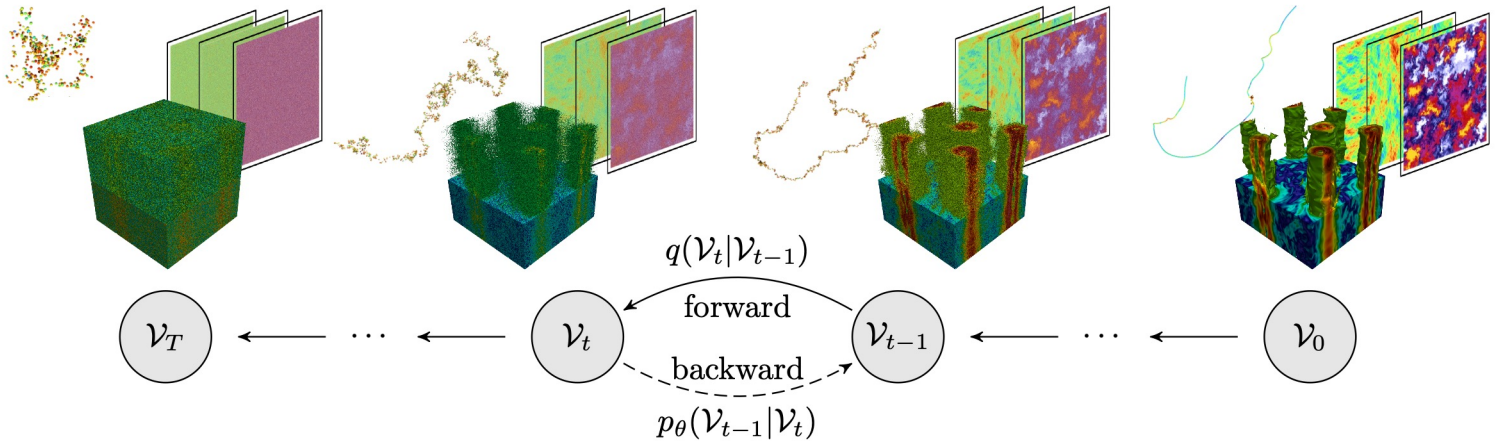
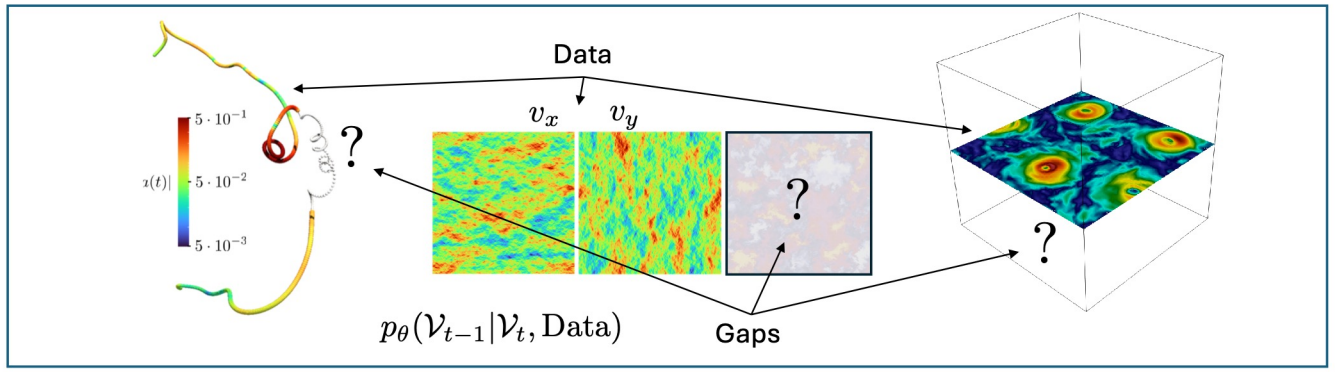
Physics-Constrained Diffusion Model for Synthesis of 3D Turbulent Data





Michele Buzzicotti

Dept. Physics & INFN, University of Rome Tor Vergata

https://mbuzzico.github.io/

michele.buzzicotti@roma2.infn.it

CREDITS:
Luca Biferale, **Mauro Sbragaglia**,
Tianyi Li, **Fabio Bonaccorso** (Uni. Tor Vergata, IT),
Luca Centurioni (Scripps San Diego, CA)

FIS

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

INFN

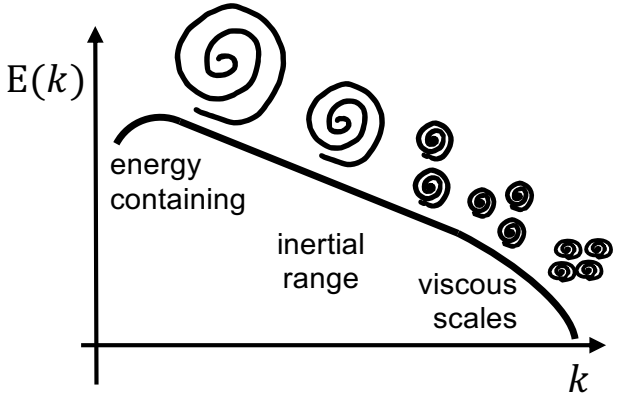
MUR Ministero dell'Università e della Ricerca

FARE RICERCA IN ITALIA FRAMEWORK PER L'ATTRAZIONE E IL RAFFORZAMENTO DELLE ECCELLENZE PER LA RICERCA IN ITALIA

TOR VERGATA UNIVERSITÀ DEGLI STUDI DI ROMA

erc European Research Council Established by the European Commission Supporting top researchers from anywhere in the world

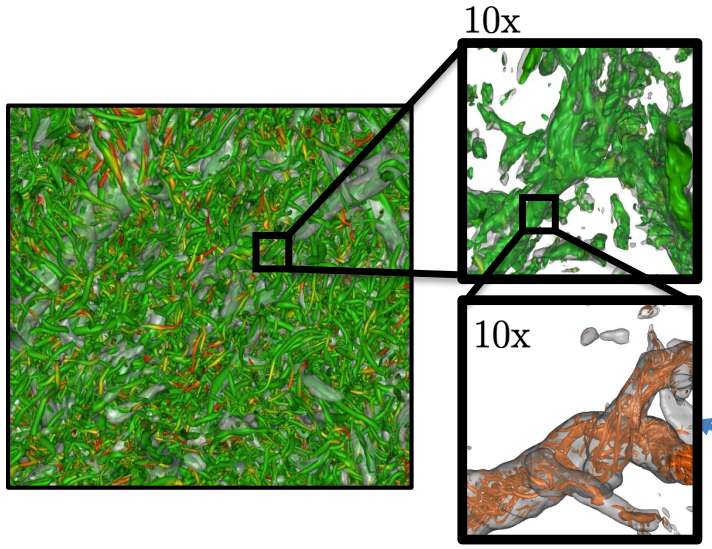
Turbulence: A Complex Problem



Navier Stokes Equations

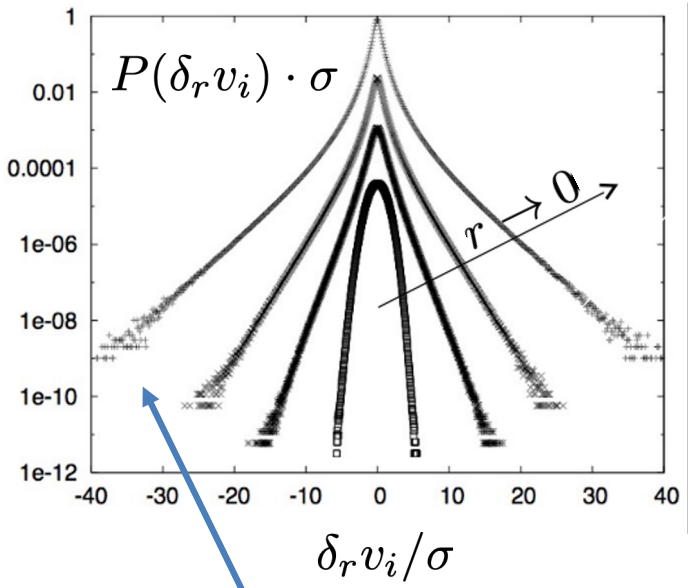
$$\begin{cases} \partial_t \boldsymbol{v} + \boldsymbol{v} \cdot \nabla \boldsymbol{v} = -\nabla p + \nu \nabla^2 \boldsymbol{v} + \boldsymbol{f} \\ \nabla \cdot \boldsymbol{v} = 0 \end{cases}$$

#1 Scale separation and effective dimensionality



#2 Big-tails, Intermittency

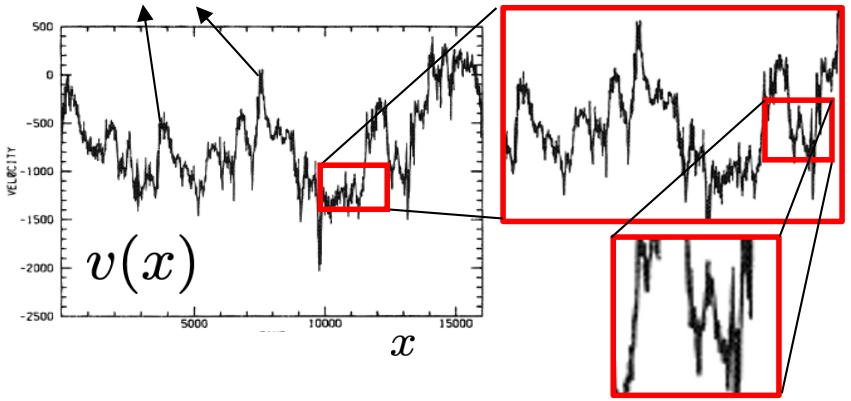
$$\delta_r v_i = \langle v_i(\boldsymbol{x} + r) - v_i(\boldsymbol{x}) \rangle$$



Small Scales
Extreme/rare events

#3 Rough velocity profile. Loss of regularity and Multifractal fluctuations

$$\delta_r v \sim r^{h(x)}$$



Wide scaling range where velocity is **Non-Differentiable**

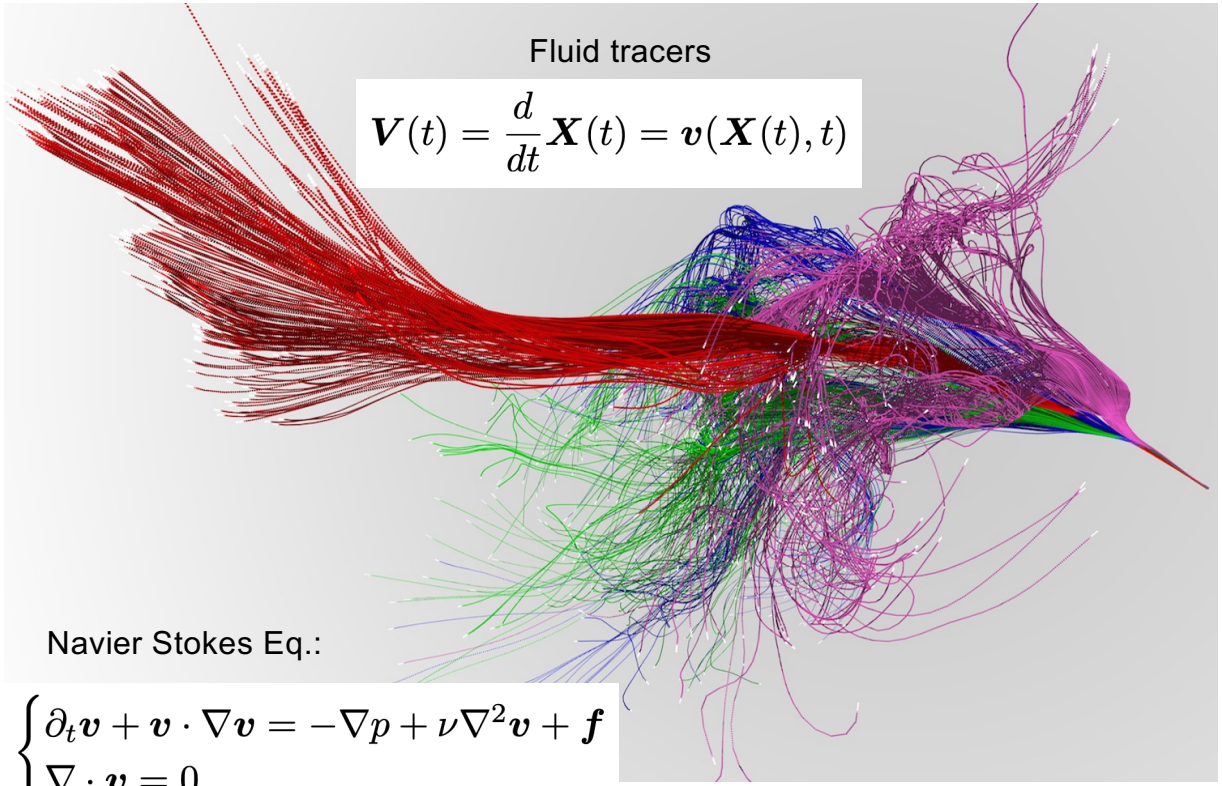
Frisch, Uriel. *Turbulence: the legacy of AN Kolmogorov*. Cambridge university press, 1995.

