L C Appel, S Kwak, F Militello, J Svensson

A Bayesian model of filamentary dynamics in MAST

A Bayesian model of filamentary dynamics in MAST

L C Appel, S Kwak, F Militello, J Svensson

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`

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.

1. Introduction

Nuclear fusion offers the promise of an unlimited power source. Six decades of research activities in thermonuclear magnetic confinement fusion have achieved nuclear fusion close to break-even conditions characterised by Q=0.8 (the ratio between input and fusion power) and a good understanding of the conditions required to achieve steady state ignition. In recent years, attention has begun to shift towards the engineering challenges of building a fusion reactor that is reliable and robust. Four fifths of the power generated in a fusion reactor is in the form of 14MeV neutrons. These neutrons deposit their energy approximately isotropically in a material volume extending beyond the first wall facing the hot plasma. However, the remaining power is used to heat the plasma, and this energy eventually deposits on the first wall in the form of isotropic radiation, inelastic collisions with neutral particles, and a localised plasma flux. The highest heat and particle flux is in the divertor region and there the maximum incident power can exceed engineering designs limits unless appropriate measures are taken, for example plasma detachment. The challenge therefore is to provide firm limits on the prediction of maximum heat flux density flowing from the plasma, and secondly to use the understanding of plasma dynamics to configure improved configurations for managing the heat flow.

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. 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 [1]. The structures were generally observed to be be aligned with field structures and became known as filaments [2]. Understanding the filament dynamics is complicated by many factors. In this region, the magnetic structure transitions from a geometry of closed field lines across a separatrix to an open field line geometry; a further feature is the field nulll known as the X-point that separates the divertor from the confinement region. A significant component of the heat and particle flux is driven by three-dimensional structures, there are strong gradients of plasma density, temperature and electric fields. In the colder plasma regions the role of the neutral particles is critical.

In early work, the general approach was to decide *a priori* the relevant measure that should be applied to the plasma fluctuation measurements[3, 4, 5]. The common analytical tools used included cross- and auto- correlation functions, moments of the distributions of the signal amplitudes, and power spectra. From the late 1990s evidence began to emerge that the characteristics of the edge fluctuations were dependent on multiple scale lengths. Multi-fractal analyses indicated that edge plasma turbulence is self-similar at short time and space scales whereas at larger scale lengths the structure of

the plasma edge turbulence is mono-fractal[5]. The conditional averaging technique[6] regarding the edge turbulance fluctuations as a superposition of coherent structures with additive zero-mean random and time independent components is now widely used[7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. It provides compelling evidence to support the view that the ion saturation current signal is dominated by large amplitude bursts[16] and has been used as an intermediate metric to compare with theoretical models [12]. Nevertheless, the conditional averaging technique has disadvantages. Its ability to identify self-similar fluctuations is at the expense of ignoring intrinsic variabilities between filaments. Experimental evidence[17] supports theoretical developments[18] regarding the coherent turbulence as self-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[19]. There is evidence of variation in the wave shape of filaments; usually the leading edge rises faster than the trailing edge[7, 17], but there are examples where the waveforms are symmetric[12]. Another issue, due to overlap of filaments, can reduce the apparent asymmetry between leading and trailing edges[17]. Furthermore, the identification of candidate fluctuations relies on selection criteria that are to some extent arbitrary. A common approach (see [8] for a typical approach amongst many) employs an amplitude threshold specified relative to the background level whereas [14] includes an additional exclusion zone to prevent filament overlap.

This paper, develops a data-driven model describing filamentary dynamics using the principles of Bayesian Inference. Similar to [17, 19], the edge turbulence data is regarded as a superposition of filaments, with an asymmetric leading and trailing edge. However, in this case, there is no restriction enforcing uniform waveforms for the filaments, and the minimum amplitude threshold providing a condition for existence of filaments is relaxed. 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 plasma core.

2. Experimental description

Results are presented from the Mega Ampere Spherical Tokamak (MAST)[20] designed to study low aspect ratio ($R/a \sim 0.85\text{m}/0.65\text{m}=1.3$), highly elongated ($\kappa$=2) plasma configurations with plasma current, I_p ≤1.5MA and toroidal field at the tokamak geometric axis, B_t ≤0.52T).

The analysis uses experimental data from the connected double null discharge 21712, during an ohmic L-mode phase (the same dataset is used by [12]). The core plasma conditions has a plasma current of Ip = 400 kA, a confining magnetic field of $B_T = 0.4$ T at R=0.66m (the resulting safety factor is $q_{95} \approx 6.2$). During the flat-top phase, the core electron temperature is $T_e \approx 650\text{eV}$ and the line-averaged density $n \approx 1.7 \times 10^{19}\text{m}^3$.
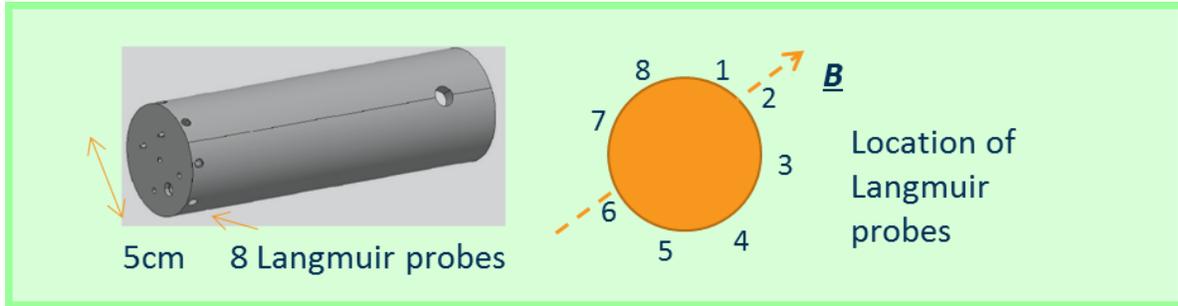
Figure 1. The left figure shows the outer shield of the MAST Gunderstrup probe indicating locations of the pins for 8 Langmuir probes arranged in a ring. The right figure indicates the labeling of the Langmuir probes and relative to the magnetic field vector.

The measurement data analysed were obtained with a Gundestrup probe[21] 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 10cm across the scape-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 -200V in order to measure the ion saturation current, $I_{sat} \sim n\sqrt{T_e}$. Figure 3 plots the time dependence of the signal from Langmuir probe 5. The radial position of the probe relative to the separatix 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 traveling through the SOL and into the plasma core.

s

3. Bayesian Model

In order to study the characteristics of the fluctuations at different radial positions, the time-series plot shown in figure 3 is divided into a set of subsets each with n_d data points. Figure 4 show details of the signal when the probe is (i) in the far SOL and (ii) in the plasma core. By inspection, the signal appear to consist of a fluctuating and a constant component. Both the constant and fluctuating components are larger in the plasma core than in the far SOL. Estimating nominal values for the fluctuation amplitude in the plasma core and SOL of (0.1, 0.4) and background of (0.05, 0.4), the ratio *fluctuation amplitude:background* increases from a value of 1 to 2 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. 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
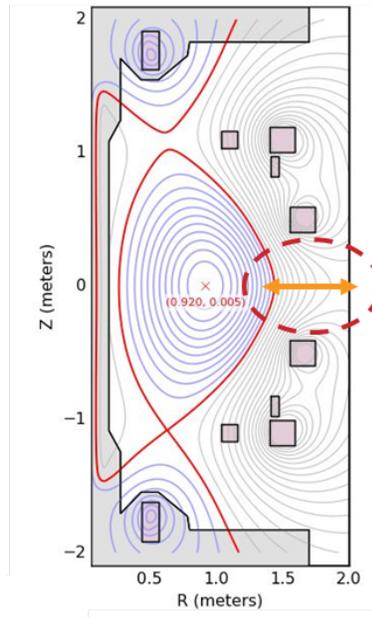
Figure 2. Contours of equilibrium poloidal flux iwith the location of the reciprocating Gunderstrup probe indicated.
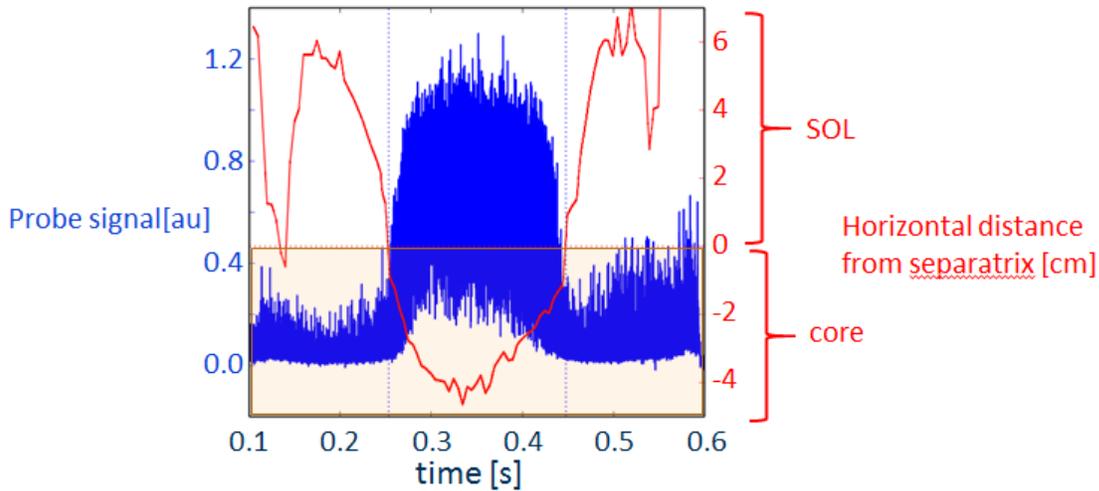
Figure 3. Time series of the ion saturation current fluctuations of Langmuir probe 5 during a single reciprocation of the Gundestrup probe assembly.

