



Hyper-spherical Reservoirs for Echo State Networks

Pietro Verzelli¹(✉), Cesare Alippi^{1,2}, and Lorenzo Livi^{3,4}

¹ Faculty of Informatics, Università della Svizzera Italiana,
69000 Lugano, Switzerland
pietro.verzelli@usi.ch

² Department of Electronics, Information and Bioengineering,
Politecnico di Milano, 20133 Milan, Italy

³ Departments of Computer Science and Mathematics,
University of Manitoba, Winnipeg, MB R3T 2N2, Canada

⁴ Department of Computer Science, College of Engineering,
Mathematics and Physical Sciences, University of Exeter,
Exeter EX4 4QF, UK

Abstract. In this paper, we propose a model of ESNs that eliminates critical dependence on hyper-parameters, resulting in networks that provably cannot enter a chaotic regime and, at the same time, denotes nonlinear behaviour in phase space characterised by a large memory of past inputs, comparable to the one of linear networks. Our contribution is supported by experiments corroborating our theoretical findings, showing that the proposed model displays dynamics that are rich-enough to approximate many common nonlinear systems used for benchmarking.

Keywords: Echo State Networks · Edge of criticality · Memory–nonlinearity tradeoff

Although the use of *Recurrent Neural Networks (RNNs)* in machine learning is boosting, also as effective building blocks for deep learning architectures, a comprehensive understanding of their working principles is still missing [4, 26]. Of particular relevance are Echo State Networks (ESNs), introduced by Jaeger [13] and independently by Maass *et al.* [16] under the name of Liquid State Machine (LSM), which emerge from RNNs due to their training simplicity. The basic idea behind ESNs is to create a randomly connected recurrent network, called reservoir, and feed it with a signal so that the network will encode the underlying dynamics in its internal states. The desired – task dependent – output is then generated by a readout layer (usually linear) trained to match the states with the desired outputs. Despite the simplified training protocol, ESNs are universal function approximators [10] and have shown to be effective in many relevant tasks [2, 3, 7, 19–22].

These networks are known to be sensitive to the setting of hyper-parameters like the Spectral Radius (SR), the input scaling and the sparseness degree [13],

which critically affect their behaviour and, hence, the performance at task. Fine tuning of hyper-parameters requires cross-validation or ad-hoc criteria for selecting the best-performing configuration. Experimental evidence and some results from the theory show that ESNs performance is usually maximised in correspondence of a very narrow region in hyper-parameter space called Edge of Chaos (EoC) [1, 6, 14, 15, 23–25, 30]. However, we comment that beyond such a region ESNs behave chaotically, resulting in useless and unreliable computations. At the same time, it is everything but trivial configuring the hyperparameters to lie on the EoC still granting a non-chaotic behaviour. A very important property for ESNs is the Echo State Property (ESP), which basically asserts that their behaviour should depend on the signal driving the network only, regardless of its initial conditions [32]. Despite being at the foundation of theoretical results [10], the ESP in its original formulation raises some issues, mainly because it does not account for multi-stability and is not tightly linked with properties of the specific input signal driving the network [17, 31, 32].

In this context, the analysis of the memory capacity (as measured by the ability of the network to reconstruct or remember past inputs) of input-driven systems plays a fundamental role in the study of ESNs [8, 9, 12, 27]. In particular, it is known that ESNs are characterized by a memory–nonlinearity trade-off [5, 11, 28], in the sense that introducing nonlinear dynamics in the network degrades memory capacity. Moreover, it has been recently shown that optimizing memory capacity does not necessarily lead to networks with higher prediction performance [18].

In a recent paper [29], we proposed an ESN model that eliminates critical dependence on hyper-parameters, resulting in models that cannot enter a chaotic regime. In addition to this major outcome, we showed that such networks denote nonlinear behaviour in phase space characterised by a large memory of past inputs (see Fig. 1): the proposed model generates dynamics that are rich-enough to approximate nonlinear systems typically used as benchmarks. Our contribution was based on a nonlinear activation function that normalizes neuron activations on a hyper-sphere. We showed that the spectral radius of the reservoir, which is the most important hyper-parameter for controlling the ESN behaviour, plays a marginal role in influencing the stability of the proposed model, although it has an impact on the capability of the network to memorize past inputs. Our theoretical analysis demonstrates that this property derives from the impossibility for the system to display a chaotic behaviour: in fact, the maximum Lyapunov exponent is always null. An interpretation of this very important outcome is that the network always operates on the EoC, regardless of the setting chosen for its hyper-parameters.

