EFDA–JET–PR(13)10

B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias
and JET EFDA contributors

Automatic Disruption Classification based on Manifold Learning for Real Time Applications on JET

Automatic Disruption Classification based on Manifold Learning for Real Time Applications on JET

B. Cannas[1], A. Fanni[1], A. Murari[2], 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]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.
Disruptions remain the biggest threat to the safe operation of Tokamaks. To efficiently mitigate the negative effects, it is now considered important not only to predict their occurrence but also to be able to determine, with high probability, the type of disruption about to occur. This paper reports the results obtained using the nonlinear Generative Topographic Map manifold learning technique for the automatic classification of disruption types. It has been tested using an extensive database of JET discharges selected from JET campaigns from C15 (year 2005) up to C27 (year 2009). The success rate of the classification is extremely high, sometimes reaching 100%, and therefore the prospects for the deployment of this tool in real time are very promising.

1. INTRODUCTION

The operational space accessible to a Tokamak is highly restricted by disruptive events, major instabilities that cause sudden losses of confinement and the abrupt termination of the discharge. In addition to affecting the execution of the research programme, they can constitute a risk for the structural integrity of the machine. The first phase of the disruption, the so-called thermal quench, can cause extremely high thermal loads on the plasma facing components and more in general on the first wall. As a consequence of the subsequent plasma abrupt current quench, large eddy currents can be induced in the vacuum vessel and surrounding structures creating forces potentially capable of producing damage to the device. Besides, during disruptions the plasma can hit the first wall and induce currents in it. The forces induced by these so called 'halo currents' can be very harmful too. Moreover, the production of relativistic (runaway) electrons during the current quench poses another threat to the integrity of the plasma facing components and the vacuum vessel. Up to now, the occurrence of disruptions has proven to be an unavoidable aspect of Tokamak operation, particularly in high performance configurations, and therefore being able to predict them and undertake mitigation actions is a top priority on the route to the next step devices.

The understanding of disruption physics is an extremely complex task. The amount of available signals in each pulse and the nonlinear relationship between various instabilities has rendered impossible up to now the development of a physical model to reliably recognize and predict the occurrence of this hazardous plasma behaviour. Therefore in the last decade, various machine learning techniques, mainly artificial neural networks and support vector machines (SVMs), have been used as an alternative approach to disruption prediction [1, 2, 3, 4]. The progress has been quite remarkable and recently a new predictor, called APODIS, has been very successfully deployed in JET real time network. Notwithstanding the considerable success rate, the obtained results have not been completely satisfactory, since predictors such as APODIS can foresee the occurrence of a disruption but are not designed to identify its type. On the other hand, to optimise the effectiveness of mitigation systems, it would be essential to predict the type of disruptive event about to occur. Indeed the best strategy to handle a disruptive plasma evolution triggered by an ITB, for example, is not necessarily the same as the one to mitigate a radiative collapse. Reliable prediction of the

disruption type would allow the control and the mitigation systems to optimise the strategy to land the plasma safely and reduce to a minimum the probability of damage to the device.

In [5] a survey of the disruption causes has been carried out over the last decade of JET operations. Each disruption has been manually analysed and associated to a particular disruption class. In particular, specific chains of events have been detected and used to classify disruptions, grouping those that follow specific paths. For JET unintentional disruptions various characteristic sequences of events have been identified. Among them, a number of clear paths could be identified that can be associated with a specific disruption class, e.g., those due to too strong an internal transport barrier and too fast a current rise. It should be noted, however, that the complexity of the disruption process makes this manual classification rather ambiguous. A few disruptions were not able to be classified at all [5]. Nevertheless, this basic work is essential to develop an automated classification able to help identifying a strategy for disruption avoidance or mitigation.

A first attempt to automatically classify disruptions at JET was proposed in [6] using pattern recognition techniques. Disruptions for training were manually classified by the authors, in collaboration with physicists at JET, in four classes: mode lock, density limit/high radiated power, H-mode/L-mode transition, and internal transport barrier plasma disruptions.

In the present paper, the challenge to automatically discriminate the type of disruption has been tackled using a nonlinear manifold learning method. Indeed, a possible approach to the discrimination of disruption types consists of identifying characteristic regions in the operational space where the plasma undergoes a disruption. However, the huge quantity of diagnostic signals that can be regarded as plasma state variables poses significant challenges to the comprehension of the information hidden in the data. Nevertheless, even if the data is embedded in a high D-dimensional space, this does not necessarily imply that its actual dimensionality is D. Assuming that the data of interest lies on an embedded, possibly nonlinear, manifold within the higher-dimensional space, the curse-of-dimensionality can be avoided and the data can be represented well in a low-dimensional subspace. To this purpose, recently, dimensionality reduction and manifold learning methods have been actively investigated [7]. Among the simplest linear methods we can cite Principal Component Analysis (PCA) [8], whereas among nonlinear methods the most popular are the Self Organizing Map (SOM) [9] and its probabilistic variant, Generative Topographic Mapping (GTM) [10].Some attempts have been performed to map the tokamak multidimensional operational space of ASDEX Upgrade using SOMs [11]. In[12] the 2D SOM of the 8-D plasma parameter space of ASDEX Upgrade has been used as disruption predictor by analysing the trajectories described over the map by the discharges under test. Recently, more in the line of investigation pursued in this paper, the use of the Geodesic Distance on a Gaussian manifold [13] provided quite encouraging results.

In [14] SOM and GTM were used to generate 2-D mappings of the 10-D JET operational space. They proved to be superior to linear techniques such as Principal Component Analysis (PCA).

In the present paper, the potentiality of the GTM mapping of the JEToperational space has been exploited to develop an automatic disruption classification of seven disruption classes identified in [5].

In particular, 243 non-intentional disruptions have been considered, that occurred in the JET campaigns from C15 to C27 (Pulse No's: 63718 – 79853).

Each disruption is projected on the map and the probabilities to belong to the different disruption classes are monitored during the time evolution, returning the class which the disrupted pulse more likely belongs to.

The paper is organized as follows: Section 2 describes the basic elements of the GTM manifold learning technique. Section 3 describes the used JET database (with the carbon wall) in quite detail, including the disruption classes and the signals given as input to the manifold learning tool. In Section 4, the mapping of the JET operational space obtained with GTM is presented, and the different disruption regions in the map are analysed with reference to their composition in terms of disruption classes. In Section 5 the results in terms of success rate of automatic classification are described and discussed. Some conclusions are drawn in Section 6.

2. MANIFOLD LEARNING TOOLS

2.1 INTRODUCTION TO MANIFOLD LEARNING

Scientists often deal with problems involving high-dimensional data. 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. Essentially, the amount of data to achieve a given spatial density of examples increases exponentially with the dimensionality of data space (empty space phenomenon).

In absence of simplifying assumptions, algorithms that operate on high-dimensional data are faced with the "curse of dimensionality" and the associated issues, resulting in very high computational requirements. For example, many machine learning algorithms slow down and get stuck in local minima. Reducing data to fewer dimensions often makes analysis algorithms more efficient and can help machine learning algorithms make more accurate predictions.

One approach to simplification is to assume that the data of interest lies on a low-dimensional manifold, embedded in the high-dimensional space. Thus, data reduced to a small enough number of dimensions can be visualized in the low dimensional embedding space. Attempting to uncover this manifold structure in a data set is referred to as *manifold learning*. It is worth mentioning that identifying the right manifold would also allow to better model the relevant physics. Therefore manifold learning has the potential not only to improve the visualization and the intuitive estimation of problems but also to qualitatively increase the understanding of the relevant physics.

In the last few years, many manifold learning techniques have been developed for dimensionality reduction. A number of supervised and unsupervised *linear* dimensionality reduction frameworks have been designed [7], which define specific procedures to choose interesting linear projections of the data such as Principal Component Analysis [8] and Grand Tour [15]. These linear methods can be powerful, but often miss important nonlinear structures in the data.

Recently, several different algorithms have been developed to perform dimensionality reduction of *nonlinear manifolds* [7]. Among them, in this paper GTM potentiality has been investigated to map the JET operational space.

