

**GEMSpa**

**USER MANUAL**

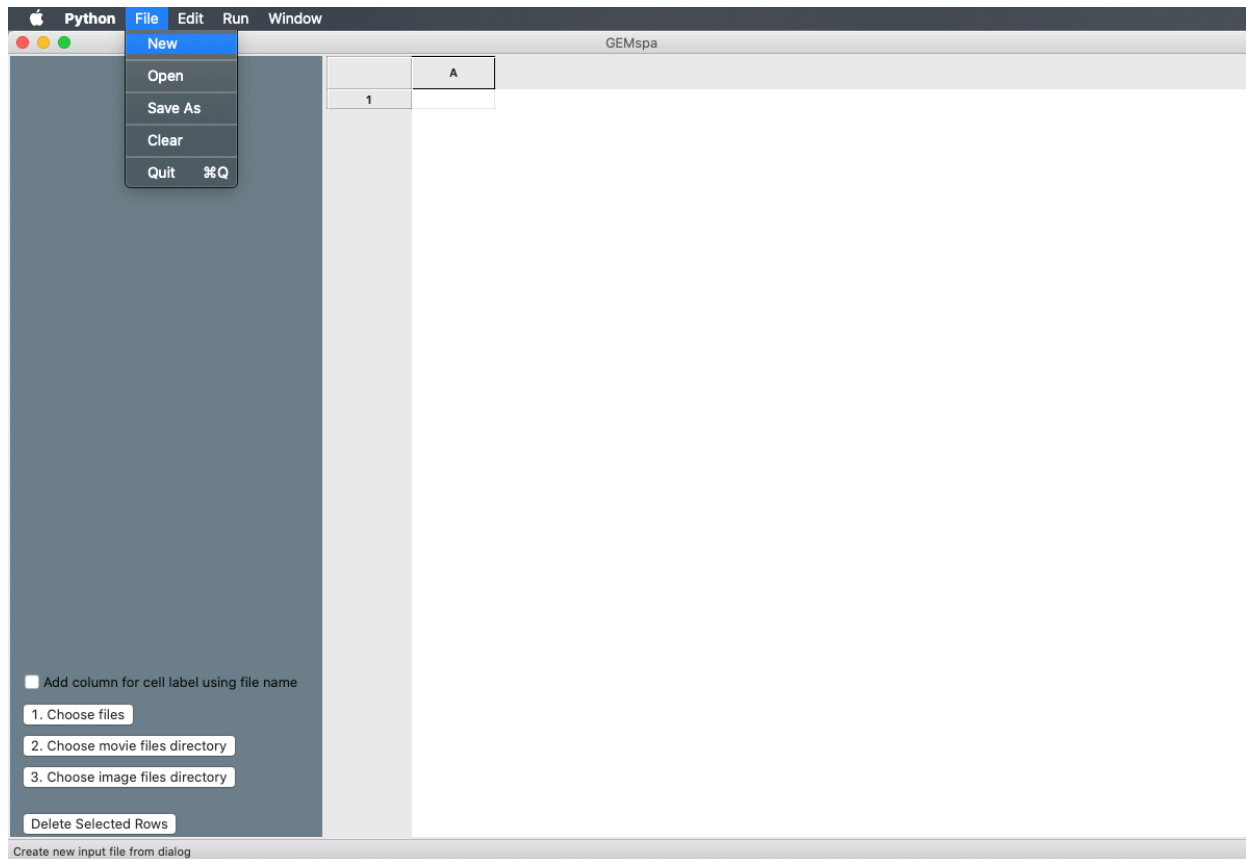
**12/3/2021**

SECTION 1: CREATING THE DATA SUBMISSION FILE	3
SECTION 2: RUNNING THE ANALYSIS	8
SECTION 3: OUTPUT	10
SECTION 4: INSTRUCTIONS FOR OPTIONAL FEATURES	12
SECTION 5: INSTALLATION	13

