

Tailor-designed models for the turbulent velocity gradient through normalizing flow

M. Carbone^{1,2,*}, V. J. Peterhans^{3,2}, A. S. Ecker^{4,2}, and M. Wilczek^{1,2†}

¹*Theoretical Physics I, University of Bayreuth, Germany*

²*Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany*

³*Faculty of Physics, University of Göttingen, Germany*

⁴*Institute of Computer Science and Campus Institute Data Science, University of Göttingen, Germany*

(Dated: March 1, 2024)

Small-scale turbulence can be comprehensively described in terms of velocity gradients, which makes them an appealing starting point for low-dimensional modeling. Typical models consist of stochastic equations based on closures for non-local pressure and viscous contributions. The fidelity of the resulting models depends on the accuracy of the underlying modeling assumptions. Here, we discuss an alternative data-driven approach leveraging machine learning to derive a velocity gradient model which captures its statistics by construction. **We use a normalizing flow to learn the velocity gradient probability density function (PDF) from direct numerical simulation (DNS) of incompressible turbulence.** Then, by using the equation for the single-time PDF of the velocity gradient, we construct a deterministic, yet chaotic, dynamical system featuring the learned steady-state PDF by design. Finally, utilizing gauge terms for the velocity gradient single-time statistics, we optimize the time correlations as obtained from our model against the DNS data. As a result, the model time realizations statistically closely resemble the time series from DNS.

Velocity gradients encode comprehensive information on the smallest scales of turbulent flows [1, 2], their statistics unveiling hallmarks of turbulence including cascades [2–5], time irreversibility [6, 7] and intermittency [8–10]. On a more applied level, velocity gradients also govern the small-scale turbulent advection, stretching, and folding of material elements [11–13], and the collisional statistics of inertial particles transported by the flow [14–16], relevant e.g. for droplet growth in turbulent warm clouds [17, 18]. Reduced-order modeling of the velocity gradient is thus appealing for a better theoretical understanding of turbulence and for practical purposes.

A widely-employed and successful approach consists of modeling the Lagrangian dynamics of the velocity gradient $\mathbf{A} \equiv \nabla \mathbf{u}$ (where $\mathbf{u}(\mathbf{x}, t)$ is a three-dimensional turbulent velocity field), following fluid particle trajectories [e.g. 19–26]. However, this single-time/single-point modeling strategy has to deal with the closure problem: **Due to the intrinsic non-locality of turbulence, the dynamical evolution of the velocity gradient at a point is not uniquely determined solely by the velocity gradient configuration at that point.** This issue is evident from the equations governing the velocity gradient dynamics

$$d_t \mathbf{A} = -\mathbf{A}^2 - \mathbf{H} + \nu \nabla^2 \mathbf{A} + \nabla \mathbf{F}, \quad \text{Tr}(\mathbf{A}) = 0, \quad (1)$$

where d_t indicates the Lagrangian derivative taken following the fluid particle trajectory, $\nabla \mathbf{F}$ represents an external forcing, Tr indicates the matrix trace and standard matrix product is implied. The pressure Hessian \mathbf{H} and the viscous Laplacian $\nu \nabla^2 \mathbf{A}$ are nonlocal functionals of the velocity gradient field [e.g. 27, 28]. These unclosed terms cannot be computed by knowing the local velocity gradient only, and they require modeling when addressed from a single-time/single-point perspective.

Due to the intrinsic randomness of equations (1), reduced-order modeling of the velocity gradient aims to

predict the statistics of $\mathbf{A}(t)$, rather than the individual time realizations. This is achieved by integrating a model Langevin equation [19, 20, 22–24, 26]

$$d_t \mathbf{A} = \mathbf{N}(\mathbf{A}) + \mathbf{\Gamma}, \quad \text{Tr}(\mathbf{A}) = 0, \quad (2)$$

for an ensemble of initial conditions, where \mathbf{N} contains a closure model for the pressure Hessian and viscous term, and $\mathbf{\Gamma}$ is a Gaussian noise, **such that the steady-state statistics of the 3×3 matrix \mathbf{A} approximate the statistics of the turbulent velocity gradient ensemble \mathbf{A} .** The modeling step from Eq. (1) to (2) implies that the right-hand side of (1) depends on the whole velocity field $\mathbf{u}(\mathbf{x}, t)$, while the right-hand side of (2) depends only on the eight-dimensional state variable \mathbf{A} . This drastic reduction of degrees of freedom generates the aforementioned closure problem.

