Simulations of sptPALM - description of the code

A simulation module was designed in order to test the validity of the sptPALM analysis. The idea was to test the analysis and understand what are the limits of the two methods for calculating the instantaneous diffusion coefficient (fit or weighted average).

The software was simulating only Brownian motion and it was possible to simulate two populations with a pre-defined diffusion coefficient for each of them. The simulation was working the following way :

1. For each frame, the program is calculating the number of proteins that are actually photoactivated. Two parameters are defined for this:
    - the total number of non-activated proteins available = **NProteinsTot**
    - the average number of proteins activated per frame = **MeanProbActivation**

    This is a Bernouilli trial where the number of trials is **NProteinsTot** and the probability of success is p = **MeanProbActivation/NProteinsTot**

2. For each newly activated proteins, their initial X,Y position is calculated using uniform distribution within a square window defined by the number of pixels **ImageSize**.
    - since the activated protein will photobleache at some point, the lifetime is calculated first assuming an exponential distribution with a mean equal to **E=MeanPhotoBleachingTime/AcquisitionTime**
    - in case there are two populations with different diffusion properties, a binomial trial is used to assigned the protein to either one of the two populations and defined its diffusion coefficient **Diff**. The ratio between the two populations is **PopulationRatio**
    - knowing the lifetime of each protein as well as its diffusion coefficient, the complete trajectory can be simulated. The idea is that for each frame, the distance **r** is calculated using a normal distribution N(0,sqrt(4Dt)).
    - finally, since the protein can blink, the trajectory is subdivided into ON/OFF states according to the parameters Ton/Toff1/Toff2. The idea is to simulate mEos2 using typical values for the ON/OFF states we measured for our experiments. Exponential distributions are again used to infer the length of each states. From this step two variables are saved :
        1. The list of positions/frame of each emitters : **DetectionList**
        2. The list of all the trajectories obtained when the emitters are ON : **Trajectories**. In order to build this trajectories, we used the parameters **MaxBlink** (maximum number of frames without detection within the same trajectory) and **MinTrajLength** (minimum number of detection within the same trajectory).

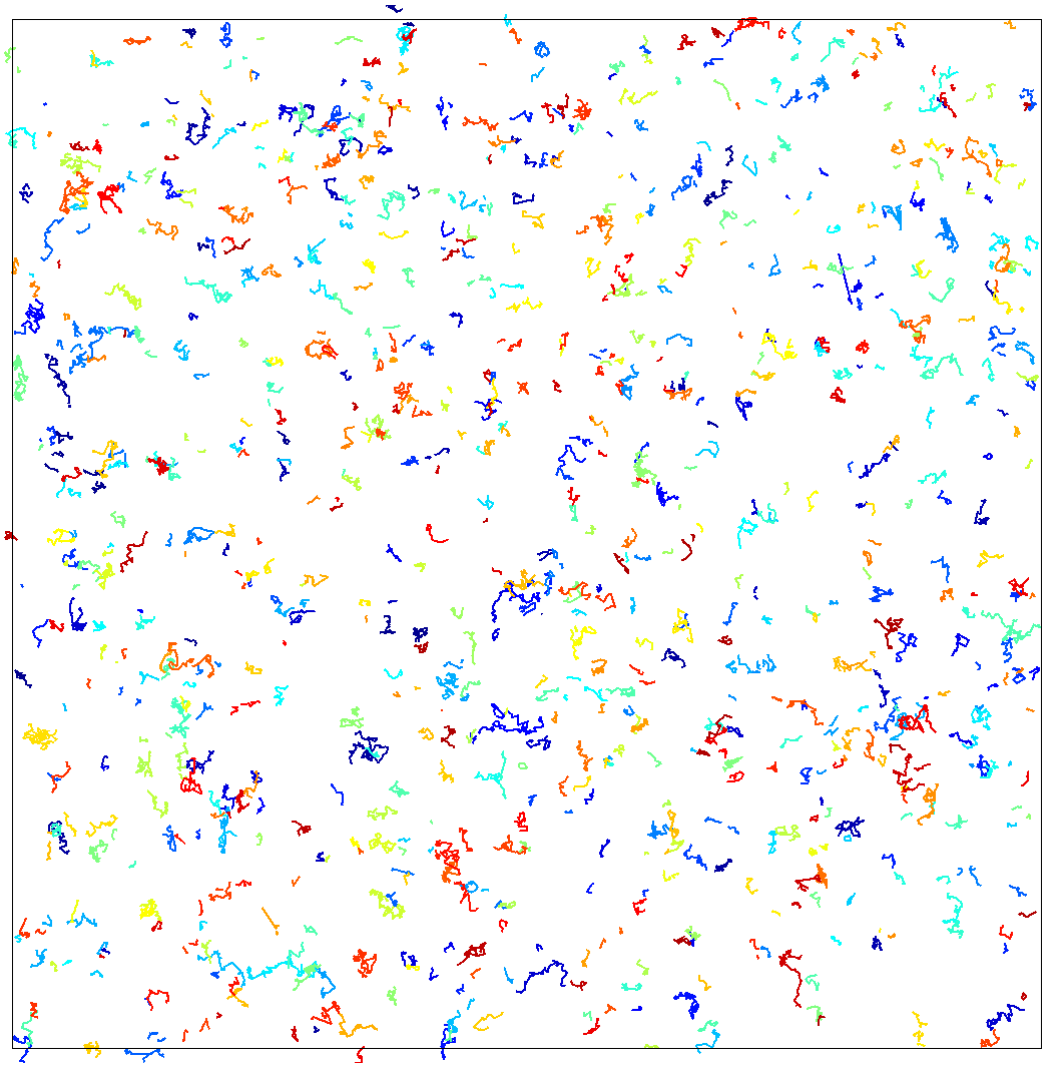
3. The trajectories are then plotted on a figure. Note that it is possible that trajectories are plotted outside the windows since they is no boundary conditions forcing the proteins to stay inside the window.

