

Statistical Characterization and Classification of Edge-Localized Plasma Instabilities

A. J. Webster^{1,2} and R. O. Dendy^{1,2,3}

¹*JET-EFDA, Culham Science Centre, Abingdon, Oxfordshire OX14 3DB, United Kingdom*

²*EURATOM/CCFE Fusion Association, Culham Science Centre, Abingdon, Oxfordshire OX14 3DB, United Kingdom*

³*Department of Physics, Centre for Fusion, Space and Astrophysics, Warwick University, Coventry CV4 7AL, United Kingdom*

(Received 7 January 2013; published 9 April 2013)

The statistics of edge-localized plasma instabilities (ELMs) in toroidal magnetically confined fusion plasmas are considered. From first principles, standard experimentally motivated assumptions are shown to determine a specific probability distribution for the waiting times between ELMs: the Weibull distribution. This is confirmed empirically by a statistically rigorous comparison with a large data set from the Joint European Torus. The successful characterization of ELM waiting times enables future work to progress in various ways. Here we present a quantitative classification of ELM types, complementary to phenomenological approaches. It also informs us about the nature of ELM processes, such as whether they are random or deterministic. The methods are extremely general and can be applied to numerous other quasiperiodic intermittent phenomena.

DOI: [10.1103/PhysRevLett.110.155004](https://doi.org/10.1103/PhysRevLett.110.155004)

PACS numbers: 52.35.Py, 05.45.Tp, 52.55.Dy

Edge-localized plasma instabilities (ELMs) [1–4] are almost ubiquitous in high-performance magnetically confined fusion plasmas. Their phenomenological properties are correlated with the quality of global energy confinement and the peak energy fluxes onto material surfaces [3–6]. Two key challenges are to statistically characterize these processes sufficiently well that a quantitative distinction between different observed classes of ELMs becomes possible and to relate this classification to the physical processes responsible for them. This will provide a test for theoretical models and is an important step towards improved estimates for the distribution of ELM waiting times and sizes, both of which must be controlled in reactor-scale magnetically confined fusion plasma experiments.

ELMs offer a rich and diverse experimental phenomenology [1–8]. There is intense theoretical research on the instabilities that may be responsible for triggering them [9], but few unifying principles have been identified. We will show that widely held experimentally motivated assumptions about the ELM process require particular statistical characteristics. Specifically, if one assumes that the likelihood of ELM occurrence increases monotonically with time elapsed since the most recent ELM, then the measured distribution of waiting times between ELMs should belong to a broad class of probability density functions (PDFs) of which the Weibull distribution [10] is a special case. This physical approach contrasts with a trial and error search for a function that best fits the data [11]. The method is easily generalized to other problems.

To test this conjecture requires the identification and selection of a large representative data set, the development and use of a reliable ELM detection algorithm, and a method to find and compare the best possible fits between data and any proposed PDF. This will provide a rigorous

basis for present and future studies. As an application of our analysis, we distinguish between type I and type III ELMs in a set of plasmas from the Joint European Torus (JET) tokamak [12], on the basis of ELM waiting time statistics alone. Whereas type III ELMs are usually smaller than type I ELMs, typically they are more frequent and the plasma's energy confinement is lower. The ELM type is presently determined by the ELM frequency's response to heating [2–4]. The physically motivated derivation for our PDF allows a clear physical interpretation of our statistical classification.

Theoretical background.—Consider the sequence and distribution of time intervals (waiting times) between ELMs. After an ELM, at $t = 0$, we discuss the statistical properties of the time of the next ELM in terms of two linked functions. We define $p(t)dt$ to be the probability that the next ELM is in the time interval $(t, t + dt)$, given that it has not yet occurred at time t . This differs crucially from the PDF of time intervals between ELMs, which we denote by $P(t)$, and gives the fraction of inter-ELM time intervals that are between t and $(t + dt)$ as $P(t)dt$. Clearly, $p(t)dt$ is a conditional probability, which multiplied by the probability that no ELM occurs between $t = 0$ and t yields the probability $P(t)dt$ of an inter-ELM time interval between t and $t + dt$. This gives the identity

$$P(t) = p(t) \left\{ 1 - \int_0^t P(y) dy \right\}, \quad (1)$$

which allows $p(t)$ to be expressed in terms of $P(t)$. Alternately, Eq. (1) can be used to show that

$$P(t) = -\frac{d}{dt} \exp \left\{ - \int_0^t p(y) dy \right\}, \quad (2)$$

giving $P(t)$ as a function of $p(t)$, with $\int_0^\infty P(t) dt = 1$. The equivalence of Eqs. (1) and (2) can be confirmed by

substituting Eq. (2) into Eq. (1) or by writing Eq. (1) as $p(t) = -(d/dt) \ln[1 - \int_0^t P(y) dy]$ and substituting into Eq. (2).

We adopt the experimentally motivated ansatz that for a short time period t_m immediately after an ELM, $p(t) = 0$, beyond which it starts to increase. The simplest dimensionless representation of this hypothesis is

$$p(t)dt = \begin{cases} 0 & t < t_m \\ \beta \left(\frac{t-t_m}{t_0}\right)^{\beta-1} \frac{dt}{t_0} & t \geq t_m, \end{cases} \quad (3)$$

where t_0 sets the time scale. With the use of Eq. (2), this gives

$$P(t)dt = \begin{cases} 0 & t < t_m \\ \beta \left(\frac{t-t_m}{t_0}\right)^{\beta-1} \exp\left[-\left(\frac{t-t_m}{t_0}\right)^\beta\right] \frac{dt}{t_0} & t \geq t_m. \end{cases} \quad (4)$$

This is a Weibull distribution [10]. It is specified by two dimensionless parameters β and $\alpha = t_m/t_0$, the time scale being set by t_0 . From a theoretical perspective, the values $\beta = 1$ and $\beta = 2$ deserve special mention. Beyond a possible time delay t_m , for $\beta = 1$, $p(t)$ is constant, corresponding to a “memoryless” process in which events occur with equal probability independent of time. The transition between $p(t)$ being a concave (decreasing derivative) and convex (increasing derivative) function is at $\beta = 2$. As β increases, events appear increasingly regular. The preceding derivation assumes that events are independent and that the process causing them is stationary. This approach generalizes easily to other physical problems through an appropriate choice of $p(t)$.

Data sets.—Equation (4) will provide a good fit to a measured sequence of waiting times when the hypothesis represented by Eq. (3) holds. Such distributions have a single maximum and require a macroscopic plasma equilibrium with a quasistationary ELM process. PDFs with additional maxima that are unlikely to have arisen from noise were discarded, as were data whose ELM type was uncertain. A search of carbon-wall JET data yielded a selection of 69 type I and 15 type III ELM data sets. The data sets each have a steady period of H mode with ELMs lasting between 3 and 6 s and plasmas with an energy confinement time typically between 0.25 and 0.4 s. The data sets are listed in the Supplemental Material [13]. The need for quasistationary ELM statistics is met by the pulse length and quality of the JET plasmas studied, which is much improved on the four data sets studied in Ref. [11].

ELM detection.—ELM detection algorithms typically examine the radiation associated with ELMs, using a threshold in amplitude to signal the start of an ELM and a similar threshold or combination of thresholds to determine when an ELM has finished [11]. In those respects, our detection algorithm is the same. The advance of the algorithm described here is that the thresholds are determined from the data in a precise and statistically invariant way so

that we do not need to reset thresholds for different sets of data. This allows statistically robust comparisons between different data sets and enables the technique to be used for non-steady-state and real-time situations if desired. Our algorithm examines the signal intensity of the Balmer-alpha radiation from deuterium (D_α) at JET’s inner divertor and proceeds in two steps. First, a scan is made of the data, obtaining for each time point the box average and standard deviation of the signal intensity for a time interval T immediately prior to that point. The average and standard deviation determine a Gaussian distribution that is subsequently used to distinguish ELMs automatically. For this study, the (D_α) signal threshold for ELM detection was for signal intensities that would only occur 1 time in 20, based on the Gaussian distribution obtained from the data preceding the measurement in question. Once the signal has fallen below the average again, the ELM is considered to have finished. We use a time interval $T = 0.41$ s that is much longer than the time between ELMs but is reasonably short compared with changes to the plasma equilibrium. For stationary pulses such as those here, with ELM waiting times $t \ll T$, results are unchanged by increasing T to the time duration of the entire data set. For cases such as these, T is independent of the data. Because we are interested in classifying ELMs by their statistical properties, here we chose the same threshold for both the type I and III data. The threshold of 1 in 20 was sufficiently sensitive for type III data but kept noise tolerable in type I data. A systematic exploration of these thresholds will be presented elsewhere.

The method just described provides a nonsubjective method to determine when the D_α signal intensity indicates an ELM. Because the study involves the detection and study of many thousands of ELMs, “incorrect” detection or omission of one or more ELMs becomes part of the experimental noise. The detection settings require only one value to be set in advance of an analysis, and because it does not need to be changed or optimized for any given set of data, it is easy and quick to analyze very large data sets. Also, because thresholds are set independently of the data, it is possible to systematically mine noisy data by varying the noise and time-scale parameters to search for patterns in data that would otherwise be obscured.

Best fit and goodness of fit.—Both the Weibull and Gaussian distributions have free parameters that must be chosen to fit the data. A simple fit is provided by using the moments of the data, e.g., average, standard deviation, and skewness, to fit the parameters. More rigorously, we can consider the likelihood function for the probability of the data given the model being considered [14] (e.g., the Weibull model W) and parameters $\bar{\lambda}$, with

$$L(\bar{\lambda}) = P(\{t_i\}|W, \bar{\lambda}), \quad (5)$$

where $P(\{t_i\}|W, \bar{\lambda})$ is the probability of observing the set of waiting times $\{t_i\}$, given the assumption of a Weibull

distribution (W), with fitting parameters $\bar{\lambda}$. The free parameters that maximize $L(\bar{\lambda})$ are their maximum likelihood (ML) estimate [14], for which the likelihood of the data (given the distribution being considered) is a maximum. In practice, the ML estimates are found by starting from the moment-fitted estimates and iterating to find $\bar{\lambda}$ that maximizes $L(\bar{\lambda})$. Given the best fits for two distributions P_A and P_B , we can compare their goodness of fit by calculating their likelihood ratio [14]

$$\Lambda(P_A, P_B) = \frac{P(\{t_i\}|P_A, \bar{\lambda}_A)}{P(\{t_i\}|P_B, \bar{\lambda}_B)}. \quad (6)$$

Under the assumption of independent $\{t_i\}$, the likelihood function and likelihood ratio can be expanded, for example, with $P(\{t_i\}|P_A, \bar{\lambda}_A) = \prod_{i=1}^n P(t_i|P_A, \bar{\lambda}_A)$. Whether P_A or P_B is a better fit to the data is determined by whether Λ is greater or less than 1.

Equation (4) has one more free parameter than a Gaussian. Thus, although Eq. (4) might provide a best fit to the data, the model might not be better, because the fit used an extra parameter. A Bayesian analysis would introduce an extra factor [14] in Eq. (6) to account for this. However, its influence will reduce as the number of ELM time intervals increases. Unless the factor is of order $1/\Lambda$, it will not affect the decision for which is the best fit. For the classification of data, the most important issue is that the PDF (not the model) is a good fit. From that perspective, the issue is not relevant. Equation (6) rigorously indicates which PDF is the best fit, and for the large number of ELMs in our analyses, Eq. (6) is sufficient to determine whether the model is significantly better or worse than a Gaussian.

An absolute measure of goodness of fit is provided by dividing the ELM waiting time axis into intervals, calculating the fraction P_i of observed ELMs in each interval i , and calculating the coefficient of variation $c_W = \langle (P_i - P_W(t_i))^2 \rangle / \langle P_W(t_i) \rangle^2$ between the observed (P_i) and the theoretical [$P_W(t_i)$] values at the midpoint t_i of the interval. This gives a normalized measure of the difference between the observed and theoretical PDFs and provides an absolute measure for goodness of fit. It has the disadvantage of being dependent upon the number of data points used to generate the P_i . Small numbers of points will make c_W susceptible to noise, increasing its value. The choice of time intervals will also affect c_W and consequently affect a fit that minimizes c_W . With enough data, this would no longer be the case, but in practice it prevents c_W from determining a unique best fit. For these reasons, we use a maximum likelihood best fit, which is unique. Similarly, if c_W is used to determine which PDF gives the best fit, the decision is in practice influenced by the choice of time intervals.

ELM classification.—A full listing of the data sets studied, the time intervals over which they were analyzed, and the results from their analysis are presented in the

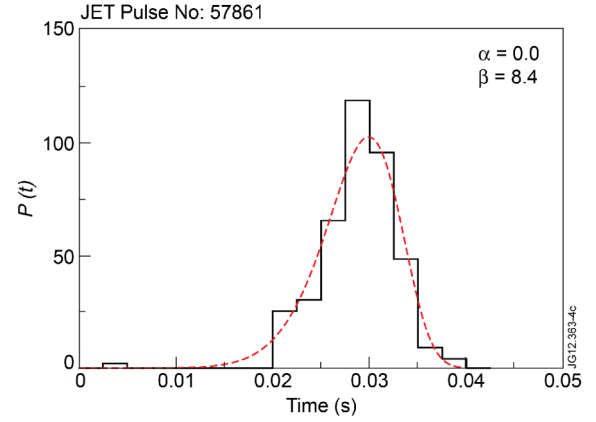


FIG. 1 (color online). Weibull (red dashed line) and experimental PDFs (black bar chart), for JET plasma No. 57861 (type I ELMs).

Supplemental Material [13]. For a data set with n ELMs, we substitute Eq. (4) for P_A and a Gaussian for P_B in Eq. (6), then calculate the geometric mean $\Lambda^{1/n}$, which will be of order 1. If $\Lambda^{1/n}$ is greater (less) than 1.0, then Λ will be much larger (smaller) for $n \gg 1$, indicating whether the Weibull fit is a better (worse) fit than a Gaussian fit. For the type I data sets $\langle \Lambda^{1/n} \rangle = 1.01 \pm 0.04$, where the error of ± 0.04 is the standard deviation and $n \sim 100$. Using time intervals of 2.5×10^{-3} s, the coefficient of variation between the fitted and observed PDFs is $\langle c_W \rangle = 0.63 \pm 0.22$ for the Weibull best fits and $\langle c_G \rangle = 0.63 \pm 0.20$ for the Gaussian best fits. For the type III data sets, $\langle \Lambda^{1/n} \rangle = 1.51 \pm 0.15$, with $n \sim 300$ or larger, $\langle c_W \rangle = 0.70 \pm 0.23$, and $\langle c_G \rangle = 1.25 \pm 0.24$. Typical examples are shown in Figs. 1 and 2. Whereas the fits are similarly good for type I ELMs, the Weibull distribution is the clear best fit for type III ELMs. Substantially improved fits are likely if outliers are removed by improved data, improved ELM detection techniques, or with some

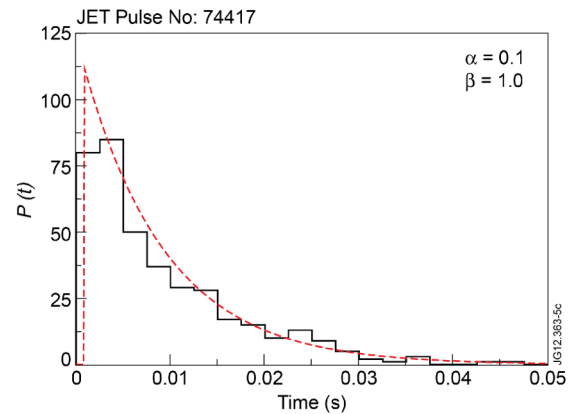


FIG. 2 (color online). Weibull (red dashed line) and experimental PDFs (black bar chart), for JET plasma No. 74417 (type III ELMs).

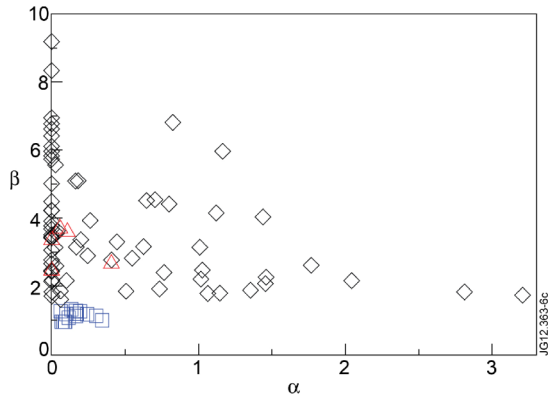


FIG. 3 (color online). Maximum likelihood best fits to Eq. (4): type I ELM database (black diamonds), type III database (blue squares), and some high-frequency ELMs (red triangles). Type III data is characterized by $\beta \sim 1$, whereas all other data has $\beta \geq 2$.

algorithm. The values of c_W and c_G can be reduced if the best fit minimizes them instead of Λ .

Figure 3 plots α and β for the type I and III ELM data sets. There is a clear clustering of type III data for $\beta = 1$ and $\alpha < 0.5$. As noted earlier, $\beta = 1$ has special significance because beyond an initial time delay t_m , it corresponds to a “memoryless” process in which the probability of an ELM is independent of time. The type I data have a wide spread in α and β , but notably β remains of order 2 or larger. As β increases, ELMs will appear increasingly regular. Therefore the type I ELMs studied here are consistent with a process whereby the probability of an ELM increases with time since the previous ELM, possibly due to the buildup of some physical quantity with time. The similarly good agreement between the Gaussian and Weibull fits allows the alternative interpretation that type I ELMs have a specific frequency that is broadened by noise and that the good fit to type III ELM data is coincidental. This is possible, although our original hypothesis is consistent with present ELM models and explains the good fit to both the type I and III data. To avoid disagreement about the classification of ELM types, our data set excludes ELMs whose type is uncertain. Therefore, there could be a continuum between classifications, not observed here. The scatter in α and β is larger for type I ELMs, possibly due to smaller ELM numbers in those data sets. If noise is not the cause, then further classification may be possible.

