PAPER • OPEN ACCESS

Assessing decadal variability of subseasonal forecasts of opportunity using explainable AI

To cite this article: Marybeth C Arcodia *et al* 2023 *Environ. Res.: Climate* **2** 045002

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
CLIMATE

PAPER

Assessing decadal variability of subseasonal forecasts of opportunity using explainable AI

Marybeth C Arcodia[1,*] , Elizabeth A Barnes[1] , Kirsten J Mayer[1,2], Jiwoo Lee[3] , Ana Ordonez[3] and Min-Seop Ahn[3,4,5]

[1] Colorado State University, Fort Collins, CO, United States of America
[2] National Center for Atmospheric Research, Boulder, CO, United States of America
[3] Lawrence Livermore National Laboratory, Livermore, CA, United States of America
[4] NASA Goddard Space Flight Center, Greenbelt, MD, United States of America
[5] ESSIC, University of Maryland, College Park, MD, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: marcodia@rams.colostate.edu

Abstract

Identifying predictable states of the climate system allows for enhanced prediction skill on the generally low-skill subseasonal timescale via forecasts with higher confidence and accuracy, known as forecasts of opportunity. This study takes a neural network approach to explore decadal variability of subseasonal predictability, particularly during forecasts of opportunity. Specifically, this work quantifies subseasonal prediction skill provided by the tropics within the Community Earth System Model Version 2 (CESM2) Large Ensemble and assesses how this skill evolves on decadal timescales. Utilizing the networks' confidence and explainable artificial intelligence, physically meaningful sources of predictability associated with periods of enhanced skill are identified. Using these networks, we find that tropically-driven subseasonal predictability varies on decadal timescales during forecasts of opportunity. Further, we investigate the drivers of the low frequency modulation of the tropical-extratropical teleconnection and discuss the implications. Analysis is extended to ECMWF Reanalysis v5 data, revealing that the relationships learned within the CESM2-Large Ensemble holds in modern reanalysis data. These results indicate that the neural networks are capable of identifying predictable decadal states of the climate system within CESM2 that are useful for making confident, accurate subseasonal precipitation predictions in the real world.

1. Introduction

The subseasonal timescale, ranging from approximately two weeks to three months, is often described as a 'desert of predictability' due to its lack of forecast skill (Vitart *et al* 2017). Despite this issue, there is heightened demand from local management officials, water managers, and stakeholders for forecasts beyond the weather timescale to allow for longer lead times for preparation (NASEM 2016, Vitart *et al* 2017, Mariotti *et al* 2019). This has led to increased efforts to improve predictability through deeper understanding of subseasonal phenomena and the large-scale features that modulate them. Subseasonal predictability stems from knowledge of initial and boundary conditions that capture the coupled interactions which constrain chaotic processes in the atmosphere (Vitart 2017, Lovejoy 2018). Forecasters rely on predictable states of the climate system that can lead to periods of higher prediction skill within the subseasonal timescale, e.g. *forecasts of opportunity*, to improve subseasonal forecasts (Albers and Newman 2019, Mariotti *et al* 2020). These forecasts can be made by strategically leveraging time periods when certain large-scale oceanic and atmospheric conditions that provide sources of subseasonal predictability are present.

Efforts to recognize when and where forecasts of opportunity will occur often focus on identifying climate processes that lead to enhanced prediction skill and/or predictability. The tropics have been

identified as a key source of midlatitude predictability via tropical-extratropical teleconnections (Hoskins and Karoly 1981, Matthews *et al* 2004, Pegion and Kirtman 2008, Lau *et al* 2012). For example, the Madden–Julian Oscillation (MJO; Madden and Julian 1971) is the dominant mode of tropical intraseasonal variability, known to influence circulation and precipitation changes around the globe (Zhang 2005, 2013, Cassou 2008, Jones and Carvalho 2012, Ling *et al* 2017, Stan *et al* 2017). Specifically, midlatitude subseasonal variability linked to the MJO affects precipitation in Alaska (Zheng *et al* 2018), the Pacific-Northwest (Bond and Vecchi 2003), and California (Arcodia *et al* 2020). Additionally, MJO teleconnections in the United States can be modulated by lower frequency modes such as ENSO (Moon *et al* 2011, Tseng *et al* 2020, Arcodia and Kirtman 2023). When active, the combined influence of the MJO and ENSO conditions can provide subseasonal forecasts of opportunity for midlatitude precipitation (Newman *et al* 2003, Johnson *et al* 2014, DeFlorio *et al* 2019, Arcodia *et al* 2020).

Seasonal-to-interannual variability, such as ENSO, and its impact on subseasonal predictability has also been explored in recent years (Lee *et al* 2019b, Huang *et al* 2021), however, lower frequency modes of variability on decadal timescales and its impact on subseasonal predictability is much less studied. Some previous work has explored how decadal variability influences seasonal midlatitude precipitation variability (Cayan *et al* 1998, Gershunov and Barnett 1998, McCabe and Dettinger 1999, Dai 2013). In addition, prior efforts have identified decadal modulation of MJO teleconnections (Kim 2020) and low frequency variability in subseasonal predictability of the North Atlantic Oscillation (Scaife *et al* 2014, O'Reilly *et al* 2017, Weisheimer *et al* 2017, Albers and Newman 2021, Luo *et al* 2022). However, this raises the question as to whether there is a subsequent impact of decadal variability on midlatitude subseasonal predictability provided by the tropics.
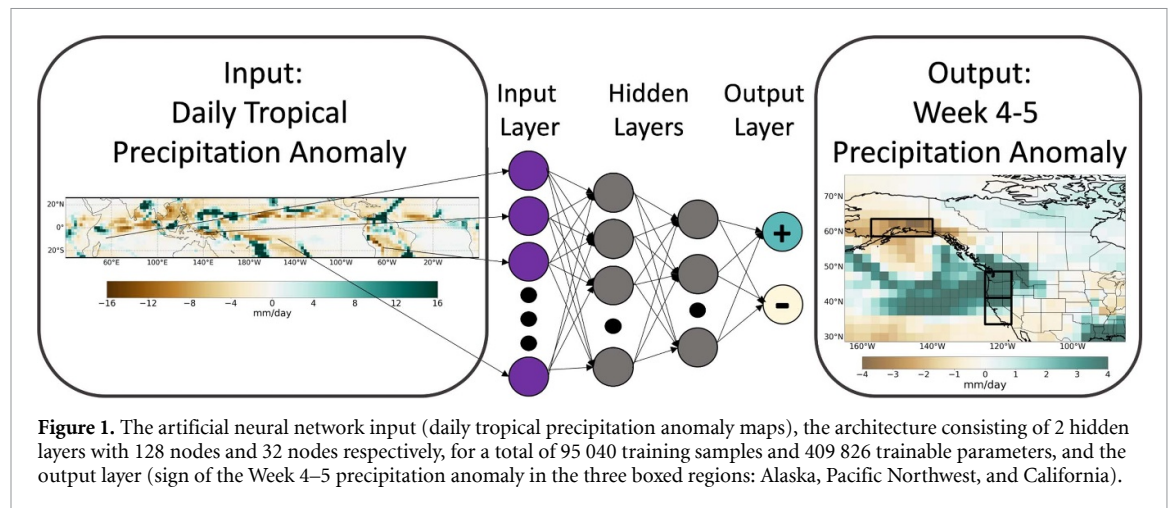
Recent improvements to subseasonal forecast skill have been attributed to identifying forecasts of opportunity based on improved initial conditions and active climate modes in official forecasting systems (Mariotti *et al* 2020). Additionally, improvements to forecasting models have enhanced subseasonal prediction skill, which can be attributed to higher resolution and better physical parameterizations (Pegion *et al* 2019, Meehl *et al* 2021). Here, we seek to explore the role of decadal variability on regional subseasonal prediction skill in the recent decades. It is important to identify and better understand the impact of decadal variability on subseasonal predictability, as this variability could affect the performance of these models in the future. Unfortunately, this investigation remains a challenging task due to the lack of historical data and historical hindcasts needed to assess variability of subseasonal predictability on decadal timescales.

To combat this issue, we utilize the Community Earth System Model Version 2—Large Ensemble (CESM2-LE; Danabasoglu *et al* 2020) to analyze 1000 years of climate model data. Specifically, we use the CESM2 historical simulation to train neural networks to quantify the subseasonal prediction skill within CESM2. The neural network approach allows us to make subseasonal predictions without the need for a forecast model. The fields of weather and climate science have benefited from applications of neural networks to learn physical relationships within the Earth system via explainability techniques (Ham *et al* 2019, McGovern *et al* 2019, Toms *et al* 2020, Antonios *et al* 2021, Gordon *et al* 2021, Martin *et al* 2022, Mayer and Barnes 2022, Straaten *et al* 2023). Here, we use shallow artificial neural networks (ANNs) to quantify U.S. West Coast subseasonal precipitation skill and identify forecasts of opportunity provided by the tropics within CESM2. We assess the low frequency variability of the subseasonal prediction skill, finding decadal variability in skill particularly during forecasts of opportunity. The European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) data are then input into the neural networks to make predictions based only on learned CESM2 relationships. The ERA5 results show skill above random chance, with enhanced skill for the forecasts of opportunity that oscillate on decadal timescales throughout the modern era.