4. The trajectories are then analyzed as usual, using the two methods. Note that for the fit of the distribution, a Gaussian is used and the binning is fixed to 0.05 in log scale.

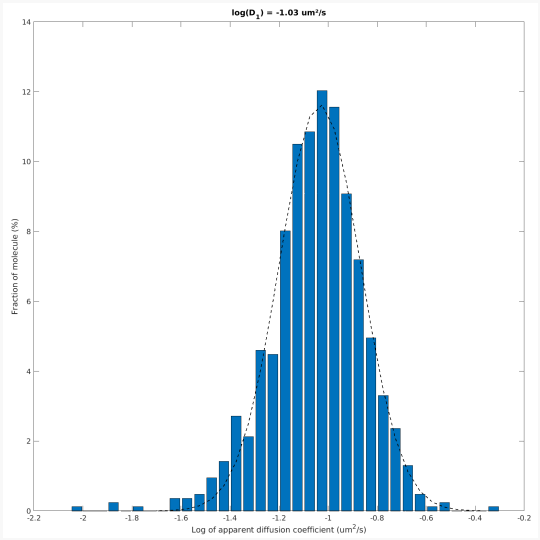
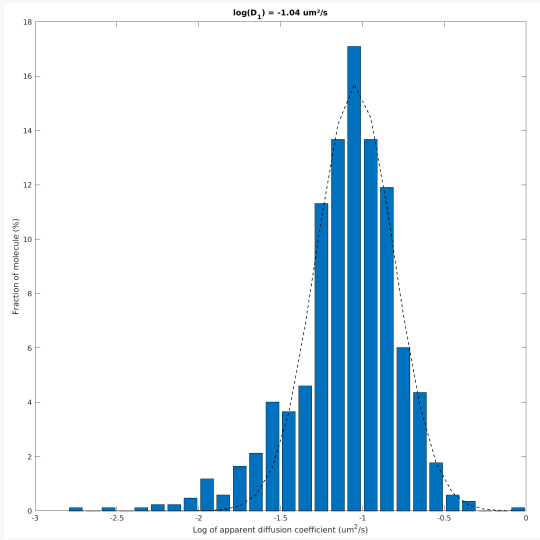
5. Finally the movies are calculated, using the results published by Hirsch et al. 2013. The software was optimized in order to increase the speed of the calculation. Briefly, the images are calculated following this procedure:
    - For each frame, an array with the same size than the image is created. The value of all the pixels is set to 0.
    - using the **DetectionList**, all the emitters that are detected in the frame are selected
    - for each emitter, the 2D distribution of the photon is calculated assuming a Gaussian distribution of standard deviation **LimitResolution**. To do so, a window of size **+/- GaussStampSize** is defined around the x/y position of the selected emitter. Using a 2D Gaussian, the number of photons assigned to each pixel is then calculated using the value of **MeanPhotons** (average number of photons emitted). *Note that we assume that the actual number of photons emitted is following a Poisson distribution.*
    - when the number of photons for each pixel is finally calculated, the A/D counts is inferred using the model derived from Hirsch et al.

First simulations

Below are the first simulation based **only on the simulated trajectories (not the images - step 1&2)**. For example, a population of proteins with a diffusion coefficient of 0.1µm²/s was first simulated - low activation (0.5) and 1000 frames.

