Review - "Boosting performance in machine learning of turbulent and geophysical flows via scale separation"

November 4, 2020

The authors are utilizing Echo State Networks to predict filtered dynamics in the perturbed Lorenz 1963 equations, the Pomeau-Manneville 89 intermittent map, and the Lorenz 1996 equations. A moving average filter is utilized for scale separation in time. The filtered dynamics are smoother and easier to predict. A residual term is added, either sampled from the training data, or based on an analytic formula derived from the moving average filter. Assuming that the filter width is smaller than the associated large timescales of the processes involved, the large scale processes can be successfully predicted. The authors claim that modeling only the spatially coarse grained and time averaged state can boost performance of ESN. However, the generalization of this argument to more realistic systems is not sufficiently supported by the results, as elaborated in the comment section below.

The idea of utilizing a moving average filter for noise reduction and scale separation, or spatial coarse graining is known. I am not sure that the novelty of the paper to apply ESNs to (spatially/time) filtered dynamics, is enough to guarantee publication in the journal. The effect of the unmodeled dynamics (the information lost during filtering) is not taken into account in the model. In most interesting applications, the effect of the unmodeled modes is the problem, and a field of study by itself (closure models in turbulence, small scale models in weather etc.).

1 Comments

1. In the three-dimensional Lorenz system, it is logical that the moving average filter produces better results. By construction, noise is added to the system. It does not come as a surprise that the ESN predicting the filtered dynamics (which are smoother) and augmented with the random residual terms, shows superior performance. However, there is no complex multiscale effect taking place, as the whole state information is given to the system (no hidden state, at least nothing is mentioned in the text about it). Moreover, as a reference time-scale, the Lyapunov time of the deterministic system is used, although the system is augmented with noise, which means that the effective Lyapunov time is in essence much shorter, as stochasticity accelerates the divergence of nearby trajectories. In any case, it is important to be critical about the conclusions drawn from this case.

2. In the Pomeau-Manneville intermittent map, it is not a surprise that the ESN cannot capture the dynamics, as they are changing very rapidly, even visually they look completely stochastic. A deterministic ESN with tanh (smooth, continuous) activation function cannot be expected to produce trajectories that look spiking/stochastic/rapidly changing. Most previous studies on ESNs were handling relatively smooth signals, and not such rapidly changing signals. At least the nature of the signal has to be taken into account in the selection of the activation function of the reservoir. Thus, it does not come as a surprise that utilizing the ESN on the time averaged dynamics and then adding a stochastic residual improves performance. As expected, the plain ESN diverges, as demonstrated also in previous studies with such non-smooth signals.

3. In the Lorenz 96 system, as demonstrated in Figure 8, the method fails to capture the long-term climate, as the dynamics predicted by the ESN are clearly different from the groundtruth.

4. In the sea-level pressure, the moving average filter ESN does not achieve any significant improvement based on the results in Figure 9.

5. In the abstract, the authors claim that "multiscale dynamics and intermittency introduce severe limitations on the applicability of recurrent neural networks, both for short-term forecasts, as well as for the reconstruction of the underlying attractor". This is shown for Echo State Networks in the document, but not in general for Recurrent Neural Networks. The argument has to be relaxed to take into account only ESNs, or a relevant reference for other RNN architectures should be given.

6. There is a contradiction in the text, in page 3, the authors state that "We aim at understanding this sensitivity in a deeper way, while assessing the possibility to reduce its impact on prediction through simple noise reduction methods", although one sentence before, they claim that they choose the ESN framework for "...its ability to forecast chaotic time series and its stability to noise". These sentences are contradicting each other. Later in the text, the authors state "Since Echo State Networks are known to be sensitive to noise (see e.g. [34]), ...".

7. The analysis of the performance of the proposed method based on different parameters e.g. intermittency of dynamics/degree of coarse graining, etc. is interesting. However, this is not adequate to warrant publication.

2 Proposal

Reject.