

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

Jamin Kurtis Rader¹ and Elizabeth A. Barnes¹

¹Colorado State University

June 23, 2023

Abstract

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 within a perfect model framework 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.

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

Jamin K. Rader¹, and Elizabeth A. Barnes¹

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

Key Points:

- An interpretable neural network provides a spatially-weighted mask for selecting optimal analogs
- Analog selected with the weighted mask offer more skillful forecasts than traditional methods for selecting analogs
- The learned mask highlights precursor regions for predicting large-scale climate anomalies in a perfect model framework

Corresponding author: Jamin K. Rader, jamin.rader@colostate.edu

Abstract

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 within a perfect model framework 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.

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.

1 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 (Kushnir et al., 2019; Merryfield et al., 2020; Towler et al., 2018). 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

43 will evolve in similar ways (Lorenz, 1969). It follows that analogs—similar looking states
44 to the initial state that is being forecast—can provide insight into how that initial state
45 will continue to evolve. The analog forecasting approach is powerful for seasonal-to-decadal
46 climate prediction (e.g., Ding et al., 2018, 2019; Menary et al., 2021; Delle Monache et
47 al., 2013; Zhang et al., 2023) and can outperform general circulation models (GCMs) ini-
48 tialized with observations, which struggle with initialization shock and climate model
49 drift (Merryfield et al., 2020; Mulholland et al., 2015).

50 A major hurdle in obtaining successful analog forecasts is that the climate system
51 is noisy and chaotic, and thus small differences between two initial states can result in
52 vast differences in their evolution (Lorenz, 1963). Thus, a successful analog forecast for
53 a particular initial climate state, which we refer to as the state of interest (SOI), requires
54 that the analogs and SOI are sufficiently similar such that their evolutions do not sig-
55 nificantly diverge during the prediction timeframe. Sufficiently similar analogs can be
56 difficult to find in the observational record since the number of independent observations
57 we have on seasonal-to-decadal scales (e.g., fewer than 100 during the satellite era) is so
58 much smaller than the number of degrees of freedom within a global geophysical field
59 (e.g., Van den Dool, 1994). While observations are in short supply, there is a wealth of
60 simulated climate data and many recent studies have used “model-analogs” (Ding et al.,
61 2018) drawn from climate model output instead (e.g., Lou et al., 2023; Peng et al., 2021;
62 Wu & Yan, 2023).

63 We refer to the library of climate model states that can be used for analog fore-
64 casting as “potential analogs.” Once a potential analog has been identified to be suffi-
65 ciently similar to the SOI we refer to it as an analog. Forecasts are made by taking the
66 mean evolution of the top-N analogs, where N is chosen by the user. There are several
67 ways to quantify the similarity between the potential analogs and the SOI. The most straight-
68 forward method is to compute the global correlation between each potential analog and
69 the SOI (e.g., Mahmood et al., 2022). Using a global correlation assumes that the sim-
70 ilarity between the maps at each grid point globally matters equally. A natural next step
71 in complexity is to compute a correlation over a region that is known to be important
72 for predictability of a given target, such as the North Pacific for predicting the Pacific
73 Decadal Oscillation (e.g., Wu & Yan, 2023). While this approach removes some regions
74 that may not be useful for determining the best analogs, it still assumes that each grid
75 point within the region is equally important and the region must be known *a priori*.

76 In the following work, we train an interpretable neural network on a proxy task that
77 is similar to the analog problem (Section 3). The network learns a weighted mask which
78 is used for determining analogs. The forecasting skill of the analogs selected using the
79 learned weighted mask is tested through a perfect model approach where climate model
80 data substitutes observations and is used to predict future climate model data. We show
81 in two examples, forecasting 5-year sea surface temperature (SST) anomalies in the North
82 Atlantic (Section 4) and wintertime SST anomalies in the tropical Pacific (i.e. El Niño
83 Southern Oscillation; Section 5), that analogs identified using the weighted mask pro-
84 vide more skillful forecasts than analogs that are identified in a way that is globally or
85 regionally uniform. In addition, we show that these masks, once generated by a neural
86 network, can be modified *post hoc* to further investigate the importance of each region
87 for seasonal-to-decadal prediction (Section 5).

88 2 Data and Metrics

89 2.1 Climate Model Data

90 We use monthly SST from the historical run of the Max Planck Institute (MPI)
91 for Meteorology Grand Ensemble (GE; Maher et al., 2019) at 2° latitude by 2° lon-
92 gitude resolution. This dataset contains 100 members and each simulates 156 years (1850-
93 2005) of the Earth’s climate with historical forcing. The MPI-GE uses the MPI Earth
94 System Model version 1.1 (ESM1.1; Giorgetta et al., 2013). Each member is initialized
95 using a different year of the preindustrial control simulation such that the differences be-
96 tween ensemble members are a product of internal variability.

97 2.2 Standardization and Selection

98 Subsets of the MPI-GE ensemble members are used for different purposes. Our li-
99 brary of potential analogs is made up of members 1-35. Members 36-50 are the SOIs for
100 training the neural network, members 51-55 are the SOIs for the early stopping valida-
101 tion set (which is used to prevent overfitting to the training data), and members 56-60
102 are the SOIs for the tuning validation set (which is used to identify optimal hyperpa-
103 rameters for the neural network). Finally, members 96-100, which are withheld until the
104 very end, are the test set for making and evaluating the analog forecasts. Details on the

105 process of tuning and training the neural network, including selecting the hyperparam-
 106 eters, can be found in Section S1.

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

112 We removed the forced signal from the climate model data by subtracting the en-
 113 semble mean of the library of potential analogs at each location and year from each set
 114 of data. After the forced signal was removed, the data was standardized by dividing by
 115 the standard deviation at each grid point across the library of potential analogs. By us-
 116 ing the library of potential analogs to calculate the forced signal and internal variance
 117 we treat the SOIs as if they are truly unseen data as we would when forecasting.

118 2.3 Metrics

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

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

122 such that a perfect prediction has a score of one, and a climatology prediction has a score
 123 of zero. Climatology is the prediction by the mean state, which is zero for this standard-
 124 ized data. Analog forecasts made using the weighted mask are compared with the fol-
 125 lowing additional baselines: a global analog forecast, a target region analog forecast, a
 126 mean target evolution forecast, and a random forecast. In the global analog forecast (tar-
 127 get region analog forecast), the analogs are selected if the unweighted MSE over the en-
 128 tire globe (target region) is the smallest. The mean target evolution forecast is based on
 129 how the targets in the input period evolve on average and is detailed in Section S2. The
 130 random forecast is made by randomly selecting targets from the library of potential analogs
 131 and using them as the prediction.

3 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 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 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 4 and 5.

4 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 (Jackson et al., 2022) and exhibit potential for skillful decadal forecasts (Hawkins et al., 2011; Sutton & Allen, 1997). SST variability in the North Atlantic has been as-

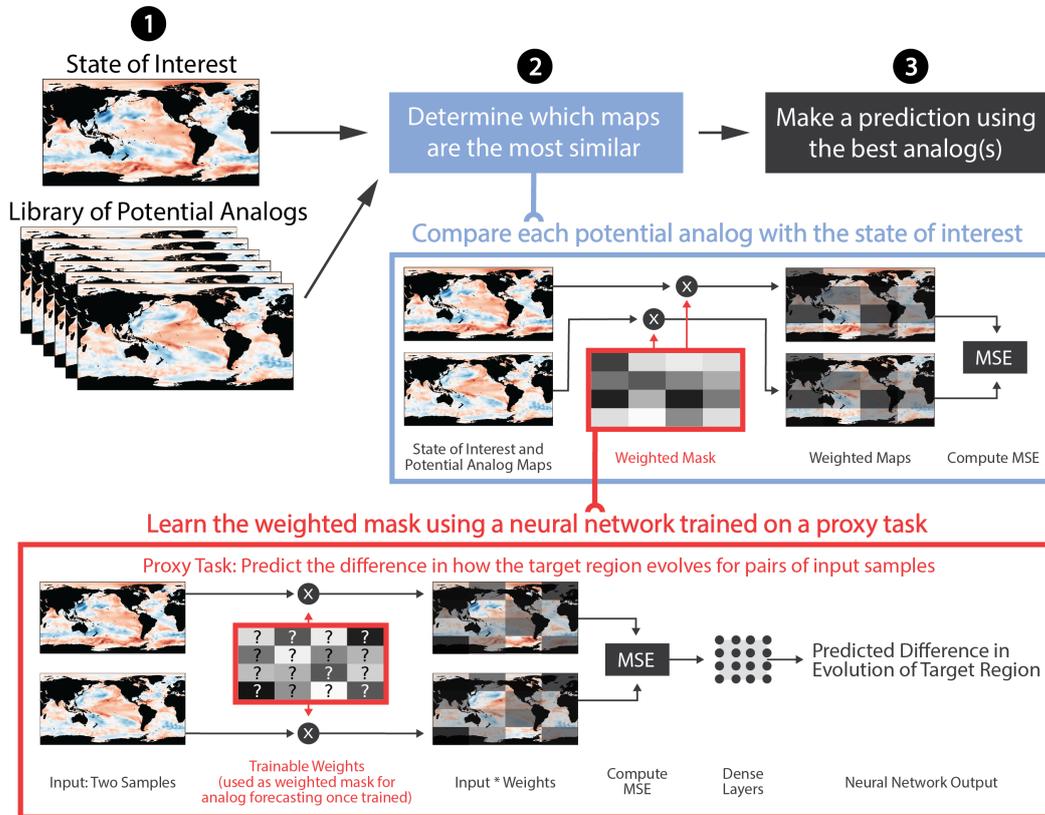


Figure 1. Optimized analog forecasting method and interpretable neural network architecture. The analog forecasting method can be described in three steps: 1) identify a state of interest and a library of potential analogs. 2) Determine which maps are the most similar. 3) Make a prediction using the 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.

163 sociated with weather and climate anomalies globally, including Atlantic hurricane fre-
164 quency and intensity (Goldenberg et al., 2001; Balaguru et al., 2018), northern hemisphere
165 precipitation (Enfield et al., 2001; Si et al., 2023), and the strength of the Asian sum-
166 mer monsoon (Shekhar et al., 2022). In this prediction problem, we use global maps of
167 SST, averaged over the previous five years, to predict the mean SST anomaly in the North
168 Atlantic (40° - 60° N, 10° - 70° W) over the following five years.

169 The weighted mask learned by the neural network is shown in Figure 2a. The Green-
170 land Sea and the gulf stream region in the western North Atlantic emerge as the most
171 important regions for identifying analogs in the MPI-GE. Over the western North At-
172 lantic, there is an area of zero weight between two areas of high weight. These may be
173 where the boundaries of persistent SST anomalies vary, and the neural network has learned
174 that the specific locations of these boundaries are not important for the prediction prob-
175 lem. Previous studies that have used an analog approach to assess North Atlantic decadal
176 predictability selected the best analogs by taking a correlation over the whole globe (Mahmood
177 et al., 2022) or the entire North Atlantic basin (Menary et al., 2021). As shown in Fig-
178 ure 2b-d, when using the weighted mask, the best analogs only have to look like the SOI
179 in the highest weight regions. An example SOI is shown in Figure 2b and its best ana-
180 log in Figure 2c. These two maps look similar in the North Atlantic, but are starkly dif-
181 ferent in the North Pacific and Indian Ocean, among other regions. Once the weighted
182 mask has been applied to the SOI (Figure 2d) and its best analog (Figure 2e), the maps
183 look nearly identical.

184 These results suggest that using uniform weights across the entire North Atlantic
185 basin, or the whole globe, may lead to a selection of analogs that are not optimized for
186 forecasting multi-year variability in the North Atlantic. Indeed, we see that this is true
187 in the skill scores shown in Figure 3a. For $1 \leq N \leq 50$, where the top-N analogs are av-
188 eraged, our weighted mask analog forecast outperforms the global and target region ana-
189 log forecasts, as well as the climatology, mean target evolution, and random baselines.
190 The skill score is lowest when only the single best analog is used for forecasting, and sub-
191 sequently improves for larger N. Given that the skill score maximizes around $N = 10$,
192 and the spread of the targets associated with the analogs (i.e. the uncertainty of the fore-
193 cast) increases with N (Figure S1), we elect to focus on results for $N = 10$ analogs. The
194 prediction by the top-10 analogs, and the spread of the targets, are shown in Figure 3b
195 for 200 years of SOIs. The analog predictions do a good job of capturing the variabil-

196 ity of North Atlantic sea surface temperatures, though they do struggle to forecast the
197 most extreme anomalies.

198 **5 Seasonal Prediction of El Niño Southern Oscillation**

199 In addition to improving multi-year forecasts of SST in the North Atlantic, the learned
200 weighted mask improves forecasts of ENSO on seasonal timescales. ENSO is the lead-
201 ing mode of global annual SST variability (Hsiung & Newell, 1983) and has an exten-
202 sive influence on global weather and climate (reviewed in Yeh et al., 2018). Analog fore-
203 casting has been applied to seasonal prediction of ENSO in several studies due to its po-
204 tential to outperform initialized GCM forecasts (e.g., Ding et al., 2018, 2019). In the fol-
205 lowing example, we use wintertime (November-March) global SST anomalies to forecast
206 SST anomalies in the Niño3.4 region (5°S-5°N, 120-170°W; Barnston et al., 1997; Han-
207 ley et al., 2003) the following winter.

