Quantifying the Effect of Climate Change on Midlatitude Subseasonal Prediction Skill Provided by the Tropics

Kirsten J Mayer[1] and Elizabeth A. Barnes[2]

[1]Colorado State University - Department of Atmospheric Science
[2]Colorado State University

November 30, 2022

Abstract

Subseasonal timescales (~2 weeks - 2 months) are known for their lack of predictability, however, specific Earth system states known to have a strong influence on these timescales can be harnessed to improve prediction skill (known as "forecasts of opportunity"). As the climate continues warming, it is hypothesized these states may change and consequently, their importance for subseasonal prediction may also be impacted. Here, we examine changes to midlatitude subseasonal prediction skill provided by the tropics under anthropogenic warming using artificial neural networks to quantify skill. The network is tasked to predict the sign of the 500hPa geopotential height for historical and future time periods in the CESM2-LE across the Northern Hemisphere at a 4 week lead using tropical precipitation. We show prediction skill changes substantially in key midlatitude regions and these changes appear linked to changes in seasonal variability with the largest differences in accuracy occurring during forecasts of opportunity.

Quantifying the Effect of Climate Change on Midlatitude Subseasonal Prediction Skill Provided by the Tropics

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 evaluate subseasonal predictability under future climate change scenarios
- In CESM2-LE, largest differences in subseasonal predictability provided by the tropics mainly occur during forecasts of opportunity
- Changes in Northern Hemisphere subseasonal prediction skill appears mainly linked to changes to seasonal variability in CESM2-LE

Corresponding author: Kirsten J. Mayer, kirsten.j.mayer@gmail.com

Abstract

Subseasonal timescales (∼2 weeks - 2 months) are known for their lack of predictability, however, specific Earth system states known to have a strong influence on these timescales can be harnessed to improve prediction skill (known as "forecasts of opportunity"). As the climate continues warming, it is hypothesized these states may change and consequently, their importance for subseasonal prediction may also be impacted. Here, we examine changes to midlatitude subseasonal prediction skill provided by the tropics under anthropogenic warming using artificial neural networks to quantify skill. The network is tasked to predict the sign of the 500hPa geopotential height for historical and future time periods in the CESM2-LE across the Northern Hemisphere at a 4 week lead using tropical precipitation. We show prediction skill changes substantially in key midlatitude regions and these changes appear linked to changes in seasonal variability with the largest differences in accuracy occurring during forecasts of opportunity.

Plain Language Summary

Predictions on 2 week to 2 month (subseasonal) timescales are important for the public and private sectors due to the increased preparation time provided to save lives and property. In the current climate, signals initiated in the tropics can overpower noise in the midlatitudes and ultimately lead to enhanced midlatitude subseasonal prediction skill. However, it has been hypothesized that increasing global temperatures due to climate change may impact these signals and their sources in the future. Therefore, it is important to understand how subseasonal predictability provided by the tropics will be affected. Here, we utilize a type of machine learning known as a neural network to investigate this question. We find that subseasonal prediction skill provided by the tropics changes throughout the Northern Hemisphere in a warmer climate and these changes appear mainly linked to changes in seasonal variability. In addition, we see that the largest differences in accuracy occur during opportunities for enhanced subseasonal prediction skill.

1 Introduction

Accurate predictions on subseasonal timescales (2 weeks - 2 months) are important for many public and private sectors such as water management and agriculture (White et al., 2021). This is because prediction on these timescales provides pivotal lead times for saving lives and property in these sectors (White et al., 2021). The tropics is of particular importance for this timescale because of intraseasonal phenomena like the Madden-Julian Oscillation (MJO; Madden & Julian, 1971, 1972). Quasi-stationary Rossby waves generated by upper level divergence associated with MJO convection (Hoskins & Ambrizzi, 1993) can modulate midlatitude circulation in the following weeks (e.g. Hoskins & Karoly, 1981; Sardeshmukh & Hoskins, 1988; Henderson et al., 2016; Baggett et al., 2017; Zheng et al., 2018) and these tropical-extratropical teleconnections are known to lead to enhanced midlatitude prediction skill on subseasonal lead times (Tseng et al., 2018). Phenomena like the El Niño Southern Oscillation (ENSO), an interannual oceanic mode in the tropical Pacific Ocean (Trenberth, 1997), can also impact subseasonal prediction. It can do so through modulation of the MJO (e.g. Hendon et al., 1999; Kessler, 2001; Pohl & Matthews, 2007) or modulation of the large-scale background state (e.g. Namias, 1986; Moon et al., 2011; Takahashi & Shirooka, 2014), and both can ultimately impact teleconnection propagation (e.g. Stan et al., 2017; Henderson & Maloney, 2018; Tseng et al., 2020; Arcodia et al., 2020) and subseasonal prediction skill (e.g. Johnson et al., 2014; L. Wang & Robertson, 2019). Therefore, when phenomena like the MJO and ENSO are present, they can provide a predictable signal above climate noise and be used to enhance subseasonal prediction skill, known as forecasts of opportunity (Mariotti et al., 2020).

The current understanding of the importance of the tropics on midlatitude subseasonal predictability is rooted in our knowledge of the historical climate. However, with the climate continuously warming, it is unclear how transferable this knowledge will be to a future, warmer climate. Therefore, research on subseasonal timescales has examined how the MJO (Maloney et al., 2018) and ENSO (Cai et al., 2021) will change in the future as well as the subsequent changes to their teleconnections (e.g. Samarasinghe et al., 2021; W. Zhou et al., 2020; Cui & Li, 2021; Meehl et al., 2007; Z.-Q. Zhou et al., 2014; Drouard & Cassou, 2019; Fereday et al., 2020; Beverley et al., 2021). It stands to reason that these changes will likely impact subseasonal predictability across the Northern Hemisphere, but little work has been done in this area (an example being Sheshadri et al., 2021). Here, we utilize the Community Earth System Model Version 2 - Large Ensemble (CESM2-LE; Rodgers et al., 2021) and simple artificial neural networks to identify changes in subseasonal predictability provided by the tropics under future warming.

In recent years, neural networks have been successfully applied to weather and climate prediction (e.g. Chapman et al., 2021; Ham et al., 2019; Gordon et al., 2021; Martin et al., 2021; Labe & Barnes, 2021; Rasp & Thuerey, 2021; Weyn et al., 2021) due to their ability to extract nonlinear relationships from large amounts of data. This makes them advantageous for learning nonlinear relationships in the climate system. In addition, recent advances in explainability techniques and their application to climate sciences demonstrate than neural networks can identify physical relationships in the Earth system (e.g. McGovern et al., 2019; Toms et al., 2020; Mayer & Barnes, 2021; Davenport & Diffenbaugh, 2021). For example, Mayer and Barnes (2021) demonstrate that neural networks can be used to identify subseasonal forecasts of opportunity through the neural network's confidence in a given prediction. They further show that the network identifies physically meaningful sources of subseasonal predictability for the North Atlantic.

Here we use artificial neural networks to quantify how subseasonal prediction skill provided by the tropics may change under future climate warming. Given the importance of forecasts of opportunity for subseasonal prediction in the current climate, we examine both total changes to overall prediction skill as well as changes to skill during forecasts of opportunity, in particular. The artificial neural networks identify subseasonal prediction skill changes across the Northern Hemisphere in the CESM2-LE. In particular, there is an increase in prediction skill over the North Atlantic and western North America as well as a decrease over the North Pacific. In addition, this approach shows that the greatest changes in skill occur during forecasts of opportunity and that these changes appear linked to changes in seasonal variability in the CESM2-LE.

2 Data and Methods

2.1 Data

Here, we examine midlatitude subseasonal prediction skill changes using the first 10 members from the Community Earth System Model Version 2 - Large Ensemble (CESM2-LE; Rodgers et al., 2021). CESM2 has both a well represented MJO (Ahn et al., 2020) and MJO teleconnections (J. Wang et al., 2022) and thus, is ideal for this analysis. We use the years 1970-2015 as our 'historical period' to represent a climate similar to today and compare it to the latter half of the century (2055-2100; 'future period') under the SSP3-7.0 climate change scenario. We find that 10 members are sufficient for this analysis as the network skill plateaus when at least 5 ensemble members are used for training, depending on location and time period (Figure S1; Text S1). While additional ensemble members could be used, we believe our conclusions would remain unaffected, as the sign of the change in prediction skill of the 20% most confident predictions remains consistent regardless of the number of members examined here.

The CESM2-LE members #1-10 are split into training (members #1-8), validation (member #9) and testing data (member #10). To simultaneously detrend and remove the seasonal cycle for each grid point, the 3rd order polynomial fit of the training and validation members' ensemble mean is subtracted from every ensemble member individually for each day of the year. We find the conclusions are insensitive to the specific members assigned to training, validation and testing (Figure S2).

We utilize the CESM2-LE tropical precipitation (28.5°S-28.5°N) and geopotential height at 500 hPa (z500; 31.25°N-88.75°N) during the extended boreal winter (November-March) since this is when MJO teleconnections tend to be strongest (Madden 1986). Tropical precipitation anomalies are computed for each member and grid point by standardizing with the training data mean and standard deviation. For computational purposes, the z500 field is partitioned into non-overlapping 5° x 5° boxes, where the average of these values is assigned to the center grid point latitude and longitude. This decreases the z500 resolution from 2.5° x 2.5° to 7.5° x 7.5°, however, given the large scale structure of z500, we do not expect the resolution reduction to impact the conclusions. The sign of the z500 anomalies are defined by subtracting the training data median from the training, validation and testing data and converting the anomalies into 0s and 1s depending on the sign (negative and positive, respectively).

Sea surface temperatures (SST) from the first 10 members of the CESM2-LE are also used to calculate the Niño 3.4 index for each member, following the *NCAR Climate Data Guide* (2020). The trend and seasonal cycle is removed simultaneously as aforementioned, and a 5 month running mean is applied prior to standardizing the SSTs with each member's mean and standard deviation. An El Niño/La Niña event is therefore defined as a standardized Niño 3.4 index value of greater/less than +/- 1σ. We use this index to examine any possible role that ENSO may play in the identified changes to subseasonal predictability.

2.2 Neural Network Architecture and Application

The neural network ingests daily tropical precipitation anomalies and makes a prediction of the sign of z500 at a given grid point at a lead of 21 days (Figure 1a). The number of input nodes is equal to the number of precipitation grid points (N=3456). The first and second layer of the network consist of 128 and 8 nodes, respectively. A softmax activation function is applied to the output layer of 2 nodes which transforms the network output into values which sum to one. These transformed values represent a network estimation of likelihood, which we refer to as 'model confidence', where the predicted category is defined as a value greater than 0.5. As shown in Mayer and Barnes (2021), when prediction skill increases with model confidence, higher model confidence can be used to identify subseasonal forecasts of opportunity.

We use this network architecture because it has some of the highest validation skill for both the historical and the future time periods in the North Atlantic and also performs well in the North Pacific (Figure S3-S4). We note that slight variations of the hyperparameters (i.e. network depth, nodes per layer, learning rate, ridge regression parameter) show similar skill. While one could optimize the architecture and hyperparameters for every gridpoint individually, we have not done this due to the considerable computational resources necessary and find it unlikely to lead to substantially different conclusions. For additional information on the network architecture and hyperparameters see Text S2.