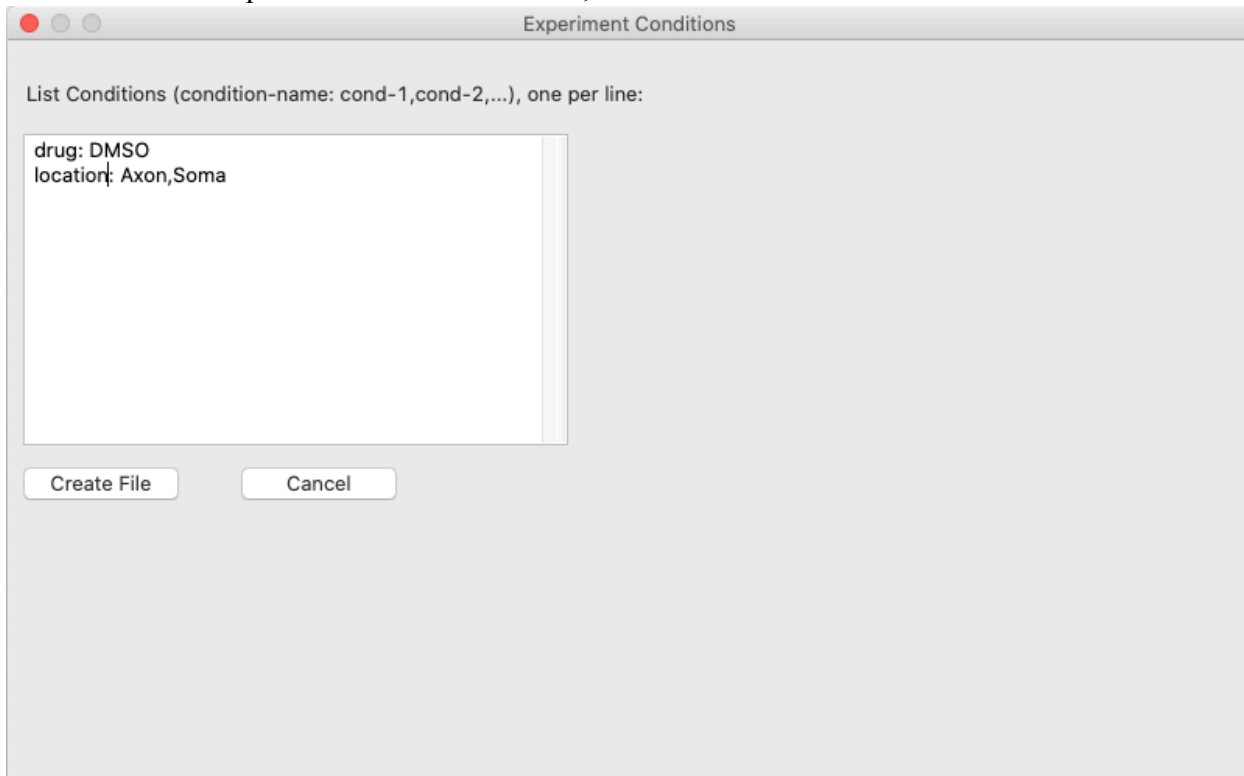
## SECTION 1: CREATING THE DATA SUBMISSION FILE

The first step for running GEMspa is to create a data submission file. This file is a tab-delimited text file with pre-defined headers that contains the list of CSV files (tracking output from MOSAIC), along with the experimental conditions for each group. This file can be created using the GEMspa GUI interface:

### Go To **File** -> **New**



The dialog box for experimental conditions will open. Enter the conditions for your experiment. Follow the format in the example below. When finished, click **Create File**:



The image shows a macOS-style dialog box titled "Experiment Conditions". It has a light gray background and a title bar with three window control buttons (red, yellow, green) on the top left. Below the title bar, the text "List Conditions (condition-name: cond-1,cond-2,...), one per line:" is displayed. A white text area with a vertical scrollbar is positioned below this text, containing the example text "drug: DMSO" and "location: Axon,Soma". At the bottom of the dialog, there are two buttons: "Create File" and "Cancel".

Experiment Conditions

List Conditions (condition-name: cond-1,cond-2,...), one per line:

drug: DMSO  
location: Axon,Soma

Create File Cancel

The column names for the data sheet in the main window will be populated, and drop-down boxes on the left side with the listing of experimental conditions will appear:

drug

DMSO

location

Axon

Soma

Add column for cell label using file name

1. Choose files

2. Choose movie files directory

3. Choose image files directory

Delete Selected Rows

	A	B	C	D	E	F	G
1	id	directory	file name	movie file dir	image file dir	drug	location

Welcome!

