Plasma Physics and Controlled Fusion

PAPER

A Bayesian model of filamentary dynamics in MAST

To cite this article: L C Appel *et al* 2020 *Plasma Phys. Control. Fusion* **62** 125002

IOP Publishing Plasma Physics and Controlled Fusion

Plasma Phys. Control. Fusion **62** (2020) 125002 (18pp) https://doi.org/10.1088/1361-6587/ab8e1a

A Bayesian model of filamentary dynamics in MAST

L C Appel[1], S Kwak[2], F Militello[1] and J Svensson[2]

[1] UKAEA-CCFE, Culham Science Centre, Abingdon, Oxon, OX14 3DB
[2] IPP, Greifswald, Germany

E-mail: lynton.appel@ukaea.uk

Received 23 January 2020, revised 10 April 2020
Accepted for publication 28 April 2020
Published 20 October 2020

Abstract

A novel approach using Bayesian inference has been implemented to interpret the filamentary dynamics measured by a Langmuir probe fixed to a reciprocating assembly on MAST. The model describes the system as a superposition of time-displaced filaments and a fixed background component. Each filament is parameterised in terms of a characteristic rise and fall time and maximum amplitude centred on local maxima in the measured data time-series. A distinctive feature of the approach is that no minimum threshold is set for the existence of filaments. It is observed that whereas large amplitude filaments are well characterised in terms of rise times, smaller amplitude filaments are often unconstrained by the data and are limited by the details of the prior. Based on these findings, a new definition for the plasma filaments is proposed based on the uncertainty in the filament rise times. The remaining filaments together with the constant background component forms a new time-dependent signal referred to as the computed *background fluctuation* signal. The characteristics of these signals (for the *plasma filaments* and for the *background fluctuations*) are reported in terms of their spatial variation as the probe moves through the SOL and into the core plasma.

Keywords: plasma physics, bayesian inference, edge plasma, tokamak physics, plasma filaments

(Some figures may appear in colour only in the online journal)

1. Introduction

The edge region of the plasma presents a unique set of challenges. It is often referred to as the plasma exhaust in recognition of the important role it fulfills to remove the fusion products from the tokamak reactor. From the earliest experiments using Langmuir probes, it was evident that this region experiences significant fluctuations. Early evidence of the 'filamentation' of D-alpha light on TFTR was reported by Zweben in 1989 [1]. Improvements in diagnostic capabilities led to a rekindled interest in imaging of the scrape-off-layer (SOL). These revealed a dynamic behaviour of plasma 'blobs' being ejected from the well-confined plasma inside the magnetic separatrix [2]. The structures were generally observed to be be aligned with field structures and became known as filaments [3]. Understanding the filament dynamics in the edge region is complicated by many factors: the magnetic topology transitions from a geometry of closed field lines across a separatrix to an open field line geometry; for divertor configurations, there is a field null (usually referred to as an X-point) on the separatrix; there are strong gradients of plasma density, temperature and electric fields; particularly in the colder plasma regions the role of the neutral particles is critical; and finally, a significant component of the heat and particle flux is carried by three-dimensional structures.

In early work, the emphasis was to characterise observed edge fluctuations in terms of standard statistical techniques [4–6]. Analytical tools used included cross- and auto- correlation functions, moments of the distributions of the signal amplitudes, and power spectra. Later experimental interpretation in terms of filaments usually relied on defining a threshold criterion. A common choice is to specify the threshold with respect to a local peak in the signal amplitude relative to the background level [7] . Supplementing this procedure [8] uses an additional exclusion zone to prevent filament overlap. Another approach is to define a selection criterion as the

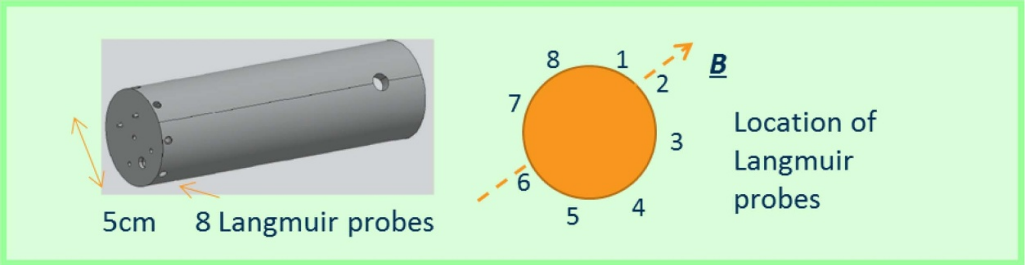

Figure 1. The left figure shows the outer shield of the MAST Gunderstrup probe indicating locations of the pins for 8 Langmuir probes arranged in a ring. The right figure indicates the labelling of the Langmuir probes and relative to the magnetic field vector.

existence of filaments within a bounded amplitude interval [9, 10]. Based on the threshold detection, studies of filamentary dynamics have been carried out using 1-D time-series data from Langmuir probes (see for example [7]), blob tracking with 2D Langmuir array data [11] and the identification of filaments in 3D turbulence simulations [12, 13]. A different approach proposed by [14] regards the measured signal as an additive sum of contributions from filaments, each with an identical pulse shape plus a Gaussian noise component. Conditional averaging sampling (CAS) [15] regards the edge turbulence fluctuations as a superposition of coherent structures with additive zero-mean random and time independent components and has been widely used for analysis of 1-D time-traces of Langmuir probe data [7, 8, 16–25]. CAS has also been applied to the analysis of 2D blob structures obtained from Langmuir probe data [10] and 3D structures computed by turbulence codes [13]. It provides compelling evidence to support the view that the ion saturation current signal is dominated by large amplitude bursts [24] and has been used as an intermediate metric to compare with theoretical models [21]. Experimental evidence [26–28] supports theoretical developments [29] regarding the coherent turbulence as similar structures with a uniform random temporal distribution consistent with a Poisson process; this work has been subsequently extended to include an attenuation source draining the filaments due to parallel generic losses [30, 31]. There is evidence of variation in the wave shape of filaments; usually the leading edge rises faster than the trailing edge [16, 27], but there are examples where the waveforms are symmetric [21]. Nevertheless, the application of CAS has a number of disadvantages. One of these is that whilst seeking to determine the averaged filament characteristics, it discards the statistical variability between filaments. Another problem is that the filaments used in CAS are generally obtained via a threshold identification procedure based on a criterion which can be difficult to justify. In addition, the overlap of filaments can reduce the asymmetry between leading and trailing edges [27].

This paper addresses the issues relating to the use of CAS and threshold detection by implementing a data-driven model describing filamentary dynamics using the principles of Bayesian Inference. Similar to [27, 31], the edge turbulence data is regarded as a superposition of filaments, with an asymmetric pulse shape. However, in this case, there is no restriction enforcing uniform waveforms for the filaments, and the minimum amplitude threshold providing a condition for

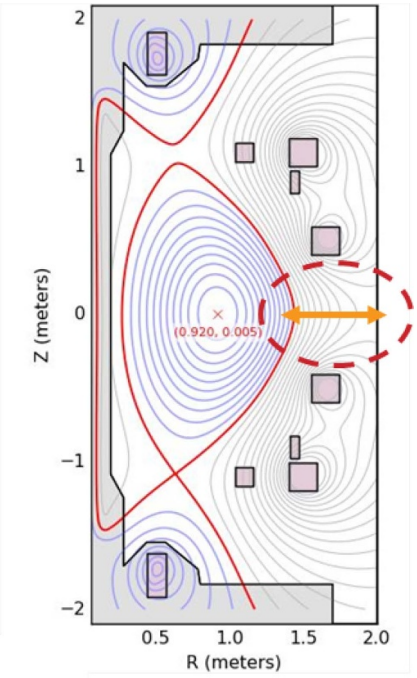

Figure 2. Contours of equilibrium poloidal flux with the location of the reciprocating Gunderstrup probe indicated. The position of the magnetic axis is indicated in the figure.

existence of filaments is relaxed. As the main purpose of this work is to demonstrate the applicability of Bayesian Inference to the study of filamentary dynamics, analysis concentrates on the time-trace data from a single Langmuir probe; it is anticipated that future work will extend the analysis to include more data signals. Section 2 describes the experimental setup. Section 3 describes a Bayesian model of the filamentary dynamics. Section 4 describes results from single Bayesian inference calculations corresponding to fixed positions and section 5 considers radial dependencies as the probe move through the SOL and into the edge of the plasma core (henceforth this region will be referred to simply as the *plasma core*).

2. Experimental description

Results are presented from the Mega Ampere Spherical Tokamak (MAST) [32] designed to study low aspect ratio ($R/a \sim$

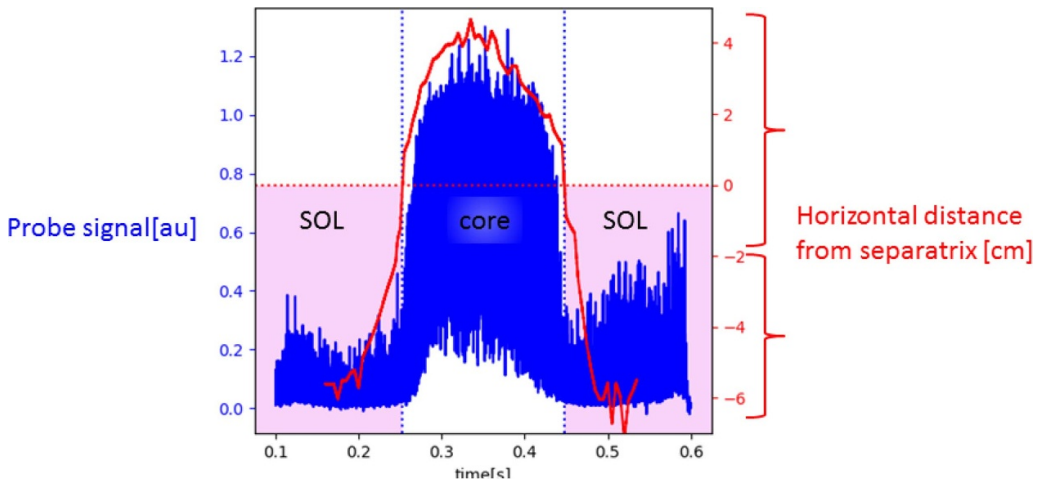

Figure 3. Time series of the ion saturation current fluctuations of Langmuir probe 5 during a single reciprocation of the Gundestrup probe assembly. The position of the probe (in red) is take from the equilibrium reconstruction and should be regarded as indicative.

0.85 m/0.65 m = 1.3), highly elongated ($\kappa = 2$) plasma configurations with plasma current, $I_p \leq$ 1.5MA and toroidal field at the tokamak geometric axis, $B_t \leq$ 0.52 T). The analysis uses experimental data from the connected double null discharge 21 712, during an ohmic L-mode phase (the same dataset is used by [21]). The core plasma conditions has a plasma current of Ip = 400 kA, a confining magnetic field of $B_T = 0.4$ T at R = 0.66 m (the resulting safety factor is $q_{95} \approx 6.2$). During the flat-top phase, the core electron temperature is $T_e \approx 650$ eV and the line-averaged density $n \approx 1.7 \times 10^{19} \mathrm{m}^{-3}$. The measurement data analysed were obtained with a Gundestrup probe [33] containing 8 separate Langmuir probes equally positioned at the end of a cylindrical assembly (see figure 1). The probe was installed on the outboard mid-plane and reciprocated a distance of 10 cm across the scrape-off-layer to just within the separatrix and back during the flat-top phase, acquiring data at 500 kHz (see figure 2). Each Langmuir probe was biased to -200 V in order to measure the ion saturation current, $I_{sat} \sim n\sqrt{T_e}$. Results presented are uncalibrated, denoted in graphs as *au* (i.e. *arbitrary units*), because we would need the collection angle of the probe to get a physically meaningful quantity. Figure 3 plots the time dependence of the signal from Langmuir probe 5. The radial position of the probe relative to the separatrix position is taken from the magnetic equilibrium reconstruction, and should be regarded as indicative. The analysis carried out in this paper is based on the data from Langmuir probe 5 in the time range $0.2 < t[s] < 0.3$ when the probe assembly is travelling through the SOL and into the the closed field line region in the edge of the plasma core. The horizontal position of the probe relative to the separatrix for specific time ranges referred to later in this paper are shown in table 1.

