Predicting slowdowns in decadal climate warming trends with explainable neural networks

Zachary M. Labe[1] and Elizabeth A. Barnes[1]

[1]Colorado State University

November 30, 2022

Abstract

The global mean surface temperature (GMST) record exhibits both interannual to multidecadal variability and long-term warming due to external climate forcing. To explore the predictability of temporary slowdowns in decadal warming, we apply an artificial neural network (ANN) to climate model data from the Community Earth System Model Version 2 Large Ensemble Project. Here, an ANN is tasked with whether or not there will be a slowdown in the rate of the GMST trend by using maps of ocean heat content at the onset. Through a machine learning explainability method, we find the ANN is learning off-equatorial patterns of anomalous ocean heat content that resemble transitions in the phase of the Interdecadal Pacific Oscillation in order to make slowdown predictions. Finally, we test our ANN on observed historical data, which further reveals how explainable neural networks are useful tools for understanding decadal variability in both climate models and observations.

Predicting slowdowns in decadal climate warming trends with explainable neural networks

Zachary M. Labe[1] and Elizabeth A. Barnes[1]

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

Key Points:

- An artificial neural network predicts the onset of slowdowns in decadal warming trends of global mean surface temperature
- Explainable AI reveals the neural network is leveraging tropical patterns of ocean heat content anomalies to make its predictions
- Transitions in the phase of the Interdecadal Pacific Oscillation are frequently associated with warming slowdown predictions in CESM2-LE

Corresponding author: Zachary M. Labe, `zmlabe@rams.colostate.edu`

Abstract

The global mean surface temperature (GMST) record exhibits both interannual to multidecadal variability and long-term warming due to external climate forcing. To explore the predictability of temporary slowdowns in decadal warming, we apply an artificial neural network (ANN) to climate model data from the Community Earth System Model Version 2 Large Ensemble Project. Here, an ANN is tasked with whether or not there will be a slowdown in the rate of the GMST trend by using maps of ocean heat content at the onset. Through a machine learning explainability method, we find the ANN is learning off-equatorial patterns of anomalous ocean heat content that resemble transitions in the phase of the Interdecadal Pacific Oscillation in order to make slowdown predictions. Finally, we test our ANN on observed historical data, which further reveals how explainable neural networks are useful tools for understanding decadal variability in both climate models and observations.

Plain Language Summary

Long-term observations reveal that Earth's average temperature is rising due to human-caused climate change. Along with this warming trend are also variations from year-to-year and even over multiple decades. This temperature variability is often tied to regional patterns of heat in the deep ocean, which can then modulate weather and climate extremes over land. In an attempt to better predict temperature variability on decadal timescales, we use a machine learning method called artificial neural networks and data from a climate model experiment, which was designed to compare climate change and variability. Here, our artificial neural network uses maps of ocean heat to predict the onset of temporary slowdowns in the rate of global warming in both the climate model and in real-world observations. We then use a visualization technique to find which areas of ocean heat that the artificial neural network is using to make its correct predictions, which are found to be mainly across the Pacific Ocean. In agreement with recent research, our study finds that new data science methods, like machine learning, can be useful tools for predicting variations in global climate.

1 Introduction

One of the most recognizable indicators of anthropogenic climate change is the global mean surface temperature (GMST) (Hansen et al., 2010; Johnson et al., 2020). It also exhibits interannual to multidecadal variability with periods of accelerations and slowdowns in the rate of decadal trends (Trenberth et al., 2002; Thompson et al., 2009; Dai et al., 2015; Maher et al., 2020). A notable example of one of these GMST slowdowns occurred in the early 2000s (Flato et al., 2013; Fyfe et al., 2013). This temporary warming slowdown ended in the mid-2010s (Mann et al., 2017; Zhang et al., 2019), and more recently, 2020 was one of the three warmest years in the observational record (Dunn et al., 2021). Although this period was commonly described as a 'hiatus' or 'pause' in global warming within scientific studies and popular media (Boykoff, 2014; Lewandowsky et al., 2016), we will refer to it here as a 'slowdown in decadal warming' (Fyfe et al., 2016), which is more consistent with our understanding of internal variability in the climate system.

Numerous mechanisms have been proposed to explain the cause of the early 2000s slowdown, as reviewed in Medhaug et al. (2017) and Xie and Kosaka (2017), but it was likely a combination of factors ranging from uncertainties in the observational data record (e.g., Cowtan & Way, 2014; Karl et al., 2015), fluctuations in radiative forcing (Schmidt et al., 2014), cooling in the eastern Pacific associated with a negative phase of the Interdecadal Pacific Oscillation (IPO) (Meehl et al., 2013; England et al., 2014; Roberts et al., 2015), anthropogenic aerosol and volcanic forcing (Santer et al., 2014; Smith et al., 2016), changes in deep ocean heat uptake (Watanabe et al., 2013), and interactions between other modes of climate variability (W. Liu & Xie, 2018). Motivated by the increasing body of literature on the causes and impacts of the early 2000s slowdown, we aim to investigate the predictability of similar GMST trends occurring in a warming climate. While decadal predictability has been explored using other statistical methods (e.g., Mann et al., 2016; Sévellec & Drijfhout, 2018), sensitivity experiments (e.g. Kosaka & Xie, 2013), and hindcasts with initialized state climate modeling frameworks (e.g., Fyfe et al., 2011; Guemas et al., 2013; Meehl et al., 2014; Meehl & Teng, 2014; Boer et al., 2016), we explore this problem through the lens of machine learning.

Deep learning methods, such as neural networks, have the ability to extract and leverage nonlinear patterns across data-intensive spatial fields, which make them promising tools for revealing new insights and sources of predictability in climate science (Reichstein et al., 2019; Barnes, Mayer, et al., 2020; Irrgang et al., 2021; Sonnewald et al., 2021). Recent work has demonstrated the utility for neural networks in identifying climate modes, teleconnections, and forecasts of opportunity for a wide variety of timescales (e.g., Wu & Hsieh, 2004; Ham et al., 2019; Toms et al., 2021; Gibson et al., 2021; Gordon et al., 2021; J. Liu et al., 2021; Mayer & Barnes, 2021; Nadiga, 2021; Tang & Duan, 2021). Further, a growing number of explainable artificial intelligence (XAI) methods have been adapted for applications in weather and climate science (McGovern et al., 2019; Toms et al., 2020), which can retrospectively trace the decisions of neural networks and compare the attribution of input features to known physical mechanisms in the Earth system. Besides evaluating trust and credibility to the machine learning prediction, XAI methods can also be used for physics-guided scientific discovery and hypothesis testing (Ebert-Uphoff & Hilburn, 2020; Toms et al., 2020; Sonnewald & Lguensat, 2021).

In this study, we use an artificial neural network (ANN) to explore the predictability of decadal warming slowdowns within a new large ensemble experiment and real-world observations. In addition to their predictability, we also use a complimentary XAI method to investigate the oceanic patterns that may provide insight to these temporary warming slowdowns.

2 Data and Methods

2.1 Climate Model Large Ensemble

For climate model data, we use a large ensemble experiment conducted by the Community Earth System Model Version 2 (CESM2; Danabasoglu et al., 2020) (see Supporting Information for more details). Specifically, we use simulations from the CESM2 Large Ensemble Community Project (CESM2-LE; Rodgers et al., 2021), which includes 100 ensemble members branched from the fully-coupled CESM2 preindustrial control (1850 radiative forcing conditions) using different atmospheric and oceanic initial states. CESM2-LE members follow historical Coupled Model Intercomparison Project Phase 6 (CMIP6) forcing from 1850 to 2014 and thereafter follow the SSP3-7.0 future radiative forcing (high emissions scenario) until 2100 (Eyring et al., 2016; O'Neill et al., 2016). We consider the first 50 ensemble members (1-50), which are prescribed with biomass burning emissions following CMIP6 protocol (Van Marle et al., 2017). In contrast, the second set of 50 ensemble members follow temporally smoothed biomass burning fluxes (51-100). As discussed in Rodgers et al. (2021), this difference in biomass burning forcing has been shown to affect large-scale climate features, including the GMST record in present day.

Due to limited data availability at the time of our analysis, we analyze only 40 ensemble members within the first subset of CESM2-LE (1-50). From these 40 members, we use monthly outputs of near-surface air temperature (T2M) and sea surface temperature (SST). We also utilize monthly ocean heat content (OHC), which is derived as the vertical heat content integral between three distinct depth layers (0–100 m, OHC100; 0–300 m, OHC300; 0–700 m, OHC700); although we focus on maps of OHC100 for the actual training of our ANN. We then apply a bilinear interpolation to all variables so that they share a common (slightly coarser) latitude by longitude grid (1.9° x 2.5°). We calculate annual means from the monthly data and use the period from 1990 to 2099 to classify slowdowns in decadal warming. To focus on warming slowdowns driven by internal variability, we remove the 40-member ensemble mean from each individual ensemble in every year and grid box for SST and OHC (Phillips et al., 2020; Maher et al., 2021).

2.2 Observations

To evaluate our ANN trained on CESM2-LE for predicting the early 2000s warming slowdown in the historical record, we use SST and T2M from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis (Hersbach et al., 2020) and OHC from the Institute of Atmospheric Physics (IAP) ocean gridded product (Cheng & Zhu, 2016; Cheng et al., 2017) (both data sets referred to here as "observations"). SSTs from ERA5 are an interpolated product between HadISST2 (Titchner & Rayner, 2014) from January 1979 to August 2007 and OSTIA (Donlon et al., 2012) from September 2007 to present. Overall, both regional and global mean time series of SST and T2M are consistent with other observational data sets (Hersbach et al., 2020; Bell et al., 2021). Gridded upper OHC from IAP also compares well with in situ measurements and is based on temperature data from the World Ocean Database (WOD; Boyer et al., 2013), which is then further bias-corrected, interpolated, and quality controlled (Li-Jing et al., 2015; Cheng et al., 2017).

In all observations, we use monthly output and bilinearly interpolate these fields onto the same 1.9° x 2.5° grid as CESM2-LE before calculating annual means. We linearly detrend each grid point for SST and OHC predictors to remove long-term warming signals and thus focus on patterns of interannual variability for slowdown predictions.

2.3 Defining Slowdowns in Decadal Warming

Figure 1 shows an example of how we define warming slowdown events in CESM2-LE and observations. While there have been numerous definitions and data sets used for identifying warming slowdown (or so-called hiatus/pause) events (e.g., Risbey et al., 2018), they are generally classified as a near-zero or negative 10-year linear trend of the GMST (Meehl et al., 2011). However, as a result of improvements to station-based observational data (e.g., Morice et al., 2021), such as in the Arctic, a reassessment of the early 2000s slowdown actually shows a positive slope (albeit close to 0°C/yr) in the latest generation of temperature data sets (Medhaug et al., 2017). This includes ERA5 reanalysis (Figure 1b). Recent studies also show that the frequency of slowdown events in CMIP5 models decreases substantially by the end of the 21st century using a negative 10-year linear trend definition (e.g., Maher et al., 2014; Li & Baker, 2016; Sévellec et al., 2016). Yet, internal variability is still projected to affect regional and global climate trends even under higher future emission scenarios (Easterling & Wehner, 2009; Li & Baker, 2016; Cassou et al., 2018; Maher et al., 2020). Therefore, we take a slowdown threshold which considers the effect of internal variability relative to the climate change signal.

First, to classify slowdown events in observations, we compute the area-weighted GMST and calculate 10-year moving linear trends beginning in 1990. We start our analysis in 1990 to avoid any multidecadal slowdown events earlier in the 20th century when the influence of the forced climate change signal may not have fully emerged (Delworth & Knutson, 2000; Papalexiou et al., 2020; Hawkins et al., 2020). We then calculate the mean slope of all decadal trend periods between 1990 and 2020 and take one standard deviation below the mean as our threshold (equating to about +0.01°C/yr, or 0.44 of the mean trends) for slowdown events in observations (black dashed line in Figure 1b). We identify four consecutive slowdown events in observations, which begin in 2002. These years are consistent with previous studies (Lewandowsky et al., 2018).

For CESM2-LE, we first compute the area-weighted GMST for the ensemble mean from all 40 members through 2099 (Figure 1a). We then calculate 10-year moving linear trends, which begin in 1990 for consistency with observations. Next we define our climate model threshold, which is a time series that is computed by multiplying the fractional slope from the observations times each of the decadal trends in the ensemble mean (forced signal) (red dashed line in Figure 1b).

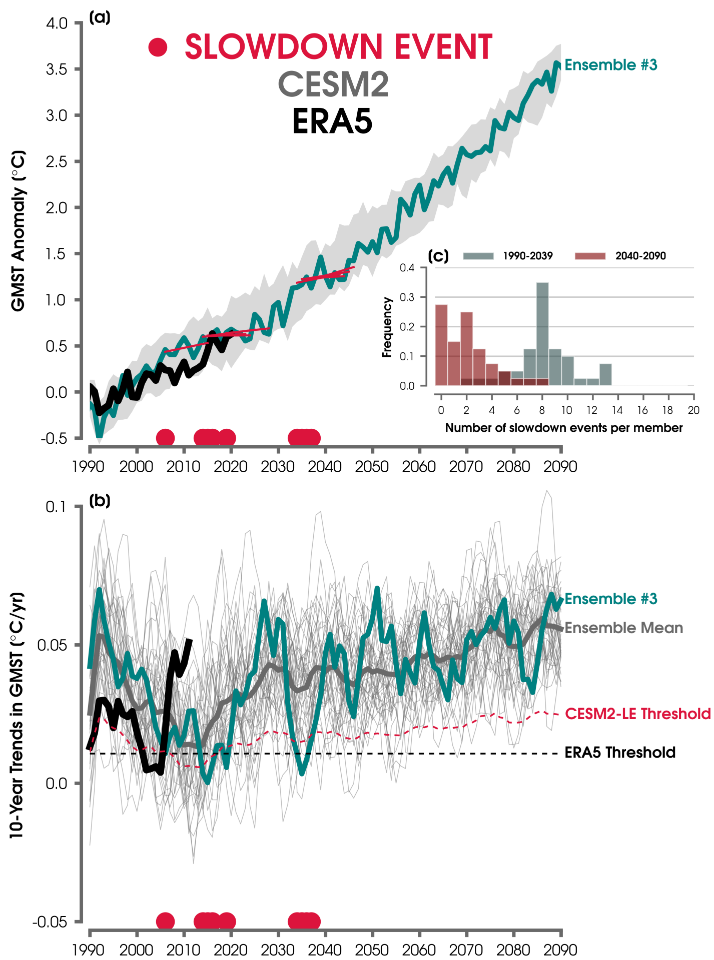

Figure 1. (a) Time series showing annual-mean GMST anomalies for one example (ensemble member) in CESM2-LE relative to a 1981-2010 baseline (blue line). The ensemble spread in annual-mean GMST anomalies is also shown in gray shading for CESM2-LE. Annual-mean GMST anomalies from ERA5 reanalysis are indicated with a black line relative to a 1981-2010 baseline. Onset of slowdown events in the example ensemble are highlighted with red dots and their associated linear trends (red lines) over each 10-year period. (b) The slope of all 10-year moving linear trends are shown for the example ensemble member compared to the other ensembles (light gray lines) and the ensemble mean (dark gray line). As in (a), red dots are shown for the onset of slowdown events in the highlighted ensemble member. Slopes of all 10-year moving linear trends are shown for ERA5 reanalysis by the black solid line. The threshold for slowdown events in CESM2-LE is shown with a red dashed line, and the threshold for slowdown events in ERA5 is shown with a black dashed line. (c) Histogram showing the frequency of slowdown events in each ensemble member over the 1990-2039 period (gray bars) and the 2040-2090 period (red bars). See Section 2.3 for more details.

