Capturing mesoscale structures in multiphase CFD simulations

Swathi Ganesh[†], Balivada Kusum Kumar[†], and Himanshu Goyal[*]

Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

E-mail: goyal@iitm.ac.in

Phone: +91 44 2257 4183

Abstract

Multiphase reactors' performance depends on the mesoscale structures formed due to multiphase hydrodynamics. Examples of mesoscale structures include gas bubbles in a fluidized bed and particle clusters in a riser. Experimental investigation of these mesoscale structures is challenging and expensive. To this end, Computational Fluid Dynamics (CFD) simulations are extensively employed; however, post-processing CFD data to capture mesoscale structures is challenging. This work develops a DBSCAN-based methodology to capture and characterize mesoscale structures from multiphase CFD simulation data. DBSCAN is an unsupervised machine-learning algorithm, which requires the value of two hyperparameters. A simple technique to calculate these hyperparameters is provided and the performance of DBSCAN is assessed on CFD-DEM simulations of bubbling fluidized beds and particle clustering. We demonstrate the computational complexity of DBSCAN to be $\mathcal{O}(n \log n)$, lower than the existing techniques, by testing its scalability on highly resolved grids (up to 100 million grid points).

[†]These authors contributed equally to this work

Keywords: DBSCAN, machine learning, fluidized bed, riser, multiphase flow, CFD

1 Introduction

Multiphase reactors are ubiquitous in the chemical industry and play an essential role in producing chemicals and energy conversion.[1,2] Bubbling fluidized beds and risers are the two commonly used multiphase reactors employing gas and solid particles. These reactors are characterized by mesoscale structures that are much larger than individual particles but much smaller than the reactor size. For example, gas bubbles in a fluidized bed and particle clusters in a riser[3] are two important mesoscale structures that significantly impact the reactor performance. In general, clusters reduce the mixing between gas and solid particles, whereas bubbles lead to gas bypass, leading to poor reactant conversion. Hence, quantitative characterization of mesoscale structures is essential to predict and optimize the performance of multiphase reactors.

Experimental investigations focused on characterizing bubble and cluster properties are challenging. The opaque nature of the solid phase makes visual investigations difficult. Bubble characteristics in thin rectangular beds are obtained using high-speed cameras and digital image analysis. However, these techniques fail in commonly used cylindrical beds. For such configurations, intrusive techniques based on optical signals, electrical pulses, and pressure fluctuations and non-intrusive techniques based on electrical capacitance, X-rays, and MRI are employed.[4] Pixel luminance of the images obtained from the non-intrusive measurements is processed through image processing software, such as MATLAB, to obtain bubble and cluster statistics.[5–7] However, these techniques are expensive and challenging for large-scale reactors and at high temperatures.

Increasing computational resources and the availability of commercial and open-source software have made Computational Fluid Dynamics (CFD) simulations an integral part of multiphase reactor investigations.[3,8–16] High performance Computing is pushing the bound-

aries of the level of detail and the reactor size simulated using CFD. Recently, CFD simulation of a fluidized bed using one billion mesh elements on 36,000 processors was performed.[17] This simulation generated 200 TB of data for 25 s of physical time, making the post-processing of the data a significant challenge. A few algorithms have recently been developed for bubble identification.[4,7,18,19] These algorithms, primarily based on image analysis, include flood-fill method,[7] maximum entropy,[18] MS3DATA,[4] and use of Visualization Toolkit (VTK).[19] Among these techniques, the time complexity of the flood-fill method and MS3DATA is known to be $\mathcal{O}(n^2)$ and $\mathcal{O}(n^{1.3})$, respectively. Computationally fast algorithms to capture mesoscale structures that are easy to implement with multiphase CFD simulations are imperative.

Machine learning algorithms are finding applications in expediting multiphase reactor simulations.[20–24] In the context of clustering data or finding patterns, unsupervised machine learning algorithms can be helpful. These algorithms create groups in data based on the similarity of data points to each other relative to the surrounding data points. Clustering algorithms can be broadly divided into hierarchical, centroid-based, density-based, and distribution-based clustering.[25] Centroid and density-based algorithms such as K-means,[26] Mean-shift,[27] OPTICS,[28] and DBSCAN[29] primarily rely on distance between points, making them suitable candidates to process mesh- or grid-based CFD simulation data.

K-means is the most popular clustering algorithm and has been used for image analysis to identify clusters in 2D riser simulations.[30–33] However, its applicability is constrained by the need for convex cluster shapes, sensitivity to initial conditions and outliers, and *a priori* knowledge of the number of clusters. These limitations of K-means can be eliminated using Density-Based Spatial Clustering of Applications with Noise[34,35] (DBSCAN), which is a density-based deterministic clustering algorithm. Unlike partition-based K-means and other hierarchical clustering methods, DBSCAN operates on the notion of density and requires minimal domain knowledge. It is highly effective in discovering arbitrary-shaped distributions in large spatial databases with outliers. These qualities make DBSCAN an

optimal choice for identifying and characterizing mesoscale structures in multiphase CFD simulations. Moreover, DBSCAN[34,35] is freely available as an open-source library in R,[36] .NET,[37] and Python,[38] allowing it to be used in a modular fashion with existing CFD codes and software.

In this work, we employ DBSCAN to identify bubbles and clusters of arbitrary shapes and sizes in the fluidized bed and riser CFD simulations, respectively. A simple approach to calculating optimal values of DBSCAN's hyperparameters is also provided. This step eliminates the time-consuming tuning process and makes DBSCAN fully automated. The adequacy of DBSCAN is assessed against test data consisting of analytical shapes and CFD-DEM simulations of a fluidized bed and particle clustering in a periodic domain. These simulations are performed using NGA,[3,39,40] a low Mach, variable density multiphase code employing the Computational Fluid Dynamics – Discrete Element Method (CFD-DEM) strategy. The computational complexity of DBSCAN is found to be close to its theoretical value of $\mathcal{O}(n \log n)$ by performing scalability analysis on highly resolved grids containing up to 100 million grid points.

2 Methodology

This section describes the DBSCAN algorithm and the key steps in its implementation. Section 2.1 describes the simulation setup NGA used to perform CFD-DEM simulations of fluidized beds and particle clustering. Section 2.2 summarizes the working of DBSCAN. A technique to calculate the optimal values of DBSCAN's hyperparameters is demonstrated in section 2.3. Pre-processing of data to implement DBSCAN is detailed in section 2.4. Finally, section 2.5 summarizes the bubble and cluster properties calculation.

2.1 Simulation setup

NGA,[39,40] a reacting multiphase flow solver, is used to perform bubbling fluidized bed and periodic riser simulations. NGA has been extensively used to investigate several multiphase flow problems, such as fluidized beds,[3,8,40] particle clustering,[41,42] turbulent flows,[43] and gas-liquid flows.[44] NGA implements a CFD-DEM approach[40] to simulate gas-solid flows, where the gas phase is modeled as a continuum and the solid particles are tracked individually. The particle tracking is based on Newton's second law of motion where particles experience forces due to gravity, drag, and collisions. Four-way coupling is considered, including the influence of particles on the gas phase and particle-particle and particle-wall collisions. The collisions are represented using the soft sphere modeling approach of Cundall and Strack.[45] A conservative immersed boundary (IB) method[46] is employed to account for the cylindrical geometry on a Cartesian mesh. More details about NGA can be found in refs.[3,39,40]

2.2 Mesoscale structure detection using DBSCAN

DBSCAN employs minimum density threshold criterion to separate regions of different densities. Each point in a spatial database $\mathcal{D}$ either belongs to a $\texttt{cluster}_D$ or is noise (outlier). Note that $\texttt{cluster}_D$ refers to a group of similar data points identified by DBSCAN and is different from a particle cluster, which is a mesoscale structure. To classify the points as a $\texttt{cluster}_D$, the neighborhood of a given radius must contain a minimum number of points, i.e., the density in the neighborhood must exceed some specified threshold. Noise points are a set of points in $\mathcal{D}$ that do not belong to any of its $\texttt{cluster}_D$.

