Exploring the Effect of Madden-Julian Oscillation on South China Precipitation with Interpretable Neural Networks

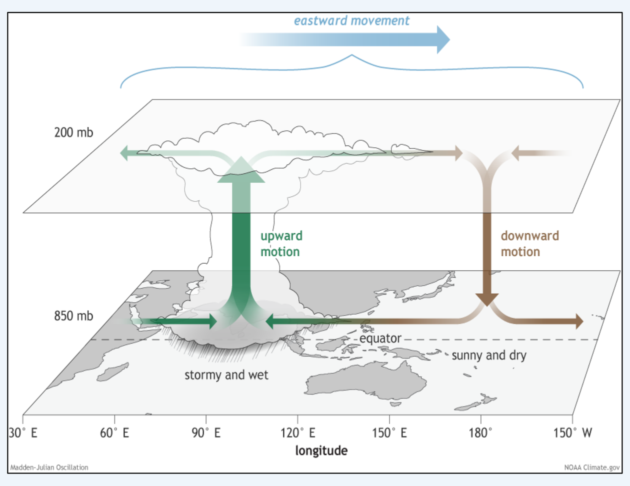
Author: Matthew Rilloraza Supervisors: Prof. Huiling Yuan, Prof. Sandro Veiga TA: Yutao Fan
Home University: University of Illinois Urbana-Champaign Host University: Nanjing University

Abstract

It is well-known that the Madden-Julian Oscillation (MJO) has deep impacts on wintertime precipitation in Southern China. Thus, deepening our understanding of the MJO and its teleconnections is crucial to accurate weather forecasting in this region. While neural network techniques can be robust and highly accurate, they (1) often lack interpretability, and (2) require vast amounts of training data to reach a high-quality result. To circumvent each of these issues, we (1) employ an inherently interpretable Neural Additive Model (NAM) and (2) utilize training data sourced from large-ensemble climate simulations. We aim to determine if there is a meaningful connection to be learned between the MJO and S. China precipitation within the simulated data. We then interpret the network to determine the most important feature contributions. While the model ensemble we produce does not outperform the null accuracy, we hope that this modeling framework can be applied to future MJO studies relating to machine learning and interpretability.

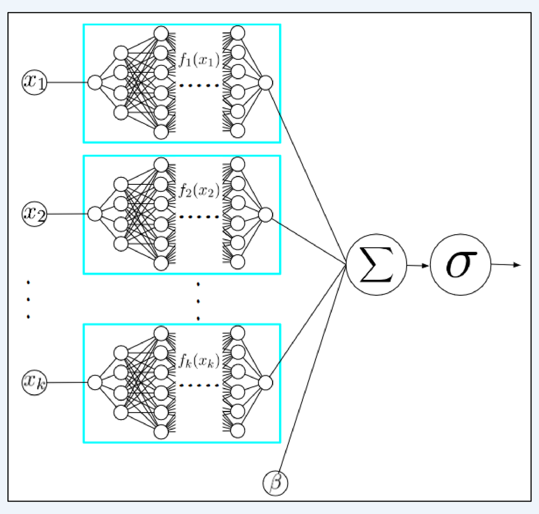
Introduction

The Madden-Julian Oscillation (MJO) is an eastward-propagating convective anomaly characterized by two primary phases: an enhanced rainfall phase and a suppressed rainfall phase. Understanding this phenomenon's teleconnections has great importance to weather forecasting in China, and it would be particularly convenient to develop a model that makes predictions based on readily-obtained MJO indices. Here, we use the OMI index, which is calculated via the principle components (PCs) of Outgoing Longwave Radiation (OLR) data.



MJO diagram.

Neural Additive Models (NAMs) provide a simple yet effective means to increase model interpretability while limiting sacrifices to model expressivity. Each feature (or subset of features) has its own isolated subnetwork (or set of subnetworks) whose contributions are linearly combined via a set of trained weights and biases. By enforcing this linear separation, the contribution of each parameter in isolation can be found exactly and does not have to be approximated. This allows for easier visualization and stricter assessments of feature importance.

