DISSERTATION

BRIDGING HUMAN AND ARTIFICIAL INTELLIGENCE FOR SKILLFUL, TRUSTWORTHY, AND INSIGHTFUL SEASONAL-TO-DECADAL CLIMATE PREDICTION

Submitted by

Jamin K. Rader

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2024

Doctoral Committee:

Advisor: Elizabeth A. Barnes

Kristen L. Rasmussen
James W. Hurrell
Camille S. Stevens-Rumann

Copyright by Jamin K. Rader 2024

All Rights Reserved

ABSTRACT

BRIDGING HUMAN AND ARTIFICIAL INTELLIGENCE FOR SKILLFUL, TRUSTWORTHY, AND INSIGHTFUL SEASONAL-TO-DECADAL CLIMATE PREDICTION

Seasonal-to-decadal climate variability is inherently difficult to predict and is intimately connected to human and natural systems worldwide. Skillful forecasts on two-month to ten-year timescales would enable proactive and informed decision-making for many industries, including fisheries, water management, and agriculture. Understanding the behavior of seasonal-to-decadal climate variability provides context for our changing environment. Neural networks, a class of artificial intelligence tools, are well-suited for exploring teleconnections, precursors, and patterns of variability, since they can identify complex relationships within immense quantities of data. Neural networks have traditionally been used as "black-box" models that produce predictions but are inherently difficult to explain. There has been a recent push to develop "interpretable" models that can be understood by human scientists. In this dissertation, I bridge human and artificial intelligence to leverage interpretable AI for *skillful*, *trustworthy*, and *insightful* prediction of seasonal-to-decadal climate variability.

First, I show how interpretable neural networks can be used to optimize a simple forecasting method, analog forecasting. This approach highlights four precursor patterns for one-year forecasts of El Niño Southern Oscillation in the Tropical Pacific, West Pacific, Baja Coast region, and Tropical Atlantic. In addition, when making five-year forecasts of observed sea surface temperature variability in the North Atlantic, this optimized analog forecasting approach rivals the performance of an initialized decadal prediction system.

Second, I design neural networks to learn patterns of internal variability and forced change. Using these neural networks, I perform climate change attribution for observed sea surface temperatures. Despite the unprecedented, record-high, global-mean sea surface temperature in 2023, our results suggest that much of this warming can be explained by internal variability, as anomalously cold conditions in 2021 and 2022 shifted to anomalously warm conditions in 2023.

Third, I use neural networks to make decadal forecasts of the likelihood that annual-global-mean temperature exceeds 1.5°C, a critical Paris Agreement temperature threshold. These forecasts predict that it is very likely that annual-global-mean temperature exceeds 1.5°C in the next decade (2024-2033), serving as a harbinger for future climate change. These forecasts are consistent with dynamical initialized prediction systems, demonstrating that neural networks can provide skillful decadal forecasts at reduced computational expense.

Neural networks are powerful tools for prediction, and facilitate deeper discovery of our chaotic, interconnected, predictable Earth.

ACKNOWLEDGEMENTS

This research was supported, in part, by National Science Foundation Division of Atmospheric and Geospace Sciences (NSF-AGS) grant 1749261, and the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0020347.

This dissertation was built on a foundation of love. I am grateful for all the support I received from my loving family, friends, teammates, and partner. My coauthors pushed me to become a better storyteller and a better scientist. The Barnes Group rejuvenated my passion for science whenever my battery got low. I'd like to thank my committee for their mentorship and support, and my advisor Libby for creating a challenging learning environment where I could safely fail.

Everything will be alright in the end.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: Introduction

The Earth—the atmosphere, the oceans, the land—is a highly complex system governed by chaotic, interconnected processes. A long history of Earth system research has resulted in major advancements in forecasting and projections, from rapid improvement in the skill of short-term weather forecasts [1] to refined representations of global climate variability and change [e.g., 2, 3]. While forecasts on seasonal-to-decadal (2 months to 10 years) timescales have also become more skillful [4, 5], there is potential for further improvement [6]. More skillful seasonal-to-decadal forecasts would enable proactive and informed decision-making for a number of sectors impacted by seasonal-to-decadal climate variability. This includes fishery management [e.g., 7], which is sensitive to seasonal-to-decadal variability in SST [e.g., 8, 9] water management in regions where there is seasonal predictability of precipitation [e.g., 10], and agriculture, which is sensitive to long-lasting temperature and precipitation extremes [e.g. 11]. As more stress is put on the environment due to human activity and human-caused climate change, seasonal-to-decadal forecasts become even more necessary for managing these systems [12].

Seasonal-to-decadal predictability primarily comes from internal climate variability in the form of teleconnections. Teleconnections are distant climate anomalies in the oceans, land, and atmosphere that are linked via atmospheric and oceanic circulations. This includes the temperature and precipitation response to large-scale patterns of ocean variability, such as El Niño Southern Ocean [ENSO; 13, 14] or Atlantic Multidecadal Variability [AMV 15, 16]. The response to external human-caused forcings, such as greenhouse gas and aerosol emissions, also adds predictability on decadal timescales as the mean state of atmosphere evolves [17]. These two sources of predictability create challenges for seasonal-to-decadal prediction: forecasts must consider the patterns of internal variability, the forced response, and the complex interactions between.

Operational seasonal-to-decadal forecasts are often based on simulations from dynamical models which have been initialized with observations [e.g., 18, 19]. Dynamical simulations are computationally expensive, and initialized forecasting requires that they are re-run with the most recent observations every year or less. Initialized prediction systems also have mean-state biases and struggle with initialization shock, which can degrade the quality of the information contained by observed initial conditions [6]. Data-driven approaches to seasonal-to-decadal prediction, including machine learning/artificial intelligence (AI) methods, have been shown to rival the prediction skill of dynamical models at reduced computational expense [e.g., 20, 21, 22]. Neural networks, in particular, are a natural fit for seasonal-to-decadal prediction, since they can learn complex relationships (e.g., teleconnections) from immense amounts of data (e.g., climate model data and observations).

While neural networks have garnered attention for their applicability to Earth system predictability [23], they have traditionally been used as "black box" models, producing predictions without explanations of their decisions [24]. However, there has been a recent push to make neural networks that explain their reasoning [25]. These "interpretable" networks deliver more than just a prediction, they provide a means for gaining new insights into our Earth system and strengthening trust in the model [e.g., 26, 27].

This dissertation lies at the intersection of seasonal-to-decadal climate prediction and interpretable AI. Within, I explore how neural networks can be used for *skillful*, *trustworthy*, and *insightful* seasonal-to-decadal climate prediction. Bridging human and artificial intelligence, these interpretable neural networks are designed to identify patterns in the Earth system in the same way climate scientists think about climate predictability. This ensures that the explanations provided by the neural networks can be understood by the humans that use them. These explanations enable deeper understanding of the present climate, the future climate, and the mechanisms that guide climate variability and predictability. In Chapter 2, I develop a neural network framework for improving seasonal-to-decadal

analog forecasts and explore the precursor patterns of ENSO and AMV. In Chapter 3, I use neural networks to disentangle internal variability and the forced response in sea surface temperature (SST) observations, and perform climate change attribution for the record-high sea surface temperature in 2023. In Chapter 4, I make decadal climate predictions of surface temperature and assess the likelihood that global-annual-mean temperature will exceed politically-relevant temperature thresholds in the near-term climate. In Chapter 5, I summarize how this dissertation advances the field of seasonal-to-decadal prediction and inspires new questions for future study.

The work in Chapter 2 has been peer-reviewed and published in Geophysical Research Letters [22], and has been reproduced in this dissertation without alteration:

Rader, J. K., and E. A. Barnes, 2023. Optimizing seasonal-to-decadal analog forecasts with a learned spatially-weighted mask. Geophysical Research Letters, 50(23), e2023GL104983. doi: 10.1029/2023GL104983.

Chapter 2: Optimizing Seasonal-to-Decadal Analog Forecasts with a Learned Spatially-Weighted Mask

2.1 Summary

Seasonal-to-decadal climate prediction is crucial for decision-making in a number of industries, but forecasts on these timescales have limited skill. Here, we develop a data-driven method for selecting optimal analogs for seasonal-to-decadal analog forecasting. Using an interpretable neural network, we learn a spatially-weighted mask that quantifies how important each grid point is for determining whether two climate states will evolve similarly. We show that analogs selected using this weighted mask provide more skillful forecasts than analogs that are selected using traditional spatially-uniform methods. This method is tested on two prediction problems using the Max Planck Institute for Meteorology Grand Ensemble: multi-year prediction of North Atlantic sea surface temperatures, and seasonal prediction of El Niño Southern Oscillation. This work demonstrates a methodical approach to selecting analogs that may be useful for improving seasonal-to-decadal forecasts and understanding their sources of skill.

2.2 Plain Language Summary

Understanding how the climate will look in one to ten years is useful for many industries, but this task is very difficult. One method for making forecasts on these timescales is called analog forecasting. In analog forecasting, a researcher finds past states in observations, or states in a climate model simulation, that look like the current state of the climate, and uses how those maps changed over time to predict how the climate will change over time. Some regions are more important for determining how a climate state will change over time, and we use a machine learning method called a neural network to identify these important regions. We find that if we only look at these important regions when determining if two climate states are similar or not, we can improve our analog forecasting skill.

2.3 Background

Forecasts on seasonal-to-decadal timescales are crucial for decision-makers in a number of industries, but forecasts on these timescales have limited skill [6, 28, 29]. Analog forecasting, predicting what will happen based on previous states with similar initial conditions, is an intuitive method for seasonal-to-decadal prediction. It is built on the premise that similar geophysical states will evolve in similar ways [30]. It follows that analogs—similar looking states to the initial state that is being forecast—can provide insight into how that initial state will continue to evolve. The analog forecasting approach is powerful for seasonal-to-decadal climate prediction [e.g., 20, 31, 32, 33, 34] and can outperform general circulation models (GCMs) initialized with observations, which struggle with initialization shock and climate model drift [6, 35].

A major hurdle in obtaining successful analog forecasts is that the climate system is noisy and chaotic, and thus small differences between two initial states can result in vast differences in their evolution [36]. Thus, a successful analog forecast for a particular initial climate state, which we refer to as the state of interest (SOI), requires that the analogs and SOI are sufficiently similar such that their evolutions do not significantly diverge during the prediction timeframe. Sufficiently similar analogs can be difficult to find in the observational record since the number of independent observations we have on seasonal-to-decadal scales (e.g., fewer than 100 during the satellite era) is so much smaller than the number of degrees of freedom within a global geophysical field [e.g., 37]. While observations are in short supply, there is a wealth of simulated climate data and many recent studies have used “model-analogs" [31] drawn from climate model output instead [e.g., 38, 39, 40].

We refer to the library of climate model states that can be used for analog forecasting as “potential analogs." Once a potential analog has been identified to be sufficiently similar to the SOI we refer to it as an analog. Forecasts are made by taking the mean evolution of the top-N analogs, where N is chosen by the user. There are several ways to quantify the

similarity between the potential analogs and the SOI. The most straightforward method is to compute the global correlation between each potential analog and the SOI [e.g., 41]. Using a global correlation assumes that the similarity between the maps at each grid point globally matters equally. A natural next step in complexity is to compute a correlation over a region that is known to be important for predictability of a given target, such as the North Pacific for predicting the Pacific Decadal Oscillation [e.g., 40]. While this approach removes some regions that may not be useful for determining the best analogs, it still assumes that each grid point within the region is equally important and the region must be known *a priori*.

In the following work, we train an interpretable neural network on a proxy task that is similar to the analog problem (Section 2.5). The network learns a weighted mask which is used for determining analogs. The forecasting skill of the analogs selected using the learned weighted mask is tested through a perfect model approach where climate model data substitutes observations and is used to predict future climate model data. We demonstrate how this method can be applied to analog forecasting through two prediction examples: forecasting 5-year sea surface temperature (SST) anomalies in the North Atlantic (Section 2.6) and wintertime SST anomalies in the tropical Pacific (i.e., El Niño Southern Oscillation; Section 2.7). In these examples, analogs identified using the weighted mask provide more skillful forecasts than analogs that are identified in a way that is globally or regionally uniform. In addition, we show that these masks, once generated by a neural network, can be modified *post hoc* to further investigate the importance of each region for seasonal-to-decadal prediction (Section 2.7).

2.4 Data and Metrics

2.4.1 Climate Model Data

We use monthly SST from the historical run of the Max Planck Institute (MPI) for Meteorology Grand Ensemble [GE; 42] at 2° latitude by 2° longitude resolution. This

dataset contains 100 members and each simulates 156 years (1850-2005) of the Earth's climate with historical forcing. The MPI-GE uses the MPI Earth System Model version 1.1 [ESM1.1; 43]. Each member is initialized using a different year of the preindustrial control simulation such that the differences between ensemble members are a product of internal variability. Learning the weighted mask requires a large number of training samples, which makes the 15,600 years provided by the MPI-GE historical simulations a natural fit for this task.

2.4.2 Standardization and Selection

Subsets of the MPI-GE ensemble members are used for different purposes. Our library of potential analogs is made up of members 1-35. Members 36-50 are the SOIs for training the neural network, members 51-55 are the SOIs for the early stopping validation set (which is used to prevent overfitting to the training data), and members 56-60 are the SOIs for the tuning validation set (which is used to identify optimal hyperparameters for the neural network). Finally, members 96-100, which are withheld until the very end, are the test set for making and evaluating the analog forecasts. Details on the process of tuning and training the neural network, including selecting the hyperparameters, can be found in Section A.1.

Each sample *i* or *j*, from the SOIs or the library of potential analogs, is composed of an input field ($I_{SOI,i}$ or $I_{analog,j}$) and a target ($T_{SOI,i}$ or $T_{analog,j}$). The input fields are one or more maps of global SST leading the targets over some earlier period (the "input period"). The targets are time- and area-mean SST anomalies over a certain region and forecast window.

We removed the forced signal from the climate model data by subtracting the ensemble mean of the library of potential analogs at each location and year from each set of data. After the forced signal was removed, the data was standardized by dividing by the standard deviation at each grid point across the library of potential analogs. By using the library of

potential analogs to calculate the forced signal and internal variance we treat the SOIs as if they are truly unseen data as we would when forecasting.

2.4.3 Metrics

We measure forecasting skill with a mean absolute error (MAE) skill score. This skill score is calculated by comparing the MAE of the analog prediction for the SOIs in the test set with the MAE of climatology, as:

$$\text{Skill Score} = 1 - \frac{MAE_{pred}}{MAE_{climo}}$$

such that a perfect prediction has a score of one, and a climatology prediction has a score of zero. Climatology is the prediction by the mean state, which is zero for this standardized data. Analog forecasts made using the weighted mask are compared with the following additional baselines: a global analog forecast, a target region analog forecast, a mean target evolution forecast, and a random forecast. In the global analog forecast (target region analog forecast), the analogs are selected if the unweighted MSE over the entire globe (target region) is the smallest. The mean target evolution forecast is based on how the targets in the input period evolve on average and is detailed in Section A.2. The random forecast is made by randomly selecting targets from the library of potential analogs and using them as the prediction. In addition to the MAE skill score, the Pearson correlation coefficient can be found in Figure A.6.

2.5 Optimized Analog Forecasting Approach

Our goal is to find optimal analogs for forecasting a specific target. To do this, we train a neural network to identify a spatially-weighted mask. This weighted mask is then multiplied by the SOI and potential analogs and the mean-squared error (MSE) between the weighted maps is used to determine how similar they are (Figure 2.1). This weighted mask should contain large values where similarity between the analogs and the SOI is

most important for predicting the target and near-zero values where similarity between the maps is not important. With this architecture, the MSE will be low if the maps agree where the mask weights are high, regardless of the differences between the maps where the mask weights are low. For the plots in this paper, the mask is normalized by dividing by the sum of the weights times the size of the input, such that the mean weight is one.

We generate the weighted mask by training a neural network on a proxy task that is tangential to our main goal. While our goal is to identify a weighted mask that is optimized for making an analog forecast, our proxy task is to predict the difference in $T_{SOI,i}$ and $T_{analog,j}$ given $I_{SOI,i}$ and $I_{analog,j}$. En route to making this prediction, the neural network must learn the weighted mask, multiply it by the two input maps, compute the MSE between these weighted maps, and finally convert the MSE into a predicted difference in the targets. This process is depicted in the red box of Figure 2.1.

Once the weighted mask has been learned, a neural network is no longer needed to make analog predictions. The weighted mask is multiplied by the SOI and each potential analog, the MSE is computed between the weighted SOI and the weighted potential analogs, and the potential analogs with the lowest MSE are used to make the analog forecast. While the proxy task is not identical to the analog problem, it provides a weighted mask that improves analog forecasting skill, as we will show in Sections 2.6 and 2.7.

2.6 Multi-year Prediction of North Atlantic Sea Surface Temperature

We first test our analog forecasting approach on a multi-year prediction of SSTs over the North Atlantic. North Atlantic SSTs exhibit clear variability on multi-annual timescales [44] and exhibit potential for skillful decadal forecasts [45, 46]. SST variability in the North Atlantic has been associated with weather and climate anomalies globally, including Atlantic hurricane frequency and intensity [47, 48], northern hemisphere precipitation [15, 49], and the strength of the Asian summer monsoon [50]. In this prediction problem,

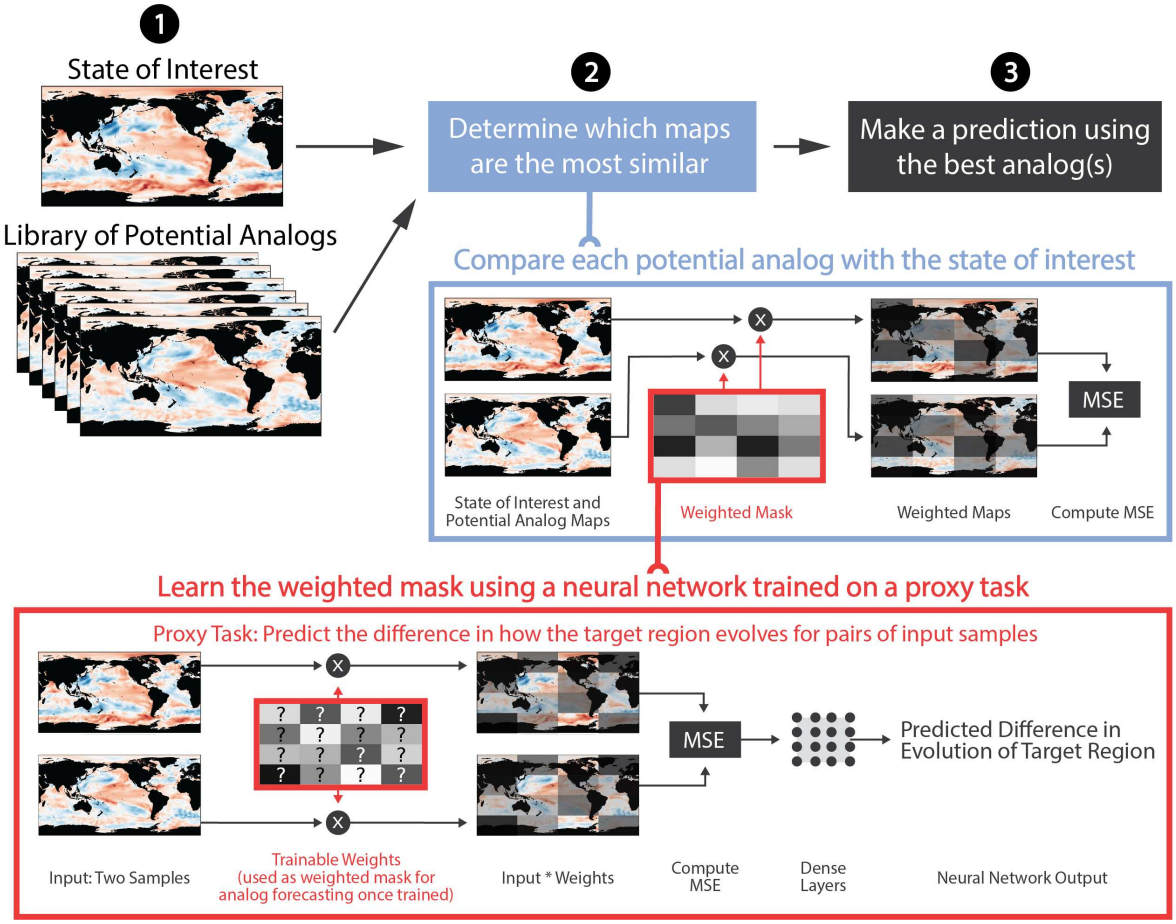

Figure 2.1: Optimized analog forecasting method and interpretable neural network architecture. The analog forecasting method can be described in three steps: 1) identify a state of interest and a library of potential analogs. 2) Determine which maps are the most similar. 3) Make a prediction using the best analog(s). In the blue box, we show our weighted-mask approach for determining the similarity of two maps. The weighted mask is multiplied by the state of interest and a potential analog before computing the mean squared error (MSE). In the red box, the interpretable neural network architecture is shown. Two input samples are multiplied by a matrix of trainable weights and the MSE is computed. This MSE is then converted to a predicted difference in the sample targets using a group of fully-connected dense layers. Note that the weighted mask has the same dimensions as the input field(s), despite the coarser resolution in this figure.

.

we use global maps of SST, averaged over the previous five years, to predict the mean SST anomaly in the North Atlantic (40°-60°N, 10°-70°W) over the following five years.

The weighted mask learned by the neural network is shown in Figure 2.2a. The Greenland Sea and the gulf stream region in the western North Atlantic emerge as the most important regions for identifying analogs in the MPI-GE. Over the western North Atlantic, there is an area of zero weight between two areas of high weight. These may be where the boundaries of persistent SST anomalies vary, and the neural network has learned that the specific locations of these boundaries are not important for the prediction problem. Previous studies that have used an analog approach to assess North Atlantic decadal predictability selected the best analogs by taking a correlation over the whole globe [41] or the entire North Atlantic basin [32]. As shown in Figure 2.2b-d, when using the weighted mask, the best analogs only have to look like the SOI in the highest weight regions. An example SOI is shown in Figure 2.2b and its best analog in Figure 2.2c. These two maps look similar in the North Atlantic, but are starkly different in the North Pacific and Indian Ocean, among other regions. Once the weighted mask has been applied to the SOI (Figure 2.2d) and its best analog (Figure 2.2e), the maps look nearly identical.

These results suggest that using uniform weights across the entire North Atlantic basin, or the whole globe, may lead to a selection of analogs that are not optimized for forecasting multi-year variability in the North Atlantic. Indeed, we see that this is true in the skill scores shown in Figure 2.3a. For $1 \leq N \leq 50$, where the top-N analogs are averaged, our weighted mask analog forecast outperforms the global and target region analog forecasts, as well as the climatology, mean target evolution, and random baselines. The skill score is lowest when only the single best analog is used for forecasting, and subsequently improves for larger N. Given that the skill score maximizes around $N = 10$, and the spread of the targets associated with the analogs (i.e. the uncertainty of the forecast) increases with N (Figure A.1), we elect to focus on results for $N = 10$ analogs. The prediction by the top-10 analogs, and the spread of the targets, are shown in Figure 2.3b for 200 years of SOIs. The

analog predictions do a good job of capturing the variability of North Atlantic sea surface temperatures, though they do struggle to forecast the most extreme anomalies.

2.7 Seasonal Prediction of El Niño Southern Oscillation

In addition to improving prediction skill, the weighted mask can be used to explore precursor patterns within the climate simulated by MPI-GE. This is a major benefit of the interpretable neural network architecture, as the weighted mask can be compared to known precursor patterns to improve trust in the weighted analog forecasts and provide new insight into Earth system predictability. Here, we extend the application of the weighted mask analog approach to seasonal forecasts of ENSO. ENSO precursors are well-studied providing an ideal case for exploring the utility of the weighted mask for predictability studies.

ENSO is the leading mode of global annual SST variability [51] and has an extensive influence on global weather and climate [reviewed in 52]. Analog forecasting has been applied to seasonal prediction of ENSO in several studies due to its potential to outperform initialized GCM forecasts [e.g., 31, 20]. In the following example, we use wintertime (November-March) global SST anomalies to forecast SST anomalies in the Niño3.4 region [5°S-5°N, 120-170°W; 53, 54] the following winter.

The weighted mask for forecasting ENSO looks markedly different from that for forecasting North Atlantic multi-year variability (Figure 2.4a). While a few regions are assigned higher weights, the weights in Figure 2.4a are much more uniform across the globe than in Figure 2.2a. The four main regions that stand out in this weighted mask have also been identified as important precursors in previous literature: the western North Pacific [e.g., 55], the Pacific Meridional Mode [e.g., 56], the Central Atlantic [e.g., 57], and the tropical Pacific itself [e.g., 58]. The skill score of the global analog forecast (Figure 2.4b) is similar to that of our weighted mask analog forecast (but always lower, see Figure A.3), which is

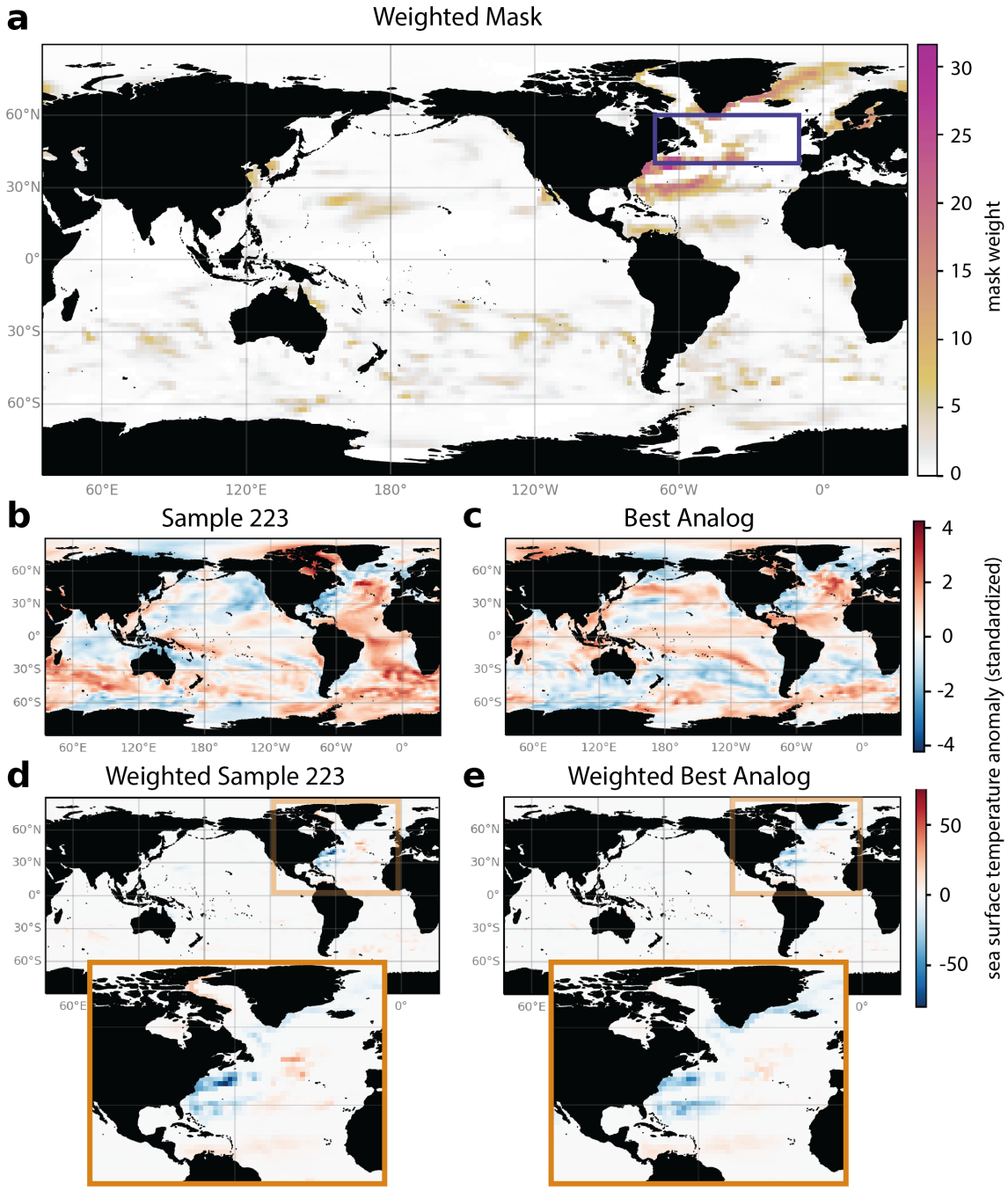

Figure 2.2: Weighted mask and example for multi-year predictions of North Atlantic SST. (a) Weighted mask, as learned by the interpretable neural network. (b) Standardized SST anomalies for a sample state of interest (SOI). (c) Standardized SST anomalies for the best analog associated with the SOI. (d) Weighted SOI. (e) Weighted best analog.

.

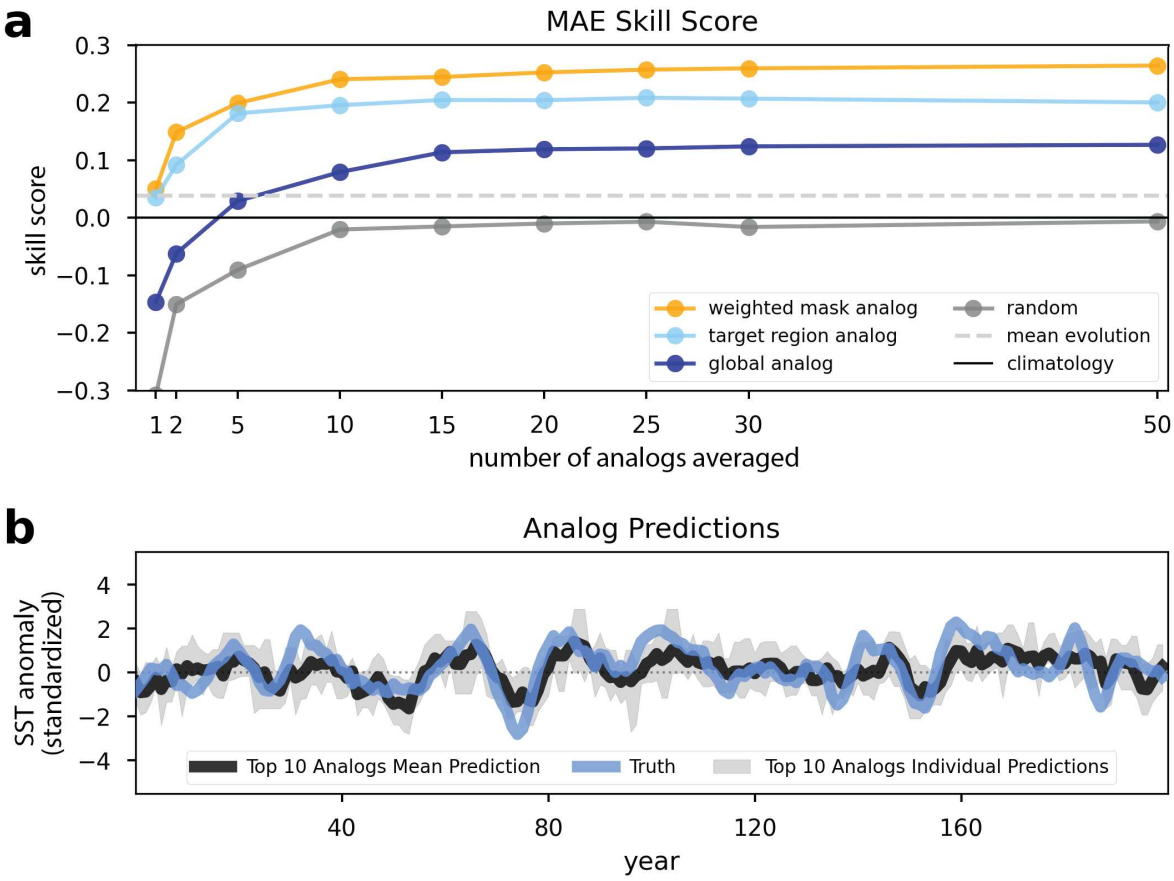

Figure 2.3: Analog forecasts of North Atlantic sea surface temperature. (a) Skill scores for our weighted mask analog forecast and other baselines. (b) Weighted mask analog forecasts for 200 years of MPI-GE simulations, including the mean prediction from the top-10 analogs, the spread of these predictions, and the truth values.

.

not surprising since the values of the weighted mask are near one for most areas of the globe.

Since the weighted mask can be manually updated *post hoc*, we use this to explore the sensitivity of the forecast skill to which regions are included in the weighted mask. Figure 2.5a shows the weighted mask for ENSO prediction (Figure 2.4a) but where the smallest 95 percent of the weights have been set to zero. Forecasts made with this "constrained" weighted mask have similar skill to the original weighted mask (as shown in Figure A.4). From the constrained weighted mask, we identify four main precursor regions for ENSO: the West Pacific (ocean grid points bounded by 0°-40°N, 100°-170°E), the Tropical Pacific (25°S-10°N, 170°E-65°W), the Baja Coast (10°N-40°N, 110°-140°W), and the Tropical Atlantic (0°-20°N, 20°-80°W).

We assess how important each precursor region is in two ways. In the first approach, we test the skill score of analog forecasting when each region is occluded from the constrained weighted mask (weights in that region are set to zero). When all four regions are included, the skill score is 0.146. Removing any of the four regions from the weighted mask results in a skill score decrease. Interestingly, removing the Tropical Atlantic results in the most drastic decrease in prediction skill. While the Tropical Atlantic has been connected to ENSO predictability—tropical Atlantic SSTs modulate the Walker Circulation and, in turn, the SST gradient of the tropical Pacific [57, 59]—it is not considered a primary driver [60]. In the second approach, we isolate each of the four regions (weights outside that region are set to zero). There is no improvement over climatology when just the Baja Coast or Tropical Atlantic is used to select analogs, and more skill when just the West Pacific or Tropical Pacific is used. However, no region alone provides anywhere near the skill that all four regions do together.

