Reconstruction of intermittent data time series as a superposition of pulses

Sajidah Ahmed,[*] Odd Erik Garcia,[†] and Audun Theodorsen[‡]

Department of Physics and Technology,

UiT The Arctic University of Norway, N-9037 Tromsø, Norway

(Dated: June 21, 2023)

Abstract

Fluctuations in a vast range of physical systems can be described as a superposition of uncorrelated pulses with a fixed shape, a process commonly referred to as a (generalized) shot noise or a filtered Poisson process. In this contribution, we present a systematic study of a novel deconvolution method to estimate the arrival times and amplitudes of the pulses from realizations of such processes. The method shows that time-series can be reconstructed for various pulse amplitude and waiting time distributions. Despite a constraint on positive-definite amplitudes, it is shown that negative amplitudes may also be reconstructed by flipping the sign of the time series. The method performs well under moderate amounts of additive noise, both white noise and colored noise having the same correlation function as the process itself. The estimation of pulse shapes from the power spectrum is accurate except for excessively broad waiting time distributions. Although the method assumes constant pulse durations, it performs well under narrowly distributed pulse durations. The most important constraint on the reconstruction is information-loss, which limits the method to intermittent processes. The ratio between the sampling time and the average waiting time between pulses must be about 1/20 or smaller for a well sampled signal. Finally, given the system forcing, the average pulse function may be recovered. This recovery is only weakly constrained by the intermittency of the process.

I. INTRODUCTION

Intermittent and seemingly random fluctuations of order unity compared to the mean value are found in a variety of nonlinear physical systems, such as turbulence in neutral fluids [1–4] and atmospheric winds [5, 6], water resources and hydrology [7, 8], complex fluids [9], fission chambers [10], physiology [11–13] and biophysics [14] and plasma turbulence, both simulations [15–18] and measurements from magnetically confined plasmas [19–27]. While such fluctuations may be extremely challenging to investigate from first principles based models, many fruitfully admit phenomenological modeling. One particularly useful reference model for time-series measurements from these systems is a stochastic model based on a superposition of localized pulses [4, 7–10, 13, 14, 18, 22, 25]. This model is called a

---

* sajidah.ahmed@uit.no

† odd.erik.garcia@uit.no

‡ audun.theodorsen@uit.no

shot noise process [28, 29] or a filtered Poisson process (FPP) [30]. If all pulses have the same functional shape and duration, the process can be written as a convolution between the pulse function (which may be considered a system response) and a random forcing. In the simplest case, the random forcing is driven by a Poisson process, giving the FPP its name.

In many cases, it is of interest to extract amplitude and waiting-time statistics of the pulses or alternatively, the average pulse function if the forcing is known. One popular family of methods is conditional averaging [31–33], used, for example, in [2, 6, 20, 22, 34–36]. Here, an amplitude threshold is set, and each time the signal crosses above this threshold, the time and amplitude of the peak are recorded along with the shape of the signal around the peak. Different authors use slightly different methods, and to the best of our knowledge, a systematic study of conditional averaging for the statistics of overlapping pulses in single-point time series is not available, although the case for non-overlapping structures has been investigated [33, 37] as has the case for two-dimensional structures with multi-point measurements [31, 38]. It is clear that both pulse overlap, and threshold requirement may influence the results and applicability of the conditional averaging method.

If the pulse function is known or may be estimated, some form of deconvolution may be performed to recover the forcing from realizations of the process. A method based on iterative deconvolution of the FPP has been shown to be robust in this problem [4, 23, 39]. This algorithm, referred to as the Richardson-Lucy (RL) algorithm for Poissonian noise [40, 41] or the iterative space reconstruction algorithm (ISRA) for normally distributed noise [42] is not new; it was originally developed for image data in astronomy [40, 41] but has seen use in several other imaging applications [42–45]. This method requires a known or estimated common pulse function for all arrivals in the time series and reproduces the forcing in the maximum-likelihood sense under normally distributed and uncorrelated additive noise.

In this contribution, we estimate pulse amplitudes and arrival times from the forcing by applying an iterative deconvolution on realizations of the process. Our main aim is to report accurately on the prospects and limitations of using this procedure as a tool for time series analysis. A wide variety of assumptions regarding pulse overlap, pulse functions, amplitude and waiting time distributions, additive noise, and correlations between these may be made in different contexts and applications. Further, noise, pulse superposition, and distribution of pulse shapes put limits on the estimation of pulse amplitudes and arrivals.

To limit the scope of this contribution, we will focus on assumptions relevant for time series measurements in turbulent fluids and plasmas, and in particular at the boundary of magnetically confined fusion plasmas which are characterized by broad and positive definite amplitude distributions, close to Poisson distributed arrivals and asymmetric, positive and exponentially decaying pulses [20–27, 34, 35, 39].

We note that the deconvolution algorithm may also be used the other way around: If the forcing is known, the pulse function may be estimated using the same algorithm. In this way, it may be used to find the system response to a controlled input forcing, if the assumption of linearity is satisfied.

This contribution is structured as follows: In Sec. II we review the stochastic model, define the base case of the model which will be modified during the following investigations, and present the deconvolution algorithm. In Sec. III the limitations of the signal reconstruction from estimated amplitudes and arrival times due to sampling and pulse overlap are investigated. In Sec. IV we assess the ability of the algorithm to reproduce various amplitude and arrival times for a known pulse function. Then, in Sec. V distortions due to additive noise, both uncorrelated and correlated, are investigated and a criterion for noise removal is established. Following this, in Sec. VI we report on the effect of estimating the pulse shape assuming a known functional form. Both the effects of over- and underestimating the pulse duration are considered, as well as the effect of a narrow distribution of pulse durations. Lastly, we turn the question around in Sec. VII and look at how well an unknown pulse function can be estimated if the pulse arrivals and amplitudes are known. Finally, we discuss the results and conclude in Sec. VIII. In this contribution, we will not consider the effects of statistical convergence: we will always use long time series to avoid large uncertainty or bias in parameter estimates and will always run the deconvolution algorithm to convergence. Effects of short time series on moment estimation were previously investigated in Refs. 46 and 47. Secs. III–VII are largely independent; the reader may consult to the problem of particular interest. The numerical implementation of this method with its library of functions is openly available on the GitHub page of the UiT Complex Systems Modelling group [48].

II. THEORY

In this section, we first review the stochastic model given by a superposition of pulses with fixed shape and duration. This is followed by a presentation of the deconvolution algorithm. The section concludes by discussing how pulse amplitudes and arrival times can be recovered from realizations of the stochastic process using the deconvolution algorithm. Throughout this manuscript, we adopt the notation that angular brackets $\langle \cdot \rangle$ refers to a theoretical mean value, while an overline $\overline{\cdot}$ refers to a sample mean. Quantities estimated by the deconvolution method use the est-subscript $\cdot_{\mathrm{est}}$.

A. The stochastic model

The basic stochastic model considered here is a superposition of pulses with a fixed shape defined as

$$\Phi_K(t) = \sum_{k=1}^{K(T)} A_k \varphi\left(\frac{t-s_k}{\tau_k}\right). \tag{1}$$

Here, $K(T)$ denotes a point process on the interval $[0, T)$ with sorted event arrival times s_k and waiting times $w_k = s_k - s_{k-1}$ with mean value $\langle w \rangle$. We will in general restrict $K(T)$ to be a *renewal* process, where the waiting times are independently and identically distributed [30]. The amplitudes A_k are randomly distributed with mean value $\langle A \rangle$, and in general assumed to be positive definite. The pulse function φ is assumed to be the same for all events but may have randomly distributed duration times τ_k. The average pulse duration time is denoted by $\tau_{\mathrm{d}} = \langle \tau \rangle$. In the following, we take all A_k, w_k and τ_k to be independent random variables, and each variable family is independently and identically distributed.

The fundamental parameter of the stochastic model is $\gamma = \tau_{\mathrm{d}}/\langle w \rangle$, referred to as the *intermittency parameter*. It describes the degree of pulse overlap and quantifies how intermittent the fluctuations are through the skewness and flatness moments [49]. In general, the influence of γ on the qualitative appearance and features of the process is the following: For γ of order unity or smaller, pulses appear well separated due to the waiting times between pulses being long compared to the duration time. When $\gamma < 1$, the process Φ will thus have a small mean value compared to the mean pulse amplitude and large relative fluctuation levels. For γ much larger than one, there is significant pulse overlap due to short waiting times and long pulse durations. This results is a process that is near normal distributed,

where Φ will have a large mean value and small relative fluctuation levels [49]. In Sec. III, we will see that it is not the intermittency parameter which determines our ability to estimate the amplitudes and arrival times of the process, but the ratio between the sampling time and the average waiting time. We write this ratio as $\gamma\theta$, where $\theta = \triangle_t/\tau_{\mathrm{d}}$ is the sampling time $\triangle_t$ normalized by the average pulse duration.

Here, some care must be taken regarding the sampling process. If we consider a sampling method which instantaneously measures a value at regular intervals, pulses which are only a few sampling times in duration may be undersampled, such that the true amplitude of the pulse is not captured. Thus the process may contain an artificially low number of pulses or have an artificially low mean value compared to an adequately sampled process. If the sampling time is an *integration time* such as a camera exposure time, all pulses will be identified. However, there may be distortions in the pulse shape due to the integration, as the integration smears out the pulse function. In either case, the deconvolution method described in Sec. II B reproduces the *sampled* time series and is not designed to make inferences regarding missing or distorted pulses. Thus, an accurate reconstruction of the true process also depends on a sufficiently high sampling rate to resolve the pulses. Our testing suggests that for the exponential pulse function in Eq. (2) (below), a sampling time of 10 times the typical pulse duration time or better is sufficient, that is, $\triangle_t/\tau_{\mathrm{d}} \leq 10^{-1}$.

We take the following as the *base case*:

- $K(T)$ is a Poisson process. Therefore, it follows that the arrival times are independent and uniformly distributed on the interval $[0, T)$, and the waiting times $w_k = s_k - s_{k-1}$ are independently and exponentially distributed with mean value $\langle w \rangle$ [30]. The mean value of the Poisson process is then given by $\langle K(T) \rangle = T/\langle w \rangle$.

- Degenerate distribution of pulse durations, $\tau_{\mathrm{d}} P_\tau(\tau) = \delta(\tau - \tau_{\mathrm{d}})$, so that all pulses have the same duration, $\tau_k = \tau_{\mathrm{d}}$.

- Fixed one-sided exponential pulse function for all events given by a jump followed by an exponential decay,

$$\varphi(x) = \begin{cases} 0, & x < 0, \\ \exp(-x), & x \geq 0. \end{cases} \tag{2}$$

.

6

- Exponentially distributed amplitudes with mean value $\langle A \rangle$, $\langle A \rangle P_A(A) = \exp\left(-A/\langle A \rangle\right)$ for $A > 0$.

Some consequences of this base case are of interest [49–52]:

- The probability distribution function (PDF) of Φ is a Gamma distribution with shape parameter $\gamma = \tau_{\mathrm{d}}/\langle w \rangle$ and scale parameter $\langle A \rangle$,

$$P_\Phi(\Phi; \gamma, \langle A \rangle) = \frac{\Phi^{\gamma-1}}{\langle A \rangle^\gamma \, \Gamma(\gamma)} \exp\left(-\frac{\Phi}{\langle A \rangle}\right), \quad \Phi > 0. \tag{3}$$

- The four lowest order moments are the mean $\langle \Phi \rangle = \gamma \langle A \rangle$, variance $\Phi_{\mathrm{rms}}^2 = \gamma \langle A \rangle^2$, skewness $S_\Phi = \langle (\Phi - \langle \Phi \rangle)^3 \rangle / \Phi_{\mathrm{rms}}^3 = 2/\sqrt{\gamma}$ and flatness (or excess kurtosis) $F_\Phi = \langle (\Phi - \langle \Phi \rangle)^4 \rangle / \Phi_{\mathrm{rms}}^4 - 3 = K_\Phi - 3 = 6/\gamma$, where K_Φ is the kurtosis.

- The frequency power spectral density (PSD) of Φ has a Lorentzian shape,

$$\Omega_\Phi(\omega) = \Phi_{\mathrm{rms}}^2 \frac{2\tau_{\mathrm{d}}}{1 + \tau_{\mathrm{d}}^2 \omega^2} + 2\pi \langle \Phi \rangle^2 \delta(\omega). \tag{4}$$

In the base case, there are three fundamental model parameters: γ, $\langle A \rangle$ and τ_{d}. From realizations of the process, the first two may be estimated from the PDF given by Eq. (3). From these follows the mean value and the standard deviation of the process. The final parameter τ_{d} may be estimated from the auto-correlation function or from the frequency PSD given by Eq. (4).

In the case that all pulses have the same duration τ_{d}, we may express the stochastic model as a convolution between the pulse function φ and a forcing f_K,

$$\Phi_K(t) = [\varphi * f_K]\left(\frac{t}{\tau_{\mathrm{d}}}\right), \tag{5}$$

where f_K is given by

$$f_K(t) = \sum_{k=1}^{K(T)} A_k \delta\left(\frac{t - s_k}{\tau_{\mathrm{d}}}\right). \tag{6}$$

Hence, one can say that Φ_K is a train of delta pulses, given by f_K, arriving according to the point process $K(T)$ which is passed through a filter φ. The FPP may be considered a linear model for highly nonlinear phenomena, where the nonlinearity has been baked into the distributions of A and s and the pulse function φ. Given an estimate of the pulse function, it is therefore possible to estimate f_K by deconvolving a realization Φ_K with the pulse function φ. Further discussions on estimation of model parameters from realizations of the process are given in Refs. 39, 46, 47, and 50.

7

B. Deconvolution method

To estimate the pulse arrival times s_k and amplitudes A_k from realizations of the stochastic process, the "iterative image space reconstruction algorithm" (ISRA), is presented in Ref. 42, and we point the reader to this publication for a more thorough discussion of the details of the algorithm. Here, we consider Φ, φ and f to be discretized with a uniform sampling time $\triangle_t$ and an odd number of data points N. Furthermore, Φ is assumed to be corrupted by uncorrelated white noise denoted X, so that we may write for $0 \leq j < N$

$$\Phi_j = (\varphi * f)_j + X_j \tag{7}$$

where f_j is a sum of Kronecker deltas weighted by the pulse amplitudes, $f_j = \sum_{k=1}^{K(T)} A_k \delta_{j,j_k}$, and by abuse of notation $\varphi_j = \varphi(j\triangle_t/\tau_{\mathrm{d}})$. We have suppressed the subscript K on Φ and f for simplicity of notation and j_k is $s_k/\triangle_t$ rounded to the nearest integer. In the following, the hat symbol $\widehat{\ }$ is used to denote a flipped vector, $\widehat{\varphi}_j = \varphi_{N-1-j}$.

The maximum-likelihood estimation applied to the above model leads to the optimization problem

$$\text{minimize } J_\Phi(f) = \tfrac{1}{2}\|\varphi * f - \Phi\|^2 \tag{8}$$

$$\text{subject to} \qquad f \geq 0. \tag{9}$$

The iteration scheme

$$f_j^{(n+1)} = f_j^{(n)} \frac{(\Phi * \widehat{\varphi})_j + b}{(f^{(n)} * \varphi * \widehat{\varphi})_j + b} \tag{10}$$

is known to converge asymptotically to the least-squares solution of the optimization problem under certain conditions on φ [42]. For our purposes, $\varphi(t) \geq 0$ and $\varphi(t = 0) > 0$ are sufficient conditions. Here, b is a free parameter chosen such that $(\Phi * \widehat{\varphi})_j + b > 0 \,\forall\, j$. The method is designed to extract a *non-negative* signal f from a measurement described by Eq. (7), where the *only* negative parts of the signal are due to noise. The effect of negative pulse amplitudes is explored in Sec. IV. We note that the standard deviation of the noise X, or alternatively the signal-to-noise ratio, plays no role in the iteration scheme.

