Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2020-39-RC1, 2020

Nonlinear Processes in Geophysics
Discussions
Open Access

Interactive comment on "Boosting performance in machine learning of geophysical flows via scale separation" *by* Davide Faranda et al.

Anonymous Referee #1

Received and published: 14 October 2020

1 General Comments

This manuscript explores the effectiveness of echo-state-networks for a hierarchy of problems. It explores 3 "toy" dynamical systems and then applies the methodology to a data driven weather prediction task. They evaluate both the equilibrium distribution (using a Xi-squared analysis) and initial value forecasts using root-mean-squared error based metrics. By these metrics, they claim that filtering the data before training an ESN generally improves these metrics in cases where the underlying dynamics are "intermittent" or show strong "coupling between timescales". For all the problems except for Lorenz 96 (L96), they pre-filter with moving averages, whereas for L96 they take advantage of the built-in scale separation between the large-scale and small-scale

variables.

Overall, I thought the results were interesting and relevant to geophysical problems which often feature intermittent and multiscale dynamics, but was not convinced that their claims were valid. See my comments below.

1. The quality of presentation should be improved
   (a) In a few cases, the color schemes used were not intelligible to colorblind readers, which significantly hampered my ability to understand their results. There are many multi-panel figures, which are explained only briefly in the text.
   (b) Notation is used inconsistently and unclearly in some places. Also, this paper introduces redundant notation. Vector and scalar quantities are not differentiated clearly.
   (c) The literature review in the introduction was incomplete in a few places. Also, for an article in a geophysical science journal, concepts like CNNs, RNNs, ESNs should all be clearly defined and differentiated from one another. The introduction sometimes incorrectly conflates these concepts.
   (d) The conclusion contains many helpful motivations that could have helped guide me through the introduction and the methods sections.
2. Their ESNs appear to fail to meaningfully reproduce the time series of the Pomeau-Manneville (fig 5) or Lorenz 96 (fig 8) examples. As with any negative result, it is unclear whether some minor methodological improvement could fix it, so I am not sure what insights these examples provides. In particular, some authors have demonstrated substantially nearly optimal performance with data-driven techniques for Lorenz 96 (Gagne, et. al. 2020) and with ESNs for similar Kuramoto-Shivashinksy model (Pathak, et. al 2018). Were the authors able to replicate the success of these previous studies?

3. The Xi-squared testing procedure seems suspect. It mixes a parametric test (Xi-squared) with a boot-strapping based test. Is there any support for this technique in the literature? It would be preferable to use a more well-known statistical test for this problem (e.g. Wilcoxon Rank sum, Kolmogorov-Smirnov).
4. The sea-level pressure example was compelling.

2 Specific Comments

Title: “Boosting performance”

This is a quibble, but “boosting” has a rather specific meaning in the machine learning literature https://en.wikipedia.org/wiki/Boosting_(machine_learning). This could be misleading.

L8: “with an optimal choice of spatial coarse grain and time filtering”

With an optimal choise of spatial coarse-graining

L20. Buchanan.

How does this PhD dissertation relate to the previous assertion. Please be more specific.

L26. Gentine

There are many other articles on parameterizations which should be mentioned e.g. (Brenowitz and Bretherton 2018, 2019; Yuval and O’Gorman 2020; Kransopalky 2005, 2013; Gettleman et. al 2020).

L27. This introduction should also mention (Rasp et. al 2020; Weyn et.al 2019) for the pure weather prediction problem

L40. “Recent examples include. . . convolutional neural ntworks, . . . ”

With the previous sentence in mind, this wording implies that convolutional neural networks are a type of RNN. I believe the references all used feed-forward architectures.

L65. “Previous results (Scher, 2018; Dueben and Bauer, 2018; Scher and Messori, 2019) suggest that RNN simulations”

Again, I don’t think these papers all studied RNNs. At least some used feed-forward architectures.

L73-90. Overall, this description does not clarify what ESNs are, and why they work outperform traditional RNNs for some problems (e.g. the vanishing gradients problem).

L90: “We estimate w_out via a ridge regression with lambda=”

How was this parameter chosen? ESNs are very sensitive to this parameters, and the optimal parameter may vary from problem to problem. This could potentially explain the poor performance on the L96 and Pomeau-Manneville examples below.

L98. “Let, U be . . . ”

For readibility, try to re-use previously introduced notation to avoid introducing too many new symbols. For instance “v” is the same as “r” in eq 1-4.

Are theses tests univariate? The equations are multivariate.

L120: “we observed excessive rejection rates”

How do you quantify this?

L121: “we use 10000 samples”

What is “a sample”. Is it a single time step of r(t) above (e.g. a K-dimensional vector)? Is it the number of timesteps or is it the number of timesteps times K? This would be clearer if described in terms of the notation used in Eqs 1-4.

L135: This formula seems odd. I would normally define predictability by computing RMSE versus the truth for a single timestep. In this case they compute the average

MSE accumulated over several timesteps. Also, this formula only makes sense for scalar u and v, but I thought we are in the vector setting?

