# A statistical approach for the automatic identification of the start of the chain of events leading to the disruptions at JET

E. Aymerich[1], A. Fanni[1], G. Sias[1], S. Carcangiu[1], B. Cannas[1], A. Murari[2], A. Pau[3], and the JET Contributors*

[1] Electrical and Electronic Engineering Dept.-University of Cagliari, Piazza D'Armi, 09123, Cagliari, Italy.

[2] Consorzio RFX-Associazione - EURATOM ENEA per la Fusione, Padova, Italy

[3] Ecole Polytechnique Fédérale de Lausanne (EPFL), Swiss Plasma Center (SPC), CH-1015 Lausanne, Switzerland

* See the author list of "E. Joffrin et al 2019 Nucl. Fusion 59 112021".

Corresponding author: enrico.aymerich@unica.it

**Abstract- This paper reports an algorithm for the automatic identification of the beginning of the chain of events leading to a disruption, evaluating the so-called reference warning time. This is the time that separates the plasma current flat top of each disrupted discharge in two parts: a non-disrupted part and a pre-disrupted one. The algorithm is based on a statistical analysis of a set of dimensionless plasma parameters computed for a set of discharges selected from the JET experimental campaigns. In every data-driven or machine learning model, such as the GTM (Generative Topographic Mapping) predictor proposed in this paper, it is indeed necessary to label the samples needed for training the model itself. The samples needed to describe the disruption free behavior are extracted from the plasma current flat top phase of the regularly terminated discharges. Whereas, to describe the disrupted space, a proper selection of the pre-disruptive phase of each disruptive discharge plays a key role on the prediction performance of the model. Moreover, as well known, such models, which are highly dependent on the training input space, may be particularly prone to degradation; hence, a regular schedule of model review and retrain must be planned. The proposed algorithm avoids the cumbersome and time-consuming manual identification of the warning times, allowing, where appropriate, to implement a continuous learning system that can be completely automated. In this paper, the automatically evaluated warning times are compared with those obtained with a manual analysis in terms of the impact on the mapping of the JET input parameter space using the GTM methodology. Moreover, the algorithm has been used to build the GTM of recent experimental campaigns, with promising results.**

## 1. Introduction

Up to now, there are not self-consistent and general physical models available to reliably identify and predict the disruptions, whereas machine learning and data-driven approaches proved to be very useful tools for disruption prediction [1-9] and classification [10-12]. However, these algorithms require a certain number of regularly terminated and disrupted input discharges to learn how to predict disruptions and, more importantly, for each disrupted discharge, they require to identify the pre-disrupted phase to describe the disrupted input space of the model [13]. The size of the training set may depend on the complexity of the algorithm; even predictors developed to learn almost from scratch [5] may require tens of disruptive shots to perform well. Considering that ITER will not sustain more than a few major disruptions [13] at full performance, the disruptive data will be basically obtained through simulations [14], with low-performance discharges or from the lower-size tokamaks [15]. Nowadays, due to the availability of more powerful computing resources, Deep Learning (DL) algorithms were used [15-16], with very promising results towards a cross-machine predictor. For instance, in [15], authors tried to generalize the algorithm performance from DIII-D to JET in the ITER like Wall configurations. They trained their model, composed by elements from Convolutional Neural Network and Recurrent Neural Networks, on DIII-D, using 1-D profiles, then tried to generalize the performance on JET. The best results, however, have been reached using "glimpses" of data, which are basically samples from the JET's training set.

Despite the quite good results obtained with such black-box approaches, it would be beneficial to be able to use the prediction models also to understand the physical mechanisms which cause the discharge to disrupt.

The identification of the chain of events leading to a disruption will allow the implementation of specific control schemes to counteract the disruptive mechanism, the synthesis of features able to better detect the beginning of this destabilizing chain of events [13, 17], and, together with the use of standardized or nondimensional parameters as inputs to the model, will allow the possibility of extending the analysis in a cross-machine framework, helping to define scaling laws or standard parameters for the detection. Several papers [13, 17, 18] had shown an increase in the algorithms' performances and the results interpretability if the training set contains information related to the events which describe the disruptive behavior. Signals or diagnostics properly describing the physics of the disruption process improved the performance of several black-box approaches, such as ones based on Deep Learning (DL) [15, 16]. Another crucial step to a better understanding of the disruptive mechanisms is the standardization of all the reference times usually considered for the prediction. In [19] the authors made a strong effort towards the standardization in this field, developing a tool for the automatic definition of important times and parameters of the disruptions, such as thermal quench and the current quench times, the time of disruption ($t_D$) and the Mode Lock time ($t_{LM}$), which is the time where the locked mode amplitude starts to rise [19]. In the same direction, in [13, 17] the authors manually identify the so-called reference warning time of a disruption, which provides a reference time to separate the plasma current flat top of each disrupted discharge in two parts: a non-disrupted part and a pre-disrupted part. This second part is defined as the phase where the chain of events leading the disruption takes place. The introduction of consistent reference warning times ($T_i$) is doubly beneficial. Firstly, the warning times allow to identify the pre-disruptive phase, which is used to describe the disrupted input space of the model. In most of the literature, this pre-disruptive phase was statistically or heuristically identified and assumed equal for all the disruptions in the data base, introducing contradictory information in the prediction model. Secondly, being the warning time strongly linked to the onset of destabilizing phenomena, the predictor response should be connected to phenomenology or precursors that characterize the various types of disruptions.

Moreover, as the goal of the disruption prediction is moving nowadays from disruption mitigation to disruption avoidance, this would be only possible if the predictor provides its response in a suitable time prior the disruption depending on the characteristics times of the disruption precursor mechanisms and on the machine, and if it allows distinguishing among the different type of destabilizing chain of events. The key of a successful prediction model is therefore the capability, for each disrupted discharge in the training set, to discriminate among the non-disrupted and pre-disruptive phases following standard and coherent criteria, linked to the observed physical mechanisms. However, this classification requires a very time-consuming manual analysis [13,17]; hence, adopting it to classify tens of thousands of shots would be highly impractical. Therefore, in this work, an algorithm for the automatic identification of the reference warning times has been developed, based on a statistical approach. It makes use of similarity measures between distributions to quantify how much a disruptive pulse is becoming dissimilar from a typical regularly terminated discharge during its time evolution. The dissimilarity is evaluated for several plasma parameters and then an optimal weighted sum is assumed as overall dissimilarity. An optimal criterium has been introduced to automatically choose, for each discharge, the reference warning time over this total dissimilarity measure.

Note that, both in the manual and in the automatic case, we assume to be able to estimate (or, better, approximate) these reference warning times by analyzing several disruption precursors. In this paper, the algorithm is based on the statistical analysis of the plasma parameters selected from the JET experimental campaigns performed from 2011 to 2013, and provides, for each disrupted discharge, the reference warning time ($T_{i\text{-}AUT}$); these times have been compared with the ones ($T_{i\text{-}MAN}$) manually computed [13,17] and extensively applied to the prediction [13, 18] and classification [17] of disruptions. The comparison has been performed in terms of performance of a prediction model based on Generative Topographic Map (GTM), which has been adopted as one of the event detectors in the PETRA system (Plasma Event TRiggering and Alarms) at JET. In particular, the performance of the GTMs as disruption predictor, built using both the manual and the automatic warning times, have been compared on the same test set selected within the same experimental campaigns. Moreover, besides the very good results of that comparison, the proposed algorithm

has been applied to more recent campaigns at JET and the performance of the updated GTM model confirms the suitability of the algorithm.
This paper is organized as follows: Section 2 discusses the considered problem and background on diagnostics and features used as inputs to the proposed algorithm. Section 3 details the data base used to assess and optimize the algorithm and to validate it. Section 4 reports a detailed description of the proposed algorithm, whereas in Section 5 the algorithm is validated referring to Generative Topographic Mapping of JET. Finally, conclusions and future development are provided in Section 6.

## 2. Background

In [13] a set of dimensionless, machine-independent, physics-based features have been synthesized, which make use of 1-D plasma profile information. These features have been used as inputs to a GTM prediction model obtaining a 2D map of the multi-dimensional parameter space of JET, where it is possible to well identify a boundary separating the region free from disruption from the disruptive region. The GTM map has been used for disruption prediction projecting the discharge on the map and triggering an alarm depending on the disruption risk associated to its different regions.
Beside the choice of proper features as input to the model, another key factor to obtain high performance, regardless the prediction model, is a proper selection of the start of the chain of events leading to the disruption. However, the determination of the sequence of events between the root cause and the final disruption is not so straightforward. Many researchers made efforts to analyze and classify manually the different chain of events which lead to disruptions, for different machines. For instance, [20] reports a summary of the statistical occurrence and the respective dependence of the events for 275 unintentional disruptions at JET during the period 2011 to 2012, after the installation of the ITER-Like Wall. In [13,17] the reconstruction of the chain of events was a key step and allowed a coherent manual identification of the warning times ($T_{i\text{-}MAN}$), which separate the non-disrupted part of a disrupted discharge from the pre-disruptive evolution of the same discharge. The identification of such warning time was based on manual analysis of physics mechanisms and chain of events leading to disruptions. Figure 1 shows the different phases of a regular terminated (green) and a disrupted (red) discharge. The split of the disrupted discharge into two well-defined phases dramatically improved the performance of the GTM predictor [13], as the input information is coherently labelled and unambiguous. Moreover, this warning time is a term of comparison for whatever disruption prediction algorithm: in other words, the alarm time provided by a predictor should not anticipate the beginning of the pre-disruptive evolution of the discharge.