Numerical testing reveals that the choice of the initial guess $f^{(0)}$ as well as the exact value of b may play a role in the rate of convergence but do not affect the result of the iteration given by Eq. (10) if b is small compared to the mean signal value. Consequently, we set the initial guess to a positive constant and the b-parameter to $b = 10^{-10} - b_{\min}$,
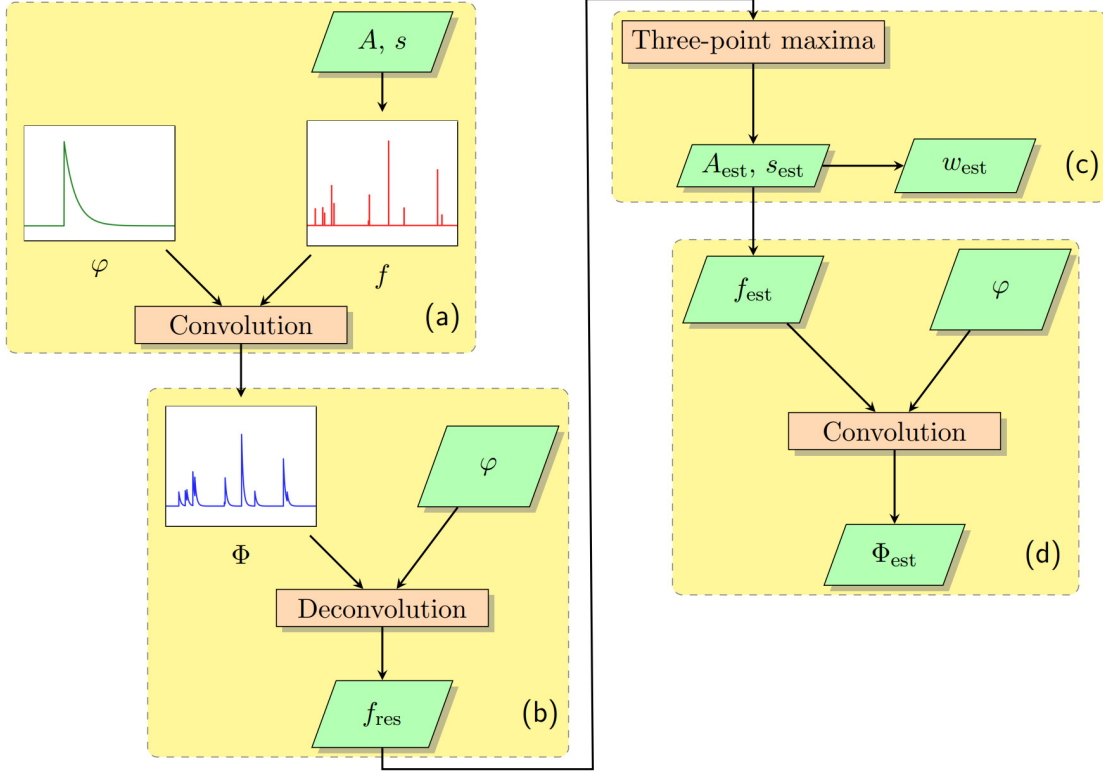
FIG. 1: Flow chart to show the procedure applied in this study of the deconvolution algorithm. The flow chart is detailed in the main text. For experimental data, the processes in boxes B, C and D would be applied. In this contribution, the focus is on the effect of C and D.

where $b_{\min} = \min[0, \min{(\Phi * \hat{\varphi})}]$. The small constant 10^{-10} is added to avoid issues with the division of numbers close to zero in the denominator in Eq. (10).

C. Extracting amplitudes and arrival times

The result of the deconvolution is the maximum-likelihood estimate of the forcing, denoted f_{res} where the subscript $\cdot_{\mathrm{res}}$ stands for 'result'. In this study, we are interested in estimating the amplitudes and arrival times that define the forcing f. For the realizations investigated in this contribution, the delta-train forcing is estimated as a series of sharply localized peaks. Therefore, a peak-finding algorithm must be applied to recover the pulse arrival times and amplitudes. We have chosen to employ a simple 3-point running maxima, so that a data point $f_{\mathrm{res},j}$ is classified as a peak if it is larger than both neighboring points:

$f_{\text{res},j} > f_{\text{res},j\pm1}$ for all $1 \le j < N - 1$. The estimated arrival time s_{est} is the location of the maxima and the estimated amplitude A_{est} is the value of the maxima.

Using the found maxima directly leads to far too many detected events for small ($\gamma < 1$) intermittency parameters (6 times the true number of events for $\gamma = 1/10$). This is due to numerical noise in sections of f_{res} without true pulses. To remove these spurious pulses, we introduce a small amplitude threshold equal to $10^{-2}\langle A \rangle$, which corresponds to the size of $|f_{\text{res}} - f|$ for $\gamma = 1/10$. The presence of this threshold does not influence the results in the absence of additive noise, as investigated below. In the presence of additive noise, a stricter amplitude threshold is placed on local maxima. This amplitude threshold is discussed further in Sec. V.

The entire reconstruction process is described by the flow diagram in Fig. 1. In box A, the synthetic signal Φ is generated. The waiting times w are randomly drawn until their sum exceeds the specified signal duration T. The arrival times s are computed from the waiting times, and a number of amplitudes A matching the arrival times are randomly drawn. The synthetically generated forcing f consists of the amplitudes A and the arrival times s, and convolving this forcing with the specified pulse function φ gives the synthetic signal Φ. In box B, we perform the ISRA specified in Eq. (10), where f_{res} is the result of the deconvolution. Here, it may be necessary to estimate φ, for example from the PSD as discussed in Sec. VI. In box C, A_{est}, s_{est} and w_{est} refer to the amplitudes, arrival times and waiting times extracted using the 3-point maxima method described above. Finally, in box D, the signal is reconstructed. Here, f_{est} refers to forcing which consists of A_{est} and s_{est}. Φ_{est} refers to the signal reconstructed by using the estimated amplitudes and arrival times. For the model realizations considered in this contribution, it will be shown that the process in box B in Fig. 1 captures f very well. Thus, the focus of this contribution is the prospects and limitations of C and D.

To check that the mean values of the estimated variables have converged, we employ a bootstrapping technique [53]. For a given data set X containing N samples, we first estimate the sample mean $\overline{X}$. Then we draw $3N/4$ random samples with replacement (that is, we are allowed to draw the same sample multiple times) and estimate the sample mean $\overline{X}_{3N/4}$. Repeating this procedure 100 times gives a data set containing 100 samples of $\overline{X}_{3N/4}$. Estimating the standard deviation of this data set gives a measure of how well $\overline{X}$ has converged. A large standard deviation would indicate that we have too few samples N,

10

and should repeat our estimate with a larger data set. We chose $3N/4$ samples instead of N for the random draw to err on the side of overestimating the standard deviation. Estimates for higher moments proceed analogously. In all cases, this produced standard deviations within 6% of the corresponding mean values, indicating well-converged statistics.

Since the parameter values (in particular γ) are varied over an order of magnitude, we report the bias in the estimate as the *ratio* between the estimated value and the true value instead of the more common *difference* between the estimated value and the true value. This lets us compare estimates for different input parameters directly. In this contribution, a deviation of 10% or more from the true value is considered significant and is indicated with *italic* numbers in the tables. This will be the basis for our discussion of the results.

III.   ESTIMATING AMPLITUDES AND WAITING TIMES

Two effects lead to loss of information of the true pulse arrivals in realizations of the process: (1) The point process $K(T)$ is continuous but as we are working with discrete time series, pulses closer than a sampling time cannot be separated. These will be counted as one pulse arrival with amplitude equal to the sum of their amplitudes, corrupting the resulting estimated amplitude distribution and number of pulses. (2) As the peaks of f_{est} may have finite widths, a peak-finding algorithm must be employed. A straightforward and permissive method, a 3-point maxima, compares each data point to its neighboring points and flags it as a peak if it is bigger than both neighbors. This method retains only the highest amplitude event if two or more pulses arrive closer than two sampling times to each other.

In App. A, it is shown that in the base case, the average number of 3-point maxima M in f as compared to the average number of pulses $\langle K \rangle$ is given approximately by

$$\frac{\langle M \rangle}{\langle K \rangle} \approx \frac{1 - \exp\left(-3\gamma\theta\right)}{3\gamma\theta}. \tag{11}$$

Here, $\theta = \triangle_t/\tau_{\mathrm{d}}$ is the normalized sampling time and $\gamma\theta = \triangle_t/\langle w \rangle$. In order to recover approximately 90% of the pulses using the 3-point maxima, $\gamma\theta < 0.075$ is necessary, while to recover approximately 95% of the pulses, $\gamma\theta < 0.035$ is necessary. This suggests an approximate threshold of $\gamma\theta \leq 1/20$ to recover most (above 90%) of the pulses. In this contribution, we have set $\triangle_t = 10^{-2}\tau_{\mathrm{d}}$, leading to an assumed approximate condition $\gamma < 5$ for reconstruction of amplitude and arrival time distributions within 10% variation in average

values. We will see that in general, $\gamma = 10$ gives too much overlap while in some cases, $\gamma \leq 1$ is required, which corresponds to approximately 98.5% pulse recovery. This threshold should be seen as a guideline, dependent on the precision and objectives required for application of the method.

The sampling and local maxima issues are illustrated in Fig. 2, which shows results from deconvolution for fixed θ and two different values of γ. Although the estimated forcing (the orange solid line) is a good approximation of the original forcing signal (the black dotted line), even for $\gamma = 10^2$, the same cannot be said of the estimated amplitudes (the orange triangles) when compared to the original amplitudes (the black triangles). In the case of $\gamma = 10$, the error is moderate and mainly caused by the 3-point maxima not identifying arrivals at neighboring points. The effect is more severe in the case $\gamma = 10^2$, where it is evident that f_{est} is given by a superposition of multiple arrivals closer than one time step of each other.

The corresponding original and reconstructed time series are presented in Fig. 3 for one-sided exponential pulses and an exponential pulse amplitude distribution. Here and in the following, $\widetilde{\Phi}$ denotes the normalized variable $\widetilde{\Phi} = \left(\Phi - \langle \Phi \rangle \right) / \Phi_{\text{rms}}$. For these high γ-values, the FPP resembles a normally distributed process. The reconstructed signal largely follows the overall path of the original signal but deviates in detail due to finite sampling rate and pulse overlap as described above. These differences may have a profound influence on the estimation of the amplitude and waiting time statistics, as investigated in detail below.

In Table I the first four moments as estimated from the reconstructed time series are presented for various values of the intermittency parameter. The estimated moments are normalized by the sample moments of the original time series. In all cases, one-sided exponential pulses with an exponential amplitude distribution are used. For $\gamma \leq 5$, there is at most 6% disagreement between the estimated and sample moments, within our allowed tolerance. For $\gamma \geq 10$, both the mean value and standard deviation are underestimated, with values below the 10% threshold. The underestimation is worse for larger intermittency parameters. Higher-order moments are wrongly estimated for $\gamma \geq 50$, as evidenced by the three values in the lower right-hand corner of the table. Thus, we conclude that Table I provides evidence for the condition $\gamma \theta \leq 1/20$ for accurate reconstruction.

The deviation in the first two moments may be explained by the following: Due to pulse overlap and the 3-point maxima, several true events are discounted. This leads to
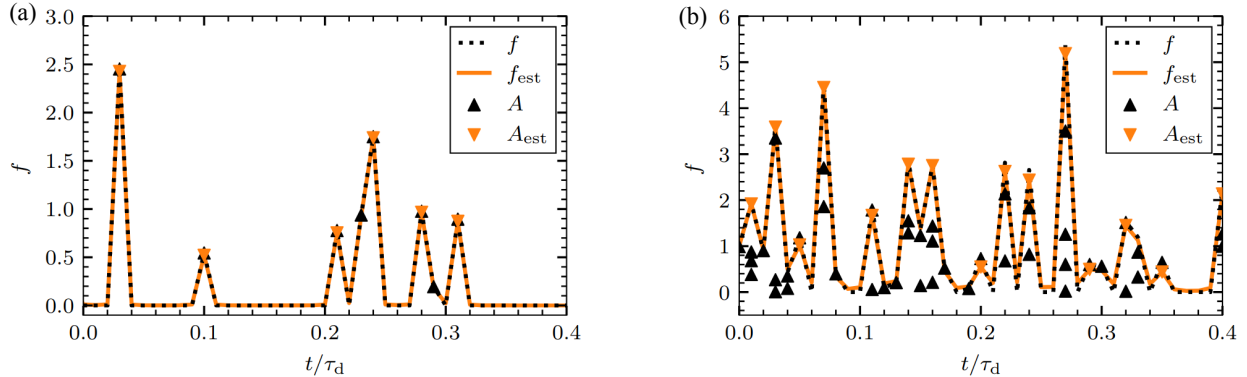
FIG. 2: Comparison between original forcing f and the estimated forcing f_{est} for two different intermittency parameters, (a) $\gamma = 10$ and (b) $\gamma = 10^2$, no noise $X = 0$, exponentially distributed pulse amplitudes and a normalized sampling time of $\triangle_t/\tau_{\text{d}} = 10^{-2}$. The exponential pulses are uncorrelated and have a uniform distribution of arrival times. The markers denote the original amplitudes A and the estimated amplitudes A_{est}.

an underestimation of the mean value and the standard deviation, more severely for higher intermittency parameters. As the discounted events are preferentially small, the small-amplitude variations are decreased, leading to a lower overall rms-level of the signal. By taking the square root of the second column and dividing it by the first column, it is seen that the relative fluctuation level, however, is robust. The deviations in the higher moments are less systematic, and a Monte-Carlo study should be carried out to put reasonable error bars on these values. The important conclusion from this table is that it supports the approximate condition $\gamma\theta \leq 1/20$ and therefore a time-consuming Monte-Carlo study for the largest intermittency parameters has not been performed here.

The effects of varying the sampling time are presented in Figs. 4 and 5. Here we keep $\gamma = 1$ fixed for various values of θ. In these figures, a realization of the FPP with normalized sampling time $\theta = 10^{-3}$ was downsampled by using only a portion of the data points in the time series. The downsampled signals were then deconvolved with a similarly downsampled pulse function, and the signals were reconstructed. In Fig. 4, the estimated forcing is presented. As expected, increasing the sampling time leads to fewer arrivals identified as not all pulses can be separated. The result is similar to the effect of keeping θ fixed and increasing γ, as expected from the theoretical prediction presented in App. A.
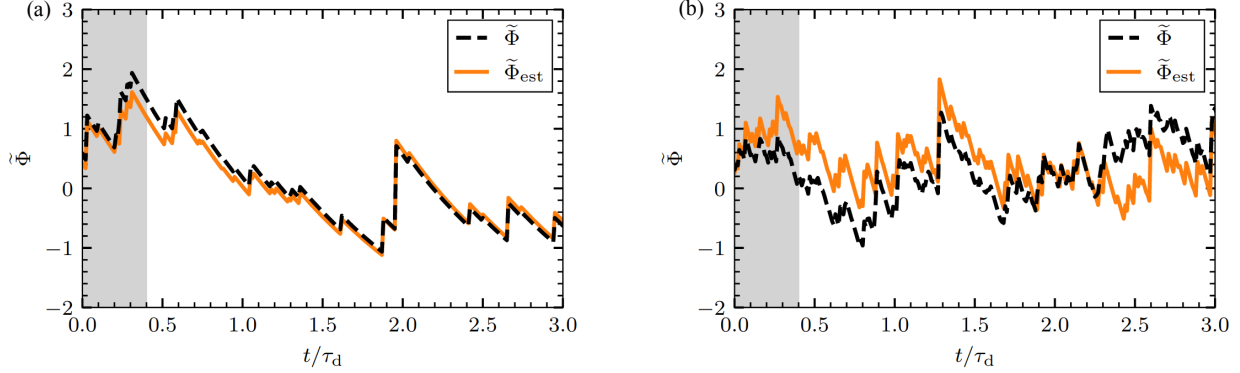
13

FIG. 3: Comparison between normalized original time series $\widetilde{\Phi}(t)$ and the normalized reconstructed time series $\widetilde{\Phi}_{\text{est}}(t)$ for two different intermittency parameters, (a) $\gamma = 10$ and (b) $\gamma = 10^2$, using amplitudes and arrival times estimated with the deconvolution method. The forcing for the first 0.4 normalized time units is presented in Fig. 2 and is gray shaded here.

γ	Moments			
	$\overline{\Phi}_{\text{est}}/\overline{\Phi}$	$\Phi^2_{\text{rms,est}}/\Phi^2_{\text{rms}}$	$S_{\Phi,\text{est}}/S_\Phi$	$(F_{\Phi,\text{est}} - 3)/(F_\Phi - 3)$
10^{-1}	1.00	0.99	1.00	1.00
1	0.99	0.97	0.99	0.99
5	0.97	0.94	0.99	0.97
10	0.94	*0.78*	0.98	0.94
50	*0.78*	*0.60*	1.02	*0.58*
10^2	*0.68*	*0.50*	*1.13*	*1.32*

TABLE I: Table showing the sample moments of the reconstructed time series for various intermittency parameters γ as compared to the sample moments of the model realization. 10^5 iterations were used for parameters $\theta = 10^{-2}$, $T/\tau_\text{d} = 10^4$ and no noise was added. The theoretical expressions for the moments are given directly below Eq. (3).

