Discrimination and visualization of ELM types based on a probabilistic description of inter-ELM waiting times

A. Shabbir[1, 2], G.Verdoolaege[1, 3], O.J.W.F.Kardaun[2], A.J. Webster[4], R.O.Dendy[4, 5], J.M. Noterdaeme[1,2] and JET-EFDA Contributors[1*]

JET-EFDA, Culham Science Centre, Abingdon, OX14 3DB, UK
[1] *Department of Applied Physics, Ghent University, Ghent, Belgium*
[2] *Max-Planck-Institut für Plasmaphysik, Garching, Germany*
[3] *LPP-ERM/KMS, Brussels, Belgium*
[4] *CCFE, Culham Science Centre, Abingdon, UK*
[5] *Department of Physics, Warwick University, Coventry, UK*

Abstract – Discrimination and visualization of different observed classes of edge-localized plasma instabilities (ELMs), using advanced data analysis techniques has been considered. An automated ELM type classifier which effectively incorporates measurement uncertainties is developed herein and applied to the discrimination of type I and type III ELMs in a set of carbon-wall JET plasmas. The approach involves constructing probability density functions (PDFs) for inter-ELM waiting times and global plasma parameters and then utilizing an effective similarity measure for comparing distributions: the Rao geodesic distance (GD). It is demonstrated that complete probability distributions of plasma parameters contain significantly more information than the measurement values alone, enabling effective discrimination of ELM types.

1. Introduction

Characterization of edge-localized modes (ELMs) and their control are crucial for ITER. Enhancement of the physical understanding of ELMs and optimization of control and mitigation schemes necessitates the discrimination of different observed classes of ELMs. We present a technique for systematic classification of ELM types based on a probabilistic description of their properties and propose this as an aid to exploratory and confirmatory analysis for theoretical models. The work can potentially provide a quantitative evaluation of various control mitigation schemes in subsequent works.

2. Experimental setup

The presented technique has been employed for the discrimination of type I and type III ELMs from a series of carbon-wall JET plasmas between the years 2000 and 2009. From the range of discharge numbers [50564, 76871], a database of 69 JET plasmas pertaining to type I ELMs, 27 JET plasmas of type III ELMs and 5 JET plasmas [66105-66109] of so-called type I high-frequency ELMs have been analysed. This is an extension of the data set used earlier by Webster *et al.* [1] for statistical characterization of ELM types. We call this dataset JET_CW ELMy database (DB1), henceforth referred as JET_CW_ELM (DB1). The analysis, in this work, has been restricted to time intervals in

[*] *See the Appendix of F.Romanelli et al., Proceedings of the 24th IAEA Fusion Energy Conference 2012, San Diego, USA.*

which the plasma conditions were quasi-stationary. Further, all experiments dealing with ELM control and mitigation techniques have been excluded.

A threshold based algorithm was developed for the extraction of inter-ELM time intervals from the measured Balmer-alpha radiation signal from deuterium (Dα) at JET's inner divertor. Inter-ELM waiting time extraction is illustrated in Figure 1.

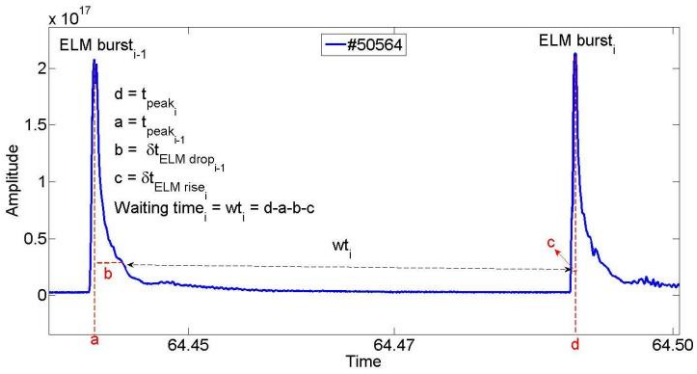

Figure 1: Illustration of the inter-ELM waiting time extraction algorithm where each discharge contains (N+1) ELM bursts and hence N waiting times.

The Weibull distribution, based on experimentally motivated assumptions, has recently been shown to be a good model for the waiting time distribution, especially for type III ELMs [1]. Hence, we use the Gaussian and 2-parameter Weibull PDFs, as illustrated in Figure 2, for describing the series of waiting times emerging from each pulse. It must be emphasised, that Webster *et al*. [1] had earlier used 3-parameter Weibull distributions in a similar context whereas this work deploys 2-parameter Weibull PDFs; hence any direct comparison between estimated parameters is ruled out between the two works.

