Subseasonal Forecasts of Opportunity Identified by an Interpretable Neural Network

Kirsten J. Mayer [1] **and Elizabeth A. Barnes** [1]

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

Key Points:

- Neural networks can be used to identify forecasts of opportunity for subseasonal prediction
- Neural network interpretability techniques pinpoint relevant tropical regions for predictions over the North Atlantic
- Clustering of neural network relevance heat maps reveals a potential new forecast of opportunity for the North Atlantic

Corresponding author: Kirsten J. Mayer, kjmayer@rams.colostate.edu

Abstract

Midlatitude prediction on subseasonal timescales is difficult due to the chaotic nature of the atmosphere and often requires the identification of favorable atmospheric conditions that may lead to enhanced skill ("forecasts of opportunity"). Here, we demonstrate that an artificial neural network can identify such opportunities for tropical-extratropical teleconnections to the North Atlantic circulation at a lead of 22 days using the network's confidence in a given prediction. Furthermore, layer-wise relevance propagation, an ANN interpretability technique, pinpoints the relevant tropical features the ANN uses to make accurate predictions. We find that layer-wise relevance propagation identifies tropical hot spots that correspond to known favorable regions for midlatitude teleconnections and reveals a potential new pattern for prediction over the North Atlantic on subseasonal timescales.

Plain Language Summary

Weather forecasting on 2 week to 2 month timescales is known for its lack of predictability due to the chaotic nature of the atmosphere. One way to improve prediction skill on these timescales involves the identification of periods of atmospheric conditions that lead to enhanced predictability ("forecasts of opportunities"). Here, we show that a neural network can accurately identify these opportunities when trying to predict the atmospheric circulation over the North Atlantic Ocean 4 weeks in advance. A neural network interpretability technique is then used to uncover what the network has "learned" to make these accurate predictions. We show that the network identifies known patterns of storminess ideal for midlatitude prediction and uncovers a possible new favorable region for enhanced prediction.

1 Introduction

Subseasonal timescales (2 weeks - 2 months) are known for their lack of predictability (Mariotti et al., 2018), yet reliable and actionable information on these timescales are required for decision making in many sectors such as public health and water management (e.g. Vitart et al., 2012; White et al., 2017). Over the past decade, there has been a substantial research effort to improve prediction on these timescales (e.g. Vitart et al., 2012; Robertson et al., 2015; Vitart et al., 2017; Pegion et al., 2019). One area of subseasonal prediction research focuses on forecasts of opportunity, the idea that certain earth system conditions provide opportunities for enhanced subseasonal prediction skill (Mariotti et al. 2020). When these opportunities arise, the information provided by the earth system's state can then be leveraged to improve forecast skill. For example, when the Madden-Julian Oscillation (MJO; Madden and Julian (1971, 1972)), a propagating tropical convective phenomenon, is active, its convective heating can lead to the excitation of quasi-stationary rossby waves (Hoskins and Ambrizzi 1993) that subsequently modulate the midlatitude circulation over the first few weeks following MJO activity (e.g., Hoskins and Karoly, 1981; Sardeshmukh and Hoskins, 1988; Henderson et al., 2016; Baggett et al., 2017; Zheng et al., 2018). When opposing convective anomalies are located over the Indian Ocean and western Pacific (defined as phases 2, 3, 6, and 7), the MJO has been shown to lead to more coherent and consistent modulations of midlatitude weather on subseasonal timescales and consequently, enhanced prediction skill (Tseng et al., 2018). The MJO is an example of forecast of opportunity identification that uses the strength and location of tropical activity to identify periods of enhanced midlatitude prediction skill.

Albers and Newman (2019) demonstrate a technique for forecast of opportunity identification through the utilization of expected skill from a linear inverse model. The study demonstrates the ability of the linear statistical model to identify forecasts of opportunity, and raises the question of whether other statistical models, such as artificial neural networks (ANNs), can identify forecasts of opportunity for subseasonal predic-

tion. ANNs are very good at nonlinear function estimation (Chen & Chen, 1995), and thus, may be able to identify both linear and nonlinear relationships that lend predictability. Recently, ANNs have been successfully applied to climate prediction such as surface temperature (Toms et al., 2019) and the El Nino Southern Oscillation (Ham et al., 2019), suggesting ANNs may be useful for indentifying subseasonal forecasts of opportunity as well.

