A robust method for quantification of preferential concentration from finite number of particles

By L. Villafañe-Roca, M. Esmaily-Moghadam, A. Banko AND J. K. Eaton

1. Motivation and objectives

The interaction of dispersed particles with a turbulent carrier flow is of interest in many engineering and environmental applications. The response of a particle to the turbulent carrier phase depends mainly on its inertia and its size relative to the time and length-scales of the flow. This dependence is represented by two parameters defined as the ratio of the particle aerodynamic time constant to a representative fluid time scale (Stokes number), and the ratio of the particle diameter to a representative fluid length-scale. The Kolmogorov time and length-scales are commonly chosen as the fluid scales. When the Stokes number is $O(1)$, particles can interact with the fluctuating velocity field of the carrier flow in such a way as to produce an in-homogeneous and intermittent distribution of particles in space. This phenomenon is referred to as preferential concentration, and it is traditionally explained by particles that are heavier than the fluid being centrifuged out of vortex cores and accumulating in regions of high strain rate (Squires & Eaton 1991; Wang & Maxey 1993). Although this mechanism is generally accepted, new theories have been proposed to describe the particle-turbulence interactions at different scales and for different values of non-dimensional parameters. Comprehensive reviews of different mechanisms for the clustering of inertial particles in turbulence can be found in recent papers by Bragg & Collins (2014) and Bragg *et al.* (2015).

Several methods exist for the quantification of preferential concentration of dispersed particles. An overview of methods commonly applied can be found in the review by Monchaux *et al.* (2012). The results from different studies are mainly compared in a qualitative sense by contrasting trends of a given index that indicates the strength of clustering. The diversity of numerical models and the complexity of experimental measurements in particle-laden flows, make quantitative comparison difficult. A quantitative metric is required to be robust against possible sources of biases, yet sufficiently sensitive to capture differences among cases. Two parameters play a major role in the quantitative evaluation of preferential concentration. These are the number of particles used to compute statistics, and the thickness of the integration volume when two-dimensional snapshots of particle positions are extracted from 2D projections of 3D particle coordinates, within a thin laser sheet or a computational domain. The variation of a metric for clustering with the number of particles does not necessarily need to be related to a physical change of the system. It is possible to define an index quantifying preferential concentration that has low sensitivity to the number of particles. However, the thickness of the integration volume has an irrevocable effect on the representation of the non-homogeneous particle distribution in a two-dimensional plane, and its influence on the quantitative assessment of preferential concentration is expected.

This report investigates the use of two-dimensional box-counting methods to generate particle concentration fields that can be used to compute statistics of preferential concentration. One of the most commonly used indexes derived from box-counting methods,

that quantifies the level of preferential concentration by comparison to a random distribution of particles, is severely affected by the number of particles used to compute statistics. This index was first proposed by Fessler *et al.* (1994), and it has found continued use in the literature. While it correctly captures the presence of particle clustering at length-scales of the order of the integral scale of the flow, and provides trends that compare well with other metrics (Aliseda *et al.* 2002), this index is sensitive to the number of particles in an image. We propose an unbiased index to capture preferential concentration that is minimally influenced by finite number of particles if sufficient samples of the spatial particle distributions are available. This index is also based on two-dimensional box-counting methods, but rather than directly computing statistics representative of the particle field from the finite number of particles in the snapshots, it uses that information to infer a property inherent to the flow that determines the level of preferential concentration. The proposed parameter enables direct comparison of the strength of preferential concentration in different systems provided that the integration volume is similar, and allows the influence of the thickness of the integration volume to be studied as an independent parameter. The new and the traditional indexes are evaluated using an experimental and numerical data base. Particle data corresponds to a fully developed turbulent duct flow of air, laden with nickel particles. The particle and fluid parameters used in the present study are such that preferential concentration of varying strength is observed in all the test cases.

The brief is organized as follows. Section 2 describes the experimental and numerical setups and the particle and flow parameters characterizing the problem. Section 3 discusses the effect of finite number of particles on the quantification of preferential concentration from two-dimensional box-counting methods, and demonstrates the capabilities of the new metric. Finally, conclusions are drawn in Section 4.

2. Experimental and numerical setups

Particle clustering is analyzed in a fully developed turbulent square duct flow of air for two Reynolds numbers based on the duct width and bulk velocities, $\mathrm{Re} = 10^4$ and 2×10^4. Both experimental and numerical data for the same range of flow and particle parameters are available. We are interested in preferential concentration within a thin rectangular volume aligned with the central plane of the test section parallel to the streamwise direction. Three-dimensional particle coordinates within that volume are projected onto a two-dimensional plane representative of the test section central plane. Only particles within the central 70% of the duct width (away from walls) are considered. In this region the mean particle concentration is constant.

2.1. *Duct flow experiments: Parameters and particle phase measurements*

The experimental setup involves a 5-meter development section with a 40 mm-wide square cross section, followed by a 0.5 m borosilicate glass test section of the same cross-sectional dimensions. The channel is vertical and the flow is downwards in the direction of gravity. Particles are injected in an upstream conditioning section. Particle dispersion is enhanced by several grids before the mixture of air and particles is accelerated through a contraction and enters the duct. The length of the development section is such that fully developed turbulence is achieved in the test section for the range of Reynolds numbers of interest, and that the particle positions and velocities at the test section are independent of their initial conditions.

