



**HAL**  
open science

# Boosting performance in Machine Learning of Turbulent and Geophysical Flows via scale separation

Davide Faranda, M. Vrac, Pascal Yiou, F. M. E. Pons, A Hamid, G Carella,  
C. G. Ngoungue Langué, S Thao, V. Gautard

## ► To cite this version:

Davide Faranda, M. Vrac, Pascal Yiou, F. M. E. Pons, A Hamid, et al.. Boosting performance in Machine Learning of Turbulent and Geophysical Flows via scale separation. 2019. hal-03341712v1

**HAL Id: hal-03341712**

**<https://hal.science/hal-03341712v1>**

Preprint submitted on 29 Oct 2019 (v1), last revised 12 Sep 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Boosting performance in Machine Learning of Turbulent and Geophysical Flows via scale separation

D. Faranda\*

*Laboratoire des Sciences du Climat et de l'Environnement,  
CE Saclay l'Orme des Merisiers, UMR 8212 CEA-CNRS-UVSQ,  
Université Paris-Saclay & IPSL, 91191 Gif-sur-Yvette, France.  
London Mathematical Laboratory, 8 Margravine Gardens, London, W68RH, UK.*

M. Vrac, P. Yiou, F.M.E. Pons, A. Hamid, G. Carella, C.G. Ngoungue Langue, S. Thao

*Laboratoire des Sciences du Climat et de l'Environnement,  
CE Saclay l'Orme des Merisiers, UMR 8212 CEA-CNRS-UVSQ,  
Université Paris-Saclay & IPSL, 91191 Gif-sur-Yvette, France.*

V. Gautard

*DRF/IRFU/DEDIP//LILAS Département d'Electronique des Detecteurs et d'Informatique pour la Physique,  
CE Saclay l'Orme des Merisiers, 91191 Gif-sur-Yvette, France.*

(Dated: October 29, 2019)

Recent advances in statistical learning have opened the possibility to forecast the behavior of chaotic systems using recurrent neural networks. In this letter we investigate the applicability of this framework to geophysical flows, known to be intermittent and turbulent. We show that both turbulence 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. We suggest that possible strategies to overcome such limitations should be based on separating the smooth large-scale dynamics, from the intermittent/turbulent features. We test these ideas on global sea-level pressure data for the past 40 years, a proxy of the atmospheric circulation dynamics.

The advent of high-performance computing has paved the way for advanced analyses of high-dimensional datasets [1, 2]. Those successes have naturally raised the question on whether it is possible to learn the dynamical behavior of a system without simulating the underlying evolution equations. Such an interest is motivated on one side by the fact that many complex systems still miss a universally accepted state equation — e.g. the brain dynamics [3], macro-economical and financial systems [4] — and, on the other, by the need of reducing the complexity of the dynamical evolution for the systems where the underlying equations are known — e.g. on geophysical and turbulent flows [5]. From a mathematical point of view, the Navier-Stokes equations, which are the cornerstone of turbulent flow dynamics, have yet to find a complete assessment for the existence/uniqueness of the solutions, which is a millennium Clay problem [6]. They are difficult to simulate for large systems such as the geophysical flows, so that approximations and parameterizations are needed for meteorological and climatological applications [7]. These difficulties are enhanced by those encountered in the modelling of phase transitions that lead to clouds and convection, which are major source of uncertainty in climate modelling [8]. Machine Learning techniques capable to learn geophysical flows dynamics could help improve those approximations and avoid running costly simulations representing all spatial/temporal scales.

Several efforts have recently been done to apply machine learning to the prediction of geophysical data [9], to learn parameterizations of subgrid processes in climate models [10–12], for nowcasting [13–15] and forecasting [16–18] of weather variables, to quantify the uncertainty of deterministic weather prediction [19]. However, these attempts mainly consist of complementing deterministic models, or short term prediction of time series data: while their utility may be substantial, they are still far from learning the dynamics of a complex system such as the Earth climate. A first great step in this direction was the use of Echo State Networks (ESN, [20]) to forecast the behavior of chaotic systems, such as the Lorenz 1963 [21] and the Kuramoto-Sivashinsky [22] dynamics. It was shown that ESN predictions of both systems attain performances comparable to those obtained with the real equations [23, 24]. Good performance of regularized ESN in the prediction of multidimensional chaotic time series was obtained, both from simulated and real data [25]. This success motivated several follow-up studies with a focus on meteorological and climate data. These are based on the idea of feeding different statistical learning algorithms with data issued from dynamical systems of different complexity, in order to study short and long-term predictability. Recent examples include equation-informed moment-matching for the Lorenz96 model [26, 27], multi-layer perceptrons to re-analysis data [28], or convolutional neural networks to simplified climate simulation models [29, 30]. All those learning models were capable to recover some short-term predictability but failed in obtaining a long term behav-

