Incorporating Uncertainty into a Regression Neural Network Enables Identification of Decadal State-Dependent Predictability in CESM2

Emily M. Gordon[1] **and Elizabeth A. Barnes**[1]

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

Key Points:

- Artificial neural networks skillfully predict sea surface temperatures on decadal timescales in CESM2.
- The networks identify predictability by assigning lower uncertainty to initial states that lead to lower prediction error.
- More predictable initial states coincide with combinations of phases of large scale decadal variability.

Corresponding author: E. M. Gordon, `emily.m.gordon95@gmail.com`

Abstract

Predictable internal climate variability on decadal timescales (2-10 years) is associated with large-scale oceanic processes, however these predictable signals may be masked by the noisy climate system. One approach to overcoming this problem is investigating state-dependent predictability - how differences in prediction skill depend on the initial state of the system. We present a machine learning approach to identify state-dependent predictability on decadal timescales in the Community Earth System Model version 2 pre-industrial control simulation by incorporating uncertainty estimates into a regression neural network. We leverage the network's prediction of uncertainty to examine state dependent predictability in sea surface temperatures by focusing on predictions with the lowest uncertainty outputs. In particular, we study two regions of the global ocean - the North Atlantic and North Pacific - and find that skillful initial states identified by the neural network correspond to particular phases of Atlantic multi-decadal variability and the interdecadal Pacific oscillation.

Plain Language Summary

As the climate warms with anthropogenic climate change, it is increasingly important to predict long term climate variability in order to prepare for possible extremes. However, the Earth's climate is chaotic and deciphering predictable long-term signals from this noisy system has proven challenging. Here we leverage times where predictable signals rise above the noise and the long-term forecasts have less error. We present a machine learning approach to identify these times when the climate is more predictable and show that these are related to particular patterns of heat in the Atlantic and Pacific Oceans.

1 Introduction

Predicting the evolution of the climate on decadal timescales (2-10 year) has far reaching implications for both climate science and society. On these timescales, changes in climate patterns are associated with the forced response to anthropogenic emissions and internal variability in ocean (Meehl et al., 2021). For example, the forced response from climate change can manifest as the steady increase of global mean temperature which provides some predictability of future temperatures. Decadal predictability of oceanic temperature variability arises from the ocean's ability to store, release and transport heat on decadal timescales. Major modes of variability in the Pacific and Atlantic Oceans are therefore linked to decadal predictability as they indicate the spatial distribution of heat in these basins. Furthermore, this internal variability in the ocean can act to either mask or amplify the forced response from climate change (Trenberth & Fasullo, 2013). The Pacific Ocean exhibits long-term variability via the interdecadal Pacific oscillation (IPO Power et al., 1999; Meehl et al., 2013) and its related mode Pacific decadal variability (PDV, Mantua et al., 1997; Y. Zhang et al., 1997). Atlantic multi-decadal variability (AMV, Enfield et al., 2001; Xie & Tanimoto, 1998) is considered the dominant form of long-term variability in the Atlantic ocean, however whether variability arises due to internal Earth system processes or external forcing is still under debate (Clement et al., 2015; Mann et al., 2021; Booth et al., 2012). Because these patterns of variability are associated with decadal predictability, decadal prediction is traditionally focused on either investigating and predicting the processes themselves, (e.g. Meehl et al., 2016; Gordon et al., 2021; R. Zhang et al., 2019), or exploring the predictability that arises from the atmospheric teleconnections driven by these patterns (e.g. R. Zhang & Delworth, 2006; Simpson et al., 2018, 2019).

As hinted at above, it is difficult to decipher the drivers of predictability in observations and historical simulations as it is influenced by the non-linear interactions between internal variability and external forcing. Studies have diagnosed predictability in pre-industrial control runs (Branstator et al., 2012), while others have deciphered predictability from internal variability in model hindcast ensembles with accompanying unforced ensembles (Yeager et al., 2018; Borchert et al., 2021). Another avenue of research has been to quantify (using various metrics) how much predictability is present in different regions of the ocean, and what the relative contributions of internal and external drivers may be(Boer, 2011; Branstator & Teng, 2010). However, predictability in the climate system can vary drastically depending on region, timescale, and initial state (Christensen et al., 2020; Meehl et al., 2021; Mariotti et al., 2020) thus studies have encouraged a shift of focus towards the concept of state-dependent predictability (Pohlmann et al., 2004; Msadek et al., 2010; Merryfield et al., 2020; Mariotti et al., 2020). This paradigm intrinsically acknowledges that some initial states lead to more predictable behavior than others. The aim is therefore to identify these more predictable initial states, as they provide the opportunity to make more skillful forecasts. State-dependent predictability has been investigated on short (subseasonal to seasonal) timescales as the identification of “forecasts of opportunity” (Albers & Newman, 2019; Mayer & Barnes, 2021) An example of an oceanic region with decadal state-dependent predictability is the North Atlantic Subpolar Gyre. It has been found that anomalously strong ocean heat transport in the North Atlantic ocean is associated with skillful predictions of sea surface temperature (SST) in the North Atlantic Subpolar Gyre for lead times up to 8 years (Brune et al., 2018; Borchert et al., 2018). So enhanced heat transport in the North Atlantic could be considered a more predictable initial state for predicting North Atlantic SSTs.

With this increased focus on state-dependent predictability, it is necessary to explore methods that can identify state-dependent predictability. Machine learning is one such method that shows promise for identifying more predictable initial states. In fact, on subseasonal timescales, classification artificial neural networks (ANNs) have been shown to objectively identify states of the Madden-Julian oscillation that lead to enhanced predictability of circulation in the North Atlantic (Mayer & Barnes, 2021) by leveraging the network’s confidence in a prediction to identify state-dependent predictability. Furthermore, on decadal timescales it has been demonstrated that ANNs can skillfully predict decadal processes (Gordon et al., 2021; Labe & Barnes, 2022) and identify states of enhanced predictability of surface temperature over land (Toms et al., 2021).

This study introduces the identification of state-dependent predictability on decadal timescales using a regression-based neural network to predict sea surface temperatures (SSTs) across the globe within the Community Earth System Model, version 2 (CESM2, Danabasoglu et al., 2020) pre-industrial control simulation. We demonstrate a powerful technique for incorporating uncertainty into the prediction of regression neural networks which has previously only been used a handful of times in climate science (Foster et al., 2021; Guillaumin & Zanna, 2021; Barnes & Barnes, 2021). We further leverage this uncertainty output to identify which initial states are associated with the lower uncertainty predictions. Lower uncertainty predictions imply more predictable inputs, hence this technique identifies state-dependent predictability. Furthermore, we link predictable initial states to major forms of variability so we are able to identify certain combinations of IPO and AMV phases that correspond to skillful decadal predictions of SSTs in CESM2.

