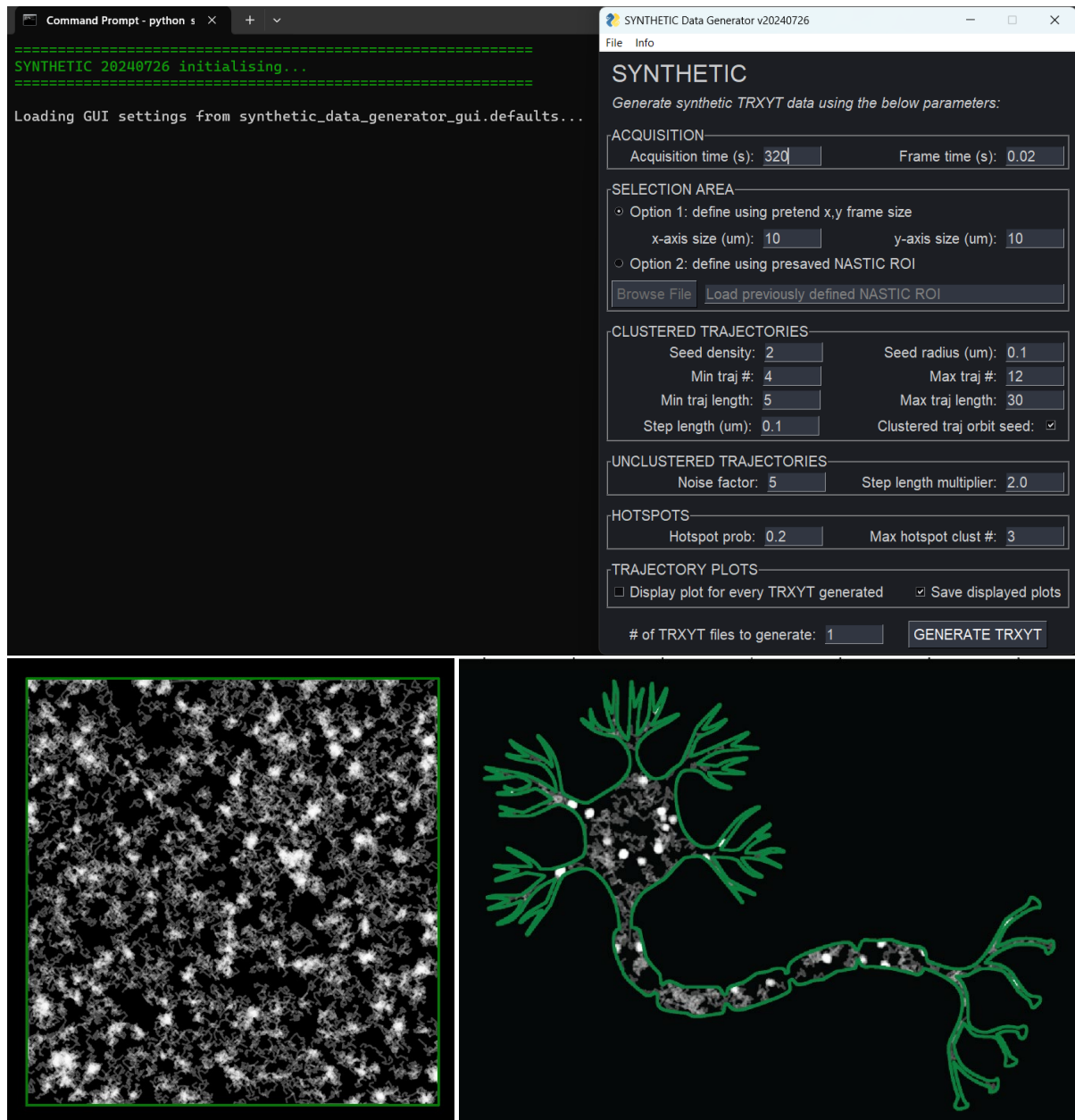


# SYNTHETIC DATA GENERATOR GUI

## USER MANUAL



Tristan Wallis [t.wallis@uq.edu.au](mailto:t.wallis@uq.edu.au)

Alex McCann [a.mccann2@uq.edu.au](mailto:a.mccann2@uq.edu.au)

Single Molecule Neuroscience Laboratory (Fred Meunier)

Queensland Brain Institute

The University of Queensland, Australia

## Introduction

The purpose of the Synthetic Data Generator GUI is to take parameters input by the user to generate synthetic trajectory files in the TRXYT format for use in NASTIC, segNASTIC and BOOSH (spatiotemporal clustering software).

Parameters that can be defined by the user in the Synthetic Data generator GUI are:

Parameter	Description	Options / user input
<b>AQUISITION</b>		
Acquisition time (s)	Pretend length of time (in seconds) to 'acquire' all frames	Any whole number (integer) larger than 0
Frame time (s)	Pretend length of time (in seconds) to 'acquire' a single frame	Any decimal number (float) larger than 0
<b>SELECTION AREA</b>		
Option 1: define using pretend x,y frame size		
x-axis size (um)	Determine selection area of using a pretend frame size (x-axis length in microns)	Any whole number (integer) larger than 0
y-axis size (um)	Determine selection area of using a pretend frame size (y-axis length in microns)	Any whole number (integer) larger than 0
Option 2: define using pre-saved NASTIC region of interest (ROI)		
Browse File	Determine the selection area using a previously generated NASTIC ROI file	roi_coordinates.tsv
<b>CLUSTERED TRAJECTORIES</b>		
Seed density	Number of spawn points (seeds) per $\text{um}^2$ of the selection area from which a single cluster will occur	Any whole number (integer) larger than 0
Seed radius (um)	Radius around each seed to make trajectories (in microns)	Any decimal number (float) larger than 0
Min traj #	Minimum number of trajectories around each seed	Any whole number (integer) larger than 4
Max traj #	Maximum number of trajectories around each seed	Any whole number (integer) larger than