---

\* Correspondence to [davide.faranda@lscce.ipsl.fr](mailto:davide.faranda@lscce.ipsl.fr)

ior coherent with the input data.

In this letter we investigate the possible origins of these limitations by analyzing two of the most striking features of geophysical flows, namely turbulence and intermittency. Turbulence affects the dynamics by introducing energy at several spatial and temporal scales. This results in non-smooth trajectories within the flow, leading to local unpredictability and increasing the number of degrees of freedom needed to describe the dynamics [31]. Intermittency triggers large fluctuations of observables of the motion in time and space [32]. Both these effects have made challenging the description of geophysical flow both for short term forecast and to reconstruct the underlying climate attractor properties.

By applying ESN to turbulent and intermittent systems, we investigate the limitations introduced by these effect in applying machine learning techniques in geophysical flows. We begin with the simplest possible setup: we simulate the effects of turbulence by artificially introducing small scales dynamics in the Lorenz 1963 equations [21] via additive noise. We investigate the Pomeau-Manneville equations [33] stochastically perturbed with additive noise to have an example of intermittent behavior. We show how the performance of ESN in predicting the behavior of the system deteriorates rapidly with increasing noise and investigate how scale separation, introduced with a moving average filter, recovers short-term predictability. The idea of using moving average for scale separation is already established for meteorological variables [34]. We choose the ESN framework following the results of [23, 24], and an established literature about its ability to forecast chaotic time series and its stability to noise. For example, [35, 36] analyse and compare the predictive performance of simple and improved ESN on simulated and observed one-dimensional chaotic time series. We aim at understanding this sensitivity in a deeper way, while assessing the possibility to reduce its impact on prediction with simple noise reduction methods. This step has a crucial importance, as noise is inevitably present in each dataset issued from real-world systems. The remaining of this letter is organised as follows: first we give an overview of the ESN method and provide the description of the systems used. Then we show the results for the perturbed Lorenz equations and the Pomeau-Manneville intermittent maps. We discuss the improvement obtained on the short-term prediction and the long-term attractor reconstruction obtained with the moving average filter (see [37]-B). We conclude by testing these ideas on atmospheric circulation data.

Echo State Networks (ESN) are implemented as follows. The code is given in [37]. Let  $x(t)$  be the  $K$ -dimensional observable originating from the dynamical system and  $r(t)$  be the  $N$  dimensional reservoir state, then:

$$r(t + dt) = \tanh(Wr(t) + W_{in}x(t)) \quad (1)$$

where  $W$  is the adjacency matrix of the reservoir: its dimensions are  $N \times N$ , and  $N$  is the number of neurons

of the reservoir. The coefficients are randomly sampled from a uniform distribution in  $[-0.5; 0.5]$ .  $W_{in}$ , with dimensions  $N \times K$ , is the weight matrix of the connections between the input layer and the reservoir and the coefficients are randomly sampled, as for  $W$ . The output of the network at time step  $t + dt$  is  $W_{out}r(t + dt) = \hat{x}(t + dt)$ , where  $\hat{x}(t + dt)$  is the ESN prediction,  $W_{out}$  with dimensions  $K \times N$ , is the weight matrix of the connections between the reservoir neurons and the output layer. We estimate  $W_{out}$  via a ridge regression [38]. In the prediction phase we have a recurrent relationship:

$$r(t + dt) = \tanh(Wr(t) + W_{in}W_{out}r(t)). \quad (2)$$

The systems we analyze are the Lorenz 1963 attractor [21] with the classical parameters, discretized with a Euler scheme and a  $dt = 0.001$  and the Pomeau-Manneville intermittent map with  $\beta = 0.91$  [33](see also [37]-C). The systems are perturbed via additive noise drawn from a Gaussian distribution with zero mean and  $\epsilon$  standard deviation. The initial conditions are randomly selected within a long trajectory of  $5 \cdot 10^6$  iterations. The training phase is fixed to  $10^5$  iterations, starting from the initial conditions, and the prediction phase to  $10^4$  iterations. These values have been derived after studying the performance of the ESN as a function of different learning windows (see [37]-C). For each dataset we run the ESN only once but for 100 different initial conditions.