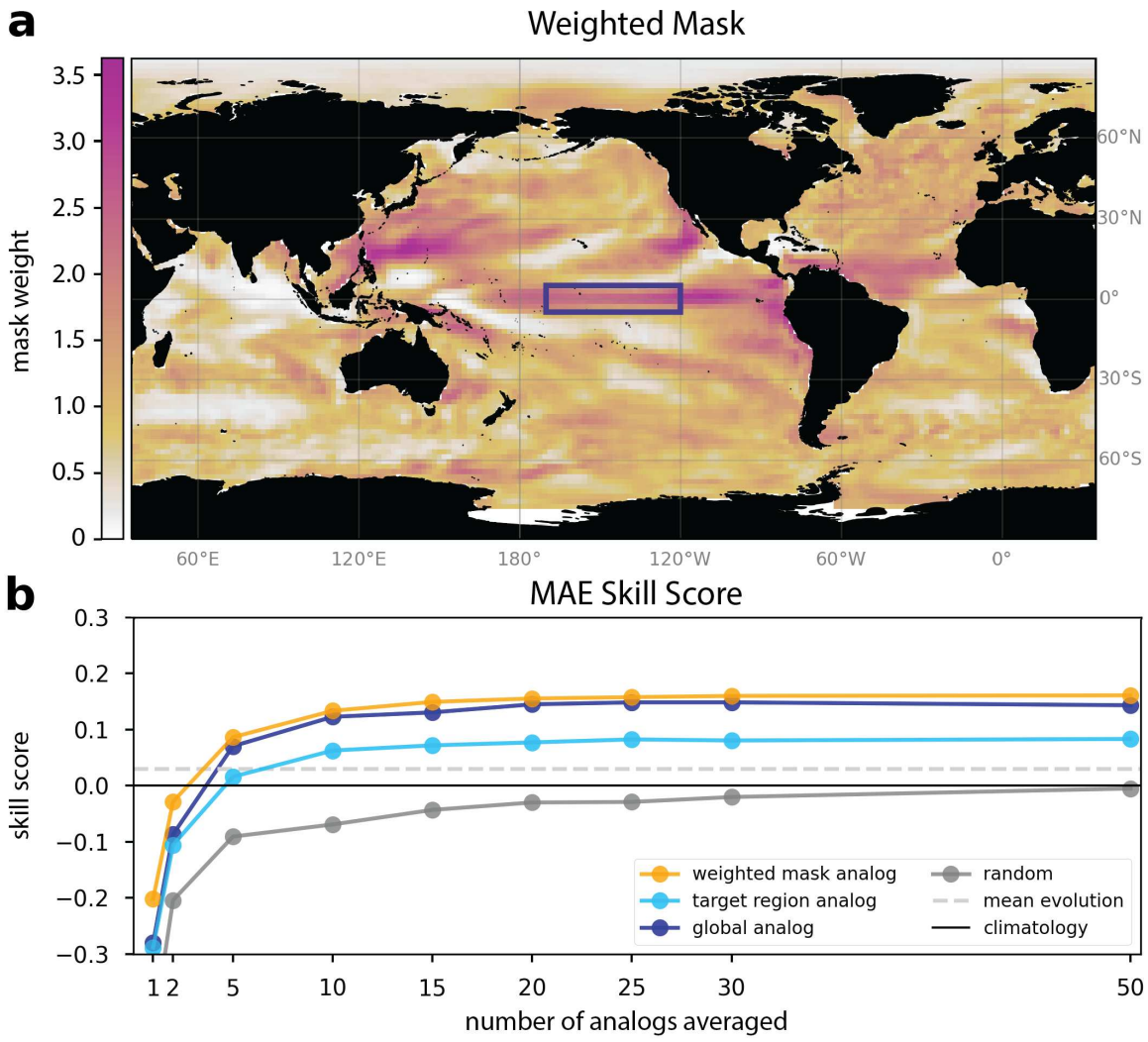

Figure 2.4: Weighted mask and skill scores for seasonal predictions of El Niño Southern Oscillation. (a) Weighted mask. (b) Skill scores for our weighted mask analog and other baselines.

.

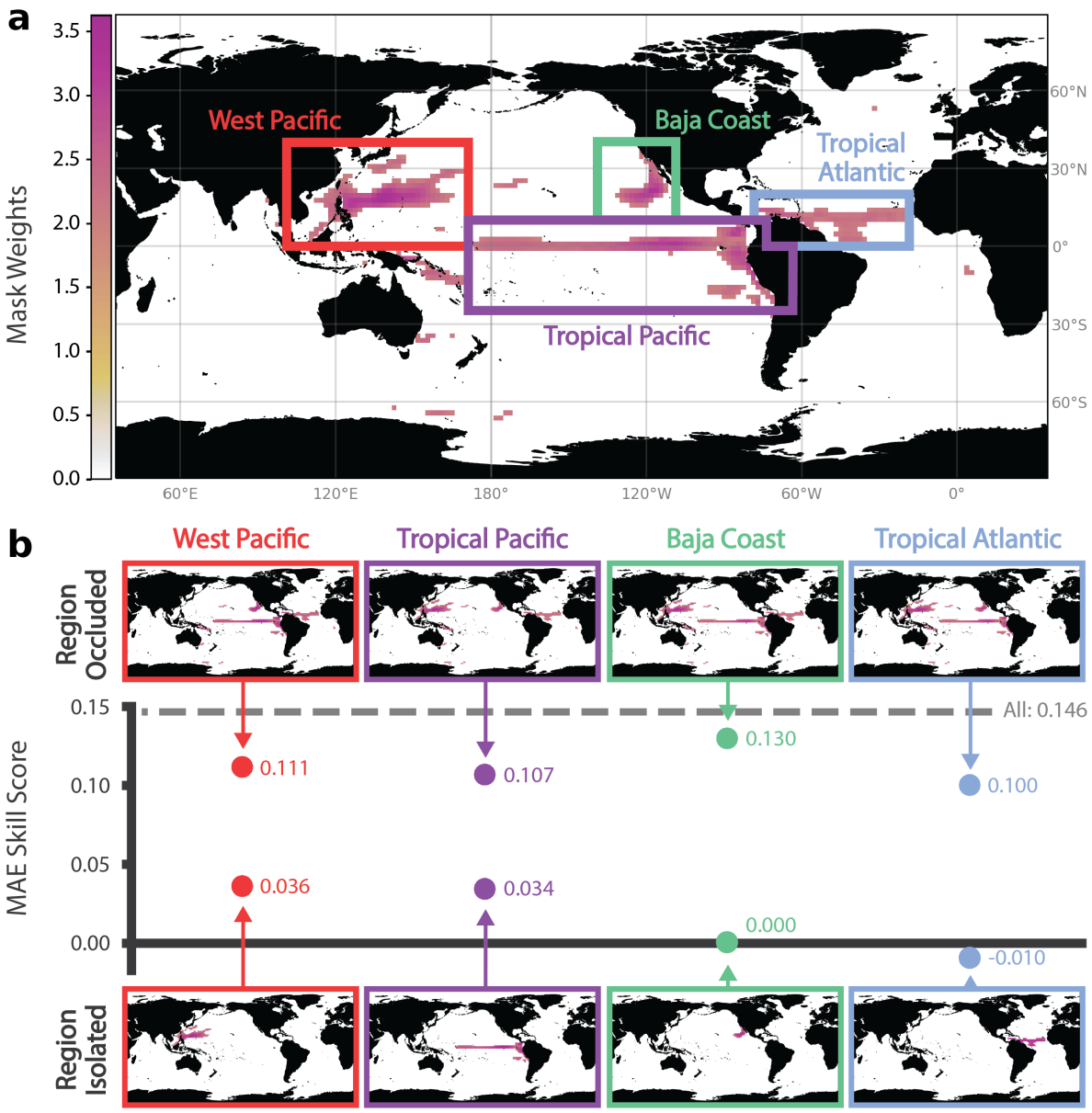

Figure 2.5: Analog forecasting skill of El Niño Southern Oscillation when various regions are occluded or isolated. (a) As in Figure 2.4a, but the lowest 95 percent of weights are set to zero. Four regions of focus are highlighted by the colored boxes. (b) Skill scores for analog forecasts when each region is occluded from the mask (top) and when the region is isolated to make a forecast (bottom).

.

2.8 Discussion and Conclusions

We have shown how an interpretable neural network can be used to identify a weighted mask that improves the selection of analogs for seasonal-to-decadal forecasting. The precursors identified in the weighted masks are not necessarily causal, but they do provide the optimal predictors for the given input. In this work we have constrained the neural network to learn one mask that represents all pathways of predictability, however allowing the network to learn different masks for different SOIs could lead to better analog forecasts.

This paper is intended to demonstrate the weighted mask approach to analog forecasting. For clarity and simplicity, we only used a single input map of SST to predict a future target SST in this work. However, this methodology is designed to identify masks for multiple inputs (e.g., different variables, time lags) as well. We provide an example of this in Figure A.5, where we include the time tendency of SST as a second input variable for the North Atlantic multi-year prediction example. Including sea surface height or ocean heat content as an additional variable [e.g., 31, 61] has the potential to improve prediction skill in the North Atlantic and tropical Pacific and would provide a unique mask for where these variables provide information beyond SST alone.

We have explored this method through a perfect model setup. As such, the identified precursors are intrinsic to the MPI-ESM and may not reflect patterns of predictability in the observed Earth system. Although there are known issues in the MPI-ESM's ability to simulate North Atlantic SSTs, including a warm bias and a weak meridional gradient [62], this learned weighted mask still acts to improve observational forecasts relative to a uniform mask. These results, and a comparison with an initialized dynamical forecast, can be found in Section A.3. Training the weighted mask on a multi-model ensemble may provide patterns that are more consistent with observations [e.g., 4, 63] and allow for enhanced analog predictions on real data. Additionally, we could train on models and observations at the same time to identify a weighted mask that is more representative of

the true Earth System. We believe that this weighted mask approach will be influential to analog forecasting moving forward.

Chapter 3: Attribution of the record-high 2023 SST using a deep-learning framework

3.1 Introduction

The global-mean sea surface temperature (SST) reached a record high in 2023, exceeding the previous record by 0.14°C and the 1981-2010 average by more than 0.5°C [64, 65]. Global-mean SST is a product of both externally forced change (e.g., human-caused warming) and variability internal to the climate system. This record was unprecedented and unpredicted: No other record had surpassed a previous record to such a large extent, and dynamical and statistical forecasting systems failed to capture such an extreme [66, 67]. Even anomalous external forcings that are not accounted for in these forecasting systems, such as from the Hunga Tonga–Hunga Ha'apai volcanic eruption in Tonga or decreased aerosol emissions from shipping, are unlikely to fully explain this rapid warming [68, 69]. This raises the question: can the 2023 SST record be attributed to an abrupt change in the forced response, or was this extreme warmth the result of internal variability?

The extent to which the record-high SST in 2023 is a product of internal variability or the forced response has major implications for our understanding of the climate system. An abrupt change in the forced response could indicate accelerating or nonlinear warming [70, 71], due to changes in SST patterns [e.g., 72] or the triggering of a climate tipping point [73]. Abrupt warming would result in major impacts on human and natural systems globally, including ecology [74, 75] and economy [76], with further implications for human well-being and behavior [77]. Attribution of the 2023 global mean SST is also critical for discussions on climate change policy [66]. International agreements, such as the United Nations Paris Agreement, have established warming targets to limit the impacts of forced climate change [78, 79]. Actively monitoring the current proximity of the climate system to these warming targets requires immediate estimation of the forced response [80].

The Intergovernmental Panel on Climate Change (IPCC) uses the 20-year rolling average temperature to estimate the global warming magnitude within climate models [81]. This requires knowledge of the climate system a decade into the future, which is not available for real-world data and thus cannot be used to estimate the global warming magnitude in observations. Traditional methods, such as linear or fourth-order polynomial fits to historical observations [e.g., 82, 83], could be used to disentangle the forced and internal components in 2023, but they lack the ability to capture nonlinearities or abrupt changes in the forced response [e.g., the temperature response to the eruption of Mount Pinatubo; 84]. Hybrid methods that combine recent observations with decadal climate forecasts [e.g., 80] can also provide near-instantaneous estimates of the current global warming magnitude, but they rely on climate projections that do not include abrupt changes to future forcing [e.g., 17]. Fingerprinting methods [e.g., 85, 86] identify distinct spatial patterns of forced change within climate model simulations which can be projected onto observations to isolate the forced component of change. Since fingerprinting is not sensitive to the time-evolution of climate change, it is a natural fit for detecting sudden shifts in the forced response. However, simulated patterns of the forced response are inconsistent with recent observed SST trends, particularly in the tropical Pacific and the Southern Ocean [87, 88], and climate model-derived fingerprints are encoded with these same biases.

In this work, we develop a deep learning method for attribution of the record-high 2023 SST. A neural network is trained to separate the internal and forced components of annual-mean SST by learning patterns of forced change and internal variability from thousands of realizations of SST simulated by climate models. Once trained, the neural network uses these learned patterns to estimate the forced and internal components of observed SST. In the following, we will show that our deep learning approach provides skillful attribution for out-of-sample climate model data, and explore the extent to which the record-high SST in 2023 can be attributed to an abrupt change in the forced response or anomalously warm internal variability.

3.2 Attribution Framework

Our study uses a neural network to separate internal variability and the forced response for annual-mean SST. Following the framework and data provided by the ongoing Forced Component Estimation Statistical Methods Intercomparison Project [ForceSMIP; 89], our neural network leverages data from five climate model initial-condition large ensembles [90, 91, 92, 93, 94]. These large ensembles simulate the climate under historical forcings (1950-2014) and three future forcing scenarios (2015-2022) [95]. MIROC-ES2L (30 members) uses the SSP2-4.5 forcing scenario, CESM2 (50 members) uses the SSP3-7.0, and MIROC6 (50 members), CanESM5 (25 members), and MPI-ESM1-2-LR (30 members) use the SSP5-8.5. All climate model data is regridded to a 2.5°x 2.5°resolution. Since each large ensemble has at least 25 members, we ensure each climate model large ensemble is equally represented in the training set by using the first 17 members for training. The following eight members are used to explore how the neural network learns patterns of the forced response and internal variability (Section B.1).

In large ensembles, the forced and internal components are known: the ensemble mean SST is the forced climate response while the residual is internal variability [96]. In this supervised learning task, we have the option to either predict the forced or internal component. We predict internal variability, calculating the forced response by subtraction, for two related reasons. First, deep learning algorithms are prone to overfitting on small training sets, and predicting the internal component provides more distinct samples to train on. While our training set comprises 6205 samples (5 climate models x 17 members x 73 years), all with unique patterns of internal variability, there is only one forced response pattern for each climate model ensemble. This leaves only 365 unique maps of the forced response (5 climate models x 73 years), and fewer truly independent samples as the forced response is highly autocorrelated. Second, patterns of internal variability are more complex than the patterns of the forced response [the forced response is often explained by just a few patterns while internal variability requires many; e.g., 97]. While internal variability

and the forced response are two sides of the same coin—the neural network need only learn one to calculate the other—targeting internal variability pushes the network to learn the behavior of internal variability, not just the forced response (Section B.1).

Our prediction task is set up as follows: The neural network is given a map of simulated annual-mean SST as input and tasked to estimate the internal variability component. The input maps are preprocessed to remove differences in the mean temperature patterns between models by first calculating the SST anomalies relative to the global mean then calculating this anomaly relative to the global mean for each climate model member (1950-2022). As is standard practice in machine learning, the input and outputs are standardized at each grid point using z-score normalization. To learn local and global patterns of internal variability and the forced response, the neural network architecture employs convolutional layers, fully-connected layers, and skip connections. The full neural network architecture and the hyperparameter tuning process are discussed in Sections B.2 and B.3.

This methodology achieves three key goals for attribution of the 2023 climate. First, our method makes estimates based on the patterns within single maps of annual-mean SST, such that attribution is not dependent on knowledge of the future climate. Second, this approach does not prescribe that the forced response must evolve smoothly and can thus capture abrupt forced change. Lastly, this approach learns complex, nonlinear patterns of both the forced response and internal variability, which allows for nuanced attribution that relies on more than just the mean forced response simulated by climate models.

We evaluate the performance of our neural network using a leave-one-out approach. Neural networks are trained on five different train/validation splits, where four climate models are used for training and one is withheld for validation. This allows us to assess our methodology on simulations of annual-mean SST that fall outside of the training data before applying the final neural network, which is trained on all five climate models, to truly out-of-sample observations. We train 10 neural networks that differ only in their initializations for each train/validation split, such that 50 neural networks are trained in

total. We use the area-weighted R^2 score (coefficient of determination) to assess the skill of our internal variability estimates for simulations in the validation set, where an R^2 score of one indicates our estimates have perfect skill and an R^2 score of zero indicates our estimates have no skill. In addition, we compare the magnitudes of true internal variability with our estimates by calculating the standard deviation at each gridpoint for both fields.

Once trained on simulated annual-mean maps of SST, the neural network takes observed annual-mean SST to estimate the contributions from internal variability and the forced response. We use the National Oceanic and Atmospheric Administration (NOAA) Extended Reconstructed SST version 5 (ERSSTv5) as our observational SST product [65]. This data is regridded to a 2.5°x 2.5°resolution to match the data in our training set. We make 10 different estimates of the internal component of observed SST using neural networks with different initializations. Our final estimate of the internal variability component is the mean of these 10 estimates, which is more skillful than the estimate from any single neural network (Figure B.1).

3.3 Results

Using the neural networks trained with leave-one-out cross-validation, we find that our approach supplies skillful estimates of internal variability for out-of-sample climate model data. Across most of the globe, the R^2 score is greater than 0.7, indicating that our neural network skillfully estimates internal variability (Figure 3.1a). In these regions of skill, the true and estimated magnitudes of internal variability are comparable, differing by less than 20%. Exceptions in the skill of our estimates of internal variability can be found in parts of the Arctic, North Atlantic, and the Southern Ocean, regions where SST variability is dominated by sea ice extent.

We compare the neural network estimate of the forced response to the true forced response when CESM2 is withheld from the training set in Figure 3.1c-f. In three regions where SST is dominated by large-scale, multi-annual climate modes—the tropical Pacific,

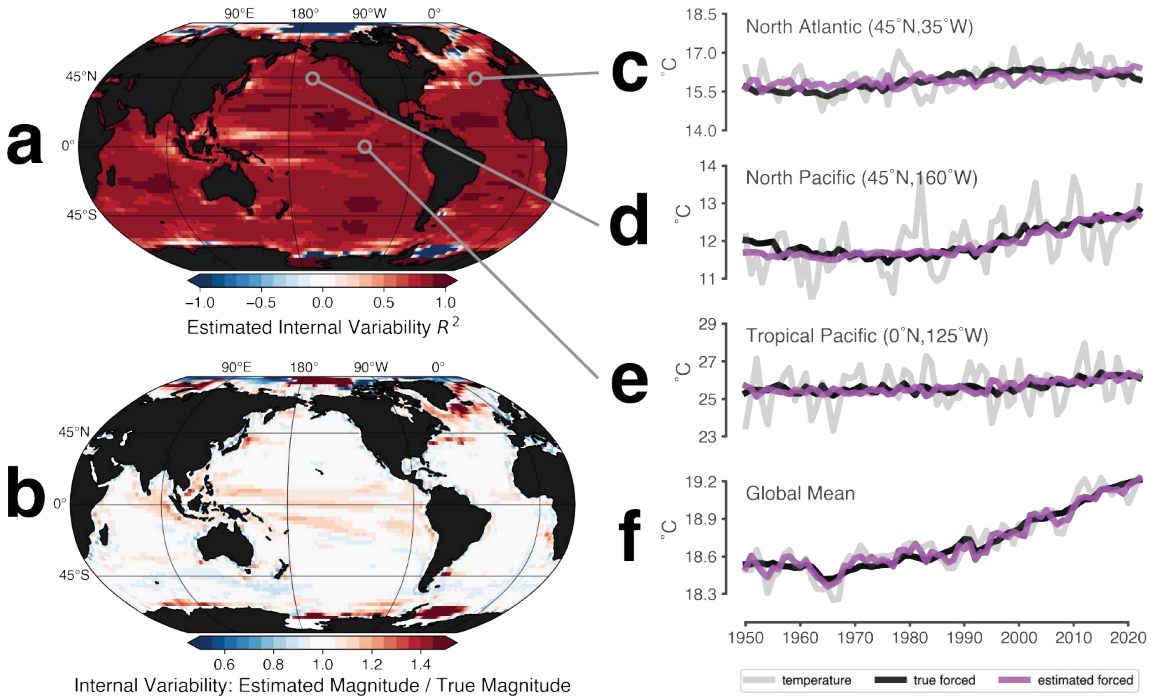

Figure 3.1: Performance of the neural network for separating the forced and internal components of global SST variability. (a) R^2 score of the estimated internal variability relative to the true internal variability averaged over the leave-one-out test sets. (b) The estimated magnitude of internal variability divided by the true magnitude of internal variability within the test sets. Ones indicate regions where the magnitude of the estimated internal variability is equal to the true magnitude of internal variability. Blues indicate regions where the estimates underestimate the magnitude of internal variability. Reds indicate regions where the estimates overestimate the magnitude of internal variability. Example estimates of the forced response for the (c) North Atlantic, (d) North Pacific, (e) Tropical Pacific, and (f) global mean.

.

the North Pacific, and the North Atlantic—our neural networks have skillfully separated the internal and forced components of variability. In addition, the global mean forced response estimate closely tracks the true forced response, including the period from 1992 to 1995 when simulations of global mean SST are cool in response to the 1991 eruption of Mount Pinatubo [98]. Additional examples when other climate models were used for validation can be found in Figure B.2.

Using neural networks trained on all five climate model large ensembles, we now estimate how much internal variability contributed to the record-high SST observed in 2023 (Figure 3.2). Our internal variability estimate reveals an anomalously warm tropical Pacific, contributing +0.04°C to global-annual-mean SST. The North and South Pacific basins are both characterized by cold SST at the eastern boundaries with warm anomalies to the west, accounting for -0.01°C and +0.01°C of the global-annual-mean SST, respectively. Warm anomalies are widespread in the Atlantic Ocean. The largest warm anomalies are in the northern subtropics, where the North Atlantic contributes +0.03°C, while the South Atlantic contributes +0.02°C. The Indian Ocean presents contrasting warm anomalies in the southwest and cold anomalies in the northeast, resulting in a basin-wide contribution of +0.01°C. The Southern Ocean accounts for -0.02°C of the global-annual-mean SST. In total, we estimate that internal variability accounted for +0.07°C on top of the forced component of annual-global-mean SST in 2023.

Estimates of the forced component reveal persistent forced warming of global-annual-mean SST over the past twenty years (Figure B.3). The forced component in 2023 is estimated to be 0.10°C warmer than the forced component in 2022, and 2022 is estimated to be 0.05°C warmer than 2021, while the estimated forced response changed by less than 0.03°C each year from 2016-2021. From 2000-2023, the forced climate is estimated to have warmed by 0.4°C, a rate of approximately 0.2°C per decade.

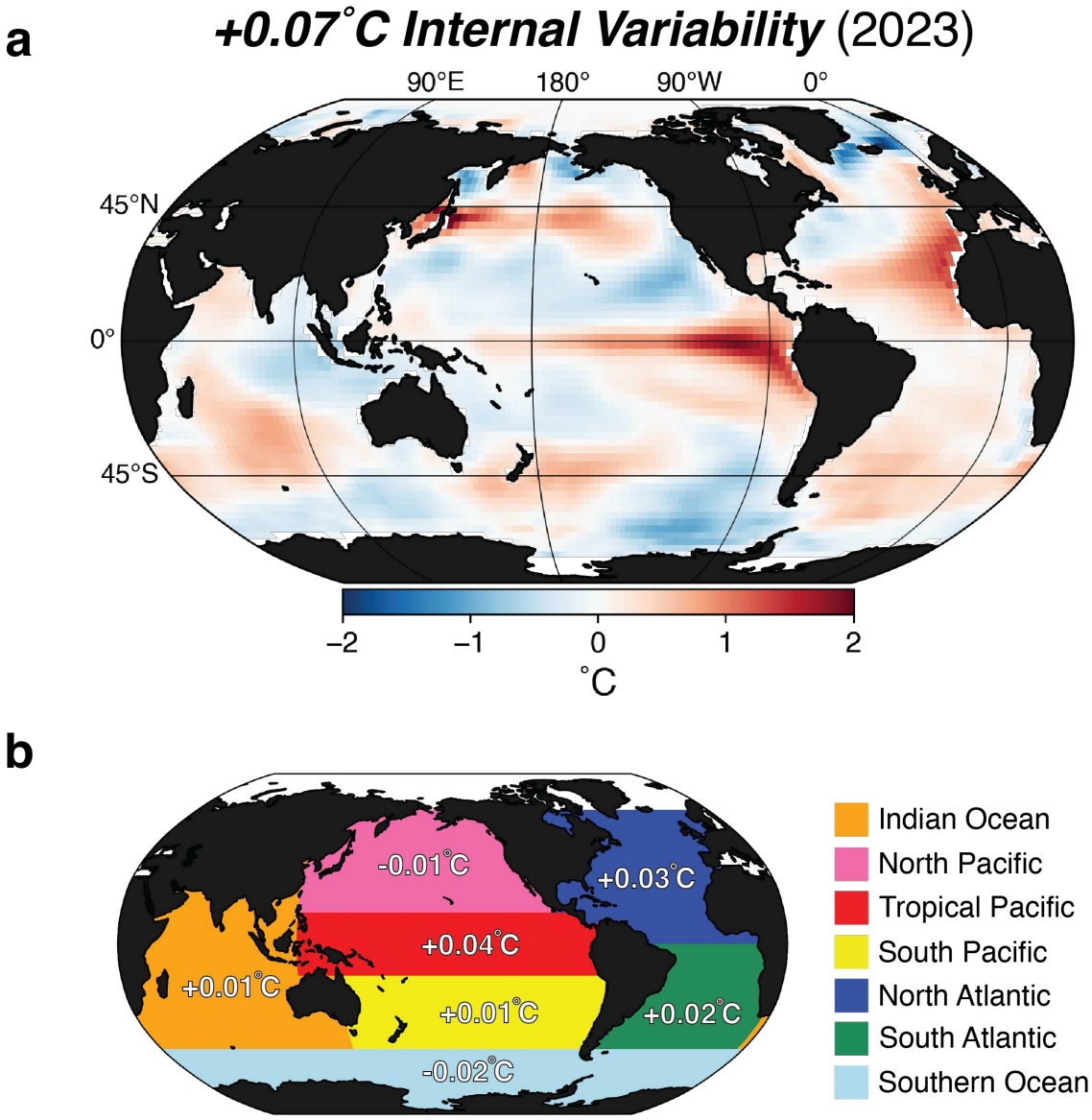

Figure 3.2: (a) Internal variability estimate for observed SST in 2023 attributes 0.07°C of the global-mean SST to internal variability. (b) Regional contributions to the global-annual-mean SST anomaly due to internal variability.

.

3.4 Discussion

Our estimates of internal variability capture multi-annual, large-scale modes of variability like El Niño Southern Oscillation [ENSO; 99, 100] in the tropical Pacific, Pacific Decadal Variability [PDV; 8, 101] in the North Pacific, and Atlantic Multidecadal Variability [AMV; 15, 16] in the North Atlantic (Figure 3.1c-e). These modes of internal variability possess characteristics that resemble the forced response [e.g., the pattern effect in the tropical Pacific and aerosol forcings in the North Atlantic; 102, 103]. Our method captures the cooling response of SST to the 1991 eruption of Mount Pinatubo, which resembled patterns of internal variability [104]. Such a response to Mount Pinatubo is seen in the observational record as well. Estimating the forced and internal components for the entire ERSSTv5 record (1940-2023) reveals forced cooling in annual-global-mean SST following the Mount Pinatubo eruption (Figure 3.3a). The identification of large-scale modes of climate variability as internal variability, and the cool SST following the eruption of Mount Pinatubo as the forced response, demonstrates that neural networks are well-suited for differentiating between forced and internal patterns of change.

This study estimates that internal variability accounted for 0.07°C of the record-high, global-mean SST in 2023. Internal variability of this magnitude does not fully explain that global-mean SST was 0.14°C warmer than the previous record set in 2016. Performing attribution on recent years' maps of global-annual-mean SST reveals a more complete story (Figure 3.3a). In the last few years, internal variability has reduced annual-global-mean SST. Global-mean internal variability in 2020, 2021, and 2022 was +0.03°C, -0.05°C, and -0.06°C, respectively (Figure B.4; Figure B.5). This can be attributed to the cold phase of ENSO, which is a strong driver for unforced variability in annual-global-mean SST [105]. From 2020 to 2022, the climate experienced a "triple dip" in cold-phase ENSO conditions [Figure B.5; 106], driving global-mean SST down and veiling the forced warming of global-mean SST (Figure 3.3a, 3.3b). Thus, the combination of continued global warming [Figure 3.3a; 107]

and the shift to warm-phase ENSO conditions [Figure 3.2; 108] caused the apparent sudden increase in global-mean SST in 2023.

In addition to the warm-phase ENSO in the tropical Pacific, the South Pacific, the Atlantic Ocean, and the Indian Ocean also added to the warm anomaly in 2023 (Figure 3.2b). South Pacific SST patterns resemble the negative phase of the Interdecadal Pacific Oscillation [IPO; 109, 110] while Indian Ocean SST patterns indicate a positive-phase Indian Ocean Dipole [IOD; 111]. The Atlantic Ocean was characterized by wide-spread warmth, in part due to the warm-phase of AMV [112]. Thus, the internally-driven warm anomalies in 2023 global-mean SST were not only a product of a warm-phase ENSO event, but also a 'perfect storm' of internal variability leading to anomalously warm conditions across the oceans worldwide. This magnitude of global-mean internal variability is not an outlier in our observational record. Our estimate that internal variability contributed +0.07°C to 2023 global-mean SST can be compared with the conditions in 2015 and 1998 (Figure 3.3c-d). Both years featured strong warm-phase ENSO events. In 2015, internal variability accounted for a +0.11°C anomaly in global-annual-mean SST, while in 1998 internal variability accounted for +0.09°C. What sets the climate of 2023 apart from the climates of 1998 and 2015, is that its warm-phase ENSO event was much weaker than the events in 1998 and 2015 [113]. However, in 2023, warm anomalies in the other ocean basins assisted in elevating the global-mean temperature.

Our results don't preclude the possibility that abrupt change has played a role in the 2023 SST. In fact, the estimated 0.1°C change in the global-annual-mean forced component from 2022 to 2023 is the largest such change in our observational record (Figure B.3). However, three periods (1987-1988, 1996-1997, 2015-2016) experienced comparable 0.07°C-0.09°C changes in the forced component. The forced component estimate changed very little from 2016 to 2021 before the rapid warming from 2021 to 2023. It is possible that the forced response of SST was concealed during this period, but temperatures continued to rise at depth [e.g., 107]. This speaks to how difficult attribution is, and emphasizes that

our results suggest a rapid change in the detectability of the forced response in maps of annual-mean maps of SST, but not necessarily an abrupt change in the forced response itself.

A novel result of this work is that the neural network can separate internal variability and the forced response when provided only a snapshot of annual-mean SST. Unlike traditional methods for detecting the forced response, such as estimating the forced response as a linear or fourth-order polynomial fit to the data, our approach does not have access to time-varying information or recent trends in the climate. This is a much harder task, and yet, this approach yields comparable skill to a fourth-order polynomial (Figure B.6). One benefit of this sample-by-sample method is that it can detect abrupt year-to-year changes in the forced climate since it does not make any assumptions about how the climate evolves over time. This method is also robust against the availability of new observations. Where re-fitting a fourth-order polynomial to observations following the release of 2024 data would cause the fourth-order polynomial, and thus the forced response estimate, to change, our trained neural networks provide consistent estimates.

While this method provides skillful attribution for annual-mean SST, there are several considerations for future study that may further improve its performance. Our method only uses annual-mean maps of SST, but monthly or seasonal maps of SST may provide the neural network with sub-annual patterns that could be useful for separating forced and internal variability. Additional fields [such as ocean heat content, precipitation or soil moisture, e.g., 86, 107] may also contain unique patterns of internal variability and forced change. However, providing the neural networks with more information is not inherently beneficial, as it may lead to overfitting. As with any data-driven method trained on climate model simulations, this method requires that climate model representations of annual-mean SST are consistent with observations. However, we have designed our neural network to learn patterns of both the forced response and internal variability (Section B.1),

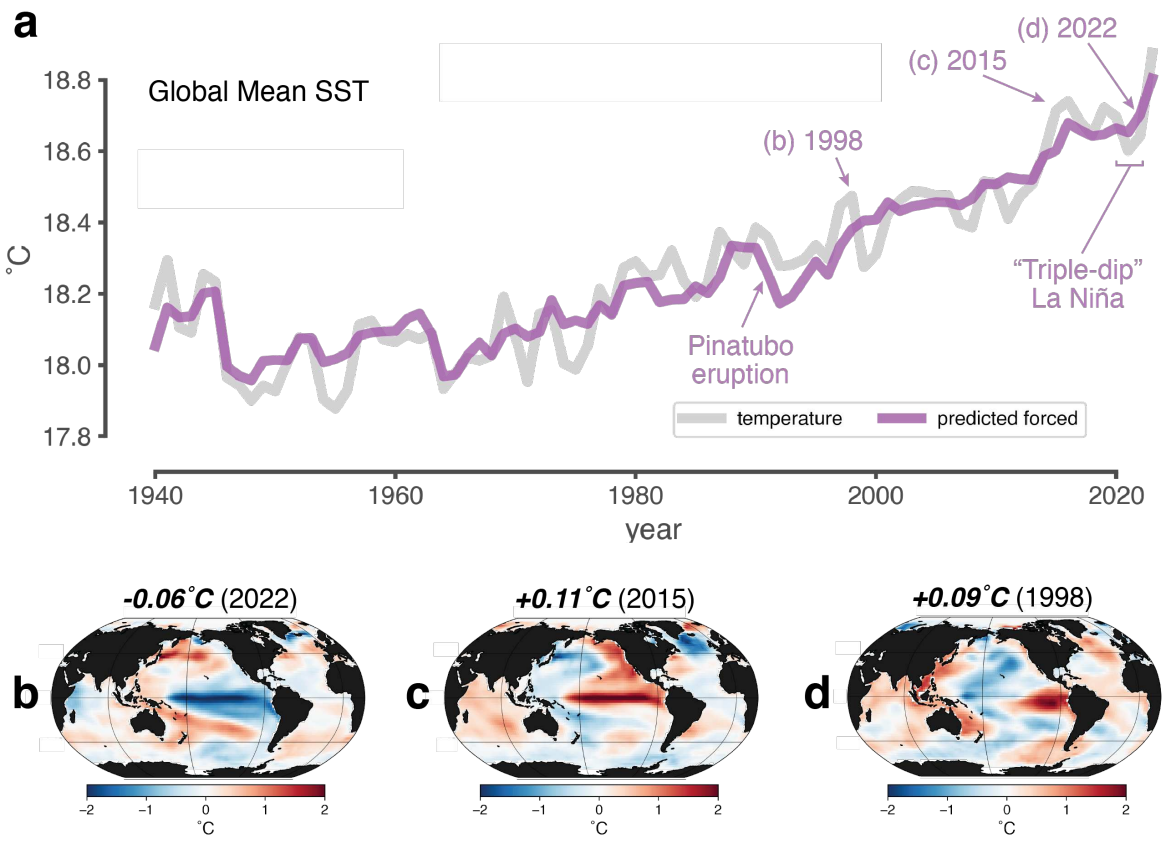

Figure 3.3: (a) Estimated forced component of observed SST, 1940-2023. Estimated internal variability for (b) 2022, (c) 2015, and (d) 1998.

.

and thus these results are more nuanced than simply projecting the simulated patterns of the forced response onto observations.

3.5 Conclusions

Annual-global-mean SST in 2023 was the warmest on record. This unprecedented extreme has prompted questions of how much of this warming can be attributed to an abrupt change in the forced response or internal variability [e.g., 66]. Our deep-learning approach finds that persistent forced warming of SST, combined with anomalously warm conditions due to internal variability, is responsible for this record high. While our method detects that the forced response of SST increased sharply between 2022 and 2023, similar trends have been observed in recent years. Therefore we cannot attribute the record-high

SST to an abrupt change in the forced response. Nevertheless, 2023 global-annual-mean SST reached a level unseen in recent history, and attribution of such extremes can provide insight into the behavior of our changing climate. Given the impacts associated with abrupt forced change on human and natural systems, future climate should be closely monitored.

