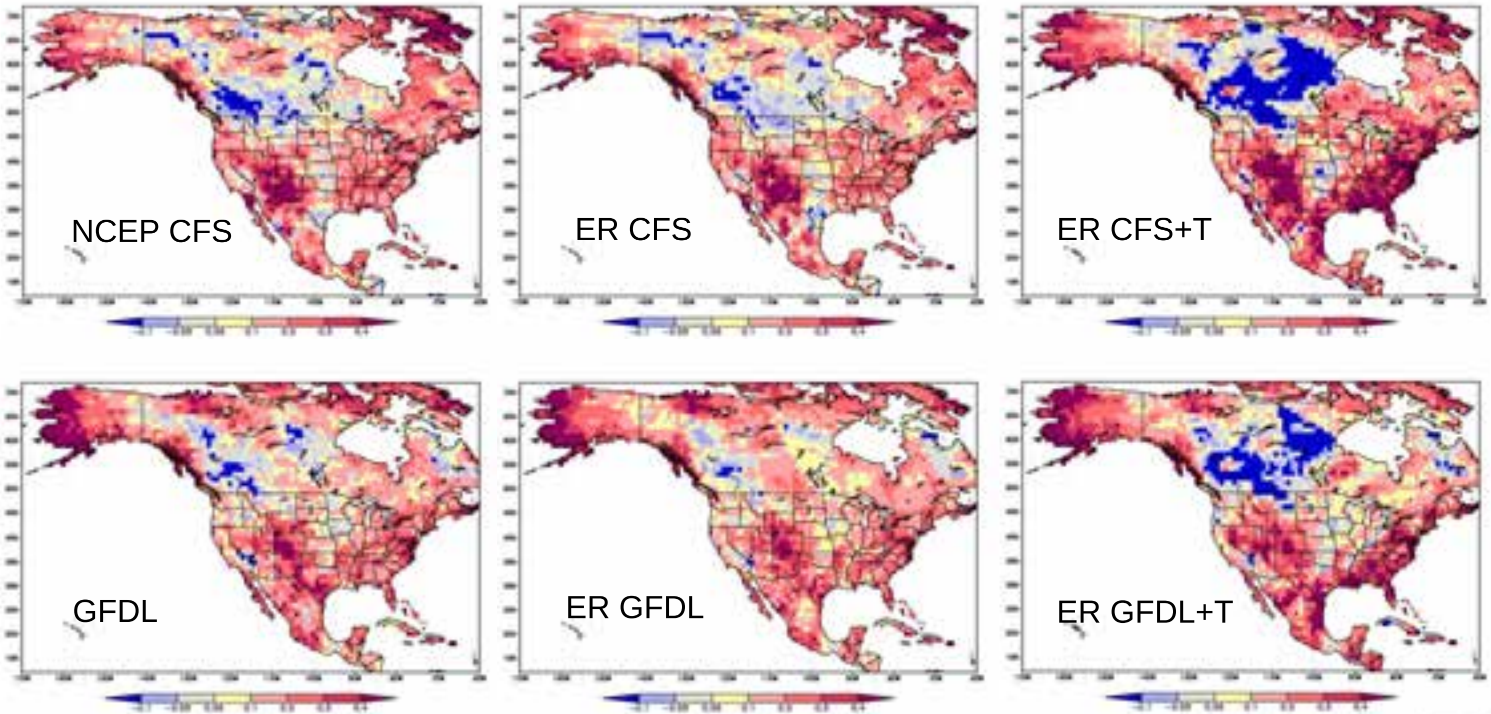


Addition of the decadal trend predictor uniformly increases the apparent **signal**, or remains constant.

No change to signal when model predicts trend or trends are small.

<u>Signal from combined Calibrate Model + Decadal Trends</u>: Anomaly Correlations for **lead-1 DJF**



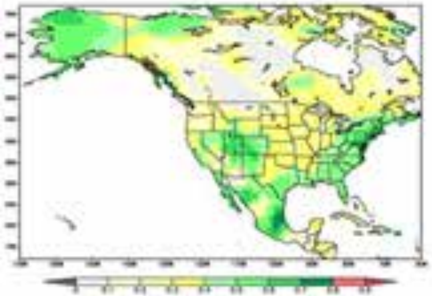


In areas with significant trend signals (bottom right), the statistical trend predictor increases the categorical skill of the probability forecasts (right).