In Fig. 5 the reconstructed signals are compared to the downsampled original time series without noise, $X = 0$. Note that the downsampled time series with $\theta = 10^{-1}$ is here visually similar to the original time series with $\theta = 10^{-3}$. The reconstruction is reasonable, even for $\theta = 1$, when compared to the downsampled signal. However, while the amplitudes and
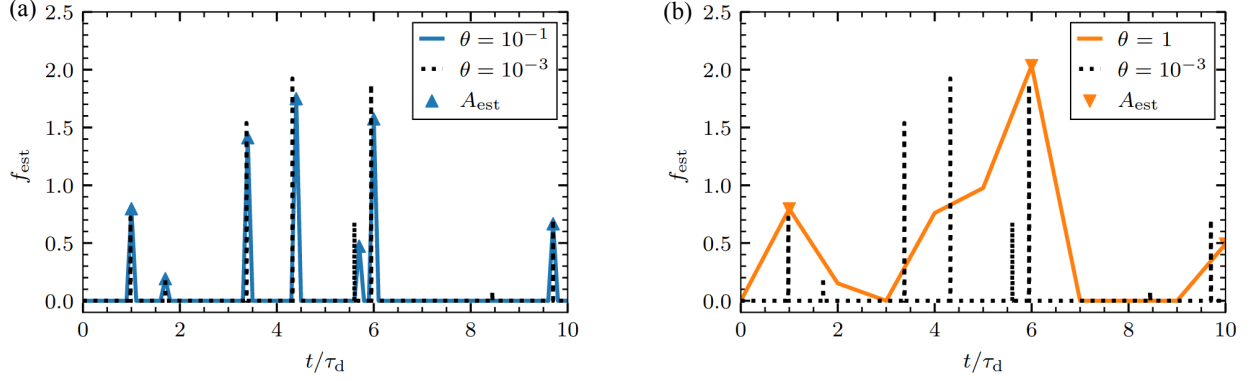
14

FIG. 4: Estimated forcing f_{est} (solid line) and estimated amplitudes (markers) for $\gamma = 1$ and two different normalized sampling times, (a) $\theta = 10^{-1}$ and (b) $\theta = 1$. The original forcing (black dotted line) with $\theta = 10^{-3}$ is shown for comparison.
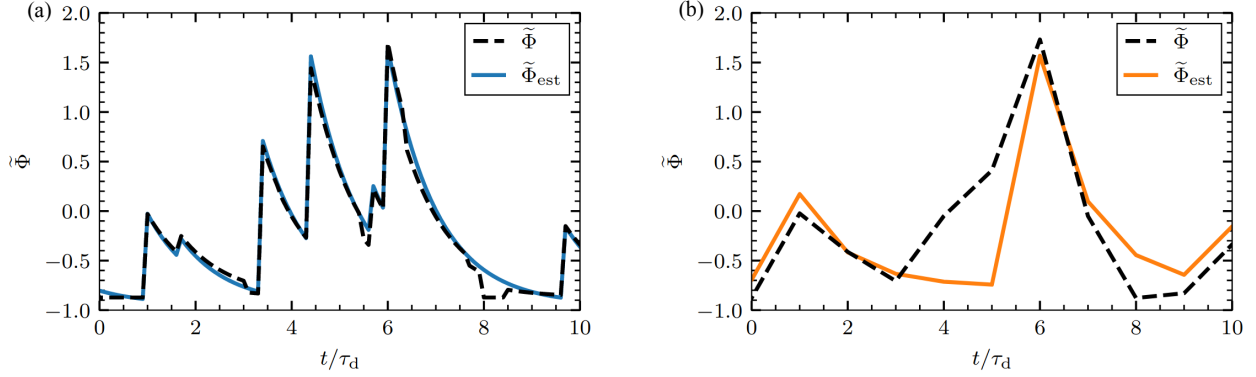

FIG. 5: Comparison between reconstructed time series using the estimated amplitudes and arrival times (solid line) and downsampled original signals (dashed line) for $\gamma = 1$ and two different normalized sampling times, (a) $\theta = 10^{-1}$ and (b) $\theta = 1$.

arrival times might be reasonably estimated in the case $\theta = 10^{-1}$, this is obviously not possible in the case $\theta = 1$, as expected.

γ	10^{-1}	1	5	10	50
Symbol	▼	◆	●	■	▲

FIG. 6: List of plot symbols for various intermittency parameters used for estimating pulse amplitude and waiting time statistics.

IV.   AMPLITUDE AND WAITING TIME DISTRIBUTIONS

In this section, we investigate the ability of the deconvolution algorithm to reconstruct the amplitude and waiting time distributions for the case where the pulse function is known. In all the following cases we use $\theta = 10^{-2}$, $T/\tau_{\mathrm{d}} = 10^4$ and we have not added any noise term. For this value of the sampling time, the condition $\gamma\theta \leq 1/20$ becomes $\gamma \leq 5$. The deconvolution ran for 10^5 iterations, leading to convergence for $\gamma < 10$ and marginal convergence for $\gamma \geq 10$.

A.   Base case

First, we consider the effects of varying γ in the base case on the estimated amplitude and waiting time distributions from the deconvolved forcing. The plot symbols used for various intermittency parameters are presented in Fig. 6. The resulting distributions are presented in Fig. 7 and the estimated parameters are presented in Table II. Here, the found number of maxima, the estimated average amplitudes, and the estimated average waiting times are compared to the theoretical number of maxima in Eq. (A6) and the theoretical mean values of the process. In the figures, the distributions are well estimated for $\gamma \leq 10$, while the mean values are within 10% of their theoretic values for $\gamma \leq 5$. The expected number of maxima is well estimated by Eq. (A6).

For $\gamma = 50$, the mean amplitude and waiting time are significantly overestimated compared to the other cases. For the amplitudes, this effect is mainly due to multiple events being added as a result of being closer than a sampling time. The resulting distribution has a shallower slope and is more concave when compared to the original distribution. For the waiting times, a significant number of true waiting times are below one sampling time (the probability $\Pr[w < \theta] \approx 0.39$ for $\gamma = 50$). As the smallest waiting time resolvable by the deconvolution method is 2θ, this introduces a cutoff in the estimated distribution which
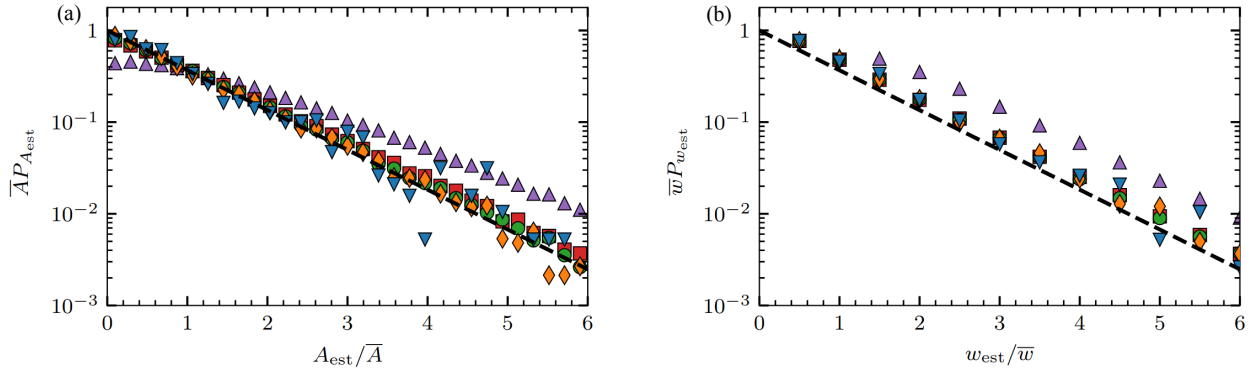
FIG. 7: Probability distributions of (a) estimated amplitudes and (b) estimated waiting times for various intermittency parameters in the base case with exponentially distributed pulse amplitudes and waiting times. In both cases, the distributions are normalized by the original sample mean values. The plot symbols are defined in Fig. 6.

increases the mean value of the estimated waiting times. Moreover, the waiting times only take values that are low-integer multiples of the sampling time. For consistency between different values of the intermittency parameter, we use the same bins for all distributions shown in these figures, chosen such that each bin contains a single integer multiple of the sampling time in the $\gamma = 50$-case. Although it is elevated, due to discounting the shortest waiting times, the distribution for $\gamma = 50$ decays with the same slope as the original distribution.

In the following, we will investigate the robustness of the method to non-exponential amplitude and waiting time distributions. We have chosen to test the Rayleigh distribution due to its Gaussian tail, the Pareto distribution as an example of a much broader distribution than the exponential, the uniform distribution for its discontinuous cutoff towards large values, and the extreme case of the degenerate distribution. Definitions of these distributions are presented in App. B.

B. Amplitude distributions

In Fig. 8 we present the amplitude distributions estimated from the deconvolution procedure for various pulse amplitude distributions and exponentially distributed waiting times. In all cases, the different symbols denote the estimated distributions corresponding to the

17

	Estimated averages	γ				
		10^{-1}	1	5	10	50
$A \sim$ Exponential	$M/\langle M \rangle$	1.02	1.01	0.98	0.98	0.95
	$\overline{A}_{\text{est}}/\overline{A}$	1.01	1.02	1.06	*1.11*	*1.58*
$w \sim$ Exponential	$\overline{w}_{\text{est}}/\overline{w}$	1.01	1.03	*1.10*	*1.19*	*2.03*

TABLE II: Ratio of (top) number of maxima in the forcing estimated from the deconvolution to the theoretical number of maxima, (middle) mean estimated amplitudes to the original sample mean and (bottom) mean estimated waiting times to the original sample waiting times. Results are from the base case and for various intermittency parameters. The corresponding distributions are presented in Fig. 7.

intermittency parameter values given in Fig. 6. The black dashed line gives the analytical distributions. Note that the Rayleigh distribution is presented with a semi-logarithmic scale, the Pareto distribution is presented with a double-logarithmic scale and the uniform and degenerate distributions are presented with a linear scale, with a semi-logarithmic scale in the inset. Table III aggregates mean values for the distributions in Fig. 8.

We see that all distributions are well estimated for $\gamma \leq 1$. The estimates for Rayleigh distributed amplitudes are visibly affected by pulse overlap for $\gamma \geq 5$, showing elevated and close to exponential tails for large amplitudes. Pareto distributed amplitudes are reliably identified for all tested intermittency parameters, likely due to the large range of probable values leading to little distortion due to pulse overlap. For uniformly distributed amplitudes, it is clear from the inset of Fig. 8(c) that while the implications of pulse overlap are visible even for $\gamma = 1$, there is a jump of order the 10^{-2} from the originally allowed values to the larger values. As a consequence, the uniform distribution is well estimated and within the variation around the straight line expected for the uniform distribution. For a degenerate distribution of pulse amplitudes, the main contribution to the estimated amplitude distribution is the expected delta peak, with corrections at higher integer values, as seen in the inset. Only for $\gamma = 50$ is there a significant contribution of normalized amplitudes larger than two and for non-integer amplitudes.

In Table III, the number of found maxima is very close to the expected number corrected
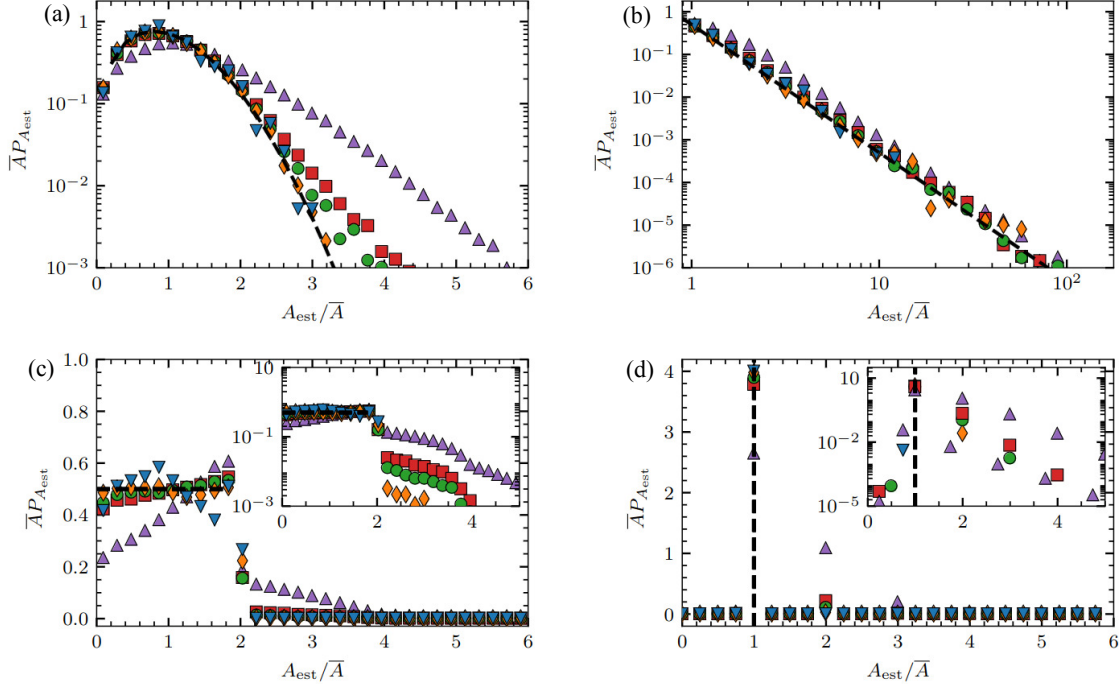
FIG. 8: Probability distributions of estimated pulse amplitudes normalized by the sample mean amplitude for various intermittency parameters and distributions of pulse amplitudes in realizations of the process. (a) $A \sim$ Rayleigh, (b) $A \sim$ Pareto(3), (c) $A \sim \mathcal{U}(0, 2)$ and (d) $A \sim$ degenerate. The insets in (c) and (d) show the distributions with semi-logarithmic scaling. In all cases, exponentially distributed waiting times were used. The black dashed lines represent the analytical amplitude distribution of the various realizations. The plot symbols are defined in Fig. 6.

by effects of discretization and taking the 3-point maxima, as discussed in Sec. III. This distortion is reflected in the deviation of the average estimated waiting time from the theoretical waiting time. There is only a small effect on the average amplitudes for $\gamma \leq 10$, irrespective of amplitude distribution.

C. Waiting time distributions

Realizations of the process have also been made for various pulse waiting time distributions and intermittency parameters, with an exponential amplitude distribution in all cases. The estimated waiting time distributions from deconvolution of these realizations are presented in Fig. 9 and mean amplitudes and waiting times as well as the number of maxima

19

P_A	Estimated averages	γ				
		10^{-1}	1	5	10	50
Rayleigh	$M/\langle M \rangle$	1.03	1.02	1.00	1.00	0.99
	$\overline{A}_{\text{est}}/\overline{A}$	1.00	1.01	1.03	1.07	*1.41*
	$\overline{w}_{\text{est}}/\overline{w}$	1.00	1.02	1.08	*1.16*	*1.95*
Pareto(3)	$M/\langle M \rangle$	1.03	1.02	1.00	1.00	0.99
	$\overline{A}_{\text{est}}/\overline{A}$	1.00	1.01	1.03	1.07	*1.45*
	$\overline{w}_{\text{est}}/\overline{w}$	1.00	1.02	1.08	*1.16*	*1.95*
$\mathcal{U}(0,2)$	$M/\langle M \rangle$	1.02	1.01	0.99	0.99	0.97
	$\overline{A}_{\text{est}}/\overline{A}$	1.00	1.01	1.04	1.08	*1.46*
	$\overline{w}_{\text{est}}/\overline{w}$	1.01	1.03	1.09	*1.17*	*1.99*
degenerate	$M/\langle M \rangle$	1.03	1.02	1.00	1.00	0.96
	$\overline{A}_{\text{est}}/\overline{A}$	1.00	1.00	1.02	1.04	*1.32*
	$\overline{w}_{\text{est}}/\overline{w}$	1.00	1.02	1.08	*1.16*	*2.01*

TABLE III: In each row: Ratio of (top) found number of maxima in the forcing estimated from the deconvolution to the theoretical number of maxima, (middle) mean estimated amplitudes to the original sample mean and (bottom) mean estimated waiting times to the original sample waiting times. Results are for the various amplitude distributions used in Fig. 8, exponentially distributed waiting times and for various intermittency parameters.

in the forcing are presented in Table IV. For $\gamma \leq 10$, all distributions show agreement with the true distribution for low ($w_{\text{est}} < 3\overline{w}$) waiting times. The mean amplitudes and waiting times are well estimated for $\gamma \leq 10$, the largest deviation being a factor 1.11 increase for uniformly distributed waiting times and $\gamma = 10$.