Heavy nickel particles of 12 μm nominal diameter are used in the experiments. Particle

Re	Bulk duct Reynolds number	10000	20000
U_{cl}	Center-line gas velocity (m/s)	3.9	7.6
Re_τ	Duct averaged friction Reynolds number	328	594
η	Kolmogorov length-scale (μ m)	110	69
τ_k	Kolmogorov time-scale (ms)	0.8	0.3
d_p/η	average ratio particle diameter to Kolmogorov length-scale	0.11	0.17
$St_k(=\tau_p/\tau_k)$	average Stokes number	5	12

TABLE 1. Most relevant flow and particle parameters

	$Re \cdot 10^3$	Φ_m low case	Φ_m moderate case
Experiments	10	1.4	11.4
Experiments	20	2.4	11.9
Simulations	10	-	6
Simulations	20	-	8

TABLE 2. Mass loading ratios

size distributions are measured using a Coulter counter. While the measured probability distribution function (PDF) of particle sizes is relatively narrow, its tails cover a range from 7 to 20 μm. The inertia of the particles is characterized by the aerodynamic particle response time, τ_p, which is about 3.6 ms for the nominal diameter. The density of the nickel particles is $\rho_p = 8900$ kg/m^3. Particle Reynolds numbers are smaller than one. Average Kolmogorov time and length-scales are based on the average turbulent dissipation rate derived from the measured streamwise pressure gradient. By increasing the flow velocity, the characteristic size of the smallest eddies in the flow and their characteristic turnover time are reduced. Although the particle characteristics remain constant, both the ratios of particle diameter to Kolmogorov length-scale and Stokes number increase. For the flow conditions of interest, particles remain smaller than the Kolmogorov length-scale and average Stokes numbers are 5 and 12 for $Re = 10^4$ and 2×10^4, respectively. Table 1 gives a summary of the relevant parameters.

Tests were performed at two levels of loading. The mass loading ratio Φ_m, defined as the ratio of the mass flow of particles to the mass flow of air, was varied from about 2% to 11% for both Reynolds numbers (Table 2). Due to the high density of Nickel, the corresponding volume loadings (Φ_v, ratio of volume filled by the particles to that of the air) are lower than 10^{-6} and 10^{-5}, which are within the dilute regime.

Fields of instantaneous particle positions are measured by optical techniques based on light scattered by particles contained within a planar laser sheet. A pulsed Nd:YAG laser is used as light source. A combination of cylindrical lenses transforms the beam into a vertical laser sheet aligned with the central plane that crosses the full width of the test section. A synchronized high-resolution camera is placed perpendicular to the laser sheet and takes uncorrelated images of the particle phase. Each snapshot represents a single 5-ns pulse from the laser and visualizes the particles that are in the volume illuminated by the laser. The optical setup ensures a constant height and thickness of the laser sheet over the imaged region. This is essential to avoid spatial biases on the number of particles identified in different regions of the images due to variations in the

thickness of the illuminated volume and illumination intensity. The knife-edge technique is used to measure the laser sheet thickness at different positions. The thickness defined as the $1/e^2$ width of the Gaussian beam profile is about 1.4 mm. Two thousand images for each test case are acquired with a CCD camera (TSI PowerView Plus 4MP, 12bits, 2048×2048 pixels) with a spatial resolution of approximately 15 μm/px.

Raw images are cropped to a reduced region of interest of 2.9×2 cm (wall-normal and streamwise directions, respectively) to ensure spatial homogeneity of the mean particle concentration field. Image processing routines are used to reduce each digital image to a list of two-dimensional particle centroid coordinates. The processing procedure includes background subtraction, different steps of Gaussian and median filters, binary conversion, and area-based centroid identification for each individual particle. Although processing routines are optimized to the present images and particle number densities, statistics of particles at scales smaller than about 0.1-0.2 mm are subject to biases against very high particle concentrations due to particle overlapping. This is the case in any two-dimensional optical particle identification technique since images capture three-dimensional information projected as a two-dimensional image and particles at different depths can overlap in the image plane. In addition, diffraction of light scattered by the particles through the camera lenses is responsible for particles in the image plane appearing larger than just the size corresponding to the magnification factor from the object plane (Adrian & Westerweel 2011). This can cause particle images to overlap even if they do not in their projection on the object plane. Filters used at the image processing stage can reduce to some extent the overlapping caused by light diffraction. In the case of polydisperse particles, light diffraction makes it impossible to discriminate particles of different sizes from their images.

