EUROFUSION WPJET1-PR(16) 14855

J Vega et al.

Real-time anomaly detection for disruption prediction: the JET case

Preprint of Paper to be submitted for publication in
Nuclear Fusion

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Real-time anomaly detection for disruption prediction: the JET case

J. Vega[1], A. Murari[2], S. Dormido-Canto[3], R. Moreno[1], A. Pereira[1], S. Esquembri[4]

and JET Contributors*

EUROfusion Consortium, JET, Culham Science Centre, Abingdon, OX14 3DB, UK

[1]Laboratorio Nacional de Fusión, CIEMAT, Madrid, Spain

[2]Consorzio RFX (CNR, ENEA, INFN, Universitá di Padova, Acciaierie Venete SpA), Corso Stati Uniti 4, 35127 Padova, Italy

[3]Dpto. Informática y Automática - UNED, Madrid, Spain

[4]Technical University of Madrid (UPM), Madrid, Spain

*See the Appendix of F. Romanelli et al., Proceedings of the 25th IAEA Fusion Energy Conference 2014, Saint Petersburg, Russia

Abstract

This article shows the development of a new kind of real-time disruption predictor that is based on detecting anomalies in the data flow. The new predictor neither depends on data from past discharges nor is based on signal amplitude thresholds. In JET, using only the locked mode signal, the new predictor shows results comparable to the JET APODIS predictor but without the need of a training process with past data. The predictor has been tested with JET discharges in the range 82460-87918. This range corresponds to all ITER-like Wall experimental campaigns (2011 – 2014). The discharge dataset consists of 1738 non-disruptive discharges and all unintentional disruptions (566 disruptive shots). The results show 8.98% of false alarms, 10.60% of missed alarms, 3.18% of tardy detections, 83.57% of valid alarms, 2.65% of premature alarms and average anticipation time of 389 ms. These rates are compared in the article with the results of the JET APODIS predictor and the JET disruption predictor based on crossing a threshold of the locked mode signal amplitude.

1. Introduction

Disruptions are unavoidable events in present tokamaks. To counteract their detrimental effects, mitigation techniques can be envisaged to reduce forces, to alleviate heat loads during the thermal quench and to avoid runaway electrons. Examples of mitigation methods can be the injection of a significant amount of gases through fast

valves [1, 2], killer pellets [3, 4] or Electron Cyclotron Resonance Heating injection [5, 6]. However, disruption mitigation techniques depend crucially on accurate disruption predictors, whose alarms have to be triggered with enough anticipation time.

There are two key expressions in the above paragraph: '*accurate disruption predictors*' and '*enough anticipation time*'. The former has to be understood in terms of high success rates (> 95% is required for ITER) and low false alarm rates (<5 % for ITER). '*Enough anticipation time*' stresses the fact that alarms are useful only if the anticipation time is greater than the characteristic time of the available mitigation tool. This characteristic time is the sum of the time needed to fire technical systems after an alarm plus the plasma response time to the mitigation action. For the sake of clarity, the 'plasma response time' has to be understood as the delay between the time in which the plasma starts to be 'perturbed' by a mitigation method and the time in which the mitigation begins to be effective.

In all tokamaks, a common signal used for disruption prediction is the locked mode (LM) amplitude. When macroscopic instabilities start locking to the wall, the LM amplitude grows during the slowing down of the plasma rotation (the particular implementation of LM measurements in JET can be found in http://users.euro-fusion.org/pages/mags/equilibrium/eq-coil-loop/saddle-loop/saddle-loop.htm). Usually, during a running experiment, if the LM amplitude crosses a certain threshold, which is set-up prior to the discharges, an alarm is triggered. This threshold is selected in a manual way and it is chosen depending on the characteristics of the experimental programme: the LM threshold is set lower or higher depending on the potential danger of the possible disruptions.

Predictors based on general machine learning methods (Support Vector Machines (SVM) [7], fuzzy logic (FL) [8] or Artificial Neural Networks (ANN) [9, 10]) provide a more intelligent way of recognizing a forthcoming disruption without the need of manual selection of thresholds. The price for this is to carry out a generally very expensive (in computational terms) training process. This training process splits the multi-dimensional operational space into two zones (disruptive and non-disruptive) and determines the separation frontier between them (fig. 1). In a running discharge, the plasma state is determined on a periodic basis and is represented by a point in the operational space. When the point is contained in the non-disruptive zone, the plasma is considered to be in a safe mode. However, when the point appears in the disruptive zone, an alarm has to be triggered.

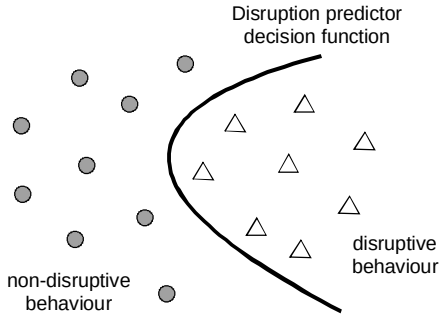

Fig. 1: A disruption predictor is implemented as an automatic classification system. During a discharge, data are sent to the classifier on a periodic basis with the objective of determining in which part of the separation frontier the plasma is located.

The experience in the development of disruption predictors based on machine learning shows that the choice of a specific algorithm (for example, SVM, FL or ANN) is less important than the selection of a set of signal features to properly represent the operational space. Typically, the operational space is represented by an M-dimensional feature space that is partially mapped by the N samples (or feature vectors) provided in the training process: $\mathbf{x}_j \in \mathbb{R}^M$, $j = 1,...,N$. It is clear that the greater N the more knowledge about the space can be learnt. However, only a large N does not guarantee a good knowledge. The set of N samples has to cover the operational space as much as possible to reduce the unexplored zones to a minimum.

The M components of the feature space have to be of distinctive nature to distinguish between disruptive and non-disruptive behaviours. These features are usually amplitudes of plasma quantities. However, it has been shown [11, 12] that the inclusion of information connected with the frequency domain of the signals improves drastically the predictions. In general, each individual feature is a single value that can be related to either a plasma quantity, or a control signal or the result of a (perhaps complex) data processing (with either physics or control quantities).

As mentioned, the outcome of the training process is the separation frontier (or decision function) between disruptive and non-disruptive behaviours. This frontier is expressed in form of a typically complex equation that relates (in a highly non-linear way) the components of the operational space. Due to, on the one hand, the complex form of the equation and, on the other hand, the potential complexity of the data processing to define the individual features, this equation usually is not valid for a direct interpretation of physics behaviours. However, during the evolution of the discharges, it is extremely useful for the real-time recognition of forthcoming disruptions (although the interpretation of the physical reasons why the predictions are correct remains problematic).

In JET, the Advanced Predictor Of DISruptions (APODIS) [13] is a machine learning system that outperforms the prediction capability of the Locked Mode Predictor based on a Threshold criterion (LMPT). During the JET ITER-like wall (ILW) campaigns (September 2011 – October 2014), the APODIS success rate and average warning time are about 82% and 274 ms respectively. The equivalent values for the LMPT are 67% and 255 ms. The reason for these different results is the exhaustive training process carried out for APODIS instead of the manual selection of thresholds that is used with the LMPT. It should be noted that APODIS was trained with almost 10,000 JET discharges with C wall between April 2007 and October 2009, where more than 900 h of CPU time in a high performance computer were needed. APODIS is in operation with the JET ILW and, so far, no re-training has been necessary. But

collecting these huge training datasets is not a realistic strategy in the next generation of experiments such as ITER or DEMO.

