Can We improve the CPC's long-lead El Niño forecasts?

Pao-Shin Chu, Hanpei Zhang, David Unger* and Luke He*
Department of Atmospheric Sciences
University of Hawaii

*Climate Prediction Center, NCEP, NOAA

Presented at the CWB, March 13, 2018

The Climate Prediction Center (CPC) of NOAA/NCEP is responsible for issuing monthly, subseasonal and seasonal prediction for the US and also for El Niño and others

- For systematically forecast ENSO, the effort began in 1982 when a very strong El Niño developed.

- It was not known that this event had already developed until several months later because of poor monitoring network back then.

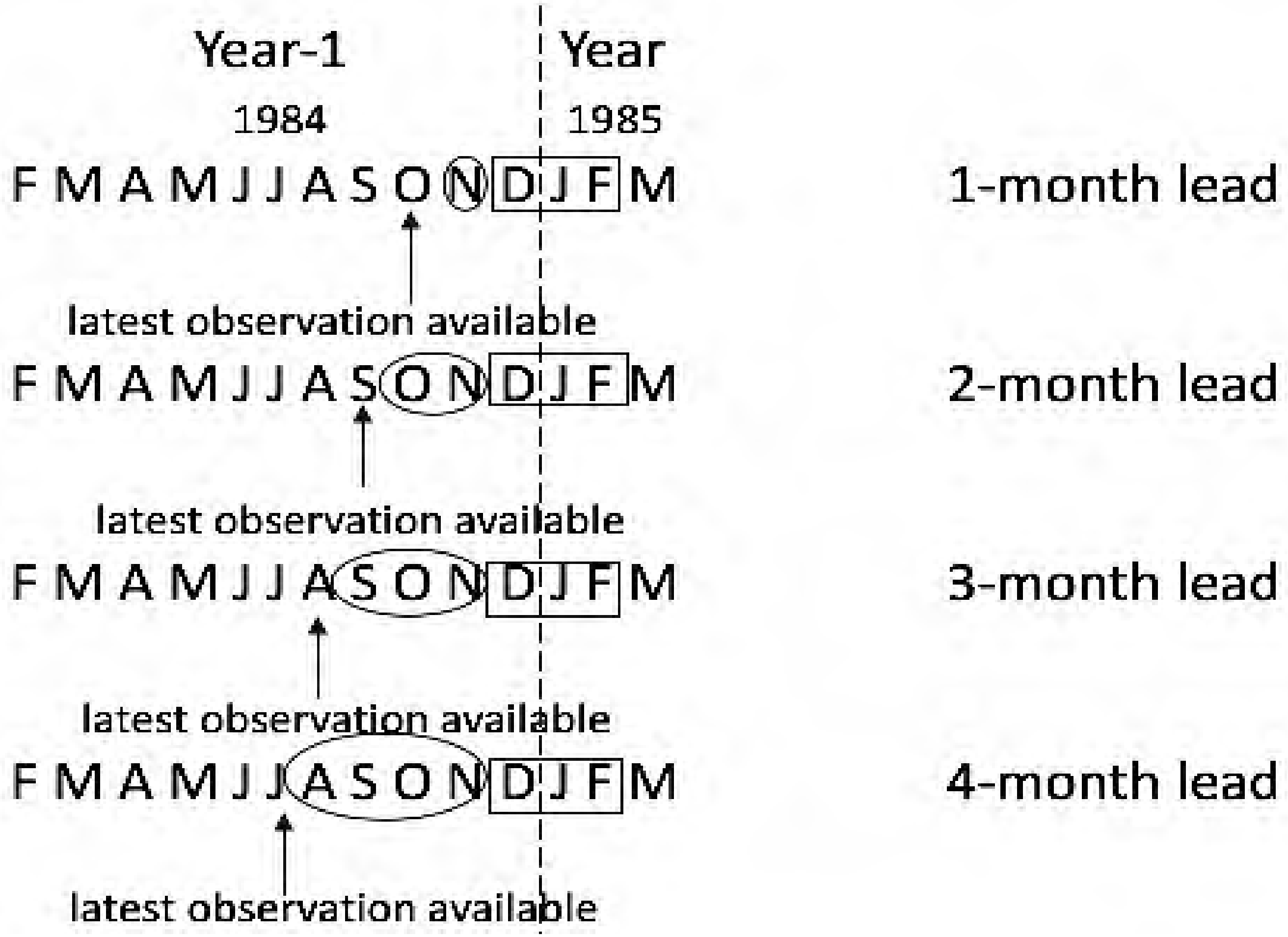
CPC Nino 3.4 (5N-5S, 90W-150W) SSTs

- Historical hindcast (retrospective or re-forecast) data for Nino 3.4 SSTs from three statistical models (CA, CCA, Markov) and one dynamical model (CFSv2) from lead 1 to 7 months; data run from 1982 to 2016.

- Received CFSv2 forecasts from Unger last October but they were uncorrected.

- Cross Validation: repeatedly omit a few observations from the data, reconstruct the model, and then estimate the omitted cases. Use Leave-Three-Out Cross Validation (LTOCV).

Lead times


Year-1
Year
1984
1985
F M A M J J A S O N D J F M
1-month lead
latest observation available
F M A M J J A S O N D J F M
2-month lead
latest observation available
F M A M J J A S O N D J F M
3-month lead
latest observation available
F M A M J J A S O N D J F M
4-month lead
latest observation available

Statistical models

1. CA: Constructed analog (CA) is a statistical forecast that is a linear combination of past observed anomaly patterns in the predictor fields such that the combination is as close as desired to the current state (Van den Dool 1994).

2. CCA: Canonical Correlation Analysis (CCA) is an empirical statistical method that finds patterns of predictors and predictand that maximize the correlation between them (Barnston et al. 1994, Chu and He, 1994, He and Barnston 1996).

3. MKV: Markov Model (MKV) is built in a reduced multivariate empirical orthogonal function (MEOF) space of the sea surface temperature anomaly, sea level and wind stress anomaly fields (Xue et al. 2000).

BMA

- Bayesian model averaging method is a statistical method for postprocessing the ensembles and producing probabilistic forecasts from ensembles.

- Why BMA? BMA is different from other model averaging methods in that it not only describes the uncertainties associated with each model simulation but also provides the diverse capabilities of different models

BMA method can assess the performance of the individual models and assign weights to models - Model Performance

We aim to use Bayesian Model Averaging (BMA) as a method that the weighted estimate is a better predictor of the observed system behavior than any of the individual models of the ensemble – Better Predictor

Bayesian Model Averaging (BMA)

Some examples:

1. Applications of Bayesian model averaging in the reconstruction of past climate change using PMIP3/CMIP5 multimodel ensemble simulation. Fang and Li, J. Climate 2016

2. Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. Schepen et al., J. Climate, 2014

3. Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. Tebaldi et al., J. Climate, 2005

For a suite of climate models k=1, 2, …, K, the joint PDF of y conditional on y_k is given by a weighted average of the individual model density as (Raftery et al. 2005)

$$f(y \mid y_k, k = 1, \ldots, K) = \sum_{k=1}^{K} w_k \cdot f_k(y \mid y_k) \qquad (1)$$

where y is the observed SST, y_k the hindcast SST from model k, w_k the BMA weight for model k and is a nonnegative value that satisfies $\sum_{k=1}^{K} w_k = 1$. The weight reflects how well model k fits the observation. The conditional density for individual model k is assumed to be Gaussian distributed.

Bayes' theorem

θ : parameter

Classical statistics: θ a constant
Bayesian inference: θ a random quantity, P(θ|y)

y: data

P(y|θ): likelihood function
π(θ): prior probability distribution

$$P(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} P(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

Posterior distribution characterizes the current best information regarding uncertainty about θ

The prior distribution of the weights is specified by a symmetric Dirichlet as (Schepen et al. 2014)

$$\mathrm{p}(w_k, k=1,\ldots,K) \propto \sum_{k=1}^{K} (w_k)^{\alpha-1} \qquad (2)$$

where α is known as the concentration parameter and equals to $1.0 + \frac{\alpha_0}{K}$.