2.2. *Duct flow simulations*

The numerical test cases reproduce the experimental conditions. Point particle DNS simulations are performed by solving the incompressible Navier-Stokes equations in an Eulerian frame and the Lagrangian particle equation of motion using a Stokes drag force. Four-way coupling interactions are included. The back-reaction force from the particle phase is projected onto the fluid phase grid, and the hard-sphere collision model is used for the deterministic particle-particle and particle-wall interactions, assuming a coefficient of restitution of unity. The computational domain is a square duct that is 40 mm $\times$ 40 mm in cross section and 240 mm in length. A separate study demonstrated that periodic boundary conditions in the streamwise direction can capture the correct physics in terms of the establishment of the mean particle concentration distribution. Flow and particle parameters for the simulations are equivalent to those described for the experiments (Table 1). The data correspond to the same two Reynolds numbers and particle properties, and polydispersity is replicated by including particles of five different sizes sampled from a discrete version of the measured particle size PDF. The diameter of each class is computed from a geometric sequence such that the entire range of experimental distribution is covered. Numerical results are available only for one level of loading for each Reynolds number, and those differ slightly from the experimental data. Mass loading ratios are 6% and 8% for $\mathrm{Re} = 10^4$ and 2×10^4, respectively. The slight discrepancy between experimental and simulated conditions is due to the difficulty in imposing *a-priori* the mass flux of particles through the domain in the simulations. The computed mass flux depends on the coupling between gas and particle velocities, turbophoretic effects, and particle collisions.

The numerical results are post-processed to mimic the experimental data. We extract

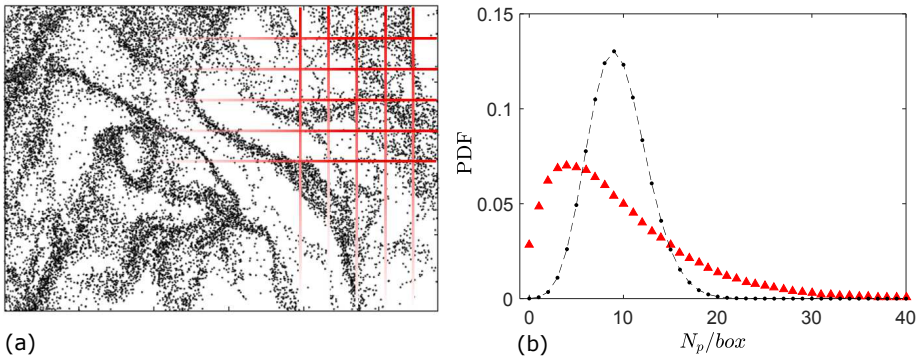

FIGURE 1. (a) Snapshot of particle centroid locations with a partial superimposed sketch of a square box grid. (b) PDF of number of particles per box of size 1.8 mm from multiple snapshots (triangles), and Poisson distribution with the same mean number of particles (dot-dashed line). Experimental data, $\mathrm{Re} = 10^4$, Φ_m=1.4%.

point particle coordinates within a volume of 2.9 cm × 2 cm × 1.4 mm. To improve the statistics, we take advantage of the symmetry of the test section (two symmetric perpendicular central planes), of the homogeneity of the mean particle concentration across the test section width, and of the homogeneity along the length of the numerical domain. This process provides, for a given time step, particle coordinates at several independent integration volumes equivalent in terms of particle distribution to the experimental data at the test section central plane.

3. Measurement of preferential concentration by box-counting methods

Box-counting is a method of collecting data for analyzing complex patterns by breaking a data set into smaller and smaller pieces (boxes), and analyzing the pieces at each smaller scale. The origin of the method is linked to fractal analysis, and its intent is to quantify fractal scaling. Box-counting methods are also traditionally used in the study of particle-laden flows to measure the level of preferential concentration of a system. Their relatively low computational cost makes them particularly attractive as the total number of particles increases. In addition, they are also used to generate local particle concentration fields from the Lagrangian information of the particle positions. Continuous representations of the particle density fields are useful to evaluate statistical descriptions of clusters and voids.

For the purpose of the present study we limit the scope of the paper to box-counting procedures to evaluate ensemble statistics derived from a large number (about 2500) of 2D maps of particle locations for a given test case. The box-counting method used here divides each snapshot into a regular grid of N square boxes of a certain size, as shown in Figure 1(a), and counts the number of particles inside each box through all the available snapshots. This information is used to compute the PDF of the number of particles per box for that particular box size. Triangles in Figure 1(b) represent the PDF for a box size length of 1.8 mm as obtained from the experiments at $\mathrm{Re} = 10^4$ and Φ_m=1.4%. The analysis is repeated for different box sizes, obtaining a distinct PDF for each box size.

The PDFs of number of particles per box are then used to quantify the magnitude of clustering at different scales. Since the ensemble statistics are obtained from discrete samples, point particles, they are sensitive to the size of the population. A metric com-

puted from those statistics could then be affected by the number of particles in the images. This complicates the quantitative comparison of results, making it difficult to evaluate whether a certain variation is due to finite particle number or the variation of some parameter of the flow. Furthermore, preferential concentration statistics may also be sensitive to the thickness of the integration volume. The thickness of the laser sheet in the experiments is a parameter with a relatively large uncertainty, so it is advisable to quote the uncertainty on preferential concentration introduced by the laser sheet thickness uncertainty. However, such uncertainty quantification can lead to erroneous results, because they are based on a metric that is sensitive to the number of particles, and a change in laser sheet thickness leads to a change in the total number of particles per image. In this section we describe a new index for quantification of preferential concentration that shows no dependence on the number of particles in each image if sufficient snapshots are available. The experimental and numerical data are used to analyze the sensitivity to finite particles and integration volume thickness of a traditional and new indexes for evaluating preferential concentration.