amplitude and, the amplitude of the ith 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:

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

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 values of $\{t_0^{(i)}\}$ for the N filaments are the
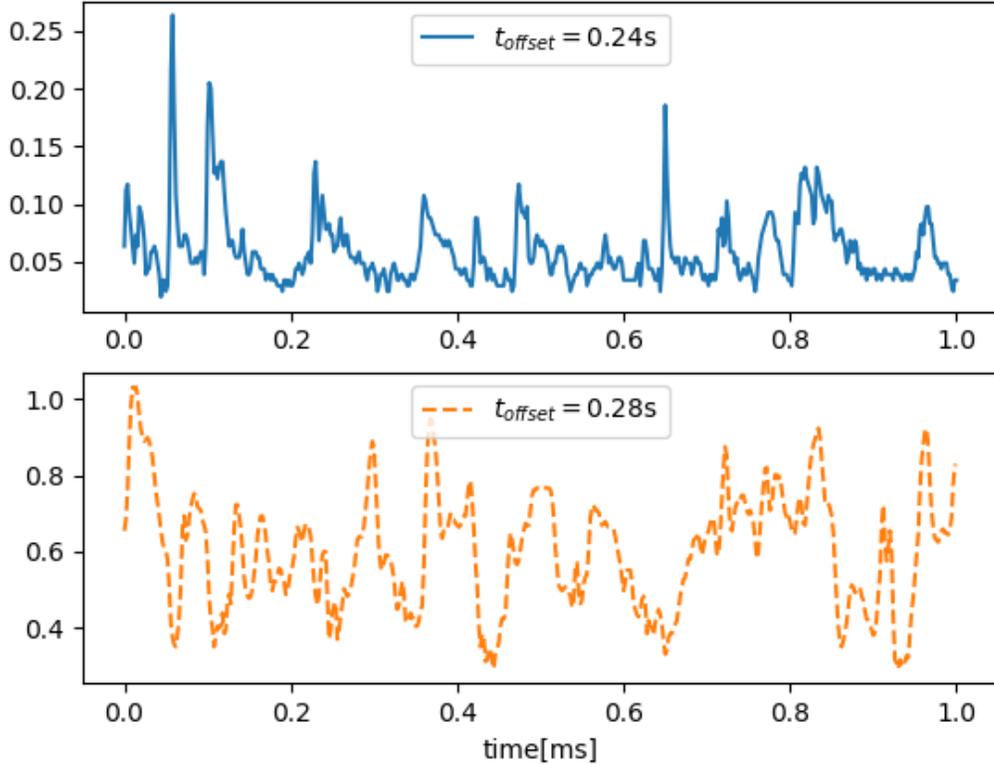
Figure 4. Experimental data from Langmuir probe 5: (a) top probe is in outer SOL; (b) bottom probe is in core plasma.

the set of times corresponding to consecutive data points $\{d_{i-1}, d_i, d_{i+1}\}$ in which d_i is a local maximum. This choice effectively sets a minimum threshold at the data digitisation level. The model has 3N+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[22] of well-posedness (These are the following conditions: (i) a solution exists; (ii) the solution is unique and (iii) the solution's behavior 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[23]. 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 [24, 23, 25]. 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:

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

where D and H are multivariate distributions for the data and the model parameters 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*). This result 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 [26, 27]) yields the proportionality condition:

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

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

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

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[23]. This is straightforward to demonstrate for a uniform prior and a normally distributed likelihood with a diagonal covariance matrix. In this case

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

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 $3N + 2$.

The Bayesian inference problem has been implemented in the Minerva Bayesian analysis framework[28]. Details are described in the appendix. The model calculations carried out in this paper compute inferences using measurement data partitioned into 0.4ms 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 $0.242 < t[s] < 0.243$ (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 $\sigma = 6 \times 10^{-3}$. Over all data subsets, the MAP background signal is observed to be zero for 96% of cases (this includes data subsets where the probe has crossed into the core plasma region), whereas the MAP value of σ is ~10% of the quiescent signal level. Figure 6 plots the probability distribution $P(B|D)$. The figure confirms that whereas the MAP-value of the background is zero (ie. the value of B at the peak of the distribution), 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 samples maintaining a value of $\sigma \sim 10\%$ of the average signal level.

To explore more fully the characteristics of the posterior, MCMC samplng was undertaken. Figure 8 plots samples from the distribution $P(\log_{10} \tau_1^{(i)}|D)$ for each of the 32 filaments plotted against their respective MAP amplitude. The color 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 color 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; filaments for which the range of sampled $\log \tau_1$ values is less than 3 appear in the upper panel whilst those with a larger range are in the lower panel. The figure reveals that for larger amplitude filaments $-6 \lesssim \log \tau_1 \lesssim -4$, whilst smaller amplitude filaments have a considerably larger range of $\log \tau_1$, in many cases extending to the limits of the prior distribution. Figure 9 is a similar plot for $\log \tau_2$. (The placement of the filaments whether in upper or lower panel is still determined by the variation of $\log \tau_1$.) Similar to before, the variation of $\log \tau_2$ 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[12]. Figure 10 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 $\log \tau_1$.) The results indicate that filaments with small MAP amplitudes mainly have sampled amplitude in the range $0 \leq A \lesssim 2A_{MAP}$, ie the fractional uncertainties are $\sim 100\%$ and include the possibilitiy of non-existence. In contrast, larger amplitude filaments mainly have smaller fractional

uncertainties apart from filments 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[29]. Furthermore, the filaments are presumed to pass through and interact with a quiescent plasma background[30]. However, the interpretation of experimental data often involves applying a smoothing function, for example using the method of time-window averaging [31], 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 8 and 9 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 the core, whereas other "filaments" form part of a fluctuating background. In particular, filaments originating from the core are defined as those in which the range of $\log \tau_1$ of samples from the posterior distribution satisfies $\delta \log \tau_1 < 3$. Figure 11 shows the computed signal components of the filaments originating from the core (hereafter these will be the only ones to be referred to as *filaments*) and the background fluctuations for $0.242 < t[s] < 0.243$ (the probe is in the SOL). The top plot shows the structures that are identified as filaments, and those that are part of the background fluctuations. The second plot from the top shows the MAP computed signal components for the filaments and for the background. The bottom two plots are the signals computed using parameter values sampled from the posterior distribution. The signals are sorted in decreasing value of the posterior probability of each sample; signals with highest probability are plotted above those with lower probability. From the plots it is evident that the maximal uncertainty in the computed filament and background fluctuations is already explored using only 20% of the data set closest to the MAP (ie using samples with the highest 20% vlaues of $P(H|D)$).

The same analysis has been carried out for data subsets throughout the SOL and into the plasma core. Figure 12 shows the an example at $0.28 < t[s] < 0.2804$ when the probe is in the core plasma region. Details of the parameters are the same as figure 11. Compared to the earlier time, both filament and background components are larger, and the filaments dominate the background throughout the time-range. The uncertainty in the computed filament and background signals are similar in both cases.

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

 (i) The MAP background is usually zero, and the expectation value from the background has a small finite value.

(ii) The model uncertainty σ is approximately 10% of the average signal.
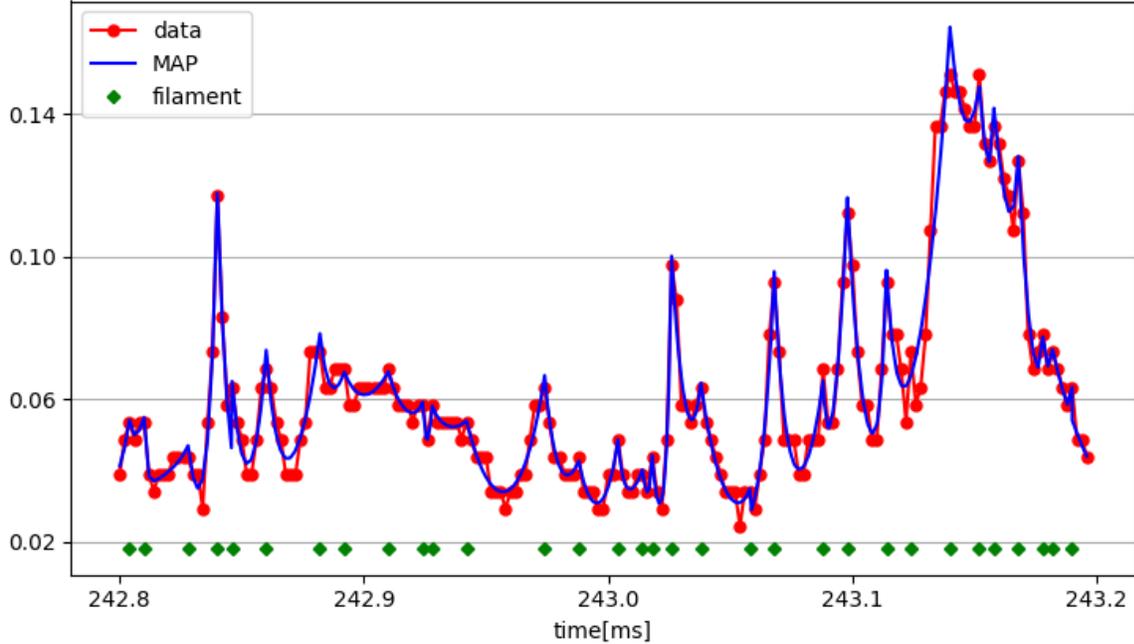
Figure 5. Experimental data and computed values using MAP parameter values in the SOL. The green diamonds are the locations of the filaments identified according to the criterion defined in section 3.

