Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-23-SC1, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Nonlinear Processes
in Geophysics
Discussions
Open Access

Interactive comment on "Generalization properties of neural networks trained on Lorenzsystems" *by* S. Scher and G. Messori

Sebastian Scher

sebastian.scher@misu.su.se

Received and published: 26 June 2019

Dear Reviewer,

Thank you for your thorough and very constructive review. We will write a detailed reply after receiving the comments from the other reviewers. However, to clarify a few points and to aid the other reviewers, we briefly reply to your main points and outline how we plan to address them.

You are absolutely right in that all conclusions from our paper do only apply for feed-forward networks. We realize that we should have pointed this out already in the abstract and potentially in the title. In the revised version, we will mention this very clearly, and also discuss it more verbose in the main text/conclusion. Additionally, we plan to

extend the analysis of neuron-activations to larger neural network models.

"The argumentation in Section 2, about whether the network learns only one or many mappings for different regions is inconsistent. The two representations are mathematically equivalent. The impression the reader gets about what the authors are trying to express, is whether different parts of the network are re-sponsible for different (local dynamics) parts of the training data"

This is indeed what we wanted to express. We agree that the mathematical notation might be misleading, and we will either remove it or explain it better in the revised version.

"The authors do not explain the training procedure and how they cope against overfitting in the CNN applied to Lorenz-95. Especially in the low data regime, the absence of measures against overfitting can have a detrimental influence on the performance on the test dataset."

Thanks for pointing out that we forgot to include the exact training procedure of the CNN for the Lorenz95 in the text. We used the last 10 percent of each training set as validation data, and controlled overfitting via monitoring validation loss (the training is stopped when the validation loss has not improved for more than 4 epochs, or when 30 epochs were reached.) We will explain this more clearly in the revised version.

"Since the neural network is forecasting a deterministic system with full state in- formation, the prediction accuracy reported in the Appendix on page 17, seems quite low. In the provided plots, the networks seems to be forecasting inaccu- rately, as the difference in the plots even at early timesteps is obvious."

It is true that the Lorenz95 system is a deterministic system, but it is a chaotic deterministic system. In fact it was explicitly designed in order to have chaotic behaviour for the study of predictability-limits (see Lorenz 1996). For the lorenz95, this predictability limit is at a forecast-time of roughly 2-3 time-units. This is usually expressed in terms of initial condition uncertainty (if there is a very small error in the initial conditions, after reaching the predictability limit, the forecast will be only as good as a random forecast). However, it also translates to model uncertainty: if a surrogate-model of the system is not absolutely perfect, it will not be able to make good forecasts behind the predictability limit. Therefore, it is expected that any type of surrogate model (like a neural network) won't be able to do well after this predictability limit (which is exactly what we see). This is also seen in other studies were the error of neural-network forecasts on the lorenz95 system increases rapidly with forecast-time (e.g. Vlachas et al 2018). We do however realize that this is quite un-intuitive for readers not familiar with atmospheric predictability studies, and we will add discussion on this in the revised manuscript. Additionally, we now realize that panel c) of figure A1 might be misleading, and might have caused your comment that *"In the provided plots, the networks seems to be forecasting inaccurately, as the difference in the plots even at early timesteps is obvious."* The plot shows the evolution of the lorenz95 model and the network climate, however they were not initialized with the same state. The plot was intended not to show forecast-performance, but to demonstrate that long runs of the CNN do look realistic. However, this is not obvious from the caption, and we apologize for the confusion. In the revised version we will show plots of runs that are actually initialized from exactly the same state, then the plots can be used to analyze forecast performance as well. We will also make this clearer in the caption.

Lorenz, E. N.: Predictability: A problem partly solved, in: Proc. Seminar on predictability, vol. 1, 1996.

P. R., Byeon, W., Wan, Z. Y., Sapsis, T. P., and Koumoutsakos, P.: Data-driven forecasting of high-dimensional chaotic sys-tems with long short-term memory networks, Proc. R. Soc. A, 474, 20170 844, https://doi.org/10.1098/rspa.2017.0844, http://rspa.royalsocietypublishing.org/content/474/2213/20170844, 2018

"The statement "... the trajectories of the network forecast simply point back to- wards the region included in the training." regarding the behavior of the neural network in

regions of the phase space not included in the training data, seems rather arbitrary. Since the neural network is not trained in these regions the be- havior can be anything."

What we meant to say is that we actually observe that in our trained networks, the network forecasts initialized outside the training phase space do point back to the training phase space. You are of course right that as an a-priori assumption this would be rather arbitrary. We will make it more clear in the revised version that this is an empirical observation.

Sebastian Scher

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2019-23, 2019.