Chapter 4: Data-driven predictions of the likelihood that annual global mean temperature will exceed 1.5°C, 2.0°C in the next decade

4.1 Summary

In accordance with the United Nations Paris Agreement, 194 nations have set a collective goal to restrict global warming to "well below 2.0°C" relative to preindustrial temperatures and pursue actions to limit warming to 1.5°C. The year 2023 saw the annual global mean temperature approach, and by some estimates exceed, 1.5°C. Given the climate extremes experienced in 2023, we seek to quantify the likelihood that a single year exceeds the Paris Agreement thresholds in the near-term climate. Our analysis uses neural networks that are trained on climate model simulations to make predictions based on historical observations. We find that it is *likely* (81%-99%) that the global temperature will exceed 1.5°C at least once over the next ten years (2024-2033), though it is *unlikely* ($\leq$12%) that the temperature will exceed 2.0 degrees. In addition, we are *virtually certain* (98%) to experience the warmest year on record by 2033. Our confidence in these results is reinforced by the neural networks' skillful observational hindcasts and physically consistent explanations of the networks' decisions. The global temperature in 2023 was extremely warm relative to recent variability, and our forecast suggests a high likelihood that 1.5°C will be crossed at least once in the next decade, signaling the approach of a 1.5°C climate.

4.2 Significance

A 1.5°C increase in global temperature has significant consequences for ecosystems and societies worldwide. To mitigate these impacts, 194 nations have signed the United Nations Paris Agreement, aiming to hold global warming well below 2.0°C and pursue 1.5°C. Thus, the timeline of when global temperatures will reach these temperature thresholds is of high

scientific and political interest. Given the climate extremes associated with the first single year to exceed 1.5°C in 2023, our study uses neural networks to explore the likelihood that the global temperature will exceed various temperature thresholds for at least one year in the next decade. Our framework predicts a very high likelihood of record-breaking global temperatures and signals the approach of a 1.5°C climate.

4.3 Introduction

In the 2010 Cancun Agreement, the United Nations Framework Convention on Climate Change (UNFCCC) identified a goal to restrict global warming to 1.5°C above preindustrial temperatures [78]. In 2015, this global warming threshold was reaffirmed by the UN Paris Agreement which set the goal of restricting warming to "well below 2.0°C" while pursuing actions to limit warming to 1.5°C [79]. Since the Paris Agreement, considerable attention has been given to identifying the impacts and risks that surround global warming magnitudes of 1.5°C, including the Intergovernmental Panel on Climate Change (IPCC) Special Report on Global Warming of 1.5°C [SR15; 114]. Impacts associated with 1.5°C warming include sea-level rise [115, 116], extreme precipitation [117, 118, 119], drought [120, 121], heatwaves [117, 122], and changes to terrestrial and marine ecosystems [123, 124].

The timeline of global warming is thus of high scientific and political interest [80, 125, 126]. It provides a timeline for invoking various measures of mitigation and adaptation [127] and informs decision-makers on the extent to which we have committed to irreversible climate impacts [e.g., 128] and what emissions pathways are required to meet the Paris Agreement goals [e.g., 129, 130]. This timeline is particularly relevant now. The year 2023 was the warmest in the modern record, and the annual global temperature approached, and by some estimates exceeded, 1.5°C [131]. This annual global temperature record was accompanied by major impacts globally, including heatwaves [132, 133], record-breaking monthly temperatures [134, 135], and record-low sea ice extent in the Antarctic [136, 137]. The annual global temperature is the product of both the forced warming trend and

temperature variability internal to the climate system [66], and while the 2023 global mean temperature does not alone indicate that the mean state of the climate has approached or breached the 1.5°C threshold [80] it may serve as a harbinger for future warming [19, 125, 129].

The World Meteorological Organization (WMO), under the recommendation of the World Climate Research Programme, develops predictions for the near-term global climate using a multi-model ensemble of climate models that are initialized with observations [19]. These initialized forecasts are made in real time and allow current knowledge of Earth system variability—"internal variability"—to guide the forecasts alongside external forcings. Predictions of the state of the global climate for the next 5 years, including the probability that annual global temperatures exceed 1.5°C at least once, are released annually via the Global Annual to Decadal Climate Update. These near-term forecasts are framed within the context of the Paris Agreement's 1.5°C threshold of warming to "provide a warning that 1.5°C warming of the mean climate is being approached" [19].

Climate models initialized with observations have difficulties retaining the important information granted by those initial conditions due to the impacts of initialization shock and climate model drift [6, 29]. Data-driven approaches offer cost-effective alternatives to initialized decadal prediction systems while simultaneously circumventing some of these issues intrinsic to initialized climate models [41].

In this work, we take advantage of the plethora of open-access climate model simulations to make data-driven predictions of the likelihood that global temperature thresholds are exceeded for at least one year in the near-term climate. Similar to the WMO approach, these thresholds may be reached through a combination of the externally forced response and internal climate variability. Our predictions are made with neural networks that use the current state of temperature to predict the maximum annual global temperature over the next one, five, and ten years. Our neural networks are trained entirely on climate model simulations and are then used to produce forecasts for the historical climate based on

out-of-sample observational inputs. Through this process, our data-driven method mimics the dynamical forecasts used in the WMO Global Annual to Decadal Climate Update, as our method uses patterns learned from climate model physics to predict the future maximum annual global temperature given observed initial conditions. We use these forecasts to assess the likelihood that any single year's annual global mean temperature will exceed the Paris Agreement temperature thresholds in the near term.

4.4 Data-Driven Prediction Framework

Using machine learning, data from climate model simulations, and historical climate observations, our analysis predicts the likelihood that 1.5°C, 1.7°C, and 2.0°C will be exceeded for at least one year in the next N = 1, 5, 10 years. The neural networks in this study make two predictions: 1) a "base" prediction, which is based only on the relationship between the global mean temperature over the past 10 years and the maximum annual global mean temperature over the next N years, and 2) a "final" prediction, which modifies the base prediction using recent initial conditions of temperature (Figure 4.1). Our temperature initial conditions are maps of annual mean temperature the year preceding the N-year forecast window. The predictions comprise two parameters, a mean (μ) and a standard deviation (σ), which describe the central prediction and its uncertainty according to a Gaussian distribution. The difference in the final prediction from the base prediction is discussed in terms of a shift factor ($\Delta\mu$, how μ changes with the addition of initial conditions) and an uncertainty scaling factor (ϵ, what fraction of the base prediction uncertainty remains with the addition of initial conditions). Integrating over the predicted Gaussian for all values greater than a temperature threshold produces a likelihood that the threshold will be exceeded for a given forecast (Figure 4.1b). (See Section 4.9 for further details.)

Our neural networks are trained on climate models from the Coupled Model Intercomparison Project Phase 6 [CMIP6; 138]. We train on both the "high" SSP3-7.0 forcing

scenario as well as on the "intermediate" SSP2-4.5 forcing scenario [which have similar global forcings over the next decade; 139], and find that the results are consistent (see Chapter C). Despite differences between climate models, such as in the spatial pattern and time evolution of temperature, the neural network produces skillful predictions for climate models withheld from the training set (Figure C.1). To produce forecasts for the observed Earth system, we then input observations of temperature into the trained neural networks. To maintain consistency with the vocabulary of initialized dynamical forecasts, we hereafter refer to this as "initializing the network with observations."

4.5 Neural Network Hindcasts

We produce hindcasts by initializing the neural network with temperature observations from 1980-2022. Our observational hindcasts skillfully predict the maximum annual global temperature for one-, five-, and ten-year prediction windows (Figure 4.2). The mean absolute error (MAE) of the final neural network predictions is 0.06°C for one-year hindcasts, 0.08°C for five-year hindcasts, and 0.07°C for ten-year hindcasts, meaning that the network's central estimates are close to the true observed values for all three prediction windows. The errors in the central estimates are well captured by the prediction uncertainties. Across all of the predictions, only one year falls outside the range of two standard deviations from the central estimate: the one-year prediction of 1992, initialized with observations from 1991 (Figure 4.2Ab). Mount Pinatubo erupted in 1991 causing a decrease in global mean temperatures thereafter [140, 141] and our neural network is not explicitly designed to predict sporadic geological activity and its climate response.

This architecture, which makes a base prediction and a final prediction, separates the contributions of the global mean climate and current temperature patterns to the forecasts of maximum annual global temperature. This introduces an element of interpretability to our neural networks in that the networks inherently provide an explanation for their decisions [25]. Comparing the base and final predictions reveals how the prediction changed once the

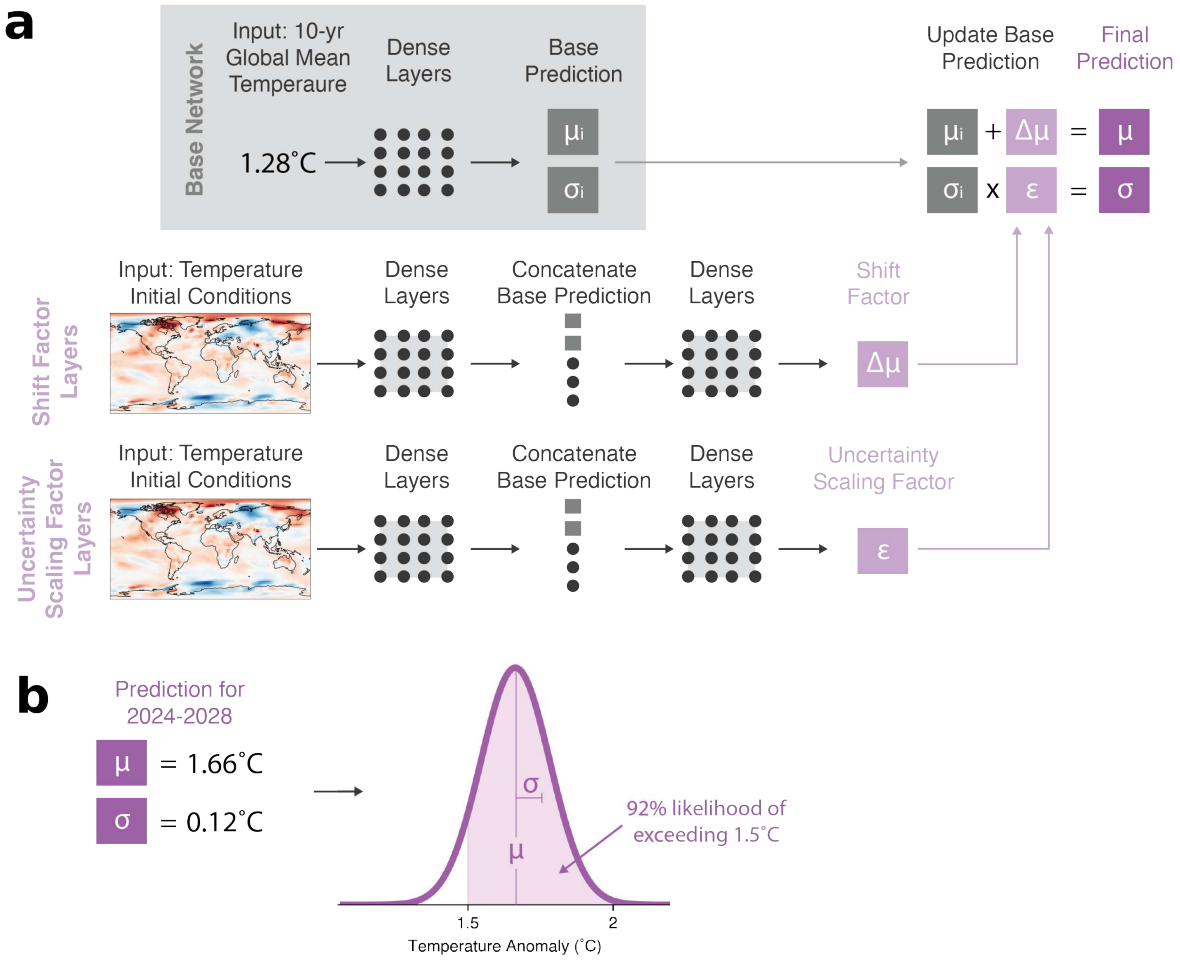

Figure 4.1: (a) Final neural network architecture, including the base network which is trained first on the global annual mean temperature over the previous 10 years, and the shift factor and uncertainty scaling factor layers of the final network which update the base prediction using the spatial patterns of temperature during the most recent year. (b) Depiction of the process for converting a neural network prediction into a likelihood. Here we show the likelihood the maximum temperature will exceed 1.5°C for the period 2024-2028.

network was allowed to update its prediction given recent initial conditions of temperature. We use two metrics to assess the skill of our predictions: MAE (°C), the mean absolute error between the central prediction (μ) and the truth, and loss (unitless), which is the negative log likelihood of the probability density of the predicted Gaussian evaluated at the truth [142]. In the following, we find that initial conditions improve the skill of our one-year forecasts, and to a lesser extent our five-year forecasts, but do little to improve the skill of our ten-year forecasts.

One-year hindcasts of global mean near-surface air temperature are highly sensitive to the initial state of the climate the prior year (Figure 4.2A). The final network hindcasts track the true temperatures much more closely than the base network hindcasts, as both the loss (-1.08) and MAE (0.06°C) indicate improved performance on observations. Including the temperature initial conditions as a predictor for the final network leads to $\pm$0.2°C shifts ($\Delta\mu$) from the base prediction and half of the uncertainty ($\epsilon\approx$0.5) of the base prediction (Figure C.2a). The final network is able to more confidently ($\epsilon<1$) predict some of the largest departures from the base network prediction, including the anomalously warm 1998 and 2016 El Niño years, and the anomalously cold 2011 La Niña year (Figure 4.2Ac-e). Curiously, the final prediction does not predict a shift in the central estimate despite the fact that the year 2023 is much warmer than the base prediction (Figure 4.2Af). Other agencies failed to predict the 2023 temperature anomaly as well [66]. While there have been theories that the warmer-than-expected global mean temperature in 2023 was the result of anomalous external forcings, such as the Hunga Tonga–Hunga Ha'apai volcanic eruption in Tonga or decreased aerosol emissions from shipping, this is still an open area of research [66, 68, 112].

Five-year hindcasts of maximum global mean temperature are slightly improved by maps of annual mean temperature one year prior to the forecast window (Figure 4.2B). In observations, recent temperature patterns are responsible for $\pm$0.15°C shifts ($\Delta\mu$) from the base prediction and approximately 85% of the uncertainty ($\epsilon\approx$0.85) of the base prediction

(Figure C.2b). Over the 1980-2020 hindcast period, the final network predictions were not an improvement over the base network prediction, as the loss is nearly identical (-0.84). However, we only have about eight independent samples in our observations (40 years divided by 5-year forecasts), so the apparent lack of improvement may be due to the limited sample size. The loss on the validation set (1,300 samples) decreases from -0.62 for the base network to -0.70 for the final network (Figure C.1c,d).

While initial conditions modify the hindcasts for one- and five-year prediction windows, predictions of the ten-year maximum temperature are minimally impacted (Figure 4.2C). The loss for the base and final networks are almost identical for both the 1980-2020 observations (Figure 4.2Ca,b) and the validation set (Figure C.1e,f). Though the final neural network has access to spatial temperature information, it does little ($\Delta\mu$<0.06, ϵ=1) to modify the base prediction (Figure C.2c).

4.6 Neural Network Forecasts for 2024-2033

The neural network hindcasts show that our neural networks make skillful predictions for the observed Earth system. Using the same neural networks we create near-term climate forecasts for the real world. We initialize the neural network with temperature observations from 2023 (Figure 4.3a) to produce forecasts of the likelihood of exceeding annual global temperature thresholds over the next year (2024), the next five years (2024-2028), and the next ten years (2024-2033). We express these forecasts in relation to preindustrial temperatures by using the Berkeley Earth estimate that 2023 was 1.54°C warmer than preindustrial temperatures and express likelihood using the IPCC guidance for consistent treatment of uncertainties [143]. Our forecasts suggest that it is *likely* (70%) that the global mean temperature in 2024 exceeds 1.5°C, though it is *unlikely* (4%) and *exceptionally unlikely* (<1%) that the global mean temperature in 2024 exceeds 1.7°C and 2.0°C, respectively (Figure 4.3b). In addition, this analysis predicts that it is *as likely as not* (53%) that 2024 will be the warmest year on record.

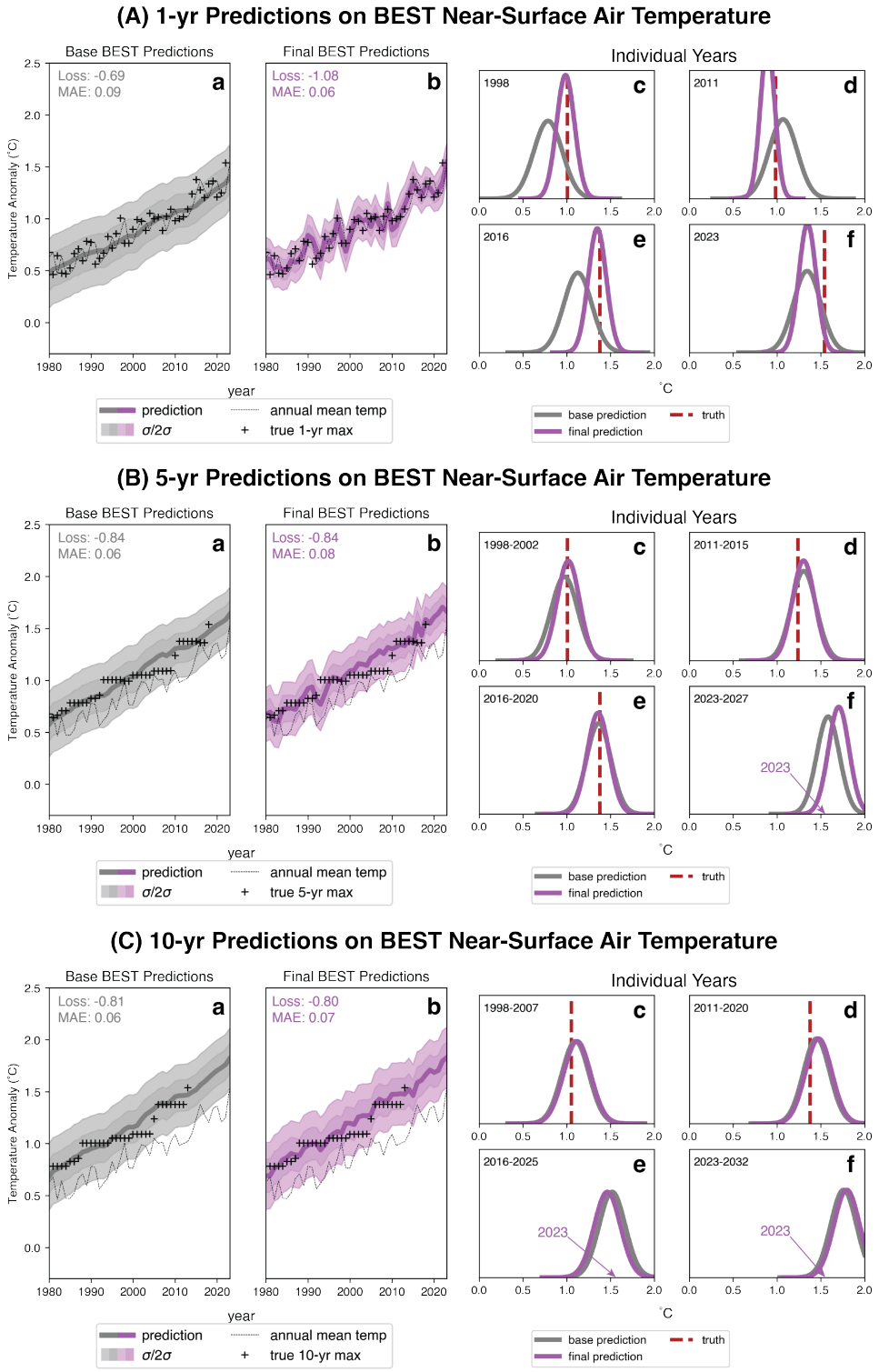

Figure 4.2: Hindcasts of the (A) 1-year, (B) 5-year, and (C) 10-year maximum annual global mean temperature for (a) the base network, and (b) the full network initialized with temperature data from BEST. Hindcast predictions for periods beginning in (c) 1998, (d) 2011, (e) 2016, and (f) 2023. The global mean temperature for 2023 (1.54°C) is indicated on the plots for prediction periods that cannot yet be verified.

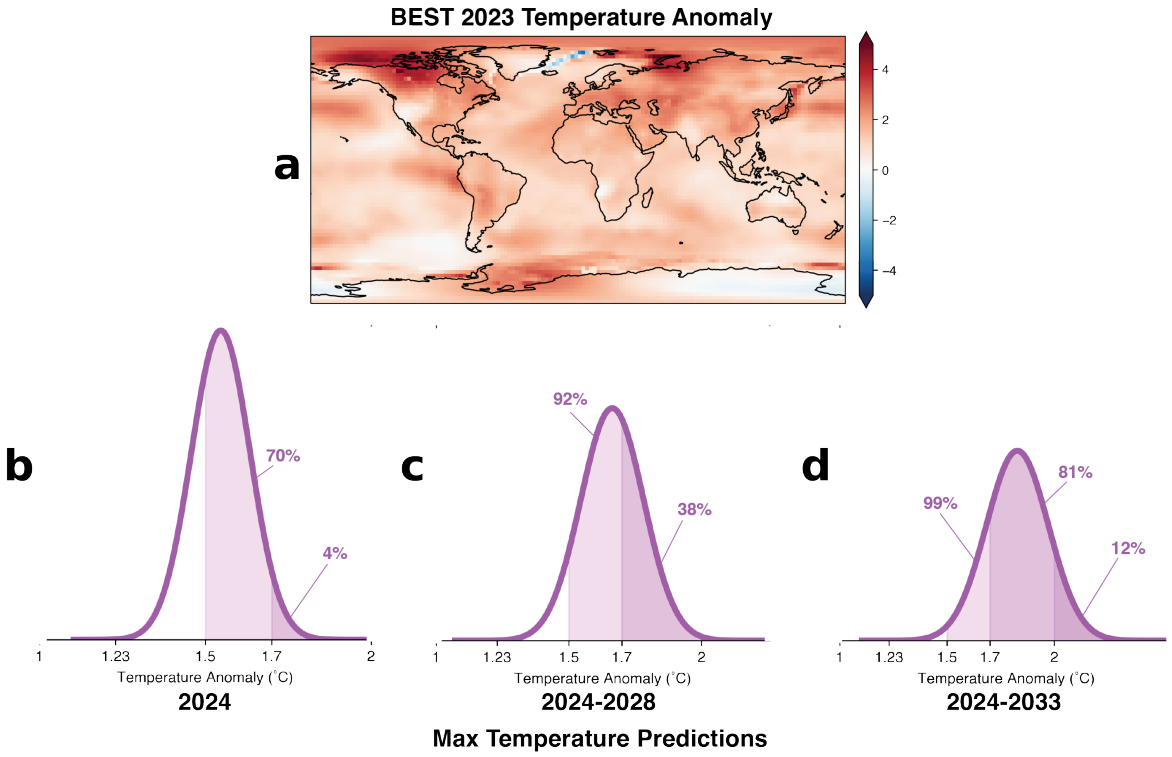

Figure 4.3: (a) BEST mean temperature for 2023 relative to the 1980-2010 mean. Predictions of maximum annual global mean temperature over the next one (b), five (c), and ten (d) years. Likelihoods of exceeding 1.5°C, 1.7°C, and 2.0°C are indicated on b-d.

Looking out over the next five years (Figure 4.3c), it appears *very likely* (92%) that at least one annual global temperature between 2024 and 2028 will exceed 1.5°C, though it is still *unlikely* (38%) that the annual temperature will exceed 1.7°C, and *exceptionally unlikely* (<1%) that the annual temperature will exceed 2.0°C (Figure 4.3c). However, it is *likely* (85%) that one year in the next five will be the warmest since 1850.

Over the next decade (2024-2033), it is *virtually certain* (99%) that at least one year will exceed 1.5°C (Figure 4.3d). The annual global temperature is *likely* (81%) to surpass 1.7°C in the next ten years and it is *very likely* (98%) that we will experience the warmest year on record, but it is still *unlikely* (12%) that the global temperature will surpass 2.0°C.

These results come with the consideration that they are based on the Berkeley Earth estimate of preindustrial temperature, which is just one of many estimates. While the Paris Agreement does not explicitly define which measure of "preindustrial temperature"

should be used to calculate global warming to date, the IPCC and many other entities use the 1850-1900 period to define the preindustrial mean global temperature [144, 81]. However, even defining the preindustrial temperature as the global mean temperature from 1850-1900 leaves considerable uncertainty. Estimates across agencies suggest that 2023 was anywhere between 1.34°C and 1.54°C warmer than the 1850-1900 average [131]. This 0.2°C range can lead to very different mitigation and adaptation timelines [145] and, in the context of temperature forecasting, it can lead to different likelihoods of exceeding specific temperature thresholds.

Other products' estimates of the 1850-1900 baseline temperature are warmer than Berkeley Earth's (these estimates are provided in Materials and Methods). The forecasted likelihoods of exceeding 1.5°C, 1.7°C, and 2.0°C given the six agencies' estimates of the 1850-1900 global temperature are shown in Table 4.1. Across these estimates, there is a 4%-70% likelihood the global temperature exceeds 1.5°C in 2024. The likelihood that the 2024 global temperature exceeds 1.7°C is $\leq$4%, and the likelihood it exceeds 2.0°C is $<$1%. In the next five years (2024-2028), the ranges of likelihood that at least one year exceeds 1.5°C, 1.7°C, and 2.0°C are 38%-92%, 2%-38%, and $<$1%, respectively. In the next ten years (2024-2033), the ranges of likelihood that at least one year exceeds 1.5°C, 1.7°C, and 2.0°C are 81%-99%, 31%-81%, and $\leq$12%.

These results are robust to the choice of observational data set (Figure C.3), the choice of the forcing scenario used for training (Figure C.4), and hyperparameter decisions, such as the train/validation split and the seed used to initialize the networks' weights and biases (Text S1-S2). In addition, we use an explainable artificial intelligence (XAI) method to further explain the neural networks' decisions and increase our trust in these forecasts. The Integrated Gradients XAI method [146, 147] identifies which regions of the globe are most important for the neural networks' predictions. The resulting XAI heatmap reveals regions in the temperature initial conditions that contributed to a shift (either warmer or cooler) in the final network's prediction (Figure 4.4). For the one-year forecast of 2024, the

tropical Pacific contributes to a warmer prediction. For the five-year forecast for 2024-2028, the North Atlantic and the North Pacific contribute to a warmer prediction, while the Weddell Sea and the Greenland Sea comprise areas that contribute to both warmer and cooler predictions. XAI heatmaps highlighting which regions are responsible for changes in the uncertainty of the final prediction can be found in Figure C.5.

4.7 Discussion

In this work, we use neural networks that are trained on climate model simulations of temperature to make near-term forecasts of the maximum annual global temperature over the next one, five, and ten years, using observed temperature data as input. Our neural networks take two inputs, the global mean temperature over the last ten years and the spatial patterns of temperature over the previous year. By using recent temperature observations as input, our method allows initial conditions to inform the forecasts, which aligns with the procedure of initialized decadal prediction.

There are some clear benefits to this data-driven approach. Whereas initialized decadal prediction systems are computationally costly to run, these data-driven networks can be trained on a standard workstation using publicly available climate model output. In addition, this method learns patterns from a suite of 30 climate models with different behaviors and representations of the climate system. Unlike forecasts from a single initialized climate model, the neural network may identify distinct patterns within subsets of the CMIP6 simulations. The neural network may then use patterns from climate models that most resemble those seen in observations (e.g., in their representation of global temperature variability) rather than allowing each climate model to contribute equally. Thus, our method is more nuanced than taking the average prediction across 30 climate models.

While the Paris Agreement sets a goal to restrict global warming to 1.5°C and well below 2.0°C relative to preindustrial temperatures, it does not explicitly define the preindustrial temperature baseline to use. Estimates of the 1850-1900 preindustrial temperature, and

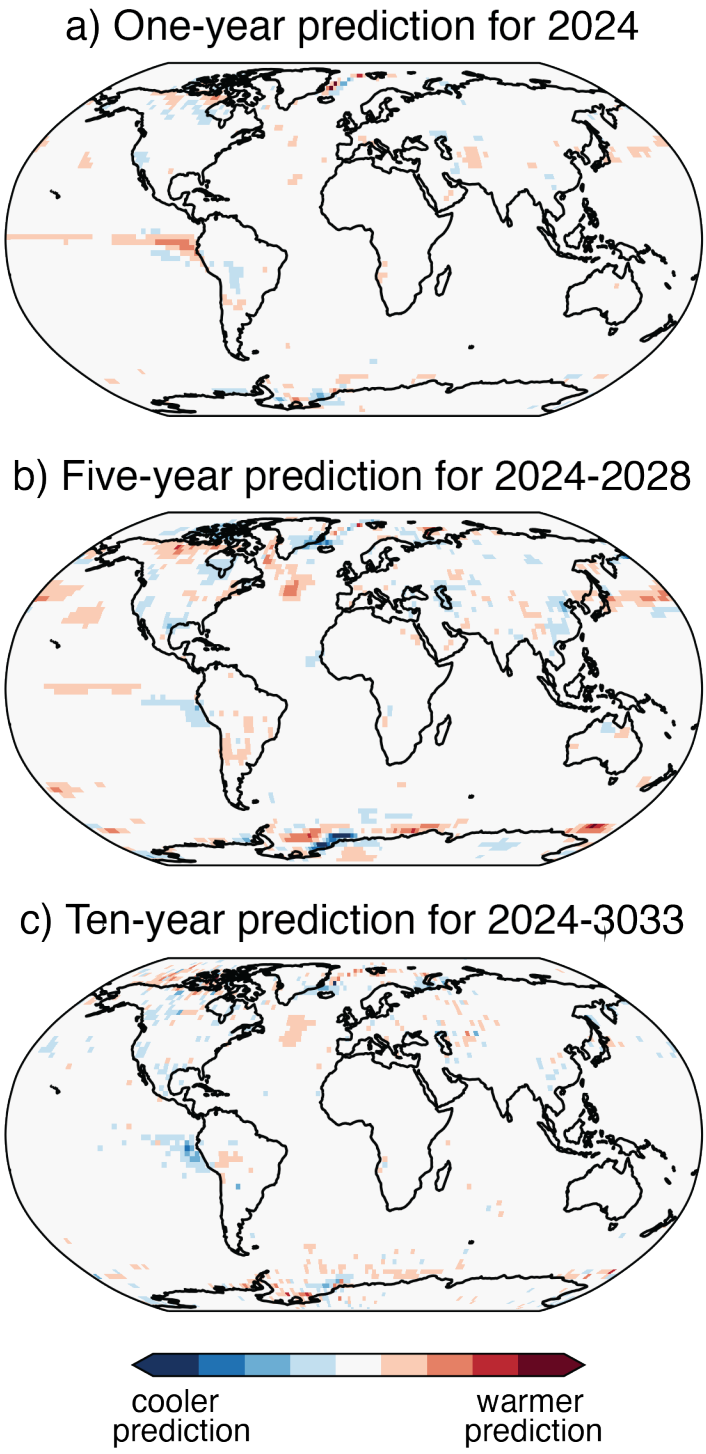

Figure 4.4: XAI heatmaps for forecasts of maximum global temperature over the next (a) one, (b) five, and (c) ten years. Reds indicate regions where the climate patterns contribute to a warmer central prediction and blues indicate regions where the climate patterns contribute to a cooler central prediction.

estimates of the global warming to date, vary considerably between products. However, regardless of which estimate of the preindustrial temperature we use for our forecasts, the general conclusion remains the same. We find it *likely* (81%-99%) that 1.5°C will be exceeded for at least one year in the next decade, though it is *unlikely* ($\leq$12%) that 2.0°C will be exceeded. These forecasted likelihoods indicate that future years will break the global temperature record set in 2023. Furthermore, the high likelihood that at least one year will exceed 1.5°C, paired with the fact that 2023 was extremely warm relative to recent variability, signals the approach of a 1.5°C climate [125].

Our method performs well on historical observations, and there is still a possibility for further improvements. Decadal prediction systems use several atmospheric and oceanic variables to initialize their dynamical models. In this work, we have used only patterns of near-surface air temperature to initialize our data-driven forecasts. Near-surface air temperature, which is highly correlated with sea surface temperature and also provides terrestrial temperature information, is useful for predictions on multi-annual timescales [6]. Oceanic variables, such as ocean heat content or sea surface height, could contribute to further predictive skill [e.g., 20, 148]. In addition, our initial conditions comprise annual mean fields of temperature, though each season may provide different information that is useful for predicting maximum annual global temperature. We leave these refinements of the input variables for future exploration.

Our confidence in these predictions, which comes from the skill of the neural networks on validation data and historical observations (Figure C.1; Figure 4.2), is further reinforced by the fact that the explanations of the neural networks' decisions are consistent with knowledge of the climate system. Explanations of the neural networks' decisions are important for assessing the trustworthiness of the predictions (that is, to verify that the neural network is making the right predictions for the right reasons) as well as to further our scientific understanding [24, 25, 26]. Here, our explanations come from two sources: interpretability provided by the neural network architecture and an XAI method. The

Table 4.1: Likelihoods of exceeding 1.5°C, 1.7°C, and 2.0°C over the next one, five, and ten years, from the final network. These likelihoods vary based on each agency's estimate of the 1850-1900 preindustrial baseline temperature. Regardless of the baseline, the neural network predicts we are *likely* to exceed 1.5°C, but *very unlikely* to exceed 2.0°C, in the next 10 years.

	2024					
Threshold	BEST	ECMWF	HadCRUT	WMO	NASA	NOAA
1.5°C	70%	44%	35%	31%	8%	4%
1.7°C	4%	<1%	<1%	<1%	<1%	<1%
2.0°C	<1%	<1%	<1%	<1%	<1%	<1%

	2024-2029					
Threshold	BEST	ECMWF	HadCRUT	WMO	NASA	NOAA
1.5°C	92%	81%	76%	73%	48%	38%
1.7°C	38%	21%	16%	14%	4%	2%
2.0°C	<1%	<1%	<1%	<1%	<1%	<1%

	2024-2033					
Threshold	BEST	ECMWF	HadCRUT	WMO	NASA	NOAA
1.5°C	99%	97%	96%	95%	86%	81%
1.7°C	81%	68%	63%	61%	39%	31%
2.0°C	12%	6%	4%	4%	<1%	<1%

explanations provided by the interpretable element of the network architecture reveals that initial conditions are more useful for our one-year forecasts than our ten-year forecasts. This is consistent with our knowledge that chaos in the Earth system reduces the predictability added by initial conditions at longer leads [36]. Furthermore, the regions highlighted by the XAI method are consistent with known precursors for global mean temperature on multi-annual timescales. These include El Niño Southern Oscillation in the tropical Pacific [105] [e.g., 149, 150, 112], Atlantic Multidecadal Variability in the North Atlantic [151, 15, 16] [e.g., 152, 153, 154], and Pacific Decadal Variability in the North Pacific [8, 101] [e.g. 150, 153, 154, 155]. The Weddell and Greenland seas may too be related to Atlantic Multidecadal Variability [e.g., 156, 22], via teleconnections such as through the Atlantic Meridional Overturning Circulation [157]. These are also regions that exhibit significant temperature variability related to changes in sea ice extent [158, 159], which may be useful for our neural network predictions.