Two internal parameters in DBSCAN need to be specified: *eps* and $minPts$. Based on these two parameters, every point in $\mathcal{D}$ is labeled as a core point C_i, border point B_i, or noise point O_i (outlier). Figure 1 shows a schematic of the key concepts of DBSCAN. DBSCAN follows a two-step approach given the parameters *eps* and $minPts$. First, it selects a point P from $\mathcal{D}$ and considers all neighbors within *eps* radius to be the members of the same cluster as P (these points are referred to as direct density-reachable). If the *eps* radius consists

of at least $minPts$ data points, then P is labeled as a core point. All points within the same $\texttt{cluster}_D$ set are density-connected, and non-core points are termed border points. Finally, points that are not density-reachable from any core points of any $\texttt{cluster}_D$ are called outliers.[34]

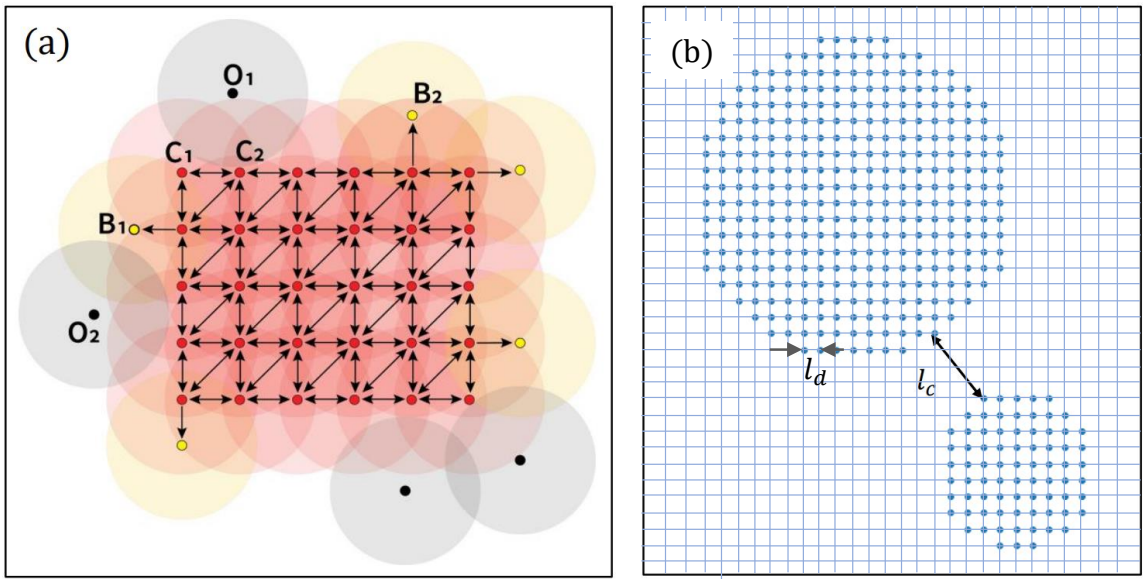

Figure 1: (a) Illustration of the principle behind DBSCAN using $minPts = 4$ as an example. Every core point C_i has at least four points ($minPts = 4$) within a circle (red region) of eps radius, including C_i itself. All density connected points B_i are the border points inside yellow circles, as there are less than four points in its neighborhood of eps radius. These are not core points but are density reachable from C_i and belong to the $\texttt{cluster}_D$. Points O_i are outliers or noise within a circle (grey region) as they are neither core points nor are density reachable from core points. (b) Schematic of two $\texttt{clusters}_D$ formed on a 2D grid of spacing l_d; the shortest distance between the $\texttt{clusters}_D$ is l_c.

2.3 Estimation of DBSCAN parameters

DBSCAN algorithm requires the values of two parameters: eps and $minPts$. A few general methods[29,34,35] exist in the literature providing broad estimates of eps and $minPts$. We demonstrate a straightforward technique to estimate the adequate values of these parameters for 2D and 3D CFD simulations.

Consider two $\texttt{clusters}_D$ on a uniform grid with grid spacing or mesh size being l_d, and the shortest distance between the $\texttt{clusters}_D$ be l_c (see Figure 1b). A lower bound for

eps could be the maximum distance between two adjacent points, equal to $\sqrt{2}l_d$ for a 2D square grid and $\sqrt{3}l_d$ for a 3D cubic grid. These values correspond to the distance between the diagonally opposite points in a square and a cube, respectively. The upper bound of *eps* should be lower than l_c to identify two nearby $\texttt{clusters}_D$ as separate entities. Thus, an adequate value of *eps* should lie between $\sqrt{2}l_d$ or $\sqrt{3}l_d$ and l_c. Since l_c varies significantly in CFD simulations due to the dynamic nature of bubbles and clusters formation and disappearance, *eps* based on the lower bound is more appropriate. Based on these observations, we evaluate *eps* as

$$eps = \begin{cases} \sqrt{2} f l_d & \text{for 2D} \\ \sqrt{3} f l_d & \text{for 3D} \end{cases} \tag{1}$$

where $f > 1$.

To determine an optimal value of f, a parametric study consisting of two $\texttt{clusters}_D$ (circles for the 2D database and spheres for the 3D database) with radii of 2 cm and 1.5 cm is conducted. f is varied between 1 and 2. The minimum gap l_c between the $\texttt{clusters}_D$ is varied, such that $l_d < l_c < 10l_d$. All sides of the computational domain are set to 10.2 cm and discretized using different grid resolutions (grid spacing from 0.5 mm to 2 mm). DBSCAN correctly captures the $\texttt{clusters}_D$ when $1 < f < 1.3$. We use $f = 1.15$ in all the cases in this work.

Once *eps* is calculated, *minPts* is estimated as the number of data points in a circle (sphere for 3D) of radius *eps*:

$$minPts = \begin{cases} \rho_N \cdot eps^2 & \text{for 2D} \\ \rho_N \cdot eps^3 & \text{for 3D} \end{cases} \tag{2}$$

where ρ_N is the number density of the dataset. In CFD simulations, $\rho_N \sim 1/l_d^2$ for 2D and $1/l_d^3$ for 3D dataset is the ratio of the number of grid points and the area (volume for 3D) of

the computational domain.

2.4 Data pre-processing

Fluid volume fraction ε_f or particle volume fraction, $\varepsilon_p = 1 - \varepsilon_f$, data obtained from CFD-DEM simulations is used to identify bubbles and clusters. A threshold or cut-off value is defined for the fluid volume fraction ε_f^* and particle volume fraction ε_p^* for bubble and cluster detection, respectively. The threshold values are such that in a bubble $\varepsilon_f > \varepsilon_f^*$, whereas in a cluster $\varepsilon_p > \varepsilon_p^*$. In the literature, ε_f^* is estimated to be in the range of 0.7 to 0.85.[4,7,19,47] We use an intermediate value of $\varepsilon_f^* = 0.75$ for bubble detection. The following expression[48] is used to estimate ε_p^*:

$$\varepsilon_p^* = \mu(\varepsilon_p) + n\sigma(\varepsilon_p) \tag{3}$$

where $\mu(\varepsilon_p)$ and $\sigma(\varepsilon_p)$ are the mean and standard deviation of ε_p, respectively. n is an integer between 1 and 3.[48–50] An intermediate value $n = 2$ is used in this work.

At every grid point, ε_f or ε_p value is replaced by δ, such that

$$\delta = \begin{cases} 1 & \text{if } \varepsilon_f > \varepsilon_f^* \text{ for bubble detection} \\ 1 & \text{if } \varepsilon_p > \varepsilon_p^* \text{ for cluster detection} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Positions of the grid points with $\delta = 1$ are considered for DBSCAN, discarding the remaining data. This step leads to a significant reduction in the memory and computing requirements as the number of grid points that are part of bubbles or clusters is much smaller than the total grid points in multiphase CFD simulation. Section 6 discusses the impact of this pre-processing step on the calculation time in more detail.

2.5 Calculation of bubble and cluster properties

After DBSCAN identifies $\texttt{clusters}_D$, four $\texttt{cluster}_D$ properties are calculated: centroid, area A_b (volume V_b in 3D), equivalent diameter d_b, and chord length CL. DBSCAN assigns unique labels to the core and border points of identified $\texttt{clusters}_D$. The centroid of a $\texttt{cluster}_D$ is calculated as

$$(\bar{x}, \bar{y}) = \left(\frac{1}{n_c} \sum_{i=1}^{n_c} x_i, \frac{1}{n_c} \sum_{i=1}^{n_c} y_i \right) \text{for 2D} \tag{5}$$

$$(\bar{x}, \bar{y}, \bar{z}) = \left(\frac{1}{n_c} \sum_{i=1}^{n_c} x_i, \frac{1}{n_c} \sum_{i=1}^{n_c} y_i, \frac{1}{n_c} \sum_{i=1}^{n_c} z_i \right) \text{for 3D} \tag{6}$$

where x_i, y_i, and z_i are the core points, and n_c is the number of core points in a $\texttt{cluster}_D$. Area-equivalent (volume-equivalent) diameter d_b is calculated as the diameter of a circle (sphere) having the same area (volume) as the $\texttt{cluster}_D$:

$$d_b = \sqrt{\frac{4A_b}{\pi}} \text{ for 2D} \tag{7}$$

$$d_b = \sqrt[3]{\frac{6V_b}{\pi}} \text{ for 3D} \tag{8}$$

Here $A_b = n_c l_d^2$ and $V_b = n_c l_d^3$.

