DAMTP 3rd Term Report - Reduced Order Models for Stochastic Atmospheric Processes

I. J. S. Shokar - Supervised by Professor P. H. Haynes & Professor R. R. Kerswell

1 Introduction and Existing Literature

1.1 Research Motivation

Climate and weather forecasting has developed significantly with the increase in data available [1], however operational forecast models of geophysical fluid dynamics, namely Global Circulation Models (GCMs) [2], are complex and computationally expensive, with many run on the world's largest supercomputers [3]. This computational constraint requires many processes that take place at spatial or temporal resolution smaller than that of the model to be parameterised. Traditionally these parameterisations have been physics-based, however we look to produce a data-driven modelling approach in the form of Reduced Order Models (ROMs) of atmospheric processes that provide the same information but at a reduced computational expense.

1.2 Beta-Plane System

The complexity of GCMs makes studying processes in isolation more difficult so we use an idealised model that provides an analogue for variability of tropospheric mid-latitudinal dynamics, the driver for European weather - the Beta-Plane system [4]. The system lies on a 2D plane, considering the troposphere as a shallow layer when compared to the horizontal scale. A lack of baroclinicity due to the absence of stratification results in the requirement for small-scale eddies, that generate turbulence, to be parameterised by a stochastic forcing, ξ, which is rapidly decorrelating, spatially homogeneous and random in space and time. The equation of motion is:

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = \xi - \mu\zeta + \nu_n \nabla^{2n} \zeta \tag{1.1}$$

where ζ is relative vorticity, defined by $\zeta = \nabla^2 \psi$, with ψ the stream-function. μ is the linear damping rate, with ν the hyper-viscosity rate, order n and J is the Jacobian, with β, the Rossby parameter, being the coefficient of a first-order approximation of the Coriolis parameter.

Interested in zonally-oriented flows, Figure 1 shows a latitude-time plot of the zonal mean zonal velocity, $\bar{u}(y,t)$, where the velocity field is $(u, v) = (-\partial_y \psi, \partial_x \psi)$, where jets form, displaying merging, nucleating and wandering behaviour. Jet streams have a major influence over regional weather patterns, transporting quantities such as momentum and heat, as well as tracers, such as ozone and water vapour [5], motivating studying their behaviour.

1.3 Data-Driven Modelling

Data-driven methods fit a model via training on data as opposed to numerical modelling based on physical equations. In this research we implement Deep Learning (DL) [6], a subset of Machine Learning, that uses layers of artificial neurons and representation learning to obtain the correct model fit. Research has been conducted to use DL to replace model-based parameterisations- utilising the computational speed-up that DL offers [7]. These have looked at emulating various different atmospheric processes with DL such as radiation schemes [8, 9], convection [10] and cloud resolving models [11], however these are deterministic models, not able to capture the stochasticity that drives dynamics in Beta-Plane system.

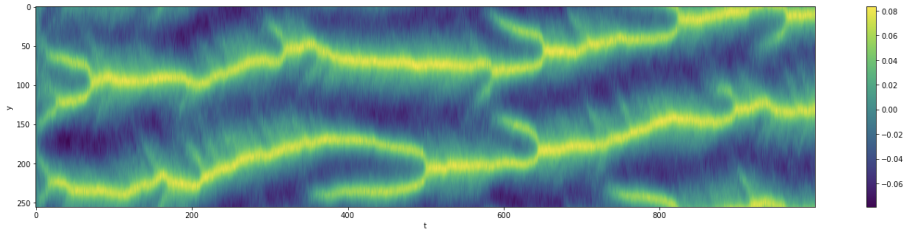

Figure 1: *An example latitude-time plot of the Beta-Plane system displaying $\bar{u}(y,t)$ over 1,000 observable time-steps from a numerical integration.*

Stochastic parameterisations have been introduced to represent probability distributions of sub-grid-scale tendencies [12] and have been shown to reduce the systematic error [13], while providing better estimates of uncertainty than that of multi-simulator ensembles. Recently, Generative Adversarial Networks (GANs), which have been very effective for super-resolution tasks, and resultingly statistical downscaling tasks [14, 15] as well as now-casting realistic precipitation maps [16], have been used for stochastic parameterisations [17]. However, GANs use an adversarial training method [18], making them black-box-models, motivating our approach to produce interpretable ROMs that provide stochastic parameterisations.