This prior distribution is intended to constrain the effect of sampling errors due to limited sample size and thus stabilizes the weights.

By invoking equation (1) and assuming there are T independent events, the

posterior distribution of the weights for each time step t = 1, …, T, is thus

$$\mathrm{p}(\mathrm{w}_{\mathrm{k}}, k = 1,\ldots,K \mid y^{t}, y_{k}^{\;t}, t = 1,\ldots,T, k = 1,\ldots,K) \tag{3a}$$

$$\propto p\left(w_k, k = 1,\ldots,K\right) \cdot \prod_{t=1}^{T} \sum_{k=1}^{K} w_k \cdot f_k(y^t \mid y_k^t) \tag{3b}$$

Thus, the BMA is a finite mixture model. Making use of the symmetric and conjugate Dirichlet prior, equation (3b) becomes

$$A = \prod_{k=1}^{K} (w_k)^{\alpha-1} \prod_{t=1}^{T} \sum_{k=1}^{K} w_k \cdot f_k(y^t \mid y_k^t) \tag{4}$$

Bayesian Model Averaging

Expectation-Maximization (EM) algorithm to derive the maximum likelihood estimation for model parameters

The maximum-likelihood estimators are the most probable values for the parameters, given the observed data. In other words, it is the value of the parameter that makes the observed data most likely to have been observed.

Expectation-Maximization (EM) algorithm

$$o_k^{t,(j+1)} = \frac{w_k^{(j)} \cdot f_k(y^t \mid y_k^t)}{\sum_{k=1}^{K} w_k^{(j)} \cdot f_k(y^t \mid y_k^t)} \longrightarrow$$

E step (expectation) calculates membership probability ($o_k^{t,(j+1)}$) for the time t and model k

$$w_k^{(j+1)} = \frac{\frac{1}{T}\sum_{t=1}^{T} o_k^{t,(j+1)} + (\alpha - 1)/T}{1 + k(\alpha - 1)/T} \longrightarrow$$

M step (maximization) calculates new weights ($w_k^{t,(j+1)}$)

The BMA weights are then estimated iteratively with the above two equations until the algorithm reaches a convergence of the logarithm of term *A* (posterior distribution of the weights)

- It starts with an initial guess for the weights. In this study, a uniform distribution is given to all weights in iteration "0."

- In the E-step, the membership probability is estimated using the current guess for the parameters (w_k) and the forecast PDF conditional on model k. In the M-step, the weights are estimated given the current value of the membership probability, the concentration parameter, sample size (T), and the number of models used (K).

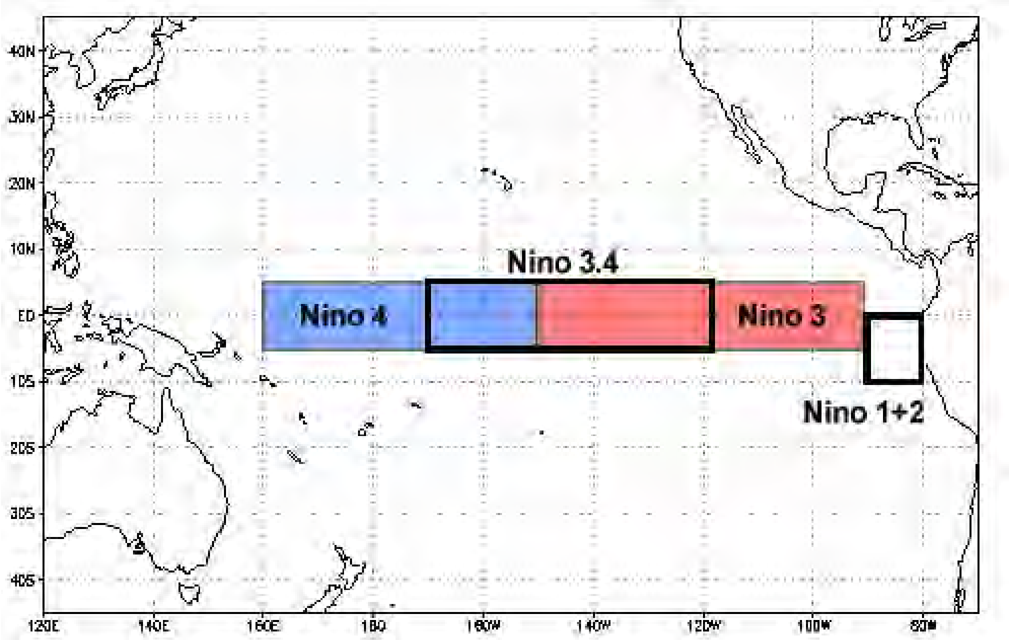
- The BMA weights are then estimated iteratively with Equations (5) and (6) until the algorithm reaches a convergence of the logarithm of term *A* in Equation (4). The EM algorithm alternates between an expectation and a maximization step. The convergence is defined as the change of the logarithm of posterior "*A*" between two consecutive iterations is no greater than a predefined small tolerance.

1. Historical SST hindcasts for Niño 3.4 (5°N-5°S) (170°W-120°W)

2. Statistical models : CA, CCA, MKV, CFSv2 (CPC)

3. 3 months running average from 1982 to 2016 (DJF JFM FMA MAM AMJ MJJ JJA JAS ASO SON OND NDJ)