The chord length CL is defined as the maximum length of a `cluster` and calculated using the property of Convex Hull computed using the Qhull library.[51] The algorithm identifies a convex enclosure and returns the unique and minimal convex set of points containing all the points of a data group. The largest distance between any two points in the Convex Hull set is considered to be CL.

3 Assessment of the methodology

The accuracy of DBSCAN and property estimation is assessed for 2D and 3D test datasets consisting of circles, ellipses, spheres, and ellipsoids. These shapes of different sizes with

varying l_d and l_c are considered for a robust assessment case. The computational domain having a side length of 10.2 cm is discretized using different grid resolutions $l_d \in (0.5, 0.8, 1.4)$ mm. The hyperparameters of DBSCAN (*eps* and *minPts*) are calculated using the methodology described in section 2.3. DBSCAN accurately identifies the $\texttt{clusters}_D$ for all the test cases. Figure 2 shows the $\texttt{cluster}_D$ identified in a few 2D (first row) and 3D (second row) datasets by DBSCAN.

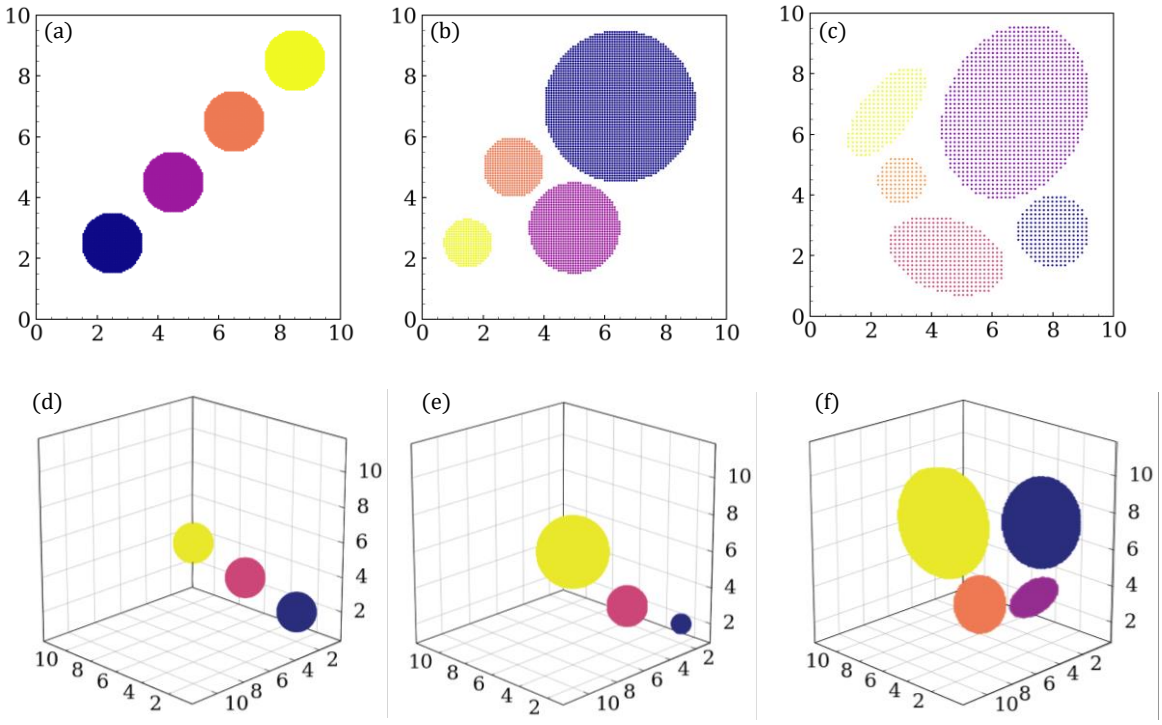

Figure 2: Application of DBSCAN for 2D and 3D analytical shapes of various sizes in a domain of side length 10.2 cm. **First row: 2D dataset** with (a) l_d=0.05 cm, l_c=0.83 cm, *eps*=0.08 cm, *minPts*=3 (b) l_d=0.08 cm, l_c=0.27 cm, *eps*=0.13 cm, *minPts*=3, and (c) l_d=0.14 cm, l_c=0.40 cm, *eps*=0.23 cm, *minPts*=3. **Second row: 3D dataset** with (d) l_d=0.05 cm, l_c=1.43 cm, *eps*=0.10 cm, *minPts*=8 (e) l_d=0.08 cm, l_c=1.28 cm, *eps*=0.16 cm, *minPts*=8, and (f) l_d=0.14 cm, l_c=0.52 cm, *eps*=0.28 cm, *minPts*=8. For all the cases $f = 1.15$. Note that the shortest distance between two shapes is reported as l_c.

$\texttt{Cluster}_D$ properties: centroid, chord length, area, and volume, are calculated as described in section 2.5 and compared with their analytical values. The corresponding maximum relative percentage error is provided in Table 1. The error in calculating 3D $\texttt{cluster}_D$ properties is lower than that for 2D $\texttt{clusters}_D$ due to the larger number of points per clus-

ter available for analysis. Moreover, the error primarily arises due to the discrete nature of the data and is maximum for the lowest grid resolution case. DBSCAN is tested for more complex shapes in section S1 of the Supporting Information.

Table 1: Error associated with the calculation of $\texttt{cluster}_D$ properties using DBSCAN for test datasets.

Property	Maximum % Error in 2D case	Maximum % Error in 3D case
Centroid	0.3	0.2
Area or Volume	2.9	0.7
Chord Length	2.8	0.9

In a multiphase flow simulation, large variations could exist in the size of bubbles or clusters. Applicability of DBSCAN is tested for varying relative sizes of $\texttt{cluster}_D$. Two $\texttt{cluster}_D$ in 2D and 3D are considered and the size of one $\texttt{cluster}_D$ is varied to have a size ratio between 1 and 100. This size ratio covers a wide variation in mesoscale structure sizes in multiphase CFD simulations. DBSCAN accurately identified the two $\texttt{clusters}_D$ for all the size ratios tested. More details about this test case and schematic of the computational domain are provided in section S2 of the Supporting Information.

In the following two sections, DBSCAN is employed to capture bubbles (section 4) and clusters (section 5) in CFD-DEM simulations of a bubbling fluidized bed and particle clustering in a periodic domain, respectively.

4 Capturing bubbles in fluidized bed simulations

The fluidized bed geometry simulated by Lu et al.[7] is implemented in this work. The fluidized bed is modeled as a vertical cylindrical pipe (for 3D simulations) and a rectangular domain (for 2D simulations). The simulations are performed using NGA.[39,40] Section 2.1 provides more details about the simulation setup. Figure 3 shows the 2D and 3D computational domains and Table 2 provides the simulation parameters. A grid independence study is performed to evaluate the sensitivity of CFD-DEM simulations on the grid size. The details

are provided in section S3 of the Supporting Information. A grid size of $3.5d_p$ and $2d_p$ is used for the 2D and 3D simulations, respectively. For the 2D simulations, we adopt the grid size used in the 2D simulations of Lu et al.[7] For the 3D simulations, we use a finer grid ($2d_p$) to test DBSCAN on a different grid resolution. Moreover, a finer grid allows using more processors, making the simulations faster. For both grid sizes, simulation results are independent of the grid.