The next step is to add your list of CSV files (MOSAIC output files) for each group. Use the drop-down boxes to choose the conditions and click on the button **1. Choose files**. Then, select a set of CSV files and they will be added to the data sheet.

### Grouping data by cell ID:

If you would like to group your data by cell (for example, if you have imaged the same cells at multiple time points and want to see results for each cell over time), GEMspa provides a mechanism to include that information in the output table and visualize results organized in this way. Check the **Add column for cell label using file name** check box *before beginning to add your CSV files to the table*, and see [SECTION 4](#) for instructions regarding the required naming convention for the files.

The screenshot shows the GEMspa application window. On the left, there are two panels: 'drug' with a dropdown menu set to 'DMSO' and 'location' with a dropdown menu set to 'Axon'. Below these is a checkbox labeled 'Add column for cell label using file name' which is currently unchecked. Underneath the checkbox are three buttons: '1. Choose files', '2. Choose movie files directory', and '3. Choose image files directory'. At the bottom left is a button labeled 'Delete Selected Rows'. The main area of the application is a table with the following structure:

	A	B	C	D	E	F	G
1	id	directory	file name	movie file dir	image file dir	drug	location
2	1	/Users/sai	Traj_DMSO_AxonRight_2h_008.csv			DMSO	Axon
3	2	/Users/sai	Traj_DMSO_AxonLeft_2h_013.csv			DMSO	Axon
4	3	/Users/sai	Traj_DMSO_AxonLeft_2h_012.csv			DMSO	Axon

At the bottom of the window, there is a 'Welcome!' message.

Continue to add all CSV files for your experiment by selecting from the drop-down boxes to choose the conditions and clicking on **1. Choose files**.

## Optional Extra Features

At any point while adding your CSV files, you may select a set of files from the list and associate a directory with additional files by clicking one of the following buttons (see [SECTION 4](#) for more details on these features and the required file naming conventions):

- **2. Choose movie files directory** – a directory containing nd2 or tiff GEM movie files for reading time step and scale directly from the meta data
- **3. Choose image files directory** – a directory containing image files for making rainbow tracks and/or ImageJ ROI files/MASK images for filtering MOSAIC tracks by region.

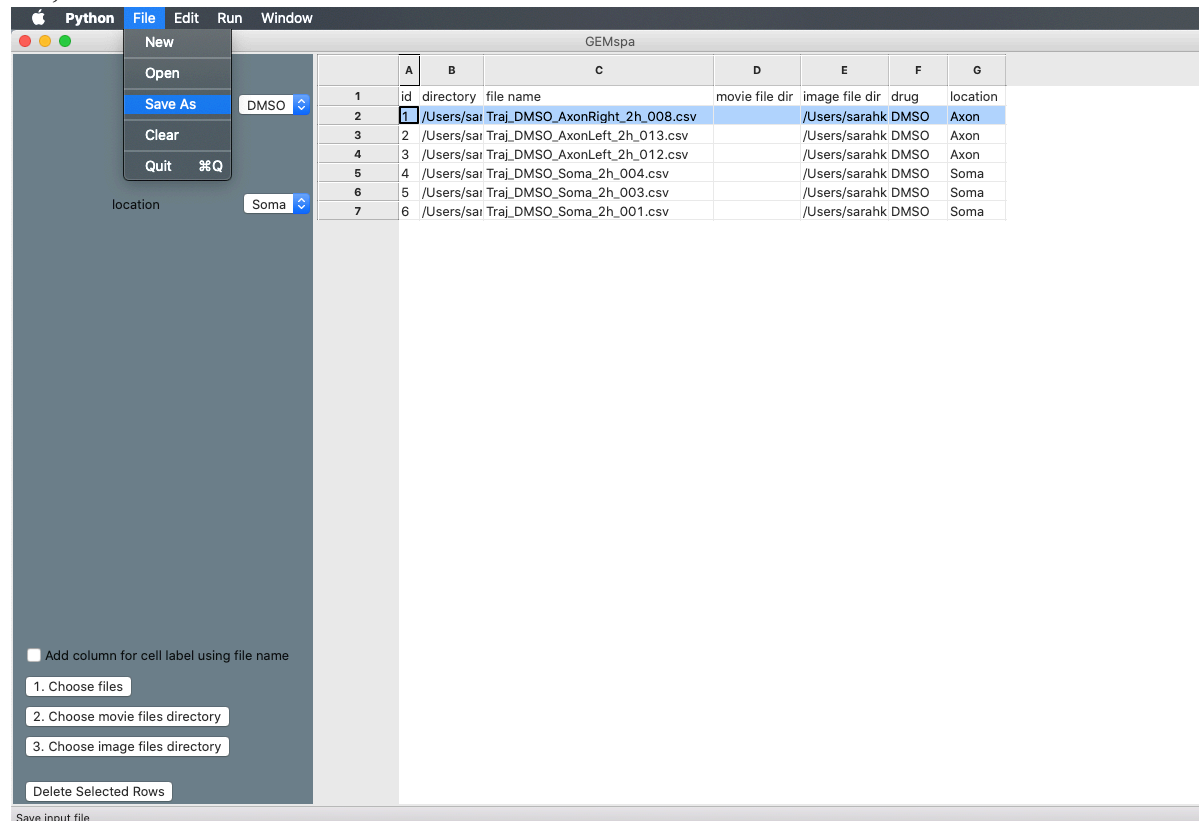
*Note that you must select the set of CSV files in the matrix for which you'd like to add these directories. The program will not automatically add them for all CSV files in the list. (in case files are split between multiple directories). Each time a list of CSV files are added using the **1. Choose files** button, that set will be automatically selected for you.*