(iii) The values of $\log \tau_1^{(i)} \approx \log \tau_2^{(i)}$ for each filament in a particular subset region (ie. the filaments are symmetrical).

(iv) We postulate that *true* filaments are well defined structures that originate from the core whereas the remaining "filaments" form part of a fluctuating background. The *true* filaments have larger MAP amplitude values compared to other filaments.

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 13 plots the variation of the computed filament and background fluctuatios with time as the probe moves from the SOL into the plasma core. The signals are computed using the MAP parameters and have been averaged in 2ms time-intervals. The plots show the average value and range of values in each time-interval. It is evident that both the computed filament signal and computed background signal exhibits large fluctuations, however, the bin-to-bin variation of the averaged signal is noticeably smaller. Both the background and filament signals increase as the probe approaches the plasma core. The filament signal increases strongly around t=0.26s, whereas the increase of the background is more gradual. One interpretation is that filaments are being generated at the location where the profile steepens. From figure 3
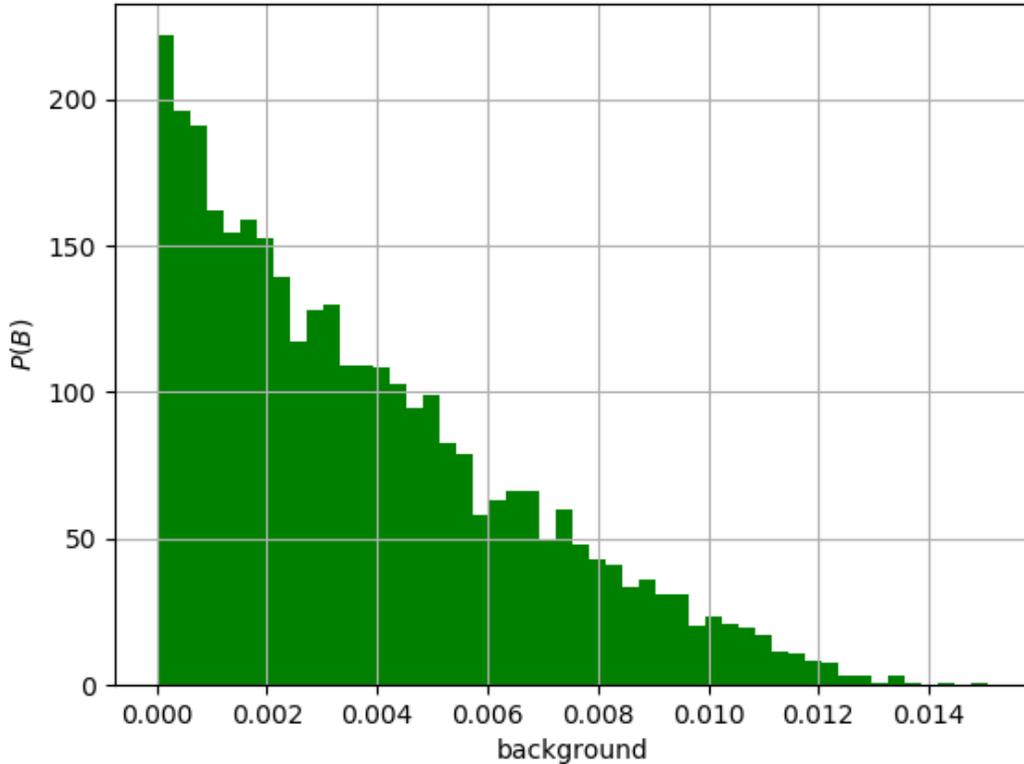
Figure 6. The distribution $P(B|D)$ assembled with 10^4 MCMC samples for the data subset $0.242 < t[s] < 0.243$ (probe is in SOL region).

this is clearly in the proximity of the magnetic separatrix, however, uncertainty in the computed magnetic configuration means that it is not possible to identify whether the source is in the plasma core or SOL. Further evidence for the existence of a filament source is provided by calculations of the probability distribution functions (pdf) of the computed filament signal at various time-intervals shown in figure 14. These display a remarkable degree of self-similarity during the times $0.195 < t[s] < 0.263$ and then a sudden change for $t[s] > 0.263$. In comparison, the pdf functions of the computed background fluctuations for the same time-intervals shown in figure 15 exhibit a more gradual evolution of the pdf structure. Figure 16 plots the pdf of filament waiting times at various times intervals. The dashed line indicates an exponential distribution that would be indicative of a Poisson process. The results show that the distribution of filament waiting times in the SOL (but not in the plasma core) are Poisson distributed. A similar series of plots for the background fluctuations is shown in figure 17; in this case the waiting times are Poisson distributed in the SOL and core plasma.

Figure 18 plots the variation in waiting times for the computed filaments and computed 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
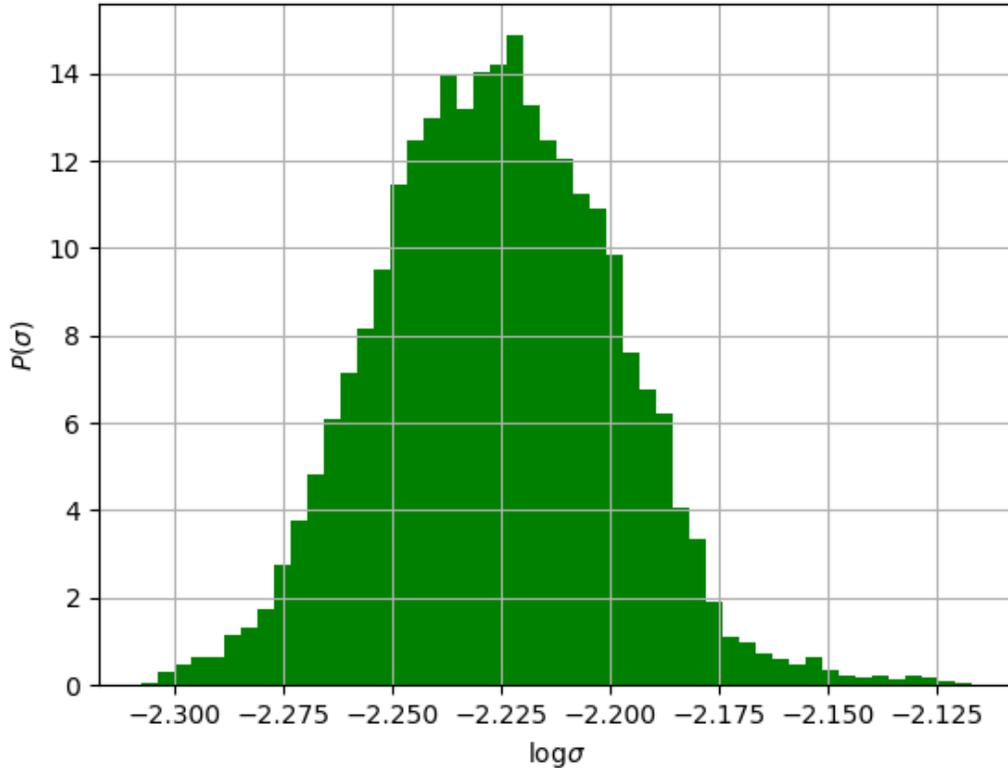
Figure 7. The distribution $P(\sigma|data)$ assembled with 10^4 MCMC samples for the data subset $0.242 < t[s] < 0.243$ (probe is in the SOL region).

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; (ii) the maximum background fluctuations exhibit a reduction in the proximity of the separatrix (around t=0.25s); (iii) maximum waiting times of background fluctuations are less in the SOL than in the plasma core. (It should be noted as shown in figure 17 that the background fluctuations are Poisson-distributed in both regions.)

