Physics-informed data-driven reconstruction of turbulent wall-bounded flows from planar measurements

Gurpreet S. Hora[1] (गुरप्रीत एस. होरा) , Pierre Gentine[2], Mostafa Momen[3] (مصطفى مومن), and Marco G. Giometto[1]†

[1]Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, USA

[2]Department of Earth and Environmental Engineering, Earth Institute, Columbia University, New York, NY 10027, USA

[3]Department of Civil and Environmental Engineering, University of Houston, Houston, TX, 77204, USA

Obtaining accurate and dense three-dimensional estimates of turbulent wall-bounded flows is notoriously challenging and this limitation negatively impacts geophysical and engineering applications such as weather forecasting, climate predictions, air quality monitoring, and flow control. This study introduces a physics-informed variational autoencoder model that reconstructs realizable three-dimensional turbulent velocity fields from two-dimensional planar measurements thereof. Physics knowledge is introduced as soft and hard constraints in the loss term and network architecture, respectively, to enhance model robustness and leverage inductive biases alongside observational ones. The performance of the proposed framework is examined in a turbulent open-channel flow application. The model excels in precisely reconstructing the dynamic flow patterns at any given time and location, including turbulent coherent structures, while also providing accurate time- and spatially-averaged flow statistics. The model outperforms state-of-the-art classical approaches for flow reconstruction such as the linear stochastic estimation method. By incorporating physical constraints, it can offer more accurate predictions of small-scale flow structures and maintain better consistency with the fundamental equations governing the system when compared to a purely data-driven approach.

Key words: Deep Learning, Variational Autoencoder, Direct Numerical Simulation, Open-Channel Flow

1. Introduction

Advancing our understanding and ability to model turbulent flows is critical for accurate weather forecasting (Skamarock *et al.* 2008) and climate projection (Mochida & Lun 2008; Toparlar *et al.* 2015), to improve urban sustainability and resilience (Chen *et al.* 2012; Casola 2019; Krayenhoff *et al.* 2020; Kameyama *et al.* 2020), and to design more performant and reliable engineering systems (Cheikh & Momen 2020; Chung *et al.* 2021). Historically, scientific discoveries in the field of turbulence have been achieved through computational methods (Scotti *et al.* 1993; Moser *et al.* 1999; Bou-Zeid *et al.* 2005; Chung & Pullin 2009; Lee & Moser 2015), experimental techniques (Champagne *et al.* 1967; Raupach *et al.* 1980; Gong & Ibbetson 1989; Prasad 2000; Elsinga *et al.* 2006; Elsinga & Marusic 2010;

† Email address for correspondence: mg3929@columbia.edu

Westerweel *et al.* 2013), and field observations (Menzies & Hardesty 1989; Gal-Chen *et al.* 1992; Rotach *et al.* 2005).

From laboratory and field observation perspectives, various sensing technologies, including hot-wire anemometry (Kuo & Corrsin 1971), particle image velocimetry (PIV) (Adrian 2005), three-dimensional (3-D) sonic anemometers (Foken & Wichura 1996), distributed temperature sensing technique (Thomas *et al.* 2012), thermal infrared cameras (Christen *et al.* 2012), and light detection and ranging (LiDAR) (Grund *et al.* 2001), have been employed to gain insight on turbulent flows. Despite their utility, these approaches are typically limited to point or planar measurements. For example, in-situ eddy covariance stations only provide a few discrete points in space due to the high cost of 3-D sonic anemometers and their involved deployment and maintenance procedures. Likewise, LiDAR and PIV measurements are generally limited to two-dimensional (2-D) planes. This sparse spatial coverage often proves inadequate for completely characterizing a turbulent flow.

Motivated by this limitation, several techniques have been proposed to reconstruct 3-D turbulent flows from sparse measurements (Van Doorne & Westerweel 2007; Ganapathisubramani *et al.* 2008; Vétel *et al.* 2010; Seo *et al.* 2016; Chandramouli *et al.* 2019; Bauweraerts & Meyers 2021). Traditional approaches have relied on the Taylor frozen-turbulence hypothesis (Lin *et al.* 2001; Van Doorne & Westerweel 2007), flow homogeneity assumptions (Chandramouli *et al.* 2019), intrinsic and extrinsic camera parameters (Pavlík *et al.* 2017), and techniques such as linear stochastic estimation (LSE) (Murray & Ukeiley 2003; Podvin *et al.* 2006). For instance, Ganapathisubramani *et al.* (2008) utilized a cinematographic stereoscopic PIV technique to collect time-resolved measurements. By applying Taylor's hypothesis, this study reconstructed a 3-D quasi-instantaneous velocity field for the turbulent jet. This allowed them to compute first- and second-order velocity statistics within a volume, providing valuable information about the statistical properties and flow behavior. However, a significant limitation of these traditional approaches is their reliance on assumptions that are not universal across flow systems and often require ad-hoc tuning, making it challenging to generalize across flow phenomena.

Reconstructing 3-D objects from 2-D images has been a long-standing and ill-posed inverse problem in computer vision (Han *et al.* 2019). Recent deep learning (DL)-based generative models have shown promising progress in reconstructing 3-D scenes of an object from a single or multiple images (Choy *et al.* 2016; Tulsiani *et al.* 2017; Henzler *et al.* 2018; Biffi *et al.* 2019; Tahir *et al.* 2021; Mildenhall *et al.* 2021; Yu *et al.* 2021). Particularly, convolutional neural network (CNN)-based Autoencoders (AEs) and their variants have emerged as an effective solution for predicting 3-D objects from a set of 2-D images (Choy *et al.* 2016; Fan *et al.* 2017; Biffi *et al.* 2019; Tahir *et al.* 2021; Tucsok *et al.* 2022). These models utilize an encoder-decoder architecture commonly employed for compression or dimensionality reduction tasks (Glaws *et al.* 2020; Theis *et al.* 2022). For example, AE has been successfully applied to compress images (Cheng *et al.* 2018) and large-scale turbulent flow simulations (Glaws *et al.* 2020). In the traditional AE framework, the encoder module extracts a low-dimensional representation, known as a latent space, from the input data. Subsequently, the decoder module reconstructs the original input from this latent space representation. When applied to the task of 3-D reconstruction, the encoder module generates a unique low-dimensional representation of 2-D input images, while the decoder module generates a corresponding 3-D representation. These capabilities provide a promising avenue for reconstructing 3-D turbulent flow fields from sparsely distributed information. Moreover, contrary to the traditional reconstruction methods mentioned above, DL-based approaches do not depend heavily on explicit assumptions about the flow systems, thus offering a fruitful pathway for accurate and reliable predictions.

Recently, DL has driven numerous advances in fluid dynamics, including accelerating numerical simulations (Kochkov *et al.* 2021; Jeon *et al.* 2022), developing improved turbulence closure models (Ling *et al.* 2016; Cheng *et al.* 2022), predicting the spatiotemporal behavior of turbulent flows (Lee & You 2019; Cai *et al.* 2021), surrogate modeling of urban boundary layer flows (Hora & Giometto 2024), estimating skin-friction drag over ocean surfaces (Yousefi *et al.* 2024*a*,*b*), and turbulence spectral enrichment efforts (Liu *et al.* 2020*a*; Kim *et al.* 2021). Despite these successes, black-box DL models can produce physically inconsistent or implausible predictions because they do not inherently incorporate the physical laws governing the phenomena under consideration (Raissi *et al.* 2019*b*). To address this challenge, the physics-informed neural network (PINN) methodology has been introduced. PINNs incorporate fundamental physical symmetries and domain knowledge into the DL architecture and loss function, thereby constraining the model to adhere to physics constraints alongside observational data, enhancing its performance and robustness (Karniadakis *et al.* 2021; Cuomo *et al.* 2022).

In recent years, there has been a growing interest in utilizing PINNs to reconstruct turbulent flow fields from spatially limited and noisy measurements (Cai *et al.* 2021; Yousif *et al.* 2023*a*; Clark Di Leoni *et al.* 2023). For example, Cai *et al.* (2021) introduced a method for inferring buoyancy-driven velocity and pressure fields from snapshots of 3-D temperature from Schlieren imaging. Their approach is based on the Raissi *et al.* (2019*b*) methodology and proved capable of capturing natural convection over an espresso cup. In Yousif *et al.* (2023*a*), the authors employed physics-informed generative adversarial networks (Goodfellow *et al.* 2014) to reconstruct 3-D turbulent flow from using 2-D cross-plane measurements. Their study aimed to propose a framework for reconstructing 3-D flows from PIV measurements while reducing the storage costs associated with extensive datasets obtained from experiments and high-fidelity simulations. Similarly, Clark Di Leoni *et al.* (2023) presented a method for reconstructing full and structured Eulerian velocity and pressure fields from 3-D sparse and noisy particle track laboratory measurements. They compared the accuracy of PINN with the constrained cost minimization method (Agarwal *et al.* 2021), demonstrating the superior performance of PINN in reconstructing velocity and pressure fields, even for noisy measurements. Collectively, these studies highlight the potential of PINN in accurately reconstructing turbulent flow fields from limited and noisy measurements.

Building from these works, this study introduces a physics-informed variational autoencoder (PVAE) model that accurately reconstructs 3-D turbulent flow fields from 2-D planar flow measurements obtained at a single wall-parallel plane. The choice of a variational autoencoder (VAE)-based model is motivated by its proven ability to reconstruct 3-D objects from 2-D images and capture high-frequency details with accuracy, particularly in tasks such as image transformation and super-resolution problems (Liu *et al.* 2020*b*,*c*; Tahir *et al.* 2021; Chira *et al.* 2023). This characteristic is especially valuable when tackling multiscale flow phenomena. Generative adversarial networks can serve as a viable alternative to VAEs for turbulence synthesis, as they have shown considerable promise in this area (Stengel *et al.* 2020; Kim *et al.* 2021; Drygala *et al.* 2022). However, generative adversarial networks often face significant challenges, including notoriously difficult training processes (Mescheder *et al.* 2017; Gui *et al.* 2021), optimization/training instability (Salimans *et al.* 2016), vanishing gradient problem (Sinn & Rawat 2018), and mode dropping/collapse issues (Bau *et al.* 2019). In this work, we show that the simpler PVAE approach excels in the considered task.

The proposed PVAE model is designed to learn a mapping $\mathcal{F}$ that takes 2-D velocity field measurements at the pre-specified wall-normal plane as input and estimates the corresponding 3-D velocity field in the computational domain, as shown in figure 1. The mapping $\mathcal{F}$ can be defined as

$$\mathbf{u}^{\mathrm{rec}} = \mathcal{F}(\mathbf{u}^{\mathrm{meas}}), \tag{1.1}$$

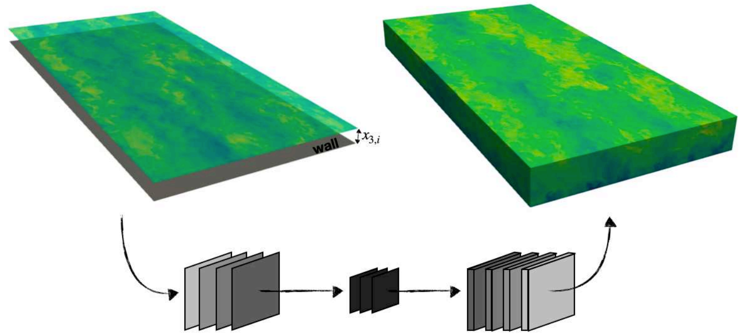

Figure 1: The schematic illustrates the process of using a Variational Autoencoder (VAE)-based generative model to reconstruct a three-dimensional (3-D) flow field $\boldsymbol{u}^{\mathrm{rec}}(x_1, x_2, x_3)$ from a given one wall-parallel plane at height $x_3^* = x_{3,i}$, i.e., $\boldsymbol{u}^{\mathrm{meas}}(x_1, x_2, x_3^*)$ where x_1, x_2, and x_3 denote three Cartesian coordinates. The framework predicts all three velocity components over the considered plane, but only the streamwise component is shown here for simplicity. A conceptual schematic of the network architecture is also shown.

where $\mathbf{u}^{\mathrm{meas}} = (u_1^{\mathrm{meas}}, u_2^{\mathrm{meas}}, u_3^{\mathrm{meas}})$ denotes the measured velocity field at a given wall-parallel plane located at $x_3 = h/2$ (unless otherwise specified), where h is the half channel width and $\mathbf{u}^{\mathrm{rec}} = (u_1^{\mathrm{rec}}, u_2^{\mathrm{rec}}, u_3^{\mathrm{rec}})$ denotes the reconstructed velocity in the 3-D spatial domain. The residuals of the governing equations, including the incompressible Navier-Stokes equations and enstrophy, are incorporated as regularization terms into the loss function of the black-box VAE model. Additionally, boundary conditions (BCs) are embedded into the architecture of the VAE using padding operations. Unlike previous studies that require cross-plane or 3-D dense measurements (Cai *et al.* 2021; Yousif *et al.* 2023*a*; Clark Di Leoni *et al.* 2023), the proposed model only needs three components of the velocity field at one wall-parallel plane. Moreover, the physical realizability of the reconstructed flow field using PVAE is examined and compared against a black box counterpart, i.e., a VAE and a traditional fluid dynamics LSE technique. Both DL models, namely the PVAE and the VAE, as well as LSE method, are trained and evaluated using data from direct numerical simulations (DNS) of turbulent channel flow at a friction Reynolds number of $Re_\tau = u_\tau h/\nu = 250$, where u_τ is the so-called friction velocity, and ν is the kinematic molecular viscosity of the fluid.

