International Journal of Multiphase Flow 191 (2025) 105313

Contents lists available at ScienceDirect

International Journal of Multiphase Flow

journal homepage: www.elsevier.com/locate/ijmulflow

Research Paper

Inertial particle cluster dynamics in the core region of a turbulent duct flow

Tuhin Bandopadhyay [a,*], Alvaro Tomas Gil [a,b], Laura Villafañe [a]

[a] Department of Aerospace Engineering, University of Illinois, Urbana-Champaign, IL, USA
[b] Technical School of Design Engineering, Universitat Politècnica de València, Valencia, 46022, Spain

A R T I C L E I N F O

Keywords:
Particle clustering
DBSCAN
Cluster dynamics
Lifetime
Particle velocity divergence

A B S T R A C T

The spatial and temporal characteristics of clusters of small and heavy particles in fully developed turbulence are studied in the central region of a vertical duct flow. A series of time-resolved three-dimensional particle positions from direct numerical simulations is used as dataset. New methods are presented to define and track cluster structures that reduce the processing time of large datasets with large number of particles. Clusters are identified using a density-based clustering algorithm, and only a reduced set of all particles is used to correlate clusters in consecutive time instances exploiting morphological information. This approach eliminates the need for particle tracking when individual particle dynamics are not sought or known a priori. A hierarchy of clusters is evidenced with size and time of presence probability distributions exhibiting power-law behaviors. Clusters across all sizes are seen to predominantly evolve by breaking up into smaller parts and merging to form larger structures, with a significant number of structures also losing particles until disintegrating. The time clusters persist positively correlates, on average, with their size. The instantaneous volume rate of change of clusters scales with the turnover time of equivalent size eddies in the inertial range, supporting that clustering is driven by multi-scale turbulent eddies for $\mathrm{St}_\eta > \mathcal{O}(1)$. Clusters shrink on average at a faster pace than they grow, which is related to positive average particle velocity divergence within clusters, which increases with increasing local particle concentration.

1. Introduction

Small heavy particles dispersed in turbulent flows preferentially concentrate as they interact with the wide range of flow scales in turbulence (Fessler et al., 1994; Balachandar and Eaton, 2010; Bragg et al., 2015). Owing to their small but larger than fluid elements' inertia, particles segregate forming regions of high particle concentration that remain coherent for timescales larger than the flow dissipative scales, termed clusters (Monchaux et al., 2012). The size and particle number density distribution within these cluster structures, as well as their evolution within the turbulent flow, influence many natural and industrial processes. In heterogeneous (solid-gas) reactions involving char particles in turbulent oxidizers, densely clustered char regions can significantly reduce conversion rates by impeding oxygen transfer to their core (Haugen et al., 2018). Conversely, clustering can enhance chemical reaction rates by increasing surface contact between reactive phases (Reade and Collins, 2000). The spatio-temporal coherence of particle clusters influences the modulation of the underlying turbulent field (Hwang and Eaton, 2006; Xu and Subramaniam, 2010; Tom et al., 2022), and affects heat transfer within and radiation transmission through the mixture (Banko et al., 2020; Banko, 2018).

In the case of small particles with particle Reynolds numbers smaller than one and at low volume loadings, the regime of focus of this work, the Stokes number based on the dissipative flow scales, St_η, has been conventionally used as the main non-dimensional number indicative of particle preferential concentration. The Stokes number is defined as the ratio of the particle response time, τ_p, to a representative flow timescale, τ_l. The 'centrifuge mechanism' is one mechanistic explanation for preferential concentration, with inertial particles accumulating in strain-dominated flow regions and being centrifuged out of vortex cores (Maxey, 1987; Squires and Eaton, 1991; Eaton and Fessler, 1994). Preferential concentration in isotropic turbulence is most intense when St_η is close to unity, with the inhomogeneous distribution of particles persisting for St_η of $\mathcal{O}(0.1) - \mathcal{O}(100)$. For $\mathrm{St}_\eta > \mathcal{O}(1)$, preferential concentration can still be explained considering the influence of multi-scale turbulent eddies in the 'centrifuge mechanism'. Particles with Stokes number based on the turnover time of an l-scale eddy much smaller than one, $\mathrm{St}_l = \tau_p/T_l \ll 1$, preferentially sample the coarse-grained fluid velocity gradient tensor at scale l, with $T_l \sim \epsilon^{-1/3} l^{2/3}$ the corresponding eddy-turnover time in the inertial range (Bragg and Collins, 2014; Bragg et al., 2015). An alternative 'sweep-stick' mechanism by which

* Corresponding author.
E-mail address: tuhinb2@illinois.edu (T. Bandopadhyay).

https://doi.org/10.1016/j.ijmultiphaseflow.2025.105313
Received 31 January 2025; Received in revised form 8 May 2025; Accepted 1 June 2025
Available online 25 June 2025

particles stick to zero-acceleration points that are swept and clustered by large-scale motions (Yoshimoto and Goto, 2007; Goto and Vassilicos, 2008; Coleman and Vassilicos, 2009) has also been used to explain particle clustering. The equivalence between both theories when $\mathrm{St}_l \ll 1$ in the inertial range of turbulence was demonstrated in Bragg and Collins (2014).

The geometrical properties of cluster structures, defined in space as connected regions of particle concentration above a predefined threshold, display self-similarity across a wide range of scales from the dissipative to the flow integral scales. In particular, probability density functions (PDF) of clusters area and volumes display power-law decays (Monchaux et al., 2010, 2012; Sumbekova et al., 2017; Uhlmann and Chouippe, 2017; Baker et al., 2017; Banko, 2018; Mora et al., 2023; Dejoan and Monchaux, 2013), and power-laws indicative of fractal behavior are also observed for cluster perimeter versus area and surface area versus volume relations (Aliseda et al., 2002; Monchaux et al., 2012; Baker et al., 2017; Banko, 2018). These power-laws have been attributed to the self-similar nature of the turbulent eddies responsible for preferential accumulation, although there is no consensus on the exact values of the power-law exponents, and their dependence on flow and particle parameters or flow configuration is unclear, with most works focused on homogeneous turbulence.

The dynamics of clusters have received little attention compared to their averaged statistical properties. The temporal autocorrelation of local concentration (Tagawa et al., 2012) and of the standard deviation of a cluster concentration (Lian et al., 2019) were used in prior works to characterize the time-scales of preferential concentration in homogeneous turbulence. It was observed that particles lighter than the fluid remained clustered for longer than heavy or neutrally buoyant particles, and that clusters of both light and heavy particles persisted longer than the time-scales of the flow structures they correlated with (Tagawa et al., 2012). Using poly-disperse heavy droplets (Lian et al., 2019), cluster time-scales were seen to increase with turbulent Reynolds number, and to a lesser extent with increasing Stokes number. Most recently, the temporal coherence of three-dimensional clusters in homogeneous turbulence was studied using time-resolved particle data from one-way coupled DNS simulations with point-particle models (Liu et al., 2020). Probability density functions of cluster's lifetimes for two St_η and varying the settling parameter S_ν (by varying the gravitational acceleration) all for the same turbulent flow, were observed to follow exponential distributions with mean lifetimes from about 3 to 9 times τ_η. A positive correlation was observed between cluster's lifetime and volume. Overall, the lifetime increased mildly with increasing Stokes number for the same gravitational field, and likewise when the gravitational acceleration was increased for the same Stokes number. The authors noted that un-clustered particle clouds contracted exponentially at a faster rate prior to cluster formation than separated after cluster death, and that cluster formation and persistence coincided with small-scale turbulence activity rather than with intense fluctuations. The faster cloud contraction rate is consistent with faster backward-in-time than forward-in-time relative particle-pair dispersion (Sawford et al., 2005; Bragg et al., 2016), even if particle-pair dispersion and dispersion of clouds of particles of diverse size and shapes cannot be directly compared. The instantaneous local concentration rate of change was analyzed in Oujia et al. (2020) by quantifying the rate of change of Voronoi cell volumes tracked between consecutive time-steps. The relative local concentration rate of change provides the divergence of the particle velocity field. Positive velocity divergence was observed on average in regions with high local concentration, indicating positive particle divergence or active cluster destruction, and average negative velocity divergence or particles approaching in regions with low concentration. Interestingly, average positive velocity divergence values in high concentration regions were much larger than average negative velocity divergence values in regions of low concentration for the one-way coupled homogeneous turbulence point particle DNS data at varying St_η analyzed. The lower average negative velocity divergence contrasts with the faster contraction rates reported in Liu et al. (2020) at first sight, but both observations cannot be directly compared with one work analyzing local instantaneous rates of change and the other averaging over complete particle clouds that precede and follow a coherent cluster structure.

Analyzing cluster dynamics requires defining cluster structures in instantaneous particle distributions that are then correlated in successive instants. The two most common methods to identify clusters from discrete particle positions are box counting techniques and Voronoi tessellations. Box-counting is attractive for its simplicity but leads to implicit loss of resolution in dilute flows, with large box sizes required for meaningful particle number density fields (Villafañe-Roca et al., 2016; Esmaily et al., 2020). Voronoi tessellations have been widely adopted since their first application to particle-laden flows (Monchaux et al., 2010). They do not rely on a predefined grid length scale and directly provide the local particle concentration, inversely proportional to the Voronoi cell area or volume, if two- or three-dimensional data, respectively. Clusters are defined as connected cells smaller than a certain threshold which is obtained from comparison of the cell size probability distribution to that of particles distributed following a random Poisson process with equal average number density. As such, they also do not rely on arbitrary thresholds. Voronoi methods can get time-consuming in large three-dimensional datasets with millions of particles (Obligado et al., 2020), which can be palliated with more efficient tesselation implementations than those relying on the Quickhull algorithm such as `Voro++` (Rycroft, 2009). Once clusters are identified irrespective of the method used, connecting clusters between successive time realizations can be done by connecting cluster labels pre-assigned to each particle on each instantaneous field if particle track information is available, as is often the case from numerical simulations. Provided changes on the particle spatial distributions are small between successive steps, i.e. the time between realizations is sufficiently small, cross-correlation of Voronoi cells properties can also be exploited for tracking (Monchaux et al., 2010; Lian et al., 2019). If individual particle tracks or their instantaneous velocities are not known a priori, and only instantaneous particle positions are available, tracking all individual particles or Voronoi cells can get computationally intensive for large number of particles, numerous time instances, and three-dimensional data. The processing time for the simplest two-frame nearest-neighbor tracking method using the efficient KD-tree algorithm scales as $\mathcal{O}(\mathrm{N_p}\log(\mathrm{N_p}))$ (Bentley, 1975), where $\mathrm{N_p}$ is the number of particles, with additional multi-step track verification needed for tracking accuracy.

This work focuses on the dynamics of particle clusters in the core region of a fully developed turbulent flow laden with heavy particles smaller than the dissipative scales. It leverages data from DNS simulations with point particle Lagrangian models of a vertical turbulent square-duct flow, following the work by Esmaily et al. (2020). The dataset consists of three-dimensional particle coordinates, equidistant in time $0.5\tau_\eta$ for a total of 91 times the dissipative eddy turnover time. Only the central 60% of the duct cross-section is used in the present analysis, away from the regions where time averaged flow and particle velocities, and particle concentrations, present spatial gradients. On average, 4.2 million particles are present within the considered volume on each time realization. Novel methods are presented for cluster identification and temporal tracking that lighten the computational cost of analyzing cluster dynamics in large datasets. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm (Ester et al., 1996) was adopted as an alternative to Voronoi tessellations. Among the several data classification approaches based on density of data points (Xu and Tian, 2015), DBSCAN is particularly well suited for particle-laden flows as it identifies clusters of data, points in space in this case, of arbitrary shape (Schubert et al., 2017) and in an arbitrary number. Clusters are tracked between consecutive time instances using a reduced set of data points that preserves the cluster shape, without the need of tracking all particles or a priori knowledge of individual particle velocities. This lessens the processing time and makes the cluster tracking methods particularly appealing to datasets consisting exclusively of particle coordinates, as is often the case in experimental data. The dynamics of cluster structures are statistically examined, including the relative frequencies of different cluster birth and death processes, the time they remain coherent, and the correlation between cluster size and temporal characteristics. Instantaneous cluster volume growth rates are also analyzed for insights into the flow timescales affecting the evolution of clusters in the inertial range. Studying the evolution of macroscopic cluster quantities, such as volume, geometric characteristics, or the total number of particles, is important for understanding particle clustering and for developing simplified models of dynamical systems. However, a local perspective remains necessary to interpret cluster evolution results and relate them to particle dynamics. Instantaneous local relative concentration rates of change can be easily computed from the cluster tracking outputs. They reflect the local particle field divergence, which supports the cluster growth/shrink rate results.