4.8 Conclusions

The UN Paris Agreement set an international goal to restrict global warming to "well below 2.0°C" relative to preindustrial temperatures while pursuing actions to limit warming to 1.5°C [79]. Using neural networks trained on climate model simulations, we create data-driven forecasts to predict the likelihood that a single annual global mean temperature anomaly will exceed these temperature thresholds in the near-term climate. Despite the fact that the global temperature in 2023 was extremely warm relative to recent variability [66], our forecasts indicate that annual global temperature will reach new extremes in the next decade. Confidence in these forecasts is reinforced by skillful predictions on out-of-sample historical observations and physically consistent explanations of the neural networks' decisions. Furthermore, these findings are robust across estimates of preindustrial temperature, forcing scenarios, and observational products. Our predictions of the likelihood that annual global temperature anomalies will exceed specific temperature thresholds do not indicate that the UN Paris Agreement will be exceeded in the next decade, as the Paris Agreement thresholds generally refer to the average temperature anomaly over 20 years [e.g., 80, 81]. However, our forecasts do indicate the proximity of the annual global mean temperature to the Paris Agreement threshold [125] and signal the approach of a 1.5°C climate. Given the substantial impacts on global ecosystems and societies that are likely to occur with long-term warming of 1.5°C [114], our results suggest a trend towards intensified climate risks in the coming years.

4.9 Materials and Methods

4.9.1 Datasets

We use near-surface air temperature data from simulations of historical and future forcing scenarios made available by the CMIP6 [138]. The CMIP6 archives the state-of-the-art climate model simulations used in the IPCC Assessment Report 6 [81]. The simulations of temperature under historical forcings cover the years 1850-2014; the future forcing

scenarios cover the years 2015-2100. We use two future forcing scenarios, SSP3-7.0 and SSP2-4.5, which are identified as the high and intermediate forcing scenarios by the IPCC [95].

In addition to simulated climate data, we use historical temperatures from three observational data sets in this study. These are the Berkeley Earth Surface Temperature [BEST; 160], the European Centre for Medium-Range Weather Forecasts Reanalysis version 5 [ERA5; 161] and the Goddard Institute for Space Studies Surface Temperature product version 4 [GISTEMPv4; 162, 163]. We re-grid all data to 2.5°latitude by 2.5°longitude using the ESMF_regrid package in the NCAR Command Language.

4.9.2 Data Preparation

We use simulations from 30 climate models, one realization each, for training the neural network. This ensures that our training set covers a wide range of climate models, which have different biases and representations of forced change and internal variability. We use 20 climate model simulations as our training set, while 10 others are withheld as a validation set. The climate models, as well as whether they appeared in the training or validation set, can be found in Section C.1. We concatenate the historical and SSP3-7.0 forcing scenarios and train on the period 1950-2080. This training period includes annual global temperature anomalies ranging from -0.1°C to 5.2°C. Our neural networks are designed to make forecasts for the near-term climate (2024-2033) and hindcasts for the historical climate (1980-2023), and this range of years in the training set guarantees that we capture the lowest annual global mean temperature in the 1980-2023 historical observations (0.46°C) as well as our highest temperature threshold (2.0°C). In addition to results for neural networks trained on the historical and SSP3-7.0 forcing scenarios, which are presented in the main text, we also show the results for a training set comprising the historical and SSP2-4.5 forcing scenarios in Figure C.4.

For both the inputs and outputs of the neural network, we normalize the data by computing the anomalies for each climate model or observational data set relative to its own 1980-2010 mean. Thus, the neural network's prediction of maximum global mean temperature is relative to the 1980-2010 mean. We then convert the predictions to a departure from the preindustrial baseline temperature from 1850-1900 by adding the estimated temperature difference between 1850-1900 and 1980-2010. As discussed in the next section, this estimate varies across observational products.

4.9.3 Annual Global Mean Temperature Anomalies

We use the period 1850-1900 to determine the preindustrial temperatures as it is a period before the majority of anthropogenic warming [144] and was used by the IPCC Assessment Report 6 [81]. Given the sparsity of surface temperature observations in the period 1850-1900, different products have different estimates of the 1850-1900 global mean temperature. We use six different estimates of the 1850-1900 temperature which come from the following agencies and products: the U.S. National Oceanographic and Atmospheric Administration (NOAA) Merged Land Ocean Global Surface Temperature Analysis version 5 [NOAAGlobalTemp v5; 164, 165], the U.S. National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies Surface Temperature product version 4 [GISTEMPv4; 162, 163], the World Meteorological Organization (WMO) average of six datasets [166], Met Office Hadley Centre Climatic Research Unit Global Surface Temperature version 5 [HadCRUT5; 167], the European Centre for Medium-Range Weather Forecasting (ECMWF) Reanalysis version 5 [ERA5; 161], and BEST [160, 131]. We use the NOAA estimate that the 1980-2010 period was 0.57°C warmer than 1850-1900, the NASA estimate of 0.60°C, the WMO estimate of 0.68°C, the Met Office Hadley Centre estimate of 0.69°C, the ECMWF estimate of 0.71°C, and the BEST estimate of 0.77°C [131].

4.9.4 Neural Networks

We train our neural networks on two inputs: the ten-year global mean temperature and the map of annual mean temperature directly preceding the forecast window. The base neural network (Section 2) takes only the ten-year global mean temperature as input. This is passed through trainable fully connected dense layers en route to outputting two values, a mean (μ_b) and a standard deviation (σ_b), representing the mean and standard deviation of a Gaussian distribution. As shown in Figure C.6, this Gaussian distribution is appropriate for predicting global maximum temperature over the one-, five-, and ten-year forecast windows. Once trained, the weights of the base network layers are frozen. The base network is a sub-model of the final neural network, and its output is used en route to the final prediction (Figure 4.1). The final network comprises two main components which update the base network prediction: the shift factor layers and the uncertainty scaling factor layers.

The shift factor layers take the annual mean map of temperature as input. This input is passed through a set of dense layers. The base predictions (μ_b, σ_b) are concatenated to the final layer of this first set of dense layers, then fed through a second set of dense layers. At the end of the second set of dense layers, the network outputs a shift factor, $\Delta\mu$. The same procedure is used for the uncertainty scaling factor layers en route to the uncertainty scaling factor output ϵ. The final prediction from the neural network uses both the base prediction (μ_b, σ_b) and the shift and scaling factors ($\Delta\mu$, ϵ) to produce a final prediction where the final prediction mean $\mu = \mu_b + \Delta\mu$ and the final prediction standard deviation $\sigma = \epsilon\sigma_b$. The layers used to learn $\Delta\mu$ and ϵ are trained separately to prevent the network from learning one parameter at the expense of the other. This training procedure and the neural network architecture are further discussed in Section C.1.

The neural networks are coded and trained using Tensorflow 2.7.0 and Tensorflow-Probability 0.15.0. The neural networks underwent extensive hyperparameter tuning, which is described in further detail in Section C.2. The neural network forecasts vary

slightly for different architectures and hyperparameter choices. The predictions from the neural networks chosen for this paper are representative of the results across hyperparameter choices while also performing well on the validation set of climate models and on observational hindcasts (Figure C.7). The hyperparameters and search space for each of the neural networks can be found in Tables C.2 and C.3.

4.9.5 XAI Heatmaps

We use the XAI method Integrated Gradients [146, 147] to explain what temperature initial conditions are important for the forecasts of global maximum temperature for 2024 and beyond. Integrated Gradients is an attribution method which attempts to predict how the temperature at each grid point contributes to the final prediction relative to the prediction for a reference state. We identify our reference state by linearly regressing observed maps of annual mean temperature on the ten-year global mean temperature. Our reference state is the map corresponding with the ten-year (2014-2023) global mean temperature. Positive regions in the heatmap can be interpreted as highlighting where temperature anomalies in the initial conditions cause the final network prediction to be warmer than the reference state. The interpretation is opposite where the heatmap is negative.

4.9.6 Data, Methods, and Software Availability

Code is available on GitHub at https://github.com/jaminrader/TemperatureThresholds2024.

Chapter 5: Conclusion

This dissertation explores how interpretable neural networks can be used for *skillful*, *trustworthy*, and *insightful* seasonal-to-decadal climate prediction. Chapter 2 outlines a methodology that epitomizes the bridging of human and artificial intelligence (AI) in climate science. In this work, I combined an intuitive forecasting method, analog forecasting, with a neural network to improve predictions on seasonal-to-decadal timescales. This neural network learned a weighted mask, which allowed analogs to be selected based on regions that are important for predictability while ignoring regions that are not. I found that the weighted analogs deliver more *skillful* forecasts than traditional model-analog methods, while providing inherent interpretability in the form of the weighted mask. In addition to an improved forecast, the weighted mask highlighted important precursor regions of seasonal-to-decadal climate variability, contributing valuable *insight* into the behavior of our teleconnected climate system. These precursors were consistent with known drivers of enhanced predictability, building *trust* that these improvements to analog forecasting are consistent with real-world physics.

The work in Chapter 2 has opened several avenues for future research. While this neural network-informed analog forecasting approach highlights precursor regions, it does not explicitly describe why these precursors enhance predictability. The weighted mask in this work is fixed, requiring that the neural network identify the same precursor regions regardless of the climate state. Given that different phases of internal variability may have unique drivers (e.g., a warm-phase ENSO event may have different precursors than a cold-phase ENSO event), state-dependent weighted masks may allow for even further improved analog forecasting. Indeed, these results have already prompted others to investigate how neural network-learned state-dependent masks can improve analog forecasts and help us understand the dynamics behind predictable patterns of climate variability in our atmosphere and oceans [e.g., 168]. Model analogs have also recently been

extended to other problems, such as understanding the drivers of fire weather conditions in the United States [169], suggesting that this approach could be used to identify precursors of climate extremes.

Chapter 3 extended the idea of bridging human and artificial intelligence to the very current issue of global warming. Motivated by the ForceSMIP workshop [89], I used neural networks to separate the forced and internal components of climate variability and provide attribution for the record-high SST in 2023. This chapter found that internal variability was responsible for warm anomalies in the Pacific, Atlantic, and Indian Oceans, which, on top of the global warming signal, led to the unprecedented warmth in 2023. This methodology demonstrated that neural networks can *skillfully* learn time-evolving patterns of internal and forced change from single maps of annual-mean SST. The patterns of internal variability identified by the neural network, and comparison with similar past events, provide *insightful* and *trustworthy* context for the current state of the climate.

While the Paris Agreement establishes an international goal to restrict the magnitude of global warming to well below 2.0°C, there is not a widely agreed-upon way to assess the current magnitude of global warming. Active monitoring of the climate is necessary for informing climate change policy and mitigation efforts. Chapter 3 emphasized that neural networks may be used to detect climate change since they can identify complex patterns and separate internal variability from the forced response. However, for neural networks to be used for climate monitoring, they must be developed with interpretability, such that the global warming estimates can be trusted by the scientific community and beyond. Future work incorporating novel machine learning techniques [e.g., 170] and additional climate fields, may produce unique perspectives for our changing environment.

Transitioning from the present state of the climate to the future state of the climate, Chapter 4 provided data-driven predictions of the likelihood that annual-global-mean surface temperature will exceed 1.5°C (and 2.0°C) relative to preindustrial conditions. This work showed that neural networks trained on publicly available climate model data

can provide *skillful* forecasts of annual-global-mean temperature out to 10 years. The interpretable component of the neural network architecture provided *insight* into how initial conditions inform decadal forecasts. While initial conditions constrained forecasts of annual-global-mean temperature out to five years, they did little to improve ten-year forecasts. *Trust* in these results comes from the fact that they're consistent with chaos theory and recent literature [36, 17].

The data-driven forecasts of annual-global-mean temperature contained in Chapter 4 are remarkably similar to the dynamical forecasts of the World Meteorological Organization (WMO). This data-driven approach suggests that there is a 73% likelihood that global-mean temperature exceeds 1.5°C in the next five years, and an 85% chance that we experience the warmest year on record. The WMO predicts 80% and 86%, respectively [171]. This provides two lines of evidence that annual-global-mean temperature will likely exceed 1.5°C in the next decade. The fact that data-driven forecasts are far less computationally expensive, but give the same results, demonstrates that operational forecasting could include interpretable neural networks, both for making and explaining predictions out to 10 years.

Neural networks are powerful tools for prediction. When properly designed, they are *skillful*, *insightful*, and *trustworthy*. This dissertation showed that interpretable neural networks can identify teleconnections and precursor patterns in the ocean and atmosphere, separate internal variability and the forced response, and make faithful estimates of climate out to ten years, at minimal computation expense. Interpretable neural networks create new opportunities for seasonal-to-decadal prediction, enabling skillful forecasting and a more complete understanding of the climate system we live in. Building and using interpretable neural networks requires intention and intuition, thus it is imperative that AI climate research include domain scientists trained in climate dynamics and change. Interpretable neural networks facilitate deeper discovery of our chaotic, interconnected, predictable Earth.

Bibliography

[1] Richard B Alley, Kerry A Emanuel, and Fuqing Zhang. Advances in weather prediction. *Science*, 363(6425):342–344, January 2019. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aav7274.

[2] Reindert J Haarsma, Malcolm J Roberts, Pier Luigi Vidale, Catherine A Senior, Alessio Bellucci, Qing Bao, Ping Chang, Susanna Corti, Neven S Fučkar, Virginie Guemas, and Others. High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6. *Geoscientific Model Development*, 9(11):4185–4208, 2016.

[3] Stephan Rasp, Michael S Pritchard, and Pierre Gentine. Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci. U. S. A.*, 115(39):9684–9689, September 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1810286115.

[4] Ben P Kirtman, Dughong Min, Johnna M Infanti, James L Kinter, Daniel A Paolino, Qin Zhang, Huug van den Dool, Suranjana Saha, Malaquias Pena Mendez, Emily Becker, Peitao Peng, Patrick Tripp, Jin Huang, David G DeWitt, Michael K Tippett, Anthony G Barnston, Shuhua Li, Anthony Rosati, Siegfried D Schubert, Michele Rienecker, Max Suarez, Zhao E Li, Jelena Marshak, Young-Kwon Lim, Joseph Tribbia, Kathleen Pegion, William J Merryfield, Bertrand Denis, and Eric F Wood. The north american multimodel ensemble: Phase-1 Seasonal-to-Interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Am. Meteorol. Soc.*, 95(4):585–601, April 2014. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-12-00050.1.

[5] D M Smith, R Eade, A A Scaife, L-P Caron, G Danabasoglu, T M DelSole, T Delworth, F J Doblas-Reyes, N J Dunstone, L Hermanson, V Kharin, M Kimoto, W J Merryfield, T Mochizuki, W A Müller, H Pohlmann, S Yeager, and X Yang. Robust skill of decadal climate predictions. *npj Climate and Atmospheric Science*, 2(1):1–10, May 2019. ISSN 2397-3722, 2397-3722. doi: 10.1038/s41612-019-0071-y.

[6] William J Merryfield, Johanna Baehr, Lauriane Batté, Emily J Becker, Amy H Butler, Caio A S Coelho, Gokhan Danabasoglu, Paul A Dirmeyer, Francisco J Doblas-Reyes, Daniela I V Domeisen, Laura Ferranti, Tatiana Ilynia, Arun Kumar, Wolfgang A Müller, Michel Rixen, Andrew W Robertson, Doug M Smith, Yuhei Takaya, Matthias Tuma, Frederic Vitart, Christopher J White, Mariano S Alvarez, Constantin Ardilouze, Hannah Attard, Cory Baggett, Magdalena A Balmaseda, Asmerom F Beraki, Partha S Bhattacharjee, Roberto Bilbao, Felipe M de Andrade, Michael J DeFlorio, Leandro B Díaz, Muhammad Azhar Ehsan, Georgios Fragkoulidis, Sam Grainger, Benjamin W Green, Momme C Hell, Johnna M Infanti, Katharina Isensee, Takahito Kataoka, Ben P Kirtman, Nicholas P Klingaman, June-Yi Lee, Kirsten Mayer, Roseanna McKay, Jennifer V Mecking, Douglas E Miller, Nele Neddermann, Ching Ho Justin Ng, Albert Ossó, Klaus Pankatz, Simon Peatman, Kathy Pegion, Judith Perlwitz, G Cristina Recalde-Coronel, Annika Reintges, Christoph Renkl, Balakrishnan Solaraju-Murali, Aaron Spring, Cristiana Stan, Y Qiang Sun, Carly R Tozer, Nicolas Vigaud, Steven Woolnough, and Stephen Yeager. Subseasonal to decadal prediction: Filling the Weather–Climate gap. *Bull. Am. Meteorol. Soc.*, 101(9):767–770, September 2020. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-19-0037.A.

[7] Desiree Tommasi, Charles A Stock, Michael A Alexander, Xiaosong Yang, Anthony Rosati, and Gabriel A Vecchi. Multi-Annual climate predictions for fisheries: An assessment of skill of sea surface temperature forecasts for large marine ecosystems. *Frontiers in Marine Science*, 4, 2017. ISSN 2296-7745. doi: 10.3389/fmars.2017.00201.

[8] N J Mantua, S R Hare, Y Zhang, and others. A pacific interdecadal climate oscillation with impacts on salmon production *. *Bulletin of the*, 1997.

[9] Patrick Lehodey, Arnaud Bertrand, Alistair J Hobday, Hidetada Kiyofuji, Sam McClatchie, Christophe E Menkès, Graham Pilling, Jeffrey Polovina, and Desiree Tom-

masi. ENSO impact on marine fisheries and ecosystems, November 2020. ISSN 2328-8779.

[10] Flavio Lehner, Andrew W Wood, Dagmar Llewellyn, Douglas B Blatchford, Angus G Goodbody, and Florian Pappenberger. Mitigating the impacts of climate nonstationarity on seasonal. *Geophysical Research Letterrs*, 2017. doi: 10.1002/2017GL076043.

[11] Andrej Ceglar and Andrea Toreti. Seasonal climate forecast can inform the european agricultural sector well in advance of harvesting. *npj Climate and Atmospheric Science*, 4(1):1–8, August 2021. ISSN 2397-3722, 2397-3722. doi: 10.1038/s41612-021-00198-3.

[12] Jess Melbourne-Thomas, Desiree Tommasi, Marion Gehlen, Eugene J Murphy, Jennifer Beckensteiner, Francisco Bravo, Tyler D Eddy, Mibu Fischer, Elizabeth Fulton, Mayya Gogina, Eileen Hofmann, Maysa Ito, Sara Mynott, Kelly Ortega-Cisneros, Anna N Osiecka, Mark R Payne, Romeo Saldívar-Lucio, and Kim J N Scherrer. Integrating human dimensions in decadal-scale prediction for marine social–ecological systems: lighting the grey zone. *ICES J. Mar. Sci.*, 80(1):16–30, December 2022. ISSN 1054-3139. doi: 10.1093/icesjms/fsac228.

[13] Andréa S Taschetto, Caroline C Ummenhofer, Malte F Stuecker, Dietmar Dommenget, Karumuri Ashok, Regina R Rodrigues, and Sang-Wook Yeh. ENSO atmospheric teleconnections, November 2020. ISSN 2328-8779.

[14] J Bjerknes. ATMOSPHERIC TELECONNECTIONS FROM THE EQUATORIAL PACIFIC. *Mon. Weather Rev.*, 97(3):163–172, March 1969. ISSN 0027-0644, 1520-0493. doi: 10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2.

[15] David B Enfield, Alberto M Mestas-Nuñez, and Paul J Trimble. The atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.*, 28(10):2077–2080, May 2001. ISSN 0094-8276, 1944-8007. doi: 10.1029/2000gl012745.

[16] R Zhang, R Sutton, G Danabasoglu, and others. A review of the role of the atlantic meridional overturning circulation in atlantic multidecadal variability and associated climate impacts. *Reviews of*, 2019. doi: 10.1029/2019RG000644.

[17] Gerald A Meehl, Jadwiga H Richter, Haiyan Teng, Antonietta Capotondi, Kim Cobb, Francisco Doblas-Reyes, Markus G Donat, Matthew H England, John C Fyfe, Weiqing Han, Hyemi Kim, Ben P Kirtman, Yochanan Kushnir, Nicole S Lovenduski, Michael E Mann, William J Merryfield, Veronica Nieves, Kathy Pegion, Nan Rosenbloom, Sara C Sanchez, Adam A Scaife, Doug Smith, Aneesh C Subramanian, Lantao Sun, Diane Thompson, Caroline C Ummenhofer, and Shang-Ping Xie. Initialized earth system prediction from subseasonal to decadal timescales. *Nature Reviews Earth & Environment*, 2(5):340–357, April 2021. ISSN 2662-138X, 2662-138X. doi: 10.1038/s43017-021-00155-x.

[18] Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui-Ya Chuang, Mark Iredell, Michael Ek, Jesse Meng, Rongqian Yang, Malaquías Peña Mendez, Huug van den Dool, Qin Zhang, Wanqiu Wang, Mingyue Chen, and Emily Becker. The NCEP climate forecast system version 2. *J. Clim.*, 27(6):2185–2208, March 2014. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-12-00823.1.

[19] Leon Hermanson, Doug Smith, Melissa Seabrook, Roberto Bilbao, Francisco Doblas-Reyes, Etienne Tourigny, Vladimir Lapin, Viatcheslav V Kharin, William J Merryfield, Reinel Sospedra-Alfonso, Panos Athanasiadis, Dario Nicoli, Silvio Gualdi, Nick Dunstone, Rosie Eade, Adam Scaife, Mark Collier, Terence O'Kane, Vassili Kitsios, Paul Sandery, Klaus Pankatz, Barbara Früh, Holger Pohlmann, Wolfgang Müller, Takahito Kataoka, Hiroaki Tatebe, Masayoshi Ishii, Yukiko Imada, Tim Kruschke, Torben Koenigk, Mehdi Pasha Karami, Shuting Yang, Tian Tian, Liping Zhang, Tom Delworth, Xiaosong Yang, Fanrong Zeng, Yiguo Wang, François Counillon, Noel

Keenlyside, Ingo Bethke, Judith Lean, Jürg Luterbacher, Rupa Kumar Kolli, and Arun Kumar. WMO global annual to decadal climate update: A prediction for 2021–25. *Bull. Am. Meteorol. Soc.*, 103(4):E1117–E1129, April 2022. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-20-0311.1.

[20] Hui Ding, Matthew Newman, Michael A Alexander, and Andrew T Wittenberg. Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophys. Res. Lett.*, 46(3):1721–1730, February 2019. ISSN 0094-8276, 1944-8007. doi: 10.1029/2018gl080598.

[21] Daniel J Vimont, Matthew Newman, David S Battisti, and Sang-Ik Shin. The role of seasonality and the ENSO mode in central and east pacific ENSO growth and evolution. *J. Clim.*, 35(11):3195–3209, June 2022. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-21-0599.1.

[22] Jamin K Rader and Elizabeth A Barnes. Optimizing seasonal-to-decadal analog forecasts with a learned spatially-weighted mask. *Geophys. Res. Lett.*, 50(23), December 2023. ISSN 0094-8276, 1944-8007. doi: 10.1029/2023gl104983.

[23] Christopher Irrgang, Niklas Boers, Maike Sonnewald, Elizabeth A Barnes, Christopher Kadow, Joanna Staneva, and Jan Saynisch-Wagner. Towards neural earth system modelling by integrating artificial intelligence in earth system science. *Nature Machine Intelligence*, 3(8):667–674, August 2021. ISSN 2522-5839, 2522-5839. doi: 10.1038/s42256-021-00374-3.

[24] Benjamin A Toms, Elizabeth A Barnes, and Imme Ebert-Uphoff. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *J. Adv. Model. Earth Syst.*, 12(9):e2019MS002002, September 2020. ISSN 1942-2466. doi: 10.1029/2019ms002002.

[25] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5): 206–215, May 2019. ISSN 2522-5839, 2522-5839. doi: 10.1038/s42256-019-0048-x.

[26] Dino Pedreschi, Fosca Giannotti, Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, and Franco Turini. Meaningful explanations of black box AI decision systems. *AAAI*, 33(01):9780–9784, July 2019. ISSN 2374-3468, 2374-3468. doi: 10.1609/aaai.v33i01.33019780.

[27] Elizabeth A Barnes, Randal J Barnes, Zane K Martin, and Jamin K Rader. This looks like that there: Interpretable neural networks for image tasks when location matters. *Artificial Intelligence for the Earth Systems*, 1(3), July 2022. ISSN 2769-7525. doi: 10.1175/AIES-D-22-0001.1.

[28] Yochanan Kushnir, Adam A Scaife, Raymond Arritt, Gianpaolo Balsamo, George Boer, Francisco Doblas-Reyes, Ed Hawkins, Masahide Kimoto, Rupa Kumar Kolli, Arun Kumar, Daniela Matei, Katja Matthes, Wolfgang A Müller, Terence O'Kane, Judith Perlwitz, Scott Power, Marilyn Raphael, Akihiko Shimpo, Doug Smith, Matthias Tuma, and Bo Wu. Towards operational predictions of the near-term climate. *Nat. Clim. Chang.*, 9(2):94–101, January 2019. ISSN 1758-678X. doi: 10.1038/s41558-018-0359-7.

[29] Erin Towler, Debasish PaiMazumder, and James Done. Toward the application of decadal climate predictions. *J. Appl. Meteorol. Climatol.*, 57(3):555–568, March 2018. ISSN 1558-8424, 1558-8432. doi: 10.1175/JAMC-D-17-0113.1.

[30] Edward N Lorenz. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, 26(4):636–646, July 1969. ISSN 0022-4928, 1520-0469. doi: 10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

[31] Hui Ding, Matthew Newman, Michael A Alexander, and Andrew T Wittenberg. Skillful climate forecasts of the tropical Indo-Pacific ocean using Model-Analogs. *J. Clim.*, 31(14):5437–5459, July 2018. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-17-0661.1.

[32] Matthew B Menary, Juliette Mignot, and Jon Robson. Skilful decadal predictions of subpolar north atlantic SSTs using CMIP model-analogues. *Environ. Res. Lett.*, 16(6):064090, June 2021. ISSN 1748-9326. doi: 10.1088/1748-9326/ac06fb.

[33] Luca Delle Monache, F Anthony Eckel, Daran L Rife, Badrinath Nagarajan, and Keith Searight. Probabilistic weather prediction with an analog ensemble. *Mon. Weather Rev.*, 141(10):3498–3516, October 2013. ISSN 0027-0644, 1520-0493. doi: 10.1175/mwr-d-12-00281.1.

[34] Liping Zhang, Thomas L Delworth, Xiaosong Yang, Yushi Morioka, Fanrong Zeng, and Feiyu Lu. Skillful decadal prediction skill over the southern ocean based on GFDL SPEAR Model-Analogs. *Environ. Res. Commun.*, 5(2):021002, February 2023. ISSN 2515-7620. doi: 10.1088/2515-7620/acb90e.

[35] David P Mulholland, Patrick Laloyaux, Keith Haines, and Magdalena Alonso Balmaseda. Origin and impact of initialization shocks in coupled Atmosphere–Ocean forecasts. *Mon. Weather Rev.*, 143(11):4631–4644, November 2015. ISSN 0027-0644, 1520-0493. doi: 10.1175/MWR-D-15-0076.1.

[36] Edward N Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20(2):130–141, March 1963. ISSN 0022-4928, 1520-0469. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

[37] H M Van den Dool. Searching for analogues, how long must we wait? *Tellus A*, 46(3):314–324, May 1994. ISSN 0280-6495, 1600-0870. doi: 10.1034/j.1600-0870.1994.t01-2-00006.x.

[38] jiale Lou, Matthew Newman, and Andrew Hoell. Multi-decadal variation of ENSO forecast skill since the late 1800s. February 2023.

[39] Weiying Peng, Quanliang Chen, Shijie Zhou, and Ping Huang. CMIP6 model-based analog forecasting for the seasonal prediction of sea surface temperature in the offshore area of china. *Geoscience Letters*, 8(1):1–8, March 2021. ISSN 2196-4092, 2196-4092. doi: 10.1186/s40562-021-00179-7.

[40] Yanling Wu and Xiaoqin Yan. Evaluating changes in the multiyear predictability of the pacific decadal oscillation using model analogs since 1900. *J. Mar. Sci. Eng.*, 11 (5):980, May 2023. ISSN 2077-1312, 2077-1312. doi: 10.3390/jmse11050980.

[41] Rashed Mahmood, Markus G Donat, Pablo Ortega, Francisco J Doblas-Reyes, Carlos Delgado-Torres, Margarida Samsó, and Pierre-Antoine Bretonnière. Constraining low-frequency variability in climate projections to predict climate on decadal to multi-decadal timescales – a poor man's initialized prediction system. *Earth Syst. Dyn.*, 13(4):1437–1450, October 2022. ISSN 2190-4979, 2190-4987. doi: 10.5194/esd-13-1437-2022.

[42] N Maher, S Milinski, L Suarez-Gutierrez, and others. The max planck institute grand ensemble: enabling the exploration of climate system variability. *Journal of Advances*, 2019. doi: 10.1029/2019MS001639.

[43] Marco A Giorgetta, Johann Jungclaus, Christian H Reick, Stephanie Legutke, Jürgen Bader, Michael Böttinger, Victor Brovkin, Traute Crueger, Monika Esch, Kerstin Fieg, Ksenia Glushak, Veronika Gayler, Helmuth Haak, Heinz-Dieter Hollweg, Tatiana Ilyina, Stefan Kinne, Luis Kornblueh, Daniela Matei, Thorsten Mauritsen, Uwe Mikolajewicz, Wolfgang Mueller, Dirk Notz, Felix Pithan, Thomas Raddatz, Sebastian Rast, Rene Redler, Erich Roeckner, Hauke Schmidt, Reiner Schnur, Joachim Segschneider, Katharina D Six, Martina Stockhause, Claudia Timmreck, Jörg Wegner, Heinrich

Widmann, Karl-H Wieners, Martin Claussen, Jochem Marotzke, and Bjorn Stevens. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J. Adv. Model. Earth Syst.*, 5(3): 572–597, July 2013. ISSN 1942-2466. doi: 10.1002/jame.20038.

[44] Laura C Jackson, Arne Biastoch, Martha W Buckley, Damien G Desbruyères, Eleanor Frajka-Williams, Ben Moat, and Jon Robson. The evolution of the north atlantic meridional overturning circulation since 1980. *Nature Reviews Earth & Environment*, 3(4): 241–254, March 2022. ISSN 2662-138X, 2662-138X. doi: 10.1038/s43017-022-00263-2.

[45] Ed Hawkins, Jon Robson, Rowan Sutton, Doug Smith, and Noel Keenlyside. Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach. *Clim. Dyn.*, 37(11-12):2495–2509, December 2011. ISSN 0930-7575, 1432-0894. doi: 10.1007/s00382-011-1023-3.

[46] R T Sutton and M R Allen. Decadal predictability of north atlantic sea surface temperature and climate. *Nature*, 388(6642):563–567, August 1997. ISSN 0028-0836. doi: 10.1038/41523.

[47] S B Goldenberg, C W Landsea, A M Mestas-Nunez, and W M Gray. The recent increase in atlantic hurricane activity: causes and implications. *Science*, 293(5529): 474–479, July 2001. ISSN 0036-8075. doi: 10.1126/science.1060040.

[48] Karthik Balaguru, Gregory R Foltz, and L Ruby Leung. Increasing magnitude of hurricane rapid intensification in the central and eastern tropical atlantic. *Geophys. Res. Lett.*, 45(9):4238–4247, May 2018. ISSN 0094-8276, 1944-8007. doi: 10.1029/2018gl077597.

[49] Dong Si, Aixue Hu, Dabang Jiang, and Xianmei Lang. Atmospheric teleconnection associated with the atlantic multidecadal variability in summer: assessment of the

CESM1 model. *Clim. Dyn.*, 60(3):1043–1060, February 2023. ISSN 0930-7575, 1432-0894. doi: 10.1007/s00382-022-06331-z.

[50] Mayank Shekhar, Anupam Sharma, A P Dimri, and Sampat Kumar Tandon. Asian summer monsoon variability, global teleconnections, and dynamics during the last 1,000 years. *Earth-Sci. Rev.*, 230:104041, July 2022. ISSN 0012-8252. doi: 10.1016/j.earscirev.2022.104041.

[51] Jane Hsiung and Reginald E Newell. The principal nonseasonal modes of variation of global sea surface temperature. *J. Phys. Oceanogr.*, 13(10):1957–1967, October 1983. ISSN 0022-3670, 1520-0485. doi: 10.1175/1520-0485(1983)013<1957:TPNMOV>2.0.CO;2.

[52] S W Yeh, W Cai, S K Min, M J McPhaden, and others. ENSO atmospheric teleconnections and their response to greenhouse gas forcing. *Reviews of*, 2018. doi: 10.1002/2017RG000568.

[53] Anthony G Barnston, Muthuvel Chelliah, and Stanley B Goldenberg. Documentation of a highly ENSO-related sst region in the equatorial pacific: Research note. *Atmosphere-Ocean*, 35(3):367–383, September 1997. ISSN 0705-5900. doi: 10.1080/07055900.1997.9649597.

[54] Deborah E Hanley, Mark A Bourassa, James J O'Brien, Shawn R Smith, and Elizabeth R Spade. A quantitative evaluation of ENSO indices. *J. Clim.*, 16(8):1249–1258, April 2003. ISSN 0894-8755, 1520-0442. doi: 10.1175/1520-0442(2003)16<1249:AQEOEI>2.0.CO;2.

[55] Shih-Yu Wang, Michelle L'Heureux, and Hsin-Hsing Chia. ENSO prediction one year in advance using western north pacific sea surface temperatures. *Geophys. Res. Lett.*, 39(5), March 2012. ISSN 0094-8276. doi: 10.1029/2012GL050909.

[56] Dillon J Amaya. The pacific meridional mode and ENSO: a review. *Current Climate Change Reports*, 5(4):296–307, December 2019. ISSN 2198-6061. doi: 10.1007/s40641-019-00142-x.

[57] M Martín-Rey, B Rodríguez-Fonseca, and others. Atlantic opportunities for ENSO prediction. *Geophys. Res. Lett.*, 2015. ISSN 0094-8276. doi: 10.1002/2015GL065062.

[58] Antonietta Capotondi and Prashant D Sardeshmukh. Optimal precursors of different types of ENSO events. *Geophys. Res. Lett.*, 42(22):9952–9960, November 2015. ISSN 0094-8276, 1944-8007. doi: 10.1002/2015gl066171.

[59] Hui Ding, Noel S Keenlyside, and Mojib Latif. Impact of the equatorial atlantic on the el niño southern oscillation. *Clim. Dyn.*, 38(9):1965–1972, May 2012. ISSN 0930-7575, 1432-0894. doi: 10.1007/s00382-011-1097-y.

[60] Chunzai Wang. A review of ENSO theories. *Natl Sci Rev*, 5(6):813–825, October 2018. ISSN 2095-5138. doi: 10.1093/nsr/nwy104.

[61] E M Gordon and E A Barnes. Incorporating uncertainty into a regression neural network enables identification of decadal State-Dependent predictability in CESM2. *Geophys. Res. Lett.*, 2022. ISSN 0094-8276.

[62] Dmitry V Sein, Matthias Gröger, William Cabos, Francisco J Alvarez-Garcia, Stefan Hagemann, Joaquim G Pinto, Alfredo Izquierdo, Alba de la Vara, Nikolay V Koldunov, Anton Yu Dvornikov, Natalia Limareva, Evgenia Alekseeva, Benjamin Martinez-Lopez, and Daniela Jacob. Regionally coupled atmosphere-ocean-marine biogeochemistry model ROM: 2. studying the climate change signal in the north atlantic and europe. *J. Adv. Model. Earth Syst.*, 12(8):e2019MS001646, August 2020. ISSN 1942-2466. doi: 10.1029/2019ms001646.