Table 1. Horizontal distance from separatrix of the Gunderstrup probe at specific time intervals (negative values are when the probe is within the separatrix). The values are taken from the magnetic equilibrium reconstruction and should be regarded as indicative.

time range [ms]	distance from separatrix[cm]
$240 < t < 241$	2.51 ± 0.03
$242.8 < t < 243.2$	2.308 ± 0.01
$280 < t < 280.4$	-2.89 ± 0.02
$280 < t < 281$	2.90 ± 0.03
$195 < t < 210$	4.9 ± 0.3
$218 < t < 228$	3.8 ± 0.3
$249 < t < 255$	0.2 ± 1.1
$255 < t < 261$	-1.1 ± 0.2
$261 < t < 263$	-1.4 ± 0.1
$263 < t < 265$	-1.7 ± 0.1
$265 < t < 267$	-1.8 ± 0.2
$267 < t < 269$	-1.9 ± 0.2
$269 < t < 273$	-2.2 ± 0.3
$273 < t < 280$	-2.7 ± 0.2

3. Bayesian model

In order to study the characteristics of the fluctuations at different radial positions, the time-series plot shown in figure 3 was partitioned into 263 subsets each with $n_d = 200$ data points of duration 0.4 ms in the time range $0.1948 < t[s] < 0.3$ during when the probe is moving through the SOL and into the plasma core. Figure 4 shows details of the signal when the probe is (i) in the far SOL and (ii) in the plasma core. By inspection, the signal can be described as the summation of a fluctuating component and a constant component. Both the constant and fluctuating components are larger in the plasma core than in the far SOL. The average signals in the SOL and plasma core are 0.059 and 0.601 respectively; the maximum fluctuation amplitudes in the SOL and plasma core are 0.263 and 1.03 respectively. From these we determine that the ratio *peak fluctuation amplitude:background* increases from a value of 1.7 to 4.5 between the plasma core and the far SOL.

The time-dependent behaviour shown in figure 4 suggests a model that assumes the signal is composed of a superposed set of time-displaced filaments above a constant background. For simplicity, each filament is assumed to have a positive amplitude; previous work has demonstrated that superposition

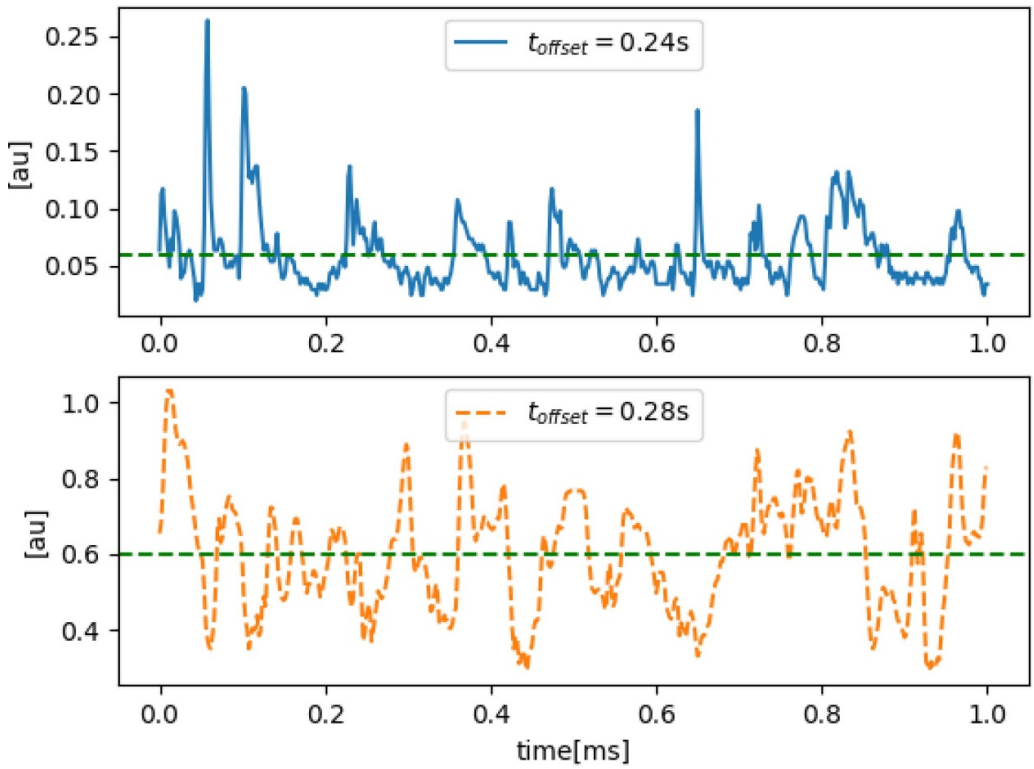

Figure 4. Ion saturation current data from Langmuir probe 5: (a) top probe is in outer SOL; (b) bottom probe is in core plasma. The average signal value in each case is denoted by a green dashed line. The t_{offset} refers to the position in the time-series from which the data is taken (see figure 3). For indicative radial positions see table 1.

of closely-spaced filaments can give the impression of filaments with negative amplitude [29]. Each filament is represented with two exponential functions for the leading and trailing edges. Denoting the rise time, fall time, the time of maximum filament amplitude and, the amplitude of the i^{th} filament as $\tau_1^{(i)}$, $\tau_2^{(i)}$, $t_0^{(i)}$, and $A^{(i)}$ respectively, the computed signal is the superposition of the background and N filaments:

$$f(t) = B + \sum_{i=1}^{N} A^{(i)} \left(\exp\left[-\frac{t_0^{(i)} - t}{\tau_1^{(i)}} \right] H(t_0^{(i)} - t) + \exp\left[-\frac{t - t_0^{(i)}}{\tau_2^{(i)}} \right] H(t - t_0^{(i)}) \right) \quad (1)$$

where H is the Heaviside function in which $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x \geq 0$. In each data subset, the set of times $\{t_0^{(i)}\}$ for the N filaments correspond to local maxima in the data, which for the i^{th} point centred at the m^{th} data value has the property for three consecutive data points $\{d_{m-1}, d_m, d_{m+1}\}$ that $(d_{m-1} - d_m)(d_{m+1} - d_m) > 0$. This choice effectively sets a minimum threshold at the data digitisation level. During the course of carrying out the current work, different strategies for defining filaments were chosen. In the end it was decided to use the above definition as the one guaranteed to identify the most number of filaments. If there was more information available from other diagnostics, it would be possible to be more discriminating. However, as will be described in the following section, when using this definition it is possible to apply a filter based on the statistics of a Bayesian inference calculation to discriminate between two different classes of of filaments.

The complete model has 3 N+1 free parameters and the task is to solve the system consistent with the n_d data points, D. This is a non-linear inverse problem, which is not guaranteed to fulfill Hadamard's postulates [34] of well-posedness (These are the following conditions: (i) a solution exists; (ii) the solution is unique and (iii) the solution's behaviour changes continuously with the initial conditions.). Determining a practical solution usually involves invoking a regularisation, however, issues relating to uniqueness and algorithmic stability can remain [35]. An alternative approach adopted in this article is to apply Bayesian inference. The deterministic problem is replaced by a statistical problem for which multiple solutions are anticipated. The Bayesian approach has been shown to be well-posed for many classes of inverse problem [35–37]. The method has an additional attraction in the context of data analysis because the model is defined explicitly together with any approximations, which would conventionally be contained in regularisations.

The method of Bayesian inference arises from the Bayes' rule:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \quad (2)$$

where D and H are multivariate distributions for the data and the model parameters respectively; $P(H|D)$ and $P(D|H)$ are conditional probabilities, meaning the 'probability of H given

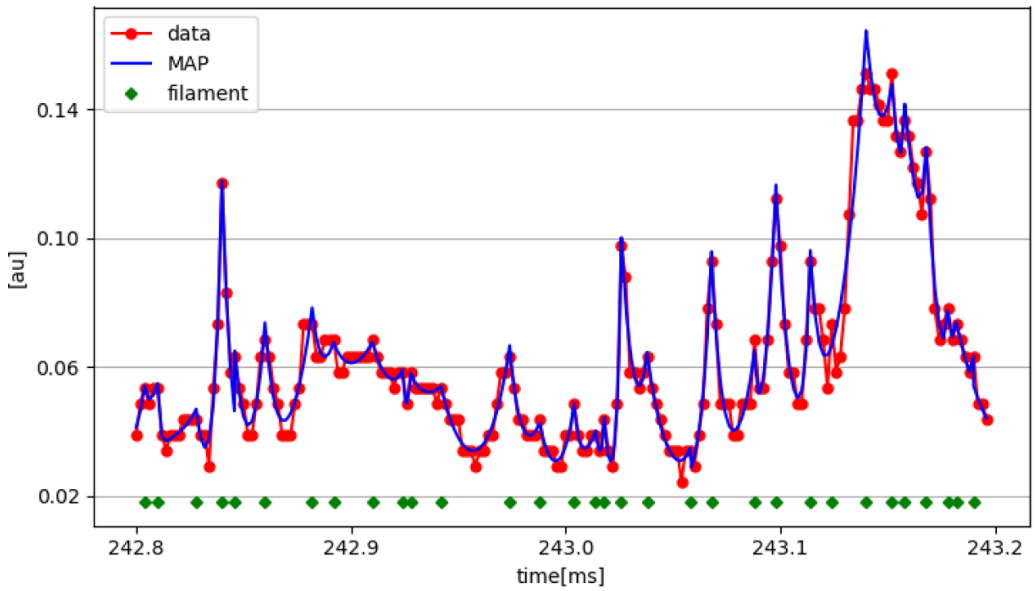

Figure 5. Experimental data and computed values using MAP parameter values in the SOL. The green diamonds are the locations of the filaments identified according to the criterion defined in section 3. For indicative radial positions see table 1.

D' and the 'probability of D given H' respectively. The equation is a manifestation of the joint distribution $P(D,H)$ and shows that the reciprocal conditional probabilities $P(D|H)$ and $P(H|D)$ (known as the *likelihood* and *posterior* respectively) are related via a scaling of the probabilities P(D) and P(H) (known as the *evidence* and the *prior*). Equation (2) is key to obtaining inverse solutions as it enables the computation of model parameters from a set of observations. Excluding the denominator term $P(D)$ which is independent of the model parameters and consequently unimportant unless ranking models (commonly referred to as model selection [38, 39]) yields the proportionality condition:

$$P(H|D) \quad \propto \quad P(D|H)P(H) \tag{3}$$

The left hand term $P(H|D)$ is the posterior distribution. In this paper we take the proportionality constant to be unity, i.e.

$$P(H|D) \quad = \quad P(D|H)P(H) \tag{4}$$

The maximum of $P(H|D)$ is known as the the *Maximum a Posteriori* (MAP) value; in the case where the distribution $P(D|H)$ and $P(H)$ are multivariate Normal distributions, the MAP solution coincides with the least-squares solution [35]. This is straightforward to demonstrate for a uniform prior and a normally distributed likelihood with a diagonal covariance matrix. In this case

$$\ln P(H|D) = \ln P(D|H) = -n_d \ln \sqrt{2\pi}\sigma - \sum_{k=1}^{n_d} \frac{(d_k - f(t_k))^2}{2\sigma^2} \tag{5}$$

where σ is the *model error*, taken to be constant for all datum samples. It is evident that the MAP solution occurs when the second term in the equation is minimized; this is the well known ''least-squares'' condition.

The conventional interpretation of σ is to regard it as the error or *uncertainty* in the data. However, in a Bayesian context the situation is more subtle. The level of agreement between model and data values is determined by the quality of the model. This has two important consequences. Firstly, an inadequate agreement between model and data motivates model enhancements. Secondly, the value of σ must reflect the ability of the model to fit the data and as this is not known prior to computing the inference, σ should be included as an additional degree of freedom, increasing the total number of free parameters to 3 N+2. A further advantage of including σ as a free parameter is that the value of σ inferred automatically ensures that the model will not overfit the data.

The Bayesian inference problem has been implemented in the Minerva Bayesian analysis framework [40]. Details are described in the appendix. The model calculations carried out in this paper compute inferences using measurement data partitioned into 0.4 ms segments. The number of filaments (N) identified at the outset of each simulation is typically around 30.

