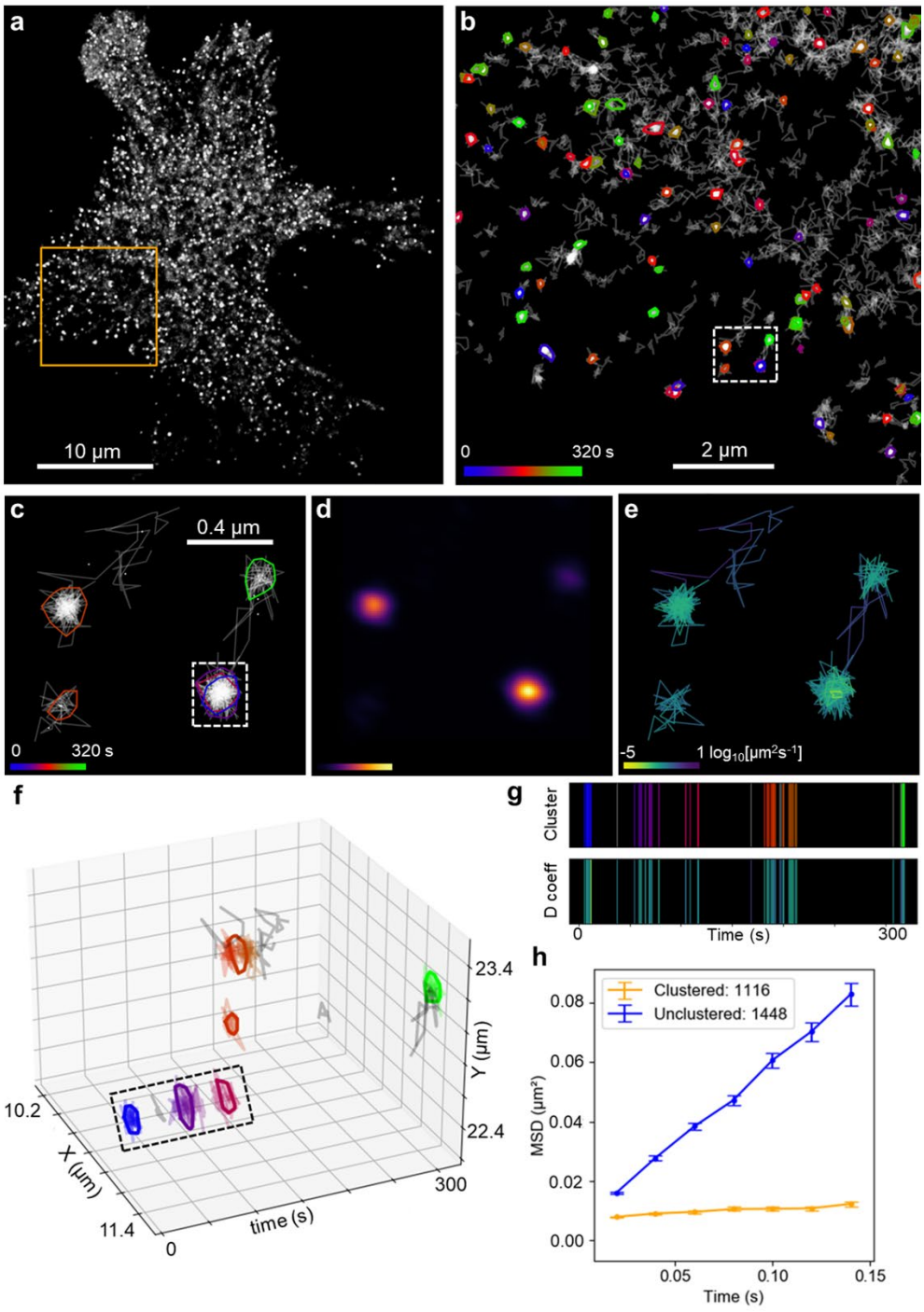
NASTIC USER MANUAL



Tristan Wallis t.wallis@uq.edu.au

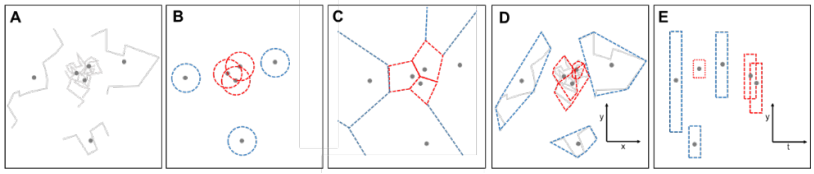
Single Molecule Neuroscience Laboratory (Fred Meunier)

Queensland Brain Institute

The University of Queensland, Australia

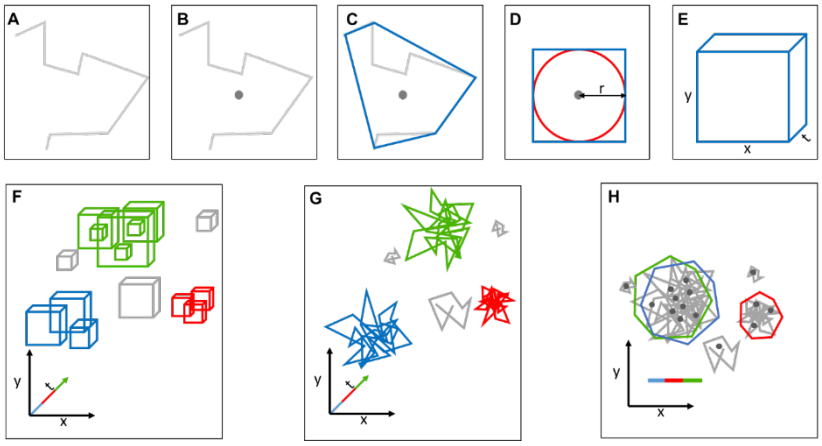
Introduction

NASTIC = nanoscale spatiotemporal indexing clustering, which represents the QBI Single Molecule Neuroscience lab's take on molecular clustering in live cells. NASTIC works by establishing whether entire molecular trajectories overlap in both space and time, and then building up clusters based on this potential overlap. NASTIC takes a different approach than widely used clustering algorithms which rely on DBSCAN or Voronoi tessellation to establish thresholds for molecular detections. NASTIC also differs markedly from these algorithms in that it also determines whether molecules are interacting in time.



Schematic representation of clustering algorithms. (A) Molecular trajectory data, with each trajectory spatial centroid indicated with a dot. **(B)** DBSCAN. Multiple molecule centroids present within a defined radius are considered clustered (red circles). The most effective radius (**ε**) and minimum number of centroids within it (MinPts) are determined empirically. **(C)** Voronoi tessellation. Tiles are drawn around each centroid such that the distance from any point within the tile is closer to its centroid than to any other centroid. Molecular centroids with tile areas less than an empirically determined threshold (red) are considered clustered. **(D)** Spatial indexing. Clustered molecules are determined by overlapping 2D bounding regions (red) defining the spatial extent of each molecular trajectory. **(E)** Spatiotemporal indexing. Each trajectory bounding region is assigned an arbitrary "thickness" in the time dimension. Overlapping 3D bounding regions represent spatiotemporally clustered molecules.

NASTIC relies on the R-tree spatial indexing algorithm, which is widely used in databases, mapping and especially videogames, where it is used to establish whether in-game objects, such as bullets and bad guys, are interacting. R-tree and other -tree algorithms such as Quad-tree, Oct-tree etc are highly optimised, which allows large numbers of interactions to be calculated quickly in order to not slow the game down unduly. We leverage this ability to rapidly establish whether trajectory bounding boxes are overlapping.



Schematic representation of spatiotemporal indexing clustering workflow. (A) Molecular trajectory composed of individual detections. **(B)** Spatiotemporal centroid representing the trajectory's average position in space and time. **(C)** Convex Hull (blue) defining the approximate spatial extent of the trajectory. **(D)** Simplified 2D spatial bounding box (blue square) based on the approximate radius (r) of the Convex Hull (red circle). **(E)** 3D spatiotemporal bounding box of user defined "thickness" in the time dimension. **(F)** R-tree spatiotemporal index of all trajectory bounding boxes. Discrete clusters of overlapping bounding boxes are indicated in colour, unclustered boxes in grey. **(G)** 3D clusters of trajectories associated with overlapping bounding boxes. **(H)** 2D representation of clustered trajectories. Coloured polygons represent the spatial Convex Hull of all detections comprising each of the clustered trajectories. Clusters are coloured according to the averaged detection time of their component trajectories, allowing assignment of overlapping clusters (green and blue) occupying the same spatial extent at different times.

NASTIC allows us to determine not only whether a trajectory overlaps with another trajectory in space, but also whether the overlap occurs in time. The idea being that trajectories which overlap spatiotemporally may represent clusters of molecules interacting at a certain space and time on the plasma membrane. Additionally, spatiotemporal metrics also allow us to

determine whether clusters repeatedly form and reform on the same region of a plasma membrane – hotspots. The temporal component of NASTIC, literally gives us another dimension of molecular dynamics data to play with.

NASTIC was conceived by Fred Meunier and Tristan Wallis during the first great Coronavirus lockdown of 2020, and subsequently prototyped and implemented as a Python GUI by Tristan Wallis 2020-2021. Sophie Hou assisted with aspects of GUI debugging. Rachel Gormal and Merja Joensuu provided valuable testing-to-destruction feedback and feature requests, and Nela Durisic provided critical MSD input. None of us are trained computer science graduates! So, while very effort has been made to create a functional, useful and stable program, bugs and missing features are bound to crop up. Your input with bug reports and feature requests for both the program and this documentation is very important to help NASTIC mature.

NASTIC is released under a Creative Commons licence. This is a very common open source licence which means you are free to use or modify it, and give/sell it to others to use/modify, under the proviso that you or they a) don't release changed versions under a more restrictive license, and b) properly acknowledge the original author (me).