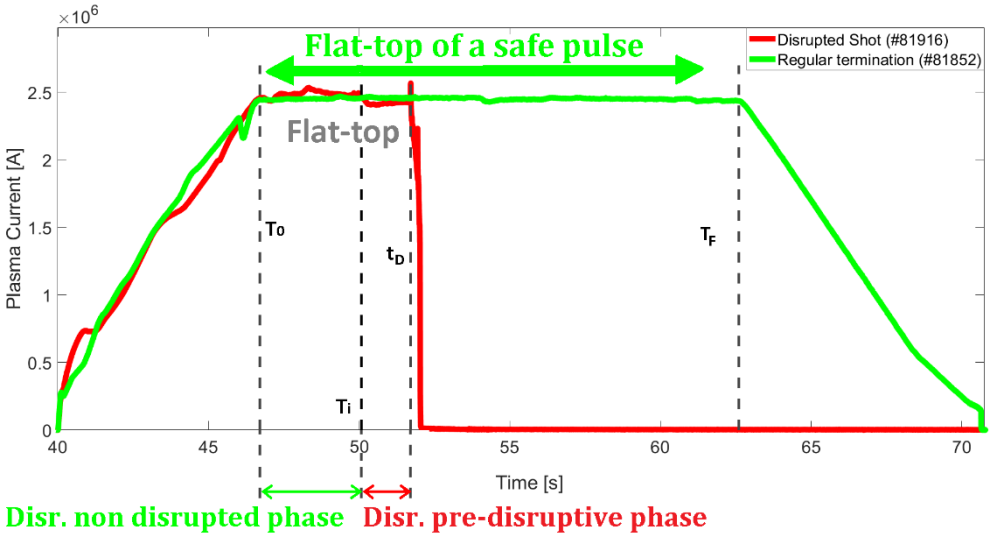


**Figure 1**. The plasma current evolution for a regularly terminated (#81852, in green) and a disrupted (#81916, in red) pulse. The flat-top of the disrupted shot is split in two: the non-disruptive phase, before the reference warning time ($T_i$)

and the pre-disruptive one, after $T_i$. $T_0$ is the starting time of the flat-top phase, whereas $T_F$ and $t_D$ (disruption time) are the flat-top phase ending times for a regular termination and a disruptive pulse, respectively.

### *2.1. Diagnostics and feature engineering*

The main purpose of the present paper is to present an algorithm for the automatic detection of the warning times using a limited set of diagnostic signals containing information on the spatial distributions of some relevant plasma properties, such as the Electron Temperature, the Electron Density and the Plasma Radiation. Moreover, the 0-D time series of the fraction of radiated power and the internal inductance have been used. In the following, the considered plasma parameters are explained in more detail together with the performed feature engineering, which is crucial to provide the algorithm with a set of inputs capable of separating the non-disrupted and pre-disruptive plasma behaviors. Another aim of the feature selection is the deletion of irrelevant or redundant information.

The profiles represent the time evolution of fundamental plasma quantities such as the electron temperature, the electron density and the radiation. In [17] these 1-D profiles have been processed to synthesize physics-based indicators called "peaking factors" (PFs), which have been used as input features of a GTM based disruption predictor [13]. The peaking factors demonstrated very useful to discriminate between a non-disruptive plasma state and a disruptive one. To compute the peaking factor of the electron temperature profile, either the ECE (Electron Cyclotron Emission) or the High-Resolution Thomson Scattering (HRTS) diagnostics can be used. The temperature peaking factors ($Te_{pf}$) based on the ECE diagnostic, in a not negligible number of cases, were found to be affected by the cut-off of several channels [21]. The effect was sometimes marginal, other times was heavily compromising the calculation of the peaking factor itself. This is the reason why the analysis of this work computes the peaking factors using only the HRTS profiles although the HRTS has a lower time resolution than the ECE. The HRTS diagnostic provides 63 data points with a repetition rate of 20 light pulses per second (20Hz). The spatial resolution of the measurements for the core region and the pedestal is respectively of 1.6 cm and 1 cm. The peaking factors, for both density and temperature, have been considered as features defined with a "core versus all" metric; they are computed as the ratio between the mean value of the considered radial profile (temperature, radiation, density) around the magnetic axis and the mean value of the measurements over the entire radius. The radial interval to define the "core" with respect to the magnetic axis is the 25% of the radial coordinate (the minor radius for poloidal mid-plane measurements) in the case of electron temperature ($Te_{pf}$) and density ($Ne_{pf}$) peaking factors.

Regarding the radiated power, in [17] the same authors computed the peaking factors using the main-vessel bolometric camera with a horizontal view of the plasma cross-section (Bolo H). The camera collects the radiation along 24 channels, 8 of which cross the divertor region and the region adjacent to the divertor with 8 cm separation between the channels' axes. The other 16 channels cover the entire plasma. A simple pinhole structure is used to define the lines-of-sight of the camera [22]. Two different peaking factors have been derived splitting the information carried out by the global poloidal radiation distribution. We firstly define the core as 4 channels of the Bolometer, from 13rd to 16th, and the divertor as 8 channels, from 1st to 8th. Then, we compute the two peaking factors, the $Rad_{pf\_CVA}$ and the $Rad_{pf\_XDIV}$: the first one is the ratio between the average radiation in the core, and the average radiation in the entire plasma excluding the divertor area. The $Rad_{pf\_XDIV}$ is instead computed as the ratio between the average radiation in the divertor and the average radiation in entire plasma excluding the area of core (20 channels). These peaking factors describe the two main mechanisms involving a radiation collapse: the accumulation of high-Z impurities in the plasma core as opposed to edge-radiative collapse. As analyzed in [19], the main events destabilizing a discharge are quite different, as well as the time scales of the corresponding chain of events. Nevertheless, the two destabilizing factors are not mutually exclusive; despite developing on different time scales, there are cases where both are simultaneously affecting the discharge. In general, even though in many cases it is possible to find clear examples of a well-defined chain of events corresponding to a specific disruption type, in other cases there is an interplay of more than one mechanism destabilizing the discharge. Hence, the synthesis of more detailed and targeted indicators

goes in the direction of a more accurate and flexible avoidance and prediction system. Finally, this split allows, in a sense, a decoupling of the two behaviors, improving the resolution of each of them.
Within the profile indicators also the internal inductance $Li$ has been considered as representative of the current density profile.
In addition, a further dimensionless parameter has been considered, i.e., the fraction of radiated power divided by the total input power ($P_{FRAC}$), which is a well-known indicator of the power balance.
For the considered discharges, the signals have been uniformly sampled with a time step of 2ms, then they were processed with a casual median filter of 40 ms width.

## 3. Data Base

To build the data base, both disrupted and regular terminated discharges have been selected from experimental campaign performed at JET from 2011 to 2016, after the installation of the ITER-Like Wall. Only the discharges where all the signals, needed to compute the features described in the previous subsection, were available and consistent have been selected. Moreover, the discharges caused by a Vertical Displacement Event, the ones terminated by massive gas injection and those in limiter configuration were excluded. In the present work, the analysis of the pulses refers to the flat-top phase; the ramp-up and the ramp-down have not been considered.
The data base includes two sets; the first set contains 132 disrupted and 115 regularly terminated discharges within the ITER Like Wall (ILW) experimental campaigns performed at JET from 2011 to 2013 and already considered in [13], for which the warning times were manually identified ($T_{i\text{-}MAN}$). In the following we refer to it as C28-C30 data set. This first set has been used to perform the statistical analysis and to assess and optimize our algorithm. In order to test the generalization capability of the algorithm, a second data set has been selected, which includes 29 disrupted and 41 regularly terminated pulses within the more recent (2016) campaigns both in baseline and hybrid scenarios (we refer to this second set as C36 data set). In this case, the suitability of the algorithm to correctly identify the pre-disruptive phase of the disrupted discharges has been evaluated in terms of the composition of the GTM that maps the more recent input space, i.e., in terms of its capability to discriminate between disrupted and non-disrupted regions.
Note that, the limited number of pulses in the C36 data set satisfying the previously described requirements is due to the increased number of mitigated disruption by MGI.
In order to have a look of the operational scenarios between the two datasets, Figure 2 compares the distribution of their main plasma parameters for the regularly terminated discharges: plasma current, $I_p$, toroidal field, $B_T$, normalized beta, $\beta_N$, total input power, line integrated density, edge safety factor $q_{95}$. It can be seen that they occupy roughly the same ranges of values but with slightly different distributions.

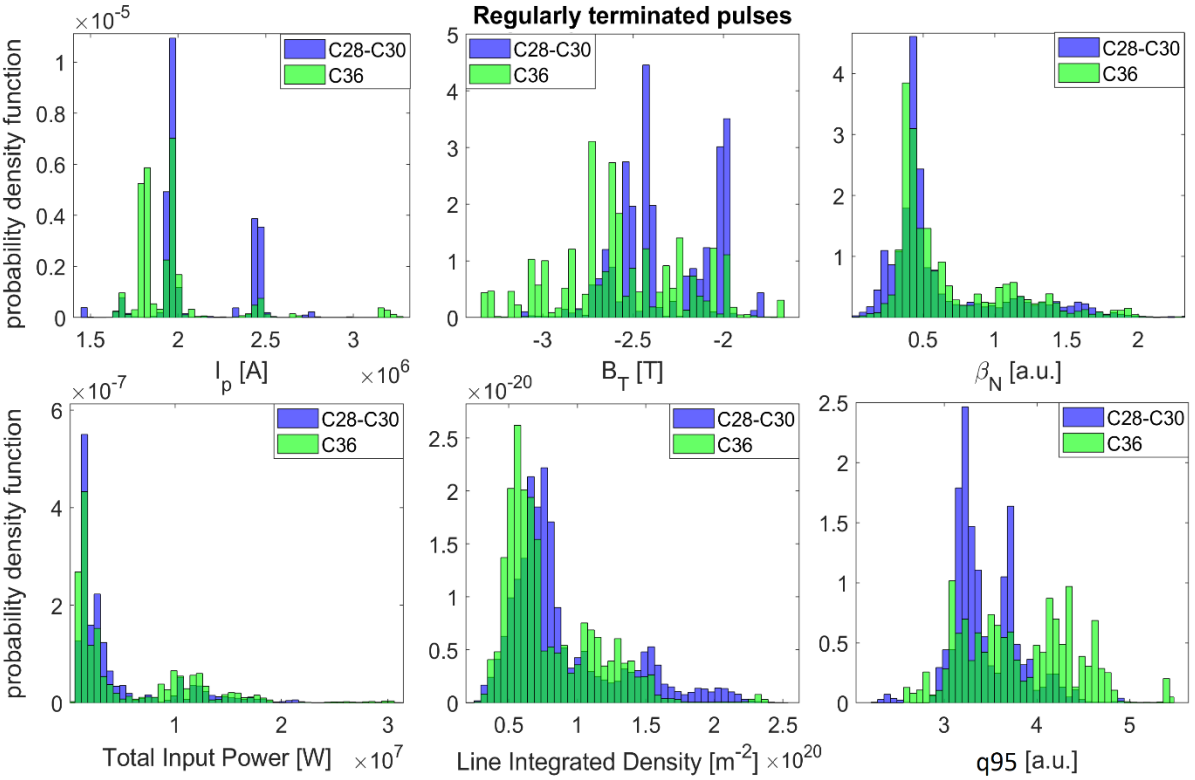


**Figure 2.** Probability density functions of the main parameters of the regularly terminated discharges in C28-C30 (blue) versus those in C36 (green) data sets for (from top left to bottom right): plasma current, toroidal field, normalized beta, total input power, line integrated density and edge safety factor $q_{95}$.