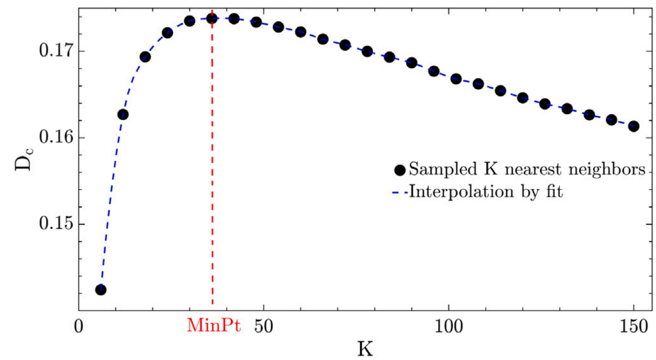
The manuscript is organized as follows. Section 2 introduces the new methodology for cluster identification and temporal tracking, including definitions for the various formation and disintegration processes that clusters may experience, and describes the dataset. All results are presented in Section 3 starting by a comparison of clusters identified using Voronoi tessellations and DBSCAN, and the computational cost of cluster identification and tracking. Cluster birth and death event statistics are presented next, followed by cluster coherence times, and instantaneous rates of change of macroscopic and local quantities-volume, and concentration, respectively. Concluding remarks highlight the main findings and discuss future analysis enabled and evidenced by this work.

2. Methods

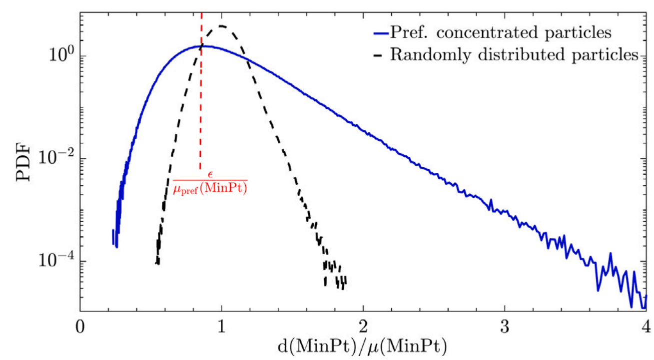
2.1. Cluster identification

Density-based Spatial Clustering of Applications with Noise (DBSCAN) is a data clustering method that groups data points into clusters based on a density threshold criterion (Ester et al., 1996). Compared to other unsupervised clustering algorithms, DBSCAN can identify any non-specified number of clusters of multiple sizes and shapes, is robust to noise or outliers, and is computationally fast. A brief description of DBSCAN is presented here, the reader is referred to Ester et al. (1996) for further details. The density threshold criterion is specified by two parameters, MinPt and ϵ, corresponding to the minimum number of data samples that must be contained within a hypersphere of radius ϵ around a data point for that point to be labeled as a core point. The algorithm begins by randomly examining the ϵ-neighborhood around a data sample, p. If the number of particles $N_\epsilon(p)$ is larger than MinPt, a cluster C is created that includes p as a core point and all its $N_\epsilon(p)$ neighbors. In the next step, the ϵ-neighborhood for all the points that have been assigned to the cluster C, $q \in C$, is analyzed. For each q point, if $N_\epsilon(q) > \mathrm{MinPt}$, all the $N_\epsilon(q)$ points that are not already in C are added and subsequently their ϵ-neighborhoods are analyzed. The algorithm proceeds until no new points can be added to that cluster and then continues to examine the remaining unprocessed points in the dataset to form new clusters. The result is the original list of data points with added labels identifying the cluster they have been assigned to, or labeled as noise if not assigned to any cluster. Clusters can contain core and non-core points, the latter necessarily located near the cluster edges without sufficient neighbors. When a non-core point is present within the ϵ-neighborhood of core points in different clusters, it is generally assigned to the first cluster meeting the criteria. To ensure the cluster identification results are deterministic, in the current work all non-core points are excluded from any cluster.

Data samples are in this application lists of particle positions, one list for each time instant. Three-dimensional data is used in this work,



(a) Deviation between the K-nearest neighbors distance distribution for preferentially concentrated and random particles, Eq. 1. The dash-line indicates the K = MinPt yielding $D_{c,max}$.



(b) MinPt – th nearest neighbor non-dimensional distance PDFs for preferentially concentrated and randomly distributed particles. Mean distance values are $\mu_{\mathrm{pref}}(\mathrm{MinPt}) = 8.6\eta$, $\mu_{\mathrm{random}}(\mathrm{MinPt}) = 9.7\eta$. The vertical dash-line indicates the smallest distance at which the distributions intersect, ϵ.

Fig. 1. Determination of DBSCAN parameters MinPt (a) and ϵ (b).

but the same methodology applies to two-dimensional data. The two input parameters needed for DBSCAN, MinPt and ϵ, are uniquely defined for any given dataset using a subset of uncorrelated particle lists, as many as needed for converged statistics, and by comparison to a set of particles with equal average number density distributed in space by a random Poisson process. As in any clustering identification method, the input parameters need to be evaluated only once for a dataset, and vary if flow or particle parameters change, as they depend on the average number density and degree of preferential concentration.

MinPt is defined as the number $n = K$ that yields the maximum deviation between the nth nearest neighbor distance probability distribution computed from the uncorrelated particle lists, and that of a random Poisson process. The deviation is quantified for varying $n = K$ number of nearest neighbors by $D_c(K)$, Eq. (1), where $\sigma(K)$ is the standard deviation of Kth nearest neighbor PDF for the preferentially distributed particles, and $\mu_{\mathrm{random}}(K)$, $\sigma_{\mathrm{random}}(K)$ are the mean and standard deviation, respectively, for the K-nearest neighbor PDF of the randomly distributed case. The analytical expressions for the probability density functions of Kth nearest neighbor distance for randomly distributed particles, as well as those for the mean and standard deviation are included in Appendix.

$$D_c(K) = \frac{\sigma(K) - \sigma_{random}(K)}{\mu_{random}(K)} \tag{1}$$

For preferentially concentrated particles $\mu(K) \leq \mu_{random}(K)$, with mean values converging to the same value for large K, of the order of the total number of particles. In contrast, $\sigma(K) > \sigma_{random}(K)$, resulting in $D_c \geq 0$. D_c behaves similarly to the box-counting preferential concentration index (Fessler et al., 1994) if K is interpreted as the box-size. For small and very large counts of nearest neighbors K both distributions have similar standard deviations and $D_c(K)$ is small. Fig. 1(a) shows

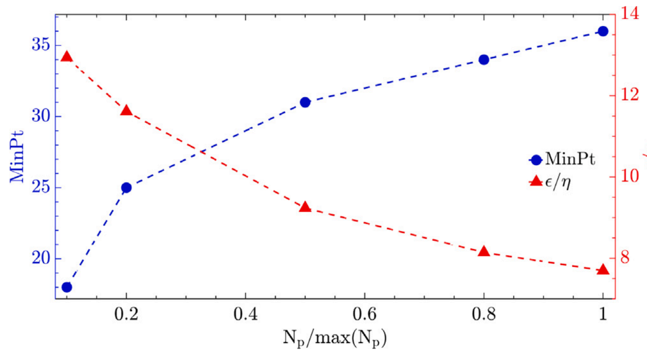


Fig. 2. Variation of $MinPt$ (left axis) and ϵ (right axis) for increasing number of particles in the dataset. The number of particles is reduced by random-sub-sampling.

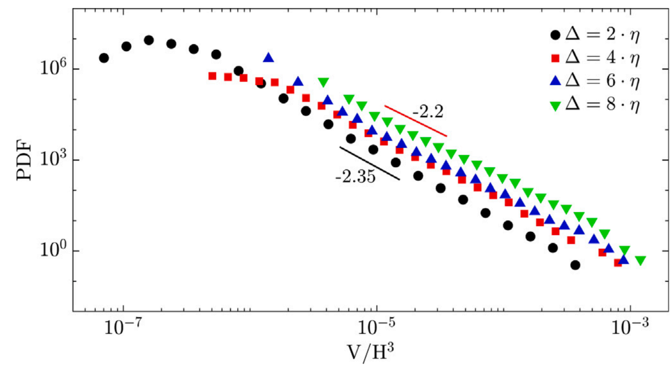


Fig. 3. PDFs of cluster volumes for varying grid size Δ. Cluster average inter-particle distance is $d_{s,c} = 1.9\eta$.

$\mathrm{D_c}$(K) from 50 uncorrelated three-dimensional particle distributions, each with about 4.2 million particles. The chosen MinPt is the value of K that yields the largest $\mathrm{D_c}$, i.e. $\mathrm{D_{c,max}} = \mathrm{D_c}(\mathrm{MinPt})$.

The distance parameter, ϵ, is defined as the smallest distance at which the MinPt-nearest neighbors PDFs for preferentially concentrated and randomly distributed particles intersect, as shown in Fig. 1(b), where the distance parameter is made non-dimensional using the mean distance value of the MinPt-nearest neighbor distribution for the preferentially concentrated particles. This criterion resembles the cell size threshold used by Voronoi based methods for cluster identification (Monchaux et al., 2010). Cluster identification results were seen to be little sensitive to small variations of ϵ, up to ±10% (not shown), for the dataset used in this work.

The proposed method serves to uniquely identify the input parameters to DBSCAN, MinPt and ϵ, which together serve as threshold criteria to define regions of high preferential concentration. As such, they depend on the particle and flow parameters and need to be re-evaluated following the procedure described if the average particle number density, the degree of preferential concentration, or both, vary, i.e. if the loading conditions, flow or particle parameters are varied. Fig. 2 shows the variation of the optimal MinPt and ϵ parameters if the average particle loading decreases, which is accomplished by randomly sub-sampling the particle field. This exercise is representative of a case in which the flow and particle parameters remain the same, and so does the degree of clustering, while the particle volume loading is altered, i.e. without two-way coupling effects being sufficient as to modify the flow and consequently the particle cluster structures. Using the optimal MinPt and ϵ for each new particle field with random sub-sampling provided in all cases the same cluster structures. The ability of the threshold parameters to adapt without user defined input parameters to the new particle field is precisely what makes this methodology, and Voronoi-based methods, robust for cluster identification. They are able to identify cluster structures that can be compared across cases with different average number densities, which is essential for preferential concentration comparisons (Villafañe-Roca et al., 2016).

DBSCAN is used in this work exclusively to identify regions of high particle concentration, or clusters, but it can be similarly used to identify regions of low particle concentration, or voids. It suffices to define a new set of input parameters, MinPt' and ϵ', and invert the clustering criterion by considering that particles are grouped into voids if the number of neighbors within an ϵ'-neighborhood is less than a threshold MinPt'. A particle P would be considered a void particle if its local number density falls below the specified threshold.

2.2. Cluster volume determination

A uniform Cartesian grid to which cluster particle positions are mapped is used to compute cluster volumes as the sum of connected cells of unit volume Δ^3 containing cluster particles. The choice of grid size does not affect the cluster particles already identified but it has an effect on the computed cluster volumes, with geometrical cluster properties defined from connected cells containing particles, which are sensitive to the number of particles, whether regular or irregular space tessellations are used. The simpler regular grid was chosen here because of the large number of cluster particles in the dataset, which limits the error on the volume estimates. The chosen value of Δ followed from a sensitivity analysis in which Δ was varied in the range $[2\eta - 8\eta] \approx [\mathrm{d_{sc}}, 4\mathrm{d_{sc}}]$ as shown in Fig. 3, with $d_{s,c} = 1.9\eta$ the mean inter-particle distance within cluster particles. For $\Delta < d_{s,c}$, many cells within the less concentrated cluster regions are devoid of particles and the volume is underestimated. Conversely, large values of Δ with many particles per cell overestimate cluster volumes by oversimplifying the cluster surface morphology. Power-laws are observed for over two decades for the probability density functions of cluster volumes shown in Fig. 3, suggesting self-similarity of cluster sizes as reported in prior works. The volume of the smallest and most frequent clusters for each value of Δ shown is that of the unit cell, with PDFs shifting to the right for increasing grid size as expected. The smallest grid size, on the order of the average inter-cluster particle distance, presents the largest deviation in the power-law exponent and the largest relative difference in average cluster volumes; the latter due to the presence of empty cells within clusters. With minimal differences between grid sizes 4 to 6η, $\Delta = 4\eta \approx 2d_{s,c}$ was chosen for this work.

