Supporting Information for "Oceanic harbingers of PDO predictability detected by neural networks"

E. M. Gordon[1], E. A. Barnes[1], J. W. Hurrell[1]

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

Contents of this file

Introduction

Here we provide a short overview of neural networks, along with the specifications of the artificial neural network (ANN) used in this study. We also describe the rationale behind the choice of a 30 month lead time followed by various statistics of the three ANNs used. Lastly we include supplementary figures to support our discussion and conclusions.

Text S1: Neural Network Overview

A general description of an artificial neural network (ANN) is thus: the neural network learns from some training data to map an input to some output, with hidden weights and connections optimized in the training process, and an activation function which allows for non-linearities. The network is trained for a set number of passes though the training data (called epochs), updating hidden weights based on minimizing the so-called loss function. The ANN architecture and training procedure in this study has been optimized for the specific problem that we consider. The use of regularization, dropout layers, training epoch and sample weights were carefully chosen to balance accuracy, but prevent over-fitting. Values used are included in Table S1. A more in-depth description of ANNs, as well as a broad background on their application to climate studies can be found in Toms, Barnes, and Ebert-Uphoff (2020).

Text S2: Rationale behind 30 month lead time Our ANN learns to predict whether a PDO phase transition will occur within some cut-off time. Consider an input such that by the time of the output, a transition has occurred (i.e. the true output is 1). If, for example, the lead time is 30 months and the transition occurred 29 months after the input, then this would be classified transition however it would be difficult for the ANN to

guess as it is similar to inputs where transitions occur at 31 months (which are classified persistence). The accuracy of the ANN dramatically decreases for samples where the transition occurs within around 3 months of the lead time. On the other hand, we want to focus on transitions occurring at least 12 months after input in order to benchmark our networks against previous work. Hence, in order to optimize for the accuracy of samples with transitions at least 12 months after input, retain good general accuracy, and a reasonable cut-off for recognizing persistence, we choose a lead time of 30 months (2.5 years).

Text S3: Summary of the 'best' neural networks

In order to find the best models for our problem setup we have trained 60 neural networks of the identical architecture, each with a different random seed. Note this seed is the same for both initializing the neural network and for choosing the transition samples to grab from the training/validation data. We train many models because we do not use all of the available data in the training process. This, along with the inherent randomness in the ANN training process can result in variation in the ANNs's accuracy. The random seed is set and recorded before the training/validation data is selected and the model is trained.

In Figure S1 we show various statistics of each individual neural network. The left panel compares the total accuracy of each ANN (x axis) with its persistence recall (percentage of the time that when persistence occurs, the ANN guesses persistence, y axis). This plot shows the difficulty in guessing persistence for this particular problem, with no ANNs above 56% recall. We comment on the reason for this in the main. As persistence appears

to be more difficult for the ANNs to learn, we designate the 'best' ANNs as those that combine high accuracy and high persistence recall. These are indicated in each plot by the pink dots.

The right panel demonstrates the ANNs's ability to predict transitions that occur 12-27 months after input, with total accuracy on the x axis and 12-27 month transition recall (percentage of the time that when a transition occurs 12-27 months after input, the ANN predicts the transition) on the y axis. This shows that the NNs we have designated as the 'best' (again in pink dots) have recall of 12-27 month transitions of around 62%-65%. While these are not the best ANNs for this task in particular, we choose them for this study as they are the best at *both* persistence and transitions, with their recall implying they have learned both, and are least likely to be over-fit.

In Figure S2 we show the confusion matrices for the best three ANNs described above. These demonstrate how the ANNs perform at the classification task on the validation data (1110 samples; 555 persistence, 555 transitions). Each row is the actual class the samples belong to, while the columns show how the ANN designated them, i.e. the top row are samples that are *true* persistence while the left column is the samples that were *predicted* as persistence. This means the main diagonal is where the ANN was correct and the off-diagonal is where the ANN was wrong. The number in each box is the number of samples placed in that category e.g. the top left box is number of samples with actual persistence *and* the ANN predicted persistence, whereas the bottom left is where an actual transition occurred but the ANN predicted persistence. In all cases, the ANNs were better

at correctly predicting transitions than persistence while the largest source of inaccuracy is due to the ANNs predicting transitions when the true class is persistence.