Separately, we calculate the GMST for each ensemble member and then their 10-year moving linear trends (Figure 1b). We define a warming slowdown event by comparing each ensemble member to check if their 10-year trend is below the climate model threshold. By defining a threshold as a fraction of the mean trend, we take into account the greater influence of the forced signal later in the 21st century. However, this projected warming still reduces the number of slowdown events after 2040 in CESM2-LE (Figure 1c).

2.4 Artificial Neural Network

For this analysis, we adopt a neural network architecture that is designed to receive input maps of OHC100 and output whether the next 10 years will observe a decadal warming slowdown. A schematic of our ANN can be found in Figure S1, and the architecture parameters are outlined in the Supporting Information.

In addition to seeing if warming slowdown events are predictable, we are also interested in the sources of predictability in fields of anomalous OHC100. To attempt to understand the ANN's decision-making process, we use a method of XAI called layer-wise relevance propagation (LRP; Bach et al., 2015; Montavon et al., 2017, 2018). The utility of LRP has been demonstrated in a wide range of weather and climate applications (e.g., Barnes, Toms, et al., 2020; Davenport & Diffenbaugh, 2021; Gordon et al., 2021; Labe & Barnes, 2021; Sonnewald & Lguensat, 2021), and an overview for the geosciences can be found in Toms et al. (2020). In short, prior to the softmax, a single prediction output is propagated backward through the ANN after freezing the model weights and biases. LRP then returns a vectorized spatial map, which shows the feature relevance for every input sample's latitude and longitude pixel. Therefore, by design, we have a unique LRP heatmap for every input sample of OHC100. Throughout this study, regions of higher relevance can be interpreted as more important for the ANN's prediction. We implement the $\mathrm{LRP_z}$ rule for back propagation, which was found by Mamalakis et al. (2021) to be a well performing XAI method using a benchmark climate data set similar to ours. To improve interpretation and reduce the amount of noise in the LRP heatmaps, we only focus on positive areas of relevance, which are features that contribute positively to the ANN's prediction output.

3 Results

3.1 Predicting Slowdown Trends in a Large Ensemble

Figure 2a shows the results of our ANN for each CESM2-LE ensemble member in the testing data set from 1990 through 2090 (i.e., 2090-2099 is the last complete decade of data). Given the large class imbalance, we focus on the F1 score (balancing precision and recall), rather than categorical accuracy, to evaluate the performance of our ANN for correctly identifying slowdown events. Figure S5 provides a collection of skill metrics for our testing data. Overall, the network achieves a F1 score of 40% and performs better than random chance (10.4%). While our ANN sometimes struggles with correctly classifying slowdown events, especially those that occur simultaneously in a row, it generally classifies at least one 10 year period during these extended events. This skill suggests that the ANN is learning information from OHC100 that corresponds to slowdown periods in CESM2-LE.

We test the robustness of our results by training 100 ANNs with unique random initialization seeds and different combinations of ensemble members used for training, validation, and testing data. The F1 score of our single seed ANN falls within the $\approx$ 85th percentile of this distribution, and additional metric scores are shown in Figure S6 for the 100 ANNs. Since there are differences in the frequency of slowdown events in each ensemble member, we also checked if there is a relationship between the accuracy of testing predications compared to the number of training slowdown events each ANN learned for the 100 iterations. However, this does not have a significant effect on the skill of our ANN (Figure S7).

3.2 Sources of Predictability for Slowdowns

To understand the sources of skill for the ANN's correct slowdown predictions in CESM2-LE, we turn to composite maps of LRP. Recall that LRP traces the decision-making process of a neural network, where higher relevance corresponds to greater importance for the ANN to make its final prediction. While we have LRP heatmaps for every input of annual-mean OHC100, we focus on correct predictions by the ANN in the testing data set. Figure 2b shows the LRP composite for all correct slowdown predictions. We find higher relevance in the off-equatorial regions of the eastern Pacific, especially in the regions of the North/South Pacific Meridional Modes (Amaya, 2019). There are also patches of higher relevance across portions of the Indian Ocean, south Atlantic, and south Pacific, which suggests that the ANN is leveraging other regional patterns of OHC to make predictions. Notably, there is no relevance for a thin band along the equator in the area of the El Niño-Southern Oscillation (ENSO). Figure 2d shows the corresponding LRP composite for no slowdowns in decadal warming, which is nearly a mirror image of Figure 2b. This may be a product of the setup of our binary classification problem, and therefore the LRP maps reveal the regions that the ANN is using to make this determination (i.e., yes or no slowdown).

We compare these LRP maps to composites of the raw (normalized) OHC data that were input to the network for correct slowdown predictions (Figure 2c) versus correct no slowdown predictions (Figure 2e). Now we find striking differences between the two OHC patterns. The composite of OHC100 for the slowdown predictions reveal an IPO-like spatial pattern with cold pools in the west-central North Pacific and west-central South Pacific and warm anomalies in the Southern Ocean and eastern Pacific. We also see the signature of a positive Indian Ocean Dipole (IOD; Saji et al., 1999) and a dipole pattern of OHC100 anomalies between the southern Atlantic and north-central Atlantic. Some studies have shown that a positive IOD can be a precursor for a rapid transition to a cooler equatorial Pacific by modulating the strength of the Walker Circulation (Izumo et al., 2010; Le et al., 2020; Yoo et al., 2020).

Figure S8 shows maps of OHC at other vertical depth levels for the slowdown predictions compared to 5-10 years after the start of the slowdown decade. We find a similar spatial pattern of SSTs (Figure S8a), but a stronger cold pool at deeper depths, which appears to be propagating eastward in the equatorial Pacific (Figure S8c-d). In contrast, we find a negative IPO-like pattern for the composites at the end of the slowdown decade (Figure S8e-h). This finding is in agreement with earlier studies that showed slowdowns in decadal warming often corresponded to trends toward a negative phase of the IPO within CMIP5 models (e.g., Maher et al., 2014). Given this evolution of events and the patterns of LRP relevance, it is feasible that the ANN is learning OHC anomalies associated with transitions in the state of the IPO.

To directly assess the IPO in the maps of OHC100, we compute the unfiltered IPO Tripole Index (normalized) following Henley et al. (2015) using annual-mean SSTs from CESM2-LE (Figure S9). For this study, we are more interested in the interannual vari-

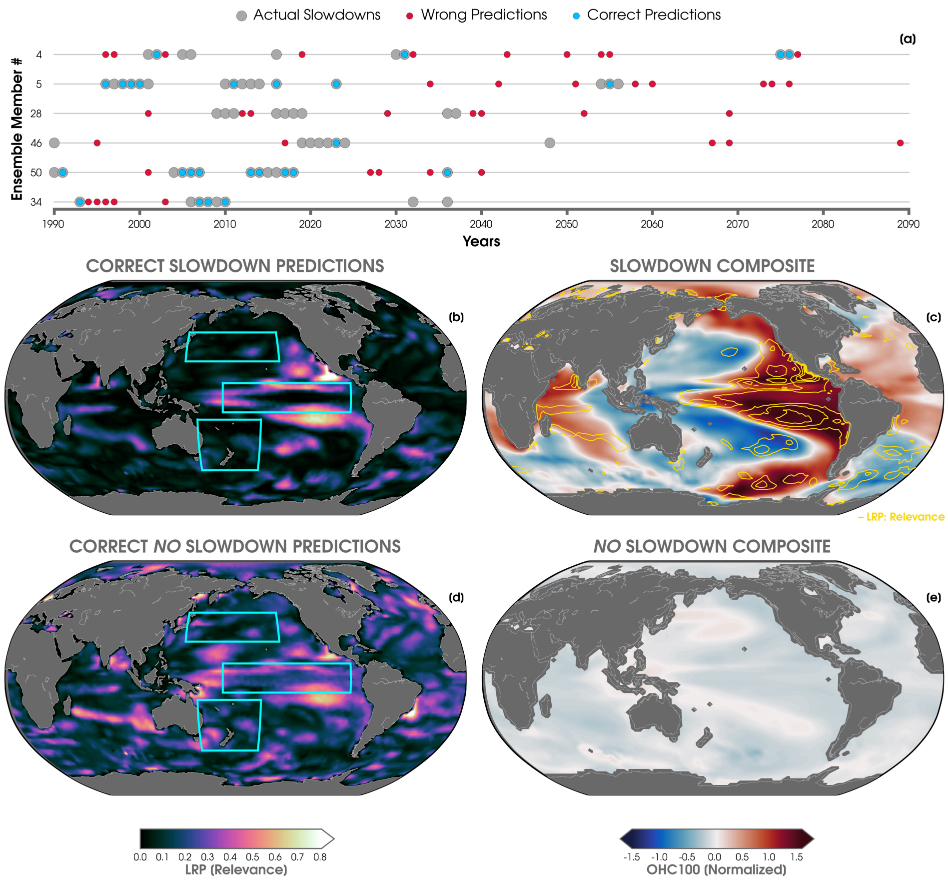

Figure 2. (a) Time series showing the results in each ensemble member of the testing data for the onset of actual slowdown events (gray dots), incorrect slowdown predictions by the ANN (red dots), and correct slowdown predictions by the ANN (blue dots). (b) LRP composite heatmap for the correct slowdown predictions by the ANN (testing data). Higher LRP values indicate greater relevance for the ANN's prediction. LRP values are normalized by the maximum relevance in the composite for visualization purposes. Blue boxes highlight regions of the Tripole Index for the IPO (Henley et al. (2015); 25°N-45°N and 140°E-145°W, 10°S-10°N and 170°E-90°W, 50°S-15°S and 150°E-160°W). (c) Composite of normalized OHC100 for correct slowdown predictions. Yellow contour lines are overlaid to show relevance from the LRP composite in (b). (d) As in (b), but for correct predictions of no slowdowns in decadal warming. (e) As in (c), but for correct predictions of no slowdowns in decadal warming.

ability of the IPO spatial pattern, seeing that we input annual-mean maps of OHC100 into the ANN. As expected from the composite analysis in Figure 2, we find that correct predictions of slowdowns generally correspond to highly positive phases of the IPO index. To demonstrate this point, we select one ensemble member and compare its annual IPO index to the frequency of the slowdowns classifications in the distribution of 100 unique ANNs (Figure S10). We find slowdown predictions often correspond to a positive IPO index in this ensemble member, but also importantly, not every positive IPO results in the prediction of a slowdown event.

To further confirm that the ANN is learning additional spatial information than simply a reflection of the IPO-like pattern of OHC anomalies, we set up a logistic regression problem by inputting only the value of the IPO index in CESM2-LE to predict whether a slowdown event will occur over the next 10 years (F1 score = 0.28). Thus, we find that using global maps of OHC100 as inputs to the fully-connected ANN provides more skillful predictions of warming slowdown events.

3.3 Predicting Slowdown Trends in Observations

Lastly, we test the utility of our neural network for capturing the observed early 2000s slowdown by inputting maps of OHC100 from observations, which are first linearly detrended and then normalized by their own mean and standard deviation at every grid point. Figure 3a shows the frequency of classifying slowdown events for each input map of observed annual-mean OHC100 over the distribution of the 100 unique ANNs. During the overlapping period with the actual early 2000s slowdown, the decade from 2003 to 2012 is classified as a warming slowdown in 31% of the ANNs, and the decade from 2004 to 2013 is classified as a warming slowdown in 50% of the ANNs.

To understand the patterns of anomalous OHC100 that the ANN is using to make a prediction in observations, we evaluate a LRP composite map from a single seed ANN (which correctly predicted two slowdown events in the early 2000s) in Figure S11a. Similar to the LRP composites using CESM2-LE (Figure 2), we find areas of higher relevance in the equatorial western Pacific, south-central Atlantic, and patches in the Indian Ocean. The ANN also predicts the onset of slowdown events mainly during positive phases of the IPO (Figure 3b). Although this correlation is not always the case (e.g., during the positive IPO event in the early 1990s), which again suggests that our ANN is leverag-

