

## Seasonal Prediction of North American Temperature and Precipitation Using the Calibration, Bridging, and Merging (CBaM) Method

Sarah Strazzo<sup>1,2</sup>, Dan Collins<sup>1</sup>, Andrew Schepen<sup>3</sup>, Q.J. Wang<sup>4</sup>, Emily Becker<sup>1,2</sup>, and Liwei Jia<sup>1,2</sup>

<sup>1</sup>Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD

<sup>2</sup>Innovim, LLC, Greenbelt, MD

<sup>3</sup>CSIRO Land and Water, Dutton Park, Queensland, Australia

<sup>4</sup>University of Melbourne, Parkville, Victoria, Australia

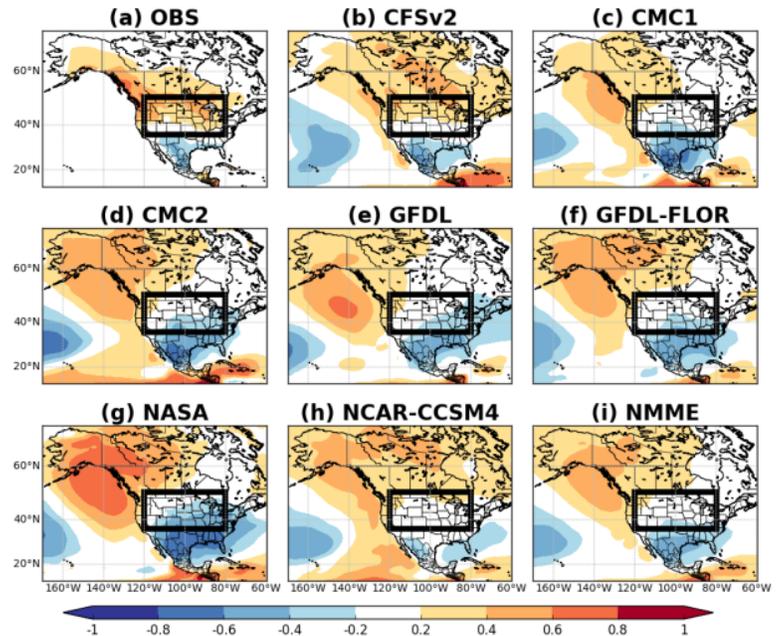
### 1. Introduction

NOAA's operational seasonal outlooks rely upon forecasts from the North American Multi-Model Ensemble (NMME; Kirtman *et al.* 2014). While the NMME remains a state-of-the-art tool for seasonal forecasting, recent research suggests that the NMME does not consistently reproduce the observed relationship between the El Niño/Southern Oscillation (ENSO) and temperature over North America (Chen *et al.* 2017). As an example, Fig. 1 shows the 1-month lead forecast December–January (DJF) ENSO–temperature teleconnection pattern from the NMME (1b–i) compared to the observed pattern (1a). Several of the NMME member models fail to represent the observed teleconnection pattern over the NMME hindcast period (1982–2010). For example, the CMC2, GFDL, GFDL-FLOR, and NASA models suggest a negative correlation ( $-0.6 < r < -0.2$ ) between ENSO and temperature over portions of the Midwestern United States where the observed correlation is positive ( $0.2 < r < 0.6$ ).

Given the short duration of the NMME hindcast period and the observed variability in both ENSO events and remote responses to ENSO events during this period, it is difficult to evaluate the skill of model teleconnection patterns. Regardless, recent results support the notion that such discrepancies contribute to reduced forecast skill (Chen *et al.* 2017). Prior work has applied Bayesian post-processing to general circulation model (GCM) output to both calibrate GCM forecasts and correct for model misrepresentation of relevant teleconnections (Schepen *et al.* 2014, 2016). Here we test this Bayesian post-processing method – known as Calibration, Bridging, and Merging (CBaM) – using the NMME hindcast dataset with the ultimate goal of improving forecast skill over regions where the NMME misrepresents the ENSO teleconnection.