Example correct network predictions for the testing ensemble member #10 are shown in Figure 1(b-c) for the historical (left column) and the future (right column) periods in (b) the North Pacific and (c) the North Atlantic. The color denotes the sign of the prediction and the darker colors denote the (20% most) confident predictions. The vertical grey shading indicates periods of ENSO events. Figure 1(b-c) demonstrates that

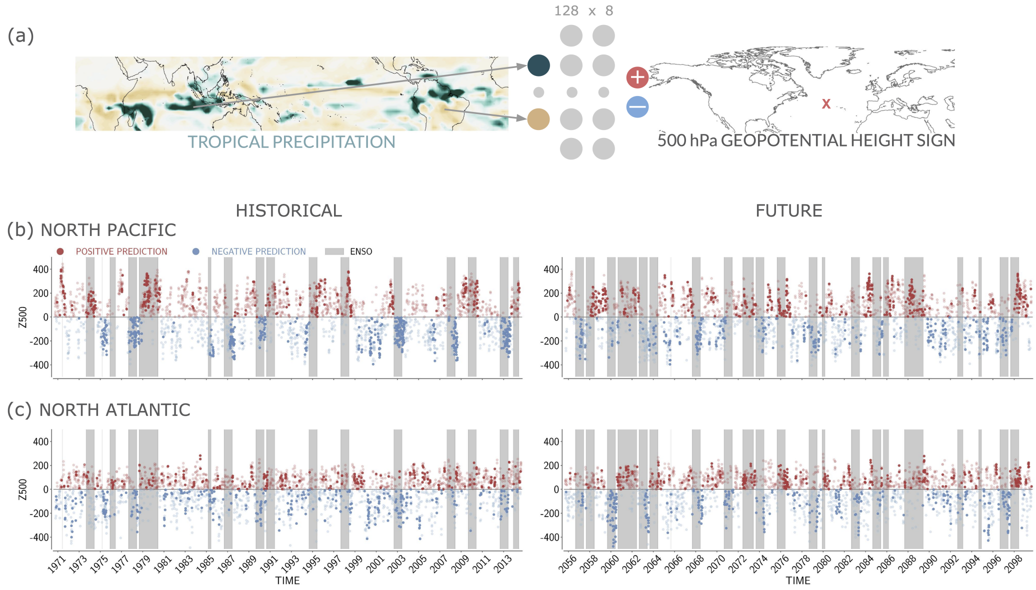

Figure 1. (a) The artificial neural network input (tropical precipitation), architecture (first hidden layer: 128 nodes, second hidden layer: 8 nodes) and output (sign of z500hPa at a location 'x'). (b,c) Timeseries of the *correct* sign predictions of z500 in ensemble member #10 for the historical (left column) and future (right column) for (b) the North Pacific and (c) the North Atlantic. Red (blue) dots indicate positive (negative) predictions. Darker dots denote the 20% most confident predictions, and the grey shading indicates when the standardized Niño 3.4 index exceeds +/-1σ.

the networks can accurately and confidently predict both sign anomalies. In addition, it shows a possible relationship between confident subseasonal predictions and ENSO events, but the amount which confident predictions coincide with ENSO events depends on location and time period. This relationship will be addressed further in section 3.2.

3 Results

3.1 Changes in Subseasonal Prediction Skill

To examine how subseasonal prediction skill provided by tropical-extratropical teleconnections changes in a warmer climate, 100 networks are trained for the North Pacific (41.25°N, 205°E) and the North Atlantic (41.25°N, 325°E) for both the historical and future periods. These two locations are chosen because they encompass regions known to be significantly impacted by MJO (e.g. Mori & Watanabe, 2008; Cassou, 2008; Lin et al., 2009) and ENSO (e.g. Wallace & Gutzler, 1981; Zhang et al., 1996) teleconnections, which subsequently have North American and European impacts. The 100 networks are created by varying their random seed to test the sensitivity of the network to the random initialized weights.

Accuracies binned by various model confidence thresholds are shown in Figure 2. Accuracy increases with model confidence (moving from left to right), suggesting the network is identifying forecasts of opportunity for these regions. The North Pacific (Figure 2a) has higher accuracy compared to the North Atlantic (Figure 2c), likely due to the strong influence of tropical phenomena like the MJO and ENSO in modulating the circulation in the North Pacific (e.g. Wallace & Gutzler, 1981; Zhang et al., 1996; Mori & Watanabe, 2008; Roundy et al., 2010; Riddle et al., 2013). In the future, subseasonal prediction skill increases in the North Atlantic (Figure 2c) and decreases in the North Pacific (Figure 2a) in the CESM2-LE, and this is most evident at higher confidence values. If one examines the accuracy for all (100% most confident) predictions, the North Atlantic and North Pacific accuracies exhibit almost no difference between the two time periods. It is when we focus on the higher confidence predictions that a clear signal emerges. In other words, the changes in subseasonal prediction skill are most evident during forecasts of opportunity in these regions.

Histograms of the accuracies at the 20% most confident threshold (Figure 2 b,d) further show that the future period has substantially shifted away from the historical period in both regions. The majority of the future North Atlantic accuracies exceed the 95th percentile of the historical accuracies, and all of the future North Pacific accuracies lie below the 5th percentile of the historical accuracies.

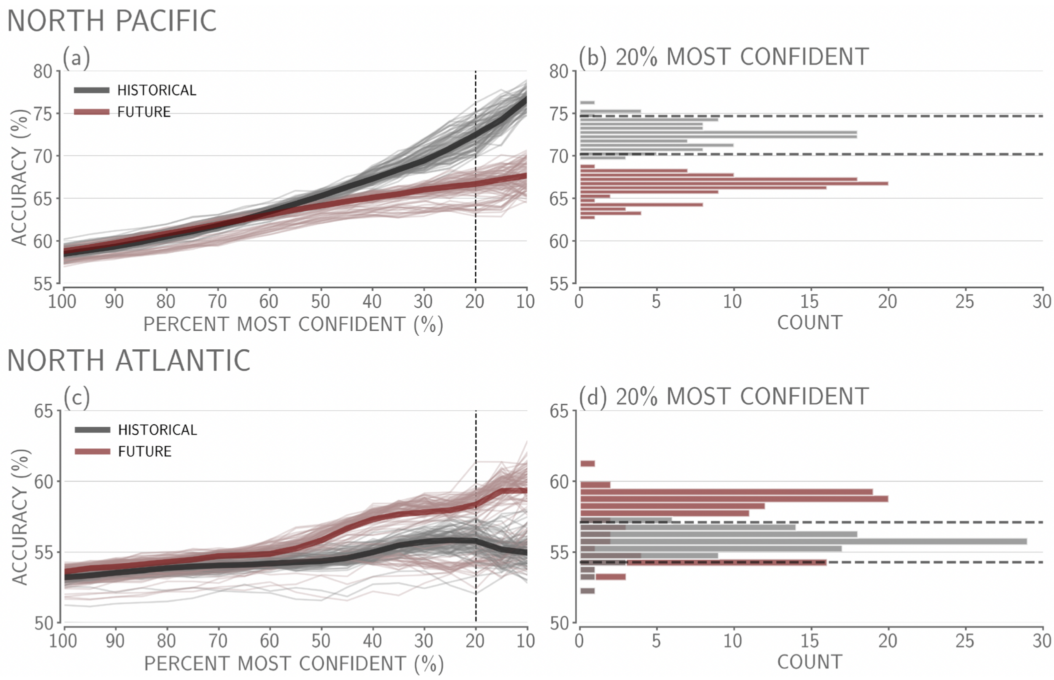

Figure 2. (a,c) Accuracy versus confidence for 100 trained networks in the North Pacific and the North Atlantic from testing member #10. Testing samples are subset so that random chance for all predictions is 50%. Thick grey and red lines denote the median accuracy across the 100 networks at each confidence threshold. Vertical black dashed lines indicate the 20% most confident predictions. (b,d) Histograms of the 100 accuracies at the 20% most confident threshold, using a bin size of 0.5%. Horizontal grey dashed lines indicate the 5th and 95th percentile bounds of the historical accuracies at the 20% most confident level.

To explore whether the results in Figure 2 hold for other regions, we train 10 neural networks for each grid point and time period across the Northern Hemisphere. We train 10 networks instead of 100 for computational efficiency. To test whether these changes in skill in the North Atlantic and North Pacific could be seen with only 10 networks, we conducted a bootstrapping analysis (Text S3; Figure S5) following the method used to create Figure 3, and find that 10 networks are sufficient for identifying these changes. Figure 3 shows the resulting mean testing accuracy of the top three of the 10 networks for each location. The top three networks are defined as the networks with the three highest 20% most confident validation accuracies. We use the top three networks so that the mean accuracies for each region are not as influenced by models that learn very little or not at all.

For all predictions (Figure 3a-b) and 20% most confident predictions ("confident predictions" from here on; Figure 3d-e), the locations of highest skill are in regions associated with the Pacific/North America pattern (PNA; Wallace & Gutzler, 1981). The higher accuracies over PNA regions suggests the network is most likely identifying forecasts of opportunity associated with teleconnections from the MJO and/or ENSO (e.g. Wallace & Gutzler, 1981; Zhang et al., 1996; Mori & Watanabe, 2008; Roundy et al., 2010; Riddle et al., 2013). In the future period (Figure 3b,e), there is an additional region of higher accuracies spread across Asia and the North Atlantic. Overall, the confident predictions have higher accuracies than all predictions, indicating that higher model confidence predictions exhibit greater skill.

In the future, spatially coherent increases in skill are seen across Asia, along the west coast of North America, across the southern United States and throughout the North Atlantic (Figure 3c,f) while decreases are seen over the North Pacific, Canada and western Europe. While the change in skill over East Asia is substantial, it appears that the overall skill in East Asia for both time periods does not harness any subseasonal variability, but rather comes about exclusively from seasonal variability or longer timescales (Figure S8-S9). As a result, these changes in skill are not addressed further here. The difference plots for both all and the confident predictions (Figure 3c,f) have similar spatial patterns of changes in accuracy, however, the confident predictions show the largest changes in skill. Specifically, the absolute maximum change in skill for all predictions is about 5% while the absolute maximum change in skill for confident predictions is about 10%. This further demonstrates that the greatest changes to subseasonal prediction skill provided by the tropics occur during forecasts of opportunity across the Northern Hemisphere, consistent with Figure 2.

