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

Jamin K. Rader[1], and Elizabeth A. Barnes[1]

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

Contents of this file

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.

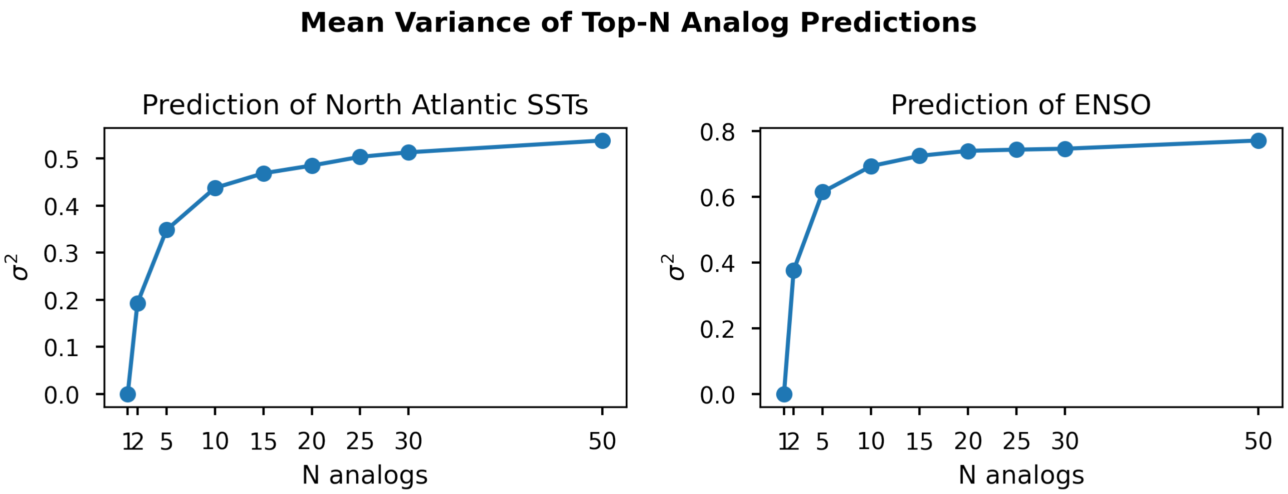

Figure S1. Mean variance of the targets associated with the top-N analogs across all testing samples.