To test the efficacy of an ANN for subseasonal forecast of opportunity identification, here we input outgoing longwave radiation (OLR) anomalies into an ANN and task the network to predict the sign of 500 hPa geopotential height (z500) anomalies over the North Atlantic 22 days later (e.g. Week 4). OLR is used to explore the ability of an ANN to identify known relationships between the MJO and the North Atlantic via tropical-extratropical teleconnections (e.g. Cassou, 2008; Henderson et al., 2016). We demonstrate that an ANN can identify subseasonal forecasts of opportunity related to tropical OLR, and through an ANN interpretability technique, can further identify relevant features for these enhanced subseasonal predictions. Specifically, we demonstrate the ability of an ANN to identify known MJO-like OLR patterns as well as a possible new tropical OLR pattern associated with predictable behavior of the North Atlantic circulation on subseasonal timescales.

2 Data and Methods

2.1 Data

We use daily mean OLR (1980-2019) from the National Center for Atmospheric Research/National Oceanic and Atmospheric Administration (NCAR/NOAA; Liebmann and Smith (1996)) and daily mean z500 (1980-2019) from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim reanalysis (ERA-I; Dee et al. (2011)). MJO teleconnections tend to be stronger during boreal winter (Madden, 1986), and therefore, the extended boreal winter months (November-February) are used for the OLR fields. Since we task the network to predict the sign of the z500 anomaly 22 days following a given OLR field, March is also included in the z500 analysis.

The annual cycle is removed from both the z500 and OLR data. For z500, the annual cycle is removed by subtracting the daily climatology over the record (1979-2019). A Fast Fourier Transform high-pass filter is then applied to the anomalies to remove seasonal oscillations (frequencies smaller than $\frac{1}{120 days}$) to ensure the network focuses on subseasonal anomalies. The median of the z500 anomalies for the training data (see 2.2.1) is then subtracted to obtain an equal number of positive and negative values. These anomalies are then converted into 0s and 1s depending on the sign (negative or positive, respectively). For OLR, the annual cycle is removed by subtracting the first 3 harmonics of the daily climatology from the raw field. The first 3 harmonics are used instead of the daily mean because OLR is a noisier field than z500.

2.2 Methods

2.2.1 Artificial Neural Network Architecture

A two layer ANN (Figure 1) is tasked to ingest tropical OLR and predict the *sign* of the z500 anomaly over the North Atlantic (40°N, 325°E; red dot in Figure 1) 22 days later. The North Atlantic is chosen for this analysis since the MJO is known to force circulation anomalies over this region on subseasonal timescales and thus allows us to explore the utility of an ANN in the context of a well known problem (e.g. Cassou, 2008; Roundy et al., 2010; Henderson et al., 2016). Each input sample to the ANN consists of vectorized daily anomalous OLR from 30°N to 20°S and 45 to 210°E, where the number of input nodes is equal to the number of OLR grid points (N = 1407). The ANN then outputs two values that describe the categorical prediction, positive or negative sign of z500, given the initial OLR input image. The softmax activation function is applied to

this final layer and transforms the two output values such that they sum to 1. The output then represents an estimation of the likelihood that an input belongs to a particular category. We refer to this estimation of likelihood as "model confidence". A more confident prediction will, therefore, have a predicted category value closer to 1. We define forecasts of opportunities as the top 10% most confident predictions by the network, although we explore alternative percentages as well.

The ANN architecture consists of two hidden layers of 128 and 8 nodes, respectively, and both use the rectified linear activation function. The final layer includes 2 nodes and uses the softmax activation function. The batch size is set to 256 samples (i.e. OLR vectorized images) and the ANN is trained for 50 epochs unless the validation loss increases for two epochs in a row. If this occurs, the ANN stops training early and restores the model's best weights to reduce overfitting. It is found that 50 epochs is sufficient for training as the ANN rarely completes all 50 epochs.

The data used to train and test the ANN is composed of three groups: training, validation, and testing. Training and validation data are used during training, where training data is used to update the weights and biases of the ANN and the validation data is used to evaluate the model. The testing data is data that has never been "seen" by the ANN to evaluate the ability of the ANN to generalize to new data. For this analysis, the ANN validation data is from November 2007 through February 2011 (N = 481) and the testing data is from November 2017 through February 2019 (N = 240). The remaining extended boreal winter (NDJF) data are used for training (November 1979 - February 2007 and November 2011 - February 2016; N = 4450; see Figure 3). To choose a model for the following analysis, ANN training is repeated for a variety of validation years. Different four year chucks are removed from the training data and set aside to use as validation. For each four year chunk, the ANN was trained 20 times with random initialized weights. We find that our conclusions remain the same when a different trained ANN is used. We present one model with reasonably high accuracy here and using the training, validation, and testing groups outlined above.

