Multi-model Prediction on Subseasonal Time-scales at the US NOAA Climate Prediction Center:

Approaches to Calibration and the Identification of Forecasts of Opportunity

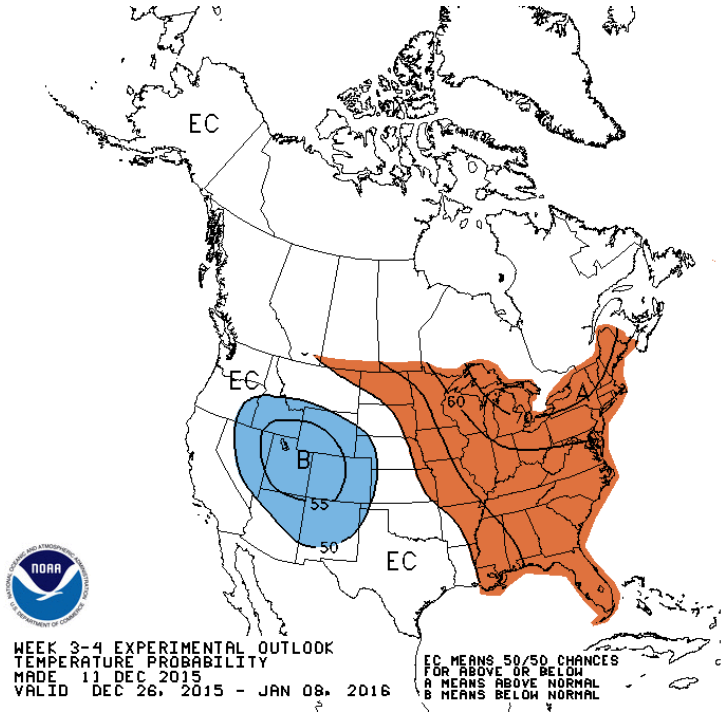
Dan C Collins, Sarah Strazzo, Emerson LaJoie, Emily Becker, and Jon Gottschalck

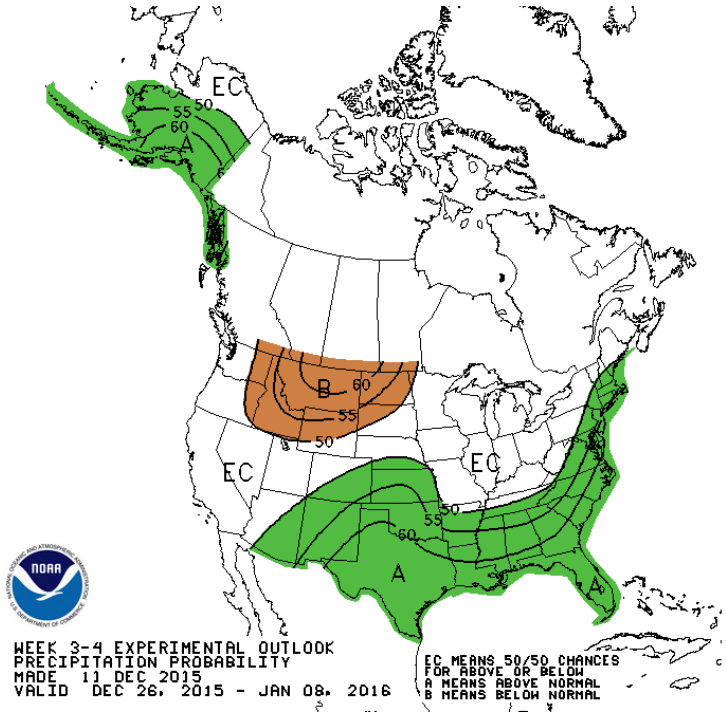
NOAA Climate Prediction Center

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

Above and below normal

Examining extremes and hazards into week 3&4

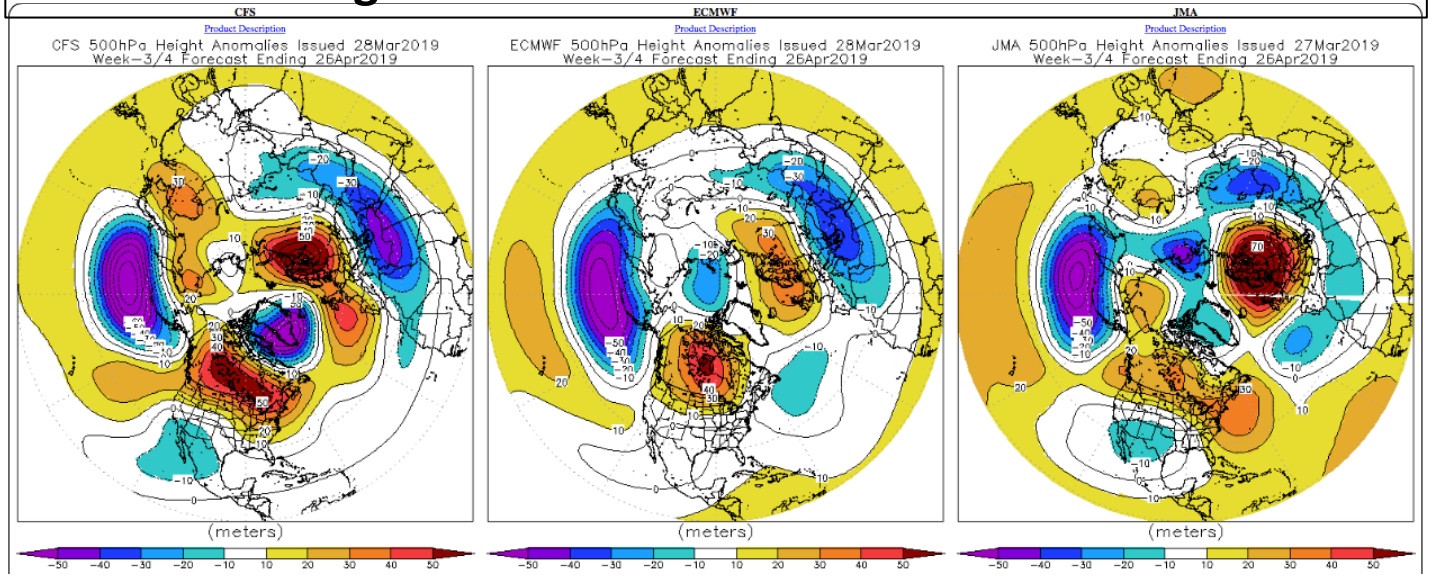




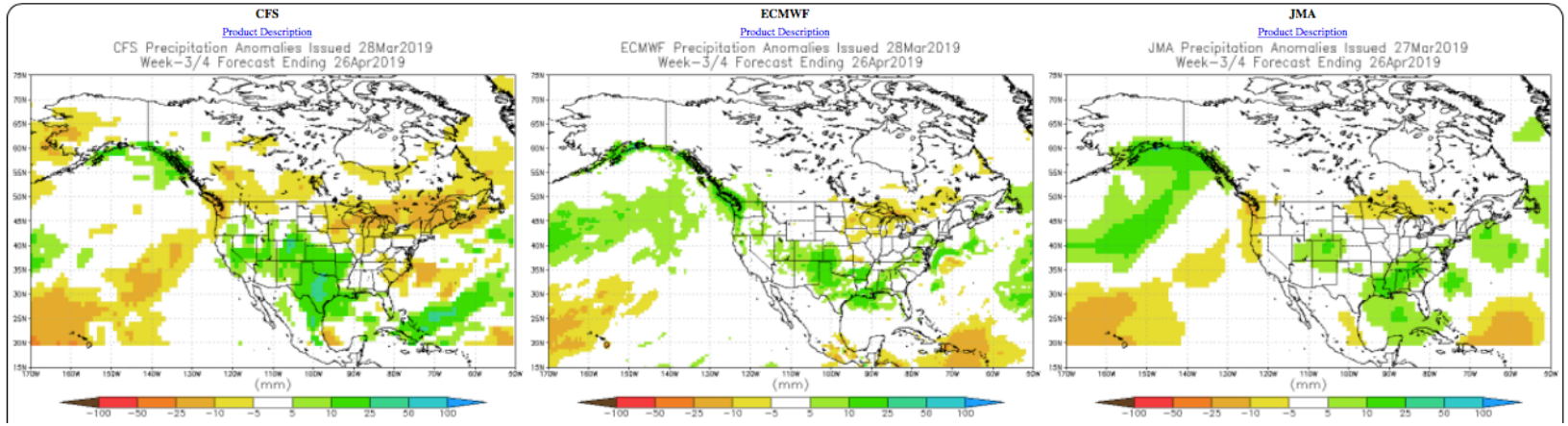
- Probability of above and below normal temperature and precipitation
- Use a combination of dynamical and statistical model forecasts
- MME guidance plays a primary role in the subseasonal forecast

