

NASTIC WRANGLER USER MANUAL

NASTIC Wrangler

Load Files Plot Data

Condition 1

Browse Select condition 1 directory

Find files Files to include:

Condition 1 File names

Condition 2

Browse Select condition 2 directory

Find files Files to include:

Condition 2 File names

LOAD DATA

NASTIC Wrangler

Load Files Plot Data

Condition 1

Shorthand name to show on plot: CONT

Select color Condition 1

Condition 2

Shorthand name to show on plot: STIM

Select color Condition 2

☒ Outlier Removal - median filtering outlier reduction for aggregate data

Data to Plot:

☒ Aggregate Data - aggregates individual cluster metrics across all samples (N = total # of clusters)

☒ Average Data - averages cluster metrics for each sample (N = # of samples)

☒ 2D PCA analysis - determine the overall relationships between samples

PLOT DATA

Tristan Wallis t.wallis@uq.edu.au

Alex McCann a.mccann2@uq.edu.au

Single Molecule Neuroscience Laboratory (Fred Meunier)

Queensland Brain Institute

The University of Queensland, Australia

Introduction

The NASTIC family of programs output a large amount of metrics related to spatiotemporal clustering in super-resolution datasets. While these are useful and edifying by themselves, the real point of the exercise is to be able to establish whether these metrics change between experimental conditions (for example stimulated vs unstimulated cells). This can't be done within NASTIC itself, which only works on a single dataset at a time. Instead, the comparative stuff is handed off to NASTIC Wrangler, rocking a simple GUI which allows you to load the multiple NASTIC-generated metrics files from two experimental conditions and perform comparative analyses on the metrics.

Primitive versions of NASTIC Wrangler were written by Tristan Wallis as the original NASTIC suite matured. To coincide with the publication of NASTIC in Nature Communications, the comparative analysis routines were overhauled, and the NASTIC Wrangler GUI was largely rewritten by Alex McCann. As for the rest of NASTIC, this software is released under a Creative Commons licence.

Computer requirements

NASTIC is a Python script and requires Python 3.8 or later, and a number of python modules to run. If you've got as far as using NASTIC Wrangler then I can safely assume you downloaded the NASTIC suite from GitHub (which includes NASTIC Wrangler), and have Python and required modules installed for NASTIC. If not, please use the following command line to install them:

```
python -m pip install colorama matplotlib matplotlib-venn numpy pandas pillow  
freesimplegui rtree scikit-learn scipy seaborn statsmodels
```

```
C:\Users\uqamcc11>python -m pip install colorama matplotlib matplotlib-venn numpy  
pandas pillow freesimplegui rtree scikit-learn scipy seaborn statsmodels
```

At the time of writing, we used the below Python module versions:

scipy	v1.13.1
numpy	v1.23.2
matplotlib	v3.8.4
matplotlib-venn	v0.11.7
freesimplegui	v5.2.0.post1
rtree	v1.0.0
scikit-learn	v1.1.2
statsmodels	v0.13.2
colorama	v0.4.6
pandas	v1.4.4
pillow	v9.2.0
seaborn	v0.12.0

If you are having problems, you can try installing the specific versions that we used. To do this, open a new instance of the command line and copy-paste in the following:

```
python -m pip install colorama==0.4.6 matplotlib==3.8.4 matplotlib-venn==0.11.7  
numpy==1.23.2 pandas==1.4.4 pillow==9.2.0 freesimplegui==5.2.0.post1 rtree==1.0.0 scikit-  
learn==1.1.2 scipy==1.13.1 seaborn==0.12.0 statsmodels==0.13.2
```

```
C:\Users\uqamcc11>python -m pip install colorama==0.4.6 matplotlib==3.8.4 matplotlibli  
b-venn==0.11.7 numpy==1.23.2 pandas==1.4.4 pillow==9.2.0 freesimplegui==5.2.0.post1  
rtree==1.0.0 scikit-learn==1.1.2 scipy==1.13.1 seaborn==0.12.0 statsmodels==0.13.2
```