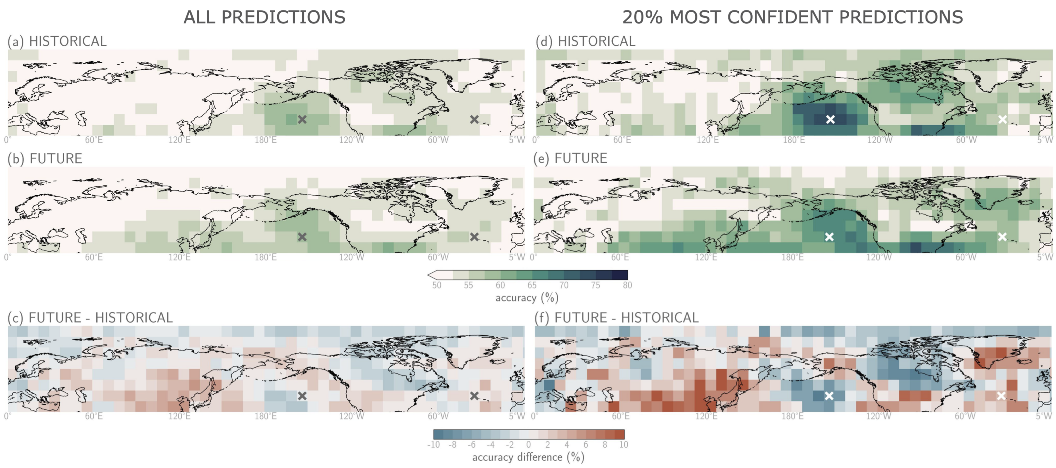

Figure 3. (a,b,d,e) Mean testing accuracy of the best 3 models for (a,b) all and (b,e) the 20% most confident predictions. (c,f) Difference in accuracy between the future and the historical time periods for (c) all and (f) the 20% most confident predictions, where red (blue) indicates an increase (decrease) in accuracy in the future. The grey and white 'x' indicate the North Pacific and North Atlantic regions (from left to right) used in Figures 1,2.

3.2 Tropical Drivers of Changing Midlatitude Skill

Seasonal variability can have a large influence on subseasonal variability and prediction skill. In the tropics, ENSO can modulate the MJO (e.g. Hendon et al., 1999; Kessler, 2001; Pohl & Matthews, 2007) and the basic state (e.g. Namias, 1986; Moon et al., 2011; Takahashi & Shirooka, 2014), and ENSO teleconnections can (de)constructively interfere with MJO teleconnections (e.g. Stan et al., 2017; Henderson & Maloney, 2018; Tseng et al., 2020; Arcodia et al., 2020). Recent studies have identified possible changes to both MJO and ENSO variability (Maloney et al., 2018; Cai et al., 2021) as well as their teleconnections (e.g. Fredriksen et al., 2020; Beverley et al., 2021; W. Zhou et al., 2020; Samarasinghe et al., 2021) under future climate warming. Thus, the changes in midlatitude subseasonal prediction skill seen in Figures 2 and 3 could be a reflection of changes to subseasonal variability, seasonal variability, or through a combination of changes to both.

We find that the increase in skill along the west coast of North America and in the North Atlantic is supported by previous research on MJO and ENSO teleconnections in a warmer climate. In particular, the subseasonal skill increase along the west coast of North America (Figure 3f) appears to be associated with a north-eastward shift of higher accuracies over the North Pacific in the future (Figure 3d-e). This is consistent with research showing that PNA patterns initiated by ENSO (e.g. Meehl & Teng, 2007; Meehl et al., 2007; Müller & Roeckner, 2008; Kug et al., 2010; Z.-Q. Zhou et al., 2014; Fredriksen et al., 2020; Beverley et al., 2021) and the MJO (Wolding et al., 2017; W. Zhou et al., 2020; Jenney et al., 2021; J. Wang et al., 2022) are projected to shift eastward in a warmer climate in a variety of climate models, including CESM2 (Fredriksen et al., 2020; J. Wang et al., 2022). In the North Atlantic, increased skill is also consistent with research suggesting that the North Atlantic may become more sensitive to MJO teleconnections (Samarasinghe et al., 2021) and that the ENSO-NAO teleconnection may strengthen (Drouard & Cassou, 2019; Fereday et al., 2020) in the future. The decrease in skill over the North Pacific is also consistent with recent research using a variety of CMIP6 models that suggests the ENSO teleconnection amplitude over the North Pacific may weaken in a warmer climate (e.g. Fredriksen et al., 2020; Beverley et al., 2021).

To gain insight into the neural network's identified sources of predictability, we apply explainable AI to create heatmaps of the relevant regions of the input tropical precipitation the network uses to make confident and correct predictions (see Text S4; Bach et al., 2015; Montavon et al., 2019). In the North Pacific and North Atlantic, the network tends to focus on the tropical equatorial Pacific, typically associated with ENSO (Figure S6). In the North Pacific, the future decrease in skill is associated with a decrease in relevance of the ENSO region (Figure S6a-d). For the North Atlantic, the future increase in skill is associated with an increase in relevance of the ENSO region (Figure S6e-h). These explainability results suggest that the changes in subseasonal prediction skill may be related to changes in the importance of the ENSO region (i.e. seasonal variability), even though both subseasonal and seasonal variability are contributing to the total skill (Figure S8-S9). This changing role of ENSO in both regions is also evident in the prediction timeseries in Figure 1. In the North Atlantic (Figure 1c), the confident predictions in the historical period are scattered throughout the years, whereas in the future period, the confident predictions correspond more frequently with ENSO events (darker dots mainly occur in the grey shading). The opposite is seen for the North Pacific (Figure 1b). Given the results of this analysis, we next examine if the changes in midlatitude subseasonal prediction skill are related to changes in ENSO teleconnections.

We analyze the relationship between ENSO teleconnections and subseasonal prediction skill changes across the Northern Hemisphere by calculating how often a positive z500 anomaly occurs 21 days following an El Niño/La Niña event (Figure 4). This metric quantifies the consistency of specific teleconnections following ENSO events. Over the North Pacific, the consistency of the z500 sign following both ENSO phases decreases (Figure 4c,f), suggestive of a reduction in the influence of ENSO teleconnections. Fur-

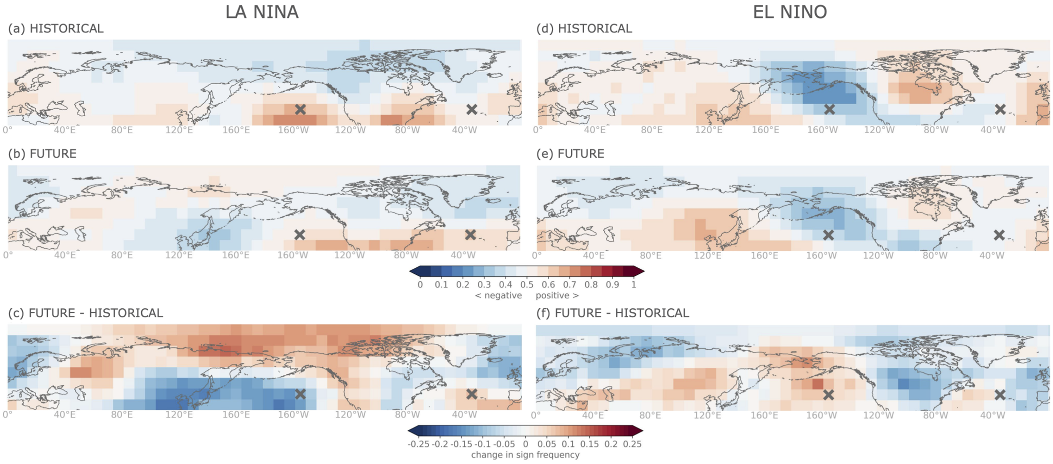

Figure 4. (a,b,d,e) Frequency of a positive sign anomaly 21 days following a standardized Niño 3.4 Index value of greater/less than +/- 1σ. Values greater (less) than 0.5 frequency indicate that positive (negative) sign anomalies are more frequent. (c,f) Difference in frequency between the future and historical time period. The left (right) column is for La Niña (El Niño). The grey 'x' indicate the North Pacific and North Atlantic regions (from left to right) used in Figures 1,2.

thermore, the large decrease in skill over Canada (Figure 3f) aligns with the decrease in El Niño teleconnection consistency in the future (Figure 4f). Over the North Atlantic, there is a slight increase in ENSO teleconnection consistency which may be related to the projected strengthening of the ENSO-NAO teleconnection in the future (Drouard & Cassou, 2019; Fereday et al., 2020). Lastly, the increase in skill along the west coast of North America (Figure 3f) aligns with an increase in consistency of La Niña teleconnections (Figure 4c). Thus, we hypothesize that the substantial changes in subseasonal prediction skill in regions across the Northern Hemisphere are connected to changes in ENSO teleconnections in the CESM2-LE.

We provide further evidence of the role of seasonal variability in changes to subseasonal prediction skill through an additional neural network analysis in the North Pacific and North Atlantic. We filter out 60+ day variability from the z500 anomalies (Text S5) to remove low-frequency signals such as those from ENSO teleconnections. With this filtering, there is almost no change in skill between the historical and future period in the North Pacific (Figure S7c-d). This demonstrates that the decrease in skill in this region is mainly a result of changes to seasonal variability. In the North Atlantic, the increase in skill is still seen, but to a reduced degree when the lower frequencies are removed (Figure S7e-f). This suggests that seasonal variability is playing a role in subseasonal prediction skill changes in this region, however, there is also likely a contribution from subseasonal variability to these changes. This is consistent with research that suggests the North Atlantic may become more sensitive to MJO teleconnections in the future (Samarasinghe et al., 2021).

The influence of seasonal variability on subseasonal prediction skill changes can be further examined in the North Pacific and North Atlantic by training the neural networks to instead predict the sign of unfiltered z500 anomalies on seasonal lead times. In the North Pacific, we find that *changes* in skill at 60 and 90 day leads are similar to that for a lead of 21 days. This again implies that the changes in subseasonal prediction skill seen in the North Pacific are due to changes in seasonal variability. In the North Atlantic, the change in skill for the seasonal lead time is larger than the 21 day lead time. This difference in the change suggests that the network is focusing on different sources of predictability for the 21 day prediction compared to the 60 or 90 day predictions, implying again that the change in skill in the North Atlantic is not purely due to seasonal variability changes in the future (Figure S8-S9).

4 Conclusions

While accurate subseasonal predictions are important for society (White et al., 2021), this timescale is known to exhibit limited predictability (Vitart et al., 2017). One method to improve prediction skill on subseasonal timescales is to utilize Earth system states which are known to provide enhanced subseasonal predictability when they are present (forecasts of opportunity; Mariotti et al., 2020). Previous research has examined how specific Earth system states important for subseasonal prediction (e.g. MJO and ENSO) and their teleconnections may change in a warmer climate (e.g. Maloney et al., 2018; Cai et al., 2021; J. Wang et al., 2022). To address whether these projected changes ultimately impact subseasonal predictability, we use the CESM2-LE and simple artificial neural networks to quantify and understand how subseasonal predictability provided by the tropics may change in a warmer climate. We find that there are changes to subseasonal prediction skill across the Northern Hemisphere and the largest differences in skill mainly occur during forecasts of opportunity.

Our results are supported by recent research on changes to MJO and ENSO teleconnections. In particular, the increase in skill along the west coast of North America is consistent with the projected eastward shift of MJO and ENSO teleconnections in the future (e.g. Jenney et al., 2021; J. Wang et al., 2022; Fredriksen et al., 2020; Beverley et al., 2021). In addition, our results suggest there is a contribution from both subseasonal and seasonal variability changes to the increase in prediction skill in the North Atlantic. This is consistent with research suggesting the North Atlantic becomes more sensitive to the MJO (Samarasinghe et al., 2021) and ENSO (Drouard & Cassou, 2019; Fereday et al., 2020) in the future. We also identify a substantial decrease in skill over the North Pacific and from our analysis, hypothesize that this decrease is mainly driven by a reduced influence of ENSO teleconnections to this region in the future. Overall, while both MJO and ENSO teleconnections are projected to change in the future, our analysis demonstrates that changes to ENSO and its teleconnections (e.g. seasonal variability) at least partially explain substantial changes in subseasonal prediction skill across the North Hemisphere in the CESM2-LE. Changes to subseasonal variability may still play a role in changes to subseasonal prediction skill in certain locations (e.g. North Atlantic), but further work is needed to understand and quantify its contribution.