In this paper, we use three different indicators of performance of the ESN: i) an indicator based on the  $\chi^2$  distance to assess whether the ESN forecast has reproduced a distribution of variable compatible with the invariant density of the system (see [37]-B). Here, by the term invariant density we refer to the physical measure of the discretized attractors, often referred to as the Sinai-Bowen-Ruelle measure [39]. The null hypothesis to be rejected is that the ESN generated data have the same distribution as those generated using the equations. We define  $\phi$  as the rate of failure of this test averaged over 100 initial conditions. ii) a measure of the predictability horizon of the ESN forecast compared to the equations. For this purpose we use the root mean square error (RMSE) and define the saturation time  $\tau_s$  as the first time that RMSE exceeds a certain threshold  $s$  (defined as  $1/10$  of the average distance on the attractor. Results are independent on  $s$  provided that it is smaller than the average distance on the attractor and few orders of magnitude larger of the numerical precision.) iii) the initial error  $\eta = RMSE(t=1)$ . Equipped with these indicators we analyze two sets of simulations performed with and without moving average. The moving average operation is the integral of  $x(t)$  between  $t$  and  $t - w$ , where  $w$  is the window size of the moving average.

For the Lorenz attractor, we perform an ensemble of 100 ESN forecast varying both the noise intensity  $\epsilon$  and the network size  $N$ . In Fig. 1(a,c,e) ESN forecasts are run without moving average while in Fig. 1(b,d,f) we apply moving average with  $w = 10dt$  ( $dt = 0.001$ ). Panels (a,b) show  $\phi$ , the rate of failure of the  $\chi^2$  test averaged

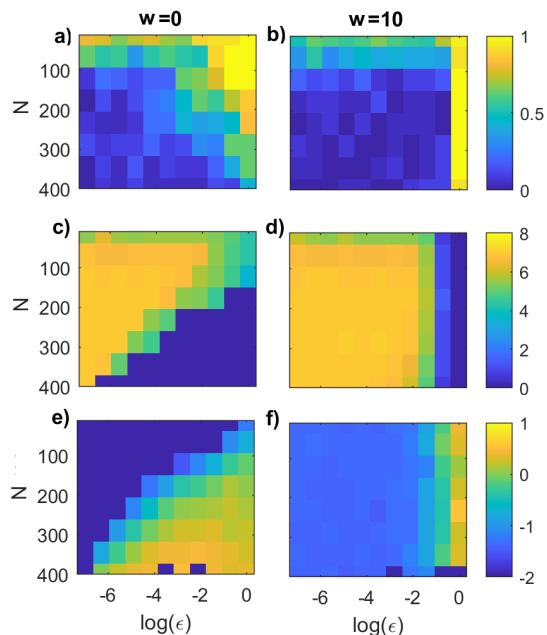


FIG. 1. Lorenz 1963 analysis for increasing noise intensity  $\epsilon$  ( $x$ -axes), and number of neurons  $N$  ( $y$ -axes). The colorscale represents:  $\phi$  the rate of failure of the  $\chi^2$  test (size  $\alpha = 0.05$ ) (a-b); the logarithm of saturation time  $\log(\tau_{s=1})$  (c,d); the logarithm of initial error  $\log(\eta)$  (e,f). All the values are averages over 100 realizations. Panels (a,c,e) refers to results without moving average, panels (b,d,f) with averaging window  $w = 10dt$ .

over the 100 initial conditions and computed considering only the  $x$  variable of the Lorenz 1963 system (results do not depend on the chosen variable). For increasing noise intensity and for small reservoir sizes the performances without moving average rapidly get worse. The moving average operation with  $w = 10$  greatly improves the three indicators except when the noise is too large (e.g.  $\epsilon = 1$ ). In this case, for any number of neurons, the moving average operation smooths too much the data. Furthermore the results exhibit only a weak dependence on the reservoir size. In [37] (section C and Figs S1-S4) we give additional details on those computations.