Rayleigh distributed waiting times display an elevated, exponential-like tail towards large values. Despite this, the mean values are better estimated than in the base case, possibly due to the clear peak of the Rayleigh distribution, ensuring most events are well separated. The distortion to the Rayleigh distribution for $\gamma = 50$ follows the pattern seen for exponentially distributed waiting times, as discussed in the introduction to Sec. IV. Again, we have

chosen bins such that each bin contains one integer multiple of the sampling time in the case $\gamma = 50$. We see that the tail of the distribution is elevated with respect to the original Rayleigh distribution and with respect to the lower γ cases. The inflated mean values in Table IV are due to the loss of waiting times shorter than two time steps and corresponding enhanced pulse overlap.

Pareto distributed waiting times are identified in all cases, and the corresponding mean amplitudes and waiting times are almost perfectly estimated. This is due to pulse overlap not being a factor, even at these high intermittency levels: For $w \sim \mathrm{Pareto}(3)$, there is a cutoff $w/\tau_\mathrm{d} \geq 1/2\gamma$, see App. B. With $\theta = 10^{-2}$, $\gamma = 50$ corresponds to a minimal waiting time of 2 data points. There is an arrival at least every second data point, and this can just about be resolved by the 3-point maxima. A higher γ would result in an artificial cutoff in the same manner as for the Rayleigh distributed waiting times.

For the degenerate distribution, $\theta = 10^{-2}$ and $\gamma = 50$ corresponds to a (degenerate) waiting time of two data points, again just at the edge of resolvability. Here, the (small) probability of larger waiting times is likely due to very small amplitudes not being picked up by the deconvolution or being removed by the 10^{-2} numerical noise threshold. As seen in Table IV, this has some effect on the average waiting time, but hardly any effect on the average amplitudes.

For the uniform distribution, the deconvolution cutoff $w/\overline{w} \approx w/\langle w \rangle \geq 2\Delta_t/\langle w \rangle = 2\theta\gamma$. In the $\gamma = 50$-case, this condition is $w/\overline{w} \geq 1$ and is evident in Fig. 9(c): the probability of waiting times below this value is practically zero, and there is a corresponding positive probability for values larger than the original maximal value. The effect on the mean amplitude and waiting time is comparable to the effect in the case of Rayleigh distributed waiting times.

Comparing Table IV to Table II we see that the theoretical $\langle M \rangle$ is only valid for exponentially distributed waiting times, although the uniformly distributed waiting times are within the 10% margin for all γ. The most severe deviation is for Pareto distributed waiting times, where the number of events found is wrong by approximately a factor $1/\gamma$. This error is to be expected, as the exponential shape of the waiting time distribution, non-zero probability of evens closer than two time steps and the memory-less property of the Poisson process were central to the calculations in App. A. For the non-exponential waiting time distributions, at least two of these assumptions are false.
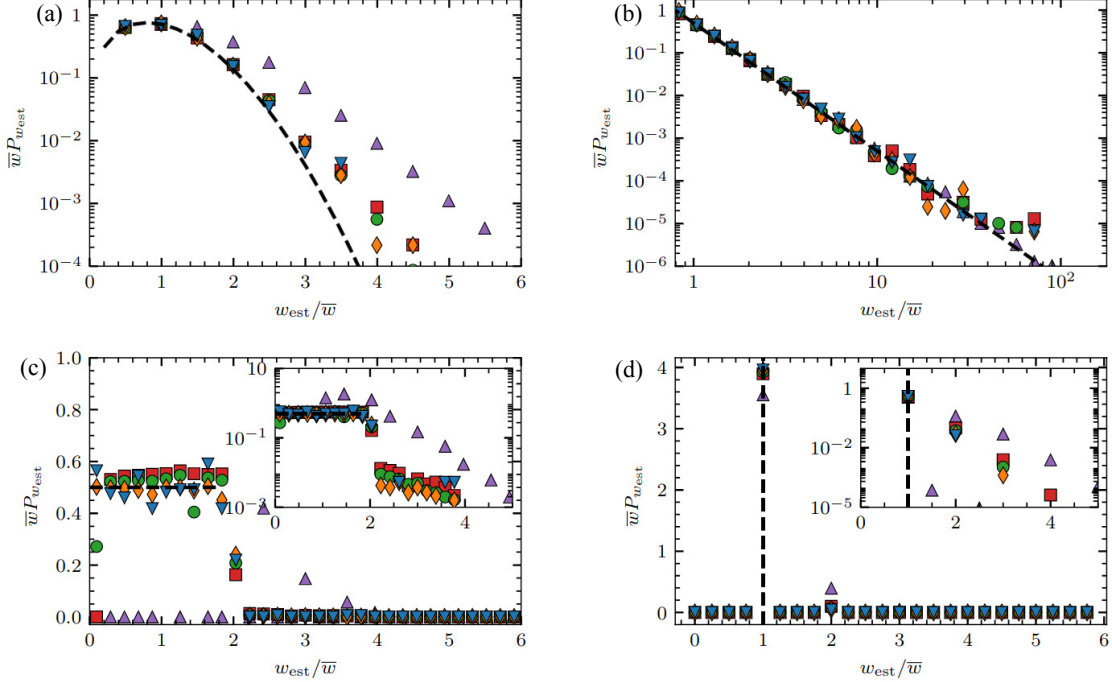
FIG. 9: Probability distributions of estimated waiting times normalized by the sample mean waiting time for various intermittency parameters and distributions of waiting times in realizations of the process. (a): $w \sim$ Rayleigh, (b): $w \sim$ Pareto(3), (c): $w \sim \mathcal{U}(0,2)$ and (d): $w \sim$ degenerate. The insets in (c) and (d) show the distributions with semi-logarithmic scaling. In all cases, exponentially distributed amplitudes were used. The black dashed lines represent the analytical waiting time distribution of the various realizations. The plot symbols are defined in Table 6.

These results may indicate that the deconvolution procedure distorts amplitudes more than waiting times. Even if a cluster of arrivals is counted as a single arrival by the estimation algorithm, the time between such clusters is not distorted much by the algorithm, and the tail in the waiting time distribution should be well estimated, if there is no cutoff for large waiting times. The most significant source of error is the loss of events closer than two time steps.

D. Negative pulse amplitudes

All the amplitude distributions investigated above were positive definite. Now, we consider realizations with zero mean, normally distributed amplitudes to see if they are repro-

22

P_w	Estimated averages	γ				
		10^{-1}	1	5	10	50
Rayleigh	$M/\langle M \rangle$	1.01	1.02	1.05	*1.11*	*1.21*
	$\overline{A}_\mathrm{est}/\overline{A}$	1.01	1.01	1.01	1.02	*1.28*
	$\overline{w}_\mathrm{est}/\overline{w}$	1.01	1.01	1.02	1.05	*1.60*
Pareto(3)	$M/\langle M \rangle$	*9.76*	1.01	*0.21*	*0.11*	*0.04*
	$\overline{A}_\mathrm{est}/\overline{A}$	1.01	1.01	1.01	1.01	1.01
	$\overline{w}_\mathrm{est}/\overline{w}$	1.01	1.01	1.01	1.01	1.01
$\mathcal{U}(0,2)$	$M/\langle M \rangle$	1.01	1.01	1.02	1.05	1.07
	$\overline{A}_\mathrm{est}/\overline{A}$	1.01	1.02	1.03	1.06	*1.33*
	$\overline{w}_\mathrm{est}/\overline{w}$	1.01	1.02	1.06	*1.11*	*1.64*
degenerate	$M/\langle M \rangle$	0.99	1.00	1.06	*1.13*	*1.72*
	$\overline{A}_\mathrm{est}/\overline{A}$	1.01	1.01	1.01	1.01	1.03
	$\overline{w}_\mathrm{est}/\overline{w}$	1.01	1.01	1.02	1.03	*1.13*

TABLE IV: Each row shows the ratio of (top) found number of maxima in the forcing estimated from the deconvolution to the theoretical number of maxima, (middle) mean estimated amplitudes to the original sample mean and (bottom) mean estimated waiting times to the original sample waiting times. Results are for the various waiting time distributions used in Fig. 9, exponentially distributed amplitudes and for various intermittency parameters.

ducible, and to investigate the robustness of the deconvolution to negative pulse amplitudes in the process. Symmetric Laplace distributed amplitudes were also tested with similar results but are not presented here. As the deconvolution algorithm works under the constraint that the forcing is non-negative, we first straightforwardly estimate the amplitudes and arrival times. This gives estimates of the positive amplitudes with corresponding arrival times. Then we multiply the signal by -1 and redo the deconvolution, giving estimates of the negative amplitudes with corresponding arrival times. In Fig. 10, the results are presented for normally distributed amplitudes. In Fig. 10(a), the resulting amplitude distri-
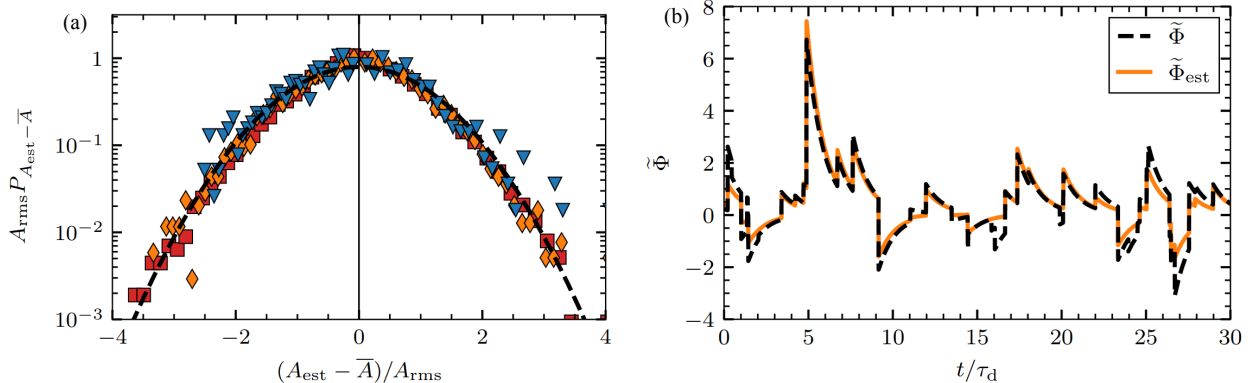
FIG. 10: Reconstruction of a realization with normally distributed pulse amplitudes. (a) Probability distribution function of the amplitudes of found events from the deconvolution. Refer to Fig. 6 for the legend. The analytical distribution (dashed line) is shown for reference, whereas the vertical solid lines show the separation between the negative found events and positive found events. (b) Reconstructed time series (solid line) using estimated amplitudes and arrival times for $\gamma = 1$. The original time series (dashed line) is shown for comparison.

butions are presented, where the vertical black line indicates the separation between positive and negative amplitudes. The overall shape of the normal distribution is well recovered. In Fig. 10(b), an example of the reconstructed signal is compared to the original signal. It is clear that while large pulse amplitudes, both positive and negative, are identified, the method struggles for small amplitudes. Adding positive and negative pulses in quick succession leads to a signal shape which cannot be recognized by the deconvolution as a sum of just positive or just negative amplitudes, and this effect is more severe for small amplitudes.

E.   Conclusion

In this most favorable case of a known pulse function and no noise added, the deconvolution method performs well overall, only limited by pulse overlap, as expected from the theory. Different distributions of pulse amplitudes and waiting times are well reproduced and mean amplitudes and waiting times are reliably estimated with an error of less than 10% for intermittent signals, $\gamma \leq 5$, lending credence to the approximate condition for pulse recovery $\gamma\theta \leq 1/20$. Even signals with both positive and negative pulse amplitudes allow

24
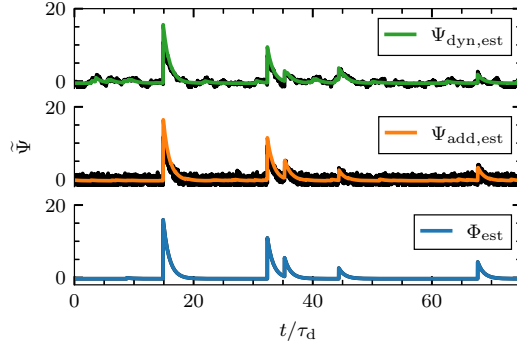
reconstruction of large amplitude events.

FIG. 11: Original time series (black lines) and reconstructed time series shown by the colored (gray) lines from the deconvolution for no noise (lower panel), additive noise (middle panel) and dynamical noise (upper panel). The intermittency parameter was set to $\gamma = 10^{-1}$, while the noise to signal ratio was set to $\epsilon = 1$.

V. EFFECTS OF NOISE

In this section, we investigate the effects of adding normally distributed noise to the FPP in two different ways. *Additive* noise consists of uncorrelated noise, while *dynamical* noise has the same correlation function as the base case FPP, achieved by convolving uncorrelated noise with the pulse function. Some effects of these forms of noise, including methods for estimating parameters, are discussed further in Refs. 50 and 52. For both noise types, we use the base case FPP with uncorrelated, one-sided exponential pulses with exponentially distributed amplitudes. The noise to signal ratio is defined as $\epsilon = X_{\mathrm{rms}}^2/\Phi_{\mathrm{rms}}^2$, where X_{rms} is the standard deviation of the noise process. Here and in the following, X_{add} refers to additive noise, X_{dyn} refers to dynamic noise and $\Psi = \Phi + X$ refers to the process with noise.

In Fig. 11 we present excerpts of the original and reconstructed time series for both types of noise and the case with no noise. In all cases, the same realization of the FPP was used. It is evident that additive noise, at least for the given noise variance and FPP intermittency, does not lead to major distortions in the reconstruction. This is expected, as the deconvolution algorithm takes normally distributed, additive noise into account and does not depend on the noise level. In Fig. 11 a few small, spurious events can be seen, likely where the noise by chance approximately reproduces the pulse shape. For dynamical noise, we see more significant spurious events as the algorithm can no longer reliably separate the noise from the small amplitude pulses in the signal. This is seen in Fig. 12, where estimated

26

amplitude and waiting time distributions are presented, using only the small amplitude threshold discussed in Sec. II C to avoid spurious arrivals. It is clear that many noise events are identified as pulses, and that this also influences the waiting time distribution. In all cases, the tail of the amplitude distribution follows the expected exponential decay. For low γ and ϵ, the effect of noise is largely concentrated in excessively many small-amplitude events which is reflected in the sharper decay of the waiting time distributions as compared to the original exponential distribution. For high γ and ϵ, the estimated amplitudes and waiting times are moderately affected, and in particular, dynamical noise has little effect on these distributions. In the following, we will discuss how to improve the results by removing the events with the smallest amplitudes.

Based on the exponential pulse amplitude and normal noise distributions, we have that $X_{\mathrm{rms}} = \sqrt{\gamma\epsilon}\langle A \rangle$, so we introduce a threshold where we reject all events with amplitude less than this value. From realizations of the process, these parameters may be estimated from the moments, probability distribution or characteristic function as described in Refs. 50 and 52. To simplify the analysis, we take γ, ϵ and $\langle A \rangle$ as given for setting the threshold level. In Figs. 13 and 14 we present estimated amplitude and waiting time distributions after applying the amplitude criterion $A_{\mathrm{est}} > \sqrt{\gamma\epsilon}\langle A \rangle$, discarding all events with smaller amplitudes. Note that for these figures, we normalize by the sample mean of the estimated values instead of the sample mean of the original values to highlight the similarity in distribution. The result is that we have good agreement in distribution in all cases of pulse overlap and noise levels investigated. Removing the arrivals due to noise also realigns the waiting time distributions to the expected exponential.