Using the CESM2-LE, we show that neural networks are a useful tool for identifying and understanding future changes in predictability. In addition, we find that changes in subseasonal prediction skill across the Northern Hemisphere are often largest during forecasts of opportunity, suggesting that future research on prediction skill changes should focus on periods of enhanced predictability. While this research addresses changes in boreal wintertime subseasonal predictability provided by the tropics, future research should also examine how other seasons and sources of predictability may be affected in a warmer climate. This could include identifying possible changes to the importance of the stratosphere for subseasonal prediction or changes to boreal summer subseasonal predictability due to changes to the importance of the boreal summer intraseasonal oscillation (B. Wang & Rui, 1990). Furthermore, although this work examines subseasonal predictability changes by the end of the century, examining how quickly these changes may be detected is also worthy of study. Ultimately, this research demonstrates the utility of neural networks to quantify and gain physical insight into changes in subseasonal predictability in future climates.

Open Research

CESM2 Large Ensemble data (precipitation, SSTs and z500) are provided by the University Corporation for Atmospheric Research/National Center for Atmospheric Research (`https://www.cesm.ucar.edu/projects/community-projects/LENS2/data-sets.html`).

Acknowledgments

This research is partially funded by the NSF Graduate Research Fellowship under grant 006784 and partially funded by the Regional and Global Model Analysis program area of the U.S. Department of Energy's Office of Biological and Environmental Research as part of the Program for Climate Model Diagnosis and Intercomparison project.

The authors declare that they have no conflict of interest.

References

Ahn, M., Kim, D., Kang, D., Lee, J., Sperber, K. R., Gleckler, P. J., . . . Kim, H. (2020, June). MJO propagation across the maritime continent: Are CMIP6 models better than CMIP5 models? *Geophys. Res. Lett.*, *47*(11), 741.

Arcodia, M. C., Kirtman, B. P., & Siqueira, L. S. P. (2020). How MJO teleconnections and ENSO interference impacts US precipitation. *J. Clim.*.

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.

Baggett, C. F., Barnes, E. A., Maloney, E. D., & Mundhenk, B. D. (2017, July). Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophys. Res. Lett.*, *44*(14), 2017GL074434.

Beverley, J. D., Collins, M., Hugo Lambert, F., & Chadwick, R. (2021, August). Future changes to el niño teleconnections over the north pacific and north america. *J. Clim.*, *34*(15), 6191–6205.

Cai, W., Santoso, A., Collins, M., Dewitte, B., Karamperidou, C., Kug, J.-S., . . . Zhong, W. (2021, August). Changing el niño–southern oscillation in a warming climate. *Nature Reviews Earth & Environment*, *2*(9), 628–644.

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

Cassou, C., Kushnir, Y., Hawkins, E., Pirani, A., Kucharski, F., Kang, I.-S., & Caltabiano, N. (2018, March). Decadal climate variability and predictability: Challenges and opportunities. *Bull. Am. Meteorol. Soc.*, *99*(3), 479–490.

Chapman, W. E., Monache, L. D., Alessandrini, S., Subramanian, A. C., Martin Ralph, F., Xie, S.-P., . . . Hayatbini, N. (2021, October). Probabilistic predictions from deterministic atmospheric river forecasts with deep learning. *Mon. Weather Rev.*, *-1*(aop).

Cui, J., & Li, T. (2021, October). Changes in MJO characteristics and impacts in the past century. *J. Clim.*, *-1*(aop), 1–1.

Davenport, F. V., & Diffenbaugh, N. S. (2021, July). Using machine learning to analyze physical causes of climate change: A case study of U.S. midwest extreme precipitation. *Geophys. Res. Lett.*.

Drouard, M., & Cassou, C. (2019, December). A modeling- and Process-Oriented study to investigate the projected change of ENSO-Forced wintertime teleconnectivity in a warmer world. *J. Clim.*, *32*(23), 8047–8068.

Fereday, D. R., Chadwick, R., Knight, J. R., & Scaife, A. A. (2020, October). Tropical rainfall linked to stronger future ENSO-NAO teleconnection in CMIP5 models. *Geophys. Res. Lett.*, *n/a*(n/a), e2020GL088664.

Fredriksen, H.-B., Berner, J., Subramanian, A. C., & Capotondi, A. (2020, November). How does el niño–southern oscillation change under global warming—a first look at CMIP6. *Geophys. Res. Lett.*, *47*(22).

Friedman, J. H. (2012, July). Fast sparse regression and classification. *Int. J. Forecast.*, *28*(3), 722–738.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.

Gordon, E. M., Barnes, E. A., & Hurrell, J. W. (2021, November). Oceanic harbingers of pacific decadal oscillation predictability in CESM2 detected by neural networks. *Geophys. Res. Lett.*, *48*(21).

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

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.

Hendon, H. H., Zhang, C., & Glick, J. D. (1999, August). Interannual variation of the Madden–Julian oscillation during austral summer. *J. Clim.*, *12*(8), 2538–2550.

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.

Jenney, A. M., Randall, D. A., & Barnes, E. A. (2021, July). Drivers of uncertainty in future projections of Madden–Julian oscillation teleconnections. *Weather Clim. Dynam.*, *2*(3), 653–673.

Johnson, N. C., Collins, D. C., Feldstein, S. B., L'Heureux, M. L., & Riddle, E. E. (2014, February). Skillful wintertime north american temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. *Weather Forecast.*, *29*(1), 23–38.

Kessler, W. S. (2001, July). EOF representations of the Madden–Julian oscillation and its connection with ENSO. *J. Clim.*, *14*(13), 3055–3061.

Kingma, D. P., & Ba, J. (2014, December). Adam: A method for stochastic optimization.

Kug, J.-S., An, S.-I., Ham, Y.-G., & Kang, I.-S. (2010, May). Changes in el niño and la niña teleconnections over north Pacific–America in the global warming simulations. *Theor. Appl. Climatol.*, *100*(3), 275–282.

Labe, Z. M., & Barnes, E. A. (2021, November). *Predicting slowdowns in decadal climate warming trends with explainable neural networks.*

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., & 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.

Maloney, E. D., Adames, Á. F., & Bui, H. X. (2018, December). Madden–Julian oscillation changes under anthropogenic warming. *Nat. Clim. Chang.*, *9*(1), 26–33.

Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2021, March). Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset.

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.*.

Martin, Z. K., Barnes, E. A., & Maloney, E. D. (2021). *Using simple, explainable neural networks to predict the Madden-Julian oscillation.*

Mayer, K. J., & Barnes, E. A. (2021, May). Subseasonal forecasts of opportunity identified by an explainable neural network. *Geophys. Res. Lett.*.

McGovern, A., Lagerquist, R., Gagne, D. J., Eli Jergensen, G., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019, November). Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Am. Meteorol. Soc.*, *100*(11), 2175–2199.

Meehl, G. A., Tebaldi, C., Teng, H., & Peterson, T. C. (2007, October). Current and future U.S. weather extremes and el niño. *Geophys. Res. Lett.*, *34*(20).

Meehl, G. A., & Teng, H. (2007, October). Multi-model changes in el niño teleconnections over north america in a future warmer climate. *Clim. Dyn.*, *29*(7-8), 779–790.

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.

Moon, J.-Y., Wang, B., & Ha, K.-J. (2011, September). ENSO regulation of MJO teleconnection. *Clim. Dyn.*, *37*(5), 1133–1149.

Mori, M., & Watanabe, M. (2008). The growth and triggering mechanisms of the PNA: A MJO-PNA coherence. *. 2*, *86*(1), 213–236.

Müller, & Roeckner. (2008). ENSO teleconnections in projections of future climate in ECHAM5/MPI-OM. *Climate Dynamics*, *31*, 533–549.

Namias, J. (1986, July). Persistence of flow patterns over north america and adjacent ocean sectors. *Mon. Weather Rev.*, *114*(7), 1368–1383.

Ncar climate data guide. (2020). `https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni`. (Accessed: 2022-2-11)

Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25). Determination press San Francisco, CA.

Pohl, B., & Matthews, A. J. (2007, June). Observed changes in the lifetime and amplitude of the Madden–Julian oscillation associated with interannual ENSO sea surface temperature anomalies. *J. Clim.*, *20*(11), 2659–2674.

Rasp, S., & Thuerey, N. (2021, February). Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for WeatherBench. *J. Adv. Model. Earth Syst.*, *13*(2).

Riddle, E. E., Stoner, M. B., Johnson, N. C., L'Heureux, M. L., Collins, D. C., & Feldstein, S. B. (2013, April). The impact of the MJO on clusters of wintertime circulation anomalies over the north american region. *Clim. Dyn.*, *40*(7), 1749–1766.

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., ... Others (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics Discussions*, 1–22.

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.

Samarasinghe, S. M., Connolly, C., Barnes, E. A., Ebert-Uphoff, I., & Sun, L. (2021, March). Strengthened causal connections between the MJO and the north atlantic with climate warming. *Geophys. Res. Lett.*, *48*(5).

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.

Sheshadri, A., Borrus, M., Yoder, M., & Robinson, T. (2021, December). Midlatitude error growth in atmospheric GCMs: The role of eddy growth rate. *Geophys. Res. Lett.*, *48*(23).

Simpson, I. R., Deser, C., McKinnon, K. A., & Barnes, E. A. (2018, October). Modeled and observed multidecadal variability in the north atlantic jet stream and its connection to sea surface temperatures. *J. Clim.*, *31*(20), 8313–8338.

Stan, C., Straus, D. M., Frederiksen, J. S., Lin, H., Maloney, E. D., & Schumacher, C. (2017, December). Review of Tropical-Extratropical teleconnections on intraseasonal time scales: The subseasonal to seasonal (S2S) teleconnection Sub-Project. *Rev. Geophys.*, *55*(4), 902–937.

Takahashi, C., & Shirooka, R. (2014, September). Storm track activity over the north pacific associated with the Madden-Julian oscillation under ENSO conditions during boreal winter. *J. Geophys. Res.*, *119*(18), 10,663–10,683.

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020, September). Physically interpretable neural networks for the geosciences: Applications to earth system variability. *J. Adv. Model. Earth Syst.*, *12*(9).

Trenberth, K. E. (1997). The definition of el nino. *Bull. Am. Meteorol. Soc.*, *78*(12), 2771–2778.

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. A. (2020, May). The consistency of MJO teleconnection patterns on interannual time scales. *J. Clim.*, *33*(9), 3471–3486.

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.

Wallace, J. M., & Gutzler, D. S. (1981, April). Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Weather Rev.*, *109*(4), 784–812.