The remainder of the paper is organized as follows. The numerical algorithm and dataset are described in §2.1. §2.2 introduces the details of the PVAE and VAE models. The LSE approach is introduced in §2.3. In §3, we assess the performance of the proposed reconstruction methods and their performance against the LSE. Findings are discussed in §4 and conclusions are drawn in §5.

2. Methodology

2.1. *Numerical setup and high-fidelity dataset*

Training and test datasets are obtained from the DNS of turbulent open-channel flow simulations. We solve the incompressible Navier-Stokes equations in rotation form, namely,

$$\frac{\partial u_i}{\partial t} + u_j \left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i} \right) = -\frac{1}{\rho} \frac{\partial p^*}{\partial x_i} + \nu \frac{\partial u_i^2}{\partial x_j^2} + f_1 \delta_{i1} \ \text{ in } \Omega \times [0, T] \ , \tag{2.1}$$

$$\frac{\partial u_i}{\partial x_i} = 0 \ \text{ in } \Omega \times [0, T] \ , \tag{2.2}$$

where x_i is the i$^{\text{th}}$ component of the position vector $\boldsymbol{x} = (x_1, x_2, x_3)$, x_1, x_2, and x_3 denote the streamwise, cross-stream, and wall-normal directions, respectively, u_i is the velocity component in the i$^{\text{th}}$ direction, $p^* = p + \rho u_j^2$ is a modified pressure, ν is the kinematic viscosity of the fluid, $f_1 \delta_{i1}$ is a constant pressure gradient driving the flow in the x_1 direction, ρ is a constant fluid density, $\Omega = [0, L_1] \times [0, L_2] \times [0, h]$ and defines the computational domain, t is time, and T is the integration time. Periodic boundary conditions apply in the wall-parallel (x_1, x_2) directions, and a no-slip boundary condition is enforced at the lower surface $\Gamma_{bottom} = \{\boldsymbol{x} | x_3 = 0\}$, and a free slip boundary condition at the top of the domain $\Gamma_{top} = \{\boldsymbol{x} | x_3 = h\}$.

Equations are discretized via a pseudospectral collocation approach (Orszag 1969) in the wall-parallel direction and a second-order centered staggered finite differences scheme in the wall-normal direction. Nonlinear terms are fully dealiased using the 3/2 rule (Canuto *et al.* 2007). A fully explicit second-order Adams Bashforth method is used for time integration, and a fractional step method is used to solve resulting algebraic equations (Chorin 1968). For more details on the algorithm, please refer to Albertson & Parlange (1999). Over the past two decades, this solver has been extensively used to conduct fluid dynamics research and validated against field and laboratory measurements (see, e.g., Meneveau *et al.* 1996; Porté-Agel *et al.* 2000; Porté-Agel 2004; Lu & Porté-Agel 2010; Hultmark *et al.* 2013; Shah & Bou-Zeid 2014; Pan *et al.* 2014; Fang & Porté-Agel 2015; Anderson *et al.* 2015; Giometto *et al.* 2016, 2017).

As part of this study, a DNS of an open-channel flow is conducted at $Re_\tau = 250$. The computational domain $\Omega = [0, 2\pi h] \times [0, \frac{4}{3}\pi h] \times [0, h]$ is discretized using $128 \times 128 \times 288$ collocation nodes in the x_1, x_2, and x_3 direction, respectively. 6×10^3 instantaneous snapshots of the three velocity components and pressure fields are collected every $T^+ = T u_\tau / h = 10^{-3}$ after the flow has fully developed. To reduce the computational cost of training the DL model and graphics processing unit (GPU) memory requirements, DNS results are interpolated from the $128 \times 128 \times 288$ grid onto a coarser $128 \times 128 \times 64$ grid. This is achieved by applying a top-hat filter and sub-sampling at twice the new Nyquist frequency.

The interpolated dataset is partitioned into training, validation, and test sets, comprising 80%, 10%, and 10% of the total snapshots, respectively. The training set is utilized for model training, while the validation set is employed to monitor model performance during the training and hyperparameter optimization phase. The test set is reserved for the assessment of the model performance. The PVAE, VAE, and LSE approach all use the same dataset for the training and evaluation.

For the efficient training of the DL models, it is also recommended to scale data using pre-processing techniques such as min-max normalization or standardization, typically within the range $[-1, 1]$ or $[0, 1]$, as suggested in Goodfellow *et al.* (2016). For this, we standardize the data by subtracting the $(\mu = \langle \overline{z} \rangle)$ and dividing it by standard deviation $(\sigma = \langle (z - \mu)^2 \rangle^{\frac{1}{2}})$

of the training dataset, i.e., $\hat{z} = \frac{z-\mu}{\sigma}$ to accelerate the training process. Throughout the study, $\overline{(\cdot)}$ is used to denote time-averaging or ensemble averaging (depending on the context), $\langle\cdot\rangle$ is temporal and spatial averaging along the wall-parallel directions, time fluctuations are written as $(\cdot)'$ (therefore $\overline{(\cdot)'} = 0$) and $(\cdot)'_{RMS}$ denotes the root mean square (RMS) of the fluctuation.

2.2. *Deep learning*

This study aims to find a mapping $\mathcal{F}$ that maps 2-D planar wall measurements to a 3-D velocity field (see equation 1.1). To approximate the target mapping $\mathcal{F}$, we propose two DL models: VAE and PVAE, which we discuss in the following sub-sections §2.2.1 and §2.2.2, respectively. To assess the effectiveness and viability of the proposed DL models, we also compare their performance against a traditional LSE approach, described in §2.3.

2.2.1. *Variational autoencoder*

To model the mapping $\mathcal{F}$, we employ a CNN-based VAE model. This method has proven effective in processing grid-like data across diverse domains such as computer vision (Liu *et al.* 2020*b*,*c*), biomedical imaging (Wei & Mahmood 2020), and scientific machine learning (Wang *et al.* 2021*b*; Solera-Rico *et al.* 2024). VAE models are generative models consisting of two parts: an encoder module that maps the input data to parameters of the probability distribution of the latent space, and a decoder module that reconstructs the quantity of interest from the realization drawn from this latent space (Kingma & Welling 2014). During the training phase of VAE, the objective is to minimize the loss function known as the evidence lower bound (ELBO), which consists of two terms: the reconstruction loss ($\mathcal{L}_{\text{data}}$) and the regularization loss ($\mathcal{L}_{\text{reg}}$). The reconstruction loss aims to minimize the discrepancy between the ground truth and reconstructed data, while the regularization term guides the network toward aligning the probability distribution generated by the encoder with the prior distribution over the latent space. The $\mathcal{L}_{\text{data}}$ leverages standard mean-squared error loss function, and $\mathcal{L}_{\text{reg}}$ is implemented using the Kullback-Leibler (KL) divergence (Joyce 2011), which measures the difference between two probability distributions. For more information, we recommend readers refer to Kingma & Welling (2014).

The probabilistic encoder (a.k.a., recognition model) and decoder (a.k.a., generative model) network is represented as $q_\phi(z|\boldsymbol{X})$ and $p_\theta(\hat{\boldsymbol{X}}|z)$, respectively, where z represents latent space, $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$ are the input and output, ϕ and θ are the parameters of the encoder and decoder network, respectively. Typically, VAE assumes a prior distribution as an isotropic multi-variate Gaussian distribution $p(z) = \mathcal{N}(0, \boldsymbol{I})$ and a posterior as a Gaussian distribution with diagonal co-variance $q_\phi(z|\boldsymbol{X}) = \mathcal{N}[\mu_\phi(\boldsymbol{X}), diag(\sigma^2_\phi(\boldsymbol{X}))]$, where μ_ϕ and σ_ϕ are the encoded parameters of latent space probability distribution. The advantage of this assumption is that it provides an exact expression for $\mathcal{L}_{\text{reg}}$ and promotes stable training of VAEs. The objective function of the VAE network ($\mathcal{L}_{\text{VAE}}$) can be expressed as follows:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{data}} + \mathcal{L}_{\text{reg}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{X}_i)^2 + \frac{1}{2}\sum_{i=1}^{k}(\sigma^2_{\phi,i} + \mu^2_{\phi,i} - 1 - \ln(\sigma^2_{\phi,i})), \tag{2.3}$$

where X_i and $\hat{X}_i$ are the actual and predicted values, k is the dimension of latent space and n is the number of training data samples. This choice of $\mathcal{L}_{\text{VAE}}$ also aligns with the existing several VAE-based studies of turbulent flows (Wang *et al.* 2021*b*; Kang *et al.* 2022; Solera-Rico *et al.* 2024).

The overall architecture of the residual block and VAE model are illustrated in figures 2 and 3, respectively. The choice of utilizing residual block in the VAE model is motivated by the

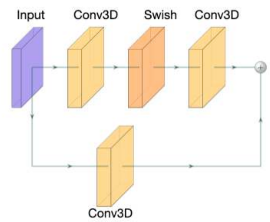

Figure 2: Architecture of the residual block. Conv3D denotes the 3-D convolutional layer, Swish represents the non-linear layer, and the (+) operator represents the elementwise addition operation of a tensor.

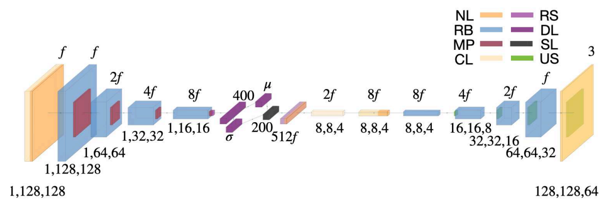

Figure 3: Diagram showcases the architecture of the VAE network, illustrating the number of channels (c) and stride (s) for each convolutional layer and residual block. The values for the number of filters (f), kernel size (k), and stride (s) are uniformly set to 42, 3, and 1, respectively. At each cuboidal block, the top quantity represents the number of filters/channels, and the bottom quantity represents the volumetric spatial dimensions of the output from the corresponding layers. Various components and operations are denoted as follows: NL represents the Swish non-linearity, RB represents the residual block, MP represents the max pooling layer, CL represents the convolutional layer, RS represents the reshape layer, DL represents the dense layer, SL represents the sampling layer, and US represents the upsampling layer.

susceptibility of deep neural networks to the vanishing gradient problem (Hochreiter 1998). Empirically, residual blocks with skip connections have been shown to alleviate this issue, ensuring efficient training (He *et al.* 2016). Many existing studies have leveraged residual block-based DL models to model turbulent flows (for example, Xuan & Shen 2023). As shown in figure 2, the output of the residual block is the sum of the linear projection of the original input and the output of two convolutional and one non-linear layers. The choice of non-linearity in the DL model is critical as it can significantly impact the model's training process and performance on a given task or objective (Goodfellow *et al.* 2016). We have compared the performance of several types of non-linearity, including the hyperbolic tangent function and the sigmoid function (Goodfellow *et al.* 2016), rectified linear units (ReLU) (Nair & Hinton 2010), and Swish (Ramachandran *et al.* 2017). The Swish non-linearity outperformed other non-linearities in the reconstruction task for the considered flow system; thus, we selected it for the present work. This choice aligns with previous research on DL models for turbulent flows (Raissi *et al.* 2019*a*; Hora & Giometto 2024).

The input fed to the VAE model consists of velocity measurements acquired at the wall-parallel plan, resulting in input dimensions of $1 \times 128 \times 128 \times 3$. As shown in figure 3, the encoder module of VAE consists of one convolutional layer, four residual blocks and max pooling layers, a dense layer, a sampling layer, a reshape layer, and swish non-linearity. The stack of residual blocks and pooling layer extracts the feature maps and compresses the input of size $1 \times 128 \times 128 \times 3$ to $1 \times 16 \times 16 \times 8f$, where f is set to 42. Next, we employ the dense layer to encode the extracted feature maps into the parameter space $(\boldsymbol{\mu}_\phi, \boldsymbol{\sigma}_\phi)$ of the latent space distributions. For our work, the dimension of the latent space is set to 200 as it ensures good reconstruction accuracy. To obtain a realization of latent space (z), we utilize a sampling layer $\boldsymbol{z} \sim q_\phi(\boldsymbol{z}|\boldsymbol{X})$ which leverages reparametrization approach described in Kingma & Welling (2014); Rezende *et al.* (2014). In the sampling layer, we first sample a random vector $\boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$ and then generate $\boldsymbol{z}$ as

$$z = \mu_\phi(\boldsymbol{x}) + \sigma_\phi(\boldsymbol{x}) \odot \boldsymbol{u}, \tag{2.4}$$

where $\odot$ is the element-wise product. This approach allows for the differentiation of the ELBO with respect to the model parameters and enables the use of gradient-based optimization methods.