208 The weighted mask for forecasting ENSO looks markedly different from that for
209 forecasting North Atlantic multi-year variability (Figure 4a). While a few regions are as-
210 signed higher weights, the weights in Figure 4a are much more uniform across the globe
211 than in Figure 2a. The four main regions that stand out in this weighted mask have also
212 been identified as important precursors in previous literature: the western North Pacific
213 (e.g., S.-Y. Wang et al., 2012), the Pacific Meridional Mode (e.g., Amaya, 2019), the Cen-
214 tral Atlantic (e.g., Martín-Rey et al., 2015), and the tropical Pacific itself (e.g., Capot-
215 tondi & Sardeshmukh, 2015). The skill score of the global analog forecast (Figure 4b)
216 is similar to that of our weighted mask analog forecast (but always lower, see Figure S3),
217 which is not surprising since the values of the weighted mask are near one for most ar-
218 eas of the globe.

219 Since the weighted mask can be manually updated *post hoc*, we use this to explore
220 the sensitivity of the forecast skill to which regions are included in the weighted mask.
221 Figure 5a shows the weighted mask for ENSO prediction (Figure 4a) but where the small-
222 est 95 percent of the weights have been set to zero. Forecasts made with this “constrained”
223 weighted mask have similar skill to the original weighted mask (as shown in Figure S4).
224 From the constrained weighted mask, we identify four main precursor regions for ENSO:
225 the West Pacific (ocean grid points bounded by 0°-40°N, 100°-170°E), the Tropical Pa-

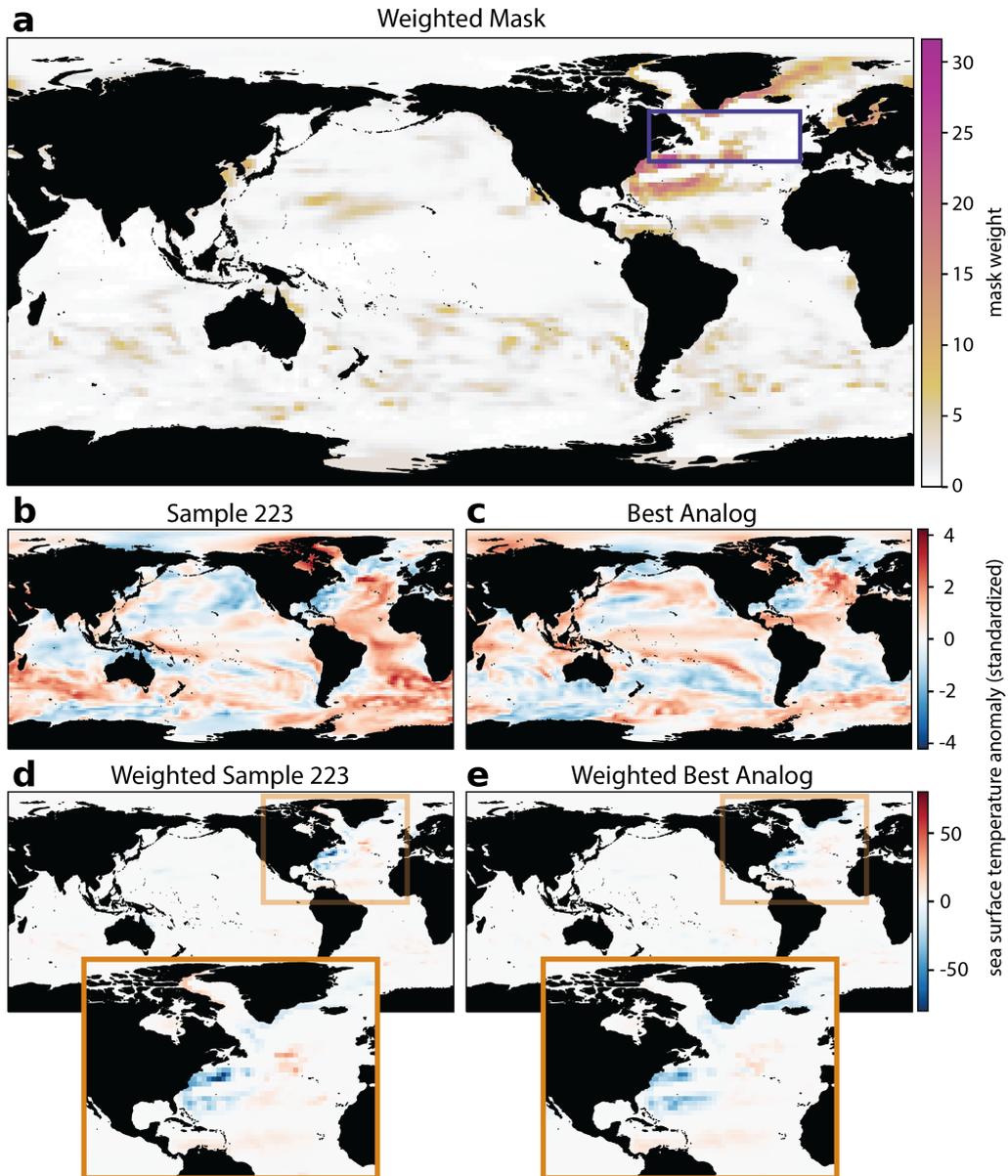


Figure 2. Weighted mask and example for multi-year predictions of North Atlantic SST. (a) Weighted mask, as learned by the interpretable neural network. (b) Standardized SST anomalies for a sample state of interest (SOI). (c) Standardized SST anomalies for the best analog associated with the SOI. (d) Weighted SOI. (e) Weighted best analog.

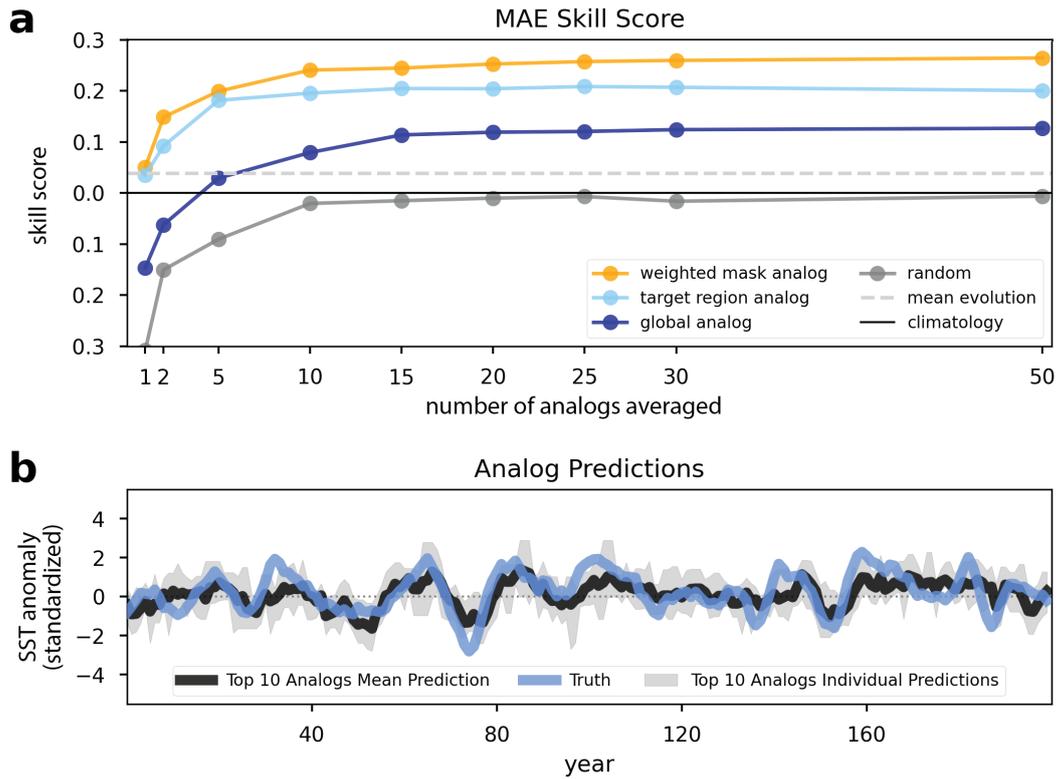


Figure 3. Analog forecasts of North Atlantic sea surface temperature. (a) Skill scores for our weighted mask analog forecast and other baselines. (b) Weighted mask analog forecasts for 200 years of MPI-GE simulations, including the mean prediction from the top-10 analogs, the spread of these predictions, and the truth values.

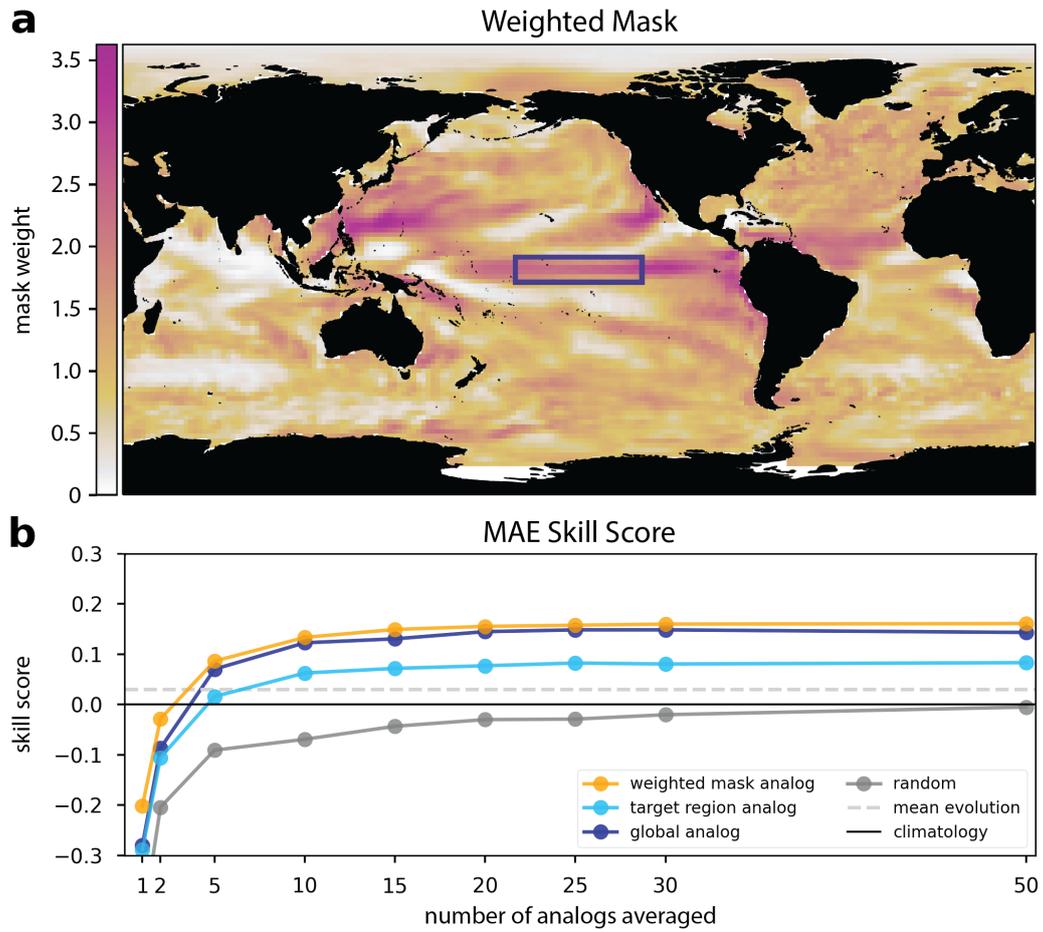


Figure 4. Weighted mask and skill scores for seasonal predictions of El Niño Southern Oscillation. (a) Weighted mask. (b) Skill scores for our weighted mask analog and other baselines.

226 cific (25°S-10°N, 170°E-65°W), the Baja Coast (10°N-40°N, 110°-140°W), and the Trop-
227 ical Atlantic (0°-20°N, 20°-80°W).

228 We assess how important each precursor region is in two ways. In the first approach,
229 we test the skill score of analog forecasting when each region is occluded from the con-
230 strained weighted mask (weights in that region are set to zero). When all four regions
231 are included, the skill score is 0.146. Removing any of the four regions from the weighted
232 mask results in a skill score decrease. Interestingly, removing the Tropical Atlantic re-
233 sults in the most drastic decrease in prediction skill. While the Tropical Atlantic has been
234 connected to ENSO predictability (e.g., Martín-Rey et al., 2015), it is not considered a
235 primary driver (C. Wang, 2018). In the second approach, we isolate each of the four re-
236 gions (weights outside that region are set to zero). There is no improvement over clima-
237 tology when just the Baja Coast or Tropical Atlantic is used to select analogs, and more
238 skill when just the West Pacific or Tropical Pacific is used. However, no region alone pro-
239 vides anywhere near the skill that all four regions do together.

240 6 Discussion and Conclusions

241 We have shown how an interpretable neural network can be used to identify a weighted
242 mask that improves the selection of analogs for seasonal-to-decadal forecasting. The pre-
243 cursors identified in the weighted masks are not necessarily causal, but they do provide
244 the optimal predictors for the given input. In this work we have constrained the neu-
245 ral network to learn one mask that represents all pathways of predictability, however al-
246 lowing the network to learn different masks for different SOIs could lead to better ana-
247 log forecasts.

248 While we only used a single input map of SST to predict a future target SST in
249 this work, this neural network architecture can be used for many other forecasting ap-
250 proaches. For example, a combination of multiple variables can be used as the predic-
251 tors (such as SST and sea surface height, as in Ding et al., 2018) or other geophysical
252 variables may be selected as the target (e.g., predicting precipitation over land). In ad-
253 dition, one may also include variables at multiple lead times to capture the time tendency
254 of the climate system. We show results that include SST tendency as an input for the
255 North Atlantic multi-year prediction example in Figure S5.

