

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

Emily M Gordon^{1,1} and Elizabeth A. Barnes^{1,1}

¹Colorado State University

November 30, 2022

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 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.

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

4 **Emily M. Gordon¹ and Elizabeth A. Barnes¹**

5 ¹Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

6 **Key Points:**

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

dictability 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-

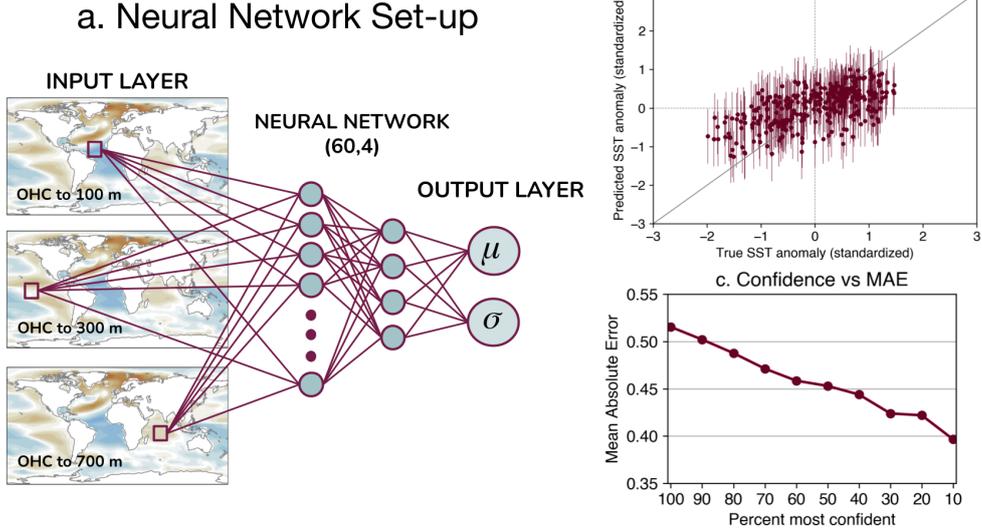


Figure 1. a. Schematic of the artificial neural network architecture. b. Scatter plot of predicted SST anomaly (y axis) vs true SST anomaly (x axis). Dots represent predicted μ values, while vertical lines represent the 1σ range. c. Prediction mean absolute error (MAE) as a function of prediction confidence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

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

124 2.2 Artificial Neural Network

125 Artificial neural networks (ANNs) are used to predict the average SST anomaly
 126 at a lead time of 1-5 years and 3-7 years, i.e. the ANN predicts the average 60 month
 127 SST anomaly in the next 12-72 months, or 36-96 months respectively. In this experiment
 128 the ANN is trained to predict the SST evolution in the CESM2 pre-industrial control,
 129 so for example, one input sample is OHC information from a specific time step in the
 130 control run, and the output prediction is the average SST anomaly over the next 12-72
 131 months in the control run. A schematic of our neural network architecture is provided
 132 in Figure 1a and a brief overview of ANNs for geoscience applications can be found in
 133 e.g. Toms et al. (2020). The predictors are three OHC grids, where each grid is OHC
 134 integrated to a different depth (100 m, 300 m and 700 m). We chose varying depths of
 135 OHC because each contains information corresponding to different forms of climate vari-
 136 ability. For example, the upper levels of the ocean integrate atmospheric forcing, and hence
 137 capture atmospheric variability as well as surface ocean dynamics (Frankignoul & Has-

138 selmann, 1977). The variability in lower levels of the ocean is guided by a combination
 139 of slow moving ocean circulation and the incorporation of mixed layer processes via the
 140 annual cycle in the thermocline (Alexander & Deser, 1995). By inputting three OHC depths
 141 into the neural network, it can theoretically combine different oceanic and atmospheric
 142 processes to make its predictions. The three ocean grids are vectorized with points over
 143 land removed resulting in a total 7947 input pixels. This input is connected to a hidden
 144 layer of 60 nodes which is then connected to another hidden layer of 4 nodes (see Fig.1).
 145 In this network, all layers are densely connected meaning all nodes in the previous layer
 146 are connected to all the nodes in the next layer. Furthermore, all nodes in the hidden
 147 layers use the rectified linear unit (ReLU) activation function. Finally this second layer
 148 is connected to the output layer of two nodes which serve as the parameters of the pre-
 149 dicted conditional distribution (see details in the next paragraph). Here the distribution
 150 is a normal distribution as we found allowing skewness did not significantly improve the
 151 network’s performance (not shown).

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

173 2.3 AMV and IPO indices

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

183 3 Results

184 3.1 Evaluating Performance

185 In this study, 10 networks (identical architecture, only varying the initial network
 186 random seed) are trained at each SST grid point in the ocean and we show the results
 187 of the best neural network at each grid point. To designate the “best” network, we se-

188 lect the ANN with the lowest mean absolute error (MAE, difference between predicted
 189 μ and true y) on the 10% of samples with the lowest σ predictions in the validation set.
 190 This designation leverages a fundamental characteristic of a network that has learned
 191 predictability in the data: prediction error should decrease as predicted σ decreases. We
 192 demonstrate this idea in Figure 1c where we show a network trained to predict SST in
 193 1-5 years in the North Atlantic (52.5°N, 325°E). Along the x-axis, we threshold by in-
 194 creasing confidence with the y-axis showing corresponding MAE for those predictions.
 195 For all samples, the MAE is ~ 0.52 however for the 40% most confident predictions the
 196 MAE has dropped to 0.46. For the 10% most confident predictions, the MAE has dropped
 197 further to ~ 0.39 implying the network has learned samples that lead to more predictable
 198 SST anomaly. We hence refer to lower σ predictions as more confident predictions, or
 199 more predictable inputs. For some grid points, all networks fail to learn anything, mean-
 200 ing they always predict an SST anomaly of zero (or very close to zero). These networks
 201 are removed before analysis, resulting in 30% of networks (525/1709) removed for lead
 202 years 1-5, and 39% (675/1709) for lead years 3-7.

203 3.2 Predicting SST

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

218 The prediction skill for lead years 3-7 is shown in Fig 2b and highlights similar re-
 219 gions as being more predictable as in lead years 1-5. Furthermore, there does not seem
 220 to be a substantial loss in skill between these two lead times. This, coupled with the spa-
 221 tial distribution of prediction skill, suggests that the ANNs are learning physical rela-
 222 tionships to make their predictions.

223 To contextualize the predictions of the ANNs, we benchmark them against a sim-
 224 ple persistence model. The persistence model predicts that the SST anomaly will be un-
 225 changed so that the SST anomaly at the time of input remains the same at the time of
 226 prediction. We calculate the MAE for the persistence model and subtract it from the
 227 MAE of the ANNs ($\Delta\text{MAE} = \text{MAE}_{ANN} - \text{MAE}_{persistence}$), and plot the results in Fig-
 228 ure 2g and 2h. In regions where ΔMAE is negative, the ANN outperforms persistence
 229 (i.e. has lower error). These regions are illustrated in warm colors in Figure 2g and 2h
 230 and illustrates that the ANNs trained in this study out-perform persistence in all loca-
 231 tions and at both lead times. These regions were all found to be significant to $\alpha = 0.05$
 232 using a one-sided Wilcoxon signed-rank test. The greatest improvement in skill above
 233 persistence occurs in the cold tongue region of the Equatorial Pacific. This is unsurpris-
 234 ing as this region exhibits large interannual variability due to the El Nino Southern Os-
 235 cillation, and hence persistence performs poorly in this region. Also notable, the improve-
 236 ments over persistence do not necessarily align with grid points where the networks achieve
 237 lowest MAE. This is a fingerprint of regional decadal variability, that regions with longer
 238 memory (e.g. the mid-latitude North Atlantic) are better modeled by persistence, but
 239 in these cases our networks still out-perform persistence.