Table 1 reports the nondimensional plasma parameters considered to develop the proposed algorithm for the automatic identification of the warning times $T_{i\text{-}AUT}$, which are the same used to develop the GTM prediction model. The last column of the table reports the weights assigned to the parameters as a result of the algorithm optimization, which will be detailed in the following. Literature [13, 17] proved that the selected features discriminate well between regularly terminated and disrupted pulses.

**Table 1.** Plasma parameters: parameter names, Acronyms, optimized weights.

| Parameter name | Acronym | Weight |
|---|---|---|
| Peaking Factor of Temperature | $Te_{pf}$ | 1 |
| Peaking Factor of Electron Density | $Ne_{pf}$ | 1 |
| Peaking Factor of the Radiation (excluding the contribution of the X-point/divertor region) | $Rad_{pf\_CVA}$ | 0.8 |
| Peaking Factor of the Radiation (excluding the contribution of the core region) | $Rad_{pf\_XDIV}$ | 0.5 |
| Internal Inductance | $L_i$ | 1 |
| Fraction of the Radiated Power | $P_{FRAC}$ | 0.7 |

Figures 3 and 4 report these features for the regularly terminated discharge # 83747 and the disrupted discharge #81916 respectively. The disrupted discharge is a high-Z impurity accumulation (or Radiation Peaking RPK) disruption [12, 23] with warning time $T_{i\text{-}MAN}$ manually set at 50.07 (highlighted with a red vertical line in Figure 4).

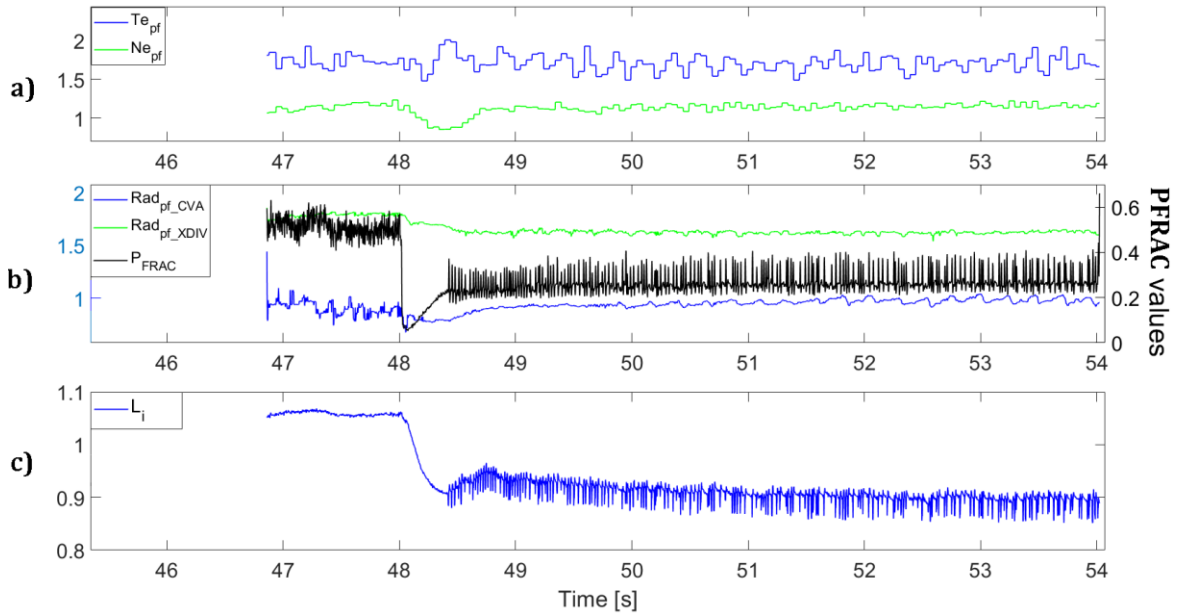


**Figure 3.** The input features for the algorithm for the JET regularly terminated discharge #83747: a) the peaking factors of the temperature ($Te_{pf}$, in blue) and density ($Ne_{pf}$, in green); b) the radiation peaking factors with the metric "Core Vs All" ($Rad_{pf\_CVA}$, in blue), which excludes the divertor, and with metric "Edge Vs All" ($Rad_{pf\_XDIV}$, in green), which excludes the core, and the Power Fraction ($P_{FRAC}$, in black); c) the internal inductance ($L_i$, in green).

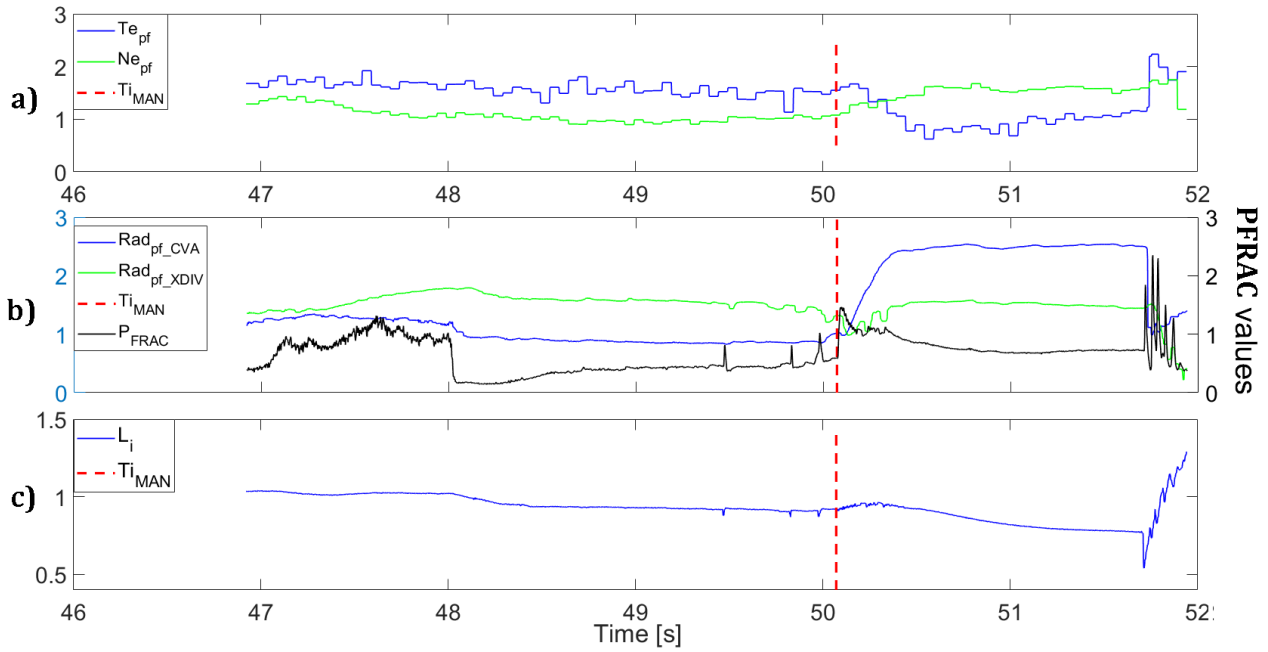


**Figure 4.** The input features for the algorithm for the JET disrupted discharge #81916: a) the peaking factors of the temperature ($Te_{pf}$, in blue) and density ($Ne_{pf}$, in green); b) the radiation peaking factors with the metric "Core Vs All" ($Rad_{pf\_CVA}$, in blue), which excludes the divertor, and with metric "Edge Vs All" ($Rad_{pf\_XDIV}$, in green), which excludes the core, and the Power Fraction ($P_{FRAC}$, in black); c) the internal inductance ($L_i$, in green). A vertical red line marks the manually detected warning time $T_{i\text{-}MAN}$.

It can be noted that, in the regularly terminated discharge the variation range of the signals is generally smaller than in the disrupted one; while this remark may be valid in most of the cases, it is not necessarily true for all the discharges. Moreover, looking at Figure 4, it can be seen that the peaking factors characterize well the typical RPK evolution: the $Ne_{pf}$ shows an increase of the density in the plasma core correlated with a temperature drop. Moreover, the peaking factor of radiation at the core rises, as well as the overall fraction of radiated power, while the internal inductance starts to decrease. This chain of events starts from the penetration

of high-Z atoms in the core that produces a change in the kinetic and current profiles that eventually leads to a destabilization of the MHD equilibrium in the plasma. The proposed algorithm weighs the variations in these signals' distributions to identify the start of the chain of events leading to disruption. This is done by comparing the distribution of each signal in the regularly terminated discharges in different time instants with the distribution of the same parameter of the single disrupted discharge, as detailed in the next section.