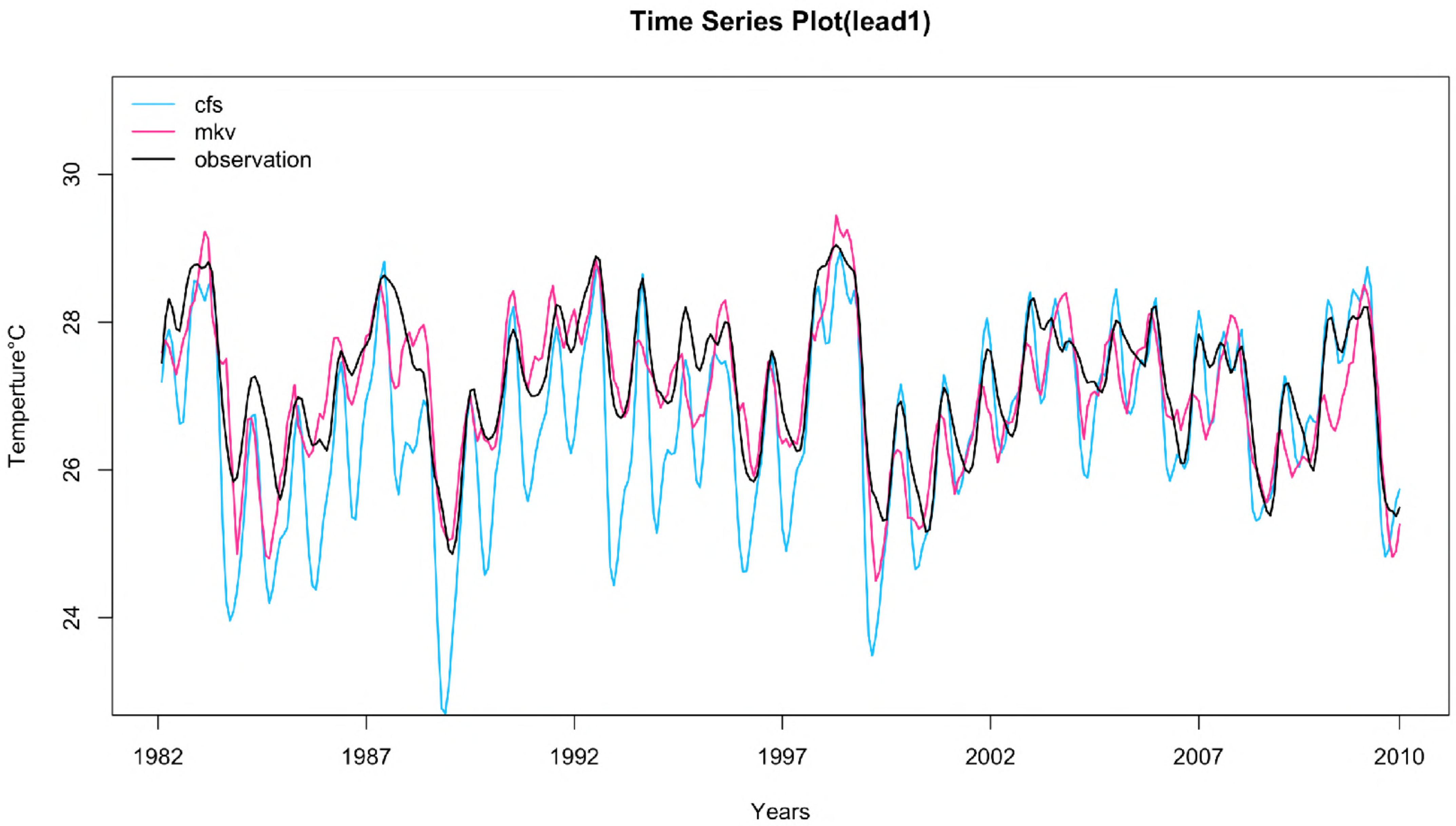
4. Lead time from 1-month to 7-month



- CFSv1 was the coupled atmosphere-ocean-land model for seasonal prediction and implemented into operations at NCEP in August 2004

- CFSv2 was implemented in March 2011 and made retrospective forecasts from 1982 to 2010 and onward for real time subseasonal and seasonal predictions

- Some issues with CFSv2 data



Advanced Microwave Sounding Unit satellite obs in 1999

The cold bias also occurred in other equatorial oceans

David Unger sent me the calibrated CFv2 reforecasts in February 2018

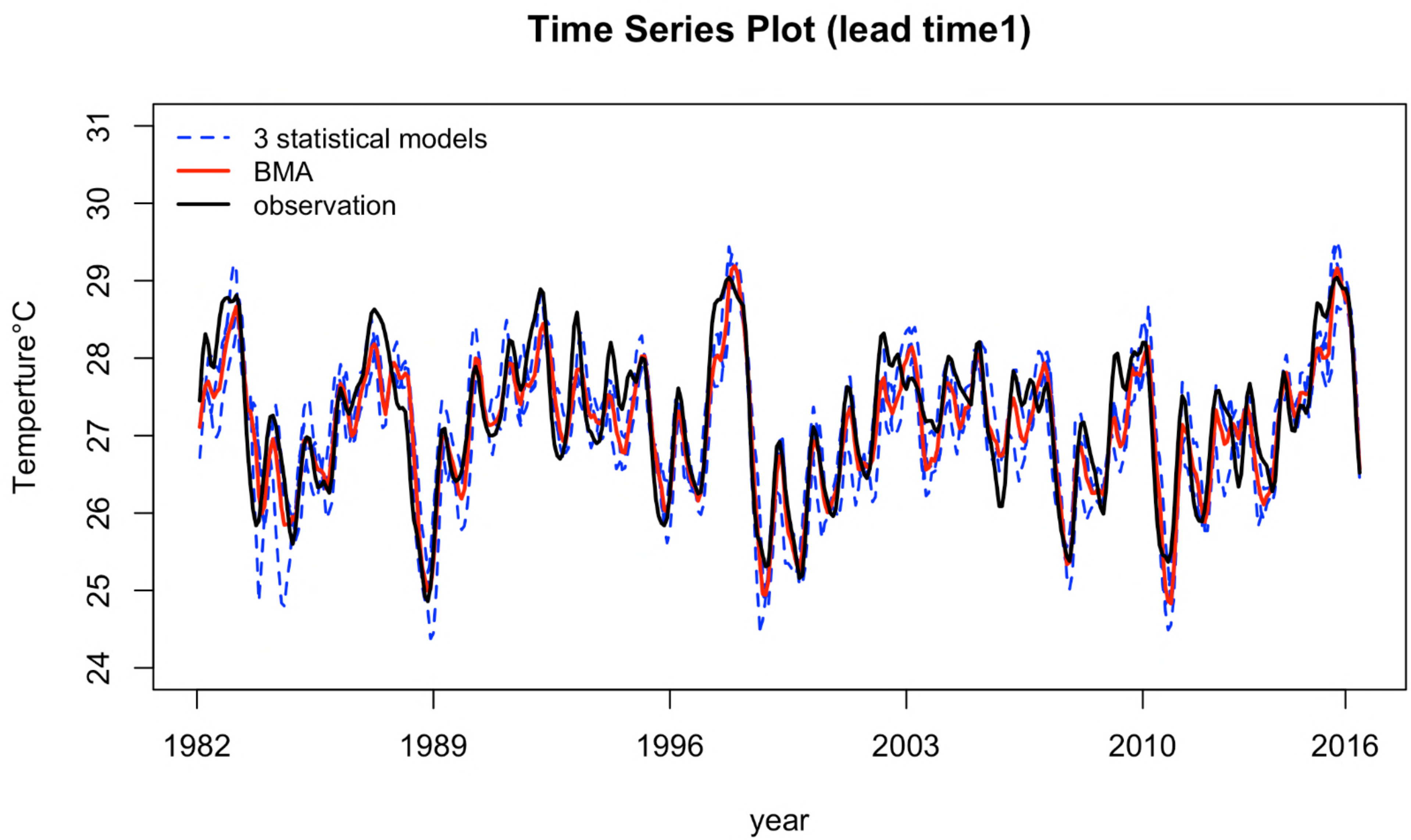
- Let mean forecast of 1982-1998 be F98 and mean observation during the same period be O98.

- Let mean forecast during 1999-2010 be F10 and mean observation during the same period be O10.

