MODELED EXPERIMENT

Dataset#1, Dataset#2, Dataset#3, and Dataset#4 are four modeled datasets consisting of lists of durations from an experiment similar to the one you are conducting with your FCS setup. The modeled durations have been parsed into files containing equal numbers of durations, similar to the data files collected with the LabView software on your data acquisition computers.

Below is a description of the simulated experiment. All the data was modeled in Matlab using the program *model_durations.m*. The simulated data files and the Matlab program will be provided to you

Goals of the experiment

(a) Collect durations that result from the photon emission of fluorophores diffusing into the confocal volume.

(b) Calculate an autocorrelation function (ACF) from this data. The attributes of the fluorophore and FCS system are as follows:

How often the fluorophore emits a photon: 0.001 ms[†]

How often a background signal occurs: 1 ms[†]

> This time-scale comes from background data collected by one of groups from Chem 184, Spring 2007 (M. Etemadi, S. Achariya, A. Aycinena)

How long a fluorophore stays in the confocal volume: 0.1 ms, ±0.05 ms[*]

[†]These are modeled as Poisson processes.

[*]The length of time that a fluorophore spends in the confocal volume was randomly selected from a Gaussian distribution with mean 0.1 ms and standard deviation 0.05 ms. Thus, it is necessary to detect numerous events in order to determine the mean time.

The experimental approach

For each data set, I collected ~30 data files that each contained a fixed number of durations. Using the Matlab program *calc_acf*, the ACF is calculated for each individual data file, and an average ACF is calculated by averaging the ACF function from each file.

Below, I have written out my thought process while I tried to optimize the analysis of my modeled data.

NOTES on collecting the data:

(1) **Length of data file**: when the ACF is calculated by *calc_acf*, the function gets noisier as τ gets larger. This is because there is little data that is separated by large values of τ. The maximum value of τ for which the ACF can be calculated is equal to the file length; this is also the value of τ for which the ACF is noisiest.

Because I am trying to see a drop off in the ACF at ~0.1 ms. I want to determine the ACF accurately out to $\tau > 0.1$ ms, let's say ~0.5-1 ms. Which means that I want the total length of the file to extend even further. I collected files with total lengths that are ~1.5-3 ms.

(2) **Number of duration events to collect**: the number of durations you want to collect varies with your fluorophore concentration. To collect 1.5 ms of data will require more duration events for a more concentrated fluorophore sample.

As I "diluted" my fluorophore sample, I decreased the number of events per file that I collected. In this way, I maintained a similar total file length from data set to data set.

(3) **Mean value of collected durations**: this is determined by the concentration of the fluorophore sample as well as by background counts from your system. You cannot resolve the ACF to values of τ smaller than your smallest durations. Thus, the mean value of collected durations gives you an idea of the values of τ for which you can confidently calculate the ACF.

Because I am trying to see a drop off in the ACF at 0.1 ms, the mean of the collected durations must be significantly less than 0.1 ms. I aimed for the mean durations to be < 0.01 ms.

(4) **Number of files to collect**: more files will allow you to average over more data to arrive at your ACF, and this averaging quashes stochastic fluctuations in the calculated ACF. On the other hand – if you are collecting data over long periods of time, photobleaching can begin to affect your data, causing your first data file to look drastically different from your last file. Furthermore, *calc_acf* takes longer to calculate the ACF as you increase the number of files.

I chose to collect approximately thirty files at all data conditions.

IN SUMMARY – The **file length** sets the maximum τ for which you can calculate the ACF. The **mean duration** gives an indication of the smallest τ for which the ACF can be resolved. The **number of events** to collect varies for different fluorophore concentrations, and should be changed to maintain an appropriate **file length**.

WARNING – Of course, if you collect 100 data files with mean duration of 0.001 ms for 100 ms, you can calculate a great ACF. However, it will take the

computer an incredibly long time to deal with the ~10,000,000 data points. So – be smart about how much data you collect.

The data files

Below are descriptions of the four data sets I "collected." Dataset#1 is the most concentrated fluorophore sample, and each data set is progressively more dilute (as indicated by the variable: *"How often a fluorophore is detected"* in the descriptions below). Data sets, each containing ~30 files, will be provided to you in class.