3.1. *Quantification of preferential concentration by comparison to a random particle distribution*

If particles are placed randomly in space, i.e. with equal probability at any location, then a box counting procedure will produce a Poisson distribution for the number of particles per box. For this reason, statistical measurements of preferential concentration are generally compared to the Poisson distribution of random particles that has the same average number of particles per box. The Poisson distribution is defined as

$$P_{poisson}(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \tag{3.1}$$

where, for a fixed box size, λ is the mean number of particles per box, and $P(n)$ is the probability of finding n particles in a box. The dashed line in Figure 1(b) represents the PDF that would have been obtained if the particles for the same test case represented by the triangular markers were randomly distributed. The comparison between the two PDFs provides an indication of how turbulence affects the particle concentration field. Fessler *et al.* (1994) introduced a parameter quantifying the strength of preferential concentration defined as the difference between the standard deviation of the measured distribution, σ_s, and that of the random distribution with the same mean, $\sigma_{poisson}$, divided by the mean λ_s (Eq. (3.2)). The subscript s indicates statistics computed from the sample population.

$$D = \frac{\sigma_s - \sigma_{poisson}}{\lambda_s}. \tag{3.2}$$

The parameter D can yield positive values, negative values, or zero. The extreme case of an equispaced distribution of particles, such that all boxes contain the same number, produces a Dirac's delta probability distribution yielding negative values of D. A random distribution without statistical error will yield D equal to zero. When preferential concentration is strong, boxes containing either no particles or a very large number are much more probable than they are for randomly distributed particles, because they accumulate in clusters and leave behind large depleted regions. In the later case large positive values of D are obtained. The value of D depends on the box size used. Large box sizes encompassing several structures will incorporate regions of both high and low concentrations and the corresponding PDF will appear nearly Poisson. The distribution of D for varying box sizes presents a peak. The box size at which D is maximized has been

associated in the literature with the scale at which preferential concentration is maximal, and the magnitude of D is used to compare the strength of clustering for different particle and flow parameters. Aliseda *et al.* (2002) compared the output of this parameter to that of a similar one introduced by Wang & Maxey (1993), defined in Eq. (3.3). Aliseda et al. showed that both parameters provided similar trends when computed from the same set of their experimental data.

$$D_2 = \sum_{n=1}^{N_p} (P(n) - P_{poisson}(n))^2. \tag{3.3}$$

Figure 2 shows the parameter D evaluated from experiments for box sizes l/H ranging from 0.01 to 0.25, with l being the box side length. In the left plot of Figure 2(a), circular and triangular markers correspond to Re $= 10^4$ and 2×10^4, respectively, and empty and filled markers represent the low and moderate loadings. For all cases, D reaches a maximum and decays as expected. We therefore conclude from the comparison that increasing Reynolds number and decreasing Φ_m decreases the strength of clustering. Also, we infer that preferential concentration is higher and is maximized at smaller scales as the particle mass loading is increased.

The above conclusions must be reexamined in light of the sensitivity of the parameter D to the number of particles in the images. The original data for Re $= 2 \times 10^4$ at Φ_m=12% contain about 12000 particles per image. We have randomly downsampled each of the snapshots of that test case to two lower mean number of particles, N_p, of 8100 and 1900, and recalculated D for varying box sizes. As shown in Figure 2(b) by the distributions of filled markers, the position of the maximum moves to lower box sizes and its value increases as more particles are used to compute D. The flow scales are not modified by randomly resampling each image; there are simply fewer particle samples. This dependence of D on N_p indicates that the position of the maximum is not necessarily related to a turbulence length-scale. Previous reports using box-counting methods to measure preferential concentration (Fessler *et al.* (1994), Aliseda *et al.* (2002), Wood *et al.* (2005)), all involved relatively low N_p, due in part to the limitations of the optical techniques used, so they can suffer from finite sampling errors.

The index D can be used to compare the strength of clustering of different data sets, provided they contain a similar average N_p per box. Data for two mass loadings resampled to the same N_p is shown in Figure 2(b). The clustering decreases as the mass loading is increased, the opposite of what is observed from Figure 2(a). The decrease of the intensity of preferential concentration with mass loading in the experiments has been corroborated by other indicators of preferential concentration including the radial distribution function (RDF), and a metric obtained from Voronoi diagram analysis (Monchaux *et al.* (2010)), for which the sensitivity to N_p has also been analyzed. Although not reported in here, the RDF has been found to be insensitive to N_p, whereas the metric derived from Voronoi analysis showed severe biases due to the number of particles.