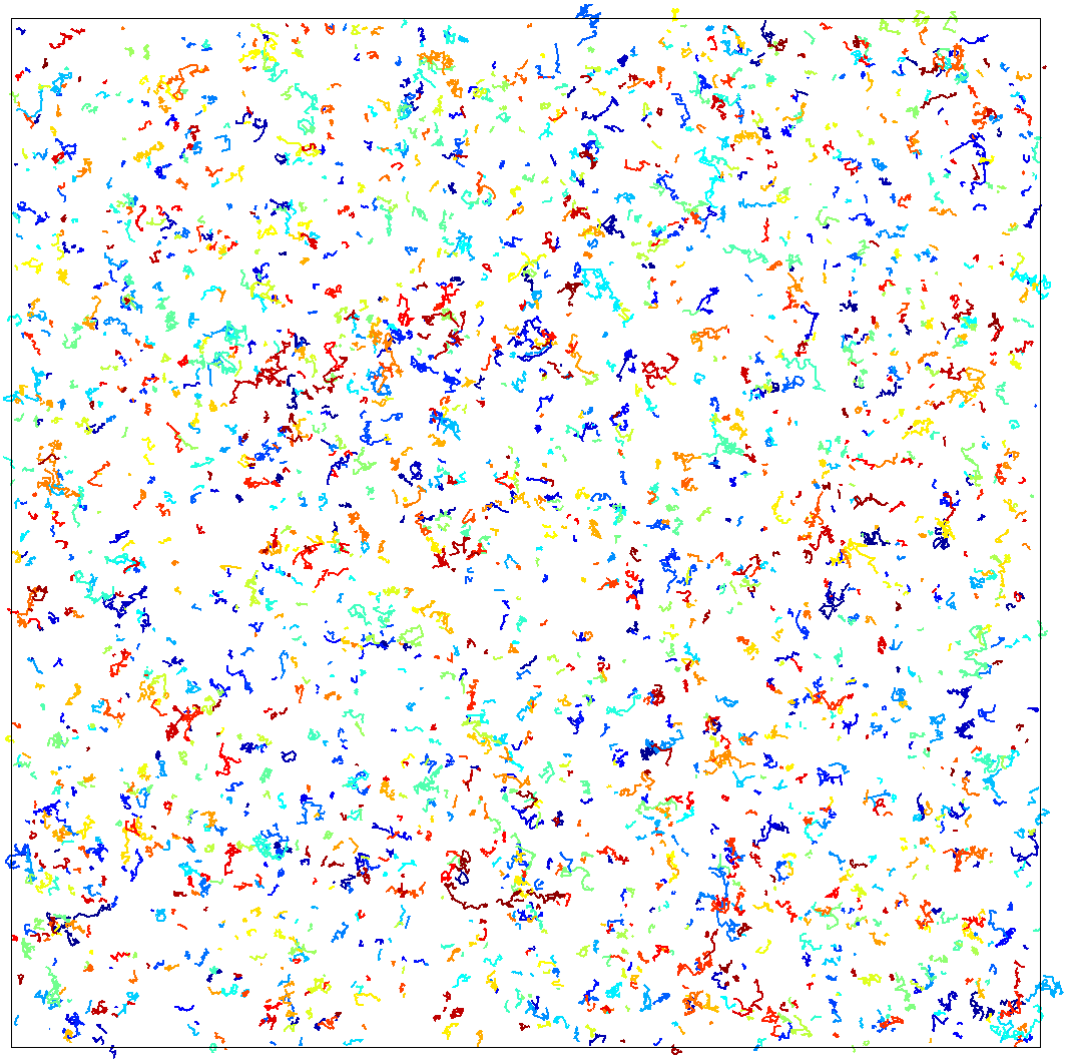


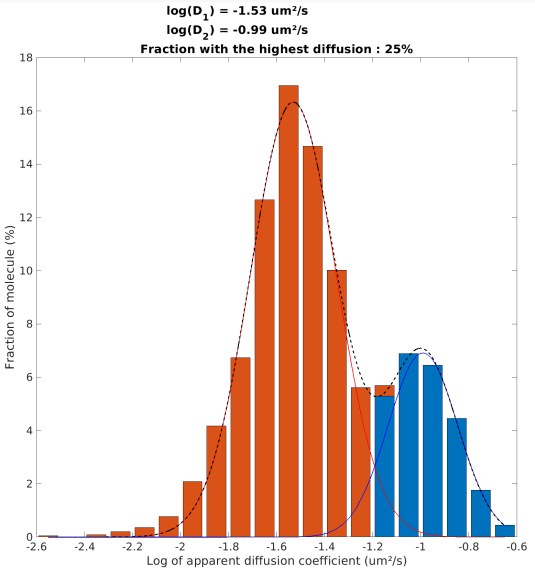
The trajectories are plotted and the resulting diffusion coefficient distributions are displayed below. There are few differences between the two, the weighted average appearing slightly broader. **Note that the functions that are used to calculate the MSD and the diffusion coefficients are the same than the ones used to analyze the MTT/TrackMate data.**

	
Fit method	Weighted average method

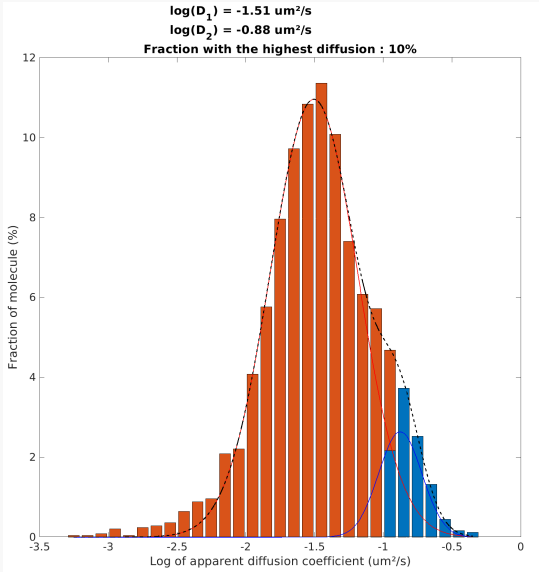
Another example where two populations were mixed :

