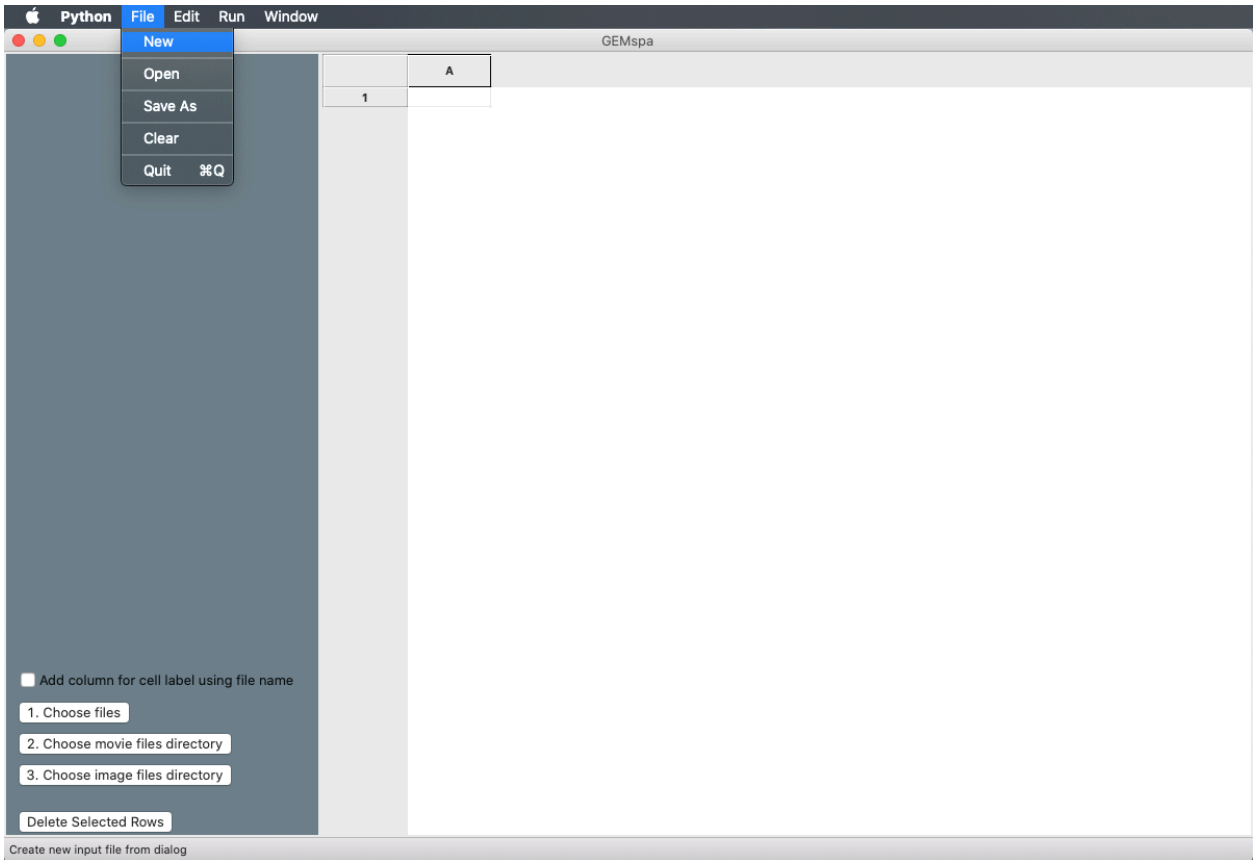
GEMspa USER MANUAL

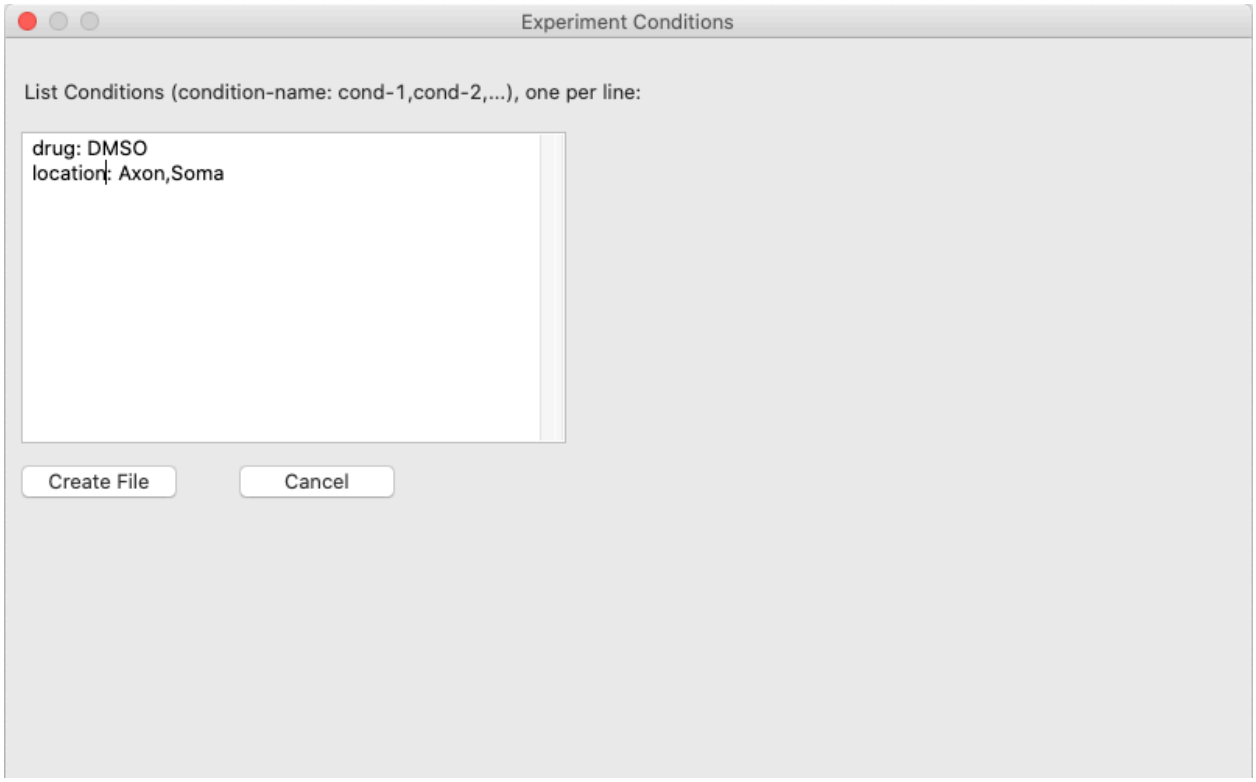
SECTION 1: CREATING THE DATA SUBMISSION FILE

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:

