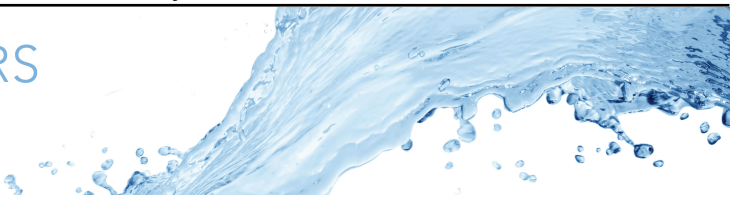
J. Fluid Mech. (2026), *vol.* 1043, A40, doi:10.1017/jfm.2026.12041



Lagrangian proper orthogonal decomposition

Stefano Brizzolara[1] **and Ron Shnapp**[2]

[1]Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ, USA

[2]Mechanical Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Corresponding author: Stefano Brizzolara, stefano.brizzolara@icloud.com

(Received 27 April 2026 UTC; revised 17 July 2026 UTC; accepted 9 August 2026 UTC)

We introduce a modal representation for Lagrangian trajectories in turbulence, termed Lagrangian proper orthogonal decomposition. An ensemble of particle trajectories is used to construct velocity time series, which are normalized independently for each trajectory to isolate fluctuations. Principal component analysis is then applied to the resulting dataset, with temporal instances defining the feature space. The method is tested on trajectories from both direct numerical simulations of homogeneous isotropic turbulence and a three-dimensional particle-tracking experiment, showing that the leading modes exhibit similar structures and energy distributions in both cases. Truncated reconstructions are obtained by combining modes and coefficients, rescaling the fluctuations and integrating in time. For trajectories of the order of the integral time scale, single-particle dispersion and curvature statistics are accurately reproduced using a limited number of modes ($\sim O(10)$), whereas capturing the tails of acceleration distributions and higher-order Lagrangian structure functions requires a larger set ($\approx$ 40–80). Longer trajectories require progressively more modes for accurate reconstruction. These results suggest a possible route to data-driven generation of synthetic particle trajectories via stochastic sampling of the modal Lagrangian dynamics.

Key words: turbulent flows

1. Introduction

Proper orthogonal decomposition (POD) is a widely used technique for extracting coherent structures from complex flow data. Originally introduced in fluid mechanics by Lumley (1967), it decomposes a flow field into a set of orthogonal modes that capture most of the kinetic energy. In this framework, a fluctuating velocity field is expressed as a sum of spatial, orthogonal modes, each modulated by time-dependent coefficients, whose order is

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

determined by their energy content (Weiss 2019). In its discrete form, POD is equivalent to principal component analysis (PCA), which consists in solving an eigenvalue problem for the data covariance matrix (Karhunen 1947).

POD has been primarily developed and applied within an Eulerian framework. In this setting, the method is applied to a set of flow realizations, which correspond to instantaneous snapshots of the velocity field measured at fixed spatial locations. Each realization is represented by the spatial distribution of the velocity field, including all components, arranged into a single state vector. The decomposition is obtained by solving the eigenvalue problem associated with the data covariance operator. The resulting eigenvectors define an orthogonal set of spatial modes, while the corresponding eigenvalues quantify their kinetic energy. Two equivalent formulations of the Eulerian POD are commonly used. In the first – the snapshot method (Sirovich 1987) – the covariance matrix is constructed from correlations between different realizations. In the second case, the covariance is constructed in physical space, thereby correlating spatial degrees of freedom across the dataset.

Despite POD's widespread use in Eulerian problems in fluid mechanics, applications in Lagrangian settings – where fluid properties are prescribed along fluid parcel positions (Monin & Yaglom 2013) – remain limited. Schiødt *et al.* (2022) developed a particle-POD method for inertial particles in decaying turbulence. Hassanian (2025) attempted to compare the distribution of POD mode energy from a Lagrangian experiment with Kolmogorov spectra, although how these are related is not yet clear. In the context of Lagrangian dynamics, Shinde & Gaitonde (2021) developed a POD routine called Lagrangian modal analysis that decomposes Lagrangian fields (the flow map) to identify coherent structures using space as a variable. The authors demonstrated that Lagrangian modal analysis can identify Lagrangian coherent structures in a manner similar to finite-time Lyapunov exponent analysis. In other contexts, PCA has been applied to particle trajectories, particularly in soft matter. For example, Chen *et al.* (2015) employed PCA at the image level to characterize particle motion, rather than as a modal decomposition of individual trajectories.

Since the introduction of Lagrangian experimental and numerical techniques roughly 30 years ago (Yeung & Pope 1989; Maas *et al.* 1993; Malik *et al.* 1993), the Lagrangian viewpoint has revealed numerous important phenomena related to dispersion (Biferale *et al.* 2005; Shnapp & Liberzon 2018; Tan & Ni 2022; Shnapp *et al.* 2023), energy transfer (Meneveau & Lund 1994; Pumir *et al.* 2001) and intermittency (Biferale *et al.* 2008; Benzi *et al.* 2010) in turbulence, as well as properties of inertial particles therein (Bec *et al.* 2010). Nevertheless, our understanding of how the nonlinear multiscale interactions inherent to turbulent flows lead to these phenomena remains in its infancy. This work is motivated by the notion that a Lagrangian modal decomposition could help in this respect.

In this paper, we introduce a method, Lagrangian POD (LPOD), to decompose Lagrangian trajectories from a turbulent flow into orthogonal modes. We demonstrate its applicability using trajectories from both a direct numerical simulation (DNS) of homogeneous and isotropic turbulence (HIT) and a three-dimensional particle tracking (3D-PTV) experiment of quasi-HIT in a von Kármán flow. With an appropriate data arrangement and normalization, relatively long trajectories can be reconstructed using a limited number of modes, which appear to have a similar functional form in both the DNS and the experimental data. We also show that the method faithfully captures the intermittency of Lagrangian acceleration, velocity differences and curvature. The method has the potential to advance modelling and data compression in the context of the Lagrangian description of turbulent flows.

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

2. Methods

2.1. *The LPOD routine*

We consider the Lagrangian trajectories of tracer particles of a homogeneous, isotropic and stationary turbulent flow, which are defined as the solutions of the following equation:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{v}(\boldsymbol{x}(t), t), \qquad \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{2.1}$$

where $\boldsymbol{x}_0$ and t_0 denote the initial position and initial time, respectively, and $\boldsymbol{v}$ is the Eulerian velocity field. The corresponding Lagrangian velocity is given by $\boldsymbol{u}(\boldsymbol{x}_0, t_0; \tau) = \boldsymbol{v}(\boldsymbol{x}(t_0 + \tau), t_0 + \tau)$, where $\tau = t - t_0$ is the time lag along the trajectory. The assumptions of statistical homogeneity and stationarity imply that the statistics of the Lagrangian velocity do not depend on $\boldsymbol{x}_0$ or t_0, so these dependencies are henceforth omitted. A discrete ensemble of N_p trajectories is considered, each represented by the velocity signal,

$$\frac{\mathrm{d}\boldsymbol{x}^{(p)}}{\mathrm{d}t} = \boldsymbol{u}^{(p)}(\tau), \qquad p = 1, \ldots, N_p, \qquad \tau \in [0, \tau_{max}] \tag{2.2}$$

where p labels the trajectories in the ensemble and τ_{max} is the duration of all trajectories. For each trajectory p and velocity component $i = 1, 2, 3$, we define the temporal mean and standard deviation:

$$\mu_i^{(p)} = \frac{1}{\tau_{max}} \int_0^{\tau_{max}} u_i^{(p)}(\tau)\, \mathrm{d}\tau = \frac{x_i^{(p)}(\tau_{max}) - x_i^{(p)}(0)}{\tau_{max}}, \tag{2.3}$$

$$\sigma_i^{(p)} = \left(\frac{1}{\tau_{max}} \int_0^{\tau_{max}} \left(u_i^{(p)}(\tau) - \mu_i^{(p)}\right)^2 \mathrm{d}\tau \right)^{1/2}. \tag{2.4}$$

The normalized velocity fluctuation is then defined as

$$\tilde{u}_i^{(p)}(\tau) = \left(u_i^{(p)}(\tau) - \mu_i^{(p)}\right)/\sigma_i^{(p)}. \tag{2.5}$$

Notably, as $\mu_i^{(p)}$ is defined per particle, the normalized velocity fluctuations do not impose any net displacement across the time τ_{max}.