Information about each set of files is posted below, along with comments about each. The data is in the same format as your data, so feel free to play around with *calc_acf* to calculate an ACF from each data set. Vary the length and binning of the calculated ACF, and see how this affects your plot.

I have also created a data set that is simply the background signal without any events. This data set is also described below.

Dataset#1 – 32 files
How often a fluorophore is detected: 0.01 ms
400 events are distributed throughout the files
Mean duration: 0.0011 ms
Mean file length: 1.33 ms / 1200 durations per file
    COMMENTS – It takes a *long* time to calculate an ACF from this data. The mean duration is smaller than is necessary to calculate a good ACF.

Dataset#2 – 27 files
How often a fluorophore is detected: 0.1 ms
200 events are distributed throughout all the files
Mean duration: 0.0021 ms
Mean file length: 1.45 ms / 700 durations per file

Dataset#3 – 32 files
How often a fluorophore is detected: 1 ms
100 events are distributed throughout all the files
Mean duration: 0.011 ms
Mean filelength: 3.32 ms / 300 durations per file
    COMMENTS – This data set meets the criteria described above.

Dataset#4 – 37 files
How often a fluorophore is detected: 10 ms
10 events are distributed throughout all the files
Mean duration: 0.050 ms
Mean file length: 1.49 ms / 30 durations per file
    COMMENTS – There should be 1 or fewer events for each data file. This is single molecule data, and will likely be too dilute to get a good ACF. Furthermore, the size of the mean durations is on the same order as the 0.1 ms drop off in ACF you are trying to observe. Thus, you may not have the appropriate resolution in τ.

Background– 100 files

Mean duration: 0.99 ms

Mean file length: 29.76 ms / 30 durations per file

> *COMMENTS* – More files were collected because there is less
> information about the ACF from a single background data file relative to
> a data file from the datasets above.

Analysis of data

Data was analyzed using the *calc_acf* program in Matlab.

In order to observe the drop in the ACF at 0.1 ms, the ACF was calculated out to 1 ms. Below are the Matlab commands used and the resulting ACF plots for 3-4 different binnings of the calculated ACF. Following analysis of a data set, I have included any comments on the plots. I analyzed the first 30 data files from all data sets except for Dataset#2, which only had 27 files.

Dataset#1

Analyzing this data with *calc_acf* took too long…and I didn't have the patience to wait for the program to finish. This is too much data.

Dataset#2

Command: *[t, a] = calc_acf(0.001, 10^(-8));*   (Bin size: 10 ns)
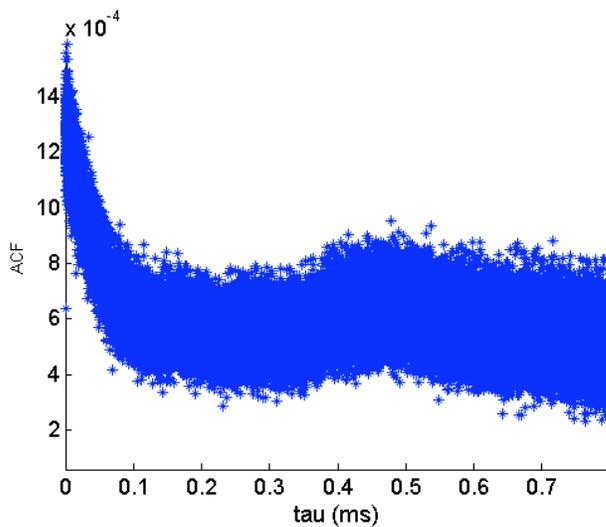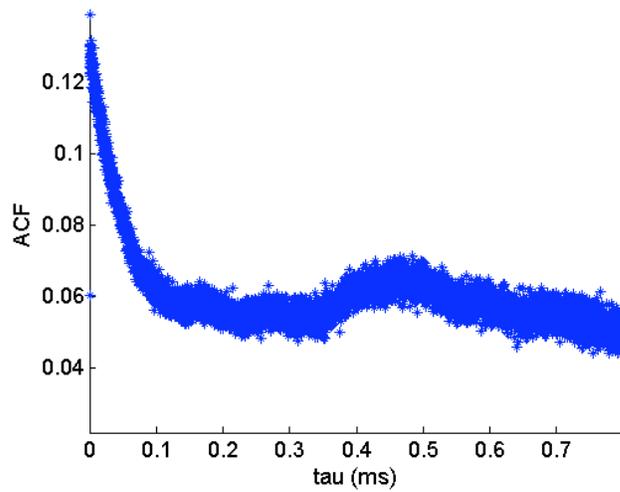
Command: *[t, a] = calc_acf(0.001, 10^(-7));* (Bin size: 100 ns)