Data requirements

NASTIC Wrangler requires you to have your data in two separate directories, one for each experimental condition. For example, let's say you'd NASTIC analysed 3 files from unstimulated control samples and 3 from stimulated samples. Most likely this data was already in separate directories, but if it wasn't you could move the output directories created by NASTIC so that you had the following directory structure:

```
/control/sample1_NASTIC_20230607-123456/metrics.tsv
/control/sample2_NASTIC_20230607-134562/metrics.tsv
/control/sample3_NASTIC_20230607-143456/metrics.tsv
```

```
/stimulation/sample1_NASTIC_20230608-091234/metrics.tsv
/stimulation/sample2_NASTIC_20230608-114523/metrics.tsv
/stimulation/sample3_NASTIC_20230608-123412/metrics.tsv
```

NASTIC Wrangler will actually search recursively for metrics.tsv files, so they could be buried in additional directories:

```
/control/another_directory/sample1_NASTIC_20230607-123456/metrics.tsv
/control/sample2_NASTIC_20230607-134562/metrics.tsv
/control/another_directory/and_another_one/sample3_NASTIC_20230607-143456/metrics.tsv
```

Metrics files

Regardless of whether you've used NASTIC, segNASTIC or BOOSH, the metrics.tsv files contain largely the same information, and differ only in reporting the different parameters that each program used.

```
NASTIC: NANOSCALE SPATIO TEMPORAL INDEXING CLUSTERING - Tristan Wallis t.wallis@uq.edu.au
TRAJECTORY FILE: C:/Users/uqamcc11/Documents/2024/PYTHON/2024_04_DATA/synthetic_col1.trxyt
ANALYSED: 20240730-140145
TRAJECTORY LENGTH CUTOFFS (steps): 5 - 1000
SELECTION DENSITY: 0.0
ACQUISITION TIME (s): 320
FRAME TIME (s): 0.02
TIME THRESHOLD (s): 20
RADIUS FACTOR: 1.2
CLUSTER THRESHOLD: 3
CLUSTER MAX RADIUS (um): 0.2
MSD FILTER THRESHOLD (um^2): None
SELECTION AREA (um^2): 111.88687310381138
SELECTED TRAJECTORIES: 2108
CLUSTERED TRAJECTORIES: 1541
UNCLUSTERED TRAJECTORIES: 567
VAR CONFINED TRAJECTORIES: 1653
VAR UNCONFINED TRAJECTORIES: 455
TOTAL CLUSTERS: 269
CLUSTERED TRAJECTORIES AVERAGE INSTANTANEOUS DIFFUSION COEFFICIENT (um^2/s): 0.005660185944639138
UNCLUSTERED TRAJECTORIES AVERAGE INSTANTANEOUS DIFFUSION COEFFICIENT (um^2/s): 0.26903002105329465
HOTSPOTS (CLUSTER SPATIAL OVERLAP AT 1/2 AVERAGE RADIUS): 14
TOTAL CLUSTERS IN HOTSPOTS: 28
AVERAGE CLUSTERS PER HOTSPOT: 2.0
PERCENTAGE OF CLUSTERS IN HOTSPOTS: 10.409

MSD CURVE DATA:
TIME (S): 0.02 0.04 0.06 0.08
UNCLUST MSD (um^2): 0.005999831369649686 0.011451052639560358 0.01694689784117027 0.022141632632847363
UNCLUST SEM: 7.642671590308276e-05 0.00022541218679897047 0.0004207143744525531 0.0006489729527946453
CLUST MSD (um^2): 0.0034137781718262717 0.003492490053222695 0.0036231789097534383 0.0037533893285046204
CLUST SEM: 2.371546476182559e-05 4.042534757295474e-05 6.53426733642238e-05 9.576686450205503e-05
ALL MSD (um^2): 0.0041093626894571425 0.005633147067669306 0.007206930633716126 0.008699373177443109
ALL SEM: 3.6700796208572195e-05 0.0001022613710685705 0.0001778930355808477 0.00025867155924025185

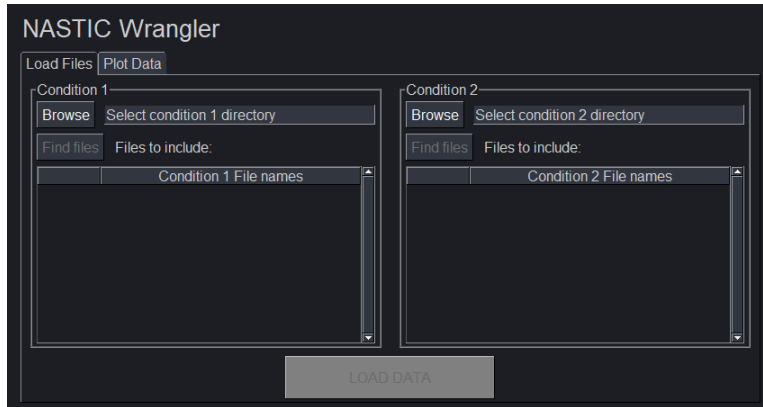
INDIVIDUAL CLUSTER METRICS:
CLUSTER MEMBERSHIP LIFETIME (s) AVG MSD (um^2) AREA (um^2) RADIUS (um) DENSITY (traj/um^2) RATE (traj/sec) AVG TIME (s)
1 6 10.1600000000000025 0.0032329336334323074 0.047451449965210925 0.12289941268241329 126.44502969664585 0.5905511811023607 275.055
2 7 13.3899999999999958 0.003067482504363558 0.022868702248484765 0.08531901317928017 306.09520050329076 0.5227781926811069 117.54142857142857
3 3 3.3599999999999994 0.0024030538250742473 0.004695380940077673 0.03865987807308326 638.9257950070337 0.892857142857143 62.22666666666667
```