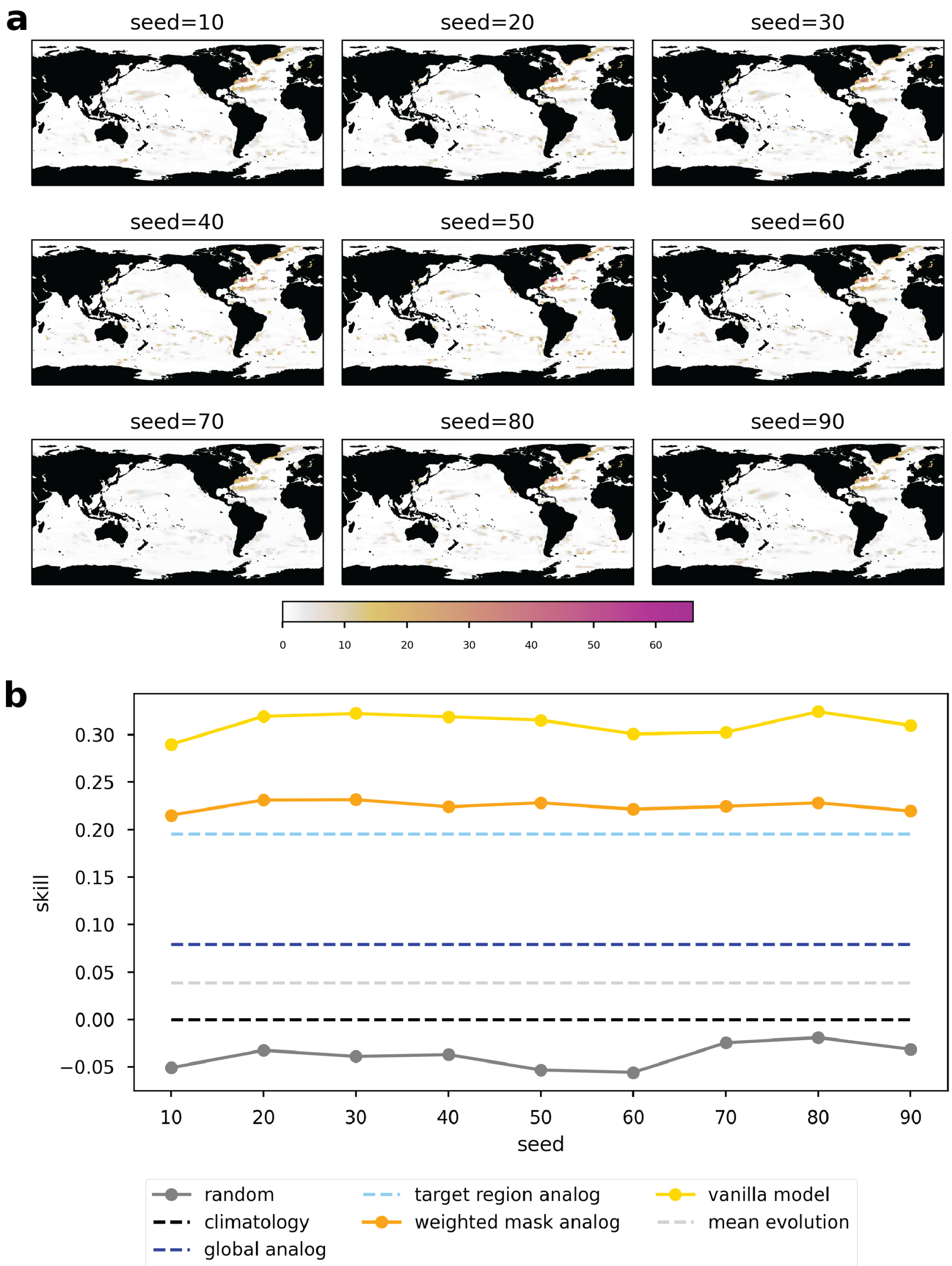

Figure S2. EXP-NorAtl: results for neural networks trained on nine different seeds. (a) Nine neural networks trained on different seeds show striking consistency in their weighted masks. (b) Skill scores for the average of the top-10 analogs. In all cases, the highest skill comes from the vanilla model,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