References

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002002. Retrieved 2021-05-04, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002002 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS002002) doi: https://doi.org/10.1029/2019MS002002

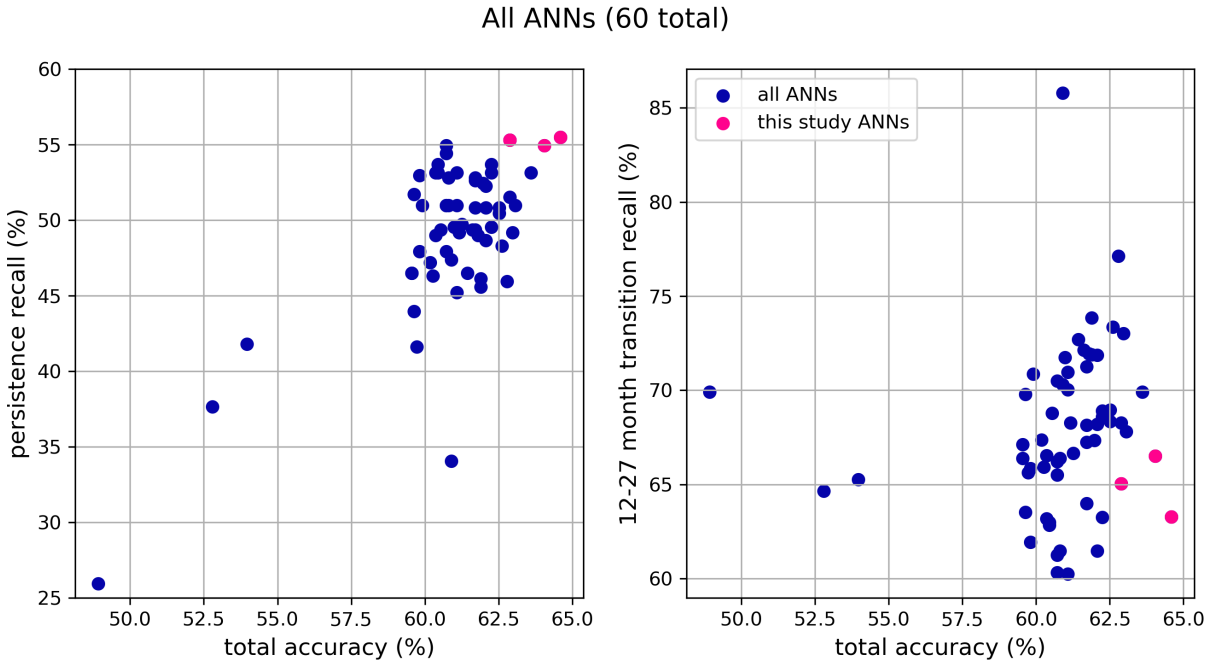

Figure S1. (left) Comparison of total accuracy (horizontal) and persistence recall (vertical) for all ANNs trained. Blue dots are all ANNs with pink dots representing the ANNs used in the study. (right) Comparison of total accuracy (horizontal) and 12-27 month transition recall.

