14th International Symposium on Turbulence and Shear Flow Phenomena (TSFP14)
Heidelberg, Germany, July 28–31, 2026

RECONSTRUCTING TURBULENT CHANNEL FLOW FIELDS WITH A MACHINE-LEARNED PRIOR DISTRIBUTION

Frederik Aerts
Mechanical Engineering
KU Leuven
Leuven, Belgium
frederik.aerts1@kuleuven.be

Dirk Nuyens
Computer Science
KU Leuven
Leuven, Belgium
dirk.nuyens@kuleuven.be

Johan Meyers
Mechanical Engineering
KU Leuven
Leuven, Belgium
johan.meyers@kuleuven.be

ABSTRACT

When turbulent flow reconstruction is regarded as a stochastic estimation problem, the reconstructed flow field must follow the conditional distribution of the field given the observations – or the posterior distribution in Bayesian terminology. Although classical methods, such as linear stochastic estimation (Adrian *et al.*, 1987), can robustly reconstruct ensemble-averaged conditional turbulent structures, recent developments in probabilistic machine learning may enable the reconstruction of realistic instantaneous conditional structures. In this work, we train a machine-learned prior distribution with a flow-based generative model to reconstruct such three-dimensional structures in a turbulent channel flow. We show that the ensemble statistics of the spatially three-dimensional flow fields sampled from the machine-learned prior closely resemble those of a direct numerical simulation. By comparing the reconstruction error to an estimated expected reconstruction error based on linear stochastic estimation, the reconstructed fields are shown to be largely consistent with the observations. Moreover, the reconstructed flow fields preserve the correct ensemble statistics in the unobserved region, indicating that the machine-learned prior can reconstruct realistic instantaneous conditional turbulent structures.

INTRODUCTION

Reconstructing velocity fluctuations from limited observations remains a challenging task due to the intricate nature of high-dimensional turbulence (Arun *et al.*, 2023). One of the first approaches to address this challenge was linear stochastic estimation (Adrian *et al.*, 1987), which reconstructs the conditional average of the flow field given the observations, with a procedure that is numerically equivalent to approximating the conditional probability distribution with a Gaussian. The conditional average of the flow field is optimal in terms of mean-square reconstruction error. However, the resulting so-called 'conditional eddies' do not necessarily capture the instantaneous structures present in the actual flow field, as they represent simply the average field associated with the observed event (Adrian, 1996).

Probabilistic machine learning models, such as Generative Adversarial Networks (GANs) and diffusion models, have shown great promise in capturing instantaneous turbulent flow fields, including their non-Gaussian features. They have for example been applied to the generation and reconstruction of 2D slices of turbulent flow fields (Li *et al.*, 2023*a*,*b*; Vishwasrao *et al.*, 2025), and to reconstructing 3D slices from 2D observations (e.g. Yousif *et al.*, 2023) or other observational setups that were prespecified during training (e.g. Rybchuk *et al.*, 2023). More recently, diffusion models have been used to reconstruct spatio-temporal (Du *et al.*, 2024) and spatially three-dimensional turbulence (Amorós-Trepat *et al.*, 2026) in arbitrary observational setups without requiring to retrain the model. When comparing GANs and diffusion models for turbulent flow reconstruction, Li *et al.* (2023*b*) not only showed that diffusion models outperform GANs, but also that there exists a trade-off between obtaining a reliable mean-square error and good statistical reconstruction properties. However, when probabilistic machine learning models are used to reconstruct instantaneous realizations of turbulent flow fields that are consistent with the observations, the mean-square error with the true field should in fact be interpreted probabilistically.

In this work, we examine the use of flow-based generative models (Lipman *et al.*, 2024) to represent the probability distribution of (discretized) instantaneous 3D turbulent flow fields and reconstruct them in arbitrary observational setups, without requiring to retrain the model. To this end, we consider a turbulent channel flow at a friction Reynolds number of $\mathrm{Re}_\tau = 180$. Although flow-based generative models have recently been adopted for reconstructing 2D slices of turbulent channel flow fields from wall measurements (Parikh *et al.*, 2026), we apply it – to the best of our knowledge – for the first time to the generation and reconstruction of spatially 3D fields. Specifically, we address three questions: (1) how accurate does the (unconditional) prior represent ensemble statistics of the turbulent channel flow fields, (2) how does the reconstruction error relate to the expected mean-square error when interpreted probabilistically, and (3) how accurate are the ensemble statistics of the unobserved parts of the reconstructed velocity fields.

The contribution is organized as follows. First, we present the general flow reconstruction problem and derive the expected reconstruction error for different kinds of stochastic estimators. Then, the machine-learned distribution is introduced, together with an algorithm to conditionally sample it. After outlining the setup of the turbulent channel flow reconstruction problem, the three questions from above are addressed.

FLOW RECONSTRUCTION

Similar to Adrian *et al.* (1987), our aim is to reconstruct the turbulent structures in the limit of negligible measurement error. Before introducing the machine-learned distribution, the reconstruction problem at hand is presented, as well as the corresponding linear stochastic estimator, and an estimate for the expected obtainable reconstruction error.