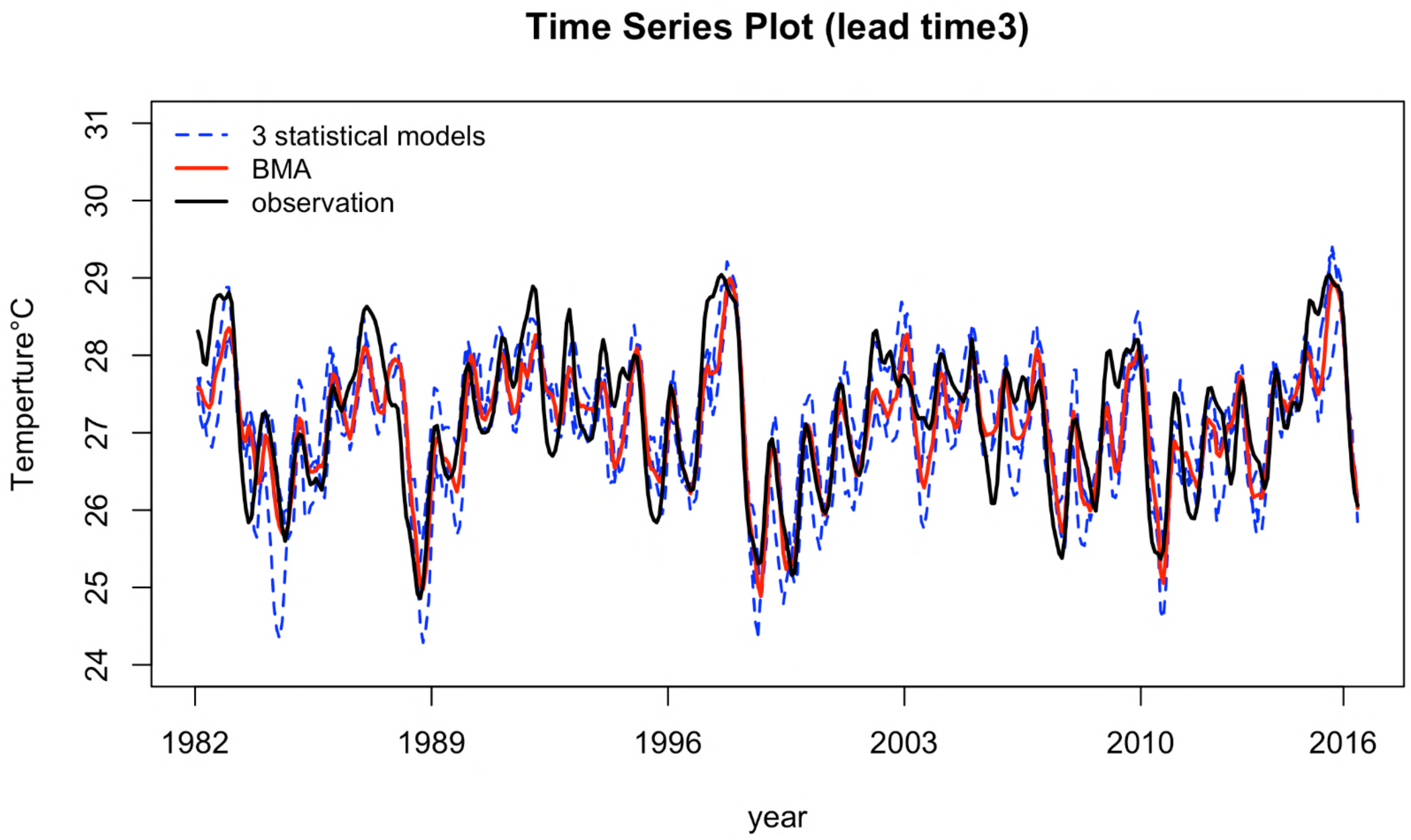
- Then correction factor = (F98-O98) – (F10-O10)

- The bias corrected forecast prior to 1999 = Forecast – correction factor

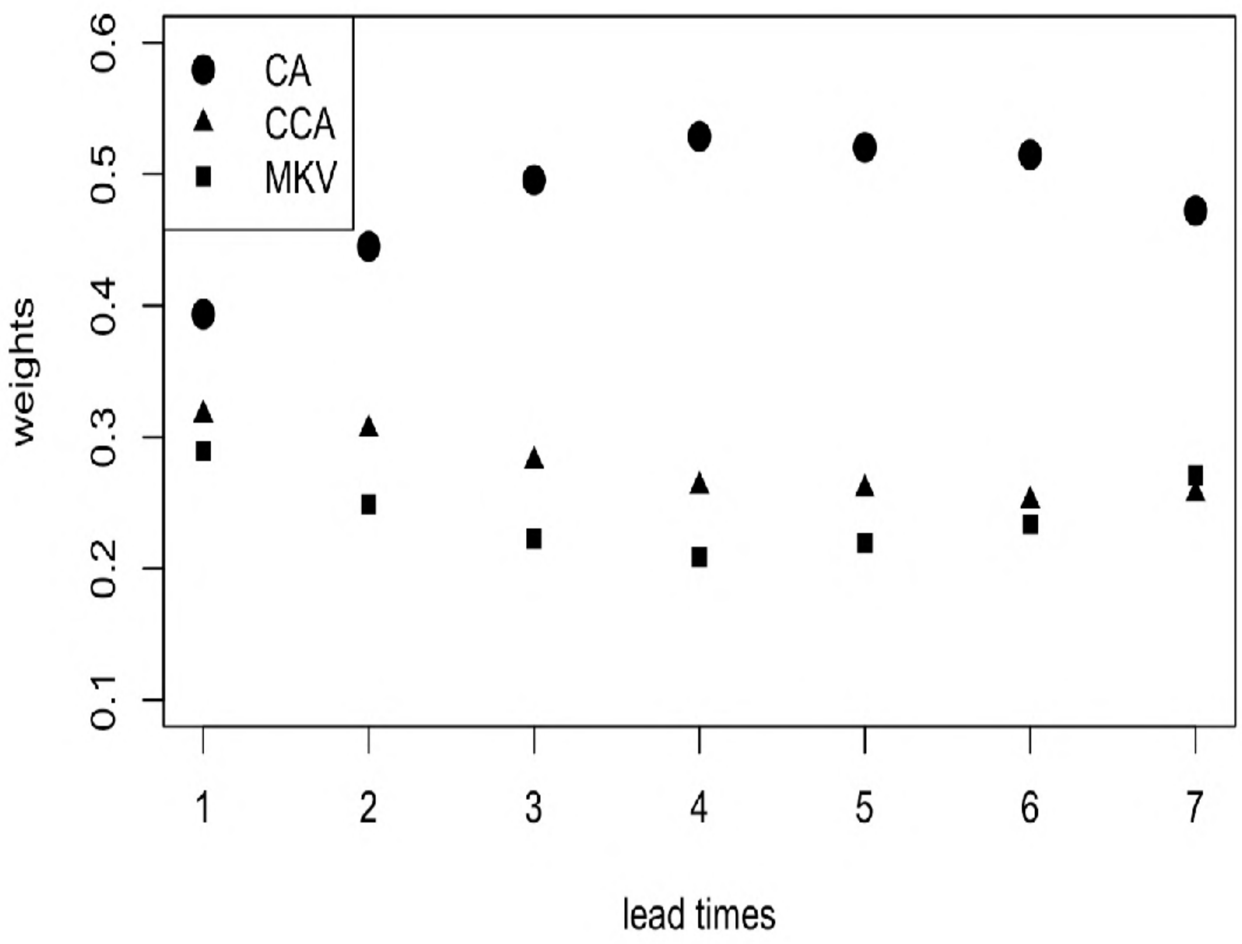
  (mean forecast bias prior to 1999 = mean forecast bias after 1998)
- Also applied a 1-2-1 smoother across lead time to prevent abrupt bias correction changes
- We are currently working on the bias corrected CFSv2 forecasts



Time Series Plot (lead time1)
3 statistical models
BMA
observation
Temperture°C
31
30
29
28
27
26
25
24
1982
1989
1996
2003
2010
2016
year


Time Series Plot (lead time3)
3 statistical models
BMA
observation
Temperture°C
31
30
29
28
27
26
25
24
1982
1989
1996
2003
2010
2016
year

BMA weights for each of three models versus lead times



Skill Score

Forecast skill is usually presented as a skill score, which can be interpreted as a percentage improvement over the reference forecasts.

$$SS_{ref} = \frac{A - A_{ref}}{A_{pref} - A_{ref}} \times 100\% \qquad \text{(S1)}$$

Equation (S1) can be constructed by using the MAE, MSE or RMSE as the underlying accuracy statistics (A). In our study, MSE values are used as accuracy statistics. The skill score is thus expressed as

$$\mathrm{SS} = \frac{MSE - MSE_{clim}}{0 - MSE_{clim}} = 1 - \frac{MSE}{MSE_{clim}} \qquad \text{(S2)}$$

$$MSE_{clim} = \frac{1}{n}\sum_{k=1}^{n}(\bar{o} - o_k)^2 \qquad \text{(S3)}$$

where the perfect forecasts have MSE=0 so SS=1. MSE_{clim} is the climatological mean square error values of the Nino 3.4 SSTs.