5.1. discussion

Section 5 presented results for the radial variaton of the computed background fluctuations and the computed plasma filaments. It was observed in figure 16 that the pdf of waiting times of filaments exhbited a signfiicant 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 generating new filaments. In contrast, observed variations in the pdfs of the signals across the plasma shown in figure 14 can be explained both by the presence of sinks (for
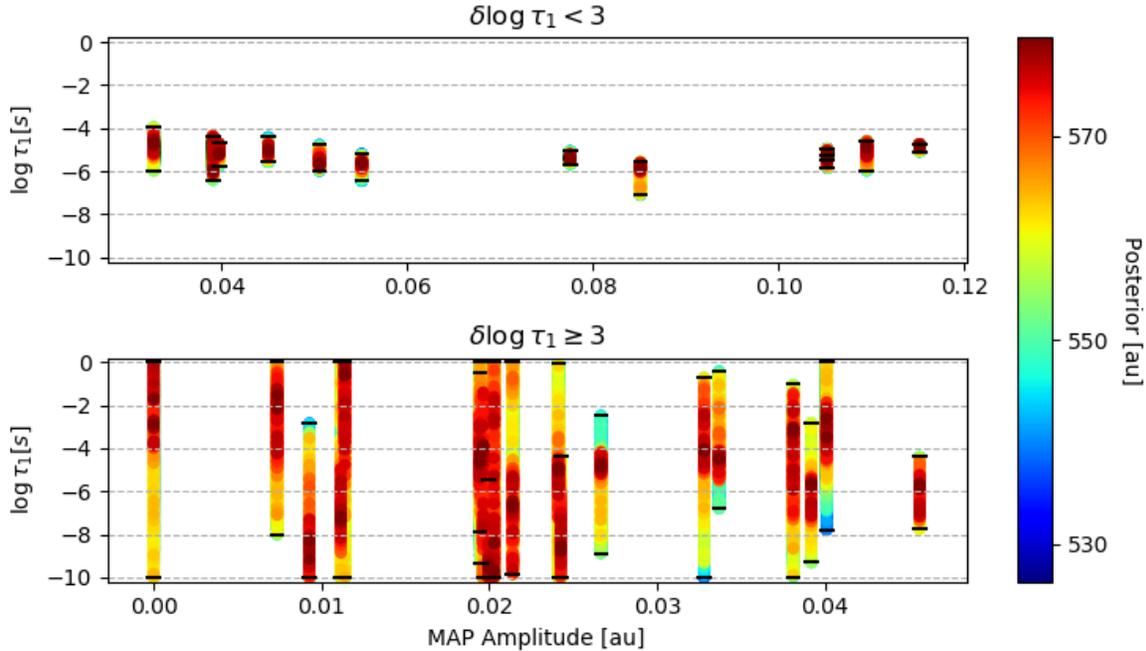
Figure 8. Samples from the distribution $P(\log_{10} \tau_1^{(i)}|D)$ for each filament in the data subset $0.242 < t[s] < 0.243$ plotted against the respective MAP amplitude. There are 32 separate filaments and the sample values for each filament are delimited by its respective MAP value. The upper panel shows filaments in which the range of sample values for $\log \tau_1 < 3$, whereas the lower panel is where the range $\log \tau_1 \geq 3$. Coincidentally the MAP amplitudes of filaments in the upper panel are all greater than those in the lower panel.

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 17 to be Poisson-distributed throughout the plasma core and the SOL. These results are consistent with a source generating random uncorrelated fluctuations in the core plasma and the SOL. 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 across field lines. However, the further observation shown in figure 18 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 plasma filaments shown in figure16 is Poisson-like in the SOL and exhibits more symmetric Gaussian characteristics in the plasma core. The results are more uncertain than those for the background, explained by the smaller number of observed events. Nevertheless the results are consistent with previous
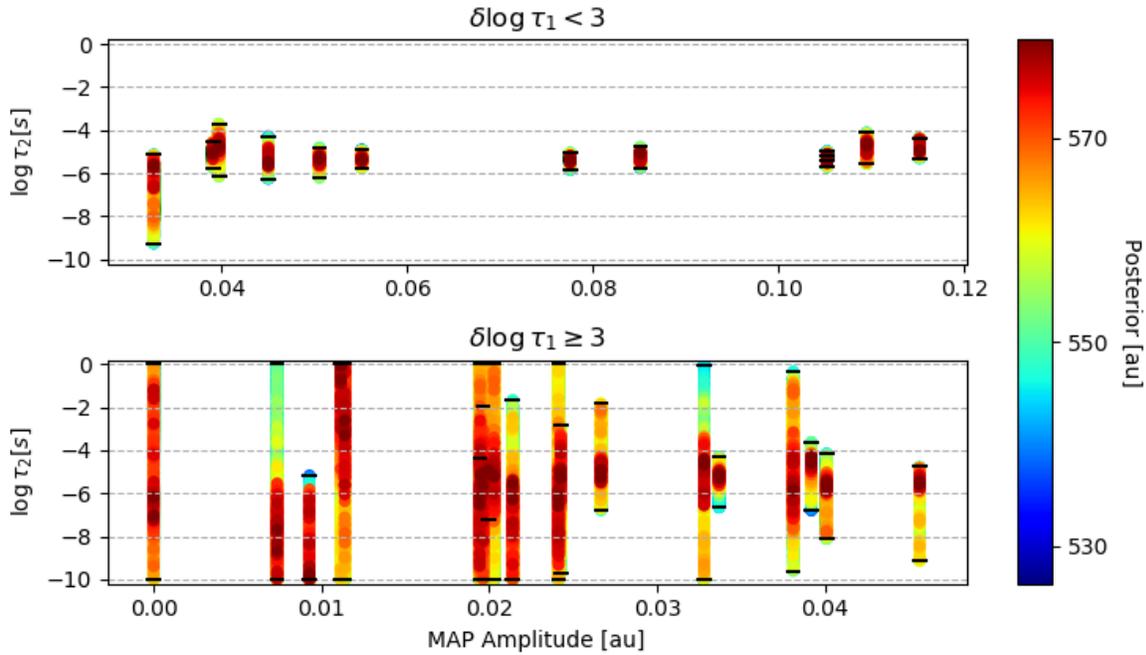
Figure 9. Samples from the distribution $P(\log_{10} \tau_2^{(i)}|D)$ for each filament in the data subset $0.242 < t[s] < 0.243$ plotted against the respective MAP amplitude. For further explanation about the interpretation of this figure, refer to the caption in figure 8
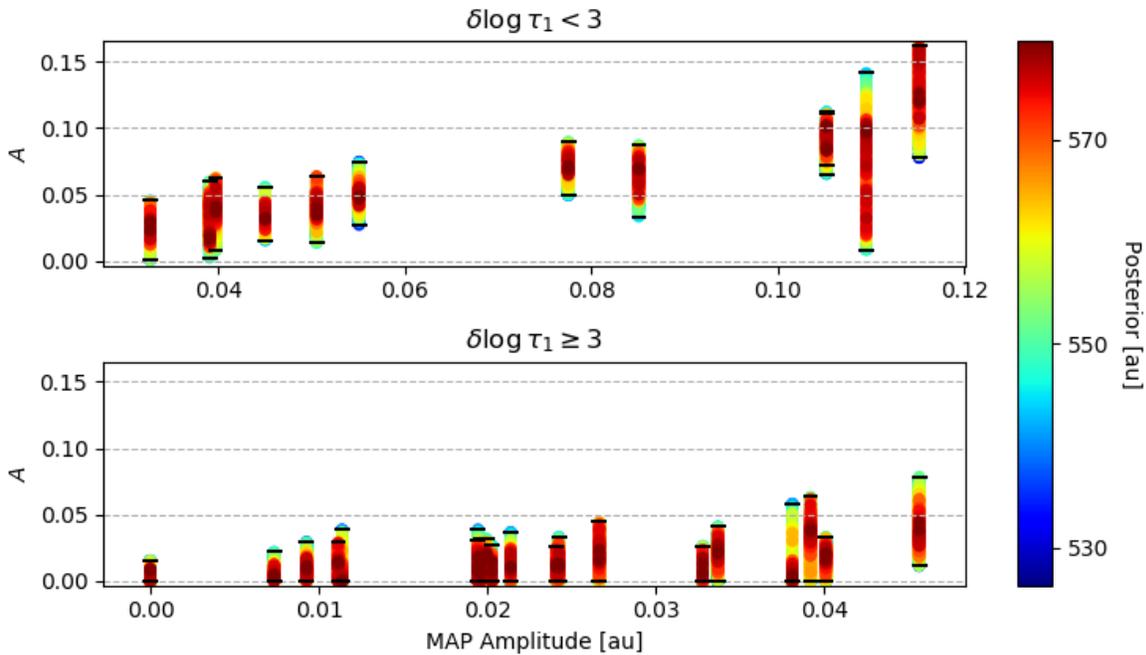
Figure 10. Samples from the distribution $P(A^{(i)}|D)$ for each filament in the data subset $0.242 < t[s] < 0.243$ plotted against the respective MAP amplitude. For further explanation about the interpretation of this figure, refer to the caption in figure 8
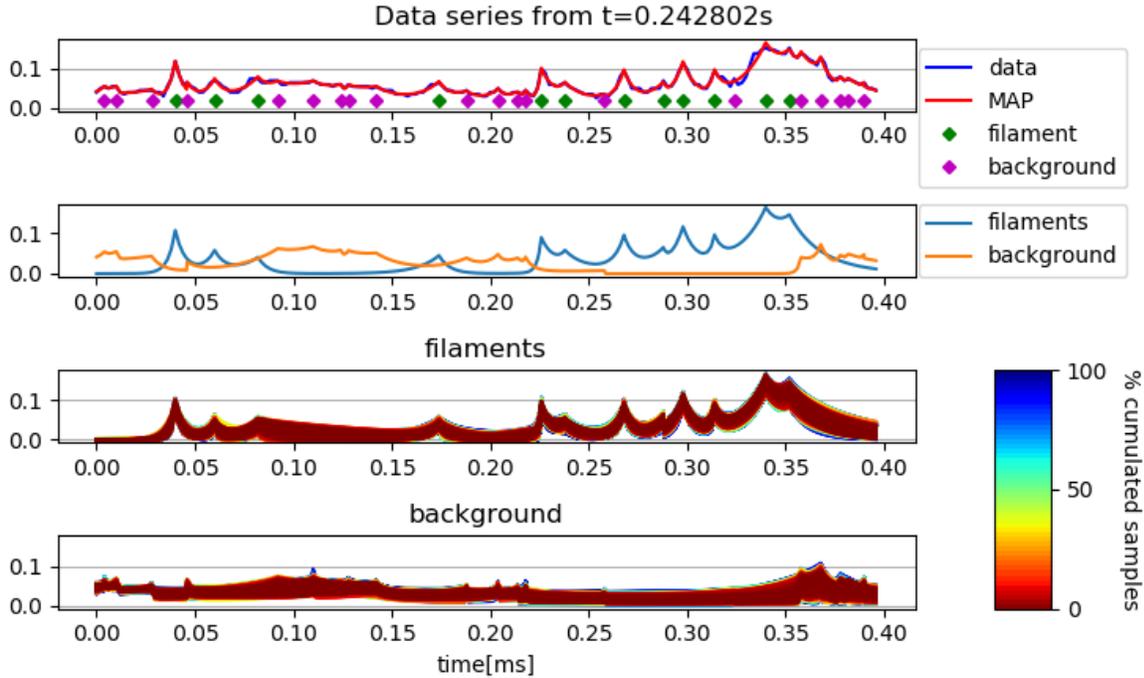
Figure 11. Interpration in terms of plasma filaments and a fluctuating background component for the data subset $0.242 < t[s] < 0.243$. Figures from top to bottom: (i) the plasma filaments are indicated in green and the background filaments in purple. The measured data and computed signal using MAP parameters 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 using MCMC sample values provide an indication of the uncertainties.