In Tabs. V and VI the estimated mean amplitudes and waiting times for all cases presented in Figs. 13 and 14 are presented. To accurately assess mean values of amplitudes and waiting times, the threshold must be taken into account. The mean value of the truncated amplitudes is, assuming exponentially distributed amplitudes, just the threshold subtracted from the sample mean value of the estimated amplitudes. Therefore, we report $\left(\overline{A}_{\mathrm{est}} - \sqrt{\gamma\epsilon}\,\langle A \rangle \right)/\overline{A}$. For waiting times, pulses are rejected if they have an amplitude $A < \sqrt{\gamma\epsilon}\langle A \rangle$, thus the number of found pulses after thresholding is reduced by a factor $1 - \Pr[A < \sqrt{\gamma\epsilon}\langle A \rangle] = \exp(-\sqrt{\gamma\epsilon})$. Here, the last equality holds for exponentially distributed amplitudes. This in turn implies that the estimated waiting time is prolonged and should be multiplied by this same factor, leading us to report $\exp(-\sqrt{\gamma\epsilon})\overline{w}_{\mathrm{est}}/\overline{w}$. Note that this argument relies on

an assumption of no pulse overlap and will be less accurate for intermittency parameters of order unity or larger.

In the case of the amplitudes, presented in Table V, we recover the mean value well when we correct for the threshold used, although additive noise leads to underestimation of the amplitudes while dynamic noise leads to overestimation of amplitudes in most cases investigated. The estimated waiting times, presented in Table VI, are in general much more affected by the noise than the amplitude distribution, reaching close to twice the original sample mean value. In both cases, it appears like moderate pulse overlap, $\gamma = 1$, is more affected by noise than either low or high pulse overlap. It may be that for low pulse overlap, noise is effectively filtered out, while for high degrees of pulse overlap, the threshold is so restrictive (see below) that only the largest amplitude pulses are identified. We also recall that $\overline{w}_{\mathrm{est}}/\overline{w} = 1.19$ is expected for $\gamma = 10$. Thus, the errors due to noise dominate for low and moderate γ, while for high γ, the errors due to pulse overlap are more significant.

In Fig. 15, we present the reconstructed time series from the deconvolution in the most extreme case considered, $\gamma = 10$ and $\epsilon = 1$. The original forcing (without noise) is compared to the forcing estimated from the signal with additive noise (orange dotted lines) and dynamical noise (green dashed lines) in Fig. 15(a). The horizontal black dashed line gives the amplitude threshold $\sqrt{\gamma\epsilon}$. It is clear that many small amplitude pulses are rejected as a result of the thresholding. Also note that additive noise appears to more severely affect large amplitudes than dynamical noise. In particular, note the large, spurious peak at normalized time 6.6. The reconstructed signals using the estimated forcing are compared to the original signal without noise in Fig. 15(b). It is clear that since only the largest amplitude pulses are identified, the method fails to reconstruct realizations of the process.

We have shown that the deconvolution recovers the forcing well in the case of normally distributed noise. To estimate amplitudes and arrival times, an amplitude threshold must be introduced. For high degrees of pulse overlap, this threshold significantly affects signal reconstruction, but by properly taking the threshold into account, we may recover the tails of the amplitude and waiting time distributions as well as estimating their mean values. The mean amplitude is accurately recovered, while the mean of the estimated waiting times is within a factor two of the original sample waiting time, even for severe noise levels.
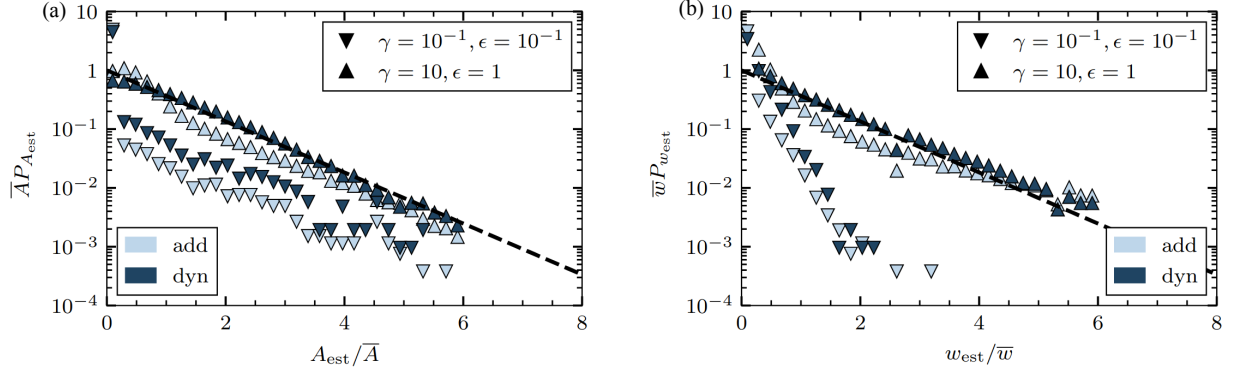
FIG. 12: Probability distributions of (a) estimated amplitudes and (b) estimated waiting times without thresholding. The black dashed line gives an exponential distribution. The symbols represent the values of γ and ϵ used, whereas the colors correspond to the noise type.
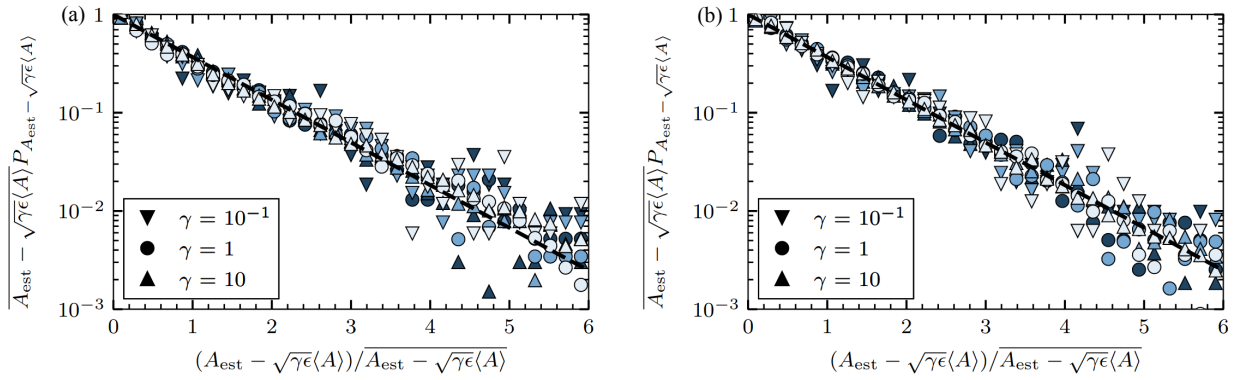

FIG. 13: Probability distributions of estimated amplitudes with thresholding in the presence of (a) additive noise and (b) dynamical noise. The colored markers represents the different noise to signal ratios ϵ, corresponding to different intermittency parameters γ represented by the shape of the markers. Light blue (light gray) refers to $\epsilon = 10^{-1}$, medium blue (medium gray) refers to $\epsilon = 1/2$ and dark blue (dark gray) represents $\epsilon = 1$. The black dashed line is an exponential distribution.

VI.   ESTIMATING THE PULSE DURATION

For any given measurement time series, the pulse function and duration may not be known and so must be estimated in order to apply the deconvolution method. For pulses with fixed duration arriving in accordance to a Poisson process, the pulse function can be obtained from the frequency PSD [51]. Alternatively, the conditionally averaging method
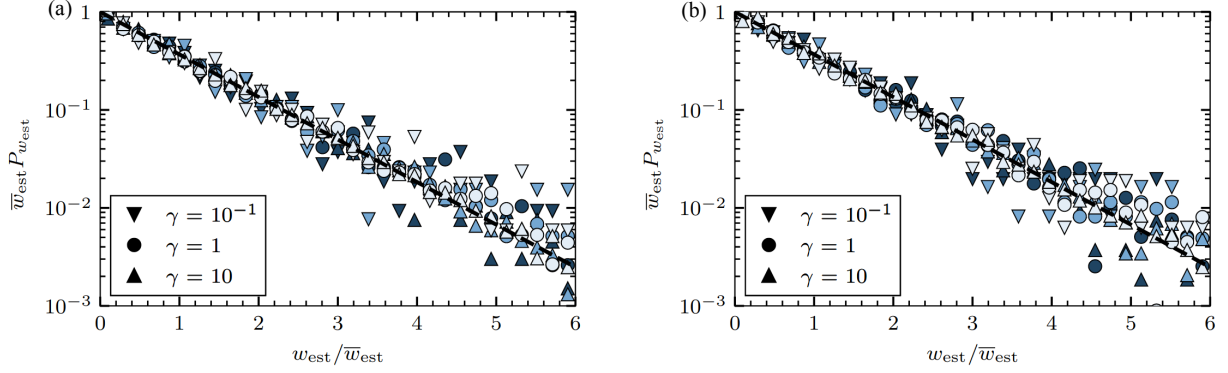
29

FIG. 14: Probability distributions of estimated waiting times with thresholding in the presence of (a) additive noise and (b) dynamical noise. The colored markers represents the different noise to signal ratios ϵ, corresponding to different intermittency parameters γ represented by the shape of the markers. Light blue (light gray) refers to $\epsilon = 10^{-1}$, medium blue (medium gray) refers to $\epsilon = 1/2$ and dark blue (dark gray) represents $\epsilon = 1$. The black dashed line is an exponential distribution.

	ϵ	γ		
		10^{-1}	1	10
Additive	10^{-1}	0.90	0.90	*0.89*
	$1/2$	0.90	*0.87*	*0.89*
	1	0.92	*0.84*	0.94
Dynamic	10^{-1}	1.01	1.03	1.07
	$1/2$	1.00	1.02	1.05
	1	0.98	1.02	1.04

TABLE V: Estimated average amplitudes in the case of amplitude noise thresholding corrected to $\left(\overline{A}_{\text{est}} - \sqrt{\gamma \epsilon} \langle A \rangle \right) / \overline{A}$ corresponding to Fig. 13.

for large-amplitude events may be used in the case of weak pulse overlap [22, 39, 54].

In this section, we investigate how different pulse duration and waiting time distributions distort the estimated pulse function and thereby the estimated amplitude and waiting time distributions. We will estimate the average duration from the frequency power spectra of the process, which, in contrast to the auto-correlation function, are very robust to a distribution

30

	ϵ	γ		
		10^{-1}	1	10
	10^{-1}	1.02	*1.24*	*1.36*
Additive	$1/2$	*1.17*	*1.63*	*1.36*
	1	*1.32*	*1.89*	*1.22*
	10^{-1}	*1.12*	*1.25*	*1.25*
Dynamic	$1/2$	*1.25*	*1.55*	*1.41*
	1	*1.37*	*1.82*	*1.52*

TABLE VI: Estimated average waiting times in the case of amplitude noise thresholding corrected to $\exp(-\sqrt{\gamma\epsilon})\overline{w}_{\text{est}}/\overline{w}$ corresponding to Fig. 14.
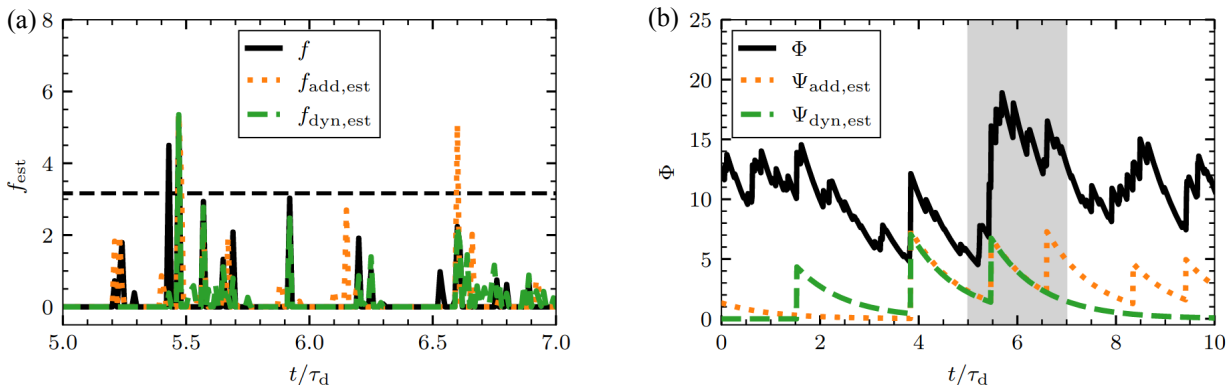

FIG. 15: (a) Excerpt of the original forcing (black solid line) with $\gamma = 10$ and $\epsilon = 1$, compared to the estimated forcing using additive noise (orange dotted line) and dynamical noise (green dashed line) where the noise threshold is shown (horizontal black dashed line). The reconstructed time series (b) is shown using the estimated amplitudes and arrival times from the 3-point maxima. The shaded background corresponds to the time axis shown on the left.

of pulse durations [51]. We restrict ourselves to the one-sided exponential pulse function, as it has a clearly identifiable Lorentzian power spectrum. We further restrict ourselves to deviations from the base case which produce reasonable Lorentzian-like power spectra. This means that degenerate and uniform waiting time distributions are not considered, as they produce pronounced peaks in the spectra. Pareto distributed waiting times produce
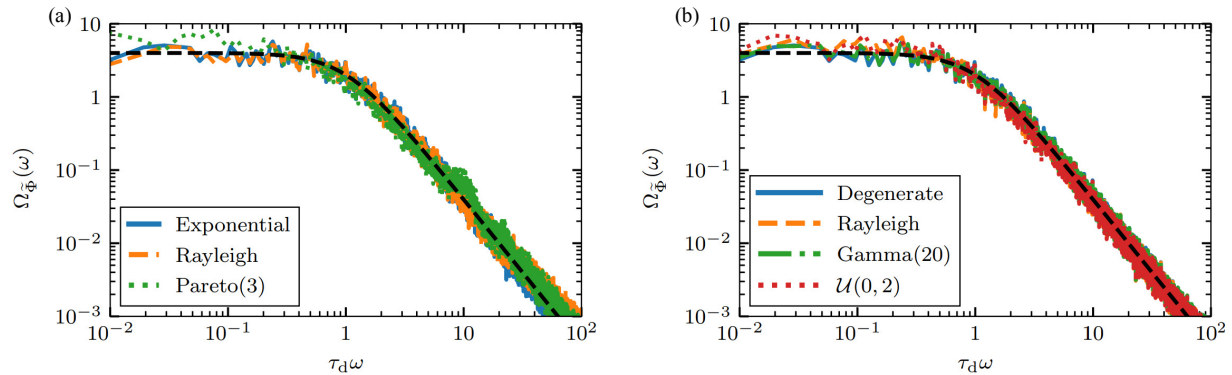
FIG. 16: Power spectral densities of normalized original time series for $\gamma = 10$ with (a) fixed pulse duration and different waiting time distributions and (b) exponential waiting time distribution and different pulse duration distributions. The black dashed line is the Lorentzian spectrum for the base case.

mild deviations in the tail, which are considered acceptable. Exponentially distributed pulse durations will not be considered due to the significantly increased zero-frequency value in the PSD [51], and neither will Pareto distributed durations due to their drastic effects on the power spectra [55]. Both uniformly distributed and Rayleigh distributed durations give Lorenzian-like power spectra. To include a narrow (compared to the Rayleigh), unimodal distribution that is still positive definite, we use a Gamma distribution with shape parameter 20.

In Fig. 16, we present the PSD of the normalized synthetic time series without added noise, for the various selected distributions of durations and waiting times. The intermittency parameter has very little visible effect on the power spectra, so figures for other intermittency parameters are not presented. Indeed, for the base case with a degenerate distribution of pulse durations and an exponential waiting time distribution, the PSD of the normalized process $\widetilde{\Phi}$ does not depend on the intermittency parameter [51]. The theoretical expectation for the base case is apparently very close to the spectra for all cases presented in Fig. 16. This will be quantified by estimates of the average duration in Secs. VI B and VI C.

A. Wrongly estimated pulse duration

Before investigating the effect of a distribution of pulse durations and waiting times on the deconvolution method, we consider the isolated effect of misidentifying the correct duration for the base case process. Here, we deconvolve a base case realization with a pulse function with a wrong duration time in the case of no noise added. We will keep the notation $\tau_{\mathrm{d}} = \langle \tau \rangle$ for the true duration of the process and use $\tau_{\mathrm{d,\,wrong}}$ for the assumed wrong duration.