Wang, B., & Rui, H. (1990, March). Synoptic climatology of transient tropical intraseasonal convection anomalies: 1975–1985. *Meteorol. Atmos. Phys.*, *44*(1), 43–61.

Wang, J., Kim, H., & DeFlorio, M. J. (2022). Future changes of PNA-like MJO teleconnections in CMIP6 models: underlying mechanisms and uncertainty. *Journal of Climate*, 1–40.

Wang, L., & Robertson, A. W. (2019, May). Week 3–4 predictability over the united states assessed from two operational ensemble prediction systems. *Clim. Dyn.*, *52*(9), 5861–5875.

Welch, B. L. (1947). The generalisation of student's problems when several different population variances are involved. *Biometrika*, *34*(1-2), 28–35.

Weyn, J. A., Durran, D. R., Caruana, R., & Cresswell-Clay, N. (2021, June). Subseasonal forecasting with a large ensemble of deep-learning weather prediction models. *J. Adv. Model. Earth Syst.*.

White, C. J., Domeisen, D. I. V., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., ... Wilson, R. G. (2021, November). Advances in the application and utility of subseasonal-to-seasonal predictions. *Bull. Am. Meteorol. Soc.*, *-1*(aop), 1–57.

Wolding, B. O., Maloney, E. D., Henderson, S., & Branson, M. (2017, March). Climate change and the madden-julian oscillation: A vertically resolved weak temperature gradient analysis. *J. Adv. Model. Earth Syst.*, *9*(1), 307–331.

Zhang, Y., Wallace, J. M., & Iwasaka, N. (1996, July). Is climate variability over the north pacific a linear response to ENSO? *J. Clim.*, *9*(7), 1468–1478.
Zheng, C., Kar-Man Chang, E., Kim, H.-M., Zhang, M., & Wang, W. (2018, August). *Impacts of the Madden–Julian oscillation on Storm-Track activity, surface air temperature, and precipitation over north america.*
Zhou, W., Yang, D., Xie, S.-P., & Ma, J. (2020, July). Amplified Madden–Julian oscillation impacts in the Pacific–North america region. *Nat. Clim. Chang.*, *10*(7), 654–660.
Zhou, Z.-Q., Xie, S.-P., Zheng, X.-T., Liu, Q., & Wang, H. (2014, December). Global Warming–Induced changes in el niño teleconnections over the north pacific and north america. *J. Clim.*, *27*(24), 9050–9064.

Supporting Information for “Quantifying the Effect of Climate Change on Midlatitude Subseasonal Prediction Skill Provided by the Tropics”

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

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

Contents of this file

1. Text S1 to S6

2. Figures S1 to S9

Corresponding author: Kirsten J. Mayer, kirsten.j.mayer@gmail.com

Overview In the supporting information, we provide details on the robustness of our results to changes in the number of training ensemble members and variations in the neural network architecture and hyperparameters. The sensitivity of the results to the choice of members used for validation and testing are examined as well. We also include information about the bootstrapping analysis, and provide additional information on the neural network explainability technique and the seasonal filtering results, along with the corresponding figures. Confidence versus accuracy diagrams for seasonal predictions are also included.

Text S1: Network Sensitivity to the Number of Training Members To test whether 8 training ensemble members (members #1-8) are sufficient for this analysis, 100 neural networks are trained with different sized training sets, starting with only 1 member and increasing to 8 members iteratively (moving left to right in Figure S1). Figure S1a,b includes the accuracies of the testing member (#10) for all predictions and Figure S1c,d includes the accuracies for the corresponding 20% most confident predictions. In the North Pacific (Figure S1a,c), the skill for the historical *and* future periods plateaus at about 5 training members for both all and the most confident predictions. The North Atlantic (Figure S1b,d) shows more skill variability with training size, but generally maintains the same range of skill for each period when 3 or more ensemble members are used. The skill variability in the North Atlantic may be related to multidecadal variability in the ensemble members (e.g. Simpson et al. 2018). While the Pacific is also impacted by longer timescales, it is more prominently impacted by decadal variability instead (e.g.

Cassou et al. 2018). Therefore, if a lower multidecadal predictability state in the North Atlantic is dominating the majority of the training years for a given ensemble member, this could ultimately impact how well the network can learn.

Text S2: Network Architecture and Sensitivity to Hyperparameters The neural network used in this study consists of two layers of 128 and 8 nodes, respectively. The rectified linear unit "relu" activation function is applied to the hidden layers. Categorical cross entropy is used for the loss function and the batch size is set to 256 samples. Adam (Kingma and Ba 2014) is used as an optimizer with a learning rate of 0.001. We reduce the learning rate exponentially by $e^{-0.1}$ for each epoch after 10 epochs to assist the network in minimizing the loss. To reduce overfitting on the training data, ridge regression ($L_2 =$ 1.0; Friedman 2012) is applied to the first hidden layer and early stopping is implemented. Ridge regression is used to direct the network to account for spatial autocorrelation within the input field (tropical precipitation). Early stopping monitors the validation prediction accuracy, so when the validation prediction accuracy does not increase for more than 20 epochs, the network stops training and reverts back to the network weights from 20 epochs before. Otherwise, the network concludes training at 100 epochs. We find that a patience of 20 epochs is useful for this problem to reduce overfitting since the network never trains for the full 100 epochs when early stopping is implemented. The output layer consists of 2 nodes and uses the softmax activation function. The softmax activation function converts the output into two numbers which sum to 1 and can be interpreted as the likelihood of a given prediction, referred to as "model confidence". A more detailed description of network training for a similar artificial neural network is provided in the supporting information of Mayer and Barnes (2021) and additional information on artificial neural networks in general can be found in Nielsen (2015) and Goodfellow et al. (2016).

To test the sensitivity of our conclusions to the network architecture and hyperparameter choice, the learning rate, ridge regression parameter, nodes per layer and the number or layers were all varied and the validation accuracy compared (Figure S3-S4). Figure S3 (S4) shows results for the North Pacific (Atlantic), where the validation member #9 accuracy of 10 trained models with different initial weights are shown for each hyperparameter variation and time period. The network hyperparameters and architecture for this analysis were ultimately chosen because it has some of the highest validation skill for both the historical and the future time period in the North Atlantic, but also performs well in the North Pacific (Figure S3-S4). We initially focus on the skill of the network in the North Atlantic because it is more difficult for the network to predict than the North Pacific. We also see that slight variations of these hyperparameters show similar skill to the network chosen.

We note that for the North Pacific hyperparameter sweep (Figure S3), validation member #9 shows a decrease in skill for all predictions between the historical and the future period which is not seen with the testing member (Figure 2a). We believe that the decrease in skill in the validation data between the two time periods is likely a result of slight overfitting of the validation during the historical time period due to its use for early stopping (not shown).

Text S3: Accuracy Bootstrapping Analysis Due to the computational costs of training 100 networks for each grid point in the Northern Hemisphere, 10 neural networks are trained for each location instead. To check whether these changes identified in the North Pacific and North Atlantic with 100 networks can be seen using only 10 networks, and to provide a reference of the magnitude of significant skill changes for the other grid points in Figure 3, we used the 100 models trained for both the North Pacific and North Atlantic to conduct a bootstrapping analysis.

For each location, from the 100 models trained, 10 models are randomly selected and the top three networks are chosen, defined using the three highest 20% most confident validation accuracies. The mean of the 20% most confident testing accuracies is then calculated for these three models, identical to the method used to calculate the testing accuracy for each grid point in Figure 3. This is repeated 1000 times for each time period with the resulting distributions plotted in Figure S5. For each region, we find that the direction of change in skill for 10 networks is the same as that for 100 networks, and the future accuracy is statistically different than the historical time period using a one-sided Welch's t-test (Welch 1947) at a 95% confidence level (p-value < 0.0001). For the North Pacific (Atlantic), we test whether the future period is statistically less (greater) than the historical. Therefore, we find that 10 networks is sufficient for identifying these subseasonal prediction skill changes.

Text S4: Layer-wise Relevance Propagation Layer-wise Relevance Propagation (LRP; Bach et al. 2015; Montavon et al. 2019) is a neural network (attribution) explainability technique that creates a heatmap of the estimated "relevance" of the input for a given prediction. Here, we use the LRP_z rule, which has been shown to perform well for specific geoscience applications (Mamalakis et al. 2021). For an individual prediction, LRP_z backpropagates relevance information from an output node through the network to create a heatmap of the estimated regions of the input that the network found most relevant for its prediction, where positive (negative) relevance denotes positive (negative) contributions to the final output. The softmax activation function is removed before back propagation and the heatmap for each prediction is normalized by dividing by the absolute maximum relevance value in that map. Figure S6 shows the average of the correct and confident predictions' heatmaps for an example neural network. We find that other networks produce similar LRP maps to this example.

In both the North Pacific and North Atlantic, the differences in relevance between the two time periods are most evident in the equatorial Pacific. In the North Pacific, the relevance of this region is generally reduced and the focus shifts westward in the future period. In the North Atlantic, the relevance of this region increases in the future, mainly over the western equatorial Pacific. The change in relevance of the equatorial Pacific corresponds with the change in prediction skill for both regions and suggests that the network's changing focus in the equatorial Pacific is related to the changes in subseasonal prediction skill.

Text S5: Seasonal Filtering Analysis To further examine the possible role of seasonal variability influencing future subseasonal prediction skill, we task the neural network to predict the sign of z500 anomalies using only z500 variability on *shorter* than 60 day (subseasonal) timescales. The z500 anomalies are filtered by removing the forward 60 day running mean. This filtering is used to direct the network to focus on tropical precipitation specifically related to midlatitude subseasonal variability in the z500 anomalies. Thus, changes to prediction skill between the two time periods are a result of changes in the ability to specifically predict midlatitude variability with shorter than 60 day periods. We use this approach to identify if the changes in skill are mainly related to changes in midlatitude subseasonal variability or if the changes are related purely to changes in seasonal variability, or a combination of the two. We note that this filtering analysis could have been conducted for the main paper, however, we wanted to retain seasonal variability information to identify possible skill changes that could be seen in a typical subseasonal forecast. For each time period and region, 100 networks are again trained and their accuracies across model confidence thresholds are computed (Figure S7).

Overall, the removal of seasonal variability reduces the information the network can use for its predictions, so the filtering leads to a decrease in skill for both time periods compared to the unfiltered predictand. In the North Pacific (Figure S7a-b), there is virtually no difference between the historical and future period when seasonal variability is removed because the historical skill decreases more than the future skill, resulting in similar accuracies across model confidence thresholds. This implies that the historical period relies more on seasonal variability for subseasonal prediction than the future period, consistent with the LRP analysis. The lack of skill change between the two time periods also implies that the *change* in subseasonal prediction skill seen in the unfiltered analysis is related to midlatitude seasonal variability instead of subseasonal. In the North Atlantic, we see that the future period still has higher prediction skill compared to the historical, although, the overall skill for both time periods is reduced (Figure S7c-d). A reduction in skill for both time periods is expected because the LRP maps suggest that both time periods rely, at least partially, on the ENSO regions for the predictions. However, even with midlatitude seasonal variability removed from z500, there is still an increase in skill from the historical to the future time period over the North Atlantic, suggesting there are other shorter timescale variability contributors to the increase in midlatitude subseasonal prediction skill in the future.