observations [12], and can be interpreted that there is a source of plasma 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 18 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 pdf of the background fluctuation signal shown in figure 15 progressively deviates from a Poisson-like distribution in the plasma core going outwards towards the far SOL. (If the distributions had been Poisson at each position, this would have suggested that there was a sink that was proportional to the signal amplitude.) It is evident that the signals at the mean value (ie around value zero on the x-axis) are overly damped in the far SOL compared to large-amplitude signals. A possible explanation for this is larger amplitude background filaments are incorrectly categorised as filaments
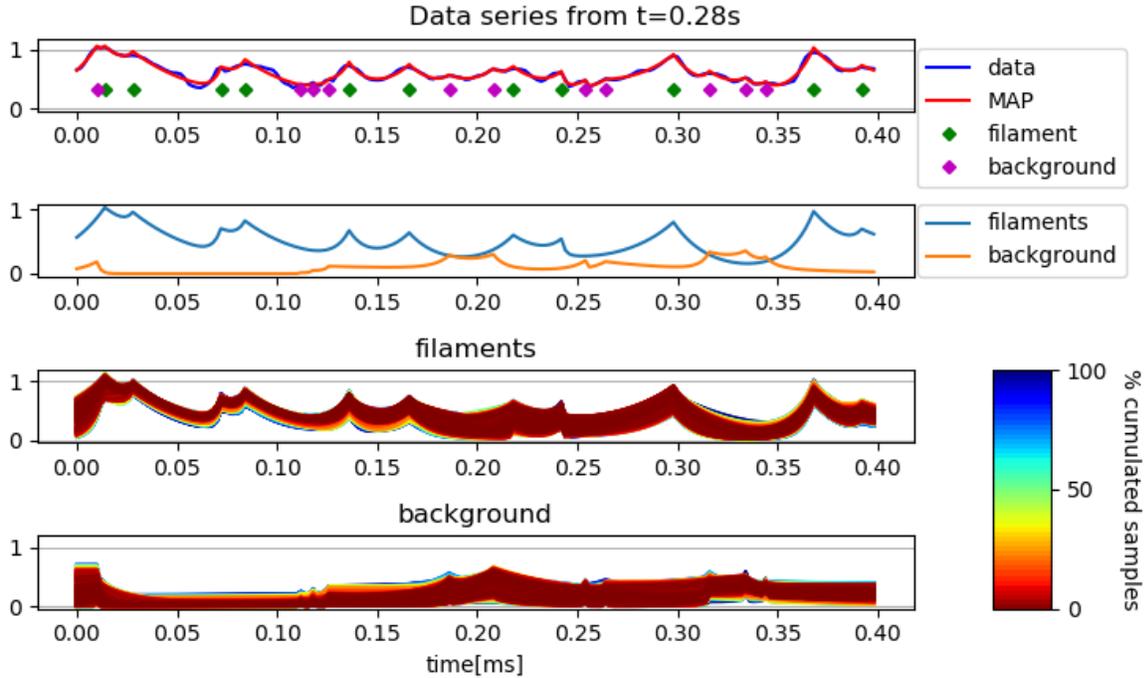
Figure 12. Interpration in terms of plasma filaments and a fluctuating background component for the data subset $0.28 < t[s] < 0.2804$. Figures from top to bottom: (i) the plasma filaments are indicated in green and the background filaments in purple. The measured data and computed signal using MAP parameters 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 using MCMC sample values provide an indication of the uncertainties.

originating from the plasma core. The further observation shown in figure 13 that the mean background fluctuations decay in the SOL can be explained by parallel transport providing a sink.

The pdfs of the plasma filaments signals shown in figure 14 are similar in form to the pdfs of the waiting times throughout the plasma core and SOL. 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 Poisson-distributed. The further observation shown in figure 13 that the the mean filament amplitude decays in the SOL can be explained by parallel transporting, although it is interesting that the rate of decay is less for the filaments than for the background fluctuations.

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
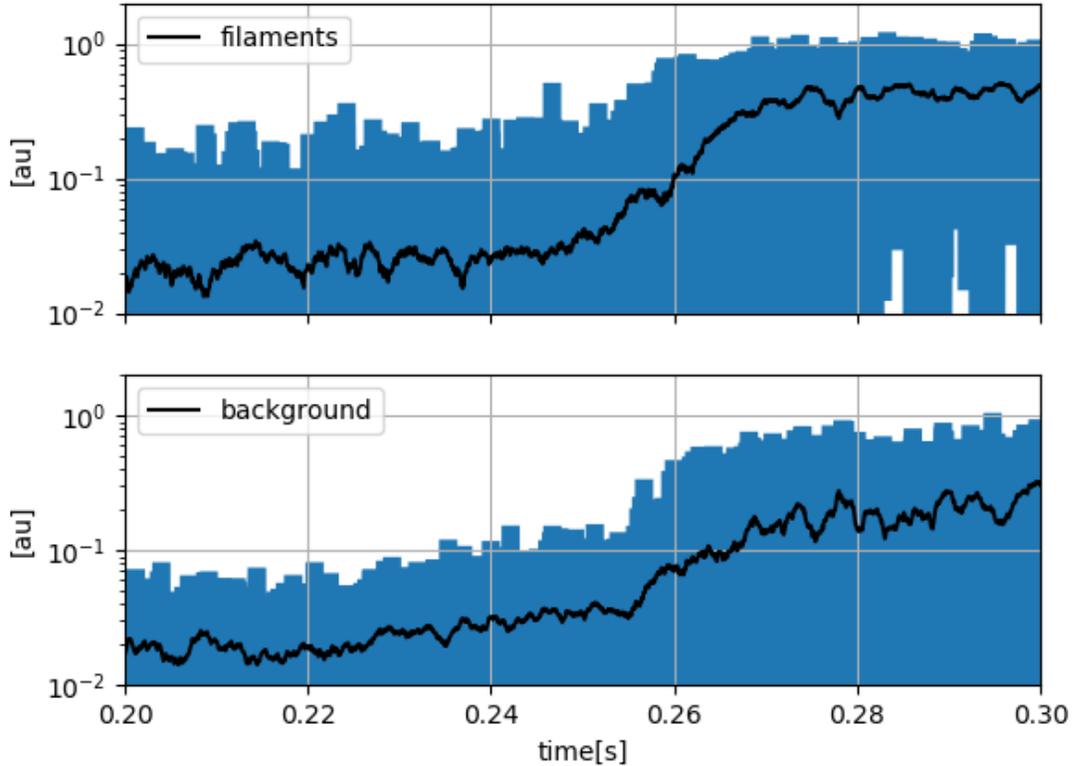
Figure 13. Time-variation of computed filament and fluctuation amplitudes using MAP parameters. The black lines are the average values, and the shaded regions indicate the range of amplitudes in time-moving subsamples of 2ms duration.