Many phenomenological models have been successful in qualitatively predicting the statistical dynamics of the velocity gradient in isotropic turbulence [e.g. 19–26, 29, 30]. More recently, data-driven approaches based on machine learning [31, 32] or lookup tables [33] have reached accuracy levels comparable to the most advanced phenomenological closures. These recent approaches for data-driven models [31–33] approximate/learn the unclosed terms arising from (1). Statistically, this amounts to estimating the average time derivative of the velocity gradient conditional on the configuration of the velocity gradient itself, i.e. $\langle d_t \mathbf{A}(t) | \mathbf{A} \rangle$, and then integrating it in time to generate velocity gradient time realizations. A major issue in this procedure is that an accurate a-priori estimation of the conditional time derivatives does not guarantee that the a-posteriori generated velocity gradient time series are realistic. This a-priori vs. a-posteriori issue, also known from Large Eddy Simulations [34], leads not only to inaccuracies in the predictions but may even imply the blow-up of the gradient realizations

over time. Diverging trajectories are either neglected [32] or damped via ad-hoc terms [26] to reach a non-trivial statistically steady state. Another open problem of the current phenomenological and data-driven models is that they can faithfully approximate single-time statistics, performing well on certain marginal PDFs of the relatively high-dimensional velocity gradient PDF, while leaving room for improvement on other marginal distributions and two-time statistics (e.g. [22–24, 32]). On a more general level, recent alternatives that have the potential of improving the classical reduced-order modeling include Lagrangian Large Eddy Simulations [35] and time series generation by diffusive models [36, 37] or Generative Adversarial Networks [38, 39].

In this work, we propose an approach to construct velocity gradient models that yield the correct velocity gradient statistics of homogeneous isotropic turbulence by design. Earlier attempts to formulate such constrained models consisted of imposing certain statistics such as the log-normality of the velocity gradient magnitude [19, 29]. Instead, here we constrain the model to yield the full distribution of the velocity gradient: **We train a neural network to predict the PDF of the velocity gradient, $f(\mathbf{A})$, from a DNS dataset employing a normalizing flow [40, 41]. Then we construct a dynamical system in the form of Eq. (2) that has the learned PDF as the steady-state probability density distribution.** In the following, we focus on a deterministic reduced-order model capable of maintaining a chaotic statistically steady state (i.e. Eq. (2) without noise). This eliminates the need for any ad-hoc stochastic forcing, which is required in existing models to keep non-trivial steady-state statistics [e.g. 22–24, 26, 31–33]. Our deterministic model is associated with a Liouville equation for the conservation of probability in the phase space

$$\partial_t f + \partial_{ij} (N_{ij} f) = 0, \quad (3)$$

where ∂_{ij} denotes tensor derivative with respect to \mathcal{A}_{ij} , and summation over repeated indices is implied. Since we know the intended steady-state solution $f(\mathbf{A})$, we can derive an expression for the vector field \mathbf{N} such that Eq. (3) yields that solution by construction, namely

$$N_{ij}(\mathbf{A}; \{\psi\}) = \partial_{pq} T_{ijpq} + T_{ijpq} \partial_{pq} \log f, \quad (4)$$

where $\mathbf{T}(\mathbf{A}; \{\psi\})$ is any smooth tensor function of the velocity gradient such that $T_{ijpq} = -T_{pqij}$, and $\{\psi\}$ is a set of model parameters. For stochastic generalizations of our approach, the same can be done by replacing \mathbf{T} with $\mathbf{T} + \mathbf{D}/2$ in Eq. (4), where $\mathbf{D}(\mathbf{A})$ is the correlation of the noise $\boldsymbol{\Gamma}$ in Eq. (2).