2.2 GENERATIVE TOPOGRAPHIC MAPPING

Generative Topographic Mapping belongs to the class of the so called "generative models", which try in a certain way to model the distribution of the data by defining a density model with low intrinsic dimensionality in the data space. On the other hand, GTM also makes reference to an algorithm which tries to model the distribution of data in terms of latent variables; this approach is based on a nonlinear transformation from latent space to data space with which a constrained mixture of Gaussians is obtained whose parameters are optimized through the Expectation Maximization algorithm [10].

GTM defines a mapping from the latent space (*L*-dimensional space) into the data space (*D*-dimensional space). For visualization purposes, the resulting mapping in the high dimensional space has to be transposed into the low dimensional latent space, which is therefore chosen to be 2 or 3-dimensional. The inversion of the mapping is performed by employing the Bayes'theorem, which allows calculating the posterior probability in the latent space.

So, given a dataset in the data space $\boldsymbol{T} = (\mathbf{t}_1,....,\mathbf{t}_N)$ the first step is to map the latent space, which consists of a regular grid of points $\boldsymbol{X} = (\mathbf{x}_1,...., \mathbf{x}_K)$, into the data space through a parameterized nonlinear function y (x, **W**), where **W** is the matrix of parameters representative of the mapping (see figure 1).

The objective of the GTM is to define a probability distribution over the *D*-dimensional space in terms of latent variables:

$$p(\mathbf{t}) = \int p(\mathbf{t}|\mathrm{x})\, p(\mathbf{x})\, dx \tag{1}$$

Since data in reality only approximately lies on a low dimensional manifold embedded in the data space, a certain noise has been included in the observed data which will be modelled by a radially symmetric Gaussian probability density function centred on the transformed latent points. Thus the distribution of , for a given and, is a spherical Gaussian centered on $y(\mathbf{x},\mathbf{W})$

$$p(\mathbf{t}|x,\ \mathbf{W}\ \beta) = \left(\frac{\beta}{2\delta}\right)^{D/2} \bullet e^{\{-\frac{\beta}{2}\ \|y(\mathrm{x};\mathbf{W})-\mathrm{t}\|^2\}} \tag{2}$$

where the inverse of the parameter is the noise variance. The distribution in **t-**space, for a given value of **W**, could then be obtained by integration over the **x-**distribution. Since the integral is generally not analytically tractable, the latent variable distribution is replaced by a prior distribution consisting of a superposition of delta functions, each one associated to one of the nodes of the regular grid in the latent space

$$p(\mathbf{x}) = \frac{1}{K} \bullet \Sigma_{k=1}^{K} \delta(x - x_k) \tag{3}$$

Substituting (2) and (3) in (1), the distribution function in the data space becomes:

$$p(\mathbf{t}|x, \mathbf{W}, \beta) = \frac{1}{K} \cdot \Sigma_{k=1}^{K} p(\mathbf{t}|x_k, \mathbf{W}, \beta) \tag{4}$$

The suggested approach is to use radial basis function (RBF), such as for example Gaussians, to perform the nonlinear mapping between the latent space and the data space. The mapping can be expressed by a linear regression model, where the mapping function is expressed as a linear combination of these basis functions:

$$y(\mathrm{x}, \mathbf{W}) = \mathbf{W} \cdot \Phi(\mathrm{x}) \tag{5}$$

is a matrix of weight parameters and M is the number of the basis functions.
The adaptive parameters of the model are $\mathbf{W}$ and β. Since the GTM represents a parametric probability density model, it can be fitted to the data set by maximum likelihood, e.g. maximizing the log likelihood function. This can be performed, e.g., using the expectation-maximization algorithm.

The posterior probability $p(\mathbf{x}_k|\mathbf{t})$ of the k-th point $\mathbf{x}$ in the latent space is computed using Bayes' theorem:

$$p(\mathrm{x}_k|\mathbf{t}) = \frac{p(\mathbf{t}|\mathrm{x}_k, \mathbf{W}^*, \beta^*) \cdot p(\mathrm{x}_k)}{\Sigma_{k'} p(\mathbf{t}|\mathrm{x}_k, \mathbf{W}^*, \beta^*) \cdot p(\mathrm{x}_{k'})} \tag{6}$$

For visualizing all the points of the dataset, it is possible to plot the mean (or the mode) of the posterior probability distribution in the latent space. The mean position $\mathbf{x}_{\mathrm{mean}}(t)$ in the latent space is calculated by averaging the coordinates of all nodes taking the responsibilities as weighting factors. In figure 1, the data point $\mathbf{t}^*$ is represented in the latent space as the mean weighted by the posterior probabilities.

Being the mapping defined by the nonlinear function $y(\mathbf{x}; \mathbf{W})$ smooth and continuous, 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 points still close in the data space.

With respect to other manifold learning algorithms, such as Self Organizing Map algorithm, GTM defines explicitly 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.

3. THE DATA BASE OF DISRUPTION TYPES

At JET, disruptions have always been monitored and a global database is available. For many of the disruptive shots, in addition to the time of the disruption, also disruption classes, which are associated to typical chain-of-events, were identified [5].

§In this paper, 243 disruptive discharges belonging to campaigns performed at JET from C15 (year 2005) and up to C27 (year 2009), in the range between Pulse No's: 63718 and 79853, have been considered. Table I reports the seven disruption types identified in the database, and their acronyms, given by [5].

Moreover, in the same table, the number of shots in each class, and the percentage of occurrence in the database, are reported. All these shots took place with the carbon wall.

This established classification is based on the macroscopic symptoms exhibited by the discharges prior to the disruption and allows comparing the results of the proposed automated clustering with the expert classification.

The plasma quantities used to automatically classify these discharges are listed in Table II. The choice of these quantities is mainly due to their availability in real-time and their relation to plasma stability. A near identical set of quantities has already been used in literature [14, 1] and in any case they can be easily made available for real-time applications.

A time instant $t_{pre\text{-}disr}$ has to be defined for the disrupted discharges, which discriminates between the non-disruptive and the disruptive phase. In this paper $t_{pre\text{-}disr}$ has been assumed equal for all the discharges, and it has been set equal to 210ms, a practical time for JET disruptions. The choice of using a unique $t_{pre\text{-}disr}$ for all disruptive pulses is widely shared in literature and in different machines [3, 4, 16-18].

Hence, the dataset for each disruptive pulse consists of the 10 signals made of 210 points each (one sample every 1 ms), in the time interval $[t_D - 210 \div t_D]$, where t_D is the disruption time.

In the aforementioned interval, also 1467 safe discharges have been selected, in order to map the operational space of JET.

The choice of the safe and disruptive shots has been made by means of a statistical analysis [14] in order to identify eventual anomalous signals and a not negligible number has been found to be unusable because of the excessive presence of outliers or a time evolution with no physical meaning, probably due to a fault of the corresponding diagnostic during the acquisition.

Moreover, by analysing the distributions of the signal values, a proper range of variation for each signal has been assumed to clean the data [14]. Data reduction has been performed on the safe samples in order to obtain a balanced data set of safe and disrupted samples [14]. The resulting final database is composed of: 222 flat-top disruptions (38900 samples) and 1467 safe discharges (239965 samples).

Note that, although large outliers have been removed, the selected signals could still contain erroneous data. Thus, the system performance we are going to present in the following sections takes into account the eventual fails of the diagnostics.

4. MAPPING OF JET OPERATIONAL SPACE

In [14] the GTM of the 10-D operational space of JET has been obtained using the data base described in the previous Section 3. This mapis reported in figure 2.

The number of clusters in the GTM has been selected making reference to the SOM presented in [14], where the map dimension was chosen by optimizing some performance indexes commonly used in the literature to evaluate how appropriate the clustering performed by the SOM is [19]. Moreover, limiting the number of clusters preserves the generalization capability of the map. It is mandatory to choose the map dimension to maximize its capacity to discriminate among patterns with different features, at the same time preserving a high generalization capability. A good trade-off between these requirements was achieved with 4998 clusters [14]. For the sake of comparison, also for the GTM, a regular grid of 70x70 (4900) cells (comparable with respect to the number of SOM units) has been considered.

