Physics-informed machine learning of the Lagrangian dynamics of velocity gradient tensor

Yifeng Tian [*] and Daniel Livescu [†]

Computational Physics and Methods Group, Computer, Computational and Statistical Sciences Division (CCS-2), Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Michael Chertkov [‡]

Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721, USA

Reduced models describing the Lagrangian dynamics of the velocity gradient tensor (VGT) in homogeneous isotropic turbulence (HIT) are developed under the physics-informed machine learning (PIML) framework. We consider the VGT at both Kolmogorov scale and coarse-grained scale within the inertial range of HIT. Building reduced models requires resolving the pressure Hessian and subfilter contributions, which is accomplished by constructing them using the integrity bases and invariants of the VGT. The developed models can be expressed using the extended tensor basis neural network (TBNN) introduced by Ling *et al.* [J. Fluid Mech. **807**, 155 (2016)]. Physical constraints, such as Galilean invariance, rotational invariance, and incompressibility condition, are thus embedded in the models explicitly. Our PIML models are trained on the Lagrangian data from a high-Reynolds number direct numerical simulation (DNS). To validate the results, we perform a comprehensive out-of-sample test. We observe that the PIML model provides an improved representation for the magnitude and orientation of the small-scale pressure Hessian contributions. Statistics of the flow, as indicated by the joint PDF of second and third invariants of the VGT, show good agreement with the "ground-truth" DNS data. A number of other important features describing the structure of HIT are reproduced by the model successfully. We have also identified challenges in modeling inertial range dynamics, which indicates that a richer modeling strategy is required. This helps us identify important directions for future research, in particular towards including inertial range geometry into the TBNN.

I. INTRODUCTION

Properties of the velocity gradient tensor (VGT), at both small scale and coarse-grained scale from the inertial interval, determine many important characteristics of turbulence [1]. The bare VGT represents the Kolmogorov (viscous) scale of turbulence and statistics of the energy dissipation, while the VGT coarse grained at a larger scale provides information about respective velocity increments and energy cascade. The aforementioned statistics are highly intermittent and rich in information about dynamics and transfer of energy and interaction among turbulent eddies [2,3]. The analysis of the VGT allows one to shed light on the mechanism of energy cascade, which

---

[*]yifengtian@lanl.gov

[†]livescu@lanl.gov

[‡]chertkov@arizona.edu

can be characterized by the geometrical features and highly nonlinear interactions of vorticity and strain, via the so-called vortex stretching and strain self-amplification [4–6]. Perry and Chong [7] and Chong *et al.* [8] have proposed an approach to classify local flow topology and structure using invariants of the VGT. The Lagrangian dynamics of the VGT have been studied following the evolution of the invariants along the conditional mean trajectories (CMTs) of the flow. The analysis has been performed using data extracted from direct numerical simulation (DNS) corresponding to various fully developed turbulent flows, such as isotropic turbulence, turbulent boundary layer, mixing layers, and more complex flows, such as shock-turbulence interaction (e.g., Refs. [9–18]). To accurately represent Lagrangian dynamics of the VGT, reduced empirical models have been proposed to capture the important dynamics without solving the full Navier-Stokes equations. The empirical model reduction approach is justified by our (so far principal) inability to resolve the nonlocal nature of the turbulence dynamics in a mathematically accurate way, which is true with regard to both Eulerian and Lagrangian attempts of building reduced models rigorously. Meneveau [1] provided a comprehensive review of Lagrangian approaches towards building reduced empirical models of turbulence. The restricted Euler equation (REE) [19,20] is the simplest Lagrangian model, and has formed a basis for further exploration. It was shown in Cantwell [21] that the REE allows exact integration, also revealing the model's significant handicap—an unphysical blow-up (singularity) in finite time, due to misrepresentation of the pressure Hessian contribution to the VGT dynamics. A number of models were proposed to resolve the finite-time singularity problem, such as linear diffusion model [22], stochastic diffusion model [23], Lagrangian tetrad model [24], Lagrangian linear diffusion model [25], multiscale model [26], and recent fluid deformation (RFD) approximation [27,28]. More recently, Wilczek and Meneveau [29] have modeled the pressure Hessian (and viscous contributions which were also omitted in the REE modeling) utilizing a random Gaussian field, and then the RFD model was extended in Johnson and Meneveau [30], leading to the recent deformation of Gaussian fields (RDGF) model. Models based on expansion about tensor bases have been explored in Refs. [31,32], where the scalar coefficients (of the tensor basis expansion) are learned from DNS data and/or derived from physical constraints [33]. Johnson and Meneveau [3] and Pereira *et al.* [34] developed models to describe small-scale intermittency in HIT at high Reynolds numbers. We emphasize that these models were theoretically motivated and justified by DNS and experiment verified physical hypotheses, such as local dependence of the pressure Hessian term on the VGT, which were then subjected to physical constraints, such as incompressibility, symmetries, etc. This modeling status quo, which seemed to be hard to improve, got a remarkable new boost from disparate fields of machine learning (ML), artificial intelligence (AI), and data science. Neural networks (NNs) have gained significant attention due to their powerful capability in expressing complex nonlinear functions and the ease of parameter training through the automatic differentiation (AD) functionality provided by open-source ML packages. Empowered by these NNs, even though they are generally application-agnostic, researchers are now starting to use them to validate and improve scientific hypotheses and advance the modeling of complex physics problems.

To the best of our knowledge, there have been surprisingly few attempts to model Lagrangian dynamics using NNs or other ML methods, despite the rich information Lagrangian statistics can provide. This is in contrast to the surge of NN activity in recent years in the field of Eulerian modeling of turbulence. In particular, major efforts were devoted to the development of closure models for Reynolds averaged Navier-Stokes (RANS) and large eddy simulations (LES) using innovative NN architectures, e.g., Ling *et al.* [35], Maulik *et al.* [36]. A very comprehensive overview of many contributions to this field can be found in the review paper by Duraisamy *et al.* [37]. To mention some of these contributions (which are mostly related to this paper): a tensor basis neural network (TBNN) embedding physical constraints, such as Galilean invariance and rotational invariance, into the closure model was developed in Ling *et al.* [35]; physics-informed machine learning (PIML) models infusing the NN with known physical constraints were developed in, e.g., Refs. [38,39]. Other than developing closure models for RANS and LES, researchers have been experimenting with novel ML approaches to learn turbulence dynamics. In this regard,

and just to name a few, we mention Mohan *et al.* [40], where a convolutional long short term memory (ConvLSTM) NN was developed to learn spatiotemporal turbulence dynamics; studies of super-resolution allowing to reconstruct turbulence fields using under-resolved data [41]; and neural ordinary differential equation (neural ODE) for turbulence forecasting [42].

In this paper we combine the body of work in statistical hydrodynamics with ML tools to develop a physics-informed, interpretable model for Lagrangian dynamics of turbulence at both fully resolved/unfiltered and coarse-grained/filtered levels. Specifically, we build upon the original architecture of TBNN [35] and extend TBNN with embedded physical constraints to model the nonlocal pressure Hessian and subfilter dynamics within the Lagrangian description of the VGT.

The rest of the paper is organized as follows. Section II provides technical background, e.g., discussing governing equations for Lagrangian dynamics of the VGT. PIML strategy for modeling closure terms, including physical interpretation of the underlying parameterization, are described in Sec. III. Section IV is devoted to the description of the "ground-truth" DNS data used to train and test the models. The details of training and testing of the constructed PIML model are provided in Sec. V. Performance of the PIML models for unfiltered and also coarse-grained dynamics is assessed through a series of comprehensive tests in Secs. VI and VII, respectively. Finally, conclusions and path forward are discussed in Sec. VIII.

II. GOVERNING EQUATIONS

A. Unfiltered velocity gradient tensor

The Lagrangian dynamics of the VGT in incompressible turbulence can be derived from the Navier-Stokes (NS) equations:

$$\frac{\partial u_i}{\partial t} + u_k \frac{\partial u_i}{\partial x_k} = -\frac{\partial P}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_k \partial x_k}, \tag{1}$$

where u_i, P, and ν denote velocity vector component, pressure, and kinematic viscosity, respectively. The velocity gradient tensor is defined as

$$A_{ij} = \frac{\partial u_i}{\partial x_j}. \tag{2}$$

The Lagrangian dynamics of the VGT and the incompressibility-enforcing Poisson equation can then be derived by applying spatial derivatives to Eq. (1):

$$\frac{dA_{ij}}{dt} = \frac{\partial A_{ij}}{\partial t} + u_k \frac{\partial A_{ij}}{\partial x_k} = -A_{ik}A_{kj} - \frac{\partial^2 P}{\partial x_i \partial x_j} + \nu \frac{\partial^2 A_{ij}}{\partial x_k \partial x_k}, \tag{3a}$$

$$\frac{\partial^2 P}{\partial x_k \partial x_k} = -A_{mn}A_{nm}. \tag{3b}$$

Using Eq. (3b), we then rearrange Eq. (3a) into the following traceless form:

$$\frac{dA_{ij}}{dt} = E_{ij} + H_{ij} + T_{ij}, \tag{4a}$$

$$E_{ij} = -\left(A_{ik}A_{kj} - \frac{1}{3}A_{mn}A_{nm}\delta_{ij}\right), \tag{4b}$$

$$H_{ij} = -\left(\frac{\partial^2 P}{\partial x_i \partial x_j} - \frac{1}{3}\frac{\partial^2 P}{\partial x_k \partial x_k}\delta_{ij}\right), \tag{4c}$$

$$T_{ij} = \nu \frac{\partial^2 A_{ij}}{\partial x_k \partial x_k}, \tag{4d}$$

where the matrices/tensors $\boldsymbol{E}$, $\boldsymbol{H}$, and $\boldsymbol{T}$ are referred to as restricted Euler (RE), nonlocal pressure Hessian, and viscous contributions to the Lagrangian dynamics, respectively. The RE equation is derived from Eq. (4a) by removing the pressure Hessian term and the viscous term. The resulting RE dynamics is closed. Obviously, the removal of the $\boldsymbol{E}$ and $\boldsymbol{H}$ contributions is not justified and closing VGT dynamics by modeling these terms, i.e., expressing them via the VGT, $\boldsymbol{A}$, or related local characteristics explicitly, is extremely challenging.

B. Coarse-grained velocity gradient tensor

Not only Eq. (4a) is not closed, it is also spatially local, therefore not showing (at least not explicitly) Lagrangian dynamics (and related statistics) of the VGT at different spatial scales. In other words, the local description is missing an important aspect of turbulence consisting in the understanding of how statistics of key characteristics and underlying phenomena, such as energy cascade and dynamics of particle separation, change with scale.

To address this apparent lack of scale information in Eq. (4a), we apply a spatial filter (in this study we use a Gaussian filter for extracting the related quantities) to the NS Eq. (1) and derive the transport equation for the coarse-grained/filtered velocity:

$$\frac{\partial \widetilde{u}_i}{\partial t} + \widetilde{u}_k \frac{\partial \widetilde{u}_i}{\partial x_k} = -\frac{\partial \widetilde{P}}{\partial x_i} + \nu \frac{\partial^2 \widetilde{u}_i}{\partial x_k \partial x_k} - \frac{\partial \tau_{ik}}{\partial x_k}, \tag{5}$$

where the tilde symbols, e.g., $\tilde{u}$, identifies filtered variables. The filter is characterized by its scale and typically selected to be within the inertial range of turbulence, which is bounded from above by the energy-containing scale and from below by the viscous (Kolmogorov) scale. The subfilter contributions, i.e., these associated with scales smaller than the filtering scale, are identifed with a prime symbol e.g., u', so that $u = \tilde{u} + u'$. τ, entering the filtered Eq. (5), is the so-called subfilter stress. The subfilter stress contribution on the right-hand side of Eq. (5) originates from the nonlinearity of the original NS Eq. (1) and makes the resulting filtered equations not closed.