Reconstruction problem

Consider observations $\boldsymbol{y} = \mathbf{M}\boldsymbol{u} + \boldsymbol{\varepsilon}$ of a discretized velocity field $\boldsymbol{u} \in \mathbb{R}^n$, where $\mathbf{M} \in \mathbb{R}^{p\times n}$ is a mask matrix that contains p rows of the identity matrix, and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I}_p)$ represents the measurement noise with variance σ_ε^2. The full velocity field $\boldsymbol{u}$ may be reconstructed from the observations $\boldsymbol{y}$ through Bayesian updating. In that case, a prior distribution of the velocity field $\pi(\boldsymbol{u})$ is updated, after observing $\boldsymbol{y}$, to a posterior distribution $p(\boldsymbol{u}|\boldsymbol{y})$ using Bayes' theorem

$$p(\boldsymbol{u}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{u})\pi(\boldsymbol{u}) \tag{1}$$

where $p(\boldsymbol{y}|\boldsymbol{u})$ is the likelihood of the observations $\boldsymbol{y}$ given the velocity field $\boldsymbol{u}$. In the limit of vanishing measurement noise, the likelihood becomes the Dirac delta distribution $p(\boldsymbol{y}|\boldsymbol{u}) = \delta(\boldsymbol{y} - \mathbf{M}\boldsymbol{u})$, such that the posterior equals the conditional prior $\pi(\boldsymbol{u}_u|\boldsymbol{u}_o = \boldsymbol{y})$, where $\boldsymbol{u}_o := \mathbf{M}\boldsymbol{u}$ is the observed part of the velocity field, $\boldsymbol{u}_u := \mathbf{M}^\perp \boldsymbol{u}$ is the unobserved part of the velocity field, and the complementary mask $\mathbf{M}^\perp \in \mathbb{R}^{(n-p)\times n}$ contains the remaining $n - p$ rows of the identity matrix. Hence, the aims are to adequately represent the prior distribution $\pi(\boldsymbol{u})$ and to efficiently sample it conditionally.

Linear stochastic estimation

Linear Stochastic Estimation (LSE) approximates the conditional average of the velocity field given the observations $\langle \boldsymbol{u}_u | \boldsymbol{u}_o = \boldsymbol{y}\rangle$, by expanding it as a Taylor series around $\boldsymbol{y} = 0$ and only keeping the first order terms (Adrian, 1996). Numerically, the linear stochastic estimator is equivalent to the posterior mean when presuming a Gaussian prior on the velocity field. When ordering the state vector as $\boldsymbol{u} = [\boldsymbol{u}_u, \boldsymbol{u}_o]^\top$, such a Gaussian prior is given by

$$\pi_{\mathrm{LSE}}\left(\begin{bmatrix}\boldsymbol{u}_u\\ \boldsymbol{u}_o\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\langle\boldsymbol{u}_u\rangle\\ \langle\boldsymbol{u}_o\rangle\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{uu} & \boldsymbol{\Sigma}_{uo}\\ \boldsymbol{\Sigma}_{ou} & \boldsymbol{\Sigma}_{oo}\end{bmatrix}\right), \tag{2}$$

where the covariance matrices are defined as $\boldsymbol{\Sigma}_{uo} = \langle(\boldsymbol{u}_u - \langle\boldsymbol{u}_u\rangle)(\boldsymbol{u}_o - \langle\boldsymbol{u}_o\rangle)^\top\rangle$ and $\langle\cdot\rangle$ represents an ensemble average. In that case, the conditional prior distribution (or posterior) is given by

$$\pi_{\mathrm{LSE}}(\boldsymbol{u}_u \mid \boldsymbol{u}_o = \boldsymbol{y}) = \mathcal{N}\Big(\boldsymbol{u}_u; \boldsymbol{\mu}_{u|o}(\boldsymbol{y}), \boldsymbol{\Sigma}_{u|o}\Big), \tag{3a}$$

$$\boldsymbol{\mu}_{u|o}(\boldsymbol{y}) = \langle\boldsymbol{u}_u\rangle + \boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}(\boldsymbol{y} - \langle\boldsymbol{u}_o\rangle), \tag{3b}$$

$$\boldsymbol{\Sigma}_{u|o} = \boldsymbol{\Sigma}_{uu} - \boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{ou}. \tag{3c}$$

When only the velocity fluctuations $\boldsymbol{u}' = \boldsymbol{u} - \langle\boldsymbol{u}\rangle$ are to be reconstructed, the classical LSE reconstruction operator is retrieved as $\langle \boldsymbol{u}'_u \mid \boldsymbol{u}'_o = \boldsymbol{y}'\rangle = \boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{y}'$. In principle, the Gaussian posterior could also be sampled to yield an ensemble of reconstructed velocity fields with instantaneous structures that are consistent with the observations. However, essential turbulent features such as spectral energy transfer and vortex stretching cannot be represented by Gaussian random fields – necessitating the use of a more realistic prior for posterior sampling.

Expected reconstruction error

The mean-square reconstruction error for a reconstruction algorithm based on posterior sampling can be expressed in terms of the first- and second-order ensemble statistics of the true and reconstructed fields. The true unobserved part of the velocity field $\boldsymbol{u}_{u|o}$ necessarily follows the true conditional distribution $\pi_{true}(\boldsymbol{u}_u \mid \boldsymbol{u}_o = \boldsymbol{y})$ with an ensemble mean $\langle\boldsymbol{u}_{u|o}\rangle(\boldsymbol{y})$ and covariance $\langle\boldsymbol{u}'_{u|o}\boldsymbol{u}'^{\top}_{u|o}\rangle(\boldsymbol{y})$ that can generally depend on $\boldsymbol{y}$. Let the reconstruction $\tilde{\boldsymbol{u}}_{u|o}(\boldsymbol{y})$ be drawn independently from a distribution with mean $\langle\tilde{\boldsymbol{u}}_{u|o}\rangle(\boldsymbol{y})$ and covariance $\langle\tilde{\boldsymbol{u}}'_{u|o}\tilde{\boldsymbol{u}}'^{\top}_{u|o}\rangle(\boldsymbol{y})$. Then the mean-square error is given by

