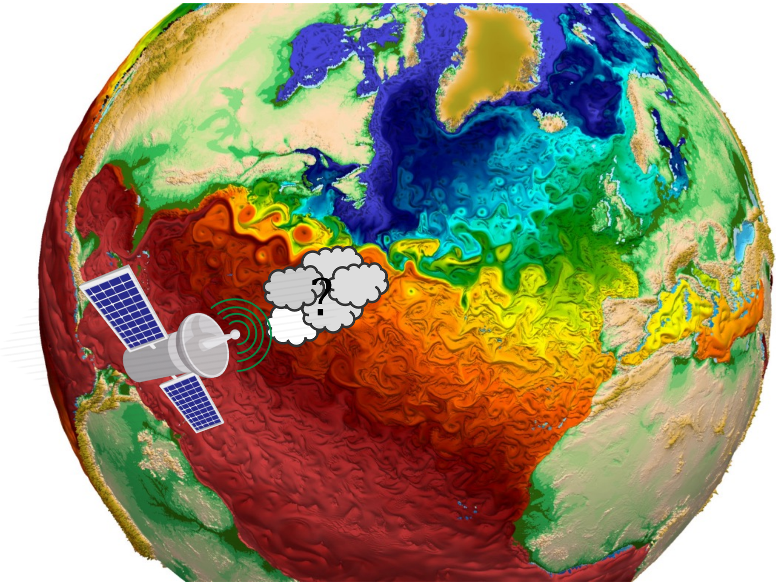
AQTIVATE

WORKSHOP: Data-Driven and Model-Based Tools For Complex Flows and Complex Fluids
Rome - Dept. Physics University of Rome Tor Vergata June/3-7/2024

Stochastic Multi-Scale Reconstruction of Turbulent Flows with Data-Driven Generative Models



Los Alamos National Laboratory (CC BY-NC-ND 2.0)

Michele Buzzicotti

Dept. Physics & INFN, University of Rome Tor Vergata

https://mbuzzico.github.io/

michele.buzzicotti@roma2.infn.it

Luca Biferale (Uni. Tor Vergata, IT),
Fabio Bonaccorso, **Tianyi Li** (Uni. Tor Vergata, IT),
Hussein Aluie, Ben Storer (Uni. Rochester, NY),
Alessandra Lanotte (CNR-Nanotec, IT)

TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

INFN

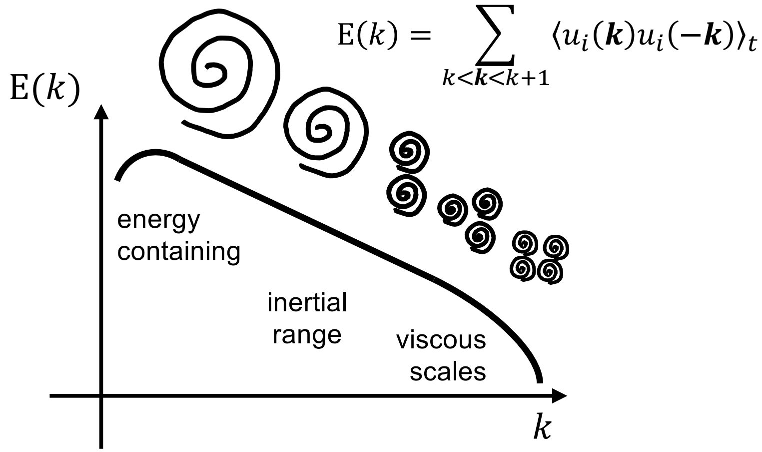
MUR
Ministero dell'Università e della Ricerca

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

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

Turbulent flows

A multiscale problem

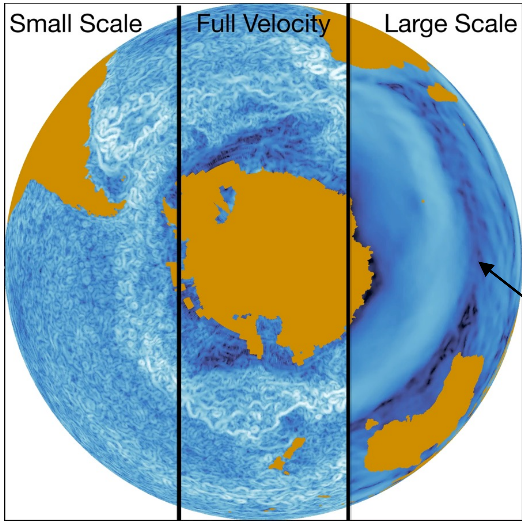


“From a distance complex objects can appear simple, but a closer look always reveals structures. We are awash in complexity at every scale we observe {…} As scientists we ask how much detail is required to explain the observations? Must we track every water molecule to explain the ocean? What leanghtscales are important to understand?”

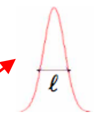
Announcement of the 2021 Nobel Prize in Physics

Coarse graining;
a tool for decomposing complex flows

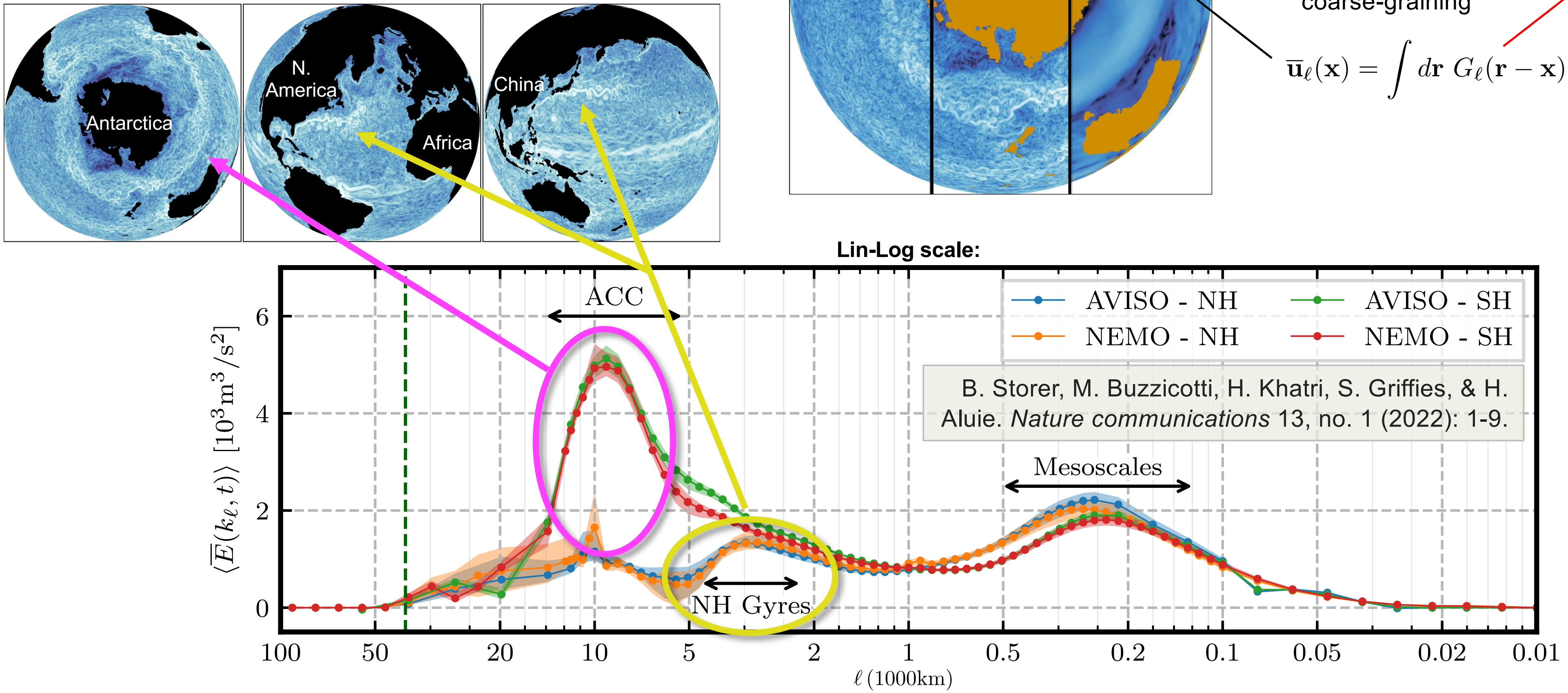
How can we do multiscale flow decomposition on complex geometries?



"coarse-graining"



$$\overline{\mathbf{u}}_\ell(\mathbf{x}) = \int d\mathbf{r}\; G_\ell(\mathbf{r}-\mathbf{x})\,\mathbf{u}(\mathbf{r})$$

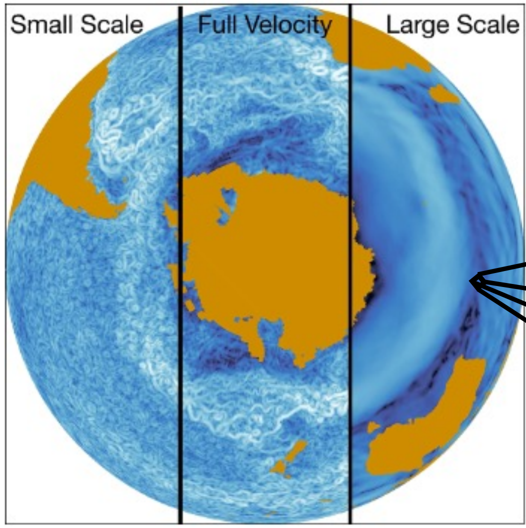


Coarse graining;
a tool for decomposing complex flows

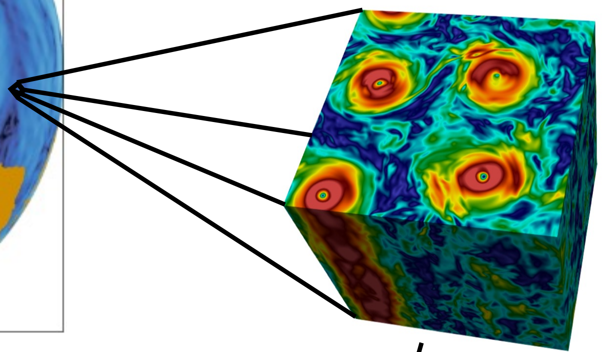
"Small-scales" ($\ell < 100\ km$) are neither resolved nor understood (3D Turbulence)

Turbulence is the most important unsolved problem of statistical classical physics "R. Feynman"