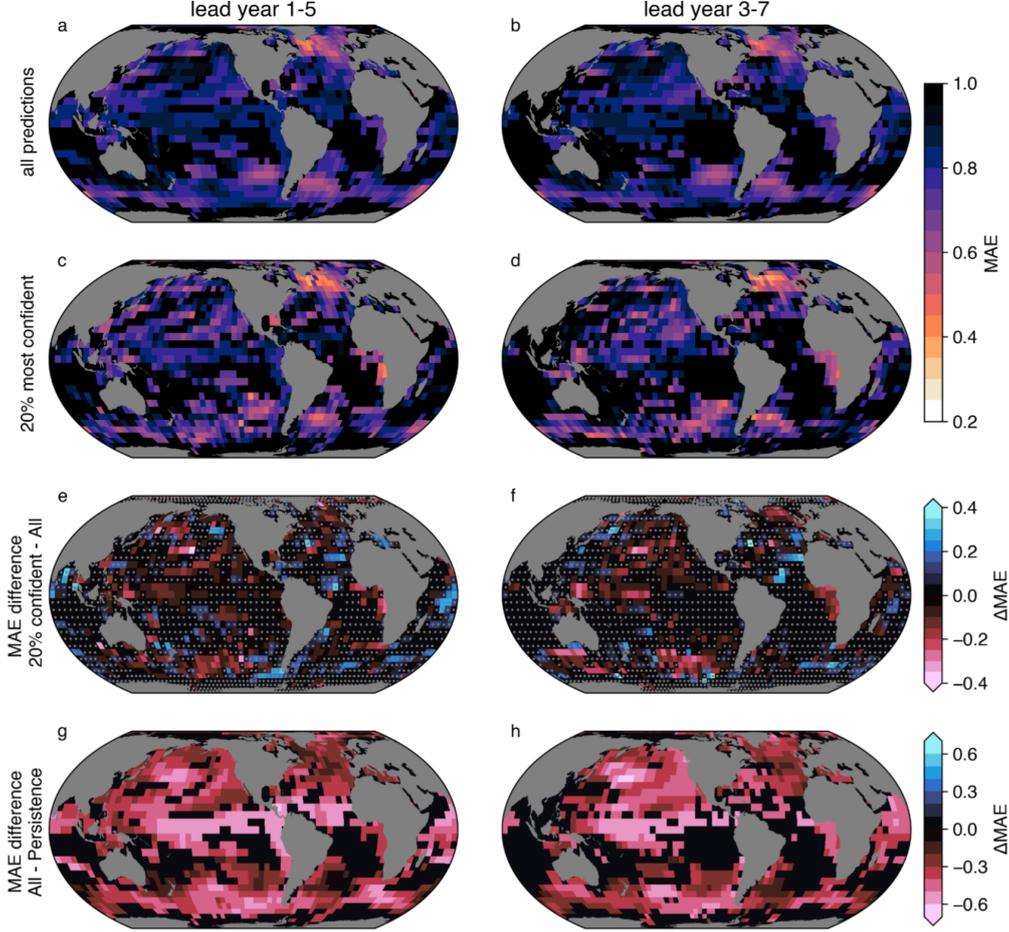


Figure 2. Evaluation of ANN prediction error. The left column is the prediction error for lead years 1-5, and the right column is for lead years 3-7. Panel a and panel b are mean absolute error (MAE) for all predictions in the testing set (i.e. all samples, $N=3400$). Panel c and panel d show MAE for 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 $MAE_{persistence}$ ($MAE_{ANN} - MAE_{persistence}$) in the testing set.

240

3.3 Identifying State-Dependent Predictability

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

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.

259

3.4 Investigating Skillful Decadal Predictions

260

261

262

263

264

265

266

267

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).

268

269

270

271

272

273

274

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

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

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

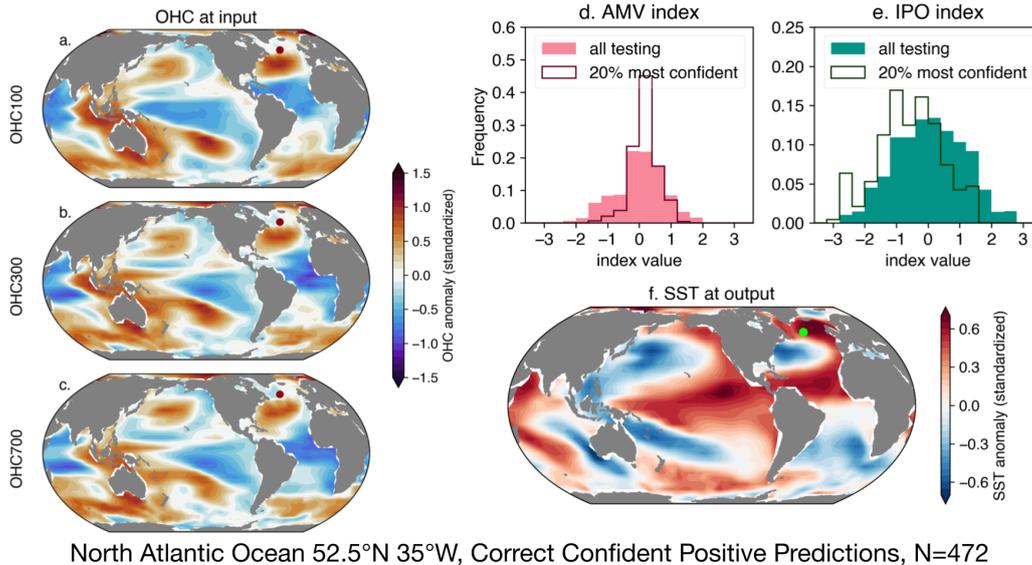


Figure 3. State-dependent predictability identified in the North Atlantic for predicting average SST anomaly at lead time 1-5 years. Panels a-c: Composite of OHC inputs for confident predictions of positive SST anomaly in a point in the North Atlantic (red dot). Panel d: histogram of AMV index for testing 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).

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

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

312 We now perform a similar analysis for an ANN trained to predict SST in 1-5 years
 313 at a point in the North Pacific (42.5°N, 175°E). In Figure 4 we show the results for the
 314 20% most confident negative predictions. For this region, 632 out of the 680 most con-
 315 fident samples were predictions of negative anomaly, implying the ANN designated neg-