Further, we have a decoder model (see figure 3) to decode the 3-D velocity fields from the latent space. It consists of dense layers, which transform the latent space into a high-dimensional space $\in R^{21504}$ ($512f$ in the figure 3), and a reshape layer that transforms the output of the dense layer into a tensor of rank four, which represents the spatial feature maps. Further, we employ four residual blocks followed by upsampling layers to perform non-linear transformation and increase the feature map resolution. Finally, we have a convolution layer with a linear activation function that outputs the tensor of size $128\times128\times64\times3$, representing the 3-D velocity field. It is important to note that the input and output of the VAE model are based on the same time instant, with each instant considered independently; consequently, no temporal information is utilized in the reconstruction process. In the next section, we will describe our proposed PVAE model.

2.2.2. *Physics-informed variational autoencoder*

As mentioned in §1, purely data-driven DL models may excel at fitting data from high-fidelity models and observations. However, their predictions may lack physical consistency and lead to poor generalization performance. To improve the black box VAE model's performance and physical realizability, we constrain the VAE network to match additional quantities and satisfy selected physical symmetries of the system. In PVAE, the VAE is constrained to match the enstrophy (E) of the reference DNS case and to satisfy the momentum (equation 2.1) and continuity equations (equation 2.2). These biases are introduced as soft constraints into the loss function via three separate regularization terms, namely $\mathcal{L}_\mathrm{E}$, $\mathcal{L}_\mathrm{C}$, and $\mathcal{L}_\mathrm{M}$, whereas BCs are enforced as hard constraints. To enforce the residual of the physical constraint into the loss function, we employ a normalized error based on L_n norm, i.e., e_n, defined as

$$e_n = \frac{||\mathbf{X} - \hat{\mathbf{X}}||_n^n}{||\mathbf{X}||_n^n}, \tag{2.5}$$

$\mathbf{X}$ and $\hat{\mathbf{X}}$ are actual and predicted quantity of interest, $||\mathbf{X}||_n^n = \frac{1}{N}\sum_{i=1}^{N} X_i^n$ and n is the order of the norm. To ensure consistency with the mean-squared error used as a reconstruction loss in $\mathcal{L}_\mathrm{VAE}$ (see equation 2.3), we set $n = 2$. During our experimentation, we found that e_2 also ensures stable training of the PVAE model. A detailed discussion of each physical constraint is provided in the following.

Turbulence is a broadband phenomenon characterized by power-law velocity spectra with negative exponents at high wavenumbers, meaning that the kinetic energy of small-scale motions is much smaller than that of larger scales. Accurately capturing high-wavenumber variations in the flow field is expected to be challenging when using DL models (Lippe *et al.* 2024). Since small-scale motions are the main contributors to E, $\mathcal{L}_{\mathrm{E}}$ biases the PVAE model towards high wavenumber modes, thus enriching the prediction with small-scale information. $\mathcal{L}_{\mathrm{E}}$ is defined as

$$\mathcal{L}_{\mathrm{E}} = e_2(\mathrm{E}^{\mathrm{DNS}}, \mathrm{E}^{\mathrm{rec}}) \ , \tag{2.6}$$

where $\mathrm{E} = \omega_i \omega_i$ is defined as an enstrophy field, $\omega_i = \epsilon_{ijk}(\partial u_k/\partial x_j)$ is the vorticity tensor, ϵ_{ijk} is the Levi-Civita symbol, and $\mathrm{E}^{\mathrm{DNS}}$ and $\mathrm{E}^{\mathrm{rec}}$ are the DNS and PVAE reconstructed enstrophy fields, respectively.

The flow system considered in this study involves incompressible fluids; therefore, it must satisfy the M and C equations. To achieve this goal, the residuals of the M and C equations are introduced as a soft constraint in the loss function to penalize deviations from these equations, following the methodology used in previous works (Raissi *et al.* 2019*b*; Gao *et al.* 2021; Clark Di Leoni *et al.* 2023). The constraining of the PVAE model to learn the residual functions will bias the model to comply with the M and C equations (see equations 2.1 and 2.2), thereby improving the realizability of the generated predictions. $\mathcal{L}_{\mathrm{M}}$ and $\mathcal{L}_{\mathrm{C}}$ can be defined as

$$\mathcal{L}_{\mathrm{M}} = e_2(\mathrm{M}(\mathbf{u}^{\mathrm{DNS}}, p^{\mathrm{DNS}}), \mathrm{M}(\mathbf{u}^{\mathrm{rec}}, p^{\mathrm{DNS}})), , \tag{2.7}$$

$$\mathcal{L}_{\mathrm{C}} = e_2(\mathrm{C}(\mathbf{u}^{\mathrm{DNS}}), \mathrm{C}(\mathbf{u}^{\mathrm{rec}})), \tag{2.8}$$

where $\mathrm{M}(\mathbf{u}, p)$ and $\mathrm{C}(\mathbf{u})$ represents the residual of the M and C equation, respectively and p^{DNS} represents the reference DNS pressure field.

The loss function ($\mathcal{L}_{\mathrm{PVAE}}$) minimized for the PVAE model during training is a combination of a content loss ($\mathcal{L}_{\mathrm{content}}$) and a physics loss ($\mathcal{L}_{\mathrm{physics}}$) and is defined as

$$\mathcal{L}_{\mathrm{physics}} = (1 - \lambda_{\mathrm{C}})\mathcal{L}_{\mathrm{M}} + \lambda_{\mathrm{C}}\mathcal{L}_{\mathrm{C}}, \tag{2.9}$$

$$\mathcal{L}_{\mathrm{content}} = (1 - \lambda_E)\mathcal{L}_{\mathrm{VAE}} + \lambda_{\mathrm{E}}\mathcal{L}_{\Omega}, \tag{2.10}$$

$$\mathcal{L}_{\mathrm{PVAE}} = (1 - \lambda_{\mathrm{P}})\mathcal{L}_{\mathrm{content}} + \lambda_{\mathrm{P}}\mathcal{L}_{\mathrm{physics}}, \tag{2.11}$$

where λ_{C}, λ_{E}, and λ_{P} are the regularization constant for $\mathcal{L}_{\mathrm{C}}$, $\mathcal{L}_E$, and $\mathcal{L}_{\mathrm{physics}}$, respectively used to balance each term of the $\mathcal{L}_{\mathrm{PVAE}}$. $\mathcal{L}_{\mathrm{E}}$, $\mathcal{L}_{\mathrm{M}}$ and $\mathcal{L}_{\mathrm{C}}$ comprise both spatial and temporal derivatives. Spatial derivatives are computed via a second-order accurate centered finite-difference scheme. A non-trainable convolution kernel is engineered to evaluate spatial derivatives based on the finite difference scheme (Gao *et al.* 2021; Xuan & Shen 2023). While we also explored using higher-order schemes, we found no significant improvement in the reconstruction accuracy and hence settled for second order. One plausible explanation is that the majority of the reconstructed structures are low-wavenumber modes, for which the second-order accurate scheme suffices in resolving them (Xuan & Shen 2023). To calculate the pressure gradient ($\partial p/\partial x_i$) and acceleration ($\partial u_i/\partial t$) terms in the M equation, flow fields from the DNS dataset are leveraged. This is appropriate because the pressure and velocity information is only required during the training phase, and once the network is trained, it can reconstruct flow fields without any additional information. It is worth noticing that the residuals of the governing equations in $\mathcal{L}_{\mathrm{M}}$ and $\mathcal{L}_{\mathrm{C}}$ are calculated against the residuals of the governing equations on reference DNS data instead of a null tensor. This is so because the DNS is initially carried out using a pseudo-spectral approach and, therefore, may not satisfy the governing equations with zero residuals on the finite-difference stencil. We use the residuals on the DNS data to ensure fair comparisons of the violation of the governing equations. Lastly, BCs are enforced as a hard constraint through a padding operation–

specifically, by adding extra pixels or ghost cells around the edges of the input data, as described in Gao *et al.* (2021). For instance, to enforce Dirichlet BCs, a constant padding value is uniformly applied around the input data.

2.2.3. *Model parameters*

In §2.2.1 and §2.2.2, we described the VAE and PVAE architectures; this section discusses corresponding model hyperparameters and the training setup.

The performance and computational cost of the CNN-based models are influenced by various parameters, including kernel sizes, number of kernel/channels (feature maps) in the convolutional layer, and the downsampling and upsampling ratio of max pooling and upsampling layer, among others. A typical choice for kernel size lies between 3 to 7 to capture fine-grained details and extract relevant features of the dataset (Simonyan & Zisserman 2014; Xuan & Shen 2023). The number of channels also affects the performance of the model; ideally, more channels can increase the model's capacity to capture and describe a wider variety of features (Goodfellow *et al.* 2016). To balance model performance and computational cost, we set the kernel size to 3 and the number of channels to a multiple of f, where f is set to 42 (see figure 3). The upsampling and downsampling ratio is kept as two, which is a common choice (Kang *et al.* 2022).

For the PVAE model, the optimal values of $\lambda_{\mathrm{E}}, \lambda_{\mathrm{C}}$, and λ_{P}, are determined through a Grid-based hyperparameter search approach (Goodfellow *et al.* 2016). To avoid the trivial solution of zero fields becoming a local minimum, these parameter values are restricted to the $[0, 0.50]$ range, as suggested in Subramaniam *et al.* (2020). Although not explicitly shown here, hyperparameter analysis revealed that $\lambda_{\mathrm{E}} = 0.25, \lambda_{\mathrm{P}} = 0.25, \lambda_{\mathrm{C}} = 0.50$ is optimal for the PVAE model, as it achieved the best overall performance on the validation dataset. We also acknowledge that other hyperparameter tuning methods, such as random search or Bayesian optimization, could be explored to improve the model's performance further (Goodfellow *et al.* 2016).

The VAE and PVAE model with the aforementioned setup is designed using the TensorFlow ML library (Abadi *et al.* 2016). Trainable parameters are randomly initialized using realization drawn from a uniform distribution (Glorot & Bengio 2010), as done before by the authors (Hora & Giometto 2024; Yousefi *et al.* 2024*b*). The learning rate is kept constant at 5×10^{-4} throughout the training and trained end-to-end by backpropagation using the Adam optimizer (Kingma & Ba 2014). Due to GPU memory limitations, an effective mini-batch size of 100 is employed. The number of epochs chosen is 5000 epochs, and it is based on the observation that extending training beyond this point did not yield significant improvements in the reconstruction accuracy on the validation dataset.

2.3. *Linear stochastic estimation*

To better assess the performance of the proposed DL model, we will compare its predictions against a more traditional approach based on the LSE technique (Adrian & Moin 1988). Traditionally employed for the extraction of coherent structures in turbulent flows (Christensen & Adrian 2001), the LSE method has more recently found applications in reconstructing velocity fields across various scenarios. These applications include off-wall plane velocity fluctuation reconstruction in turbulent open-channel flow using wall-shear-stress components and pressure measurements (Guastoni *et al.* 2021), reconstructing 3-D velocity fields from surface velocity and elevation measurements (Xuan & Shen 2023), and characterizing vorticity fields (Wang *et al.* 2021*a*). Consequently, the LSE method offers a suitable benchmark for our study.

The LSE method involves estimating a linear expression comprising empirical parameters from the measured quantifies to infer unknown quantities. In the context of this work, the

LSE is employed to estimate the 3-D velocity field fluctuations from 2-D measurements. Mathematically, this operation is defined as

$$u_i^{\prime,\mathrm{rec}} = Q_{ij} u_j^{\prime,\mathrm{meas}}, \quad j = 1, 2, 3, \tag{2.12}$$

where Q_{ij} is a linear operator. This linear operator, Q_{ij}, is determined by minimizing the mean squared difference between the reconstructed and corresponding DNS velocity fields. For the reconstructed velocity field $\mathbf{u'}^{,\mathrm{rec}}$ and the corresponding DNS field $\mathbf{u'}^{,\mathrm{DNS}}$, the mean squared discrepancy D can be expressed in an integral form as

$$\mathrm{D} = \frac{\int\int\int_V (\mathbf{u'}^{,\mathrm{rec}} - \mathbf{u'}^{,\mathrm{DNS}})(\mathbf{u'}^{,\mathrm{rec}} - \mathbf{u'}^{,\mathrm{DNS}})}{\int\int\int_V dV}. \tag{2.13}$$

The optimal value of Q_{ij} is determined by minimizing $\overline{\mathrm{D}}$ (Wang *et al.* 2021*a*). Further, this optimization problem can be formulated as

$$Q_{ij}\overline{u_j^{\prime,\mathrm{meas}}(\mathbf{r})u_m^{\prime,\mathrm{meas}}(\mathbf{r'})} = \overline{u_i^{\prime,\mathrm{DNS}}u_m^{\prime,\mathrm{meas}}(\mathbf{r'})}, \quad m = 1, 2, 3. \tag{2.14}$$

Here, $\mathbf{r} \in V$ is a position vector, $\overline{u_j^{\mathrm{meas}}(\mathbf{r})u_m^{\mathrm{meas}}(\mathbf{r'})}$ can be defined as averaged measured velocity distribution given an event of $u_m^{i,\mathrm{meas}}$ occurring at $\mathbf{r'}$ and similarly, $\overline{u_i^{\mathrm{DNS}}u_m^{\mathrm{meas}}(\mathbf{r'})}$ as conditionally averaged velocity distribution for the event u_m^{meas} (Wang *et al.* 2021*a*; Xuan & Shen 2023). In general, equation 2.14 leads to a linear system with a large number of variables directly linked to the discretization of the spatial domain. This characteristic renders the computational solution of such a linear system impractical. To circumvent this limitation, and as suggested by Wang *et al.* (2021*a*) and Xuan & Shen (2023), we leveraged the wall-parallel homogeneity inherent in open-channel flow. This enabled us to decouple equation 2.12 in the wall-normal direction, making it possible to independently evaluate it for each wall-parallel ($x_1 - x_2$) plane. For more information on the LSE approach, we recommend readers to refer Wang *et al.* (2021*a*) and Xuan & Shen (2023).