$$\mathrm{MSE} = \mathbb{E}_y\left[\langle\|\boldsymbol{u}_{u|o} - \tilde{\boldsymbol{u}}_{u|o}(\boldsymbol{y})\|_2^2\rangle\right], \tag{4a}$$

$$= \mathbb{E}_y\Big[\mathrm{Tr}\langle\boldsymbol{u}'_{u|o}\boldsymbol{u}'^{\top}_{u|o}\rangle(\boldsymbol{y}) + \mathrm{Tr}\langle\tilde{\boldsymbol{u}}'_{u|o}\tilde{\boldsymbol{u}}'^{\top}_{u|o}\rangle(\boldsymbol{y}) + \|\langle\boldsymbol{u}_{u|o}\rangle(\boldsymbol{y}) - \langle\tilde{\boldsymbol{u}}_{u|o}\rangle(\boldsymbol{y})\|_2^2\Big]. \tag{4b}$$

From this expression, it is seen that the expected MSE is obtained by taking the conditional mean, i.e., $\tilde{\boldsymbol{u}}_{u|o} = \langle\boldsymbol{u}_{u|o}\rangle(\boldsymbol{y})$, as stated before.

To reconstruct realistic instantaneous structures that are consistent with the observations, we should ideally sample the reconstruction from the true conditional prior distribution. In that case, the expected MSE equals twice the trace of the conditional covariance. If the reconstruction is taken as the mean of an ensemble of M samples of the true conditional distribution, its covariance is given by $\langle\boldsymbol{u}'_{u|o}\boldsymbol{u}'^{\top}_{u|o}\rangle(\boldsymbol{y})/M$, such that the expected MSE equals

$$\mathrm{MSE}_{true}(M) = \left(1 + \frac{1}{M}\right)\mathrm{Tr}\left(\mathbb{E}_{\boldsymbol{y}}\left[\langle\boldsymbol{u}'_{u|o}\boldsymbol{u}'^{\top}_{u|o}\rangle(\boldsymbol{y})\right]\right). \tag{5}$$

This formula recovers both the MSE of a single instantaneous reconstruction, sampled from the true conditional distribution, for $M = 1$ and the MSE of the conditional mean for $M \to \infty$.

The true conditional covariance depends on the observation $\boldsymbol{y}$, through the conditional mean $\langle\boldsymbol{u}_{u|o}\rangle(\boldsymbol{y})$, which is unknown. Because we typically only have one realization of the field in the unobserved region per realization of the observed region available, we cannot compute $\langle\boldsymbol{u}_{u|o}\rangle(\boldsymbol{y})$ from data. Therefore, we will approximate it with the best linear estimator, as given by LSE. In that case, the MSE is approximately given by

$$\mathrm{MSE}_{true}(M) \approx \left(1 + \frac{1}{M}\right)\mathrm{Tr}\left(\boldsymbol{\Sigma}_{u|o}\right), \tag{6}$$

which is exact for truly Gaussian fields. For illustrative purposes, consider $u_u, u_o \in \mathbb{R}$ to be jointly Gaussian distributed with equal variance σ^2 and correlation ρ. In that case, the MSE is given by $(1 + 1/M)\sigma^2(1 - \rho^2)$, which goes to zero if the observed and unobserved part are perfectly correlated. Hence, the spatial correlation plays a vital role in the evolution of the reconstruction error.

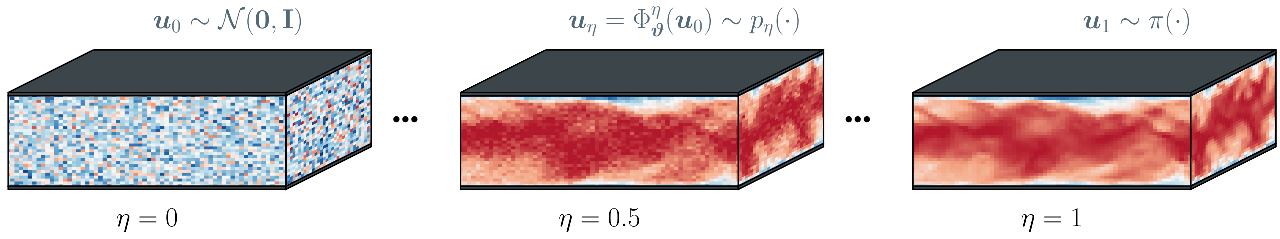

Figure 1. Schematic representation of the sampling process with the flow-based generative model: the flow $\boldsymbol{u}_\eta = \Phi^\eta_{\boldsymbol{\vartheta}}(\boldsymbol{u}_0)$ deterministically maps an initial condition $\boldsymbol{u}_0$, sampled from a standard Gaussian, to a turbulent channel flow field $\boldsymbol{u}_1$ at fictitious time $\eta = 1$.

MACHINE-LEARNED PRIOR

To reconstruct realistic instantaneous turbulent velocity fields from partial observations, we train a flow-based generative model in order to represent more a realistic prior distribution. The remainder of this section outlines the learned prior, and a new method to conditionally sample it.