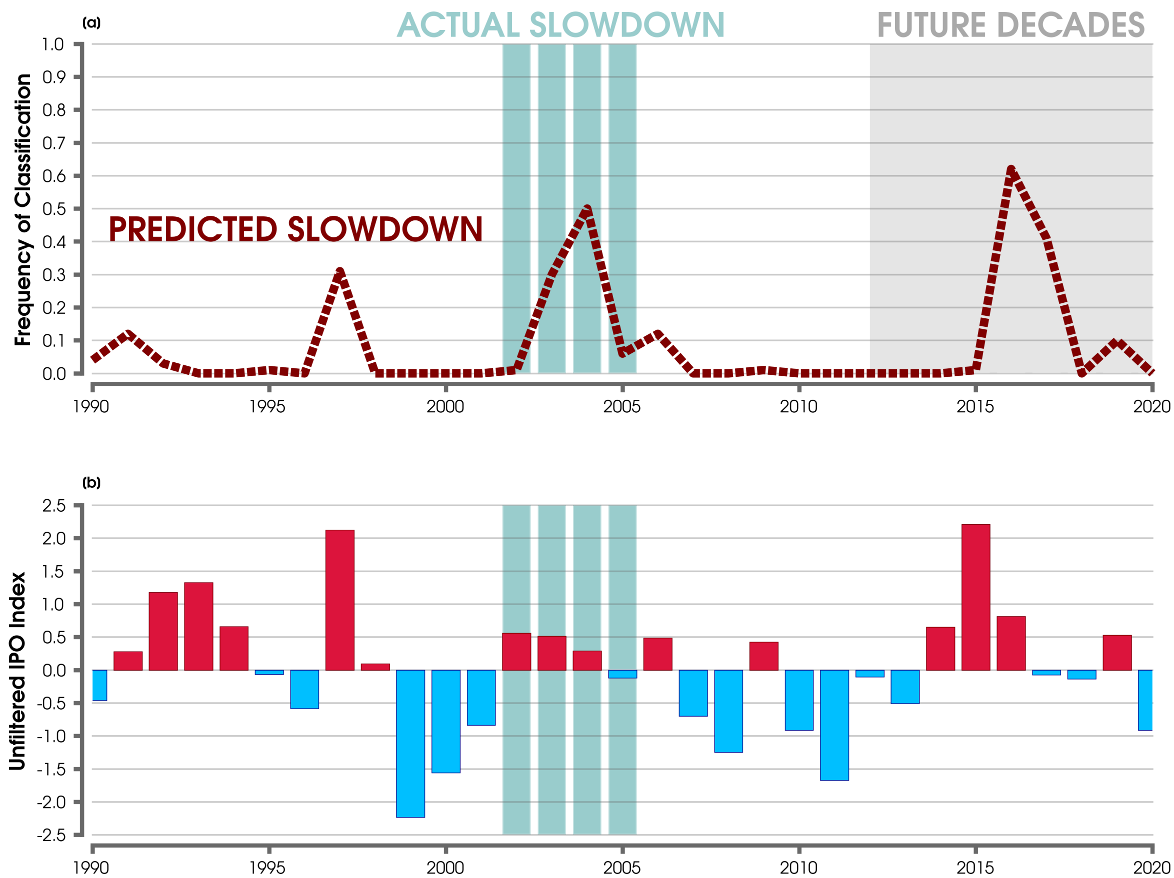

Figure 3. (a) Time series showing the frequency of slowdown onset predictions after inputting observations into 100 ANNs constructed from different combinations of training, testing, and validation data. Green bars show the onset of actual slowdown events in observations. Gray shading indicates 10-year trend periods that extend into the future (e.g., 2012-2021). (b) Time series of the unfiltered Tripole IPO Index (normalized) for each year in observations (red/blue bars). Green bars show the onset of actual slowdown events in observations.

ing additional spatial information than simply the IPO pattern to make predictions. This is also supported by our interpretation of the LRP maps, which show higher relevance regions across the western Pacific and not necessarily the canonical IPO/PDO patterns (Parker et al., 2007; Newman et al., 2016) (Figure S11a).

At the time of our analysis, the last complete decade of GMST observations covers the decade of 2011 to 2020 (Figure 1a). However, since we only need OHC prior to predicting the future 10 years, we can also explore warming slowdown events starting in 2012 (Figure 3a). For these future predictions, the period from 2016 to 2025 is classified as a warming slowdown in 62% of the unique ANNs, and 2017 to 2026 is classified for 41% of them. 2016 was characterized by the dissipation of an extreme El Niño event into a weak La Niña state (Santoso et al., 2017), and the GMST also set a new record high for that respective year (Aaron-Morrison et al., 2017). Similarly, the IPO index also shows a transition from a highly positive phase in 2015 to a neutral or negative phase in the following years through 2020 (Figure 3b). Composites of normalized SST and OHC for 2016 and 2017 show anomalously warm subsurface waters just off the equator in the eastern Pacific and cold pools in the tropical Indo-Pacific and north-central Pacific (Figure S12). Comparing the LRP composite map over 2016 and 2017 with the raw OHC100 anomalies (Figure S11b and Figure S12b), we find higher relevance outlining the warm anomalies in the eastern Pacific and patches of relevance in the Indian Ocean and southern Pacific. The LRP composite for the future slowdown prediction in Figure S11b is more similar to those outlined in CESM2-LE (e.g., Figure 2b), which may provide insight for why the ANN more confidently predicts a slowdown compared to the earlier 2000s event.

4 Summary and Conclusions

In this study, we show the utility of ANNs for predicting temporary slowdowns in the decadal warming of GMST. Although our ANN is trained on climate model data from CESM2-LE, we find that it also produces skillful predictions of the early 2000s warming slowdown in observational data. We further compliment our ANN with a machine learning explainability method (LRP) to attempt to understand where the neural network is looking to make its correct predictions. The LRP maps reveal that the ANN is mainly using off-equatorial anomalies of OHC100 to predict the onset of a decadal warming slowdown. These patterns suggest that the ANN may be learning precursors for transitions to a negative phase of the IPO, although the topic of cross-basin atmosphere-ocean interactions remains an active area of study (Cai et al., 2019; Power et al., 2021). However, we note that decadal warming slowdowns can also occur due to external forcing (e.g., aerosols) or other modes of climate variability (von Känel et al., 2017; Medhaug et al., 2017).

Finally, we note a few important caveats for this work. First, we train our ANN on a large ensemble from only one climate model (CESM2), and thus our results may be influenced by model biases. The results may also be sensitive to the external forcing, such as from the protocol for simulating present-day biomass burning and using the SSP3-7.0 emissions scenario (Rodgers et al., 2021). Further, we only set up our ANN to receive single maps of annual-mean OHC100 as inputs. It is conceivable that the skill of the ANN may improve with additional input data, such as from maps using other OHC levels, which may be provide the ANN more information to identify precursors to decadal warming slowdowns. It may also be useful to combine maps of OHC at different lead times, which was recently demonstrated by Gordon et al. (2021) for predicting transitions in the phase of the PDO. The value of adding more complexity to the ANN architecture will be left for future work. Regardless, even our simple ANN demonstrates that temporary warming slowdowns may have some predictability from Pacific climate variability.

Open Research

Climate model data used in this study are freely available from the CESM2 Large Ensemble Project (`https://www.cesm.ucar.edu/projects/community-projects/LENS2/data-sets.html`). Observations used in this study are from Institute of Atmospheric Physics (IAP) ocean heat content (`http://www.ocean.iap.ac.cn/`), and monthly atmospheric reanalysis data are also freely available from ERA5 (`https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview`).

Conflict of Interest

The Authors declare no conflicts of interest relevant to this study.

Acknowledgments

This work was supported by NOAA MAPP grant NA19OAR4310289 and by the Regional and Global Model Analysis program area of the U.S. Department of Energy's Office of Biological and Environmental Research as part of the Program for Climate Model Diagnosis and Intercomparison project. We thank Dr. John Fasullo (NCAR) for providing fields of OHC from the CESM2 Large Ensemble. We also would like to acknowledge the CESM2 Large Ensemble Community Project, supercomputing resources provided by the IBS Center for Climate Physics in South Korea, and the Cheyenne supercomputer which is operated by the Computational and Information Systems Laboratory (CISL) at NCAR (doi:105065/D5RX99HX) and supported by the National Science Foundation (NSF).

References

Aaron-Morrison, A. P., Ackerman, S. A., Adams, N. G., Adler, R. F., Albanil, A., Alfaro, E. J., … Romanovsky, V. E. (2017, aug). State of the Climate in 2016. *Bulletin of the American Meteorological Society*, *98*(8), Si–S280. Retrieved from `https://journals.ametsoc.org/view/journals/bams/98/8/2017bamsstateoftheclimate.1.xml` doi: 10.1175/2017BAMSSTATEOFTHECLIMATE.1

Amaya, D. J. (2019, sep). The Pacific Meridional Mode and ENSO: a Review. *Current Climate Change Reports*, *5*(4), 296–307. Retrieved from `https://link.springer.com/article/10.1007/s40641-019-00142-x` doi: 10.1007/S40641-019-00142-X

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015, jul). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, *10*(7), e0130140. Retrieved from `http://www.hfsp.org/,` doi: 10.1371/journal.pone.0130140

Barnes, E. A., Mayer, K., Toms, B., Martin, Z., & Gordon, E. (2020, dec). Identifying Opportunities for Skillful Weather Prediction with Interpretable Neural Networks. *arXiv*. Retrieved from `http://arxiv.org/abs/2012.07830`

Barnes, E. A., Toms, B., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2020, sep). Indicator Patterns of Forced Change Learned by an Artificial Neural Network. *Journal of Advances in Modeling Earth Systems*, *12*(9). Retrieved from `https://onlinelibrary.wiley.com/doi/10.1029/2020MS002195` doi: 10.1029/2020MS002195

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., … Thépaut, J.-N. (2021). The ERA5 Global Reanalysis: Preliminary Extension to 1950. *Quarterly Journal of the Royal Meteorological Society*. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4174` doi: 10.1002/QJ.4174

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., … Eade, R. (2016, oct). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3751–3777. doi: 10.5194/GMD-9-3751-2016

Boyer, T., Antonov, J., Baranova, O., Coleman, C., Garcia, H., Grodsky, A., . . . Zweng, M. (2013). World Ocean Database 2013. *NOAA Atlas NESDIS 72*, *NOAA Atlas*.

Boykoff, M. T. (2014, feb). Media discourse on the climate slowdown. *Nature Climate Change*, *4*(3), 156–158. Retrieved from https://www.nature.com/articles/nclimate2156 doi: 10.1038/nclimate2156

Cai, W., Wu, L., Lengaigne, M., Li, T., McGregor, S., Kug, J. S., . . . Chang, P. (2019, mar). Pantropical climate interactions. *Science*, *363*(6430). Retrieved from https://www.science.org/doi/abs/10.1126/science.aav4236 doi: 10.1126/SCIENCE.AAV4236/ASSET/D6EE1BDF-355B-4FAF-9816-980B943B28E6/ASSETS/GRAPHIC/363_AAV4236_F5.JPEG

Cassou, C., Kushnir, Y., Hawkins, E., Pirani, A., Kucharski, F., Kang, I.-S., & Caltabiano, N. (2018, mar). Decadal Climate Variability and Predictability: Challenges and Opportunities. *Bulletin of the American Meteorological Society*, *99*(3), 479–490. Retrieved from https://journals.ametsoc.org/view/journals/bams/99/3/bams-d-16-0286.1.xml doi: 10.1175/BAMS-D-16-0286.1

Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., & Zhu, J. (2017, mar). Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, *3*(3). Retrieved from https://www.science.org/doi/abs/10.1126/sciadv.1601545 doi: 10.1126/SCIADV.1601545

Cheng, L., & Zhu, J. (2016). Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variations. *Journal of Climate*, *29*(15). doi: 10.1175/JCLI-D-15-0730.1

Cowtan, K., & Way, R. G. (2014, jul). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.2297 doi: 10.1002/QJ.2297

Dai, A., Fyfe, J. C., Xie, S.-P., & Dai, X. (2015, apr). Decadal modulation of global surface temperature by internal climate variability. *Nature Climate*

Change, *5*(6), 555–559. Retrieved from https://www.nature.com/articles/nclimate2605 doi: 10.1038/nclimate2605

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020, feb). 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/10.1029/2019MS001916 doi: 10.1029/2019MS001916

Davenport, F. V., & Diffenbaugh, N. S. (2021, aug). Using Machine Learning to Analyze Physical Causes of Climate Change: A Case Study of U.S. Midwest Extreme Precipitation. *Geophysical Research Letters*, *48*(15), e2021GL093787. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GL093787 doi: 10.1029/2021GL093787

Delworth, T. L., & Knutson, T. R. (2000, mar). Simulation of Early 20th Century Global Warming. *Science*, *287*(5461), 2246–2250. Retrieved from https://www.science.org/doi/abs/10.1126/science.287.5461.2246 doi: 10.1126/SCIENCE.287.5461.2246

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012, jan). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, *116*, 140–158. doi: 10.1016/J.RSE.2010.10.017

Dunn, R. J. H., Aldred, F., Gobron, N., Miller, J. B., Willett, K. M., Ades, M., ... Zotta, R. M. (2021, aug). Global Climate [in "State of the Climate in 2020"]. *Bulletin of the American Meteorological Society*, *102*(8), S11–S142. Retrieved from https://journals.ametsoc.org/view/journals/bams/102/8/BAMS-D-21-0098.1.xml doi: 10.1175/BAMS-D-21-0098.1

Easterling, D. R., & Wehner, M. F. (2009, apr). Is the climate warming or cooling? *Geophysical Research Letters*, *36*(8), 8706. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2009GL037810 doi: 10.1029/2009GL037810

Ebert-Uphoff, I., & Hilburn, K. (2020, dec). Evaluation, Tuning, and Interpretation of Neural Networks for Working with Images in Meteorological Applications. *Bulletin of the American Meteorological Society*, *101*(12), E2149–E2170. Retrieved from https://journals.ametsoc.org/view/journals/bams/101/12/BAMS-D-20-0097.1.xml doi: 10.1175/BAMS-D-20-0097.1

England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., … Santoso, A. (2014, feb). Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nature Climate Change*, *4*(3), 222–227. Retrieved from https://www.nature.com/articles/nclimate2106 doi: 10.1038/nclimate2106

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

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., … Rummukainen, M. (2013). IPCC AR5 - Chapter 9: Evaluation of Climate Models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 741–866.