		the min traj # and smaller than 10000
Min traj length	Minimum number of trajectory steps	Any whole number (integer) larger than 5
Max traj length	Maximum number of trajectory steps	Any whole number (integer) larger than the min traj length
Step length (um)	Maximum step length within a trajectory (in microns)	Any decimal number (float) larger than 0
Clustered traj orbit seed	Clustered trajectories orbit their spawn point rather than random walking	True / False
<b>UNCLUSTERED TRAJECTORIES</b>		
Noise factor	Noise factor multiplier used to calculate the number of unclustered trajectories per seed	Any whole number (integer) larger than 0
Step length multiplier	Step length multiplier used to calculate unclustered trajectory step length	Any decimal number (float) larger than 0
<b>HOTSPOTS</b>		
Hotspot prob	Chance of a given seed point generating multiple spatially overlapping but temporally distinct clusters	Any decimal (float) larger than 0
Max hotspot clust #	Maximum number of temporal clusters at a given hotspot	Any whole number (integer) larger than 1

## Computer requirements

The NASTIC suite (which includes the Synthetic Data Generator GUI) 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 GUI are:

pysimplegui	v4.60.4
colorama	v0.4.6
scipy	v1.13.1
numpy	v1.23.2
scikit-learn	v1.1.2

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 pysimplegui==4.60.4 colorama==0.4.6 scipy==1.13.1 numpy==1.23.2 scikit-learn==1.1.2
```

```
C:\Users\uqamcc11>python -m pip install pysimplegui==4.60.4 colorama==0.4.6 scipy==1.13.1 numpy==1.23.2 scikit-learn==1.1.2
```

## Updates

The Synthetic Data Generator GUI 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](https://github.com/tristanwallis/smlm_clustering/releases)

## Running the Synthetic Data Generator GUI

### Step 1 – run the script

If all is installed properly then double clicking the `synthetic_data_generator_GUI.py` icon will launch a console, followed shortly after by the GUI.

Console:

```
Command Prompt - pytl  x  +  v  -  □  x

=====
SYNTHETIC 20240726 initialising...
=====

Loading GUI settings from synthetic_data_generator_gui.defaults...
|
```

Graphical user interface (GUI):

SYNTHETIC Data Generator v20240726

File Info

### SYNTHETIC

Generate synthetic TRXYT data using the below parameters:

#### ACQUISITION

Acquisition time (s):  Frame time (s):

#### SELECTION AREA

☒ Option 1: define using pretend x,y frame size  
x-axis size (um):  y-axis size (um):   
☐ Option 2: define using presaved NASTIC ROI

#### CLUSTERED TRAJECTORIES

Seed density:  Seed radius (um):   
Min traj #:  Max traj #:   
Min traj length:  Max traj length:   
Step length (um):  Clustered traj orbit seed: ☒

#### UNCLUSTERED TRAJECTORIES

Noise factor:  Step length multiplier:

#### HOTSPOTS

Hotspot prob:  Max hotspot clust #:

#### TRAJECTORY PLOTS

☐ Display plot for every TRXYT generated ☒ Save displayed plots

# of TRXYT files to generate:

## ACQUISITION

ACQUISITION	
Acquisition time (s): <input type="text" value="320"/>	Frame time (s): <input type="text" value="0.02"/>

### Step 2 – define the acquisition time

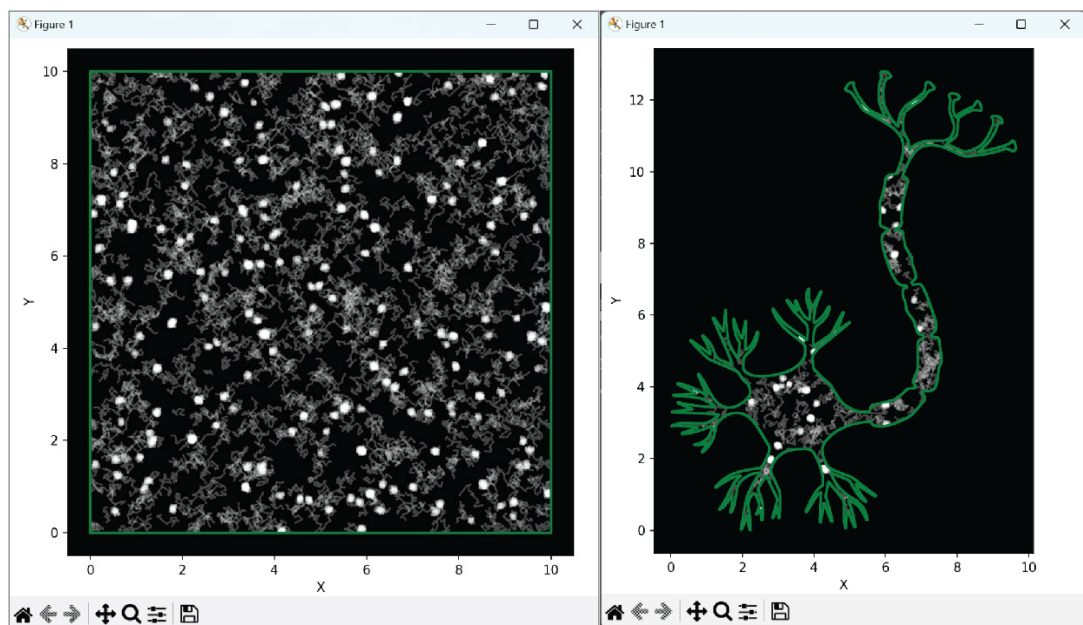
Click on the text box next to 'Acquisition time (s):' and type in the total amount of time (in seconds) taken to 'acquire' all of the pretend frames that will make up the synthetic data acquisition. The value must be a whole number greater than 0.

### Step 3 – define the frame time

Click on the text box next to 'Frame time (s):' and type in the amount of time (in seconds) taken to 'acquire' a single frame in the pretend time series. The value must be a decimal number greater than 0.

## SELECTION AREA

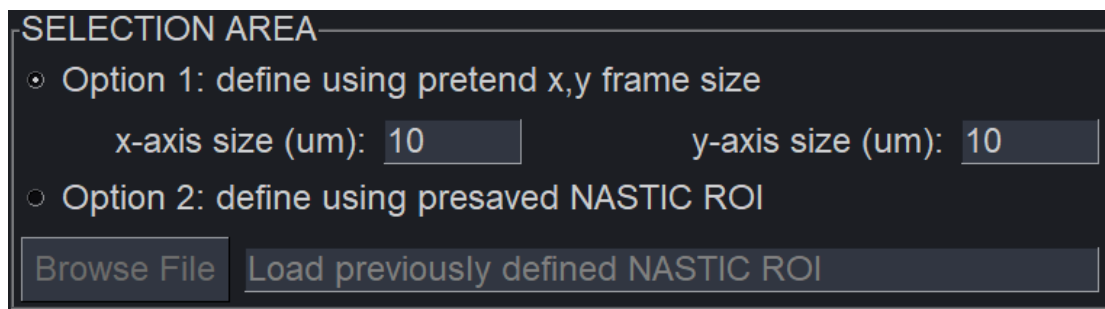
### Step 4 – choose the method used to define the selection area



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).

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

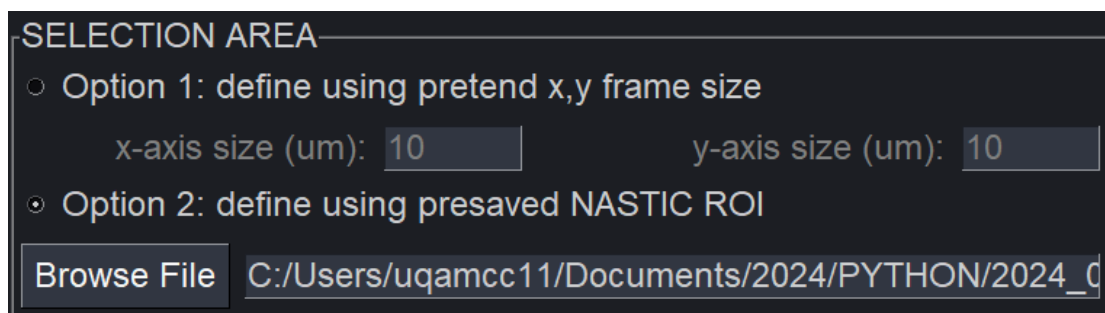


The screenshot shows a dark-themed dialog box titled 'SELECTION AREA'. It contains two radio button options. The first option, 'Option 1: define using pretend x,y frame size', is selected and highlighted with a blue circle. Below this option are two input fields: 'x-axis size (um):' with the value '10' and 'y-axis size (um):' with the value '10'. The second option, 'Option 2: define using presaved NASTIC ROI', is not selected. At the bottom, there are two buttons: 'Browse File' and 'Load previously defined NASTIC ROI'. The 'Browse File' button is disabled (grayed out), while the 'Load previously defined NASTIC ROI' button is active.

To define the selection area (region of interest; ROI) using pretend x and y-axis lengths, click on the radio option on the left of 'Option 1: define using pretend x,y frame size'.

This will enable the 'x-axis size (um)' and 'y-axis size (um)' parameters, and disable the 'Browse File' button that is used for browsing NASTIC ROI files.

#### Option 2: define using presaved NASTIC ROI



The screenshot shows the same 'SELECTION AREA' dialog box, but now 'Option 2: define using presaved NASTIC ROI' is selected and highlighted with a blue circle. The 'x-axis size (um):' and 'y-axis size (um):' input fields are now disabled (grayed out). The 'Browse File' button is now active, while the 'Load previously defined NASTIC ROI' button is disabled. The text box next to the 'Browse File' button contains the file path 'C:/Users/uqamcc11/Documents/2024/PYTHON/2024\_d'.

To define the selection area (region of interest; ROI) using an ROI that was previously saved in NASTIC, click on the radio option on the left of 'Option 2: define using presaved NASTIC ROI'.

This will enable the 'Browse File' button that is used for browsing NASTIC ROI files, and disable the 'x-axis size (um)' and 'y-axis size (um)' parameters.

Next, click on the now enabled 'Browse File' button to open a search window. Browse to a directory that contains a NASTIC 'roi\_coordinates.tsv' file, click on the ROI file of interest, then click 'Open' to confirm.

The filename of the selected ROI file will be shown in the text box next to the 'Browse File' button.

## CLUSTERED TRAJECTORIES

CLUSTERED TRAJECTORIES	
Seed density: <input type="text" value="2"/>	Seed radius (um): <input type="text" value="0.1"/>
Min traj #: <input type="text" value="4"/>	Max traj #: <input type="text" value="12"/>
Min traj length: <input type="text" value="5"/>	Max traj length: <input type="text" value="30"/>
Step length (um): <input type="text" value="0.1"/>	Clustered traj orbit seed: <input checked="" type="checkbox"/>

### Step 5 – define the seed density

To define the number of seeds (spawn points) per micron squared (um<sup>2</sup>) from which clustered trajectories will spawn, click on the text box next to 'Seed density:' and type in a whole number (integer) that is larger than 0.

### Step 6 – define the seed radius

To define the radius (in microns) from the centre of the seed from which clustered trajectories will spawn, click on the text box next to 'Seed radius (um):' and type in a decimal number (float) larger than 0.

### Step 7 – define the minimum number of clustered trajectories

To define the minimum number of clustered trajectories that will be generated (in total), click on the text box next to 'Min traj #' and type in a whole number (integer) larger than 4.

### Step 8 – define the maximum number of clustered trajectories

To define the maximum number of clustered trajectories that will be generated (in total), click on the text box next to 'Max traj #' and type in a whole number (integer) than is larger than the minimum number of clustered trajectories, and smaller than 10000.

