A generative machine learning surrogate model of plasma turbulence

B Clavier, D Zarzoso, E Frénod, Diego Del-Castillo-Negrete

A generative machine learning surrogate model of plasma turbulence.

B. Clavier* and D. Zarzoso
Aix-Marseille University, CNRS, Centrale Marseille, M2P2 UMR 7340, Marseille

D. del-Castilllo-Negrete
Oak Ridge National Laboratory, Oak Ridge, TN 37831-8071, United States of America

E. Frénod
Université Bretagne Sud, LMBA UMR 6205, Vannes

State-of-the-art techniques in generative artificial intelligence are employed for the first time to construct a surrogate model for plasma turbulence that enables long time transport simulations. The proposed GAIT (Generative Artificial Intelligence Turbulence) model is based on the coupling of a convolutional variational auto-encoder, that encodes precomputed turbulence data into a reduce latent space, and a deep neural network and decoder that generate new turbulence states 400 times faster than the direct numerical integration. The model is applied to the Hasegawa-Wakatani (HW) plasma turbulence model, that is closely related to the quasigeostrophic model used in geophysical fluid dynamics. Very good agreement is found between the GAIT and the HW models in the spatio-temporal Fourier and Proper Orthogonal Decomposition spectra as well as in the flow topology characterized by the Okubo-Weiss decomposition. Agreement is also found in the probability distribution function of particle displacements and the effective turbulent diffusivity.

Turbulence is ubiquituous in nature and industrial systems. The atmosphere of planets [1], oceanic currents [2], ionized gases in stars [3], the solar tachocline [4], and the solar wind [5] are some examples of systems where turbulence can be encountered. Other examples include turbulent flows behind airfoils and in combustion chambers where turbulence is known to play an important role in mixing [6]. Independently of its nature or underlying mechanism, turbulence represents without any doubt one of the greatest challenges in numerical modelling due to the vast range of spatio-temporal scales involved and the unpredictable behaviour.

In this Letter we focus on turbulence in magnetized plasmas of interest to controlled nuclear fusion and astrophysics. The pressing need to understand and predict the role of turbulence in particles and energy transport in fusion plasmas has motivated the development of a large number of computational tools ranging from fluid to gyro-fluid and gyro-kinetic models. However, due to the spatiotemporal multiscale properties of plasma turbulence these computations are very time consuming and in some cases impractical or unaffordable. To overcome this limitation we propose the use of state-of-the-art artificial intelligence (AI) methods to accelerate turbulence simulations. Although our focus is on plasma physics, as we will describe below, the turbulence model of interest contains as a special case a model extensively used in geophysical fluid dynamics. Recently, a variety of machine learning techniques have been used to accelerate and validate plasma physics simulations as well as to develop data-driven surrogate models [7–13] Of particular interest is the potential of AI to create new data from existing data using generative deep-learning techniques. These techniques, which have been used in chemistry to automatically design new molecules [14] and fluid mechanics [15], form the basis of our proposed GAIT (Generative Artificial Intelligence Turbulence) surrogate model.

Once computed, a turbulence field can be used to perform transport studies of passive tracers, e.g., impurities. In the simplest setting, the basic idea is to solve the transport equation in a time interval $(0, t_{max})$ for a large ensemble of particles using the electromagnetic fields computed from the numerical solution of a turbulence model (e.g., gyro-kinetics or gyro-fluid). The main limitation encountered in this type of studies is that the t_{max} of relevance for transport studies is typically much larger than the time range for which the turbulence model can be solved, due to limited computational resources. One way to overcome this limitation is by using surrogate models, like the GAIT model proposed here, that once trained in a time range $(0, t_{train})$ can be used to produce, with negligible computational cost, turbulent states in the transport scale range $(0, t_{max})$ where $t_{train} \ll t_{max}$.