Fyfe, J. C., Gillett, N. P., & Zwiers, F. W. (2013, aug). Overestimated global warming over the past 20 years. *Nature Climate Change*, *3*(9), 767–769. Retrieved from https://www.nature.com/articles/nclimate1972 doi: 10.1038/nclimate1972

Fyfe, J. C., Meehl, G. A., England, M. H., Mann, M. E., Santer, B. D., Flato, G. M., … Swart, N. C. (2016, feb). Making sense of the early-2000s warming slowdown. *Nature Climate Change*, *6*(3), 224–228. Retrieved from https://www.nature.com/articles/nclimate2938 doi: 10.1038/nclimate2938

Fyfe, J. C., Merryfield, W. J., Kharin, V., Boer, G. J., Lee, W.-S., & von Salzen, K. (2011, nov). Skillful predictions of decadal trends in global mean surface temperature. *Geophysical Research Letters*, *38*(22), 22801. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011GL049508 doi: 10.1029/2011GL049508

Gibson, P. B., Chapman, W. E., Altinok, A., Delle Monache, L., DeFlorio, M. J., & Waliser, D. E. (2021, aug). Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts. *Communications Earth & Environment*, *2*(1), 1–13. Retrieved from https://www.nature.com/articles/s43247-021-00225-4 doi: 10.1038/s43247-021-00225-4

Gordon, E. M., Barnes, E. A., & Hurrell, J. W. (2021, nov). Oceanic Harbingers of Pacific Decadal Oscillation Predictability in CESM2 Detected by Neural Networks. *Geophysical Research Letters*, *48*(21), e2021GL095392. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GL095392 doi: 10.1029/2021GL095392

Guemas, V., Doblas-Reyes, F. J., Andreu-Burillo, I., & Asif, M. (2013, apr). Retrospective prediction of the global warming slowdown in the past decade. *Nature Climate Change*, *3*(7), 649–653. Retrieved from https://www.nature.com/articles/nclimate1863 doi: 10.1038/nclimate1863

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019, sep). Deep learning for multi-year ENSO forecasts. *Nature =*, *573*(7775), 568–572. Retrieved from https://www.nature.com/articles/s41586-019-1559-7 doi: 10.1038/s41586-019-1559-7

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010, dec). GLOBAL SURFACE TEMPERATURE CHANGE. *Reviews of Geophysics*, *48*(4), 4004. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2010RG000345 doi: 10.1029/2010RG000345

Hawkins, E., Frame, D., Harrington, L., Joshi, M., King, A., Rojas, M., & Sutton, R. (2020, mar). Observed Emergence of the Climate Change Signal: From the Familiar to the Unknown. *Geophysical Research Letters*, *47*(6). Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL086259 doi: 10.1029/2019GL086259

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K. (2015). A Tripole Index for the Interdecadal Pacific Oscillation. *Climate Dynamics*. doi: 10.1007/s00382-015-2525-1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020, may). The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society*. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803 doi: 10.1002/qj.3803

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., &

Saynisch-Wagner, J. (2021, aug). Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence*, *3*(8), 667–674. Retrieved from `https://www.nature.com/articles/s42256-021-00374-3` doi: 10.1038/s42256-021-00374-3

Izumo, T., Vialard, J., Lengaigne, M., De Boyer Montegut, C., Behera, S. K., Luo, J. J., ... Yamagata, T. (2010, feb). Influence of the state of the Indian Ocean Dipole on the following year's El Niño. *Nature Geoscience*, *3*(3), 168–172. Retrieved from `https://www.nature.com/articles/ngeo760` doi: 10.1038/ngeo760

Johnson, N. C., Amaya, D. J., Ding, Q., Kosaka, Y., Tokinaga, H., & Xie, S. P. (2020, may). Multidecadal modulations of key metrics of global climate change. *Global and Planetary Change*, *188*, 103149. doi: 10.1016/J.GLOPLACHA.2020.103149

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., ... Zhang, H.-M. (2015, jun). Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, *348*(6242), 1469–1472. Retrieved from `https://www.science.org/doi/abs/10.1126/science.aaa5632` doi: 10.1126/SCIENCE.AAA5632

Kosaka, Y., & Xie, S.-P. (2013, aug). Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, *501*(7467), 403–407. Retrieved from `https://www.nature.com/articles/nature12534` doi: 10.1038/nature12534

Labe, Z. M., & Barnes, E. A. (2021). Detecting climate signals using explainable AI with single-forcing large ensembles. *Journal of Advances in Modeling Earth Systems*, *13*(6), 1–18. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002464` doi: 10.1029/2021MS002464

Le, T., Ha, K. J., Bae, D. H., & Kim, S. H. (2020, oct). Causal effects of Indian Ocean Dipole on El Niño–Southern Oscillation during 1950–2014 based on high-resolution models and reanalysis data. *Environmental Research Letters*, *15*(10), 1040b6. Retrieved from `https://iopscience.iop.org/article/10.1088/1748-9326/abb96d/meta` doi: 10.1088/1748-9326/ABB96D

Lewandowsky, S., Cowtan, K., Risbey, J. S., Mann, M. E., Steinman, B. A., Oreskes, N., & Rahmstorf, S. (2018, dec). The 'pause' in global warming in historical context: (II). Comparing models to observations. *Environmental Research Letters*, *13*(12), 123007. Retrieved from https://iopscience.iop.org/article/10.1088/1748-9326/aaf372/meta doi: 10.1088/1748-9326/AAF372

Lewandowsky, S., Risbey, J. S., & Oreskes, N. (2016, may). The "Pause" in Global Warming: Turning a Routine Fluctuation into a Problem for Science. *Bulletin of the American Meteorological Society*, *97*(5), 723–733. Retrieved from https://journals.ametsoc.org/view/journals/bams/97/5/bams-d-14-00106.1.xml doi: 10.1175/BAMS-D-14-00106.1

Li, T. W., & Baker, N. C. (2016, jul). Detecting Warming Hiatus Periods in CMIP5 Climate Model Projections. *International Journal of Atmospheric Sciences*, 1–7. doi: 10.1155/2016/9657659

Li-Jing, C., Jiang, Z., & Abraham, J. (2015, jan). Global Upper Ocean Heat Content Estimation: Recent Progress and the Remaining Challenges. *Atmospheric and Oceanic Science Letters*, *8*(6), 333–338. Retrieved from https://www.tandfonline.com/action/journalInformation?journalCode=taos20 doi: 10.3878/AOSL20150031

Liu, J., Tang, Y., Wu, Y., Li, T., Wang, Q., & Chen, D. (2021, oct). Forecasting the Indian Ocean Dipole With Deep Learning Techniques. *Geophysical Research Letters*, *48*(20), e2021GL094407. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GL094407 doi: 10.1029/2021GL094407

Liu, W., & Xie, S. P. (2018, jun). An ocean view of the global surface warming hiatus. *Oceanography*, *31*(2 Special Issue), 72–79. doi: 10.5670/OCEANOG.2018.217

Maher, N., Gupta, A. S., & England, M. H. (2014). *Drivers of decadal hiatus periods in the 20th and 21st centuries* (Vol. 41) (No. 16). doi: 10.1002/2014GL060527

Maher, N., Lehner, F., & Marotzke, J. (2020, may). Quantifying the role of internal variability in the temperature we expect to observe in the coming decades. *Environmental Research Letters*, *15*(5), 054014. Retrieved from https://iopscience.iop.org/article/10.1088/1748-9326/ab7d02/meta doi: 10.1088/1748-9326/AB7D02

Maher, N., Milinski, S., & Ludwig, R. (2021, apr). Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics*, *12*(2), 401–418. Retrieved from https://esd.copernicus.org/articles/12/401/2021/ doi: 10.5194/esd-12-401-2021

Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2021, mar). Neural Network Attribution Methods for Problems in Geoscience: A Novel Synthetic Benchmark Dataset. *arXiv*. Retrieved from http://arxiv.org/abs/2103.10005

Mann, M. E., Miller, S. K., Rahmstorf, S., Steinman, B. A., & Tingley, M. (2017, aug). Record temperature streak bears anthropogenic fingerprint. *Geophysical Research Letters*, *44*(15), 7936–7944. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GL074056 doi: 10.1002/2017GL074056

Mann, M. E., Steinman, B. A., Miller, S. K., Frankcombe, L. M., England, M. H., & Cheung, A. H. (2016, apr). Predictability of the recent slowdown and subsequent recovery of large-scale surface warming using statistical methods. *Geophysical Research Letters*, *43*(7), 3459–3467. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2016GL068159 doi: 10.1002/2016GL068159

Mayer, K. J., & Barnes, E. A. (2021, may). Subseasonal Forecasts of Opportunity Identified by an Explainable Neural Network. *Geophysical Research Letters*, *48*(10), e2020GL092092. Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2020GL092092 doi: 10.1029/2020GL092092

McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019, nov). Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, *100*(11), 2175–2199. Retrieved from http://journals.ametsoc.org/bams/article-pdf/100/11/2175/4876688/bams-d-18-0195{_}1.pdf doi: 10.1175/BAMS-D-18-0195.1

Medhaug, I., Stolpe, M. B., Fischer, E. M., & Knutti, R. (2017, may). Reconciling controversies about the 'global warming hiatus'. *Nature*, *545*(7652), 41–47. Retrieved from https://www.nature.com/articles/nature22315 doi: 10.1038/nature22315

Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A., & Trenberth, K. E. (2011, sep). Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change*, *1*(7), 360–364. Retrieved from https://www.nature.com/articles/nclimate1229 doi: 10.1038/nclimate1229

Meehl, G. A., Hu, A., Arblaster, J. M., Fasullo, J., & Trenberth, K. E. (2013, sep). Externally Forced and Internally Generated Decadal Climate Variability Associated with the Interdecadal Pacific Oscillation. *Journal of Climate*, *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., & Teng, H. (2014, mar). CMIP5 multi-model hindcasts for the mid-1970s shift and early 2000s hiatus and predictions for 2016–2035. *Geophysical Research Letters*, *41*(5), 1711–1716. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2014GL059256 doi: 10.1002/2014GL059256

Meehl, G. A., Teng, H., & Arblaster, J. M. (2014, sep). Climate model simulations of the observed early-2000s hiatus of global warming. *Nature Climate Change*, *4*(10), 898–902. Retrieved from https://www.nature.com/articles/nclimate2357 doi: 10.1038/nclimate2357

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., & Müller, K. R. (2017, may). Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition*, *65*, 211–222. doi: 10.1016/j.patcog.2016.11.008

Montavon, G., Samek, W., & Müller, K. R. (2018, feb). *Methods for interpreting and understanding deep neural networks* (Vol. 73). Elsevier Inc. doi: 10.1016/j.dsp.2017.10.011

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., … Simpson, I. R. (2021, feb). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1029/2019JD032361https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JD032361https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019JD032361 doi: 10.1029/2019JD032361

Nadiga, B. T. (2021, apr). Reservoir Computing as a Tool for Climate Predictability Studies. *Journal of Advances in Modeling Earth Systems*, *13*(4), e2020MS002290. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020MS002290 doi: 10.1029/2020MS002290

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., ... Smith, C. A. (2016, jun). The Pacific Decadal Oscillation, Revisited. *Journal of Climate*, *29*(12), 4399–4427. Retrieved from http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0508.1 doi: 10.1175/JCLI-D-15-0508.1

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... Sanderson, B. M. (2016, sep). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. doi: 10.5194/GMD-9-3461-2016

Papalexiou, S. M., Rajulapati, C. R., Clark, M. P., & Lehner, F. (2020, oct). Robustness of CMIP6 Historical Global Mean Temperature Simulations: Trends, Long-Term Persistence, Autocorrelation, and Distributional Shape. *Earth's Future*, *8*(10), e2020EF001667. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020EF001667 doi: 10.1029/2020EF001667

Parker, D., Folland, C., Scaife, A., Knight, J., Colman, A., Baines, P., & Dong, B. (2007, sep). Decadal to multidecadal variability and the climate change background. *Journal of Geophysical Research: Atmospheres*, *112*(D18), 18115. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2007JD008411 doi: 10.1029/2007JD008411

Phillips, A. S., Deser, C., Fasullo, J., Schneider, D. P., & Simpson, I. R. (2020). *Assessing Climate Variability and Change in Model Large Ensembles: A User's Guide to the Climate Variability Diagnostics Package for Large Ensembles Version 1.0* (Tech. Rep.). Boulder, Colorado: NCAR. Retrieved from https://opensky.ucar.edu/islandora/object/manuscripts:1001 doi: 10.5065/h7c7-f961

Power, S., Lengaigne, M., Capotondi, A., Khodri, M., Vialard, J., Jebri, B., ... Henley, B. J. (2021, oct). Decadal climate variability in the tropical Pacific: Characteristics, causes, predictability, and prospects. *Science*, *374*(6563). Retrieved from https://www.science.org/doi/abs/10.1126/science.aay9165 doi: 10.1126/SCIENCE.AAY9165

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019, feb). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. Retrieved from `https://www.nature.com/articles/s41586-019-0912-1` doi: 10.1038/s41586-019-0912-1

Risbey, J. S., Lewandowsky, S., Cowtan, K., Oreskes, N., Rahmstorf, S., Jokimäki, A., & Foster, G. (2018, dec). A fluctuation in surface temperature in historical context: reassessment and retrospective on the evidence. *Environmental Research Letters*, *13*(12), 123008. Retrieved from `https://iopscience.iop.org/article/10.1088/1748-9326/aaf342/meta` doi: 10.1088/1748-9326/AAF342

Roberts, C. D., Palmer, M. D., McNeall, D., & Collins, M. (2015, feb). Quantifying the likelihood of a continued hiatus in global warming. *Nature Climate Change*, *5*(4), 337–342. Retrieved from `https://www.nature.com/articles/nclimate2531` doi: 10.1038/nclimate2531

Rodgers, K., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., ... Yeager, S. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics Discussions*, 1–22. doi: 10.5194/ESD-2021-50

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999, sep). A dipole mode in the tropical Indian Ocean. *Nature 1999*, *401*(6751), 360–363. Retrieved from `https://www.nature.com/articles/43854` doi: 10.1038/43854

Santer, B. D., Bonfils, C., Painter, J. F., Zelinka, M. D., Mears, C., Solomon, S., ... Wentz, F. J. (2014, feb). Volcanic contribution to decadal changes in tropospheric temperature. *Nature Geoscience 2014 7:3*, *7*(3), 185–189. Retrieved from `https://www.nature.com/articles/ngeo2098` doi: 10.1038/ngeo2098

Santoso, A., Mcphaden, M. J., & Cai, W. (2017, dec). The Defining Characteristics of ENSO Extremes and the Strong 2015/2016 El Niño. *Reviews of Geophysics*, *55*(4), 1079–1129. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017RG000560` doi: 10.1002/2017RG000560

Schmidt, G. A., Shindell, D. T., & Tsigaridis, K. (2014, feb). Reconciling warming trends. *Nature Geoscience*, *7*(3), 158–160. Retrieved from `https://www`

.nature.com/articles/ngeo2105 doi: 10.1038/ngeo2105

Sévellec, F., & Drijfhout, S. S. (2018, aug). A novel probabilistic forecast system predicting anomalously warm 2018-2022 reinforcing the long-term global warming trend. *Nature Communications*, *9*(1), 1–12. Retrieved from https://www.nature.com/articles/s41467-018-05442-8 doi: 10.1038/s41467-018-05442-8

Sévellec, F., Sinha, B., & Skliris, N. (2016, aug). The rogue nature of hiatuses in a global warming climate. *Geophysical Research Letters*, *43*(15), 8169–8177. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2016GL068950 doi: 10.1002/2016GL068950

Smith, D. M., Booth, B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., … Thompson, V. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*(10). doi: 10.1038/nclimate3058

Sonnewald, M., & Lguensat, R. (2021, aug). Revealing the Impact of Global Heating on North Atlantic Circulation Using Transparent Machine Learning. *Journal of Advances in Modeling Earth Systems*, *13*(8), e2021MS002496. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002496 doi: 10.1029/2021MS002496

Sonnewald, M., Lguensat, R., Jones, D. C., Dueben, P. D., Brajard, J., & Balaji, V. (2021, jul). Bridging observations, theory and numerical simulation of the ocean using machine learning. *Environmental Research Letters*, *16*(7), 073008. Retrieved from https://iopscience.iop.org/article/10.1088/1748-9326/ac0eb0/meta doi: 10.1088/1748-9326/AC0EB0