Operational model guidance: NCEP CFS, ECMWF & JMA

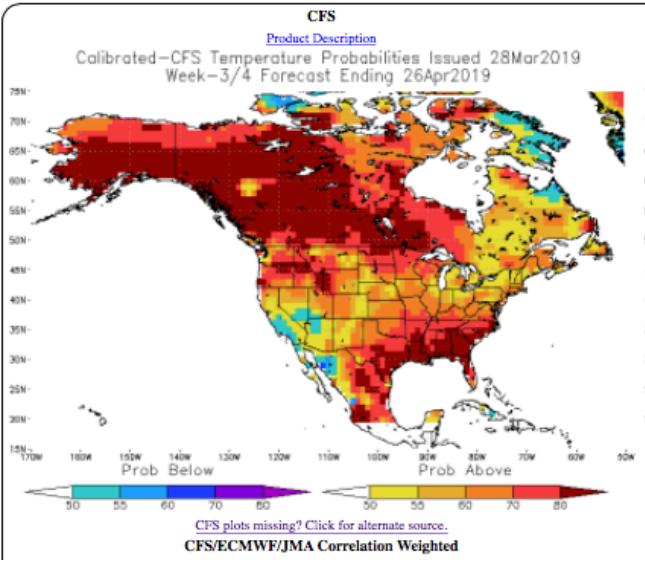
- **500 hPa height anomalies**

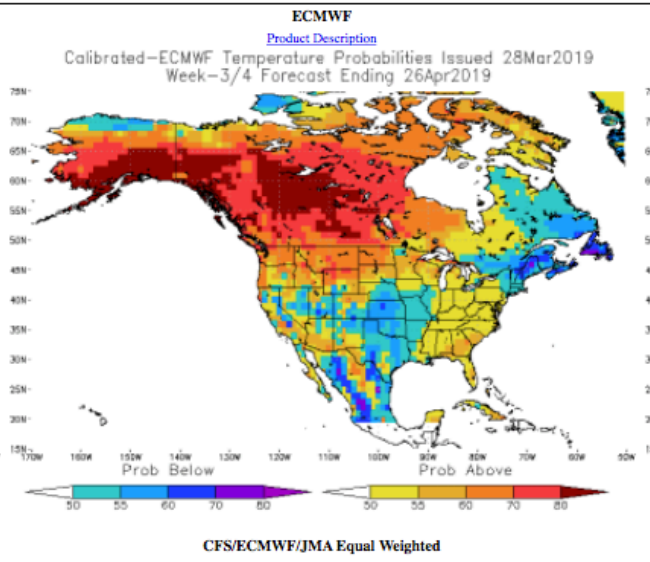


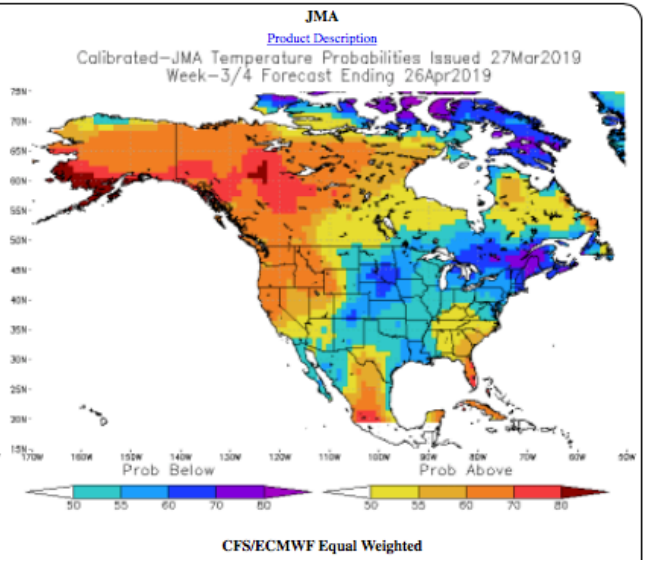
- **Precipitation anomalies**



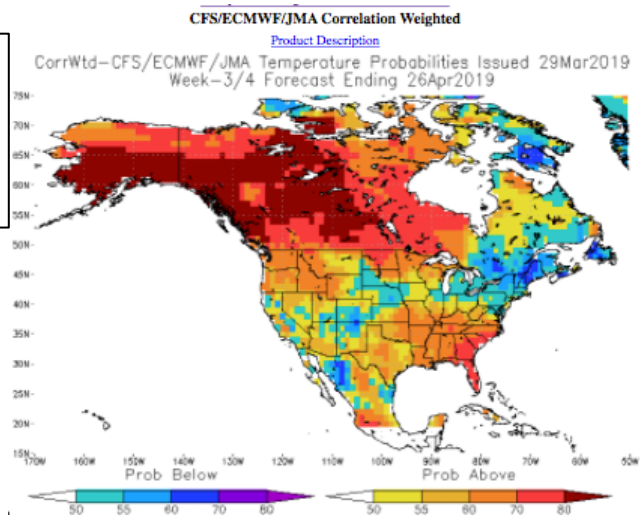
- Each dynamical model is bias corrected using model hindcast
- Calibrated PDFs made using hindcast skill* (heterogeneous hindcast samples)







