EFDA–JET–PR(14)16

B. Cannas, P.C. de Vries, A. Fanni, A. Murari, A. Pau, G. Sias
and JET EFDA contributors

Advances in Automatic Disruption Classification in JET

Advances in Automatic Disruption Classification in JET

B. Cannas[1], P.C. de Vries[2], A. Fanni[1], A. Murari[3], A. Pau[1], G. Sias[1]
and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]Electrical and Electronic Engineering Dept. - University of Cagliari, Italy
[2]ITER Organization, Route de Vinon sur Verdon, 13115 St. Paul lez Durance, France
[3]Consorzio RFX-Associazione EURATOM ENEA per la Fusione, I-35127 Padova, Italy
** See annex of F. Romanelli et al, "Overview of JET Results",*
(24th IAEA Fusion Energy Conference, San Diego, USA (2012)).

Preprint of Paper to be submitted for publication in
Nuclear Fusion

ABSTRACT

The new full-metal ITER-like wall (ILW) at JET was found to have a deep impact on the physics of disruptions at JET. In order to develop disruption classification, the 10-dimensional operational space of JET with the new ILW has been explored using the Generative Topographic Mapping method (GTM). The 2-dimensional map has been exploited to develop an automatic disruption classification of several disruption classes manually identified. In particular, all the non-intentional disruptions have been considered, that occurred in the JET from 2011 to 2013 performed with the new wall (JET-ILW). A statistical analysis of the plasma parameters describing the operational spaces of JET with Carbon wall (JET-C) and JET-ILW has been performed and some physical considerations have been made on the difference of these two operational spaces and the disruption classes which can be identified. The performance of the JET-ILW GTM classifier is tested in a real-time fashion in conjunction with a disruption predictor presently operating at JET with good results (above 90%). Moreover, to validate and analyse the results another reference classifier has been developed, based on the kNearest Neighbour technique. Finally, in order to verify the reliability of the performed classification, a conformal predictor has been developed which is based on non-conformity measures.

1. INTRODUCTION

Avoidance or mitigation of disruptions is of primary importance in order to preserve the integrity of tokamak machines because disruptions could result in large forces or extreme heat loads. Hence, understanding of disruptive phenomena is particularly important in designing and operating new experimental devices such as ITER, which will have the task of demonstrating the feasibility of fusion energy production from a technical and engineering point of view.

These considerations motivate a strong interest in developing methods and techniques that minimize both the number and the severity of disruptions. The latter can be accomplished by achieving an early detection of a disruptive event such that mitigating actions can be triggered. Therefore, it would be helpful to distinguish the cause of the disruption, because different disruption classes may require a different reactions or mitigation strategies.

The work presented in this paper fits in the broad framework of machine learning techniques that have been exploited as an alternative approach to automatic disruption classification at JET.

Machine learning methods have been extensively used in the field of disruption prediction. In particular, several contributions have been presented using neural networks (NN) in different tokamaks [1]. One of the major drawbacks of the NN approaches is that the network performance normally deteriorates when new plasma configurations are presented to the network. Improvements, from this point of view, might be possible using Novelty Detection techniques [2]. Another successful experience in JET is represented by the real-time Advanced Predictor Of DISruptions (APODIS) [3].

In [4, 5] the authors investigated the possibility of improving the previous black box approaches, which are blind, or non-explanatory, by a process called manifold learning, which finds low dimensional structures in high dimensional data caused by constraints on the data itself.

In [4] the mapping of the multi-dimensional plasma parameter space of ASDEX Upgrade has been performed using a 2D Self Organizing Map (SOM).

In [5], the high dimensional operational space of JET has been analyzed and described using different linear projection methods such as Principal Component Analysis, and non-linear manifold learning techniques such as SOM and GTM. The 2D SOM and/or GTM maps allowed identifying characteristic regions of the plasma scenario and discriminating between regions with high risk of disruption and those with low risk of disruption.

Fewer efforts have been made to apply machine learning techniques to disruption classification, even if being able not only to predict but also classify the type of disruption will enable one to better choose the appropriate mitigation strategy.

The first attempt to automatically classify disruptions at JET was described in [6] using pattern recognition techniques. Disruptions for training were manually classified by some of the authors, in collaboration with physicists at JET, in four classes.

It has to be highlighted that, manually classifying disruption type is essential to develop any automated classification system. In [7] and [8] both the proposed automatic disruption classifiers were based on the manual classification proposed in [9] for the discharges occurring during the JET operations with the Carbon Wall from 2000 to 2010. In [9] specific chains-of-events that led to disruption have been identified and used to classify disruptions, grouping those that follow specific paths. Sometimes these paths are clear and unique, while others could follow near similar courses. Moreover, several different problems may occur simultaneously, eventually leading to a disruption. Hence, not always an unambiguous manual classification is possible.

In [5] the potentiality of the GTM mapping of the JET-C operational space has been exploited to develop an automatic disruption classifier of seven disruption types classified in [9], showing a great potential in terms of classification success rate (exceeding 97%).

In [8] a clustering method, based on the geodesic distance on a probabilistic manifold, has been applied to the JET-C disruption database. The developed technique identifies the type of disruption with 85% confidence, several hundreds of ms before the thermal quench.

The new full-metal ITER-like wall at JET was found to have a deep impact on the physics of disruptions at JET. Such impact has been analyzed in [10, de Vries APS 2013] where it has been stressed that the main difference between JET-C and JET-ILW is the lengthening of the current quench due to lower radiation and higher temperatures during the disruption, which increases the impulse to the vessel and conducts a larger fraction of energy to the wall. This is aggravated by the fact that the ILW is more vulnerable to heat loads.

Regarding the disruption causes, differences between JET-ILW and JET-C have been identified in [10, de Vries APS 2013] for 2011 and 2012 campaigns. The predominant effect of the ILW on disruption causes was the change in density limit, more disruptions due to error field locked mode, and a new class of disruptions, due to accumulation of high-Z impurities. The error field locked modes became more common with the JET-ILW because the density could drop significantly in case of failure of the gas injection system, allowing these modes to grow, while with the JET-C the density would remain higher, due to wall recycling. Accumulation of high-Z impurity has been observed in special cases with the JET-C. However, with the JET-ILW it becomes the predominant disruption cause at JET [de Vries APS 2013].

In the present paper, a statistical analysis on JET-C and JET-ILW disruptions have been performed to investigate how the modification of disruption physics in the JET-ILW experiments eventually influences the operational space of JET. The analysis showed the necessity to develop a specialized GTM map of the JET-ILW 10-dimensional plasma parameter space for disruption classification purposes. Results of the mapping have been reported showing the suitability of the proposed method for the classification task, simulating the on-line application in conjunction with APODIS prediction system. Moreover, the potentiality of the method in giving useful physics insight in the development of disruptions has been discussed.

Furthermore, in order to corroborate the obtained results, those obtained with another classifier based on kNN have been presented. Finally, in order to verify the reliability of the classification, a conformal predictor has been developed which provides information on the level of confidence of the proposed classification.

2. MACHINE LEARNING METHODS

Today the large amount of data available from fusion experiments and their character of high-dimensionality make it particularly difficult to handle, process, and extract properly what is really important among all the available information. In fact very often data sets consist not only of a huge number of examples, but are also characterized by a consistent number of features necessary to exhaustively represent the behavior of a certain phenomenon. Obviously not all the features have necessarily the same level of importance, or it can happen that some of them are redundant or completely useless in relation to a specific objective. This is a key point for several reasons: first of all, even if computer power is continuously increasing, there is a computational limit to the amount of data which can be handled because of the complexity of the algorithms and the required hardware memory. Furthermore, high-dimensionality makes data very difficult to interpret, which is a common scientific problem. The most obvious issue is visualization; when the data dimension is greater than three they cannot be visualized and it becomes harder to perceive similarities and dissimilarities between different variables. Furthermore, the sampling of the space is harder due to the high number of possible data samples, and one has to take into account also the aspect of the computational burden required by pattern recognition, classification and prediction algorithms. Therefore, reducing the quantity of relevant features in a data set is a fundamental step for the subsequent application of powerful data-analysis and machine learning techniques. In the literature a wide range of methods to approach the aforementioned issues are proposed. In the following, the machine learning methods used in this paper for feature extraction, data reduction, data visualization (mapping) and classification are briefly described.

2.1 GENERATIVE TOPOGRAPHIC MAPPING

Generative Topographic Mapping belongs to the class of the so called “generative models”, which try in a various ways to model the distribution of the data by defining a density model with low intrinsic dimensionality in the data space. Through a nonlinear mapping from the latent space to the data space, the GTM generates a mixture of Gaussians, whose centers are constrained to lie on a low dimension space embedded in the high-dimensional one and has to be fitted to the data. This is usually achieved through a form of the Expectation Maximization algorithm by maximizing the likelihood or the log-likelihood function of the model [12].

In a certain way, GTM has been inspired by the SOM algorithm [13], attempting to overcome its limitations. In particular, SOM does not define a density model and the convergence of the prototype vectors are not based on the optimization of an objective function such as the likelihood function, in fact the preservation of the neighborhood structure is not guaranteed. Being a generative latent model, GTM basically tries to find a representation in terms of a small number of latent variables: in order to be able to visualize the lower dimensional representation of the data, the latent variable dimension must be two or three. Since the mapping is defined from the latent space to the data space, for visualization purposes an inversion of the mapping itself is required and this is achieved computing the posterior probability in the latent space through the Bayes' theorem.

However, a single data point corresponds to a probability distribution in the latent space, not just to a single point; therefore, usually condensed information such as the mean or the mode of the posterior distribution are made as reference. The nonlinear mapping between the latent space and the data space can be expressed by a linear regression model: one of the suggested approaches is to use a linear combination of radial basis functions (RBFs), such as for example Gaussians.

Similarly to the SOM algorithm, GTM can be applied for data clustering and topology preservation. Being the mapping defined by a smooth and continuous nonlinear function, the topographic ordering of the latent space will be preserved in the data space, in the sense that points close in the latent space will be mapped onto nodes still close in the data space. Summarizing, GTM explicitly defines a density model (given by the mixture distribution) in the data space, and it allows overcoming several problems, in particular the ones related to the objective function (log likelihood) to be maximized during the training process, and the convergence to a (local) maximum of such an objective function, that is guaranteed by the Expectation Maximization algorithm.

