Velocity gradient prediction using parameterized Lagrangian deformation models

Criston Hyett, Yifeng Tian, Michael Woodward, Misha Stepanov, Chris Fryer, Michael Chertkov, Daniel Livescu

August 14, 2023

Abstract

We seek to efficiently predict the statistical evolution of the velocity gradient tensor (VGT) by creating local models for the pressure Hessian. Previous work has identified physics-informed machine learning (PIML) to be adept in this prediction; of note in this class of models is the Tensor Basis Neural Network (TBNN) for its embedded physical constraints and demonstrated performance. Simultaneously, phenomenological models were advanced by approximating the local closure to the pressure Hessian via deformation models using the history of the VGT. The latest in this series of models is the Recent Deformation of Gaussian Fields (RDGF) model. In this work, we combine the (local in time) PIML approach with the phenomenological idea of inclusion of recent deformation to create a data-driven Lagrangian deformation model. We compare the model performance to both the TBNN and the RDGF models, and provide data-driven hypotheses regarding the Lagrangian upstream assumptions made in the RDGF model.

1 Introduction

The velocity gradient tensor (VGT) describes many important aspects of turbulence. It displays characteristic non-Gaussian statistics including intermittency, describes the deformation rate of a fluid volume, and encapsulates alignment between strain and vorticity [**?**]. The VGT has been the subject of much study and with the exponential growth of computation, direct numerical simulations that resolve the smallest scales of turbulence have provided enormous VGT datasets. Even more recently, machine learning (ML), particularly physics-informed machine learning (PIML), have shown remarkable ability to glean predictive capability from these large datasets, suggesting patterns exist that we have not yet recognized.

Of particular note in recent years is the Recent Deformation of Gaussian Fields model (RDGF) [**?**], wherein the authors postulate a closure of the VGT evolution equations via choosing a Gaussian upstream condition that is deformed according to the current VGT. This hypothesis is in the same spirit as the Tetrad model from Chertkov et.al [**?**] that proposed a local closure to the VGT equations, by closing the equations for the VGT using deformation of a small fluid element.

In the realm of machine learning, this work is inspired by Tian et.al[**?**], that used a highly structured, so-called Tensor Basis Neural Network (TBNN) architecture to close the equations. This network structure was inspired by theoretical work by [**?**], [**?**], [**?**]. It has been shown to explicitly respect Galilean and rotational invariance, enforce incompressibility, and after training, can predict many quantities of interest in the statistics of the VGT.

Our goals in this paper are to advance the phenomenology via data analysis, and leverage this to improve the TBNN methodology to better predict the statistical evolution of the VGT.

2 Previous Work

As isotropic turbulence is an inherently statistical process, and models of the pressure Hessian

$$P = \frac{\partial^2 p}{\partial x_i \partial x_j} \tag{1}$$

are attempting to locally close nonlocal equations, it is useful to decompose the prediction into mean and fluctuating parts

$$P = \bar{P} + P' \tag{2}$$

where $\bar{P}$ is the mean, and P' represents the fluctations about this mean. When discussing the models below, the focus is on the (deterministic) modeling techniques used for $\bar{P}$. Accurate representation of the distribution of P' is important, but we first seek to glean as much predictive capability from $\bar{P}$, and only after will enrich the models with models of P'. To simplify notation then, we will drop the overbar in any predictions, and instead write $\hat{P}$ to denote an approximation of the mean.

Throughout, we denote vectors with lower-case Latin letters, 2nd-rank tensors with upper-case Latin letters, and constants using lower-case Greek alphabet. The two exceptions to this rule are the second and third scalar invariants of the velocity gradient tensor, denoted by Q, R respectively, which we maintain in congruence with the literature.

We follow the Einstein summation notation, that repeated indices imply summation, e.g. $A_{ii} = \mathrm{tr}(A)$, $u_k \frac{\partial u_i}{\partial x_k} = u \cdot \nabla u$, $\frac{\partial^2 u_i}{\partial x_k \partial x_k} = \nabla^2 u$, etc.

For all tensors in the paper, indices will be denoted by one or more of $i, j, k, l, m, n \in [1, 2, 3]$.

2.1 Governing Equations for VGT

Navier-Stokes defines the evolution of a velocity field $u(x, t)$, given a pressure field $p(x, t)$

$$\frac{\partial u_i}{\partial t} + u_k \frac{\partial u_i}{\partial x_k} = -\frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_k \partial x_k} \tag{3}$$

The VGT is defined by $A_{ij} = \frac{\partial u_i}{\partial x_j}$, so we apply spatial derivatives to eq(3), and use the definition of material derivative to obtain an ODE for the velocity gradient tensor (defined in the Lagrangian frame)

$$\frac{dA_{ij}}{dt} = \frac{\partial A_{ij}}{\partial t} + u_k \frac{\partial A_{ij}}{\partial x_k} = -A_{ik} A_{kj} - \frac{\partial^2 p}{\partial x_i \partial x_j} + \nu \frac{\partial^2 A_{ij}}{\partial x_k \partial x_k} \tag{4}$$