One can also apply the filtering procedure to Eq. (3a), therefore arriving at the following equation governing Lagrangian dynamics of the coarse-grained VGT:

$$\frac{dM_{ij}}{dt} = \widetilde{E}_{ij} + \widetilde{H}_{ij} + \widetilde{T}_{ij} + \Pi_{ij}, \tag{6a}$$

$$\widetilde{E}_{ij} = -\left(M_{ik} M_{kj} - \frac{1}{3} M_{mn} M_{nm} \delta_{ij} \right), \tag{6b}$$

$$\widetilde{H}_{ij} = -\left(\frac{\partial^2 \widetilde{P}}{\partial x_i \partial x_j} - \frac{1}{3} \frac{\partial^2 \widetilde{P}}{\partial x_k \partial x_k} \delta_{ij} \right), \tag{6c}$$

$$\widetilde{T}_{ij} = \nu \frac{\partial^2 M_{ij}}{\partial x_k \partial x_k}, \tag{6d}$$

$$\Pi_{ij} = \Pi_{ij}^r + \Pi_{ij}^s = -\left(\frac{\partial^2 \tau_{ik}}{\partial x_j \partial x_k} - \frac{1}{3} \frac{\partial^2 \tau_{ik}}{\partial x_j \partial x_k} \delta_{ij} \right) + u_k' \frac{\partial M_{ij}}{\partial x_k}, \tag{6e}$$

$$\Pi_{ij}^r = -\left(\frac{\partial^2 \tau_{ik}}{\partial x_j \partial x_k} - \frac{1}{3} \frac{\partial^2 \tau_{ik}}{\partial x_j \partial x_k} \delta_{ij} \right), \tag{6f}$$

$$\Pi_{ij}^s = u_k' \frac{\partial M_{ij}}{\partial x_k}, \tag{6g}$$

where $\boldsymbol{M}$ is our notation for the filtered component of $\boldsymbol{A}$. Notice the appearance of a new (if compared with the original, unfiltered equation) subgrid contribution, $\boldsymbol{\Pi}$, on the right-hand side of Eq. (6a). Furthermore, taking the divergence of the filtered version of the momentum equations and using commutation of the filtering operation with respect to spatial derivative, one arrives at the

following filtered version of the Poisson equation:

$$\frac{\partial^2 \widetilde{P}}{\partial x_k \partial x_k} = -M_{mn}M_{nm} - \frac{\partial^2 \tau_{ik}}{\partial x_i \partial x_k}. \tag{7}$$

This equation has additional terms on the right-hand side compared to Eq. (3b). To make the Lagrangian dynamics of the coarse-grained VGT closed, i.e., expressed via M, one obviously needs to model functional dependence of the pressure Hessian, viscous, and subfilter contributions on M.

III. PIML STRATEGY

In this section, we describe our approach to modeling the generally unclosed unfiltered and coarse-grained Lagrangian dynamics, described by Eqs. (4a) and (6a), via a combination of NNs and physical constraints.

A. Physical foundation of PIML model

1. Nonlocal pressure Hessian

Formal solution for nonlocal pressure Hessian, expressed as a convolution of the second invariant of the VGT, can be obtained by solving Eq. (3b) with the Green's function of the Laplacian operator [43]:

$$H_{ij}(\boldsymbol{x}) = \iiint \frac{\delta_{ij} - \hat{r}_i \hat{r}_j}{2\pi r^3} Q(\boldsymbol{x} + \boldsymbol{r}) \, d\boldsymbol{r}, \tag{8a}$$

$$Q = -\frac{1}{2} A_{mn} A_{nm}, \tag{8b}$$

where $\boldsymbol{r}$ indicates the displacement vector with respect to the point of interest $\boldsymbol{x}$, and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ is the unit displacement vector. Equations (8) show that the nonlocal pressure Hessian depends on the spatial distribution of the second invariant of the velocity gradient tensor, $Q(\boldsymbol{x} + \boldsymbol{r})$.

In this work, we model the nonlocal pressure Hessian term by a surrogate parameterized expression for the spatial distribution of the second invariant Q. We substitute $Q(\boldsymbol{x} + \boldsymbol{r})$, similar to the approach used in Ref. [31], by a function of A. Assuming locality of the relation, which is admittedly a bold hypothesis, the most general form of $Q(\boldsymbol{x} + \boldsymbol{r})$ is the Taylor series expansion:

$$Q(\boldsymbol{x} + \boldsymbol{r}) = a(\boldsymbol{r}) + b_{ij}(\boldsymbol{r})A_{ij} + c_{ijkl}(\boldsymbol{r})A_{ij}A_{kl} + \cdots . \tag{9}$$

Next, and following Lawson and Dawson [31], we enforce in Eq. (9) rotational invariance, i.e., $Q(\boldsymbol{x} + \boldsymbol{r}|A) = Q(\boldsymbol{x} + \boldsymbol{V}\boldsymbol{r}|\boldsymbol{V}A\boldsymbol{V}^T)$, and then arrive at

$$Q(\boldsymbol{x} + \boldsymbol{r}) = b_1 \tau_0^{-2} + b_2(\hat{\boldsymbol{r}}^T A \hat{\boldsymbol{r}}) + b_3 \mathrm{Tr}(A^2) + b_4 \mathrm{Tr}(AA^T) + b_5(\hat{\boldsymbol{r}}^T A^2 \hat{\boldsymbol{r}})$$
$$+ b_6(\hat{\boldsymbol{r}}^T AA^T \hat{\boldsymbol{r}}) + b_7(\hat{\boldsymbol{r}}^T A^T A \hat{\boldsymbol{r}}) + b_8(\hat{\boldsymbol{r}}^T A \hat{\boldsymbol{r}})^2 + \cdots , \tag{10}$$

where τ_0 is the reference timescale and b_i for $i = 1, \ldots,$ are scalar functions of r, assumed related to the two-point correlation functions of Q with different moments of the VGT. In the models analyzed so far, the expansion in Eq. (10) is usually truncated at the second level; see, e.g., [29,31]. In this study, we take advantage of the ML approach and consider full expansion, then extracting functional forms of the coefficients through training.

Equation (10) could be combined with Eq. (8) to perform the integration over $\boldsymbol{r}$. We use the strain rate tensor S and rotation tensor W to denote the symmetric and skew-symmetric parts of the VGT, where $A = S + W$ and $A^T = S - W$. Our final model for the nonlocal pressure Hessian becomes

$$\boldsymbol{H} = \sum_{m,n=0}^{\infty} a_{mn} S^m W^n. \tag{11}$$

This expression coincides with the model for effective viscosity proposed in Pope [44] to represent the anisotropy tensor of Reynolds stresses in the RANS Eulerian approach. The coefficients in Eq. (11) represent weighted integrals of the two-point correlation function of Q with different moments of the VGT [31]. Originally, i.e., in Ref. [44], the equation was further transformed utilizing the symmetric integrity bases and invariants of the S and R tensors and then using the Caley-Hamilton theorem to reduce the high-order products of the tensor to linear combinations of the integrity bases:

$$H = \sum_{n=1}^{10} g_s^{(n)}(\lambda_1, \ldots, \lambda_5) T_s^{(n)}, \tag{12}$$

where $g_s^{(n)}$ are scalar functions of the invariants. In total, there are five independent invariants and 10 integrity bases [44]:

$$\lambda_1 = \text{Tr}(S^2), \quad \lambda_2 = \text{Tr}(W^2), \quad \lambda_3 = \text{Tr}(S^3),$$
$$\lambda_4 = \text{Tr}(W^2 S), \quad \lambda_5 = \text{Tr}(W^2 S^2) \tag{13}$$

$$
\begin{aligned}
T_s^{(1)} &= S, \quad T_s^{(2)} = SW - WS, \\
T_s^{(3)} &= S^2 - \tfrac{1}{3} I \text{Tr}(S^2), \quad T_s^{(4)} = W^2 - \tfrac{1}{3} I \text{Tr}(W^2), \\
T_s^{(5)} &= WS^2 - S^2 W, \quad T_s^{(6)} = W^2 S + SW^2 - \tfrac{2}{3} \text{Tr}(SW^2), \\
T_s^{(7)} &= WSW^2 - W^2 SW, \quad T_s^{(8)} = SWS^2 - S^2 WS, \\
T_s^{(9)} &= W^2 S^2 + S^2 W^2 - \tfrac{2}{3} I \text{Tr}(S^2 W^2), \quad T_s^{(10)} = WS^2 W^2 - W^2 S^2 W.
\end{aligned} \tag{14}
$$

We emphasize that in all the previous studies known to the authors, the functional form of $g^{(n)}$ was greatly simplified and the integrity bases considered were typically limited to second-order [29,31], i.e., the first four bases $T_s^{(1)}, \ldots, T_s^{(4)}$. Then, $g_s^{(n)}(\lambda_1, \ldots, \lambda_5)$ was either derived from a modeled two-point correlation function or reconstructed from experimental or numerical data. In this work, we use a ML model, or more specifically a NN, to learn the functional form of $g_s^{(n)}$.

2. Subfilter contribution

The subfilter contribution Π in the coarse-grained dynamics of the VGT can be divided into two parts: subfilter stress Π^r and subfilter fluctuations Π^s. We take different approaches to model the two terms.

The subfilter stress contribution is nonlocal and asymmetric in nature, and we use an approach similar to the one proposed in Ref. [44] to model the term. We expand the integrity bases to account for the skew-symmetric part and thus consider

$$\Pi^r = \sum_{n=1}^{10} g_s^{(n)}(\lambda_1, \ldots, \lambda_5) T_s^{(n)} + \sum_{n=1}^{6} g_a^{(n)}(\lambda_1, \ldots, \lambda_5) T_a^{(n)}. \tag{15}$$

There are six skew-symmetric bases, and they can be derived following Ref. [45]:

$$
\begin{aligned}
T_a^{(1)} &= W, \quad T_a^{(2)} = SW + WS, \\
T_a^{(3)} &= S^2 W + WS^2, \quad T_a^{(4)} = W^2 S - SW^2, \\
T_a^{(5)} &= W^2 S^2 - S^2 W^2, \quad T_a^{(6)} = SW^2 S^2 - S^2 W^2 S.
\end{aligned} \tag{16}
$$

The subfilter fluctuations come from the spatiotemporal scales that are smaller than the filter size. Specifically, the subfilter contribution to the coarse-grained dynamics is associated with the effects of backscattering, i.e., transfer of energy from scales smaller than the filter scale upscales,

which goes against the average transfer of energy towards smaller scales. We choose to model these backscattering contributions by a stochastic Langevin process:

$$\Pi_{ij}^s = C_{ijkl}dW_{kl}, \tag{17}$$

where dW stands for the Wiener process. The covariance matrix C is fixed and expected to be learned from the "ground-truth" data.

B. Neural network architecture

It is well known since at least the 1990s (see, e.g., Ref. [46]) that NNs are capable of representing an arbitrary smooth function. This theoretical concept received a tremendous boost empirically during the last decade with the so-called deep learning, which made the approach practically feasible, and moreover a powerful tool applicable to a great variety of applications. In this work, we take advantage of the modern deep learning approach to model different contributions to the Lagrangian dynamics of the VGT. We also test the benefits of injecting known physics/constraints into NNs, by comparing the proposed physics-informed machine learning (PIML) strategy to an application ignorant NN approach, which we refer to as physics-blind (PB).

1. Extended tensor basis neural network

To learn the functions contributing to Eq. (11), we follow the approach of Ling *et al.* [35], where a tensor basis neural network (TBNN) was developed. The TBNN maps functions of invariants and integrity bases to another target tensor. The method was used in Ling *et al.* [35] and in a number of follow-up papers to develop ML models for RANS and LES. In this work, we use an extended TBNN approach to learn our physics-informed Lagrangian dynamics of the VGT.