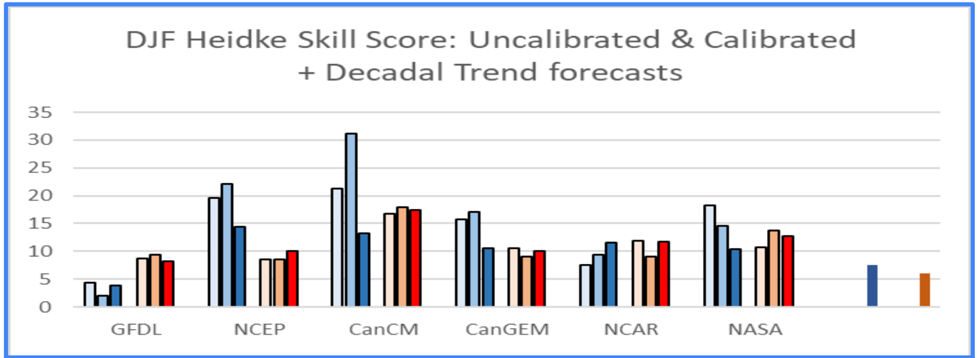
Areas with weak trends are better predicted by the calibrated model forecasts (center).

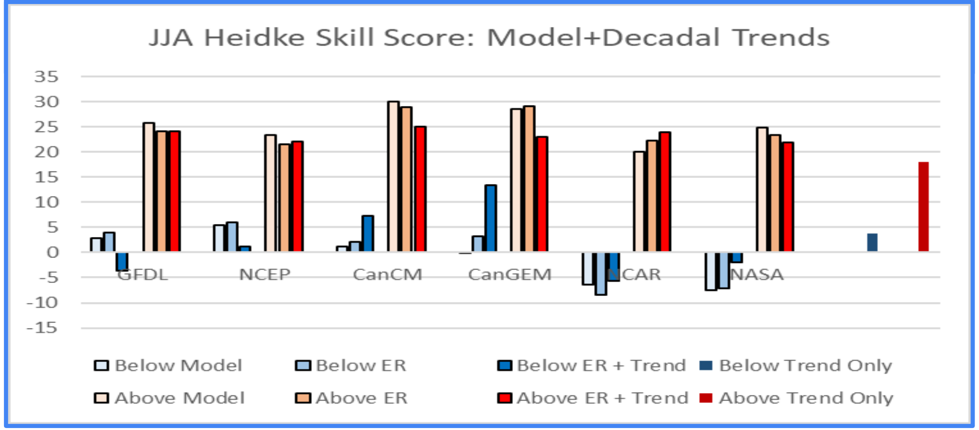
<u>Heidke Skill Score (SON)</u>: 1) Raw ensemble probabilities 2) ER calibrated model 3) ER + Trend calibration

- Leave-3 years out cross validation



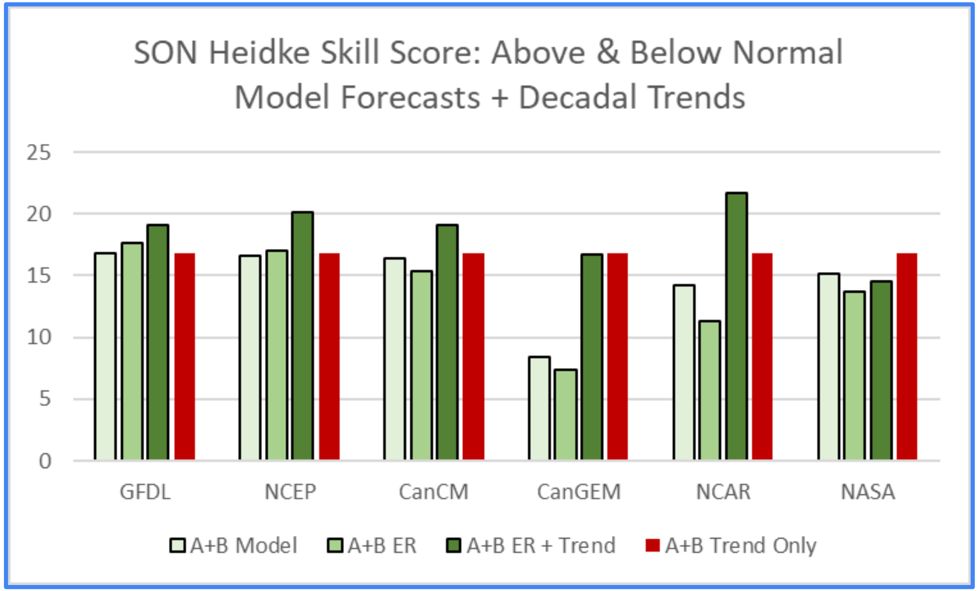
Below Normal (blue) and Above Normal (red) Mean Heidke Skill Score and the Calibration of the NMME models using decadal and interannual signals (DJF, Contiguous U.S. + Alaska, 2006-2020)