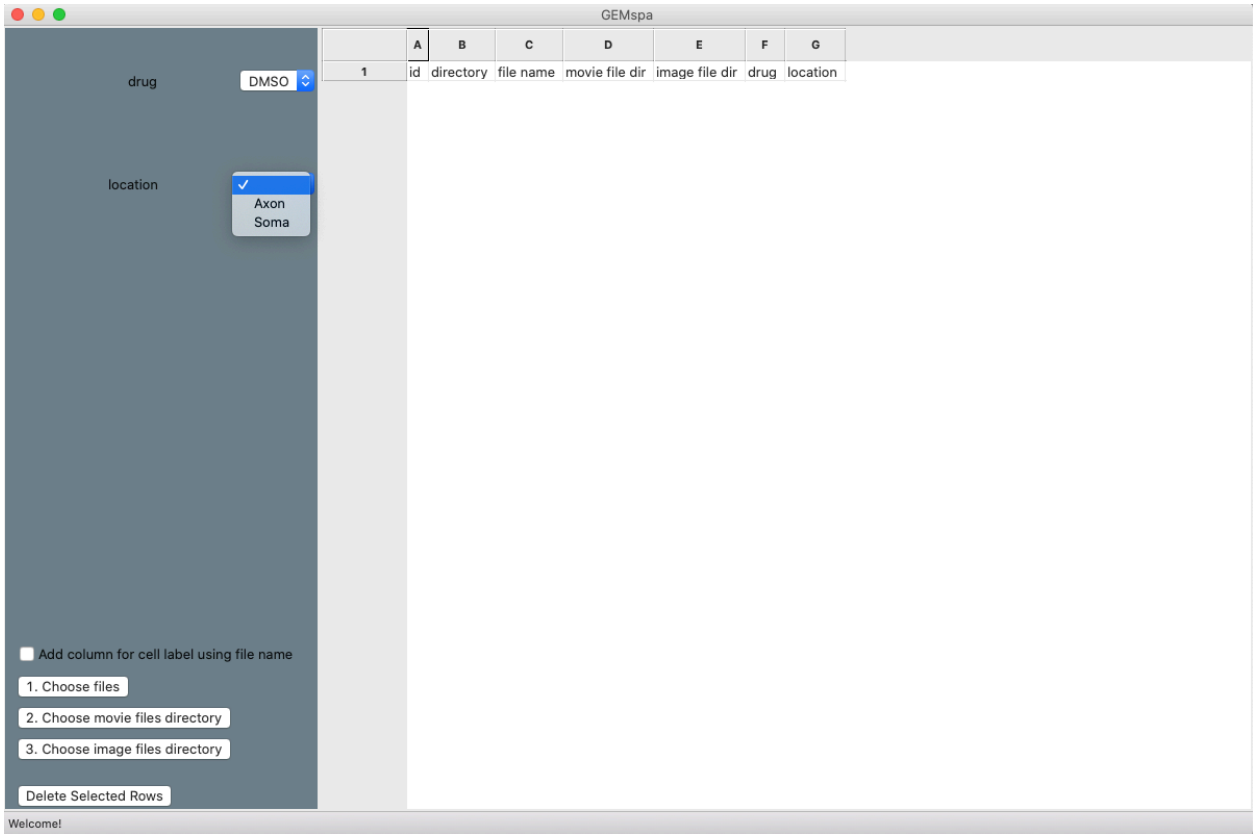
Go To **File -> New**



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**:

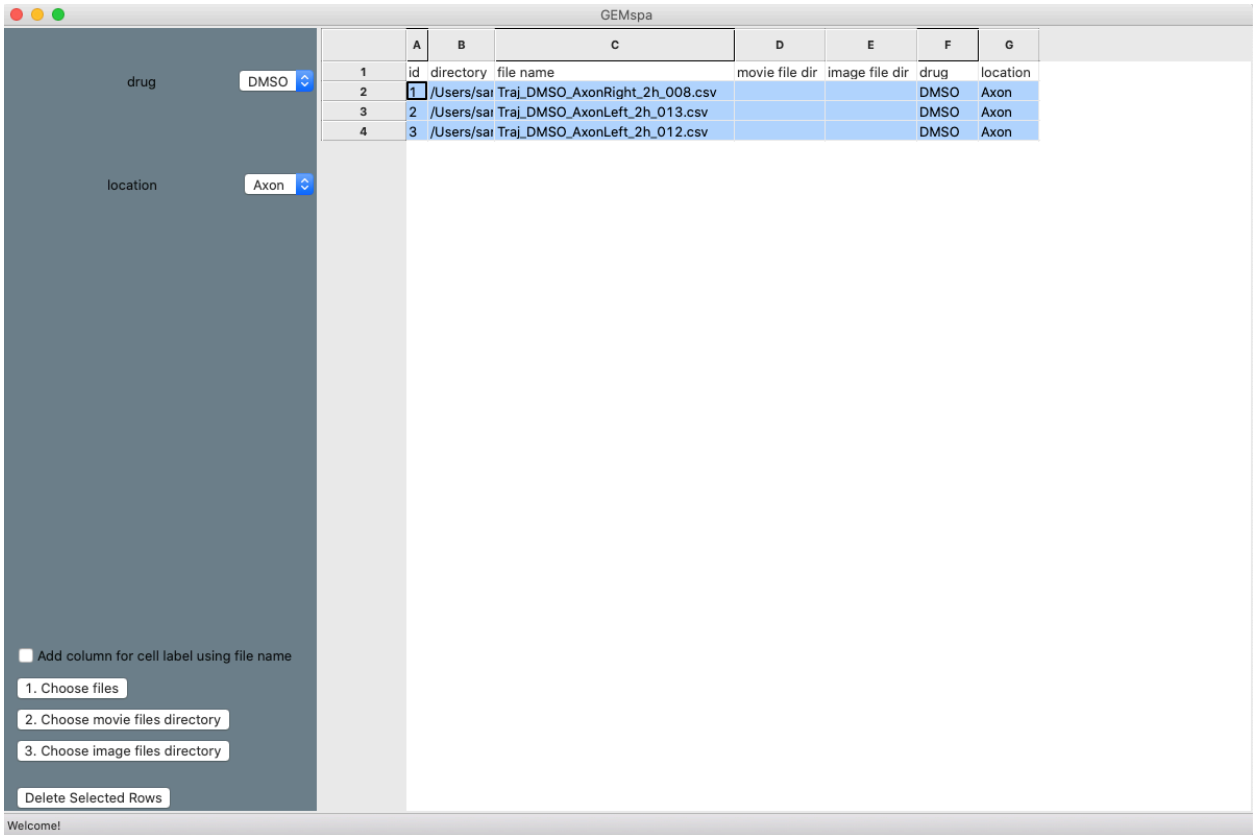


The column names for the data sheet in the main window will be populated, and drop-down boxes on the left side with the listing of experimental conditions will appear:



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.