2 Data and Methods

2.1 Data

We use sea surface temperature (SST) and ocean heat content (OHC) output from the CESM2 pre-industrial control run for the Coupled Model Intercomparison Project phase 6 (CMIP6; Eyring et al., 2016). OHC is interpolated to a $4^\circ \times 4^\circ$ grid. We train ANNs at each SST grid point so SST is interpolated to a $5^\circ \times 5^\circ$ grid which captures the regional variation in predictability while not being too computationally demanding. We use monthly output of the 2000 year run with the first 100 years removed to allow the ocean circulation to spin-up. Both OHC and SST are then de-seasonalized by remov-

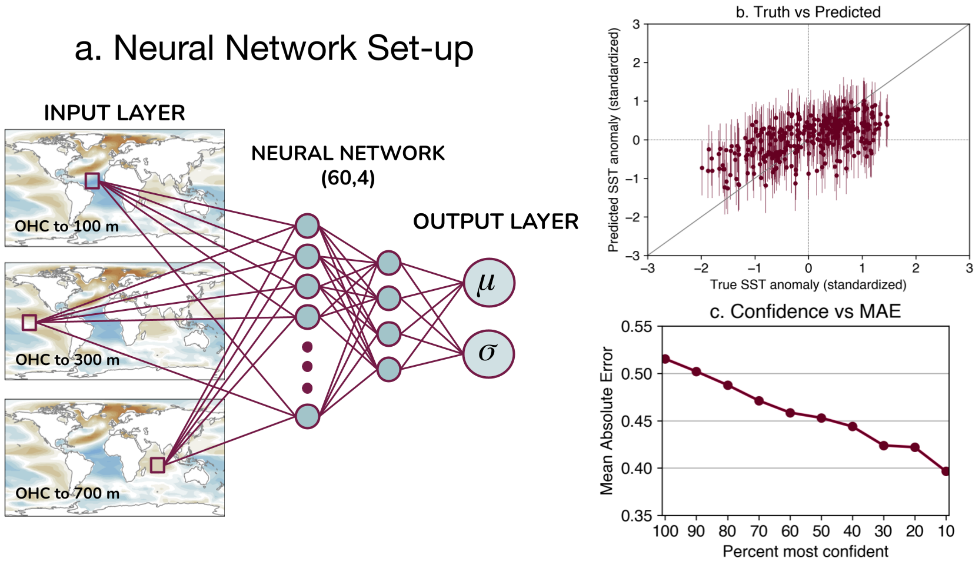

Figure 1. a. Schematic of the artificial neural network architecture. b. Scatter plot of predicted SST anomaly (y axis) vs true SST anomaly (x axis). Dots represent predicted μ values, while vertical lines represent the 1σ range. c. Prediction mean absolute error (MAE) as a function of prediction confidence (see text). Both b. and c. utilize the same network trained to predict SST in the North Atlantic Ocean (52.5°N, 325°E).

.

ing the mean annual cycle from each grid point. Furthermore, to account for model drift, after deseasonalizing we calculate the third degree polynomial trend via least squares and subtract this from each grid point. This means that each variable's statistics are approximately stationary for the remaining 1900 years of data. OHC is smoothed using a 60 month backward running mean to smooth high frequency variability. We divide the preprocessed data into training, validation and testing. The first 70% (~1300 years) is used for training, the next 15% (~300 years) for validation and the last 15% (~300 years) for testing. We calculate the mean and standard deviation for every point on both the OHC and SST grids in the training set. We then use these values to standardize all of the training, validation and testing data.

2.2 Artificial Neural Network

Artificial neural networks (ANNs) are used to predict the average SST anomaly at a lead time of 1-5 years and 3-7 years, i.e. the ANN predicts the average 60 month SST anomaly in the next 12-72 months, or 36-96 months respectively. In this experiment the ANN is trained to predict the SST evolution in the CESM2 pre-industrial control, so for example, one input sample is OHC information from a specific time step in the control run, and the output prediction is the average SST anomaly over the next 12-72 months in the control run. A schematic of our neural network architecture is provided in Figure 1a and a brief overview of ANNs for geoscience applications can be found in e.g. Toms et al. (2020). The predictors are three OHC grids, where each grid is OHC integrated to a different depth (100 m, 300 m and 700 m). We chose varying depths of OHC because each contains information corresponding to different forms of climate variability. For example, the upper levels of the ocean integrate atmospheric forcing, and hence capture atmospheric variability as well as surface ocean dynamics (Frankignoul & Hasselmann, 1977). The variability in lower levels of the ocean is guided by a combination of slow moving ocean circulation and the incorporation of mixed layer processes via the annual cycle in the thermocline (Alexander & Deser, 1995). By inputting three OHC depths into the neural network, it can theoretically combine different oceanic and atmospheric processes to make its predictions. The three ocean grids are vectorized with points over land removed resulting in a total 7947 input pixels. This input is connected to a hidden layer of 60 nodes which is then connected to another hidden layer of 4 nodes (see Fig.1). In this network, all layers are densely connected meaning all nodes in the previous layer are connected to all the nodes in the next layer. Furthermore, all nodes in the hidden layers use the rectified linear unit (ReLU) activation function. Finally this second layer is connected to the output layer of two nodes which serve as the parameters of the predicted conditional distribution (see details in the next paragraph). Here the distribution is a normal distribution as we found allowing skewness did not significantly improve the network's performance (not shown).

We use the $-\log(p)$ loss function described by e.g. Barnes et al. (2021) which we will summarize briefly. For each input, the network outputs two values, μ and σ. To calculate loss, μ and σ are used to construct a conditional distribution, d and the negative log likelihood function is calculated at the true value (y_{true}), i.e. loss = -log($p(y_{true}|d)$). This means that the neural network can decrease loss (decrease -log($p(y_{true}|d)$)) in different ways: either with a low σ value and μ that is close to y_{true}, or predict a larger σ value with μ that is further from y_{true}, or both. The neural network is therefore not penalized for high error predictions as long as it also guesses a correspondingly high σ value, that is, if it recognizes an input is less predictable by assigning a high σ value. The predictions of such an ANN are illustrated in Figure 1b, where we show an example scatter plot of prediction vs truth from an ANN trained to predict SST anomaly in the North Atlantic Subpolar Gyre. Note that we can plot both the predicted anomaly value (μ, colored dots) and an uncertainty range, with the error bars indicating the $\pm 1\sigma$ range predicted by the ANN. The ANN is trained using the training set, with the validation set evaluated at the end of each epoch. The results presented in this study are from the testing set. During training, we use a learning rate of 1×10^{-4} with stochastic gradient descent for up to 1000 epochs with early stopping when validation loss did not decrease for 100 epochs. To implement regularization, we include a dropout layer between the input layer and first hidden layer in training. We found that a high rate of dropout (80% dropout rate in this experiment) forced the ANN to learn information more slowly and greatly reduced over-fitting on the validation set.