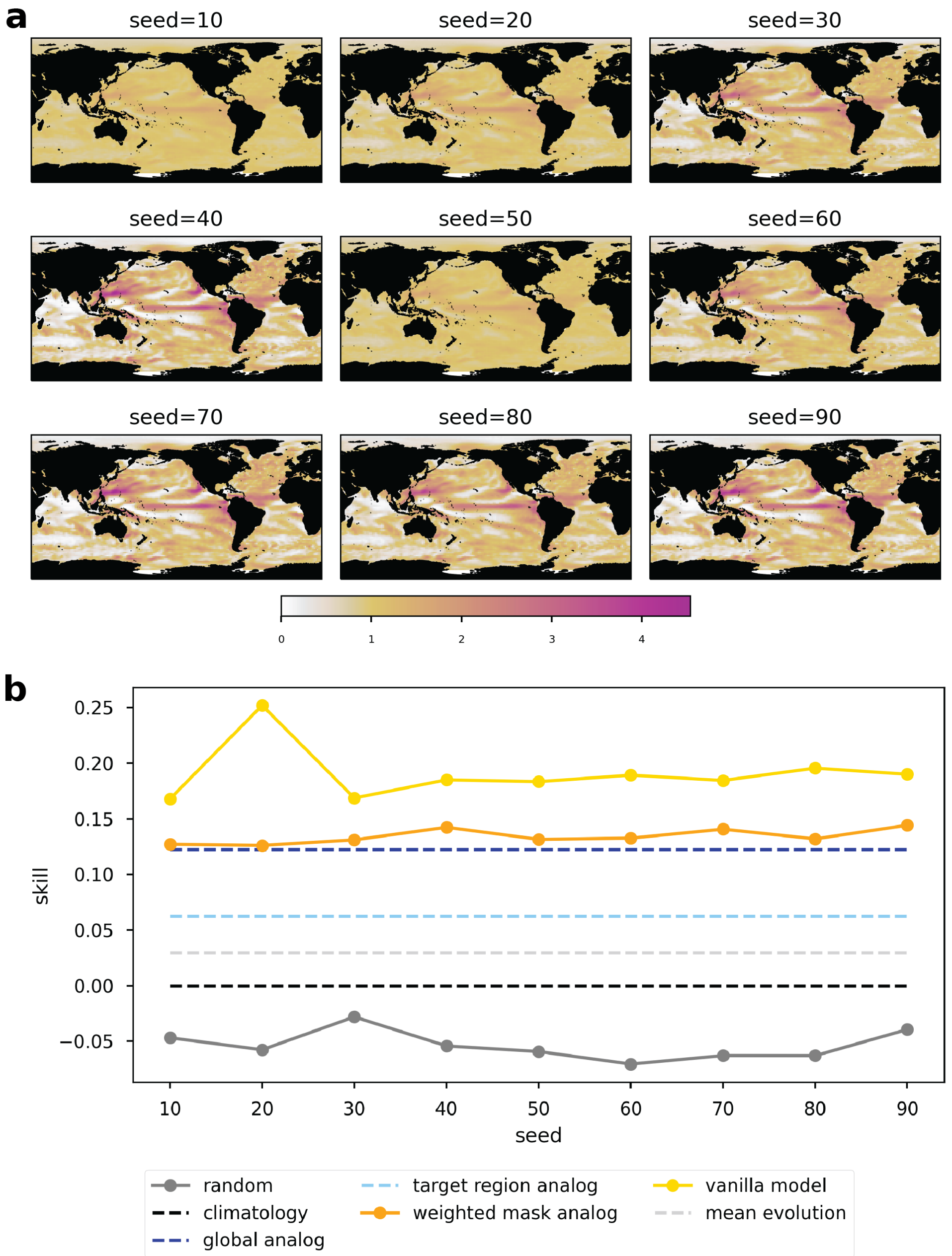

Figure S3. EXP-Niño: results for neural networks trained on nine different seeds. (a) Changing the seed used for the neural network training results in slight variation in the weighted mask. However, all weighted masks highlight the central tropical Pacific, western Pacific, Baja coast, and central Atlantic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