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.

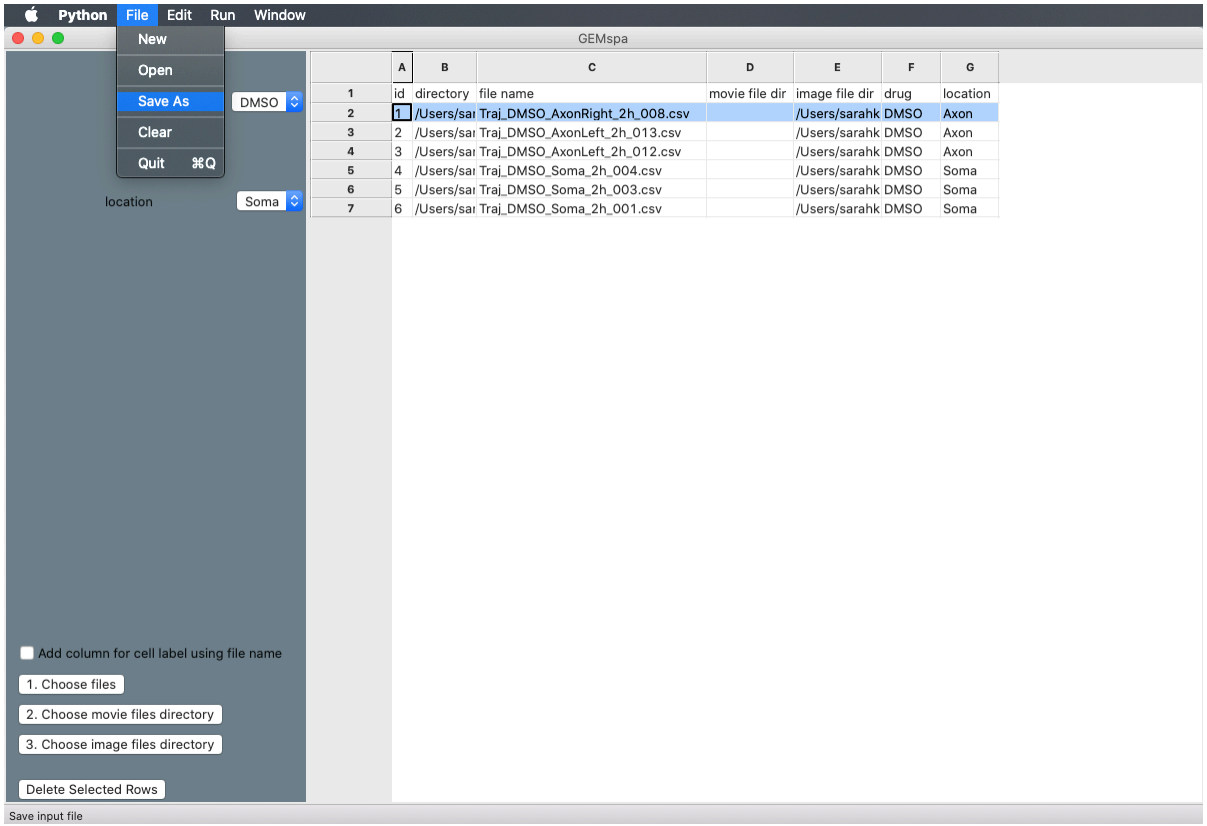
Note that you must select the set of CSV files in the matrix for which you'd like to add these directories. The program will not automatically add them for all CSV files in the list. (in case files are split between multiple directories). Each time a list of CSV files are added using the **1. Choose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