Correlation weighted MME:
Above and below median
2-m temperature



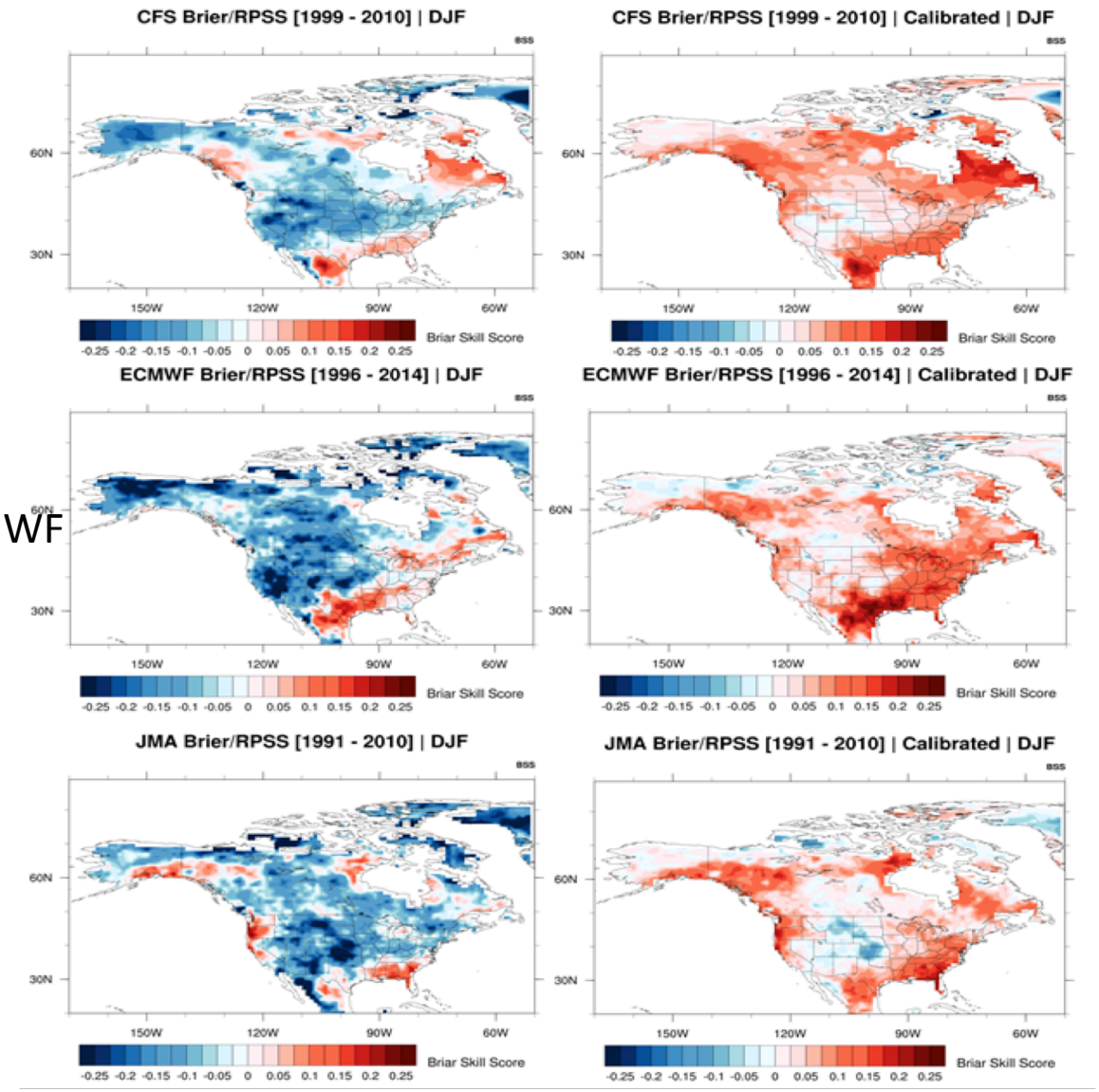
*Ensemble Regression,
Unger et al., 2009

Brier Skill Scores of DJF temperature forecasts :

CFS

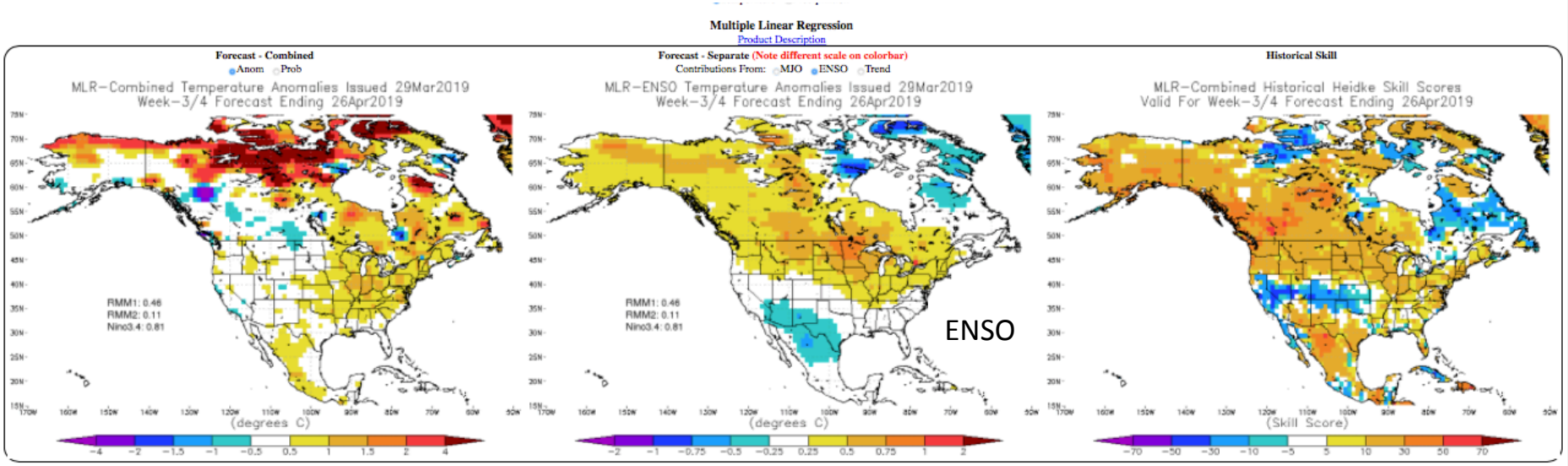
ECMWF

JMA



No calibration (left) and calibrated (right)

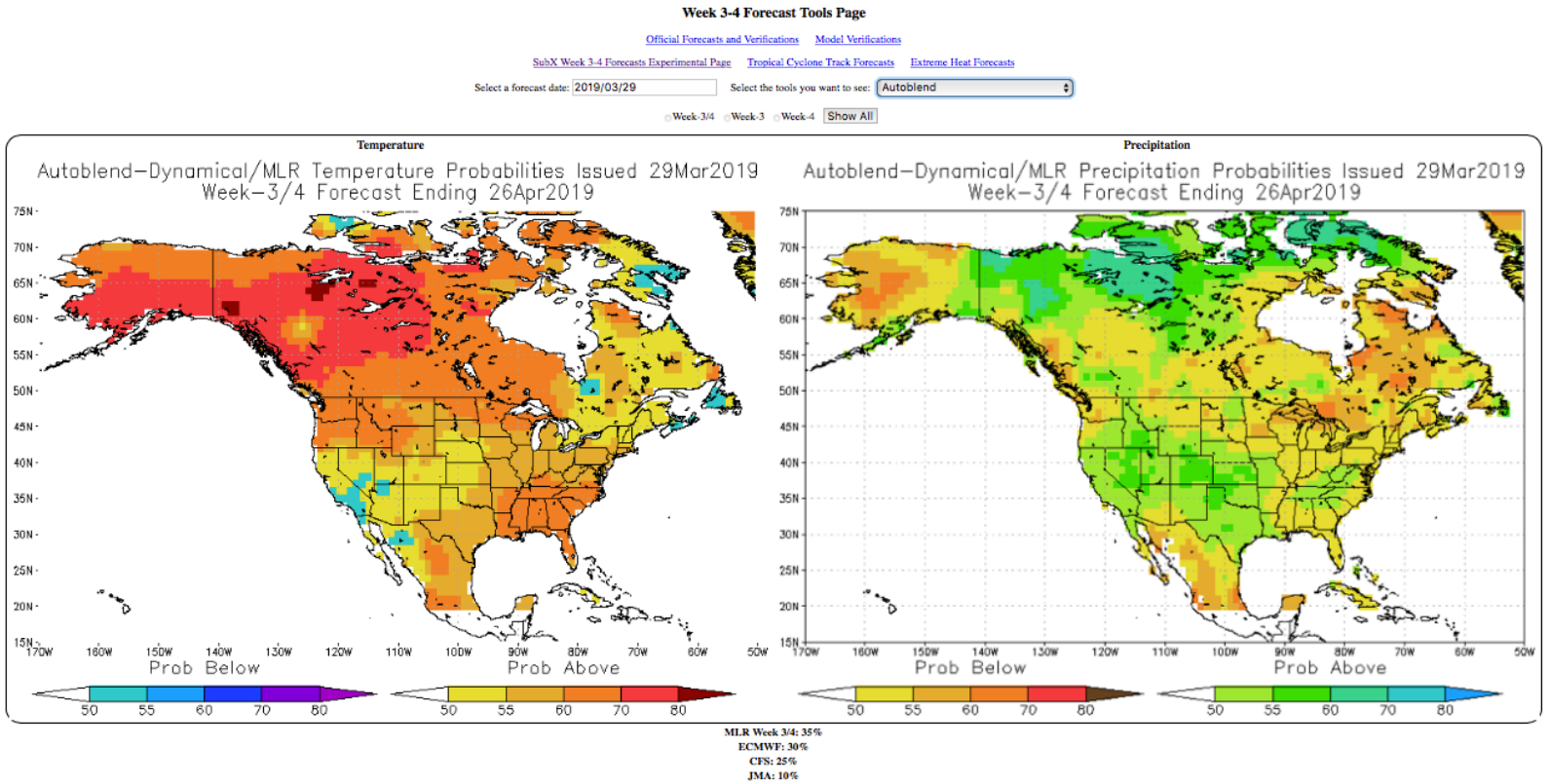
Statistical model guidance: Multiple Linear Regression