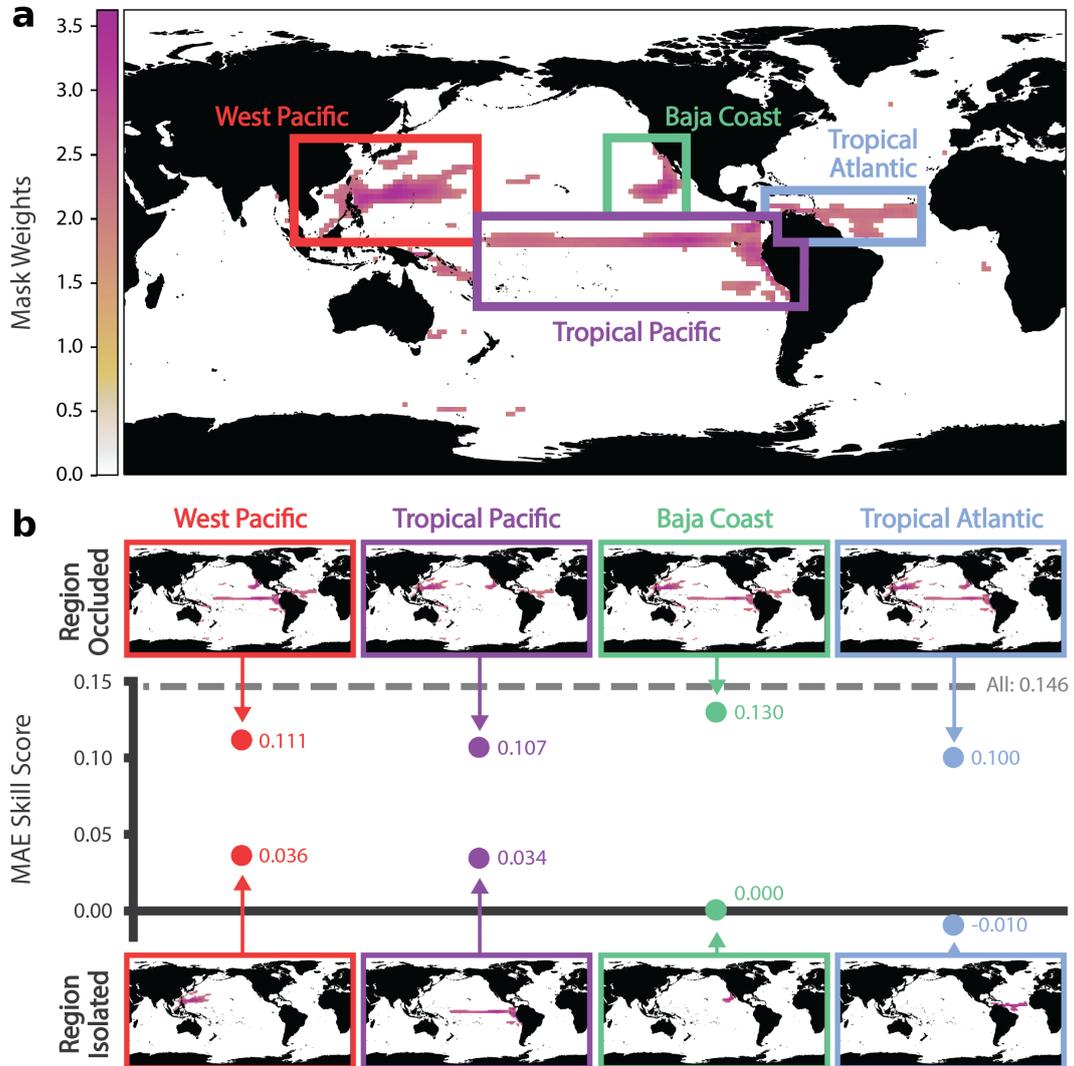


Figure 5. Analog forecasting skill of El Niño Southern Oscillation when various regions are occluded or isolated. (a) As in Figure 4a, but the lowest 95 percent of weights are set to zero. Four regions of focus are highlighted by the colored boxes. (b) Skill scores for analog forecasts when each region is occluded from the mask (top) and when the region is isolated to make a forecast (bottom).

256 We have explored this method through a perfect model setup. As such, the iden-
257 tified precursors are intrinsic to MPI-ESM1.1 and may not reflect patterns of predictabil-
258 ity in the observed Earth system. Training the weighted mask on a multi-model ensem-
259 ble may provide patterns that are more consistent with observations (e.g. Kirtman et
260 al., 2014; Rader et al., 2022) and allow for enhanced analog predictions on real data. Ad-
261 ditionally, we could train on models and observations at the same time to identify a weighted
262 mask that is more representative of the true Earth System. We believe that this weighted
263 mask approach will be influential to analog forecasting moving forward.

264 Open Research Section

265 The data used in this study, simulations from the Max Planck Institute for Mete-
266 orology Grand Ensemble, are publicly available at [https://esgf-data.dkrz.de/projects/mpi-](https://esgf-data.dkrz.de/projects/mpi-ge/)
267 [ge/](https://esgf-data.dkrz.de/projects/mpi-ge/). The weighted masks, and all python code used to generate the data and figures in
268 this paper, can be found at XXX REVIEWERS, THIS CAN CURRENTLY BE FOUND
269 AT <https://github.com/jaminrader/WeightedMaskAnalogForecasting> AND WILL BE
270 UPLOADED TO ZENODO WHEN FINISHED XXX.

271 Acknowledgments

272 This material is based upon work supported by the U.S. Department of Energy,
273 Office of Science, Office of Advanced Scientific Computing Research, Department of En-
274 ergy Computational Science Graduate Fellowship under Award Number DE-SC0020347,
275 and by grant AGS-2210068 from the National Science Foundation. The authors would
276 like to thank colleagues at the NOAA Physical Sciences Laboratory for their helpful feed-
277 back.

278 References

- 279 Amaya, D. J. (2019, December). The pacific meridional mode and ENSO: a re-
280 view. *Current Climate Change Reports*, 5(4), 296–307. doi: 10.1007/s40641
281 -019-00142-x
- 282 Balaguru, K., Foltz, G. R., & Leung, L. R. (2018, May). Increasing magnitude of
283 hurricane rapid intensification in the central and eastern tropical atlantic. *Geo-*
284 *phys. Res. Lett.*, 45(9), 4238–4247. doi: 10.1029/2018gl077597
- 285 Barnston, A. G., Chelliah, M., & Goldenberg, S. B. (1997, September). Doc-

- 286 umentation of a highly ENSO-related sst region in the equatorial pa-
287 cific: Research note. *Atmosphere-Ocean*, *35*(3), 367–383. doi: 10.1080/
288 07055900.1997.9649597
- 289 Capotondi, A., & Sardeshmukh, P. D. (2015, November). Optimal precursors of dif-
290 ferent types of ENSO events. *Geophys. Res. Lett.*, *42*(22), 9952–9960. doi: 10
291 .1002/2015gl066171
- 292 Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B., & Searight, K. (2013,
293 October). Probabilistic weather prediction with an analog ensemble. *Mon.*
294 *Weather Rev.*, *141*(10), 3498–3516. doi: 10.1175/mwr-d-12-00281.1
- 295 Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2018, July). Skill-
296 ful climate forecasts of the tropical Indo-Pacific ocean using Model-Analogs. *J.*
297 *Clim.*, *31*(14), 5437–5459. doi: 10.1175/JCLI-D-17-0661.1
- 298 Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2019, February).
299 Diagnosing secular variations in retrospective ENSO seasonal forecast skill
300 using CMIP5 model-analogs. *Geophys. Res. Lett.*, *46*(3), 1721–1730. doi:
301 10.1029/2018gl080598
- 302 Enfield, D. B., Mestas-Nuñez, A. M., & Trimble, P. J. (2001, May). The at-
303 lantic multidecadal oscillation and its relation to rainfall and river flows
304 in the continental U.S. *Geophys. Res. Lett.*, *28*(10), 2077–2080. doi:
305 10.1029/2000gl012745
- 306 Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M.,
307 ... Stevens, B. (2013, July). Climate and carbon cycle changes from 1850 to
308 2100 in MPI-ESM simulations for the coupled model intercomparison project
309 phase 5. *J. Adv. Model. Earth Syst.*, *5*(3), 572–597. doi: 10.1002/jame.20038
- 310 Goldenberg, S. B., Landsea, C. W., Mestas-Nunez, A. M., & Gray, W. M. (2001,
311 July). The recent increase in atlantic hurricane activity: causes and implica-
312 tions. *Science*, *293*(5529), 474–479. doi: 10.1126/science.1060040
- 313 Hanley, D. E., Bourassa, M. A., O’Brien, J. J., Smith, S. R., & Spade, E. R. (2003,
314 April). A quantitative evaluation of ENSO indices. *J. Clim.*, *16*(8), 1249–1258.
315 doi: 10.1175/1520-0442(2003)16(1249:AQEOEI)2.0.CO;2
- 316 Hawkins, E., Robson, J., Sutton, R., Smith, D., & Keenlyside, N. (2011, December).
317 Evaluating the potential for statistical decadal predictions of sea surface tem-
318 peratures with a perfect model approach. *Clim. Dyn.*, *37*(11-12), 2495–2509.

- 319 doi: 10.1007/s00382-011-1023-3
- 320 Hsiung, J., & Newell, R. E. (1983, October). The principal nonseasonal modes of
321 variation of global sea surface temperature. *J. Phys. Oceanogr.*, *13*(10), 1957–
322 1967. doi: 10.1175/1520-0485(1983)013<1957:TPNMOV>2.0.CO;2
- 323 Jackson, L. C., Biastoch, A., Buckley, M. W., Desbruyères, D. G., Frajka-Williams,
324 E., Moat, B., & Robson, J. (2022, March). The evolution of the north at-
325 lantic meridional overturning circulation since 1980. *Nature Reviews Earth &
326 Environment*, *3*(4), 241–254. doi: 10.1038/s43017-022-00263-2
- 327 Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q.,
328 ... Wood, E. F. (2014, April). The north american multimodel ensem-
329 ble: Phase-1 Seasonal-to-Interannual prediction; phase-2 toward developing
330 intraseasonal prediction. *Bull. Am. Meteorol. Soc.*, *95*(4), 585–601. doi:
331 10.1175/BAMS-D-12-00050.1
- 332 Kushnir, Y., Scaife, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., ...
333 Wu, B. (2019, January). Towards operational predictions of the near-term
334 climate. *Nat. Clim. Chang.*, *9*(2), 94–101. doi: 10.1038/s41558-018-0359-7
- 335 Lorenz, E. N. (1963, March). Deterministic nonperiodic flow. *J. Atmos. Sci.*, *20*(2),
336 130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2
- 337 Lorenz, E. N. (1969, July). Atmospheric predictability as revealed by natu-
338 rally occurring analogues. *J. Atmos. Sci.*, *26*(4), 636–646. doi: 10.1175/
339 1520-0469(1969)26<636:APARBN>2.0.CO;2
- 340 Lou, Newman, M., & Hoell, A. (2023, February). *Multi-decadal variation of ENSO
341 forecast skill since the late 1800s*. doi: 10.21203/rs.3.rs-2544766/v1
- 342 Maher, N., Milinski, S., Suarez-Gutierrez, L., & others. (2019). The max planck in-
343 stitute grand ensemble: enabling the exploration of climate system variability.
344 *Journal of Advances*. doi: 10.1029/2019MS001639
- 345 Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., Delgado-Torres, C.,
346 Samsó, M., & Bretonnière, P.-A. (2022, October). Constraining low-frequency
347 variability in climate projections to predict climate on decadal to multi-decadal
348 timescales – a poor man’s initialized prediction system. *Earth Syst. Dyn.*,
349 *13*(4), 1437–1450. doi: 10.5194/esd-13-1437-2022
- 350 Martín-Rey, M., Rodríguez-Fonseca, B., & others. (2015). Atlantic opportunities for
351 ENSO prediction. *Geophys. Res. Lett.* doi: 10.1002/2015GL065062