The LPOD is defined as a Karhunen–Loève expansion of these normalized Lagrangian velocity fluctuations,

$$\tilde{u}_i^{(p)}(\tau) = \sum_{k=1}^{\infty} a_{i,k}^{(p)}\, \phi_k(\tau), \tag{2.6}$$

where $\phi_k(\tau)$ are temporal modes and $a_{i,k}^{(p)}$ are the corresponding modal coefficients. The modes satisfy the orthogonality condition, namely $\int_0^{\tau_{max}} \phi_k(\tau)\, \phi_\ell(\tau)\, \mathrm{d}\tau = \delta_{k\ell}$. A truncated reconstruction using the first K modes is obtained as

$$u_i^{(p,K)}(\tau) = \mu_i^{(p)} + \sigma_i^{(p)} \sum_{k=1}^{K} a_{i,k}^{(p)}\, \phi_k(\tau), \tag{2.7}$$

and the corresponding trajectory obtained by the truncated reconstruction is then computed by time integration:

$$\boldsymbol{x}^{(p,K)}(\tau) = \boldsymbol{x}^{(p)}(0) + \int_0^{\tau} \boldsymbol{u}^{(p,K)}(s)\, \mathrm{d}s. \tag{2.8}$$

In practice, the velocity is available at discrete times $\tau_n = n\Delta\tau$, with $n = 1, \ldots, N_t$, where $\Delta\tau$ is the time step with which the trajectory is sampled, or, equivalently,

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

the inverse of the sampling frequency. The expansion is therefore approximated by applying PCA to the matrix of sampled normalized signals $U_{(p,i),n} = \tilde{u}_i^{(p)}(\tau_n)$ whose rows correspond to stacked trajectory velocity components and whose columns correspond to time samples. The eigenvectors of the temporal covariance matrix provide the discrete approximation of the modes $\phi_k(\tau_n)$, while the projections of the signals onto these modes give the coefficients $a_{i,k}^{(p)}$. A GitHub repository with Python implementation of the LPOD algorithm is available at Shnapp & Brizzolara (2026).

2.2. *Datasets description*

We employ data from both a DNS of HIT and a 3D-PTV experiment. The DNS trajectories are integrated in a 1024^3 HIT flow, $Re_\lambda = 433$, from the Johns Hopkins Turbulence Database – a widely used open database for exploring turbulence dynamics (Li *et al.* 2008). Trajectories are 688 frames long and are integrated with a time step of $\Delta\tau = 0.045\tau_\eta$, where τ_η is the Kolmogorov time scale. The total length of the trajectories corresponds to $\tau_{max} = 30.77\tau_\eta = 0.68T_I$, where T_I is the integral time scale of the flow. The numerical dataset contains 5120 trajectories. Trajectory data were downloaded directly from the online database via a MATLAB interface. Detailed information about this open database is available from Li *et al.* (2008).

The experimental trajectories are obtained from a low-seeding 3D-PTV experiment conducted in a von Kármán turbulence tank, which produces approximately zero-mean HIT at $Re_\lambda \approx 150$. The set-up is made of a circular 0.2 m diameter Poly(methyl methacrylate) (PMMA) cylinder enclosed inside a square tank, both filled with filtered tap water. The flow is driven by two smooth disks rotating in opposite directions at 3 Hz. Trajectories of 50 µm polyamide tracers (PSP50, Dantec Dynamics) have been identified using the MyPTV open-source software (Shnapp 2022). The trajectories obtained through 3D-PTV vary in length, with shorter trajectories more likely and longer ones rarer. Unless otherwise specified, we considered trajectory chunks of 200 frames, which correspond to an ensemble of 2607 trajectories. The trajectories are sampled with a time step of $\Delta\tau = 0.09\tau_\eta$. The total length of the trajectories corresponds to $\tau_{max} = 18.95\ \tau_\eta = 1.34\ T_I$. Further details about the experimental dataset and the evaluation of the integral and Kolmogorov scales are available in Appendix A.

2.3. *Statistical observables*

We introduce a set of statistical observables to assess the performance of truncated LPOD trajectory reconstructions, including modal properties, single-particle Lagrangian statistics, and geometric characteristics of the trajectories.

For the modal representation, we compute the projection coefficients associated with each mode, along with the corresponding variance explained, which quantifies the energy content of each principal component. To characterize Lagrangian transport, we first consider two dispersion-related observables: the mean-squared particle displacement and the deviation from ballistic motion, respectively:

$$\langle \delta \boldsymbol{x}^2 \rangle = \langle [\boldsymbol{x}(\tau) - \boldsymbol{x}_0]^2 \rangle, \tag{2.9a}$$

$$\langle \delta_b \boldsymbol{x}^2 \rangle = \langle [\boldsymbol{x}(\tau) - (\boldsymbol{x}_0 + \boldsymbol{u}(0)\,\tau)]^2 \rangle. \tag{2.9b}$$

The mean-squared displacement exhibits ballistic scaling for $\tau \ll T_I$, namely $\langle \delta \boldsymbol{x}^2 \rangle \sim u_{rms}^2 \tau^2$ (Taylor 1922). The deviation from ballistic motion isolates the contribution of velocity fluctuations (Shnapp 2019) and is expected to scale as $\langle \delta_b \boldsymbol{x}^2 \rangle \sim \epsilon \tau^3$ in the inertial

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

range (Obukhov 1959), similarly to the scaling of the mean square separation between particle pairs (Richardson 1926). More details about the derivation of this scaling can be found in Appendix B material and in Monin & Yaglom (2013), equation (24.13).

We further consider the Lagrangian velocity increments, namely

$$\delta \boldsymbol{u}(\tau) = \boldsymbol{u}(\tau) - \boldsymbol{u}(0) \tag{2.10}$$

and their longitudinal component, defined as

$$\delta u_{\parallel} = \frac{\delta \boldsymbol{u}(\tau) \cdot (\boldsymbol{x}(\tau) - \boldsymbol{x}_0)}{|\boldsymbol{x}(\tau) - \boldsymbol{x}_0|}. \tag{2.11}$$

The pth-order absolute value and longitudinal structure functions are defined as

$$S_p = \langle |\delta \boldsymbol{u}(\tau)|^p \rangle, \tag{2.12}$$

$$S_p^{\parallel}(\tau) = \langle \delta u_{\parallel}(\tau)^p \rangle. \tag{2.13}$$

In the inertial range, the second-order structure functions ($p = 2$) are expected to scale linearly with time, namely $S_2^{\parallel} \sim \epsilon \tau$ (Monin & Yaglom 2013).

To challenge the capability of LPOD in reconstructing higher-order statistics associated with intermittency, we consider the scaling of the fourth- and sixth-order structure functions versus the second-order one, as well as the logarithmic derivative of S_4 versus S_2, namely

$$\xi_4(\tau) = \frac{\mathrm{d}\ \log(S_4)}{\mathrm{d}\ \log(S_2)}. \tag{2.14}$$

This latter quantity is a proxy for intermittency and has been found to exhibit a universal functional form across several datasets and different Reynolds numbers (Arnéodo *et al.* 2008; Toschi & Bodenschatz 2009).

An additional strongly intermittent observable is the absolute value of the Lagrangian acceleration, namely the material derivative of the Lagrangian velocity:

$$a(\tau) = |\boldsymbol{a}(\tau)|, \quad \boldsymbol{a}(\tau) = \frac{\mathrm{d}\boldsymbol{u}(\tau)}{\mathrm{d}\tau}. \tag{2.15}$$

The probability distribution of $|\boldsymbol{a}|$ is heavy-tailed and, at sufficiently high Reynolds number, it is believed to possess a universal shape when normalized by the Kolmogorov scales (Toschi & Bodenschatz 2009).