As a tractable example to illustrate and test the GAIT model we consider the extensively used Hasegawa-Wakatani (HW) model [16]

$$\partial_t n + [\phi, n] = C\,(\phi - n) - \kappa \partial_y \phi - \mu \nabla^2 n \qquad (1)$$

$$\partial_t \Omega + [\phi, \Omega] = C\,(\phi - n) - \mu \nabla^2 \Omega \qquad (2)$$

where ϕ is the electrostatic potential, n is the density, $\Omega = \nabla_\perp^2 \phi$ is the vorticity, C is the adiabaticity coefficient, μ is a hyper-diffusion coefficient, $[A, B] = \partial_x A \partial_y B - \partial_y A \partial_x B$ is the canonical Poisson bracket, and $\kappa = -\partial_x \log n_0$ is the drive of the instability underlying the turbulence with n_0 the background density. The magnetic field is assumed time-independent, uniform and perpendicular to the (x, y) plane, i.e. $\mathbf{B} = B\mathbf{e}_z$.

* benoit.clavier@univ-amu.fr

Time is normalized to the cyclotron frequency ω_{c0}, and (x, y) to the Larmor radius ρ_s. In applications to controlled nuclear fusion in toroidal magnetic confinement devices, this model is used to describe edge turbulence in a poloidal cross-section perpendicular to the toroidal magnetic field, and has been very recently employed in the context of developing data-driven reduced models [17].

In the special case n =constant, Eqs. (1)-(2) reduce to the quasigeostrophic β-plane equation, also known as Charney equation, modeling Rossby waves in the atmosphere and the oceans [18]. In this reduction, the magnetic field corresponds to the earth's rotation, the electrostatic potential to the fluid streamfunction, and the plasma density gradient to the variation of the Coriolis force with latitude. This close analogy opens the possibility of applying methods and ideas in magnetized plasmas to geophysical flows and viceversa, see for example Ref. [19] and references therein. In particular, since our methodology focuses on the construction of a surrogate model for the potential, and not the density, it is directly applicable to geophysical fluid dynamics turbulence.

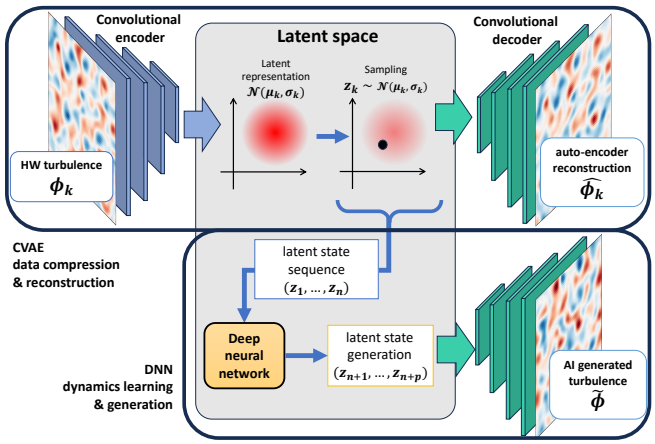

FIG. 1: Schematic representation of the architecture of the proposed generative machine learning surrogate model for plasma turbulence. [For a dynamic representation see video in the supplemental material]

To train the GAIT model, Eqs.(1)-(2) with $C = 1$, $\kappa = 1$, $B = 1$, $\mu = 10^{-3}$ and $k_0 = 0.15$ were solved numerically using a pseudo-spectral method with a $(N_x, N_y) = 512 \times 512$ spatial grid and a 4th order Runge-Kutta scheme for the time advance with $\Delta t = 2 \cdot 10^{-2} \omega_{c0}^{-1}$ to guarantee convergence and stability. A total of $N_s = 8000$ snapshots of the electrostatic potential were obtained by saving the numerical data at a rate $\Delta t_{saving} = \omega_{c0}^{-1}$, i.e. every 50 numerical time steps. To avoid transient effects, snapshots were saved after turbulence has saturated, which in this case corresponded to $\omega_{c0} t \simeq 200$.