## Deleting Rows

You may delete row(s) added by mistake with the **Delete Selected Rows** button. *Be careful not to delete the header row!*

## Saving

When finished adding all of your data, it is important to SAVE this file as a **.txt** file (**File->Save As**). You will need to select the saved version of this file when running the analysis: GEMspa reads directly from the saved file, NOT from the data in the editor.



## SECTION 2: RUNNING THE ANALYSIS

After creating and saving the data sheet for your experiment, you are ready to run the analysis. In the GEMspa interface, click on **Run -> GO**. The Run Dialog Box will open. Here it has been annotated with brief descriptions for the parameters. Section 4 provides further details for the parameters associated with the optional features: movies with uneven time-steps, rainbow tracks, ROIs for filtering tracks



Run GEM Analysis

Enter directory to save the results:

/Users/sarahkeeqan/Dropbox/mac\_files/holtlab/qems\_code/GemSp  Folder for results

Enter filename for input file:

/Users/sarahkeeqan/Dropbox/mac\_files/holtlab/qems\_code/GemSp  Data sheet created in previous step.

Time between frames (s):  Frame rate of the GEM movie.

Scale (microns per px):  Length scale of the GEM movie.

Min. track length (effective Diff):  Filter tracks < this length when finding eff-D for individual tracks.

Max t-lag (effective Diff):  Cutoff in t-lag ( $\tau$ ) to be used for fitting equation (1) for D.

Min. track length (ensemble average):  Filter tracks < this length when calculating time-ensemble average MSD.

Max t-lag (ensemble average, anomalous exp)  Cutoff in t-lag ( $\tau$ ) to be used for fitting TE-MSD with equation (2) for K,  $\alpha$ .

Min track length (step size/angles):  Tracks less than this length will be excluded for step sizes/angles calculation

Max t-lag (step size/angles):  Steps sizes/angles will be calculated for t-lag ( $\tau$ ) from 1 to this maximum.

Time step tolerance (uneven time steps) (s):  Difference in time step permitted from the minimum (for uneven time steps).

Min D for filtered plots:  For plots labeled “\_filtered”, tracks with Deff below min are removed.

Max D for filtered plots:  For plots labeled “\_filtered”, tracks with eff-D above max are removed.

Max D for rainbow tracks:  eff-D > this value will be set to this value.

Max step size for rainbow tracks (microns):  Step sizes > this value will be set to this value.

Prefix for image file name:  Optional prefix for image file name (can be left blank if no prefix)

☐ Use movie files to read scale/time-step  Check for uneven time steps  Draw rainbow tracks on image files  Use ImageJ ROI files to filter tracks

MSD = (2d) mean square displacement

MSD( $\tau$ ) =  $4D\tau$  eq. 1, D (eff-D) = effective Diffusion coefficient ( $\text{um}^2/\text{s}$ )

MSD( $\tau$ ) =  $4K\tau^\alpha$  eq. 2,  $\alpha$  = anomalous exponent, K = generalized Diffusion coefficient ( $\text{um}^2/\text{s}^\alpha$ )

Once you have chosen your parameters, click “Run Analysis.” GEMspa will begin to run. A log file will be output in the results directory where you can view the status of the run and any errors encountered. Eventually, the full output will be saved to the results directory. See SECTION 3 for details on all output.