- 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

Automated Blend: Combining statistical and calibrated dynamical models

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



- Experimenting with methods for determining weights, and…
- Combining calibrated forecasts to make a calibrated blend

SubX vs. S2S archives

- SubX protocol for hindcast years & frequency (at least weekly)
  - S2S has heterogeneous hindcasts

- SubX combines operational (NCEP, ECCC) and experimental models
  - S2S models are operational

- SubX forecast data is available in near real time to any user, through the IRI Data Library:
  http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/

  - S2S forecasts available on a 3-week delay

SubX : The Subseasonal Prediction Experiment

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

- Hindcasts (1999-2015)
  - More than **18 months** of **weekly** real-time forecasts
  - 7 **operational or experimental** ensemble models

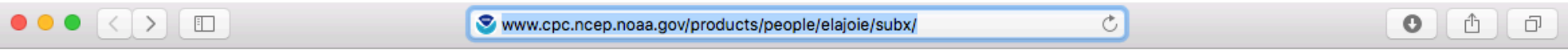
- Facilitating MME and examining the value to subseasonal
  - Model calibration and multi-model ensemble combination
  - Addition of new experimental systems (e.g. NCAR CESM2)
  - Assessing the added value to operational models

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

1 week of SubX	Jan 2	Jan 3	Jan 4	Jan 5	Jan 6	Jan 7	Jan 8	**Jan 9** **Forecast Day**	**Week 3-4 Outlook: Jan 24 – Feb 06**
Day of the week & 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: Sat + 13 days (Fri) → WK34**
Center-Model	---------- **Forecast Initialization Period** ----------								
ECCC-GEM 4 members 32 days								**Forecast Day**	**Hindcast & Real-time forecast**
EMC-**GEFS** 11 members 35 days								**Forecast Day**	**Variable hindcast days**
ESRL-**FIMv2** 4 members 32 days								**Forecast Day**	**Variable real-time & hindcast days**
NASA-GEOS 4 members 45 days								**Forecast Day**	
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 in next year: CESM-46LCAM5 10 members 45 days								**Forecast Day**	***MME for week 3-4 collected for each week in the hindcast***
CESM-30LCAM5 10 members 45 days								**Forecast Day**	

Courtesy of E. LaJoie

CPC SubX guidance



SubX Week 3/4 forecasts

SubX : The Subseasonal Prediction Experiment Project

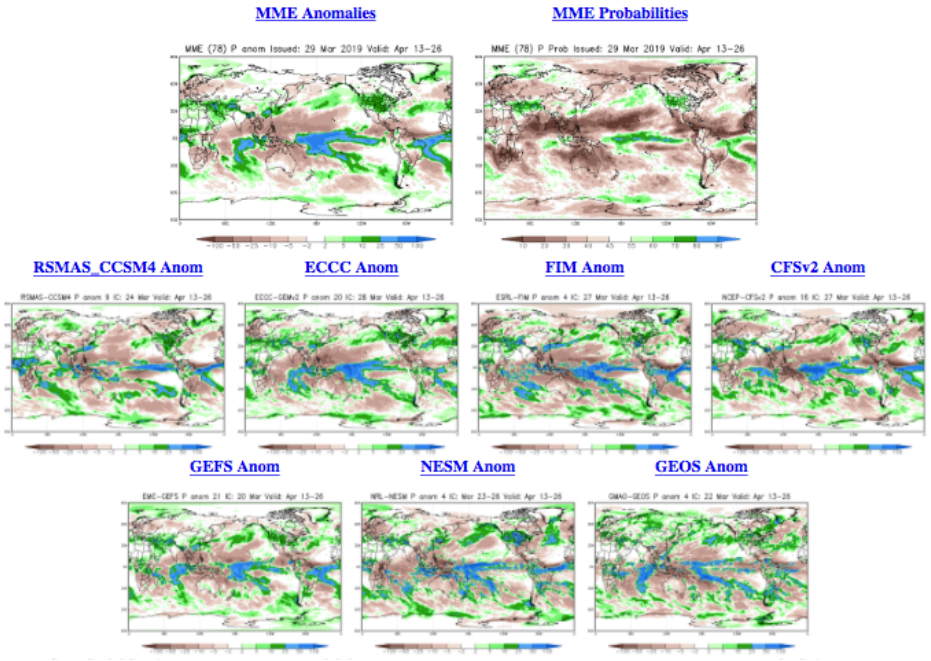
IRI Data Library

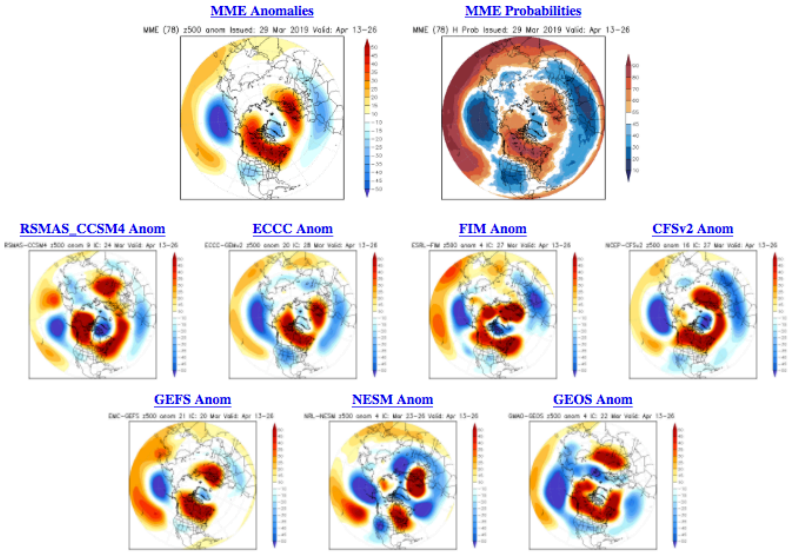
Week 3/4 Operational Model Forecasts

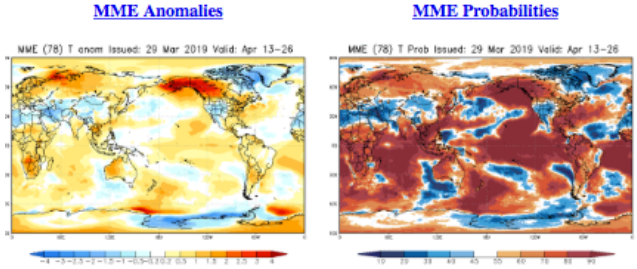
North America	Global
500-hPa height	500-hPa height
2-m Temperature	2-m Temperature
Precipitation	Precipitation