As illustrated in Fig. 1 the proposed GAIT model consists of a convolutional variational auto-encoder (CVAE) coupled to a deep neural network (DNN). CVAEs are

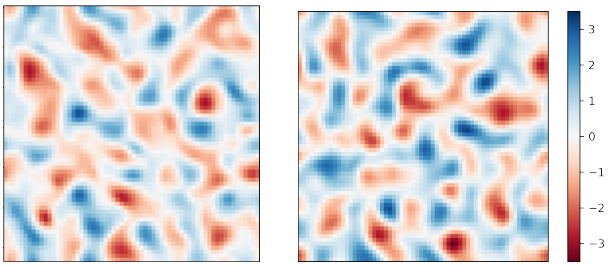

FIG. 2: Snapshots of electrostatic potential obtained from the direct numerical integration of Hasegawa-Wakatani model (left) and the proposed GAIT (Generative Artificial Intelligence Turbulence) model (right).

probabilistic generative machine learning models composed of an encoder, $\mathcal{E}$, and a decoder, $\mathcal{D}$ [20, 21]. Similar architecture has been employed to compress diagnostic measurements characterizing a plasma state [8]. In our problem $\mathcal{E}$ is a convolutional neural network (CNN) consisting of five successive 2D convolution layers (with variable stride, kernel sizes, and filters) that map snapshots of the electrostatic potential, ϕ, represented in an $N_x \times N_y$ grid of the (x, y) plane, to points in a reduced N-dimensional latent space of parameters of a variational Gaussian distribution $\mathcal{N}(\mathbf{z}|\mu, \sigma)$, with $\mathbf{z} \in \mathbb{R}^N$, $\mu \in \mathbb{R}^N$ and $\sigma \in \mathbb{R}^N$. For training purposes, the 512×512 turbulence data was down-sampled to a $N_x \times N_y = 64 \times 64$ resolution. For the dimension of the latent space we used $N = 64$, resulting in a encoder data compression factor of $N_x N_y / N = 64$. The decoder, $\mathcal{D}$, is designed as a mirror image of the encoder with transposed convolution layers that progressively unfold N-dimensional latent vectors into $N_x \times N_y$ dimensional images, such that $\mathcal{D}[\mathcal{E}[\phi]] = \hat{\phi} \approx \phi$. The goal of the training is to find an optimal $\mathcal{E}$ that compresses the data preserving the maximum information and an optimal $\mathcal{D}$ that has the minimum reconstruction error. This is achieved by minimizing the loss function

$$L_{\text{CVAE}} = \sum_i^{N_s} w_\phi \left\| \phi_i - \hat{\phi}_i \right\|^2 + w_{\nabla\phi} \left\| \nabla\phi_i - \nabla\hat{\phi}_i \right\|^2 + w_{\mathbb{KL}} \mathbb{KL} \left(\mathcal{N}(\mathbf{z}_i|\mu_i, \sigma_i), \mathcal{N}(\mathbf{z}_i|\mathbf{0}, \mathbf{1}) \right) \quad (3)$$

where $\{\phi_i\}_{=1}^{N_s}$ and $\{\hat{\phi}_i\}_{=1}^{N_s}$ are the N_s training and the N_s reconstructed snapshots respectively, $\mathbf{z}_i = \mathcal{E}(\phi_i)$, and $\mathbb{KL}$ is the Kulback-Leibler divergence [22]. The first two terms of Eq. (3) are the reconstruction loss of ϕ and $\nabla\phi$ in the L2 norm, weighted by the hyper-parameters w_ϕ and $w_{\nabla\phi}$. Although in principle one can set $w_{\nabla\phi} = 0$, we have found that using $w_{\nabla\phi} \neq 0$ leads to faster learning and a reduction of the error in the evaluation of the electric field, $\mathbf{E} = -\nabla\phi$, which for particle transport studies is a more relevant field than ϕ itself. The third term in Eq.(3) uses the Kulback-Leibler divergence to quantify the difference between the normal distributions of the