Using the incompressibility condition,

$$\frac{\partial u_i}{\partial x_i} = 0 \implies A_{ii} = 0 \tag{5}$$

we can take the trace of eq(4) to find

$$\frac{\partial^2 p}{\partial x_k \partial x_k} = -A_{ij}A_{ji} \tag{6}$$

which determines the trace of the pressure Hessian. Thus, we can define the so-called deviatoric (or nonlocal) pressure Hessian as

$$H_{ij} := -\left(\frac{\partial^2 p}{\partial x_i \partial x_j} - \frac{1}{3}\frac{\partial p}{\partial x_k \partial x_k}\delta_{ij}\right) \tag{7}$$

and denote the purely local contributions as the so-called "Restricted Euler"

$$E_{ij} := -\left(A_{ik}A_{kj} + \frac{1}{3}A_{mn}A_{nm}\delta_{ij}\right) \tag{8}$$

Finally, letting the viscous term be denoted by

$$T_{ij} := \nu \frac{\partial^2 A_{ij}}{\partial x_k \partial x_k} \tag{9}$$

We can write the ODE for the Lagrangian VGT as

$$\frac{dA_{ij}}{dt} = E_{ij} + H_{ij} + T_{ij} \tag{10}$$

2.2 Pressure Hessian

In the first attempt to locally close the VGT equations, Viellefosse [**?**] and Cantwell [**?**] studied the "Restricted Euler" dynamics:

$$\frac{dA_{ij}}{dt} \approx E_{ij} \tag{11}$$

and found the resulting system had a finite-time singularity.

Further studies confirm that the main inhibitor to this unphysical singularity is the deviatoric pressure Hessian. As suggested by fig(1), while the viscous term is dynamically important to damp turbulent fluctuations, the main modeling challenge is the deviatoric pressure Hessian.

In the case of incompressible turbulence we can write the formal (nonlocal) solution for the deviatoric pressure Hessian as a spatial integral [**?**] in the Eularian frame

$$H_{ij}(\boldsymbol{x}) = \iiint \frac{\delta_{ij} - \hat{r_i}\hat{r_j}}{2\pi r^3} Q(\boldsymbol{x} + \boldsymbol{r}) d\boldsymbol{r} \tag{12}$$

where $Q = \frac{1}{2}\mathrm{tr}(A^2)$, and $\boldsymbol{x} \in \mathbb{R}^3$.

The challenge then, is to obtain an approximation of this integral using only local information.

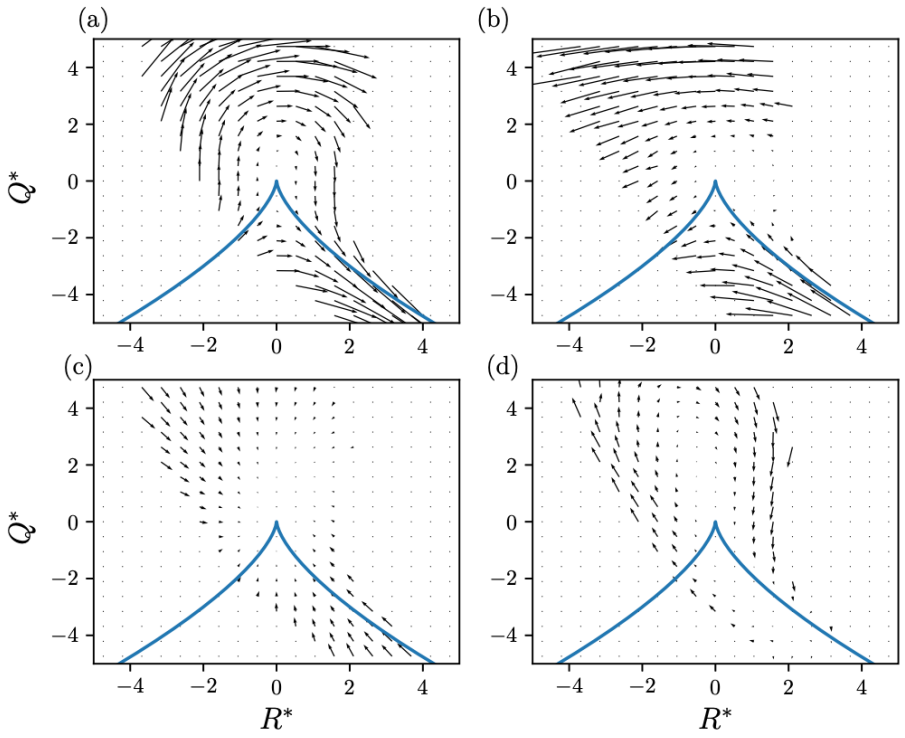

Figure 1: The ground truth conditional mean tangents (CMTs) in the Q-R phase plane arising from (a) Restricted Euler (b) deviatoric pressure Hessian (c) viscous (d) sum of all terms. Taken from Tian et.al, TODO generate this figure.

2.3 Tensor Basis Neural Network

Following Lawson & Dawson [?], an expansion of the integral in eq(12) is proposed, first as a Taylor series expansion in $Q(\boldsymbol{x}+\boldsymbol{r})$

$$\hat{H} = \sum_{m,n=0}^{\infty} \alpha_{mn} S^m W^n \tag{13}$$

where

$$S = \frac{1}{2}(A + A^T) \qquad W = \frac{1}{2}(A - A^T) \tag{14}$$

Finally we can reduce from an infinite sum using Cayley-Hamilton, and expand via the tensor basis[?],[?]