## SECTION 3: OUTPUT

GEMspa will output the following data files in tab-delimited format (.txt):

**summary.txt** – results summarized for each input csv file or ROI. Includes columns in the input data sheet and additional columns as follows:

**roi**: if filtering by ROIs, the ROI name is listed here. If a mask image was used, then the label of the region is listed here. A labeled mask image is output to the **image file dir**

**D\_median**: median eff-D over all tracks in the file/ROI; for each track the eff-D is calculated by fitting eq. 1 to the MSD values

**D\_median\_filtered**: median eff-D over all tracks where  $\min \leq \text{eff-D} \leq \max$  from Run parameters (“Max/Min D for filtered plots”)

**D\_mean**: mean eff-D (as above)

**D\_mean\_filtered**: mean filtered eff-D (as above)

**num\_tracks**: number of tracks for this file/ROI

**num\_tracks\_D**: count of tracks with track length  $\geq \min$  from Run parameters (“Min. track length (effective Diffusion)”)

**area**: if filtering by ROIs, the ROI area (in microns)

**ensemble\_D**: eff-D calculated by fitting eq. 1 the time-ensemble average MSD values

**ensemble\_r\_sq**:  $r^2$  for the fit of eq. 1

**ensemble\_loglog\_K**: generalized Diffusion coefficient calculated from fitting eq. 2 to the time-ensemble average MSD values

**ensemble\_aexp**: anomalous exponent calculated from fitting eq. 2 to the time-ensemble average MSD values

**ensemble\_loglog\_r\_sq**:  $r^2$  for the fit of eq. 2

**group**: concatenation of the conditions columns for this row (used for making/labeling plots)

**all\_data.txt** – results for each track in each file or ROI. Includes the columns in the input data sheet, with additional columns as follows:

**roi** (see above)

**D\_median** (see above)

**D\_median\_filtered** (see above)

**D\_mean** (see above)

**D\_mean\_filtered** (see above)

**D**: eff-D calculated by fitting eq. 1 to the MSD values of the current track

**err**: one standard deviation error on the estimated parameter (D)

**r\_sq**:  $r^2$  for the fit

**rmse**: root mean square error (residuals)

**track\_len**: the length of the current track

**D\_max\_tlag**: max tlag considered for fitting eq. 1

**group** (see above)

**all\_data\_angles.txt** – all angles listed by t-lag, for each file/ROI. Relative angle defined as in Burov, et al. PNAS 2013.

**all\_data\_step\_sizes.txt** – all steps sizes (distance between successive points in each trajectory) listed by t-lag, for each file/ROI. In microns.

**all\_data\_track\_lengths\_and\_Rg.txt** – for each track, the corresponding track length and radius of gyration is listed. Radius of gyration as described in: Elliot, et al. Physical Chemistry Chemical Physics 2011.

**NGP.txt** – Non-gaussian parameter calculated for the displacement distributions for t-lag = 1, 2, 3. It is related to the 4<sup>th</sup> order moment of the distribution (kurtosis), which is a measure of the “tailed-ness” of a distribution. (For Brownian motion, displacement distribution should be Gaussian.)

GEMspa will also output the following plots in pdf format:

**summary\_D\_median.pdf** – median Deff is shown per file/ROI for each sample group as a box plot

**summary\_D\_median\_filtered.pdf** – same as above, but filtered based on min/max Deff from Run parameters

If Rainbow tracks option is selected, GEMspa will output the following images:

**<image\_name>\_tracks\_Deff.tif** – “rainbow tracks” drawn on the designated image file. Tracks are colored by Deff.

**<image\_name>\_tracks\_ss.tif** – “rainbow tracks” drawn on the designated image file. Tracks are colored by step size.