2.2 KNEAREST NEIGHBOUR

The kNearest Neighbours algorithm (k-NN) is a reference non-parametric method used for classification and regression. It represents one of the simpler but at the same time more used learning algorithms. An object can be classified on the base of its neighbors classification by a majority vote with the object being assigned to the class with the higher number of neighbors among the knearest ones. kNN is defined as an instance-based classifier, unlike GTM for example, which defines a generative latent model. There are several implementations of this algorithm, such as the weighted version for taking into account the different importance of the neighbors on the base of the distance to the test unlabeled point.

The kNN technique requires the definition of a similarity measure, or in other words a distance measure. The most commonly used metric is the Euclidean distance, but also other metrics such as Hamming distance [14] or Mahalanobis distance [15] can be used depending on structure and properties of the data of interest.

k-NN is a simple and flexible technique whose drawbacks are well known, as for example the application of the majority-voting criterion for classification when the dataset is strongly unbalanced in terms of the different classes. In this case, the class with higher frequency of occurrence can distort the majority vote among knearest neighbors. One solution to overcome this problem is to take into account the distance of each of the knearest neighbors with a weighted sum multiplying for a factor proportional to the inverse of the distance from the considered point to the test unlabeled point.

The method has some strong consistency results. In particular, the algorithm is guaranteed to yield an error rate no worse than twice the Bayes Manifold learning algorithms error rate if the amount of data tends to infinity [16]. Bayes error rate is referred to the optimal decision boundary that provides the lowest probability of error for a classifier, given a distribution of data [17].

2.3 CONFORMAL PREDICTORS

Conformal predictors belong to the wide family of machine learning algorithms that can be applied for prediction and classification purposes. Unlike others methods, they have the peculiarity to provide together with prediction or classification also the corresponding level of confidence [18, 19]. The theory of conformal predictions is based on the principles of algorithmic randomness, and on the Kolmogorov complexity of an i.i.d. (identically independently distributed) sequence of data instances.

Conformal predictors can be used together with any method of prediction, such as support vector machines, neural networks, decision trees, or nearest neighbour classifiers. Recently, a method based on membership functions has been proposed to extend their use also to Fuzzy Logic classifiers [20]. To determine the confidence level for the classification of a new object, it is necessary to estimate how different a new object is from the old examples: to this purpose, usually a nonconformity score is calculated on the base of a defined nonconformity measure.

Let us consider N successive ordered pairs $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$, where $\mathbf{z_i} = (\mathbf{x}_i, y_i)$ represents the generic example, which consists of an object xi and the corresponding label yi. Both the object and the label belong to measurable spaces, respectively the object and the label space. Conformal prediction requires firstly the definition of a nonconformity measure, which measures how different a new example is from old examples. A bag of size $n \in N$ is a collection of n elements and can be given in any order. In the following, a bag of size n will be indicated with the notation. The first step of the conformal prediction algorithm is the computation of the nonconformity scores for any object of the given bag on the base of a defined nonconformity measure A:

$$\alpha_i := A\left(\left\langle z_1, \ldots, z_{i\ 1}, z_{i+1}, \ldots, z_n \right\rangle, z_i\right) \tag{1}$$

Nonconformity scores have not an absolute value, being relative to the particular case considered for the given bag of objects $\langle z_1, \ldots, z_n \rangle$. Therefore, in order to generalize and give a measure of how unusual an element z_i is with respect to the other elements of the bag, its score must be compared with the one of all the other objects. This can be done for example by computing the so-called *p-value*, which is defined by the fraction:

$$p\text{–}value \;=\; \frac{\#\left|\left\{j = j,\ldots,n : \alpha_j \geq \alpha_i\right\}\right|}{n} \tag{2}$$

This fraction, which is the *p-value* for z_i, can assume values between *1/n* and *1*, and represents the normalized number of examples belonging to the bag at least as nonconforming as z_i. The closer to its lower bound *1/n* the *p-value* is the more nonconforming the object z_i is with respect to the other elements of the bag. If *n* is large enough, a high level of nonconformity may define an outlier for the considered class.

In the framework of the classification with conformal predictors, the *p-values* have a double function: they are used to assign the class to a new element and, at the same time, on the base of their values, it is possible to define the goodness and the reliability of the classification itself. Thus, if a new object of unknown label to be classified on the base of the defined nonconformity measure into one of *N* available classes is considered, the conformal predictor will assign to the new object the label with the highest *p-value*. The reliability of the prediction is quantified by two parameters, confidence and credibility, defined as:

$$\begin{aligned} &\mathit{Credibility} = \mathit{Largest\ p\text{-}value}\ \ (\max_j(p_j), j = 1,\ldots,N). \\ &\mathit{Confidence} = 1 - 2^{nd}\ \mathit{largest\ p\text{-}value} \end{aligned} \tag{3}$$

The values of credibility and confidence are indicative of the reliability with which the classification is provided. In particular, assuming that each class is statistically well represented in the training set, a low value of credibility means that the new object (test) is not representative of any class of objects in the bag (training set). Another important point is represented by the fact that the maximum *p-value* is not necessarily defined in a unique way, in the sense that the maximum *p-value* could be attributed to more than one class. This is a case of ambiguity, which means the conformal predictor for the given training set, on the base of the defined nonconformity measure, is not able to discriminate among the classes which the maximum *p-value* is associated with.

As it has been anticipated at the beginning of this section, the nonconformity score can be computed in different ways. For the classification purpose of this work, the conformal predictor is based on the nearest neighbour technique. When a new example $z_n = (\mathbf{x}_n, y_n)$ has to be classified, the nearest-neighbour technique finds the object $\mathbf{x}_i$ of the training set closest to the new one ($\mathbf{x}_n$) and assigns its label y_i to the label y_n to be predicted, but it doesn’t provide any information about the confidence of the prediction. On the other hand, conformal predictors measure the nonconformity of the new example with respect to the old ones belonging to the training set quantifying the goodness of the prediction. In particular, for all the possible classes, they compare the distance of the nearest object $\mathbf{x}_i$ with the same label previously attributed, with the distance of the nearest neighbour with a different label, computing the so-called nonconformity scores:

$$\alpha_{\mathrm{i}} = \frac{\min\left\{\left|x_j - x_{\mathrm{i}}\right| : 1 \le j \le n \;\&\; j \ne i \;\&\; y_i = y_j\right\}}{\min\left\{\left|x_j - x_i\right| : 1 \le j \le n \;\&\; j \ne i \;\&\; y_i \ne y_j\right\}} = \frac{\text{distance to z's nearest neighbour with the same label}}{\text{distance to z's nearestneighbour with a different label}} \tag{4}$$

3. AUTOMATIC CLASSIFICATION OF THE JET CARBON WALL DISRUPTIONS

In [7] the GTM of the 10D operational space of JET with Carbon Wall has been used to develop a disruption classifier of seven disruption classes manually classified in [9].

In particular, 243 non-intentional disruptions occurred on JET in the experimental campaigns from 2005 up to 2009, in the shot range between 63718 and 79853, have been considered. In the aforementioned interval, also 1467 safe discharges have been selected. The plasma quantities used to described the operational space are [5]: the plasma current (I_p); the poloidal beta (b_p); the Model Lock Amplitude (*LM*); the Safety Factor at 95% of Poloidal Flux (q_{95}); the Total Input Power (P_{tot}); the Plasma Internal Inductance (l_i); the Plasma Centroid Vertical Position (Z_{cc}); the Line Integrated Plasma Density (ne_{lid}); the Stored Diamagnetic Energy Time Derivative (dW_{dia}/dt); the Total Radiated Power (P_{rad}).

Each signal has been sampled at 1 kHz, and a “safe” label has been associated with each sample of the safe discharges whereas a “disrupted” label has been associated with the last 210 samples of the disruption terminated discharges (one sample every 1 ms in the time interval [$\mathrm{t_D}$210 ms - $\mathrm{t_D}$], where $\mathrm{t_D}$ is the disruption time [7]). Then, a data reduction has been performed for the safe discharges to reduce the huge amount of safe samples and to balance the data set of safe and disrupted samples.

In [9] the non-intentional disruptions in the considered JET-C campaigns have been analysed and associated with particular disruption classes by detecting specific chains-of-events and grouping those that follow definite paths. In particular, the following seven classes have been identified: problems during the Auxiliary Power Shut-Down (ASD); Greenwald Limit (GWL); Impurity Control problem (IMC); Internal Transport Barrier (ITB); Low Density and Low ‘*q*’ (LON); Density Control problem (NC); Neo-Classical Tearing Mode (NTM). It should be noted that the complexity of the disruption process could make this manual classification rather ambiguous and a few disruptions were not able to be classified at all [9]. Nevertheless, this work was essential to develop an automated classification able to help identifying a strategy for disruption avoidance or mitigation.

Making reference to this manual classification, a label corresponding to the disruption types can be associated with each disruptive sample. In Figure 1, the 2D GTM of the 10D JET-C operational space is reported, making reference to the Mode representation [12]. In the GTM, the latent space is a discrete grid of nodes (or cells). The arrangement of nodes is a two-dimensional regular spacing in a 70x70 rectangular grid. Each map unit in the GTM can be associated with a particular composition characterized by a coloured symbol, as shown in the legend in Figure 1.

Beyond the data analysis and the characterization of the operational space, also the potential of such mapping techniques for the disruption classification has been exploited, in order to figure out at least in the feature space, if it is possible to distinguish regions where a certain class results to be predominant with respect to the others. In the case of the GTMs, Figure 1 shows that some classes are quite widespread all over the disruptive regions in the operational space, but also regions where a specific class results to be predominant with respect to the others can be found. Thus, there is not only a well-defined separation between disruptive and non-disruptive regions, but also the possibility to characterize certain regions with a higher probability for a certain class with respect to the others. For example, it can be seen that disruptions due to too strong ITBs, which are characterized by a well-defined physics, are projected in the lower right corner of the GTM, while several regions are interested mostly by both NCs and IMCs.

