Modeling the El Niño Southern Oscillation with Neural Differential Equations

L.T. Giorgini [1] **S.H. Lim** [1] **W. Moon** [1 2] **N. Chen** [3] **J.S. Wettlaufer** [4]

Abstract

We use a Neural Ordinary Differential Equation to model and predict the seasonal to interannual variability of El Niño Southern Oscillation (ENSO). We train our neural network model using partial observations involving only sea surface temperature data. Our approach is computationally inexpensive, it reproduces the main seasonal features of ENSO, and exhibits robust predictions skills.

1. Introduction

El Niño Southern Oscillation (ENSO) is the largest interannual variability phenomenon in the tropical Pacific and its influence extends beyond the tropics to higher latitudes via atmospheric and oceanic teleconnections. Therefore, it has significant impacts on global climate predictions. However, because of the intricate interplay between the stochastic forcing from atmospheric transients and the nonlinear air-sea interactions that drive ENSO phenomena (1; 2), building models that are capable of skillful ENSO forecasts remains a major challenge in climate science.

Here we use a Neural Ordinary Differential Equation (NODE) (3) approach to model ENSO phenomena. Our NODE model successfully captures the essential features of ENSO, the two most important features of which are (i) "phase locking" to the seasonal cycle (4), which is the concentration of abnormal ENSO events during November and December, and (ii) the predictability barrier (5), which is related to the increase of model uncertainties during spring. Finally, our model exhibits robust short-term prediction skills, outperforming more complex and computationally expensive models on the same tasks.

[1]Nordita, Royal Institute of Technology and Stockholm University - Stockholm 106 91, Sweden [2]Department of Mathematics, Stockholm University - Stockholm 106 91, Sweden [3]Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA [4]Yale University, New Haven, Connecticut 06520, USA. Correspondence to: Ludovico Theo Giorgini < ludovico.giorgini@su.se >.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

2. Data

We assume that only a single time series, the Nino3.4 sea surface temperature (SST) index data, is available for the entire training procedure of the neural network (NN). The data consists of monthly SST anomalies [1] derived from the temperature analyses of the Bureau National Operations Centre (BNOC) at the Australian Bureau of Meteorology (10). These data are measurement values averaged over the Nino3.4 region (5N-5S, 170W-120W), starting from the pre-industrial era and spanning the entire period 1870-2016.

3. Methods

We assume that the temporal evolution of ENSO can be represented by an autonomous continuous-time dynamical system of the form

$$\dot{x}(t) = F(x(t)), \tag{1}$$

where $x : [0, T] \to R^D$, $T > 0$ is the state of the system and $F : R^D \to R^D$ a deterministic force field.

It was shown in (6) that ENSO exhibits chaotic behavior and evolves inside a strange attractor (7). We denote $\nu \leq D$ as the box counting dimension of such attractor A. Because of Whitney's embedding theorem (8), A can be embedded in an Euclidean space with dimension $m > 2\nu$. The diffeomorphism ϕ that maps $A \to R^m$ can be chosen, according to the Takens theorem (9), as

$$\phi(x(t)) = [x_\alpha(t), x_\alpha(t-\Delta\tau), ..., x_\alpha(t-(m-1)\Delta\tau)] \tag{2}$$

for every $\alpha \in \{1, \dots, D\}$ This means that for each α we can construct a function $G^\alpha : R^m \to R^D$ such that

$$\dot{x}_\beta = G^\alpha(x_\alpha(t), x_\alpha(t-\Delta\tau), ..., x_\alpha(t-(m-1)\Delta\tau))_\beta \tag{3}$$

[1]The SST anomalies are defined as the difference between the monthly sea surface temperature data and the average monthly temperature. Let $T(i)$ be the temperature relative to the i-th month, with $i \in 12N$ and N the total number of years in which the sea surface temperature data are available. The average temperature corresponding to each month is then

$$\bar{T}_m = \frac{1}{N}\sum_{i=1}^{N} T_{m+12i} \quad m \in [1, 12],$$

and hence we define the SST anomalies as $x_i = T_i - \bar{T}_{i \bmod [12]+1}$.

for every $\alpha, \beta \in \{1, \ldots, D\}$. Using Eq. (3) for $\alpha = \beta$, we can write the temporal evolution of this process $y : [0, T] \to R$ as

$$\dot{y}(t) = G(y(t), y(t - \Delta\tau), ..., y(t - (m-1)\Delta\tau)), \quad (4)$$

where we have dropped the indices.