2.3 AMV and IPO indices

We compute the AMV and IPO indices within CESM2 using the deseasoned and detrended SST data. For the AMV index, we calculate the monthly mean SST anomaly over the North Atlantic ocean (0°N to 80°N, 280°E to 360°E) and then standardize by removing the mean and dividing by the standard deviation. Note we do not de-trend by the global mean SST as recommended by Trenberth and Shea (2006) because the control run lacks a forced long term warming trend and model drift was removed during preprocessing. We calculate the IPO index following the tripole index proposed by Henley et al. (2015). We include plots of the spatial AMV and IPO patterns in CESM2 and the method for calculating IPO index in the Supplement.

3 Results

3.1 Evaluating Performance

In this study, 10 networks (identical architecture, only varying the initial network random seed) are trained at each SST grid point in the ocean and we show the results of the best neural network at each grid point. To designate the "best" network, we select the ANN with the lowest mean absolute error (MAE, difference between predicted μ and true y) on the 10% of samples with the lowest σ predictions in the validation set. This designation leverages a fundamental characteristic of a network that has learned predictability in the data: prediction error should decrease as predicted σ decreases. We demonstrate this idea in Figure 1c where we show a network trained to predict SST in 1-5 years in the North Atlantic (52.5°N, 325°E). Along the x-axis, we threshold by increasing confidence with the y-axis showing corresponding MAE for those predictions. For all samples, the MAE is $\sim$0.52 however for the 40% most confident predictions the MAE has dropped to 0.46. For the 10% most confident predictions, the MAE has dropped further to $\sim$0.39 implying the network has learned samples that lead to more predictable SST anomaly. We hence refer to lower σ predictions as more confident predictions, or more predictable inputs. For some grid points, all networks fail to learn anything, meaning they always predict an SST anomaly of zero (or very close to zero). These networks are removed before analysis, resulting in 30% of networks (525/1709) removed for lead years 1-5, and 39% (675/1709) for lead years 3-7.

3.2 Predicting SST

We ensure that the ANNs are learning to skillfully predict SSTs on decadal timescales in CESM2 by examining prediction error in the testing data at each grid point. Fig. 2a is the MAE for ANN predictions for the testing set for lead years 1-5, with black indicating grid points where all 10 networks failed to learn anything. These regions are largely in the Southern Hemisphere subtropics, The lowest MAEs are found in the North Atlantic Ocean and the Southern Ocean around South America. This spatial distribution of prediction skill (including regions where the networks failed) broadly agrees with that found to be attributable to internal variability in the decadal hindcast studies using the CESM1 decadal prediction large ensemble (Yeager et al., 2018; Christensen et al., 2020).These studies use a different model version (CESM1 vs CESM2), and the simulations include the effects of external forcing since 1850. However, the widespread agreement of spatially varying predictability suggests the results in Figure 2 are not a result of experiment design or network architecture but are rather due to differences in predictability between regions.

The prediction skill for lead years 3-7 is shown in Fig 2b and highlights similar regions as being more predictable as in lead years 1-5. Furthermore, there does not seem to be a substantial loss in skill between these two lead times. This, coupled with the spatial distribution of prediction skill, suggests that the ANNs are learning physical relationships to make their predictions.

To contextualize the predictions of the ANNs, we benchmark them against a simple persistence model. The persistence model predicts that the SST anomaly will be unchanged so that the SST anomaly at the time of input remains the same at the time of prediction. We calculate the MAE for the persistence model and subtract it from the MAE of the ANNs (ΔMAE = MAE$_{ANN}$− MAE$_{persistence}$), and plot the results in Figure 2g and 2h. In regions where ΔMAE is negative, the ANN outperforms persistence (i.e. has lower error). These regions are illustrated in warm colors in Figure 2g and 2h and illustrates that the ANNs trained in this study out-perform persistence in all locations and at both lead times. These regions were all found to be significant to $\alpha = 0.05$ using a one-sided Wilcoxon signed-rank test. The greatest improvement in skill above persistence occurs in the cold tongue region of the Equatorial Pacific. This is unsurprising as this region exhibits large interannual variability due to the El Nino Southern Oscillation, and hence persistence performs poorly in this region. Also notable, the improvements over persistence do not necessarily align with grid points where the networks achieve lowest MAE. This is a fingerprint of regional decadal variability, that regions with longer memory (e.g. the mid-latitude North Atlantic) are better modeled by persistence, but in these cases our networks still out-perform persistence.