$$\hat{H} = \sum_{n=1}^{10} g^{(n)}(\lambda_1, \ldots, \lambda_5) T^{(n)} \tag{15}$$

with $g^{(n)}$ scalar functions of the invariants

$$\lambda_1 = \mathrm{tr}(S^2) \quad \lambda_2 = \mathrm{tr}(W^2) \quad \lambda_3 = \mathrm{tr}(S^3) \quad \lambda_4 = \mathrm{tr}(W^2 S) \quad \lambda_5 = \mathrm{tr}(W^2 S^2) \tag{16}$$

and the tensor basis given by:

$$T^{(1)} = S \qquad T^{(2)} = SW - WS \tag{17}$$

$$T^{(3)} = S^2 - \frac{1}{3} I \cdot \mathrm{tr}(S^2) \qquad T^{(4)} = W^2 - \frac{1}{3} I \cdot \mathrm{tr}(W^2) \tag{18}$$

$$T^{(5)} = WS^2 - S^2 W \qquad T^{(6)} = W^2 S + SW^2 - \frac{2}{3} I \cdot \mathrm{tr}(SW^2) \tag{19}$$

$$T^{(7)} = WSW^2 - W^2 SW \qquad T^{(8)} = SWS^2 - S^2 WS \tag{20}$$

$$T^{(9)} = W^2 S^2 + S^2 W^2 - \frac{2}{3} I \cdot \mathrm{tr}(S^2 W^2) \qquad T^{(10)} = WS^2 W^2 - W^2 S^2 W \tag{21}$$

This formulation reduces the challenge to finding the functions of known scalars, i.e., learning the $g^{(n)}$'s as shown in fig(2).

This network architecture ensures the output tensor $\hat{H}$ is symmetric, traceless, and the network itself is rotationally and Galilean invariant. A technical note, since we are using the L_2 loss, $\hat{H}$ is only an approximation to the mean deviatoric pressure Hessian if the statistics of the pressure Hessian are normal. As we expect turbulent statistics to be non-Gaussian, the choice of loss is driven more by mathematical convenience.

Tian et.al [?] showed ability to train the network, and demonstrated state-of-the-art performance on a variety of relevant physical metrics: eigenvector alignments, Q-R conditional mean tangents (CMTs), and *a posteriori* tests evolving an initially Gaussian field to fully developed turbulence as evidenced by characteristic teardrop-shaped Q-R probability distributions (PDFs).

2.4 Recent Deformation of Gaussian Fields

Johnson and Meneveau [?] introduced the RDGF model following a long line of postulating that the deviatoric part of the pressure Hessian could be locally modeled using information of the time history of the deformation tensor [?],[?],[?].

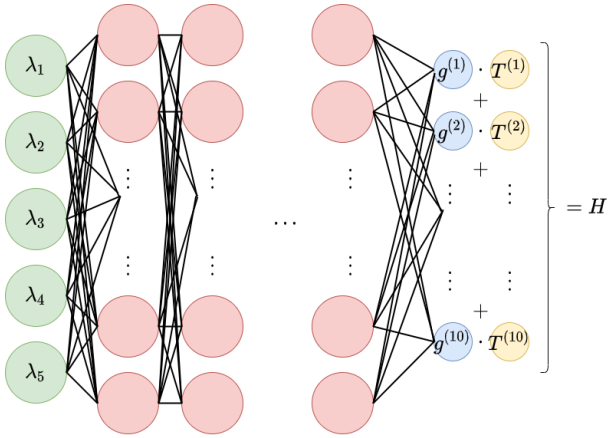

Figure 2: The architecture of the TBNN. The invariants and tensor basis elements are calculated from each sample of the VGT, the invariants are then used as input to a fully connected network, and the resulting $g^{(i)}$ multiply the corresponding tensor basis elements $T^{(i)}$, before summing into the prediction $\hat{H}$

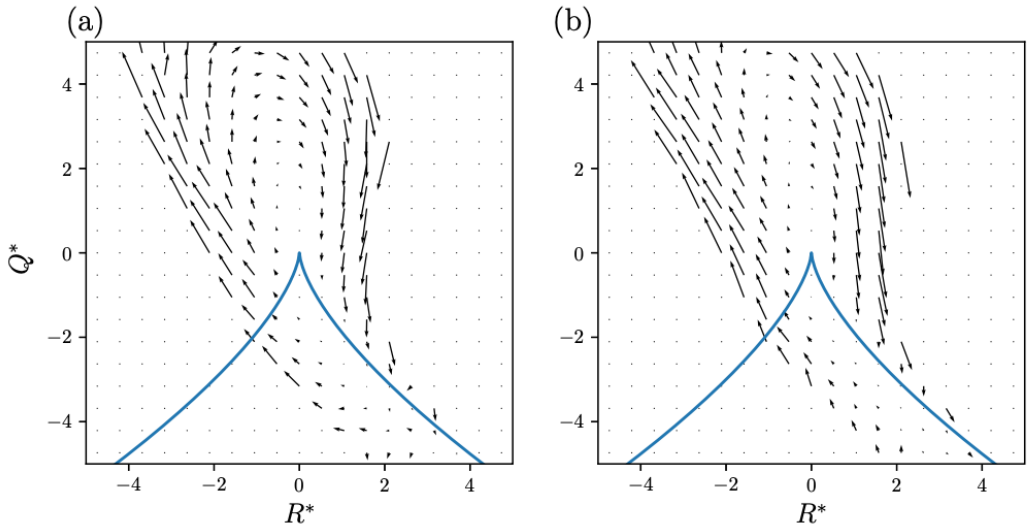

Figure 3: Results from Tian et.al[**?**] showing Q-R CMTs from (a) DNS data and (b) trained TBNN

Chertkov et.al's Tetrad Model [**?**], attempted to capture this deformation directly by following a small fluid element approximated by four Lagrangian particles, initialized in an isotropic configuration. This model avoided the finite time singularity resulting from using only the restricted Euler term[**?**]. Attempts to fuse this low-dimensional model of a fluid element with data-driven methods for prediction of the coarse-grained VGT are ongoing [**?**],[**?**]. Chevillard & Meneveau used the same isotropic Lagrangian upstream condition as in the Tetrad Model when developing their Recent Fluid Deformation (RFD) model. Most recently, Johnson & Meneveau enriched the approximation of the upstream condition to that of a Gaussian field using high fidelity direct numerical simulation (DNS) data[**?**].