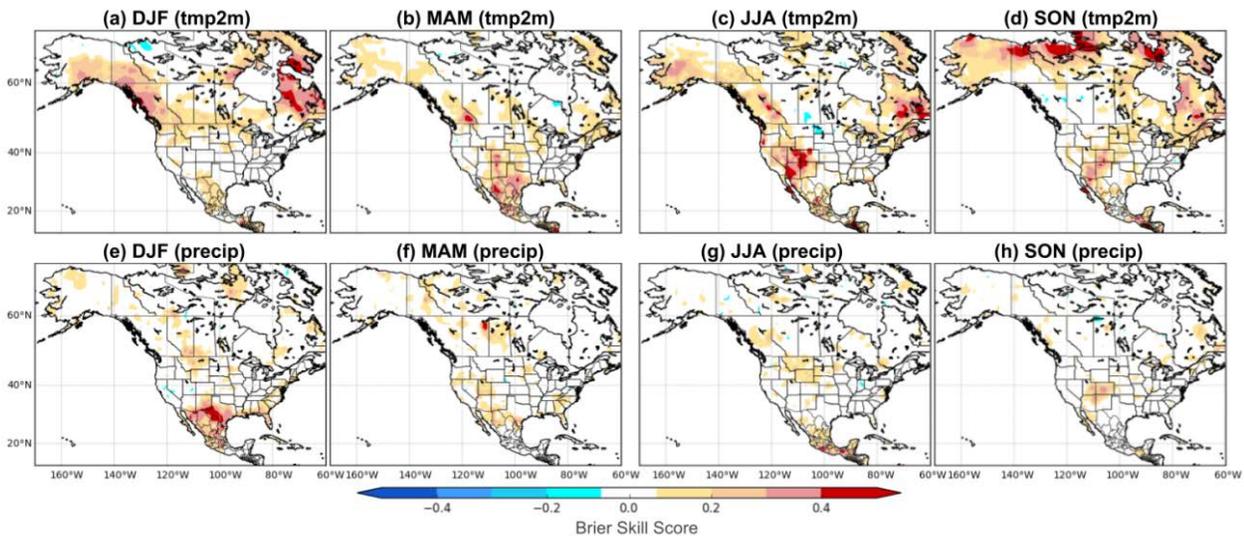
### 2. Data and methods

The CBaM method for post-processing dynamical model forecasts employs Bayesian joint probability (BJP) modeling (Wang *et al.* 2009). Using NMME reforecasts and observed SST, 2-m temperature, and