Trying to avoid the use of huge amount of discharges in the training process, a recent alternative has been the development of disruption predictors from scratch [14, 15]. Disruption predictors from scratch are high learning rate predictors whose learning process starts after the first disruption. They are retrained after each missed alarm by adding disruptive and non-disruptive examples to the existing training dataset. In JET with the ILW, these adaptive predictors show success rates about 83% and average warning times of 244 ms [15].

It should be emphasized that the '*learning from scratch*' approach is not only relevant for the start of a new device (like ITER or DEMO) but also in the case of existing devices, when they experience significant changes in the operational space. Examples of this can be JET from the carbon wall to the ITER-like wall operations or from DD to DT operations. Such changes may cause a drop in the success rate of more traditional predictors but, with continuous learning, systems based on '*learning from scratch*' can recover fairly quickly.

A more advanced option for disruption prediction would be the use of intelligent predictors that start their learning process with each new discharge and without the need for previous information from past discharges. The objective is to learn how a safe plasma evolution is and to trigger an alarm when an anomalous behaviour appears. Of course, the anomaly has to be closely related to a disruption. This is a novel approach for disruption prediction [16, 17] that completely avoids, firstly, the requirements of having a large database of disruptive and non-disruptive discharges (ITER or DEMO cannot wait for hundreds of disruptions to have a reliable predictor) and, secondly, carrying out very expensive computational trainings to determine good enough data-driven models.

However, it is important to point out that the use of previous knowledge is also possible with this type of predictors. There can be additional information gained from past experience that can be used to trigger the alarms at proper times. In this article, this kind of information is called '*privileged knowledge*' (PK). The PK can be obtained from theoretical models, data-driven models, simple inspection of past discharges or whatever sources of information related to the specific implementation of the anomalous behaviour recognition. It should be noted that PK can be either available or not and if so, it can be used or not. In other words, the concept of predictors based on anomaly detection, which is presented in this article, means that these predictors have to predict just with the information acquired during each running discharge regardless of the existence of privileged knowledge. The PK has to be understood as an additional help to improve the predictions performed by the underlying methodology for detecting anomalies.

This article describes a disruption predictor for JET that is based on a locked mode signal that does not need previous shots for training purposes. The predictor is

based on the automatic recognition of changes (anomaly detections) in data streams through the identification of outliers in the data flow. Due to this reason, the predictor is called Single signal Predictor based on Anomaly Detection (SPAD). In SPAD, the locked mode samples are processed in time windows 32 ms long and the outputs of this processing form the data flow to be sequentially analysed. Near a disruption, the data generating model changes as the data are streamed and this change is detected and used to trigger an alarm.

Section 2 explains the on-line data stream setting used for SPAD. Section 3 presents a conceptual view of SPAD. Section 4 introduces the locked mode signal as the fundamental quantity for the present SPAD version. Section 5 is devoted to describing the data processing and section 6 discusses the results in JET.

2. On-line learning during discharge production

In general, data stream applications have to deal with the changing characteristics of the data. As most decision-making tasks rely on the timeliness and relevance of their supporting data, the changing nature of the data creates tremendous challenges for many learning algorithms and data mining techniques. On the other hand, changes in the data may convey interesting time-dependent information and knowledge. Problems driven by the changing characteristics of data include change detection techniques, quantification of changes and the ability to build accurate models for the changing data. Examples of applications for change detection include network monitoring, video surveillance, Internet security and anomaly/fraud/intrusion detection.

Typically, in an on-line data streaming setting, data are observed sequentially and a decision on the identification of any kind of change in the data has to be made '*on-the-fly*'. Focusing the attention on disruption prediction, an anomaly detector system has to learn in each new discharge the evolution of a safe behaviour and to fire an alarm when changes in the data stream are detected. It is important to note that the production of discharges under different scenarios can generate different classes of non-disruptive plasmas. Therefore, in principle, the safe evolutions have to be learnt in every discharge.

An essential point in the application of anomaly detection to recognize a forthcoming disruption is to be sure that the change in the data corresponds to the identification of a disruptive event. Otherwise, lots of false alarms would be triggered and the development of interesting plasma scenarios would be very cumbersome.

Also, it is necessary to mention the requirements to be met by a disruption predictor based on anomaly detection. First, it is important to note that the sequential data are read only once. Second, the delay between a true alarm and its detection has to be minimal. Third, it should be noted that the number of both missed alarms and false alarms must be minimal. Last but not least, data streams should be handled efficiently

from a computational point of view, which is crucial for the real-time implementation of the predictor.

3. SPAD predictor: conceptual view

The detection of anomalies can be implemented by detecting outliers in a data stream. From a conceptual point of view, the detection of anomalies can be carried out through the recognition of changes in the data flow. Fig. 2 shows the conceptual design of the anomaly detection used in this article. Let's assume the real-time acquisition of input samples is on a periodic basis with period Δt. From time t_0, there are *n* samples that follow the same data distribution (unknown, but the same one). From $t_0 + n \cdot \Delta t$, the data distribution changes. It should be noted that the distribution is again unknown, but it is different from the previous one. The objective is to determine as soon as possible when the change has been produced.

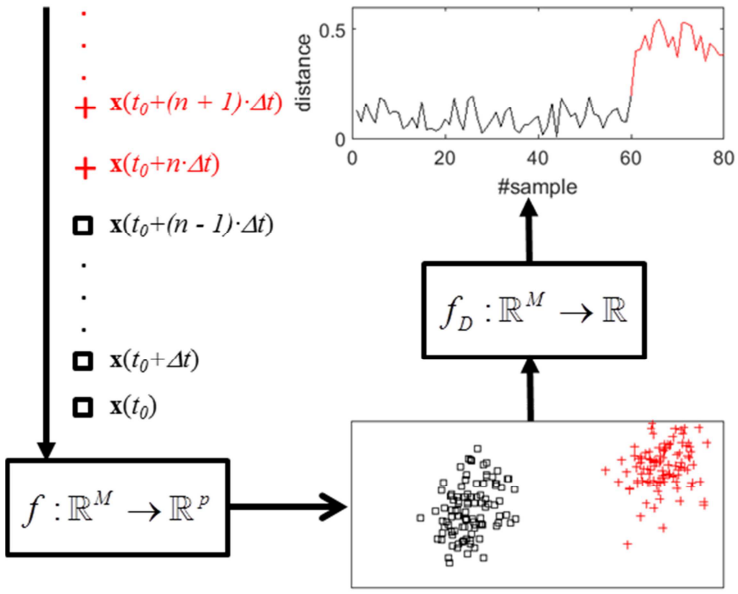

Fig. 2: The objective of the anomaly detector is to group all input samples into a single cluster through a mapping function *f*, which has to be determined for every particular implementation of the detector. Only when a second cluster appears, an alarm has to be triggered. The alarm is triggered when an outlier is identified in the temporal evolution of the distance between each sample and the centroid of all the previous ones.