This manual should help you to become a world leader in spatiotemporal cluster analysis. In addition, NASTIC has popup help for most of its buttons and fields which should help you further. If in doubt, just contact me at t.wallis@uq.edu.au

Computer requirements

NASTIC is a Python script and requires Python 3.8 or later to run. Python is available for most computer platforms so you can run NASTIC on Windows, Linux and Mac, or even on a Raspberry Pi if you were so inclined. NASTIC will not run on the older version of Python 2.7 which is still lingering on a lot of computer systems. You are strongly encouraged to either visit https://www.python.org and download and install the latest version* or hassle someone in IT to do it for you. You/they will also need to install a number of Python modules, which NASTIC uses to do a lot of the heavy lifting. This is simple to do from a command line.

```
Microsoft Windows [Version 10.0.19042.1052]
(c) Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>python -m pip install matplotlib numpy pandas pysimplegui rtree scipy scikit-learn seaborn
```

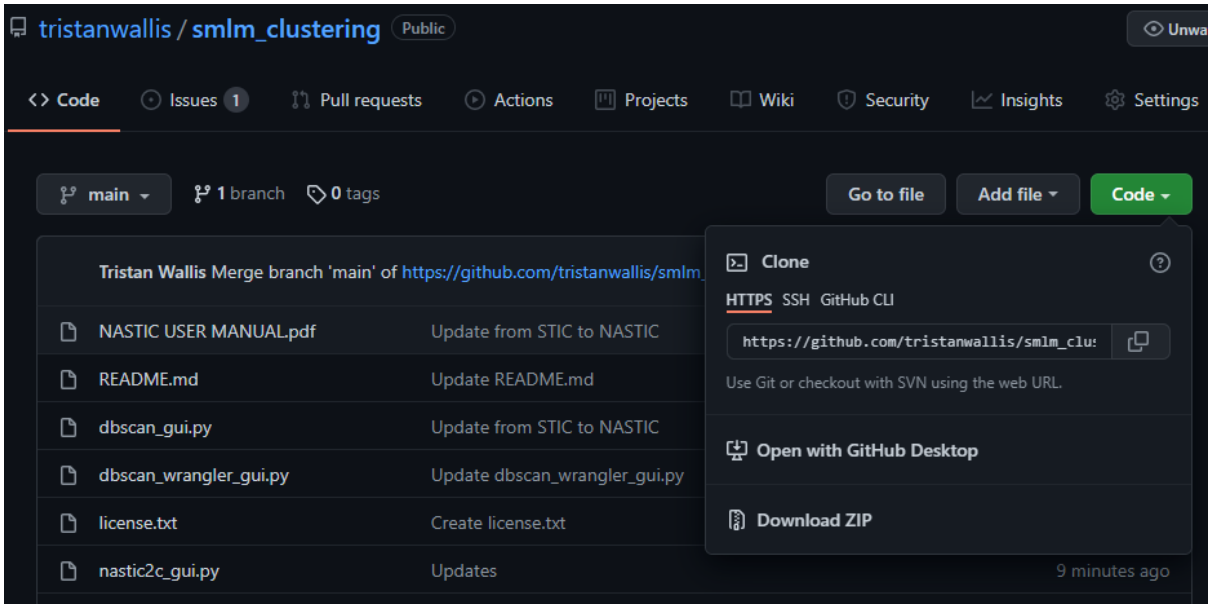
NASTIC is a multithreaded application which means that the intensive number crunching is farmed out to however many cores your computer has. You'll therefore theoretically get better performance on a 4Ghz computer with 8 cores than you will on a 5Ghz computer with 4 cores. But even on modest i5 laptop such as mine most analyses should take less than a minute.

IMPORTANT: Because of the multithreading, NASTIC does not play properly on virtual computers which dynamically assign the user virtual cores depending on usage etc. You will need to run it on a physical computer.

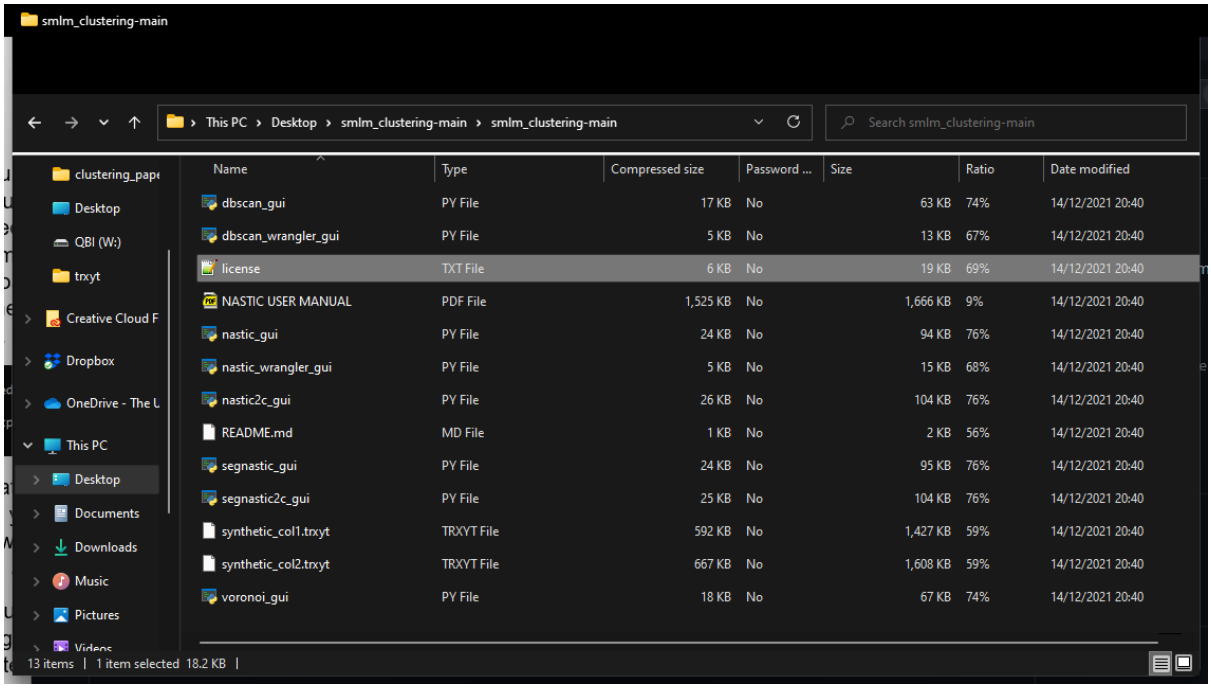
***ALSO IMPORTANT:** At the time of writing the latest version of Python is **3.10**. Bafflingly, at this stage none of the modules are compiled, which means that the pip commands as detailed above will probably fail because your computer will try to unsuccessfully build the modules. Consider using Python **3.8** or **3.9**.

Obtaining NASTIC

The latest greatest versions of NASTIC and related clustering software are always available on Github: https://github.com/tristanwallis/smlm_clustering. The total download is just a few Mb so it should only take 10 seconds to download and ensure that you're not running a superseded version. Just click on the green "Code" button and then "Download zip".



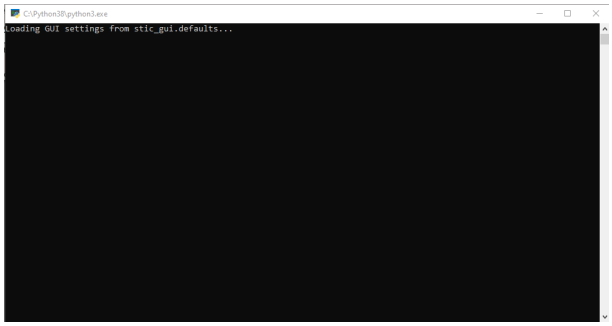
This will then download smlm_clustering_main.zip

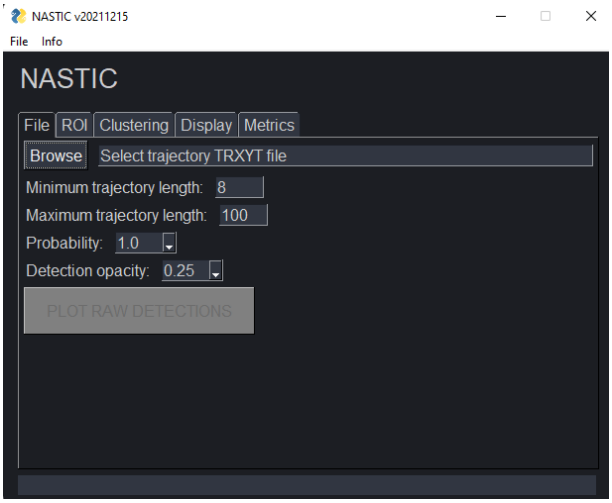


Extract these files into the directory of your choice and you should be good to go!

Running NASTIC

Provided that you've installed Python and the required modules correctly, double clicking on **NASTIC_gui.py** should open a command shell, load a brief splash screen and then the main program itself. The command shell is where NASTIC will print various useful information as it runs. If you close the shell you'll close NASTIC too.





NASTIC will display its version number (just the datestamp of its last changes) in the window title, so you'll always know what version you are running.

In order to simplify the potentially confusing clustering workflow, NASTIC is divided into a number of functional tabs, which it will automatically switch to as appropriate:

- **File:** select trajectory data, screen it based on trajectory length, and display the raw detections
- **ROI:** draw one or more rectangular or free hand regions of interest on the raw detection display (or load in previously saved ROIs) and select encompassed trajectories with optional density screening
- **Clustering:** enter the appropriate parameters for spatiotemporal clustering
- **Display:** plot the results of the clustering with a large range of options for trajectory, centroid, cluster and hotspot display, and optionally export high resolution images for publication
- **Metrics:** visualise and save a wide range of cluster metrics

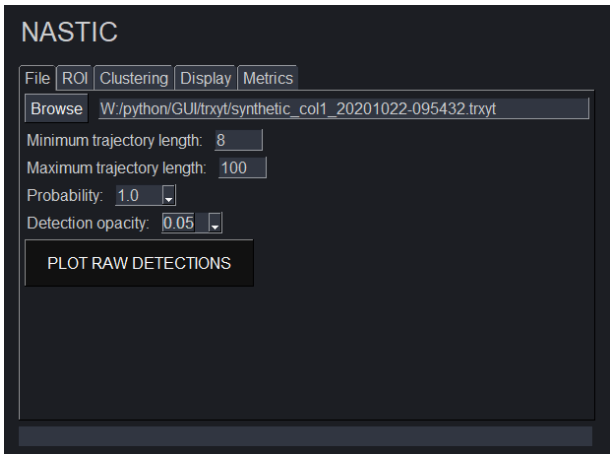
Each of these tabs will be detailed below.

File Tab

NASTIC works on simple text files where each line contains the info for a single detection: trajectory number, X co-ordinate (µm), Y co-ordinate (µm) and time (sec), separated by spaces (not commas or tabs). These files are generated by Matlab from PalmTracer output, and should have a **trxyt** suffix.