A drift of the form of Eq. (4) ensures that f is a steady-state distribution of our model, and this holds true for any tensor \mathbf{T} that is anti-symmetric in its index pairs. Therefore, imposing the single-time steady PDF leaves the gauge term \mathbf{T} undetermined, that is, there exists a

class of models with the same single-time distribution but different multi-time statistics. Within this class, we can select the model that produces realizations of the velocity gradient which are statistically closest to the realizations in the DNS dataset. **To this end, we optimize the conditional time derivatives together with the time correlations of the velocity gradient.** For statistically isotropic turbulence, we can exploit additional gauge terms that affect velocity gradient multi-time correlations while leaving the single-time statistics unchanged for kinematic reasons [26, 42]. Those gauge terms stem from the particular functional form of an isotropic velocity gradient PDF, which depends only on the five invariants formed through the strain rate $\mathbf{S} = (\mathbf{A} + \mathbf{A}^\top)/2$ and the vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{u}$ [43, 44]

$$\begin{aligned} \mathcal{I}_1 &= \text{Tr}(\mathbf{S}^2) & \mathcal{I}_2 &= \|\boldsymbol{\omega}\|^2/2 & \mathcal{I}_3 &= \text{Tr}(\mathbf{S}^3) \\ \mathcal{I}_4 &= \boldsymbol{\omega} \cdot \mathbf{S} \boldsymbol{\omega} & \mathcal{I}_5 &= \boldsymbol{\omega} \cdot \mathbf{S}^2 \boldsymbol{\omega}. \end{aligned} \quad (5)$$

The tensors $\tilde{\mathbf{T}}$ which leave the dynamics of the invariants (5) unchanged satisfy the kinematic condition

$$\tilde{T}_{ijpq} \partial_{pq} \mathcal{I}_k = 0, \quad (6)$$

and will help us optimize the velocity gradient multi-time correlations while preserving its single-time statistics.

Now that we have established the basic theory behind the method, let us discuss its practical implementation. The proposed approach relies on a parameterization of the velocity gradient PDF, a technical step that we accomplish through a normalizing flow [40, 41]. To facilitate the implementation of the normalizing flow, we unroll the components of the 3×3 traceless matrix \mathcal{A}_{ij} into an eight-dimensional vector \mathbf{a} . Then, starting from an ensemble of eight-dimensional Gaussian vectors $\mathbf{a}^{(0)}$, we perform a sequence of real, non-volume-preserving, quasi-linear transformations [45]. In each transformation layer, we update one variable through an affine transformation, whose scaling $\Lambda^{(K)}$ and bias $b^{(K)}$ are computed from the remaining seven variables. Then we repeat the sequence of transformations four times for each variable

$$a_i^{(K)} = \begin{cases} \exp(\Lambda^{(K)}) a_i^{(K-1)} + b^{(K)} & \text{if } i = K \pmod{8} \\ a_i^{(K-1)} & \text{otherwise,} \end{cases} \quad (7)$$

from transformation layer $K = 1$ to $K = 32$. The scaling functions $\Lambda^{(K)}(\mathbf{a}^{(K-1)}; \{\theta^{(K)}\})$ and the bias functions $b^{(K)}(\mathbf{a}^{(K-1)}; \{\theta^{(K)}\})$ in Eq. (7) depend on all the gradient components except the one updated across that layer, so that $\partial a_i^{(K)} / \partial a_j^{(K-1)} = \exp(\Lambda^{(K)})$ for $i = K \pmod{8}$ and $i = j$. The scaling and bias functions consist of two neural networks for each layer, featuring two hidden linear layers with 64 neurons, Softplus activation function and parameters $\{\theta^{(K)}\}$. The transformed eight-dimensional vector $\mathbf{a}^{(32)}$ is mapped back to a 3×3 traceless matrix, $\mathbf{a}^{(32)} \Leftrightarrow \mathbf{A}$, and the ensemble of matrices \mathbf{A}