Each sample in the data set has been associated to a label: a safe state is associated to each non-disruptive sample, whereas a disruptive state is associated to each disruptive sample. A colour has been associated to each cluster of the map, depending on its class membership (see figure 2): safe clusters, containing only samples from safe discharges, are blue; disruptive clusters, containing only samples in the last 210ms of disrupted discharges, are red; mixed clusters, containing both safe and disrupted clusters, are grey; and empty clusters are white. Each colour, which is representative of a particular cluster composition, corresponds to a different disruption risk.

The reduction of 10-D data to 2-D space and the grouping of similar data items together, allows one the visualization of the plasma parameter space and the extraction of useful information on characteristic regions of the plasma operational space and their associated risk of disruption. As can be noted from figure 2, GTM presents a large safe region (blue), some disruptive regions (red), well separated from the safe region by mixed and empty regions.

The results in [14] clearly show as nonlinear manifold learning methods, such as SOM and GTM, have high efficiency for data visualization. SOM has a limited computational complexity and many functions are available to visualize, analyse and understand the high dimensional plasma parameter space. Conversely, GTM presents a better discriminating capability, but it has high computation complexity and needs adequate hardware. For the analysis presented here, a double 6-core computer with 32 Gb RAM has been used.

4.1 ANALYSIS OF THE DISRUPTED REGIONS

The temporal sequence of the samples in a discharge can be projected on the map, depicting the movement of the operating point during a discharge. Following the trajectory in the map, it will be possible to eventually recognize the proximity to an operational region where the risk of an imminent disruption is high.

In the present paper, effective real time strategies have been developed to use the JET mapping for classification purposes.

An analysis has been made to find whether the different disruption classes lie in confined regions of the map, i.e., whether the different disrupted regions of the map are associated to particular disruption classes. To this purpose, making reference to the manual disruption classification as reported in [5] a label (corresponding to the disruption types reported in table I) has been associated to each sample of a disruption.

Monitoring the evolution of each disruptive discharge on the GTM it has been found that many of them evolves within the same region. However, some regions can contain samples belonging to different disruption classes, as can be seen in figure 3, where the Auxiliary power shut-down (ASD) and Density control problem (NC) classes are represented. In particular, the clusters in the GTM, which contains samples of ASD (figure 3a) and NC (figure 3b) disruptions, are marked with different shades of bright red and green, in such a way to identify at the same time the class of disruption and the percentage of samples of the considered class with respect to the total number of disruptive samples. ASDs and NCs are two among the most numerous classes of disruption in the considered database. Qualitatively, it can be seen that the two classes mainly occupy different areas in the maps.

The other class with a high frequency of occurrence, both in the considered database and in the totality of non-intentional disruptions on JET, is the Impurity control problem (IMC). These last three classes are quite widespread all over the disruptive regions in the operational space, even if we can find regions where a specific class results to be predominant with respect to the others (see figure 3). This can be seen also making reference to figure 4 where the GTM is visualized using the so-called "pie planes". In such visualization, each cluster is represented by a pie chart describing the percentage composition in terms of number of samples belonging to safe and disruptive shots. The samples belonging to safe discharges are represented in blue, while the ones belonging to disruptive discharges are diversified according to the colour code reported on the legend in the same figure, with reference to the different classes of disruptions. From this figure, it can be seen for example that the clusters in the regions marked with boxes contain samples mainly coming from IMC and NC disrupted discharges.

This very heterogeneous picture could be partially due to the uncertainty of the manual classification or, more likely, to the complexity of the chain-of-events that the disruptions follow during their temporal evolution.

For example, the well-known mechanism leading to an edge cooling disruption could take place because of different reasons, such as a too high edge density or a high impurity density at the edge. In the case of density control problem (NC) and impurity control problem (IMC) disruptions, the two processes could be quite distinct even if both characterized by a high level of radiation. In particular, for a density limit disruption, radiation can be poloidally asymmetric and the instability is often linked to the stability of the divertor detachment and to the formation of MARFEs. While in the case of radiative collapse by impurities, the radiation collapse is poloidally symmetric, shrinking the plasma column and increasing the plasma inductance [20] [21].

The previous considerations are confirmed by looking again at figure 4, which reports also the Component Planes of the internal inductance and the locked mode. The Component Plane representation expresses the relative component distribution of the input data on the 2-D map [14]. From these figures, it can be seen that the regions marked with boxes are mainly characterized by density control problem (NC) and impurity control problem (IMC) disruptions, and show high values of internal inductance and locked mode. The presence of both classes in the described region could be due to the connection with high radiation, even if the processes that lead to disruption are different. In this case the signals contained in the database do not seem to allow a further distinction of the two classes.

Unlike the previously analyzed classes, disruptions due to too strong an internal transport barriers (ITB), occupy a confined region in the right bottom corner in GTM map (see figure 5).
Disruptions due to too strong internal transport barriers (ITB) constitute an important class of disruptions to which one of the shortest duration of the chain-of-events is associated. Being fast, they result to be particularly difficult to detect and typically exhibit the highest energies and heat loads. As it has been shown in [5], disruptions due to too strong internal transport barriers are characterized by the highest ratio between the plasma energy at the time of the disruption and the maximum energy during the last one second of plasma. Plasmas characterized by internal transport barriers exhibit radially localized regions of improved confinement with steep pressure gradients in the plasma core, which in turn could drive instabilities that lead to a disruption. In relation to the achievement of continuous operation it is well known that a large fraction of bootstrap current is necessary, and that discharges exhibiting the formation of ITBs are favourable to this aim. Experimentally, the presence of such a current fraction is usually associated to high-discharges with a weakly positive or negative magnetic shear in the central region of the plasma column. High values of q are probably due to the fact that advanced scenarios are typically run at q=5 and 6.

In figure 5, the clusters where the samples of the disruptions due to too strong internal transport barriers (ITB) are grouped, are visualized. The different cyan shades represent different percentages of samples of the considered class with respect to the total number of disrupted samples in the same cluster. It can be seen that disruptions due to too strong internal transport barriers mainly occupies the region marked with the boxes in figures 5. These regions have also been represented by means of the component planes of the poloidal beta and the safety factor. As expected, disruptions due to too strong ITBs are characterized by high values of these two parameters.

5. RESULTS OF AUTOMATIC DISRUPTION CLASSIFICATION

Each cluster in the map contains samples coming from different disruption types and/or safe samples. By following on the map the temporal sequence of the samples of a disruption (the last 210 samples of the disrupted shots), each sample will be associated to a cluster. For each sample and each class, a *class membership* can be defined, being the percentage of samples of the considered class in the cluster to which the sample belongs, with respect to the total number of disruptive samples in the cluster itself.

In figure 6 the temporal evolution of the class membership of the seven classes (Class membership functions) during the JET discharge No.66313 is reported for the GTM.

As it can be noted, for the majority of samples of this shot, the greatest class membership value corresponds to impurity control problem disruption (IMC), which is the same class assigned to this shot in [5].This is true for a relatively long interval before the disruption time. Note that, during this pulse excessive Neon is introduced into the plasma during a phase with no auxiliary heating, resulting in a radiative collapse. As Neon also increases the density significantly, this could justify the presence of relatively high NC class membership values close to the disruption in figure 6.
The previous analysis shows the power and the versatility of the proposed techniques; the different classes of disruption tend to aggregate according to the self-organization of the map in such a way that each class results to be predominant with respect to the others in particular regions of the operational space. In order to perform an automatic disruption classification using the GTM output and to quantify its effectiveness, a proper classification criterion has been introduced.

In particular, to classify a disruptive shot a majority voting algorithm has been adopted based on the class membership of each class in the whole time interval before the disruption (210ms).

In table III, the results obtained by applying the majority voting to all the 222 disrupted pulses are reported in terms of percentage success rate. A pulse has been considered correctly classified if the automatic system produces the same classification given in [5]. As it can be seen, the success rate of GTM is high for all the considered classes, reaching in some cases even the percentage of 100%.

5.1 DISCUSSION OF THE RESULTS

Even if the analysis of the previous section clearly shows the potentiality of the developed tool, it is important to identify the limits of its discriminating capability in the present configuration.