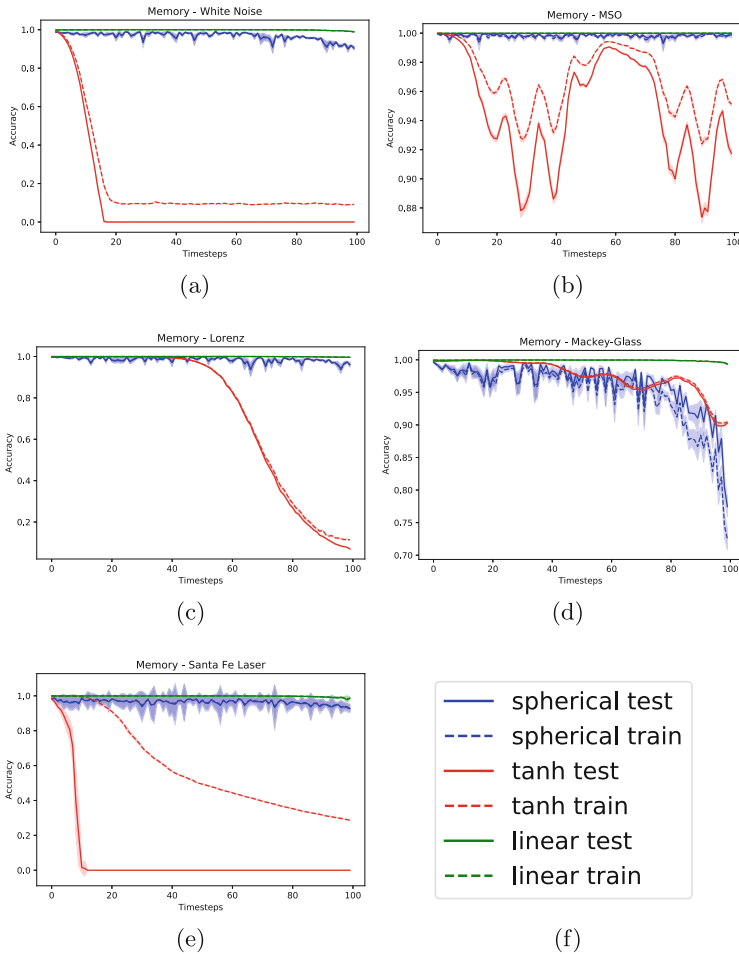


Fig. 1. Results of the experiments on memory for different benchmarks. Panel (a) displays the white noise memorization task, (b) the MSO, (c) the x -coordinate of the Lorenz system, (d) the Mackey-Glass series and (e) the Santa Fe laser dataset. As described in the legend (f), different line types account for results obtained on training and test data. The shaded areas represent the standard deviations, computed using 20 different realization for each point.

References

1. Bertschinger, N., Natschläger, T.: Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput.* **16**(7), 1413–1436 (2004). <https://doi.org/10.1162/089976604323057443>
2. Bianchi, F.M., Scardapane, S., Løkse, S., Jenssen, R.: Reservoir computing approaches for representation and classification of multivariate time series. arXiv preprint [arXiv:1803.07870](https://arxiv.org/abs/1803.07870) (2018)