		CA	CCA	MKV	BMAcv
	ss	0.728	0.725	0.662	0.815
lead1	RMSE	0.477	0.479	0.532	0.393
	cc	0.874	0.852	0.853	0.912
	ss	0.686	0.661	0.577	0.765
lead2	RMSE	0.513	0.533	0.595	0.444
	cc	0.850	0.813	0.806	0.884
	ss	0.649	0.588	0.500	0.716
lead3	RMSE	0.542	0.588	0.647	0.488
	cc	0.829	0.767	0.762	0.856
	ss	0.605	0.518	0.441	0.669
lead4	RMSE	0.575	0.635	0.685	0.527
	cc	0.804	0.720	0.726	0.829
	ss	0.552	0.457	0.401	0.625
lead5	RMSE	0.613	0.675	0.710	0.562
	cc	0.775	0.678	0.697	0.803
	ss	0.491	0.403	0.375	0.579
lead6	RMSE	0.655	0.709	0.725	0.595
	cc	0.741	0.641	0.671	0.774
	ss	0.429	0.364	0.356	0.529
lead7	RMSE	0.694	0.732	0.737	0.630
	cc	0.701	0.611	0.648	0.736

A comparison between BMA deterministic forecast and each individual model forecast from lead one to seven. The BMA deterministic forecast was performed by leaving three out cross validation.

For all leads, the BMA has the highest skill score, lowest RMSE, and highest correlation relative to 3 individual models. The CA has a better skill than CCA and MKV. Forecast results and the weights for each model are in agreement. The model with a higher weight is also the model with better forecasts.

CFSv1 CFSv2

3-mo lead

6-mo lead

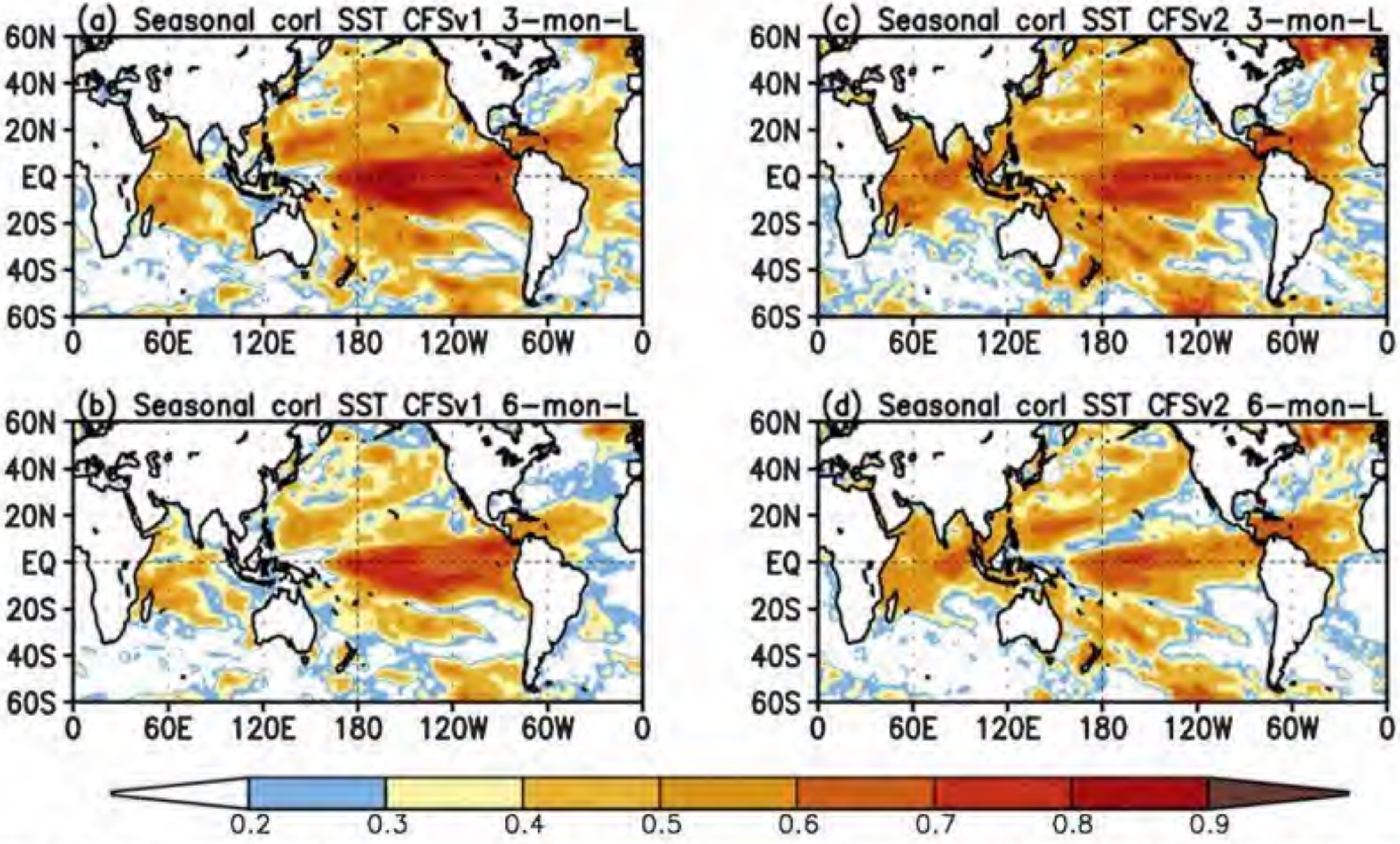
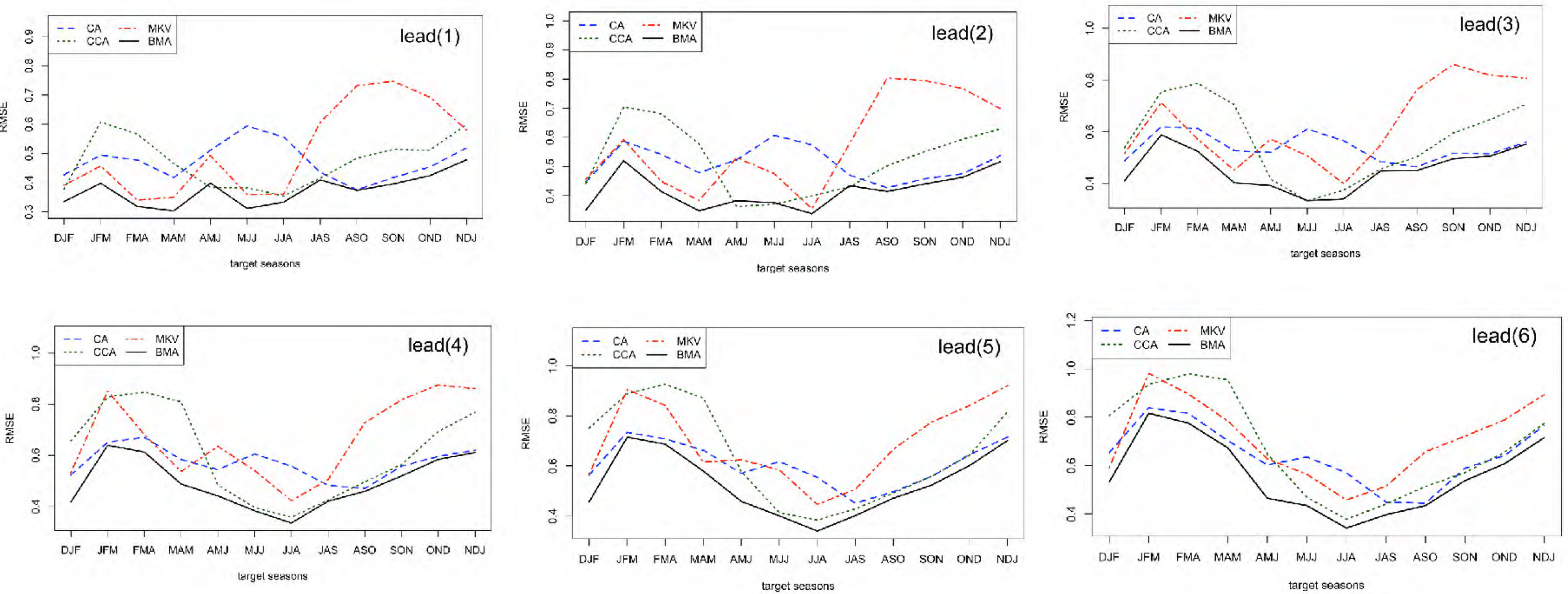