We represent the mapping G as a neural network parameterized by a learnable vector θ. Because the optimal choice of the embedding dimension m and the time interval $\Delta\tau$ are unknown, we take them to be learnable parameters. Given an observed discrete time series $y = (y^{obs}(\Delta t), ...y^{obs}(N\Delta t))$ of length N, where $\Delta t = \frac{1}{q}\Delta\tau$ with $q \in N$, our aim is to find $\hat{\theta}(n)$, defined as

$$\hat{\theta}(n) = \arg\min_{\theta} \mathcal{L}^n(y^{obs}, y_\theta^{pred}), \quad (5)$$

where

$$\begin{aligned} &\mathcal{L}^n(y^{obs}, y_\theta^{pred}) \\ &= \left[\frac{1}{N - n - (m-1)q} \sum_{i=n+(m-1)q}^{N} \right. \\ &\left. \left(y^{obs}(i\Delta t) - \int_{(i-n)\Delta t}^{i\Delta t} ds\, G_\theta\left(s; y_0(j\Delta t)\right)\right)^2\right]^{\frac{1}{2}}, \end{aligned} \quad (6)$$

is the loss function measuring the discrepancy between the observed time series and model predictions, which are obtained by numerically integrating Eq. (4) for n time steps. In the above, $y_0(j\Delta t) := (y^{obs}(j\Delta t), y^{obs}(j\Delta t - \Delta\tau), ..., y^{obs}(j\Delta t - (m-1)\Delta\tau)$, and we use the forward Euler scheme for the numerical integration.

4. Results

We consider three training settings corresponding to minimizing the loss function $\mathcal{L}^n$, with $n = 20, 40, 48$. Each of these setting corresponds to fixing a timescale where the parameters θ are optimized.

For each setting, we train our NN, varying the values of the embedding dimension m and time interval $\Delta\tau$. We measure the predictive performance of the model in terms of the pattern correlation and the root mean square error, normalized with respect to the standard deviation of the SST anomalies and subtracted to one (we shall call this quantity RMSE), between predictions and observations. We repeat this procedure using two different values of the hyperparameter n: $n = 20$ and $n = 40$ (see Fig. (1)). The goal of this test is to tune the hyperparameter n to improve the predictive performance of the algorithm and gain insight into the dynamical characteristics of ENSO. Note that the embedding dimension m provides an upper bound for the dimension ν of the underlying system ($\nu < \frac{m}{2}$). Fig. (1) shows that the optimal embedding dimension is relatively low (less than 5 for the case of $n = 20$ and less than 7 for the case of $n = 40$). This shows that ENSO is effectively a low dimensional system (see e.g. (11) and references therein). For a given n, the optimal value for the time delay as a function of the embedding dimension reveals the dominant time scale for reliable predictions.

Fig. (2) shows the correlation and RMSE between observations and predictions, in terms of the lead and starting months. For comparison, we consider three different models (n=20, 40 and 48), shown in the first, second and third rows respectively. All three models exhibit the characteristic features of ENSO described above. First, the correlation exhibits an abrupt reduction for predictions starting during the boreal spring due to the observed lowering of the signal to noise ratio in the boreal summer. Second, the RMSE exhibits maximal values for predictions in the boreal winter, during which the signal to noise ratio reaches its maximal value.

In the first column of Fig. (3) we show the correlation and the RMSE as a function of the delay between predictions and observations. The second column of Fig. (3) shows the predictions for different values of the time delay. For the first three panels we use the parameters trained with $n = 20$, and for the last panel we use the parameters trained with $n = 40$. We compare the predictive performance of our model to that obtained using kernel analog forecasting (KAF) (13), the linear inverse model (LIM) (12), the General Circulation Models (GCMs) and the convolutional neural network (CNN) model of (14).

In panel (a) of Fig. (3) we show the correlation and RMSE as a function of time delay for the three NNs trained using $n = 20$, $n = 40$ and $n = 48$, corresponding to a training period of 6, 10 and 12 months respectively. If we use each of these models for predictions, included in the time range where the NN is trained, we see clearly that our models considerably outperform the LIM model and increase the time horizon for skillful forecasts by approximately 5 months. This is represented by the correlation threshold of 0.5. Our models significantly outperform the KAF model for predictions of less than 6 months. The KAF model exhibits a performance similar to that of the LIM; for longer-term predictions the improvement is only marginal.