Two-dimensional measurements of preferential concentration from three-dimensional particle fields depend on the thickness of the integration volume considered. This dependency can be easily evaluated using numerical data by considering the thickness as a variable parameter. In order to understand the effect of the integration volume thickness on the observed level of clustering, a measure of preferential concentration independent of the N_p should be used. Variation of D due to the three-dimensionality of the particle field is coupled to its dependence on N_p as the thickness is varied. On the other hand, resampling data from different thicknesses such that N_p per image stays the same changes the

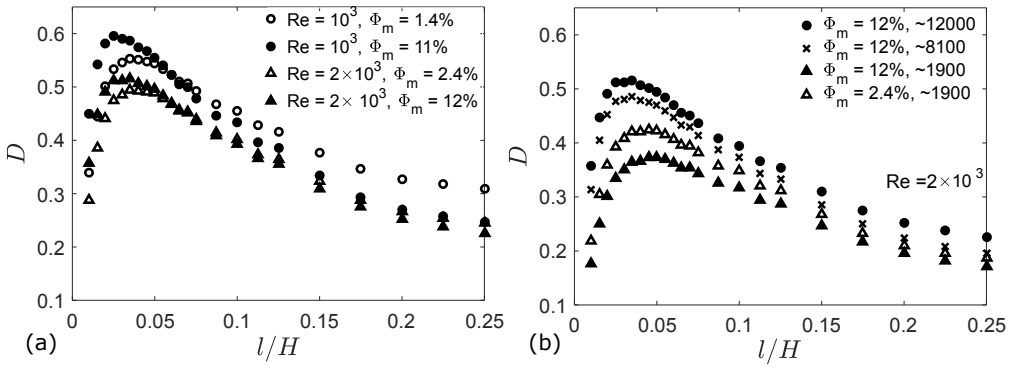

FIGURE 2. Variation of preferential concentration (D) for different box sizes: (a) different Reynolds numbers and mass loadings without resample, and (b) $\mathrm{Re} = 2 \times 10^4$ and two loadings, number of particles per snapshots downsampled.

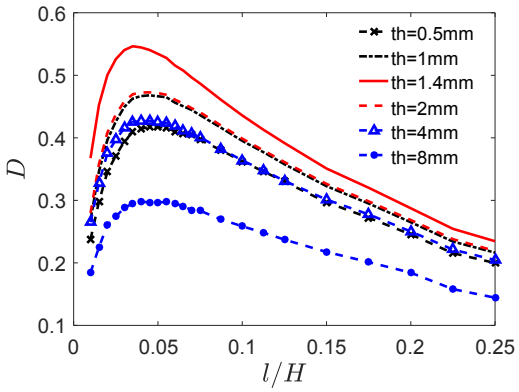

FIGURE 3. Variation of preferential concentration (D) with thickness.

degree of sampling of the 3D structures. Figure 3 shows D for varying box size obtained from the DNS data at $\mathrm{Re} = 10^4$ for five values of thicknesses, without resampling. An increase in thickness produces competing effects on the value of D. It changes both the averaging volume on which concentration is measured, and the total number of particles. For large thicknesses N_p increases. However the increase in thickness is also equivalent to an increase in box size: regions of clusters and voids are averaged and the distribution approaches a random distribution. For smaller values of thickness, the reduction of N_p dominates and results in a decrease of D due to insufficient numbers of particles. The observed trend is non-monotonic with the highest values of D for thicknesses near of 1.4 mm. Sensitivity of D to N_p obscures the physical variation of preferential concentration produced by other parameters.

3.2. *New method for quantification of preferential concentration*

Our objective in this section is to introduce a metric for quantification of preferential concentration that is minimally sensitive to the number of particles. To achieve this objective, we formulate an index in terms of a property of the system which is inherently linked to the underlying flow. The analysis is based on the idea that this inherent property of the flow is independent of a particular realization of particle positions, and hence is independent of the number of particles.

Consider the data sets that were discussed in the previous section. To simplify the derivation, we fix the box size in what follows. Regardless of the size of the box, our index will be insensitive to the number of particles. Hence, by imposing no limitation on the box size, we ensure that the new index remains generic.

All the data acquired from an experiment can be represented by a set of numbers $\mathcal{N} = \{n_1, \cdots, n_{N_t}\}$, where $N_t = N_b \times N_s$ is the total number of boxes in all the snapshots (i.e. the product of the number of boxes in one snapshot N_b multiplied by the number of snapshots N_s) and n_i for $i = 1, \cdots, N_t$ is the number of particles in one of the boxes in one of the snapshots. Therefore, all the data acquired from all the snapshots and all boxes during an experiment is contained in $\mathcal{N}$. From now on, we only rely on $\mathcal{N}$ to further develop our analysis. $\mathcal{N}$ depends on the underlying flow denoted by $\boldsymbol{u}(\boldsymbol{x}, t)$, where $\boldsymbol{u}$ is the Eulerian fluid velocity field which is a function of position $\boldsymbol{x}$ and time t. $\mathcal{N}$ also depends on the time sequence at which snapshots were acquired $\mathcal{T} = \{t_1, \cdots, t_{N_s}\}$, where N_s is the number of snapshots defined above. Additionally, $\mathcal{N}$ depends on the initial position of the particles at $t = 0$ denoted by $\mathcal{X} = \{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_{N_p}\}$, where N_p is the total number of particles. Based on these definitions, given $\boldsymbol{u}$, $\mathcal{T}$, and $\mathcal{X}$, one can determine $\mathcal{N}$ since the dynamics are deterministic.