It is worth noting that disruption classes are defined on the base of the typical chain-of-events, as reported in [5]. In particular, the classification is mostly based on the middle track of these chains. In this work, the automatic classification has been developed taking into account only the last 210 ms of the disruptive discharges. Thus, depending on the length of the typical chain-of-events, it could happen that the classifier is not able to entirely pick up the phenomenology which characterizes a certain class. On the other hand, it could even happen that in the final stage of the discharge the indication about the class changes, as if the disruption is evolving from a certain class to another one. This is basically due to the fact that several different paths can converge towards very similar destabilization of modes that lead in the end to the disruption. Hence, this limits the discrimination capability of the classification system when approaching to the disruption time.

For example, in our classification, a not uncommon phenomenon is observed for density control problem (NC) and impurity control problem (IMC) disruptions, which initially evolve in a region where they could be correctly classified with a high level of confidence, and then evolve in the auxiliary shut down (ASD) class when approaching the disruption time.

In figure 7 the class membership functions for the Pulse No: 67322, manually classified in [5] as

NC, are shown. As it can be seen, the discharge initially evolves in clusters where all the samples belong to the NC class, while, as it approaches to the disruption time, it moves towards clusters where the majority of the samples belongs to ASD class. Note that, an ASD disruption is basically a density control problem during/after the switch of the heating system. Hence, at first glance, the traces of an ASD and of the considered disruption would follow very similar paths.

Another cause that limits the discrimination capability of the classification system is that some classes can exhibit very similar values of some parameters. This is the case of Neoclassical Tearing Mode (NTM) disruptions and those due to too strong internal transport barriers (ITBs) in particular operating scenarios.

In JET several experiments have been carried out for the beta limit assessment, varying the pressure and the q profiles, ranging from Hybrid to ITB configuration, in order to investigate advanced scenarios. During these experiments, among the other instabilities, NTMs with m/n=2/1, 3/2, 4/3 have been observed [22]. In these conditions the value of q_{95} is about 4-5, while the currents assume values in a range slightly above 1MA.

Figure 8a shows the clusters occupied by the Pulse No: 72670, manually classified as NTM. Such a discharge evolves in the right bottom corner of the GTM. As shown in figure 4 (see pie planes representations), the majority of disruptions in this area is due to NTMs and too strong ITBs.

Figure 8b shows the class membership during the time evolution of the shot. As predicted by the previous considerations, NTM and ITB classes are characterized by a comparable value of the class membership functions. In this case, the adopted criterion allows one to correctly classify the disruption as NTM, even if with a reduced level of confidence.

As it can be noted by the component planes for β_p, P_{tot}, and I_p, the right bottom covering both the maps is characterized by high values of poloidal beta (figure 5) and total input power (figure 8c), as well as low values of plasma current (figure 8d).

Discrimination capability could be improved by considering further information, e.g., for example the measurement of pressure profile peaking.

Two other very interesting cases, are represented by the Pulse No: 79772, manually classified in [5] as NC, and the shot No. 79770 manually classified in [5] as LON. In figure 9a and 9b the class membership functions returned by the GTM are reported.

Such discharges were performed for investigating the physics of ELM control with magnetic perturbation fields (EFCC). Our system classified the Pulse No: 79770 as LON, as in [5], whereas, regarding the Pulse No: 79772, classified as NC in [5], the GTM recognizes the presence of the NC class for the whole considered time interval, but the highest class membership is associated to the Low q and low density problem disruption (LON) class. The pulse in the final phase is indeed characterized by low values of the edge safety factor and disrupted at. Actually, during this pulse, when the NBI was switched off, erroneous density control gave rise to too fast a density decrease, leading to too low a density and an error field locked mode.

CONCLUSIONS

The Generative Topographic Mapping (GTM) manifold learning technique has been applied to the problem of determining the disruption type at JET. The aim is to classify the disruptions, so that the mitigation strategy and the plasma landing can be properly optimized, to reduce the impact on the integrity of the device. In this respect, GTM shows a great potential in terms of classification success rate, reaching in some cases even the percentage of 100%.

It has to be emphasized the possibility to identify a low dimensional space where not only there is a natural discrimination between safe and disruptive regions, but also that the different classes of disruptions tend to occupy different regions, that was absolutely not obvious so far.

The excellent results obtained with GTM motivate the deployment of this tool in the real time digital network (ATM) of JET. This is already planned for the next campaigns. Probably some retraining will be necessary to optimize the performance of the map for the new configuration of the machine, which has now a full metallic wall. Disruption prediction is out of the scope of the present paper. However, the potentiality of the available map suggests us the possibility to track the temporal sequence of the samples on the map, depicting the movement of the operating point during a discharge. Following the trajectory in the map, it will be possible to recognize an operational region where the risk of an imminent disruption is high.

With reference to more long term aspects, the systematic analysis of the disruption types performed with the manifold learning tool indicates quite clearly that the plasma can evolve through different states before the disruption time and that it is therefore a bit too reductive to associate a single disruption type to a disruptive discharge. It would be more appropriate to associate a series of different disruptive configurations to every disruptive discharge, so that its evolution towards the disruption time could be better analyzed. This would allow not only improving the classification capability of the manifold leaning tool but would also permit to use it to better study the physics of the disruptions.

ACKNOWLEDGEMENT

The authors would like to thank Peter de Vries for providing the manual classification of the disruptions and for the useful discussions. This work was supported by the Euratom Communities under the contract of Association between EURATOM/ENEA. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

REFERENCES

[1]. Cannas B., Fanni A., Zedda M.K., Sonato P. and JET EFDA Contributors A prediction tool for real-time application in the disruption protection system at JET, *Nuclear Fusion* **47** (2007) 1559–69.

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

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

[4]. Rattá G.A., Vega J., Murari A., Vagliasindi G., Johnson M.F., De Vries P.C. and JET EFDA Contributors An advanced disruption predictor for JET tested in a simulated real-time environment, *Nuclear Fusion* **50** (2010) 025005.

[5]. de Vries P.C., Johnson M.F., Alper B., Buratti P., Hender T.C., Koslowski H.R., Riccardo V., and JET-EFDA Contributors Survey of disruption causes at JET, *Nucl. Fusion* **51** (2011) 053018.

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

[7]. Lee J.A., Verleysen M. *Nonlinear Dimensionality Reduction* (Springer) 2007.

[8]. Jolliffe I.T., *Principal Component Analysis* Springer-Verlag 1986.

[9]. Kohonen MT *Self-Organization and Associative Memory* (Springer-Verlag, New York) 1989.

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

[11]. Camplani M., Cannas B., Fanni A., Pautasso G., Sias G., Sonato P. and the Asdex-Upgrade Team Tracking of the Plasma States in a Nuclear Fusion Device using SOMs, *Neural Computing and Applications* **20** (2011) 851-63.

[12]. Aledda R., Cannas B., Fanni A., Pautasso G., Sias G. and the ASDEX Upgrade Team Mapping of the ASDEX Upgrade Operational Space for Disruption Prediction, *IEEE Trans. on Plasma Science* **40** (2012) 570 – 576.

[13]. Murari A., Boutot P., Vega J., Gelfusa M., Moreno R., Verdoolaege G., de Vries P.C. and JET-EFDA Contributors Clustering based on the Geodesic Distance on Gaussian Manifolds for the Automatic Classification of Disruptions, *Nucl. Fusion* **53** (2013) doi:10.1088/0029-5515/53/3/033006.

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

[15]. Asimov D. The Grand Tour: a Tool for Viewing Multidimensional Data, *SIAM Journal on Scientific and Statistical Computing* **6** (1985) 128- 143.

[16]. Murari A., Vega J., Rattà G.A., Vagliasindi G., Johnson M.F., Hong S.H. and JET-EFDA Contributors Unbiased and non-supervised learning methods for disruption prediction at JET, *Nuclear Fusion* **49** (2009) 055028-39.

[17]. Rattá G.A., Vega J., Murari A., Johnson M.F. and JET EFDA Contributors Disruption Prediction at JET with a Combination of Exploratory Data Analysis and Supervised Method, Proc. Topical Conf. High Temperature Plasma Diagnostic (Albuquerque, New Mexico) 2008.

[18]. Zhang Y., Pautasso G., Kardaun O., Tardini G., Zhang X. D. and the ASDEX Upgrade Team

Prediction of ASDEX Upgrade disruptions using discriminant analysis, *Nuclear Fusion* **51** (2011) 063039-41.