In Fig. 17, the estimated amplitude and waiting time distributions are presented for various assumed $\tau_{\mathrm{d,\,wrong}}$. The intermittency parameter for all realizations is $\gamma = 10$. Similar results are present, but with a weaker effect, for $\gamma = 1$. We observe deviations from the exponential distribution of waiting times when $\tau_{\mathrm{d,\,wrong}} > \tau_{\mathrm{d}}$ and deviations for an exponential amplitude distribution for the two most extreme cases with $\tau_{\mathrm{d,\,wrong}}/\tau_{\mathrm{d}} = 1/10$ and $\tau_{\mathrm{d,\,wrong}}/\tau_{\mathrm{d}} = 10$. This is complemented by tables VII and VIII where the estimated average waiting time and amplitude are compared to their respective original sample mean values. Large $\tau_{\mathrm{d,\,wrong}}/\tau_{\mathrm{d}}$ leads to overestimation of the average waiting time and this is more pronounced for larger intermittency parameters. Large $\tau_{\mathrm{d,\,wrong}}/\tau_{\mathrm{d}}$ also leads to underestimation of the average amplitude and this is most pronounced for small intermittency parameters. Small $\tau_{\mathrm{d,\,wrong}}/\tau_{\mathrm{d}}$ has very little effect on the average waiting time but leads to overestimation of the average amplitude, in particular for large intermittency parameters.

We interpret these results as follows. The deconvolution preserves the integral of the signal (at least before the 3-point maxima is applied) and the integral of the pulse function is equal to the duration time. Therefore, overestimating or underestimating the pulse duration leads to decreased or increased mass in the estimated forcing, respectively. Increased mass in the forcing raises the zero-level of the entire forcing, increasing the amplitudes but not causing any events to be lost. Decreased mass in the forcing can only be achieved by decreasing pulse amplitudes, which also eliminates some pulses entirely. Hence, overestimating the duration time leads to lost pulse arrivals, while underestimating the duration mainly leads to increased pulse amplitudes, which is moderate unless the underestimation is extreme or there is significant pulse overlap. These effects suggest that in application of the deconvolution method, where the pulse duration must be estimated, one should favor doing the deconvolution with a slightly lower duration than the estimated one. Noise, different distributions of amplitudes, waiting times, or durations, or correlations between these
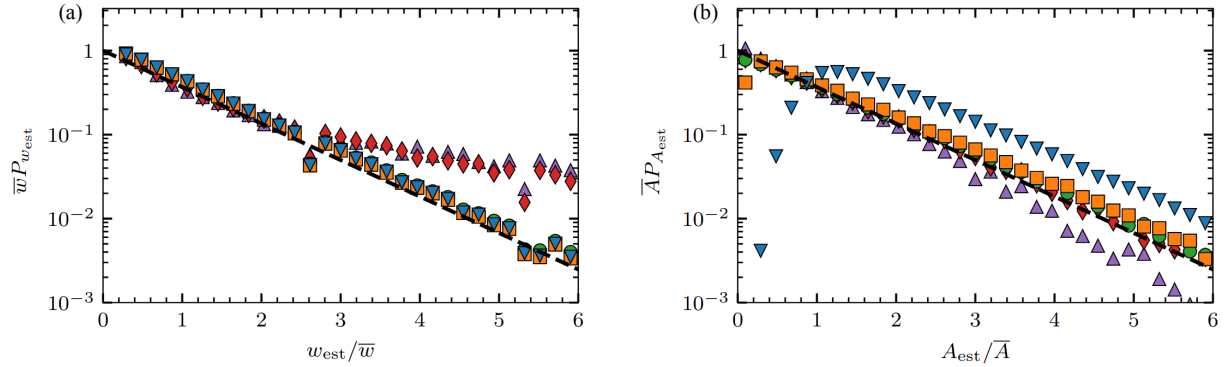
FIG. 17: Probability distribution function of estimated (a) waiting times and (b) amplitudes for $\gamma = 10$ and various assumed values of $\tau_{d,\,wrong}$. The black dashed lines shows an exponential distribution. ▼ represents $\tau_{d,\,wrong}/\tau_d = 1/10$, ■ shows the data for $\tau_{d,\,wrong}/\tau_d = 1/2$, ● shows the data for $\tau_{d,\,wrong}/\tau_d = 1$ ♦ shows the data for $\tau_{d,\,wrong}/\tau_d = 2$ and ▲ represents $\tau_{d,\,wrong}/\tau_d = 10$.

γ	$\tau_{d,\,wrong}/\tau_d$							
	10^{-1}	$1/2$	1	$11/10$	$5/4$	$3/2$	2	10
10^{-1}	1.01	1.01	1.01	1.02	1.04	1.06	*1.12*	*1.78*
1	1.02	1.01	1.03	1.09	*1.18*	*1.33*	*1.60*	*4.53*
10	*1.17*	*1.16*	*1.19*	*1.28*	*1.44*	*1.71*	*2.22*	*9.22*

TABLE VII: Table of the estimated average pulse waiting times normalized by the mean of the sample waiting times, $\overline{w}_{est}/\overline{w}$, for different pulse durations at different intermittency values.

random variables could change this conclusion, however.

B. Effect of waiting time distribution

Here, we will consider how various waiting time distributions affect the estimated pulse duration, and in turn how this influences the estimated amplitude and waiting time statistics using the deconvolution method. In Table IX, the estimated pulse duration for the cases in Fig. 16(a) are presented. These estimated durations were found by performing a least-square minimization using the curve_fit function of the SciPy module in Python, in the normalized

γ	$\tau_{\mathrm{d,wrong}}/\tau_{\mathrm{d}}$							
	10^{-1}	$1/2$	1	$11/10$	$5/4$	$3/2$	2	10
10^{-1}	1.02	1.01	1.01	0.97	0.91	*0.82*	*0.70*	*0.25*
1	1.10	1.01	1.02	1.00	0.97	0.93	*0.86*	*0.50*
10	*1.97*	*1.20*	*1.11*	*1.11*	1.10	1.09	1.07	*0.88*

TABLE VIII: Table of the estimated average pulse amplitudes normalized by the sample mean amplitudes, $\overline{A}_{\mathrm{est}}/\overline{A}$, for different pulse durations at different intermittency values.

$\tau_{\mathrm{d,est}}/\tau_{\mathrm{d}}$	γ		
	10^{-1}	1	10
$w \sim$ Exponential	1.00	1.08	0.96
$w \sim$ Rayleigh	0.93	1.04	0.99
$w \sim$ Pareto(3)	*1.51*	*1.41*	*1.42*

TABLE IX: Ratio between estimated pulse duration $\tau_{\mathrm{d,est}}$ and the pulse τ_{d} for different waiting time distributions, corresponding to the spectra presented in Fig. 16(a).

frequency range $\tau_{\mathrm{d}}\omega$ between 10^{-1} to 10^{2}. The pulse duration is well estimated for the case of exponentially and Rayleigh distributed waiting times. The pulse amplitude and waiting time distributions estimated from the deconvolution method are not affected by the duration estimate in these cases, and are therefore not presented.

Pareto distributed waiting times gives an overestimation of the pulse duration, and we expect this to have effects on the subsequently estimated amplitude and waiting time distributions using the deconvolution method. The case $\gamma = 10^{-1}$ has the largest overestimation and will therefore be investigated in more detail. In Fig. 18, we present estimated amplitude and waiting time distributions from deconvolution for pulse durations τ_{d}, $\tau_{\mathrm{d,est}}$ and $\tau_{\mathrm{d,est}}/2$. The first is included as a baseline reference, while the third is included to demonstrate the effect of using a shorter duration than the estimated one. The amplitude distribution is not visibly affected, but the estimated waiting time distributions show elevated tails at large waiting times ($> 10\overline{w}$) and in the case where we used $\tau_{\mathrm{d,est}}$, the distribution is elevated for
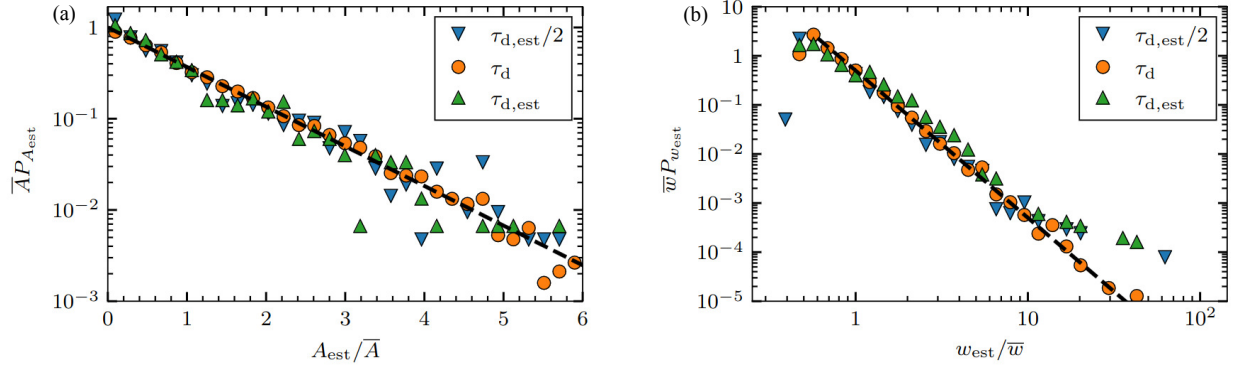
FIG. 18: Estimated (a) amplitude distributions and (b) waiting time distributions for Pareto distributed waiting times, $w \sim \text{Pareto}(3)$, and $\gamma = 10^{-1}$. These plots compare the results from deconvolution performed using the true duration (orange circles) and the estimated duration (green triangles) from Table IX. The black dashed lines are the references of the input distributions.

Estimated averages	τ_{d}	$\tau_{\text{d, est}}/2$	$\tau_{\text{d, est}}$
$\overline{A}_{\text{est}}/\overline{A}$	1.01	0.91	0.88
$\overline{w}_{\text{est}}/\overline{w}$	1.01	0.91	1.28

TABLE X: Comparison of rescaled estimated parameters from Fig. 18 for Pareto waiting times, $\gamma = 10^{-1}$, using different duration time values in the deconvolution.

waiting times greater than $\overline{w}$.

The mean values of all distributions presented in Fig. 18 are presented in Table X. The mean amplitudes are consistently underestimated, both for $\tau_{\text{d, est}} > \tau_{\text{d}}$ and for $\tau_{\text{d, est}}/2 < \tau_{\text{d}}$, in contrast to the results in Table VIII. We see that the deconvolution significantly overestimates $\overline{w}_{\text{est}}$ for $\tau_{\text{d, est}}$, and underestimates $\overline{w}_{\text{est}}$ for $\tau_{\text{d, est}}/2$. In agreement with the results reported in the previous section, using a slightly lower pulse duration than the estimated one improves the estimated average amplitude and waiting time.

C. Distribution of pulse durations

Consider now the situation where there is a distribution of pulse durations. The deconvolution method assumes all pulses to have the same duration, so it is of interest to investigate

P_τ	Estimated averages	γ		
		10^{-1}	1	10
Rayleigh	$\overline{A}_{\text{est}}/\overline{A}$	*0.86*	*0.88*	*0.88*
	$\overline{w}_{\text{est}}/\overline{w}$	0.99	1.03	1.00
Gamma(20)	$\overline{A}_{\text{est}}/\overline{A}$	0.96	0.97	1.06
	$\overline{w}_{\text{est}}/\overline{w}$	1.02	1.05	*1.16*
$\mathcal{U}(0,2)$	$\overline{A}_{\text{est}}/\overline{A}$	*0.80*	*0.83*	*0.80*
	$\overline{w}_{\text{est}}/\overline{w}$	1.00	1.01	0.92

TABLE XI: Estimated averages of the amplitude and waiting times using different pulse duration distributions with the same average duration for all cases.

how it performs for a distribution of pulse durations. We first establish a baseline for the performance of the method using the theoretical average duration time. In Table XI, we present estimated average amplitudes and waiting times for three different pulse duration distributions. Table II describes the results for a degenerate distribution of pulse durations. While Gamma distributed durations with shape parameter 20 is narrow enough for the parameters to be well estimated, in both the Rayleigh and uniform cases, the amplitudes are underestimated for all intermittency parameters. For small intermittency parameters, this agrees with the results in Table VIII: the amplitudes of pulses with duration greater than τ_{d} are accurately reconstructed, while amplitudes of pulses with durations smaller than τ_{d} are underestimated. Pulse overlap modifies this relationship, leading to more robustly underestimated amplitudes for the case of randomly distributed durations.

The estimated average duration from fitting to the PSD is presented in Table XII for various duration distributions and intermittency parameters. The case of degenerately distributed pulse durations is equivalent to the case of exponentially distributed waiting times presented in Table IX. Again, a Gamma(20) distribution of durations does not appear to affect the results, but the average pulse duration is overestimated in both of the other cases. The case of uniformly distributed durations give the largest deviation in the estimated average pulse duration and will again be investigated in more detail.

The estimated pulse amplitude and waiting time distributions for the case of uniformly

$\tau_{\mathrm{d,est}}/\overline{\tau_{\mathrm{d}}}$	γ		
	10^{-1}	1	10
$\tau_{\mathrm{d}} \sim \mathrm{Rayleigh}$	*1.43*	*1.27*	*1.28*
$\tau_{\mathrm{d}} \sim \mathrm{Gamma}(20)$	1.02	1.10	0.98
$\tau_{\mathrm{d}} \sim \mathcal{U}(0,2)$	*1.54*	*1.45*	*1.39*

TABLE XII: Ratio between estimated pulse duration $\tau_{\mathrm{d,est}}$ and sample mean of the durations $\overline{\tau}_{\mathrm{d}}$ using different duration time distributions.
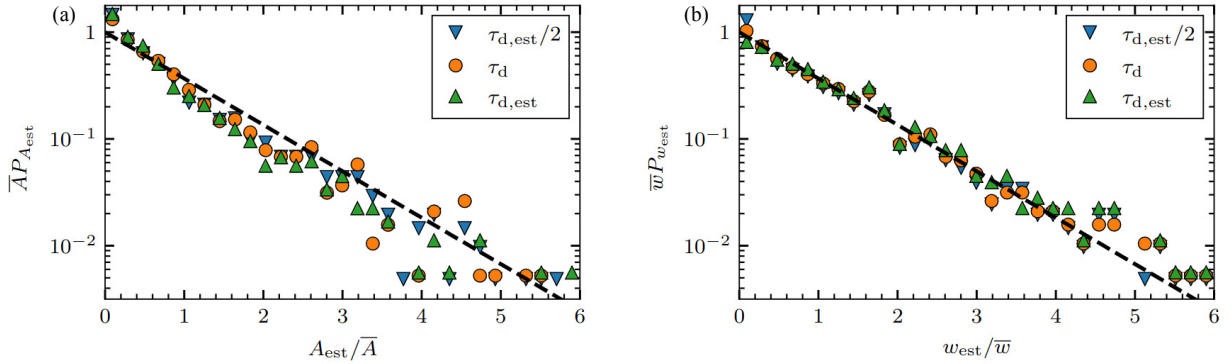

FIG. 19: Estimated (a) amplitude distributions and (b) waiting time distributions for uniformly distributed duration times, $\tau \sim \mathcal{U}(0,2)$ and $\gamma = 10^{-1}$, using various assumed constant duration times. The black dashed lines are the references of the input distributions.

distributed durations are presented in Fig. 19, and the estimated average amplitudes and waiting times are presented in Table XIII. Again, we compare results from deconvolution with pulses using τ_{d}, $\tau_{\mathrm{d,est}}$ and $\tau_{\mathrm{d,est}}/2$ as pulse duration. Both the amplitude and waiting time distributions are well estimated. The estimated average waiting time is within 10% of the original sample mean value for both estimated pulse durations used for deconvolution. However, the deconvolution with the reduced estimated pulse duration captures the average amplitude better than the full estimated duration time, consistent with the previous cases.

Estimated averages	τ_d	$\tau_\mathrm{d,est}/2$	$\tau_\mathrm{d,est}$
$\overline{A}_\mathrm{est}/\overline{A}$	*0.80*	*0.80*	*0.72*
$\overline{w}_\mathrm{est}/\overline{w}$	1.00	0.93	1.08

TABLE XIII: Table showing the rescaled estimated parameters from Fig. 19 for uniformly distributed pulse durations using different average duration values in the deconvolution for $\gamma = 10^{-1}$. The corresponding estimated duration can be found in Table XII.

D. Conclusion

In this section, we have investigated the effect of estimating the pulse duration time from the PSD in the case of non-exponentially distributed waiting times, as well as for a distribution of pulse durations. If the shape of the power spectrum is similar to that of a single pulse (as it will be for uncorrelated pulses with an exponential waiting time distribution), the pulse duration is accurately estimated and there are no issues in applying the deconvolution method. Likewise, a narrow distribution of pulse durations (Gamma(20) in our case) give reliable estimates of amplitudes and waiting times. However, broadly distributed pulse durations will lead to errors in the estimated averages, even if the average duration time is known exactly. It is again demonstrated that using a smaller duration time than the one estimated from the frequency spectrum is preferable, but did not improve on using the mean duration time.