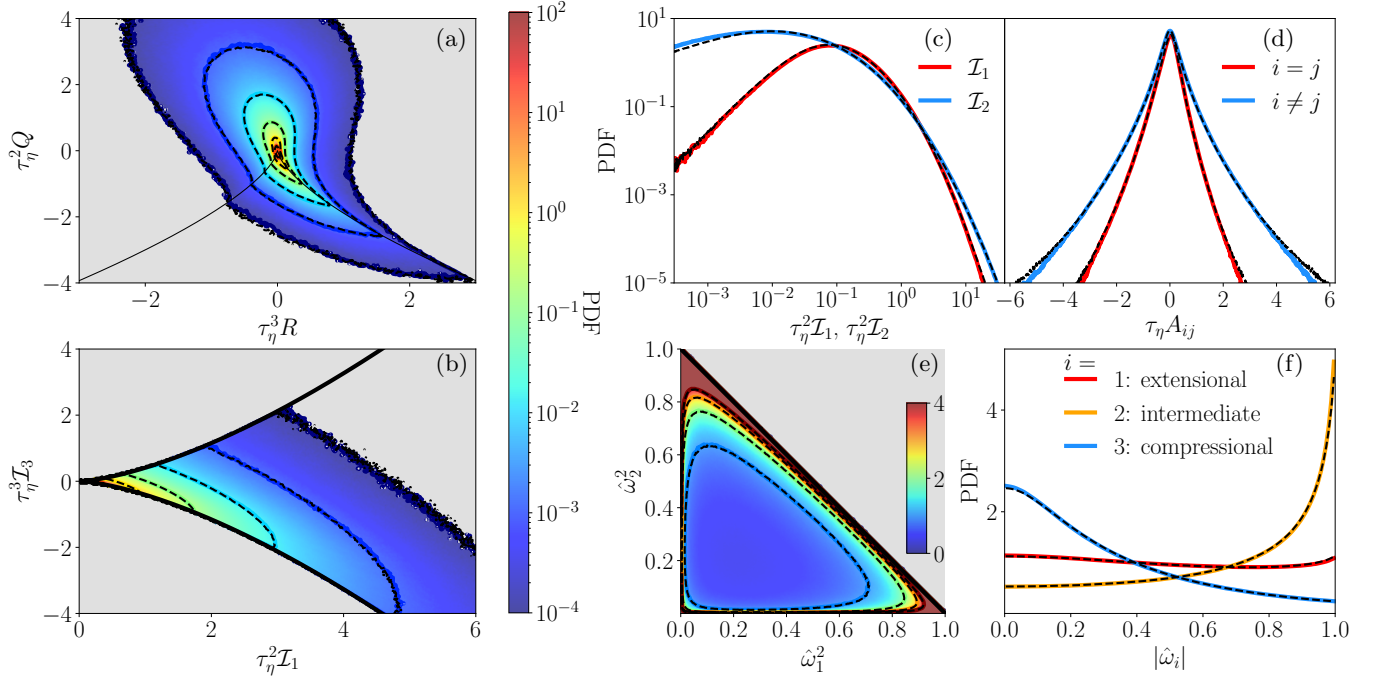


FIG. 1. Single-time velocity gradient distributions from the DNS (colormaps and colored solid lines) and from our model (black dashed lines). Panel (a) shows the joint PDF of principal invariants of the velocity gradient, $Q = -\text{Tr}(\mathbf{A}^2)/2$ and $R = -\text{Tr}(\mathbf{A}^3)/3$, panel (b) shows the joint PDF of the strain invariants $\mathcal{I}_1 = \text{Tr}(\mathbf{S}^2)$ and $\mathcal{I}_3 = \text{Tr}(\mathbf{S}^3)$. Panel (c) shows the PDF of the strain-rate and vorticity magnitudes; the PDFs of the longitudinal and transverse velocity gradient components are shown in panel (d). Panels (e) and (f) display the PDFs of the normalized vorticity components in the strain eigenframe, $\hat{\omega}_i = \mathbf{v}_i \cdot \boldsymbol{\omega} / \|\boldsymbol{\omega}\|$, where \mathbf{v}_i are the strain-rate eigenvectors (index 1 denoting the most extensional direction and 3 the most compressional). All dimensional quantities are nondimensionalized by the Kolmogorov time scale, $\tau_\eta = 1/\sqrt{2\mathcal{I}_1}$.

aims to be statistically similar to the ensemble of turbulent velocity gradients \mathbf{A} , through a maximum likelihood criterion. Owing the quasi-linearity of the transformation (7), the log-likelihood $\log f^{(K)}$ of the ensemble of matrices $\mathbf{A}^{(K)}$ changes across a layer as

$$\log f^{(K)}(\mathbf{A}^{(K)}) = \log f^{(K-1)}(\mathbf{A}^{(K-1)}) + \Lambda^{(K)}. \quad (8)$$

We find the optimal parameters in the transformations (7) and the corresponding parameterization of the velocity gradient PDF by maximizing the log-likelihood of the turbulent velocity gradient ensemble from the DNS

$$\{\bar{\theta}\} = \underset{\{\theta\}}{\text{argmax}} \langle \log f(\mathbf{A}; \{\theta\}) \rangle, \quad (9)$$

where $f \equiv f^{(32)}$ is the target velocity gradient PDF, \mathbf{A} is an element of the turbulent velocity gradient ensemble from the DNS, the angle brackets denote ensemble average computed over the DNS dataset, and $\{\theta\}$ stands for all the parameters featured in the transformations (7).