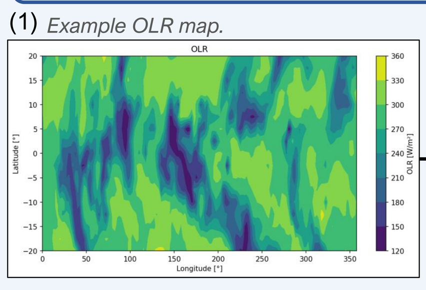

NAM architecture for binary classification.

The Coupled Model Intercomparison Project 6 (CMIP6) is a global initiative through which researchers around the world contribute climate simulation data. Data derived from CMIP iterations have been widely used in climate research since the project's initiation in 1995. Previous studies have shown that machine learning models are able to learn meaningful patterns from CMIP6, due in part to the large wealth of data available. Here, we utilize OLR and precipitation data derived from 14 models to compute simulated MJO indices and define the model target, respectively.

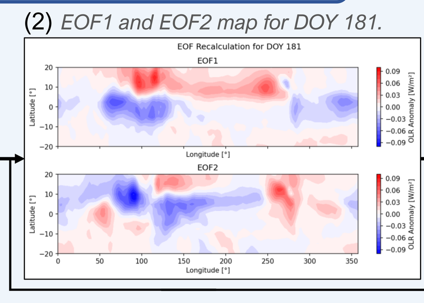
Methods

Compute CMIP6 MJO Index (OMI)

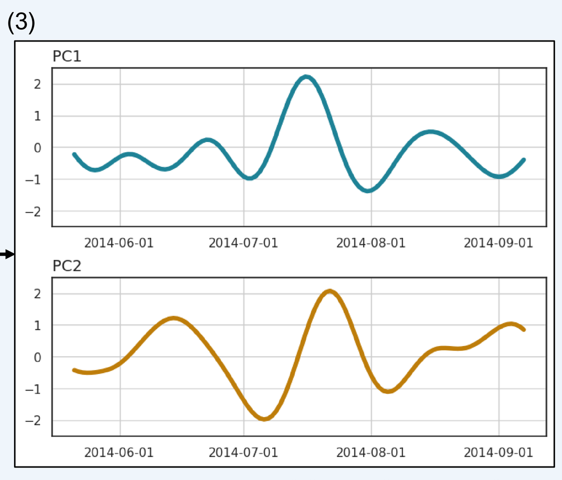
(1) *Example OLR map.*



(2) *EOF1 and EOF2 map for DOY 181.*



(3)

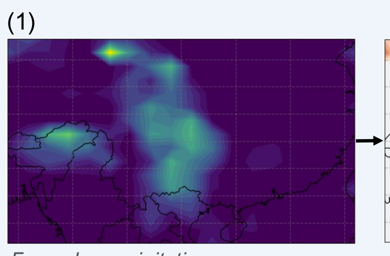


Example time series of OMI PC1 and PC2.

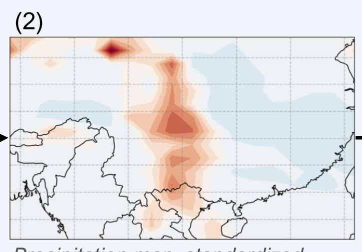
1. Filter CMIP6 OLR data and regrid to +/- 20 deg latitude band
2. Compute empirical orthogonal functions (EOFs)
3. Apply EOFs to OLR data to calculate the OMI index, which consists of the first two PCs

Compute CMIP6 Precipitation Anomaly

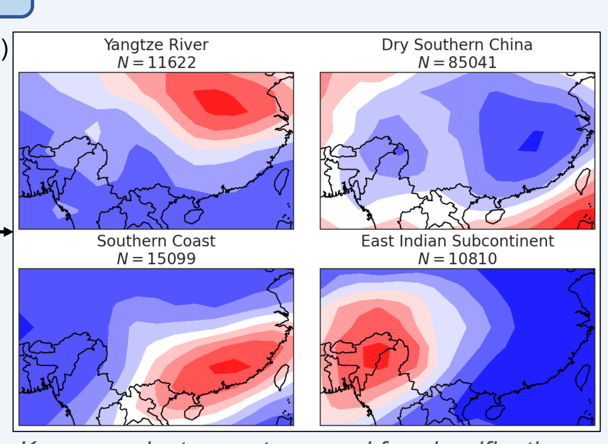
(1)


Example precipitation map.