Next, consider a hypothetical scenario in which we are able to make several realizations of $\mathcal{N}$ by only varying $\mathcal{X}$. Thus, in all these realizations $\boldsymbol{u}$ and $\mathcal{T}$ remain unchanged. To make a physical interpretation of this hypothetical scenario, one may think of a one-way coupled simulation where only the initial position of the particles is changed and the remaining parameters are kept fixed, including the duration of the simulation, the initial condition of the flow, and the time at which snapshots are acquired. We denote the realizations of $\mathcal{N}$ obtained in this manner by $\mathcal{M}^1, \cdots, \mathcal{M}^{N_r}$. The number of realizations N_r is taken to be very large to ensure statistical convergence. Note that $|\mathcal{M}^i| = |\mathcal{N}| = N_t$, where $|\bullet|$ denote the dimension of the set. All those realization also contain N_t entries, namely $m_i^j \in \mathcal{M}^j$, number of particles in box i (here box referes to any box in any snapshot as $i = 1, \cdots, N_t$) from realization j. This hypothetical exercise, although unreproducible in reality due to the chaotic nature of turbulence and four-way coupling effects, is discussed to clarify our derivation.

Next, we consider the number of particles in a particular box i and build a distribution from its different realizations. This will produce a set $\mathcal{P}_i = \{m_i^1, \cdots, m_i^{N_r}\}$. The distribution of $\mathcal{P}_i$ can be inferred from its construction and is Poisson: 1) members of $\mathcal{P}_i$ are determined based on the probability of having a particle in a box with a fixed average rate, and 2) members of $\mathcal{P}_i$ are obtained from independent realizations. Since the mean and variance of a Poisson distribution are identical

$$\langle (m_i - \langle m_i \rangle)^2 \rangle = \langle m_i \rangle, \quad i = 1, \cdots, N_t. \tag{3.4}$$

In Eq. (3.4), $\langle \bullet \rangle$ denotes ensemble-averaging over different realizations $\langle m_i \rangle = N_r^{-1} \sum_{j=1}^{N_r} m_i^j$.

Since different realizations were based on different $\mathcal{X}$, $\langle m_i \rangle$ depends only on the underlying flow, i.e. $\boldsymbol{u}$ and $\mathcal{T}$, and does not depend on the intial position of the particles $\mathcal{X}$. We take advantages of this independence to construct a metric that is robust to variations of the number of particles. More specifically, we compute the coefficient of variance (standard deviation divided by the mean) of the set $\langle \mathcal{M} \rangle = \{\langle m_1 \rangle, \cdots, \langle m_{N_t} \rangle\}$,

$$D_c = \frac{\sqrt{\mathbb{E}[(\langle m \rangle - \mathbb{E}[\langle m \rangle])^2]}}{\mathbb{E}[\langle m \rangle]}, \tag{3.5}$$

as our metric, where $\mathbb{E}[\bullet]$ is the mean computed from all boxes $\mathbb{E}[m] = N_t^{-1} \sum_{i=1}^{N_t} m_i$.

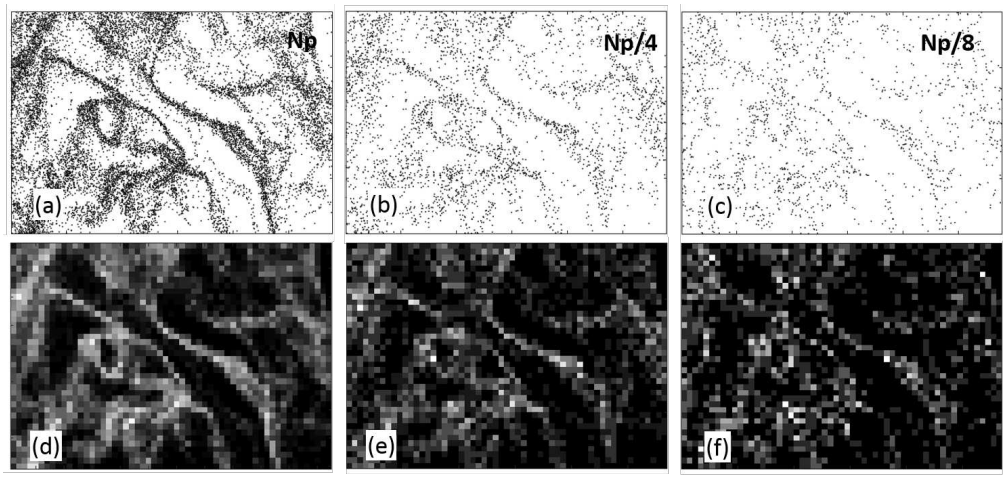

FIGURE 4. Snapshot of particle positions: (a) original N_p , (b) randomly downsampled to $N_p/4$ and (c) $N_p/8$. Particle density field corresponding to resampled images for a box size of 1 mm (d-f).