Please do not change the formatting of the files in any way.

Usage

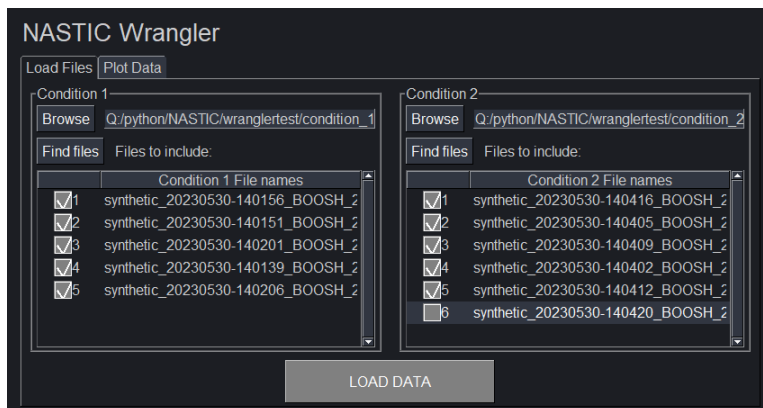
If all is installed properly then double clicking the **nastic_wrangler_gui.py** icon will launch the program.

Load Files tab:



Operation is very straightforward:

1. Select the top-level directory for each condition using “Browse”
2. Click the “Find files” button and NASTIC Wrangler will recursively search the directory for all metrics.tsv files and will populate a list
3. Select/deselect files using the tick box (use ~1s delay between clicks)
4. Click “LOAD DATA”