(2)


Precipitation map, standardized.

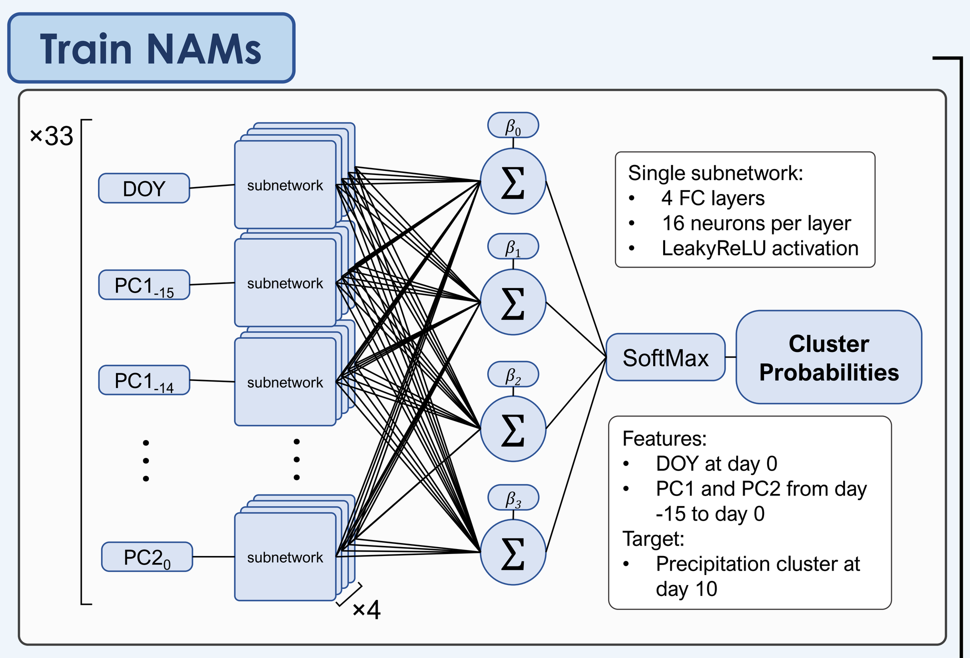
(3)



K-means cluster centers used for classification.

1. Obtain CMIP6 precipitation data
2. Standardize on day of year to obtain annual anomaly
3. Apply k-means clustering to obtain precipitation classes

Train NAMs



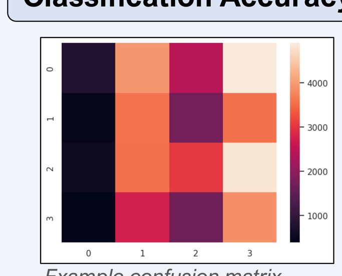
×20 (ensemble)

Training Parameters:
- Cross-Entropy loss (class-weighted)
- Adam optimizer
- LR reduction on learning plateau
- Xavier uniform initialization
- Early stopping (typically activated in ~10-20 epochs)

- *N* = 52,531
  - Dry Southern China cluster undersampled to 15000
- Training set: 80%
  - N_{train} = 42024

Evaluate and Interpret Models

Classification Accuracy


Example confusion matrix.

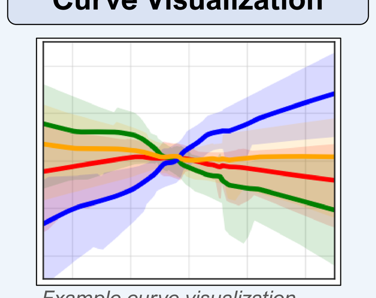
Permutation Importance

- After shuffling feature *j*, we record the factor by which the accuracy decreases, defined as the Permutation Importance (PI).
- We expect more important features to have a greater influence on accuracy and thus a greater PI.
- This is repeated multiple times to reach a stable result.

$$\mathrm{PI}_j = \frac{A - A_j}{A}$$

A = baseline accuracy (percentage correct)
A_j = accuracy after permuting feature j

Curve Visualization


Example curve visualization.

