UiT The Arctic University of Norway

Faculty of Science and Technology

Department of Physics and Technology

Conditional averaging of overlapping pulses

Rolf Annar Berg Nilsen

FYS-3900 Master's thesis in physics - 60 ECTS May 2023

Abstract

Conditional averaging is a signal processing method used to study turbulent fluctuations in a variety of fields. The method, in its simplest form, works by finding peaks in a signal that fulfill a certain size threshold. Equally sized excerpts of the signal around every peak are then cut out and averaged. This yields the average shape of the events that fulfill the condition. Based on the peak finding within the method one also obtains the amplitudes and waiting times between the conditional events. the aim of this thesis is to test if these statistics can be used to estimate underlying properties of the signals we are looking at. We use signals created by superposing pulses with the same shape at different times and with different amplitudes decided by input probability distributions, and compare the inputs used to make the signals with the outputs of conditional averaging. By changing the input distributions, we can alter the degree of pulse overlap within the signals, and thus see for which degrees of pulse overlap conditional averaging successfully reproduces the underlying statistics. We will also investigate the methods robustness in the face of noise, studying how well different conditions recover the underlying pulses, while also attempting to establish if additional conditions aimed at reducing noise effects are necessary.

Our results conclude that conditional averaging generally works well when predicting the shape of the underlying pulses, even in the face of pulse overlap. Noise severely affects the pulse shape estimates without any additional noise-reducing conditions. The noise reducing condition that we investigate here is one where we enforce a minimum distance between peaks. This additional condition results in greatly mitigating the effects of noise., leaving us to conclude that this it should always be used in addition to the size threshold condition to achieve the most robust results. The amplitude distribution estimates work well if there are not too many overlapping structures. However, the estimate break down even in the case of a moderate degree of pulse overlap, defined roughly as when there is on average one pulse arrival per pulse duration time in the signal. The estimates resemble the tail of the signals probability distribution largely independent of input amplitudes. Comparing the estimates from different input waiting time distributions, we conclude that the method fails to predict the underlying distribution, even in the case of little pulse overlap. We loosely define this as one pulse arrival per ten pulse durations on average within the signal. When pulses overlap more often, the method predicts exponentially distributed waiting times independent of input distribution. This leads to the main conclusion of the thesis; that the use of conditional averaging should in general be limited to estimating the average shape of underlying events.

Finally, the use of conditional averaging in previous works is also discussed. We find that the conclusions based on the average pulse shape estimate are valid, as authors use the method in regimes of pulse overlap where this estimate is still accurate. Contrary to this, conclusions made from amplitude and waiting time distributions are often made at degrees of pulse overlap where we have demonstrated the method to give erroneous results. This might lead authors to make conclusions based on incorrect information.

Acknowledgements

First and foremost I would like to thank my supervisor, associate professor Audun Theodorsen, for excellent guidance and great advice that always led to improvements within my work.

My thanks also go out to professors Odd Erik Garcia and Hans L. Pécseli for sharing their valuable insights into previous work done on conditional averaging.

And finally, I would also like to thank my parents for providing the most tasteful of support in the form of delicious and encouraging meals throughout the entire year. Without them my Sunday dinners would be significantly less toothsome.

Contents

1 Introduction

What appears to be randomly arriving coherent fluctuations on orders of magnitude comparable to the mean are observed in many nonlinear systems in many fields and on a variety of scales. They are found in turbulent winds [1,2], in the flow of rivers [3,4], in turbulent thermal convection of neutral fluids [5–7], in complex fluids [8], in neutron flux measurements within fission chambers [9], in measurements of plasma turbulence from magnetically confined plasmas [10–13], and many more [14–17]. One technique that has been used to analyze such fluctuations is conditional averaging [18–20] for its capabilities on extracting information such as the average fluctuation structure, amplitude distribution of the fluctuations and waiting time distribution of the waiting times between fluctuations. In its simplest form this routine picks out events every time a signal exceeds a given threshold, picks out a fixed size section of the signal around the maximum within that threshold crossing, and averages all these excerpts to find what the average structure that crosses the threshold looks like, in addition to recording the event amplitudes and time between each event. Earlier studies on the method have been focused on the effects of noise and has only been concerned with non-overlapping structures [4, 20, 21], and on 2D structures [18, 22]. The regime of non-overlapping structures is certainly not guaranteed in measurements of intermittent fluctuations, and little is known about the relation between the results of the conditional averaging and the underlying statistics of the signal when we cannot guarantee this. The goal of this thesis is thus to investigate and further the understanding on conditional averaging and its abilities to recover the underlying statistics of the signal.

One widely used way of modelling intermittent fluctuations is with a filtered Poisson process [4, 8, 9, 13, 23, 24]. The process can be described by a superposition of uncorrelated pulses where parameters such as the pulse shape, distribution of pulse amplitudes, duration times and waiting times between pulses can be randomly distributed. In addition, this process easily allows for different degrees of pulse overlap, making it the perfect test-bed to test if the results of conditional averaging can be used as accurate estimations for the underlying statistics of the FPP. Pulse overlap will be a central metric in this thesis and with it we hope to determine in what regimes conditional averaging give reliable results. Achieving this would help our understanding of experimental data to uncover if the results of the method represents something inherent about the fluctuations it analyzes or if it is an "unfortunate" property of the method and should be disregarded as such.

This thesis is structured as follows: In chapter 2 we describe the conditional averaging routine, how it works, what conditions we will use, why we use them, and what information we can record with the method. In chapter 3 we will describe the various distributions used to generate realizations of the FPP, some useful transformations of random variables used in the thesis and then we will discuss the FPP before defining a base case of the FPP used for testing and a little section on the limit of no pulse overlap and how it may affect some of the threshold statistics we are looking at. In chapter 4 we will go over the experimental results where we will first examine the effects of different pulse shapes in section 4.1. We will then look at the effects of two types of noise on the conditional averaged waveform in section 4.1.2. In section 4.2 we will look at the effects of different amplitude distributions to see if the amplitude distribution of the conditional average resembles the conditional amplitude distribution of the input amplitude, and investigating how different amplitude distributions affect the pulse shape. Then, in section 4.3, we will examine the effects of different waiting time distributions, much like we did for amplitudes, before discussing some combination cases where both the amplitude- and waiting time distributions are varied for confirmation or debunking of earlier results. After discussing all the results, we will discuss the some works where conditional averaging have been used to estimate properties about real-life data in chapter 5. Here, we will go into if their uses of conditional averaging are good ones or not based on the observed degree of pulse overlap within the signals. Finally, we will go over all the main results in a concluding manner in chapter 6, establishing the uses of conditional averaging before giving some outlooks on how this work could be furthered.

2 Conditional averaging

Conditional averaging is a technique that picks out events from a signal based on a given condition and averages them, outputting the "average event" and information such as where the events were found in the signal or how large the events were. The condition should not be too complex, as too complex conditions may lead to a self-fulfilling prediction. For example, say you want to learn if a signal you have is made up of lots of individual triangle pulses. One condition to pick out such pulses is to say that the condition requires a linear rise, followed by a linear fall over a certain timescale. The average event picked out by placing such a condition on the technique may then output the triangle shape you were expecting. But one cannot say if that average event resembles some intrinsic property of the signal or if it simply picked out the parts of the signal which looked triangular. This is why it is preferable that the condition we require should be a simple one that is not biased toward any specific shape. In this chapter we will introduce the conditional averaging method and the various conditions used.

Before illustrating we introduce some important notation. Φ is the signal. $\langle\Phi\rangle$ is the signal mean and Φ_{rms} is the standard deviation of the signal. $\tilde{\Phi}$ is the normalization we will use and it is defined as

$$\tilde{\Phi} = \frac{\Phi - \langle\Phi\rangle}{\Phi_{\text{rms}}}. \tag{2.1}$$

Thus the normalized signal $\tilde{\Phi}$ has a mean of 0 and a unit standard deviation.

To illustrate how a conditional average works we can look at figure 1. Here a simple amplitude threshold is the condition. At first we normalize our signal according to equation 2.1. The normalized signal is shown in the first figure. Then we pick an amplitude threshold, in this case one, and find where the signal crosses the threshold (red in the second figure). Within each crossing segment we identify the largest peak (lime). We then pick out equally big windows (red in the third figure) around each peak as shown in the last figure. Within these windows we take the signal values (orange) and average them so that we get the conditional average. In this thesis we also normalize each event by its amplitude before averaging, giving us the average shape of events that fulfill the condition. If one does not normalize each event before averaging the average itself becomes amplitude biased, meaning the shape of the larger events will be more expressed in the waveform result. If one has a good reason to expect the larger events to better represent underlying information about the signal this method might provide better results. However, if one cannot make such an assumption an amplitude unbiased method might be better suited.

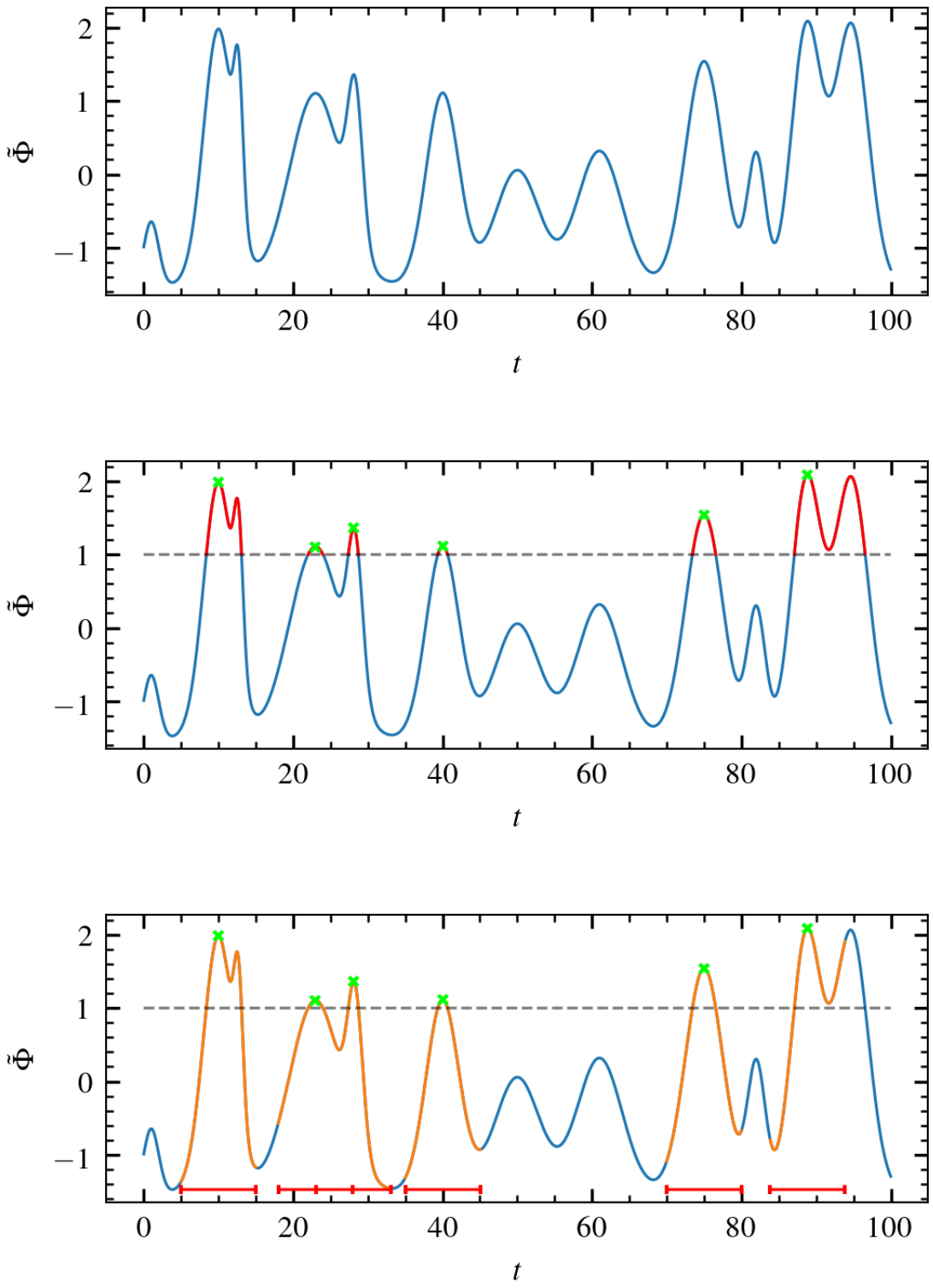


Figure 1: Step wise illustration of how conditional averaging picks out events with a simple amplitude threshold. The reference signal is constructed as a sum of individual Gaussian pulses arriving at different times with different amplitudes and different duration times. The arrivals, amplitudes and duration times of each pulse have been carefully chosen to illustrate many aspects of the peak finding in conditional averaging.

In this thesis we will study two main conditions, one being the simple amplitude threshold described earlier, and the other being a prominence threshold. The prominence threshold picks out large local peaks compared to the signal around it, allowing one to pick up large events that may occur beneath the more commonly used amplitude threshold. The reason for investigating this additional threshold condition is that we hypothesize that this method might be able to better pick out true underlying events in the signal. We also hope that the prominence threshold is inherently more resistant to the effects of noise, as a requirement of pulses being locally large hoppefully results in less noise fluctuations being registered as conditional events.

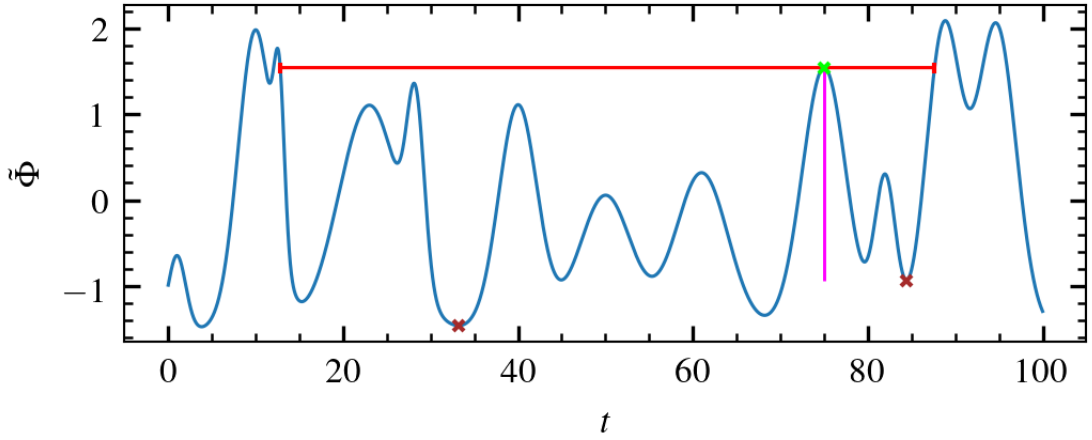


Figure 2: Example of how prominence is calculated.

In figure 2 we see an example of how prominence is calculated. The peak selected for the example is shown in lime. We start by extending a line (red) from the peak on either side until it hits the slope of a taller peak or the end of the signal, creating a window. Within this window we pick out the lowest point on both sides of the peak (brown). The taller one of these is chosen as a reference (brown right), and the prominence is the distance between this point and the peak (magenta). To better illustrate the differences between the threshold conditions and how they affect the peak finding, we first need to explain how the prominence threshold works in the conditional average method.

To explain how the prominence condition works in conditional averaging we use an illustration. Looking at figure 3 we start with the same normalized signal as before. All the peaks (lime) in the signal are found, and their prominence values (magenta and yellow) are calculated according to the previous description. We then place a condition on the minimum required prominence, in this case one, and use only the peaks with a prominence value above this threshold (magenta) to calculate the conditional average with the same use of taking equal windows around each peak (red) and using the signal values in those windows (orange) for the average. When normalizing with this method the prominence value of each conditional event is used to normalize it and the events are shifted so that all the peak points coincide before taking the average.

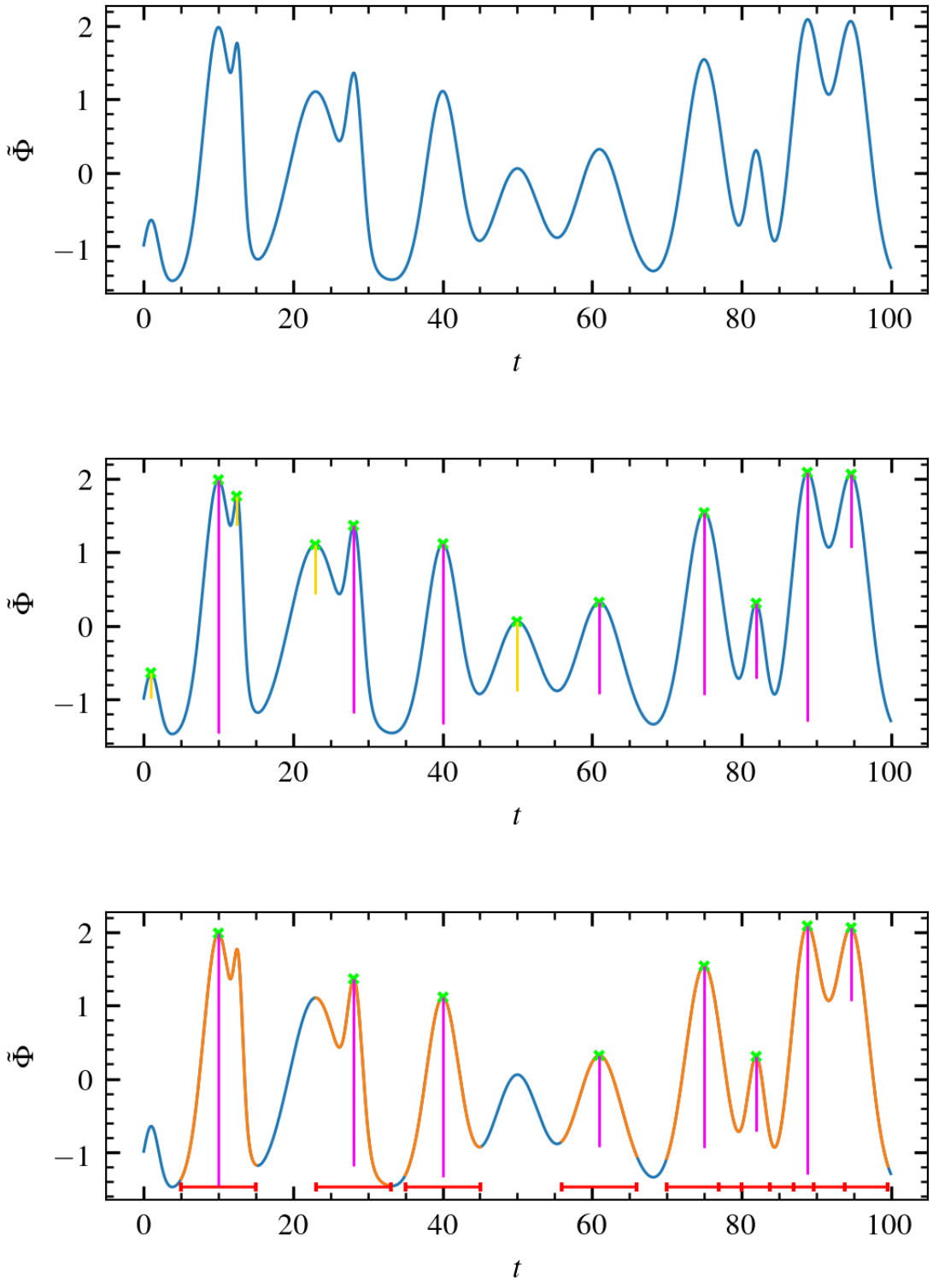


Figure 3: Step wise illustration of how conditional averaging picks out events with a prominence threshold.

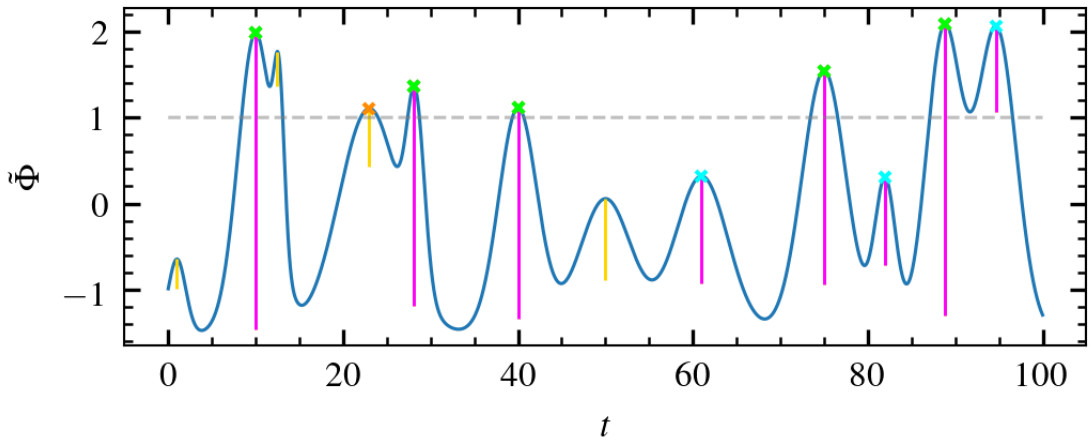


Figure 4: Example of how different peaks are chosen based on different size thresholds. The dashed grey line is the amplitude threshold. The yellow lines are prominences below the prominence threshold and the magenta lines are prominences above the prominence threshold. Peaks that are picked up only when using the amplitude threshold are shown in orange, only by the prominence threshold in cyan and by both threshold conditions in lime.

Figure 4 shows the difference in the type of peaks that are picked up with the different threshold conditions. The common peaks (lime) are the large ones that cross the amplitude threshold (grey) and that has a prominence that is larger than the prominence threshold (magenta). The peaks that only the amplitude threshold picks up (orange) are the peaks that cross the amplitude threshold, but have prominence values smaller than the prominence threshold (yellow). The peaks that are only picked up by the prominence threshold (cyan) are peaks with a large prominence, but do not cross the amplitude threshold such as the third and fifth peaks from the right. Peaks that have high enough prominence values, but are also within one amplitude threshold crossing of another peak, are also only picked up with a prominence threshold. The peaks that neither method pick up are the peaks with prominence values less than the threshold and that do not cross the amplitude threshold, like the two peaks with yellow prominences below the grey line. Another type of peak that neither pick up are low prominence peaks that are close to a taller peak, but within one amplitude threshold crossing, such as the third peak from the left. What we see from this comparison is that there can exist local large amplitude events that are not picked up by the standard amplitude threshold, but that we would find with a prominence threshold. The prominence condition also enforces an inherent distance between peaks, as a peak must decay for a certain time before its prominence value can become big enough to be registered as a conditional event. We hope that this property creates statistical independence between events, making additional conditions that are created for just this purpose unnecessary.

An additional condition that is one that states that two separate averaging windows (shown in red in figure 1 and 3) are not allowed to overlap. What this condition does is that is makes no single part of the signal appear multiple times in the average. The thought behind this additional condition is that it introduces some forced independence between events, removing any unwanted correlations that may arise from including multiple segments of the signal multiple times in the average. This additional condition will be referred to as `overlap=True` throughout the rest of this thesis.

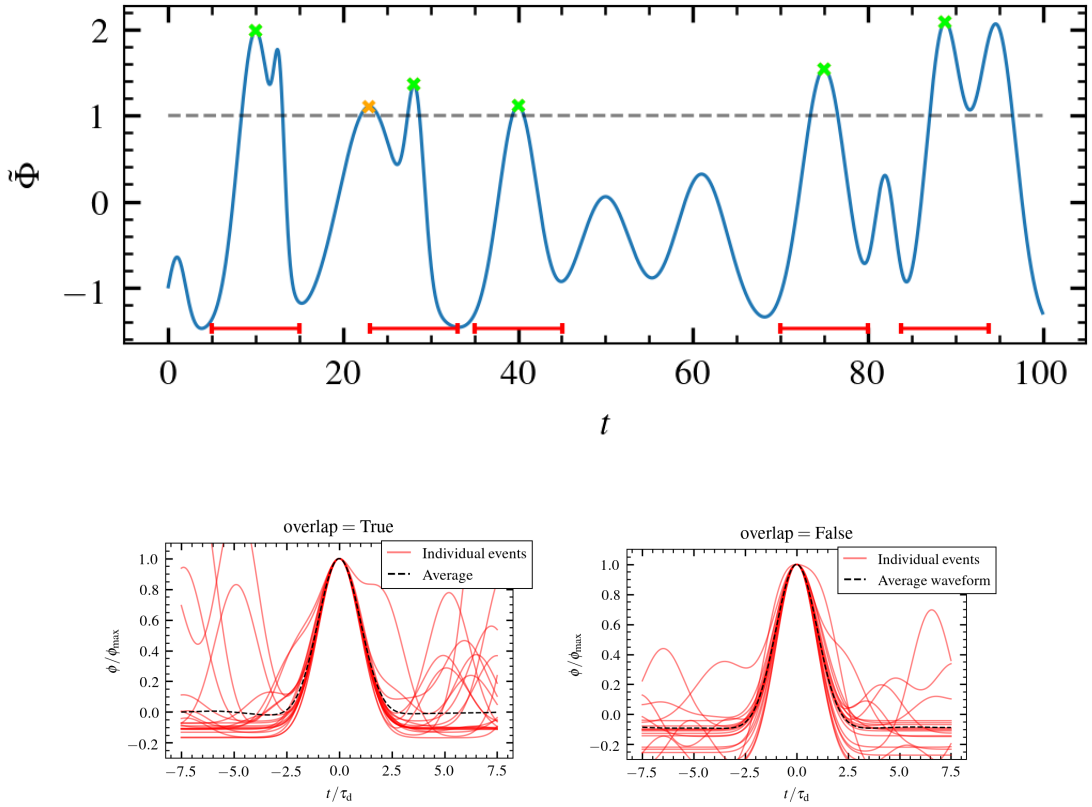


Figure 5: Example of the difference and without the `overlap=False` condition in the peak finding (top). And an example of a conditional averaged waveform both with (bottom right) and without (bottom left) the `overlap=False` condition. The black dashed line showing the average and the red lines showing individual contributions in the average. The signal that the conditional averaging was used on consisted of a superposition of Gaussian pulses with degenerately distributed duration times with a mean of one, exponential waiting times with a mean of 10 and exponentially distributed amplitudes with a mean of 1.

Looking at figure 5 we can see what the `overlap=False` condition does. In the top figure we see the peaks piked out by the amplitude threshold condition together with the `overlap=False` condition (lime) and the peaks picked out without the `overlap=False` condition (orange). The reason for that the orange peak is not included when we use the `overlap=False` condition is that it falls within the window (red) of the larger lime peak to the right of it. In the two bottom figures we also see the conditionally averaged waveform both with and without the `overlap=False` condition. If we do not use the additional condition, we see that in the individual events the occurrence of larger neighbouring events is more frequent than in the case where we use the `overlap=False` condition. We see in the averaged waveform that this condition lowers the observed mean (the value to which the pulses decay to). Without the `overlap=False` condition the average decays to the true mean of 0, but with the `overlap=False` condition we impose a bias that reduces the amount of large neighbouring events in the individual conditional events that are picked up. This naturally occurs as we intentionally choose the largest peak within an averaging window. Keep in mind that the `overlap=False` condition does not eliminate the occurrence of larger neighbouring events in the individual conditional events.

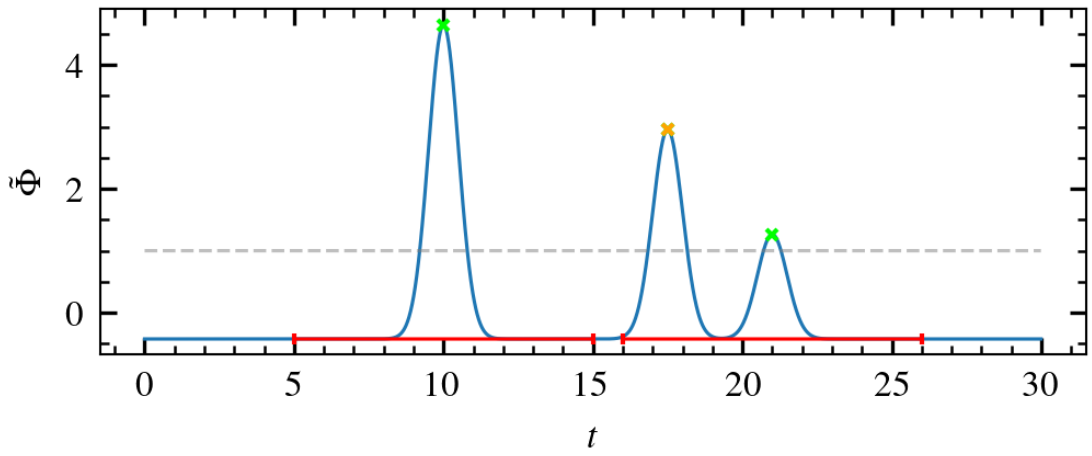


Figure 6: Example of how a neighbouring peak may be larger than the conditional event within an individual sample when using the `overlap=False` condition.

In figure 6 we see an example of how neighbouring peaks may still be larger than the conditional event within an individual sample when using the `overlap=False` condition. The orange peak is larger than the lime peaks which become conditional events. The reason it is not registered is because its averaging window would overlap with the leftmost tallest peak. It still appears in the average within the rightmost peaks individual sample, thus as a taller peak than the conditional event itself.

The reason we are looking at both the amplitude threshold and the prominence threshold is that we want to compare them, both with and without the `overlap=False` condition. Does the prominence threshold inherently provide statistical independence between events without the `overlap=False` condition? Is either method better than the other at predicting underlying amplitude and waiting time distributions? Is the `overlap=False` condition necessary for either method? These are all interesting questions that we will attempt to answer within this thesis but the main goal remains to provide further insight into how the results of conditional averaging fair when it comes to predicting underlying information about test signals.

3 Stochastic modelling

3.1 The filtered Poisson process

To test the results of conditional averaging we need to be able to control the parameters that conditional averaging is said to estimate, letting us compare estimates against the input information. One stochastic process that fits this bill perfectly is the filtered Poisson process (FPP). In this chapter we will introduce this stochastic process and some important aspects of it.

The filtered Poisson process can be described as a sum of uncorrelated pulses on a given time interval. If the number of pulses are K and the interval has a length of T then the process Φ can be described by [23]

$$\Phi_K(t) = \sum_{k=1}^{K(T)} A_k \phi\left(\frac{t - t_k}{\tau_k}\right) \tag{3.1}$$

where A_k are the amplitudes of each pulse, t_k is the arrival time of each pulse, τ_k is the duration time of each pulse and ϕ is the pulse function. As this stochastic process is easily manipulated it is the ideal test bed for conditional averaging. The outputs of conditional averaging are an average pulse shape, the amplitudes of all the events found and the waiting time between those events. As we see from 3.1 we can draw amplitudes from any given distribution to use as the amplitude distribution. We can specify the pulse function and then compare the input together with the conditionally averaged waveform. And we can construct the arrival times from different waiting time distributions by cumulatively summing them up, then comparing the waiting time distribution with the one from the conditional averaging.

Going further we need to define some useful parameters of the model. $\langle A \rangle$ is the average pulse amplitude, $\tau_\mathrm{w} = \langle T_\mathrm{W} \rangle = T/\langle K \rangle$ is the average waiting time between pulses and $\tau_\mathrm{d} = \langle \tau_k \rangle$ is the average duration time of the pulses. The ratio of the average pulse duration and the average waiting time is defined as the intermittency parameter [23]

$$\gamma = \frac{\tau_\mathrm{d}}{\tau_\mathrm{w}} \tag{3.2}$$

This parameter determines how intermittent a signal is. For low values of gamma there is a large waiting time between pulses compared to the duration time and there will be little overlap between pulses, while for large values of γ there will on average be a large amount of pulses per duration time, and thus much pulse overlap. Thus, γ can also be viewed as the parameter that determines the degree of pulse overlap. Varying this parameter will be the central focus of this thesis. Earlier work on conditional averaging has been done where the pulses were not allowed to overlap [4,20,21] and will be comparable to the work done here for small values of γ. The main goal of this thesis will be to determine the regimes of γ where conditional averaging give useful results that relate to the input distributions and pulse functions.

With these important parameters in mind we can define some of the moments of the process [23].

The expected value

$$\langle \Phi \rangle = \frac{\tau_\mathrm{d}}{\tau_\mathrm{w}} I_1 \langle A \rangle = \gamma I_1 \langle A \rangle \tag{3.3}$$

Where the integer moments of the pulse function, I_n, are defined as

$$I_n = \int_{-\infty}^{\infty} \mathrm{d}x [\phi(x)]^n \tag{3.4}$$

The second moment and the rms value

$$\langle \Phi^2 \rangle = \langle \Phi \rangle^2 + \gamma I_2 \langle A^2 \rangle \tag{3.5}$$

$$\Phi_\mathrm{rms}^2 = \gamma I_2 \langle A^2 \rangle \tag{3.6}$$

The skewness and flatness can be obtained from the cumulants an are [25]

$$S_\Phi = \frac{1}{\gamma^{1/2}} \frac{I_3}{I_2^{3/2}} \frac{\langle A^3 \rangle}{\langle A^2 \rangle^{3/2}} \tag{3.7}$$

$$F_\Phi = 3 + \frac{1}{\gamma} \frac{I_4}{I_2^2} \frac{\langle A^4 \rangle}{\langle A^2 \rangle^2} \tag{3.8}$$

We see that for small γ both the mean and the rms value tend to be small but the relative fluctuation level given by

$$\frac{\Phi_{\text{rms}}^2}{\langle\Phi\rangle^2} = \frac{1}{\gamma}\frac{I_2}{I_1^2}\frac{\langle A^2\rangle}{\langle A\rangle^2} \tag{3.9}$$

becomes large for small γ indicating that the process varies a lot in relation to its mean for very intermittent realizations. We also see that when γ becomes large the relative fluctuation level, the skewness and the excess kurtosis ($F_\Phi - 3$) become small while the mean and the rms-values get large. This indicates that the distribution becomes more symmetric and the flatness approaches that of a normal distribution as γ increases.

3.1.1 The base case

As there are numerous ways the different input distributions and pulse functions the FPP can take, and as we can only look at a limited number of cases, we choose one base case of the FPP which has some nice analytical expressions for a number of statistics. The base case is defined as

$$A \sim \text{Exp}(1) \tag{3.10}$$
$$\tau \sim \text{Degenerate}(1) \tag{3.11}$$
$$T_{\text{w}} \sim \text{Exp}\left(\frac{1}{\gamma}\right) \tag{3.12}$$
$$\phi(x) = \begin{cases} \exp(-x), & x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.13}$$

where we have exponentially distributed amplitudes and waiting times, meaning we have uniformly distributed arrival times and the number of pulses given a certain time interval is thus distributed according to a Poisson distribution [26]. The duration time distribution is degenerate with a mean of one and the pulse function is chosen to be a one-sided exponential pulse.