- The first one (30%) with D=0.1µm²/s (log(D)=-1)
- The second one (70%) with D=0.03µm²/s (log(D)= -1.5)





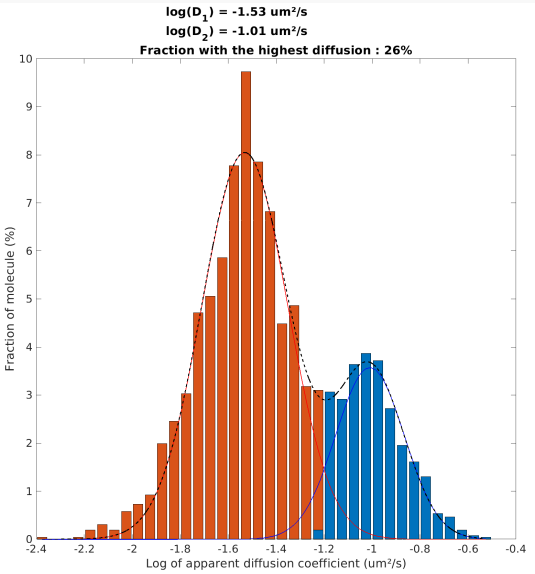
Fit method - bin distance = 0.1



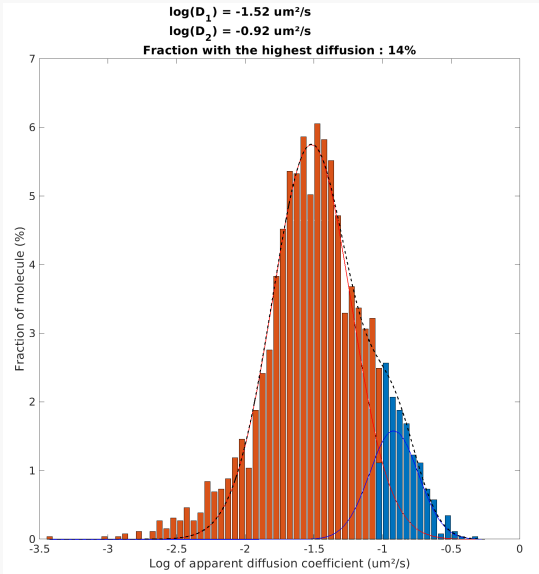
Weighted average method - bin distance = 0.1

In that case, it appears again that the calculation with the fit method is more accurate, the distribution being sharper, allowing for a better estimation of the ratio between the two populations when the diffusion coefficients are close. The same simulation was analyzed after changing the bin size of the histogram from 0.1 to 0.05 but it did not improve the results, the weighted average method leading to a distribution too broad for a proper separation of the two populations.

These results already means than the **Fit distribution, though slow, is more accurate and should be preferentially used when we expect several populations of proteins with different diffusion coefficients**.



Fit method - bin distance = 0.05



Weighted average method - bin distance = 0.05

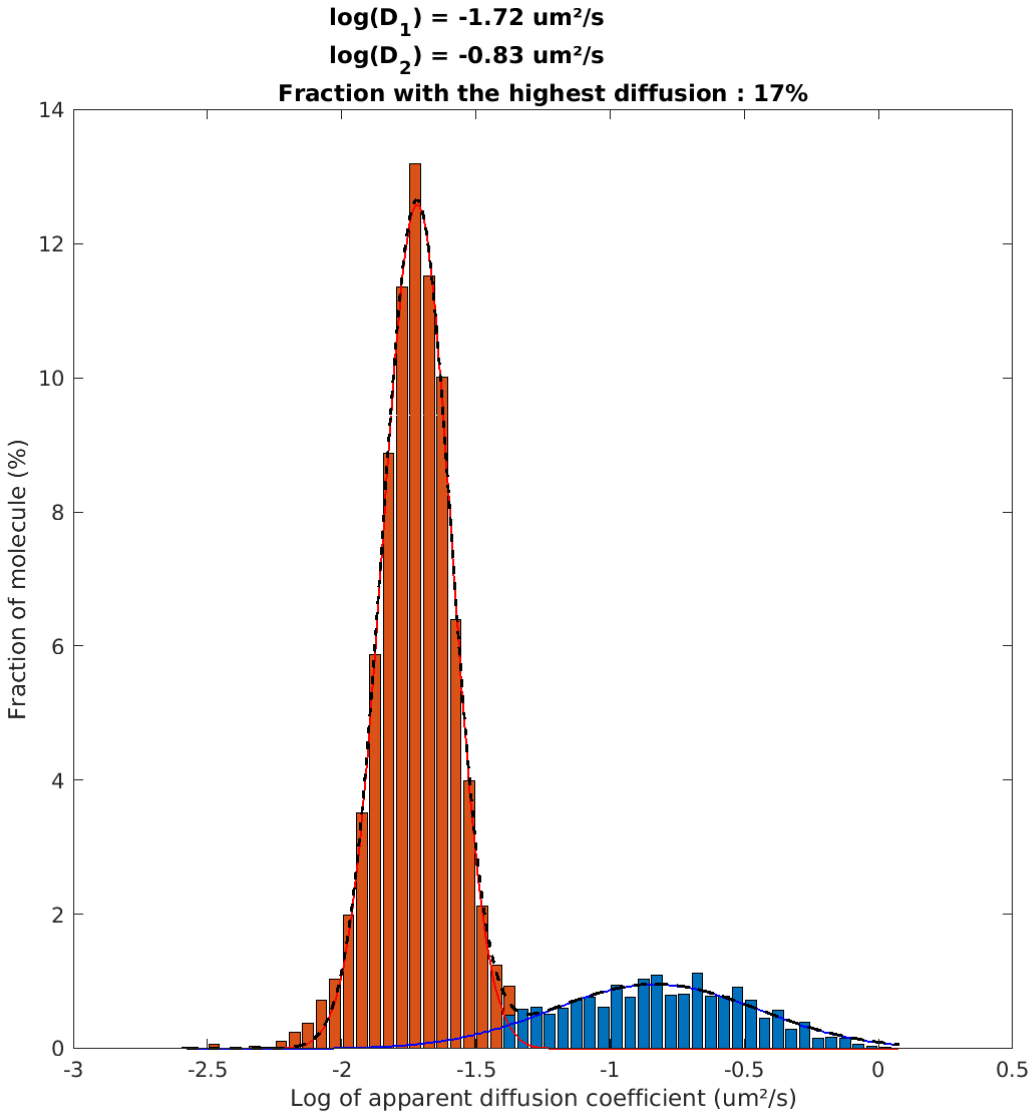
Simulation of a sptPALM movie and MTT analysis