3. Results

In this section, we evaluate the performance of the PVAE, VAE, and LSE models. A qualitative and quantitative comparison of the reconstructed velocity field, using instantaneous pseudo-color maps alongside corresponding probability density function and vortical structures using the Q criterion, is presented in §3.1. Next, we examine the energy spectra and auto-correlation of the reconstructed flow and compare them to DNS results in §3.2. The space- and time-averaged wall-normal profiles of turbulent flow statistics are presented in the §3.3. Moreover, we conduct a physical realizability analysis focusing on the residual of the momentum and continuity equations, elaborated upon in §3.4. Finally, in §3.5, we explore the impact of measurement on the reconstruction accuracy of the models.

3.1. *Reconstructed instantaneous flow field*

In this and the subsequent §3.2, we use a single snapshot from the test dataset as an illustrative example to compare the reconstructed flow field with the ground truth DNS results. To begin with, we test the ability of the PVAE to reconstruct the 3-D instantaneous velocity field fluctuations using pseudocolor maps, as these are a valuable starting point for characterizing discrepancies between reference and predicted fields. To this end, instantaneous streamwise ($u_1^{\prime +}$), cross-stream ($u_2^{\prime +}$), and wall-normal ($u_3^{\prime +}$) velocity fluctuations from the PVAE are compared against corresponding DNS, VAE, and LSE predictions over a chosen wall-parallel

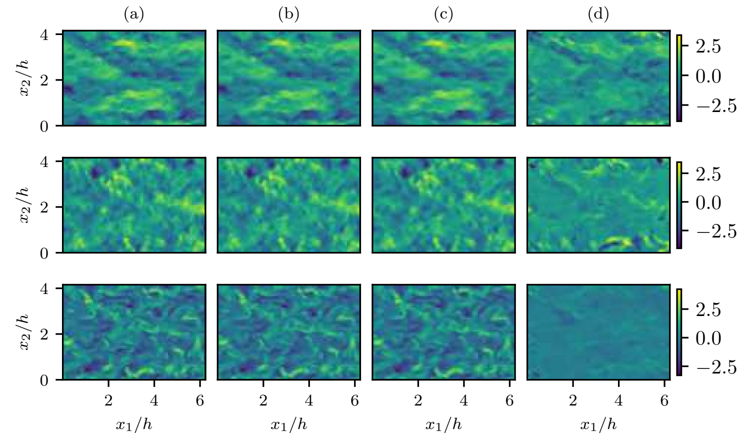

Figure 4: Instantaneous snapshot of normalized streamwise $u_1'^+$ (top), cross-stream $u_2'^+$ (middle), and the vertical $u_3'^+$ (bottom) velocity fluctuations at height $x_3 = 2h/5$. Reference direct numerical simulation (DNS), physics-informed variational autoencoder (PVAE), VAE, and linear stochastic estimation (LSE) results are shown in panels corresponding to columns (a), (b), (c), and (d), respectively. h is the height of the computational domain, and x_1 and x_2 are streamwise and cross-streamwise directions. The superscript + indicates a quantity scaled in inner units using the fluid viscosity ν and the friction velocity u_τ.

plane in figure 4. All velocity fluctuation components are scaled in inner units using ν and u_τ. The rationale for examining velocity fluctuations is that these provide an intuitive picture of the spatial structure of the flow while also enabling comparison with the LSE approach (which can only predict fluctuations as described in §2.3). It is evident from figure 4 that predictions from both the PVAE and VAE models at the (arbitrarily chosen) $x_3 = 2h/5$ plane are in remarkable agreement with the reference DNS data and surpass the performance of the traditional LSE approach. For the instantaneous streamwise velocity fluctuation component $u_1'^+$, the reference DNS solution is characterized by streamwise-elongated, high- and low-speed streaks flanking each other in the cross-stream direction. The DL models excel in reconstructing both the spatial variability and the magnitude of these flow features. The spanwise $u_2'^+$ and vertical $u_3'^+$ velocity fluctuations feature modes of variability that are relatively more compact in space when compared to those of the $u_1'^+$ field, which are expected to pose a challenge for data-driven approaches (Xuan & Shen 2023). Nonetheless, based on visual inspection, the DL methods again exhibit remarkable accuracy in capturing these instantaneous fields. Interestingly, the addition of physical constraints does not appear to yield any apparent improvement in the structure of the predicted instantaneous flow field. LSE predictions appear more homogeneous on the considered plane, resulting in a loss of critical spatial variability details. This behavior is particularly noticeable in the representation of high wavenumber modes. Although not shown here, the LSE performance significantly degrades at planes more distant from the $x_3 = h/2$ sampling location—a finding consistent with that from other studies (Suzuki & Hasegawa 2017; Xuan & Shen 2023). The upcoming discussion will focus exclusively on the DL approaches.

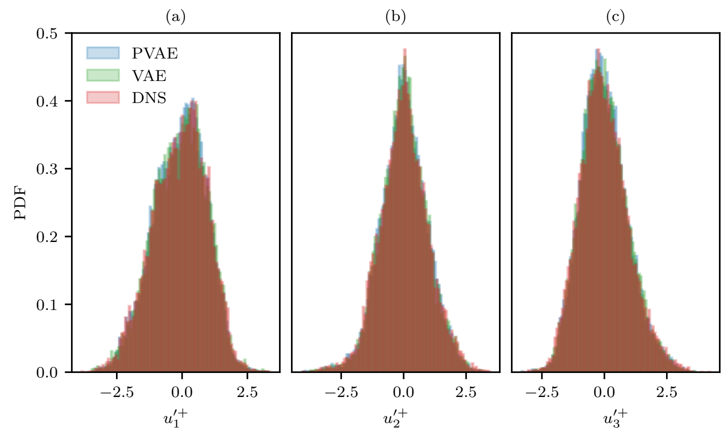

Figure 5: Empirical probability density function (PDF) of $u_1'^+$ (a), cross-stream $u_2'^+$ (b), and the wall-normal $u_3'^+$ (c) velocity fluctuations for the DNS (red), VAE (green), and PVAE (blue) models at $x_3 = 2h/5$.

To gain further insight into the spatial variability of the flow fields, figure 5 presents histograms of the $u_1'^+$, $u_2'^+$, and $u_3'^+$ fields evaluated over the aforementioned plane. Panels a, b, and c in figure 5 correspond to $u_1'^+$, $u_2'^+$, and $u_3'^+$ fluctuation fields, respectively. The histogram uses blue color for PVAE, green for VAE, and red for the ground truth DNS results. The figure demonstrates that both the PVAE and VAE models successfully capture the majority of the DNS velocity variability. However, larger discrepancies compared to the reference DNS data are observed near the mode of the distributions, representing the most frequent events. The results again indicate that incorporating physical constraints does not improve the predictions of the DL models.

The predictive capabilities of the PVAE model are further assessed in terms of coherent vortex structures, the building blocks of turbulence. Coherent vortices are defined as flow regions with long-lasting vorticity concentration ω, allowing for a local roll-up of the surrounding fluid (Lesieur 1997). These structures play a crucial role in transporting mass, energy, and momentum within turbulent flows and have been the focus of sustained research in the past decades (Robinson 1991). Analyzing vortical structures in the predictions of DL models is also important because a mean squared error function is utilized to minimize the discrepancy between the reconstructed velocity field and the corresponding ground truth data (see §2.2). However, this error minimization on the velocity field does not inherently ensure an accurate representation of vortical structures in the reconstructed flow fields. In figure 6, we present selected isosurfaces of the Q-criterion (Dubief & Delcayre 2000) obtained from the reference DNS, the PVAE, and the VAE models. These visualizations effectively demonstrate the presence of hairpin heads and tails, corroborating findings from previous studies (Scott *et al.* 1991). Notably, the vortical structures observed in the reconstructed instantaneous velocity fields of the DL models exhibit a remarkable resemblance to the DNS data, indicating that the proposed formulations can accurately reproduce the salient features

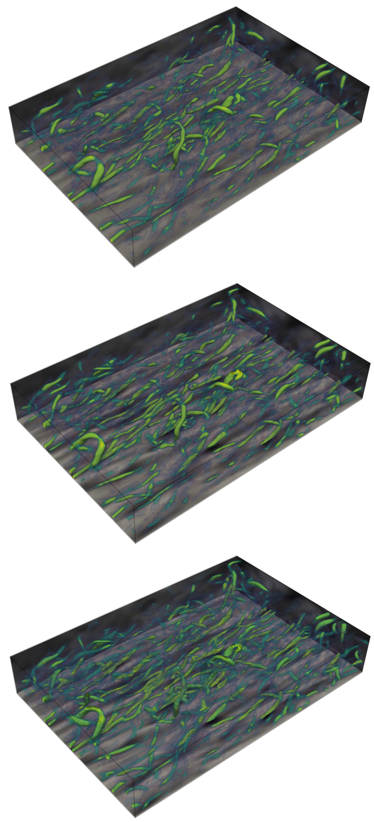

Figure 6: Instantaneous turbulent vortical structures extracted using Q criterion for the DNS (top), PVAE (middle), and VAE (bottom) cases. Different colors and transparency denote isosurfaces with different Q magnitudes.

of boundary-layer turbulence. Upon visual inspection, it is also apparent that the PVAE and VAE flow fields are qualitatively similar, and the inclusion of physical constraints does not appear to enhance the representation of the vortical structure.

To provide a quantitative measure of model accuracy, we next compare predictions in terms of a normalized mean-squared error (e_2) (see equation 2.5). Focusing on the $x_3 = 2h/5$ plane from figure 4, the PVAE (VAE) e_2 error is 2.8% (3.3%) for $u_1'^+$, 4.8% (5.4%) for $u_2'^+$, and 5.8% (7.6%) for the $u_3'^+$ velocity fluctuation components. The error in the streamwise velocity component ($u_1'^+$) is relatively lower than that for the cross-stream and wall-normal components ($u_2'^+$ and $u_3'^+$). This phenomenon has also been noted in recent work by Yousif

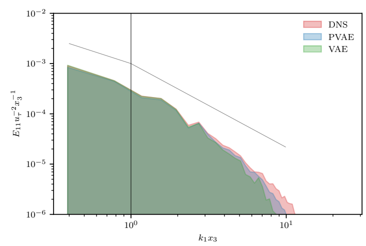

Figure 7: Vertical structure of the normalized one-dimensional energy spectra of streamwise velocity ($E_{11} u_\tau^{-2} x_3^{-1}$) components at height $x_3 = 2h/5$. The solid gray line depicts the $(k_1 x_3)^{-1}$ production range and $(k_1 x_3)^{-5/3}$ in the inertial subrange scaling.

et al. (2023*a*), where it was suggested that such a disparity may be attributed to the dominance of the streamwise velocity field in pressure-driven wall-bounded flows (Pope 2000), which consequently becomes the primary focus of DL models during their predictions. Another plausible reason for the variable accuracy in predicting velocity field is the use of a cumulative loss function (Hora & Giometto 2024). Hora & Giometto (2024) noted that the cumulative loss function does not impose specific constraints on individual predicted quantities, leading to variable accuracy in flow statistics predictions. Nevertheless, in the considered plane, errors are modest, and the overall accuracy of model predictions would be suitable for most geophysical and engineering applications assuming the same error magnitudes at higher Re_τ.

Based on the above analysis, it can be concluded that both the PVAE and VAE models can qualitatively reconstruct the instantaneous velocity field at the unseen wall-parallel plane and outperform the LSE approach. For a more quantitative assessment of model performance, we next examine the reconstructed velocity spectra and two-point correlation statistics.

3.2. *Energy spectra, and two-point correlations*

One-dimensional streamwise spectra of streamwise velocity (E_{11}) are shown in figure 7. The spectra corresponding to the PVAE, DNS, and VAE models are represented using blue, red, and green colors, respectively. It is apparent from the figure that the DL models accurately capture the energy distribution of large-scale structures (small k_1). However, notable differences between the DNS and DL models appear as the k_1 values increase, particularly for the VAE model. These discrepancies highlight a limitation in accurately learning and reproducing small-scale flow variability. This behavior can be explained via the frequency principle (F-principle). According to this principle, when a DL model is trained using the mean-squared objective function, low-frequency information is usually learned with greater accuracy when compared to high-frequency information (Xu *et al.* 2019; Zhang *et al.* 2022). Further, the error analysis using e_2 (see equation 2.5) of PVAE (VAE) reveals that for the area under the curve of streamwise energy E_{11} spectra in the production ($k_1 x_3 < 1$) and inertial and dissipation subranges ($k_1 x_3 \geqslant 1$) are 0.4% (0.1%) and 1.3% (3.1)%, respectively. The enstrophy loss term introduced in the PVAE model yields an apparent improvement in this sense, but local (in k_1) errors remain substantial, corroborating the argument made in Beucler *et al.* (2021) that soft constraints may not enable the DL model to satisfy the physics exactly. Overall, results indicate that the PVAE performs marginally better than the VAE model and that both DL models successfully capture the overall kinetic energy of the flow.