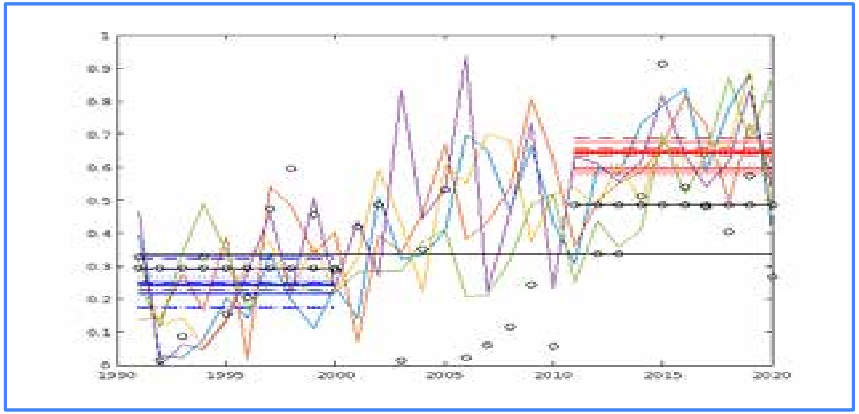
- Calibration using ER+T compared to ER or raw model probability forecasts show generally mixed results
- Generally small changes in skill
- Global averaged Heidke Skill Scores show consistent improved skill

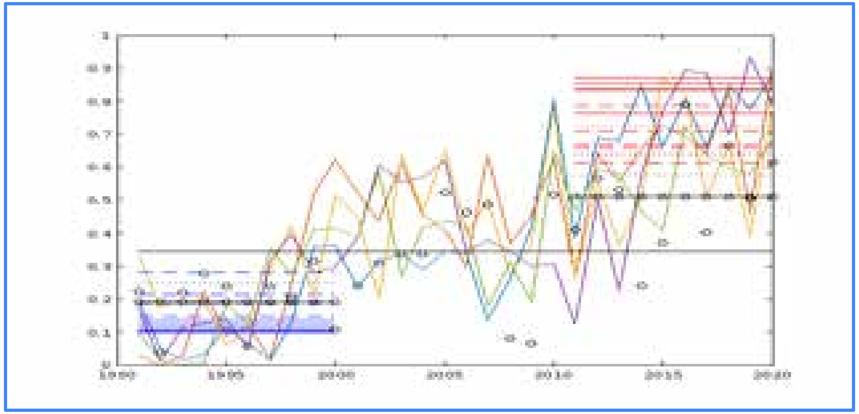
Skill and the Calibration of the NMME models using decadal and interannual signals (SON, 2006-2020, A+B)



- Dynamical models alone have similar skill to a statistical trend forecast (red bars)
- While the Ensemble Regression (ER) + Decadal Trend calibrated models for SON nearly always show improved skill
- This implies skill in both the decadal trends and other model signals

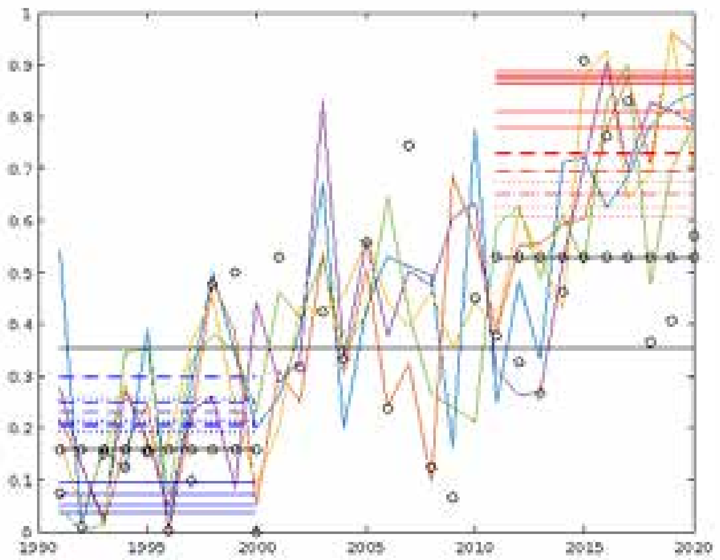
Frequency of Above-normal Forecasts & Observed (circles) (DJF/JJA)