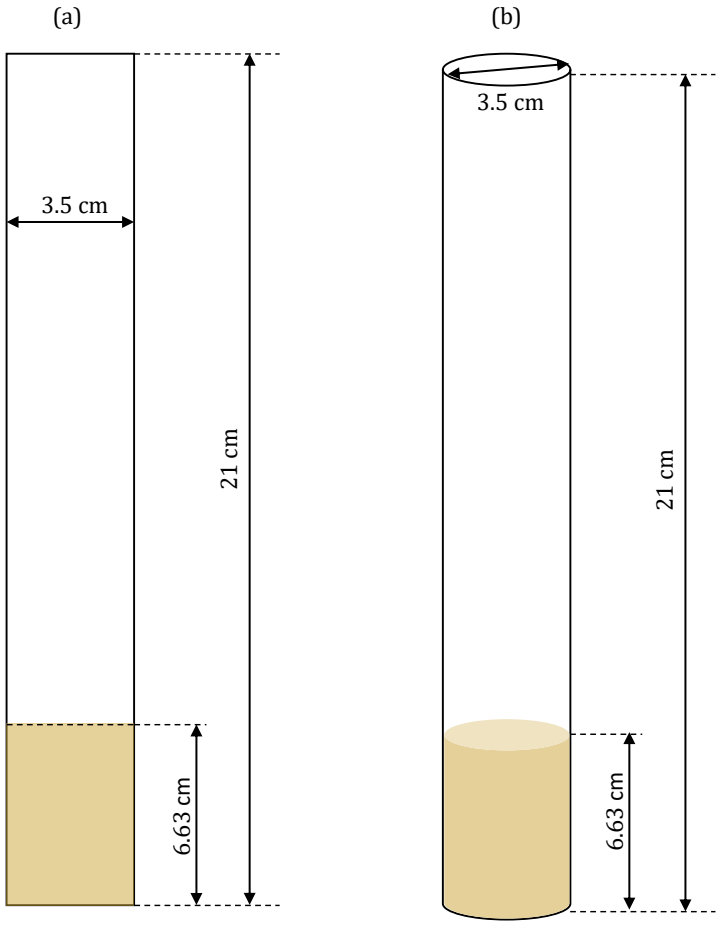

Figure 3: Computational domain of CFD-DEM simulations of a (a) 2D fluidized bed and (b) 3D fluidized bed geometry.

Air at room temperature is injected from the bottom of the bed to achieve fluidization. The time required to achieve a statistically steady state is around 1.1 seconds (three times the flow through time). The simulations are performed for nine seconds with a time step of 10

Table 2: Parameters for 2D and 3D fluidized bed simulations.

Parameter	**Value**
Fluidized bed height [cm]	21
Solid bed height [cm]	6.63
Bed width/diameter [cm]	3.5
Mesh $n_x \times n_y \times n_z$	$120 \times 20 \times 1$ (2D); $210 \times 35 \times 35$(3D)
Mesh size l_d [mm]	1.75 (2D); 1 (3D)
Particle diameter [μm]	500
Particle density [kg/m^3]	2580
Number of particles	10152 (2D); 586560 (3D)
Restitution coefficient	0.9
Wall restitution coefficient	0.75
Friction coefficient	0.3
Inlet gas velocity [m/s]	0.57 ($3U_{mf}$)

μs, and the output is saved every 0.01 seconds to have sufficient data for bubble statistics. The 2D and 3D simulations required 16 and 89 hours on a high-performance computing cluster using 20 and 48 cores (Intel Xeon Skylake 6148), respectively.

DBSCAN is used to capture gas bubbles and estimate their properties as described in section 2. Figure 4 schematizes the steps involved in bubbles identification by DBSCAN for 2D fluidized bed case. Figure 4a shows the simulated fluidized bed at a given instant during fluidization. The fluid volume fraction ε_f at a grid point is calculated via the volume averaging operation based on a filtering kernel.[40] The data pre-processing step described by eq. 4 in section 2.4 converts ε_f data into δ (0 for $\varepsilon_f \leq \varepsilon_f^*$ and 1 for $\varepsilon_f > \varepsilon_f^*$) represented by black ($\delta = 0$, emulsion phase) and white ($\delta = 1$, bubble phase) regions in Figure 4b. Two bubbles can be identified based on visual inspection. The position of grid points with $\delta = 1$, shown as black regions in Figure 4c, is passed to DBSCAN, discarding the remaining data (with $\delta = 0$). This step leads to a significant saving in the time required by DBSCAN as the number of grid points within bubbles is much smaller than the number of grid points outside bubbles. In this work, the bubble phase constitutes around 10% and 15% of the total grid points in the 2D and 3D fluidized bed simulations, respectively. Figure 4d shows the two bubbles identified by DBSCAN along with their convex hull boundary and chord. DBSCAN

required around 0.01 s to identify bubbles from the 2D simulation data at a given time.

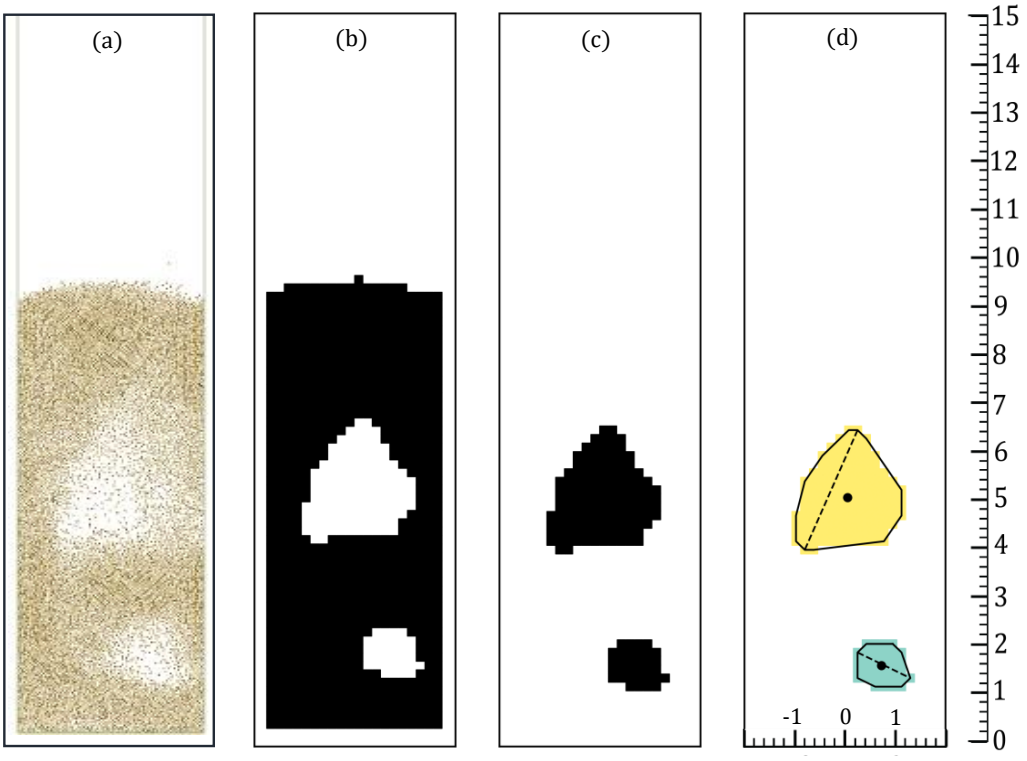

Figure 4: Application of DBSCAN to capture bubbles in a 2D fluidized bed CFD-DEM simulation. (a) The instantaneous position of solid particles in the fluidizing region at 4 s; (b) After converting ε_f into δ during data pre-processing, $\delta = 0$ ($\varepsilon_f \leq \varepsilon_f^*$) in the black region and $\delta = 1$ ($\varepsilon_f > \varepsilon_f^*$) in the white region; (c) Position of grid points with $\delta = 1$ is passed to DBSCAN; (d) Two bubbles identified by the DBSCAN algorithm (using $eps = 0.28$ cm and $minPts = 3$) shown in yellow and green. For the yellow bubble, the area, centroid, and chord length are 3.98 cm^2, $(0.05, 5.09)$ cm, and 2.67 cm, respectively. For the green bubble, the area, centroid, and chord length are 1.02 cm^2, $(0.73, 1.65)$ cm, and 1.17 cm, respectively.

Next, DBSCAN is tested for the 3D CFD-DEM fluidized bed simulation. Figure 5a shows the isosurface of bubbles ($\varepsilon_f = 0.75$) identified by Ansys EnSight – simulation data visualization software. The DBSCAN predictions are shown in Figure 5b, along with the convex hull boundaries and the chord of each bubble. These visual observations confirm the accuracy of DBSCAN in capturing bubbles in 2D and 3D fluidized bed simulations. DBSCAN required around 1.5 s to identify bubbles from the 3D simulation data at a given time. A few additional comparisons of DBSCAN predictions and Ansys Ensight are presented in section S4 of the Supporting Information.