Tang, Y., & Duan, A. (2021, oct). Using deep learning to predict the East Asian summer monsoon. *Environmental Research Letters*. Retrieved from https://iopscience.iop.org/article/10.1088/1748-9326/ac34bc/meta doi: 10.1088/1748-9326/AC34BC

Thompson, D. W. J., Wallace, J. M., Jones, P. D., & Kennedy, J. J. (2009, nov). Identifying Signatures of Natural Climate Variability in Time Series of Global-Mean Surface Temperature: Methodology and Insights. *Journal of Climate*, *22*(22), 6120–6141. Retrieved from https://journals.ametsoc.org/view/journals/clim/22/22/2009jcli3089.1.xml doi: 10.1175/2009JCLI3089.1

Titchner, H. A., & Rayner, N. A. (2014, mar). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, *119*(6), 2864–2889. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JD020316 doi: 10.1002/2013JD020316

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020, sep). Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability. *Journal of Advances in Modeling Earth Systems*, *12*(9). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2019MS002002 doi: 10.1029/2019MS002002

Toms, B. A., Barnes, E. A., & Hurrell, J. W. (2021, jun). Assessing Decadal Predictability in an Earth-System Model Using Explainable Neural Networks. *Geophysical Research Letters*, *48*(12), e2021GL093842. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GL093842 doi: 10.1029/2021GL093842

Trenberth, K. E., Caron, J. M., Stepaniak, D. P., & Worley, S. (2002, apr). Evolution of El Niño–Southern Oscillation and global atmospheric surface temperatures. *Journal of Geophysical Research: Atmospheres*, *107*(D8), AAC 5–1. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2000JD000298 doi: 10.1029/2000JD000298

Van Marle, M. J., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A. L., Field, R. D., … Van Der Werf, G. R. (2017, sep). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750-2015). *Geoscientific Model Development*, *10*(9), 3329–3357. doi: 10.5194/GMD-10-3329-2017

von Känel, L., Frölicher, T. L., & Gruber, N. (2017, aug). Hiatus-like decades in the absence of equatorial Pacific cooling and accelerated global ocean heat uptake. *Geophysical Research Letters*, *44*(15), 7909–7918. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GL073578 doi: 10.1002/2017GL073578

Watanabe, M., Kamae, Y., Yoshimori, M., Oka, A., Sato, M., Ishii, M., … Kimoto, M. (2013, jun). Strengthening of ocean heat uptake efficiency associated with the recent climate hiatus. *Geophysical Research Letters*, *40*(12), 3175–3179.

Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/grl.50541 doi: 10.1002/GRL.50541

Wu, A., & Hsieh, W. W. (2004, jan). The nonlinear Northern Hemisphere winter atmospheric response to ENSO. *Geophysical Research Letters*, *31*(2). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2003GL018885 doi: 10.1029/2003GL018885

Xie, S.-P., & Kosaka, Y. (2017, mar). What Caused the Global Surface Warming Hiatus of 1998–2013? *Current Climate Change Reports 2017 3:2*, *3*(2), 128–140. Retrieved from https://link.springer.com/article/10.1007/s40641-017-0063-0 doi: 10.1007/S40641-017-0063-0

Yoo, J. H., Moon, S., Ha, K. J., Yun, K. S., & Lee, J. Y. (2020, aug). Cases for the sole effect of the Indian Ocean Dipole in the rapid phase transition of the El Niño–Southern Oscillation. *Theoretical and Applied Climatology*, *141*(3-4), 999–1007. Retrieved from https://link.springer.com/article/10.1007/s00704-020-03265-6 doi: 10.1007/S00704-020-03265-6/FIGURES/8

Zhang, C., Li, S., Luo, F., & Huang, Z. (2019, oct). The global warming hiatus has faded away: An analysis of 2014–2016 global surface air temperatures. *International Journal of Climatology*, *39*(12), 4853–4868. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.6114 doi: 10.1002/JOC.6114

References From the Supporting Information

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th usenix symposium on operating systems design and implementation, osdi 2016.*

Agarap, A. F. (2018, mar). Deep Learning using Rectified Linear Units (ReLU). *arXiv*. Retrieved from http://arxiv.org/abs/1803.08375

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*.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015, jul). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, *10*(7), e0130140. Retrieved from `http://www.hfsp.org/,` doi: 10.1371/journal.pone.0130140

Capotondi, A., Deser, C., Phillips, A. S., Okumura, Y., & Larson, S. M. (2020, dec). ENSO and Pacific Decadal Variability in the Community Earth System Model Version 2. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2019MS002022. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS002022` doi: 10.1029/2019MS002022

Chen, H.-C., Fei-Fei-Jin, Zhao, S., Wittenberg, A. T., & Xie, S. (2021, dec). ENSO Dynamics in the E3SM-1-0, CESM2, and GFDL-CM4 Climate Models. *Journal of Climate*, *34*(23), 9365–9384. Retrieved from `https://journals.ametsoc.org/view/journals/clim/34/23/JCLI-D-21-0355.1.xml` doi: 10.1175/JCLI-D-21-0355.1

Crameri, F. (2018, jan). Scientific colour maps. *Zenodo*. Retrieved from `https://zenodo.org/record/4153113` doi: 10.5281/ZENODO.4153113

Crameri, F., Shephard, G. E., & Heron, P. J. (2020, dec). The misuse of colour in science communication. *Nature Communications*, *11*(1), 1–10. Retrieved from `https://doi.org/10.1038/s41467-020-19160-7` doi: 10.1038/s41467-020-19160-7

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., ... Yeager, S. G. (2012). The CCSM4 ocean component. *Journal of Climate*, *25*(5). doi: 10.1175/JCLI-D-11-00091.1

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020, feb). 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/10.1029/2019MS001916` doi: 10.1029/2019MS001916

Fasullo, J. T. (2020, aug). Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geoscientific Model Development*, *13*(8), 3627–3642. doi: 10.5194/GMD-13-3627-2020

Friedman, J. H. (2012, jul). Fast sparse regression and classification. *International Journal of Forecasting*, *28*(3), 722–738. doi: 10.1016/j.ijforecast.2012.05.001

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., … Mills, M. J. (2019, jul). High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters*, *46*(14), 8329–8337. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GL083978` doi: 10.1029/2019GL083978

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.*

Green, D. A. (2011). A colour scheme for the display of astronomical intensity images. *Bulletin of the Astronomical Society of India*, *39*(2).

Harris, C. R., Jarrod Millman, K., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020, sep). Array programming with NumPy. *Nature*, *585*(7825), 357. Retrieved from `https://doi.org/10.1038/s41586-020-2649-2` doi: 10.1038/s41586-020-2649-2

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K. (2015). A Tripole Index for the Interdecadal Pacific Oscillation. *Climate Dynamics*. doi: 10.1007/s00382-015-2525-1

Hunter, J. D. (2007, may). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 99–104. doi: 10.1109/MCSE.2007.55

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., … Marshall, S. (2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9). doi: 10.1175/BAMS-D-12-00121.1

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., … Vertenstein, M. (2015, aug). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. Retrieved from `http://journals.ametsoc.org/doi/10.1175/BAMS-D-13-00255.1` doi: 10.1175/BAMS-D-13-00255.1

Lecun, Y., Bengio, Y., & Hinton, G. (2015, may). *Deep learning* (Vol. 521) (No. 7553). Nature Publishing Group. Retrieved from `https://www.nature.com/articles/nature14539` doi: 10.1038/nature14539

Meehl, G. A., Arblaster, J. M., Bates, S., Richter, J. H., Tebaldi, C., Gettelman, A., … Strand, G. (2020, sep). Characteristics of Future Warmer Base States in CESM2. *Earth and Space Science*, *7*(9), e2020EA001296. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020EA001296 doi: 10.1029/2020EA001296

Montavon, G., Samek, W., & Müller, K. R. (2018, feb). *Methods for interpreting and understanding deep neural networks* (Vol. 73). Elsevier Inc. doi: 10.1016/j.dsp.2017.10.011

NCAR. (2019). *The NCAR Command Language (Version 6.6.2).* Boulder, Colorado. Retrieved from http://dx.doi.org/10.5065/D6WD3XH5 doi: http://dx.doi.org/10.5065/D6WD3XH5

Nesterov, Y. (1983). A method for unconstrained convex minimization problem with the rate of convergence o(1/kˆ2). *Doklady AN USSR*, *269*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Ruder, S. (2016, sep). An overview of gradient descent optimization algorithms. *arXiv*. Retrieved from http://arxiv.org/abs/1609.04747

Schulzweida, U. (2019, feb). CDO User Guide. *Zenodo*. Retrieved from https://zenodo.org/record/2558193 doi: 10.5281/ZENODO.2558193

Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., … Yeager, S. (2010). The Parallel Ocean Program (POP) reference manual: Ocean component of the Community Climate System Model (CCSM). *Rep. LAUR-01853*, *141*.

Thyng, K., Greene, C., Hetland, R., Zimmerle, H., & DiMarco, S. (2016, sep). True Colors of Oceanography: Guidelines for Effective and Accurate Colormap Selection. *Oceanography*, *29*(3), 9–13. Retrieved from https://tos.org/oceanography/article/true-colors-of-oceanography-guidelines-for-effective-and-accurate-colormap doi: 10.5670/oceanog.2016.66

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3). doi: 10.1038/s41592-019-0686-2

Zender, C. S. (2008). Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). *Environmental Modelling and Software*, *23*(10-11). doi: 10.1016/j.envsoft.2008.03.004

Supporting Information for "Predicting slowdowns in decadal climate warming trends with explainable neural networks"

Zachary M. Labe[1] and Elizabeth A. Barnes[1]

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

Contents of this file

Corresponding author: Zachary M. Labe (zmlabe@rams.colostate.edu)

Text S1: Community Earth System Model Version 2 (CESM2)

CESM2 uses a nominal 1° horizontal resolution and includes 32 vertical levels with a model top at 2.26 hPa. Components for CESM2 include an atmosphere model from Community Atmosphere Model version 6 (CAM6; Danabasoglu et al., 2020) and an ocean model from Parallel Ocean Program Version 2 (POP2; Smith et al., 2010; Danabasoglu et al., 2012), which are further coupled to interactive ice, land, and ocean biogeochemistry models. Additional details on model development can be found in Danabasoglu et al. (2020). Overall, CESM2 scores well in comparison to other Coupled Model Intercomparison Project Phase 6 (CMIP6) models (e.g., Fasullo, 2020) and includes numerous improvements to cloud microphysics, the ocean surface boundary layer, and land processes over the previous model generation (CESM1; Hurrell et al., 2013; Kay et al., 2015). Future projections of global mean surface temperature (GMST) in CESM2 generally fall in the upper range of CMIP6 models, which is likely due to a higher equilibrium climate sensitivity (Gettelman et al., 2019; Meehl et al., 2020). Representation of the El Niño-Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) in CESM2 compare fairly well to observations, but there are still some large differences in simulated amplitude and spatial patterns (Capotondi et al., 2020; Chen et al., 2021).

Text S2: Artificial Neural Network Architecture

Here, we provide an overview of the artificial neural network (ANN) used in our analysis (Figure S1). Our input layer receives vectorized maps of annual-mean ocean heat content in the 0-100 m depth (OHC100) from the CESM2 Large Ensemble Community Project (CESM2-LE), where each unit represents one grid box (13248 units per map from 92 latitudes by 144 longitudes).

The input vector is then fed into two hidden layers with 30 nodes each, and our output layer contains two nodes (yes or no for a decadal warming slowdown). We apply the rectified linear unit (ReLU; $f(x) = max(0, x)$; Agarap, 2018) to our hidden nodes and include a softmax operator in the output layer. The softmax function remaps the output values so that they sum to one and can then be interpreted as the ANN's confidence for each prediction output. For example, the winning predicted category (i.e., yes or no slowdown) will have a confidence value greater than 0.5. Our ANN uses a categorical cross entropy loss function. Overall, this general setup is commonly used for many neural network classification problems (e.g., Lecun et al., 2015; Goodfellow et al., 2016). Given the large class imbalance (i.e., fewer number of slowdown training samples compared to non-slowdowns), we find an ANN architecture of this complexity achieves a reasonable F1 score (harmonic mean of the ANN's precision and recall) (Figures S2-S3).

Before training our ANN, we standardize our maps of OHC100 by subtracting the mean and dividing by the standard deviation separately at every grid point and across all years for the training ensemble members (13248 units). Specifically, we train our ANN using 70% of the climate model data (28 ensemble members), validate on 15% (6 ensemble members), and test on the remaining 15% (6 ensemble members). During training, we use the stochastic gradient descent optimizer and turn on the Nesterov momentum parameter (set to 0.9) (Nesterov, 1983; Ruder, 2016). Our learning rate is set to 0.001, and the batch size is 128. While we set the ANN to train using 500 epochs, we apply early stopping on the validation loss to prevent overfitting. In the other words, the ANN is finished training if the validation loss does not improve for 10 epochs in a row. Using this approach, our ANN generally reaches no more than 35 epochs and is restored to the iteration with the best model weights.

To further account for overfitting, we apply L_2 ridge regularization (Friedman, 2012) to the weights of the first hidden layer. Our L_2 parameter is set to 0.5 after exploring several different combinations of ANN architectures, hyperparameters, and random initialization seeds (Figures S2-S3). Ridge regularization ensures the ANN is not sensitive to outlier weights, which helps to consider any spatial autocorrelation in the input fields of OHC100. Finally, we assign class weights in the loss function, since there is a large class imbalance with only 16 or fewer slowdown events per individual ensemble member (Figure 1c). This parameters tells the model to pay more attention to the underrepresented class during the training process. Figure S4 shows the results of ANNs using a range of class weights compared to the original class imbalance (approximately 8.8 to 1). For the main figures and analysis presented here, we selected a smaller fraction to be applied to the balanced class weights (approximately 4.4 to 1).

Text S3: Open Software/Tools

Preprocessing and regridding were completed using NCL v6.2.2 (NCAR, 2019), NCO v4.9.3 (Zender, 2008), and CDO v1.9.8 (Schulzweida, 2019). Figures and main analysis were completed using open source Python v3.7.6, Numpy v1.19 (Harris et al., 2020), SciPy v1.4.1 (Virtanen et al., 2020), Matplotlib v3.2.2 (Hunter, 2007), and colormaps provided by cmocean v2.0 (Thyng et al., 2016), Palettable's cubehelix v3.3.0 (Green, 2011), and Scientific v7.0.0 (Crameri, 2018; Crameri et al., 2020). Additional Python packages used for development of the ANN and LRP visualizations include TensorFlow v1.15.0/v2.4.0 (Abadi et al., 2016), Scikit-learn v0.24.2 (Pedregosa et

al., 2011), and iNNvestigate v1.0.8 (Alber et al., 2019). Computer code for the ANN architecture and exploratory data analysis is available at `https://github.com/zmlabe/predictGMSTrate` (*note that a DOI archival repository will be provided using Zenodo if this paper is considered for publication*). References for the data sets are provided throughout the study. Lastly, we would like to thank all the scientists, software engineers, and administrators who contributed to the development of CESM2.