At this point NASTIC Wrangler will read all the selected metrics.tsv files, report stuff to the terminal, and will switch to the “Plot Data” tab.

```
Reading Condition 1 files...
Condition 1 files selected for analysis:
-----
File 1 Q:/python/NASTIC/wranglertest/condition_1/synthetic_20230530-140156_BOOSH_20230530-140706/metrics.tsv
File 2 Q:/python/NASTIC/wranglertest/condition_1/synthetic_20230530-140151_BOOSH_20230530-140640/metrics.tsv
File 3 Q:/python/NASTIC/wranglertest/condition_1/synthetic_20230530-140201_BOOSH_20230530-140727/metrics.tsv
File 4 Q:/python/NASTIC/wranglertest/condition_1/synthetic_20230530-140139_BOOSH_20230530-140618/metrics.tsv
File 5 Q:/python/NASTIC/wranglertest/condition_1/synthetic_20230530-140206_BOOSH_20230530-140755/metrics.tsv
Condition 1 files read (5 files)

Reading Condition 2 files...
Condition 2 files selected for analysis:
-----
File 1 Q:/python/NASTIC/wranglertest/condition_2/synthetic_20230530-140416_BOOSH_20230530-141051/metrics.tsv
File 2 Q:/python/NASTIC/wranglertest/condition_2/synthetic_20230530-140405_BOOSH_20230530-140905/metrics.tsv
File 3 Q:/python/NASTIC/wranglertest/condition_2/synthetic_20230530-140409_BOOSH_20230530-140924/metrics.tsv
```

Plot Data tab:

The screenshot shows the 'NASTIC Wrangler' application window with the 'Plot Data' tab selected. The interface is divided into two main sections for 'Condition 1' and 'Condition 2'. For Condition 1, the shorthand name is 'CONT' and the selected color is 'Condition 1' (blue). For Condition 2, the shorthand name is 'STIM' and the selected color is 'Condition 2' (orange). Below these, there are checkboxes for 'Outlier Removal' (checked), 'Aggregate Data' (checked), 'Average Data' (checked), and '2D PCA analysis' (checked). Each checkbox has a corresponding description: 'Outlier Removal' is 'median filtering outlier reduction for aggregate data', 'Aggregate Data' is 'aggregates individual cluster metrics across all samples (N = total # of clusters)', 'Average Data' is 'averages cluster metrics for each sample (N = # of samples)', and '2D PCA analysis' is 'determine the overall relationships between samples'. At the bottom right, there is a 'PLOT DATA' button.

5. Give each condition a shorthand name that will show up in subsequent plots
6. Select a colour to use for plotting each condition
7. Use the tick boxes to select what you want to plot
8. Click the “PLOT DATA” button. NASTIC Wrangler will save all plotted data to a dated timestamped directory in the same directory as NASTIC Wrangler itself: **nastic_wrangler_output_YYYYMMDD-HHMMSS**.
9. Paste your plots into your manuscript and submit it to Nature.

Notes:

- If a new directory is selected using “Browse”, the “LOAD DATA” and “PLOT DATA” buttons will become disabled, as the files in the newly selected directory will first have to be found using the “Find files” button before they can be read using “LOAD DATA” and subsequently analysed using “PLOT DATA”.
- If after pressing the “LOAD DATA” button, the user changes the list of files to include for analysis (by unticking previously included files and/or ticking previously excluded files), the “PLOT DATA” button will become disabled until the new list of files to include in analysis is read by pressing the “LOAD DATA” button.
- Shorthand names to show on plots: must be different and cannot be blank as this causes issues with how the plots are displayed. If you really don’t want a label to appear on the plots, you can set the shorthand name of one condition to a single space, and the shorthand name of the other condition to a double space, so that they appear blank on the plots but will not interfere with the generation of the plots.

Plot types

Aggregate data: These plots combine all the metrics from all of the clusters observed across all your files for a given condition. These plots are saved as PNG formatted to `nastic_wranger_output_YYYYMMDD-HHMMSS/aggregate_plots`

Metrics:

Area: average cluster area

Avmsd: average cluster MSD (at t=0)