Command: *[t, a] = calc_acf(0.001, 10^(-6));* (Bin size: 1 μs)
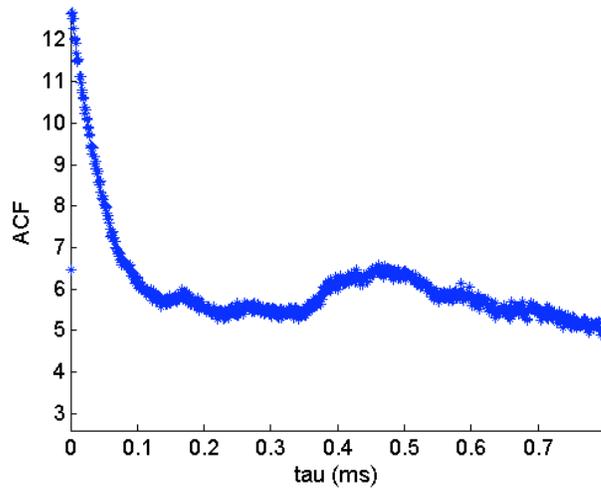
Command: *[t, a] = calc_acf(0.001, 10^(-5));* (Bin size: 10 μs)
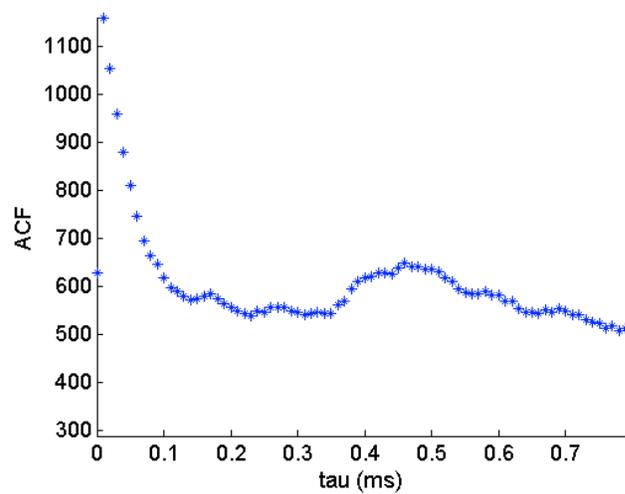
Comments on Dataset#2:

- Increasing the size of the bins reduced the fluctuations in the calculated ACF. However, a bin size of 10 μs resulted in very few values of the ACF throughout the drop at τ = 0.1 ms.

Dataset#3

Command: *[t, a] = calc_acf(0.001, 10^(-8));*   (Bin size: 10 ns)
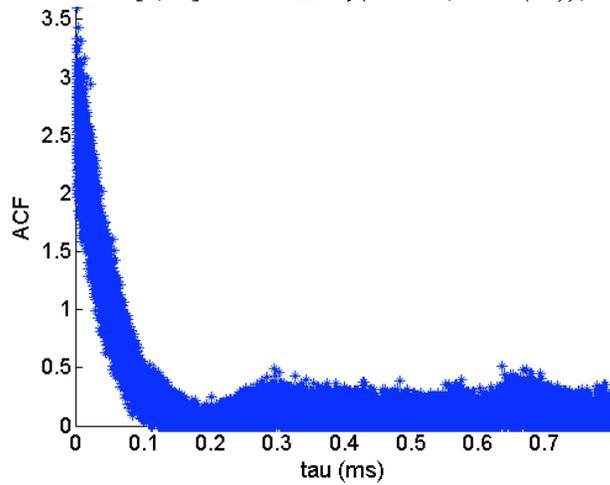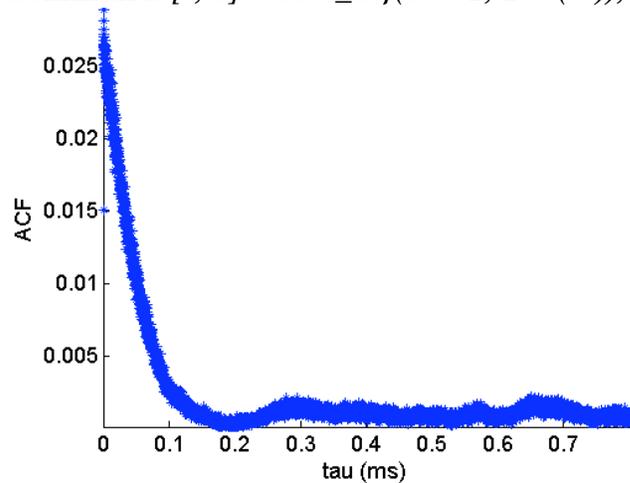
Command: *[t, a] = calc_acf(0.001, 10^(-7));*   (Bin size: 100 ns)