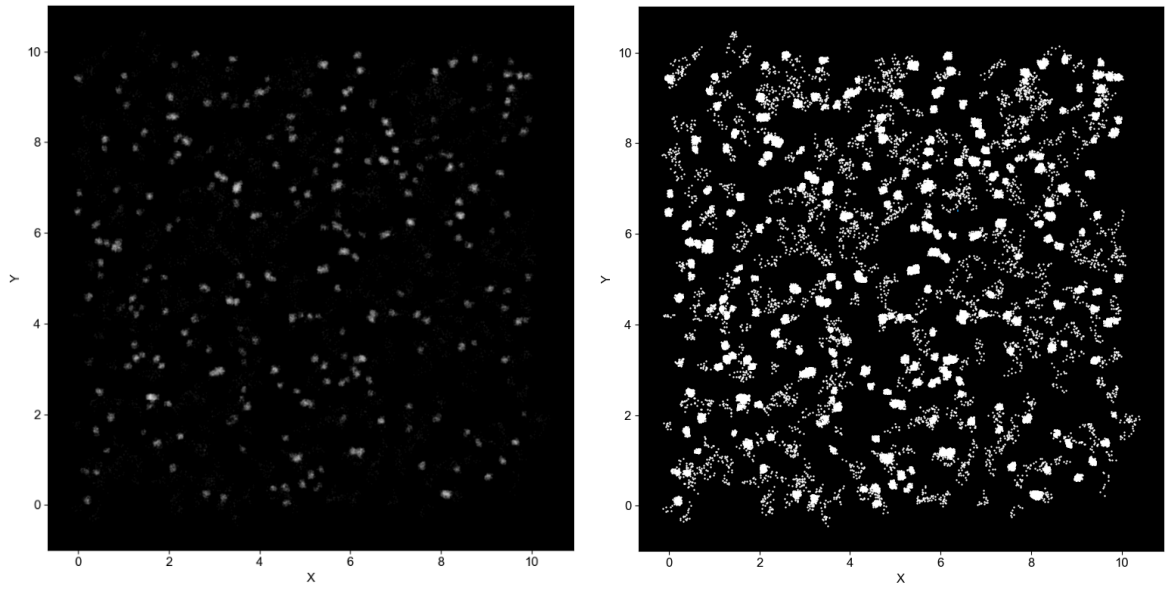
TRXYT files should contain no headers or comments, just the data:

```
1 9.0117 39.86 0.02
1 8.9603 39.837 0.04
1 9.093 39.958 0.06
1 9.0645 39.975 0.08
2 9.1191 39.932 0.1
2 18.9266 39.915 0.12
Etc
```



Clicking the **"Browse"** button will open a file selector which will allow you to choose TRXYT files (only). Once you've selected a file its path will be displayed, and you can then opt to change the length filters which by default screen out really short (largely uninformative) trajectories and really long ones (which are usually background noise).

In order to plot this raw data NASTIC needs to know what % of the raw detections to use (default Probability = 1.0 → 100%) and how opaque each detection should be on the plot. If you have high density data then too high probability and opacity values might result in a solid white plot which could make subsequent ROI assignment tricky. The below images are 0.05 and 1.0 opacity respectively.



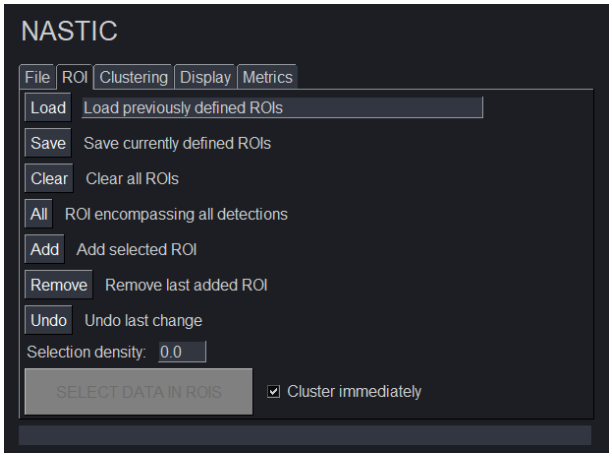
Clicking **"PLOT RAW DETECTIONS"** will plot the raw detections, and the program will automatically switch to the **ROI** tab. You won't be able to click **"PLOT RAW DETECTIONS"**

until you have actually specified an input file. In fact, all of the buttons in NASTIC will only be active at the appropriate stages of the analysis.

IMPORTANT NOTE: It appears that some Matlab processing of trajectory data converts trajectory numbers > 99999 into scientific notation with insufficient decimal points. eg 102103 to 1.0210e+05, 102104 to 1.0210e+05. This can cause multiple trajectories to be incorrectly merged into a single trajectory. For trajectories > 99999 NASTIC empirically determines whether detections are within 0.32u of each other, and assigns them into a single trajectory accordingly. For trajectories <99999 it honours the existing trajectory number.

IMPORTANT NOTE 2: Selecting a probability of < 1 will result in fewer detections being displayed, but internally NASTIC is still working with the entire set of detections and will select them according to your subsequent ROI. So if you load a dataset of 10000 detections comprising 1000 trajectories and select a probability of 0.5, only 5000 detections will be displayed. If you subsequently use an ROI encompassing all detections, NASTIC will then select the original 1000 trajectories.

ROI TAB

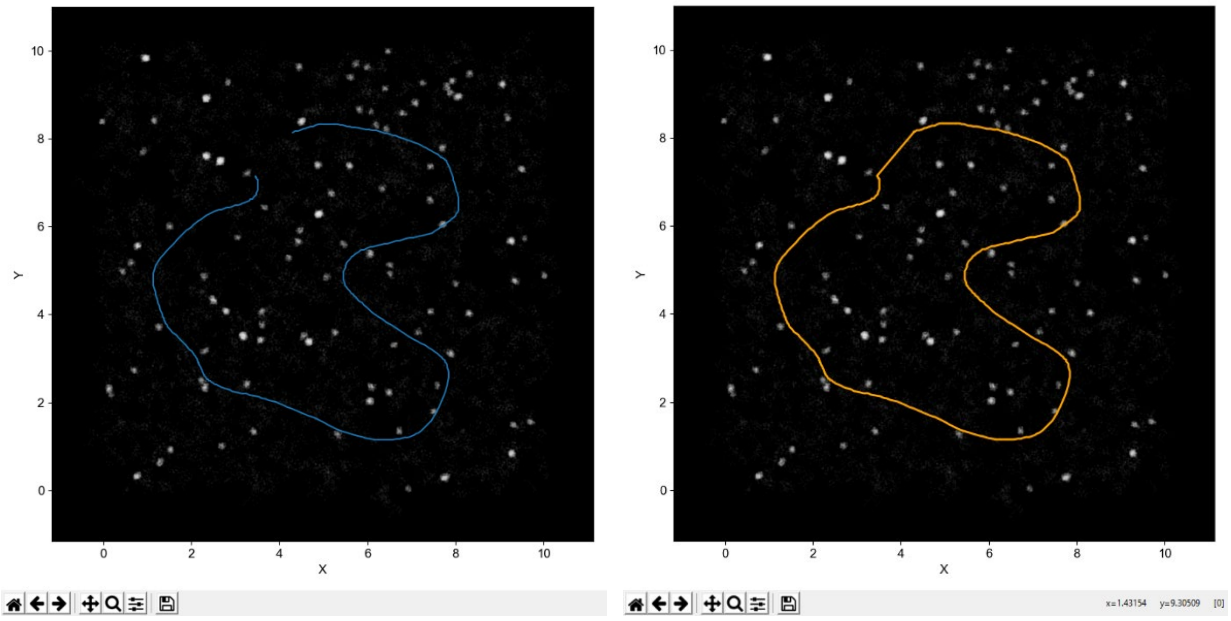


This tab allows you to select single or multiple regions of interest from the raw detections plot.

Clicking **"Load"** will allow you to load in saved ROIs from previous NASTIC analyses.

Clicking **"Save"** will allow you to save the current ROI(s) to a datestamped YYYYMMDD_roi_coordinate.tsv file in the same directory as NASTIC itself. If you are doing a complex series of ROIs, I recommend saving them as you go to prevent potential aggravation. ROIs will also be saved automatically when you save the analysis later on.

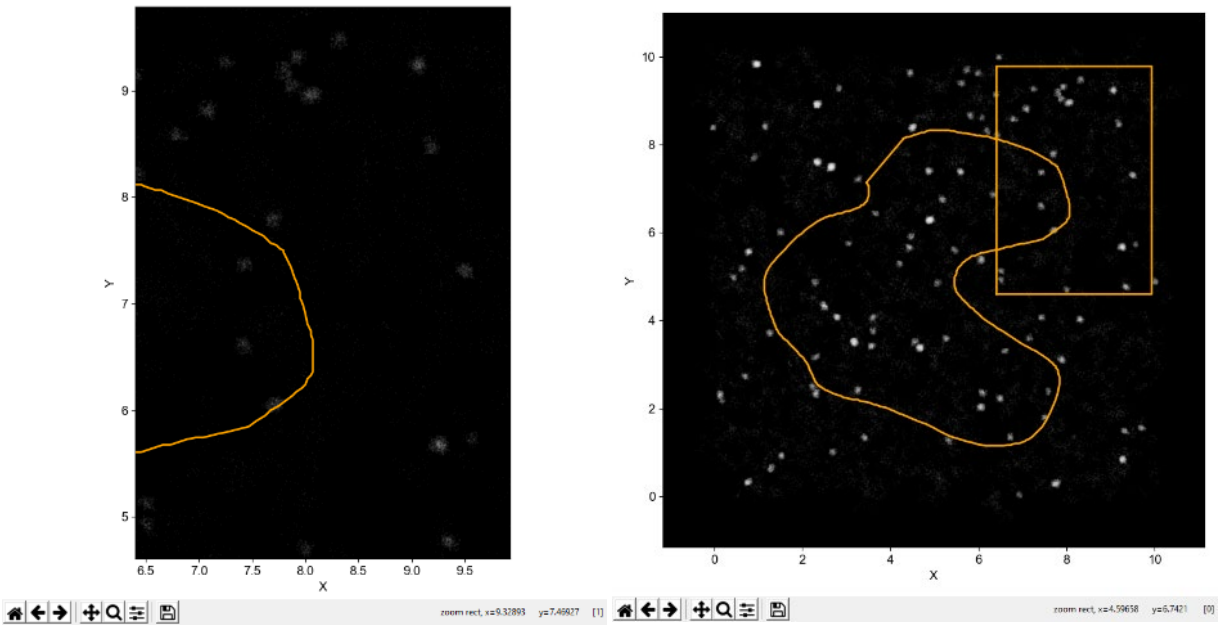
Defining your own ROIs can be done either using freehand drawing, or using a rectangular shape. You can add as many ROIs as you'd like, and it doesn't matter if they overlap as the trajectories will only be selected once in overlap areas.