Avtime: average cluster time centroid

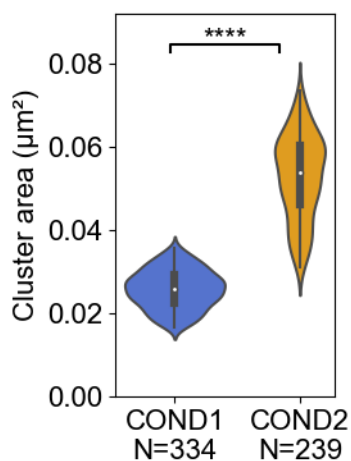
Density: average density of trajectories within clusters

Lifetime: average cluster lifetime

Membership: average number of trajectories per cluster

Radius: average cluster radius

Rate: average rate of detections within clusters



Data is plotted as a violin plot (not individual data points) to allow assessment of data distribution. The white dot in the centre of the plot shows the median of the distribution, the thicker dark bar shows the interquartile range of the data, and the thinner dark bar shows the data that extends to 1.5 times the interquartile range. The significance, as determined by t-test, is plotted as asterisks where ns = no significance, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

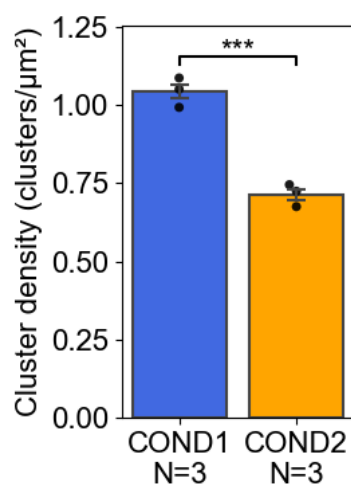
You can optionally remove outliers from aggregate data prior to plotting, using median filtering. Median filtering works by calculating the mean for the data, establishing how far each value in the data deviates from the mean, and then dividing each deviation by the mean of all the deviations. Those values whose deviation/mean of deviation is > 2.5 are excluded.

Aggregate metrics will highlight trends in clustering behaviour between conditions, but the large N means that there is more chance of statistical significance.

Average data: These plots derive an average for each metric across the number of samples analysed. These plots are saved to **nastic_wranger_output_YYYYMMDD-HHMMSS/average_plots** and their filenames are self-explanatory. Please note that while there a large number of average metrics, not all of them are particularly informative and they are included for completeness. For example, Selection_area is down to user input and does not reflect any biology.

Metrics:

Average_cluster_area
 Average_cluster_lifetime
 Average_cluster_membership
 Average_cluster_radius
 Average_cluster_rate
 Average_cluster_time_centroid
 Average_clusters_per_hotspot
 Average_density_of_clustered_trajectories
 Average_MSD_of_clustered_trajectories
 Cluster_density
 Clustered_trajectories
 Clustered_trajectories_Inst_diff_coeff
 Hotspots
 Percentage_clustered_trajectories
 Percentage_clusters_in_hotspots
 Selected_trajectories
 Selection_area
 Total_clusters
 Total_clusters_in_hotspots
 Unclustered_trajectories
 Unclustered_trajectories_Inst_diff_coeff



Data is presented as a bar plot of the average with datapoints plotted as a swarm plot, and significance displayed as above.

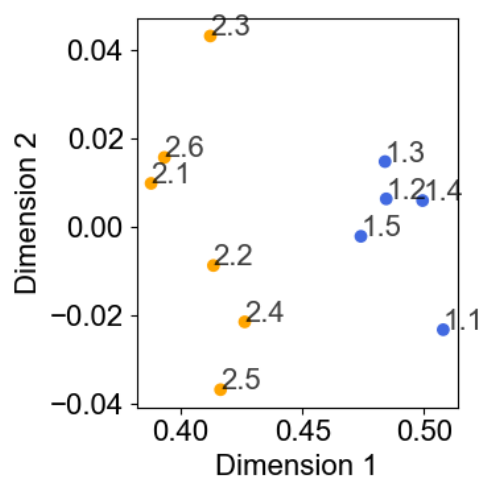
All of the aggregate and average plotted data is also saved to **nastic_wranger_output_YYYYMMDD-HHMMSS/processed_metrics.tsv** so that you may import it into other software for analysis and plotting.

