

Using simple, explainable neural networks to predict the Madden-Julian oscillation

Zane K. Martin¹

Elizabeth A. Barnes¹

Eric Maloney¹

¹ Department of Atmospheric Science, Colorado State University, Fort Collins, CO

Corresponding author: Zane Martin, zkmartin@colostate.edu

Draft submitted to

Geophysical Research Letters

2021

Key points

1. Simple machine learning frameworks present a new, efficient, flexible tool to predict and study the Madden-Julian oscillation (MJO)
2. Shallow artificial neural networks (ANNs) predict a real-time MJO index out to ~17 days in winter and ~10 days in summer
3. Varying ANN input and using explainable artificial intelligence methods offer insights into the MJO and key regions for prediction skill

Abstract: Few studies have utilized machine learning techniques to predict or understand the Madden-Julian oscillation (MJO), a key source of subseasonal variability and predictability. Here we present a simple framework for real-time MJO prediction using shallow artificial neural networks (ANNs). We construct two ANN architectures that predict a real-time MJO index using latitude-longitude maps of tropical variables. These ANNs make skillful MJO predictions out to ~17 days in October-March and ~10 days in April-September, and efficiently capture aspects of MJO predictability found in more complex, computationally-expensive models. Varying model input and applying ANN explainability techniques further reveal sources and regions important for ANN prediction skill. This simple machine learning framework can be more broadly adapted and applied to predict and understand other climate phenomena.

Plain Language Summary: The Madden-Julian oscillation (MJO) -- a large-scale, organized pattern of wind and rain in the tropics -- is important for making weather and climate predictions weeks to months into the future. Many different models have been used to study the MJO, but few works have examined how machine learning and artificial intelligence methods might inform our ability to predict and understand the MJO. In this work, we show how two different types of machine learning models, called artificial neural networks, perform at predicting the MJO. We demonstrate that simple artificial neural networks make skillful MJO predictions beyond 1-2 weeks into the future, and perform

better than simpler statistical methods. We also highlight how neural networks can be used to explore sources of prediction skill and inform understanding of the MJO, via changing what variables the model uses and applying techniques that illuminate important regions for skillful predictions. Because our neural networks perform relatively well, are simple to implement, are computationally affordable, and can be used to inform understanding, we believe these methods may be more broadly applicable to study other important climate phenomena aside from just the MJO.

1. Introduction

The Madden-Julian oscillation (MJO), a planetary-scale, eastward-propagating coupling of tropical circulation and convection (Madden and Julian 1971, 1972; Zhang 2005), is a key source of subseasonal-to-seasonal (S2S) predictability (Vitart et al. 2017; Kim et al. 2018). Skillful MJO prediction has important societal implications (Meehl et al. 2021; Vitart et al. 2017; Kim et al. 2018), and extensive research has explored using both empirical/statistical models or initialized dynamical forecast models to predict and study the MJO (e.g. Waliser 2012; Vitart et al. 2017; Kim et al. 2018; Meehl et al. 2021; and references therein). Before the late 2000s, statistical models showed superior MJO prediction skill (~2 weeks; Waliser 2012; Kang and Kim 2010), but dynamical models have continually improved and several now skillfully predict the MJO beyond one month (Vitart 2014; Vitart 2017; Kim et al. 2018). In contrast, statistical MJO modeling has stagnated, and statistical MJO models still almost exclusively use traditional linear methods (e.g. Kang and Kim 2010). New approaches are needed to drive statistical MJO modeling forward.

Non-linear machine learning (ML) techniques have proven skillful at predicting various climate and weather phenomena (Gagne et al. 2014; Lagerquist et al. 2017; McGovern et al. 2017; Weyn et al. 2019; Rasp et al. 2020; Ham et al. 2019; Mayer and Barnes 2021), but limited work has considered MJO prediction. Recent ML studies have identified the MJO (Toms et al. 2020b), reconstructed past MJO behavior (Dasgupta et al. 2020), or bias-corrected dynamical model output of MJO indices (Kim et al. 2021), but only one study to our knowledge has examined MJO prediction solely using ML (Love and Matthews 2009). It is thus timely to establish ML frameworks for predicting the MJO and begin to quantify ML model performance. Equally important is demonstrating that ML models are useful for more than just prediction: they invite experimentation and can inform physical understanding of the MJO. We highlight this underappreciated aspect of ML modeling through experiments changing input data, the exploration of ML model architecture, and the application of tools from the field of explainable AI (XAI; McGovern et al. 2019; Toms et al. 2020a; Mamalakis et al. 2021).

This paper addresses three aspects of using machine learning to study the MJO: (1) developing ML frameworks, (2) analyzing ML model performance, and (3) demonstrating how ML can inform physical understanding. We prioritize simple techniques (i.e. shallow, fully-connected artificial neural networks; ANNs) to

establish a benchmark for future ML modeling, ensure our approach is broadly accessible, and facilitate applying XAI tools. Further, while our focus is the MJO, the concept and methods we describe are widely transferable to other areas in Earth science, and may help inform ML modeling of other climate phenomena.

1. Data and Methodology

(a) *Data*

The predictors of our ANN models are latitude-longitude maps of processed tropical variables from 20°N-20°S. The predictand is the Real-time Multivariate MJO index (Wheeler and Hendon 2004) which tracks the MJO using an empirical orthogonal function (EOF) analysis of outgoing longwave radiation (OLR), and zonal wind at 850 and 200 hPa (Wheeler and Hendon 2004). The index consists of two time series (“RMM1” and “RMM2”) that together represent the strength and location of the MJO. Plotted on a 2-D plane, the RMM phase angle describes the location, or “phase”, of the MJO (e.g. Fig. 1), while the RMM amplitude ($\sqrt{RMM1^2 + RMM2^2}$) measures MJO strength. RMM has known limitations (Roundy et al. 2009; Straub 2013) and other MJO indices exist (Kiladis et al. 2014; Kikuchi et al. 2012; Ventrice et al. 2013), but RMM represents a logical starting point as it is a widely-used benchmark index suitable for real-time forecasts.