2.3. Cluster tracking

High particle loadings and local number densities facilitate the identification of clusters and the quantification of their geometrical properties, but penalize the temporal analysis of clusters from the dynamics of their constituent particles when individual particle tracks or their instantaneous velocities are not known a priori. If Lagrangian quantities are not known but also not sought, correlating cluster structures across consecutive time instants does not necessarily require computationally costly particle tracking algorithms. Cluster structures can be tracked using a reduced set of all cluster particles that best represent the cluster boundary and its skeleton, and exploiting similarities in the local distribution of those particles in successive time steps. As in any tracking method, shorter time intervals between consecutive particle fields warrant higher accuracy, but the temporal resolution requirements are less stringent when compared to those needed for particle tracking. If in addition, the number of particles within clusters is sufficiently large, it is possible to simplify the morphological tracking search from three- to two-dimensions, while still capturing the three-dimensional cluster dynamics, and without loss of resolution.

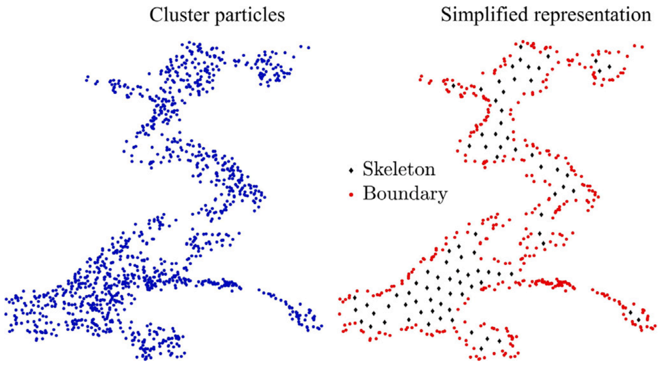


Fig. 4. Planar projection of all particles within a cluster on a δ_h thickness slice (left). Boundary and skeleton points representing the same cluster (right).

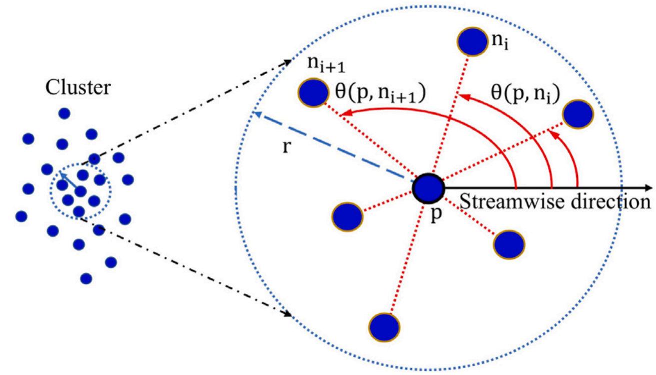


Fig. 5. Sketch of the definition of angular particle distribution used to identify boundary particles.

2.3.1. Data structure simplification

For each instantaneous particle distribution, DBSCAN outputs three-dimensional cluster particle positions, each particle with a cluster membership label. Instead of connecting particle membership labels to clusters in successive realizations directly in three-dimensions, the correlation search is simplified by dividing the domain in N_s non-overlapping thin slices, and connecting particle cluster labels across two-dimensional projections on each slice. On each planar projection only two distinct subsets of particles are used: particles defining the boundary of the bisected cluster, and particles defining the internal cluster skeleton, Fig. 4. Thus, the three pre-processing steps required on each instantaneous realization prior to tracking are sorting particles into planar projections, identifying boundary particles, and defining skeleton points.

Projection: Cluster particles are sorted into non-overlapping contiguous parallel thin volumes of thickness δ_h. Their center-plane to which particles are projected spans the full domain in the x and y, streamwise and wall parallel directions, respectively. Effectively, each two-dimensional projection slice simulates the data obtained by a camera imaging the light scattered by particles illuminated by a laser sheet of thickness δ_h. Small values of δ_h are desired to better resolve the three-dimensional cluster surface morphology while satisfying constraints on the minimum δ_h imposed by the particle displacement in the out of plane direction between realizations, and by the average inter-particle distance within clusters d_{sc}. The latter follows the same arguments as those involved in the determination of the grid size for calculating cluster volumes and is more stringent in the present work than the out-of-plane particle displacement constraint, given the duct flow configuration. A sensitivity analysis confirmed that the same cluster structures are identified in the two-dimensional projections for small variations of δ_h around the chosen value of $\delta_h = \Delta = 2d_{s,c} \approx 4\eta$. This observation agrees with the results reported in Monchaux (2012) from experiments in homogeneous turbulence for varying laser thickness, where the degree of preferential concentration was relatively unaffected for δ_h between 2η and 6η.

Cluster boundary: Boundary particles are identified for each two-dimensional projection and cluster label separately. For each cluster particle p, the angular distribution of distance vectors to surrounding particles is computed within a neighborhood radius $r = \epsilon$ of p, with ϵ the radius of search used in DBSCAN. Angles are defined by the distance vector with respect to the streamwise direction as shown in Fig. 5 schematic. Because particles located at the edges of two-dimensional clouds are not surrounded by particles in all directions, the angular separation between adjacent distance vectors can be exploited to identify boundary particles. As such, p is a boundary particle if $\Delta\theta_{p,i} = \theta(p, n_{i+1}) - \theta(p, n_i) > \delta_\theta$, that is, if the list of angular distances contains any entry for which the separation is larger than a threshold, with δ_θ defined as the average angular separation computed for all clusters. The threshold corresponded to $\delta_\theta = \pi/5$ in this work, with no significant variation in the number or distribution of boundary particles found varying δ_θ in the range $[\pi/5, \pi/2]$. The maximum angular separation for boundary particles, $\Delta\theta^{BP}_{p,max}$, informs about local cluster perimeter curvature.

Skeleton points: Skeleton points are a subset of particles within each two-dimensional cluster, similarly spaced independently of the local concentration, and the ones tracked between consecutive realizations to determine cluster dynamics. To identify them, all non-boundary particles within each cluster are first ranked based on their distance to their nearest boundary particle. The one furthest to any boundary is chosen as the first skeleton point, and all particles closer to it than a threshold distance k_c are eliminated. The process is repeated recurrently, choosing as next skeleton point the particle furthest from any boundary point of those remaining, until no more non-boundary particles remain. The threshold k_c regulates the number density of skeleton points, with large values reducing the number of skeleton points and therefore the tracking requirements, but also leading to increased uncertainty in cluster dynamics as narrow and short cluster branches may be devoid of skeleton points. On the other hand, there are no benefits to choosing k_c below the slice thickness as two-dimensional projections of three-dimensional positions cannot accurately resolve distances in the out-of-plane direction below δ_h. Since tracking in successive particle fields also relies on the distribution of nearby boundary points and not only on the distance between skeleton points, cluster tracking results are not very sensitive to small variations of k_c around a value that satisfies the prior guidelines. This was corroborated by varying $k_c \in [d_{s,c}, 4d_{s,c}]$. 10% of the total number of cluster particles were identified as skeleton points for the chosen $k_c = 4d_s \approx 8\eta = 2\delta_h$, greatly reducing the number of particles to track.

2.3.2. Temporal connectivity

For each of the N_s planar projections independently, skeleton points are paired between consecutive realizations, exploiting information about their distance to boundary points instead of exclusively on nearest neighbor searches. Cluster morphology information is encoded in a distance vector d_i^j of fixed length n generated for each skeleton point j, where n is the number of angular sectors of width $2\pi/n$ dividing the two-dimensional space around the skeleton point, and the streamwise direction is taken as angular reference. Each vector entry contains the distance between the skeleton point and the closest boundary particle within the corresponding angular sector. The number of angular sectors depends on the average distance between boundary points and between skeleton and boundary points, and thus is related to the average inter-particle distance within clusters $d_{s,c}$ and the choice of k_c. It should be chosen as large as possible while ensuring there are none or few empty entries on average in the distance vectors. $n = 20$ met that criterion in

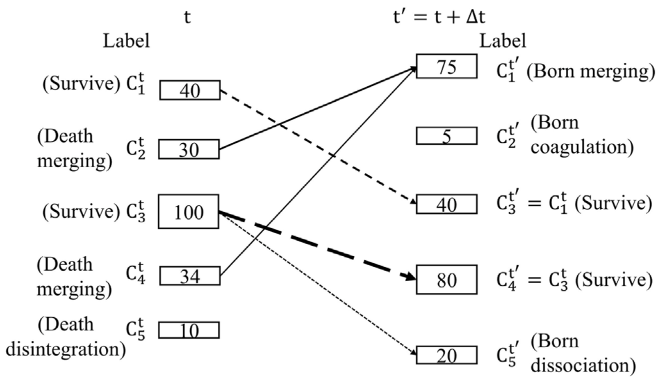


Fig. 6. Schematic exemplifying cluster connectivity and dynamic events classification. Left and right columns indicate clusters identified in two consecutive time realizations, with numbers in rectangular boxes representing the number of skeleton points on each cluster. Lines connect clusters with paired skeleton points.

this work, with cluster connectivity results invariant for n in the tested range 15 to 30.

For each skeleton point at time t potential matches at time $t' = t + \Delta t$ are those skeleton points that lie within a circular region r_{focus} of the average particle displacement $\langle \boldsymbol{u_f} \rangle \Delta t$, where $\langle \boldsymbol{u_f} \rangle$ is the local time-averaged flow velocity. The radius of search is defined for each skeleton point j at time t as

$$r_{focus,i} = \sqrt{u'^2} \cdot \Delta t + \overline{d^j} \tag{2}$$

where $\sqrt{u'^2}$ is the local time-averaged fluid velocity fluctuation, and $\overline{d^j}$ is the average of all entries in the distance vector d_i^j for the j skeleton point. Of all candidate skeleton points j' at t', the best match for point j at t is that with minimum Euclidean distance between their distance vectors, i.e the j' that minimizes $\left\| d_i^{j'} - d_i^j \right\|$. This ensures boundary points are similarly distributed around the paired skeleton points in time consecutive realizations.

Since each cluster particle retains its original cluster label assigned during cluster identification, once connectivity between skeleton points in consecutive times is determined in all N_s projections, the deformation and dynamics of a given cluster in three-dimensions can be evaluated.

2.4. Definition of cluster dynamics events

A cluster can undergo different processes as it evolves between time-instants t and $t' = t + \Delta t$, examples of which are shown schematically in Fig. 6, with lines indicating clusters with skeleton points that have been paired. Each cluster has a number label assigned during identification and a number of skeleton points (shown in a rectangular box), that is used to discriminate whether a cluster is considered the same or a different structure as it loses or gains particles. In addition to a cluster continuing as the same identifiable structure (surviving), six different cluster dynamic events can be defined, three resulting in the birth of new clusters and three equivalent ones that result in cluster death. A new cluster can emerge by coagulation of non-cluster particles, by merging of smaller clusters, or by dissociating from a larger cluster. Likewise, a cluster can die by disintegrating into non-cluster particles, by dissociating into smaller clusters, or by merging into a bigger cluster.

To unequivocally determine which of these dynamic processes are occurring during each time interval, the presence of paired skeleton points alone is not sufficient when clusters share them with more than one cluster in the following or preceding time instance. Forward and backward in time connectivity of skeleton points is required, and a threshold of 50% in common skeleton points is used to deem a cluster as the same structure, as done in Liu et al. (2020).

Table 1
Relevant flow and particle parameters.