We next examine the spatial coherence of the predicted flow field via the two-point auto-

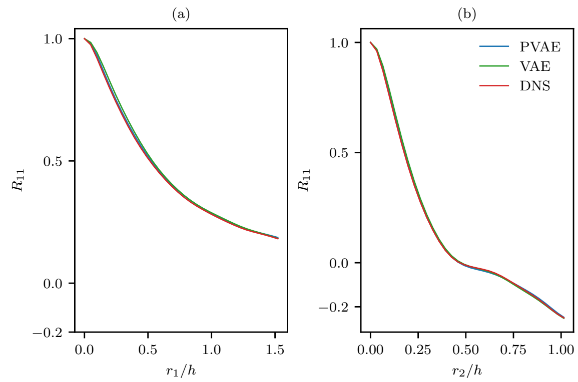

Figure 8: One-dimensional spatial autocorrelation (R_{11}) of streamwise velocity at height $x_3 \approx 2h/5$ along the streamwise (a) and cross-stream direction (b).

correlation R_{11} function along the streamwise (x_1) and cross-stream directions (x_2), shown in figure 8. From a physical perspective, it is notable that R_{11} remains finite within the considered range of r_1/h (figure 8, a), implying that the selected computational domain is not large enough for the flow to decorrelate completely. Additionally, the spatial autocorrelation plot reveals negative lobes in the cross-stream direction (figure 8, b), highlighting the presence of streamwise-elongated high- and low-momentum streaks, flanking each other in the cross-stream direction (Zhou *et al.* 1999). This quantity relates to the velocity spectrum, and we can observe that the PVAE model does an excellent job of predicting such a profile, suggesting that the proposed formulation can effectively capture the large-scale structure of the flow field—the ones primarily contributing to flow coherence. Moreover, the VAE model also aligns remarkably well with the DNS profile, underscoring that integration of physical constraints does not significantly enhance model performance in this specific aspect.

3.3. *One-dimensional flow statistics*

The wall-normal structure of flow statistics has been the subject of sustained research in the past decades, owing to the key role they play in controlling surface drag as well as mass and energy exchanges across a range of applications (Nagib & Chauhan 2008). Figure 9 compares the DL and DNS predictions in terms of normalized streamwise velocity (a) and root-mean-square (RMS) velocity (b,c,d) profiles. Focusing on $\langle u_1^+ \rangle$, it is apparent that the proposed PVAE and VAE models accurately predict such a quantity. DL-based RMS profiles are also in great agreement with corresponding DNS data, indicating that the DL algorithms are able to correctly capture second-order moments of the velocity field. The maximum percentage error of the PVAE (VAE) model occurs at the peak of the profiles, with a $\max(e_1)$ error of 5.4% (3.9%) for $\langle u'^+_{1,\mathrm{RMS}} \rangle$, 6.6% (4.2%) for $\langle u'^+_{2,\mathrm{RMS}} \rangle$, and 10.2% (6.5%) for $\langle u'^+_{3,\mathrm{RMS}} \rangle$.

All in all, this and the previous sections have shown that both the PVAE and VAE models can reconstruct 3-D flow fields from 2-D planar measurements that are in excellent agreement with corresponding DNS results in terms of coherent structures, velocity spectra, spatial flow coherence, and one-dimensional profiles of velocity statistics. However, we note that although physical constraints such as momentum and mass conservation have been

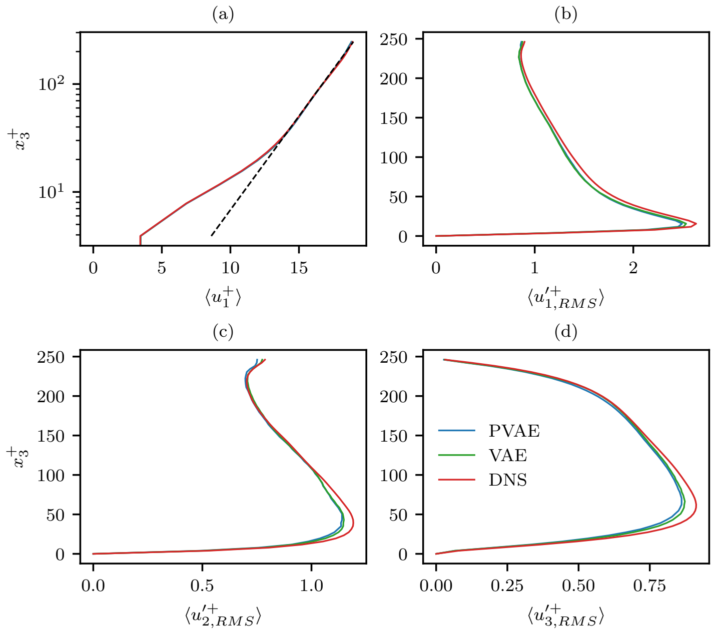

Figure 9: Vertical structure of normalized mean streamwise velocity $\langle u_1^+ \rangle$ (a), streamwise $\langle u'^+_{1,\mathrm{RMS}} \rangle$ (b), cross-stream $\langle u'^+_{2,\mathrm{RMS}} \rangle$ (c), and wall-normal $\langle u'^+_{3,\mathrm{RMS}} \rangle$ (d) root mean square (RMS) velocity fluctuations. $\langle \cdot \rangle$ denotes the averaging operation in time and along coordinates of statistical homogeneity (x_1, x_2).

incorporated into the objective function of the PVAE model, this has not led to significant improvements in terms of model performance when compared to the VAE formulation, with the exception of velocity spectra. This finding suggests that the "physics-less" VAE model might have approximately learned these constraints during the training process, which would justify its accuracy. To gain further insight into this, the next section examines the ability of the proposed models to conserve mass and momentum—the constraints explicitly enforced in the PVAE.

3.4. *Physical realizability*

The considered flow system is governed by the incompressible Navier-Stokes and mass-conservation equations, i.e., M and C equations (equations 2.1 and 2.2). Compliance with these symmetries is, hence, an important requisite in the model assessment. To determine how well the proposed DL algorithms adhere to the conservation equations throughout the training phase, the physics residuals $\mathcal{L}_\mathrm{M}$ and $\mathcal{L}_\mathrm{C}$ (equations 2.7 and 2.8) are presented against the number of training epochs. These residuals, depicted in figure 10, function as

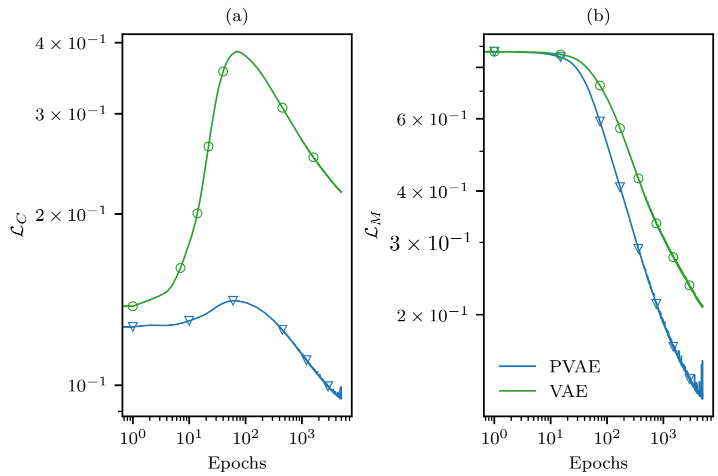

Figure 10: Continuity ($\mathcal{L}_C$) (a) and momentum($\mathcal{L}_M$) (b) regularization term against number of epochs. Epochs indicate the number of iterations used in the learning process.

indicators of the deviation from the governing equations. Upon completing the training phase, the $\mathcal{L}_C$ and $\mathcal{L}_M$ losses associated with the PVAE model are approximately half of those recorded for the VAE counterpart. The incorporation of physical constraints also yields a more rapid convergence of the corresponding loss terms, which is especially apparent for the $\mathcal{L}_M$ loss. This result underscores the important role played by physical constraints in guiding the learning process toward solutions that are more consistent with the governing equations. However, it is pertinent to note that when considering the scale of magnitude, losses from the VAE model are still within a comparable range to those of the PVAE, thereby justifying the commendable performance of the former in accurately capturing flow statistics. Similar to the findings in the previous sections, we observe that adding physical laws as a soft constraint to the DL model does not ensure that the predictions satisfy these laws exactly, and thus, there is a need to apply them as a hard constraint (Beucler *et al.* 2021).

In terms of computing time, training the PVAE and VAE models using four NVIDIA RTX A6000 GPUs took approximately five and four days, respectively. Given the modest increase in computational cost associated with training the PVAE, introducing physical constraints may be justified for applications that demand higher accuracy in small-scale feature reconstructions and better consistency with the underlying governing equations.

3.5. *Impact of measurement plane on reconstruction accuracy*

Results presented in the preceding sections are representative of reconstructed flows based on planar measurements sampled at $x_3 = h/2$. However, information may be available at different wall-normal distances, and the choice of sampling plane may impact the predictive accuracy of the proposed model. To investigate model sensitivity to the measurement location, two additional sampling locations are here considered, namely $x_3 \in \{h/4, 3h/4\}$, or

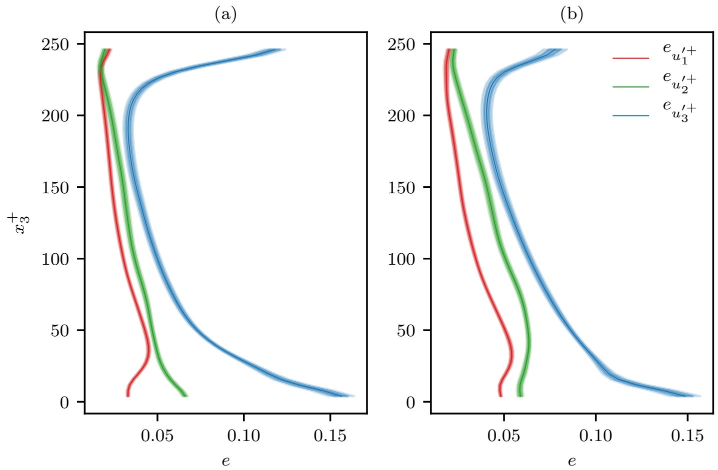

Figure 11: Mean (μ) and standard deviation (σ) of the reconstruction error (e) for the PVAE (a) and VAE (b) model across different sampling planes. Three sampling planes are considered for evaluating these statistics, namely $x_3 = \{h/4, h/2, 3h/4\}$. Solid lines depict $\mu(e)$, $\pm\sigma(e)$ (dark color), and $\pm 2\sigma(e)$ (light color). The analysis excludes the bottom and top boundaries due to the presence of zero values.

equivalently, $x_3^+ = \{62.5, 187.5\}$. A PVAE model with fixed architecture and hyperparameters is trained for each of these sampling locations. For comparison, a corresponding VAE model is also trained for each of these sampling locations, yielding a total of four additional DL predictions.

Model performance is evaluated in a statistical sense via comparison of reconstruction error as a function of x_3^+ for the streamwise ($u_1'^+$), cross-stream ($u_2'^+$), and wall-normal ($u_3'^+$) velocity fluctuations. Figure 11 presents the results of such an analysis. In this figure, the reconstruction error $e(x_3^+)$ of a velocity component is defined as

$$e(x_3^+) = \frac{1}{n}\sum_{k=1}^{n} \frac{\int\int_{x_1,x_2} (u_j'^{,\text{rec,k}} - u_j'^{,\text{DNS,k}})^2 dx_1 dx_2}{\int\int_{x_1,x_2} (u_j'^{,\text{DNS,k}})^2 dx_1 dx_2}, \tag{3.1}$$

where u_j' is the j^{th} component of velocity fluctuations, k is the k^{th} sample, and n is the total number of samples in the test dataset. Further, the mean and standard deviation of $e(x_3^+)$, at each wall-normal location are defined as $\mu = \frac{1}{3}\sum_{i=1}^{3} e_i(x_3^+)$ and $\sigma = \sqrt{\frac{1}{3}\sum_{i=1}^{3}[e_i(x_3^+) - \mu]^2}$, respectively. Here, $e_i(x_3^+)$ represents the reconstruction error corresponding to the DL models trained at the i^{th} sampling plane. Shaded regions in figure 11 depict the error variability amongst models trained using varied sampling locations. As shown in figure 11, the overall reconstruction error of both models is within 7% for $u_1'^+$ and $u_2'^+$ and can reach up to about 15% for the $u_3'^+$ velocity fluctuation. The PVAE slightly outperforms the VAE in predicting velocity fluctuations overall. Close to the wall, both models exhibit relatively larger reconstruction errors for the $u_1'^+$ and $u_2'^+$ velocity fluctuation components, with a gradual improvement as we move towards the channel half width. The behavior of $u_3'^+$ deviates from this pattern, displaying relatively higher error values near both the wall and the free surface. This behavior can be easily explained when considering the small magnitude of wall-normal velocity fluctuation in the vicinity of the lower and top boundaries, which magnify relative errors. The bottom and top boundaries are rigid lids, which significantly dampen $u_3'^+$ velocity fluctuations and lead to energy redistribution to the wall-parallel planes. What is also apparent from the figure both models feature small σ values (fraction of a percent), indicating that model performance is insensitive to the choice of sampling location at the considered Reynolds number. In summary, the proposed PVAE model is highly accurate and is expected to yield consistent performance irrespective of the wall-parallel sampling plane. The inclusion of physical constraints also yields a modest but consistent improvement in performance when compared to the black box approach.