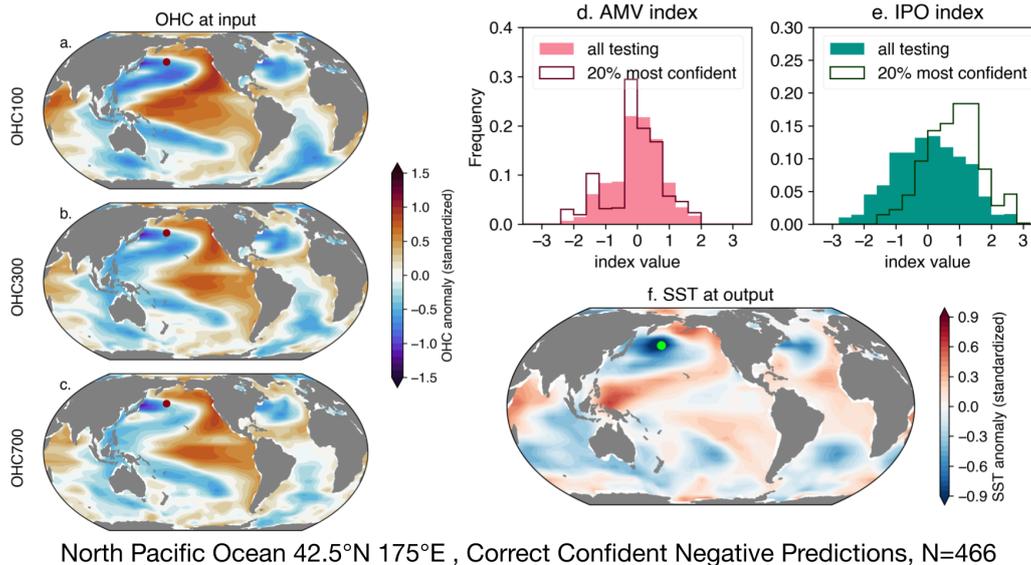


Figure 4. As Figure 3 but for the North Pacific

316 active predictions as more confident. Again we plot only the correct predictions, result-
 317 ing in 466 samples in these composites. Fig 4a-c shows the composite OHC inputs for
 318 confident negative predictions, and the major signal appears to be a positive IPO/PDV
 319 pattern in all panels. It is likely the ANN utilized this pattern to make these confident
 320 negative predictions from the ANN explainability heat-maps (see Text S1 and Figure S3).
 321 This is supported by the histogram of the IPO index in Fig 4e which shows the distri-
 322 bution of IPO phase in the confident samples is shifted such that confident samples sig-
 323 nificantly coincide with the positive phase of the IPO. There is no such strong signal in
 324 the AMV index (Fig 4d). Lastly, the confident predictions appear to relate to persistence
 325 in the positive IPO phase because the composite map of SST at output (Fig 4f) shows
 326 an IPO pattern in the Pacific Ocean. The largest SST anomalies are in the north Pa-
 327 cific mid-latitudes, in the traditional PDV region. From this, we posit that skillful pre-
 328 dictions of SST in the North Pacific are associated with persistence in the positive phase
 329 of IPO (i.e. negative SST anomaly at the predicted grid point). Here, the ANN prefer-
 330 entially identifies negative SST predictions as skillful, perhaps implying that persistence
 331 in the positive phase of IPO is more predictable than persistence of the negative phase.
 332 We posit that this difference in predictability is due to the underlying non-linear mech-
 333 anisms governing IPO dynamics and particularly the asymmetry in the dynamics gov-
 334 erning ENSO events (Choi et al., 2013; Okumura & Deser, 2010). Further investigation
 335 of this is an avenue for future work.

336 4 Discussion & Conclusion

337 We show that artificial neural networks (ANNs) skillfully predict SST evolution on
 338 decadal timescales and that they can objectively identify decadal state-dependent pre-
 339 dictability due to internal variability in the North Pacific and North Atlantic Oceans.
 340 Specifically, we use a regression neural network where the predictions take the form of
 341 a conditional normal distribution which we leverage to isolate predictions that are more
 342 likely to have lower error. This approach allows us to investigate possible contributing
 343 mechanisms to decadal SST predictability, particularly Atlantic multi-decadal variabil-
 344 ity and the interdecadal Pacific oscillation (AMV and IPO, Figs 3 and 4). We chose to
 345 model the conditional distributions as normal distributions as alternatives did not sig-

nificantly 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

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

399 References

- 400 Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X.
 401 (2016, November). Tensorflow: A system for large-scale machine learning.
 402 In *12th USENIX symposium on operating systems design and implementation*
 403 (*OSDI 16*) (pp. 265–283). Savannah, GA: USENIX Association. Retrieved
 404 from [https://www.usenix.org/conference/osdi16/technical-sessions/](https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi)
 405 [presentation/abadi](https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi)
- 406 Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., ...
 407 Kindermans, P.-J. (2019). Investigate neural networks! *Journal of Machine*
 408 *Learning Research*, 20(93), 1–8. Retrieved from [http://jmlr.org/papers/](http://jmlr.org/papers/v20/18-540.html)
 409 [v20/18-540.html](http://jmlr.org/papers/v20/18-540.html)
- 410 Albers, J. R., & Newman, M. (2019, November). A Priori identification of skill-
 411 ful extratropical subseasonal forecasts. *Geophys. Res. Lett.*, 46(21), 12527–
 412 12536. Retrieved from [https://onlinelibrary.wiley.com/doi/10.1029/](https://onlinelibrary.wiley.com/doi/10.1029/2019GL085270)
 413 [2019GL085270](https://onlinelibrary.wiley.com/doi/10.1029/2019GL085270) doi: 10.1029/2019gl085270
- 414 Alexander, M. A., & Deser, C. (1995, January). A Mechanism for the Recurrence
 415 of Wintertime Midlatitude SST Anomalies. *J. Phys. Oceanogr.*, 25(1), 122–
 416 137. Retrieved from [https://journals.ametsoc.org/view/journals/phoc/](https://journals.ametsoc.org/view/journals/phoc/25/1/1520-0485_1995_025_0122_amftro_2_0_co_2.xml?tab_body=fulltext)
 417 [25/1/1520-0485_1995_025_0122_amftro_2_0_co_2.xml?tab_body=fulltext](https://journals.ametsoc.org/view/journals/phoc/25/1/1520-0485_1995_025_0122_amftro_2_0_co_2.xml?tab_body=fulltext)
 418 [-display](https://journals.ametsoc.org/view/journals/phoc/25/1/1520-0485_1995_025_0122_amftro_2_0_co_2.xml?tab_body=fulltext) doi: 10.1175/1520-0485(1995)025<0122:AMFTRO>2.0.CO;2
- 419 Barnes, E. A., & Barnes, R. J. (2021, December). Controlled abstention neural
 420 networks for identifying skillful predictions for regression problems. *J. Adv.*
 421 *Model. Earth Syst.*, 13(12). Retrieved from [https://onlinelibrary.wiley](https://onlinelibrary.wiley.com/doi/10.1029/2021MS002575)
 422 [.com/doi/10.1029/2021MS002575](https://onlinelibrary.wiley.com/doi/10.1029/2021MS002575) doi: 10.1029/2021ms002575
- 423 Barnes, E. A., Barnes, R. J., & Gordillo, N. (2021, September). *Adding Uncer-*
 424 *tainty to Neural Network Regression Tasks in the Geosciences*. Retrieved from
 425 <http://arxiv.org/abs/2109.07250>
- 426 Boer, G. J. (2011, March). Decadal potential predictability of twenty-first century
 427 climate. *Clim. Dyn.*, 36(5), 1119–1133. Retrieved from [https://doi.org/10](https://doi.org/10.1007/s00382-010-0747-9)
 428 [.1007/s00382-010-0747-9](https://doi.org/10.1007/s00382-010-0747-9) doi: 10.1007/s00382-010-0747-9
- 429 Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., & Bellouin, N.
 430 (2012, April). Aerosols implicated as a prime driver of twentieth-century North
 431 Atlantic climate variability. *Nature*, 484(7393), 228–232. Retrieved from
 432 <http://dx.doi.org/10.1038/nature10946> doi: 10.1038/nature10946
- 433 Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., &
 434 Mignot, J. (2021, February). Improved decadal predictions of north Atlantic
 435 subpolar gyre SST in CMIP6. *Geophys. Res. Lett.*, 48(3). Retrieved from
 436 <https://onlinelibrary.wiley.com/doi/10.1029/2020GL091307> doi:
 437 [10.1029/2020gl091307](https://onlinelibrary.wiley.com/doi/10.1029/2020GL091307)
- 438 Borchert, L. F., Müller, W. A., & Baehr, J. (2018, September). Atlantic Ocean
 439 Heat Transport Influences Interannual-to-Decadal Surface Temperature Pre-
 440 dictability in the North Atlantic Region. *J. Clim.*, 31(17), 6763–6782. Re-
 441 trieved from [https://journals.ametsoc.org/view/journals/clim/31/17/](https://journals.ametsoc.org/view/journals/clim/31/17/jcli-d-17-0734.1.xml)
 442 [jcli-d-17-0734.1.xml](https://journals.ametsoc.org/view/journals/clim/31/17/jcli-d-17-0734.1.xml) doi: 10.1175/JCLI-D-17-0734.1
- 443 Branstator, G., & Teng, H. (2010, December). Two Limits of Initial-Value
 444 Decadal Predictability in a CGCM. *J. Clim.*, 23(23), 6292–6311. Re-
 445 trieved from [https://journals.ametsoc.org/view/journals/clim/23/](https://journals.ametsoc.org/view/journals/clim/23/23/2010jcli3678.1.xml)
 446 [23/2010jcli3678.1.xml](https://journals.ametsoc.org/view/journals/clim/23/23/2010jcli3678.1.xml) doi: 10.1175/2010JCLI3678.1