The DNS dataset consists of the velocity gradient realizations along 2^{18} fluid particle Lagrangian trajectories from a DNS of three-dimensional, incompressible, statistically steady and isotropic turbulence. The velocity gradient time realizations span $64\tau_\eta$, and are sampled with a time step $\tau_\eta/60$ (where τ_η is the Kolmogorov time

scale). The DNS resolves 1024^3 Fourier modes with a maximum resolved wavenumber $k_{\text{max}} = 3/\eta$ (where η is the Kolmogorov scale) and Reynolds number based on the Taylor microscale $\text{Re}_\lambda = 220$. The normalizing flow implementation relies on Pytorch [46]. We employ the Adam optimizer [47] with a learning rate of 5×10^{-5} and running average coefficients $\beta_1 = 0.9$, $\beta_2 = 0.999$. We train the model on 40 GPUs for 5×10^4 training steps, on batches of 1024 velocity gradient time realizations.

The normalizing flow yields an accurate parameterization of the turbulent velocity gradient PDF, as well as an invertible transformation (7) capable of mapping an ensemble of Gaussian random matrices onto an ensemble resembling the turbulent velocity gradients. This is in the same spirit of Gaussian mapping closures in fluid dynamics [48, 49]. We can now sample the learned steady-state PDF by using the transformation (7) or by integrating our model (2) with vector field (4), and the results coincide. Figure 1 shows a comparison between velocity gradient statistics from DNS data and our model. The model captures several hallmark statistics of small-scale turbulence with unprecedented accuracy, including: The PDF of the velocity gradient (Fig. 1a) and strain-rate principal invariants (Fig. 1b), the distribution of the strain-rate and vorticity magnitude (Fig. 1c) and the ve-