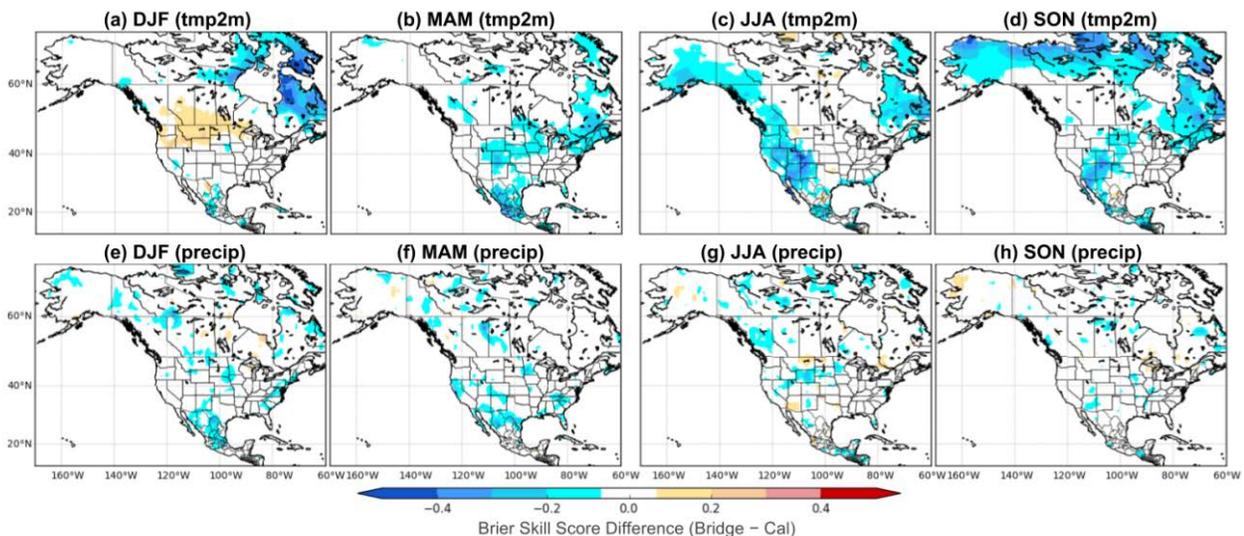


**Fig. 1** Correlation between DJF Niño3.4 SST anomalies and 2-m temperature over North America for the NMME hindcast period 1982–2010, a) observation, b)–h) 1-month lead forecast by ensemble members, and i) by NMME.

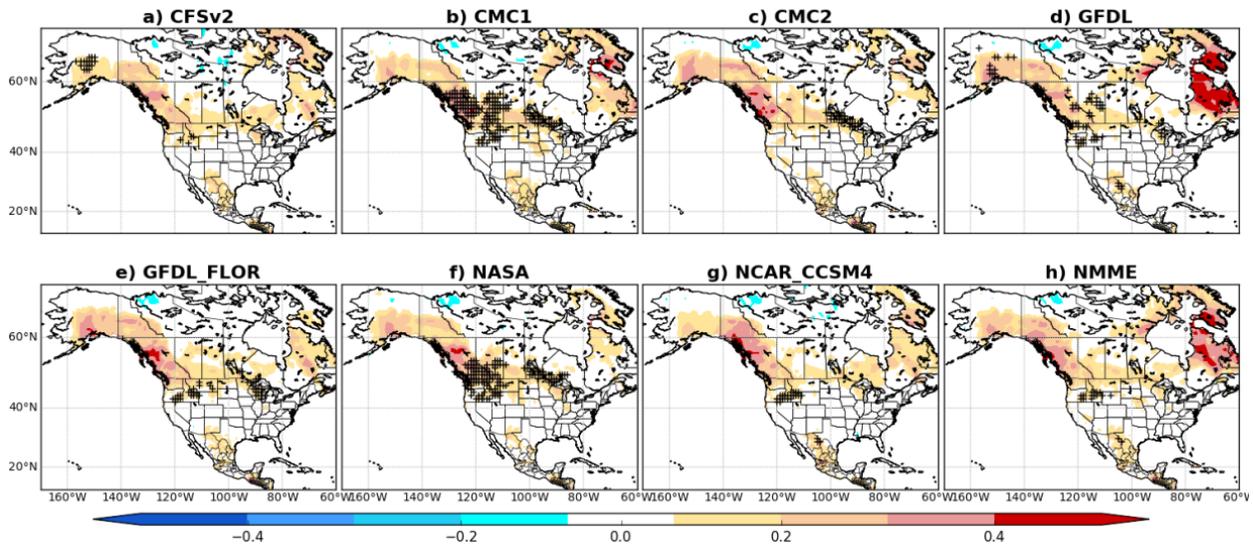
precipitation data from the Optimum Interpolation SST (OISST), Global Historical Climatology Network+ Climate Anomaly Monitoring System (GHCN+CAMS), and Climate Prediction Center Merged Analysis of Precipitation (CMAP) datasets, respectively, we develop two types of Bayesian joint probability models: one to model the relationship between NMME-forecast temperature (precipitation) and observed temperature (precipitation) at each grid point, and another to model the relationship between NMME-forecast Niño3.4 anomalies and observed temperature or precipitation at each grid point. We refer to the former as a calibration model and the latter as a bridging model. Bivariate normal distributions are at the core of a Bayesian joint probability model. Prior to Bayesian joint probability model development, data transformations are applied to the temperature and precipitation data to ensure that they conform to a bivariate normal distribution. Markov Chain Monte Carlo sampling is employed to obtain a numerical sample of the posterior distribution of the bivariate normal parameters (mean and covariance). For each calibration (bridging) model, we sample 1000 bivariate normal parameters from which we derive an ensemble of 1000 calibrated (bridged) forecasts.