Text S6: Seasonal Predictions To check whether the neural networks are using more than seasonal information for their predictions, we train 100 neural networks for East Asia, the North Pacific and the North Atlantic for leads of 60 and 90 days (Figure S8-S9). East Asia is also analyzed here because of the unexpected increase in subseasonal prediction skill in the future (Figure 3). By training the networks at seasonal lead times, we can assess whether the prediction skill in each region *only* comes from seasonal variability. In other words, if only seasonal variability is contributing to the prediction skill, there should be no difference in skill between a lead of 21 days and a lead of 60 or 90 days.

We see that in East Asia (Figure S8a-b, S9a-b) the neural networks have similar skill whether trained at a lead of 21, 60 or 90 days. This suggests that the skill seen at 21 days is likely skill from seasonal variability alone. On the other hand, the North Pacific (Figure S8c-d, S9c-d) and the North Atlantic (Figure S8e-f, S9e-f) both show higher skill at a lead of 21 days, suggesting that in these regions the neural network is using more than seasonal variability for its predictions. Lastly, Figures S8 and S9 demonstrate that the *changes* in skill between the historical and future periods at a lead of 21 days (Figure 2), are similar to those for seasonal lead predictions, particularly in the North Pacific. In the North Atlantic, the change in skill is larger for the seasonal lead predictions than the 21 day lead. This implies that the networks for the 21 day lead prediction use sources of predictability other than seasonal variability to make predictions, ultimately impacting how much the skill changes between time periods. Overall, this analysis again suggests that seasonal variability is playing a role in the changes to subseasonal prediction skill, but the magnitude of the seasonal influence varies by region.

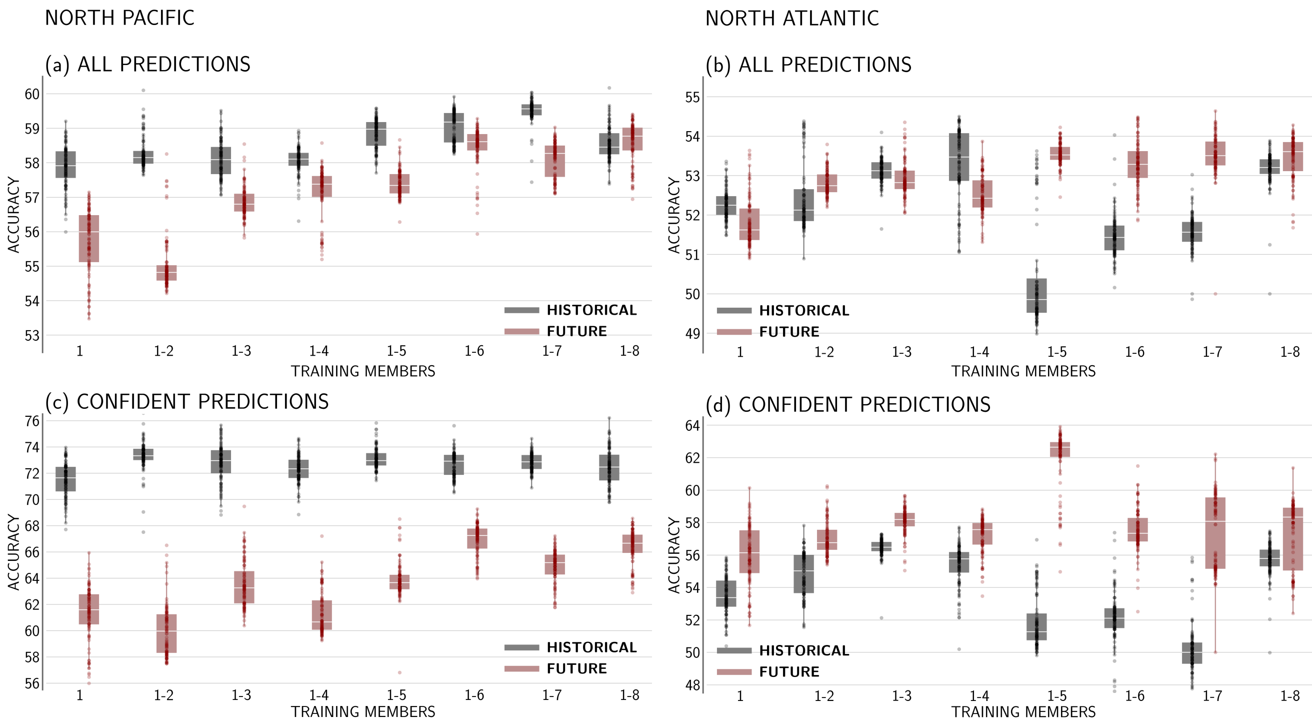

Figure S1. Box and whisker plots of (a,b) all prediction and (c,d) the 20% most confident prediction accuracies for testing ensemble member #10 for the (a,c) North Pacific and (b,d) North Atlantic using increasing numbers of ensemble members for training. Training members #1-8 are used for the main analysis. The black (red) denotes the historical (future) period and the x-axis are the members used to train. The dots indicate individual accuracy for each of the 100 models trained. The white line across each box is the median of the models and the edges of the boxes are the 25th and 75th percentiles.

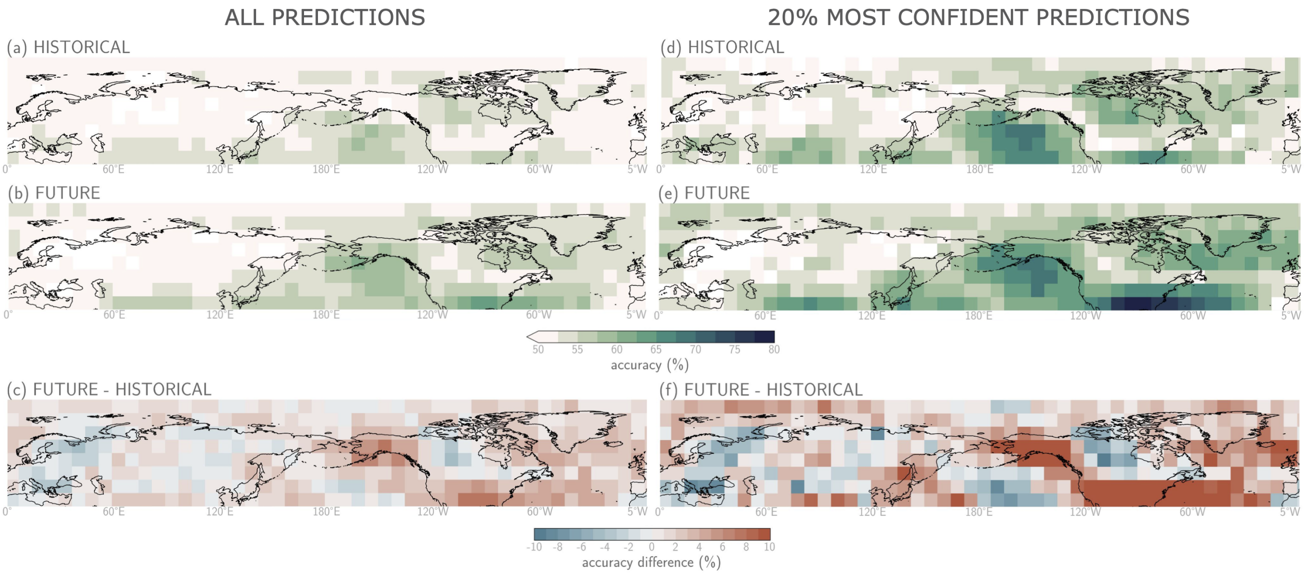

Figure S2. As in main text Figure 3, but with ensemble members #3-10 for training, member #2 for validation and member #1 for testing.

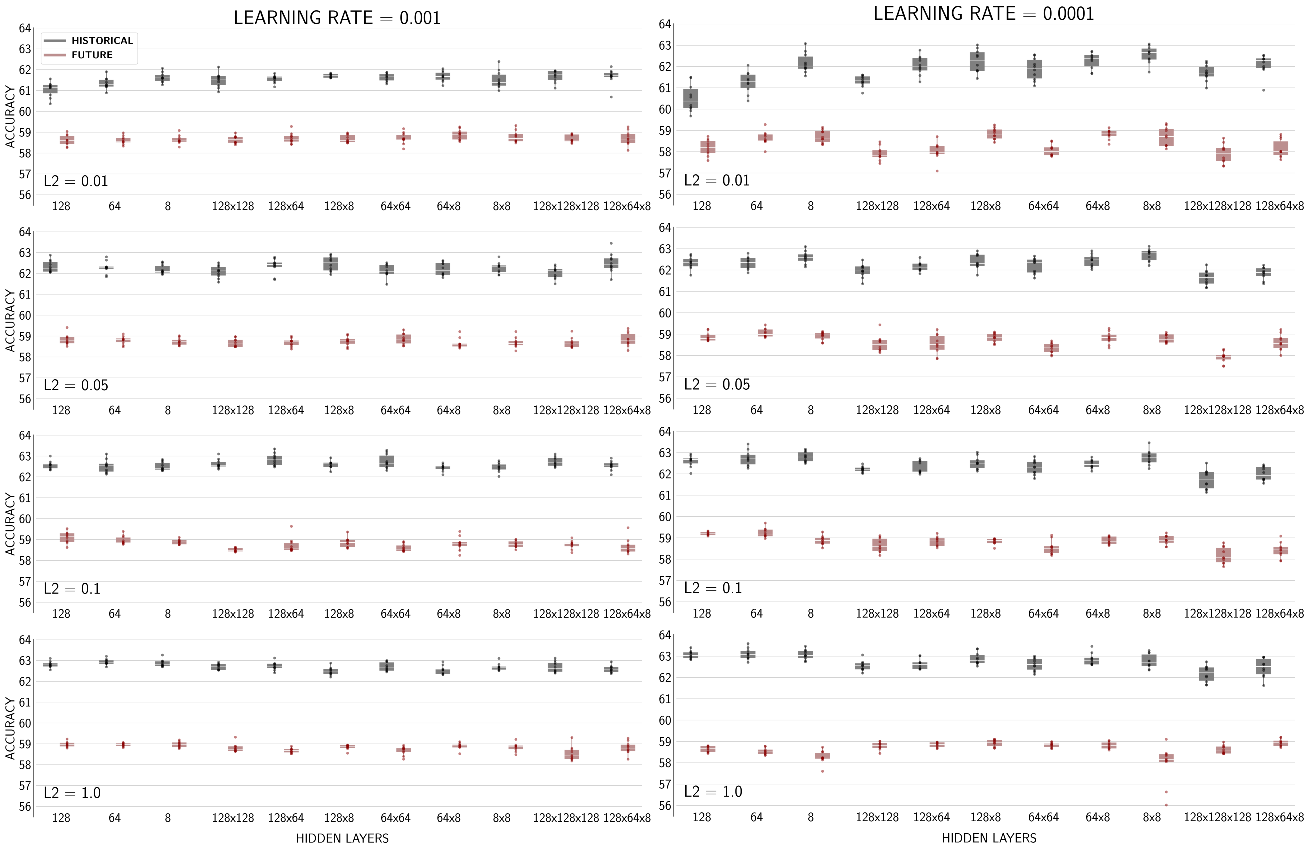

Figure S3. Validation (member #9) box and whisker plots of accuracies for 10 trained models in the North Pacific for variations combinations of the learning rate, ridge regression (L2), nodes per layer, and number of layers. Networks accuracies for a learning rate of 0.001 (0.0001) are in the left (right) column. Ridge regression values (denoted in the bottom left of each figure) increase from top to bottom and the network depth increases from left to right, where the number(s) represent the number of nodes per layer.