Simulation #1 : Single population - D=0.01µm²/s

A single population was simulated for 3000 frames with the following parameters :

- Mean number of emitted photons = 100
- Average number of activated proteins per frame = 3
- All the other parameters are set to the initial values.

A first analysis was performed using the weighted average of the MSD. Not surprisingly, two populations were detected due to the high number of mis-connections from MTT.



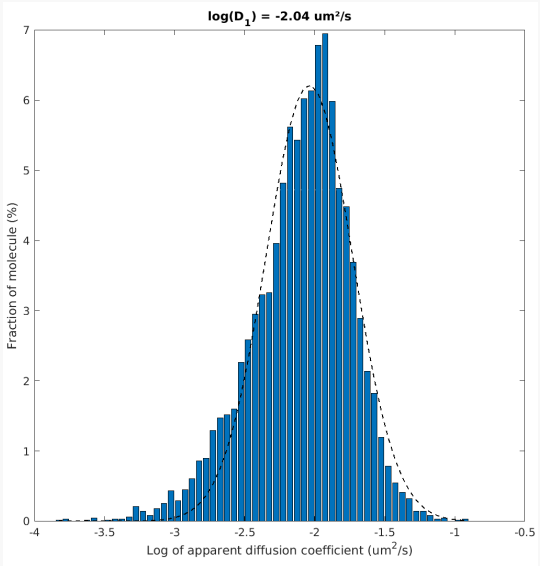
A second analysis was then performed, defining the maximum distance between two consecutive detections at 0.2µm (the average distance between two consecutive detections being 30nm in our conditions). The analysis was then performed using the two analysis methods (see below). Surprisingly, there are several differences between the two distributions :

- for the fit method, the distribution is very broad but the average value is very close to the expected value of D

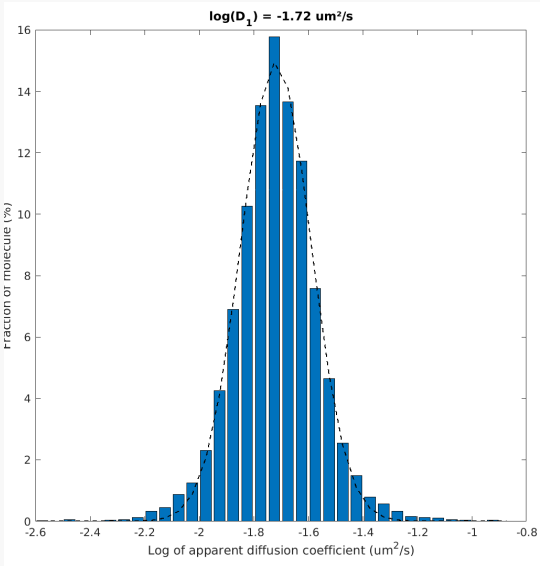
- for the average method, the distribution is sharper with a standard deviation similar to the one estimated directly from the simulated positions. However, the average diffusion coefficient is slightly shifted toward the higher values (0.019µm²/s instead of 0.01µm²/s).

The same simulation was repeated using 200 for the average number of emitted photons. The idea was to check whether the average method was more sensitive to SNR, something that is excepted and was already described by Uphoff et al. 2014 (a correction factor should be added to the diffusion coefficient and is directly related to the precision of localization, therefore the SNR).