As previously mentioned, each node in the map is related to samples coming from different classes. By projecting onto the map the temporal evolution of a discharge, each sample results to be associated with a node. For each sample and each class, a class membership can be defined on the base of the percentage of samples of the considered class in the node to which the sample is associated, with respect to the total number of disruptive samples in the node itself. In order to classify a disruptive shot, a majority voting algorithm has been adopted based on the class membership of each class in a prefixed time interval before the disruption. In [7] the classification has been performed in the last 210 ms of the disrupted pulses and the automatic classification was in very good agreement with respect to the manual classification, as reported in Table I.

4. JET-ILW VERSUS JET-C OPERATIONAL SPACE

After the installation of the new ILW it was first attempted to project the disruptions of the JET-ILW campaigns onto the GTM trained with the JET-C discharges, but the performance of the map in classifying the new disruptions significantly deteriorated for certain classes (especially for IMC), probably because of the fact that the operational space, or at least, the considered feature space changed. Therefore, a detailed analysis has been performed to investigate how the modification of the disruption physics, recognised in [10], in the JET-ILW experiments with respect to the JET-C ones, eventually influences the classification space of JET.

As mentioned in the introduction, the most common disruptions during the first phase of operation with the ILW, were those due to accumulation of high-Z impurities, mainly W, and as a consequence excessive core radiation. Originally, for JET-C operations, the class IMC was proposed to deal with disruptions due to impurity control problems. However, for JET-ILW operations it was found that, within the IMC class, a distinct sub-class existed, related particularly to the control of high-Z impurities. Then, the later sub-class has been identified as a new separated disruption class [11] which in this paper is labelled IMC_high-Z.

The original training to detect the IMC class was based on JET-C data and, in these cases, the IMC disruptions were mainly due to low-Z impurities and linked to large edge radiation, resulting in the shrinking of the plasma column, yielding the growth of instabilities that disrupt the plasma. Conversely, the new IMC_high-Z class has features that are quite distinct from the IMC class, such as accumulation of high-Z material, strong core radiation and the formation of hollow temperature profiles, which result in the flattening of the current density profile, yielding again an onset of instabilities [11].

Such disruptions were rare with the JET-C and hence previously not identified as separate class [9]. The root cause of the disruptions due to high-Z impurity accumulation may lie in the edge, where sputtered material enters the plasma, although a clear cause is not often found. For these reasons, disruptions related to high-Z impurity control have been considered separately as a new class. To evaluate whether this modification in the physics of the disruptions has changed the disruption operational space, a statistical analysis has been performed on JET-C and JET-ILW disruption classes. In Table II, the composition of the databases for both the JET-C and the JET-ILW is reported. For JET-C the database consists of 243 non intentional disruptions occurred from 2005 to 2009; for JET-ILW it consists of 149 non intentional disruptions occurred from 2011 to 2013. In Table II the distribution and the occurrence for the different classes are reported. As it can be seen from Table II ("JET-C" and "JET-ILW" columns) and Figure 2, the composition of the two data bases is quite different: in particular, disruptions due to Greenwald limit or due to too strong ITB are no longer present in the new campaigns, whereas the number of disruptions due to IMC consistently increased, as earlier reported [de Vries APS 2013].

Moreover, by considering the new impurity control problem disruption class, the disruptions distribution slightly modifies (see "JET-ILW with IMC_high-Z" column in the same Table II): 81 of the 109 IMC disruptions and one of the NTM become IMC_high-Z. The assignment to the different classes is based on the manual classification described in [de Vries APS 2013].

A statistical analysis has been then performed on the plasma parameters describing the JET-C and the JET-ILW operational spaces. In Figure 3 the probability density functions (pdf) of four plasma parameters related to the last 210ms of the IMC disruptions for the JET-C (red lines) and JET-ILW (grey dashed lines), and IMC_high-Z for JET-ILW (blue dashed lines) are reported: (a) Plasma current Ip; (b) Safety Factor at 95% of Poloidal Flux q95; (c) Plasma Internal Inductance li; (d) Line Integrated Plasma Density nelid. The analysis gives us interesting information in particular for the new IMC class, confirming that a new GTM is needed to represent the JET-ILW operational space. From Figure 3, it can be seen that it is quite difficult to discriminate among classes just from the distribution of the signals. In fact it is well known that what really matters is the combination of the parameters.

Moreover, for the new IMC class the pdf of the internal inductance is shifted towards lower values, whereas the pdf of the electron density is shifted toward higher values. This is a direct indication of the impact of the high-Z material on the core density and its radiation, flattening the current density profile, thus lowering the internal inductance. Further analyses can be performed to compare different behavior of disruption classes passing from JET-C to JET-ILW.

Regarding the density control problem and the impurity control problem classes, Figure 4 reports the probability density functions of I_p and l_i for the last 210 ms of IMC and NC disruptions with JET-C, whereas Figure 5 reports the distributions of the same signals for the IMC, IMC_high-Z and NC disruptions with JET-ILW.

From Figure 5, it can be seen that, for the JET-ILW, both I_p and l_i signals result to be quite different, especially if we compare NC and IMC_high-Z classes. In particular, the new impurity control problem type basically occurs for lower values of plasma inductance, mainly as a results of the flattening or the hollowing of the current profiles. Regarding the plasma current, it can be seen that no NC disruptions occur above 2 MA: note that high values of I_p, in the case of IMC_high-Z disruptions, are probably due to the typical ranges of currents used in the attempt to control high-Z impurity accumulation. Therefore, in this case, the distributions are showing the statistical evidence of the considered databases and not a direct dependence of high values of I_p with high-Z impurities. Conversely, for the JET-C, NC and IMC disruptions share the same region in the operational space [Cannas NF 2013]. This is confirmed also looking at Figure 4, where the probability density functions of I_p and l_i are more or less overlapped.

5. MAPPING OF THE JET-ILW OPERATIONAL SPACE

Starting from the previous statistical analysis and the physical considerations on the new disruption class, a new GTM has been trained to represent the JET-ILW operational space. The training set consists of the last 210ms of the 149 non intentional JET-ILW disruptions (29137 samples); the resulting GTM has a latent space of 36x36 grid of nodes built using 81 radial basis functions (Gaussian shape) with a 1.5 width. In Figure 6 (a) the Mode Representation of the GTM is reported. Figure 6 (b) shows the GTM Pie Plane representation. In such visualization, each node is represented by a pie chart describing the percentage composition in terms of number of samples belonging to the different classes. The samples are diversified according to the colour code reported on the legend in the same figure, with reference to the different classes of disruptions. Both representations highlight a high level of separation among the different classes.

Figures 7 (a) and (b) show the same map (Mode (a) and Pie Plane (b) representations), trained with the same training parameters, where the IMC_high-Z class has been introduced.

In Table III, the level of separation of the different classes is reported in terms of percentage of samples of each class which is projected into nodes entirely composed by samples of the considered class.

Table IV reports the same information of Table III, but with the new impurity type class in addition. It can be seen that the new class, IMC_high-Z, is even better separated with respect to the other classes. In fact, it is interesting to observe that, coherently with what has been found for the JET-C operational space, the main contribution to the nodes shared by samples of density control problem and impurity control problem disruptions is given by the old “IMC” class, whereas the overlapping on the map presented by the new impurity type is mainly with the IMC class itself.

Further useful information can be obtained by looking at the component planes of some signals. The component plane representation expresses the relative component distribution of the input data on the 2D map [7], allowing to identify also by visualization eventual similar patterns or particular behaviours for certain classes. As an example, the differences in terms of the plasma current and the internal inductance for the density control problem and the impurity control problem classes can be easily pointed out by analysing the corresponding component planes shown in Figure 8.

Similar considerations to those made for the probabilities density functions of Figure 5 can be done: in particular, making reference to the GTM map in Figure 7, it is easy to see how impurity control problem disruptions occur typically for higher values of the plasma current and lower values of the internal inductance. These tools, together with the statistical analysis, can provide efficiently non-trivial information of a complex multidimensional space, which usually is quite hard to get with classical methods.

In order to test the performance in classification of the new maps, a real time application has been simulated in conjunction to APODIS: the majority voting algorithm has been applied to the class membership function of a time window of respectively 32 or 64ms right before the time in which APODIS triggers the alarm. Note that, in several cases APODIS gives the alarm significantly in advance with respect to the thermal quench, even hundreds of ms in advance.

Table V reports the performance of the real time automatic classification achieved by the GTM trained considering the classes previously defined for the JET-C. As it can be seen, the global success rate is quite high, reaching values above the 90%, so in very good agreement with the manual classification. The classification performance slightly deteriorates when the new class is considered, as shown in Table VI. This is mainly due to the difficulty to discriminate the new class from the previous impurity control problem one, at least on the base of the selected plasma parameters. Other signals, such as core radiation or radiation peaking, should be included to better discriminate the two IMC classes, but such signals are not always reliable for all the disruptions in the data base.

6. VALIDATION AND COMPARISON

In order to validate and analyse the results obtained with GTM, another reference classifier has been developed based on kNN, which uses as kernel the Mahalonobis distance. This is a reference instant based classifier, unlike GTM that builds a generative latent model. In this case, the majority voting is applied to the *k* closest points in the high dimensional space and it can be interpreted also in terms of Bayes' formalism. Table VII reports the performance of the kNN classifier for the classes identified for the JET-C.

Table VIII shows the kNN performance when the IMC_high-Z class is considered. Also in this case, the global performance is above 90% when the new impurity control problem class is not considered, whereas the performance deteriorates when the new class is considered.

The class-membership function gives us useful information. As an example, in Figure 9 the class-memberships of the Pulse No: 82867 is reported for both GTM and kNN; it results to be an IMC disruption according to the manual classification. It is possible to note a transition among different classes and in particular that between NCs and IMCs or vice versa, which is not uncommon both for JET-C and JET-ILW disruptions. It means that the characteristics of the disruption process change in time, and are detected differently long before the disruption and closer to the disruption time. Note that APODIS alarm is triggered almost two seconds before the thermal quench. It is also very important to point out that both the classifiers converge to the same results, even if, in this specific case, we can observe that for GTM based classifiers the phase where we can associate the highest probability to the correct class is about 400 ms, whereas in the kNN is more than 700ms.

In Figure 10 the time evolution of some of the available signals is reported for the same discharge (No. 82867) with reference to the time window analysed in Figure 9.