- 352 Menary, M. B., Mignot, J., & Robson, J. (2021, June). Skilful decadal predictions
 353 of subpolar north atlantic SSTs using CMIP model-analogues. *Environ. Res.*
 354 *Lett.*, *16*(6), 064090. doi: 10.1088/1748-9326/ac06fb
- 355 Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho,
 356 C. A. S., ... Yeager, S. (2020, September). Subseasonal to decadal predic-
 357 tion: Filling the Weather–Climate gap. *Bull. Am. Meteorol. Soc.*, *101*(9),
 358 767–770. doi: 10.1175/BAMS-D-19-0037.A
- 359 Mulholland, D. P., Laloyaux, P., Haines, K., & Balmaseda, M. A. (2015, Novem-
 360 ber). Origin and impact of initialization shocks in coupled Atmosphere–
 361 Ocean forecasts. *Mon. Weather Rev.*, *143*(11), 4631–4644. doi: 10.1175/
 362 MWR-D-15-0076.1
- 363 Peng, W., Chen, Q., Zhou, S., & Huang, P. (2021, March). CMIP6 model-
 364 based analog forecasting for the seasonal prediction of sea surface temper-
 365 ature in the offshore area of china. *Geoscience Letters*, *8*(1), 1–8. doi:
 366 10.1186/s40562-021-00179-7
- 367 Rader, J. K., Barnes, E. A., Ebert-Uphoff, I., & others. (2022). Detection of forced
 368 change within combined climate fields using explainable neural networks. *Jour-
 369 nal of Advances*.
- 370 Shekhar, M., Sharma, A., Dimri, A. P., & Tandon, S. K. (2022, July). Asian summer
 371 monsoon variability, global teleconnections, and dynamics during the last 1,000
 372 years. *Earth-Sci. Rev.*, *230*, 104041. doi: 10.1016/j.earscirev.2022.104041
- 373 Si, D., Hu, A., Jiang, D., & Lang, X. (2023, February). Atmospheric telecon-
 374 nection associated with the atlantic multidecadal variability in summer:
 375 assessment of the CESM1 model. *Clim. Dyn.*, *60*(3), 1043–1060. doi:
 376 10.1007/s00382-022-06331-z
- 377 Sutton, R. T., & Allen, M. R. (1997, August). Decadal predictability of north at-
 378 lantic sea surface temperature and climate. *Nature*, *388*(6642), 563–567. doi:
 379 10.1038/41523
- 380 Towler, E., PaiMazumder, D., & Done, J. (2018, March). Toward the application of
 381 decadal climate predictions. *J. Appl. Meteorol. Climatol.*, *57*(3), 555–568. doi:
 382 10.1175/JAMC-D-17-0113.1
- 383 Van den Dool, H. M. (1994, May). Searching for analogues, how long must we wait?
 384 *Tellus A*, *46*(3), 314–324. doi: 10.1034/j.1600-0870.1994.t01-2-00006.x

- 385 Wang, C. (2018, October). A review of ENSO theories. *Natl Sci Rev*, 5(6), 813–825.
386 doi: 10.1093/nsr/nwy104
- 387 Wang, S.-Y., L’Heureux, M., & Chia, H.-H. (2012, March). ENSO prediction one
388 year in advance using western north pacific sea surface temperatures. *Geophys.*
389 *Res. Lett.*, 39(5). doi: 10.1029/2012GL050909
- 390 Wu, Y., & Yan, X. (2023, May). Evaluating changes in the multiyear predictability
391 of the pacific decadal oscillation using model analogs since 1900. *J. Mar. Sci.*
392 *Eng.*, 11(5), 980. doi: 10.3390/jmse11050980
- 393 Yeh, S. W., Cai, W., Min, S. K., McPhaden, M. J., & others. (2018). ENSO atmo-
394 spheric teleconnections and their response to greenhouse gas forcing. *Reviews*
395 *of*. doi: 10.1002/2017RG000568
- 396 Zhang, L., Delworth, T. L., Yang, X., Morioka, Y., Zeng, F., & Lu, F. (2023, Febru-
397 ary). Skillful decadal prediction skill over the southern ocean based on GFDL
398 SPEAR Model-Analogs. *Environ. Res. Commun.*, 5(2), 021002. doi: 10.1088/
399 2515-7620/acb90e

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

Jamin K. Rader¹, and Elizabeth A. Barnes¹

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

Key Points:

- An interpretable neural network provides a spatially-weighted mask for selecting optimal analogs
- Analog selected with the weighted mask offer more skillful forecasts than traditional methods for selecting analogs
- The learned mask highlights precursor regions for predicting large-scale climate anomalies in a perfect model framework

Corresponding author: Jamin K. Rader, jamin.rader@colostate.edu

12 **Abstract**

13 Seasonal-to-decadal climate prediction is crucial for decision-making in a number of in-
14 dustries, but forecasts on these timescales have limited skill. Here, we develop a data-
15 driven method for selecting optimal analogs for seasonal-to-decadal analog forecasting.
16 Using an interpretable neural network, we learn a spatially-weighted mask that quan-
17 tifies how important each grid point is for determining whether two climate states will
18 evolve similarly. We show that analogs selected using this weighted mask provide more
19 skillful forecasts than analogs that are selected using traditional spatially-uniform meth-
20 ods. This method is tested on two prediction problems within a perfect model frame-
21 work using the Max Planck Institute for Meteorology Grand Ensemble: multi-year pre-
22 diction of North Atlantic sea surface temperatures, and seasonal prediction of El Niño
23 Southern Oscillation. This work demonstrates a methodical approach to selecting analogs
24 that may be useful for improving seasonal-to-decadal forecasts and understanding their
25 sources of skill.

26 **Plain Language Summary**

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

37 **1 Background**

38 Forecasts on seasonal-to-decadal timescales are crucial for decision-makers in a num-
39 ber of industries, but forecasts on these timescales have limited skill (Kushnir et al., 2019;
40 Merryfield et al., 2020; Towler et al., 2018). Analog forecasting, predicting what will hap-
41 pen based on previous states with similar initial conditions, is an intuitive method for
42 seasonal-to-decadal prediction. It is built on the premise that similar geophysical states

43 will evolve in similar ways (Lorenz, 1969). It follows that analogs—similar looking states
44 to the initial state that is being forecast—can provide insight into how that initial state
45 will continue to evolve. The analog forecasting approach is powerful for seasonal-to-decadal
46 climate prediction (e.g., Ding et al., 2018, 2019; Menary et al., 2021; Delle Monache et
47 al., 2013; Zhang et al., 2023) and can outperform general circulation models (GCMs) ini-
48 tialized with observations, which struggle with initialization shock and climate model
49 drift (Merryfield et al., 2020; Mulholland et al., 2015).

50 A major hurdle in obtaining successful analog forecasts is that the climate system
51 is noisy and chaotic, and thus small differences between two initial states can result in
52 vast differences in their evolution (Lorenz, 1963). Thus, a successful analog forecast for
53 a particular initial climate state, which we refer to as the state of interest (SOI), requires
54 that the analogs and SOI are sufficiently similar such that their evolutions do not sig-
55 nificantly diverge during the prediction timeframe. Sufficiently similar analogs can be
56 difficult to find in the observational record since the number of independent observations
57 we have on seasonal-to-decadal scales (e.g., fewer than 100 during the satellite era) is so
58 much smaller than the number of degrees of freedom within a global geophysical field
59 (e.g., Van den Dool, 1994). While observations are in short supply, there is a wealth of
60 simulated climate data and many recent studies have used “model-analogs” (Ding et al.,
61 2018) drawn from climate model output instead (e.g., Lou et al., 2023; Peng et al., 2021;
62 Wu & Yan, 2023).

63 We refer to the library of climate model states that can be used for analog fore-
64 casting as “potential analogs.” Once a potential analog has been identified to be suffi-
65 ciently similar to the SOI we refer to it as an analog. Forecasts are made by taking the
66 mean evolution of the top-N analogs, where N is chosen by the user. There are several
67 ways to quantify the similarity between the potential analogs and the SOI. The most straight-
68 forward method is to compute the global correlation between each potential analog and
69 the SOI (e.g., Mahmood et al., 2022). Using a global correlation assumes that the sim-
70 ilarity between the maps at each grid point globally matters equally. A natural next step
71 in complexity is to compute a correlation over a region that is known to be important
72 for predictability of a given target, such as the North Pacific for predicting the Pacific
73 Decadal Oscillation (e.g., Wu & Yan, 2023). While this approach removes some regions
74 that may not be useful for determining the best analogs, it still assumes that each grid
75 point within the region is equally important and the region must be known *a priori*.

76 In the following work, we train an interpretable neural network on a proxy task that
 77 is similar to the analog problem (Section 3). The network learns a weighted mask which
 78 is used for determining analogs. The forecasting skill of the analogs selected using the
 79 learned weighted mask is tested through a perfect model approach where climate model
 80 data substitutes observations and is used to predict future climate model data. We show
 81 in two examples, forecasting 5-year sea surface temperature (SST) anomalies in the North
 82 Atlantic (Section 4) and wintertime SST anomalies in the tropical Pacific (i.e. El Niño
 83 Southern Oscillation; Section 5), that analogs identified using the weighted mask pro-
 84 vide more skillful forecasts than analogs that are identified in a way that is globally or
 85 regionally uniform. In addition, we show that these masks, once generated by a neural
 86 network, can be modified *post hoc* to further investigate the importance of each region
 87 for seasonal-to-decadal prediction (Section 5).

88 2 Data and Metrics

89 2.1 Climate Model Data

90 We use monthly SST from the historical run of the Max Planck Institute (MPI)
 91 for Meteorology Grand Ensemble (GE; Maher et al., 2019) at 2° latitude by 2° lon-
 92 gitude resolution. This dataset contains 100 members and each simulates 156 years (1850-
 93 2005) of the Earth’s climate with historical forcing. The MPI-GE uses the MPI Earth
 94 System Model version 1.1 (ESM1.1; Giorgetta et al., 2013). Each member is initialized
 95 using a different year of the preindustrial control simulation such that the differences be-
 96 tween ensemble members are a product of internal variability.

97 2.2 Standardization and Selection

98 Subsets of the MPI-GE ensemble members are used for different purposes. Our li-
 99 brary of potential analogs is made up of members 1-35. Members 36-50 are the SOIs for
 100 training the neural network, members 51-55 are the SOIs for the early stopping valida-
 101 tion set (which is used to prevent overfitting to the training data), and members 56-60
 102 are the SOIs for the tuning validation set (which is used to identify optimal hyperpa-
 103 rameters for the neural network). Finally, members 96-100, which are withheld until the
 104 very end, are the test set for making and evaluating the analog forecasts. Details on the

105 process of tuning and training the neural network, including selecting the hyperparam-
 106 eters, can be found in Section S1.

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

112 We removed the forced signal from the climate model data by subtracting the en-
 113 semble mean of the library of potential analogs at each location and year from each set
 114 of data. After the forced signal was removed, the data was standardized by dividing by
 115 the standard deviation at each grid point across the library of potential analogs. By us-
 116 ing the library of potential analogs to calculate the forced signal and internal variance
 117 we treat the SOIs as if they are truly unseen data as we would when forecasting.

118 2.3 Metrics

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

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

122 such that a perfect prediction has a score of one, and a climatology prediction has a score
 123 of zero. Climatology is the prediction by the mean state, which is zero for this standard-
 124 ized data. Analog forecasts made using the weighted mask are compared with the fol-
 125 lowing additional baselines: a global analog forecast, a target region analog forecast, a
 126 mean target evolution forecast, and a random forecast. In the global analog forecast (tar-
 127 get region analog forecast), the analogs are selected if the unweighted MSE over the en-
 128 tire globe (target region) is the smallest. The mean target evolution forecast is based on
 129 how the targets in the input period evolve on average and is detailed in Section S2. The
 130 random forecast is made by randomly selecting targets from the library of potential analogs
 131 and using them as the prediction.

132 **3 Optimized Analog Forecasting Approach**

133 Our goal is to find optimal analogs for forecasting a specific target. To do this, we
 134 train a neural network to identify a spatially-weighted mask. This weighted mask is then
 135 multiplied by the SOI and potential analogs and the mean-squared error (MSE) between
 136 the weighted maps is used to determine how similar they are (Figure 1). This weighted
 137 mask should contain large values where similarity between the analogs and the SOI is
 138 most important for predicting the target and near-zero values where similarity between
 139 the maps is not important. With this architecture, the MSE will be low if the maps agree
 140 where the mask weights are high, regardless of the differences between the maps where
 141 the mask weights are low. For the plots in this paper, the mask is normalized by divid-
 142 ing by the sum of the weights times the size of the input, such that the mean weight is
 143 one.

144 We generate the weighted mask by training a neural network on a proxy task that
 145 is tangential to our main goal. While our goal is to identify a weighted mask that is op-
 146 timized for making an analog forecast, our proxy task is to predict the difference in $T_{SOI,i}$
 147 and $T_{analog,j}$ given $I_{SOI,i}$ and $I_{analog,j}$. En route to making this prediction, the neural
 148 network must learn the weighted mask, multiply it by the two input maps, compute the
 149 MSE between these weighted maps, and finally convert the MSE into a predicted dif-
 150 ference in the targets. This process is depicted in the red box of Figure 1.

151 Once the weighted mask has been learned, a neural network is no longer needed
 152 to make analog predictions. The weighted mask is multiplied by the SOI and each po-
 153 tential analog, the MSE is computed between the weighted SOI and the weighted po-
 154 tential analogs, and the potential analogs with the lowest MSE are used to make the ana-
 155 log forecast. While the proxy task is not identical to the analog problem, it provides a
 156 weighted mask that improves analog forecasting skill, as we will show in Sections 4 and
 157 5.