In this case the mean and rms values are

$$\langle\Phi\rangle = \gamma \tag{3.14}$$
$$\Phi_{\text{rms}} = \sqrt{\gamma} \tag{3.15}$$

as $I_n = \frac{1}{n}$ and $\langle A^n\rangle = n!$ for this case. The probability density function for the base case is a gamma distribution

$$p_\Phi(\Phi) \sim \Gamma(\gamma, 1) \tag{3.16}$$

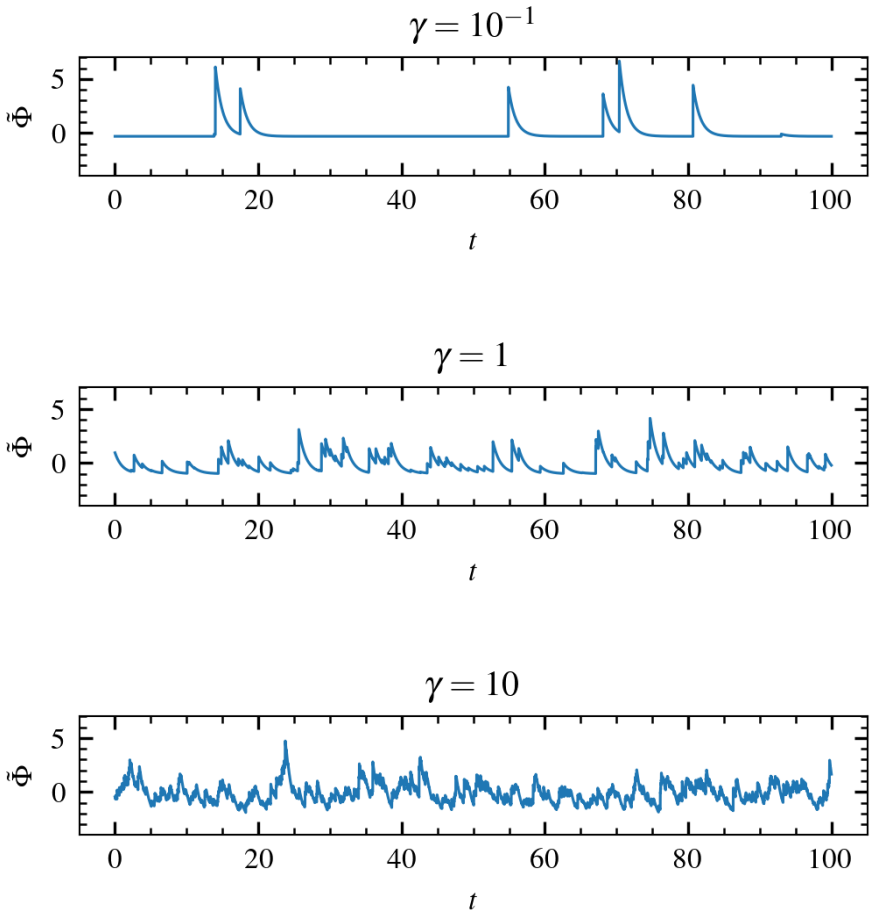


Figure 7: Example of realizations of the base case with different degrees of intermittency.

In figure 7 we see examples of the normalized base case at different levels of intermittency. In the $\gamma = 10^{-1}$ signal we easily see the distinct pulses that are superposed to make up the signal. This will serve as our high intermittency case for which we expect conditional averaging to provide the most reliable results. In the $\gamma = 1$ signal we see that the overlap of individual pulses is more prevalent, but one can still distinguish individual pulses from one another. Looking at the $\gamma = 10$ signal we see the effects of pulse overlap, here it is much harder to tell what the underlying pulse shape is, or if there even is one. Intuitively based on this decline in being able to tell underlying pulse information, one can hypothesize that the results from conditional averaging would decline in quality with increasing pulse overlap in the signals.

3.1.2 Noise

Noise is an unavoidable element we encounter if we do any form of analysis on real life data, which is why it is also one of the aspects of conditional averaging that has been studied previously. For example [20] focused on creating a more complicated condition and averaging scheme to reduce the effects of noise. However, the underlying structures were not allowed to overlap. As real life data may include the effects of overlapping underlying structures and noise at the same time it is important to investigate how conditional averaging performs under the variation of both these parameters.

In this thesis we will concern ourselves with two different forms of noise. One is simply additive or observational noise, analogous to measurement noise. This is denoted by [27]

$$\Omega(t) = \Phi(t) + \sigma N(t) \tag{3.17}$$

where σ denotes the noise intensity and $N(t)$ is a normally distributed random process with a man

of zero and a unit standard deviation. The other noise type is dynamical noise and is denoted by [27]

$$\Psi(t) = \Phi(t) + \sigma Y(t) \tag{3.18}$$

where σ again denotes the noise intensity parameter and $Y(t)$ is an Ornstein-Uhlenbeck process with zero mean and unit standard deviation. Both noise types are independent of $\Phi(t)$ but the difference is that $N(t)$ varies around the mean on a timescale defined by the sampling time, whereas $Y(t)$ varies on the same timescales as the process itself as they have the same auto-correlation function [27]. With these types of noise in mind we can define the relative noise level ε

$$\varepsilon = \left(\frac{X_{\text{rms}}}{\Phi_{\text{rms}}}\right)^2 = \frac{\sigma^2}{\gamma\langle A\rangle^2} \tag{3.19}$$

where X represents either of the noise processes σN or σY. It is this parameter we will use to control the amount of noise as we generate realizations of these processes.

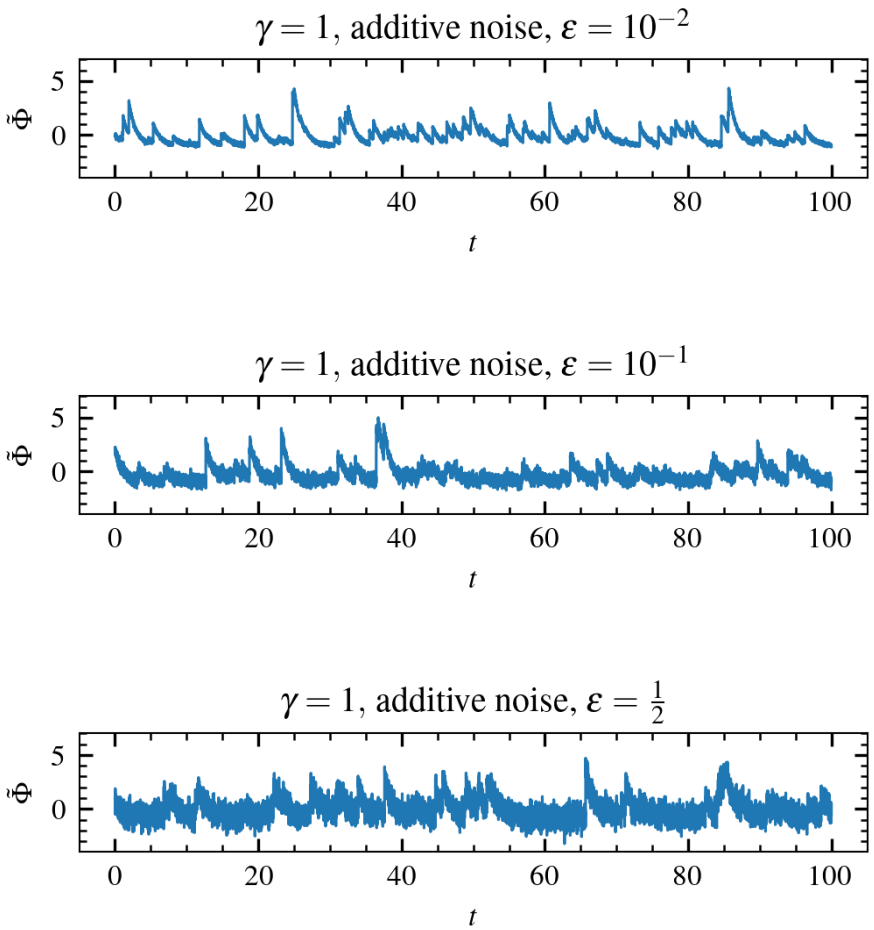


Figure 8: Example of realizations of the base case with $\gamma = 1$ and additive noise for different degrees of the relative noise level.

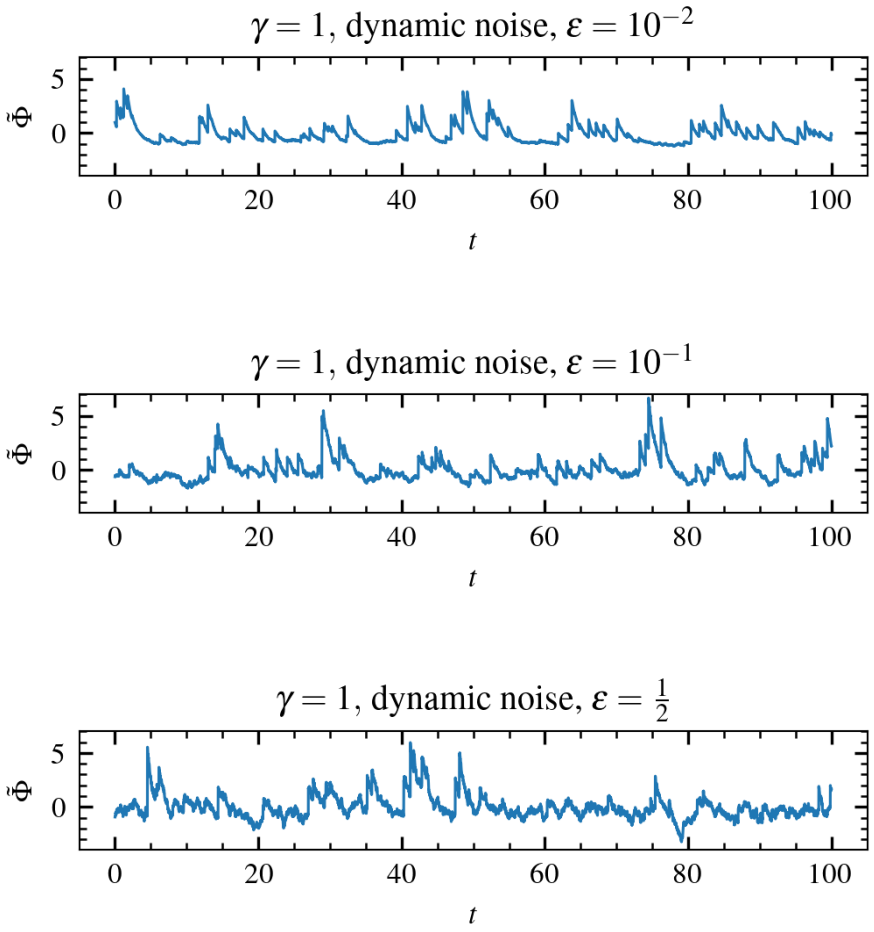


Figure 9: Example of realizations of the base case with $\gamma = 1$ and dynamic noise for different degrees of relative noise level.

In figures 8 and 9 we see examples of the base case with the different types of noise at different levels of ε with $\gamma = 1$. We observe that additive noise is the noise type that affects our visual ability to distinguish individual pulses in the signal the most. While dynamic noise does affect it as well we are still able to pick out more pulses at higher degrees of ε. With this in mind, we might expect the results of conditional averaging work better with dynamic noise. Conditional averaging picks out and averages the large amplitude events within the signals, with the hope being that this procedure averages out the fluctuations created by noise, giving us a waveform that resembles the underlying pulse shape.

3.1.3 The high intermittency case

In the case where $\gamma << 1$ we have a highly intermittent signal. In this situation we assume no pulse overlap as the probability of overlapping events become vanishing as γ becomes small if we have uniformly distributed arrivals. In this ideal scenario all threshold crossings are determined solely by the amplitude distribution. The amplitude distribution from the conditional average should then match the theoretical conditional amplitude distribution given by B.5. For exponential amplitudes this expression becomes

$$p_{A_c}(A_c) = \frac{1}{\langle A \rangle} \exp\left(-\frac{A_c - c}{\langle A \rangle}\right) \tag{3.20}$$

for the pure threshold method with the base case the threshold is given by

$$c = 2.5\Phi_{\text{rms}} + \langle \Phi \rangle = 2.5\sqrt{\gamma} + \gamma \tag{3.21}$$

and the conditional amplitude distribution becomes

$$p_{A_c}(A_c) = \exp(-A_c + 2.5\sqrt{\gamma} + \gamma) \tag{3.22}$$

on the interval $[2.5\sqrt{\gamma}+\gamma, \infty]$. The waiting time distribution of events above the threshold is then also determined by the amplitude distribution. For uniformly distributed arrivals the events above threshold are also uniformly distributed, but the mean time between events will be larger as there will be fewer events above the threshold than the total events. The ratio determining the new mean waiting time will then be given by the amplitude distribution and the threshold. The new mean is thus given by

$$\langle T_{\mathrm{w}c} \rangle = \langle T_{\mathrm{w}} \rangle (1 - P_A(c)) = \langle W \rangle \exp(-(2.5\sqrt{\gamma} + \gamma)). \tag{3.23}$$

In the base case where $\tau_{\mathrm{d}} = 1$ this can also be expressed as

$$\langle T_{\mathrm{w}c} \rangle = \frac{1}{\gamma} \exp(-(2.5\sqrt{\gamma} + \gamma)). \tag{3.24}$$

And the waiting times between events above the threshold is distributed exponentially according to

$$T_{\mathrm{w}\,c} \sim \mathrm{Exp}\left(\frac{1}{\gamma} \exp(-(2.5\sqrt{\gamma} + \gamma))\right). \tag{3.25}$$

4 Results and discussion

In this chapter we will examine the different estimates that conditional averaging gives us. To do this we will construct different realizations of the filtered Poisson process as described in section 3.1 and use conditional averaging to attempt to recover what we used to generate the realizations. The inputs we will attempt to recover is the pulse function ϕ through the conditionally averaged waveform, the duration time τ_d from the waveform estimate, the conditional amplitude distribution P_A through the amplitude and prominence values of the conditional events, and the waiting time distribution P_W through the times between conditional events. We will examine two different threshold conditions, the amplitude threshold and the prominence threshold to see if one better recovers the inputs than the other. The role of the `overlap=True` condition will also be investigated to see if its property of providing more independent events is a useful one, or if it is a superfluous condition.

All the analysis in this section is done with the intermittency parameter γ in mind. It determines the degree of intermittency of the signals and therefore the degree of overlap between individual pulses. We can therefore examine the results of conditional averaging as the degree of pulse overlap within the signal increases. When we have highly intermittent signals, equivalent to a small γ-value, the individual events that make up the signals are more likely to "stand alone" and be clearly distinguishable from one another if one simply visually inspects the signal. In this case we expect conditional averaging to accurately recover the input parameters of the signals as the average will be dominated by actual underlying events crossing the chosen threshold. As γ increases the signals become less intermittent with more pulse overlap and threshold crossings become more and more due to individual events overlapping. In this transitional intermittency, where some crossings are caused by individual events, and some by overlap, we wish to test conditional averaging to see how overlap affects the estimates. As γ increases even more the probability of threshold crossings due to individual events vanishes and the conditional average will be wholly dominated by overlap induced crossings. We also wish to examine the results of conditional averaging in this regime. Because of this we chose to mainly investigate three different orders of the intermittency parameter. $\gamma = 10^{-1}$ as the highly intermittent case, $\gamma = 1$ as the transitory stage of intermittency and $\gamma = 10$ as the overlap dominated case. This can also be motivated from experimental data measuring plasma turbulence in the scrape-off layer of toroidal fusion reactors where one has used the filtered Poisson process to describe such fluctuations. Here γ is often found to be somewhere in between 1 and 10 [11, 27, 28], making any results from this investigation highly relevant for this field as both the FPP-framework and conditional averaging is used here.

The section is structured as follows. At first, we will look into how well conditional averaging recovers the underlying shape and duration time of the pulses by varying the true pulse shape, ϕ in section 4.1.1. We will also investigate the effects of overlap on estimated pulse symmetry. Afterwards we explore the effects of noise on the average waveform and find the main effect of the `ovelap=False` condition in section 4.1.2. Furthermore, in section 4.1.3, we will examine signals where the amplitudes and waveforms of individual events are correlated to see if conditional averaging can pick up such a correlation by varying the size of the threshold. In section 4.2 we examine different input amplitude distributions to reveal if the conditional average can recover them in their conditional forms. We will also study how different amplitude distributions affect the waveform. Finally we will investigate different waiting time distributions and their effects on the estimates obtained from conditional averaging in section 4.3.

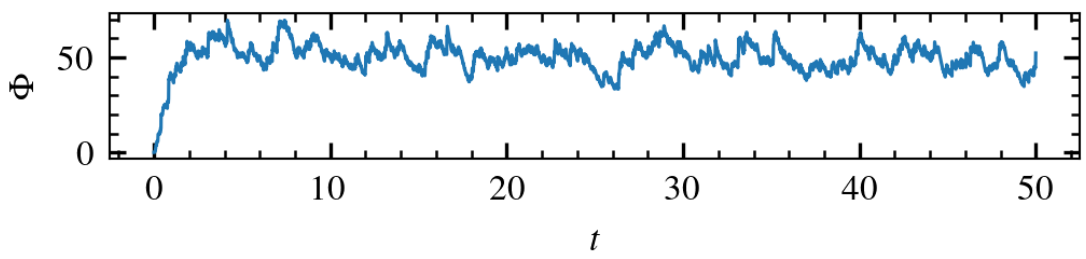


Figure 10: Example of a realizations of the base case with $\gamma = 50$ at the build-up phase of the signal.

The simulation parameters common for all signals are as follows. The total duration of the signals is $T = 10^5$, the signal is generated in sample increments of $\Delta = 10^{-2}$, making each signal have 10^7 data points. Each signal was initially generated with $T' = 1.002 \cdot 10^5$ where 10^2 points was stripped of each end to eliminate possible end effects such as removing the non-stationary section where the signal builds up as observed in figure 10. This results in $\gamma = 10^{-1}$ signals consisting of on average 10^4 individual pulses, $\gamma = 1$ signals consisting of 10^5 individual pulses and $\gamma = 10$ signals on average consisting of 10^6 individual pulses when the duration time of all the pulses is $\tau_\mathrm{d} = 1$. This ensures that the signals are long enough that we should not encounter any large problems related to a lack of data.

4.1 Waveform analysis

The main use of conditional averaging has been its use to obtain an estimate for the shape of underlying events within a given signal. It is therefore obvious that we should study this ability, as discerning its robustness is important to be able to distinguish good results from bad ones. To do this we will examine the averaged waveform results from conditional averaging for the amplitude and prominence threshold conditions both with and without the `overlap=False` condition for different degrees of intermittency. We will explore the methods ability to recover different pulse functions, both symmetric and asymmetric. In addition to exploring the waveform estimate in the presence of noise within the signals. We hope to find out conclusively where the waveform estimates break down and how intermittency affects different underlying pulse shapes in the conditional average.

4.1.1 Different pulse functions

Signals as constructed from a filtered Poisson process have different statistical properties based on the underlying pulse function [23]. Therefore predicting this waveform is an important task that can help us understand more about the signals we study. Conditional averaging estimates this pulse shape, making it important to test this property to see if it can be trusted at increasing degrees of pulse overlap. To test this we will construct realizations of the FPP with different input pulse functions, get the conditionally averaged waveform from taking a conditional average, then finding the best fit of the input pulse function to the waveform to be able to compare the input parameters of the pulse to the fitted parameters from the waveform

To generate the signals used in the pulse function analysis the base case 3.1.1 parameters were used but with different pulse shapes. Meaning the true amplitude and waiting time distributions are both exponential together with the time axis being normalized by the true duration time.

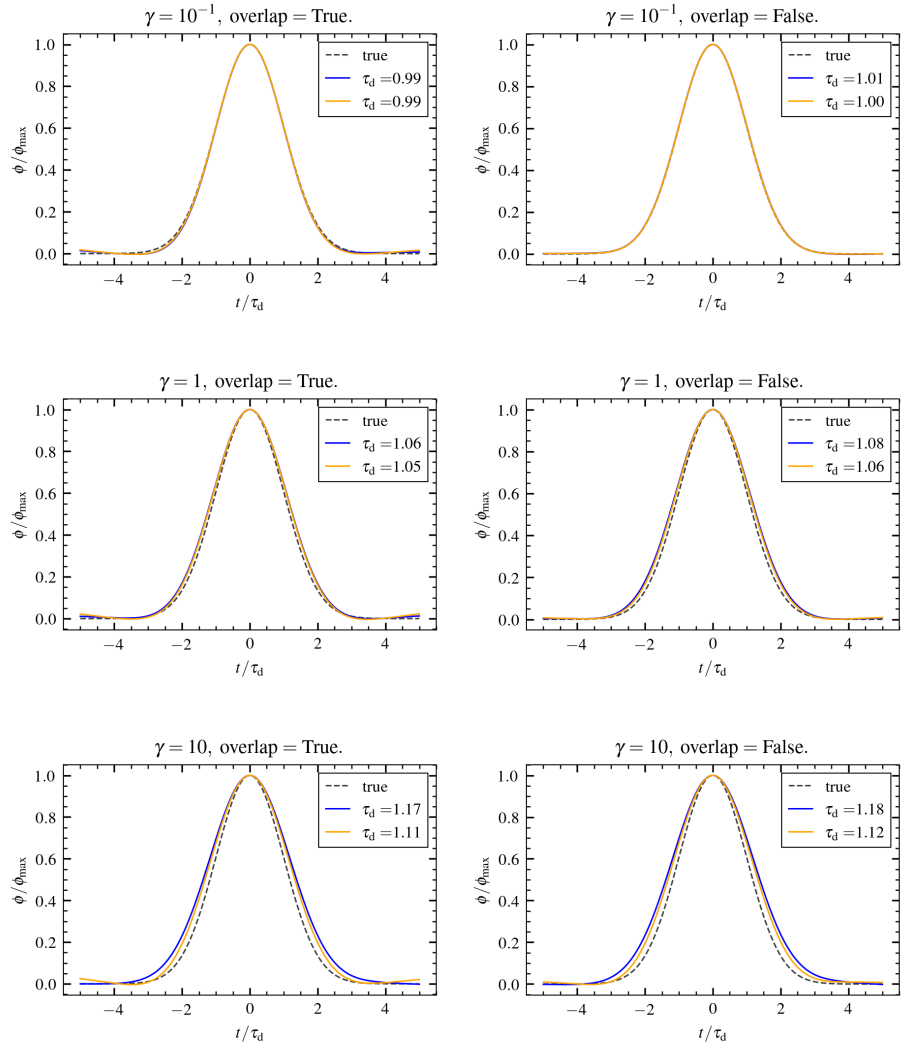


Figure 11: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A Gaussian pulse function was used to generate the signal.

Looking at figure 11 we see that the conditionally averaged waveform overlaps nicely with the true pulse function for $\gamma = 10^{-1}$ with there being no significant difference between the estimates of either condition. The `overlap=False` condition also seem to have little effect. This is again the case for $\gamma = 1$ where the estimates differ little from each other, however the estimated duration time is consistently larger than the true value for all cases. This becomes even more apparent in the $\gamma = 10$ case where all estimated duration times are significantly larger than the true one. For this case it also seems like the prominence condition yields more accurate estimates independent of allowing overlap. From a general standpoint one could say all these waveforms look Gaussian, so if anything the method accurately picks out the general shape even though it tends to overestimate the duration time of the pulses when $\gamma > 10^{-1}$.

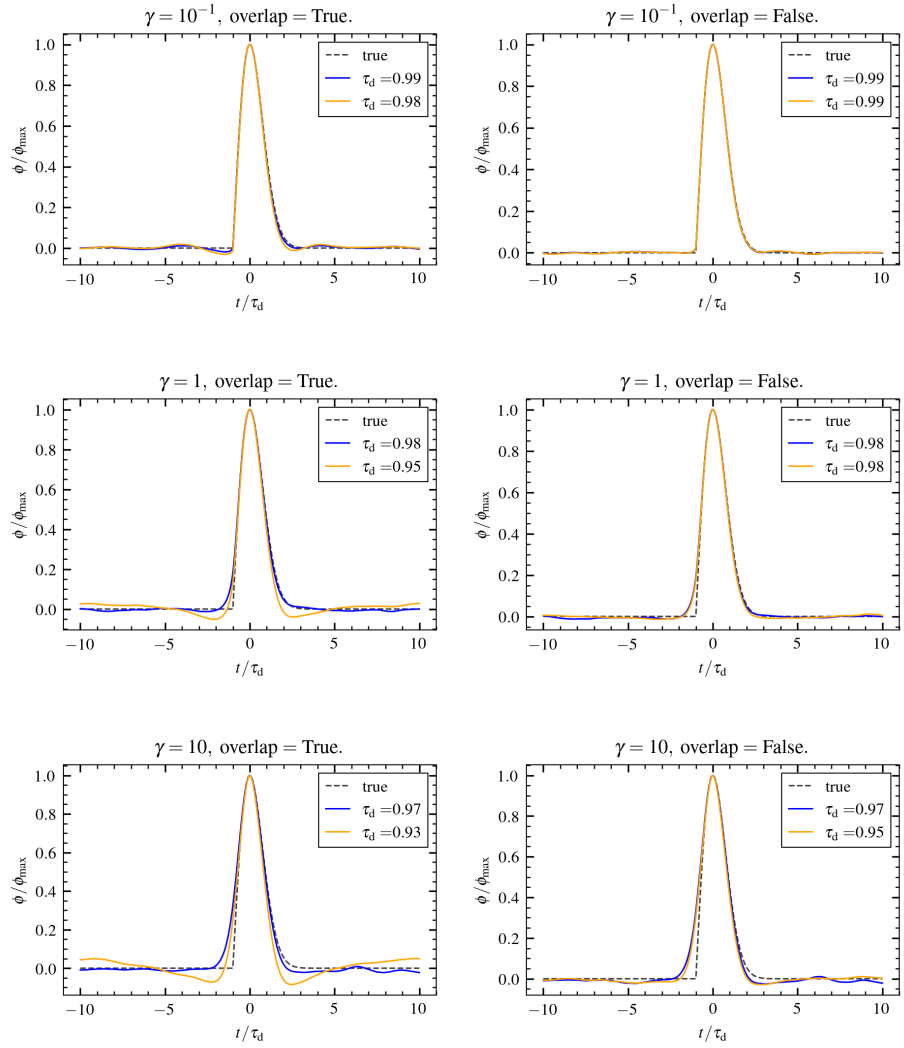


Figure 12: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A Rayleigh pulse function was used to generate the signal.

In figure 12 we see the results of conditional averaging used on a signal with a Rayleigh pulse function as shown in appendix D.8. For $\gamma = 10^{-1}$ both conditions provide waveforms that overlap nicely with the true pulse function. However there is "waviness" in the waveforms where overlapping windows are allowed which is not there with the `overlap=False` condition. This may be an effect caused by multiple parts of the signal being included more than once in the average, due to not using the `overlap=False` condition, however there is no strong evidence for this. The shape is a Rayleigh pulse in the most intermittent case, with a clear difference between the polynomial rise and the Gaussian fall of the pulse. As γ increases this shape become harder to discern. When $\gamma = 1$ this difference between a rise and a fall is less visible, and one could not certainly claim this to be a Rayleigh pulse. The fall looks to be accurately represented, but the rise is slower, giving us something more symmetric in the end. The prominence method also produces two clear troughs about two duration times away from the center of the pulse when overlapping windows are allowed. With `overlap=False` there is no clear difference between the waveforms, as they both yield the same shape. If we look at the $\gamma = 10$ case, we see much of what we saw in the $\gamma = 1$ case but more exaggerated. The pulses are almost completely symmetric in all cases, making it look more like a Gaussian pulse. The troughs of the prominence condition with overlap allowed are also more distinct, while still being located at about the same spot. If we look at the estimated duration times they are consistently lower for all γ, however it is clear from the $\gamma = 1$ and $\gamma = 10$ cases that one might not consider a Rayleigh pulse at first glance, making these duration time estimates

somewhat flawed.

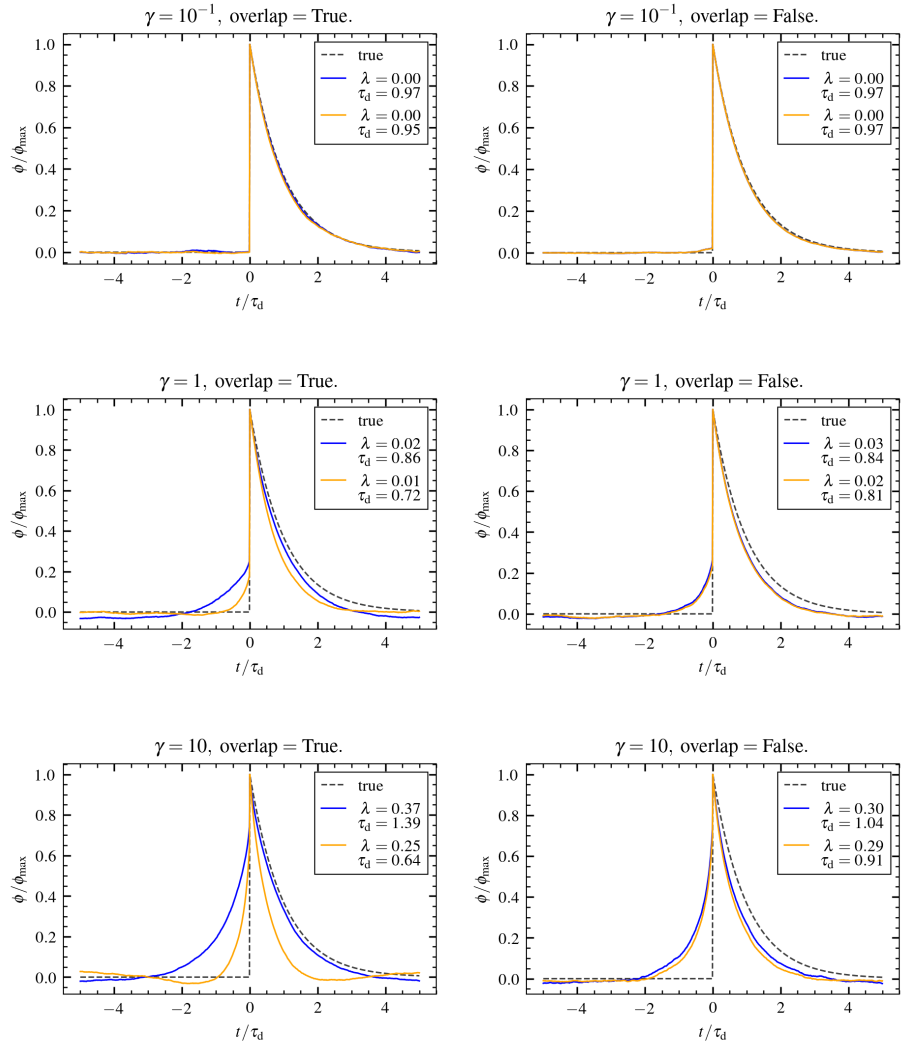


Figure 13: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A one-sided exponential pulse function was used to generate the signal.

Taking the non-symmetric pulse to an extreme we look at the base case in section 3.1.1 with its one-sided exponential pulse from appendix D.1 in figure 13. For $\gamma = 10^{-1}$ we see an accurate representation of the true pulse function for all variations of the condition. The asymmetry parameter λ describes the asymmetry in a double exponential pulse shape as described in appendix D.2, where $\lambda \to 0$ corresponds to a one-sided exponential pulse and $lambda = \frac{1}{2}$ corresponds to a perfectly symmetric double exponential pulse. The asymmetry parameter is accurately estimated and the duration time is only slightly underestimated with no significant difference between the methods. However, issues arise when $\gamma = 1$ both conditions yield a waveform that rises slowly followed by the expected steepness of a one-sided pulse. The fall is then consistently faster for all conditions. The resulting estimated duration time is then smaller as the fall time is shorter. When using the `overlap=False` condition the methods differ little, but in the case of allowing overlap there is a clearer difference. The amplitude threshold condition gives a waveform with both a longer rise and fall time compared to the prominence condition. From the general shape one could still make an educated guess on that the signal is made up of one-sided pulses, as the instant rise is still captured in the waveform. Moving on to the $\gamma = 10$ signal we see that this instant rise of the pulse is completely washed out, yielding what looks like a two sided pulse. This is also reflected in the λ estimates. Again, the fall time is underestimated in all cases, but as the pulses are more symmetric the rise is grossly overestimated. For the prominence condition with allowing

overlap we see some hints of the troughs that we saw for the Rayleigh pulse. They are again located at about two duration times away from the center, and this bias might be what causes this condition to underestimate duration times. This may also be the situation for the $\gamma = 1$ case for the prominence condition with overlap. With the `overlap=False` condition the methods have the same general shape, but they still do not resemble the true pulse function any better as one would not guess the true pulse would be a one-sided exponential pulse just from looking at the conditional average.

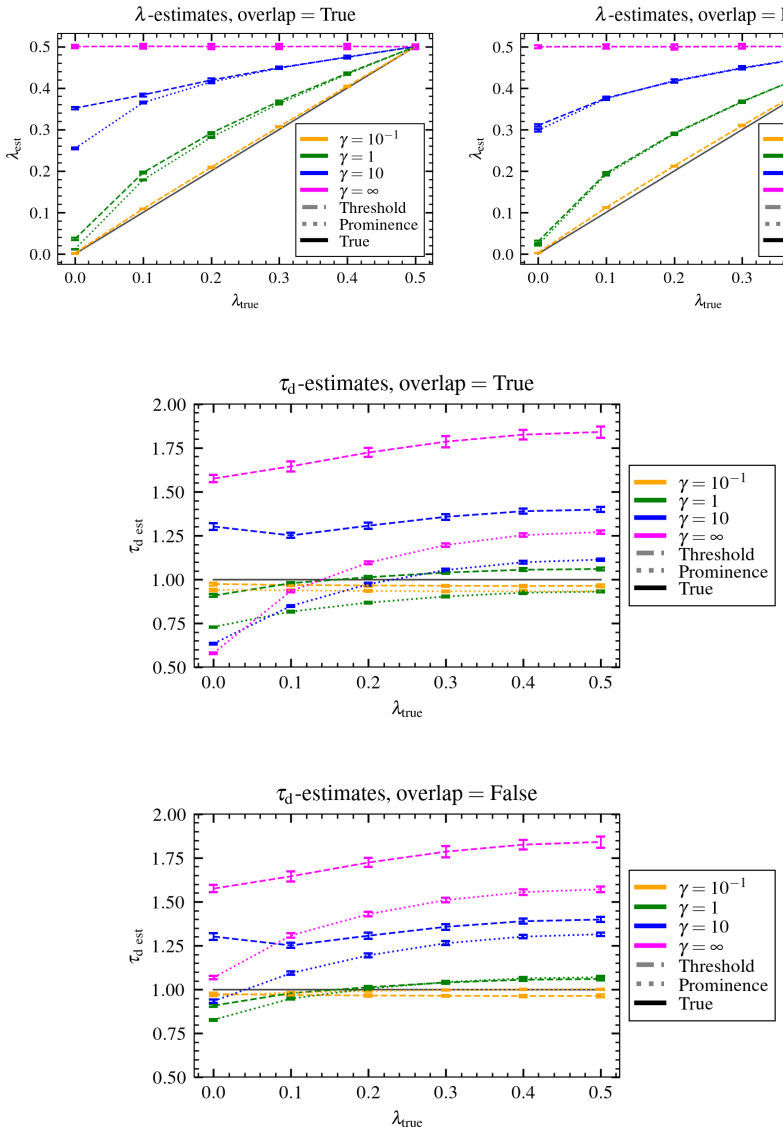


Figure 14: Asymmetry and duration time estimates from the conditionally averaged waveform of different realizations of the FPP with different asymmetry parameters. Each point on the graph represents an average of 100 estimates from 100 realizations of the stochastic process.

In figure 14 we see how asymmetry and duration time estimates from conditional averaging develop with increasing γ for two sided exponential pulses. This figure is based on some excellent earlier work done by Rasmus Nordal [29]. The $\gamma = \infty$ case is approximated numerically by convolving a signal consisting of uncorrelated Gaussian noise together with the double exponential pulse shape. What we observe is that the `overlap=False` condition improves the duration time estimates at $\gamma = 10^{-1}$ and $\gamma = 1$. The asymmetry estimates differ little between the threshold conditions, especially together with the `overlap=False` condition. We also see a general trend in that duration times are often estimated to be larger depending on the asymmetry of the input pulse function, supporting the claim that symmetric pulses generally tend to lead to a larger duration

time estimate from conditional averaging. We also see clearly that the effects of pulse overlap lead to more symmetric estimates, with the $\gamma = \infty$ case yielding completely symmetric estimates independent of input pulse.