- 447 Branstator, G., Teng, H., Meehl, G. A., Kimoto, M., Knight, J. R., Latif, M., &
 448 Rosati, A. (2012, March). Systematic Estimates of Initial-Value Decadal
 449 Predictability for Six AOGCMs. *J. Clim.*, *25*(6), 1827–1846. Retrieved
 450 from [https://journals.ametsoc.org/view/journals/clim/25/6/
 451 jcli-d-11-00227.1.xml](https://journals.ametsoc.org/view/journals/clim/25/6/jcli-d-11-00227.1.xml) doi: 10.1175/JCLI-D-11-00227.1
- 452 Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J. (2018, Septem-
 453 ber). Time dependency of the prediction skill for the North Atlantic sub-
 454 polar gyre in initialized decadal hindcasts. *Clim. Dyn.*, *51*(5), 1947–1970.
 455 Retrieved from <https://doi.org/10.1007/s00382-017-3991-4> doi:
 456 10.1007/s00382-017-3991-4
- 457 Choi, K.-Y., Vecchi, G. A., & Wittenberg, A. T. (2013, December). ENSO Tran-
 458 sition, Duration, and Amplitude Asymmetries: Role of the Nonlinear Wind
 459 Stress Coupling in a Conceptual Model. *J. Clim.*, *26*(23), 9462–9476. Re-
 460 trieved from [https://journals.ametsoc.org/view/journals/clim/26/23/
 461 jcli-d-13-00045.1.xml](https://journals.ametsoc.org/view/journals/clim/26/23/jcli-d-13-00045.1.xml) doi: 10.1175/JCLI-D-13-00045.1
- 462 Christensen, H. M., Berner, J., & Yeager, S. (2020, September). The Value of
 463 Initialization on Decadal Timescales: State-Dependent Predictability in the
 464 CESM Decadal Prediction Large Ensemble. *J. Clim.*, *33*(17), 7353–7370. Re-
 465 trieved from [https://journals.ametsoc.org/jcli/article/33/17/7353/
 466 348619/The-Value-of-Initialization-on-Decadal-Timescales](https://journals.ametsoc.org/jcli/article/33/17/7353/348619/The-Value-of-Initialization-on-Decadal-Timescales) doi:
 467 10.1175/JCLI-D-19-0571.1
- 468 Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Rädel, G.,
 469 & Stevens, B. (2015, October). The Atlantic Multidecadal Oscillation with-
 470 out a role for ocean circulation. *Science*, *350*(6258), 320–324. Retrieved
 471 from <https://science.sciencemag.org/content/350/6258/320> doi:
 472 10.1126/science.aab3980
- 473 Danabasoglu, G. (2019). *Ncar cesm2 model output prepared for cmip6 cmip his-*
 474 *torical*. Earth System Grid Federation. Retrieved from [https://doi.org/10
 475 .22033/ESGF/CMIP6.7627](https://doi.org/10.22033/ESGF/CMIP6.7627) doi: 10.22033/ESGF/CMIP6.7627
- 476 Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K.,
 477 Edwards, J., ... Strand, W. G. (2020). The Community Earth System Model
 478 Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2),
 479 e2019MS001916. Retrieved from [https://agupubs.onlinelibrary.wiley
 480 .com/doi/abs/10.1029/2019MS001916](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001916) doi: 10.1029/2019MS001916
- 481 Enfield, D. B., Mestas-Nuñez, A. M., & Trimble, P. J. (2001). The Atlantic Mul-
 482 tidecadal Oscillation and its relation to rainfall and river flows in the conti-
 483 nental U.S. *Geophys. Res. Lett.*, *28*(10), 2077–2080. Retrieved from [http://
 484 agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL012745](http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL012745) doi:
 485 10.1029/2000GL012745
- 486 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., &
 487 Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project
 488 Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model
 489 Development*, *9*(5), 1937–1958. Retrieved from [https://gmd.copernicus
 490 .org/articles/9/1937/2016/](https://gmd.copernicus.org/articles/9/1937/2016/) doi: 10.5194/gmd-9-1937-2016
- 491 Fang, S.-W., Khodri, M., Timmreck, C., Zanchettin, D., & Jungclaus, J. (2021,
 492 December). Disentangling internal and external contributions to Atlantic
 493 multidecadal variability over the past millennium. *Geophys. Res. Lett.*,
 494 *48*(23). Retrieved from [https://onlinelibrary.wiley.com/doi/10.1029/
 495 2021GL095990](https://onlinelibrary.wiley.com/doi/10.1029/2021GL095990) doi: 10.1029/2021gl095990
- 496 Fenske, T., & Clement, A. (2022, February). No internal connections detected be-
 497 tween low frequency climate modes in north Atlantic and north Pacific basins.
 498 *Geophys. Res. Lett.* Retrieved from [https://onlinelibrary.wiley.com/
 499 doi/10.1029/2022GL097957](https://onlinelibrary.wiley.com/doi/10.1029/2022GL097957) doi: 10.1029/2022gl097957
- 500 Foster, D., Gagne, D. J., II, & Whitt, D. B. (2021, December). Probabilistic ma-
 501 chine learning estimation of ocean mixed layer depth from dense satellite and

- 502 sparse in situ observations. *J. Adv. Model. Earth Syst.*, 13(12). Retrieved
 503 from <https://onlinelibrary.wiley.com/doi/10.1029/2021MS002474> doi:
 504 10.1029/2021ms002474
- 505 Frankignoul, C., & Hasselmann, K. (1977). Stochastic climate models, Part II
 506 Application to sea-surface temperature anomalies and thermocline vari-
 507 ability. *Tell'Us*, 29(4), 289–305. Retrieved from <http://onlinelibrary>
 508 [.wiley.com/doi/abs/10.1111/j.2153-3490.1977.tb00740.x](http://onlinelibrary.wiley.com/doi/abs/10.1111/j.2153-3490.1977.tb00740.x) doi:
 509 10.1111/j.2153-3490.1977.tb00740.x
- 510 Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger,
 511 M., ... Stevens, B. (2013, July). Climate and carbon cycle changes from
 512 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison
 513 Project phase 5. *J. Adv. Model. Earth Syst.*, 5(3), 572–597. Retrieved from
 514 <http://doi.wiley.com/10.1002/jame.20038> doi: 10.1002/jame.20038
- 515 Gordon, E. M. (2022, June). *emily-gordy/Decadal-SST-prediction: Decadal SST pre-*
 516 *diction, revised GRL submission*. Zenodo. Retrieved from [https://doi.org/](https://doi.org/10.5281/zenodo.6646950)
 517 [10.5281/zenodo.6646950](https://doi.org/10.5281/zenodo.6646950) doi: 10.5281/zenodo.6646950
- 518 Gordon, E. M., Barnes, E. A., & Hurrell, J. W. (2021, November). Oceanic
 519 harbingers of Pacific decadal oscillation predictability in CESM2 de-
 520 tected by neural networks. *Geophys. Res. Lett.*, 48(21). Retrieved from
 521 <https://onlinelibrary.wiley.com/doi/10.1029/2021GL095392> doi:
 522 10.1029/2021gl095392
- 523 Guillaumin, A. P., & Zanna, L. (2021, September). Stochastic-deep learning pa-
 524 rameterization of ocean momentum forcing. *J. Adv. Model. Earth Syst.*,
 525 13(9). Retrieved from [https://onlinelibrary.wiley.com/doi/10.1029/](https://onlinelibrary.wiley.com/doi/10.1029/2021MS002534)
 526 [2021MS002534](https://onlinelibrary.wiley.com/doi/10.1029/2021MS002534) doi: 10.1029/2021ms002534
- 527 Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K.
 528 (2015, December). A Tripole Index for the Interdecadal Pacific Oscillation.
 529 *Clim. Dyn.*, 45(11), 3077–3090. Retrieved from [https://doi.org/10.1007/](https://doi.org/10.1007/s00382-015-2525-1)
 530 [s00382-015-2525-1](https://doi.org/10.1007/s00382-015-2525-1) doi: 10.1007/s00382-015-2525-1
- 531 Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J.,
 532 ... Marshall, S. (2013, September). The Community Earth System Model:
 533 A Framework for Collaborative Research. *Bull. Am. Meteorol. Soc.*, 94(9),
 534 1339–1360. Retrieved from [https://journals.ametsoc.org/view/journals/](https://journals.ametsoc.org/view/journals/bams/94/9/bams-d-12-00121.1.xml)
 535 [bams/94/9/bams-d-12-00121.1.xml](https://journals.ametsoc.org/view/journals/bams/94/9/bams-d-12-00121.1.xml) doi: 10.1175/BAMS-D-12-00121.1
- 536 Labe, Z. M., & Barnes, E. A. (2022, February). *Predicting slowdowns in*
 537 *decadal climate warming trends with explainable neural networks*. Retrieved
 538 from <http://www.essoar.org/doi/10.1002/essoar.10508874.2> doi:
 539 10.1002/essoar.10508874.2
- 540 Mann, M. E., Steinman, B. A., Brouillette, D. J., & Miller, S. K. (2021, March).
 541 Multidecadal climate oscillations during the past millennium driven by vol-
 542 canic forcing. *Science*, 371(6533), 1014–1019. Retrieved from [http://](http://dx.doi.org/10.1126/science.abc5810)
 543 dx.doi.org/10.1126/science.abc5810 doi: 10.1126/science.abc5810
- 544 Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997,
 545 June). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon
 546 Production. *Bull. Am. Meteorol. Soc.*, 78(6), 1069–1080. Retrieved
 547 from [http://journals.ametsoc.org/bams/article/78/6/1069/55942/](http://journals.ametsoc.org/bams/article/78/6/1069/55942/A-Pacific-Interdecadal-Climate-Oscillation-with)
 548 [A-Pacific-Interdecadal-Climate-Oscillation-with](http://journals.ametsoc.org/bams/article/78/6/1069/55942/A-Pacific-Interdecadal-Climate-Oscillation-with) doi: 10.1175/
 549 [1520-0477\(1997\)078<1069:APICOW>2.0.CO;2](http://journals.ametsoc.org/bams/article/78/6/1069/55942/A-Pacific-Interdecadal-Climate-Oscillation-with)
- 550 Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C.,
 551 ... Albers, J. (2020, May). Windows of Opportunity for Skillful Forecasts
 552 Subseasonal to Seasonal and Beyond. *Bull. Am. Meteorol. Soc.*, 101(5), E608–
 553 E625. Retrieved from [http://journals.ametsoc.org/bams/article/101/](http://journals.ametsoc.org/bams/article/101/5/E608/345558/Windows-of-Opportunity-for-Skillful-Forecasts)
 554 [5/E608/345558/Windows-of-Opportunity-for-Skillful-Forecasts](http://journals.ametsoc.org/bams/article/101/5/E608/345558/Windows-of-Opportunity-for-Skillful-Forecasts) doi:
 555 [10.1175/BAMS-D-18-0326.1](http://journals.ametsoc.org/bams/article/101/5/E608/345558/Windows-of-Opportunity-for-Skillful-Forecasts)
- 556 Mayer, K. J., & Barnes, E. A. (2021). Subseasonal Forecasts of Opportunity