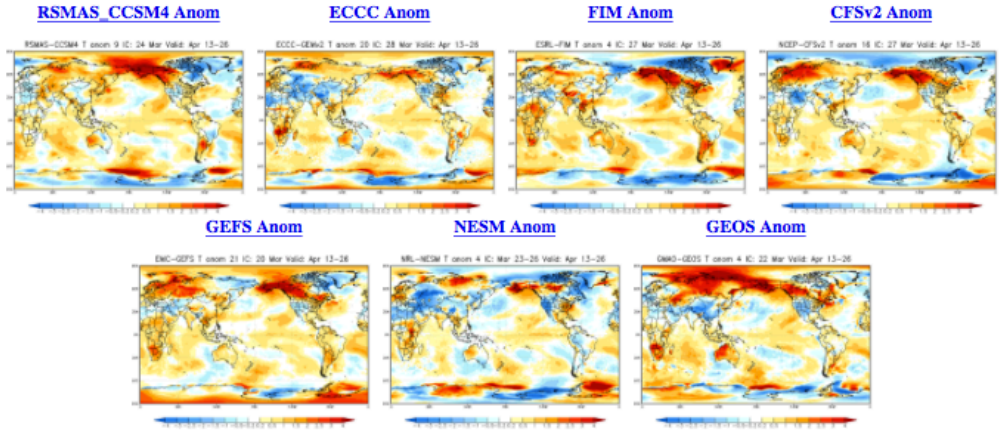
http://www.cpc.ncep.noaa.gov/products/people/elajoie/subx/

SubX Week 3-4 guidance





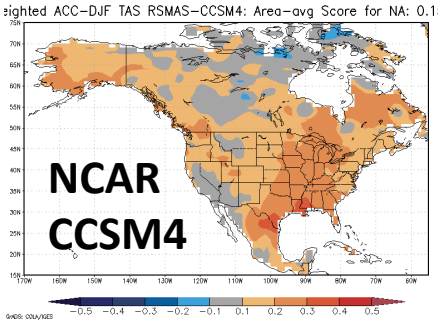


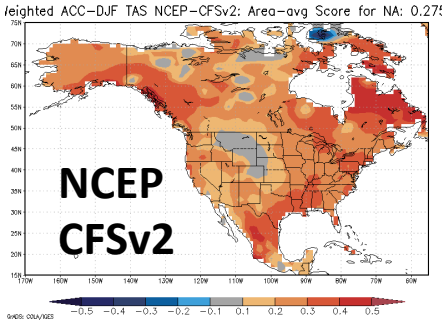


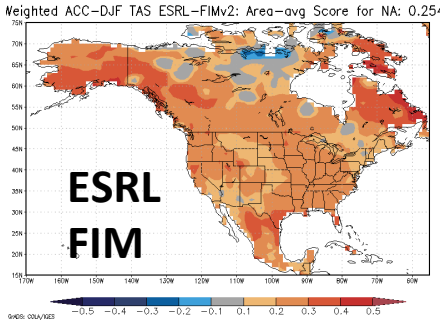
Courtesy of E. LaJoie

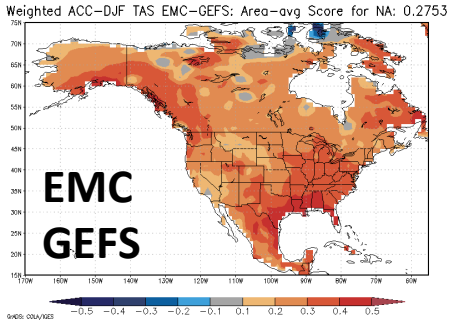
Anomaly Correlation by model & MME (DJF)

- **MME outperforms any individual model**

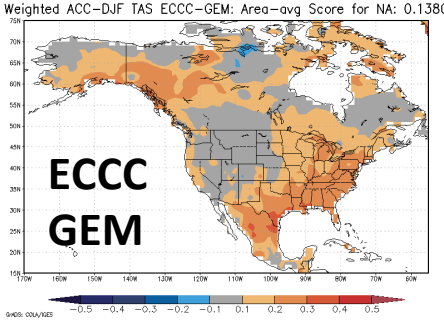


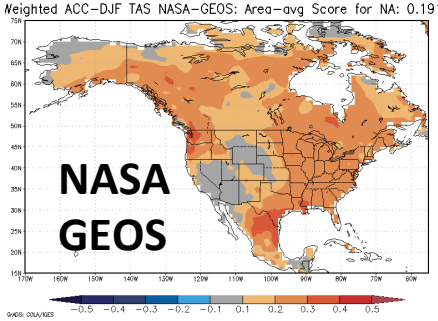


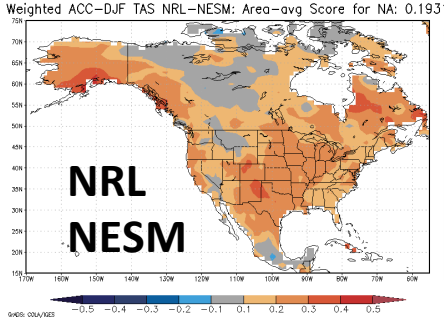


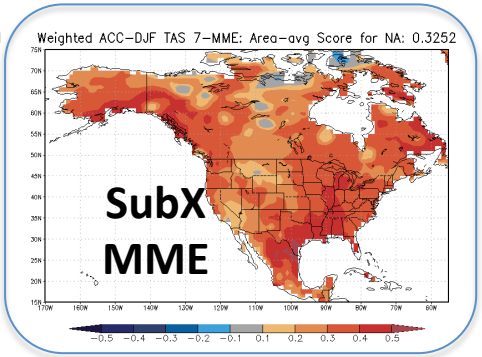


Individual Model and MME Anomaly Correlations









Courtesy of E. LaJoie

Calibration of ensembles to obtain reliable probabilities

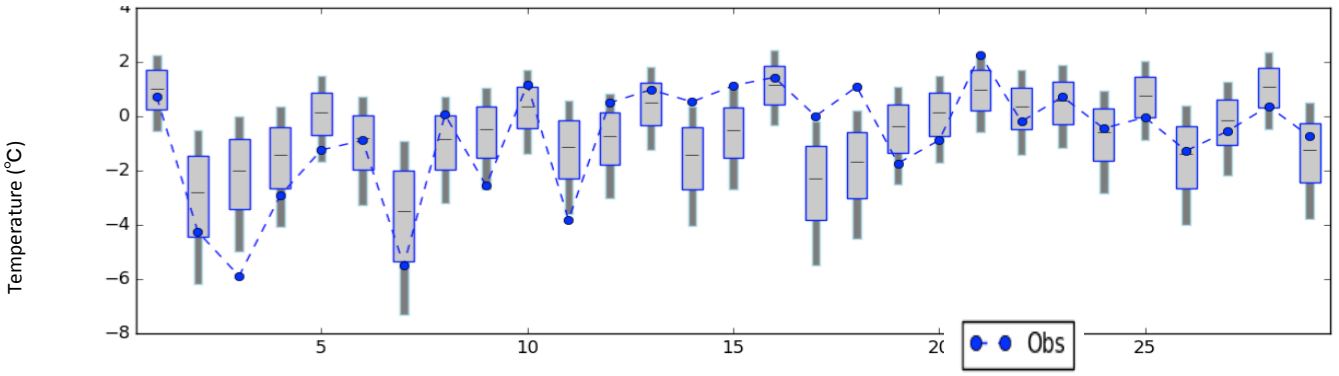
… while retaining skill

- 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 uses **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.
    - Mini-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