Flow-based generative model

Flow-based generative models learn a transformation of a prespecified probability distribution, typically a standard Gaussian, to a target probability distribution – in our case the prior distribution of the velocity field $\pi(\boldsymbol{u})$. To do so, they define a set of Ordinary Differential Equations (ODEs) for a state $\boldsymbol{u}_\eta \in \mathbb{R}^n$ that depends on a fictitious time coordinate η:

$$\frac{\mathrm{d}\boldsymbol{u}_\eta}{\mathrm{d}\eta} = \boldsymbol{f}_{\boldsymbol{\vartheta}}(\boldsymbol{u}_\eta, \eta), \tag{7}$$

through an η-dependent generator $\boldsymbol{f}_{\boldsymbol{\vartheta}}(\boldsymbol{u}_\eta, \eta)$, represented by a neural network with parameters $\boldsymbol{\vartheta}$. The flow $\boldsymbol{u}_\eta = \Phi^\eta_{\boldsymbol{\vartheta}}(\boldsymbol{u}_0)$, defined by the ODEs, deterministically maps an initial condition $\boldsymbol{u}_0$ to its state at time η. For an ensemble of initial conditions $\boldsymbol{u}_0^{(k)}$ sampled from a distribution $p_0(\cdot)$, the flow $\boldsymbol{u}_\eta^{(k)} = \Phi^\eta_{\boldsymbol{\vartheta}}(\boldsymbol{u}_0^{(k)})$ generates a probability path p_η with $\boldsymbol{u}_\eta^{(k)} \sim p_\eta(\cdot)$. Hence, the goal is to learn a generator $\boldsymbol{f}_{\boldsymbol{\vartheta}}(\boldsymbol{u}_\eta, \eta)$, typically defined over an interval $\eta \in [0, 1]$, so that $p_1(\boldsymbol{u}) \approx \pi(\boldsymbol{u})$.

The generator is learned by minimizing the conditional flow matching loss for a conditional optimal transport path (Lipman *et al.*, 2024), which linearly interpolates between noise and turbulent channel flow fields in this case. We use the U-Net architecture implemented by OpenAI (Dhariwal & Nichol, 2021) to parametrize the generator. Once training is done, the flow-based generative model can produce samples that approximately follow the prior distribution $\pi(\boldsymbol{u})$. This is achieved by drawing initial conditions $\boldsymbol{u}_0^{(k)} \sim p_0(\cdot)$ from the base distribution, and integrating the ODEs defined in Eq. (7) using a suitable numerical time integration scheme. This procedure is depicted schematically in Figure 1. In what follows, we employ a fourth-order Runge-Kutta scheme with a total of 20 time steps.

Reconstruction algorithm

To sample turbulent velocity fields from the conditional prior instead of from the full prior, the generator must be adapted such that the reconstructed fields equal the observations in the observed part of the field and are consistent with the observations in the unobserved part. Inspired by the heuristic RePaint algorithm for image inpainting with diffusion models by Lugmayr *et al.* (2022), we adapt the generation process in a manner that can be motivated as follows. We introduce a Markov process with the same fictitious time coordinate η, discretized as $\eta_n = \frac{n}{N}$, that yields the posterior (or conditional prior) distribution at $\eta_N = 1$:

$$\begin{aligned} p(\boldsymbol{u}_N|\boldsymbol{y}) = \int \cdots \int \prod_{n=0}^{N-1} p(\boldsymbol{u}_{u,n+1}|\boldsymbol{u}_{o,n+1}, \boldsymbol{u}_n, \boldsymbol{y}) \\ \times p(\boldsymbol{u}_{o,n+1}|\boldsymbol{u}_n, \boldsymbol{y})\, p(\boldsymbol{u}_0|\boldsymbol{y})\, \mathrm{d}\boldsymbol{u}_0 \cdots \mathrm{d}\boldsymbol{u}_{N-1}. \end{aligned} \tag{8}$$

Since the flow field in the observed region $\boldsymbol{u}_{o,N}$ must match the measurements $\boldsymbol{y}$ exactly, we prescribe the same linearly interpolating process as on which the model was trained to the unobserved part of the field $p(\boldsymbol{u}_{o,n}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{u}_{o,n}; \eta_n \boldsymbol{y}, (1-\eta_n)^2 \mathbf{I})$, which defines $p(\boldsymbol{u}_{o,n+1}|\boldsymbol{u}_n, \boldsymbol{y})$ and ensures that $p(\boldsymbol{u}_{o,N}|\boldsymbol{y}) = \delta(\boldsymbol{u}_{o,N} - \boldsymbol{y})$ and $p(\boldsymbol{u}_{o,0}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{u}_{o,0}; \mathbf{0}, \mathrm{I})$. The best we can do for $p(\boldsymbol{u}_{u,n+1}|\boldsymbol{u}_{o,n+1}, \boldsymbol{u}_n, \boldsymbol{y})$ is to replace it with $p(\boldsymbol{u}_{u,n+1}|\boldsymbol{u}_n)$, which is given by the discretized flow-based generative model, with $p(\boldsymbol{u}_{u,0}) = \mathcal{N}(\boldsymbol{u}_{u,0}; \mathbf{0}, \mathbf{I})$.