RDGF models the conditional average of the (full) pressure Hessian tensor as:

$$\hat{P}_{ij} = \frac{\partial y_k}{\partial x_i}\frac{\partial^2 p}{\partial y_k \partial y_l}\frac{\partial y_l}{\partial x_j} = D^{-1}_{ki}\tilde{P}_{kl}D^{-1}_{lj}, \tag{22}$$

where $\tilde{P}$ is the pressure Hessian tensor/matrix evaluated along the Lagrangian trajectory upstream in time. The deformation tensor, defined as a sensitivity of a particle observed at the position x at the moment of time t to its initial position y at the previous moment of time,

$$D_{ij} = \frac{\partial x_i}{\partial y_j}, \tag{23}$$

evolves according to

$$\frac{dD_{ij}}{dt} = A_{ik}D_{kj}, \qquad \text{with } D_{ij}(0) = \delta_{ij}, \tag{24}$$

assuming isotropic initial condition (without loss of generality for sufficiently far removed upstream times). Solution of Eq. (24) is defined via the time-ordered exponential

$$D(t) = \mathrm{Texp}\left(\int_0^t dt' A(t')\right) = \lim_{N\to\infty}\prod_{i=0}^{N}\left(e^{A(t_i)\Delta t}\right), \tag{25}$$

where $t_i = i\Delta t$ and $\Delta t = t/N$. If we take $N = 1$ and set Δt to be the smallest temporal scale of turbulence (also called the Kolmogorov, or viscous scale) we arrive at the RDGF approximation for the deformation tensor

$$\hat{D}(x,t) = \exp\left(A(x,t)\Delta t\right) \approx D(\Delta t). \tag{26}$$

All that is left is to determine the upstream pressure Hessian $\tilde{P}$. The RDGF model uses a Gaussian field closure for $\tilde{P}$,

$$\hat{\tilde{P}}_{ij} = \frac{1}{3}\tilde{P}_{kk}\delta_{ij} + \langle\tilde{P}^{(d)}_{ij}|A\rangle_{\text{Gaussian}} \tag{27}$$

where $\langle\tilde{P}^{(d)}|A\rangle_{Gaussian}$ is the mean deviatoric pressure Hessian in a Gaussian field, found by Wilczek & Meneveau[**?**]. Using the notation of the tensor basis [Eqs. (17-21)], they found

$$\langle\tilde{P}^{(d)}_{ij}|A\rangle_{\text{Gaussian}} = \gamma T^{(2)} + \alpha T^{(3)} + \beta T^{(4)} \tag{28}$$

with the value of the coefficients found from a mix of theoretical and numerical experiments

$$\alpha = -\frac{2}{7}, \quad \beta = -\frac{2}{5}, \quad \gamma \approx 0.08. \tag{29}$$

Introducing the deformed Gaussian field approximation

$$\hat{G}_{ij} := \hat{D}^{-1}_{mi} \langle \tilde{P}^{(d)}_{mn} | A \rangle_{\text{Gaussian}} \hat{D}^{-1}_{nj}, \tag{30}$$

and a short-time approximation to the inverse of the left Cauchy-Green tensor

$$\hat{C}^{-1}_{ij} = \hat{D}^{-1}_{ki} \hat{D}^{-1}_{kj} \tag{31}$$

We enforce the prediction in Eq. (27) to have the correct trace, and arrive at the following RDGF prediction for the component of the pressure Hessian conditioned to the VGT

$$\hat{P}_{ij} = 2Q \frac{\hat{C}^{-1}_{ij}}{\hat{C}^{-1}_{kk}} + \hat{G}_{ij} - \frac{\hat{C}^{-1}_{ij}}{\hat{C}^{-1}_{kk}} \hat{G}_{ll}. \tag{32}$$

3 Methodology

This section is a work-in-progress. I talk more about this in the prospectus, and welcome comments and ideas.

3.1 Proposed Models

Our approach is to utilize the idea of biasing the statistics of the conditional deviatoric pressure Hessian using the history of the VGT (as in the RDGF, RFD, and Tetrad models), while avoiding postulating the exact functional form; instead allowing it to be data-driven using PIML (as in TBNN model).

We do this by augmenting the inputs to the feed-forward portion of the TBNN, the idea being to bias the statistics of the invariants, while believing that the tensors derived from the local-in-time VGT are sufficient to span the function space.

3.1.1 Temporal Convolution

The simplest of our proposed models augments the set of invariants using a learned temporal convolution

$$\hat{H} = \sum_{n=1}^{10} g^{(n)}_{\theta}(\lambda_1, \ldots, \lambda_5, C) T^{(n)} \tag{33}$$

with

$$C(t) = \sum_{t_f}^{t_f - \Delta t \cdot n} \zeta_n A_n \tag{34}$$

where ζ_n are learned simultaneously with the parameterization of the scalar g functions.