## 4. Automatic detection of the warning time

### *4.1 Statistical analysis*

A univariate statistical analysis has been firstly performed to evaluate the power of each selected feature in discriminating between disruptive and non-disruptive behavior. This analysis has been performed on the first set of discharges of the data base (C28-C30 data set). Figure 5 reports the probability density functions (*pdf*) of the six parameters in Table 1 for the non-disruptive pulses (blue) versus the non-disrupted phase of the disruptive pulses (red). Here, the manual selected warning times have been used to discriminate between the non-disrupted and pre-disruptive phases of the disrupted discharges. The results of the analysis, reported in Figure 5, refer to phases that can be considered in a non-disrupted condition. It can be observed that there is an overlap between the *pdf* of the parameters of non-disrupted discharges and the non-disrupted phase of disrupted ones. Figure 6 reports the *pdf* of the parameters of the non-disruptive pulses (blue) versus the pre-disruptive phase of the disrupted pulses (red) for the same parameters in Figure 5. Looking at Figure 6, it can be seen that the parameters distribute differently during the unstable phase (i.e., after the warning time $T_{i\text{-}MAN}$) with a wider range of parameter values. Moreover, the *pdfs* of the pre-disruptive phases of the disrupted discharges shift with respect to the stable phases. The orange arrows in Figure 6 highlight the shifts.
Summarizing, during the non-disrupted phase of the disrupted discharges the distribution of the parameters is very similar to the distribution of the regularly terminated discharges, while during the pre-disruptive phase, the values are distributed quite differently.
The main idea of the proposed algorithm is to introduce distance/similarity measures between these probability density functions when the reference warning time varies, in order to automatically identify the moment when a disrupted discharge starts its pre-disruptive evolution. For instance, Figure 7 compares the distribution of the temperature peaking factor of the non-disrupted pulses (blue) in the database with the *pdf* of a window of 500 ms, centered at different time instants, of the disruptive discharge #81916. From a) to d) the time instant is getting closer and closer to the time of disruption, where in c) the time instant is the closest to the manually selected warning time ($T_{i\text{-}MAN}$) (about 50.07s) [13]. The time evolution in Figure 7 clearly shows that, approaching to the actual warning time, the overlap of the two distributions reduced.

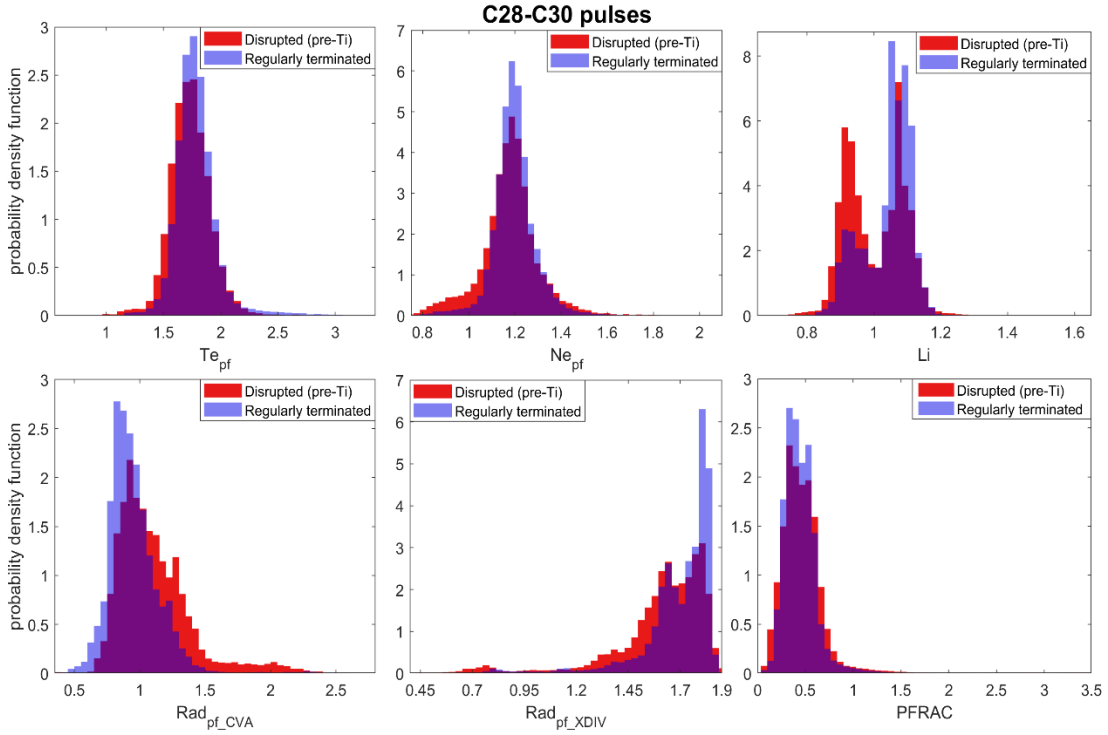


**Figure 5.** C28-C30 data set: Probability density functions of the parameters of the regularly terminated pulses (blue) versus the non-disrupted phase of the disrupted pulses (red) for (from top left to bottom right): electron temperature peaking factor, electron density peaking factor, internal inductance, radiation at the core peaking factor, radiation at the edge peaking factor, fraction of radiated power.

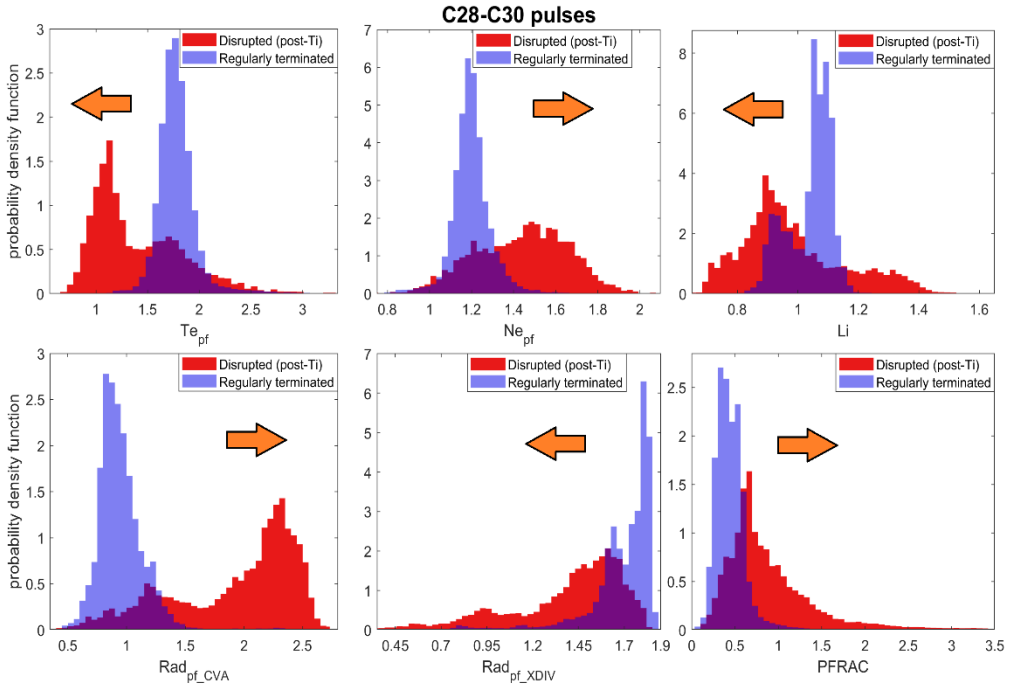


**Figure 6.** C28-C30 data set: Probability density functions of the parameters of the regularly terminated pulses (blue) versus the pre-disruptive phase of the disrupted pulses (red) for (from top left to bottom right): electron temperature peaking factor, electron density peaking factor, internal inductance, radiation at the core peaking factor, radiation at the edge peaking factor, fraction of radiated power. The shift of the distributions is marked with an orange arrow.

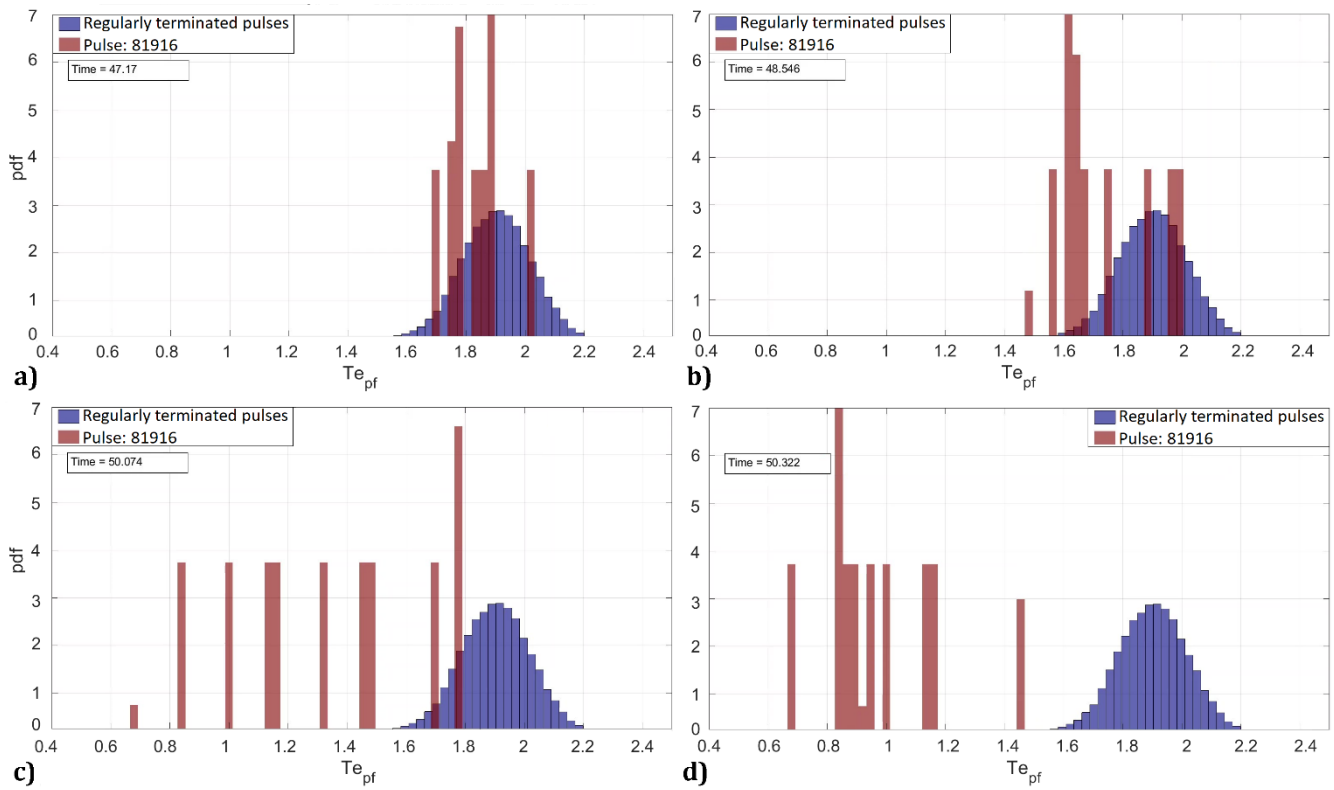


**Figure 7.** Probability density functions of the temperature peaking factor ($Te_{pf}$) of the regularly terminated pulses (blue) in the C28-C30 data set versus the *pdf* of a 500 ms window, centered at different time instants (indicated on each subplot), of the disruptive discharge #81916. From a) to d) the time instant is getting closer and closer to the time of disruption, where in c) the time instant is the closest to the manually selected warning time ($T_{i\text{-}MAN}$) [13].