Finally, we consider the curvature as a proxy of for the trajectories' geometry:

$$\kappa = \frac{|\boldsymbol{u} \times \boldsymbol{a}|}{|\boldsymbol{u}|^3}. \tag{2.16}$$

At sufficiently high Reynolds number, the curvature distribution exhibits a power-law behaviour that scales linearly for small curvatures and as $\kappa^{-2/5}$ for large curvatures (Braun *et al.* 2006; Xu *et al.* 2007).

3. Results

3.1. *LPOD modes and their energy*

We start by presenting the results of LPOD applied to both the numerical and experimental datasets described in § 2. Figure 1(*a*) shows the first 10 modes of the expansion, where the time is normalized with the trajectory duration and the intensity with the first value (the value at $\tau = 0$). In both the experiment and the simulation, progressively higher-order

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

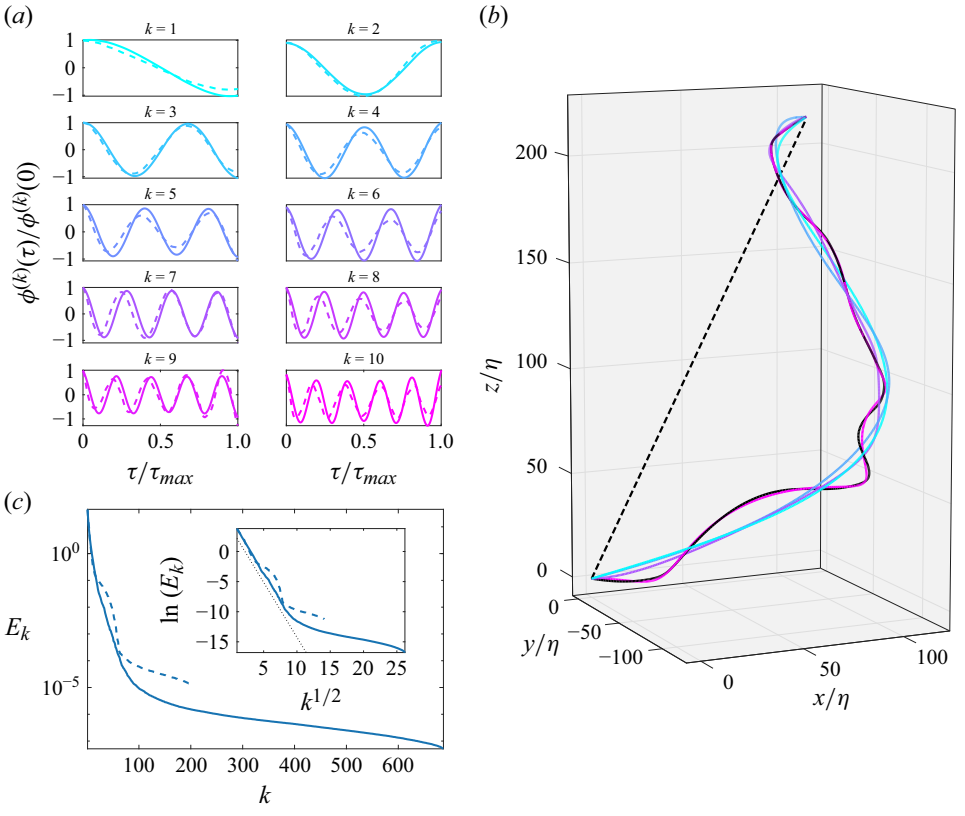


Figure 1. (*a*) First 10 LPOD modes: solid lines denote DNS data and dashed lines 3D-PTV data. (*b*) Example DNS trajectory reconstruction using an increasing number of modes (2, 4, 8 and 16); the dashed line shows the ballistic reconstruction from the mean velocity, and the black line the original trajectory. (*c*) Energy distribution (explained variance) normalized such that $\sum_{k=1}^{N_t} E_k = 100$; solid and dashed lines correspond to DNS and 3D-PTV data, respectively. The inset shows the same data in square-log coordinates to highlight the subexponential decay.

modes capture higher-frequency velocity fluctuations. By comparing the leading modes obtained from the experiments (dashed lines) and the simulation (solid line), we observe that the mode shapes are similar in both cases. Figure 1(*b*) shows an example of truncated reconstruction for a topologically complex trajectory from the DNS data. Although truncated at the 16th mode, the reconstruction of this example trajectory qualitatively matches the ground truth. The similarity between the modes from the two datasets is an important feature of LPOD, as it has significant potential for data reconstruction, compression and modelling. This result is, in fact, expected, as a well-known feature of the orthogonal decomposition is that statistically stationary signals decompose into the harmonic basis of a discrete Fourier transform (George 1988). Since the flow considered here is stationary, homogeneous and isotropic (nearly so in the experiment), the trajectories are expected to follow stationary statistics, which is encapsulated in the modes we obtained. Deviations from ideal HIT behaviour are evident in the differences between the two datasets, which become progressively more pronounced as k increases. These discrepancies arise because the experimental flow is sampled with finite-size particles and the DNS dataset is not perfectly stationary, situations that cannot be represented by purely periodic modes (Hodžić *et al.* 2024). Notably, orthogonal decomposition strategies are also

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

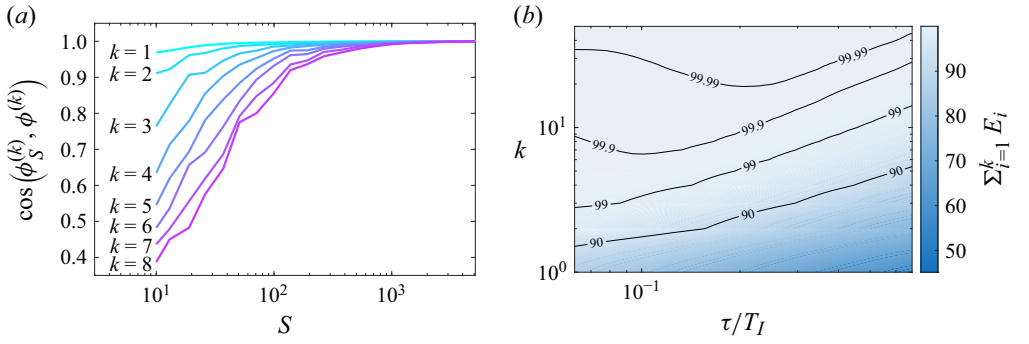


Figure 2. (*a*) Convergence profile of the first eight modes, namely the cosine between the kth mode computed using the whole dataset ($\phi^{(k)}$) and the ones computed with a subsample of S trajectories ($\phi_S^{(k)}$); each profile is computed by averaging the results of 100 bootstrap resamples of the dataset. (*b*) Heatmap of the total energy captured by the first k modes for different trajectory durations τ; the solid lines are isolines corresponding to 90 %, 99 %, 99.9 % and 99.99 % of the total energy. Both graphs are obtained using the DNS dataset.

very relevant for inhomogeneous or non-stationary flows or those that are homogeneous in some ways but not others, as discussed by George (1988).

Figure 1(*c*) shows the percentage of energy content associated with each mode (its eigenvalue) for the experiments and DNS data. As with PCA analyses, most of the energy in LPOD is concentrated in the first few modes, with the first 10 modes accounting for 99 % of the kinetic energy. As highlighted by the inset of figure 1(*c*), the mode energy decreases subexponentially, approximately as $e^{-(k/C)^{1/2}}$, where C is a constant, up to $k \approx 100$. For k larger than this, the deviation between the experimental and numerical cases becomes more significant. Several factors may lead to discrepancies between datasets for high-order modes. First, the two datasets are characterized by heavily different Reynolds numbers ($Re_\lambda = 433$ for the simulations and $Re_\lambda \approx 150$ for the experiments). Second, the temporal sampling rates differ substantially: the simulation data are sampled at approximately 10 times the frequency of the experimental data when expressed in Kolmogorov time units. Third, the total trajectory length differs between the two datasets in both Kolmogorov time units and integral time units (see § 2.2). As a result, the experimental trajectories contain fewer samples (200) than the simulated ones (688), leading to a correspondingly smaller number of modes. Since the explained energy is expressed as a percentage of the total variance, this difference necessarily affects the distribution of energy among the highest-order modes.