2. Data and methods

2.1. Climate model data
Training ANNs to learn predictive subseasonal signals amongst noisy climate data is a data-intensive process (Rosa *et al* 2020). Thus, we take advantage of historical simulations performed as part of the CESM-2 LE (CESM2-LE; Danabasoglu *et al* 2020). Specifically, we analyze ten members which were selected to sample the four pre-selected initial states for different AMOC states and include smoothed biomass burning. Results were ultimately not sensitive to the initialized AMOC state (not shown) and so the ten members should be generally viewed as ten largely independent realizations of the same historical climate. We found ten ensemble members to be sufficient, which is consistent with Mayer and Barnes (2021) who found ten members sufficient for a similar problem. Additional information about the ensemble members used can be found in the Supplemental Table 1.

Figure 1. The artificial neural network input (daily tropical precipitation anomaly maps), the architecture consisting of 2 hidden layers with 128 nodes and 32 nodes respectively, for a total of 95 040 training samples and 409 826 trainable parameters, and the output layer (sign of the Week 4–5 precipitation anomaly in the three boxed regions: Alaska, Pacific Northwest, and California).

Daily precipitation data are interpolated from a 1 by 1 degree resolution to 2.5 by 2.5 degree resolution via bilinear interpolation and selected from 1850 to 1949. We use data from November-March since boreal winter is known to have strong tropical-extratropical teleconnections in the Northern Hemisphere (Weickmann 1983, Soulard *et al* 2021). The use of daily data allows networks to learn relevant low frequency features within the input without a priori smoothing. Daily precipitation anomalies are calculated by subtracting the linear trend of the ensemble mean for each day of year from each grid point independently to retain only internal variability. Similar analyses were performed using the pre-industrial control runs which yielded comparable behavior (not shown), confirming the effective removal of the forced response.

We train a set of ANNs (more details provided in section 2.3) to ingest a map of daily tropical precipitation anomalies and classify the sign (i.e. positive or negative) of the midlatitude subseasonal precipitation anomaly (figure 1). The inputs (predictors) to the neural network are maps of the daily tropical precipitation anomalies from 26° S to 26° N where each grid point represents an individual node in the input layer. The outputs (predictands) are the sign of precipitation averaged over days 22–35 (Week 4–5) after the input day. For example, the input map for 1 November 1850 is used to predict the sign of the averaged 23 November–6 December 1850 precipitation anomaly in the three predicted regions. Next, the input map for 2 November 1850 is used to predict the sign of the 24 November–7 December 1850 anomaly, and so on. The output data is area-averaged over coastal Alaska (58.75°–63.75° N; 202.5°–220° E), the Pacific Northwest (41.25°–48.75° N; 235°–242.5° E), and California (33.75°–41.25° N; 235°–242.5° E) as shown in figure 1.

Ten sets of experiments are conducted. For each experiment, ten neural networks are trained on 8 ensemble members (the training set) with 100 years of data each, and each of the ten networks is initialized with a different random seed to sample uncertainty in the network initialization and ensure robustness of the results. Each trained neural network is then validated on one of the remaining members and tested on the other. The training, validation, and testing members are then reorganized so a new set of eight members are used to train the ten neural networks (each initialized with ten random seeds) that are validated on one member, and tested on the remaining member. This process is repeated 10 times, so that each ensemble member is used as a test member. With ten random initializations per network, this results in a set of 100 trained neural networks.

Prior to training the neural networks, the input data are normalized by subtracting the mean and dividing by the standard deviation of each grid point across the training/validation data. As a note, the daily anomalies are calculated with the training and validation members used for each set of neural networks to ensure that no information about the testing data is given to the neural networks during training. The output data is normalized by removing the median of the training/validation output data, then converted to binary to represent the sign of the anomalies (0 s for a negative prediction; 1 for a positive prediction). The training and the validation output data are class balanced by randomly removing samples so that there are equal numbers of 0 s and 1 s.

2.2. Reanalysis data

We also use precipitation from the ERA5 (Hersbach *et al* 2020) dataset to evaluate the neural networks trained on the CESM2-LE. We use November-March and interpolate the precipitation fields to 2.5 by 2.5 degree resolution from a 0.25 by 0.25 degree resolution via bilinear interpolation in the same way as the CESM2 data. However, due to data availability, we take the 00:00 time step at each grid point and our ERA5 analysis spans a later time period: 1959–2021. The trend is calculated by fitting a 3rd order polynomial to

each day of year to capture any climate change signal, particularly in recent history. However, we found results to be fairly insensitive to the method of computing the anomalies (e.g. removing a linear vs. a polynomial trend). The ERA5 daily precipitation anomalies are then calculated via subtraction of the trend. While removing the trend here does not capture the true forced response as was done with the CESM2 ensemble mean, removing the trend is sufficient for removing the forced response for a reanalysis dataset where multiple ensemble members are not available. We tested the sensitivity of the choice of trend and dataset used to compute anomalies for each dataset, and found the results to be largely insensitive to these choices. Prior to inputting the ERA5 fields to the trained network, the data are normalized by subtracting the mean of the ERA5 data and dividing by the standard deviation of the ERA5 data at each grid point. This is done to ensure that the reanalysis and CESM2 data are in data-relative units prior to being input into the neural network. The standardized ERA5 data are converted to binary labels after the median is removed.

2.3. Neural network architecture
The ANN architecture is depicted in figure 1. The ANN consists of two hidden layers, the first with 128 nodes and the second with 32 nodes, both of which use a rectified linear activation function (ReLU). The final output layer consists of two nodes, representing our two classes: a negative or positive sign. This final layer uses the softmax function so that the two output values sum to one and thus represent an estimation of probability. This probability is then used to choose the predicted output, where the output with a probability greater than 0.5 is chosen as the selected class. This probability is also deemed our 'network confidence' following the methodology of Mayer and Barnes (2021, 2022).

The network is trained using a categorical cross-entropy loss function with a batch size of 512 samples, a ridge regression coefficient of 5.0 (to minimize overfitting), and a learning rate of 6.7×10^{-6}. Early stopping is also applied if validation loss does not decrease after 25 epochs. The 10 random seeds used are 210, 47, 33, 133, 410, 692, 64, 99, 910, 92. Hyperparameter tuning was performed using the KerasTuner (O'Malley *et al* 2019) to find the optimal set of hyperparameters across all regions. The same set of hyperparameters is used for all 100 neural networks. Additional information on the hyperparameter tuning can be found in the supplemental table 2.
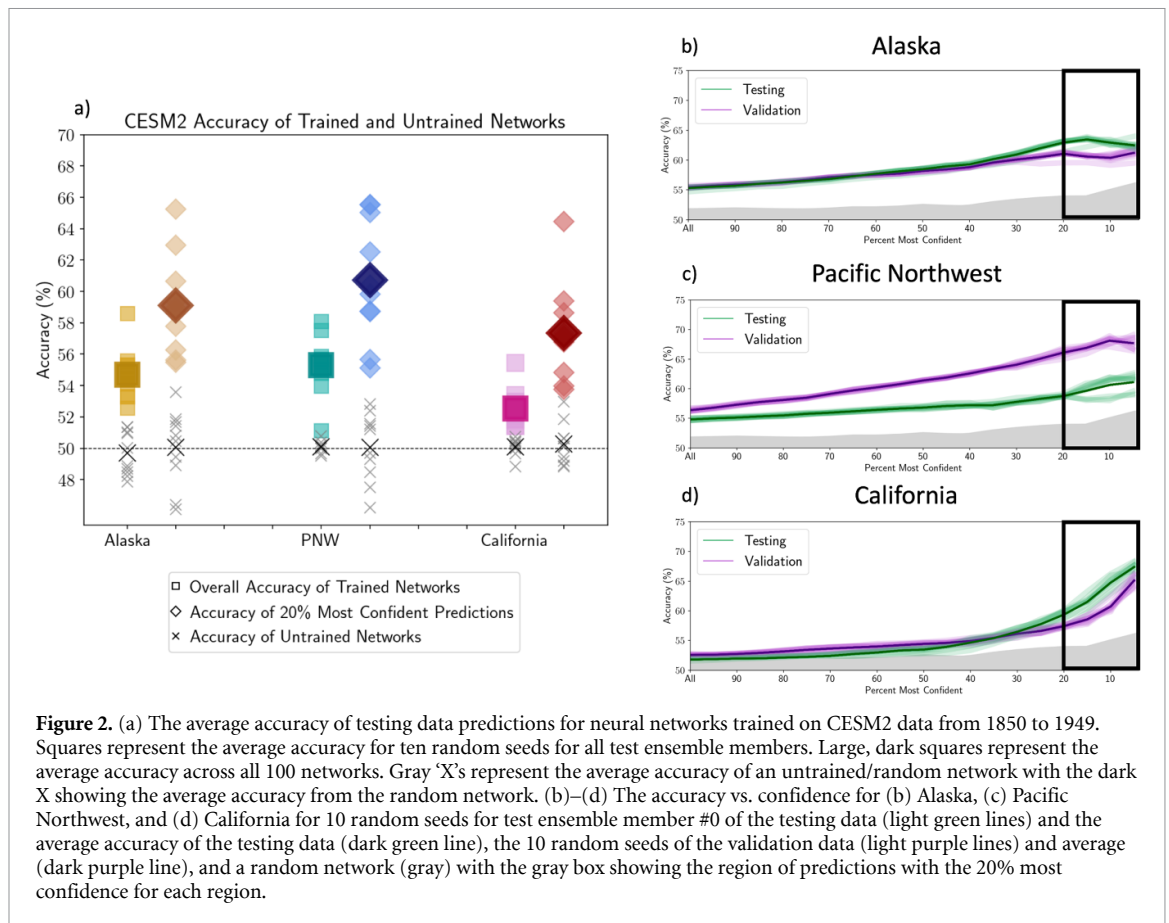
2.4. Explainable AI
Explainable artificial intelligence (XAI) methods have been utilized to open the 'black box' of machine learning to gain insights about the climate system (McGovern *et al* 2019, Antonios *et al* 2021, 2022, Gordon *et al* 2021, Mayer and Barnes 2021, Sonnewald *et al* 2021, Pegion *et al* 2022, Labe and Barnes 2022). Here, we use an XAI attribution method known as integrated gradients (Sundararajan *et al* 2017). Integrated gradients is an extension of the Input*Gradient method derived from the gradient/saliency method (Baehrens *et al* 2010) which explains the network's decision by computing the first partial derivative of the network's output with respect to the input. An input has a high relevance value if the network used that feature to make its prediction. Integrated gradients is used here as it is a nonlinear version of the Input*Gradient method and can identify properties of nonlinear relationships (Bommer *et al* 2023).