The approximated conditional prior can now be sampled hierarchically through Eq. (8), starting from $\boldsymbol{u}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. At each fictitious timestep η_n with $n = 0 \ldots N-1$, we compute

$$\boldsymbol{u}_{o,n+1} = \frac{n+1}{N}(\boldsymbol{y} - \boldsymbol{u}_{o,0}) + \boldsymbol{u}_{o,0}, \tag{9a}$$

$$\boldsymbol{u}_{u,n+1} = \boldsymbol{u}_{u,n+1} + \mathbf{M}^\perp \boldsymbol{f}_{\boldsymbol{\vartheta}}(\boldsymbol{u}_n, \eta_n)/N, \tag{9b}$$

$$\boldsymbol{u}_{n+1} = \mathrm{concat}(\boldsymbol{u}_{u,n+1}, \boldsymbol{u}_{o,n+1}). \tag{9c}$$

Consequently, an ensemble of instantaneous velocity field reconstructions given the observations $\boldsymbol{y}$ is obtained by starting from different $\boldsymbol{u}_0 \sim \mathcal{N}(0, \mathbf{I})$ and simulating the ODEs of Eq. (9). Since $\boldsymbol{u}_{o,N} = \boldsymbol{y}$, all reconstructed fields exactly match the observations for the observed part. In what follows, we use $N = 400$ steps when sampling the conditional machine-learned prior.

CASE SETUP

The methodology is applied to a turbulent channel flow, based on a collection of snapshots from a Direct Numerical Simulation (DNS). The downstream half of the channel is masked and should be reconstructed from the observed upstream half as depicted in Figure 2. Since the velocity field in the downstream unobserved block decorrelates from the observed region when going farther in the unobserved region, the reconstruction error is expected to increase gradually with the distance from the observed region – rendering this observational setup well-suited to examine the consistency of the observed and reconstructed regions. The remainder of this section details the DNS and the numerical procedure required to estimate the expected reconstruction error.

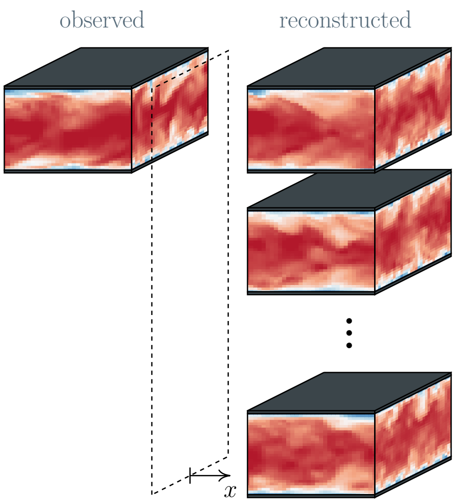

Figure 2. Schematic of the observational setup for the flow reconstruction problem.

Direct numerical simulation

We perform a Direct Numerical Simulation (DNS) of pressure-driven channel flow at friction Reynolds number $\mathrm{Re}_\tau = 180$ using the pseudospectral code *SPWind* (cf. Meyers & Sagaut (2007)). The flow is governed by the incompressible Navier–Stokes equations in a domain of size $8\pi\delta \times 2\delta \times 3\pi\delta$ along the streamwise (x), wall-normal (y), and spanwise (z) directions, where $\delta = 1$ is the half-channel height. The domain is periodic in the streamwise and spanwise directions (x and z), with a grid resolution of $\Delta x^+ \times \Delta z^+ = 11.78 \times 5.89$ or $N_x \times N_z = 384 \times 288$. In the wall-normal direction, a fourth-order energy conservative scheme (Verstappen & Veldman, 2003) is used on a grid resolution of $\Delta y^+ \in [0.17, 3.1]$ obtained for $N_y = 256$ with a hyperbolic tangent stretching function. After an initial spin-up, the simulation is ran for $Tu_\tau/\delta = 50$ time units and snapshots are taken at $\Delta t u_\tau/\delta = 0.25$.

The requirements for the training data differ from those for the DNS. The learned statistics should be homogeneous in the streamwise and spanwise direction, and not periodic as in the DNS. Therefore, the DNS domain is cut in boxes of size $2\pi\delta \times 2\delta \times \pi\delta/2$, which no longer satisfy the periodic boundary conditions. The velocity fields are then filtered and downsampled by a factor two, such that $N_s = 4824$ snapshots $\{\boldsymbol{u}_s\}_{s=1}^{N_s}$ are obtained with $\boldsymbol{u}_s \in \mathbb{R}^n$ and $n = 3N_xN_yN_z$ where $N_x = 48, N_y = 128$, and $N_z = 24$. Of those snapshots, the first 80% is used for training, the next 10% for validation, and the final 10% for testing out of sample.

Estimating the expected reconstruction error

The mean-square error on each velocity component at each streamwise, spanwise, and wall-normal location in the unobserved region is approximately proportional to the corresponding diagonal entry of the conditional covariance matrix $\boldsymbol{\Sigma}_{u|o}$ (cf. Eq. (6)). The expected MSE can be estimated from the snapshots using similar techniques as for the LSE operator.

In the considered measurement setup, the spanwise direction is statistically homogeneous, such that the spanwise Fourier-transformed covariance matrix diagonalizes per wavenumber k_z in that direction (Tinney *et al.*, 2006). Therefore, the conditional covariance (Eq. (3c)) can be computed per wavenumber k_z, using the spanwise Fourier transformed velocity fields $\hat{\boldsymbol{u}}_{o,s}^{(k_z)} \in \mathbb{C}^r$ and $\hat{\boldsymbol{u}}_{u,s}^{(k_z)} \in \mathbb{C}^{m-r}$ with $r = p/N_z$ and $m = n/N_z = 3N_xN_y$. Note that the snapshots on the smaller domains are not longer exactly periodic in the spanwise direction, such that the discrete Fourier transform introduces spectral leakage. Since the reconstruction error is then also homogeneous in the spanwise direction, we directly compute the spanwise average of the conditional covariance matrix as $\langle \boldsymbol{\Sigma}_{u|o} \rangle_z = \sum_{k_z} \boldsymbol{\Sigma}_{u|o}^{(k_z)}/N_z$, by using Parseval's theorem on each of its components.