As suggested by the energy distribution differences found between the two datasets, the performance of LPOD may depend on the trajectory duration. On the one hand, longer trajectories are expected to require a larger number of modes to be accurately represented; on the other hand, estimating higher-order modes is expected to require increasingly large ensembles of trajectories in the dataset. To quantify the effect of the dataset size on the representation of the modes, we report in figure 2(*a*) the convergence profiles of the first eight modes, quantified by the cosine similarity between the kth eigenfunction computed using the full dataset and that obtained from a random subset of trajectories. As the sample size increases, this similarity approaches unity, indicating convergence of the estimated mode. As expected, higher-order modes (large k) converge more slowly and therefore require more trajectories to be reliably identified. To quantify the effect of trajectory duration, we compute the cumulative explained energy captured by the first k modes as a function of trajectory length. Figure 2(*b*) shows isolines of cumulated explained energy, demonstrating that progressively longer trajectories require an increasing number

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

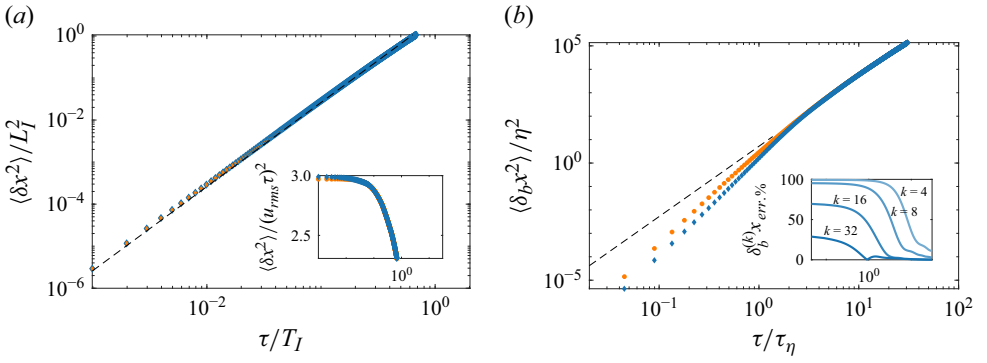


Figure 3. (*a*) Mean-square displacement obtained by the DNS dataset; the inset shows the data normalized by the ballistic scaling. (*b*) Mean square of the deviation-from-ballistic motion; the inset shows the relative reconstruction error $\delta_b^{(k)} x_{err.\%} = 100 \times |\langle \delta_b^{(k)} x^2 \rangle - \langle \delta_b x^2 \rangle| / \langle \delta_b x^2 \rangle$ computed with progressively more modes. In both panels, orange and blue markers denote the original data and the 16-mode truncation, respectively.

of modes to capture the same fraction of the total energy. For instance, five modes capture approximately 99 % of the energy for trajectories of duration 0.2, T_I, whereas trajectories of duration 0.6, T_I require approximately 10 modes to achieve the same level of cumulated energy.

3.2. *Reconstructing Lagrangian statistics*

We present the Lagrangian statistics defined in § 2.3 for both the original and reconstructed trajectories. Figure 3 shows Lagrangian single-particle dispersion statistics. Considering the mean-squared displacement, figure 3(*a*) shows that the truncated reconstruction done with 16 modes is indistinguishable from the original signal. This is expected, since the mean velocity for each trajectory is removed prior to the decomposition and reintroduced during reconstruction. As a result, the ballistic contribution to the displacement is preserved exactly, so that, in the ballistic regime, the mean-squared displacement is recovered, even in the absence of fluctuating modes. To better assess the limits of the truncated representation on the single-particle dispersion process, we consider the deviation from ballistic motion $\langle \delta_b x^2 \rangle$, shown in figure 3(*b*). This quantity is more sensitive to small-scale velocity fluctuations and is therefore more difficult to reconstruct, particularly at short time lags. Nevertheless, a relatively small number of modes (16) is sufficient to recover the correct behaviour for lags exceeding a few Kolmogorov time scales, when the data are expected to exhibit inertial-range scaling. To assess the impact of the number of modes on the reconstruction of $\langle \delta_b x^2 \rangle$, we evaluate the relative reconstruction error, defined as the ratio between the difference of the reconstructed and original signals and the original signal itself. This is shown in the inset of figure 3(*b*) as a function of both the time lag and the number of retained modes, demonstrating that adding modes reduces the relative error at progressively shorter time lags (smaller scales).

The second-order longitudinal structure function is accurately reconstructed for lags larger than a few Kolmogorov time scales using 16 modes only (figure 4*a*). The inset shows the relative reconstruction error, demonstrating that reconstruction improves at progressively shorter lags as more modes are added. Figure 4(*b*) shows the probability density functions (PDFs) of the longitudinal normalized velocity increments for increasing time lags. As the lag increases, the distributions approach Gaussian behaviour. Reconstructions based on a limited number of modes (16) underestimate the

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

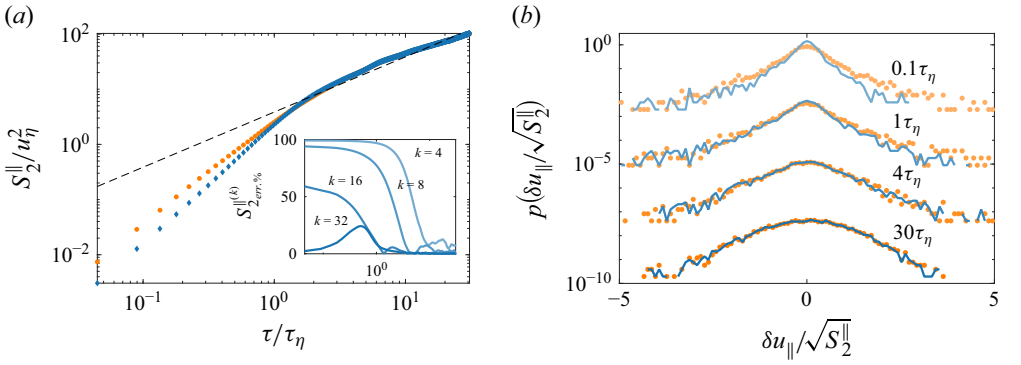


Figure 4. (*a*) Second-order longitudinal structure function computed from the DNS dataset (orange circles denote the original data, whereas blue diamonds are obtained from the first 16 modes); the inset shows the relative reconstruction error $S_{2\,err.\,\%}^{\parallel(k)} = 100 \times |S_2^{\parallel(k)} - S_2^{\parallel}|/S_2^{\parallel}$ computed with progressively more modes. (*b*) The PDF of the velocity increments from the DNS dataset. The time lag varies from 0.1 to 30 τ_η. The reconstructed data (blue solid lines) are obtained using the first 16 modes; the orange circles are the original data. The data are normalized with the structure function computed from the original data for both cases.