Other data are from three sources: OLR data is from the NOAA Interpolated OLR dataset (Liebmann and Smith 1996), sea-surface temperature (SST) is from the NOAA OI SST V2 High Resolution dataset (Reynolds et al. 2007), and all other variables are from ERA-5 reanalysis (Hersbach et al. 2020). Additional data from the ERA-20C dataset (Poli et al. 2016) is used in the Supplemental Material, as described therein. Data are daily from January 1, 1979 (1982 for SST) to December 31, 2019, and interpolated onto a common 2.5° x 2.5° grid. Data are pre-processed similarly to the RMM input variables by subtracting the daily climatology’s mean and first three seasonal-cycle harmonics, and a previous 120-day mean (Wheeler and Hendon 2004). Variables are not averaged latitudinally because we are interested in how the 2-D structure is utilized by the ANNs (though sensitivity tests exploring latitudinal averaging are discussed in Supplemental Text S2). We also normalize each variable by subtracting the tropics-wide, all-time mean and dividing by the tropics-wide, all-time standard deviation at each grid point.

1. *Artificial Neural Networks*

We explored two ANN architectures to study the MJO (Fig. 1): a “regression model” and a “classification model”. We present a high-level overview of both architectures here; for details see Supplemental Text S1.

Both ANN architectures input the processed latitude-longitude maps of one or more variables from a single day, and output information about the RMM index N days into the future (Fig. 1). A separate ANN is trained for each N from 0

to 20 days. The difference between the regression and classification ANNs is the nature of their outputs. The regression ANN (not to be confused with a linear regression model) outputs numerical values of RMM1 and RMM2 (Figs. 1a, S1). The classification model outputs the probability that the MJO is in each of nine classes: active (RMM amplitude > 1) in one of the eight canonical MJO phases (Wheeler and Hendon 2004) or weak. The classification ANN outputs a probability for each class, and the prediction is the class with the highest probability (e.g. Figs. 1b, S2).

Both the regression and classification ANNs have a single hidden layer of 16 nodes, and use ridge regularization and early stopping to avoid overfitting (see Supplemental Text S1). They are trained on daily data from June 1, 1979 (1982 for SST) to December 31, 2009. For the classification ANN, since weak MJO days are the most common class (~39% of all days) we avoid class imbalance by randomly subsampling weak MJO days during training so they are 11% of all training days. Weak days are not subsampled over the validation period. Both ANNs are evaluated over the period January 1, 2010 to November 30, 2019, with forecasts initialized daily.

ANN performance is slightly improved if the models are trained separately on different seasons (e.g. Supplemental Text S1, Fig. S3), which allows the ANNs to learn more season-specific patterns. This is likely important for the MJO due to its seasonal shifts in behavior, strength, and structure (Hendon and Salby 1994; Hendon et al. 1999; Zhang and Dong 2004), and we found splitting the data into two six-month periods (October-March, or “boreal winter”, and April-September, or “boreal summer”) provided a good trade-off between seasonal specificity and number of training samples.

Finally, in some instances we trained multiple ANNs for the same seasons and lead times, varying only the random initial training weights to ensure convergence of our results and quantify sensitivity to initialization. For more information regarding the details of our data, ANN architecture, and ANN parameters see Supplemental Text S1. Sensitivity tests varying ANN parameters and input data pre-processing are discussed in Supplemental Text S2.

To assess model skill in the regression ANN, we utilize the bivariate correlation coefficient (BCC; Vitart et al. 2017; Kim et al. 2018), with a value greater than 0.5 denoting skill. In the classification ANN, skill is measured using the accuracy predicting the MJO phase. ANN performance is also assessed relative to a statistical “persistence model” that forecasts the average counter-clockwise propagation of the RMM index, holding RMM amplitude fixed from its initial state (see Supplemental Text S1). Thus, the persistence model forecasts represent the average RMM behavior corresponding to the slow eastward propagation of MJO signals.

1. *Layer-wise Relevance Propagation*

To demonstrate how the classification ANN correctly captures regions of importance for predicting the MJO, we use an ANN explainability technique called

layer-wise relevance propagation (LRP; Bach et al. 2015; Samek et al. 2016; Montavon et al. 2019). LRP has begun to gain traction in Earth science as a tool for understanding ANNs (Toms et al. 2020a; Toms et al. 2020b; Barnes et al. 2020; Mayer and Barnes 2021; Mamalakis et al. 2021; Madakumbura et al. 2021), and more technical detail on LRP is provided in Supplemental Text S1 and the above-cited works. Broadly, LRP is an algorithm applied to a trained ANN. After a prediction is made, LRP back-propagates that prediction’s output through the ANN in reverse. Ultimately, LRP returns a vector of the same size as the input (here a latitude-longitude map), where the returned quantity, termed the “relevance”, shows which input nodes (i.e. which latitude-longitude points) were most important in determining that prediction.

1. Artificial Neural Network Performance

(a) Overall Model Performance

In this subsection we focus on the regression and classification ANNs that input three variables simultaneously: OLR, zonal wind at 850 hPa, and zonal wind at 200 hPa (Fig. 1). This input combination is among the best-performing across the experiments we conducted, and are the same variables that comprise RMM. Additional inputs are discussed in Section 3.2 and Supplemental Text S2.

Overall, the winter and summer regression ANNs show prediction skill, respectively, of ~17 days and ~11 days (Fig. 2a), with small spread across a 10-member ensemble. This is at the forefront of statistical MJO prediction techniques (e.g. Kang and Kim 2010; Kim et al. 2018), which is impressive given the simplicity of the ANNs, the real-time nature of the model, and the lack of substantial input pre-processing. Both winter and summer regression ANNs have a higher BCC than the persistence model after 2-3 days and make skillful predictions out to longer leads than the persistence model. This demonstrates that the ANNs learn not only to identify the MJO and propagate it east, but also capture more nuanced MJO behavior. The higher skill in winter versus summer is consistent with results in most dynamical models (e.g., Vitart 2017), indicating that ANNs are able to reproduce aspects of MJO predictability seen in more complex dynamical modeling set-ups. The regression ANN skill also shows relatively small sensitivity to initial MJO phase (Fig. S4), with somewhat higher skill (~18-19 days) across MJO events initialized in phases 1-3 and lower skill (~14-15 days) for phases 6 and 8. Whether this change is statistically significant, and what may explain it, was not examined in detail.