4. Discussion

This section offers a perspective on the previous findings and investigates the cause of observed discrepancies in the proposed model predictions. §3.1 demonstrated that the DL models effectively reconstructed the 3-D velocity flow fields from 2-D planar wall-parallel measurements and outperformed the traditional LSE approach. However, the accuracy of DL models varies depending on the specific flow variable. In particular, DL models showed higher accuracy in capturing the streamwise (u_1) velocity field compared to the cross-stream (u_2) and wall-normal (u_3) components. Results in §3.2 also showed that DL models accurately captured the large-scale structures; however, they faced challenges in representing the small-scale structure. The integration of the enstrophy constraint in the PVAE model enhanced its ability to capture small-scale structures, though slight discrepancies still existed locally for large wavenumbers. Further, in §3.4, we found that the inclusion of physical constraints, namely, incompressible Navier-Stokes, improved the consistency of the reconstructed flow fields with the governing equations, though the predictions did not fully satisfy these constraints. The following paragraphs will provide further insights into these findings and discuss strategies to address these limitations and improve model performance.

As discussed in §2, both the DL and LSE approaches primarily use convolution operations. However, the DL models demonstrated superior performance compared to the LSE approach in reconstructing 3-D velocity fields from 2-D measurements (see §3.1). The success of DL models could be attributed to their use of non-linear transformation. The DL models employed a two-step transformation process: first, the input was transformed linearly using convolutional layers, followed by the application of Swish functions to introduce non-linear effects. In contrast, the transformation in the LSE was purely linear, using the Q operator (see §2.3). Additionally, the DL model includes upsampling and downsampling layers, which adjust the resolution or number of grid points in the outputs of the convolutional layers. This allows the filters/kernels in different convolutional layers to efficiently process features at varying spatial scales, improving the model's ability to capture a wide range of scales present in the data. The combination of convolutional, non-linearity, upsampling, and downsampling layers enabled the DL models to approximate a complex $\mathcal{F}$ mapping (as described in equation 1.1), which the LSE model could not achieve. This finding is consistent with numerous studies that have demonstrated the superior performance of non-linear DL models over linear methods in addressing turbulent flow problems (Guastoni *et al.* 2021; Xuan & Shen 2023).

The quantitative error analysis using the normalized mean square error metric in §3.1 showed that the DL models captured the streamwise (u_1) velocity field with greater accuracy compared to the cross-stream (u_2) and wall-normal (u_3) components. During the training phase of DL models, we utilized standard mean-squared error function as a $\mathcal{L}_{\text{data}}$ (equivalent to the error of the kinetic energy) which is an aggregated measure of the reconstruction accuracy for the different velocity component. In the considered open-channel flow setup, the Reynolds stress tensor is highly anisotropic—a property stemming from the distinct distribution of u_1, u_2, and u_3. Anisotropy is the root cause of the observed imbalance in the loss terms, ultimately impacting the accuracy of the model in capturing the different velocity components (Clark Di Leoni *et al.* 2023). One plausible approach to alleviate this issue is to separate the $\mathcal{L}_{\text{data}}$ (equation 2.3) into three components with different weights, i.e.,

$$\mathcal{L}_{\text{data}} = \sum_{j=1}^{3} \beta_j \frac{1}{n} \sum_{i=1}^{n} (u_{j,i}^{\text{DNS}} - u_{j,i}^{\text{rec}})^2 , \tag{4.1}$$

where β_j is an independent hyperparameter utilized to balance each term of $\mathcal{L}_{\text{data}}$ (Clark Di Leoni *et al.* 2023). The value of β_j's can be determined by incorporating them as trainable parameters of the model (Xiang *et al.* 2022).

In §3.2, we analyzed energy spectra to evaluate the capability of DL models to reconstruct flow scales in turbulent channel flow. Both DL models accurately captured the large-scale structures; however, notable differences between the DNS and DL model predictions were apparent for the small-scale structures, particularly for the VAE model. The limitations of the DL models in capturing small-scale structures can be attributed to the F-principle. According to this principle, when training a DL model, it tends to learn the low-frequency components more accurately and quickly while exhibiting relatively poorer performance with high-frequency components, as discussed in existing studies (Xu *et al.* 2019; Zhang *et al.* 2022). The marginally better performance of the PVAE model over the VAE model in capturing small-scale structures could be attributed to the enstrophy constraint. During the training of the PVAE model, enstrophy was added as a soft constraint, which led to better prediction of small-scale structures. However, this approach still resulted in some minor discrepancies, as soft constraints do not precisely satisfy physical laws (Beucler *et al.* 2021). Recently, Lippe *et al.* (2024) proposed a partial differential equations (PDE)-Refiner approach which enhances DL models' ability to accurately model structures corresponding to all wavenumbers. Therefore, employing a PDE-Refiner approach for the reconstruction of velocity fields could potentially enhance model performance and more accurately capture small-scale structures.

In §3, the findings revealed that the incorporation of the momentum and mass conservation as a soft constraint alongside the observation data did not yield any discernible advantage in the reconstruction accuracy. However, it was observed that for the flow system under consideration, the physical constraints enhanced the physical realizability of the DL model (see §3.4). To introduce these physical principles as soft constraints in the loss term, we introduced λ_* as new hyperparameters. For this study, we employed a grid-based hyperparameter search approach to identify the optimal values. However, alternative strategies could be explored to determine these additional hyperparameters, potentially yielding improved results. For example, they could be updated based on the analysis of the Hessian of the loss function (Wang *et al.* 2021*c*) or integrated into the model as trainable parameters (Xiang *et al.* 2022). An alternative approach is to enforce governing equations as hard constraints, which can enable deep learning models to precisely adhere to physical laws, as demonstrated by Beucler *et al.* (2021); Gao *et al.* (2021).

In §3, DL models and LSE approach are trained and tested with the dataset corresponding to open-channel flow at $Re_\tau = 250$. We expect the model to perform accurately for Re_τ values below 250; however, for Re_τ higher than 250, performance is likely to degrade due to the emergence of finer small-scale structures at higher Re_τ (Pope 2000). To address this, transfer learning–a technique that adapts a model trained on one task for a related task (Weiss *et al.* 2016)–could be employed to extend model capabilities to higher Re_τ (see, e.g., Guastoni *et al.* 2021; Yousif *et al.* 2022, 2023*a*,*b*). Instead of training a model from scratch for higher Re_τ, transfer learning allows the reuse of pre-learned features from a lower Re_τ case to accelerate the training and reduce data requirements for higher Re_τ cases (see, for example, Guastoni *et al.* 2021; Yousif *et al.* 2023*b*).

5. Conclusions

This study has proposed a PVAE model reconstructing 3-D flow fields from 2-D wall-parallel measurements in an open-channel flow at $Re_\tau = 250$. Physics-based constraints, including momentum and continuity equations, enstrophy, and boundary conditions, have been incorporated into the loss function and architecture of the DL model as soft and hard constraints to improve the model's performance. The reconstruction abilities of the PVAE have been compared against a corresponding black-box VAE (no physics constraints) and a more traditional LSE reconstruction method. Model assessment has focused on reconstructed instantaneous 3-D flow fields and coherent structures, velocity histograms, energy spectra, two-point correlations, one-dimensional first- and second-order flow statistics, and the residual with respect to the governing equations. This analysis focused on assessing the benefits of the proposed DL architecture over traditional methodologies as well as the impact of physical constraints on model accuracy and robustness.

A qualitative analysis based on visual inspection of the reconstructed instantaneous snapshots, histograms, and vorticity structures of velocity fields has indicated that PVAE model predictions are in remarkable agreement with reference DNS velocity, albeit with minor discrepancies. Visual inspection of the reconstructed instantaneous snapshots against the corresponding DNS results has also demonstrated that the model is in remarkable agreement with the reference solution and outperforms the classical LSE approach. A quantitative error analysis has indicated that the streamwise (u_1) velocity field is captured with higher accuracy when compared to the cross-stream (u_2) and wall-normal (u_3) components. As briefly mentioned in §4, one plausible reason is the use of a cumulative loss function. To mitigate this issue, $\mathcal{L}_{\text{data}}$ can be separated into three components, with β_j employed to balance individual term within $\mathcal{L}_{\text{data}}$.

Evaluation of energy spectra and two-point autocorrelation has further confirmed that the PVAE performs well in capturing large-scale flow structures, with minor discrepancies for high wavenumber modes. Further, aside from minor discrepancies in the reconstructed peak velocity RMSs, double-averaged flow statistics were also found to be in very good agreement with corresponding DNS data, demonstrating that the proposed model has learned the energetic scales of the flow—the main contributors to the mean and RMS velocity statistics. While the addition of physical constraints did not lead to apparent improvements in terms of large-scale features and double-averaged flow profiles, it did improve the ability of the model to capture small-scale structures and the physical realizability of the reconstructed flow fields (see figure 7). Notably, the inclusion of physical constraints reduced the residual on the momentum (continuity) equation by $\approx 56\%$ (41%). In terms of computational cost, introducing physical constraints leads to a modest 20% increase in the computational cost when using an equivalent number of training samples and epochs—a cost that may be justified in applications requiring accuracy in terms of physical realizability. Lastly, it has been shown that predictions from the PVAE model are insensitive to the sampling-plane location for the considered flow system. We note that this might not hold true at higher Reynolds numbers.

Overall, this study demonstrates that PVAE models can accurately reconstruct 3-D open channel flow at $Re_\tau = 250$ from 2-D wall-parallel measurements at arbitrary wall-normal distance from the surface, surpassing the performance of traditional LSE techniques. This capability is of interest to the engineering and geophysics communities, given the aforementioned challenges associated with performing dense measurements of 3-D turbulent flow fields in both laboratory and full-scale environments. The proposed formulation can also assist in compressing 3-D data into a convenient 2-D framework for data archival, yielding storage reduction. Although we focused on a specific case involving neutrally stratified turbulent open-channel flow, the proposed approach can be easily extended to more complex flow systems where complex surface morphologies, thermal stratification, and other flow physics are involved.

Acknowledgment. The authors acknowledge insightful suggestions and feedback from Dr. Carl Vondrick, which have helped refine the methodology of this study.

Funding. The authors acknowledge support from the Data Science Institute at Columbia University. Hora also acknowledges support from the Texas Advanced Computing Center via a Frontera Computational Science Fellowship. The work is also based upon work supported by of, or in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-22-1-0178. Giometto and Gentine acknowledge funding from the funding from NSF through the Learning the Earth with Artificial Intelligence and Physics (LEAP) Science and Technology Center (STC) (Award #2019625). This work also used the Anvil supercomputer at Purdue University through allocation ATM180022 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which National Science Foundation supports grants #2138259, #2138286, #2138307, #2137603, and #2138296.

Declaration of interests. The authors report no conflict of interest.

Data availability statement. Scripts and datasets generated as part of this study are available from the corresponding author on reasonable request.

Author ORCIDs.
G. S. Hora, https://orcid.org/0009-0004-7894-5175;
M. G. Giometto, https://orcid.org/0000-0001-9661-0599.

REFERENCES

Abadi, Martín, Barham, Paul, Chen, Jianmin, Chen, Zhifeng, Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Irving, Geoffrey, Isard, Michael & others 2016 {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pp. 265–283.

Adrian, Ronald J 2005 Twenty years of particle image velocimetry. *Experiments in fluids* **39**, 159–169.

Adrian, Ronald J & Moin, Parviz 1988 Stochastic estimation of organized turbulent structure: homogeneous shear flow. *Journal of Fluid Mechanics* **190**, 531–559.

Agarwal, Karuna, Ram, Omri, Wang, Jin, Lu, Yuhui & Katz, Joseph 2021 Reconstructing velocity and pressure from noisy sparse particle tracks using constrained cost minimization. *Experiments in Fluids* **62**, 1–20.

Albertson, John D & Parlange, Marc B 1999 Surface length scales and shear stress: Implications for land-atmosphere interaction over complex terrain. *Water Resources Research* **35**, 2121–2132.

Anderson, William, Li, Qi & Bou-Zeid, Elie 2015 Numerical simulation of flow over urban-like topographies and evaluation of turbulence temporal attributes. *Journal of Turbulence* **16**, 809–831.

Bau, David, Zhu, Jun-Yan, Wulff, Jonas, Peebles, William, Strobelt, Hendrik, Zhou, Bolei & Torralba, Antonio 2019 Seeing what a gan cannot generate. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4502–4511.

Bauweraerts, Pieter & Meyers, Johan 2021 Reconstruction of turbulent flow fields from lidar measurements using large-eddy simulation. *Journal of Fluid Mechanics* **906**, A17.