Alexakis, Alexandros, and Luca Biferale. "Cascades and transitions in turbulent flows." *Physics Reports* 767 (2018): 1-101.

#4 Lagrangian Description of Turbulence: Why?

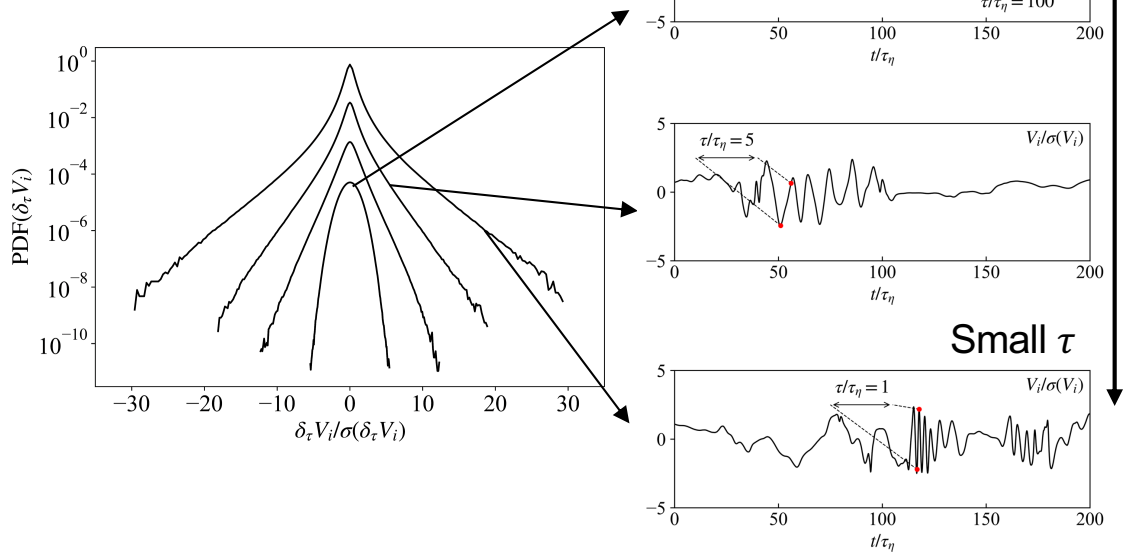
(I) TESTING APPLICATIONS/MODELS:

(A) Single, Two and Multi-Particles Dispersion
(B) Particles' Breaking/Aggregation
(C) Multi-time Multi-scales Turbulent Fluctuations

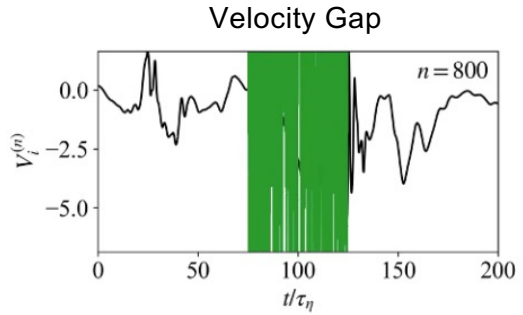


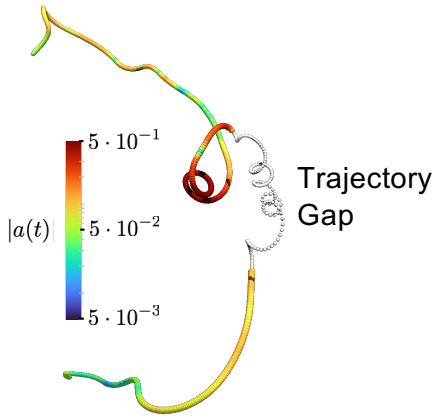
Non-Self-Similar Statistics

$\delta_\tau V_i(t) = V_i(t+\tau) - V_i(t)$



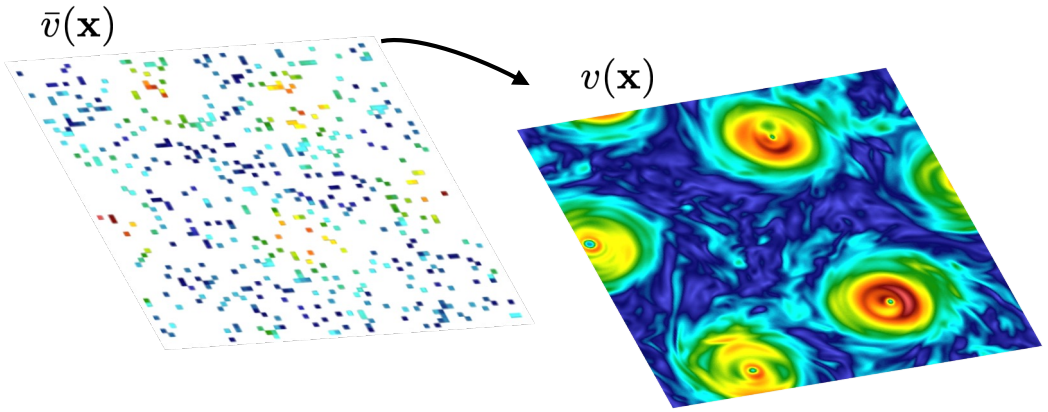
(II) Reconstruction of Gappy Data





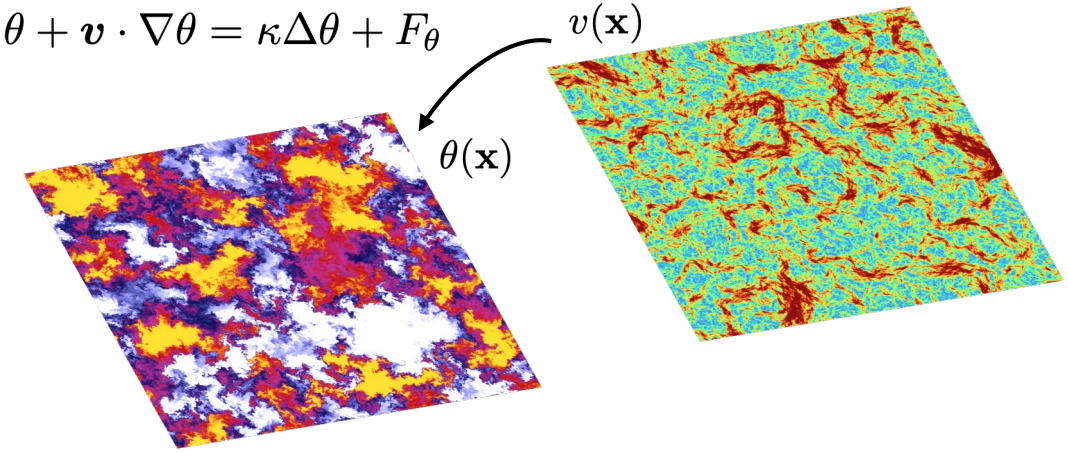
Full-State Reconstruction

Sparse Observations



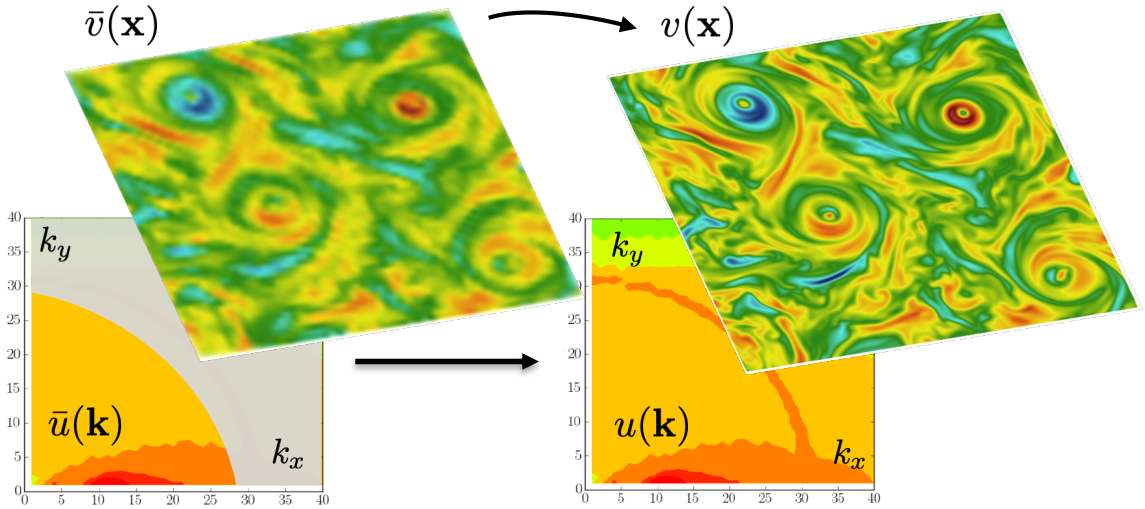
Inverse Problems

$$\begin{cases} \partial_t \boldsymbol{v} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\nabla p + \nu\Delta\boldsymbol{v} - \beta\boldsymbol{v} + \mathbf{F}_v \\ \partial_t\theta + \boldsymbol{v}\cdot\nabla\theta = \kappa\Delta\theta + F_\theta \end{cases}$$



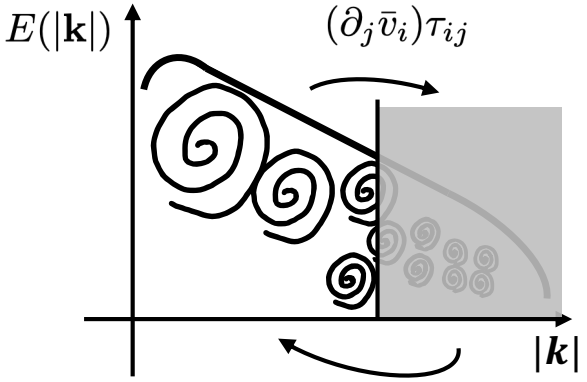
Small-Scales Reconstruction

Super Resolution

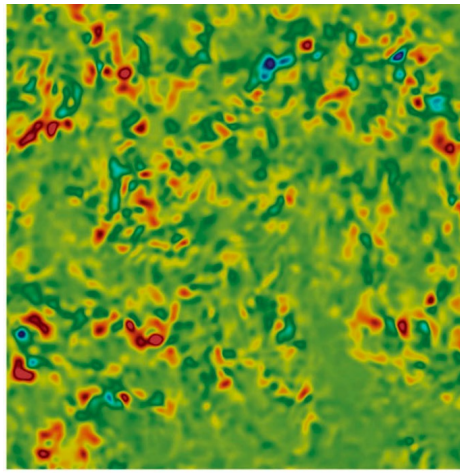


Sub-Grid Closure

$$\tau_{ij} = \overline{v_i v_j} - \bar{v}_i \bar{v}_j$$



$$\frac{1}{2}\partial_t(\bar{v}_i\bar{v}_i) + \ldots = (\partial_j\bar{v}_i)\tau_{ij}$$