NASTIC defaults to using the freehand selector. Just hold down the mouse and draw. You don't have to close the shape. Clicking the **"Add"** button will close the shape and add the ROI in yellow.

To switch to a rectangle selector, just click on the little magnifying glass icon on the bottom of the display window.

The mouse cursor will change to a crosshair. Hold the mouse to select the region you want, and when you release it the display will zoom in on the selected area. Click the **"Add"** button to add this zoomed area and return to the full sized image.



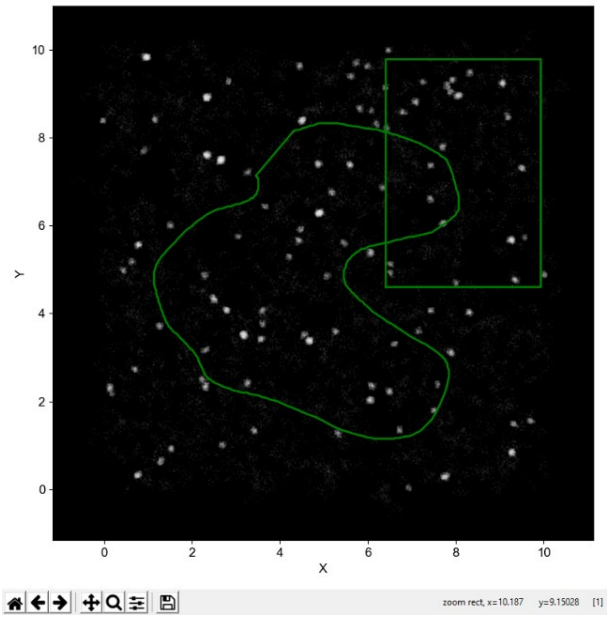
Clicking the **"Remove"** button will allow you to sequentially remove the last added ROI. You can click it more than once to remove more ROIs. This also applies to ROIs loaded in using the "Load" button.

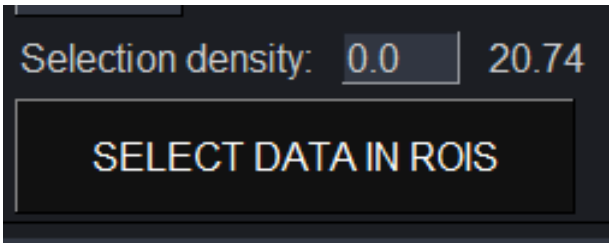
"Clear" will completely clear the ROIs.

"All" will create a single ROI which encompasses the full spatial extent of all the detections.

"Undo" will undo the last change you made to the ROIs. If you have 5 laboriously selected ROIs and accidentally press "Clear" or "All", you can click this button and get back to those 5 ROIs.

You can load a previous set of ROIs, and add more ROIs, etc etc etc. Basically there is a lot of flexibility to allow you to just analyse the areas you think are important. Once you've decided on the perfect set of ROIs then click the **"SELECT DATA IN ROIS"** button, which will now be active. If the "Cluster immediately" box is ticked, trajectories will be selected in ROIs and immediately clustered. If the box is unticked then the trajectories will just be selected. The ROIs will turn green to indicate this.





You'll see a figure next to the **Selection densit**y box, this is the number of selected trajectories per μm^2. You can adjust this down by entering the value you want in the Selection density box. Clicking **"SELECT DATA IN ROIS"** will reselect the trajectories and randomly drop some to bring the number down close to your entered value. Eg you could enter 15, and the system will return 14.89 trajectories/ μm$^{2.}$ This feature can be useful if analysing data from multiple acquisitions with varying trajectory density, as it can allow you to adjust all acquisitions to a common density.

If you enter 0 in the box then **"SELECT DATA IN ROIS"** will just go back to selecting everything.

NASTIC will automatically switch to the **Clustering** tab once trajectories have been selected.

Clustering Tab

The values on this tab are at the heart of R-tree based spatiotemporal clustering. By default NASTIC uses sensible values, which generally give good results with synthetic datasets which represent the kind of density and clustering typically seen in PC12 sptPALM experiments.

Acquisition time: The length of the acquisition in seconds. The program works internally using seconds, not frames.

Frame time: The time between each frame in sec. In our lab the data is typically acquired at 50Hz, so each frame is 0.02 s (20ms). A 320 s acquisition is thus 16000 of these frames.

Time threshold: This is the time "thickness" in seconds that defines whether trajectories are considered to be interacting in time. A time threshold of 20 sec means that two clusters are considered as clustered if they overlap in space, and are within 20 seconds of each other. Each trajectory will be examined to determine if there are other trajectories up to 10 seconds earlier, and up to 10 seconds later. If you set a time threshold of 640 seconds, then NASTIC will essentially only report spatial clustering (NASIC?) and will return clusters more or less the same as DBSCAN or Voronoi.

Radius factor: NASTIC does not use the convex hull of the trajectory, but rather calculates the ideal radius of the convex hull and builds a box based on this. Unless the trajectory moved within a perfectly circular area, this box does not fully represent the extent of the molecule. So we can expand the box to better account for this. A value of 1.2 works well for starters. You can take it up to 1.5, possibly even 2 if you have a lot of directed trajectories or low density trajectories. But this will of course potentially start showing more spurious overlap. Alternatively, if you are using very high density data you can use a radius factor of <1. If your data is really high density you should consider either changing the selection density in the ROI tab, or perhaps using **segNASTIC**, which clusters based on trajectory segment overlap rather than whole trajectory bounding box overlap.

Cluster threshold: How many trajectories have to overlap in order to be considered a cluster. 3 is a good start. In high density data you may want to consider increasing this. Before you start complaining, be aware that a cluster threshold like this is at the core of DBSCAN and Voronoi too.

Cluster size screen: The general consensus is that protein nanoclusters are usually less than 0.1µm in radius. NASTIC clusters much larger than this are more likely to be random overlap of higher mobility trajectories. These can be screened out by setting this value appropriately. If you set it much below 0.15µm you might risk throwing out valid clusters, so use this wisely.

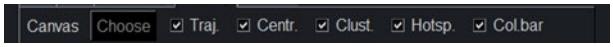
MSD screen: This setting may be useful to screen out high mobility background trajectories which could otherwise contribute to large diffuse clusters. Ticking this box will cause NASTIC to measure the MSD for all trajectories, and establish the average MSD. It will then reject those trajectories whose MSD is greater than the average. Use this carefully – for datasets with lots of low mobility trajectories the average MSD will be lower, and you might end up not using trajectories with low mobility but still higher than the average.

“CLUSTER SELECTED DATA” will apply your values and use spatiotemporal indexing clustering on the selected trajectories. If **“Plot immediately”** is ticked then the clusters will be automatically displayed after the analysis is complete.

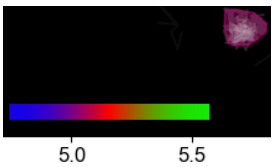
Display Tab

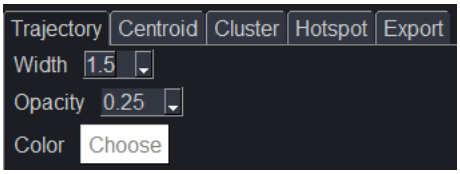
NASTIC gives you a large amount of flexibility to display your clustered data, and therefore has a few extra subtabs that you can access. At any time, you can change various parameters and hit the **"PLOT CLUSTERED DATA"** button to update the display.

You can use the controls on the bottom  of the display window to zoom in (make sure you've selected the magnifying glass and changed the cursor to a crosshair), the arrows to move back or forward through the various zoom, and the home icon to go back to full view.



Choose the background colour of the plot, and whether you want trajectories, centroids, clusters and hotspots displayed. The **"colorbar"** checkbox will plot a colorbar onto the display, where blue = earlier in the acquisition and green = later in the acquisition.

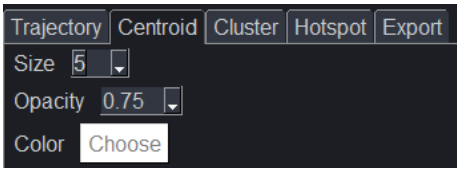




Width: how thick the trajectory plot lines are

Opacity: 0 = completely transparent, 1 = solid

Color: all trajectories will be plotted in this color.

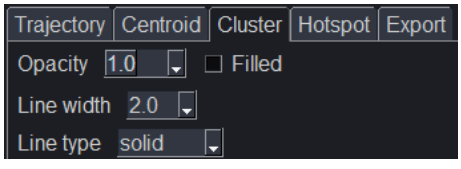


Size: how large the trajectory centroid will appear

Opacity: 0 = completely transparent, 1 = solid

Color: All centroids will have the same color, regardless of whether they are in clusters or not.

Each cluster represents a convex hull boundary around the greatest extend of the component detections. The color of the cluster represents its approximate time in the acquisition



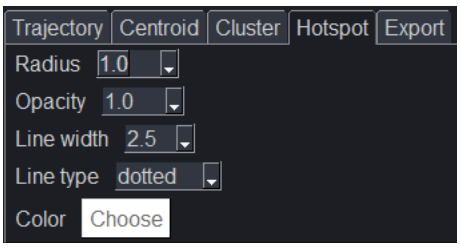
Opacity: 0 = completely transparent, 1 = solid

Filled: The cluster will be filled with solid color at half the opacity.

Line width: how thick the trajectory border lines are

Line type: Style of the trajectory border line

Hotspots represent regions where clusters occupy the same spatial region, but at different times. This tab will allow plotting a convex hull encompassing the detections of all overlapping clusters.



Radius: NASTIC calculates the average radius of all clusters. Setting radius to 1 means that two clusters whose centroids are closer than the average radius will be considered as a hotspot. <1 = clusters have to be closer than the average radius, >1 = clusters can be further apart than the average radius

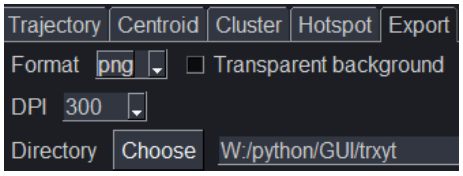
Opacity: 0 = completely transparent, 1 = solid

Line width: thickness of the hotspot border line

Line type: Style of the hotspot border line

Color: all hotspots will be plotted in this color.

At any time you can export a high quality file representing the contents of the main plot window, by clicking the **"SAVE PLOT"** button. These images may be more suitable for publication etc.



Format: eps, pdf, png, ps, svg. Depending on what software you want to use for figure wrangling, png is a common lossless bitmap format which will open in anything.

Transparent background: The image will ignore the background canvas color and will be output with a transparent background. This could come in handy if you want to make a fancy image with a gradient background for instance.