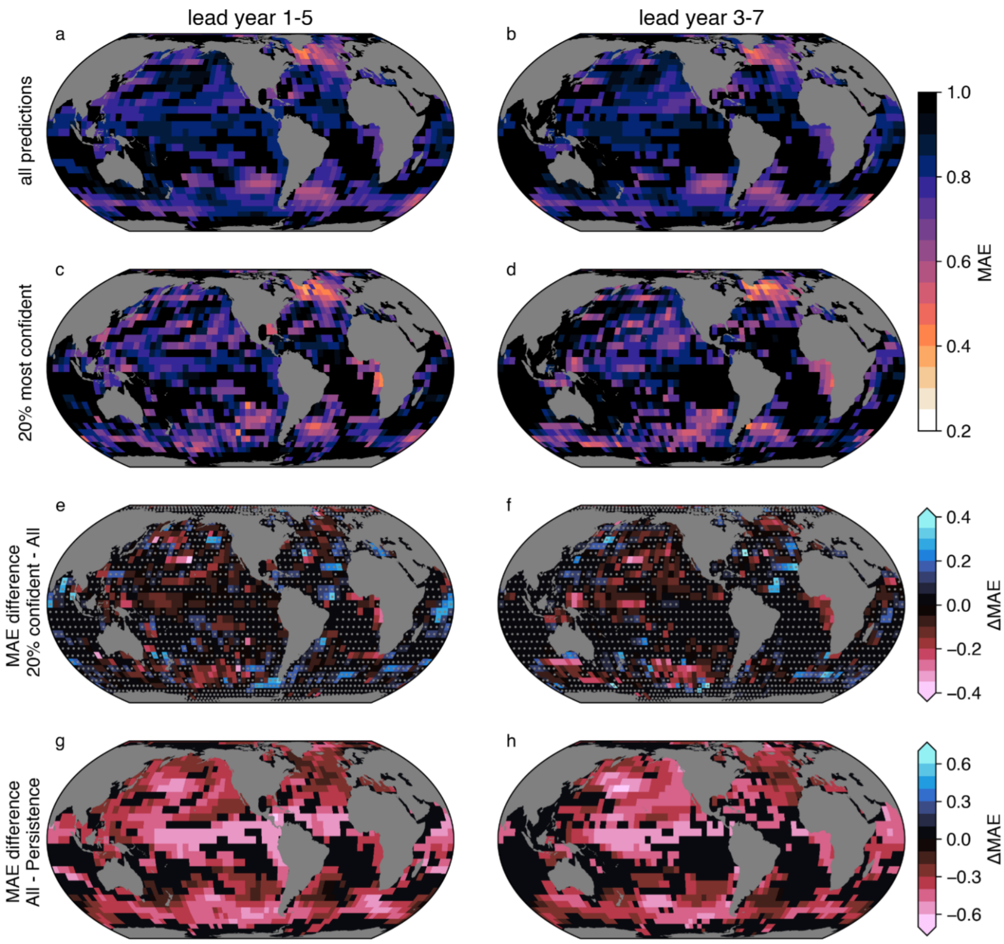

Figure 2. Evaluation of ANN prediction error. The left column is the prediction error for lead years 1-5, and the right column is for lead years 3-7. Panel a and panel b are mean absolute error (MAE) for all predictions in the testing set (i.e. all samples, N=3400). Panel c and panel d show MAE for only the 20% most confident predictions in the testing set as identified using the ANNs's uncertainty (N=680). Panel e and panel f are the differences between the 20% most confident predictions and all predictions (e.g. panel e = panel c – panel a). Stippling indicates areas where the skill improvement is not statistically significant to $\alpha = 0.05$. Panel f and panel g are the difference between MAE_{ANN} and $\mathrm{MAE}_{persistence}$ ($\mathrm{MAE}_{ANN} - \mathrm{MAE}_{persistence}$) in the testing set.

3.3 Identifying State-Dependent Predictability

The predictive power of ANNs for decadal prediction is now demonstrated by using them to identify state-dependent predictability. In Figure 2c and 2d we plot the MAE for only the 20% most confident predictions (20% lowest predicted σ) by the ANN for each SST grid point. That is, ANN objectively identifies more predictable initial states, and we do not directly use knowledge of the ground truth to identify these predictions. To aid in visualization, we also plot the difference in MAE between the 20% confident predictions and all predictions in Figure 2e. When comparing the most confident predictions with all predictions at lead years 1-5 (Figure 2e), MAE is largely reduced for more confident predictions in the mid-latitudes, implying that more confident predictions are associated with smaller prediction errors in these locations. Similarly for lead year 3-7 (Fig. 2f), we see that sorting for the most confident predictions leads to reduced error in most locations. For those regions where error increases, this is likely due to the network learning predictability in the testing and validation data that does not generalize to the testing data which either suggests over-fitting or unaccounted-for model drift. Interestingly, at both lead times, some regions that show very little skill across all predictions exhibit large increases in skill when considering only the most confident predictions (e.g. central Pacific and the Gulf of Guinea), demonstrating that a region may be considered not predictable when in fact it is just not *always* predictable.

3.4 Investigating Skillful Decadal Predictions

By using ANN predictions to identify state dependent predictability, we can also investigate oceanic patterns that lead to predictability. Here we examine the predictions of two ANNs trained to predict SSTs in the North Atlantic and North Pacific oceans to investigate processes that are contributing to enhanced prediction skill in these regions. In the following analysis we single out two particular grid points to investigate SST predictability but the results are largely unchanged for the directly adjacent grid cells. Here, we show results for the testing data but these results are consistent throughout the control run (see supplementary material).

Figure 3 shows the 20% most confident predictions of positive SST anomaly for a point in the North Atlantic Sub-Polar Gyre from the testing set (52.5°N, 325°E). We single out positive predictions because the ANN's confident predictions are preferentially positive (583 positive predictions out of 680 confident testing samples, where 680 is 20% of the testing set), implying that the ANN detects that particular positive predictions lead to lower uncertainty. As predictions are preferentially positive, this is evidence that the ANN is detecting state-dependent predictability in the North Atlantic

We plot the correct and confident positive predictions to ensure we are analyzing the correct signals that contribute to predictability. This leaves 472 samples. Fig 3a – 3c show the composite of OHC input maps for correct and confident positive predictions to investigate the initial states that lead to predictability. At all three OHC levels there is a positive OHC anomaly in the subtropical to mid-latitude Atlantic Ocean. We verify that this signal was likely utilized by the ANN in its predictions by using an ANN explainability technique to investigate the input regions that are important to the network's prediction (see Text S1 and Figure S2). This shows the positive OHC anomaly in the North Atlantic at all three OHC levels was highlighted as contributing to the ANN's decisions. As the positive heat anomaly is slightly south of the predicted grid point, this could indicate northward heat transport to achieve a positive prediction. The composite SST anomaly in Fig 3f shows the positive anomaly is around the predicted grid point in the North Atlantic which implies that this anomaly has moved northward from the initial state (i.e. northward from the positive OHC anomaly in the subtropical North Atlantic in Fig 3a). From this evidence, we posit that the skillful SST prediction is preceded by a positive heat anomaly in North Atlantic ocean, which is transported into the