The end result is an estimation of D_c from the real experimental measurements, i.e. $\mathcal{N}$. Although $\langle \mathcal{M} \rangle$ is not known, it can be estimated from $\mathcal{N}$. The total number of particles and boxes are fixed among all realizations and is the same as that of experiment, hence $\mathbb{E}[m^1] = \cdots = \mathbb{E}[m^{N_r}] = \mathbb{E}[\langle m \rangle] = \mathbb{E}[n] = \lambda_s$. If $\mathcal{N}$ is constructed based on sufficient statistical samples, another realization of the same experiment would produce the same statistics. Since $\mathcal{M}^j$ is a realization of the experiment, thus variance computed based on $\mathcal{M}^j$ should be equal to that of $\mathcal{N}$ within the statistical margin of error. Therefore, we use $\sigma_s^2 = \mathbb{E}[(n - \mathbb{E}[n])^2]$ to estimate

$$\langle \mathbb{E}[(m - \mathbb{E}[m])^2] \rangle \approx \sigma_s^2. \tag{3.6}$$

Combining Eqs. (3.4)–(3.6), and noting that $\langle \bullet \rangle$ and $\mathbb{E}[\bullet]$ operators are commutable, one can show

$$D_c \approx \left(\frac{\sigma}{\lambda}\right)_c = \sqrt{\frac{\sigma_s^2 - \lambda_s}{\lambda_s^2}}, \tag{3.7}$$

where $(\sigma/\lambda)_c$ is a corrected coefficient of variance that depends only on σ_s and λ_s, and therefore, can be computed directly from the experimental results $\mathcal{N}$. Equation (3.7) is the main result of this analysis.

To test the proposed metric, we consider a snapshot and perform downsampling. This is reflected in Figure 4, where the total number of particles identified in a certain snapshot from the experiments (a) has been randomly downsampled to reduce N_p by factors of 4 (b) and 8 (c). Figure 4(d-f) shows the corresponding 2D fields of number of particles per box for a box size of 1 mm. Regions of high and low N_p/box create a pattern that becomes blurred as N_p is reduced. However one can still identify the pattern.

The effectiveness of the proposed formulation in correcting for the finite number of particles is demonstrated in Table 3. The coefficients of variance with and without correction, as well as the conventional index D, are tabulated. The box size is 1 mm, $\mathrm{Re} = 10^4$, and $\Phi_m = 12\%$, in this case. The coefficient of variance without correction σ_s/λ_s and the conventional index D vary significantly as the number of particles is changed, whereas the

	N_p	$N_p/2$	$N_p/4$	$N_p/8$
$(\sigma_s - \sigma_p)/\lambda_s$	0.50	0.42	0.32	0.26
σ_s/λ_s	0.91	1.01	1.15	1.43
$(\sigma/\lambda)_c$	0.81	0.82	0.80	0.82

TABLE 3. Parameter D, sample and corrected coefficients of variance. $\mathrm{Re} = 10^4$, $\Phi_m = 12\%$.

new index $(\sigma/\lambda)_c$ produces a similar value at all N_p considered. Therefore, the corrected coefficient of variance provides a converged measure of clustering regardless of the number of particles in the image if sufficient statistical samples (e.g. snapshots) are available. As N_p/box decreases (i.e. lower N_p or smaller box size), the number of required samples for statistical convergence N_t increases. Note that for an infinite number of particles, the corrected coefficient of variance reduces to the coefficient of variance without correction (Eq. 3.8).

$$\lim_{\lambda_s \to \infty} \sqrt{\frac{\sigma_s^2 - \lambda_s}{\lambda_s^2}} = \frac{\sigma_s}{\lambda_s}. \tag{3.8}$$

The conventional index D from Eq. (3.2) can be expressed as the difference between the coefficients of variance of the sample and that of the Poisson distribution with the same mean. In the limit of infinite number of particles, the coefficient of variance of a Poisson distribution is zero, and the index D converges to the same value as the corrected coefficient of variance (Eq. 3.9).

$$\lim_{\lambda_s \to \infty} \frac{\sigma_s}{\lambda_s} - \frac{\sigma_p}{\lambda_p} = \lim_{\lambda_s \to \infty} \frac{\sigma_s}{\lambda_s} - \frac{1}{\sqrt{\lambda_p}} = \frac{\sigma_s}{\lambda_s}. \tag{3.9}$$

For a finite number of particles, the subtraction of the Poisson coefficient of variance can be interpreted as a correction intending to remove the random component of the distribution. However, this correction is too strong when the number of particles are finite. As a result, D underpredicts preferential concentration due to the insufficient number of particles.

The corrected coefficient of variance is used in Figure 5(a) to compare the strength of preferential concentration at different scales for the same cases as in Figure 2(a). In all cases, preferential concentration decreases as the box size is increased. As argued before, for large box sizes regions of high and low concentration are encompassed in each box, and the standard deviation of the number of particles per box decreases. However, there is no local maximum. Higher resolution, thinner laser sheet thickness, and smaller ratio of particle diameter to Kolmogorov length-scale are required to analyze preferential concentration at smaller length-scales. Preferential concentration decreases as Reynolds number and mass loading are increased. These trends are in agreement with those obtained from D when evaluated on snapshots with a similar number of particles. Note that the new index did not require resampling of the number of particles per image in order to identify the correct trends. The insensitivity of the corrected coefficient of variance to N_p is reflected in Figure 5(b) for $\mathrm{Re} = 2 \times 10^4$.