It can be shown that the mean-square reconstruction error in snapshot-based LSE equals the projection error on the subspace spanned by the snapshots, with rank N_s for linearly independent snapshots. Therefore, the reconstruction error vanishes when the number of measurements equals the number of snapshots, although the true reconstruction problem is still underdetermined when there are less measurements than state variables. To partially correct for this negative bias, we estimate the conditional covariance for each wavenumber as

$$\boldsymbol{S}_{u|o}^{(k_z)} = \frac{N_s - 1}{N_s - 1 - r}\left(\hat{\boldsymbol{S}}_{uu}^{(k_z)} - \hat{\boldsymbol{S}}_{uo}^{(k_z)} \hat{\boldsymbol{S}}_{oo}^{(k_z)^{-1}} \hat{\boldsymbol{S}}_{ou}^{(k_z)}\right) \qquad (10)$$

where the factor $(N_s - 1)/(N_s - 1 - r)$ accounts for the reduced degrees of freedom in the Wishart distribution of the Schur complement of $\hat{\boldsymbol{S}}^{(k_z)} = \sum_{n=1}^{N_s} \hat{\boldsymbol{u}}_s^{(k_z)} \hat{\boldsymbol{u}}_s^{(k_z)*}/(N_s - 1) \in \mathbb{C}^{m\times m}$ (Muirhead, 1982) and the asterix denotes the complex conjugate. In order to keep $r < N_s - 1$, we further downsample the velocity field in the vertical direction by a factor 4 in the wall-normal direction when estimating the expected reconstruction error. Because of the stretched grid in the wall-normal direction, the second order statistics are found to be still adequately represented in that case.

RESULTS AND DISCUSSION

Before assessing the use of the machine-learned conditional prior for flow reconstruction, the prior itself is validated by examining its ensemble statistics. Then, its conditional sampling quality for flow reconstruction is examined by comparing the reconstruction error with the error estimates based on LSE, and by examining the ensemble statistics of the reconstructed fields in the unobserved region.

Ensemble statistics of machine-learned prior

The ensemble statistics of the machine-learned prior are compared with those of the DNS test data, using an ensemble of 500 generated velocity fields. As seen in Figure 3a-c, the mean velocity profile and Reynolds stresses are accurately reproduced, despite their inhomogeneity and anisotropy. The spatial coherency of the turbulent channel flow fields is also reproduced, as indicated by the good agreement between the streamwise one-dimensional energy spectra at different distances from the wall in Figure 3d. At $y^+ = 9.1$, the energy content of the small scales is inflated in both the velocity fields from test data and those sampled from the machine-learned prior. Since the snapshots are not truly periodic, this is likely the result of spectral leakage in the discrete Fourier transform. Figures 3e and 3f show that the local spatial coherency is also replicated, as the probability distributions of the longitudinal and transverse streamwise velocity increments agree with those from DNS. Moreover, their heavy tails and the significant skewness of longitudinal velocity increments show that the non-Gaussian features of the velocity field are also represented in the machine-learned prior.