The architecture of the extended TBNN we use in this study is shown in Fig. 1. Inputs of the multilayer deep NN are the five invariants λ_i introduced above. The input layer is connected to a series of hidden layers via the rectified linear unit (ReLU) activation function, which is used to introduce nonlinearities into the ML model. The first output layer is the scalar output with two branches, each representing scalar functions $g_s^{(n)}$ and $g_a^{(n)}$. The scalar output is then combined with the corresponding symmetric and skew-symmetric integrity bases to produce the final tensor output. The extended TBNN expands the tensor bases where the output tensor is residing. While retaining the classical features of TBNN, such as Galilean invariance and rotational invariance, we also guarantee in this architecture enforcement of the incompressibility constraint by modeling the deviatoric/traceless part of the dynamics.

The choice of the hyper parameters of the deep NN, such as the number of layers and number of nodes per layer, is proven to be critical. The NN needs to be deep enough to represent a sufficiently wide class of functions, however also not too deep to avoid overfitting. We have established, by trial and error, that a NN with eight layers and 40 nodes per layer in the hidden layer achieves a good balance. The scalar output layer has either 10 or 16 nodes depending on the types of integrity bases used. The last layer is a "physical" layer with no trainable parameters. The training is performed using the "Adam" optimization method [47] with a standard (quadratic) loss function. We use a varying learning rate, which decreases from 10^{-3} to 10^{-6} throughout the process of training.

2. Physics-blind deep neural network

In order to demonstrate advantages of the proposed PIML model, we construct a naive PB model which does not assume any prior physics input. A natural choice of the NN structure for the PB model is the out-of-the-box deep NN, which is a multi-layer, densely connected, feed-forward neural network. This deep NN gets nine components of the nondimensionalized velocity gradient tensor as input and it outputs nine components of the nonlocal pressure Hessian. In terms of the hyperparameters, we choose to use a NN with eight layers and 40 nodes within each layer, resulting
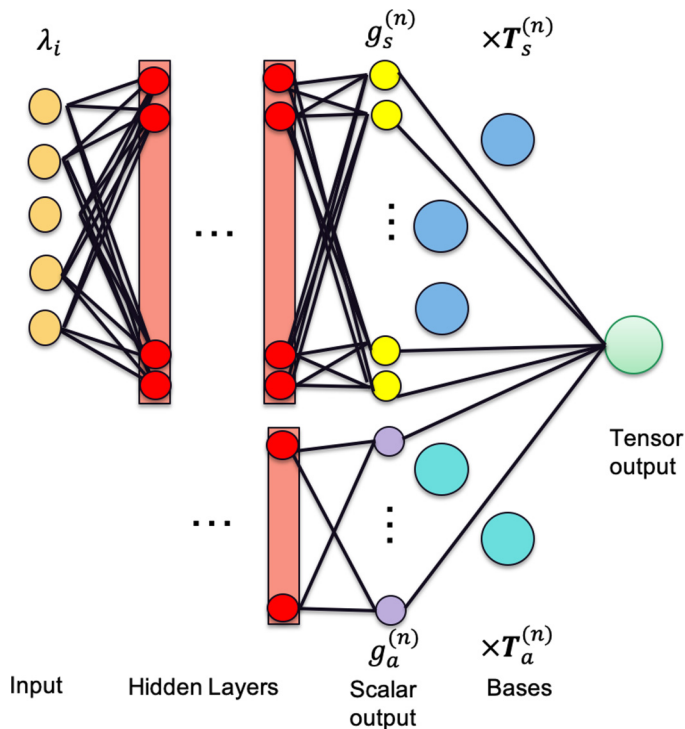
FIG. 1. Architecture of the extended tensor basis neural network used in this work, which is based on the TBNN structure used in Ling *et al.* [35] and extended to include skew-symmetric tensor bases.

in the number of parameters comparable to the number of trainable parameters within the PIML model. The learning rates of the PB model and the PIML model are the same.

IV. "GROUND-TRUTH" DATA

The "ground-truth" data are generated from the Eulerian DNS solution of the incompressible NS equations. Isotropic turbulence is generated on a 1024^3 grid using the pseudospectral method. A large-scale linear forcing term is applied to prevent turbulence from decaying. Time advancement is achieved through Adam-Bashforth-Moulton method. The Taylor Reynolds number when the turbulence reaches a statistically steady state is approximately 250. See Refs. [48,49] for more details on the numerical method. The coarse-grained statistics are obtained by applying a Gaussian filter to the Eulerian data with a chosen filter size l_Δ.

Only Eulerian data are required for the training purpose. In this work, we have sampled up to 4 million data points randomly spanned over space and time, but no improvement is observed when the training set is larger than 0.5 million. Then, turbulence statistics of interest, including statistics of the coarse-grained VGT, pressure Hessian, viscous term, and subfilter term, are calculated using sixth-order Lagrange interpolation from the Eulerian grid to the random locations. The data are then nondimensionalized using the corresponding timescales of the resolved VGT (i.e., Kolmogorov time scale for unfiltered VGT). The invariants and the tensor bases, which are inputs of the extended TBNN, are then calculated using the dimensionless VGT. The train-to-test split ratio is chosen to be 80%:20%.

The Lagrangian data following trajectories of fluid particles are also evaluated and used for diagnostics. The fluid particles are treated as noninertial particles and their velocity is assumed to equal the local flow velocity. The velocity of the fluid particle is interpolated from the Eulerian
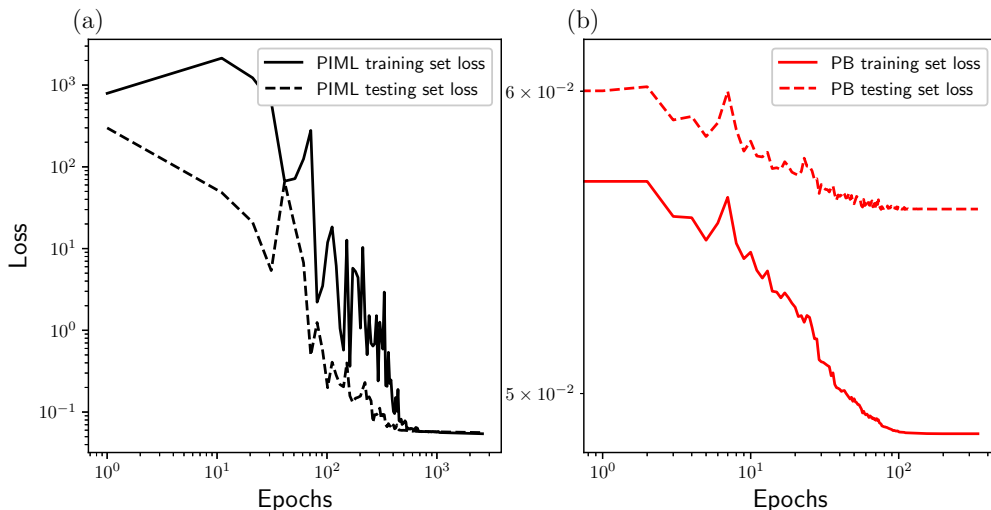
FIG. 2. The decay of both the training and testing loss for the (a) PIML model and (b) PB model.

grid using cubic splines [50]. Trajectories of the fluid particles and the temporal evolution of various statistics are recorded and stored to be used later for training and validation.

V. TRAINING AND TESTING OF THE ML MODELS

We implement and train the PIML model and the PB model using TensorFlow [51] and Keras [52] open-source ML library. The parameters (weights and biases of the NN layers) in both models are randomly initialized using the Glorot normal initialization method. The data set is randomly divided into minibatches of size 4096, therefore, each training epoch takes about 122 iterations. The parameters of the NN are updated by minimizing the quadratic loss function:

$$\text{Loss} = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} \|\boldsymbol{H}_{i,\text{model}} - \boldsymbol{H}_{i,\text{DNS}}\|_F^2, \tag{18}$$

using the Adam optimization method. Here in Eq. (18), $\| \cdot \|_F$ denotes the Frobenius norm of tensors. The initial learning rate is set to 10^{-3} and it is then reduced gradually to 10^{-6} throughout the training process. Training is terminated when both training and testing losses are saturated. Figure 2 shows the training and testing losses for both PIML and PB models as a function of the number of epochs. We observe that the initial loss for the PIML model is much larger than one observed in the PB model. This is due to the fact that the high-order tensor basis is highly intermittent, and therefore random initialization of the parameters results in a large loss. Since the initial loss of the PIML model is orders of magnitude larger than one in the PB model, it takes more epochs for the PIML model to fully converge. At the final training epoch, the PIML model reaches a similar training/testing loss of $4.80 \times 10^{-2}/4.90 \times 10^{-2}$. On the other hand, the final testing loss is 5.58×10^{-2} for the PB model, that is larger than the training loss of PIML. This shows that injecting the known physical constraints into the NN structure can provide physics-based regularization for training of the ML models.

Another advantage of the physics-informed approach is that the physical constraints and invariance can be automatically preserved by the build-in NN structure. On the other hand, when working with the PB model, we blindly rely on the NN to discover the underlying physical constraints and invariance. In Fig. 3 we provide comparison of the testing errors of the incompressibility constraints
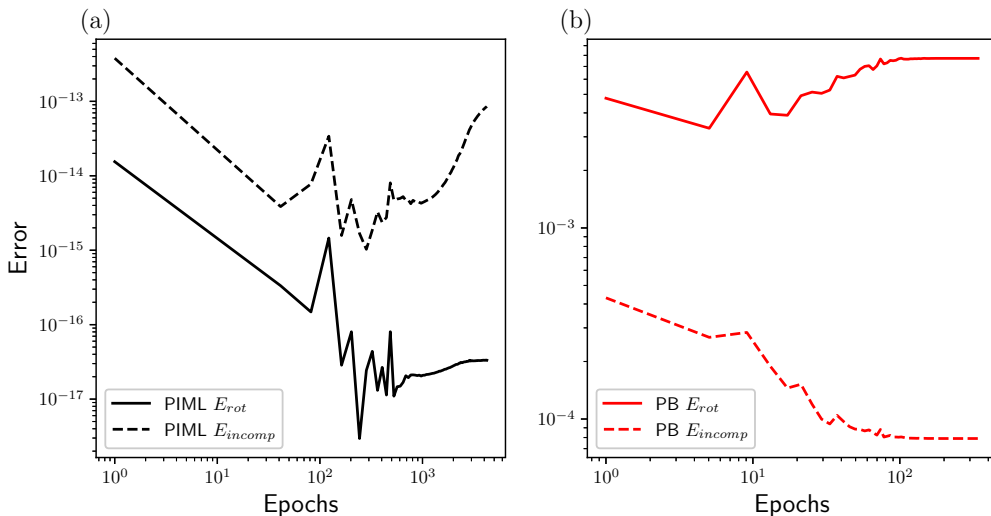
FIG. 3. The testing error of the incompressibility constraint and rotational invariance of the (a) PIML and (b) PB models.

(E_{incomp}) and rotational invariance (E_{rot}) from the PIML and PB models. The errors are calculated using the following formulas:

$$E_{\mathrm{incomp}} = \frac{1}{n_{\mathrm{test}}} \sum_{i=1}^{n_{\mathrm{test}}} |\mathrm{Tr}(\boldsymbol{H})|, \tag{19a}$$

$$E_{\mathrm{rot}} = \frac{1}{n_{\mathrm{test}}} \sum_{i=1}^{n_{\mathrm{test}}} \|\boldsymbol{\Omega} f(\boldsymbol{A}) - f(\boldsymbol{\Omega} \boldsymbol{A})\|_F^2, \tag{19b}$$

where $\boldsymbol{\Omega}$ denotes a random rotation operator and f denotes the ML models. As expected, the PIML model conserves both incompressibility constraints and rotational invariance throughout the training and respective errors are around the machine precision. On the other hand, the PB model is unable to fully capture the underlying invariance.

VI. UNFILTERED VELOCITY GRADIENT TENSOR

The PIML model described in Sec. III is trained using fully resolved data for the VGT and the pressure Hessian statistics. In this section, the performance of the trained PIML Lagrangian model is analyzed from a number of perspectives. The unclosed nonlocal pressure Hessian is in the prime focus of the analysis. A stochastic ODE is then constructed using the newly learned model to check if the statistics is reproducible.