Beucler, Tom, Pritchard, Michael, Rasp, Stephan, Ott, Jordan, Baldi, Pierre & Gentine, Pierre 2021 Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters* **126**, 098302.

Biffi, Carlo, Cerrolaza, Juan J, Tarroni, Giacomo, de Marvao, Antonio, Cook, Stuart A, O'Regan, Declan P & Rueckert, Daniel 2019 3d high-resolution cardiac segmentation reconstruction from 2d views using conditional variational autoencoders. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1643–1646. IEEE.

Bou-Zeid, Elie, Meneveau, Charles & Parlange, Marc 2005 A scale-dependent lagrangian dynamic model for large eddy simulation of complex turbulent flows. *Physics of fluids* **17**, 025105.

Cai, Shengze, Wang, Zhicheng, Fuest, Frederik, Jeon, Young Jin, Gray, Callum & Karniadakis, George Em 2021 Flow over an espresso cup: inferring 3-d velocity and pressure fields from tomographic background oriented schlieren via physics-informed neural networks. *Journal of Fluid Mechanics* **915**, A102.

Canuto, Claudio, Hussaini, M Yousuff, Quarteroni, Alfio & Zang, Thomas A 2007 *Spectral methods: fundamentals in single domains*. Springer Science & Business Media.

Casola, Linda 2019 *Enhancing urban sustainability with data, modeling, and simulation: proceedings of a workshop*. National Academies Press.

Champagne, Francis Henry, Sleicher, CA & Wehrmann, OH 1967 Turbulence measurements with inclined hot-wires part 1. heat transfer experiments with inclined hot-wire. *Journal of Fluid Mechanics* **28** (1), 153–175.

Chandramouli, Pranav, Memin, Etienne, Heitz, Dominique & Fiabane, Lionel 2019 Fast 3d flow reconstructions from 2d cross-plane observations. *Experiments in Fluids* **60**, 1–21.

Cheikh, Mohamad Ibrahim & Momen, Mostafa 2020 The interacting effects of storm surge intensification and sea-level rise on coastal resiliency: a high-resolution turbulence resolving case study. *Environmental Research Communications* **2** (11), 115002.

Chen, Fei, Bornstein, Robert, Grimmond, Sue, Li, Ju, Liang, Xudong, Martilli, Alberto, Miao, Shiguang, Voogt, James & Wang, Yingchun 2012 Research priorities in observing and modeling urban weather and climate. *Bulletin of the American Meteorological Society* **93**, 1725–1728.

Cheng, Yu, Giometto, Marco G, Kauffmann, Pit, Lin, Ling, Cao, Chen, Zupnick, Cody, Li, Harold, Li, Qi, Huang, Yu, Abernathey, Ryan & others 2022 Deep learning for subgrid-scale turbulence modeling in large-eddy simulations of the convective atmospheric boundary layer. *Journal of Advances in Modeling Earth Systems* **14**, e2021MS002847.

Cheng, Zhengxue, Sun, Heming, Takeuchi, Masaru & Katto, Jiro 2018 Deep convolutional autoencoder-based lossy image compression. In *2018 Picture Coding Symposium (PCS)*, pp. 253–257. IEEE.

Chira, Darius, Haralampiev, Ilian, Winther, Ole, Dittadi, Andrea & Liévin, Valentin 2023 Image super-resolution with deep variational autoencoders. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pp. 395–411. Springer.

Chorin, Alexandre Joel 1968 Numerical solution of the navier-stokes equations. *Mathematics of computation* **22**, 745–762.

Choy, Christopher B, Xu, Danfei, Gwak, JunYoung, Chen, Kevin & Savarese, Silvio 2016 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *European conference on computer vision*, pp. 628–644. Springer.

Christen, Andreas, Meier, Fred & Scherer, Dieter 2012 High-frequency fluctuations of surface temperatures in an urban environment. *Theoretical and Applied Climatology* **108**, 301–324.

Christensen, Kenneth T & Adrian, Ronald J 2001 Statistical evidence of hairpin vortex packets in wall turbulence. *Journal of Fluid Mechanics* **431**, 433–443.

Chung, Daniel, Hutchins, Nicholas, Schultz, Michael P & Flack, Karen A 2021 Predicting the drag of rough surfaces. *Annual Review of Fluid Mechanics* **53**, 439–471.

Chung, D & Pullin, DI 2009 Large-eddy simulation and wall modelling of turbulent channel flow. *Journal of fluid mechanics* **631**, 281–309.

Clark Di Leoni, Patricio, Agarwal, Karuna, Zaki, Tamer A, Meneveau, Charles & Katz, Joseph 2023 Reconstructing turbulent velocity and pressure fields from under-resolved noisy particle tracks using physics-informed neural networks. *Experiments in Fluids* **64**, 95.

Cuomo, Salvatore, Di Cola, Vincenzo Schiano, Giampaolo, Fabio, Rozza, Gianluigi, Raissi, Maziar & Piccialli, Francesco 2022 Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing* **92** (3), 88.

Drygala, Claudia, Winhart, Benjamin, di Mare, Francesca & Gottschalk, Hanno 2022 Generative modeling of turbulence. *Physics of Fluids* **34** (3).

Dubief, Yves & Delcayre, Franck 2000 On coherent-vortex identification in turbulence. *Journal of turbulence* **1**, 011.

Elsinga, GE & Marusic, I 2010 Evolution and lifetimes of flow topology in a turbulent boundary layer. *Physics of Fluids* **22** (1).

Elsinga, Gerrit E, Scarano, Fulvio, Wieneke, Bernhard & van Oudheusden, Bas W 2006 Tomographic particle image velocimetry. *Experiments in fluids* **41** (6), 933–947.

Fan, Haoqiang, Su, Hao & Guibas, Leonidas J 2017 A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 605–613.

Fang, Jiannong & Porté-Agel, Fernando 2015 Large-eddy simulation of very-large-scale motions in the neutrally stratified atmospheric boundary layer. *Boundary-Layer Meteorology* **155**, 397–416.

Foken, Th & Wichura, Bodo 1996 Tools for quality assessment of surface-based flux measurements. *Agricultural and forest meteorology* **78** (1-2), 83–105.

Gal-Chen, Tzvi, Xu, Mei & Eberhard, Wynn L 1992 Estimations of atmospheric boundary layer fluxes and other turbulence parameters from doppler lidar data. *Journal of Geophysical Research: Atmospheres* **97** (D17), 18409–18423.

Ganapathisubramani, B, Lakshminarasimhan, K & Clemens, NT 2008 Investigation of three-dimensional structure of fine scales in a turbulent jet by using cinematographic stereoscopic particle image velocimetry. *Journal of fluid mechanics* **598**, 141–175.

Gao, Han, Sun, Luning & Wang, Jian-Xun 2021 Phygeonet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state pdes on irregular domain. *Journal of Computational Physics* **428**, 110079.

Giometto, Marco Giovanni, Christen, Andreas, Egli, Pascal Emanuel, Schmid, Manuel F, Tooke, RT, Coops, NC & Parlange, Marc B 2017 Effects of trees on mean wind, turbulence and momentum exchange within and above a real urban environment. *Advances in Water Resources* **106**, 154–168.

Giometto, Marco G, Christen, Andreas, Meneveau, Charles, Fang, Jiannong, Krafczyk, M & Parlange, Marc B 2016 Spatial characteristics of roughness sublayer mean flow and turbulence over a realistic urban surface. *Boundary-layer meteorology* **160**, 425–452.

Glaws, Andrew, King, Ryan & Sprague, Michael 2020 Deep learning for in situ data compression of large turbulent flow simulations. *Physical Review Fluids* **5**, 114602.

Glorot, Xavier & Bengio, Yoshua 2010 Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings.

Gong, Wanmin & Ibbetson, Alan 1989 A wind tunnel study of turbulent flow over model hills. *Boundary-Layer Meteorology* **49**, 113–148.

Goodfellow, Ian, Bengio, Yoshua & Courville, Aaron 2016 *Deep Learning*. MIT Press, http://www.deeplearningbook.org.

Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron & Bengio, Yoshua 2014 Generative adversarial nets. *Advances in neural information processing systems* **27**.

Grund, Christian J, Banta, Robert M, George, Joanne L, Howell, James N, Post, Madison J, Richter, Ronald A & Weickmann, Ann M 2001 High-resolution doppler lidar for boundary layer and cloud research. *Journal of Atmospheric and Oceanic Technology* **18** (3), 376–393.

Guastoni, Luca, Güemes, Alejandro, Ianiro, Andrea, Discetti, Stefano, Schlatter, Philipp, Azizpour, Hossein & Vinuesa, Ricardo 2021 Convolutional-network models to predict wall-bounded turbulence from wall quantities. *Journal of Fluid Mechanics* **928**, A27.

Gui, Jie, Sun, Zhenan, Wen, Yonggang, Tao, Dacheng & Ye, Jieping 2021 A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering* **35** (4), 3313–3332.

Han, Xian-Feng, Laga, Hamid & Bennamoun, Mohammed 2019 Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era. *IEEE transactions on pattern analysis and machine intelligence* **43**, 1578–1604.

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing & Sun, Jian 2016 Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Henzler, Philipp, Rasche, Volker, Ropinski, Timo & Ritschel, Tobias 2018 Single-image tomography: 3d volumes from 2d cranial x-rays. In *Computer Graphics Forum*, , vol. 37, pp. 377–388. Wiley Online Library.

Hochreiter, Sepp 1998 The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **6**, 107–116.

Hora, Gurpreet S & Giometto, Marco G 2024 Surrogate modeling of urban boundary layer flows. *Physics of Fluids* **36** (7).

Hultmark, Marcus, Calaf, Marc & Parlange, Marc B 2013 A new wall shear stress model for atmospheric boundary layer simulations. *Journal of the Atmospheric Sciences* **70**, 3460–3470.

Jeon, Joongoo, Lee, Juhyeong & Kim, Sung Joong 2022 Finite volume method network for the acceleration of unsteady computational fluid dynamics: Non-reacting and reacting flows. *International Journal of Energy Research* .

Joyce, James M 2011 Kullback-leibler divergence. In *International encyclopedia of statistical science*, pp. 720–722. Springer.

Kameyama, Emi, Saunders, Jennifer & Carrero-Martínez, Franklin 2020 Advancing urban sustainability in china and the united states: Proceedings of a workshop .

Kang, Yu-Eop, Yang, Sunwoong & Yee, Kwanjung 2022 Physics-aware reduced-order modeling of transonic flow via β-variational autoencoder. *Physics of Fluids* **34** (7).

Karniadakis, George Em, Kevrekidis, Ioannis G, Lu, Lu, Perdikaris, Paris, Wang, Sifan & Yang, Liu 2021 Physics-informed machine learning. *Nature Reviews Physics* **3**, 422–440.

Kim, Hyojin, Kim, Junhyuk, Won, Sungjin & Lee, Changhoon 2021 Unsupervised deep learning for super-resolution reconstruction of turbulence. *Journal of Fluid Mechanics* **910**, A29.

Kingma, DP & Welling, M 2014 Auto-encoding variational bayes. 2nd international conference on learning representations (iclr2014). *Preprint, submitted December* **23**, 2013.

Kingma, Diederik P & Ba, Jimmy 2014 Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Kochkov, Dmitrii, Smith, Jamie A, Alieva, Ayya, Wang, Qing, Brenner, Michael P & Hoyer, Stephan 2021 Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences* **118**, e2101784118.

Krayenhoff, E Scott, Jiang, Timothy, Christen, Andreas, Martilli, Alberto, Oke, Timothy R, Bailey, Brian N, Nazarian, Negin, Voogt, James A, Giometto, Marco G, Stastny, Austine & others 2020 A multi-layer urban canopy meteorological model with trees (bep-tree): Street tree impacts on pedestrian-level climate. *Urban Climate* **32**, 100590.

Kuo, Albert Yi-Shuong & Corrsin, Stanley 1971 Experiments on internal intermittency and fine-structure distribution functions in fully turbulent fluid. *Journal of Fluid Mechanics* **50** (2), 285–319.

Lee, Myoungkyu & Moser, Robert D 2015 Direct numerical simulation of turbulent channel flow up to. *Journal of fluid mechanics* **774**, 395–415.

Lee, Sangseung & You, Donghyun 2019 Data-driven prediction of unsteady flow over a circular cylinder using deep learning. *Journal of Fluid Mechanics* **879**, 217–254.

Lesieur, Marcel 1997 Turbulence in fluids kluwer academic publishers.

Lin, Ching-Long, Chai, Tianfeng & Sun, Juanzhen 2001 Retrieval of flow structures in a convective boundary layer using an adjoint model: Identical twin experiments. *Journal of the atmospheric sciences* **58** (13), 1767–1783.

Ling, Julia, Kurzawski, Andrew & Templeton, Jeremy 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics* **807**, 155–166.

Lippe, Phillip, Veeling, Bas, Perdikaris, Paris, Turner, Richard & Brandstetter, Johannes 2024 Pde-refiner: Achieving accurate long rollouts with neural pde solvers. *Advances in Neural Information Processing Systems* **36**.

Liu, Bo, Tang, Jiupeng, Huang, Haibo & Lu, Xi-Yun 2020*a* Deep learning methods for super-resolution reconstruction of turbulent flows. *Physics of Fluids* **32**, 025105.