1.4 Reduced Order Modelling

Here ROMs are models with a smaller number of degrees-of-freedom, corresponding to reduced computational cost. One form of ROM, Manifold Learning [19] finds the form of an inertial manifold on which the underlying dynamics of the system lie, with research showing that DL, which fits a series of non-linear functions to the data [20], captures complex patterns more effectively than methods such as Proper Orthogonal Decomposition (POD) that only recover linear manifolds. POD has previously been coupled with time evolution models [21], such as Recurrent Neural Networks (RNN) [22], and we assumed that coupling an Autoencoder, a DL model with an information bottleneck, with such models should produce a far better ROM than POD at equivalent order.

Our previous study showed that an Autoencoder produced a reconstruction error 50 times smaller than POD ($RMSE_{Autoencoder} = 0.00018$, $RMSE_{POD} = 0.0091$, with the size of the latent space $h = 16$), however when coupled to an RNN the model could not could not effectively give predictions, as an Autoencoder searches for a stationary manifold, while the dynamics of the Beta-Plane system are driven stochastically. Others have looked methods for finding a manifold for stochastic dynamical systems [23, 24, 25, 26], however we believed that a variational DL model, one that samples from a probability distribution for each latent variable, could better model the system.

2 Recent Work

2.1 Numerical Reduced Order Modelling: Quasilinear Zonal Wavenumber Truncation

One form of numerical ROMs is a quasilinear (QL) approximation [27]. QL only permits interactions between the mean-flow and the Reynolds stresses, omitting eddy-eddy interactions. Using a Reynolds Decomposition, where $u = \bar{u} + u'$, we obtain the Reynolds Stress term, u', and given our study of zonal flows, the mean flow defined as the zonal average, $\bar{u}$. This lack of scattering leads prevents cascades in the zonal wavenumbers, leading us to believe that a wavenumber truncation in this direction could be made, that still retained the coherent structures. Other research had shown that it was valid to make a QL approximation of the Beta-Plane system [28, 29, 30].

This truncation reduces the number of grid points over which the integration (taking place over the same square, two-dimensional, doubly-periodic domain: $(x, y) \in [-\pi, \pi)^2$) is made. The full system uses

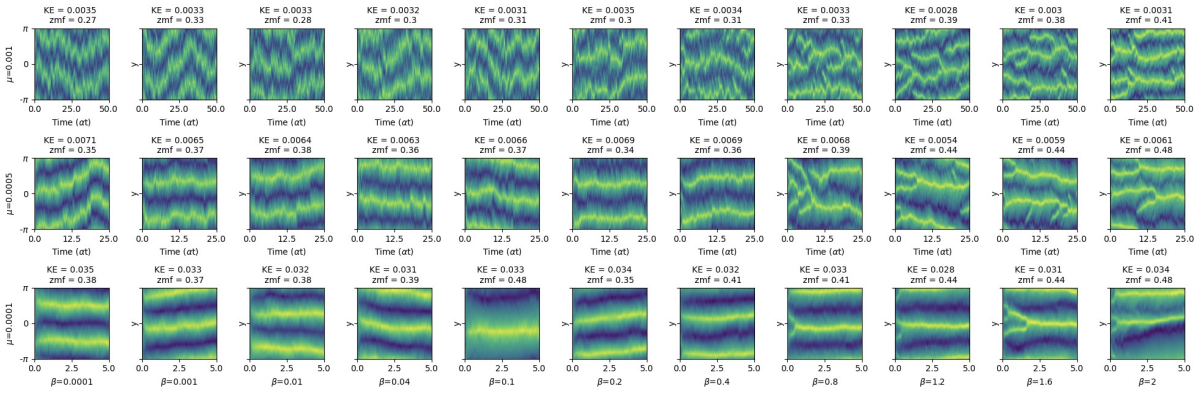

Figure 2: *A collection of latitude-time plots for a QL system where only 4 zonal wavenumbers were forced, with merging and nucleating events most clearly observed when $\mu = 0.001$ and $0.8 < \beta < 1.6$. In order to evaluate whether we were in the correct parameter regime to make comparisons to the full QL system, we vary both β, and the damping rate, μ. μ was varied to control the KE of the system, due to numerical instabilities that occurred when ε, the energy injection rate, was outside of a confined range.*