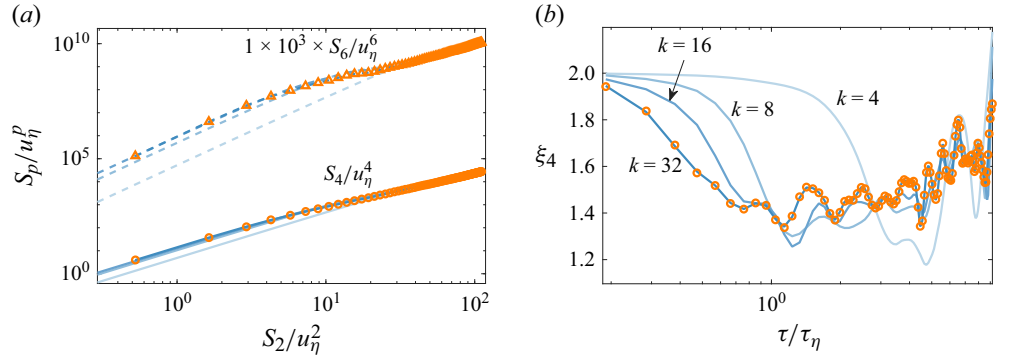


Figure 5. (*a*) Fourth- (S_4, orange circles) and sixth-order (S_6, orange triangles) structure functions plotted against the second-order one (S_2); continuous and dashed lines are the reconstructed signals of S_4 and S_6, respectively; the colour intensifies with the number of modes, from 4 (lighter blue) to 32 (darker blue); S_6 is multiplied by a factor 1000 for clarity. (*b*) Logarithmic derivative of S_4 versus S_2, as defined in (2.14), versus time lag. The orange circles denote the original signal, while the blue lines correspond to reconstructions using 4, 8, 16 and 32 modes (colours from light to dark blue). The curves are obtained from experimental data, using trajectories of duration $\sim 10\,\tau_\eta$ (100 frames, half the duration of the long trajectories used elsewhere) to improve statistics.

tails at short time lags when the distribution is strongly non-Gaussian, whereas they are accurate for longer time lags as Gaussianity is approached.

To test the LPOD capabilities with regard to intermittency, we compute the fourth- and sixth-order structure functions. To this purpose, we employed the experimental dataset, extending it by using trajectories as short as $10\,\tau_\eta$ (100 frames). In this form, the experimental dataset contains approximately 32 500 trajectories (see further details in Appendix A), enough to compute the scalings of higher-order structure functions. Figure 5(*a*) shows that the scaling of S_4 and S_6 versus S_2 can be correctly reconstructed by using 32 modes. Similarly, figure 5(*b*) demonstrates that 32 modes are sufficient to reconstruct the signal of the logarithmic derivative of S_4 with respect to S_2 (equation (2.14)), although the original signal is noisy due to the still moderate number of trajectories

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

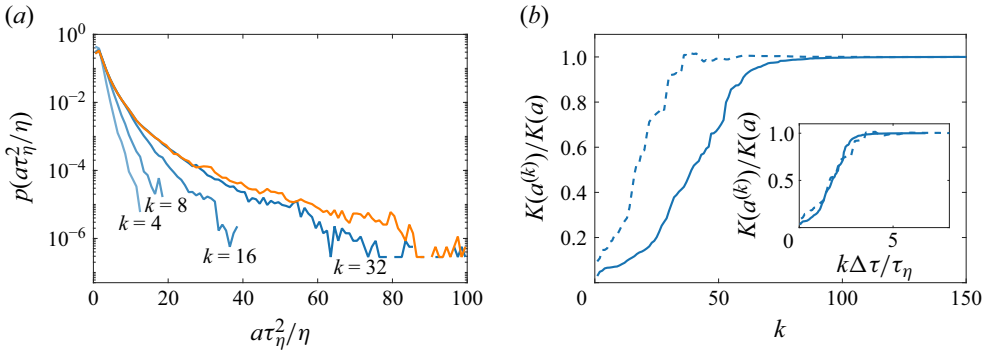


Figure 6. (*a*) The PDF of the trajectory acceleration computed from the DNS data. The orange line denotes the original signal, while the blue lines correspond to reconstructions using 4, 8, 16 and 32 modes. (*b*) Kurtosis $\mathcal{K}$ of the reconstructed acceleration normalized by the original signal's Kurtosis; dashed and solid lines indicate experimental and DNS data, respectively; the inset shows the same curves with the mode number multiplied by $\Delta\tau/\tau_\eta$.

in the dataset. As demonstrated in § 3.1, the shorter trajectories employed in this analysis are easier to reconstruct, while longer trajectories are expected to require a larger number of modes. Furthermore, the present trajectory duration, approximately $10\tau_\eta$, is too short to robustly identify an inertial-range scaling regime, since contamination from dissipative effects – in fact, linked to intermittency – remain significant at these time lags (Arnéodo *et al.* 2008).

Figure 6(*a*) shows the probability distribution of the trajectory acceleration for the DNS data. The acceleration distribution, as the longitudinal structure function at small time lags, requires high-order modes to be captured accurately. Indeed, the variance of the acceleration is proportional to the longitudinal structure function for vanishing time lags (see (24.12) in Monin & Yaglom 2013), which has been shown to require a relatively high number of modes to be accurately reconstructed (see the inset of figure 4*a*). However, even by using 32 modes, the tail of the acceleration distribution is hard to correctly capture: this discrepancy is due to the inherent difficulties of compressing the strongly intermittent signal of the Lagrangian acceleration for these relatively long numerical trajectories. To better highlight this aspect, figure 6(*b*) shows the ratio of the Kurtosis of the reconstructed acceleration to that of the original data: up to 80 modes are required to fully recover the extreme acceleration events in the DNS data (continuous line), whereas a smaller number (≈ 40) suffices for the experimental data (dashed line). As such, approximately 0.1 % and 1.0 % of the trajectories' turbulent kinetic energy is needed to resolve acceleration intermittency in the numerical and experimental data, respectively. Here, the difference between the experiments and the DNS data is likely due to difference in Reynolds number ($Re_\lambda \approx 150$ in the experiments versus 430 in the DNS), trajectory length ($\tau_{max} = 19.76\tau_\eta$ in the experiments versus $30.85\tau_\eta$ in the DNS), and sampling frequency ($10/\tau_\eta$ in the experiments versus $25/\tau_\eta$ in the DNS). We observe that, by normalizing the mode number k by the normalized sampling frequency $\tau_\eta/\Delta\tau$, the DNS and experimental Kurtosis collapse on the same master curve (see the inset of figure 6*b*).

Figure 7 shows the probability distribution of the trajectory curvature for the DNS data. From this perspective, the reconstructed and original signals appear indistinguishable, indicating that the first 16 modes accurately capture the trajectories' geometry. This result is surprising, as the curvature contains both the velocity and acceleration in its definition, the latter being difficult to reconstruct with few modes, especially concerning the tails

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

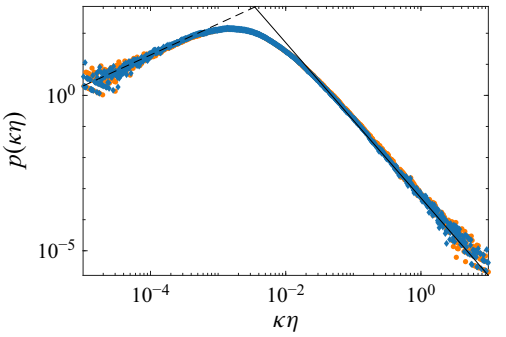


Figure 7. The PDF of the trajectory curvature (DNS dataset). Orange and blue markers are the original and reconstructed (16 modes) data, respectively, while the dashed and solid lines indicate the dimensional scalings κ^1 and $\kappa^{-5/2}$ for small and large curvatures, respectively. Data are logarithmically binned.

of the distribution. A possible explanation for this observation is that curvature is more correlated with low velocity than with high acceleration (Xu *et al.* 2007), making it more sensitive to large scale flow features than to intermittent acceleration events; then, as our method normalizes the velocity fluctuations it make it more robust for faithfully reconstructing this geometrical aspect.

4. Discussion and outlook

In this paper, we introduced a new method for analysing Lagrangian trajectories using PCA, which can be viewed as a Lagrangian analogue of standard POD. To the best of our knowledge, the only attempt to perform POD on Lagrangian trajectories was by Schiødt *et al.* (2022), but with a different approach. In their work, the authors normalized the data with ensemble averages, which are not suited for isolating the fluctuating velocities of each trajectory and do not preserve the total displacement over the full trajectory length. In contrast, our normalization ensures that the truncated reconstructions always begin and end at the same locations as the original trajectories. This is well illustrated by figure 1(*b*), which shows that the total drift along each trajectory is preserved by construction, regardless of the number of modes used. Encoding this normalization in the data also renders the single particle dispersion accurately reconstructible with an extremely low number of modes, as shown in figure 3(*a*), because, for lags shorter than the integral time scale, the process is ballistic (Taylor 1922). However, while this approach preserves the initial and final positions by construction, it does so at the cost of storing the temporal mean and standard deviation for each trajectory separately.