mapped snapshots in the latent space with mean μ_i and variance σ_i and a normal distribution with zero mean and unit variance. Minimizing this difference forces the encoder to map the different images into a compact set with small variability in the latent space, a property that is critical for the success of the generation of new turbulence data based on the training data. The training of the CVAE used 8000 snapshots of the potential and was implemented using an Adam optimizer with a learning rate $l_r = 10^{-3}$ and 5000 epochs. Since the numerical simulations are performed in a double periodic domain, the standard convolution algorithms (e.g., those in TensorFlow) needed to be modified so the encoder automatically captures the periodicity of the input data and the decoder generates periodic images from points in the latent space.

Once the CVAE is trained, the sequence of $N_s = 8000$ points in the latent space, $\{\mathbf{z}_1, \mathbf{z}_2, \ldots \mathbf{z}_{N_s}\}$, corresponding to the evolution in time of the turbulence snapshots $\{\phi_1(x,y), \phi_2(x,y), \ldots \phi_{N_s}(x,y)\}$, are used to train a multi-layer DNN (Deep Neural Network) to predict $\mathbf{z}_{N_{s+i}}$, given $\mathbf{z}_{N_{s+i-1}}$. For this, we shuffle and split the pairs into training and testing datasets, with a test-size of 0.30, and train the DNN for 500 epochs using Adam optimizer with a learning rate $l_r = 10^{-3}$. The DNN architecture consisted of three hidden layers and was implemented using Tensorflow and the API Keras using the loss function $L_{\rm DNN} = \sum_i \|\mathbf{z}_{i+1} - DNN(\mathbf{z}_i)\|^2$. Numerical results show that the electrostatic energy, $|\nabla\phi_i|^2$ of a turbulence snapshot increases with the euclidean norm of the corresponding latent space vector $\mathbf{z_i} = \mathcal{E}\phi_i$. Therefore, to avoid strong variations in the electrostatic energy, the DNN uses bounded *tanh* activation functions that preclude the generation of outliers during the generation process. The turbulence AI generation process is completed by using the decoder to map the new points in the latent space into new snapshots of turbulence, $\{\tilde{\phi}_{N_s+1}(x,y), \ldots \tilde{\phi}_{N_s+N_g}(x,y)\}$ with $\tilde{\phi}_j(x,y) = \mathcal{D}[\tilde{\mathbf{z}}_j]$. As shown in Fig.2, the images of the AI generated turbulent states are practically indistinguishable from those generated from the direct numerical simulation of the Hasegawa-Wakatani turbulence model in Eq. (1)-(2). However, from the physics perspective it is critical to go beyond the observed qualitative agreement of the images and perform systematic tests using quantitative metrics. The rest of the letter is devoted to this goal. Unless otherwise stated, the weights used in the loss function were set to: $w_\phi = 1$, $w_{\nabla\phi} = 10$, and $w_{\mathbb{KL}} = 0.01$. These values were selected to reach an optimal balance between the compression properties of the encoder in the latent space and the reconstruction error of the decoder.

The first test compares the Fourier power spectra in space of the electric field turbulent fluctuations, $\overline{|\widehat{\delta E}(k)|}$, where $\widehat{\delta E}(k,t)$ is the Fourier transform in space of $\delta E(x,y,t) = E(x,y,t) - \langle E \rangle$, $\langle f \rangle$ denotes spatial average, $\overline{f}$ denotes time average, E is the magnitude of the electric field, $E = |-\nabla\phi|$, and $k = \sqrt{k_x^2 + k_y^2}$. The second test compares the Fourier power spectra in time of the electric field turbulent fluctuations, $\left\langle |\widetilde{\delta E}(\omega)| \right\rangle$, where $\widetilde{\delta E}(x,\omega)$ is the Fourier transform in time of $\delta E(x,y,t)$. Figure 3 shows the quantitative between the HW and the GAIT model in the Fourier power spectra in space and time of the electric field turbulent fluctuations.