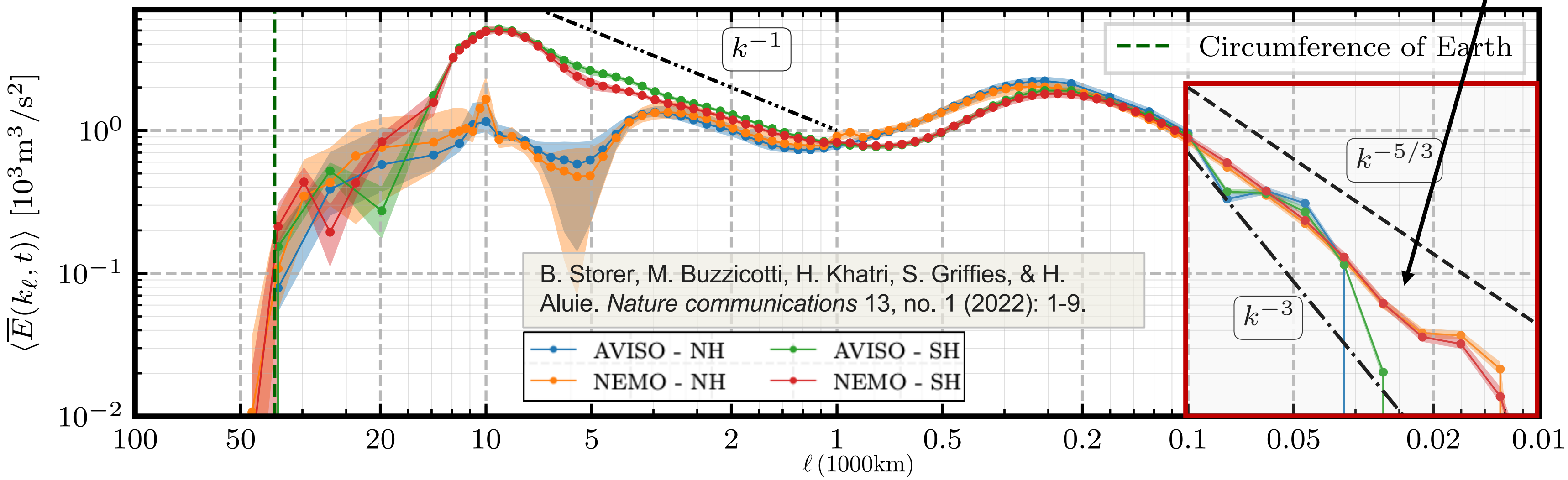
- **Non-linear Dynamics**
- **Non-perturbative out-of-equilibrium Statistical Field Theory**
- **Chaotic Dynamics**
- **Anomalous Scaling**
- **Billions/Trillions of Degrees of Freedom**



3D Turbulence
Non-Gaussian fluctuations over
~ several decades of scales



Log-Log scale:



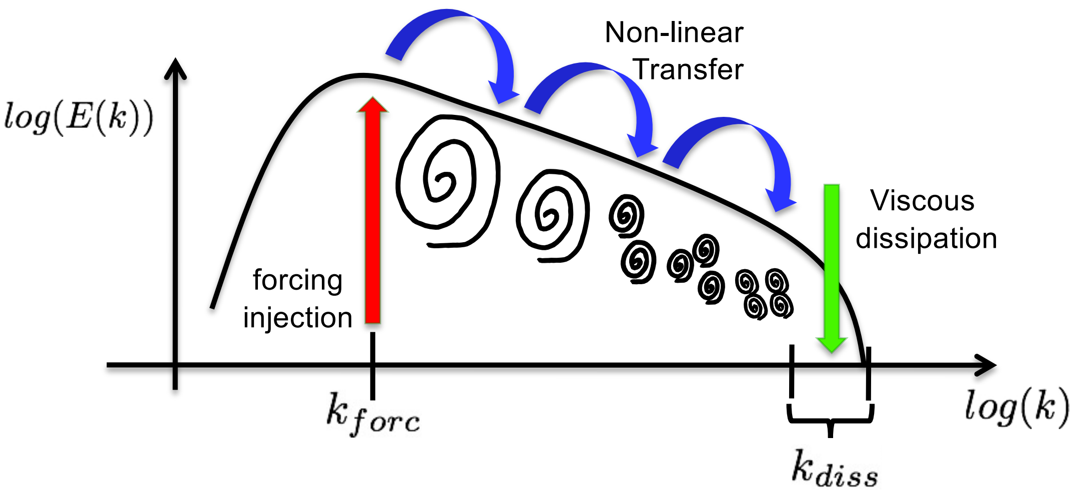
Leonardo da Vinci (~ 1500): “doue la turbolenza dell acqua si genera **[injected]**; doue la turbolenza dell acqua si mantiene **[advected]** plugho; doue la turbolenza dell acqua si posa **[dissipated]**”

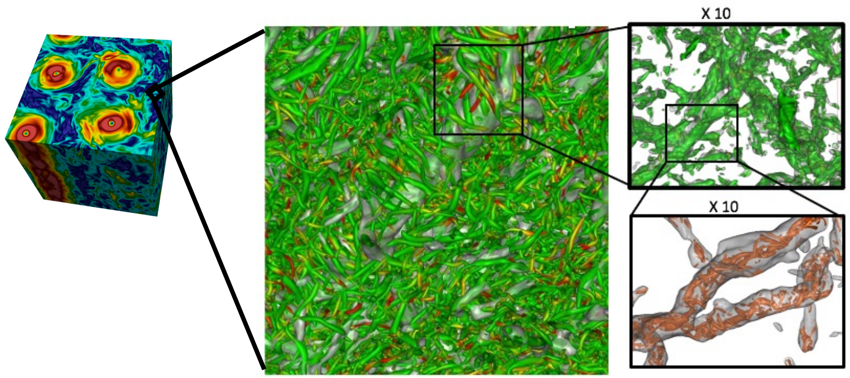
Navier-Stokes Equations:

$$m\vec{a} = \vec{F}$$

$$\partial_t \vec{v} + (\vec{v} \cdot \vec{\partial})\vec{v} = -\vec{\partial} P + \nu \Delta \vec{v} + \vec{f}$$

acceleration pressure viscosity external forcing





Multiscale nature of 3D Turbulent flows
K. Burger et al #84174 APS DFD. ***2015***

Leonardo da Vinci (~ 1500): "doue la turbolenza dell acqua si genera **[injected]**; doue la turbolenza dell acqua si mantiene **[advected]** plugho; doue la turbolenza dell acqua si posa **[dissipated]**"

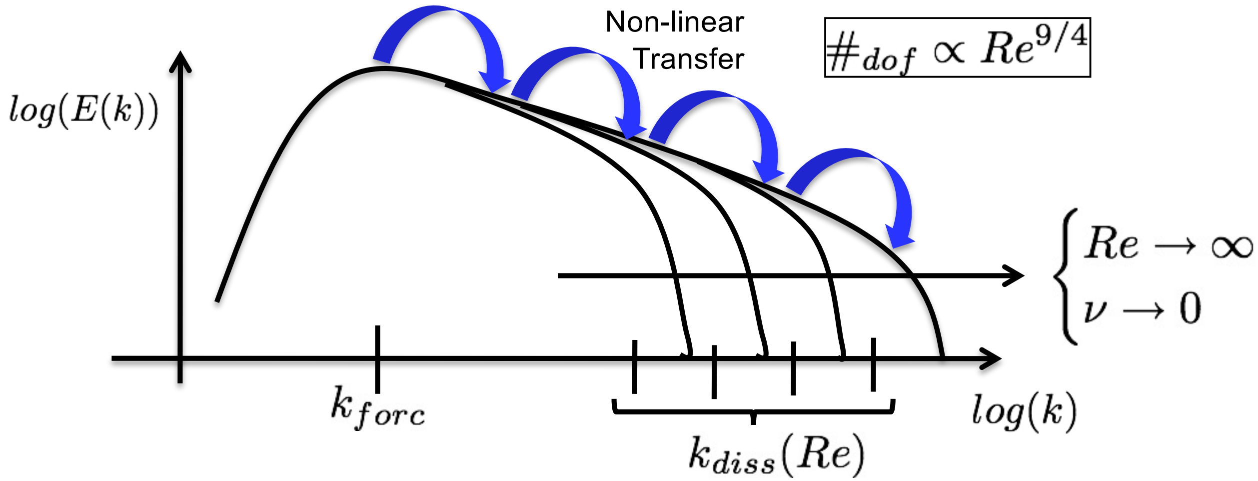
Navier-Stokes Equations:

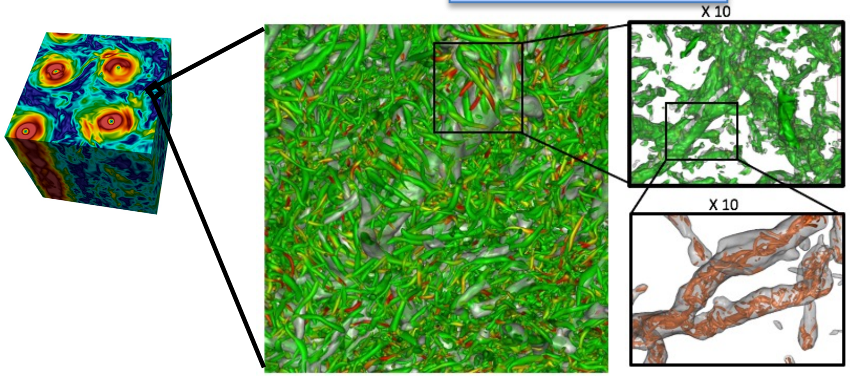
$$\begin{cases} \hat{t} = t/t_0 \\ \hat{x} = x/l_0 \\ \hat{v} = v/v_0 \end{cases}$$

control parameter: $\boxed{Re = \frac{l_0 v_0}{\nu}}$

$$\partial_t \vec{v} + (\vec{v} \cdot \vec{\partial})\vec{v} = -\vec{\partial} P + \frac{1}{Re} \Delta \vec{v} + \vec{f}$$

Multi-Scale!