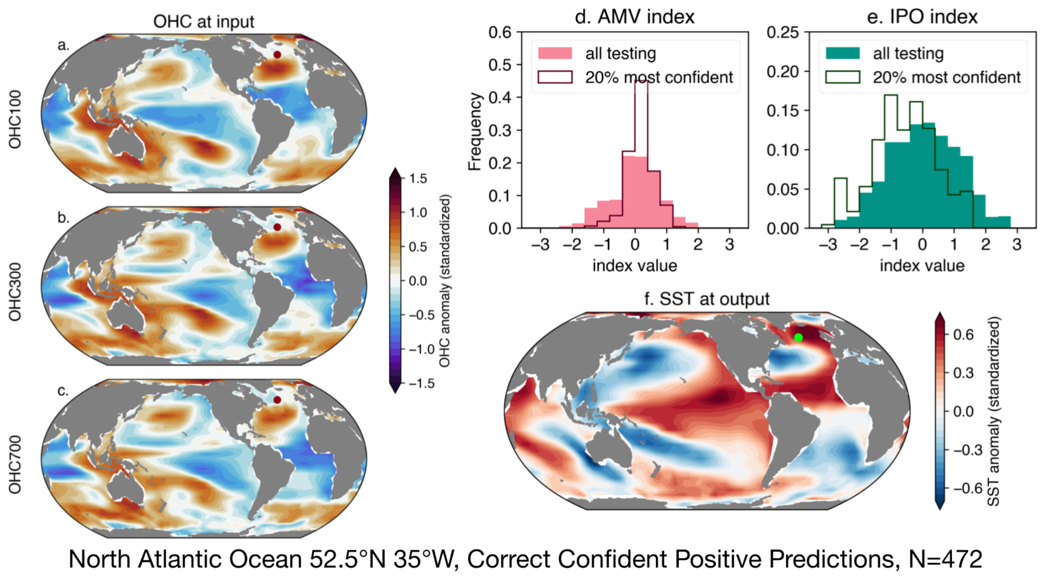

Figure 3. State-dependent predictability identified in the North Atlantic for predicting average SST anomaly at lead time 1-5 years. Panels a-c: Composite of OHC inputs for confident predictions of positive SST anomaly in a point in the North Atlantic (red dot). Panel d: histogram of AMV index for testing data (dark pink) and most confident predictions (light pink). Panel e: as panel d but for IPO index. Panel f: Composite of SST map for confident predictions of SST in the North Atlantic (green dot).

gyre region. This is consistent with Borchert et al. (2018) who identified periods of enhanced heat transport in the mid-latitude as a state of increased predictability of SSTs in the North Atlantic subpolar gyre for up to 8 years.

As an analogue for oceanic variability, we also consider the phases of the AMV and IPO during periods of increased network confidence. In Fig 3d we present the distribution of the AMV index during the entire testing period (pink shading, mean = 0.00) with the solid line showing the distribution for only 20% confident predictions which has a mean of 0.16. From this, it appears that confident predictions are most likely to occur during positive AMV. When randomly drawing 20% of the samples from the AMV distribution in testing, the likelihood of a mean of 0.16 occurring is less than 1%. This implies that more skillful SST predictions in the North Atlantic Sub-Polar Gyre coincide with northward heat transport from the subtropics (from 3a-c and f) coupled with the positive phase of AMV (from 3d). This is consistent with previous results by e.g. Christensen et al. (2020); Borchert et al. (2018). In 3e, we show the distribution of IPO phase for the testing data (green shading, mean = 0.05) and 20% most confident predictions outlined with the solid line, with a mean of -0.58. The likelihood drawing a mean of -0.58 from the IPO testing distribution is less than 1% which suggests that the negative phase of the IPO contributes to the predictability of North Atlantic SSTs. This is also apparent in Fig 3a-c which all show the negative IPO pattern in the Pacific Ocean. This may indicate some inter-basin teleconnection that contributes to the predictability of North Atlantic SSTs.

We now perform a similar analysis for an ANN trained to predict SST in 1-5 years at a point in the North Pacific (42.5°N, 175°E). In Figure 4 we show the results for the 20% most confident negative predictions. For this region, 632 out of the 680 most confident samples were predictions of negative anomaly, implying the ANN designated neg-

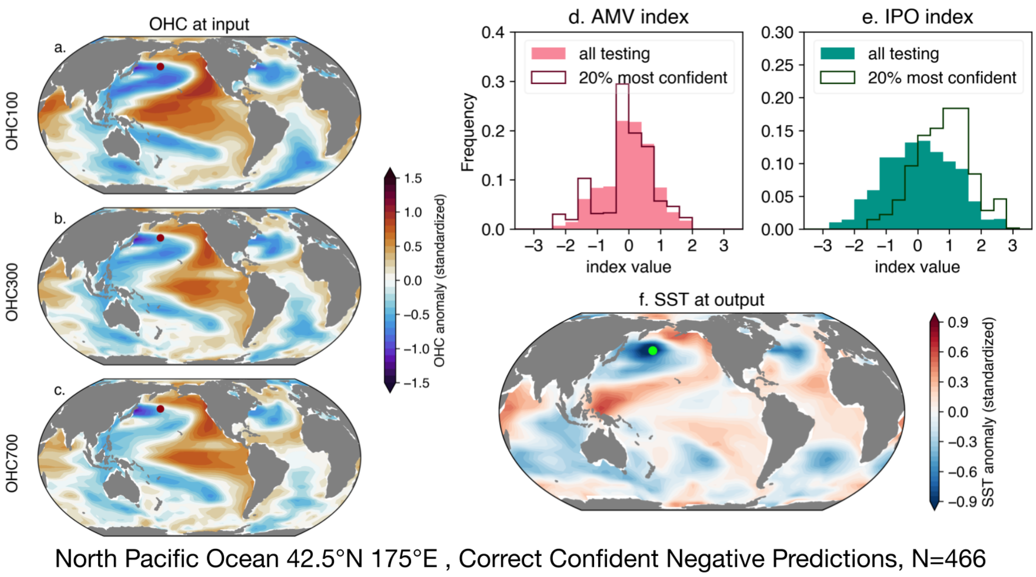

Figure 4. As Figure 3 but for the North Pacific

ative predictions as more confident. Again we plot only the correct predictions, resulting in 466 samples in these composites. Fig 4a-c shows the composite OHC inputs for confident negative predictions, and the major signal appears to be a positive IPO/PDV pattern in all panels. It is likely the ANN utilized this pattern to make these confident negative predictions from the ANN explainability heat-maps (see Text S1 and Figure S3). This is supported by the histogram of the IPO index in Fig 4e which shows the distribution of IPO phase in the confident samples is shifted such that confident samples significantly coincide with the positive phase of the IPO. There is no such strong signal in the AMV index (Fig 4d). Lastly, the confident predictions appear to relate to persistence in the positive IPO phase because the composite map of SST at output (Fig 4f) shows an IPO pattern in the Pacific Ocean. The largest SST anomalies are in the north Pacific mid-latitudes, in the traditional PDV region. From this, we posit that skillful predictions of SST in the North Pacific are associated with persistence in the positive phase of IPO (i.e. negative SST anomaly at the predicted grid point). Here, the ANN preferentially identifies negative SST predictions as skillful, perhaps implying that persistence in the positive phase of IPO is more predictable than persistence of the negative phase. We posit that this difference in predictability is due to the underlying non-linear mechanisms governing IPO dynamics and particularly the asymmetry in the dynamics governing ENSO events (Choi et al., 2013; Okumura & Deser, 2010). Further investigation of this is an avenue for future work.