As can be seen in Figure 10, a locked mode grows at t = 13.79s, around which a rapid change of the density occurs, followed by a quench of the temperature that, in the subsequent phases, recovers up to the final thermal quench at t=55.73s. Both PTN and APODIS trigger the alarm when the mode locks (see Figure 9) and for both classifiers the discharge evolves as a NC disruption up to the final phase where is correctly classified as IMC, according to the manual classification. Thus, given the complex behaviours that often characterize the evolution of a discharge, it is important to know the reliability of the classification. Literature provides recent methods, such as the conformal predictors, which allow us to take into account also this aspect. To this purpose, a conformal predictor has been developed which is based on non-conformity measures. Note that, conformal predictors have the advantage to provide a measure of the reliability of the classification, even if the well-known constraints related to the computation time restrict their application in real time.

Regarding classification, the conformal predictors can provide the level of reliability of classification itself with two parameters: the credibility and the confidence, which are defined on the base of the *p-values* (see section 2.3). In Figure 11 the class membership provided by the classifier is reported together with the credibility and the confidence levels for the Pulse No: 82867. As can be seen, the credibility, which is the parameter with more variability, is quite low for all the initial phase, and then it rises constantly during the last 400ms, according to the results obtained with the GTM based classifier.

The credibility, even if low in the phase where the conformal predictor assigns the label corresponding to the NC class, is mostly above 0.05, which in literature [21] is often used as threshold for trusting or not a prediction (right side of Figure 11). In general, if the credibility is less than 5%, the considered samples are not representative of the training set, or in other words, they cannot be considered as generated independently from the same distribution. In particular, the credibility falls under the considered threshold in correspondence of the transition between NC and IMC classes. This behaviour could depend on a rapid reconfiguration or a change in the considered parameters' space. Further analysis should be done to clarify this point.

In Figure 12, the class membership function obtained with the GTM (a) and with the kNN (b) based classifiers are reported for the Pulse No: 82569, which has been manually classified as IMC disruption.

It can be noted that, in addition to the agreement in the classification provided by the two methods, the confidence level plotted in Figure 13 remains very high for a long phase. In fact, looking at the projection on the GTM map (see Figure 14), the discharge is evolving in a limited region of the operational space, and this means that the parameters are not changing too much in the considered time interval, at least up to the last phases just before the disruption. This is confirmed by the time evolution of some of the considered plasma parameters, as can be seen in Figure 15.

In Figure 16 an example is shown of a discharge (Pulse No: 82669) that disrupted due to impurity accumulation, i.e. the IMC_high-Z class. Figures 17 and 18 report the classmembership functions calculated through the GTM and the kNN classifiers, and through the conformal predictor, respectively, for the aforementioned pulse. In this case, the accumulation of *W* occurs after a step-down of the Neutral Beam Injection power [10], and the hollowing of the temperature profile can be observed. Eventually the instabilities that are triggered by the broadening of the current density profile lock and a disruption takes place. All the three predictors classify the pulse as IMC_high-Z according to the manual classification. Furthermore, it is interesting to see that when the mode locks there are “jumps” in the classmembership calculated by the conformal predictor, and the credibility in the corresponding time interval drops almost to zero. In the interval prior to the locked mode, again the three classifiers clearly recognize the new impurity type.

CONCLUSIONS

The challenge to automatically discriminate the type of disruption at JET both in the Carbon wall campaigns and in the ITER Like wall ones has been tackled using a GTM manifold learning method. The disruption classes in the JET-ILW have been deeply analysed and compared with those in the JET-C. In particular, the probability density functions of the different plasma parameters highlight the different behaviours of the new impurity control problem disruptions, due to tungsten accumulation in the core of the plasma column, with respect to the old IMC ones. Moreover, the statistical analysis showed the variation of the JET-ILW operational space with respect to that with JET-C.

For this reason, a new GTM map has been trained for JET-ILW. The latter has been used to simulate a real time behaviour of the GTM classifier in conjunction with the prediction system APODIS, which is successfully working on line at JET. The obtained results assess the suitability of the GTM based classifier for real time applications with very good results: the prediction success rate is quite high (above 90%) according to the manual classification. However, even if still high, the performance worsened when the new IMC class is introduced, because it is quite difficult to distinguish this new class from the previously defined IMC class. Furthermore, in order to validate and analyse the obtained results, another reference classifier has been developed based on kNN that uses as kernel the Mahalanobis distance. The performance of the reference classifier is still above 90%, but, also for it, the success rate deteriorates when the new IMC class is introduced. These excellent results motivate the deployment of this tool in the real time digital network (ATM) of JET.

Several visualization tools have been developed for the GTM such as Pie Plane representation or Component Plane representation, which make possible to extract relevant information that confirms the physical characteristics of the different classes. Monitoring the evolution of each disruptive discharge on the GTM, a class membership has been defined by means of which it is possible to perform a statistical analysis of the transitions among different classes.

Finally, in order to verify the reliability of the performed classification, a conformal predictor has been developed which is based on non-conformity measures. The obtained results indicate the suitability of the conformal predictors to assess the reliability of the GTM classification even if the computational time allows their use only in an off line fashion. Unlike kNN and Conformal predictors, GTM model can be exploited for data visualization purposes [5, 7], allowing the analysis of the operational space where the relevant physics takes place.

Summarizing, the developed tools are able to provide physics insight of a complex multidimensional space by allowing to pick up changes in the plasma parameters space or transitions among different states during the evolution of a discharge. They give the possibility, furthermore, to efficiently retrieve relations and dependencies among the parameters, making easier to find out particular behaviours often hidden by the high dimensionality of the data itself.

Future work will be devoted to integrate and refine the proposed approach by considering different weights for certain parameter on the base of conditions or rules to be defined through both physical and statistical considerations. Such integration, together, eventually, with the introduction of constraints, could be fundamental to take into account also additional information such as stability limits. This would give rise to the "supervision" of an unsupervised system through physics and statistic.

ACKNOWLEDGEMENT

This work was supported by EURATOM and 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.

REFERENCES

[1]. B. Cannas, A. Fanni, G. Pautasso, G. Sias and P. Sonato (2010) An adaptive real-time disruption predictor for ASDEX Upgrade, Nuclear Fusion **50** 075004.

[2]. Cannas, R. S. Delogu, A. Fanni, P. Sonato, M. K. Zedda and JET EFDA contributors (2007) Support Vector Machines for disruption prediction and novelty detection at JET Fusion Engineering and Design, **82**, 5, 14, 1124 - 1130.

[3]. J. Vega, S. Dormido-Canto, J. M. López, A. Murari, J. M. Ramírez, R. Moreno, M. Ruiz, D. Alves, R. Felton and JET-EFDA Contributors (2013) Results of the JET real-time disruption predictor in the ITER-like wall campaigns, Fusion Engineering and Design 88 1228-1231.

[4]. R. Aledda, B. Cannas, A. Fanni, G. Sias, G. Pautasso, and the ASDEX Upgrade Team (2012) Mapping of the ASDEX Upgrade Operational Space for Disruption Prediction, IEEE Transactions on Plasma Science, **40**, 3.

[5]. Cannas B., Fanni A., Murari A., Pau A., Sias G. and JET EFDA Contributors (2013) Manifold Learning to Interpret JET High-dimensional Operational Space Plasma Physics and Controlled Fusion **55** doi:10.1088/0741-3335/55/4/045006.

[6]. Cannas, F. Cau, A. Fanni, P. Sonato, M.K. Zedda, andJET-EFDA contributors (2006) Automatic disruption classification at JET: comparison of different pattern recognition techniques, Nucl. Fusion, **46**, 699–708.

[7]. B Cannas, A Fanni, A Murari, A Pau, G Sias, and JET EFDA Contributors (2013) Automatic Disruption Classification based on Manifold Learning for Real Time Applications on JET, Nuclear Fusion **53** 093023.

[8]. A. Murari, P. Boutot, J. Vega, M. Gelfusa, R. Moreno, G. Verdoolaege, P.C. de Vries and JET-EFDA Contributors (2013) Clustering based on the geodesic distance on Gaussian manifolds for the automatic classification of disruptions, Nuclear Fusion **53** 033006.

[9]. de Vries P.C., Johnson M.F., Alper B., Buratti P., Hender T.C., Koslowski H.R., Riccardo V. and JET-EFDA Contributors (2011) Survey of disruption causes at JET, Nuclear Fusion **51** 53018.

[10]. deVries P.C. et al. (2012) The impact of the ITER-like wall at JET on disruptions Plasma Physics and Controlled Fusion **54** 124032

[11]. de Vries et al. (2014) The influence of an ITER-like wall on disruptions at JET Physics of Plasmas 21, 056101.

[12]. Bishop C., Svensén M., Williams C. (1998) GTM: The generative topographic mapping Neural Computation **10** 215–34.

[13]. Kohonen M.T. (1989) Self-Organization and Associative Memory Springer-Verlag, New York.

[14]. Hamming, Richard W. (1950), Error detecting and error correcting codes, Bell System Technical Journal **29** (2): 147–160, MR 0035935.

[15]. Mahalanobis (1936), On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India 2 (1): 49–55. Retrieved 2012–05–03.

[16]. T.M. Cover, P.E. Hart (1967) Nearest neighbor pattern classification IEEE Transactions on Information Theory **13** (1): 21-27.

[17]. R.O. Duda, P.E. Hart, D.G. Stork, (2001) Pattern Classification, Wiley, 2nd Edition

[18]. G. Shafer, V. Vovk (2008) A Tutorial on Conformal Prediction Journal of Machine Learning Research **9** 371- 421.

[19]. V. Vovk, A. Gammerman, G. Shafer (2010) Algorithmic Learn- ing in a Random World, Springer.

[20]. A. Murari, J. Vega, D. Mazon, T. Courregelongue (2014) Preliminary numerical investigations of conformal predictors based on fuzzy logic classifiers Annals of Mathematics and Artificial Intelligence January Doi:10.1007/s10472-014-9399-5 Springer International Publishing.

[21]. A. Gammerman and V. Vovk (2007) Hedging Predictions in Machine Learning, Computer Journal **50**, 151–163.

Class	ASD	GWL	IMC	LON	NC	NTM	TOT
Success Rate	100	100	99	100	100	92	97

Table I: Success rates of the automatic classification performed by GTM.