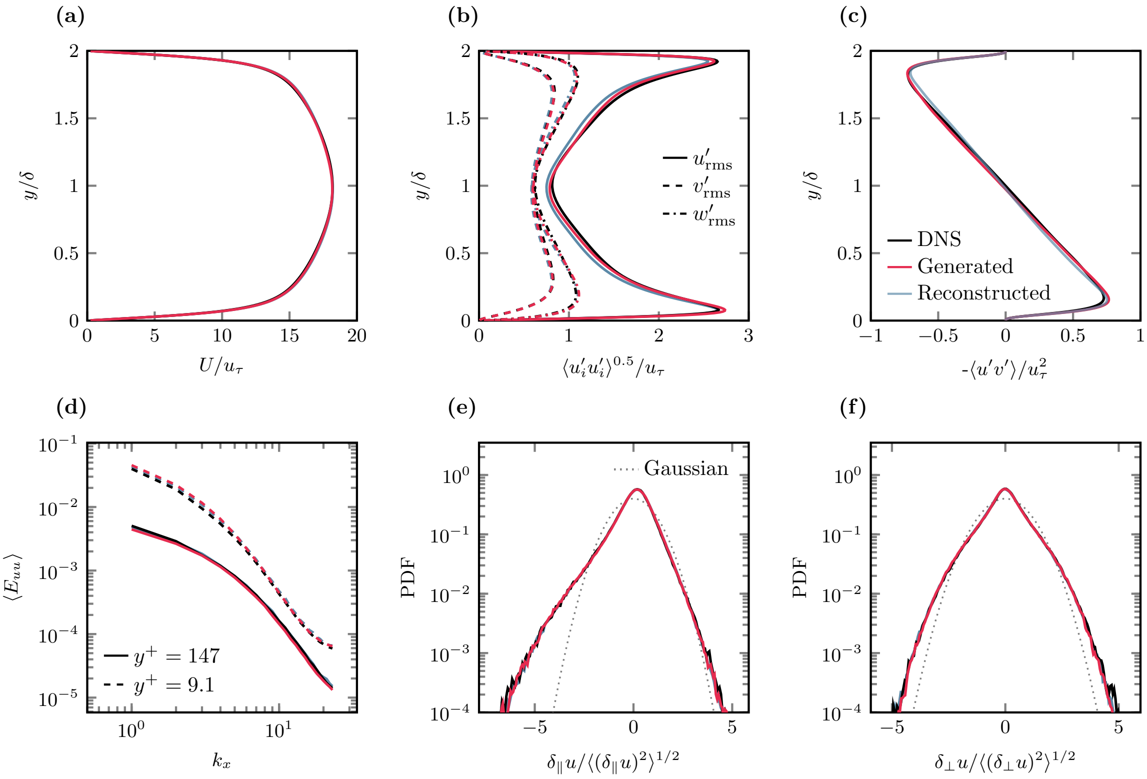

Figure 3. Comparison of ensemble statistics derived from 500 generated (and reconstructed) flow fields and those from the DNS test data. The comparison includes the (a) mean velocity profile, (b) root-mean-square velocity fluctuations, (c) Reynolds shear stress, (d) one-dimensional energy spectrum in the streamwise direction, (e) distribution of the longitudinal streamwise velocity increments at $y = 0.19\delta$, and (f) distribution of the transverse streamwise velocity increments in the spanwise direction at $y = 0.19\delta$.

Comparison of the reconstruction error

Now we reconstruct masked downstream blocks of turbulent velocity fields from the DNS test data. To this end, ensembles of 50 reconstructed flow fields are generated for each of in total 50 flow fields taken from the test data. Since the variance of the velocity components varies with the wall-normal position (cf. Figure 3b), we compute the following summarizing mean-square error for each velocity component

$$\langle e_u^2(x) \rangle = \frac{1}{\pi\delta} \int_0^{2\delta} \left(\int_0^{\pi/2} \frac{\left\langle [(\tilde{u}(\boldsymbol{x}) - u(\boldsymbol{x})]^2 \right\rangle}{\langle u'(\boldsymbol{x}) u'(\boldsymbol{x}) \rangle} \, \mathrm{d}z \right) \mathrm{d}y. \quad (11)$$

When computing the integral in the wall-normal direction, we however neglect the error in the region closer than $\Delta y^+ = 1$ to the wall, because of the vanishing variance. The expected mean squared error is computed in a similar fashion, where the numerator in Eq. (11) is given by the diagonal elements of the conditional covariance matrix from Eq. (6) and the spanwise average is computed through Parseval's theorem.

The expected and obtained height-averaged and normalized mean-square errors are compared in Figure 4. As anticipated, the expected reconstruction error increases gradually with the depth into the unobserved region x, as defined in Figure 2. Close to the observed region ($x \to 0$), the velocity field is still strongly correlated, whereas further on, it becomes uncorrelated from the observed part of the velocity field. It is seen that the mean-square reconstruction error obtained with the machine-learned prior evolves towards the same bounds: once the variance for the ensemble average of the reconstructions and twice the variance for the ensemble members. The discrepancy between the expected and obtained error is partly due to the negative bias of the estimated error and partly due to the approximations made in the conditional sampling algorithm. Although Figure 4 only shows the reconstruction error when 50% of the channel is observed, the error evolves toward the same bounds when only one upstream slice or 2.1% of the channel is observed (not shown).

Ensemble statistics of the unobserved part

The ensemble statistics of the unobserved part of the velocity fields are also compared with those from the DNS test data and the velocity fields generated from the unconditional prior in Figure 3. All statistics, except the energy spectra, are computed in the unobserved downstream half of the channel. Interestingly, all reconstructed velocity fields display practically the same one-point first- and second-order statistics, one-dimensional streamwise energy spectra, and velocity increment distributions. Moreover, essential non-Gaussian features, such as the negative skewness of the longitudinal streamwise velocity increments, which is related to the interscale energy flux (cf. Kolmogorov's four-fifths law), are reproduced in the unobserved region as well. Hence, the considered flow-based generative model performs well both in terms of the consistency (mean squared error) as in terms of the reality (ensemble statistics) of the instantaneous conditional structures.

CONCLUSIONS

Probabilistic machine learning models offer unique new ways to approximate the (conditional) prior distribution of turbulent flow fields, central to turbulent flow reconstruction when regarded as a stochastic estimation problem. We train a flow-

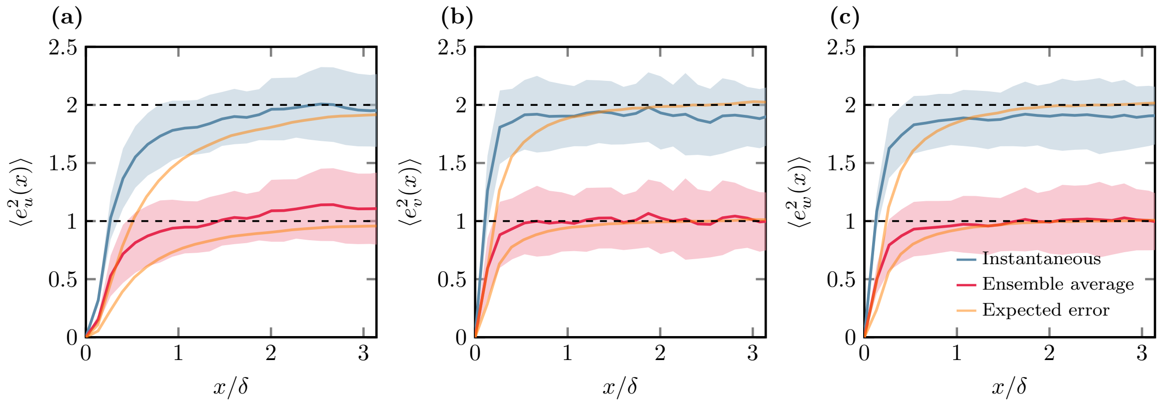

Figure 4. Comparison of the height-averaged and normalized mean-square reconstruction error obtained with the machine-learned prior and estimated based on linear stochastic estimation for the (a) streamwise, (b) wall-normal, and (c) spanwise velocity component. The averaged error over 50 distinct observations is given as a solid line with a band of ± one standard deviation.

