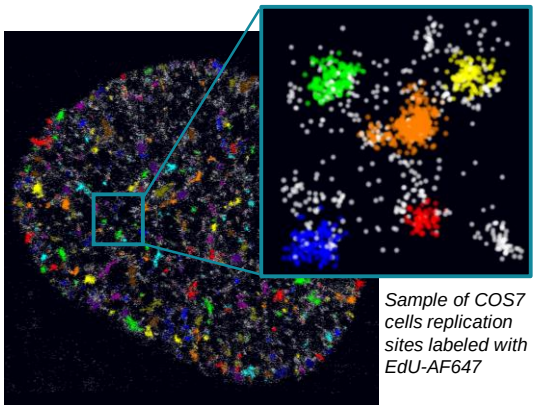
How to analyze clusters in SMLM data

Single-molecule localization microscopy is particularly adapted for clustering analysis, considering that the final dataset is not simply an image, but the coordinates of individual molecules. Many applications require clustering analysis: clathrins, nuclear pores, bacteria, DNA replication sites… biological entities tend to accumulate and form 3D clusters whose size, density and distribution are worth quantifying. These factors can then describe qualitatively a biological structure, and they can be compared quantitatively between conditions, for example to measure the effect of a given drug on the size of clusters, etc. Here, we describe the main algorithms developed to analyze clusters of single molecules.



Sample of COS7 cells replication sites labeled with EdU-AF647

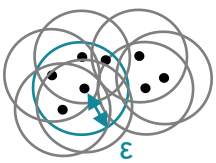
How can I isolate individual clusters ?

Several algorithms enable the identification of clusters in a dataset, including:

DBSCAN (Density Based Spatial Clustering of Applications with Noise) requires two input parameters: a distance ε and a minimum number of neighbors MinPts. For each localization in the dataset, the algorithm searches whether it has enough neighbors MinPts within the distance ε. If yes, it considers the localization as part of a cluster, etc.

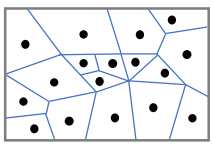
Voronoi partitions the image into polygons, where each polygon contains one, and only one, localization. The area/volume of the polygon is indicative of the density of localizations: a dense region will have small polygons while a low-density region will have big polygons. The user can choose a density threshold, above which localizations are considered as part of a cluster.

DBSCAN
Ester et al. 1996

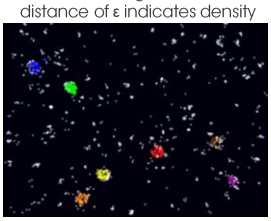


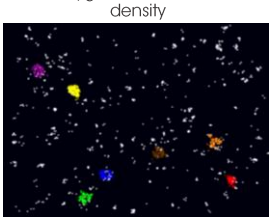
Number of neighbors within a distance of ε indicates density

Voronoi segmentation (SR-Tesseller)
Levet et al. 2015



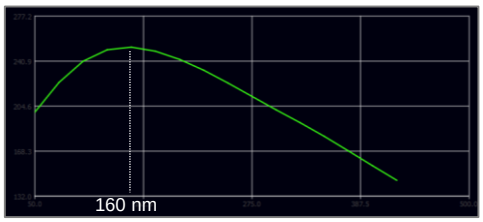
Polygon size indicates density





In practice

How do I know if there are clusters in my dataset?
The K Ripley function evaluates whether a population of localizations is aggregated or randomly distributed based on a neighborhood analysis.



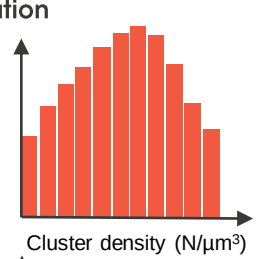
The bell-shaped curve indicates the presence of aggretated datapoints and provides an estimate of size of these aggregates.

Abbelight SAFe modules and NEO software offer features covering:

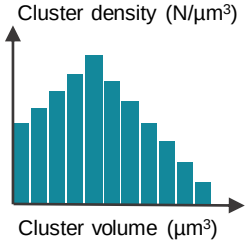
- Descriptive spatial statistics
- Optimized clustering algorithms using smart and fast spatial partitioning of data:
  - K Ripley function
  - DBSCAN algorithm
  - Voronoi segmentation
- Quantitative cluster info: number of clusters, cluster surface area/volume, density number of localizations, centroids, radius of gyration
- 3D real-time interactive visualization
- Export of quantitative results

Quantification

Clusters	1	2	3
Centroid X Y Z (µm)	 669 526 0.313	 654 514 0.302	 678 525 0.308
Number of localizations	277	249	256
Volume (µm³)	0.015	0.019	0.017



Cluster density (N/µm³)



Cluster volume (µm³)

contact@abbelight.com