The NN model trained with $n = 20$ outperforms much more advanced and computationally expensive models, such as the GCMs and the CNN of (14), for predictions 6 months into the future (see Fig.(3)(b)). For longer-term predictions, which means prediction beyond the spring predictability barrier, the performance of our model degrades noticeably when compared with that of the other two models. This is due the fact that our NNs are trained using only the data from the Nino3.4 region instead of the data from the entire Pacific

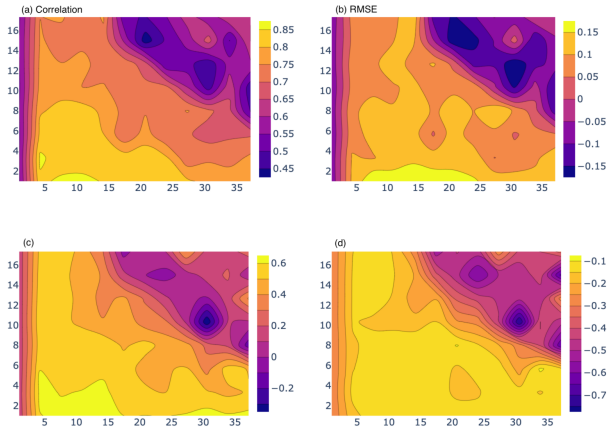

Figure 1. Correlation and RMSE between predictions and observations for different values of the embedding dimension and time interval for two values of the hyperparameter; $n = 20$ for panels (a,b) and $n = 40$ for panels (c,d).

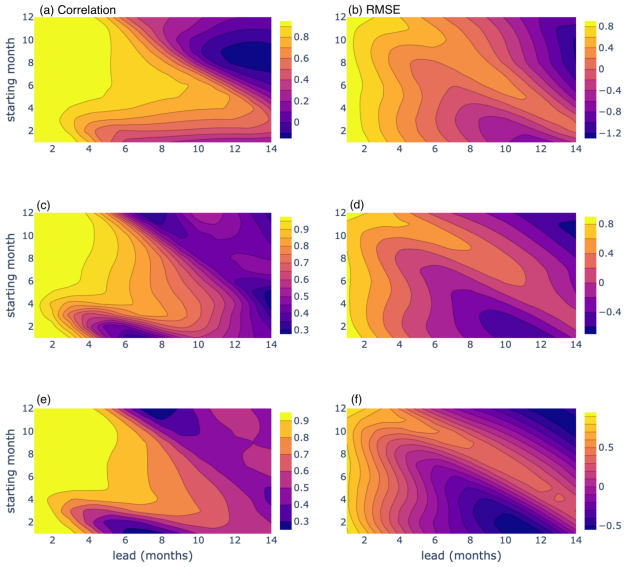

Figure 2. Pattern correlation and RMSE in function of the lead (expressed in months) for different starting months. Three different NN have been used for the comparison, $n = 20$ for panels (a,b), $n = 40$ for panels (c,d) and $n = 48$ for panels (e,f).

as used by the other models. The lack of the more extensive data limits the ability of our NNs to learn correlation patterns spanning broad regions of the Pacific, thereby significantly impacting prediction on long timescales.

5. Conclusions

In this paper we have developed a Neural Ordinary Differential Equation (NODE) framework to accurately model the El Niño Southern Oscillation (ENSO). By learning the optimal embedding dimension and the associated time delay in our approach, we have reproduced the key seasonal features of ENSO. When compared to other methods, our method is computationally less costly and yet exhibits robust short-term predictive performance. Our work demonstrates the promising potential of using the NODE framework for climate modeling.

References

[1] R. Kleeman, and A. M. Moore, J. Atmos. Sci. **54**, 753 (1997).

[2] S. I. An, and F. F. Jin, J. Clim. **17**, 2399 (2004).

[3] Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. (2018). Neural ordinary differential equations. arXiv preprint arXiv:1806.07366.

[4] E. Tziperman, M. A. Cane, S. E. Zebiak, Y. Xue, and B. Blumenthal, J. Clim. **11**, 2191 (1998).

[5] P.J. Webster, and S. Yang, Quart. J. Roy. Met. Soc. **118**, 877 (1992).

[6] E. Tziperman, et al., Science **264** 72 (1994).

[7] T. Bohr, P. Bak, M. H. Jensen Phys. Rev. A. **30** 1970 (1984).

[8] H. Whitney, Ann. Math., 220 (1944).

[9] F. Takens, Dyn. syst. turb., Springer, Berlin, Heidelberg, 366 1981.