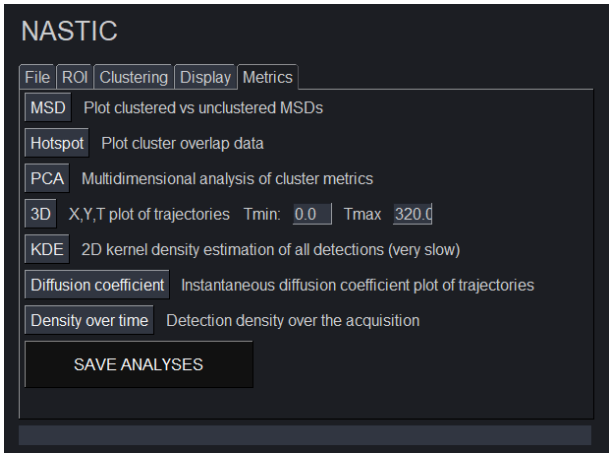
DPI: Dots per inch. 50 = lower resolution, 1200 = very high resolution but very large file size

Directory: Where to save the image.

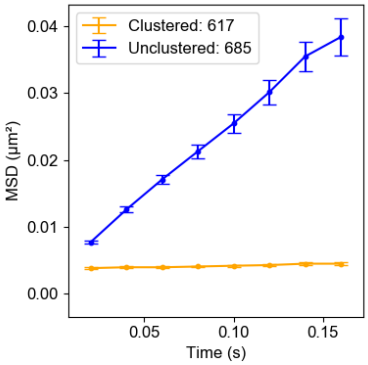
Rather than use the mouse to zoom in on a region, you can specify it exactly by entering the corner co-ordinates for the zoom box. This can come in handy if you are eg comparing NASTIC, DBSCAN and VORONOI (the GUIs of these programs all allow you to zoom like this) of a particular dataset and want to visualise how the clusters differ in the same area. Hit **"PLOT CLUSTERED DATA"** to update the image once you've entered the co-ordinates. You can enter 0 in all boxes to revert to the full plot of all selected trajectories.

"Metrics immediately" will automatically switch to the Metrics tab once the plot has completed. This can be useful if you are analysing large number of TRXYTs all with the same parameters, and don't care much about the plotted cluster image. Ticking **"Cluster immediately" + "Plot immediately" + "Metrics immediately"** and clicking **"SELECT DATA IN ROIS"** in the ROI tab will selected trajectories, cluster them, plot the output, and immediately switch to the Metrics tab so that you can save the analyses or examine them further.

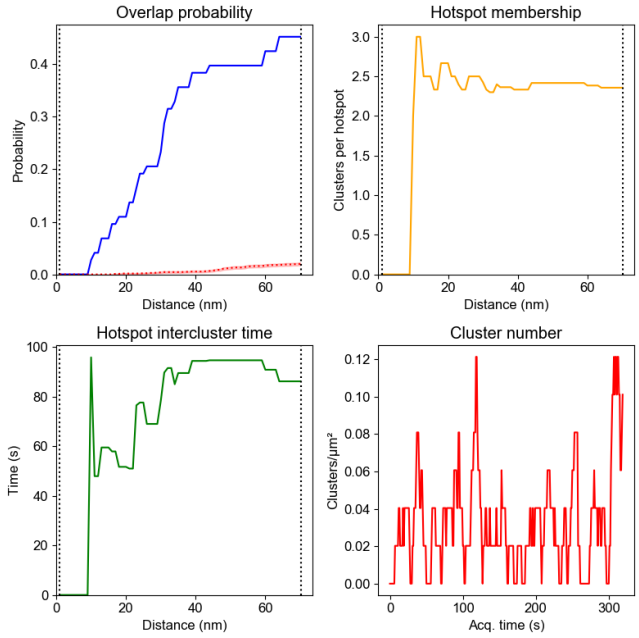
Metrics Tab



This tab will allow you to examine various cluster metrics to allow further assessment of the data. Because NASTIC does temporal clustering, many of these metrics are not available when using DBSCAN or Voronoi.

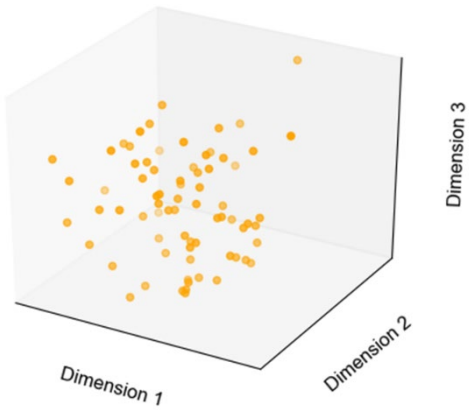


MSD: Will plot Mean square displacement vs time for clustered trajectories and unclustered trajectories. Ideally we want to see clustered detections with lower mobility as per the picture on the right.

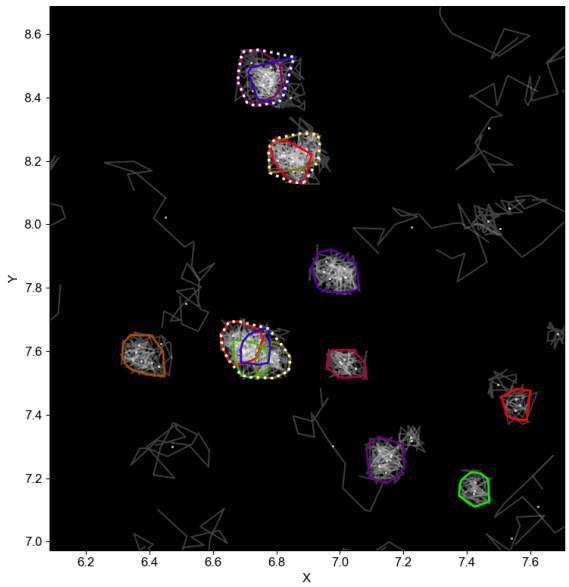


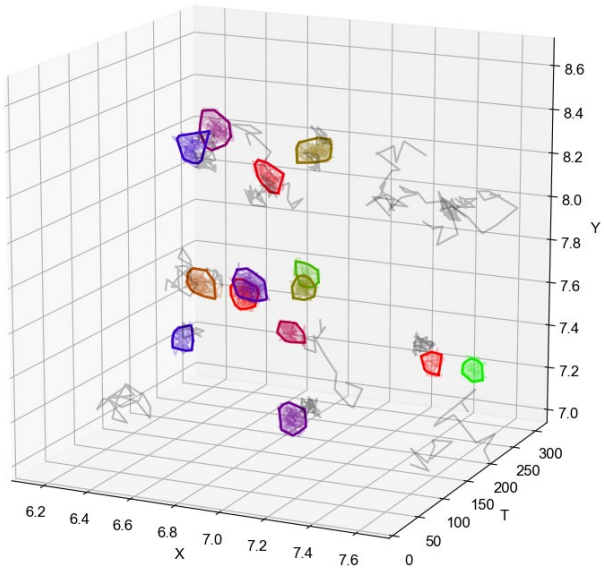
Hotspot: Because we can get spatial and temporal clustering information using NASTIC, we are able to determine the presence of "hotspots" where clusters repeatedly form and dissociate on the same region of the membrane. Overlap probability indicates the likelihood of this. Firstly, the average radius of all clusters is determined. Then, the spatial centroid of each cluster is determined. DBSCAN is then performed multiple times using an epsilon between 0 and the average cluster radius. At each epsilon the number of centroids with another centroid within epsilon is calculated, and converted to a probability such that 1 = all centroids have a neighbor within epsilon. In the figure on the left approximately 45% of all clusters have another cluster within 70nm (the average cluster radius), and ~ 20% have a cluster within ½ the average radius. Remember, they may overlap in space, but occupy a different point in time. Additional information can be extracted from the DBSCAN clusters, such as the number of clusters in each hotspot, and the time between these clusters. The cluster number plot allows you to determine whether the overall number of clusters at a given point in time changed over the course of the acquisition. This could be useful in a system where cells were stimulated ½ way through an acquisition for instance.

PCA: Each cluster is assigned multiple metrics such as size, number of trajectories, average MSD, lifetime etc etc. Dimensionality reduction using principal component analysis (PCA) allows you to determine whether there are subpopulations of clusters with distinctly different metrics. Usually this is not the case, but one could imagine a situation where an analysis of a cell with two distinct membrane regions resulted in a differently behaving clusters. These might show up as distinct groups of dots on the PCA plot.



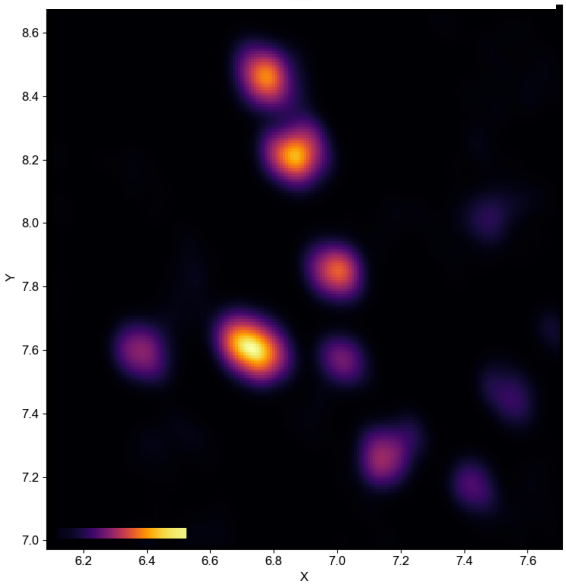
3D: NASTIC works by treating superres data as 3 dimensional [x, y, time]. In the final 2D plots, the time dimension is represented as color. If you click on the "3D" the trajectories on the main display plot will be rendered into 3D space, which will not only allow you to get a better feel for how the clustering picked up things in space and time, but might also make for some pretty pictures for publication. By default the 3D plot displays trajectories over the full acquisition time window. **Tmin** and **Tmax** can be used to adjust the time window to display, allowing you to zoom in on a particular time period.