[63] J K Rader, E A Barnes, I Ebert-Uphoff, and others. Detection of forced change within combined climate fields using explainable neural networks. *Journal of Advances*, 2022.

[64] Lijing Cheng, John Abraham, Kevin E Trenberth, Tim Boyer, Michael E Mann, Jiang Zhu, Fan Wang, Fujiang Yu, Ricardo Locarnini, John Fasullo, Fei Zheng, Yuanlong Li, Bin Zhang, Liying Wan, Xingrong Chen, Dakui Wang, Licheng Feng, Xiangzhou Song, Yulong Liu, Franco Reseghetti, Simona Simoncelli, Viktor Gouretski, Gengxin Chen, Alexey Mishonov, Jim Reagan, Karina Von Schuckmann, Yuying Pan, Zhetao Tan, Yujing Zhu, Wangxu Wei, Guancheng Li, Qiuping Ren, Lijuan Cao, and Yayang Lu. New record ocean temperatures and related climate indicators in 2023. *Adv. Atmos. Sci.*, 41(6):1068–1082, June 2024. ISSN 1861-9533. doi: 10.1007/s00376-024-3378-5.

[65] Boyin Huang, Peter W Thorne, Viva F Banzon, Tim Boyer, Gennady Chepurin, Jay H Lawrimore, Matthew J Menne, Thomas M Smith, Russell S Vose, Huai-Min Zhang, and Others. NOAA extended reconstructed sea surface temperature (ERSST), version 5. *NOAA National Centers for Environmental Information*, 30(8179-8205):25, 2017.

[66] Gavin Schmidt. Climate models can't explain 2023's huge heat anomaly — we could be in uncharted territory. http://dx.doi.org/10.1038/d41586-024-00816-z, March 2024. Accessed: 2024-3-27.

[67] Nick J Dunstone, Doug M Smith, Chris Atkinson, Andrew Colman, Chris Folland, Leon Hermanson, Sarah Ineson, Rachel Killick, Colin Morice, Nick Rayner, Melissa Seabrook, and Adam A Scaife. Will 2024 be the first year that global temperature exceeds 1.5°c? *Atmos. Sci. Lett.*, June 2024. ISSN 1530-261X. doi: 10.1002/asl.1254.

[68] M R Schoeberl, Y Wang, R Ueyama, A Dessler, G Taha, and W Yu. The estimated climate impact of the hunga Tonga-Hunga ha'apai eruption plume. *Geophys. Res. Lett.*, 50(18), September 2023. ISSN 0094-8276, 1944-8007. doi: 10.1029/2023gl104634.

[69] Tianle Yuan, Hua Song, Lazaros Oreopoulos, Robert Wood, Huisheng Bian, Katherine Breen, Mian Chin, Hongbin Yu, Donifan Barahona, Kerry Meyer, and Steven

Platnick. Abrupt reduction in shipping emission as an inadvertent geoengineering termination shock produces substantial radiative warming. *Communications Earth & Environment*, 5(1):1–8, May 2024. ISSN 2662-4435, 2662-4435. doi: 10.1038/s43247-024-01442-3.

[70] R B Alley, J Marotzke, W D Nordhaus, J T Overpeck, D M Peteet, R A Pielke, Jr, R T Pierrehumbert, P B Rhines, T F Stocker, L D Talley, and J M Wallace. Abrupt climate change. *Science*, 299(5615):2005–2010, March 2003. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1081056.

[71] Doug McNeall, Paul R Halloran, Peter Good, and Richard A Betts. Analyzing abrupt and nonlinear climate changes and their impacts. *Wiley Interdiscip. Rev. Clim. Change*, 2(5):663–686, September 2011. ISSN 1757-7780, 1757-7799. doi: 10.1002/wcc.130.

[72] M J Alessi and M A A Rugenstein. Surface temperature pattern scenarios suggest higher warming rates than current projections. *Geophys. Res. Lett.*, 2023. ISSN 0094-8276.

[73] Paul D L Ritchie, Joseph J Clarke, Peter M Cox, and Chris Huntingford. Overshooting tipping point thresholds in a changing climate. *Nature*, 592(7855):517–523, April 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03263-2.

[74] Monica G Turner, W John Calder, Graeme S Cumming, Terry P Hughes, Anke Jentsch, Shannon L LaDeau, Timothy M Lenton, Bryan N Shuman, Merritt R Turetsky, Zak Ratajczak, John W Williams, A Park Williams, and Stephen R Carpenter. Climate change, ecosystems and abrupt change: science priorities. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 375(1794):20190105, March 2020. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2019.0105.

[75] David N Bengston, Jason Crabtree, and Teppo Hujala. Abrupt climate change: Exploring the implications of a wild card. *Futures*, 124:102641, December 2020. ISSN 0016-3287. doi: 10.1016/j.futures.2020.102641.

[76] Simon Dietz, James Rising, Thomas Stoerk, and Gernot Wagner. Economic impacts of tipping points in the climate system. *Proc. Natl. Acad. Sci. U. S. A.*, 118(34), August 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2103081118.

[77] Robyn Wagoner. In response to abrupt climate change. https://archives.evergreen.edu/masterstheses/Accession86-10MES/Wagoner_RMESthesis2015.pdf, 2015. Accessed: 2024-6-17.

[78] UNFCCC. The cancun agreements: Outcome of the work of the ad hoc working group on long-term cooperative action under the convention CP.16. Technical Report FCCC/CP/2010/7/Add.1, 2010.

[79] UNFCCC. Adoption of the paris agreement. i: Proposal by the president. draft decision CP.21. Technical Report FCCC/CP/2015/L. 9/Rev. 1, Geneva, 2015, 2015.

[80] Richard A Betts, Stephen E Belcher, Leon Hermanson, Albert Klein Tank, Jason A Lowe, Chris D Jones, Colin P Morice, Nick A Rayner, Adam A Scaife, and Peter A Stott. Approaching 1.5 °c: how will we know we've reached this crucial warming mark? *Nature*, 624(7990):33–35, December 2023. ISSN 0028-0836, 1476-4687. doi: 10.1038/d41586-023-03775-z.

[81] June-Yi Lee, Jochem Marotzke, Govindasamy Bala, Long Cao, Susanna Corti, John P Dunne, Francois Engelbrecht, Erich Fischer, John C Fyfe, Christopher Jones, and Others. Future global climate: scenario-based projections and near-term information. In *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, pages 553–672. Cambridge University Press, 2021.

[82] E Hawkins and R Sutton. Time of emergence of climate signals. *Geophys. Res. Lett.*, 39(1), January 2012. ISSN 0094-8276, 1944-8007. doi: 10.1029/2011gl050087.

[83] Leela M Frankcombe, Matthew H England, Michael E Mann, and Byron A Steinman. Separating internal variability from the externally forced climate response. *J. Clim.*, 28(20):8184–8202, October 2015. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-15-0069.1.

[84] D E Parker, H Wilson, P D Jones, and others. The impact of mount pinatubo on worldwide temperatures. *Journal of ...*, 1996. doi: 10.1002/(SICI)1097-0088(199605)16:5<487::AID-JOC39>3.0.CO;2-J.

[85] Benjamin D Santer, Jeffrey F Painter, Carl A Mears, Charles Doutriaux, Peter Caldwell, Julie M Arblaster, Philip J Cameron-Smith, Nathan P Gillett, Peter J Gleckler, John Lanzante, Judith Perlwitz, Susan Solomon, Peter A Stott, Karl E Taylor, Laurent Terray, Peter W Thorne, Michael F Wehner, Frank J Wentz, Tom M L Wigley, Laura J Wilcox, and Cheng-Zhi Zou. Identifying human influences on atmospheric temperature. *Proc. Natl. Acad. Sci. U. S. A.*, 110(1):26–33, January 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1210514109.

[86] Céline J W Bonfils, Benjamin D Santer, John C Fyfe, Kate Marvel, Thomas J Phillips, and Susan R H Zimmerman. Human influence on joint changes in temperature, rainfall and continental aridity. *Nat. Clim. Chang.*, 10(8):726–731, July 2020. ISSN 1758-678X. doi: 10.1038/s41558-020-0821-1.

[87] Sarah M Kang, Yue Yu, Clara Deser, Xiyue Zhang, In-Sik Kang, Sun-Seon Lee, Keith B Rodgers, and Paulo Ceppi. Global impacts of recent southern ocean cooling. *Proc. Natl. Acad. Sci. U. S. A.*, 120(30):e2300881120, July 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2300881120.

[88] Robert C J Wills, Yue Dong, Cristian Proistosecu, Kyle C Armour, and David S Battisti. Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophys. Res. Lett.*, 49(17), September 2022. ISSN 0094-8276, 1944-8007. doi: 10.1029/2022gl100011.

[89] R J Wills, C Deser, K McKinnon, A Phillips, and others. Towards probabilistic climate projections based on the CMIP6 multi-model ensemble. 2024.

[90] Keith B Rodgers, Sun-Seon Lee, Nan Rosenbloom, Axel Timmermann, Gokhan Danabasoglu, Clara Deser, Jim Edwards, Ji-Eun Kim, Isla R Simpson, Karl Stein, Malte F Stuecker, Ryohei Yamaguchi, Tamás Bódai, Eui-Seok Chung, Lei Huang, Who M Kim, Jean-François Lamarque, Danica L Lombardozzi, William R Wieder, and Stephen G Yeager. Ubiquity of human-induced changes in climate variability. *Earth Syst. Dyn.*, 12(4):1393–1411, December 2021. ISSN 2190-4979, 2190-4987. doi: 10.5194/esd-12-1393-2021.

[91] Karl-Hermann Wieners, Marco Giorgetta, Johann Jungclaus, Christian Reick, Monika Esch, Matthias Bittner, Veronika Gayler, Helmuth Haak, Philipp de Vrese, Thomas Raddatz, and Others. MPI-M MPI-ESM1. 2-LR model output prepared for CMIP6 ScenarioMIP ssp126. 2019.

[92] Neil Cameron Swart, Jason N S Cole, Viatcheslav V Kharin, Mike Lazare, John F Scinocca, Nathan P Gillett, James Anstey, Vivek Arora, James R Christian, Yanjun Jiao, and Others. CCCma CanESM5 model output prepared for CMIP6 ScenarioMIP. 2019.

[93] Kaoru Tachiiri, Manabu Abe, Tomohiro Hajima, Osamu Arakawa, Tatsuo Suzuki, Yoshiki Komuro, Koji Ogochi, Michio Watanabe, Akitomo Yamamoto, Hiroaki Tatebe, Maki A Noguchi, Rumi Ohgaito, Akinori Ito, Dai Yamazaki, Akihiko Ito, Kumiko Takata, Shingo Watanabe, and Michio Kawamiya. MIROC MIROC-ES2L model output prepared for CMIP6 ScenarioMIP, August 2019.

[94] H Tatebe and M Watanabe. MIROC MIROC6 model output prepared for CMIP6 ScenarioMIP. *Earth System Grid Federation. https://doi. org/10. 22033/ESGF/CMIP6*, 5603, 2018.

[95] Brian C O'Neill, Elmar Kriegler, Kristie L Ebi, Eric Kemp-Benedict, Keywan Riahi, Dale S Rothman, Bas J van Ruijven, Detlef P van Vuuren, Joern Birkmann, Kasper Kok, Marc Levy, and William Solecki. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Glob. Environ. Change*, 42:169–180, January 2017. ISSN 0959-3780. doi: 10.1016/j.gloenvcha.2015.01.004.

[96] C Deser, F Lehner, K B Rodgers, T Ault, T L Delworth, P N DiNezio, A Fiore, C Frankignoul, J C Fyfe, D E Horton, J E Kay, R Knutti, N S Lovenduski, J Marotzke, K A McKinnon, S Minobe, J Randerson, J A Screen, I R Simpson, and M Ting. Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Chang.*, 10(4):277–286, March 2020. ISSN 1758-678X. doi: 10.1038/s41558-020-0731-2.

[97] Robert C Wills, Tapio Schneider, John M Wallace, David S Battisti, and Dennis L Hartmann. Disentangling global warming, multidecadal variability, and el niño in pacific temperatures. *Geophys. Res. Lett.*, 45(5):2487–2496, March 2018. ISSN 0094-8276, 1944-8007. doi: 10.1002/2017GL076327.

[98] Andrew G Pauling, Cecilia M Bitz, and Kyle C Armour. The climate response to the mt. pinatubo eruption does not constrain climate sensitivity. *Geophys. Res. Lett.*, 50 (7), April 2023. ISSN 0094-8276, 1944-8007. doi: 10.1029/2023gl102946.

[99] J Bjerknes. A possible response of the atmospheric hadley circulation to equatorial anomalies of ocean temperature. *Tell'Us*, 18(4):820–829, January 1966. ISSN 0804-6042, 0040-2826. doi: 10.3402/tellusa.v18i4.9712.

[100] Wenju Cai, Agus Santoso, Matthew Collins, Boris Dewitte, Christina Karamperidou, Jong-Seong Kug, Matthieu Lengaigne, Michael J McPhaden, Malte F Stuecker, Andréa S Taschetto, Axel Timmermann, Lixin Wu, Sang-Wook Yeh, Guojian Wang, Benjamin Ng, Fan Jia, Yun Yang, Jun Ying, Xiao-Tong Zheng, Tobias Bayr, Josephine R Brown, Antonietta Capotondi, Kim M Cobb, Bolan Gan, Tao Geng, Yoo-Geun Ham, Fei-Fei Jin, Hyun-Su Jo, Xichen Li, Xiaopei Lin, Shayne McGregor, Jae-Heung Park, Karl Stein, Kai Yang, Li Zhang, and Wenxiu Zhong. Changing el niño–southern oscillation in a warming climate. *Nature Reviews Earth & Environment*, 2(9):628–644, August 2021. ISSN 2662-138X, 2662-138X. doi: 10.1038/s43017-021-00199-z.

[101] Matthew Newman, Michael A Alexander, Toby R Ault, Kim M Cobb, Clara Deser, Emanuele Di Lorenzo, Nathan J Mantua, Arthur J Miller, Shoshiro Minobe, Hisashi Nakamura, Niklas Schneider, Daniel J Vimont, Adam S Phillips, James D Scott, and Catherine A Smith. The pacific decadal oscillation, revisited. *J. Clim.*, 29(12): 4399–4427, June 2016. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-15-0508.1.

[102] Yue Dong, C Proistosescu, K Armour, and D Battisti. Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a green's function approach: The preeminence of the western pacific. *J. Clim.*, July 2019. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-18-0843.1.

[103] Minhua Qin, Aiguo Dai, and Wenjian Hua. Aerosol-forced multidecadal variations across all ocean basins in models and observations since 1920. *Sci Adv*, 6(29): eabb0425, July 2020. ISSN 2375-2548. doi: 10.1126/sciadv.abb0425.

[104] Gerald A Meehl, Haiyan Teng, Nicola Maher, and Matthew H England. Effects of the mount pinatubo eruption on decadal climate prediction skill of pacific sea surface temperatures. *Geophys. Res. Lett.*, 42(24), December 2015. ISSN 0094-8276, 1944-8007. doi: 10.1002/2015gl066608.

[105] K E Trenberth and T J Hoar. El niño and climate change. *Geophys. Res. Lett.*, 1997. ISSN 0094-8276. doi: 10.1029/97GL03092.

[106] Xiaofan Li, Zeng-Zhen Hu, Michael J McPhaden, Congwen Zhu, and Yunyun Liu. Triple-dip la niñas in 1998–2001 and 2020–2023: Impact of mean state changes. *J. Geophys. Res.*, 128(17), September 2023. ISSN 0148-0227. doi: 10.1029/2023jd038843.

[107] Andrea Storto and Chunxue Yang. Acceleration of the ocean warming from 1961 to 2022 unveiled by large-ensemble reanalyses. *Nat. Commun.*, 15(1):545, January 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-44749-7.

[108] S P Raghuraman, B Soden, A Clement, and others. The 2023 global warming spike was driven by el niño/southern oscillation. https://egusphere.copernicus.org/preprints/2024/egusphere-2024-1937/, 2024. Accessed: 2024-7-11.

[109] S Power, T Casey, C Folland, A Colman, and V Mehta. Inter-decadal modulation of the impact of ENSO on australia. *Clim. Dyn.*, 15(5):319–324, May 1999. ISSN 0930-7575. doi: 10.1007/s003820050284.

[110] Hanna Heidemann, Tim Cowan, Scott B Power, and Benjamin J Henley. Statistical relationships between the interdecadal pacific oscillation and el niño–southern oscillation. *Clim. Dyn.*, 62(3):2499–2515, March 2024. ISSN 0930-7575, 1432-0894. doi: 10.1007/s00382-023-07035-8.

[111] N H Saji and T Yamagata. Possible impacts of indian ocean dipole mode events on global climate. *Clim. Res.*, 25:151–169, 2003. ISSN 0936-577X, 1616-1572. doi: 10.3354/cr025151.

[112] Kexin Li, Fei Zheng, Jiang Zhu, and Qing-Cun Zeng. El niño and the AMO sparked the astonishingly large margin of warming in the global mean surface temperature in 2023. *Adv. Atmos. Sci.*, January 2024. ISSN 1861-9533. doi: 10.1007/s00376-023-3371-4.

[113] NOAA/CPC (National Oceanic and Atmospheric Administration/Climate Prediction Center). Cold & warm episodes by season. https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php. Accessed: 2024-7-10.

[114] Myles Allen, Mustafa Babiker, Yang Chen, and Heleen C de Coninck. IPCC SR15: Summary for policymakers. In *IPCC Special Report Global Warming of 1.5 C*. Intergovernmental Panel on Climate Change, 2018.

[115] Robert M DeConto, David Pollard, Richard B Alley, Isabella Velicogna, Edward Gasson, Natalya Gomez, Shaina Sadai, Alan Condron, Daniel M Gilford, Erica L Ashe, Robert E Kopp, Dawei Li, and Andrea Dutton. The paris climate agreement and future sea-level rise from antarctica. *Nature*, 593(7857):83–89, May 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03427-0.

[116] S Jevrejeva, L P Jackson, A Grinsted, D Lincke, and B Marzeion. Flood damage costs under the sea level rise with warming of 1.5 °c and 2 °c. *Environ. Res. Lett.*, 13(7): 074014, July 2018. ISSN 1748-9326. doi: 10.1088/1748-9326/aacc76.

[117] E M Fischer and R Knutti. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Clim. Chang.*, 5(6):560–564, April 2015. ISSN 1758-678X. doi: 10.1038/nclimate2617.

[118] Sihan Li, Friederike E L Otto, Luke J Harrington, Sarah N Sparrow, and David C H Wallom. A pan-South-America assessment of avoided exposure to dangerous extreme precipitation by limiting to 1.5 °c warming. *Environ. Res. Lett.*, 15(5):054005, May 2020. ISSN 1748-9326. doi: 10.1088/1748-9326/ab50a2.

[119] Sanjit Kumar Mondal, Jinglong Huang, Yanjun Wang, Buda Su, Zbigniew W Kundzewicz, Shan Jiang, Jianqing Zhai, Ziyan Chen, Cheng Jing, and Tong Jiang. Changes in extreme precipitation across south asia for each 0.5 °c of warming from

1.5 °c to 3.0°c above pre-industrial levels. *Atmos. Res.*, 266(105961):105961, March 2022. ISSN 0169-8095, 1873-2895. doi: 10.1016/j.atmosres.2021.105961.

[120] Lei Gu, Jie Chen, Jiabo Yin, Sylvia C Sullivan, Hui-Min Wang, Shenglian Guo, Liping Zhang, and Jong-Suk Kim. Projected increases in magnitude and socioeconomic exposure of global droughts in 1.5 and 2 °c warmer climates. *Hydrol. Earth Syst. Sci.*, 24(1):451–472, January 2020. ISSN 1027-5606, 1607-7938. doi: 10.5194/hess-24-451-2020.

[121] Wenbin Liu, Fubao Sun, Wee Ho Lim, Jie Zhang, Hong Wang, Hideo Shiogama, and Yuqing Zhang. Global drought and severe drought-affected populations in 1.5 and 2 °c warmer worlds. *Earth Syst. Dyn.*, 9(1):267–283, March 2018. ISSN 2190-4979, 2190-4987. doi: 10.5194/esd-9-267-2018.

[122] Thomas L Frölicher, Erich M Fischer, and Nicolas Gruber. Marine heatwaves under global warming. *Nature*, 560(7718):360–364, August 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0383-9.

[123] William W L Cheung, Gabriel Reygondeau, and Thomas L Frölicher. Large benefits to marine fisheries of meeting the 1.5°c global warming target. *Science*, 354(6319): 1591–1594, December 2016. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aag2331.

[124] Pete Smith, Jeff Price, Amy Molotoks, Rachel Warren, and Yadvinder Malhi. Impacts on terrestrial biodiversity of moving from a 2°c to a 1.5°c target. *Philos. Trans. A Math. Phys. Eng. Sci.*, 376(2119), May 2018. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2016.0456.

[125] D M Smith, A A Scaife, E Hawkins, and others. Predicted chance that global warming will temporarily exceed 1.5 C. *Geophys Res Lett*, 2018. doi: 10.1029/2018GL079362.

[126] Joeri Rogelj, Oliver Fricko, Malte Meinshausen, Volker Krey, Johanna J J Zilliacus, and Keywan Riahi. Understanding the origin of paris agreement emission uncertainties. *Nat. Commun.*, 8:15748, June 2017. ISSN 2041-1723. doi: 10.1038/ncomms15748.

[127] Daniela Jacob, Lola Kotova, Claas Teichmann, Stefan P Sobolowski, Robert Vautard, Chantal Donnelly, Aristeidis G Koutroulis, Manolis G Grillakis, Ioannis K Tsanis, Andrea Damm, Abdulla Sakalli, and Michelle T H van Vliet. Climate impacts in europe under +1.5°c global warming. *Earths Future*, 6(2):264–285, February 2018. ISSN 2328-4277. doi: 10.1002/2017ef000710.

[128] David I Armstrong McKay, Arie Staal, Jesse F Abrams, Ricarda Winkelmann, Boris Sakschewski, Sina Loriani, Ingo Fetzer, Sarah E Cornell, Johan Rockström, and Timothy M Lenton. Exceeding 1.5°c global warming could trigger multiple climate tipping points. *Science*, 377(6611):eabn7950, September 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abn7950.

[129] Sonia I Seneviratne, Joeri Rogelj, Roland Séférian, Richard Wartenburger, Myles R Allen, Michelle Cain, Richard J Millar, Kristie L Ebi, Neville Ellis, Ove Hoegh-Guldberg, Antony J Payne, Carl-Friedrich Schleussner, Petra Tschakert, and Rachel F Warren. The many possible climates from the paris agreement's aim of 1.5 °c warming. *Nature*, 558(7708):41–49, June 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0181-4.

[130] Andy Wiltshire, Dan Bernie, Laila Gohar, Jason Lowe, Camilla Mathison, and Chris Smith. Post COP26 : does the 1.5°c climate target remain alive? *Weather*, 77(12): 412–417, December 2022. ISSN 0043-1656, 1477-8696. doi: 10.1002/wea.4331.

[131] Robert Rohde. Global temperature report for 2023. https://berkeleyearth.org/global-temperature-report-for-2023/, January 2024. Accessed: 2024-2-13.

[132] Cheng Qian, Yangbo Ye, Jiacheng Jiang, Yangyang Zhong, Yuting Zhang, Izidine Pinto, Cunrui Huang, Sihan Li, and Ke Wei. Rapid attribution of the record-breaking heatwave event in north china in june 2023 and future risks. *Environ. Res. Lett.*, 19 (1):014028, December 2023. ISSN 1748-9326. doi: 10.1088/1748-9326/ad0dd9.

[133] Marc Lemus-Canovas, Damián Insua-Costa, Ricardo M Trigo, and Diego G Miralles. Record-shattering 2023 spring heatwave in western mediterranean amplified by long-term drought. *npj Climate and Atmospheric Science*, 7(1):1–8, January 2024. ISSN 2397-3722, 2397-3722. doi: 10.1038/s41612-024-00569-6.

[134] Mika Rantanen, Jouni Räisänen, and Joonas Merikanto. A method for estimating the effect of climate change on monthly mean temperatures: September 2023 and other recent record-warm months in helsinki, finland. *Atmos. Sci. Lett.*, February 2024. ISSN 1530-261X. doi: 10.1002/asl.1216.

[135] Johannes Laimighofer and Herbert Formayer. Climate change contribution to the 2023 autumn temperature records in vienna. *Sci. Rep.*, 14(1):4213, February 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-54822-2.

[136] Ariaan Purich and Edward W Doddridge. Record low antarctic sea ice coverage indicates a new sea ice state. *Communications Earth & Environment*, 4(1):1–9, September 2023. ISSN 2662-4435, 2662-4435. doi: 10.1038/s43247-023-00961-9.

[137] Babula Jena, S Kshitija, C C Bajish, John Turner, Caroline Holmes, Jeremy Wilkinson, Rahul Mohan, and M Thamban. Evolution of antarctic sea ice ahead of the record low annual maximum extent in september 2023. *Geophys. Res. Lett.*, 51(7), April 2024. ISSN 0094-8276, 1944-8007. doi: 10.1029/2023gl107561.

[138] Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the coupled model intercomparison

project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, 9 (5):1937–1958, May 2016. ISSN 1991-959X, 1991-9603. doi: 10.5194/gmd-9-1937-2016.

[139] M Gidden, K Riahi, Steven J Smith, S Fujimori, Gunnar Luderer, E Kriegler, D V van Vuuren, Maarten van den Berg, Leyang Feng, David Klein, K Calvin, J Doelman, S Frank, Oliver Fricko, M Harmsen, T Hasegawa, P Havlík, Jérôme Hilaire, R Hoesly, Jill Horing, A Popp, E Stehfest, and Kiyoshi Takahashi. Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geosci. Model Dev.*, November 2018. ISSN 1991-959X, 1991-9603. doi: 10.5194/GMD-12-1443-2019.

[140] J Hansen, A Lacis, R Ruedy, and others. Potential climate impact of mount pinatubo eruption. *Geophys. Res. Lett.*, 1992. ISSN 0094-8276. doi: 10.1029/91GL02788.

[141] T Canty, N R Mascioli, M D Smarte, and R J Salawitch. An empirical model of global climate–part 1: A critical evaluation of volcanic cooling. *Atmos. Chem. Phys.*, 13(8): 3997–4031, 2013.

[142] Elizabeth A Barnes, Randal J Barnes, and Nicolas Gordillo. Adding uncertainty to neural network regression tasks in the geosciences. *arXiv*, September 2021.

[143] Michael D Mastrandrea, Katharine J Mach, Gian-Kasper Plattner, Ottmar Edenhofer, Thomas F Stocker, Christopher B Field, Kristie L Ebi, and Patrick R Matschoss. The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Clim. Change*, 108(4):675, August 2011. ISSN 0165-0009, 1573-1480. doi: 10.1007/s10584-011-0178-6.

[144] Ed Hawkins, Pablo Ortega, Emma Suckling, Andrew Schurer, Gabi Hegerl, Phil Jones, Manoj Joshi, Timothy J Osborn, Valérie Masson-Delmotte, Juliette Mignot, Peter Thorne, and Geert Jan van Oldenborgh. Estimating changes in global temperature

since the preindustrial period. *Bull. Am. Meteorol. Soc.*, 98(9):1841–1856, September 2017. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-16-0007.1.

[145] Andrew P Schurer, Michael E Mann, Ed Hawkins, Simon F B Tett, and Gabriele C Hegerl. Importance of the Pre-Industrial baseline in determining the likelihood of exceeding the paris limits. *Nat. Clim. Chang.*, 7(8):563–567, August 2017. ISSN 1758-678X. doi: 10.1038/nclimate3345.

[146] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3319–3328. PMLR, 2017.

[147] Antonios Mamalakis, Elizabeth A Barnes, and Imme Ebert-Uphoff. Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artificial Intelligence for the Earth Systems*, 1(4), October 2022. ISSN 2769-7525. doi: 10.1175/AIES-D-22-0012.1.

[148] Emily M Gordon, Elizabeth A Barnes, and James W Hurrell. Oceanic harbingers of pacific decadal oscillation predictability in CESM2 detected by neural networks. *Geophys. Res. Lett.*, 48(21), November 2021. ISSN 0094-8276, 1944-8007. doi: 10.1029/2021gl095392.

[149] Victor E Privalsky and Donald T Jensen. Assessment of the influence of ENSO on annual global air temperatures. *Dyn. Atmos. Oceans*, 22(3):161–178, July 1995. ISSN 0377-0265. doi: 10.1016/0377-0265(94)00400-Q.

[150] Emily M Gordon, Elizabeth A Barnes, and Frances V Davenport. Separating internal and forced contributions to near term SST predictability in the CESM2-LE. *Environ. Res. Lett.*, 18(10):104047, 2023.

[151] R A Kerr. A north atlantic climate pacemaker for the centuries. *Science*, 288(5473): 1984–1985, June 2000. ISSN 0036-8075. doi: 10.1126/science.288.5473.1984.

[152] Sergey Kravtsov and Christopher Spannagle. Multidecadal climate variability in observed and modeled surface temperatures. *J. Clim.*, 21(5):1104–1121, March 2008. ISSN 0894-8755, 1520-0442. doi: 10.1175/2007JCLI1874.1.

[153] Zhenhao Xu, Gang Huang, Fei Ji, and Bo Liu. Multi-scale variability features of global sea surface temperature over the past century. *Frontiers in Marine Science*, 10, 2023. ISSN 2296-7745. doi: 10.3389/fmars.2023.1238320.

[154] Patrick T Brown, Wenhong Li, Eugene C Cordero, and Steven A Mauget. Comparing the model-simulated global warming signal to observations using empirical estimates of unforced noise. *Sci. Rep.*, 5:9957, April 2015. ISSN 2045-2322. doi: 10.1038/srep09957.

[155] Zachary M Labe and Elizabeth A Barnes. Predicting slowdowns in decadal climate warming trends with explainable neural networks. *Geophys. Res. Lett.*, 49(9), May 2022. ISSN 0094-8276, 1944-8007. doi: 10.1029/2022gl098173.

[156] Michael P Meredith, Loïc Jullion, Peter J Brown, Alberto C Naveira Garabato, and Matthew P Couldrey. Dense waters of the weddell and scotia seas: recent changes in properties and circulation. *Philos. Trans. A Math. Phys. Eng. Sci.*, 372(2019):20130041, July 2014. ISSN 1364-503X. doi: 10.1098/rsta.2013.0041.

[157] M W Buckley and J Marshall. Observations, inferences, and mechanisms of the atlantic meridional overturning circulation: A review. *Rev. Geophys.*, 2016. ISSN 8755-1209. doi: 10.1002/2015RG000493.

[158] A Germe, M N Houssais, C Herbaut, and others. Greenland sea sea ice variability over 1979–2007 and its link to the surface atmosphere. *J. Geophys. Res.*, 2011. ISSN 0148-0227. doi: 10.1029/2011JC006960.

[159] Avinash Kumar, Juhi Yadav, and Rahul Mohan. Seasonal sea-ice variability and its trend in the weddell sea sector of west antarctica. *Environ. Res. Lett.*, 16(2):024046, February 2021. ISSN 1748-9326. doi: 10.1088/1748-9326/abdc88.

[160] Robert A Rohde and Zeke Hausfather. The berkeley earth land/ocean temperature record. *Earth Syst. Sci. Data*, 12(4):3469–3479, December 2020. ISSN 1866-3508, 1866-3516. doi: 10.5194/essd-12-3469-2020.

[161] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, 146(730):1999–2049, July 2020. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.3803.

[162] J Hansen, R Ruedy, M Sato, and K Lo. Global surface temperature change. *Rev. Geophys.*, 2010. ISSN 8755-1209. doi: 10.1029/2010RG000345.

[163] Nathan J L Lenssen, Gavin A Schmidt, James E Hansen, Matthew J Menne, Avraham Persin, Reto Ruedy, and Daniel Zyss. Improvements in the GISTEMP uncertainty model. *J. Geophys. Res.*, 124(12):6307–6326, June 2019. ISSN 0148-0227. doi: 10.1029/2018jd029522.

[164] Boyin Huang, Matthew J Menne, Tim Boyer, Eric Freeman, Byron E Gleason, Jay H Lawrimore, Chunying Liu, J Jared Rennie, Carl J Schreck, Fengying Sun, Russell

Vose, Claude N Williams, Xungang Yin, and Huai-Min Zhang. Uncertainty estimates for sea surface temperature and land surface air temperature in NOAAGlobalTemp version 5. *J. Clim.*, 33(4):1351–1379, February 2020. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-19-0395.1.

[165] Huai-Min Zhang, J Lawrimore, Boyin Huang, M Menne, Xungang Yin, Ahira Snchez-Lugo, B Gleason, R Vose, D Arndt, J Rennie, and Claude N Williams. Updated temperature data give a sharper view of climate trends. *Eos*, July 2019. ISSN 0096-3941, 2324-9250. doi: 10.1029/2019EO128229.

[166] World Meteorological Organization. WMO confirms that 2023 smashes global temperature record. https://wmo.int/news/media-centre/wmo-confirms-2023-smashes-global-temperature-record, January 2024. Accessed: 2024-2-13.

[167] C P Morice, J J Kennedy, N A Rayner, J P Winn, E Hogan, R E Killick, R J H Dunn, T J Osborn, P D Jones, and I R Simpson. An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *J. Geophys. Res.*, 126(3), February 2021. ISSN 0148-0227. doi: 10.1029/2019jd032361.

[168] Kinya Toride, Matthew Newman, Andrew Hoell, A Capotondi, Jakob Schlör, and Dillon Amaya. Using deep learning to identify initial error sensitivity of ENSO forecasts. *ArXiv*, abs/2404.15419, 2024. ISSN 2331-8422. doi: 10.48550/arXiv.2404.15419.

[169] Jiale Lou, Youngji Joh, and Thomas Delworth. The role of long-term trends and internal variability in altering fire weather conditions in the western united states. 2024.

[170] Constantin Bône, Guillaume Gastineau, Sylvie Thiria, Patrick Gallinari, and Carlos Mejia. Separation of internal and forced variability of climate using a U-Net. *J. Adv. Model. Earth Syst.*, 16(6), June 2024. ISSN 1942-2466. doi: 10.1029/2023ms003964.

[171] World Meteorological Organization. Global temperature is likely to exceed 1.5°c above pre-industrial level temporarily in next 5 years. https://wmo.int/news/media-centre/global-temperature-likely-exceed-15degc-above-pre-industrial-level-temporarily-next-5-years, June 2024. Accessed: 2024-7-12.

Appendix A: Supporting Information for Chapter 2

A.1 Neural Network Training and Hyperparameter Tuning

The interpretable neural network architecture, shown in the blue box of Figure 2.1, is composed as follows.

1) The neural network receives two input samples, such as two global maps of sea surface temperature (SST), which are associated with two targets, such as the SST anomaly in the North Atlantic over the following five years.

2) The input samples are each multiplied by an array of trainable weights that have the same dimensions as the inputs. Each input sample is multiplied by identical trainable weights.

3) The mean squared error (MSE) between the two input*weights layers is calculated.

4) The computed MSE is fed into a series of fully-connected dense layers. These dense layers are intended to find a relationship between the weighted MSE and the absolute difference between the targets associated with each of the inputs (which is the predictand for this neural network task).

There are four main tunable parameters for the interpretable neural network: the learning rate, the L2 regularization applied to the mask (acts to smooth out the weights and reduce overfitting), the size of the dense layers, and the activation function for the dense layers.