This model avoids biasing the prediction with phenomenological theory, but at the cost of a computational overhead. However, because of the simple structure, the learned ζ's may inform a temporal correlation in the VGT that is useful in the prediction of the deviatoric pressure Hessian. An extreme case of very short time correlations would reinforce the hypothesis presented in RDGF, while longer temporal correlations would support the modeling approaches of e.g., [**?**].

3.1.2 Time-Ordered Exponential

Finally, we may take an intermediate step between the two and consider, as a hypothesis, replacing C in Eq. (33) by the time-ordered exponential

$$W(t) = \text{Texp}\left(\int_0^t dt' A(t')\right) \tag{35}$$

Misha says: Criston, I have attempted to clean notations ... X is already introduced as an initial position for a Lagrangian particle ... I am suggesting to use C for the temporal convolution and $W(t)$ for the time-ordered exponential.

3.2 Data Analysis

In this section we perform numerical studies on DNS data to evaluate assumptions and choices for model hyperparameters.

3.2.1 Ground Truth Data

Our data is DNS of forced isotropic turbulence with $Re = 240$. Eularian data is generated, random points are sampled, VGT and pressure Hessian are constructed at these points using interpolation, and Lagrangian trajectories of these quantities are advanced using the Eularian velocity fields. We have 122k samples, each advanced for 1000 timesteps with $\Delta t = 3e - 4$, representing approximately half an inertial eddy turnover time. For each sample and each timestep, we have measured VGT, PH, and viscous terms.

3.2.2 Cramer Functions of Strain Rate Eigenvalues

Coarsely stated, the Cramer function (also more generally called the rate function) determines how quickly large deviations from a mean diminish. If we believe in the ability to close the equations for the deviatoric pressure Hessian locally, it is apparent that there must exist a relation between temporal correlations of the strain-rate tensor, and temporal correlations of the deviatoric pressure Hessian.

3.2.3 Gaussian Upstream Condition

The Gaussian upstream condition present in RDGF is convenient, but for short deformation time is known to be inaccurate. Via analysis of Lagrangian data, we can evaluate how long the deformation time must be to transform the Gaussian upstream condition to a distribution representative of fully developed turbulence. In particular, we can evaluate

$$\min_{\Delta t} \left\| P_{gt}^{(d)} - P'^{(d)}_{\text{Gaussian}}(\Delta t) \right\| \tag{36}$$

where $P'^{(d)}_{\text{Gaussian}}(\Delta t)$ is the true deformation of the Gaussian upstream condition, by the ground truth VGT over a time Δt.

4 Results

4.1 Hyperparameter Tuning

One important deviation from the guiding paper is in data normalization. Tian et.al, normalized the VGT and PH using an empirical measure of the timescale

$$\tau \approx \langle \left\| S^2 \right\| \rangle^{-1} \tag{37}$$

In our work, we confirmed it was vital to performance to normalize the VGT

$$A' = \tau A \tag{38}$$

but found the network performance suffered considerably if the PH was normalized. One hypothesis is that this is an artifact of initialization of the network parameters. The PH should be normalized by τ^2, which for $Re = 240$ $\tau \approx 3e-3$. If one uses this normalization, the network does train, but it seems to start very far away from the optimal. It is possible this could remedied by scaling the initialization of the linear output layer as well as individualizing that layer's learning rate - this was not performed.

In calculation of the $g^{(i)}$ functions, we use a fully connected feed-forward network with 5 hidden layers of 50 nodes apiece with relu activation functions and a linear output layer. This network is trained using the ADAM optimizer with learning rate $5e-3$ decaying to a minimum of $1e-6$ via factors of 2 when the loss reaches a plateau.

4.2 Result reproduction

After implementing this network architecture and finding the hyperparameters above, the results in figure(4) were found. These results agree well with the original paper, and suggest the hyperparameter choices were accurate.

4.3 Temporal Convolution TBNN

Preliminary results suggest that the additional historical information of the VGT is in fact useful for the prediction of the PH. In figure(5), results of training the temporal convolution TBNN (convTBNN) are shown, improving both in the loss and eigenvector alignment metrics.

An interesting aspect of the convTBNN is the ability to inspect the kernel length - loosely the length over which the network found it useful to set weights away from zero. Figure(6) displays the weights as a function of kernel entry number. The sampling in time occurs every $20\Delta t$, where $\Delta t = 3e-4$ and the Kolmogorov timescale is $\tau \approx 3e-3$, so that the sampling occurs about every 2τ. We see that only the last three weights are nonzero, suggesting that a history of about 6τ is useful in prediction.

5 Conclusion

In this work (in progress), we reproduced the current state-of-the-art Lagrangian model to predict the PH, contextualized it with recent phenomenological advancements, and advanced the methodology by introducing and evaluating a new predictive model.