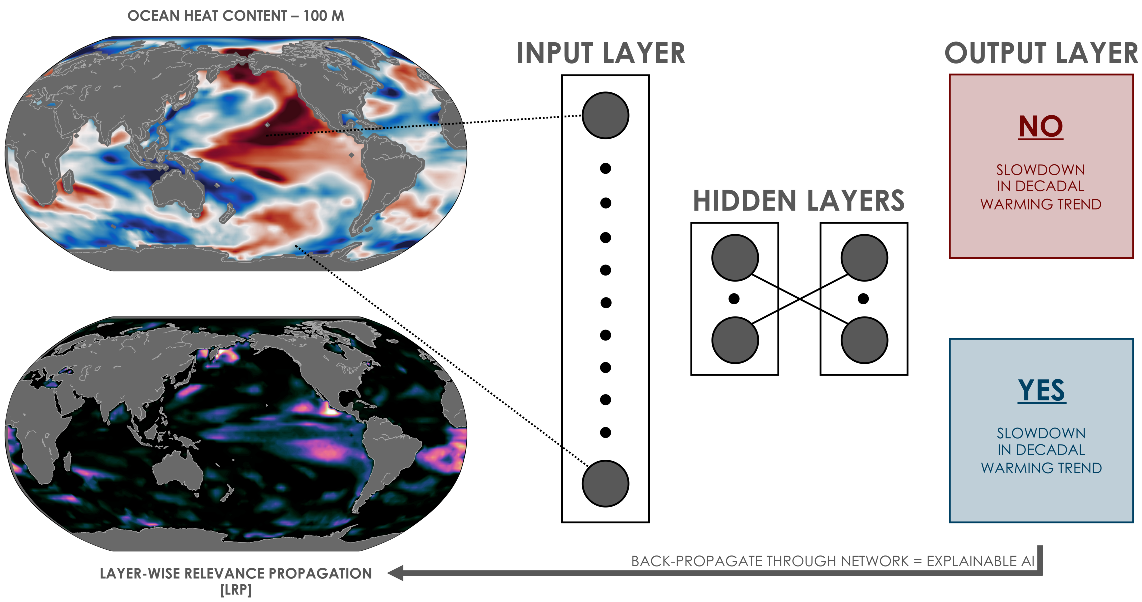

Figure S1. Schematic of the artificial neural network (ANN) used in this study for predicting the onset of a slowdown in decadal warming trend (output layer) from a global map of annual mean ocean heat content in the 0-100 m depth (input layer). The ANN consists of two hidden layers that both contain 30 hidden nodes. The output layer includes a softmax activation function. An example heatmap using layer-wise relevance propagation (LRP; Bach et al., 2015; Montavon et al., 2018) is also illustrated here. LRP highlights the regions of greater relevance for the ANN to decide whether a slowdown event will occur for the next 10 years.

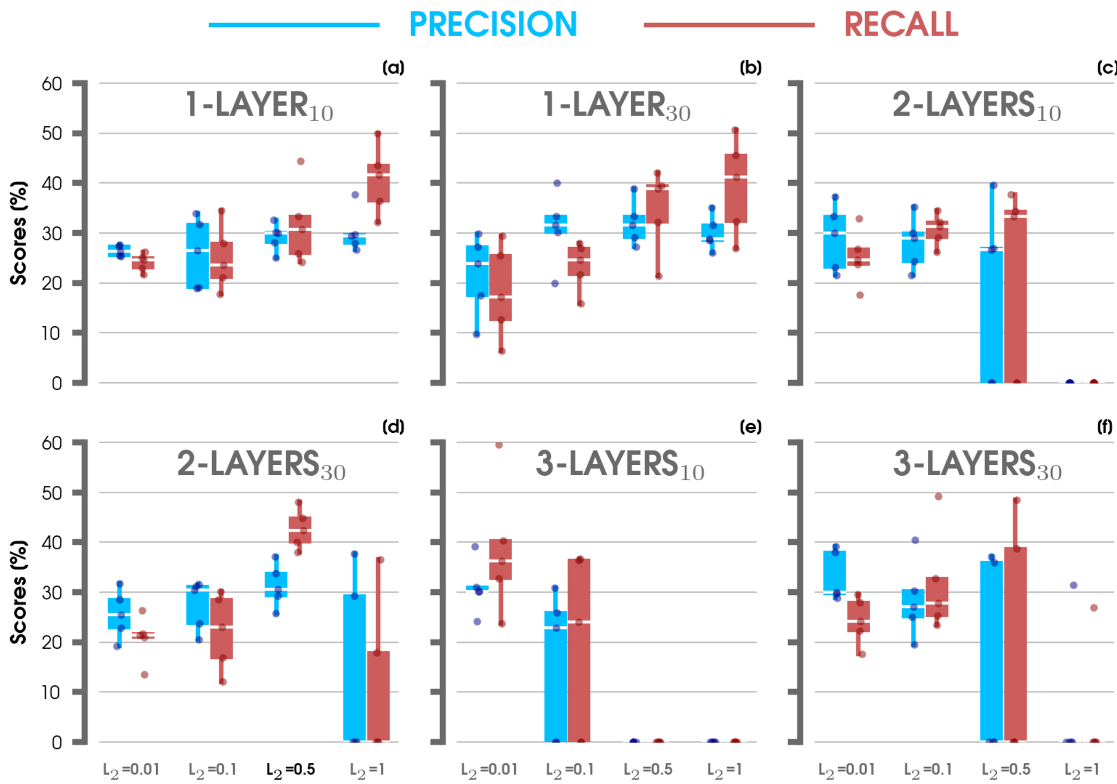

Figure S2. Box-and-whisker plots showing the precision (blue) and recall (red) scores for validation data. Results are shown for ANN architectures using (a) 1 hidden layers of 10 nodes, (b) 1 hidden layers of 30 nodes, (c) 2 hidden layers of 10 nodes each, (d) 2 hidden layers of 30 nodes each, (e) 3 hidden layers of 10 nodes each, (f) and 3 hidden layers of 30 nodes each (f). Each architecture also compares scores for different L_2 regularization values (0.01, 0.1, 0.5, 1). The box-and-whisker distributions of ANNs are comprised of 5 iterations (different combinations of training, testing, and validation data and random initialization seeds). The architecture used in the main analysis is labeled in bold for 2 hidden layers of 30 nodes each (subplot d).

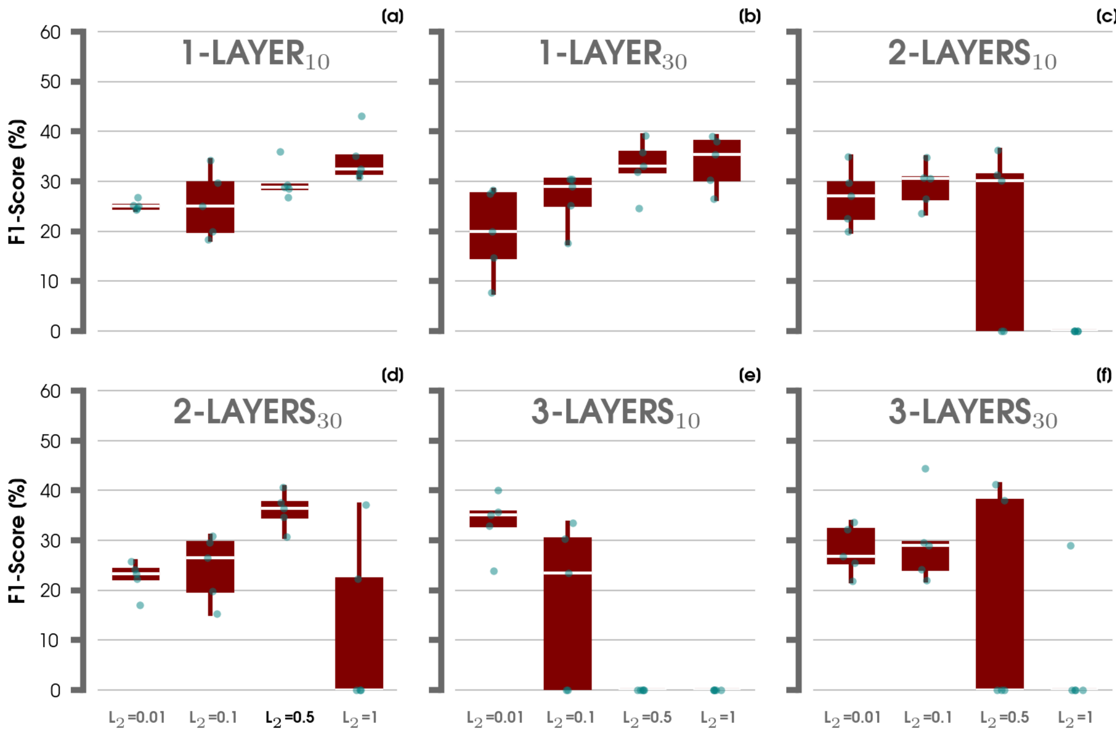

Figure S3. Box-and-whisker plots showing the F1 score for validation data. Results are shown for ANN architectures using (a) 1 hidden layers of 10 nodes, (b) 1 hidden layers of 30 nodes, (c) 2 hidden layers of 10 nodes each, (d) 2 hidden layers of 30 nodes each, (e) 3 hidden layers of 10 nodes each, (f) and 3 hidden layers of 30 nodes each (f). Each architecture also compares scores for different L_2 regularization values (0.01, 0.1, 0.5, 1). The box-and-whisker distributions of ANNs are comprised of 5 iterations (different combinations of training, testing, and validation data and random initialization seeds). The architecture used in the main analysis is labeled in bold for 2 hidden layers of 30 nodes each (subplot d).

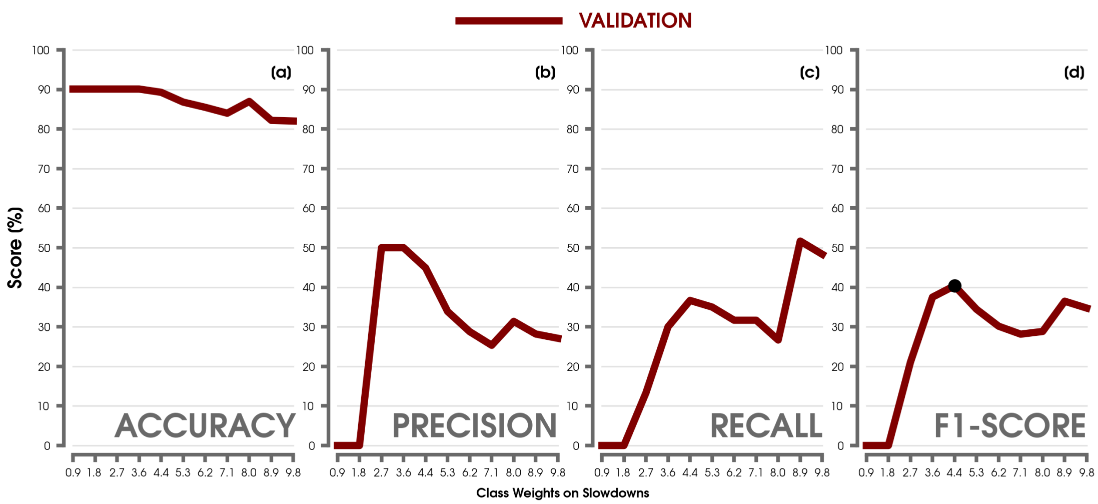

Figure S4. (a) Accuracy, (b) precision, (c) recall, (d) and F1 scores for validation data in the ANN architecture used throughout the paper, but with different class weights on slowdown events. The class weight used in the main analysis is shown with a marker for the F1 score.

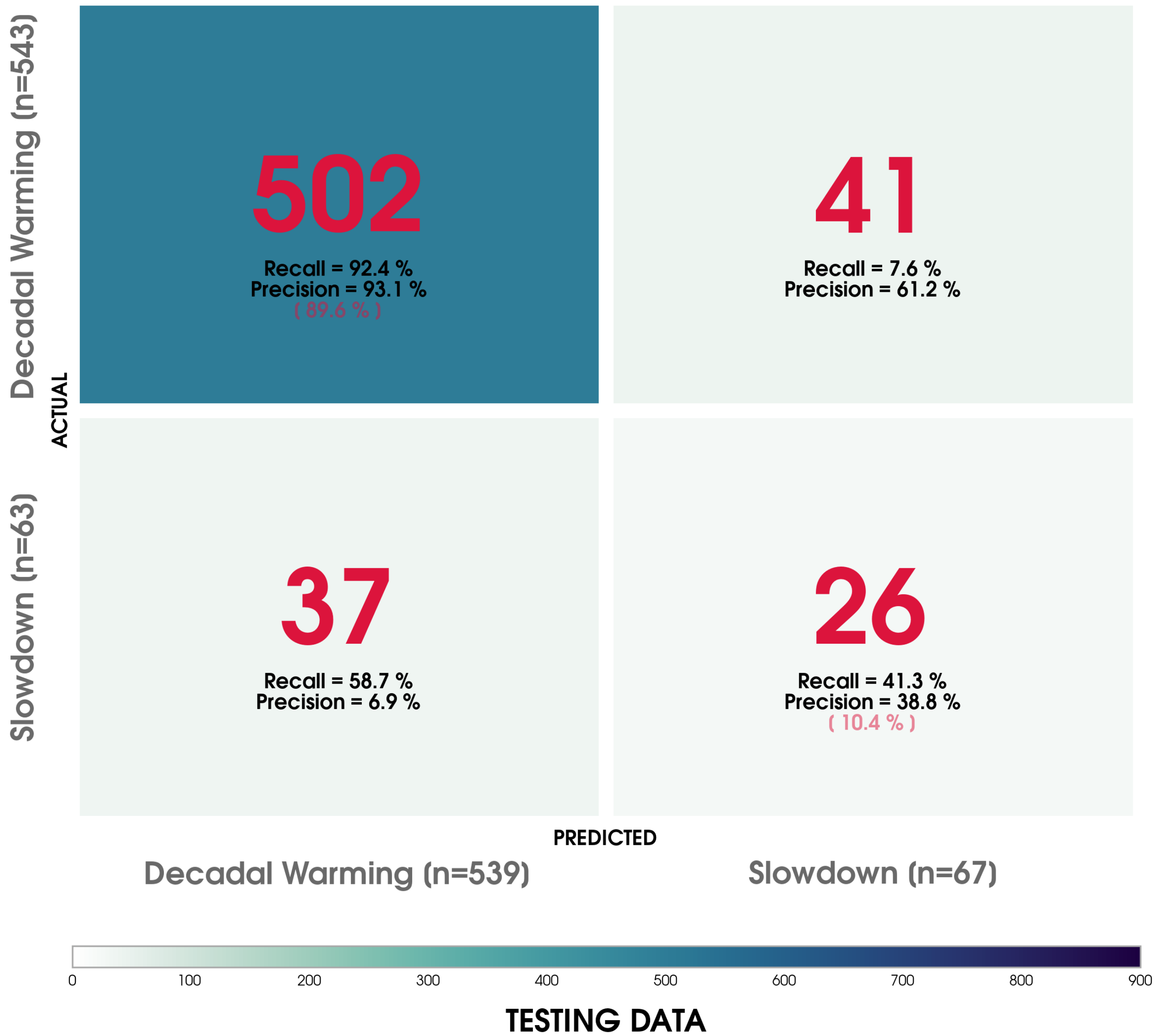

Figure S5. Confusion matrix of validation data for all predictions. The shading and large red values inside each box represents the sample size (n) for each classification category bin. The small red percentage value in the top-left and bottom-right boxes is the random chance probability for picking a slowdown event or not. Scores for recall and precision are also shown for each category bin.