An important choice when using integrated gradients is the baseline, from which all attributions are computed (Antonios *et al* 2021). For this study, we choose a baseline input of zeros, since the daily anomalies are centered around zero by construction. We note that the corresponding network output for a zero input is roughly 50% probability for both classes, implying that our baseline for an input of all zeros is uninformative when it comes to precipitation prediction. This behavior is preferable for classification problems, and is intuitive for our problem in that zero anomalies (i.e. the mean) hold no predictive information (Sundararajan *et al* 2017, Antonios *et al* 2023).

3. Results

3.1. Forecasts of opportunity
The neural networks are designed to take a map of daily tropical precipitation anomalies as input and classify the sign of the Week 4–5 precipitation anomaly in a particular U.S. region. Here, we focus our results on three regions known to be influenced by tropical teleconnections on subseasonal timescales: Alaska, the Pacific Northwest, and California. For each experiment (identified by the member used for testing), 10 neural networks are initialized with different random seeds and the average test accuracy across the ten networks is computed and shown in figure 2(a) (squares). The mean of the 10 averaged accuracies, i.e. the mean of all 100 trained neural networks, is also shown in figure 2(a) (large, dark squares). These accuracies are compared to an *untrained* network in which the testing data was used to make predictions with neural networks that were initialized with random weights, but not trained. The accuracy of the untrained networks can then be compared to that of the trained networks to assess if the trained networks resulted in higher

Figure 2. (a) The average accuracy of testing data predictions for neural networks trained on CESM2 data from 1850 to 1949. Squares represent the average accuracy for ten random seeds for all test ensemble members. Large, dark squares represent the average accuracy across all 100 networks. Gray 'X's represent the average accuracy of an untrained/random network with the dark X showing the average accuracy from the random network. (b)–(d) The accuracy vs. confidence for (b) Alaska, (c) Pacific Northwest, and (d) California for 10 random seeds for test ensemble member #0 of the testing data (light green lines) and the average accuracy of the testing data (dark green line), the 10 random seeds of the validation data (light purple lines) and average (dark purple line), and a random network (gray) with the gray box showing the region of predictions with the 20% most confidence for each region.

accuracy than the untrained networks or random chance. For each of our three regions, the average accuracy of the untrained networks (crosses) is approximately 50%, as expected. The average accuracy from the trained (squares) are distinctly higher than the average accuracy of the untrained networks, indicating that the neural networks exhibit higher prediction skill than random chance.

We leverage the network's confidence in each prediction to identify forecasts of opportunity. In figures 2(b)–(d), the confidence values associated with each prediction for one experiment per region are ranked by percentile, ranging from 'All' (e.g. 100% of the predictions; left) to the top 10% (right). Shading indicates the accuracy versus confidence for each of the untrained networks initialized with a different random seed, with the darker line indicating the average of the random seeds.

For each region, the accuracy increases as confidence increases (figures 2(b)–(d)). This is particularly evident when compared against the accuracy as a function of confidence for the untrained networks (gray lines). The accuracy of the untrained networks remains fairly constant as confidence increases, with a slight increase in accuracy for the top 10% of predictions due to higher variance from a smaller sample size. Supplemental figure S1 shows the same as figures 2(b)–(d) but for all ten experiments for each region. Since the accuracy of our trained neural networks increases as confidence increases, the most confident predictions are also those predictions with the highest accuracy, and thus, are evidence of forecasts of opportunity. From here on out, we define forecasts of opportunity as the 20% most confident predictions, however, our conclusions are insensitive to different confidence intervals selected ($\pm$10%). The accuracy of the forecasts of opportunity are plotted in figure 2(a) as diamonds for the three predictand regions, with the average of the accuracies consistently higher for the forecasts of opportunity (diamonds) than the average accuracy for all predictions (squares).

We then assess the output to determine if the sign of the predicted anomaly varies within a season, ensuring the networks were making subseasonal predictions, rather than seasonal predictions. A time series showing the subseasonal variability of the predictions is shown in figure 3, in which the predictions vary between negative and positive within one season. We highlight forecasts of opportunity with darker/larger circles, showing that the network tends to be most confident in either a negative or positive prediction for each season. For example, in the 1875–1876 winter the most confident predictions were all positive, but in the 1876–1877 winter, the confident predictions were all negative. Therefore, the networks demonstrate subseasonal variability in the anomaly sign prediction, but the most confident predictions are typically of the same sign within a season.
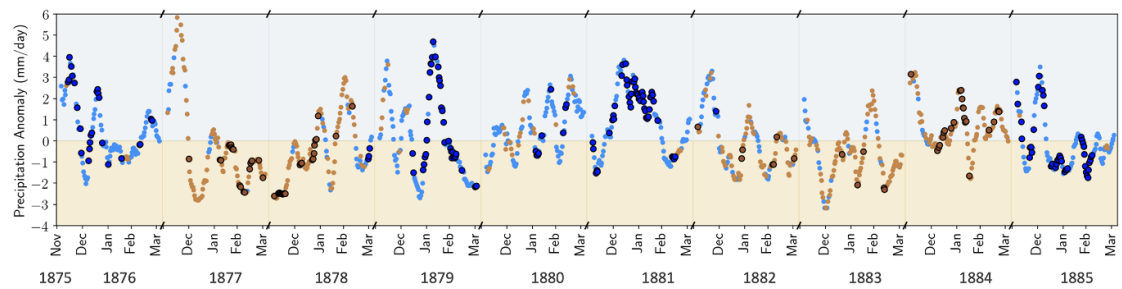
Figure 3. Time series of precipitation anomalies in Alaska from test ensemble member #9. Brown dots indicate a negative anomaly prediction. Blue dots indicate a positive anomaly prediction. Brown dots in the negative region shaded brown indicate a correctly predicted negative anomaly. Blue dots in the positive region shaded blue indicate a correctly predicted positive anomaly. Darker/larger dots outlined in black indicate the 20% most confident predictions as identified by the neural network.
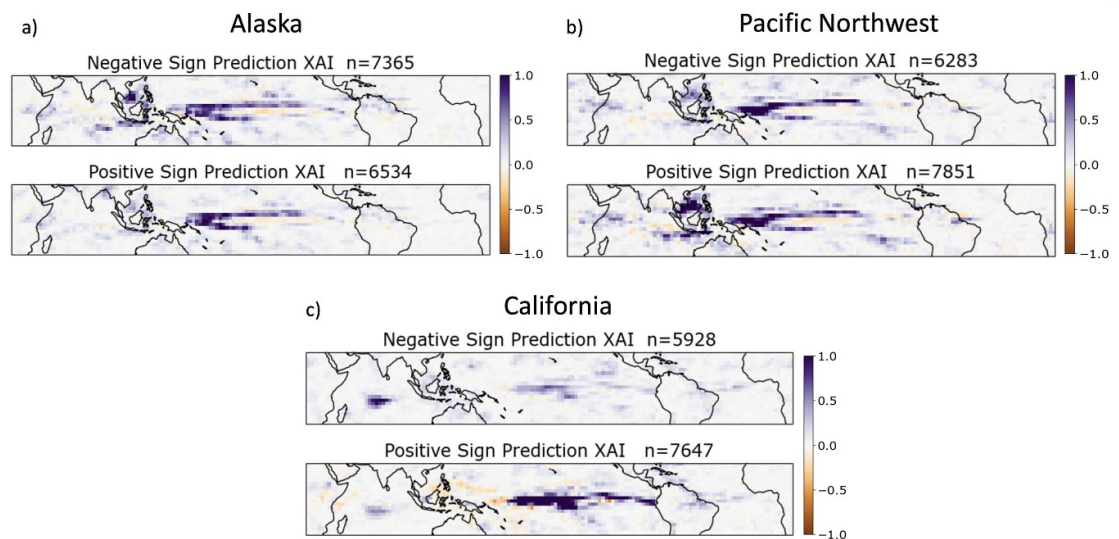

Figure 4. Integrated gradients XAI composited heatmaps for (a) Alaska, (b) Pacific Northwest, and (c) California for the 20% most confident and correct negative (top) and positive (bottom) predictions. The number (n) represents the number of samples in each composite. Darker purple colors represent grid cells that were more relevant to the prediction, orange/brown colors represent adverse contributions to the predictions. Colorbars indicate a unitless measure of relevance.

These results demonstrate that the neural networks perform better than random chance and can identify features in the input that makes the network more confident (and accurate) in its prediction. Next, we explore the sources of predictability the neural network is utilizing to make its confident and correct predictions in each region. We do this to ensure the neural network is identifying physically meaningful regions for subseasonal precipitation prediction along the West Coast of the U.S. before further exploring if the accuracy of these networks varies on decadal timescales.