DISRUPTIONS		JET-C		JET-ILW		JET-ILW with IMC_high-Z	
Labels	Classes	num	num %	num	num %	num	num %
ASD	Auxiliary Power Shut-Down	50	20,58	2	1,34	2	1,34
GWL	Greenwald Limit	9	3,70	0	0,00	0	0,00
IMC	Impurity Control Problem	83	34,16	109	73,15	28	18,79
IMC_high-Z	New Impurity Control Problem	0	0,00	0	0,00	82	55,03
ITB	Internal Transport Barrier	10	4,12	0	0,00	0	0,00
LON	Low density and low q	12	4,94	7	4,70	7	4,70
NC	Density Control problem	58	23,87	22	14,77	22	14,77
NTM	Neo-Classical Tearing Mode	21	8,64	9	6,04	8	5,37

Table II: Composition of the JET-C and JET-ILW non intentional disruption data bases.

Classes	Class samples (%)
ASD	15,86
IMC	93,51
LON	68,16
NC	77,57
NTM	60,38

Table III: Discrimination capability of the GTM model for the considered classes.

Classes	Class samples (%)
ASD	15,86
IMC	72,90
LON	68,16
NC	77,57
NTM	55,36
IMC_high-Z	91,18

Table IV: Discrimination capability of the GTM model with the IMC_high-Z.

	GLOBAL	ASD	IMC	LON	NC	NTM
GTM 32ms	93	100	94	67	100	86
GTM 64ms	94	100	95	67	100	86

Table V: Success rates of the real time automatic classification performed by GTM on the classes identified for the JET-C.

	GLOBAL	ASD	IMC	LON	NC	NTM	IMC_high-Z
GTM 32ms	87	100	68	67	100	83	93
GTM 64ms	86	100	71	67	100	83	89

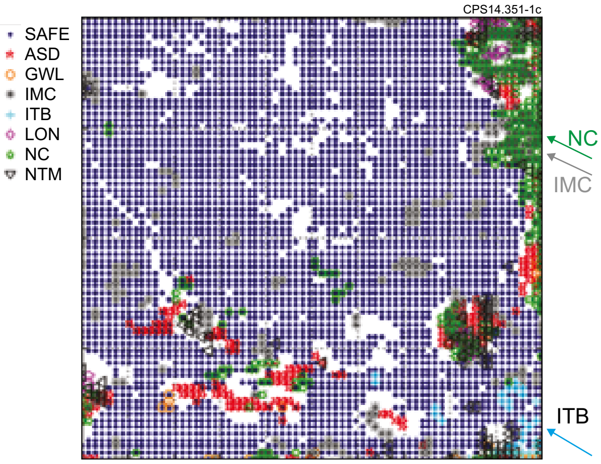
Table VI: Success rates of the real time automatic classification performed by GTM considering the IMC_high-Z disruption class.

	GLOBAL	ASD	IMC	LON	NC	NTM	IMC_high -Z
k-NN 32ms	91	100	82	71	95	83	95
k-NN 64ms	88	100	82	71	90	83	91

Table VII: Success rates of the real time automatic classification performed by kNN classifier considering the classes identified for the JET-C.

	GLOBAL	ASD	IMC	LON	NC	NTM
k-NN 32ms	93	100	95	71	90	86
K-NN 64ms	92	100	95	71	86	86

Table VIII: Success rates of the real time automatic classification performed by kNN classifier considering the IMC_high-Z class.



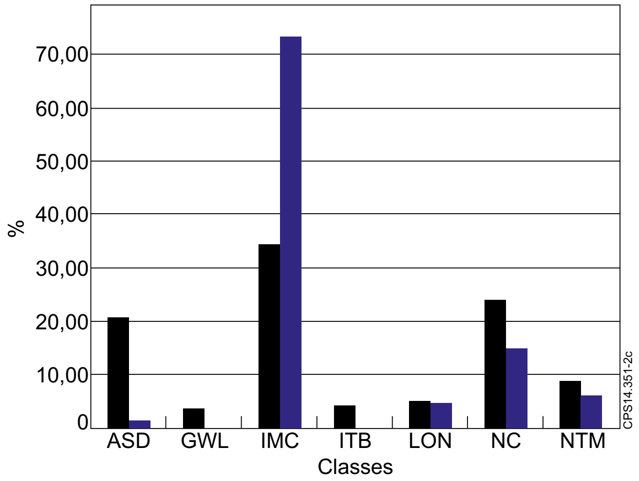


Figure 1: 2D GTM of the 10D JET-C operational space (Mode Representation). The safe nodes are blue, the disruptive nodes are represented with different colours and symbols as indicated in the legend, empty nodes are white.

Figure 2: Distribution of disruptions in the JET-C (black) and JET-ILW (blue) campaigns.

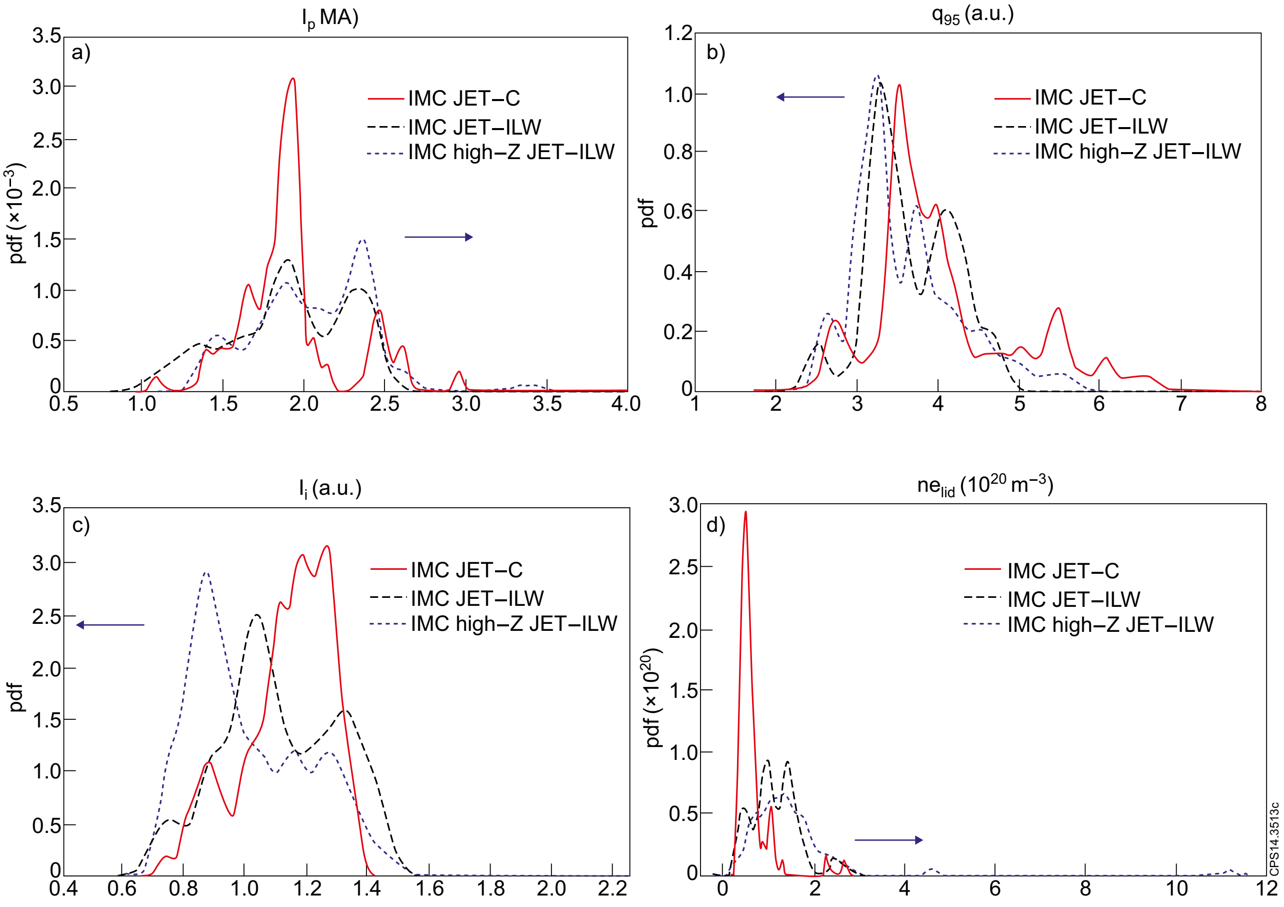


Figure 3: Probability density functions of: (a) Plasma current (I_p); (b) Safety Factor at 95% of Poloidal Flux (q_{95}); (c) Plasma Internal Inductance (li); Line Integrated Plasma Density (ne_{lid}).

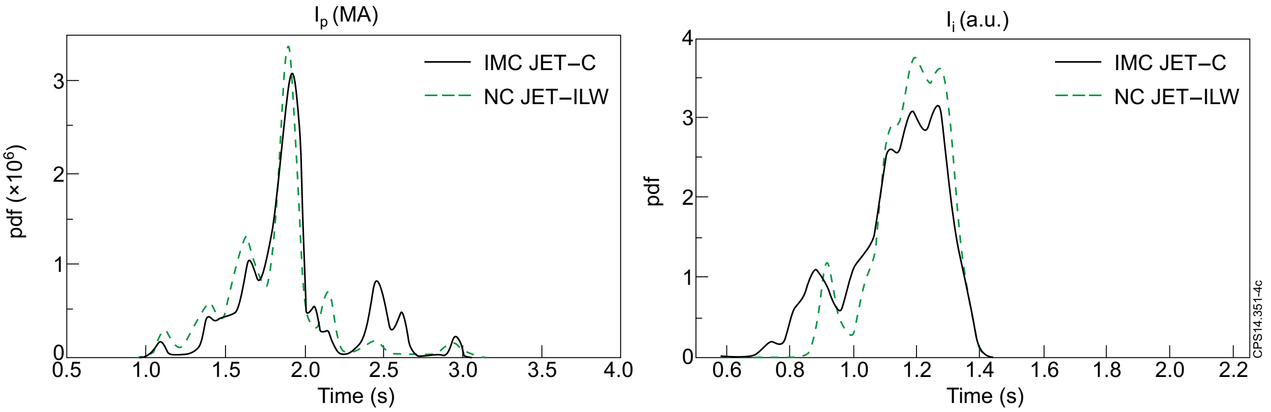


Figure 4: Probability density functions of Ip (left side) and li (right side) for the IMC (grey) and NC (green) disruptions with JET-C.