The analysis for the Pomeau-Manneville map (definition is provided in [37]-C) is shown in Fig. 2. We first observe that the forecast of the intermittent dynamics of the Pomeau-Manneville map is much more complicated than for the Lorenz system. This is a consequence of the intermittent behavior of this system. The ESN can never simulate an intermittent behavior, for all combinations of noise intensity and reservoir sizes. This is reflected in the behavior of the indicators. Independently from the network size considered, in the deterministic limit, the ESN fails to reproduce the invariant density in 80% of the cases ( $\phi \simeq 0.8$ ). For intermediate noise intensities  $\phi > 0.9$  (Fig. 2-a). The saturation time  $\tau_{s=1}$  for the short term forecast is small (Fig. 2c) as well as the initial error (Fig. 2e).

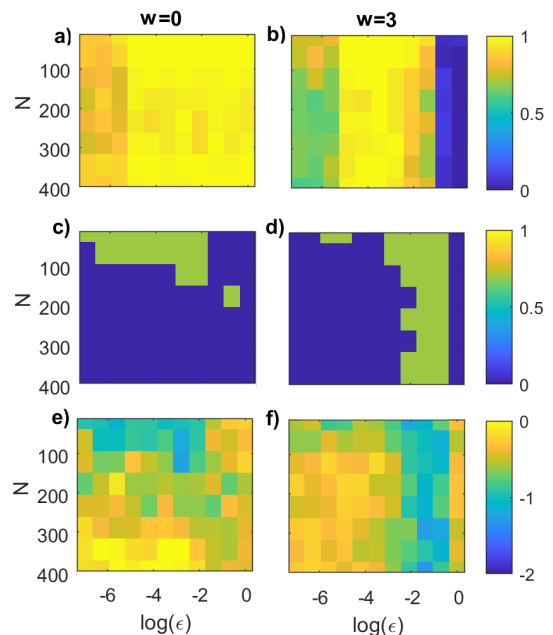


FIG. 2. Analysis of the Pomeau-Manneville system for increasing noise intensity  $\epsilon$  ( $x$ -axes), and number of neurons  $N$  ( $y$ -axes). The color scale represents:  $\phi$  the rate of failure of the  $\chi^2$  test (size  $\alpha = 0.05$ ) (a-b); the logarithm of saturation time  $\log(\tau_{s=1})$  (c,d); the logarithm of initial error  $\log(\eta)$  (e,f). All the values are average over 100 realizations. Panels (a,c,e) refers to results without moving average, panels (b,d,f) with averaging window  $w = 3$ .

The moving average procedure with  $w = 3$  (see [37]-C for a justification of this value) partially improves the performances (Fig. 2b,d,f) and it enables ESN to simulate an intermittent behavior (Fig. 3). Panel (a) shows the intermittent behavior of the data generated with the ESN trained on moving averaged data of Pomeau-Manneville system (red) and compare to the target time series (blue). Panel (b) displays the spectra which show in both cases a power law decay, typical of turbulent phenomena. Although the intermittent behavior is captured, this realization of ESN shows a different distribution of invariant density (not shown), concentrated around  $x = 0.5$  for the ESN prediction and around  $x = 0$  for the original data.

We now test the effectiveness of the moving average procedure in learning the behavior of turbulent and intermittent systems on climate data issued by reanalysis projects. We use data from the National Centers for Environmental Prediction (NCEP) version 2 [Donner la reference, e.g. Kisler et al., BAMS]. We consider the global sea-level pressure (SLP) fields every  $dt = 6$  h from 1979 to 31/08/2019 ([37]-C). Since these data are spatially extended, we need to investigate also the effect of spatial resolution on ESN performance. We find that higher predictability and better correspondence of long term behavior is achieved for a spatial coarse grain factor  $cg = 4$  (see [37]-C and Figs. S6-S7). Focusing on  $cg = 4$ , we extract randomly from the data a day in the first 25 years,