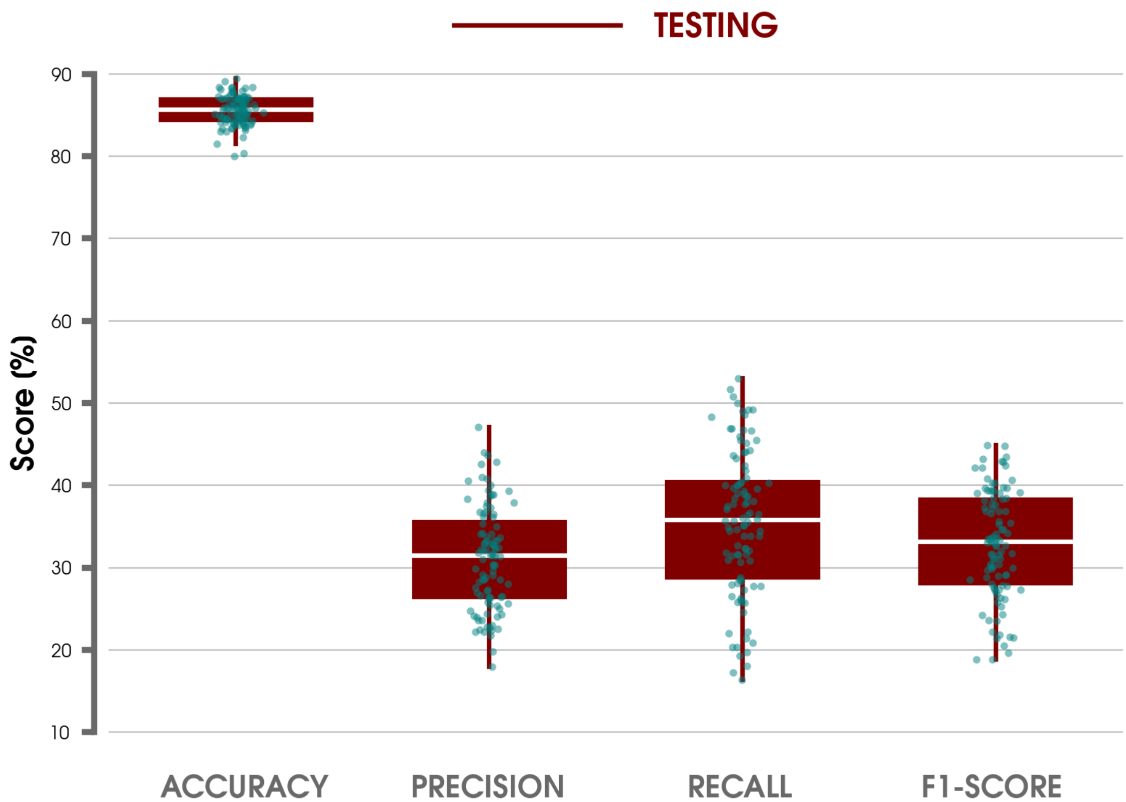

Figure S6. Box-and-whisker plots showing the accuracy, precision, recall, and F1 scores for the ANN architecture used throughout the paper after considering 100 different combinations of training, testing, and validation data and random initialization seeds.

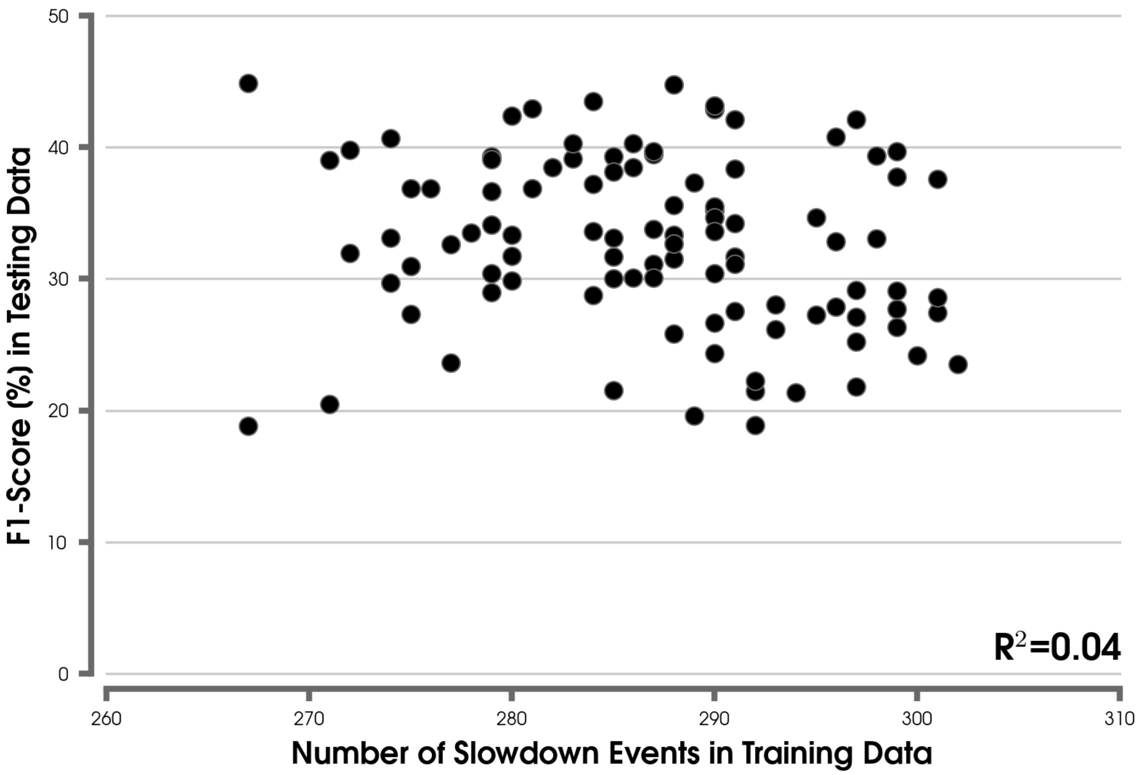

Figure S7. Scatter plot showing the number of slowdown events in training data compared to the F1 score for testing data in 100 ANNs using different combinations of training, testing, and validation data and random initialization seeds.

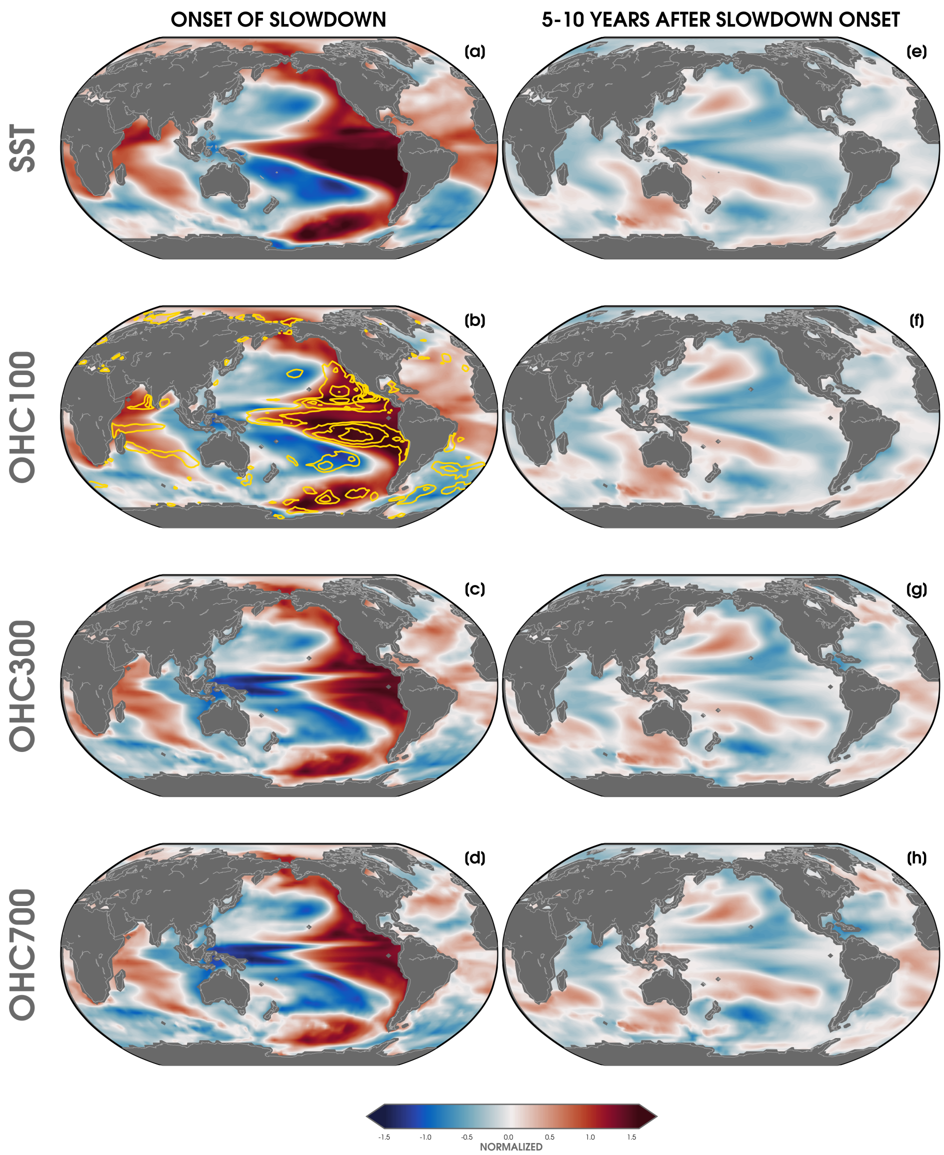

Figure S8. (a) Composite of normalized sea surface temperature (SST) for correct slowdown predictions by the ANN. (b) As in (a), but for ocean heat content in the 0-100 m layer (OHC100). Yellow contour lines are overlaid to show relevance from the LRP composite in main Figure 2b. (c) As in (a), but for ocean heat content in the 0-300 m layer (OHC300). (d) As in (a), but for ocean heat content in the 0-700 m layer (OHC700). (e-h) As in (a-d), but for composites of 5-10 years after the correct slowdown predictions.

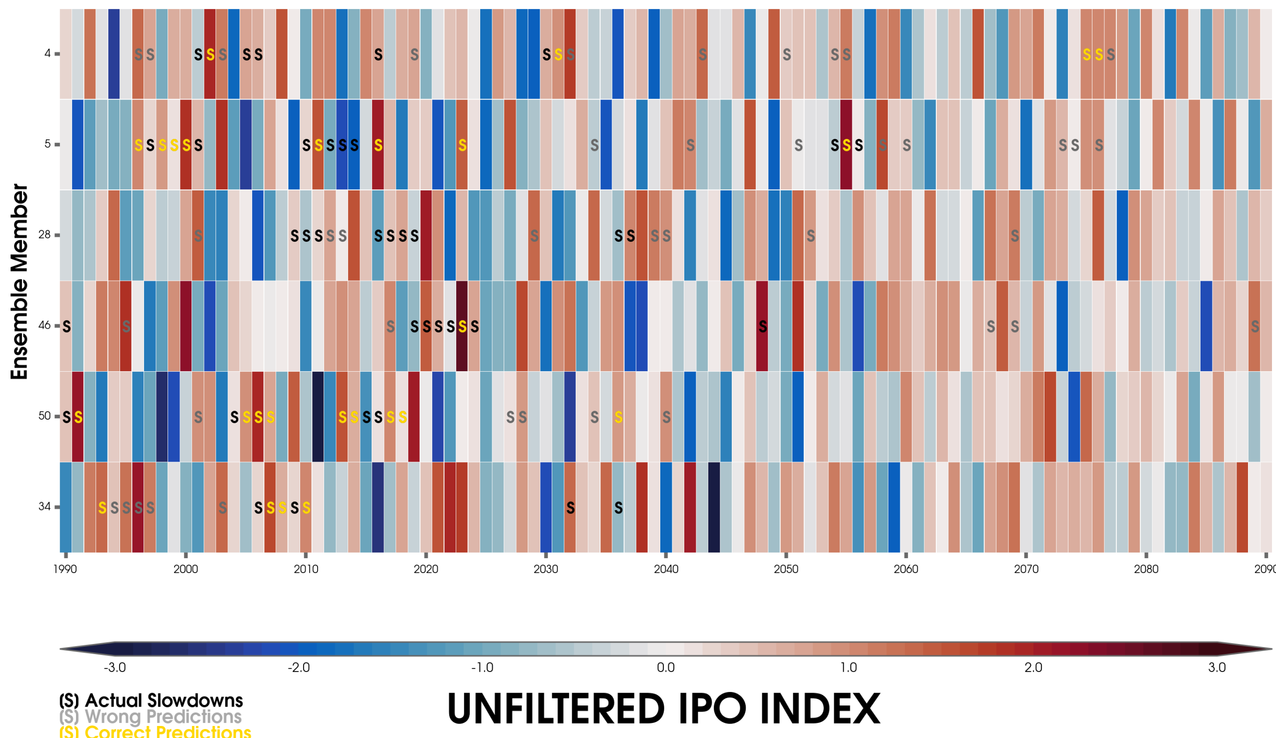

Figure S9. Unfiltered Tripole IPO Index (normalized) for each year of the six ensemble members in the testing data. Correct predictions by the ANN for the onset of slowdown events are highlighted with a yellow 'S' in each ensemble member, wrong slowdown predictions by the ANN are highlighted with a gray 'S' in each ensemble member, and all other actual slowdown events are indicated with a black 'S' in each ensemble member.