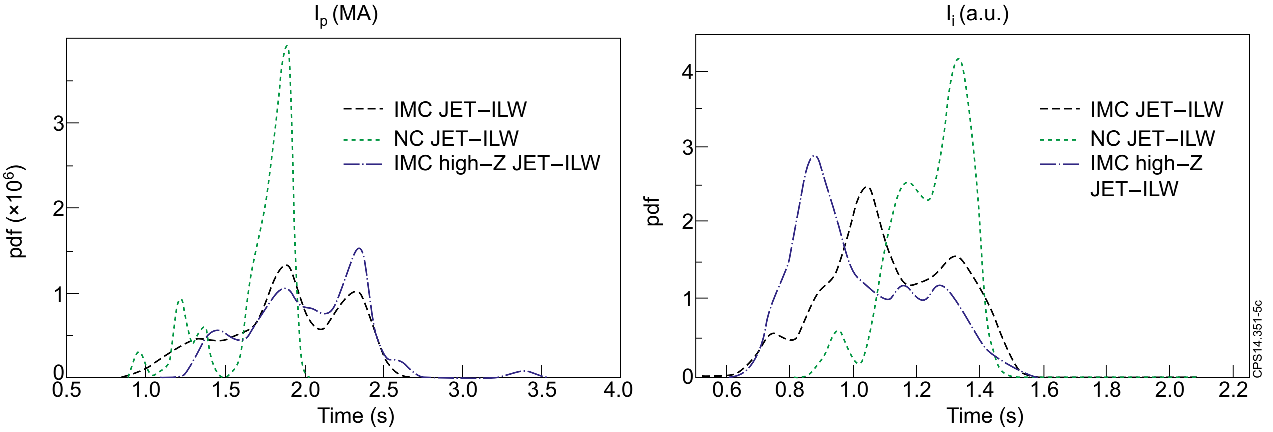


Figure 5: Probability density functions of Ip (left side) and li (right side) for the IMC (dashed grey), IMC_high-Z (dashed blue) and NC (dashed green) disruptions with JET-ILW.

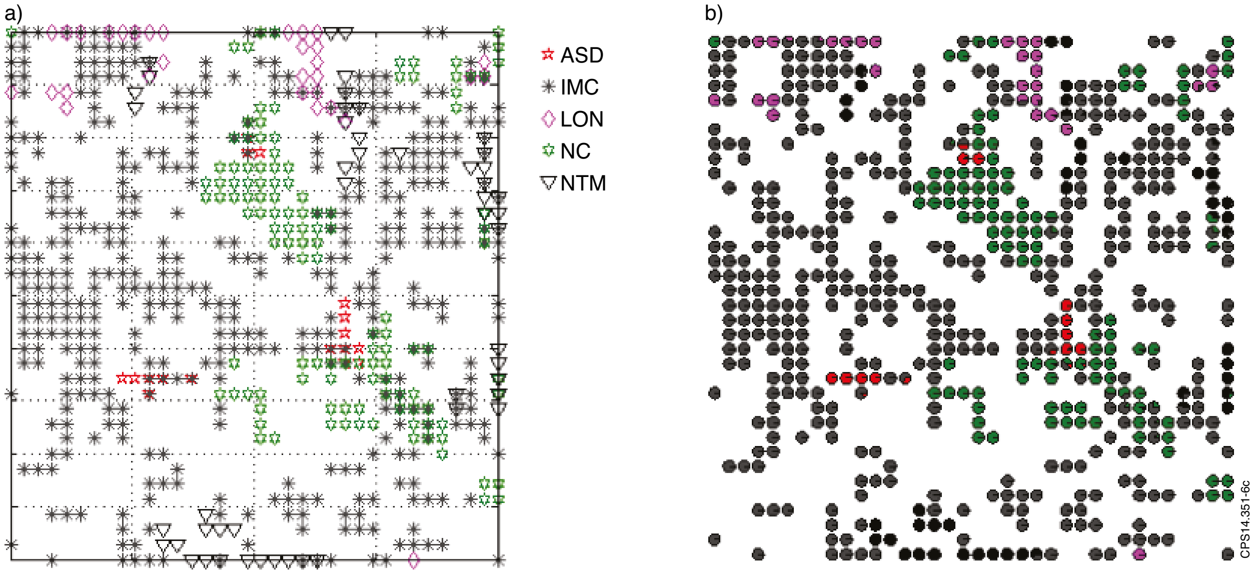


Figure 6: 2D GTM of the 10D JET-ILW operational space: (a) Mode Representation. The nodes are represented with different colours and symbols as indicated in the legend, empty nodes are white; (b) Pie Plane Representation. The nodes composition in terms of the five different classes of disruptions is represented according to the colour code reported on the legend.

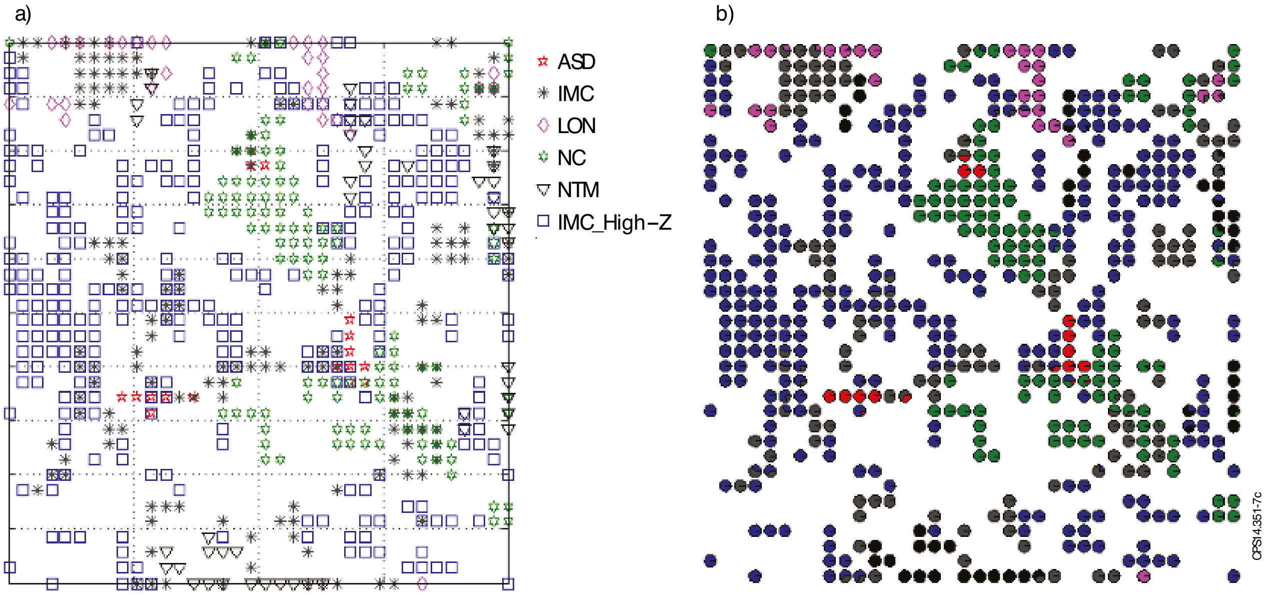


Figure 7:2D GTM of the 10D JET-ILW operational space with the IMC_high-Z disruption class: (a) Mode Representation. The nodes are represented with different colours and symbols as indicated in the legend, empty nodes are white; (b) Pie Plane Representation. The nodes composition in terms of the six different classes of disruptions is represented according to the colour code reported on the legend.

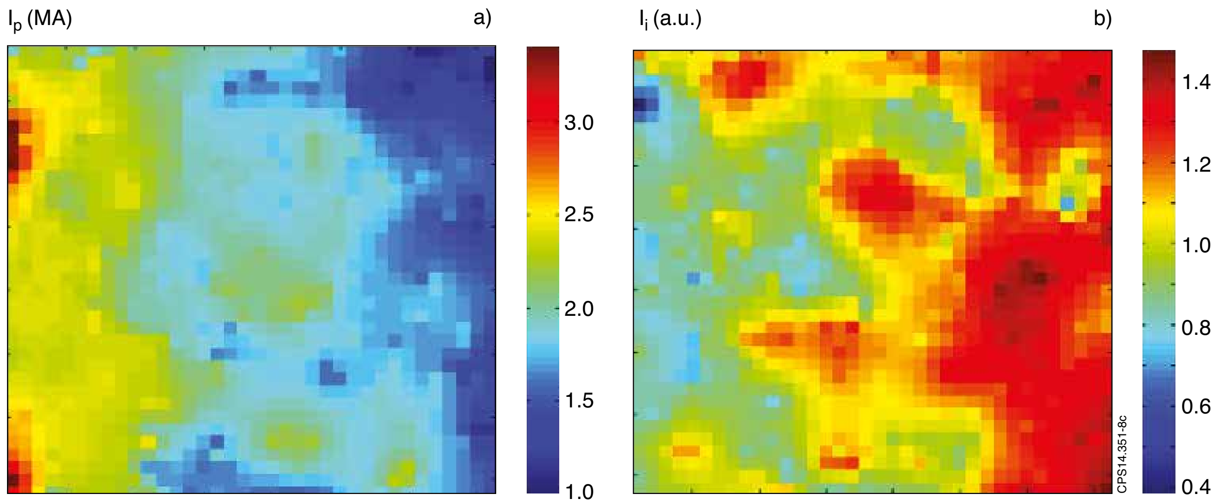


Figure 8: Component planes of the plasma current (left side) and the plasma internal inductance (right side).