Command: *[t, a] = calc_acf(0.001, 10^(-6));*   (Bin size: 1 µs)
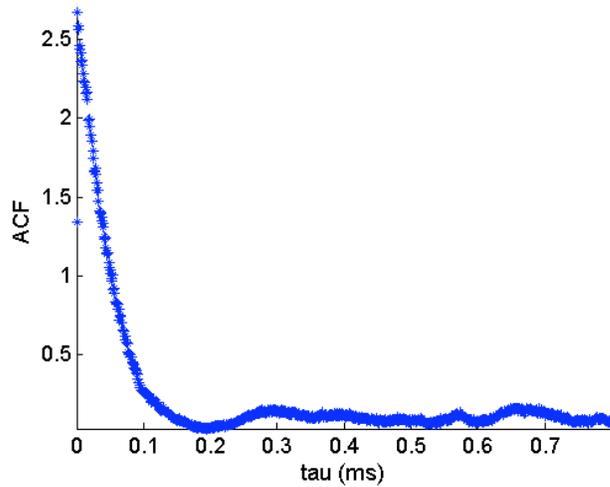
Command: *[t, a] = calc_acf(0.001, 10^(-5));*   (Bin size: 10 µs)
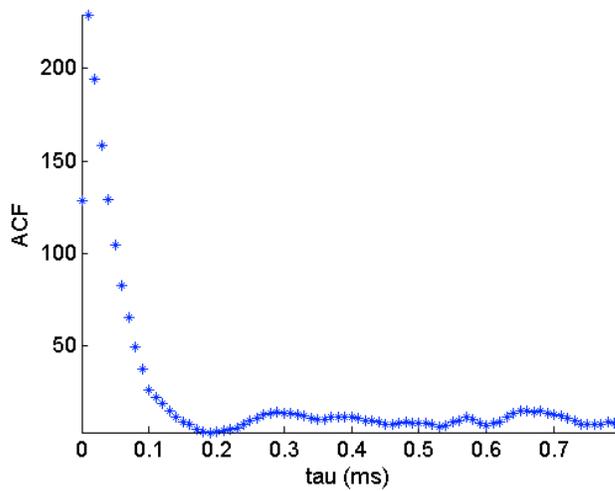
*Comments on Dataset#3*
- Compared to Dataset#2, the ACF is closer to zero following the drop at $\tau = 0.1$ ms.

Dataset#4

Command: *[t, a] = calc_acf(0.001, 10^(-7));*   (Bin size: 100 ns)
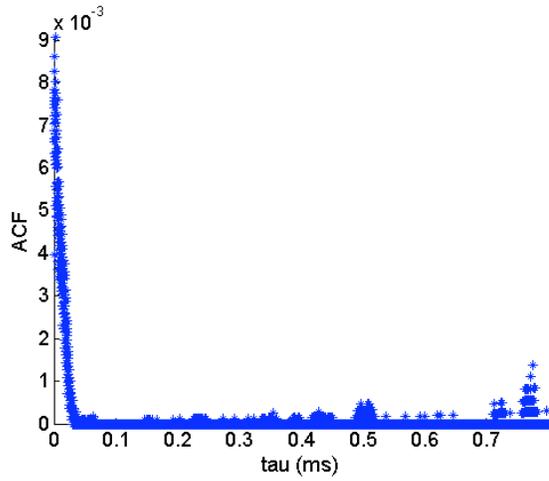