3.2. Identifying sources of subseasonal predictability

Explainable AI techniques are powerful tools that allow us to not only explain the network's decision-making process, but also to identify sources of predictability. Here, we use the XAI technique, integrated gradients, described in section 2.4 to locate tropical sources of subseasonal predictability for U.S. West Coast precipitation. We calculate integrated gradients heatmaps for each input map that correspond to a correct forecast-of-opportunity prediction, e.g. a prediction that was correct and in the 20% most confident range. The heatmaps are composited separately for negative and positive predictions, shown in figure 4. Heatmaps were also clustered using k-means clustering using 2–4 clusters (not shown), but sources of predictability were similar across the clusters compared to the composited (1 cluster) heatmaps.

The XAI heatmaps shown in figure 4 depict the regions of the input that are most relevant to the network's correct, confident predictions. Regions with the highest values (e.g. darkest shading) highlight potential sources of predictability. In both Alaska and the Pacific Northwest, the tropical western Pacific region is important for the networks' predictions, with off-equatorial bands extending east into the central Pacific. This pattern is similar to the precipitation patterns related to ENSO documented in CESM2

(Capotondi *et al* 2020). Furthermore, the heatmaps reveal that the northern Maritime Continent region is relevant for negative sign predictions in Alaska and positive sign predictions in the Pacific Northwest. The Maritime Continent region is often associated with the MJO (Zhang 2005, Ahn *et al* 2020, Daehyun *et al* 2020, Kang *et al* 2021), although the MJO-associated rainfall is typically more centered around the equator than the XAI heatmap highlighted region. We note that MJO convection is well represented in CESM2 (Danabasoglu *et al* 2020), and do not attribute the off-equatorial region to be associated with a bias in CESM2. Since the neural networks are tasked with predicting precipitation anomalies at 21–35 d lead time, and it generally takes about 14 d for Rossby wave trains to be established from the tropics to the extratropics (Hoskins and Karoly 1981), the expected teleconnection would be that associated with the previous one or two phases of the MJO. The Maritime Continent region is associated with active convection in MJO phases 4 and 5, which is consistent with Pacific Northwest positive precipitation anomalies in phases 2 and 3. This region is also associated with suppressed convection in phases 8 and 1, consistent with negative Pacific Northwest precipitation anomalies in phases 6 and 7, with variations based on early or late winter season (Bond and Vecchi 2003).

For predictions over California, a major tropical source of predictability is located predominantly in the central equatorial Pacific and is most relevant for the positive sign predictions, whereas the central Indian Ocean is found to be a more prominent source of predictability for negative sign predictions. This region is associated with active convection during MJO phase 2 and 3 which is consistent with negative rainfall anomalies during MJO phases 4 and 5 in California particularly during El Niño events (Arcodia *et al* 2020).
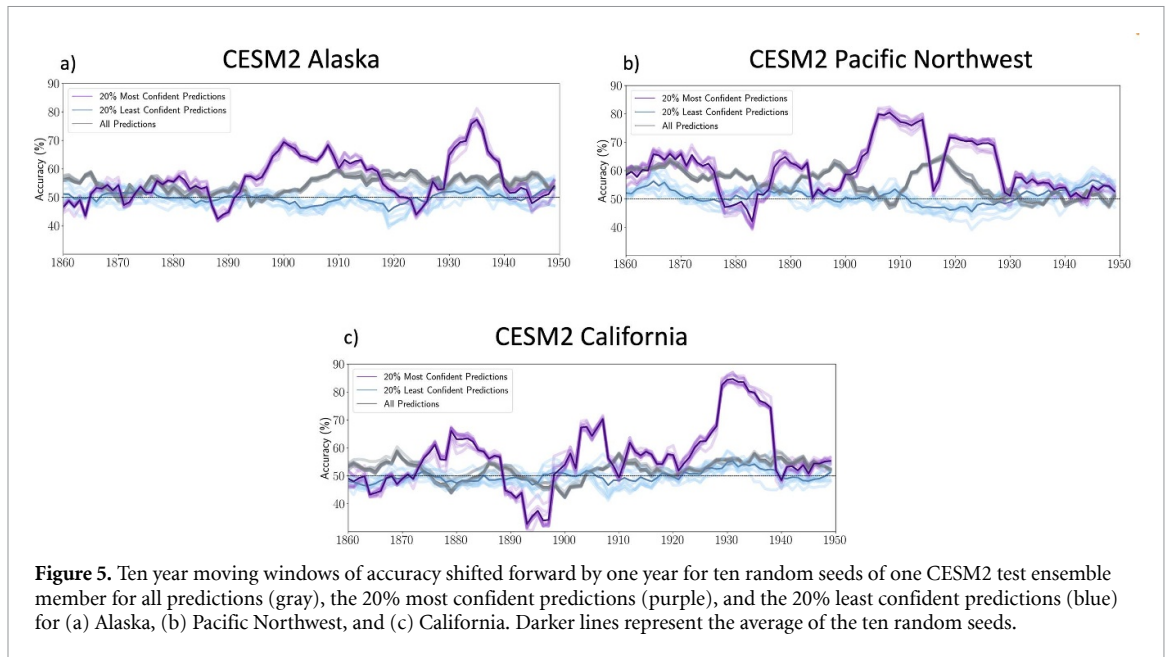
Additionally, we computed the composites of integrated gradients heatmaps for the incorrect 20% most confident predictions (supplementary figure S2). We found the regions providing adverse contributions to the predictions (yellow/brown colors) to be the inverse of those providing relevant contributions to the 20% most confident and *correct* predictions (see figure 4). Thus, the sources of predictability provide relevant information for the network to make confident and correct predictions, but also are found to be the source regions responsible for the network to make a confident but incorrect prediction.

3.3. Decadal variability of subseasonal predictability

The XAI analysis has identified physically relevant tropical sources of predictability that allow us to build trust in the neural networks predictions and its ability to identify forecasts of opportunity. These predictions are next assessed to determine if the subseasonal predictability quantified by the networks varies on decadal timescales, with periods of higher and lower skill than the overall averaged accuracy. The accuracy of the networks' predictions is calculated over a ten year moving window (figure 5), as opposed to the full 100 year period for which the accuracy was calculated in figure 1. For example, the accuracy of all test members for each experiment is first calculated over a ten year period (1851–1860), then shifted forward one year (1852–1861), etc to create the low-frequency time series of the accuracy in each region (gray lines). The time series is centered around the ten year period for which the accuracy is calculated (e.g. the point corresponding to 1860 was calculated over 1856–1865). Figure 5 shows the accuracy time series for one test ensemble member for each region, with each individual line representing one random seed and the dark line representing the average over the random seeds. Supplemental figure figures S2–S4 shows the time series for each of the ten testing ensemble members. The same moving averaging process is repeated for the 20% most confident predictions (purple lines) and the 20% least confident predictions (blue lines). That is, the 20% most confident and least confident samples are selected from each ten year window, and then the accuracy of those samples is computed.

The variability of the 20% least confident predictions (blue lines) in figure 5, show little decadal variability with accuracies remaining consistent around 50% (e.g. random chance). Decadal variability of the 20% most confident predictions (purple lines) is much more pronounced than the accuracy of all predictions (gray lines), suggesting the decadal variability in accuracy from all predictions is mainly from the decadal variability of the accuracy of confident predictions. For all three regions, the ten year averaged accuracy ranges between approximately 40%–80%. Based on the sources of predictability identified in figure 4, the forecasts of opportunity occur when certain conditions are present in the highlighted regions, but the accuracy of these forecasts of opportunity varies on decadal timescales. For example, during the late 1890s in Alaska, the ten year average accuracy was over 70%, but the accuracy dropped to almost 40% (worse than random chance) during the late 1880s. Similar low frequency variability in accuracy is found in the Pacific Northwest and California during forecasts of opportunity. This variability implies that there is a low frequency modulation of subseasonal prediction skill for midlatitude precipitation driven by the tropics.

We investigate the drivers of the low frequency variability of the subseasonal prediction skill, focusing explicitly on forecasts of opportunity. Although the 20% most confident predictions have the highest accuracy overall (figure 2), these predictions also exhibit the largest variability with periods of both the highest and lowest skill (figure 5). The daily tropical precipitation input maps corresponding to the 20%

Figure 5. Ten year moving windows of accuracy shifted forward by one year for ten random seeds of one CESM2 test ensemble member for all predictions (gray), the 20% most confident predictions (purple), and the 20% least confident predictions (blue) for (a) Alaska, (b) Pacific Northwest, and (c) California. Darker lines represent the average of the ten random seeds.

most confident predictions were composited by sign of the true anomaly in the predicted region and by maps which resulted in correct and incorrect predictions, resulting in four composite maps for each region (figure 6). In California, correct predictions for a negative anomaly correspond to negative precipitation anomalies in the tropical Pacific, typically associated with a La Niña precipitation pattern (Dai and Wigley 2000). Correct predictions for a positive anomaly correspond to the opposite pattern with positive tropical precipitation anomalies, typically associated with an El Niño precipitation pattern (Dai and Wigley 2000). However, when the network is confident and *incorrect*, the network uses the learned relationship between the tropics and the sign of mid latitude precipitation, but the true sign is now opposite. In other words, when a La Niña-type pattern is present in the equatorial Pacific, the network will confidently predict a negative anomaly, but the true anomaly will be positive, and vice versa for an El Niño-type pattern. A similar relationship is found for Alaska while an opposite-signed relationship is found in the Pacific Northwest: an El Niño-like pattern corresponds to a negative sign prediction and a La Niña-like pattern corresponds to a positive sign prediction. This contrasting relationship with ENSO between California and the Pacific Northwest precipitation anomalies is consistent with previous literature on ENSO teleconnections (Evans *et al* 1998, Diaz *et al* 2001). These results suggest that the network has confidently learned a relationship between tropical and extratropical precipitation that is only present during specific low frequency time periods.