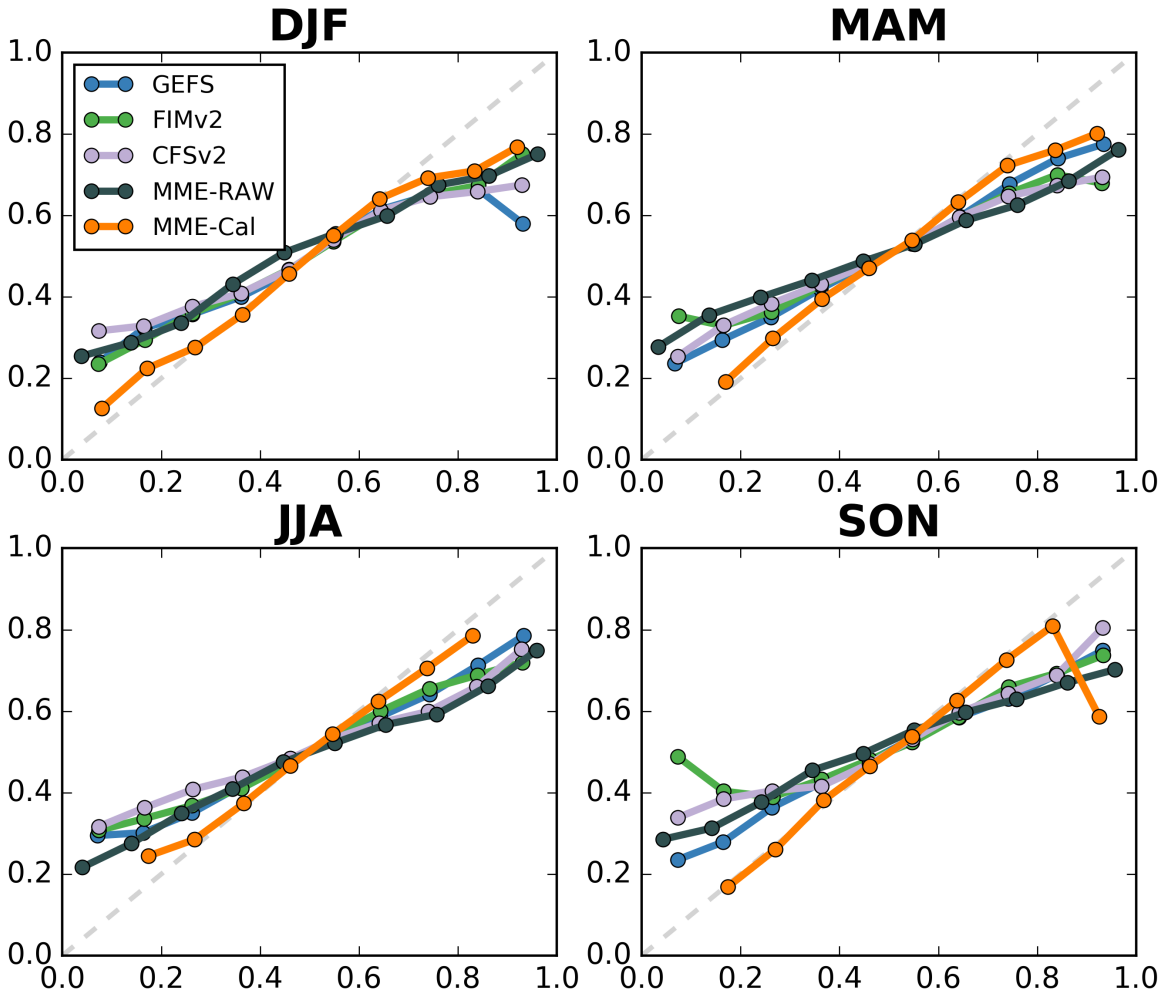
1 week of SubX	Jan 2	Jan 3	Jan 4	Jan 5	Jan 6	Jan 7	Jan 8	**Jan 9** **Forecast Day**	**Week 3-4 Outlook: Jan 24 – Feb 06**
Day of the week & Days to Target Dates	Fri 22:35	Sat 21:34	Sun 20:33	Mon 19:32	Tues 18:31	Wed 17:30	Thurs 16:29	**Fri 15:28**	**2 weeks: Sat + 13 days (Fri) → WK34**
Center-Model	---------- **Forecast Initialization Period** ----------								
ECCC-GEM 4 members 32 days								**Forecast Day**	**Hindcast & Real-time forecast**
EMC-**GEFS** 11 members 35 days								**Forecast Day**	**Variable hindcast days**
ESRL-**FIMv2** 4 members 32 days								**Forecast Day**	**Variable real-time & hindcast days**
NASA-GEOS 4 members 45 days								**Forecast Day**	
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 in next year: CESM-46LCAM5 10 members 45 days								**Forecast Day**	***MME for week 3-4 collected for each week in the hindcast***
CESM-30LCAM5 10 members 45 days								**Forecast Day**	

3-Model Mini-MME

Courtesy of E. LaJoie

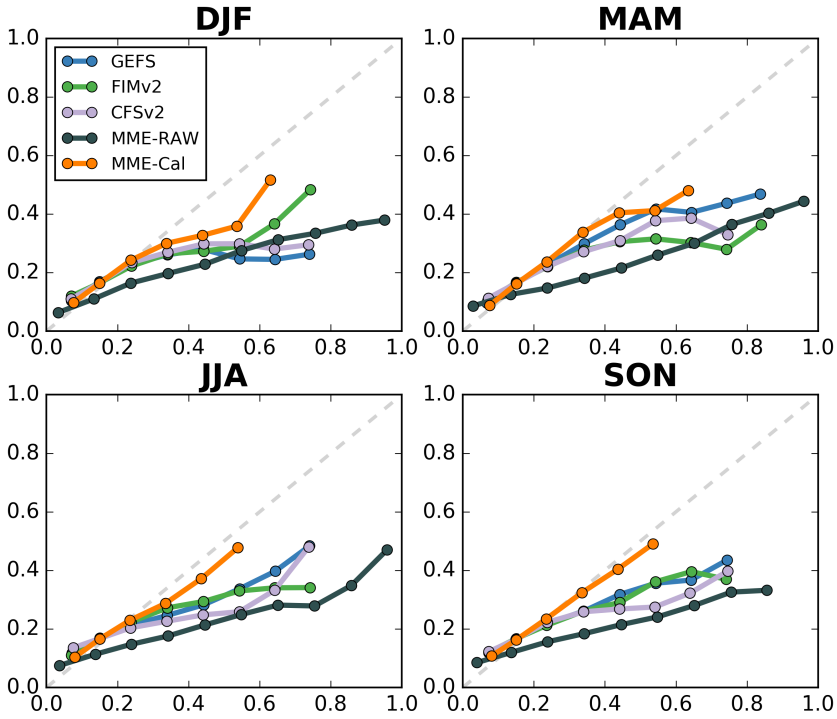
Above / below normal temperature reliability