- 557 Identified by an Explainable Neural Network. *Geophys. Res. Lett.*, 48(10),
 558 e2020GL092092. Retrieved from [http://agupubs.onlinelibrary.wiley](http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL092092)
 559 [.com/doi/abs/10.1029/2020GL092092](http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL092092) doi: 10.1029/2020GL092092
- 560 Meehl, G. A., Hu, A., Arblaster, J. M., Fasullo, J., & Trenberth, K. E. (2013,
 561 September). Externally Forced and Internally Generated Decadal Climate
 562 Variability Associated with the Interdecadal Pacific Oscillation. *J. Clim.*,
 563 26(18), 7298–7310. Retrieved from [https://journals.ametsoc.org/](https://journals.ametsoc.org/view/journals/clim/26/18/jcli-d-12-00548.1.xml)
 564 [view/journals/clim/26/18/jcli-d-12-00548.1.xml](https://journals.ametsoc.org/view/journals/clim/26/18/jcli-d-12-00548.1.xml) doi: 10.1175/
 565 JCLI-D-12-00548.1
- 566 Meehl, G. A., Hu, A., & Teng, H. (2016, June). Initialized decadal prediction
 567 for transition to positive phase of the Interdecadal Pacific Oscillation. *Nat.*
 568 *Commun.*, 7(1), 11718. Retrieved from [http://www.nature.com/articles/](http://www.nature.com/articles/ncomms11718)
 569 [ncomms11718](http://www.nature.com/articles/ncomms11718) doi: 10.1038/ncomms11718
- 570 Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes,
 571 F., ... Xie, S.-P. (2021, April). Initialized Earth System prediction from
 572 subseasonal to decadal timescales. *Nature Reviews Earth & Environment*,
 573 2(5), 340–357. Retrieved from [https://www.nature.com/articles/](https://www.nature.com/articles/s43017-021-00155-x)
 574 [s43017-021-00155-x](https://www.nature.com/articles/s43017-021-00155-x) doi: 10.1038/s43017-021-00155-x
- 575 Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho,
 576 C. A. S., ... Yeager, S. (2020, June). Current and Emerging Developments in
 577 Subseasonal to Decadal Prediction. *Bull. Am. Meteorol. Soc.*, 101(6), E869–
 578 E896. Retrieved from [https://journals.ametsoc.org/view/journals/](https://journals.ametsoc.org/view/journals/bams/101/6/bamsD190037.xml)
 579 [bams/101/6/bamsD190037.xml](https://journals.ametsoc.org/view/journals/bams/101/6/bamsD190037.xml) doi: 10.1175/BAMS-D-19-0037.1
- 580 Msadek, R., Dixon, K. W., Delworth, T. L., & Hurlin, W. (2010, October). As-
 581 sessing the predictability of the Atlantic meridional overturning circulation
 582 and associated fingerprints. *Geophys. Res. Lett.*, 37(19). Retrieved from
 583 <http://doi.wiley.com/10.1029/2010GL044517> doi: 10.1029/2010gl044517
- 584 Okumura, Y. M., & Deser, C. (2010, November). Asymmetry in the Duration of
 585 El Niño and La Niña. *J. Clim.*, 23(21), 5826–5843. Retrieved from [https://](https://journals.ametsoc.org/view/journals/clim/23/21/2010jcli3592.1.xml)
 586 journals.ametsoc.org/view/journals/clim/23/21/2010jcli3592.1.xml
 587 doi: 10.1175/2010JCLI3592.1
- 588 Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., & Tschuck, P. (2004,
 589 November). Estimating the Decadal Predictability of a Coupled AOGCM. *J.*
 590 *Clim.*, 17(22), 4463–4472. Retrieved from [https://journals.ametsoc.org/](https://journals.ametsoc.org/view/journals/clim/17/22/3209.1.xml)
 591 [view/journals/clim/17/22/3209.1.xml](https://journals.ametsoc.org/view/journals/clim/17/22/3209.1.xml) doi: 10.1175/3209.1
- 592 Power, S., Casey, T., Folland, C., Colman, A., & Mehta, V. (1999, May). Inter-
 593 decadal modulation of the impact of ENSO on Australia. *Clim. Dyn.*, 15(5),
 594 319–324. Retrieved from <https://doi.org/10.1007/s003820050284> doi: 10
 595 [.1007/s003820050284](https://doi.org/10.1007/s003820050284)
- 596 Schulzweida, U. (2019, October). *Cdo user guide*. Retrieved from [https://doi.org/](https://doi.org/10.5281/zenodo.3539275)
 597 [10.5281/zenodo.3539275](https://doi.org/10.5281/zenodo.3539275) doi: 10.5281/zenodo.3539275
- 598 Simpson, I. R., Deser, C., McKinnon, K. A., & Barnes, E. A. (2018, October). Mod-
 599 eled and Observed Multidecadal Variability in the North Atlantic Jet Stream
 600 and Its Connection to Sea Surface Temperatures. *J. Clim.*, 31(20), 8313–8338.
 601 Retrieved from [https://journals.ametsoc.org/view/journals/clim/31/](https://journals.ametsoc.org/view/journals/clim/31/20/jcli-d-18-0168.1.xml)
 602 [20/jcli-d-18-0168.1.xml](https://journals.ametsoc.org/view/journals/clim/31/20/jcli-d-18-0168.1.xml) doi: 10.1175/JCLI-D-18-0168.1
- 603 Simpson, I. R., Yeager, S. G., McKinnon, K. A., & Deser, C. (2019, August).
 604 Decadal predictability of late winter precipitation in western Europe through
 605 an ocean–jet stream connection. *Nat. Geosci.*, 12(8), 613–619. Retrieved
 606 from <https://www.nature.com/articles/s41561-019-0391-x> doi:
 607 [10.1038/s41561-019-0391-x](https://www.nature.com/articles/s41561-019-0391-x)
- 608 Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically Interpretable
 609 Neural Networks for the Geosciences: Applications to Earth System Variabil-
 610 ity. *Journal of Advances in Modeling Earth Systems*, 12(9), e2019MS002002.
 611 Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/>