By analysing the statistics of truncated reconstructions, we show that LPOD's performance depends on the time lag considered, on the trajectories' length, and on the Reynolds number. For a fixed Reynolds number and trajectory duration, quantities defined at intermediate and long lags are accurately reconstructed with a small number of modes, whereas observables based on short-time lags, such as acceleration and short-time velocity increments, require a significantly larger number of modes to converge. This reflects the strong intermittency (see, e.g. La Porta *et al.* (2001) and Mordant *et al.* (2001)) associated with these quantities, which requires higher frequencies and thus higher-order modes to be accurately captured. In general, higher-order statistical observables, in particular the fourth- and sixth-order structure functions, require more modes to be accurately reconstructed, allegedly a consequence of the intrinsic L_2 variance optimality of LPOD.

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

By plotting the normalized acceleration Kurtosis as a function of the number of retained modes, we observe that scaling the mode number with the sampling frequency normalized by the Kolmogorov time collapses the data on the same master curve. This suggests that, if the data were sampled at the same normalized frequency, the number of modes needed to reconstruct the acceleration signals would have been Reynolds number and trajectory-length-independent. In other words, the degree of complexity needed for reconstructing acceleration statistics seems to be proportional to the length of the trajectories in terms of Kolmogorov times, regardless of whether this is set by the Reynolds number of the flow or by their length.

Finally, LPOD may serve as a starting point for modelling, particularly for generating synthetic trajectories. This task has been traditionally addressed with stochastic models (see, e.g. Thomson (1987), Pope & Chen (1990), Sawford (1991), Wilson *et al.* (2009) and Shnapp *et al.* (2020)) and recently using deep-learning techniques, e.g. diffusion networks (see, e.g. Li *et al.* 2024*a*,*b*, 2025) or data-informed stochastic models based on neural networks (de Wit *et al.* 2026). Applying LPOD to two different dataset of isotropic turbulence generated with different techniques and forcing conditions, we showed that the shape of the modes appears to only weakly depend on the turbulence under consideration, and may thus be universal – albeit only two cases (one DNS and one experiment) are reported here; this suggests that learning the joint probability distribution of the modes coefficient and stochastic sampling from it may be sufficient to construct synthetic Lagrangian turbulence. However, further investigation and testing of non-homogeneous, non-stationary and non-isotropic turbulent flows are needed to assess the mode shapes in different contexts.

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

Appendix A. Characterization of the experimental dataset

In the experiment, the flow is sustained in a von Kármán turbulence generator using filtered tap water inside a cylindrical volume, 18 cm in height and 18 cm in diameter. The flow is generated by two counter-rotating smooth discs, driven via two step motors to rotate at 150 revolutions per minute. Four synchronized cameras are used to image tracer particles in the flow (50 μm, PSP50 from Dantec Dynamic), viewing approximately 1500 particles per frame, located within a rectangular volume measuring $100 \times 100 \times 50$ mm at the centre of the cylinder – see figure 8(*c*) for a snapshot of the three-dimensional trajectories. The velocity in the region of measurement is, to a good approximation, zero-mean with Gaussian fluctuations whose root mean square is approximately $u_{rms} \approx 2.87 \times 10^{-2}$ m s^{-1} with 10 % difference between the vertical and horizontal directions. In agreement with previous studies (see, e.g. Chevillard *et al.* (2003) and Shnapp (2019)), the PDF of Lagrangian velocity differences relaxes from a heavy-tailed distribution to a Gaussian distribution across the inertial range (figure 8*b*), demarcating small-scale intermittency. The energy dissipation rate can be computed following different approaches. Considering the second-order Lagrangian structure function with a standard value of the constant $C_0 = 3$, we obtain $\epsilon \approx 8.2 \times 10^{-4}$ m^2 s^{-3}, see figure 8(*a*). Considering the equilibrium formulation (Frisch & Donnelly 1996), we obtain $\epsilon = C_\epsilon \, {u_{rms}}^3 / L_I = 6.4 \times 10^{-4}$ m^2 s^{-3} using $C_\epsilon = 0.5$, where the integral length scale obtained by fitting the Eulerian correlation function is $L_I = 18.6$ mm (see figure 9*a*). Considering the third-order Eulerian longitudinal structure function and the 4/5 law (Frisch & Donnelly 1996), we obtain $\epsilon = 4.7 \times 10^{-4}$ m^2 s^{-3} (fit shown in figure 9*b*).

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

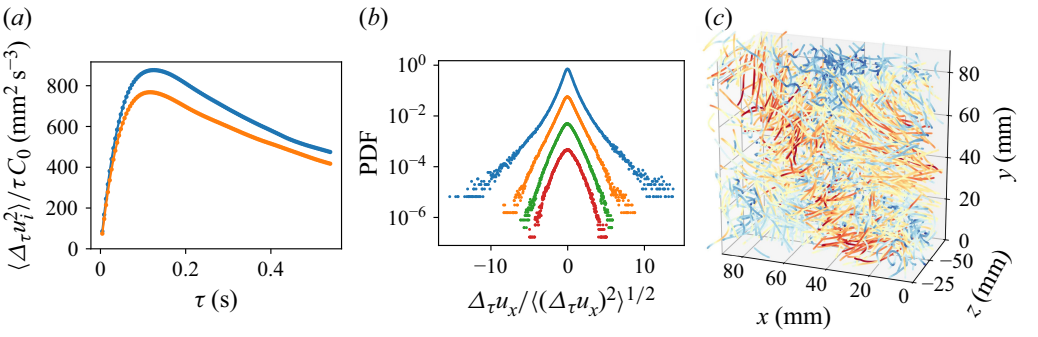


Figure 8. (*a*) Lagrangian second-order structure functions for the x (top) and y (bottom) velocity components, normalized with Kolmogorov scaling with $C_0 = 3$; the peaks allow to estimate the dissipation rate as $\epsilon \approx 820\,\mathrm{mm^2\,s^{-3}}$. (*b*) The PDFs of Lagrangian velocity increments, normalized by their standard deviation and vertically translated for $\tau/\tau_\eta \approx 0.6$, 3, 6 and 12, demonstrating the change of the PDF with time lag, commonly referred to as small-scale intermittency. (*c*) A Rendering of trajectories from the experiment, where the shade indicates the velocity magnitude.

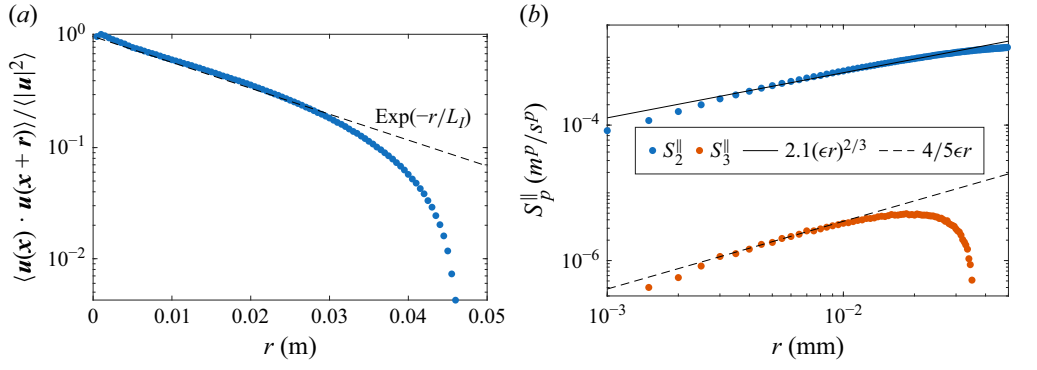


Figure 9. (*a*) Eulerian velocity autocorrelation function. (*b*) Second- and third-order Eulerian longitudinal structure functions.