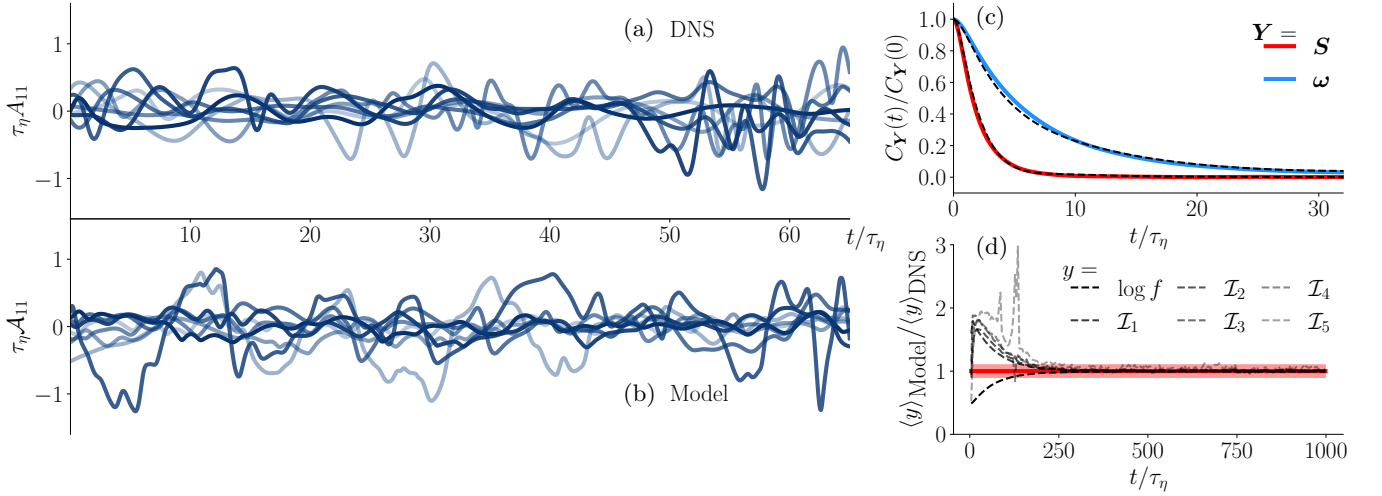


FIG. 2. Comparison of DNS and model trajectories, time correlations and model convergence. Panels (a) and (b) show sample realizations of the longitudinal velocity gradient component along fluid particle trajectories from the DNS and our model. Panel (c) displays the normalized time correlations of the strain rate and vorticity from the DNS (colored solid lines) and our model (black dashed lines). Panel (d) shows the convergence of our model (transparent black dashed lines) from Gaussian initial conditions to turbulent-like statistics, as quantified by the log-likelihood of the velocity gradient ensemble and the average velocity gradient invariants, as functions of time. The red line indicates the DNS reference, with the transparent region denoting a 10% accuracy interval. All dimensional quantities are nondimensionalized by the Kolmogorov time scale, $\tau_\eta = 1/\sqrt{\langle 2\mathcal{I}_1 \rangle}$.

locity gradient Cartesian components (Fig. 1d), as well as the statistical alignments between the vorticity and the strain-rate eigenframe (Fig. 1e-f). The quantitative agreement between the model and the DNS data on several marginal distributions from the velocity gradient PDF shows that we have gained access to the full velocity gradient PDF.

Our data-driven approach decouples the single-time statistics modeling from the optimization of the multi-time statistics. **The anti-symmetric tensor $\mathbf{T}(\mathcal{A}; \{\psi\})$ governs the velocity gradient multi-time statistics**, and we represent it by a neural network featuring three hidden linear layers with 128 neurons, Softplus activation function and parameters $\{\psi\}$. To optimize the parameters of \mathbf{T} , we numerically integrate the model equations (2) for an ensemble of initial conditions sampled from the DNS data, $\mathcal{A}(0) \equiv \mathcal{A}(0)$, and require that the time realizations $\mathcal{A}(t) = \mathcal{A}(0) + \int_0^t ds \mathbf{N}(\mathcal{A}(s); \{\psi\})$ have time correlations and conditional time derivatives similar to the DNS,

$$\{\bar{\psi}\} = \underset{\{\psi\}}{\operatorname{argmin}} \left(\int_0^T dt \|\langle \mathcal{A}_{ij}(0) \mathcal{A}_{pq}(t) - A_{ij}(0) A_{pq}(t) \rangle\|^2 + \beta \sum_{k=1}^5 \left\langle \|\partial_{i_j} \mathcal{I}_k(\mathcal{A}) (d_t A_{ij} - N_{ij}(\mathcal{A}; \{\psi\}))\|^2 \right\rangle \right) \quad (10)$$

where β is a constant weighting parameter. For the minimization (10), we employ the Adam optimizer with the parameters reported above for Eq. (9), a time span $T = 8\tau_\eta$ and weighting factor $\beta = 10^{-5}$. The training converges within 10^3 train steps.

The optimization (10) results in a network \mathbf{T} , thus a tensor function \mathbf{N} (see Eq. (4)), such that the time correlations from the model approximate the time correlations from the DNS data while preserving the previously learned single-time PDF $f(\mathcal{A})$ by construction. Integrating the model (2) with the drift (4) optimized according to the criterion (10) can be understood as a particular Monte Carlo sampling of the PDF f aided by normalizing flow [50], in which the sampling trajectories $\mathcal{A}(t)$ have the meaning of actual time realizations. Although the model was not designed to capture individual velocity gradient time realizations, the time series generated by the model look qualitatively similar to those from DNS, as displayed in Fig. 2(a-b). By visual inspection, the DNS trajectories $A_{11}(t)$ and model trajectories $\mathcal{A}_{11}(t)$ have similar amplitudes with large and negative excursions occurring more frequently than positive ones, and the duration of such excursions is comparable. Also quantitatively, the auto-correlation functions of strain $C_S = \langle S_{ij}(0) S_{ij}(t) \rangle$ and vorticity $C_\omega = \langle \omega_i(0) \omega_i(t) \rangle$ from the model match well those of the DNS data, as Fig. 2(c) shows.