We find the neural networks are highly confident and accurate in its predictions during periods when the tropical-extratropical teleconnection is present from the tropical sources of predictability to the predictand midlatitude region. However, when there is a learned signal in the tropical source of predictability but that teleconnection is not present, the network will still be confident, but inaccurate in its predictions (figure 6). Based on these results, we can conclude that the teleconnection from the tropics to the extratropics is modulated on low frequency timescales, driving the periods of high and low subseasonal prediction skill.

3.4. Predictability within ERA5

The neural networks described thus far were trained, validated, and tested on the CESM2-LE for the time period between 1850 and 1949. The network's task was straightforward in that it was trained on 100 years of data and tested on data from the same climate model. Since climate models are known to have biases and deficiencies (Danabasoglu *et al* 2020), we explore whether the relationships learned by the network in the CESM2-LE historical simulation apply to the ERA5 reanalysis data. That is, we do not retrain the network on the ERA5 data, but rather, we use the CESM2-trained network to make predictions on the ERA5 anomalies. We make predictions on daily ERA5 data over the years 1959–2021, and the averaged accuracies across all 100 trained networks are shown in figure 7(a).

We analyze the ERA5 testing data in the same way as the CESM2 data, and find the accuracies are higher than random chance in each of the three locations. As expected, the accuracy over all of the ERA5 predictions is lower than the accuracies of the CESM2 data since the network was trained on the CESM2 data alone. Interestingly, California has notably the highest skill amongst the three predictand regions (figure 7). The
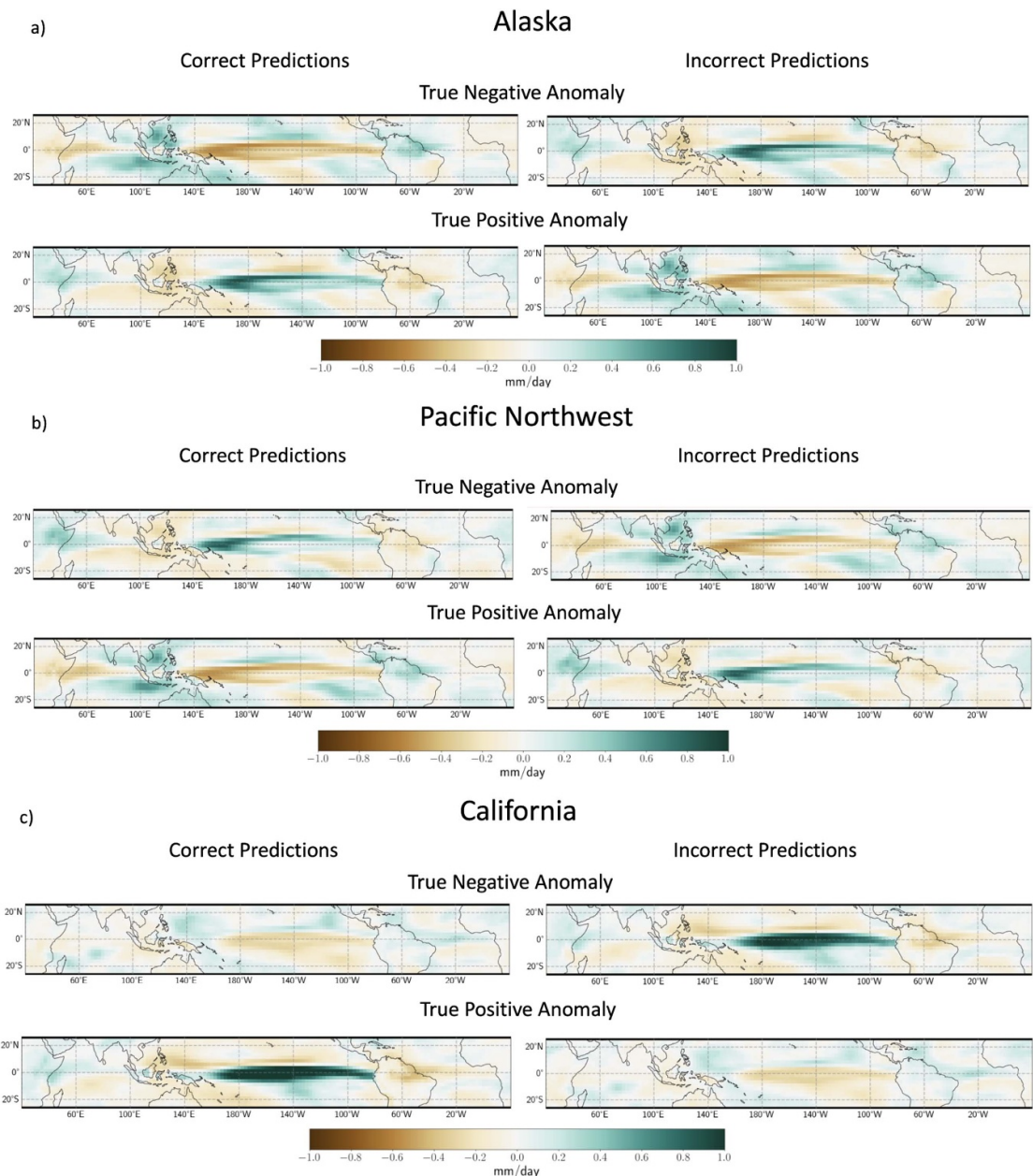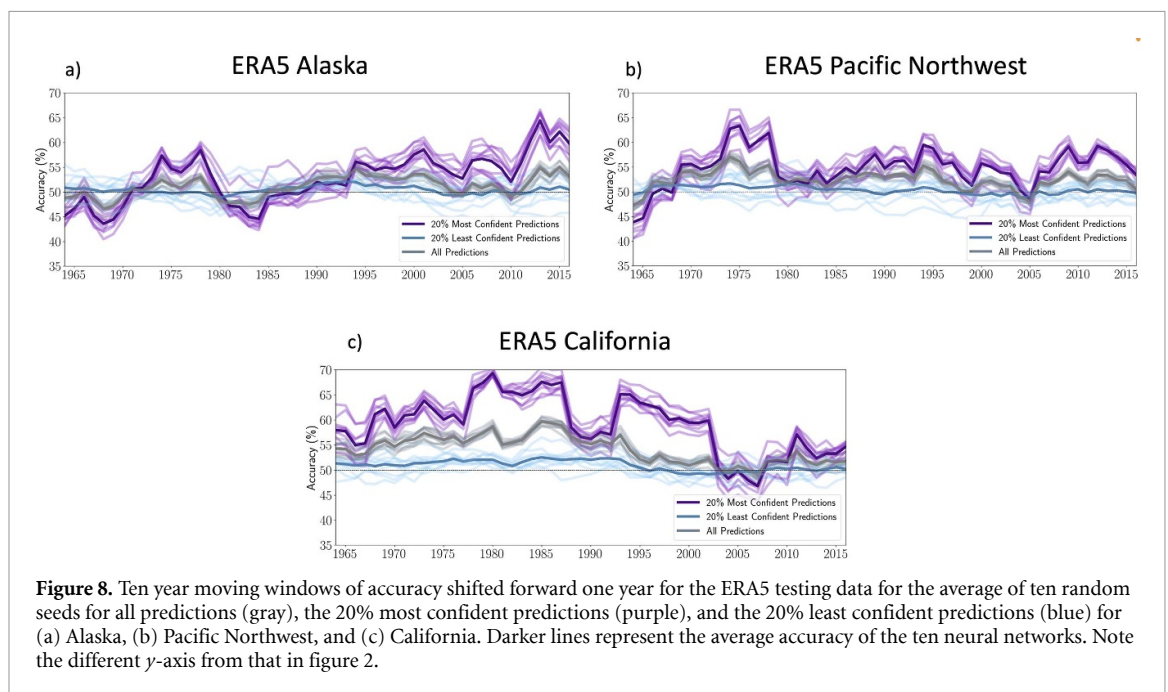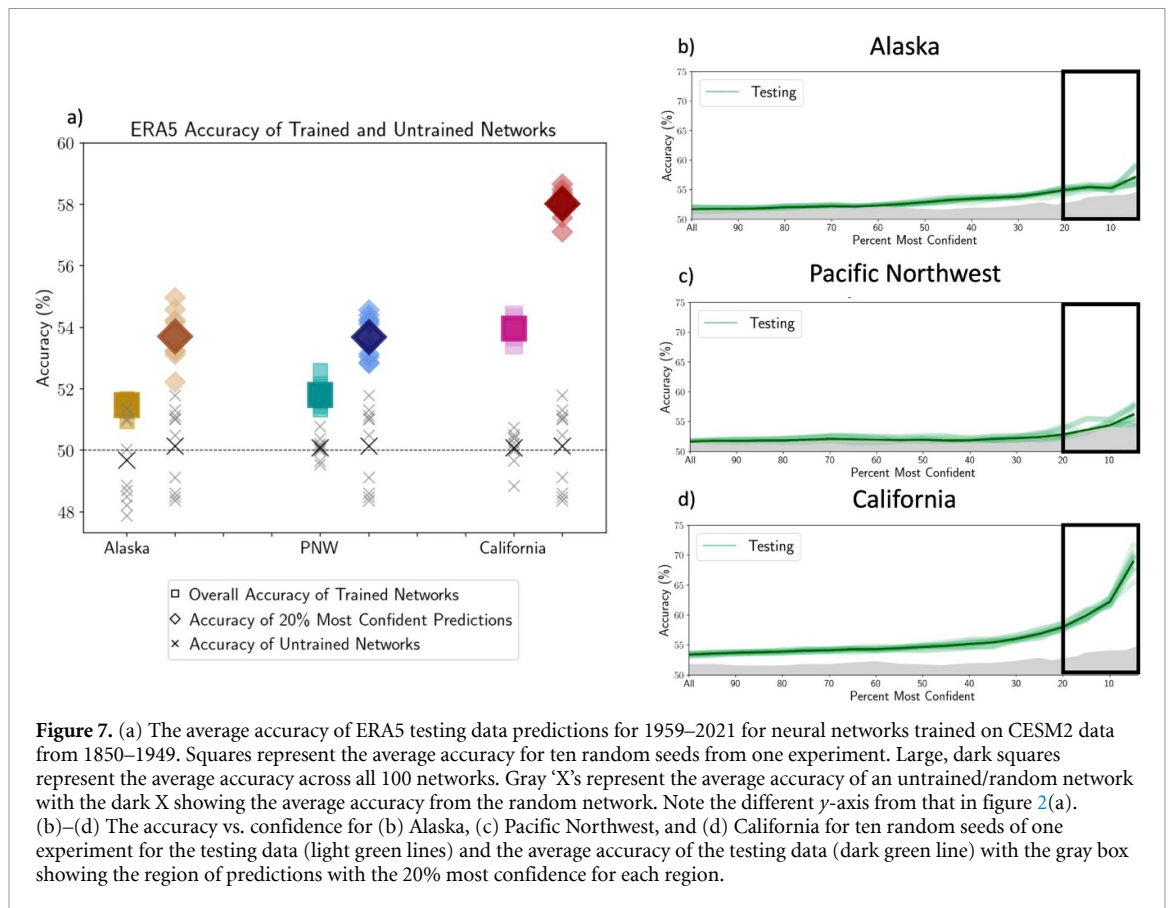
Figure 6. Composite maps of the input precipitation anomalies (mm day-1) for days that corresponded the 20% most confident and correct predictions (left) and incorrect predictions (right) for a true negative anomaly (top) and a true positive anomaly (bottom) for (a) Alaska, (b) Pacific Northwest, and (c) California using the CESM-2 data.