### *4.2 The algorithm*

As discussed in [13,17], the selection of the warning time $T_{i\text{-}MAN}$ required a tedious and time-consuming analysis of several events and parameters, additional to the ones used as inputs for the proposed algorithm, and not necessarily available in real time. In this paper, a *Warning Time Indicator* (*WTI*) has been built that can be used to automatically detect the warning time in the disrupted discharges.

As previously mentioned, the algorithm is based on the comparison of the distributions of the selected plasma parameters in the regularly terminated and in the disrupted discharges. In particular, it is assumed that, before the onset of the chain of events leading to disruption (before the actual warning time $T_i$), the distributions of the parameters in the disruptive discharges are close to those of the regularly terminated ones, whereas they become more and more dissimilar while approaching the disruption time. Hence, for each plasma feature in Table 1, the distribution of the regularly terminated pulses (*SAFE_distr*) has been considered as the reference distribution. Then, for each discharge and for each time instant *t*, the algorithm scans every parameter from the beginning to the end of the flat-top, identifying two different distributions:

- *LEFTpart_distr*: the distribution before *t*
- *RIGHTpart_distr*: the distribution after *t*

and computes the distance/similarity between these two distributions to the *SAFE_distr*.

Note that, for time instants at the beginning (at the end) of the flat-top, a very small number of samples is available for the *LEFTpart_distr* (*RIGHTpart_distr*). This creates a border effect at the beginning (at the end), which has been partly compensated by padding the first 125 ms of the initial and final part of each signal. The padding has been done by simply replicating the respective part of the signal, so that at the beginning of the flat-top and at its end, the distributions could be represented by more values.

In order to evaluate the distance/similarity, several metrics have been considered [24], based both on the computation of misclassification probability, such as Bhattacharya, Hellinger, Kullback-Leigler Divergence

and Matusita and on the computation of the distribution similarities, such as those belonging to the inner product family. Among all the tested metrics, in this paper, the final choice was the Cosine similarity metric, which basically implements the normalized inner product:

$$s_{Cos} = \frac{\sum_{i=1}^{B} P_i Q_i}{\sqrt{\sum_{i=1}^{B} P_i^2} \sqrt{\sum_{i=1}^{B} Q_i^2}}$$

where, $P$ and $Q$ are the two probability density functions, each composed by the same number B of bins.
This metric is itself normalized between 0 and 1 and allows to add the measures referred to different parameters without rescaling them regardless of their range of variation.
Hence, two similarity measures have been evaluated for each parameter: the similarity of the left part of the discharge with the disruption-free input space (*LEFTpart_simil*) and the similarity of the right part of the discharge again with the same disruption-free input space (*RIGHTpart_simil*). For a disrupted discharge, when approaching the actual warning time $T_i$, it is expected that the right part distribution has similarity value close to 0, and the left part has similarity value close to 1. In fact, in such a case, in the left part the discharge is still in the non-disrupted phase, whereas in the right part it already shows a disruptive behavior.
These similarity measures are normalized with respect to the similarity of the whole flat-top phase (*Total_simil*), then the values are truncated to 1; this adjustment makes the algorithm work for the shots where the signal range is very different from the non disruptive one, even during the non-disrupted phase.
Subsequently, the normalized left part similarity is subtracted from the normalized right part similarity and the negative values are truncated to 0.
Then, the standard deviation of each plasma parameter is computed in a sliding window of 500 ms width, when the flat-top phase lasts more than 500ms, otherwise it is set equal to half flat-top length. Since the parameters may have different ranges, they are normalized between 0 and 1 before computing the standard deviation.
For each plasma parameter, an indicator is evaluated by weighing its standard deviation with the difference of the similarities. Hence, the parameter variations which do not produce a destabilization of the discharge are neglected.
Figure 8 shows, as an example, the construction of the indicator for the $Rad_{pf\_CVA}$ signal of the pulse #81916. Figure 8a) reports the signal $Rad_{pf\_CVA}$ (blue) and the same signal padded at the beginning and at the end (red dashed line) to avoid border effects processing the signal. Figure 8b) reports the normalized left part similarity (in blue), the normalized right part similarity (in red), and the difference between the blue and red signals (in yellow), where negative values are truncated to 0. Figure 8c) reports the $Rad_{pf\_CVA}$ standard deviation computed in the sliding window (red) and the $Rad_{pf\_CVA}$ indicator (in blue), computed as a time by time product between the yellow signal in Figure 8b) and the standard deviation.

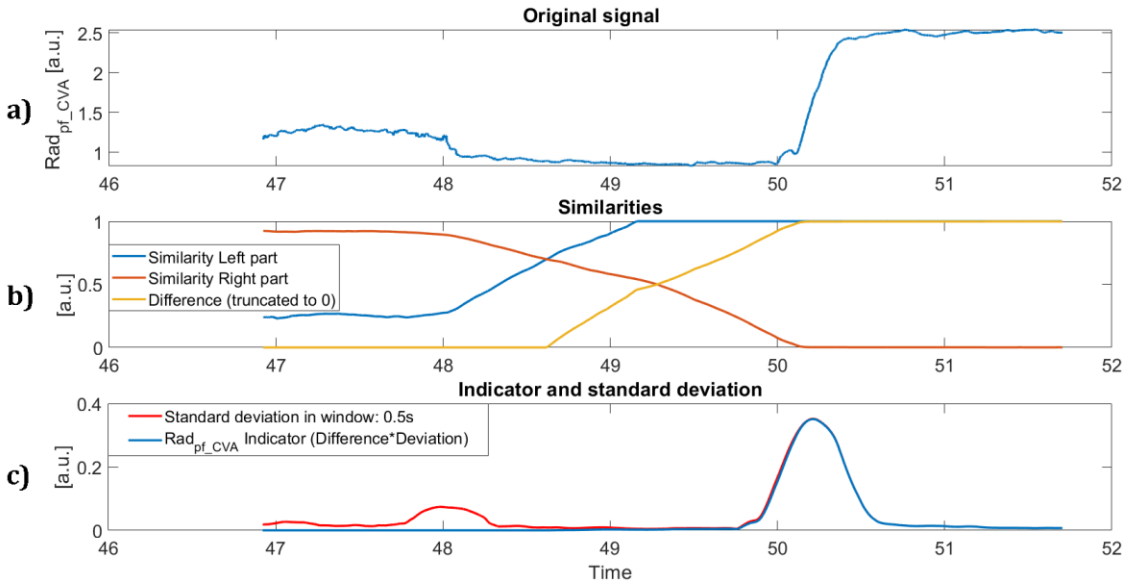

**Figure 8.** Construction of the indicator for the parameter $Rad_{pf\_CVA}$, of the disrupted shot #81916: a) $Rad_{pf\_CVA}$ (blue), and $Rad_{pf\_CVA}$ padded at the beginning and at the end (red dashed); b) normalized *LEFTpart_simil* (blue), normalized *RIGHTpart_simil* (red), and their difference (yellow), where negative values are truncated to 0; c) standard deviation computed in a sliding window of variable length, adjusted depending on the signal length (maximum value is 0.5s) (red) and the indicator (blue).

It can be noted that, at around 48 s, the original signal varies and produces some peaks in the windowed standard deviation; these variations of the signal, on the other hand, are not moving the signal distribution outside the non disruptive disruption one: this determines a low value of the similarity difference and hence a low value of the indicator for the $Rad_{pf\_CVA}$. This is not true for the following variation at around 50s, which is the time when there is the beginning of the chain of event leading to the disruption. The indicator highlights the points where there is both a variation from the disruption-free input space and a variation in the signal trend. This is the reason why, in Figures 8c, the indicator grows at around 50.3 s and then drops afterwards, due to the drop of the standard deviation.

Finally, an overall indicator (*Warning Time Indicator* or *WTI*) is evaluated as the weighted sum of the single plasma parameter indicators. To set the parameter weights an optimization procedure has been performed, as described in the next subsection. Table 1 (last column) shows the finally adopted weights.

Figure 9 shows the *WTI* for the regularly terminated discharge #83747(a) and for the disrupted discharge #81916 (b), already considered in Figure 3 and Figure 4. Note the different range of variation.