4. Filamentary characteristics

Figure 5 plots the measured and computed signals using the MAP parameters of a single Bayesian Inference calculation for the data subset $0.2428 < t[s] < 0.2432$ (at this time the probe is in the SOL). The MAP parameters were computed using the method of Hooke and Jeeves as explained in the appendix. The number of filaments identified is 32 (this is the number of local maxima in accordance with the criterion stated in section 3) and there is good agreement between measured and computed signals. For this time range, the background $B = 0$, and the model error is $\sigma = 6 \times 10^{-3}$ corresponding to ~ 10% of the average signal ~ 0.06 (see figure 4). Over all data subsets, the MAP background signal is observed to be

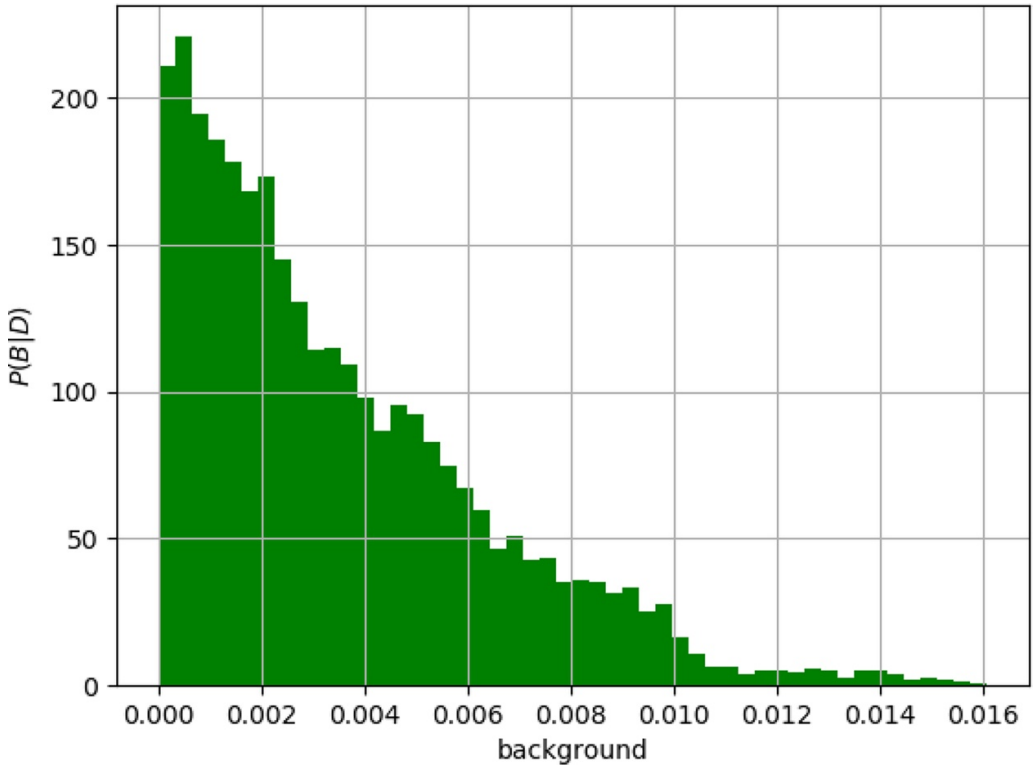

Figure 6. The distribution $P(B|D)$ assembled with 10^4 MCMC samples for the data subset $0.242\,8 < t[s] < 0.243\,2$ (probe is in SOL region).

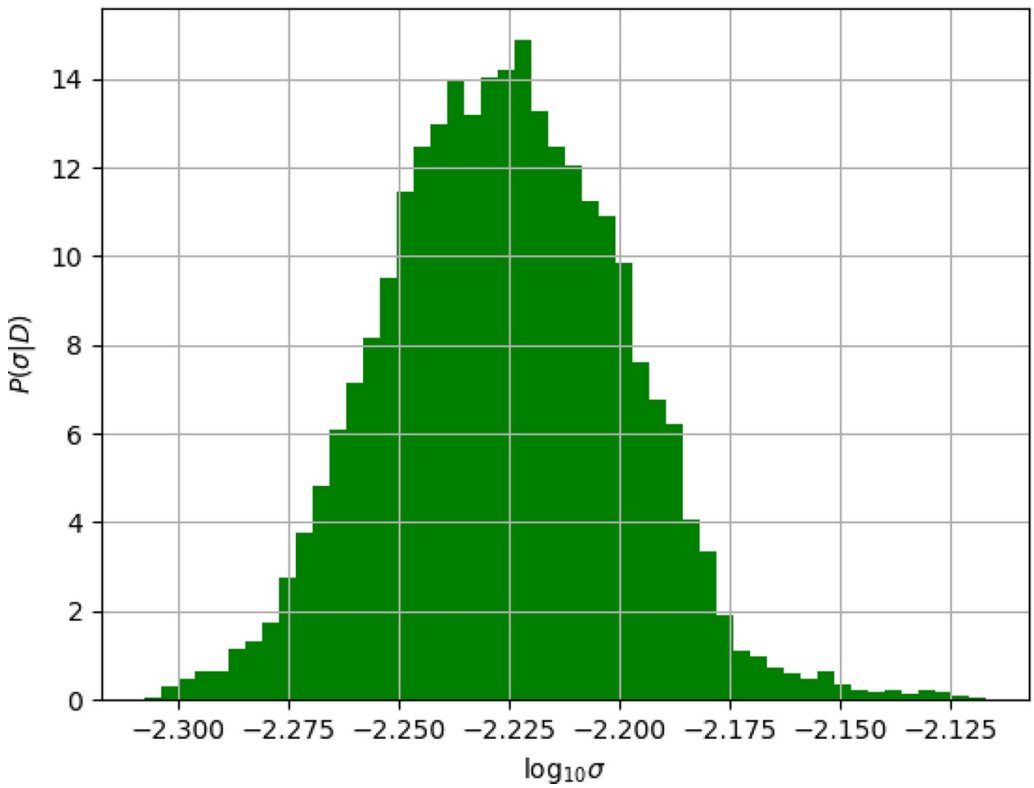

Figure 7. The distribution $P(\sigma|D)$ assembled with 10^4 MCMC samples for the data subset $0.242\,8 < t[s] < 0.243\,2$ (probe is in the SOL region).

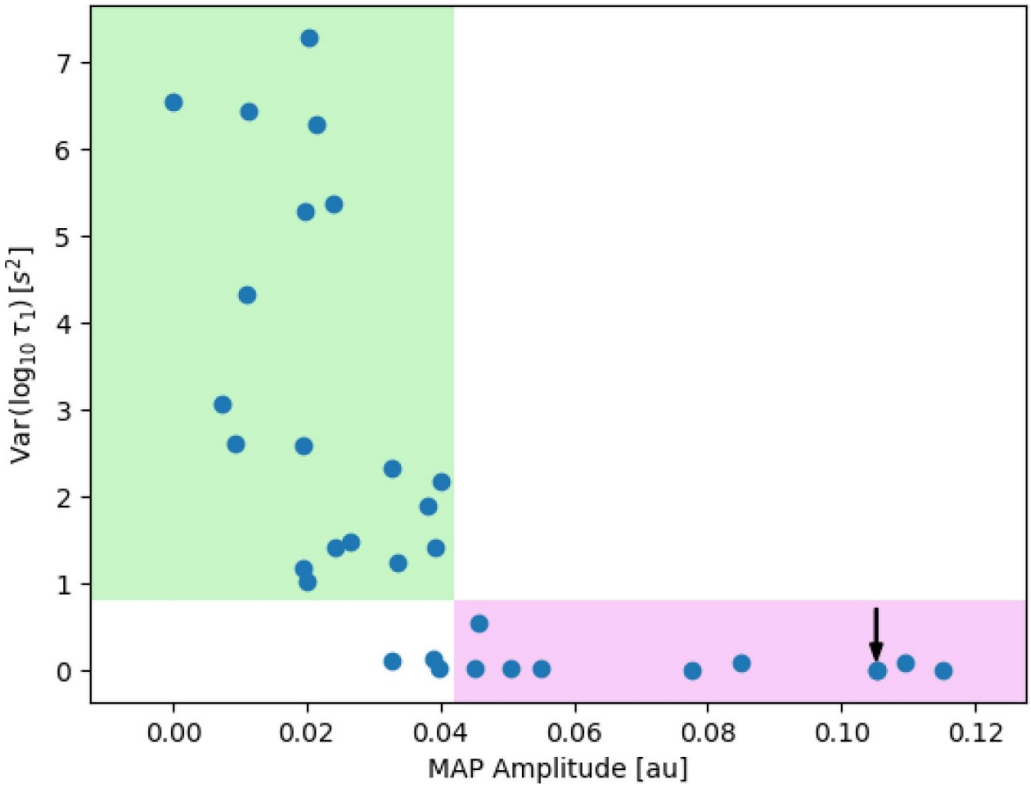

Figure 8. Variance of sampled values of $log_{10}\tau_1$ for each filament in the data subset $0.2428 < t[s] < 0.2432$ plotted against its respective MAP amplitude (the probe is in the SOL). The results are for 32 separate filaments; the arrow indicates results of two filaments that are indistinguishable on the plot. The shaded regions indicate (i) filaments with large amplitude and small variance (violet) (ii) filaments with small amplitude and large variance (green).

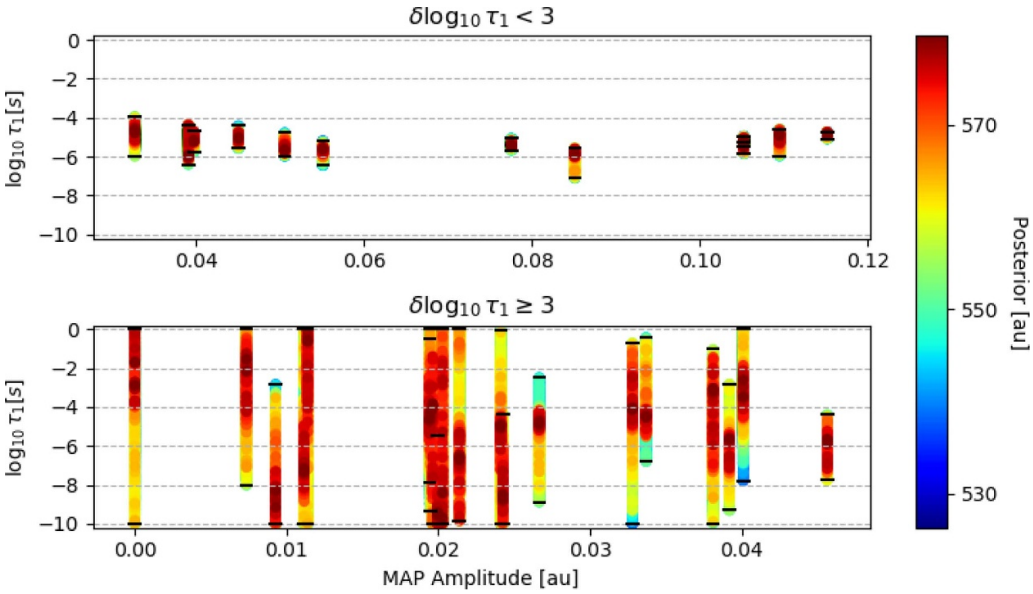

Figure 9. Samples from the distribution $P(\log_{10}\tau_1^{(i)}|D)$ for each of 32 filament in the data subset $0.2428 < t[s] < 0.2432$ plotted against its respective MAP amplitude (the probe is in the SOL). The parameter $\delta\log_{10}\tau_1$ denotes the range of sampled values of $\log_{10}\tau_1$. Results of some filaments are overlapping.

zero for 96% of cases. (This includes data subsets where the probe has crossed into the core plasma region.) Figure 6 plots the probability distribution $P(B|D)$. The figure confirms that whereas the MAP-value of the background (i.e. the value of B at the peak of the distribution), is zero there is nevertheless a small non-zero expectation value. A similar result is observed for other data subsets. The probability distribution $P(\sigma|D)$ is shown in Figure 7 and is symmetric with the majority of

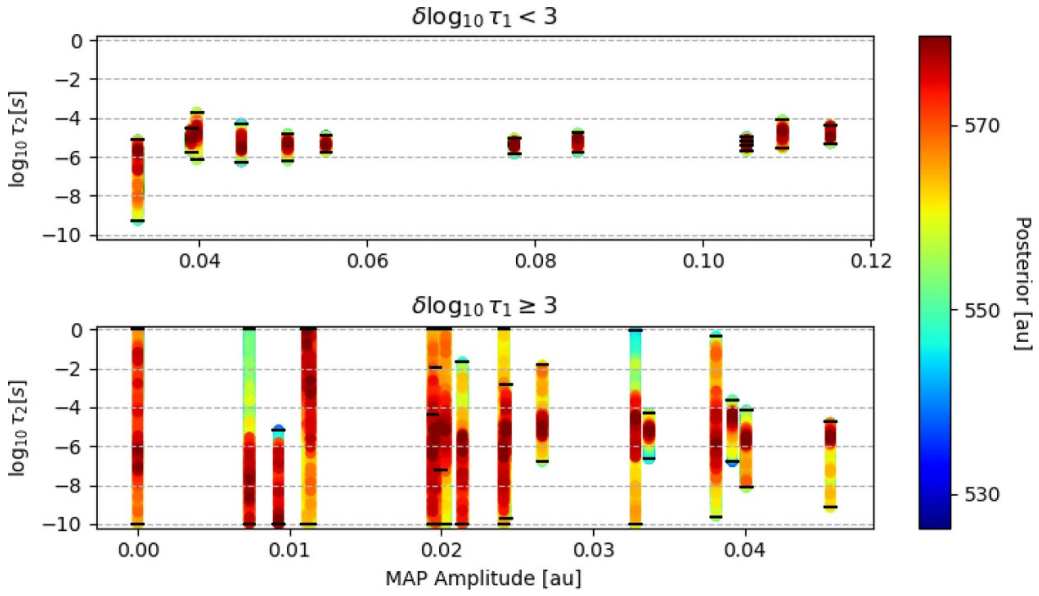

Figure 10. Samples from the distribution $P(\log_{10}\tau_2^{(i)}|D)$ for each of 32 filament in the data subset $0.2428 < t[s] < 0.2432$ plotted against the respective MAP amplitude (the probe is in the SOL). Results of some filaments are overlapping.