In general, the samples $\mathbf{x}(t)$ can be multi-dimensional feature vectors, *i.e.* $\mathbf{x}(t) \in \mathbb{R}^M$, where *M* is the number of vector components (or features). The anomaly detector maps the input samples into an output space $\mathbb{R}^p$ in such a way that the mapping function $f : \mathbb{R}^M \rightarrow \mathbb{R}^p$ is able to group the output samples $f(\mathbf{x}(t)) \in \mathbb{R}^p$ into two separated clusters (black and red clusters in fig. 2). The first cluster is made up of

output samples $f\left(\mathbf{x}(t)\right), t \in \left[t_0, t_0 + (n-1)\cdot\Delta t\right]$ and the second cluster contains output samples $f\left(\mathbf{x}(t)\right), t \geq t_0 + n\cdot\Delta t$. The detection process has to recognise when the output samples start to be *far* from the initial cluster and then, an alarm is triggered.

On the one hand, it is important to point out that, typically, there can be a high number of mapping functions and, therefore, multiple alternatives can be tested. On the other hand, the recognition of output samples *far* from the initial cluster implies the use of *distances* in the mathematical sense. In other words, the output samples $f\left(\mathbf{x}(t)\right)$ have to be mapped according to $f_D : \mathbb{R}^p \rightarrow \mathbb{R}$, where f_D is a distance. In particular, the objective of SPAD is to follow the temporal evolution of the distance between each new output sample and the centroid defined by all previous output samples. An anomaly is recognised when an outlier appears, *i.e.* after a sudden rise of the distance (see sample number 60 in the top plot of fig. 2). Of course, any SPAD implementation has to recognise the anomaly in real-time.

4. SPAD predictor and the locked mode

There are two main factors in the development of the SPAD predictor: the disruption recognition criterion and the temporal resolution of the predictions. The first one is the key of the method to maintain low rates of *false alarms*, low rates of *missed alarms*, low rates of *premature alarms*, low rates of *tardy detections* and high rates of *valid alarms*. For the sake of clarity, definitions of the previous concepts for JET were provided in [15] and are included in appendix 1.

With regard to the temporal resolution of the predictions, triggering an alarm as soon as possible plays a central role in order to achieve the largest possible warning time.

As mentioned in section 2, it is extremely important that the anomaly detected in the data stream really corresponds to an incoming disruption. A single quantity closely related to disruptions is the locked mode signal. Therefore, in a first approach, this signal will be the only one used for the implementation of a disruption predictor based on anomaly detection.

It should be noted that SPAD has to be different and to perform better than a standard LMPT, which is based on triggering an alarm when the locked mode amplitude just crosses a threshold. The experience on JET, in the development of disruption predictors, reveals the importance of using not only features in the time domain but also in the frequency domain [13, 14, 15]. The inclusion of features related to the frequency domain implies that the basic element to extract features (in this case from the locked mode signal) has to be a time window (characterized by its temporal length) with a number of digitized samples (characterized by the sampling rate). Due to the good results of APODIS with time windows of 32 ms and sampling rates of 1 kSamples/s,

SPAD will use the same parameters. However, the resolution of 32 ms to make predictions can be improved. To avoid the use of higher sampling rates, a sliding window mechanism has been implemented [18]. In the SPAD case, the latest 32 samples of the locked mode signal are processed every 2 ms (fig. 3).

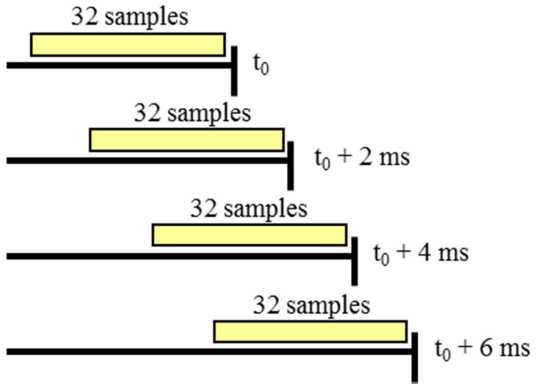

Fig. 3: Sliding window mechanism to achieve a prediction resolution of 2 ms without increasing the sampling rate (1 kSamples/s).

5. SPAD data processing

From section 3 and fig. 2, the steps for the SPAD implementation are clear. Firstly, data collected in time windows 32 ms long have to be processed in such a way that during a non-disruptive evolution of a discharge, the output samples of the mapping $f:\mathbb{R}^{32}\rightarrow\mathbb{R}^{p}$ form a single cluster. In the presence of an incoming disruption, the output samples should appear far from the cluster.

A possible mapping is to compute the wavelet transform [19] of the 32 samples of every time window. The choice of the wavelet transform provides many advantages: data compression, computing efficiency and simultaneous time and frequency representation. In particular, the Haar wavelet family has been selected.

Wavelet transforms process data at different resolutions or decomposition levels. In each decomposition level, two sets of coefficients are obtained: approximation coefficients and detail coefficients. The approximation coefficients are a low pass filter of the signal whereas the detail coefficients are a high pass filter (fig. 4).

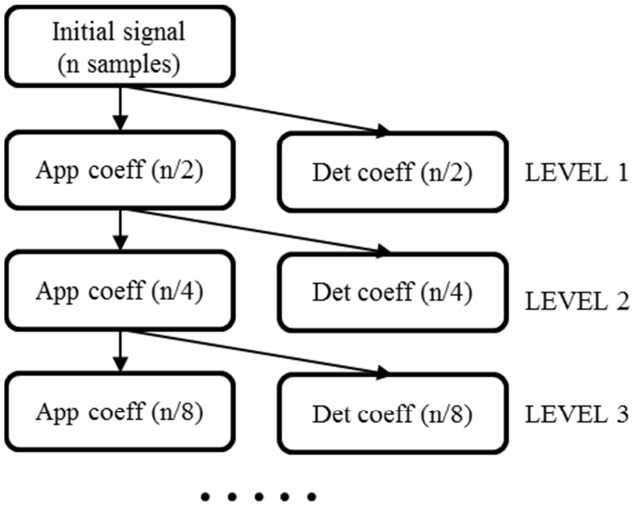

Fig. 4: Schematic view of the Haar wavelet transform. Each decomposition level reduces by half the number of coefficients.

Taking into account the 32 samples of the locked mode signal in each time window, let's compute the approximation coefficients of a Haar wavelet transform at level 4. This means a mapping $f:\mathbb{R}^{32}\to\mathbb{R}^{2}$, or in other words, to compress the information into 2 points in each time window. Therefore, every 2 ms, the latest 32 samples are processed with the Haar wavelet transform and the output samples are vectors of dimension 2, where the two components are just the approximation coefficients of the transform. Fig. 5a shows the scatterplot of these output samples in a JET non-disruptive discharge. It should be noted that in this bi-dimensional space, the points remain together in a single cluster. Fig. 5b shows the scatterplot corresponding to a disruptive discharge. In this particular discharge, the latest 68 points of the scatterplot before the disruption, which correspond to a time of 136 ms, are clearly outside of the initial cluster. When this starts to happen, an alarm has to be activated. It should be noted that the warning time for this discharge is 136 ms.