A different neural network architecture is tuned for each prediction problem. The prediction problems/experiments are: EXP-Niño, predicting NDJFM Niño3.4 SST anomalies given global NDJFM SST one year prior, and EXP-NorAtl, predicting 5-year SST anomalies in the North Atlantic given global SSTs in the five years prior. In Figure A.5, we also show results for EXP501 - predicting 5-year SST anomalies in the North Atlantic given global SSTs in the five years prior and the difference between the global SSTs in the five years prior and the period 3-7 years prior (i.e. the sea surface temperature tendency). The same

hyperparameters that were tuned for EXP-NorAtl are used for EXP501.

To tune each experiment the following procedure was performed:
1) Tune the neural network using the constants in Table A.1 and the hyperparameter search space in Table A.2. Train 100 total models and assess their loss on validation data (not used for training or early stopping). This is the base hyperparameter search, and will be used to constrain the search space for more tuning.
2) Identify the top-10 models in terms of validation loss from the base hyperparameter search. Constrain the hyperparameter space to the ranges of hyperparameters that appeared in these 10 best models. This constrained hyperparameter space is referred to as the "refined" hyperparameter space. For the dense layers, all configurations are retained that have a number of trainable parameters captured by the minimum and the maximum number of trainable parameters within the dense layers (not including the input weights) of the 10 best models.
3) Tune the neural network by training 100 new models using a random search of the refined hyperparameter space in Tables A.3- A.4.

The hyperparameters associated with the model with the best validation loss in the refined search were used for the results. These are shown in Tables A.5- A.6. The results for models using the random seed of 0, which is important for the initialization of weights and the random selection of samples for neural network training, are shown in the main text. Additionally, the results for models trained with the random seeds of 10, 20, 30, 40, 50, 60, 70, 80, and 90 can be found here in Figures A.1 and A.2.

A.2 Neural Network Overview

The description of the mean target evolution baseline was withheld from the main text, and is instead supplied here. We make a mean target evolution forecast by first binning

the samples in the training set based on the target values *during the input period*. The mean evolution of each bin is determined by taking all the samples within that bin and calculating the mean target value during the forecast window. The mean target evolution forecast is made by then identifying which bin each sample from the test set falls into, and using the mean evolution of that bin as the prediction.

In addition to the baselines in the main text, we present one additional baseline in the supplement: the skill of a “vanilla model." The vanilla model is your typical feed forward artificial neural network. Given a state of interest as input, the vanilla model is tasked to predict the target. It is not constrained to follow the analog framework.

A.3 Application to Observations

We apply the weighted mask, learned on the MPI-GE, to predict the SST in the North Atlantic over five-year periods in observations. Here, we use the European Centre for Medium-Range Weather Forecasts Reanalysis version 5, ERA5, as observations [161]. Since we separated MPI-GE into its forced and unforced components, and we do not know the true forced signal in observations, we elected not to remove the forced signal within observations. We standardized the observations relative to the MPI-GE mean and standard deviation. The predictions for observations of 5-year North Atlantic SST is shown in Figure A.7. These results are compared with the decadal forecasts of CFSv2-2011 which was initialized for hindcasts every fifth year between 1960 and 2005 [18]. In this limited sample size, we find that the weighted analogs do a decent job of predicting observations relative to the initialized CFSv2 hindcasts (Table A.7). ERA5 SST data can be found on Copernicus Climate Data Store (https://cds.climate.copernicus.eu) and CFSv2 data was provided by the Earth System Grid Federation (https://esgf-node.llnl.gov/projects/esgf-llnl/).

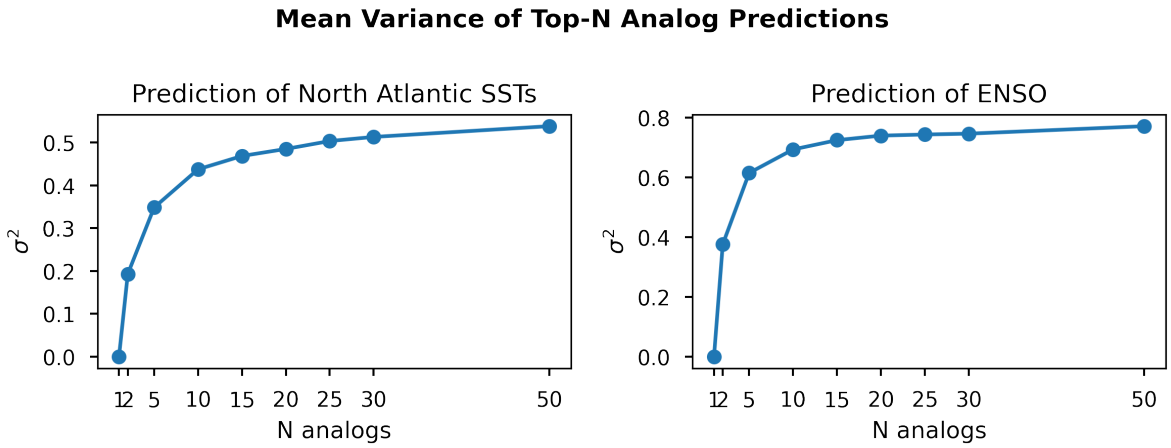

Figure A.1: Mean variance of the targets associated with the top-N analogs across all testing samples.

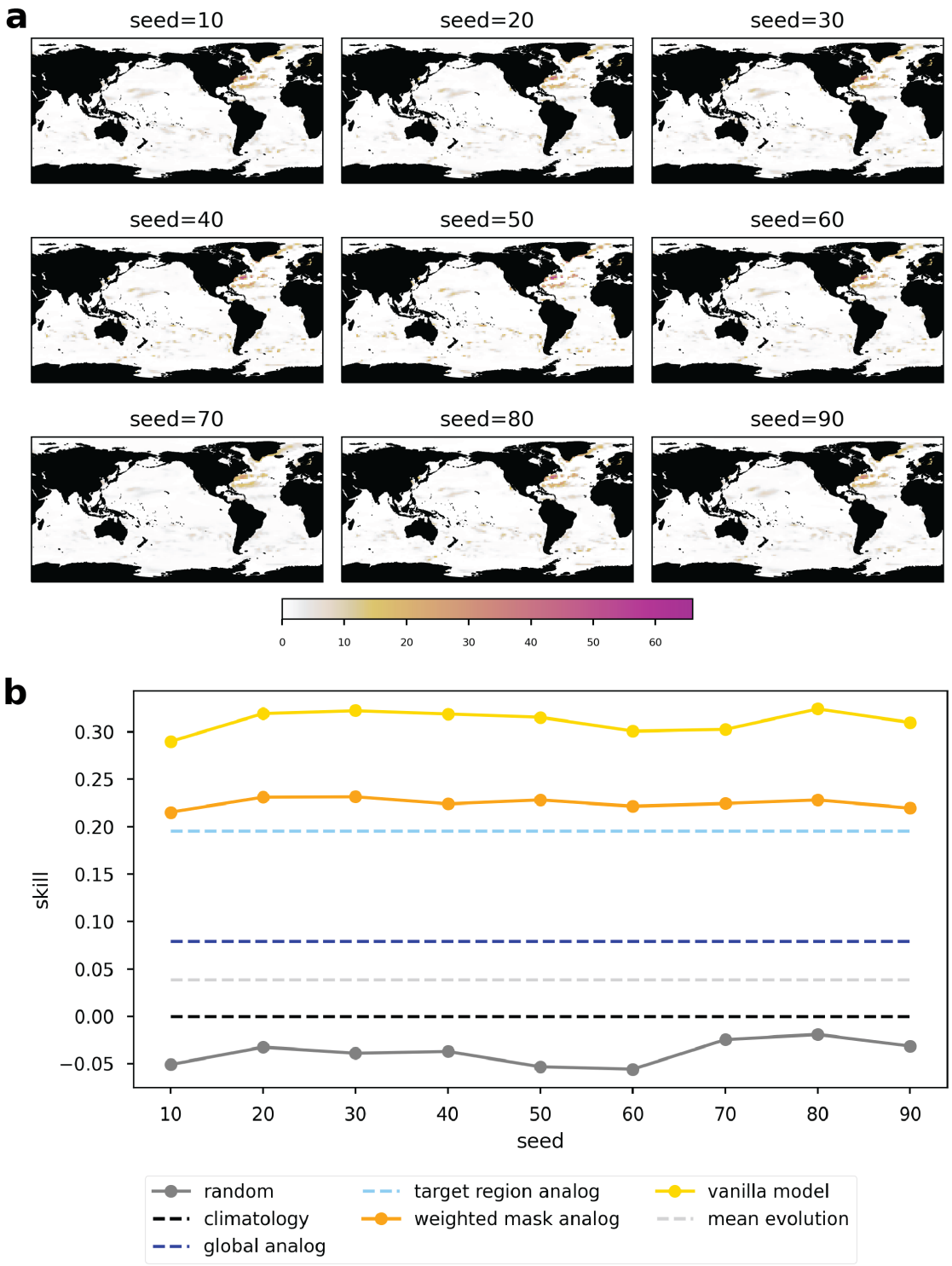

Figure A.2: EXP-NorAtl: results for neural networks trained on nine different seeds. (a) Nine neural networks trained on different seeds show striking consistency in their weighted masks. (b) Skill scores for the average of the top-10 analogs. In all cases, the highest skill comes from the vanilla model, followed by the analog models. The masked analog outperforms the baselines discussed in the main text.

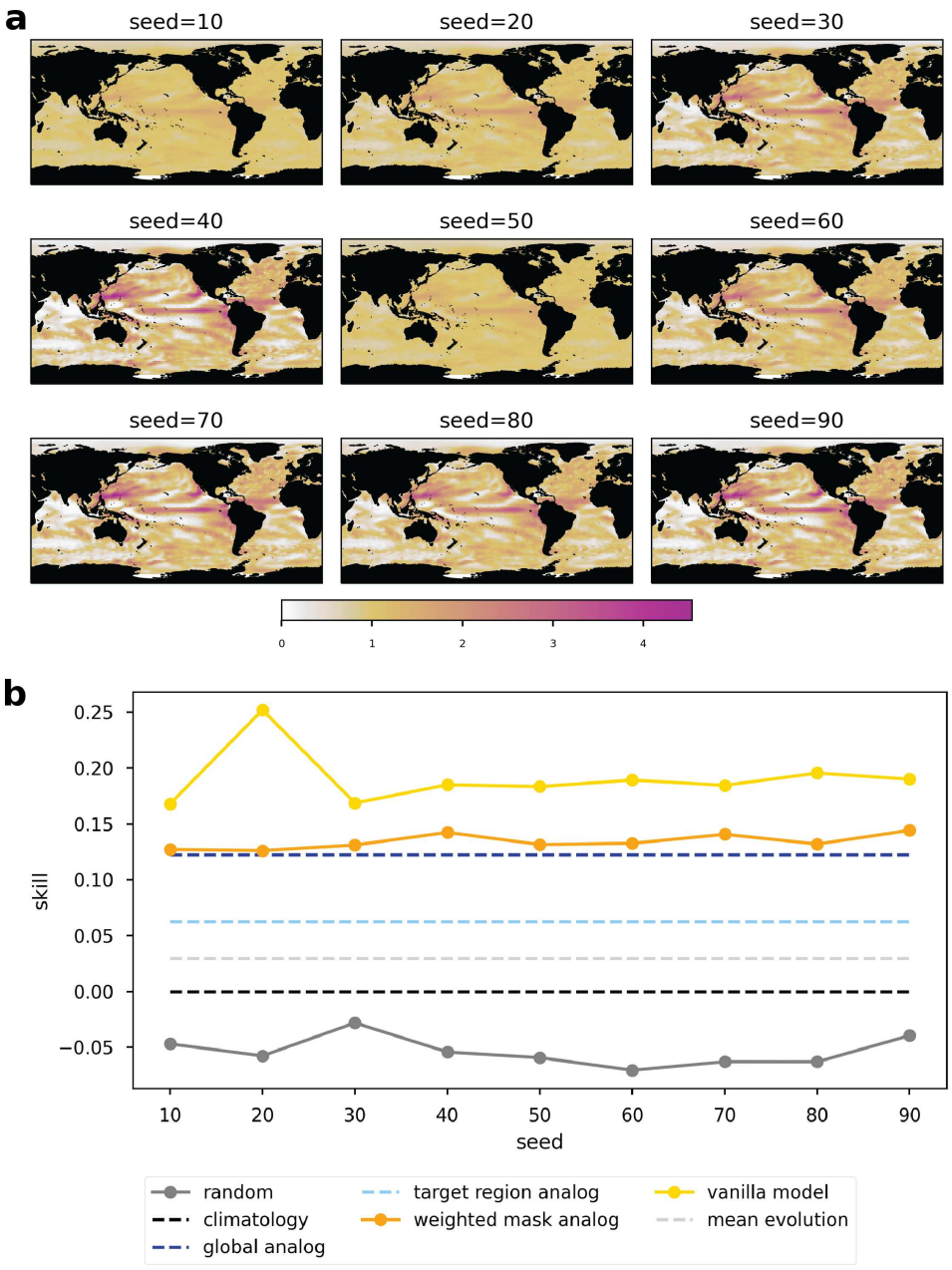

Figure A.3: EXP-Niño: results for neural networks trained on nine different seeds. (a) Changing the seed used for the neural network training results in slight variation in the weighted mask. However, all weighted masks highlight the central tropical Pacific, western Pacific, Baja coast, and central Atlantic as the most important (though to varying degrees). (b) Skill scores for the average of the top-10 analogs. The vanilla model outperforms the weighted mask analog model across the board. In all cases the masked analog outperforms the baselines discussed in the main text.

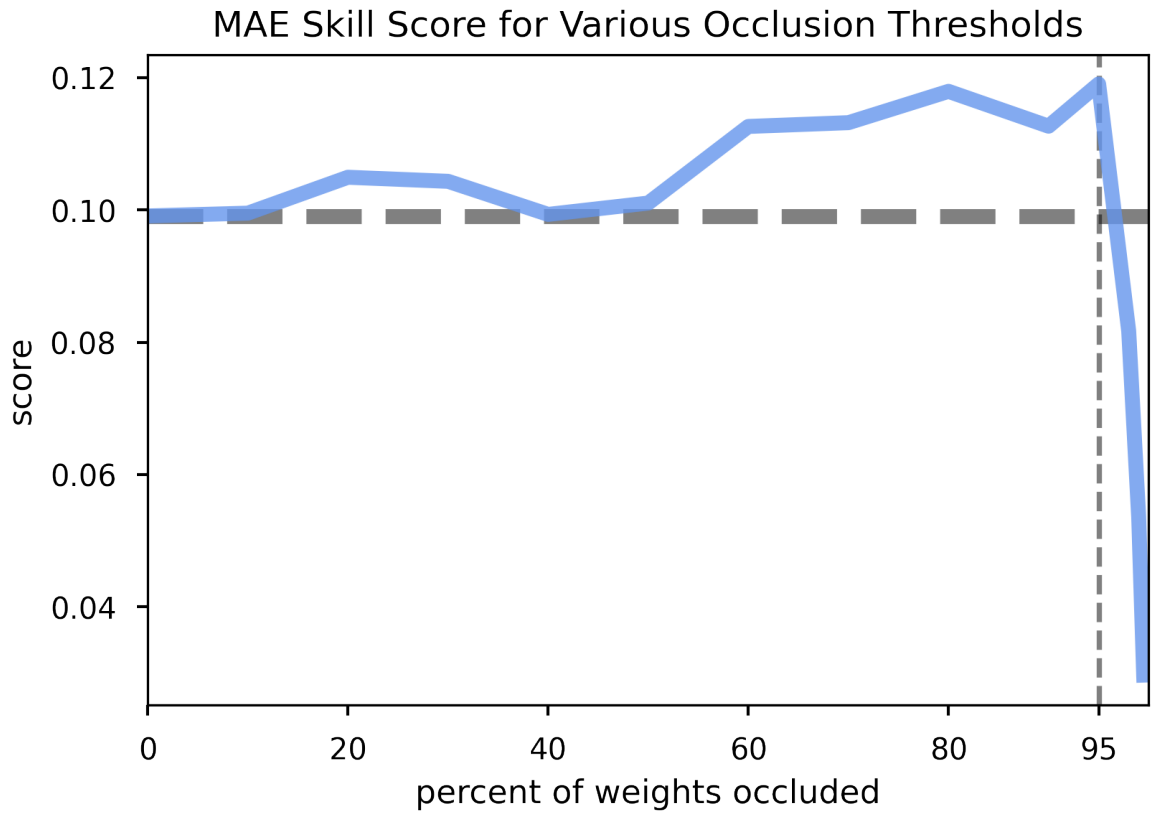

Figure A.4: The skill of an analog forecast for EXP-Niño using the weighted mask when the smallest weights are set to zero. The horizontal line indicates the forecasting skill before the weighted mask has been altered. The vertical line indicates the forecasting skill using a weighted mask where the smallest 95 percent of the weights have been set to zero. Removing the smallest weights does not have much of an impact on forecasting skill, and may even improve it. These results are for the validation data set.

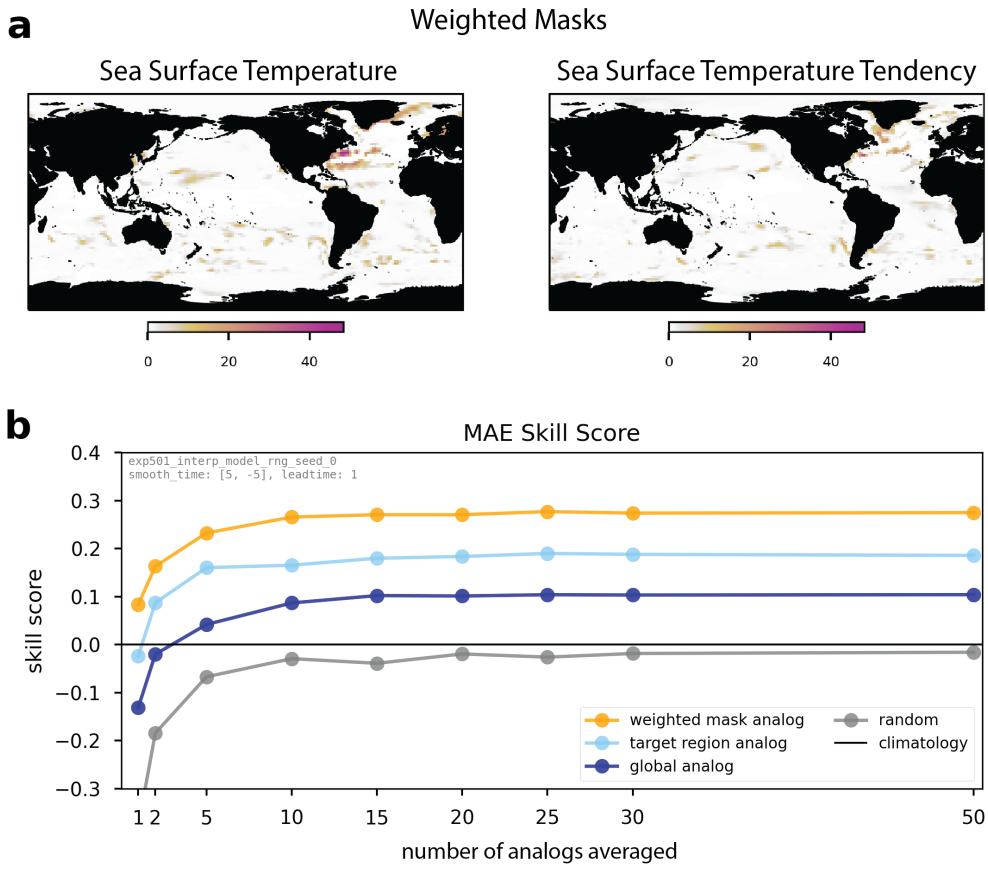

Figure A.5: (a) Weighted masks for EXP501. (b) Skill scores for EXP501 versus various baselines. Adding an SST tendency input field did not notably impact forecasting skill in this problem.

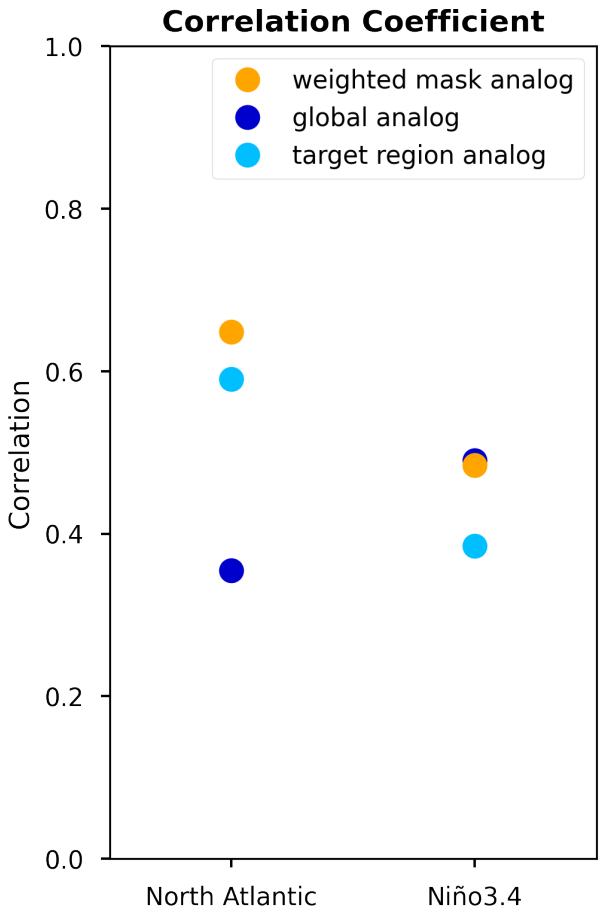

Figure A.6: Correlation coefficients for predictions made by analog forecasts for EXP-NorAtl and EXP-Niño.

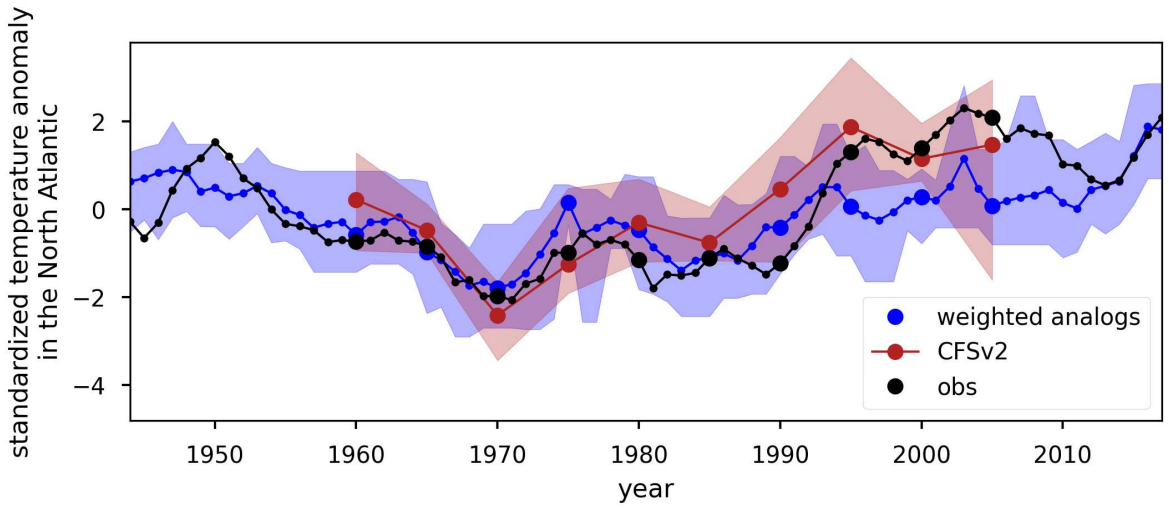

Figure A.7: The mean prediction from the top 10 MPI-GE weighted analogs and four members of CFSv2 for multi-year prediction of the observed (ERA5) North Atlantic. The large dots indicate the predictions during years that CFSv2 was initialized while the small dots indicate the intermediate years. The shading reflects the minimum and maximum predictions of the 10 weighted analogs and the four CFSv2 members.

Analog Members	0 through 34
Training SOI Members	35 through 49
Validation SOI Members (early stopping)	50 through 54
Validation SOI Members (tuning)	55 through 60
Testing SOI Members	95 through 99
Loss Function	MSE
Early Stopping Patience (epochs)	50
Early Stopping Minimum Delta	0.0005
Maximum # of Epochs	5,000
Validation Batch Size	2,500
Mask Activation Function *	relu
Mask Initial Value *	ones
Dense Layer Weight and Biases Initial Values	random normal

Table A.1: Constant values for all neural networks trained.

Dense Layers	0-3 Layers, with 1, 2, 5, 10, 20, 50 or 100 nodes in all layers.
Activation Function	elu, relu, tanh
L2 Regularization applied to Mask * L2 Regularization applied to Input ^	0.0, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1
Learning Rate	0.01, 0.001, 0.0001

Table A.2: Base hyperparameter search space for identifying the best neural network architecture for each experiment.

Interpretable Analog Model	
Dense Layers	1 Layer with 5, 10, 20, 50, 100 nodes 2 Layers with 2, 5, 10, 20, 50 nodes 3 Layers with 1, 2, 5, 10, 20, 50 nodes
Activation Function	elu, relu, tanh
L2 Regularization applied to Mask	0.0
Learning Rate	0.01, 0.001, 0.0001
Vanilla Model	
Dense Layers	1 Layer with 5, 10, 20, 50, 100 nodes 2 Layers with 2, 5, 10, 20, 50 nodes 3 Layers with 2, 5, 10, 20 nodes
Activation Function	elu, relu
L2 Regularization applied to Input	0.0, 1e-5, 1e-4, 1e-3, 1e-2
Learning Rate	0.01, 0.001, 0.0001
Vanilla Analog Model	
Dense Layers	1 Layer with 5, 10, 20, 50, 100 nodes 2 Layers with 2, 5, 10, 20, 50, 100 nodes 3 Layers with 2, 5, 10, 20, 50, 100 nodes
Activation Function	elu, relu
L2 Regularization applied to Input	0.0, 1e-5, 1e-4
Learning Rate	0.01, 0.001, 0.0001

Table A.3: Refined hyperparameter search space for EXP-Niño (seasonal prediction of El Niño Southern Oscillation).

Interpretable Analog Model	
Dense Layers	1 Layer with 5, 10, 20, 50, 100 nodes 2 Layers with 2, 5, 10, 20, 50, 100 nodes 3 Layers with 1, 2, 5, 10, 20, 50 nodes
Activation Function	elu, relu, tanh
L2 Regularization applied to Mask	0.0
Learning Rate	0.01, 0.001, 0.0001
Vanilla Model	
Dense Layers	1 Layer: 1, 2, 5, 10, 20, 50, 100 nodes 2 Layers with 1, 2, 5, 10, 20, 50 nodes 3 Layers with 1, 2, 5, 10, 20 nodes
Activation Function	elu, relu, tanh
L2 Regularization applied to Input	0.0, 1e-5, 1e-4
Learning Rate	0.01, 0.001, 0.0001
Vanilla Analog Model	
Dense Layers	1 Layer with 5, 10, 20, 50, 100 nodes 2 Layers with 5, 10, 20, 50, 100 nodes 3 Layers with 2, 5, 10, 20, 50, 100 nodes
Activation Function	elu, relu
L2 Regularization applied to Input	0.0, 1e-5, 1e-4, 1e-3
Learning Rate	0.01, 0.001, 0.0001

Table A.4: Refined hyperparameter search space for EXP-NorAtl (decadal prediction of the North Atlantic).

Interpretable Analog Model	
Dense Layers	[20, 20]
Activation Function	tanh
L2 Regularization applied to Mask	0.0
Learning Rate	0.0001
Vanilla Model	
Dense Layers	[2, 2]
Activation Function	relu
L2 Regularization applied to Input	1e-5
Learning Rate	0.0001
Vanilla Analog Model	
Dense Layers	[50, 50]
Activation Function	relu
L2 Regularization applied to Input	0.0
Learning Rate	0.0001

Table A.5: Chosen hyperparameters for EXP-Niño (seasonal prediction of El Niño Oscillation).

Interpretable Analog Model	
Dense Layers	[20, 20]
Activation Function	elu
L2 Regularization applied to Mask	0.0
Learning Rate	0.01
Vanilla Model	
Dense Layers	[100]
Activation Function	relu
L2 Regularization applied to Input	0.0
Learning Rate	0.0001
Vanilla Analog Model	
Dense Layers	[100, 100]
Activation Function	relu
L2 Regularization applied to Input	0.0
Learning Rate	0.0001

Table A.6: Chosen hyperparameters for EXP-NorAtl (decadal prediction of the North Atlantic). These were also the hyperparameters used for EXP501 (decadal prediction of the North Atlantic with a time lag input–see Figure A.5).

	MAE Skill Score	Correlation
Weighted Analog	0.35	0.73
Uniform Analog	0.15	0.55
CFSv2 Mean	0.45	0.86
CFSv2 Member 1	0.45	0.83
CFSv2 Member 2	0.31	0.86
CFSv2 Member 3	-0.13	0.44
CFSv2 Member 4	0.14	0.69

Table A.7: MAE skill score and Pearson's corrleation coefficient for observational predictions using MPI-GE analogs with the weighted mask, compared to a globally uniform analog and CFSv2 (without volcanic forcing). Note that the skill score and correlation is calculated for every fifth year between 1960 and 2005 because CFSv2 was only initialized in these years. The skill scores and correlations for CFSv2 with volcanic forcing, which are not included here, were lower than CFSv2 without volcanic forcing.

Appendix B: Supporting Information for Chapter 3

B.1 Learning Patterns of Internal Variability and the Forced Response

We assess the extent to which our neural network architecture has learned patterns of internal variability by creating a validation set using eight of the members from each climate model ensemble. In these eight ensembles, we shuffle the internal variability across time at each grid point. In this way, we remove coherent spatial patterns of internal variability, such that the maps of annual-mean SST look like noise on top of the forced response (Figure B.7). We then make estimates for these maps with shuffled internal variability using our trained neural networks. Figure B.8 shows the skill of our neural networks using leave-one-out cross-validation on the validation sets without shuffling and the validation sets with shuffling. The neural networks perform poorly on these maps with shuffled internal variability, suggesting that the neural network is learning patterns of internal variability. This is not the case for a simple linear model with a skip connection, which has comparable skill for both the maps with physically consistent patterns of internal variability and shuffled internal variability, suggesting it has only learned patterns of the forced response. There are other reasons that a neural network may not perform well on maps with shuffled internal variability, since they fall outside of the training distribution. However, it provides one line of evidence that the neural network has learned patterns of internal variability.

B.2 Neural Network Architecture

Our neural network architecture consists of the following layers:

- Input
- Encoding1 - Convolution 3x3
- Encoding2 - Convolution 3x3
- Pool1 - MaxPool 2x2

- Encoding3 - Convolution 3x3
- Encoding4 - Convolution 3x3
- Pool2 - MaxPool 2x2
- Dense1 - 100 Dense Nodes
- Dense2 - 100 Dense Nodes
- Dense3 - 100 Dense Nodes
- Dense4 - 100 Dense Nodes
- Dense5 - 100 Dense Nodes
- Reshape - Reshape to Dimensions of Pool2
- Concatenate1 - Concatenate [Reshape1, Pool2]
- Upsampling1 - Upsample 2x2
- Decoding1 - Transpose Convolution 3x3
- Decoding2 - Transpose Convolution 3x3
- Concatenate2 - Concatenate [Decoding2, Pool1]
- Upsampling2 - Upsample 2x2
- Decoding3 - Transpose Convolution 3x3
- Decoding4 - Transpose Convolution 3x3
- Concatenate3 - Concatenate [Decoding4, Input]
- Decoding5 - Transpose Convolution 1x1

All Dense layers used a tanh activation function, and their weights and biases were initialized using a random normal distribution. All 3x3 Convolutions and Transpose Convolutions consisted of 32 filters and a stride of one, and used a relu activation function, the weights were initialized using a Glorot uniform distribution and the biases were initialized at zero. The Decoding5 layer used a linear activation function and one filter, but was otherwise the same as the other Decoding layers.

B.3 Neural Network Tuning

We explored a vast array of neural network architectures for this study which are shown in Figure B.9 and Table B.1. Our goal was to select a neural network architecture that was both skillful on out-of-sample climate model data and learned patterns of both internal variability and the forced response. Since climate models have similar biases in the forced response, learning patterns of internal variability may allow the neural network to learn more nuanced patterns than just the simulated forced response (see Figure B.1).

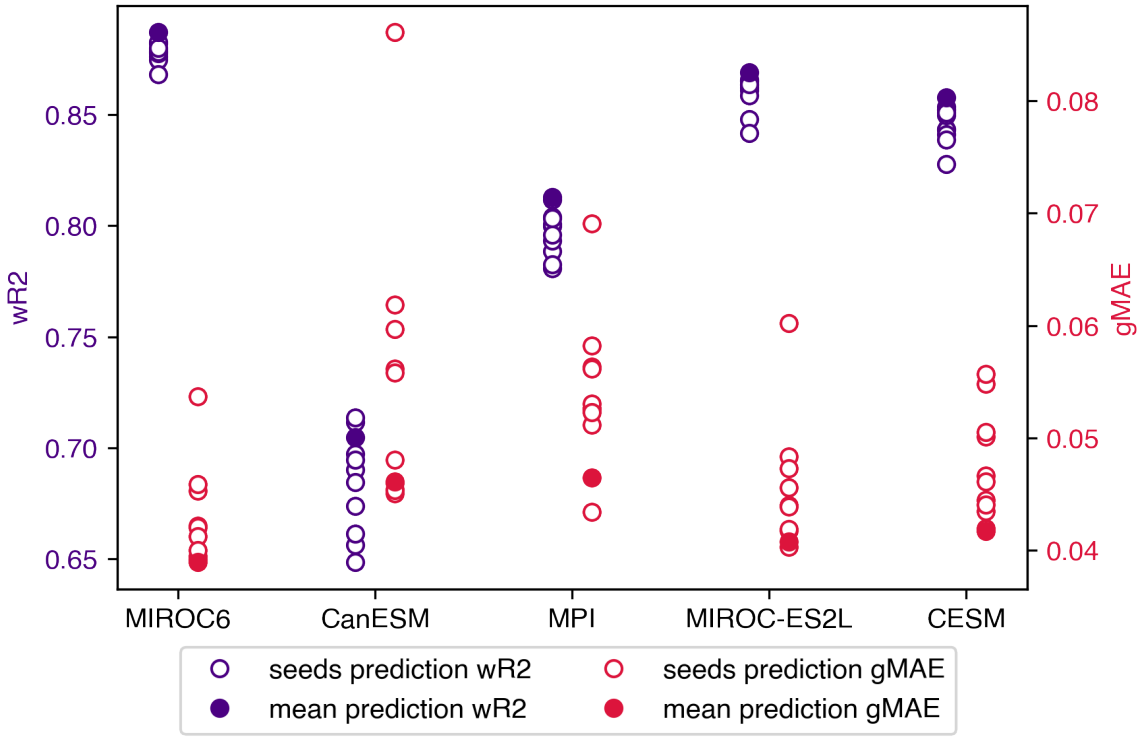

Figure B.1: Area-weighted R^2 scores (wR^2) and MAE of the global mean temperature (gMAE) for estimates from individual seeds and the mean across seeds. Generally, taking the mean estimate across seeds allows for better performance than using any single individual seed. Note that a higher wR^2 is more skillful than a lower wR^2, while a lower gMAE is more skillful than a higher gMAE.

