Dynamical error metrics for machine learning forecasting

Zhou Fang [a], **Gianmarco Mengaldo** [a]

[a] *National University of Singapore, Singapore, 119077* zhou.fang@u.nus.edu, mpegim@nus.edu.sg

* Presenting author

1. Introduction

Forecasting the future states of a system is a fundamental challenge in applied science and engineering, with applications as diverse as atmosphere and ocean dynamics, astrophysics, biology, and finance, to cite a few. Traditionally, this problem has been approached using numerical methods that approximate the solution of the governing equations [1, 2, 3, 4, 5], thereby providing the future evolution of the underlying system,

More recently, the emergence of machine learning (ML) has led to a paradigm shift in forecasting. ML models learn complex patterns directly from data, often without enforcing the underlying equations, providing an alternative to equation-based models. ML models have demonstrated the ability to achieve accurate forecasts for both canonical dynamical systems [6] and real-world applications, such as weather [7, 8, 9, 10] and climate [11, 12, 13]. However, they often struggle to accurately capture fine-scale structures in long-term predictions [14] due to their known spectral bias. Additionally, they may exhibit instability or unphysical behavior, limiting their reliability in high-fidelity applications.

Several promising strategies have emerged to address these challenges, including physics-informed ML approaches [15], as well as explicit physical constraints that enforce the conservation of key physical quantities [13]. Nevertheless, functioning as a black box, determining whether the model forecasts adhere to established physical principles remains challenging. Commonly used evaluation metrics primarily measure the difference between predictions and actual target values (e.g., mean squared error and its variants [16, 17]), without assessing the fidelity of the forecasts under other metrics.

An important aspect that is critically underexplored is the dynamical consistency of the forecasts. In this work, we focus on this latter aspect, whereby we use dynamical indices (DI), namely instantaneous dimension (d) and inverse persistence (θ) derived from dynamical systems theory [18] to assess dynamical consistency of ML forecasts.

We analyze the dynamical differences in both direct and recursive ML forecasts across widely used model architectures, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks [19], Transformers [20, 21], and Graph Neural Networks (GNNs) [22] for 3 synthetic datasets Lorenz 63, Kuramoto–Sivashinsky (KS) equation and Kolmogorov flow. Additionally, for weather forecasting, we consider state-of-the-art models, specifically the Transformer-based Pangu-Weather [8] and the GNN-based GraphCast [7].

We show that d and θ are closely related to commonly used error measures, such as mean squared error (MSE). Moreover, we propose d- and θ- based error metrics, which provide complementary information on the dynamical consistence of ML forecasts.

2. Results and discussion

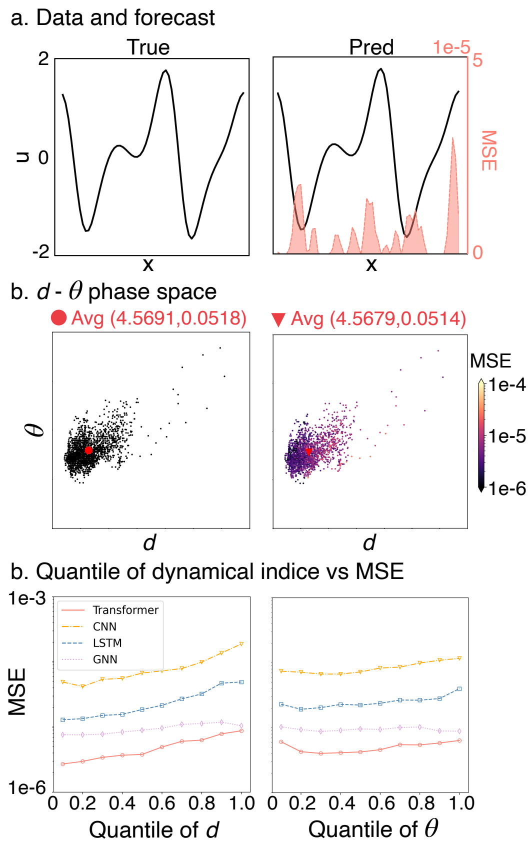

Fig. 1: Direct forecast error and dynamical properties of KS equation

For single-step forecasting, figure 1 presents an overview of (a) the datasets and forecast samples, (b) the d-θ phase space, and (c) the relationship between dynamical characteristics and MSE on the KS equation dataset. Notably, we observe that MSE tends to increase for high values of d and θ, indicating that greater dynamical complexity (high d) and lower persistence (high θ) are strong predictors of larger forecast errors. This trend also holds for other standard error metrics, such as mean absolute error (MAE) and root mean square error (RMSE), and extends to the other datasets investigated. For recur-