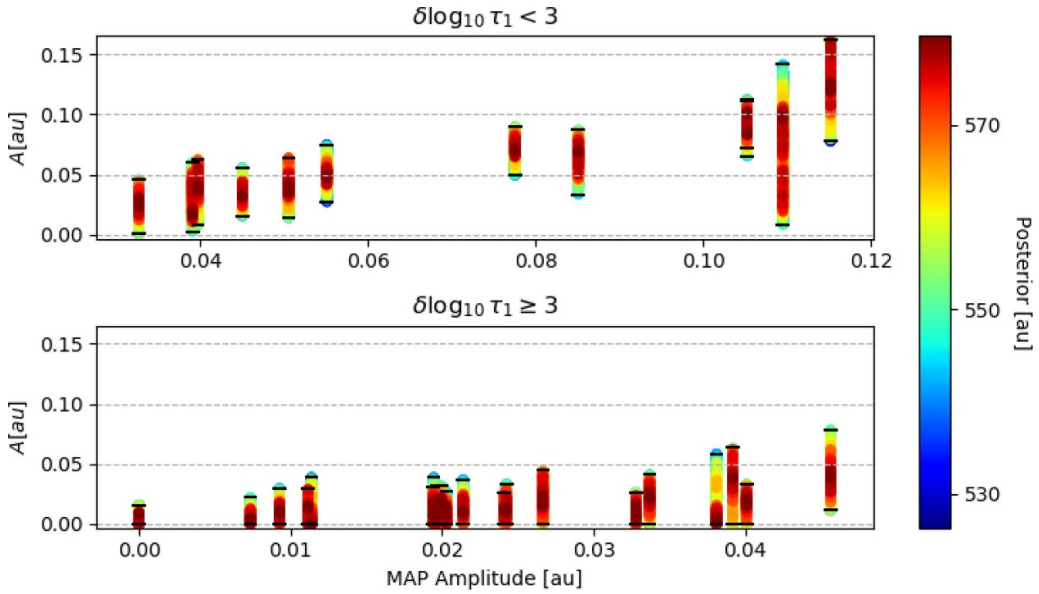

Figure 11. Samples from the distribution $P(A^{(i)}|D)$ for each of 32 filament in the data subset $0.2428 < t[s] < 0.2432$ plotted against the respective MAP amplitude (the probe is in the SOL). Results of some filaments are overlapping.

samples maintaining a value of $\sigma \sim 10\%$ of the average signal level.

To explore more fully the characteristics of the posterior, MCMC sampling (see the appendix for details) was undertaken. Figure 8 shows the variance for each of 32 filaments identified in the data subset $0.2428 < t[s] < 0.2432$ plotted versus its respective MAP amplitude. An interesting result is that the filaments cluster mainly into two classes: (i) large amplitude filaments with a small variance and (ii) small amplitude filaments with a large variance. Figure 9 plots details of the sampled distributions. The colour indicates the value of the posterior distribution of each sample with sample points having larger posterior probability values plotted over any with lower values. On the premise that the sample density is related to the value of $P(H|D)$ the colour can be regarded as an indicator of the density of sample points, however, as the samples are from a marginalised posterior distribution, the precise interpretation will depend on the details of the posterior distribution. Black vertical bars indicate the limits of the sample ranges and help to distinguish between filaments in the cases where the MAP amplitudes are similar. The figure has two separate plots in accordance with the separate classes identified in figure 8: filaments for which the range of sampled $\log_{10}\tau_1$ values (denoted as $\delta\log_{10}\tau_1$) satisfies $\delta\log_{10}\tau_1 < 3$ appear in the upper panel whilst those in which $\delta\log_{10}\tau_1 \geq 3$ are in the lower panel. The figure reveals

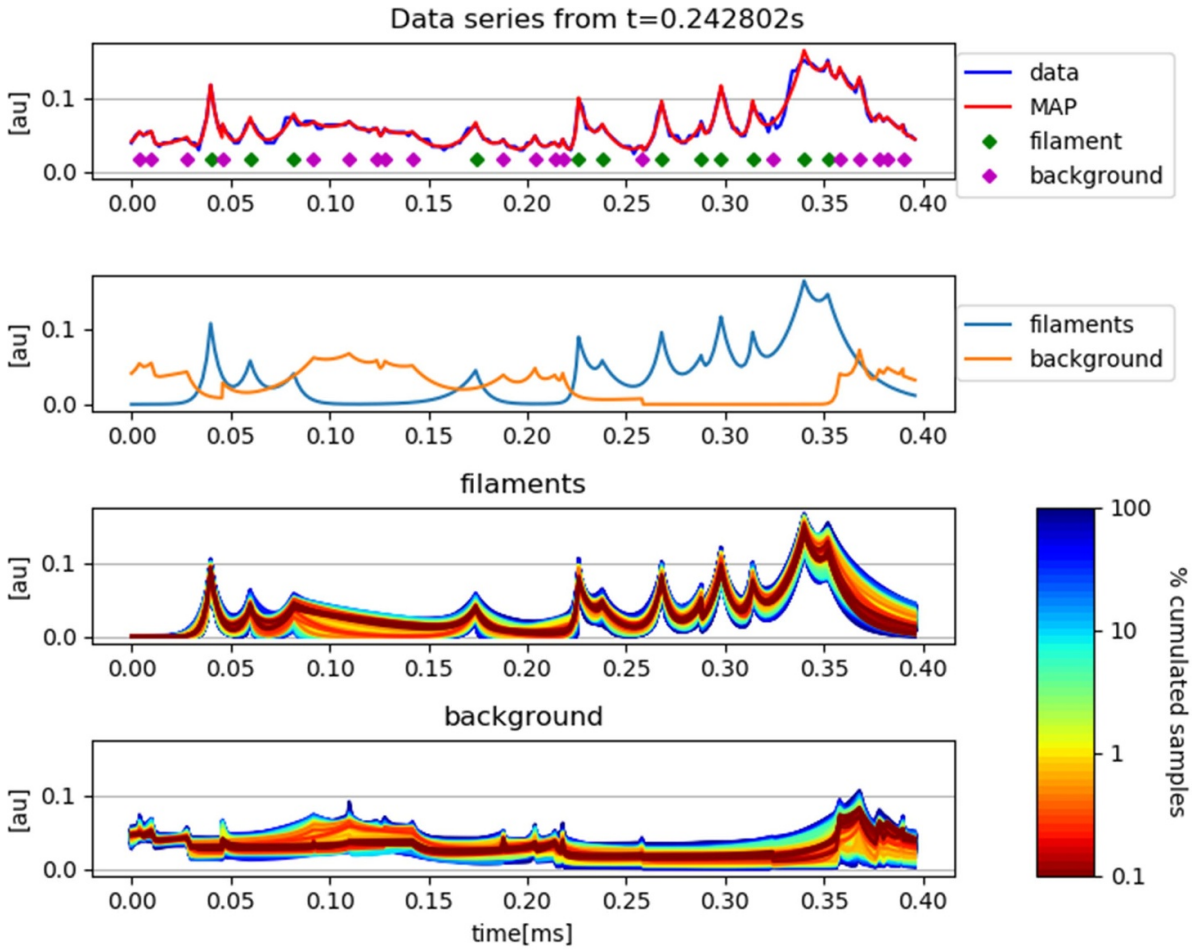

Figure 12. Interpretation in terms of plasma filaments and a fluctuating background component for the data subset $0.2428 < t[s] < 0.2432$ (the probe is in the SOL). Figures from top to bottom: (i) the plasma filaments are indicated in green and the background filaments in purple. The measured data and computed signal using MAP parameters are shown; (ii) The computed signal for the filaments and the background fluctuations using MAP parameters are shown; (iii) and (iv) multiple computed signals for the filaments and background fluctuations over-plotted using MCMC sampled parameter values; calculations using parameters with high probability appear in preference to those with lower probability.

that for larger amplitude filaments $-6 \lessapprox \log_{10}\tau_1 \lessapprox -4$, whilst smaller amplitude filaments have a considerably larger range of $\log_{10}\tau_1$, in many cases extending to the limits of the prior distribution. Figure 10 is a similar plot for $\log_{10}\tau_2$. (The placement of the filaments whether in the upper or lower panel is still determined by the variation of $\log_{10}\tau_1$.) As before, the variation of $\log_{10}\tau_2$ correlates strongly with the filament MAP amplitude. Furthermore, the rise (τ_1) and fall (τ_2) time scales are in the same range suggesting that the filaments are essentially symmetric; this result concurs with previous analyses of filaments on MAST [21]. Figure 11 is a similar plot for the sampled filament amplitude. (The placement of the filaments whether in upper or lower panel is still determined by the variation of $\log_{10}\tau_1$.) The results indicate that filaments with small MAP amplitudes mainly have sampled amplitude in the range $0 \leq A \lessapprox 2A_{MAP}$, ie the fractional uncertainties are ~ 100% and include the possibility of non-existence (defined as the range of sampled amplitudes divided by A_{MAP}). In contrast, larger amplitude filaments mainly have smaller fractional uncertainties apart from filaments that are closely spaced in time.