Section 2.2: It is unclear why this moving average is described here. It would be clearer if the introduction had introduced a broad outline of the paper.

248: “Performances are again better when using the exact formula (Figure 4b,e,h) than using the residuals δu (Figure 4c,f,i).”

It would be helpful to refer to Eq 11 here.

Line 250: “ESN simulations do not reproduce the intermittency in the average of the target signal. They only show some second order intermittency in the fluctuations.”

Is “the average” supposed to mean “the moving average” rather than “time average”? Is “second order intermittency?”. Is this a formal concept?

L270. Forward Euler time steppers are notoriously inaccurate. Why not use a more advanced time stepper (e.g. Runge Kutta) for better accuracy? There are many convenient software packages for integrating ODEs with better schemes (e.g. ode45 in MatLab).

What is N? It must be network size, but given all the notational changes it is hard to be sure.

Line331: “We show the results using the residuals (Eq. 9)”

Why not show the results with the “exact method” (Eq. 11)? It seems the earlier results implied this technique was more effective.

Figure 10 b-d. These panels all look different. I don’t see much reason to prefer panel d to c. Could the authors present a more convincing visualization for the claimed improvement of the moving average filter? Maybe a single power-spectra plot would be more succinct, especially since the author’s don’t comment on the timing the high-frequency vs low-frequency results.

Line 373. “For the Lorenz 1996 mode, we did not apply a moving average filter to the data,. . . ”

It would have been nice to see this motivation described in Section 3.

3 Technical corrections

L73. ‘Reservoir compution"’

There is a missing quote.

L74. “The principle of Reservoir computing”

Does “Reservoir” need to be capitalized here? If so, I would expect “computing” to be capitalized as well. “reservoir” is not always capitalized in this manuscript.

L76. “In our study *ESNs* are implemented”

L77. "The code is given *in the* appendix

L97: “to this purpose” –> “for this purpose”

L239: “we find the best match. . . are obtained for w=3”

Correct “are” to “is”.

249: “Figure 5a)”

Remove the parenthesis

Line275. “Figure 6.b,d)”

This should read “Figure 6 b,d”. Figures should be referred to with a consistent convention.

L288. “distance T”.

Do the authors mean Σ? T is the length of the time series.

Figure 8: The text in this graphic is fuzzy. Please save at a higher resolution.

Figure 2a: This plot has too many curves. Red-green is bad for colorblind readers. It is hard to see the author's point.

Figure 3, 4: These colorscales are not legible for colorblind readers. I could not interpret these figures and relied on the author's textual description of the results. I suggesting using "viridis" or another sequential colorbar.

4 References

Pathak, J., Hunt, B., Girvan, M., Lu, Z., & Ott, E. (2018). Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach. Physical Review Letters, 120(2). https://doi.org/10.1103/physrevlett.120.024102

Gagne, D. J., II, Christensen, H. M., Subramanian, A. C., & Monahan, A. H. (2020). Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model. Journal of Advances in Modeling Earth Systems, 12(3). https://doi.org/10.1029/2019ms001896

Yuval, J. & O'Gorman, P. A. Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nat. Commun.* **11**, 3295 (2020)

Brenowitz, N. D. & Bretherton, C. S. Spatially Extended Tests of a Neural Network Parametrization Trained by CoarseâĂŘgraining. *J. Adv. Model. Earth Syst.* (2019) doi:10.1029/2019MS001711

Krasnopolsky, V. M., Fox-Rabinovitz, M. S. & Chalikov, D. V. New Approach to Calculation of Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Longwave Radiation in a Climate Model. *Mon. Weather Rev.* **133**, 1370–1383 (2005)

Krasnopolsky, V. M., Fox-Rabinovitz, M. S. & Belochitski, A. A. Using Ensemble of Neural Networks to Learn Stochastic Convection Parameterizations for Climate and Numerical Weather Prediction Models from Data Simulated by a Cloud Resolving Model. *Advances in Artificial Neural Systems* **2013**, e485913 (2013)

Brenowitz, N. D. & Bretherton, C. S. Prognostic Validation of a Neural Network Unified Physics Parameterization. *Geophys. Res. Lett.* **17**, 2493 (2018)

O'Gorman, P. A. & Dwyer, J. G. Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events. *J. Adv. Model. Earth Syst.* **10**, 2548–2563 (2018)

Gettleman et. al. Machine Learning the Warm Rain Process.

Rasp, S. *et al.* WeatherBench: A benchmark dataset for dataâĂŘdriven weather forecasting. *J. Adv. Model. Earth Syst.* (2020) doi:10.1029/2020MS002203

Weyn, J. A., Durran, D. R. & Caruana, R. Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500âĂŘhPa Geopotential Height From Historical Weather Data. *J. Adv. Model. Earth Syst.* **11**, 2680–2693 (2019)

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2020-39, 2020.