**Fig. 2** Brier skill scores from fully merged CBaM post-processed NMME probabilistic forecasts of below normal temperature (2a–d) and precipitation (2e–h) over the NMME hindcast period.



**Fig. 3** Difference maps depicting the difference between bridged and calibrated Brier skill scores for probabilistic forecasts of below normal temperature (3a–d) and precipitation (3e–h) over the NMME hindcast period. Warm shading represents areas for which bridging yielded higher Brier skill scores than calibration.



**Fig. 4** Shading indicates Brier skill scores associated with 1-month lead merged CBA forecasts of below normal DJF 2-m temperature for each of the NMME member models and the multi-model mean. Hatching denotes grid cells for which bridging statistically significantly improves forecast skill, where statistical significance is determined using a bootstrap method.

We develop separate bridging and calibration models for each NMME member model of grid point, initial time, and lead; and apply leave-one-year-out cross validation to test the method over the hindcast period. The final step in the CBA method is to merge the bridged and calibrated forecasts using Bayesian model averaging. The resulting merged forecast represents the weighted mean of the calibrated and bridged forecasts.

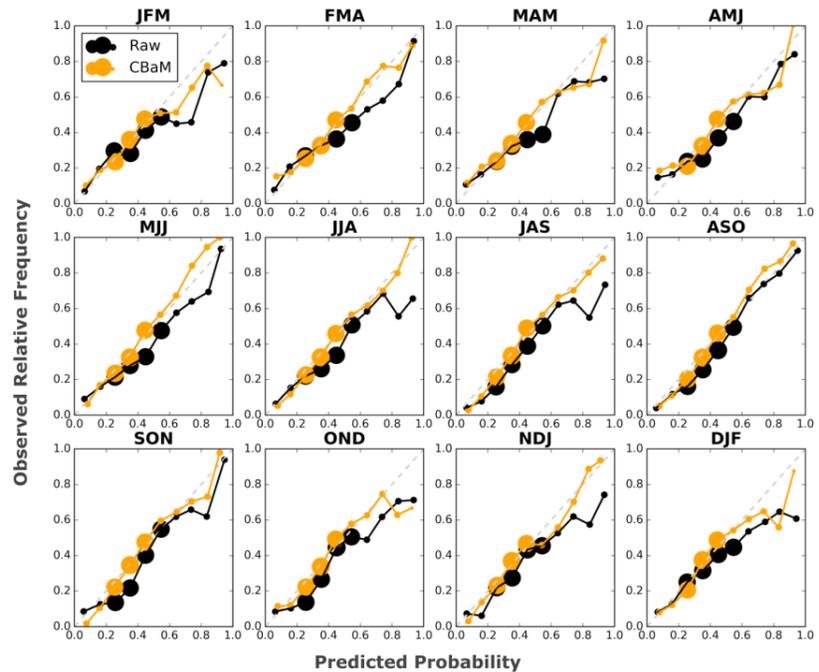
Finally, we assess probabilistic forecast skill using Brier skill scores and reliability diagrams. For brevity, we only present results for probabilistic forecasts of below normal temperature and precipitation.