N=256 grid-points in x and y, however we here impose that the number of grid points in x, the zonal direction, is N_x and retaining full resolution y is $N_y = 256$ with $N_x < N_y = 256$. The goal was to both reduce computation of the model, to more efficiently produce training examples for a DL model, as well to have a fewer degrees-of-freedom representation of the system, potentially making the system more analytically tractable.

As the full QL system is forced by an annulus in spectral space, with a principal forcing wavenumber $\mathbf{k} = 16$, we must alter the forcing structure given the reduced zonal wavenumbers. After observing a lack of jet merging or nucleation events when only forcing single zonal wavenumber forced (associated forced wavenumbers $|k_y| < 16$), as well as two zonal wavenumbers, we implemented a forcing structure which projected the shape of the annulus forcing onto our reduced set of wavenumbers in x, the details of which are omitted for this report. Under this scheme we found that a minimum of 4 zonal wavenumbers were required to see behaviour like that of the full system, as seen in Figure 2, which corresponded to a computational speed up of 4.3 times.

2.2 Deep Learning: Variational Autoencoder

Secondarily, we used a Variational Autoenoder (VAE) [32] whereby the latent representation is mapped to a Gaussian distribution, defined by a mean, μ, and standard deviation, σ, for each latent variable, rather than a fixed vector as with an Autoencoder, allowing for a sample $z = \mu + \sigma\epsilon$ where $\epsilon \sim \mathcal{N}(0, 1)$ to be taken from the distribution. The loss function of the Autoencoder is that of the Mean Square Error (MSE), while that most commonly used for a VAE is the evidence lower bound (ELBO), which is a is a combination of the negative log-likelihood function (which under an assumption of Gaussian distribution is equivalent to the MSE) and the Kullback-Leibler (KL) divergence, which is a measure of loss when the distribution $q(x)$ is used to approximate $p(x)$:

$$-ELBO = MSE(p, q) + \beta D_{KL}(p, q) = \sum_{i=1}^{N} \|p(x_i) - q(x_i)\|^2 + \beta \mathbb{E}_q[logp(x_i) - logq(x_i)] \tag{2.1}$$

where N is the training batch size and $\mathbb{E}$, the expectation function. Here a β term is introduced to provide a weighting between these two measures of loss, and a value of $\beta = 0.01$, was used in the following results. A schematic for the VAE can be seen in Figure 3.

The VAE is coupled to a time evolution model in the encoder block, here an RNN [33], that takes in a time-series of length l as initial conditions. The model encodes the input and makes a prediction of the possible next time-step via sampling in latent space, before that is decoded to provide a forecast. The length of l can be varied, with $l = 1$ corresponding to a Markovian model. This prediction is then used in the input sequence for prediction of the next time-step, allowing the model to make predictions beyond the initial data provided.

Figure 4 shows that the ensemble of predictions provide plausible forecasts, when compared to the truth. Of course, the truth represents just one possible trajectory from a single realisation of the stochastic forcing in the numerical integration, with different realisations potentially not exhibiting the merging event at $t = 46$. Each prediction took 4 seconds to produce 50 future time-steps, while the numerical integration takes ∼27 minutes to produce a time-series over the same period, ∼400 times faster.