Data-Driven Generative Models

Standard Gaussian

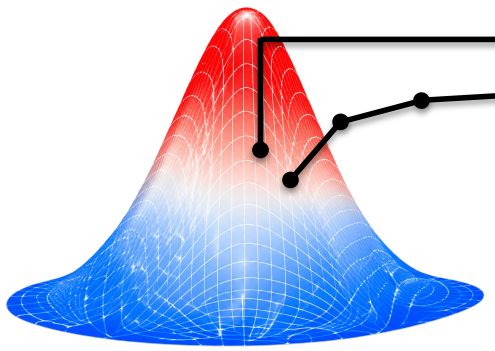
$\mathcal{N}(0, I)$

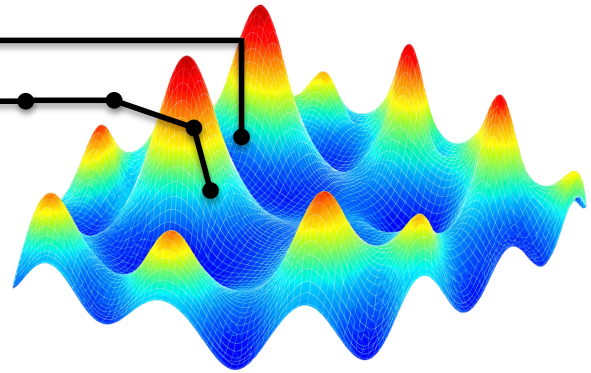
Data Distribution

$p_{\text{data}}(x)$

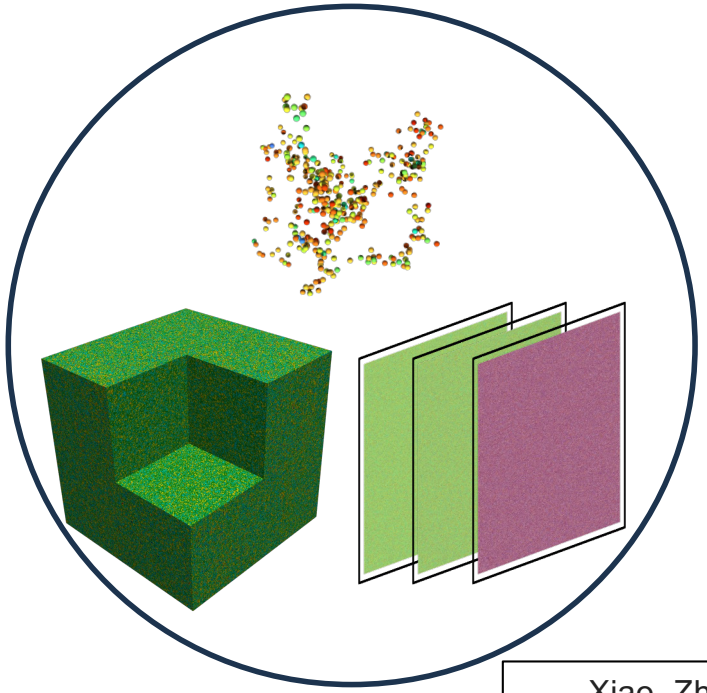
VAE, GAN

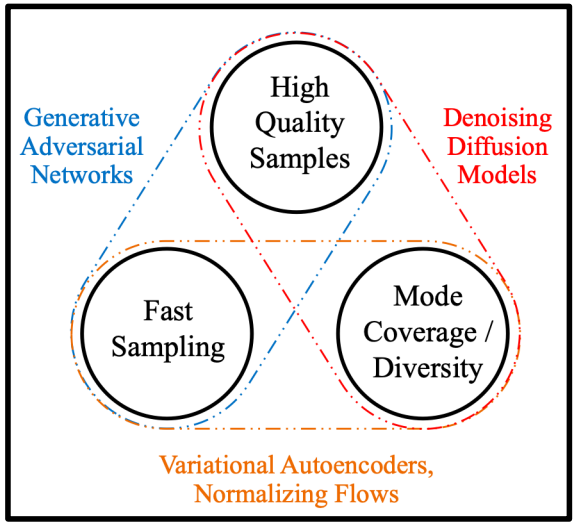
Norm Flows

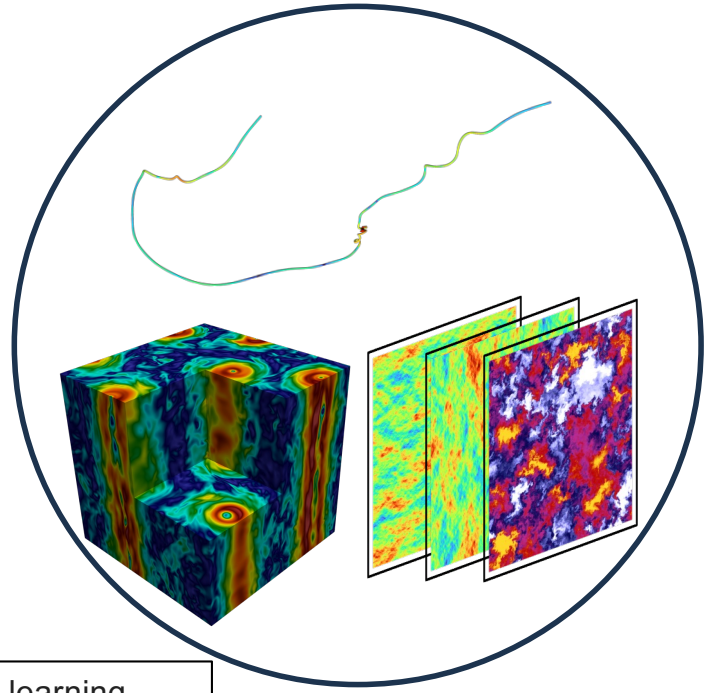




Generative Models learn a MAP
from a simple distribution into the
Data Distribution

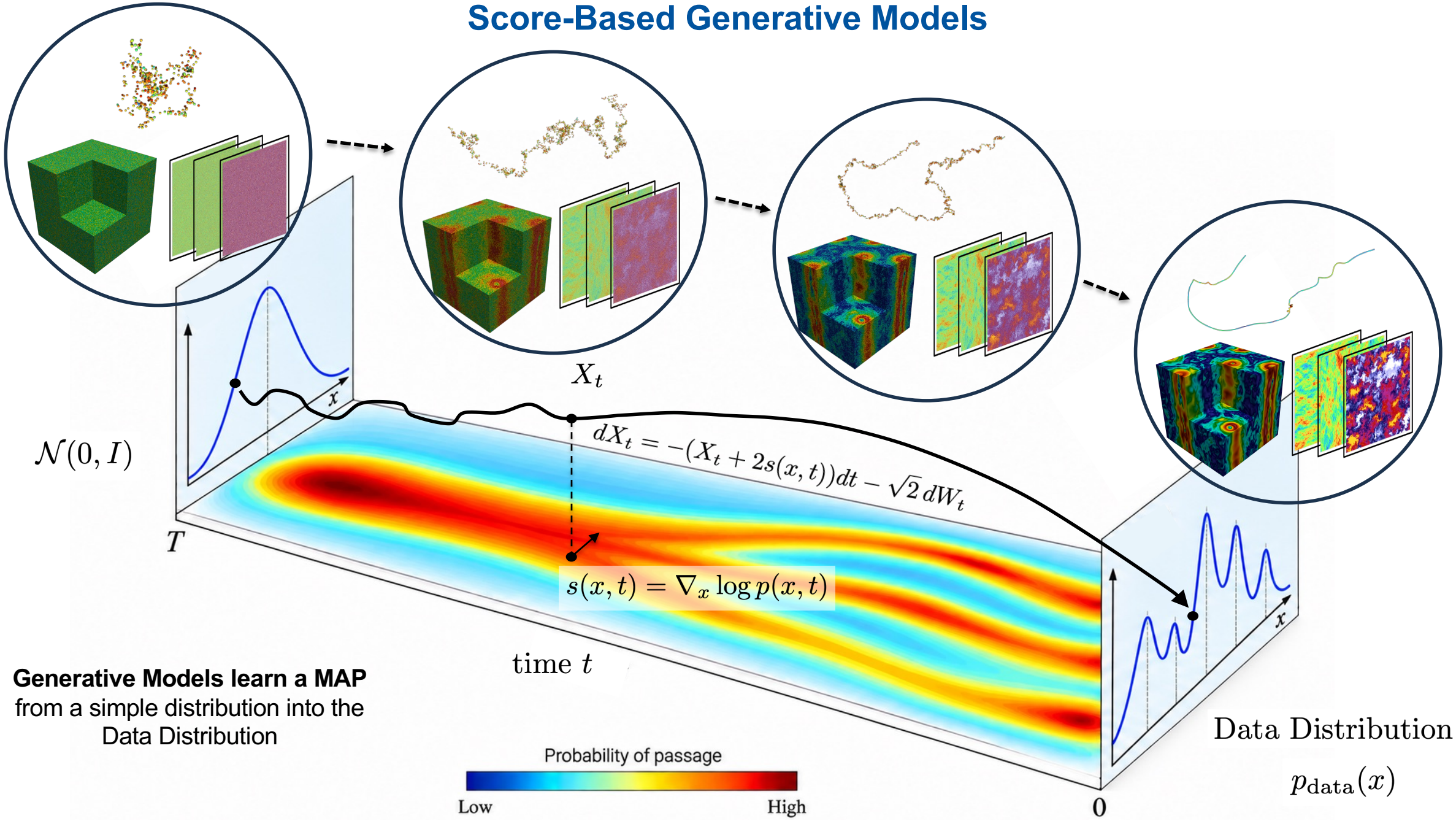






Xiao, Zhisheng, Karsten Kreis, and Arash Vahdat. "Tackling the generative learning trilemma with denoising diffusion gans." *arXiv preprint arXiv:2112.07804* (2021).

Score-Based Generative Models



Generative Models learn a MAP
from a simple distribution into the
Data Distribution

Forward process:

Ornstein-Uhlenbeck diffusion progressively maps the data distribution onto an i.i.d. Gaussian distribution in the long-time limit.

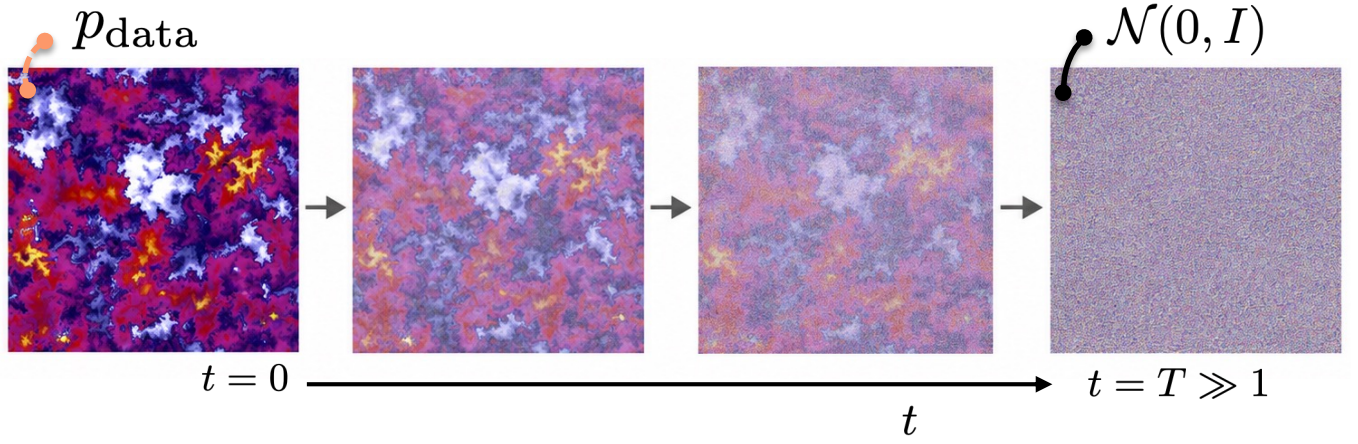
OU process

$$\begin{cases} dX_t = -X_t\,dt + \sqrt{2}\,dW_t \\ X_0 \sim p_{\text{data}} \end{cases}$$

Wiener process

Forward marginal

$$\begin{cases} p(x,t) = \int p(x,t\,|\,x_0,0)\,p_{\text{data}}(x_0)\,dx_0 \\ p(x,t\,|\,x_0,0) = \mathcal{N}\left(e^{-t}x_0,\ (1-e^{-2t})I\right) \end{cases}$$



Backward (generative) process:

Reverse-time Langevin dynamics, where the learned score function defines the drift that progressively reverses the forward diffusion.