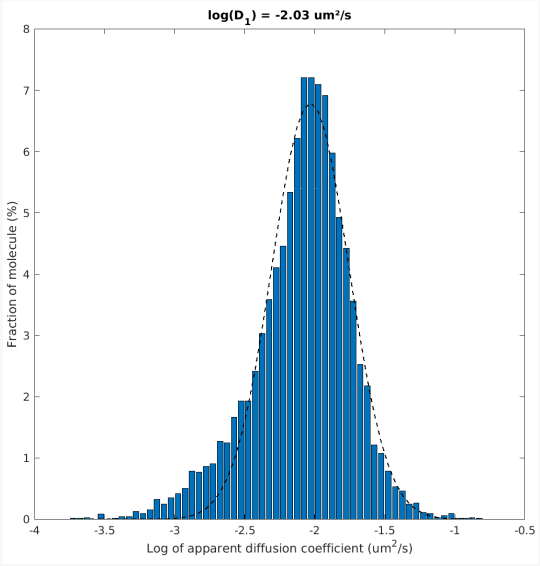
As expected, for a higher number of photons, the result does not change for the fit method but for the weighted average, the diffusion coefficient is closer to the theoretical value (0.0135µm²/s)



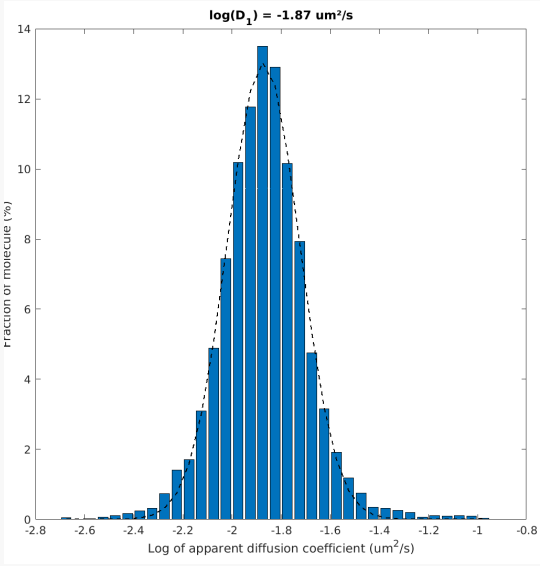
Fit method - Average number of photons = 100



Weighted average method

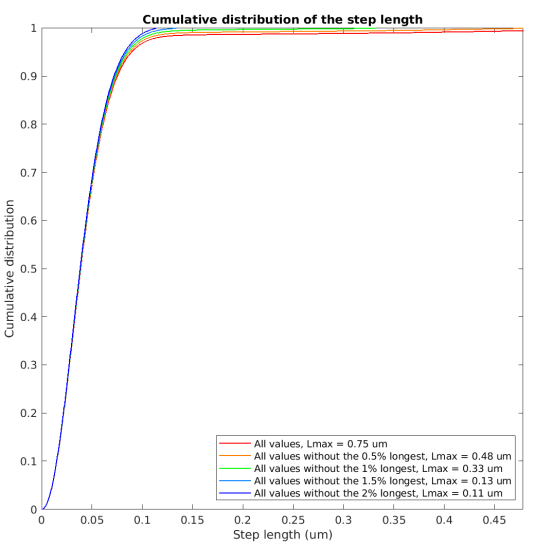


Fit method - Average number of photons = 200

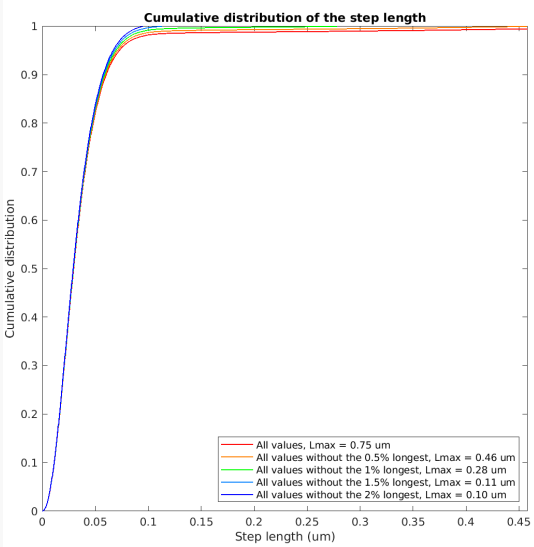


Weighted average method

When looking at the cumulative distribution of the single step distances, the effect of the number of photons (therefore the SNR) is clearly observed. For an average number of photons of 200, the mean distance is 30nm while for 100 photons, this distance increases to 40nm, **suggesting that the detection is partially dominated by the noise.**