Throughout the manuscript we consider the energy dissipation rate as obtained from the third-order structure function, namely $\epsilon = 4.7 \times 10^{-4}\,\mathrm{m^2\,s^{-3}}$, which provides $\tau_\eta = 0.046\,\mathrm{s}$, $\eta = 2.14 \times 10^{-4}\,\mathrm{m}$ and the Taylor microscale $\lambda = u_{rms}\sqrt{15(\nu/\epsilon)} = 5.1 \times 10^{-3}\,\mathrm{m}$. We consider as integral time scale $T_I = L/u_{rms} = 0.65\,\mathrm{s}$. The corresponding Reynolds numbers are $Re_\lambda \approx 140$ and $Re_L \approx 530$. These estimates are the one that allows the most direct comparison with the numerical dataset (see Li *et al.* 2008).

As discussed in § 3.2, the 3D-PTV approach yields a set of trajectories with varying lengths. Shorter trajectories are more common than longer trajectories. Figure 10 shows the histogram of the number of trajectories per duration, binned per frame, demonstrating that the most likely trajectories have a duration of $5 - 10\,\tau_\eta$. As shown by the figure, longer trajectories are exponentially less frequent. The red line represents the number of trajectories longer than a given duration. Only ≈ 2500 trajectories longer than $20\tau_\eta$ are available; while considering trajectories as short as $10\tau_\eta$, the dataset can be extended to 32 500 trajectories.

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

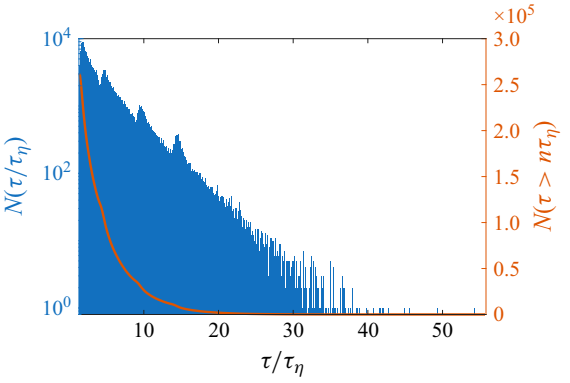


Figure 10. Number of trajectories per trajectory duration. Blue is the number of trajectories of a given duration, while the red curve is the cumulative sum, namely the number of trajectories longer than n times the Kolmogorov time scale.

Appendix B. Derivation of the scaling for the mean-square of the deviation-from-ballistic

We report the derivation of the mean-square of the deviation from ballistic motion presented by Obukhov (1959) and reported by Monin & Yaglom (2013), equation (24.13). We consider the deviation-from-ballistic vector $\delta_b \boldsymbol{x} = \boldsymbol{x}(\tau) - \boldsymbol{x}(t) - \boldsymbol{u}(t)\,\tau$ and derive both sides respect to τ:

$$\frac{\mathrm{d}\,\delta_b \boldsymbol{x}}{\mathrm{d}\,\tau} = \boldsymbol{u}(t+\tau) - \boldsymbol{u}(t). \tag{B1}$$

By multiplying both sides of (B1) by $\delta_b \boldsymbol{x}$, we obtain

$$\delta_b \boldsymbol{x} \cdot \frac{\mathrm{d}\,\delta_b \boldsymbol{x}}{\mathrm{d}\tau_2} = [\boldsymbol{u}(t+\tau_2) - \boldsymbol{u}(t)] \cdot \int [\boldsymbol{u}(t+\tau_1) - \boldsymbol{u}(t)]\,\mathrm{d}\tau_1. \tag{B2}$$

The left-hand side is equal to $(1/2)(\mathrm{d}\,\delta_b x^2/\mathrm{d}\,\tau_2)$, so integrating the above with respect to τ_2:

$$\delta_b x^2 = 2 \int [\boldsymbol{u}(t+\tau_2) - \boldsymbol{u}(t)]\,\mathrm{d}\tau_2 \cdot \int [\boldsymbol{u}(t+\tau_1) - \boldsymbol{u}(t)]\,\mathrm{d}\tau_1. \tag{B3}$$

Combining the integral and averaging gives

$$\left\langle \delta_b x^2 \right\rangle = 2 \iint \langle [\boldsymbol{u}(t+\tau_1) - \boldsymbol{u}(t)] \cdot [\boldsymbol{u}(t+\tau_2) - \boldsymbol{u}(t)] \rangle\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2. \tag{B4}$$

Expanding the terms in parentheses by computing the scalar product and using the isotropy of the velocity statistics yields

$$\left\langle \delta_b x^2 \right\rangle = 6 \iint [R(\tau_2 - \tau_1) - R(\tau_2) - R(\tau_1) + R(0)]\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 \tag{B5}$$

where $R(\tau - t) = \langle u(t+\tau)u(t) \rangle$ is the two-times velocity correlation function. The velocity correlations can be written in terms of the structure function, since, by definition, $S_2 = 2(R(t) - R(\tau - t))$. This yields

$$\left\langle \delta_b x^2 \right\rangle = 3 \iint [S_2(\tau_1) + S_2(\tau_2) - S_2(|\tau_2 - \tau_1|)]\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2, \tag{B6}$$

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

where the absolute value $|\tau_2 - \tau_1|$ is due to the symmetry of the correlation function. Within the inertial range of turbulence, we can plug the Kolmogorov scaling ansatz $S_2 = C_0\epsilon t$ where C_0 is the Lagrangian, single-component, Kolmogorov constant and get

$$\left\langle \delta_b x^2 \right\rangle = 3\, C_0\epsilon \iint (\tau_1 + \tau_2 - |\tau_2 - \tau_1|)\ \mathrm{d}\tau_1\, \mathrm{d}\tau_2 = 3\, C_0\epsilon \iint \min(\tau_1, \tau_2)\, \mathrm{d}\tau_1\, \mathrm{d}\tau_2, \quad \text{(B7)}$$

which, computing the integral leads to

$$\left\langle \delta_b \boldsymbol{x}^2 \right\rangle = \frac{C_0}{2}\epsilon\tau^3, \qquad \tau_\eta \ll \tau \ll T_I. \quad \text{(B8)}$$

For short times ($\tau \ll \tau_\eta$), we can approximate $\delta_b \boldsymbol{x}$ through a Taylor expansion as

$$\delta_b \boldsymbol{x} = \frac{1}{2}\langle \boldsymbol{a}^2 \rangle \tau^2 + O(\tau^3) \quad \text{(B9)}$$

which, using the scaling $\langle \boldsymbol{a}^2 \rangle \sim \eta/\tau_\eta^2$ leads to a power four scaling:

$$\langle \delta_b \boldsymbol{x}^2 \rangle \sim \eta^2 \left(t/\tau_\eta\right)^4. \quad \text{(B10)}$$

References

Arnéodo, A. *et al.* 2008 Universal intermittent properties of particle trajectories in highly turbulent flows. *Phys. Rev. Lett.* **100** (25), 254504.

Bec, J., Biferale, L., Lanotte, A.S., Scagliarini, A. & Toschi, F. 2010 Turbulent pair dispersion of inertial particles. *J. Fluid Mech.* **645**, 497–528.

Benzi, R., Biferale, L., Fisher, R., Lamb, D.Q. & Toschi, F. 2010 Inertial range Eulerian and Lagrangian statistics from numerical simulations of isotropic turbulence. *J. Fluid Mech.* **653**, 221–244.

Biferale, L., Bodenschatz, E., Cencini, M., Lanotte, A.S., Ouellette, N.T., Toschi, F. & Xu, H. 2008 Lagrangian structure functions in turbulence: a quantitative comparison between experiment and direct numerical simulation. *Phys. Fluids* **20** (6), 065103.

Biferale, L., Boffetta, G., Celani, A., Devenish, B.J., Lanotte, A. & Toschi, F. 2005 Lagrangian statistics of particle pairs in homogeneous isotropic turbulence. *Phys. Fluids* **17** (11), 115101.

Braun, W., De Lillo, F. & Eckhardt, B. 2006 Geometry of particle paths in turbulent flows. *J. Turbul.* **7**, N62.

Chen, H.Y., Liégeois, R., de Bruyn, J.R. & Soddu, A. 2015 Principal-component analysis of particle motion. *Phys. Rev. E* **91** (4), 042308.