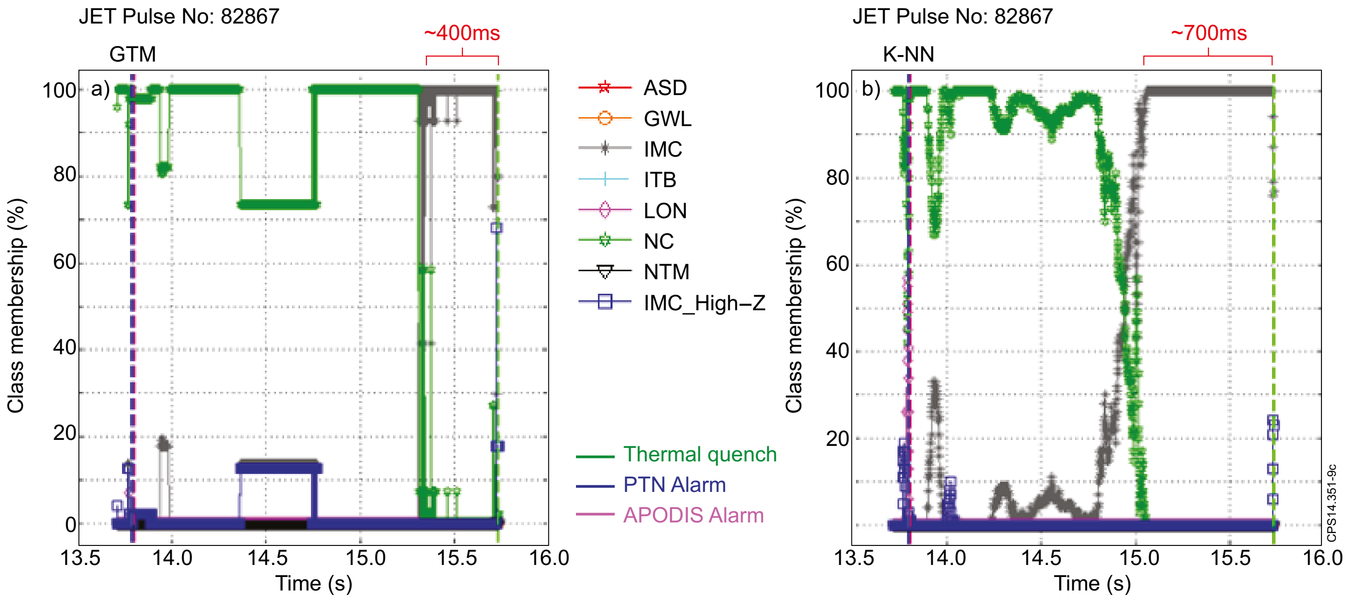


Figure 9: Class-membership functions of the Pulse No. 82867 (IMC) for GTM (left side) and kNN (right side). The vertical green line identifies the thermal quench, the blue line the JET Pulse Termination Network (PTN) alarm, and the pink line the APODIS alarm.

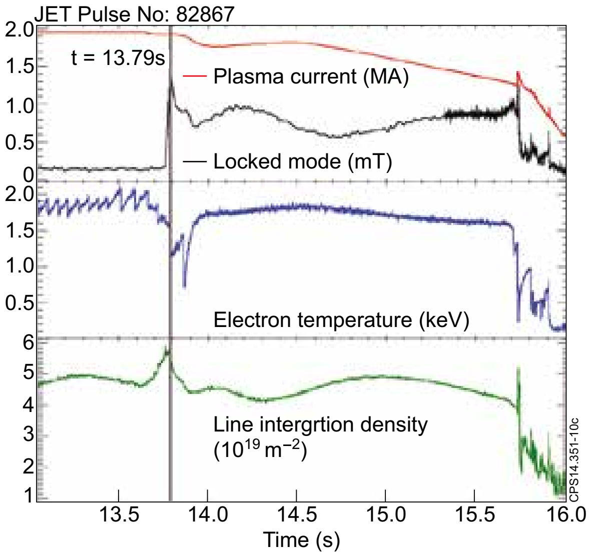


Figure 10: Time evolution of a) plasma current, b) central electron temperature from Electron Cyclotron Emission (ECE) measurements, c) line integrated density and d) locked mode amplitude for the current fat-top phase of the Pulse No: 82867; the vertical green line represents the time of the locked mode (t = 13.79s) that triggers the PTN.

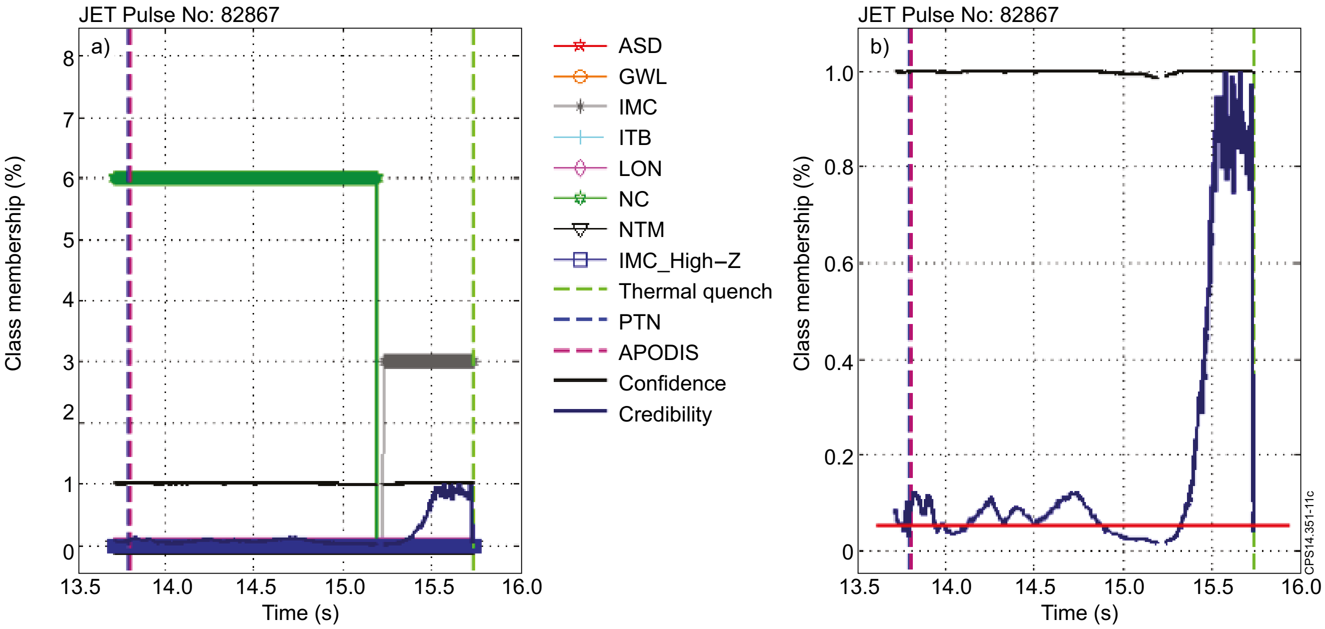


Figure 11: Left side: class-membership provided by the conformal predictor for the Pulse No. 82867, credibility (blue) and confidence level (black). The vertical green line identifies the thermal quench, the blue line the PTN alarm, and the pink line the APODIS alarm. Right side: zoom representing the credibility (blue), the confidence level (black) and the threshold of 0.05 (red).

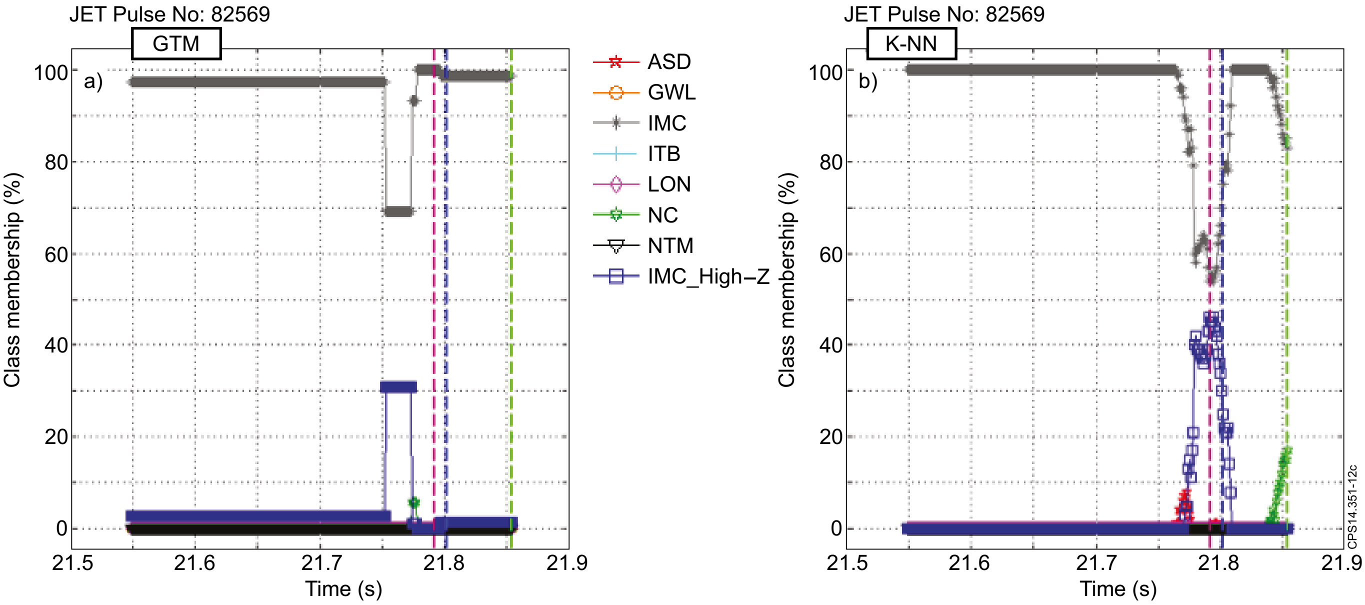


Figure 12: Class-membership functions of the Pulse No. 82569 (IMC) for GTM (left side) and kNN (right Figure 13: Class-membership provided by the conformal predictor for the Pulse No. 82569, credibility (blue) and confidence level (black). The vertical green line identifies the thermal quench, the blue line the PTN alarm, and the pink line the APODIS alarm.