The other pulse functions that were analysed was the Lorentz pulse in appendix D.5, the gamma pulse in appendix D.7, the triangle pulse in appendix D.4 and the box pulse in appendix D.6. The figures from these can be found in appendix F.1. In general, from looking at different pulse shapes we can conclude that conditional averaging works very well in reproducing the pulse shape when the signal consists of many distinct pulses that tend not to overlap with each other (when γ is small). There is one exception for this and that is the box pulse as shown in appendix F.1 in figure 62. The reason for this pulse failing so drastically under the effects of pulse overlap is because of the characteristic way box pulses tend to overlap. If successive box pulses arrive closely the rise of the resulting conditional event will look like a staircase with uneven steps. These steps will continue to go up until the either the last pulse have arrived or until the duration time of the first pulses have passed. This results in a staircase shape somewhat inversely mirrored (the rise of the first ascending step becomes the decline of the first descending step) on the other side of the peak. When many such staircase events are averaged out the staircase shapes disappear and we get the gradual slopes we observe leading to a sharp peak in the conditionally averaged waveform.

When pulse overlap becomes more prevalent the waveform results of conditional averaging, independent of conditions, become worse. The degree of error is dependent on the true pulse shape. Conditional averaging works well for symmetric pulses with no instant rises or falls up to $\gamma = 1$ when estimating pulse parameters. The shape of the pulse is also in general preserved for $\gamma = 10$, however the duration of the pulses tend to be overestimated by 10-40% for this regime. The `overlap=False` condition also significantly improves the prominence condition across the board, as this condition yields characteristic troughs in the waveform for higher γ if overlapping windows are allowed. These troughs in turn influences the waveform, artificially reducing the apparent duration times. The `overlap=False` condition also influences the amplitude threshold condition, however not to the same degree, and it is not clear if this yields improved results from simply looking at the produced waveforms. Looking at the non-symmetric pulses however we see that the error is much more distinct as the pulse overlap in the signals increase. When γ increases the conditionally averaged waveforms become more and more symmetric. This influences different pulses in different ways, such as a one-sided exponential pulse still being clear when $\gamma = 1$ because of its instant rise, but an average from a signal consisting of Rayleigh pulses may be misinterpreted as a Gaussian pulse even for $\gamma = 1$. When $\gamma = 10$ the asymmetry is almost completely washed out independent of the chosen threshold type and enforcement of window overlap. The estimated duration time is also consistently lower than the true value for the asymmetric pulses when $\gamma \geq 1$.

To conclude the section we have shown that the conditionally averaged waveforms are heavily dependent on the underlying pulse shape. The method works better at higher degrees of pulse overlap for symmetric pulse functions than asymmetric ones. Asymmetry in the estimate is also lost with higher γ-values. This may lead to incorrect conclusions based on the waveform, meaning one should ideally estimate asymmetry in conjunction with other methods to not be led astray by the effects of pulse overlap.

4.1.2 The effects of noise on the waveform

Noise is a phenomenon any data analyst will come across, making it a necessary point of study for any signal processing method. In this section we will look at the average waveform from conditional averaging for the prominence and amplitude threshold conditions, both with and without the `overlap=False` condition on the averaging windows. We will concern ourselves with two types of noise, observational, also known as additive noise and dynamic noise as described in section 3.1.2. An attempt at finding the best threshold condition will be made and we will establish if the `overlap=False` condition is required or not.

To generate the signals used in the noise analysis the base case 3.1.1 was used with the two noise types described in 3.1.2. Meaning the amplitudes and waiting times are exponentially distributed, the duration time is one, and the pulse function is a one-sided exponential pulse.

From the results in the previous section we know that the waveforms become largely symmetric and give unreliable results when $\gamma = 10$, so we will mainly concern ourselves with $\gamma = 10^{-1}$ and $\gamma = 1$ in this section. If the figures for $\gamma = 10$ are of interest, they can be found in appendix F.2.

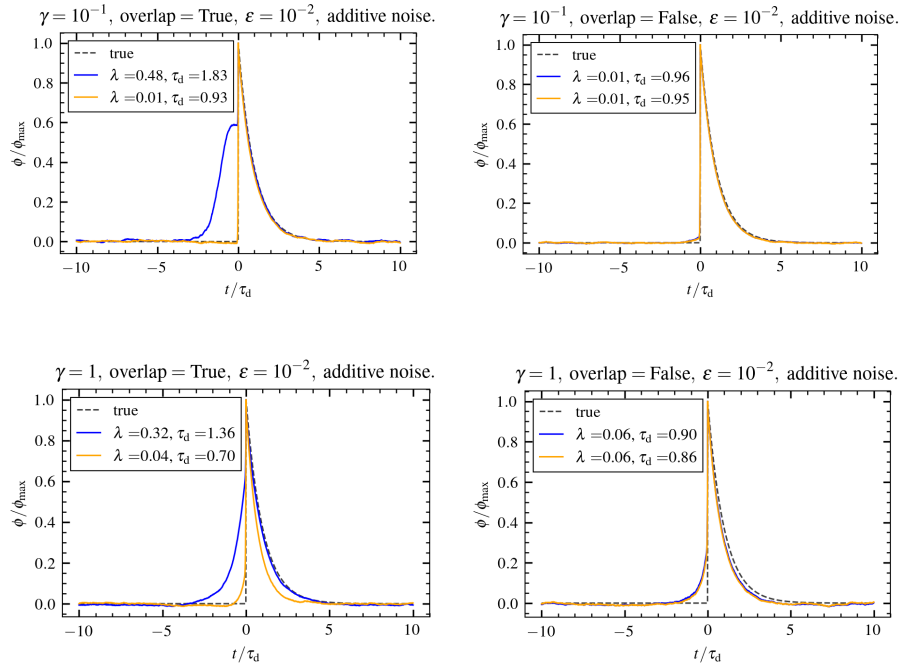


Figure 15: $\varepsilon = 10^{-2}$ Additive noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top) and 1 (bottom). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

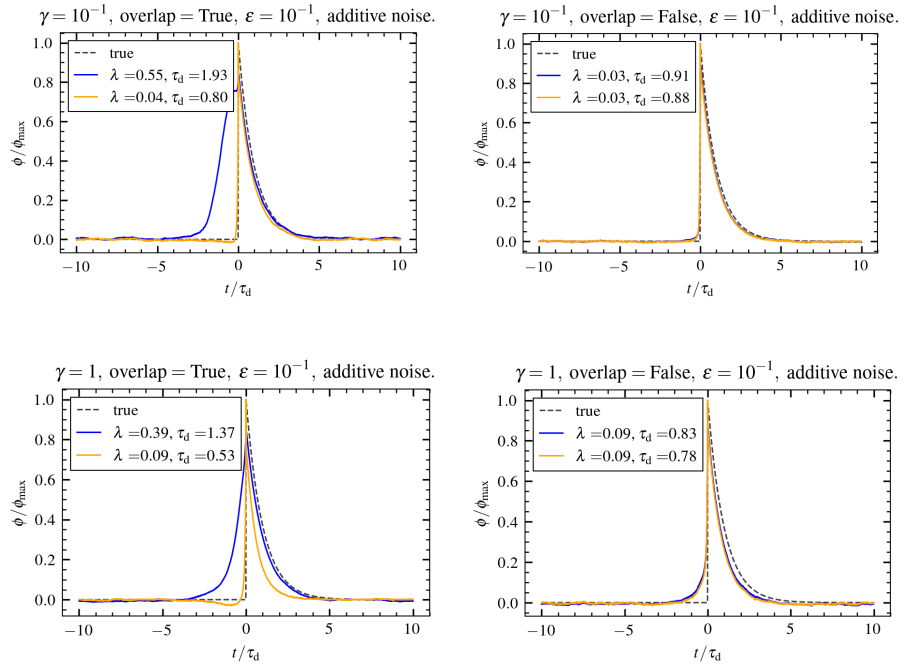


Figure 16: $\varepsilon = 10^{-1}$ Additive noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top) and 1 (bottom). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

In figures 15 and 16 we see the waveform results of conditional averaging used on signals with $\epsilon = 10^{-2}$ and $\epsilon = 10^{-1}$. The figures share the same traits, making it natural to discuss them together. Here we see that the waveform from the amplitude condition is not reproduced accurately if we allow overlapping windows, The fall is rather well reproduced, but the rise is completely wrong. The prominence condition is not affected by this and resembles mostly the same shapes as we saw in section 4.1 without noise. The `overlap=False` condition drastically improves the results from the amplitude threshold while at the same time removing the troughs in the prominence method that we still see hints of. For $\varepsilon = 10^{-1}$ we also see hints of the noise in the waveform itself where the peak has this small jump which comes from the additive noise for both the $\gamma = 10^{-1}$ and $\gamma = 1$ cases.

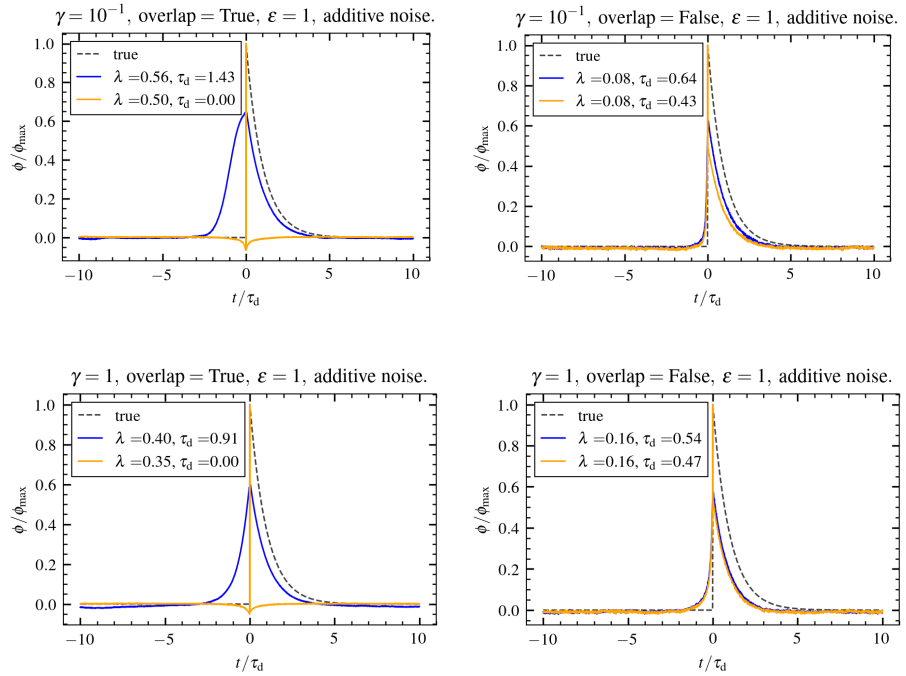


Figure 17: $\varepsilon = 1$ Additive noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top) and 1 (bottom). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

The effects of noise are further exacerbated when the noise to signal fluctuation ratio, ε, is increased as we see in figure 17 where both the prominence and the amplitude threshold methods break down completely without the added `overlap=False` condition. The reasons for breakdown are however different. The amplitude threshold method breaks down because of noise jitter just before and just after the crossing of the threshold caused by an underlying event. To explain this we can look to figure 5. Say the second lime peak from the left is an underlying event, then small peaks caused by noise around the threshold level will cause the signal to dip up and down around the threshold during the rise and the fall of the underlying pulse. This will result in many peaks like the orange peak that are close to a taller peak, but are being registered because it dips below the amplitude threshold before the taller peak. The prominence method breaks down because the noise itself fluctuates on scales comparable to the signal, making noise events much more prominent. There are as many noise events as there are signal data points (10^7), meaning there are significantly many more prominent noise events than prominent underlying signal events. This leads to the noise events dominating the average, resulting in a simple spike on the average waveform. However, the `overlap=False` condition again drastically improves the waveform estimate, where they resemble more of the structures we saw in 13 but with a clearly expressed noise jump at the center. There is also little difference between the two threshold conditions when using the `overlap=False` condition, making us unable to say one is better than the other. The parameter estimates are also clearly affected by this jump, meaning one should be careful when estimating pulse parameters from a signal with additive noise, with this care being more important with increasing ε.

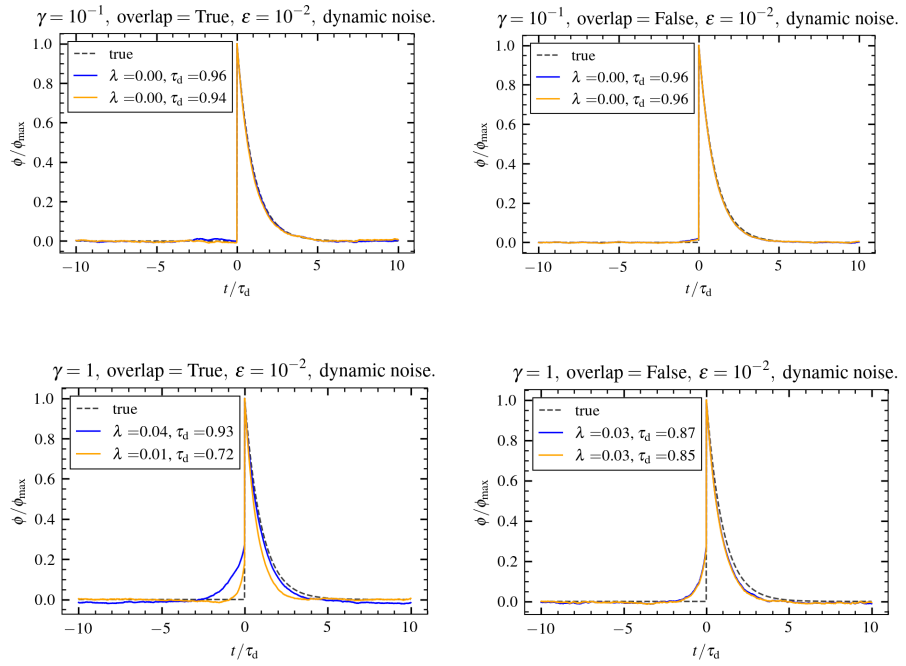


Figure 18: $\varepsilon = 10^{-2}$ Dynamic noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top) and 1 (bottom). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

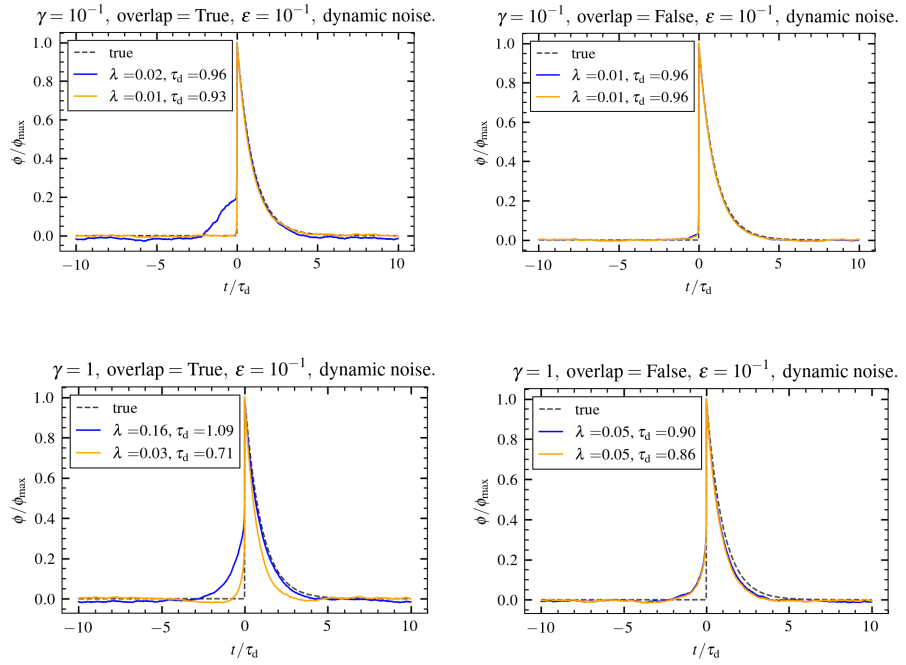


Figure 19: $\varepsilon = 10^{-1}$ Dynamic noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top) and 1 (bottom). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

For dynamical noise in figure 18 the results largely resemble what we saw in figure 13. Both methods are viable for low γ and the same issues of an underestimated fall time together with a small rise before the true instant rise being apparent in the $\gamma = 1$ case. But again, we see that with more noise comes more issues in figure 19. The prominence waveform is not affected by the noise in comparison to the amplitude threshold waveform. Even for small γ without enforcing overlap we again see a small bump before the steep rise with this bump looking more exponential in the $\gamma = 1$ case. However, this issue is again mitigated by the `overlap=False` condition, yielding similar results for both threshold conditions.

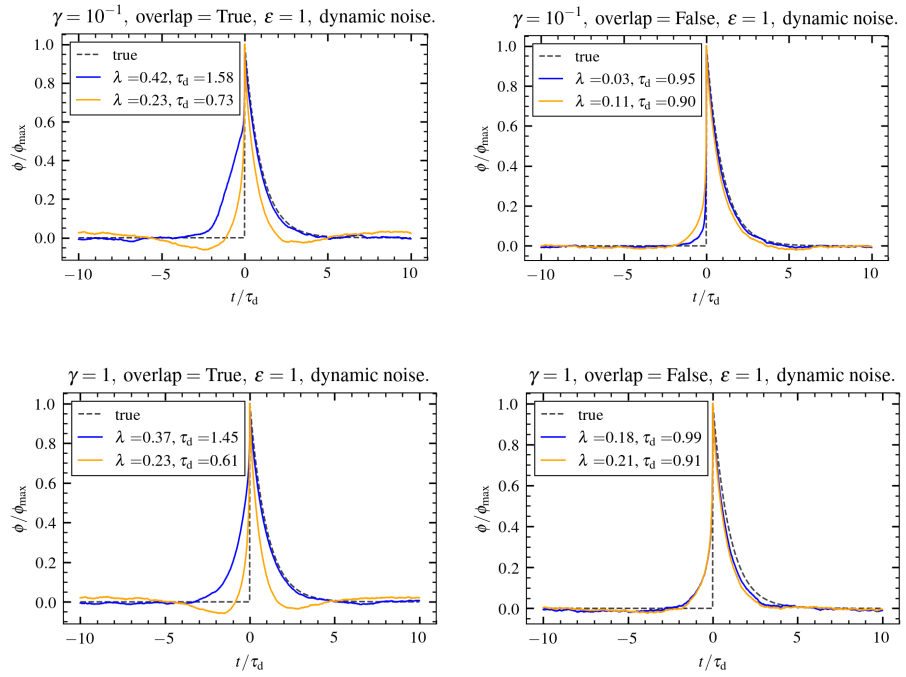


Figure 20: $\varepsilon = 1$ Dynamic noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top) and 1 (bottom). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

In figure 20 we see the the waveform results of conditional averaging when $\varepsilon = 1$ for dynamical noise. There is not much different from the $\varepsilon = 10^{-1}$ case, however the $\gamma = 1$ case with `overlap=False` looks more symmetric than the $\varepsilon = 10^{-1}$ case which is also reflected in the asymmetry estimate λ.

After having looked at both additive and dynamic noise the most obvious takeaway is that the additional `overlap=False` condition improves the waveform estimate for any degree of intermittency. The only case where one might argue it does not make a difference is in the dynamic noise case with $\varepsilon = 10^{-2}$ and $\gamma = 10^{-1}$. We hoped that the prominence condition would inherently be insensitive to noise because of a preference for an inherent distance between peaks because of the nature of decay between events and how prominence is defined. This is supported for all but the $\varepsilon = 1$ case of additive noise. The waveforms look similar to the ones observed in section 4.1 without the `overlap=False` condition. However the prominence conditions issues of the troughs persists, making the `overlap=False` condition necessary despite what we had hoped. These troughs show up only in the cases without `overlap=False` and represents events that are close together with significant prominence each, thus the slow decay and rise before and after the troughs respectively are a result of the neighbouring peaks that average out to a generally larger level than what would be the case if we use the `overlap=False` condition. The amplitude threshold method is significantly improved result wise with `overlap=False`. Without this condition the method picks up lots of noise jitters that just rises above amplitude threshold, often during the rise or decay of an underlying event. An interesting observation is that the different noise types also come with different shapes on the waveform for this condition which might be because of the dynamical noise fluctuates on a longer timescale than the additive noise, making the threshold crossings fewer

and therefore affecting the average in a different way. Additive noise also looks to be the type that affects the average the most, with the distinct jump on $t = 0$ only showing for this type of noise for larger ε. We have also observed that the threshold conditions have negligible differences between them when using the `overlap=False` condition. Comparing our high intermittent cases with the results of earlier studies [20, 21] we arrive at the same conclusion. Conditional averaging is well suited to reproduce the shape of underlying events in the face of noise if one introduces an additional condition that ensures some robustness against picking up the same underlying event multiple times due to noise fluctuations. The `overlap=False` condition being one example of such a condition as long as one chooses a window size that is not too small. From these results we can conclude that the additional `overlap=False` condition is necessary to obtain the most accurate waveform compared to the true pulse shape, especially when there is noise in the signal. This is relevant independent of the intermittency as results are improved for both highly intermittent and non-intermittent signals with the `overlap=False` condition.

4.1.3 Mixed waveforms

Underlying events may have different shapes depending on their sizes. This may come in the form of completely different shapes at different amplitudes, perhaps indicating a mix of different processes, or in the form of the symmetry varying with amplitude, possibly displaying an underlying property of the process. With conditional averaging we can adjust the threshold to look for different shapes between different thresholds, possibly uncovering such a correlation. The filtered Poisson process also allows for such a correlation when generating realizations as we can use the distributed amplitudes to determine the pulse shape as well.

To correlate the amplitude distribution with the pulse function the asymmetry parameter λ of the double exponential pulse function was made a function of the amplitudes according to

$$\lambda(A) = \begin{cases} 0.5, & 0 \leq A < a \\ 0.4, & a \leq A < b \\ 0.3, & b \leq A < c \\ 0.2, & c \leq A < d \\ 0.1, & d \leq A < e \\ 0, & e \leq A. \end{cases} \tag{4.1}$$

What this relationship does is that it makes pulses more asymmetric as the amplitudes grow, with completely symmetric pulses at the lowest amplitudes and one-sided pulses at largest amplitudes.

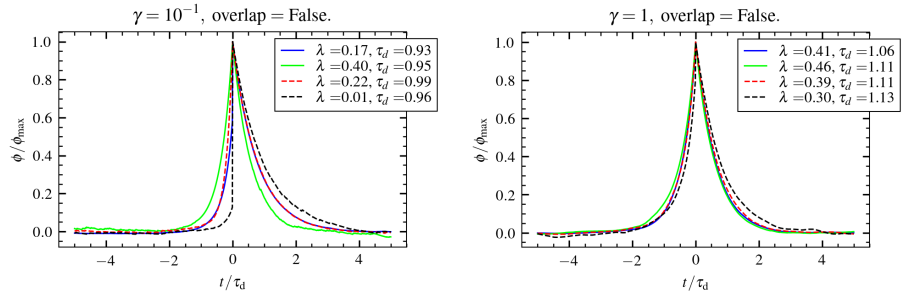


Figure 21: Conditionally averaged waveforms and the estimated λ and τ_d for $\gamma = 10^{-1}$ (left) and 1 (right). The thresholds are > 2.5 (blue), $2 - 4$ (lime), $4 - 6$ (red) and $6 - 8$ (black) in units of the signals rms value. The a, b, c, d, e values were $2, 4, 4, 6, 6$ in units of the signals rms value plus the mean of the signal. The input amplitude distribution was Exp(1).

In figure 21 the results of conditional averaging between different amplitude thresholds are shown for a signal with waveforms correlated with amplitudes so that one there is simply one waveform within each interval, $\lambda = 0.5$ between 0 and 2 standard deviations, $\lambda = 0.4$ between 2 and 4 standard deviations, $\lambda = 0.2$ between 4 and 6 standard deviations and $\lambda = 0$ for any events

with amplitudes above 6 standard deviations. We see from the waveforms that in the $\gamma = 10^{-1}$-case the averaged waveform estimates coincide well with the pulses placed within the threshold windows. However, in the $\gamma = 1$-case this distinction is not as clear. There is an overall decrease between the increasing threshold windows, but not as steep as the true relationship between pulses and waveform asymmetry. It is evident that the effects of overlap diminish our abilities to discern the true asymmetry parameter, even for the most idealized example at $\gamma = 1$. For $\gamma = 10$ it is hard to create such an idealized scenario with thresholds above two standard deviations as the probability of singular events crossing the threshold vanishes at this γ.

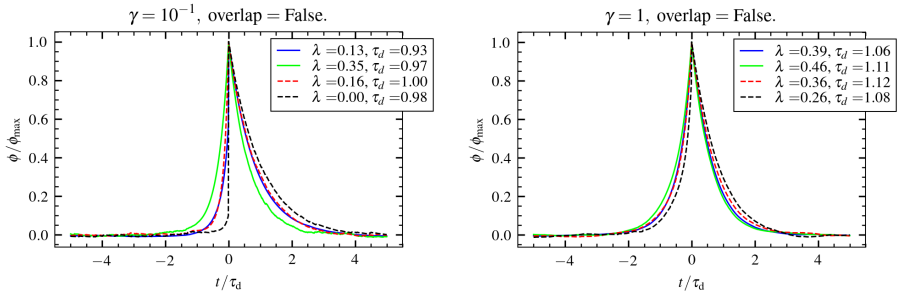


Figure 22: Conditionally averaged waveforms and the estimated λ and τ_d for $\gamma = 10^{-1}$ (left) and 1 (right). The thresholds are > 2.5 (blue), $2 - 4$ (lime), $4 - 6$ (red) and $6 - 8$ (black) in units of the signals rms value. The a, b, c, d, e values were $2, 3, 4, 5, 6$ in units of the signals rms value plus the mean of the signal. The input amplitude distribution was Exp(1).

Similarly in figure 22 we have an idealized example with certain pulses carefully placed in different threshold windows depending on their amplitude. Between 0 and 2 standard deviations the pulses are symmetric with $\lambda = 0.5$, within 2-4 standard deviations there are both $\lambda = 0.4$ and $\lambda = 0.3$ pulses, between 4 and 6 standard deviations there are $\lambda = 0.2$ and $\lambda = 0.1$ pulses, while anything above 6 standard deviations is a one-sided $\lambda = 0$ pulse. In the figure for $\gamma = 10^{-1}$ we see the results of conditional averaging between these threshold windows, and we see that the asymmetry estimates resemble an average between the waveforms within each threshold window, as we would expect from the previous example. However just as in the previous figure we see that this accuracy quickly vanishes as $\gamma = 1$ where we are only able to discern a decreasing relation between higher amplitudes and the asymmetry parameter without accurately estimating the true waveforms.

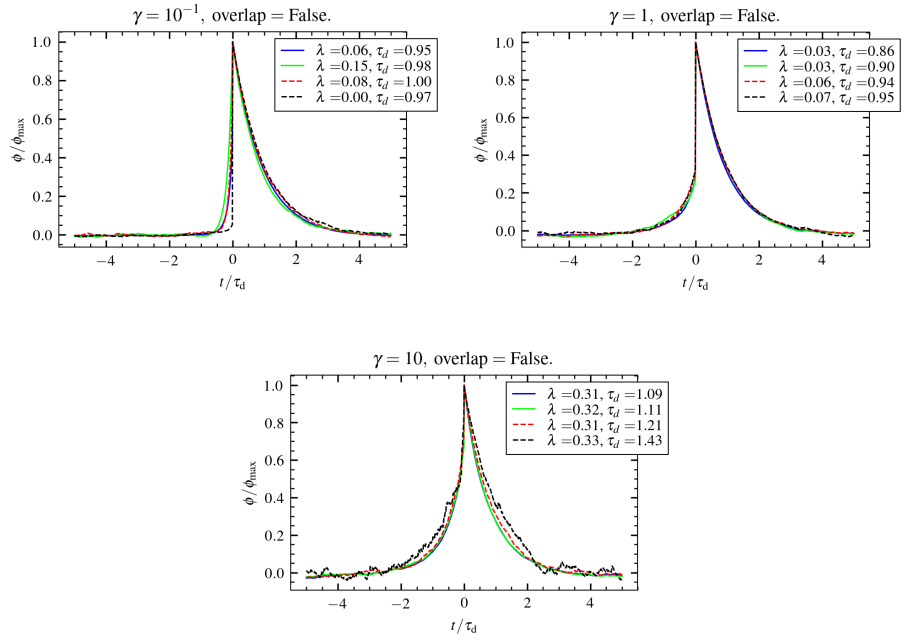


Figure 23: Conditionally averaged waveforms and the estimated λ and τ_d for $\gamma = 10^{-1}$ (left), 1 (right) and 10 (bottom). The thresholds are > 2.5 (blue), $2 - 4$ (lime), $4 - 6$ (red) and $6 - 8$ (black) in units of the signals rms value. The intervals were spaced so that an equal amount of events were made for each value of λ. The input amplitude distribution was Exp(1).

It is highly unlikely that one would come across such highly idealized examples. To construct another example we make it equally likely for any event to be in any asymmetry bracket while still making the correlation be amplitude dependent. This was done, giving us an equal amount of events for each λ-value and the results of conditional averaging used on such a realization can be seen in figure 23. In the $\gamma = 10^{-1}$ case we can see that the averaged waveforms give us again this λ-average in each bracket, showing us that for this example most of the threshold crossings are due to the higher amplitude events with an asymmetry of $\lambda = 0.2$ and below. From this, one would not be able to conclude the true distribution of an equal number of events in each bracket, and perhaps infer that the signal consists of largely asymmetric events. In the $\gamma = 1$ case the distinction is even harder to make where most threshold crossing stem from the $\lambda = 0$-events and overlap effects, making the more symmetric events mostly undetected by conditional averaging. This makes the underlying distribution impossible to detect in this case, even suggesting symmetry increases slightly for larger pulses. For $\gamma = 10$ the waveform estimates about the same independent of threshold bracket. The effects of pulse overlap dominate the averages, making this method of detecting differences in waveforms on different amplitude scales entirely unreliable at this γ-scale.

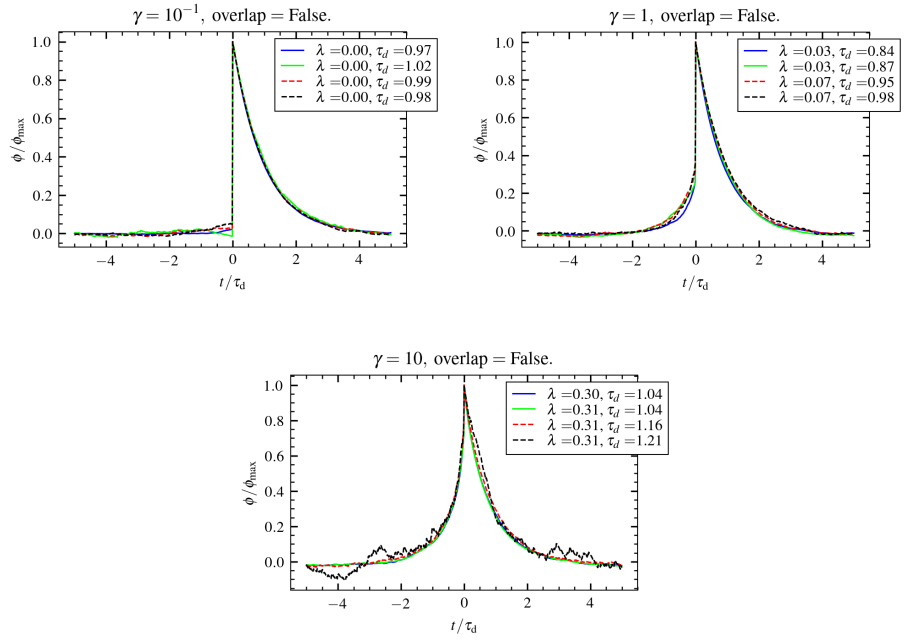


Figure 24: Conditionally averaged waveforms and the estimated λ and τ_d for $\gamma = 10^{-1}$ (left), 1 (right) and 10 (bottom). The thresholds are > 2.5 (blue), $2-4$ (lime), $4-6$ (red) and $6-8$ (black) in units of the signals rms value. There is no correlation between pulse shapes and amplitudes in the signal. The input amplitude distribution was Exp(1).

As a reference we also use this method on a signal with no correlation between amplitudes and pulse shape in figure 24. We see little difference in the estimated asymmetries between the threshold brackets, leading us to conclude that increasing pulse overlap does not lead to observed correlations between amplitudes and the shape of the waveform. The duration time estimates change depending on the bracket for $\gamma = 1$ and $\gamma = 10$. This may be due to the effects of overlap being normalized out to different degrees depending on the amplitude of the conditional events. If we for simplicity assume that overlap affects the pulse shape by raising or lowering the rise and tail of underlying large amplitude events by a fixed amount, then normalizing based on peak amplitudes will make the effect of overlap appear stronger for the smaller conditional events within the average, while lager events will appear with more suppressed effects of overlap in the average due to normalizing by a bigger peak value.

From looking at amplitude correlated pulse-shapes we have observed that conditional averaging only works reliably in the highly intermittent case where one obtains more accurate asymmetry estimates within each threshold bracket. This was clear from the highly idealized cases with the correlation being tailored to the specific threshold windows. However, the methods ability to accurately pick out the true waveform within these brackets disappear even for moderately intermittent signals with $\gamma = 1$ where one is only able to uncover the general decreasing trend between symmetry and increasing amplitudes. For $\gamma = 10$ we observed that the ability to pick out any difference in waveforms based on amplitude thresholds is largely lost, as the effects of pulse overlap cause almost all threshold crossings at this intermittency scale. Using conditional averaging in this way for signals where overlap is even somewhat apparent is problematic for accurate waveform representations, and one should only use it to find general trends in the correlation at when γ is on the order of one. For highly intermittent signals this way of predicting the underlying correlation works fine. Although if the correlation is more continuous in nature one might have to divide up the intervals even more, making even longer signals necessary for statistically robust results as smaller and more brackets will inevitably have fewer events within them per bracket.

4.2 Amplitude distribution analysis

Conditional averaging as we use it picks out peaks in a signal based on different threshold conditions. One important piece of information about the peaks are their heights which have been used to estimate the amplitude distributions of the underlying events of the signal [13, 28, 30, 31]. In this section we will examine the empirical amplitude and prominence distributions obtained through conditional averaging using the amplitude and prominence thresholds respectively. The distributions will be compared to the amplitude distributions used to generate the signals and the signals conditional distribution above the set threshold as well. As we previously found the `overlap=False` condition is vital for good waveform results all the analysis in this section will be done with that condition as well.