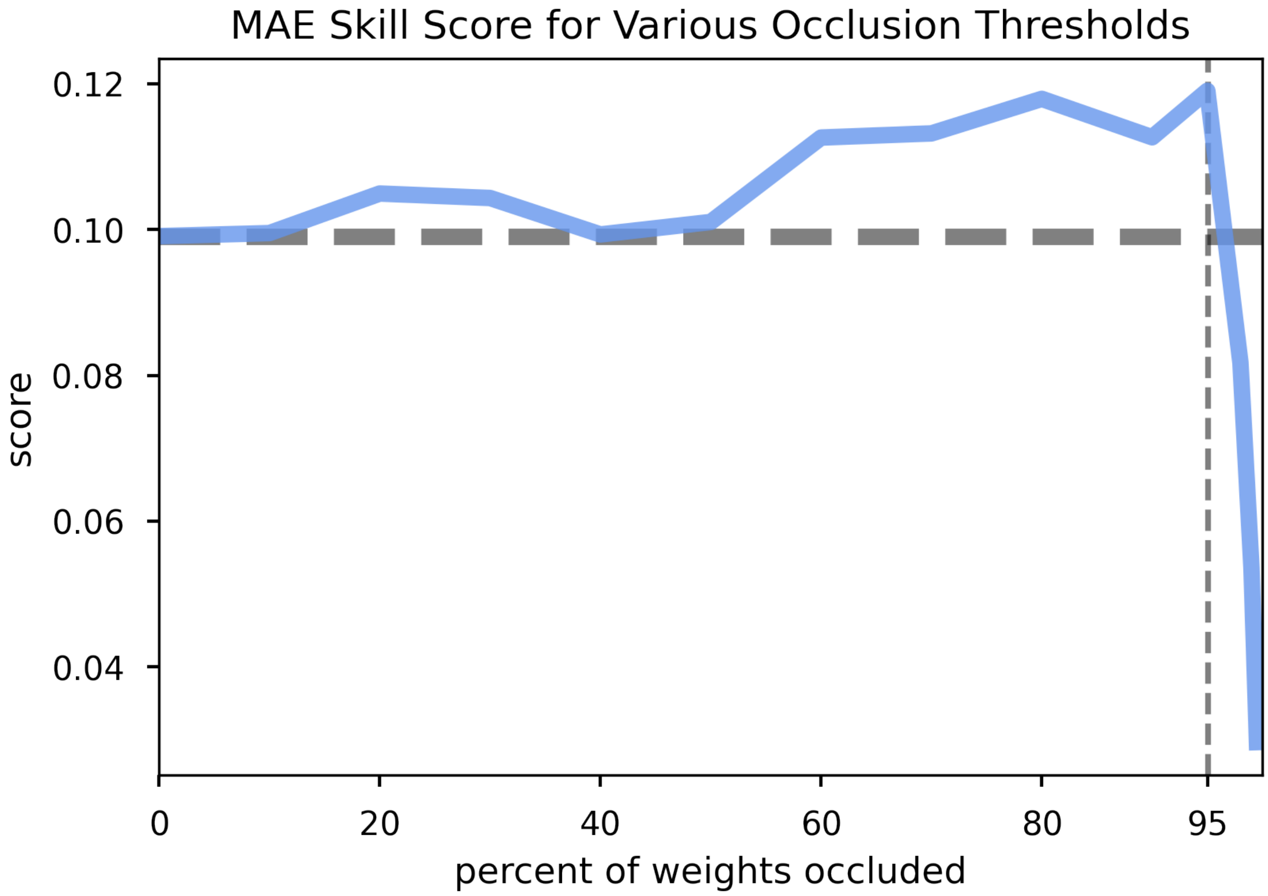

Figure S4. The skill of an analog forecast for EXP-Niño using the weighted mask when the smallest weights are set to zero. The horizontal line indicates the forecasting skill before the weighted mask has been altered. The vertical line indicates the forecasting skill using a weighted mask where the smallest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

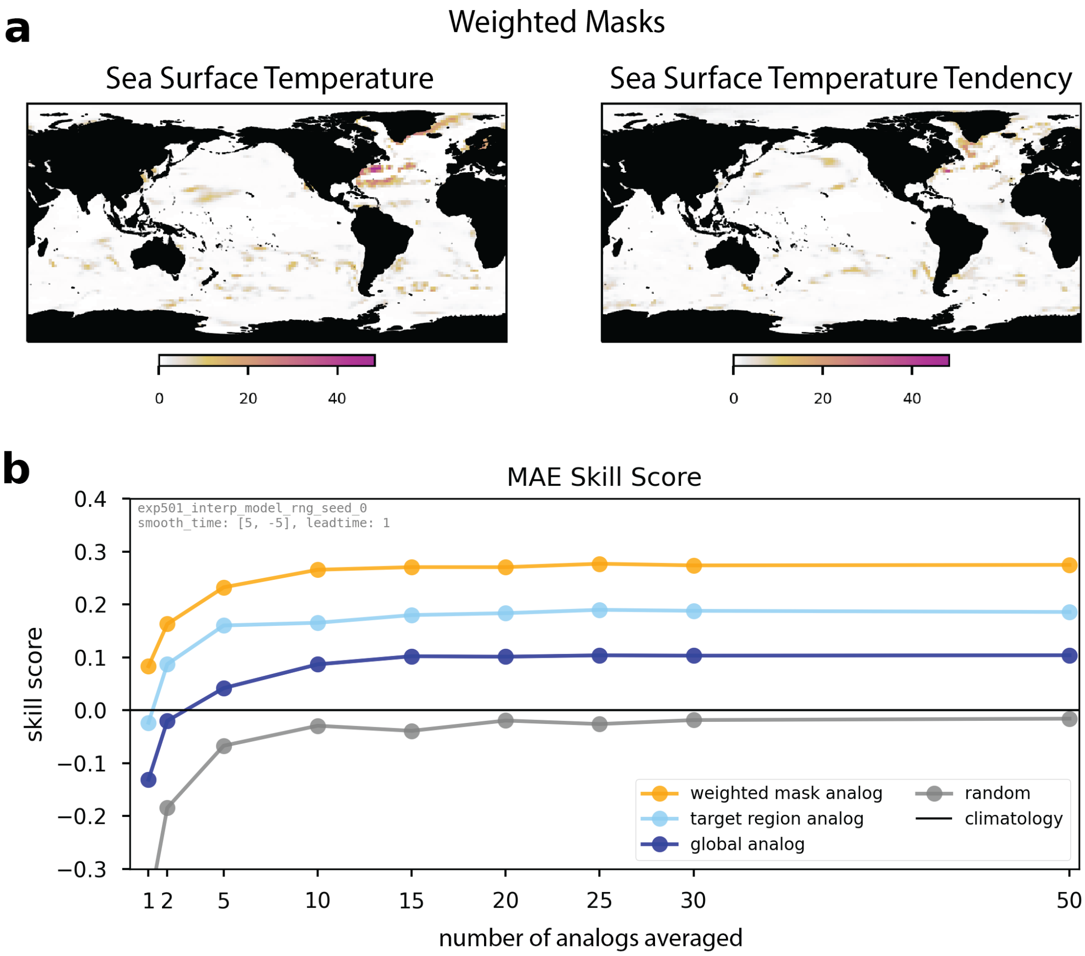

Figure S5. (a) Weighted masks for EXP501. (b) Skill scores for EXP501 versus various baselines. Adding an SST tendency input field did not notably impact forecasting skill in this problem.

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).