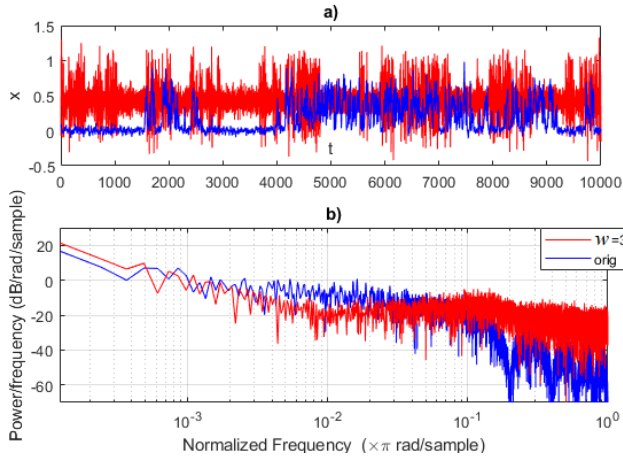


FIG. 3. Pomeau-Manneville ESN simulation obtained using the moving average time series with  $w = 3$  showing an intermittent behavior (red) and compared to the target trajectory (blue). Evolution of trajectories in time (a) and Fourier power spectra (b).

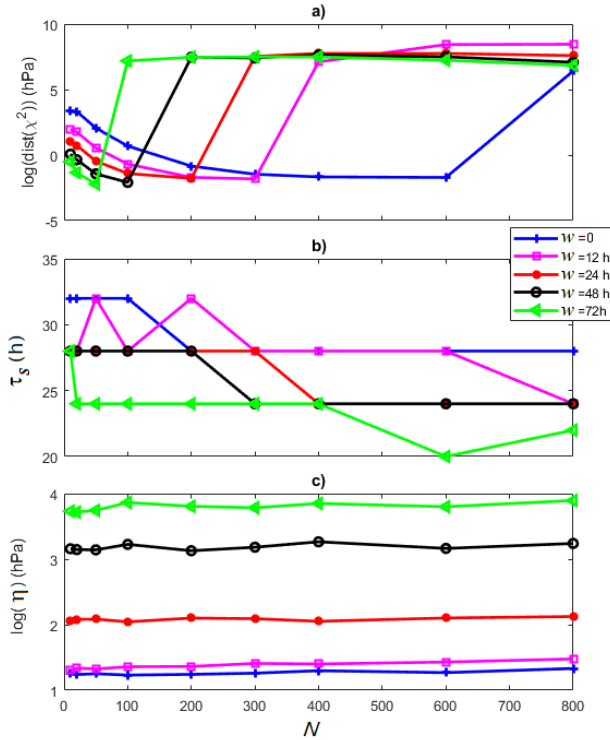


FIG. 4. ESN performance for the NCEP SLP 1979-2019 data. a)  $\chi^2$  distance; (b) the saturation time  $\tau_s$  in hours with  $s = 1.5$  hPa. (c) the logarithm of the initial error  $\eta$  in hPa. Averaging window sizes  $w$  are shown with different colors.

and then we train the ESN with 10 years of SLP data and start a long prediction of 5 years just after the training. We consider the same performance indicators introduced before. However, rather than the rate of failure of the  $\chi^2$  test, we show the  $\chi^2$  distance averaged over 100 long forecasts (Figure 4-a), for the global mean SLP. Results

are shown for different  $w$  (colors) and network sizes  $N$ . Fig. 4-b,c) show respectively  $\tau_s$  in hours, as a measure of the predictability horizon of ESN forecasts, and  $\log(\eta)$ . The results show a great dependence on  $N$  and  $w$ . Indeed the the behavior of the three indicators is not monotonic with  $N$ . Depending on  $w$ , there is a network size which causes a loss of prediction power due to overfitting. The moving average procedure improves the performance but only for small network sizes. Visual inspection shows that the best combination is found to be for  $w=12$  h and  $N = 200$ . We compare in details the results obtained for two 10 years prediction with  $w = 0$  h and  $w = 12$  h at  $N = 200$  and  $cg = 4$  fixed. At the beginning of the forecast time (Video S1), the target field (Video S1-a) is close to both what is obtained with  $w = 0$  h (Video S1-b) and  $w = 12$  h (Video S1-c). However, when looking after 24000 h (Video S2), of course we do not expect to see agreement among the three datasets. Indeed we are well beyond the predictability horizon. However, we remark that the dynamics for the run with  $w = 0$  h is steady: positions of cyclones and anticyclones barely evolve with time. Instead, the run  $w = 12$  h shows a richer dynamical evolution with generation and annihilation of cyclones. This visual impression is confirmed by the distribution and spectral analysis presented in ([37] Fig. S8).