PCA: These plots use Principal Component Analysis where each sample datapoint represents a number of normalised average metrics as described above.
nastic_wranger_output_YYYYMMDD-HHMMSS/pca.png
nastic_wranger_output_YYYYMMDD-HHMMSS/pca_labels.png

Metrics used:

Percentage Clustered Trajectories
Var Confined Trajectories*
Var Unconfined Trajectories*
Clustered Trajectories Average Instantaneous Diffusion Coefficient (Um^2/S)
Unclustered Trajectories Average Instantaneous Diffusion Coefficient (Um^2/S)
Average Clusters Per Hotspot
Percentage Of Clusters In Hotspots
Cluster Density (Clusters/ Um^2)

* If VAR analysis was performed



This plot allows the user to determine qualitatively if there are differences between the conditions, and whether there may be outlier samples which can be removed (using sample tick boxes as described above) for future analyses.

1.1 = condition 1, sample 1

2.3 = condition 2, sample 3

The plot is available with/without labels

Menu

Default settings:

Similarly to the other NASTIC programs, the first time you run NASTIC Wrangler it will create a file called **nastic_wrangler_gui.defaults** which contains the default values for the plots that are generated. You can load it into a text editor or spreadsheet to view it if you'd like.

Short name condition 1	COND1
Short name condition 2	COND2
Color condition 1	royalblue
Color condition 2	orange
Plot average data	True
Plot aggregate data	True
Outlier removal	True
Plot 2D PCA data	True

Values that can be changed in NASTIC Wrangler and saved to this file include: shorthand names and colors given to condition 1 and 2 (shown on plots), which plots will be generated (Average, Aggregate and 2D PCA), and whether Outlier Removal will be performed on Aggregate data. If you wish to keep these values (especially color) consistent across datasets, "**File >> Save settings**" will save them to **nastic_wrangler_gui.defaults**. The next time you start NASTIC Wrangler, they'll be loaded in as the defaults. If you

wish to reset these values, "**File >> Default settings**" will restore the defaults shown in the above table. If you don't save them, the next time you load NASTIC you will have to reselect the values that you used previously. Similarly, if you started with your preferred values, and changed settings, "**File >> Load settings**" will restore your last saved settings. If settings for previously plotted data were not saved, the colors and shorthand names used, as well as whether outlier reduction was used, can be found in the corresponding **processed_metrics.tsv** output file, and then copied (without spaces) into the defaults file.

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

Help:

Clicking on "**Info >> Help**" will show a popup window (left) which explains what each of the buttons do, as a useful reminder when using NASTIC Wrangler.

Updates:

"**Info >> Updates**" will open the NASTIC GitHub release page in your default browser. The latest release of NASTIC is always at the top of the page. Compare the date in your NASTIC window with the release date, and update as necessary. NASTIC development is ongoing, and you are strongly encouraged to keep up to date for bug fixes and feature improvements.

Help

This program allows the visualisation of the metrics.tsv files produced by NASTIC and SEGNASTIC. Comparison bar plots for each metric and statistical significance (t-test) are shown.

Load Files tab:

- Browse: Select a directory for each condition.
- Find files: Recursively search directories for metrics.tsv files.
- Untick files to exclude from analysis (use ~1s delay between clicks).
- LOAD DATA: Load selected files and extract information.

Plot Data tab:

- Shorthand name: for each condition to appear on plots.
- Select color: for each condition to appear on plots.
- Outlier removal: uses median filtering to remove outliers from aggregate data.
- Aggregate data: aggregates individual cluster metrics across all samples (N = total number of clusters).
- Average data: averages cluster metrics for each sample (N = number of samples).
- 2D PCA analysis allows you to determine the overall relationships between samples.
- PLOT DATA: plot (selected) aggregate, average and PCA. Also saves timestamped TSV of raw data used for plots.

Tristan Wallis, Alex McCann 20230608

OK