A. *A priori* results

Aside from the Frobenius norm of the tensor, which is used as the loss function in training, an important diagnostic of the predicted pressure Hessian is associated with the alignment of its eigenvectors and magnitudes of its eigenvalues with respect to characteristics extracted from the "ground-truth" data. Directions of the eigenvectors are known to play an important role in the dynamics of the VGT, which are related to flow topology, vortex stretching mechanism, dissipation, etc. In this section, the PIML model of the nonlocal pressure Hessian is developed using the extended TBNN. It is then compared with a number of prior attempts to model the nonlocal pressure Hessian. Specifically, we juxtapose our model to the empirical models developed in Refs. [29–31],
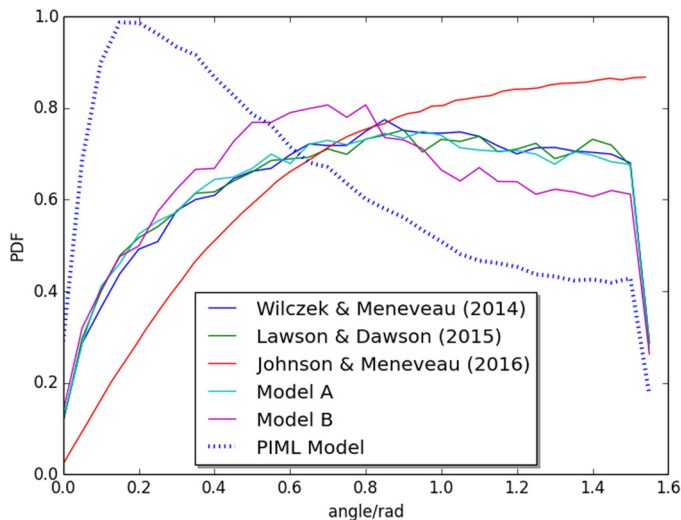
FIG. 4. PDFs of the angle (in radians) between the predicted e_1 eigenvector and DNS data. The PDF peaks at 0 when the predicted e_1 eigenvector is aligned with the DNS data and it peaks at $\pi/2$ in the case of misalignment.

which rely on the aforementioned arguments for the pressure Hessian and subgrid terms locality, but do not use a NN to represent functional dependence of the modeled terms on the VGT. To emphasize the role of higher-order functions in our local modeling, we also introduce hierarchy of models by constructing two simplified ML models: model A, for which $g_s^{(n)}$ are set to constants (rather than functions of λ_i), and model B, for which $g_s^{(n)}$ are set to linear functions of λ_i.

Symmetries of the nonlocal pressure Hessian allow us to reduce the discussion to an analysis of its three distinct real eigenvalues (denoted as $e_1 < e_2 < e_3$) and respective eigenvectors. Figure 4 shows the PDF of the angle between predicted and "ground-truth" directions of the e_1 eigenvector, corresponding to the most negative eigenvalue. If the PDF peaks at $(\pi/2)$ rad, it indicates that the model predictions are mostly misaligned with the "ground-truth" data. If the PDF peaks at 0 rad, the model eigenvector is likely to be aligned with the "ground-truth" data. Perfect alignment with a single-valued PDF of 0 rad is impossible because of the approximate nature of the empirical model (we approximate a clearly nonlocal problem with a local model). Figure 4 shows that the only model that can capture the alignment is our PIML model. Second in performance is the simplified ML model B (with $g_s^{(n)}$ assumed to be linear in λ_i). Even though the performance of the linearized PIML models (both ML model A and ML model B) are somehow better than that of other non-ML models, the improvement is limited. We relate the poor performance of the non-ML models in the alignment test to the fact that the models ignore high-order functions of the invariants and bases. Simplified ML models A and B improve the representation of the functional space; however, in only a limited way as the coefficients are restricted to being constants or linear functions. The ability of our basic PIML model to learn sufficiently many higher-order functions makes it significantly better in performance than other models. The limitations of the non-ML based models were foreshadowed in Lawson and Dawson [31].

The eigenvalues of the nonlocal pressure Hessian, known to play an important role in explaining physical details of the energy cascade, type of the fluid element deformation, vorticity-strain alignment, and vortex stretching are tested next. By construction, the trace of the nonlocal pressure Hessian tensor is zero, so that the sum of the eigenvalues is also zero, $e_1 + e_2 + e_3 = 0$. We denote these eigenvalues such as e_3 is largest (mostly positive) eigenvalue, e_1, is the smallest (mostly negative) eigenvalue, and e_2 is the intermediate eigenvalue and can be either negative or

TABLE I. Correlation coefficients of the eigenvalues.

Models	corr($e_{1,\mathrm{model}}, e_{1,\mathrm{DNS}}$)	corr($e_{2,\mathrm{model}}, e_{2,\mathrm{DNS}}$)	corr($e_{3,\mathrm{model}}, e_{3,\mathrm{DNS}}$)
Wilczek and Meneveau [29]	0.69	0.11	0.71
Lawson and Dawson [31]	0.69	0.04	0.72
Johnson and Meneveau [30]	0.13	0.06	0.12
Model A	0.70	−0.07	0.73
Model B	0.61	0.09	0.67
PIML model	0.95	0.79	0.94

positive. Table I shows correlation coefficients between model prediction and DNS (ground truth) for the three eigenvalues. Evidently, the PIML model shows a significant improvement over all other models for all the eigenvalues, therefore giving a better prediction of the dynamics. Another observation is that the correlation coefficients corresponding to e_1 and e_3 are generally much higher than those for e_2. We relate it to the fact that by construction e_2 is smaller in magnitude, and it can also be positive or negative, therefore making it harder to predict its mean value accurately.

The advantage of the PIML model based on TBNN is that, unlike in the case of the black-box PB model, it allows to interpret the trained model by extracting the scalar coefficients and compare them with physics-based models. The general functional form of previous physics-based models used in the past [29,31,32] can be written as

$$\boldsymbol{H} = \delta\boldsymbol{S} + \gamma(\boldsymbol{SW} - \boldsymbol{WS}) + \alpha\big(\boldsymbol{S}^2 - \tfrac{1}{3}\boldsymbol{I}\mathrm{Tr}(\boldsymbol{S}^2)\big) + \beta\big(\boldsymbol{W}^2 - \tfrac{1}{3}\boldsymbol{I}\mathrm{Tr}(\boldsymbol{W}^2)\big). \qquad (20)$$

The main tuning parameters for the model are δ, γ, α, and β, which can be viewed as the scalar coefficients $g_s^{(1)}$ to $g_s^{(4)}$. In Fig. 5 we compare the PDFs of the four scalar coefficients extracted from the PIML model and the ML model A, which are computed using the testing set, with the corresponding tuning parameters taken from previous studies, i.e., based on the Gaussian closure [29] and on the second-order basis expansion [31]. Note that model A is based on the same TBNN structure, however with 10 scalar functions $g_s^{(n)}$ set constant. We observe that the learned scalar coefficients in the model A are very close to the parameters in Lawson and Dawson [31], but the coefficients are very different from those derived using the Gaussian closure model. This is because the parameters in Lawson and Dawson [31] are also obtained by matching with DNS data using second-order basis expansion. In the case of the PIML model, the predicted scalar coefficients are no longer constant—they are nonlinear functions of the invariants. Figure 5 shows that by expanding the functional space of the scalar coefficients using a NN, the model is able to capture nontrivial nonlinear dependence on the invariants originating from the projection of high-order polynomials to the low-order bases. This nontrivial nonlinear dependence shifts the peak values of the PDFs away from the constant coefficients in model A. We also note that the PDFs of the learned scalar coefficients are generally narrow in the center but also showing wide tails. This is likely due to the high intermittency of the high-order invariants.

We then compare contributions from different integrity bases using the average Frobenius norm of $g_s^{(n)}\boldsymbol{T}^{(n)}$ for different models. As shown in Fig. 6, the Frobenius norm of contributions from the first four integrity bases are comparable across the PIML model, model A, and Lawson and Dawson [31], and they are also larger than those seen in the Gaussian closure model. For high-order bases, model A is able only to capture nontrivial contributions from $\boldsymbol{T}^{(5)}$, while the PIML model, with the functional space enriched by a NN, is able to capture all the high-order contributions. Even though it seems that the average contributions (as measured by the Frobenius norm) are small for $\boldsymbol{T}^{(n)}, n \in 5, \ldots, 10$, they are highly intermittent and may play an important role in improving the orientation of the eigenvectors, as observed in Fig. 4.

This analysis confirms that by expanding the functional space using TBNN, the PIML model is capable of predicting rather accurately orientation/eigenvectors and magnitude/eigenvalues of the nonlocal pressure Hessian. Encouraged by these results, let us take a step further and study how the
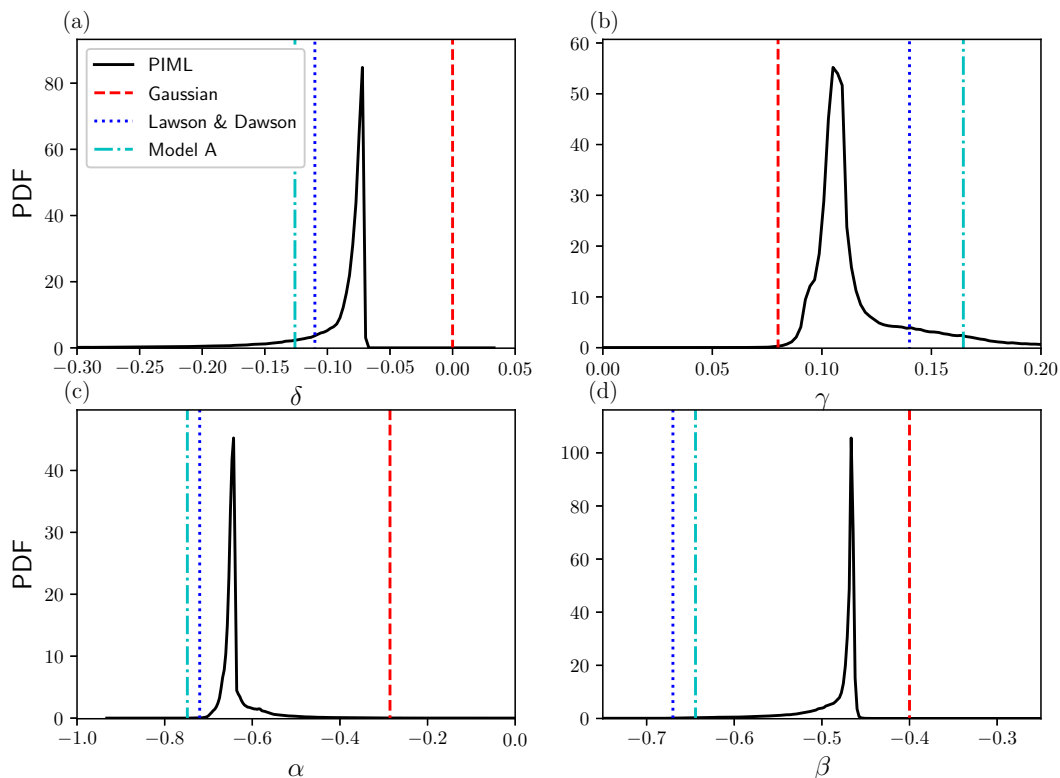
FIG. 5. PDFs of the coefficients of the tensor bases from the PIML model, ML model A, Gaussian closure [29], and Ref. [31].
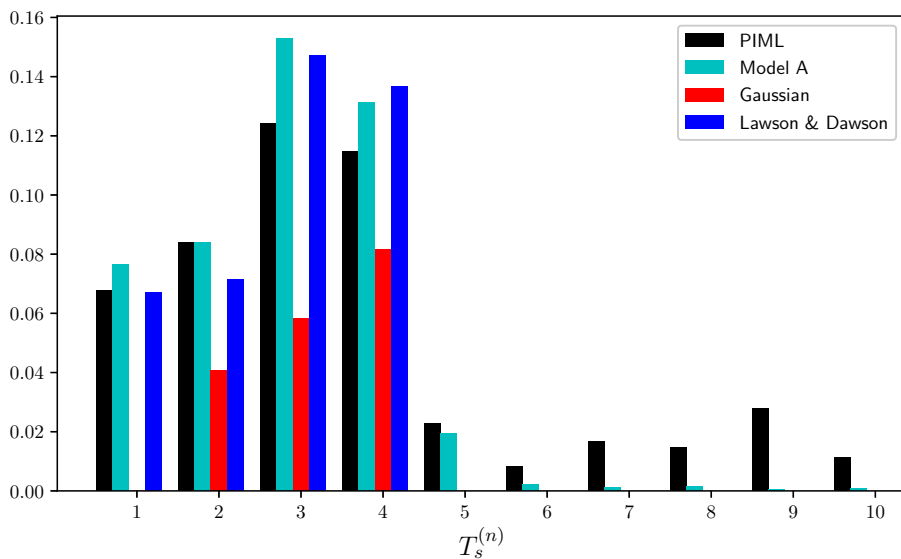