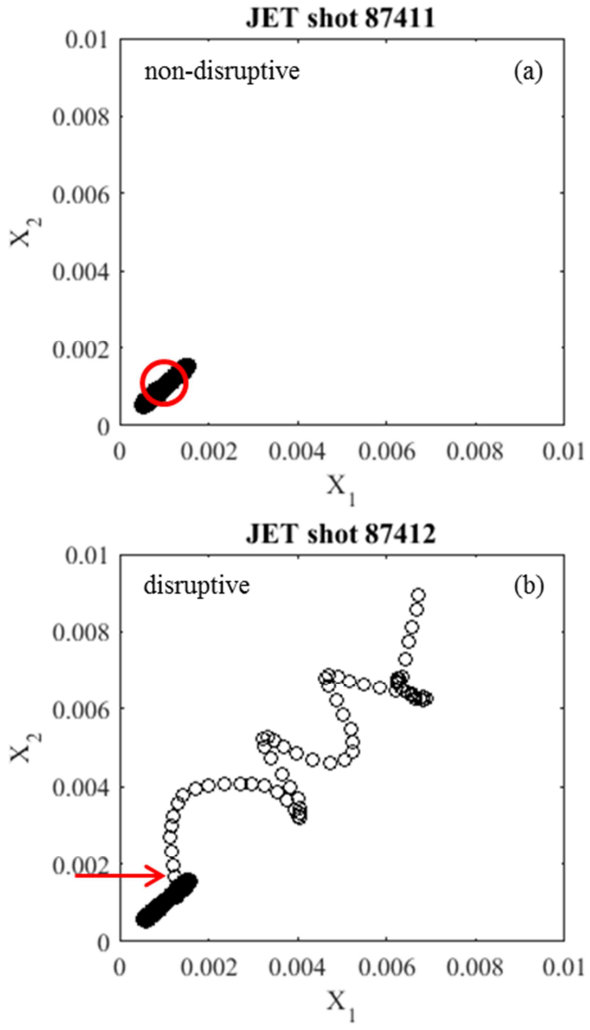

Fig. 5: Points are plotted every 2 ms and a compact cluster structure is observed in non-disruptive phases of the discharges. This structure is broken in the proximity of a disruption (the first outlier is marked by a red arrow in fig. 5b). The Euclidean distance cannot be used to detect outliers: lots of false alarms would appear as the Euclidean iso-distance contours are circles (Fig. 5a).

Fig. 6 shows the plasma current and the locked mode signals of the JET discharges that are shown in fig. 5. These shots will be used along this article as examples of the SPAD data processing.

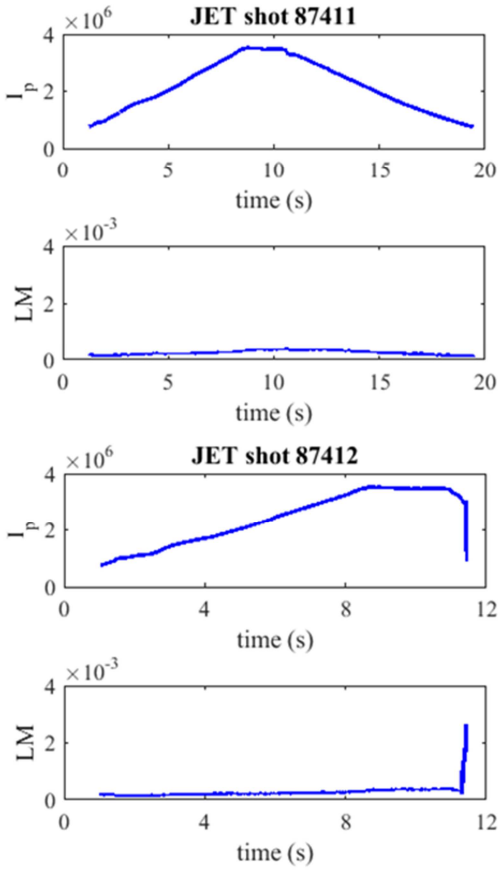

Fig. 6: Temporal evolution of the plasma current and the locked mode signals of JET discharges 87411 (non-disruptive) and 87412 (disruptive).

Coming back to fig. 5, it should be noted that the alarm must be triggered the first time that a point in the bi-dimensional space X_1-X_2 is *far enough* from the cluster centre. The Euclidean distance does not work to detect this condition because the Euclidean iso-distance contours are circles and, therefore, this distance does not take into account the pattern of positive covariance that is shown in the scatterplots (figs. 5a and 5b). However, the Mahalanobis distance [20] does adjust for covariance (fig. 7) and it can be used to recognise anomalous points, *i.e.* points clearly outside from the compact cluster. Therefore, the identification of outliers requires following the temporal evolution of the Mahalanobis distance between each new point and the centroid of all previous points. When this distance is recognised as an outlier, it means that a disruption is approaching and an alarm must be triggered.

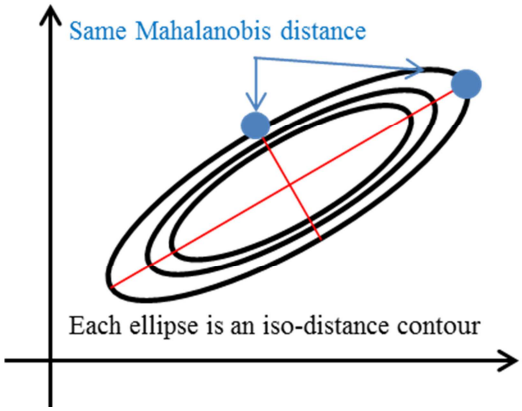

Fig. 7: The Mahalanobis distance takes into account the data covariance. The Mahalanobis iso-distance contours are ellipses and, therefore, the two blue points have the same distance to the centroid.

It should be noted that the process of recognising outliers in the Mahalanobis distances can be carried out in every discharge without the need of previous information from past discharges, which is the objective of this type of predictor. In the particular implementation for JET, when the plasma current exceeds the threshold of 750 kA in each discharge, the next 100 ms are used to compute the Haar wavelet transform of the locked mode signal every 2 ms with the latest 32 samples. In this way, an initial compact cluster of 50 points is formed. After the formation of this initial cluster and with a period of 2 ms, the latest 32 samples of the locked mode signal are analysed with the Haar wavelet transform (level 4 of decomposition). After obtaining each new bi-dimensional point, the Mahalanobis distance between it and the cluster centroid defined by all bi-dimensional points is computed.

Fig. 8 shows the temporal evolution of the Mahalanobis distance with a time resolution of 2 ms in both a non-disruptive discharge (fig. 8a) and a disruptive one (fig. 8b). First of all, it is important to note the different scales in both discharges. Also, it should be emphasized that the different scales are not related to the disruptive or non-disruptive character of the discharges. This means that just the Mahalanobis distance of any point to the cluster centroid is not able to distinguish anomalies in the data flow. The anomaly has to be recognised when the Mahalanobis distance is an outlier in relation to the distances of the previous points. Therefore, it is necessary to define an outlier factor, K_M, that identifies outlier distances. To this end, the criterion to identify outliers at a time t_P is related to Mahalanobis distances greater than or equal to K_M standard deviations from the baseline model of each discharge:

$$\left| \frac{D_{Mahalanobis}(t_P) - mean\left(D_{Mahalanobis}(t \le t_P)\right)}{std\left(D_{Mahalanobis}(t \le t_P)\right)} \right| \ge K_M . \qquad (1)$$

In the present version of SPAD, privileged knowledge (as explained in the introduction) has been used. About 20 JET disruptions were analysed to look for a proper value for K_M. From this analysis, K_M has been set empirically to a value of 10.