Parameter	value
Bulk velocity, U_b [m/s]	7.7
Re_H	$2.1 \cdot 10^4$
Kolmogorov length scale, η [m]	$7 \cdot 10^{-5}$
Kolmogorov time scale, τ_η [s]	$0.33 \cdot 10^{-3}$
Convective timescale, $\tau_c (= H/U_b)$ [s]	$5.2 \cdot 10^{-3}$
$\bar{d}_p/\eta$	0.17
ρ_p/ρ	7416.7
Stokes number, $\bar{St}_\eta$	12
Re_p	0.4
Volume loading, ϕ_v	$2.7 \cdot 10^{-5}$

The examples in Fig. 6 evidenced the need for thresholds and bidirectional connectivity. C_3^t shares skeleton points with $C_4^{t'}$ and $C_5^{t'}$, which could be considered two new clusters in t' with C_3^t losing its identity, or it could be considered that since $C_4^{t'}$ shares more skeleton points with C_3^t these are the same structure that survives while a new smaller one $C_5^{t'}$ is born at t' from dissociation. Using the 50% bidirectional connectivity threshold, the latter is considered to occur as C_3^t and $C_4^{t'}$ show 100% and 80% forward and backward connectivity, respectively, while C_3^t and $C_5^{t'}$ show 100% forward but only 20% backward connectivity. Similarly, C_2^t and C_4^t, both of similar size, are seeing to merge into $C_1^{t'}$ which is considered a new cluster born by merging as it does not have 50% of its skeleton points in common with any of the prior two. A cluster at t can share skeleton points with multiple clusters at t′, if it fails the 50% common skeleton points threshold with any of them it is considered to have died by dissociation with the clusters at t' labeled as born by dissociation from a larger cluster. Clusters not connected to any cluster at the subsequent time instant and not exiting through domain boundaries have died by disintegration, and analogously born by coagulation if not connected to any cluster in the prior time instance and have not just entered through the domain boundaries.

Using these definitions of clusters' survival, birth, and death across a long time series, it is possible to map the complete history of each cluster within the domain. Clusters that cross the domain boundaries are ignored to avoid biases due to partial time tracks, and only those clusters that are born and die within the domain are used for the cluster dynamics analysis. The total time a cluster survives is defined as '*time of presence* ($t_{presence}$)'. It is often a time in between other dynamic events, as at birth it may keep some of the particles from a prior cluster that died by dissociation or merging, and once it dies some of its particles may continue clustered on a new cluster with a different label. To avoid possible biases on $t_{pres.}$ statistics, '*lifetime* (t_{life})' is considered as the time of presence for the subset clusters that die exclusively by disintegration within the domain, which cannot be counted multiple times with successive births and deaths by merging and dissociating.

2.5. Description of the dataset

A time-resolved series of three-dimensional particle positions from direct numerical simulations with point particle models for the discrete phase is leveraged in this work. The simulations reproduced a vertical turbulent square-duct flow of air laden with small heavy particles, replicating a companion experiment. A brief summary of the dataset is presented here, with details regarding the numerical framework and joint numerical and experimental setup described in Esmaily et al. (2020). The Reynolds number based on the bulk flow velocity and duct width is $Re_H = 2.1 \cdot 10^4$. The carrier fluid is air with $\rho = 1.2$ kg/m^3 and $\mu = 1.88 \cdot 10^{-5}$ Pa s. The dissipative time and length scales reported in Table 1 were obtained from the mean pressure drop along the development length. Nickel particles were used with $\rho_p = 8900$ kg/m^3, particle diameters ranging between 4 and 24 μm, and an average diameter $\bar{d}_p = 12$ μm. Particle parameters in Table 1 are based on the nominal average diameter. Particles are smaller than the dissipative

Table 2
Overall comparison of cluster particles and clusters identified by DBSCAN and Voronoi.

Metric	Voronoi	DBSCAN
No. cluster particles/ No. particles, in %	32	36
Avg. no. of clusters per realization	22 500	1410
Average cluster size ($V^{1/3}$)	0.01H	0.02H

flow scales, justifying the use of point-particle models, and the average Stokes number based on the Kolmogorov scale is 12. Simulations accounted for gravity, particle polydispersity, binary particle collisions using a hard sphere model, and particle reaction force on the fluid. The particle drag force used the Schiller–Naumann drag correlation for finite Reynolds numbers (Naumann and Schiller, 1935). The particle Reynolds number was 0.4 in average, with $\mathrm{Re_p} = \left\|\mathbf{u}_\mathrm{p} - \mathbf{u}_\mathrm{f}\right\| \mathrm{d_p}/\nu$, where $\mathbf{u}_p$ is the particle velocity and $\mathbf{u}_f$ is the undisturbed fluid velocity at the particle location, which was estimated by interpolating the local fluid velocity at the particle location while minimizing the error in drag force calculation (Esmaily and Horwitz, 2018). The particle mass loading ratio is 20%, which corresponds to a volume loading of about $2.7 \cdot 10^{-5}$. An iterative matching procedure was used to find the number of particles of each size required to replicate a given particle size distribution and mass loading ratio (Esmaily et al., 2020).

Only results in the central 60% of the square duct width were used in this work to avoid regions of non-uniform average flow and particle quantities. The domain length in the streamwise direction is 6.75 duct widths. In the reduced domain analyzed, each instantaneous realization contains 4.2 million particles on average. The time interval between consecutive particle distributions available for the analysis is $\Delta \mathrm{t} = 0.03\tau_\mathrm{c} = 0.5\tau_\eta$, with τ_c the convective time based on the duct width and bulk velocity. The time series covers a total time equivalent to $91.4\tau_\eta$ (200 realizations), which was seen as sufficient for converged statistics of cluster dynamics.

3. Results

3.1. Cluster statistics

Statistics of clusters identified using the new DBSCAN-based methodology are compared to those obtained using Voronoi tesselation-based methods, evaluated on the same set of uncorrelated three-dimensional particle fields. On average, 32% and 36% of all particles on each realization are identified as cluster particles by Voronoi and DBSCAN, respectively. Of all the cluster particles identified by Voronoi, 88% are also cluster particles by DBSCAN. Despite the lesser total number of cluster particles identified by Voronoi, and following cell connectivity, Voronoi methods lead to a much larger number of small clusters, which results in a smaller average cluster size compared to the clusters identified using DBSCAN. These overall comparison metrics are summarized in Table 2.

The differences in number of clusters and average cluster size between both methods are not due to fundamental differences in local number density thresholds, although they vary slightly, or to any constraint on the number of particles required to form a cluster, which none of the methods includes. They are instead rooted on the different methods to discriminate whether each individual particle in the dataset is a cluster particle or not. Voronoi tessellation-based methods label individual particles as cluster particles if the corresponding cell volume is smaller than a threshold, to then form a cluster out of cells sharing a common facet of the tessellation, i.e. by connectivity. The volume threshold is defined by the smallest volume at which the probability distribution function of all cell volumes intersects with that of a random Poisson particle distribution with the same average number density. The local concentration threshold is $C_{th,V} = 2.3C_0$, with $C_{th,V}$ setting

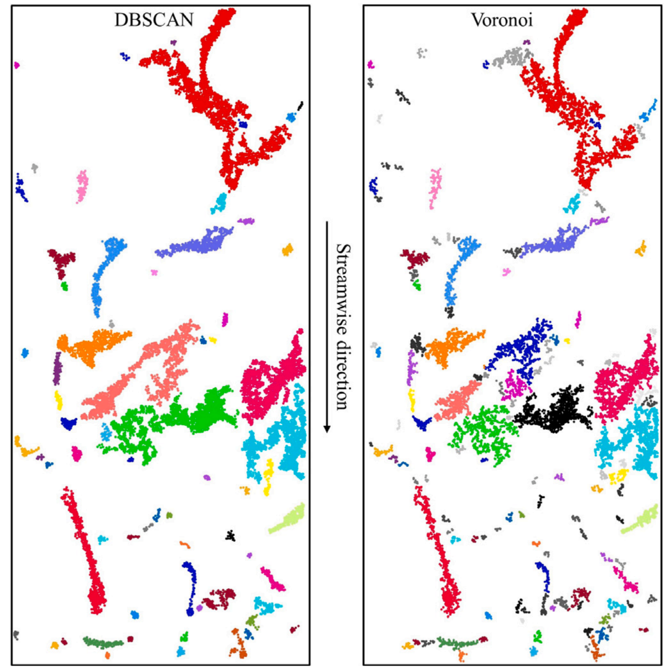
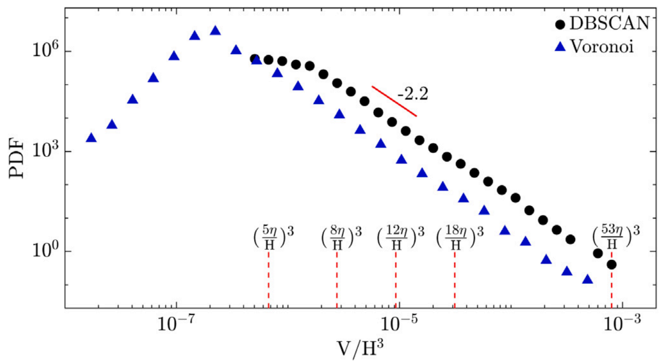


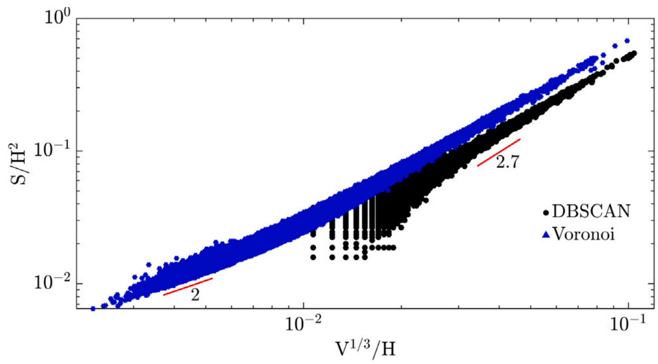
Fig. 7. Clusters identified using DBSCAN (left) and Voronoi (right) based methods. The same three-dimensional particle fields are used and only a thin slice in which particles are projected onto a 2D plane is shown for visual comparison. Different colors represent independent clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the minimum cell concentration (or maximum cell size) for the corresponding particle to be considered a cluster particle. There is no explicit constraint on the number of nearby particles but there is an implicit dependence on how neighboring particles are distributed; e.g. particles with small pairwise distance but aligned forming a filament will yield large tessellation cells and will not be considered cluster particles. This is the reason for the larger number of small clusters identified using Voronoi tessellations, and for the lesser number of cluster particles, with clusters losing connectivity and effectively splitting into smaller structures along thin branches. DBSCAN-based methods instead do not impose any requirement on the relative local arrangement of particles. A minimum local concentration is required for a particle to be discriminated as a cluster particle, MinPt within a sphere of radius ϵ around each candidate particle, but this local concentration criteria can be met by particles along a line filament as well as if particles are distributed such that the total average interparticle distance is minimized. The equivalent number density threshold using the MinPt and ϵ values defined by comparison to a random Poisson particle distribution with equal average number density is $C_{th,D} = MinPt/(4/3\pi\epsilon^3) = 2C_0$. Note that the number MinPt is only used to discriminate individual particles as cluster particles, but does not restrict in any way the total number of cluster particles forming a given cluster. It is possible for one particle P to meet the criteria of having at least MinPt particles within a sphere of radius ϵ centered at P, while the same criteria is not satisfied by any of the other particles around P, in which case DBSCAN would yield a cluster with only one cluster particle.

In conclusion, although Voronoi and DBSCAN yield similar but not identical number density thresholds (in both cases defined by comparison to a random Poisson particle distribution), the differences in overall cluster statistics are primarily due to the fact that the volume of Voronoi tessellations depends on the spatial arrangement of surrounding particles. As a result, Voronoi tessellation-based methods break into distinct clusters structures with thin branches, leading to a larger number of smaller clusters, while DBSCAN does not. This is qualitatively visible in the planar comparison shown in Fig. 7, where only clusters with



(a) PDFs of cluster volumes normalized by the duct width.



(b) Scatter plot of surface area versus unit length volume for all clusters, normalized by the duct width.

Fig. 8. Comparison of geometrical cluster properties identified using Voronoi and DBSCAN.