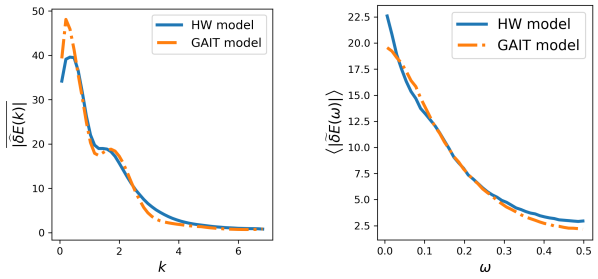

FIG. 3: Comparison of electric field turbulent fluctuations spectra in the HW and GAIT models. Left: Fourier power spectra in space, $\overline{|\widehat{\delta E}(k)|}$. Right: Fourier power spectra in time, $\left\langle |\widetilde{\delta E}(\omega)| \right\rangle$.

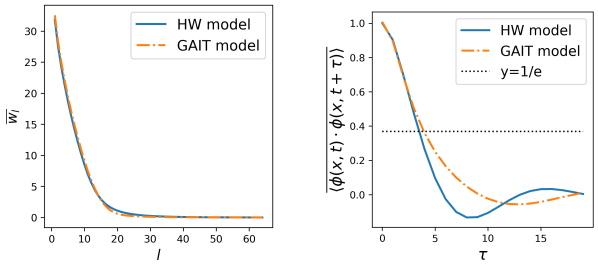

FIG. 4: Comparison of electrostatic potential fluctuations in the HW and GAIT models.(Left) Proper orthogonal decomposition spectrum. (Right) Eulerian time auto-correlation function.

A distinctive feature of 2D turbulence is the spontaneous formation and persistence of coherent structures, e.g. vortices and zonal flows [23]. Understanding these structures is important because the trapping effect of vortices and the long displacements caused by zonal flows can have a critical role in transport, see for example [24] and references therein. Accordingly, an important metric in the evaluation of the proposed AI model is how well the model captures the coherent structures of the flow. To quantify this, we use Proper Orthogonal Decomposition (POD) methods which have been extensively used in fluids [25] and plasmas [26]. Contrary to Fourier spectral analysis, POD is based on data-driven, empirical modes providing an optimal representation of the turbulence state in the energy norm. As discussed in [27] the POD mode decomposition and spectra can be used to characterize the spatiotemporal coherence of HW plasma turbulence. Following this idea, the left panel of

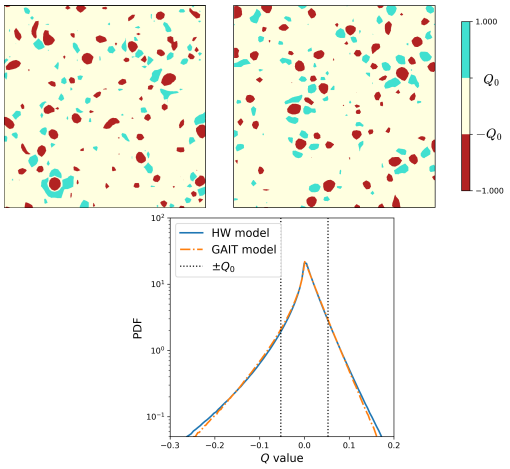

FIG. 5: Comparison of flow topology in the HW and GAIT models. Top panel: spatial distribution of Okubo-Weiss field at a fixed time, $Q(x, y, t_c)$, in the HW model (left) and the GAIT model (right). Bottom panel: comparison of probability distribution of Q values in space and time.