## SECTION 4: INSTRUCTIONS FOR OPTIONAL FEATURES

### Grouping by cell ID

If you would like to group your data by cell (for example, if you have imaged the same cells at multiple time points and want to track each cell over time), GEMspa provides a mechanism to include that information in the output table and visualize results organized in this way. While creating the data submission file ([SECTION 1](#)), check the **Add column for cell label using file name** check box *before beginning to add your CSV files to the table*. **File naming:** your CSV files must contain a suffix indicating cell ID, separated by an underscore ('\_') from the rest of the file name, e.g. '<filename>\_A.csv' for cell A, '<filename>\_B.csv' for cell B, etc. GEMspa will extract the cell ID from the file name and create a new column in the data sheet named "cell."

### Reading Time step and scale directly from GEM movie meta-data

If you have GEM movie files (nd2/oem-tif) for reading the time step and scale information from the meta-data, you can select rows from the data sheet and then click on the button **2. Choose movie files directory** to choose the directory where the movie files are located. *Note that for uneven time step filtering, only nd2 movie files are acceptable. OEM TIF files may not have complete information on individual time steps.*

### Time step filtering of tracks for uneven time steps

If your GEM movie may contain time steps that are unequal, GEMspa will filter to include only tracks (or portions of tracks) where the time between frames is within a certain range relative to the minimum time between frames for the movie. This range is defined in the Run parameter called "Time step resolution." For example, if time step resolution is 0.005s and the min. time step is 0.012s, then any frames with time step > 0.017s will be considered as invalid and these frames will be removed from a track. The largest continuous portion of a track not containing any invalid frames will be set as the new, fixed track.

### Rainbow Tracks and filtering by ROI/MASK

If you have an additional image (e.g. a DNA stain), GEMspa can draw 'rainbow tracks' on this image for visualizing both the effective Diffusion coefficient (Deff) of each track and the step-sizes. Select rows from the data sheet and then click the button **3. Choose image files directory** to choose the directory where the image files are located. If you want to limit the tracks that are analyzed from the CSV file to only tracks within an ROI, the ROI files should be placed in this directory also. Note: ROI files can be accepted as .roi/zip or as a mask (tif). If the ROI file is a mask, it is not necessary for an image file to be present. If the ROI file is .zip/roi, then the corresponding image file must be present.

The color scheme for the Rainbow Tracks is the python "jet" colormap:



The Run parameters "Max D for rainbow tracks" and "Max step size for rainbow tracks" allow you to set the top value for the dark red color. Dark blue corresponds to Deff or step size = 0.

### File naming conventions

*In order for GEMspa to correctly match the movie/image/ROI files with the corresponding CSV files, use the following naming conventions:*

CSV file: Traj\_<filename>.csv

Movie file (nd2 or OEM TIF): <filename>.tif or <filename>.nd2

Image file (2D greyscale TIF image): PRE\_<filename>.tif

ROI file (ImageJ): (PRE\_)<filename>.zip or (PRE)<filename>.roi

or

Image Mask: (PRE\_)<filename>\_mask.tif

**PRE\_:** PRE\_ is an optional prefix that can be part of the image file name. This prefix is entered in the Run dialog box. It can also be left blank. If ROI/Mask files exist, GEMspa will search for them with and without the prefix. Either naming style is acceptable.

**ROI Types:** These ROI types can be read by GEMspa: Rectangle, Rotated Rectangle, Oval, Polygon, Freehand. (Rounded Rectangle, Elliptical and Selection Brush Tool are not supported).

**Mask images:** A mask image may also be used instead of an ROI file from ImageJ. The image mask should be a 2-valued image with 0 for background and 1 or 255 as foreground.

**Output Naming:** In ImageJ, each ROI in the ROI Manager has a name, and it is possible to edit this name. These ROI names will be included in the output data of GEMspa for linking results back to the corresponding ROIs. If an image mask is used, GEMspa will output a labeled image where each region is labeled with an ID number. These ID numbers will be included in the GEMspa output tables.

## SECTION 5: INSTALLATION

- 1) Download and install python version 3.7 from <https://www.python.org/downloads/>
- 2) Clone (download) the github repository at <https://github.com/liamholtlab/GEMspa>
- 3) Create a virtual environment – an isolated python environment - for running GEMspa.

### Mac Instructions

- i) Open Terminal by going to Launchpad and typing “Terminal” in the Search Box.
- ii) Since Mac OS comes pre-installed with python, we can check to ensure the python that was installed in Step 1 is being used for our GEMspa installation.