A strength of the regression ANN is the quantitative information it provides about MJO location and strength, but a prevalent source of error is a decrease in the ANN-predicted MJO amplitude at lead times past a few days, especially in phases 4-7 (Fig. 1a, Figs. S1, S4). Minimizing this bias is an open challenge, and was one motivation for exploring a classification ANN architecture that focuses on the MJO phase alone.

Similar to the regression ANN, the classification ANN shows higher skill in winter than in summer (Fig. S3) and outperforms the persistence model after

3-4 days (Fig. 2b). It further shows accuracy well above random chance for all lead times (Figs 2b, S5), indicating the model possesses skill out to several weeks. At short lead times, where the classification model is primarily identifying the MJO (similar to Toms et al. 2020b), the phase of active MJO events are predicted with an accuracy of ~70-80% (Fig. 2b). Most incorrectly predicted active MJO events at short leads are near the boundary between two RMM phases and predictions are often incorrect by only one phase (Fig. S5). The lead-0 accuracy here is comparable to Toms et al. (2020b), despite differences in our input variables, data pre-processing, MJO index, and ANN complexity. At longer leads, accuracy decreases monotonically, but stays above 20% for active MJO events out to 20 days (Fig. 2b). In contrast to performance predicting active MJO events, the classification ANN struggles to predict weak MJO days (Fig. 2b, S5), with ~40% accuracy identifying weak days at lead 0, and worse performance at longer leads (**Fig. S5**). The improvement in ANN accuracy forecasting active MJO days versus all days is on the order of 15% (Fig. 2b), even out past 2-week lead times.

The tradeoffs between classification and regression ANN architectures make choosing a “better” model difficult. The regression model outputs more precise RMM information and is more readily comparable to existing statistical and dynamical models, but struggles to predict strong MJO amplitudes at long leads. This is true even when the regression model was re-trained using fewer weak MJO days to emphasize strong MJO events, as little change in performance was seen (Fig. S6). The classification ANN shows the opposite tendency, overestimating the percentage of active MJO days and struggling to accurately predict weak MJO events. Further, accuracy of MJO phase is not a commonly used metric of model skill, and the classification ANN (as set up here) cannot provide precise information about MJO strength and location.

Still, results for both ML architectures show that aspects of the MJO are skillfully predicted beyond two weeks in winter. A range of sensitivity tests (Supple. Text S2 and Figs. S7, S8), including increasing the amount of training data using 20th-century reanalysis, showed comparable performance, though tests were not exhaustive nor explored beyond simple ANN architectures. Also note that while our primary goal here is to study ML modeling of the MJO, these simple ANNs are not yet competitive with most S2S dynamical models (e.g. Vitart 2017; Kim et al. 2018). It remains to be seen whether future ML research might change this, but as the next section illustrates, ANNs can be used as a tool for more than just prediction, and may help spur new discoveries or generate new hypotheses.

1. *Efficiency and Utility of ANNs*

In addition to examining the overall prediction skill, we also explored how ANNs can continue to inform our understanding of the MJO and MJO predictability. For example, ANNs reproduce aspects of MJO prediction skill found in dynamical models, but they do so at a fraction of the cost and capture more than just seasonal changes in skill (Fig. 2a). Both regression and classification models

show higher skill for initially stronger MJO events (Figs. 2c, S9): the regression ANN skillfully predicts MJO events which are initially strong or very strong (RMM amplitude > 1.5) out to ~ 20 days in winter, while skill predicting weak events is only ~ 10 days. ANNs also capture more mysterious aspects of MJO predictability, such as the sensitivity to the phase of the stratospheric quasi-biennial oscillation (Marshall et al. 2017; Lim et al. 2019; Wang et al. 2019; Martin et al. 2021): the wintertime regression ANN skill during QBO easterly months is nearly 20 days, whereas during QBO westerly months skill is only 15 days (Fig. S10).

Another utility of the classification ANN is the “model confidence,” which describes the ANN’s surety that a given prediction is correct. Model confidence has clear utility for forecasters, could drive future work in probabilistic MJO prediction (Marshall et al. 2016), and may be useful in improving understanding of MJO predictability. Regardless of lead time, our classification ANNs are reliable – in the sense that ANN confidence corresponds well with model accuracy – which indicates that model confidence is a useful and meaningful output (Fig. 2d). Furthermore, ANN confidence relates to physical aspects of the MJO: confidence is closely associated with initial MJO amplitude (correlation coefficients of ~ 0.5 - 0.7 depending on lead), with higher confidence associated with higher initial MJO amplitude (Fig. 2d). Research using ANN confidence to identify predictable states of the atmosphere has recently shown promise in the context of MJO teleconnections to the extra-tropics (Barnes et al. 2020; Mayer and Barnes 2021), and thus may lead to new insights into MJO predictability or behavior.

Simple ANNs can also be used to efficiently explore the impact of certain variables on MJO prediction. We illustrate this through classification ANN experiments inputting various combinations of one to three different variables, targeting leads 0, 5, and 10 days for brevity. Overall, model performance varies widely depending on input (Fig. 3). For example, across 1-variable ANNs (Fig. 3a) 850 hPa meridional wind and sea-surface temperature (SST) models show much poorer performance than other inputs. In the case of the SST model, this suggests the ocean state alone (when processed to highlight subseasonal variability) does not contain MJO signals the ANN is able to leverage, consistent with findings that sub-seasonal SST variability does not drive the MJO (e.g. Newman et al. 2009). In the case of meridional wind, while the MJO possesses signals in meridional wind associated with Rossby wave gyres (Zhang 2005), we hypothesize that skill may be low because these signals lack the global-scale coherence seen in variables like zonal wind and OLR and captured by RMM.