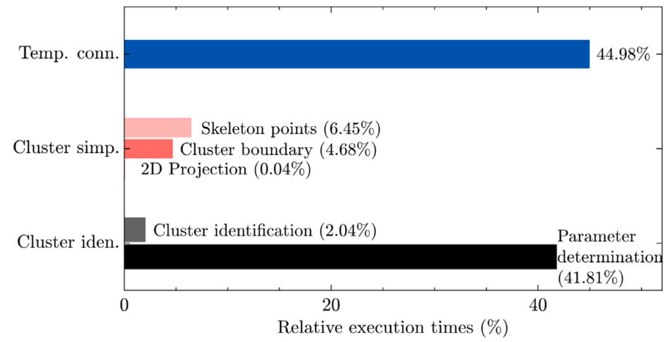


Fig. 9. Relative execution times (in % of the total reference execution time of 9233 s) for cluster identification, cluster simplification, and temporal cluster connectivity, evaluated on a 3D particle field containing 4.2 M particles.

more than 20 cluster particles are displayed. The resemblance in the morphology of the clusters identified by both methods is remarkable despite the differences on the total number and average size of cluster structures.

Cluster volumes and surface area statistics are compared in Fig. 8. The cluster volumes PDF is shifted to the right in Fig. 8(a) for DBSCAN clusters with respect to that from Voronoi with large number of small clusters in agreement with the observation that Voronoi tessellations cause clusters to split into smaller clusters. Small differences are also expected on the geometric cluster metrics compared, volume and surface area, which originate from the different methods used to define the continuous three-dimensional space that the clusters, collections of discrete particles, occupy. Voronoi tessellations directly provide cell volumes, each containing one particle, that can be added. For the clusters identified using DBSCAN a Cartesian grid is used to define the space the cluster particles occupy, as described in Section 2.2. This constrains to $\Delta^3 = (4\eta)^3$ the volume of the smallest DBSCAN clusters while Voronoi clusters can yield any volume value, thus the differences on the smallest volume counts in Fig. 8(a). Beyond the differences due to the spatial discretizations, both cluster volume PDFs showcase power-laws for about three decades with the same -2.2 exponent. Linear log–log fittings cover cluster length scales in the range 8 to 53η for DBSCAN, which extends to about 4η for Voronoi. The exponent is slightly larger than values reported in prior works for three-dimensional data in homogeneous turbulence, ranging between $-16/9$ and -2 (Dejoan and Monchaux, 2013; Uhlmann and Chouippe, 2017; Baker et al., 2017).

The similitude between clusters identified by both methods is further corroborated by the scatter plots of cluster surface versus unit length volume, shown in Fig. 8(b). Cluster surface area is computed as the sum of the cell surface areas which are not shared with any other cell in the same cluster. For the large clusters, a least square fit yields a common power-law exponent of 2.7, confirming morphological similarity of the structures identified using both methods. The non-integer exponent is characteristic of fractal structures with convoluted surfaces (Isichenko, 1992). For homogeneous turbulence, exponents ranging between 2.85 and 3 have been reported in Baker et al. (2017) for three-dimensional data. The smallest clusters, not shown in DBSCAN results because of the chosen cartesian grid resolution used to compute volume, yield an integer exponent of 2, characteristic of regular three-dimensional objects. These are clusters formed by few nearby particles that satisfy the minimum Voronoi cell volume criteria, but it has been debated whether these particles are brought to close proximity by the underlying turbulence or are the effect of randomness, as suggested in (Baker et al., 2017).

3.2. Computational performance

The relative execution times for all tasks required to identify and track clusters using the methods described in Section 2 are reported in Fig. 9 as a percentage, with the total execution time used as a reference including the sum of: (1) determination of hyper-parameters (MinPt, ϵ) and cluster identification in one three-dimensional particle field using the DBSCAN module from `scikit-learn` in Python; (2) cluster simplification for one particle field, which involves planar projections, and determination of boundary and skeleton points for all clusters in all projections; (3) temporal connectivity of clusters in two consecutive time instances. The total reference execution time for these tasks is 9233 s. Execution times were evaluated on a desktop computer with an AMD Ryzen 3700X 8-Core Processor running at 4 GHz with 64 GB of RAM. Parameter determination takes a significant amount of the total reference time, but it only needs to be performed once for a given dataset in as many three-dimensional particle fields as needed for convergence. This means that once MinPt and ϵ are fixed, the execution time for subsequent instantaneous particle distributions, tracking included, is 58% of the reference total time reported. The number of fields needed for convergence of MinPt and ϵ depends on the specific dataset. Twenty fields sufficed in this work.

Considering only cluster identification, no tracking nor the cluster simplification required for tracking, the total execution time including parameter determination is reduced to 4048 s, 43.85% of the reference. For comparison, the execution time for cluster identification on one three-dimensional field using the `Spicy` Voronoi Python module for spatial tessellation is 85% higher than with DBSCAN, including in both cases the determination of the thresholding parameters. This comparison includes the spatial tessellations (98% of the total cluster identification time), determination of cell volume threshold (negligible), classification into cluster and void cells, and cell connectivity. With Voronoi, cluster identification on each particle field (time instance) involves the same computational cost. With DBSCAN,

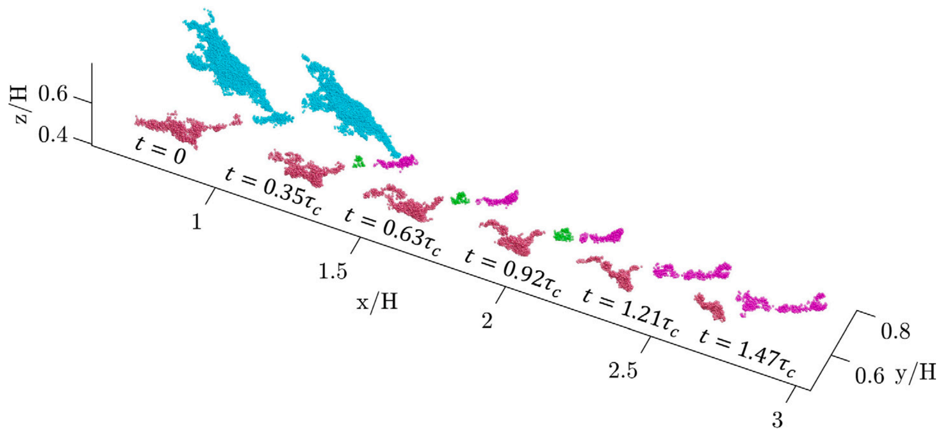


Fig. 10. Illustration of cluster dynamics. Color coded clusters move in the x-streamwise direction. Two smaller clusters are formed by dissociation from a larger cluster at $t = 0.35\tau_c$ and merge at $t = 1.21\tau_c$, while a cluster (red) slowly shrinks until disintegration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

once the hyper-parameters are determined, cluster identification on each additional particle field takes 4.6% of the prior 4048s total cluster identification time. Thus, the computational cost savings with DBSCAN when compared to Voronoi Python packages become increasingly advantageous as the number of instantaneous fields to be analyzed increases. The higher cost of Voronoi tessellations using the `Spicy` Python package is due to the Quickhull algorithm, which has a theoretical average time complexity of $\mathcal{O}(\mathrm{N_p}\log(\nu_\mathrm{p}))$, and a worst-case time complexity of $\mathcal{O}(\mathrm{N_p^2})$, with $\mathrm{N_p}$ the number of particles and ν_p the number of Voronoi cell vertices (Barber et al., 1996). The number of vertices ν_p scales non-linearly with $\mathrm{N_p}$, depending on the particles' spatial distribution. Instead, the theoretical complexity of DBSCAN is $\mathcal{O}(\mathrm{N_p}\log(\mathrm{N_p}))$ when utilizing the KD-tree algorithm (Ester et al., 1996), independently of the complexity of the particle distribution. However, more efficient implementations of Voronoi tessellations are available in C++, such as `Voro++` (Rycroft, 2009), which can offset the cost of the spatial tessellations. This library optimally exhibits $\mathcal{O}(\mathrm{N_p})$ scaling for homogeneously distributed particles, with performance degrading for highly irregular or clustered distributions up to a worst-case time complexity $\mathcal{O}(\mathrm{N_p}\cdot\mathrm{N_n})$, with $\mathrm{N_n}$ number of neighboring particles. Thus, DBSCAN compares very favorably to the Quickhull algorithm for easily implementable Voronoi tessellations in Python, and has a similar computational cost when compared to more efficient Voronoi implementations available in C++, with marginal gains for very large number of particles per field and high degree of clustering.

The methods developed in this work aim at facilitating cluster tracking in datasets with large numbers of particles and time realizations, which is independent from the method used for the prior necessary cluster identification. The same methodology used in this work for tracking can be used with clusters identified using DBSCAN or Voronoi methods. Cluster simplification accounts for 11.17% of the total reference execution time, and the subsequent temporal connectivity of skeleton points between two consecutive time instances for 44.98%, as shown in Fig. 9. The approach used for connectivity of skeleton points follows a Lagrangian particle-tracking approach that involves nearest-neighbor search, with a theoretical computational complexity of $\mathcal{O}(N_p \log N_p)$, with the KD-tree algorithm. About 36% of all particles in the domain are cluster particles, Table 2, but only 27% of those are skeleton points. The computational cost is thus significantly reduced by tracking only the skeleton points instead of all cluster particles as it would be needed with conventional cluster tracking. In addition, mapping the distribution of neighboring boundary points instead of that of neighboring particles to identify the best match for a skeleton point in successive time instants further reduces the cost.

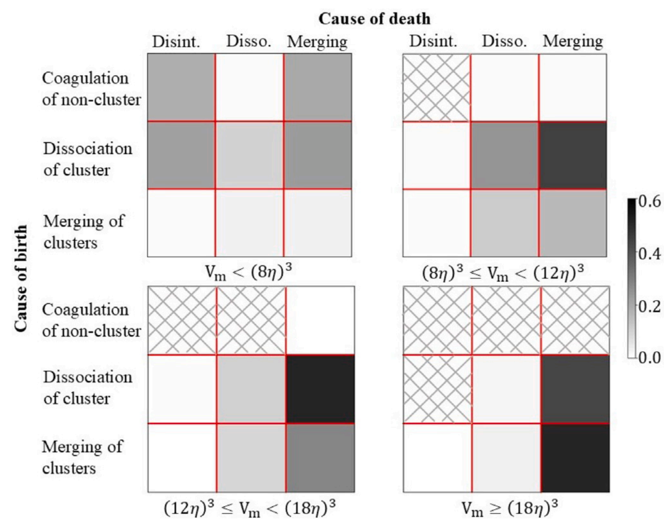


Fig. 11. Joint probabilities of cluster birth–death events conditioned on life-averaged cluster volume. Zero probability is indicated with cross-line pattern.

3.3. Cluster birth–death events

Particle clusters exhibit complex dynamics as they evolve in the turbulent flow and interact with neighboring structures. Fig. 10 shows the three-dimensional evolution of selected clusters that undergo dissociation, merging, and disintegration, as per the definitions of dynamic events described in Section 2.4. The clusters whose size extends to the boundary of the domain at any instant over their time of presence are rejected to avoid bias due to the finite domain size. Discarded clusters also include those with less than three particles, equivalent to an average volume of one grid cell, and those present exclusively in one instantaneous realization without any forward or backward connectivity.

The joint probabilities of cluster birth and death events are shown in Fig. 11. Given the differences in number of clusters of different sizes (Fig. 8(a)), the probabilities of birth–death events are further broken into four bins based on the average volume of the clusters during their lifetime, $\mathrm{V_m}$. Bin ranges are chosen such that each bin contains enough clusters for statistically significant results. The largest clusters, $\mathrm{V_m} \geq (18\eta)^3$, are observed to predominantly form and die by merging, with dissociation the second most likely event for new clusters. Since a 50% forward and backward connectivity is required for a given cluster to continue as the same entity, these are large

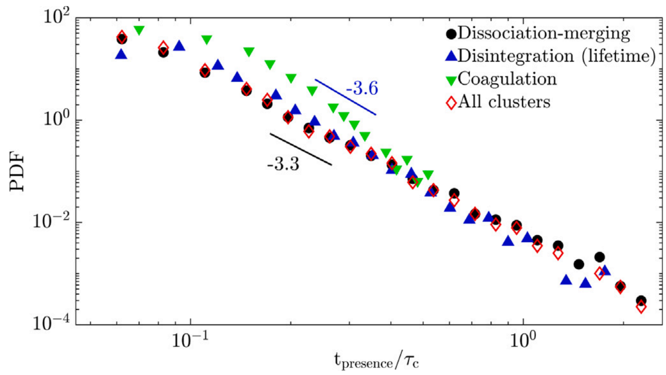


Fig. 12. PDFs of cluster time of presence for: all clusters (empty circles), clusters vanishing into non-cluster particles (lifetime, triangular markers), and clusters undergoing dissociation or merging (filled circles).