4 Discussion & Conclusion

We show that artificial neural networks (ANNs) skillfully predict SST evolution on decadal timescales and that they can objectively identify decadal state-dependent predictability due to internal variabiltiy in the North Pacific and North Atlantic Oceans. Specifically, we use a regression neural network where the predictions take the form of a conditional normal distribution which we leverage to isolate predictions that are more likely to have lower error. This approach allows us to investigate possible contributing mechanisms to decadal SST predictability, particularly Atlantic multi-decadal variability and the interdecadal Pacific oscillation (AMV and IPO, Figs 3 and 4). We chose to model the conditional distributions as normal distributions as alternatives did not significantly improve skill. We suggest that future studies investigating state-dependent predictability for other timescales and variables may benefit from the addition of skewness to the predicted conditional distributions (Barnes et al., 2021), as well as further exploring alternative network architectures to tease out additional skill.

We investigate state-dependent predictability in two regions, the North Atlantic Subpolar Gyre, and the North Pacific Ocean by identifying predictions in these regions that the ANNs assigned the lowest uncertainty and investigating the processes that correspond to these confident predictions. This study utilizes the CESM2 long control representation of the climate system and the results in the North Atlantic appear to agree with hindcast studies of Brune et al. (2018); Borchert et al. (2018); Yeager et al. (2018) which use different models to that used here (MPI-ESM; Giorgetta et al. (2013) and CESM1; Hurrell et al. (2013)). These previous studies also incorporate observations or reanalysis to evaluate the prediction skill of the decadal hindcasts. Moreover, in a study of initialized decadal hindcasts in the CMIP6 archive, Borchert et al. (2021) attribute predictable SSTs in the North Atlantic subpolar gyre to the effects of external forcing in the historical era, particularly volcanic forcing. Since our findings are consistent with the state-dependent predictability investigated in these studies, this suggests that the ANN predictions and mechanisms investigated here are likely relevant to realistic climate variability and implies a role for internal variability in North Atlantic predictability. Further investigation is left for future work.

Here we present a data-driven approach to diagnosing state-dependent predictability in an unforced model simulation. In addition to the role of North Atlantic heat transport, we find evidence for a state-dependent inter-basin teleconnection, that is, the negative phase of the IPO influencing predictability of North Atlantic SSTs (Fig 3). The drivers of predictability and variability in the North Atlantic ocean are still debated, especially the relative roles of internal variability and external forcing (Wu et al., 2011; Clement et al., 2015; R. Zhang et al., 2019; Mann et al., 2021; Fang et al., 2021; Fenske & Clement, 2022). We hence suggest that future work on decadal prediction should investigate the roles of internal variability and external forcing through the lens of state-dependent predictability.

This study emphasizes the importance of examining state-dependent predictability for decadal predictions. We stress that the *a priori* identification of more predictable initial states greatly increases prediction skill and can hence aid in estimating the evolution of the internal long-term variability of the climate system.

5 Open Research

We use CESM2 output from the pre-industrial control experiment which is freely available from Earth System Grid https://esgf-node.llnl.gov/projects/cmip6 (Danabasoglu, 2019).

Analysis was carried out in Python 3.7 and 3.9, ANNs were developed using TensorFlow (Abadi et al., 2016), while XAI heatmaps were created with iNNvestigate (Alber et al., 2019). Many color maps in this work are the from CMasher package (van der Velden, 2020) and regridding was achieved using Climate Data Operators (CDO; Schulzweida, 2019).

Code used to preprocess, generate the ANNs, and produce the figures in this work can be found at Gordon (2022).

Acknowledgments
E. M. Gordon is partially funded by Fulbright New Zealand. E. M. Gordon and E. A. Barnes are supported, in part, by NSF CAREER AGS-1749261 under the Climate and

Large-scale Dynamics program. We thank John Fasullo at the National Center for Atmospheric Research (NCAR) for diagnosing the OHC from CESM2. We would like to acknowledge high-performance computing support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Zheng, X. (2016, November). Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (pp. 265–283). Savannah, GA: USENIX Association. Retrieved from `https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi`

Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., … Kindermans, P.-J. (2019). innvestigate neural networks! *Journal of Machine Learning Research*, *20*(93), 1-8. Retrieved from `http://jmlr.org/papers/v20/18-540.html`

Albers, J. R., & Newman, M. (2019, November). A Priori identification of skillful extratropical subseasonal forecasts. *Geophys. Res. Lett.*, *46*(21), 12527–12536. Retrieved from `https://onlinelibrary.wiley.com/doi/10.1029/2019GL085270` doi: 10.1029/2019gl085270

Alexander, M. A., & Deser, C. (1995, January). A Mechanism for the Recurrence of Wintertime Midlatitude SST Anomalies. *J. Phys. Oceanogr.*, *25*(1), 122–137. Retrieved from `https://journals.ametsoc.org/view/journals/phoc/25/1/1520-0485_1995_025_0122_amftro_2_0_co_2.xml?tab_body=fulltext-display` doi: 10.1175/1520-0485(1995)025⟨0122:AMFTRO⟩2.0.CO;2

Barnes, E. A., & Barnes, R. J. (2021, December). Controlled abstention neural networks for identifying skillful predictions for regression problems. *J. Adv. Model. Earth Syst.*, *13*(12). Retrieved from `https://onlinelibrary.wiley.com/doi/10.1029/2021MS002575` doi: 10.1029/2021ms002575

Barnes, E. A., Barnes, R. J., & Gordillo, N. (2021, September). *Adding Uncertainty to Neural Network Regression Tasks in the Geosciences.* Retrieved from `http://arxiv.org/abs/2109.07250`

Boer, G. J. (2011, March). Decadal potential predictability of twenty-first century climate. *Clim. Dyn.*, *36*(5), 1119–1133. Retrieved from `https://doi.org/10.1007/s00382-010-0747-9` doi: 10.1007/s00382-010-0747-9

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., & Bellouin, N. (2012, April). Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, *484*(7393), 228–232. Retrieved from `http://dx.doi.org/10.1038/nature10946` doi: 10.1038/nature10946

Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., & Mignot, J. (2021, February). Improved decadal predictions of north Atlantic subpolar gyre SST in CMIP6. *Geophys. Res. Lett.*, *48*(3). Retrieved from `https://onlinelibrary.wiley.com/doi/10.1029/2020GL091307` doi: 10.1029/2020gl091307

Borchert, L. F., Müller, W. A., & Baehr, J. (2018, September). Atlantic Ocean Heat Transport Influences Interannual-to-Decadal Surface Temperature Predictability in the North Atlantic Region. *J. Clim.*, *31*(17), 6763–6782. Retrieved from `https://journals.ametsoc.org/view/journals/clim/31/17/jcli-d-17-0734.1.xml` doi: 10.1175/JCLI-D-17-0734.1