[19]. Dunn J Well separated clusters and optimal fuzzy partitions, *J Cybern* (1974) 495–104.

[20]. Rapp J et al Density limits in TEXTOR-94 auxiliary heated discharges, *Nuclear Fusion* **39** (1999) 765.

[21]. Borrass K. et al Recent H-mode density limit studies at JET, *Nuclear Fusion* **44** (2004) 752.

[22]. Buratti P., Buttery R.J., Chapman I.T., Crisanti F., Gryaznevich M., Hender T.C., Howell D.F., Joffrin E., Hobirk J., Litaudon X., Mailloux J. and JET-EFDA Contributors MHD stability limit analysis in JET high β_N advanced scenarios, 36th EPS Conference on Plasma Phys. Sofia, June 29 - July 3, ECA **33**, O-2.007 (2009).

Classes	Label	TOT	TOT %
Auxiliary power shut-down	ASD	50	50.58
Too strong internal transport barrier (ITB)	ITB	10	4.12
Impurity control problem	IMC	83	34.16
Low density and low 'q'	LON	12	4.94
Neo-classical tearing mode	NTM	21	8.64
Density control problems	NC	58	23.87
Greenwald limit	GWL	9	3.70
TOT		**243**	**100**

Table I: Non intentional disruptions types as defined in [5].

JPF Signals	Acronym	Unit
Plasma Current	I_p	A
Poloidal Beta	β_p	*a.u.*
Mode Lock Amplitude	LM	T
Safety Factor at 95% of Poloidal Flux	q_{95}	*a.u.*
Total Input Power	P_{tot}	W
Plasma Internal Inductance	li	*a.u.*
Plasma Centroid Vertical Position	Z_{cc}	m
Line Integrated Plasma Density	ne_{lid}	m^{-2}
Stored Diamagnetic Energy Time Derivative	dW_{dia}/dt	W
Total Radiated Power	P_{rad}	W

Table II: Set of considered signals.

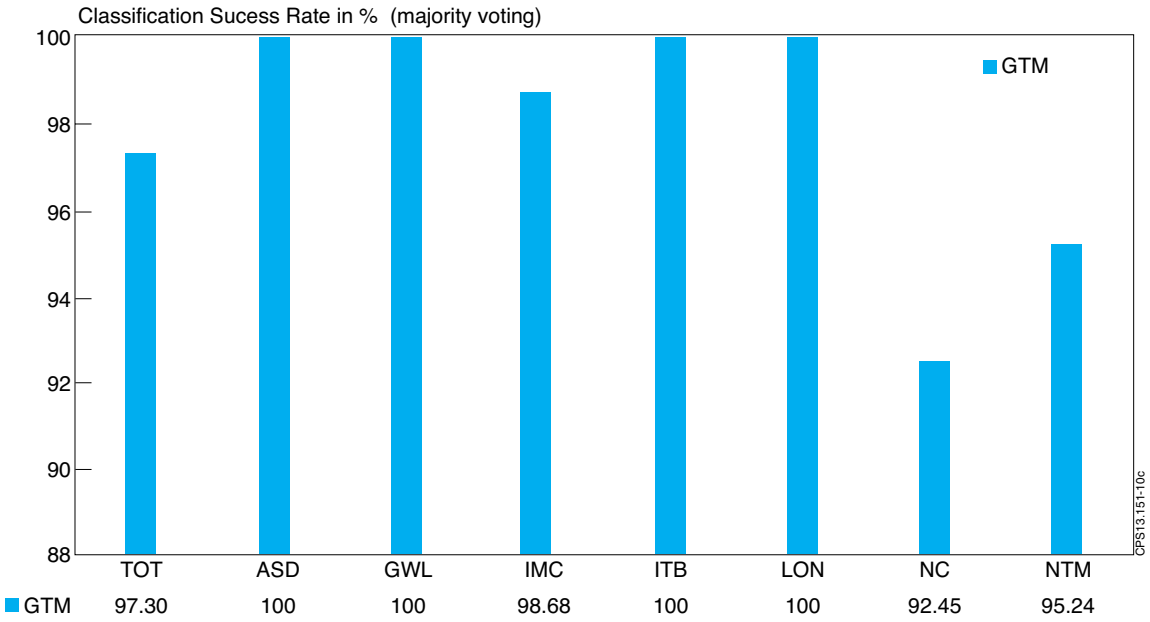

Table III: Comparison between the percentage success rates of the off-line automatic classification performed by GTM and SOM.

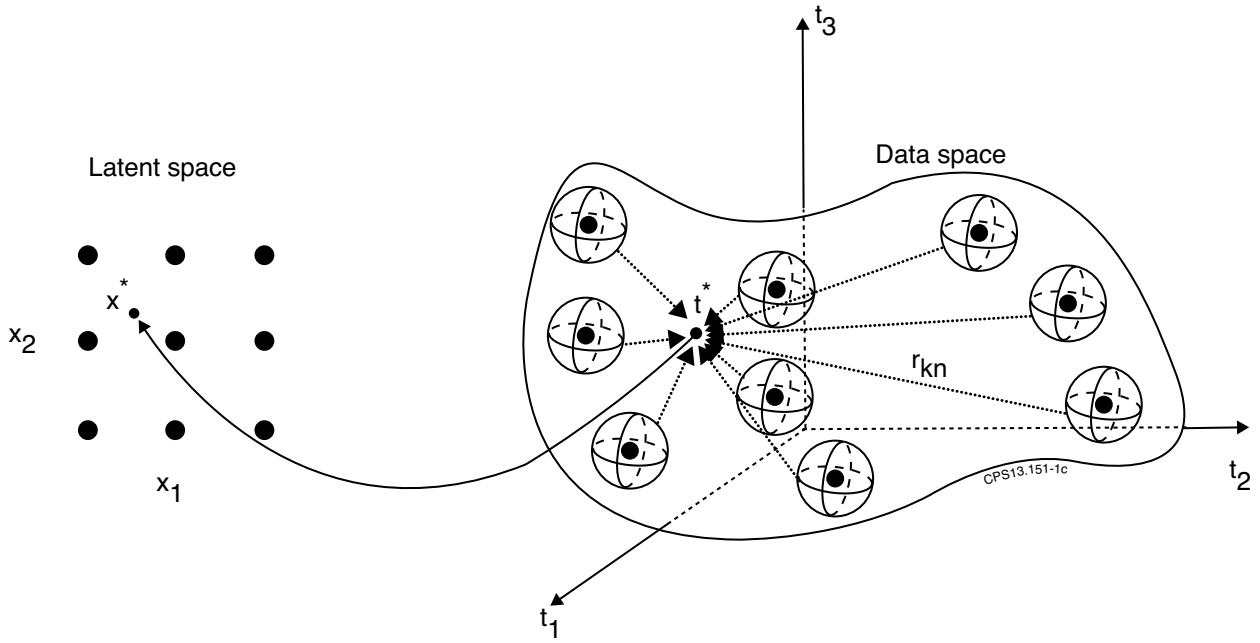

Figure 1: GTM mapping and manifold: each node located at a regular grid in the latent space is mapped to a corresponding point y(x;W) in the data space, and forms the centre of a corresponding Gaussian distribution. In the figure the correspondences between a data point in the manifold embedded in the data space and the mean of the posterior distribution in the latent space is also shown

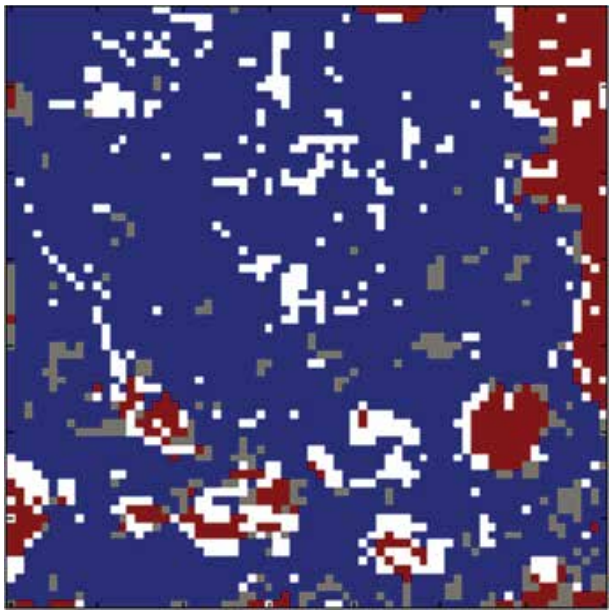

Figure 2: 2-D GTM of the 10-D JET operational space: safe clusters (blue), disruptive clusters (red), mixed clusters (grey), empty clusters (white).