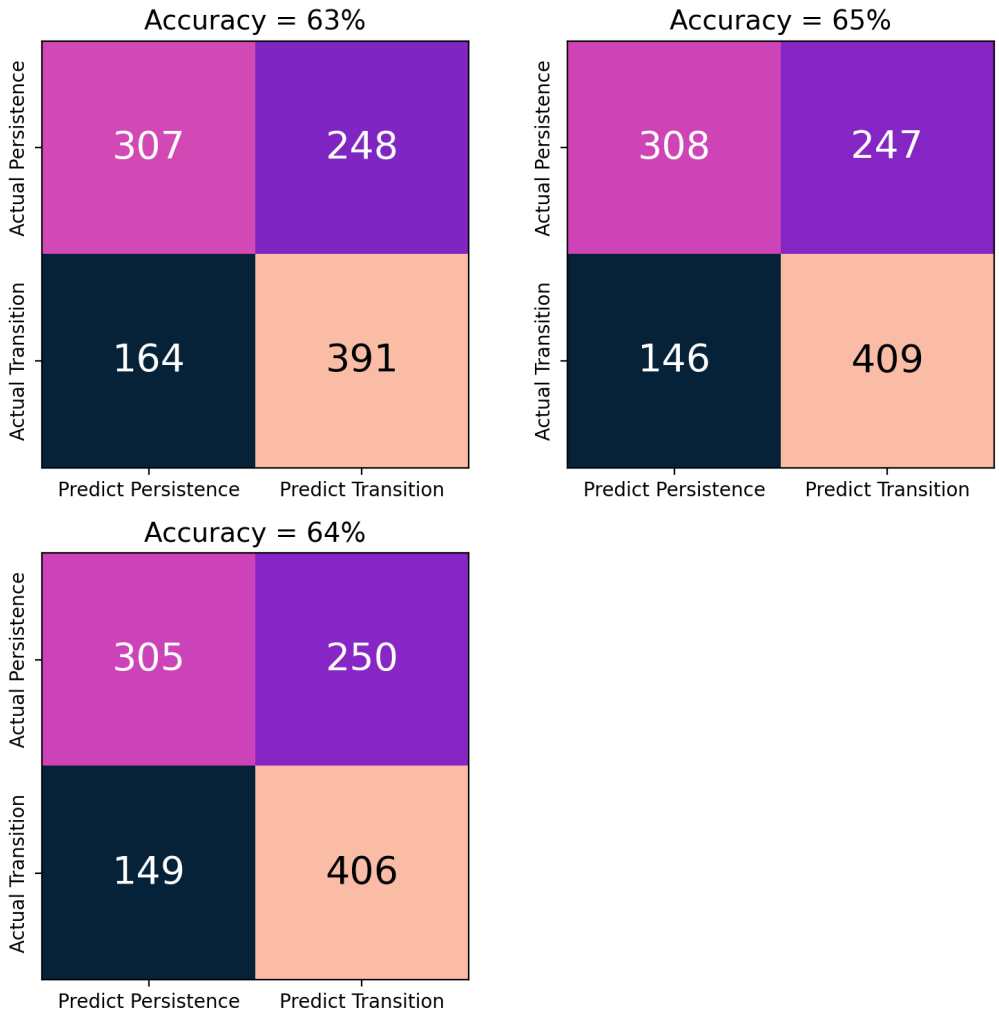

Figure S2. Confusion matrices for the 3 models used in this study. Vertical axis is the actual class and horizontal axis is the predicted class. Number of samples in each bin is printed in each square and total accuracy of each ANN in the title.