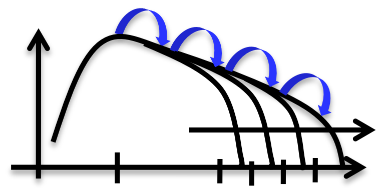
Multiscale nature of 3D Turbulent flows
K. Burger et al #84174 APS DFD. ***2015***

Navier-Stokes Equations:

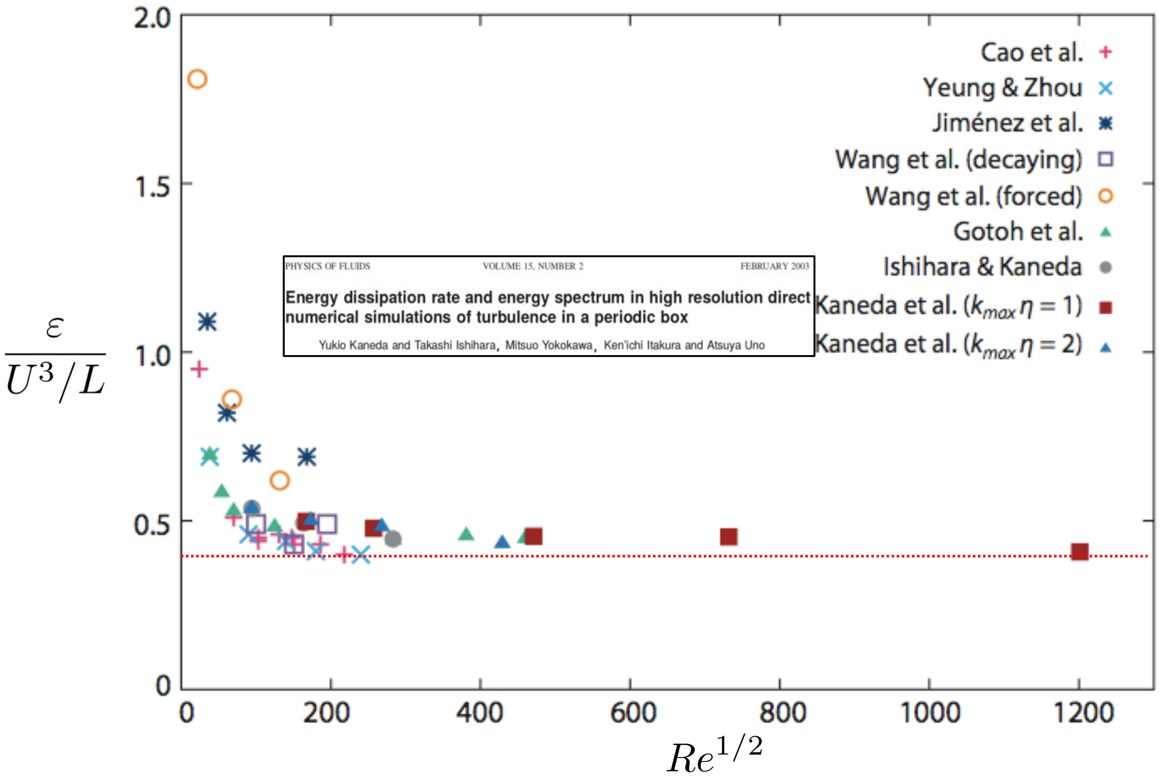
$$\boxed{Re = \frac{l_0 v_0}{\nu}} \qquad \partial_t \vec{v} + (\vec{v}\cdot\vec{\partial})\vec{v} = -\vec{\partial}P + \frac{1}{Re}\Delta\vec{v} + \vec{f}$$

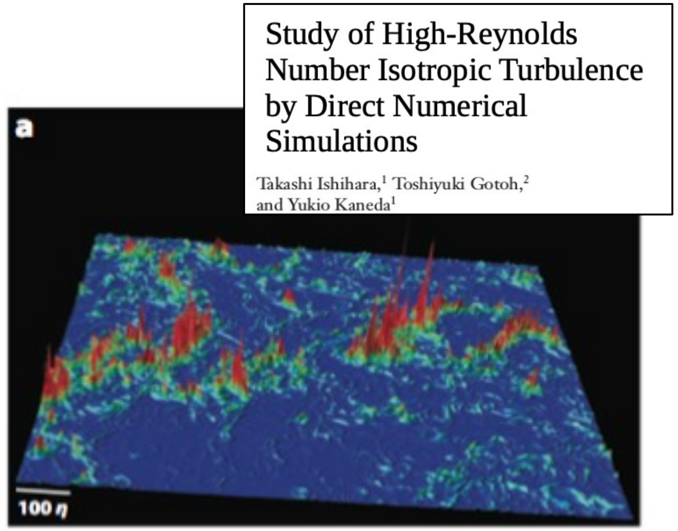
ENERGY DISSIPATION

$$\lim_{Re\to\infty} \epsilon = \lim_{\nu\to 0} \nu\langle(\partial v)^2\rangle \to const.$$



$$\begin{cases} Re \to \infty \\ \nu \to 0 \end{cases}$$

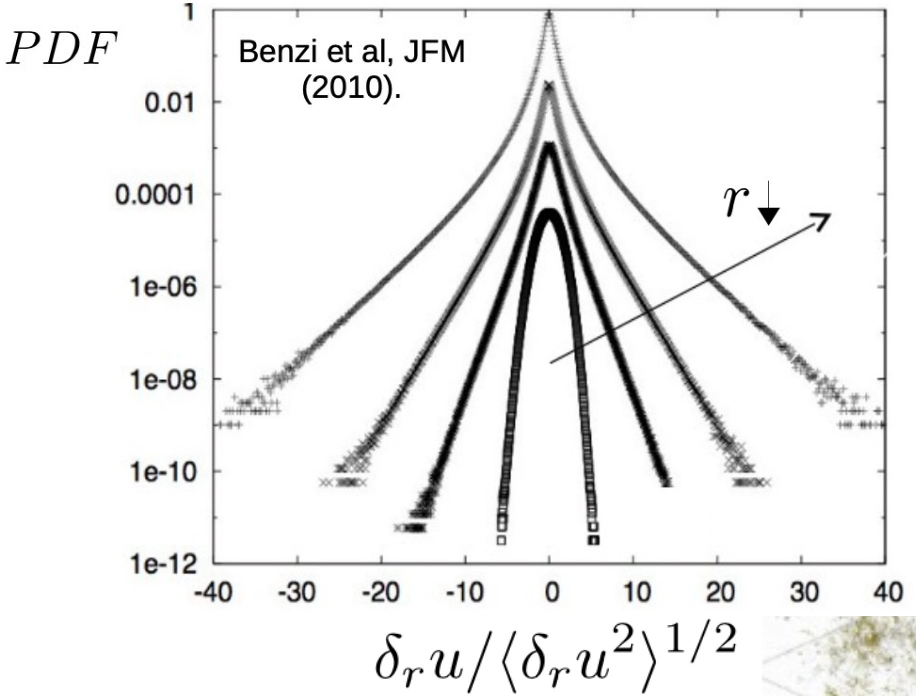


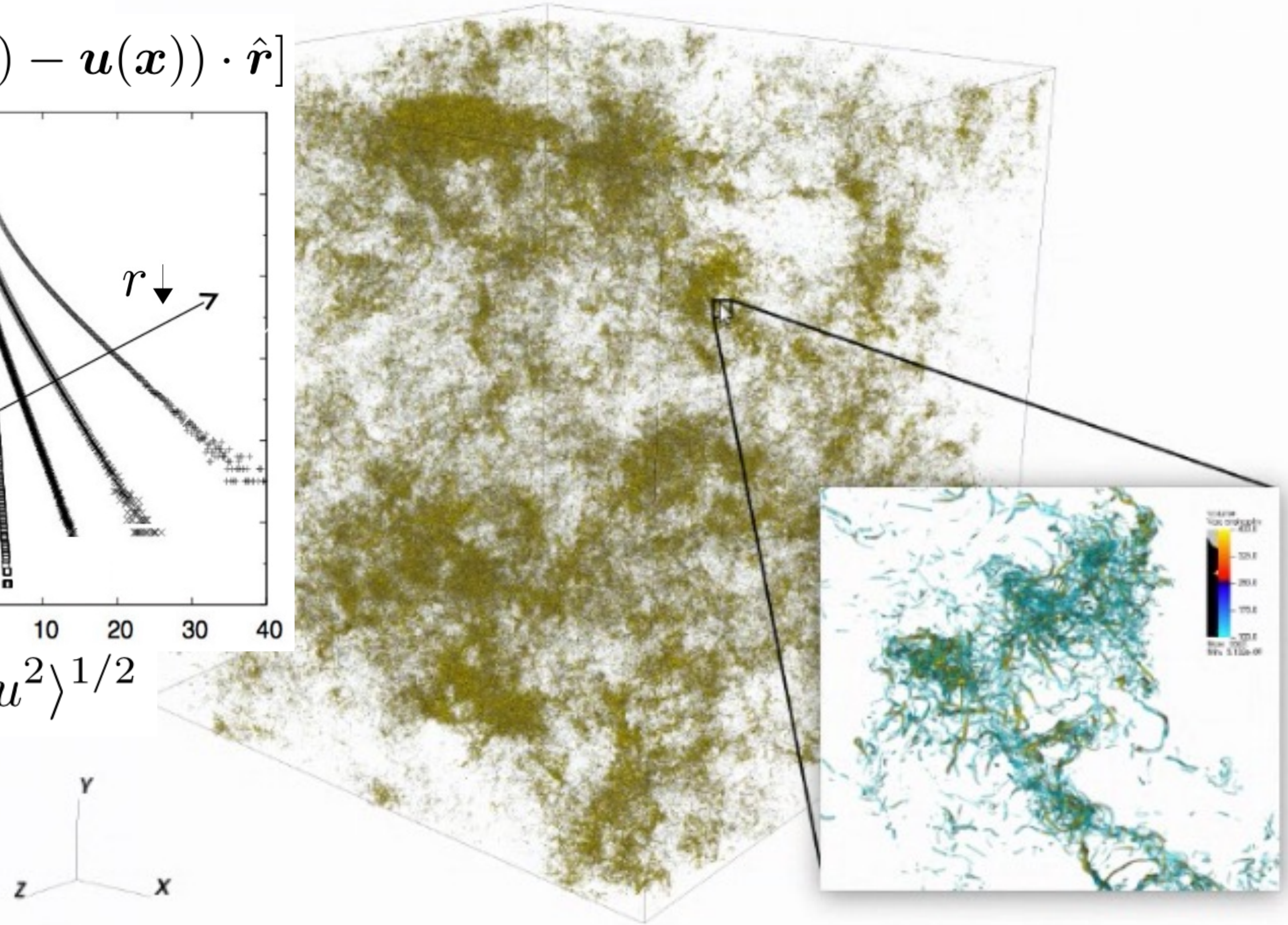


Dissipative Anomaly!