Reverse Langevin

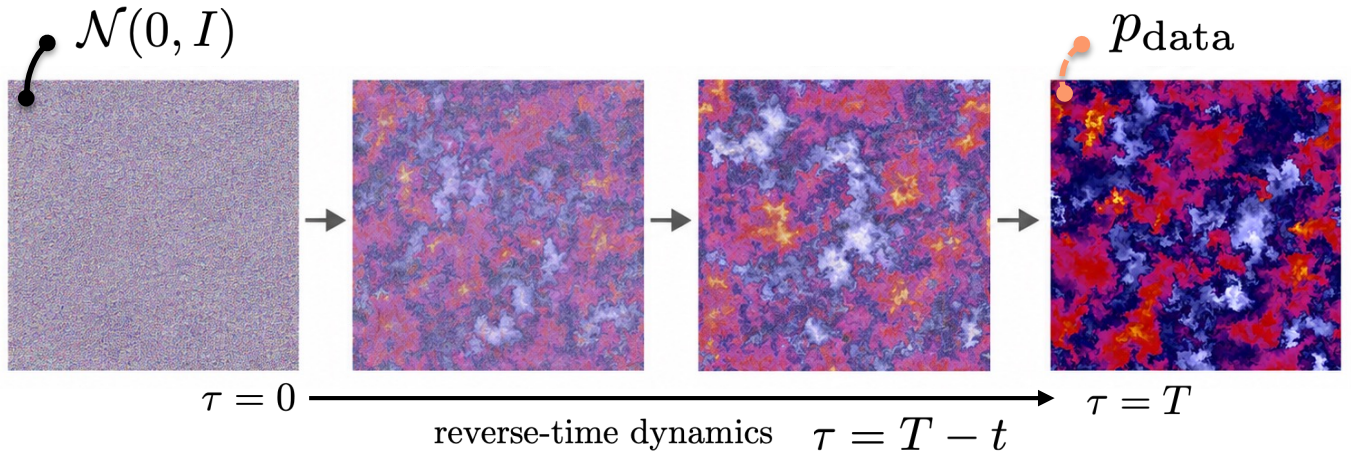
$$\begin{cases} dX_\tau = \left(X_\tau + 2\,s(x,t_t-\tau)\right)d\tau + \sqrt{2}\,dW_\tau \\ X_0 \sim \mathcal{N}(0,I) \end{cases}$$

score function

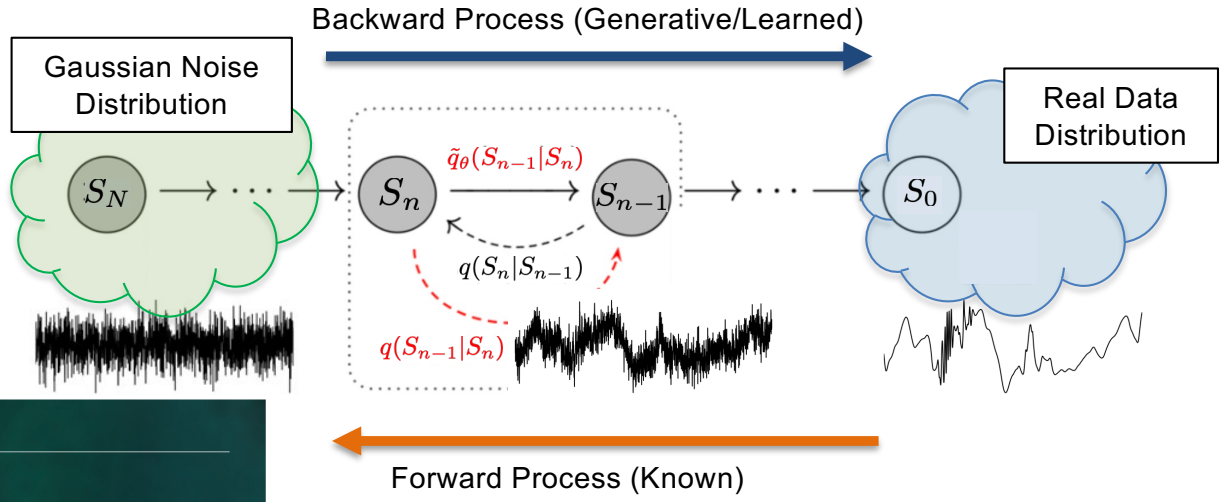
$$s(x,t) = \nabla_x \log p(x,t)$$

$$\boxed{s_\theta(x,t) \approx s(x,t)}$$

Learned via Neural-Net



Synthetic Lagrangian Turbulence by Generative Diffusion Models

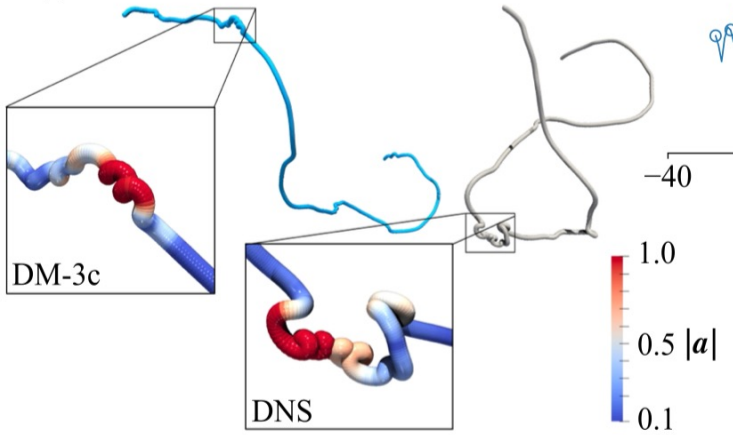


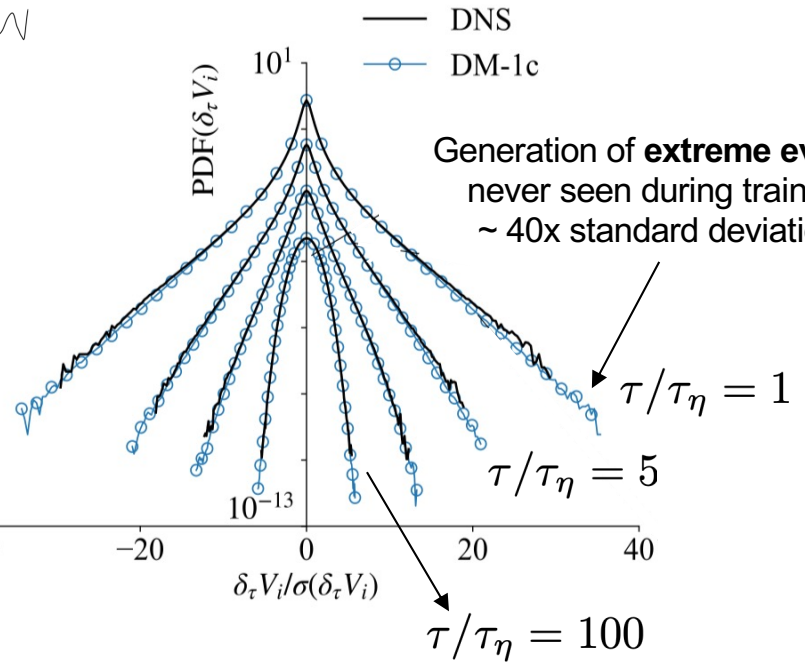
First Application of DM for the generation of turbulent data

Lagrangian particle trajectories advected by 3D fully developed Turbulent flows



Generation of the correct 3D **topological structure**





Generation of **extreme events** never seen during training ~ 40x standard deviation

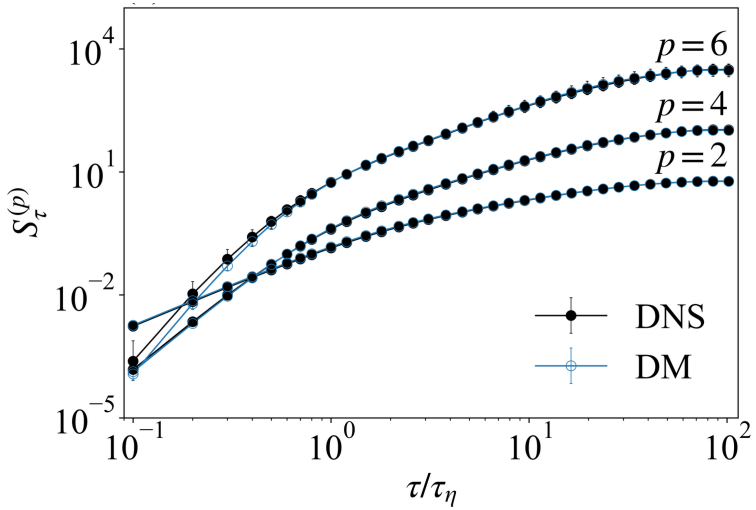
$$\delta_\tau V_i(t) = V_i(t+\tau) - V_i(t)$$

T. Li, L. Biferale, F. Bonaccorso, M. A. Scarpolini, and M. Buzzicotti, *Nature Machine Intelligence* (2024): 1-11.

Statistical Comparison with DNS data

Lagrangian Structure Functions

$$S_\tau^{(p)} = \langle (\delta_\tau V_i)^p \rangle$$

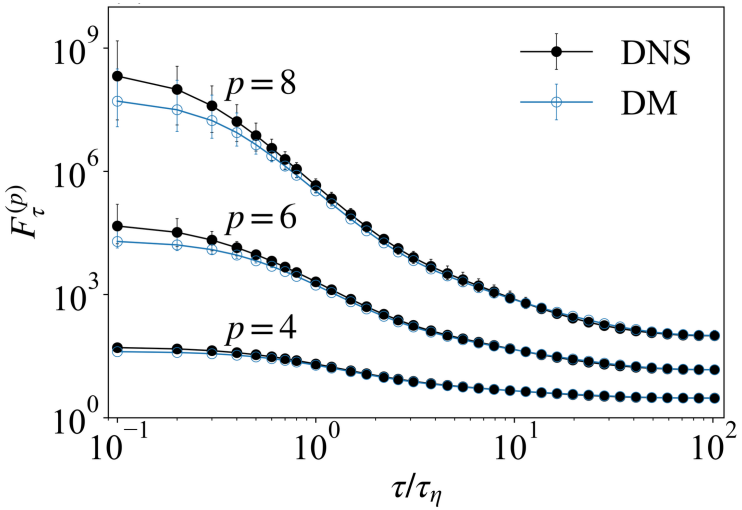


$$\delta_\tau V_i(t) = V_i(t+\tau) - V_i(t)$$

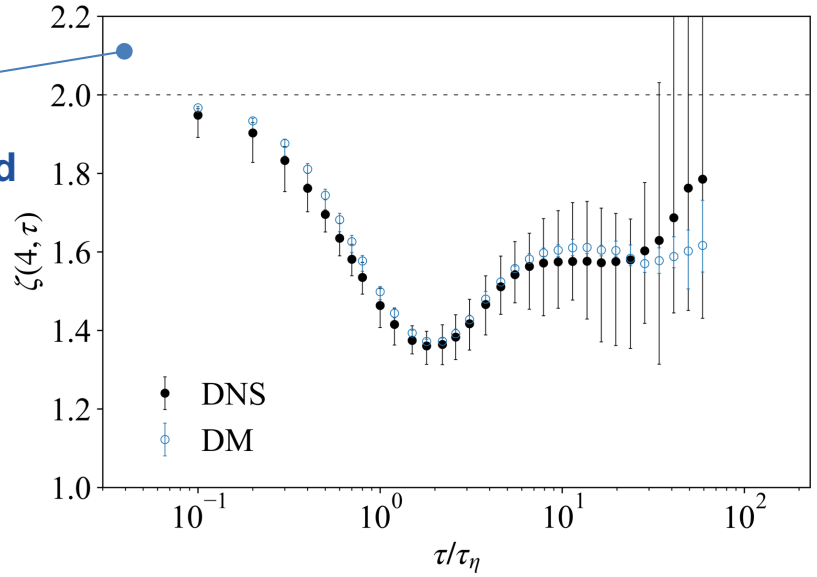
$$\zeta(4,\tau) = \frac{d \log S_\tau^{(4)}}{d \log S_\tau^{(2)}}$$