This value allows achieving good success rates simultaneously with low false alarm rates. If $K_M < 10$, the false alarm rate increases and if $K_M > 10$, both tardy detections and missed alarms increase.

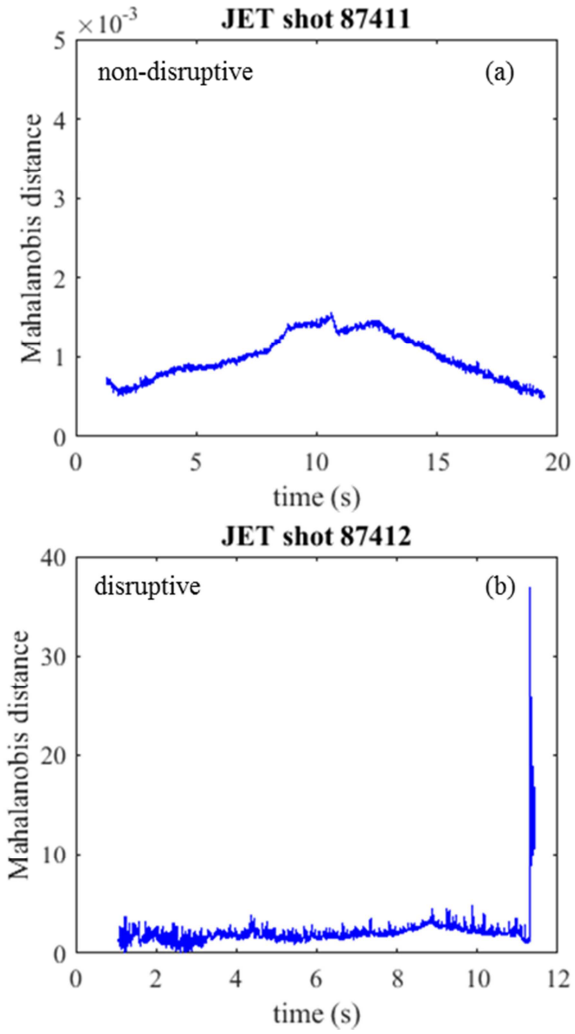

Fig. 8: The Mahalanobis distance is not enough to discriminate anomalous behaviours from regular plasma evolution. An outlier factor, eq (1), to recognise outlier distances is necessary.

Fig. 9 shows the temporal evolution of the outlier factor corresponding to the discharges of fig. 8. Fig. 9a shows an outlier factor that remains below the threshold $K_M = 10$ during the whole discharge and, therefore, no alarm is triggered. However, fig. 9b plots the temporal evolution of the outlier factor that triggers an alarm 136 ms before the disruption.

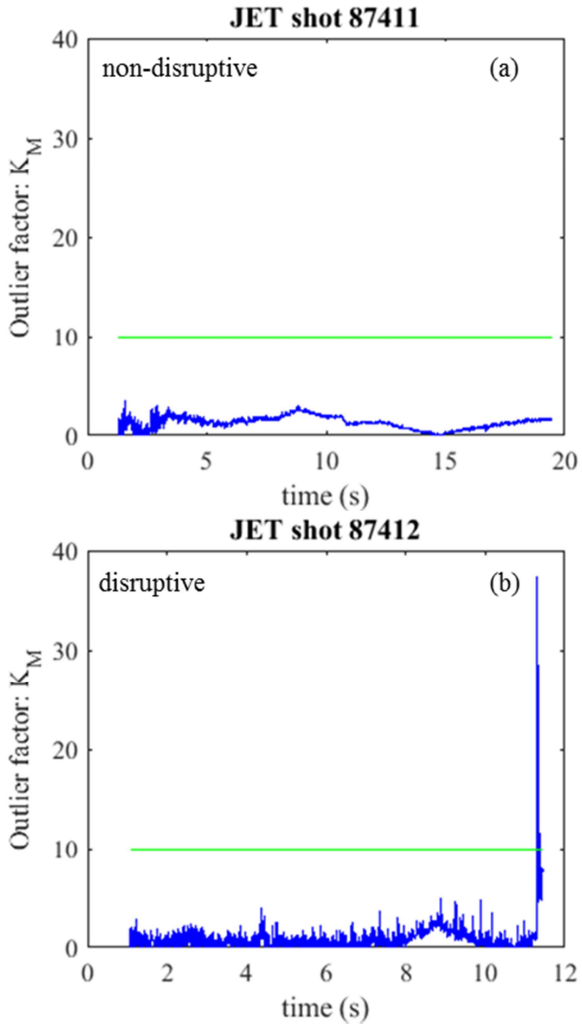

Fig. 9: Temporal evolution of the outlier factor in two different discharges.

6. SPAD results

Previous section describes the SPAD data processing for a particular mapping $f:\mathbb{R}^{32}\to\mathbb{R}^{2}$ defined by the Haar wavelet transform at the level 4 of decomposition. However, the same reasoning is valid for decomposition levels 1, 2 or 3, which correspond to mappings $f:\mathbb{R}^{32}\to\mathbb{R}^{16}$, $f:\mathbb{R}^{32}\to\mathbb{R}^{8}$ and $f:\mathbb{R}^{32}\to\mathbb{R}^{4}$ respectively. In these cases, the compact clusters are made up of output samples with dimensions 16, 8 and 4 respectively. In all cases, the Mahalanobis distance, as explained in section 5, has been used to recognise when an output sample is outside the compact cluster.

All SPAD results in this section use only the locked mode signal for disruption prediction purposes and all executions have been performed by simulating real-time conditions. SPAD has been applied to a JET database with all unintentional disruptions (566 discharges) and all non-disruptive discharges (1738) in the range 82460-87918 (ILW experimental campaigns between 2011 and 2014). In all cases, $K_M=10$.

Table 1 summarizes the prediction of disruptions by detecting anomalies in the Mahalanobis distance evolution for mappings $f:\mathbb{R}^{32}\to\mathbb{R}^{p}, 2\le p\le 16$ based on the Haar wavelet transform (approximation coefficients).

Table 1: Base mapping: Haar wavelet transform (approximation coefficients). FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

p	FA	MA	TD	VA	PA
2	9.84	11.31	3.53	81.98	3.18
4	9.55	10.95	3.36	82.86	2.83
8	8.98	10.60	3.18	83.57	2.65
16	18.24	11.31	3.00	79.68	6.01

Table 2 shows the results of mappings $f:\mathbb{R}^{32}\to\mathbb{R}^{p}, 2\le p\le 8$ based on retaining p Fourier components (CC excluded) in the processing of the locked mode signal in time windows 32 ms long.

Table 2: Base mapping: Fast Fourier transform. FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

p	FA	MA	TD	VA	PA
2	12.66	9.36	2.65	81.80	6.18
4	11.91	9.36	2.83	82.33	5.48
8	12.31	10.07	2.83	82.69	4.42

Table 3 presents a particular case of mapping $f:\mathbb{R}^{32}\to\mathbb{R}^{2}$, where the two components of the output samples are respectively the mean value of the locked mode signal in time windows of length 32 ms and the standard deviation of the Fourier spectrum (excluding CC component) of the locked mode signal in those windows. It is important to note that these components are just the features used by the APODIS predictor in JET [13].