Multiscale Statistical Properties

$$\delta_r u = [(\boldsymbol{u}(\boldsymbol{x}+\boldsymbol{r}) - \boldsymbol{u}(\boldsymbol{x})) \cdot \hat{\boldsymbol{r}}]$$





Intermittency

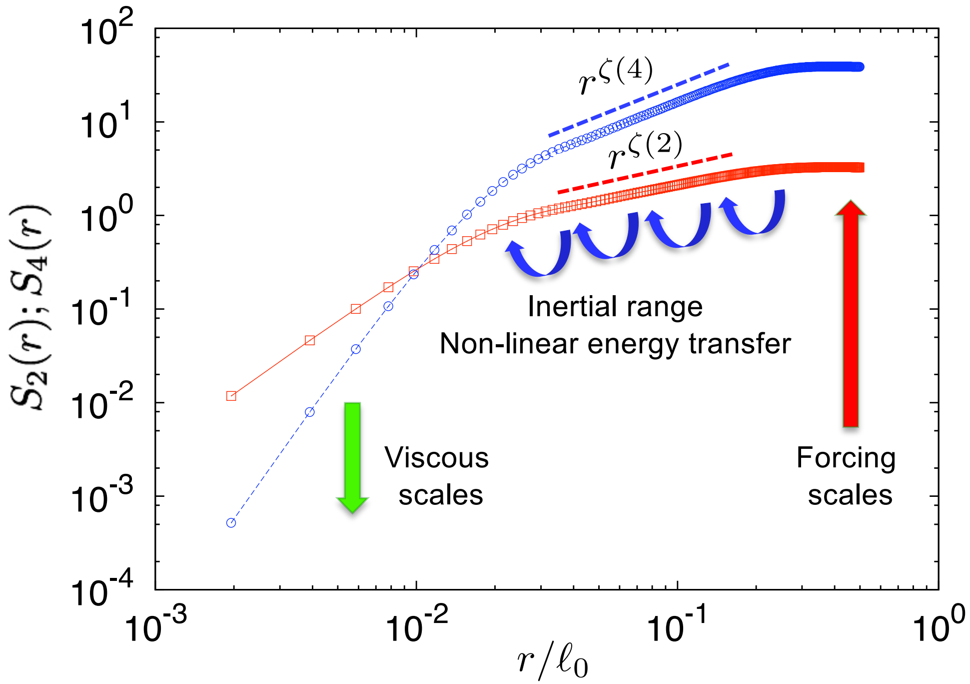
Non-Self-Similar Statistics

Multiscale Statistical Properties

$$\delta_r u = [(\boldsymbol{u}(\boldsymbol{x}+\boldsymbol{r}) - \boldsymbol{u}(\boldsymbol{x})) \cdot \hat{\boldsymbol{r}}]$$

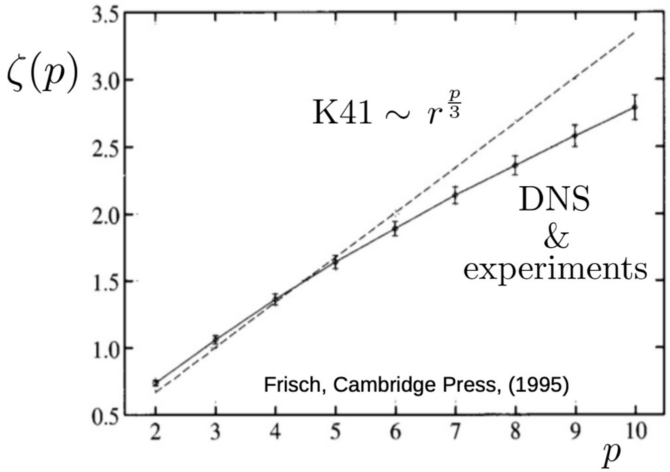
Structure Functions

$$S_p(r) = \langle (\delta_r u)^p \rangle \sim r^{\zeta(p)}$$



Inertial Range, Local Slopes

$$\zeta(p) = \frac{d \log S_p(r)}{d \log(r)}$$

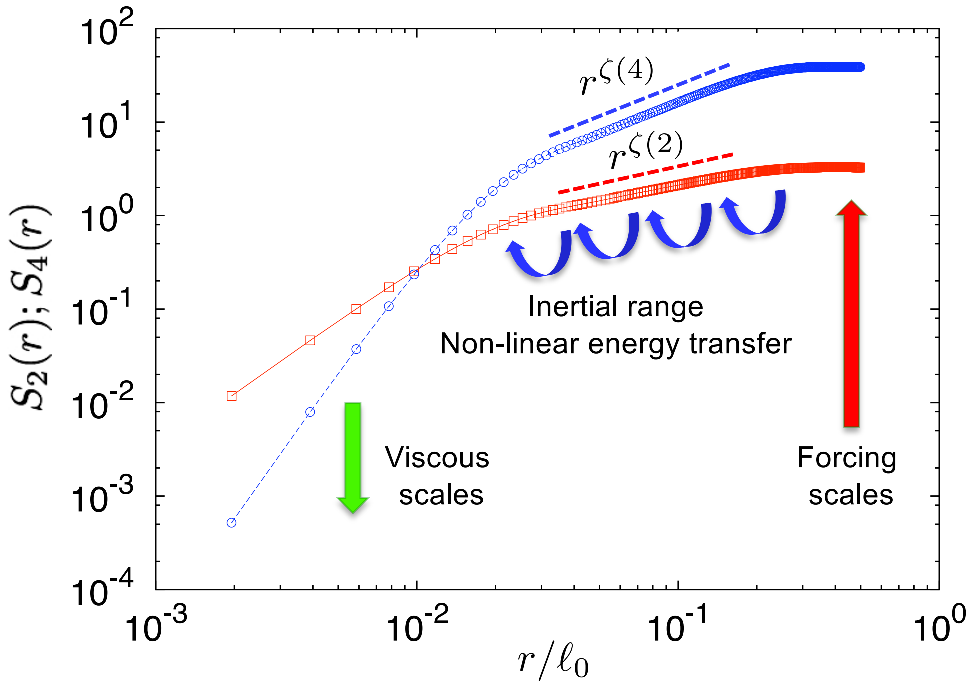


Multiscale Statistical Properties

$$\delta_r u = [(\boldsymbol{u}(\boldsymbol{x}+\boldsymbol{r}) - \boldsymbol{u}(\boldsymbol{x})) \cdot \hat{\boldsymbol{r}}]$$

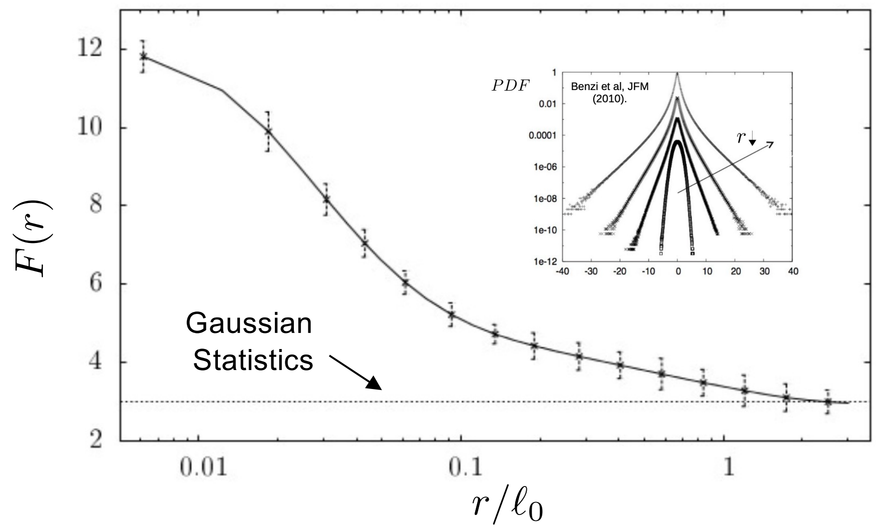
Structure Functions

$$S_p(r) = \langle (\delta_r u)^p \rangle \sim r^{\zeta(p)}$$

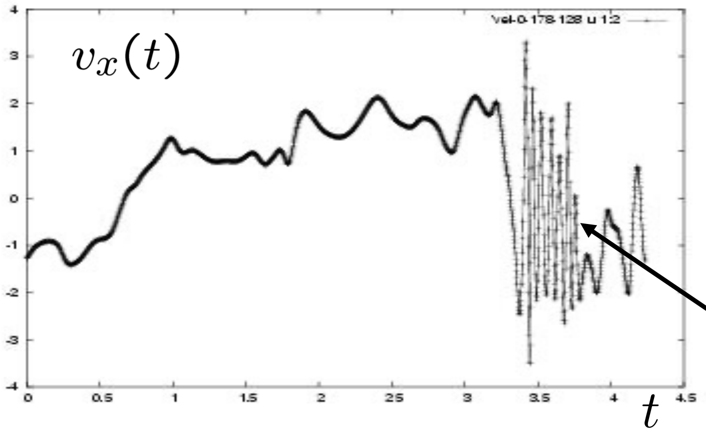


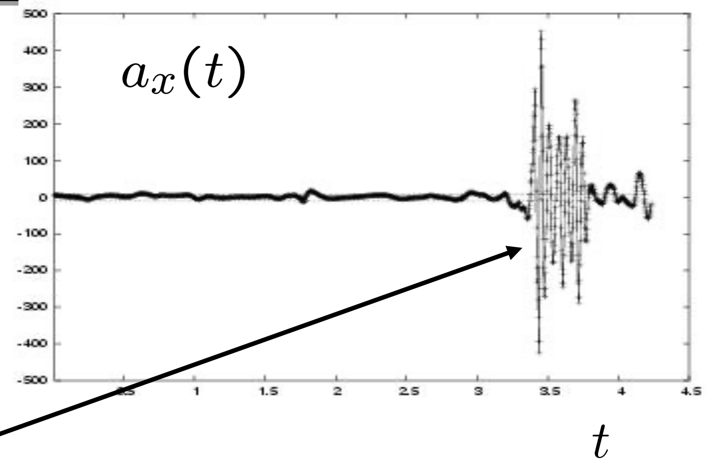
Flatness

$$F(r) = \frac{\langle (\delta_r u)^4 \rangle}{\langle (\delta_r u)^2 \rangle^2}$$

