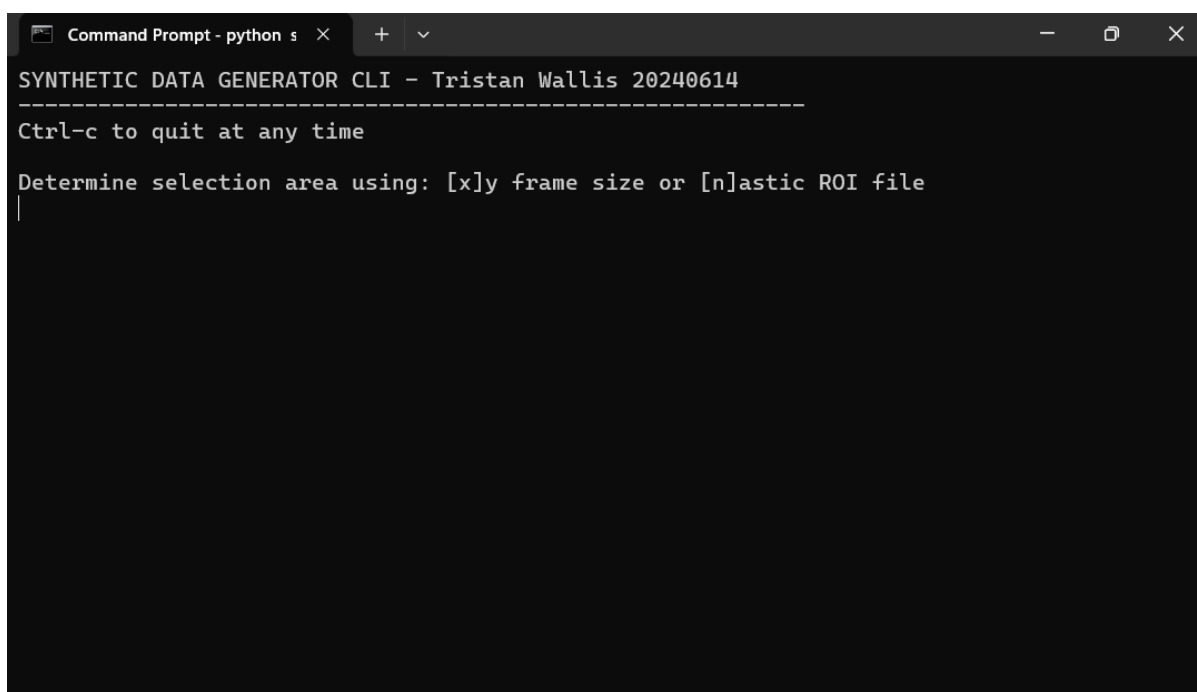# SYNTHETIC DATA GENERATOR (CLI) USER MANUAL

This user manual is for the command line (CLI) version.

For information on how to use the graphical user interface (GUI) version, please refer to the Synthetic Data Generator (GUI) User Manual.

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 Synthetic Data Generator CLI is a simple text mode interface used to generate synthetic trajectory files in the TRXYT format for use in NASTIC, segNASTIC and BOOSH (spatiotemporal clustering software).

TRXYT files are generated using pre-defined parameters located within the script itself:

```
# VARS
acquisition_time = 320 # pretend length of acquisition (default 320)
frame_time = 0.02 # sec (default 0.02)
seed_density = 2 # how many cluster seeds per um2 of selection area (default 2)
min_traj_num = 4 # min number of trajectories around each seed (default 4)
max_traj_num = 12 # max number of trajectories around each seed (default 12)
x_size = 10 # pretend microns (default 10)
y_size = 10 # pretend microns (default 10)
min_traj_length = 5 # min number of trajectory steps (default 5)
max_traj_length = 30 # max number of trajectory steps (default 30)
radius = 0.1 # radius around each seed to make trajectories (um) (default 0.1)
steplength = 0.1 # maximum step length within trajectory (um) (default 0.1)
noise_factor = 5 # number of unclustered noise trajectories per seed (default 5)
unconst = 4 # steplength multiplier of unclustered trajectories (default 4)
hotspotprobability = 0.2 # chance of a given seed point generating multiple spatially
overlapping but temporally distinct clusters (default 0.2)
hotspotmax = 3 # maximum number of temporal clusters at a given hotspot (default 3)
orbit = True # clustered trajectories orbit their spawn point rather than random
walking (default True)
```

To change these parameters, you can:

> 1) open the script in a text editor (such as Notepad++) and change the default values to the desired values before saving and rerunning the script.

> 2) use the graphical user interface (GUI) version of the Synthetic Data Generator to adjust values (see the Synthetic Data Generator GUI User Manual for more details).