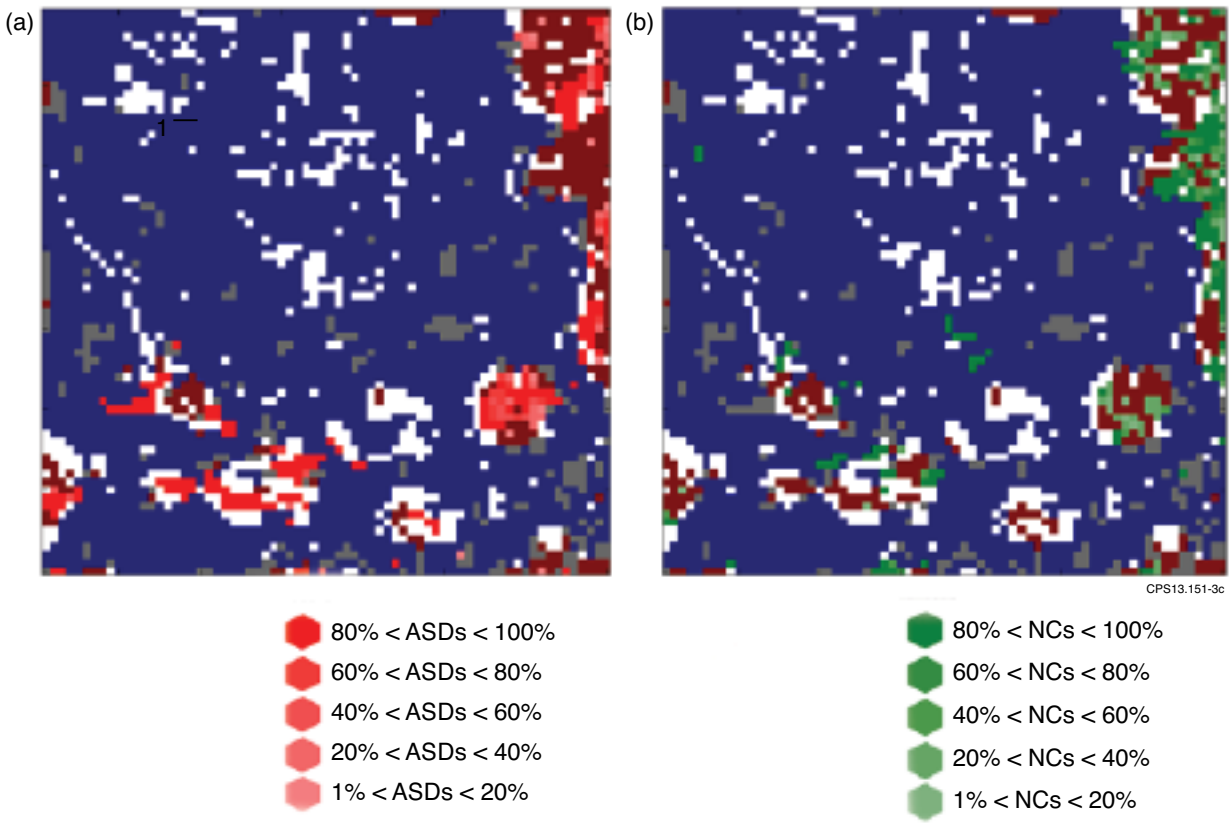

Figure 3: GTMscoloured depending on disruption class: (a) clusters marked by shades of red contain ASD samples; (b) clusters marked by shades of green contain NC samples.

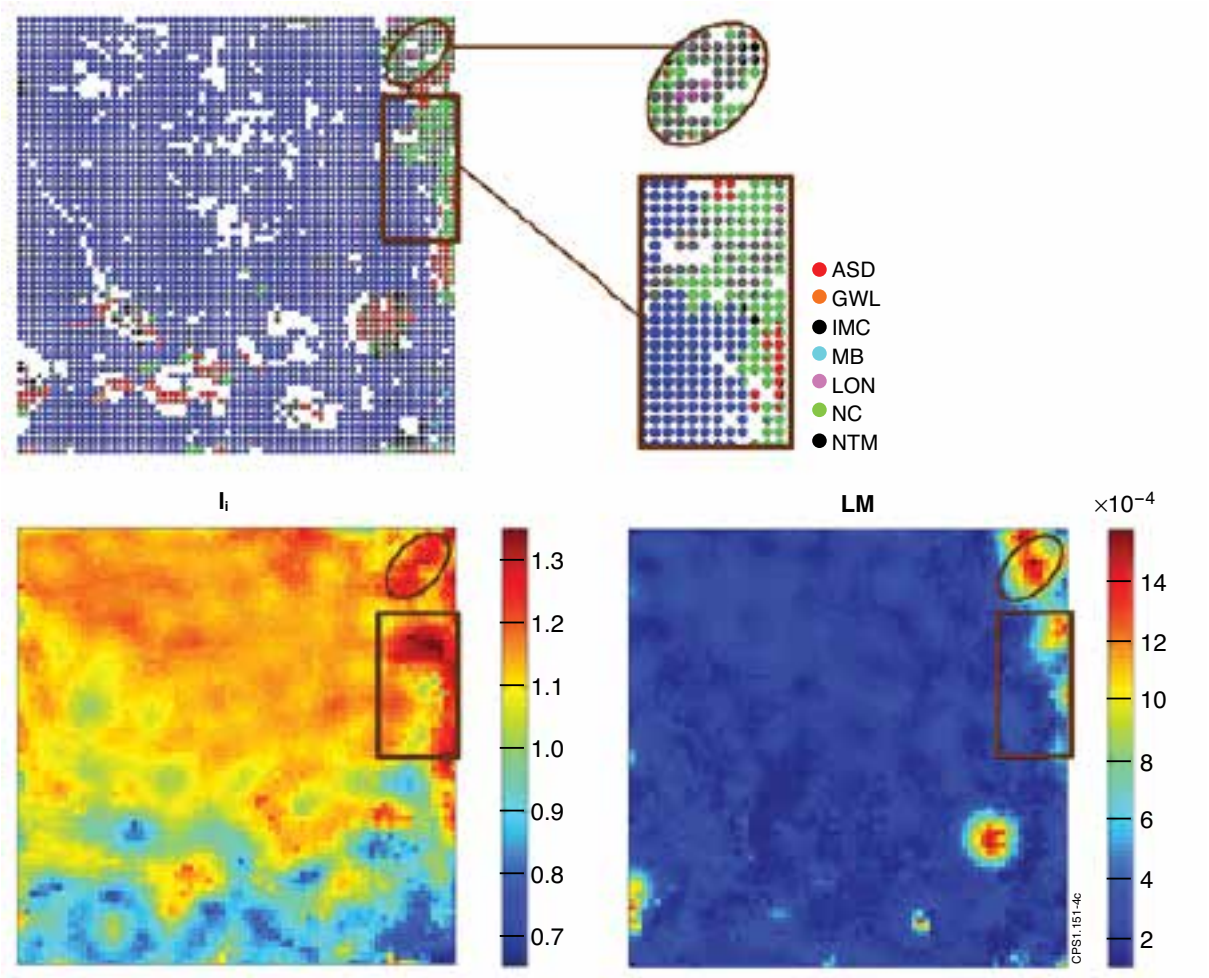

Figure 4: Top: GTM (left side) using a pie chart representation. Zoom of the regions in the boxes (right side). Bottom: GTM component planes of the internal inductance (left side) and the locked mode (right side).