The most accurate models at short leads are those that input 850 hPa and/or 200 hPa zonal winds (Fig. 3). This is consistent with literature showing that MJO circulation tends to drive the RMM index (Straub 2013; Ventrice et al. 2013), an aspect of RMM the ANN has organically learned. Interestingly, skill identifying the MJO at short leads does not necessarily imply similar performance predicting the MJO at longer leads. For example, at lead 0 the 850hPa

and 200 hPa zonal wind model has the clear highest accuracy among 2-variable models (Fig. 3b), but at lead 5 and 10 its accuracy overlaps with other configurations. Best performing models at longer leads are those that include information about zonal wind and the large-scale thermodynamic or moisture signature of the MJO, as measured for example by OLR or column water vapor. Further, RMM input variables are not always clearly superior at leads 5 and 10: a model with total column water, 200 hPa zonal wind and 200 hPa temperature performs as well as or slightly better than the model with 200 and 850 hPa zonal wind and OLR (Fig. 3c).

Finally, while more input variables tend to improve model performance (Fig. 3), tests showed no substantial improvement using 4 or more inputs (Fig. S11), at least among the variables considered here. Whether this is due to the limited complexity of our ANNs, the amount of training data, or because new, meaningful information is difficult to leverage with more variables is not known. Additional variables (perhaps with different preprocessing) will continue to be explored, but these initial tests provide a proof-of-concept for the kind of experimentation that ANNs afford.

1. ANN Explainability Through LRP

A further advantage of ANNs versus other MJO modeling frameworks is ability to apply XAI tools like LRP, which can illuminate sources of model prediction skill. As an example, Figure 4 shows wintertime composite LRP maps using the classification ANN from Section 3.1. LRP maps are shown for lead times of 0 and 10 days, composited across correct ANN predictions when the observed MJO is in phase 5 at the time of verification. Composites are further restricted to those events when model confidence exceeds the 60th percentile (see Supplemental Text S1 for more detail).

The LRP plots confirm that the classification ANN focuses on regions central to the MJO. At lead 0, OLR relevance highlights suppressed Indian Ocean convection and active conditions around the Maritime Continent (Fig. 4a,b), whereas wind fields focus on low-level westerly anomalies around the Maritime Continent (Fig. 4c,d) and upper level signals in the central and east Pacific (Fig. 4e,f). At lead 10, LRP shows how the ANN accounts for eastward MJO propagation: the maximum relevance for OLR is shifted west relative to lead 0, highlighting strong convection in the eastern Indian ocean (Fig. 4g,h). The lead-10 model also focuses on a small region of strong low-level convergence near the equatorial Maritime Continent, and upper-level easterly anomalies in the western Indian Ocean (Figs. 4i-l). LRP thus provides information about how the ANN identifies the MJO and what signals across variables are most associated with future MJO behavior. The unique information LRP provides may also be used to explore sources of MJO prediction skill under different large-scale states or for case studies of particular events.

1. Discussion & Conclusions

Motivated by a lack of recent progress in statistical MJO modeling and the abil-

ity of machine learning methods to skillfully predict other climate and weather phenomena, here we demonstrate how simple machine learning frameworks can be used to predict the MJO. We established two straightforward neural network architectures (a regression and classification approach) that use shallow ANNs to predict an MJO index. The regression ANN shows prediction skill out to ~17 days in winter and ~11 days in summer, which is high skill for a statistical approach, and both ANN architectures set benchmarks for continued ML modeling of the MJO. Simple ANNs are also efficiently able to reproduce aspects of MJO predictability found in more complex, computationally-expensive dynamical models, making them affordable tools to continue to study the MJO and MJO predictability. Explainable AI tools can also help illuminate sources and regions of ANN model skill.

This work illustrates how simple ANNs can be used not only for prediction, but also as tools for hypothesis testing and experimentation that might drive new discoveries or scientific insights. While our focus here is on the MJO, the framework we establish is widely applicable to a range of different climate phenomena, especially oscillations that can be represented as simple indices. The performance, affordability, accessibility, and explainability of simple ANNs thus recommends their continued adoption by the climate community.

Acknowledgments

Z.K.M. acknowledges support from the National Science Foundation under Award No. 2020305. E.A.B. and E.D.M. are supported, in part, by NOAA Climate Test Bed Grant NA18OAR4310296. E.D.M. also acknowledges support from NSF Climate and Large-Scale grant AGS-1841754, and NOAA CVP Grant NA18OAR4310299.

Data availability

All datasets used in this study are publicly available. The RMM index (Wheeler and Hendon 2004) is available at www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt. For reanalysis and observed data, NOAA Interpolated OLR (Liebmann and Smith 1996) is available at https://psl.noaa.gov/data/gridded/data.interp_OLR.html; NOAA OI SST V2 High Resolution (Reynolds et al. 2007) is available at <https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html>; ERA-5 reanalysis (Hersbach et al. 2020) is available at <https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset>; and ERA-20C data (Poli et al. 2016) is available at <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-20c>.

References

Bach, Sebastian, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. 2015. “On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation.” *PloS One* 10 (7): e0130140.

- Barnes, Elizabeth A., Kirsten Mayer, Benjamin Toms, Zane Martin, and Emily Gordon. 2020. “Identifying Opportunities for Skillful Weather Prediction with Interpretable Neural Networks.” *arXiv [physics.ao-ph]*. arXiv. <http://arxiv.org/abs/2012.07830>.
- Dasgupta, Panini, Abirlal Metya, C. V. Naidu, Manmeet Singh, and M. K. Roxy. 2020. “Exploring the Long-Term Changes in the Madden Julian Oscillation Using Machine Learning.” *Scientific Reports* 10 (1): 18567.
- Gagne, David John, Amy McGovern, and Ming Xue. 2014. “Machine Learning Enhancement of Storm-Scale Ensemble Probabilistic Quantitative Precipitation Forecasts.” *Weather and Forecasting* 29 (4): 1024–43.
- Ham, Yoo-Geun, Jeong-Hwan Kim, and Jing-Jia Luo. 2019. “Deep Learning for Multi-Year ENSO Forecasts.” *Nature* 573 (7775): 568–72.
- Hendon, Harry H., and Murry L. Salby. 1994. “The Life Cycle of the Madden-Julian Oscillation.” *Journal of the Atmospheric Sciences* 51 (15): 2225–37.
- Hendon, Harry H., Chidong Zhang, and John D. Glick. 1999. “Interannual Variation of the Madden-Julian Oscillation during Austral Summer.” *Journal of Climate* 12 (8): 2538–50.
- Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, et al. 2020. “The ERA5 Global Re-analysis.” *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999–2049.
- Kang, In-Sik, and Hye-Mi Kim. 2010. “Assessment of MJO Predictability for Boreal Winter with Various Statistical and Dynamical Models.” *Journal of Climate* 23 (9): 2368–78.
- Kikuchi, Kazuyoshi, Bin Wang, and Yoshiyuki Kajikawa. 2012. “Bimodal Representation of the Tropical Intraseasonal Oscillation.” *Climate Dynamics* 38 (9-10): 1989–2000.
- Kiladis, George N., Juliana Dias, Katherine H. Straub, Matthew C. Wheeler, Stefan N. Tulich, Kazuyoshi Kikuchi, Klaus M. Weickmann, and Michael J. Ventrice. 2014. “A Comparison of OLR and Circulation-Based Indices for Tracking the MJO.” *Monthly Weather Review* 142 (5): 1697–1715.
- Kim, H., Y. G. Ham, Y. S. Joo, and S. W. Son. 2021. “Deep Learning for Bias Correction of MJO Prediction.” *Nature Communications* 12 (1): 1–7.
- Kim, Hyemi, Frédéric Vitart, and Duane E. Waliser. 2018. “Prediction of the Madden-Julian Oscillation: A Review.” *Journal of Climate* 31 (23): 9425–43.
- Lagerquist, Ryan, Amy McGovern, and Travis Smith. 2017. “Machine Learning for Real-Time Prediction of Damaging Straight-Line Convective Wind.” *Weather and Forecasting* 32 (6): 2175–93.

- Liebmann, Brant, and Catherine A. Smith. 1996. “Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset.” *Bulletin of the American Meteorological Society* 77 (6): 1275–77.
- Lim, Yuna, Seok-Woo Son, Andrew G. Marshall, Harry H. Hendon, and Kyong-Hwan Seo. 2019. “Influence of the QBO on MJO Prediction Skill in the Subseasonal-to-Seasonal Prediction Models.” *Climate Dynamics*, March, 1–15.
- Love, Barnaby S., and Adrian J. Matthews. 2009. “Real-Time Localised Forecasting of the Madden-Julian Oscillation Using Neural Network Models.” *Quarterly Journal of the Royal Meteorological Society* 135 (643): 1471–83.
- Madakumbura, Gavin D., Chad W. Thackeray, Jesse Norris, Naomi Goldenson, and Alex Hall. 2021. “Anthropogenic Influence on Extreme Precipitation over Global Land Areas Seen in Multiple Observational Datasets.” *Research Square*, April. <https://doi.org/10.21203/rs.3.rs-227967/v2>.
- Madden, Roland A., and Paul R. Julian. 1971. “Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific.” *Journal of the Atmospheric Sciences* 28 (5): 702–8.
- Madden, Roland A., and Paul R. Julian. 1972. “Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period.” *Journal of the Atmospheric Sciences* 29 (6): 1109–23.
- Mamalakis, Antonios, Imme Ebert-Uphoff, and Elizabeth A. Barnes. 2021. “Neural Network Attribution Methods for Problems in Geoscience: A Novel Synthetic Benchmark Dataset.” *arXiv [physics.geo-Ph]*. arXiv. <http://arxiv.org/abs/2103.10005>.
- Marshall, Andrew G., Harry H. Hendon, and Debra Hudson. 2016. “Visualizing and Verifying Probabilistic Forecasts of the Madden-Julian Oscillation.” *Geophysical Research Letters* 43 (23): 12,278–12,286.
- Marshall, Andrew G., Harry H. Hendon, Seok Woo Son, and Yuna Lim. 2017. “Impact of the Quasi-Biennial Oscillation on Predictability of the Madden-Julian Oscillation.” *Climate Dynamics* 49 (4): 1365–77.
- Martin, Zane, Seok-Woo Son, Amy Butler, Harry Hendon, Hyemi Kim, Adam Sobel, Shigeo Yoden, and Chidong Zhang. 2021. “The Influence of the Quasi-Biennial Oscillation on the Madden-Julian Oscillation.” *Nature Reviews Earth & Environment*, June, 1–13.
- Mayer, Kirsten J., and Elizabeth A. Barnes. 2021. “Subseasonal Forecasts of Opportunity Identified by an Explainable Neural Network.” *Geophysical Research Letters*, May. <https://doi.org/10.1029/2020gl092092>.
- McGovern, Amy, Kimberly L. Elmore, David John Gagne, Sue Ellen Haupt, Christopher D. Karstens, Ryan Lagerquist, Travis Smith, and John K. Williams. 2017. “Using Artificial Intelligence to Improve Real-Time Decision-Making for

- High-Impact Weather.” *Bulletin of the American Meteorological Society* 98 (10): 2073–90.
- McGovern, Amy, Ryan Lagerquist, David John Gagne, G. Eli Jergensen, Kimberly L. Elmore, Cameron R. Homeyer, and Travis Smith. 2019. “Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning.” *Bulletin of the American Meteorological Society* 100 (11): 2175–99.
- Meehl, Gerald A., Jadwiga H. Richter, Haiyan Teng, Antonietta Capotondi, Kim Cobb, Francisco Doblas-Reyes, Markus G. Donat, et al. 2021. “Initialized Earth System Prediction from Subseasonal to Decadal Timescales.” *Nature Reviews Earth & Environment*, April, 1–18.
- Montavon, Grégoire, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. 2019. “Layer-Wise Relevance Propagation: An Overview.” In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, edited by Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller, 193–209. Cham: Springer International Publishing.
- Newman, Matthew, Prashant D. Sardeshmukh, and Cécile Penland. 2009. “How Important Is Air–Sea Coupling in ENSO and MJO Evolution?” *Journal of Climate* 22 (11): 2958–77.
- Poli, Paul, Hans Hersbach, Dick P. Dee, Paul Berrisford, Adrian J. Simmons, Frédéric Vitart, Patrick Laloyaux, et al. 2016. “ERA-20C: An Atmospheric Reanalysis of the Twentieth Century.” *Journal of Climate* 29 (11): 4083–97.
- Rasp, Stephan, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, and Nils Thuerey. 2020. “WeatherBench: A Benchmark Data Set for Data-driven Weather Forecasting.” *Journal of Advances in Modeling Earth Systems* 12 (11). <https://doi.org/10.1029/2020ms002203>.
- Reynolds, Richard W., Thomas M. Smith, Chunying Liu, Dudley B. Chelton, Kenneth S. Casey, and Michael G. Schlax. 2007. “Daily High-Resolution-Blended Analyses for Sea Surface Temperature.” *Journal of Climate* 20 (22): 5473–96.
- Roundy, Paul E., Carl J. Schreck, and Matthew A. Janiga. 2009. “Contributions of Convectively Coupled Equatorial Rossby Waves and Kelvin Waves to the Real-Time Multivariate MJO Indices.” *Monthly Weather Review* 137 (1): 469–78.
- Samek, Wojciech, Grégoire Montavon, Alexander Binder, Sebastian Lapuschkin, and Klaus-Robert Müller. 2016. “Interpreting the Predictions of Complex ML Models by Layer-Wise Relevance Propagation.” *arXiv [stat.ML]*. arXiv. <http://arxiv.org/abs/1611.08191>.
- Straub, Katherine H. 2013. “MJO Initiation in the Real-Time Multivariate MJO Index.” *Journal of Climate* 26 (4): 1130–51.

- Toms, Benjamin A., Elizabeth A. Barnes, and Imme Ebert-Uphoff. 2020a. “Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability.” *Journal of Advances in Modeling Earth Systems* 12 (9): 1.
- Toms, Benjamin A., Karthik Kashinath, Prabhat, and Da Yang. 2020b. “Testing the Reliability of Interpretable Neural Networks in Geoscience Using the Madden-Julian Oscillation.” *Geosci. Model Dev. Discuss.*, <https://doi.org/10.5194/gmd-2020-152>.
- Ventrice, Michael J., Matthew C. Wheeler, Harry H. Hendon, Carl J. Schreck, Chris D. Thorncroft, and George N. Kiladis. 2013. “A Modified Multivariate Madden–Julian Oscillation Index Using Velocity Potential.” *Monthly Weather Review* 141 (12): 4197–4210.
- Vitart, F., C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, et al. 2017. “The Subseasonal to Seasonal (S2S) Prediction Project Database.” *Bulletin of the American Meteorological Society* 98 (1): 163–73.
- Vitart, Frédéric. 2014. “Evolution of ECMWF Sub-Seasonal Forecast Skill Scores.” *Quarterly Journal of the Royal Meteorological Society* 140 (683): 1889–99.
- Vitart, Frédéric. 2017. “Madden-Julian Oscillation Prediction and Teleconnections in the S2S Database.” *Quarterly Journal of the Royal Meteorological Society* 143 (706): 2210–20.
- Waliser, Duane 2012. “Predictability and Forecasting.” In *Intraseasonal Variability in the Atmosphere-Ocean Climate System*, edited by William K-M Lau and Duane E. Waliser, 433–76. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Wang, S., M. K. Tippett, A. H. Sobel, Z. K. Martin and F. Vitart, 2019. “Impact of the QBO on Prediction and Predictability of the MJO Convection.” *Journal of Geophysical Research: Atmospheres*. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030575>
- Weyn, Jonathan A., Dale R. Durran, and Rich Caruana. 2019. “Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height from Historical Weather Data.” *Journal of Advances in Modeling Earth Systems* 11 (8): 2680–93.
- Wheeler, Matthew C., and Harry H. Hendon. 2004. “An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction.” *Monthly Weather Review* 132 (8): 1917–32.
- Zhang, Chidong. 2005. “Madden-Julian Oscillation.” *Reviews of Geophysics* 43 (2). <https://doi.org/10.1029/2004rg000158>.
- Zhang, Chidong, and Min Dong. 2004. “Seasonality in the Madden–Julian Oscillation.” *Journal of Climate* 17 (16): 3169–80.

Figures

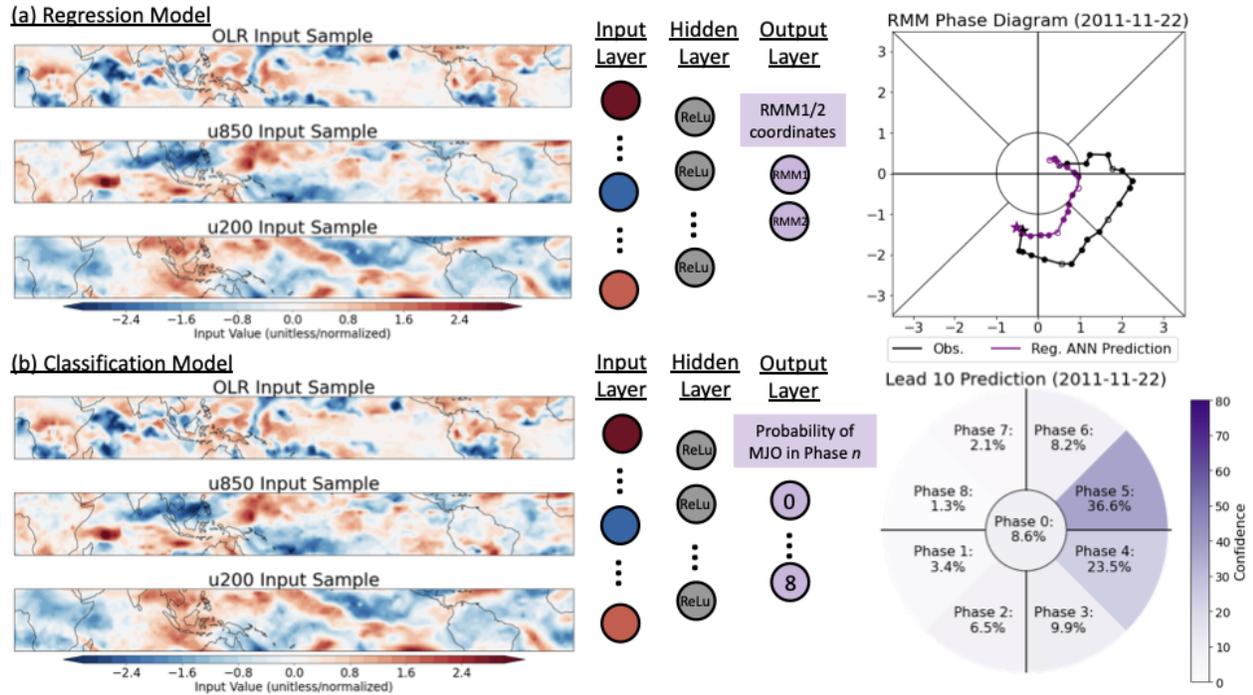


Figure 1. ANN model schematics. (a) The regression ANN; leftmost panels show a sample input of OLR and zonal wind at 850 hPa (u850) and 200 hPa (u200) from November 22, 2011. The input is passed through a 16-node hidden layer with a rectified linear unit (“ReLU”) activation function. The regression ANN outputs values of RMM1 and RMM2 at a single lead time, and separate ANNs are trained for leads from 0-20 days. An example 20-day ANN forecast (purple) versus observations (black) is shown in the rightmost panel. (b) The classification model schematic. Input is identical to the regression model, but the output is the probability the MJO is active in RMM Phase 1-8 or is inactive (“Phase 0”). An example forecast at a 10-day lead from November 22, 2011 is shown on the right.

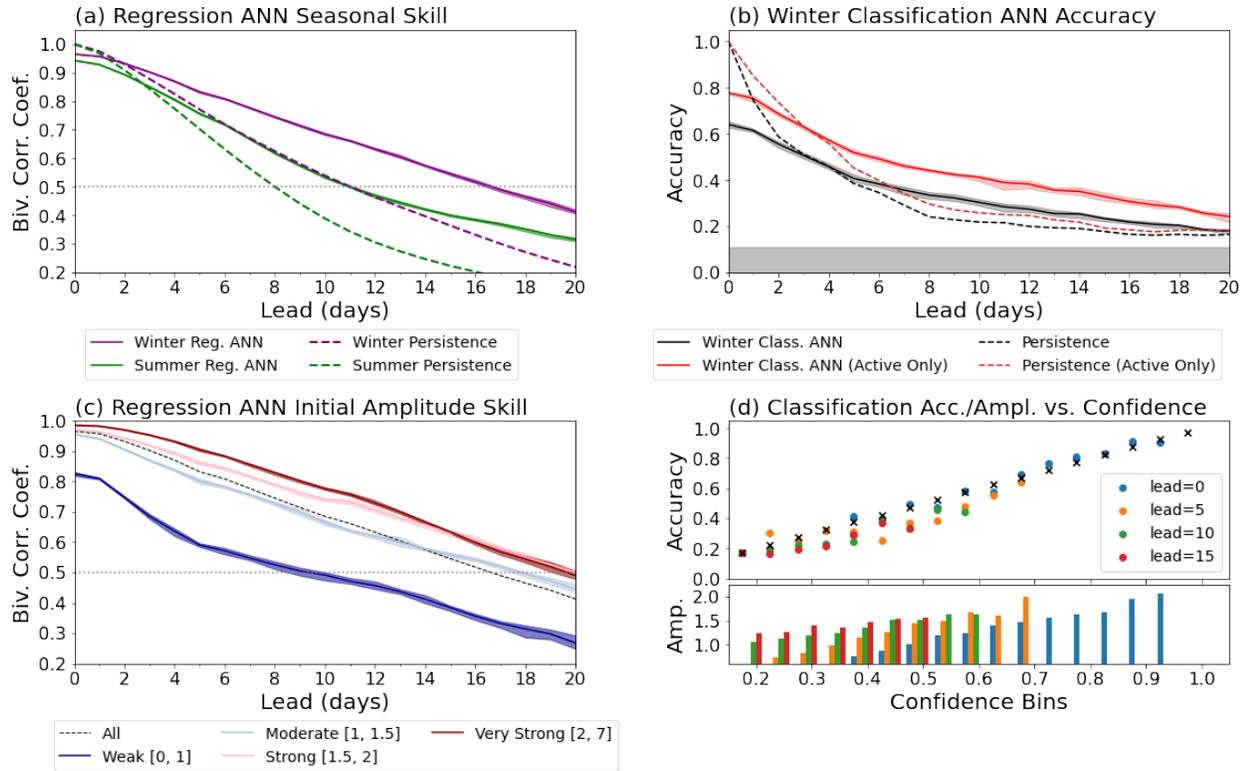


Figure 2. Regression and classification ANN performance. (a) Regression ANN and persistence model MJO prediction skill during winter (purple) and summer (green). A bivariate correlation coefficient (BCC) threshold of 0.5 denotes skill (dotted). (b) The classification ANN and persistence model accuracy during winter: black curves are accuracy across all days, and red curves are accuracy predicting only active MJO days. Shading indicates less than random chance (~11%). (c) As in panel (a), but for winter forecasts binned by observed initial MJO amplitude via the RMM amplitude ranges in the legend. (d) Winter classification ANN accuracy (top panel) and initial observed MJO amplitude (bottom panel) binned by ANN confidence (x -axis, in bins of width 0.05) at leads of 0, 5, 10, and 15 days. The black x 's indicate the one-to-one line.

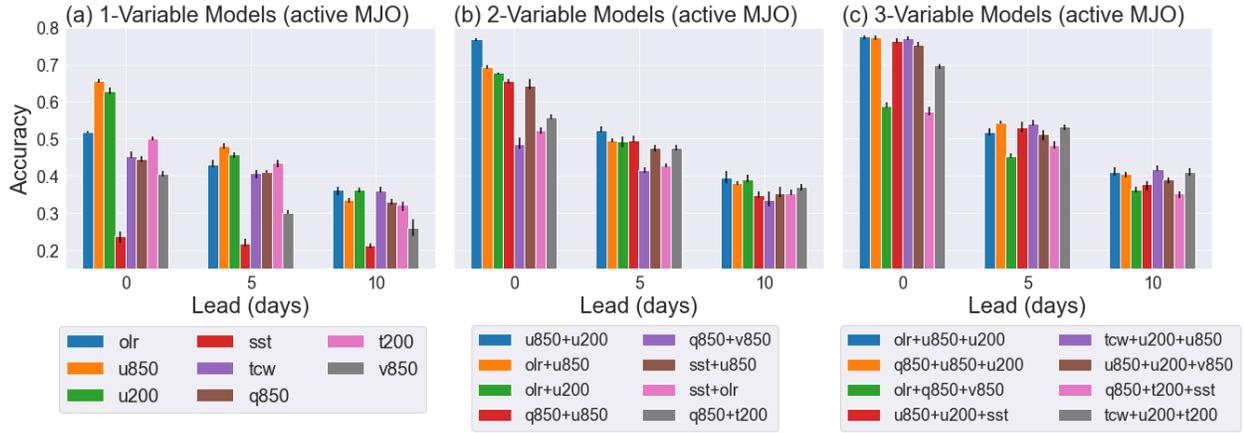


Figure 3. Sensitivity to input variables. Winter classification ANN accuracy predicting active MJO days at leads of 0, 5, and 10 days given different input variables. 1-variable (panel a), 2-variable (panel b), and 3-variable (panel c) models are shown. For each model, 5 ANNs are trained with different initial random weights (error lines). The legend short-hand refers to zonal wind (u), total column water vapor (tcw), specific humidity (q), temperature (t), and meridional wind (v), with numbers indicating the pressure level where relevant.

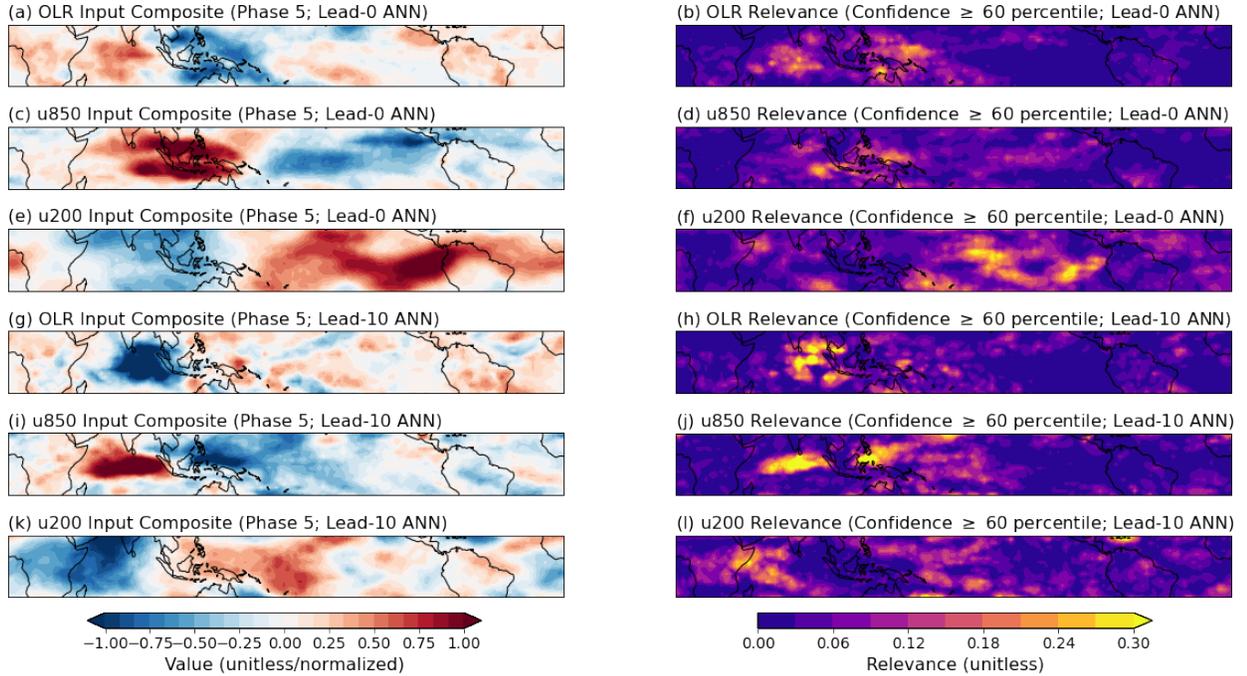


Figure 4. Layer-wise relevance propagation example. Composites of

normalized input variables (left column) and LRP relevance (right column) for correct classification ANN predictions of MJO events in Phase 5 at the time of verification. Only forecasts when model confidence exceeds the 60th percentile are included. Panels (a-f) are the lead-0 model, and (g-l) are the lead-10 model, both inputting 3 variables (OLR, and 850 and 200 hPa zonal wind).