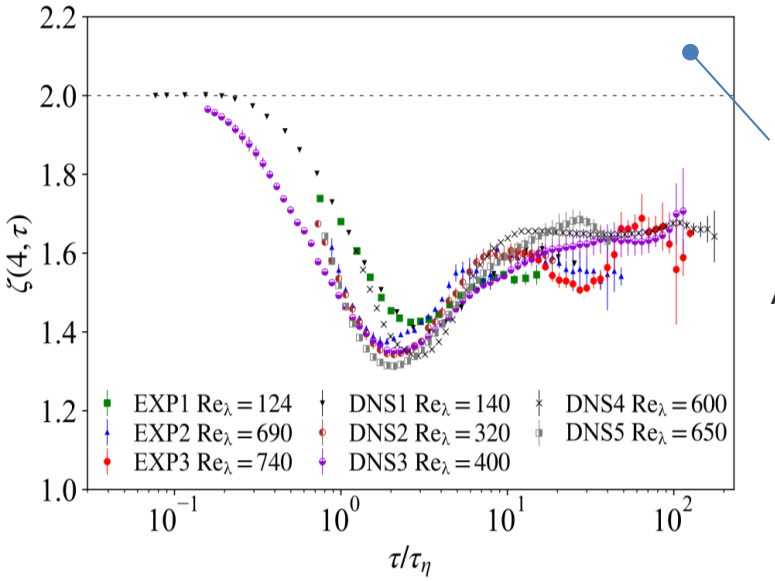
Generalized Flatness

$$F_\tau^{(p)} = S_\tau^{(p)} / [S_\tau^{(2)}]^{p/2}$$



Our Results; Training & Genereted Data



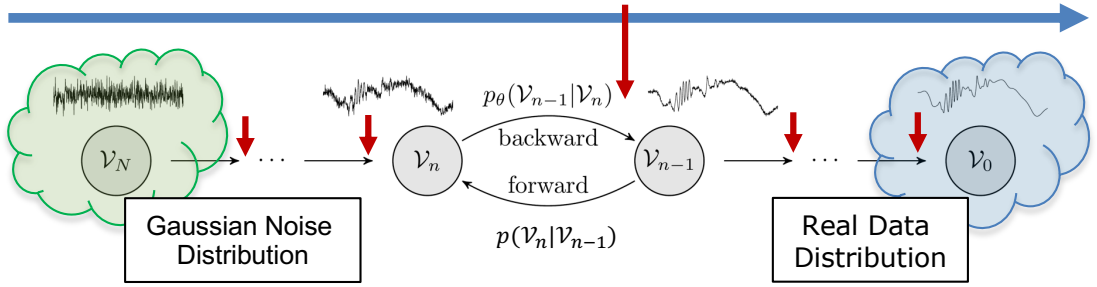


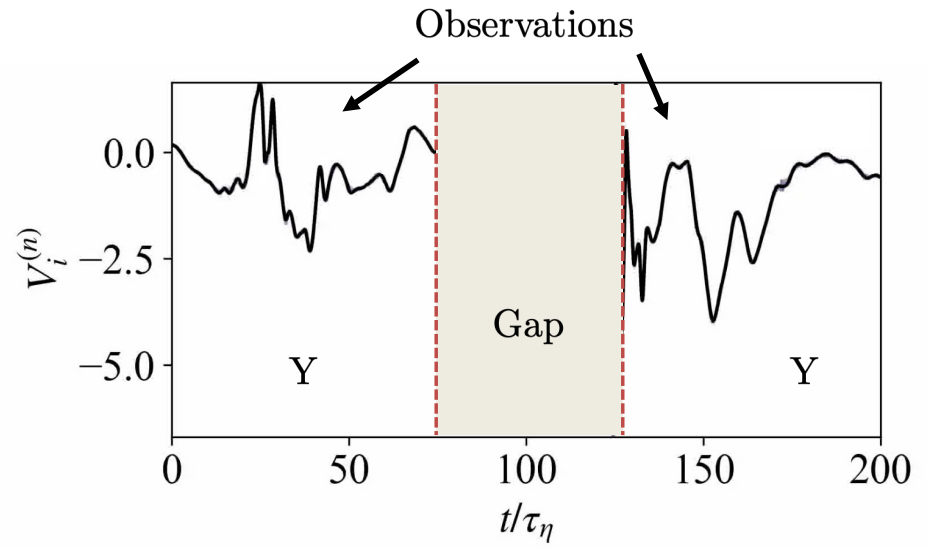
Experimental & Numerical Results

Arnéodo, A. et al. Phys. Rev. Lett. 100, 254504 (2008).

Data Reconstruction - “Guiding the Diffusion Models”

Guided Backward Process:





communications physics

Article | Open access | Published: 26 September 2025

Stochastic reconstruction of gappy Lagrangian turbulent signals by conditional diffusion models

Tianyi Li, Luca Biferale, Fabio Bonaccorso, Michele Buzzicotti ✉ & Luca Centurioni

Communications Physics **8**, Article number: 372 (2025) | Cite this article

Diffusion Posterior Sampling (DPS)

Ye, H., Lin, H., Han, J., Xu, M., Liu, S., Liang, Y., ... & Ermon, S.. Tfg: Unified training-free guidance for diffusion models. *arXiv:2409.15761*. (2024)

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C.. Diffusion posterior sampling for general noisy inverse problems. *arXiv:2209.14687*. (2022)

Unguided Backward Process:

$$\begin{cases} -dX_t = (X_t + 2\nabla_x \log p(x,t))\ dt + \sqrt{2}\, dW_t \\ X_0 \sim \mathcal{N}(0, I) \end{cases}$$

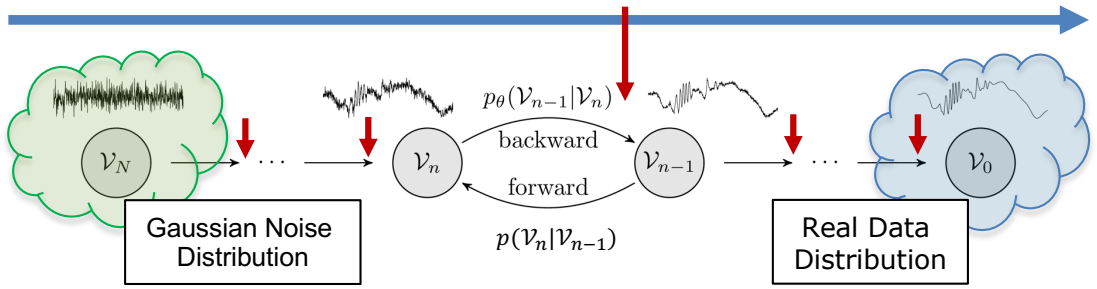
Guiding on the Observations (Y):

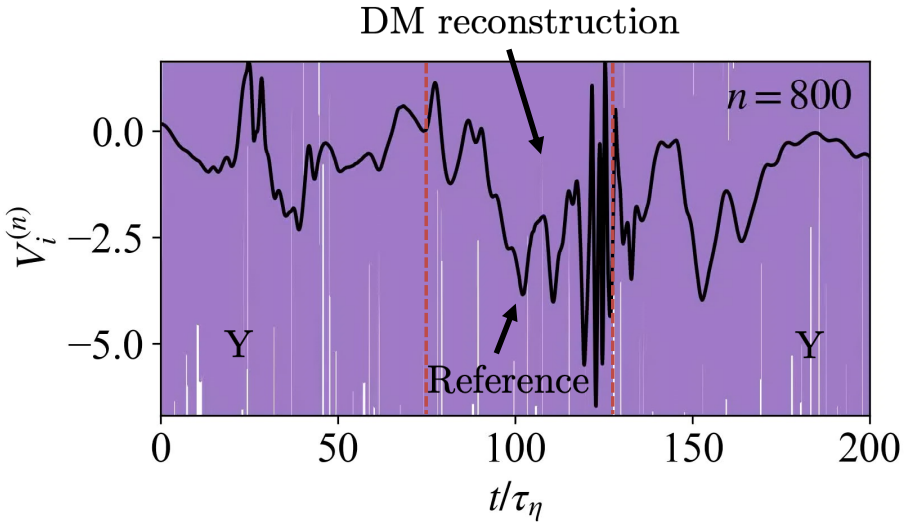
$$\frac{\partial \log p(x,t)}{\partial x} \rightarrow \frac{\partial \log p(x,t|Y)}{\partial x} \qquad p(x,t|Y) = \frac{p(Y|x,t)p(x,t)}{P_t(Y)}$$

$$\begin{cases} -dX_t = \left[X_t + 2\left(\nabla_x \log p(x,t) + \nabla_x \log p(Y|x,t)\right)\right] dt + \sqrt{2}\, dW_t \\ X_0 \sim \mathcal{N}(0, I) \end{cases}$$

Data Reconstruction - “Guiding the Diffusion Models”

Guided Backward Process:





Diffusion Posterior Sampling (DPS)

Ye, H., Lin, H., Han, J., Xu, M., Liu, S., Liang, Y., ... & Ermon, S.. Tfg: Unified training-free guidance for diffusion models. *arXiv:2409.15761*. (2024)

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C.. Diffusion posterior sampling for general noisy inverse problems. *arXiv:2209.14687*. (2022)

Unguided Backward Process:

$$\begin{cases} -dX_t = (X_t + 2\nabla_x \log p(x,t))\ dt + \sqrt{2}\, dW_t \\ X_0 \sim \mathcal{N}(0, I) \end{cases}$$

Guiding on the Observations (Y):

GUIDANCE

$$\approx s_\theta(x,t) \qquad \approx \zeta|Y - \hat{Y}_{0,\theta}(x,t)|$$

$$\begin{cases} -dX_t = \left[X_t + 2\left(\overbrace{\nabla_x \log p(x,t)}^{\approx s_\theta(x,t)} + \overbrace{\nabla_x \log p(Y|x,t)}^{\approx \zeta|Y - \hat{Y}_{0,\theta}(x,t)|}\right)\right] dt + \sqrt{2}\, dW_t \\ X_0 \sim \mathcal{N}(0, I) \end{cases}$$

communications physics

Article | Open access | Published: 26 September 2025

Stochastic reconstruction of gappy Lagrangian turbulent signals by conditional diffusion models

Tianyi Li, Luca Biferale, Fabio Bonaccorso, Michele Buzzicotti & Luca Centurioni

Communications Physics **8**, Article number: 372 (2025) | Cite this article

Reconstruction of NOAA's Global Drifter Program Data

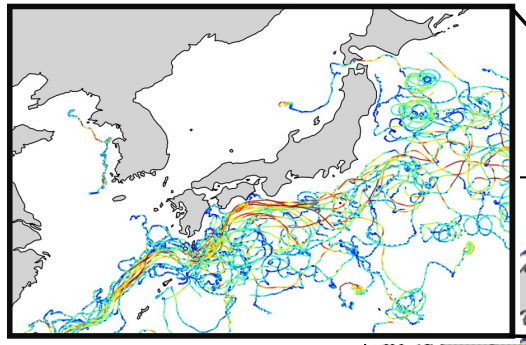
communications physics

Article | Open access | Published: 26 September 2025

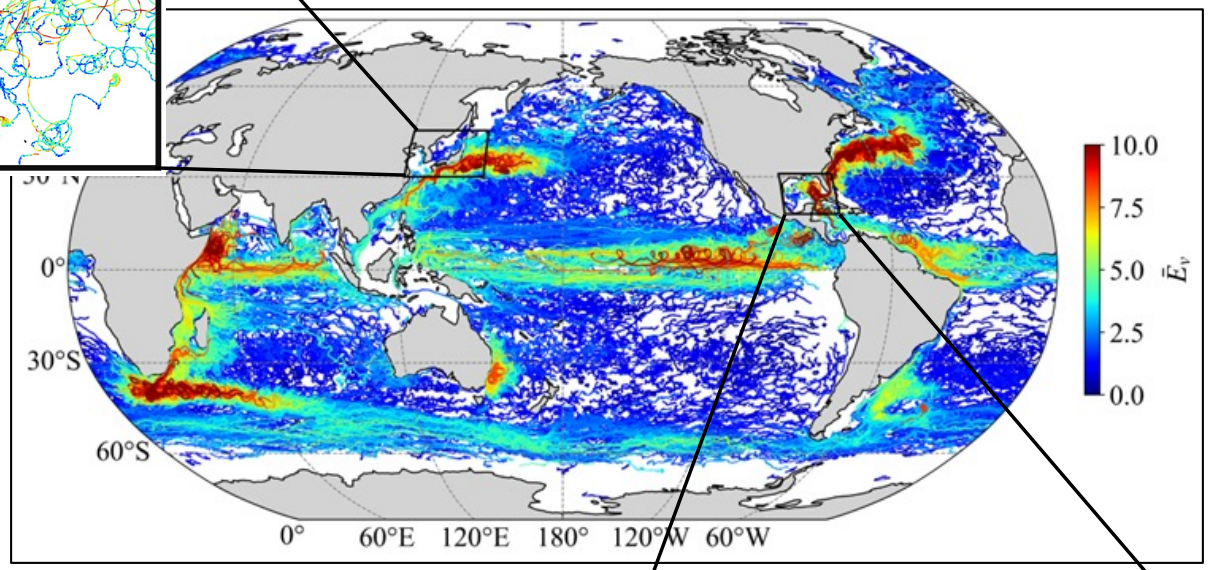
Stochastic reconstruction of gappy Lagrangian turbulent signals by conditional diffusion models