### 3. Results

As expected, we find that the fully merged temperature forecasts yield higher Brier skill scores than the fully merged precipitation forecasts (Fig. 2). We additionally find that the calibrated forecasts on average perform better than the bridged forecasts, particularly for temperature (Fig. 3). However, bridging improves DJF temperature forecasts for some regions and seasons. In particular, bridging marginally increases skill over parts of the Midwestern and Northern United States – approximately the same regions where the NMME misrepresents the ENSO-temperature teleconnection pattern.

We apply a bootstrap significance test to determine whether bridging significantly enhances 1-month lead forecast skill of DJF 2-m temperature beyond what skill is achieved through calibration (Fig. 4). Bridging statistically significantly ( $\alpha=5\%$ ) improves 1-month lead forecast skill over calibration for approximately 10% of North American CMC2 and NASA grids. The degree to which bridging enhances skill varies by NMME member model, season, and lead (not shown). Improvement from bridging is largely confined to winter months, although we find that bridging significantly enhances forecast skill for a small subset ( $\sim 5\%$ ) of grids for CMC1 forecasts of JJA temperature. Overall, these results suggest that bridging is likely a more useful tool for post-processing individual model ensembles rather than multi-model ensembles. Bridging improves forecast skill relative to calibration for fewer than 3% of NMME grids over the hindcast period. Finally, we find that the fully merged CBA forecasts are more statistically reliable than the raw NMME forecasts (Fig. 5), particularly for higher predicted probabilities. Both calibration and bridging yield statistically reliable forecasts, although calibration tends to achieve better reliability in the higher predicted probability bins (not shown).

While these results suggest that the CBaM tool may improve real-time NMME forecast skill, some caveats should be noted. As noted previously, we develop bridging models using a relatively short, 29-year data set. Given this, a small number of extreme ENSO events most strongly influences the statistical relationships. Additionally, we develop separate bridging and calibration models at each grid point and therefore do not take into account spatial correlations among grid points. Future iterations of the tool should account for such spatial autocorrelation, possibly using hierarchical Bayesian modeling. Finally, it remains to be determined how well BJP calibration and bridging models developed from hindcast data will perform using a truly statistically independent real-time forecast. We are currently testing the method to post-process real-time NMME seasonal forecasts of temperature and precipitation to determine its utility as an operational forecast tool. The experimental probabilistic real-time CBaM forecasts can be found at <http://www.cpc.ncep.noaa.gov/products/people/dcollins/cbam/>.



**Fig. 5** Reliability diagrams comparing the statistical reliability of 1-month lead raw (black) versus CBaM (orange) forecasts of 2-m temperature. The CBaM results refer to the fully merged multi-model (NMME) forecast. The raw forecast refers to the multi-model mean forecast, without any bias-correction applied. The horizontal axis denotes the predicted probability while the vertical axis denotes the observed relative frequency. The light gray dashed line corresponds to a perfectly reliable forecast. The size of the plotted circles is proportional to the number of forecast probabilities that fall into a given predicted probability bin.

## References

- Chen, L. C., H. van den Dool, E. Becker, and Q. Zhang, 2017: ENSO precipitation and temperature forecasts in the North American Multimodel Ensemble: Composite analysis and validation. *J. Climate*, **30**, 1103-1125.
- Kirtman, B. P., and Co-authors, 2014: The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, **95**, 585-601. doi: <http://dx.doi.org/10.1175/BAMS-D-12-00050.1>
- Schepen, A., Q. J. Wang, and D. E. Robertson, 2014: Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. *Mon. Wea. Rev.*, **142**, 1758-1770.
- Schepen, A., Q. J. Wang, and Y. Everingham, 2016: Calibration, bridging, and merging to improve GCM seasonal temperature forecasts in Australia. *Mon. Wea. Rev.*, **144**, 2421-2441.
- Wang, Q. J., D. E. Robertson, and F. H. S. Chiew (2009), A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resour. Res.*, **45**, W05407, doi:10.1029/2008WR007355.