Branstator, G., & Teng, H. (2010, December). Two Limits of Initial-Value Decadal Predictability in a CGCM. *J. Clim.*, *23*(23), 6292–6311. Retrieved from `https://journals.ametsoc.org/view/journals/clim/23/23/2010jcli3678.1.xml` doi: 10.1175/2010JCLI3678.1

Branstator, G., Teng, H., Meehl, G. A., Kimoto, M., Knight, J. R., Latif, M., & Rosati, A. (2012, March). Systematic Estimates of Initial-Value Decadal Predictability for Six AOGCMs. *J. Clim.*, *25*(6), 1827–1846. Retrieved from https://journals.ametsoc.org/view/journals/clim/25/6/jcli-d-11-00227.1.xml doi: 10.1175/JCLI-D-11-00227.1

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J. (2018, September). Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts. *Clim. Dyn.*, *51*(5), 1947–1970. Retrieved from https://doi.org/10.1007/s00382-017-3991-4 doi: 10.1007/s00382-017-3991-4

Choi, K.-Y., Vecchi, G. A., & Wittenberg, A. T. (2013, December). ENSO Transition, Duration, and Amplitude Asymmetries: Role of the Nonlinear Wind Stress Coupling in a Conceptual Model. *J. Clim.*, *26*(23), 9462–9476. Retrieved from https://journals.ametsoc.org/view/journals/clim/26/23/jcli-d-13-00045.1.xml doi: 10.1175/JCLI-D-13-00045.1

Christensen, H. M., Berner, J., & Yeager, S. (2020, September). The Value of Initialization on Decadal Timescales: State-Dependent Predictability in the CESM Decadal Prediction Large Ensemble. *J. Clim.*, *33*(17), 7353–7370. Retrieved from https://journals.ametsoc.org/jcli/article/33/17/7353/348619/The-Value-of-Initialization-on-Decadal-Timescales doi: 10.1175/JCLI-D-19-0571.1

Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Rädel, G., & Stevens, B. (2015, October). The Atlantic Multidecadal Oscillation without a role for ocean circulation. *Science*, *350*(6258), 320–324. Retrieved from https://science.sciencemag.org/content/350/6258/320 doi: 10.1126/science.aab3980

Danabasoglu, G. (2019). *Ncar cesm2 model output prepared for cmip6 cmip historical.* Earth System Grid Federation. Retrieved from https://doi.org/10.22033/ESGF/CMIP6.7627 doi: 10.22033/ESGF/CMIP6.7627

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001916 doi: 10.1029/2019MS001916

Enfield, D. B., Mestas-Nuñez, A. M., & Trimble, P. J. (2001). The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.*, *28*(10), 2077–2080. Retrieved from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL012745 doi: 10.1029/2000GL012745

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved from https://gmd.copernicus.org/articles/9/1937/2016/ doi: 10.5194/gmd-9-1937-2016

Fang, S.-W., Khodri, M., Timmreck, C., Zanchettin, D., & Jungclaus, J. (2021, December). Disentangling internal and external contributions to Atlantic multidecadal variability over the past millennium. *Geophys. Res. Lett.*, *48*(23). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2021GL095990 doi: 10.1029/2021gl095990

Fenske, T., & Clement, A. (2022, February). No internal connections detected between low frequency climate modes in north Atlantic and north Pacific basins. *Geophys. Res. Lett.*. Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2022GL097957 doi: 10.1029/2022gl097957

Foster, D., Gagne, D. J., II, & Whitt, D. B. (2021, December). Probabilistic machine learning estimation of ocean mixed layer depth from dense satellite and sparse in situ observations. *J. Adv. Model. Earth Syst.*, *13*(12). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2021MS002474 doi: 10.1029/2021ms002474

Frankignoul, C., & Hasselmann, K. (1977). Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability. *Tell'Us*, *29*(4), 289–305. Retrieved from http://onlinelibrary.wiley.com/doi/abs/10.1111/j.2153-3490.1977.tb00740.x doi: 10.1111/j.2153-3490.1977.tb00740.x

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., … Stevens, B. (2013, July). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, *5*(3), 572–597. Retrieved from http://doi.wiley.com/10.1002/jame.20038 doi: 10.1002/jame.20038

Gordon, E. M. (2022, June). *emily-gordy/Decadal-SST-prediction: Decadal SST prediction, revised GRL submission.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.6646950 doi: 10.5281/zenodo.6646950

Gordon, E. M., Barnes, E. A., & Hurrell, J. W. (2021, November). Oceanic harbingers of Pacific decadal oscillation predictability in CESM2 detected by neural networks. *Geophys. Res. Lett.*, *48*(21). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2021GL095392 doi: 10.1029/2021gl095392

Guillaumin, A. P., & Zanna, L. (2021, September). Stochastic-deep learning parameterization of ocean momentum forcing. *J. Adv. Model. Earth Syst.*, *13*(9). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2021MS002534 doi: 10.1029/2021ms002534

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K. (2015, December). A Tripole Index for the Interdecadal Pacific Oscillation. *Clim. Dyn.*, *45*(11), 3077–3090. Retrieved from https://doi.org/10.1007/s00382-015-2525-1 doi: 10.1007/s00382-015-2525-1

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., … Marshall, S. (2013, September). The Community Earth System Model: A Framework for Collaborative Research. *Bull. Am. Meteorol. Soc.*, *94*(9), 1339–1360. Retrieved from https://journals.ametsoc.org/view/journals/bams/94/9/bams-d-12-00121.1.xml doi: 10.1175/BAMS-D-12-00121.1

Labe, Z. M., & Barnes, E. A. (2022, February). *Predicting slowdowns in decadal climate warming trends with explainable neural networks.* Retrieved from http://www.essoar.org/doi/10.1002/essoar.10508874.2 doi: 10.1002/essoar.10508874.2

Mann, M. E., Steinman, B. A., Brouillette, D. J., & Miller, S. K. (2021, March). Multidecadal climate oscillations during the past millennium driven by volcanic forcing. *Science*, *371*(6533), 1014–1019. Retrieved from http://dx.doi.org/10.1126/science.abc5810 doi: 10.1126/science.abc5810

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997, June). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bull. Am. Meteorol. Soc.*, *78*(6), 1069–1080. Retrieved from http://journals.ametsoc.org/bams/article/78/6/1069/55942/A-Pacific-Interdecadal-Climate-Oscillation-with doi: 10.1175/1520-0477(1997)078⟨1069:APICOW⟩2.0.CO;2