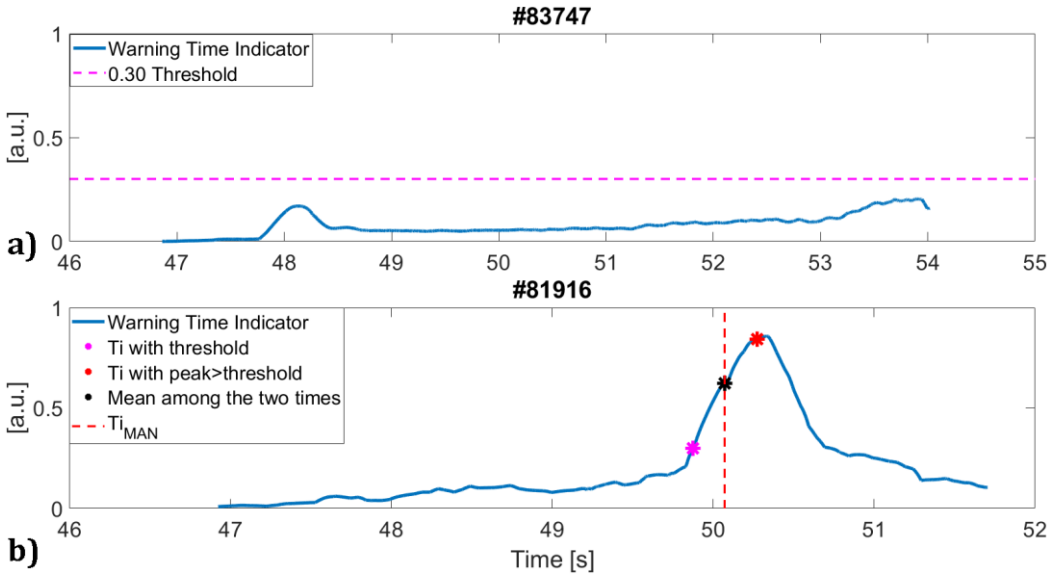


**Figure 9.** Overall Indicator: a) regularly terminated pulse #83437; b) for the disrupted pulse #81916.

Figure 10 reports the pseudo-code of the algorithm to construct the *WTI*.

```
WTIndicator = 0
For every signal:
   signal = padded_signal
   SIGNALweight = optimized weight of the indicator for this signal
   for every instant_t in time:
      #Defining the similarity
      LEFTpart_distr = distribution of signal before instant_t
      RIGHTpart_distr = distribution of signal after instant_t
      Total_simil = similarity(SIGNAL_distr,SAFE_distr)
      LEFTpart_simil = similarity(LEFT_part_distr,SAFE_distr)/Total_simil
      RIGHTpart_simil = similarity(RIGHTpart_distr,SAFE_distr)/Total_simil
      LEFTpart_simil and RIGHTpart_simil are truncated to 1
      SIMILmeasure = LEFTpart_simil-RIGHTpart_simil
      Negative values of SIMILmeasure are truncated to 0

      #Defining the window
      window_length = min(Signal_length/4, 500 ms)
      Window = [instant_t - window_length/2, instant_t + window_length/2]
      STD_win = standard deviation of signal in Window

   SIGNAL_indicator = SIMILmeasure * STD_win
   WTIndicator = WTIndicator + SIGNAL_indicator*SIGNALweight
```

**Figure 10.** Pseudo-code for the *WTI*

### *4.2.1 Automatic identification of the warning time $T_{i\text{-}AUT}$*

As expected, the ranges of variation of the *WTI* are very different among the regularly terminated and the disrupted pulses. Moreover, looking at Figure 9b, it can be noted that the *WTI* highlights the moment when the features are varying, so that a threshold can be used to identify the onset of the chain of events leading to disruption.

Figure 11 shows the distribution of the values of the *WTI* for the regularly terminated pulses in the C28-C30 data set where the value 0.3 corresponds to the 99th percentile. Using this value as a threshold on the *WTI*, a warning time of 50.02 s is obtained (magenta star in Figure 9b). Other criteria have been taken into consideration to detect the warning time, such as the time corresponding to the first local maximum of the *WTI* greater than 0.3 (red star in Figure 9b), or the mean between the previous two.

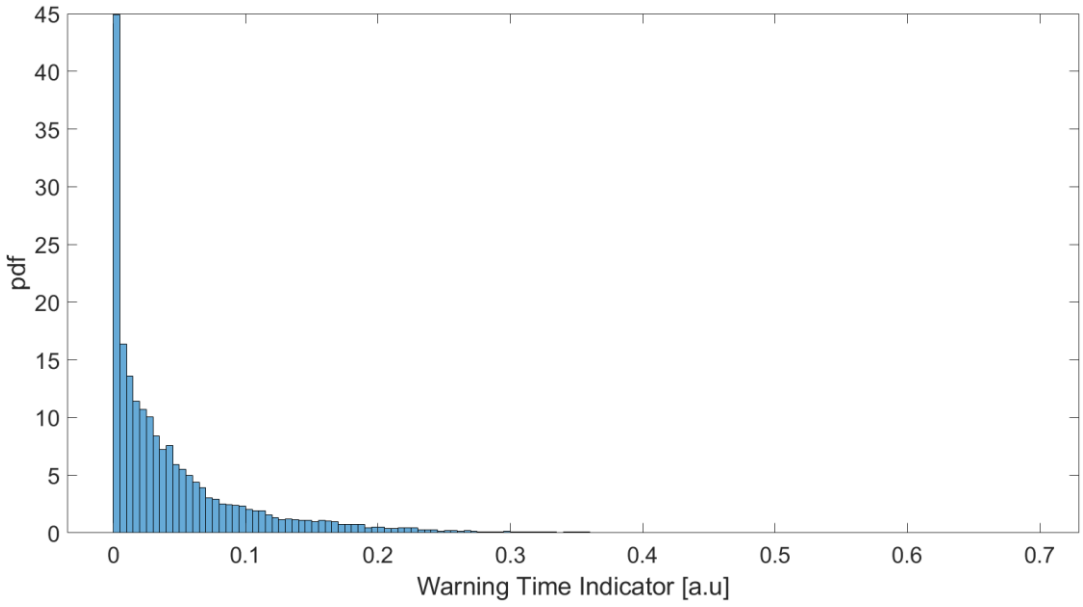


**Figure 11.** Probability density function of the WTI values for the regularly terminated pulses in the C28-C30 data set.

The best criterion is the mean between the time detected using the threshold equal to 0.3, with an assertion time of 10 samples (20 ms), and the time of the first peak of the *WTI* greater than 0.3. It has been chosen in order to maximize the degree of separability of disrupted and non-disrupted regions in the GTM map.
Moreover, in order to consider disruptive processes characterized by fast time scales, which cannot be identified through the proposed statistical method, the mode locking occurrence has been also considered. Note that, algorithms based on the Mode-Locking (ML) signal already exist and are implemented in the largest devices, to trigger an alarm and mitigate the disruption.
Finally, the warning time has been identified as the lower time between the mode locking occurrence and the time obtained with the *WTI*. In this case, the value of the *WTI* may be greatly lower than the threshold.
Assuming such criterion on the *WTI*, the corresponding warning time for the pulse #81916 is 50.075s, which is very close to the manually selected warning time $T_{i\text{-}MAN}$ (50.07 s) (see Figure 9 b where this warning time is identified by the black star, whereas $T_{i\text{-}MAN}$ corresponds to the vertical red dashed line). Furthermore, no warning time is detected for the regularly terminated discharge #81852 (see Figure 9a).

### *4.2.2 Optimization of the algorithm parameters*

As previously mentioned, the *WTI* is obtained as a weighted sum of the indicators of the plasma parameters in Table 1. Varying the weights leads to different warning times, and therefore to different GTM maps. Only three of the six weights have been optimized, namely the two peaking factors of the radiation and the radiated fraction of the total input power, because they are all expression of the plasma radiation, whereas the other three weights have been set to the maximum value (equal to one). The optimization strategy consists in exhaustively exploring the search space along the three coordinate directions (in this case each coordinate corresponds to a weight) and considering as goal of the optimization, again, the maximization of the degree of separability of disrupted and non-disrupted regions in the map, which means the minimization of the percentage of samples falling in the mixed clusters of the GTM (grey clusters in Figure 12). During the search, each weight value has been uniformly varied between 0.1 and 1 with step 0.1. The optimal weights are reported in the last column of Table 1, which correspond to the minimal percentage of samples in grey clusters equal to 21.47%.
Figure 12 a) shows the GTM ($GTM_{C28\text{-}C30\text{-}AUT}$) trained using the warning times $T_{i\text{-}AUT}$ obtained with the optimal weights reported in Table 1. The clusters in the map are colored on the basis of the node composition: the green clusters contain only samples coming from regularly terminated pulses (safe samples), the red clusters contain only samples coming from the pre-disruptive phase of the disrupted discharges (disruptive samples), whereas grey mixed clusters contain both safe and disruptive samples. The white clusters are empty. Figure 12b) reports the GTM trained using the manually identified warning times $T_{i\text{-}MAN}$ ($GTM_{C28\text{-}C30\text{-}MAN}$).
The six parameters listed in Table 1 have been used to train both the GTMs. For the sake of comparison, the GTM hyperparameters, such as the number of latent points (2500), the number of radial basis functions (400) and their variance $\sigma = 0.8$, have been assumed equal to the ones used in [13], as well as the training set, which contains the same 89 disrupted shots and 70 regular terminations used in this paper. However, unlike in [13], in Figure 12a) the pre-disruptive phase of the disrupted discharges has been identified using $T_{i\text{-}AUT}$ instead of $T_{i\text{-}MAN}$. It can be seen that, in both the maps, there is a well-defined separation between the two regions representing the disruptive (red) and non-disruptive (green) 2-D input space. Moreover, the shape and the compositions of the two maps are quite similar (see Table 2): the percentage of samples falling in the mixed grey clusters differs by about 3% and the percentage of white clusters differs less than about 1%. Hence, it is expected that the two maps have quite similar performance when used as disruption predictors, as it will be shown in the next section.

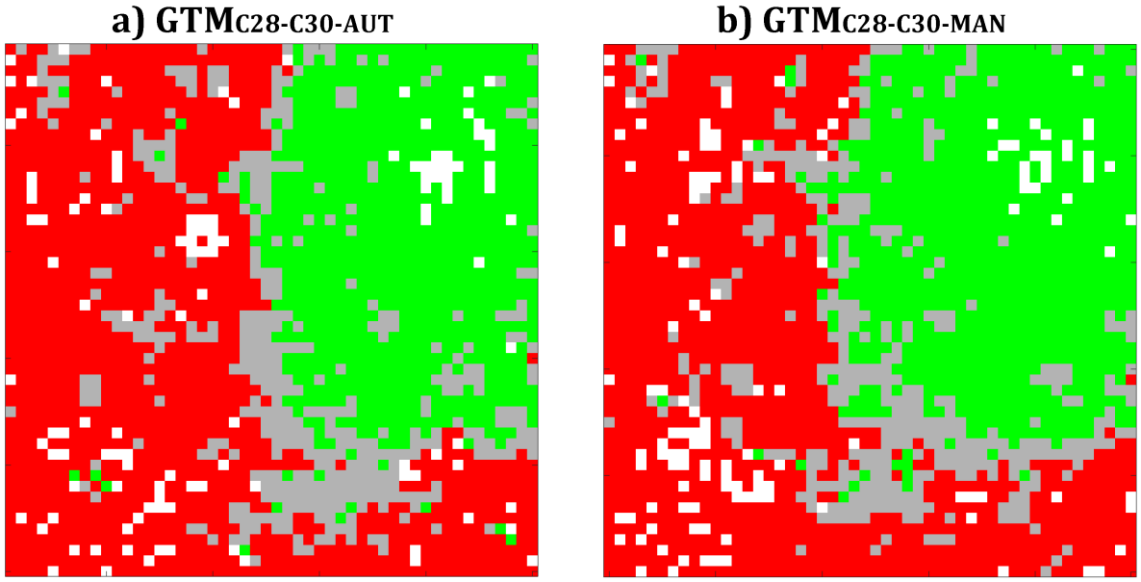


**Figure 12.** a) $GTM_{C28\text{-}C30\text{-}AUT}$ of the 6 plasma dimensionless parameters obtained using $T_{i\text{-}AUT}$ to determine the pre-disruptive samples; b) $GTM_{C28\text{-}C30\text{-}MAN}$ of the same parameters obtained using $T_{i\text{-}MAN}$. The maps are colored on the basis of the node composition: the green clusters contain only samples coming from regularly terminated pulses, the red clusters contain only samples coming from the pre-disruptive phase of the disrupted discharges, whereas grey mixed clusters contain both non-disruptive and disruptive samples. The white clusters are empty.