We have analysed the performance of ESN in reproducing the behavior of noisy and intermittent data, as those measured as observables of geophysical flows. The performance of ESN in predicting both short and long term behaviors rapidly drops when the systems are perturbed with noise. However, we found that a good predictability is partially recovered when scale separation is performed. Here we have used a simple moving-average filter and shown that a careful choice of the moving-average window can enhance predictability. The results also apply to geophysical datasets: here we analysed the atmospheric circulation, represented by an SLP field. Heuristically, the moving-average operation helps both in smoothing the signal and by providing the ESN with a wider temporal information. In some sense, this is reminiscent of the embedding procedure [40], where the signal behavior is reconstructed by providing not only information on the previous time step, but on previous times depending on the complexity. In future works, it will be interesting to use other learning architectures and other methods of separating large- from small-scale components [41–43]. Our results give a more formal framework for applications of machine learning techniques on geophysical data. For example, [12, 44] have shown the capabilities of spatial separation of scales and the interest of learning subgrid dynamics. Indeed, several difficulties encountered in the application of machine learning on climate data could be overcome if the appropriate framework is used [19, 27, 29].

## ACKNOWLEDGMENTS

We acknowledge B D’Alena, J Brajard, Balaji, B Dubrulle, R Vautard, N Vercauteren, F Daviaud, Y Sato

for useful discussions. This work is supported by the CNRS INSU-LEFE-MANU grant ”DINCLIC”.

- 
- [1] M. I. Jordan and T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, *Science* **349**, 255 (2015).
- [2] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
- [3] D. S. Bassett and O. Sporns, Network neuroscience, *Nature neuroscience* **20**, 353 (2017).
- [4] C. Quinlan, B. Babin, J. Carr, and M. Griffin, *Business research methods* (South Western Cengage, 2019).
- [5] J.-X. Wang, J.-L. Wu, and H. Xiao, Physics-informed machine learning approach for reconstructing reynolds stress modeling discrepancies based on dns data, *Physical Review Fluids* **2**, 034603 (2017).
- [6] J. A. Carlson, A. Jaffe, and A. Wiles, *The millennium prize problems* (American Mathematical Society Providence, RI, 2006).
- [7] M. Buchanan, *The limits of machine prediction*, Ph.D. thesis, Nature Publishing Group (2019).
- [8] S. Bony, B. Stevens, D. M. Frierson, C. Jakob, M. Kageyama, R. Pincus, T. G. Shepherd, S. C. Sherwood, A. P. Siebesma, A. H. Sobel, *et al.*, Clouds, circulation and climate sensitivity, *Nature Geoscience* **8**, 261 (2015).
- [9] J.-L. Wu, H. Xiao, and E. Paterson, Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework, *Physical Review Fluids* **3**, 074602 (2018).
- [10] V. M. Krasnopolsky and M. S. Fox-Rabinovitz, Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction, *Neural Networks* **19**, 122 (2006).
- [11] S. Rasp, M. S. Pritchard, and P. Gentine, Deep learning to represent subgrid processes in climate models, *Proceedings of the National Academy of Sciences* **115**, 9684 (2018).
- [12] P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, Could machine learning break the convection parameterization deadlock?, *Geophysical Research Letters* **45**, 5742 (2018).
- [13] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, Convolutional lstm network: A machine learning approach for precipitation nowcasting, in *Advances in neural information processing systems* (2015) pp. 802–810.
- [14] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, Deep learning for precipitation nowcasting: A benchmark and a new model, in *Advances in neural information processing systems* (2017) pp. 5617–5627.
- [15] M. Sprenger, S. Schemm, R. Oechslin, and J. Jenkner, Nowcasting foehn wind events using the adaboost machine learning algorithm, *Weather and Forecasting* **32**, 1079 (2017).
- [16] J. N. Liu, Y. Hu, Y. He, P. W. Chan, and L. Lai, Deep neural network modeling for big data weather forecasting, in *Information Granularity, Big Data, and Computational Intelligence* (Springer, 2015) pp. 389–408.
- [17] A. Grover, A. Kapoor, and E. Horvitz, A deep hybrid model for weather forecasting, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2015) pp. 379–386.
- [18] S. E. Haupt, J. Cowie, S. Linden, T. McCandless, B. Kosovic, and S. Alessandrini, Machine learning for applied weather prediction, in *2018 IEEE 14th International Conference on e-Science (e-Science)* (IEEE, 2018) pp. 276–277.
- [19] S. Scher and G. Messori, Predicting weather forecast uncertainty with machine learning, *Quarterly Journal of the Royal Meteorological Society* **144**, 2830 (2018).
- [20] H. Jaeger, 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**, 13 (2001).
- [21] E. N. Lorenz, Deterministic nonperiodic flow, *Journal of the atmospheric sciences* **20**, 130 (1963).
- [22] J. M. Hyman and B. Nicolaenko, The kuramoto-sivashinsky equation: a bridge between pde’s and dynamical systems, *Physica D: Nonlinear Phenomena* **18**, 113 (1986).
- [23] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, Using machine learning to replicate chaotic attractors and calculate lyapunov exponents from data, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 121102 (2017).
- [24] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, *Physical review letters* **120**, 024102 (2018).
- [25] M. Xu, M. Han, T. Qiu, and H. Lin, Hybrid regularized echo state network for multivariate chaotic time series prediction, *IEEE transactions on cybernetics* **49**, 2305 (2018).
- [26] E. N. Lorenz, Predictability: A problem partly solved, in *Proc. Seminar on predictability*, Vol. 1 (1996).
- [27] T. Schneider, S. Lan, A. Stuart, and J. Teixeira, Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, *Geophysical Research Letters* **44**, 12 (2017).
- [28] S. Scher, Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning, *Geophysical Research Letters* **45**, 12 (2018).
- [29] P. D. Dueben and P. Bauer, Challenges and design choices for global weather and climate models based on machine learning, *Geoscientific Model Development* **11**, 3999 (2018).
- [30] S. Scher and G. Messori, Weather and climate forecasting with neural networks: using general circulation models (gcms) with different complexity as a study ground,