FIG. 6. Contributions of the different tensor bases to the prediction of nonlocal pressure Hessian for the PIML model, ML model A, Gaussian closure [29], and Ref. [31].
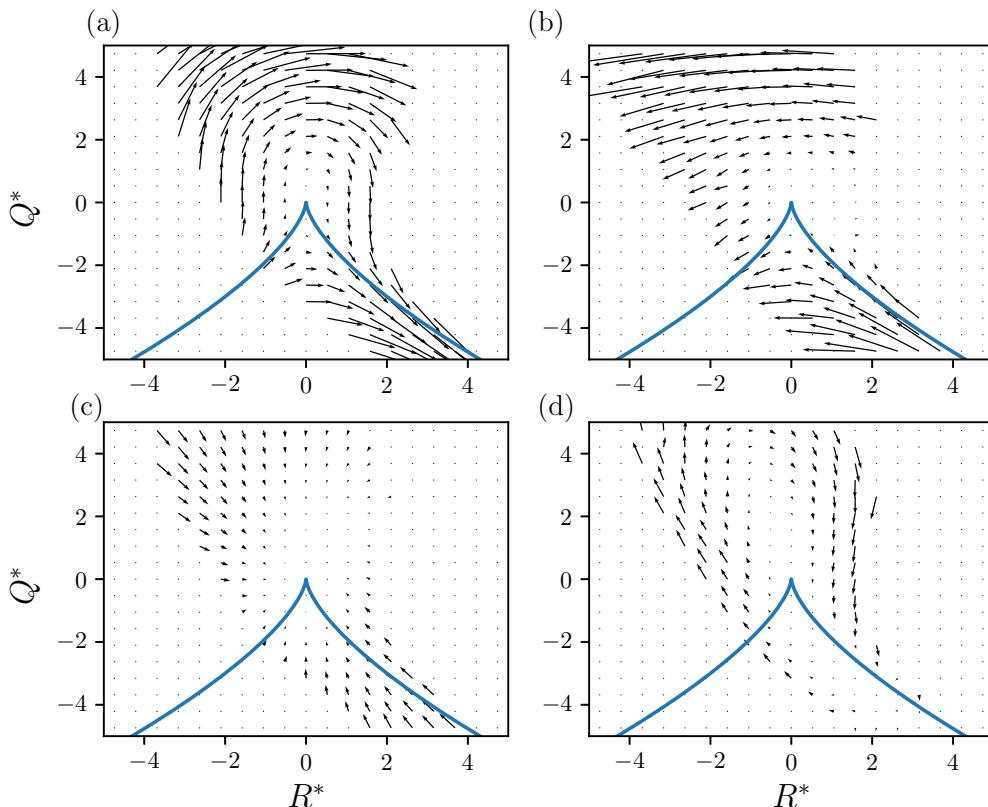
FIG. 7. Lagrangian dynamics projected onto the Q-R phase plane: (a) restricted Euler term, (b) pressure Hessian term, (c) viscous term, and (d) sum of all the terms. Zero discriminant line, $D = 27R^2 - 4Q^3 = 0$, descriptive of the RE dynamics [19], is shown in blue.

PIML model represents statistics of the VGT second invariant, Q, defined according to Eq. (8b) and the third invariant, $R = -\frac{1}{3} A_{mn} A_{nl} A_{lm}$. Specifically, we follow the Q-R plane dynamics of the VGT, widely used to describe the geometry of turbulent flows [9–15,24]. The Lagrangian dynamics, as described by Eq. (4), can then be projected onto the Q-R plane:

$$\frac{dQ}{dt} = -3R + A_{mn}H_{nm} + A_{mn}T_{nm}, \tag{21a}$$

$$\frac{dR}{dt} = \frac{2}{3}Q^2 + A_{mn}A_{nl}H_{lm} + A_{mn}A_{nl}T_{lm}. \tag{21b}$$

Dynamics (rate of change) within the Q-R plane is described, naturally, by the vector $(\frac{dQ}{dt}, \frac{dR}{dt})$, formed from the components on the left-hand sides of the equations. One can also form respective two-component vectors describing the Q-R plane dynamics associated with the RE, pressure Hessian, and viscous contributions on the right-hand sides of Eqs. (21). Figure 7 shows the dynamics from different contributions on the Q-R plane using conditional mean vector (CMV).

Let us make a quick detour and comment on the dynamics of different vectors in the Q-R plane, as extracted from the DNS and shown in Fig. 7. The restricted Euler dynamics is deterministic and, if no other terms are present, should move the "QR particle" largely along the zero discriminant line. The RE term extracted from DNS, seen in Fig. 7(a), shows this as a general trend. As shown in Fig. 7(b), the pressure Hessian term counters the RE effects in some areas and moves the
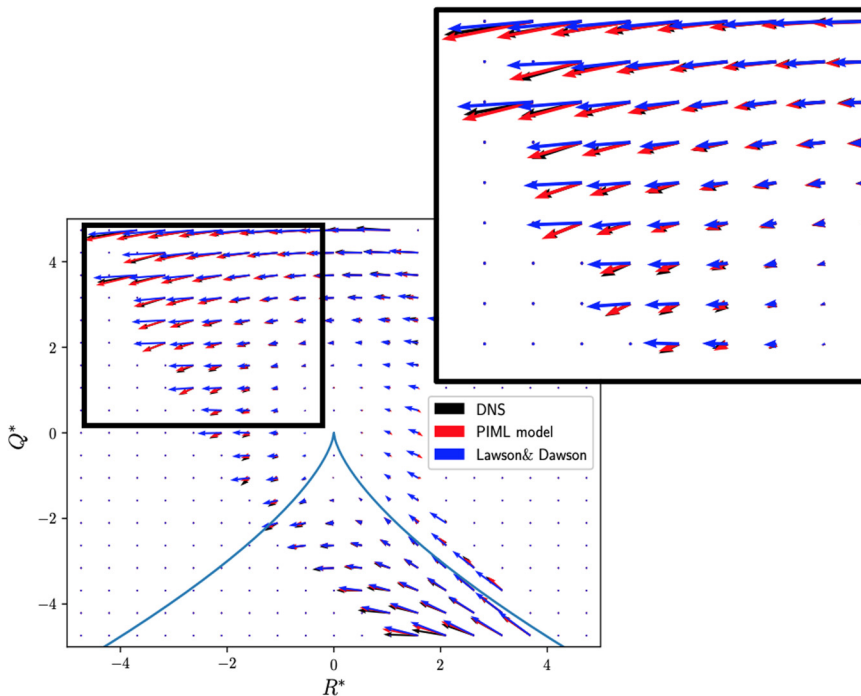
FIG. 8. Comparison of pressure Hessian contributions on the Q-R phase plane.

"QR particle" towards the left part of the Q-R plane. The QR vector associated with the viscous contribution largely points towards the origin, consistent with the fact that it aims to damp the turbulent fluctuations. The contribution summing up all the three vectors based on the right-hand side of Eqs. (21) and shown in Fig. 7(d) results in the classical clockwise motion around the zero discriminant line resulting in the tear-drop shape of the Q-R plane PDF [24].

Returning back to the analysis of the PIML model performance, we now turn to the Q-R plane comparison of PIML and also non-ML-based empirical model from Ref. [31] with DNS, shown in Fig. 8. Evidently, the PIML model improves over the non-ML-based model [31] in terms of representing both magnitude and direction of the CMV, especially in the bottom left and bottom right quadrants. Substituting predictions of the PIML model for the right-hand side terms in Eq. (21) one can extract PIML dynamics of the QR vector shown in Fig. 9. We observe that the PIML model results in a reasonably accurate reproduction of the clockwise dynamics in the Q-R plane.

Summarizing results discussed in this section—we observe that the PIML model performs better than the reduced empirical models constructed so far, in terms of reproducing reasonably well DNS results for Kolmogorov scale dynamics of the VGT tensor, expressed via magnitude and orientation of its eigenvalues and eigenvectors, and dynamics of its second and third invariants.

B. *A posteriori* analysis

After the *a priori* analysis of the PIML model, which has revealed its success in representing the DNS results, it seems natural for the next move to perform a more robust test consisting of substituting the PIML model for various terms on the right-hand side of the Lagrangian Eqs. (3a). However, the nonlocal and stochastic nature of the contributions suggest that we should also augment the deterministic PIML terms with stochastic corrections. This suggests we need to model the original unclosed Lagrangian ODE [Eq. (3a)] by a closed stochastic ODE (SODE). Inspired by previous (however, non-ML based) papers advancing this line of thought [24,29,30] we consider the
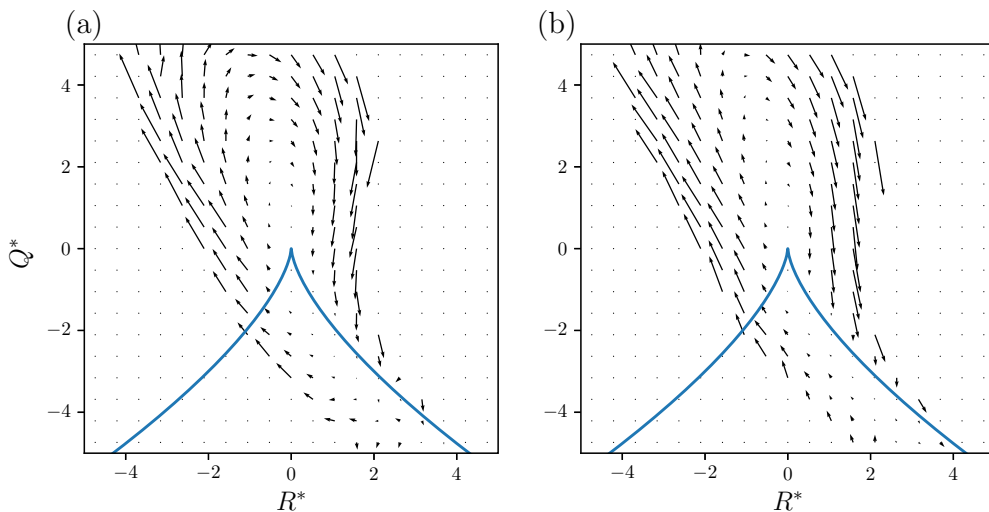
FIG. 9. Comparison of Lagrangian dynamics between (a) DNS and (b) PIML model on the Q-R phase plane.

following stochastic model:

$$\frac{dA_{ij}}{dt} = E_{ij} + H_{ij,\text{PIML}} + T_{ij,\text{model}} + b_{ijmn}dW_{mn}, \tag{22}$$

where the newly introduced b is the covariance matrix of the stochastic (Wiener/Langevin) contribution to the dynamics. We borrow the form of the covariance matrix from Ref. [30], $b_{ijmn} = -\frac{1}{3}\sqrt{\frac{D_s}{5}}\delta_{ij}\delta_{mn} + \frac{1}{2}(\sqrt{\frac{D_s}{5}} + \sqrt{\frac{D_a}{3}})\delta_{im}\delta_{jn} + \frac{1}{2}(\sqrt{\frac{D_s}{5}} - \sqrt{\frac{D_a}{3}})\delta_{in}\delta_{jm}$ and learn the tuning parameters D_a and D_s from the residual between "ground-truth" and model prediction. In addition, we use a simple linear diffusion to model the viscous contributions T, where the coefficients, defined as linear scalar function of invariants, are learned from the data.