Since our model is deterministic, it is a priori unclear whether it approaches a unique statistically steady state independent of initial conditions. To answer this question, we performed numerical tests in which we observed convergence from an ensemble of Gaussian initial conditions to the prescribed steady-state PDF. We quantify the convergence by comparing the log-likelihood of the model PDF $\langle \log f(\mathcal{A}(t)) \rangle$ to the log-likelihood of the DNS data estimated via normalizing flow. Additionally, we compare the average invariants from the model real-

izations and the DNS data. As Fig. 2(d) shows, all quantities approach the expected steady-state values for times longer than $\mathcal{O}(100\tau_\eta)$. The convergence of the deterministic system to a statistically steady state, independent of the initial conditions, is a typical feature of a chaotic system [51]. Accordingly, we find the largest Lyapunov exponent for our model to be positive, and the dynamics to be mildly dissipative, with an estimated Kaplan-Yorke dimension of 7.96. Those observations, combined with the convergence to stationary statistics, indicate deterministic chaos.

Finally, the convergence to a target distribution showcases that it is possible to construct generative machine-learning models which do not rely on mapping stochastic signals into a distribution of target objects [e.g. 36], but rather use the chaotic nature of the underlying high-dimensional mapping to mimic creativity.

Summarizing, we have developed a reduced-order deterministic velocity gradient model that captures the statistical properties of the full eight-dimensional velocity gradient PDF by design. This is achieved in a two-step procedure. First, the full PDF is learned from DNS data using a normalizing flow [40, 45]. Based on that, we construct a Liouville equation which has the learned PDF as a stationary solution. The Liouville equation is associated with a deterministic reduced-order model, which allows to generate time realizations of the process. We then use gauge freedom for the single-time statistics to optimize the time correlations and conditional time derivatives of the model to emulate actual realizations of Lagrangian velocity gradients in turbulence. The resulting deterministic reduced-order model shows signs of chaotic dynamics. For example, it features a positive Lyapunov exponent and converges to the expected statistically stationary state from Gaussian initial conditions. A more detailed analysis of the model dynamical system's properties is an interesting subject for future work.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101001081). The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project b159cb EnSimTurb. Computational resources from the Max Planck Computing and Data Facility and support by the Max Planck Society are gratefully acknowledged.

* maurizio.carbone@uni-bayreuth.de

† michael.wilczek@uni-bayreuth.de

[1] C. Meneveau, *Annu. Rev. Fluid Mech.* **43**, 219 (2011).