Two methods can be used to define the selection area (ROI):

> 1) use the x_size and y_size parameters to define a rectangular ROI (in microns)

> 2) load an ROI file previously generated using NASTIC, segNASTIC or BOOSH
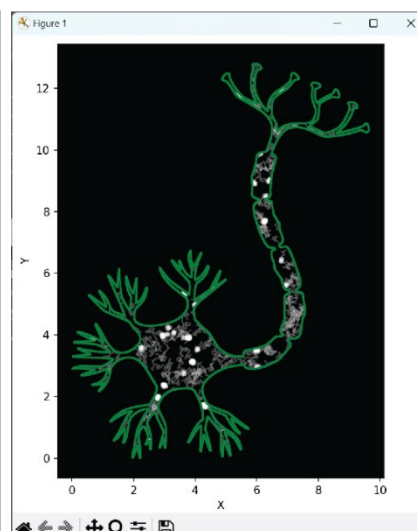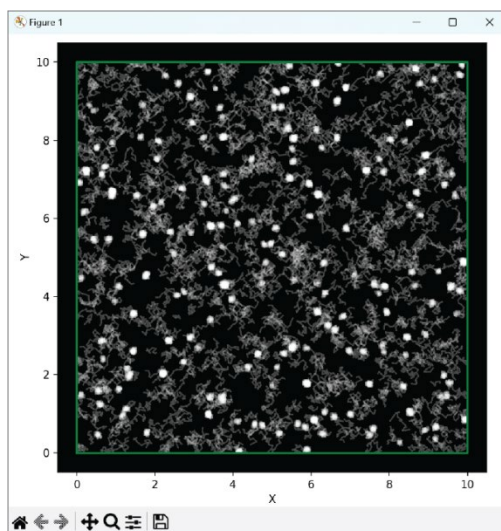
Left: example plot of trajectories that were generated using x,y frame size (Option 1).

Right: example plot of trajectories that were generated using a pre-saved NASTIC ROI file (Option 2).

Computer requirements

The NASTIC suite (which includes the Synthetic Data Generator CLI) consists of Python scripts that require Python 3.8 or later, and a number of Python modules to run. Python is available for most computer platforms so you can run it on Windows, Linux and Mac. It 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. You will also need to install a number of Python modules, which is simple to do from a command line.

The Python module versions used in the most recent version of the Synthetic Data Generator CLI are:

scipy	v1.13.1
numpy	v1.23.2
matplotlib	v3.8.4

To install these specific Python module versions, open a new instance of the command line and copy-paste in the following:

```
python -m pip install scipy==1.13.1 numpy==1.23.2 matplotlib==3.8.4
```

```
C:\Users\uqamcc11>python -m pip install scipy==1.13.1 numpy==1.23.2 matplotlib==3.8.4
```

Updates

The Synthetic Data Generator CLI along with other NASTIC family members is periodically updated with new functionalities.

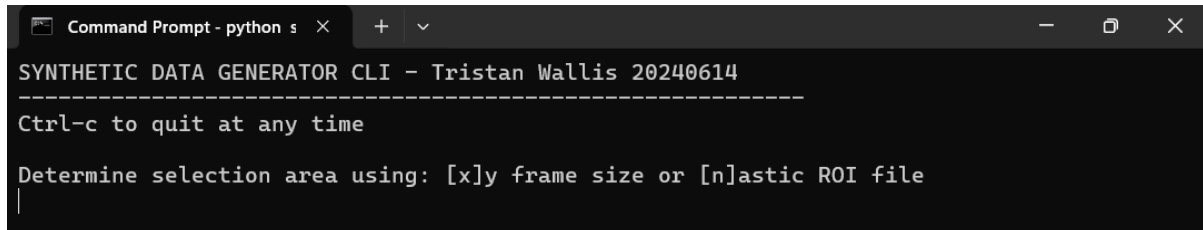Check the GitHub for updates using this link:
https://github.com/tristanwallis/smlm_clustering/releases

Running the Synthetic Data Generator CLI

Step 1 – run the script:

<u>Option 1 – double click on the script:</u>

Double click on **synthetic_data_generator_cli.py** script, and a simple text mode interface will launch in a terminal:

<u>Option 2 – run the script using the terminal:</u>

Alternatively, this can be done by opening a terminal, navigating to the location of the **synthetic_data_generator_cli.py** script, and typing the following before pressing the return key:

This will launch the simple text mode interface in the terminal:

Step 2 – choose the method used to define the selection area (ROI):

Option 1 – define using pretend x,y frame size:

To define the selection area (region of interest; ROI) using pretend x and y-axis lengths, type the letter x followed by the return key:

```
Command Prompt - python s    ×    +    ˅

SYNTHETIC DATA GENERATOR CLI - Tristan Wallis 20240614
--------------------------------------------------------
Ctrl-c to quit at any time

Determine selection area using: [x]y frame size or [n]astic ROI file
x
```
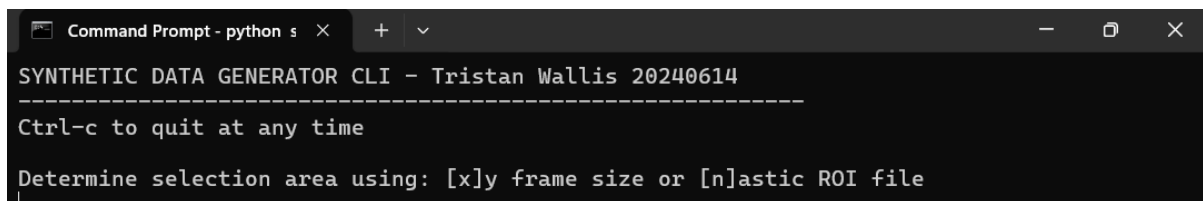
The Synthetic Data Generator will then generate metrics using the parameters that are pre-defined in the script. These metrics will be used to generate the TRXYT file.

The following information will be printed in the console:

```
Generating trajectories
------------------------------------------------------------
Generating 200 spatiotemporal cluster seeds (2 seeds/um2)...
Generating trajectories around cluster seeds...
Generating unclustered trajectory seeds...
Generating 1000 unclustered trajectories with higher mobility (5x number of cluster seeds)...

Generating metrics
------------------------------------------------------------
Total traj: 3175
Clustered traj: 2175
Total clusters: 261
Avg traj per cluster: 8.333333333333334
Avg cluster radius: 0.07325633932431765

Writing files
------------------------------------------------------------
Writing TRXYT...
Writing metrics...
TRXYT and metrics files saved to:
C:\Users\uqamcc11\Documents\2024\PYTHON\2024_04_DATA\synthetic_data_output_20240730-155243


Done!

Return key: Generate another file
Ctrl-c: Exit
```

A datestamped folder (synthetic_data_output_YYYYMMDD-hhmmss) will be generated in the same folder as the Synthetic Data Generator, and will contain the following datestamped files:

1) metrics.tsv   (synthetic_YYYYMMDD-hhmmss_metrics.tsv)

2) .trxyt   (synthetic_data_YYYYMMDD-hhmmss.trxyt

Once the files have been saved, you can either:

1) Generate another trxyt file (press the return key to go back to the start)

2) Exit the program (press Ctrl + C to exit)

<u>Option 2 – define using pre-saved NASTIC ROI file:</u>

To define the selection area (region of interest; ROI) using an ROI that was previously saved in NASTIC, type the letter n followed by the return key:

The Synthetic Data Generator will recursively search for roi_coordinates.tsv files previously generated in NASTIC, segNASTIC or BOOSH.

Located ROI files will be listed in the console:

Note: To be successfully located by the Synthetic Data Generator, the ROI files must:

1) be in the correct naming format (* = any character): *roi_coordinates*.tsv
2) be in the same folder as the Synthetic Data Generator (or within subfolder)

To select the desired ROI file, type in the number corresponding to the file followed by the return key:

The following information will be printed in the console:

```
Determining selection area
--------------------------------------------------------
NASTIC ROI file selected:
C:\Users\uqamcc11\Documents\2024\PYTHON\2024_04_DATA\roi_coordinates1.tsv
Reading ROI file...
ROI selection area: 111.88687310381138 um2

Generating trajectories
--------------------------------------------------------
Generating 224 spatiotemporal cluster seeds (2 seeds/um2)...
Generating trajectories around cluster seeds...
Generating unclustered trajectory seeds...
Generating 1120 unclustered trajectories with higher mobility (5x number of cluster seeds)...

Generating metrics
--------------------------------------------------------
Total traj: 3403
Clustered traj: 2283
Total clusters: 294
Avg traj per cluster: 7.76530612244898
Avg cluster radius: 0.06972987734654672

Writing files
--------------------------------------------------------
Writing TRXYT...
Writing metrics...
TRXYT and metrics files saved to:
C:\Users\uqamcc11\Documents\2024\PYTHON\2024_04_DATA\synthetic_data_output_20240730-161641


Done!

Return key: Generate another file
Ctrl-c: Exit
```

A datestamped folder (synthetic_data_output_YYYYMMDD-hhmmss) will be generated in the same folder as the Synthetic Data Generator, and will contain the following datestamped files:

1) metrics.tsv   (synthetic_YYYYMMDD-hhmmss_metrics.tsv)

2) .trxyt   (synthetic_data_YYYYMMDD-hhmmss.trxyt

Once the files have been saved, you can either:

1) Generate another trxyt file (press the return key to go back to the start)
2) Exit the program (press Ctrl + C to exit)