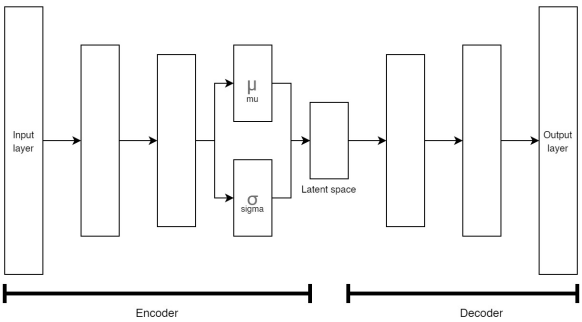

Figure 3: *A schematic showing the architecture of a VAE. [31] The mean, μ, and standard deviation, σ, are provided for each latent variable in the latent space, size h. Information from the inputs is reduced via the encoder block which reduces the number of neurons in each proceeding layer down to latent space and is then decoded to provide an output. In our model, the encoder is coupled to a RNN, repeating the encoder block for each member of the time-series, with a forecast of the underlying dynamics made in the latent space, before being decoded.*

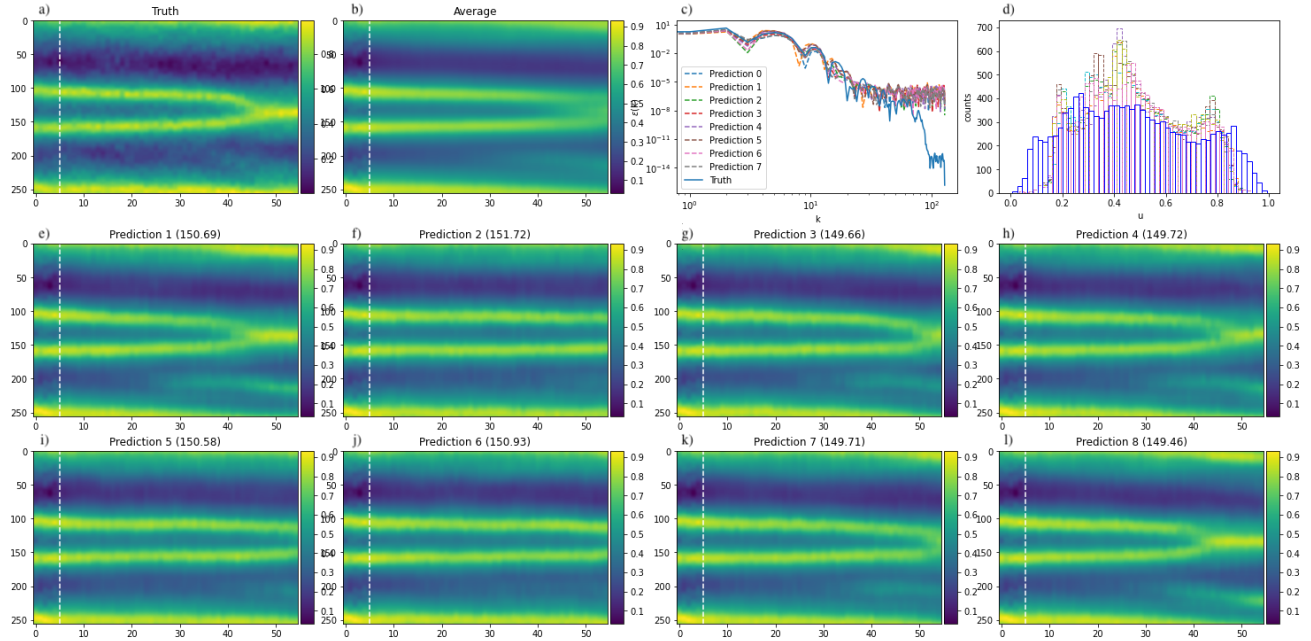

Figure 4: *A series of plots displaying the output from the VAE-RNN over a forecast period of 50 time-steps. a) Shows the Truth data obtained via numerical integration, used for verification. e-l) Show an ensemble of outputs from the VAE, having been initially shown the 5 time-steps to the left of the dotted line as initial conditions, each forecasting the next 50 time-steps. The values in brackets above each prediction show the negative-log likelihood between each prediction and the Truth b) Shows the mean over the ensemble. c) Shows a Power Spectra for the truth data as well as the ensemble; the legend applies to both c) and d) with the truth shown with the solid blue line and the ensemble members indicated by dashed lines of various colours. d) Shows a histogram of values of $\bar{u}$ over the forecast period to compare distributions.*