Chevillard, L., Roux, S.G., Lévêque, E., Mordant, N., Pinton, J.-F. & Arnéodo, A. 2003 Lagrangian velocity statistics in turbulent flows: effects of dissipation. *Phys. Rev. Lett.* **91** (21), 214502.

Frisch, U. & Donnelly, R.J. 1996 *Turbulence: The Legacy of AN Kolmogorov*. American Institute of Physics.

George, W.K. 1988 Insight into the dynamics of coherent structures from a proper orthogonal decomposition. In *International Seminar on Wall Turbulence*. University of Buffalo, SUNY.

Hassanian, R. 2025 Proper orthogonal decomposition analysis of turbulent energy spectra within a Lagrangian particle tracking framework. *J. Appl. Comput. Mech.* e19545.

Hodžić, A., Olesen, P.J. & Velte, C.M. 2024 On the discrepancies between POD and Fourier modes on aperiodic domains. *J. Engng Maths* **145** (1), 10.

Karhunen, K. 1947 Über lineare methoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae* **37**, 1.

La Porta, A., Voth, G.A., Crawford, A.M., Alexander, J. & Bodenschatz, E. 2001 Fluid particle accelerations in fully developed turbulence. *Nature* **409** (6823), 1017–1019.

Li, T., Biferale, L., Bonaccorso, F., Buzzicotti, M. & Centurioni, L. 2025 Stochastic reconstruction of gappy Lagrangian turbulent signals by conditional diffusion models. *Commun. Phys.* **8** (1), 372.

Li, T., Biferale, L., Bonaccorso, F., Scarpolini, M.A. & Buzzicotti, M. 2024*a* Synthetic Lagrangian turbulence by generative diffusion models. *Nat. Mach. Intell.* **6** (4), 393–403.

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

LI, T., TOMMASI, S., BUZZICOTTI, M., BONACCORSO, F. & BIFERALE, L. 2024*b* Generative diffusion models for synthetic trajectories of heavy and light particles in turbulence. *Intl J. Multiphase Flow* **181**, 104980.

LI, Y., PERLMAN, E., WAN, M., YANG, Y., MENEVEAU, C., BURNS, R., CHEN, S., SZALAY, A. & EYINK, G. 2008 A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *J. Turbul.* **9**, N31.

LUMLEY, J.L. 1967 The structure of inhomogeneous turbulent flows. In *Atmospheric Turbulence and Radio Wave Propagation*, pp. 166–178. Nauka.

MAAS, H.-G., GRUEN, A. & PAPANTONIOU, D. 1993 Particle tracking velocimetry in three-dimensional flows: part 1. Photogrammetric determination of particle coordinates. *Exp. Fluids* **15** (2), 133–146.

MALIK, N.A., DRACOS, T. & PAPANTONIOU, D.A. 1993 Particle tracking velocimetry in three-dimensional flows: part II: particle tracking. *Exp. Fluids* **15** (4), 279–294.

MENEVEAU, C. & LUND, T.S. 1994 On the Lagrangian nature of the turbulence energy cascade. *Phys. Fluids* **6** (8), 2820–2825.

MONIN, A.S. & YAGLOM, A.M. 2013 *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence*, vol. 2. Courier Corporation.

MORDANT, N., METZ, P., MICHEL, O. & PINTON, J.-F. 2001 Measurement of Lagrangian velocity in fully developed turbulence. *Phys. Rev. Lett.* **87** (21), 214501.

OBUKHOV, A.M. 1959 Description of turbulence in terms of Lagrangian variables. *Adv. Geophys.* **6**, 113–116.

POPE, S.B. & CHEN, Y.L. 1990 The velocity-dissipation probability density function model for turbulent flows. *Phys. Fluids A: Fluid Dyn.* **2** (8), 1437–1449.

PUMIR, A., SHRAIMAN, B.I. & CHERTKOV, M. 2001 The Lagrangian view of energy transfer in turbulent flow. *EPL (Europhys. Lett.)* **56** (3), 379–385.

RICHARDSON, L.F. 1926 Atmospheric diffusion shown on a distance-neighbour graph. *Proc. R. Soc. Lond. A Containing Papers Math. Phys. Character* **110** (756), 709–737.

SAWFORD, B.L. 1991 Reynolds number effects in Lagrangian stochastic models of turbulent dispersion. *Phys. Fluids A: Fluid Dyn.* **3** (6), 1577–1586.

SCHIØDT, M., HODŽIĆ, A., EVRARD, F., HAUSMANN, M., VAN WACHEM, B. & VELTE, C.M. 2022 Characterizing Lagrangian particle dynamics in decaying homogeneous isotropic turbulence using proper orthogonal decomposition. *Phys. Fluids* **34** (6), 063303.

SHINDE, V.J. & GAITONDE, D.V. 2021 Lagrangian approach for modal analysis of fluid flows. *J. Fluid Mech.* **928**, A35.

SHNAPP, R. 2019 *Lagrangian Experimental Investigation of Dispersion in Canopy Turbulence*. Tel Aviv University.

SHNAPP, R. 2022 MyPTV: a python package for 3D particle tracking. *J. Open Source Softw.* **7** (75), 4398.

SHNAPP, R., BOHBOT-RAVIV, Y., LIBERZON, A. & FATTAL, E. 2020 Turbulence-obstacle interactions in the Lagrangian framework: applications for stochastic modeling in canopy flows. *Phys. Rev. Fluids* **5** (9), 094601.

SHNAPP, R. & BRIZZOLARA, S. 2026 Lagrangian_POD. GitHub. Available at: https://github.com/ronshnapp/Lagrangian_POD, accessed: 2026-07-17.

SHNAPP, R., BRIZZOLARA, S., NEAMTU-HALIC, M.M., GAMBINO, A. & HOLZNER, M. 2023 Universal alignment in turbulent pair dispersion. *Nat. Commun.* **14** (1), 4195.

SHNAPP, R. & LIBERZON, A. 2018 Generalization of turbulent pair dispersion to large initial separations. *Phys. Rev. Lett.* **120** (24), 244502.

SIROVICH, L. 1987 Turbulence and the dynamics of coherent structures. I. Coherent structures. *Q. Appl. Maths* **45** (3), 561–571.

TAN, S. & NI, R. 2022 Universality and intermittency of pair dispersion in turbulence. *Phys. Rev. Lett.* **128** (11), 114502.

TAYLOR, G.I. 1922 Diffusion by continuous movements. *Proc. Lond. Math. Soc.* **2** (1), 196–212.

THOMSON, D.J. 1987 Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *J. Fluid Mech.* **180**, 529–556.

TOSCHI, F. & BODENSCHATZ, E. 2009 Lagrangian properties of particles in turbulence. *Annu. Rev. Fluid Mech.* **41** (1), 375–404.

WEISS, J. 2019 A tutorial on the proper orthogonal decomposition. In *AIAA Aviation 2019 Forum*, p. 3333. TU Berlin.

WILSON, J.D., YEE, E., EK, N. & D'AMOURS, R. 2009 Lagrangian simulation of wind transport in the urban environment. *Q. J. R. Meteorol. Soc.: A J. Atmos. Sci. Appl. Meteorol. Phys. Oceanogr.* **135** (643), 1586–1602.

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press

DE WIT, X.M., GABBANA, A., WOODWARD, M., LIN, Y.T., TOSCHI, F. & LIVESCU, D. 2026 Data-driven Mori–Zwanzig modeling of Lagrangian particle dynamics in turbulent flows. *Proc. Natl Acad. Sci. USA* **123** (13), e2525390123.

XU, H., OUELLETTE, N.T. & BODENSCHATZ, E. 2007 Curvature of Lagrangian trajectories in turbulence. *Phys. Rev. Lett.* **98** (5), 050201.

YEUNG, P.-K. & POPE, S.B. 1989 Lagrangian statistics from direct numerical simulations of isotropic turbulence. *J. Fluid Mech.* **207**, 531–586.

https://doi.org/10.1017/jfm.2026.12041 Published online by Cambridge University Press