A Simple Statistical Model of the Madden Julian Oscillation using CFS and Lagged Linear Regression

Michelle L. L'Heureux, Jon C. Gottschalck, and R. Wayne Higgins
NOAA/NWS/NCEP/Climate Prediction Center

Background:

Wheeler and Hendon (*MWR, 2004*) have created an index to track the Madden Julian Oscillation (MJO), which is represented by the first two Principal Components, **RMM1 and RMM2**, that describe the out-of-phase eastward propagating wave.

RMM1 and RMM2 can be calculated in real-time and therefore is a valuable monitoring indicator that shows the current location and amplitude of the propagating MJO wave.

Wheeler and Hendon have also created a 2-predictor **Lagged Linear Regression (LLR)** statistical model to predict the future state of RMM1 and RMM2.

NCEP **Climate Forecast System (CFS)** is a fully coupled ocean-land-atmosphere dynamical model, which became operational in August 2004.

Objective:

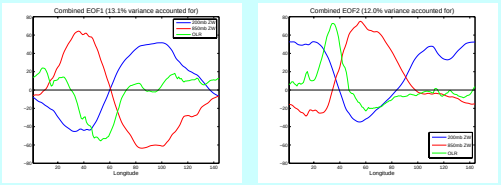
To create a 4-predictor multivariate linear equation that predicts RMM1 and RMM2 by incorporating the independent skill provided by a statistical (LLR: lagged linear regression) model and by a dynamical (CFS: Climate Forecast System) model.

Some Key Questions:

- *Can a 4-predictor model, leveraging the skill provided by LLR and CFS, improve the prediction of the MJO?*
- *Does the CFS add useful, independent skill to a simple multivariate predictive equation of the MJO?*

Data and Methods:

The MJO index (1980-2006) is calculated using the two leading structures from a combined Empirical Orthogonal Function (EOF) of 850mb and 200mb zonal wind and OLR (averaged from 15°S-15°N). Before NCEP/NCAR Reanalysis and CFS data is used in the EOF, the first four harmonics (seasonal cycle) are removed, ENSO is linearly removed, and the most recent 120-day mean is removed. Here are the resulting EOFs using Reanalysis data:



The NCEP CFS hindcast data is extracted from Dec. 1980 - Jan. 2006 for a total of 4500 members (15mem/mth * 12 mth * 25 yrs) out to a forecast period of 45 days. The data is processed as described above and is projected onto the two leading EOFs to obtain the leading Principal Components RMM1 and RMM2.

The 4-predictor multivariate model used to predict RMM1 or RMM2 is created using these four predictors: past observed (Reanalysis) values of RMM1 and RMM2 at the initial time of the forecast and also forecasted (CFS hindcast) values of RMM1 and RMM2.
[see equation in the top center panel]

Weights/coefficients in the 2-predictor vs. 4-predictor Multivariate Linear Model using CFS and LLR

2-predictor equation

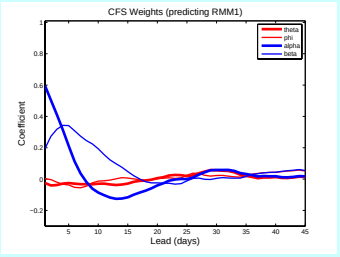
$$RMM1_{obs} = \alpha RMM1_{\substack{obs \\ or \\ cfs}} + \beta RMM2_{\substack{obs \\ or \\ cfs}}$$

obs *(observations): Reanalysis data used in LLR*

4-predictor equation

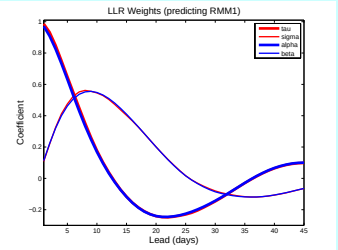
$$RMM1_{obs} = \tau RMM1_{obs} + \sigma RMM2_{obs} + \theta RMM1_{cfs} + \phi RMM2_{cfs}$$

(1) Weights for CFS (2-predictor vs. 4 predictor)



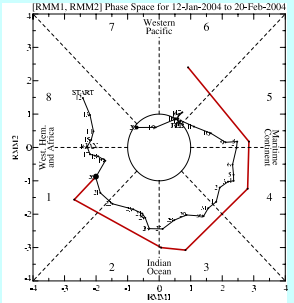
- Weights for the CFS become **near zero** when the statistical LLR model is added into the predictive equation

(2) Weights for LLR (2-predictor vs. 4 predictor)



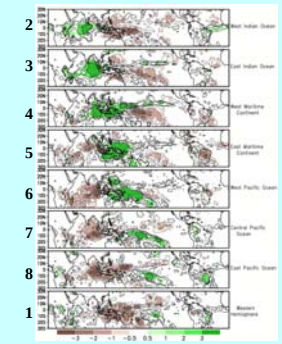
- Weights for the statistical LLR model change very little when the CFS is added into the predictive equation

Sample Results from 4-predictor Model (LLR + CFS)



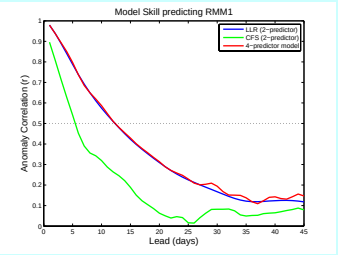
Red line: 4-predictor model (pentads)
Black Line: Observations (daily)

Nov-Mar precipitation composite based on the location (phase) of the MJO



Forecast Skill of RMM1 and RMM2

Prediction of RMM1

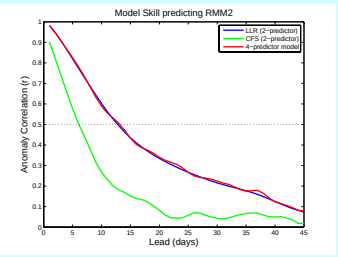


Red: Prediction of RMM1 using 4-predictor model

Blue: Prediction of RMM1 using 2-predictor LLR

Green: Prediction of RMM1 using 2-predictor CFS

Prediction of RMM2



Red: Prediction of RMM2 using 4-predictor model

Blue: Prediction of RMM2 using 2-predictor LLR

Green: Prediction of RMM2 using 2-predictor CFS

In both cases,

- Skill of the statistical LLR 2-predictor model (blue) exceeds the CFS 2-predictor model (green)
- Skill for the 4-predictor model (red) only barely exceeds the 2-predictor LLR model (blue)
- CFS is adding no useful, independent information to the 4-predictor model

Summary and Conclusions

- While the NCEP Climate Forecasting System (CFS) model partially captures the MJO signal, it does not offer useful, independent predictive skill beyond the skill obtained by a simple lagged linear regression (LLR).
- We are still waiting for a dynamical model that predicts the MJO in a way that is comparable to the CFS in predicting ENSO.
- Multiple linear regression is a powerful tool that consolidates several different techniques/models into one predictive equation that takes advantage of the independent skill of each model. While the 4-predictor model (CFS + LLR) did not provide useful skill in this case, this technique is the basis for improving and consolidating existing CPC MJO forecast tools. **ALSO SEE Qin Zhang's poster and Jon Gottschalck's talk.**