Table 3: Base mapping: APODIS features. FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

p	FA	MA	TD	VA	PA
2	13.18	9.89	3.53	81.10	5.48

Tables 4, 5 and 6 are particular cases of mappings $f:\mathbb{R}^{32}\to\mathbb{R}$. Tables 4 and 5 use the Haar wavelet transform but compressing the information content of the processing time windows into just 1 point (level 5 of decomposition). The first table uses the approximation coefficients and the second one the detail coefficients. Table 6 uses the second feature of APODIS in JET: the standard deviation of the Fourier spectrum (excluding CC component).

Table 4: Base mapping: Haar wavelet transform (approximation coefficients). FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

p	FA	MA	TD	VA	PA
1	2.88	24.91	3.89	69.96	1.24

Table 5: Base mapping: Haar wavelet transform (detail coefficients). FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

p	FA	MA	TD	VA	PA
1	9.21	11.84	3.18	83.04	1.94

Table 6: Base mapping: second APODIS feature. FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

p	FA	MA	TD	VA	PA
1	10.82	8.66	3.00	83.22	5.12

According to the tables it is clear that predictors based on anomaly detections can be candidates as disruption predictors in devices like ITER or DEMO. Their main advantage is that they do not need data from past discharges for the real-time prediction of forthcoming disruptions. However, privilege knowledge can also be used.

Focusing the attention on the tables, the maximum percentage of valid alarms (83.57%) corresponds to map the input data into 8 points with the Haar wavelet transform (table 1). The valid alarm rate is a compromise that takes into account the relevant parameters related to disruptive discharges: success rate, missed alarms, tardy detections and premature alarms. On the other hand, the smallest false alarm rate (2.88%) is shown in table 4, but this case provides a reduced rate of valid alarms and, therefore, it is not a good balance. Apart from this false alarm rate, the minimum one (8.98%) is provided again in table 1 and $p = 8$. Therefore, the case $p = 8$ of table 1 should be considered as the winner of the cases presented in the six tables.

Fig. 10 and table 7 compare the predictions carried out by SPAD (table 1, $p = 8$), APODIS and the LMPT in JET. The plot covers the whole database taken into account (566 unintentional disruptions and 1738 non-disruptive discharges during JET ILW discharges). It should be noted that the LMPT obtains the worst results and SPAD has slightly better performance than the APODIS predictor.

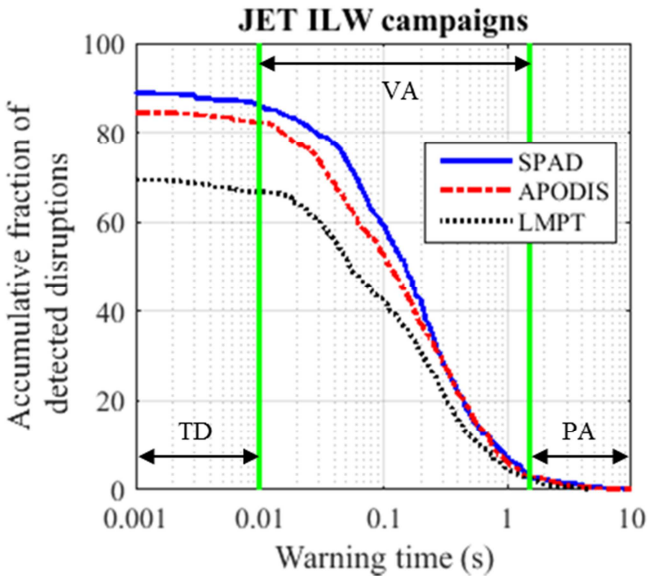

Fig. 10: Comparison between predictors. TD: tardy detections. VA: valid alarms. PA: premature alarms.

Table 7: Disruption prediction comparison among SPAD, APODIS and LMPT. FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

	MA	TD	VA	PA
SPAD	10.60	3.18	83.57	2.65
APODIS	15.38	2.47	79.15	3.00
LMPT	30.39	3.00	63.96	2.65

Figs. 11a-11d and table 8 break down JET results corresponding to the latest four experimental campaigns with the ILW. It is important to note that SPAD usually gives better results than the others, which is very important taking into account that no training with past discharges is required. Also, it should be mentioned that APODIS was trained with C-wall data and no re-training has been performed. As particular cases, in the campaign shown in fig. 11b, neither SPAD nor APODIS show tardy detections. In the campaign summarized in fig. 11d (H campaign), only SPAD shows tardy detections and the LMPT success rate is extremely low.

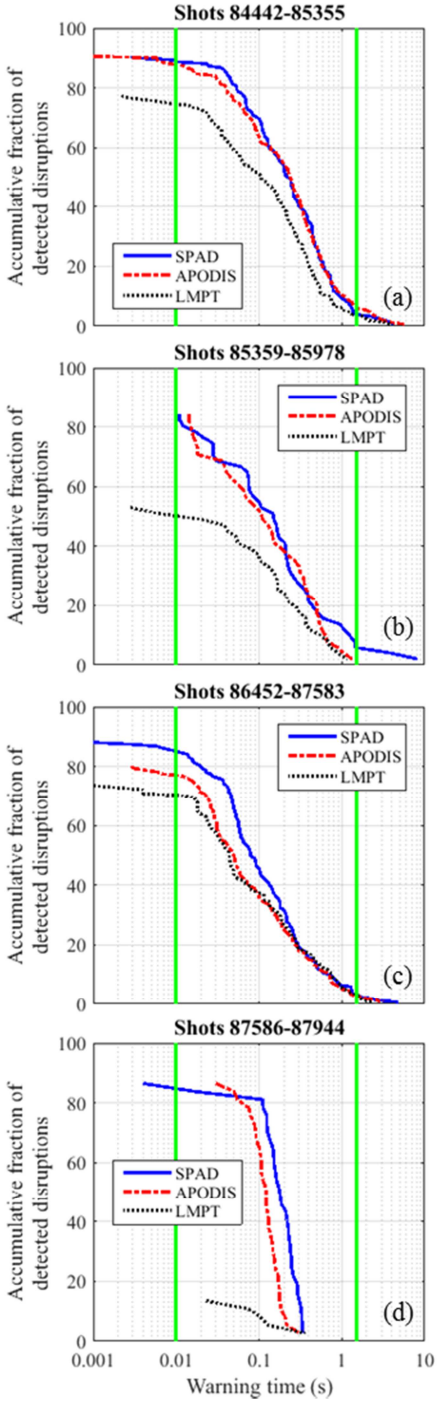

Fig. 11: Comparison between predictors in the four latest ILW campaigns of JET.