- 612 10.1029/2019MS002002 doi: 10.1029/2019MS002002
 613 Toms, B. A., Barnes, E. A., & Hurrell, J. W. (2021, June). Assessing decadal pre-
 614 dictability in an earth–system model using explainable neural networks. *Geo-*
 615 *phys. Res. Lett.*, *48*(12). Retrieved from [https://onlinelibrary.wiley.com/](https://onlinelibrary.wiley.com/doi/10.1029/2021GL093842)
 616 [doi/10.1029/2021gl093842](https://onlinelibrary.wiley.com/doi/10.1029/2021gl093842) doi: 10.1029/2021gl093842
 617 Trenberth, K. E., & Fasullo, J. T. (2013, December). An apparent hia-
 618 tus in global warming? *Earths Future*, *1*(1), 19–32. Retrieved from
 619 <https://onlinelibrary.wiley.com/doi/10.1002/2013EF000165> doi:
 620 10.1002/2013ef000165
 621 Trenberth, K. E., & Shea, D. J. (2006). Atlantic hurricanes and natural variability
 622 in 2005. *Geophys. Res. Lett.*, *33*(12). Retrieved from [http://doi.wiley.com/](http://doi.wiley.com/10.1029/2006GL026894)
 623 [10.1029/2006GL026894](http://doi.wiley.com/10.1029/2006GL026894) doi: 10.1029/2006gl026894
 624 van der Velden, E. (2020, February). CMasher: Scientific colormaps for making ac-
 625 cessible, informative and 'cmashing' plots. *The Journal of Open Source Soft-*
 626 *ware*, *5*(46), 2004. doi: 10.21105/joss.02004
 627 Wu, S., Liu, Z., Zhang, R., & Delworth, T. L. (2011, February). On the observed
 628 relationship between the Pacific Decadal Oscillation and the Atlantic Multi-
 629 decadal Oscillation. *J. Oceanogr.*, *67*(1), 27–35. Retrieved from [https://](https://doi.org/10.1007/s10872-011-0003-x)
 630 doi.org/10.1007/s10872-011-0003-x doi: 10.1007/s10872-011-0003-x
 631 Xie, S.-P., & Tanimoto, Y. (1998). A pan-Atlantic decadal climate oscillation.
 632 *Geophys. Res. Lett.*, *25*(12), 2185–2188. Retrieved from [http://](http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98GL01525)
 633 agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98GL01525 doi:
 634 10.1029/98GL01525
 635 Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C.,
 636 Meehl, G. A., ... Lovenduski, N. S. (2018, September). Predicting Near-Term
 637 Changes in the Earth System: A Large Ensemble of Initialized Decadal Predic-
 638 tion Simulations Using the Community Earth System Model. *Bull. Am. Meteo-*
 639 *rol. Soc.*, *99*(9), 1867–1886. Retrieved from [https://journals.ametsoc.org/](https://journals.ametsoc.org/bams/article/99/9/1867/70398/Predicting-Near-Term-Changes-in-the-Earth-System-A)
 640 [bams/article/99/9/1867/70398/Predicting-Near-Term-Changes-in-the](https://journals.ametsoc.org/bams/article/99/9/1867/70398/Predicting-Near-Term-Changes-in-the-Earth-System-A)
 641 [-Earth-System-A](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
 642 Zhang, R., & Delworth, T. L. (2006). Impact of Atlantic multidecadal oscillations on
 643 India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, *33*(17). Re-
 644 trieved from <http://doi.wiley.com/10.1029/2006GL026267> doi: 10.1029/
 645 2006gl026267
 646 Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G.,
 647 ... Little, C. M. (2019, June). A review of the role of the Atlantic merid-
 648 ional overturning circulation in Atlantic multidecadal variability and asso-
 649 ciated climate impacts. *Rev. Geophys.*, *57*(2), 316–375. Retrieved from
 650 <https://onlinelibrary.wiley.com/doi/10.1029/2019RG000644> doi:
 651 10.1029/2019rg000644
 652 Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997, May). ENSO-like Inter-
 653 decadal Variability: 1900–93. *J. Clim.*, *10*(5), 1004–1020. Retrieved from
 654 [https://journals.ametsoc.org/view/journals/clim/10/5/1520-0442](https://journals.ametsoc.org/view/journals/clim/10/5/1520-0442_1997_010_1004_eliv_2.0.co_2.xml)
 655 [_1997_010_1004_eliv_2.0.co_2.xml](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:
 656 ELIV>2.0.CO;2

Supporting Information for ”Incorporating Uncertainty into a Regression Neural Network Enables Identification of Decadal State-Dependent Predictability”

Emily M. Gordon¹, Elizabeth A. Barnes¹

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

Contents of this file

1. Text S1: IPO Index Calculation
2. Text S2: Neural Network Explainability
3. Figure S1: AMV and IPO index patterns
4. Figure S2: North Atlantic analysis in the training and validation data
5. Figure S3: North Pacific analysis in the training and validation data
6. Figure S4: Explainability maps for North Atlantic predictions
7. Figure S5: Explainability maps for North Pacific predictions

Introduction The text in this document (Text S1) is a description of explainable AI (XAI), and provides a discussion of XAI findings which support the conclusions in the main text. This text references Figures S2 and S3 which are the XAI analyses of Figures

3 and 4 from the main document, respectively. In Figure 1 we provide plots showing the Atlantic multi-decadal variability (AMV) and interdecadal Pacific oscillation (IPO) patterns calculated in the CESM2 long control run.

IPO Index Calculation We calculate the IPO index using the method outlined by Henley et al. (2015) and we detail here. From the deseasoned SST data we calculate the area averaged monthly SST anomalies in three boxes in the Pacific Ocean:

1. 25°N to 45°N and 140°E to 145°W
2. 10°S to 10°N and 170°E to 90°W
3. 50°S to 15°S and 150°E to 160°W

Using the numbering above, the index is calculated from the following equation:

$$\text{IPO} = \text{Box2} - 0.5 * (\text{Box1} + \text{Box3}) \quad (1)$$

The resulting pattern from projecting the IPO index onto global SSTs is plotted in Figure S1, with the boxes in these calculations outlined in purple.

Neural Network Explainability To support our results, we use neural network explainability techniques (explainable AI or XAI) to examine the decision-making process of the ANNs. The underlying goal of the XAI methods used here is to provide an indication of how each input pixel contributed to a neural network’s prediction. The methods we use here are attribution methods, in particular we use three methods, Integrated Gradient, LRP-Z (which is the same as Input times Gradient for networks with ReLU activation) and LRP-epsilon. All of these methods assign each input pixel a relevance, where positive relevance indicates that a pixel contributed to positively to an output node of interest and vice versa. For comprehensive discussion of XAI with application to climate science, and

best practices, see Mamalakis, Ebert-Uphoff, and Barnes (2021) and Mamalakis, Barnes, and Ebert-Uphoff (2022).