- Below normal is observed more frequently early in the climatology period & above normal is observed more frequently late in the climatology period
- Forecast models predict above normal less frequently than observed early in period (dotted blue lines) and more frequently than observed late in the period (dotted red lines)
- The category frequency bias in dynamical model forecasts (Kirtman et al 2014) appears a necessary component of skillful forecasts;
- Skill-calibrated dynamical model forecasts sometimes exhibit an increase in this bias towards over-predicting the dominant category (dashed red and blue lines)
- The separation of decadal trends from other signals in calibrated model forecasts leads to a reduction/increase in DJF/JJA frequency biases, while increasing total anomaly correlations.

Frequency of Above-normal Forecasts & Observed (SON): Contiguous U.S. + Alaska



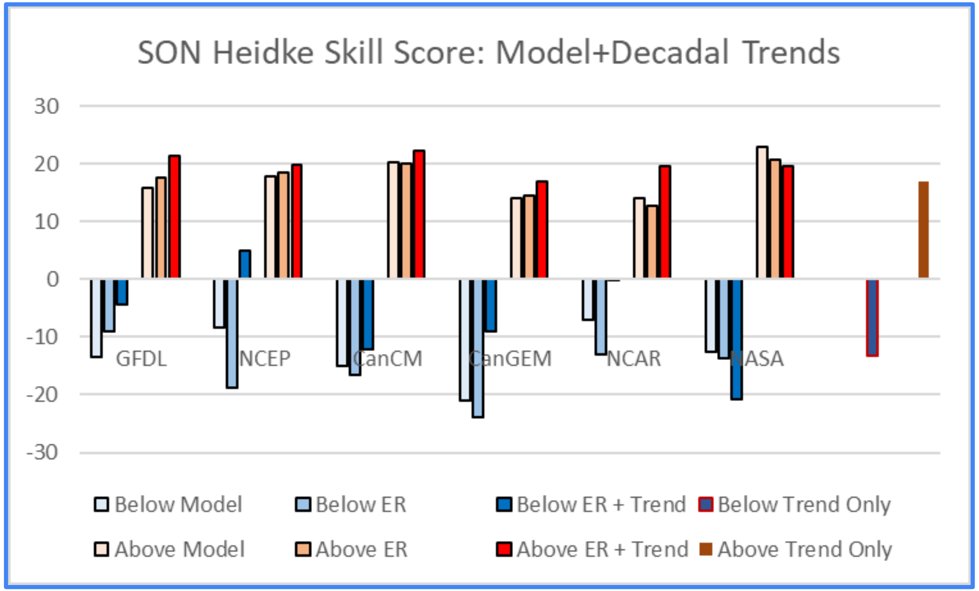
- ER Calibration increases the coverage of above normal temperatures in later years (dashed lines) relative to the raw un-calibrated model output (dotted lines)
- The combined ER + Trend calibration increases the coverage of above normal forecasts even further to nearly 90% coverage by above normal (SON and warm seasons).
- There is less potential to skillfully forecast below normal in seasons with the greatest decadal trends.

Conclusions

- Skillful decadal signals in seasonal models vary across regions and the annual cycle
- Simple statistical models of decadal signals may be more skillful than dynamical models in some locations and seasons
- Combining statistical forecasts of decadal signals with dynamical model forecasts can yield increases in skill when the decadal trend signals are large
- Forecast category frequency bias appears to be a necessary component of reliable probabilistic forecasts and decadal timescale trends, to predict changes in the observed category frequencies