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., … Albers, J. (2020, May). Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond. *Bull. Am. Meteorol. Soc.*, *101*(5), E608–E625. Retrieved from http://journals.ametsoc.org/bams/article/101/5/E608/345558/Windows-of-Opportunity-for-Skillful-Forecasts doi: 10.1175/BAMS-D-18-0326.1

Mayer, K. J., & Barnes, E. A. (2021). Subseasonal Forecasts of Opportunity

Identified by an Explainable Neural Network. *Geophys. Res. Lett.*, *48*(10), e2020GL092092. Retrieved from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL092092 doi: 10.1029/2020GL092092

Meehl, G. A., Hu, A., Arblaster, J. M., Fasullo, J., & Trenberth, K. E. (2013, September). Externally Forced and Internally Generated Decadal Climate Variability Associated with the Interdecadal Pacific Oscillation. *J. Clim.*, *26*(18), 7298–7310. Retrieved from https://journals.ametsoc.org/view/journals/clim/26/18/jcli-d-12-00548.1.xml doi: 10.1175/JCLI-D-12-00548.1

Meehl, G. A., Hu, A., & Teng, H. (2016, June). Initialized decadal prediction for transition to positive phase of the Interdecadal Pacific Oscillation. *Nat. Commun.*, *7*(1), 11718. Retrieved from http://www.nature.com/articles/ncomms11718 doi: 10.1038/ncomms11718

Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., ... Xie, S.-P. (2021, April). Initialized Earth System prediction from subseasonal to decadal timescales. *Nature Reviews Earth & Environment*, *2*(5), 340–357. Retrieved from https://www.nature.com/articles/s43017-021-00155-x doi: 10.1038/s43017-021-00155-x

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., ... Yeager, S. (2020, June). Current and Emerging Developments in Subseasonal to Decadal Prediction. *Bull. Am. Meteorol. Soc.*, *101*(6), E869–E896. Retrieved from https://journals.ametsoc.org/view/journals/bams/101/6/bamsD190037.xml doi: 10.1175/BAMS-D-19-0037.1

Msadek, R., Dixon, K. W., Delworth, T. L., & Hurlin, W. (2010, October). Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys. Res. Lett.*, *37*(19). Retrieved from http://doi.wiley.com/10.1029/2010GL044517 doi: 10.1029/2010gl044517

Okumura, Y. M., & Deser, C. (2010, November). Asymmetry in the Duration of El Niño and La Niña. *J. Clim.*, *23*(21), 5826–5843. Retrieved from https://journals.ametsoc.org/view/journals/clim/23/21/2010jcli3592.1.xml doi: 10.1175/2010JCLI3592.1

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., & Tschuck, P. (2004, November). Estimating the Decadal Predictability of a Coupled AOGCM. *J. Clim.*, *17*(22), 4463–4472. Retrieved from https://journals.ametsoc.org/view/journals/clim/17/22/3209.1.xml doi: 10.1175/3209.1

Power, S., Casey, T., Folland, C., Colman, A., & Mehta, V. (1999, May). Interdecadal modulation of the impact of ENSO on Australia. *Clim. Dyn.*, *15*(5), 319–324. Retrieved from https://doi.org/10.1007/s003820050284 doi: 10.1007/s003820050284

Schulzweida, U. (2019, October). *Cdo user guide.* Retrieved from https://doi.org/10.5281/zenodo.3539275 doi: 10.5281/zenodo.3539275

Simpson, I. R., Deser, C., McKinnon, K. A., & Barnes, E. A. (2018, October). Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperatures. *J. Clim.*, *31*(20), 8313–8338. Retrieved from https://journals.ametsoc.org/view/journals/clim/31/20/jcli-d-18-0168.1.xml doi: 10.1175/JCLI-D-18-0168.1

Simpson, I. R., Yeager, S. G., McKinnon, K. A., & Deser, C. (2019, August). Decadal predictability of late winter precipitation in western Europe through an ocean–jet stream connection. *Nat. Geosci.*, *12*(8), 613–619. Retrieved from https://www.nature.com/articles/s41561-019-0391-x doi: 10.1038/s41561-019-0391-x

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002002. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/

10.1029/2019MS002002 doi: 10.1029/2019MS002002

Toms, B. A., Barnes, E. A., & Hurrell, J. W. (2021, June). Assessing decadal predictability in an earth–system model using explainable neural networks. *Geophys. Res. Lett.*, *48*(12). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2021GL093842 doi: 10.1029/2021gl093842

Trenberth, K. E., & Fasullo, J. T. (2013, December). An apparent hiatus in global warming? *Earths Future*, *1*(1), 19–32. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/2013EF000165 doi: 10.1002/2013ef000165

Trenberth, K. E., & Shea, D. J. (2006). Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.*, *33*(12). Retrieved from http://doi.wiley.com/10.1029/2006GL026894 doi: 10.1029/2006gl026894

van der Velden, E. (2020, February). CMasher: Scientific colormaps for making accessible, informative and 'cmashing' plots. *The Journal of Open Source Software*, *5*(46), 2004. doi: 10.21105/joss.02004

Wu, S., Liu, Z., Zhang, R., & Delworth, T. L. (2011, February). On the observed relationship between the Pacific Decadal Oscillation and the Atlantic Multidecadal Oscillation. *J. Oceanogr.*, *67*(1), 27–35. Retrieved from https://doi.org/10.1007/s10872-011-0003-x doi: 10.1007/s10872-011-0003-x

Xie, S.-P., & Tanimoto, Y. (1998). A pan-Atlantic decadal climate oscillation. *Geophys. Res. Lett.*, *25*(12), 2185–2188. Retrieved from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98GL01525 doi: 10.1029/98GL01525

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., ... Lovenduski, N. S. (2018, September). Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model. *Bull. Am. Meteorol. Soc.*, *99*(9), 1867–1886. Retrieved from https://journals.ametsoc.org/bams/article/99/9/1867/70398/Predicting-Near-Term-Changes-in-the-Earth-System-A doi: 10.1175/BAMS-D-17-0098.1

Zhang, R., & Delworth, T. L. (2006). Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, *33*(17). Retrieved from http://doi.wiley.com/10.1029/2006GL026267 doi: 10.1029/2006gl026267

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., ... Little, C. M. (2019, June). A review of the role of the Atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Rev. Geophys.*, *57*(2), 316–375. Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2019RG000644 doi: 10.1029/2019rg000644

Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997, May). ENSO-like Interdecadal Variability: 1900–93. *J. Clim.*, *10*(5), 1004–1020. Retrieved from https://journals.ametsoc.org/view/journals/clim/10/5/1520-0442_1997_010_1004_eliv_2.0.co_2.xml doi: 10.1175/1520-0442(1997)010⟨1004:ELIV⟩2.0.CO;2