- Geoscientific Model Development **12**, 2797 (2019).
- [31] G. Paladin and A. Vulpiani, Degrees of freedom of turbulence, *Physical Review A* **35**, 1971 (1987).
- [32] D. Schertzer, S. Lovejoy, F. Schmitt, Y. Chigirinskaya, and D. Marsan, Multifractal cascade dynamics and turbulent intermittency, *Fractals* **5**, 427 (1997).
- [33] P. Manneville, Intermittency, self-similarity and 1/f spectrum in dissipative dynamical systems, *Journal de Physique* **41**, 1235 (1980).
- [34] R. E. Eskridge, J. Y. Ku, S. T. Rao, P. S. Porter, and I. G. Zurbenko, Separating different scales of motion in time series of meteorological variables, *Bulletin of the American Meteorological Society* **78**, 1473 (1997).
- [35] Z. Shi and M. Han, Support vector echo-state machine for chaotic time-series prediction, *IEEE Transactions on Neural Networks* **18**, 359 (2007).
- [36] D. Li, M. Han, and J. Wang, Chaotic time series prediction based on a novel robust echo state network, *IEEE Transactions on Neural Networks and Learning Systems* **23**, 787 (2012).
- [37] See Supplemental Material at [URL will be inserted by publisher] for [give brief description of material].
- [38] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the lasso and generalizations* (Chapman and Hall/CRC, 2015).
- [39] L.-S. Young, What are srj measures, and which dynamical systems have them?, *Journal of Statistical Physics* **108**, 733 (2002).
- [40] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, *Physica D: Nonlinear Phenomena* **110**, 43 (1997).
- [41] S. Wold, K. Esbensen, and P. Geladi, Principal component analysis, *Chemometrics and intelligent laboratory systems* **2**, 37 (1987).
- [42] G. Froyland, G. A. Gottwald, and A. Hammerlindl, A computational method to extract macroscopic variables and their dynamics in multiscale systems, *SIAM Journal on Applied Dynamical Systems* **13**, 1816 (2014).
- [43] F. Kwasiok, The reduction of complex dynamical systems using principal interaction patterns, *Physica D: Nonlinear Phenomena* **92**, 28 (1996).
- [44] T. Bolton and L. Zanna, Applications of deep learning to ocean data inference and subgrid parameterization, *Journal of Advances in Modeling Earth Systems* **11**, 376 (2019).