As an example, we analyzed JET plasmas 66105–66109, whose ELM frequency is typical of type III ELMs [2–4,6], but whose D_α signal is visually similar to that of type I ELMs. On the basis of Fig. 3, they are not type III ELMs.

Conclusions.—We have shown how simple experimentally motivated assumptions require a Weibull PDF for inter-ELM waiting times. The model applies to stationary processes. A search of JET data yielded 84 sufficiently long and steady plasmas to test the model, details of which are in the Supplemental Material [13]. A statistically

rigorous ELM detection technique was developed to compare the data sets from experiments many years apart. The method uses a single dimensionless threshold that is set independently of the data, and a single time period, allowing rapid objective comparisons between different data sets. The data set was analyzed, and a maximum likelihood best fit was calculated, finding a good Weibull fit to both type I and type III data. Therefore we explored whether the dimensionless fitting coefficients α and β could be used to classify the data, concluding that they can. The classification has a clear interpretation: type III ELMs are consistent with a memoryless process, but type I ELMs are consistent with the buildup of a quantity with time, leading to instability. In contrast, present ELM classification requires either a subjective judgment or experimental time to determine how ELM frequency responds to heating [2–4].

To summarize, we have shown that a rigorous statistical analysis of ELM waiting times is possible, and that it can provide a quantitative classification of ELM types and physical insight into the processes responsible for them. The methods have numerous potential future applications, especially for the longer plasma pulses planned for ITER [15]. These include data mining, use in real time and for other signals, and a quantitative characterization of the response of ELM sequences to external parameters.

Thanks to B. Alper, G. Maddison, and M. Beurskens for advice on the ELM data. This work, supported by the European Communities under the Contract of Association between EURATOM and CCFE, was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission. This work was also funded in part by the RCUK Energy Programme under Grant No. EP/I501045. This work was done under the JET-EFDA workprogramme [16].

-
- [1] M. Keilhacker, *Plasma Phys. Controlled Fusion* **26**, 49 (1984).
 - [2] H. Zohm, *Plasma Phys. Controlled Fusion* **38**, 105 (1996).
 - [3] A. Loarte *et al.*, *Plasma Phys. Controlled Fusion* **45**, 1549 (2003).
 - [4] K. Kamiya *et al.*, *Plasma Phys. Controlled Fusion* **49**, S43 (2007).
 - [5] D.C. McDonald, Y. Andrew, G.T.A. Huysmans, A. Loarte, J. Ongena, J. Rapp, and S. Saarelma, *Fusion Sci. Technol.* **53**, 891 (2008).
 - [6] J. Rapp *et al.*, *Nucl. Fusion* **49**, 095012 (2009).
 - [7] Y. Liang, *Fusion Sci. Technol.* **59**, 586 (2011).
 - [8] A. W. Degeling, Y. R. Martin, P. E. Bak, J. B. Lister, and X. Llobet, *Plasma Phys. Controlled Fusion* **43**, 1671 (2001).
 - [9] A. J. Webster, *Nucl. Fusion* **52**, 114023 (2012).
 - [10] W. Weibull, *Trans. ASME* 293 (1951).
 - [11] J. Greenough, S. C. Chapman, R. O. Dendy, and D. J. Ward, *Plasma Phys. Controlled Fusion* **45**, 747 (2003).

- [12] J. Wesson, *Tokamaks* (Oxford University Press, Oxford, 1997).
- [13] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.110.155004> for full details of the data sets studied and the results of their analysis.
- [14] D. S. Sivia, *Data Analysis A Bayesian Tutorial* (Oxford University Press, Oxford, 2005).
- [15] R. Aymar, P. Barabaschi, and Y. Shimomura for The ITER Team, *Plasma Phys. Controlled Fusion* **44**, 519 (2002).
- [16] F. Romanelli *et al.*, *Proceedings of the 24th IAEA Fusion Energy Conference 2012, San Diego, CA, USA* [International Atomic Energy Agency (IAEA), Vienna, 2012]. (All the members of the JET-EFDA collaboration appear in the appendix of this paper.)