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 sub-sampled 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 $\sim 10\%$. 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 chacteristised 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
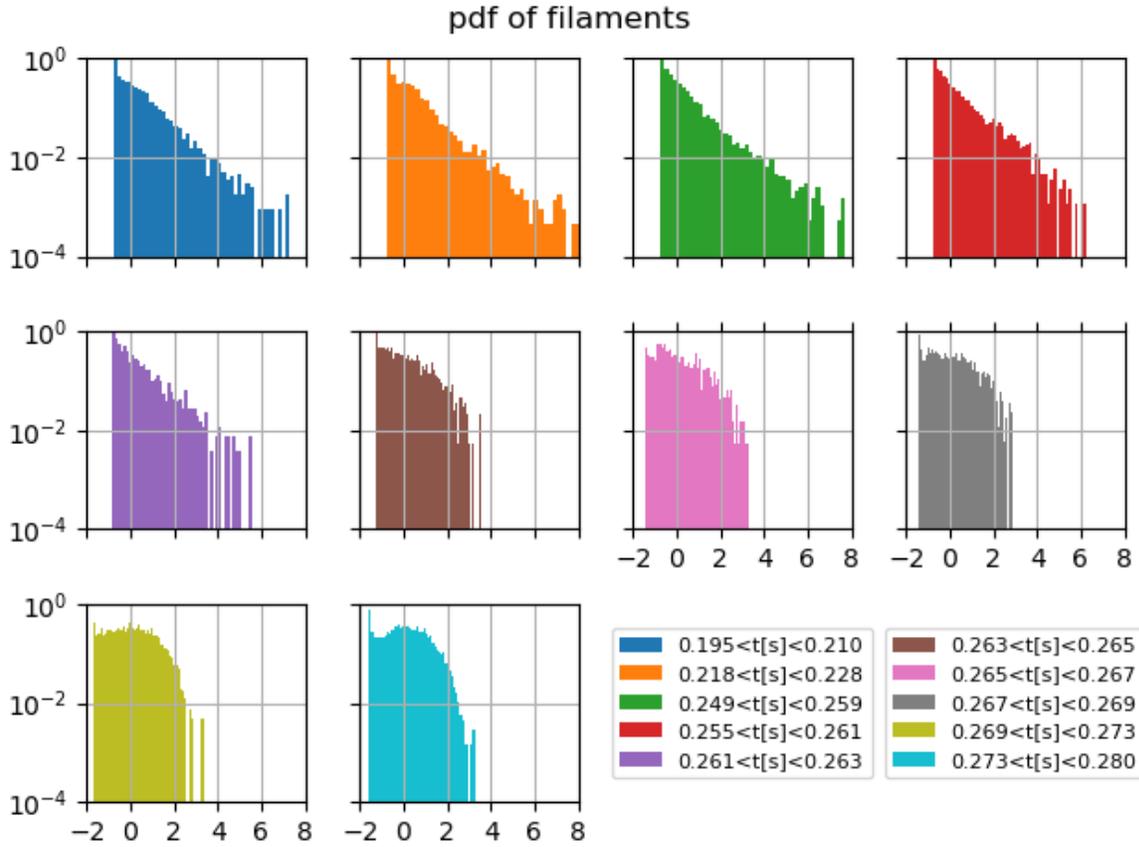
Figure 14. The pdf of the computed filament signal (f) defined as $(f - <f>)/\sigma_f$ for ten time-ranges from the SOL through to the plasma core.

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 plasma filaments are Poisson distributed. The pdf of waiting times for the background fluctuations are Poisson-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 plasma filaments is constant throughout the plasma core and SOL regions even though details of the pdf change from a Poisson-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
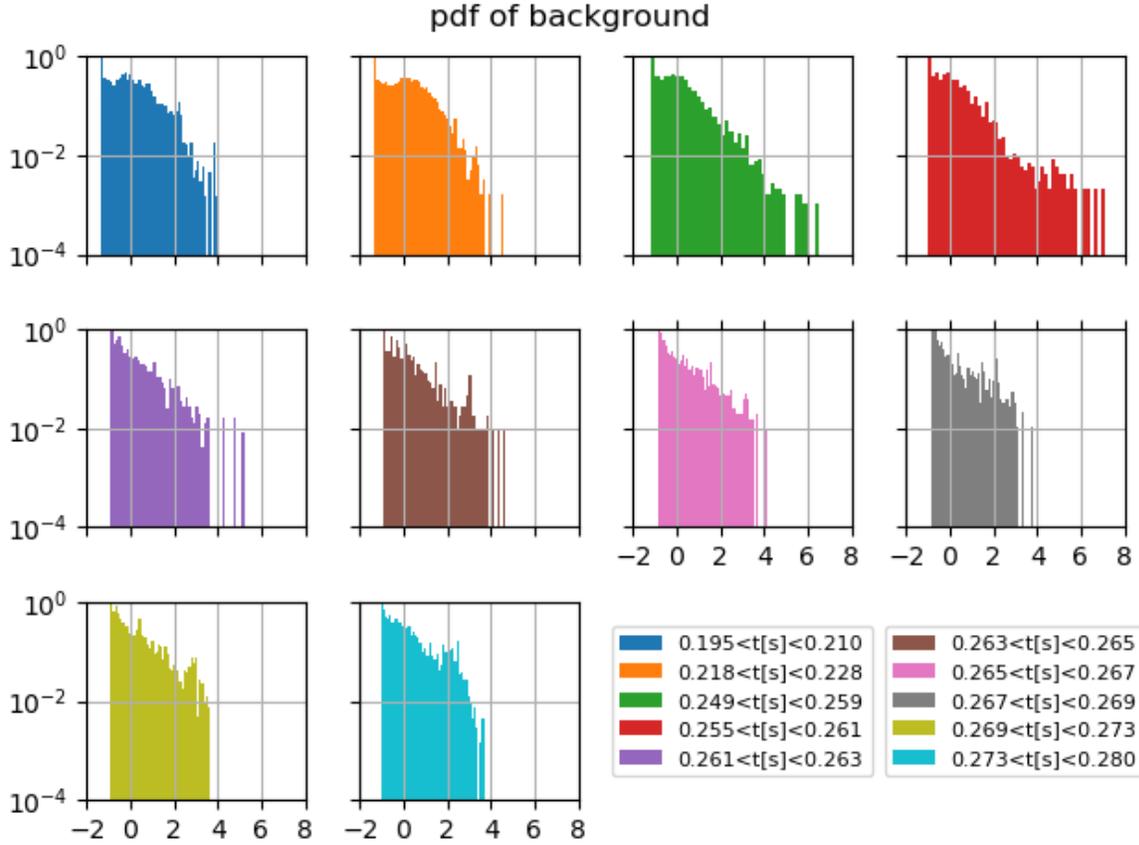
Figure 15. The pdf of the computed background signal (b) defined as $(b- <b>)/\sigma_b$ for ten time-ranges from the SOL through to the plasma core.

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 $\sim 10\%$. 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 and plasma filament are observed to decrease across the SOL region. Future work should establish a definition for these signals on the basis of theoretically-based constraints.

7. Acknowledgements

pdf of waiting times: filaments

■ 0.195<t[s]<0.210, 261 filaments, mean=57μs	■ 0.261<t[s]<0.265, 82 filaments, mean=48μs
■ 0.218<t[s]<0.228, 171 filaments, mean=59μs	■ 0.265<t[s]<0.269, 92 filaments, mean=44μs
■ 0.249<t[s]<0.259, 232 filaments, mean=43μs	■ 0.269<t[s]<0.273, 85 filaments, mean=47μs
■ 0.255<t[s]<0.261, 116 filaments, mean=52μs	■ 0.273<t[s]<0.280, 153 filaments, mean=46μs
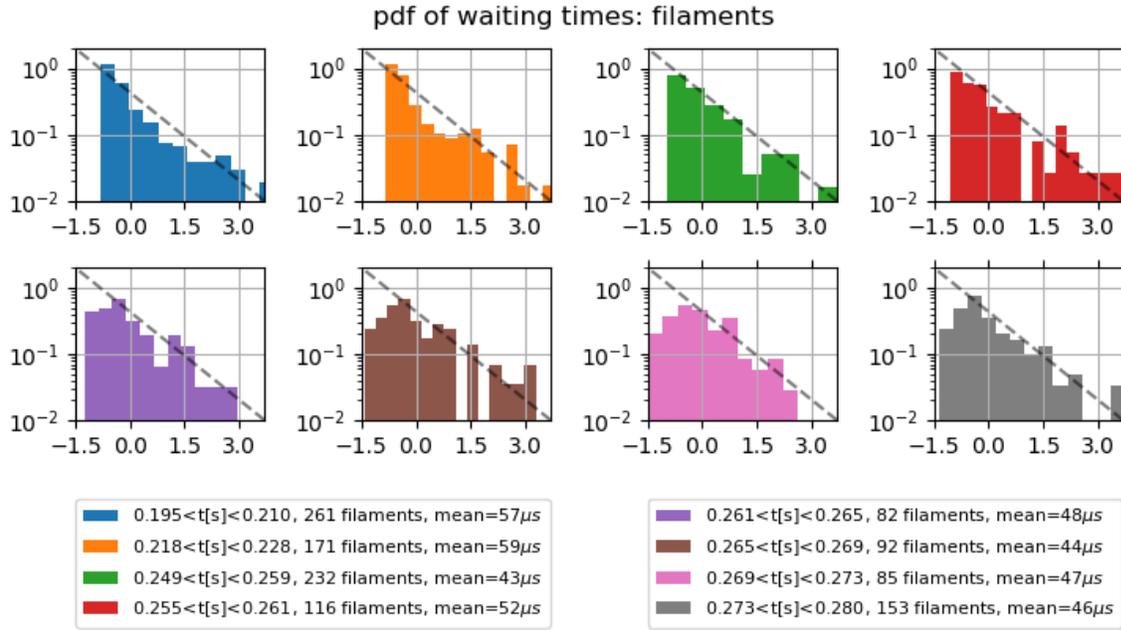
Figure 16. The pdf of the computed filament waiting times (w) defined as $(w- <w>)/\sigma_w$ for eight time-ranges from the SOL through to the plasma core. The dashed lines shows the characteristic of a Poisson distribution.