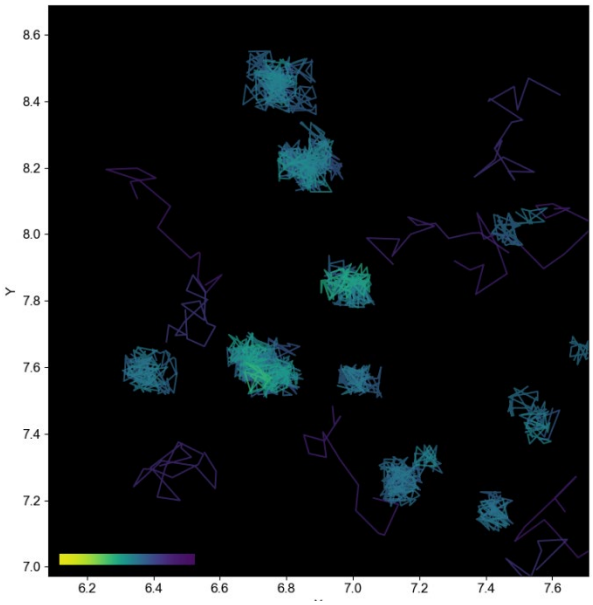


You can rotate the 3D plot by holding the left mouse button and moving the mouse appropriately. You can zoom in or out by holding the right mouse button and moving the mouse up or down.

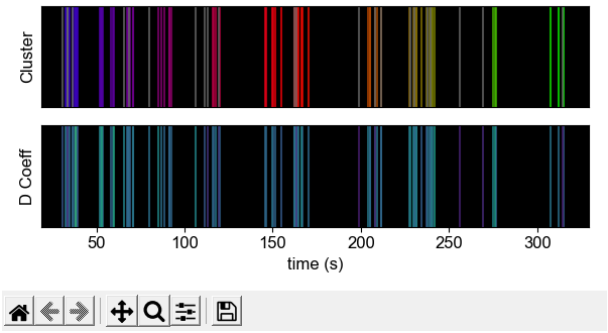
KDE: Kernel density estimation is basically a 2D histogram of the density of the detections, which shows regions of higher density as brighter color. Not only does it look pretty but it can be very useful for determining whether your clusters correspond to regions of higher detection density. This is not always the case however, since it's possible for a single trajectory with many detections to show up on KDE, but not be selected for NASTIC. KDE



analysis can be very slow with lots of detections so please don't use it to analyse your entire image - start with a 2 µm x 2 µm region and see how long that takes before increasing things.

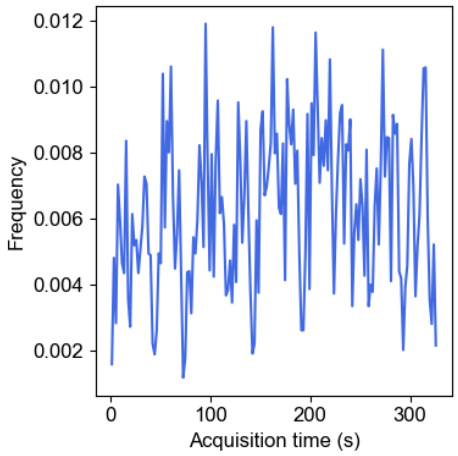


Diffusion coefficient: NASTIC calculates the instantaneous diffusion coefficient as the gradient of the first 4 steps of the MSD for each trajectory. Plotting this can help to visualise whether clusters consist of lower mobility (low diff coeff) trajectories against a background of higher mobility unclustered detections.



The data is also plotted as 1D, so that you can determine whether trajectories with low diffusion coefficient are present in clusters.

Density / time :



Divides the acquisition into 2 second bins (eg for a 320 sec aqcquisition create 160 bins) and counts the trajectories in each bin. This allows rapid determination of whether the number of molecular detections remains constant over the analysis. For acquisitions where the conditions change at some point (eg stimulation of PC12 cells), this might result in recruitment of additonal molecules to the plasma membrane, which can be visualised with this metric.

"SAVE ANALYSES" will save any open plot windows, a plot of the raw detections and ROIs, the ROI coordinates themselves, and a TSV file, **metrics.tsv,** containing the various analysis metrics to a datestamped folder in the same directory as the TRXYT you analysed. The file can be opened in any text editor or Excel.

NANOSCALE SPATIO TEMPORAL INDEXING CLUSTERING - Tristan Wallis t.wallis@uq.edu.au						
TRAJECTORY FILE:	W:/python/GUI/trxyt/synthetic_col1_20201022-095432.trxyt					
ANALYSED:	20211215-133708					
TRAJECTORY LENGTH CUTOFFS (steps):	8 - 100					
TIME THRESHOLD (s):	20					
CLUSTER THRESHOLD:	3					
RADIUS FACTOR:	1.2					
MSD FILTER THRESHOLD (um^2):	None					
CLUSTER MAX RADIUS (um):	0.25					
SELECTION AREA (um^2):	114.9425					
SELECTED TRAJECTORIES:	2023					
CLUSTERED TRAJECTORIES:	1225					
UNCLUSTERED TRAJECTORIES:	798					
TOTAL CLUSTERS:	229					
CLUSTERED TRAJECTORIES AVERAGE INSTANTANEOUS DIFFUSION COEFFICIENT (um^2/s):	0.000922					
UNCLUSTERED TRAJECTORIES AVERAGE INSTANTANEOUS DIFFUSION COEFFICIENT (um^2/s):	0.165133					
HOTSPOTS (CLUSTER SPATIAL OVERLAP AT 1/2 AVERAGE RADIUS):	9					
TOTAL CLUSTERS IN HOTSPOTS:	20					
AVERAGE CLUSTERS PER HOTSPOT:	2.222222					
PERCENTAGE OF CLUSTERS IN HOTSPOTS:	8.734					

The MSD curve data is also tabulated should the user wish to plot in Prism etc

MSD CURVE DATA:							
TIME (S):	0.02	0.04	0.06	0.08	0.1	0.12	0.14
UNCLUST MSD (um^2):	0.004916261	0.008227591	0.011649468	0.014824236	0.018202581	0.021589937	0.024744567
UNCLUST SEM:	6.92E-05	0.000200777	0.000353356	0.00052163	0.000711869	0.000912965	0.001127653
CLUST MSD (um^2):	0.003385254	0.00337221	0.003404058	0.003440583	0.003448916	0.00351042	0.003503516
CLUST SEM:	2.34E-05	2.71E-05	3.59E-05	4.54E-05	6.01E-05	7.29E-05	8.62E-05
ALL MSD (um^2):	0.003989181	0.005287481	0.006656573	0.00793102	0.009268701	0.010642133	0.011882339
ALL SEM:	3.50E-05	9.66E-05	0.000167116	0.000241649	0.00032539	0.000412598	0.000503848

The remainder of the file contains the individual and averaged metrics for the clusters identified during the analysis

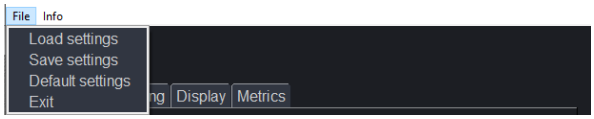
INDIVIDUAL CLUSTER METRICS:								
CLUSTER	MEMBERSHIP	LIFETIME (s)	AVG MSD (um^2	AREA (um^2)	RADIUS (um)	DENSITY (traj/ur	RATE (traj/sec)	AVG TIME (s)
1	4	4.34	0.003045833	0.014944422	0.068970699	267.658392	0.921658986	273.22
2	4	13.39	0.002938636	0.010071418	0.056620067	397.1635301	0.298730396	118.3475
3	3	3.36	0.002403054	0.004695381	0.038659878	638.925795	0.892857143	62.22666667
4	5	3.72	0.003565336	0.013906241	0.066531901	359.5508055	1.344086022	54.134
5	7	5.05	0.003329512	0.016727002	0.072968281	418.4850396	1.386138614	253.2128571
6	3	6.58	0.003079342	0.007711196	0.049543415	389.044693	0.455927052	65.48333333
7	3	7.96	0.002638875	0.007332106	0.048310265	409.1594154	0.376884422	80.73333333
8	8	9.19	0.003176084	0.018912975	0.077589863	422.9900427	0.870511425	162.43625
9	4	9.16	0.003954768	0.033445685	0.103179901	119.5968943	0.436681223	63.015
10	5	9.61	0.003301422	0.014065784	0.066912467	355.4725319	0.520291363	54.892
11	3	1.61	0.00347634	0.007758419	0.049694882	386.676738	1.863354037	221.1766667
etc	etc	etc	etc	etc	etc	etc	etc	etc
AVG	5.349344978	6.674366812	0.00337696	0.013613329	0.064836745	401.8836942	0.985935471	175.1764285
SEM	0.12320699	0.155182697	2.44E-05	0.000322412	0.000751864	5.646237537	0.082770292	5.97882413

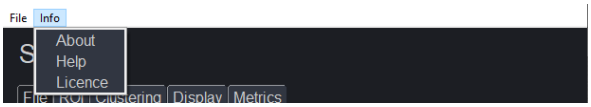
This data can be plotted as is, but the real power comes from being able to combine these metrics files from larger numbers of analyses across multiple conditions. This is handled by **NASTIC WRANGLER**, which is detailed in a separate document.

Once you have finished analysing a given TRXYT, you can flick back to the **File** tab and load another one. This will automatically close all other open windows except the main display window. You should never need to close this window, but if you do the program will let you know all about it and will open a fresh window and reset the analysis.

Menus

In addition to the tab and button functionality of NASTIC, there are some menu items you should get acquainted with.





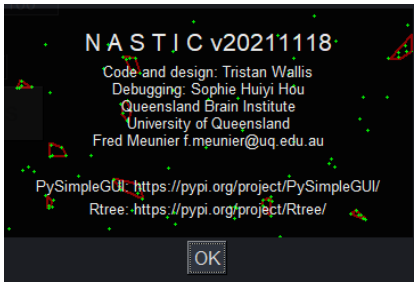
The first time you run NASTIC it will create a file called **NASTIC_gui.defaults** which contains the default values for the various steps of the analysis. You can load it into a text editor or spreadsheet to view it if you'd like.

Trajectory probability	1
Raw trajectory detection plot opacity	0.05
Selection density	0
Trajectory minimum length	8
Trajectory maximum length	100
Acquisition time (s)	320
Time threshold (s)	20
Radius factor	1.2
Cluster threshold	3
Canvas color	black
Plot trajectories	TRUE
Plot centroids	FALSE
Plot clusters	TRUE
Plot colorbar	TRUE
Trajectory line width	1.5
Trajectory line color	white
Trajectory line opacity	0.25
Centroid size	5
Centroid color	white
Centroid opacity	0.75
Cluster fill	FALSE
Cluster line width	1.5
Cluster line opacity	1
Cluster line type	solid
Plot save format	png
Plot save dpi	300
Plot background transparent	FALSE
Auto cluster	TRUE
Auto plot	TRUE
Cluster size screen	0.15
Auto metric	FALSE
MSD filter	FALSE

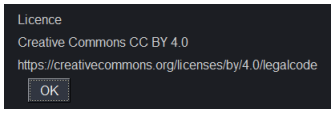
Nearly every value that you can change in NASTIC can be saved to this file. So if you arrive at a set of values that work well for a given set of data, **"File >> Save settings"** will save them to NASTIC_gui.defaults. The next time you start NASTIC, they'll be loaded in as the defaults. If you get too far out of whack, **"File >> Default settings"** will restore the defaults. But remember if you don't save them, the next time you load NASTIC you'll get your messed up values! Similarly if you started with your preferred values, and changed multiple settings, **"File >> Load settings"** will restore your last saved settings.