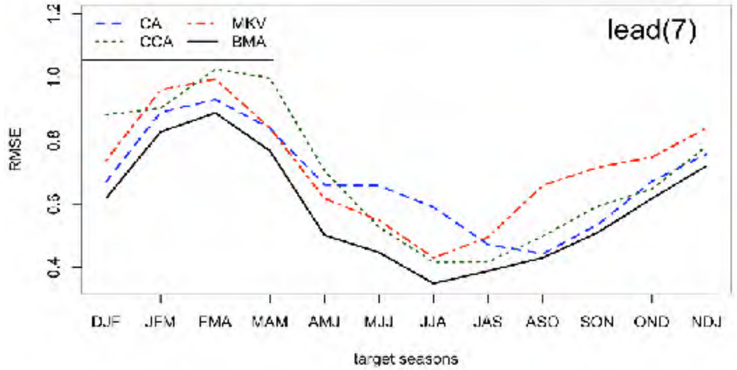
FIG. 2. Anomaly correlation of 3-month-mean SST between model forecasts and observation: (a) 3-month lead CFSv1, (b) 6-month lead CFSv1, (c) 3-month lead CFSv2, and (d) 6-month lead CFSv2. Contours are plotted at an interval of 0.1.

Saha et al., 2014

Anomaly correlation of 3-mo mean SST between forecast and obs.

“In the tropical Pacific, the CFSv2 skill is slightly lower than that of CFSv1…”

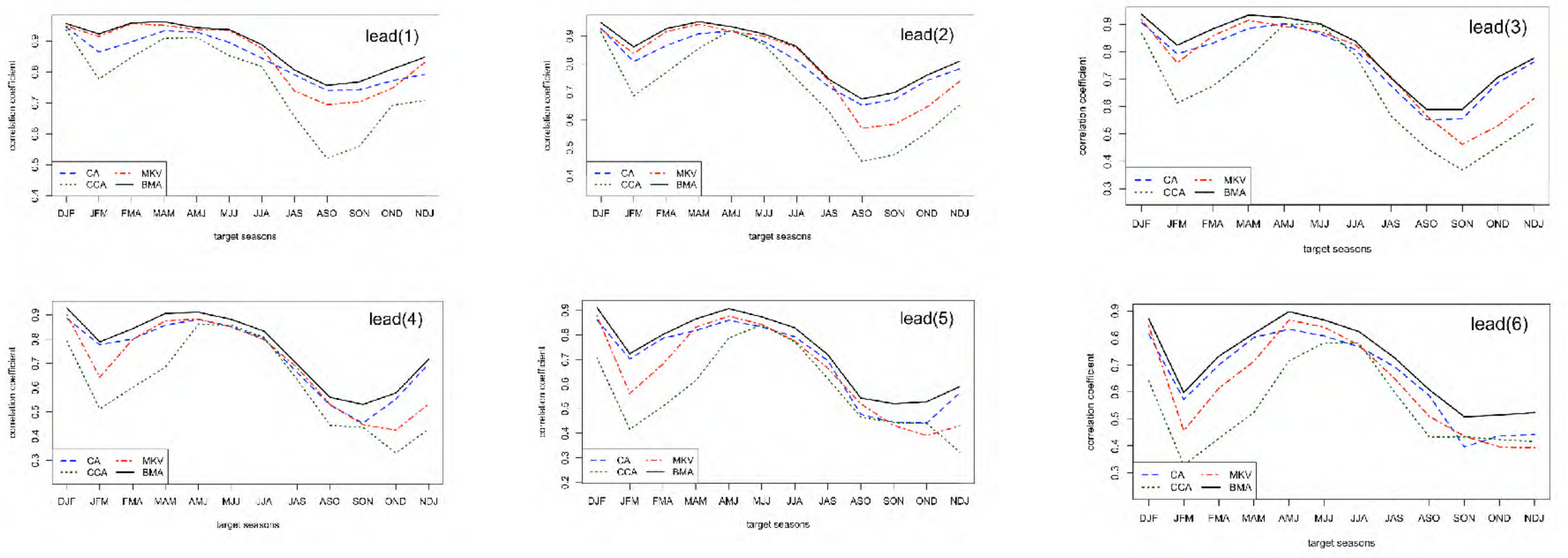


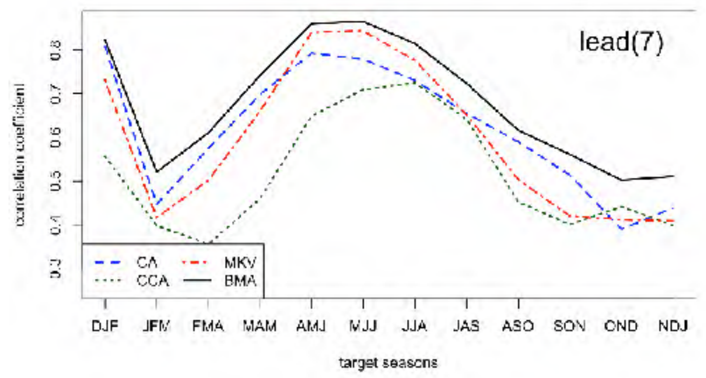


BMA deterministic forecast performance (RMSE) for different target seasons. No cross validation. 35 years from 1982 to 2016.

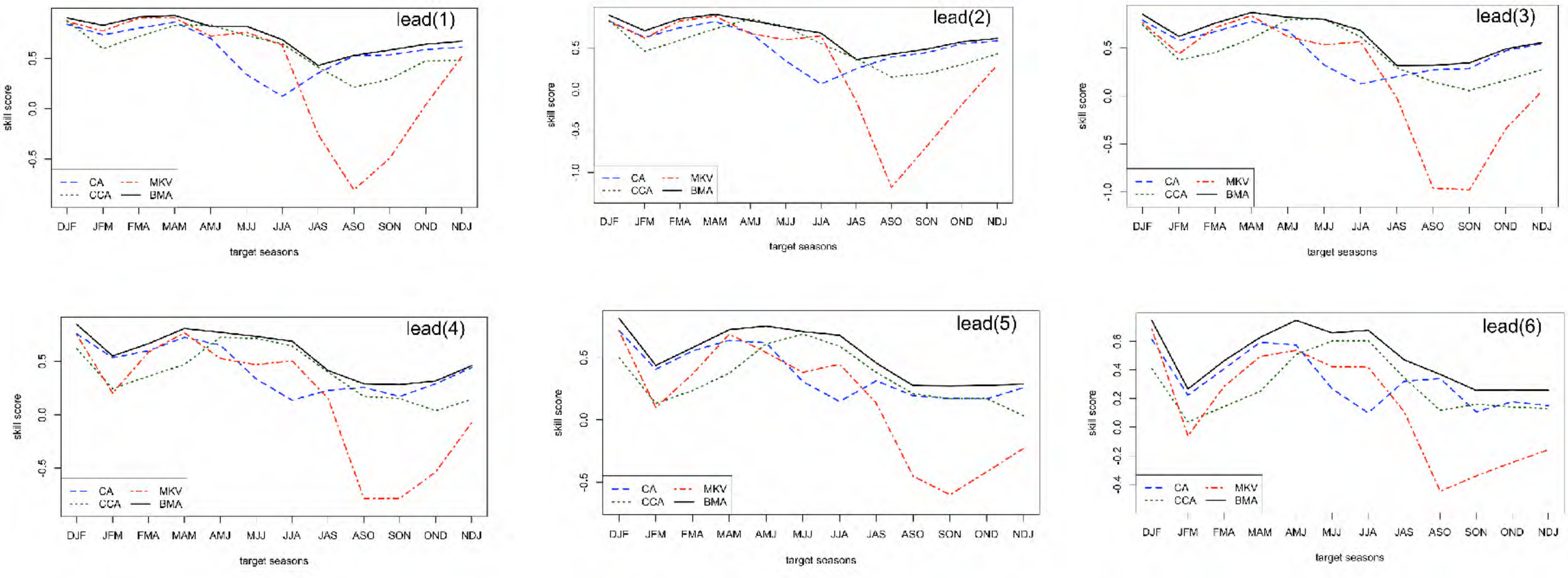
CA is not the best for all target seasons (AMJ,MJJ, JJA); Markov has the worst forecast from JAS to NDJ; CCA is worst from JFM to MAM. For BMA, the RMSE is almost always the lowest for all target seasons.