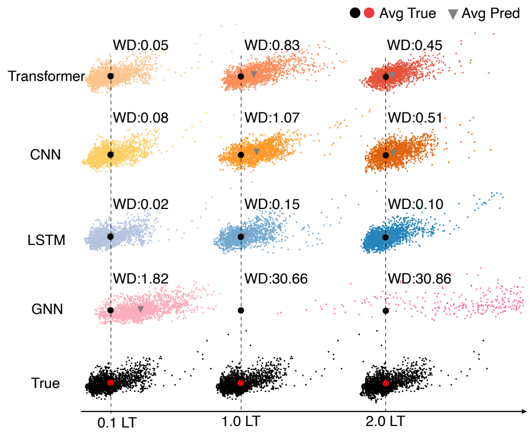

Fig. 2: The true and predicted dynamical phase space for recursive forecasts

sive forecasting, Figure 2 illustrates the true and predicted states at different forecast times within the dynamical phase space. The mean values of d and θ are indicated in the plot, along with the Wasserstein distance (WD), which quantifies the difference between the true and predicted distributions. While the predicted d and θ closely resemble those of the true system over short forecast periods, the error grows significantly as the forecast horizon extends, indicating poor dynamical consistency. Furthermore, we observe substantial variation across model architectures. For instance, the GNN fails to preserve the shape of attractor in dynamical phase space, after forecast horizon reaches 1 LT.

3. Conclusion

In this work, we investigate the dynamical consistency of ML forecasts, demonstrating that the dynamics of the underlying data is intrinsically linked to the forecast errors. In particular, higher values of d and θ are associated with increased forecast errors. Moreover, we introduced error metrics based on d and θ as a measure of dynamical consistency.

Our findings underscore the effectiveness of dynamical indices as diagnostic tools for ML forecasts, complementing traditional error metrics such as MAE, MSE, and RMSE. Our proposed approach advances the goal of better assessing the fidelity of ML forecasts by providing error metrics from a dynamical consistency perspective.

4. Data and methods

4.1 Datasets description

Lorenz 63 system The Lorenz system is a well-known chaotic dynamical system governed by the set of ordinary differential equations given in Eq. 1 [23]

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y-x)\\ \frac{dy}{dt} &= \rho(x-z)-y \\ \frac{dz}{dt} &= xy-\beta z.\end{aligned} \tag{1}$$

It is widely used as a benchmark for forecasting tasks. The Lorenz 63 system is typically simulated using the standard parameter values: $\sigma = 10$, $\rho = 28$, $\beta = 2.667$.

Kuramoto–Sivashinsky equation The one-dimensional (1D) Kuramoto–Sivashinsky (KS) equation is a fourth-order partial differential equation (PDE) of the form given in Eq. 2

$$u_t + uu_x + u_{xx} + u_{xxxx} = 0, \tag{2}$$

describing the evolution of a spatiotemporal system. Here, u represents the observable, t denotes time, and x is the spatial coordinate defined over $[0, L)$.

Kolmogorov flow Kolmogorov flow is a 2D shear flow governed by Navier-Stokes equations with periodic Kolmogorov forcing, as shown in Eq. 3

$$\begin{aligned} u_t + u\cdot\nabla u &= -\nabla p + \frac{1}{Re}\nabla^2 u + sin(ny)\hat{x} \\ \nabla u &= 0.\end{aligned} \tag{3}$$

Here, u represents the observable, t denotes time, p is the pressure, and $sin(ny)$ represents the external force at a given frequency.

Regional Weather The fifth-generation ECMWF (European Centre for Medium-Range Weather Forecasts) atmospheric reanalysis (ERA5) [24] is used as the ground truth. We consider the state-of-the-art weather forecast models Pangu-Weather[8] and GraphCast [7] trained on ERA5 data. The model forecasts are downloaded from WeatherBench2 [25].

4.2 Dynamical indices

Dynamical indices offer a mathematically rigorous and purely data-driven framework for analyzing the local, instantaneous, and state-dependent dynamical properties of complex systems [26]. This framework includes two dynamical indices: (i) the local dimension d, which provides information on the system's dynamical complexity, and (ii) θ, describing the clustering of recurring dynamical paths around a certain state, which is also known as reciprocal of the local persistence time $\Theta = \theta^{-1}$. For a more detailed mathematical formulation, we refer to [18].

Several works have shown that a specific combination of d and θ can provide useful dynamical and physical insights across various disciplines, including atmospheric sciences [18, 27, 28], oceanography [29], and fluid mechanics [30].

Acknowledgments

GM acknowledges support from MOE Tier 2: Prediction-to-Mitigation with Digital Twins of the Earth's Weather, under grant number # T2EP50221-0017.