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

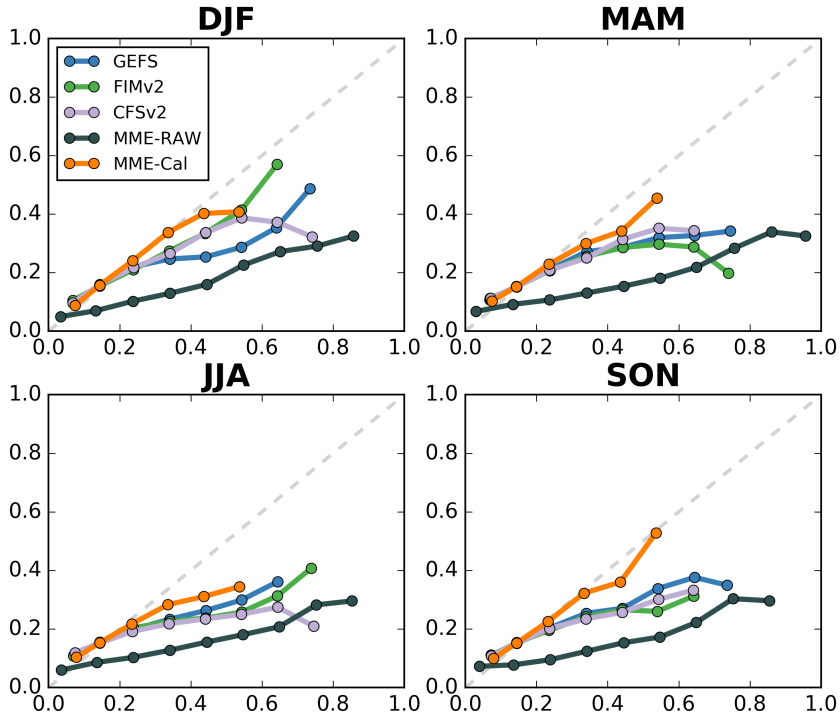


Extreme above/below normal reliability (high and low 15th percentile)

- **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

Are Calibrated Probabilistic MME forecasts more skillful than individual models and un-calibrated ?

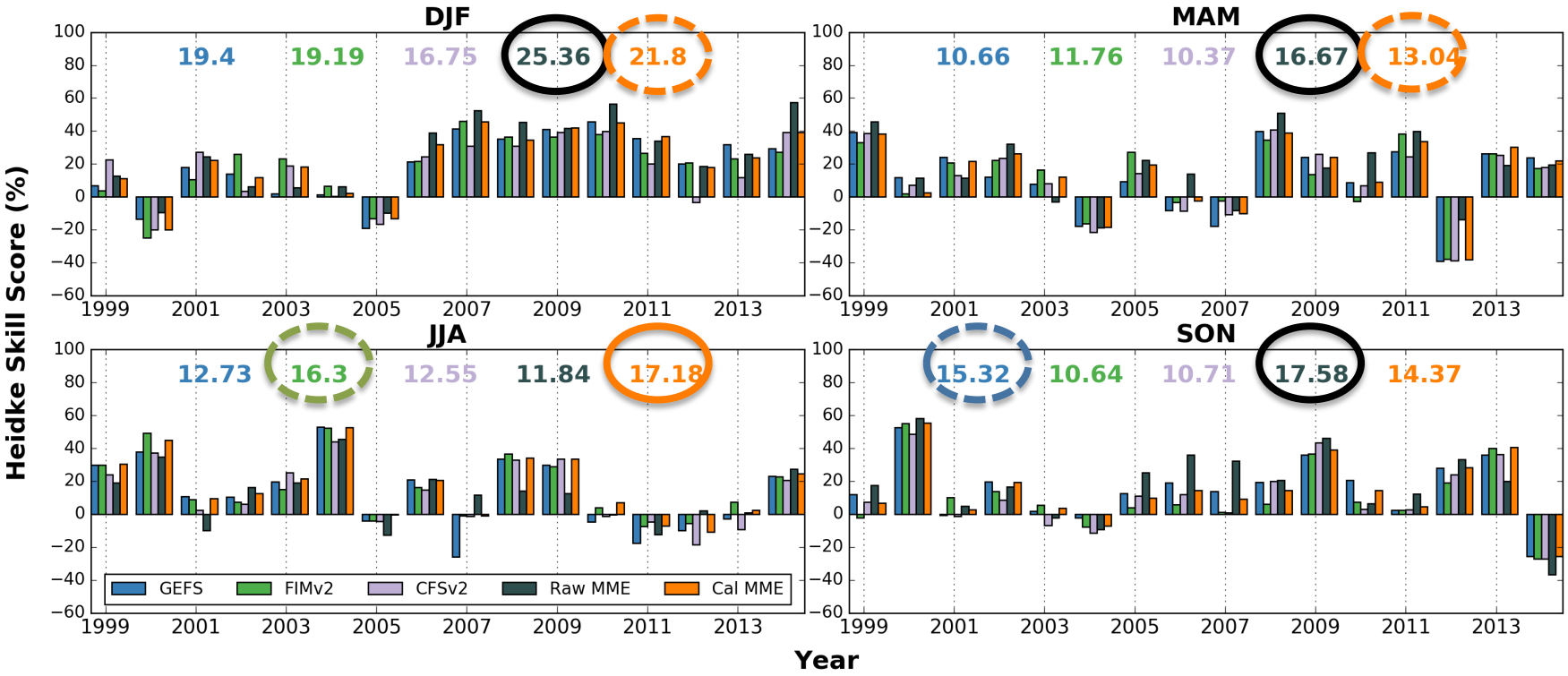
- Heidke Skill Score : Improvement of hit rate relative to random forecast of 15% frequency of occurrence