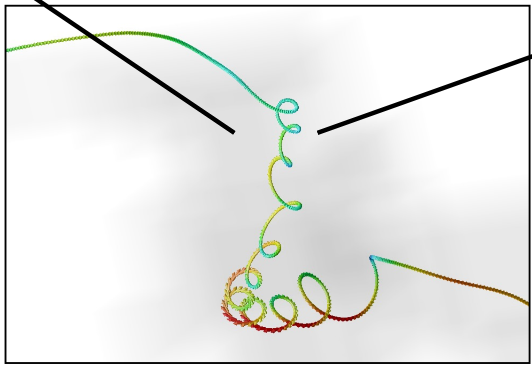


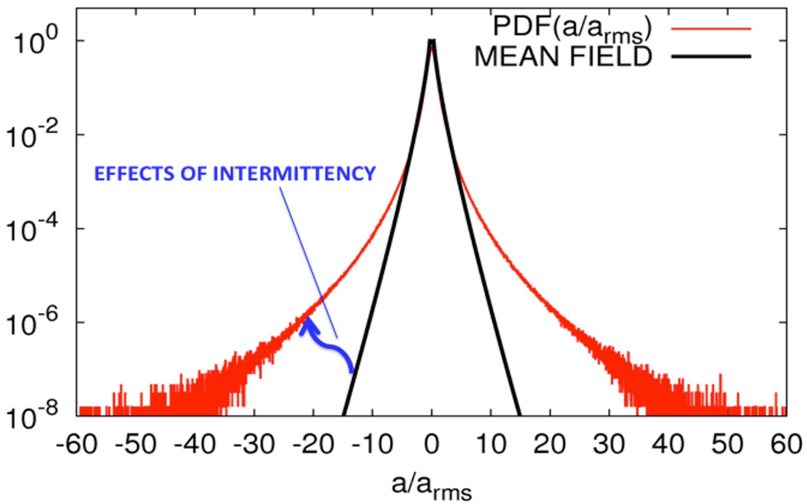
Lagrangian Turbulence





EXTREME EVENTS

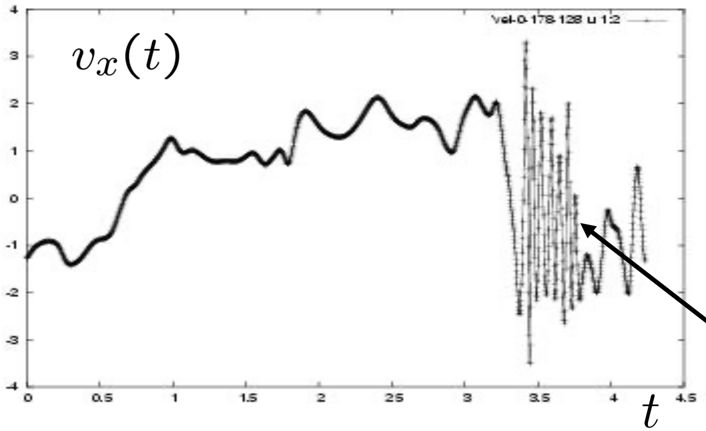




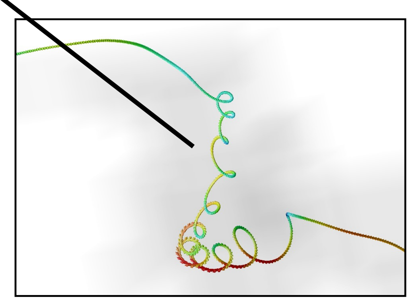
Literature:

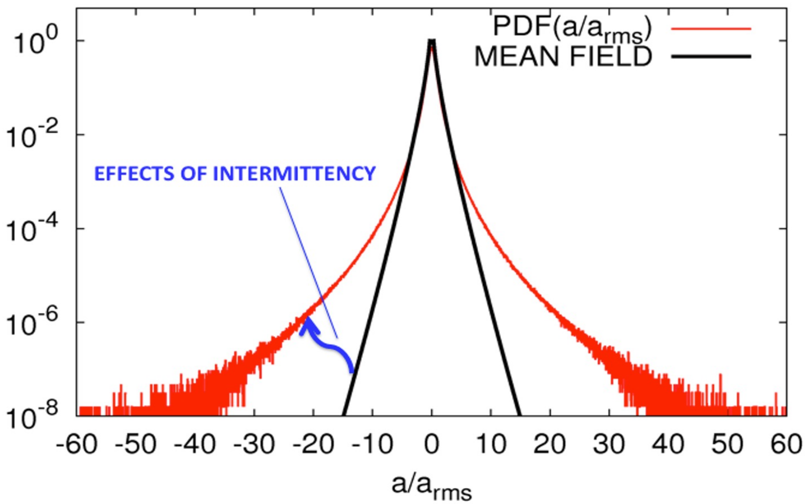
- F. Toschi and E. Bodenschatz. Lagrangian Properties of Particles in Turbulence. Annu. Rev. Fluid Mech. 41, 375 (2009)
- L Biferale, G Boffetta, A Celani, A Lanotte, F Toschi. Particle trapping in three-dimensional fully developed turbulence Physics of Fluids 17 (2), 021701 (2005)
- La Porta, G.A. Voth, A.M. Crawford, J. Alexander et al. Fluid particle accelerations in fully developed turbulence. Nature, 409(6823), 1017 (2001)
- N. Mordant, P. Metz, O. Michel and J.F. Pinton. Measurement of Lagrangian velocity in fully developed turbulence. Phys. Rev. Lett. 87(21), 214501 (2001)

Lagrangian Turbulence

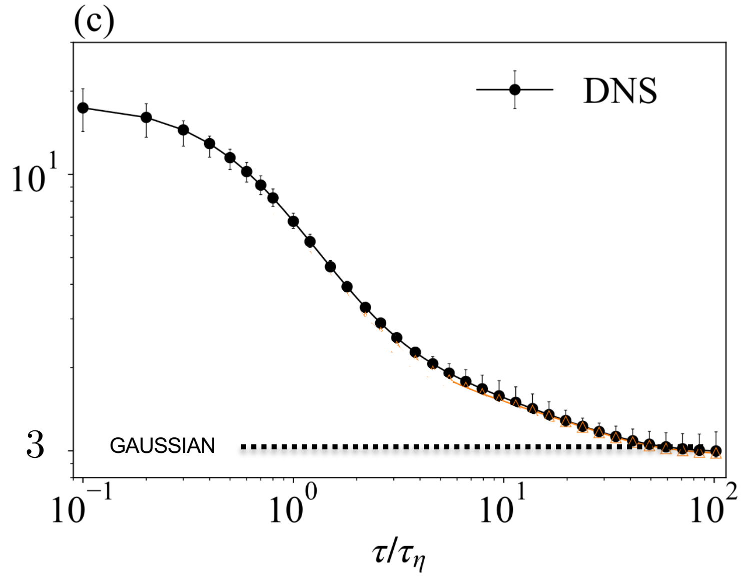


EXTREME EVENTS





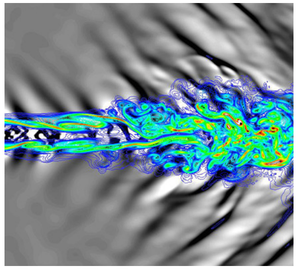
INTERMITTENCY



$$\delta_\tau V_i(t) = V_i(t+\tau) - V_i(t), \qquad S_\tau^{(p)} = \langle (\delta_\tau V_i)^p \rangle$$

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

Typical Reynolds numbers under real world conditions



laboratory flow

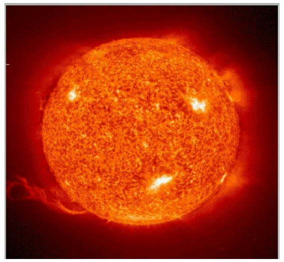
$$Re \sim 10^5 - 10^9$$

$$\#_{dof} \sim 10^{11} - 10^{20}$$

geophysical flow

$$Re \sim 10^8 - 10^{12}$$

$$\#_{dof} \sim 10^{18} - 10^{30}$$



astrophysical flow

$$Re > 10^{15}$$

$$\#_{dof} \sim \infty$$

state-of-the-art Direct Numerical Simulation:

Isotropic, homogeneous Fully Periodic Flows
Pseudo-Spectral Methods.
Resolution 12000^3 (Y. Kaneda, APS 2017)

Reynolds : 10^8,
Storage of 1 velocity configuration (double precision): 40 Tbyte
RAM requirements for time marching ~ 160 Tbyte

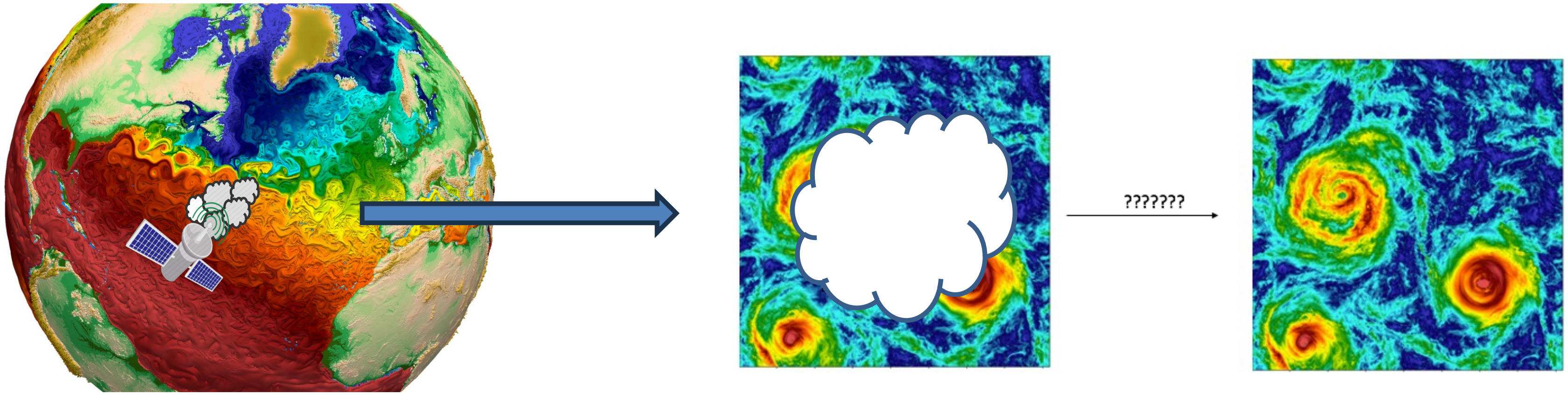
Moral: brute force Direct Numerical Simulations
able to saturate any computing power
(present and/or future EXA-SCALE -> ZETTA-SCALE): Computo ergo sum?