Table 8: Disruption prediction comparison among SPAD, APODIS and LMPT during the latest four ILW experimental campaigns of JET. FA: false alarms (%). MA: missed alarms (%). TD: tardy detections (%). VA: valid alarms (%). PA: premature alarms (%)

	MA	TD	VA	PA
SPAD (fig. 11a)	9.27	2.65	84.77	3.31
APODIS (fig. 11a)	9.27	3.31	81.46	5.96
LMPT (fig. 11a)	22.52	3.31	70.86	3.31

SPAD (fig. 11b)	15.69	0.00	80.39	3.92
APODIS (fig. 11b)	15.69	0.00	84.31	0.00
LMPT (fig. 11b)	47.06	3.92	49.02	0.00
SPAD (fig. 11c)	11.92	3.31	82.12	2.65
APODIS (fig. 11c)	19.87	3.31	74.17	2.65
LMPT (fig. 11c)	26.49	3.31	66.89	3.31
SPAD (fig. 11d)	13.52	2.70	83.78	0.00
APODIS (fig. 11d)	13.51	0.00	86.49	0.00
LMPT (fig. 11d)	86.49	0.00	13.51	0.00

7. Conclusions and directions of future work

Fig. 12 puts together the time traces of a JET disruptive discharge corresponding to I_P, *LM*, Mahalanobis distance and K_M. The plots of fig. 12 are to be compared with the ones of Fig. 13, which show the temporal evolution of the same signals for a non-disruptive discharge. The figures try to emphasize the fact that SPAD is sensible to anomalies in the temporal evolution of the locked mode signal. According to the results reported along the paper, it is clear that the simultaneous use of both the time domain and the frequency domain, even with a single quantity, provides an essential benefit versus the simple utilization of the signal amplitude (time domain). The results prove that predictors based on anomaly detections present performances which make them good candidates for devices like ITER. Indeed, in global term, they are not far from meeting the requirements of the next generation of devices. On the other hand, they present the main advantage that they do not need data from past discharges for the real-time prediction of forthcoming disruptions. However, privilege knowledge can also be used if available. Certainly, it should be noted that PK has been used in this work to empirically determine a good enough value of the outlier factor. However, work in progress is trying to determine critical outlier factors *on-the-fly* to trigger an alarm without using privilege knowledge. To this end, the framework of conformal predictors [21] is being used. An alarm will be fired when '*strange*' Mahalanobis distances appear in a discharge. The concept of '*strange*' means how conformal a Mahalanobis distance at a certain time is in relation to the distances at earlier times in the discharge. Also, different types of metric, such as geodesic, could be tested to see whether they can improve the success rate, as they did in some applications in the past [22].

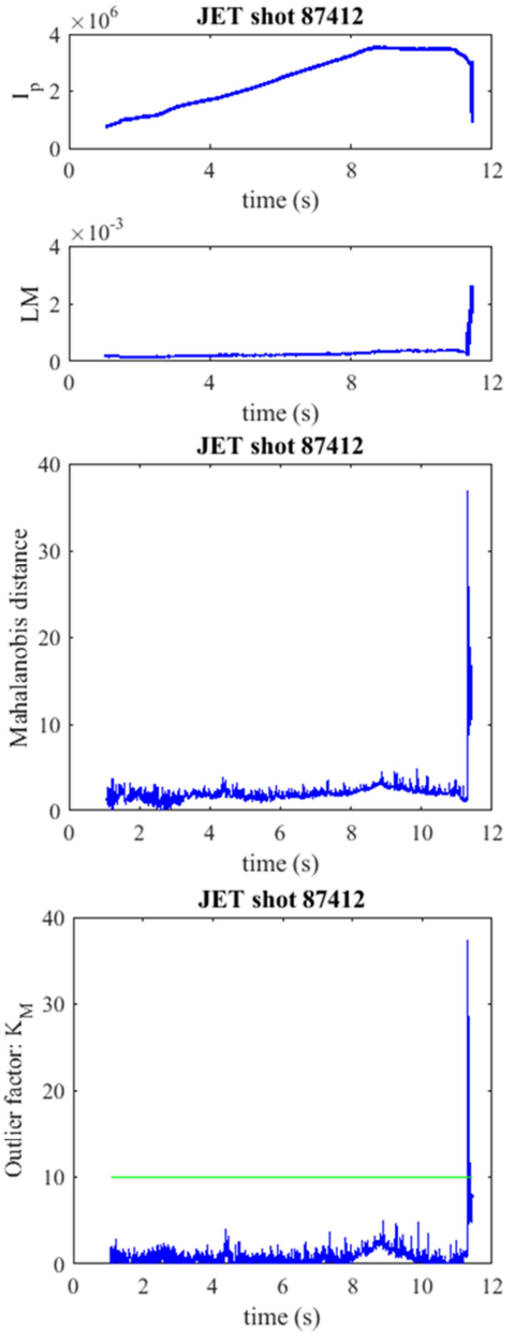

Fig. 12: Temporal evolution of the quantities used to detect anomalies in a disruptive discharge.

The other main direction of research consists of using more features as input to the classifier. Indeed the present version basically uses a single signal, the locked mode. A significant improvement in the performance is to be expected once other important quantities, such as the radiated fraction or the internal inductance were also included. A systematic investigation of the most adequate combination of signals for anomaly detection is also considered a prerequisite to the development of multi-machine predictors based on the proposed approach.

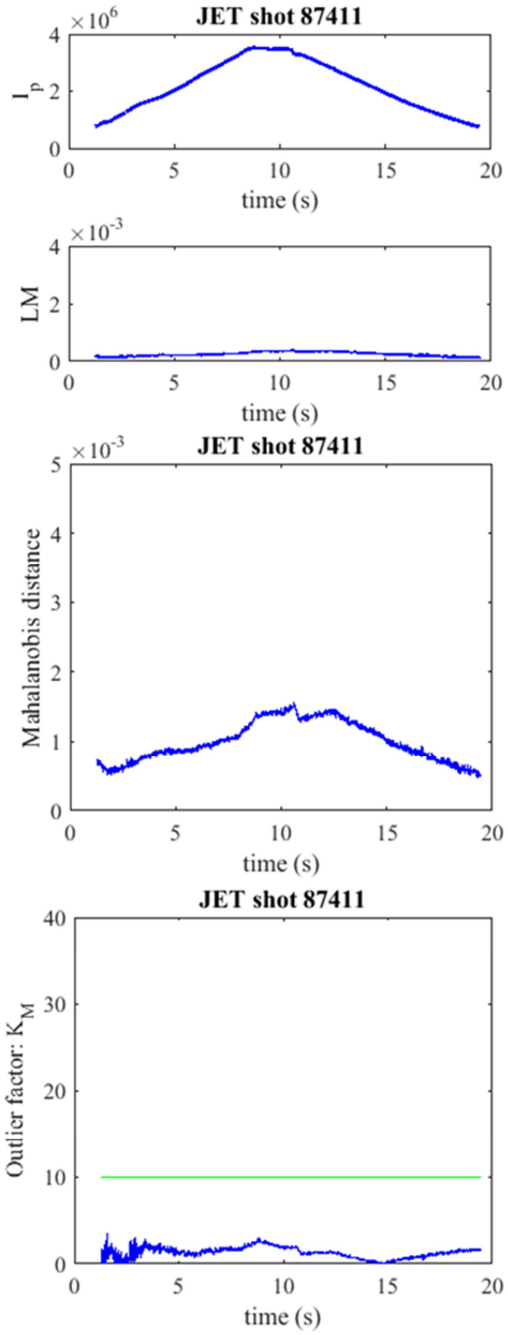

Fig. 13: Temporal evolution of the quantities used to detect anomalies in a non-disruptive discharge.