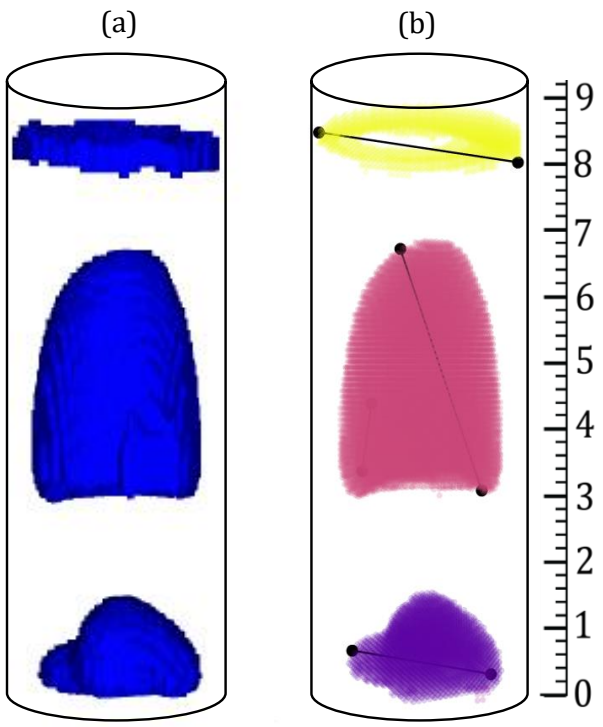

Figure 5: Application of DBSCAN to capture bubbles in a 3D fluidized bed CFD-DEM simulation. (a) Isosurface with $\varepsilon_f = 0.75$ plotted using Ansys EnSight; (b) Three bubbles identified by the DBSCAN algorithm (using $eps = 0.20$ cm and $minPts = 8$) shown in yellow, red, and purple. For the yellow bubble, the volume is 2.08 cm^3, the centroid is $(9.99, 1.65, 2.10)$ cm, and the chord length is 4.82 cm. For the red bubble, the volume is 19.76 cm^3, the centroid is $(5.81, 1.85, 2.01)$ cm, and the chord length is 4.77 cm. For the purple cluster, the volume is 4.83 cm^3, the centroid is $(1.15, 1.74, 2.08)$ cm, and the chord length is 3.28 cm.

DBSCAN results are analyzed to quantify the bubble size variation along the bed height. In the bubbling fluidization regime, bubbles tend to nucleate at the bottom of the bed and grow in size along the bed height. The variation in the bubble equivalent diameter with the bed height for 2D and 3D fluidized bed simulations is shown in Figure 6a along with the predictions of two correlations[52,53] from the literature. Predictions of DBSCAN and the correlations are in close agreement. Moreover, bubbles in the 3D simulation are slightly larger than in the 2D simulation. This observation is related to the inability of the 2D simulations to capture the coalescence effects accurately.[54]

Bubble size in a bubbling fluidized is known to follow the gamma distribution.[7,55] Figure 6b and Figure 6c show the Probability Density Functions (PDF) of bubble equivalent

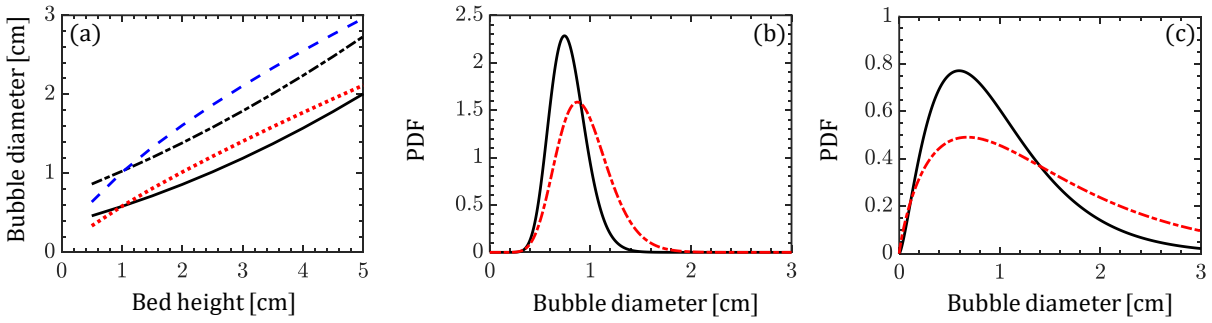

Figure 6: Bubble statistics in 2D and 3D fluidized bed simulations. (a) Variation in equivalent bubble diameter along the bed height obtained from 2D simulation (solid black line), 3D simulation (dash-dotted black line) Shen's correlation[53] (dashed blue line), Darton's correlation[52] (dotted red line). PDF of equivalent bubble diameter at 25% (solid black line) and 50% (dash-dotted red line) of the static bed height for (b) 2D simulation and (c) 3D simulation.

diameters at 25% and 50% of the static bed height for 2D and 3D fluidized bed simulations, respectively. These PDFs are obtained by fitting the gamma distribution to the bubble diameter data obtained using DBSCAN and are qualitatively similar to those obtained in the literature for similar fluidized bed configuration.[7,55] The PDFs peak decreases and width increases as the bed height increases, implying an increase in the fraction of larger bubbles.

Several investigations in the literature report a bimodal behavior of the bubble size distribution at higher bed heights. This observation is attributed to bubble breakage and coalescence.[7,55] The bimodal trend is captured by fitting the bubble diameter distribution using a second-order Gaussian distribution. Figure 7a and Figure 7b show the bubble size distribution for 2D and 3D fluidized bed simulations at 50% of the static bed height, respectively. Both 2D and 3D simulations show bi-modality in the bubble size distribution; however, the bi-modal behavior is much stronger in 3D simulations than in 2D simulations. Such differences are reported in the literature.[54]

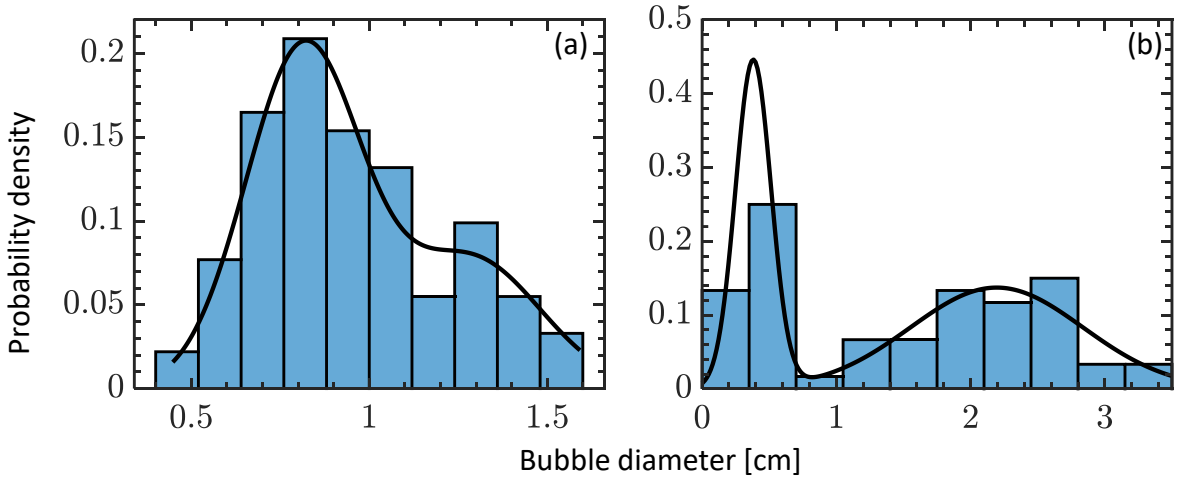

Figure 7: Equivalent bubble diameter distribution at 50% of the static bed height obtained from the CFD-DEM simulations of the (a) 2D fluidized bed and (b) 3D fluidized bed. Histograms are based on the bubble sizes estimated using DBSCAN. Solid lines represent the second-order Gaussian distribution fitted to the histogram data.

5 Capturing clusters in riser simulations

A triply periodic domain is considered to model particle clustering in a riser using the CFD-DEM approach in NGA.[39,40] This simulation configuration represents the fully-developed region of a riser far from the riser entrance and walls.[56] Section 2.1 provides more details about the CFD solver NGA and Table 3 provides the relevant simulation parameters. Particles are initially randomly distributed in the domain with a mean volume fraction of 0.015. Particles fall under gravity, entraining the gas phase, and the two-way coupling between the phases leads to particle clustering. A statistically steady state is reached in about 1 s. At this point, the average cluster size reaches a steady value. The simulation is performed for 5.5 s with a time step of 5 μs seconds, and the data is saved every 0.1 s. The simulation required 60 hours on 48 cores (Intel Xeon Skylake 6148) of a high-performance computing cluster. Consequently, DBSCAN is used to capture particle clusters and estimate their properties, as described in section 2. DBSCAN required around 3.5 s to identify clusters from the simulation data at a given time.