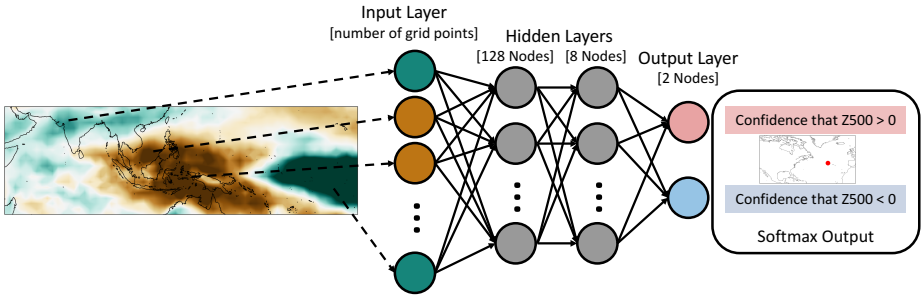

Figure 1. Artificial neural network architecture for prediction of the sign of z500 anomalies over the North Atlantic 22 days following tropical OLR anomalies. The neural network consists of two hidden layers of 128 and 8 nodes, respectively, and an output layer of two nodes (one node for each sign). The output layer uses the softmax activation function.

2.2.2 Layer-Wise Relevance Propagation (LRP)

While ANNs are a useful tool for making predictions, in doing so, they are learning *how* to make accurate predictions. Therefore, understanding the inner workings of an ANN can provide valuable information for improving prediction skill and understanding, as well as increasing user confidence in the results. Here, we utilize a relatively new

neural network interpretability technique to the geosciences called layer-wise relevance propagation (LRP; Montavon et al. (2019); Bach et al. (2015)) to extract and visualize the features the ANN learns to make accurate predictions. While Toms et al. (2020) describe the use of LRP for geoscience applications in detail, we briefly provide a high-level description here. After network training is completed, a single sample is passed through the network and a prediction is made (in our case, two output values are predicted). Our implementation of LRP then takes the highest of these values (i.e. the winning category) and back-propagates this value through the network via a series of predefined rules, ultimately distributing it across the input nodes (i.e. input gridpoints). What results is a heat map of "relevance" across the input space, where input nodes that are more relevant for the network's specific prediction for that sample are given higher relevance. This process is then repeated for every sample of interest, resulting in a unique relevance heat map for each sample. In our study, since the input layer consists of maps of OLR anomalies, the LRP heat maps are maps of the relevant tropical OLR patterns for each prediction of the circulation over the North Atlantic. These maps are discussed in detail in Section 3.2.

3 Results

3.1 Identifying Forecasts of Opportunity

ANNs with the architecture shown in Figure 1 are trained 100 times with random initialized weights to predict the sign of the z500 anomalies 22 days following the tropical OLR anomalies. Figure 2a shows the distribution of the testing prediction accuracy for all 100 models, where dark teal represents the distribution of all predictions and light teal represents the distribution of the 10% most confident predictions. The corresponding colored vertical dashed lines indicate a threshold for what is expected by random chance. To calculate the random chance accuracy threshold, 100,000 randomly generated groups (N=240 for all and N=24 for 10% most confident predictions) of zeros and ones are used to create a distribution of accuracies, and the 90^{th} percentile of this distribution is used as the random chance threshold. In Figure 2a, the top 10% most confident prediction accuracies (light teal) are shifted towards higher accuracies compared to the distribution with all predictions (dark teal). This shift in the distributions demonstrates that in general, higher model confidence leads to substantially enhanced prediction accuracy.