IMPORTANT NOTE: Periodically NASTIC is updated with new functionality that requires a new **nastic_gui.defaults** file. Newer versions of NASTIC may crash if loading an older defaults file. You are encouraged to delete your existing defaults file and run the new version of NASTIC to recreate the correct file with default settings.

Clicking **"About"** will pop up this glorious little window with a real time molecular clustering simulation. Warning this is mesmerising to look at and could waste several minutes of your life. Why did I do this? Because I could.



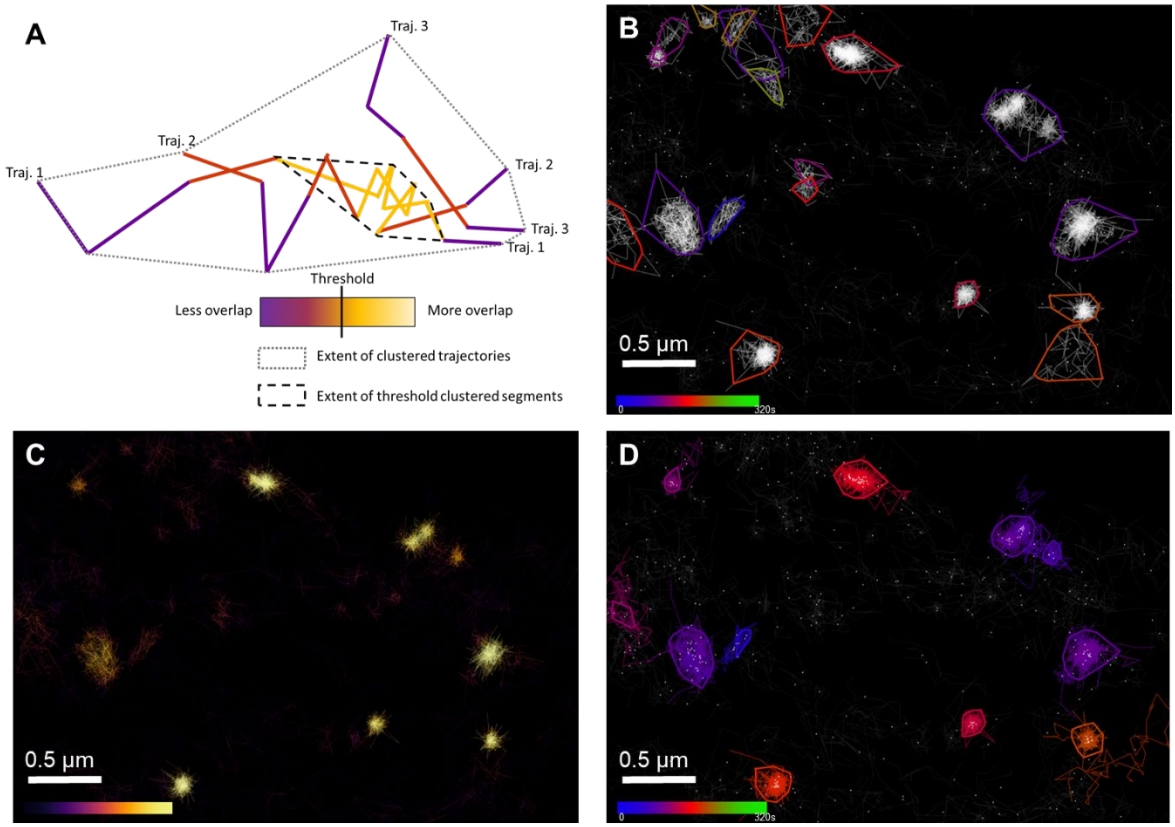
"Help" does nothing much at this stage, eventually it will link to a version of this document.



"Licence" reiterates the CC BY 4.0 concept embraced by NASTIC and allows the user to find out more about this powerful open source licence.

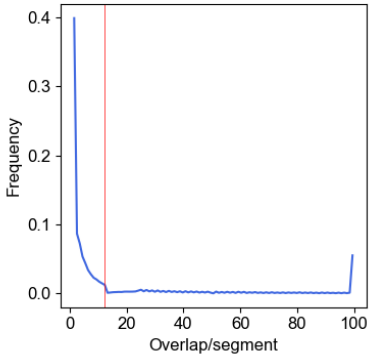
Appendix 1: Segment NASTIC

NASTIC works by establishing whether the bounding boxes of trajectories overlap in space and time. If trajectory density is very high it's possible that high mobility trajectories might still overlap, leading to spurious clustering.



NASTIC of trajectory segments (segNASTIC). (A) Schematic representation of trajectory segment thresholding, based on overlap with segments from other trajectories. (**B)** Syntaxin1a-EGFP imaged by uPAINT using Atto-647-labelled anti-GFP nanobodies in PC12 cells. Spatiotemporal clusters identified using spatiotemporal indexing of trajectory bounding boxes using ***r*** = 1.2 and ***t*** = 20 s. Each coloured cluster boundary represents the convex hull of the detections belonging to all trajectories in the cluster (**C)** Pseudo-density map of trajectory segment overlap, with each trajectory coloured according to the number of overlaps with other trajectory segments, as determined by spatiotemporal indexing of segment bounding boxes. (**D)** Spatiotemporal clusters identified using thresholded segments ***t*** = 20 s. Each coloured cluster represents the convex hull of detections belonging to the clustered segments. All trajectories containing clustered segments are shown in the same colour as the cluster.

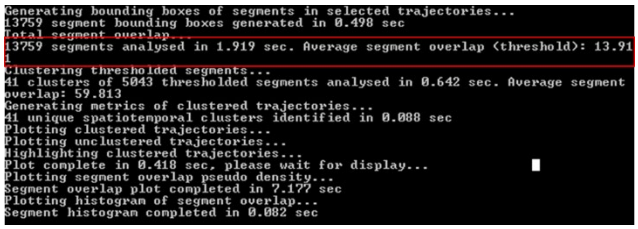
Segment NASTIC attempts to get around this by clustering based on trajectory segment overlap. Trajectory segments which overlap with many other segments are more likely to represent areas of true molecular interaction. As the above figure shows, it more accurately returns areas of interaction in high density uPAINT data. The segment version of NASTIC is found as segNASTIC_gui.py, and double clicking it loads an interface similar to NASTIC.



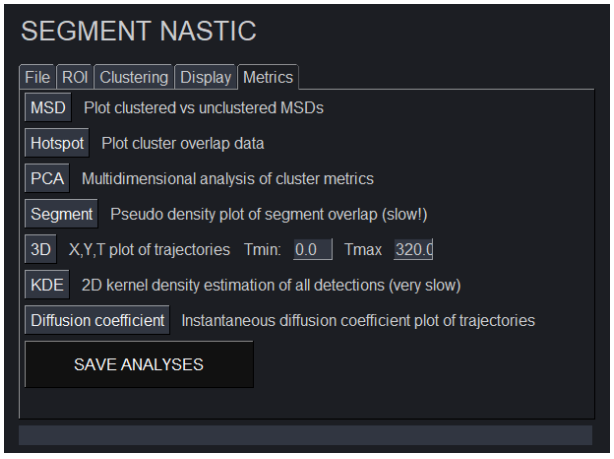
The clustering parameters of segNASTIC are slightly different to allow for the fact that it is clustering based on segment overlap. Basically segNASTIC analyses the overlap of all trajectory segments, and then works out what the average overlap is. Segments with overlap greater than this are considered as potentially clustered. It turns out that the overage overlap also corresponds to the inflection point on the distribution of overlap (red line)

Segment threshold: Clusters must contain at least this many segments whose overlap is greater than the threshold. This is roughly equivalent to the "Cluster threshold" setting in normal NASTIC. The higher it's set, the more stringent the clustering will be.

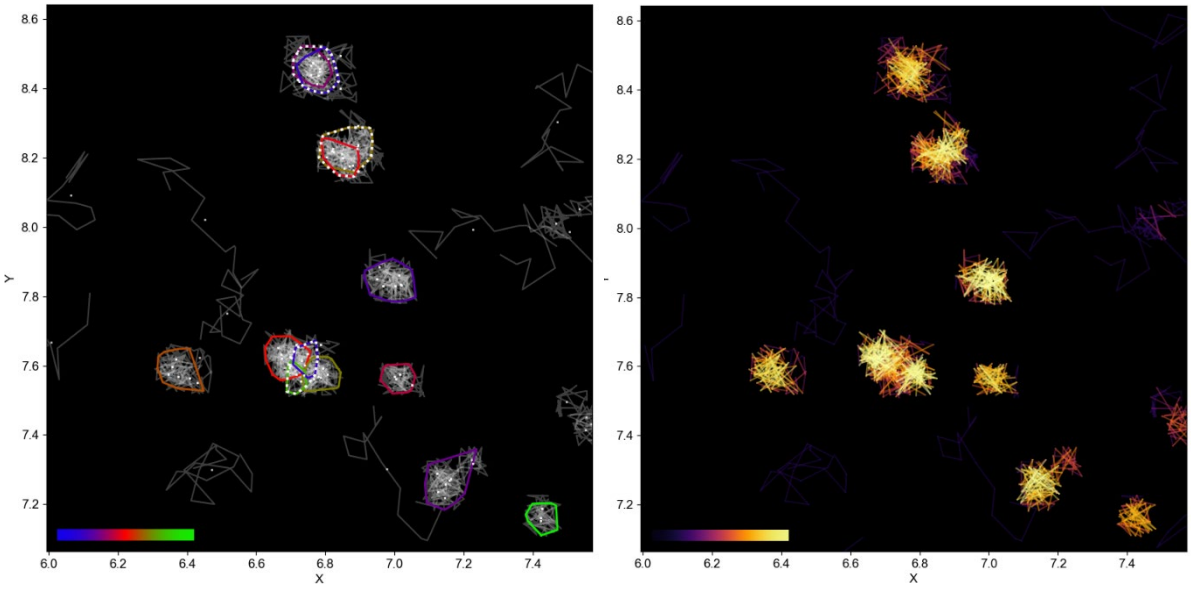
Overlap threshold override: Yes, too many occurrences of "over". Basically if this is set to 0, then the system will use the average overlap as a threshold. Otherwise, you can make it more or less stringent by entering your own value here. Obviously you can't know what the average overlap is until you've already performed an analysis using *overlap threshold override = 0*, in which case the threshold will be displayed in the shell:



In this case, you might want more stringent clustering and you could increase the threshold from 13.91 up to eg 20.