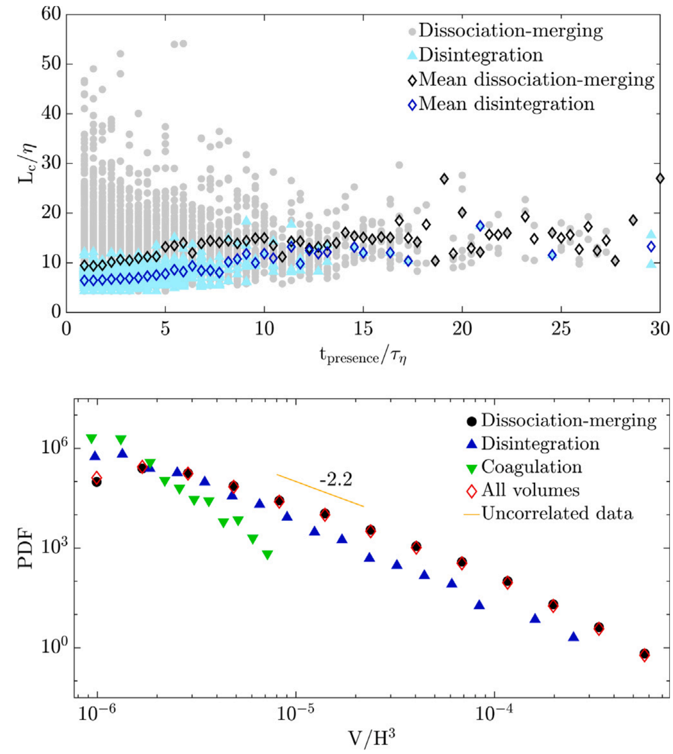


Fig. 13. Relation between characteristic cluster size and time of presence (top), and cluster volume PDFs (bottom), for different birth–death events.

clusters splitting (dissociating) into likely still large structures and coming together. Moderately large clusters, $(12\eta)^3 \leq V_m < (18\eta)^3$, show the same dominant birth and death dynamics, but the likelihood of new clusters emerging from a larger one that separates is now larger than that of clusters merging into a larger one, as this larger cluster is counted on the largest bin. The same trend is also observed for the next smaller bin, that of clusters with average volume $(8\eta)^3 \leq V_m < (12\eta)^3$. As the average volume decreases and the probability of new clusters forming by dissociation increases, also the probability of losing their cluster label identity (dying) by dissociation increases. That is, these medium- to small-size clusters not only lose small parts, but they do so losing their cluster identity. The smallest clusters, $V_m < (8\eta)^3$, typically emerge from two similarly frequent processes: from parts of larger clusters or from non-clustered particles coming together, and they die by either merging into a larger structure counted on another bin, or by disintegrating to non-clustered particles.

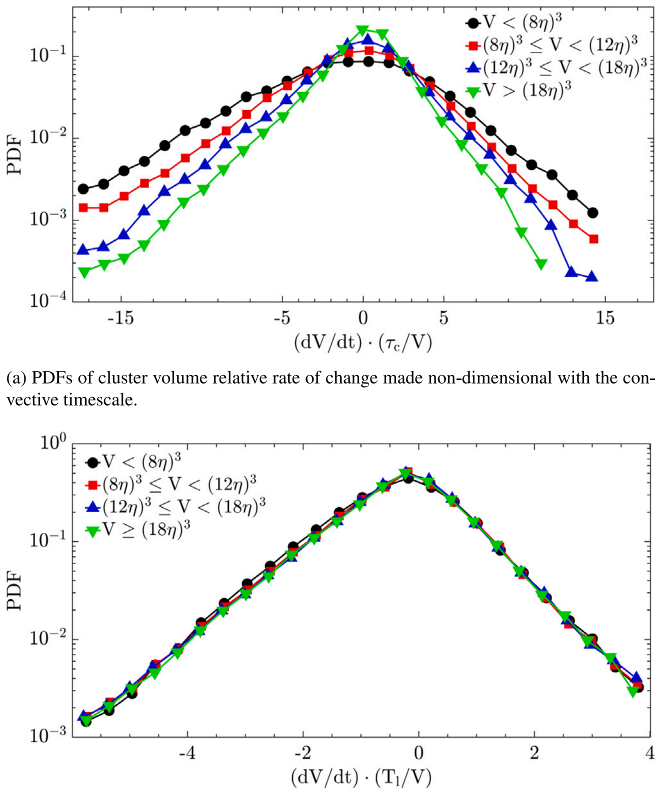
The joint probabilities highlight that clusters larger than 8η evolve mainly by losing parts that then join a larger cluster, setting a continuous cycle of dissociation and merging. Birth by dissociation from a larger cluster and death by merging onto another cluster accounts for 45% of all possible birth–death events, excluding those clusters smaller than $(8\eta)^3$ in average, and for 28% when all clusters independent of size are considered. All combinations of birth/death by dissociation or merging considered together account for 93% of the events for clusters of average volume larger than $(8\eta)^3$, and 74% for all sizes.

Clusters smaller than 8η are mostly very short-lived (discussed next) and therefore their life-average size remains small. Because they are the most numerous, number-based statistics, irrespective of structure size, can overshadow the dynamic events of larger structures, the most relevant for particle clustering in the inertial range of turbulence. The number of cells with zero probability of clusters born by coagulation increases in Fig. 11 as the life-average cluster volume increases, indicating that a cluster growing into a large structure in isolation, without interacting and eventually merging to other structures, is unlikely within the considered domain. Birth by coagulation represents 36% of all smallest clusters birth events, 14% of all birth events across sizes, and 2.5% if excluding the smallest bin of life-averaged volumes. These results differ from those in Liu et al. (2020) for homogeneous turbulence, where a major dominance of birth by coagulation and death by disintegration was observed for all Stokes and settling parameters analyzed.

Death by disintegration is also most frequent for the smallest life-averaged size clusters, with 33% of all smallest clusters death events, 12% across all sizes, and 4.5% excluding the smallest volume bin. However, although lesser in number than death by splitting or merging, the occurrence of death by disintegration is highly relevant to the temporal evolution of clusters for two reasons: (1) clusters that evolve until disintegration are observed across all sizes (as will be shown in Fig. 13); and (2) tracking these clusters allows for unbiased time of presence-volume relationships. The latter is challenged by clusters with the same constituent particles that may split and merge several times, and each time be counted as an independent birth–death event.

3.4. Cluster coherence times

Cluster coherence time, or time of persistence, is the total time elapsed between cluster birth and death, with lifetime specifically reserved for clusters vanishing by disintegration into non-cluster particles. The probability distribution of time of persistence made non-dimensional by the convective flow timescale is shown in Fig. 12 for: all clusters independently of their birth–death mechanism, clusters forming and dying by any dissociation-merging combination, clusters forming by coagulation, and those undergoing disintegration. Power-laws are observed in all cases except for the shortest-lived clusters born by coagulation. The distributions for dissociation-merging and 'all clusters' overlap displaying a −3.3 slope, while a slightly steeper linear fit of −3.6 is obtained for disintegrating clusters. The overlap is expected as dissociation-merging constitutes 74% of all death-birth events. PDFs extend from many clusters identified during only one dissipative eddy-turn over time (three consecutive fields) to few of them that persist one or two convective times, about 15 to 30 Kolmogorov time-scales. Given the finite domain length and time series, the maximum time a cluster could remain as the same entity is necessarily smaller than $5.8\tau_c$. The power-law exponents were not affected by reducing the domain length down to 20% (not shown), only the largest achievable coherence time decreased. The self-similarity of coherence times is consistent with the notion that clustering of small and heavy particles with $St_\eta > \mathcal{O}(1)$ is dominated by eddies within the inertial range of turbulence (Bragg and Collins, 2014; Bragg et al., 2015). Exponential distributions were reported instead of power-laws in Liu et al. (2020) for the coherence times of clusters in homogeneous isotropic turbulence, although it



(a) PDFs of cluster volume relative rate of change made non-dimensional with the convective timescale.

(b) PDFs of cluster volume relative rate of change made non-dimensional with the turnover time of an eddy of equivalent size, with $l \sim V^{1/3}$, $T_l = \epsilon^{-1/3}V^{2/9}$

Fig. 14. The variation of cluster volume per unit time for all clusters is shown relative to their volume. Data is binned by volume.

could be argued that both fits yield similar coefficients of determination for the range of lifetimes reported. The exponential fit allowed the authors to identify a characteristic cluster timescale, 3 to $9\tau_\eta$ for the different Stokes and settling parameters considered, of the same order as those reported from Lagrangian auto-correlations in prior works in homogeneous turbulent flows (Tagawa et al., 2012; Lian et al., 2019). Similar exponential fits would yield similar characteristic times in the present work. To our knowledge, Liu et al. (2020) is the only prior work that explored coherence times and thus the frequent comparisons.

The differences in the decay rate of coherence times PDFs for the different groups examined can be explained by the differences in the PDFs of cluster volumes for the same groups, and the relation between the average size of a cluster and the time it remains coherent, both shown in Fig. 13. Scatter plots of cluster characteristic size based on each cluster life-averaged volume, $L_C/\eta = (V_m/\eta)^{1/3}$, and their average, are shown versus cluster time of presence in Fig. 13-(top) for clusters undergoing dissociation-merging and death by disintegration separately. Dissociation-merging events are predominant across all cluster sizes, and the large scatter is expected as large and small structures lose their cluster identity as they dissociate into smaller parts or merge into a large one. Their time of persistence is a time between dynamic events, with some of their constituent particles remaining clustered although with a different cluster label. Despite the scatter, the average displays a positive trend with larger clusters surviving the longest. The same trend is evident for clusters that vanish by disintegration, with lesser scatter and smaller averaged characteristic sizes. In both cases, there is a positive relation between average cluster characteristic size and time of presence, with an equivalent linearized slope of less than 0.5 when size and time are made non-dimensional with the dissipative scales.

Uncorrelated time realizations provided a power-law PDF of cluster volumes with a slope of −2.2 (Fig. 8(a)). Fig. 13-(bottom) showcases similar cluster volume PDFs using correlated data (all volumes across the time of presence of a cluster included), with cluster volume probabilities divided into the same groups of birth–death mechanisms as before. The volume probabilities for all clusters and those undergoing dissociation and merging show a perfect overlap, with a decay rate similar to uncorrelated data although for a narrower range of volumes. The narrower range is expected as only the subset of all clusters that were born and died within the domain is considered in the dynamic analysis. The volume probability of disintegration events presents a steeper decay compared to dissociation-merging. Since a similar average trend is observed for both groups between life-averaged cluster size and time, the steeper decay translates into an also steeper time of presence slope. The same argument can be used to justify the fast decay of time of presence of the smallest clusters born by coagulation.

3.5. Volume rate of change and lifetime prediction

Provided the apparent self-similarity of macroscopic cluster characteristics is caused by the self-similarity of turbulent flow scales, the rate of change of cluster volumes should be related to the inertial range scaling. The relative volume rate of change can be computed between any two consecutive time realizations as the volume difference per unit time, made non-dimensional by the average volume between realizations and the convective time scale, Eq. (3).

$$\frac{dV}{dt}\cdot\frac{\tau_c}{V} = \frac{V(t+\Delta t)-V(t)}{\Delta t}\cdot\frac{2\tau_c}{V(t+\Delta t)+V(t)} \tag{3}$$

The probability density functions for all clusters conditioned on volume are depicted in Fig. 14(a). The distributions exhibit exponentially stretched tails for volume increase and decrease, characteristic of Laplace distributions frequently encountered in turbulence Chevillard et al. (2006). The distributions become narrower with increasing cluster volume, with larger clusters experiencing on average slower relative volume rate changes compared to smaller ones, which leads to an increase in lifetime with cluster size. The trends invert if the absolute change in volume per unit time is considered (not shown), with larger clusters on average experiencing larger changes of their absolute volume when compared to smaller clusters for the same increment of time. All PDFs have negative mean and skewness, indicating that clusters experience on average a faster rate of volume decrease than increase, i.e they shrink faster than grow.