J. von NEUMANN (1949)

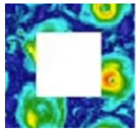
These considerations justify the view that a considerable mathematical effort towards a detailed understanding of the mechanism of turbulence is called for. The entire experience with the subject indicates that the purely analytical approach is beset with difficulties, which at this moment are still prohibitive. The reason for this is probably as was indicated above: That our intuitive relationship to the subject is still too loose — not having succeeded at anything like deep mathematical penetration in any part of the subject, we are still quite disoriented as to the relevant factors, and as to the proper analytical machinery to be used.

Under these conditions there might be some hope to 'break the deadlock' by extensive, but well-planned, computational efforts. It must be admitted that the problems in question are too vast to be solved by a direct computational attack, that is, by an outright calculation of a representative family of special cases. There are, however, strong indications that one could name certain strategic points in this complex, where relevant information must be obtained by direct calculations. If this is properly done, and the operation is then repeated on the basis of broader information then becoming available, etc., there is a reasonable chance of effecting real penetrations in this complex of problems and gradually developing a useful, intuitive relationship to it. This should, in the end, make an attack with analytical methods, that is truly more mathematical, possible.[1]

Turbulent Data GENERATION, FULL STATE RECONSTRUCTION

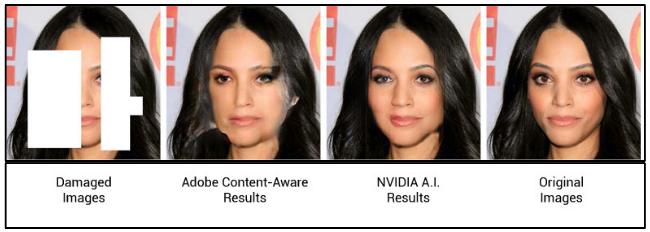


EQ. BASED

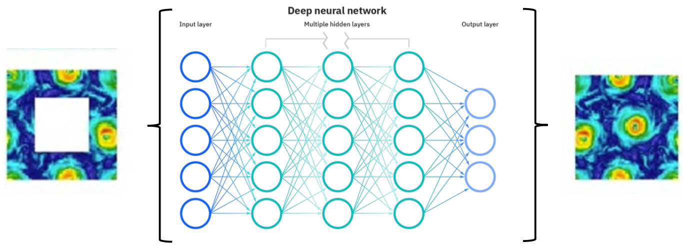
$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla}\boldsymbol{v} + 2\Omega\hat{x}_3\times\boldsymbol{v} = -\boldsymbol{\nabla} p + \nu\nabla^2\boldsymbol{v} +$$ N(v - )

Nudging

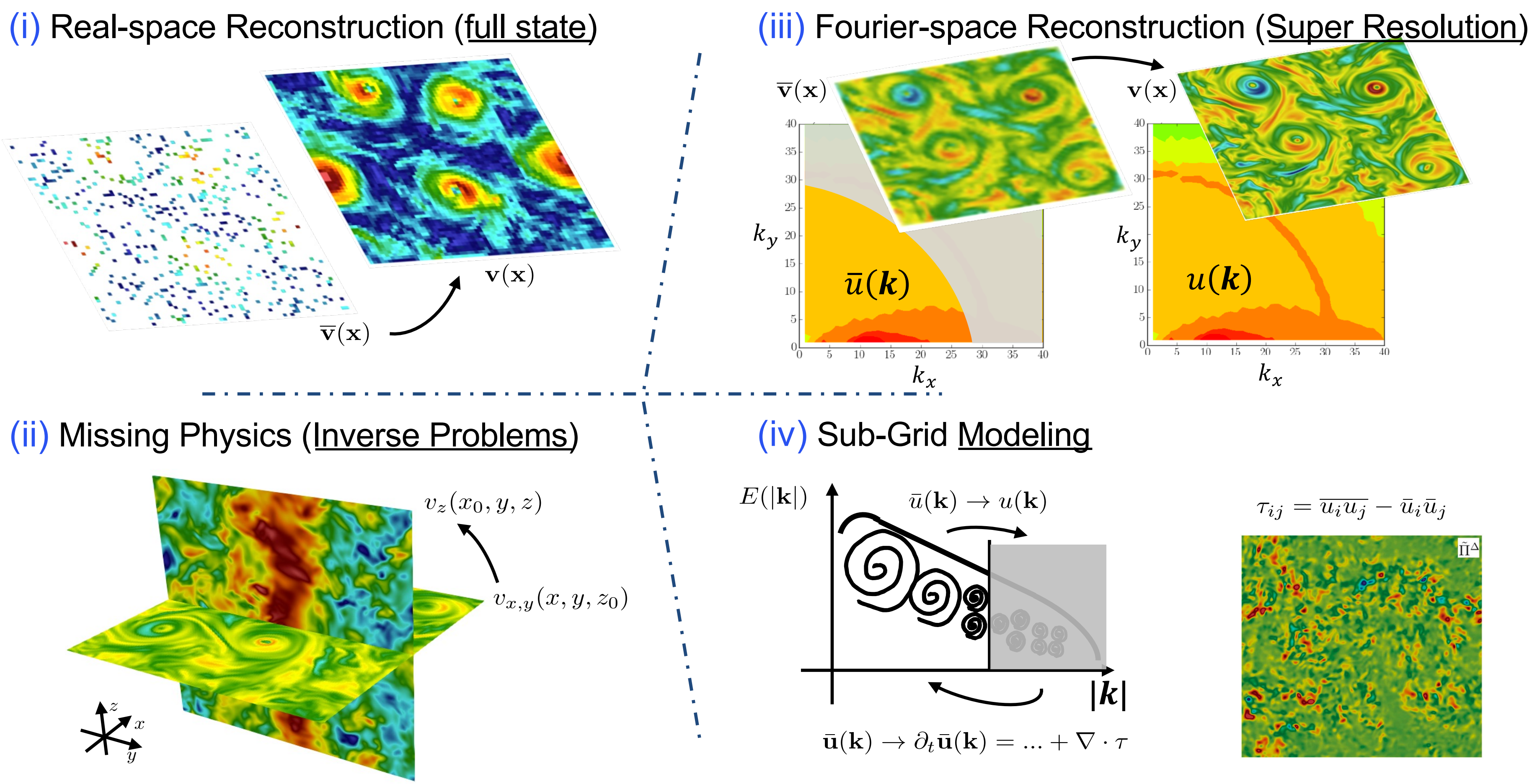
ML-DATA DRIVEN



Can we use Data-Driven Methods to reconstruct turbulent flows?



Turbulent Data GENERATION, FLOW RECONSTRUCTION

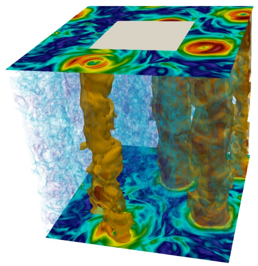


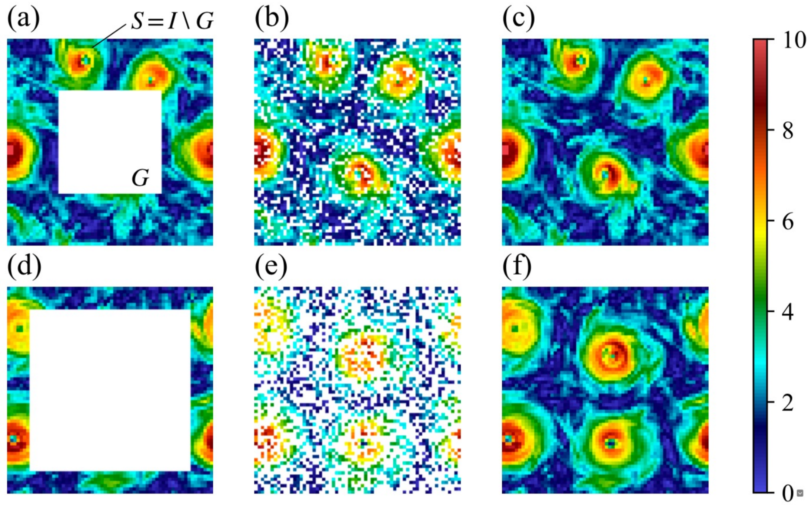
M. Buzzicotti. "Data reconstruction for complex flows using AI: recent progress, obstacles, and perspectives."
Europhysics Letters, EPL 142 23001 (2023).

Experimenting in Silico: Full State Reconstruction

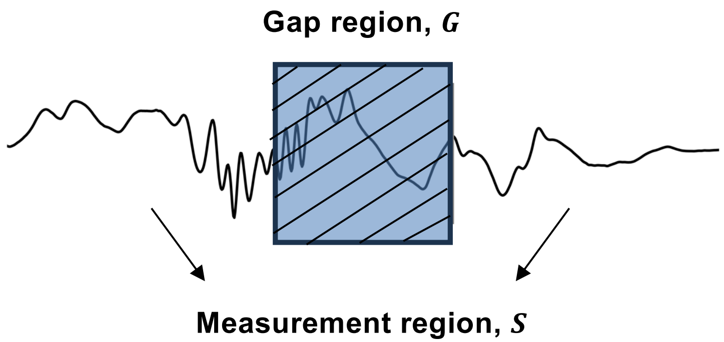
Problem Setup:

Eulerian





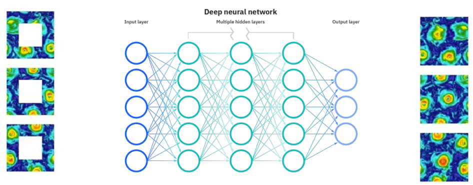
Lagrangian



DATA DRIVEN

$$u(\boldsymbol{x}) = \sum_{n=1}^{N} a_n \psi_n(\boldsymbol{x}) = \sum_{n=1}^{N'} a_n \psi_n(\boldsymbol{x}) + \sum_{n=N'+1}^{N} a_n \psi_n(\boldsymbol{x})$$

$$\int_S \left[u(\boldsymbol{x}) - \sum_{n=1}^{N'} a_n \psi_n(\boldsymbol{x})\right]^2 \mathrm{d}\boldsymbol{x}$$