3. Bianchi, F.M., Scardapane, S., Uncini, A., Rizzi, A., Sadeghian, A.: Prediction of telephone calls load using echo state network with exogenous variables. *Neural Netw.* **71**, 204–213 (2015). <https://doi.org/10.1016/j.neunet.2015.08.010>
4. Ceni, A., Ashwin, P., Livi, L.: Interpreting recurrent neural networks behaviour via excitable network attractors. *Cogn. Comput.* (2019). <https://doi.org/10.1007/s12559-019-09634-2>
5. Dambre, J., Verstraeten, D., Schrauwen, B., Massar, S.: Information processing capacity of dynamical systems. *Sci. Rep.* **2** (2012). <https://doi.org/10.1038/srep00514>
6. Gallicchio, C.: Chasing the echo state property. arXiv preprint [arXiv:1811.10892](https://arxiv.org/abs/1811.10892) (2018)
7. Gallicchio, C., Micheli, A., Pedrelli, L.: Comparison between DeepESNs and gated RNNs on multivariate time-series prediction. arXiv preprint [arXiv:1812.11527](https://arxiv.org/abs/1812.11527) (2018)
8. Ganguli, S., Huh, D., Sompolinsky, H.: Memory traces in dynamical systems. *Proc. Nat. Acad. Sci.* **105**(48), 18970–18975 (2008). <https://doi.org/10.1073/pnas.0804451105>
9. Goudarzi, A., Marzen, S., Banda, P., Feldman, G., Teuscher, C., Stefanovic, D.: Memory and information processing in recurrent neural networks. arXiv preprint [arXiv:1604.06929](https://arxiv.org/abs/1604.06929) (2016)
10. Grigoryeva, L., Ortega, J.P.: Echo state networks are universal. *Neural Netw.* **108**, 495–508 (2018). <https://doi.org/10.1016/j.neunet.2018.08.025>
11. Inubushi, M., Yoshimura, K.: Reservoir computing beyond memory-nonlinearity trade-off. *Sci. Rep.* **7**(1), 10199 (2017). <https://doi.org/10.1038/s41598-017-10257-6>
12. Jaeger, H.: Short term memory in echo state networks, vol. 5. GMD-Forschungszentrum Informationstechnik (2002)
13. Jaeger, H., Haas, H.: Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**(5667), 78–80 (2004). <https://doi.org/10.1126/science.1091277>
14. Legenstein, R., Maass, W.: Edge of chaos and prediction of computational performance for neural circuit models. *Neural Netw.* **20**(3), 323–334 (2007). <https://doi.org/10.1016/j.neunet.2007.04.017>
15. Livi, L., Bianchi, F.M., Alippi, C.: Determination of the edge of criticality in echo state networks through Fisher information maximization. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(3), 706–717 (2018). <https://doi.org/10.1109/TNNLS.2016.2644268>
16. Maass, W., Natschläger, T., Markram, H.: Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**(11), 2531–2560 (2002). <https://doi.org/10.1162/089976602760407955>
17. Manjunath, G., Jaeger, H.: Echo state property linked to an input: exploring a fundamental characteristic of recurrent neural networks. *Neural Comput.* **25**(3), 671–696 (2013). https://doi.org/10.1162/NECO_a.00411
18. Marzen, S.: Difference between memory and prediction in linear recurrent networks. *Phys. Rev. E* **96**(3), 032308 (2017). <https://doi.org/10.1103/PhysRevE.96.032308>
19. Palumbo, F., Gallicchio, C., Pucci, R., Micheli, A.: Human activity recognition using multisensor data fusion based on reservoir computing. *J. Ambient Intell. Smart Environ.* **8**(2), 87–107 (2016)
20. Pathak, J., Hunt, B., Girvan, M., Lu, Z., Ott, E.: Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys. Rev. Lett.* **120**(2), 024102 (2018)

21. Pathak, J., Lu, Z., Hunt, B.R., Girvan, M., Ott, E.: Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data. *Chaos: Interdisc. J. Nonlinear Sci.* **27**(12), 121102 (2017). <https://doi.org/10.1063/1.5010300>
22. Pathak, J., et al.: Hybrid forecasting of chaotic processes: using machine learning in conjunction with a knowledge-based model. *Chaos: Interdisc. J. Nonlinear Sci.* **28**(4), 041101 (2018). <https://doi.org/10.1063/1.5028373>
23. Rajan, K., Abbott, L.F., Sompolinsky, H.: Stimulus-dependent suppression of chaos in recurrent neural networks. *Phys. Rev. E* **82**(1), 011903 (2010). <https://doi.org/10.1103/PhysRevE.82.011903>
24. Rivkind, A., Barak, O.: Local dynamics in trained recurrent neural networks. *Phys. Rev. Lett.* **118**, 258101 (2017). <https://doi.org/10.1103/PhysRevLett.118.258101>
25. Sompolinsky, H., Crisanti, A., Sommers, H.J.: Chaos in random neural networks. *Phys. Rev. Lett.* **61**(3), 259 (1988). <https://doi.org/10.1103/PhysRevLett.61.259>
26. Sussillo, D., Barak, O.: Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.* **25**(3), 626–649 (2013). https://doi.org/10.1162/NECO_a.00409
27. Tiño, P., Rodan, A.: Short term memory in input-driven linear dynamical systems. *Neurocomputing* **112**, 58–63 (2013). <https://doi.org/10.1016/j.neucom.2012.12.041>
28. Verstraeten, D., Dambre, J., Dutoit, X., Schrauwen, B.: Memory versus non-linearity in reservoirs. In: *IEEE International Joint Conference on Neural Networks*, pp. 1–8. IEEE, Barcelona (2010)
29. Verzelli, P., Alippi, C., Livi, L.: Echo state networks with self-normalizing activations on the hyper-sphere. *arXiv preprint arXiv:1903.11691* (2019)
30. Verzelli, P., Livi, L., Alippi, C.: A characterization of the edge of criticality in binary echo state networks. In: *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE (2018). <https://doi.org/10.1109/MLSP.2018.8516959>
31. Wainrib, G., Galtier, M.N.: A local echo state property through the largest Lyapunov exponent. *Neural Netw.* **76**, 39–45 (2016). <https://doi.org/10.1016/j.neunet.2015.12.013>
32. Yildiz, I.B., Jaeger, H., Kiebel, S.J.: Re-visiting the echo state property. *Neural Netw.* **35**, 1–9 (2012). <https://doi.org/10.1016/j.neunet.2012.07.005>