In this section all signals were generated according to the base case 3.1.1 but with different amplitude distributions. Meaning the waiting times are exponentially distributed, the duration time is one and the pulse function is a one-sided exponential. In addition, the mean amplitude is always set to be one, where the scale parameter of the distribution is used to specify the mean according to how the shape parameter (if there is any) is chosen. The notation $\langle A \rangle_{\text{ind}}$ denotes the mean of the values above the threshold for each individual distribution shown. For instance for the amplitudes from the conditional average it denotes the mean of these amplitudes, while for the signal probability distribution it denotes the mean of the signal values above the same threshold and for the theoretical distribution it denotes the mean above the threshold B.6. For the prominence method the prominence distribution is shown, which is strictly not the same as the amplitude distribution, however we will investigate if this distribution is a good estimate for the amplitude distribution. The main distributions we will investigate is the exponential, Lomax, Rayleigh and degenerate distributions. The reason for choosing these distributions are the their properties. The exponential distribution is the one used in the base case and is often one that is used in the literature [13, 31]. The Lomax distribution is investigated because of its heavy tail, making larger amplitudes more likely which would help us investigate a broader range of amplitudes. The Rayleigh distribution would help us investigate if any unimodality is preserved and to see the effects of an even steeper tail than for the exponential distribution. And finally, the degenerate distribution allows for a nice idealized test case where all amplitudes are the same, and we essentially look at the effects of pulse overlap in isolation. It turns out that the amplitude distribution estimates from conditional averaging often align with the conditional probability density function of the signal itself in the case of the amplitude threshold. If it also does not follow the true distribution, then one could interpret this as the method effectively picking out random points on the signal instead of being biased toward underlying events. The points that conditional averaging cannot be seen as truly random points on the signal as the method picks out peak values, however these peak values tend to look like random points on the signal as the effects of overlap increase. This would be a major issue as conditional averaging is hypothesized to pick out individual underlying events, making conclusions drawn from the conditional average amplitude distribution estimates inaccurate. For the prominence threshold it is hard to make the same case, as what it provides is a prominence distribution. The reason we don't compare the prominence distribution to the signal distribution is that it is only the peaks in the signal have prominence values. For this same reason one cannot obtain a prominence distribution of the signal, and therefore it is hard to check if the prominence distribution estimated from conditional averaging converges to the prominence distribution of the signal itself.

This section is structured as follows. Firstly we look at the effects of different amplitude distributions in the waveform estimate, discussing distributions together in order of increasing γ. Then we discuss the amplitude estimates themselves, comparing them to each other, the underlying amplitude distribution and to the conditional distribution of the signal itself. In the end we conclude the section with the most important takeaways from this analysis.

4.2.1 Effects of different amplitude distributions on the averaged waveforms

In this section we will discuss the effects of different amplitude distributions on the conditionally averaged waveform. There is little difference between the estimates from the two methods when it comes to the waveform with the `overlap=False`, meaning we will discuss them together. We will discuss them in increasing order of γ comparing them to one another between input amplitude distributions.

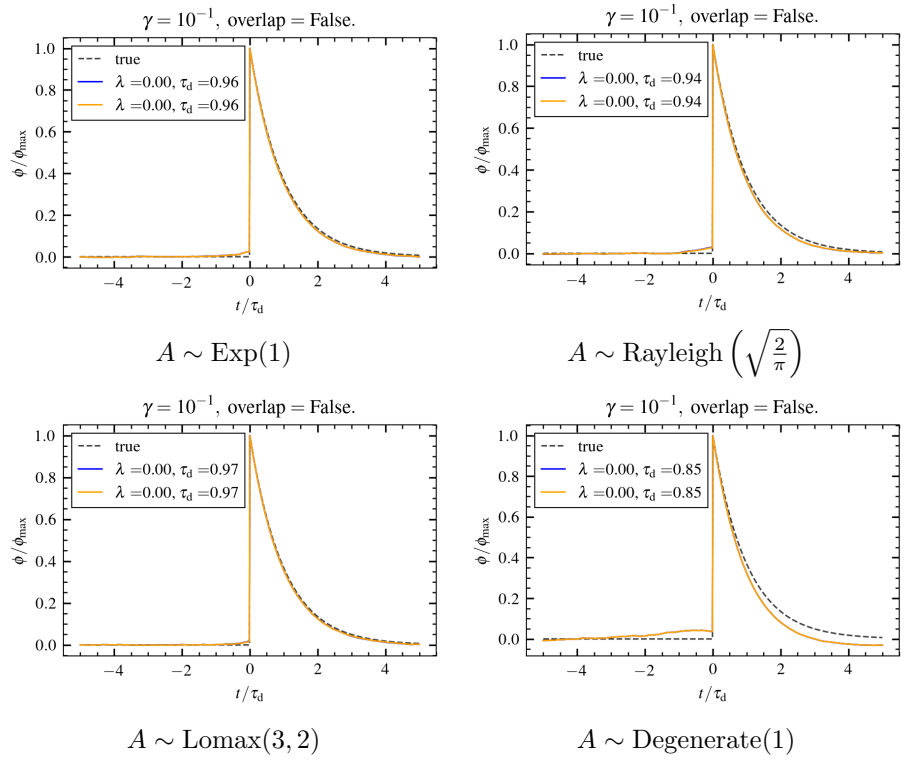


Figure 25: $\gamma = 10^{-1}$. Conditionally averaged waveform for both methods. The distributions denote the underlying amplitude distributions of each signal that the methods were used on.

Starting with the highly intermittent case in figure 25 where we see the different waveform estimates from conditional averaging used on signals with different input amplitude distributions. We that three of the four waveforms are reproduced accurately, except for the signal with degenerately distributed amplitudes. The waveform estimate in this case is affected by the underlying amplitude distribution in the form of a small, elevated region. This region may stem from a combination of the waiting time distribution and the degenerate amplitudes. In cases where two pulses are closely spaced together, as in within one window size of each other, the conditional averaging method would pick up the taller event and make an excerpt around this peak. In the case where every event has the same amplitude the taller signal value is always at the peak of the event that arrived the latest as it will arrive on top of the decay of the first event. As the excerpt is around the taller peak the earlier peak presents itself in the average as we see in this small, elevated region. With a small probability of pulse overlap this does not happen too often in highly intermittent signals, but we still see that there is some effect from it on the waveform. The reason this does not happen in the case of other distributions, despite having the same amplitude mean, is that for those distributions the latest peak within an averaging window is not always the tallest. If an earlier event is larger, then that event is most likely going to be the one picked up by the method (as long as it is not within one window size of an even taller, earlier event). The excerpt is thus taken around the earlier event, effectively lowering the "mean level before arrival" compared to the degenerate distribution where this mean is higher because all events have the same amplitude.

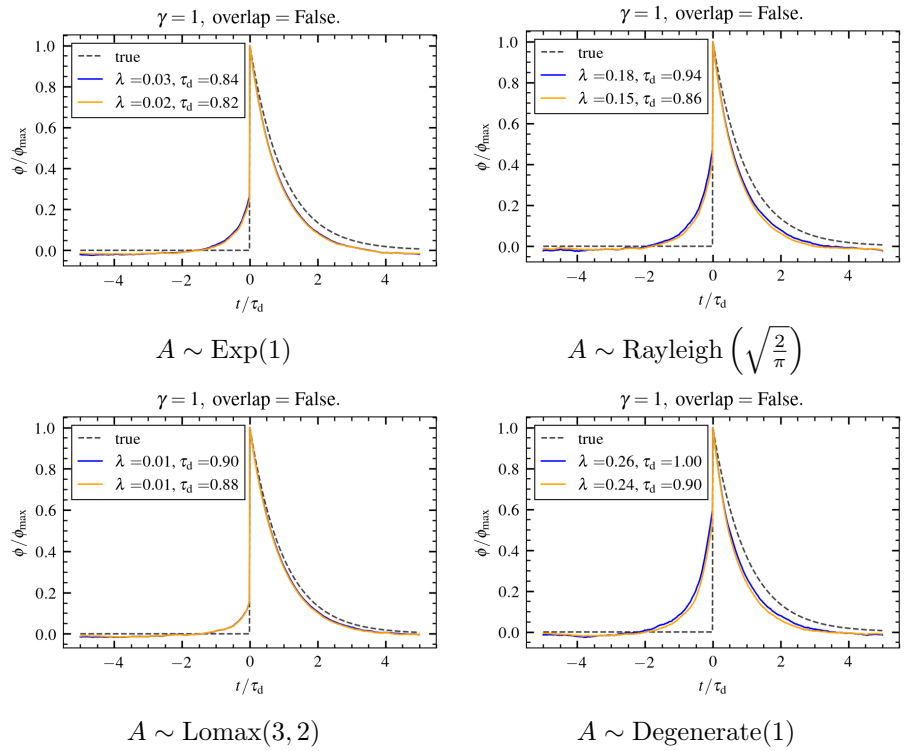


Figure 26: $\gamma = 1$. Conditionally averaged waveform for both methods. The distributions denote the underlying amplitude distributions of each signal that the methods were used on.

In figure 26 we see the estimated waveforms from conditional averaging used on signals with differently distributed amplitudes. The waveform from the exponentially distributed amplitudes resemble the same traits as we found in section 4.1, with a slight rise before the instantaneous jump of an underlying event. However, what is interesting is that this rise is evidently different in its appearance on the average waveform for all the different distributions. We start to see a trend in how the waveform estimate is affected by the input amplitude distributions. The heavier the tail of the input distribution the smaller this initial rise before the jump for this particular pulse shape. This also makes sense as in signals with with a bigger large-to-small ratio of amplitude events there would be more events that "stick out" distinctly from the rest of the signal, making the effects of pulse overlap on the waveform not as pronounced in the average. For a signal dominated by small amplitude events most threshold crossings would be due to pulse overlap, no matter if it is an amplitude or a prominence threshold, making the average being dominated by crossings of this type instead of singular events exceeding the threshold. Thus, the waveform estimate of conditional averaging might be more accurate depending the probability of large amplitude events in the signal it analyzes.

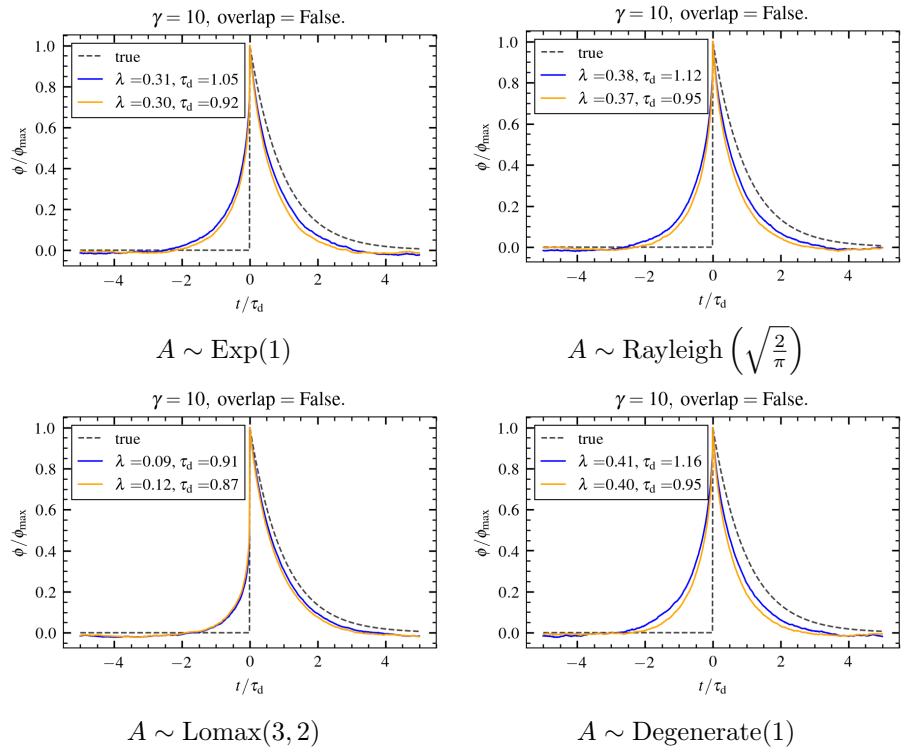


Figure 27: $\gamma = 10$. Conditionally averaged waveform for both methods. The distributions denote the underlying amplitude distributions of each signal that the methods were used on.

Looking at the conditionally averaged waveforms from signals with differently distributed amplitudes in the $\gamma = 10$ case we observe much the same as in the $\gamma = 1$ case. From our earlier waveform analysis in section 4.1 we concluded that the results of the method were largely inaccurate for asymmetric pulses at this intermittency. However here we see that the effects of pulse overlap can be mitigated by broadening the amplitude distribution, making waveforms from signals with $\gamma = 10$ appear closer to the underlying waveform. The broader the underlying distribution, the smaller the initial rise and the better the estimates. This further supports the hypothesis that the results of conditional averaging are more robust when the underlying amplitude distributions are heavy-tailed.

4.2.2 Amplitude distribution estimates

In this section we will look at signals with differently distributed underlying amplitudes to see if the estimates from conditional averaging coincide with the true amplitude distribution or not. We will also compare the amplitude estimates between the two threshold conditions, attempting to establish if either one is better suited to amplitude estimation than the other.

Highly intermittent signals $\gamma = 10^{-1}$

To start this analysis we will explore highly intermittent signals, where the effects of pulse overlap are negligible. Here we expect to accurately reproduce the underlying amplitude distribution as the large majority of conditional events will coincide with underlying events because of negligible pulse overlap.

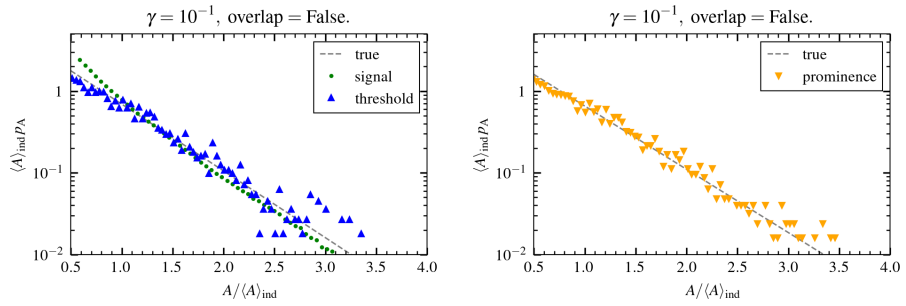


Figure 28: $\gamma = 10^{-1}$ $A \sim \text{Exp}(1)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

In figure figure 28 we see the amplitude distribution estimates from conditional averaging used on a signal with exponentially distributed amplitudes. We see that both the prominence and amplitude threshold methods closely follow the true conditional amplitude distribution in an exponential tail. However, the signals conditional distribution is also exponential, making the estimate look like good reproductions of both the signal distribution and the true amplitude distribution.

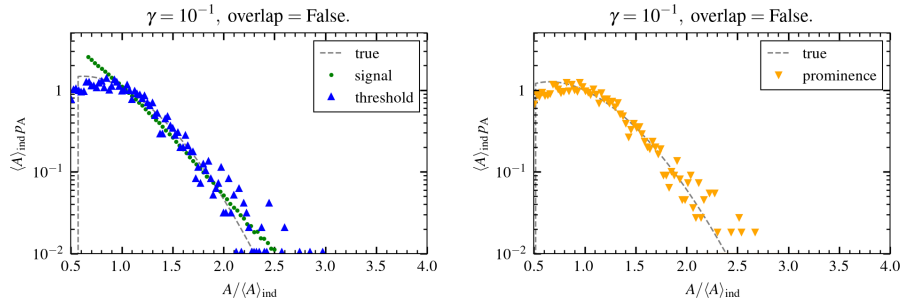


Figure 29: $\gamma = 10^{-1}$ $A \sim \text{Rayleigh}\left(\sqrt{\frac{2}{\pi}}\right)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

In figure 29 we see the amplitude distribution estimates from conditional averaging used on a signal with Rayleigh distributed amplitudes. We see that both the prominence and amplitude threshold methods closely follow the true conditional amplitude distribution initially, but as it tails off it is hard to say if the tail is Gaussian or exponential due to the lack of data. The signal distribution has an exponential tail, which could look to align with the estimates from conditional averaging for larger amplitudes, but it is still not a distinct enough difference that we can rule out which distribution the estimated distributions follow.

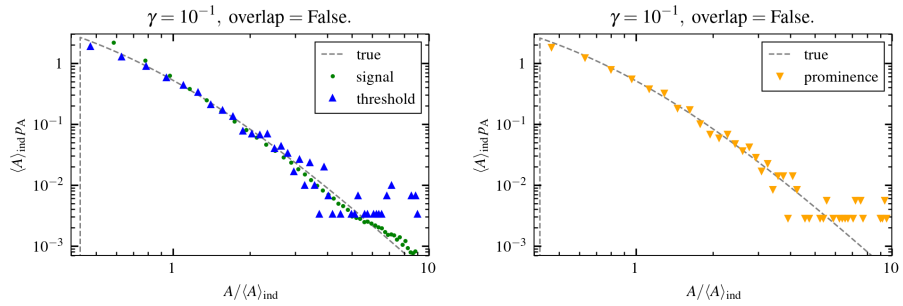


Figure 30: $\gamma = 10^{-1}$ $A \sim \text{Lomax}(3, 2)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

In figure 30 we see the estimated amplitude and prominence distributions from a signal with Lomax distributed amplitudes. The estimated amplitude distributions closely follow the true amplitude distribution, but as the signal distribution is also power-law distributed we must again conclude that the estimate from conditional averaging reproduces both the signal and the true amplitude distributions.

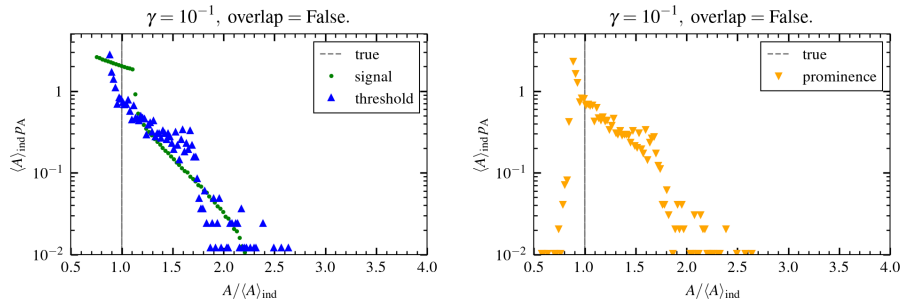


Figure 31: $\gamma = 10^{-1}$ $A \sim \text{Degenerate}(1)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

Looking at the conditional averaging results from a signal with degenerate amplitudes in figure 31 we see that the estimated amplitude distributions exhibit a clear spike which indicates that the methods mainly pick up one amplitude which is what we would expect. The reason this spike does not align with the true distribution is because of the normalization mentioned at the start of this section. The conditional signal distribution also has a sharp jump before it tails off exponentially. The reason for this jump is cases of pulse overlap. In highly intermittent signals it us unlikely that there is pulse overlap, thus the density before the jump represents the region up until the mean of the amplitudes. Anything above this amplitude mean will be unlikely because of pulse overlap, which is why we see this sharp jump in probability at this value. The estimated distributions do not match up with the signal distribution, giving us some confidence that the method might be accurate for highly intermittent signals.

From looking at the results with different input amplitude distributions in the case when $\gamma = 10^{-1}$ we can see that the amplitude estimates generally reproduce both the signal and the true distribution well.

Moderately intermittent signals $\gamma = 1$

In this part we will move on to look at signals with an intermittency of $\gamma = 1$ where individual events and the effects of pulse overlap are both apparent. It is on this intermittency that we hope to see how pulse overlap starts to affect the amplitude estimate.

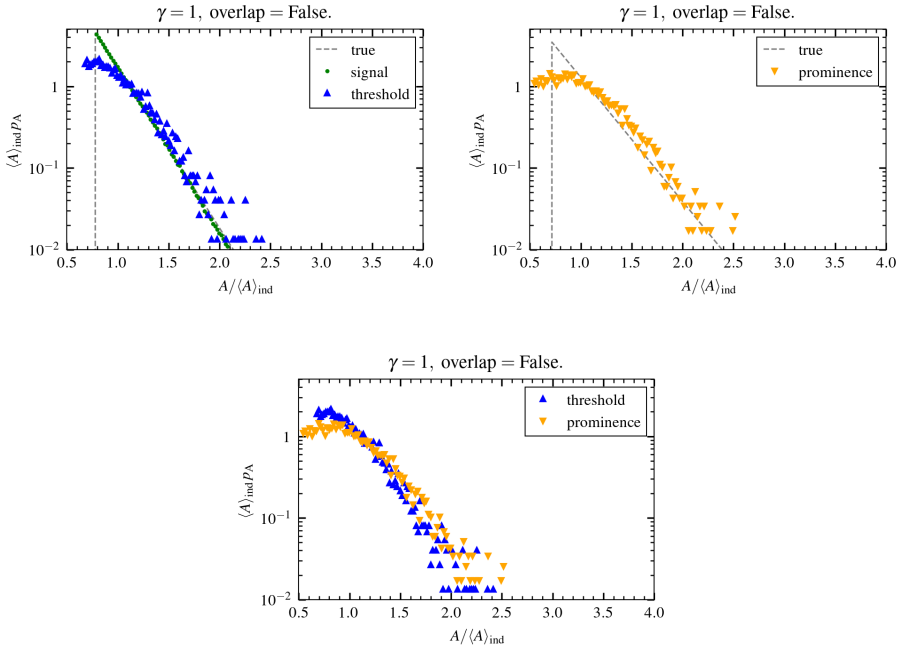


Figure 32: $\gamma = 1$ $A \sim \text{Exp}(1)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

In figure 32 we see the amplitude distribution estimates from conditional averaging used on a signal with exponentially distributed amplitudes. The distribution estimates are again exponential together with the signal distribution, giving us the same conclusion as earlier with the estimate reproducing both. We also start to see some difference between the estimates from the different types of condition, with the prominence estimate looking broader with an initial flatter region.

In figure 33 we see the amplitude estimates obtained from conditional averaging used on a signal with Rayleigh distributed amplitudes. From the amplitude threshold method, we can clearly see that the estimated distribution follows the signal distribution rather than the true distribution, supporting the claim that conditional averaging might simply pick out random points on the signal as γ increases. The estimate from the prominence method also fails to predict the true distribution, with the distribution having obtained an uni-modal shape in contrast to the Gaussian tail from the true distribution.

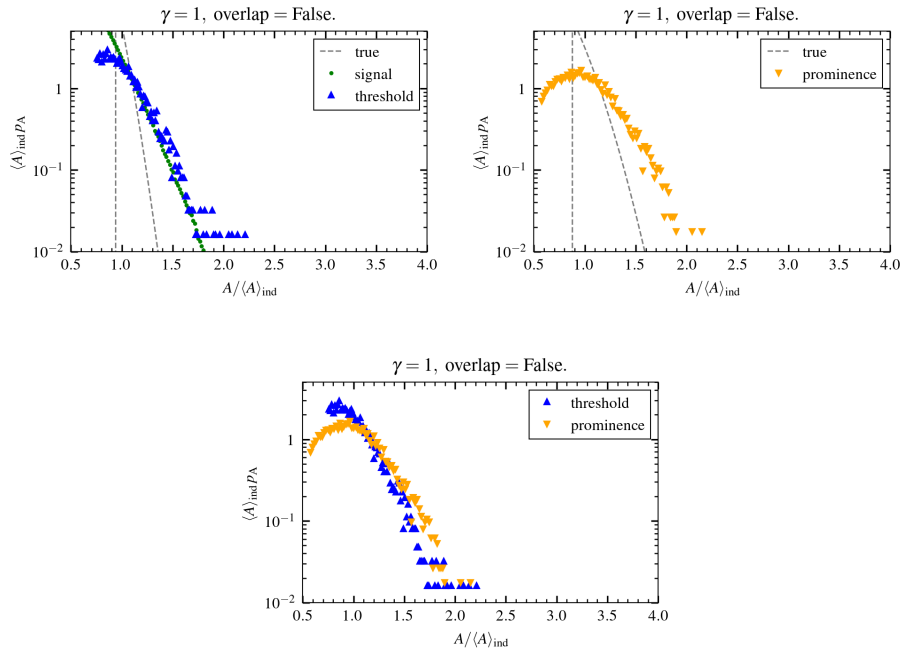


Figure 33: $\gamma = 1$ $A \sim \text{Rayleigh}\left(\sqrt{\frac{2}{\pi}}\right)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

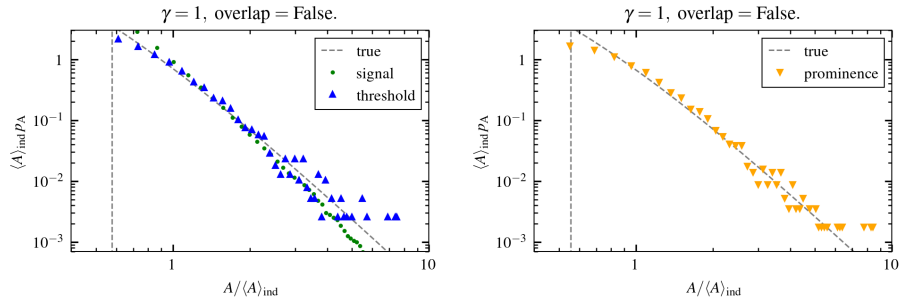


Figure 34: $\gamma = 1$ $A \sim \text{Lomax}(3, 2)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

In figure 34 we see the amplitude estimation results of conditional averaging from a signal with Lomax distributed amplitudes. Here we see the same characteristics exhibited in the $\gamma = 10^{-1}$ case in figure 30. Both the amplitude estimates follow the true distribution but also the signal distribution, making it hard to draw any new conclusions.

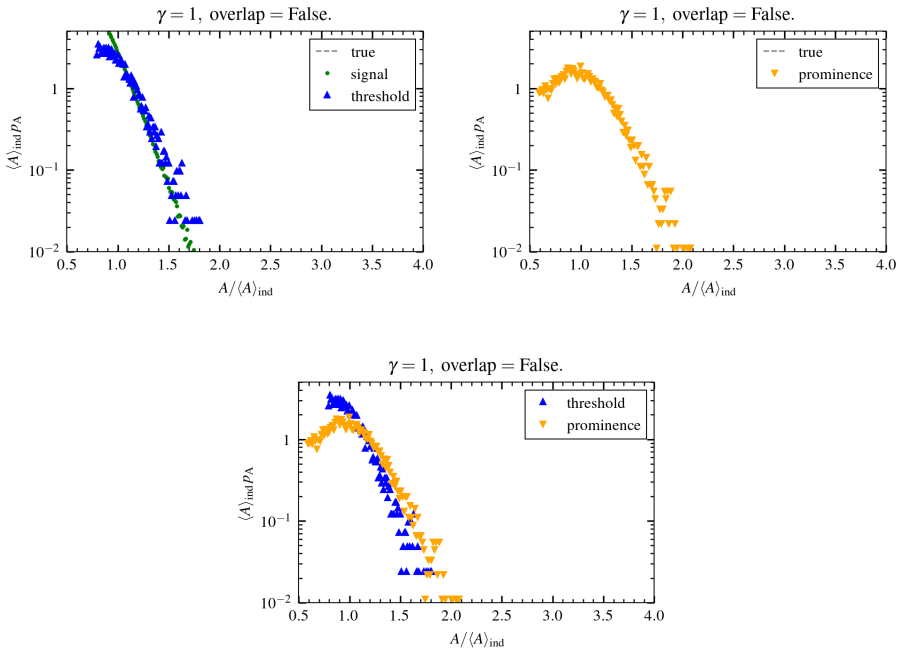


Figure 35: $\gamma = 1$ $A \sim$ Degenerate(1). Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

figure 35 shows the amplitude distribution estimates of conditional averaging used on a signal with degenerately distributed amplitudes. Examining the amplitude estimates we see that the amplitude threshold estimate, and the signal distribution is exponential. The reason we do not see the true conditional distribution here is that it is the simple fact that the threshold of $2.5\sqrt{\gamma}+\gamma = 3.5$ is too high, making no single event able to cross as all events has the same amplitude of 1. The shape of the prominence distribution is also similar to the one from figure 33 while not resembling any jumps that could indicate an underlying degenerate distribution.

As evident from the $\gamma = 1$ examination of the estimates from conditional averaging we see that the results already start to differ from the true distributions. In all cases but the exponential and Lomax distributed amplitudes we can confidently say that the amplitude threshold condition picks out random points on the signal, rather than uncovering underlying information about the distribution of individual events. The prominence condition also fails to predict the amplitude distribution in its own characteristic way, however as the signal does not have a prominence distribution to directly compare with it is hard to pinpoint the exact reason as to why it fails.

The pulse overlap dominated signals $\gamma = 10$

We will now study signals with an intermittency of $\gamma = 10$, where the effects of pulse overlap dominate. Almost all conditional events are due to a random build up due to small underlying events. Our main interest at this degree of pulse overlap is the heavy-tailed Lomax distribution. We seek to know if the finite probability of very large amplitude events will shine through in the amplitude distribution estimate, or if the effects of pulse overlap has completely taken over.

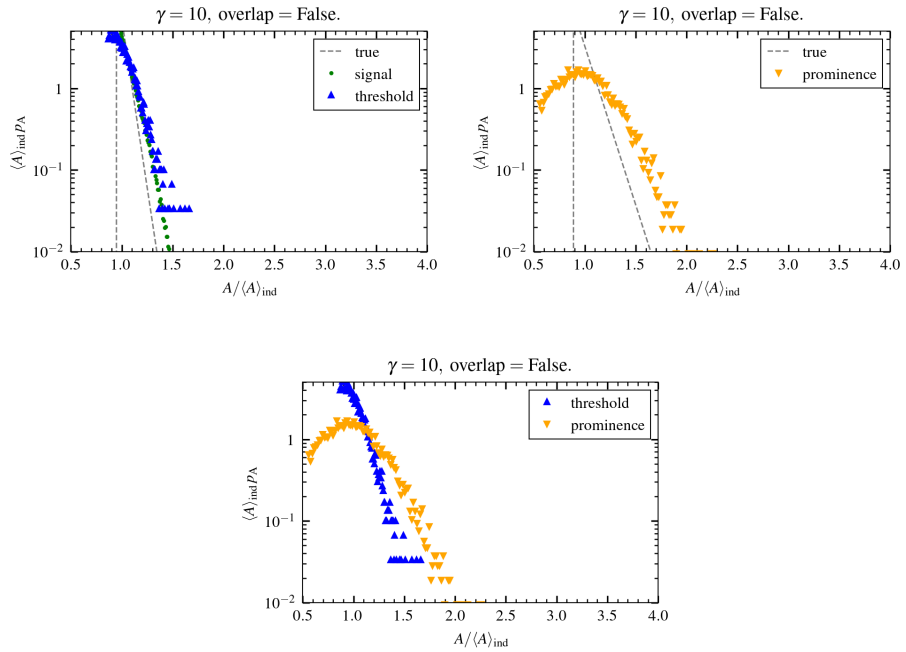


Figure 36: $\gamma = 10$ $A \sim \text{Exp}(1)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

In the case of exponentially distributed amplitudes in figure figure 36 we see that the amplitude distribution estimate from the prominence method has taken on the characteristic shape we saw earlier in some of the $\gamma = 1$ cases. The amplitude threshold estimate follows the signal distribution and looks exponential in the tail. The shift compared to the theoretical distribution stems from their individual normalizations. It is clear the the prominence method provides an inaccurate estimate for the amplitude distribution. From the general trend we have seen we are safer to assume that the reason the amplitude threshold condition predicts an exponential distribution is because of the tail of the signal distribution being exponential, rather than providing an accurate estimate of the underlying exponential amplitude distribution.

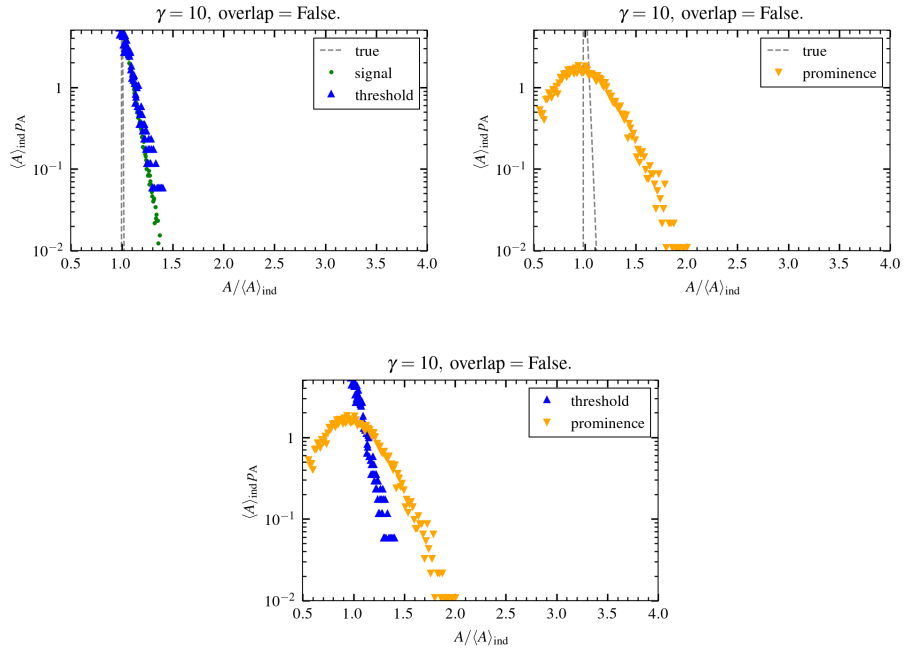


Figure 37: $\gamma = 10^{-1}$ $A \sim \text{Rayleigh}\left(\sqrt{\frac{2}{\pi}}\right)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

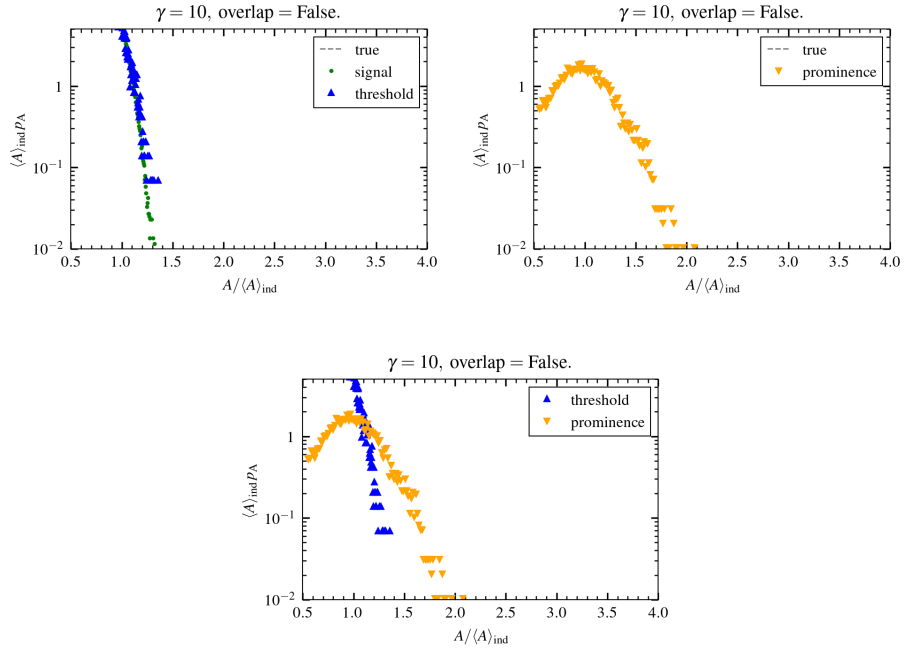


Figure 38: $\gamma = 10$ $A \sim \text{Degenerate}(1)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

In figure 37 and figure 38 we see the results of conditional averaging used on a signal with Rayleigh and degenerately distributed amplitudes respectively. The takeaway from these figures is much the same as in the $\gamma = 1$ case with both threshold conditions failing to predict the true distribution. The prominence estimate takes on the same shape and the amplitude threshold estimate closely follows the signal distribution rather than the true amplitude distribution.