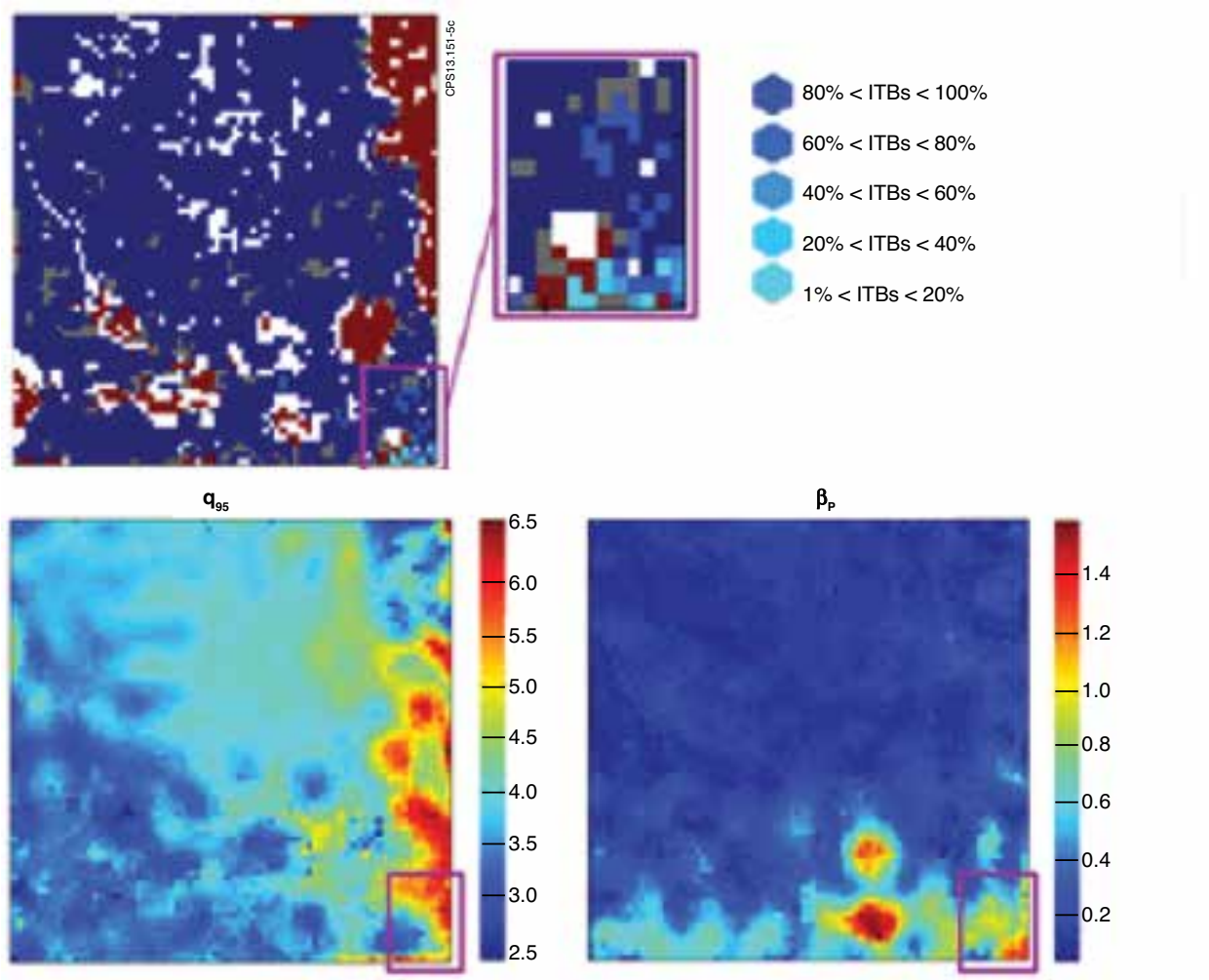

Figure 5: Analysis of the cluster composition for ITB disruptions and Component Planes of poloidal beta and safety factor in the GTM.

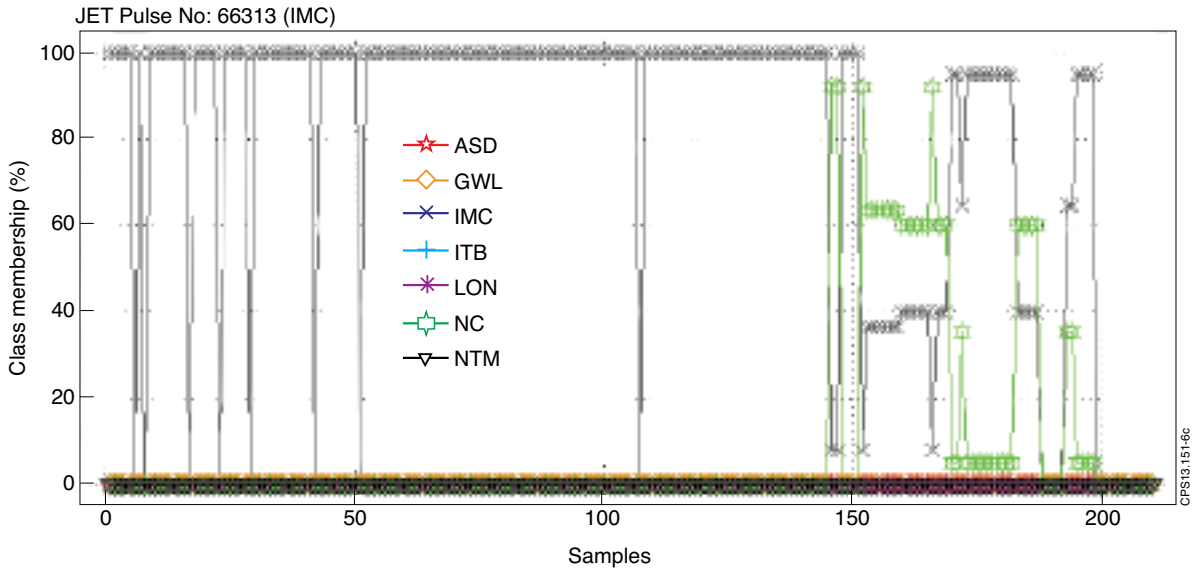

Figure 6: Class membership functions for Pulse No: 66313.

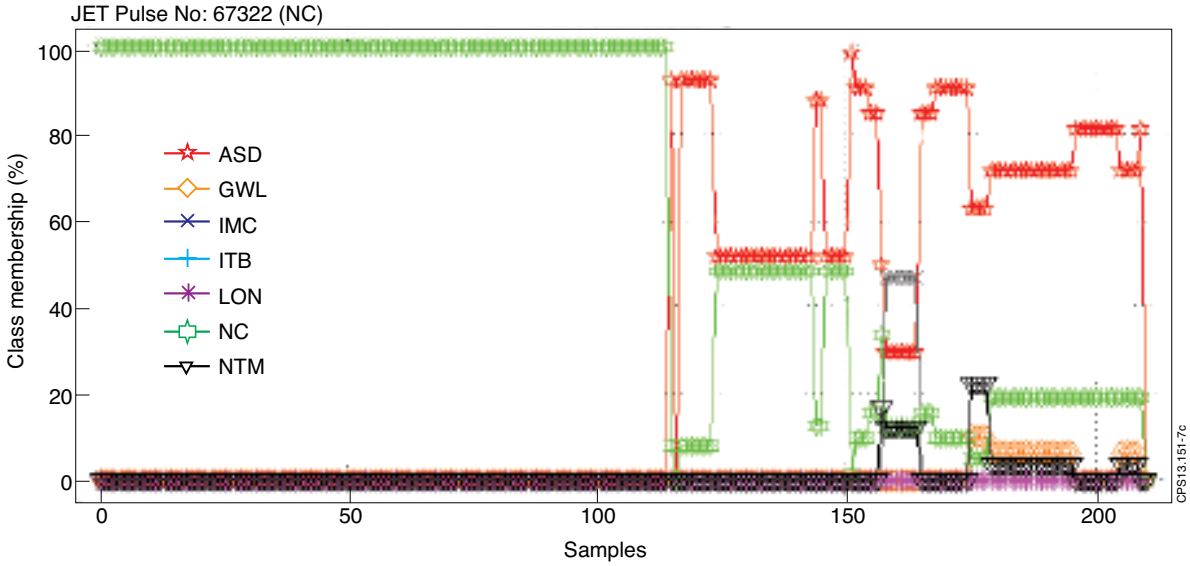

Figure 7: Class membership functions for Pulse No: 67322.