Appendix 1

False alarm: alarm triggered in a non-disruptive discharge.

Missed alarm: no alarm triggered before the disruption in a disruptive discharge.

Premature alarm: alarm triggered with a warning time greater than 1.5 s in a disruptive discharge.

Tardy detection: alarm triggered with a warning time less than 10 ms in a disruptive discharge.

Valid alarm: alarm triggered with a warning time between 10 ms and 1.5 s in a disruptive discharge.

Acknowledgments

This work was partially funded by the Spanish Ministry of Economy and Competitiveness under the Projects No ENE2012-38970-C04-01, ENE2012-38970-C04-03 and ENE2012-38970-C04-04.

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

[1] M. Lehnen, A. Alonso, G. Arnoux, N. Baumgarten, S.A. Bozhenkov, S. Brezinsek, M. Brix, T. Eich, S.N. Gerasimov, A. Huber, S. Jachmich, U. Kruezi, P.D. Morgan, V.V. Plyusnin, C. Reux, V. Riccardo, G. Sergienko, M.F. Stamp and JET EFDA contributors. “Disruption mitigation by massive gas injection in JET”. Nuclear Fusion 51 (2011) 123010 (12pp).

[2] M. Bakhtiari, G. Olynyk, R. Granetz, D.G. Whyte, M.L. Reinke, K. Zhurovich and V. Izzo. “Using mixed gases for massive gas injection disruption mitigation on Alcator C-Mod”. Nuclear Fusion 51 (2011) 063007 (9pp).

[3] G. Pautasso, K. Buchl, J. C. Fuchs, O. Gruber, A. Herrmann, K. Lackner, P. T. Lang, K. F. Mast, M. Ulrich, H. Zohm and ASDEX Upgrade Team. “Use of impurity pellets to control energy dissipation during disruption”. Nuclear Fusion, 36, 10 (1996) 1291-1297.

[4] N. Commaux, L. R. Baylor, S. K. Combs, N. W. Eidietis, T. E. Evans, C. R. Foust, E. M. Hollmann, D. A. Humphreys, V. A. Izzo, A. N. James, T. C. Jernigan, S. J. Meitner, P. B. Parks, J. C.Wesley and J. H. Yu. “Novel rapid shutdown strategies for runaway electron suppression in DIII-D”. Nuclear Fusion 51 (2011) 103001 (9pp).

[5] B. Esposito, G. Granucci, P. Smeulders, S. Nowak, J. R. Martín-Solís, L. Gabellieri, FTU and ECRH teams. “Disruption Avoidance in the Frascati Tokamak Upgrade by Means of Magnetohydrodynamic Mode Stabilization Using Electron-Cyclotron-Resonance Heating”. Phys. Rev. Lett. 100, 045006 (2008).

[6] B. Esposito, G. Granucci, M. Maraschek, S. Nowak, A. Gude, V. Igochine, E. Lazzaro, R. McDermott, E. Poli, J. Stober, W. Suttrop, W. Treutterer, H. Zohm, D. Brunetti and ASDEX Upgrade Team. “Avoidance of disruptions at high β_N in ASDEX Upgrade with off-axis ECRH”. Nuclear Fusion 51 (2011) 083051 (9pp).

[7] B. Cannas, R.S. Delogu, A. Fanni, P. Sonato, M. K. Zedda and JET-EFDA contributors. “Support vector machines for disruption prediction and novelty detection at JET”. Fusion Engineering and Design 82 (2007) 1124–1130.

[8] A. Murari, G. Vagliasindi, P. Arena, L. Fortuna, O. Barana, M. Johnson and JET-EFDA Contributors. “Prototype of an adaptive disruption predictor for JET based on fuzzy logic and regression trees”. Nuclear Fusion 48 (2008) 035010 (10pp).

[9] G. Pautasso, C. Tichmann, S. Egorov, T. Zehetbauer, O. Gruber, M. Maraschek, K.-F. Mast, V. Mertens, I. Perchermeier, G. Raupp,W. Treutterer, C.G.Windsor, ASDEX Upgrade Team. “On-line prediction and mitigation of disruptions in ASDEX Upgrade”. Nuclear Fusion 42 (2002) 100-108.

[10] B. Cannas, A. Fanni, G. Pautasso, G. Sias and P. Sonato. “An adaptive real-time disruption predictor for ASDEX Upgrade”. Nuclear Fusion 50 (2010) 075004 (12pp).

[11] G. A. Rattá, J. Vega, A. Murari, M. Johnson and JET-EFDA Contributors. “Feature extraction for improved disruption prediction analysis at JET”. Review of Scientific Instruments. 79, 10F328 (2008).

[12] G. A. Rattá, J. Vega, A. Murari, G. Vagliasindi, M. F. Johnson, P. C. de Vries and JET-EFDA Contributors. “An Advanced Disruption Predictor for JET tested in a simulated Real Time Environment” Nuclear Fusion. 50 (2010) 025005 (10pp).

[13] 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. “Results of the JET real-time disruption predictor in the ITER-like wall campaigns”. Fusion Engineering and Design 88 (2013) 1228-1231.

[14] S. Dormido-Canto, J. Vega, J. M. Ramírez, A. Murari, R. Moreno, J. M. López, A. Pereira and JET-EFDA Contributors. “Development of an efficient real-time disruption predictor from scratch on JET and implications for ITER”. Nuclear Fusion. 53 (2013) 113001 (8pp).

[15] J. Vega, A. Murari, S. Dormido-Canto, R. Moreno, A. Pereira, A. Acero and JET-EFDA Contributors. “Adaptive high learning rate probabilistic disruption predictors from scratch for the next generation of tokamaks”. Nuclear Fusion. 54 (2014) 123001 (17pp).

[16] J. Vega, R. Moreno, A. Pereira, S. Dormido-Canto, A. Murari and JET Contributors. “Advanced disruption predictor based on the locked mode signal:

application to JET". 1st EPS Conference on Plasma Diagnostics. April 14-17, 2015. Book of abstracts. Frascati, Italy.

[17] J. Vega, A. Murari, S. Dormido-Canto, R. Moreno, A. Pereira, G. A. Rattá and JET Contributors. "Disruption Precursor Detection: Combining the Time and Frequency Domains". SOFE program. 26th Symposium on Fusion Engineering (SOFE 2015). May 31st-June 4th, 2015. Austin (TX), USA.

[18] R. Moreno, J. Vega, A. Murari, S. Dormido-Canto, J. M. López, J. M. Ramírez and JET EFDA Contributors. "Robustness and increased time resolution of JET Advanced Predictor of Disruptions". Plasma Physics and Controlled Fusion. 56 (2014) 114003 (8 pp).

[19] S. Mallat, "A Wavelet Tour of Signal Processing", 2nd ed. Academic press, New York (2001).

[20] S. Theodoridis, K. Koutroumbas. "Pattern Recognition", 4th Edition. Academic Press, London (2009).

[21] V. Vovk, A. Gammerman, G. Shafer. "Algorithmic learning in a random world". Springer (2005).

[22] A. Murari et al. Nuclear Fusion 53 (3) 2013