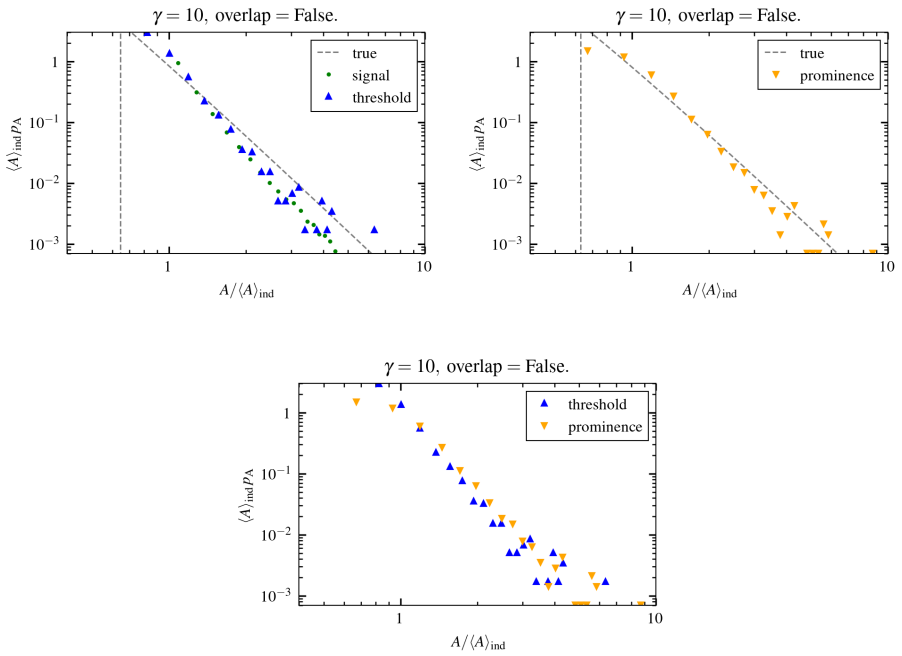


Figure 39: $\gamma = 10$ $A \sim \text{Lomax}(3, 2)$. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

The most interesting case is the results of conditional averaging used on a signal with Lomax distributed amplitudes as we see in figure 39. From the earlier $\gamma = 10^{-1}$ and $\gamma = 1$ cases concluded that the distribution matches both the signal and underlying distributions. From looking at the amplitude estimates here we do not see the characteristic shape of the prominence distribution that we have seen in the other $\gamma = 10$ cases. The tail of the amplitude estimate looks to follow the signal distribution more than the true polynomial decay of the true distribution, giving us slightly more confidence in assuming that the effects of pulse overlap cause the method to pick up random signal values. However, this is by far the best performing case compared to the lighter-tailed distributions mentioned earlier.

From the amplitude estimation results it is obvious that conditional averaging fails to reproduce the true amplitude distributions as γ increases. As the degree of pulse overlap increases the conditional events of both the amplitude and prominence thresholds become more and more due to overlap than due to individual events. This causes the events picked up by conditional averaging to appear as random points on the signal itself instead of predicting an underlying distribution of amplitudes for the individual events it tries to capture. The effects of pulse overlap is somewhat mitigated based on the distribution of amplitudes. For a distribution with a high probability of a small range of amplitudes the estimates from conditional averaging are less accurate than for a distribution where a wider range of amplitudes are more probable. This is based on the rough order that we saw the method fail in depending on amplitude distribution estimates. Degenerately distributed amplitude signals have the most inaccurate results, followed by the Rayleigh distributed cases, then the exponential and finally the Lomax distributed amplitudes which still showed decently accurate estimates at $\gamma = 10$. The main takeaway from this section is that conditional averaging already fails to predict amplitude distributions at degrees of intermittency on the order of $\gamma = 1$, showing that conclusions based on conditional averaging amplitude distribution estimates in this regime should be taken with a grain of salt. One should also be careful when using the method for more intermittent signals and at least check the amplitude distribution estimates against the signal distributions as a sanity check before concluding anything definitively.

What is also evident is that the conditionally averaged waveform is dependent on the input amplitude distribution. Heavier tails in the underlying distributions lead to more distinct events in the signals, making singular events responsible for more threshold crossings than in signals with lighter tails where most conditional events are due to pulse overlap. This might suggest that one could use this observation, in combination with another type of waveform estimate, to estimate the broadness of the underlying amplitude distribution.

4.3 Waiting time distribution analysis

Knowing the waiting times between individual events is a useful metric. Say you are building large bridge in an area susceptible to wind gusts. If you can estimate the distribution of waiting times between gusts you can design your bridge so that you avoid the bridge's resonant frequency being around the average frequency found from estimates based on an observed waiting time distribution. Or perhaps you are forecasting large avalanches in an area based on avalanche data from earlier seasons. Knowing how the times between avalanches are distributed would help predict at what times in that area is unsafe for skiing or hiking. Conditional averaging is one method that has been used to estimate waiting time distributions, and in this section we will examine its ability to do so. The filtered Poisson process is a process with exponentially distributed waiting times because of the underlying Poisson process determining where the individual events are located. This implies the locations are memoryless, making the probability of event locations independent of one another. As we want to examine other waiting time distributions than the exponential one we must move away from the underlying Poisson process for this section and instead work in the regime of a more general renewal process described in appendix C. This allows us to construct the arrival times based on other waiting time distributions. This is done by drawing random waiting times from a distribution and then cumulatively summing them up to create the arrival times. The intermittency parameter was maintained by deciding the mean of the distributions beforehand while keeping the unit duration times.

To generate the signals in this section the base case in section 3.1.1 was used with the modification to waiting times mentioned above. In this section we will examine the exponential, Rayleigh, Lomax and degenerate distributions for waiting times. The reason for this is much the same as in the amplitude analysis. The exponential distribution is the one used most in the literature, the Rayleigh distribution allows us to investigate an even steeper tailed distribution while also looking into if unimodality is recovered. The Lomax distribution lets us look into the effects of a heavy tailed distribution and the degenerate distribution allows us to see if conditional averaging can be used to pick up periodicity in the signal. All distributions are normalized by the true mean τ_w. As we keep τ_d constant this means that $\tau_w \sim \frac{1}{\gamma}$. In addition, we use the `overlap=False` condition as we have deemed it necessary in section 4.1.2. The true distributions shown are therefore conditional as `overlap=False` enforces a chosen distance between the events found, making any waiting time shorter than this distance impossible to procure.

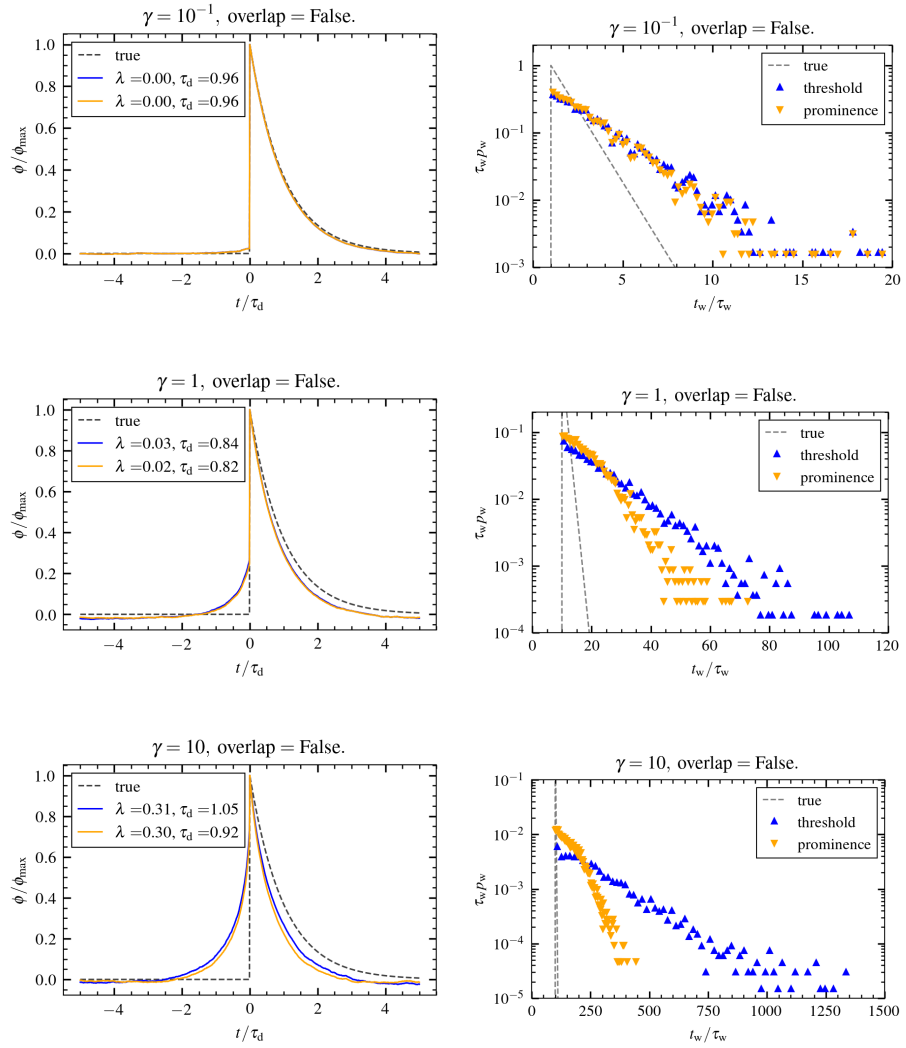


Figure 40: $T_w \sim \text{Exp}(\frac{1}{\gamma})$. Conditionally averaged waveform and waiting time distributions for both the pure threshold (blue) and prominence (orange) methods.

In figure 40 we see the waveform and waiting time estimates obtained from conditional averaging for the usual exponentially distributed waiting times. We see that the slope clearly does not match the slope of the true distribution, however this is expected as introducing a threshold would inevitably make it so that the method does not pick up all the events. This would result in us expecting to see an exponential distribution still, as the events we do pick up are still uniformly distributed if they are indeed the true events. Thus, we expect an exponential distribution but with a larger mean, which is just what we see here for $\gamma = 10^{-1}$. As γ increases this we see this trend continue for the amplitude threshold method, clear exponential tails which do not match the true distributions slope. The prominence method looks to also tail of exponentially, but with a certain break. However, observing an exponential waiting time distribution in the estimate does not necessarily indicate an underlying exponential distribution. As we concluded earlier from the amplitude distribution discussion we know that the method largely picks out random points on the signal because of overlapping structures even for $\gamma = 1$. Then analogue to this when discussing waiting times would be picking out random points on the time axis, and if these points are uniformly distributed the resulting waiting time distribution would then be exponential. Because of this it we cannot conclude either way from just looking at the base case at different degrees of intermittency.

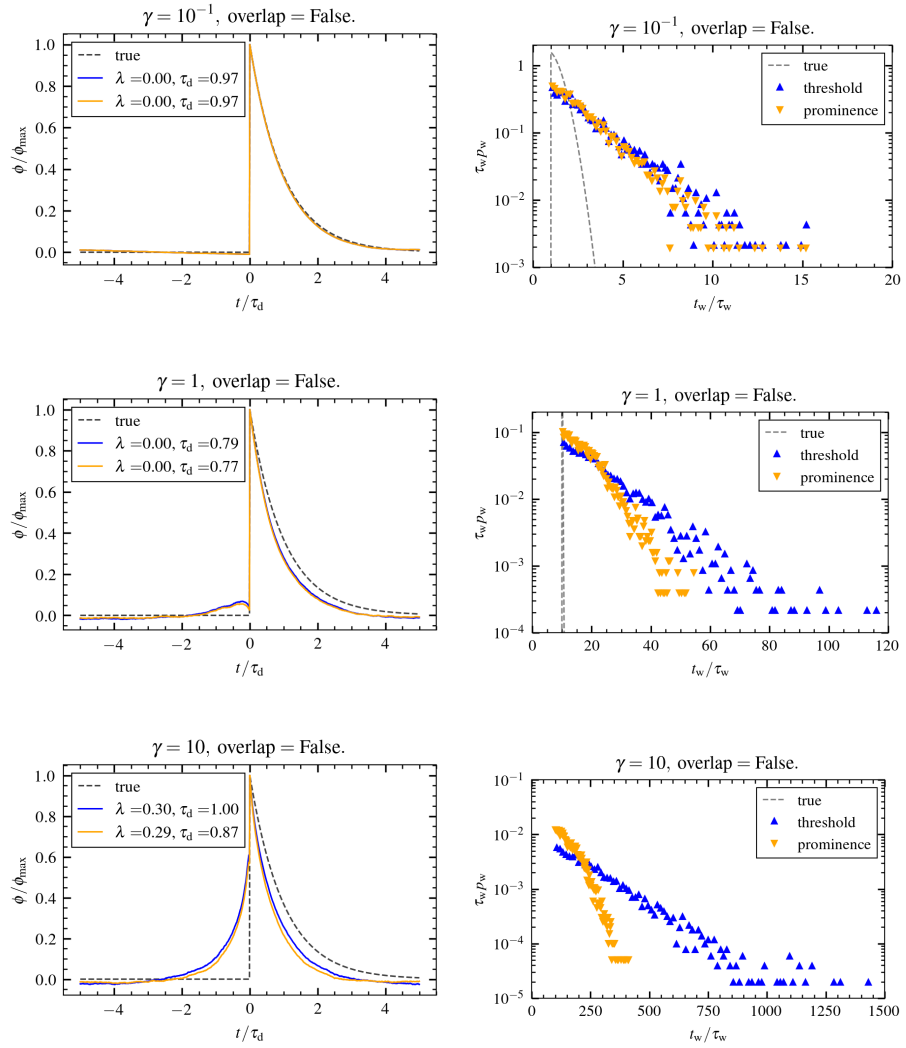


Figure 41: $T_w \sim \text{Rayleigh}\left(\sqrt{\frac{2}{\pi}}\frac{1}{\gamma}\right)$. Conditionally averaged waveform and waiting time distributions for both the pure threshold (blue) and prominence (orange) methods.

In figure 41 we see the waveform and waiting time estimates from conditional averaging used on a signal with Rayleigh distributed waiting times. At first look we see that the estimated waiting time distributions look much like the case with exponentially distributed waiting times, with there being clear exponential tails for both threshold conditions independent of γ. For $\gamma = 1$ and $\gamma = 10$ this can be explained by by the method picking up random points on the signal, resulting in exponentially distributed waiting times independent of the input distribution. However, for the $\gamma = 10^{-1}$ case we cannot jump to this same conclusion because of the inconclusive results obtained when discussing the amplitude distributions in this γ-regime. One way to explain the observed exponential waiting time distribution, even for highly intermittent signals is to look to the amplitude distribution. If we start with the amplitude threshold condition we know that the signal must cross a certain threshold for an event to be picked up. What determines this threshold crossing as high degrees of intermittency (low γ) can be assumed to be solely the amplitude distribution because of the very low degree of pulse overlap. As the amplitudes and waiting times are uncorrelated each pulse arrival has an independent probability of crossing this threshold with respect to its temporal position. Thus, the crossing positions will seemingly have exponential waiting times between in the tail of the distribution. This is not a very strong argument, nor does it hold for peaked distributions as we would still see more crossings with waiting times around this peak, simply because there would be more events with temporal spacing around this peak. However when we use the `overlap=False` condition the peak of unimodal distributions usually

become "windowed out" as the peak location lies within the window size, making the conditional distribution monotonically decreasing from the threshold point. Another interesting point is the $\gamma = 1$ waveform. Here we see a small bump before the instant rise of the pulse. The most probable explanation for this is that closely spaced events are highly unlikely for Rayleigh distributed waiting times because the probability of short waiting times is small. This makes it so that every event on average gets at least some time to decay before the next event, even with pulse overlap, giving us a small bump on the waveform. This is something we can also see in the other unimodal waiting time distributions in F.4.

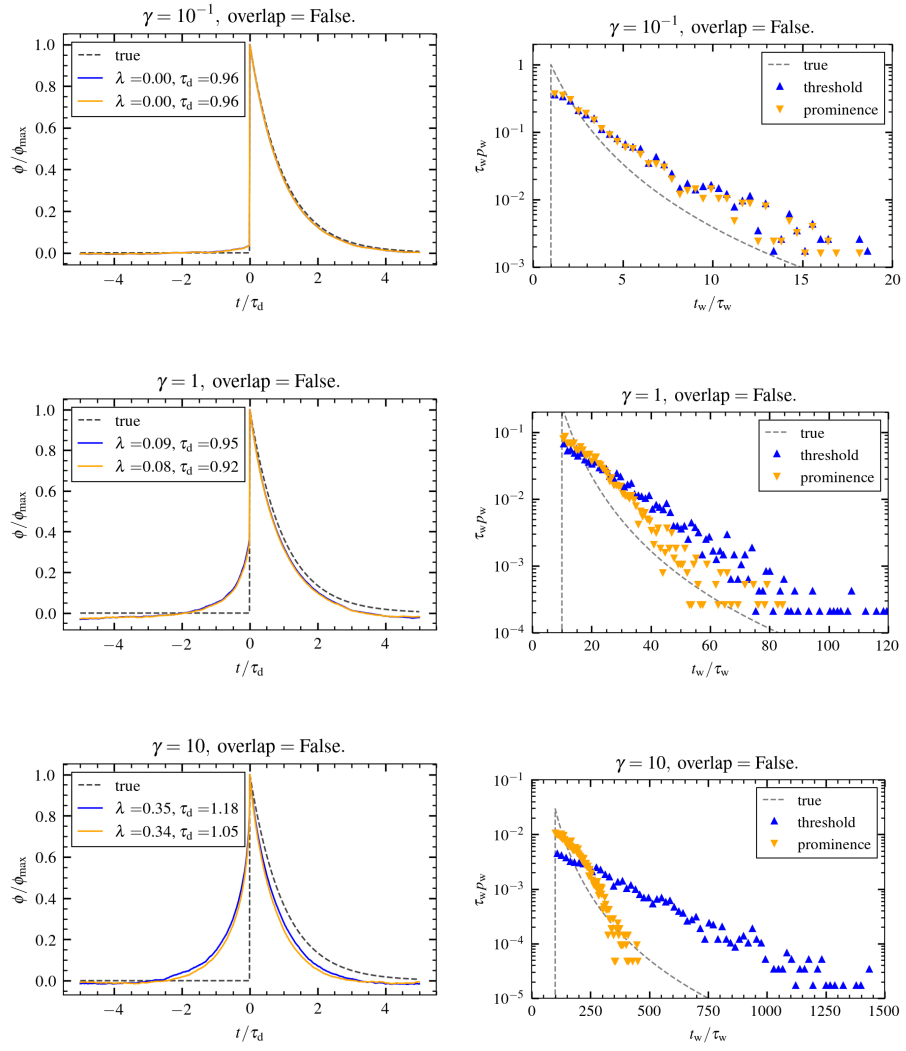


Figure 42: $\gamma = 10^{-1}$ $T_w \sim \text{Lomax}(3, \frac{2}{\gamma})$. Conditionally averaged waveform and waiting time distributions for both the pure threshold (blue) and prominence (orange) methods.

In figure figure 42 we see the waveform estimates and waiting time estimates from conditional averaging used on a signal with Lomax distributed waiting times. Again we observe much the same in the $\gamma = 1$ and $\gamma = 10$ cases with exponential tails for both methods, with a break in the prominence estimate. In the $\gamma = 10^{-1}$ case it is not entirely clear, it looks to start off exponential, but around $t_w/\tau_w = 10$ it becomes unclear. The waveform estimate is marginally better at $\gamma = 1$ than for exponentially distributed waiting times. The reason for this might be that longer waiting times are more probable, making the decay of events clearer when they occur. This might affect the average, but not to a very large degree.

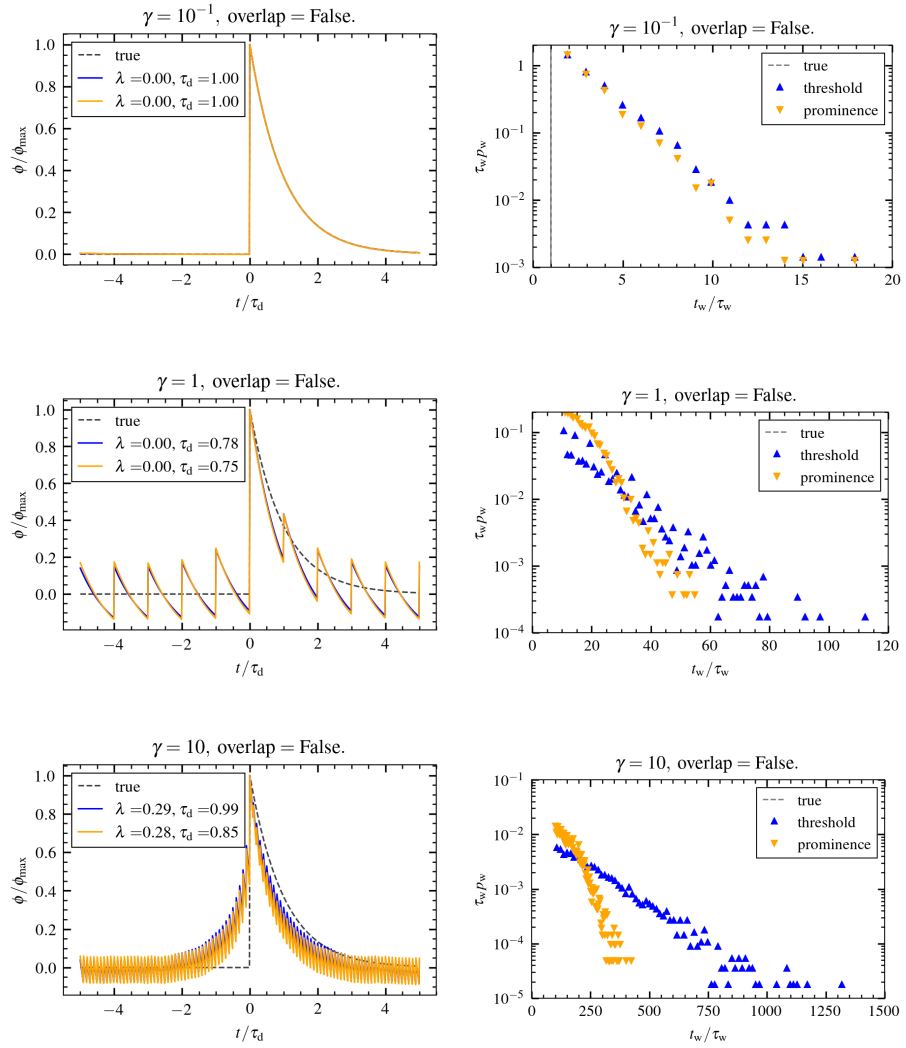


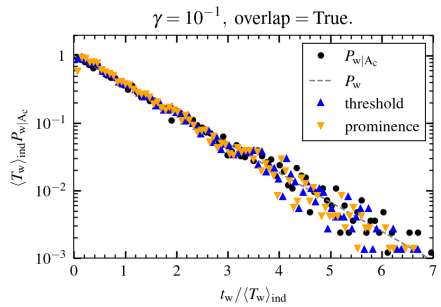
Figure 43: $\gamma = 10^{-1}$ $T_w \sim$ Degenerate($\frac{1}{\gamma}$). Conditionally averaged waveform and waiting time distributions for both the pure threshold (blue) and prominence (orange) methods.

In the case of the conditional average results from a periodic signal or degenerately distributed waiting times in figure 43 it is plain that the tails are exponential. In the case of small γ one could say that we would even expect the true distribution to be geometric. Every arrival is equally spaced from each other, making the observed waiting times multiples of the chosen waiting time. As the probability of crossing is determined solely by the amplitude distribution one could deem each crossing a success. Thus, each observed waiting time would represent a number of trials in units of the mean waiting time before a successful threshold crossing, which can be described by a geometric distribution. The waveform estimates show the effects of a periodic signal on the average pulse shape ass one can clearly see the individual pulses arriving within fixed times from each other.

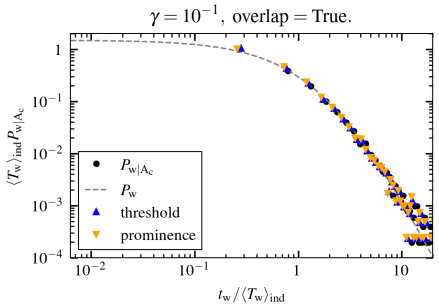
4.3.1 Effects of amplitude distributions in the highly intermittent case

As we now have seen, the waiting time distribution estimates generally do not match the true waiting time distributions, event at high intermittency. We can then ask ourselves if the method actually estimates the underlying waiting time distribution. As there is a size threshold involved what we really estimate is the underlying waiting time distribution conditioned on a certain amplitude threshold. To test this if the method estimates the underlying waiting time distribution or the amplitude conditioned waiting time distribution, we can generate highly intermittent signals and from the underlying amplitudes and waiting times we can generate this amplitude conditional

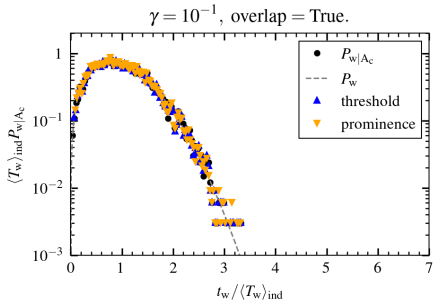
waiting time distribution by looking at the waiting times between the underlying events that cross the threshold. This can then be compared with the estimates from conditional averaging and the underlying waiting time distribution to see which is actually predicted by conditional averaging. The notation $\langle\cdot\rangle_{\text{ind}}$ again denotes the individual means of each respective distribution. Meaning the distributions are normalized by their respective means.



$T_w \sim \text{Exp}(\frac{1}{\gamma}),\ A \sim \text{Degenerate}(1)$



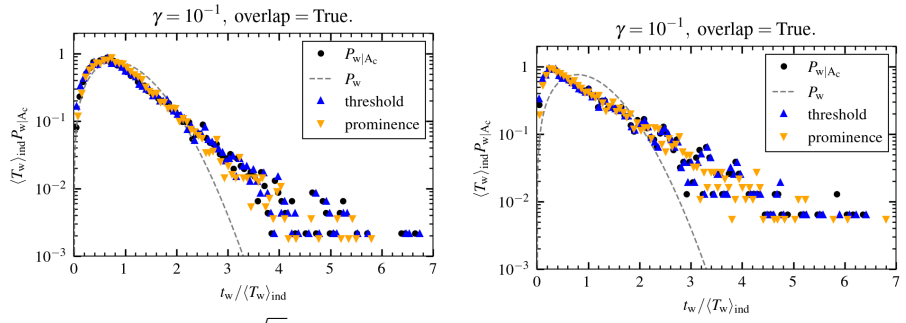
$T_w \sim \text{Lomax}(3, \frac{2}{\gamma}),\ A \sim \text{Degenerate}(1)$



$T_w \sim \text{Rayleigh}(\sqrt{\frac{2}{\pi}}\frac{1}{\gamma}),\ A \sim \text{Degenerate}(1)$

Figure 44: $\gamma = 10^{-1}$. Waiting time distribution estimates for both the pure threshold (blue) and prominence (orange) methods together with the waiting time distribution given the amplitude threshold (black) and the underlying waiting time distribution (grey).

In figure 44 we see the true amplitude conditional waiting time distribution, the estimates from conditional averaging, and the underlying waiting time distribution for different variations of underlying waiting time distributions. All the amplitude distributions are degenerately distributed, showing that all the distributions match. This is to be expected as when the amplitudes are degenerately distributed every event (except for a few cases of overlapping events) is picked up by conditional averaging. Every event also crosses the threshold, making the amplitude conditioned waiting time distribution equal to the unconditioned waiting time distribution.



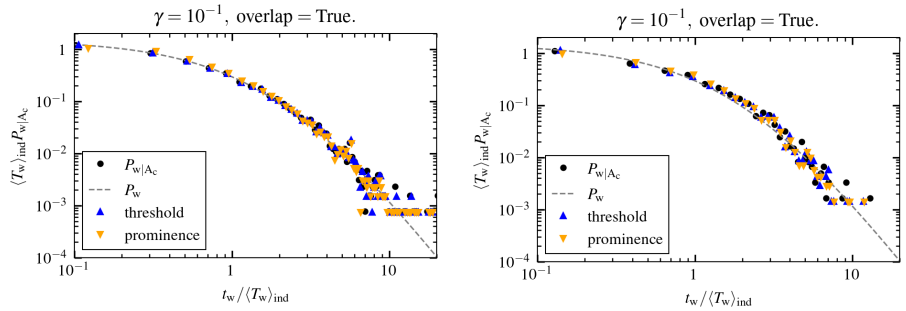
$T_w \sim \text{Rayleigh}(\sqrt{\frac{2}{\pi}\frac{1}{\gamma}}),\ A \sim \text{Rayleigh}(\sqrt{\frac{2}{\pi}})$

$T_w \sim \text{Rayleigh}(\sqrt{\frac{2}{\pi}\frac{1}{\gamma}}),\ A \sim \text{Lomax}(3, 2)$

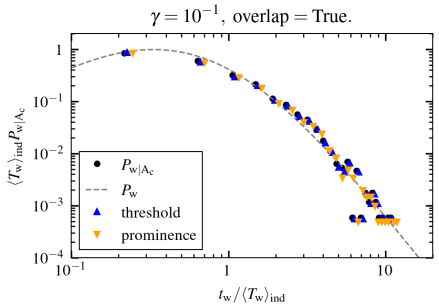
Figure 45: $\gamma = 10^{-1}$. Waiting time distribution estimates for both the pure threshold (blue) and prominence (orange) methods together with the waiting time distribution given the amplitude threshold (black) and the underlying waiting time distribution (grey).

In figure 45 we see what happens if the amplitudes are randomly distributed together with Rayleigh distributed waiting times for the conditional average estimates and the two waiting time distributions. What we see is that the tails become exponential. This can be explained analogous to how the degenerately distributed waiting times looked geometric in the distribution. The tail distributed events essentially get a random chance of crossing the threshold, and as the true tail is sharper than an exponential the differences between the events that cross the threshold become exponentially distributed because there are enough events in this range.

The counterexample to the previous figure we can see in figure 46 where the conditional average estimates and the true waiting time distributions, both the amplitude conditioned and the unconditioned one, again line up nicely. Here we have underlying heavy-tailed distributions such as the Lomax and beta prime distributions with different amplitude distributions. What we see is that the estimates generally match the true distributions. The reason that these distributions do not become exponential in their tails, despite the random chance of a threshold crossing, is because of the tails being heavier than that of an exponential. For example, say we have two events with a temporal spacing that corresponds to a tail event from an underlying Lomax distribution. If both events are picked up by conditional averaging we get the true distance which is fine. If only the first is picked up then the waiting time to the next pulse can only be longer than to that of the second event, thus conditional averaging estimates a heavy tail independent of amplitude distribution.



$T_w \sim \text{Lomax}(3, \frac{2}{\gamma}),\ A \sim \text{Rayleigh}(\sqrt{\frac{2}{\pi}})$ $T_w \sim \text{Lomax}(3, \frac{2}{\gamma}),\ A \sim \text{Lomax}(3, 2)$



$T_w \sim \beta'(3, 3, \frac{2}{3\gamma}),\ A \sim \text{Exp}(1)$

Figure 46: $\gamma = 10^{-1}$. Waiting time distribution estimates for both the pure threshold (blue) and prominence (orange) methods together with the waiting time distribution given the amplitude threshold (black) and the underlying waiting time distribution (grey).

Table 1: Table of the mean values from the different estimates and distributions shown in figures 44, 45 and 46. The two leftmost columns showing which combination of distributions and the four rightmost columns showing the means of their respective distribution or estimates.

Wait dist	Amp dist	$\langle$Amp$\rangle$	$\langle$Prom$\rangle$	$\langle P_w \rangle$	$\langle P_{w\|A_c} \rangle$
Exp	Degenerate	10.7	10.9	10.0	10.1
Lomax	Degenerate	10.7	11.1	10.0	9.9
Rayleigh	Degenerate	10.0	10.0	10.0	10.0
Rayleigh	Rayleigh	14.4	13.3	10.0	14.6
Rauleigh	Lomax	38.5	35.3	10.0	39.1
Lomax	Rayleigh	15.28	15.23	10.0	15.7
Lomax	Lomax	38.7	38.4	10.0	42.3
β'	Exp	23.2	21.8	10.0	24.0

We see in table 1 that the mean values for the estimates correspond with both the waiting time distribution and the amplitude conditioned waiting time distribution when the amplitudes are degenerately distributed. This is in agreement with what we saw earlier in figure 44. However, as soon as the amplitudes become distributed we see the means abandon the mean of the waiting time distribution with the conditional average estimates more closely matching the amplitude conditioned waiting time distribution, even in the cases where we saw a nice functional match between all distributions such as in figure 46. This supports the waiting time estimates from conditional averaging being an estimate of the amplitude conditioned waiting time distribution rather than the unconditioned waiting time distribution.

From this we can conclude that conditional averaging is really only able to estimate the true waiting time distribution if the signal is highly intermittent and the true amplitude distribution is degenerate as this is the case when the amplitude conditioned and unconditioned waiting time distributions coincide. This severely limits the use cases where one can draw reliable conclusions from the waiting time estimates.

4.3.2 Conclusion on the waiting time distribution estimates

Overall what we can see that the waiting time distribution estimates are not a good estimate for the true waiting time distribution for any degree of intermittency except for highly intermittent waiting time distributions with degenerately distributed amplitudes. This stems from the threshold condition in the conditional average and threshold crossings being determined by the amplitude distribution in addition to pulse overlap. The waveform estimate is also affected by the waiting time distribution, but in a different way than the amplitude distribution. Unimodal waiting time distributions such as the Rayleigh or gamma distributions have differently shaped distortions in the averaged waveform, as the probability of closely spaced events are so small, creating a small bump-like distortion rather than the slower rise created by monotonically decreasing distributions such as the Lomax and exponential distributions. In the case of low γ the exponential look of the tails can be somewhat explained by the independent amplitude distribution making threshold crossings look more uniformly distributed than their arrivals. In the case where pulse overlap becomes more prevalent the events picked up can be seen as random points on the signal, the time difference between these random points on the signal then become exponential. From this it is clear that conditional averaging should not be used to estimate waiting times directly, as involving a size threshold on the signal inherently involves the underlying amplitude distribution of the signal as well.

4.3.3 Tail rate estimation

As we saw looking at waiting times the tails of the distributions were exponential, one interesting piece of information is then looking at how the slope of this tail changes with increasing γ. In this section we will explore this slope for various γ-values to see if it can be used to give an estimate of the mean waiting time of the underlying signal. This is has been done in the literature [**?**, 32], making it relevant for us to see if the slope of this tail represents valuable information. The mean waiting time can be used to estimate γ based on a given τ_{d}, giving us a plot of γ_{est} against γ. For low gamma we expect waiting times between threshold crossings to be distributed according to equation 3.25 for exponentially distributed waiting times. As $\gamma \to \infty$ the FPP approaches an Ornstein-Uhlenbeck process [33]. To estimate the upper limit we use conditional averaging on a realization of this process, then estimating the exponential slope of the resulting waiting time distribution estimate.

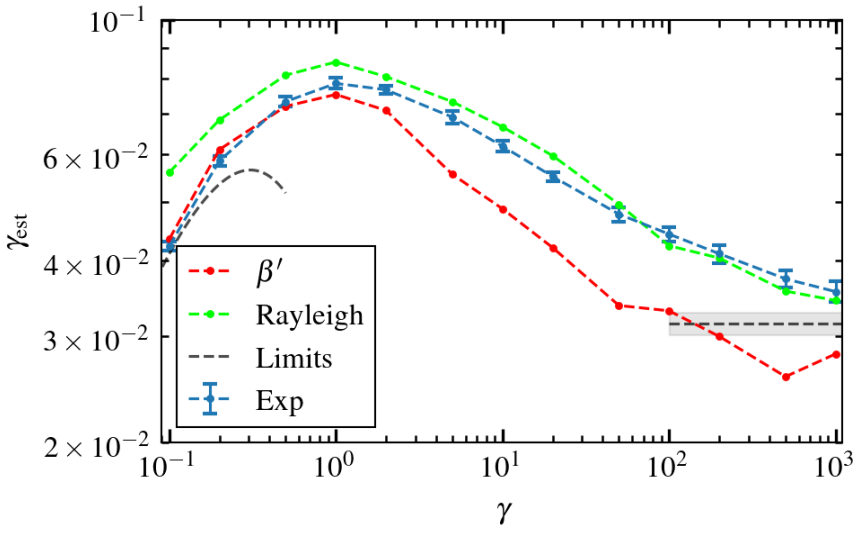


Figure 47: Decay rate estimation of the tail of waiting time distributions for different degrees of intermittency and different waiting time distributions. Each point on the blue graph is an average of estimates from ten different realizations with $T_{\text{w}} \sim \text{Exp}(\frac{1}{\gamma})$. The green graph shows tail estimates with $T_{\text{w}} \sim \text{Rayleigh}\left(\sqrt{\frac{2}{\pi}}\frac{1}{\gamma}\right)$. The red graph shows tail rate estimates with $T_{\text{w}} \sim \beta'(3, 3, \frac{2}{3\gamma})$. The left limit is the analytical limit with no pulse overlap. The right limit is from using decay rate estimation on an Ornstein-Uhlenbeck process which is the limit of the FPP as $\gamma \to \infty$ [33].