Previous studies on MAST provide good evidence that filaments are structures that have been ejected from the plasma core and travel at a relatively constant velocity across flux surfaces [41]. Furthermore, the filaments are presumed to pass through and interact with a quiescent plasma background [42]. However, the interpretation of experimental data often involves applying a smoothing function, for example using the method of time-window averaging [43], and the results of Bayesian inference calculations do not provide compelling evidence that there is a finite quiescent background. An alternative possibility is that the background itself is fluctuating. The difference between the plasma filaments and the background fluctuations would be in their structure and directional velocity, aspects that cannot be determined from a single Langmuir probe time series. The results presented in figures 9 and 10 show that many filaments, generally those with larger amplitude, have relatively well defined structures (characterised by small variations in τ_1 and τ_2) in comparison with small amplitude filaments. One interpretation is that filaments with well-defined structures have been ejected from

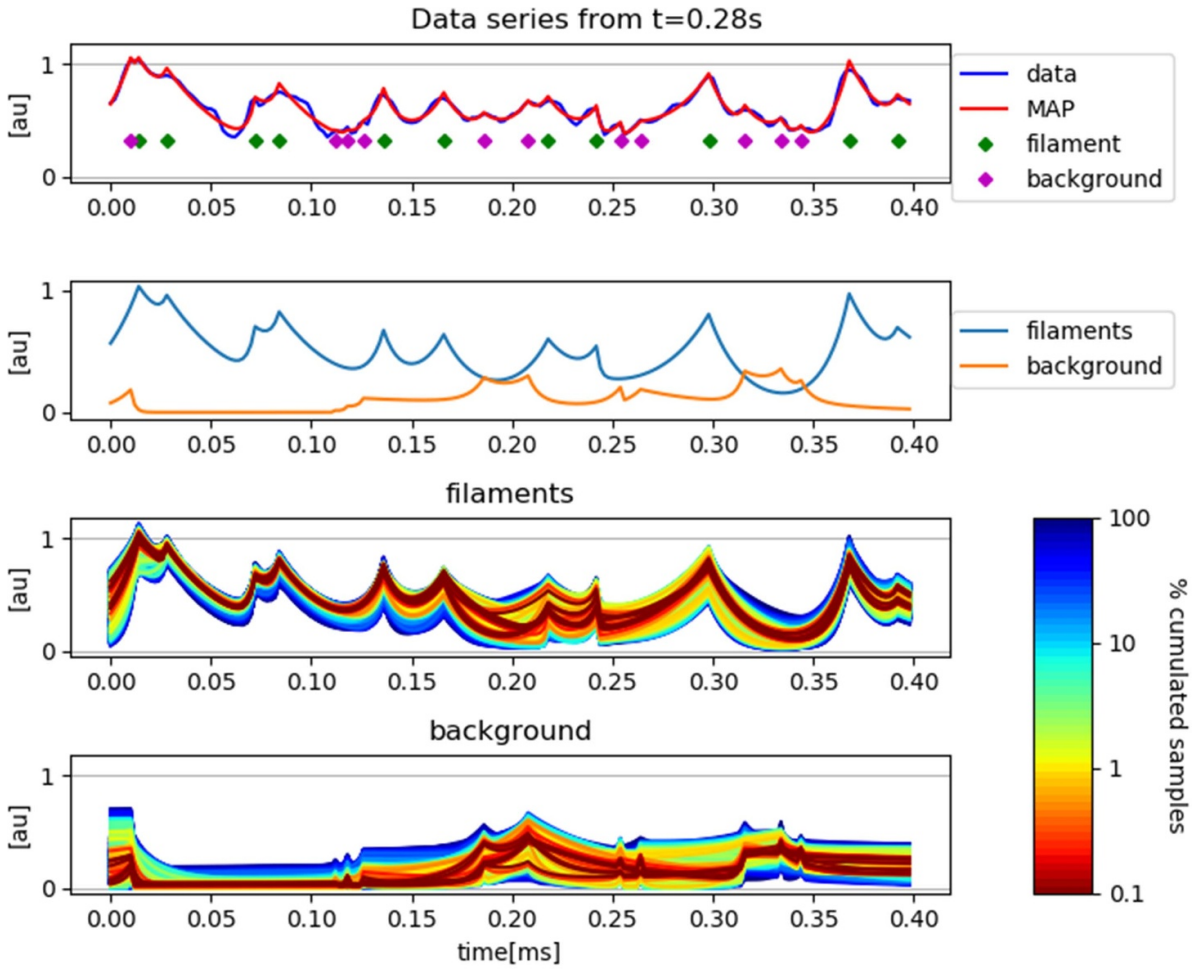

Figure 13. Details of the computed filaments and background fluctuations for the data subset $0.28 < t[s] < 0.2804$ when the probe is in the plasma core. For further explanation of the plots, refer to figure 12.

the core. *In the remainder of this paper, filaments will refer to the subset of structures identified by Bayesian Inference calculations satisfying* $\delta \log_{10} \tau_1 < 3$; *the remaining structures are considered to contribute to the background fluctuations.* Figure 12 shows the computed filaments and background fluctuations for $0.2428 < t[s] < 0.2432$ (the probe is in the SOL). The top plot shows the locations of the individual structures; also the computed and measured time-series. The second plot from the top shows the MAP computed signal components for the filaments and for the background fluctuations. The bottom two plots are the signals computed using parameter values sampled from the posterior distribution. The plots are the results of multiple calculations superimposed. The signals have been plotted in decreasing value of the posterior probability of each sample such that signals with highest probability are plotted above those with lower probability. The brown colouration of these plots indicates that the maximum uncertainty of the filament and background fluctuations is already explored using only ~ 10% of the data set with the highest values of $P(H|D)$.

The same analysis has been carried out for data subsets throughout the SOL and into the plasma core. Figure 13 shows the an example at $0.28 < t[s] < 0.2804$ when the probe is in the core plasma region. Compared to the earlier time, both filament and background fluctuation components are larger, and the filaments dominate the background fluctuations throughout the time-range. For each time range, the uncertainty in the filament and background signals are similar (this is evident from the two lower plots).

In summary, the key findings for analyses of all data subsets carried out are as follows:

(a) The MAP background is usually zero, and the expectation value from the background has a small finite value. This result is consistent with previous work that demonstrates that the background can be the constructed from the superposition of filamentary waveforms [14, 30].
(b) The model uncertainty σ is approximately10% of the average signal.
(c) The values of $\log_{10} \tau_1^{(i)} \approx \log_{10} \tau_2^{(i)}$ for each filament in a particular subset region (i.e. the filaments are symmetrical).
(d) Filaments have larger MAP amplitude values compared to the background fluctuations.
(e) We postulate that filaments are the same structures as identified in previous studies (see for example [41]); these are well defined structures that originate from the plasma core in contrast to background fluctuations which we suggest have shorter coherence lengths.

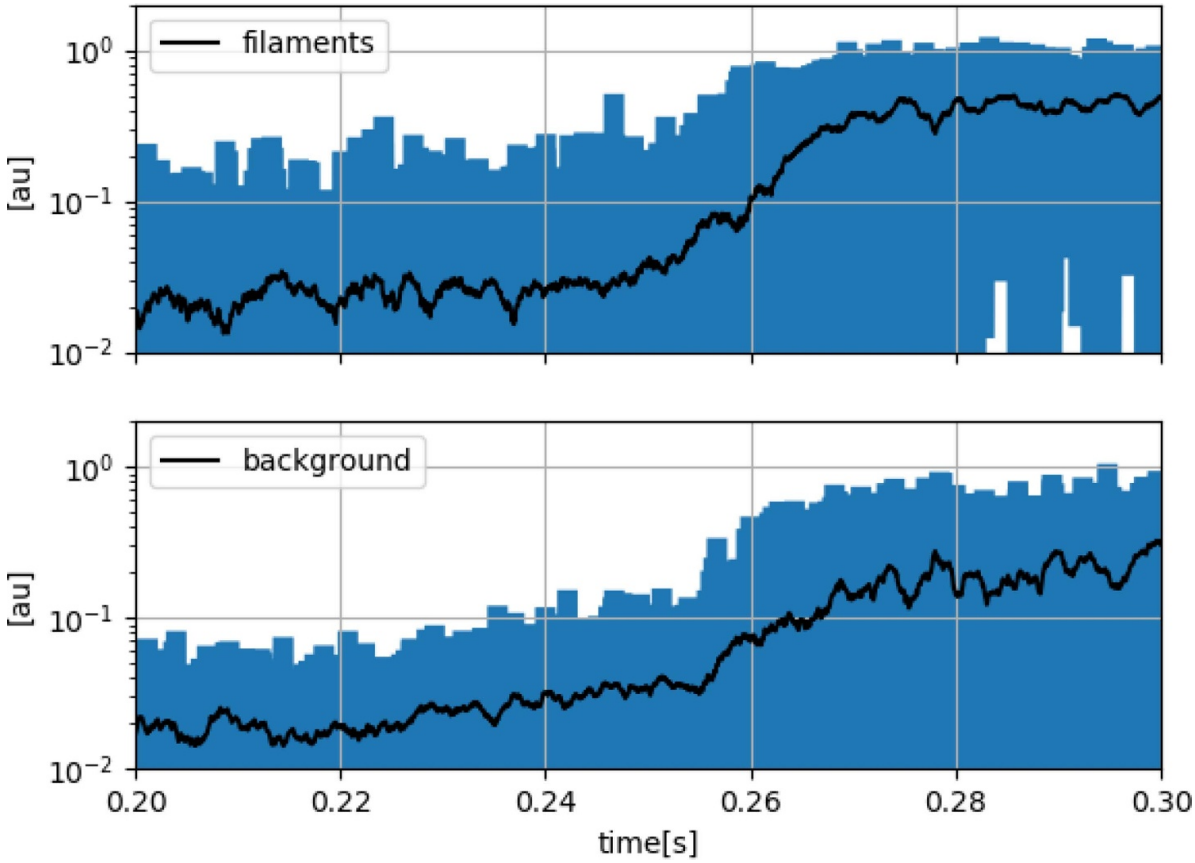

Figure 14. Time-variation of computed filament and fluctuating background amplitudes using MAP parameters. The probe is in the SOL at 0.2 s moving across the separatrix and is in the plasma core at 0.3 s. The black lines are the average values, and the shaded regions indicate the range of amplitudes in time-moving subsamples of 2 ms duration.

Separation of the filaments into two components, (filaments and a fluctuating background) is an interpretation of the model results. The underlying hypothesis is that the measured fluctuations arise from a different source and that the diagnostic probe is better able to discriminate the structure of the larger filaments than the background fluctuations. In future work, it will be interesting to understand whether further diagnostic information supports the underlying model; this would be apparent in a reduction of the variance of the posterior.

5. Radial dependencies

In this section we use the new definition of plasma filaments and background fluctuations postulated in section 4 to explore statistical variations across the SOL and into the outer plasma core region. Figure 14 plots the variation of the computed filament and background fluctuations with time as the probe moves from the SOL into the plasma core. Whereas the equilibrium reconstruction (see figure 3) indicates that the probe crosses the separatrix at ~ 0.25 s, results from the probe shown in figure 14 and subsequent figures indicates a transition change somewhat later, at ~ 0.26 s. The signals are computed using the MAP parameters and have been averaged in 2 ms time-intervals. The plots show the average and range of fluctuation amplitudes in the five 0.4 ms time-intervals that are within each 2 ms time window. It is evident that although the range of amplitudes (shown in blue) for each 2 ms time bin is large compared to the average denoted by the black line, the variation between each 2 ms time bin is noticeably smaller. Both the background and filament signals increase as the probe approaches the plasma core. Between t = 0.255 s and 0.27 s the averaged filament amplitudes increases more than the averaged background fluctuations. One interpretation for the change in amplitude is that the drain rate of filaments changes, mirroring the magnetic topology transition across the separatrix from closed field lines within the plasma core to field lines that terminate at the divertor. This is consistent with 2D numerical modeling of the filamentary dynamics in the SOL which identifies parallel transport along open field lines as a significant drain component affecting filament amplitudes [44]. Differences between the changes in the filament and background fluctuation amplitude could indicate differences in the fundamental mechanisms affecting these plasma components. As noted earlier, uncertainty in the computed magnetic configuration means that it is not possible to identify the precise location of the steepest gradient in filament amplitude at ~ 0.26 s relative to the position of the separatrix. Further evidence for the existence of a transition for the filament drain rate is provided by calculations of the probability distribution functions (pdf) of the computed filament signal at various time-intervals shown in figure 15. The shape of the