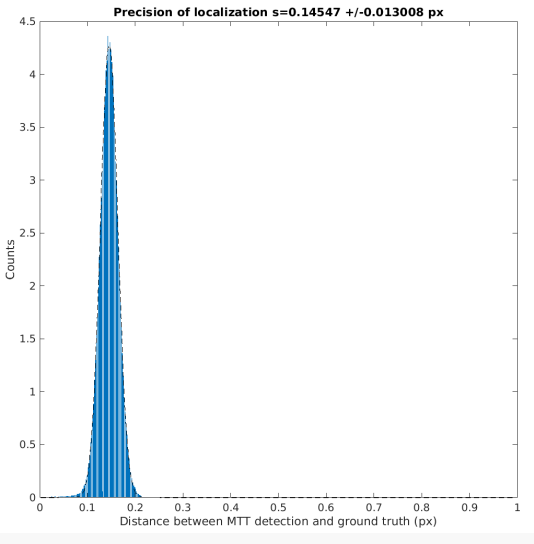
Average number of photons = 100



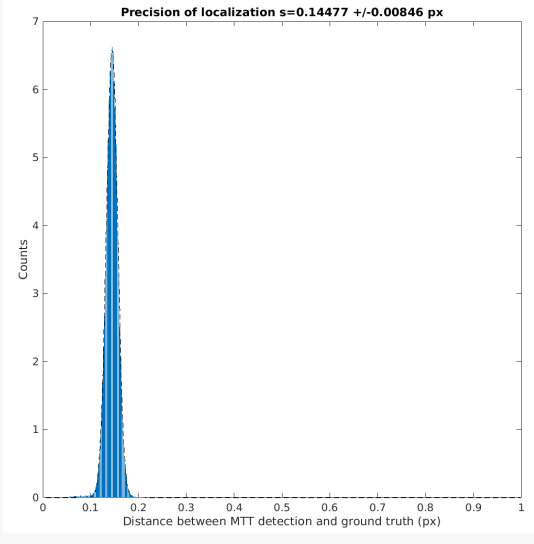
Average number of photons = 200

In order to quantify the effect of the localization precision on the diffusion coefficient estimation, we can directly calculate the localization precision by comparing the MTT detections with the ground-truth positions of the simulated emitters.

When the average number of photons is set to 100, the localization precision for the simulated movies is equal to 15+/-3nm, **which corresponds to an incertitude of 0.011µm²/s on the diffusion coefficient**. For 200 photons this values does not really change, which is a bit surprising (though a precision of localization of 15nm is actually already extremely good for PALM. We are usually more on the order of 30nm, which corresponds to an error of 0.045µm²/s ...). **In any case, what those results suggest is that the weighted average method should not be used when the diffusion coefficient is lower than 0.01µm²/s**.



100 photons

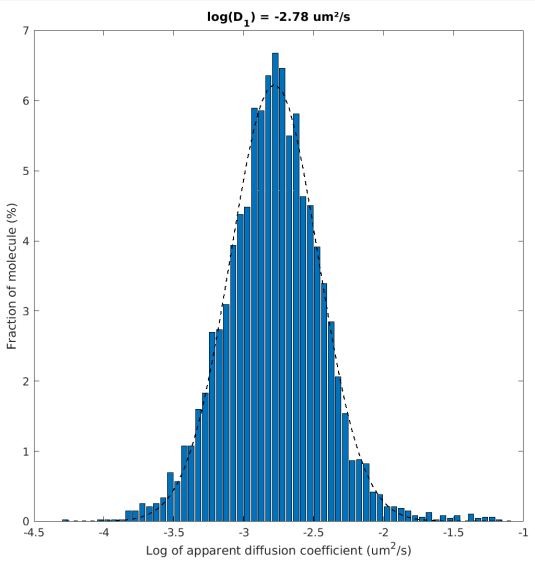


200 photons

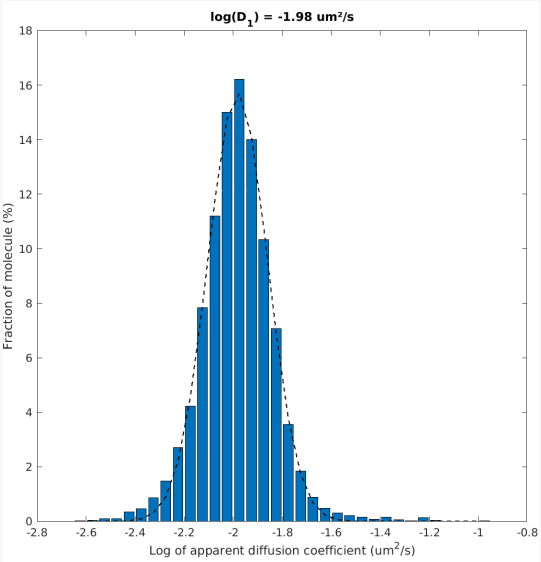
Simulation #2 : Single population - D=0.001µm²/s

A single population was simulated for 3000 frames with the same set of parameters (again mean number of emitted photons is set to 100 or 200). Again the data were analyzed using both analysis methods and the results are showing the same trend that was already observed for simulation #1, though even stronger since the diffusion coefficient is now one order of magnitude lower :