What we see in figure 47 is how the exponential tail of different signals varies with increasing γ. The first and most obvious observation is that the rates do not vary much with the variation being between 10^{-2} and 10^{-1}, not even spanning a whole decade while γ spanning over four decades. The filtered Poisson process approaches the Ornstein-Uhlenbeck (OU) process, however the rate of approach is very slow. The renewal processes with Rayleigh and beta-prime distributed waiting times have the same general shape as the FPP. It looks as if beta-prime distributed waiting times do not approach the limit of an OU-process, however as this is only one sample per point, we cannot conclude that this is not because of statistical fluctuations. All of the maxima lay at $\gamma = 1$, meaning it is in this range conditional averaging picks up the largest number of events. The main takeaway from this is that the number of events picked up by conditional averaging cannot be used as an accurate predictor of the number of underlying events.

5 The use of conditional averaging in other works

Within the fusion plasma community the filtered Poisson process has been widely used to model turbulent fluctuations of magnetically confined plasmas in the outer region of toroidal fusion devices, this region is often referred to as the scrape off layer (SOL) [24, 27, 31, 32, 34]. Conveniently for us, conditional averaging is also a widely used method to retrieve the shape, amplitude and waiting time distributions within in the same field [18, 30], letting us use estimates of intermittency to review the validity of the results obtained from conditional averaging. In this chapter we will go over 3 papers where conditional averaging has been used to estimate the waveform, the amplitude distribution and the waiting time distribution.

The first one is [32] where they looked at plasma fluctuations in the SOL of the Korea Superconducting Tokamak Advanced Research (KSTAR) device. Here the γ-values were estimated to lay between 1.7 and 2.4 with a noise to signal fluctuation ratio of $\varepsilon = 0.11$. From this we already know that the amplitude and waiting time results may not be trusted entirely.

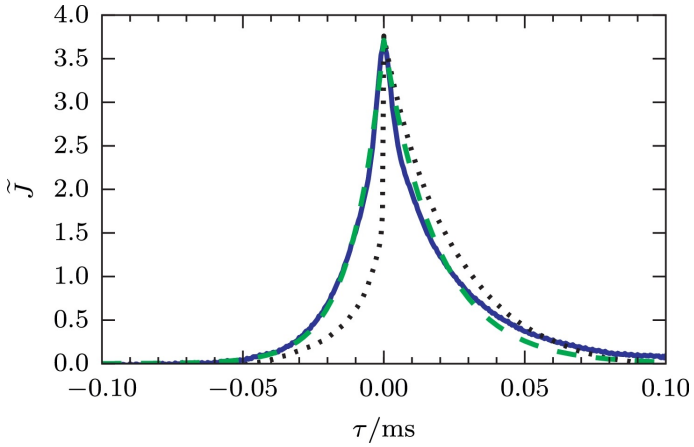


Figure 48: Conditionally averaged wave-forms with peak amplitudes larger than 2.5 times the rms value for the ion saturation current (full blue line), the synthetic data (dotted black line) and the best fit of a double-exponential pulse shape to the measurement data (dashed green line). Figure and figure description from [32] fig 11.

In figure 48 we see their results of conditional averaging on the normalized ion saturation current. In this figure they have used conditional averaging with an amplitude threshold and the `overlap=False` condition. What we quickly see is that this is the method that does not normalize each event, resulting in an averaged waveform that exceeds 1. This choice creates an inherent bias for large amplitude events, meaning if the pulse shape changes depending on pulse height then the shape of the larger pulses will be more apparent in the average. Without having compared them we can hypothesize that this bias somewhat mitigates the effects of pulse overlap as if one biases larger events then the shape of those events will be more distinct than smaller events due to overlap. To model this signal they used a one-sided exponential pulse with a duration time of 30μs, based on an estimate from the auto correlation function (ACF) of the signal, to generate a realization of the FPP with exponentially distributed amplitudes and waiting times. The conditional average of this realization we also see in figure 48. Looking at the mismatch between the conditional averages of the ion saturation current and the FPP-realization it is clear that a one-sided pulse used in the synthethic data is perhaps not the best suggestion. In the article the stated rise time is 11μs and the fall time is 19μs based on the double exponential fit to the waveform. This gives an estimated asymmetry of $\lambda = 0.37$. The duration time estimate of 30μs agrees with the estimate from the ACF. We expect the duration time estimates from conditional averaging to be relatively accurate as γ is not too high and the underlying pulse is most likely not severely asymmetric. Knowing that pulse overlap at this degree of intermittency increases symmetry in the waveform leads us to conclude that a double exponential pulse shape with asymmetry below 0.37 would be a better choice for the underlying waveform. However, without having thoroughly investigated the amplitude biased method we cannot conclude that this is for certain.

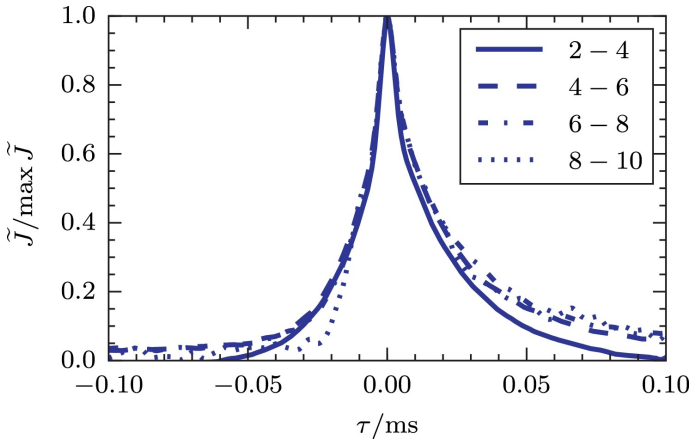


Figure 49: Conditionally averaged burst wave-forms for the ion saturation current signal with peak amplitudes in units of the rms value given by the range indicated in the legend. Figure and figure description from [32] fig 12.

Figure 49 describes the conditional averaging done between different threshold brackets, much as we did in section 4.1.3. The method is again assumed to be amplitude biased with the waveforms being normalized after the conditional averaging is done. Here the author concludes that the waveform does not change with increased amplitudes, which looks to be a reasonable conclusion. In context with using the amplitude biased method it also helps to show that a possible amplitude bias in the averaging will not skew the shape of the waveform toward higher amplitudes. This further strengthens the earlier waveform discussion because we know we are not working with an amplitude biased average. As we know, on this scale of γ we would expect to be able to reveal a general correlation trend between the waveform shape and amplitudes, but not the specific correlation. This makes this figure a valid use of conditional averaging as it is used to attempt to reveal a general correlation trend.

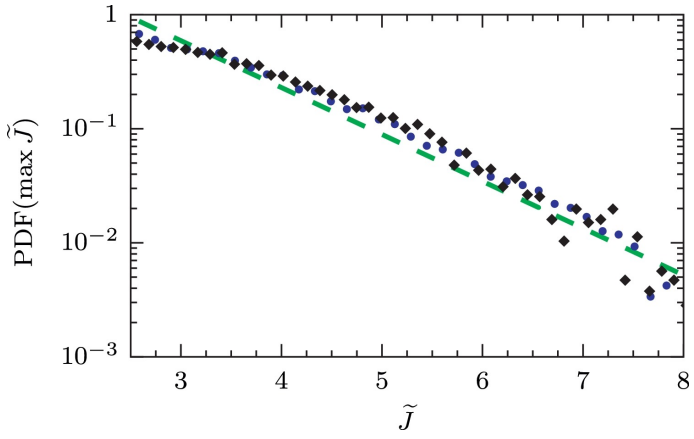


Figure 50: Probability distribution function for burst amplitudes with peak values larger than 2.5 times the rms level for the ion saturation current (blue circles), the synthetic data (black diamonds), and an exponential fit to the measurement data (dashed green line). Figure and figure description from [32] fig 13.

Figure 50 describes the amplitude distribution estimate of the fluctuations, an estimate from a realization of an FPP and an exponential fit to the fluctuation estimate. As $\gamma > 1$ in this case we know that at this point most underlying non heavy-tailed distributions will appear exponential like the tail of the signal distribution itself. Earlier in the paper the signal distribution is shown with a clear exponential tail, which from our earlier conclusions would result in an exponential estimate from the conditional average as well. The author mentions this expected exponential tail from the probability density of the signal. However we cannot rule out other underlying

amplitude distributions based on what we see here as other input amplitude distributions also result in exponential tails in the signal at this level of intermittency.

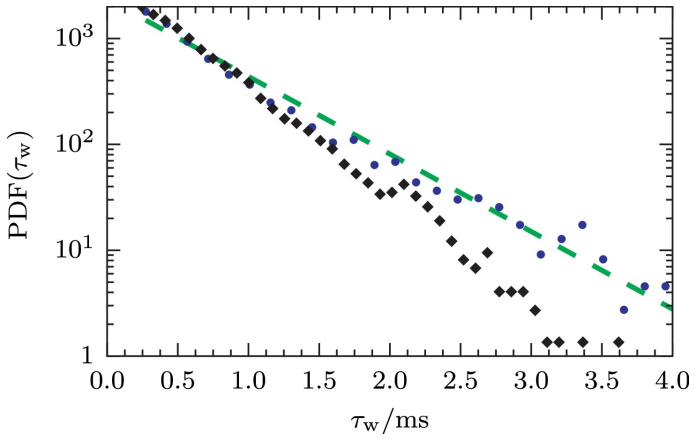


Figure 51: Probability distribution function for waiting times between large-amplitude events with peak values larger than 2.5 times the rms level for the ion saturation current (blue circles), the synthetic data (black diamonds) and an exponential fit to the measurement data (dashed green line). Figure and figure description from [32] fig 14.

In figure 51 we see the waiting time distribution obtained from using conditional averaging on the plasma fluctuation signal and a realization of the FPP, together with an exponential fit. The mean waiting time above threshold is also estimated to be 0.8μs. We know that the mean waiting time estimated from this tail is not a good measure for the true mean, as this tail rate will vary little across many decades of γ like we saw in figure 47. As $\gamma > 1$ we can safely say that any waiting time distribution here would look exponential in the conditional average, as the effects of overlap are quite a large factor at this intermittency. And if the amplitudes are distributed randomly, and not degenerate, we know that conditional averaging would not be able to reveal the true waiting time distribution even at high intermittency. This figure illustrates a use of conditional averaging within the regime where we have shown the results from the method to be unreliable. We have shown that the waiting time distribution estimate should not be used to conclude anything about the underlying waiting time distribution in almost all cases.

The second one is [13] where they looked at plasma fluctuations in the SOL of the Tokamak à configuration variable (TCV) device. The author does not directly mention it but the intermittency can be estimated by $\frac{\langle J\rangle^2}{J_{\text{rms}}^2}$ [13]. As $\frac{J_{\text{rms}}}{\langle J\rangle} \approx 0.7$ from the paper we can estimate that the intermittency of the fluctuation signal being $\gamma \approx 2$. From knowing this intermittency, we expect only the waveform estimates to be somewhat robust based on our earlier conclusions.

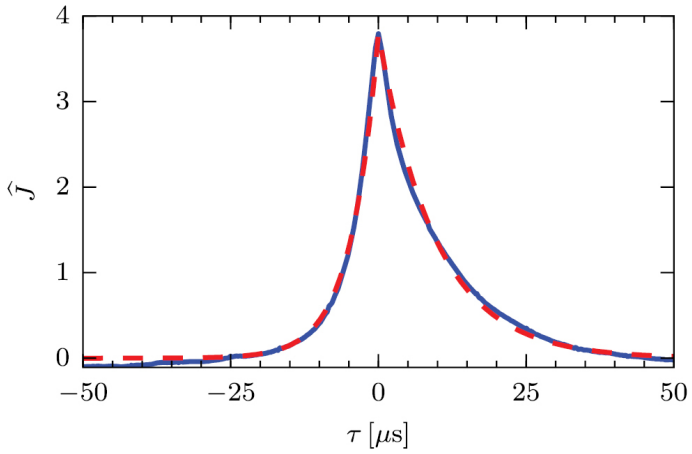


Figure 52: Conditionally averaged wave-form for the ion saturation current with peak amplitudes larger than 2.5 times the rms value (full line) together with a fitted double-exponential pulse shape (broken line). Figure and figure description from [13] fig 10.

In figure 52 we see the waveform of conditional averaging used on the fluctuating signal, together with a double-exponential fit. Initially we see that the amplitude biased method is used, and the `overlap=False` condition is also in use. With a mentioned rise time of 5μs and an estimated duration time of 15μs the estimated asymmetry is $\lambda = \frac{1}{3}$. The authors do not make any asymmetry conclusions about the underlying pulse shape, but they mention that the general shape of large amplitude bursts is a double exponential which is a perfectly valid conclusion to make in light of the findings of this thesis as we expect conditional averaging to recover the functional shape of underlying pulses at this intermittency.

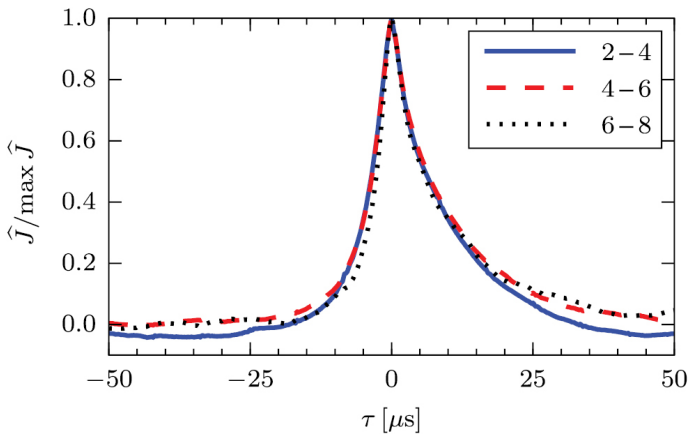


Figure 53: Conditionally averaged burst wave-form for the ion saturation current signal with peak amplitudes in units of the rms value given by the range indicated in the legend. Figure and figure description from [13] fig 12.

Figure 53 describes the results of conditional averaging done in different threshold brackets to unveil possible correlations between asymmetry and the size of underlying events. What the authors observe is that there appears to be no such correlation, which is a valid conclusion as we would expect to pick out a general trend in such a correlation at this intermittency. This is also an important test to do when using the amplitude biased method to check if the earlier waveform can be trusted.

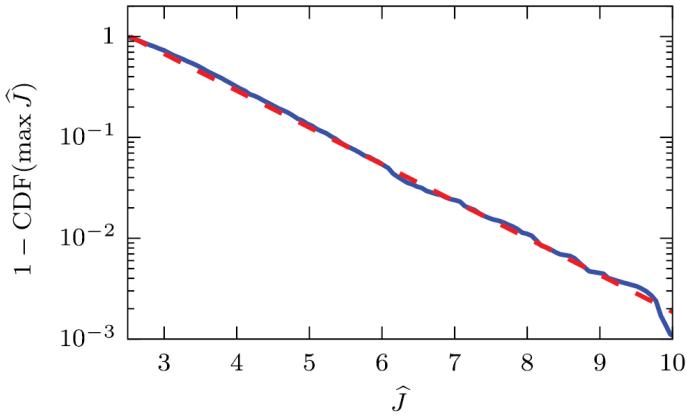


Figure 54: Complementary cumulative distribution function for ion saturation current burst amplitudes with peak values larger than 2.5 times the rms level (full line). The broken line shows the fit of a truncated exponential distribution. Figure and figure description from [13] fig 15.

The amplitude complimentary cumulative distribution (CCDF) function estimate from conditional averaging used on the fluctuating signal is shown in figure 54 together win an exponential fit. An exponential CCDF indicates an exponential PDF, meaning here the amplitudes are as well estimated to be distributed exponentially. However, as $\gamma = 2$ and the tail of the signal PDF is also exponential we would expect to find that the conditional averaging picks acts as to pick out peaks which appear as random points on the signal in this case. Other amplitude distributions also lead to exponential tails in the signal distribution at higher degrees of pulse overlap which means that one cannot draw the conclusion of the underlying pulses being exponentially distributed at this intermittency just from the conditional average estimate.

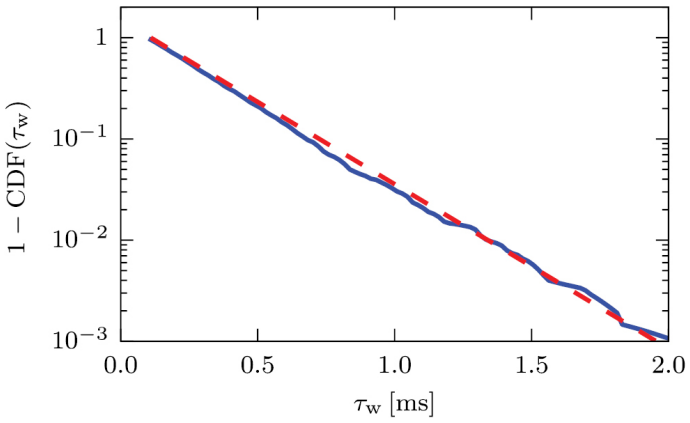


Figure 55: Complementary cumulative distribution function for waiting times between large-amplitude events in the ion saturation current signal with peak values larger than 2.5 times the rms level (full line). The broken line shows the fit of a truncated exponential distribution. Figure and figure description from [13] fig 17.

In figure 55 we see the waiting time CCDf estimate from conditional averaging together with an exponential fit. Again, we know that an exponential CCDF implies an exponential PDF. The authors use this exponential estimate to support their hypothesis that the underlying events follow a Poisson process. Most waiting time distributions, and thus other renewal processes, lead to exponential waiting time estimates from conditional averaging at intermittency levels of $\gamma > 1$. This makes this an example of conditional averaging used at intermittency where we have demonstrated the method to provide misleading results, as we cannot rule out other underlying renewal processes based on an exponential estimate.

The third one is [30] where they looked at plasma fluctuations in the SOL of the Alcator C-mod device. In this paper they looked at different signals from different runs where each run was under

different conditions to compare them and see if the FPP could describe them all. The intermittency parameter has been estimated from the probability density functions. For the ohmic low density signal (red) $\gamma = \frac{3}{4}$. The ohmic high density signal has an estimated intermittency of $\gamma = 2$. For the ELM-free H signal (blue) $\gamma = 3$ and the EDA H-mode (black) signal has $\gamma = 5$. From here on they will be referred to by their colours. We expect the most reliable results from the red signal as $\gamma < 1$ and the most unreliable results from the black signal as $\gamma = 5$.

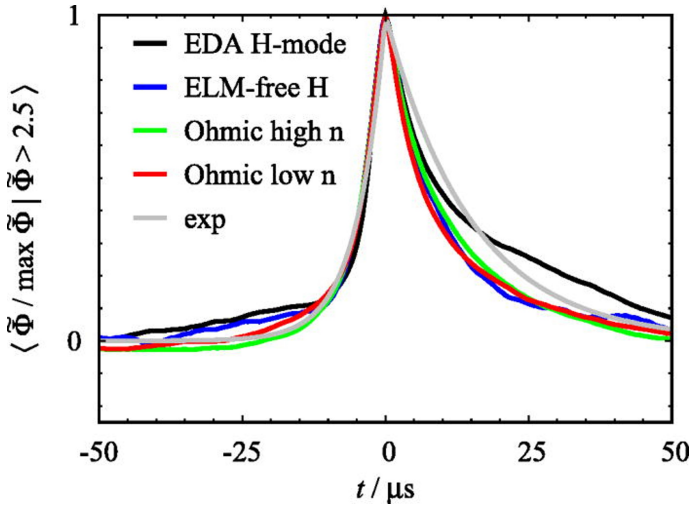


Figure 56: Conditionally averaged waveform for large-amplitude events in the GPI intensity signals measured at (R, Z)=(90.69, –2.99) cm for various plasma parameters and confinement modes. Also shown is a two-sided exponential pulse with a rise time of 5 μs and a fall time of 15 μs (grey line). Figure and figure description from [30] fig 8.

In figure 56 we see the conditionally averaged waveforms of all the different signals, together with a double exponential pulse. Here the authors used the `overlap=False` condition and the amplitude biased method. One important mention is that the method found hundreds of events for each signal, making the average more affected by outlier events than what we have worked with. The black sheep of the waveforms is fittingly the black signal with a significantly longer fall time than the other signals. As the waveforms are largely the same the authors conclude that the the underlying pulse shape of all the signals are also equal. In unsymmetrical pulses we know that pulse overlap increases symmetry in the conditionally averaged waveform. This info would lead us to conclude that the underlying pulses get less symmetric with increasing γ if the conditionally average waveform remain largely the same independent of increasing γ. However, as the authors used the amplitude biased method we cannot say this for certain, as this bias may lead to a smaller degree in loss of asymmetry depending on γ compared to the amplitude unbiased method. Conditional averaging is able to pick out different functional pulse shapes at this intermittency, making the conclusions of the authors valid. The waveforms clearly resemble a double exponential pulse independent of intermittency.

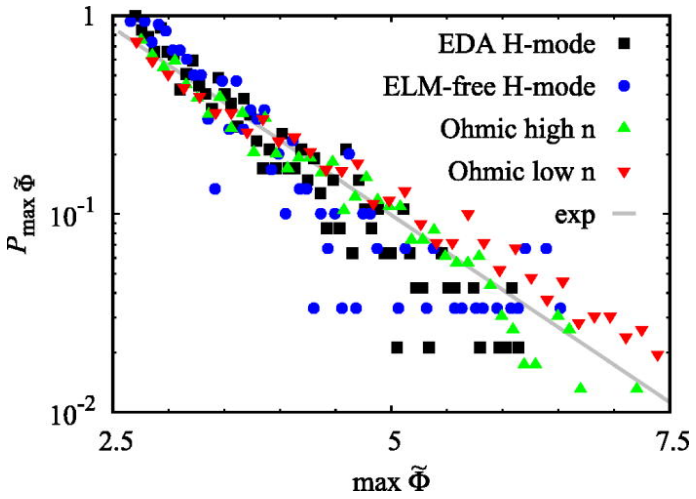


Figure 57: PDF of peak amplitudes in the GPI intensity signals above 2.5 standard deviations measured at (R, Z)=(90.69, –2.99) cm for various plasma parameters and confinement modes. The full line shows a truncated exponential distribution. Figure and figure description from [30] fig 9.

Figure 57 describes the estimated amplitude distribution of the underlying pulses for all the signals. As the author comments all follow an exponential distribution. However the PDFs of all the signals themselves also featured exponential tails, not letting us rule out other amplitude distributions at intermittency values of one and above. We can only really rule out a heavy tailed amplitude distribution for all the signals based on the presented evidence. Overlap affects the amplitude distribution estimate in such a way that what we see here is more likely a sample of peak values that look like random points on the signal itself.

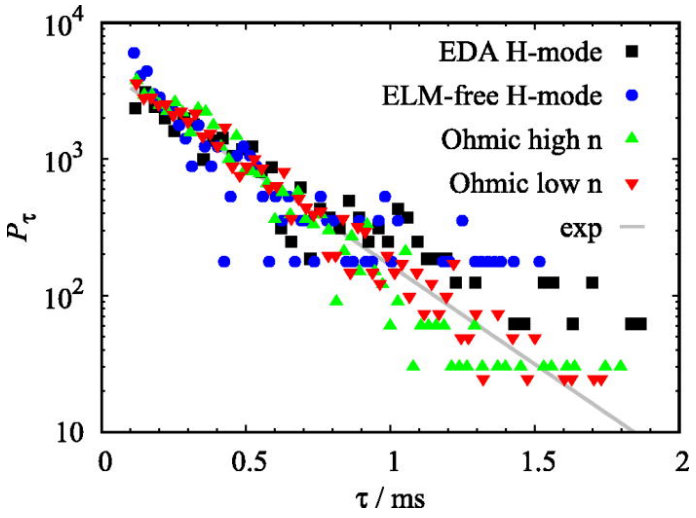


Figure 58: PDF of waiting times between large-amplitude events with peak amplitudes above 2.5 standard deviations measured at (R, Z)=(90.69, –2.99) cm for various plasma parameters and confinement modes. The full line shows a truncated exponential distribution. Figure and figure description from [30] fig 10.

The waiting time distribution estimates in figure 58 shows a clear exponential tail for all the signals. From this the author concludes that the events are uncorrelated as an underlying exponential waiting time distribution would mean an underlying Poisson process. However, based on the degree of intermittency and what we have learned, this is a case where one cannot trust the distribution estimate. Most underlying renewal processes would result in exponentially distributed waiting time estimates from conditional averaging. Both sharper decaying Gaussian tails and slower decaying polynomial decays resulted in exponential estimates from our testing in section 4.3 making waiting time distribution estimates a bad way of using conditional averaging.

What we see is that the estimates from conditional averaging are often used in intermittency regimes where those results are not actually valid. The waveform estimate is the most robust one, being especially well suited to reproduce the functional form of symmetric smooth pulse shapes. Authors rarely keep in mind that an inherent property of conditional averaging is that asymmetry in the underlying pulse shape is gradually lost with increasing pulse overlap. Amplitude estimates are often taken for granted to be exponential at higher γ-values. What we would predict at these intermittency values is that the estimate resembles the tail of the signal distribution rather than the true distribution. The cases where authors most often use conditional averaging in what we have concluded to be the the most misleading way are conclusions made from the waiting time distribution estimates. The most common conclusion that is often made is that an exponential waiting time distribution estimate from conditional averaging is a sign of an underlying Poisson process. We have shown that the waiting time distribution estimate should rarely be interpreted as a pointer toward the underlying waiting times. All distributions studied in this thesis resulted in exponential estimates from conditional averaging when $\gamma \geq 1$. The waiting time estimate from conditional averaging is also amplitude dependent as the amplitudes are one important parameter that decides if an event is picked up or not, the other being pulse overlap. For amplitude and waiting time distribution estimates one should instead look to other methods, such as the deconvolution method mentioned in [35].

6 Conclusions and outlook

Using the filtered Poisson process as a test bed for conditional averaging we have obtained valuable information about the regimes of overlap the method provides accurate, and not so accurate predictions about the underlying process. Throughout this thesis we examined two different threshold conditions, an amplitude threshold which is the standard [13, 28, 30] and a prominence threshold, to attempt to establish if one is better than the other for all the different estimates. At first, we looked into how different pulse shapes were affected by increasing intermittency. Then, moving on with the one-sided exponential pulse, we investigated the effects two different types of noise on the waveform, establishing the importance of the additional `overlap=False` condition. In the following section we examined a simple correlation between the asymmetry parameter λ from a two-sided exponential pulse shape and the amplitude distribution, exploring if conditional averaging can be used to uncover such a correlation at different levels of intermittency. We then proceeded with changing the underlying amplitude distribution and examining both the effects on the waveform and the accuracy of the estimated conditional amplitude distribution compared to the true conditional distribution. Finally, we moved away from the underlying Poisson process, working with a more general renewal process to generate signals with different waiting time distributions to attempt to establish if the estimated distribution from conditional averaging can be used as a predictor for the underlying one, and also looking at how different waiting time distributions affect the averaged waveform itself. Finally we looked into a general trend of exponential tails from the waiting time distribution analysis, investigating how the slope of this tail evolves with increasing γ.

Conditional averaging has been used to estimate all these properties [18, 27, 28, 36, 37] but without a good understanding of if the method is actually valid when pulse overlap is allowed. The method has previously been investigated in terms of noise while not allowing the underlying pulses to overlap [4, 20, 21], where the results have been robust, which is why we have sought out to mainly investigate the effects of pulse overlap on the results we obtain from conditional averaging.

After looking into different waveforms we established that conditional averaging reproduces symmetric pulse shapes to a much better degree than asymmetric pulses. We observed that as inherent pulse overlap in the signal increased, the averaged waveform converged toward a symmetric one. Increasing overlap also affected the symmetric pulses, but to a less serious degree in the form of an overestimation in the average duration time τ_d as the shape of most symmetric pulses could still clearly be distinguished from one another at $\gamma = 10$. We also found that the `overlap=False` condition is necessary for the prominence threshold to eliminate distinct troughs in the averaged waveform, stemming from closely spaced conditional events. After looking at both additive and dynamic noise we established that the `overlap=False` condition is also vital for the amplitude threshold condition if one wants accurate estimates, even for low degrees of noise-to-signal fluctuation ratios ε on highly intermittent signals. The results from correlating the asymmetry of the waveform with the size of the event lead to the conclusion that we can indeed estimate an average waveform between different threshold brackets at high intermittency, however accuracy is quickly lost with increasing pulse overlap, as even at $\gamma = 1$ we were only able to discern the general trend of decreasing asymmetry with increasing amplitudes, but not with accurate estimates of the true correlation. At $\gamma = 10$ the effects of overlap dominate, making even such a distinction of the correlation trend indiscernible.

From looking at different amplitude distributions in section 4.2 we observed an interesting correlation between the broadness of the distribution and the waveform estimates. If a signal has a relatively large probability of high amplitude events compared to low amplitude events, then more threshold crossings will be because of singular events than due to the effects of overlap, making the conditional events represent underlying events to a larger degree than in signals where there is a large probability of a small number of events. The effects of overlap for the one-sided exponential pulse shape is a small initial rise before the instantaneous rise, leading to a perceived loss in asymmetry. With heavy tailed distributions this effect is mitigated by the relatively large probability of large amplitude events and with degenerately distributed amplitudes the effects of overlap is most prevalent, leading to the quickest loss in asymmetry. Thus, the main conclusion from looking at the amplitude distributions effect on the waveform is that the broadness of the amplitude distribution mitigates the effects of pulse overlap. From looking at the amplitude distribution estimates we also observed that both methods failed at predicting the underlying distribution at $\gamma = 1$ and above. For $\gamma = 0.1$ the amplitude distribution estimates reproduced both the

signal distribution and the underlying amplitude distributions in all cases but the degenerately distributed amplitudes. The methods failed in different ways. The estimate from the prominence method converged toward the same unimodal shape for almost all input distributions. The estimate from the amplitude threshold converged toward the conditional signal distribution, allowing us to make the conclusion that the amplitude threshold leads to picking out peaks that appear as random points on the signal when the effects of overlap increase. We again observed that the distribution that allowed for the most accurate amplitude distribution estimates was the heavy-tailed Lomax distribution, showing again that broader distributions mitigate the effects of overlap. The main takeaway from this section being that conditional averaging is not generally a good way to predict the underlying amplitude distribution for anything but highly intermittent signals.

What we saw in section 4.3 is that the waiting time distribution affects both the waveform and the waiting time distribution estimate. The waveform is affected mostly by the shape of distribution in the form that purely tailed distribution creates different distortions on the waveform than unimodal distributions. The distortion in the waveform for purely tailed waiting time distributions is a small initial rise which leads to an increase in symmetry. For unimodal distributions the distortion is in the form of a small bump which stems from the low probability of very small waiting times, lowering the signal "mass" just before each pulse arrival. The waiting time distribution estimates quickly become exponential independent of the underlying distribution. In the highly intermittent case this is because of the amplitude distribution also affects the estimated waiting time distribution, as one would expect from a size-threshold condition in the method. The only situation the high intermittency estimate represents the true underlying waiting time distribution is when the amplitudes are degenerately distributed, otherwise the tail will appear exponential for most input waiting time distributions. In the cases where pulse overlap is more pronounced one can draw the analogy to the conclusion from the amplitude section of random signal points being picked out, leading to essentially uniformly distributed random arrival times which signifies an exponential waiting time distribution. This leads to the conclusion that the waiting time distribution estimates from conditional averaging should not be used unless one has a highly intermittent signal with equally sized events. From the tail rate estimation, the main takeaway is that the number of conditional events cannot accurately be used to predict the underlying amount of singular events.

After discussing how conditional averaging is used when looking at experimental data in section 5 we found that conclusions made from the waveform estimates of conditional averaging are generally valid within the intermittency regimes they have been used in. However, that is not the case for amplitude and waiting time distribution estimates. Authors interpret amplitude estimates as information on the amplitude distribution of the underlying events in γ-regimes where we know the amplitude estimate will follow the tail of the signal distribution rather than the underlying distribution. This is based on our findings effectively equivalent of concluding that the amplitudes of underlying events are distributed according to the tail of the probability density function of the signal itself. In terms of waiting time estimates authors often conclude that the underlying process is a Poisson process based on seeing an exponential estimate. We have learned that this may be misleading, as all underlying renewal processes we looked at led to exponential estimates from conditional averaging at $\gamma = 1$ and above, because of pulse overlap. It is only for a few highly intermittent niche cases that the waiting time estimates from conditional averaging coincide with the underlying distribution.

To further this work there are multiple avenues one could explore. One could combine more parameters to conduct a more comprehensive study, looking at how the conditional average estimates are affected by other combinations of underlying pulse shapes, amplitude distributions and waiting time distributions to either strengthen the conclusions in this thesis or make new ones based on observations arising from new combinations. Another road would be to explore distributed duration times, where one could investigate if duration time estimates from the conditionally averaged waveform coincide with the mean duration time of the underlying pulses. One could also explore more correlations within the signal, correlating pulse shapes with waiting times could reveal an underlying bias for large waiting times within the method. Another interesting correlation would be waiting times and amplitudes to see how that affects the results. For example, one would expect that if one makes a correlation between amplitudes and waiting times in a way that makes large waiting times equal large amplitudes then conditional averaging would yield better waveform estimates, as such a correlation would lead to more distinct events in the signal. Or the reverse, large waiting times leading to small amplitudes, which one could imagine would lead to worse waveform

estimates because of more prevalent pulse overlap around the large events that are picked up by the method. One could also inquire into the different normalizations, examining the effects of the amplitude/prominence biased methods to uncover possible aspects that make them preferable over the normalization methods. The threshold conditions can also be investigated together, to attempt to see if using both threshold conditions at the same time alleviate the effects of noise, making the `overlap=False` condition superfluous. Another way of proceeding would be to examine the if the conclusions of this thesis can be used to develop new methods. Can one use the waveform estimate to somehow determine the broadness of the underlying amplitude distribution? Can the waveform estimate be used to detect different waiting time distributions? These are all interesting lanes one could take to further explore conditional averaging.

A Probability distributions and their standardized moments

In this appendix we present the probability distributions encountered in this thesis where most are used as both amplitude and waiting time distributions when generating test signals.

A.1 The normal distribution

The location and scale parameters are μ and σ. The distribution is denoted by $\mathcal{N}(\mu, \sigma)$.

The probability density function is given by [38]

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{x-\mu}{\sigma}\right)^2\right) \tag{A.1}$$

The cumulative distribution function is given by

$$P(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right) \tag{A.2}$$

Mean	Variance	Skewness	Excess kurtosis
μ	σ	0	0

A.2 The Poisson distribution

The rate parameter is λ. The distribution is denoted by Poisson(λ).