The explainability composite maps for each region investigated in the main text is provided in Figure S4-S5. Each of the first three columns is a different method (Gradient, Input times Gradient, LRP-epsilon from left to right). We use an epsilon value of 0.01, and apply Gaussian smoothing to each explainability map to assist with visualization. Each row is a different OHC level (OHC to 100 m, OHC to 300 m, OHC to 700 m from top to bottom). The right-most column in each is the composite OHC input which acts as a reference to how the relevance patterns correspond to the physical input maps.

In Figure S4 we look at the composite explainability maps for confident predictions of positive SST anomaly in the North Atlantic ocean (green dot, same as in Figure 3 in the main text). For all three methods, red regions contributed to the neural network's positive prediction. It appears the positive OHC anomaly in the North Atlantic Ocean contributed to the positive SST prediction, especially at the lowest level of the ocean (OHC to 700m). All XAI methods show the same patterns, reducing the likelihood for spurious relevance (although not eliminating it, see (Mamalakis et al., 2021)).

In Figure S5 we look at composite explainability maps for confident predictions of negative SST anomaly in the North Pacific ocean (green dot, same as Figure 4 in the main text). Here, the blue regions imply regions that contributed to neural network's negative prediction. Here, relevance highlights that the negative anomaly in the Kuroshio region in the upper layers, coupled with the positive anomaly in the off equatorial Pacific in lowest layers most contributed to the negative prediction. This anomaly pattern is

indicative of the IPO's positive phase. Again the highlighted relevances are consistent across explainability methods.

References

- 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
- Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022, February). Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *ArXiv*. Retrieved from <http://arxiv.org/abs/2202.03407>
- Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2021, March). Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *ArXiv*. Retrieved from <http://arxiv.org/abs/2103.10005>

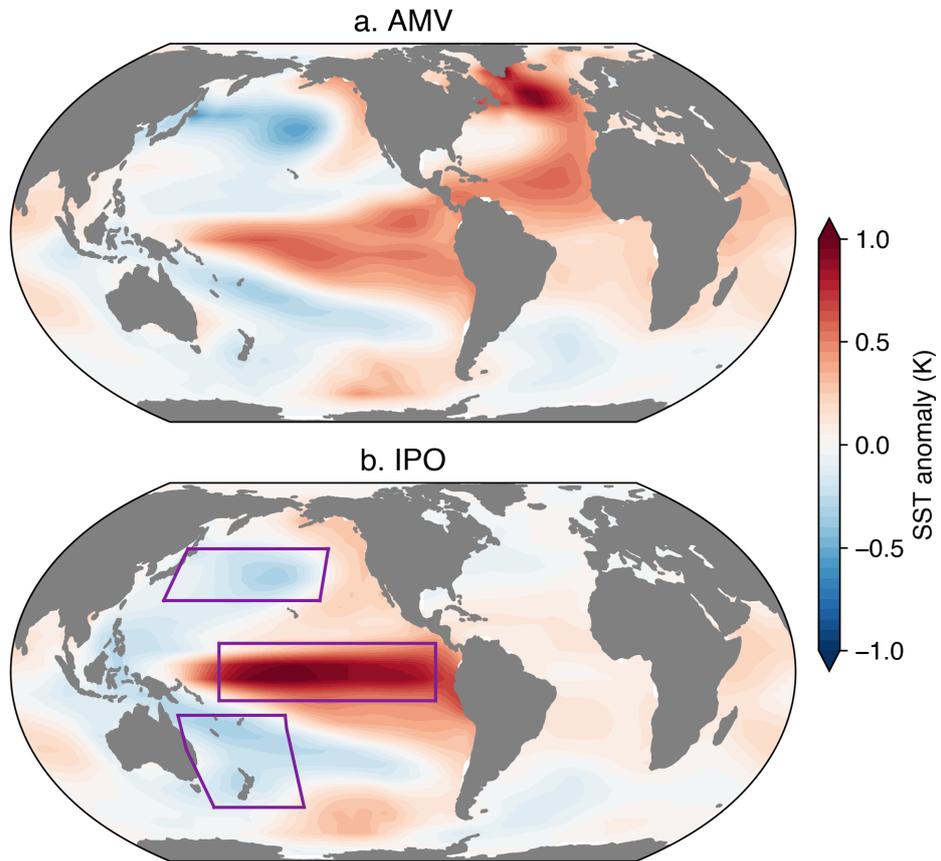


Figure S1. Patterns of large scale SST variability in CESM2 calculated using the methods discussed in Section 2.3 in Main a. AMV index projected onto global SSTs. b. IPO index projected onto global SSTs.

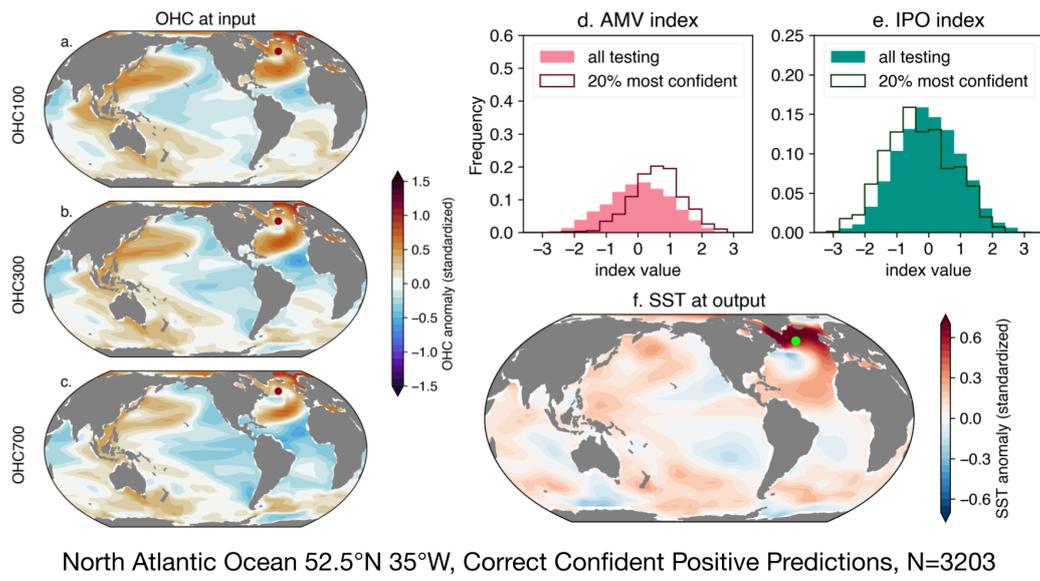


Figure S2. As Figure 3 in the main document but for the training and validation data.

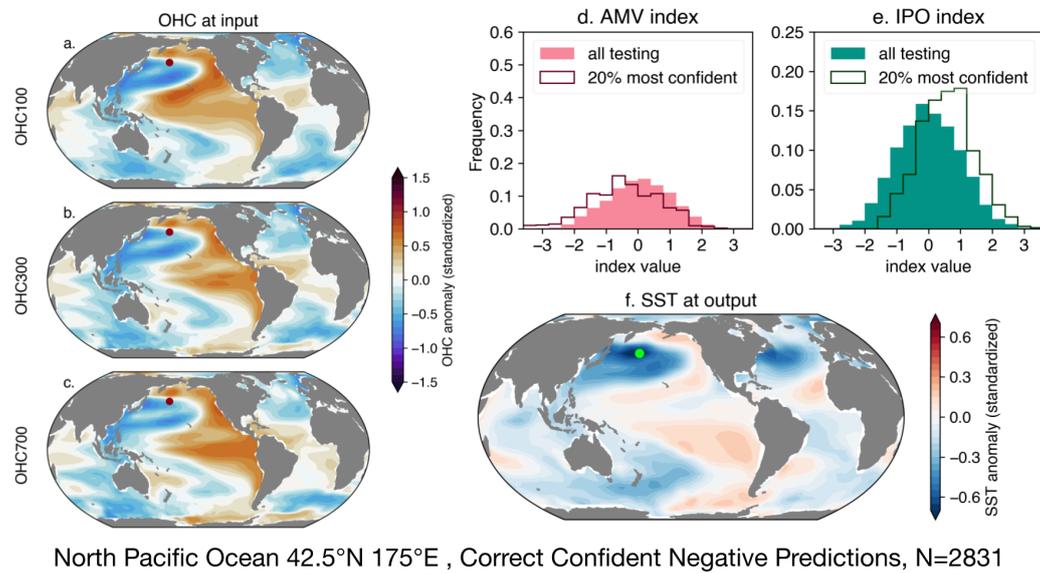


Figure S3. As Figure 4 in the main document but for the training and validation data.