VII. RECONSTRUCTION OF THE PULSE FUNCTION

In some applications, the forcing is known or may be estimated, while the pulse function (or system response to the forcing) is unknown. It is clear from Eq. (7) and the iteration scheme given by Eq. (10) that the particular interpretation of the vectors f and φ does not affect the deconvolution algorithm if the known vector satisfies the conditions of non-negativity and a positive value at $t = 0$. As such, we may consider f a known forcing and φ an unknown pulse function and obtain the deconvolution algorithm by switching the symbols f and φ in Eq. (10):

$$\varphi_j^{(n+1)} = \varphi_j^{(n)} \frac{\left(\Phi * \widehat{f}\right)_j + b}{\left(\varphi^{(n)} * f * \widehat{f}\right)_j + b}.$$

(12)

Here, we are interested in the direct result of the deconvolution, and so we do not expect the value of the intermittency parameter to significantly influence the result. This is confirmed by Fig. 20, where we present reconstructions of the one-sided exponential pulse for different intermittency parameters from realizations of the model in the base case. In all cases with a finite intermittency parameter, the pulse reconstruction is reliable, although in the inset it is seen how a lager intermittency parameter pushes up the noise floor, thereby decreasing the time window of accurate reconstruction. Still, for $\gamma = 10$, the effects are only seen two decades below the maximal pulse value. In Fig. 20, we have also indicated the case $\gamma \to \infty$ by letting the forcing signal consist of independently and identically normally distributed random variables, with a constant added to make the forcing positive. In this most severe case, the pulse function is significantly affected, with a slight rise before the peak and correspondingly a faster decay after the peak. In the following, we will only consider the case of moderate pulse overlap, given by $\gamma = 1$. For the results presented in this section, we use a portion of the synthetic time series and its known forcing, where both have a length of $2^{19} + 1$ data points. The initial guess for estimating the pulse function is an array, also of size $2^{19} + 1$ data points, containing a boxcar function centered at zero with an amplitude of one and a width of $2^{17} + 1$ data points, equivalent to about 1300 pulse duration times. All values outside the boxcar are set to zero. The boxcar is used to improve stability; allowing positive values over the entire estimated pulse function array can sometimes lead to spurious positive values at the far ends with a corresponding degradation of the pulse function in the center. Since zeros remain zero during the iteration of the algorithm, such effects are removed by
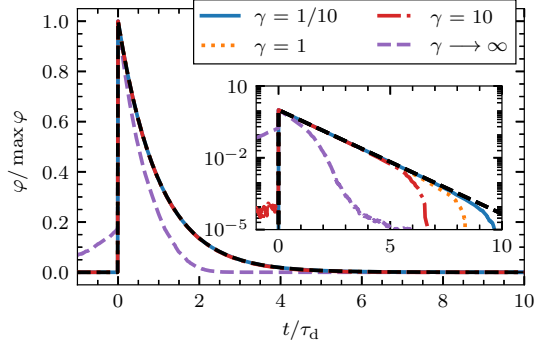
FIG. 20: Reconstruction of a one-side exponential pulse function with different intermittency parameters without noise added. The inset shows the differences between the pulse reconstructions with semi-logarithmic axes. The black dashed line shows the true one-sided exponential pulse function.

the boxcar. Note that the boxcar is still huge compared to the expected size of the pulse it contains.

In Fig. 21, reconstructed pulse functions from deconvolution of down-sampled model realizations using down-sampled forcing is presented. Here it is seen that the pulse function is not accurately reproduced if the process is undersampled. Note that for the comparison, the pulse functions have been rescaled to match the amplitude of the original pulse function. For $\theta = 10^{-1}$, the maximum value is 0.93 and for $\theta = 1$, the maximum value is 0.62. Thus, the reconstructed amplitude is affected as well as the pulse function. We note, however, that this result is sensitive to how undersampling affects the forcing, and in particular whether undersampling leads to losses of entire pulses or not. In the case presented here, loss of pulses in the forcing but not the signal is the major discrepancy between the original down-sampled forcing and the forcing estimated from the down-sampled signals.

We consider the effect of noise in Fig. 22. Additive noise leads to noise in the tail, worse for higher ϵ. Still, we reach noise rms two times the signal rms without significant deviations from the pulse function. Dynamical noise also distorts the pulse function, but not significantly, likely due to the noise in this case being convolved with the same pulse function as the forcing.

In Fig. 23, we present reconstruction of the pulse function in the case of broad distribution of durations. The deconvolution method assumes all pulses have the same duration, but is shown to accurately reproduce the pulse function with largely the correct average du-
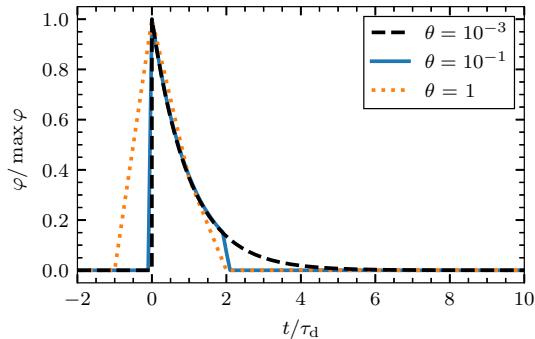
41

FIG. 21: Reconstruction of a one-sided exponential pulse function (solid colored line and dotted colored line) for variously down-sampled signals with intermittency parameter $\gamma = 1$ and no noise added. The black dashed line shows the true one-sided exponential pulse.
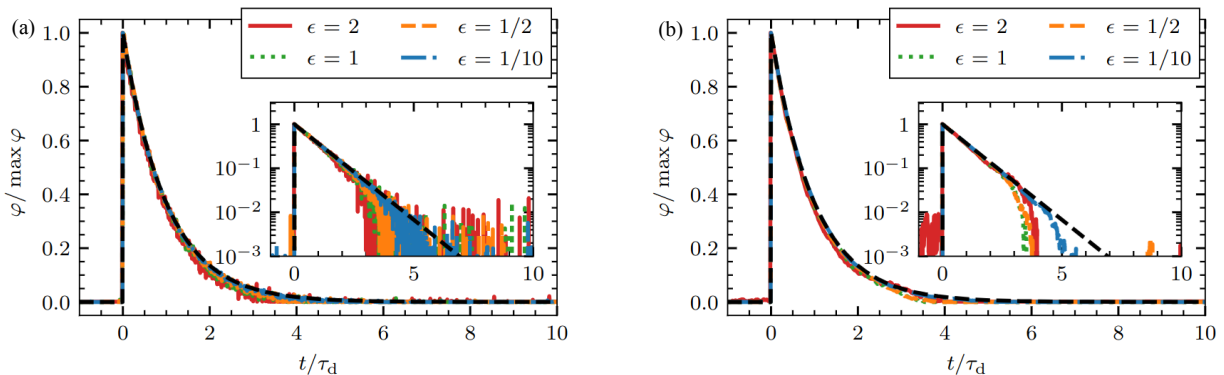

FIG. 22: Reconstruction of a one-sided exponential pulse function using the modified RL-deconvolution for $\gamma = 1$ and different values of the noise to signal ratio ϵ. (a) Additive noise. (b) Dynamic noise. The black dashed lines represent the true one-sided exponential pulse function.

ration. Narrower duration distributions (Rayleigh, Gamma(20)) were also attempted, but gave results indistinguishable from the true pulse function. Attempting to fit the result of the Pareto case to a single exponential on a linear scale gives $\tau_{\mathrm{d,est}} = 0.85$, which is easily seen in the inset not to capture the correct pulse function.

In conclusion, the deconvolution method can be used reliably in order to recover the pulse function from a given forcing. Only severe noise, undersampling or excessively broad pulse duration distributions lead to significant deviations from the average pulse function.
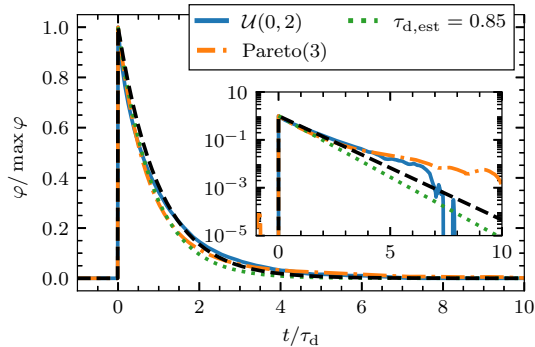
FIG. 23: Pulse reconstruction with randomly distributed pulse durations for $\gamma = 1$ and no added noise. The inset shows the same with semi-logarithmic axes. The dotted line gives the best fit to the pulse function in the Pareto case. The black dashed line is the true one-sided exponential case.

VIII. DISCUSSION AND CONCLUSIONS

In this contribution, we have presented a novel method for extracting pulse amplitudes and arrival times from realizations of a stochastic process given by a superposition of pulses with fixed shape. The method relies on the ISRA deconvolution algorithm,which produces the maximum-likelihood solution to the deconvolution problem $\Phi = \varphi * f + X$ where Φ is a known signal, φ is a known pulse or kernel function, f is the forcing to be estimated and X is normally distributed noise. Since the result of the deconvolution algorithm is the forcing time series f, a 3-point maxima is used to estimate pulse amplitudes and arrival times.

For realizations of an intermittent process with high temporal resolution (sampling time $1/20$ times the average time between pulses or better and $1/10$ times the average pulse duration time or better), amplitude and waiting time distributions are well recovered in a variety of cases. Coarser sampling or more pulse overlap both lead to several pulses being counted as one, which distorts the estimated amplitude distribution and leads to overestimation of the average amplitude and waiting time. We note that this condition on the sampling time still allows for pulse overlap and pulses which are separated by two sampling times or more are robustly separated by the algorithm. Based on studies of numerous model realizations, it is recommend to use the approximate conditions $\gamma\theta = \triangle_t/\langle w \rangle \leq 1/20$ and $\theta \leq 1/10$ to determine if the deconvolution will give reasonable estimates of mean values of pulse amplitudes and waiting times, and $\gamma\theta \leq 1/10$ if only the functional shape of the

corresponding distributions are desired.

While the deconvolution method is only designed for positive valued signals and forcings, negative signal values may be accounted for by a straightforward modification of the algorithm. If negative values are present, we may recover both positive and negative parts by using the method on both the signal and its sign reversed version separately, and combine the results. While the method is not able to accurately resolve parts of data time series where fluctuations of different sign arrive close together, parts of the signal where one sign dominates are well reconstructed.

Noise may be handled, and relies at present on introducing an amplitude threshold in the 3-point maxima. The threshold performs well recovering the tail of the amplitude and waiting time distributions. Average pulse amplitudes are estimated to within 15% of their true value, while mean waiting times are estimated to within a factor two.

If the pulse duration is not known before applying the deconvolution method, it may be estimated from the frequency PSD of the process. The spectrum is insensitive to pulse overlap and amplitude distribution and robust to non-exponential waiting time distributions as well as distributions of the pulse duration. It is demonstrated that for intermittent processes, underestimation of the pulse duration has little to no effect on the estimation of pulse amplitudes and arrivals while overestimation of the pulse duration has significant implications. Broad distributions in waiting times or durations lead to overestimation of the average duration, which in turn distorts average amplitude and waiting time estimated from the deconvolution. It is shown that in these cases, performing the deconvolution with a pulse duration half the estimated value from the PSD improves the results.

Lastly, if the forcing is known but the pulse function is unknown, the ISRA algorithm may be employed straightforwardly. We have demonstrated that the reproduction of the pulse function is excellent for all but the most severe undersampling or noise. Even if there is a narrow distribution of the pulse durations, the algorithm recovers the pulse function with the average duration time.

For real data, we advise a procedure as in Refs. 23 and 39. First, the analysis leading to estimates of the stochastic model parameters is performed. Then, the deconvolution is performed on the empirical data. Finally, synthetic data from model realizations with the estimated model parameters is made and analysed in the same manner as the empirical data. Results from analysis of the synthetic data should then be compared to the results of the

44

measurement time series. Ideally, this should be carried out as a Monte-Carlo study with multiple model realizations, demonstrating that the results of the deconvolution are within the expected errors for synthetic data.

In conclusion, the deconvolution algorithm is shown to recover amplitude and waiting time distributions from realizations of an intermittent process even in the presence of significant pulse overlap, addition of noise, and deviations from the expected pulse function. For all signals considered in this contribution, the underlying ISRA method recovers the forcing admirably and only the basic information loss associated with the finite sampling of a continuous signal affects the reconstruction of amplitude and waiting time distributions.

Appendix A: Effect of sampling on event recollection

In this section, we investigate the information loss associated with poor sampling of the filtered Poisson process (FPP). Consider a Poisson point process $K(T)$ on the interval $[0, T)$ with rate parameter $1/\gamma$. For a given realization, the arrivals of K events are uniformly distributed on the interval. The interval is discretized into N time steps of size $\triangle_t = T/N$. For reference, we note that

$$\langle K \rangle = \gamma \theta N, \tag{A1}$$

where $\theta = \triangle_t / \tau_\mathrm{d}$. We only record if events occur in a given time step, but not how many events occur. We therefore move from the process $K(T)$ to the process $F(N)$, denoting locations with events. By necessity, $F \leq K$ and $F \leq N$. For each of the N time steps, the probability of receiving events is $1 - \mathrm{Pr}[\text{No events in time } \theta]$, which from the Poisson distribution of K is $1 - \exp(-\gamma\theta)$. Therefore, the probability mass function (PMF) of F is a Binomial distribution over N trials with success probability $1 - \exp(-\gamma\theta)$,

$$P_F(f; \gamma, \theta, N) = \binom{N}{f} \exp(-\gamma\theta)^{N-f}[1 - \exp(-\gamma\theta)]^f. \tag{A2}$$

The mean value is given by

$$\langle F \rangle = N\left[1 - \exp(-\gamma\theta)\right], \tag{A3}$$

45

For $\gamma\theta \ll 1$, the exponential in Eq. (A3) can be expanded and $\langle F \rangle \approx \langle K \rangle$. However, $\gamma \approx 1/\theta$ gives $\langle K \rangle \approx N$ but $\langle F \rangle \approx 0.6N$ as many events arrive at the same discrete time location.

Let us now investigate the effect of the 3-point maxima peak finding algorithm. Letting M denote the number of maxima, we do the following approximation (here "cluster of size k" means k consecutive filled time steps with empty time steps at each end)

$$M = \sum_c \sum_m (\text{number of maxima in cluster } m \text{ of size } c)$$

$$\langle M \rangle \approx N \sum_c P[\text{cluster of size } c] \langle \text{number of maxima in clusters of size } c \rangle \qquad (A4)$$

Letting C be the cluster size, the probability of having a cluster of size c is given by

$$\Pr[C = c] = \left(1 - e^{-\gamma\theta}\right)^c \left(e^{-\gamma\theta}\right)^2. \qquad (A5)$$

To find the number of maxima per cluster, we argue as follows: As the amplitudes are independently and identically distributed, so are the values of the forcing at neighboring time steps. As such, all permutations of the forcing values in a cluster are equally likely. For $c = 1$ and $c = 2$, there is obviously just one maximum. For $c = 3$, there are 6 permutations, two of which give two maxima and the rest give one, for an average of $4/3$ maxima in the cluster. By going from two to three data points, we had a $1/3$ chance of adding an extra maximum. Adding further data points to the end of the sequence each time gives an additional $1/3$ chance of a new maxima, so the average number of maxima should increase by $1/3$ per new data point. Aided by brute force investigation of all ordered sequences up to $c = 10$, we guess that the average number of maxima in an ordered sequence of size $c \geq 2$ is $(c + 1)/3$, including the end points. Adding $1/3$ to this gives $((c + 1) + 1)/3$, the average number of maxima in a sequence of $c + 1$ data points. We therefore have that the approximation in Eq. (A4) can be written as

$$\frac{\langle M \rangle}{N} \approx \Pr[C = 1] + \sum_{c=2}^{\infty} \Pr[C = c] \frac{c + 1}{3}$$

$$\frac{\langle M \rangle}{N} \approx \frac{1 - e^{-\gamma\theta}}{3} \left[1 + e^{-\gamma\theta} + \left(e^{-\gamma\theta}\right)^2\right] \qquad (A6)$$

where we have used $\Pr[C = c]$ from Eq. (A5). This expression fits our expectations: For very small $\gamma\theta$, we mainly expect clusters of size 1, and Eq. (A6) approaches Eq. (A3). For very large $\gamma\theta$, the entire time series is likely to be filled, so we expect about $N/3$ maxima following the discussion above.
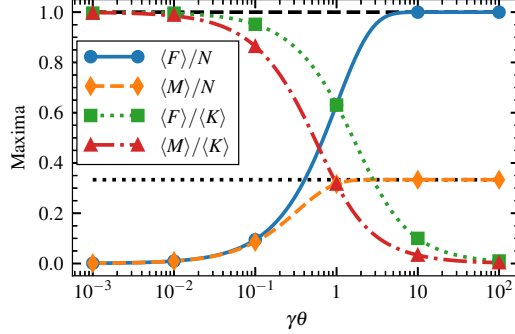
FIG. 24: Comparison of the expected number of events $\langle K \rangle$, the expected number of time steps with events $\langle F \rangle$ and the expected number of events found after a 3-point maxima $\langle M \rangle$ as a function of $\gamma\theta$. In all cases, the number of data points was $N = 10^5$ and the normalized time step was $\theta = 10^{-2}$. The lines give analytical approximations, and the symbols are results from numerical simulations.