3 Future Work

Next steps are to explore variations of architecture of the VAE-RNN, as well as alternative training objectives. This could be the use of reservoir computing [34] for the time evolution. This only requires the training of a single readout layer that maps from a reservoir network to the desired output, significantly reducing training time. Another addition could be that of elements from Physics-Informed Neural-Networks, a set of methods that use constraints when training DL models, such as physical laws, as regularisation agents that limit the space of solutions [35, 36].

A key question is that of quantifying the forecast skill, as the truth data used for comparison is only one possible realisation of the flow. Figure 4 shows the power spectra as well as a histogram for the truth data as well as the prediction ensemble, and other statistical tests will be looked at in order to verify the forecast skill. Another metric is whether the spread of forecasts from our ensemble resembles that of numerical integration for different forcing realisations, and a comparative experiment will be conducted to verify this.

As an emulator of this physical process there are questions over the VAE-RNN's numerical stability over long time-periods, as well as how the VAE will deal with forecasting rare events, with neural-networks being interpolative- fitting to the training data. To combat this, sampling methods as well a feedback loop between the model and training data generation could be explored [37].

The ROM approach provides a latent space that could yield insight into both model robustness as well as underlying system dynamics. One question that arises is that of quantification of stability of different system states - how predictable are they. We will look to use both the ability to cheaply produce an ensemble as well as the latent variables to try and identify states stability [38] to produce a method that corresponds to considering Lyapunov exponents for a stochastic dynamical system [39].

Once a robust model has been finalised, future work could include forecasting of the full velocity field of the Beta-Plane system as well as a two-layer model that can generate turbulence without stochastic forcing, comparative exploration between the two systems will be made and it may be that the 3D model is still best captured best by a stochastic ROM due to sub-grid variability. This could also include the exploration of other geophysical process models and parameterisations including those used in operational forecasts.

References

[1] Hossein Hassani, Xu Huang, and Emmanuel Silva. Big data and climate change. *Big Data and Cognitive Computing*, 3(1), 2019.

[2] MacCracken M. C. Grotch, S. L. The use of general circulation models to predict regional climatic change. *Journal of Climatet*, 4, 1991.

[3] Huize Wang and Robin Wordsworth. Extremely long convergence times in a 3d GCM simulation of the sub-neptune gliese 1214b. *The Astrophysical Journal*, 891(1):7, feb 2020.

[4] Frank B. Lipps. A note on the beta-plane approximation. *Tellus*, 1964.

[5] Cope. L. The dynamics of geophysical and astrophysical turbulence. *Thesis, DAMTP, University of Cambridge. 2021.*

[6] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, Jan 2015.

[7] Tapio Schneider, Shiwei Lan, Andrew Stuart, and João Teixeira. Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical research letters*, 44(24):12,396–12,417, 2017.

[8] F. Chevallier, F. Chéruy, N. A. Scott, and A. Chédin. A Neural Network Approach for a Fast and Accurate Computation of a Longwave Radiative Budget. *Journal of Applied Meteorology*, 37(11):1385–1397, November 1998.

[9] Vladimir M. Krasnopolsky, Michael S. Fox-Rabinovitz, and Dmitry V. Chalikov. New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model. *Monthly Weather Review*, 133(5):1370 – 1383, 2005.

[10] Pierre Gentine, M. Pritchard, Stephan Rasp, Gael Reinaudi, and Galen Yacalis. Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45:5742 – 5751, 2018.

[11] Stephan Rasp, Michael S. Pritchard, and Pierre Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, 2018.

[12] TN Palmer. Towards the probabilistic earth-system simulator: A vision for the future of climate and weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 138(665):841–861, 2012.

[13] T.N. Palmer, Roberto Buizza, F. Doblas-Reyes, T. Jung, Martin Leutbecher, G.J. Shutts, M. Steinheimer, and Antje Weisheimer. Stochastic parametrization and model uncertainty. (598):42, 10 2009.