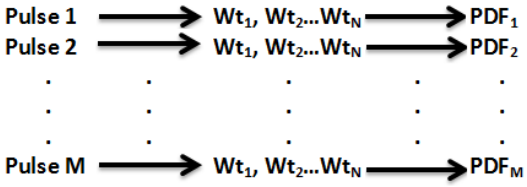

Figure 2. Each pulse is represented as a series of waiting times, followed by modelling by a suitable probability density function (PDF), where there are *M* pulses and each pulse has *N* waiting times.

Finally, density-normalized input power (<P_n>) [*keV/s*] and normalized electron temperature <T_e> [*keV*], were also included in the dataset and used in addition to inter-ELM waiting time statistics for discriminating between type I and type III ELMs. A Gaussian probability distribution was fit to time slices of the signals for these plasma parameters during ELM activity. <P_n> and <T_e> are given as follows:

$$< P_n > = \frac{P_{input}}{n_{e,(vol.avg)} * Volume} \; keV/s \text{ , where } P_{input} = P_{ohmic} + P_{NBI} + P_{ICRH} \; W$$

$$kT = < T_e > = \frac{1}{3} * \frac{W_{thermal}}{n_{e\,,(vol.avg)} * Volume} \; keV$$

3. Classification in probability spaces

The plasma parameters used for discriminating between ELM types are modelled by using suitable PDFs, hence, subsequent processing in the corresponding spaces of probability distributions is required. Classification of ELM types can be regarded as problem typical of the domain of *pattern recognition*, which essentially relies on geometric concepts, particularly distance. Thus, we employ the mathematical

framework of *information geometry*, which treats a family of PDFs as a space wherein each point represents a single PDF, allowing the calculation of the Rao geodesic distance (GD) between probability distributions [2]. A closed-form expression for the GD, existing in the case of a univariate Gaussian model, $f(x:\mu,\sigma)$, and a univariate Weibull model $f(x:\eta,\beta)$, allows accurate and fast computation of the distance between PDFs. See [3].

The *k*-nearest neighbour classifier (*k*-NN), a non-parametric distance-based technique, illustrated in Figure 3 and Figure 4, is used for the classification of ELM types [3]. With *k*-NN, samples are assigned the same class as that of the majority of their *k* nearest neighbours. The nearest neighbours are determined by the shortest distance from the test sample (the sample whose class type is yet unknown) to the samples in the training set (the samples whose class type is known).

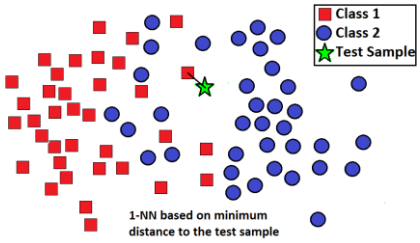

Figure 3: Illustration of 1-NN classifier. The test sample is assigned to Class 1, as the nearest neighbour of the test sample belongs to this class.

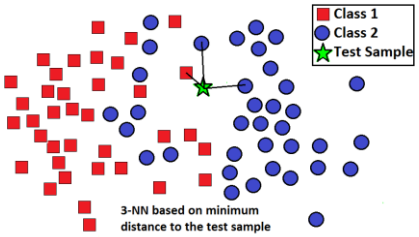

Figure 4: Illustration of 3-NN classifier. The test sample is assigned to class 2, to which the majority of the nearest neighbours belong.

The *k*-NN classifier coupled with GD between PDFs describing the plasma parameters is presented as an automated and effective classification scheme for ELM types. The performance of this scheme is compared with the *k*-NN classifier coupled with Euclidean distance (ED) between measurement values of plasma parameters, and the results are presented in the following section.

4. Classification results

The maximum likelihood best fit parameters for the PDFs of inter-ELM waiting times are illustrated in Figures 5 and 6. From visual inspection of Figure 5 it can be observed that both the mean and standard deviation of the inter-ELM waiting times are determinant of ELM type. Further, Figure 5 indicates a positive correlation between the mean of the inter-ELM waiting time and the standard deviation.

