Identifying individual granules from SIM z-stacks

Raw data

* SIM z-stack. Each z-slice is a 16-bit image. The raw data is a 3D matrix of grayscale intensity values ranging from 0-65536.

Pixel data to 3D point cloud

* Max entropy thresholding using stack histogram: Converts stack to a binary stack.
* 3D sparse binary matrix is the input to Python.
* Row, Col, and z-slice number of high fluorescent intensity regions is extracted from the sparse matrix and stored as an array of tuples.
* This is convenient for further computation because it removes the need to deal with large matrices, and allows us to consider only the regions of fluorescence within the matrix.
* In this way pixel data is converted to 3D coordinates, and the z-stack can be visualized as a 3D point cloud. Row number is the location of the high intensity fluorescent point in the x-dimension, and similarly col and z-slice number correspond to y-dimensional and z-dimensional data respectively.

Identifying clusters in the 3D point cloud, corresponding to germ plasm granules, using DBSCAN

* Assuming that the granules are spherical, a core point in the granule, should be surrounded by a 9x6 = 54 points, with √3 = 1.732 (Euclidean distance) being the distance of the farthest point from the core point.
* We could set minPts = 54 +1 (the point itself is counted) = 55 and ε = 1.732 as the parameters for DBSCAN. This would result in any point surrounded by 55 points, including the point itself, within a radius of 1.732 from it to be classified as a core point of a cluster. After core points are identified, all points within a 1.732 radius from the core points are classified as belonging to the same cluster. Each cluster corresponds to an individual granule.
* However, this could cause no clusters to be identified if granules are not perfectly spherical and vary in size.
* To avoid stringent parameter setting based on shape and size assumptions, DBSCAN is performed using a range of parameter settings.
* minPts = 3 to 55. minPts of 1,2 causes every individual point to be considered as a cluster, so 3 is the minimum. 55 is set as the maximum as it is number of nearest neighbors for a perfectly spherical granule.
* For each iteration of DBSCAN with minPts specified, a kNN distance plot is computed with k=minPts. The elbow in the graph is set as ε. The elbow is automatically found from the graph using second derivatives to find the point of maximum curvature.
* For every iteration of DBSCAN, cluster validation is performed to determine the best combination of parameters.
* In this way individual granules can be identified without *a prior* knowledge of size, shape and number of granules present.