Fig. 4 compares the time average (over the 8000 snapshots) of the POD singular values $\overline{w}_l$ as function of the rank, l, where $\phi(x, y, t) = \sum_{l=1}^{N} w_l(t)u^{(l)}(x, t)v^{(l)}(y, t)$ is the singular value decomposition of the electrostatic potential at time t. The close agreement, indicates that the GAIT model reproduces well the coherent structures of the HW model. Another metric of interest is the Eulerian correlation time of the electrostatic potential fluctuations, $\mathcal{C}(\tau) = \langle\overline{\phi(x, t)\phi(x, t + \tau)}\rangle$. The right panel in Fig. 4 shows overall agreement in this metric although the correlation in the GAIT model exhibits a slightly slower decay.

To further compare the GAIT and the HW turbulence models, we consider the Okubo-Weiss (OW) parameter, used in [28] to characterize the flow topology in HW plasma turbulence. The OW parameter is defined as $Q = s^2 - \omega^2$, where $\omega = \partial_x V_y - \partial_y V_x$ is the vorticity, $s^2 = s_1^2 + s_2^2$ is the deformation with $s_1 = \partial_x V_x - \partial_y V_y$ and $s_2 = \partial_x V_y + \partial_y V_x$ [29]. In the problem of interest here V is the magnitude of the drift velocity, $\mathbf{V} = -\nabla\phi \times B\mathbf{e}_z = (V_x, V_y) = B(-\partial_y\phi, \partial_x\phi)$. Given the threshold, $Q_0 = \sqrt{\langle\overline{Q^2}\rangle}$, the turbulent flow can then be partitioned into three topologically distinctive regions: strongly elliptic (vorticity dominated) $Q \leq -Q_0$, strongly hyperbolic (deformation dominated) $Q \geq Q_0$, and intermediate $-Q_0 < Q < Q_0$. From a transport perspective, strongly elliptic regions tend to reduce transport due to particle trapping and strongly hyperbolic regions tend to enhance mixing due to particle dispersion. The agreement between the GAIT and the HW models in the flow topology of the turbulent field is shown in Fig.5. The top panels compare the Okubo-Weiss three levels ($Q < -Q_0$, $-Q_0 < Q < Q_0$, $Q > Q_0$) decomposition of the turbulence field at a given time. The bottom panel compares the probability distribution of Q values in space and time $\{Q_{i,j}(t_k)\}$, where $i = 1, \ldots 64$, $j = 1, \ldots 64$ and $k = 1, \ldots N_s$.

As a final test we consider the transport problem consisting of solving $\frac{d\mathbf{x}}{dt} = -\nabla\phi \times B\mathbf{e}_z$ for an ensemble of 10^4 tracers initially distributed uniformly in the square domain $\{(x, y)|0 < x < 2\pi/k_0\,, 0 < y < 2\pi/k_0\}$. A Runge-Kutta method, with a timestep $\Delta t = 0.2$, was used to integrate the orbits, along with a linear interpolation in time, and a bilinear interpolation in space for the potential snapshots. When computing the Lagrangian statistics the particles' orbits are considered in the extended $\mathbb{R}^2$. That is the values of $(x_i(t), y_i(t))$ were not wrapped in the $[0, 2\pi/k_0] \times [0, 2\pi/k_0]$ double periodic domain. The main quantities of interest are the the probability distribution function of displacements $P(\Delta r)$ and its variance, $\sigma^2_{\Delta r}(t) = \left\langle (\Delta r - \langle\Delta r\rangle)^2 \right\rangle$, where $\Delta r^2 = (x\,(t) - x\,(0))^2 + (y\,(t) - y\,(0))^2$ and $\langle\cdot\rangle$ denotes ensemble average.