[10] BNOC (Bureau National Operations Centre) Operations Bulletin Number 120 "Operational Implementation of ACCESS-S1 Seasonal Prediction System (Dec 2018)".

[11] L.T. Giorgini, W. Moon, N. Chen and J.S. Wettlaufer, arXiv: 2012.09858 (2021).

[12] C. Penland, and T. Magorian, J. Climate **6**, 1067 (1993).

[13] X. Wang, J. Slawinska, and D. Giannakis, Sci. rep. **10**, 1 (2020).

[14] Y. Ham, K. Jeong-Hwan, and L. Jing-Jia, Nature **573**, 568 (2019).

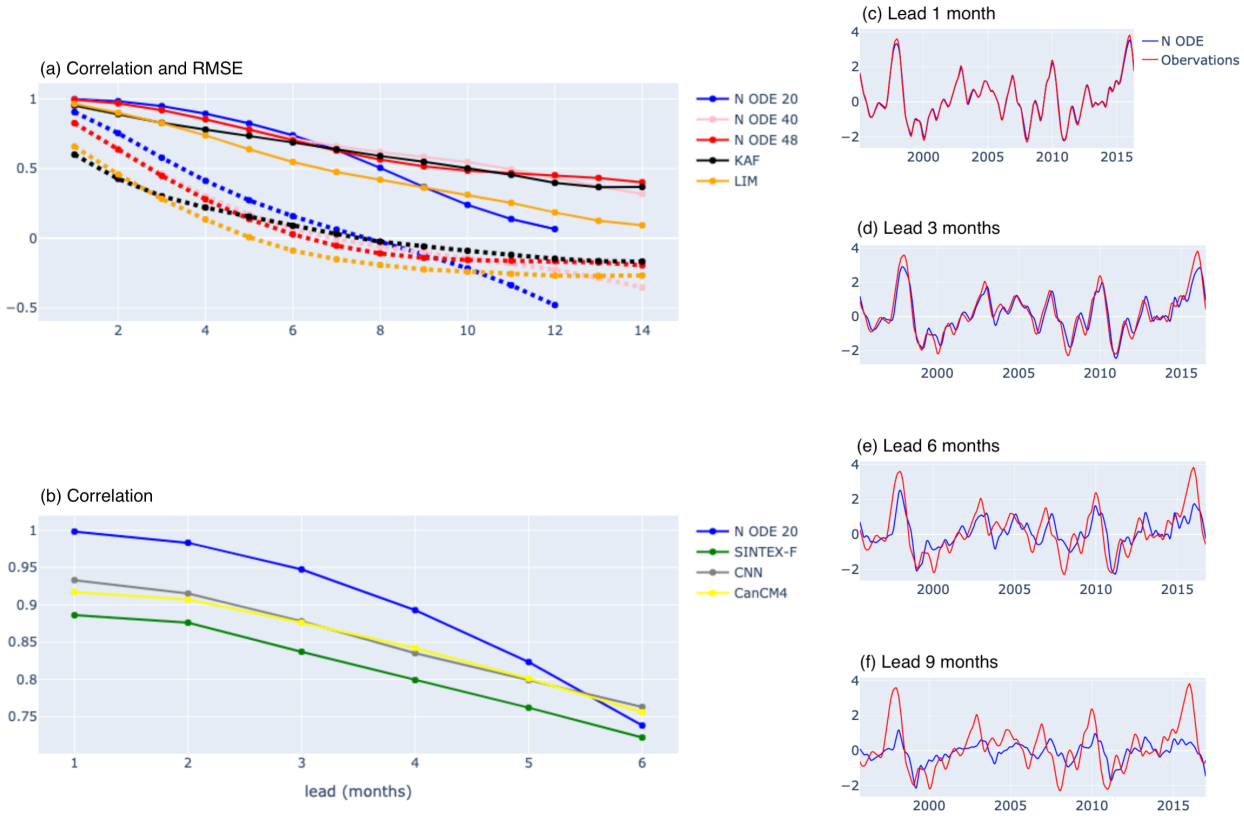

Figure 3. **Left column:** prediction performances of our Neural ODE (NODE 20,40,48, corresponding to $n = 20, 40$ and 48) compared to predictions using kernel methods (panel a) and different GCMs and a CNN model (panel b). The correlation and RMSE have been obtained using 1995-2016 as the verification period in the first figure and 1985-2016 for the correlation in the second figure. **Right column:** comparison of the predicted time series of the Nino3.4 SST index with observations.