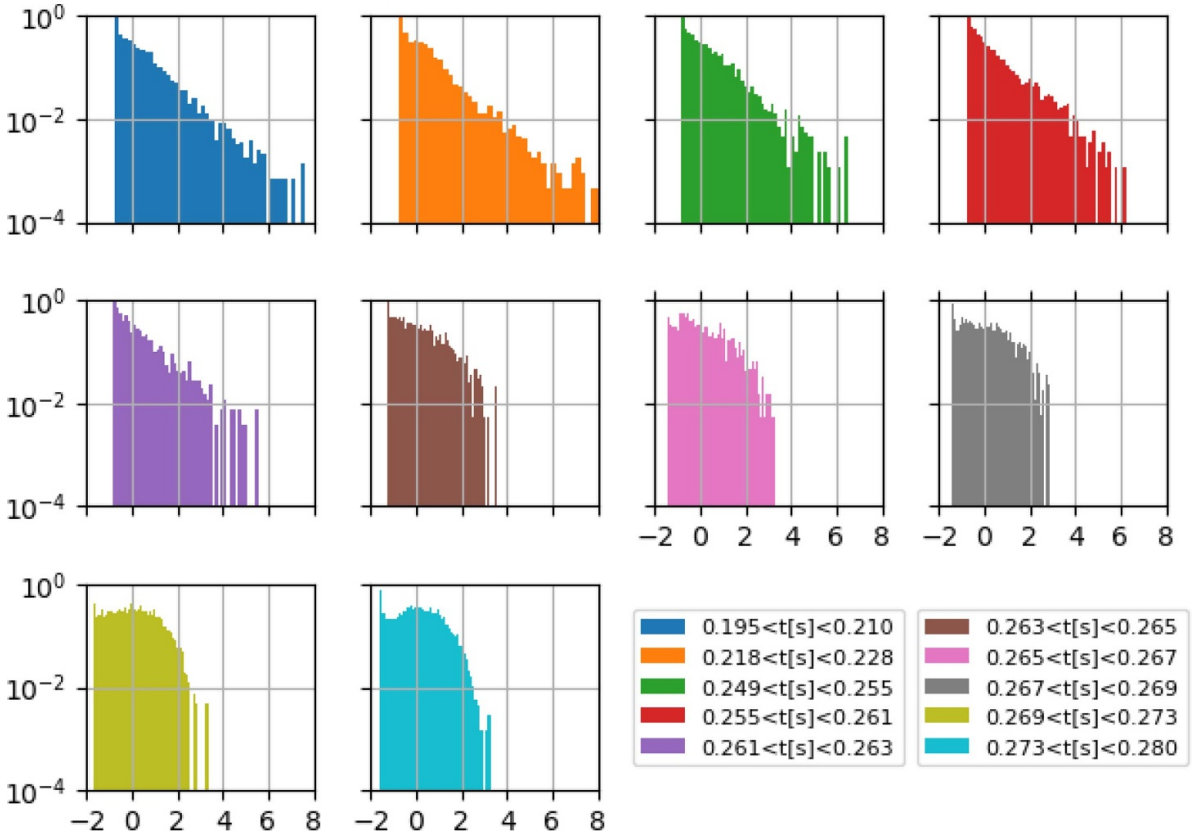

Figure 15. The pdf of the computed filament signal (f) defined as $(f- <f>)/\sigma_f$ for ten time-ranges from the SOL through to the plasma core. In this figure $<f>$ and σ_f denote respectively the time-average and standard deviation of f. Subsequent figures use the same definition with b and w replacing f . The probe is in the SOL at 0.2 s moving across the separatrix and is in the plasma core at 0.3 s. The indicative radial position corresponding to the time range of each plot is given in table 1.

pdf corresponding to different positions in the SOL during the times $0.195 < t[s] < 0.263$ are very similar. Approaching the confinement region, there is a sudden change for $t[s] > 0.263$. In comparison, the pdf of the computed background fluctuations for the same time-intervals shown in figure 16 exhibit a more gradual evolution of the pdf structure. Figure 17 plots the pdf of filament waiting times (defined as the time interval between adjacent filaments) for eight time ranges from the SOL through to the plasma core. The dashed line indicates an exponential distribution. The results indicate that the distribution of filament waiting times in the SOL (but not in the plasma core) are exponentially distributed over two decades. A similar series of plots for the background fluctuations is shown in figure 18; in this case waiting times appear to be exponentially distributed in the SOL and core plasma over two decades.

Figure 19 plots the variation in waiting times for the computed filaments and background fluctuations as the probe moves from the SOL into the plasma core. Whereas the average waiting times for background fluctuations is larger in the plasma core than in the SOL, average waiting times for filaments show very little variation with position. The maximum waiting times in both cases have some interesting features: (i) the maximum waiting times of filaments are larger in the SOL than in the core (i.e. the solid blue bars in the lower plot exhibit a rising trend from right to left); (ii) the maximum background fluctuations exhibit a reduction in the proximity of the separatrix (around t = 0.25 s); (iii) maximum waiting times of background fluctuations are less in the SOL than in the plasma core.

5.1. Discussion

Section 5 presented results for the radial variation of the computed background fluctuations and the filaments. It was observed in figure 17 that the pdf of waiting times of filaments exhibited a significant structural change in the proximity of the separatrix. Assuming that the radial filament velocity of filaments remains constant the result can be explained by the presence of a filament source in this region generating new filaments. In contrast, observed variations in the pdfs of the filaments across the plasma shown in figure 15 can be explained both by the presence of sinks (for example of parallel transport in the SOL region) and by the presence of signal sources across the plasma.

The pdf of waiting times for the background fluctuations was shown in figure 18 to be exponentially distributed throughout the plasma core and the SOL. These results are similar to those reported on other tokamaks [26, 45] and are consistent with a source generating random uncorrelated fluctuations in the core plasma and the SOL. The justification for this statement is that a probability of waiting times between

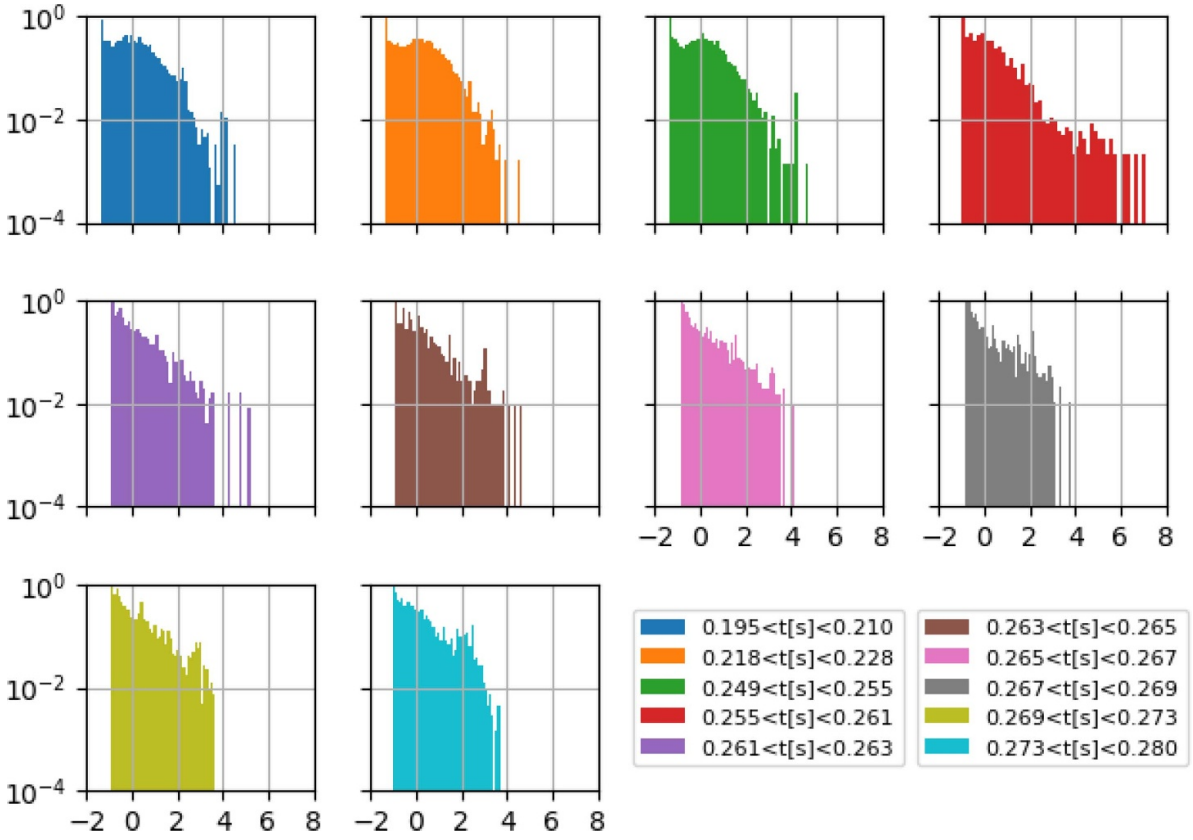

Figure 16. The pdf of the computed background signal (b) defined as $(b- < b >)/\sigma_b$ for ten time-ranges from the SOL through to the plasma core. The probe is in the SOL at 0.2 s moving across the separatrix and is in the plasma core at 0.3 s. The indicative radial position corresponding to the time range of each plot is given in table 1.

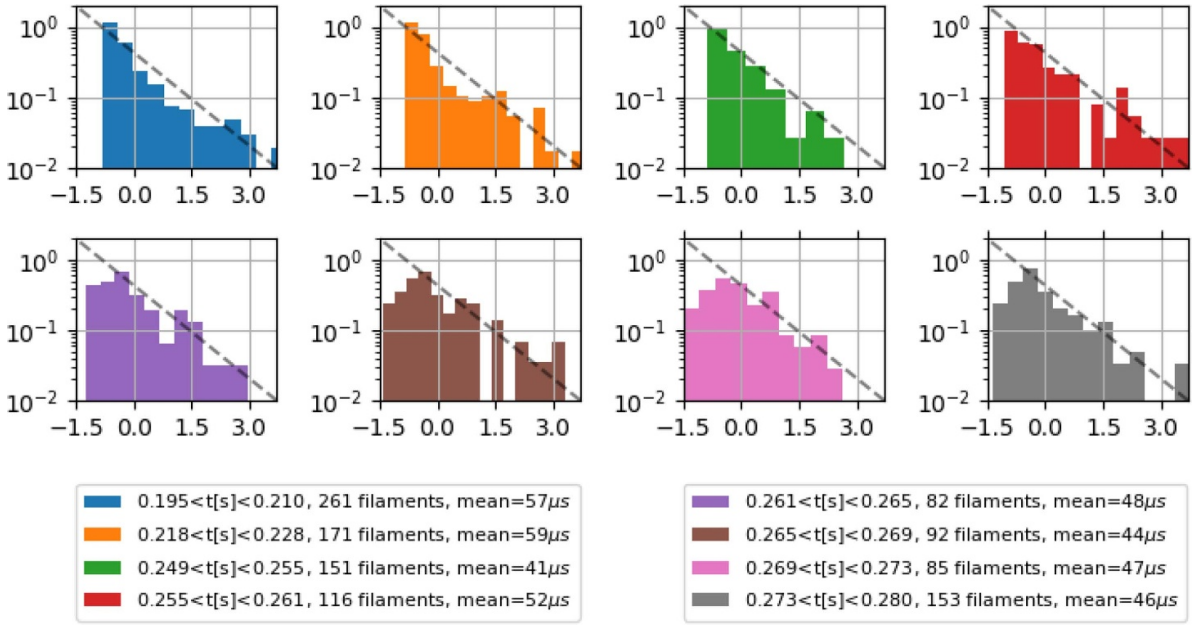

Figure 17. The pdf of the computed filament waiting times (w) defined as $(w- < w >)/\sigma_w$ for eight time-ranges from the SOL through to the plasma core. The probe is in the SOL at 0.2 s moving across the separatrix and is in the plasma core at 0.3 s. The dashed lines shows the characteristic of a distribution. The indicative radial position corresponding to the time range of each plot is given in table 1.

independent events which is exponentially distributed is the result of a classical Poisson process (of random uncorrelated events) [46, 47].

Without information about the radial velocities it is not possible to separate the source due to spontaneous generation of fluctuations and that due to radial transport of fluctuations

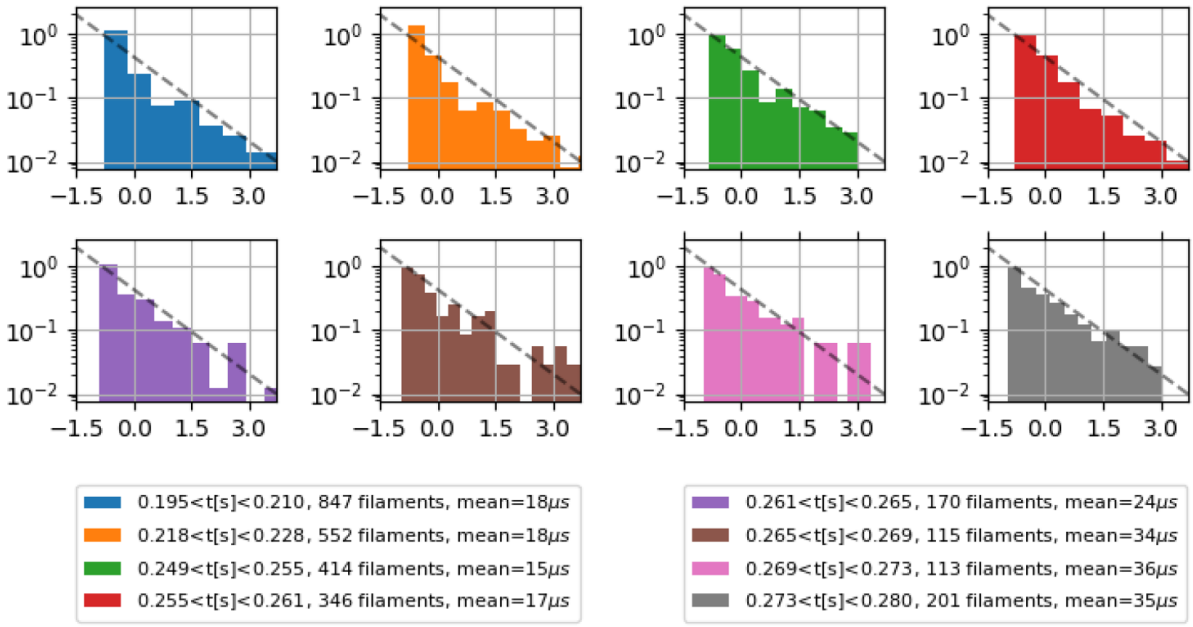

Figure 18. The pdf of the computed background fluctuation waiting times (w) defined as $(w - < w >)/\sigma_w$ for eight time-ranges from the SOL through to the plasma core. The probe is in the SOL at 0.2 s moving across the separatrix and is in the plasma core at 0.3 s. The dashed lines shows the characteristic of a exponential distribution. The indicative radial position corresponding to the time range of each plot is given in table 1.