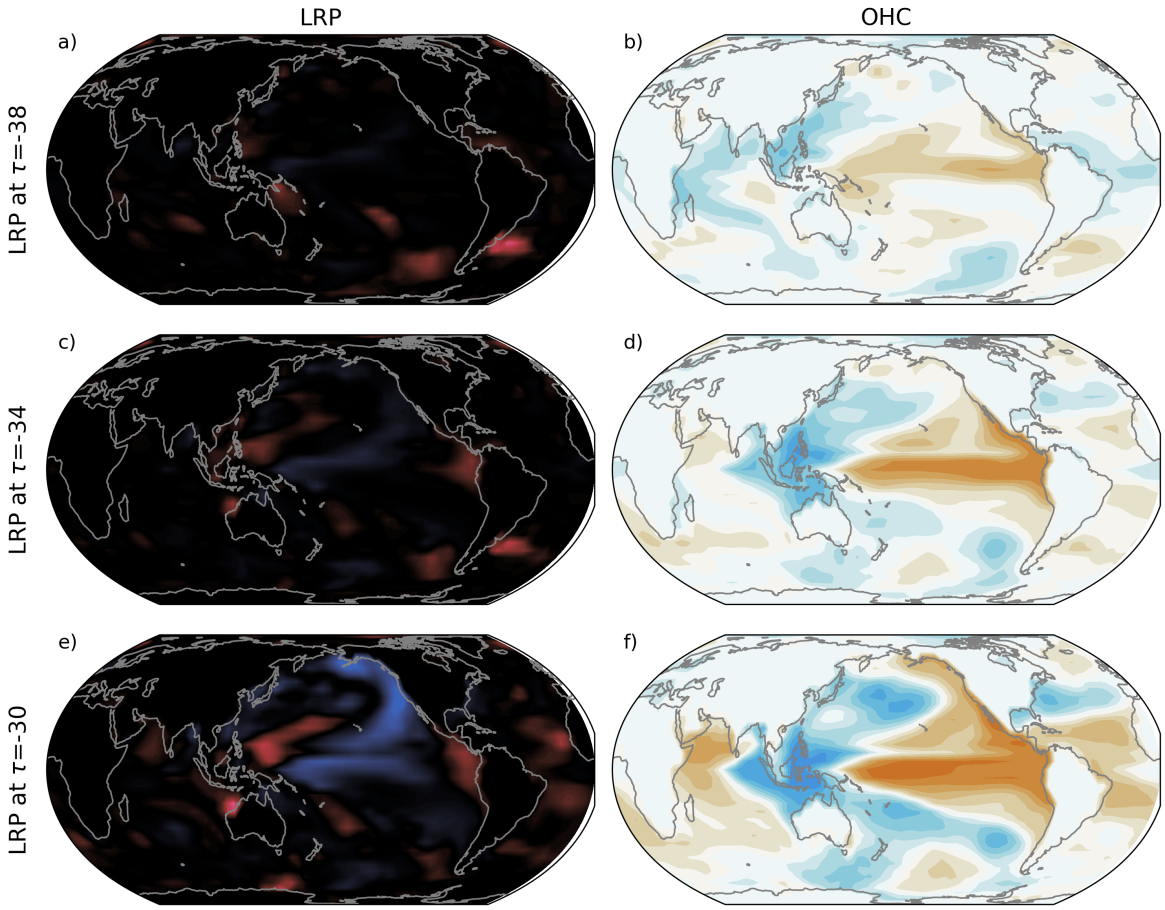

Figure S3. Left column: composite LRP maps for input maps where model correctly guesses transition from positive to negative occurs 12-27 months after final input. a) 38 months before output, c) 34 months before output, e) 30 months before output (and panel a in Figure 3). Color scale as in Figure 3. Right column: As left column but for composite OHC anomaly, with units of standard deviation at each grid point and color scale as in Figure 3.

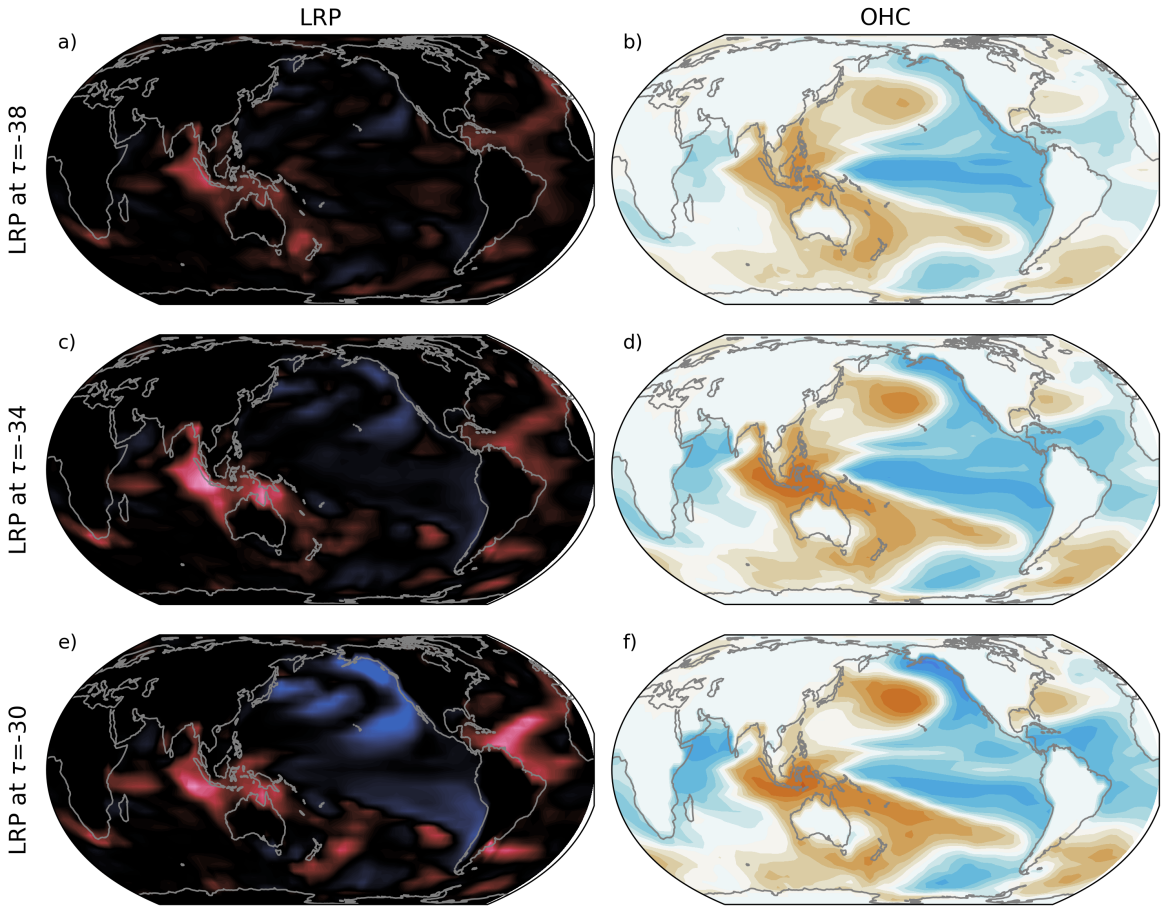

Figure S4. As Figure S4 but for negative to positive transitions.

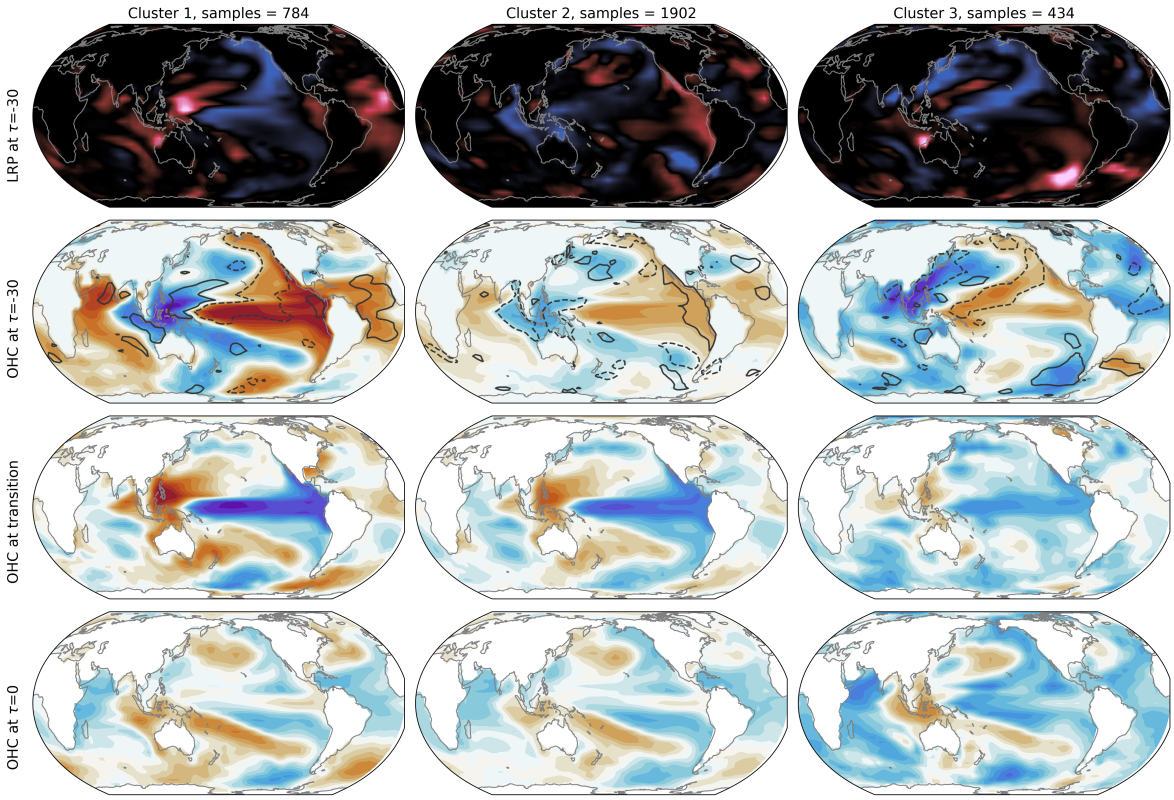

Figure S5. K-means of LRP maps when model correctly predicts positive to negative transition 12-27 months after input. Each column represents a cluster. Color scale as in Figure 3. Top row is LRP maps at month $\tau = -30$, second row is corresponding OHC with top and bottom 5% from the LRP contoured (dashed and dotted respectively as in Figure 3). The third row is OHC at the transition while the bottom row is OHC at month $\tau = 0$.

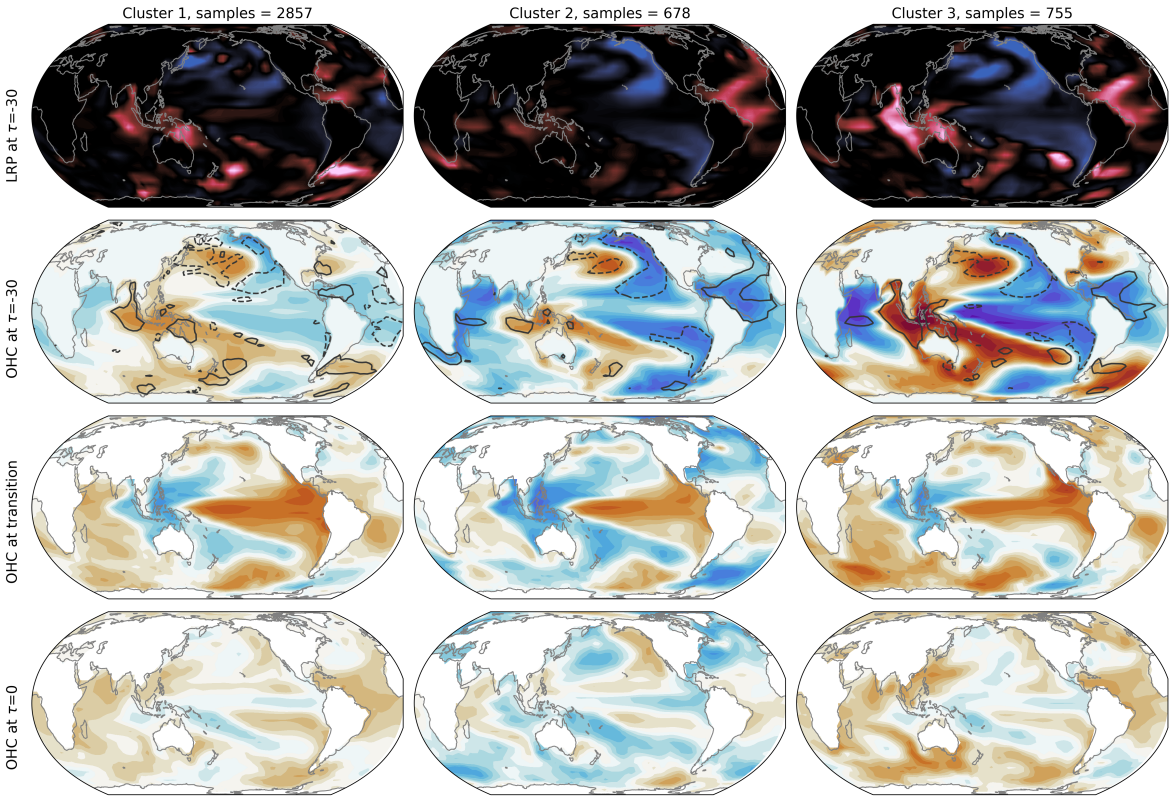

Figure S6. As Figure S5 but for negative to positive transitions

Table S1. Table of neural network specifications and accuracy for the ANNs used in this study.

Input	3 deseasoned and standardized $4° \times 4°$ OHC grids, 4 months apart
Architecture	3 vectorized OHC grids (12150 pixels total) connected to a single hidden layer with 8 nodes and rectified linear unit (ReLU) activation function, then connected to 2 output nodes representing positive and negative phase prediction with softmax activation to normalize outputs to probabilities.
Training	L2 regularization coefficient of 12 and dropout of one node per epoch on hidden layer. Adam optimization algorithm, with initial learning rate of 10^{-3}, dropping by a factor of 2 every 25 epochs. Trained for 300 epochs total. Categorical cross entropy loss function. First 1800 years (21600 samples) used for training, latter 200 years (2400 samples) used for validation (see main).
Output	Prediction of whether PDO transition occurs within 30 months of last input map.