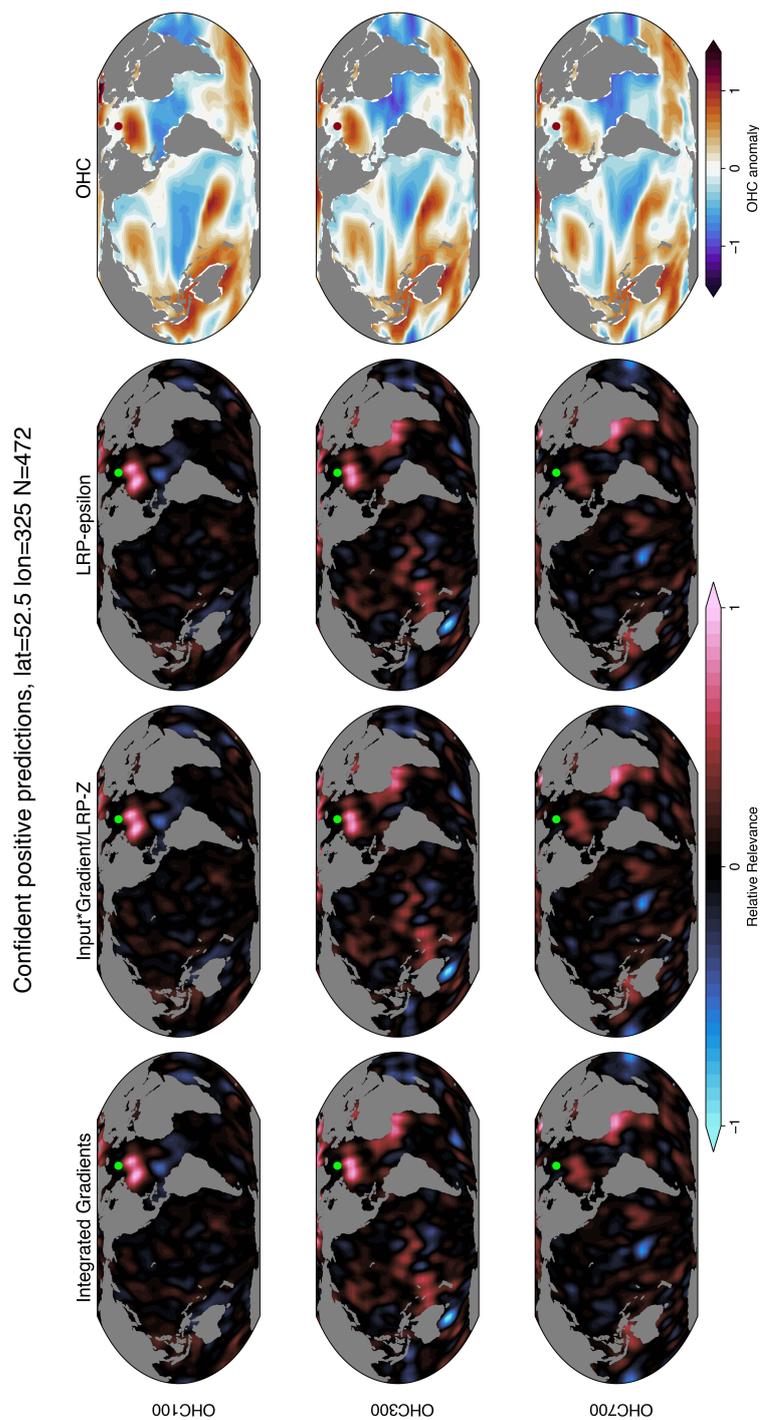


Figure S4. Composite explainability maps for predictions in Figure 3 of the main text. Each of the first three columns is a different technique (Integrated Gradients, Input times Gradient, LRP-epsilon from left to right), while each row is a different ocean layer (OHC to 100 m, OHC to 300 m, OHC to 700 m from top to bottom). The right-most column is the composite OHC input (the same as Fig 3a-c).

June 15, 2022, 4:49pm

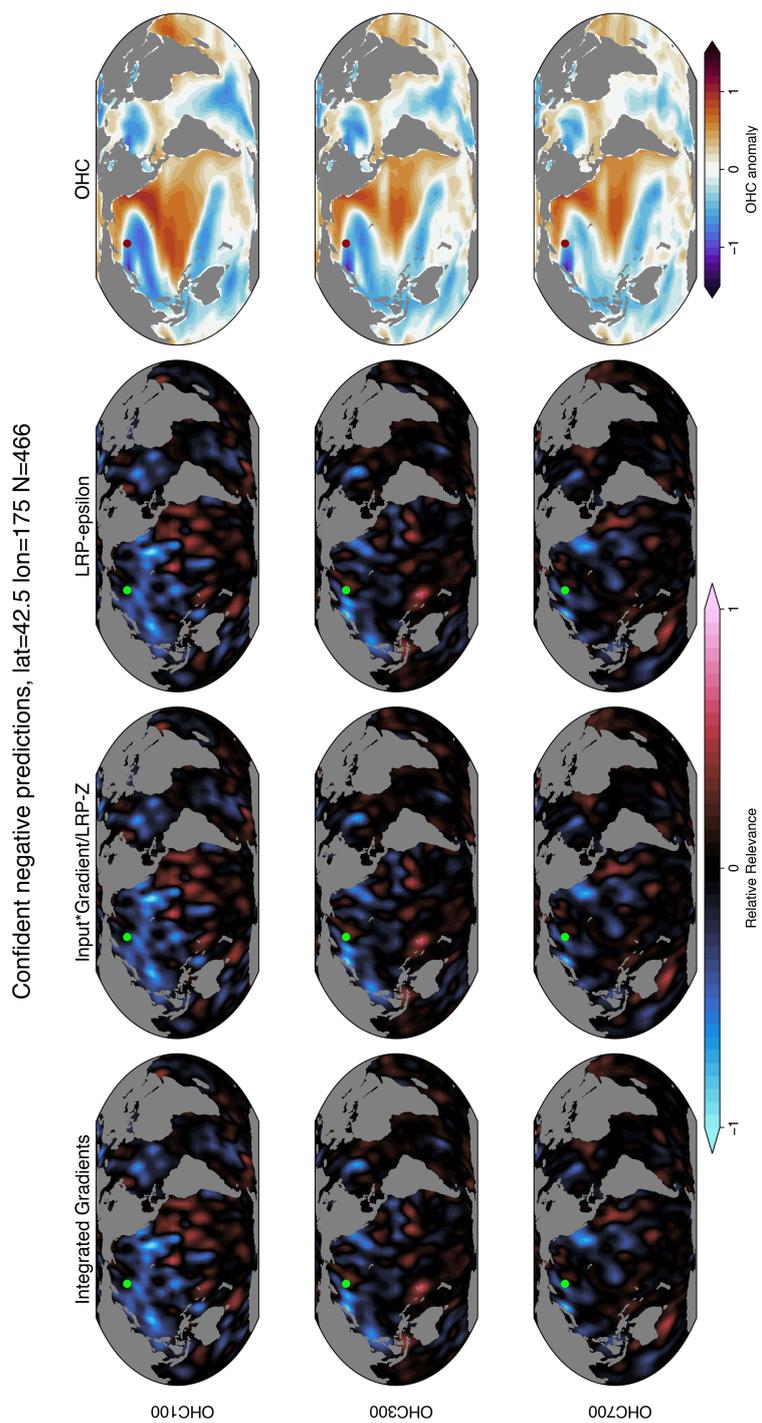


Figure S5. As Figure S4 but for North Pacific predictions in Figure 4 in the main text.