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

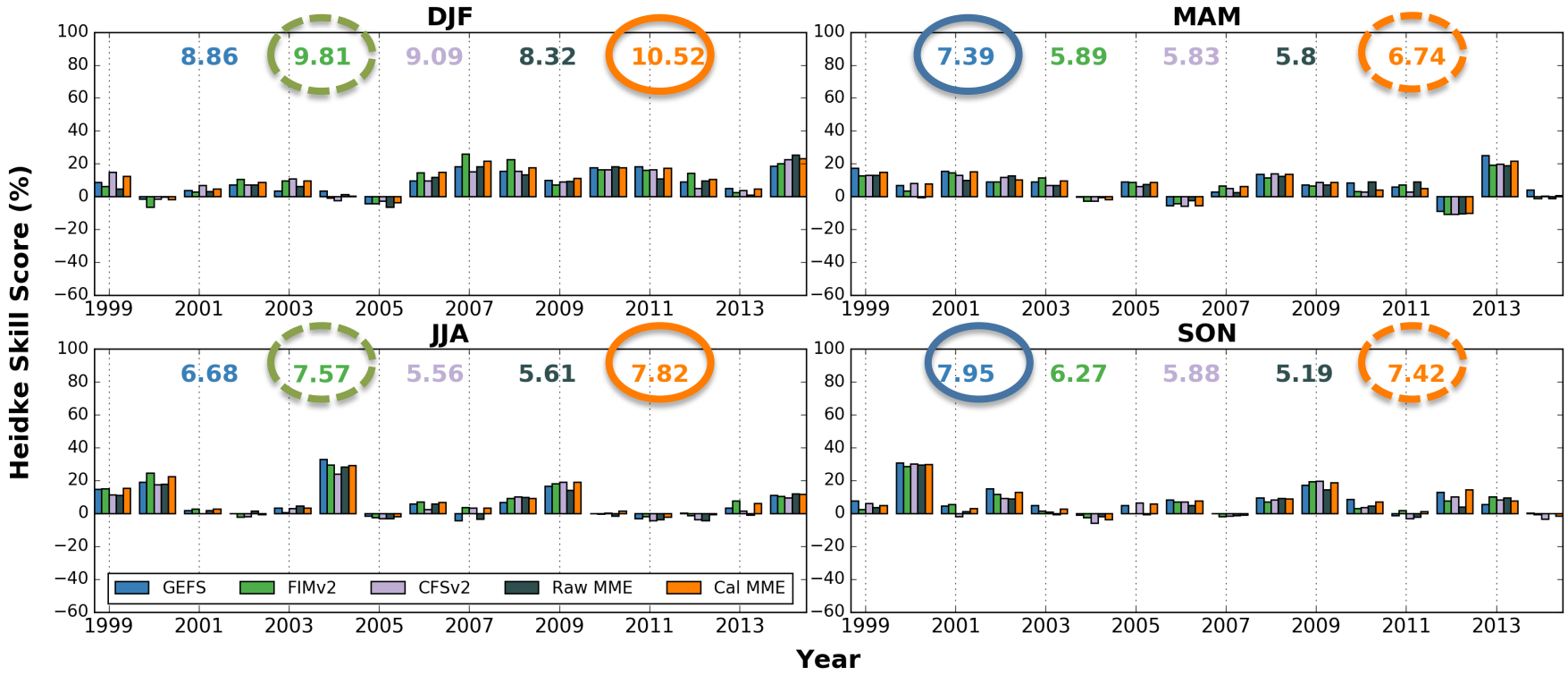


Courtesy of S. Strazzo

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**

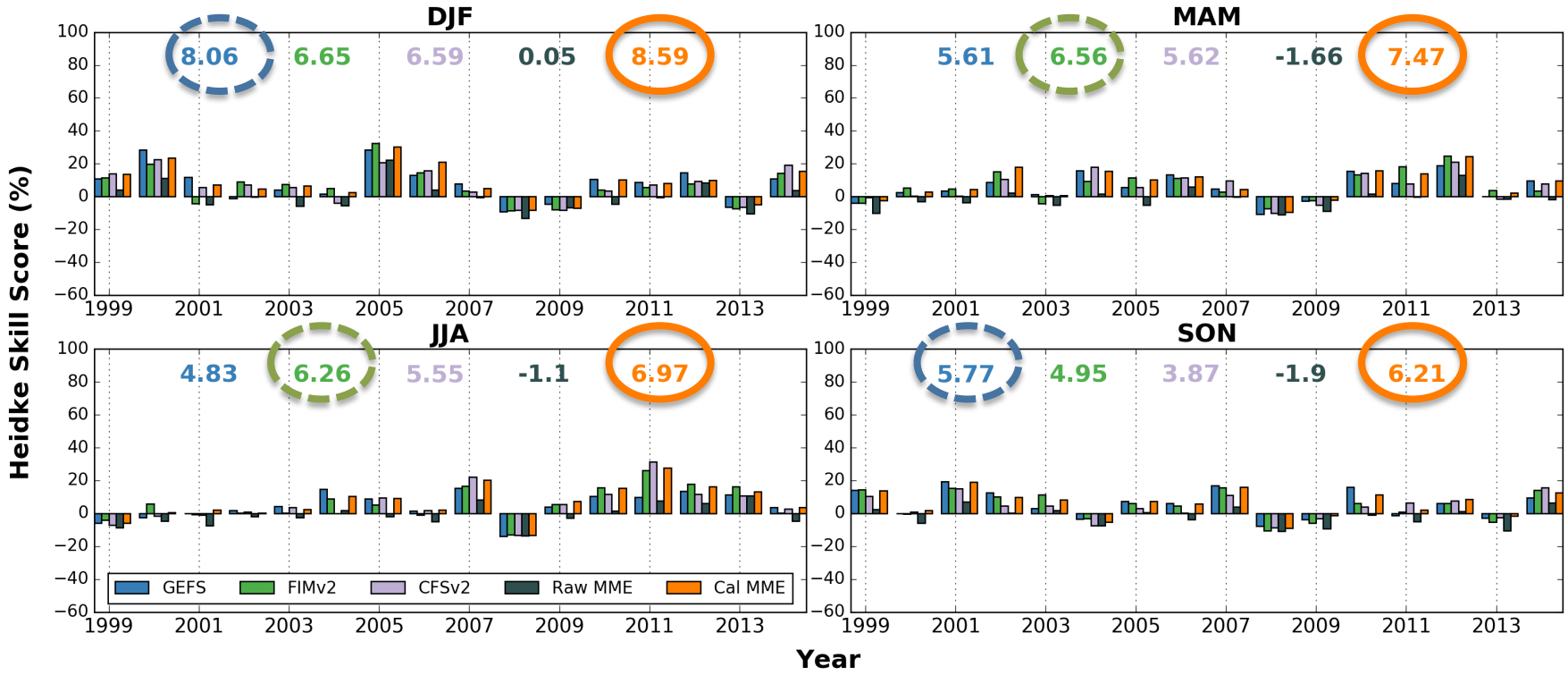


Courtesy of S. Strazzo

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 S. Strazzo

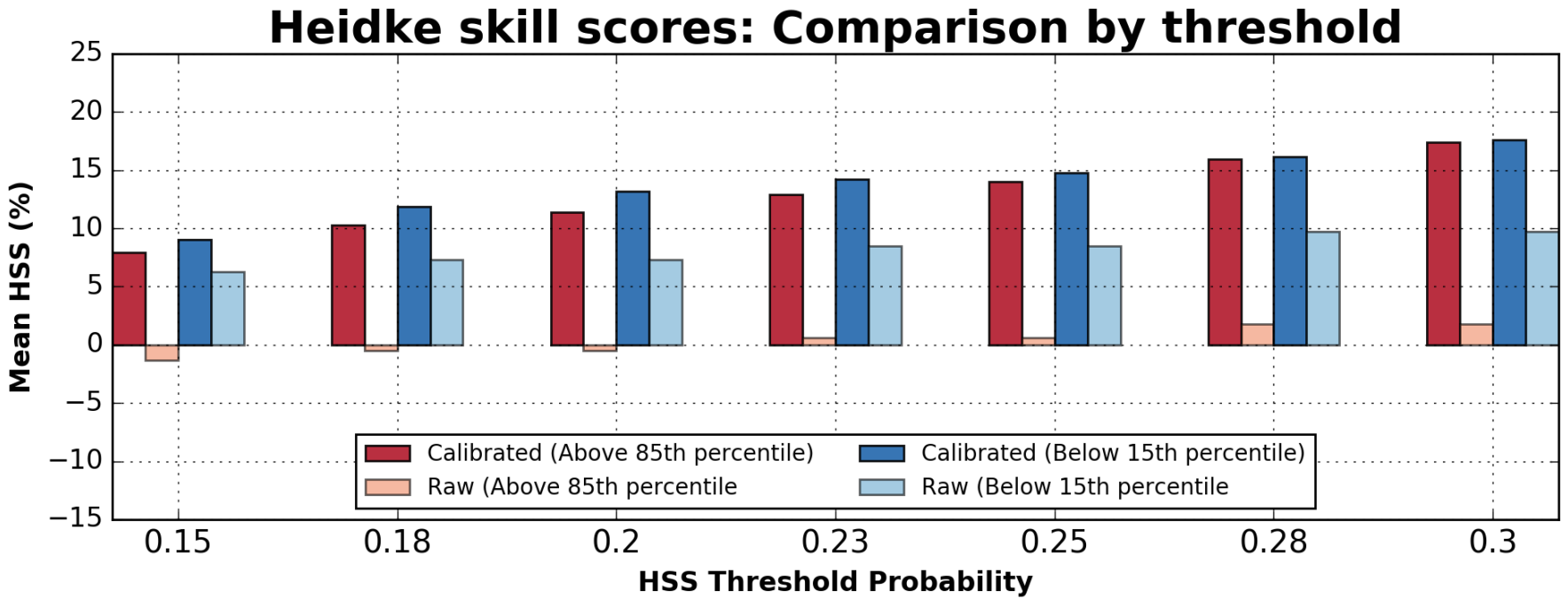
“Forecasts of Opportunity”

Can intermittent skill of forecasts be identified prior to forecast?

Extreme above / below normal temperature

- **Calibration** improves Heidke Skill Score of raw extremes forecast at all probability levels

- **Greater probability implies greater skill**



Courtesy of S. 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