ERA5 data used for California shows nearly the same skill as the CESM2 data, particularly for the most confident forecasts (figures 2(a) and 7(a) red diamonds). Figure 7 is compelling in that the networks exhibit skill above random chance, particularly for the forecasts of opportunity (diamonds). That is, the networks have identified predictable states of the climate system within CESM2, that are useful for making accurate subseasonal precipitation predictions in the real world.

We extend this analysis to assess if the ERA5 data exhibits similar decadal variability of the prediction skill as was found in the CESM2-LE historical simulation. Figure 8 shows the average accuracy from ERA5 data for each of the ten random seeds per experiment, since the same ERA5 dataset is used for testing in each of the ten sets of neural networks. The dark lines show the average of the ten experiments. As with CESM2, there is also decadal variability, in the subseasonal prediction skill of ERA5 data, specifically for forecasts of opportunity, albeit with lower magnitude in the range of accuracy than in the CESM2 data. Also of note is California's high subseasonal prediction skill during the late 1970s and early 1980s, and again in the mid-to-late 1990s. The Week 4–5 precipitation anomaly prediction skill in California provided *only from the tropics* is, on average, over 65% for over a decade. The skill then drops to roughly random chance in the 2000s. In Alaska, the 1970s and early 2000s exhibit low frequency periods of high accuracy for the 20% most

Figure 7. (a) The average accuracy of ERA5 testing data predictions for 1959–2021 for neural networks trained on CESM2 data from 1850–1949. Squares represent the average accuracy for ten random seeds from one experiment. Large, dark squares represent the average accuracy across all 100 networks. Gray 'X's represent the average accuracy of an untrained/random network with the dark X showing the average accuracy from the random network. Note the different *y*-axis from that in figure 2(a). (b)–(d) The accuracy vs. confidence for (b) Alaska, (c) Pacific Northwest, and (d) California for ten random seeds of one experiment for the testing data (light green lines) and the average accuracy of the testing data (dark green line) with the gray box showing the region of predictions with the 20% most confidence for each region.

Figure 8. Ten year moving windows of accuracy shifted forward one year for the ERA5 testing data for the average of ten random seeds for all predictions (gray), the 20% most confident predictions (purple), and the 20% least confident predictions (blue) for (a) Alaska, (b) Pacific Northwest, and (c) California. Darker lines represent the average accuracy of the ten neural networks. Note the different *y*-axis from that in figure 2.

confident predictions compared to all predictions, with skill below random chance during the 1980s and 1990s. In the Pacific Northwest, the highest accuracy for the most confident predictions is also during the 1970s, with skill equal to below random chance in the 2000s. Consistent with the results found using the CESM2-LE historical simulations, these results show the largest decadal variability in subseasonal prediction skill occur during forecasts of opportunity.

4. Discussion

The decadal variability of subseasonal predictability found here is related to a low frequency modulation of the teleconnection from the tropics to the extratropics. An in-depth analysis has been conducted exploring various climate modes as drivers of the low frequency variability. The indices for extratropical variability modes (e.g. Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Pacific North America pattern (PNA), and Atlantic Multi-decadal Oscillation (AMO) are calculated using the Program for Climate Model Diagnosis and Intercomparison Metrics Package (Gleckler *et al* 2008, 2016). In particular, we applied the Common Basis Function approach that projects the observed empirical orthogonal functions onto model anomalies (Lee *et al* 2019a, 2021). For PDO and AMO, Met Office Hadley Centre Sea Ice and sea surface temperature dataset (HadISST) version 1.1 (Rayner *et al* 2003) monthly sea surface temperature was used as reference observation. For NAO and PNA, sea-level pressure from the 20th Century Reanalysis (20CR; Compo *et al* 2011) was used as reference observation. For both observations, the modes are analyzed over the period 1900–2005.

Particular effort went into investigating the role of the PDO as the PDO is known to modulate rainfall variability along the West Coast on low frequency timescales (Newman *et al* 2016). The correlations between the ten year moving window of accuracy of the 20% most confident predictions and the ten year smoothed PDO Index are shown in Supplemental figure S6 for Alaska, but results are similar for the Pacific Northwest and California. We find that the correlations between the PDO and the accuracy of the forecasts of opportunity for different ensemble members are inconsistent, with some correlation coefficients of the same magnitude but opposite sign between test members. Similar correlation calculations between the timeseries of the low frequency accuracy and the AMO, NAO, and PNA indices are computed and, as was found with the PDO, we find little-to-no consistent relationship between decadal variability of subseasonal predictability and the AMO, NAO or PNA. Related to our findings, Oliver (2015) showed that while the relationship between the MJO and Alaska surface air temperature varies on multidecadal timescales, the PDO is not fully responsible for the variability. Oliver (2015) did find a weak negative correlation between MJO-related midlatitude air temperature variability and the PDO in observations. This suggests the correlation found in Oliver (2015) may be related to the forced response and not internal variability analyzed using the CESM2 data, but further investigation is required. Moreover, van Oldenborgh (2005) found that decadal fluctuations in ENSO teleconnections, particularly for precipitation, can be explained only by random statistical fluctuations of a constant teleconnection, suggesting that the low frequency variability may be, in part, stochastically driven.

Relationships between the PDO and ENSO have also been shown to affect low-frequency variability of North American precipitation (Gershunov and Barnett 1998, Fuentes-Franco *et al* 2016, Nicola Kay Jennifer and Capotondi 2022). Additional work looking at the frequency of the PDO, ENSO, and PDO + ENSO in-phase periods with high or low accuracy was performed, but results were not significant to show that a PDO-ENSO phase relationship was fully responsible for variability in the subseasonal midlatitude prediction skill. Based on these additional climate mode analyses, it is likely that low frequency modulation of the tropical-extratropical teleconnection is not solely related to one particular climate mode of variability.

Decadal modulation of the MJO and ENSO has been documented (Burgman *et al* 2008, Gushchina and Dewitte 2019), and teleconnectivity between the tropics and the midlatitudes has been shown to vary based on both MJO and ENSO phase (Kim 2020, Tseng *et al* 2020). Furthermore, decadal variability of MJO teleconnections in boreal winter has been shown to be associated with changes in both the MJO tropical heating and the mean state via the jetstream as a waveguide (Kim 2020). Ambrizzi *et al* (1995) also found that regions of teleconnectivity tend to be co-located with the jets. Investigation of the role of the jet as the main determinant of the teleconnectivity between tropical and extratropical precipitation on decadal timescales is left for future work. Further investigation into the drivers responsible for the low frequency teleconnection modulation could help enhance high prediction skill in the midlatitudes provided by the tropics.

5. Conclusions

ANNs are employed to quantify subseasonal midlatitude prediction skill of U.S. West Coast precipitation provided by the tropics and assess variability of this skill on decadal timescales. We find that as the confidence of the networks increases, the accuracy increases, indicating that the networks can identify more predictable states of the climate system that result in forecasts with higher skill. Thus, we leverage the network's confidence in its predictions to identify subseasonal forecasts of opportunity, defined here as the 20% most confident predictions. An XAI method, integrated gradients, identifies regions in the tropical input maps which were most relevant to a neural network making a confident and correct prediction. These highlighted locations are found in regions associated with the MJO and ENSO, consistent with previous

literature noting these climate modes can produce subseasonal forecasts of opportunity when active (Johnson *et al* 2014, DeFlorio *et al* 2019, Arcodia *et al* 2020, Tseng *et al* 2020). These results suggest the network is identifying physically meaningful sources of midlatitude subseasonal predictability.