.

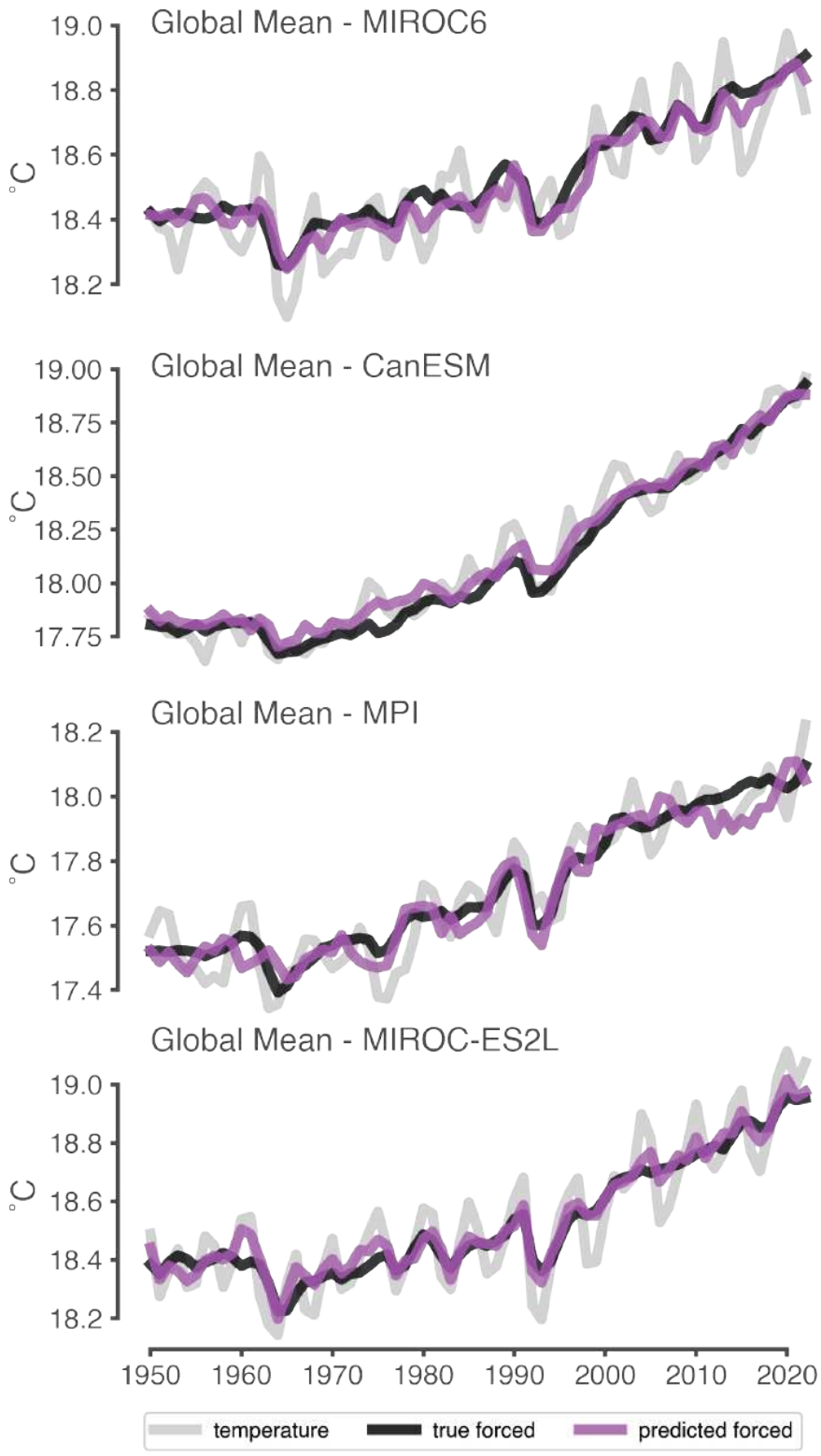

Figure B.2: Estimated forced component of global-annual-mean SST for one member from each climate model, using the neural networks trained for leave-one-out cross-validation.

.

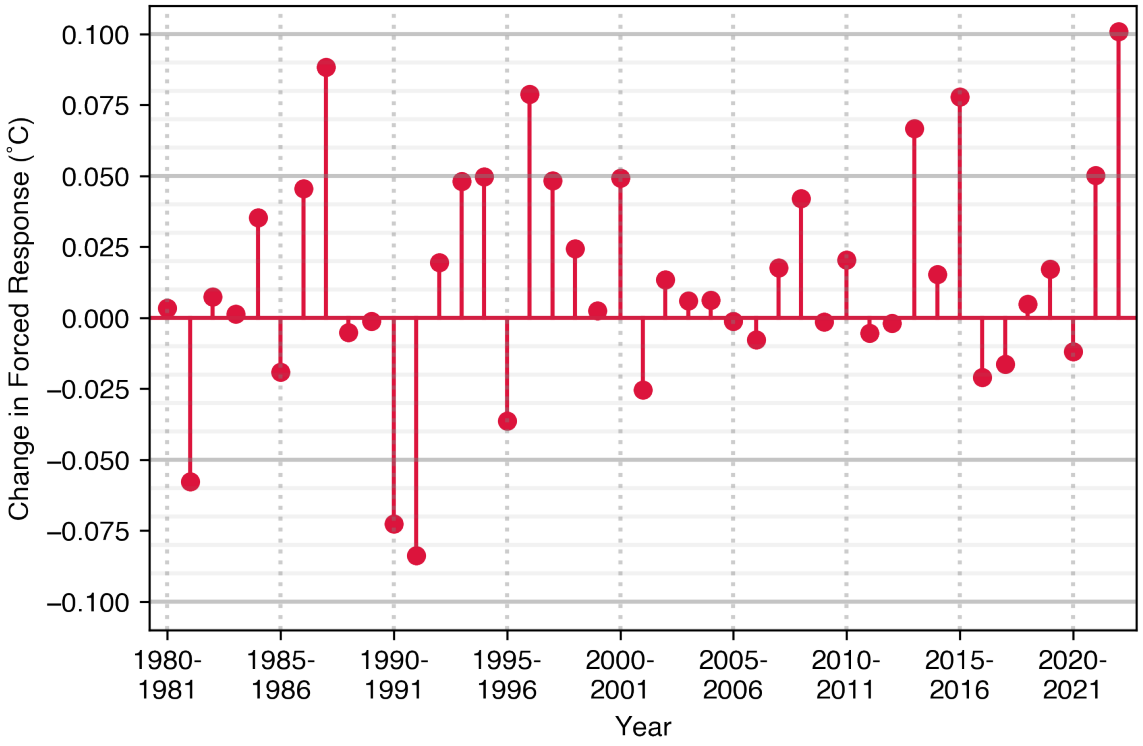

Figure B.3: Year-to-year change in the estimated forced component of global-annual-mean SST, 1980-2023.

.

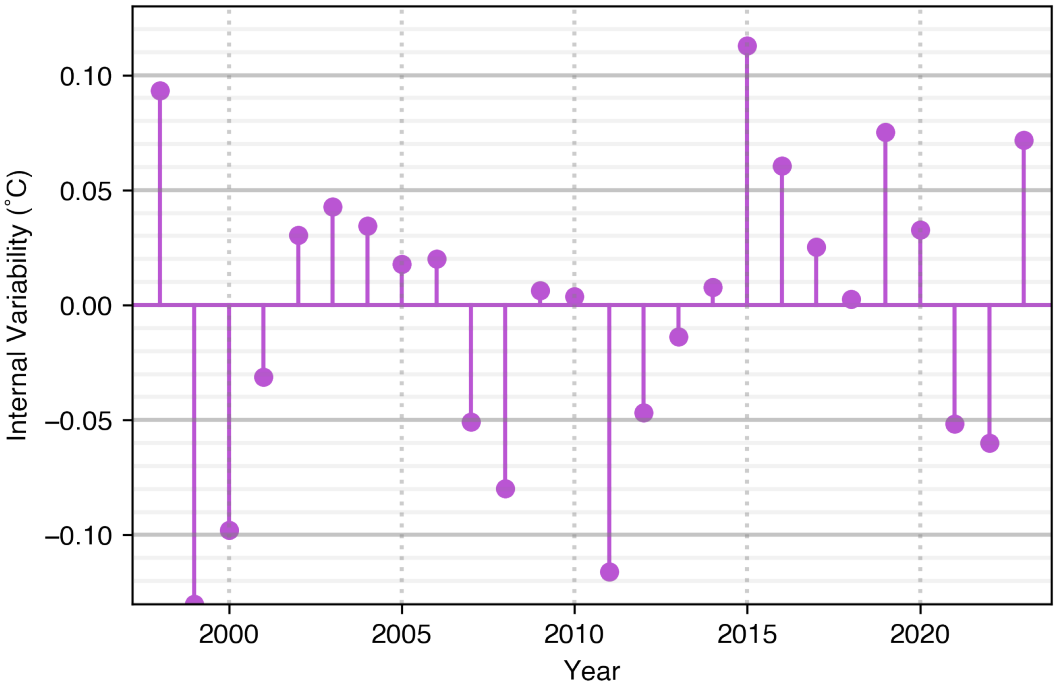

Figure B.4: Estimated internal component of global-annual-mean SST, 1998-2023.

.

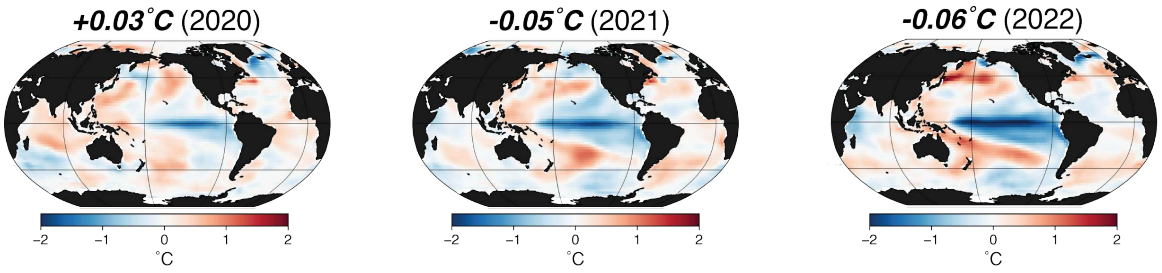

Figure B.5: Internal variability estimates for the 2020-2022 "triple-dip" of cold phase ENSO events.

.

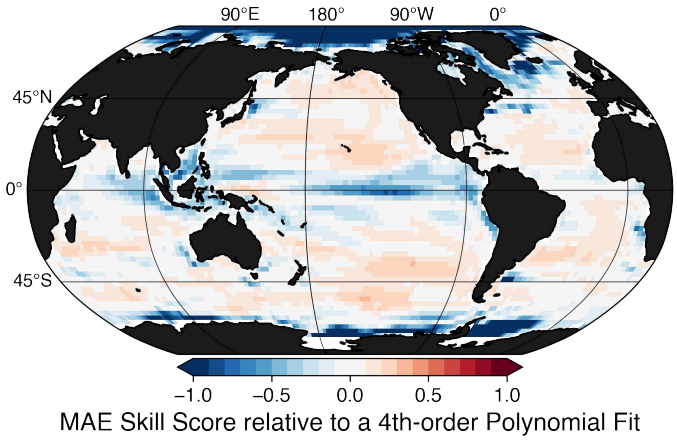

Figure B.6: MAE Skill Score relative to estimating the forced response as the fourth-order polynomial fit to all data. Values larger than zero indicate the neural network estimates of internal variability are more skillful than a fourth-order polynomial fit, while values less than zero indicate the opposite. These are assessed on all the leave-one-out validation climate models.

.

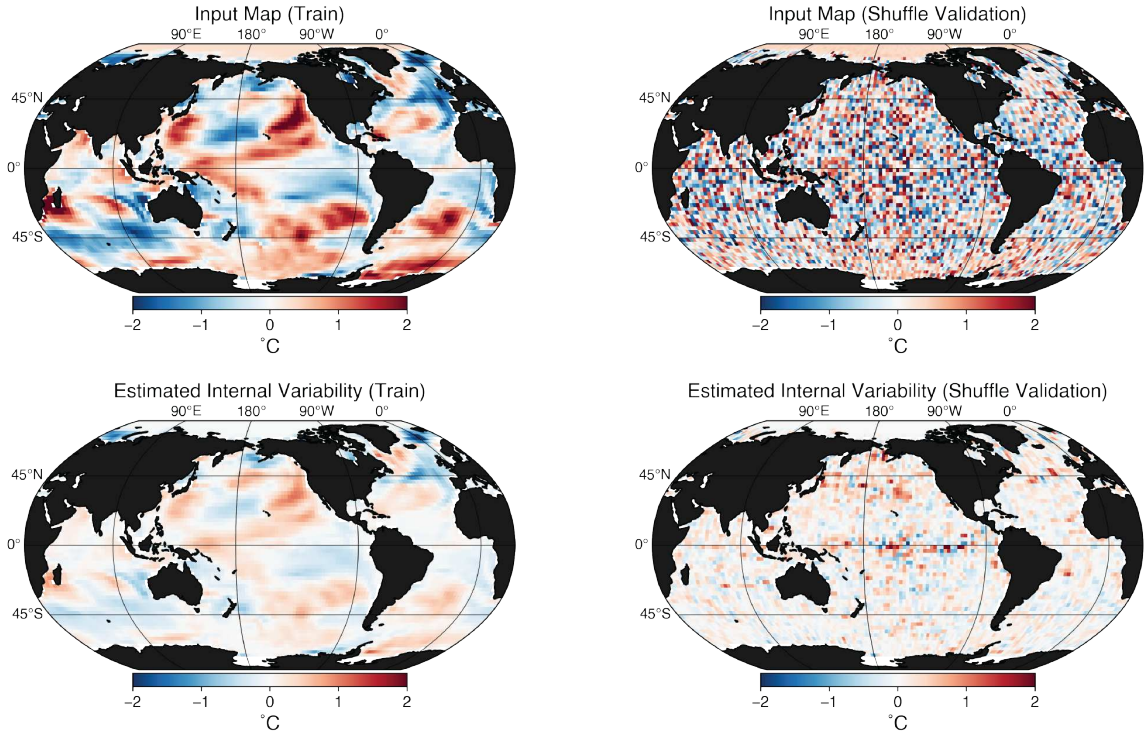

Figure B.7: Example input maps and the corresponding internal variability estimate for a sample in the training set and a sample in the validation set with internal variability shuffled.

.

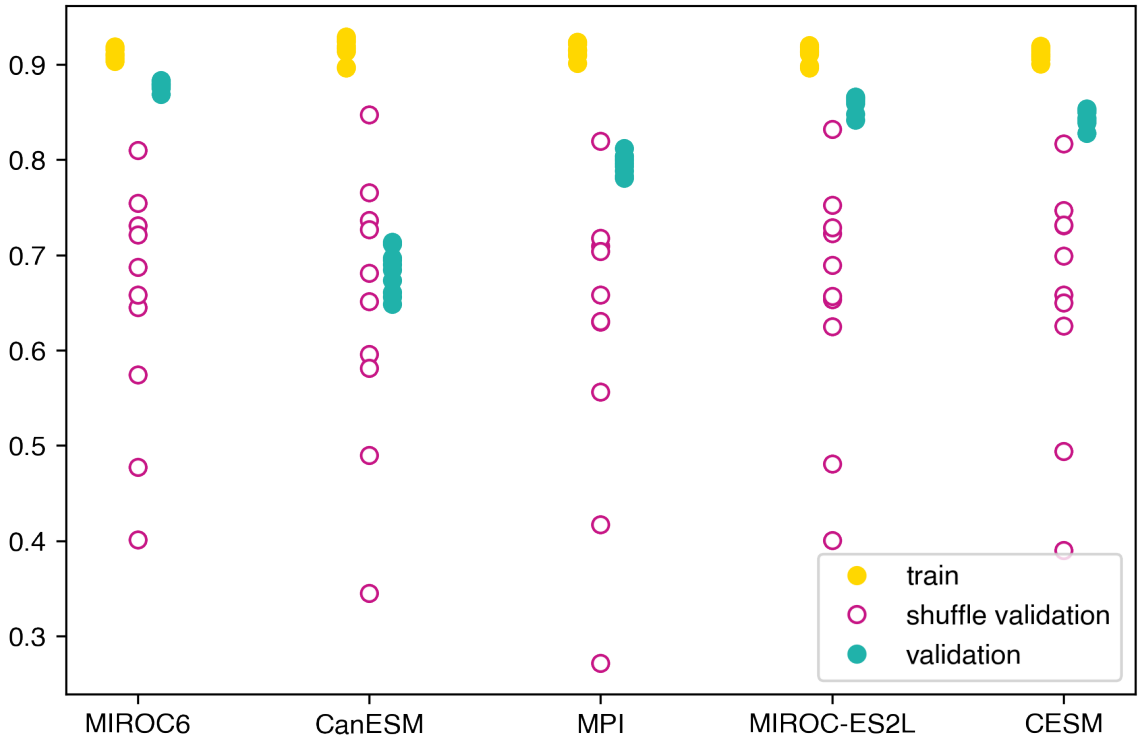

Figure B.8: Metrics for the training set, shuffle validation, and validation data using neural networks trained for leave-one-out cross-validation. The shuffle validation uses the same climate models as the training set, but eight different members, with the interval variability shuffled, as shown in Figure B.7.

.

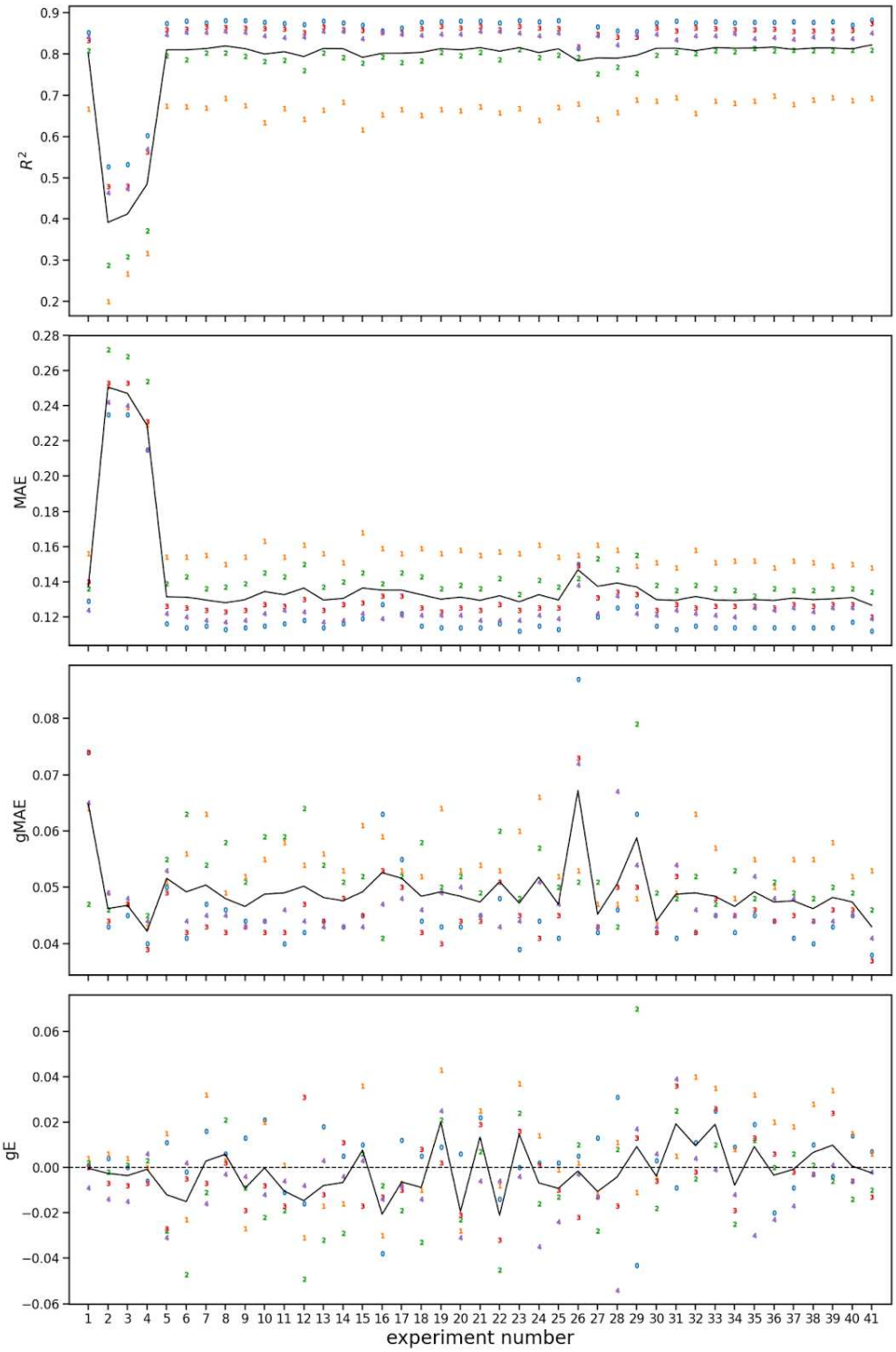

Figure B.9: Metrics for neural network tuning experiments: R^2, mean absolute error (MAE), global MAE (gMAE), and mean global error (gE). Metric values for each of the leave-one-out experiments are shown with numbers (1=MIROC6, 2=CanESM5, 3=MPI-ESM1-2-LR, 4=MIROC-ES2L, 5=CESM2), and the mean of these five are shown as a solid line. Details for each experiment can be found in Tables B.1 and B.2.

.

	activation function	learning rate	training epochs	dense layers	convolution layers	notes
1	linear		13	1 x 1	initial skip	early stop (15)
2					no layers	only dense
3					initial skip	dense + skip
4					no skips	
5					single layer	
6						all defaults
7					three layer	
8			10		three layer	
9					mean pooling	
10					linear activation	
11					tanh activation	
12					16 filters	fewer filters
13					64 filters	more filters
14					2 x 2 kernel	small kernel
15					6 x 6 kernel	large kernel
16				[100, 100, 1, 100, 100]		1 code node
17				5 x 10		shallow layers
18				5 x 1000		deep layers
19			10	5 x 1000		deep layers
20				[1000, 100, 10, 100, 1000]		encoder / decoder

Table B.1: Details of each tuning experiment, continues to Table 2. Unless otherwise noted, defaults are the following: activation function = 'tanh', learning rate = 0.001, training epochs = 5, dense layers = 5 x 100, convolution layers = [Skip, Conv, Conv, MaxPool(2), Skip], [Conv, Conv, MaxPool(2), Skip] with Conv a 32 filter, 3x3 kernel, 1 stride convolution using a 'relu' activation. The three layer convolution = [Skip, Conv, Conv, MaxPool(2), Skip], [Conv, Conv, MaxPool(2), Skip], [Conv, Conv, MaxPool(2), Skip], while the single layer convolution = [Skip, Conv, Conv, MaxPool(2), Skip].

.

21			10	[1000, 100, 10, 100, 1000]		encoder / decoder
22				7 x 100		wider layers
23			10	7 x 100		wider layers
24	linear					
25	relu					
26		0.01				high learning
27		0.0001				low learning
28			1			
29			2			
30			4			
31			6			
32			8			
33			10			
34			12			
35			14			
36			16			
37			18			
38			20			
39			22			
40			24			
41	relu		10	[1000, 100, 10, 100, 1000]	single layer	best metrics

Table B.2: Continuation of Table 1.

.

Appendix C: Supporting Information for Chapter 4

C.1 Neural Network Training

The training process consists of three parts. First, the base network layers are trained, then the shift factor layers are trained, then the uncertainty scaling factor layers are trained, all using early stopping on the validation set to prevent overfitting. If training the shift factor or uncertainty scaling factor layers resulted in a higher (worse) validation loss for all epochs of training, the neural network reset to the weights it had before training when the validation loss was better. The training process is described in further detail in the caption of Figure C.9, and the hyperparameters for each neural network are presented in Table C.2. In the following, we show which CMIP6 models were used (for two forcing scenarios, SSP3-7.0 and SSP2-4.5) as well as which models were part of the validation set (withheld from training and used for early stopping).

Climate Models for SSP3-7.0:
ACCESS-CM2, ACCESS-ESM1-5, AWI-CM-1-1-MR, BCC-CSM2-MR, CanESM5, CAS-ESM2-0, CESM2, CESM2-WACCM, CMCC-CM2-SR5, CMCC-ESM2, CNRM-CM6-1, CNRM-CM6-1-HR, FGOALS-f3-L, FGOALS-g3, GISS-E2-1-G, GFDL-ESM4, IITM-ESM, INM-CM4-8, INM-CM5-0, IPSL-CM5A2-INCA, IPSL-CM6A-LR, KACE-1-0-G, MIROC6, MIROC-ES2L, MPI-ESM1-2-LR, MRI-ESM2-0, NorESM2-LM, NorESM2-MM, TaiESM1, UKESM1-0-LL

Climate Models for SSP2-4.5:
ACCESS-CM2, ACCESS-ESM1-5, AWI-CM-1-1-MR, BCC-CSM2-MR, CanESM5, CAS-ESM2-0, CESM2, CESM2-WACCM, CMCC-CM2-SR5, CMCC-ESM2, CNRM-CM6-1, CNRM-CM6-1-HR, CNRM-ESM2-1, GFDL-CM4, GFDL-ESM4, GISS-E2-1-G, FGOALS-f3-L, FGOALS-g3, IPSL-CM6A-LR, MIROC6, MIROC-ES2L, MRI-ESM2-0, NorESM2-LM

Models in the Validation Set for One-year Predictions (2024):
ACCESS-ESM1-5, CMCC-CM2-SR5, CMCC-ESM2, CNRM-CM6-1, IPSL-CM5A2-INCA, IPSL-CM6A-LR, INM-CM4-8, INM-CM5-0, MIROC6, MPI-ESM1-2-LR

Models in the Validation Set for Five-year Predictions (2024-2028):
ACCESS-ESM1-5, CMCC-CM2-SR5, CMCC-ESM2, CNRM-CM6-1, IPSL-CM5A2-INCA, IPSL-CM6A-LR, INM-CM4-8, INM-CM5-0, MIROC6, MPI-ESM1-2-LR

Models in the Validation Set for Ten-year Predictions (2024-2033):
ACCESS-CM2, ACCESS-ESM1-5, BCC-CSM2-MR, CAS-ESM2-0, CMCC-CM2-SR5, CMCC-ESM2, FGOALS-f3-L, GFDL-ESM4, INM-CM5-0, NorESM2-MM

C.2 Neural Network Tuning

The selected hyperparameters for the neural network can be found in Table C.2. We followed the following procedure for hyperparameter tuning.

1. We trained 30 base neural networks with randomly selected hyperparameters. The training/validation split seed, which determines which GCMs fall in the training and validation splits, and the initialization seed, which determines the initial values of the weights and biases, were fixed at 0. The hyperparameter search space can be found in Table C.3. We selected the same base network hyperparameters for the one-, five-, and ten-year prediction tasks, and show in Figure C.10 that the selected hyperparameters result in one of the best models in terms of validation loss. We also show in Figure C.11 that changing the train/val split seed and the initialization seed does not largely impact the loss on observations.

2. Once the hyperparameters for the base network were selected, we trained 100 final neural networks with different hyperparameter choices, but with the same train-

ing/validation split. The hyperparameter search space for the final neural network can be found in Table C.3. We then repeated this four more times with different training/validation splits, and selected the hyperparameters that resulted in the best average loss on the validation set across the five training/validation sets (Figure C.12).

3. Once the best hyperparameters were identified, we again trained 100 more networks, across 10 different training/validation splits and 10 different initialization seeds (which determines the initial values of the weights and biases).

4. We selected the final neural network, which we use for the results in this paper, by considering how well it performs on 1980-2020 observational hindcasts (measured by loss), how well it performed on the validation data (measured by the PIT-D metric), and verified that the likelihood of exceeding temperature thresholds for the selected network was representative of all 100 trained networks as a whole. The selected network, and the metrics for all 99 networks that were not included in the main text are included in Figures C.7 and C.8.

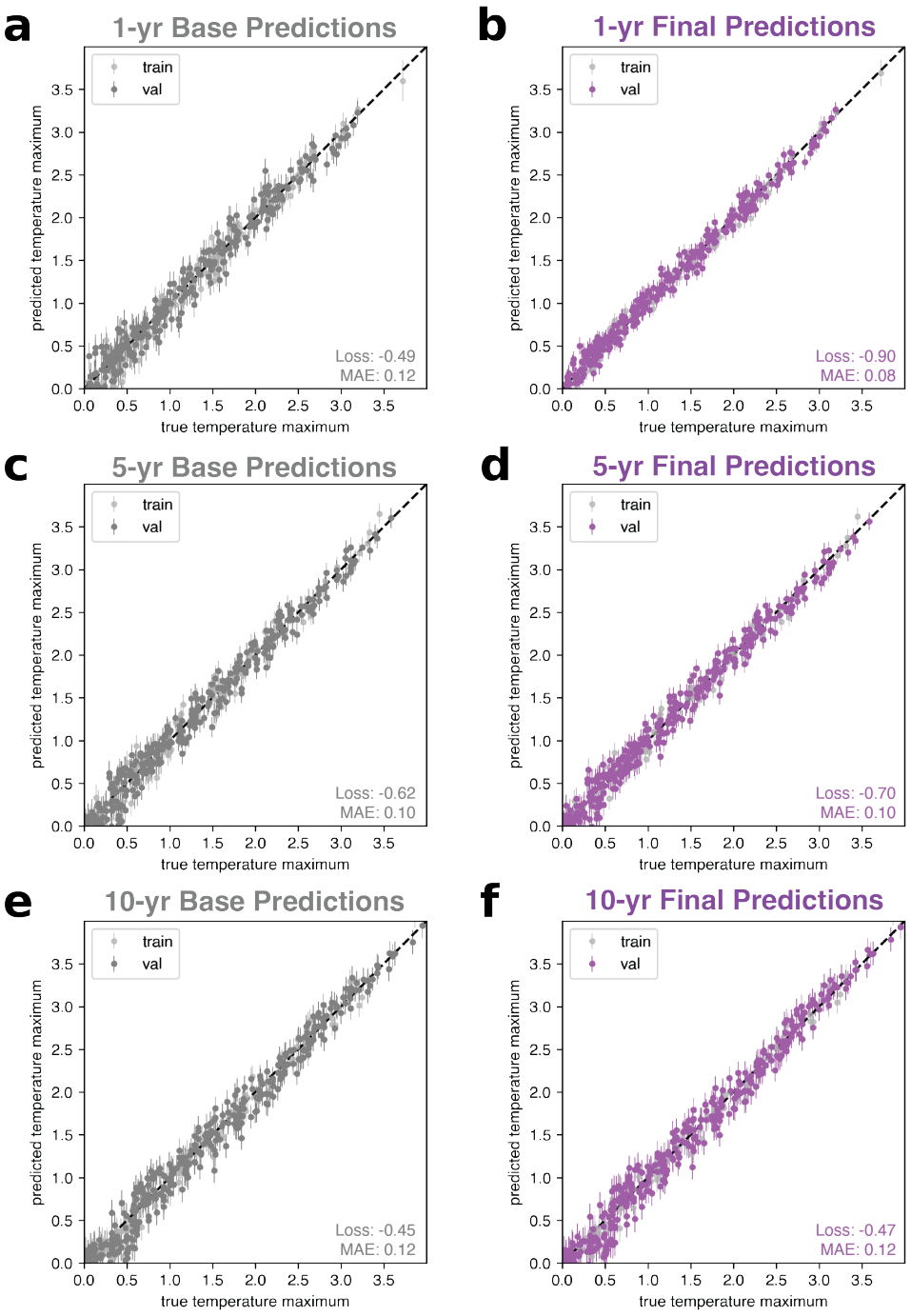

Table C.1: Prediction of the global-annual-mean maximum temperature over one- (a-b), five- (c-d), and ten- (e-f) year prediction windows for the base network (a,c,e), and the full network (b,d,f). The network performs similarly on the training and validation sets of climate models. The loss and mean absolute error of the central predictions for the validation set are presented at the bottom right of each frame.

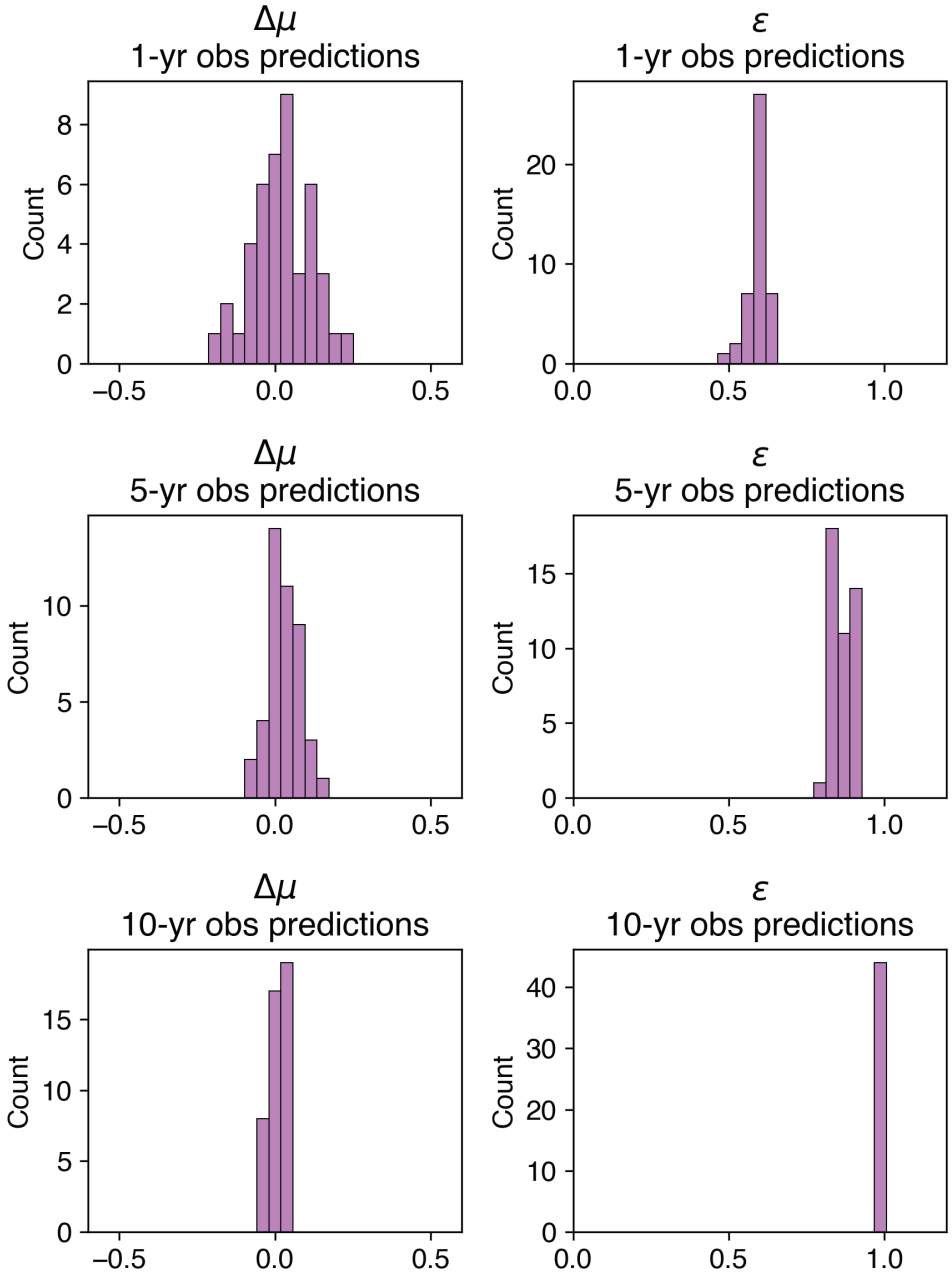

Table C.2: Updates to the base prediction by the full neural network on the testing set of GCM data. $\Delta\mu$ indicates whether the final prediction increases (>0) or decreases (<0) relative to the base prediction. ϵ indicates whether the uncertainty (σ) increases (>1) or decreases (<1) relative to the initial condition. In the one- and five-year predictions, the addition of the spatial temperature patterns to the model acts to lower uncertainty, in some cases halving . This is not the case for the ten-year predictions, wherein the model elects to stick with the standard deviation of the base prediction.