158 **4 Multi-year Prediction of North Atlantic Sea Surface Temperature**

159 We first test our analog forecasting approach on a multi-year prediction of SSTs
 160 over the North Atlantic. North Atlantic SSTs exhibit clear variability on multi-annual
 161 timescales (Jackson et al., 2022) and exhibit potential for skillful decadal forecasts (Hawkins
 162 et al., 2011; Sutton & Allen, 1997). SST variability in the North Atlantic has been as-

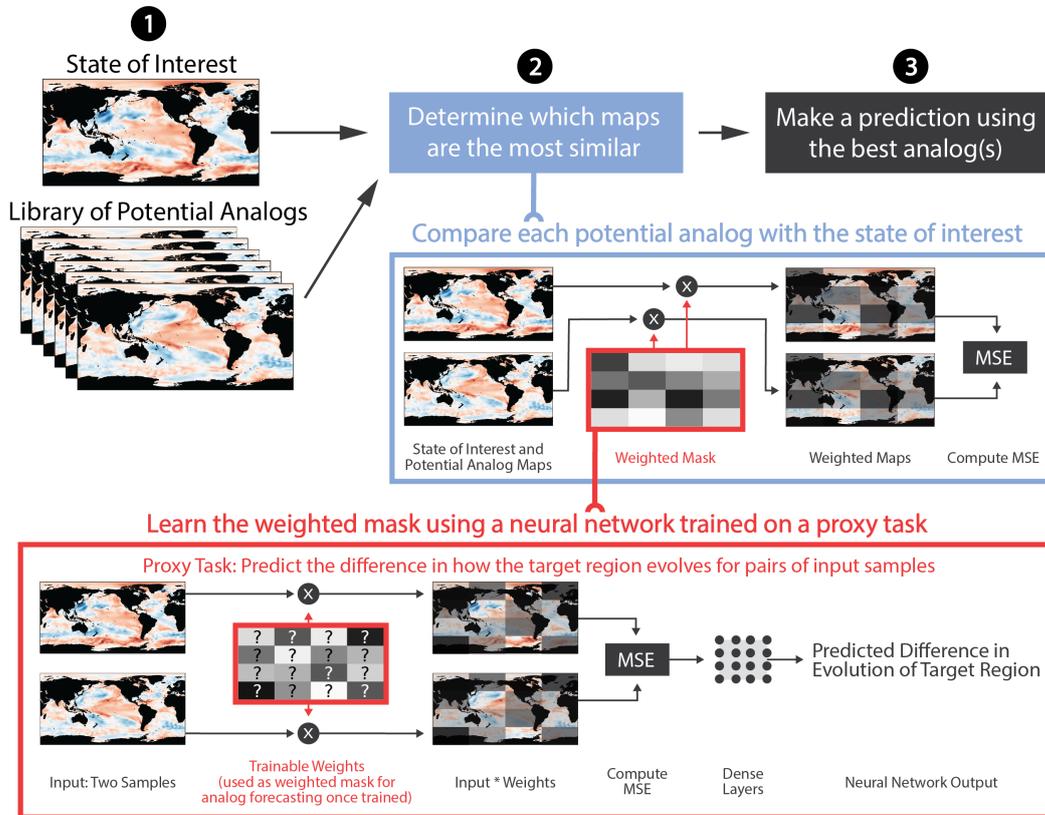


Figure 1. Optimized analog forecasting method and interpretable neural network architecture. The analog forecasting method can be described in three steps: 1) identify a state of interest and a library of potential analogs. 2) Determine which maps are the most similar. 3) Make a prediction using the 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.

163 sociated with weather and climate anomalies globally, including Atlantic hurricane fre-
164 quency and intensity (Goldenberg et al., 2001; Balaguru et al., 2018), northern hemisphere
165 precipitation (Enfield et al., 2001; Si et al., 2023), and the strength of the Asian sum-
166 mer monsoon (Shekhar et al., 2022). In this prediction problem, we use global maps of
167 SST, averaged over the previous five years, to predict the mean SST anomaly in the North
168 Atlantic (40° - 60° N, 10° - 70° W) over the following five years.

169 The weighted mask learned by the neural network is shown in Figure 2a. The Green-
170 land Sea and the gulf stream region in the western North Atlantic emerge as the most
171 important regions for identifying analogs in the MPI-GE. Over the western North At-
172 lantic, there is an area of zero weight between two areas of high weight. These may be
173 where the boundaries of persistent SST anomalies vary, and the neural network has learned
174 that the specific locations of these boundaries are not important for the prediction prob-
175 lem. Previous studies that have used an analog approach to assess North Atlantic decadal
176 predictability selected the best analogs by taking a correlation over the whole globe (Mahmood
177 et al., 2022) or the entire North Atlantic basin (Menary et al., 2021). As shown in Fig-
178 ure 2b-d, when using the weighted mask, the best analogs only have to look like the SOI
179 in the highest weight regions. An example SOI is shown in Figure 2b and its best ana-
180 log in Figure 2c. These two maps look similar in the North Atlantic, but are starkly dif-
181 ferent in the North Pacific and Indian Ocean, among other regions. Once the weighted
182 mask has been applied to the SOI (Figure 2d) and its best analog (Figure 2e), the maps
183 look nearly identical.

184 These results suggest that using uniform weights across the entire North Atlantic
185 basin, or the whole globe, may lead to a selection of analogs that are not optimized for
186 forecasting multi-year variability in the North Atlantic. Indeed, we see that this is true
187 in the skill scores shown in Figure 3a. For $1 \leq N \leq 50$, where the top-N analogs are av-
188 eraged, our weighted mask analog forecast outperforms the global and target region ana-
189 log forecasts, as well as the climatology, mean target evolution, and random baselines.
190 The skill score is lowest when only the single best analog is used for forecasting, and sub-
191 sequently improves for larger N. Given that the skill score maximizes around $N = 10$,
192 and the spread of the targets associated with the analogs (i.e. the uncertainty of the fore-
193 cast) increases with N (Figure S1), we elect to focus on results for $N = 10$ analogs. The
194 prediction by the top-10 analogs, and the spread of the targets, are shown in Figure 3b
195 for 200 years of SOIs. The analog predictions do a good job of capturing the variabil-

196 ity of North Atlantic sea surface temperatures, though they do struggle to forecast the
197 most extreme anomalies.

198 **5 Seasonal Prediction of El Niño Southern Oscillation**

199 In addition to improving multi-year forecasts of SST in the North Atlantic, the learned
200 weighted mask improves forecasts of ENSO on seasonal timescales. ENSO is the lead-
201 ing mode of global annual SST variability (Hsiung & Newell, 1983) and has an exten-
202 sive influence on global weather and climate (reviewed in Yeh et al., 2018). Analog fore-
203 casting has been applied to seasonal prediction of ENSO in several studies due to its po-
204 tential to outperform initialized GCM forecasts (e.g., Ding et al., 2018, 2019). In the fol-
205 lowing example, we use wintertime (November-March) global SST anomalies to forecast
206 SST anomalies in the Niño3.4 region (5°S - 5°N , 120°W - 170°W ; Barnston et al., 1997; Han-
207 ley et al., 2003) the following winter.

208 The weighted mask for forecasting ENSO looks markedly different from that for
209 forecasting North Atlantic multi-year variability (Figure 4a). While a few regions are as-
210 signed higher weights, the weights in Figure 4a are much more uniform across the globe
211 than in Figure 2a. The four main regions that stand out in this weighted mask have also
212 been identified as important precursors in previous literature: the western North Pacific
213 (e.g., S.-Y. Wang et al., 2012), the Pacific Meridional Mode (e.g., Amaya, 2019), the Cen-
214 tral Atlantic (e.g., Martín-Rey et al., 2015), and the tropical Pacific itself (e.g., Capot-
215 tondi & Sardeshmukh, 2015). The skill score of the global analog forecast (Figure 4b)
216 is similar to that of our weighted mask analog forecast (but always lower, see Figure S3),
217 which is not surprising since the values of the weighted mask are near one for most ar-
218 eas of the globe.

219 Since the weighted mask can be manually updated *post hoc*, we use this to explore
220 the sensitivity of the forecast skill to which regions are included in the weighted mask.
221 Figure 5a shows the weighted mask for ENSO prediction (Figure 4a) but where the small-
222 est 95 percent of the weights have been set to zero. Forecasts made with this “constrained”
223 weighted mask have similar skill to the original weighted mask (as shown in Figure S4).
224 From the constrained weighted mask, we identify four main precursor regions for ENSO:
225 the West Pacific (ocean grid points bounded by 0° - 40°N , 100° - 170°E), the Tropical Pa-

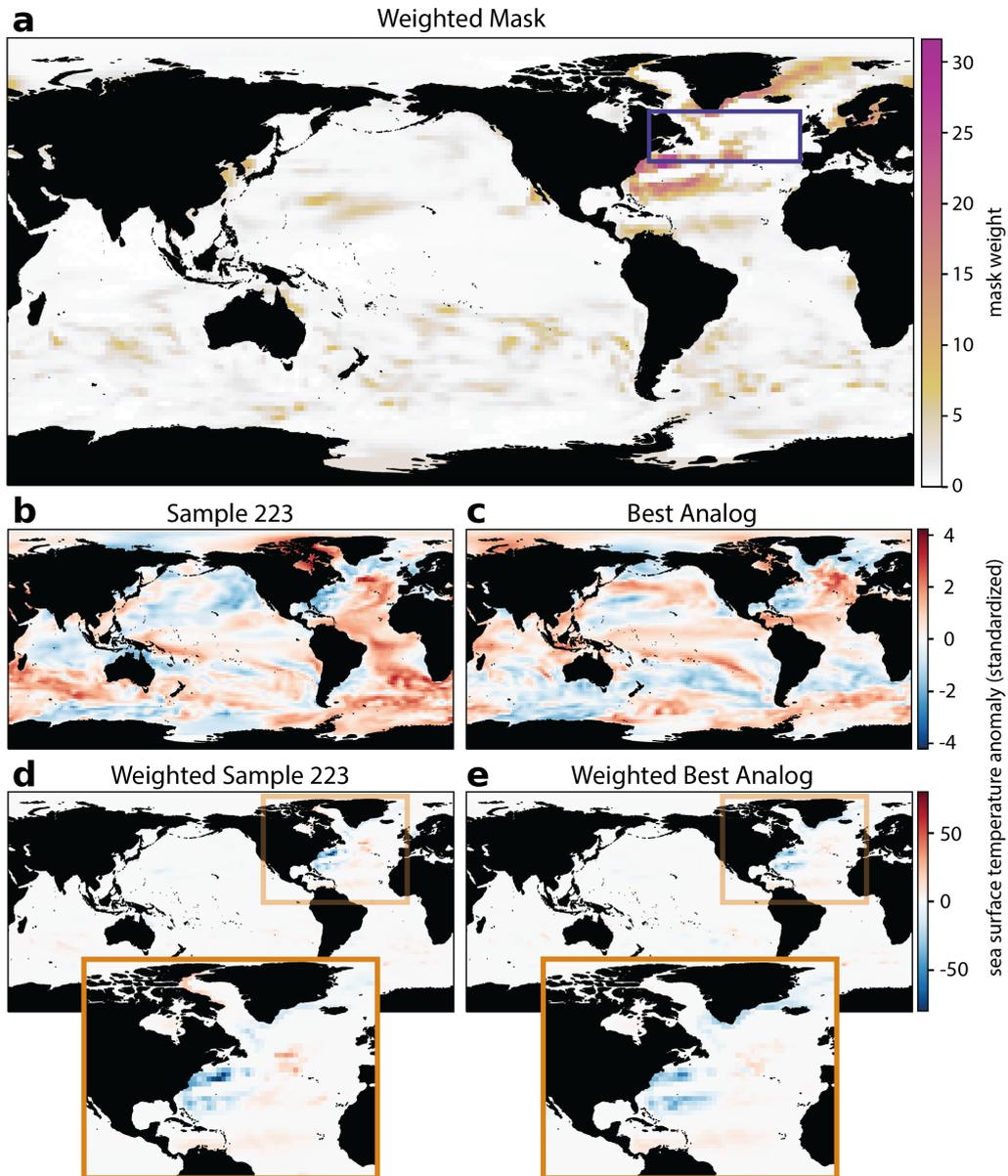


Figure 2. Weighted mask and example for multi-year predictions of North Atlantic SST. (a) Weighted mask, as learned by the interpretable neural network. (b) Standardized SST anomalies for a sample state of interest (SOI). (c) Standardized SST anomalies for the best analog associated with the SOI. (d) Weighted SOI. (e) Weighted best analog.

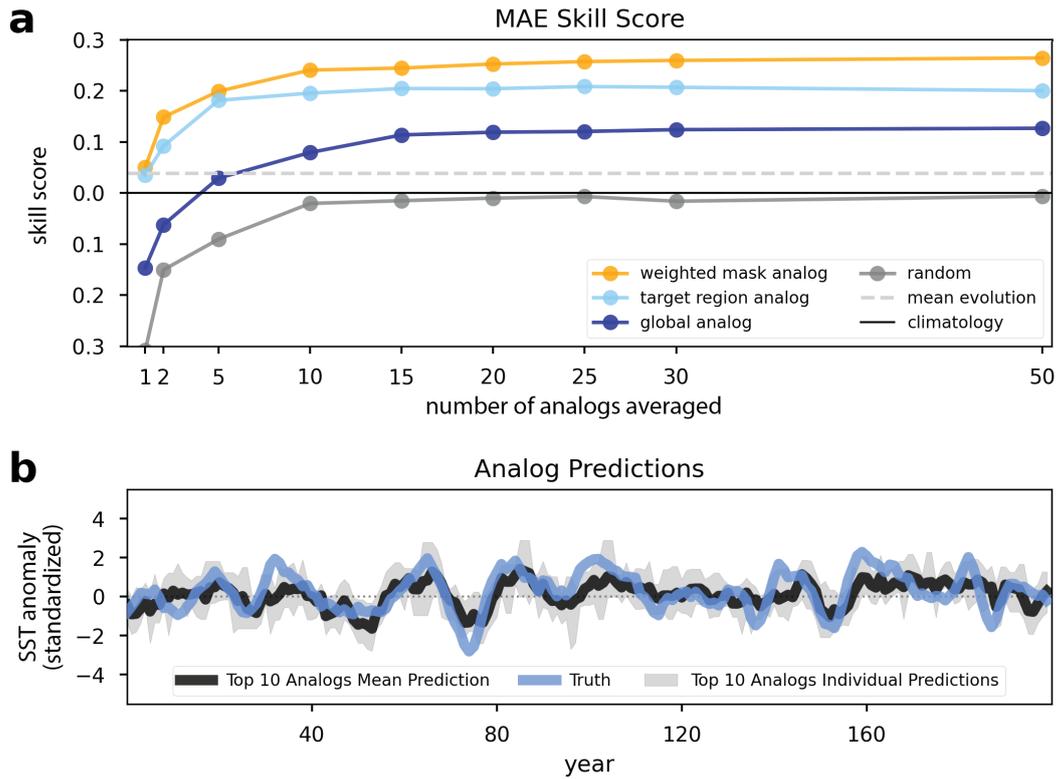


Figure 3. Analog forecasts of North Atlantic sea surface temperature. (a) Skill scores for our weighted mask analog forecast and other baselines. (b) Weighted mask analog forecasts for 200 years of MPI-GE simulations, including the mean prediction from the top-10 analogs, the spread of these predictions, and the truth values.