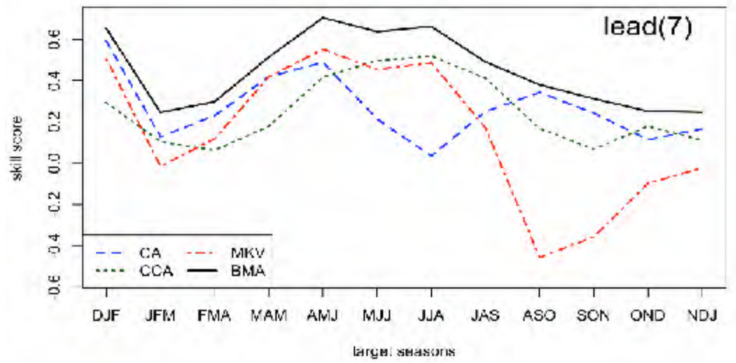
The boreal summer has lower RMSE while other seasons have larger errors.



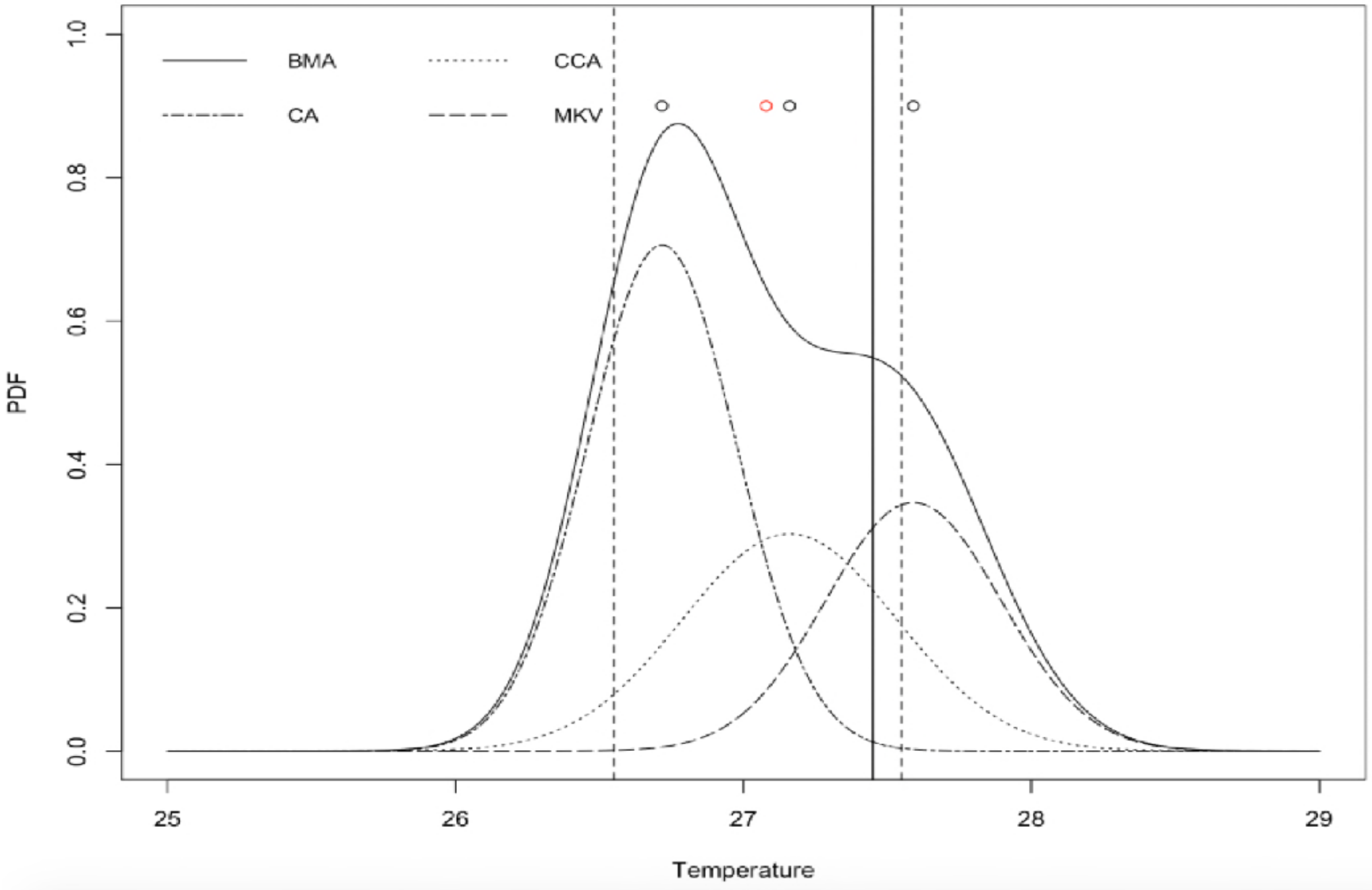


Correlation coefficients of individual model forecasts and BMA deterministic forecasts for lead 1-7. Overall, the CCA has the lowest correlations and BMA has the highest correlation for all leads and all target seasons. From lead 1 to 4, boreal autumn has lowest correlation and spring has the best correlation. At higher leads, JFM also has low correlation while AMJ/MJJ has higher correlations.





Skill score for individual model forecast and BMA deterministic forecasts for different target seasons. The Markov model has the worst sill from JAS to NDJ for all leads. The BMA performance is still robust and best for all leads.