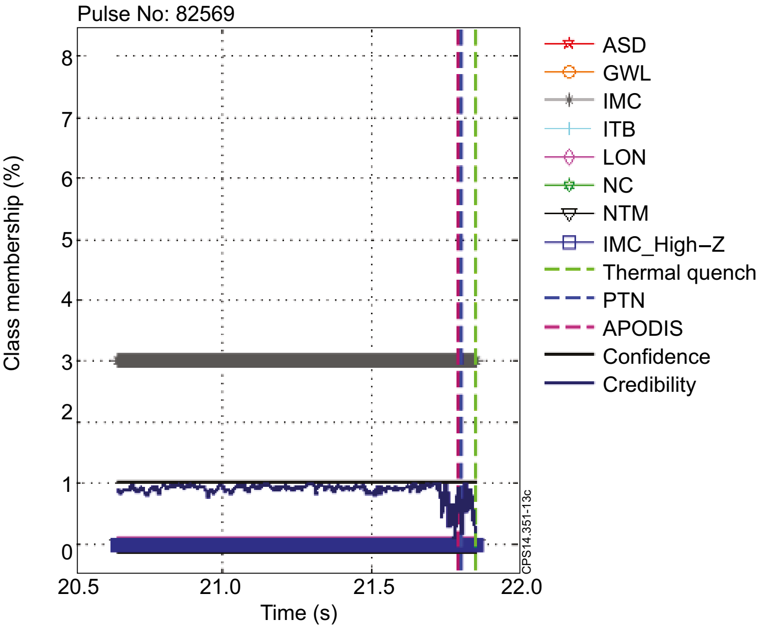


Figure 13: Class-membership provided by the conformal predictor for the Pulse No: 82569, credibility (blue) and confidence level (black). The vertical green line identifies the thermal quench, the blue line the PTN alarm, and the pink line the APODIS alarm.

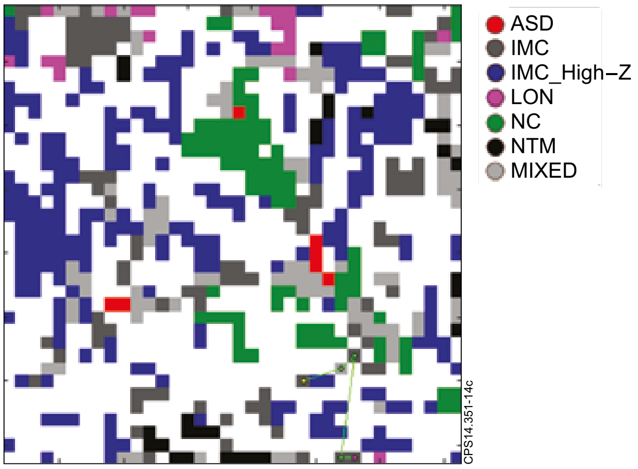


Figure 14: Projection of the Pulse No: 82569 on the GTM map. The nodes are represented with different colours as indicated in the legend, empty nodes are white; the discharge starts from the yellow dot and terminated in the magenta dot.

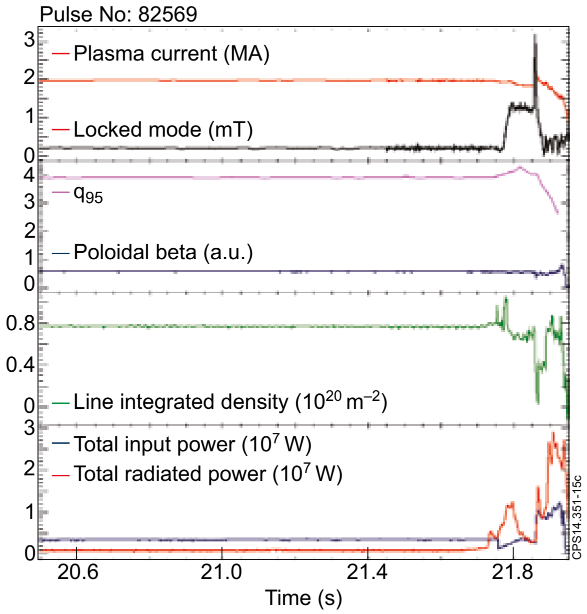


Figure 15: Time evolution of a) plasma current, b) q_{95}, c) line integrated density, d) locked mode amplitude, e) poloidal beta, f) total input power and g) total radiated power measured by bolometer for the Pulse No. 82569.

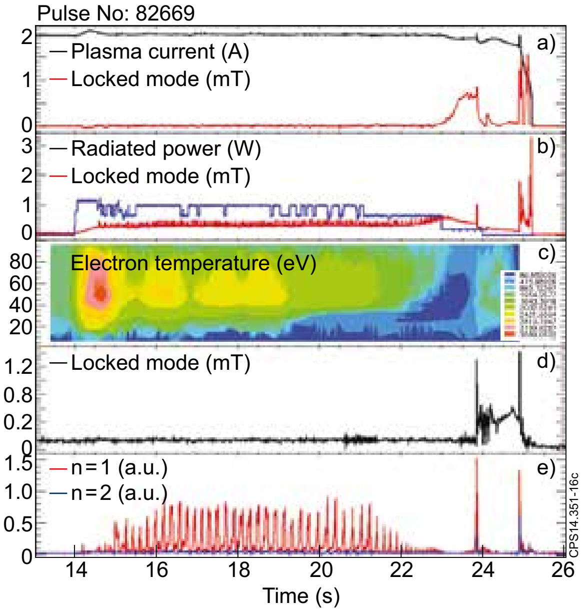


Figure 16: Example of disruption caused by impurity accumulation (discharge No. 82669).

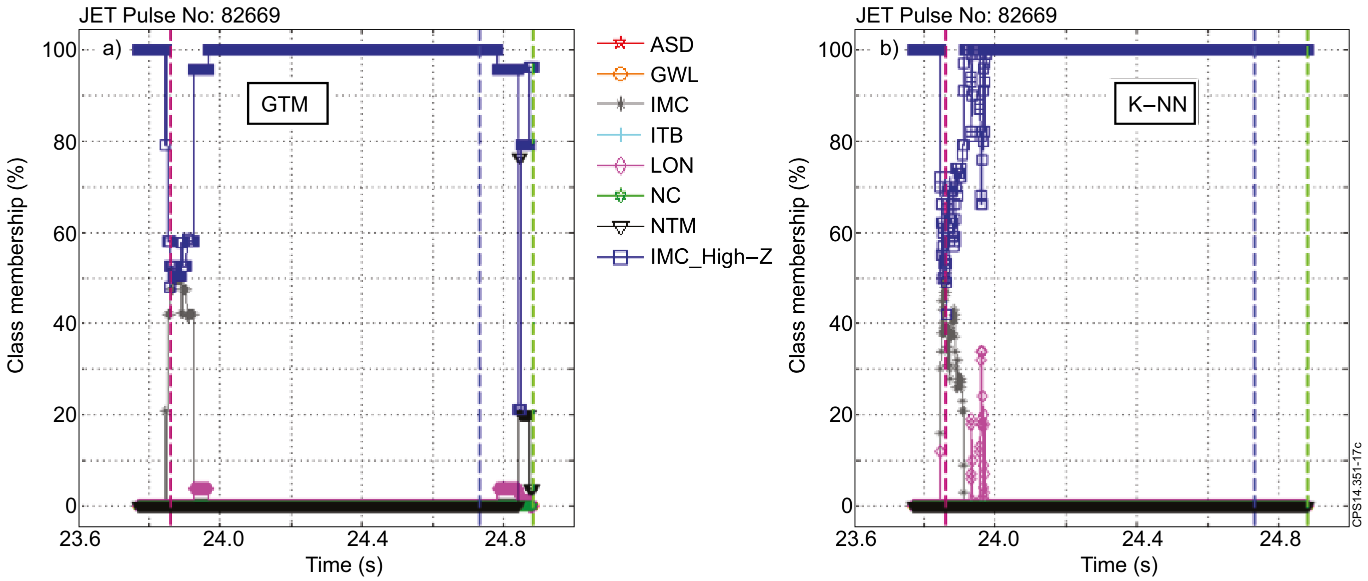


Figure 17: Class-membership functions calculated through a) GTM, b) kNN for the Pulse No: 82669.

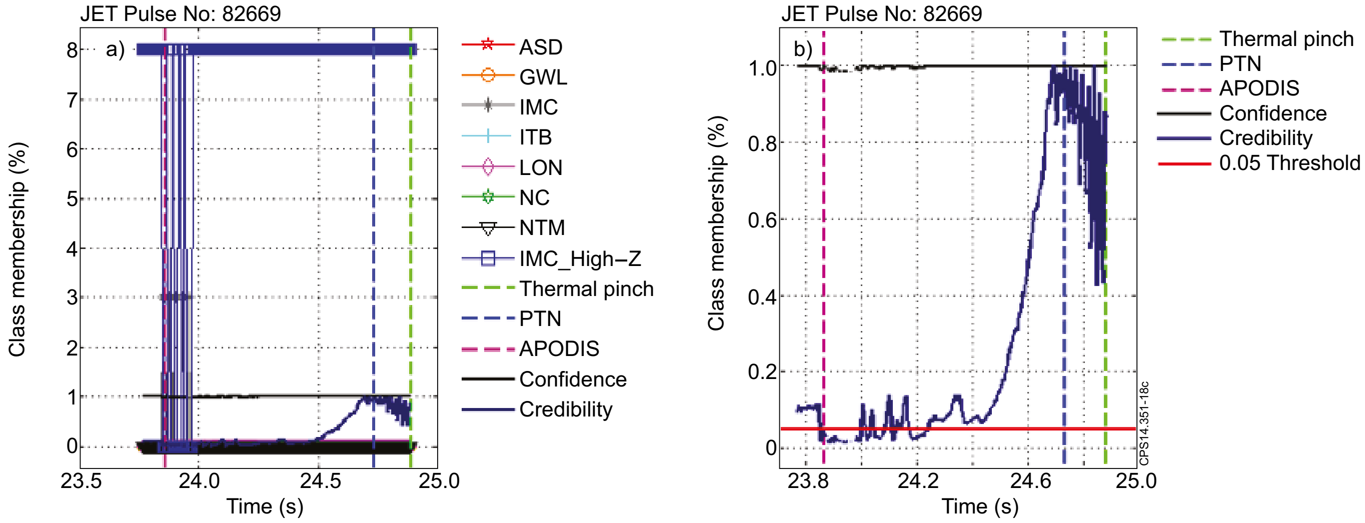


Figure 18: a) Class-membership functions calculated for the Pulse No: 82669 through the conformal predictor; in b) a zoom of a) is reported regarding the confidence level (black) and the credibility (blue).