Figure 8a shows the isosurface of clusters with $\varepsilon_p = 0.03$ (calculated using eq. 3) plotted

Table 3: Simulation parameters for particle clustering in a 3D periodic domain.

Parameter	**Value**
Dimensions [cm]	$3.072 \times 3.072 \times 3.072$
Mesh $n_x \times n_y \times n_z$	$60 \times 60 \times 60$
Mesh size l_d [μm]	512
Particle diameter [μm]	200
Particle density [kg/m^3]	500
Mean particle volume fraction	0.015
Restitution coefficient	0.8
Friction Coefficient	0.092

using Ansys EnSight. The size and shape of particle clusters are distinct from those of bubbles in a fluidized bed. The clusters identified by DBSCAN and the corresponding chords are shown in Figure 8b. A comparison of Figure 8a and b shows that DBSCAN captures most of the clusters observed in the simulation. DBSCAN predictions for particle clusters at other time stamps are provided in section S5 of the Supporting Information.

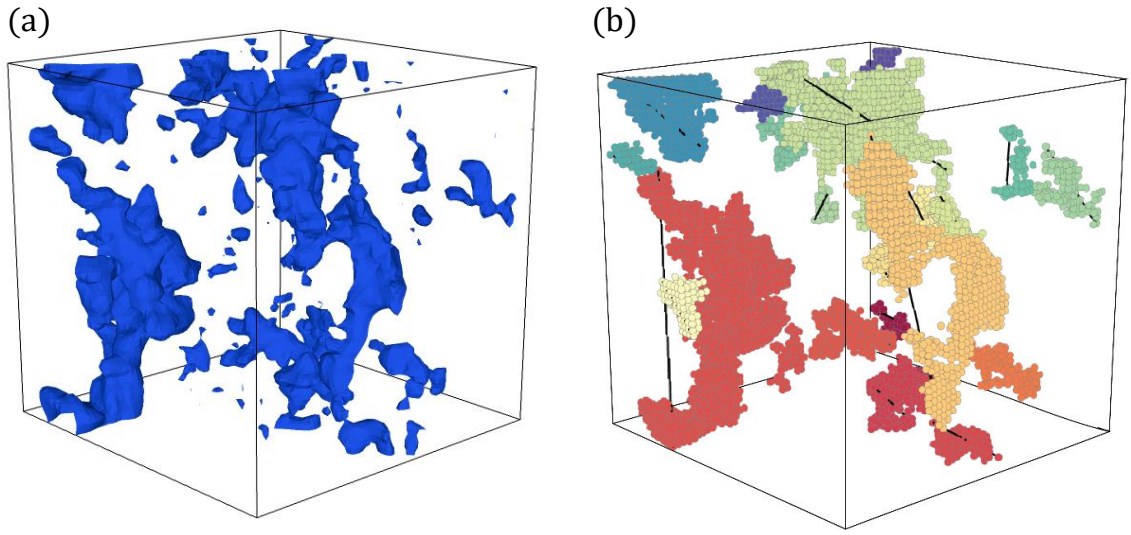

Figure 8: Application of the DBSCAN algorithm to capture particle clusters in CFD-DEM simulations of gas-solid flows in a 3D triply-periodic domain. (a) Isosurface with $\varepsilon_p = 0.03$ plotted in Ansys EnSight; (b) Clusters identified by the DBSCAN algorithm using $eps = 0.10$ cm and $minPts = 8$.

Figure 9a shows the PDFs obtained by fitting the log-normal distribution to cluster equivalent diameter (solid black line) and chord length (dashed red line) data obtained using

DBSCAN. The elongated shape of clusters makes the chord lengths larger than the equivalent diameter, as seen by the wider spread of the chord length PDF than the equivalent diameter PDF. The PDFs show that most clusters are smaller than 1 cm. However, the largest chord length is around 3.7 cm, which is close to *a priori* estimate of the cluster size $\mathcal{L} = \tau_p^2 g = 3.85$ cm. τ_p is the particle response time given as[41]

$$\tau_p = \frac{\rho_p d_p^2}{18\mu} \tag{9}$$

where ρ_p is the particle density and μ is the gas viscosity. τ_p is calculated to be 63 ms. We note that the simulation domain size is comparable to $\mathcal{L}$, constraining the cluster growth. However, the simulation configuration is adequate to test the ability of DBSCAN to capture particle clusters. Figure 9b shows the % volume of the computational domain occupied by clusters of different sizes. Each data point represents the % volume occupied by the clusters within a bin size of 0.5 cm. Although larger clusters ($>$ 1 cm) are fewer, they occupy 63% of the total volume occupied by clusters. These observations suggest the importance of accurately capturing the entire spectrum of cluster sizes and their chord length.

In this work, DBSCAN is applied to CFD-DEM simulations; however, it can be used with other CFD approaches, such as the two-fluid model where both gas and particles are solved on an Eulerian mesh. DBSCAN requires the gas or solid volume fraction at grid points to capture bubbles and clusters.

6 Calculation time

DBSCAN passes through each point in the dataset, solving the fixed-radius (*eps* in this work) near neighbors problem. The theoretical runtime complexity of the DBSCAN algorithm is $\mathcal{O}(n \log n)$ with n being the size of the dataset. We calculate the time complexity of DBSCAN using a 2D case consisting of two circles (radii: 1 cm and 2 cm) and a 3D case consisting of two spheres (radii: 2 cm and 3 cm). l_c is set to 1 cm and 1.06 cm for the 2D and 3D cases,

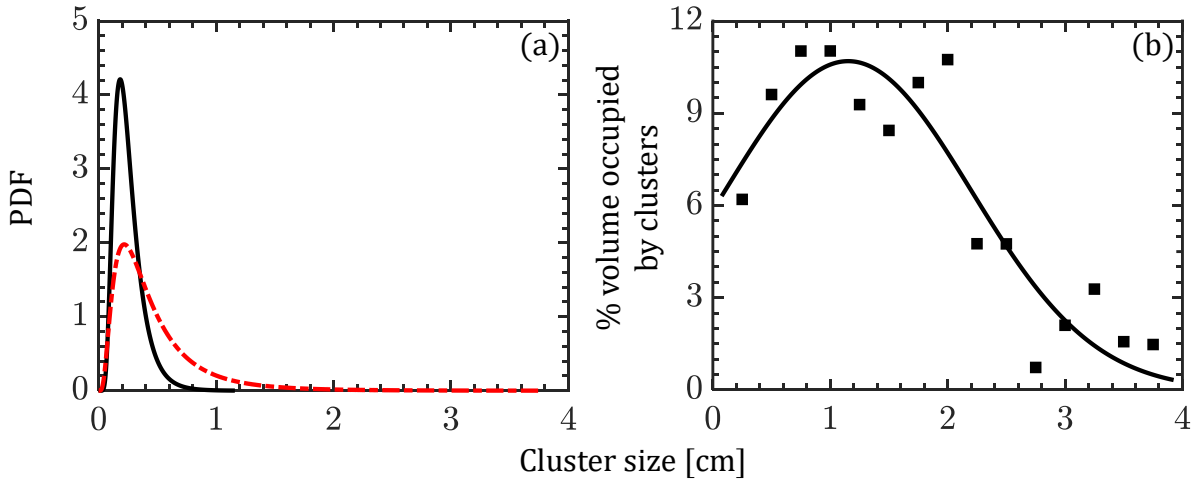

Figure 9: Distribution of cluster size and volume obtained from the CFD-DEM simulations of particle clustering using DBSCAN. (a) PDFs of the distribution of cluster equivalent diameter (solid black line) and chord length (dashed red line) obtained by fitting the log-normal distribution; (b) Distribution of the % volume of the computational domain occupied by clusters as a function of chord length (black symbols) and the Gaussian distribution fitted to the data (solid black line).

respectively. A side length of 10 cm is used for both domains. Grid size l_d is varied from 0.7 cm to 10 μm for the 2D case and 0.7 cm to 400 μm for the 3D case, increasing the grid resolution up to 100 million grid points.