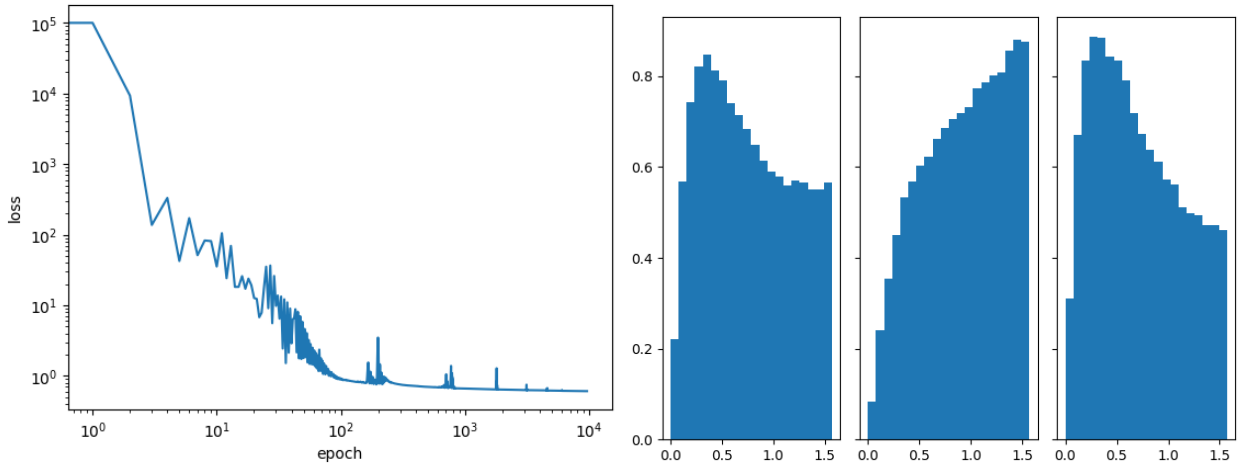

Figure 4: (left) Loss as a function of epoch for the TBNN, (right) PDFs of eigenvalue alignment for the pressure hessian, sorted from smallest to largest eigenvalue left to right. These plots agree well with results from Tian et.al's paper, suggesting the implementation is correct.

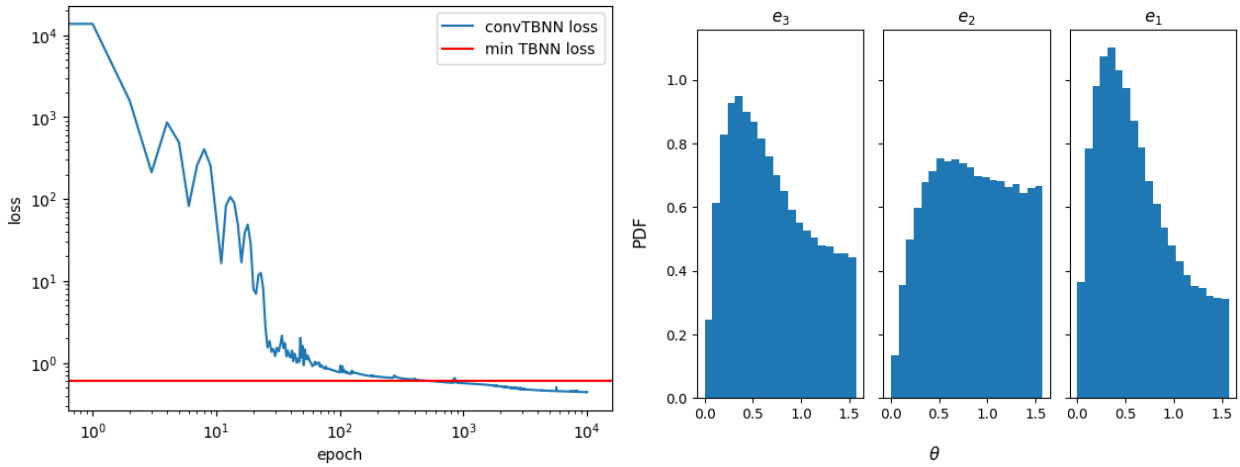

Figure 5: Preliminary results of training the convTBNN. (Left) shows the loss vs epoch, with the minimum loss obtained by the unmodified TBNN marked in horizontal red, and the convTBNN in blue. (Right) shows the alignment PDFs of eigenvectors of the PH. Here e_1 corresponds to the eigenvector with the greatest associated eigenvalue. Note that compared to fig(4), all alignment PDFs are better, shifting significant density from the misaligned (larger θ) to aligned.

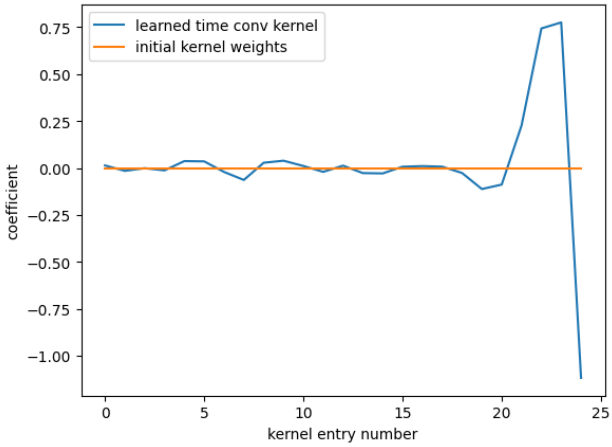

Figure 6: The weights of the learned temporal convolution kernel. This may act as a stand-in for a measure of significance for VGT history as it relates to the prediction of the PH. If these results hold across many trainings, it would suggest that only "recent" history is informative in prediction of the PH.

The TBNN is a powerful tool to capture symmetries, while allowing for the flexibility of a NN. We applied this model to homogeneous, isotropic turbulence to close the equations of the non-local pressure Hessian using only the local velocity gradient tensor. We contrasted this approach with a mathematically similar, but philosophically different approach using the RDGF model, motivated by the idea that the recent history of the VGT could better inform the prediction of the current PH.

By combining the two, we created the convTBNN, a first data-driven step towards biasing the statistics of the current PH prediction using the history of the VGT. We showed that we could outperform the local in time TBNN in two metrics, loss value and eigenvector alignment. We began to interpret the weights of the temporal convolution kernel, and related them to the hypotheses made in the phenomenological models (RFD, RDGF).