Explained Variance Importance

- We expect features with a greater amount of output variance to be more important.
- The sample variance of each feature normalized by the total variance is the Importance Ratio (IR).

$$D(f_j) = \frac{1}{n-1} \sum_i \left[f_j(x_{ij}) - \langle f_j(x_j) \rangle\right]^2$$

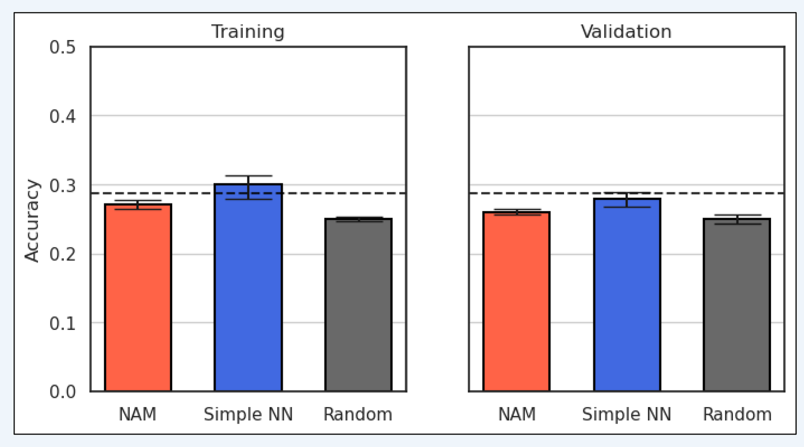
f_j = j^{th} feature network
x_{ij} = sample i of j^{th} feature
n = number of samples

$$\mathrm{IR}_j = \frac{D(f_j)}{\sum_j D(f_j)}$$

Results

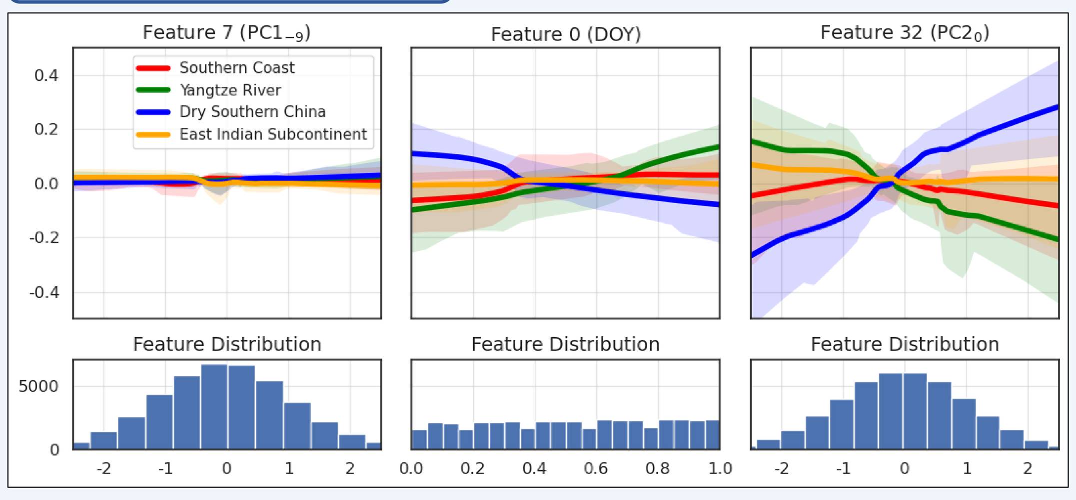
Accuracy Assessment

- NAM performance was compared to that of a simpler fully-connected neural net, as well as random guessing.
- The simple NN outperforms the NAM, but neither is able to overcome the null accuracy, defined by the majority class (horizontal dashed line).
- Both outperform a purely random model, but only by a small margin.