The new parameter is well suited to analyze the effect of integration volume thickness on the two-dimensional evaluation of preferential concentration. This effect is shown in Figure 6 where numerical results for $\mathrm{Re} = 10^4$ have been analyzed for thickness ranging

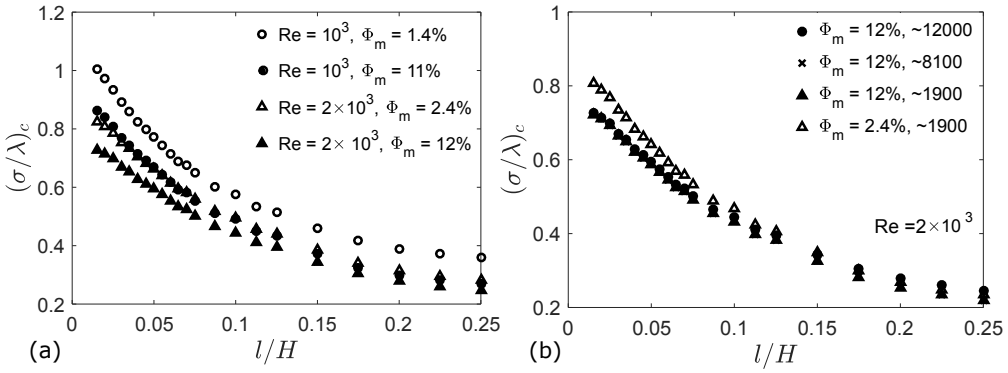

FIGURE 5. Distribution of corrected coefficient of variance for different box sizes: (a) for Re $= 10^4$ and 2×10^4, and both low and moderate loadings without resample, and (b) for Re $= 2 \times 10^4$ and both loadings with downsampling.

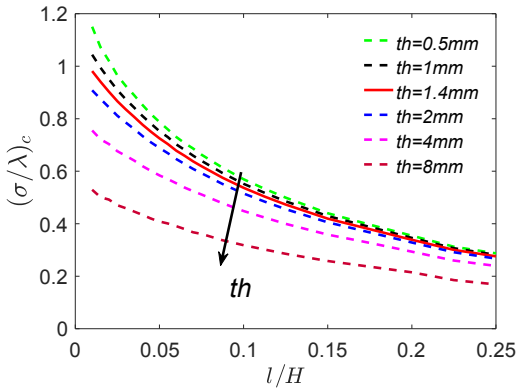

FIGURE 6. Variation of preferential concentration (corrected CV) with integration volume thickness.

from 0.5 mm to 8 mm. The total number of samples available is insufficient for thicknesses smaller than 0.5 mm to reach convergence of the corrected coefficient of variance, and results are not reported. Based on the new index, preferential concentration decreases monotonically with increasing thickness. This trend is expected due to the filtering effect in the direction normal to the image plane as the thickness is increased. This monotonic physical trend is captured by the corrected coefficient of variance, while the conventional parameter D displayed unphysical, non-monotonic behavior due to its sensitivity to the number of particles.

4. Conclusions

A new index was introduced for quantification of preferential concentration of inertial particles in turbulent flows. Using numerical and experimental data, we showed that the conventional index used to evaluate the strength of preferential concentration is significantly sensitive to the number of particles. It can also lead to wrong conclusions regarding the scale at which preferential concentration is maximal. The decay of the index at small scales is in fact an artifact due to insufficient number of particles. The

use of the conventional index for case-to-case comparison studies requires resampling the data to the same number of particles such that all cases are affected by similar biases.

To remedy this problem, we presented an statistical argument to derive a new index that is less sensitive to the number of particles, and showed that it only depends on the underlying flow. The index, which theoretically corresponds to the coefficient of variance of the probability distribution of particle locations, is estimated from a corrected coefficient of variance of the observed concentration field. The index was tested using the experimental and numerical data base, and showed minimal dependence on the number of particles. Our index provides a platform for quantitative assessment of the preferential concentration that is well suited for validation of numerical models and comparison of different data sets.

Acknowledgments

This investigation is funded by the Department of Energy's National Nuclear Security Administration (PSAAP-II-Program), Grant #DE-NA0002373-1.

REFERENCES

Adrian, R. J. & Westerweel, J. 2011 *Particle image velocimetry*. Cambridge University Press.

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.

Wang, L.-P. & Lasheras, M. R. 1993 Settling velocity and concentration distribution of heavy particles in homogeneous isotropic turbulence. *J. Fluid Mech.* **256**, 27–68.

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

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**, 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**, 023029.

Fessler, J. R., Kulick, J. D. & Eaton, J. K. 1994 Preferential concentration of heavy particles in a turbulent channel flow. *Phys. Fluids* **6**, 3742–3749.

Monchaux, R., Bourgoin, M. & Cartellier, A. 2010 Preferential concentration of heavy particles: a Voronoi analysis. *Phys. Fluids (1994-present)* **22**, 103304.

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

Squires, K. D. & Eaton, J. K. 1991 Preferential concentration of particles by turbulence. *Phys. Fluids* **3**, 1169–1178.

Sundaram, S. & Collins, L. R. 1997 Collision statistics in an isotropic particle-laden turbulent suspension. Part I. Direct numerical simulations. *J. Fluid Mech.* **335**, 75–109.

Wood, A., Hwang, W. & Eaton, J. 2005 Preferential concentration of particles in homogeneous and isotropic turbulence. *Int. J. Multiphase Flow* **31**, 1220–1230.