pdf of waiting times: background

■ 0.195<t[s]<0.210, 847 filaments, mean=18μs	■ 0.261<t[s]<0.265, 170 filaments, mean=24μs
■ 0.218<t[s]<0.228, 552 filaments, mean=18μs	■ 0.265<t[s]<0.269, 115 filaments, mean=34μs
■ 0.249<t[s]<0.259, 656 filaments, mean=15μs	■ 0.269<t[s]<0.273, 113 filaments, mean=36μs
■ 0.255<t[s]<0.261, 346 filaments, mean=17μs	■ 0.273<t[s]<0.280, 201 filaments, mean=35μs
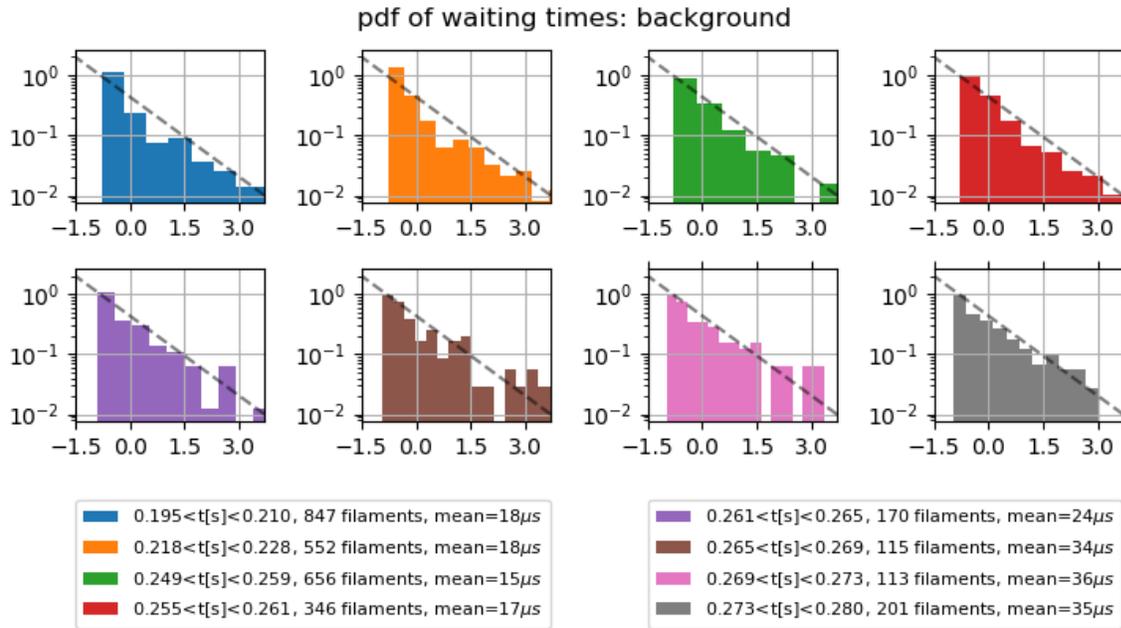
Figure 17. The pdf of the computed background fluctuation waiting times (w) defined as $(w- <w>)/\sigma_w$ for eight time-ranges from the SOL through to the plasma core. The dashed lines shows the characteristic of a Poisson distribution.
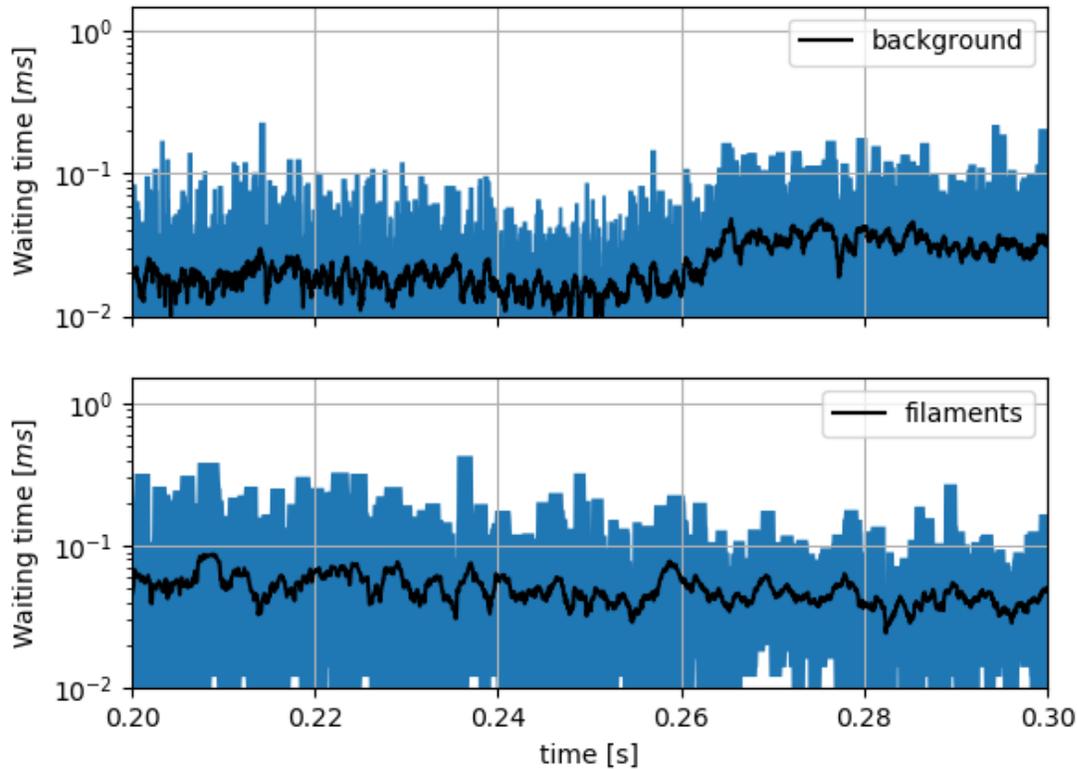
Figure 18. Time-variation of computed filament and background fluctuation waiting times. The black line are the average values, and the shaded regions indicate the range of waiting times in time-moving subsamples each containing 30 events.

References

[1] J. R. Myra, D. A. DIppolito, D. P. Stotler, S. J. Zweben, B. P. LeBlanc, J. E. Menard, R. J. Maqueda, and J. Boedo. Blob birth and transport in the tokamak edge plasma: Analysis of imaging data. *Plasma Physics*, 13:092509, 2006.

[2] A. Kirk, N. Ben Ayed, G. Counsell, B. Dudson, T. Eich, A. Herrmann, B. Koch, R. Martin, A. Meakins, S. Saarelma, R. Scannell, S. Tallents, M. Walsh, H. R. Wilson, and the MAST team. Filament structures at the plasma edge on MAST. *Plasma Physics and Controlled Fusion*, 48:B433, 2006.

[3] S.J. Zweben and R.W. Gould. Scaling of edge-plasma turbulence in the Caltech tokamak. *Nuclear Fusion*, 23(12):1625, 1983.

[4] S.J. Zweben and R.W. Gould. Structure of edge-plasma turbulence in the Caltech tokamak. *Nuclear Fusion*, 25(2):171, 1985.

[5] B.A. Carreras, V.E. Lynch, D. E. Newman, R. Balbn, J. Bleuel, M.A. Pedrosa, M. Endler, B. van Milligen, E. Snchez, and C. Hidalgo. Intermittency of plasma edge fluctuation data: Multifractal analysis. *Physics of Plasmas*, 7(8):3278, 2000.