### Step 9 – define the minimum number clustered trajectory steps

To define the minimum number steps for each clustered trajectory, click on the text box next to 'Min traj length' and type in a whole number (integer) that is larger than 5.



### Step 10 – define the maximum number clustered trajectory steps

To define the maximum number of steps for each clustered trajectory, click on the text box next to 'Max traj length' and type in a whole number (integer) that is larger than the minimum trajectory length.

### Step 11 – define the step length of clustered trajectories

To define the length of each step in clustered trajectories (in microns), click on the text box next to 'Step length (um):' and type in a decimal number (float) that is larger than 0.

### Step 12 – choose whether clustered trajectories orbit seeds or walk randomly

#### Option 1: trajectories orbit seed

CLUSTERED TRAJECTORIES	
Seed density: 2	Seed radius (um): 0.1
Min traj #: 4	Max traj #: 12
Min traj length: 5	Max traj length: 30
Step length (um): 0.1	Clustered traj orbit seed: <input checked="" type="checkbox"/>

To make clustered trajectories that stay close to their original spawn point, make sure that the checkbox next to 'Clustered traj orbit seed:' is ticked (click on the checkbox to change its status).

#### Option 2: trajectories walk randomly

CLUSTERED TRAJECTORIES	
Seed density: 2	Seed radius (um): 0.1
Min traj #: 4	Max traj #: 12
Min traj length: 5	Max traj length: 30
Step length (um): 0.1	Clustered traj orbit seed: <input type="checkbox"/>

To make clustered trajectories that move randomly after spawning at their seed point, make sure that the checkbox next to 'Clustered traj orbit seed:' is unticked (click on the checkbox to change its status).

## UNCLUSTERED TRAJECTORIES

UNCLUSTERED TRAJECTORIES	
Noise factor: <input type="text" value="5"/>	Step length multiplier: <input type="text" value="2.0"/>

### Step 13 – define the noise factor

To determine the number of unclustered trajectories using the noise factor (multiply noise factor by the number of clustered trajectories), click on the text box next to 'Noise factor:' and type in a whole number (integer) that is larger than 0.

### Step 14 – define the step length multiplier

To determine the step lengths of unclustered trajectories (multiply step length multiplier by the step length of clustered trajectories), click on the text box next to 'Step length multiplier:' and type in a decimal number (float) larger than 0.

## HOTSPOTS

HOTSPOTS	
Hotspot prob: <input type="text" value="0.2"/>	Max hotspot clust #: <input type="text" value="3"/>

### Step 15 – define the hotspot probability

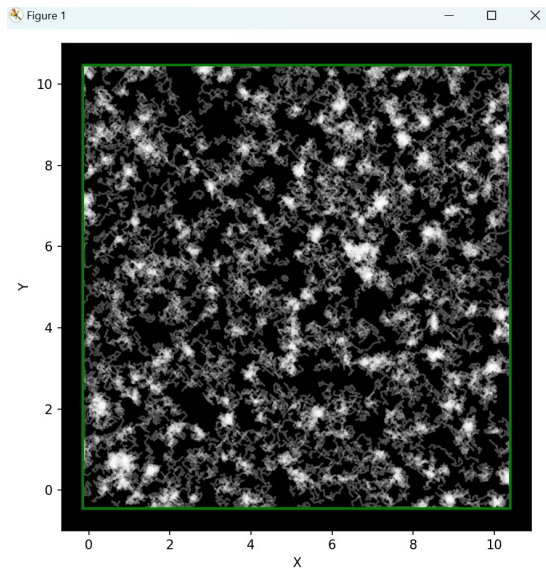
To define the chance of a given seed point generating multiple clusters that overlap in time and space (hotspots), click on the text box next to 'Hotspot prob:' and type in a decimal number that is larger than 0.

### Step 16 – define the maximum number of clusters per hotspot

To define the maximum number of temporal clusters that occur at a given hotspot, click on the text box next to 'Max hotspot clust #:', and type in a whole number that is larger than 1.

## TRAJECTORY PLOTS

TRAJECTORY PLOTS	
<input type="checkbox"/> Display plot for every TRXYT generated	<input checked="" type="checkbox"/> Save displayed plots



Example plot of generated trajectories with selection area outlined in green.

### Step 17 – select whether to display single or multiple TRXYT plots

#### Option 1: display a single representative TRXYT plot

**TRAJECTORY PLOTS**

☐ Display plot for every TRXYT generated
 ☒ Save displayed plots

To display a single representative TRXYT plot, make sure that the 'Display plot for every TRXYT generated' checkbox is unticked (click on the checkbox to change its status).

Note: selecting this option means that only the representative trajectory plot can be saved.

#### Option 2: display every TRXYT plot

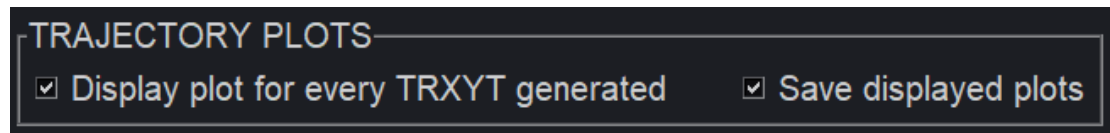
**TRAJECTORY PLOTS**

☒ Display plot for every TRXYT generated
 ☒ Save displayed plots

To display a plot for every TRXYT file that is generated, make sure that the 'Display plot for every TRXYT generated' checkbox is ticked (click on the checkbox to change its status).

### Step 18 – select whether the displayed plots will be saved

#### Option 1: save all displayed plots



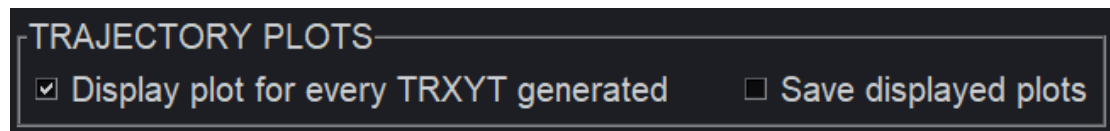
TRAJECTORY PLOTS

☒ Display plot for every TRXYT generated      ☒ Save displayed plots

To save all displayed plots, make sure that the 'Save displayed plots' checkbox is ticked (click on the checkbox to change its status).

Note: if the 'Display plot for every TRXYT generated' checkbox is unticked, only the single representative TRXYT plot will be saved.

#### Option 2: don't save any plots

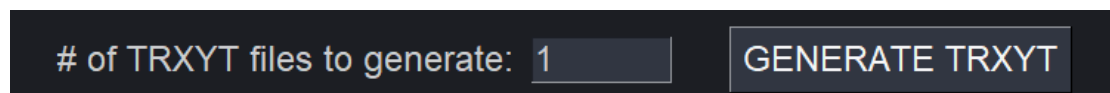


TRAJECTORY PLOTS

☒ Display plot for every TRXYT generated      ☐ Save displayed plots

To prevent plots from being saved, make sure that the 'Save displayed plots' checkbox is unticked (click on the checkbox to change its status).

### Step 19 – Select the number of TRXYT files to generate



# of TRXYT files to generate: 1      GENERATE TRXYT

To select the number of synthetic TRXYT files to generate, click on the text box next to ' # of TRXYT files to generate:' and type in a whole number that is larger than 0.

### Step 20 – Generate the TRXYT file(s)

To generate the TRXYT file(s) using the selected parameters, click on the 'GENERATE TRXYT' button.

The metrics generated by using the selected parameters will be shown in the terminal:

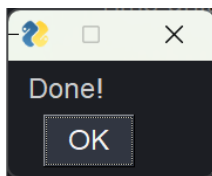
```
Command Prompt - python s X + v