- [2] P. L. Johnson and M. Wilczek, *Annu. Rev. Fluid Mech.* **56**, 463 (2024).
- [3] A. Alexakis and L. Biferale, *Phys. Rep.* **767–769**, 1 (2018).
- [4] M. Carbone and A. D. Bragg, *J. Fluid Mech.* **883**, R2 (2020).
- [5] P. L. Johnson, *Phys. Rev. Lett.* **124**, 104501 (2020).
- [6] J. Jucha, H. Xu, A. Pumir, and E. Bodenschatz, *Phys. Rev. Lett.* **113**, 054501 (2014).
- [7] A. Cheminet, D. Geneste, A. Barlet, Y. Ostovan, T. Chaabo, V. Valori, P. Debue, C. Cuvier, F. David, J.-M. Foucaut, J.-P. Laval, V. Padilla, C. Wiertel-Gasquet, and B. Dubrulle, *Phys. Rev. Lett.* **129**, 124501 (2022).
- [8] Z.-S. She, E. Jackson, and S. A. Orszag, *Nature* **344**, 226 (1990).
- [9] D. Buaria and A. Pumir, *Phys. Rev. Res.* **3**, L042020 (2021).
- [10] A. Fuchs, C. Herbert, J. Rolland, M. Wächter, F. Bouchet, and J. Peinke, *Phys. Rev. Lett.* **129**, 034502 (2022).
- [11] S. S. Girimaji and S. B. Pope, *J. Fluid Mech.* **220**, 427–458 (1990).
- [12] L. Bentkamp, T. D. Drivas, C. C. Lalescu, and M. Wilczek, *Nat. Commun.* **13**, 2088 (2022).
- [13] Y. Qi, C. Meneveau, G. A. Voth, and R. Ni, *Phys. Rev. Lett.* **130**, 154001 (2023).
- [14] E. Balkovsky, G. Falkovich, and A. Fouxon, *Phys. Rev. Lett.* **86**, 2790 (2001).
- [15] J. Bec, L. Biferale, M. Cencini, A. Lanotte, S. Musacchio, and F. Toschi, *Phys. Rev. Lett.* **98**, 084502 (2007).
- [16] P. J. Ireland, A. D. Bragg, and L. R. Collins, *J. Fluid Mech.* **796**, 617 (2016).
- [17] E. Bodenschatz, S. P. Malinowski, R. A. Shaw, and F. Stratmann, *Science* **327**, 970 (2010).
- [18] W. W. Grabowski and L.-P. Wang, *Annu. Rev. Fluid Mech.* **45**, 293 (2013).
- [19] S. S. Girimaji and S. B. Pope, *Phys. Fluids* **2**, 242 (1990).
- [20] M. Chertkov, A. Pumir, and B. I. Shraiman, *Phys. Fluids* **11**, 2394 (1999).
- [21] A. Naso and A. Pumir, *Phys. Rev. E* **72**, 056318 (2005).
- [22] L. Chevillard and C. Meneveau, *Phys. Rev. Lett.* **97**, 174501 (2006).
- [23] M. Wilczek and C. Meneveau, *J. Fluid Mech.* **756**, 191 (2014).
- [24] P. L. Johnson and C. Meneveau, *J. Fluid Mech.* **804**, 387 (2016).
- [25] P. L. Johnson and C. Meneveau, *Phys. Rev. Fluids* **2**, 072601(R) (2017).
- [26] L. A. Leppin and M. Wilczek, *Phys. Rev. Lett.* **125**, 224501 (2020).
- [27] K. Ohkitani and S. Kishiba, *Phys. Fluids* **7**, 411 (1995).
- [28] A. J. Majda and A. L. Bertozzi, *Vorticity and Incompressible Flow* (2001).
- [29] R. M. Pereira, L. Moriconi, and L. Chevillard, *J. Fluid Mech.* **839**, 430 (2018).
- [30] P. L. Johnson and C. Meneveau, *J. Fluid Mech.* **837**, 80–114 (2018).
- [31] Y. Tian, D. Livescu, and M. Chertkov, *Phys. Rev. Fluids* **6**, 094607 (2021).
- [32] D. Buaria and K. R. Sreenivasan, *Proc. Natl. Acad. Sci. U.S.A.* **120** (2023).
- [33] R. Das and S. S. Girimaji, (2023), arXiv:2304.14529.
- [34] J. Jiménez and R. D. Moser, *AIAA J.* **38**, 605 (2000).

- [35] Y. Tian, M. Woodward, M. Stepanov, C. Fryer, C. Hyett, D. Livescu, and M. Chertkov, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2213638120 (2023).
- [36] J. Ho, A. Jain, and P. Abbeel, in *Advances in Neural Information Processing Systems*, Vol. 33 (2020) pp. 6840–6851.
- [37] T. Li, L. Biferale, F. Bonaccorso, M. A. Scarpolini, and M. Buzzicotti, (2023), arXiv:2307.08529.
- [38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems*, Vol. 27 (2014).
- [39] M. Arjovsky, S. Chintala, and L. Bottou, in *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70 (2017) pp. 214–223.
- [40] E. G. Tabak and E. Vanden-Eijnden, *Commun. Math. Sci.* **8**, 217 (2010).
- [41] I. Kobyzev, S. J. Prince, and M. A. Brubaker, *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3964 (2021).
- [42] M. Carbone, M. Iovieno, and A. D. Bragg, *J. Fluid Mech.* **900**, A38 (2020).
- [43] S. Pennisi and M. Trovato, *Int. J. Eng. Sci.* **25**, 1059 (1987).
- [44] T. Lund and E. A. Novikov, *Center for Turbulence Research* (Stanford University and NASA) (1992).
- [45] L. Dinh, J. Sohl-Dickstein, and S. Bengio, in *International Conference on Learning Representations* (2017).
- [46] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, (2019), arXiv:912.01703.
- [47] D. P. Kingma and J. Ba, (2017), arXiv:1412.6980.
- [48] H. Chen, S. Chen, and R. H. Kraichnan, *Phys. Rev. Lett.* **63**, 2657 (1989).
- [49] R. H. Kraichnan, *Phys. Rev. Lett.* **65**, 575 (1990).
- [50] P. Bialas, P. Korcyl, and T. Stebel, (2023), arXiv:2308.13294.
- [51] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering* (Westview Press, 2000).