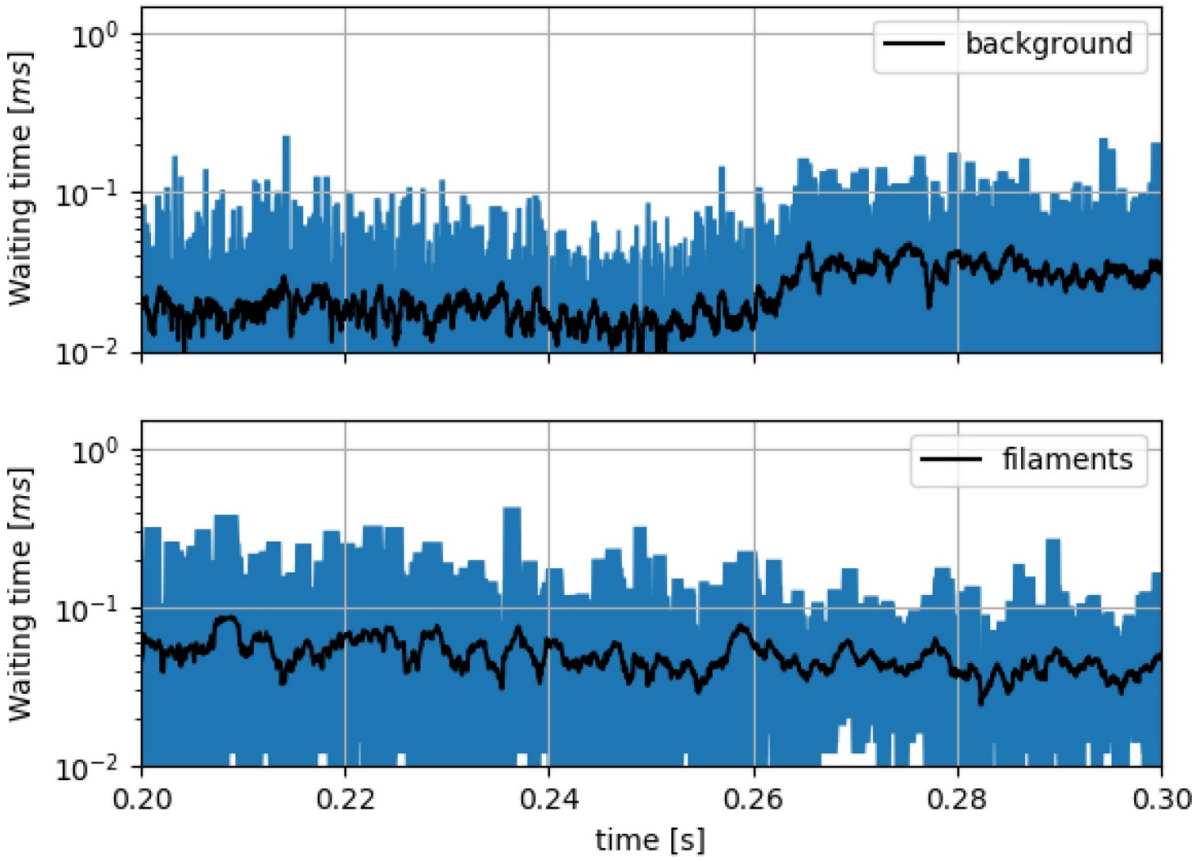

Figure 19. Time-variation of computed filament and background fluctuation waiting times. The probe is in the SOL at 0.2 s moving across the separatrix and is in the plasma core at 0.3 s. The black line are the average values, and the shaded regions indicate the range of waiting times in time-moving subsamples each containing 30 events (ie 30 waiting times).

across field lines. However, the further observation shown in figure 19 that the average waiting time for background fluctuations is constant in each region and is approximately smaller by a factor of 2 in the SOL region suggests that either the radial transport of fluctuation encounters a barrier in the proximity of the magnetic separatrix and/or the source generating fluctuations is different between the SOL and the plasma core.

The pdf of waiting times for the filaments shown in figure 17 is Exponential-like in the SOL and exhibits more symmetric Gaussian characteristics in the plasma core. The results have more scatter than those for the background, explained by the smaller number of observed events. Nevertheless the results are consistent with previous observations [21], and can be interpreted that there is a source of filaments in the proximity of the separatrix generating random uncorrelated filaments that subsequently move radial outward through the SOL region. The further observation shown in figure 19 that the average waiting time for filaments is constant across the SOL and plasma core regions is unexpected. The result may be coincidental and needs to be confirmed with more data. An observation that the average waiting time is independent of position would suggest that the dynamics in the two regions are linked and if this is true the variation in the forms of the pdf could be explained by variations in the radial velocities of the filament.

The pdfs of the filament signals shown in figure 15 are similar in form to the pdfs of the waiting times throughout the plasma core and SOL (figures 17 and 18). This suggests that changes in the structure of the pdfs are affected by differences in radial velocity and spontaneous generation of filaments at each location. In the SOL region the structure of both signal and waiting times pdf are exponentially distributed over 2 decades. The further observation shown in figure 14 that the mean filament amplitude is smaller in the SOL than in the plasma core can be explained by parallel transport. The rate of decay is less for the filaments than for the background fluctuations; this observation could be due to enhanced transport of filaments due to $E \times B$ drift [41]; furthermore, phenomena like blob spinning [2] can increase coherence and lifetime.

6. Summary and conclusions

A novel approach using Bayesian inference has been implemented to interpret the filamentary dynamics measured by a Langmuir probe fixed to a reciprocating assembly on MAST. The model describes the system as a superposition of time-displaced filaments and a fixed background component. Each filament is parameterised in terms of a characteristic rise and fall time and maximum amplitude centred on local maxima in the measured data time-series. A distinctive feature of the approach is that no minimum threshold is set for the existence of filaments. Furthermore, the model uncertainty is provided as an additional free parameter. Data is analysed in short subsampled intervals characterising fixed positions in the SOL and core plasma corresponding to the position of the reciprocating probe assembly. The results obtained achieve a fit to the model with an error of ~ 10%. The MAP background signal is found to be zero in over 95% of subsample intervals. Results of Markov chain Monte Carlo sampling of the posterior distribution provide uncertainties on all model parameters. Differences in the MAP values for the rise and fall are small compared to the computed uncertainties in these parameters. This result is in line with previously reported findings that filaments on MAST are symmetric. It is observed that whereas large amplitude filaments are well characterised in terms of rise times, smaller amplitude filaments are often unconstrained by the data and are limited by the details of the prior. Based on these findings, a new definition for the filaments is proposed based on the uncertainty in the filament rise times. The remaining filaments together with the constant background component forms a new time-dependent signal referred to as the computed *background fluctuation* signal. The characteristics of these signals (for the *plasma filaments* and for the *background fluctuations*) are reported in terms of their spatial variation as the probe moves through the SOL and into the core plasma. It is shown that the pdf of both the waiting times and signal amplitude in the SOL for the computed filaments are exponentially distributed over more than 2 decades. The pdf of waiting times for the background fluctuations are exponentially-distributed throughout the plasma core and the SOL. The mean waiting times of the background fluctuations are constant in the core plasma and SOL regions, with the magnitude in the SOL approximately 50% the magnitude in the plasma core. The mean waiting times of the filaments is constant throughout the plasma core and SOL regions even though details of the pdf change from a Exponential-like distribution in the SOL to a more symmetric distribution in the plasma core.

It is important to note that the filament-based model on which the current work is predicated is essentially an ad hoc model that is based on observations of the data rather than on theoretical foundations. Whereas it may be reasonable to suppose that the evolution of the explosive events that generate filaments could results in exponential signal components, the extension to include the background component is not substantiated. Nevertheless it is an interesting result that the fit of the model to the data is achieved to within ~ 10%. The separation of the signal components into two time-dependent fluctuating components is novel. It permits the calculation of a fluctuating background signal that is consistent with the Bayesian model, and the mean value of both background fluctuations and filaments are observed to decrease across the SOL region. Future work should establish a definition for these signals on the basis of theoretically-based constraints.

The novelty of the approach described in the paper is that it replaces the threshold for the existence of a filament based on its amplitude and waiting time with a criterion based on the probability that a filament fits a specific model. Specifically the work employs a filament model containing two exponential functions. The criterion for the existence of a filament is expressed as the probability that the filament exists with this functional form. In the future, the approach described in this work would enable a quantified assessment of different functional forms for the filaments. A further novelty of

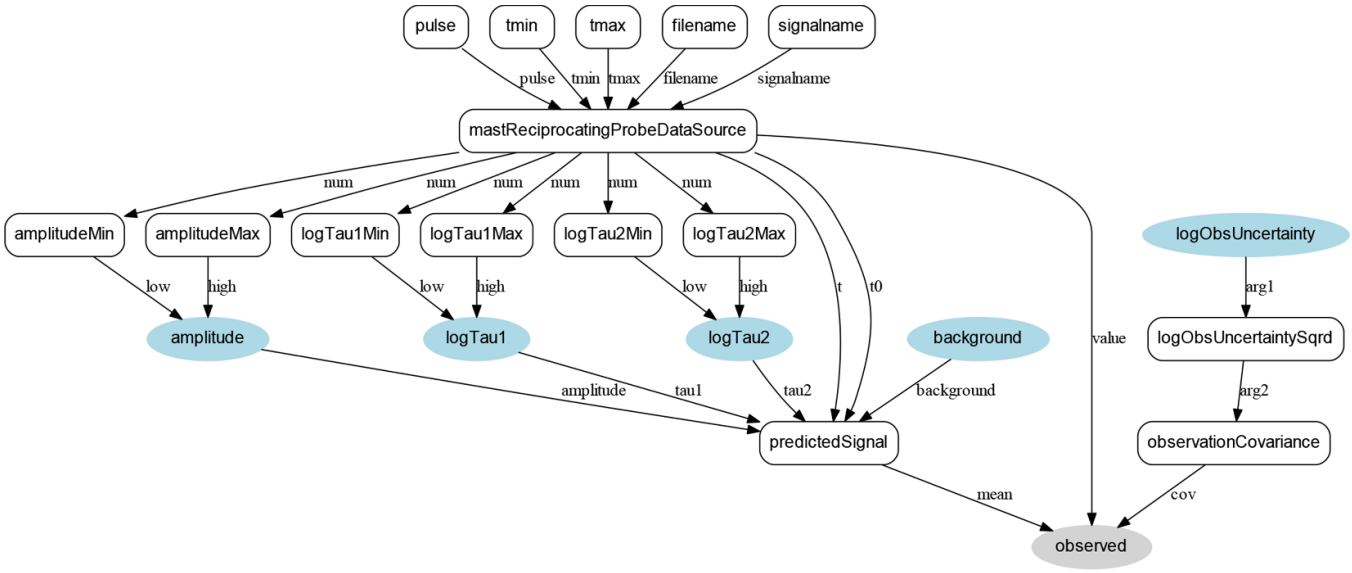

Figure A1. Graphical model employed by Bayesian inference problem.

the approach described in the paper is that the information from the background is retained as a "fluctuating background", and is not removed as is the case with the conditional averaging sampling procedure. Typical computational times on an I7 chipset are around 7 minutes for a calculation of a single Bayesian computation using a 0.4 ms data subset. This is an indicative time; execution times can be reduced using more modern chipsets and with parallellisation.

Acknowledgments

This work was funded by the RCUK Energy Programme [Grant No. EP/P012450/1]. To obtain further information on the data and models underlying this paper, whose release may be subject to commercial restrictions, please contact PublicationsManager@ukaea.uk.