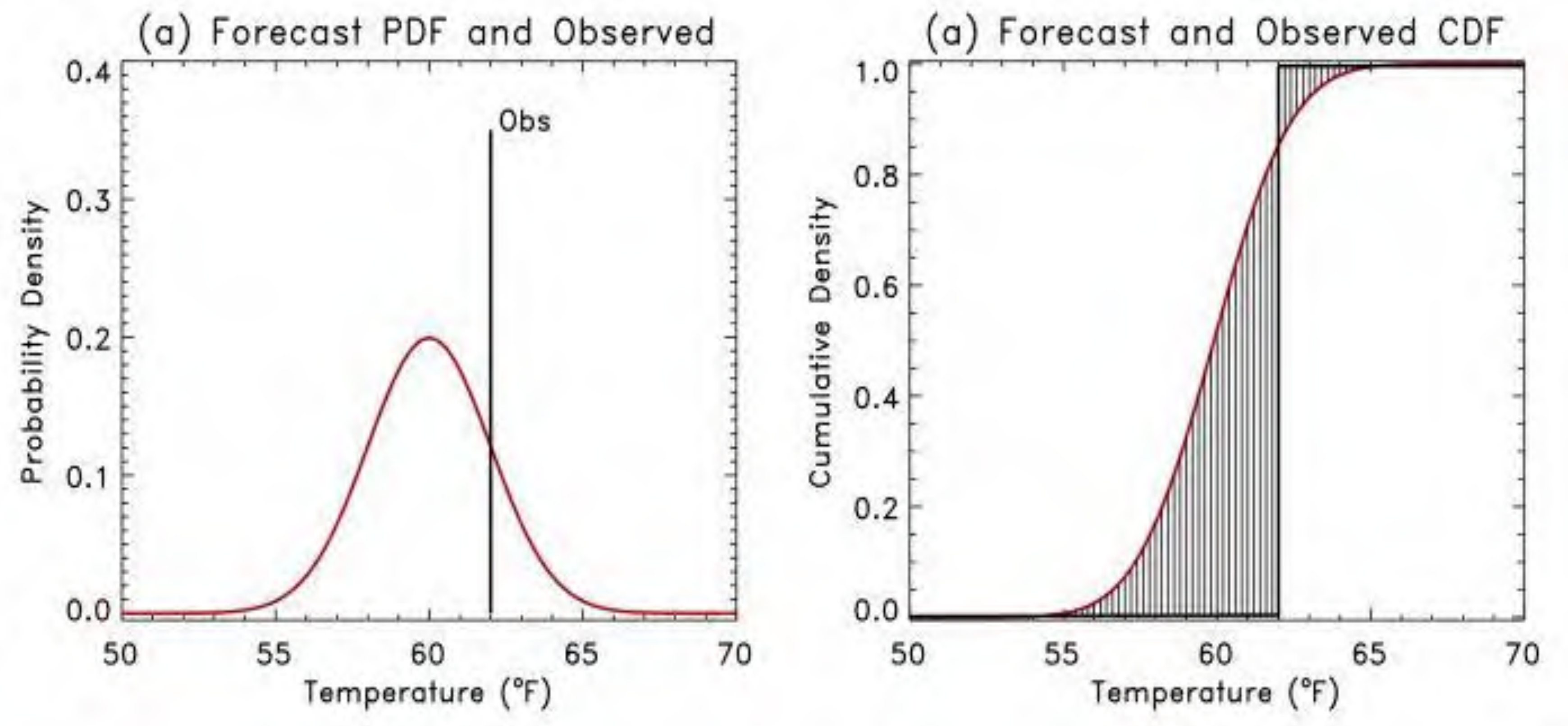
The BMA deterministic forecast is 27.2C and observation is 27.4C.

An example of the BMA probability forecast (solid pdf), individual model probability forecasts (CA, CCA, MKV), deterministic forecast from each model (solid circle) and BMA deterministic forecast (red circle) for one month lead forecast of FMA 1982. The vertical solid line is the observation. BMA probability forecast is a weighted average of each model PDF.   Dashed vertical lines are BMA 90% prediction interval.

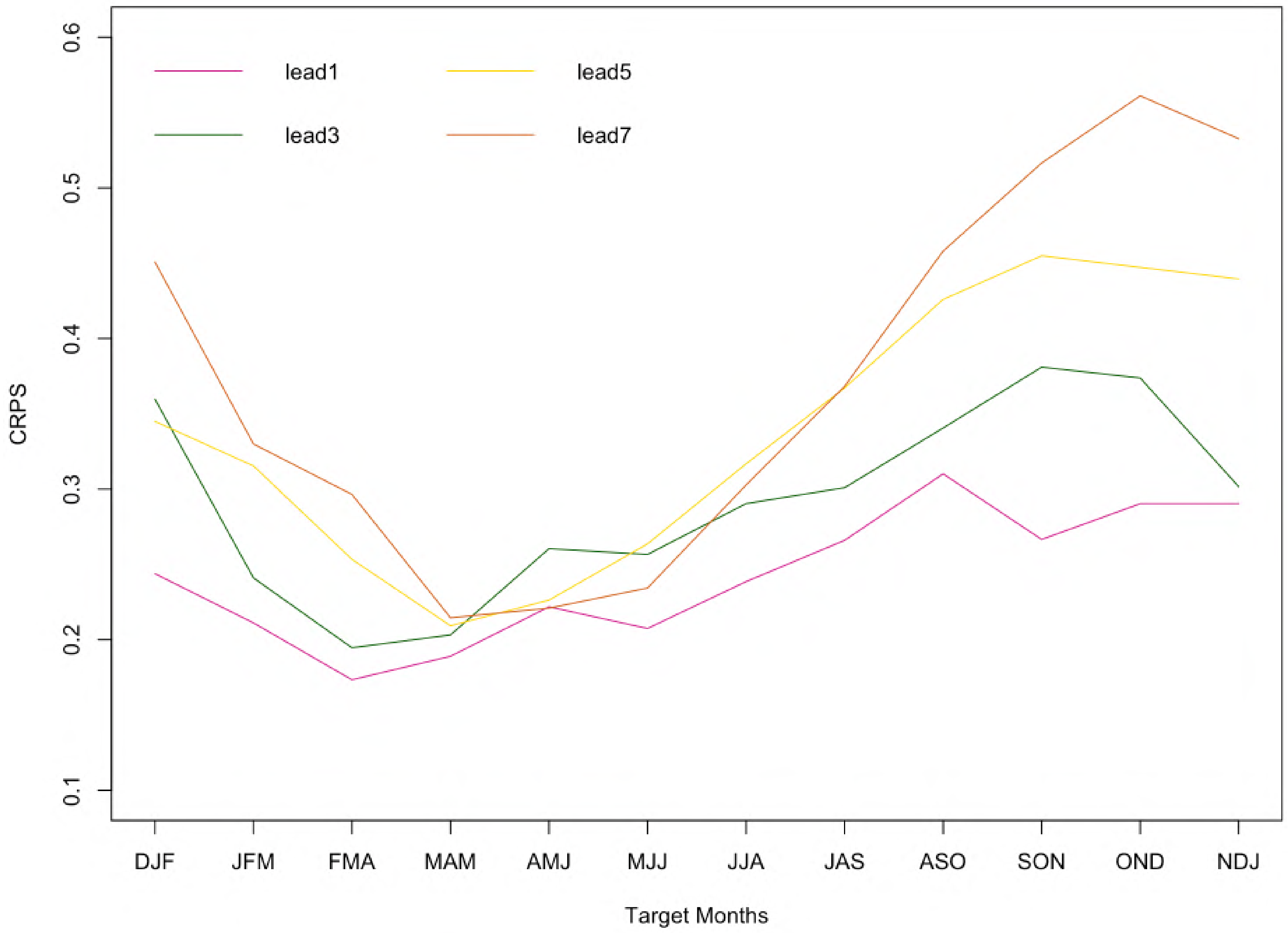
Continuous ranked probability score (CRPS)

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_0\,(y)]^{\,2} dy$$

$$F_0(y) = \begin{cases} 0, y < observed\ value \\ 1, y \geq observed\ value \end{cases}$$



The computation of the CRPS involves the total area between the CDF of the forecast and CDF of the observation. The CRPS has a negative orientation so the smaller values the better are the forecast.



BMA CRPS results of individual target seasons. The BMA CRPS results were done by separating the data into different target seasons and use leave three out cross validation method to apply BMA probability forecast to them. Each line represents the changes of CRPS of target seasons in different lead times

Summary and Future Direction

- Based on CPC’s 7-month lead-time seasonal SST forecasts for the Nino 3.4 region, BMA forecast outcomes are shown to have lower RMSE, higher skill score, and higher correlation coefficients for all leads than three statistical operational models. Will include calibrated retrospective forecasts from CFSv2 ensembles (24 members) in the future.

- BMA assigns weights to each model and can assess model performance. In general, CA has the better forecast skill relative to CCA and Markov models.

- How about subseasonal prediction and seasonal rainfall forecasts?

Thank you!