Using these networks, we find that subseasonal prediction skill provided by the tropics can vary on decadal timescales within the CESM2-LE historical simulations. The network-identified forecasts of opportunity exhibit the highest variability of accuracy, with periods of considerably higher and lower accuracy than the average of all predictions from the same time period. Thus, there are low frequency time periods in which the tropics provide large prediction skill to subseasonal midlatitude precipitation prediction. Moreover, there is a limit to the amount of prediction skill provided by the tropics on decadal timescales in this neural network framework as demonstrated by periods with forecast skill that can be worse than random chance.

The neural networks trained on CESM2 data from 1850 to 1949 were used to test ERA5 reanalysis data from 1959–2021 to examine if the relationships learned in CESM2 are found in reanalysis. The network is skillful on ERA5 data for all three regions, particularly during forecasts of opportunity, despite the neural network being trained on a different dataset and time period. This suggests CESM2 can identify real, physical relationships between tropical and extratropical precipitation useful for subseasonal predictability. Furthermore, the accuracy of the network for forecasts of opportunity also varies on decadal timescales in the ERA5 data, as was found in the CESM2 data. California exhibits the highest accuracy for the forecasts of opportunity in the ERA5 data, indicating that the relationships learned by the CESM-trained neural networks were most useful for California. While the last few decades have seen improvements in subseasonal prediction skill due to model advancements (Mariotti *et al* 2020), we find additional regional increases in subseasonal predictability provided by the tropics during forecasts of opportunity due to decadal variability.

Notably, there is a drop in skill since the late 2000s in California, which previously had two periods of high accuracy during the mid-1970s to mid-1980s and then the mid-1990s to early-2000s. These results have implications for subseasonal midlatitude prediction skill, as we have highlighted low frequency periods in which subseasonal prediction skill provided *only from the tropics* has high accuracy ($>65\%$ in ERA5) for the most confident forecasts. Conversely, there are low frequency time periods when the most confident predictions have prediction skill worse than random chance. This variability is driven by a decadal modulation in the teleconnectivity between the tropics and the extratropics—high accuracy time periods occur when the learned teleconnection from the tropics to the midlatitude region being predicted is present, and low accuracy periods occur when the teleconnection is not present. Investigating prediction skill of the teleconnectivity between the tropics and the extratropics on decadal timescales has the potential to enhance subseasonal forecast skill, particularly for forecasts of opportunity.

Data availability statement

CESM2 Large Ensemble Data are available freely to the public at www.cesm.ucar.edu/community-projects/lens2. ERA5 data are available freely at https://cds.climate.copernicus.eu/.

All Python code for processing data and figures for this analysis is available at www.github.com/mbarcodia/ERC23_paper_code and it has been converted to a permanent repository on Zenodo.

PMP source code is available on GitHub (https://github.com/PCMDI/pcmdi_metrics) with DOI number assigned by Zenodo at 10.5281/zenodo.7783324. The Twentieth Century Reanalysis (20CR) data is provided by the NOAA/Earth System Research Laboratory (ESRL)/Physical Sciences Division (PSD) from their website at www.esrl.noaa.gov/psd/. The HadISST data is available through the UK Met Office's website at www.metoffice.gov.uk/hadobs/hadisst/.

Acknowledgments

ORCID iDs

Marybeth C Arcodia ● https://orcid.org/0000-0002-7038-1932
Elizabeth A Barnes ● https://orcid.org/0000-0003-4284-9320
Jiwoo Lee ● https://orcid.org/0000-0002-0016-7199
Min-Seop Ahn ● https://orcid.org/0000-0002-3308-7793

References

Ahn M-S, Kim D, Ham Y-G and Park S 2020 Role of maritime continent land convection on the mean state and MJO propagation *J. Clim.* **33** 1659–75

Albers J R and Newman M 2019 A priori identification of skillful extratropical subseasonal forecasts *Geophys. Res. Lett.* **46** 12527–36

Albers J R and Newman M 2021 Subseasonal predictability of the North Atlantic oscillation *Environ. Res. Lett.* **16** 044024

Ambrizzi T, Hoskins B J and Hsu H-H 1995 Rossby wave propagation and teleconnection patterns in the Austral winter *J. Atmos. Sci.* **52** 3661–72

Antonios M, Barnes E A and Ebert-Uphoff I 2022 Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience *Artif. Intell. Earth Syst.* **1** e220012

Antonios M, Barnes E A and Ebert-Uphoff I 2023 Carefully choose the baseline: lessons learned from applying XAI attribution methods for regression tasks in geoscience *Artif. Intell. Earth Syst.* **2** e220058

Antonios M, Ebert-Uphoff I and Barnes E A 2021 Neural network attribution methods for problems in geoscience: a novel synthetic benchmark dataset *Environ. Data Sci.* **1** E8

Arcodia Marybeth and Kirtman Ben 2023 *Clim. Dyn.* Using simplified linear and nonlinear models to assess ENSO-modulated MJO teleconnections (https://doi.org/10.1007/s00382-023-06864-x)

Arcodia M C, Kirtman B P and Siqueira L S P 2020 How MJO teleconnections and ENSO interference impacts U.S. precipitation *J. Clim.* **33** 4621–40

Baehrens D, Schroeter T, Harmeling S, Kawanabe M, Hansen K and Mueller K-R 2010 How to explain individual classification decisions *J. Mach. Learn. Res.* **11** 1803–31

Bommer P, Kretschmer M, Hedström A, Bareeva D and Höhne M M-C 2023 Finding the right XAI method—a guide for the evaluation and ranking of explainable AI methods in climate science (arXiv:2303.00652 [cs.LG])

Bond N A and Vecchi G A 2003 The influence of the Madden–Julian oscillation on precipitation in Oregon and Washington *Weather Forecast.* **18** 600–13

Burgman R J, Schopf P S and Kirtman B P 2008 Decadal modulation of ENSO in a hybrid coupled model *J. Clim.* **21** 5482–500

Capotondi A, Deser C, Phillips A S, Okumura Y and Larson S M 2020 ENSO and Pacific decadal variability in the community earth system model version 2 *J. Adv. Model. Earth Syst.* **12** e2019MS002022

Cassou C 2008 Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic Oscillation *Nature* **455** 523–7

Cayan D R, Dettinger M D, Diaz H F and Graham N E 1998 Decadal variability of precipitation over Western North America *J. Clim.* **11** 3148–66

Compo G P *et al* 2011 The twentieth century reanalysis project *Q. J. R. Meteorol. Soc.* **137** 1–28

Daehyun K, Kim D, Ahn M-S, Neale R, Lee J and Gleckler P J 2020 The role of the mean state on MJO simulation in CESM2 ensemble simulation *Geophys. Res. Lett.* **47** e2020GL089824

Dai A 2013 The influence of the Inter-Decadal Pacific oscillation on US precipitation during 1923–2010 *Clim. Dyn.* **41** 633–46

Dai A and Wigley T M L 2000 Global patterns of ENSO-induced precipitation *Geophys. Res. Lett.* **27** 1283–6

Danabasoglu G *et al* 2020 The community earth system model version 2 (CESM2) *J. Adv. Model. Earth Syst.* **12** e2019MS001916

DeFlorio M J *et al* 2019 Experimental subseasonal to seasonal (S2S) forecasting of atmospheric rivers over the Western United States *J. Geophys. Res.* **124** 11242–65

Diaz H F, Hoerling M P and Eischeid J K 2001 ENSO variability, teleconnections and climate change *Int. J. Climatol.* **21** 1845–62

Evans M N, Fairbanks R G and Rubenstone J L 1998 A proxy index of ENSO teleconnections *Nature* **394** 732–3

Fuentes-Franco R, Giorgi F, Coppola E and Kucharski F 2016 The role of ENSO and PDO in variability of winter precipitation over North America from twenty first century CMIP5 projections *Clim. Dyn.* **46** 3259–77

Gershunov A and Barnett T P 1998 Interdecadal modulation of ENSO teleconnections *Bull. Am. Meteorol. Soc.* **79** 2715–26

Gleckler P J, Taylor K E and Doutriaux C 2008 Performance metrics for climate models *J. Geophys. Res. Atmos.* **113**

Gleckler P, Doutriaux C, Durack P, Taylor K, Zhang Y, Williams D, Mason E and Servonnat J 2016 A more powerful reality test for climate models *Eos* **97**

Gordon E M, Barnes E A and Hurrell J W 2021 Oceanic harbingers of pacific decadal oscillation predictability in CESM2 detected by neural networks *Geophys. Res. Lett.* **48** e2021GL095392

Gushchina D and Dewitte B 2019 Decadal modulation of the relationship between intraseasonal tropical variability and ENSO *Clim. Dyn.* **52** 2091–103

Ham Y-G, Kim J-H and Luo J-J 2019 Deep learning for multi-year ENSO forecasts *Nature* **573** 568–72

Hersbach H *et al* 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049

Hoskins B J and Karoly D J 1981 The steady linear response of a spherical atmosphere to thermal and orographic forcing *J. Atmos. Sci.* **38** 1179–96

Huang H, Patricola C M, Bercos-Hickey E, Zhou Y, Rhoades A, Risser M D and Collins W D 2021 Sources of subseasonal to seasonal predictability of atmospheric rivers and precipitation in the Western United States *J. Geophys. Res.* **126** e2020JD034053

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

Jones C and Carvalho L M V 2012 Spatial–intensity variations in extreme precipitation in the contiguous United States and the Madden–Julian oscillation *J. Clim.* **25** 4898–913