**Table 2**. GTMs composition (using $T_{i\text{-}AUT}$ and $T_{i\text{-}MAN}$)

| GTM | % safe samples belonging to safe (green) clusters | % disr. samples belonging to disr. (red) clusters | % samples in the grey clusters | % empty clusters |
|---|---|---|---|---|
| $GTM_{C28\text{-}C30\text{-}AUT}$ | 75.25 | 81.51 | 21.47 | 4.92 |
| $GTM_{C28\text{-}C30\text{-}MAN}$ | 79.17 | 84.74 | 18.12 | 5.52 |

## 5. Algorithm validation and results

Figure 13 reports the cumulative warning time, that is the difference between disruption time and the reference warning time $T_{i\text{-}MAN}$ (in black) and $T_{i\text{-}AUT}$ (in magenta). As can be noted, they follow quite the same trend confirming the validity of the proposed algorithm. Note that, in the construction of the algorithm, the warning times $T_{i\text{-}MAN}$ have not been used. They were considered only as benchmarks values to evaluate the performance of the algorithm.

The same Figure 13 reports the cumulative alarm time provided by the two GTMs in Figure 12 when used as disruption predictors on the entire C28-C30 data set adopting the same multiple condition alarm scheme in [13], shown in Figure 14. The cumulative alarm time distribution reports the fraction of the shots that has an alarm time larger than a selected value. In particular, Figure 13 reports, in blue, the cumulative alarm time provided by the GTM trained with the manually detected warning times $T_{i\text{-}MAN}$ and, in orange (dashed), the alarms obtained using $T_{i\text{-}AUT}$. The cumulative alarm times are almost overlapping with comparable prediction performance: the GTM trained with $T_{i\text{-}AUT}$ presents one missed alarm (0.7%), one tardy detection (a detection is considered tardy if the warning time is less than 10 ms), and 3 false alarms (2.6%) on the entire dataset, whereas the GTM trained with $T_{i\text{-}MAN}$ has one missed alarm, one tardy detection and 6 false alarms (6%) on the same dataset. Note that, the slight differences with the results in [13] are due to a slightly different definition of the peaking factors and a different choice of the parameters in the alarm scheme.

Figure 13 reports also the cumulative Locked Mode time, evaluated as the difference between disruption time and Locked mode onset time (in green). Note that, the alarm time is well in advance with respect to the time needed by the disruption mitigation valve (DMV, highlighted with a red vertical dashed line in Figure 13) to

intervene, with more than 55% of the discharges predicted more than 1 second before the disruption time. Furthermore, very often, the proposed predictor is able to activate an alarm well in advance with respect to the Locked Mode trigger.

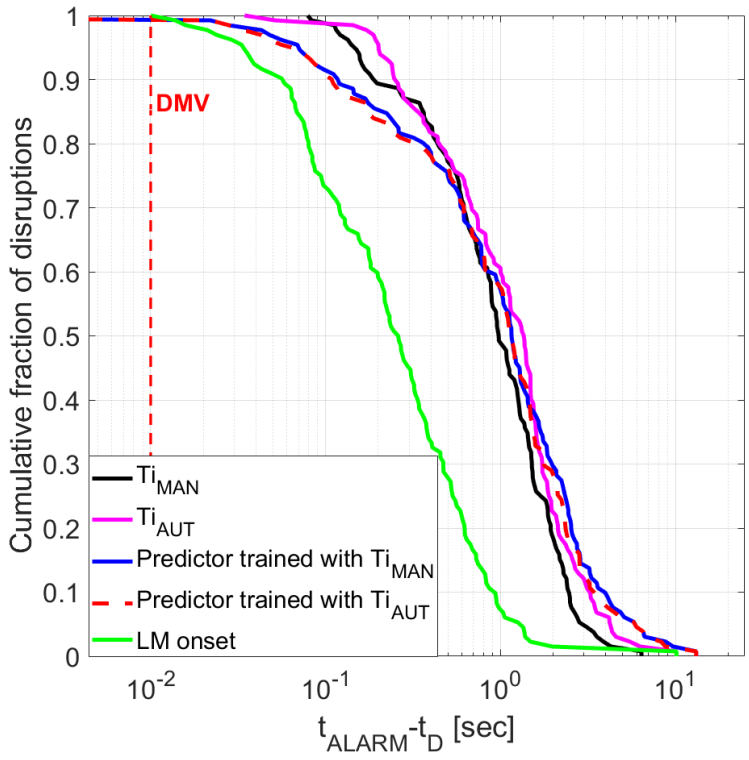


**Figure 13.** Cumulative warning time distributions for all the disrupted discharges in the training and test set of C28-C30 data set (the red vertical dashed line points out the DMV time, which allows to identify tardy detections).

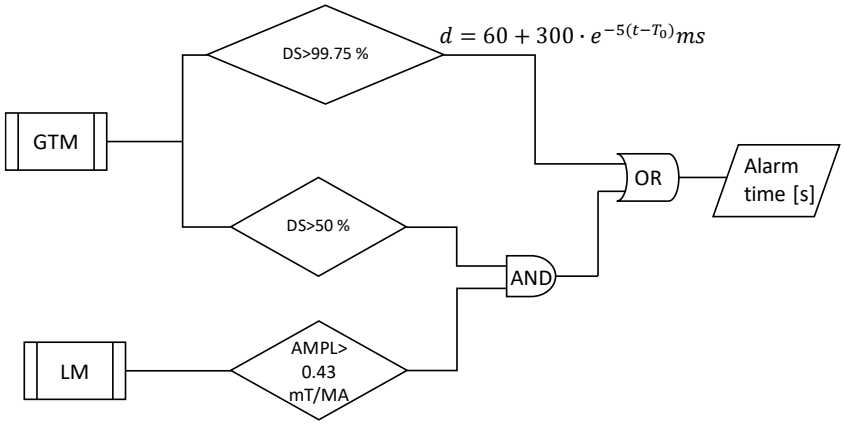


**Figure 14.** Multiple condition alarm scheme of the disruption predictor [13]. *DS* is the percentage of disrupted samples in the cluster where the discharge trajectory stays for at least $d$ consecutive milliseconds ($d$ is the assertion time). $T_0$ is the starting point of the flat-top.

The generalization capability of $GTM_{C28\text{-}C30\text{-}AUT}$ as disruption predictor has been evaluated on the C36 data set by projecting the 29 disrupted and 41 regular terminated discharges on the map. As expected, the prediction performance deteriorates with about 86% correct disruption predictions (1 missed alarm and 2 tardy detections) and 12% false alarms. Note that, 3 of the 5 false alarms are triggered by an abnormal increase of $P_{FRAC}$ due to interruption of the additional heating system and could be avoided by inhibiting GTM response when this event occurs. On the other hand, we did not observe any premature detection of disrupted discharges generated by this issue. For the two tardy detections, it is observed a very late locked mode as disruption cause.

The deterioration is commonly observed in whatever data-based model, and so in the present case, due to the variation of the GTM input parameters in the more recent campaigns. Figure 15 reports the probability density functions of the input plasma parameters of the regularly terminated discharges in C28-C30 (blue) versus those in C36 (green) data sets. From Figure 15 it can be seen that, even if the ranges of variation of the 6 considered input parameters are not so different, their distributions are quite different, especially for what concern the peaking factor of the radiation at the divertor ($Rad_{pf\_XDIV}$). Hence the need to regularly update the GTM model, when the 6-D input parameter space changes. To automatize the update, the automatic identification of the reference warning time *Ti* is mandatory.