At the Terminal prompt (\$), type:

which python3

This command should print the location of python 3.7 that was just installed (usually it is installed at:

/Library/Frameworks/Python.framework/Versions/3.7/bin/python3.7). You can check the timestamp on the file by typing:

```
ls -l /Library/Frameworks/Python.framework/Versions/3.7/bin/python3.7
```

If this does not appear to be the correct installation, you can try:

```
which python3.7
```

In this case, use the name python3.7 in step b. below.

- iii) Install and activate a virtual environment (an isolated installation of python) and install the python packages needed for GEMSpa. In the Terminal, type:

```
python3 -m venv gemspa-python  
source gemspa-python/bin/activate  
python3 -m pip install -r <path-to-GEMSpa>/requirements.txt
```

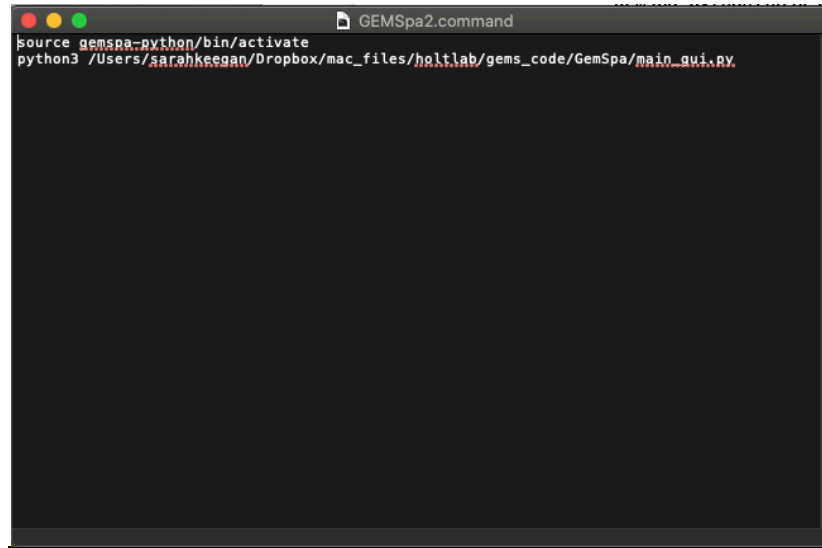
gemspa-python can be any name you want to call the virtual environment – just don't use any special characters or spaces in the name. <path-to-GEMSpa> is the full path to the location where you downloaded GEMSpa.

- iv) Now you are ready to run GEMSpa. GEMSpa can be run from the Terminal by typing:

```
gemspa-python/bin/python3 <path-to-GEMSpa>/main_gui.py
```

- v) You can also create a Desktop Shortcut so you don't have to open Terminal and type in the commands manually:

Open a blank text file with TextEdit. Change the type to "Plain Text" (Format -> "Make Plain Text"). Enter the 2 commands from Step c above (each on a separate line). Save this file as "GEMSpa.command" on your Desktop. Edit the extension so that there is no ".txt" extension. It must end in ".command." Here is an example:



```
source gems-pa-python/bin/activate
python3 /Users/sarahkeegan/Dropbox/mac_files/hgltlab/gems_code/GemSpa/main_gui.py
```

In order to “execute” this file, you will need to grant it execute permissions. Open Terminal, and type:

```
chmod a+x ~/Desktop/GEMSpa.command
```

### **Windows Instructions**

- i) Open Command Prompt by clicking Start and then typing “cmd” into the Search Box.
- ii) Install and activate a virtual environment (an isolated installation of python) and install the python packages needed for GEMSpa. In the Command Prompt, type:

```
python3 -m venv gems-pa-python
gems-pa-python\Scripts\activate.bat
python3 -m pip install -r <path-to-GEMSpa>/requirements.txt
```

gems-pa-python can be any name you want to call the virtual environment – just don’t use any special characters or spaces in the name. <path-to-GEMSpa> is the full path to the location where you downloaded GEMSpa.

- vi) Now you are ready to run GEMSpa. GEMSpa can be run from the Command Prompt by typing:

```
gems-pa-python\Scripts\python3 <path-to-GEMSpa>/main_gui.py
```

- iii) You can also create a Desktop Shortcut so you don’t have to open Terminal and type in the commands manually: