Data-driven surrogate models for climate modeling: application of echo state networks to the multi-scale Lorenz system as a test case

Anonymous Authors[1]

Abstract

Understanding the effects of climate change relies on physics driven computationally expensive climate models which are still imperfect owing to ineffective subgrid scale parametrization. An effective way to treat these ineffective parametrization of largely uncertain subgrid scale processes are data-driven surrogate models with machine learning techniques. These surrogate models train on observational data capturing either the embeddings of their (subgrid scale processes') underlying dynamics on the large scale processes or to simulate the subgrid processes accurately to be fed into the large scale processes. In this paper an extended version of the Lorenz 96 system is studied, which consists of three equations for a set of slow, intermediate, and fast variables, providing a fitting prototype for multi-scale, spatio-temporal chaos, and in particular, the complex dynamics of the climate system. In this work, we have built a data-driven model based on echo state networks (ESN) aimed, specifically at climate modeling. This model can predict the spatio-temporal chaotic evolution of the Lorenz system for several Lyapunov timescales. We show that the ESN model outperforms, in terms of the prediction horizon, a deep learning technique based on recurrent neural network (RNN) with long short-term memory (LSTM) and an artificial neural network by factors between 3 and 10. The results suggest that ESN has the potential for being a powerful method for surrogate modeling and data-driven prediction for problems of interest to the climate community.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1. Introduction

There has been a lot of effort put into deep learning in recent times to parameterize complex non-linear physical processes in climate models especially for clouds and deep convection (Rasp et al., 2018; Gentine et al., 2018) . These efforts arise from the ever demanding need of immense computational expense to resolve subgrid scale processes such as clouds, gravity waves, submesoscale eddies, sea ice etc., (Collins et al., 2006). Machine learning can provide a solution to this challenge by providing data-driven surrogate models of the subgrid-scale physics needing high resolution (Reichstein et al., 2019). In such cases either the surrogate model feeds the parametrized values into the climate model, or the effects of subgrid-scale processes on the large scale circulation in the climate model is learned through observation or simulation of the large scale data itself.

In this paper we propose a data-driven solution to the multi-scale climate system which is represented by the 3-tier Lorenz 96 equations (Thornes et al., 2017). The proposed data-driven model (details in section 3) is restricted to train only on the low frequency large amplitude variable over a training period and expected to freely predict this variable in future time steps by learning the hidden embeddings of the intermediate and high frequency variables that are coupled with the large scale variable (see section 2) . This system exhibits fully chaotic multi-scale dynamics and serves as a fitting prototype to test a modified echo state network (Jaeger, 2001), which has been previously found to be promising in systems exhibiting chaotic dynamics (Pathak et al., 2018). We show that our network provides predictions up to multiple Lyapunov exponents (or earth-days) and outperforms state-of-the-art deep learning approaches such as RNN-LSTM and previously reported prediction horizon by fully connected neural networks (Dueben & Bauer, 2018).

2. Multi-Scale Lorenz 96 System

The 3-tier multi-scale Lorenz system is governed by the following non-linear ODEs:

$$\frac{dX_k}{dt} = X_{k-1}\left(X_{k+1} - X_{k-2}\right) + F - \frac{hc}{b}\Sigma_j Y_{j,k}$$

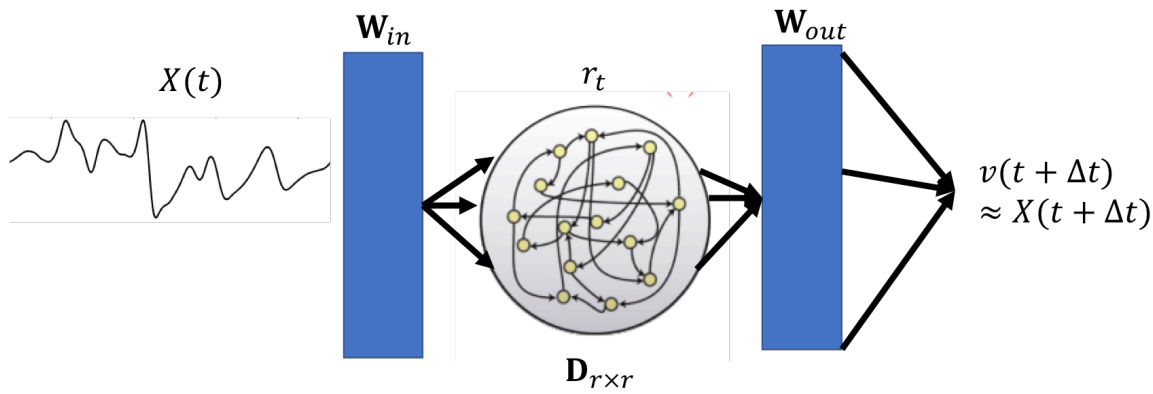

Figure 1. Schematic of the echo state network with sparse reservoir D_r, input layer of weights $\mathbf{W_{in}}$ and output layer of weights $\mathbf{W_{out}}$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - cY_{j,k} + \frac{hc}{b}X_k - \frac{he}{d}\Sigma_i Z_{i,j,k}$$

$$\frac{dZ_{i,j,k}}{dt} = -edZ_{i-1,j,k}\left(Z_{i+1,j,k} - Z_{i-2,j,k}\right) - geZ_{i,j,k} + \frac{he}{d}Y_{j,k}$$

In these equations, $F = 20$ is a large scale forcing term that makes the system highly chaotic; while $b = c = e = d = g = 10$ and $h = 1$ ensures a variation in scale of amplitude and frequency of the three variables. The vector field X has 8 elements while Y and Z have 64 and 512 elements respectively. This configuration ensures that this system has large amplitude and low variability in X analogous to the large scale circulation (mean flow) in the atmosphere while the Y and Z variables would represent the high frequency, low amplitude synoptic eddies and baroclinic eddies respectively. Here, we develop a data-driven surrogate model with an echo state network which can only train on the low frequency large scale variable, X , which is the easiest to observe and can learn the embedded dynamics of Y and Z affecting X. It then freely predicts the evolution of X in further time steps without the need of incorporating any knowledge of Y and Z in the model (although all three variables are coupled in the equations).

3. Echo State Network

Following the promising results shown by (Pathak et al., 2018) we develop an echo state network for the purpose of forecasting X in the Lorenz 96 system. The echo state network (Jaeger, 2001) is a recurrent neural network that consists of a large reservoir of size $D_r \times D_r$ with sparsely connected nodes modeled as an Erdős-Reyni graph whose adjacency matrix ($\mathbf{A}$) is drawn from an uniform random distribution between $[-1, 1]$ and has a state of $r(t) \in \Re^{D_r}$. The network takes in the input time series, $X(t)$ that is fed into the network via a layer of fixed weights $\mathbf{W_{in}}$ and is read out of the network through an output layer of $\mathbf{W_{out}}$. $\mathbf{W_{out}}$ is the only trainable layer in this network making the training process orders of magnitude faster than conventional RNNs that are trained via backpropagation through time. The equations governing the training process is as follows:

$$r(t+\delta t) = tanh\left(\mathbf{A}r(t) + \mathbf{W_{in}}X(t)\right)$$

$$\mathbf{W_{out}} = \arg\min_{\mathbf{W_{out}}} ||\mathbf{W_{out}}r(t) - X(t)||$$

$$v(t) = \mathbf{W_{out}}X(t)$$

$$v(t) = X(t)$$

$\mathbf{W_{out}}$ in our case is calculated by using the ridge regression algorithm while a basis function expansion is performed on the state matrix $\mathbf{r}$ by squaring every odd column (a quadratic non-linearity in the system makes this choice suitable for this problem).The relative simplicity of the echo state network makes it an ideal tool for surrogate modeling while outperforming RNN-LSTMs in terms of prediction horizon as shown in the next section.

4. Results and Discussions

In order to generate the data, the ODEs in section 2 is solved using a fourth-order Runge-Kutta solver with $dt = 0.005$ similar to (Dueben & Bauer, 2018). The ESN is trained on 10^5 examples of X each sampled at $dt = 0.005$. The optimal $\mathbf{W_{out}}$ is used to predict further in time while the state vector $r(t)$ is updated at each prediction time step. The results shown in both Figure 2 (A) and 3 conclude that the ESN has forecasting skills up to 12 Lyapunov exponents or 2.5 MTU (model time units) for an arbitrary initial condition

outperforming state-of-the-art LSTM networks (the LSTM network was constructed to predict $X(t+\delta t)$ from the previous k time steps where k was optimized with hyperparameter tuning) and previously reported neural networks (Dueben & Bauer, 2018) by twice as many MTUs. Figure 2 (B) shows the relative L_2 error ($||X(t) - X_{pred}(t)||/\langle||X(t)||\rangle$ where $\langle.\rangle$ denotes temporal averaging) averaged over 50 initial conditions remain robust up to 5 Lyapunov exponents for any initial condition while suggesting that some initial conditions are more difficult to predict from than others. This is expected in chaotic systems with multiple attractors since the regime at which the network is situated during the beginning of forecast is difficult to estimate *a-priori*.

Since the predictions are real time, it would allow us to probe further into the capability of surrogate models. For example one can extend this framework to predict Y instead of X and feed this value into the equations governing X. The equations can then be integrated in time with the time step corresponding to the large scale variable. This can be used as an alternative to the current practice of assuming a polynomial fit in Y to be fed into X (Thornes et al., 2017). The potential shown by these relatively simple-to-train networks open new windows to explore more complicated ESNs that can integrate the feature extraction skills of convolutional neural networks and chaotic-dynamics-emulating skills of ESNs that have been previously reported to be absent in LSTMs (Vlachas et al., 2018). These relatively inexpensive networks can augment modeling efforts in the climate community especially with the shift in dynamics owing to climate change. Largely uncertain parametrization processes can be replaced by observation-trained surrogates that perform forecasting in real-time. Understanding the physics of climate change that requires years of research effort can be captured (although in a black box setting) effectively through observational data on which these machine learning algorithms are trained. Future efforts would be needed to fully understand the dynamics of these trained black-boxes.