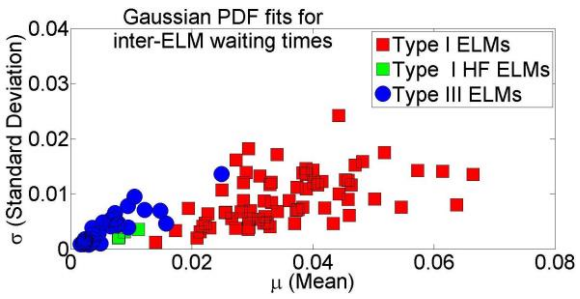

Figure 5: Maximum likelihood best fit parameters for Gaussian PDFs for each ELM type.

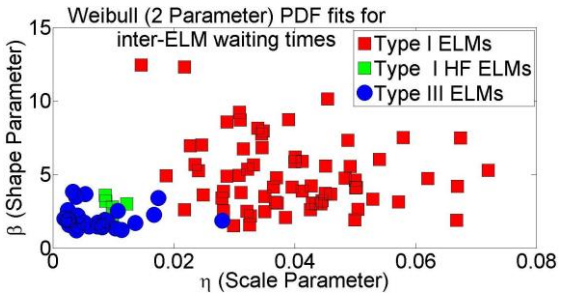

Figure 6: Maximum likelihood best fit parameters for Weibull (2P) PDFs for each ELM type.

Classification of ELM types was performed using a 1-NN classifier (10-fold cross-validated). The success rate (SR) is defined as the percentage of correct classifications, i.e. the percentage of type I and type III ELMs correctly classified. The results are shown in Table 1. Class-wise success rates (SR_I and SR_{III}) are also indicated. It can be readily observed from Table 1, that the success rates using the GD are significantly higher than with the Euclidean distance measure (ED), hence validating that the probabilistic description of plasma parameters contains significantly more information than single measurement values (or averages) alone. Thus, the distribution of inter-ELM waiting times is a crucial predictor for ELM types. Weibull (2P) PDFs give a marginally higher success rate than Gaussian PDFs. This can be attributed to their quality of being a better fit to the type III ELM waiting times [1]. Addition of the global plasma parameters ($<T_e>$ and $<P_n>$) to the predictor set also brings a modest improvement in success rates.

Distance	Predictor	SR	SR_I	SR_{III}
GD	Gaussian PDFs for waiting times	89.31 (0.06)	95.04 (0.06)	83.38 (0.12)
	Weibull PDFs for waiting times	90.50 (0.08)	96.21 (0.05)	84.78 (0.18)
	Gaussian PDFs for waiting times, $<T_e>$, $<P_n>$	91.38 (0.07)	98.77 (0.02)	84.00 (0.12)
ED	Average waiting times	83.50 (0.07)	92.94 (0.02)	74.06 (0.15)
	Weibull PDFs for waiting times	87.36 (0.06)	95.03 (0.02)	79.70 (0.10)
	Average waiting times, $<T_e>$, $<P_n>$	85.12 (0.07)	92.87 (0.03)	77.37 (0.14)

Table 1: Success rates (SR) using a 1-NN classifier based on GD and ED. Success rates for each class (SR_I and SR_{III}) individually are also listed. The standard deviation of each result is mentioned in parentheses.

5. Conclusions and outlook

An automated discriminator between ELM types has been presented and it has been shown that a probabilistic description of plasma parameters, in conjunction with the Rao geodesic distance as a proper PDF similarity measure, improves classification performance.

A natural initial extension to the current work is the deployment of alternate PDFs (such as 3-parameter Weibull) which potentially better capture the underlying statistics of inter-ELM waiting times. In addition, space-resolved pedestal parameters, compared to global parameters, may enhance the discrimination performance. The developed technique will also be applied for classifying additional ELM types, such as type II ELMs, and mapping them in the machine operational space. Furthermore, the method will be used for quantifying ELM properties between various operational regimes (e.g. carbon wall vs. metallic wall) and for inter-machine comparison of ELM behavior, after normalization w.r.t the machine confinement time. Finally the developed method may also contribute to systematic quantification of the effectiveness of ELM mitigation schemes.

Acknowledgements – This work was supported by EURATOM and 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.

References

[1] A.J.Webster *et al.*, Phys. Rev. Lett., **110,** 155004 (2013)**.**

[2] G.Verdoolaege *et al.*, Plasma Phys. Control. Fusion, **54**, art. no. 124006 (6 pp.), 2012.

[3] J.M. Oller *et al.*, Sankhya: The Indian Journal of Statistics, **49**, Series A, Pt. 1 pp.17-23, 1987.