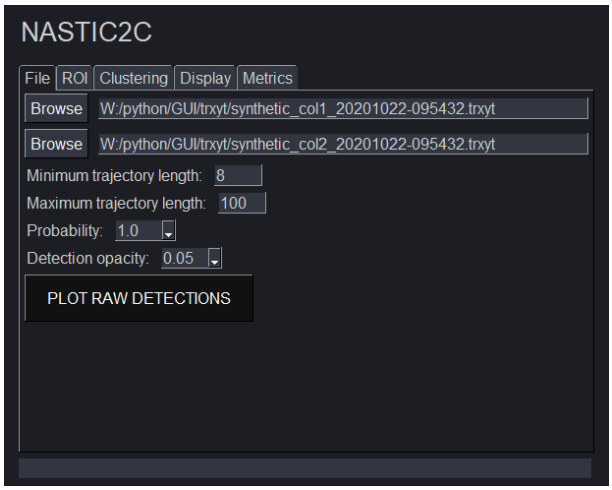
Once you've performed the analysis, the metrics are largely the same except for the inclusion of the "Segment" option, which allows you to plot the segment overlap density and the segment overlap distribution (as shown above).


8.6
8.4
8.2
8.0
7.8
7.6
7.4
7.2
Y
6.0
6.2
6.4
6.6
6.8
7.0
7.2
7.4
X

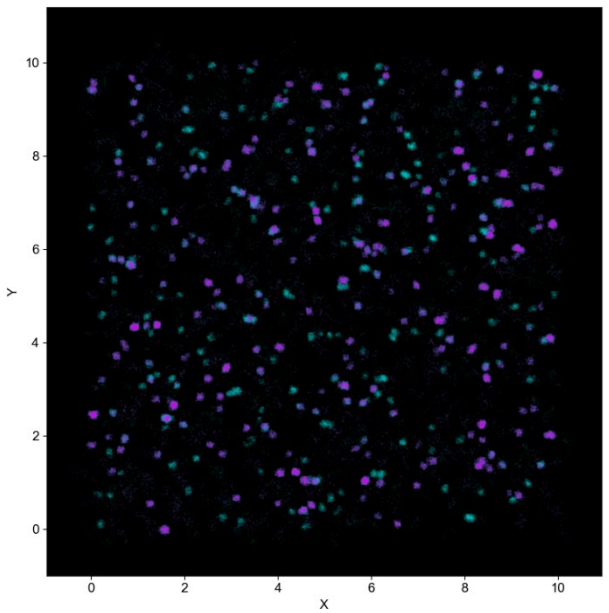
Appendix 2: Two color NASTIC

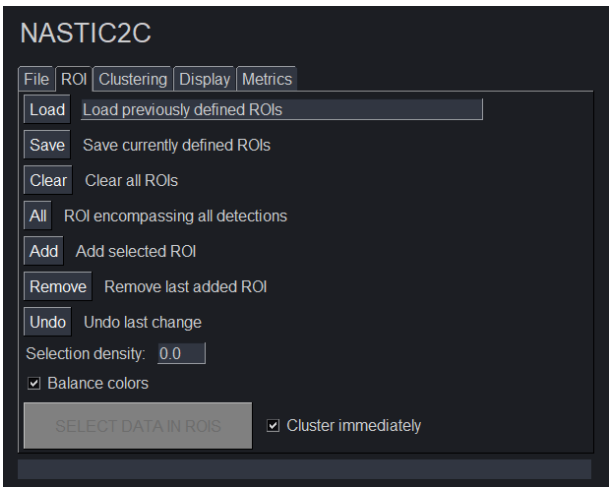
The spatiotemporal indexing approach used in NASTIC can readily be extended to two color analyses, in which two target molecules are independently labelled with different fluorophores and SMLM data simultaneously acquired at two different wavelengths to generate two datasets. Two-color NASTIC (NASTIC2C) is performed by combining the datasets and establishing spatiotemporal overlap of all trajectories. The type (color) information for each trajectory in the combined datasets is retained, which allows calculation of the relative contribution of each molecular type to the resulting clusters, and is used to inform the graphic output. In addition to identifying clusters of each molecular type, NASTIC2C can identify mixed clusters with varying ratios of molecule 1 and 2, and resolve areas of spatiotemporal overlap. A segment version of twocolor NASTIC (segNASTIC2C) is also available, which may be useful for high density two color data.

The two color versions of NASTIC And segNASTIC have similar interfaces to the single color versions, with allowances for the fact that two datasets are used.



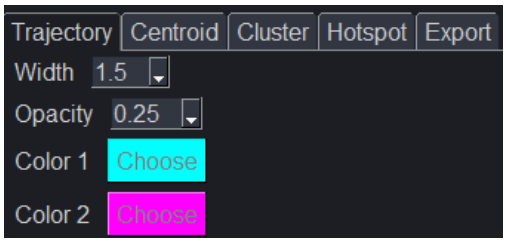
Raw detections are plotted in two colors (cyan and magenta by default) and ROIs are selected using the same mechanism.



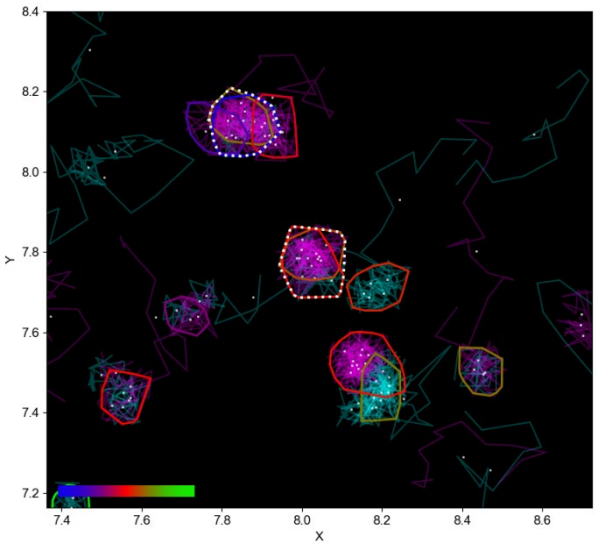


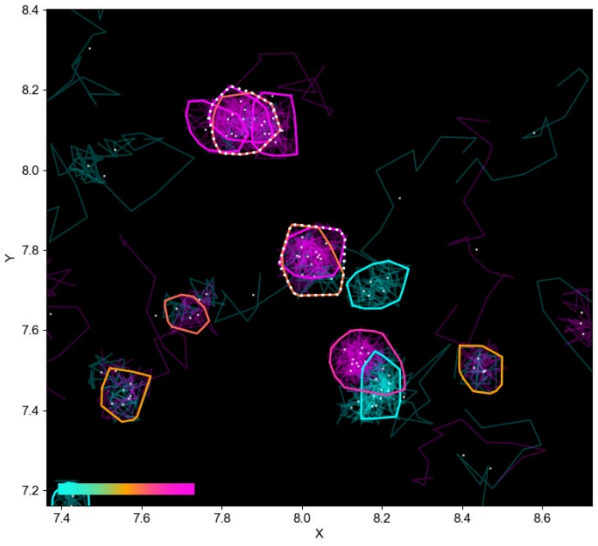
“Balance colors” will randomly remove trajectories from the higher density dataset so that both datasets contain the same number of trajectories in the ROI/s

The display tab allows you to select two colors for trajectory display. These colours will be used for raw detection display also.

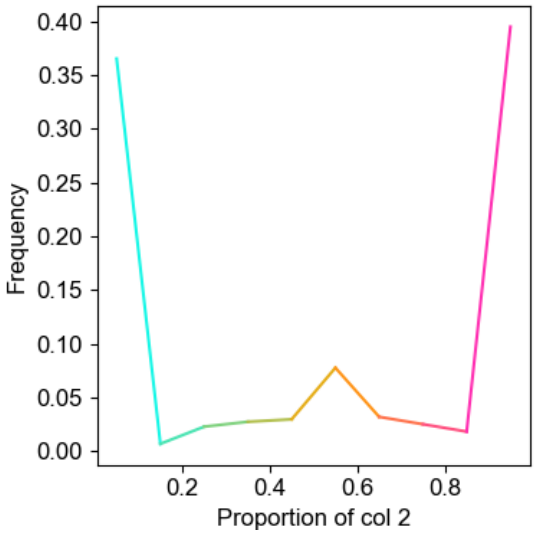


Clusters may be displayed either according to their average time in the acquisition (as for NASTIC and segNASTIC) or according to the relative ratios of molecule 1 and molecule 2.





The metrics returned are largely the same, with the addition of a **"2 color metrics"** button which will plot the distribution of color ratio across all detected clusters.



FAQ

Q: Why did you even bother when there are other clustering programs available?

A: Because our laboratory is particularly interested in the temporal aspects of nanomolecular clustering and these are difficult to extract with current solutions

Q: Why won't NASTIC even start when I double click it?

A: Have you or your IT support installed Python 3.8+, and assigned the correct PATH so the system can find it?

Q: Why does NASTIC crash with missing module errors?

A: Have you correctly installed the required modules using PIP?

Q: I updated NASTIC, why is it now crashing with weird variable errors?

A: The new version probably doesn't like your existing **nastic_gui.defaults** file. Delete it and try again.

Q: Why does NASTIC run slowly on my shiny new Mac?

A: Because Macs and Python don't always play together properly, particularly with GUI Python programs.

Q: Why won't NASTIC run on our server?

A: Because it needs access to a physical computer to do its multicore processing

Q: Why does NASTIC occasionally crash after I've analysed many files?

A: Because I'm a second rate programmer. There, I came out and said it. If you're a real programmer and want to help optimise NASTIC, please do.

Q: What's with the stupid open source licence?

A: Open source licences actually empower both the author and the user, unlike most commercial licences where the vendor holds all the cards. You can install NASTIC on as many computers as you want, give copies to your friends, sell it, modify it, collaborate with me and others to improve it, or even release your own forked version of the code with a better name. What you can't do is release a version without acknowledging the author of the original concept and code (me), and you can't release your version under a more restrictive licence. Sounds fair to me.

Q: How do I report bugs to you?

A: Firstly, thanks for wanting to do so. Please let me know:

- Your operating system
- Your Python version
- The NASTIC version
- The nature of the bug
- Steps taken to reproduce the bug (if possible)