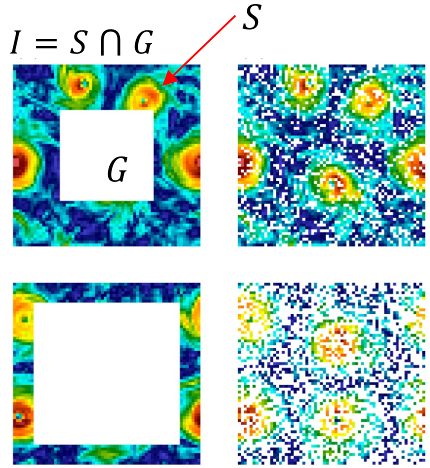
1. EQUATION FREE
PRINCIPAL ORTHOGONAL DECOMPOSITION
GAPPY-POD & EXTENDED POD
GAUSSIAN PROCESS REGRESSION

2. EQUATION FREE
GENERATIVE ADVERSARIAL NETWROKS
DIFFUSION MODELS

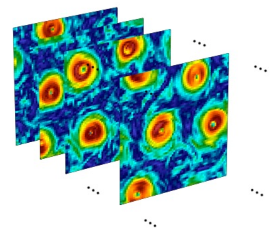
GAPPY-POD (PRINCIPAL ORTHOGONAL DECOMPOSITION)

DATA DRIVEN

$$u_G(\boldsymbol{x}) = \sum_{n=1}^{N'} a_n^* \psi_n(\boldsymbol{x}) + \sum_{n=N'+1}^{N} a_n \psi_n(\boldsymbol{x}) \qquad (\boldsymbol{x} \in G)$$



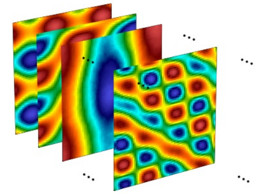
Training dataset $u^1(\boldsymbol{x}), u^2(\boldsymbol{x}), \cdots$



$$K(\boldsymbol{x},\boldsymbol{y}) = \langle u_I(\boldsymbol{x}) u_I(\boldsymbol{y}) \rangle \ (\boldsymbol{x},\boldsymbol{y} \in I)$$

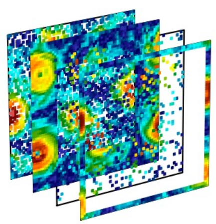
$$\int_I K(\boldsymbol{x},\boldsymbol{y}) \psi_n(\boldsymbol{y}) d\boldsymbol{y} = \lambda_n \psi_n(\boldsymbol{x})$$

POD modes $\psi_1(\boldsymbol{x}), \psi_2(\boldsymbol{x}), \cdots$



Reconstruction Process

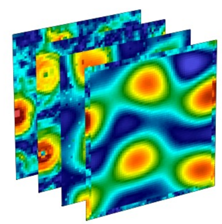
Gappy data $u^1(\tilde{\boldsymbol{x}}), u^2(\tilde{\boldsymbol{x}}), \cdots$



Compute expansion coeff. $a_1, a_2, \cdots$

$$\tilde{E} = \int_S d\boldsymbol{x} \left(u_S(\boldsymbol{x}) - \sum_{n=1}^{N'} a_n \psi_n(\boldsymbol{x}) \right)^2$$

$$a^* = \operatorname*{argmin}_{\boldsymbol{a}} \tilde{E}$$



Reconstructed field
optimized wrt the
mean square error
on the measurements

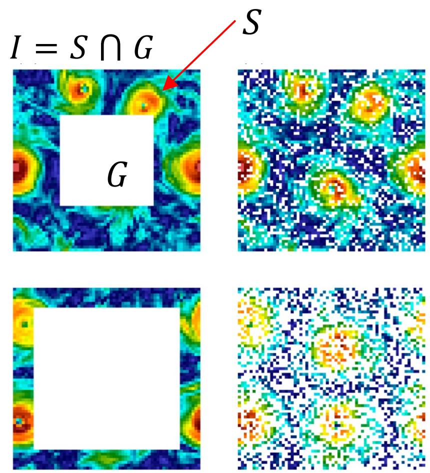
Li, T., Buzzicotti, M., Biferale, L., Bonaccorso, F., Chen, S., & Wan, M. (2023). *Journal of Fluid Mechanics*, *971*, A3.

EXTENDED POD (PRINCIPAL ORTHOGONAL DECOMPOSITION)

DATA DRIVEN

<ins>“Training” Phase</ins>

$$u_S(\boldsymbol{x}) = \sum_{n=1}^{N_S} b_n \phi_n(\boldsymbol{x}) \leftarrow$$ **Exact POD decomposition on *S***



$$K(\boldsymbol{x}, \boldsymbol{y}) = \langle u_S(\boldsymbol{x}) u_S(\boldsymbol{y}) \rangle \ (\boldsymbol{x}, \boldsymbol{y} \in S)$$

$$\int_S K(\boldsymbol{x}, \boldsymbol{y}) \phi_n(\boldsymbol{y}) d\boldsymbol{y} = \lambda_n \phi_n(\boldsymbol{x})$$

$$\phi_n(\boldsymbol{x}) = \frac{\langle b_n u_S(\boldsymbol{x}) \rangle}{\lambda_n}$$

average over the training dataset

Orthogonal POD modes on the support, *S*

EPOD modes computed replacing the velocity in the GAP region (over the training dataset)

$$\Rightarrow \quad \phi_n^e(\boldsymbol{x}) = \frac{\langle b_n u_G(\boldsymbol{x}) \rangle}{\lambda_n}$$

EPOD are no longer an orthogonal base

<ins>Reconstruction Process</ins>

$$b_n = \int_S u_S(\boldsymbol{x}) \phi_n(\boldsymbol{x}) d\boldsymbol{x} \quad \Rightarrow \quad u_G(\boldsymbol{x}) = \sum_{n=1}^{N_S} b_n \phi_n^e(\boldsymbol{x})$$

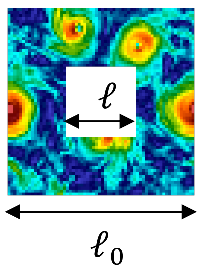
exact coeff. computed on the measure support

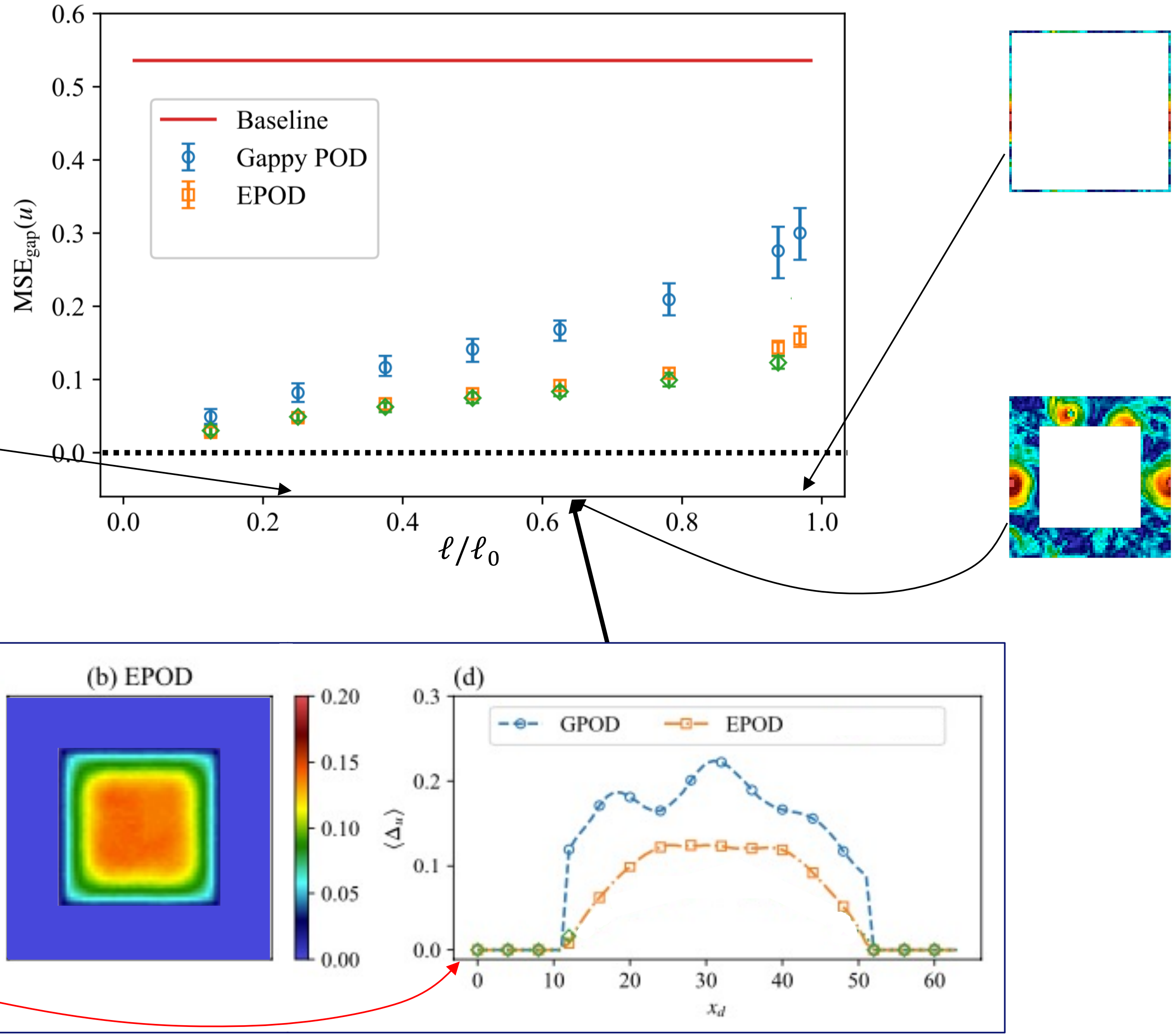
Approximated solution inside the GAP

Extended proper orthogonal decomposition: a tool to analyse correlated events in turbulent flows. Borée, J. Experiments in fluids, 35(2), 188-192.