Generating trxyt file # 1
-----
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)...
Total traj: 3061
Clustered traj: 2061
Total clusters: 260
Avg traj per cluster: 7.926923076923077
Avg cluster radius: 0.07336383526946079
Writing trxyt...
Writing metrics...

Generating trxyt file # 2
-----
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)...
Total traj: 3030
Clustered traj: 2030
Total clusters: 248
Avg traj per cluster: 8.15
Avg cluster radius: 0.0735254788824443
Writing trxyt...
Writing metrics...
Plotting trajectories...
Plotting selection area...

DONE!
```

A 'Done!' popup will show once the GUI has finished generating the files:



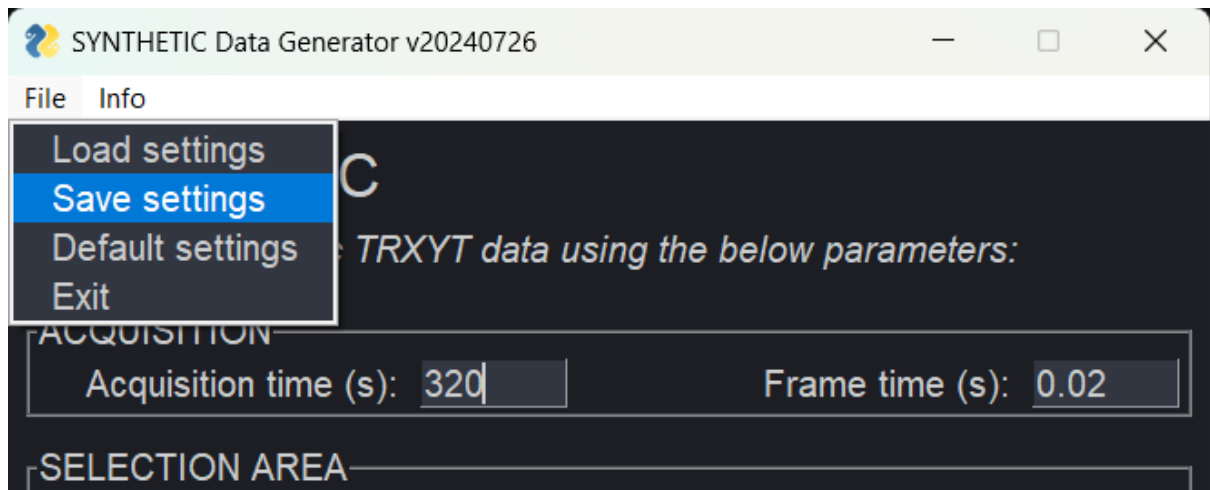
A date-stamped folder will be generated in the same location as the `super_res_data_wrangler_gui.py` script (`synthetic_data_output_YYYYMMDD-hhmmss`). This folder contains:

- Each TRXYT file generated (date-stamped and numbered; `synthetic_data_YYYYMMDD-hhmmss_1.trxyt`)
- A metrics file for each TRXYT file generated (`synthetic_data_YYYYMMDD-hhmmss_1_metrics.tsv`)
- If applicable: trajectory plot(s) with the selection area outlined in green (`synthetic_data_YYYYMMDD-hhmmss_1.png`)

## Save settings

The default values for each of the parameters is saved in a defaults file called 'synthetic\_data\_generator\_gui.defaults'.

To save the selected parameters to the defaults file for use next time the GUI is opened, click on the 'File >> Save settings' option in the Menu bar (top left hand corner of the GUI).

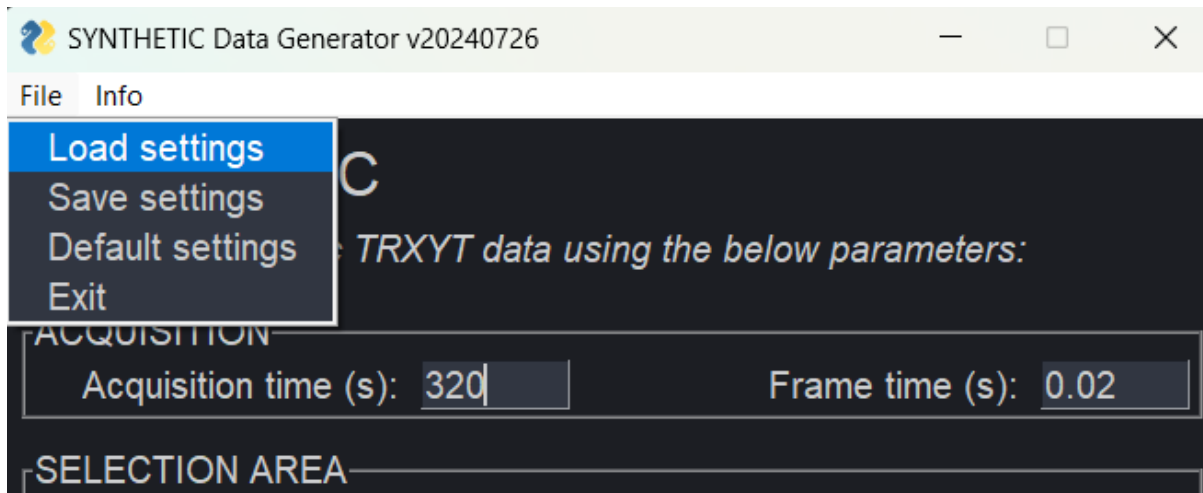


The following message will be displayed in the console:

```
Saving GUI settings to synthetic_data_generator_gui.defaults...
```

## Load settings

If you wish to load the previously saved settings, click 'File >> Load settings' in the menu bar.

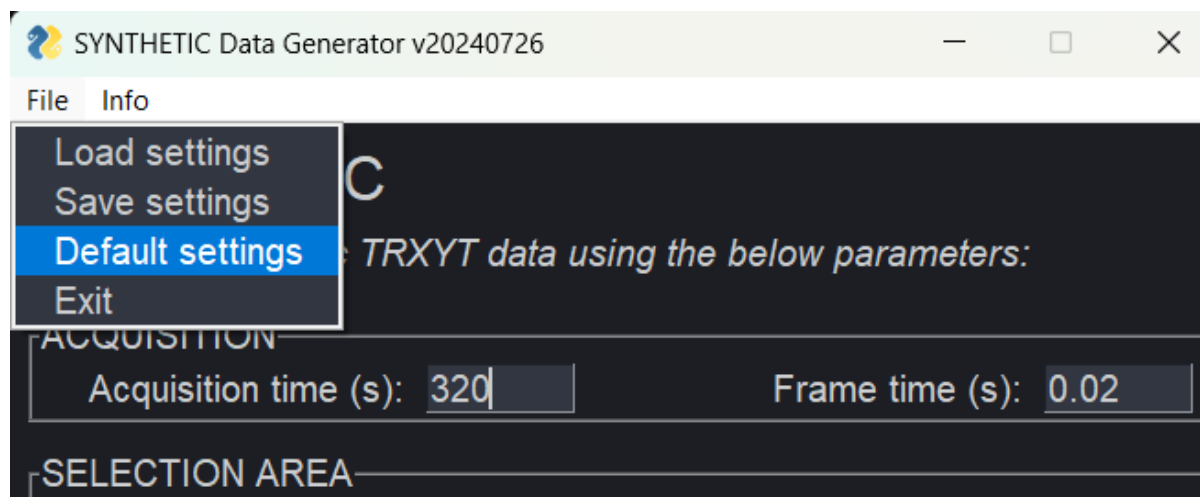


The following message will be displayed in the console:

```
Loading GUI settings from synthetic_data_generator_gui.defaults...
```

## Default settings

If you wish to restore values to the original default values, click 'File >> Default settings' in the menu bar. Subsequently click 'File >> Save settings' to save the original default values as the default settings.



The following message will be displayed in the console:

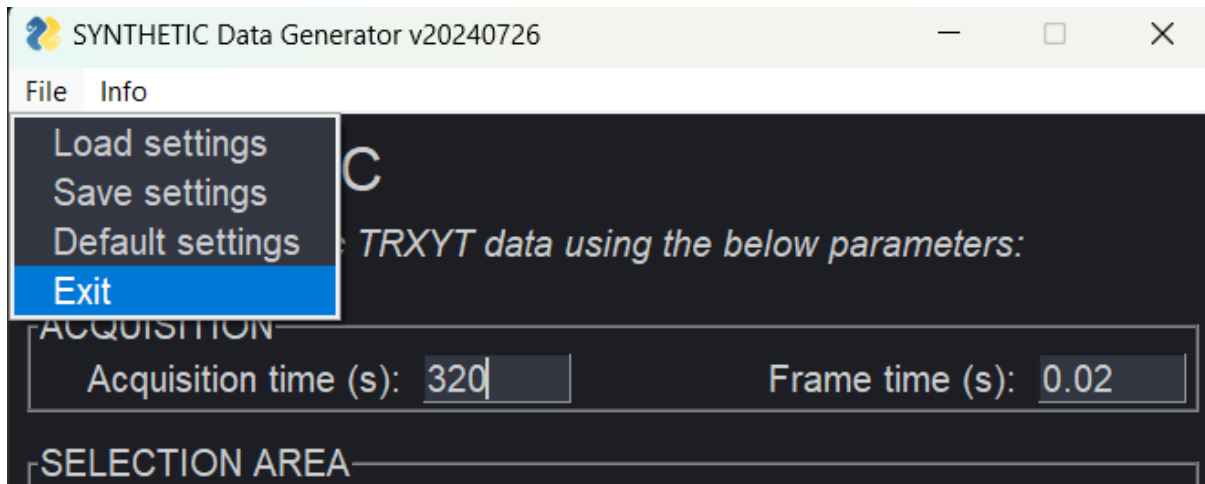
```
Using default GUI settings...  
Saving GUI settings to synthetic_data_generator_gui.defaults...
```

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



## Exiting the program

The program can be exited either by clicking 'File >> Exit' in the menu bar, clicking the 'x' in the top right corner of the window, or by typing Ctr + C in the terminal.

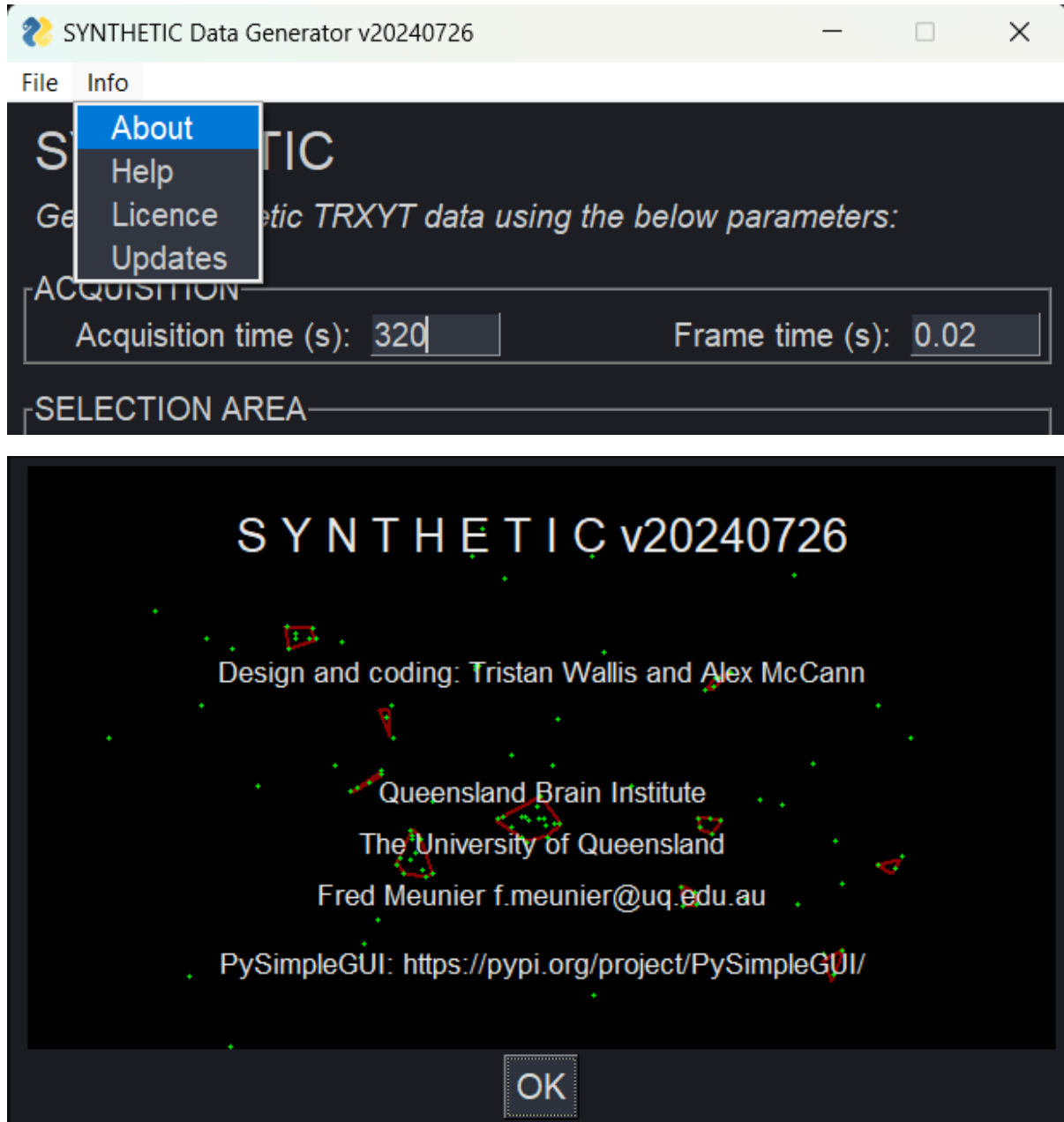


The GUI will close, and the console will display the following message:

```
Exiting...
```

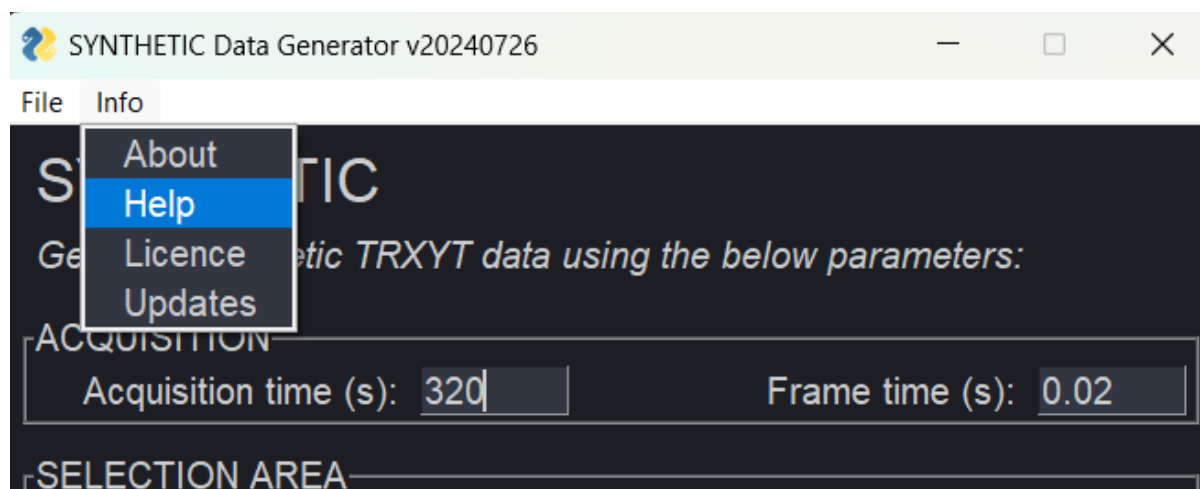
## About

Clicking 'Info >> About' in the menu bar brings up a splash screen showing the about information for the GUI.



## Help

Detailed information about each parameter can be viewed by clicking 'Info >> Help' in the menu bar of the GUI.



### Help

This program generates synthetic .trxyt files based on parameters defined by the user, for use in nanoscale spatiotemporal indexing clustering (NASTIC and segNASTIC) or 3D DBSCAN (BOOSH).

ACQUISITION - parameters defining the time dimensions of the pretend acquisition data.

SELECTION AREA - parameters defining the spatial dimensions of the pretend acquisition data.

Option 1: use x-axis size (um) and y-axis size (um)

Option 2: use previously saved NASTIC ROI file (roi\_coordinates.tsv).

CLUSTERED TRAJECTORIES - parameters defining the behaviour of clustered trajectories.

Seed = point at which a single cluster will form.

Step length = step size within a clustered trajectory.

UNCLUSTERED TRAJECTORIES - parameters defining the behaviour of unclustered trajectories.

Noise factor = value multiplied by the seed number to get the number of unclustered trajectories per seed

Step length multiplier = value multiplied by the clustered trajectory step length to obtain the step size within unclustered trajectories.

HOTSPOTS - parameters defining the behaviour of hotspots.

Hotspot = point at which multiple spatially overlapping but temporally distinct clusters form.

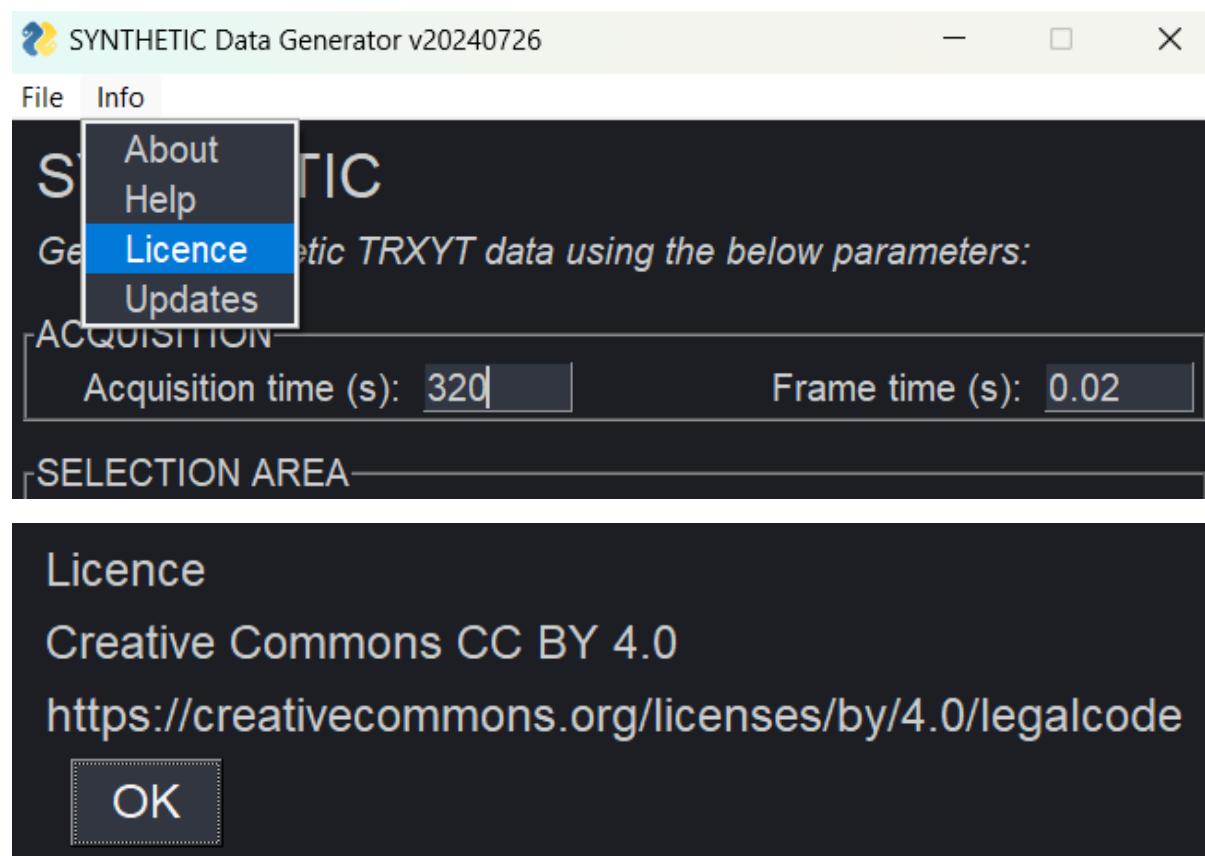
GENERATE TRXYT - generates synthetic .trxyt files based on the selected parameters, and a corresponding metrics.tsv file containing the parameters that were selected, and the metrics which they generated. Both file types are saved in the same directory as the synthetic\_data\_generator\_gui.py python script, in a folder that is generated with the naming format 'synthetic\_data\_output\_YYYYMMDD-HHMMSS'.

Tristan Wallis, Alex McCann 20240726

OK

## License

Clicking 'Info >> Licence' in the menu bar brings up the license information.



## Updates

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

Clicking 'Info >> Updates' will open a webpage for the GitHub.