based generative model to reconstruct instantaneous three-dimensional conditional structures in a turbulent channel flow, for the first time. To this end, we adapted the RePaint algorithm by Lugmayr *et al.* (2022) to work with flow-based generative models. First, we show that the machine-learned prior matches the ensemble statistics from DNS, including the heavy tails of velocity increment distributions and negative skewness of the longitudinal velocity increments, which is crucial for spectral energy transfer. Secondly, we assess the consistency of the reconstructed fields by comparing the reconstruction error with an estimated expected reconstruction error, based on linear stochastic estimation (Adrian *et al.*, 1987). It is shown that the reconstruction error evolves spatially towards the expected bounds, although at a faster rate. Thirdly, we confirm that the ensemble statistics of the reconstructed fields in the unobserved region still match those from DNS. Consequently, the machine-learned prior is shown to enable the reconstruction of an ensemble of realistic instantaneous conditional eddies, complementing the conditionally averaged eddies from linear stochastic estimation. Future work will systematically compare higher-order ensemble statistics as well as terms in the turbulent kinetic energy budget to further substantiate these claims.

REFERENCES

Adrian, R., Moin, P. & Moser, R. 1987 Stochastic estimation of conditional eddies in turbulent channel flow. *Proceedings of the 1987 Summer Program of the Stanford Center of Turbulence Research* .

Adrian, R. J. 1996 Stochastic Estimation of the Structure of Turbulent Fields. In *Eddy Structure Identification* (ed. J. P. Bonnet), pp. 145–195. Vienna: Springer Vienna.

Amorós-Trepat, M., Medrano-Navarro, L., Liu, Q., Guastoni, L. & Thuerey, N. 2026 Guiding diffusion models to reconstruct flow fields from sparse data. *Physics of Fluids* **38** (1), publisher: AIP Publishing.

Arun, R., Bae, H. J. & McKeon, B. J. 2023 Towards real-time reconstruction of velocity fluctuations in turbulent channel flow. *Physical Review Fluids* **8** (6), 064612.

Dhariwal, P. & Nichol, A. 2021 Diffusion Models Beat GANs on Image Synthesis. ArXiv:2105.05233 [cs].

Du, P., Parikh, M. H., Fan, X., Liu, X.-Y. & Wang, J.-X. 2024 Conditional neural field latent diffusion model for generating spatiotemporal turbulence. *Nature Communications* **15** (1), 10416, publisher: Nature Publishing Group.

Li, T., Buzzicotti, M., Biferale, L., Bonaccorso, F., Chen, S. & Wan, M. 2023*a* Multi-scale reconstruction of turbulent rotating flows with proper orthogonal decomposition and generative adversarial networks. *Journal of Fluid Mechanics* **971**, A3.

Li, T., Lanotte, A. S., Buzzicotti, M., Bonaccorso, F. & Biferale, L. 2023*b* Multi-Scale Reconstruction of Turbulent Rotating Flows with Generative Diffusion Models. *Atmosphere* **15** (1), 60.

Lipman, Y., Havasi, M., Holderrieth, P., Shaul, N., Le, M., Karrer, B., Chen, R. T. Q., Lopez-Paz, D., Ben-Hamu, H. & Gat, I. 2024 Flow Matching Guide and Code. ArXiv:2412.06264 [cs].

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R. & Gool, L. V. 2022 RePaint: Inpainting using Denoising Diffusion Probabilistic Models. ArXiv:2201.09865 [cs].

Meyers, J. & Sagaut, P. 2007 Is plane-channel flow a friendly case for the testing of large-eddy simulation subgrid-scale models? *Physics of Fluids* **19** (4), 048105.

Muirhead, R. J. 1982 *Aspects of Multivariate Statistical Theory*, 1st edn. Wiley.

Parikh, M. H., Fan, X. & Wang, J.-X. 2026 Conditional flow matching for generative modelling of near-wall turbulence with quantified uncertainty. *Journal of Fluid Mechanics* **1029**, A32.

Rybchuk, A., Hassanaly, M., Hamilton, N., Doubrawa, P., Fulton, M. J. & Martínez-Tossas, L. A. 2023 Ensemble flow reconstruction in the atmospheric boundary layer from spatially limited measurements through latent diffusion models. *Physics of Fluids* **35** (12), 126604.

Tinney, C. E., Coiffet, F., Delville, J., Hall, A. M., Jordan, P. & Glauser, M. N. 2006 On spectral linear stochastic estimation. *Experiments in Fluids* **41** (5), 763–775.

Verstappen, R. & Veldman, A. 2003 Symmetry-preserving discretization of turbulent flow. *Journal of Computational Physics* **187** (1), 343–368.

Vishwasrao, A., Gutha, S. B. C., Cremades, A., Wijk, K., Patil, A., Gorle, C., McKeon, B. J., Azizpour, H. & Vinuesa, R. 2025 Diff-SPORT: Diffusion-based Sensor Placement Optimization and Reconstruction of Turbulent flows in urban environments. ArXiv:2506.00214 [physics].

Yousif, M. Z., Yu, L., Hoyas, S., Vinuesa, R. & Lim, H. 2023 A deep-learning approach for reconstructing 3D turbulent flows from 2D observation data. *Scientific Reports* **13** (1), 2529.