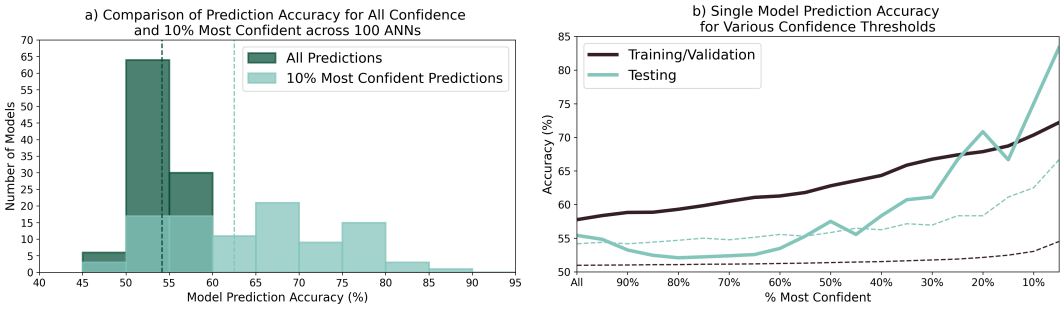

Figure 2. (a) Histograms of testing prediction accuracy for 100 trained ANNs. The dark teal represents the histogram of all prediction accuracies and the light teal represents the histogram for the 10% most confident prediction accuracies. The dark teal and light teal dashed lines in (a) are the maximum accuracies expected by random chance at the 90% confidence level for the corresponding colored histogram (see text for details). (b) Accuracy of one particular model as a function of the percent most confident predictions for training and validation (black) and testing (light teal) data. The dashed lines indicate the maximum accuracies expected by random chance at the 90% confidence level for the corresponding colored lines (see text for details).

We chose one model from Figure 2a to further understand how accuracy varies when a different percent model confidence is used (Figure 2b). The solid lines represent the accuracy across various model confidence values for training and validation (black) and testing (light teal) data sets. Together, Figure 2a and b illustrate that model confidence and prediction accuracy generally increase together and therefore, can be used to identify forecasts of opportunities, or periods of enhanced prediction skill. From this analysis, the 10% most confident predictions are chosen to define FOOs since this threshold has one of the largest accuracy differences from random chance while still retaining 10% of the samples.

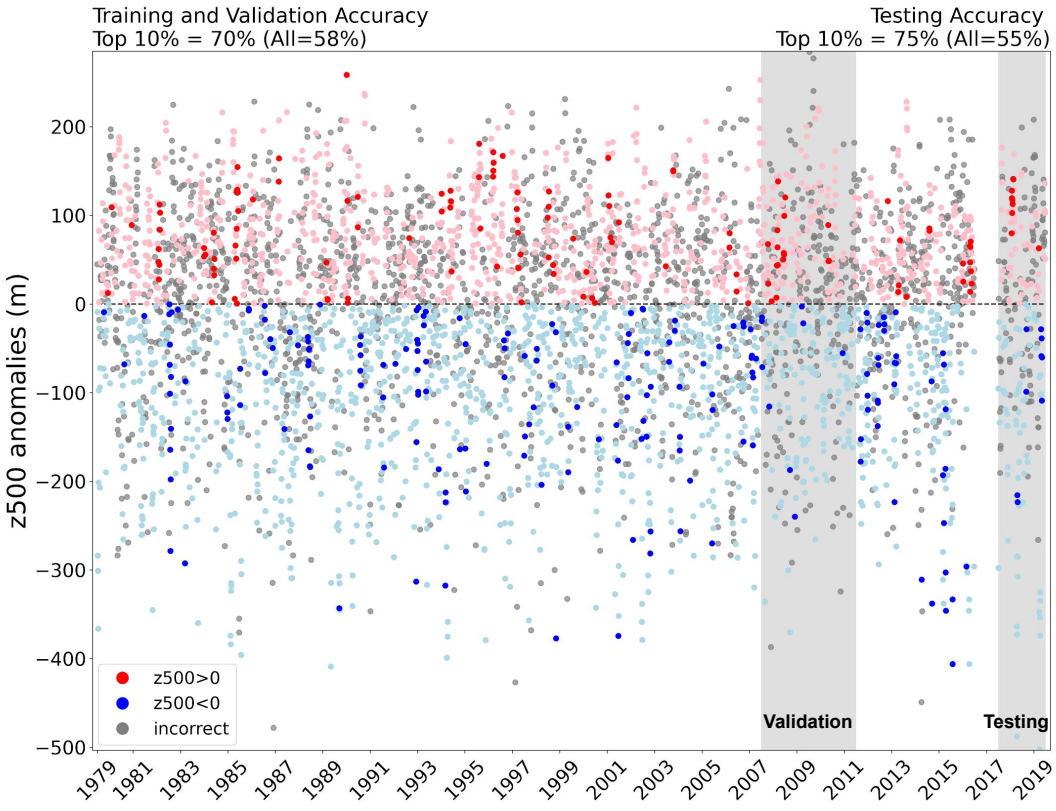

Figure 3. Timeseries of z500 anomalies shaded by the sign of the ANN predictions. Blue dots represent correct negative predictions, red dots represent correct positive predictions, and dark colored dots indicate forecasts of opportunities (i.e. 10% most confident predictions). Grey dots represent incorrect predictions. The vertical grey shading from 2007-2011 highlights the time period used for validation and the vertical grey shading from 2017-2019 highlights the time period used for testing. The accuracies for training and validation as well as testing data for forecasts of opportunities and all predictions are given in the top left and right, respectively.

