Testing Consistency of Track Selection and Filter Comparison

The track selection process was modified in order to achieve higher consistency of the separation estimations, and a new track filter was deployed in order to reduce the human effort in the process. However the process still involves a manual step, which could introduce systematic bias and therefore the separation estimates of tracks selected by different people need to be compared. Although the new process has better justification, because it involves more prior knowledge about the FLImP algorithm and the experimental setup, results selected with the new filter need to be tested against the results of the old filter.

In order to make useful comparison between different track selections and different filters it is necessary to know how many tracks are required to distinguish between two separations.

There are 3 tests which need to be made :

1. How many tracks can produce reliable results?

    1.1. How many tracks are required to estimate separation with some pre-defined confidence interval of tracks, which come from the same type of oligomer.

    1.2. How many tracks are required to find out to from which type of oligomer a particular type of track is generated.

2. If different people select tracks from the same data set and analyse them, would the analysis conclude the same separations?

3. If the same data set is analysed with the old filter and the new filter, would the results be the same and if not which result represents the truth better?

The first question can be answered by simulating tracks with the same separation, then using random sub-sets of tracks with different sizes to estimate the separations and the confidence intervals. Similar simulation can be done with two different separations, which correspond to real oligomers and use random subsets to determine how many tracks are required to reliably find out to from which type of oligomer a track has been generated.

The second question can be answered by analysing the same data set by different people and comparing the results. Similarly the third question can be answered by analysing the same datasets with the old and new filter and compare the results.

Finally the performance of the old and the new filter will be compared in terms of number of tracks manually reviewed, number of tracks analysed and number of tracks used. The comparison will be on the same data sets.