[14] Jussi Leinonen, Daniele Nerini, and Alexis Berne. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7211–7223, 2021.

[15] Jussi Leinonen, Daniele Nerini, and Alexis Berne. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7211–7223, sep 2021.

[16] Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson Ellen Clancy, Alberto Arribas†, and Shakir Mohamed. Skillful precipitation nowcasting using deep generative models of radar. *Nature*, 597:672–677, 2021.

[17] David John Gagne, Hannah M. Christensen, Aneesh C. Subramanian, and Adam H. Monahan. Machine learning for stochastic parameterization: Generative adversarial networks in the lorenz '96 model. *Journal of Advances in Modeling Earth Systems*, 12(3), mar 2020.

[18] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[19] Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature Communications*, 12(1):1417–1417, 2021.

[20] Alec J. Linot and Michael D. Graham. Deep learning to discover and predict dynamics on an inertial manifold. *CoRR*, abs/2001.04263, 2020.

[21] Clarence W. Rowley, Tim Colonius, and Richard M. Murray. Model reduction for compressible flows using pod and galerkin projection. *Physica D: Nonlinear Phenomena*, 189(1):115–129, 2004.

[22] Zhenghai Wang, Dunhui Xiao, Fangxin Fang, R. Govindan, and Christopher Pain. Model identification of reduced order fluid dynamics systems using deep learning. *International Journal for Numerical Methods in Fluids*, 86, 07 2017.

[23] Todd L. Parsons and Tim Rogers. Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology, 2015.

[24] Xue-Mei Li. Stochastic differential equations on noncompact manifolds: moment stability and its topological consequences. 2019.

[25] Sebastian Ullmann, Christopher Müller, and Jens Lang. Stochastic galerkin reduced basis methods for parametrized linear convection–diffusion–reaction equations. *Fluids*, 6(8), 2021.

[26] Xingye Kan, Jinqiao Duan, Ioannis G. Kevrekidis, and Anthony J. Roberts. Simulating stochastic inertial manifolds by a backward-forward approach. *SIAM Journal on Applied Dynamical Systems*, 12(1):487–514, 2013.

[27] J. B. Marston and S. M. Tobias. Recent developments in theories of inhomogeneous and anisotropic turbulence, 2022.

[28] Kaushik Srinivasan and William Young. Zonostrophic instability. *Journal of Atmospheric Sciences*, 69:1633–1656, 05 2012.

[29] F. Bouchet, J. B. Marston, and T. Tangarife. Fluctuations and large deviations of reynolds stresses in zonal jet dynamics. *Physics of Fluids*, 30(1):015110, 2018.

[30] Navid C. Constantinou, Brian F. Farrell, and Petros J. Ioannou. Emergence and equilibration of jets in beta-plane turbulence: Applications of stochastic structural stability theory. *Journal of the Atmospheric Sciences*, 71(5):1818 – 1842, 2014.

[31] Muhammad Ardi. Using Variational Autoencoder (VAE) to Generate New Images.

[32] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013.

[33] Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. Timevae: A variational auto-encoder for multivariate time series generation, 2021.

[34] 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. *Phys. Rev. Lett.*, 120:024102, Jan 2018.

[35] 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, January 2021.

[36] Dmitrii Kochkov, Jamie A. Smith, Ayya Alieva, Qing Wang, Michael P. Brenner, and Stephan Hoyer. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21):e2101784118, 2021.

[37] Dario Lucente, Joran Rolland, Corentin Herbert, and Freddy Bouchet. Coupling rare event algorithms with data-based learned committor functions using the analogue markov chain, 2021.

[38] Jacob Page, Michael P. Brenner, and Rich R. Kerswell. Revealing the state space of turbulence using machine learning. *Phys. Rev. Fluids*, 6:034402, Mar 2021.

[39] Olivier Martin. Lyapunov exponents of stochastic dynamical systems. *Journal of Statistical Physics*, 41(1-2):249–261, October 1985.