The probability mass function is given by [38]

$$p(n) = \exp(\lambda)\frac{(\lambda)^n}{n!}, \quad n = 0, 1, 2... \tag{A.3}$$

The cumulative distribution function is given by

$$P(x) = \exp(-\lambda)\sum_{j=0}^{\lfloor n \rfloor}\frac{\lambda^j}{j!} \tag{A.4}$$

Mean	Variance	Skewness	Excess kurtosis
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$

A.3 The uniform distribution

The support parameters are a and b. The distribution is denoted by $\mathcal{U}(a, b)$. The probability density function is given by [38]

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{elsewhere} \end{cases} \tag{A.5}$$

The cumulative distribution function is given by

$$P(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 0, & \text{elsewhere} \end{cases} \tag{A.6}$$

Mean	Variance	Skewness	Excess kurtosis
$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	0	$-\frac{6}{5}$

A.4 The exponential distribution

The scale parameter of the distribution is β. The distribution is denoted by Exp(β)..

The probability density function is given by [38]

$$p(x;\beta) = \begin{cases} \frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.7}$$

The cumulative distribution function is given by

$$P(x;\beta) = \begin{cases} 1 - \exp\left(-\frac{x}{\beta}\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.8}$$

Mean	Variance	Skewness	Excess kurtosis
β	β^2	2	6

A.5 The gamma distribution

The shape and scale parameters of the distribution are α and β. The distribution is denoted by $\Gamma(\alpha, \beta)$.

The probability density function is given by [38]

$$p(x;\alpha,\beta) = \begin{cases} \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}}\exp\left(-\frac{x}{\beta}\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.9}$$

The cumulative distribution function is given by

$$P(x;\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)}\gamma(\alpha, \frac{x}{\beta}), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.10}$$

Mean	Variance	Skewness	Excess kurtosis
$\alpha\beta$	$\alpha\beta^2$	$\frac{2}{\sqrt{\alpha}}$	$\frac{6}{\alpha}$

A.6 The Rayleigh distribution

The scale parameter of the distribution is σ. The distribution is denoted by Rayleigh(σ).

The probability density function is given by [39]

$$p(x;\alpha,\beta) = \begin{cases} \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.11}$$

The cumulative distribution function is given by

$$P(x;\alpha,\beta) = \begin{cases} 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.12}$$

Mean	Variance	Skewness	Excess kurtosis
$\sigma\sqrt{\frac{\pi}{2}}$	$\frac{4-\pi}{2}\sigma^2$	$\frac{2\sqrt{\pi}(\pi-3)}{(4-\pi)^{3/2}}$	$-\frac{6\pi^2-24\pi+16}{(4-\pi)^2}$

A.7 The Lomax distribution

The shape and scale parameters of the distribution are α and β. The distribution is denoted by Lomax(α, β). This is also known as the Pareto type II distribution.

The probability density function is given by [40]

$$p(x;\alpha,\beta) = \begin{cases} \frac{\alpha}{\beta}\left(1+\frac{x}{\beta}\right)^{-\alpha-1}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.13}$$

The cumulative distribution function is given by

$$P(x;\alpha,\beta) = \begin{cases} 1-\left(1+\frac{x}{\beta}\right)^{-\alpha}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.14}$$

Mean	Variance	Skewness	Excess kurtosis
$\begin{cases} \frac{\beta}{\alpha-1}, \alpha > 1 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{\beta^2\alpha}{(\alpha-1)^2(\alpha-2)}, \alpha > 2 \\ \infty, 1 < \alpha \leq 2 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{2(1+\alpha)}{\alpha-3}\sqrt{\frac{\alpha-2}{\alpha}}, \alpha > 3 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{6(\alpha^3+\alpha^2-6\alpha-2)}{\alpha(\alpha-3)(\alpha-4)}, \alpha > 4 \\ \text{Undefined elsewhere} \end{cases}$

A.8 The degenerate distribution

The parameter of the distribution is k. The distribution is denoted by Degenerate(k).

The probability density function is given by

$$p(x;k) = \delta(x-k) \tag{A.15}$$

where δ is the Dirac delta function.

The cumulative distribution function is given by

$$P(x;k) = \begin{cases} 1, & x \geq k \\ 0, & x < k \end{cases} \tag{A.16}$$

Mean	Variance	Skewness	Excess kurtosis
k	0	Undefined	Undefined

A.9 The beta prime distribution

The shape and scale parameters of the distribution are α, β and q. The distribution is denoted by $\beta'(\alpha, \beta, q)$.

The probability density function is given by [41]

$$p(x;\alpha,\beta,q) = \begin{cases} \frac{(\frac{x}{q})^{\alpha-1}(1+\frac{x}{q})^{-(\alpha+\beta)}}{qB(\alpha,\beta)}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.17}$$

The cumulative distribution function is given by

$$P(x;\alpha,\beta) = \begin{cases} I_{\frac{x}{q+x}}(\alpha,\beta), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.18}$$

Mean	Variance	Skewness
$\begin{cases} \frac{q\alpha}{\beta-1}, \beta > 1 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{q^2\alpha(\alpha+\beta+1)}{(\beta-2)(\beta-1)^2}, \beta > 2 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{2(2\alpha+\beta-1)}{\beta-3}\sqrt{\frac{\beta-2}{\alpha(\alpha+\beta-1)}}, \beta > 3 \\ \text{Undefined elsewhere} \end{cases}$

A.10 The inverse gamma distribution

The shape and scale parameters of the distribution are α and β. The distribution is denoted by $\Gamma^{-1}(\alpha, \beta, q)$.

The probability density function is given by [42]

$$p(x;\alpha,\beta,q) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)}(\frac{1}{x}^{\alpha+1})\exp\left(-\frac{\beta}{x}\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.19}$$

The cumulative distribution function is given by

$$P(x;\alpha,\beta) = \begin{cases} \frac{\Gamma(\alpha,\frac{\beta}{x})}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.20}$$

Mean	Variance	Skewness	Excess kurtosis
$\begin{cases} \frac{\beta}{\alpha-1}, \alpha > 1 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{4\sqrt{\alpha-2}}{\alpha-3}, \alpha > 3 \\ \text{Undefined elsewhere} \end{cases}$	$\begin{cases} \frac{6(5\alpha-11)}{(\alpha-3)(\alpha-4)}, \alpha > 4 \\ \text{Undefined elsewhere} \end{cases}$

B Transformation and normalization

In this appendix we go into the different normalizations used in this thesis and the transformation rules that apply when performing them.

It is often useful to normalize data before working with. This makes the analysis more universal as you can more easily compare results between data sets. The normalization we are going to use the most is

$$\tilde{X} = \frac{X - \langle X \rangle}{X_{\text{rms}}} \tag{B.1}$$

To work within the regime of this normalization we need to make sure we correctly transform the random variable. If X is a continuous random variable with probability density function $p_X(x)$ on the interval $A \subseteq \mathbb{R}$ and the transformed random variable is $Y = g(X)$ where g is a strictly increasing or decreasing function, the probability density function of Y becomes [38]

$$p_Y(y) = p_X(g^{-1}(y))|\frac{\mathrm{d}}{\mathrm{d}y}\left(g^{-1}(y)\right)| \tag{B.2}$$

defined on the interval $B = \{y = g(x) : x \in A\}$. Thus a linear transformation $Y = aX + b$ yields the transformed probability density function [38]

$$p_Y(y) = \frac{1}{|a|}p_X\left(\frac{y-b}{a}\right). \tag{B.3}$$

Making the probability density function of our normalized random variable [38]

$$p_{\tilde{X}}(\tilde{x}) = X_{\text{rms}}p_X(X_{\text{rms}}\tilde{x} + \langle X \rangle). \tag{B.4}$$

As this thesis is primarily concerned with the conditional average it is also useful to define the conditional probability distribution so we can compare with analytical expressions. The usual condition used here will be a simple threshold value. If we again let X be our random continuous variable defined on the interval $[a, b]$ where $a < c < b$, with PDF $p_X(x)$ then the conditional PDF given that the condition is $X \geq c$ is

$$p_{X_c}(x_c) = p_X(x|x \geq c) = \frac{p_X(x_c)}{1 - P_X(c)} \tag{B.5}$$

defined on the interval $[c, b]$. F is the cumulative distribution function and S is the survival function. If the threshold c is placed above the upper limit of support for the random variable then the probability of finding values above the threshold is obviously zero. The mean above the threshold is then defined as

$$\langle X_c \rangle = \int_c^b p_{X_c}(x_c)\mathrm{d}x_c = \frac{1}{1 - P_X(c)}\int_c^b p_X(x)\mathrm{d}x \tag{B.6}$$

C The renewal process

In this appendix we present the general renewal process. This is relevant as we are working with different waiting time distributions within this thesis.

A renewal process can be defined in the following way [43]. Let the waiting times between events w_k be independent identically distributed (IID) with a finite positive expectation value, and the arrivals times t_k of the events be defined as

$$t_k = w_1 + w_2 + ... + w_{k-1} + w_k, \quad k \geq 1. \tag{C.1}$$

Then

$$N(T) = \max\{k : t_k \leq T\}, \quad T \geq 0 \tag{C.2}$$

is the renewal process which tells us the number of arrivals up until time T.

C.1 The Poisson process

The Poisson process $\{N(t), t \geq 0\}$ is a renewal process and can be defined in the following way [26].

- $N(0) = 0$.
- The process has stationary and independent increments.
- $\mathbb{P}(N(t) = n) = \exp(\lambda t)\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2...$

Where λ is the rate parameter in the Poisson distribution. The Poisson process and the exponential distribution are closely related. Let w_1 be the time until the first event of a Poisson process and w_n be the time between the $(n-1)$th and the nth event, then $W \sim \text{Exp}(\frac{1}{\lambda})$ [26]. Another important relation referred to in this thesis is the relation between the Poisson process and the uniform distribution. Let $t_n \geq 0$ be the arrival times of events from a Poisson process, then the arrival times within an interval $[0, T)$ are uniformly distributed according to $\mathcal{U}(0, T)$ [26]. To summarize the relation between the Poisson process and the distributions can be stated as follows. On a finite temporal interval $[0, T)$

- The number of events are distributed according to $N \sim \text{Poisson}(\lambda T)$.
- The arrivals are distributed according to $t_n \sim \mathcal{U}(0, T)$.
- The time between arrivals are distributed according to $W \sim \text{Exp}(\frac{1}{\lambda})$.

D Pulse functions

All the pulse functions are defined in a way so that the first integer moment $I_1 = 1$. Thus we can seek inspiration in unimodal probability density functions when define different pulse functions as the integral over all possible values for any valid probability density function is 1.

D.1 The one-sided exponential pulse

The one-sided exponential pulse function is given by

$$\phi(x) = \begin{cases} \exp(-x), & x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{D.1}$$

D.2 The two sided exponential pulse

The double exponential pulse function is given by

$$\phi(x; \lambda) = \begin{cases} \exp\left(\frac{x}{\lambda}\right), & x \leq 0 \\ \exp\left(-\frac{x}{1-\lambda}\right), & x > 0 \end{cases} \tag{D.2}$$

Where λ is the asymmetry parameter and can take values in the interval (0, 1).

D.3 The Gaussian pulse

The Gaussian pulse function is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{D.3}$$

D.4 The triangle pulse

The triangle pulse function is given by

$$\phi(x) = \begin{cases} 1 - |x|, & |x| < 1 \\ 0, & \text{otherwise} \end{cases} \tag{D.4}$$

D.5 The Lorentz pulse

The Lorentz pulse function is given by

$$\phi(x) = \frac{1}{\pi\,(1 + x^2)} \tag{D.5}$$

D.6 The box pulse

The box pulse function is given by

$$\phi(x) = \begin{cases} 1, & |x| < 1 \\ 0, & \text{otherwise} \end{cases} \tag{D.6}$$

D.7 The gamma pulse

The gamma pulse function is given by

$$\phi(x;\alpha) = \begin{cases} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x), & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{D.7}$$

where α is a shape parameter, just as in the PDF of the gamma distribution and Γ is the gamma function.

D.8 The Rayleigh pulse

The Rayleigh pulse function is given by

$$\phi(x;\alpha) = \begin{cases} x \exp\left(-\frac{x^2}{2}\right), & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{D.8}$$

E Special functions

In this appendix we present the non-elementary functions used to define some of the probability distributions earlier.

- The gamma function is defined as

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-x) \mathrm{d}t. \tag{E.1}$$

- The beta function is defined as

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} \mathrm{d}t \tag{E.2}$$

- The error function is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{(\pi)}} \int_0^x \exp\left(-t^2\right) \mathrm{d}t \tag{E.3}$$

F Additional figures

In this appendix we present additional figures that was used to support the results in section 4.

F.1 Additional waveforms

In this appendix more figures used in the waveform analysis in section 4.1 is presented. The reason why they are included here, and not in the discussion, is because of the traits of conditional averaging that they demonstrate were already demonstrated by other figures.

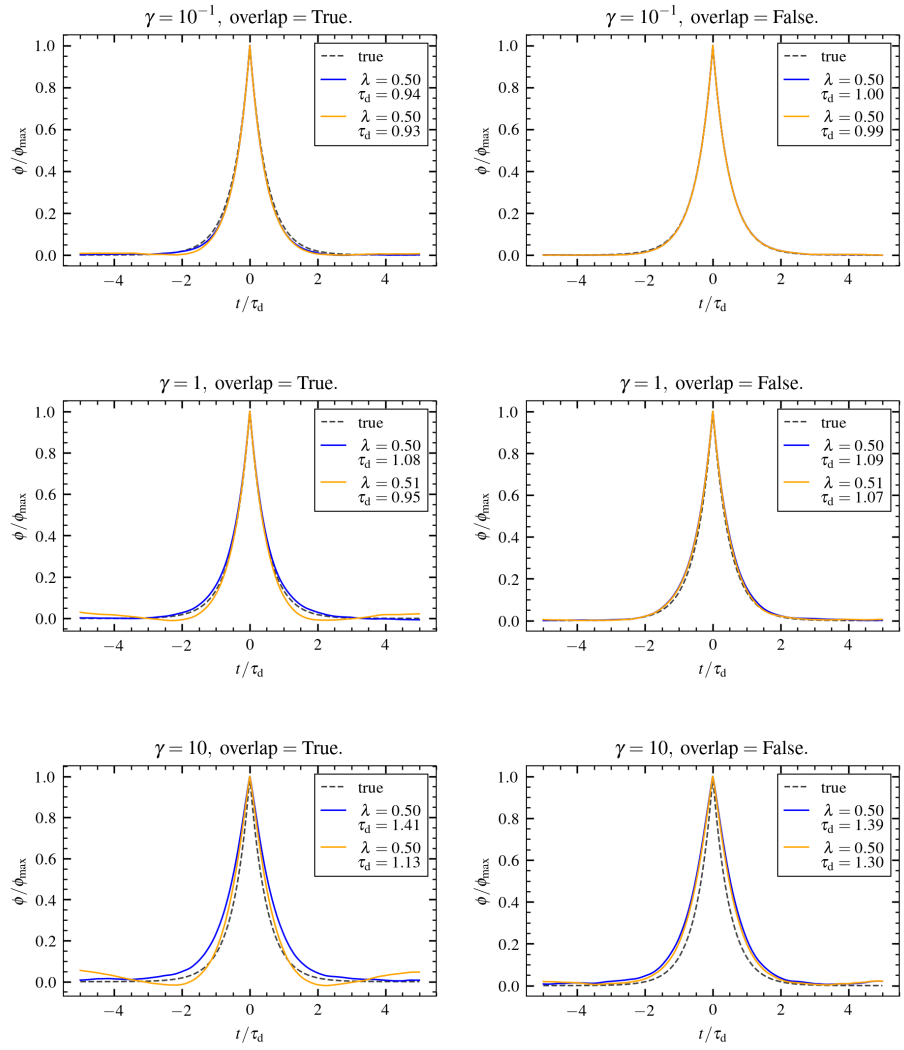


Figure 59: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A two sided symmetric exponential pulse function was used to generate the signal.

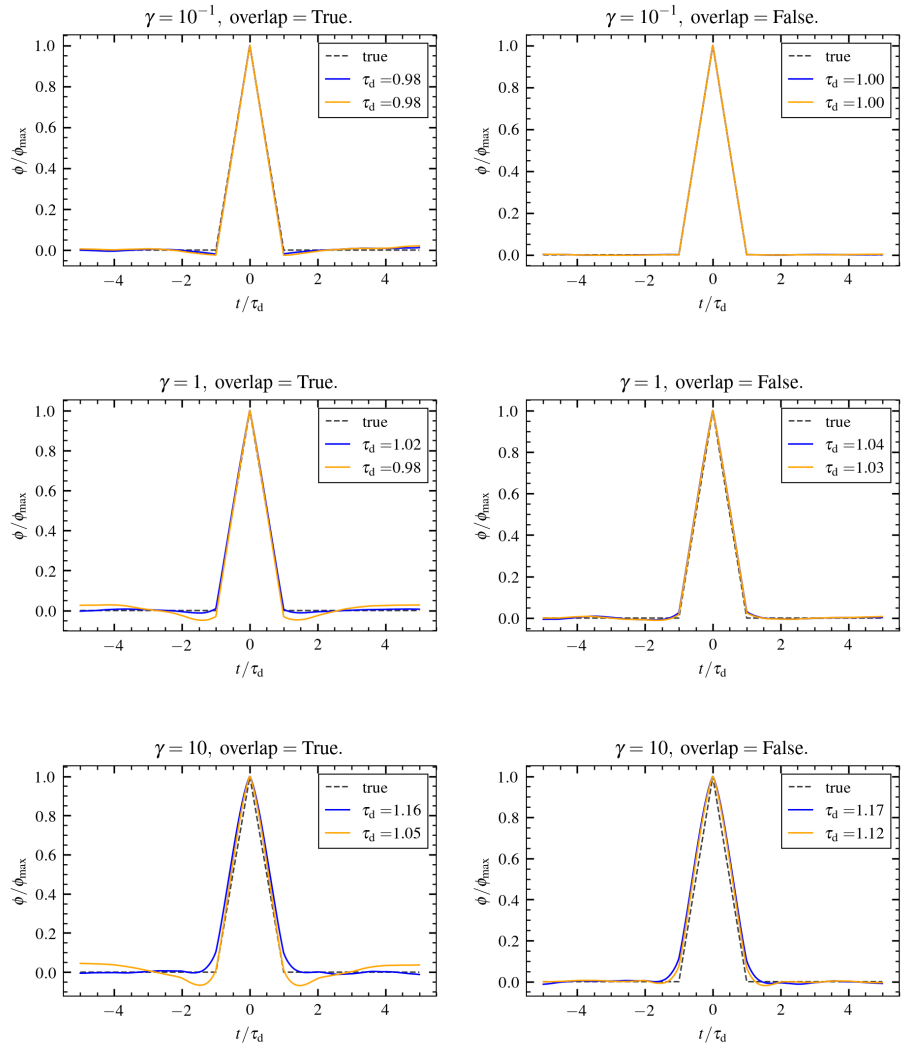


Figure 60: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A triangular pulse function was used to generate the signal.

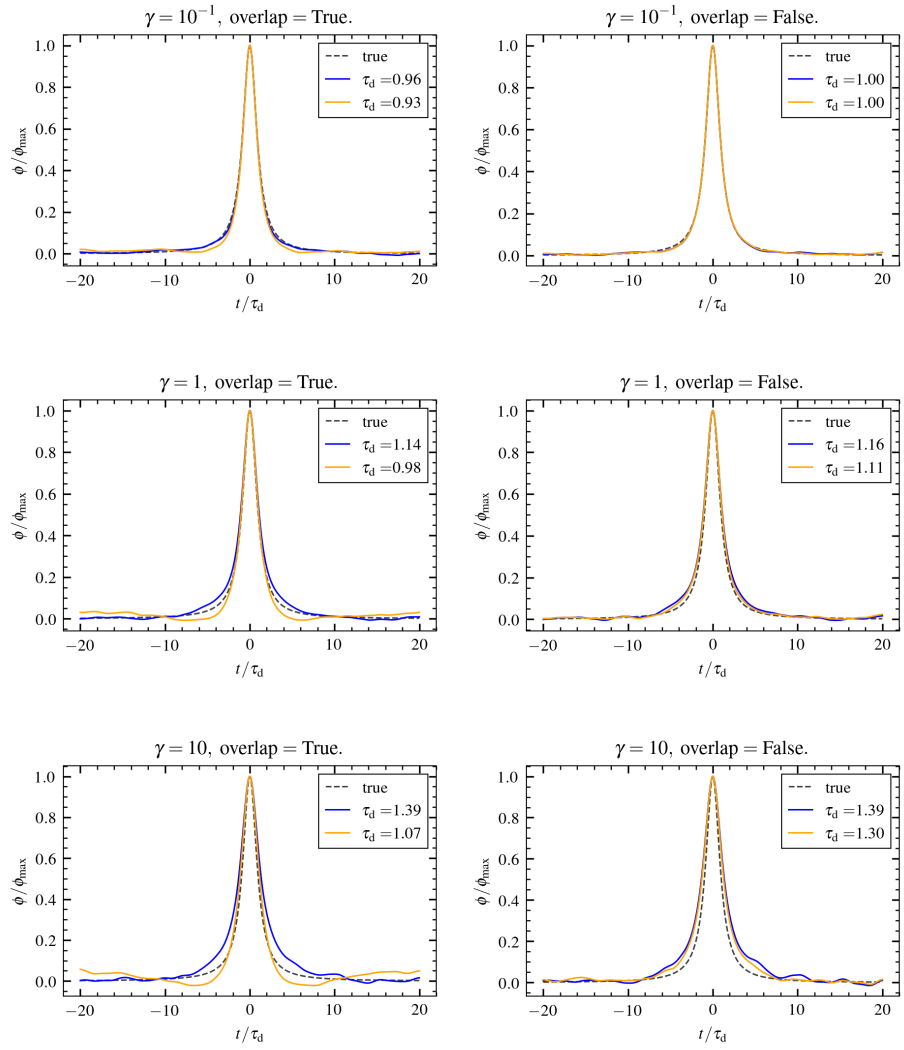


Figure 61: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A Lorentz pulse function was used to generate the signal.

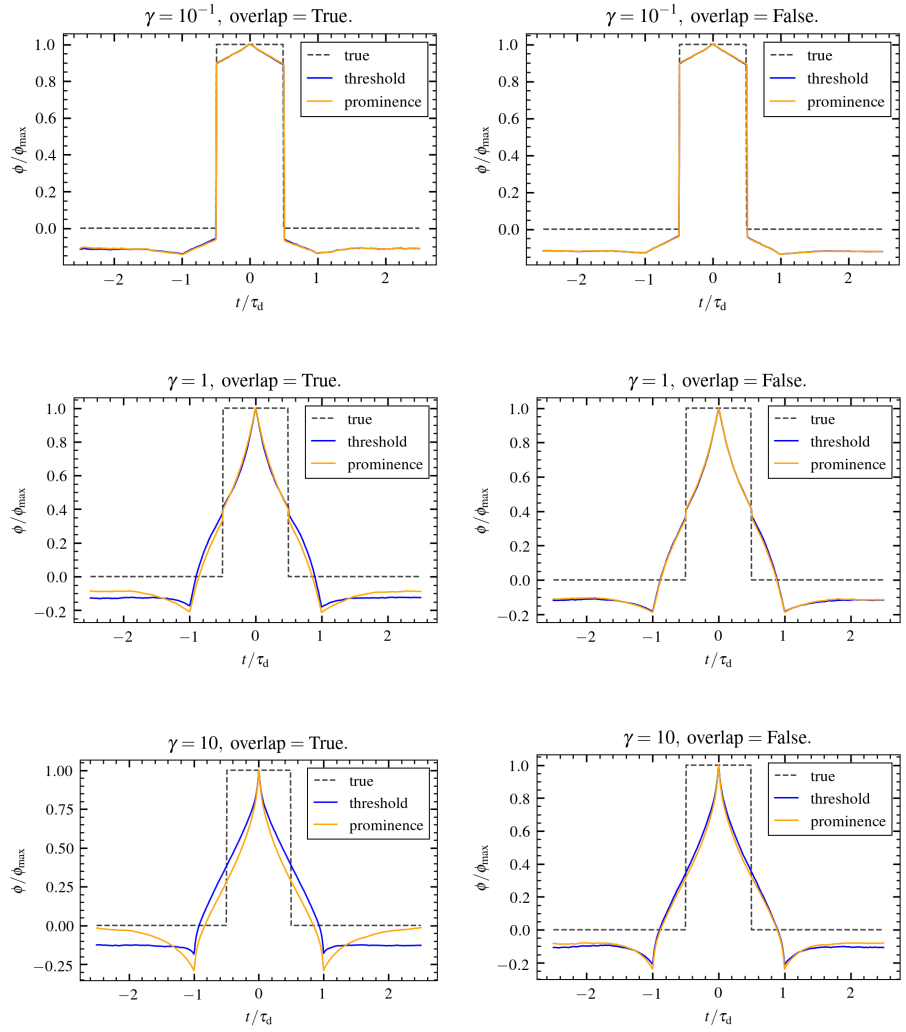


Figure 62: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A box pulse function was used to generate the signal.

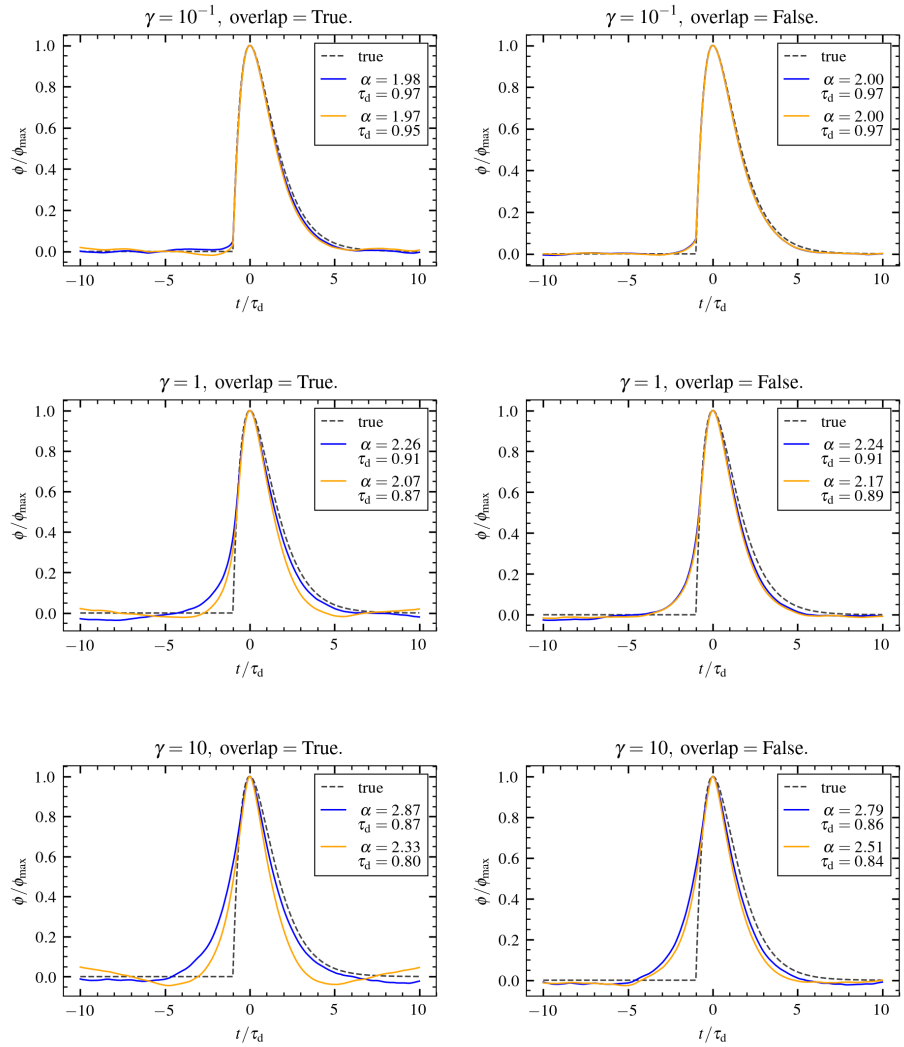


Figure 63: Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). γ =0.1 (top), 1 (middle) and 10 (bottom). A gamma pulse function with $\alpha = 2$ was used to generate the signal.

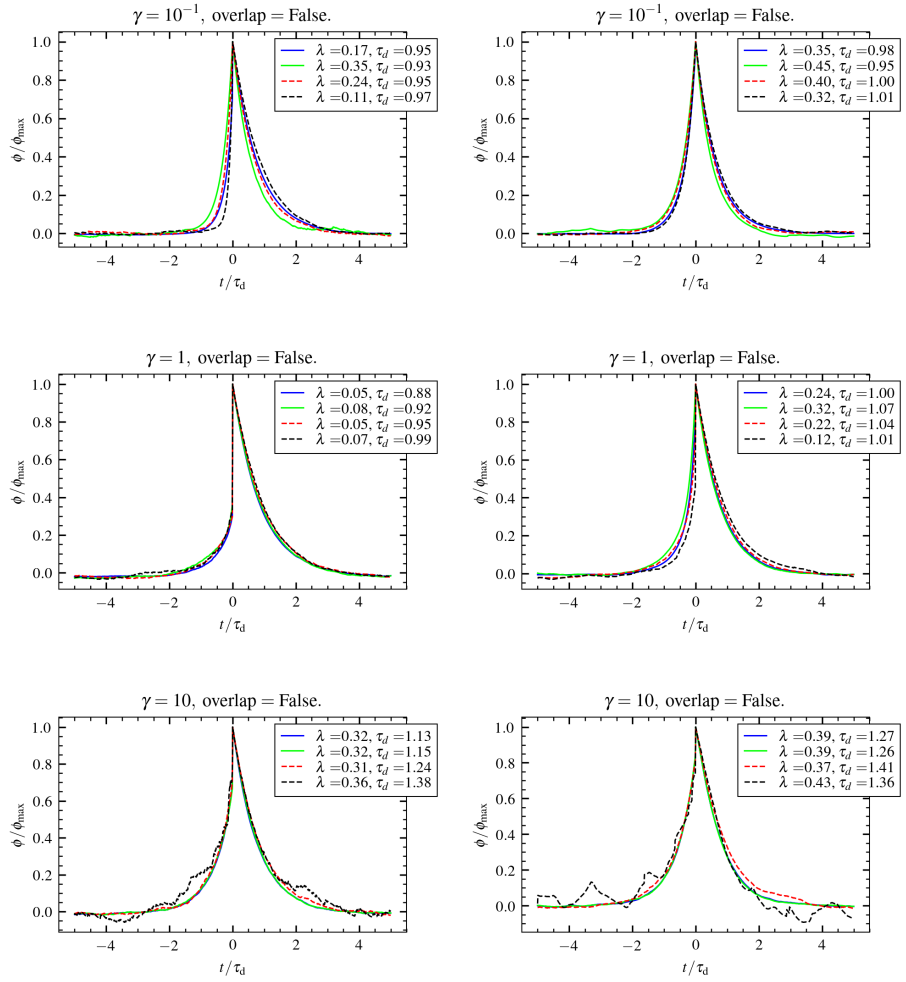


Figure 64: Conditionally averaged waveforms and the estimated λ and τ_d for $\gamma = 10^{-1}$ (top), 1 (middle) and 10 (bottom). The thresholds are > 2.5 (blue), $2-4$ (lime), $4-6$ (red) and $6-8$ (black) in units of the signals rms value. The left figures has the a, b, c, d, e values of $1/2, 1, 3/2, 2, 5/2$ (more assymetric events) and the right figures has the values $1, 2, 3, 4, 5$ (fewer assymetric events). The input amplitude distribution was Exp(1).

F.2 Additional waveforms with noise

In this appendix we present additional figures from the waveform analysis with noise in section 4.1.2. The reason why they are included here, and not in the discussion, is because of the traits of conditional averaging that they demonstrate were already demonstrated by other figures.

