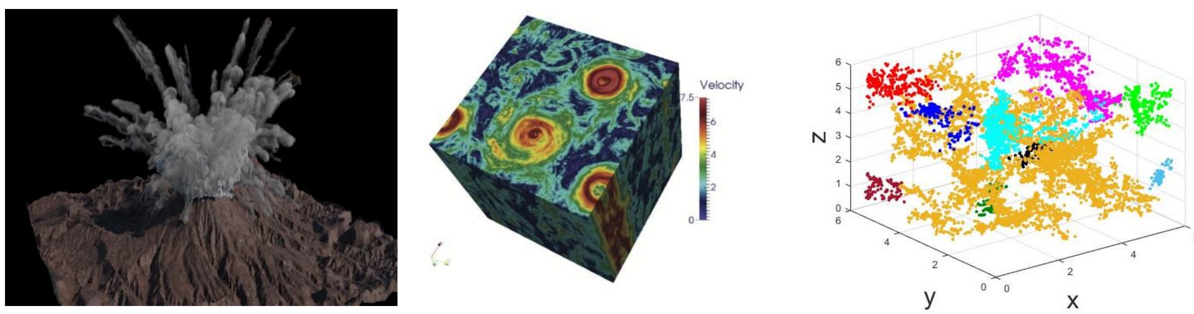
Clustering of particles in turbulent flow using Voronoi tessellation

Graduate Students: Reza Momenifar, Jonathan Holt[1] (Civil & Environmental Engineering)
Faculty sponsor: Prof. Simon tsz fung Mak
Undergraduate course: Machine Learning and Data Mining (STA 325L-001), Fall 2020

Description: For the fall 2020 Data Expedition initiative, we propose a dataset in the context of particle clustering that contains topological features of clusters including cluster size (volume), number of particles and their positions in each cluster. The dataset is extracted from many numerical simulations in 3D space, performed in our Theoretical and Computational Fluid Dynamics Group, which models the distributions of a large number of particles lying under an idealized turbulence in a cubic box. In these simulations, three independent control parameters representing the properties of turbulent flow and particles are varied and particles' position and other dynamical properties of particles and flow fields (e.g. velocity) are stored. In our recent study [1], which was focused on local analysis of particle-clustering in turbulence, the Voronoi tessellation analysis was performed and clusters (based on its definition in this context) were identified and so the existing dataset was enriched with new information pulled from this analysis. This proposal guides students to practice an abridged version of this study, which is briefly illustrated in figure 1.



a-) volcanic eruption b-) Numerical simulations c-) Cluster visulization
Figure 1: particle clustering in turbulent flow

Size of dataset: This Voronoi analysis has been implemented for 120 different simulations, where each simulation contains 20 different time steps to the aim of statistical convergence. Therefore, 2400 datasets containing multiple columns including topological and dynamical features of clusters are available. For the purpose of this proposal, we can offer any number of these datasets, in which the number of rows (indicating number of clusters) varies from 10,000 to 3,000,000 (depending on the type of simulation[2]) and each row contains count and position of particles in a cluster along with the cluster volume.

Potential classroom exercises:

1- What are the applications of Voronoi tessellation analysis and how it can bring insight into the distribution of clusters?
   a- Students become familiar with definition of Voronoi tessellation and examples of its application in marketing, robotics, statistics and data analysis, computer graphics, geophysics, epidemiology, Astronomy, and meteorology.
   b- Students gain intuition into how to perform this analysis to identify clusters in the context of particle clustering.

---

[1] mohammadreza.momenifar@duke.edu and Jonathan.Holt@duke.edu
[2] We can easily shrink number of rows by randomly selecting a fraction of them, if a dataset seems too large.

c- Students load data (position of particles within clusters) to visualize the clusters.
d- Considering cluster volume as a random variable, students compute its probability distribution function (PDF).
e- Students plot the PDF of cluster volumes for multiple simulations representing different control parameters. Given that size of cluster volumes vary significantly in different simulations (multi-scale analysis), students may need to redo the previous step and adjust the bin size so that they obtain a smooth distribution.
f- Students are trained how to interpret the distributions from both the statistical and physical points of view.

2- How can one gain insight into the topology (e.g. size, orientation) of clusters?
  a- Considering cluster volume as a random variable, students compute its mode, first and second moments along with standard deviation.
  b- Students plot the quantities in the previous step for multiple simulations representing different control parameters.
  c- Students plot the number of particles in each cluster against the volume of the cluster to find the relationship between them.
  d- Students are trained how to perform the singular value decomposition (SVD) using the position of particles in each cluster and find the principal axes and principal values of each cluster to gain insight on the spatial orientation of clusters (e.g. the direction of the greatest spread of particle distances from the cluster centroid).
  e- Students are trained how to interpret the results in previous steps and validate their interpretation with the visualization of clusters.
3- (Optional) How do different properties of particles and turbulent flow affect the topology of clusters?
  a- Students are introduced to some basic concepts in fluid dynamics to understand why particles form a cluster and how different properties of turbulence and particles can affect the general features of clusters (e.g. size, shape and orientation)

Techniques:

- **a- Coding techniques**: Student can analyze the datasets using Python, MATLAB or any other programming languages
- **b- Computational techniques**: multi-scale statistical analysis, Probability Distribution Function (PDF), statistical moments, Singular Value Decomposition (SVD)

Sources:

1- Momenifar, M., Bragg, A.~D.\ 2019.\ Local analysis of the clustering, velocities and accelerations of particles settling in turbulence.\ arXiv e-prints arXiv:1908.00341.
2- Ireland, Peter J., et al. "Highly parallel particle-laden flow solver for turbulence research." *Computers & Fluids* 76 (2013): 170-177.
3- Momenifar, Mohammadreza, Rohit Dhariwal, and Andrew D. Bragg. "Influence of Reynolds number on the motion of settling, bidisperse inertial particles in turbulence." Physical Review Fluids 4.5 (2019): 054301.
4- Ireland, Peter J., Andrew D. Bragg, and Lance R. Collins. "The effect of Reynolds number on inertial particle dynamics in isotropic turbulence. Part 2. Simulations with gravitational effects." Journal of Fluid Mechanics 796 (2016): 659-711.