.

References

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Briegleb, B. P., Bitz, C. M., Lin, S.-J., and Zhang, M. The formulation and atmospheric simulation of the community atmosphere model version 3 (cam3). *Journal of Climate*, 19(11): 2144–2161, 2006.

Dueben, P. D. and Bauer, P. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10):3999–4009, 2018.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G. Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11):5742–5751, 2018.

Jaeger, H. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34): 13, 2001.

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical Review Letters*, 120(2):024102, 2018.

Rasp, S., Pritchard, M. S., and Gentine, P. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39): 9684–9689, 2018.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195, 2019.

Thornes, T., Düben, P., and Palmer, T. On the use of scale-dependent precision in earth system modelling. *Quarterly Journal of the Royal Meteorological Society*, 143(703): 897–908, 2017.

Vlachas, P. R., Byeon, W., Wan, Z. Y., Sapsis, T. P., and Koumoutsakos, P. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2213): 20170844, 2018.

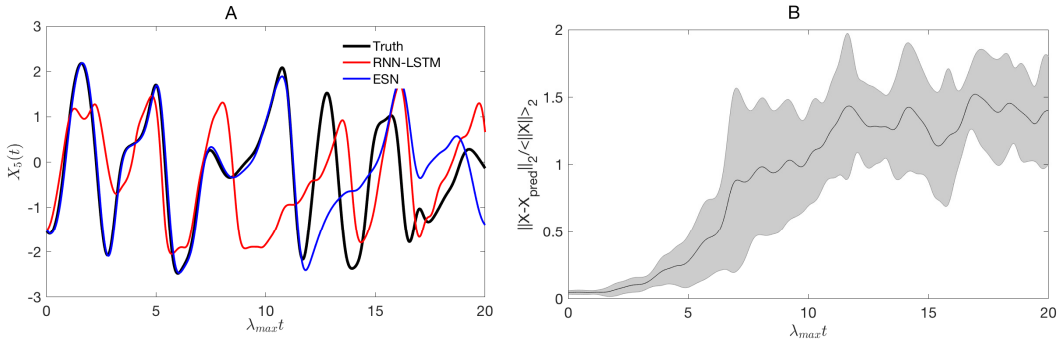

Figure 2. (A) Prediction of ESN as compared to RNN-LSTM against the true time series for X_1 with the x-axis scaled as $\lambda_{max}t$ where $\lambda_{max} = 4.5$ is the maximum Lyapunov exponent of the system (B) Mean relative L2 norm error of the predicted ESN output with time averaged over 50 initial conditions with the standard deviation shown in the shading.

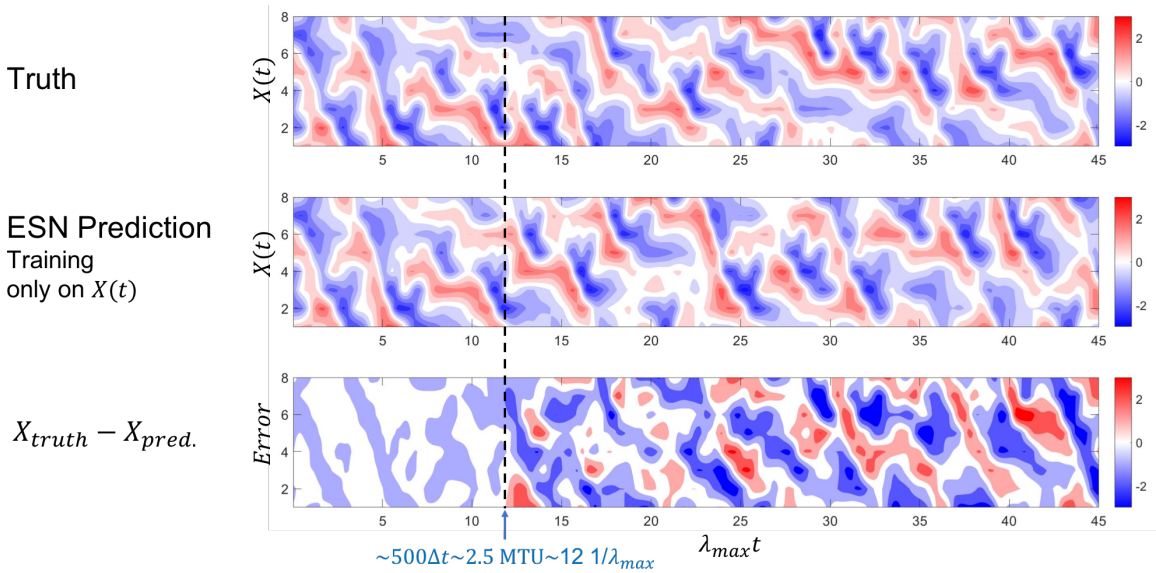

Figure 3. Prediction of ESN for all 8 elements of vector field X when trained only on X. Prediction horizons reach up to 12 Lyapunov exponents which is approximately 2.5 model time units of the system.