Point-wise Mean Squared Error

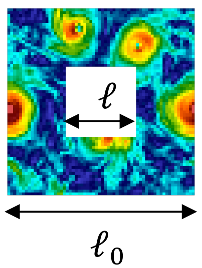
$$MSE_{gap} = \frac{1}{\mathrm{E}_k} \int_G \left(u_{true}(\boldsymbol{x}) - u_{pred}(\boldsymbol{x})\right)^2 d\boldsymbol{x}$$

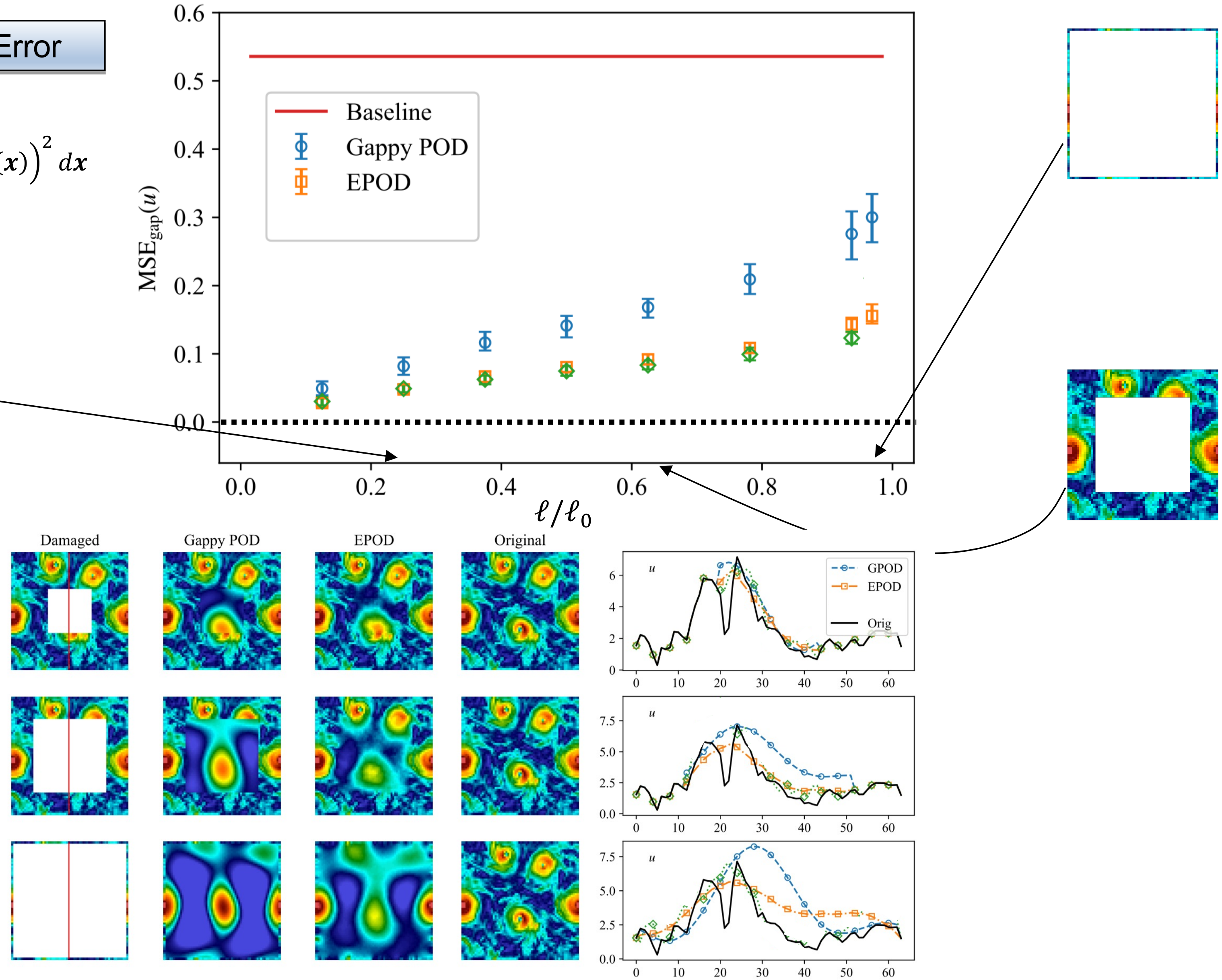




Point-wise Mean Squared Error

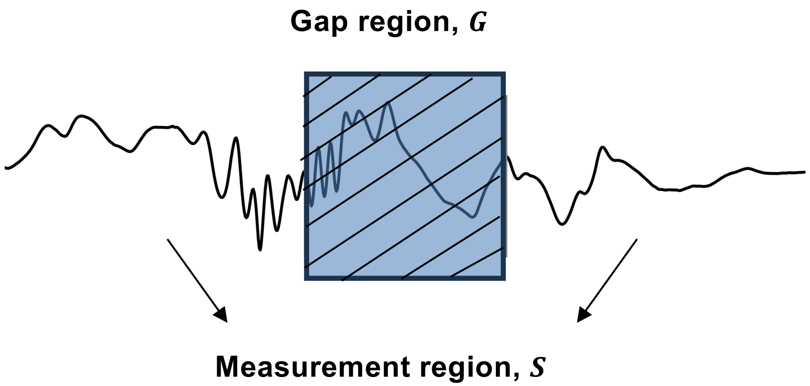
$$MSE_{gap} = \frac{1}{\mathrm{E}_k} \int_G \left(u_{true}(\boldsymbol{x}) - u_{pred}(\boldsymbol{x})\right)^2 d\boldsymbol{x}$$





Gaussian Process Regression (GPR)

DATA DRIVEN



Measuring the covariance on Gap-free datasets:

$$\mathcal{V}_S = \{V(t_{s_1}), V(t_{s_2}), \ldots, V(t_{s_{N(S)}}) \,|\, t_{s_i} \in S\}$$

$$\mathcal{V}_G = \{V(t_{g_1}), V(t_{g_2}), \ldots, V(t_{g_{N(G)}}) \,|\, t_{g_i} \in G\}$$

"Gaussianity" assumption

$$\begin{bmatrix} \mathcal{V}_S \\ \mathcal{V}_G \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} C_{SS} & C_{SG} \\ C_{GS} & C_{GG} \end{bmatrix}\right)$$

Covariances;
measure-measure
measure-gap
gap-gap

$$(C_{SS})_{ij} = \langle V(t_{s_i}) V(t_{s_j}) \rangle$$
$$(C_{SG})_{ij} = \langle V(t_{s_i}) V(t_{g_j}) \rangle$$
$$(C_{GG})_{ij} = \langle V(t_{g_i}) V(t_{g_j}) \rangle$$

(**Training Process**)

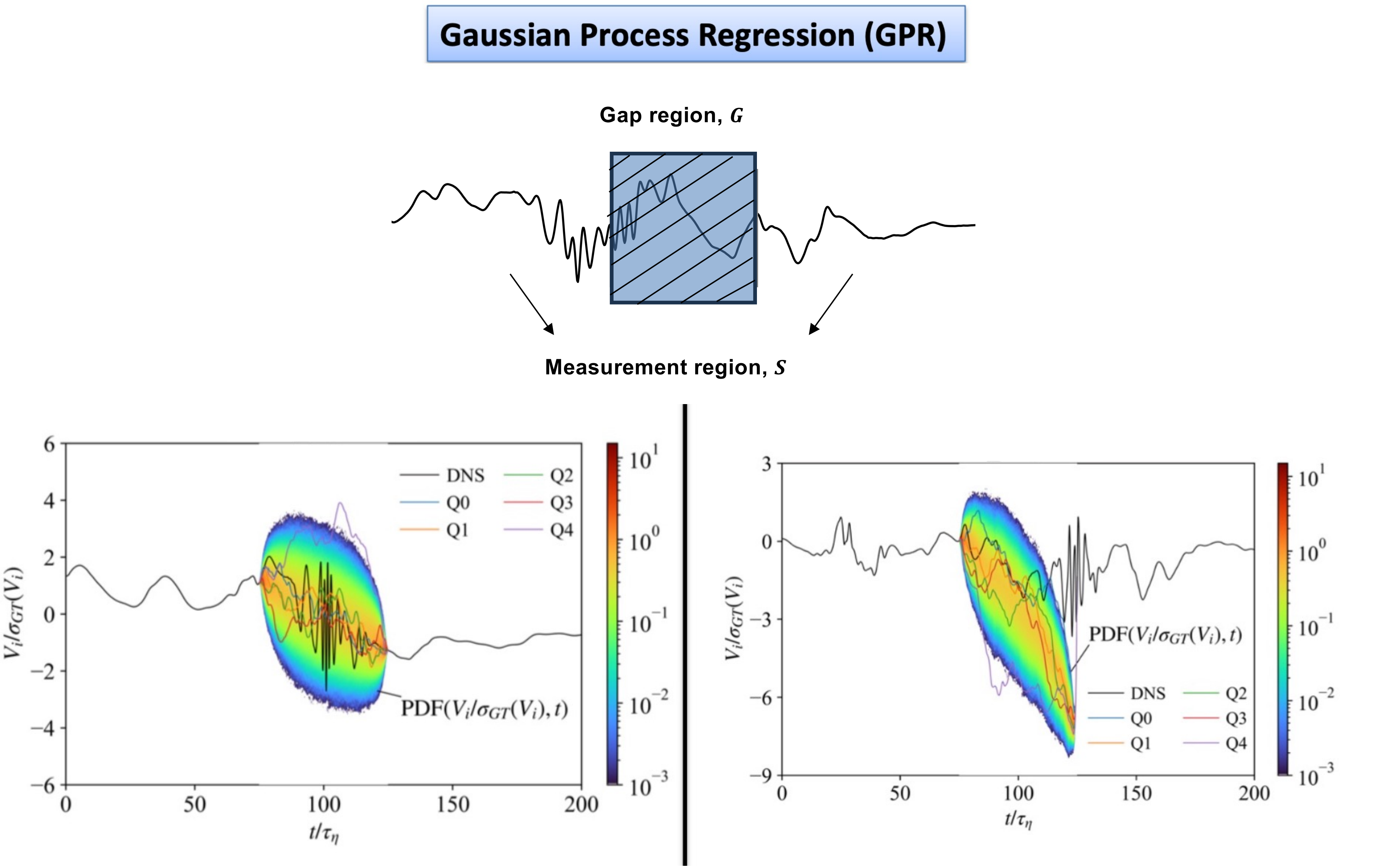
Reconstruction process conditioning on $\mathcal{V}_S$:

$$\mathcal{V}_G | \mathcal{V}_S \sim \mathcal{N}(\mu_G, \Sigma_{GG})$$

$$\mu_G = C_{GS} C_{SS}^{-1} \mathcal{V}_S$$

$$\Sigma_{GG} = C_{GG} - C_{GS} C_{SS}^{-1} C_{SG}$$

Sampling from a Gaussian whose mean and variance are given by the covariance matrices evaluated on GAP-free data.



Gaussian Process Regression (GPR)
Gap region, G
Measurement region, S
DNS
Q0
Q1
Q2
Q3
Q4
PDF($V_i/\sigma_{GT}(V_i), t$)
$V_i/\sigma_{GT}(V_i)$
t/τ_η

Can data-driven solutions be improved with machine learning methods?