In Fig. 24, we compare the analytical results of this section with results from a Monte-Carlo study. The lines give the analytical predictions, while the points give mean values of 20 realizations of the base case with $N = 10^5$ and $\theta = 10^{-2}$ for various γ. One standard deviation would give error bars smaller than the plot symbols. Eq. (A6) is an excellent approximation to the average number of maxima for large N, and that the full distribution of the number of events is narrow, justifying the use of $\langle M \rangle$ instead of the full distribution.

For the deconvolution method, $\langle M \rangle / \langle K \rangle$ is the most significant number, as this gives a measure of how well we may hope to recreate the individual pulses in the time series. The lower this ratio is, the more separate pulses are counted as one. If we had a way of avoiding the 3-point maxima, we would still have the coarse-graining introduced by discretizing the time series. In this ideal case, $\langle F \rangle / \langle K \rangle$ is the important ratio.

Appendix B: Definition of distributions

In this section, we list the distributions used in this manuscript. In all cases, we give the distributions in terms of their mean value $\langle X \rangle = \mu$, so that for any realization $\{X_k\}_{k=1}^K$ listed here, we may get amplitudes, waiting times or pulse durations by setting $\mu = \langle A \rangle$, $\mu = \tau_{\mathrm{w}}$, or $\mu = \tau_{\mathrm{d}}$ respectively. Unless otherwise noted, all distributions used are positive definite, so we only give the distributions for $x > 0$. For $x < 0$, we have $P_X(x) = 0$.

a. *The exponential distribution* is given by

$$P_X(x) = \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right). \tag{B1}$$

We denote an exponentially distributed random variable by $X \sim \mathrm{Exp}$

b. *The gamma distribution* is given by

$$P_X(x; k) = \frac{k^k}{\Gamma(k)\mu^k} x^{k-1} \exp\left(-\frac{kx}{\mu}\right). \tag{B2}$$

This distribution has one free parameter, the shape parameter k. For $k = 1$, this coincides with the exponential distribution. In the main text, the gamma distribution is denoted by $X \sim \mathrm{Gamma}(k)$

c. *The Rayleigh distribution* is given by

$$P_X(x) = \frac{\pi x}{2\mu^2} \exp\left(-\frac{\pi x^2}{4\mu^2}\right). \tag{B3}$$

We denote a Rayleigh distributed random variable by $X \sim \mathrm{Rayleigh}$

d. *The Pareto distribution* is given by

$$P_X(x) = \frac{(\alpha - 2)^{\alpha-1}}{(\alpha - 1)^{\alpha-2}} \frac{\mu^{\alpha-1}}{x^\alpha}, \quad x \geq \frac{\alpha - 2}{\alpha - 1}\mu. \tag{B4}$$

As we require a well-defined mean for all random variables, we demand $\alpha > 2$. We denote a Pareto distributed random variable by $X \sim \mathrm{Pareto}(\alpha)$. Note that following this definition, the PDF decays as $x^{-\alpha}$, while in the standard definition, it is the cumulative distribution function which decays as $x^{-\alpha}$.

e. *The degenerate distribution* is given by

$$P_X(x) = \delta(x - \mu). \tag{B5}$$

We denote a degenerately distributed random variable by $X \sim \mathrm{degenerate}$.

f. *The uniform distribution* is given by

$$P_X(x) = \begin{cases} \frac{1}{2\mu}, & 0 \leq x \leq 2\mu \\ 0, & \text{else.} \end{cases} \tag{B6}$$

Note that this is the broadest possible non-negative uniform distribution with mean μ. We denote the uniform distribution by $X \sim \mathcal{U}(0, 2)$.

Appendix C: Pearson correlation coefficients

In this appendix, we present Pearson correlation coefficients between the true forcing f and the forcing f_{est} from estimated amplitudes and arrival times. These results are presented for completion, in general the conclusions that may be drawn from these tables are the same as may be drawn from the distributions and mean values in the main text. To guide the eye and highlight this correspondence, we have marked correlations below 0.9 in italics.

Comparing Tables XVII (for noisy signals without thresholding) and XVIII (for noisy signals with thresholding), we see that the thresholding, although it improves the estimate of the amplitude- and waiting time distributions, uniformly lead to larger differences between f and f_{est}. If the thresholding were able to separate between events and noise, we would expect the thresholding to improve the correlation. From the Tables, it is evident that it removes many true events as well.

	γ				
	1/10	1	5	10	50
$A \sim$ Exponential	0.95	0.93	0.92	0.91	*0.86*
$w \sim$ Exponential	0.95	0.93	0.92	0.91	*0.86*

TABLE XIV: Table of the Pearson correlation coefficients between the true forcing and the forcing from estimated amplitudes and arrivals corresponding to the the base case, Fig. 7 and Table II.

	γ
	10^{-1}	1	5	10	50
Rayleigh	0.94	0.92	0.91	0.90	*0.82*
Pareto(3)	0.92	0.93	0.96	0.92	0.90
$\mathcal{U}(0,2)$	0.94	0.92	0.91	0.90	*0.82*
degenerate	0.93	0.92	0.90	*0.88*	*0.77*

TABLE XV: Pearson correlation coefficients between the true forcing and the forcing from estimated amplitudes and arrivals for different amplitude distributions, corresponding to Fig. 8 and Table III.

	γ
	10^{-1}	1	5	10	50
Rayleigh	0.90	0.93	0.92	0.92	*0.87*
Pareto(3)	0.92	0.93	0.92	0.92	0.91
$\mathcal{U}(0,2)$	0.91	0.93	0.91	0.91	*0.87*
degenerate	1.00	1.00	0.93	0.93	0.90

TABLE XVI: Pearson correlation coefficients between the true forcing and the forcing from estimated amplitudes and arrivals for different waiting time distributions, corresponding to Fig. 9 and Table IV

	ϵ	γ		
		10^{-1}	1	10
Additive	10^{-1}	0.99	0.95	0.74
	1/2	0.97	0.83	0.43
	1	0.95	0.73	0.30
Dynamic	10^{-1}	1.00	0.99	0.97
	1/2	0.98	0.93	0.88
	1	0.96	0.88	0.78

TABLE XVII: Table showing the values of the Pearson correlation coefficients between the true forcing and the estimate forcing for different noise to signal ratios, corresponding to Fig. 12.

	ϵ	γ		
		10^{-1}	1	10
Additive	10^{-1}	0.95	0.89	0.67
	1/2	0.93	0.77	0.37
	1	0.92	0.68	0.24
Dynamic	10^{-1}	0.95	0.91	0.85
	1/2	0.94	0.85	0.65
	1	0.93	0.78	0.49

TABLE XVIII: Pearson correlation coefficients between the true forcing and the forcing using estimated amplitudes and arrival times with thresholding, for various intermittency parameters and noise to signal ratios corresponding to Figs. 13 and 14 and Tables V and VI.

γ	$\tau_{d,\,wrong}/\tau_d$				
	10^{-1}	$1/2$	1	2	10
10^{-1}	0.95	0.95	0.95	0.94	*0.77*
1	0.92	0.93	0.93	*0.88*	*0.61*
10	*0.87*	0.91	0.91	*0.84*	*0.53*

TABLE XIX: Pearson correlation coefficients between the true forcing and the estimated amplitudes and arrival times using 3-point maxima for various pulse duration times and intermittency values. This table corresponds to Fig. 17 and Tables VII and VIII.

[1] P. F. Salipante, S. E. Meek, and S. D. Hudson, Soft Matter **14**, 9020 (2018).

[2] S.-S. Ding, H.-M. Li, W.-D. Yan, and J.-Q. Zhong, Phys. Rev. Fluids **4**, 023501 (2019).

[3] Y. Wang, X. He, and P. Tong, J. Fluid Mech. **874**, 263 (2019).

[4] G. Decristoforo, A. Theodorsen, and O. E. Garcia, Phys. Fluids **32**, 085102 (2020).

[5] L. Kristensen, M. Casanova, M. S. Courtney, and I. Troen, Boundary-Layer Meteorol. **55**, 91 (1991).

[6] R. Narasimha, S. R. Kumar, A. Prabhu, and S. V. Kailas, Philos. Trans. A. Math. Phys. Eng. Sci. **365**, 841 (2007).

[7] P. Claps, A. Giordano, and F. Laio, Adv. Water Resour. **28**, 992 (2005).

[8] M. Lefebvre, Stat. Probab. Lett. **78**, 3274 (2008).

[9] J.-B. Thomazo, B. Le Révérend, L.-L. Pontani, A. M. Prevost, and E. Wandersman, Proc. Natl. Acad. Sci. **118**, e2020402118 (2021).

[10] Z. Elter, C. Jammes, I. Pázsit, L. Pál, and P. Filliatre, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. **774**, 60 (2015).

[11] J. R. Segal, B. Ceccarelli, R. Fesce, and W. P. Hurlbut, Biophys. J. **47**, 183 (1985).

[12] R. Fesce, J. R. Segal, and W. P. Hurlbut, J. Gen. Physiol. **88**, 25 (1986).

[13] M. J. E. Richardson, Phys. Rev. E **98**, 042405 (2018).

[14] V. Hakim and J. Ranft, Phys. Rev. E **101**, 012411 (2020).

[15] O. E. Garcia, V. Naulin, A. H. Nielsen, and J. Juul Rasmussen, Phys. Plasmas **12**, 062309 (2005).

[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, Phys. Plasmas **20**, 012303 (2013).

[17] J. Anderson and B. Hnat, Phys. Plasmas **24**, 062301 (2017).

[18] G. Decristoforo, F. Militello, T. Nicholas, J. Omotani, C. Marsden, N. Walkden, and O. E. Garcia, Phys. Plasmas **27**, 122301 (2020).

[19] A. J. Wootton, J. Nucl. Mater. **176-177**, 77 (1990).

[20] G. Y. Antar, G. Counsell, Y. Yu, B. Labombard, and P. Devynck, Phys. Plasmas **10**, 419 (2003).

[21] D. A. D'Ippolito, J. R. Myra, and S. J. Zweben, Phys. Plasmas **18**, 060501 (2011).

[22] O. E. Garcia, R. Kube, A. Theodorsen, J.-G. Bak, S.-H. Hong, H.-S. Kim, the KSTAR Project Team, and R. A. Pitts, Nucl. Mater. Energy **12**, 36 (2017).

[23] R. Kube, A. Theodorsen, O. E. Garcia, D. Brunner, B. LaBombard, and J. L. Terry, J. Plasma Phys. **86**, 905860519 (2020).

[24] F. Militello, P. Tamain, W. Fundamenski, A. Kirk, V. Naulin, and A. H. Nielsen, Plasma Phys. Control. Fusion **55**, 025005 (2013).

[25] A. Theodorsen, O. E. Garcia, J. Horacek, R. Kube, and R. A. Pitts, Plasma Phys. Control. Fusion **58**, 044006 (2016).

[26] A. Theodorsen, O. E. Garcia, R. Kube, B. LaBombard, and J. L. Terry, Nucl. Fusion **57**, 114004 (2017).

[27] N. R. Walkden, A. Wynn, F. Militello, B. Lipschultz, G. Matthews, C. Guillemaut, J. Harrison, and D. Moulton, Nucl. Fusion **57**, 036016 (2017).

[28] S. O. Rice, Bell Syst. Tech. J. **23**, 282 (1944).

[29] S. O. Rice, Bell Syst. Tech. J. **24**, 46 (1945).

[30] E. Parzen, *Stochastic Processes* (Society for Industrial and Applied Mathematics, Philadelphia, 1999).

[31] H. L. Pécseli and J. Trulsen, Phys. Fluids B Plasma Phys. **1**, 1616 (1989).

[32] F. J. Øynes, H. L. Pécseli, and K. Rypdal, Phys. Rev. Lett. **75**, 81 (1995).

[33] D. Block, I. Teliban, F. Greiner, and A. Piel, Phys. Scr. **T122**, 25 (2006).

[34] D. Rudakov, J. Boedo, R. Moyer, P. Stangeby, J. Watkins, D. Whyte, L. Zeng, N. Brooks, R. Doerner, T. Evans, M. Fenstermacher, M. Groth, E. Hollmann, S. Krasheninnikov, C. Lasnier, A. Leonard, M. Mahdavi, G. McKee, A. McLean, A. Pigarov, W. Wampler, G. Wang, W. West, and C. Wong, Nucl. Fusion **45**, 1589 (2005).

[35] J. A. Boedo, J. R. Myra, S. Zweben, R. Maingi, R. J. Maqueda, V. A. Soukhanovskii, J. W. Ahn, J. Canik, N. Crocker, D. A. D'Ippolito, R. Bell, H. Kugel, B. Leblanc, L. A. Roquemore, and D. L. Rudakov, Phys. Plasmas **21**, 042309 (2014).

[36] R. Kube, O. E. Garcia, A. Theodorsen, D. Brunner, A. Q. Kuang, B. LaBombard, and J. L. Terry, Plasma Phys. Control. Fusion **60**, 065002 (2018).

[37] I. Teliban, D. Block, A. Piel, and F. Greiner, Plasma Phys. Control. Fusion **49**, 485 (2007).

[38] H. Johnsen, H. L. Pécseli, and J. Trulsen, Phys. Fluids **30**, 2239 (1987).

[39] A. Theodorsen, O. E. Garcia, R. Kube, B. LaBombard, and J. L. Terry, Phys. Plasmas **25**, 122309 (2018), arXiv:1802.05052.

[40] W. H. Richardson, J. Opt. Soc. Am. **62**, 55 (1972).

[41] L. B. Lucy, Astron. J. **79**, 745 (1974).

[42] F. Benvenuto, R. Zanella, L. Zanni, and M. Bertero, Inverse Probl. **26**, 025004 (2010).

[43] F. Dell'Acqua, G. Rizzo, P. Scifo, R. A. Clarke, G. Scotti, and F. Fazio, IEEE Trans. Biomed. Eng. **54**, 462 (2007).

[44] F. Dell'Acqua, P. Scifo, G. Rizzo, M. Catani, A. Simmons, G. Scotti, and F. Fazio, Neuroimage **49**, 1446 (2010).

[45] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, Microsc. Res. Tech. **69**, 260 (2006).

[46] R. Kube and O. E. Garcia, Phys. Plasmas **22**, 012502 (2015).

[47] O. E. Garcia, R. Kube, A. Theodorsen, and H. L. Pécseli, Phys. Plasmas **23**, 052308 (2016).

[48] https://github.com/uit-cosmo/ (2022).

[49] O. E. Garcia, Phys. Rev. Lett. **108**, 265001 (2012).

[50] A. Theodorsen, O. E. Garcia, and M. Rypdal, Phys. Scr. **92**, 054002 (2017).

[51] O. E. Garcia and A. Theodorsen, Phys. Plasmas **24**, 032309 (2017).

[52] A. Theodorsen and O. E. Garcia, Plasma Phys. Control. Fusion **60**, 034006 (2018).

[53] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Application*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, 1997).

[54] O. E. Garcia, R. Kube, A. Theodorsen, B. LaBombard, and J. L. Terry, Phys. Plasmas **25**, 056103 (2018).

[55] M. A. Korzeniowska, A. Theodorsen, M. W. Rypdal, and O. E. Garcia (2023), in review, Phys. Rev. Research.