The convective timescale used to make the relative volume rate change non-dimensional is a constant factor in this case. If instead the turnover time of an eddy of the same size as the cluster is used, the PDFs become independent of cluster volume, Fig. 14(b). The characteristic time of an eddy in the inertial range is given by $T_l \approx \epsilon^{-1/3}l^{2/3}$, where ϵ is the average kinetic energy dissipation rate, and V/T_l can be interpreted as the volume rate of change of a turbulent eddy of size $l \sim V^{1/3}$. The collapse of the PDFs suggests that the decay and growth rates of the clusters are indeed driven by the underlying self-similar turbulent eddies. It also implies that the average relative cluster volume rate of change, regardless of its size, is inversely proportional to the corresponding eddy turnover time, i.e. $\langle dV/V\,dt\rangle \propto 1/T_l$. A relationship between the volume of a cluster and its initial volume follows from the exponential tails of the distribution, with different average rates for clusters growing or shrinking. The least-square fit of the left exponential tail provides a mean decay rate of $\langle dV/dt\rangle = -1.02V/T_l$. Relating T_l to the cluster volume and integrating yields the absolute average change in volume of a shrinking cluster during a given time interval as a function of the average flow dissipation rate, Eq. (4).

$$V_t^{2/9} - V_0^{2/9} = -\frac{2}{9}\frac{1.02}{\epsilon^{-1/3}}(t-t_0) \tag{4}$$

For clusters shrinking until disintegrating ($V_t = 0$), without interacting with other structures, Eq. (4) provides an average relation to

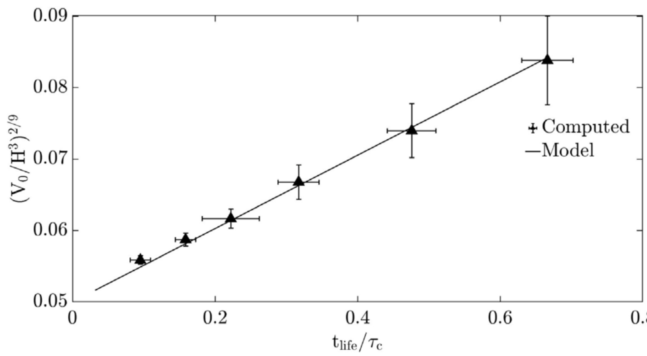


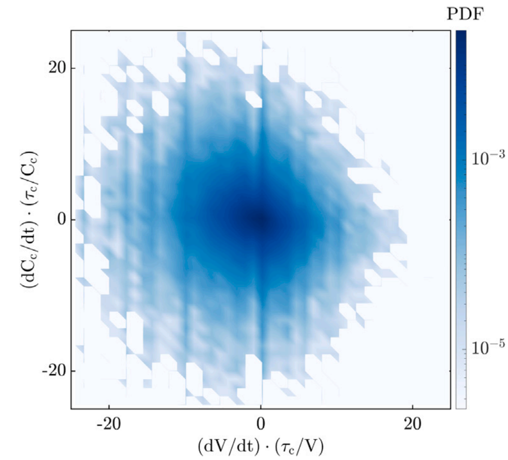
Fig. 15. Initial volume - lifetime relationship for shrinking clusters disintegrating. Error bars indicate the standard deviation for each lifetime bin from the data. The solid line corresponds to the average relation, Eq. (4).

estimate the lifetime of a cluster based on its initial volume, V_0 at t_0. This relation is shown in Fig. 15 together with averaged initial volume-lifetime pairs extracted from the data, made non-dimensional by the channel width and convective time. The error bars indicate the standard deviation of the initial volumes and lifetimes for each lifetime bin used to reduce the data. With both model and sampled data emerging from the same dataset, the agreement serves as verification. Eq. (4) provides a similar average relation as that shown in Fig. 13-(top) in terms of life-time average volume.

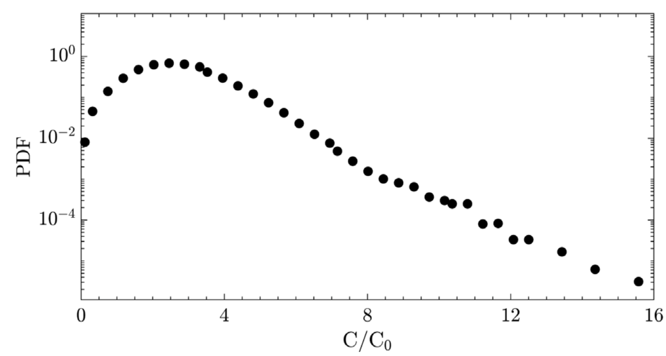
3.6. The role of particle concentration on cluster evolution

Cluster averaged particle concentration PDFs (not shown) are wide, with PDFs conditioned in cluster size all yielding similar average values despite the smaller tails of larger structures. Consequently, the instantaneous rate of change of cluster averaged concentration shows no correlation with that of cluster volume, as reflected in the joint probability scatter plot in Fig. 16(a). This indicates that cluster-averaged concentration is not a good quantity of interest to understand the temporal evolution of cluster structures, and that the number of particles within clusters, and their rate of change, are on average linearly related to cluster volumes, and their rate of change. A local perspective into the evolution of regions of dissimilar concentration, which can vary largely within clusters, brings insights into the contraction/expansion of the particle field, which can play an important role in the clusters' longevity. The probability density function of local concentration around all cluster particles, computed within a sphere of radius $r = 4\eta$, is shown in Fig. 16(b). Values up to 16 times the domain averaged concentration C_0 are observed, with maximum values that depend on the filter size r. The contraction or expansion of the particle field at the filter scale r can be easily evaluated using the connectivity map between skeleton points in consecutive time instances exploited for cluster tracking. The instantaneous concentration rate of change is defined analogously to the volume rate of change, with the local concentration used to compute relative rates defined, as before, as the average between consecutive realizations.

Distributions of relative concentration rate of change made non-dimensional with the convective timescale are shown in Fig. 17, binned on local concentration. For increasing concentration, the mean and skewness of the distributions become more negative. The heavy right tails representing increasing local concentrations (positive rate) drop faster than the left tails as the concentration increases. The same distributions made non-dimensional by the domain average concentration, C_0, display inverted relative trends (not shown) with tails flattening for increasing local concentration, but the negative skewness and high kurtosis remain. An important observation is that local particle concentration within clusters decreases on average at a faster rate than



(a) Joint PDF of cluster averaged concentration and volume rates of change.



(b) PDF of local filtered concentration surrounding cluster particles, filter size $r = 4\eta$.

Fig. 16. Cluster averaged (top) and local (bottom) statistics evidencing the lack of correlation between average concentration and cluster volume rates of change, and the large variation of particle number density within clusters, which is masked by cluster average concentrations.

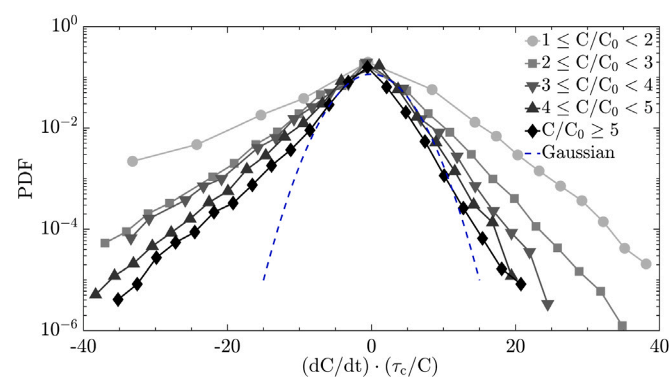


Fig. 17. Relative local concentration rate of change PDFs conditioned on concentration.

it increases, pointing to particles in the clustered regions separating at a faster rate than they converge, i.e. cluster destruction. Similar observations were made in Oujia et al. (2020) from one-way coupled inertial point particle simulations in homogeneous turbulence using a different method to evaluate the local rate of change of concentration. The authors exploited the inverse proportionality between the volume of a Voronoi cell and the local concentration, and used the relative contraction and expansion of tracked Voronoi cells during a short time increment to compute the local concentration rate of change, which equals the particle velocity divergence with opposite sign. Negative average values of relative concentration rate of change (positive average values of particle divergence) were observed for small cell volumes, which correspond to highly concentrated regions in clusters, and vice versa for large cell volumes. With increasing Stokes number, larger values for both the negative and positive sides of the PDFs were reported, more intense cluster formation for large, and destruction for small, Voronoi cell volumes. With simulations not accounting for collisions or reaction force by particles on fluid, crossing particle trajectories due to caustics were presented as possible reason. The numerical dataset used in the current work includes particle collisions and reaction force by particles on fluid. Furthermore, the change of concentration between consecutive time instances is evaluated on a constant volume sphere ($V \approx (4\eta)^3$) centered on a tracked skeleton cluster particle. Thus, while Voronoi-based methods provide the material derivative of the local concentration by tracking the volume change of a cell with constant mass (one particle), the current method evaluates a coarse-grained instantaneous concentration rate of change within a constant volume control volume. Despite the differences on the methods to evaluate local concentration rates of change, on the flow and particle parameters, and on the real effects simulated, the qualitative similarities remain, pointing to a robust asymmetry of the local concentration rate of change with negative average values in cluster regions with high local concentration.

The average negative and negatively skewed local concentration rate of change probability distribution resembles the equally negatively averaged and skewed cluster volume rate of change distribution, even if they refer to local and macroscopic quantities, respectively. Combined, they indicate that clusters shrink on average at a faster rate as their regions of high concentration see their concentration diminished at a faster rate, i.e. shrinking clusters lose particles at a faster rate. Faster rates of decrease of local and global properties are balanced by slower but more numerous growth events, such that the cluster averaged rate of concentration increase or decrease over all clusters is the same. This is, the total number of cluster particles in the domain remains statistically constant as expected in a fully developed flow.

It should be remarked that the results presented here exclusively consider particles that are in cluster regions at the time of evaluation, and not those in more dilute regions not identified as cluster particles. Although not shown, it has been confirmed using Voronoi tracking in two-dimensional projections that non-cluster particles exhibit heavy positive tails of relative concentration rate of change, with positive skewness and positive mean values. Considering all particles in the domain, irrespective of local concentration, yielded a symmetric distribution of concentration rate of change. This equilibrium is expected considering the fully developed statistically stationary flow. The symmetry of the cluster-averaged concentration rate of change in Fig. 16(a) also verifies this global equilibrium.

4. Concluding remarks

The temporal evolution of clusters of inertial particles was studied in the central region of a fully developed turbulent duct flow using a series of three-dimensional particle fields equispaced by half the dissipative timescale. The large number of particles and instantaneous particle fields motivated the development of novel methods for identifying and tracking particle clusters. These tools enabled the statistical analysis of cluster spatial and temporal characteristics.

Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was shown ideally suited to identify particle clusters, defined as regions of high particle concentration. A methodology was developed to uniquely define the two input parameters to DBSCAN, that serve as criterion to identify particle clusters, from the dataset. As in other cluster identification methods using box-counting and Voronoi tessellations, the criterion used to discern regions highly concentrated particles was derived from the comparison of the specific particle field of interest and that produced by a random Poisson process, both with equal average number density. Structures identified using the new methodology are in agreement with those obtained from Voronoi tesselation-based methods. While this work focused on particle clusters, DBSCAN can similarly be extended to identify regions of low particle concentration, or voids. DBSCAN is a robust alternative that can significantly reduce the computational time for cluster identification in large datasets when compared to space tesselation implementations using the standard Quickhull algorithm.

A novel cluster tracking methodology was introduced that reduces the computational cost when compared to that of tracking all individual cluster particles. Clusters are reduced to internal skeleton points, for which connectivity is determined across time-consecutive particle fields, and boundary particles, that define the cluster surface and are essential to the morphological tracking used to map skeleton points. In addition, this work leveraged the large number of particles to reconstruct three-dimensional cluster temporal evolution from planar tracking, further simplifying the data processing. The methods described can be used on two- or three-dimensional particle fields, and are advantageous from a computational standpoint for large series of time-resolved data with high particle loading when cluster evolution is aimed for, rather than Lagrangian particle statistics.