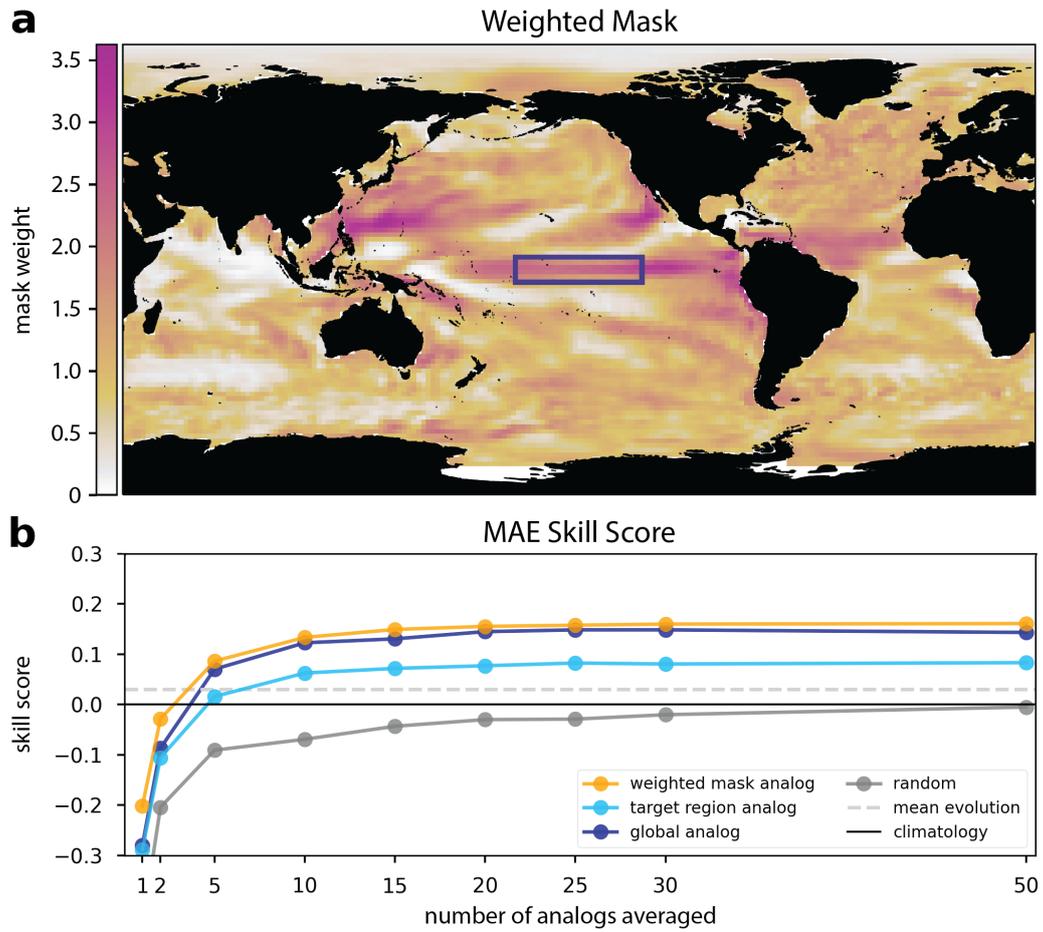


Figure 4. Weighted mask and skill scores for seasonal predictions of El Niño Southern Oscillation. (a) Weighted mask. (b) Skill scores for our weighted mask analog and other baselines.

226 cific (25°S-10°N, 170°E-65°W), the Baja Coast (10°N-40°N, 110°-140°W), and the Trop-
227 ical Atlantic (0°-20°N, 20°-80°W).

228 We assess how important each precursor region is in two ways. In the first approach,
229 we test the skill score of analog forecasting when each region is occluded from the con-
230 strained weighted mask (weights in that region are set to zero). When all four regions
231 are included, the skill score is 0.146. Removing any of the four regions from the weighted
232 mask results in a skill score decrease. Interestingly, removing the Tropical Atlantic re-
233 sults in the most drastic decrease in prediction skill. While the Tropical Atlantic has been
234 connected to ENSO predictability (e.g., Martín-Rey et al., 2015), it is not considered a
235 primary driver (C. Wang, 2018). In the second approach, we isolate each of the four re-
236 gions (weights outside that region are set to zero). There is no improvement over clima-
237 tology when just the Baja Coast or Tropical Atlantic is used to select analogs, and more
238 skill when just the West Pacific or Tropical Pacific is used. However, no region alone pro-
239 vides anywhere near the skill that all four regions do together.

240 6 Discussion and Conclusions

241 We have shown how an interpretable neural network can be used to identify a weighted
242 mask that improves the selection of analogs for seasonal-to-decadal forecasting. The pre-
243 cursors identified in the weighted masks are not necessarily causal, but they do provide
244 the optimal predictors for the given input. In this work we have constrained the neu-
245 ral network to learn one mask that represents all pathways of predictability, however al-
246 lowing the network to learn different masks for different SOIs could lead to better ana-
247 log forecasts.

248 While we only used a single input map of SST to predict a future target SST in
249 this work, this neural network architecture can be used for many other forecasting ap-
250 proaches. For example, a combination of multiple variables can be used as the predic-
251 tors (such as SST and sea surface height, as in Ding et al., 2018) or other geophysical
252 variables may be selected as the target (e.g., predicting precipitation over land). In ad-
253 dition, one may also include variables at multiple lead times to capture the time tendency
254 of the climate system. We show results that include SST tendency as an input for the
255 North Atlantic multi-year prediction example in Figure S5.

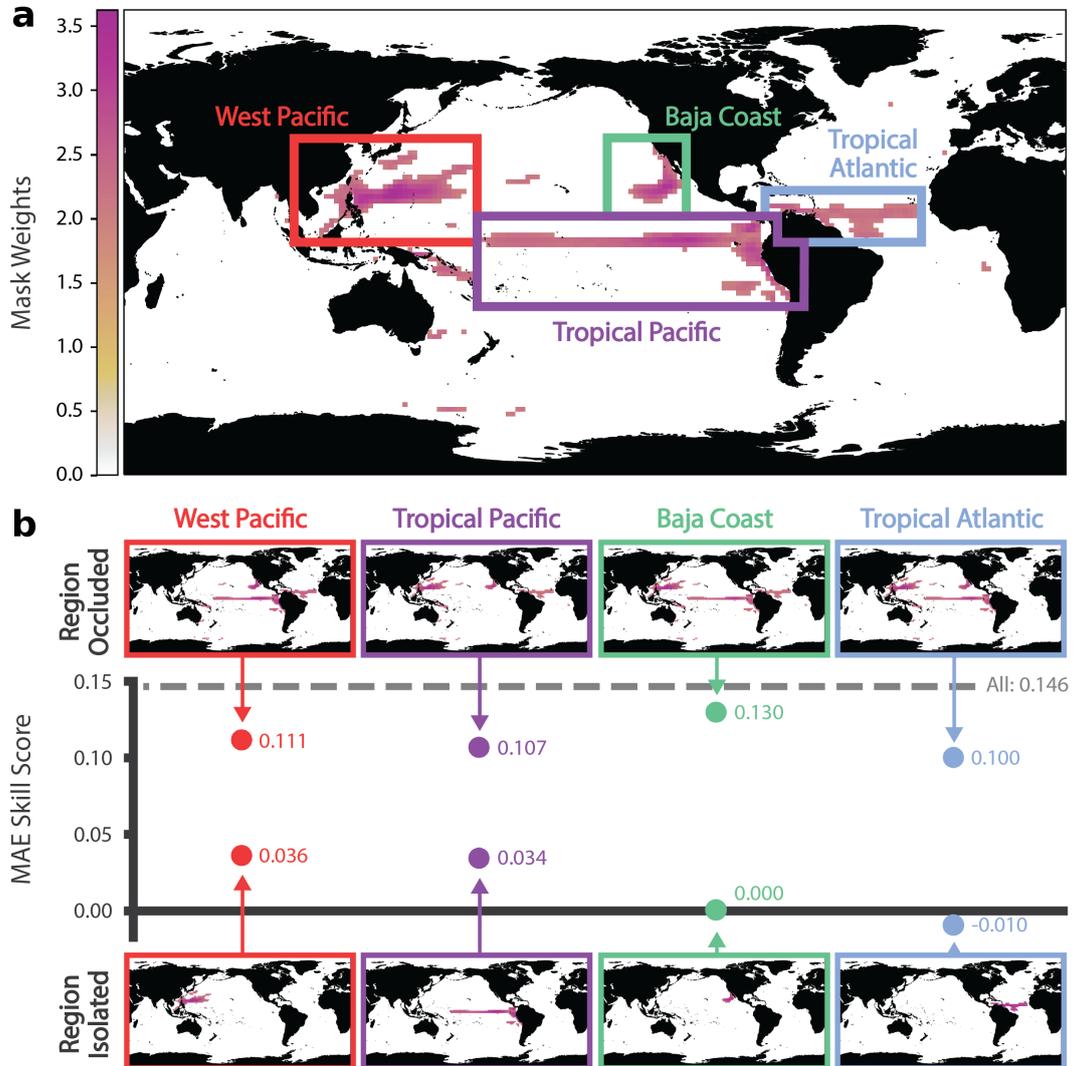


Figure 5. Analog forecasting skill of El Niño Southern Oscillation when various regions are occluded or isolated. (a) As in Figure 4a, but the lowest 95 percent of weights are set to zero. Four regions of focus are highlighted by the colored boxes. (b) Skill scores for analog forecasts when each region is occluded from the mask (top) and when the region is isolated to make a forecast (bottom).

256 We have explored this method through a perfect model setup. As such, the iden-
257 tified precursors are intrinsic to MPI-ESM1.1 and may not reflect patterns of predictabil-
258 ity in the observed Earth system. Training the weighted mask on a multi-model ensem-
259 ble may provide patterns that are more consistent with observations (e.g. Kirtman et
260 al., 2014; Rader et al., 2022) and allow for enhanced analog predictions on real data. Ad-
261 ditionally, we could train on models and observations at the same time to identify a weighted
262 mask that is more representative of the true Earth System. We believe that this weighted
263 mask approach will be influential to analog forecasting moving forward.

264 Open Research Section

265 The data used in this study, simulations from the Max Planck Institute for Mete-
266 orology Grand Ensemble, are publicly available at [https://esgf-data.dkrz.de/projects/mpi-
267 ge/](https://esgf-data.dkrz.de/projects/mpi-ge/). The weighted masks, and all python code used to generate the data and figures in
268 this paper, can be found at XXX REVIEWERS, THIS CAN CURRENTLY BE FOUND
269 AT <https://github.com/jaminrader/WeightedMaskAnalogForecasting> AND WILL BE
270 UPLOADED TO ZENODO WHEN FINISHED XXX.

271 Acknowledgments

272 This material is based upon work supported by the U.S. Department of Energy,
273 Office of Science, Office of Advanced Scientific Computing Research, Department of En-
274 ergy Computational Science Graduate Fellowship under Award Number DE-SC0020347,
275 and by grant AGS-2210068 from the National Science Foundation. The authors would
276 like to thank colleagues at the NOAA Physical Sciences Laboratory for their helpful feed-
277 back.