Tianyi Li, Luca Biferale, Fabio Bonaccorso, Michele Buzzicotti ✉ & Luca Centurioni

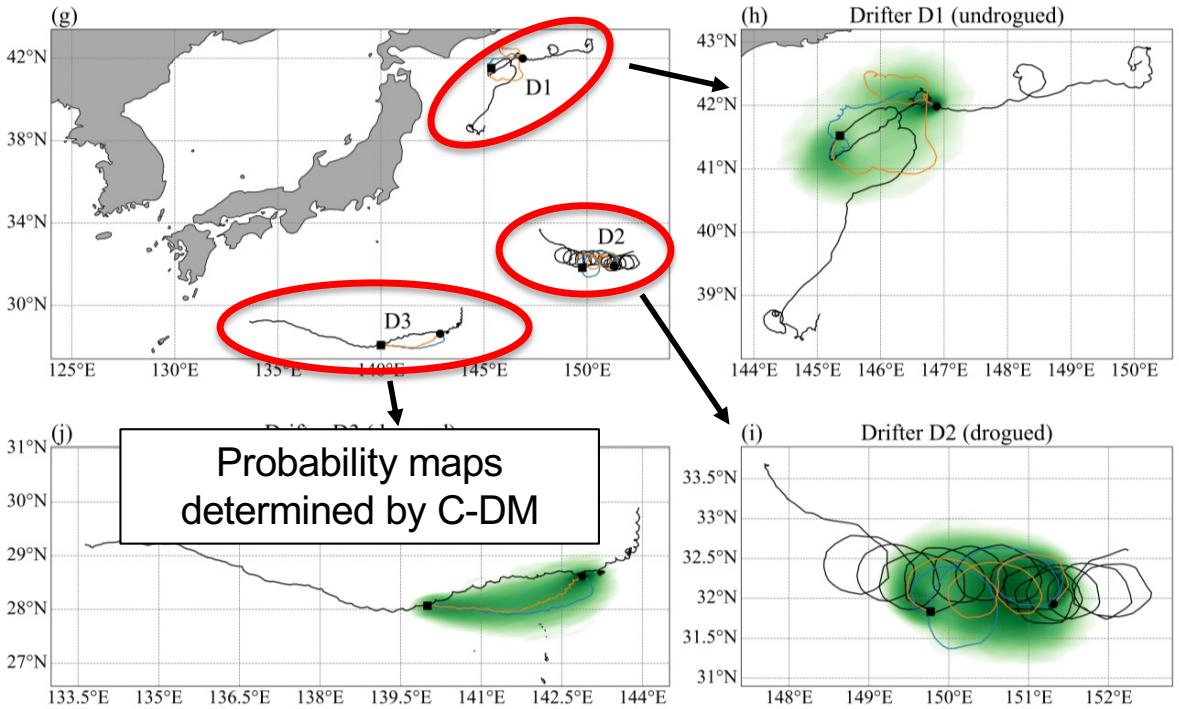
Communications Physics **8**, Article number: 372 (2025) | Cite this article



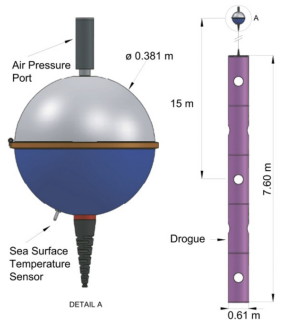
Data: ~115k trajectories 60 days long

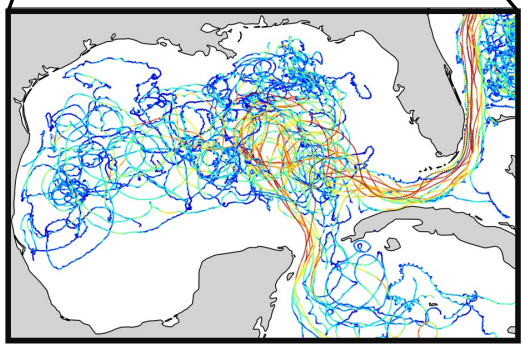


Reconstruction of a 15 days gap



Probability maps determined by C-DM

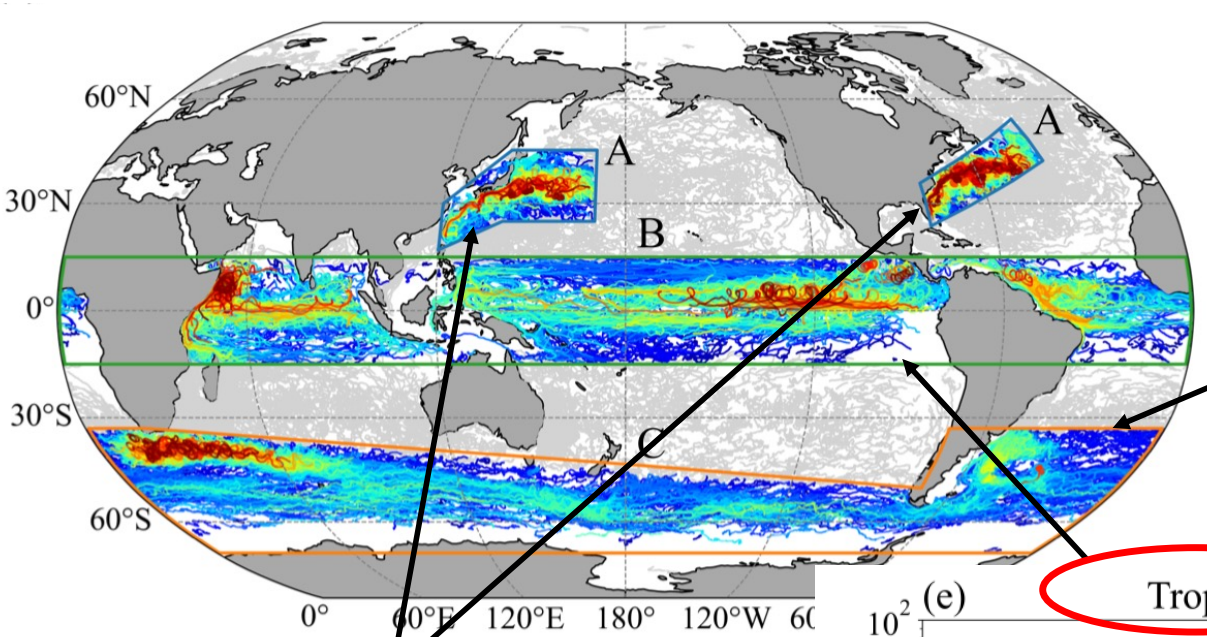




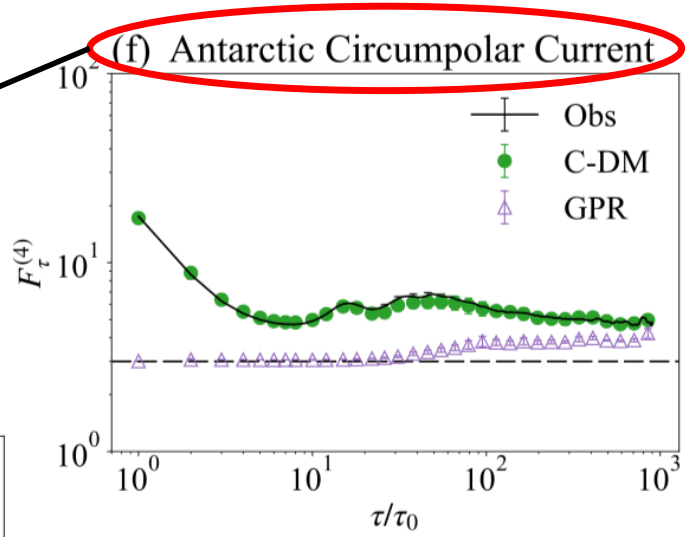
Communications Physics 8.1 (2025): 372.

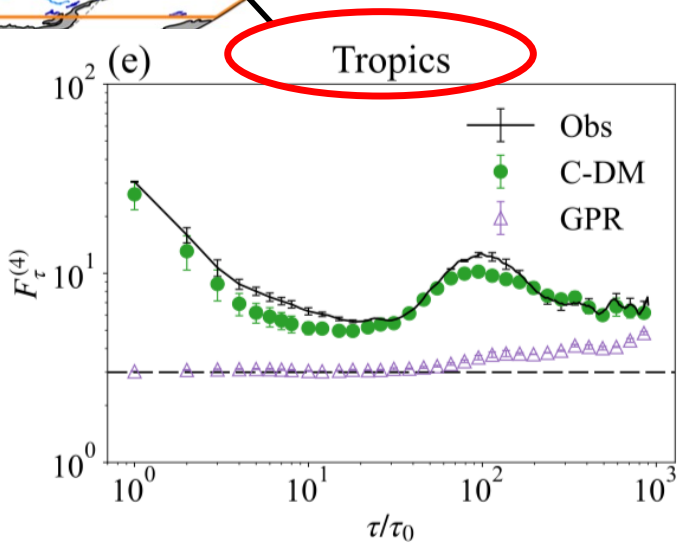
Stochastic Reconstruction of Gappy Lagrangian Turbulent Signals by Conditional Diffusion Models T. Li, L. Biferale, F. Bonaccorso, M. Buzzicotti, L. Centurioni

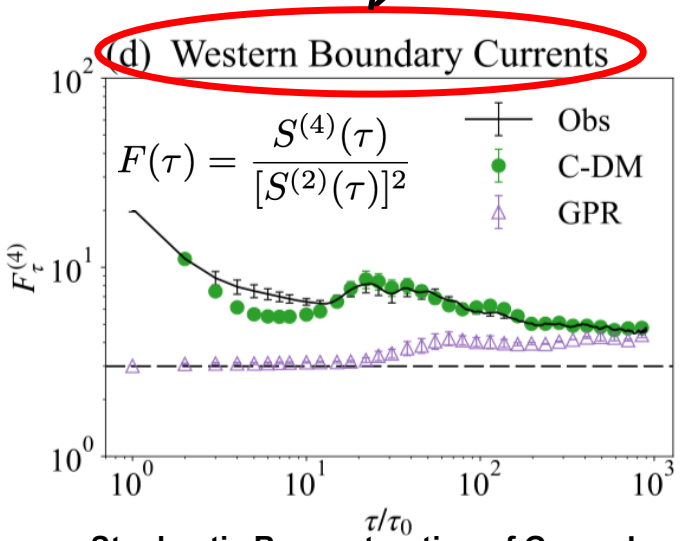
Reconstruction of NOAA's Global Drifter Program Data