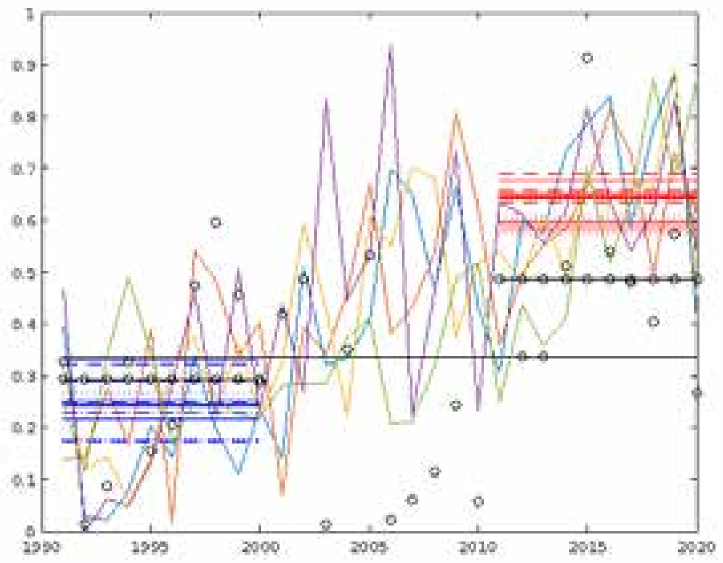
Extra Slides

Skill and the Calibration of the NMME models using decadal and interannual signals (SON, 2006-2020, A or B)

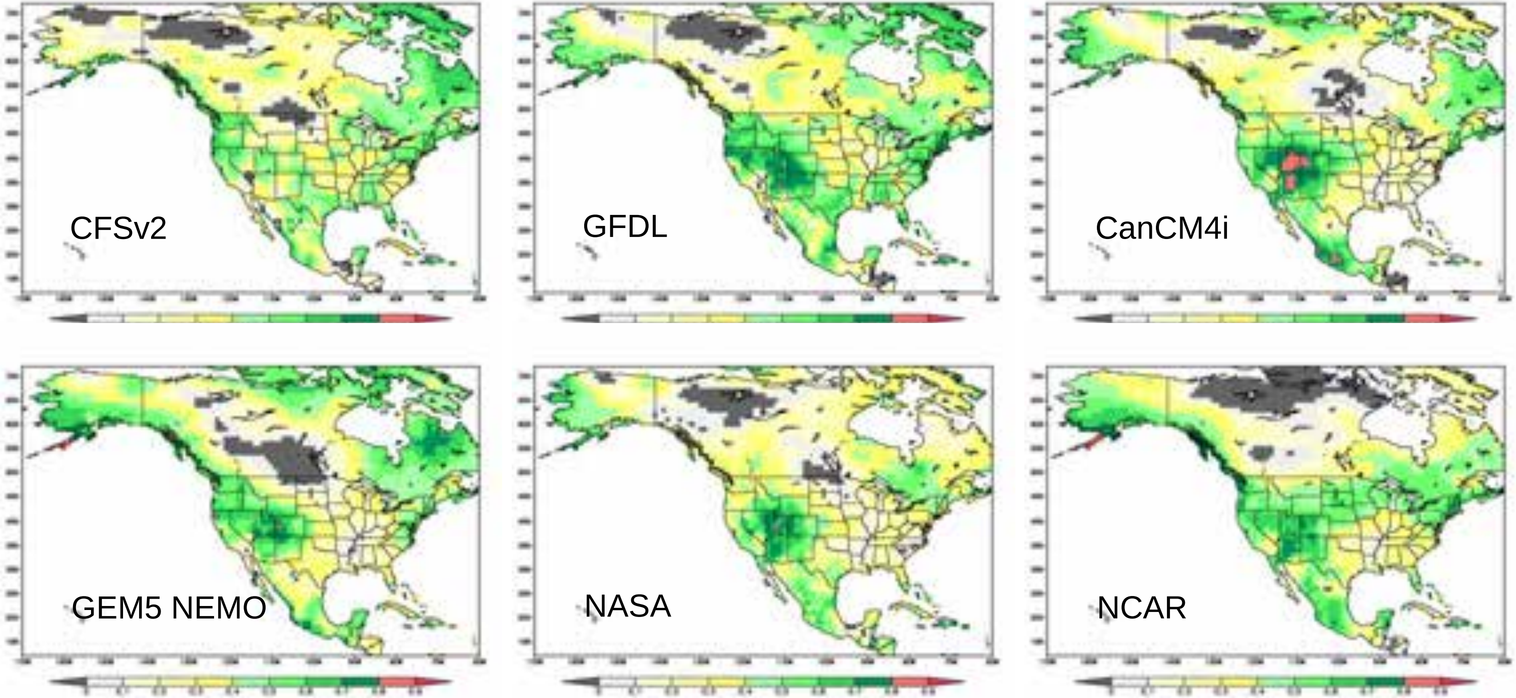


- Dynamical models alone have similar skill to a statistical trend forecast
- While the Ensemble Regression (ER) + Decadal Trend calibrated models for SON nearly always show improved skill
- This implies skill in both the decadal trends and other signals