Appendix A.

Figure A1 is a graphical representation of the Bayesian model implemented in Minerva. The ellipses denote probability distributions. The components of the prior are shaded in blue: $P(A)$, $P(\log_{10}\tau_1)$, $P(\log_{10}\tau_2)$ and $P(B)$ and $P(\log_{10}\sigma)$ to denote that these correspond to the free parameters in the model whereas the likelihood $P(D|H)$ is shaded in grey. All component distributions are multivariate apart from $P(B)$ and $P(\sigma)$. The arrows shown in the figure follow the conventions of a Directed Acyclic Graphical model [48] corresponding to a factorization of the joint probability distribution i.e. from the graph, the joint distribution is identified as

$$P(D,A,\log_{10}\tau_1,\log_{10}\tau_2,B\log_{10}\sigma) = P(D|H)P(A)P(\log_{10}\tau_1)P(\log_{10}\tau_2)P(B)P(\log_{10}\sigma) \quad \text{(A1)}$$

with the likelihood $P(D|H) \equiv P(D|A,\log_{10}\tau_1,\log_{10}\tau_2,B,\log_{10}\sigma)$

A detail of the model is that parameters relating to τ_1, τ_2 and σ are represented using the logarithm to the base 10. The reason is that it enables all free parameters to have similar magnitudes which generally improves computational tractability of high dimensional sampling problems. The specific choice of parameter can be compensated by appropriate transformation of the random variable distribution [49]. In the present case we employ random uniform distributions excluding negative background and amplitudes, but otherwise providing large bounds in order to minimize the information content of the prior distributions:

$$\begin{aligned} -10 &\leq \log_{10}\tau_1^{(i)}[s] \leq 0 \forall i \in \mathbb{N}, 1 \leq i \leq N \\ -10 &\leq \log_{10}\tau_2^{(i)}[s] \leq 0 \forall i \in \mathbb{N}, 1 \leq i \leq N \\ 0 &\leq A^{(i)} \leq 3 \forall i \in \mathbb{N}, 1 \leq i \leq N \\ 0 &\leq B \leq 3 \\ -4 &\leq \log_{10}\sigma \leq -1 \end{aligned}$$

The remaining elements in figure A1 are either parameters that can be set for a specific run (for example, *pulse*, *tmin* and *tmax*) or a computational step (for example *mastReciprocatingDataSource* reads and prepares data whereas *predicted* computes the signal for a particular set of parameters).

The MAP solutions are obtained using the method of Hooke and Jeeves [50]. Exploration of the posterior distribution is carried out using Markov chain Monte Carlo (MCMC) sampling [51]. The MCMC algorithm utilizes an adaptive algorithm to compute the covariance of the jump distribution which is rescaled following an initial "burn-in" phase to ensure an acceptance condition of around 0.75 [52]. The burn-in phase and subsequent posterior exploration phases each used 5×10^6 samples with 10^4 (ie 1:100) sample points stored for subsequent processing.

ORCID iDs

L C Appel https://orcid.org/0000-0001-9259-6240
F Militello https://orcid.org/0000-0002-8034-4756

References

[1] Zweben S J and Medley S S 1989 Visible imaging of edge fluctuations in the TFTR tokamak *Phys. Fluids B: Plasma Phys.* **1** 2058
[2] Myra J R, D'Ippolito D A, Stotler D P, Zweben S J, LeBlanc B P, Menard J E, Maqueda R J and Boedo J 2006 Blob birth and transport in the tokamak edge plasma: Analysis of imaging data *Plasma Phys.* **13** 092509
[3] Kirk A *et al* 2006 The MAST team. Filament structures at the plasma edge on MAST *Plasma Phys. Control. Fusion* **48** B433
[4] Zweben S J and Gould R W 1983 Scaling of edge-plasma turbulence in the Caltech tokamak *Nucl. Fusion* **23** 1625
[5] Zweben S J and Gould R W 1985 Structure of edge-plasma turbulence in the Caltech tokamak *Nucl. Fusion* **25** pg171
[6] Carreras B A 2000 Intermittency of plasma edge fluctuation data: Multifractal analysis *Phys. Plasmas* **7** 3278
[7] Garcia O E *et al* 2007 Fluctuations and transport in the tcv scrape-off layer *Nucl. Fusion* **47** 667
[8] Carralero D *et al* 2014 The ASDEX Upgrade Team. An experimental investigation of the high density transition of the scrape-off layer transport in ASDEX Upgrade *Nucl. Fusion* **54** 123005
[9] Furno I *et al* 2008 Mechanism for blob generation in the TORPEX toroidal plasma *Phys. Plasmas* **15** 055903
[10] Fedorczak N, Gallo A, Tamain P, Bufferand H, Ciraolo G and Ghendrih P 2018 On the dynamics of blobs in scrape-off layer plasma: Model validation from two-dimensional simulations and experiments in Tore Supra *Contrib. Plasma Physics* **58** 471
[11] Müller S H, Diallo A, Fasoli A, Furno I, Labit B, Plyushchev G, Podestà M and Poli F M 2006 Probabilistic analysis of turbulent structures from two-dimensional plasma imaging *Phys. Plasmas* **13** 100701
[12] Churchill R M, Chang C S, Ku S and Dominski J 2017 Pedestal and edge electrostatic turbulence characteristics from an XGC1 gyrokinetic simulation *Plasma Phys. Control. Fusion* **51** 105014
[13] Nespoli F *et al* 2019 3D structure and dynamics of filaments in turbulence simulations of WEST diverted plasmas *Nucl. Fusion* **59** 096006
[14] Theodorsen A, Garcia O E, Kube R, LaBombard B and Terry J L 2018 Universality of poisson-driven plasma fluctuations in the Alcator C-Mod scrape-off layer *Phys. Plasmas* **25** 122309
[15] Filippas V, Bengtson R D, Meier G-X M andi Li, Ritz C and Powers E J 1985 Conditional analysis of floating potential fluctuations at the edge of the texas experimental tokamak upgrade (TEXT-U) *Phys. Plasmas* **2** 839
[16] Boedo J A *et al* 2001 Transport by intermittent convection in the boundary of the DIII-D tokamak *Phys. Plasmas* **8** 8426
[17] Zweben S J, Boedo J A, Grulke O, Hidalg C, LaBombard B, Maqueda R J, Scarin P and Terry J L 2007 Edge turbulence measurements in toroidal fusion devices *Plasma Phys. Control Fusion* **49** S1
[18] Silva C *et al* 2009 Intermittent transport in the JET far-SOL *J. Nucl. Mater.* **390-391** 785–8
[19] Silva C *et al* 2009 *Plasma Phys. Control. Fusion* **51** 105001
[20] Müller H W *et al* 2011 Latest investigations on fluctuations, ELM filaments and turbulent transport in the SOL of ASDEX Upgrade *Nucl. Fusion* **51** 073023
[21] Militello F *et al* 2013 Experimental and numerical characterization of the turbulence in the scrape-off layer of MAST *Plasma Phys. Control. Fusion* **55** 025005
[22] Boedo J A *et al* 2014 NSTX Team. Edge transport studies in the edge and scrape-off layer of the national spherical torus experiment with langmuir probes *Phys. Plasmas* **4** 042309
[23] Grulke O, Terry J L, Cziegler I, LaBombard B and Garcia O E 2014 Experimental investigation of the parallel structure of fluctuations in the scrape-off layer of alcator cmod *Nucl. Fusion* **54** 043012
[24] Garcia O E, Horacek J and Pitts R A 2015 Intermittent fluctuations in the tcv scrape-off layer *Nucl. Fusion* **55** 2002
[25] Tanaka H *et al* 2015 Multi-pin langmuir probe measurement for identification of blob propagation characteristics in the Large Helical Device *J. Nucl. Mater.* **463** 761–4
[26] Theodorsen A, Garcia O E, Horacek J, Kube R and Pitts R A 2016 Scrape-off layer turbulence in TCV: evidence in support of stochastic modelling *Plasma Phys. Control. Fusion* **58** 044006
[27] Garcia O E, Kubea R, Theodorsena A, Bak J-G, Hong S-H and Kim H-S 2017 KSTAR Project Team and Pitts R A SOL width and intermittent fluctuations in KSTAR *Nucl. Mater. Energy* **12** 36–43
[28] Kube R *et al* 2018 *Plasma Phys. Control. Fusion* **60** 065002
[29] Garcia O E 2012 Stochastic modeling of intermittent scrape-off layer plasma fluctuations *Phys. Rev. Lett.* **108** 265001
[30] Garcia O E, Kube R, Theodorsen A and Pécseli H L 2016 Stochastic modelling of intermittent fluctuations in the scrape-off layer: Correlations, distributions, levelcrossings and moment estimation *Phys. Plasmas* **23** 052308
[31] Militello F and Omotani J T 2016 Scrape off layer profiles interpreted with filament dynamics *Nucl. Fusion* **56** 104004
[32] Chapman I T *et al* 2015 Overview of MAST results *Nucl. Fusion* **55** 104008
[33] MacLatchy C S, Boucher C, Poirier D A, and Gunn J 1992 *Rev. Sci. Instrum.* **63** 3923
[34] Hadamard J 1902 Sur les problèmes aux dérivées partielles et leur signification physique *Princeton Univ. Bull.* **13** 49–52
[35] Stuart A M 2010 Inverse problems: A bayesian perspective *Acta Numerica* **19** 451–559
[36] Cotter S L, Dashti M and Stuart A M 2010 Approximation of bayesian inverse problems for pdes *SIAM J. Numer. Anal* **48** 322–45
[37] Iglesias M A, Lin K and Stuart A M 2014 Well-posed bayesian geometric inverse problems arising in subsurface flow *Inverse Prob.* **30** 114001
[38] MacKay D J C 2003 *Information Theory, Inference and Learning Algorithms* (Boston, PA: Cambridge University Press)
[39] Mukherjee P, Parkinson D and Liddle A R 2006 A nested sampling algorithm for cosmological model selection *Astrophys.* J. **638** L51—-L54
[40] Svensson J and Werner A and 2008 JET-EFDA contributors. Current tomography for axisymmetric plasmas *Plas. Phys. Cont. Fus.* **50** 085002
[41] Militello F *et al* 2016 Multi-code analysis of scrape-off layer filament dynamics in MAST *Plasma Phys. Control. Fusion* **58** 105002
[42] Easy L, Militello F, Omotani J, Dudson B, Havlíčková E, Tamain P, Naulin V and Nielsen A H 2014 Three dimensional simulations of plasma filaments in the scrape off layer: A comparison with models of reduced dimensionality *Phys. Plasmas* **21** 122515
[43] Militello F, Garzotti L, Harrison J, Omotani J T, Scannell R, Allan S, Kirk A, Lupelli I, Thornton A J and 2016 The MAST team. Characterisation of the L-mode scrape off layer in MAST: decay lengths *Nucl. Fusion* **56** 106006

[44] Militello F, Farley T, Mukhi K, Walkden N and Omotani J T 2018 A two-dimensional statistical framework connecting thermodynamic profiles with filaments in the scrape off layer and application to experiments *Phys. Plasmas* **25** 056112
[45] Garcia O E, Fritzner S M, Kube R, Cziegler I, LaBombard B and Terry J L 2013 Intermittent fluctuations in the alcator C-Mod scrape-off layer *Phys. Plasmas* **20** 055901
[46] Johnson J L 2003 *Probability and Statistics for Computer Science* (New York: Wiley-Interscience)
[47] Mainard F, Gorenflo R and Vivo A 2007 Beyond the poisson renewal process: A tutorial survey *J. Computat. Appl. Math.* **205** 725–35
[48] Thulasiraman K and Swamy M N S 1992 *Graphs: Theory and Algorithms* (New York: Wiley)
[49] Drew J H, Evans D L, Glen A G and Leemis L 2008 Computational probability: algorithms and applications in the mathematical sciences Chapter 4 *Int. Series in Operations Research and Management Science* Springer vol 117
[50] Gottfried B S and Weisman J 1973 *Introduction to Optimization Theory* (Englewood Cliffs, NJ: Prentice Hall)
[51] Brooks S, Gelman A, Jones G L and Meng X-L eds 2011 *Handbook of Markov Chain Monte Carlo* (London: Chapman and Hall)
[52] Roberts G O and Rosenthal J S 2001 Optimal scaling for various Metropolis-Hastings algorithms *Stat. Sci.* **16** 351–67