II) Statistical reconstruction of Drifter's velocity







C-DM can recover data non-homogeneity In the reconstructed signal

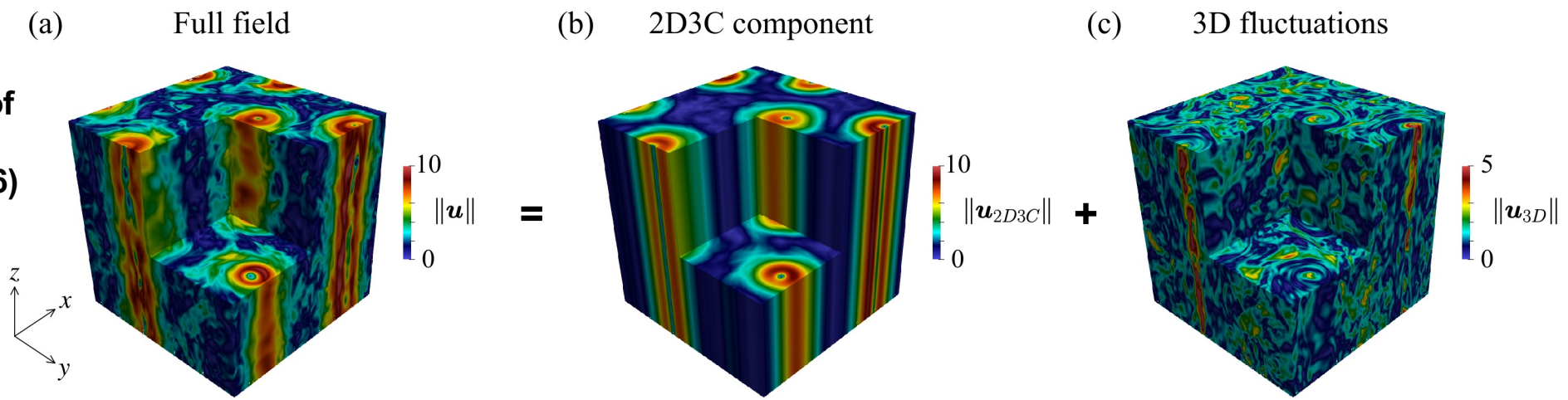
Communications Physics 8.1 (2025): 372.
Stochastic Reconstruction of Gappy Lagrangian Turbulent Signals by Conditional Diffusion Models T. Li, L. Biferale, F. Bonaccorso, M. Buzzicotti, L. Centurioni

3D Rotating Turbulence:

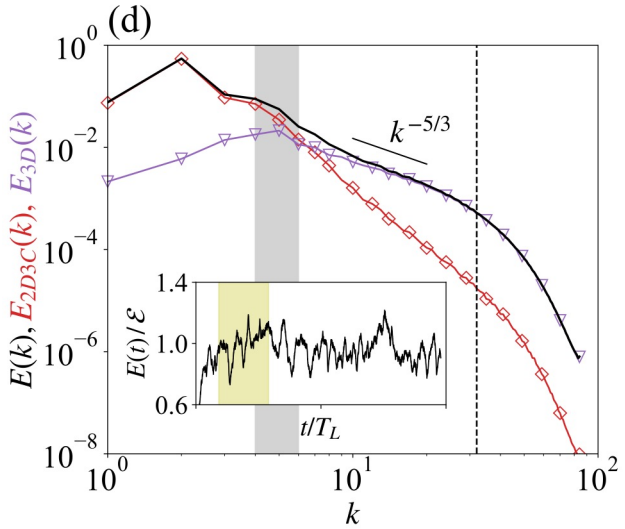
$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + 2\boldsymbol{\Omega}\times\boldsymbol{u} = -\nabla\tilde{p} + \nu\Delta\boldsymbol{u} + \boldsymbol{f}$$

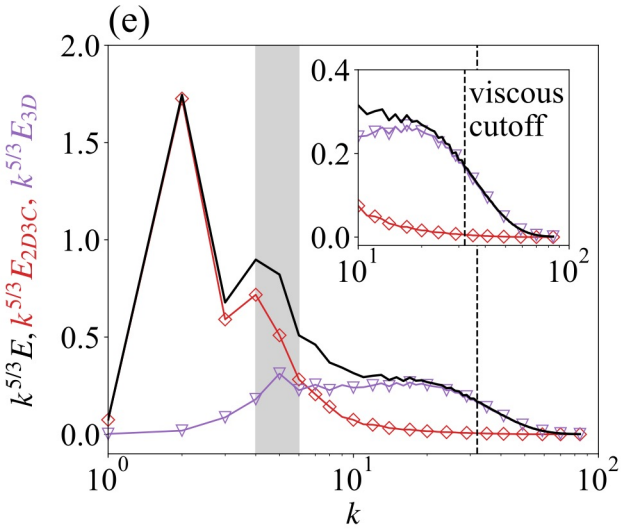
$$\boldsymbol{u}(x,y,z) = \boldsymbol{u}_{2D3C}(x,y) + \boldsymbol{u}_{3D}(x,y,z), \qquad \boldsymbol{u}_{2D3C}(x,y) = \frac{1}{L_0}\int_0^{L_0} \boldsymbol{u}(x,y,z)\,dz$$

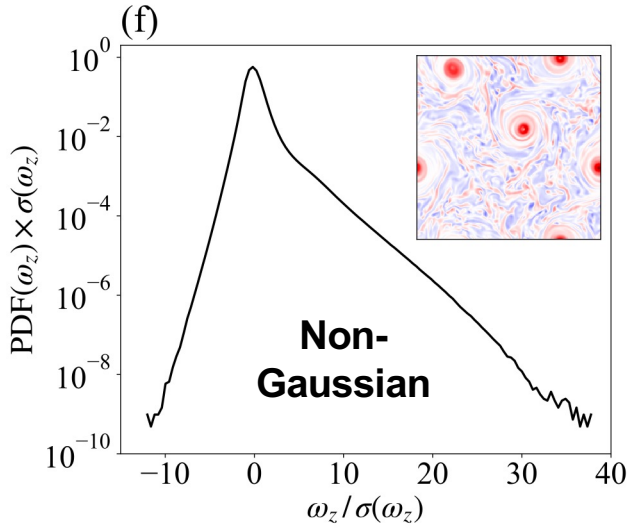
Large Number of Degrees of Freedom (~ 10^6)



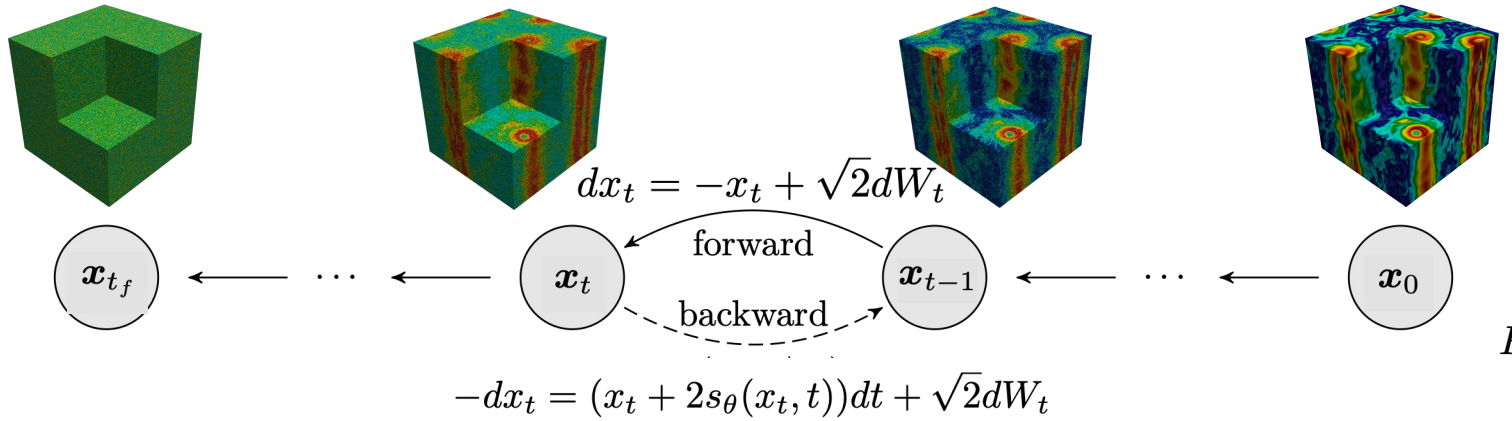
Anisotropic







Diffusion Models: Standard Training / Progressive Training / Physics Constrained



Samples Drawn from:

$P_{t_f}(x_{t_f}) \equiv \mathcal{N}(\mathbf{0}, \mathbf{I})$

Samples Drawn from:

$P_{\text{data}}(x) \approx \{\boldsymbol{u}_i(\boldsymbol{x})\}_{i=1}^{N}$

$$-dx_t = (x_t + 2s_\theta(x_t, t))dt + \sqrt{2}dW_t$$

1) Standard Training

2) Progressive Training

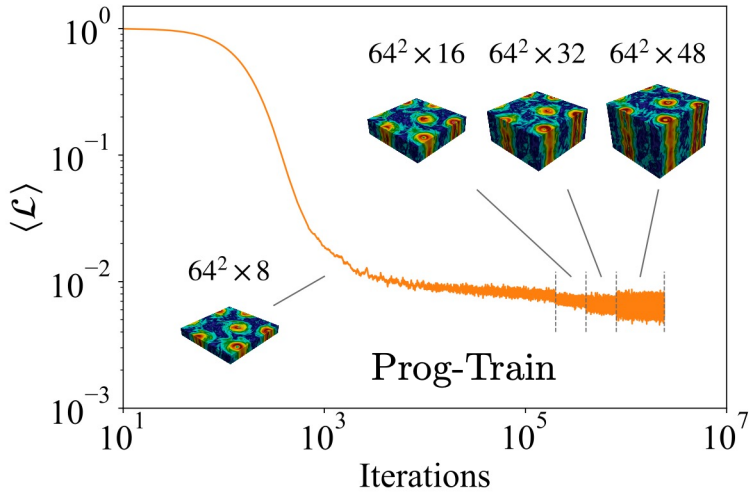
3) Physics-Constrained generation

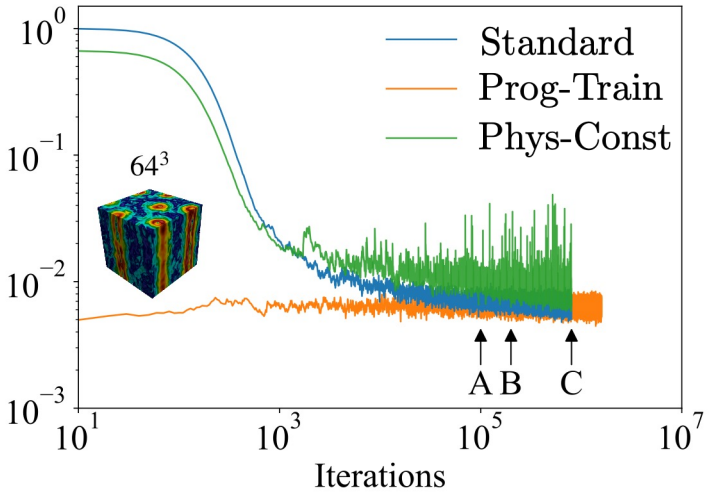
Key idea: Tweedie's Formula

$$\int P_{\text{data}}(x_0 \mid x_t)\, x_0\, dx_0 = e^t\, x_t + \sigma_t^2 \underbrace{\nabla_{x_t} \log P_t(x_t)}_{\approx s_\theta(x,t)}$$

$$\langle x_0 \rangle \approx e^t x_t + \sigma_t^2 s_\theta(x_t, t) = \hat{x}_{\theta,0}$$

$$s_\theta^*(x_t, t) = \frac{1}{\sigma_t^2}(\mathcal{P}(\hat{x}_{\theta,0}) - e^t x_t)$$





We impose the following physics constrains on all the model estimations:

1) Zero Mean 2) Divergence-free

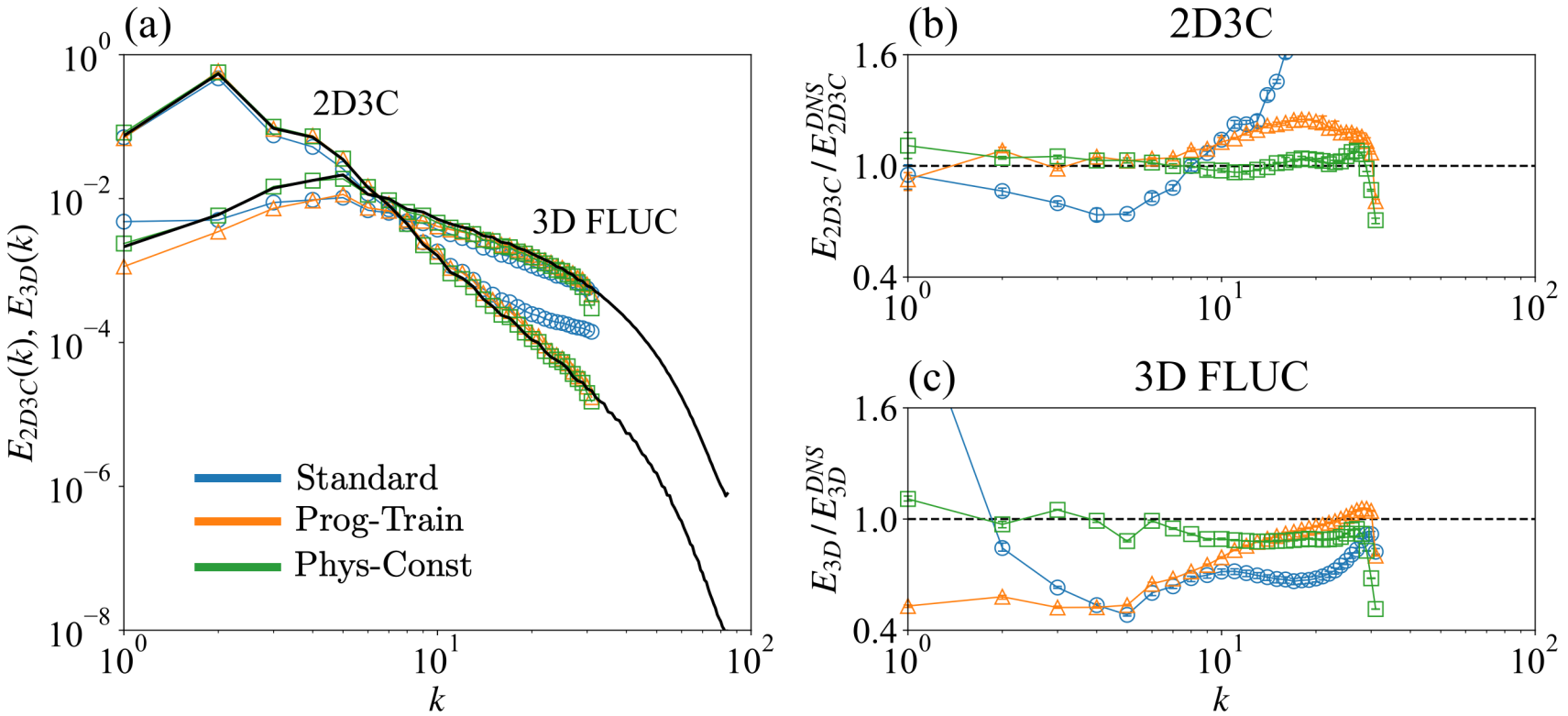
Nature Communications (Accepted, 2026)

Physics-constrained diffusion model for synthesis of 3D turbulent data. T. Li, L. Biferale, F. Bonaccorso, M. Buzzicotti

Score Based Generative Models: Standard / Progressive Training / Physics Constrained

Energy Spectrum

$$E_{3D}(k) = \frac{1}{2} \sum_{\substack{k \le |\boldsymbol{k}'| < k+1 \\ k'_z \neq 0}} \sum_{i \in \{x,y,z\}} \left|\hat{u}_i(\boldsymbol{k}')\right|^2 \longleftarrow E(k) \longrightarrow E_{2D3C}(k) = \frac{1}{2} \sum_{\substack{k \le |\boldsymbol{k}'| < k+1 \\ k'_z = 0}} \sum_{i \in \{x,y,z\}} \left|\hat{u}_i(\boldsymbol{k}')\right|^2$$



Nature Communications (Under revision, 2026)

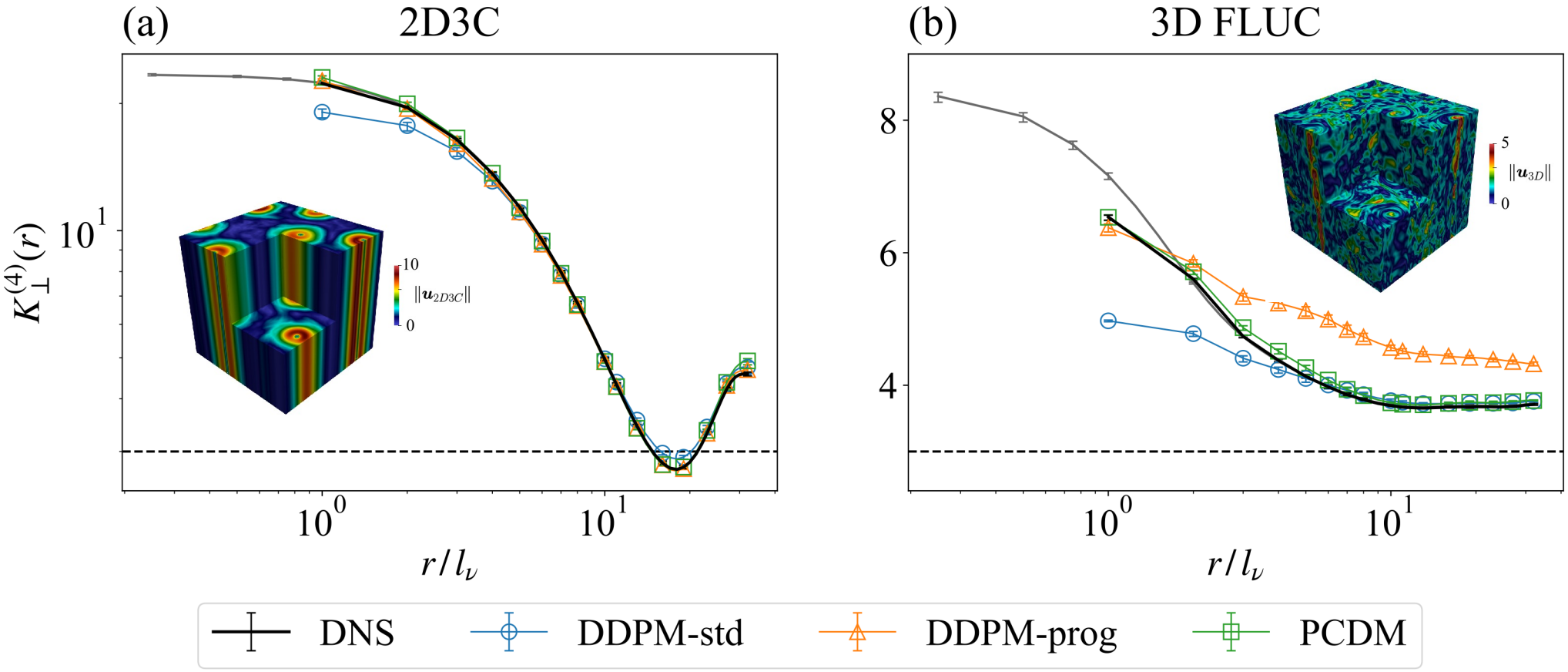
A Generative Diffusion Framework for Physically Consistent 3D Turbulence. T. Li, L. Biferale, F. Bonaccorso, M. Buzzicotti

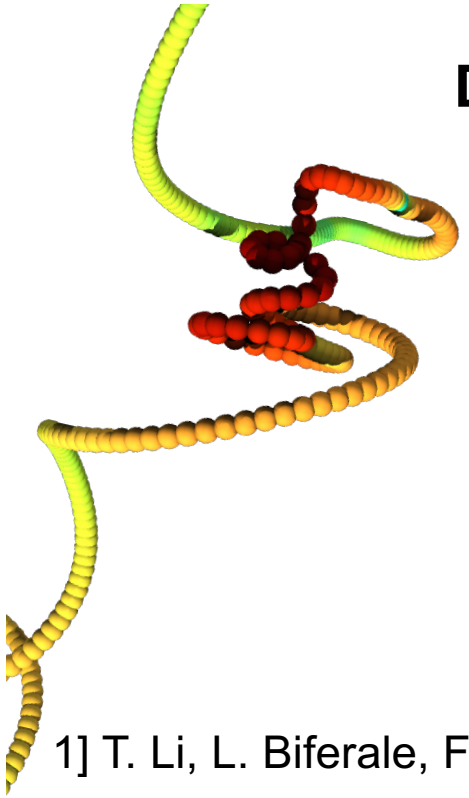
Score Based Generative Models: Standard / Progressive Training / Physics Constrained

Kurtosis of velocity increments

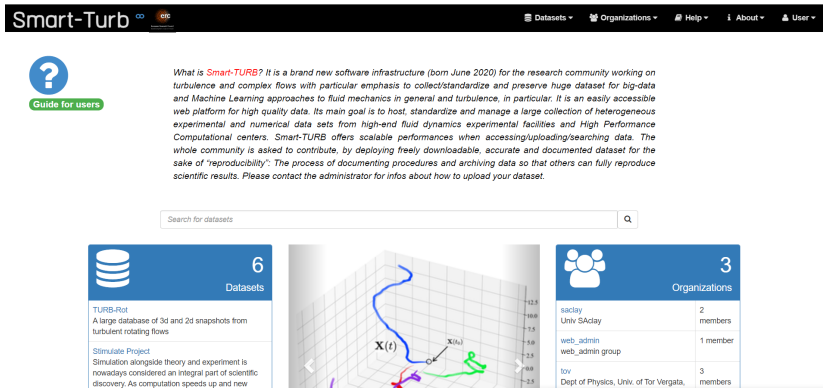
$\delta u_\perp(\boldsymbol{r}) = [\boldsymbol{u}(\boldsymbol{x}+\boldsymbol{r}) - \boldsymbol{u}(\boldsymbol{x})] \cdot \hat{\boldsymbol{t}}$ → orthogonal to both, $\boldsymbol{r}$, and the rotation axis

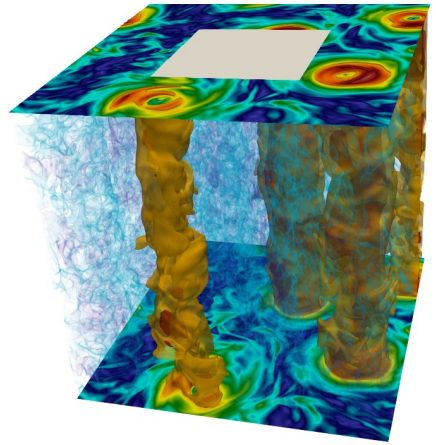
Kurtosis: $K_\perp^{(4)}(r) = \dfrac{\langle [\delta u_\perp(r)]^4 \rangle}{\langle [\delta u_\perp(r)]^2 \rangle^2}$





Data & code are open source: smart-turb.roma2.infn.it





our works on Lagrangian & Eulerian data generation

1] T. Li, L. Biferale, F. Bonaccorso, M. A. Scarpolini & M. Buzzicotti, Synthetic Lagrangian turbulence by generative diffusion models. *Nature Machine Intelligence*, 6(4), 393-403.

2] T. Li, L. Biferale, F. Bonaccorso, M. Buzzicotti & L. Centurioni. Stochastic reconstruction of gappy Lagrangian turbulent signals by conditional diffusion models. *Communications Physics*, *8*(1), 372.

3] Li, T., Buzzicotti, M., Bonaccorso, F., & Biferale, L. (2026). Physics-constrained diffusion model for synthesis of 3D turbulent data. arXiv preprint arXiv:2603.12834. (*Accepted on Nature Communications*).

4] H. Narula, T. Li, M. Buzzicotti, L. Biferale & P. Perlekar. (2026). Gappy Reconstruction of Bubbly Flows by Guided Diffusion Models. arXiv:2606.29843. (*Accepted on PRX Intelligence*).