Frequency of Above-normal Forecasts & Observed (DJF)



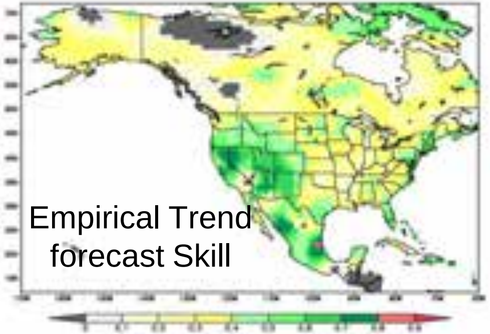
- Below normal is observed more frequently early in the climatology period & above normal is observed more frequently late in the climatology period
- Forecast models predict above normal less frequently than observed early in period (dotted blue lines) and more frequently than observed late in the period (dotted red lines)
- The category frequency bias in dynamical model forecasts (Kirtman et al 2014) appears a necessary component of skillful forecasts;
- Skill-calibrated dynamical model forecasts exhibit an increase in this bias towards over-predicting the dominant category (dashed red and blue lines)
- The separation of decadal trends from other signals in calibrated model forecasts leads to a reduction/increase in DJF/JJA frequency biases, while increasing total anomaly correlations.

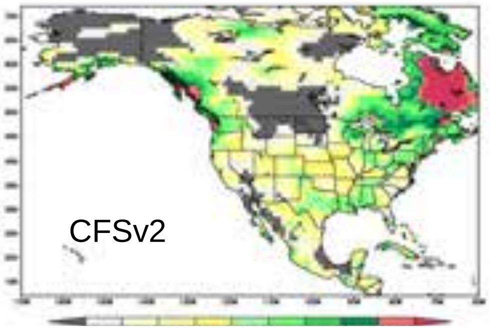


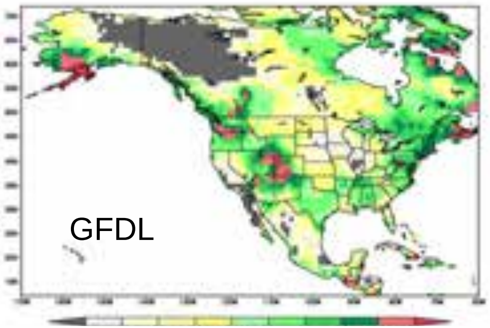
NMME model skill patterns are consistent with observed trends (lower right).

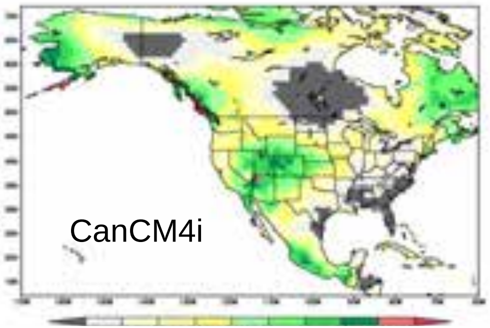
Apparent independent skill in some regions.

<u>**Model Skill**</u>: **Anomaly Correlations** to GHCN-CAMS observations for **lead-1 JJA**:



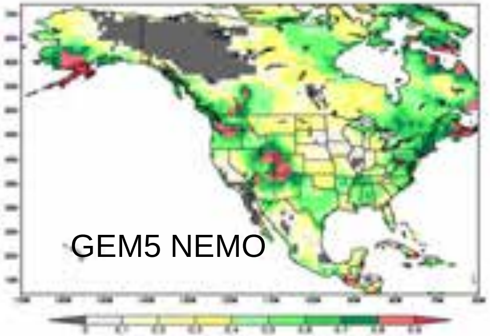


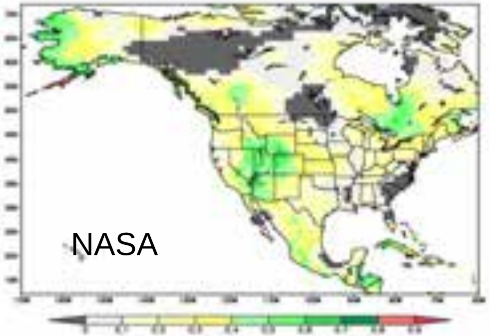


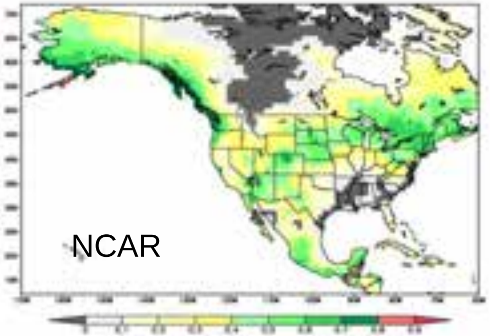


Model skill independent of empirical trend forecast varies.