First, DBSCAN is used while including the grid points inside and outside the circles. For this case, Figure 10a shows the time required by DBSCAN to identify $\texttt{clusters}_D$ for the 2D (red dashed line) and 3D (black dash-dotted line) datasets. Next, DBSCAN is implemented after discarding the data outside the circles, which is the strategy used in this work. The corresponding time required by DBSCAN is shown in Figure 10b. For both cases, the time complexity is close to $\mathcal{O}(n \log n)$ much faster than that of existing techniques.[4,7] Moreover, discarding the data outside the domains of interest resulted in an 85% reduction in the data, which is comparable to the reduction in the time required. In multiphase reactors, only a small fraction of space is covered by mesoscale structures. Thus, discarding the data outside these structures leads to a significant reduction in compute time in contrast to the techniques that process the entire grid data. For all the cases, the time required to calculate the $\texttt{clusters}_D$ properties is small in comparison to the DBSCAN execution.

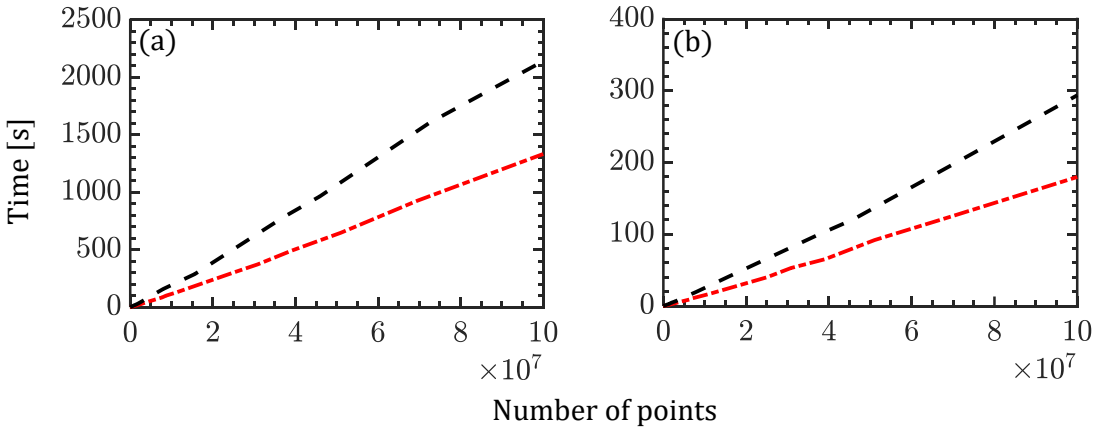

Figure 10: Scalability of DBSCAN for 2D and 3D computational domains. (a) for all the data in the domain (inside and outside the circles); (b) for the data within the circles. Dash-dotted red line: 2D case with l_d varying from 1.2 cm to 10 μm for a fixed $l_c = 1$ cm; Dashed black line: 3D case with l_d varying from 1.2 cm to 400 μm for a fixed $l_c = 1.06$ cm.

7 Data Availability and Reproducibility Statement

The numerical data from Figures 2–10, and Figures S1–S9 are available as a .zip file in the Supplementary Material. CFD simulation data is verified by comparing the bubble statistics with the literature. The developed DBSCAN-based methodology is verified by comparing its predictions with the CFD data visualizations using Ensight software.

8 Conclusion

Quantifying the mesoscale structures present in multiphase reactors, such as bubbles in a fluidized bed and clusters in a riser, is imperative in designing efficient multiphase reactors. For this purpose, multiphase CFD simulations are becoming common, complementing limited and expensive experimental measurements. However, large-scale CFD simulations require computationally fast post-processing tools to capture and analyze the mesoscale structures.

This work demonstrates the suitability of DBSCAN, a freely available unsupervised machine learning algorithm, to identify gas bubbles and particle clusters in 2D and 3D reactor simulations. DBSCAN calculates the distance between the nearest points to identify groups

such that for each point in a group, the neighborhood of a given radius contains at least a minimum number of points. DBSCAN is assessed for analytical shapes and CFD-DEM simulations of bubbling fluidized beds and particle clustering. Bubble and cluster properties, such as centroid, equivalent diameter, and longest chord length are calculated. These properties obtained at consecutive time frames can be further used to trace the bubble movement and calculate the bubble velocity as described in the literature. The computational complexity of DBSCAN is calculated using a high-resolution mesh and is found to be close to its theoretical value of $\mathcal{O}(n \log n)$, which is much faster than the available bubble and cluster detection algorithms. Moreover, DBSCAN only requires the gird information that lies within the bubbles/clusters, significantly reducing the calculation time. Along with capturing bubbles and clusters in fluidized beds and risers, DBSCAN can be easily extended to capture mesoscale structures in other multiphase flows, such as bubbles in a bubble column and preferential concentration of droplets or particles in a turbulent flow.

Acknowledgment

The authors greatly acknowledge National Supercomputing Mission (award number: DST / NSM/R&D_HPC_Applications/2021/15) and Science and Engineering Research Board (award number: SRG/2021/002255) for funding this research. National Supercomputing Mission (NSM) provided computing resources of 'PARAM Shivay' at the Indian Institute of Technology (BHU), Varanasi, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and the Department of Science and Technology (DST), Government of India.

References

(1) Utikar, R. P.; Ranade, V. V. Intensifying multiphase reactions and reactors: strategies and examples. ACS Sustainable Chemistry & Engineering **2017**, 5, 3607–3622.

(2) Goyal, H.; Sadula, S.; Vlachos, D. G. Microwave heating of slurries. Chemical Engineering Journal **2021**, 417, 127892.

(3) Goyal, H.; Desjardins, O.; Pepiot, P.; Capecelatro, J. A computational study of the effects of multiphase dynamics in catalytic upgrading of biomass pyrolysis vapor. AIChE Journal **2018**, 64, 3341–3353.

(4) Bakshi, A.; Altantzis, C.; Bates, R. B.; Ghoniem, A. F. Multiphase-flow Statistics using 3D Detection and Tracking Algorithm (MS3DATA): Methodology and application to large-scale fluidized beds. Chemical Engineering Journal **2016**, 293, 355–364.

(5) Shen, L.; Johnsson, F.; Leckner, B. Digital image analysis of hydrodynamics two-dimensional bubbling fluidized beds. Chemical engineering science **2004**, 59, 2607–2617.

(6) Movahedirad, S.; Molaei Dehkordi, A.; Banaei, M.; Deen, N.; van Sint Annaland, M.; Kuipers, J. Bubble size distribution in two-dimensional gas–solid fluidized beds. Industrial & engineering chemistry research **2012**, 51, 6571–6579.

(7) Lu, Y.; Huang, J.; Zheng, P. A CFD–DEM study of bubble dynamics in fluidized bed using flood fill method. Chemical Engineering Journal **2015**, 274, 123–131.

(8) Goyal, H.; Pepiot, P. A compact kinetic model for biomass pyrolysis at gasification conditions. Energy & Fuels **2017**, 31, 12120–12132.

(9) Ambekar, A. S.; Schwarzmeier, C.; Rüde, U.; Buwa, V. V. Particle-resolved turbulent flow in a packed bed: RANS, LES, and DNS simulations. AIChE Journal **2023**, 69, e17615.

(10) Ding, K.; Xiong, Q.; Zhong, Z.; Zhong, D.; Zhang, Y. CFD simulation of combustible solid waste pyrolysis in a fluidized bed reactor. Powder Technology **2020**, 362, 177–187.

(11) Stefanidis, G. D.; Merci, B.; Heynderickx, G. J.; Marin, G. B. CFD simulations of steam cracking furnaces using detailed combustion mechanisms. Computers & chemical engineering **2006**, 30, 635–649.

(12) Bala Chandran, R.; De Smith, R. M.; Davidson, J. H. Model of an integrated solar thermochemical reactor/reticulated ceramic foam heat exchanger for gas-phase heat recovery. International Journal of Heat and Mass Transfer **2015**, 81, 404–414.

(13) Uglietti, R.; Micale, D.; La Zara, D.; Goulas, A.; Nardi, L.; Bracconi, M.; van Ommen, J. R.; Maestri, M. A combined experimental and multiscale modeling approach for the investigation of lab-scale fluidized bed reactors. Reaction Chemistry & Engineering **2023**,

(14) Wang, S.; Luo, K.; Fan, J. CFD-DEM coupled with thermochemical sub-models for biomass gasification: Validation and sensitivity analysis. Chemical Engineering Science **2020**, 217, 115550.

(15) Bracconi, M.; Maestri, M.; Cuoci, A. In situ adaptive tabulation for the CFD simulation of heterogeneous reactors based on operator-splitting algorithm. AIChE Journal **2017**, 63, 95–104.