Kang D, Kim D, Ahn M-S and An S-I 2021 The role of the background meridional moisture gradient on the propagation of the MJO over the maritime continent *J. Clim.* **34** 6565–81

Kim H 2020 *Mechanisms for the Decadal Change of the MJO Teleconnection over the Northern Hemisphere in Boreal Winter* (Graduate School of UNIST) (available at: https://scholarworks.unist.ac.kr/handle/201301/31779)

Labe Z M and Barnes E A 2022 Predicting slowdowns in decadal climate warming trends with explainable neural networks *Geophys. Res. Lett.* **49** e2022GL098173

Lau W K M, Waliser D E and Roundy P E 2012 Tropical–extratropical interactions *Intraseasonal Variability in the Atmosphere-Ocean Climate System* ed W K-M Lau and D E Waliser (Springer) pp 497–512

Lee J, Sperber K R, Gleckler P J, Bonfils C J W and Taylor K E 2019a Quantifying the agreement between observed and simulated extratropical modes of interannual variability *Clim. Dyn.* **52** 4057–89

Lee J, Sperber K R, Gleckler P J, Taylor K E and Bonfils C J W 2021 Benchmarking performance changes in the simulation of extratropical modes of variability across CMIP generations *J. Clim.* **34** 6945–69

Lee R W, Woolnough S J, Charlton-Perez A J and Vitart F 2019b ENSO modulation of MJO teleconnections to the north Atlantic and Europe *Geophys. Res. Lett.* **46** 13535–45

Ling J, Li C, Li T, Jia X, Khouider B, Maloney E, Vitart F, Xiao Z and Zhang C 2017 Challenges and opportunities in MJO studies *Bull. Am. Meteorol. Soc.* **98** ES53–56

Lovejoy S 2018 Spectra, intermittency, and extremes of weather, macroweather and climate *Sci. Rep.* **8** 12697

Luo B, Luo D, Dai A, Simmonds I and Wu L 2022 Decadal variability of winter warm Arctic-cold Eurasia dipole patterns modulated by Pacific decadal oscillation and Atlantic multidecadal oscillation *Earth's Future* **10** e2021EF002351

Madden R A and Julian P R 1971 Detection of a 40–50 day oscillation in the zonal wind in the Tropical Pacific *J. Atmos. Sci.* **28** 702–8

Mariotti A *et al* 2020 Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond *Bull. Am. Meteorol. Soc.* **101** E608–25

Mariotti A, Barnes E A, Chang E K-M, Lang A, Dirmeyer P A, Pegion K, Barrie D and Baggett C 2019 Bridging the weather-to-climate prediction gap *Eos* **100**

Martin Z K, Barnes E A and Maloney E 2022 Using simple, explainable neural networks to predict the Madden–Julian oscillation *J. Adv. Model. Earth Syst.* **14** e2021MS002774

Matthews A J, Hoskins B J and Masutani M 2004 The global response to tropical heating in the Madden–Julian oscillation during the northern winter *Q. J. R. Meteorol. Soc.* **130** 1991–2011

Mayer K J and Barnes E A 2021 Subseasonal forecasts of opportunity identified by an explainable neural network *Geophys. Res. Lett.* **48** e2020GL092092

Mayer K J and Barnes E A 2022 Quantifying the effect of climate change on midlatitude subseasonal prediction skill provided by the tropics *Geophys. Res. Lett.* **49** e2022GL098663

McCabe G J and Dettinger M D 1999 Decadal variations in the strength of ENSO teleconnections with precipitation in the Western United States *Int. J. Climatol.* **19** 1399–410

McGovern A, Lagerquist R, John Gagne D, Jergensen G E, Elmore K L, Homeyer C R and Smith T 2019 Making the black box more transparent: understanding the physical implications of machine learning *Bull. Am. Meteorol. Soc.* **100** 2175–99

Meehl G A, Richter J H, Teng H, Capotondi A, Cobb K, Doblas-Reyes F, Donat M G, England M H, Fyfe J C, Han W and Kim H 2021 Initialized earth system prediction from subseasonal to decadal timescales *Nat. Rev. Earth Environ.* **2** 340–57

Moon J-Y, Wang B and Ha K-J 2011 ENSO regulation of MJO teleconnection *Clim. Dyn.* **37** 1133–49

National Academies of Sciences, Engineering, and Medicine, Division on Earth and Life Studies, Ocean Studies Board, Board on Atmospheric Sciences and Climate, and Committee on Developing a U.S. Research Agenda to Advance Subseasonal to Seasonal Forecasting 2016 *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts* (National Academies Press)

Newman M *et al* 2016 The pacific decadal oscillation, revisited *J. Clim.* **29** 4399–427

Newman M, Sardeshmukh P D, Winkler C R and Whitaker J S 2003 A study of subseasonal predictability *Mon. Weather Rev.* **131** 1715–32

Nicola Kay Jennifer M and Capotondi A 2022 Modulation of ENSO teleconnections over North America by the Pacific decadal oscillation *Environ. Res. Lett.* **17** 114005

O'Malley T, Bursztein E, Long J, Chollet F, Jin H and Invernizzi L 2019 *KerasTuner* (available at: https://github.com/keras-team/keras-tuner)

O'Reilly C H, Heatley J, MacLeod D, Weisheimer A, Palmer T N, Schaller N and Woollings T 2017 Variability in seasonal forecast skill of Northern hemisphere winters over the twentieth century *Geophys. Res. Lett.* **44** 5729–38

Oliver E C J 2015 Multidecadal variations in the modulation of Alaska wintertime air temperature by the Madden–Julian oscillation *Theor. Appl. Climatol.* **121** 1–11

Pegion K *et al* 2019 The subseasonal experiment (SubX): a multimodel subseasonal prediction experiment *Bull. Am. Meteorol. Soc.* **100** 2043–60

Pegion Kathy, Becker E J and Kirtman B P 2022 Understanding predictability of daily southeast U.S. precipitation using explainable machine learning *Artif. Intell. Earth Syst.* **1** e220011

Pegion K and Kirtman B P 2008 The impact of air–sea interactions on the predictability of the tropical intraseasonal oscillation *J. Clim.* **21** 5870–86

Rayner N A, Parker D E, Horton E B, Folland C K, Alexander L V, Rowell D P, Kent E C and Kaplan A 2003 Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century *J. Geophys. Res. Atmos.* **108**

Rosa J P S, Guerra D J D, Horta N C G, Martins R M F and Lourenço N C C 2020 Overview of artificial neural networks *Using Artificial Neural Networks for Analog Integrated Circuit Design Automation* ed J P S Rosa, D J D Guerra, N C G Horta, R M F Martins and N C C Lourenço (Springer International Publishing) pp 21–44

Scaife A A *et al* 2014 Skillful long-range prediction of European and North American winters *Geophys. Res. Lett.* **41** 2514–9

Sonnewald M, Lguensat R, Jones D C, Dueben P D, Brajard J and Balaji V 2021 Bridging observations, theory and numerical simulation of the ocean using machine learning *Environ. Res. Lett.* **16** 073008

Soulard N, Lin H, Derome J and Yu B 2021 Tropical forcing of the circumglobal teleconnection pattern in boreal winter *Clim. Dyn.* **57** 865–77

Stan C, Straus D M, Frederiksen J S, Lin H, Maloney E D and Schumacher C 2017 Review of tropical-extratropical teleconnections on intraseasonal time scales *Rev. Geophys.* **55** 902–37

Straaten C V, Whan K, Coumou D, van den Hurk B and Schmeits M 2023 Correcting sub-seasonal forecast errors with an explainable ANN to understand misrepresented sources of predictability of european summer temperatures *Artif. Intell. Earth Syst.* **1** 1–49

Sundararajan M, Taly A and Yan Q 2017 Axiomatic attribution for deep networks *Proc. Machine Learning Research* vol 70, ed D Precup and Y W Teh pp 3319–28

Toms B A, Barnes E A and Ebert-Uphoff I 2020 Physically interpretable neural networks for the geosciences: applications to earth system variability *J. Adv. Model. Earth Syst.* **12** e2019MS002002

Tseng K-C, Maloney E and Barnes E A 2020 The consistency of MJO teleconnection patterns on interannual time scales *J. Clim.* **33** 3471–86

van Oldenborgh G J 2005 Searching for decadal variations in ENSO precipitation teleconnections *Geophys. Res. Lett.* **32**

Vitart F 2017 Madden–Julian oscillation prediction and teleconnections in the S2S database *Q. J. R. Meteorol. Soc.* **143** 2210–20

Vitart F *et al* 2017 The subseasonal to seasonal (S2S) prediction project database *Bull. Am. Meteorol. Soc.* **98** 163–73

Weickmann K M 1983 Intraseasonal circulation and outgoing longwave radiation modes during Northern hemisphere winter *Mon. Weather Rev.* **111** 1838–58

Weisheimer A, Schaller N, O'Reilly C, MacLeod D A and Palmer T 2017 Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic oscillation (NAO) and their potential value for extreme event attribution *Q. J. R. Meteorol. Soc.* **143** 917–26

Zhang C 2005 Madden–Julian oscillation *Rev. Geophys.* **43**

Zhang C 2013 Madden–Julian oscillation: bridging weather and climate *Bull. Am. Meteorol. Soc.* **94** 1849–70

Zheng C, Kar-Man Chang E, Kim H, Zhang M and Wang W 2018 Impacts of the Madden–Julian oscillation on storm-track activity, surface air temperature, and precipitation over North America *J. Climate* **31** 6113—-34