278 References

- 279 Amaya, D. J. (2019, December). The pacific meridional mode and ENSO: a re-
280 view. *Current Climate Change Reports*, 5(4), 296–307. doi: 10.1007/s40641
281 -019-00142-x
- 282 Balaguru, K., Foltz, G. R., & Leung, L. R. (2018, May). Increasing magnitude of
283 hurricane rapid intensification in the central and eastern tropical atlantic. *Geo-
284 phys. Res. Lett.*, 45(9), 4238–4247. doi: 10.1029/2018gl077597
- 285 Barnston, A. G., Chelliah, M., & Goldenberg, S. B. (1997, September). Doc-

- 286 umentation of a highly ENSO-related sst region in the equatorial pa-
287 cific: Research note. *Atmosphere-Ocean*, *35*(3), 367–383. doi: 10.1080/
288 07055900.1997.9649597
- 289 Capotondi, A., & Sardeshmukh, P. D. (2015, November). Optimal precursors of dif-
290 ferent types of ENSO events. *Geophys. Res. Lett.*, *42*(22), 9952–9960. doi: 10
291 .1002/2015gl066171
- 292 Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B., & Searight, K. (2013,
293 October). Probabilistic weather prediction with an analog ensemble. *Mon.*
294 *Weather Rev.*, *141*(10), 3498–3516. doi: 10.1175/mwr-d-12-00281.1
- 295 Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2018, July). Skill-
296 ful climate forecasts of the tropical Indo-Pacific ocean using Model-Analogs. *J.*
297 *Clim.*, *31*(14), 5437–5459. doi: 10.1175/JCLI-D-17-0661.1
- 298 Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2019, February).
299 Diagnosing secular variations in retrospective ENSO seasonal forecast skill
300 using CMIP5 model-analogs. *Geophys. Res. Lett.*, *46*(3), 1721–1730. doi:
301 10.1029/2018gl080598
- 302 Enfield, D. B., Mestas-Nuñez, A. M., & Trimble, P. J. (2001, May). The at-
303 lantic multidecadal oscillation and its relation to rainfall and river flows
304 in the continental U.S. *Geophys. Res. Lett.*, *28*(10), 2077–2080. doi:
305 10.1029/2000gl012745
- 306 Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M.,
307 ... Stevens, B. (2013, July). Climate and carbon cycle changes from 1850 to
308 2100 in MPI-ESM simulations for the coupled model intercomparison project
309 phase 5. *J. Adv. Model. Earth Syst.*, *5*(3), 572–597. doi: 10.1002/jame.20038
- 310 Goldenberg, S. B., Landsea, C. W., Mestas-Nunez, A. M., & Gray, W. M. (2001,
311 July). The recent increase in atlantic hurricane activity: causes and implica-
312 tions. *Science*, *293*(5529), 474–479. doi: 10.1126/science.1060040
- 313 Hanley, D. E., Bourassa, M. A., O’Brien, J. J., Smith, S. R., & Spade, E. R. (2003,
314 April). A quantitative evaluation of ENSO indices. *J. Clim.*, *16*(8), 1249–1258.
315 doi: 10.1175/1520-0442(2003)16<1249:AQEOEI>2.0.CO;2
- 316 Hawkins, E., Robson, J., Sutton, R., Smith, D., & Keenlyside, N. (2011, December).
317 Evaluating the potential for statistical decadal predictions of sea surface tem-
318 peratures with a perfect model approach. *Clim. Dyn.*, *37*(11-12), 2495–2509.

- 319 doi: 10.1007/s00382-011-1023-3
- 320 Hsiung, J., & Newell, R. E. (1983, October). The principal nonseasonal modes of
321 variation of global sea surface temperature. *J. Phys. Oceanogr.*, *13*(10), 1957–
322 1967. doi: 10.1175/1520-0485(1983)013<1957:TPNMOV>2.0.CO;2
- 323 Jackson, L. C., Biastoch, A., Buckley, M. W., Desbruyères, D. G., Frajka-Williams,
324 E., Moat, B., & Robson, J. (2022, March). The evolution of the north at-
325 lantic meridional overturning circulation since 1980. *Nature Reviews Earth &
326 Environment*, *3*(4), 241–254. doi: 10.1038/s43017-022-00263-2
- 327 Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q.,
328 ... Wood, E. F. (2014, April). The north american multimodel ensem-
329 ble: Phase-1 Seasonal-to-Interannual prediction; phase-2 toward developing
330 intraseasonal prediction. *Bull. Am. Meteorol. Soc.*, *95*(4), 585–601. doi:
331 10.1175/BAMS-D-12-00050.1
- 332 Kushnir, Y., Scaife, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., ...
333 Wu, B. (2019, January). Towards operational predictions of the near-term
334 climate. *Nat. Clim. Chang.*, *9*(2), 94–101. doi: 10.1038/s41558-018-0359-7
- 335 Lorenz, E. N. (1963, March). Deterministic nonperiodic flow. *J. Atmos. Sci.*, *20*(2),
336 130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2
- 337 Lorenz, E. N. (1969, July). Atmospheric predictability as revealed by natu-
338 rally occurring analogues. *J. Atmos. Sci.*, *26*(4), 636–646. doi: 10.1175/
339 1520-0469(1969)26<636:APARBN>2.0.CO;2
- 340 Lou, Newman, M., & Hoell, A. (2023, February). *Multi-decadal variation of ENSO
341 forecast skill since the late 1800s*. doi: 10.21203/rs.3.rs-2544766/v1
- 342 Maher, N., Milinski, S., Suarez-Gutierrez, L., & others. (2019). The max planck in-
343 stitute grand ensemble: enabling the exploration of climate system variability.
344 *Journal of Advances*. doi: 10.1029/2019MS001639
- 345 Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., Delgado-Torres, C.,
346 Samsó, M., & Bretonnière, P.-A. (2022, October). Constraining low-frequency
347 variability in climate projections to predict climate on decadal to multi-decadal
348 timescales – a poor man’s initialized prediction system. *Earth Syst. Dyn.*,
349 *13*(4), 1437–1450. doi: 10.5194/esd-13-1437-2022
- 350 Martín-Rey, M., Rodríguez-Fonseca, B., & others. (2015). Atlantic opportunities for
351 ENSO prediction. *Geophys. Res. Lett.* doi: 10.1002/2015GL065062

- 352 Menary, M. B., Mignot, J., & Robson, J. (2021, June). Skilful decadal predictions
353 of subpolar north atlantic SSTs using CMIP model-analogues. *Environ. Res.
354 Lett.*, *16*(6), 064090. doi: 10.1088/1748-9326/ac06fb
- 355 Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho,
356 C. A. S., ... Yeager, S. (2020, September). Subseasonal to decadal predic-
357 tion: Filling the Weather–Climate gap. *Bull. Am. Meteorol. Soc.*, *101*(9),
358 767–770. doi: 10.1175/BAMS-D-19-0037.A
- 359 Mulholland, D. P., Laloyaux, P., Haines, K., & Balmaseda, M. A. (2015, Novem-
360 ber). Origin and impact of initialization shocks in coupled Atmosphere–
361 Ocean forecasts. *Mon. Weather Rev.*, *143*(11), 4631–4644. doi: 10.1175/
362 MWR-D-15-0076.1
- 363 Peng, W., Chen, Q., Zhou, S., & Huang, P. (2021, March). CMIP6 model-
364 based analog forecasting for the seasonal prediction of sea surface temper-
365 ature in the offshore area of china. *Geoscience Letters*, *8*(1), 1–8. doi:
366 10.1186/s40562-021-00179-7
- 367 Rader, J. K., Barnes, E. A., Ebert-Uphoff, I., & others. (2022). Detection of forced
368 change within combined climate fields using explainable neural networks. *Jour-
369 nal of Advances*.
- 370 Shekhar, M., Sharma, A., Dimri, A. P., & Tandon, S. K. (2022, July). Asian summer
371 monsoon variability, global teleconnections, and dynamics during the last 1,000
372 years. *Earth-Sci. Rev.*, *230*, 104041. doi: 10.1016/j.earscirev.2022.104041
- 373 Si, D., Hu, A., Jiang, D., & Lang, X. (2023, February). Atmospheric telecon-
374 nection associated with the atlantic multidecadal variability in summer:
375 assessment of the CESM1 model. *Clim. Dyn.*, *60*(3), 1043–1060. doi:
376 10.1007/s00382-022-06331-z
- 377 Sutton, R. T., & Allen, M. R. (1997, August). Decadal predictability of north at-
378 lantic sea surface temperature and climate. *Nature*, *388*(6642), 563–567. doi:
379 10.1038/41523
- 380 Towler, E., PaiMazumder, D., & Done, J. (2018, March). Toward the application of
381 decadal climate predictions. *J. Appl. Meteorol. Climatol.*, *57*(3), 555–568. doi:
382 10.1175/JAMC-D-17-0113.1
- 383 Van den Dool, H. M. (1994, May). Searching for analogues, how long must we wait?
384 *Tellus A*, *46*(3), 314–324. doi: 10.1034/j.1600-0870.1994.t01-2-00006.x

- 385 Wang, C. (2018, October). A review of ENSO theories. *Natl Sci Rev*, 5(6), 813–825.
386 doi: 10.1093/nsr/nwy104
- 387 Wang, S.-Y., L’Heureux, M., & Chia, H.-H. (2012, March). ENSO prediction one
388 year in advance using western north pacific sea surface temperatures. *Geophys.*
389 *Res. Lett.*, 39(5). doi: 10.1029/2012GL050909
- 390 Wu, Y., & Yan, X. (2023, May). Evaluating changes in the multiyear predictability
391 of the pacific decadal oscillation using model analogs since 1900. *J. Mar. Sci.*
392 *Eng.*, 11(5), 980. doi: 10.3390/jmse11050980
- 393 Yeh, S. W., Cai, W., Min, S. K., McPhaden, M. J., & others. (2018). ENSO atmo-
394 spheric teleconnections and their response to greenhouse gas forcing. *Reviews*
395 *of*. doi: 10.1002/2017RG000568
- 396 Zhang, L., Delworth, T. L., Yang, X., Morioka, Y., Zeng, F., & Lu, F. (2023, Febru-
397 ary). Skillful decadal prediction skill over the southern ocean based on GFDL
398 SPEAR Model-Analogs. *Environ. Res. Commun.*, 5(2), 021002. doi: 10.1088/
399 2515-7620/acb90e

Supporting Information for “Optimizing Seasonal-to-Decadal Analog Forecasts with a Learned Spatially-Weighted Mask”

Jamin K. Rader¹, and Elizabeth A. Barnes¹

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

Contents of this file

1. Text S1 to S2
2. Figures S1 to S4
3. Tables S1 to S6

Text S1: Neural Network Training and Hyperparameter Tuning

The interpretable neural network architecture, shown in the blue box of Figure 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 S4, 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 S1 and the hyperparameter search space in Table S2. 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 S3-S4.

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 S5-S6. 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 S1 and S2.

Text S2: Baselines

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.

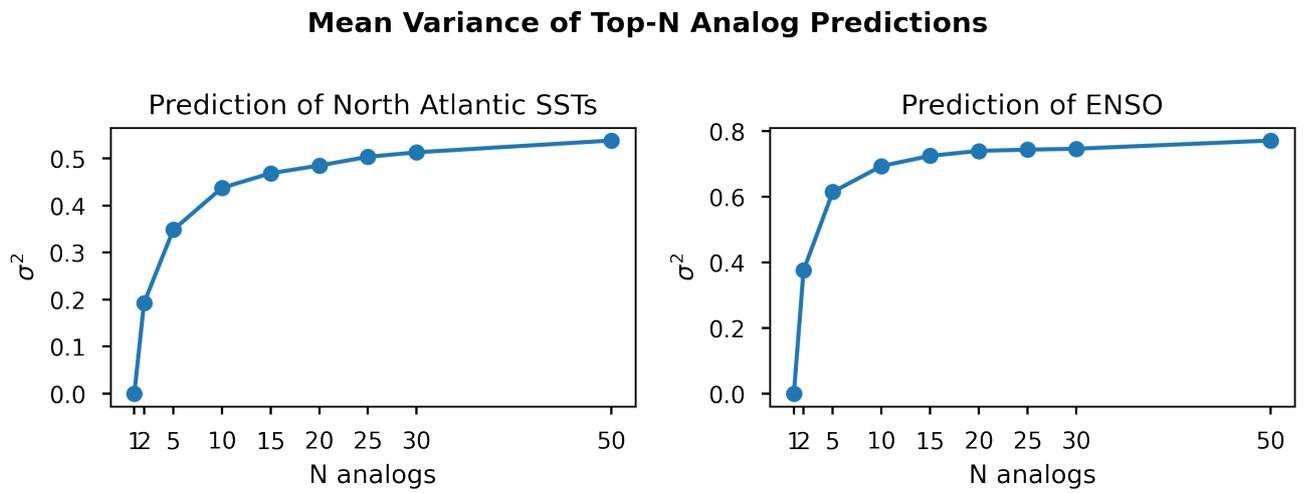


Figure S1. Mean variance of the targets associated with the top-N analogs across all testing samples.

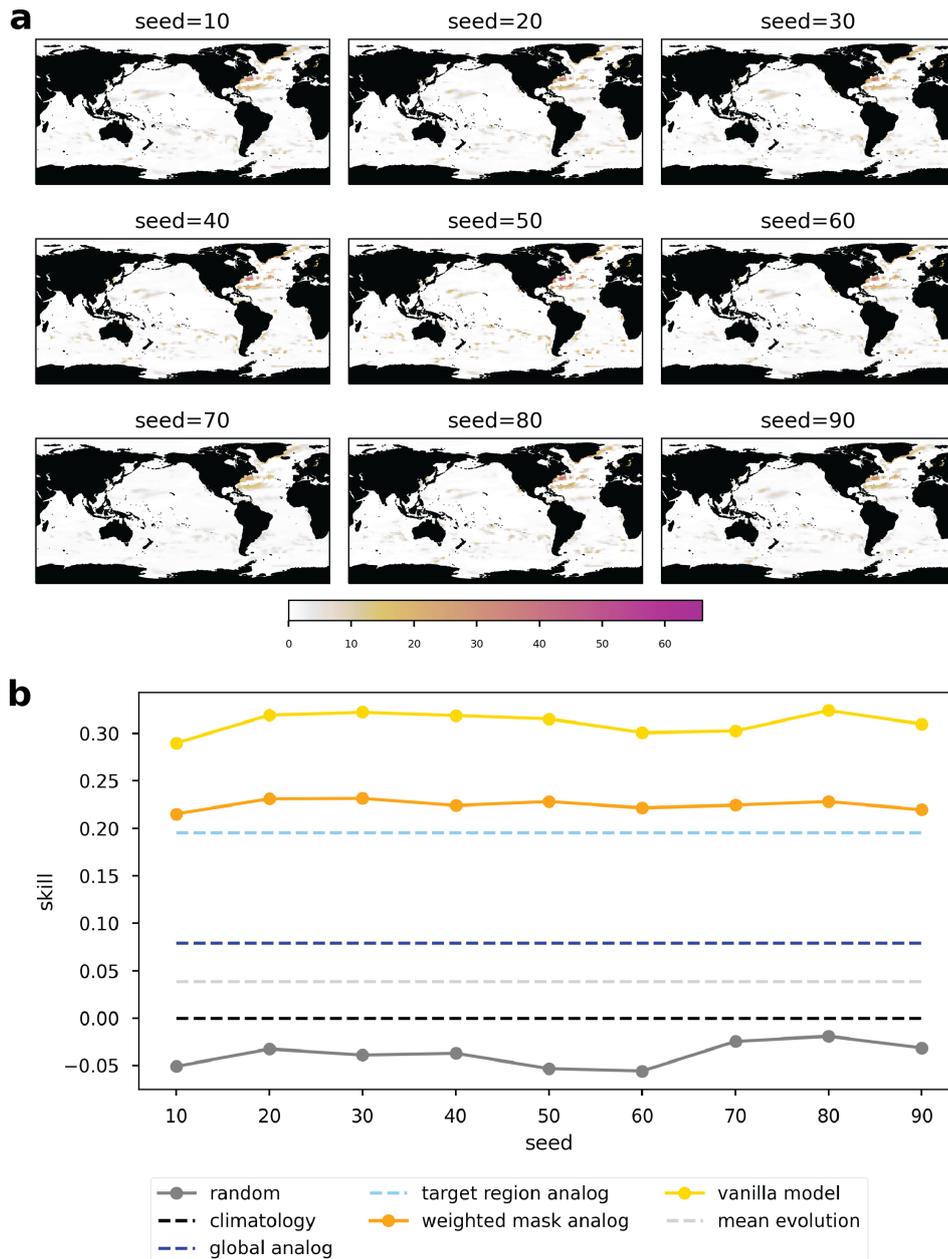


Figure S2. EXP-NorAtl: results for neural networks trained on nine different seeds. (a) Nine neural networks trained on different seeds show striking consistency in their weighted masks. (b) Skill scores for the average of the top-10 analogs. In all cases, the highest skill comes from the vanilla model, followed by the analog models. The masked analog outperforms the baselines discussed in the main text.