Further work needs to be done to ensure these results hold under evaluation of richer metrics such as the Q-R conditional mean tangents, and evaluate the accuracy and stability of the resulting differential equation for the VGT.

6 Appendix

6.1 Interpretability

Previous work by the authors [**?**] suggested there may be a latent space in the invariants when predicting the PH using the TBNN. The existence of a latent space could be important to the ability to interpret the results, e.g. searching for a functional form or providing physical insight; or if we believe in the statistical independence of the input (in our case there are many reasons to) it suggests the network struggles to exploit the additional information. In the case of [**?**] it seems to be the latter.

The proposed latent space is indeed persistent when training the network having normalized the PH using τ^2, in particular, the latent space claimed consists only of

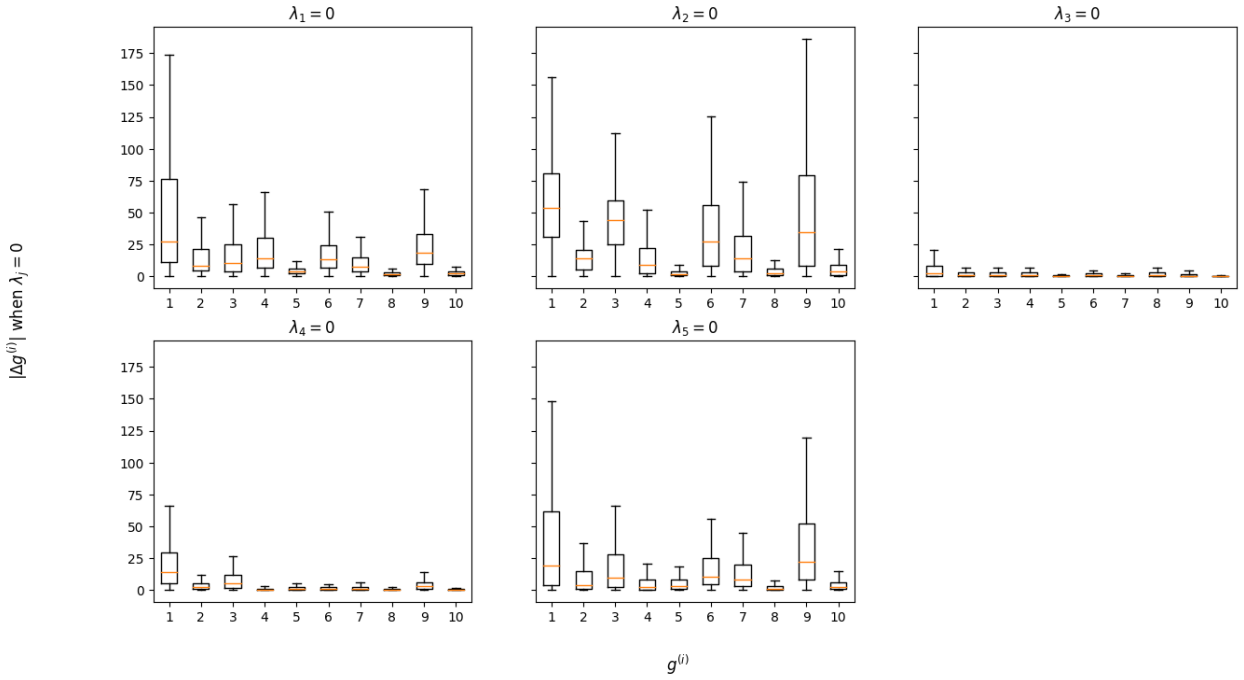

Figure 7: Distributions of sensitivity of $g^{(i)}$ to variability of λ_j. Previous work performed suggested that λ_{3-5} were unimportant to the prediction using the TBNN. This result was predicated upon the normalization of the PH using τ^2. While our plot reinforces that $\lambda_{3,4}$ do provide an order of magnitude smaller correction, the sensitivity of the network to λ_5 is of the same order of magnitude as that to $\lambda_{1,2}$. This suggests that by *not* normalizing the PH, we are able to exploit additional information in the invariants.

the low-order invariants. The hypothesis by the author was that the latent space emerged as a numerical artifact of strong normalization of the high-order invariants ($\lambda_{3,4} \to \tau^3\lambda_{3,4}$, and $\lambda_5 \to \tau^4\lambda_5$). But as we show in figure(7), there exists leading-order sensitivity to λ_5 (and not insignficant sensitivity to $\lambda_{3,4}$) when the PH is *not* normalized. This may indicate that the low-order latent space was a numerical artifact resulting from very small weights in the *output layer*, combined with strong normalization of high-order invariants.

To explore this question further, we applied an autoencoder to the set of invariants, as well as calculated the mutual information between the invariants.

The mutual information calculation, shown in table(1), indicates that while the invariants are not strictly independent, they are also far from highly correlated. The maximum value - biased away from the ideal value of 10 here by the k-nearest neighbors algorithm with $k = 5$, detailed in [**?**].