To have a preliminary assessment of the stochastic model validity we perform *a posteriori* test, consisting of seeding a large number (one million) of particles initialized with a random Gaussian VGT and then evolving each particle according to the SODE Eqs. (22). In this experiment, we use a second-order predictor-corrector method to advance the particles in time. The total integration is performed for half eddy turnover time, where the eddy turnover time is calculated using the formula l/u'_{rms}, where l is the integral length scale of the flow. Figure 10(b) shows the Q-R plane of the PDF
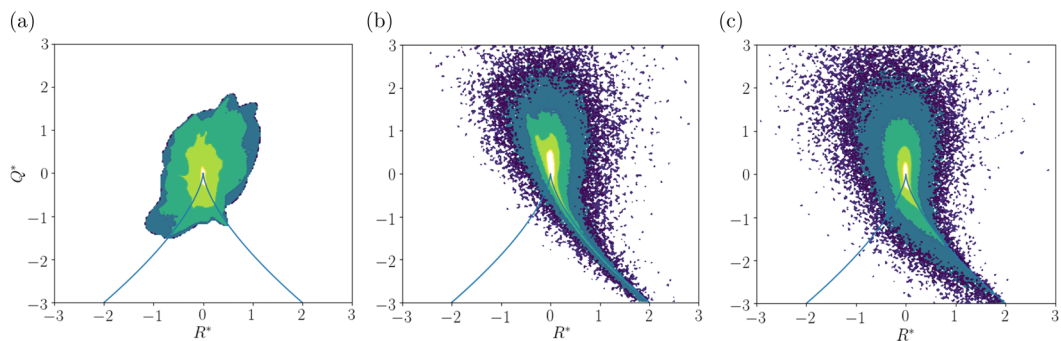

FIG. 10. Q-R plane PDF. (a) Random initialization (Gaussian, traceless VGT), (b) PIML SODE prediction, and (c) DNS.
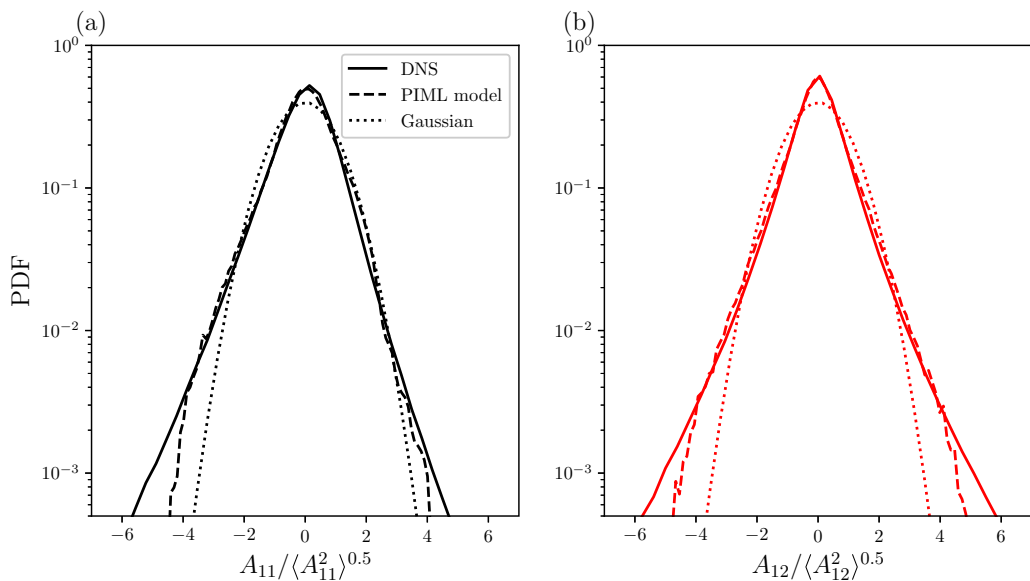
FIG. 11. Standardized PDFs of the longitudinal component: (a) A_{11} and transverse component; (b): A_{12} of the VGT.

which we get for this experiment, as compared with the "ground-truth" data shown in Fig. 10(c) and the result we get assuming a Gaussian i.i.d. statistics for the components of the VGT. Obviously, the Gaussian example shown in Fig. 10(a) (which we use to initiate the SODE PIML dynamics) is completely off in terms of the shape and spread of the joint distribution. However, comparison of the resulting SODE map of the Gaussian seed, shown in Fig. 10(b), with the "ground-truth"
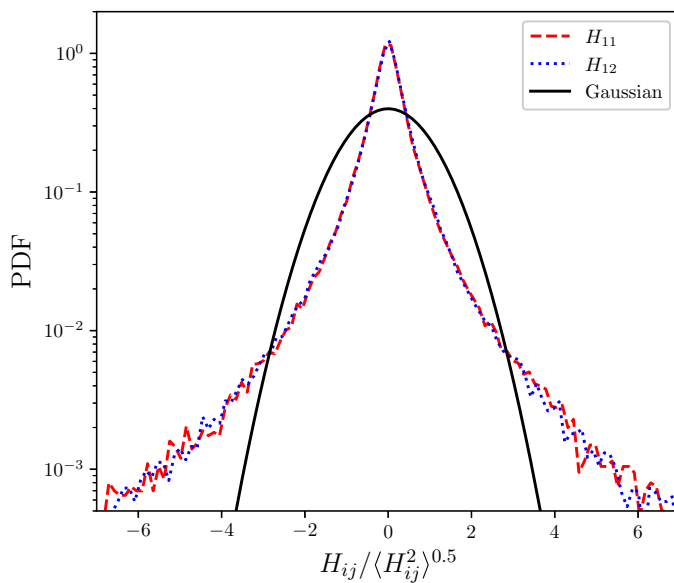

FIG. 12. Standardized PDFs of the PIML model error of the longitudinal H_{11} and transverse H_{12} components of pressure Hessian.
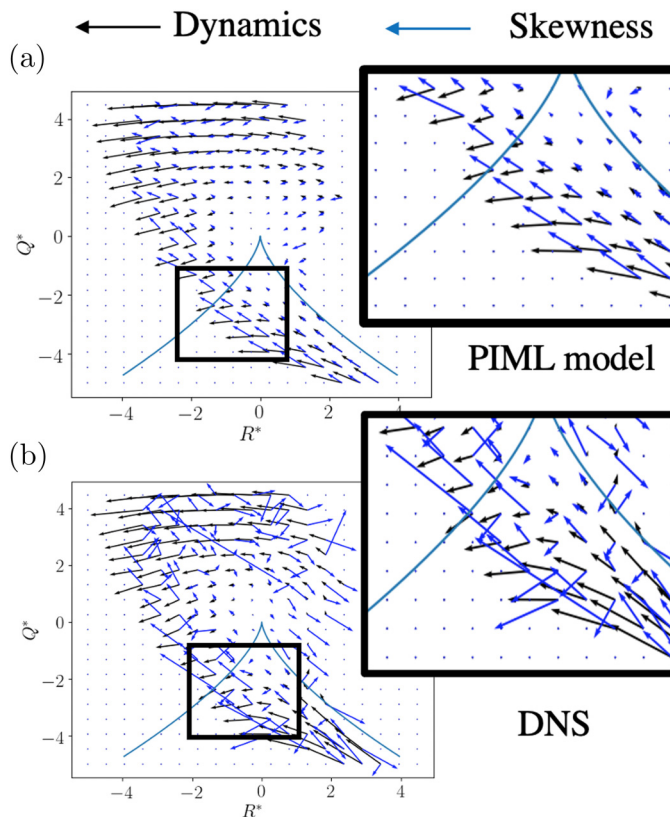
FIG. 13. CMV of the PIML model prediction on the Q-R phase plane. Both the mean (black arrows) and skewness (blue arrows) are shown with zoomed-in view of the third and fourth quadrants.

data is satisfactory. The two share the similar tear-drop shape attributed to certain preferred local configurations associated with vortex stretching, and similar increased spread related to the large intermittency of small-scale turbulence.

To further quantify the statistical properties of the VGT generated from the *a posteriori* test, we compare the standardized PDFs of the longitudinal, A_{11}, and transverse, A_{12}, components directly with those calculated from the DNS data, as well as their random Gaussian initialization in Fig. 11. Evidently, our PIML SODE is capable, given the initial Gaussian field, of reproducing non-Gaussian features of the VGT. In particular, the skewed PDF of the longitudinal component can be captured, even though the skewness factor is underpredicted (-0.3 for the PIML model and -0.55 for DNS). For the transverse component, the predicted PDF is symmetric (skewness factor around 0), which is in full agreement with the DNS data.

Upon further (more detailed) examination of the PIML results, we note that the PIML SODE predicts a narrower tail along the zero discriminant line compared to the DNS. We also observe that in the PIML-based experiments there are fewer particles ending in the left-bottom and right-bottom quadrants of the Q-R plane. As shown in Fig. 7, the contribution from the pressure Hessian plays a major role in moving QR particles into the left-bottom and left-up quadrant. Besides, it is known from Tian *et al.* [15] that the pressure Hessian has the largest magnitude, as compared with other contributions to the dynamics, and it is also highly non-Gaussian (intermittent). These observations suggest that even though the stochastic PIML model is capable of predicting rather accurately the mean value of the nonlocal pressure Hessian, it fails to represent correctly its higher-order moments, e.g., skewness, kurtosis, and tails of the pressure Hessian distribution. To make this
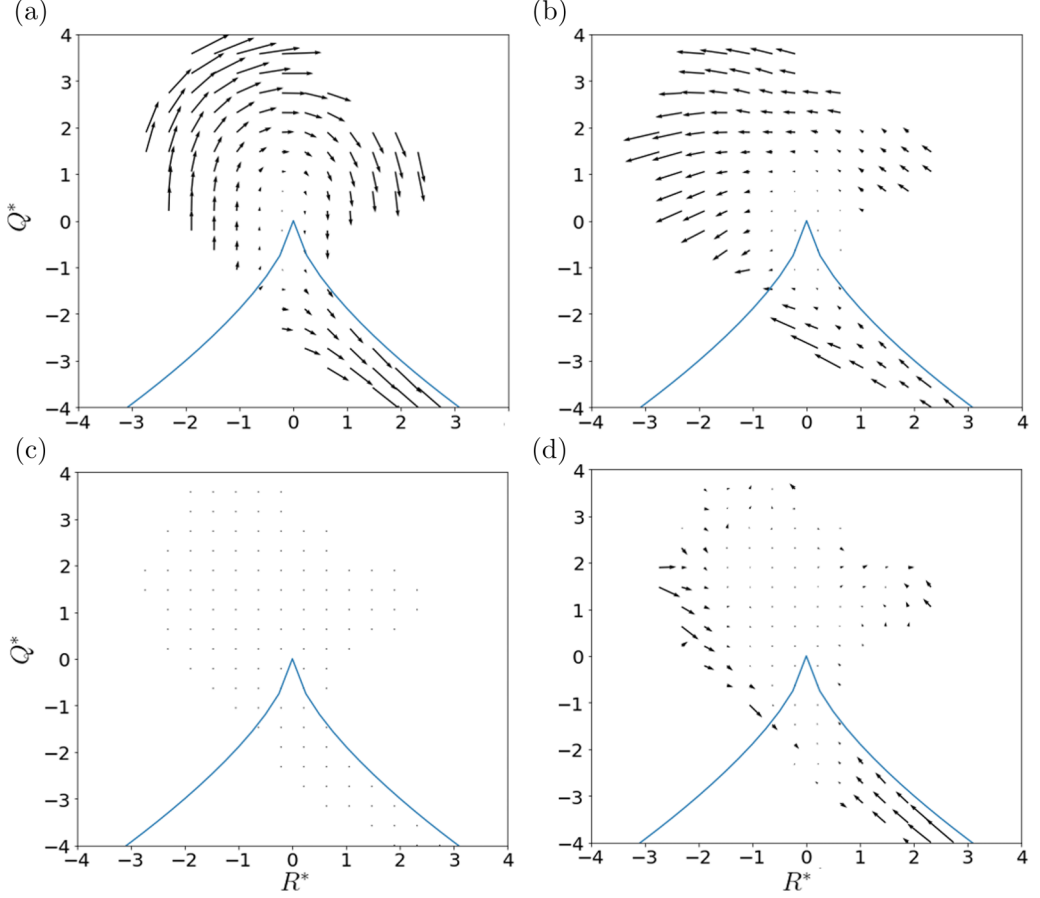
FIG. 14. CMV of the coarse-grained VGT dynamics with filter scale selected to be, $l_\Delta = \pi/8$. (a) restricted Euler, (b) pressure Hessian, (c) viscous term, and (d) subfilter contribution.

statement clear, we show standardized error in the SODE-PIML prediction in Fig. 12. The error is calculated using the residual between pressure Hessian predicted by PIML model and "ground-truth" DNS, $\boldsymbol{H}_{\mathrm{DNS}}$-$\boldsymbol{H}_{\mathrm{PIML}}$. As expected, the PDF shows tails that are much wider than one would expect from nonintermittent (Gaussian) statistics. We conclude that SODE-PIML misses rare events pushing more QR particles into the left quadrants of the QR plane. Misrepresentation of statistics by SODE-PIML is further quantified in Fig. 13, where we show in addition to the mean vector (black), the respective vector associated with the skewness (blue) of the Q-R plane vector associated with the right-hand side of Eqs. (4a). We clearly see here a rather significant mismatch in skewness between the SODE-PIML and the DNS results observed in the left part of the Q-R plane.

Several factors could contribute to the fact that the PIML model fails to predict these high-order non-Gaussian statistics: the choice of the loss function and the choice of stochastic terms. By using the quadratic loss function and a Gaussian stochastic term, the error is essentially assumed to be Gaussian and therefore high-order statistics cannot be represented. Future work on developing better loss functions and stochastic term will be discussed in the last section. Another caveat of the constructed SODE is that the viscous contributions are approximated using a simple linear diffusion model, which is generally not enough to damp high-order nonlinear contributions from the ML models. This is a common issue of all the ML models. It is generally difficult to devise a high-order nonlinear damping term, like the one proposed in Ref. [32], to prevent instability of the
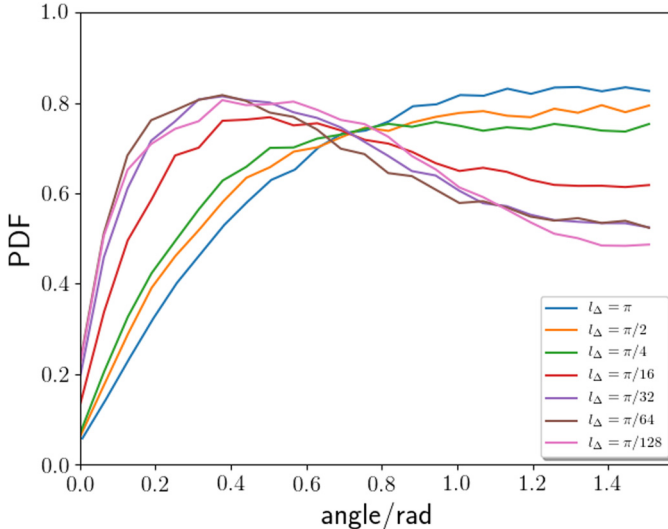
FIG. 15. PDF of the angle (radian) between the predicted e_1 eigenvector and DNS data for various filter sizes.

Lagrangian particle evolution. We are working on improving stability of the Lagrrangian dynamics taking advantage of a number of recently developed ML techniques.

VII. *A PRIORI* ANALYSIS OF COARSE-GRAINED VELOCITY GRADIENT TENSOR

We will start the discussion of the process of coarse graining with a disclaimer. Coarse-graining/filtering of the Lagrangian dynamics of the VGT adds a "geometrical" complexity to the problem, in the sense that even if it is spatially isotropic, i.e., filtering is over a ball of a fixed size, the ball will evolve acquiring a complicated, typically multiscale shape. This is the foundation for some of the advanced physics-based Lagrangian models [24,30]. Having acknowledged the complexity of this effect associated with the Lagrangian description, we will not address it here in full. Instead, we limit the description to only spatial filtering over a static and isotropic volume and focus instead on extending modeling of the small-scale VGT based on locality to that of the coarse-grained VGT on a fixed volume.

We start this discussion with an analysis, illustrated in Fig. 14, of what DNS shows for different terms on the right-hand side of Eqs. (6a) where the scale of coarse graining is selected to be $l_\Delta = \pi/8$, i.e., somewhere within the inertial interval limited by the energy-containing scale, estimated by π, and the viscous/Kolmogorov scale, estimated by $\pi/512$. Note that the velocity forcing used here, following Refs. [48,49], enforces the Kolmogorov scale at the onset, while energy is added at wave numbers $k \leqslant |1.5|$. We observe that the QR-flow of the coarse-grained RE term and the coarse-grained pressure Hessian term are similar to the respective unfiltered terms. We also observe that the viscous contribution becomes small and its "eddy dissipation" role is taken by the subfilter term.

Now we turn our attention to the focal point of this section—empirical modeling with a NN of the pressure Hessian and subfilter contributions. The coarse-grained pressure Hessian is modeled using the same TBNN with the symmetric integrity bases as used above in the unfiltered (single-point) case. We consider coarse graining at different scales independently, therefore train the TBNN models for each scale anew. Figure 15 shows the alignment between model prediction and DNS for the pressure Hessian coarse-grained at different scales. We note that the predicted alignment is not as good as that in the unfiltered PIML model. This means that adding a length scale to

TABLE II. Correlation coefficients of the eigenvalues for different scales of coarse graining.

Filter size	$\mathrm{corr}(e_{1,\mathrm{model}}, e_{1,\mathrm{DNS}})$	$\mathrm{corr}(e_{2,\mathrm{model}}, e_{2,\mathrm{DNS}})$	$\mathrm{corr}(e_{3,\mathrm{model}}, e_{3,\mathrm{DNS}})$
$l_\Delta = \pi$	0.17	0.21	0.31
$l_\Delta = \pi/2$	0.35	0.07	0.33
$l_\Delta = \pi/4$	0.50	0.008	0.51
$l_\Delta = \pi/16$	0.61	0.06	0.63
$l_\Delta = \pi/32$	0.66	0.12	0.67
$l_\Delta = \pi/64$	0.71	0.16	0.72
$l_\Delta = \pi/128$	0.72	0.20	0.73

the dynamics makes it harder for predicting the orientation. As expected, the model performance converges towards the unfiltered model as the filter size decreases. Table II shows correlation coefficients between eigenvalues resulting from the PIML model and DNS data coarse-grained at different scales. We note that the performance of the same TBNN architecture becomes worse as the filter size increases. This analysis suggests that the TBNN architecture is less suitable for learning coarse-grained dynamics as the scale of coarse graining departs significantly from the viscous scale.

Figure 16 shows PDFs of prediction error for PIML models of different coarse-grained scales. Interestingly, the error becomes more and more Gaussian-like when the filter size increases. This indicates that the Langevin type SODE model is better suited for modeling coarse-grained dynamics than those occurring at small scales.

The subfilter contribution is modeled using the extended TBNN framework with all 16 integrity bases (10 symmetric and six skew symmetric) included. Figure 17 compares the CMV of subfilter contribution for coarse-graining size $l_\Delta = \pi/8$. The PIML model can generally predict the qualitative behavior of the dynamics, but there still exists a large error in the magnitude, especially in the third quadrant. On top of that, the lack of explicit length scale information in the model makes it harder to generalize. This indicates an improved modeling strategy is required for coarse-grained dynamics.
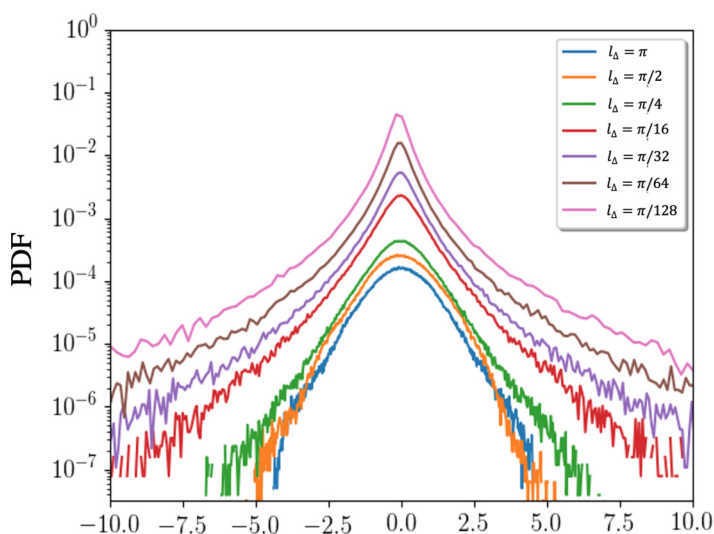

FIG. 16. PDF of the error of PIML model prediction for various filter sizes. The curves are renormalized to avoid clustering.
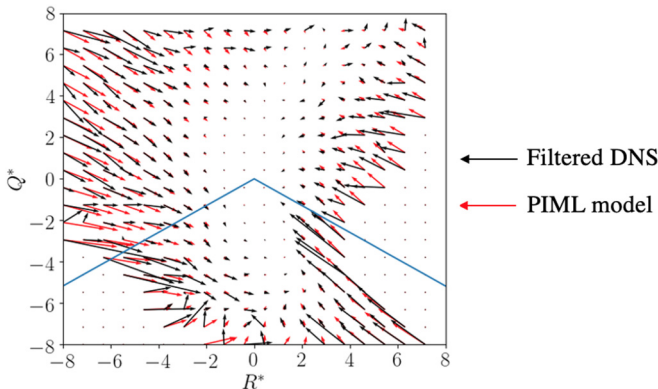
FIG. 17. Comparison of subfilter CMV between the PIML model prediction and DNS. PIML model prediction is shown as red and DNS data are shown as black. The chosen coarse-graining size is $\pi/8$.

VIII. CONCLUSIONS AND FUTURE WORK

Modeling the velocity gradient tensor (VGT), evaluated locally at a point in the Eulerian and/or Lagrangian frames, and/or averaged over scales within the inertial range of turbulence, is the focus of this manuscript. We are not the first to study this interesting object, which is known to be important for understanding quantitatively and qualitatively many principal features of developed incompressible turbulence, such as geometry and statistics of energy transfer, vorticity-strain alignment, mixing, etc. In this work, we aim at developing a reduced empirical model of a stochastic ODE type describing the VGT evolution at both small scales and coarse-grained scales. We take advantage of the emerging Machine Learning techniques/tools and inject them with physics-inspired and mathematics-based constraints to develop a mixed approach: physics informed machine learning. This allowed us to go much deeper than before in terms of quality achieved in reproducing the VGT with a reduced model. We show in this paper that the PIML approach applied to the Lagrangian evolution of the coarse-grained VGT is advantageous over existing methods. To achieve this conclusion, it required overcoming a number of obstacles which we find useful to list and discuss again below.

A major difficulty in modeling dynamics of the VGT in incompressible flow is related to its nonlocality of pressure contributions inherited from the principal nonlinearity of the microscopic description. We tackle this problem by proposing an empirical model for the spatial distribution of the second invariant of the VGT that satisfies Galilean invariance, rotational invariance, and incompressibility condition. After utilizing an ansatz for nonlocal pressure Hessian, the PIML model is reformatted into a combination of functions of invariants and integrity bases. This functional form falls naturally into the TBNN architecture with physically interpretable parameterization. After training the model with DNS data, we arrived at an improved representation of the magnitude and orientation of pressure Hessian eigenvales and eigenvectors. In contrast to physics-blind ML model, we can interpret such improvement by analyzing the statistical properties of scalar coefficients for different tensor bases. A stochastic ODE is constructed to model the full evolution of VGT with a simplified constant covariance matrix for the stochastic term. An *a posteriori* test is then conducted by temporally evolving the SODE from a random Gaussian initial field. The classical tear-drop shape of the *Q-R* joint distribution is reproduced, but the tail is more elongated compared to the DNS data. Further diagnostics show that the non-Gaussian statistics of the pressure Hessian cannot be fully represented by the PIML model.

For the coarse-grained VGT dynamics, a similar approach is adopted using the extended TBNN framework, which expands the integrity bases for skew-symmetric tensor prediction. The PIML model is trained using filtered DNS data with different filter sizes. With increasing filter size,

the dynamics become more Gaussian-like, which makes stochastic modeling more reasonable. However, the performance of the model, i.e., orientation and magnitude of the eigenvectors and eigenvalues of the coarse-grained VGT, is worse compared to the unfiltered case. This is likely because the current extended TBNN is not able to embed length-scale information in the model.

We also acknowledge that our current PIML model simplifies the spatial complexity due to the single-point modeling strategy. We recognize that better spatial modeling should be Lagrangian, injecting into consideration the dynamics of the coarse-grained volume.

To correct the problems with the VGT-only reduced modeling encountered in this paper, we have started to work in Ref. [53] on extending it. In particular, and acting consistently with the tetrad model approach developed in Ref. [24], we work on including into Lagrangian description geometrical characteristics, such as the shape of the coarse-graining fluid blob coevolving with the VGT. When training on high-Reynolds number DNS data, we expect the improved model to learn the universal inertial-range Lagrangian dynamics. The small-scale VGT dynamics with a high dependence on Reynolds number can then be modeled using the interscale interaction and energy transfer. We also work on improving and tuning the ML approaches. Thus, we have developed a Bayesian learning approach towards constructing SODE by learning (and not postulating as above) the covariance matrix (possibly dependent on both the coarse-graining shape and scale and on the coarse-grained VGT). We also envision exploring in the future ways to provide a better representation for the high-order/non-Gaussian statistics of turbulence. Injecting non-Wiener/non-Langevin noise and working with a richer family of loss functions (used to describe mismatch between DNS and empirical results) may help to tackle the problem of better representation of the turbulence and the inherited intermittency.

[1] C. Meneveau, Lagrangian dynamics and models of the velocity gradient tensor in turbulent flows, Annu. Rev. Fluid Mech. **43**, 219 (2011).

[2] G. L. Eyink, Multi-scale gradient expansion of the turbulent stress tensor, J. Fluid Mech. **549**, 159 (2006).

[3] P. L. Johnson and C. Meneveau, Turbulence intermittency in a multiple-time-scale Navier-Stokes-based reduced model, Phys. Rev. Fluids **2**, 072601(R) (2017).

[4] N. A. K. Doan, N. Swaminathan, P. A. Davidson, and M. Tanahashi, Scale locality of the energy cascade using real space quantities, Phys. Rev. Fluids **3**, 084601 (2018).

[5] M. Carbone and A. D. Bragg, Is vortex stretching the main cause of the turbulent energy cascade? J. Fluid Mech. **883**, R2 (2020).

[6] P. L. Johnson, Energy Transfer from Large to Small Scales in Turbulence by Multiscale Nonlinear Strain and Vorticity Interactions, Phys. Rev. Lett. **124**, 104501 (2020).

[7] A. E. Perry and M. S. Chong, A description of eddying motions and flow patterns using critical-point concepts, Annu. Rev. Fluid Mech. **19**, 125 (1987).

[8] M. S. Chong, A. E. Perry, and B. J. Cantwell, A general classification of three-dimensional flow fields, Phys. Fluids A: Fluid Dyn. **2**, 765 (1990).

[9] M. S. Chong, J. Soria, A. E. Perry, J. Chacin, B. J. Cantwell, and Y. Na, Turbulence structures of wall-bounded shear flows found using DNS data, J. Fluid Mech. **357**, 225 (1998).

[10] J. Martin, A. Ooi, M. S. Chong, and J. Soria, Dynamics of the velocity gradient tensor invariants in isotropic turbulence, Phys. Fluids **10**, 2336 (1998).

[11] A. Ooi, J. Martin, J. Soria, and M. S. Chong, A study of the evolution and characteristics of the invariants of the velocity-gradient tensor in isotropic turbulence, J. Fluid Mech. **381**, 141 (1999).

[12] L. Wang and X. Y. Lu, Flow topology in compressible turbulent boundary layer, J. Fluid Mech. **703**, 255 (2012).

[13] Y. B. Chu and X. Y. Lu, Topological evolution in compressible turbulent boundary layers, J. Fluid Mech. **733**, 414 (2013).

[14] P. Bechlars and R. D. Sandberg, Evolution of the velocity gradient tensor invariant dynamics in a turbulent boundary layer, J. Fluid Mech. **815**, 223 (2017).

[15] Y. Tian, F. A. Jaberi, and D. Livescu, Density effects on the post-shock turbulence structure and dynamics, J. Fluid Mech. **880**, 935 (2019).

[16] R. Das and S. S. Girimaji, Characterization of velocity-gradient dynamics in incompressible turbulence using local streamline geometry, J. Fluid Mech. **895**, A5 (2020).

[17] J. Tom, M. Carbone, and A. D. Bragg, Exploring the turbulent velocity gradients at different scales from the perspective of the strain-rate eigenframe, J. Fluid Mech. **910**, A24 (2021).

[18] M. Carbone, M. Iovieno, and A. Bragg, Symmetry transformation and dimensionality reduction of the anisotropic pressure Hessian, J. Fluid Mech. **900**, A38 (2020).

[19] P. Vieillefosse, Local interaction between vorticity and shear in a perfect incompressible fluid, J. Phys. **43**, 837 (1982).

[20] P. Vieillefosse, Internal motion of a small element of fluid in an inviscid flow, Phys. A **125**, 150 (1984).

[21] B. J. Cantwell, Exact solution of a restricted euler equation for the velocity gradient tensor, Phys. Fluids A: Fluid Dyn. **4**, 782 (1992).

[22] J. Martin, C. Dopazo, and L. Valiño, Dynamics of velocity gradient invariants in turbulence: Restricted Euler and linear diffusion models, Phys. Fluids **10**, 2012 (1998).

[23] S. S. Girimaji and S. B. Pope, A diffusion model for velocity gradients in turbulence, Phys. Fluids A: Fluid Dyn. **2**, 242 (1990).

[24] M. Chertkov, A. Pumir, and B. I. Shraiman, Lagrangian tetrad dynamics and the phenomenology of turbulence, Phys. Fluids **11**, 2394 (1999).

[25] E. Jeong and S. S. Girimaji, Velocity-gradient dynamics in turbulence: Effect of viscosity and forcing, Theor. Comput. Fluid Dyn. **16**, 421 (2003).

[26] L. Biferale, L. Chevillard, C. Meneveau, and F. Toschi, Multiscale Model of Gradient Evolution in Turbulent Flows, Phys. Rev. Lett. **98**, 214501 (2007).

[27] L. Chevillard and C. Meneveau, Lagrangian Dynamics and Statistical Geometric Structure of Turbulence, Phys. Rev. Lett. **97**, 174501 (2006).

[28] L. Chevillard, C. Meneveau, L. Biferale, and F. Toschi, Modeling the pressure Hessian and viscous Laplacian in turbulence: Comparisons with direct numerical simulation and implications on velocity gradient dynamics, Phys. Fluids **20**, 101504 (2008).

[29] M. Wilczek and C. Meneveau, Pressure Hessian and viscous contributions to velocity gradient statistics based on Gaussian random fields, J. Fluid Mech. **756**, 191 (2014).

[30] P. L. Johnson and C. Meneveau, A closure for Lagrangian velocity gradient evolution in turbulence using recent-deformation mapping of initially Gaussian fields, J. Fluid Mech. **804**, 387 (2016).

[31] J. M. Lawson and J. R. Dawson, On velocity gradient dynamics and turbulent structure, J. Fluid Mech. **780**, 60 (2015).

[32] L. A. Leppin and M. Wilczek, Capturing Velocity Gradients and Particle Rotation Rates in Turbulence, Phys. Rev Lett. **125**, 224501 (2020).

[33] R. Betchov, An inequality concerning the production of vorticity in isotropic turbulence, J. Fluid Mech. **1**, 497 (1956).

[34] R. M. Pereira, L. Moriconi, and L. Chevillard, A multifractal model for the velocity gradient dynamics in turbulent flows, J. Fluid Mech. **839**, 430 (2018).

[35] J. Ling, A. Kurzawski, and J. Templeton, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, J. Fluid Mech. **807**, 155 (2016).

[36] R. Maulik, O. San, A. Rasheed, and P. Vedula, Subgrid modelling for two-dimensional turbulence using neural networks, J. Fluid Mech. **858**, 122 (2019).

[37] K. Duraisamy, G. Iaccarino, and H. Xiao, Turbulence modeling in the age of data, Annu. Rev. Fluid Mech. **51**, 357 (2019).

[38] R. King, O. Hennigh, A. Mohan, and M. Chertkov, From deep to physics-informed learning of turbulence: Diagnostics, arXiv:1810.07785.

[39] J.-X. Wang, J.-L. Wu, and H. Xiao, Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data, Phys. Rev. Fluids **2**, 034603 (2017).

[40] A. Mohan, D. Daniel, M. Chertkov, and D. Livescu, Compressed convolutional LSTM: An efficient deep learning framework to model high fidelity 3D turbulence, arXiv:1903.00033.

[41] K. Fukami, K. Fukagata, and K. Taira, Super-resolution reconstruction of turbulent flows with machine learning, J. Fluid. Mech. **870**, 106 (2019).

[42] G. D. Portwood, P. P. Mitra, M. D. Ribeiro, T. M. Nguyen, B. T. Nadiga, J. A. Saenz, M. Chertkov, A. Garg, A. Anandkumar, A. Dengel *et al.*, Turbulence forecasting via neural ode, arXiv:1911.05180.

[43] K. Ohkitani and S. Kishiba, Nonlocal nature of vortex stretching in an inviscid fluid, Phys. Fluids **7**, 411 (1995).

[44] S. B. Pope, A more general effective-viscosity hypothesis, J. Fluid Mech. **72**, 331 (1975).

[45] Q.-S. Zheng, On the representations for isotropic vector-valued, symmetric tensor-valued and skew-symmetric tensor-valued functions, Int. J. Eng. Sci. **31**, 1013 (1993).

[46] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Netw. **4**, 251 (1991).

[47] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[48] M. Petersen and D. Livescu, Forcing for statistically stationary compressible isotropic turbulence, Phys. Fluids **22**, 116101 (2010).

[49] D. Daniel, D. Livescu, and J. Ryu, Reaction analogy based forcing for incompressible scalar turbulence, Phys. Rev. Fluids **3**, 094602 (2018).

[50] P. K. Yeung and S. B. Pope, An algorithm for tracking fluid particles in numerical simulations of homogeneous turbulence, J. Comput. Phys. **79**, 373 (1988).

[51] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.

[52] F. Chollet *et al.*, Keras, https://keras.io (2015).

[53] C. Hyett, M. Chertkov, Y. Tian, and D. Livescu, Machine learning statistical Lagrangian geometry of turbulence, in *APS Division of Fluid Dynamics Meeting Abstracts*, pp. S01–024 (2020).