[6] V. Filippas, R.D. Bengtson, M. Meier G-X. Li andi, Ch.P. Ritz, and E.J. Powers. Conditional analysis of floating potential fluctuations at the edge of the texas experimental tokamak upgrade (TEXT-U). *Physics of Plasmas*, 2:839, 1985.

[7] J.A. Boedo, D. Rudakov, R. Moyer, S. Krasheninnikov, D. Whyte, G. McKee, G. Tynan,

M. Schaffer, P.Stangeby, P. West, S. Allen, T. Evans, R. Fonck, E. Hollmann, A. Leonard, A. Mahdavi, G. Porter, M. Tillack, and G. Antar. Transport by intermittent convection in the boundary of the DIII-D tokamak. *Physics of Plasmas*, 8:8426, 2001.

[8] O. E. Garcia et al. *Nuclear Fusion*, 47:667, 2007.

[9] S.J. Zweben, J.A. Boedo, O. Grulke, C. Hidalg, B. LaBombard, R.J. Maqueda, P. Scarin, and J. L. Terry. Edge turbulence measurements in toroidal fusion devices. *Plasma Phys. Control Fusion*, 49:S1, 2007.

[10] C. Silva et al. Intermittent transport in the JET far-SOL. *Journal of Nuclear Materials*, 390-391:785–788, 2009.

[11] H.W. Müller et al. Latest investigations on fluctuations, ELM filaments and turbulent transport in the SOL of ASDEX Upgrade. *Nuclear Fusion*, 51(7):073023, 2011.

[12] F. Militello et al. Experimental and numerical characterization of the turbulence in the scrape-off layer of MAST. *Plasma Physics and Controlled Fusion*, 55:025005, 2013.

[13] 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, D. L. Rudako, and NSTX Team. Edge transport studies in the edge and scrape-off layer of the national spherical torus experiment with langmuir probes. *Physics of Plasmas*, 4:042309, 2014.

[14] D. Carralero, G. Birkenmeier, H.W. Müller, P. Manz, P. deMarne, S.H. Müller, F. Reimold, U. Stroth, M. Wischmeier, E. Wolfrum, and The ASDEX Upgrade Team. An experimental investigation of the high density transition of the scrape-off layer transport in ASDEX Upgrade. *Nuclear Fusion*, 54:123005, 2014.

[15] O. Grulke, J.L. Terry, I. Cziegler, B. LaBombard, and O.E. Garcia. Experimental investigation of the parallel structure of fluctuations in the scrape-off layer of alcator cmod. *Nuclear Fusion*, 54(4):043012, 2014.

[16] O.E. Garcia, J. Horacek, and R.A. Pitts. Intermittent fluctuations in the tcv scrape-off layer. *Nuclear Fusion*, 55(6):2002, 2015.

[17] Garcia O.E., R. Kubea, Theodorsena A., Bak J.-G., Hong S.-H., Kim H.-S., KSTAR Project Team, and Pitts R.A. SOL width and intermittent fluctuations in KSTAR. *Nuclear Materials and Energy*, 12:36–43, 2017.

[18] O.E. Garcia. Stochastic modeling of intermittent scrape-off layer plasma fluctuations. *Phys. Rev. Lett.*, 108:265001, 2012.

[19] F. Militello and J.T. Omotani. Scrape off layer profiles interpreted with filament dynamics. *Nuclear Fusion*, 56:104004, 2016.

[20] I.T.Chapman et al. Overview of MAST results. *Nuclear Fusion*, 55:104008, 2015.

[21] MacLatchy C.S., Boucher C., Poirier D.A., , and Gunn J. *Review Of Scientific Instruments*, 63(8):3923, 1992.

[22] Jaques Hadamard. Sur les problmes aux drives partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

[23] A.M. Stuart. Inverse problems: A bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[24] Cotter S.L., Dashti M., and A.M. Stuart. Approximation of bayesian inverse problems for pdes. *SIAM J. Numer. Anal*, 48(1):322–345, 2010.

[25] Iglesias M.A., Lin K., and Stuart A.M.. Well-posed bayesian geometric inverse problems arising in subsurface flow. *Inverse Problems*, 30:114001, 2014.

[26] D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Boston, PA, 2003.

[27] Mukherjee P., Parkinson D., and Liddle A.R. A nested sampling algorithm for cosmological model selection. *Astrophysical Journal*, 638:L51–L54, 2006.

[28] Svensson J., WernerA., and JET-EFDA contributors. Current tomography for axisymmetric plasmas. *Plas. Phys. Cont. Fus.*, 50:085002, 2008.

[29] F. Militello, N.R. Walkden, T. Farley, W.A. Gracias, J. Olsen, F. Riva, L. Easy, N. Fedorczak, I. Lupelli, J. Madsen, A.H. Nielsen, P. Ricci, P. Tamain, and J. Young. Multi-code analysis of
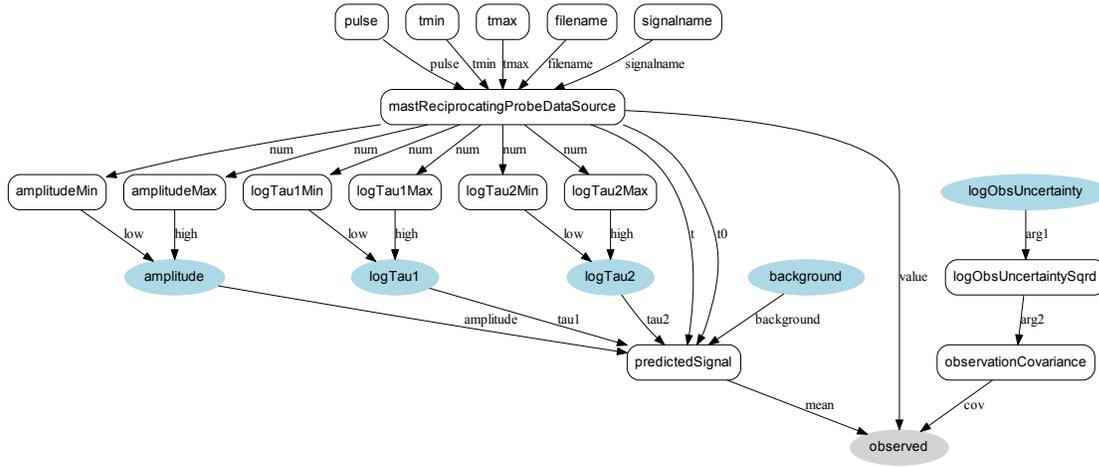
Figure 19. Graphical model employed by Bayesian inference problem.

scrape-off layer filament dynamics in MAST. *Plasma Physics and Controlled Fusion*, 58:105002, 2016.

[30] L. Easy, F. Militello, J. Omotani, B. Dudson, E. Havlkov, P. Tamain, V. Naulin, and A. H. Nielsen. Three dimensional simulations of plasma filaments in the scrape off layer: A comparison with models of reduced dimensionality. *Physics of Plasmas*, 21:122515, 2014.

[31] F. Militello, L. Garzotti, J. Harrison, J.T. Omotani, R. Scannell, S. Allan, A. Kirk, I. Lupelli, A.J. Thornton, and the MAST team. Characterisation of the L-mode scrape off layer in MAST: decay lengths. *Nuclear Fusion*, 56:106006, 2016.

[32] K. Thulasiraman and M.N.S. Swamy. *Graphs: Theory and Algorithms.* John Wiley, 1992.

[33] J.H. Drew, D.L. Evansand A.G.Glen, and L. Leemis. *Computational Probability: Algorithms and Applications in the Mathematical Sciences, chapter4.*

[34] B.S. Gottfried and J. Weisman. *Introduction to Optimization Theory.* Prentice Hall, 1973.

[35] G.O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16:351–367, 2001.

8. Appendix

Figure 19 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 gray. 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[32] corresponding to a factorization of the joint probability distribution ie. from the graph, the joint distribution is identified as

$$P(D, A, \log_{10} \tau_1, \log_{10} \tau_2, B \log_{10} \sigma) = P(D|H)P(A)P(\log \tau_1)P(\log \tau_2)P(B)P(\log \sigma) \quad (6)$$

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[33]. 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:

$$-7 \leq \log_{10} \tau_1^{(i)}[s] \leq -3 \quad \forall i \in \mathbb{N}, 1 \leq i \leq N$$
$$-7 \leq \log_{10} \tau_2^{(i)}[s] \leq -3 \quad \forall i \in \mathbb{N}, 1 \leq i \leq N$$
$$0 \quad \leq A^{(i)} \leq 3 \quad \forall i \in \mathbb{N}, 1 \leq i \leq N$$
$$0 \quad \leq B \leq 3$$
$$-4 \leq \log_{10} \sigma \leq -1$$

The remaining elements in figure 19 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[34]. Markov chain Monte Carlo (MCMC) sampling is carried out with an adaptive algorithm to compute the covariance of the jump distribution which is rescaled following an intiial burn-in phase to ensure an acceptance condition of around 0.75 [35]. The burn-in phase and subsequent posterior exploration phases each used 5×10^6 with 10^4 (ie 1:100) sample points stored for subsequent processing.