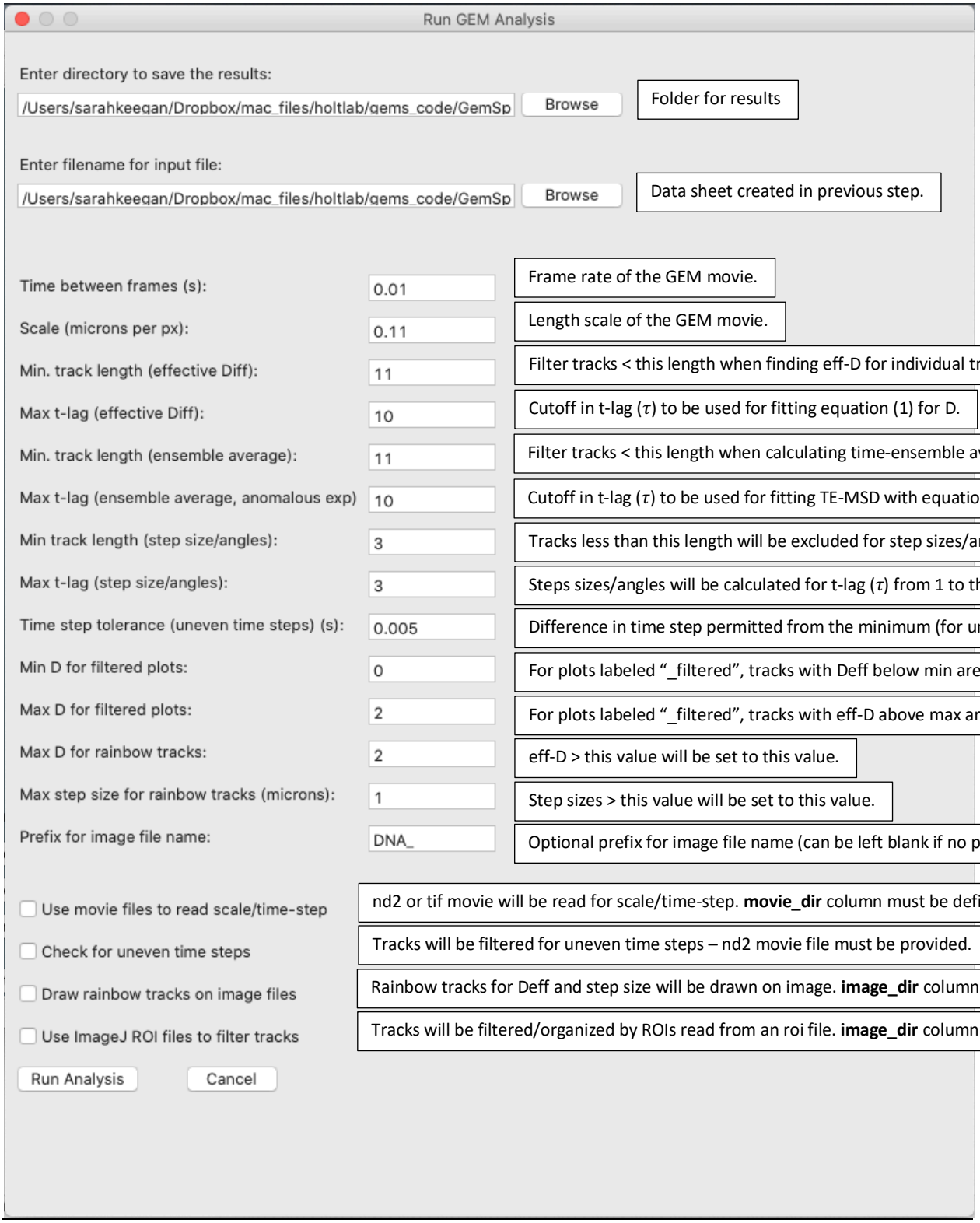
Saving

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



MSD = (2d) mean square displacement

MSD(τ) = 4Dτ eq. 1, D (eff-D) = effective Diffusion coefficient (um^2/s)

MSD(τ) = 4Kτ^{α} eq. 2, α = anomalous exponent, K = generalized Diffusion coefficient (um^2/s^{α})

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 <= eff-D <= 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 >= 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^{th} 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.*

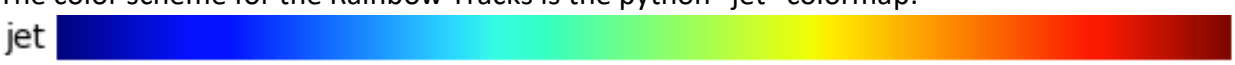
Time step filtering of tracks for uneven time steps

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.