Figure 6 shows that both the HW and the GAIT models exhibit the expected diffusive scaling $\sigma^2_{\Delta r}(t) \sim t$ and Gaussian profiles of $P(\Delta r)$. However, for the parameters used in the results reported above, $(w_\phi, w_{\nabla\phi}, w_{\mathbb{KL}}) = (1, 10, 0.01)$, labeled as "case 1" in the figure, the value of the diffusivity $D = \sigma^2_{\Delta r}/2t$ in the GAIT model, $D_{GAIT,1} = 1.28$, tends to be bigger than the diffusivity obtained using the HW model, $D_{HW} = 0.933$. This discrepancy can be resolved by using $(w_\phi, w_{\nabla\phi}, w_{\mathbb{KL}}) = (0, 10, 0.1)$. This case, labeled as "case 2" in the figure, exhibits excellent agreement with $D_{GAIT,2} = 0.981$. Most importantly, the GAIT model preserves the same transport properties when extrapolated way beyond the range of training and solution of the HW model.

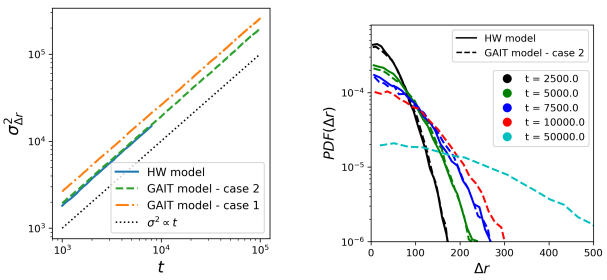

FIG. 6: Comparison of radial displacements, Δr, of tracers in HW and GAIT models. Left: variance $\sigma^2_{\Delta r}(t)$ as function of t in log-log scale. Right: Probability distribution function of displacements at different times.

In the calculations reported in this letter, the numerical integration of the HW model to produce 10000 snapshots of training data took 50 hours on a V100 GPU node of the French Jean-Zay super computer. On the same node, the training of the CVAE took 1 hour, and the training of the DNN less than two minutes. The generation of 10000 snapshots of turbulent states with the

GAIT model took about 7.5 min. That is, once trained, the GAIT model can generate turbulent states at a rate 50 hrs/ 7.5 min $\sim$ 400 times faster than the direct numerical integration.

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. This work has received financial support from the AIM4EP project (ANR-21-CE30-0018), funded by the French National Research Agency (ANR), and from the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy under Contract No. DE-AC05-00OR22725. All the simulations and training of neural networks reported here were performed on HPC resources of IDRIS under the allocations 2021-A0100512455, 2022-AD010512455R1 and 2023-A0140514165 made by GENCI.

[1] J. C. Wyngaard, Atmospheric turbulence, Annual Review of Fluid Mechanics **24**, 205 (1992).
[2] A. E. Gargett, Ocean turbulence, Annual Review of Fluid Mechanics **21**, 419 (1989).
[3] V. Canuto and J. Christensen-Dalsgaard, Turbulence in astrophysics: Stars, Annual review of fluid mechanics **30**, 167 (1998).
[4] E. Spiegel and J.-P. Zahn, The solar tachocline, Astronomy and Astrophysics (ISSN 0004-6361), vol. 265, no. 1, p. 106-114. **265**, 106 (1992).
[5] R. Bruno and V. Carbone, The solar wind as a turbulence laboratory, Living Reviews in Solar Physics **10**, 1 (2013).
[6] N. Peters, Turbulent combustion, Measurement Science and Technology **12**, 2022 (2001).
[7] 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).
[8] B. Zhu, M. Zhao, H. Bhatia, X.-q. Xu, P.-T. Bremer, W. Meyer, N. Li, and T. Rognlien, Data-driven model for divertor plasma detachment prediction, Journal of Plasma Physics **88**, 895880504 (2022).
[9] P. Rodriguez-Fernandez, A. White, A. Creely, M. Greenwald, N. Howard, F. Sciortino, and J. Wright, Vitals: a surrogate-based optimization framework for the accelerated validation of plasma transport codes, Fusion Science and Technology **74**, 65 (2018).
[10] R. M. Churchill, M. D. Boyer, and S. C. Cowley, Accelerating fusion energy with ai, in *Artificial Intelligence for Science: A Deep Learning Revolution* (World Scientific, 2023) pp. 271–284.
[11] R. Greif, F. Jenko, and N. Thuerey, Physics-preserving ai-accelerated simulations of plasma turbulence, arXiv preprint arXiv:2309.16400 (2023).
[12] M. Yang, P. Wang, D. del Castillo-Negrete, Y. Cao, and G. Zhang, A pseudo-reversible normalizing flow for stochastic dynamical systems with various initial distributions, arXiv preprint arXiv:2306.05580 (2023).
[13] C. J. McDevitt, A physics-informed deep learning model of the hot tail runaway electron seed, Physics of Plasmas **30** (2023).
[14] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules, ACS central science **4**, 268 (2018).
[15] Y. Wang, A. Solera-Rico, C. S. Vila, and R. Vinuesa, Towards optimal β-variational autoencoders combined with transformers for reduced-order modelling of turbulent flows, International Journal of Heat and Fluid Flow **105**, 109254 (2024).
[16] A. Hasegawa and M. Wakatani, Plasma edge turbulence, Physical Review Letters **50**, 682 (1983).
[17] C. Gahr, I.-G. Farcas, and F. Jenko, Learning physics-based reduced models from data for the hasegawa-wakatani equations, arXiv preprint arXiv:2401.05972 (2024).
[18] J. Pedlosky, *Geophysical fluid dynamics*, 2nd ed. (Springer New York, NY, 1987).
[19] D. del Castillo-Negrete, Chaotic transport in zonal flows in analogous geophysical and plasma systems, Physics of Plasmas **7**, 1702 (2000), https://pubs.aip.org/aip/pop/article-pdf/7/5/1702/19273666/1702_1_online.pdf.
[20] D. P. Kingma and M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).
[21] M. Cheng, F. Fang, C. Pain, and I. Navon, An advanced hybrid deep adversarial autoencoder for parameterized nonlinear fluid flow modelling, Computer Methods in Applied Mechanics and Engineering **372**, 113375 (2020).
[22] S. Kullback and R. A. Leibler, On information and sufficiency, The annals of mathematical statistics **22**, 79 (1951).
[23] J. Sommeria, Experimental study of the two-dimensional inverse energy cascade in a square box, Journal of fluid mechanics **170**, 139 (1986).
[24] D. del Castillo-Negrete, Chaotic transport in zonal flows in analogous geophysical and plasma systems, Physics of Plasmas **7**, 1702 (2000).
[25] G. Berkooz, P. Holmes, and J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annual review of fluid mechanics **25**, 539 (1993).
[26] P. Beyer, S. Benkadda, and X. Garbet, Proper orthogonal decomposition and galerkin projection for a three-dimensional plasma dynamical system, Physical Review E **61**, 813 (2000).
[27] S. Futatani, S. Benkadda, and D. del Castillo-Negrete, Spatiotemporal multiscaling analysis of impurity transport in plasma turbulence using proper orthogonal decomposition, Physics of Plasmas **16**, 042506 (2009), https://pubs.aip.org/aip/pop/article-

pdf/doi/10.1063/1.3095865/15930324/042506_1_online.pdf.
[28] B. Kadoch, D. del Castillo-Negrete, W. J. T. Bos, and K. Schneider, Lagrangian conditional statistics and flow topology in edge plasma turbulence, Physics of Plasmas **29**, 102301 (2022), https://pubs.aip.org/aip/pop/article-pdf/doi/10.1063/5.0098501/16570139/102301_1_online.pdf.
[29] D. Elhmaïdi, A. Provenzale, and A. Babiano, Elementary topology of two-dimensional turbulence from a lagrangian viewpoint and single-particle dispersion, Journal of Fluid Mechanics **257**, 533–558 (1993).