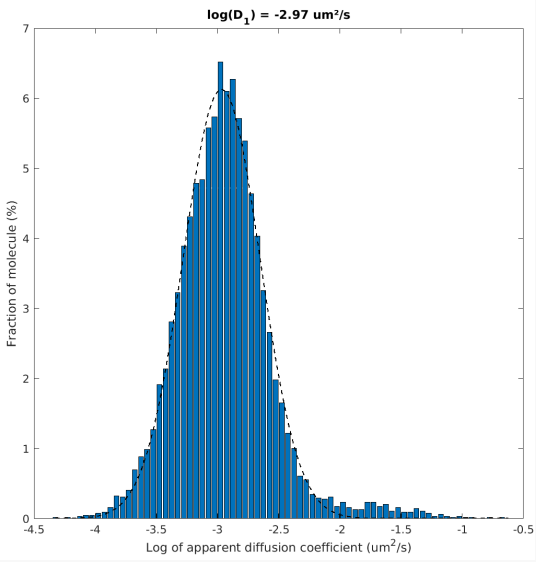
- the fit method appears again more accurate, even if the average value for the diffusion coefficient is slightly off (0.0017µm²/s) when the number of photons is low. However, when the number of photons is set to 200, the estimation of the diffusion coefficient is again accurate. This means that when the number of photons is set to 100, the localization precision is too low and does not allow anymore a good estimation of the molecules displacement.
- for the weighted average however, the estimation is off by an order of magnitude (D = 0.0105 µm²/s) when the number of photons is 100. When the number of emitted photons is set to 200, the results are slightly better and D = 0.0048µm²/s.



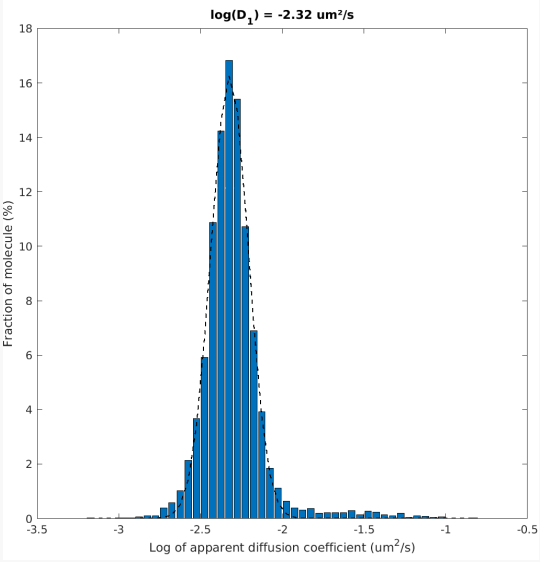
Fit method - Average number of photons = 100



Weighted average method



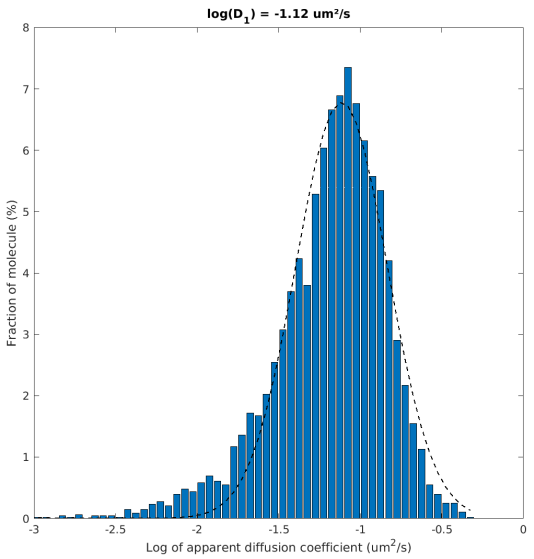
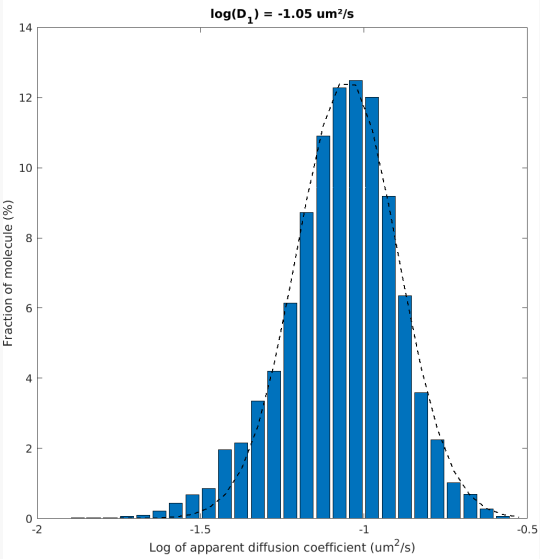
Fit method - Average number of photons = 200



Weighted average method

Simulation #3 : Single population - D=0.1µm²/s

In that case, both methods are returning the expected diffusion coefficient. The reason is that for D=0.1µm²/s, the average displacement between two consecutive detection is well above the localization precision.

	
Fit method - Average number of photons = 100	Weighted average method

Conclusion:

- The simulation is working properly and is faster than before.
- Using MTT/TrackMate, we can analyze the simulated movies and found back the proper values for the diffusion coefficients. However,
    1. For diffusion coefficient below **0.01µm²/s the fit method is much more accurate**. The weighted average method is too sensitive to the localization precision and is returning wrong estimation of the diffusion coefficient.
    2. For diffusion coefficient above 0.01µm²/s, both methods can be used. **However, the standard deviation of the distribution is narrower when using the weighted average, which can be important to keep in mind if several populations are co-existing within the sample.**