Tools for Subseasonal and Seasonal Forecasts: Model Post-Processing, Calibration and Consolidation

Dan C Collins

NOAA Climate Prediction Center

Outline

- Outlooks use a combination of Dynamical and Statistical Models
- Dynamical Models, Ensembles, Multi-model ensembles (MME)
  - Represent the major physical processes of the climate system
- Post-processing
  - Correcting dynamical model forecasts using observations
  - Bias correction of the mean and variance of model output
- Calibration
  - Using the skill of verified hindcasts or forecasts to calibrate probabilities in forecasts
  - Reliability :

    Forecasted probabilities represent frequency that a forecasted event occurs (i.e. skill)
- Statistical models
  - Use observed datasets, including reanalysis, to identify predictable signals for the future climate state from the current climate state

Ensemble Models

- Each model has own biases and skill
- Ensembles: Models run repeatedly for a single set of initial conditions
- Ensemble models predict the likely outcomes
- Ensemble mean forecast anomalies tell us the signal or predictable component of the forecast
- Ensemble spread tells us the noise or unpredictable component of the forecast

Methods to make Probability forecasts (using dynamical models)

1) Count ensemble members above threshold, "Raw model probabilities"
   - Model climatological mean removed
   - Bias corrected, sometimes for both mean and variance

2) Calibration: Ensemble Regression[1], Bayesian Joint Probability[2], Probability Anomaly Correlation[3] methods:
   - Removes bias
   - Correct the variance to be similar to observations
   - Correct for skill
   - Make a probability distribution

[1]Unger et al 2009

[2]Wang et al 2009

[3]van den Dool et al 2017

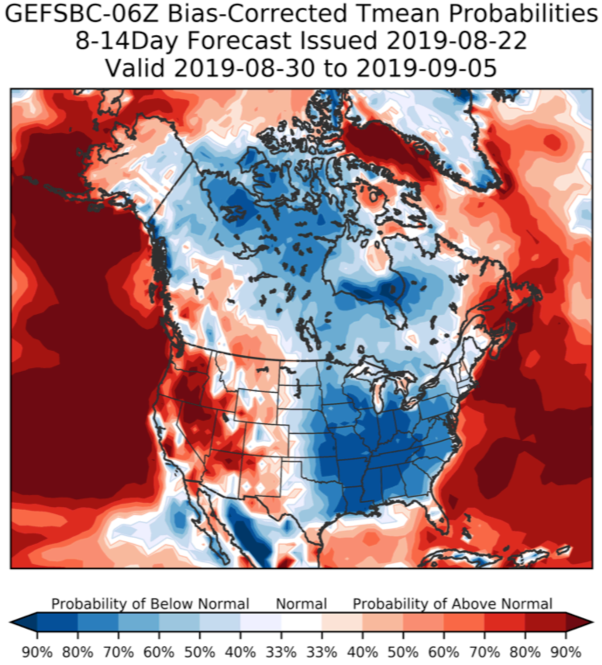
Week 2 Outlook

NAEFS

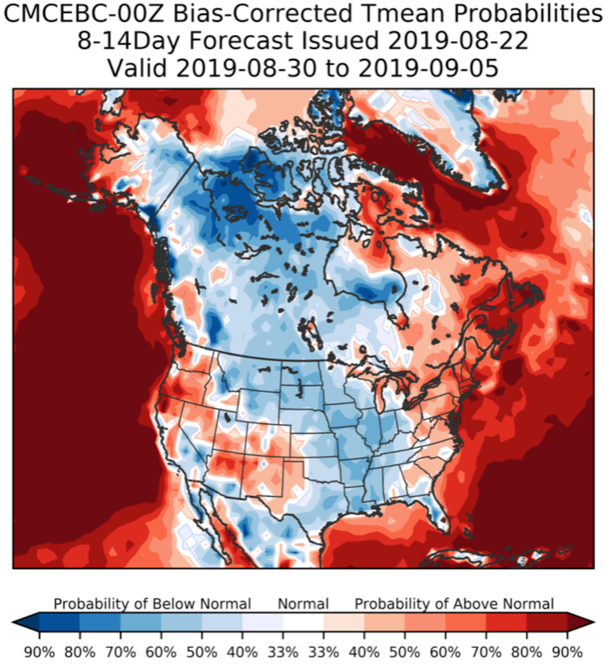
North American Ensemble Forecast System

NAEFS Forecast Probabilities in support of CPC's Week-2 Outlook: Temperature

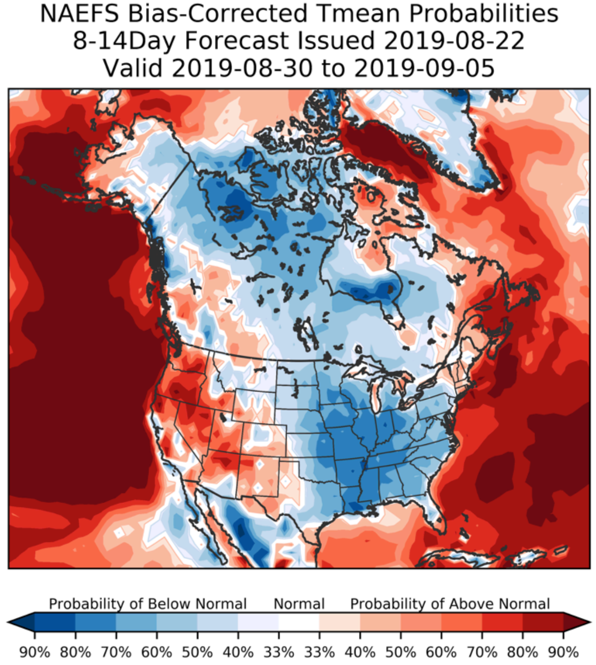
Bias Corrected from recent forecast verifications



NOAA GEFS

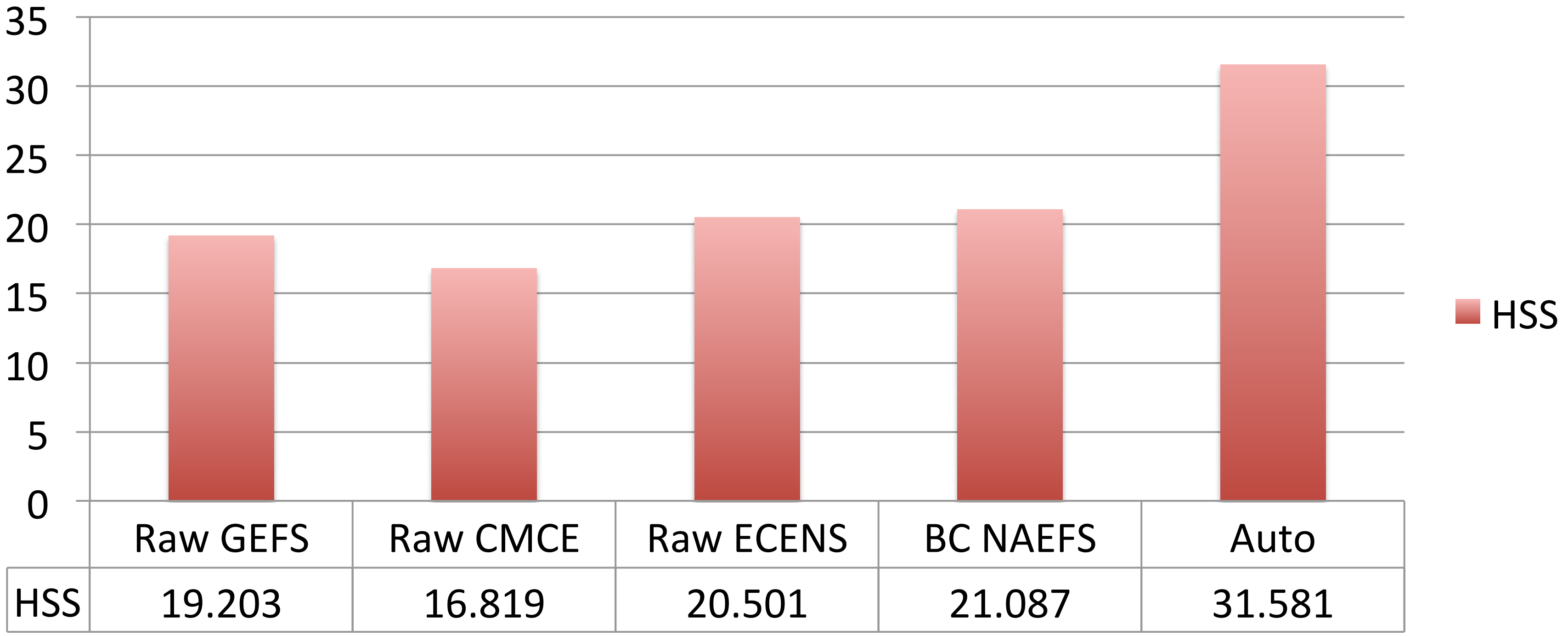


Environment Canada

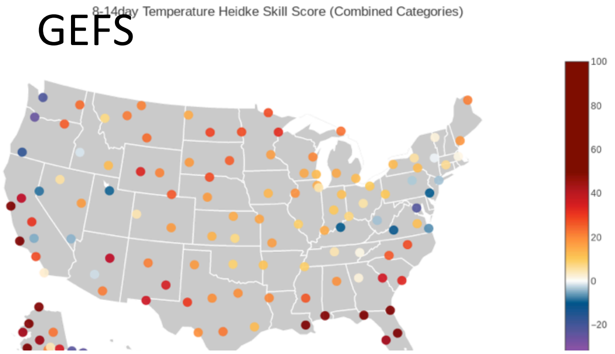


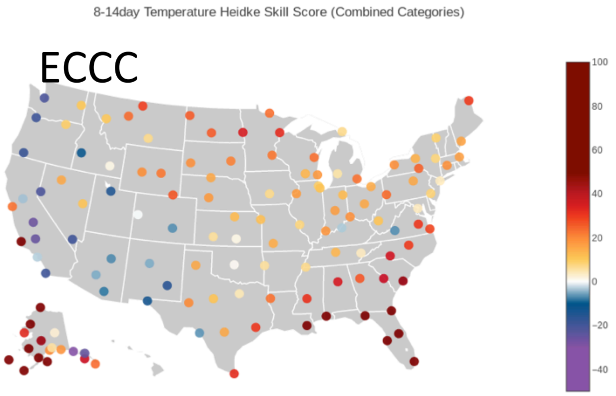
NAEFS

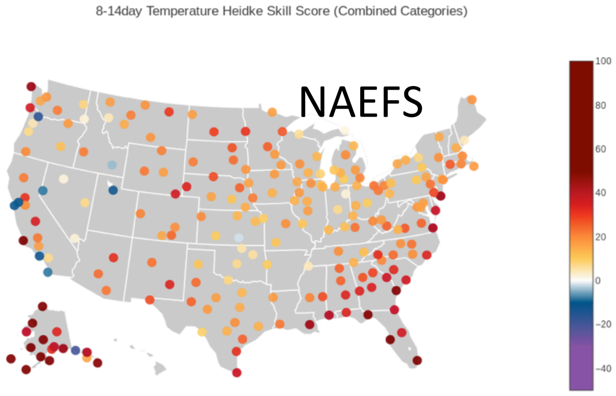
Week-2 365-day HSS Summary



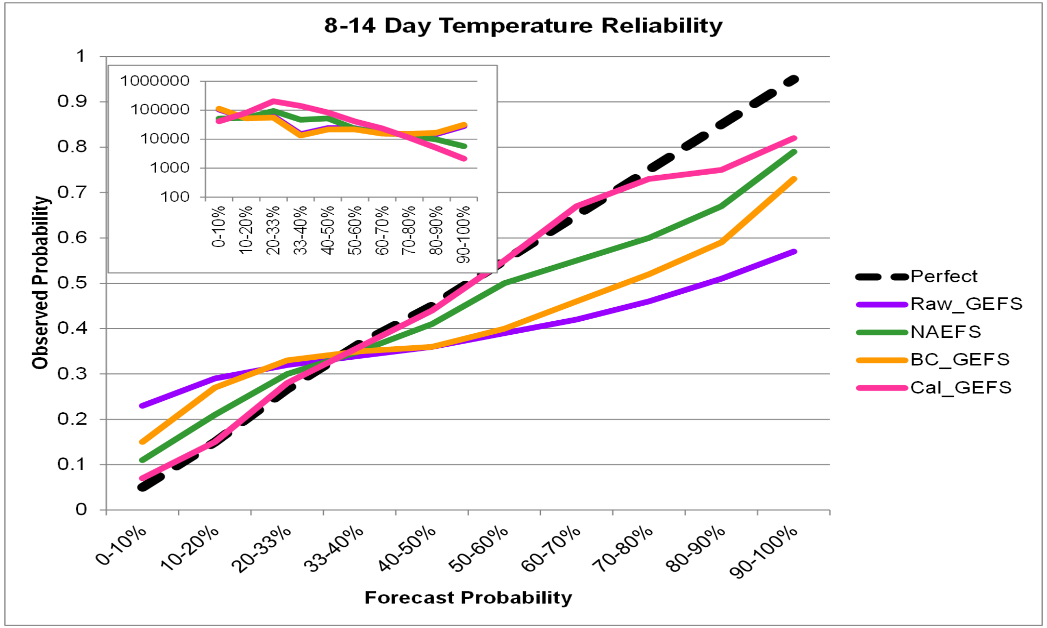
Spatial Maps of Heidke Skill Score for GEFS, ECCC, and NAEFS







Week-2 reliability of reforecast-calibrated GEFS probabilities compared to NAEFS, bias-corrected GEFS, and uncorrected GEFS



Reliability of week-2 probability forecasts including:
raw GEFS, bias-corrected GEFS, NAEFS MME, and reforecast-calibrated GEFS.

Reforecasts & Calibration

Calibration of Probability Forecasts using <u>Ensemble Regression</u>

1) Derive regression equation
   - Between the ensemble mean and the observations
   - Apply to each ensemble member (i.e. each model run).

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

2) Expected residual error for each member
   - Residual error is what is left when subtracting the average ensemble spread from the average mean square error (MSE).

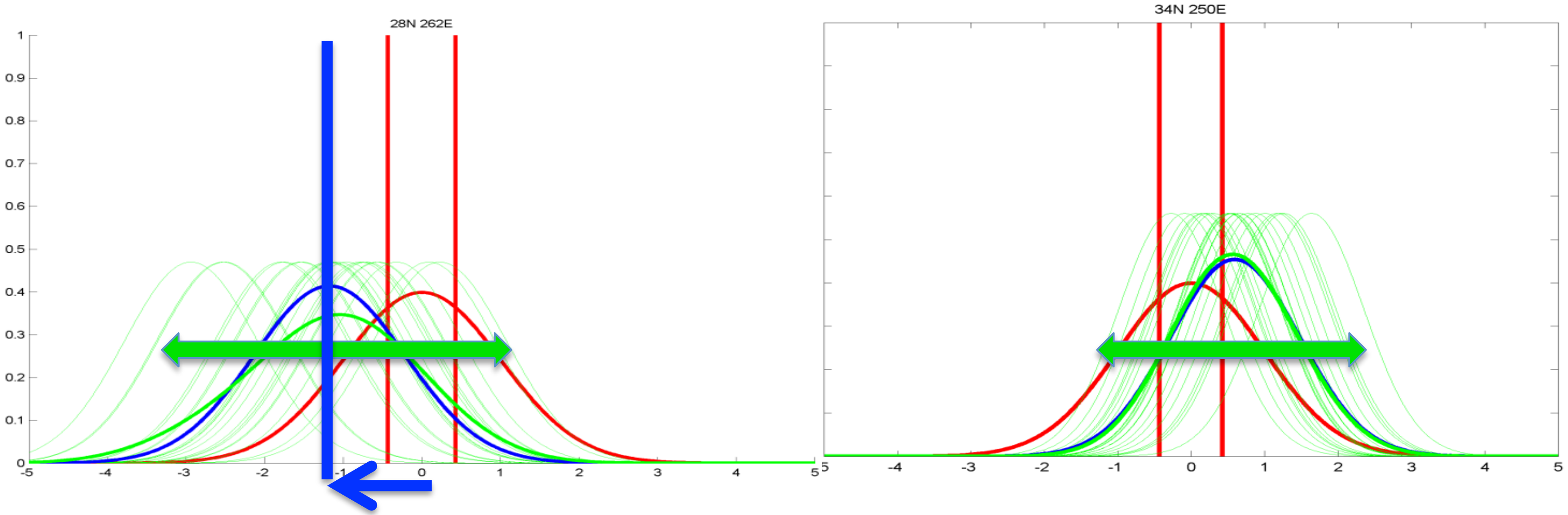
$$\sigma_\varepsilon^2 = [MSE] = \sigma_{ens}^2 + \varepsilon^2$$

$$\varepsilon^2 = \sigma_\varepsilon^2 - \sigma_{ens}^2$$

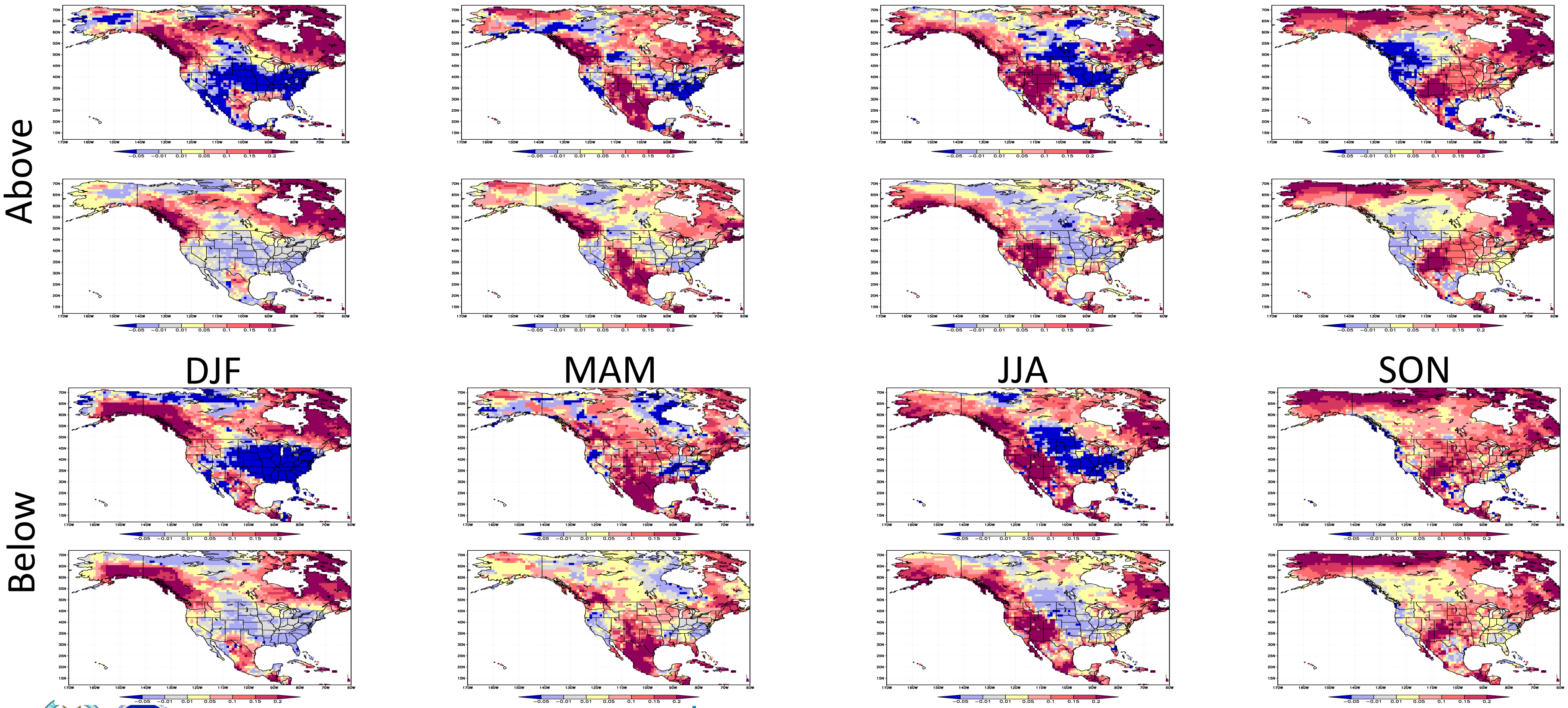
$$[MSE] = \sigma_{obs}^2\,(1 - R_m^2)$$

- Ensemble spread is the average difference between ensemble members (i.e. the standard deviation of all the model runs in an ensemble forecast)
- Ensemble regression allows variations in the spread of ensemble members to determine the width of the PDF and the uncertainty in the forecast

Errors not accounted for by the spread are accounted for by Gaussian error distributions



Brier Skill Scores for above normal (top 2 rows) and below normal (bottom 2 rows) for calibrated (2nd and 4th row) and count-based model probabilities

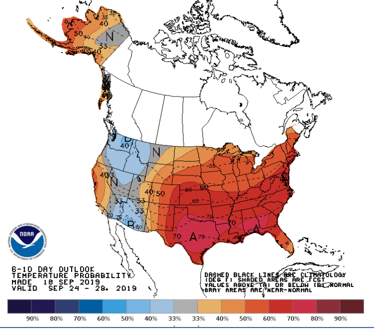


Extended Range Consolidation

Week-2 Consolidation: Reforecast-calibrated ECMWF + NCEP GEFS

Current Auto Forecast

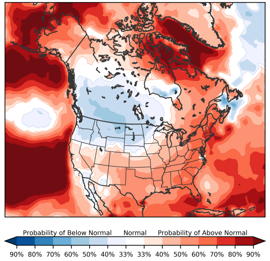
(Tercile Only; Subjective Weighting)



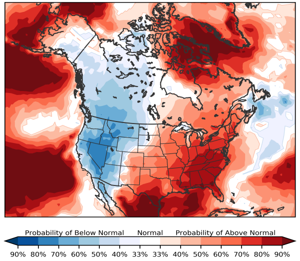
Consolidation Reforecast Inputs (GEFS and ECENS)

(Full distribution; Reliable Probabilties)

GEFS-LEGACY-00Z Rfcst-Cal Tmean Probabilities
8-14Day Forecast Issued 2019-09-18
Valid 2019-09-26 to 2019-10-02



ECENS-00Z Rfcst-Cal Tmean Probabilities
8-14Day Forecast Issued 2019-09-18
Valid 2019-09-26 to 2019-10-02



Key Facts

Parameters: Mean Temperature / Total Precipitation

Valid Period: 6-10 Day / 8-14 Day

Calibration Method: Ensemble Regression

Weighting Method: Counts

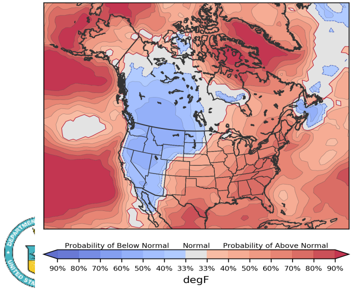
Output Types: Probabilities, Full Fields, Anomalies, Weights

Probability Thresholds: Full distribution

Consolidation Probabilistic Output

(Temporally and Spatially Weighted)

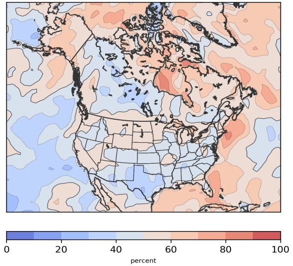
Consolidated 8-14 day tmean issued 20190918
valid 20190926 - 20191002



Consolidation Dynamic Weights

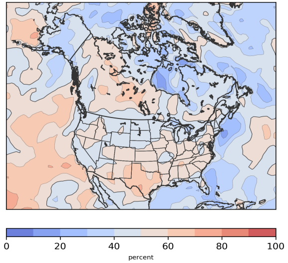
GEFS

gefs-reforecast weights for 8-14 day tmean
issued 20190918 valid 20190926 - 20191002



ECENS

ecens-reforecast weights for 8-14 day tmean
issued 20190918 valid 20190926 - 20191002

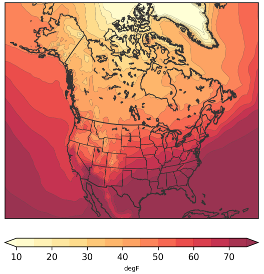


Full Field Output

Available for full distribution

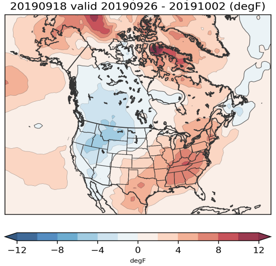
50th Percentile

Full field 8-14 day tmean for forecast probability 50.0%
Issued 20190918 valid 20190926 - 20191002 (degF)



Anomalies

Anomalies for forecast probability 50.0% issued
20190918 valid 20190926 - 20191002 (degF)

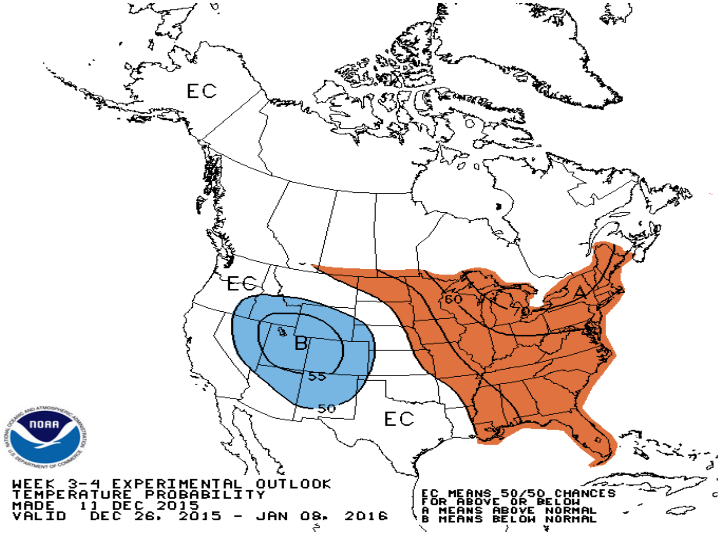


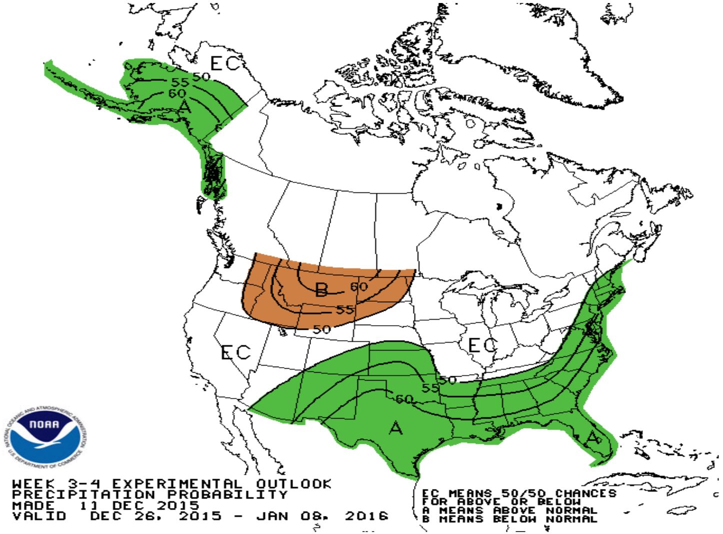
Week 3-4 Outlook

MME forecasts in support of week 3-4 US NOAA operational outlooks:

Above and below normal

Examining extremes and hazards into week 3&4

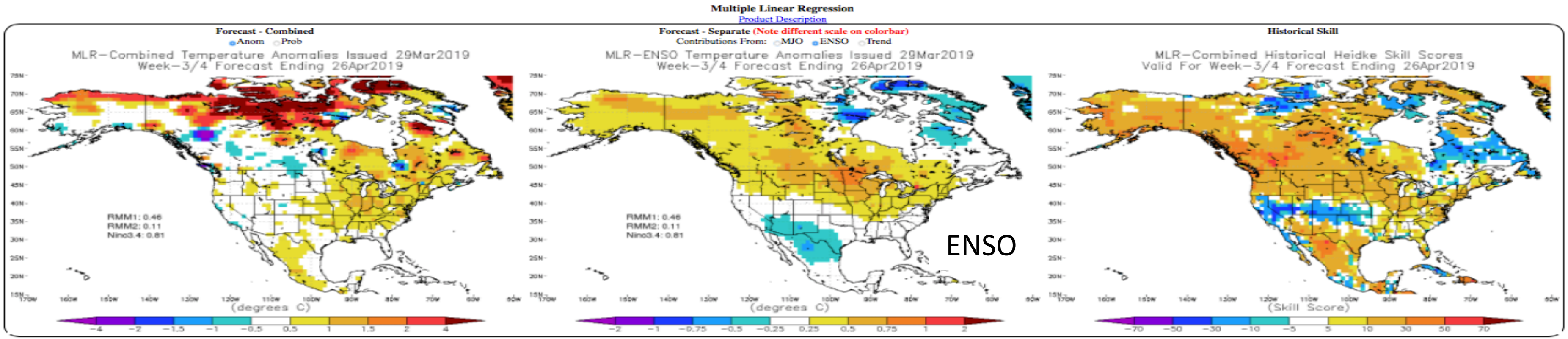




- Probability of above and below normal temperature and precipitation
- Use a combination of dynamical and statistical model forecasts
- MME guidance plays a primary role in the subseasonal forecast

Statistical Models

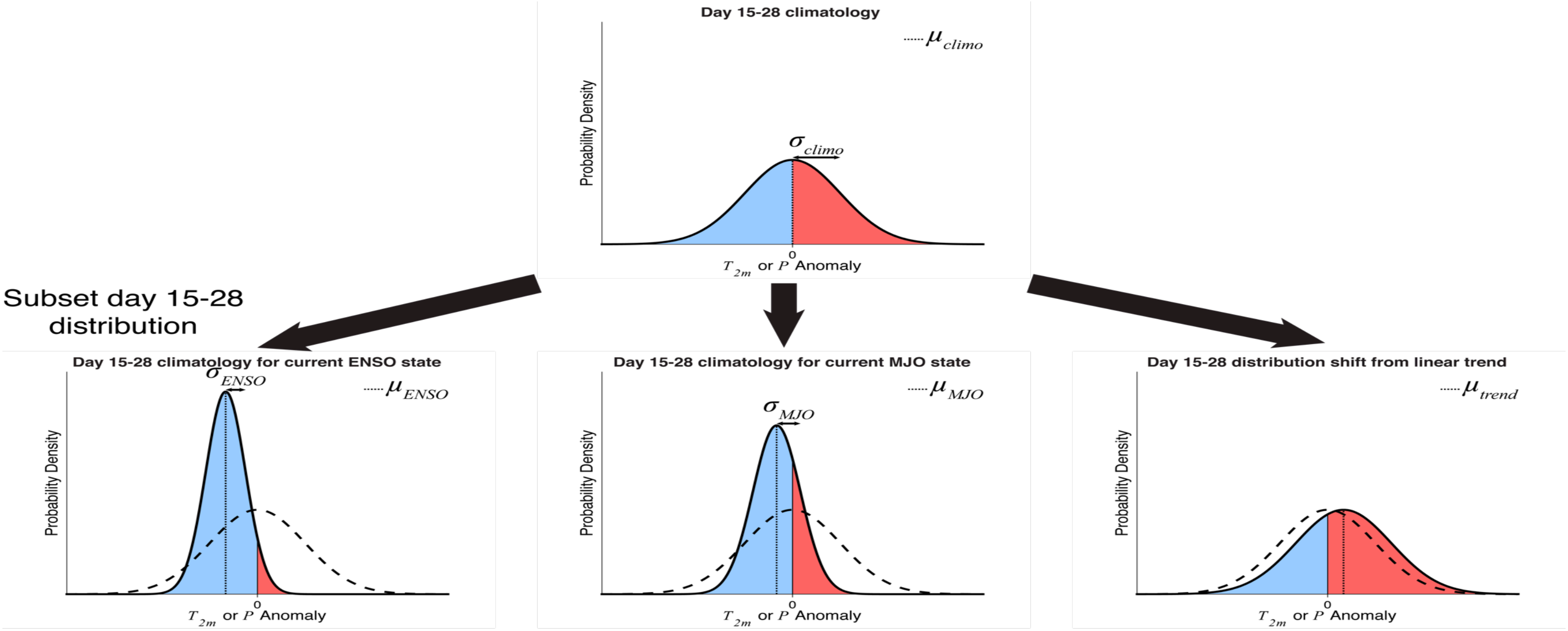
Statistical model guidance: Multiple Linear Regression



- Based on initial state of MJO and ENSO: RMM and Nino 3.4 index predictors
- Regression of climate indices to local temperature and precipitation anomalies
- Comparable skill to dynamical models

PHASE MODEL MJO + ENSO + Decadal Trend

Johnson et al. (2014 – Weather and Forecasting)



- Days 15-28 forecasts of T_{2m}, P generated empirically from ENSO, MJO base states.
  - ENSO: Niño 3.4; MJO: *Wheeler and Hendon (2004)*

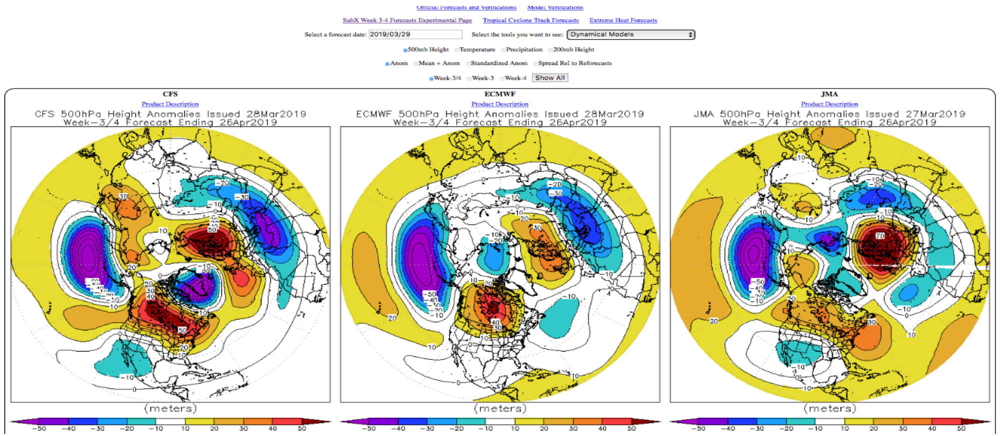
MULTIPLE LINEAR REGRESSION Model

- Predictors: All standardized (1982-2013 base) over running 3-month period
  - Daily Niño 3.4 index (OISSTv2)
  - Daily Wheeler-Hendon (04) RMM1, RMM2
  - Linear trend
- Predictand: T_{2m} or P anomaly
  - Gaussian PDF then constructed using adjusted climatological variance:

$$\sigma_{adj} = \sigma_{climo}\sqrt{1 - R^2}$$

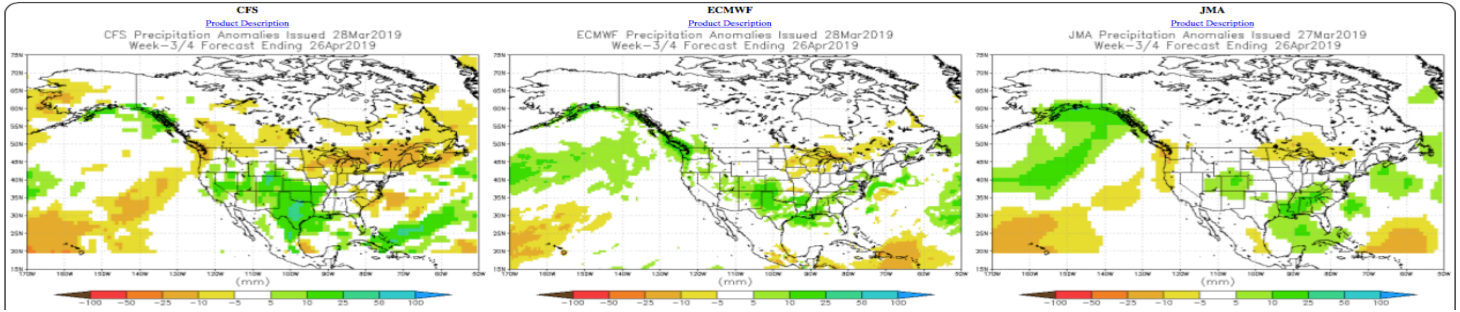
Week 3-4 Dynamical Models

Operational model guidance: NCEP CFS, ECMWF & JMA

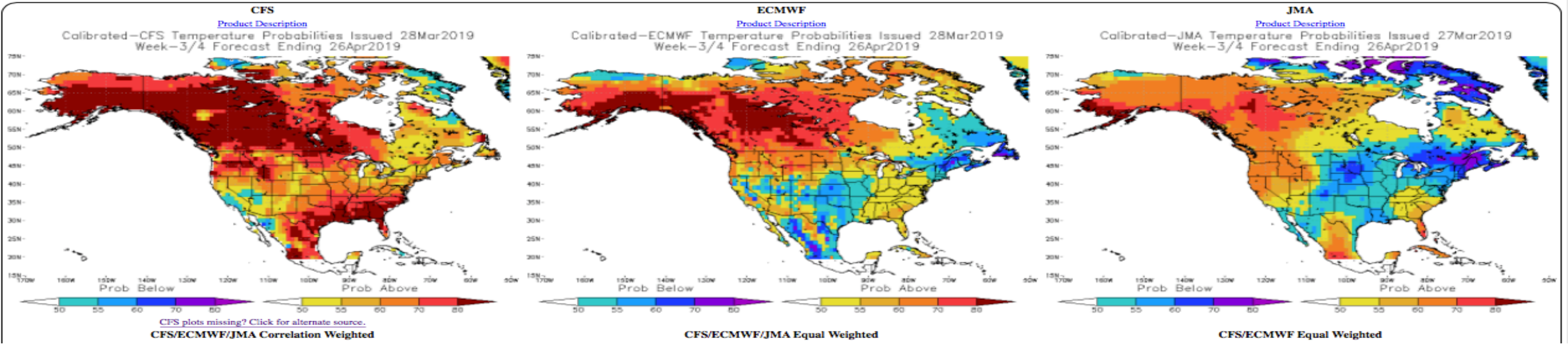


- 500 hPa height anomalies

- Precipitation anomalies

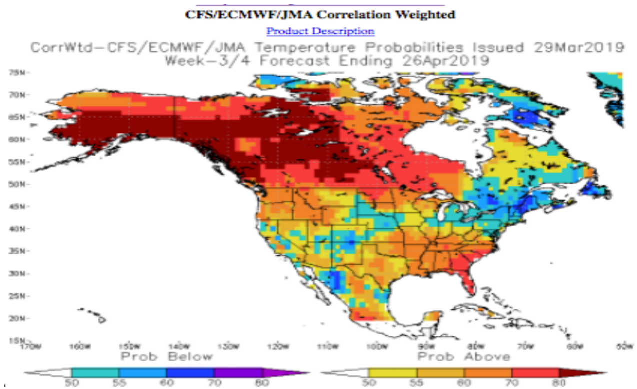


- Each dynamical model is bias corrected using model hindcast
- Calibrated PDFs made using hindcast skill of each model (ECMWF, CFSv2, JMA)



<u>Correlation weighted MME:</u>
Above and below median
2-m temperature

$$Weight_m = R_m^2 \,/ \sum_{m=1 to M} R_m^2$$



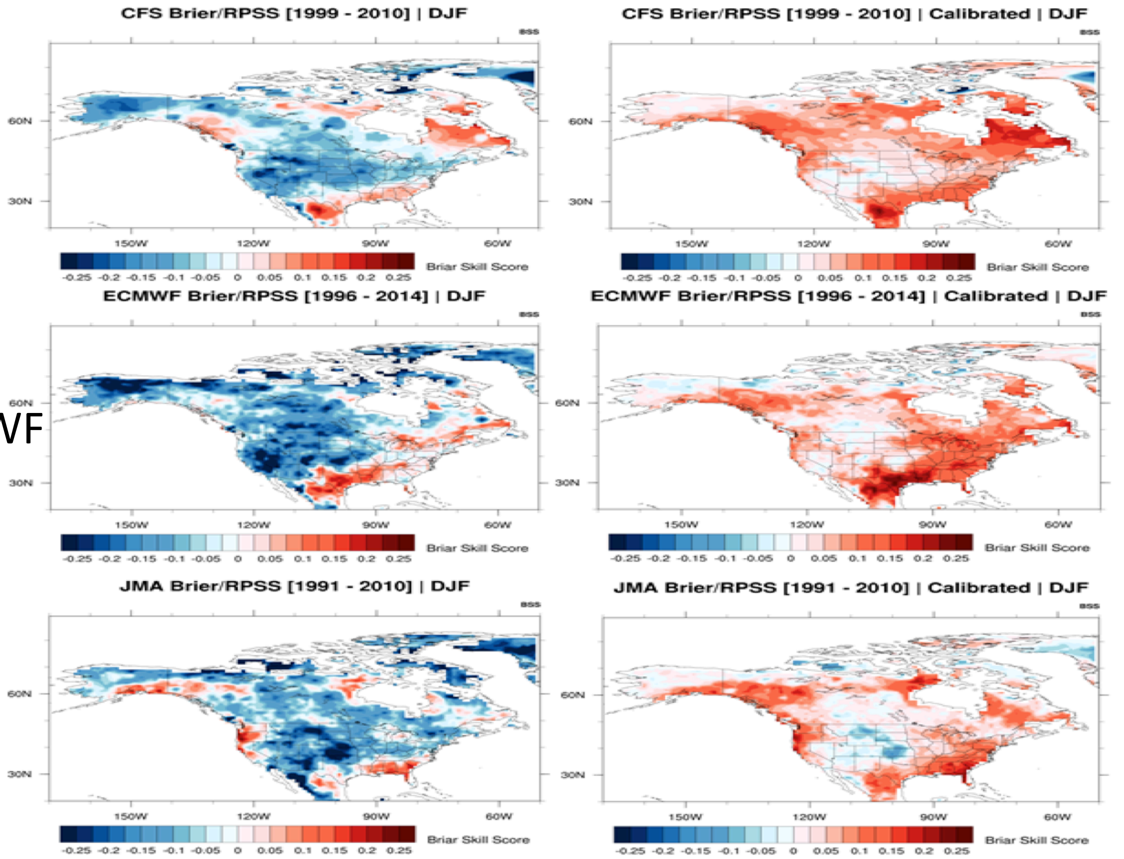
Brier Skill Scores of DJF temperature forecasts :

Not calibrated (left) & calibrated (right)

CFS

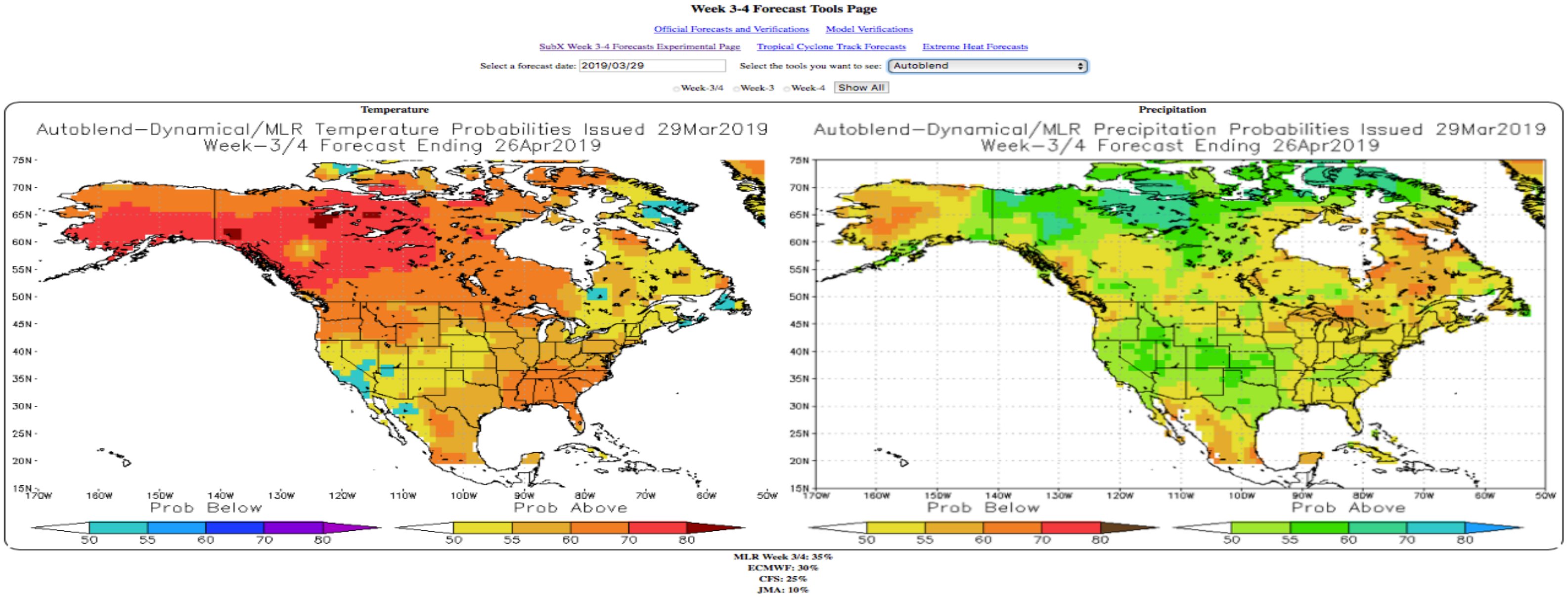
ECMWF

JMA

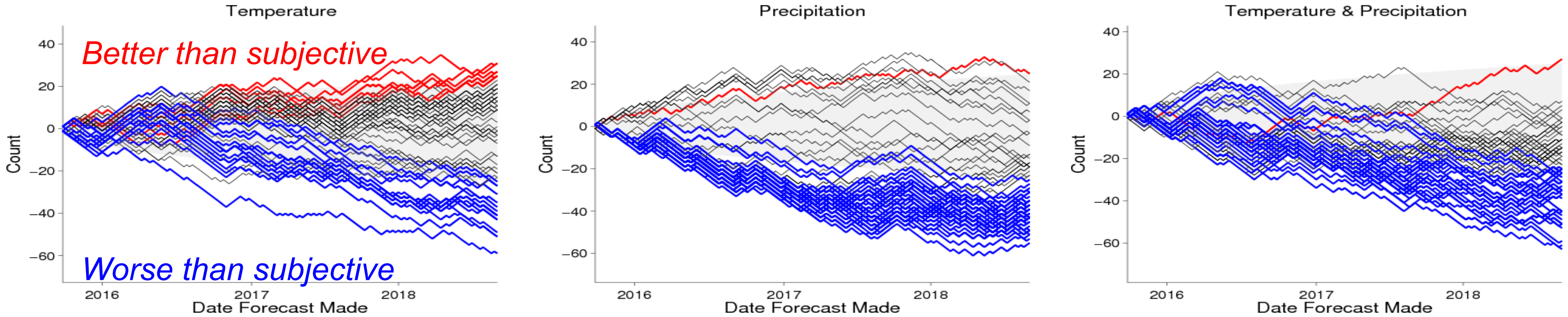


Automated Blend: Combining statistical and calibrated dynamical models

$$w_1{*}MLR \; + \; w_2{*}ECMWF \; + \; w_3{*}NCEP\_CFS \; + \; w_4{*}JMA$$



- Experimenting with objectively determining weights



Possible Weights by Tool				Average HSS		
CFS	EC	JMA	MLR	Temp	Precip	T&P
20%	30%	10%	40%	**29.8***	5.2	**17.5***
10%	30%	20%	40%	**32.1***	2.9	**17.5**
20%	40%	0%	40%	27.6	**6.8***	**17.2**
10%	40%	20%	30%	**31.2***	3.0	**17.1**
0%	40%	30%	30%	**32.2***	1.4	16.8
20%	10%	30%	40%	**31.5***	0.9	16.2
20%	10%	40%	30%	**31.6***	-0.3	15.7
30%	10%	40%	20%	**29.3***	-0.4	14.4
25%	30%	10%	35%	28.6	5.2	16.9
			CPC	25.8	1.4	13.6

9/2015-9/2018 Results

Bold: Heidke Skill Score improved relative to subjective consolidation (2nd to last row).

*: Significantly improved relative to subjective consolidation based on sign test (figures at top).

Yellow shading: HSS improved relative to CPC outlooks (last row).

Courtesy of D. Harnos

SubX : The Subseasonal Prediction Experiment

- Providing a protocol, database and test bed for hindcast and real-time subseasonal forecasts

- Hindcasts (1999-2015)
  - More than **2 years** of **weekly** real-time forecasts
  - 7 **operational or experimental** ensemble models

- Generating MME and examining the value to subseasonal forecasts
  - Model calibration and multi-model ensemble combination
  - Adding new experimental systems (e.g. NCAR CESM2)
  - Assessing the added value to operational models

- Supporting NOAA National Weather Service /Climate Prediction Center, **Week 3-4 Outlooks**

CPC SubX guidance

www.cpc.ncep.noaa.gov/products/people/elajoie/subx/

SubX Week 3/4 forecasts

SubX : The Subseasonal Prediction Experiment Project

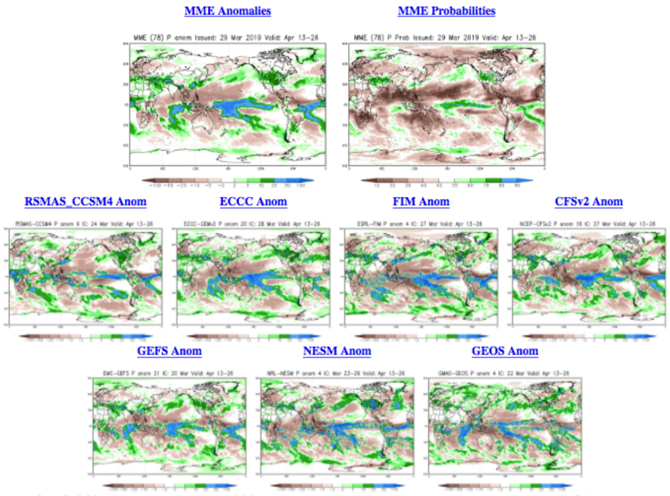
IRI Data Library

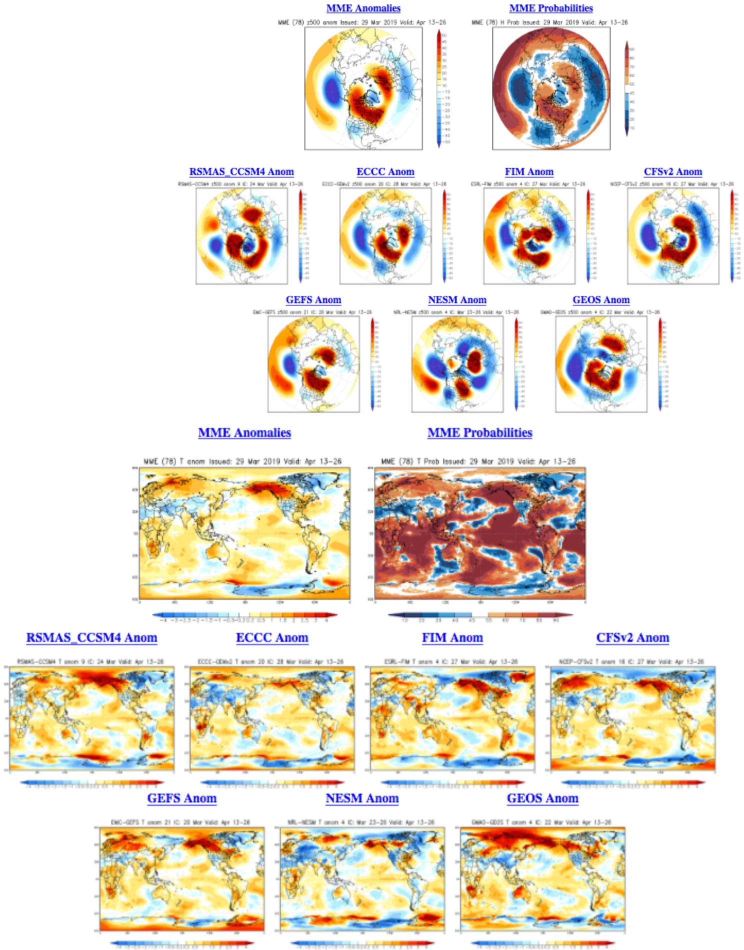
Week 3/4 Operational Model Forecasts

North America	Global
500-hPa height	500-hPa height
2-m Temperature	2-m Temperature
Precipitation	Precipitation

http://www.cpc.ncep.noaa.gov/products/people/elajoie/subx/

SubX Week 3-4 guidance

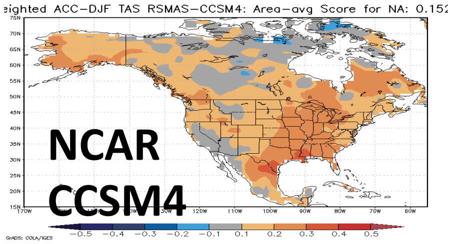


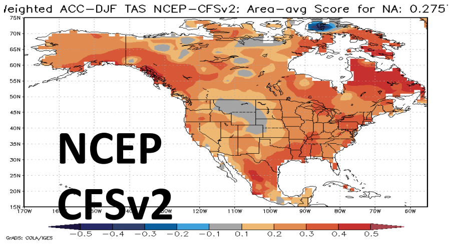


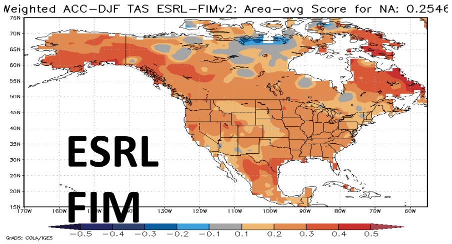
Courtesy of Emerson LaJoie

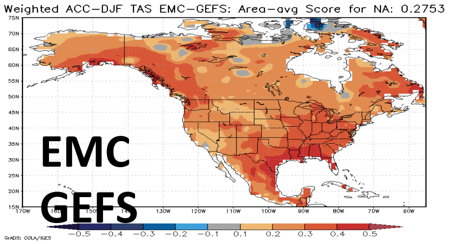
Anomaly Correlation by model & MME (DJF)

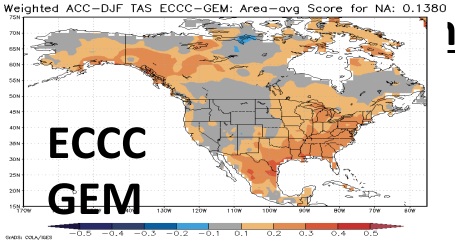
- MME outperforms any individual model

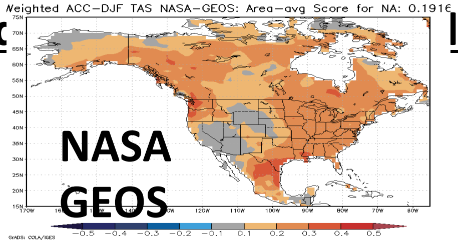


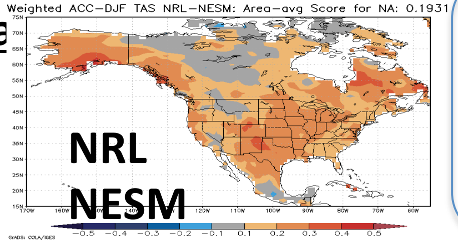


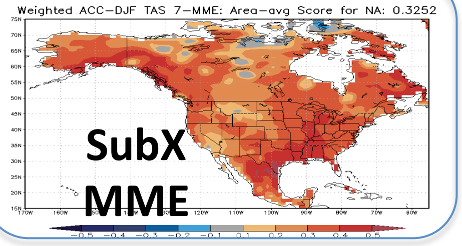




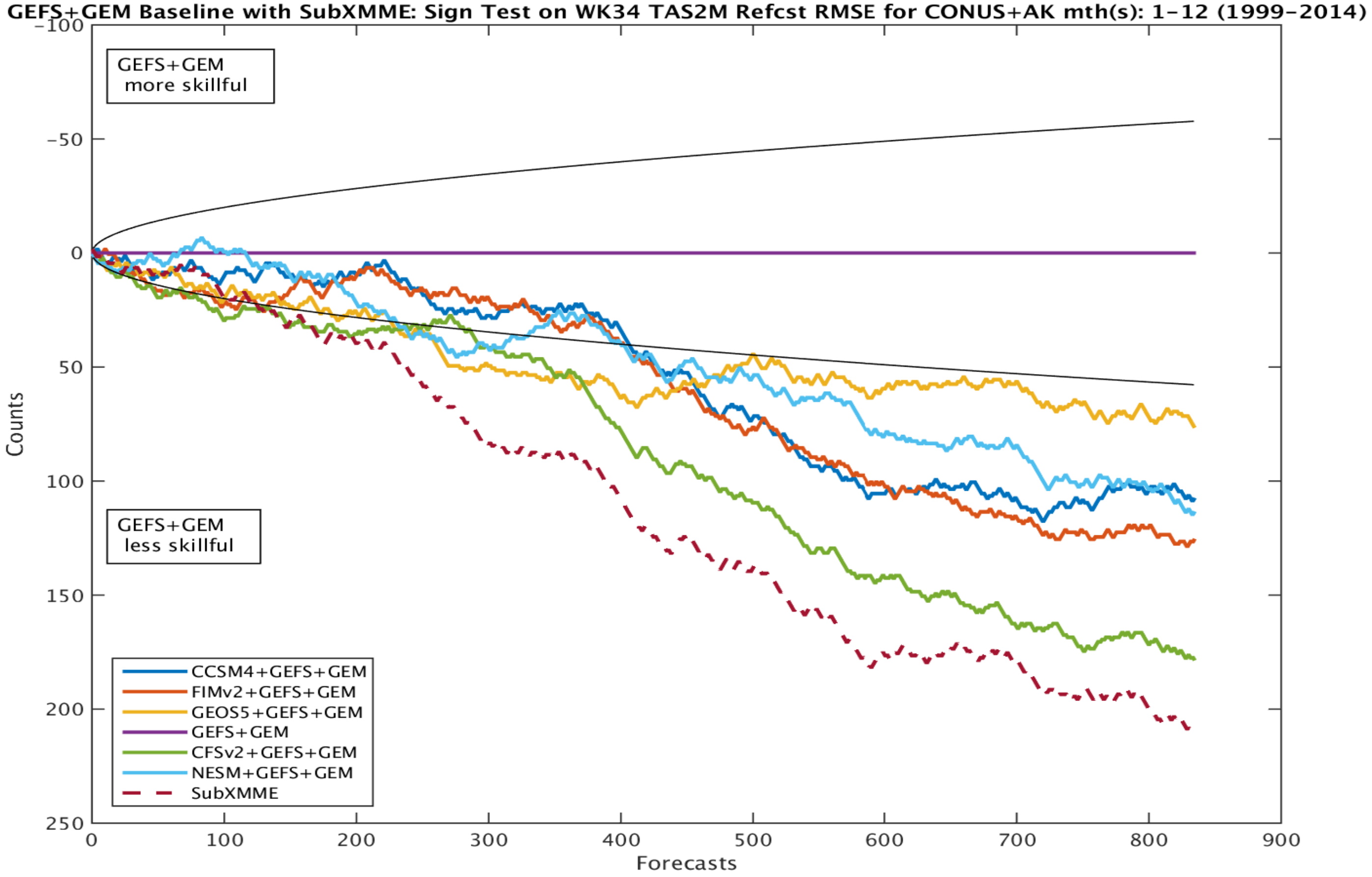




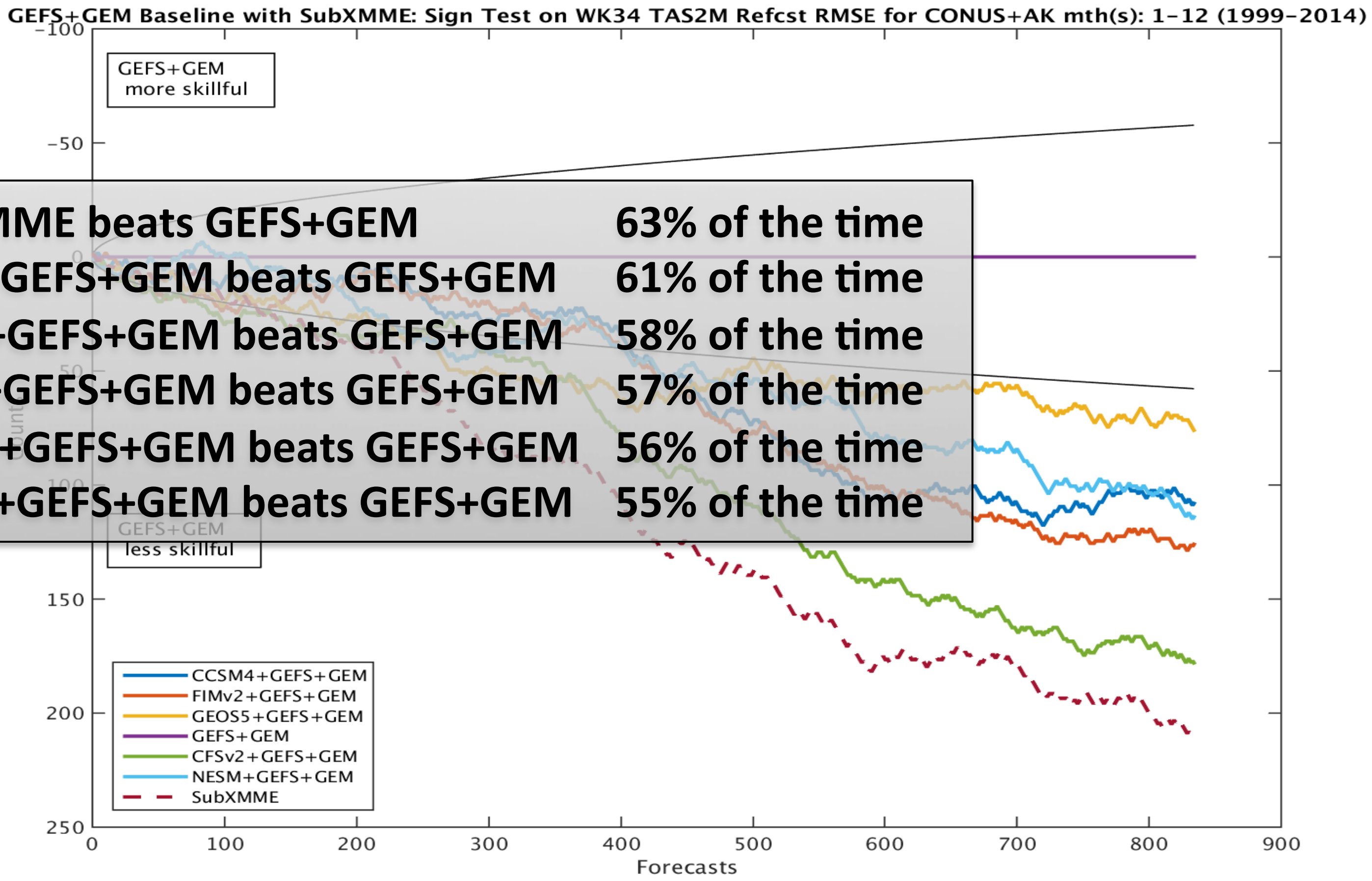




Courtesy of E. LaJoie



Courtesy of Emerson LaJoie



SubX MME beats GEFS+GEM	**63% of the time**
CFSv2+GEFS+GEM beats GEFS+GEM	**61% of the time**
FIMv2+GEFS+GEM beats GEFS+GEM	**58% of the time**
NESM+GEFS+GEM beats GEFS+GEM	**57% of the time**
CCSM4+GEFS+GEM beats GEFS+GEM	**56% of the time**
GEOS5+GEFS+GEM beats GEFS+GEM	**55% of the time**

Courtesy of Emerson LaJoie

Calibration of ensembles to obtain reliable probabilities

… while retaining skill

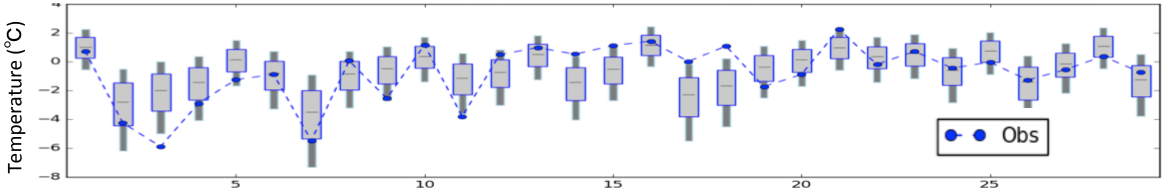
- Do forecast probabilities from an MME represent frequency of occurrence?

- Can an individual model be calibrated and produce the same skill and reliability as an MME?

Bayesian Joint Probability (BJP) Model

Strazzo et al., 2018

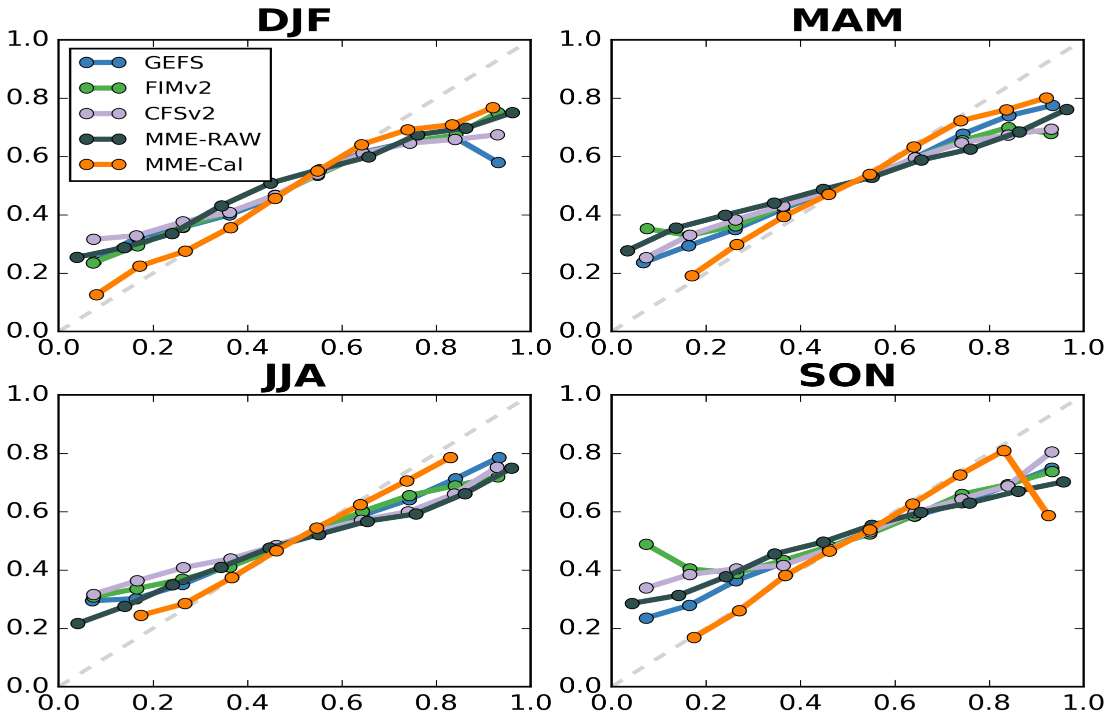
- Calibration using **Bayesian Joint Probability (BJP)** modeling (Wang et al. 2009).
  - Predictor (e.g., CFSv2 2-m T) and predictand (e.g., observed 2-m T) modeled using a bivariate normal distribution, where the distribution parameters are not fixed.
  - Individual **calibration** BJP models are developed for each SubX model ensemble mean, grid point, lead, and season.
    - MME is simple average of 3 ensemble model probabilities.
  - **Ensemble Regression (EReg) baseline** used at CPC (Unger et al. 2009).
- BJP generates a statistical ensemble by sampling from the posterior distribution of the bivariate normal parameters (n = 1000).



BJP forecast of DJF 2-m temperature for a single grid point

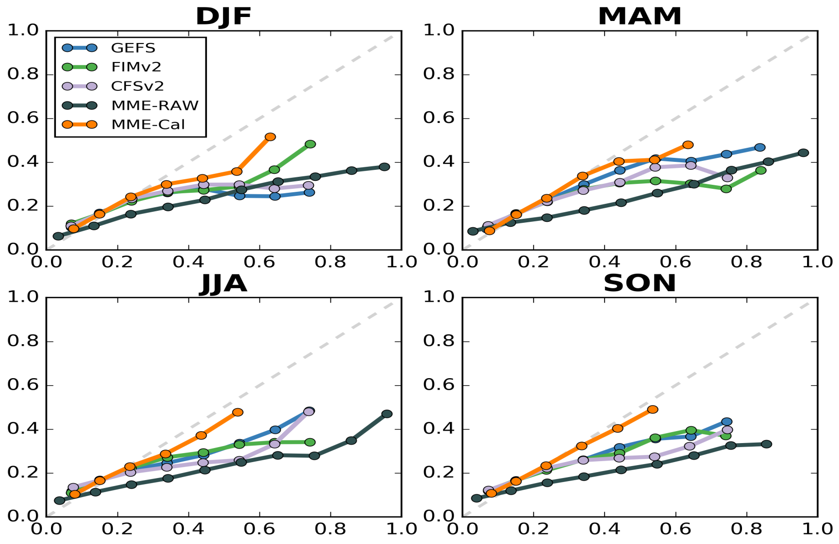
Above / below normal temperature reliability

- **Calibrated MME more reliable** than calibrated **GEFS**, **FIMv2** or **CFSv2,** (small ensemble size), or **MME member count (raw)** probability in all seasons

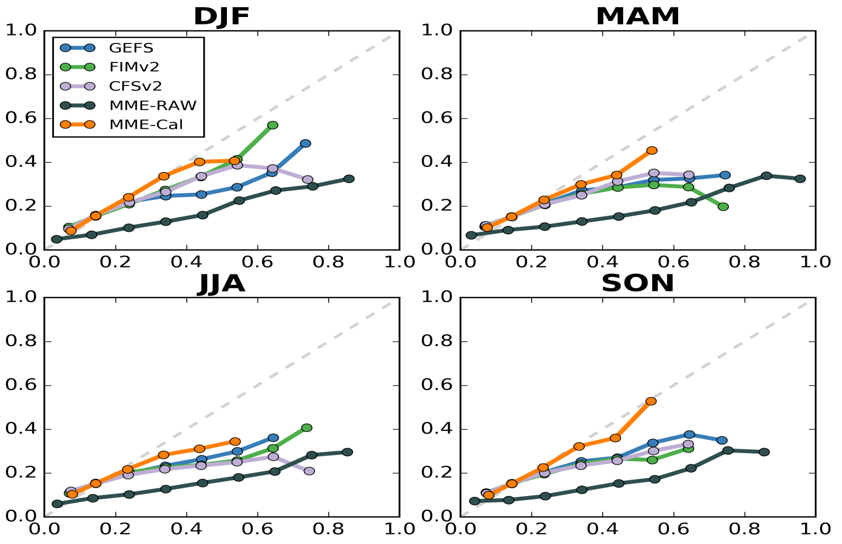


Extreme above/below normal reliability (high and low 15th percentile)

- **Calibrated MME** essential to reliability of probabilities of extremes
- **Raw MME** has much less reliable probabilities
- Individual calibrated **SubXGEFS**, **FIMv2** or **CFSv2** are less reliable than **MME**



Extreme **below** normal

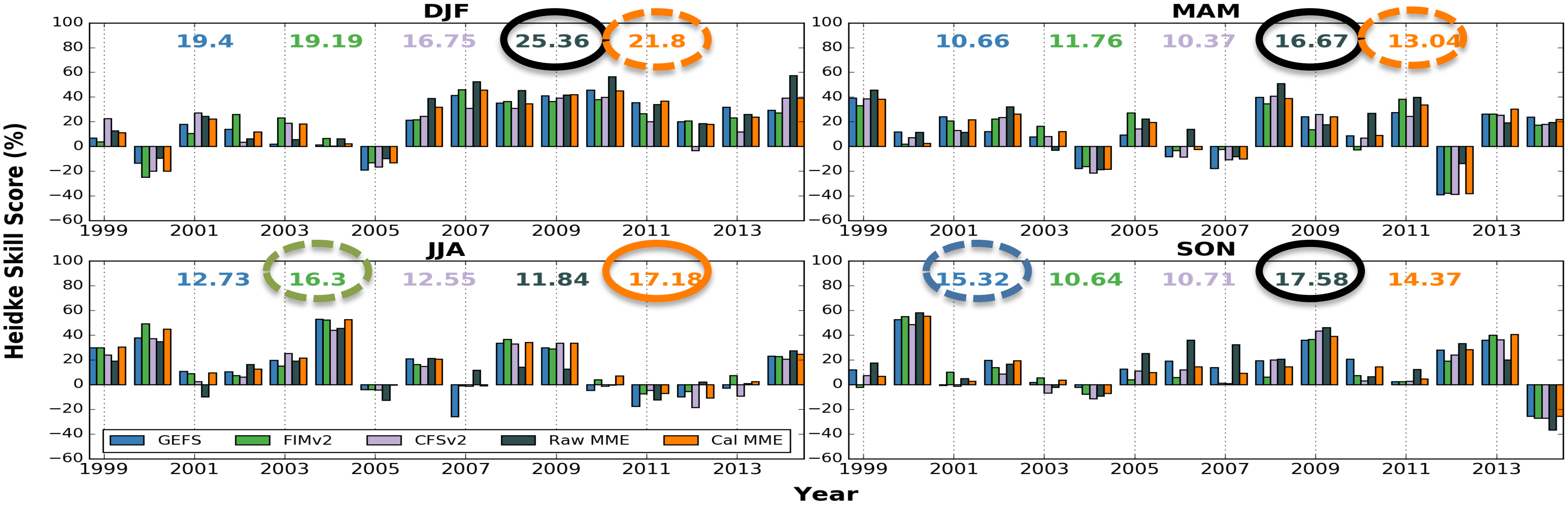


Extreme **above** normal

Above/below median Heidke Skill Score (% Hit rate improvement)

1st & 2nd ranked models

- **Raw mini-MME** has less reliable probabilities but occasionally better hit rate
- MME more skillful in most months & years than GEFS. FIMv2 or CFSv2

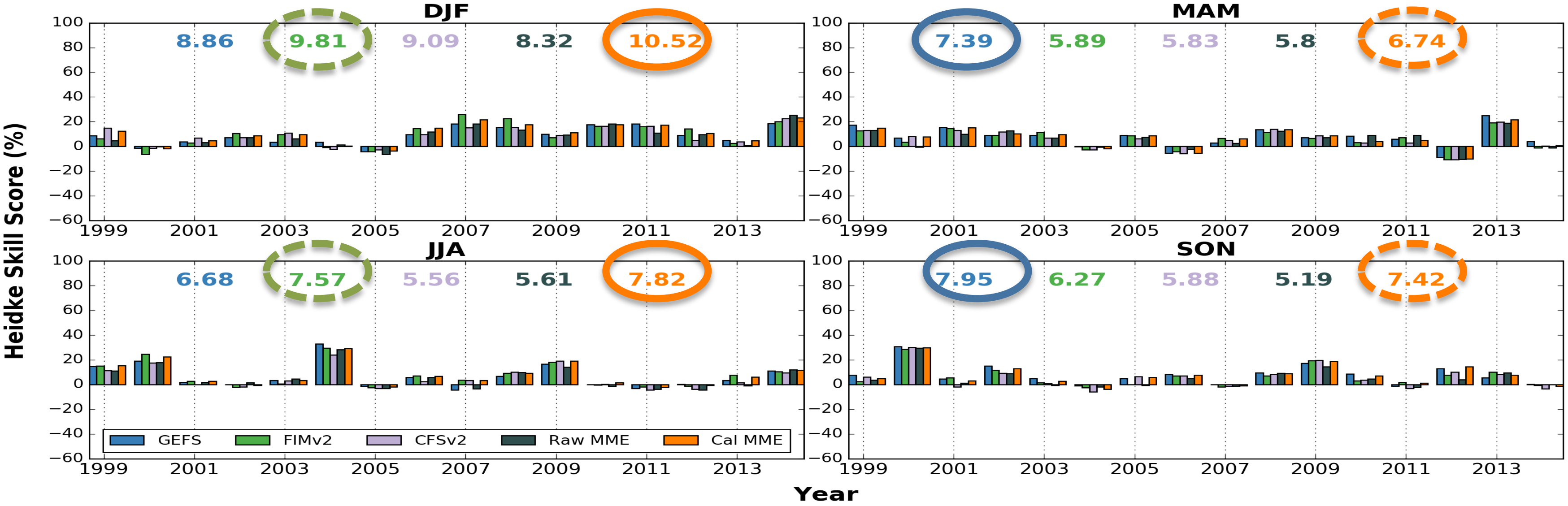


Courtesy of S. Strazzo

Extreme below normal Heidke Skill Score

1st & 2nd ranked models

- Calibration of **raw mini-MME probabilities** improves overall Heidke Skill Score
- **Raw mini-MME** has less reliable probabilities AND lower hit rate
- MME more skillful in most months & years than **GEFS**, **FIMv2** or **CFSv2**

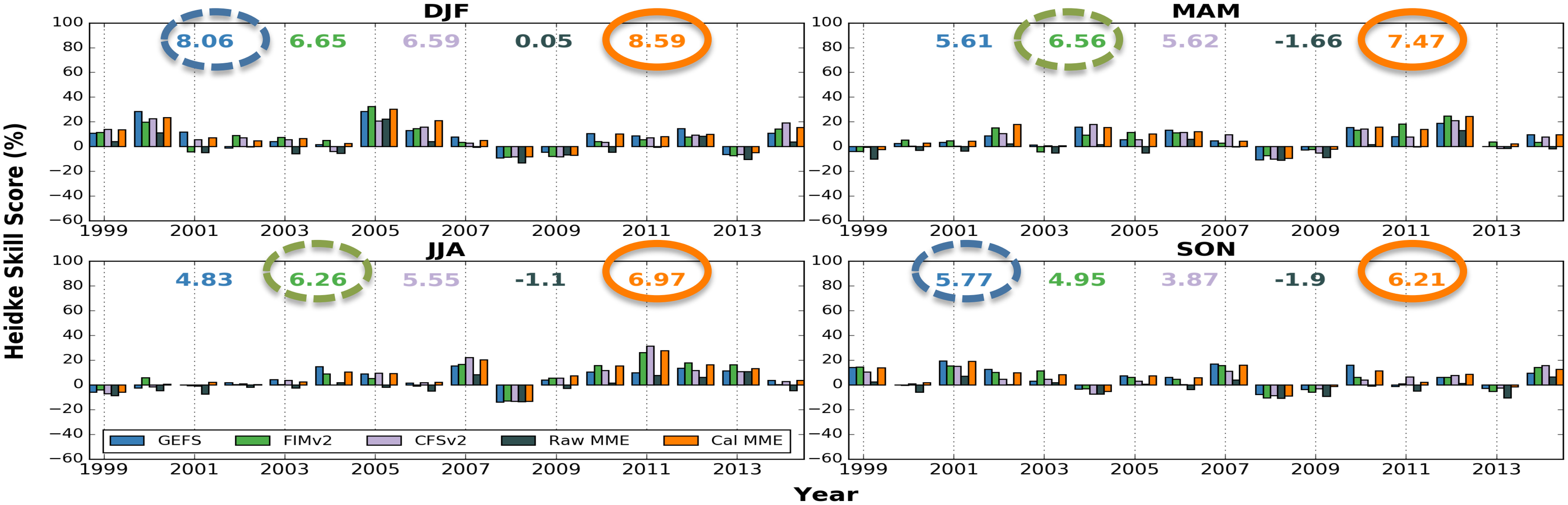


Courtesy of Sarah Strazzo

Extreme above normal Heidke Skill Score

1st & 2nd ranked models

- Calibration of **raw mini-MME probabilities** *improves overall Heidke Skill Score*
- **Raw mini-MME** has less reliable probabilities AND lower hit rate
- MME more skillful in most months / years than GEFS, FIMv2 or CFSv2



Courtesy of Sarah Strazzo

Summary

- MME of calibrated models produced reliable probabilities
- Calibration improves probabilistic skill (Brier and RPSS)
  - MME improves skill over individual models
- Higher probabilities represent periods of greater skill for extremes, or forecasts of opportunity
- Future work:
  - Optimize MME combination weighting
  - Identify conditional skill and forecasts of opportunity using possibly weather regimes or climate modes of variability

Monthly and Seasonal Outlooks

Strategies to Seasonal Climate Prediction

- Statistical forecasts
- Ensembles of dynamical predictions: Identifying signals
- Multi-model ensembles: Canceling systematic errors
- Statistical correction of systematic model errors or **Model post-processing**
- Hybrid statistical-dynamical forecasts from climate mode forecasts by dynamical models

Calibration, Bridging, and Merging

Calibration, Bridging, and Merging (CBaM)

Calibration, **Bridging**, and Merging (CBaM)

Dynamical model forecast of a relevant climate index (e.g., Niño 3.4)

Statistical post-processing →

Statistically bridged forecast of North American 2-m temperature

Calibration, Bridging, and **Merging** (CBaM)

Statistically bridged forecast of North American 2-m temperature

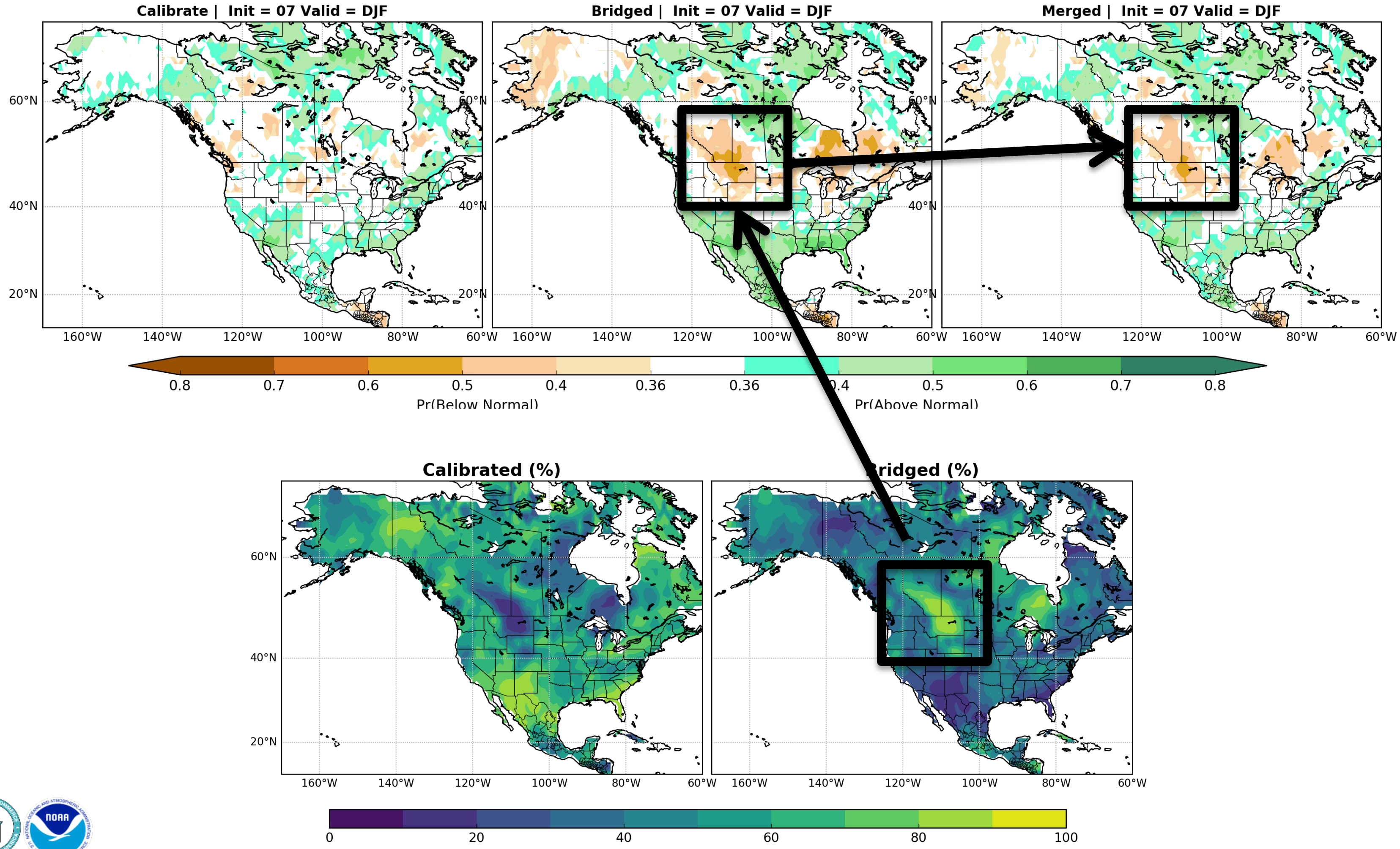
✖ W

✚

Weighted merging of forecasts based on performance in hindcast period

Statistically corrected (calibrated) forecast of North American 2-m temperature

✖ W

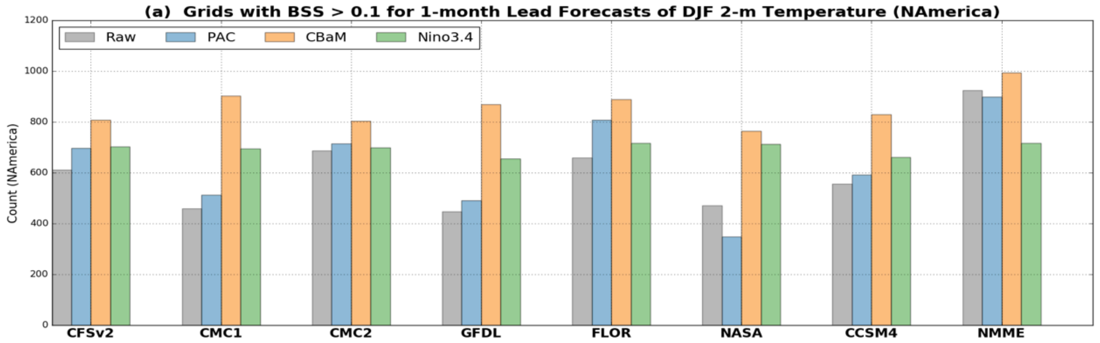


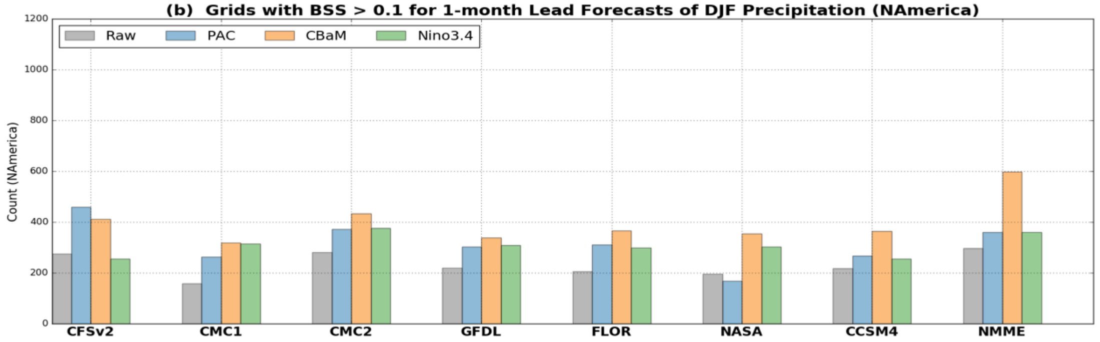
Calibrate | Init = 07 Valid = DJF
Bridged | Init = 07 Valid = DJF
Merged | Init = 07 Valid = DJF
60°N
40°N
20°N
160°W
140°W
120°W
100°W
80°W
60°W
0.8
0.7
0.6
0.5
0.4
0.36
0.36
0.4
0.5
0.6
0.7
0.8
Pr(Below Normal)
Pr(Above Normal)
Calibrated (%)
Bridged (%)
0
20
40
60
80
100

CBAM increases coverage of positive skill

North America grid points with Brier Skill Scores > 10%

CBaM in orange
PAC in blue





PAC calibrated NMME

Probability Anomaly Correlation (PAC)

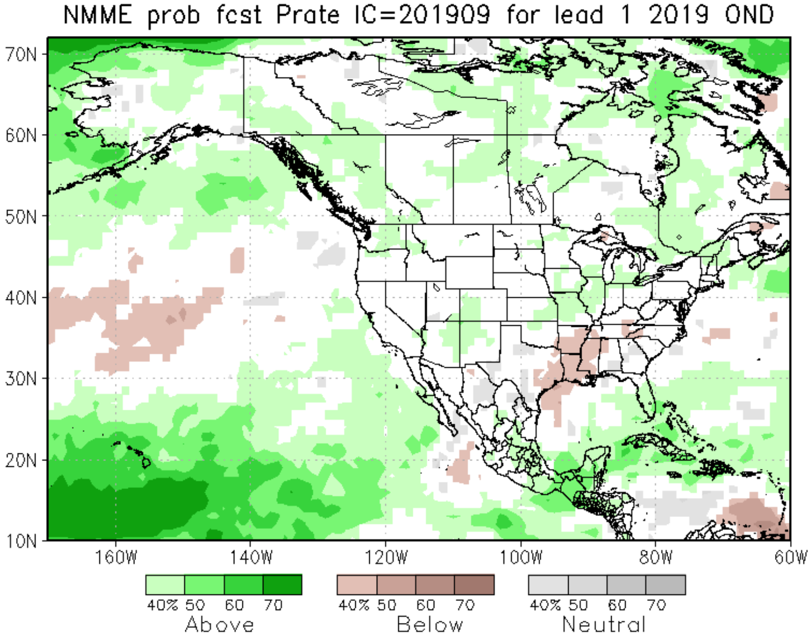
- Regression relating hindcast probability anomalies to observed probability anomalies
(+66.7 % when tercile is observed, -33.3% when not observed)

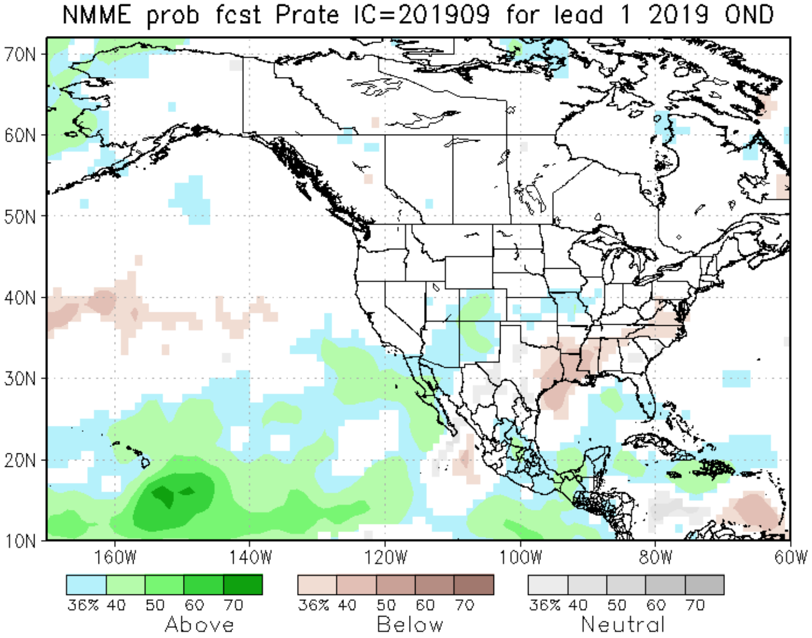
- Minimizes the Brier score (i.e., MSE for probabilities)

- Multiplies the forecast probabilities by regression coefficient

- van den Dool et al. (2017) (*Weather and Forecasting*) for more details

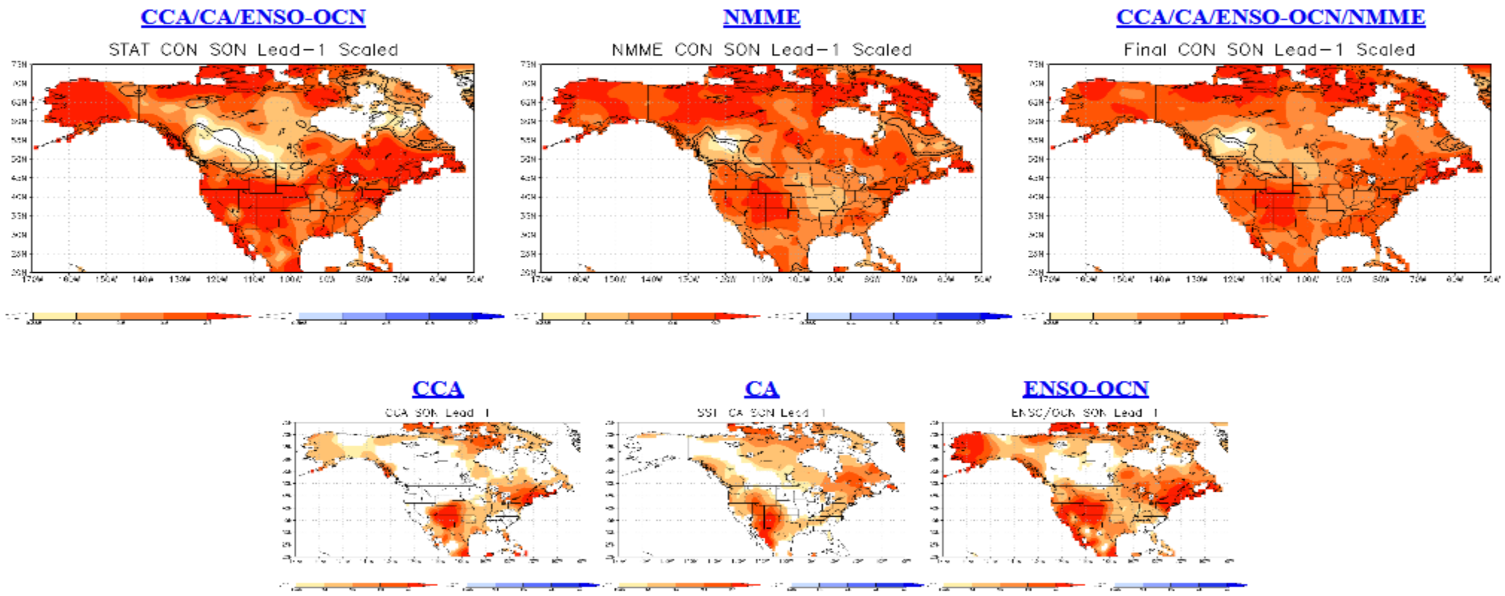
PAC calibrated NMME (right) damps ensemble count probabilities (left) based on skill





Probability Anomaly Correlation Consolidation

Seasonal Consolidation of Temperature and Precipitation, Using PAC



- Multiple statistical and dynamical tools are weighted according to Probability Anomaly Correlation skill, and combined into one map
- Combined forecast consolidation is calibrated using PAC in a final step (insuring reliability)

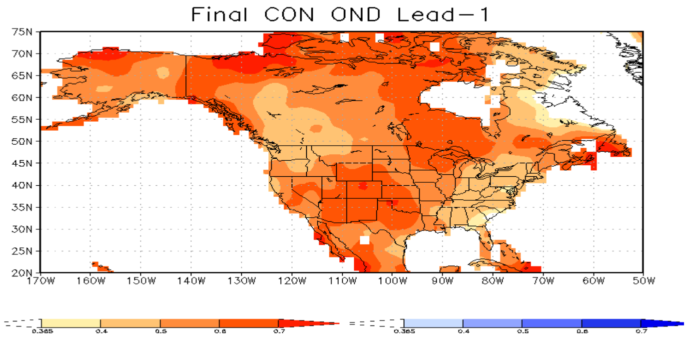
Consolidated Seasonal Forecast Tool

- Probabilistic seasonal forecast for temperature and precipitation created using probability anomaly correlation (PAC) for calibration of forecast inputs from various statistical and dynamical models.

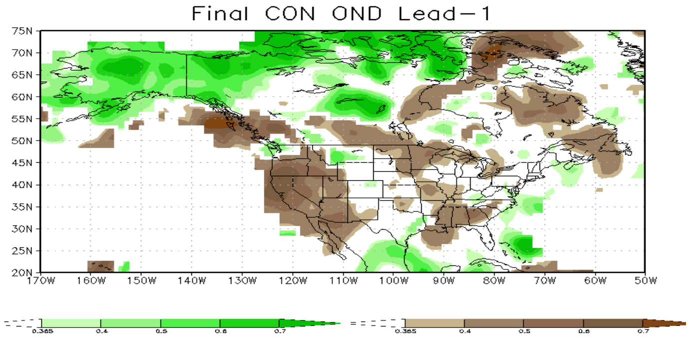
- Statistical and dynamical models are each consolidated as separate categories first, then combined for a final, hybrid statistical/dynamical probabilistic forecast.

- Output graphics for consolidated forecast plus individual member contributions can be found at http://www.cpc.ncep.noaa.gov/pacdir/ncca.html

Temperature

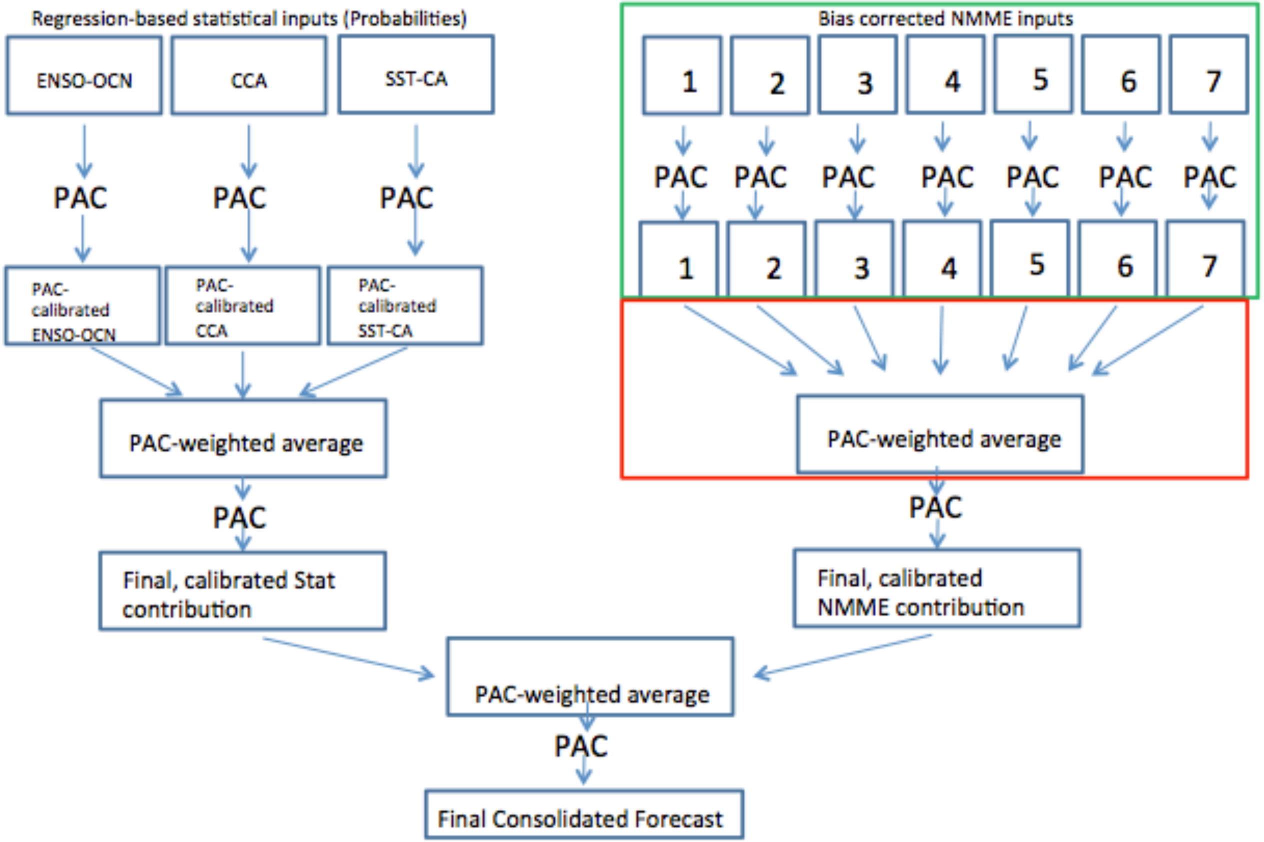


Precipitation



Seasonal Consolidation Flowchart

For Both Temperature and Precipitation (For each lead, season, above/below tercile)



Tools Summary Information

- Using Ensemble Prediction Systems & MME
- Calibrating probability forecasts such that they are reliable
    - Probability approximates the frequency that observations are in the correct category
- Consolidation:

  Combining dynamical models, accounting for complimentary skill
- Utilizing skill related to climate state
    - e.g. MJO or ENSO
    - Statistical Models & Hybrid Statistical-Dynamical (Bridging) models

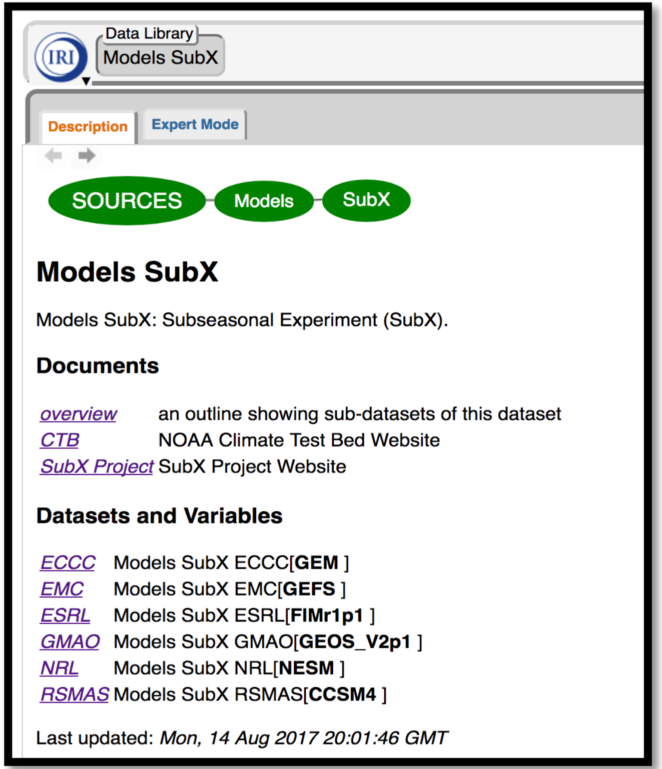
Thank you

Additional Material

Week of Hindcast Dates and Target Dates	Jan 1	Jan 2	Jan 3	Jan 4	Jan 5	Jan 6	Jan 7	Jan 8 Forecast Day	Week 3-4 Outlook: Jan 22 – Feb 05
Day of the week and Days to Target Dates	Fri 22:35	Sat 21:34	Sun 20:33	Mon 19:32	Tues 18:31	Wed 17:30	Thurs 16:29	Fri 15:28	2 weeks from Sat + 13 days → WK34
Center-Model	---------------------- Reforecast Grab Period -------------------								
ECCC-GEM 4 members 32 days	○	○	○	○	○	○	●	Forecast Day	● = Realtime
EMC-GEFS 11 members 35 days						●		Forecast Day	
ESRL-FIMv2 4 members 32 days						●		Forecast Day	
NASA-GEOS 4 members 45 days	○	○	○	○	○	○	○	Forecast Day	* GEOS5 roves in Realtime
NCEP-CFSv2 4 members 44 days						●		Forecast Day	
NRL-NESM 4 lagged members 45 days		●	●	●	●			Forecast Day	
RSMAS-CCSM4 3 members 45 days	○	○	●	○	○	○	○	Forecast Day	
Coming Soon: NCAR-CESM 10 members 45 days				●				Forecast Day	

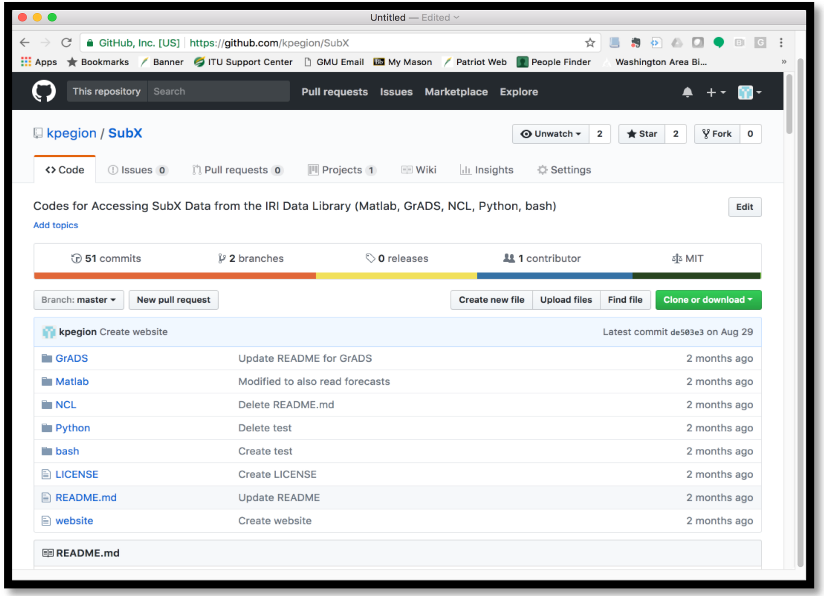
Real-time and Re-forecast Database

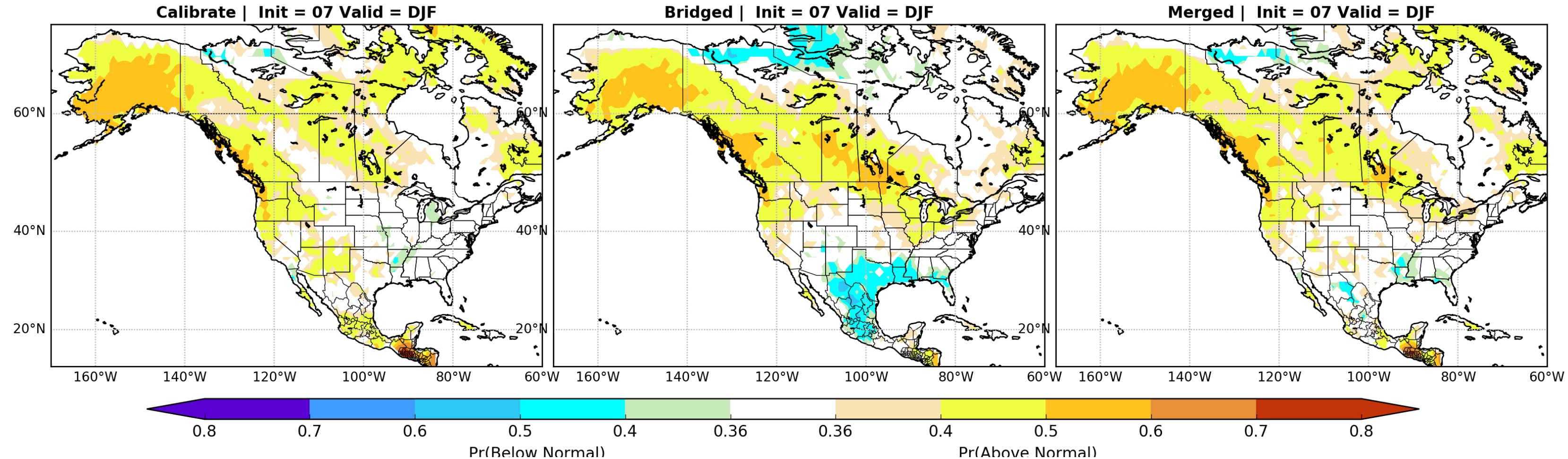
Data publicly available from the IRI Data Library

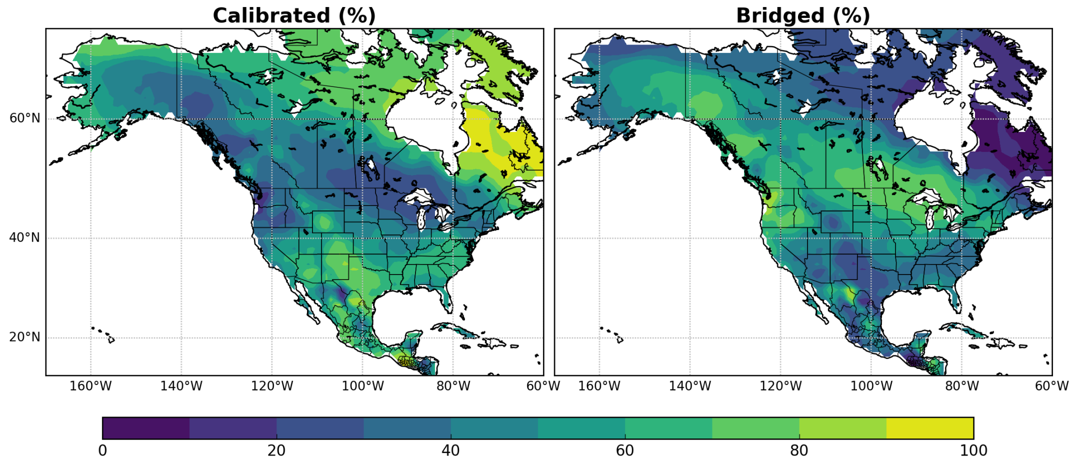


http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/

Codes to download on github




Calibrate | Init = 07 Valid = DJF
Bridged | Init = 07 Valid = DJF
Merged | Init = 07 Valid = DJF
60°N
40°N
20°N
160°W 140°W 120°W 100°W 80°W 60°W
0.8 0.7 0.6 0.5 0.4 0.36 0.36 0.4 0.5 0.6 0.7 0.8
Pr(Below Normal)
Pr(Above Normal)


Calibrated (%)
Bridged (%)
60°N
40°N
20°N
160°W 140°W 120°W 100°W 80°W 60°W
0 20 40 60 80 100

BSS: NMME calibrated, bridged, & merged DJF forecasts