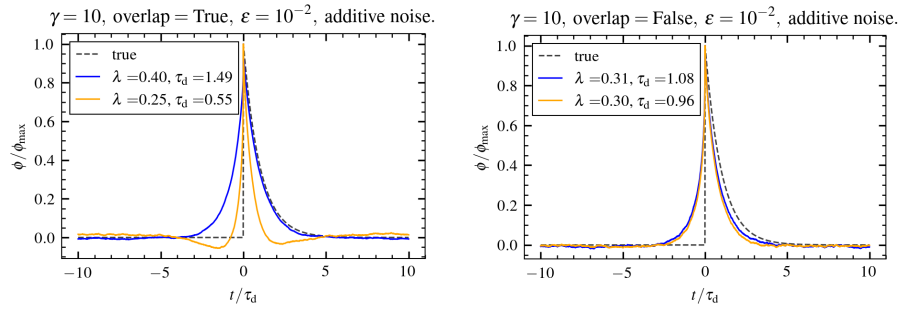


Figure 65: $\varepsilon = 10^{-2}$ Additive noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

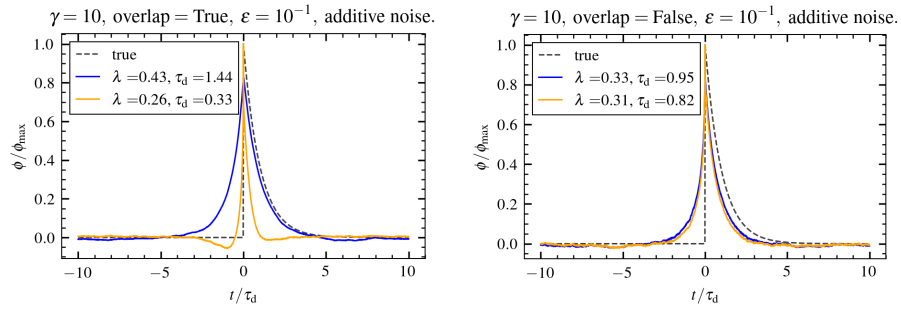


Figure 66: $\varepsilon = 10^{-1}$ Additive noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

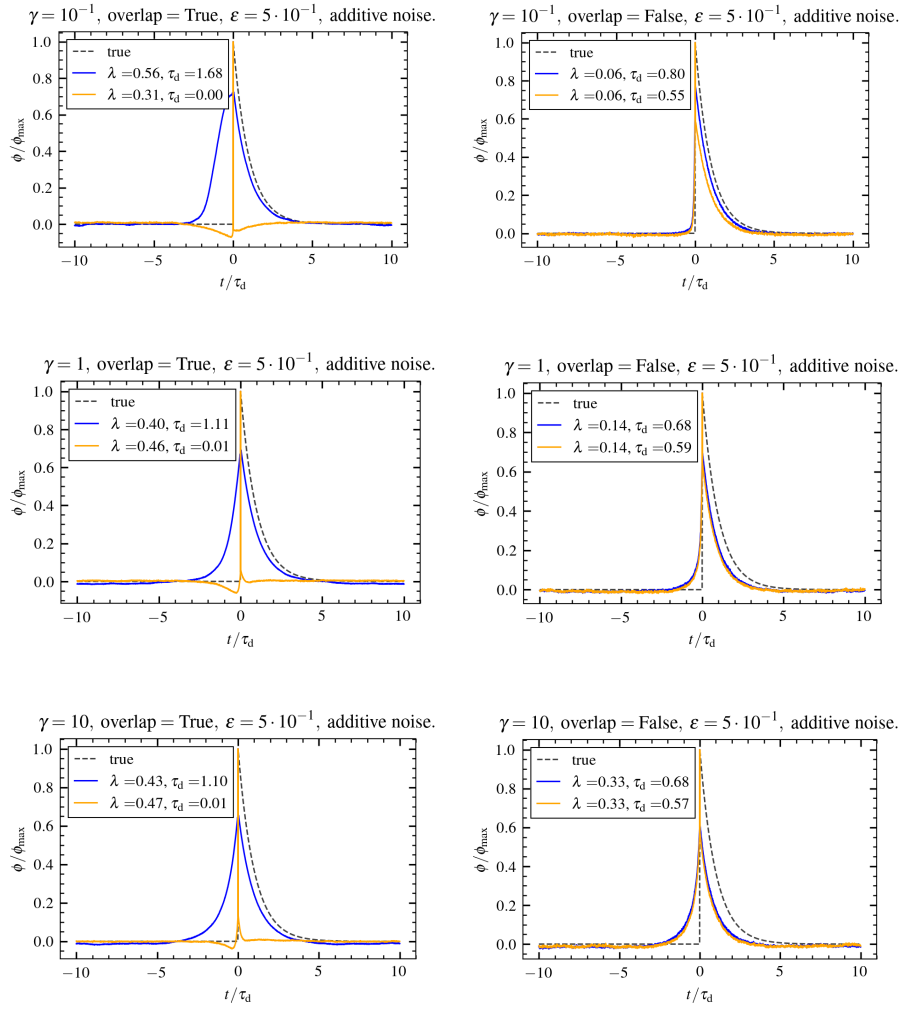


Figure 67: $\varepsilon = 0.5$ Additive noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

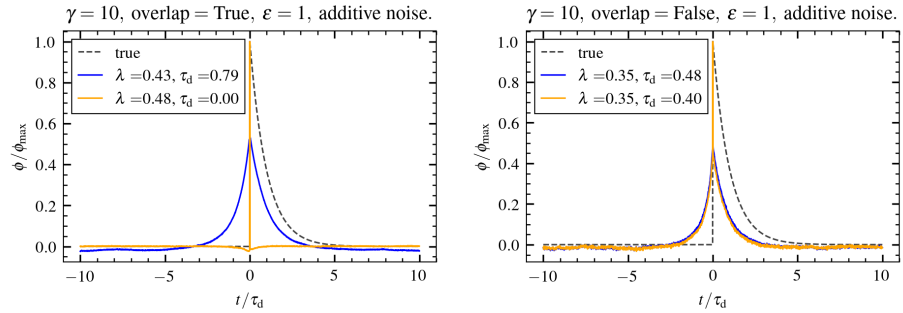


Figure 68: $\varepsilon = 1$ Additive noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

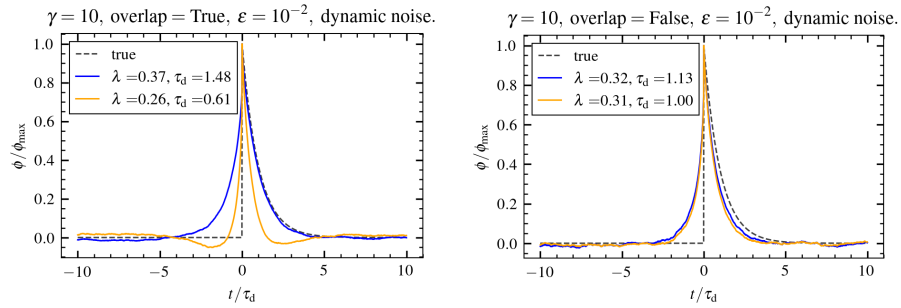


Figure 69: $\varepsilon = 10^{-2}$ Dynamic noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

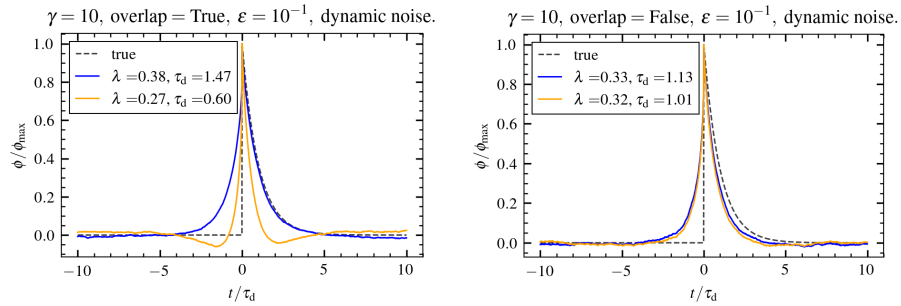


Figure 70: $\varepsilon = 10^{-1}$ Dynamic noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

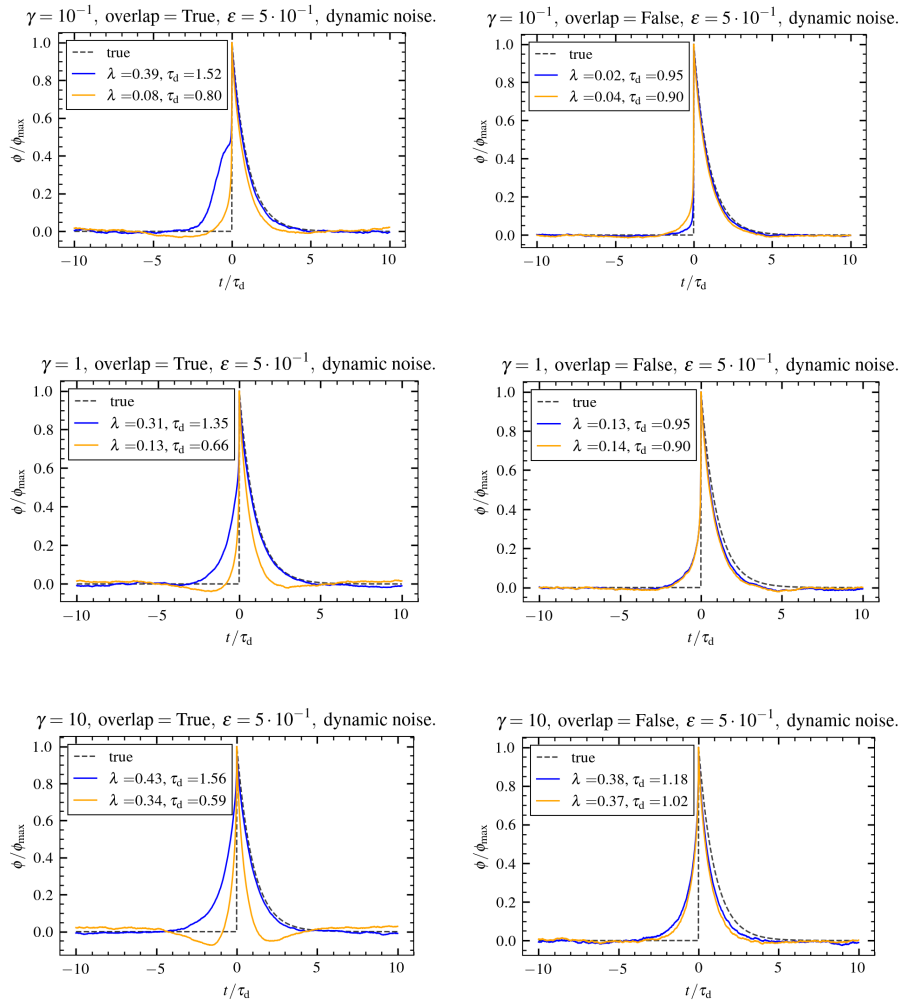


Figure 71: $\varepsilon = 0.5$ Dynamic noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

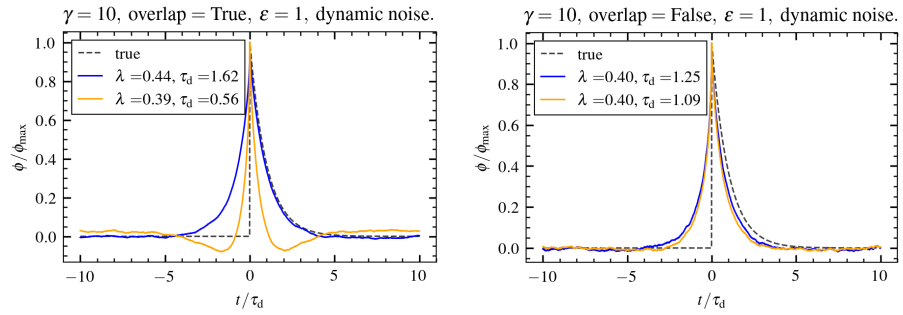


Figure 72: $\varepsilon = 1$ Dynamic noise. Conditionally averaged waveform for the threshold method (blue) and the prominence method (orange), without windowing (left) and with windowing (right). A double exponential pulse function with $\lambda = 0$ was used to generate the signal.

F.3 Additional amplitude distributions

In this appendix we present additional figures from the amplitude analysis in section 4.2

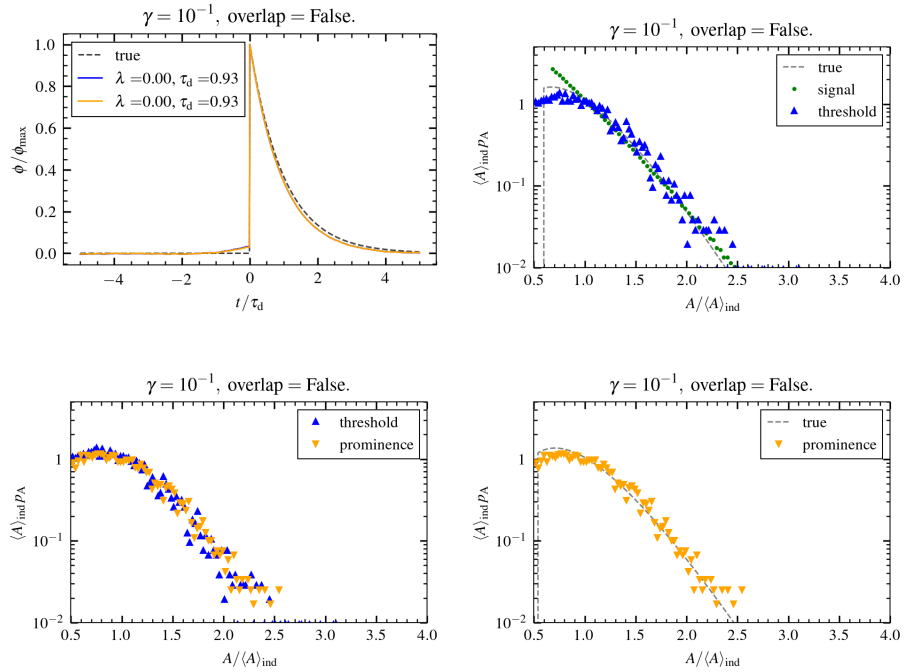


Figure 73: $\gamma = 10^{-1}$ $A \sim \Gamma(5, \frac{1}{5})$. Conditionally averaged waveform for both methods. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

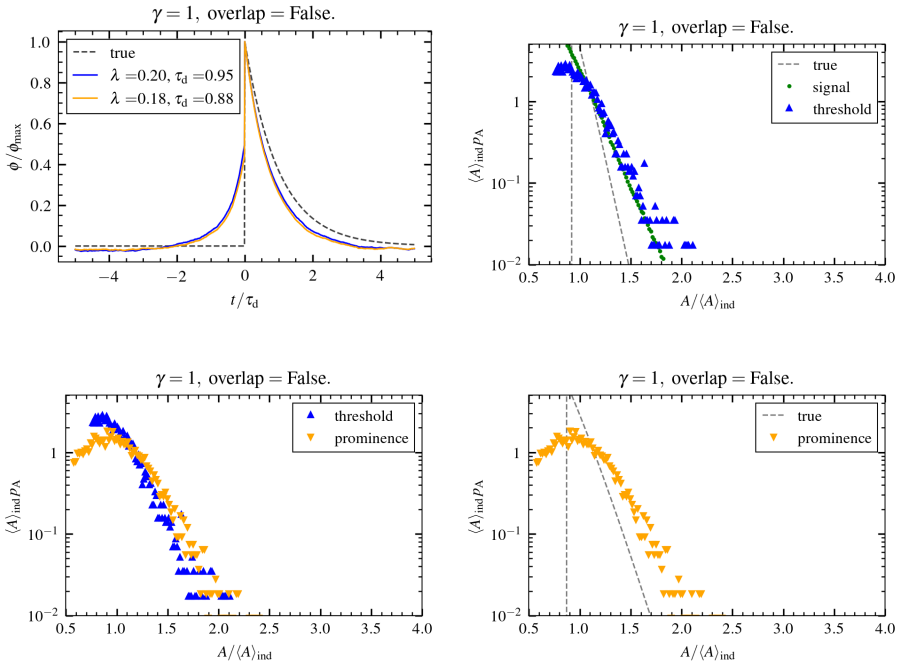


Figure 74: $\gamma = 1$ $A \sim \Gamma(5, \frac{1}{5})$. Conditionally averaged waveform for both methods. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

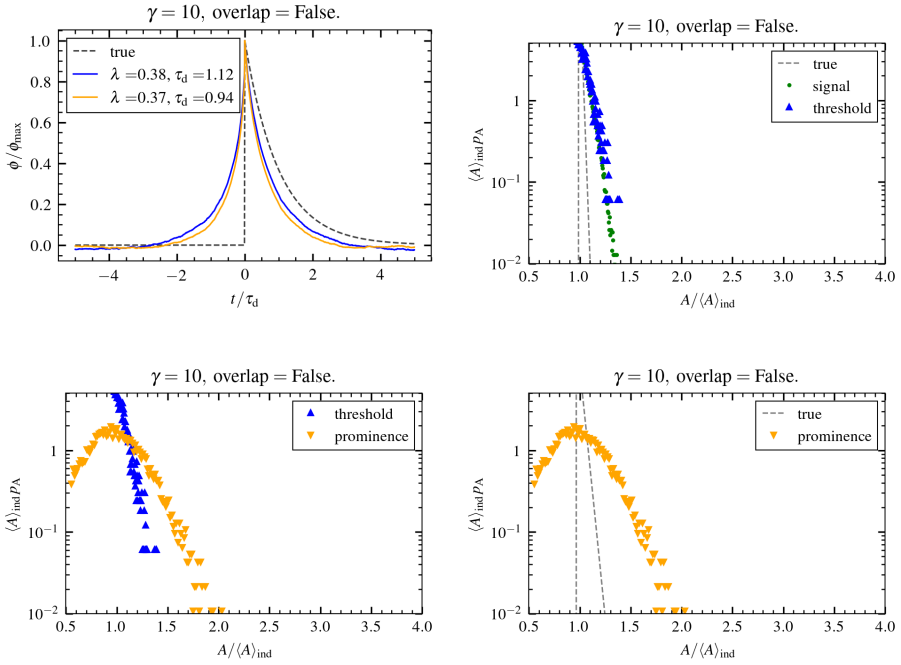


Figure 75: $\gamma = 10$ $A \sim \Gamma(5, \frac{1}{5})$. Conditionally averaged waveform for both methods. Amplitude distribution for the pure threshold (blue) method, prominence distribution (orange) for the prominence method and conditional signal distribution (green).

F.4 Additional waiting time distributions

In this appendix we present additional figures from the waiting time distribution analysis in section 4.3. The reason why they are included here, and not in the discussion, is because of the traits of conditional averaging that they demonstrate were already well demonstrated by other figures.

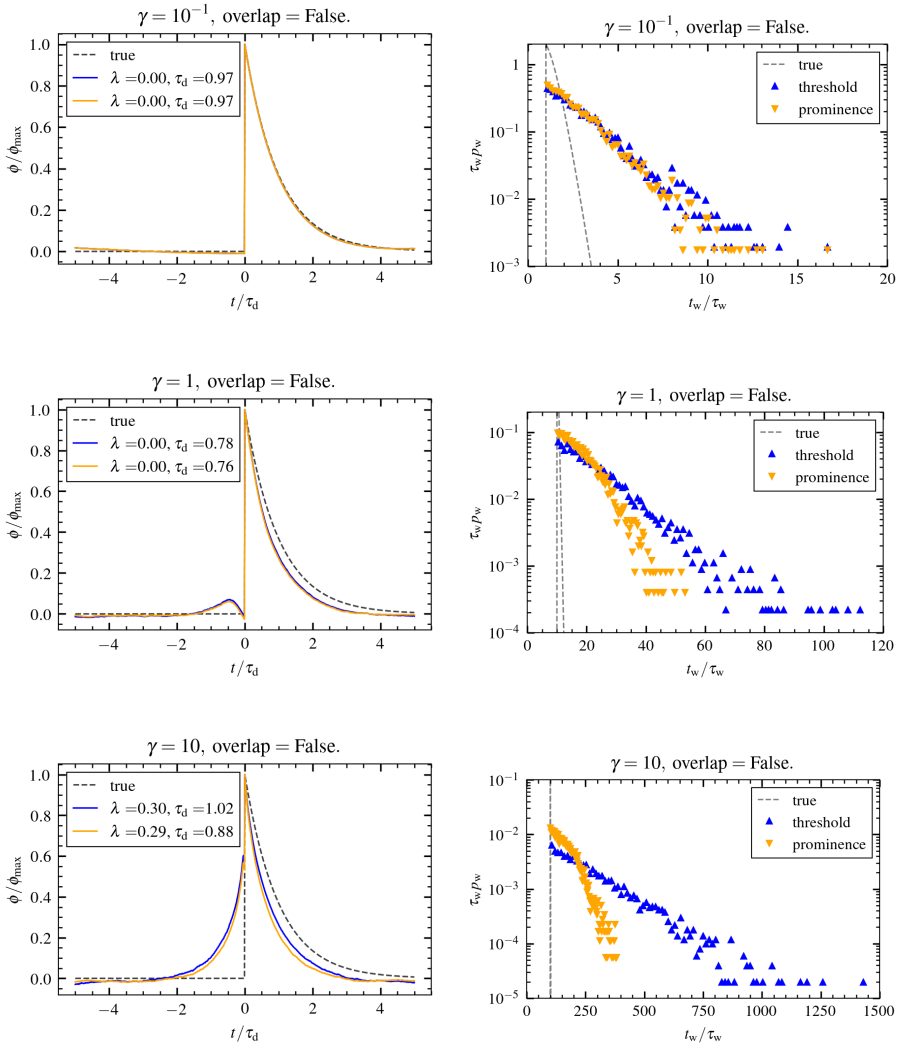


Figure 76: $\gamma = 10^{-1}$ $T_w \sim \Gamma(5, \frac{1}{5\gamma})$. Conditionally averaged waveform and waiting time distributions for both the pure threshold (blue) and prominence (orange) methods.

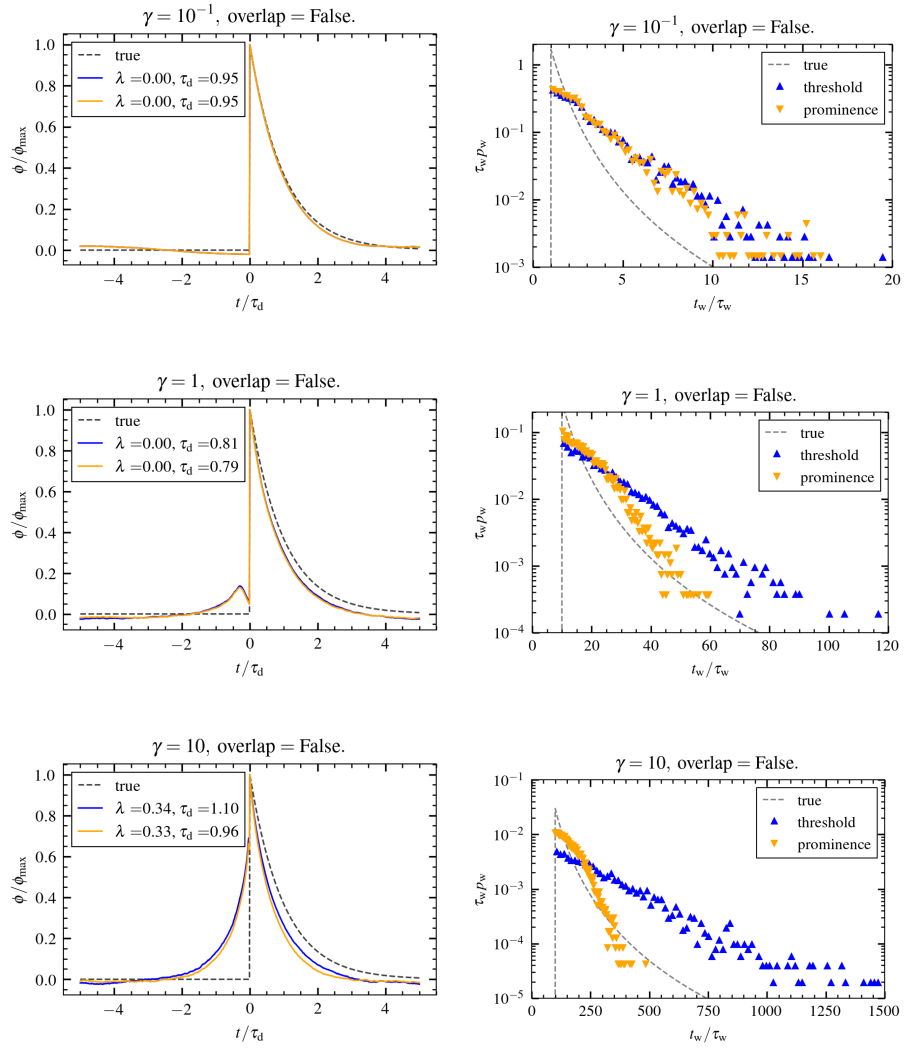


Figure 77: $T_w \sim \Gamma^{-1}(3, \frac{2}{\gamma})$. Conditionally averaged waveform and waiting time distributions for both the pure threshold (blue) and prominence (orange) methods.

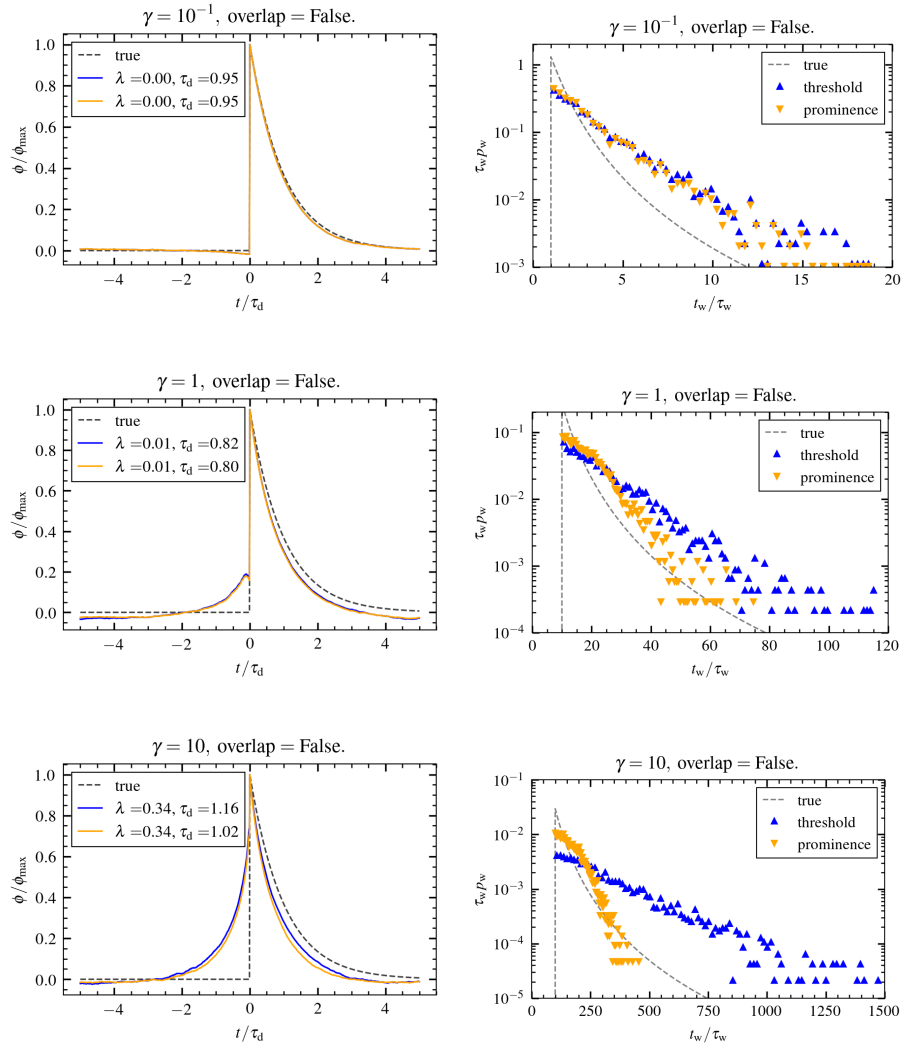


Figure 78: $T_w \sim \beta'(3, 3, \frac{2}{3\gamma})$. Conditionally averaged waveform and waiting time distributions for both the pure threshold (blue) and prominence (orange) methods.

G Python code

In this appendix links to the GitHub pages where the main code used in this project is located is presented.

- The self produced code https://github.com/RolfNi/Master
- The filtered Poisson process framework, statistical analysis, and conditional averaging code by the UiT Complex Systems Modelling group https://github.com/uit-cosmo/

References

[1] L. Kristensen, M. Casanova, M. S. Courtney, and I. Troen. In search of a gust definition. *Boundary-Layer Meteorology*, 55(1-2):91–107, April 1991. doi:10.1007/bf00119328.

[2] Roddam Narasimha, S. Rudra Kumar, A Prabhu, and S.V Kailas. Turbulent flux events in a nearly neutral atmospheric boundary layer. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1852):841–858, January 2007. doi:10.1098/rsta.2006.1949.

[3] Mario Lefebvre. Generalized filtered poisson processes and application in hydrology. *Statistics &amp Probability Letters*, 78(18):3274–3276, December 2008. doi:10.1016/j.spl.2008.06.020.

[4] P. Claps, A. Giordano, and F. Laio. Advances in shot noise modeling of daily streamflows. *Advances in Water Resources*, 28(9):992–1000, September 2005. doi:10.1016/j.advwatres.2005.03.008.

[5] Shan-Shan Ding, Hui-Min Li, Wen-Dan Yan, and Jin-Qiang Zhong. Temperature fluctuations relevant to thermal-plume dynamics in turbulent rotating rayleigh-bénard convection. *Physical Review Fluids*, 4(2), February 2019. doi:10.1103/physrevfluids.4.023501.

[6] Yin Wang, Xiaozhou He, and Penger Tong. Turbulent temperature fluctuations in a closed rayleigh–bénard convection cell. *Journal of Fluid Mechanics*, 874:263–284, July 2019. doi:10.1017/jfm.2019.405.

[7] G. Decristoforo, A. Theodorsen, and O. E. Garcia. Intermittent fluctuations due to lorentzian pulses in turbulent thermal convection. *Physics of Fluids*, 32(8):085102, August 2020. doi:10.1063/5.0012017.

[8] Jean-Baptiste Thomazo, Benjamin Le Révérend, Lea-Laetitia Pontani, Alexis M. Prevost, and Elie Wandersman. A bending fluctuation-based mechanism for particle detection by ciliated structures. *Proceedings of the National Academy of Sciences*, 118(31), July 2021. doi:10.1073/pnas.2020402118.

[9] Zs. Elter, C. Jammes, I. Pázsit, L. Pál, and P. Filliatre. Performance investigation of the pulse and campbelling modes of a fission chamber using a poisson pulse train simulation code. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 774:60–67, February 2015. doi:10.1016/j.nima.2014.11.065.

[10] Ghassan Y. Antar, Glenn Counsell, Yang Yu, Brian Labombard, and Pascal Devynck. Universality of intermittent convective transport in the scrape-off layer of magnetically confined devices. *Physics of Plasmas*, 10(2):419–428, February 2003. doi:10.1063/1.1536166.

[11] R. Kube, A. Theodorsen, O. E. Garcia, D. Brunner, B. LaBombard, and J. L. Terry. Comparison between mirror langmuir probe and gas-puff imaging measurements of intermittent fluctuations in the alcator c-mod scrape-off layer. *Journal of Plasma Physics*, 86(5), October 2020. doi:10.1017/s0022377820001282.

[12] D. A. D'Ippolito, J. R. Myra, and S. J. Zweben. Convective transport by intermittent blob-filaments: Comparison of theory and experiment. *Physics of Plasmas*, 18(6):060501, June 2011. doi:10.1063/1.3594609.

[13] A Theodorsen, O E Garcia, J Horacek, R Kube, and R A Pitts. Scrape-off layer turbulence in TCV: evidence in support of stochastic modelling. *Plasma Physics and Controlled Fusion*, 58(4):044006, January 2016. doi:10.1088/0741-3335/58/4/044006.

[14] Vincent Hakim and Jonas Ranft. Lifetime of a structure evolving by cluster aggregation and particle loss, and application to postsynaptic scaffold domains. *Physical Review E*, 101(1), January 2020. doi:10.1103/physreve.101.012411.

[15] O. E. Garcia, V. Naulin, A. H. Nielsen, and J. Juul Rasmussen. Turbulence and intermittent transport at the boundary of magnetized plasmas. *Physics of Plasmas*, 12(6):062309, June 2005. doi:10.1063/1.1925617.

[16] H. Karimabadi, V. Roytershteyn, M. Wan, W. H. Matthaeus, W. Daughton, P. Wu, M. Shay, B. Loring, J. Borovsky, E. Leonardis, S. C. Chapman, and T. K. M. Nakamura. Coherent structures, intermittent turbulence, and dissipation in high-temperature plasmas. *Physics of Plasmas*, 20(1):012303, January 2013. doi:10.1063/1.4773205.

[17] Johan Anderson and Bogdan Hnat. Statistical analysis of hasegawa-wakatani turbulence. *Physics of Plasmas*, 24(6):062301, June 2017. doi:10.1063/1.4984985.

[18] H. L. Pécseli and J. Trulsen. A statistical analysis of numerically simulated plasma turbulence. *Physics of Fluids B: Plasma Physics*, 1(8):1616–1636, August 1989. doi:10.1063/1.858940.

[19] Frank J. Øynes, Hans L. Pécseli, and Kristoffer Rypdal. Fluctuations in a magnetized toroidal plasma without rotational transform. *Physical Review Letters*, 75(1):81–84, July 1995. doi:10.1103/physrevlett.75.81.

[20] D Block, I Teliban, F Greiner, and A Piel. Prospects and limitations of conditional averaging. *Physica Scripta*, T122:25–33, January 2006. doi:10.1088/0031-8949/2006/t122/007.

[21] I Teliban, D Block, A Piel, and F Greiner. Improved conditional averaging technique for plasma fluctuation diagnostics. *Plasma Physics and Controlled Fusion*, 49(4):485–497, March 2007. doi:10.1088/0741-3335/49/4/011.

[22] H. Johnsen, H. L. Pe

cseli, and J. Trulsen. Conditional eddies in plasma turbulence. *Physics of Fluids*, 30(7):2239, 1987. doi:10.1063/1.866158.

[23] O. E. Garcia and A. Theodorsen. Auto-correlation function and frequency spectrum due to a super-position of uncorrelated exponential pulses. *Physics of Plasmas*, 24(3):032309, March 2017. doi:10.1063/1.4978955.

[24] M. Zurita, W. A. Hernandez, C. Crepaldi, F. A. C. Pereira, and Z. O. Guimarães-Filho. Stochastic modeling of plasma fluctuations with bursts and correlated noise in TCABR. *Physics of Plasmas*, 29(5):052303, May 2022. doi:10.1063/5.0081281.

[25] A Theodorsen and O E Garcia. Probability distribution functions for intermittent scrape-off layer plasma fluctuations. *Plasma Physics and Controlled Fusion*, 60(3):034006, January 2018. doi:10.1088/1361-6587/aa9f9c.

[26] Onno J. Boxma and Uri Yechiali. scpp/scp oisson processes, December 2007. doi:10.1002/9780470061572.eqr055.

[27] A. Theodorsen, O.E. Garcia, R. Kube, B. LaBombard, and J.L. Terry. Relationship between frequency power spectra and intermittent, large-amplitude bursts in the alcator c-mod scrape-off layer. *Nuclear Fusion*, 57(11):114004, August 2017. doi:10.1088/1741-4326/aa7e4c.

[28] R Kube, A Theodorsen, O E Garcia, B LaBombard, and J L Terry. Fluctuation statistics in the scrape-off layer of alcator c-mod. *Plasma Physics and Controlled Fusion*, 58(5):054001, March 2016. doi:10.1088/0741-3335/58/5/054001.

[29] R Nordal. Report on conditional averaging, 2022.

[30] O. E. Garcia, R. Kube, A. Theodorsen, B. LaBombard, and J. L. Terry. Intermittent fluctuations in the alcator c-mod scrape-off layer for ohmic and high confinement mode plasmas. *Physics of Plasmas*, 25(5):056103, May 2018. doi:10.1063/1.5018709.

[31] R Kube, O E Garcia, A Theodorsen, D Brunner, A Q Kuang, B LaBombard, and J L Terry. Intermittent electron density and temperature fluctuations and associated fluxes in the alcator c-mod scrape-off layer. *Plasma Physics and Controlled Fusion*, 60(6):065002, April 2018. doi:10.1088/1361-6587/aab726.

[32] O.E. Garcia, R. Kube, A. Theodorsen, J.-G. Bak, S.-H. Hong, H.-S. Kim, the KSTAR Project Team, and R.A. Pitts. SOL width and intermittent fluctuations in KSTAR. *Nuclear Materials and Energy*, 12:36–43, August 2017. doi:10.1016/j.nme.2016.11.008.

[33] A. Theodorsen and O. E. Garcia. Level crossings and excess times due to a superposition of uncorrelated exponential pulses. *Physical Review E*, 97(1), January 2018. doi:10.1103/physreve.97.012110.

[34] F. Militello and J.T. Omotani. Scrape off layer profiles interpreted with filament dynamics. *Nuclear Fusion*, 56(10):104004, August 2016. doi:10.1088/0029-5515/56/10/104004.

[35] A. Theodorsen, O. E. Garcia, R. Kube, B. LaBombard, and J. L. Terry. Universality of poisson-driven plasma fluctuations in the alcator c-mod scrape-off layer. *Physics of Plasmas*, 25(12):122309, December 2018. doi:10.1063/1.5064744.

[36] O.E. Garcia, J. Horacek, and R.A. Pitts. Intermittent fluctuations in the TCV scrape-off layer. *Nuclear Fusion*, 55(6):062002, May 2015. doi:10.1088/0029-5515/55/6/062002.

[37] O. E. Garcia, S. M. Fritzner, R. Kube, I. Cziegler, B. LaBombard, and J. L. Terry. Intermittent fluctuations in the alcator c-mod scrape-off layer. *Physics of Plasmas*, 20(5):055901, May 2013. doi:10.1063/1.4802942.

[38] Ronald E Walpole, Raymond H Myers, Sharon L Myers, and Keying E Ye. *Probability & statistics for engineers & scientists, global edition*. Pearson Education, London, England, 9 edition, July 2016.

[39] Athanasios Papoulis and S Unnikrishna Pillai. *Probability, random variables and stochastic processes with errata sheet (int'l ed)*. McGraw-Hill Professional, New York, NY, 4 edition, January 2002.

[40] David Vose. *Risk Analysis*. John Wiley & Sons, Chichester, England, 3 edition, March 2008.

[41] Marcelo Bourguignon, Manoel Santos-Neto, and Mário de Castro. A new regression model for positive random variables with skewed and long tail. *METRON*, 79(1):33–55, March 2021. doi:10.1007/s40300-021-00203-y.

[42] Andrew G. Glen. On the inverse gamma as a survival distribution. In *International Series in Operations Research &amp Management Science*, pages 15–30. Springer International Publishing, December 2016. doi:10.1007/978-3-319-43317-2_2.

[43] Geoffrey Grimmett and David Stirzaker. *Probability and random processes*. Oxford University Press, London, England, 3 edition, May 2001.