References

[1] Yvon Maday and Anthony T. Patera. Spectral element methods for the incompressible navier-stokes equations. *State-of-the-art surveys on computational mechanics*, 4:71–143, 1989.

[2] Thomas J. R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover Publications, 2000.

[3] George E. Karniadakis and Spencer J. Sherwin. *Spectral/hp Element Methods for Computational Fluid Dynamics*. Oxford University Press, 2005.

[4] Randall J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*. Society for Industrial and Applied Mathematics (SIAM), 2007.

[5] Alfio Quarteroni and Alberto Valli. *Numerical Approximation of Partial Differential Equations*. Springer, 2008.

[6] Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical review letters*, 120(2):024102, 2018.

[7] Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023.

[8] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, 2023.

[9] Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan Weyn, Haiyu Dong, Anna Vaughan, Jayesh K. Gupta, Kit Tambiratnam, Alex Archibald, Elizabeth Heider, Max Welling, Richard E. Turner, and Paris Perdikaris. Aurora: A foundation model of the atmosphere. *arXiv preprint arXiv:2405.13063*, 2024.

[10] Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, et al. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025.

[11] Xiao Wang, Siyan Liu, Aristeidis Tsaris, Jong-Youl Choi, Ashwin M Aji, Ming Fan, Wei Zhang, Junqi Yin, Moetasim Ashfaq, Dan Lu, et al. Orbit: Oak ridge base foundation model for earth system predictability. In *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–11. IEEE, 2024.

[12] Oliver Watt-Meyer, Brian Henn, Jeremy McGibbon, Spencer K Clark, Anna Kwa, W Andre Perkins, Elynn Wu, Lucas Harris, and Christopher S Bretherton. Ace2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *arXiv preprint arXiv:2411.11268*, 2024.

[13] Xin Wang, Juntao Yang, Jeff Adie, Simon See, Kalli Furtado, Chen Chen, Troy Arcomano, Romit Maulik, and Gianmarco Mengaldo. Condensnet: Enabling stable long-term climate simulations via hybrid deep learning models with adaptive physical constraints. *arXiv preprint arXiv:2502.13185*, 2025.

[14] Dibyajyoti Chakraborty, Arvind T Mohan, and Romit Maulik. Binned spectral power loss for improved prediction of chaotic systems. *arXiv preprint arXiv:2502.00472*, 2025.

[15] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[16] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[17] Andrei Botchkarev. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint arXiv:1809.03006*, 2018.

[18] Davide Faranda, Gabriele Messori, and Pascal Yiou. Dynamical proxies of north atlantic predictability and extremes. *Scientific reports*, 7(1):41278, 2017.

[19] Hasim Sak, Andrew W Senior, Françoise Beaufays, et al. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Interspeech*, volume 2014, pages 338–342, 2014.

[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[21] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[22] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[23] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[24] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

[25] Stephan Rasp, Stephan Hoyer, Alexander Merose, Ian Langmore, Peter Battaglia, Tyler Russel, Alvaro Sanchez-Gonzalez, Vivian Yang, Rob Carver, Shreya Agrawal, Matthew Chantry, Zied Ben Bouallegue, Peter Dueben, Carla Bromberg, Jared Sisk, Luke Barrington, Aaron Bell, and Fei Sha. Weatherbench 2: A benchmark for the next generation of data-driven global weather models, 2023.

[26] Valerio Lucarini, Davide Faranda, Jorge Miguel Milhazes de Freitas, Mark Holland, Tobias Kuna, Matthew Nicol, Mike Todd, Sandro Vaienti, et al. *Extremes and recurrence in dynamical systems*. John Wiley & Sons, 2016.

[27] Chenyu Dong, Gabriele Messori, Davide Faranda, Adriano Gualandi, Valerio Lucarini, and Gianmarco Mengaldo. Multiscale dynamical indices reveal scale-dependent atmospheric dynamics. *arXiv preprint arXiv:2412.10069*, 2024.

[28] Chenyu Dong, Davide Faranda, Adriano Gualandi, Valerio Lucarini, and Gianmarco Mengaldo. Revisiting the predictability of dynamical systems: a new local data-driven approach. *arXiv preprint arXiv:2409.14865*, 2024.

[29] Guangpeng Liu, Fabrizio Falasca, and Annalisa Bracco. Dynamical characterization of the loop current attractor. *Geophysical Research Letters*, 48(24):e2021GL096731, 2021.

[30] Gabriele Messori, Nili Harnik, Erica Madonna, Orli Lachmy, and Davide Faranda. A dynamical systems characterisation of atmospheric jet regimes. *Earth System Dynamics Discussions*, 2020:1–23, 2020.