The autoencoder structure is a fully connected, feed forward neural network,

$$AE_\theta : \mathbb{R}^5 \to \mathbb{R}^5 \text{ via } \mathbb{R}^5 \to \mathbb{R}^4 \to \mathbb{R}^h \to \mathbb{R}^4 \to \mathbb{R}^5 \tag{39}$$

where h is the imposed latent dimension, illustrated in fig[8]. The nonlinear activation functions of AE_θ are the so-called "leaky-relu". We implement the "robust max-min" normalization, relying on the interquartile distance - namely, letting Q_1, Q_3 be the first

$I(\lambda_i, \lambda_j)$	λ_1	λ_2	λ_3	λ_4	λ_5
λ_1		0.26350727	1.40593169	0.40473788	0.72127464
λ_2			0.1968969	0.80102301	0.95679744
λ_3				0.43549834	0.49519878
λ_4					0.83990126
λ_5					

Table 1: A scaled mutual information $I(\lambda_i, \lambda_j)$, between invariants. Note the maximum value is near 10, the method used here introduces a bias however, namely via k-nearest neighbors with $k = 5$, as discussed in [**?**]

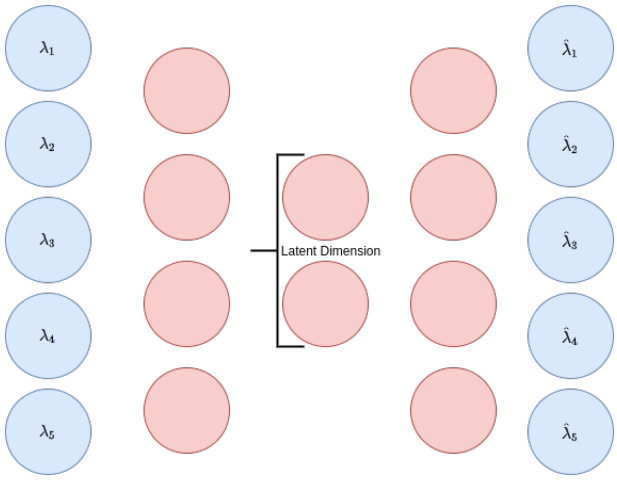

Figure 8: Autoencoder structure applied to the invariant input data.

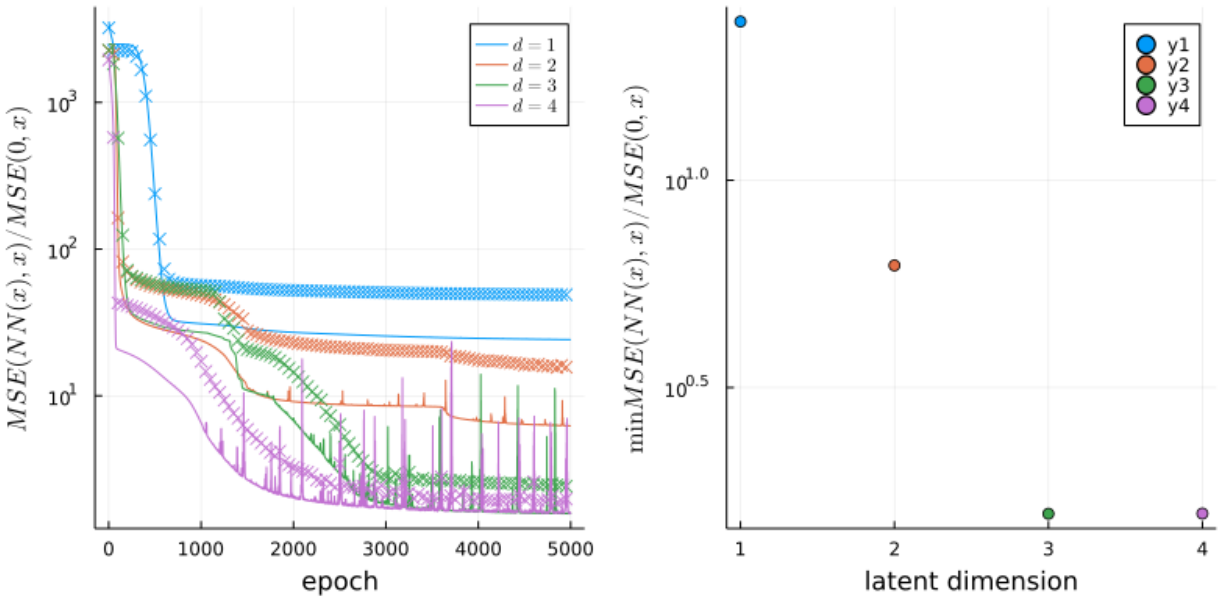

Figure 9: Autoencoder applied to the invariant input data, in an attempt for a nonlinear projection onto a latent space. (Left) shows the evolution of the loss according to the imposed latent dimension, where 'x's mark test loss, while solid lines mark training loss. (Right) shows the minimum loss as a function of latent dimension. Notice the two jumps at $h = 2$ and $h = 3$ - suggesting a sequence of increasing fidelity.

and third quartiles, we normalize our invariants as

$$\tilde{\lambda}_i = \frac{\lambda - Q_1}{Q_3 - Q_1} \tag{40}$$

This is intentionally different than the normalization used in the TBNN, as normalizing via timescale and optimizing using the mean-squared-error gives a very strong preference towards reconstructing only the low-order invariants. The robust max-min normalization is an attempt to set each on an equal footing, but itself is subject to biases introduced by differences in distribution shapes.

We use the ADAM optimizer with learning rate $\eta = 1e-3$, partitioning the dataset of 100k samples into 75% training, 25% test. The results are shown in fig[9] and may suggest a small reduction in dimension (i.e., $5 \to 3$). These results match the leading order contributions to the sensitivities of the $g^{(i)}$ as shown in fig(7), but we caution this suggestion as the low-order basis elements are still significantly influenced.