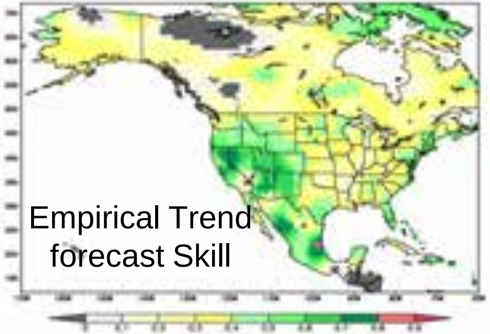
e.g., CFSv2, GFDL, GEM5, and NCAR all show independent skill over the Northeast CONUS

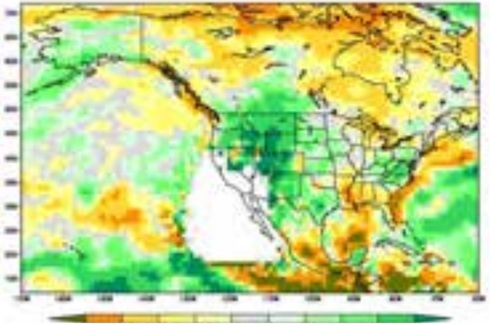




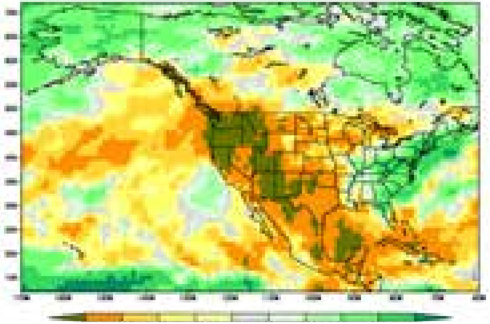


Model Skill independent of the trend predictor: Beta of forecasts accounting for trend ; lead-1 JJA

