Live-cell cluster analysis: a brief user manual

The overall methodology is presented in the original paper (Griffie, Owen et al SMALL Methods) https://doi.org/10.1002/smtd.201800008 . *This user manual focuses on running the analysis.*

Data formatting:

Both step 1 and 2 enable a reliable and consistent cluster analysis with comparable quality of output.

1: The first step consists in estimating the MPR (minimum point requirement), using fixed-cell data analysis as calibration (number of clusters and percentage of unclustered localisations in particular). The MPR is data set specific and should thus be estimated for each condition. This will ensure the optimal temporal resolution is reached for reliable cluster analysis over all the conditions studied. The full methodology is explained in the original paper.

2: Once the MPR has been estimated, it can be translated in acquisition frames: how many acquisition frames are required to fulfil the MPR? As the number of localisations may vary from one frame to the other, we recommend following a 90% confidence interval (CI) or above (i.e. setting the number of frames to accumulate so that 90% of the generated maps for analysis are above the MPR to avoid discontinuous tracks). The number of frames required can vary from one ROI to the other, and is therefore ROI dependant.

At that stage you will be able to access your temporal resolution for each ROI i.e. the sum of acquisition frames required to achieve the MPR with the user defined CI. For instance, if the MPR is set at 100 points and localisations extracted from 10 frames (30ms) are needed to reach the MPR 90% of time. My temporal resolution will be 10 x 30ms.

3: The last step consists in generating the MPR maps to be analysed by concatenating the previously estimated number of frames with a sliding window approach (add and delete one frame at a time). *We recommend generating about 1000 MPR maps long movie to get reliable descriptors whilst minimizing the computational cost.*

Each MPR map should contain: (x,y) coordinates, the uncertainties and the frame ID from which each localisation has been extracted.

(If this seems complicated, contact juliette.griffie@kcl.ac.uk for formatting codes and associated manual).

4: Similarly to fixed-cell 2D and 3D Bayesian cluster analysis, all the MPR maps to be analysed are saved as data.txt in independent folders, placed in a "movie" folder along with the config.txt file.

An example "Movie 1" folder is provided*. The ROI is set at 2 x 2 μm^2 in our example. The config.txt has to be updated if the ROI size is changed. More details can be found on (Griffie et al Nature Protocols).*

5: Copy the codes from the package (run_live.R, internal_live.R and the postprocessing_live.R)

Running the analysis

1: Open R studio, set the working directory and open the "Run_live.R" code (*If you are ensure, more details can be found in Griffie et al Nature Protocols*).

2: l2 and 3 must be completed accurately depending on your data sets, e.g.:

```
foldernames=c("Movie 1")
MPR<<-77
```

3: Run the code (click on "source")

In addition to running the analysis, cluster descriptors over time such as number of clusters, cluster radii in nm, and % of unclustered localisations will be generated in csv files at that stage.

Post-processing

1: Open "postprocessing_live.R" and complete lines 2-5 accurately

```
foldernames=c("Movie 1")
Starting_frame_id<<-2304
nb_folder<<-20
MPR<<-77
```

"nb_folders" is the total number of MPR maps to process and "Starting_frame_id" is the ID of the first acquisition frame of the movie (found in the data.txt file of folder "1". In the example "Movie 1", it's 2304).

2: Run the code (click on "source")

A Contribution.csv file will be generated. It contains each cluster characteristics tracked over time:

```
#1 frame
#2 cluster ID
#3 Cluster position x
#4 Cluster position y
#5 number of molecules attributed to the cluster
#6 cluster radius (nm)
#7 % of localisations within that cluster coming from newly added points
#8 % of localisations within that cluster coming from points previously attributed to the unclustered pool
#9 % of localisations within that cluster coming from points previously attributed to same cluster
#10 % of localisations within that cluster coming from cluster-cluster interactions: splitting
#11 % of localisations within that cluster coming from cluster-cluster interactions: merging
```

The contribution.csv allows following the clusters over time (both in characteristics and in space), as their attributed IDs remain constant from one frame to the other. **However, it does not connect discontinuous tracks or provide MSD curves. If you would like to have a ready to go tracking routine to extract these information from the "contribution.csv" file, contact juliette.griffie@kcl.ac.uk)**