ADVANCING EARTH AND SPACE SCIENCES

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2024JD042917

A Regimes-Based Approach to Identifying Seasonal State-Dependent Prediction Skill

Kyle Shackelford[1], **Charlotte A. DeMott**[1], **Peter Jan van Leeuwen**[1], **and Elizabeth A. Barnes**[1]

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

Key Points:

- A methodology for detecting state-dependent prediction skill is applied to seasonal prediction of North Atlantic atmospheric circulation
- Neural networks identify four SST regimes that improve seasonal prediction relative to baseline forecasting methods
- The comprehensive approach to seasonal prediction identifies prediction skill beyond ENSO and outperforms a linear ENSO-based model

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Shackelford,
Kyle.Shackelford@colostate.edu

Citation:

Shackelford, K., DeMott, C. A., van Leeuwen, P. J., & Barnes, E. A. (2025). A regimes-based approach to identifying seasonal state-dependent prediction skill. *Journal of Geophysical Research: Atmospheres*, *130*, e2024JD042917. https://doi.org/10.1029/2024JD042917

Received 7 NOV 2024
Accepted 5 APR 2025

Author Contributions:

Conceptualization: Kyle Shackelford, Charlotte A. DeMott, Peter Jan van Leeuwen, Elizabeth A. Barnes
Formal analysis: Kyle Shackelford
Investigation: Kyle Shackelford
Methodology: Kyle Shackelford, Charlotte A. DeMott, Peter Jan van Leeuwen, Elizabeth A. Barnes
Supervision: Charlotte A. DeMott, Peter Jan van Leeuwen
Visualization: Kyle Shackelford
Writing – original draft: Kyle Shackelford

Abstract

Subseasonal-to-decadal atmospheric prediction skill attained from initial conditions is typically limited by the chaotic nature of the atmosphere. However, for some atmospheric phenomena, prediction skill on subseasonal-to-decadal timescales is increased when the initial conditions are in a particular state. In this study, we employ machine learning to identify sea surface temperature (SST) regimes that enhance prediction skill of North Atlantic atmospheric circulation. An ensemble of artificial neural networks is trained to predict anomalous, low-pass filtered 500 mb height at 7–8 weeks lead using SST. We then use self-organizing maps (SOMs) constructed from 9 regions within the SST domain to detect state-dependent prediction skill. SOMs are built using the entire SST time series, and we assess which SOM units feature confident neural network predictions. Four regimes are identified that provide skillful seasonal predictions of 500 mb height. Our findings demonstrate the importance of extratropical decadal SST variability in modulating downstream ENSO teleconnections to the North Atlantic. The methodology presented could aid future forecasting on subseasonal-to-decadal timescales.

Plain Language Summary

Accurate weather and climate forecasting on timescales longer than 2 weeks is an ongoing challenge in the atmospheric science community. However, improved forecasting on these timescales is important for advancing our physical understanding of the climate system and informing public-sector decisions like disaster preparedness. To address this challenge, some researchers focus on specific atmospheric and oceanic conditions (e.g., El Niño) that may extend the duration of skillful forecasts. Here, we identify patterns of sea surface temperature that improve the forecast accuracy of atmospheric conditions over the North Atlantic Ocean on seasonal timescales (7–8 weeks). Atmospheric conditions over the North Atlantic Ocean have a profound impact upon the weather in western Europe and eastern North America, and improved seasonal forecasts in this region could mitigate the societal effects of extreme events. Our method improves forecast accuracy relative to traditional forecasting methods, and provides an adaptable framework that could be applied to other atmospheric forecasting problems.

1. Introduction

Accurate, probabilistic weather prediction on seasonal timescales is critical for informing society-relevant decisions such as water allocation and disaster preparation, but occurs on timescales where initial conditions provide limited predictive skill. Instead, prediction skill on these time scales comes from lower frequency modes of weather and climate variability, such as the state of the stratosphere (Baldwin & Dunkerton, 2001), El Niño Southern Oscillation (ENSO; e.g., Alexander et al., 2002), high latitude snow cover (Gastineau et al., 2017), and soil moisture (Esit et al., 2021). On seasonal timescales, ENSO modification of tropical convection influences global weather and climate through Rossby wave trains that extend into the midlatitudes (e.g., Hoskins & Karoly, 1981). The downstream influence of ENSO teleconnections on the extratropics enhances atmospheric seasonal predictability in difficult to forecast regions, such as the North Atlantic (Bjerknes, 1969).

The North Atlantic Oscillation (NAO) is a dominant mode of climate variability in the northern hemisphere midlatitudes, with variability on timescales of weeks to decades (e.g., Hurrell et al., 2003; Wallace & Gutzler, 1981). Defined by pressure differences between the subpolar Icelandic Low and subtropical Azores High, the NAO exerts a sizable influence on European and North American weather (Scaife et al., 2014). ENSO is frequently cited as a source of NAO prediction skill on seasonal time scales (e.g., Brönnimann et al., 2007), but much debate remains about the exact nature of this relationship (e.g., Mezzina et al., 2020).

Dynamically, El Niño events trigger an atmospheric bridge between the tropical Pacific and North Pacific, deepening the Aleutian Low and promoting wavenumber 1 growth and vertical wave propagation into the

Writing – review & editing:
Kyle Shackelford, Charlotte A. DeMott,
Peter Jan van Leeuwen, Elizabeth
A. Barnes

stratosphere (Alexander et al., 2002). Vertically propagating waves disrupt the polar stratospheric vortex, which imprints on the troposphere as enhanced blocking over the Greenland region 2−3 weeks later (Alexander et al., 2002), projecting negatively onto the NAO. Additionally, anomalous tropical convection during ENSO events excites Rossby waves that propagate into the extratropics, resulting in a north/south dipole-like pressure structure over eastern North America that resembles the negative NAO phase (Brönnimann et al., 2007). Conversely, La Niña events are frequently associated with a positive phase of the NAO, resulting in an oft cited inverse linear relationship between ENSO phase and NAO phase (e.g., Scaife et al., 2014). However, the exact nature of this relationship remains uncertain, with studies citing the longitudinal position of ENSO sea surface temperature (SST) anomalies (Zhang, Wang, et al., 2019) and intrinsic low frequency NAO variability (Weisheimer et al., 2019) as confounding factors in seasonal prediction. Furthermore, recent studies identify a distinction between the projection of ENSO−driven teleconnections onto the NAO index, and the dynamics of the NAO itself (Mezzina et al., 2020).

Additional modes of SST variability, such as the Indian Ocean Dipole (Hardiman et al., 2020) and Pacific Decadal Oscillation (Brönnimann et al., 2007; Hall et al., 2019), are also identified as potential sources of NAO predictability on seasonal timescales. Recent findings demonstrate that the PDO modifies the strength and location of the Aleutian Low during ENSO events, but primarily focus on subsequent impacts on temperature and precipitation over North America (Dong & Dai, 2015; Dong et al., 2018; Maher et al., 2022). Previous studies investigating NAO seasonal predictability generally consider different modes of SST variability in isolation, rather than collectively, with an emphasis on the linear ENSO−NAO relationship.

There has been a long tradition of connecting predictability on subseasonal-to-decadal timescales to specific initial states (e.g., Palmer & Hagedorn, 2006). Recently, there has been a renewed focus in this area related to the maturing of machine-learning methods (Gordon & Barnes, 2022; Mariotti et al., 2020; Mayer & Barnes, 2021). So-called “forecasts of opportunity,” or state-dependent predictability, describe a configuration of the initial state that improves prediction skill of a given phenomena beyond the background prediction skill, thus increasing forecast confidence. Machine learning methods have proven particularly useful for diagnosing state-dependent predictability. For example, Mayer and Barnes (2021) use classification-based neural networks to identify patterns of tropical convection that enhance North Atlantic atmospheric predictability on subseasonal timescales, and Gordon and Barnes (2022) use regression-based neural networks to identify low frequency modes of ocean variability that enhance SST predictability on decadal timescales. Additionally, Roberts et al. (2020) use unsupervised learning to cluster large-scale atmospheric circulations into weather regimes to identify subseasonal forecasts of opportunity.

This study offers a holistic, regimes-based approach to state-dependent seasonal prediction of North Atlantic atmospheric circulation using neural networks and self-organizing maps (SOMs). We train an ensemble of neural networks to predict a conditional Gaussian distribution of 2-week smoothed North Atlantic 500 mb heights at a 7–8 weeks lead using SSTs over 10°S to 60°N (Sections 2.2 and 3.1). The standard deviation of the predicted distribution is used to assess the relative confidence of each prediction, with a smaller predicted standard deviation deemed more confident than a larger predicted standard deviation. SST modes linked to enhanced seasonal predictability of North Atlantic atmospheric circulations form the basis of the SOMs, which are trained over the entire SST time series (Section 2.3). To identify SST regimes that enhance seasonal prediction skill, we assess which SOM units most frequently feature confident neural network predictions and pinpoint four distinct regimes of enhanced prediction skill (Sections 3.2 and 3.3). Two regimes capture ENSO-driven SST variability and conform to the linear ENSO−NAO relationship documented in previous studies, instilling confidence in the methodology. The two remaining regimes illustrate cases that oppose the linear ENSO−NAO model, suggesting sources of prediction skill in SST beyond ENSO.

2. Data and Methods

2.1. Data

We use 82 years (1940–2021) of hourly SST over 10°S to 60°N and 500 mb heights (Z500) over the northwest Atlantic Ocean (50°W–70°W; 52°N–70°N) from the ERA-5 reanalysis data set (Hersbach et al., 2020). Our selected Z500 region (50°W–70°W; 52°N–70°N) targets anomalous Greenland blocking corresponding to negative NAO events in which height anomalies occur closer to the North American continent (Davini et al., 2012). Targeting this region provides a more specific forecast than targeting a broader region spanning Greenland, as the associated surface impacts will be largest over North America.

SST is interpolated temporally from hourly to daily resolution and spatially from a $0.25° \times 0.25°$ grid to a $5° \times 5°$ grid. At each grid cell, SST is de-seasonalized by subtracting the mean annual cycle and then linearly detrended to remove the climate change signal. Z500 is interpolated temporally from hourly to daily resolution and an area-weighted mean is applied over the North Atlantic target area to obtain a single Z500 value for each day. The Z500 annual cycle is noisier than SST, thus Z500 is de-seasonalized by subtracting the first four harmonics of the annual cycle and then linearly detrended to remove the climate change signal. We then apply a 15-day forward running mean at day 53 to the Z500 time series to filter out daily to weekly timescales and instead highlight lower frequency variability. Hereafter, the 15-day forward running mean Z500 between days $+53-+67$ will be referred to as week $7-8$ Z500. During training/validation/testing split, we compute the mean and standard deviation of Z500 and SST at each grid cell for each unique training set, and use these values to standardize the training, validation, and testing data to have zero mean and unit variance. In Section 4, we use ERA5 500 mb heights for composite analysis that are regridded to a $2° \times 2°$ grid, but otherwise preprocessed in the same manner as Z500.

2.2. Neural Network Setup

We follow the methodology outlined in Barnes et al., 2021, wherein regression-based artificial neural networks (ANNs) are used to predict the parameters of a conditional distribution of a target variable. In this study, we employ ANNs that receive SST over 10°S–60°N as input and output the mean (μ) and standard deviation (σ) of a conditional Gaussian distribution of week $7-8$ Z500 (e.g., Barnes et al., 2021; Nix & Weigend, 1994; Wang et al., 2016). All grid cells over land are removed from the input, leaving 625 SST values in the ANN input layer. The input layer is densely connected to the first hidden layer containing 32 nodes, which is densely connected to a second hidden layer containing 16 nodes. The second hidden layer is connected to an output layer of two nodes predicting the μ and σ of a conditional distribution of the 15-day running mean Z500 averaged over 50°W–70°W and 52°N–70°N (Figure 1a). The hidden layers contain ReLu activation functions, while the output layers contain linear activation functions. Dropout is implemented between the input layer and hidden layer, as well as between the first hidden layer and second hidden layer, with a dropout rate of 0.8. We tested various dropout rates and found neural networks to be prone to overfitting in the absence of high dropout. ANNs train for 60 epochs, with a learning rate of $1e-4$.

Evaluating state-dependent prediction skill across an individual ANN can introduce biases through how the data are split between training and validation, whereas evaluating ensemble members with varied training and validation data ensures that detection of state-dependent prediction is robust to sampling bias. Thus, an ensemble of 15 ANNs is trained featuring identical hyperparameters, but training/validation split and initial weights differ amongst ensemble members. To generate testing data, we implement a "leave one out" approach, where a sliding 5-year window is applied across the time series to provide testing data, with individual years outside a given 5-year period assigned to the training/validation data sets at a 70%/30% split. In this way, predictions are generated for each day between 1940 and 2021. All analysis in this paper is performed on the testing data using the ANN ensemble mean. More information about our ANN setup can be found in Text S1 in Supporting Information S1.

2.3. Identifying State-Dependent Forecast Skill Using Self-Organizing Maps

Self-organizing maps (SOMs) are a form of unsupervised machine learning for clustering complex data sets (Kohonen, 1982). Previous studies have implemented SOMs to detect modes of decadal climate variability and predictability (Gu & Gervais, 2021), to diagnose atmospheric circulations associated with extreme weather (Cassano et al., 2006), and to assess changes to arctic air masses under climate change (Gervais et al., 2016). The SOM algorithm is based on a set of cluster centers, called units, typically projected on a 2-dimensional map. Each unit is associated with a weight vector, which is the position vector of the unit in the input space. The position vectors are initialized randomly, or using the leading principle components on the underlying data. While units in the map space stay fixed, training consists in moving their weight vectors toward the input data. Specifically, for each input sample, we determine the best matching unit (BMU), as defined by the minimum Euclidean distance between the sample and the weight vector. The weight vector of the BMU and neighborhood units are then updated to more closely match the input samples at a rate determined by a learning rate parameter. In this way, units adjust to the data points and form the data clusters. User-selected hyperparameters that determine the 2-D

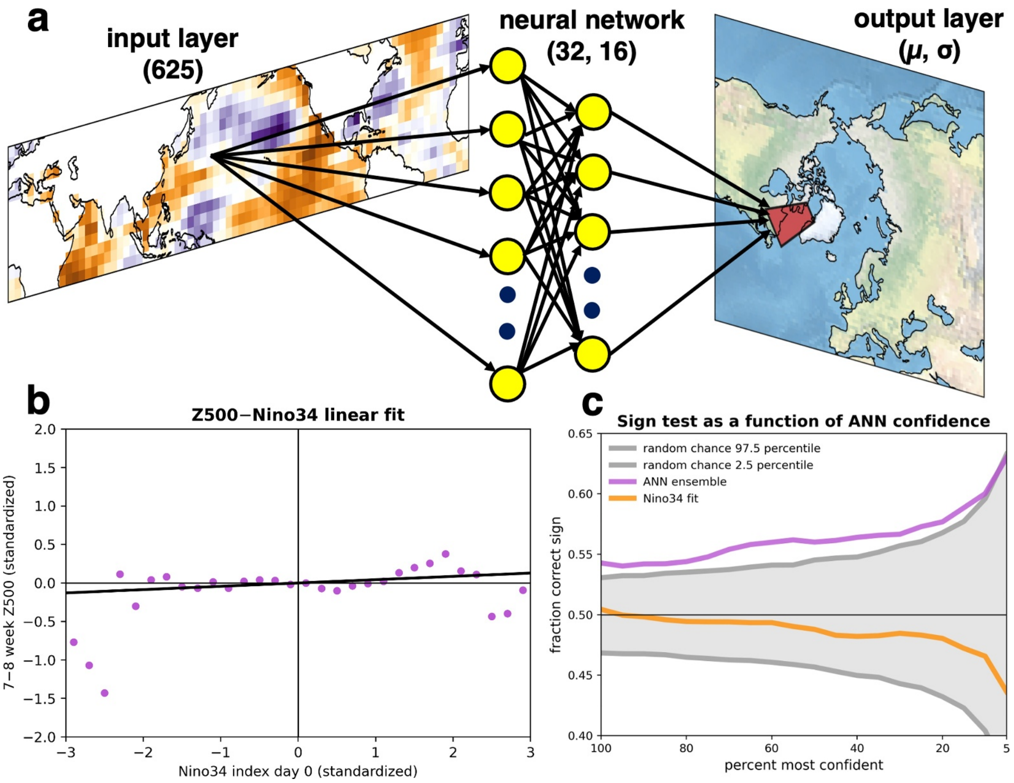
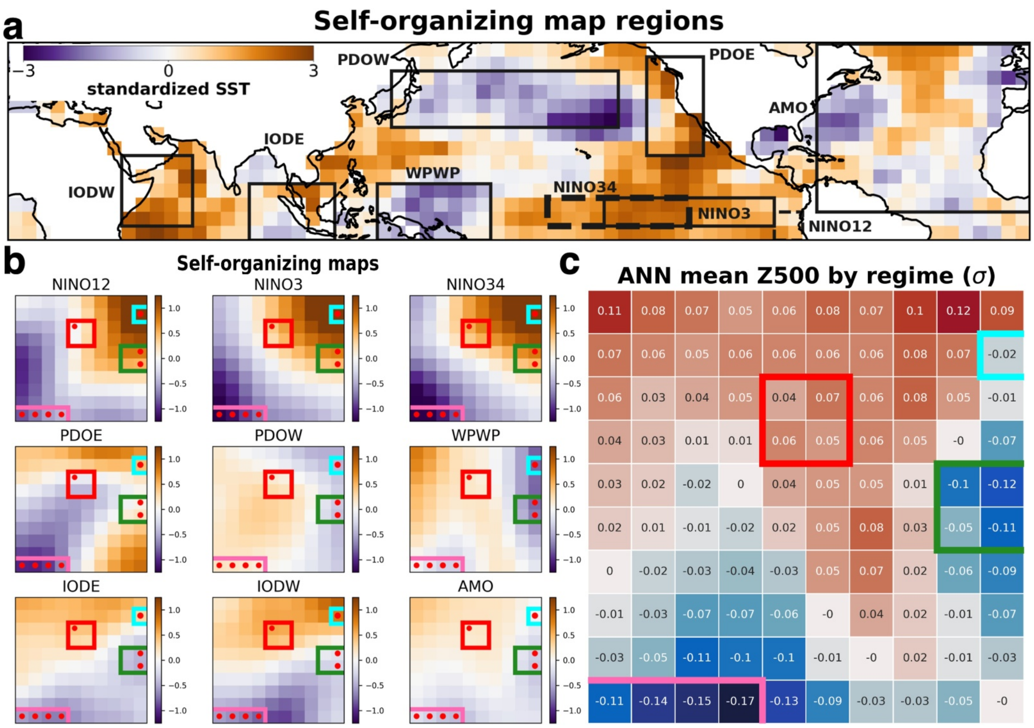


Figure 1. (a) Illustration of ANN architecture. (b) Ordinary least squares linear fit between standardized Nino34 index at day 0 and standardized Z500 at week 7−8 (black), and mean Z500 (purple) binned by 0.2 σ intervals in the Nino34 index. The Nino34–Z500 correlation coefficient is 0.043. (c) Fraction correct sign anomaly prediction (with respect to Z500 median) as a function of ANN confidence threshold for: ANN ensemble mean (purple), Nino34 linear fit (orange), and the 2.5 and 97.5 random chance percentiles (gray) generated through Monte Carlo simulation.

plane dimensions and how nodes move toward samples are specific to given applications and are defined in Supporting Information S1. The 2-D plane allows for effective visualization, as shown in Figure 2b.

We use SOMs to detect SST regimes that enhance seasonal prediction skill of North Atlantic atmospheric circulation. First, we identify modes of SST variability known to influence North Atlantic circulation: El Niño Southern Oscillation (ENSO; e.g., Alexander et al., 2002), the Indian Ocean Dipole (IOD; e.g., Hardiman et al., 2020), the Pacific Decadal Oscillation (PDO; e.g., Bronnimann et al., 2007), the West Pacific Warm Pool (WPWP), and the Atlantic Multidecadal Oscillation (AMO; e.g., Sutton & Hodson, 2003). Second, we take area-weighted means of regional SST to approximate the variability associated with these modes, leaving nine mean variables: Nino12, Nino3, Nino34, the eastern and western components of the PDO (PDOE and PDOW), the eastern and western component of the IOD (IODE and IODW), WPWP, and AMO (Figure 2a). Next, we apply self-organizing maps to time series of the nine SST regions over the entire record (1940−2021) to visualize the covariability of the individual modes, such that every daily SST timestep is assigned to the SOM unit that best matches anomalous SST in each SOM region on that day. More information on SOMs, their training process, and relevant hyperparameters can be found in Supporting Information S1 (Text S3).

To detect SST regimes that enhance Z500 prediction skill, we assess the SST input associated with the most confident ANN predictions, based on the standard deviation of the Z500 prediction as determined by the ANN. We define the 20% most confident predictions as the predictions where ensemble mean standard deviation output falls below the 20th percentile standard deviation output, imposing no requirement that these predictions be "correct." After isolating the most confident predictions, we determine the SOM BMU for the SST input associated with each confident ANN prediction, and obtain a count of confident ANN predictions originating from each unit in the SOM grid. We then normalize each SOM unit count of confident ANN predictions by the total count of ANN predictions originating from that unit to compute the frequency that predictions with SST input



0.11	0.08	0.07	0.05	0.06	0.08	0.07	0.1	0.12	0.09
0.07	0.06	0.05	0.06	0.06	0.06	0.06	0.08	0.07	-0.02
0.06	0.03	0.04	0.05	0.04	0.07	0.06	0.08	0.05	-0.01
0.04	0.03	0.01	0.01	0.06	0.05	0.06	0.05	-0	-0.07
0.03	0.02	-0.02	0	0.04	0.05	0.05	0.01	-0.1	-0.12
0.02	0.01	-0.01	-0.02	0.02	0.05	0.08	0.03	-0.05	-0.11
0	-0.02	-0.03	-0.04	-0.03	0.05	0.07	0.02	-0.06	-0.09
-0.01	-0.03	-0.07	-0.07	-0.06	-0	0.04	0.02	-0.01	-0.07
-0.03	-0.05	-0.11	-0.1	-0.1	-0.01	-0	0.02	-0.01	-0.03
-0.11	-0.14	-0.15	-0.17	-0.13	-0.09	-0.03	-0.03	-0.05	-0

Figure 2. (a) SST regions (outlined) used to generate self-organizing maps (SOMs) overlaid an example SST field (example SST field is assigned to the SOM unit in the upper rightmost corner). Mean SST is taken over the box corresponding to each region. (b) SOMs for all SST time steps (1940–2021). Red dots indicate units where >40% of testing samples fall within the 20% most confident predictions, and colored boxes correspond to SST regimes of enhanced prediction skill (see: Section 3.3). Pink outlined SOM units correspond to regime 1, cyan outlined units correspond to regime 2, green outlined units correspond to regime 3, and red outlined units correspond to regime 4. (c) ANN predicted mean of the Z500 distribution by regime.

originating from a given SOM unit fall within the 20% most confident ANN predictions. SOM units where this frequency exceeds 40% are marked with a red circle in Figure 2b, and the frequencies for each SOM unit can be seen in Figure S3 in Supporting Information S1.

3. Results

In this section, we discuss the performance of the ANN (Section 3.1) and the identification of state-dependent SST regimes that enhance Z500 seasonal prediction skill (Sections 3.2 and 3.3).

3.1. ANN Skill

We compute the mean absolute error (MAE) between the predicted mean of the Z500 distribution and the true Z500 to assess ANN performance. For all samples across the ANN ensemble, MAE is 0.756 σ. As ANN predicted standard deviation decreases, that is, ANN confidence increases, ANN MAE initially increases, before decreasing below the ANN baseline MAE for the top 5% most confident predictions (Figure S2 in Supporting Information S1), which comprises approximately 67 independent forecasts. We speculate that the initial increase in MAE with increasing ANN confidence reflects the relatively weak predictive signal in the data, rather than an overconfident or overfitting ANN. When instead looking at how often the ANN predicts the correct sign anomaly (with respect to the Z500 median), the ANN predicts the correct sign anomaly increasingly frequently with increasing ANN confidence (Figure 1c). This indicates the ANN ensemble is often predicting the correct sign anomaly for more confident predictions, and that the modest increase in MAE is reflective of the extreme nature of true Z500 for more confident predictions.

To evaluate ANN skill, we compare ANN performance to a hierarchy of baseline models (see Table 1). At the lowest level, we consider a persistence model that predicts the same week 7–8 Z500 as the Z500 at time of input (day 0), and results in a MAE of 1.08, a 42% increase in MAE relative to the ANN ensemble. We next consider a

Table 1
ANN and Baseline Models Mean Absolute Error and Percent of Predictions That Are the Same Sign as True Label (With Respect to True Label Median) for All Predictions and 5% Most Confident ANN Predictions

	Mean absolute error		% correct sign	
Model	All	Top 5%	All	Top 5%
$\text{ANN}_{ensemble}$ mean	0.756	0.751	54.3%	63.0%
MLR	0.763	0.758	53.6%	61.5%
Z500−Nino34	0.760	0.792	50.4%	43.6%
Climatology	0.761	0.795	N/A	N/A
Persistence	1.075	1.097	50.7%	53.1%

climatology model that always predicts the mean value, which is zero for the standardized Z500 data. The climatology model results in a MAE of ∼0.76, a modest 0.62% increase relative to the ANN ensemble. Finally, we consider a linear Nino34 model that predicts standardized Z500 using the ordinary least squares fit between standardized Z500 and standardized Nino34 index (Figure 1b). The linear Nino34 model only slightly improves upon climatology, resulting in a MAE of ∼0.76, or 0.57% increase relative to the ANN ensemble. We also implement a sign test to evaluate the frequency in which anomalous Z500 predictions are the correct sign (positive or negative) with respect to the median true Z500. For all testing samples, the ANN ensemble mean predicts the correct sign Z500 54.3% of the time, outperforming both the persistence model (50.7%) and linear Nino34 model (50.4%). Conditioning on the top 5% most confident predictions widens this performance gap, with the ANN predicting the correct sign 63% of the time, compared to 53.1% for persistence and 43.6% for the linear Nino34 model (Table 1). We speculate that the poor performance of the Nino34 model for higher confidence ANN predictions results from its inability to account for the modification of ENSO teleconnections by other SST modes (see: Section 4). To evaluate the added skill from the complex ANN approach, we also consider a multiple linear regression model (MLR), using the SOM regions outlined in the following section as predictors. We iteratively withhold 5-year testing periods for MLR to provide a suitable comparison to the ANN and avoid overfitting. The MLR baseline generally outperforms persistence, climatology, and the linear Nino34 model, but underperforms the ANN (Table 1), indicating that there is additional skill gained from the complex, nonlinear approach.

3.2. Identifying SST Regimes That Enhance Z500 Seasonal Predictability

We implement the SOM methodology outlined in Section 2.3 to detect SST patterns that enhance seasonal predictability of North Atlantic Z500. Figure 2a depicts the regions used to generate the SOM weight vectors and Figure 2b shows the SOMs, with dots denoting the units that most frequently feature the 20% most confident predictions across the ANN ensemble. A few things within Figure 2 are worth discussing: first, nearby units within the SOM indeed exhibit similar sign and magnitude SST anomalies when compared to distant units; second, nearby units within the SOM also exhibit similar sign and magnitude of predicted Z500 (Figure 2c); third, SOM units that enhance Z500 predictability tend to be focused on the extrema of the SOMs. Additionally, the sign and relative magnitude of predicted Z500 by SOM unit resembles the sign and magnitude of true Z500 by SOM unit (Figure S4 in Supporting Information S1), particularly at the extrema of the SOM where ensemble predictions are more confident and predicted Z500 is most extreme. Sammon mapping is used to visualize the SOM topology (Sammon Jr, 1969), which demonstrates that SOM units are distinct and ordered (Figure 3), reflective of a stable training process and small topological error $(Error_{topological} = 0.015)$.

Observing the SOM principle that nearby units are more similar than distant units, we group some adjacent units into four SST regimes where Z500 seasonal predictability is enhanced. Figure 2 reveals these four distinct regions in the SOMs as grouped units where predictions are more confident (Figure 2b), SST within each regime is of similar sign and magnitude (Figure 2b), and predicted Z500 is of similar sign and magnitude (Figure 2c). Although regimes 2 (cyan), 3 (green), and 4 (red) appear close in the 2-D SOM grid (Figure 2b), the SOM topology demonstrates that the relative Euclidean distance between the three regimes is large (Figure 3), and the regimes exhibit distinct SST patterns and anomalous Z500 (Section 3.3). Some units assigned to the four regimes of enhanced prediction skill fail to eclipse the 40% threshold of most confident predictions (Section 2.3), but still feature a higher frequency of most confident predictions relative to the average unit (Figure S3 in Supporting Information S1). For this reason, and because these units feature similar SST and predicted Z500 as the adjacent, most confident units, they are incorporated within the respective regimes.

3.3. SST Regimes of Enhanced Seasonal Prediction Skill

In this section, the SST regimes that enhance seasonal predictability of North Atlantic atmospheric circulation are discussed: a mature-to-decaying La Niña regime (Section 3.3.1), a strong El Niño regime (Section 3.3.2), and two regimes describing periods of ENSO transition (Section 3.3.3).

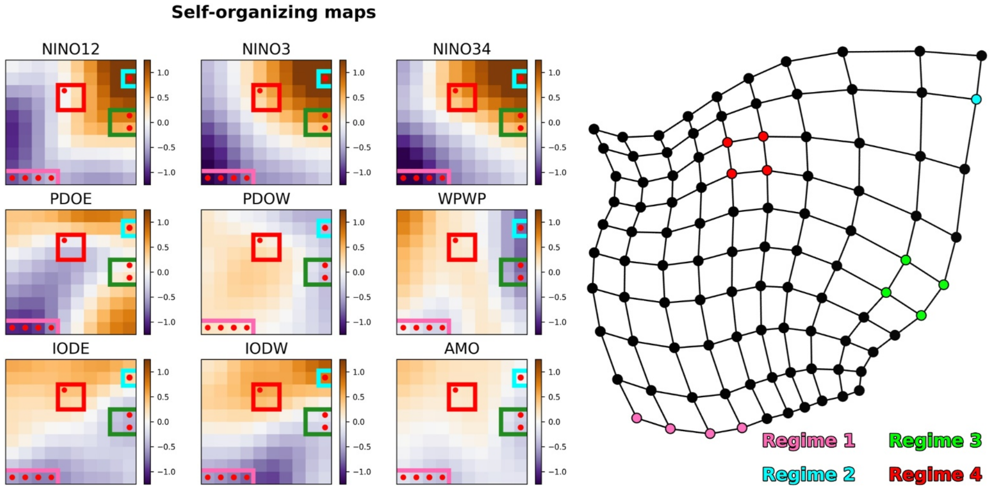


Figure 3. SOMs (left) and Sammon map (right). The Sammon map depicts the relative Euclidean distance between units within the corresponding self-organizing maps. The outlined regimes of enhanced prediction skill in the SOM (left) are shaded and labeled in the corresponding Sammon map.

3.3.1. Regime 1: Mature-to-Decaying La Niña

The regime 1 SST pattern (pink-enclosed SOM nodes in the lower left panels of Figure 2b) depicts a mature-to-decaying La Niña, with below average SST in the tropical Pacific, above average SST in the Maritime Continent region, a −PDO signal in the North Pacific and below normal SSTs across the Indian Ocean (Figure 4b). The corresponding predicted and true Z500 for regime 1 conforms to the established linear ENSO−Z500 connection, presenting a clear, negative signal in anomalous Z500, as 94% of all predictions and 62% of all true labels are negative (Figure 4a). Figure 4c depicts the regime 1 mean input×gradient, a machine learning explainability technique (XAI) in which sample inputs are multiplied by the local network gradient in an attempt to quantify the marginal contribution of a given input to the prediction (Mamalakis et al., 2022). The mean XAI pattern

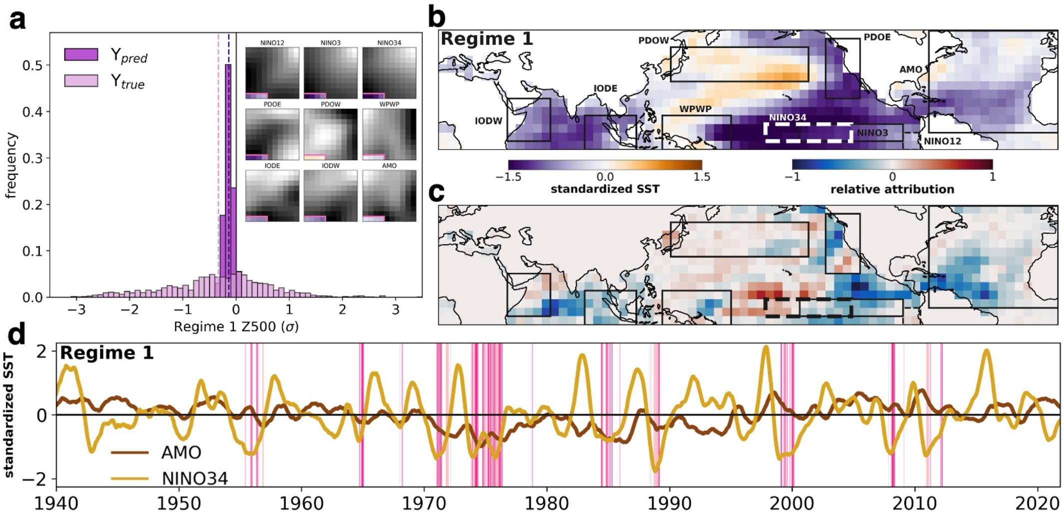


Figure 4. (a) ANN predictions (dark) versus true labels (light) for regime 1 SST samples, with inlaid SOMs highlighting regime 1 units. Dashed vertical lines represent mean predicted Z500 (dark) and mean true Z500 (light). (b) Mean SST pattern for regime 1 samples, with SOM regions overlaid. (c) Mean input*gradient for regime 1 samples, normalized by the absolute value of the largest magnitude input*gradient, with SOM regions overlaid. (d) Time series of low-pass filtered (365-day running mean) Nino34 region standardized SST (light brown) and AMO region standardized SST (dark brown), with time periods corresponding to regime 1 (pink) shaded. Lighter shading indicates a time step within a given regime corresponding to lower than average week 7−8 Z500, while darker shading indicates a time step within a given regime corresponding to higher than average week 7−8 Z500.

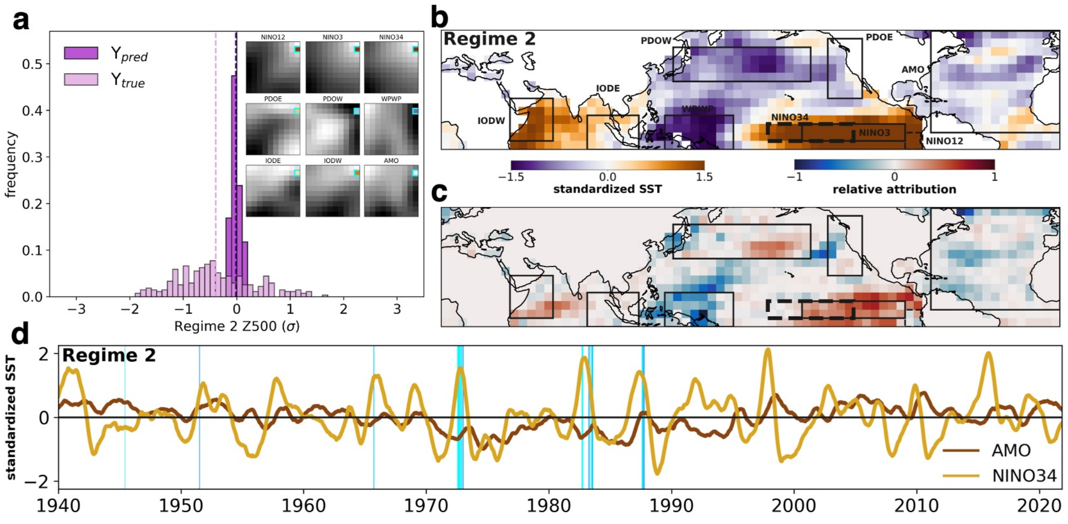


Figure 5. As in Figure 4, but for regime 2.

associated with regime 1 supports the regions selected to generate the SOMs, with large relative attribution assigned in and around the ENSO, PDO, and IODE regions, as well as the subtropical Atlantic. While the connection between La Niña events and lower than normal heights over the North Atlantic is well documented (Zhang, Wang, et al., 2019), the identification of this regime and the corresponding Z500 signal lends credibility to our regimes-based methodology.

3.3.2. Regime 2: Atypical Strong El Niño

Interestingly, the regimes-based analysis identifies an El Niño regime that enhances Z500 seasonal predictability (Figure 2b), but corresponds to lower than normal Z500 in both ANN predictions and true labels (Figure 5a). We refer to this regime as regime 2 (cyan-enclosed SOM node in the upper right panels in Figure 2b). Regime 2 depicts a moderate-to-strong El Niño, but with an unclear PDO signature, and lower SSTs in the WPWP and subtropical Atlantic relative to other strong El Niño regimes (Figure 5b). The corresponding predicted and true Z500 for regime 2 reveals a clear, negative signal in anomalous Z500 that opposes the established linear ENSO-Z500 connection, as 67% of all predictions and 72% of all true labels are negative (Figure 5a). There is a slight signal for extreme, negative Z500 in regime 2 true labels, but ANN predictions are generally modestly negative. Regime 2 generally occurs during the "shoulders" of extreme El Niño events, as well as during less intense El Niño events (e.g., 1951–1952; Figure 5d). XAI suggests that SST anomalies in the WPWP, North Pacific, and North Atlantic negatively contribute to predicted Z500 and interrupt the established linear ENSO−Z500 teleconnection (Figure 5c).

3.3.3. Regimes 3 and 4: ENSO Transition and Weak El Niño

The regime 3 SST pattern (green-enclosed SOM nodes in the center right panels of Figure 2b) broadly depicts SSTs associated with a period of ENSO transition (Figure 6). More specifically, the regime 3 SST pattern often corresponds to emerging El Niño events that coincide with a negative phase of the AMO, as 96% of regime 3 occurrences coincide with −AMO (Figure 6d). The emerging El Niño conditions are reflected in the weakly positive SST anomalies across all ENSO regions, while the Indian Ocean displays SST patterns more consistent with a La Niña state (Figure 6b). The corresponding predicted and true Z500 for regime 3 opposes the linear ENSO−Z500 relationship (Figure 6a), suggesting there may be a lag in this relationship during ENSO transitions. Regime 3 XAI indicates that WPWP, eastern Indian Ocean, subtropical Atlantic, and Kuroshio extension region are the primary sources of enhanced prediction skill (Figure 6c). Positive relative attribution values across the Nino34 region indicates that regime 3 ANN output, which is negative, is largely attributable to anomalous SST outside the ENSO regions.

Regime 4 SSTs (red-enclosed SOM notes in the center panels of Figure 2b) also depict periods of ENSO transition and ENSO neutral states, with weak-to-moderate positive SST anomalies across all ENSO regions (Figure 7). The

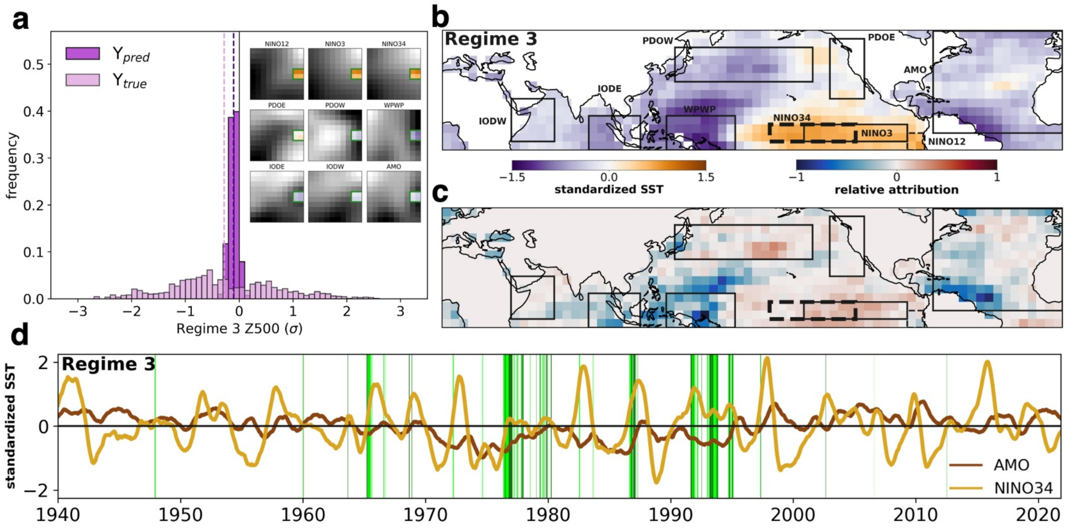


Figure 6. As in Figure 4, but for regime 3.

corresponding predicted and true Z500 for regime 4 reveals a slight signal for positive Z500 (Figure 7a). Regime 4 SSTs sometimes occur during El Niño to La Niña transitions (Figure 7d), and Indian Ocean SSTs and the weakly positive Z500 signal again supports the idea of lagging ENSO teleconnections. Regime 4 mean XAI reveals relatively larger, positive attribution values in the tropical and subtropical Pacific when compared to regime 3, suggesting that the ENSO region is a larger source of prediction skill for regime 4 than regime 3. However, regime 4 attribution values in the Indian and Atlantic Oceans, PDOE, and Kuroshio extension are largely the opposite sign of regime 3 attribution values in the same regions, signaling that there may be additional sources of prediction skill outside ENSO.

4. Prediction Skill Beyond ENSO

While regimes 1 and 4 conform to the established ENSO−Z500 relationship, the identification of regimes 2 and 3 as enhancing Z500 prediction skill demonstrates that sources of prediction skill exist outside the ENSO region. For all regime 2 samples ($N = 272$), the linear ENSO−Z500 model described in Section 3.1 predicts the correct sign Z500 anomaly 28% of the time, whereas the ANN predicts the correct sign Z500 anomaly 53% of the time (Table 2). Similarly, for all regime 3 samples ($N = 1107$), the linear ENSO−Z500 model predicts the correct sign

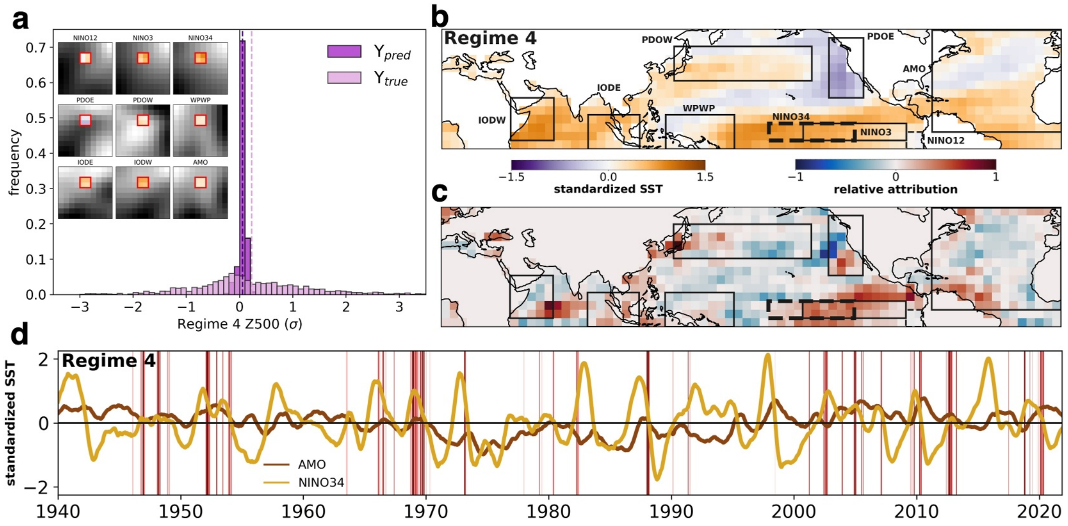


Figure 7. As in Figure 4, but for regime 4.

Table 2
ANN and Linear Nino34 Fit Mean Absolute Error and Percent Correct Sign (With Respect to the Median) for Regimes 1–4

Regime	Method	MAE	% correct sign
Regime 1	ANN	0.777	62.4%
	Nino34	0.781	57.7%
Regime 2	ANN	0.656	52.6%
	Nino34	0.734	28.3%
Regime 3	ANN	0.824	63.6%
	Nino34	0.878	38.0%
Regime 4	ANN	0.832	58.9%
	Nino34	0.832	56.7%

Z500 anomaly 38% of the time, whereas the ANN predicts the correct sign Z500 anomaly 64% of the time (Table 2). Furthermore, the ANN reduces MAE by 11% for regime 2 predictions and 6% for regime 3 predictions relative to the linear Nino34 model. In contrast, the ANN offers limited improvement over the linear ENSO−Z500 model for regimes 1 and 4, which exhibit the frequently cited linear relationship between ENSO and Z500.

Machine learning explainability (XAI) indicates that positive ENSO SST anomalies contribute toward higher than normal Z500 for regime 2 (Figure 5c). However, discrepancies in the corresponding Z500 signal between regime 2 and other strong El Niño regimes (Figure 2c) suggests that a linear ENSO−Z500 relationship may only exist for canonical El Niño periods that coincide with a +PDO. A +PDO SST pattern is typically reflective of a deeper Aleutian Low, which enhances wind speeds and, consequently, surface fluxes over the central north Pacific while simultaneously advecting relatively warmer air over western North America (e.g., Alexander et al., 2002; Dow et al., 2024). A deeper Aleutian Low drives downstream El Niño teleconnections over North America through promotion of wave number 1 growth and vertical propagation, as well as through a midlatitude Rossby wave pattern. The absence of a coherent +PDO signal during strong El Niños (e.g., regime 2) may reflect a relatively weaker Aleutian Low, and thus, an interruption of conventional El Niño teleconnections. To assess this, we evaluate composite 500 mb heights averaged over days 53–67 following regime 2 occurrences, as well as for the three strong El Niño regimes to the left, above, and diagonally above regime 2 in the SOM grid (Figure 2b). The three strong El Niño regimes surrounding regime 2 are hereafter referred to as "canonical El Niño regimes." For canonical El Niño regimes, composite heights display a deeper than normal Aleutian Low and traditional El Niño teleconnection pattern over North America and the North Atlantic (Figure 8). However, regime 2 composite heights display higher than normal heights over the Aleutian Low region and a dipole-like height structure over the western North Atlantic. Maher et al. (2022) identified a

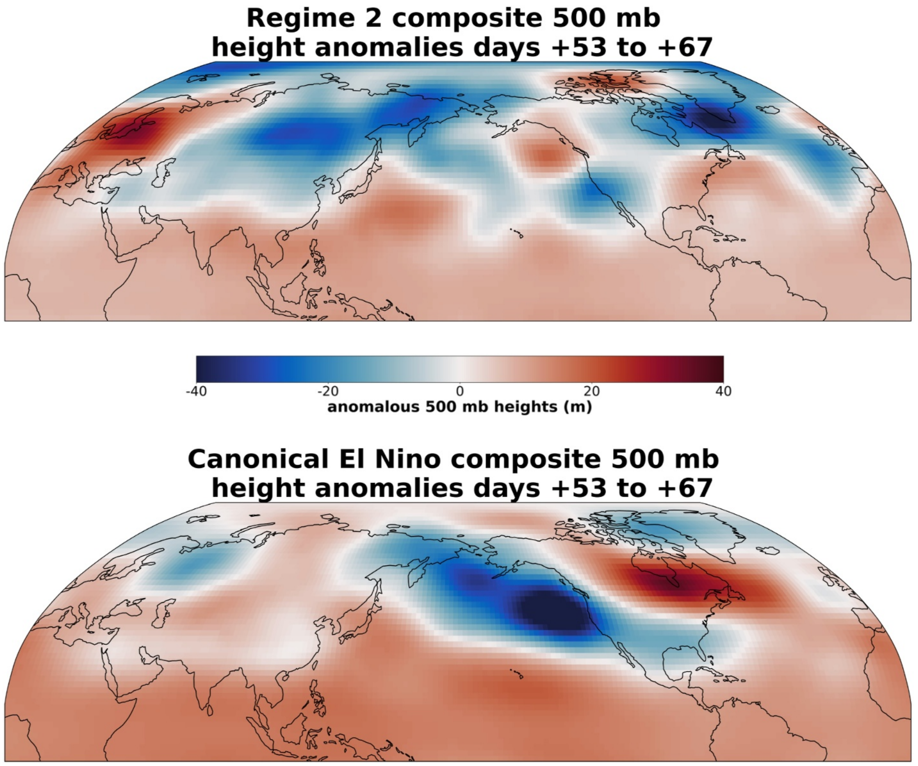


Figure 8. Composite ERA5 500 mb height anomalies averaged over days +53 to +67 following regime 2 and canonical El Niño regime occurrences.

similar PDO modification of Aleutian Low strength during ENSO events, and corresponding modification of ENSO impacts on North American temperature and precipitation, using single model initial-condition large ensembles. Our findings indicate PDO modification of El Niño teleconnections extends to North Atlantic circulation patterns.

In addition to the negative attribution from PDOE SST anomalies, regime 2 XAI suggests a linkage between extreme negative WPWP SST anomalies and lower than normal Z500. Regime 2 composite 500 mb height anomalies offer a physical explanation for this linkage, in the form of a wavenumber 2 Rossby wave signature extending northeast from the West Pacific to the North Atlantic (Figure 8). Notably, less extreme negative WPWP SST anomalies during canonical El Niño regimes contribute only modestly to predicted Z500 (not shown). The weaker WPWP negative contributions to Z500 during canonical El Niño regimes are overwhelmed by positive contributions from the ENSO regions, indicating that this potential linkage is sensitive to a WPWP SST threshold. Abid et al. (2021) also identified a distinction in El Niño teleconnections, specifically between boreal fall and boreal winter. They find that positive IOD events generate a wave train response that projects positively onto the NAO and dominates the boreal fall teleconnection, whereas the canonical El Niño wave train that projects negatively onto the NAO dominates during boreal winter. While regime 2 also occurs most frequently during boreal fall (59% of occurrences) and projects positively onto the NAO, our findings suggest a weaker Aleutian Low and possibly WPWP-induced wave train drive the teleconnection discrepancy, rather than an IOD-induced wave train.

Discrepancies in Z500 signal between regimes 3 and 4 appear partially connected to lagging ENSO teleconnections to remote SSTs (e.g., Indian Ocean), and consequently, Z500. Regime 3 Indian Ocean SSTs are consistent with an El Niño event in its infancy, while regime 4 Indian Ocean SSTs are consistent with a mature to decaying El Niño event (e.g., Alexander et al., 2002). XAI identifies the Indian Ocean as an important source of prediction skill for both regimes (Figures 6c and 7c). However, both regimes also occur during ENSO neutral periods and in weak-to-moderate El Niño events, indicating that lagging ENSO teleconnections likely do not account for the Z500 discrepancy, alone. The Atlantic Ocean is another region of SST differences between the two regimes, as regime 3 occurs almost exclusively during −AMO and regime 4 occurs more frequently during +AMO periods. XAI also highlights the Atlantic Ocean as an important source of prediction skill for both regimes, suggesting that Atlantic decadal SST variability may modulate the ENSO−Z500 relationship during weak-to-moderate El Niños. Zhang, Mei, et al. (2019) find that the AMO enhances ENSO teleconnections to the North Atlantic when the two are "in-phase" and diminishes the teleconnections when the two are "out-of-phase," which is consistent with Z500 discrepancies between regimes 3 and 4.

Notably, regime 3 composite SSTs depict a weak-to-moderate El Niño and +PDO and regime 4 composite SSTs depicts a weak-to-moderate El Niño and −PDO (Figures 6b and 7b). The corresponding Z500 signal for regimes 3 and 4 opposes the PDO modulation of El Niño teleconnections discussed above for regime 2, and appears instead to be driven by AMO and Indian Ocean SST variability. Meanwhile, SOM units corresponding to canonical El Niño regimes depict strong El Niños alongside −AMO (not shown). Composite Z500 following canonical El Niño events depicts a deep Aleutian Low and higher than normal Z500, suggesting the PDO exhibits a larger influence on ENSO−Z500 relationship than the AMO for strong El Niños. Collectively, this demonstrates the importance of accounting for event-to-event variability, and considering co-variability between ENSO, PDO, and AMO signals for seasonal Z500 prediction.

5. Conclusions and Summary

We present a methodology for identifying state-dependent prediction skill using artificial neural networks (ANNs) and self-organizing maps (SOMs), and apply this methodology to identify SST regimes that enhance seasonal predictability of North Atlantic atmospheric circulation. Following Barnes et al. (2021), regression-based ANNs receive SST as input and output the mean and standard deviation of a conditional Gaussian distribution of 500 mb heights (Z500) over the North Atlantic 7–8 weeks later (Figure 1a). Thus, ANN output inherently contains information on the confidence of a given prediction, as indicated by the standard deviation, or width, of the Z500 distribution (Barnes et al., 2021; Gordon & Barnes, 2022). We exploit this to separately evaluate more confident ANN predictions, and demonstrate that the ANN ensemble increasingly outperforms baseline metrics as ANN confidence increases (Figure 1c; Table 1). Prominent modes of SST variability with previously documented influence on North Atlantic atmospheric circulation are used to generate the SOMs

(Section 2.3), and four SST regimes are identified by assessing which SOM units the SST input associated with the most confident ANN predictions are assigned to (Section 3.2, Figure 2).

Many previous studies identify an approximately linear relationship between El Niño southern oscillation (ENSO) and Z500 in our target region (e.g., Scaife et al., 2014), but debate remains about how event-to-event ENSO variability may modify this relationship (Abid et al., 2021; Zhang, Wang, et al., 2019). Our results suggest that ENSO events with conventional remote SSTs may be more likely to conform to the linear ENSO−Z500 relationship (Section 3.3). Specifically, we find that strong El Niño events concurrent with a positive phase of the Pacific Decadal Oscillation (PDO) promote a tropical−extratropical teleconnection leading to higher than normal Z500, whereas strong El Niño events concurrent with a −PDO promote lower than normal Z500. Composite analysis suggests this is connected to PDO modification of the intensity of the Aleutian Low (Figure 8) and subsequent extratropical Rossby wave pattern. This supports recent findings that Pacific Decadal Variability strengthens ENSO teleconnections when the two are in-phase and weakens ENSO teleconnections when the two are out-of-phase (Dong & Dai, 2015; Dong et al., 2018; Maher et al., 2022), and that this modification is more apparent during El Niño events (Dong et al., 2018). Additionally, our findings indicate that the ENSO−Z500 relationship during strong El Niños is sensitive to the magnitude of anomalous negative West Pacific Warm Pool SSTs (Section 4). During ENSO transition periods and weak-to-moderate El Niño events, we find predictability of Z500 related to the combined SST anomalies of the West Pacific Warm Pool, the North Atlantic, and the North Pacific. Sources of Z500 prediction skill beyond ENSO during weak ENSO events and ENSO neutral conditions convey the importance of a holistic approach to seasonal Z500 prediction.

While our results demonstrate the importance of considering SST variability beyond ENSO in seasonal prediction of North Atlantic atmospheric circulation, the sample size evaluated in this study is limited by the observational record. Thus, we may not capture the full range of co-variability amongst the nine SOM SST regions. The regimes identified here are constrained to those that occur frequently enough to allow for increased ANN confidence in its associated predictions. Additionally, regime 2 features limited unique occurrences over the observational record (Figure 5d). However, the Aleutian Low response to regime 2 SST is documented here and by previous work (e.g., Maher et al., 2022), increasing our confidence in this finding. In contrast, there is notable event-to-event variability in the Z500 wave train response to regime 2 events, but more persistent regime 2 events exhibit more robust wavenumber 2 wave train patterns extending from the west Pacific (not shown).

ANN predictions for the four regimes of enhanced prediction skill lends credence to the methodology outlined in this paper. We impose no requirement that predictions be “correct” when evaluating the most confident predictions, yet ANN predictions and true labels exhibit similar mean Z500 values for all four SST regimes of enhanced prediction skill (Figures 4a–7a). Additionally, two identified regimes, conventional El Niño and La Niña cases, are associated with well-documented physical connections between input (SST) and output (Z500), which builds confidence in the methodology. Explainable machine learning techniques indicate that the predicted output for all four regimes is largely attributable to anomalous SST within the regions used to construct the SOMs (Figures 4c–7c). This affirms the SST regions we selected to construct the SOMs, and also demonstrates the importance of a priori knowledge of potential relationships between input and output when considering this methodology.

We view the regimes-based methodology detailed in this paper as an adaptable framework with the ability to identify state-dependent prediction skill in atmospheric and oceanic problems on subseasonal-to-decadal timescales. The methodology is well-suited for many subseasonal-to-seasonal prediction problems, such as the timing and intensity of the Asian monsoon (e.g., Kim et al., 2012), regional droughts and heat waves (e.g., Hao et al., 2018), cold air outbreaks (e.g., Albers et al., 2022), and more straightforward problems like single location precipitation and temperature forecasts. On decadal time scales, slowly varying ocean modes could be targeted (e.g., Pacific Decadal Variability). In practice, the methodology could also be applied to shorter time scale problems, for example, ingredients-based forecasts of severe weather (Mazurek et al., 2025).

Data Availability Statement

ERA-5 reanalysis data on single levels (e.g., SST) and pressure levels (e.g., 500 mb heights) is publicly available from the Copernicus Climate Data Store: https://cds.climate.copernicus.eu/. ANNs were developed using

Tensorflow (Abadi et al., 2016) and SOMs were developed using MiniSom (Vettigli, 2018). Climate Data Operators (CDO; Schulzweida, 2019) were used for data interpolation.

Acknowledgments

We thank Kristian Strommen and one anonymous reviewer for their substantial contributions to this paper during the peer review process. We also would like to acknowledge and thank Eric Maloney for insightful comments on preliminary results from this study. This work is supported by NSF Physical Oceanography division (Grant 1924659) and is supported, in part, by NOAA Grant NA22OAR4310621 and NSF Grant RISE-2019758 (EAB).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, M., et al. (2016). Tensorflow: A system for large-scale machine learning [Software]. *12th USENIX Symposium on Operating Systems Design and Implementation* (pp. 265–283). https://doi.org/10.48550/arXiv.1605.0869

Abid, M. A., Kucharski, F., Molteni, F., Kang, I.-S., Tompkins, A. M., & Almazroui, M. (2021). Separating the Indian and Pacific Ocean impacts on the Euro-Atlantic response to ENSO and its transition from early to late winter. *Journal of Climate*, *34*(4), 1531–1548. https://doi.org/10.1175/JCLI-D-20-0075.1

Albers, J. R., Newman, M., Hoell, A., Breeden, M. L., Wang, Y., & Lou, J. (2022). The February 2021 cold air outbreak in the United States: A subseasonal forecast of opportunity. *Bulletin of the American Meteorological Society*, *103*(12), E2887–E2904. https://doi.org/10.1175/BAMS-D-21-0266.1

Alexander, M. A., Bladé, I., Newman, M., Lanzate, J. R., Lau, N.-C., & Scott, J. D. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *Journal of Climate*, *15*, 2205–2231. https://doi.org/10.1175/1520-0442(2002)015⟨2205:TABTIO⟩2.0.CO;2

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. https://doi.org/10.1126/science.1063315

Barnes, E. A., Barnes, R. J., & Gordillo, N. (2021). Adding uncertainty to neural network regression tasks in the geosciences. https://doi.org/10.48550/arXiv.2109.07250

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review*, *97*(3), 163–172. https://doi.org/10.1175/1520-0493(1969)097⟨0163:ATFTEP⟩2.3.CO;2

Brönnimann, S., Xoplaki, E., Casty, C., Pauling, A., & Luterbacher, J. (2007). ENSO influence on Europe during the last centuries. *Climate Dynamics*, *28*, 181–197. https://doi.org/10.1007/s00382-006-0175-z

Cassano, J. J., Uotila, P., & Lynch, A. (2006). Changes in synoptic weather patterns in the polar regions in the twentieth and twenty-first centuries, Part 1: Arctic. *International Journal of Climatology*, *26*(8), 1027–1049. https://doi.org/10.1002/joc.1306

Davini, P., Cagnazzo, C., Neale, R., & Tribbia, J. (2012). Coupling between Greenland blocking and the North Atlantic Oscillation pattern. *Geophysical Research Letters*, *39*(14), L14701. https://doi.org/10.1029/2012GL052315

Dong, B., & Dai, A. (2015). The influence of the Interdecadal Pacific Oscillation on temperature and precipitation over the globe. *Climate Dynamics*, *45*(9–10), 2667–2681. https://doi.org/10.1007/s00382-015-2500-x

Dong, B., Dai, A., Vuille, M., & Timm, O. E. (2018). The influence of the interdecadal Pacific oscillation on temperature and precipitation over the globe. *Journal of Climate*, *31*(18), 7337–7361. https://doi.org/10.1175/JCLI-D-17-0663.1

Dow, W. J., McKenna, C. M., Joshi, M. M., Blaker, A. T., Rigby, R., & Maycock, A. C. (2024). Sustained intensification of the Aleutian low induces weak tropical Pacific Sea surface warming. *Weather and Climate Dynamics*, *5*(1), 357–367. https://doi.org/10.5194/wcd-5-357-2024

Esit, M., Kumar, S., Pandey, A., Lawrence, D. M., Rangwala, I., & Yeager, S. (2021). Seasonal to multi-year soil moisture drought forecasting. *npj Climate and Atmospheric Sciences*, *4*(16), 181–197. https://doi.org/10.1038/s41612-021-00172-z

Gastineau, G., García-Serrano, J., & Frankignoul, C. (2017). The influence of autumnal Eurasian snow cover on climate and its link with Arctic Sea ice cover. *Journal of Climate*, *30*(19), 7599–7619. https://doi.org/10.1175/JCLI-D-16-0623.1

Gervais, M., Atallah, E., Gyakum, J. R., & Tremblay, L. B. (2016). Arctic air masses in a warming world. *Journal of Climate*, *29*(7), 2359–2373. https://doi.org/10.1175/JCLI-D-15-0499.1

Gordon, E. M., & Barnes, E. A. (2022). Incorporating uncertainty into a regression neural network enables identification of decadal state-dependent predictability in CESM2. *Geophysical Research Letters*, *49*(15), e2022GL098635. https://doi.org/10.1029/2022GL098635

Gu, Q., & Gervais, M. (2021). Exploring North Atlantic and North Pacific decadal climate prediction using self-organizing maps. *Journal of Climate*, *34*(1), 123–141. https://doi.org/10.1175/JCLI-D-20-0017.1

Hall, R. J., Hua-Liang, W., & Edward, H. (2019). Complex systems modelling for statistical forecasting of winter North Atlantic atmospheric variability: A new approach to North Atlantic seasonal forecasting. *Quarterly Journal of the Royal Meteorological Society*, *145*(723), 2568–2585. https://doi.org/10.1002/qj.3579

Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, *56*(1), 108–141. https://doi.org/10.1002/2016RG000549

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Knight, J. R., Davies, P., et al. (2020). Predictability of European winter 2019/20: Indian Ocean Dipole impacts on the NAO. *Atmospheric Science Letters*, *21*(12), e1005. https://doi.org/10.1002/asl.1005

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis [Dataset]. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hoskins, B. J., & Karoly, D. (1981). The steady linear response of a spherical atmosphere to thermal and orographic forcing. *Journal of the Atmospheric Sciences*, *38*, 1179–1196. https://doi.org/10.1175/1520-0469(1981)038⟨1179:TSLROA⟩2.0.CO;2

Hurrell, J., Kushnir, Y., Ottersen, G., & Visbeck, M. (2003). An overview of the North Atlantic oscillation. The North Atlantic Oscillation: Climatic significance and. *Environmental Impact*, *134*, 1–35. https://doi.org/10.1029/134GM01

Kim, H. M., Webster, P. J., Curry, J., & Toma, V. E. (2012). Asian summer monsoon prediction in ECMWF system 4 and NCEP CFSV2 retrospective seasonal forecasts. *Climate Dynamics*, *39*(12), 2975–2991. https://doi.org/10.1007/s00382-012-1470-5

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. https://doi.org/10.1007/BF00337288

Maher, N., Kay, J. E., & Capotondi, A. (2022). Modulation of ENSO teleconnections over North America by the Pacific decadal oscillation. *Environmental Research Letters*, *17*(11), 114005. https://doi.org/10.1088/1748-9326/ac9327

Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2022). Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *Environmental Data Science*, *1*, e8. https://doi.org/10.1017/eds.2022.7

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., et al. (2020). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, *101*(5), E608–E625. https://doi.org/10.1175/BAMS-D-18-0326.1

Mayer, K. J., & Barnes, E. A. (2021). Subseasonal forecasts of opportunity identified by an explainable neural network. *Geophysical Research Letters*, *48*(10), e2020GL092092. https://doi.org/10.1029/2020GL092092

Mazurek, A. C., Hill, A. J., Schumacher, R. S., & McDaniel, H. J. (2025). Can ingredients-based forecasting be learned? Disentangling a random forest's severe weather predictions. *Weather and Forecasting*, *40*(2), 237–258. https://doi.org/10.1175/WAF-D-23-0193.1

Mezzina, B., García-Serrano, J., Bladé, I., & Kucharski, F. (2020). Dynamics of the ENSO teleconnection and NAO variability in the North Atlantic–European late winter. *Journal of Climate*, *33*(3), 907–923. https://doi.org/10.1175/JCLI-D-19-0192.1

Nix, D., & Weigend, A. S. (1994). Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*. https://doi.org/10.1109/ICNN.1994.374138

Palmer, T., & Hagedorn, R. (2006). *Predictability of weather and climate*. Cambridge University Press. https://doi.org/10.1017/CBO9780511617652

Roberts, A. R., Vigaud, N., Yuan, J., & Tippett, M. K. (2020). Toward identifying subseasonal forecasts of opportunity using North American weather regimes. *Monthly Weather Review*, *148*(5), 1861–1875. https://doi.org/10.1175/MWR-D-19-0285.1

Sammon, J. W., Jr. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, *C-18*(5), 401–409. https://doi.org/10.1109/t-c.1969.222678

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*(7), 2514–2519. https://doi.org/10.1002/2014GL059637

Schulzweida, U. (2019). CDO user guide [Software]. https://doi.org/10.5281/zenodo.3539275

Sutton, R. T., & Hodson, D. L. R. (2003). Influence of the ocean on North Atlantic climate variability 1871–1999. *Journal of Climate*, *16*, 3296–3313. https://doi.org/10.1175/1520-0442(2003)016⟨3296:IOTOON⟩2.0.CO;2

Vettigli, G. (2018). MiniSom: Minimalistic and numpy-based implementation of the self organizing map [Software]. https://github.com/JustGlowing/minisom/

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, *109*, 784–812. https://doi.org/10.1175/1520-0493(1981)109⟨0784:TITGHF⟩2.0.CO;2

Wang, Z., Li, H., Ouyang, W., & Wang, X. (2016). Learnable histogram: Statistical context features for deep neural networks. *arXiv*. https://doi.org/10.48550/arXiv.1804.09398

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., & Palmer, T. (2019). 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. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 917–926. https://doi.org/10.1002/qj.2976

Zhang, W., Mei, X., Geng, X., Turner, A. G., & Jin, F.-F. (2019). A nonstationary ENSO–NAO relationship due to AMO modulation. *Journal of Climate*, *32*(1), 33–43. https://doi.org/10.1175/JCLI-D-18-0365.1

Zhang, W., Wang, Z., Stuecker, M. F., Turner, A. G., Jin, F.-F., & Geng, X. (2019). Impact of ENSO longitudinal position on teleconnections to the NAO. *Climate Dynamics*, *52*(1–2), 257–274. https://doi.org/10.1007/s00382-018-4135-1