Continuing with the trained ANN from Figure 2b, Figure 3 shows the predictions made by the ANN as a function of time. When evaluating the network with the training and validation data, the prediction accuracy for all predictions is 58% and for the top 10% most confident predictions is 70%. For the testing data, the prediction accuracy for all predictions is 55% and for the top 10% most confident predictions is 75%. In addition, Figure 3 shows that the forecast of opportunities (dark colored dots) are not clustered over specific time periods or years, and generally do not favor a particular sign. Furthermore, the ANN not only accurately predicts large anomalies but can also pre-

dict days with smaller anomalies. This figure, therefore, illustrates that the ANN is successfully identifying forecasts of opportunities over the record.

3.2 Tropical Sources of Predictability

We have shown that ANNs can identify forecasts of opportunities using model confidence; however, understanding where this enhanced skill originates is critical for improving physical understanding as well as gaining trust in the network's predictions. To do so, layer-wise relevance propagation is used to identify the OLR patterns that lead the ANN to make correct predictions (see Section 2.2.2). The shading in Figure 4 shows the regions the network found relevant, on average, to make confident and correct positive (Figure 4a-d) or negative (Figure 4e-h) z500 predictions. The contours correspond to the average OLR anomalies for these confident and correct predictions. The correct 10% most confident predictions from the training, validation and testing data sets are combined for this LRP analysis. All three sets of data are used instead of only testing data because all data sets have similar accuracies and LRP values (not shown). Thus, including all the data increases the sample sizes for the analysis.

The average LRP heat map for the correct forecasts of opportunity of positive sign predictions (Figure 4a) indicates two hot spots, one over the southern Indian Ocean into the southern Maritime Continent and the other over the western Pacific. The average LRP heat map for the correct forecasts of opportunity of negative sign predictions (Figure 4b) indicates five hot spots, one over the southern Maritime Continent, another over the western Pacific Ocean, the third into the central Pacific and the fourth and fifth over Vietnam and into the South China Sea as well as Saudi Arabia. For both sign predictions, the hot spots near the southern Maritime Continent and over the western Pacific have opposing signed OLR anomalies (contours) and straddle 140°E. These dipoles of convection over the Indian Ocean through the Maritime Continent and over the western Pacific have similar structures to phase 4-5 and phase 1,7-8 of the MJO (Wheeler & Hendon, 2004). This structure of OLR is consistent with previous research of MJO teleconnections over the North Atlantic for average lead times of 10-14 and 15-19 days (e.g. Henderson et al., 2016; Henderson & Maloney, 2018; Tseng et al., 2018). In addition, this dipole structure is known to lead to higher pattern consistency of teleconnections in the midlatitudes (Tseng et al., 2019), which has been shown to lead to enhanced prediction skill (Tseng et al., 2018). Rossby waves initiated by the MJO tend to be quasi-stationary, which suggests that these OLR anomalies may also correspond to 22 day leads as well. This Maritime Continent and western Pacific Ocean dipole highlighted by LRP is therefore consistent with previous research and demonstrates that the ANN has learned physically relevant structures.

Negative sign predictions also have three more hot spots that are not seen for the positive sign predictions. In the next section, it is shown that the central Pacific hot spot is just an eastward shifted version of the one in the western Pacific and therefore, is also likely associated with the MJO. In addition, the hot spot over Saudi Arabia is hypothesized to be associated with the two-way relationship between the North Atlantic Oscillation (NAO) and the MJO. On the other hand, the hot spot over Vietnam and the South China Sea is discussed as a possible new region relevant for enhanced subseasonal prediction.

3.2.1 K-means Clustering of LRP Maps

To further distinguish the relevant regions for the ANN's predictions, k-means clustering (Hartigan & Wong, 1979) is applied to the LRP maps (Figure 4b-d, f-h). This analysis reveals that the composite LRP maps for each sign (Figure 4a,e) actually consist of multiple distinct patterns used by the ANN. For positive sign predictions (Figure 4b,c,d), each cluster is associated with a hot spot located between the central Indian ocean and the maritime continent. Cluster 1 and 2 (Figure 4b,c), have hot spots around the east-

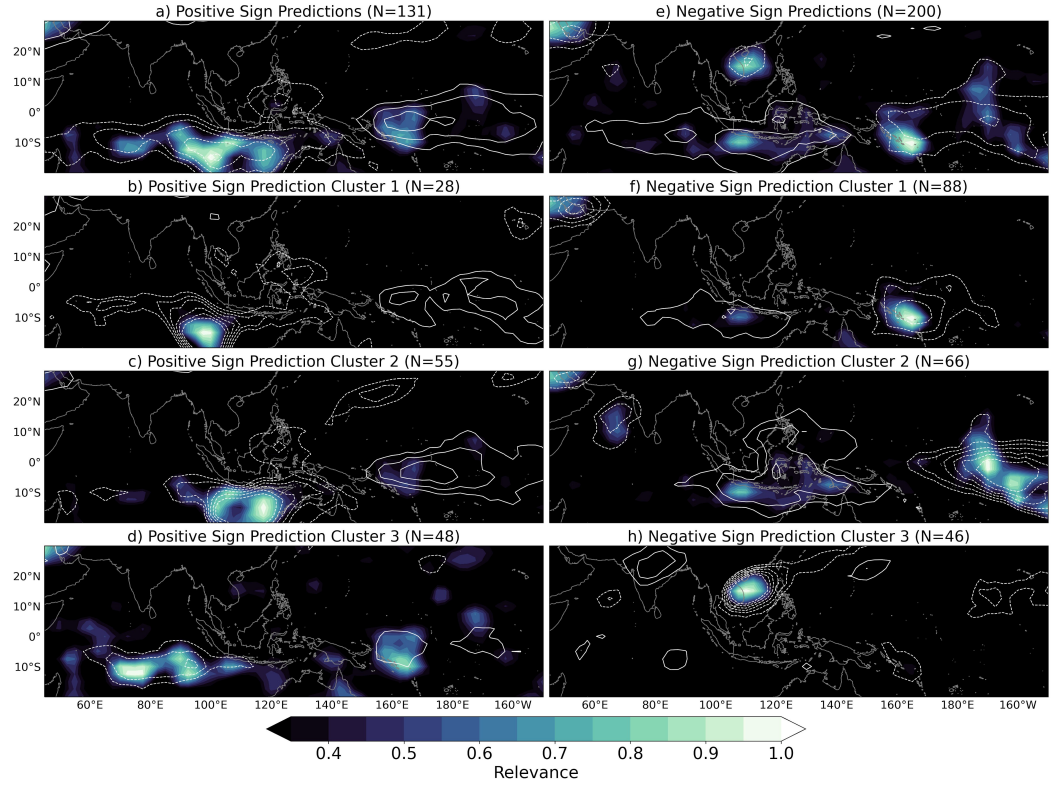

Figure 4. Shading denotes smoothed composites of layer-wise relevance propagation fields for all correct forecasts of opportunities for (a) positive sign and (e) negative sign predictions across training, validation and testing periods. The associated three k-means clusters of LRP for (b-d) positive sign predictions and (f-h) negative sign predictions are also shown. Contours represent the corresponding smoothed OLR anomalies where solid lines are positive values ($0.3-1.5\frac{W}{m^2}$) and dashed lines are negative values (-1.5− -0.3$\frac{W}{m^2}$) and the contour interval is $0.2\frac{W}{m^2}$.

ern and central maritime continent which are both associated with negative OLR anomalies. While not highlighted by LRP, each negative OLR anomaly region is accompanied by a region of positive OLR anomalies over the western Pacific. This suggests the model is identifying an MJO-like pattern, but focuses exclusively on the enhanced convection region. This is consistent with recent research that suggests that convection over the Indian Ocean dominates the formation of a positive NAO anomaly (Shao et al., 2020). On the other hand, cluster 3 (Figure 4d) has two main hot spots that highlight a convective dipole. One is located over the south-central Indian ocean and associated with a negative OLR anomaly, and the other is located in the western Pacific and is associated with a positive OLR anomaly. As previously mentioned, these regions lead to more consistent midlatitude teleconnections (Tseng et al., 2018), and therefore, this cluster supports previously identified tropical OLR regions and patterns ideal for enhanced prediction skill on subseasonal timescales.

For cluster 1 and 2 of the negative sign predictions (Figure 4f,g), there are two hot spots, one east of the Maritime Continent and the other over the Pacific Ocean, where cluster 2 is shifted slightly east of cluster 1. As with the positive sign predictions, each hot spot is associated with negative OLR anomalies, except the regions of positive OLR anomalies are also weakly highlighted by the LRP analysis. Even so, the LRP analysis mainly highlights the western Pacific region, and suggests that the network largely finds

the regions of enhanced convection most relevant. Unexpectedly, there is also a hot spot located over Saudi Arabia in both cluster 1 and 2. This hot spot appears to only be important when this MJO-like dipole structure is present. To the authors' knowledge, this region has not been shown to be important for tropical-extratropical teleconnections to the North Atlantic. However, previous research has shown that there is a two-way relationship between the MJO and NAO. Following the NAO, there tends to be a significant modulation of the tropical upper troposphere zonal wind over the Atlantic-Africa region (Lin et al., 2009). This modulation has been hypothesized to play a role in MJO initialization (Lin et al., 2009; Lin & Brunet, 2011). Since the NAO can persist over many weeks, the network may be identifying an influence of the NAO on the MJO and back to the NAO. We leave a deeper exploration of this possible mechanism to future work.

Unlike the other clusters, the third cluster (Figure 4h) has one hot spot over Vietnam and the South China Sea. We hypothesize that this region is physically important as it is located south of the subtropical jet entrance region and is associated with large negative OLR anomalies. Rossby waves can be generated through advection of vorticity by upper level divergence associated with negative OLR anomalies (Sardeshmukh & Hoskins, 1988), and since this hot spot region is close to the entrance region of the jet, these waves can more easily become trapped within the jet and directed east before exiting (Hoskins & Karoly, 1981; Hoskins & Ambrizzi, 1993). Based on these known tropical-extratropical teleconnection dynamics, it is likely that this hot spot over Vietnam and the South China Sea is a new pattern identified by the ANN.

4 Conclusions

Improving subseasonal prediction accuracy and understanding requires identifying opportunities that can lead to enhanced predictability (e.g. Mariotti et al., 2020). Here, we show that an artificial neural network can identify forecasts of opportunity for subseasonal prediction using the network's confidence in its prediction. In addition, we demonstrate that layer-wise relevance propagation can extract knowledge gained by the ANN to identify relevant physical tropical features important for the predictions. K-means clustering of the LRP maps further provides insight into multiple distinct patterns used by the ANN for enhanced prediction and reveals a possible new forecast of opportunity for prediction over the North Atlantic.

The hot spots identified by the ANN provide a stepping stone to further our understanding of tropical-extratropical teleconnections. For example, lagged composite analysis or simplified models can be used to further explore the physical mechanisms behind enhanced midlatitude predictability associated with these regions. In addition, analysis of the incorrect predictions made by the ANN may also be useful for improving our understanding of ideal tropical patterns for enhanced subseasonal prediction. Finally, while our application is focused on subseasonal prediction, the approach outlined here should be applicable to predictions across timescales. Ultimately, this paper demonstrates that ANNs are not only a useful tool for prediction, but can also be used to gain physical insight into predictability and subsequently, improve prediction skill.

Acknowledgments

This research is partially funded by the National Science Foundation Graduate Research Fellowship under the grant number 006784 supporting Kirsten J. Mayer and partially funded by the National Science Foundation Harnessing the Data Revolution through supporting Elizabeth A. Barnes with grant 1934668.

The authors declare that they have no conflict of interest.

Data availability: ERA-I reanalysis data are provided by the European Centre for Medium-Range Forecasts (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/ era-

interim; Dee et al., 2011). The interpolated OLR data is provided by the NOAA/OAR/ESRL PSL, Boulder, CO, USA (https://psl.noaa.gov/data/gridded/data.interp_OLR.html; Liebmann and Smith, 1996).

References

Albers, J. R., & Newman, M. (2019, November). A priori identification of skillful extratropical subseasonal forecasts. *Geophys. Res. Lett.*, *46*(21), 12527–12536.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015, July). On Pixel-Wise explanations for Non-Linear classifier decisions by Layer-Wise relevance propagation. *PLoS One*, *10*(7), e0130140.

Cassou, C. (2008, September). Intraseasonal interaction between the Madden-Julian oscillation and the north atlantic oscillation. *Nature*, *455*(7212), 523–527.

Chen, T., & Chen, H. (1995). Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Trans. Neural Netw.*, *6*(4), 911–917.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Others (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, *137*(656), 553–597.

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019, September). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.*, *28*(1), 100–108.

Henderson, S. A., & Maloney, E. D. (2018, July). The impact of the Madden–Julian oscillation on High-Latitude winter blocking during el niño–southern oscillation events. *J. Clim.*, *31*(13), 5293–5318.

Henderson, S. A., Maloney, E. D., & Barnes, E. A. (2016, June). The influence of the Madden–Julian oscillation on northern hemisphere winter blocking. *J. Clim.*, *29*(12), 4597–4616.

Hoskins, B. J., & Ambrizzi, T. (1993, June). Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci.*, *50*(12), 1661–1671.

Hoskins, B. J., & Karoly, D. J. (1981, June). The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, *38*(6), 1179–1196.

Liebmann, B., & Smith, C. (1996). Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Am. Meteorol. Soc.*, *77*, 1275–1277.

Lin, H., & Brunet, G. (2011, January). Impact of the north atlantic oscillation on the forecast skill of the Madden-Julian oscillation: IMPACT OF NAO ON MJO FORECAST. *Geophys. Res. Lett.*, *38*(2).

Lin, H., Brunet, G., & Derome, J. (2009, January). An observed connection between the north atlantic oscillation and the Madden–Julian oscillation. *J. Clim.*, *22*(2), 364–380.

Madden, R. A. (1986). Seasonal variations of the 40-50 day oscillation in the tropics. *J. Atmos. Sci.*, *43*(24), 3138–3158.

Madden, R. A., & Julian, P. R. (1971, July). Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *J. Atmos. Sci.*, *28*(5), 702–708.

Madden, R. A., & Julian, P. R. (1972, September). Description of Global-Scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, *29*(6), 1109–1123.

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., ... Albers, J. (2020, January). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Am. Meteorol. Soc.*.

Mariotti, A., Ruti, P. M., & Rixen, M. (2018, March). Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate and Atmospheric Science*, *1*(1), 1–4.

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K.-R. (2019). Layer-Wise relevance propagation: An overview. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable AI: Interpreting, explaining and visualizing deep learning* (pp. 193–209). Cham: Springer International Publishing.

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., ... Kim, H. (2019, October). The subseasonal experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Am. Meteorol. Soc.*, *100*(10), 2043–2060.

Robertson, A. W., Kumar, A., Peña, M., & Vitart, F. (2015, March). Improving and promoting subseasonal to seasonal prediction. *Bull. Am. Meteorol. Soc.*, *96*(3), ES49–ES53.

Roundy, P. E., MacRitchie, K., Asuma, J., & Melino, T. (2010, August). Modulation of the global atmospheric circulation by combined activity in the Madden–Julian oscillation and the el niño–southern oscillation during boreal winter. *J. Clim.*, *23*(15), 4045–4059.

Sardeshmukh, P. D., & Hoskins, B. J. (1988, April). The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, *45*(7), 1228–1251.

Shao, X., Straus, D. M., Li, S., Swenson, E., Yadav, P., & Song, J. (2020). Forcing of the MJO-related indian ocean heating on the intraseasonal lagged NAO.

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2019, December). Physically interpretable neural networks for the geosciences: Applications to earth system variability.

Toms, B. A., Barnes, E. A., Maloney, E. D., & Heever, S. C. (2020, April). The global teleconnection signature of the Madden-Julian oscillation and its modulation by the Quasi-Biennial oscillation. *J. Geophys. Res. D: Atmos.*, *125*(7), 1.

Tseng, K.-C., Barnes, E. A., & Maloney, E. D. (2018, January). Prediction of the midlatitude response to strong Madden-Julian oscillation events on S2S time scales: PREDICTION OF Z500 AT S2S TIME SCALES. *Geophys. Res. Lett.*, *45*(1), 463–470.

Tseng, K.-C., Maloney, E., & Barnes, E. (2019, January). The consistency of MJO teleconnection patterns: An explanation using linear rossby wave theory. *J. Clim.*, *32*(2), 531–548.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., ... Zhang, L. (2017, January). The subseasonal to seasonal (S2S) prediction project database. *Bull. Am. Meteorol. Soc.*, *98*(1), 163–173.

Vitart, F., Robertson, A. W., & Anderson, D. L. T. (2012, January). Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *WMO Bull.*, *61*(61).

Wheeler, M. C., & Hendon, H. H. (2004, August). An All-Season Real-Time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Weather Rev.*, *132*(8), 1917–1932.

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., ... Zebiak, S. E. (2017, July). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Met. Apps*, *24*(3), 315–325.