Liu, Zhi-Song, Siu, Wan-Chi & Chan, Yui-Lam 2020*b* Photo-realistic image super-resolution via variational autoencoders. *IEEE Transactions on Circuits and Systems for video Technology* **31**, 1351–1365.

Liu, Zhi-Song, Siu, Wan-Chi, Wang, Li-Wen, Li, Chu-Tak & Cani, Marie-Paule 2020*c* Unsupervised real image super-resolution via generative variational autoencoder. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 442–443.

Lu, Hao & Porté-Agel, Fernando 2010 A modulated gradient model for large-eddy simulation: application to a neutral atmospheric boundary layer. *Physics of Fluids* **22**, 015109.

Meneveau, Charles, Lund, Thomas S & Cabot, William H 1996 A lagrangian dynamic subgrid-scale model of turbulence. *Journal of fluid mechanics* **319**, 353–385.

Menzies, Robert T & Hardesty, R Michael 1989 Coherent doppler lidar for measurements of wind fields. *Proceedings of the IEEE* **77** (3), 449–462.

Mescheder, Lars, Nowozin, Sebastian & Geiger, Andreas 2017 The numerics of gans. *Advances in neural information processing systems* **30**.

Mildenhall, Ben, Srinivasan, Pratul P, Tancik, Matthew, Barron, Jonathan T, Ramamoorthi, Ravi & Ng, Ren 2021 Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65** (1), 99–106.

Mochida, Akashi & Lun, Isaac YF 2008 Prediction of wind environment and thermal comfort at pedestrian level in urban area. *Journal of wind engineering and industrial aerodynamics* **96**, 1498–1527.

Moser, Robert D, Kim, John & Mansour, Nagi N 1999 Direct numerical simulation of turbulent channel flow up to reτ= 590. *Physics of fluids* **11** (4), 943–945.

Murray, Nathan E & Ukeiley, Lawrence S 2003 Estimation of the flowfield from surface pressure measurements in an open cavity. *AIAA journal* **41**, 969–972.

Nagib, Hassan M & Chauhan, Kapil A 2008 Variations of von kármán coefficient in canonical flows. *Physics of fluids* **20** (10).

Nair, Vinod & Hinton, Geoffrey E 2010 Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.

Orszag, Steven A 1969 Numerical methods for the simulation of turbulence. *The Physics of Fluids* **12**, II–250.

Pan, Ying, Follett, Elizabeth, Chamecki, Marcelo & Nepf, Heidi 2014 Strong and weak, unsteady reconfiguration and its impact on turbulence structure within plant canopies. *Physics of Fluids* **26**, 2003–2017.

Pavlík, David, Uruba, Václav & Kopeckỳ, Václav 2017 Reconstruction of 3d piv data in complicated experimental arrangements. In *EPJ Web of Conferences*, , vol. 143, p. 02090. EDP Sciences.

Podvin, Bérengre, Fraigneau, Yann, Lusseyran, François & Gougat, Pierre 2006 A reconstruction method for the flow past an open cavity. *Journal of Fluids Engineering* .

Pope, Stephen B 2000 *Turbulent flows*. Cambridge university press.

Porté-Agel, Fernando 2004 A scale-dependent dynamic model for scalar transport in large-eddy simulations of the atmospheric boundary layer. *Boundary-Layer Meteorology* **112**, 81–105.

Porté-Agel, Fernando, Meneveau, Charles & Parlange, Marc B 2000 A scale-dependent dynamic model for large-eddy simulation: application to a neutral atmospheric boundary layer. *Journal of Fluid Mechanics* **415**, 261–284.

Prasad, Ajay K 2000 Stereoscopic particle image velocimetry. *Experiments in fluids* **29** (2), 103–116.

Raissi, Maziar, Babaee, Hessam & Givi, Peyman 2019*a* Deep learning of turbulent scalar mixing. *Physical Review Fluids* **4** (12), 124501.

Raissi, Maziar, Perdikaris, Paris & Karniadakis, George E 2019*b* Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* **378**, 686–707.

Ramachandran, Prajit, Zoph, Barret & Le, Quoc V 2017 Searching for activation functions. *arXiv preprint arXiv:1710.05941* .

Raupach, MR, Thom, AS & Edwards, I 1980 A wind-tunnel study of turbulent flow close to regularly arrayed rough surfaces. *Boundary-Layer Meteorology* **18**, 373–397.

Rezende, Danilo Jimenez, Mohamed, Shakir & Wierstra, Daan 2014 Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pp. 1278–1286. PMLR.

Robinson, Stephen K 1991 Coherent motions in the turbulent boundary layer. *Annual review of fluid mechanics* **23**, 601–639.

Rotach, MW, Vogt, Roland, Bernhofer, Christian, Batchvarova, Ekaterina, Christen, Andreas, Clappier, Alain, Feddersen, B, Gryning, S-E, Martucci, G, Mayer, Helmut & others 2005 Bubble–an urban boundary layer meteorology project. *Theoretical and Applied Climatology* **81**, 231–261.

Salimans, Tim, Goodfellow, Ian, Zaremba, Wojciech, Cheung, Vicki, Radford, Alec & Chen, Xi 2016 Improved techniques for training gans. *Advances in neural information processing systems* **29**.

Scott, Dennis O, Sorenson, Lori R, Steele, Karla L, Puckett, DeAnna L & Lunte, Craig E

1991 In vivo microdialysis sampling for pharmacokinetic investigations. *Pharmaceutical research* **8**, 389–392.

Scotti, Alberto, Meneveau, Charles & Lilly, Douglas K 1993 Generalized smagorinsky model for anisotropic grids. *Physics of Fluids A: Fluid Dynamics* **5** (9), 2306–2308.

Seo, Jeonghwa, Seol, Dong Myung, Han, Bumwoo & Rhee, Shin Hyung 2016 Turbulent wake field reconstruction of vlcc models using two-dimensional towed underwater piv measurements. *Ocean engineering* **118**, 28–40.

Shah, Stimit & Bou-Zeid, Elie 2014 Very-large-scale motions in the atmospheric boundary layer educed by snapshot proper orthogonal decomposition. *Boundary-Layer Meteorology* **153**, 355–387.

Simonyan, Karen & Zisserman, Andrew 2014 Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* .

Sinn, Mathieu & Rawat, Ambrish 2018 Non-parametric estimation of jensen-shannon divergence in generative adversarial network training. In *International Conference on Artificial Intelligence and Statistics*, pp. 642–651. PMLR.

Skamarock, WC, Klemp, JB, Dudhia, J, Gill, DO, Barker, DM, Wang, W & Powers, JG 2008 A description of the advanced research wrf version 3 (national center for atmospheric research, boulder, co). *Tech. Rep.*. NCAR Technical Note NCAR/TN-4751STR.

Solera-Rico, Alberto, Sanmiguel Vila, Carlos, Gómez-López, Miguel, Wang, Yuning, Almashjary, Abdulrahman, Dawson, Scott TM & Vinuesa, Ricardo 2024 β-variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nature Communications* **15** (1), 1361.

Stengel, Karen, Glaws, Andrew, Hettinger, Dylan & King, Ryan N 2020 Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences* **117** (29), 16805–16815.

Subramaniam, Akshay, Wong, Man Long, Borker, Raunak D, Nimmagadda, Sravya & Lele, Sanjiva K 2020 Turbulence enrichment using physics-informed generative adversarial networks. *arXiv preprint arXiv:2003.01907* .

Suzuki, Takao & Hasegawa, Yosuke 2017 Estimation of turbulent channel flow at based on the wall measurement using a simple sequential approach. *Journal of Fluid Mechanics* **830**, 760–796.

Tahir, Rohan, Sargano, Allah Bux & Habib, Zulfiqar 2021 Voxel-based 3d object reconstruction from single 2d image using variational autoencoders. *Mathematics* **9**, 2288.

Theis, Lucas, Shi, Wenzhe, Cunningham, Andrew & Huszár, Ferenc 2022 Lossy image compression with compressive autoencoders. In *International conference on learning representations*.

Thomas, Christoph K, Kennedy, Adam M, Selker, John S, Moretti, Ayla, Schroth, Martin H, Smoot, Alexander R, Tufillaro, Nicholas B & Zeeman, Matthias J 2012 High-resolution fibre-optic temperature sensing: A new tool to study the two-dimensional structure of atmospheric surface-layer flow. *Boundary-layer meteorology* **142**, 177–192.

Toparlar, Yasin, Blocken, Bert, Vos, P v, Van Heijst, GJF, Janssen, WD, van Hooff, Twan, Montazeri, Hamid & Timmermans, HJP 2015 Cfd simulation and validation of urban microclimate: A case study for bergpolder zuid, rotterdam. *Building and environment* **83**, 79–90.

Tucsok, Matthew, Gazani, Sara Hatami, Gupta, Kashish & Najjaran, Homayoun 2022 3d reconstruction from 2d images: A two-part autoencoder-like tool. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 538–543. IEEE.

Tulsiani, Shubham, Zhou, Tinghui, Efros, Alexei A & Malik, Jitendra 2017 Multi-view supervision for single-view reconstruction via differentiable ray consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2626–2634.

Van Doorne, CWH & Westerweel, J 2007 Measurement of laminar, transitional and turbulent pipe flow using stereoscopic-piv. *Experiments in Fluids* **42**, 259–279.

Vétel, Jérôme, Garon, André & Pelletier, Dominique 2010 Vortex identification methods based on temporal signal-processing of time-resolved piv data. *Experiments in fluids* **48**, 441–459.

Wang, Chengyue, Gao, Qi, Wang, Biao, Pan, Chong & Wang, Jinjun 2021*a* Vortex-to-velocity reconstruction for wall-bounded turbulence via the field-based linear stochastic estimation. *Journal of Fluid Mechanics* **922**, A18.

Wang, Jing, He, Cheng, Li, Runze, Chen, Haixin, Zhai, Chen & Zhang, Miao 2021*b* Flow field prediction of supercritical airfoils via variational autoencoder based deep learning framework. *Physics of Fluids* **33** (8).

Wang, Sifan, Teng, Yujun & Perdikaris, Paris 2021*c* Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing* **43**, A3055–A3081.

Wei, Ruoqi & Mahmood, Ausif 2020 Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey. *Ieee Access* **9**, 4939–4956.

Weiss, Karl, Khoshgoftaar, Taghi M & Wang, DingDing 2016 A survey of transfer learning. *Journal of Big data* **3**, 1–40.

Westerweel, Jerry, Elsinga, Gerrit E & Adrian, Ronald J 2013 Particle image velocimetry for complex and turbulent flows. *Annual Review of Fluid Mechanics* **45**, 409–436.

Xiang, Zixue, Peng, Wei, Liu, Xu & Yao, Wen 2022 Self-adaptive loss balanced physics-informed neural networks. *Neurocomputing* **496**, 11–34.

Xu, Zhi-Qin John, Zhang, Yaoyu & Xiao, Yanyang 2019 Training behavior of deep neural network in frequency domain. In *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part I 26*, pp. 264–274. Springer.

Xuan, Anqing & Shen, Lian 2023 Reconstruction of three-dimensional turbulent flow structures using surface measurements for free-surface flows based on a convolutional neural network. *Journal of Fluid Mechanics* **959**, A34.

Yousefi, Kianoosh, Hora, Gurpreet S, Yang, Hongshuo & Giometto, Marco 2024*a* Data-driven met-ocean model for offshore wind energy applications. In *Journal of Physics: Conference Series*, , vol. 2767, p. 052005. IOP Publishing.

Yousefi, Kianoosh, Hora, Gurpreet Singh, Yang, Hongshuo, Veron, Fabrice & Giometto, Marco G 2024*b* A machine learning model for reconstructing skin-friction drag over ocean surface waves. *Journal of Fluid Mechanics* **983**, A9.

Yousif, Mustafa Z, Yu, Linqi, Hoyas, Sergio, Vinuesa, Ricardo & Lim, HeeChang 2023*a* A deep-learning approach for reconstructing 3d turbulent flows from 2d observation data. *Scientific Reports* **13**, 2529.

Yousif, Mustafa Z, Yu, Linqi & Lim, HeeChang 2022 Physics-guided deep learning for generating turbulent inflow conditions. *Journal of Fluid Mechanics* **936**, A21.

Yousif, Mustafa Z, Zhang, Meng, Yu, Linqi, Vinuesa, Ricardo & Lim, HeeChang 2023*b* A transformer-based synthetic-inflow generator for spatially developing turbulent boundary layers. *Journal of Fluid Mechanics* **957**, A6.

Yu, Alex, Ye, Vickie, Tancik, Matthew & Kanazawa, Angjoo 2021 pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4578–4587.

Zhang, Fan, Hu, Haibao, Ren, Feng, Zhang, Heng & Du, Peng 2022 Reconstructing turbulent velocity information for arbitrarily gappy flow fields using the deep convolutional neural network. *Physics of Fluids* **34**, 127117.

Zhou, Jigen, Adrian, Ronald J, Balachandar, Sivaramakrishnan & Kendall, TM1693393 1999 Mechanisms for generating coherent packets of hairpin vortices in channel flow. *Journal of fluid mechanics* **387**, 353–396.