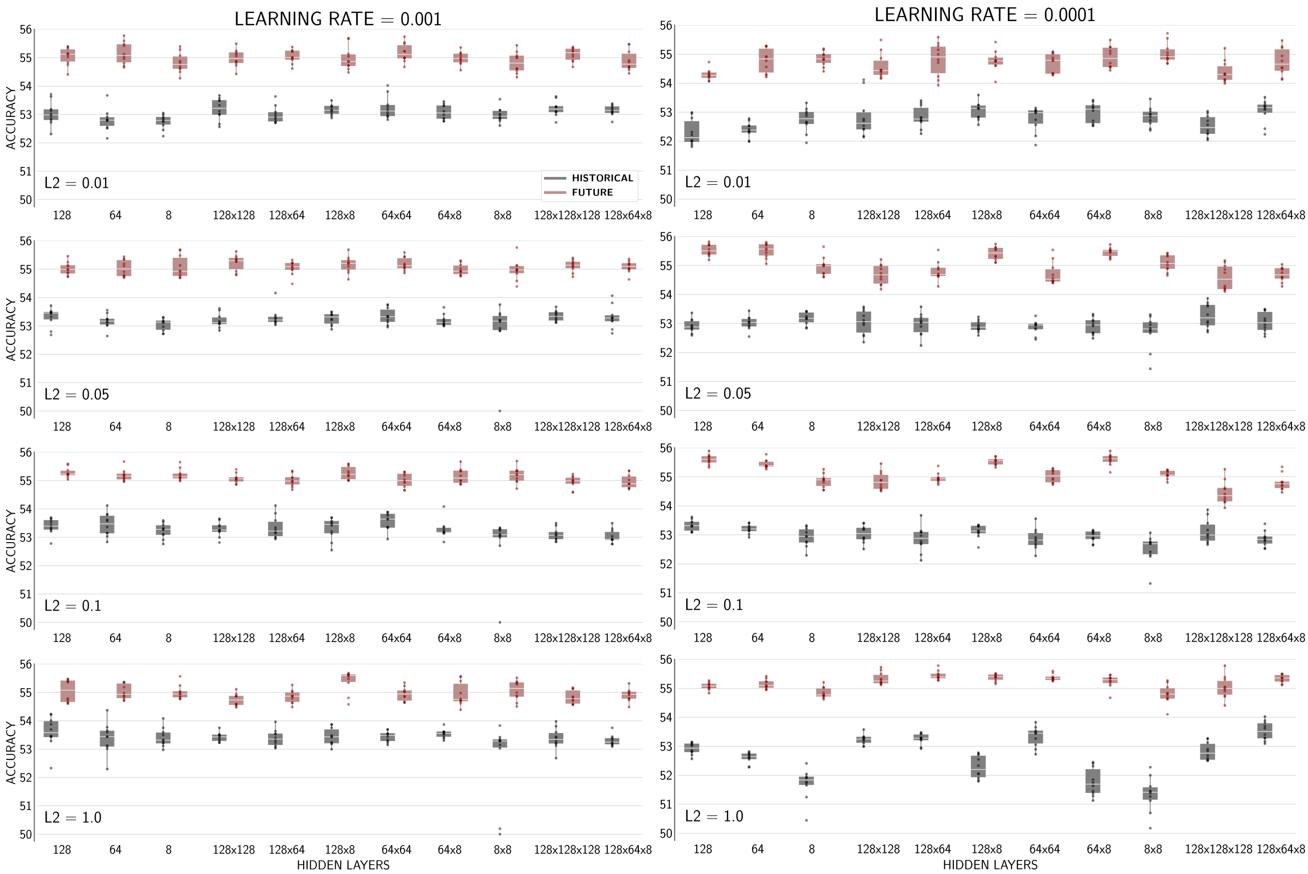

Figure S4. As in Figure S2, but for the North Atlantic.

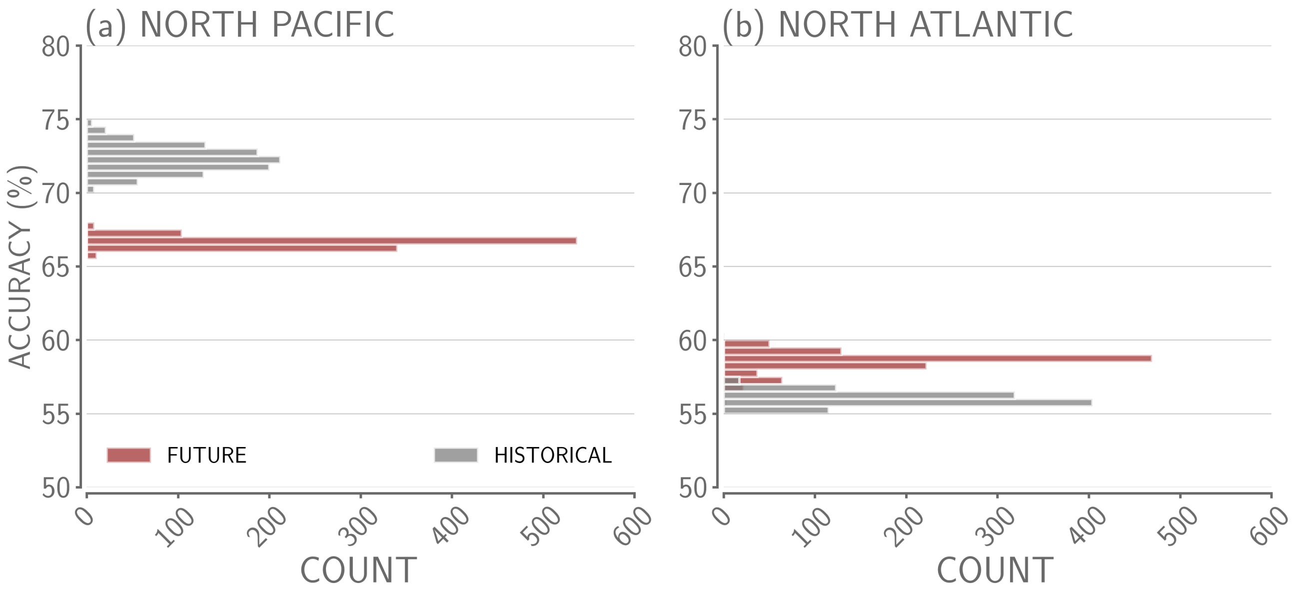

Figure S5. Histograms of bootstrapped top 3 models' mean 20% most confident testing accuracies with a bin size of 0.5% for (a) the North Pacific and (b) the North Atlantic, where grey and red refer to the historical and future, respectively.

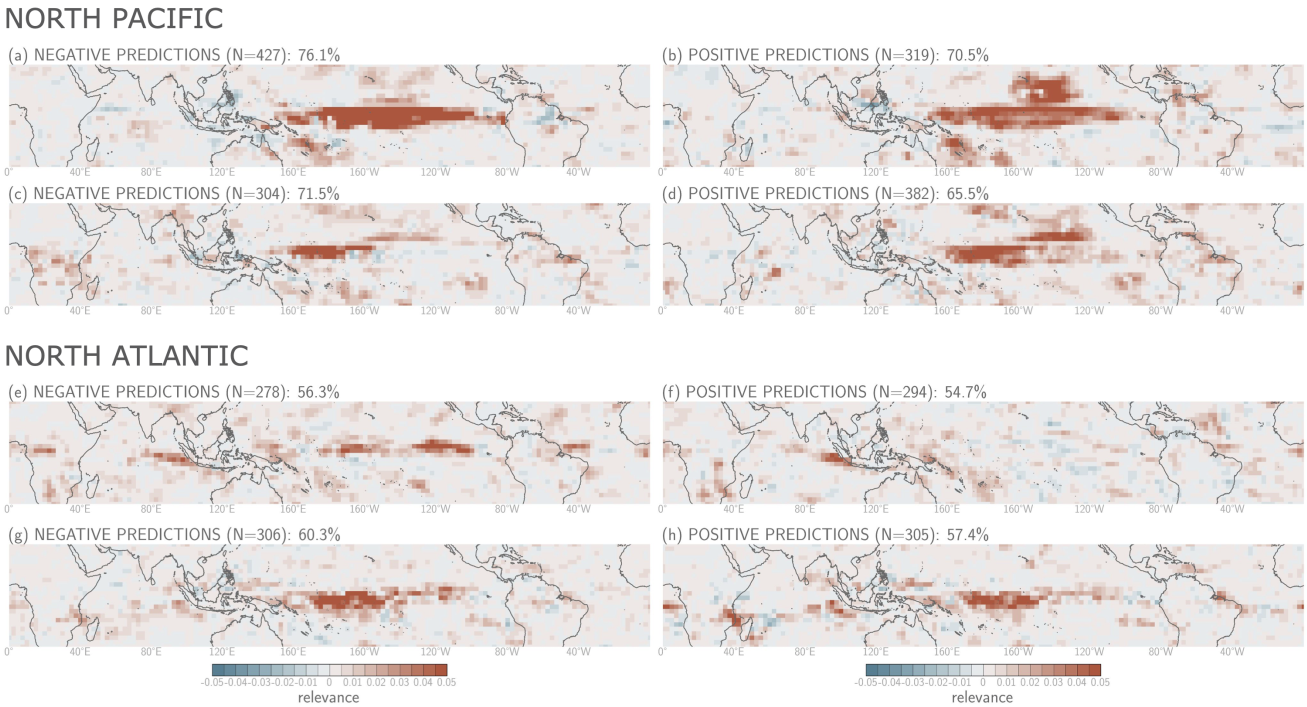

Figure S6. Example average layer-wise relevance plots for the 20% most confident and correct predictions in the North Pacific (a-d) and the North Atlantic (e-h). The top two panels for each locations (a-b, e-f) are the historical period and the bottom two panels for each location (c-d, g-h) are the future period. The left column includes heatmaps for the negative predictions and the right column includes heatmaps for the positive predictions. Red (blue) colors indicate the location had a positive (negative) contribution to the correct prediction. The percentage at the top of each panel is the conditional accuracy for each sign prediction and 'N' is the number of samples in each average.