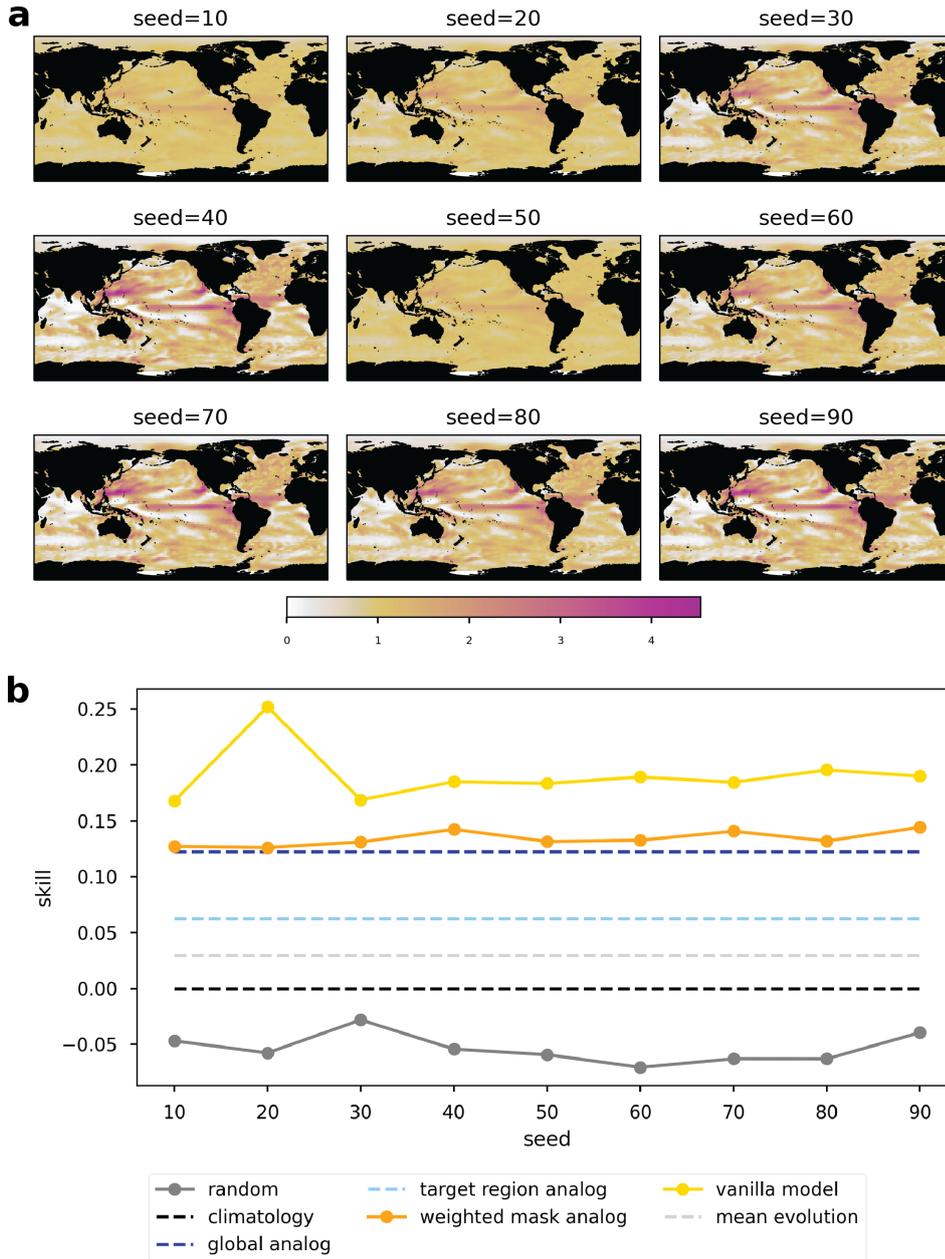


Figure S3. EXP-Niño: results for neural networks trained on nine different seeds. (a) Changing the seed used for the neural network training results in slight variation in the weighted mask. However, all weighted masks highlight the central tropical Pacific, western Pacific, Baja coast, and central Atlantic 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.

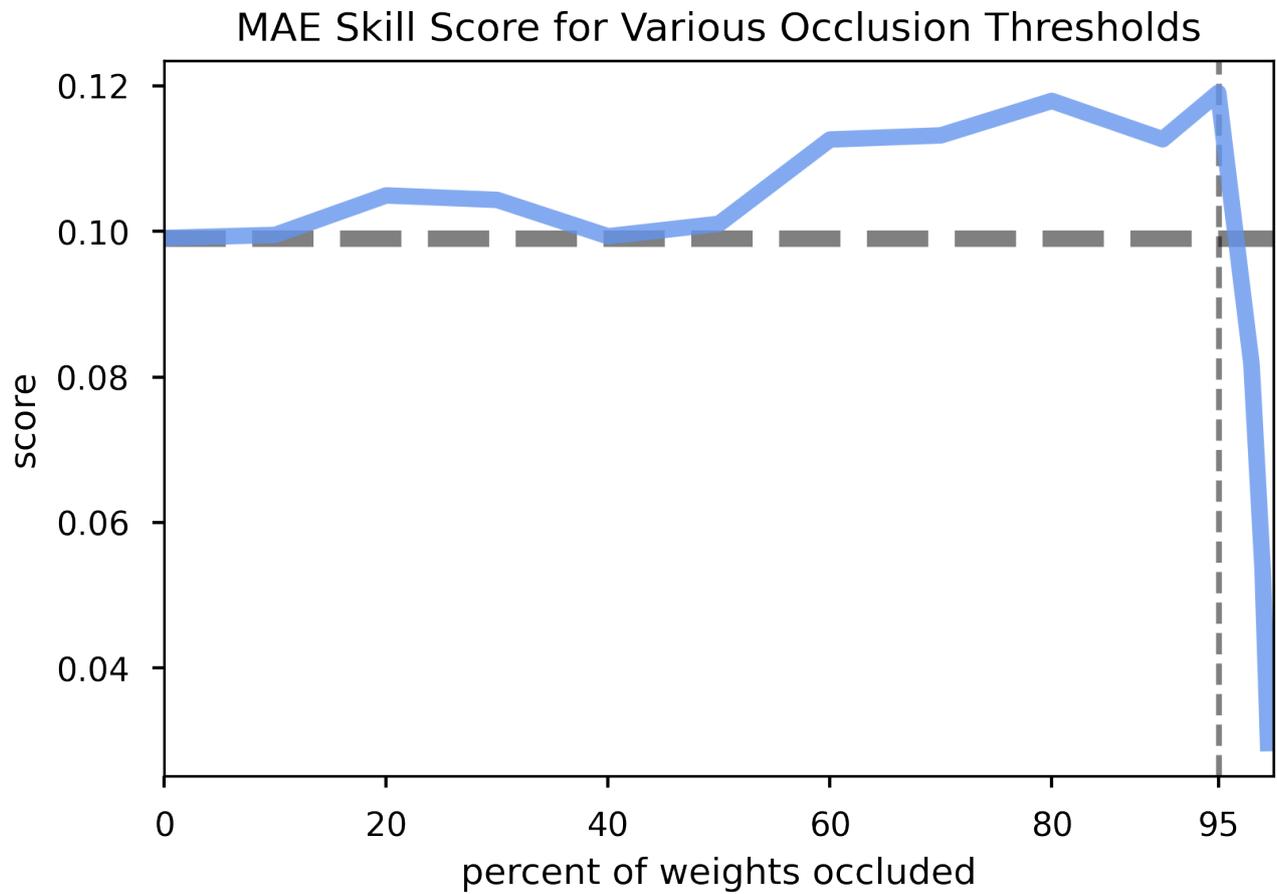


Figure S4. The skill of an analog forecast for EXP-Niño using the weighted mask when the smallest weights are set to zero. The horizontal line indicates the forecasting skill before the weighted mask has been altered. The vertical line indicates the forecasting skill using a weighted mask where the smallest 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.

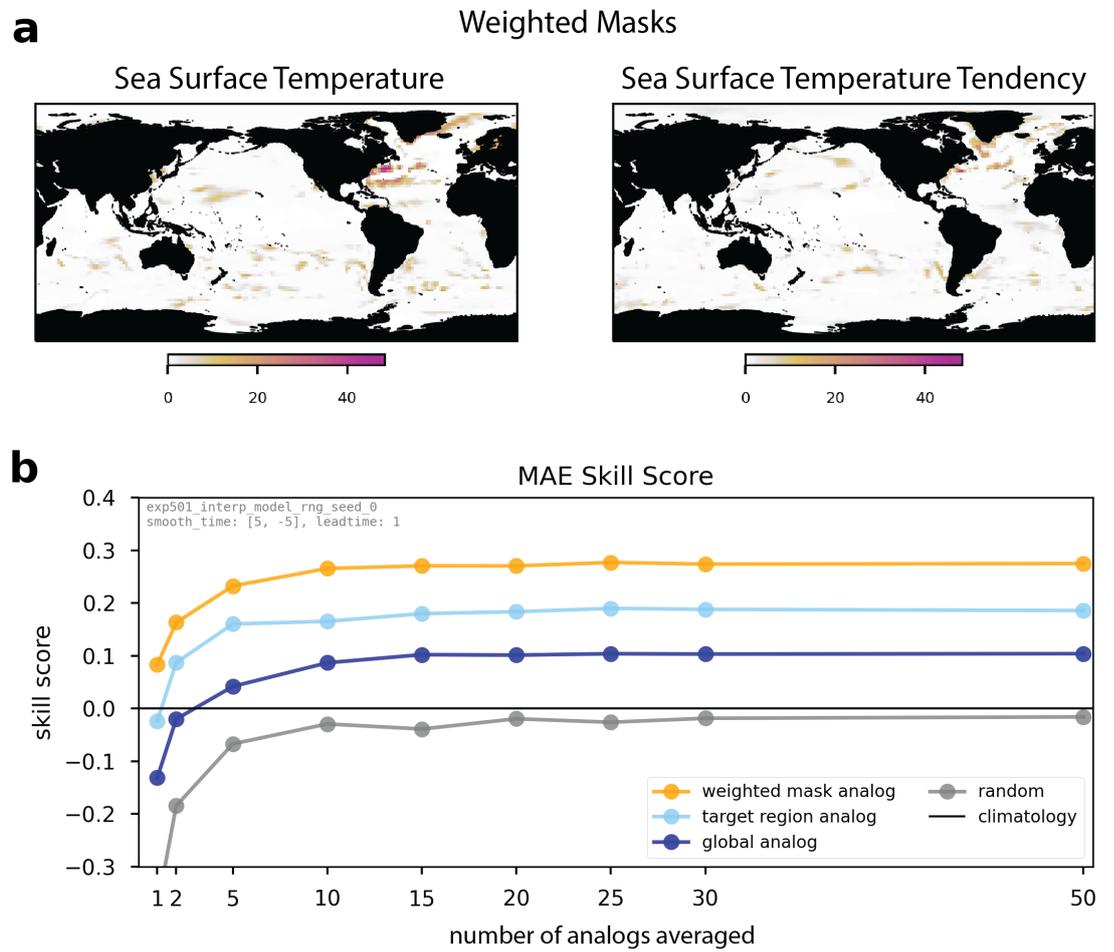


Figure S5. (a) Weighted masks for EXP501. (b) Skill scores for EXP501 versus various baselines. Adding an SST tendency input field did not notably impact forecasting skill in this problem.

| | |
|--|---------------|
| 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 S1. 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 S2. 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 S3. 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 S4. 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 S5. 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 S6. 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 S4).