Error bars indicate 5th and 95th percentiles. Dashed line represents portion of majority class in dataset, which represents the null accuracy

Curve Visualization

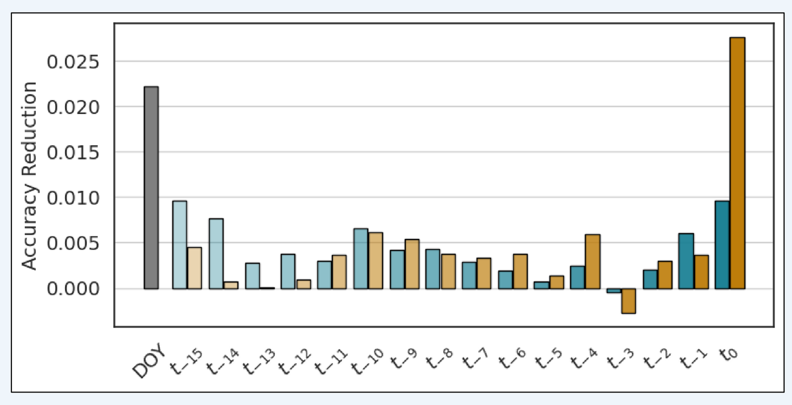


Raw logit curves for individual features. Fill bounds indicate 5th and 95th percentiles across model ensemble. Feature distribution is taken from training set. DOY was min-max normalized between 0 and 1.

- Features such as $PC1_{-9}$ have less of an influence on the output decision than those with a higher output range, such as DOY and $PC2_0$.
- We see more nuance in the decision curve near the centers of feature distributions in the training set.

Permutation Importance

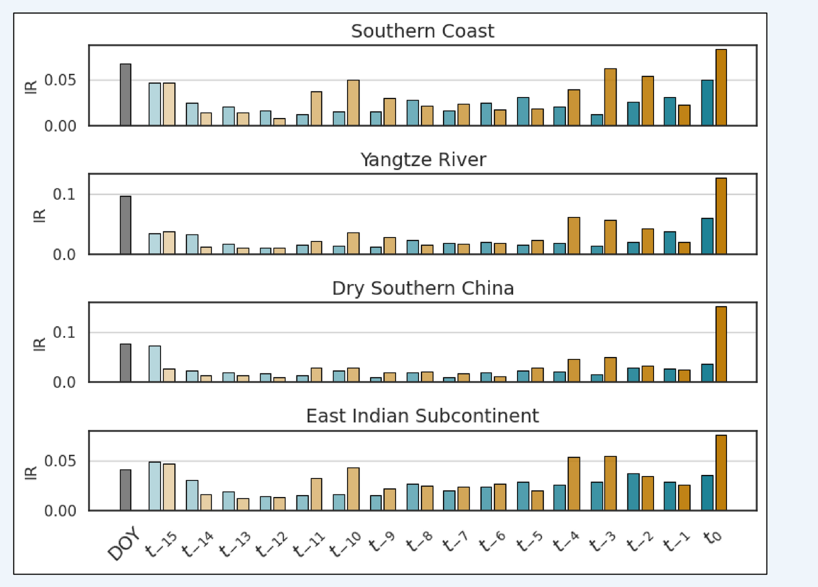
- Most important features: $PC2_0$, DOY, $PC1_0$
- Features closer in time to the precipitation prediction date seem more important, but this behavior drops off at around t = -3



Permutation importance. Larger values are more important. Blue represents PC1, brown PC2. Pairs of bars represent the same lag time. Darker shading indicates a later time.

Explained Variance Importance

- Again, we see $PC2_0$ appearing and DOY appearing as the most important features.
- The importance structure in larger lag times is less defined.
- Since each cluster has its own decision curve, we are able to separate this importance metric per cluster. Between cluster importance ratios, we see a similar pattern.
- Permutation analysis and explained variance analysis seem to agree on which features are the most important.