For the fully developed particle-laden channel flow under study, with small inertial particles and a Stokes number of 12 based on the dissipative scales, particle clusters across a wide range of sizes majorly evolve by breaking (dissociating) into smaller parts and merging to form larger structures. To a lesser extent, clusters of all sizes are also seen to lose particles until disintegration, while clusters growing in isolation are limited to small and short-lived structures. A hierarchy of clusters is present, whose size and time of presence probability distributions display power-law like probability density functions. The self-similarity of cluster sizes (volume or area) has been reported in multiple prior works from uncorrelated particle fields in homogeneous and fully developed turbulent flows, and has been attributed to the self-similarity of eddies in the inertial range of turbulence, even though the power-law exponent value and its relation to the flow and particle parameters remain elusive. The like-wise power-law behavior of cluster time of presence (or coherence time) is reported here for the first time. Independently analyzing size and coherence time statistics for structures undergoing dissociation-merging, and those evolving until disintegration without interacting with other clusters, shows the same type of statistical distributions, although with steeper power-laws for dissociating clusters with lesser number of large structures.

The instantaneous relative volume rate of change of clusters is seen to scale with the turnover time of eddies in the inertial range with equal reference length, leading to volume independent rates of change. This finding agrees with the notion of clustering driven by multi-scale turbulent eddies for $\mathrm{St}_\eta > \mathcal{O}(1)$, with particles with Stokes number based on the turnover time of an l-scale eddy $\mathrm{St}_l = \tau_p/T_l \ll 1$ preferentially sampling the coarse-grained fluid velocity gradient tensor at scale l, with $T_l \sim \epsilon^{-1/3} l^{2/3}$ (Bragg and Collins, 2014; Bragg et al., 2015). When integrated, the shrinking rate provides an analytical relation between cluster initial volume and expected lifetime that agrees with the values sampled from the time history of disintegrating clusters. Growth and shrinking rate relations, together with cluster size probability distributions, can be used for simplified models of cluster dynamics.

Average instantaneous cluster growth and shrink rates were found to be asymmetric, with clusters shrinking at a faster pace. Similarly, negatively skewed probability distributions were identified for local rates of change of relative particle concentration within clusters, with skewness that increases in absolute value for increasing local concentration. Negative local rate of change of relative concentration translates to positive local particle velocity divergence, which implies that particles within clusters separate at a faster rate than they approach, explaining the faster rate of volume decrease compared to growth. Despite these asymmetries in the temporal evolution of clusters, there is an equilibrium on the total number of cluster particles at any given time, as expected in a fully developed turbulent flow.

This work statistically explored the dynamics of particle clusters in fully developed turbulence from a macroscopic perspective, with some insights from instantaneous contraction and expansion rates of filtered particle concentration within clusters. Further analysis considering relative particle velocities conditioned in concentration, and local flow quantities, is required to explain the asymmetry in the probability distributions of volume and local concentration rates of change. Extending the analysis to other flow and particle parameters is also essential to evaluate the role of particle inertia and gravitational effects on the cluster dynamics, and generally, particle clustering in the inertial range of turbulence.

CRediT authorship contribution statement

Tuhin Bandopadhyay: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Alvaro Tomas Gil:** Methodology, Conceptualization. **Laura Villafañe:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the PSAAP II program at Stanford University, and Dr. Jaravel specifically, who provided the numerical data used in this work.

Appendix. Kth nearest neighbor distance pdf for random particle distribution

The probability density function of Kth nearest neighbor distance for randomly distributed particles in three-dimensions with average number density λ obeys the following analytical expression (Illian et al., 2008),

$$f_K(d) = 4\pi\lambda d^2 e^{-\frac{4}{3}\pi\lambda d^3} \frac{\left(\frac{4}{3}\pi\lambda d^3\right)^{K-1}}{(K-1)!} \tag{A.1}$$

where d is the distance to the Kth nearest neighbor. The mean and standard deviation of the randomly distributed particles are given by the following expressions, where $\Gamma(\cdot)$ is the Gamma function.

$$\mu_{random}(K) = \left(\frac{3}{4\pi\lambda}\right)^{1/3} \frac{\Gamma(K+1/3)}{\Gamma(K)} \tag{A.2}$$

$$\sigma_{random}(K) = \sqrt{\left(\frac{3}{4\pi\lambda}\right)^{2/3} \frac{\Gamma(K+2/3)}{\Gamma(K)} - \left(\left(\frac{3}{4\pi\lambda}\right)^{1/3} \frac{\Gamma(K+1/3)}{\Gamma(K)}\right)^2} \tag{A.3}$$

Data availability

Data will be made available on request.

References

Aliseda, A., Cartellier, A., Hainaux, F., Lasheras, J.C., 2002. Effect of preferential concentration on the settling velocity of heavy particles in homogeneous isotropic turbulence. J. Fluid Mech. 468, 77–105.

Baker, L., Frankel, A., Mani, A., Coletti, F., 2017. Coherent clusters of inertial particles in homogeneous turbulence. J. Fluid Mech. 833, 364–398.

Balachandar, S., Eaton, J.K., 2010. Turbulent dispersed multiphase flow. Annu. Rev. Fluid Mech. 42, 111–133.

Banko, A.J., 2018. Radiation Absorption by Inertial Particles in a Turbulent Square Duct Flow (Ph.D. thesis). Stanford University.

Banko, A.J., Villafañe, L., Kim, J.H., Eaton, J.K., 2020. Temperature statistics in a radiatively heated particle-laden turbulent square duct flow. Int. J. Heat Fluid Flow 84, 108618.

Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls. ACM Trans. Math. Softw. (TOMS) 22 (4), 469–483.

Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. Commun. ACM 18 (9), 509–517.

Bragg, A.D., Collins, L.R., 2014. New insights from comparing statistical theories for inertial particles in turbulence: I. Spatial distribution of particles. New J. Phys. 16 (5), 055013.

Bragg, A.D., Ireland, P.J., Collins, L.R., 2015. Mechanisms for the clustering of inertial particles in the inertial range of isotropic turbulence. Phys. Rev. E 92 (2), 023029.

Bragg, A.D., Ireland, P.J., Collins, L.R., 2016. Forward and backward in time dispersion of fluid and inertial particles in isotropic turbulence. Phys. Fluids 28 (1).

Chevillard, L., Castaing, B., Lévêque, E., Arnéodo, A., 2006. Unified multifractal description of velocity increments statistics in turbulence: Intermittency and skewness. Phys. D: Nonlinear Phenom. 218 (1), 77–82.

Coleman, S., Vassilicos, J., 2009. A unified sweep-stick mechanism to explain particle clustering in two-and three-dimensional homogeneous, isotropic turbulence. Phys. Fluids 21 (11).

Dejoan, A., Monchaux, R., 2013. Preferential concentration and settling of heavy particles in homogeneous turbulence. Phys. Fluids 25 (1), 013301.

Eaton, J.K., Fessler, J., 1994. Preferential concentration of particles by turbulence. Int. J. Multiph. Flow 20, 169–209.

Esmaily, M., Horwitz, J., 2018. A correction scheme for two-way coupled point-particle simulations on anisotropic grids. J. Comput. Phys. 375, 960–982.

Esmaily, M., Villafane, L., Banko, A., Iaccarino, G., Eaton, J., Mani, A., 2020. A benchmark for particle-laden turbulent duct flow: A joint computational and experimental study. Int. J. Multiph. Flow 132, 103410.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd. Vol. 96, pp. 226–231.

Fessler, J.R., Kulick, J.D., Eaton, J.K., 1994. Preferential concentration of heavy particles in a turbulent channel flow. Phys. Fluids 6.

Goto, S., Vassilicos, J., 2008. Sweep-stick mechanism of heavy particle clustering in fluid turbulence. Phys. Rev. Lett. 100 (5), 054503.

Haugen, N.E.L., Krüger, J., Mitra, D., Løvås, T., 2018. The effect of turbulence on mass transfer rates of small inertial particles with surface reactions. J. Fluid Mech. 836, 932–951.

Hwang, W., Eaton, J.K., 2006. Homogeneous and isotropic turbulence modulation by small heavy particles. J. Fluid Mech. 564, 361–393.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley & Sons.

Isichenko, M.B., 1992. Percolation, statistical topography, and transport in random media. Rev. Modern Phys. 64, 961–1043. http://dx.doi.org/10.1103/RevModPhys.64.961, URL https://link.aps.org/doi/10.1103/RevModPhys.64.961.

Lian, H., Chang, X., Hardalupas, Y., 2019. Time resolved measurements of droplet preferential concentration in homogeneous isotropic turbulence without mean flow. Phys. Fluids 31 (2), 025103.

Liu, Y., Shen, L., Zamansky, R., Coletti, F., 2020. Life and death of inertial particle clusters in turbulence. J. Fluid Mech. 902.

Maxey, M.R., 1987. The gravitational settling of aerosol particles in homogeneous turbulence and random flow fields. J. Fluid Mech. 174, 441–465.

Monchaux, R., 2012. Measuring concentration with Voronoï diagrams: the study of possible biases. New J. Phys. 14 (9), 095013.

Monchaux, R., Bourgoin, M., Cartellier, A., 2010. Preferential concentration of heavy particles: A Voronoï analysis. Phys. Fluids 22 (10), 103304.

Monchaux, R., Bourgoin, M., Cartellier, A., 2012. Analyzing preferential concentration and clustering of inertial particles in turbulence. Int. J. Multiph. Flow 40, 1–18.

Mora, D.O., Aliseda, A., Cartellier, A., Obligado, M., 2023. Characterizing inertial particle clustering with 1D signals. Meas. Sci. Technol. 35 (2), 025302. http://dx.doi.org/10.1088/1361-6501/ad031f.

Naumann, Z., Schiller, L., 1935. A drag coefficient correlation. Z. Ver. Dtsch. Ing. 77 (318), e323.

Obligado, M., Cartellier, A., Aliseda, A., Calmant, T., de Palma, N., 2020. Study on preferential concentration of inertial particles in homogeneous isotropic turbulence via big-data techniques. Phys. Rev. Fluids 5 (2), 024303.

Oujia, T., Matsuda, K., Schneider, K., 2020. Divergence and convergence of inertial particles in high-Reynolds-number turbulence. J. Fluid Mech. 905, A14.

Reade, W.C., Collins, L.R., 2000. Effect of preferential concentration on turbulent collision rates. Phys. Fluids 12 (10), 2530–2540.

Rycroft, C.H., 2009. VORO++: A three-dimensional voronoi cell library in C++. Chaos: Interdiscip. J. Nonlinear Sci. (ISSN: 1054-1500) 19 (4), 041111.

Sawford, B.L., Yeung, P., Borgas, M.S., 2005. Comparison of backwards and forwards relative dispersion in turbulence. Phys. Fluids 17 (9).

Schubert, E., Sander, J., Ester, M., Kriegel, J.-P., 2017. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Trans. Database Syst. 42 3.

Squires, K.D., Eaton, J.K., 1991. Preferential fselfconcentration of particles by turbulence. Phys. Fluids A: Fluid Dyn. 3 (5), 1169–1178.

Sumbekova, S., Cartellier, A., Aliseda, A., Bourgoin, M., 2017. Preferential concentration of inertial sub-Kolmogorov particles: the roles of mass loading of particles, Stokes numbers, and Reynolds numbers. Phys. Rev. Fluids 2 (2), 024302.

Tagawa, Y., Mercado, J.M., Prakash, V.N., Calzavarini, E., Sun, C., Lohse, D., 2012. Three-dimensional Lagrangian Voronoï analysis for clustering of particles and bubbles in turbulence. J. Fluid Mech. 693, 201–215.

Tom, J., Carbone, M., Bragg, A.D., 2022. How does two-way coupling modify particle settling and the role of multiscale preferential sweeping? J. Fluid Mech. 947, A7.

Uhlmann, M., Chouippe, A., 2017. Clustering and preferential concentration of finite-size particles in forced homogeneous-isotropic turbulence. J. Fluid Mech. 812, 991–1023. http://dx.doi.org/10.1017/jfm.2016.826.

Villafañe-Roca, L., Esmaily-Moghadam, M., Banko, A., Eaton, J., 2016. A robust method for quantification of preferential concentration from finite number of particles. In: Center for Turbulence Research Annual Research Briefs. pp. 123–135.

Xu, Y., Subramaniam, S., 2010. Effect of particle clusters on carrier flow turbulence: A direct numerical simulation study. Flow Turbul. Combust. 85 (3), 735–761.

Xu, D., Tian, Y., 2015. A comprehensive survey of clustering algorithms. Ann. Data Sci. (2015) 2, 165–193.

Yoshimoto, H., Goto, S., 2007. Self-similar clustering of inertial particles in homogeneous turbulence. J. Fluid Mech. 577, 275–286.