Command: *[t, a] = calc_acf(0.001, 10^(-6));*   (Bin size: 1 μs)
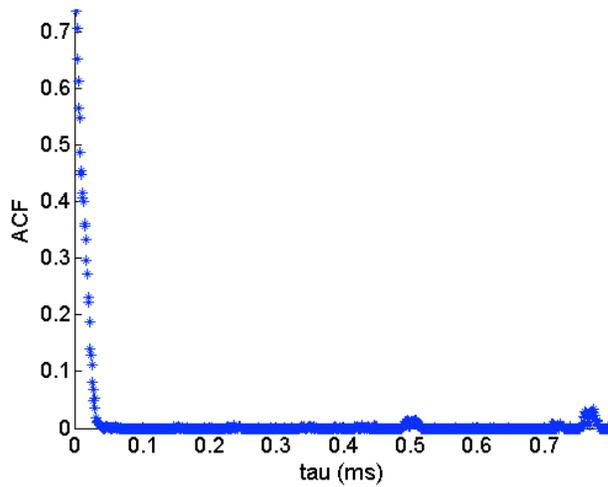
Command: *[t, a] = calc_acf(0.001, 10^(-5));*   (Bin size: 10 μs)
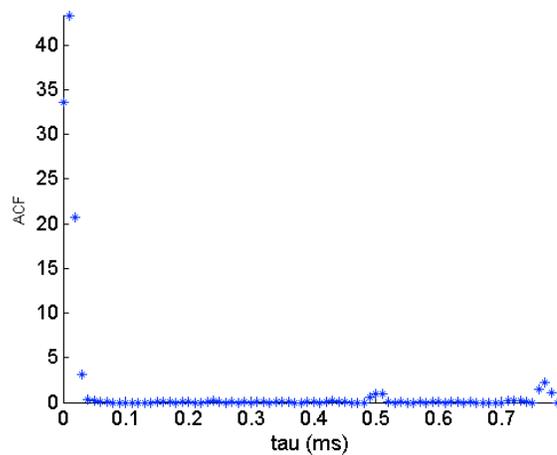
Comments on Dataset#4:

- It does not appear that the ACF is dropping at $\tau = 0.1$ ms. Instead, it seems to drop at 0.5 ms. This is likely because we have so few events in our data, and thus there is a large error on the mean time that a fluorophore spends in the confocal volume. We need more events than are in this data set.
- There is not enough data to make it worthwhile to bin the data with a 10 ns bin-size.

In general, Dataset#2 and Dataset#3 allowed me to effectively observe the drop in the calculated ACF. Binning sizes of 0.1-1 µs yielded plots of the ACF that were the most clear (i.e. the fluctuations were suppressed by the binning, but there were still numerous data points throughout the drop in the ACF).

Background

As a control – I analyzed the 100 background signal files:

Command: *[t, a] = calc_acf(0.001, 10^(-5));*   (Bin size: 10 µs)
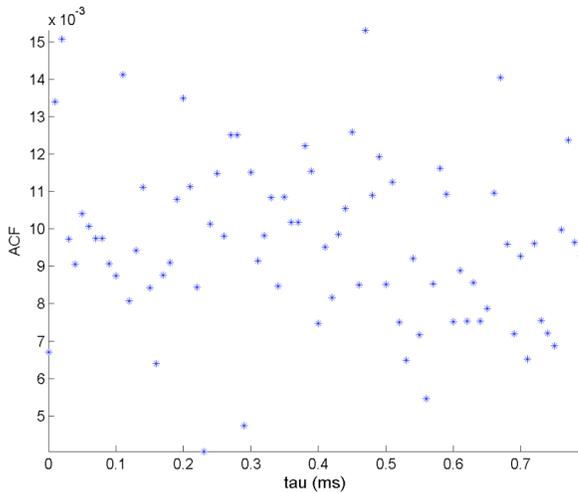
Comments:
- As expected, there is no indication of a drop at $\tau = 0.1$ ms.