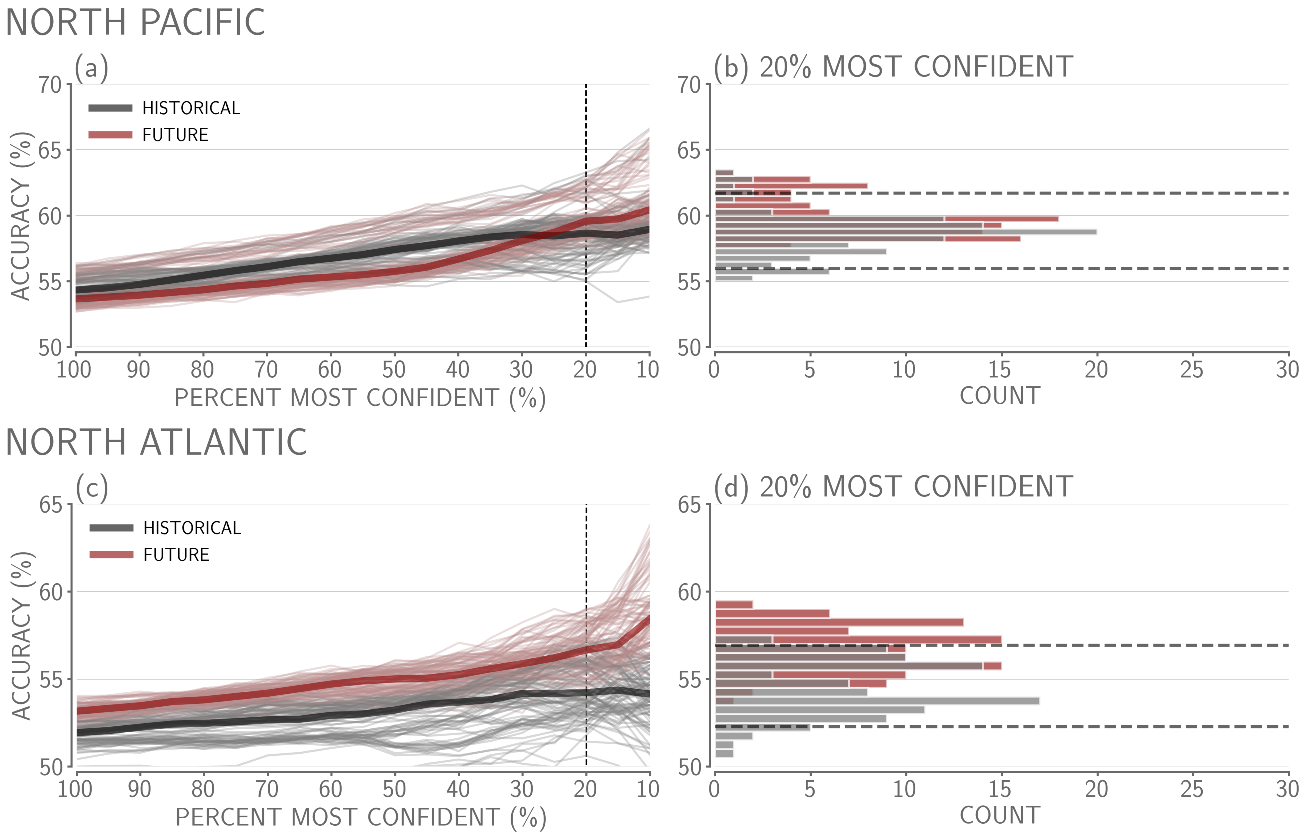

Figure S7. As in Figure 2 in the main text, but with 60+ day z500 anomaly variability removed from the predictand.

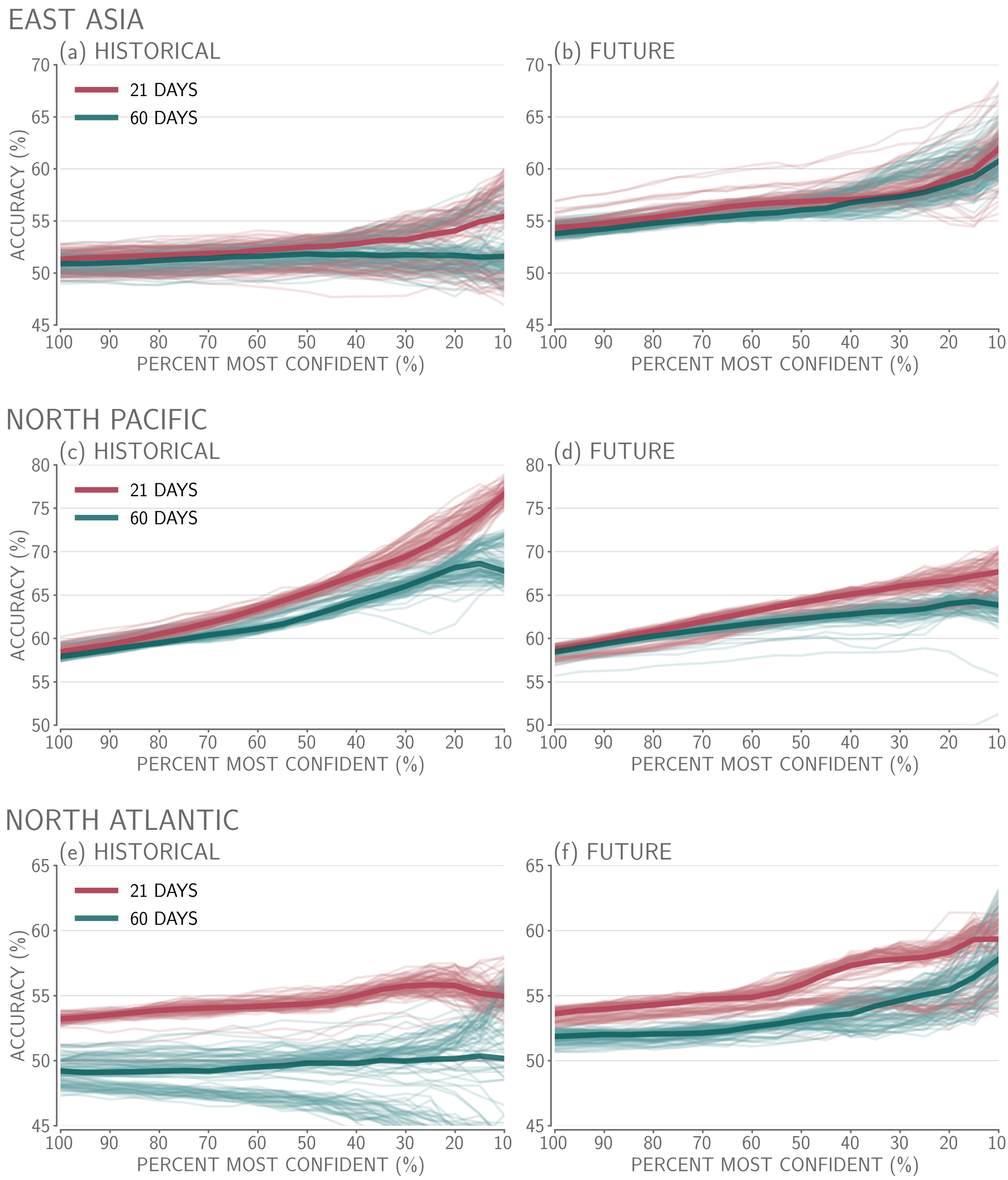

Figure S8. Accuracy versus confidence for 100 trained networks for the (left) historical and (right) future time period at leads of 21 (pink) and 60 (teal) days in (a,b) East Asia, (c,d) the North Pacific and (e,f) the North Atlantic. Accuracies are calculated using the testing member #10 and the thicker lines denote the median accuracy across the 100 networks at each confidence threshold. The pink lines are the same as the red/grey lines included in Figure 2 for the respective location and time period.

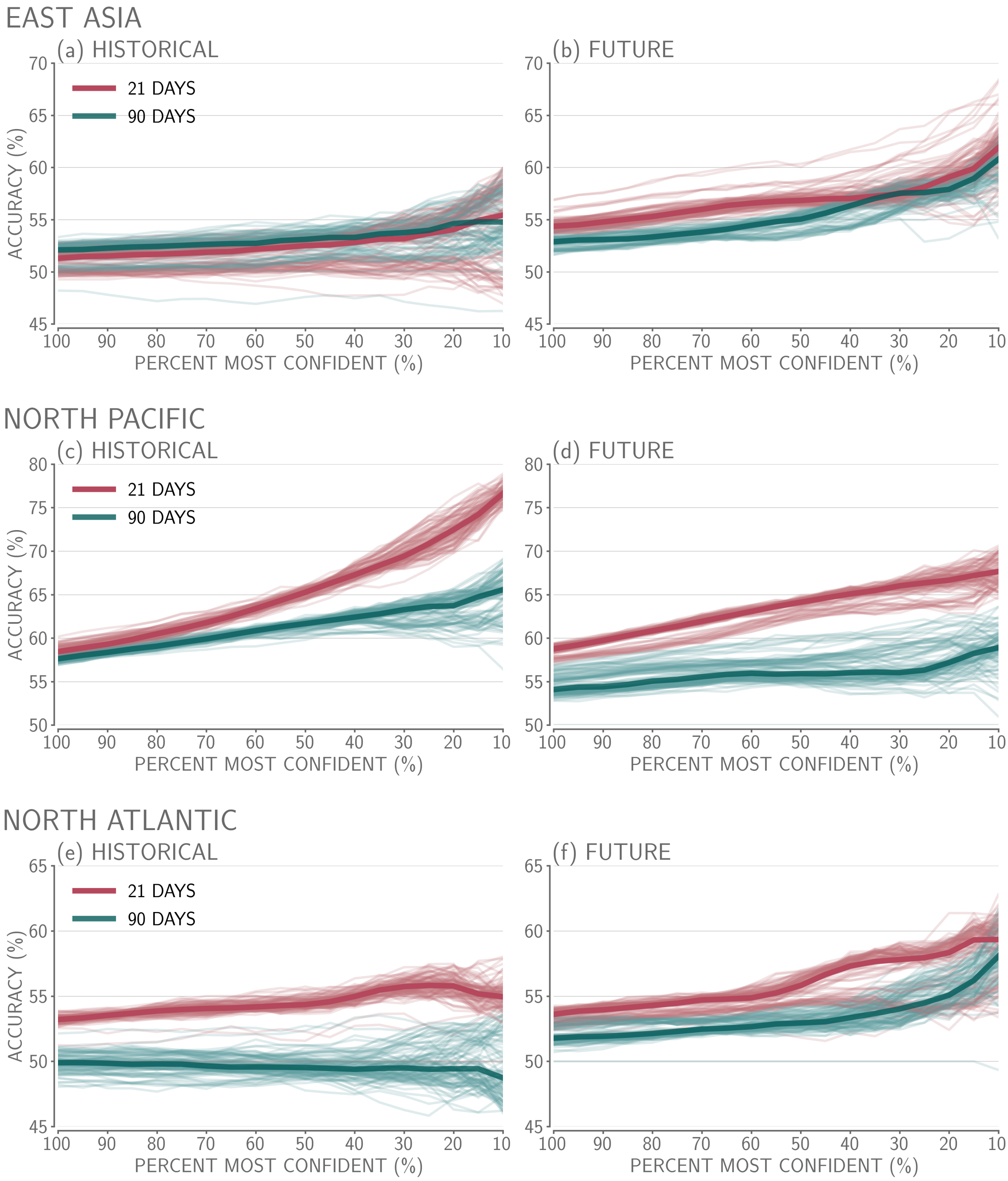

Figure S9. As in Figure S8, but for a lead of 90 days.