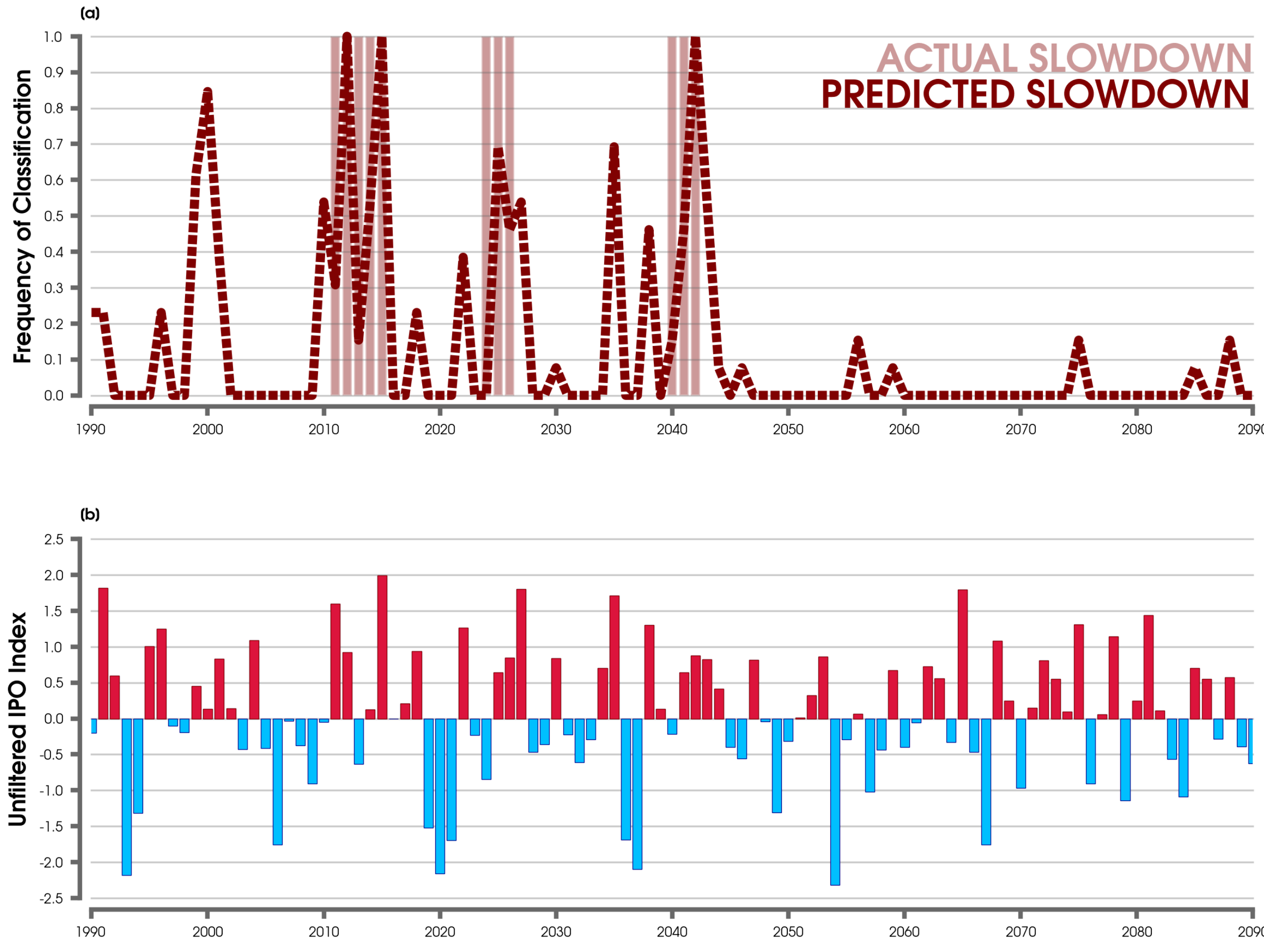

Figure S10. (a) Time series showing the frequency of slowdown onset predictions for one ensemble member realization (in testing data) using 13 ANNs constructed from different combinations of training, testing, and validation data (dashed dark red line). Light red bars show the onset of actual slowdown events in the ensemble member. (b) Time series of the unfiltered Tripole IPO Index (normalized) for each year in the same ensemble member (red/blue bars).

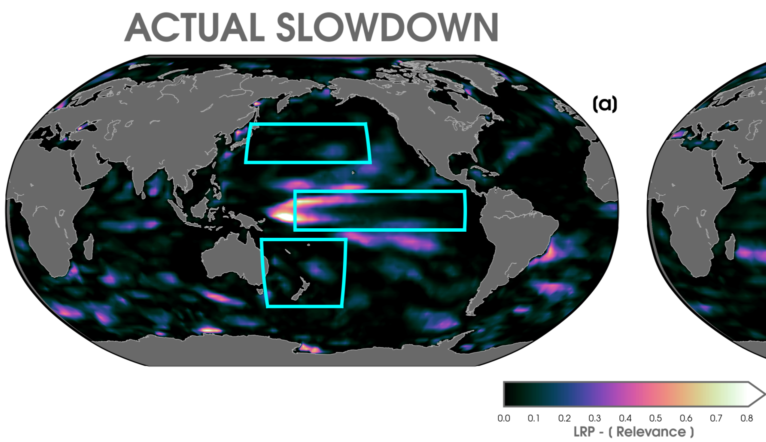

Figure S11. (a) LRP composite heatmap for the correct slowdown predictions by the ANN in observations. (b) As in (a), but for the ANN slowdown predictions during the future 10-year trend periods. Higher LRP values indicate greater relevance for the ANN's prediction. LRP values are normalized by the maximum relevance in the composite for visualization purposes. Blue boxes highlight regions of the Tripole Index for the IPO (Henley et al. (2015); 25°N-45°N and 140°E-145°W, 10°S-10°N and 170°E-90°W, 50°S-15°S and 150°E-160°W).

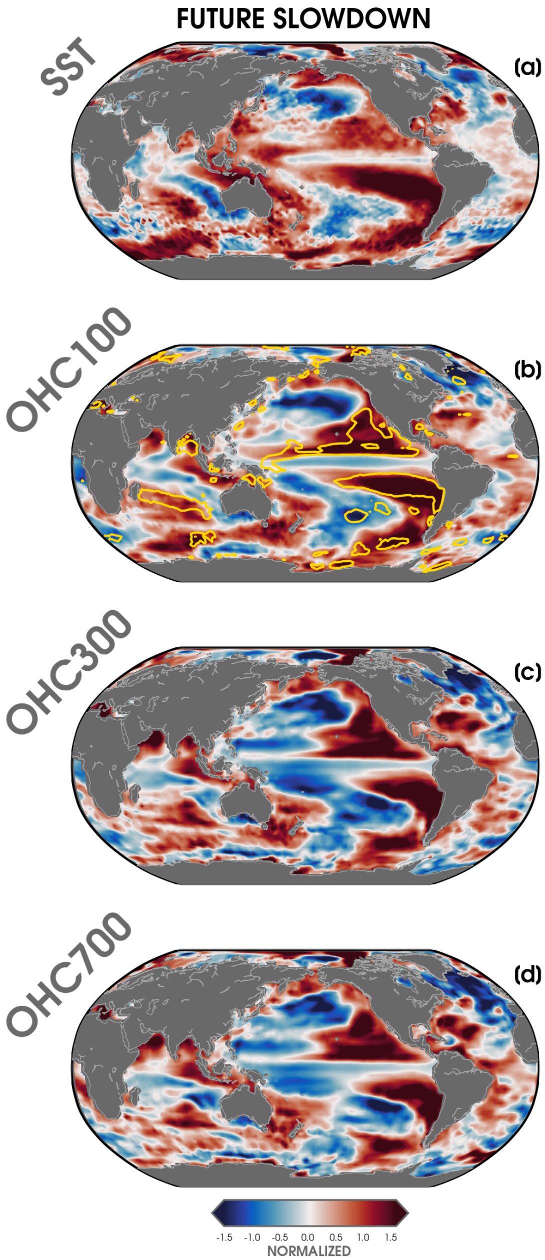

Figure S12. (a) Composite of normalized sea surface temperature (SST) for the future slowdown predictions after testing observations with the ANN. (b) As in (a), but for ocean heat content in the 0-100 m layer (OHC100). Yellow contour lines are overlaid to show relevance from the LRP composite in Figure S11b. (c) As in (a), but for ocean heat content in the 0-300 m layer (OHC300). (d) As in (a), but for ocean heat content in the 0-700 m layer (OHC700).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th usenix symposium on operating systems design and implementation, osdi 2016.*

Agarap, A. F. (2018, mar). Deep Learning using Rectified Linear Units (ReLU). *arXiv*. Retrieved from http://arxiv.org/abs/1803.08375

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*.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015, jul). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, *10*(7), e0130140. Retrieved from http://www.hfsp.org/, doi: 10.1371/journal.pone.0130140

Capotondi, A., Deser, C., Phillips, A. S., Okumura, Y., & Larson, S. M. (2020, dec). ENSO and Pacific Decadal Variability in the Community Earth System Model Version 2. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2019MS002022. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019MS002022 doi: 10.1029/2019MS002022

Chen, H.-C., Fei-Fei-Jin, Zhao, S., Wittenberg, A. T., & Xie, S. (2021, dec). ENSO Dynamics in the E3SM-1-0, CESM2, and GFDL-CM4 Climate Models. *Journal of Climate*, *34*(23), 9365–9384. Retrieved from https://journals.ametsoc.org/view/journals/clim/34/23/JCLI-D-21-0355.1.xml doi: 10.1175/JCLI-D-21-0355.1

Crameri, F. (2018, jan). Scientific colour maps. *Zenodo*. Retrieved from https://zenodo.org/record/4153113 doi: 10.5281/ZENODO.4153113

Crameri, F., Shephard, G. E., & Heron, P. J. (2020, dec). The misuse of colour in science communication. *Nature Communications*, *11*(1), 1–10. Retrieved from https://doi.org/10.1038/s41467-020-19160-7 doi: 10.1038/s41467-020-19160-7

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., ... Yeager, S. G. (2012). The CCSM4 ocean component. *Journal of Climate*, *25*(5). doi: 10.1175/JCLI-D-11-00091.1

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020, feb). 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/10.1029/2019MS001916 doi: 10.1029/2019MS001916

Fasullo, J. T. (2020, aug). Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geoscientific Model Development*, *13*(8), 3627–3642. doi: 10.5194/GMD-13-3627-2020

Friedman, J. H. (2012, jul). Fast sparse regression and classification. *International Journal of Forecasting*, *28*(3), 722–738. doi: 10.1016/j.ijforecast.2012.05.001

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., ... Mills, M. J. (2019, jul). High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters*, *46*(14), 8329–8337. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GL083978 doi: 10

.1029/2019GL083978

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.*

Green, D. A. (2011). A colour scheme for the display of astronomical intensity images. *Bulletin of the Astronomical Society of India*, *39*(2).

Harris, C. R., Jarrod Millman, K., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020, sep). Array programming with NumPy. *Nature*, *585*(7825), 357. Retrieved from https://doi.org/10.1038/s41586-020-2649-2 doi: 10.1038/s41586-020-2649-2

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K. (2015). A Tripole Index for the Interdecadal Pacific Oscillation. *Climate Dynamics*. doi: 10.1007/s00382-015-2525-1

Hunter, J. D. (2007, may). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 99–104. doi: 10.1109/MCSE.2007.55

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., ... Marshall, S. (2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9). doi: 10.1175/BAMS-D-12-00121.1

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., ... Vertenstein, M. (2015, aug). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. Retrieved from http://journals.ametsoc.org/doi/10.1175/BAMS-D-13-00255.1 doi: 10.1175/BAMS-D-13-00255.1

Lecun, Y., Bengio, Y., & Hinton, G. (2015, may). *Deep learning* (Vol. 521) (No. 7553). Nature Publishing Group. Retrieved from `https://www.nature.com/articles/nature14539` doi: 10.1038/nature14539

Meehl, G. A., Arblaster, J. M., Bates, S., Richter, J. H., Tebaldi, C., Gettelman, A., ... Strand, G. (2020, sep). Characteristics of Future Warmer Base States in CESM2. *Earth and Space Science*, *7*(9), e2020EA001296. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020EA001296` doi: 10.1029/2020EA001296

Montavon, G., Samek, W., & Müller, K. R. (2018, feb). *Methods for interpreting and understanding deep neural networks* (Vol. 73). Elsevier Inc. doi: 10.1016/j.dsp.2017.10.011

NCAR. (2019). *The NCAR Command Language (Version 6.6.2).* Boulder, Colorado. Retrieved from `http://dx.doi.org/10.5065/D6WD3XH5` doi: http://dx.doi.org/10.5065/D6WD3XH5

Nesterov, Y. (1983). A method for unconstrained convex minimization problem with the rate of convergence o(1/k^2). *Doklady AN USSR*, *269*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Ruder, S. (2016, sep). An overview of gradient descent optimization algorithms. *arXiv*. Retrieved from `http://arxiv.org/abs/1609.04747`

Schulzweida, U. (2019, feb). CDO User Guide. *Zenodo*. Retrieved from `https://zenodo.org/record/2558193` doi: 10.5281/ZENODO.2558193

Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., ... Yeager, S. (2010).

The Parallel Ocean Program (POP) reference manual: Ocean component of the Community Climate System Model (CCSM). *Rep. LAUR-01853*, *141*.

Thyng, K., Greene, C., Hetland, R., Zimmerle, H., & DiMarco, S. (2016, sep). True Colors of Oceanography: Guidelines for Effective and Accurate Colormap Selection. *Oceanography*, *29*(3), 9–13. Retrieved from `https://tos.org/oceanography/article/true-colors-of-oceanography-guidelines-for-effective-and-accurate-colormap` doi: 10.5670/oceanog.2016.66

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3). doi: 10.1038/s41592-019-0686-2

Zender, C. S. (2008). Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). *Environmental Modelling and Software*, *23*(10-11). doi: 10.1016/j.envsoft.2008.03.004