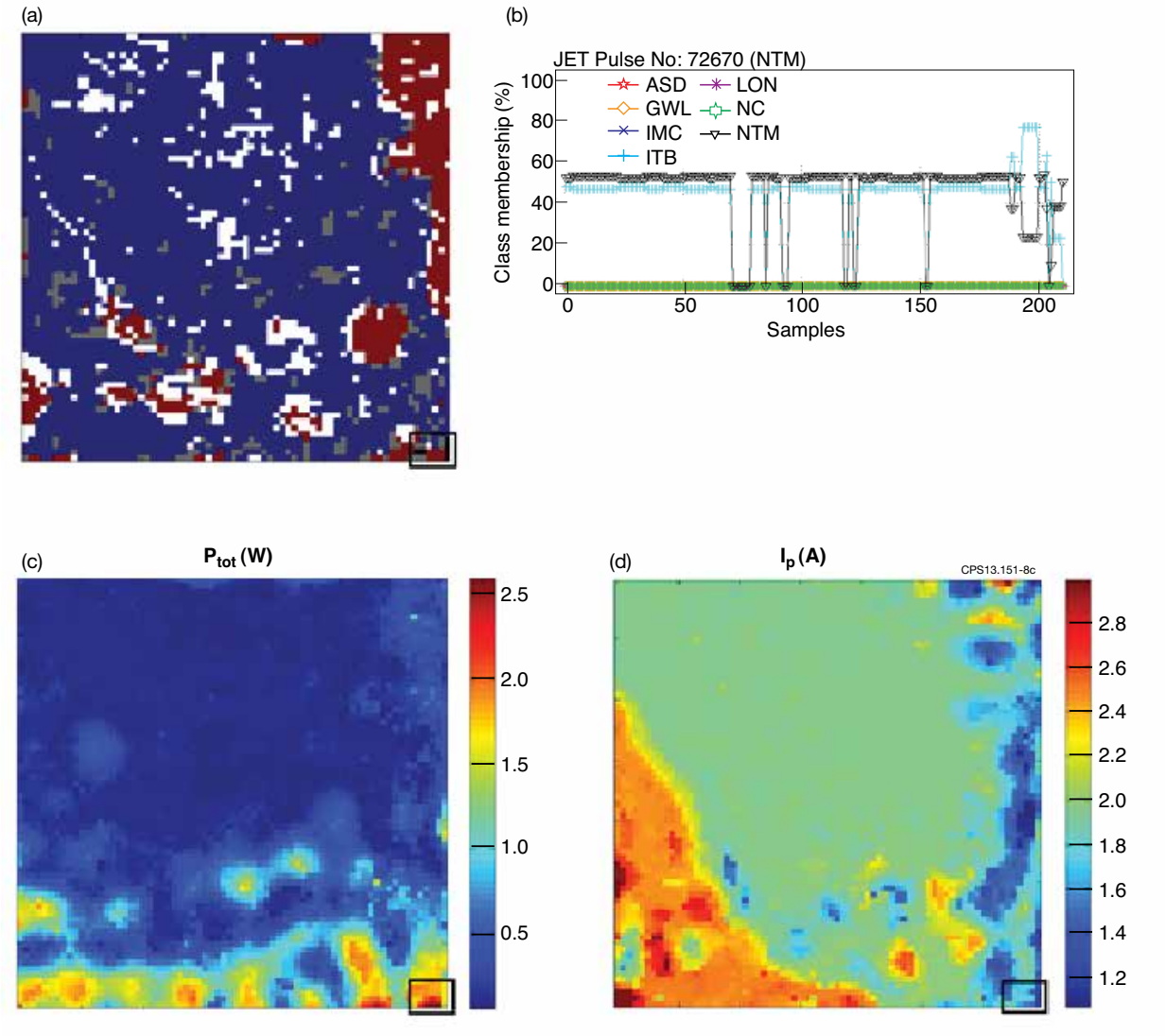

Figure 8 : (a) clusters (black) occupied in the GTM by the Pulse No: 72670; (b) Class membership functions. (c-d): component planes of the total input power (left) and the plasma current (right).

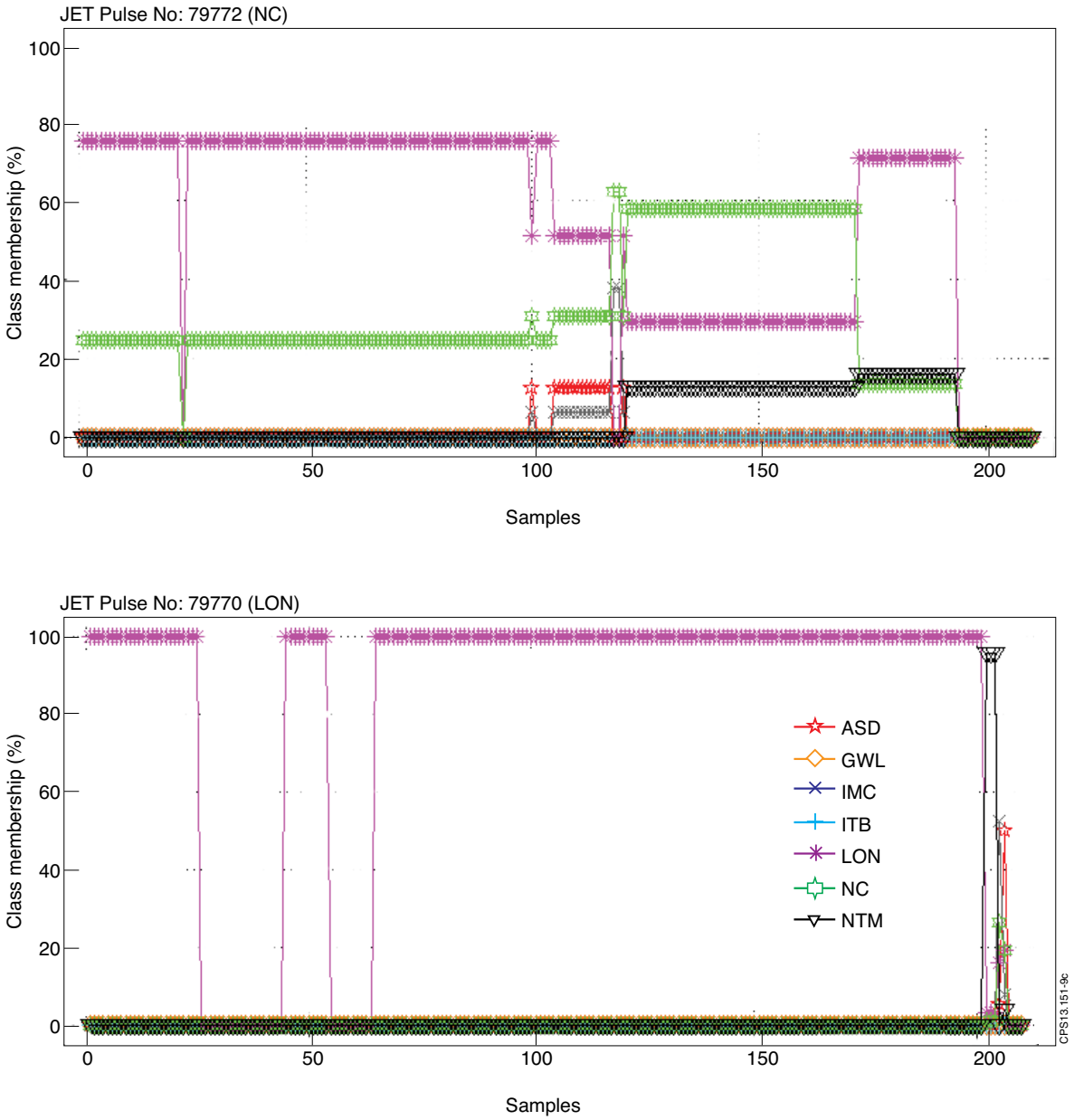

Figure 9: Class membership functions for Pulse No's: 9772(a) and 79770 (b)