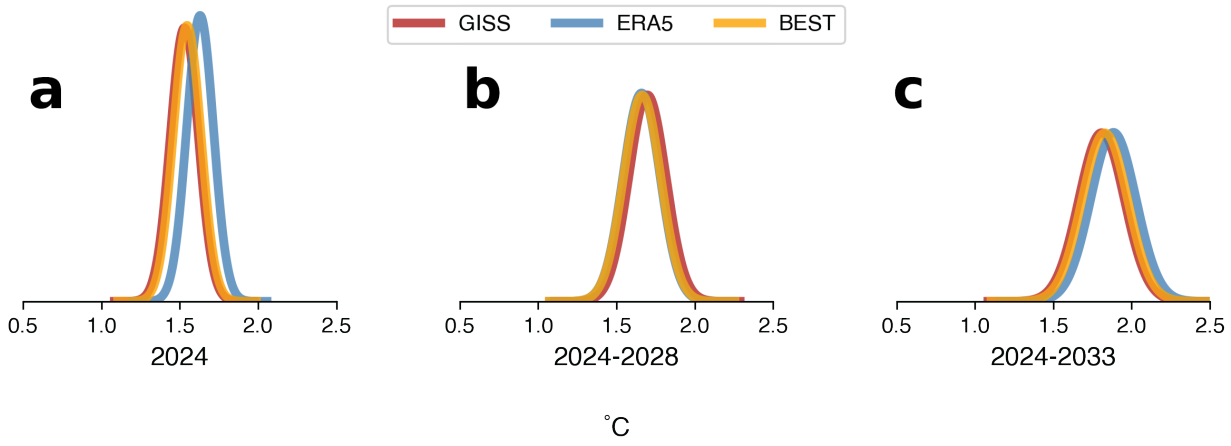

Table C.3: Neural networks initialized in 2023 with three observational data sets predicting for (a) 2024, (b) 2024-2028, (c) 2024-2033. Our results are robust across observational data sets.

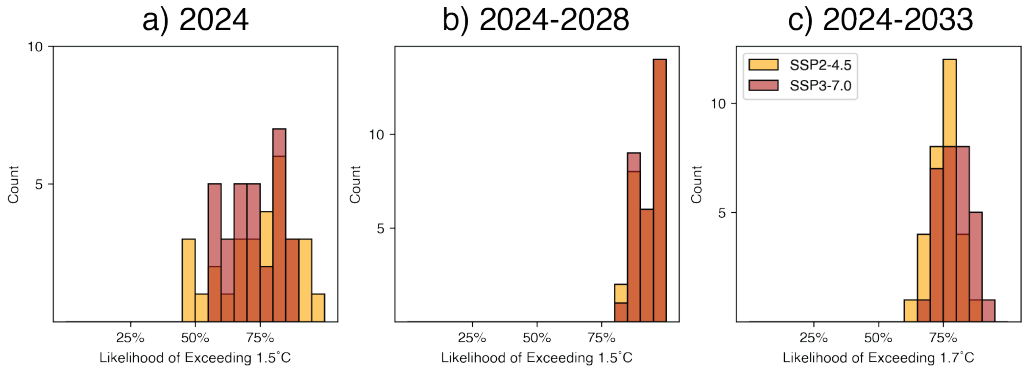

Table C.4: Likelihood of exceeding temperature thresholds across 30 different train/validation splits and initialization seeds for (a) 2024, (b) 2024-2028, (c) 2024-2033. The results for SSP2-4.5 are shown in orange and the results for SSP3-7.0 are shown in red. The similar distributions of predicted likelihoods for SSP2-4.5 and SSP3-7.0 confirm that our results are robust across these forcing scenarios. We use 23 climate models for SSP2-4.5, 15 for training and 8 for validation (Section C.1).

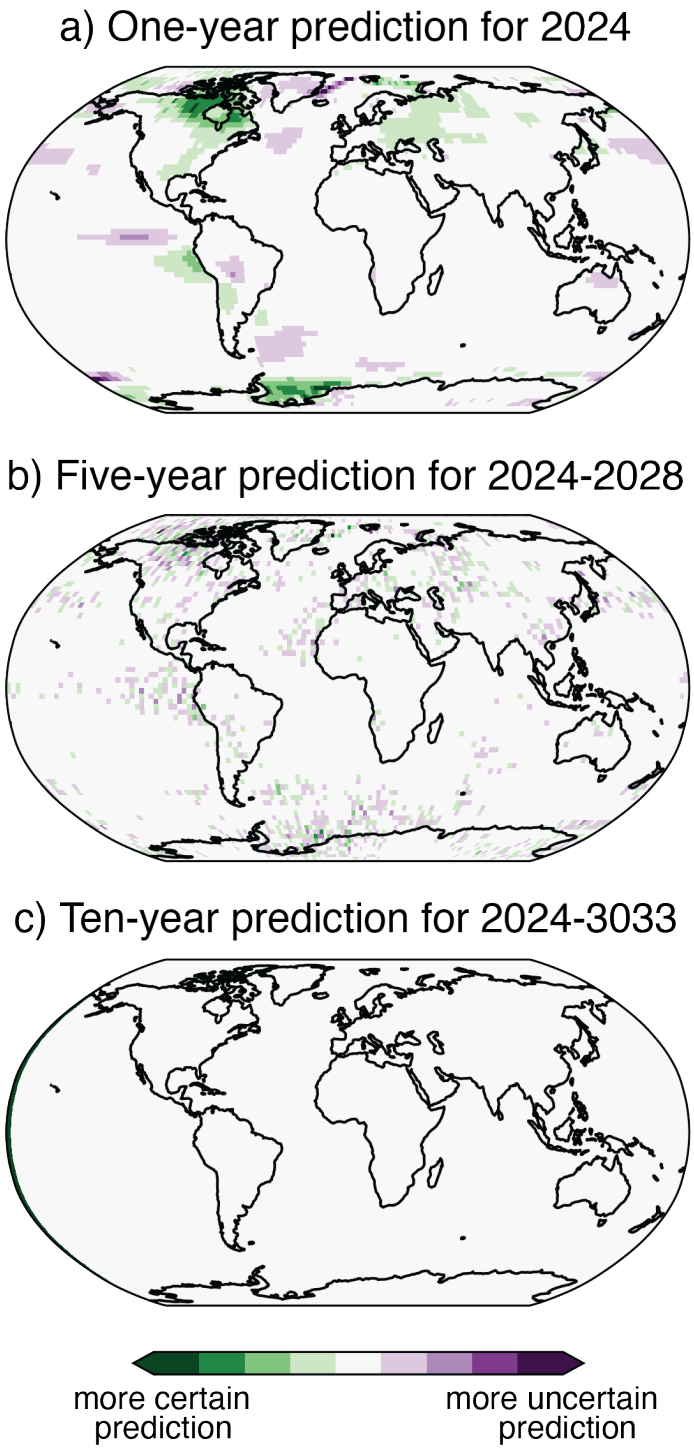

Table C.5: Attribution heatmaps for real-time forecasts of maximum global temperature over the next (a) one, (b) five, and (c) ten years. Reds indicate regions where the climate patterns contribute to a more certain prediction and blues indicate regions where the climate patterns contribute to a more uncertain prediction.

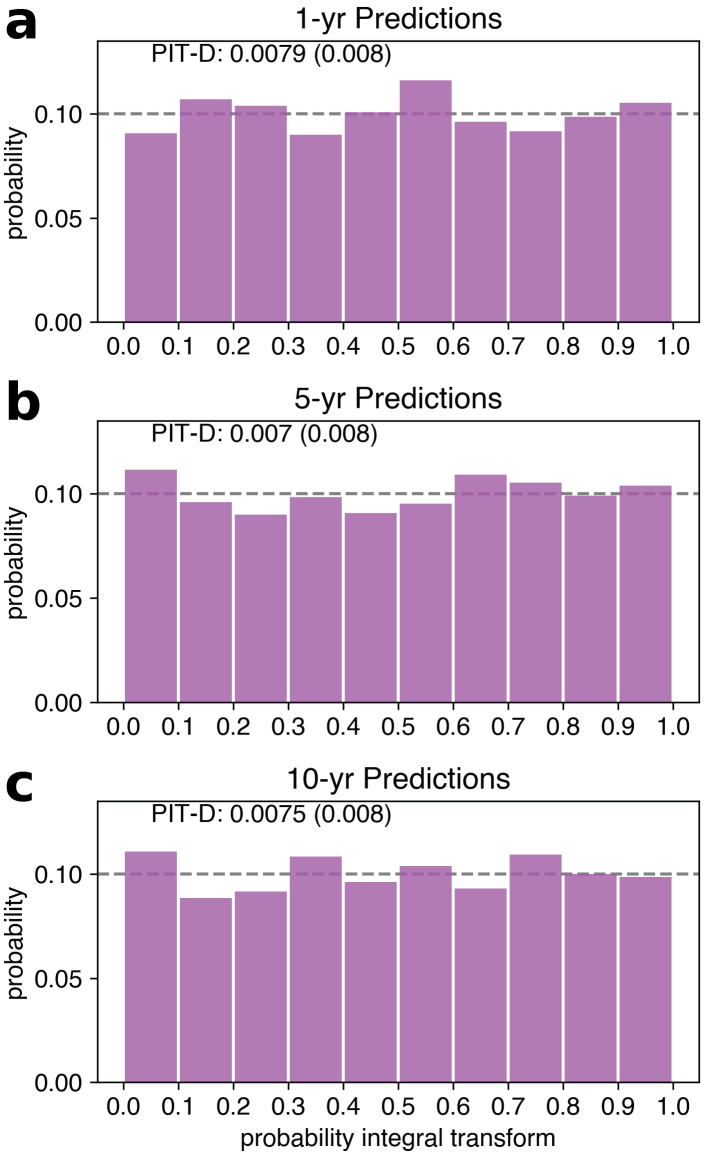

Table C.6: Probability integral transform (PIT) histogram on the validation data for the three chosen models for 1-, 5-, and 10-year predictions of maximum global near-surface temperature.

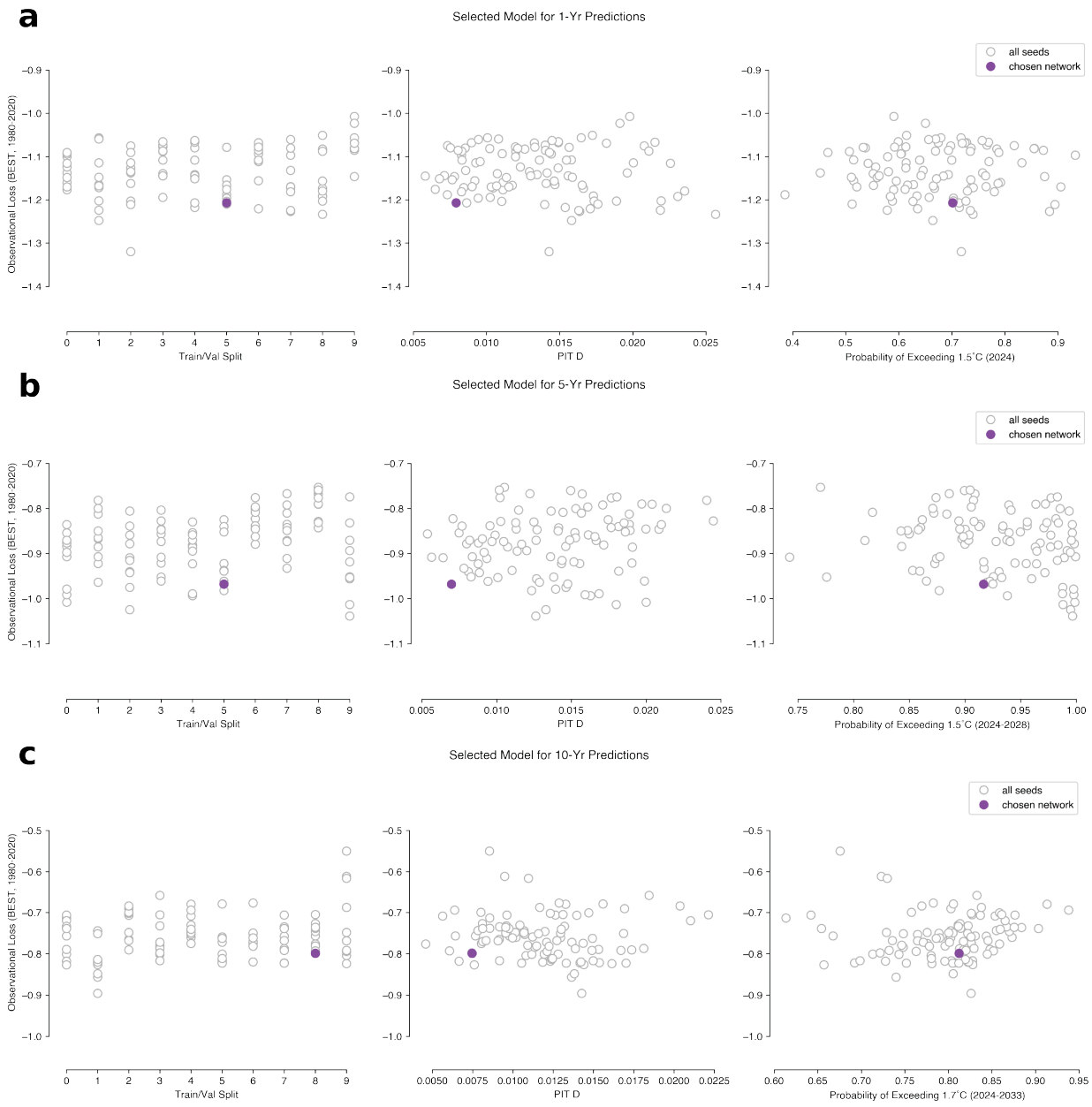

Table C.7: Loss on BEST (1980-2020) for networks trained on 10 different train/val splits and 10 different initialization seeds for (a) one-, (b) five-, and (c) ten-year predictions. The chosen network is shown in solid purple. (Left) The distribution of losses for each train/val split. (Center) The PIT D values on validation associated with each trained network. (Right) The probability of exceeding 1.5°C or 1.7°C.

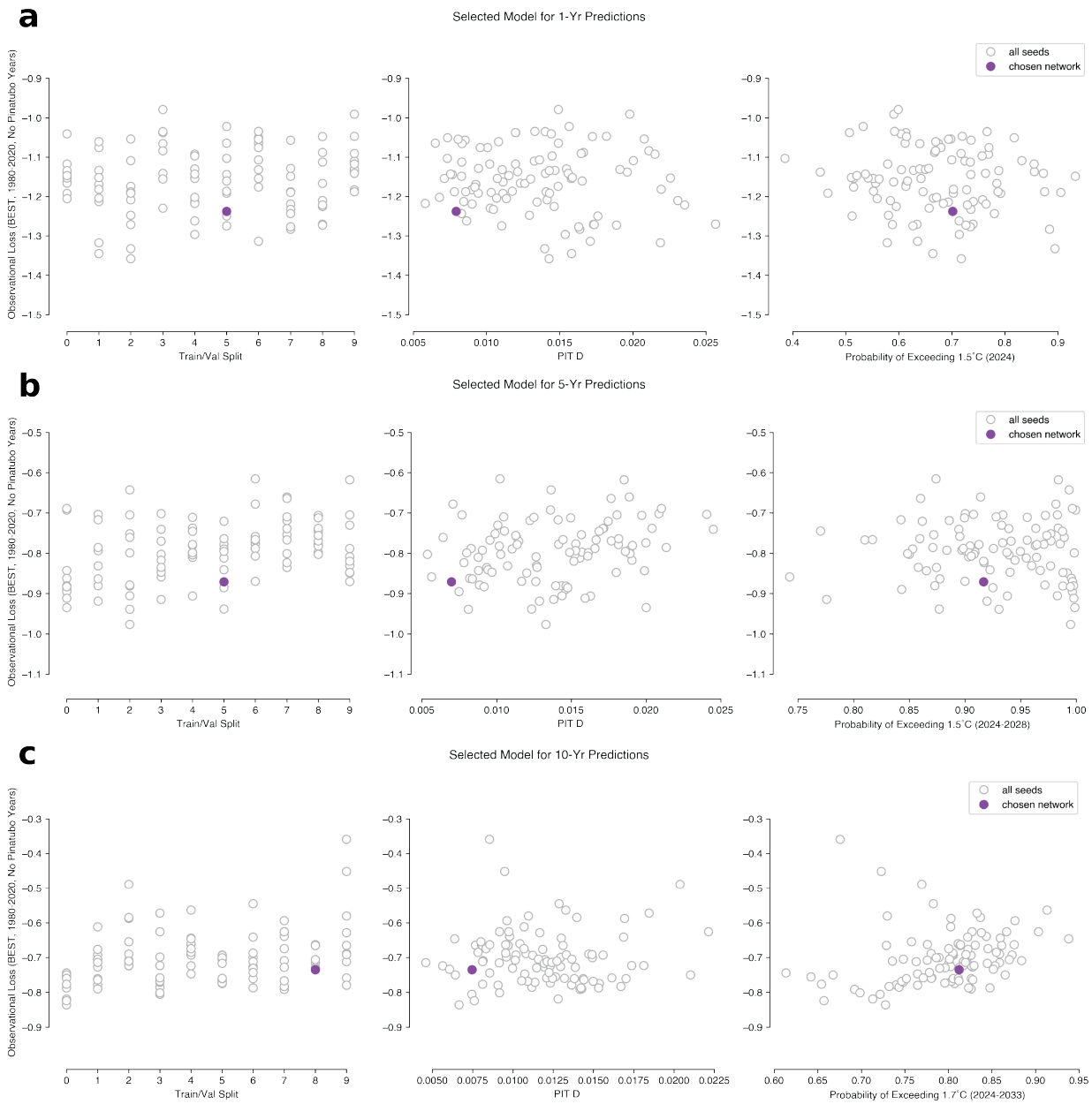

Table C.8: Same as Figure C.7, but the years impacted by the eruption of Mount Pinatubo (1991-1995) are excluded from the calculation of loss.

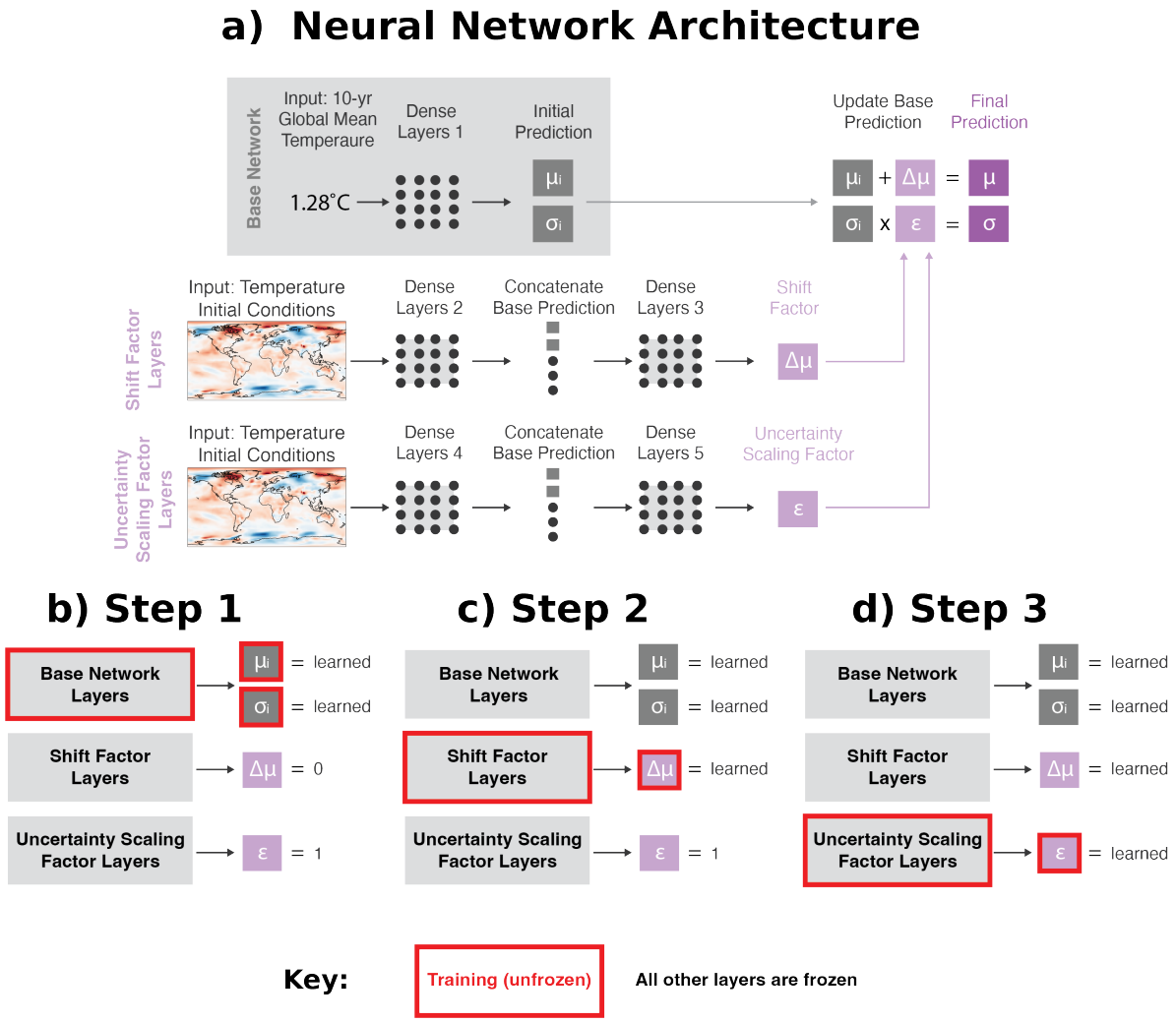

Table C.9: (a) Same as Figure 1a, but with the dense layers numbered for consistency with the hyperparameter choices in Table C.2. (b-d) The training process consists of three parts, where red indicates the layers that are unfrozen and thus the weights are updating during training, while all other layers are frozen. (b) The base network layers are trained while the shift factor and uncertainty scaling factor layers are frozen. The shift factor ($\Delta\mu$) is initialized at zero, and the uncertainty scaling factor (ϵ) is initialized at one, such that the final prediction (μ, σ) is equal to the initial prediction (μ_b, σ_b). (c) The shift factor layers are then trained, while the initial network and uncertainty scaling factor layers are frozen. During training, the neural network learns to modify the central prediction μ by modifying $\Delta\mu$, where $\mu = \mu_b + \Delta\mu$. (d) The uncertainty scaling factor layers are then trained, while the base network and shift factor layers are frozen. During training, the neural network learns to modify the uncertainty of the prediction σ by modifying ϵ, where $\sigma = \epsilon\sigma_b$.

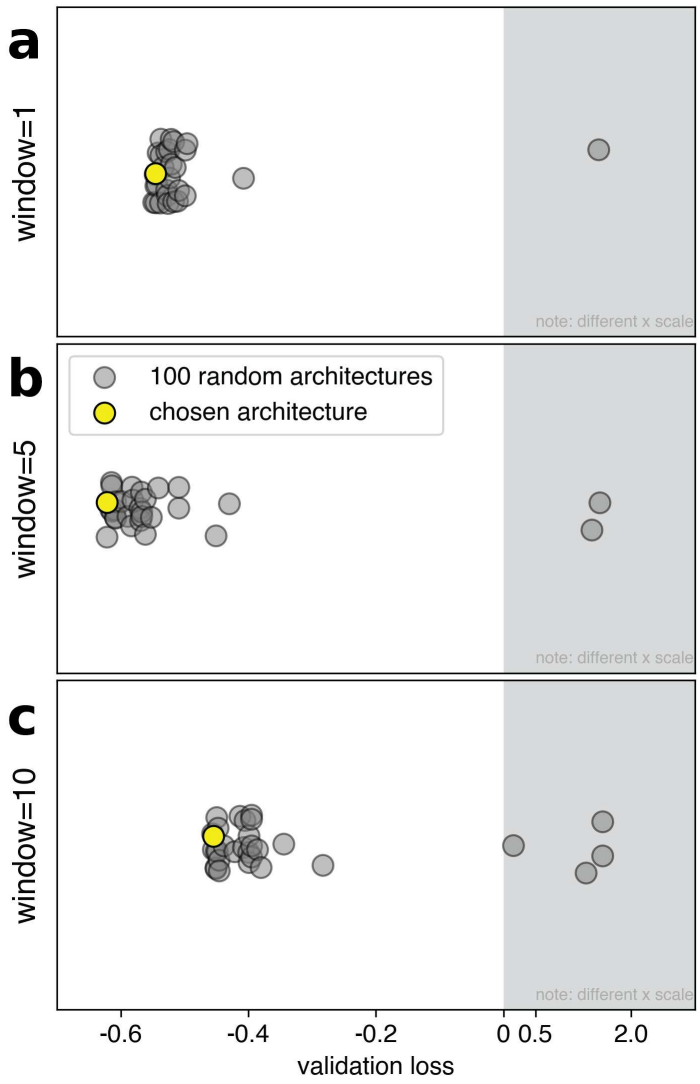

Table C.10: Validation loss for 100 different architectures for the base network for (a) one-, (b) five-, and (c) ten-year prediction windows. Each architecture is trained with the same initial weights and training/validation split. The chosen architecture, in yellow, is the same for all three prediction windows (1, 5, and 10 years). Note that the gray shaded region has a logarithmic scale. The chosen architecture, and the full hyperparameter tuning space, can be found in Table C.3.

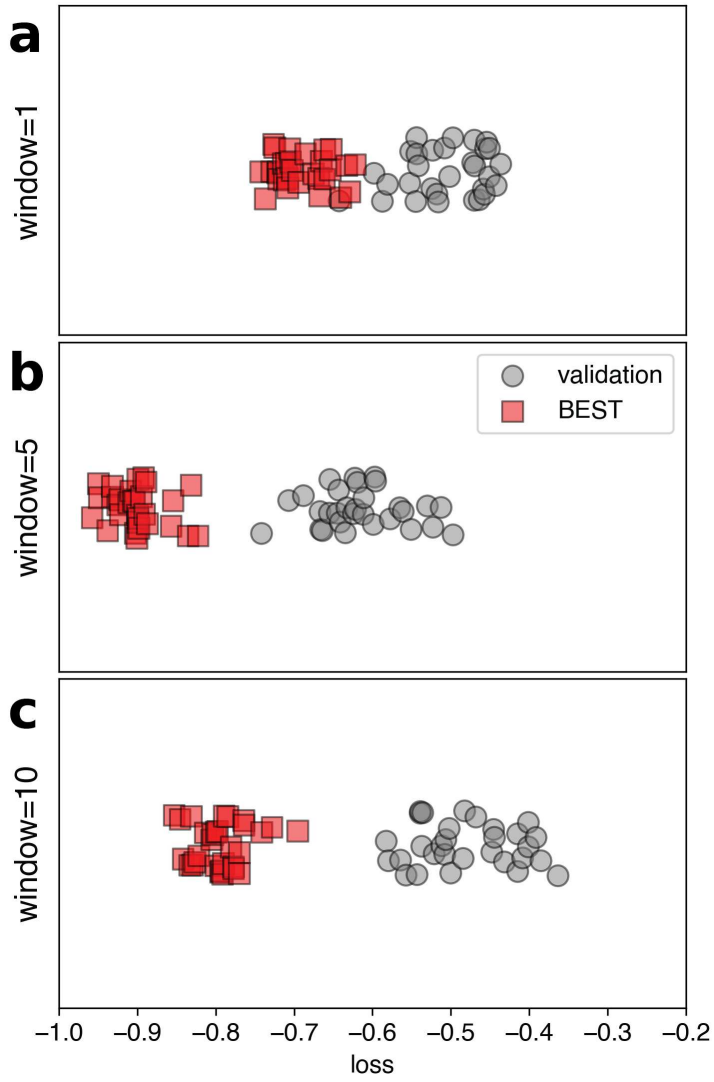

Table C.11: Validation and observations loss for 30 different train/validation splits for the base network predicting over (a) one-, (b) five-, and (c) ten-year windows. While the validation loss is sensitive to train/val split, the loss on observations (BEST; 1980-2022) is still high.

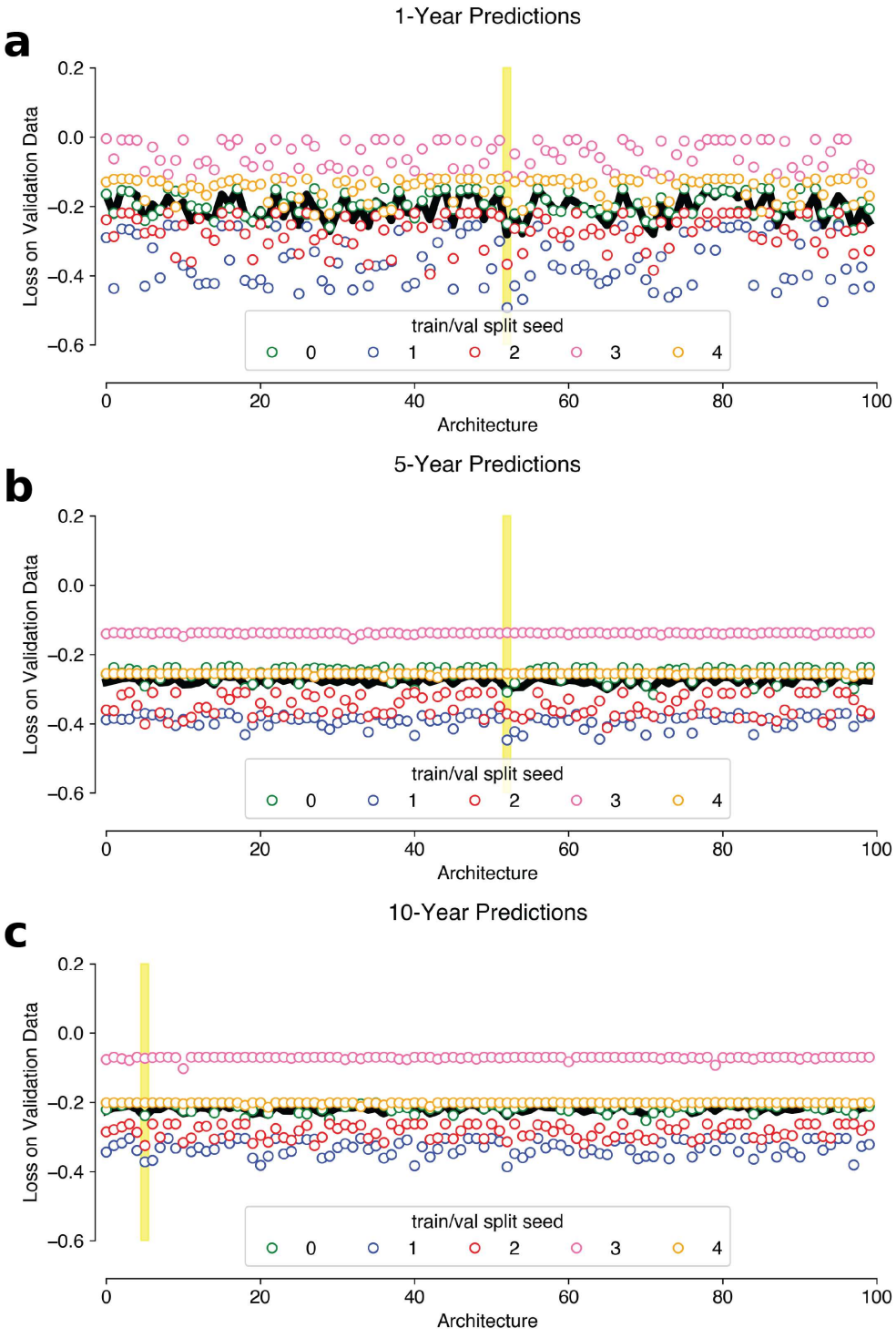

Table C.12: Validation loss for 100 different architectures, across 5 different seeds, for the final network. The 5 seeds result in different initial weights and train/val splits. The chosen architecture, highlighted in yellow, had the lowest mean validation loss across the 5 seeds. The chosen architectures, and the hyperparameter tuning space, can be found in Tables C.2 and C.3.

	2024		2024-2028		2024-2033	
	Final	Base	Final	Base	Final	Base
1.5˚C	*70%*	*26%*	*92%*	*86%*	*99%*	*99%*
1.7˚C	*4%*	*3%*	*38%*	*32%*	*81%*	*79%*
2.0˚C	*<1%*	*<1%*	*<1%*	*<1%*	*12%*	*11%*

Table C.1: Likelihoods of exceeding 1.5°C, 1.7°C, and 2.0°C over the next one, five, and ten years, from the final network and the base network. These likelihoods use the Berkeley Earth estimate that the global mean temperature in 2023 was 1.54°C higher than 1850-1900 temperatures.

	10 years	5 years	1 year
Batch Size	64	64	64
Max # of Epochs	25,000	25,000	25,000
Early Stopping Patience	50	50	50
Base Network Learning Rate	0.01	0.01	0.01
Final Network Learning Rate	0.0001	0.001	
Dense Layers 1 (# layers x # nodes and activation)	3x2, gelu	3x2, gelu	3x2, gelu
Dense Layers 2/4 (# layers x # nodes and activation)	2x100, relu	3x100, relu	3x100, relu
Dense Layers 3/5 (# layers x # nodes and activation)	3x5, relu	2x50, relu	2x50, relu
Ridge Parameter for first layer of Dense Layers 2/4	0.1	0.01	0.01
Train/Test Split Seed	8319	3770	3770
Initialization Seed	8	7	6

Table C.2: Hyperparameters for the neural networks used in the main text. Refer to Section C.1 and Figure C.9 for further description of the network architecture.

	Choices
Base Network Learning Rate	0.1, 0.01, 0.001
Final Network Learning Rate	0.01, 0.001, 0.0001
# layers for each set of Dense Layers	0, 1, 2, 3 (0 indicates a linear network)
# nodes per layer for each set of Dense Layers	1, 2, 5, 10, 20, 50, 100
Activation for Dense Layers 1	relu, elu, gelu, tanh
Activation for Dense Layers 2-5	relu, elu, tanh
Ridge Parameter for first layer of Dense Layers 2/4	0.1, 0.01, 0.001, 0

Table C.3: Hyperparameter search space for the neural networks used in the main text. Refer to Section C.1, Section C.2 and Figure C.9 for further description of the network architecture and tuning procedure.