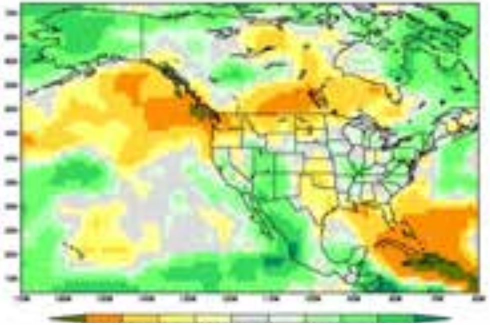




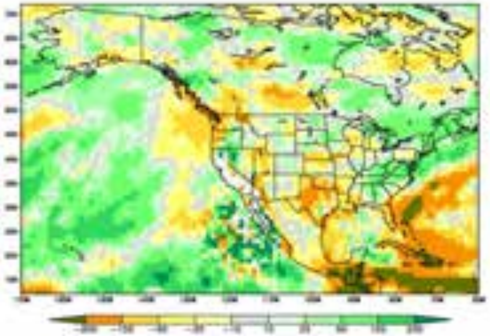
CFSv2



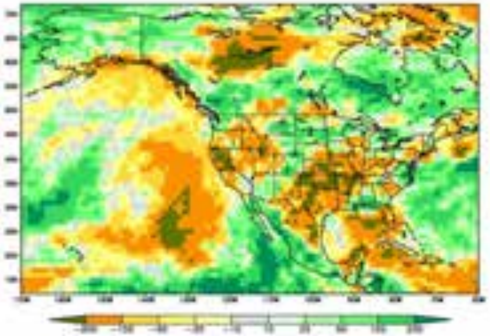
GFDL



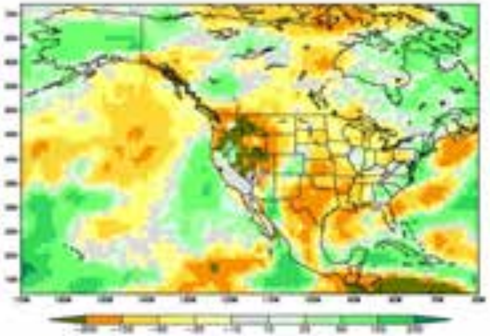
CanCM4i



GEM5 NEMO



NASA

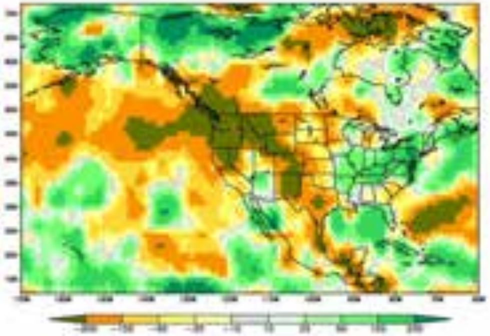


NCAR

NMME precipitation trends vary widely among models.

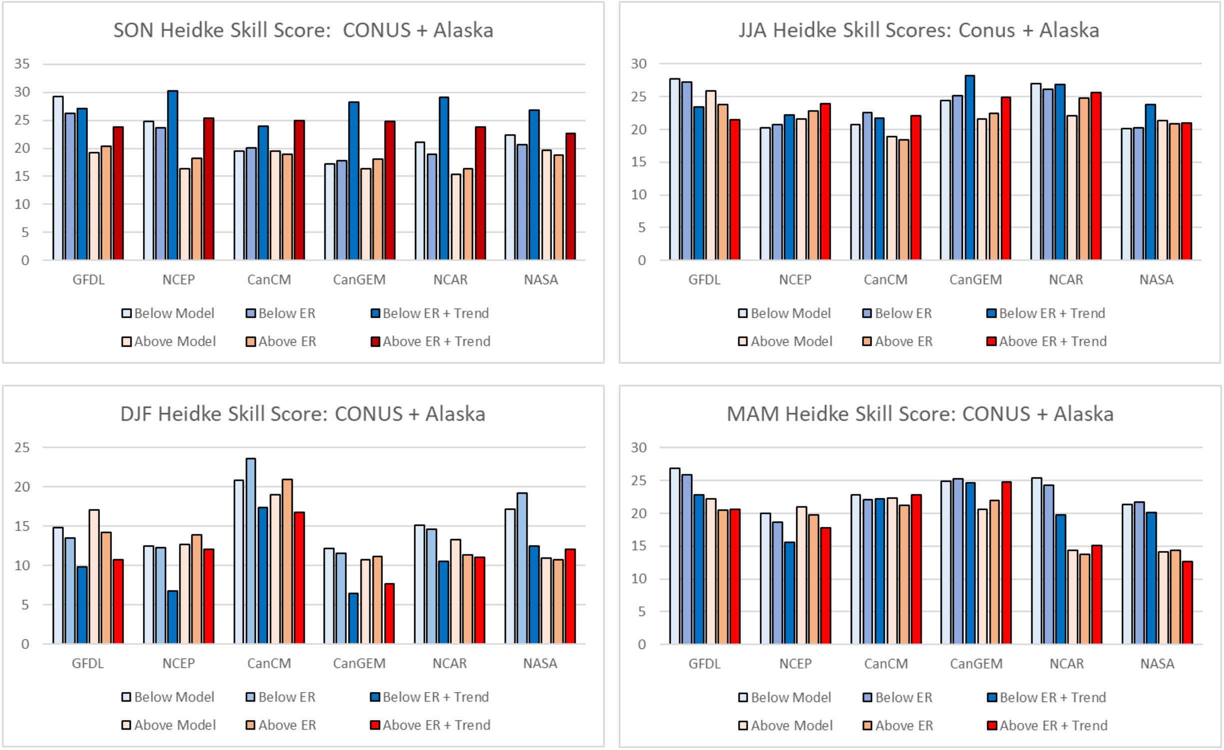
Some consistencies and some inconsistencies with observations during summer months.

Observed



<u>Precipitation</u> Linear Trend based on **NMME model forecasts** versus CMAP observation trends for **lead-1 JJA** forecasts from May initial conditions:

Skill and the Calibration of the NMME models using decadal and interannual signals (SON, MAM)



- The ER+T calibrated models for SON nearly always show improved skill
- ER+T total average skill mixed result
- Global average show consistent improved skill

- Ensemble regression allows variations in the spread of ensemble members to determine the uncertainty of each forecast
- Errors not accounted for by the spread are accounted for by the addition of Gaussian distributions