Explained variance importance. Larger values are more important. Blue represents PC1, brown PC2. Pairs of bars represent the same lag time. Darker shading indicates a later time.

Discussion and Conclusion

A very small pattern could be learned from the dataset, but not in a way that was significantly above the null accuracy. Trained models achieved an accuracy better than random guessing but did not outperform a model that simply learns to predict the majority class. This could indicate that the current data processing framework does not represent the link between MJO and S. China precipitation well. Also, previous work has shown that for certain CMIP6 models, the MJO and its teleconnections are not simulated accurately.
Future work: Experiment with different MJO indices (such as the Real-Time Multivariate MJO (RMM) index) and other data processing strategies. Use varied data sources within CMIP6.

While allowing for greater interpretability, a simpler architecture was able to outperform NAMs in terms of pure accuracy. This simpler architecture only consisted of a few fully-connected layers (and contained interaction terms) yet outperformed the NAMs, which had a much higher parameter count. However, because of the modular nature of NAMs, unique subnetwork architectures could be used to take advantage of more advanced models while maintaining interpretability.
Future work: Experiment with alternative or hybrid NAM architectures, such as a Generalized Additive Model with Structured Interactions (GAMI-net), which allow for pairwise interactions between the most important features.

Despite relatively poor model performance, there is still an importance structure to be found, and decision curves are still meaningful. DOY and the latest lag-time PCs were found to have the highest importance in both metrics used. And within this most-important feature set, a collection relatively consistent decision curves can be found across the different models in the ensemble. If the pattern between MJO and S. China precipitation were more pronounced in this dataset, NAMs could prove useful in not only determining feature importance but also for direct visualization of feature contributions.
Future work: Apply the interpretable aspects of NAMs on different datasets.

Acknowledgements

I would like to thank Prof. Huiling Yuan, Prof. Sandro Veiga, and Yutao Fan for their guidance and support during this research project.

Many thanks to all of the faculty, TAs, and program coordinators who have all made GripS possible.

References

Agarwal, Rishabh, Levi Melnick, Nicholas Frosst, Xuezhou Zhang, Ben Lengerich, Rich Caruana, and Geoffrey Hinton. "Neural Additive Models: Interpretable Machine Learning with Neural Nets." arXiv, October 24, 2021. http://arxiv.org/abs/2004.13912.

Gibson, Peter B., William E. Chapman, Alphan Altinok, Luca Delle Monache, Michael J. DeFlorio, and Duane E. Waliser. "Training Machine Learning Models on Climate Model Output Yields Skillful Interpretable Seasonal Precipitation Forecasts." Communications Earth & Environment 2, no. 1: 159. https://doi.org/10.1038/s43247-021-00225-4.

Hoffmann, Christoph G., George N. Kiladis, Maria Gehne, and Christian Von Savigny. "A Python Package to Calculate the OLR-Based Index of the Madden- Julian-Oscillation (OMI) in Climate Science and Weather Forecasting." Journal of Open Research Software 9, no. 1: 9. https://doi.org/10.5334/jors.331.

Kiladis, George N., Juliana Dias, Katherine H. Straub, Matthew C. Wheeler, Stefan N. Tulich, Kazuyoshi Kikuchi, Klaus M. Weickmann, and Michael J. Ventrice. "A Comparison of OLR and Circulation-Based Indices for Tracking the MJO." Monthly Weather Review 142, no. 5: 1697–1715. https://doi.org/10.1175/MWR-D-13-00301.1.

Mayer, Kirsten J., William E. Chapman, and William A. Manriquez. "Exploring the Relative Importance of the MJO and ENSO to North Pacific Subseasonal Predictability." Geophysical Research Letters 51, no. 10: e2024GL108479. https://doi.org/10.1029/2024GL108479.

Yang, Zebin, Aijun Zhang, and Agus Sudjianto. "GAMI-Net: An Explainable Neural Network Based on Generalized Additive Models with Structured Interactions." Pattern Recognition 120: 108192.

Email: mrr12@illinois.edu (school) - mrillo@pm.me (personal) WeChat ID: uM1eLpcW