(16) Zhao, Y.; Chew, J. W. Discrete element method study on hopper discharge behaviors of binary mixtures of nonspherical particles. AIChE Journal **2020**, 66, e16254.

(17) Neau, H.; Pigou, M.; Fede, P.; Ansart, R.; Baudry, C.; Mérigoux, N.; Laviéville, J.; Fournier, Y.; Renon, N.; Simonin, O. Massively parallel numerical simulation using up to 36,000 CPU cores of an industrial-scale polydispersed reactive pressurized fluidized bed with a mesh of one billion cells. Powder Technology **2020**, 366, 906–924.

(18) Sobrino, C.; Acosta-Iborra, A.; Izquierdo-Barrientos, M.; De Vega, M. Three-dimensional two-fluid modeling of a cylindrical fluidized bed and validation of the maximum entropy method to determine bubble properties. Chemical Engineering Journal **2015**, 262, 628–639.

(19) Buchheit, K.; Altantzis, C.; Bakshi, A.; Jordan, T.; Van Essendelft, D. The BubbleTree toolset: CFD-integrated algorithm for Lagrangian tracking and rigorous statistical analysis of bubble motion and gas fluxes for application to 3D fluidized bed simulations. Powder Technology **2018**, 338, 960–974.

(20) Sharma, K. G.; Kaisare, N. S.; Goyal, H. A recurrent neural network model for biomass gasification chemistry. Reaction Chemistry & Engineering **2022**, 7, 570–579.

(21) Beck, D. A.; Carothers, J. M.; Subramanian, V. R.; Pfaendtner, J. Data science: Accelerating innovation and discovery in chemical engineering. AIChE Journal **2016**, 62, 1402–1416.

(22) Tanudjaja, H. J.; Chew, J. W. Application of machine learning-based models to understand and predict critical flux of oil-in-water emulsion in crossflow microfiltration. Industrial & Engineering Chemistry Research **2022**, 61, 8470–8477.

(23) Bracconi, M. Intensification of catalytic reactors: A synergic effort of Multiscale Modeling, Machine Learning and Additive Manufacturing. Chemical Engineering and Processing-Process Intensification **2022**, 181, 109148.

(24) Hough, B. R.; Beck, D. A.; Schwartz, D. T.; Pfaendtner, J. Application of machine learning to pyrolysis reaction networks: Reducing model solution time to enable process optimization. Computers & Chemical Engineering **2017**, 104, 56–63.

(25) Xu, D.; Tian, Y. A comprehensive survey of clustering algorithms. Annals of Data Science **2015**, 2, 165–193.

(26) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. series c (applied statistics) **1979**, 28, 100–108.

(27) Cheng, Y. Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence **1995**, 17, 790–799.

(28) Ankerst, M.; Breunig, M. M.; Kriegel, H.-P.; Sander, J. OPTICS: Ordering points to identify the clustering structure. ACM Sigmod record **1999**, 28, 49–60.

(29) Sander, J.; Ester, M.; Kriegel, H.-P.; Xu, X. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. Data Mining and Knowledge Discovery **1998**, 2, 169–194.

(30) Yin, S.; Zhong, W.; Song, T.; Lu, P.; Chen, Y. Clusters identification and meso-scale structures in a circulating fluidized bed based on image processing. Advanced Powder Technology **2019**, 30, 3010–3020.

(31) Yin, S.; Song, T.; Chen, T.; Chen, Y.; Lu, P.; Guo, Q. Study of cluster characteristics in a circulating fluidized bed riser. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects **2020**, 42, 1553–1564.

(32) Wang, C.; Lan, X.; Sun, Z.; Han, M.; Gao, J.; Ye, M.; Zhu, J. Cluster Identification by a k-means Algorithm-Assisted Imaging Method in a Laboratory-Scale Circulating Fluidized Bed. Industrial & Engineering Chemistry Research **2021**, 61, 942–956.

(33) Wang, T.; Deng, A.; He, Y.; Wu, B.; Gao, R.; Tang, T. Artificial intelligence-based approach for cluster identification in a CFB riser. Chemical Engineering Science **2022**, 118379.

(34) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X., et al. A density-based algorithm for discovering clusters in large spatial databases with noise. kdd. 1996; pp 226–231.

(35) Schubert, E.; Sander, J.; Ester, M.; Kriegel, H. P.; Xu, X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Transactions on Database Systems (TODS) **2017**, 42, 1–21.

(36) Hahsler, M.; Piekenbrock, M.; Doran, D. dbscan: Fast density-based clustering with R. Journal of Statistical Software **2019**, 91, 1–30.

(37) Turner, S. NuGet – DBSCAN. `https://www.nuget.org/packages/DBSCAN/`.

(38) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **2011**, 12, 2825–2830.

(39) Desjardins, O.; Blanquart, G.; Balarac, G.; Pitsch, H. High order conservative finite difference scheme for variable density low Mach number turbulent flows. Journal of Computational Physics **2008**, 227, 7125–7159.

(40) Capecelatro, J.; Desjardins, O. An Euler–Lagrange strategy for simulating particle-laden flows. Journal of Computational Physics **2013**, 238, 1–31.

(41) Capecelatro, J.; Pepiot, P.; Desjardins, O. Numerical characterization and modeling of particle clustering in wall-bounded vertical risers. Chemical Engineering Journal **2014**, 245, 295–310.

(42) Capecelatro, J.; Pepiot, P.; Desjardins, O. Numerical investigation and modeling of reacting gas-solid flows in the presence of clusters. Chemical Engineering Science **2015**, 122, 403–415.

(43) Kasbaoui, M. H.; Koch, D. L.; Desjardins, O. The rapid distortion of two-way coupled particle-laden turbulence. Journal of Fluid Mechanics **2019**, 877, 82–104.

(44) Kuhn, M. B.; Desjardins, O. Experimentally validated high-fidelity simulations of a liquid jet in supersonic crossflow. International Journal of Multiphase Flow **2022**, 156, 104195.

(45) Cundall, P. A.; Strack, O. D. A discrete numerical model for granular assemblies. geotechnique **1979**, 29, 47–65.

(46) Meyer, M.; Devesa, A.; Hickel, S.; Hu, X. Y.; Adams, N. A. A conservative immersed interface method for large-eddy simulation of incompressible flows. Journal of Computational Physics **2010**, 229, 6300–6317.

(47) Rüdisüli, M.; Schildhauer, T. J.; Biollaz, S. M.; Wokaun, A.; van Ommen, J. R. Comparison of bubble growth obtained from pressure fluctuation measurements to optical probing and literature correlations. Chemical engineering science **2012**, 74, 266–275.

(48) Wang, J.; Ge, W.; Li, J. Eulerian simulation of heterogeneous gas–solid flows in CFB risers: EMMS-based sub-grid scale model with a revised cluster description. Chemical Engineering Science **2008**, 63, 1553–1571.

(49) Qi, X.-B.; Zeng, T.; Huang, W.-X.; Zhu, J. J.; Shi, Y.-F. Experimental study of solids holdups inside particle clusters in CFB risers. Sichuan University Engineering Science Edition **2005**, 37, 46–50.

(50) Sharma, A. K.; Tuzla, K.; Matsen, J.; Chen, J. C. Parametric effects of particle size and gas velocity on cluster characteristics in fast fluidized beds. Powder Technology **2000**, 111, 114–122.

(51) Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H. The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software (TOMS) **1996**, 22, 469–483.

(52) Darton, R.; LaNauze, R.; Davidson, J.; Harrison, D. Bubble growth due to coalescence in fluidised beds. Trans Inst Chem Eng **1977**, 55, 274 – 280.

(53) Shen, L.; Johnsson, F.; Leckner, B. Digital image analysis of hydrodynamics two-dimensional bubbling fluidized beds. Chemical Engineering Science **2004**, 59, 2607–2617.

(54) Bakshi, A.; Altantzis, C.; Bershanska, A.; Stark, A.; Ghoniem, A. On the limitations of 2D CFD for thin-rectangular fluidized bed simulations. Powder Technology **2018**, 332, 114–119.

(55) Busciglio, A.; Vella, G.; Micale, G.; Rizzuti, L. Analysis of the bubbling behaviour of 2D gas solid fluidized beds: Part I. Digital image analysis technique. Chemical Engineering Journal **2008**, 140, 398–413.

(56) Beetham, S.; Capecelatro, J. Biomass pyrolysis in fully-developed turbulent riser flow. Renewable Energy **2019**, 140, 751–760.