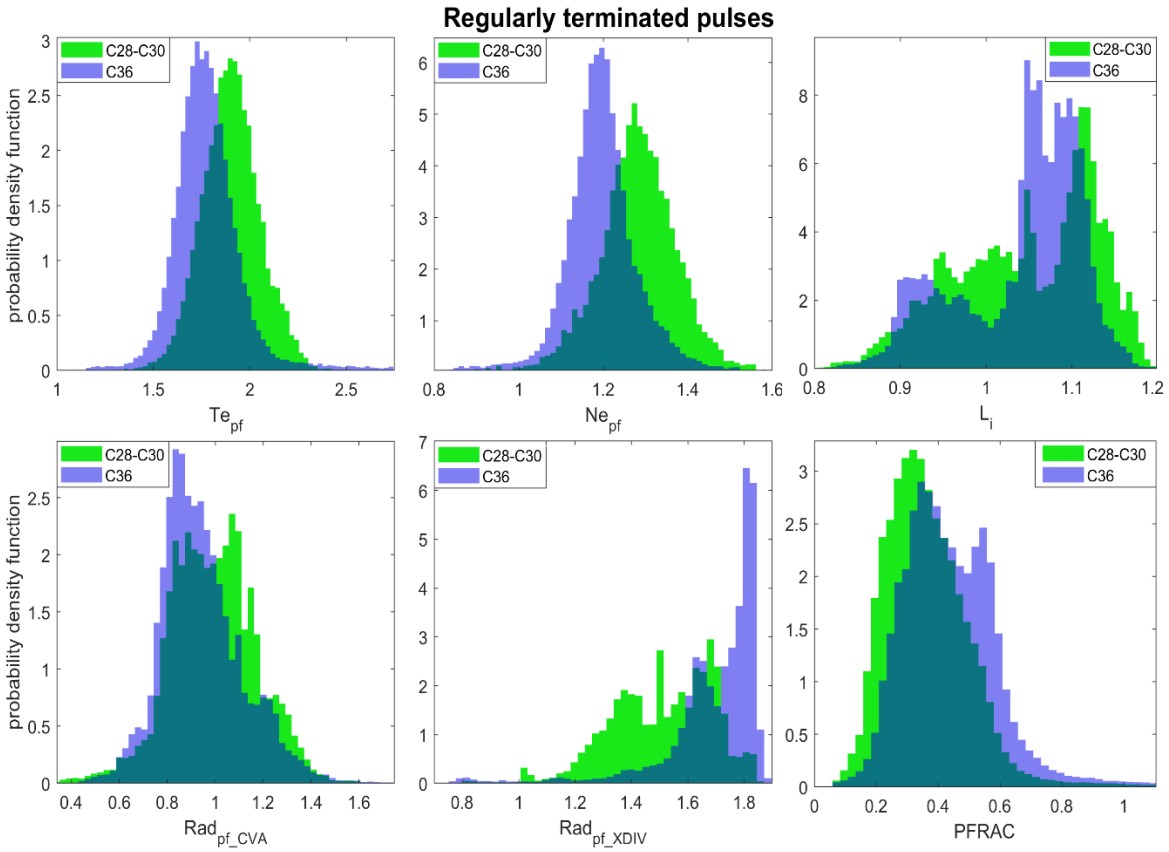


**Figure 15.** Probability density functions of the parameters of the regularly terminated discharges in C28-C30 (blue) versus those in C36 (green) data sets for (from top left to bottom right): electron temperature peaking factor, electron density peaking factor, internal inductance, radiation at the core peaking factor, radiation at the edge peaking factor, fraction of radiated power.

To this purpose, the warning times can be evaluated using the proposed algorithm avoiding the complex and time-consuming manual analysis.

To confirm the robustness of the algorithm for automatically determining the warning times, a statistical analysis of the selected plasma parameters has been performed on the discharges in the C36 dataset. Figure 16 reports the *pdf* of the selected parameters for the regularly terminated pulses (blue) versus the non-disrupted phase of the disruptive pulses (red), whereas Figure 17 reports the *pdf* of non-disruptive pulses (blue) versus the pre-disruptive phase of the disrupted pulses (red). Looking at Figure 17, it can be seen that, similarly to what observed in Figure 6, the *pdfs* of the pre-disruptive phases of the disrupted discharges shift with respect to the non-disrupted phases. The orange arrows in Figure 17 highlight the shifts. Hence, the previously proposed algorithm has been used to evaluate the warning times, $T_{i\text{-}AUT}$, in the disrupted discharges of the C36 dataset.

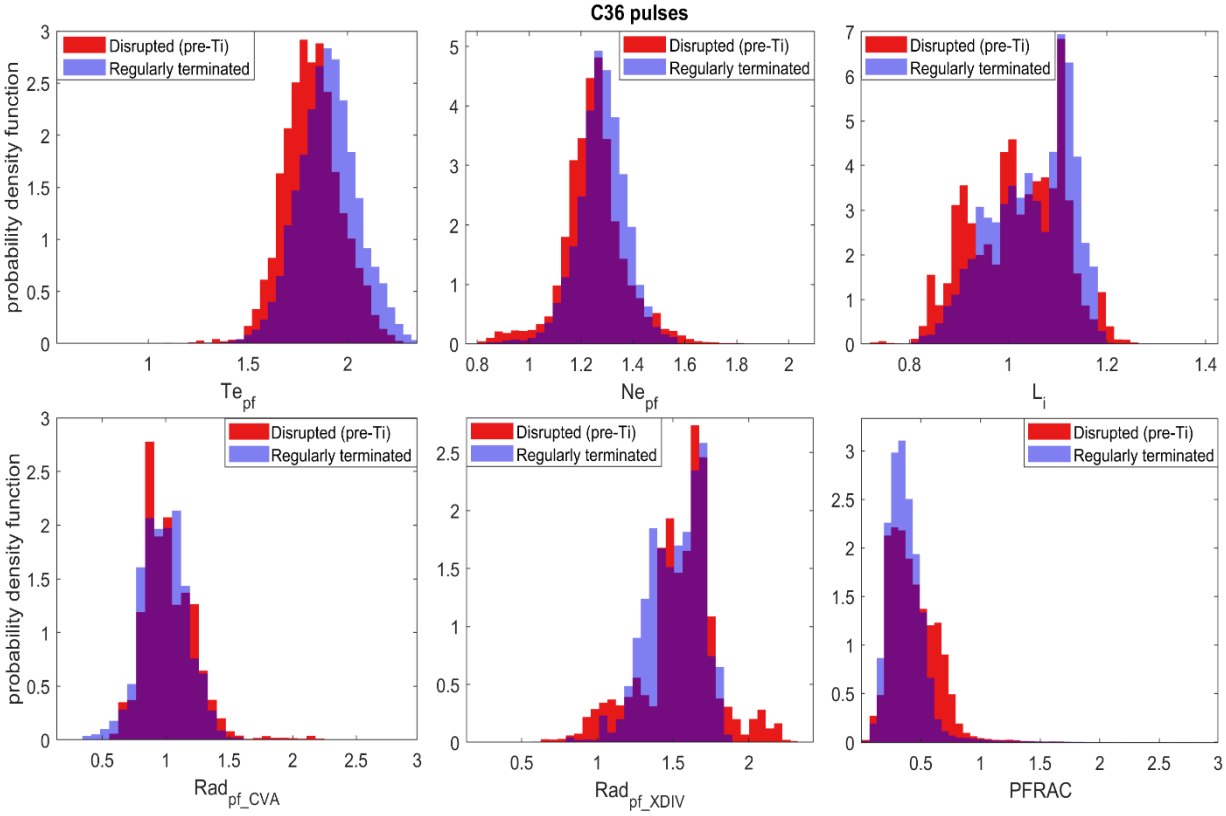


**Figure 16.** C36 data set: Probability density functions of the parameters of the regularly terminated pulses (blue) versus the non-disrupted phase (selected with the Ti-$_{AUT}$) of the disruptive pulses (red) for (from top left to bottom right): electron temperature peaking factor, electron density peaking factor, internal inductance, radiation at the core peaking factor, radiation at the edge peaking factor, fraction of radiated power.

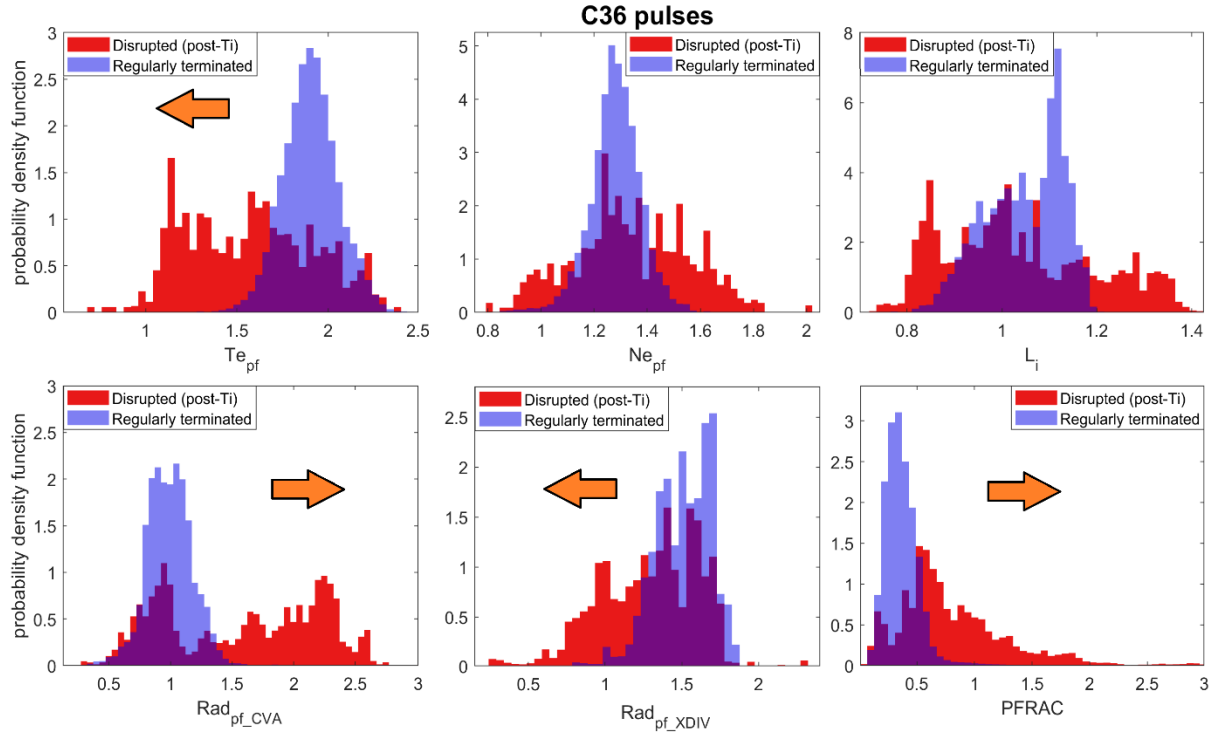


**Figure 17.** C36 data set: Probability density functions of the parameters of the regularly terminated pulses (blue) versus the pre-disruptive phase (selected with the Ti-$_{AUT}$) of the disrupted pulses (red) for (from top left to bottom right): electron temperature peaking factor, electron density peaking factor, internal inductance, radiation at the core peaking factor,

radiation at the edge peaking factor, fraction of radiated power. The shift of the distributions is marked with an orange arrow.

To validate the obtained $T_{i\text{-}AUT}$, a new GTM ($GTM_{C36\text{-}AUT}$) has been trained using all the pulses in the C36 data set, except two disruptions where the $T_{i\text{-}AUT}$ was not detected by the proposed algorithm.
Note that, being the GTM an unsupervised algorithm, the data are mapped only exploiting their intrinsic properties. The optimal GTM hyperparameters are the following: number of latent points = 1024, number of radial basis functions = 784, variance $\sigma$ = 1.2.
A very important measure of the map properties is the Unified distance matrix (U-matrix) [25]. This matrix, a standard way of representing the Self-Organizing Maps, visualizes the Euclidean distance among adjacent clusters of the map by using different shades of grey. A darker shading between clusters corresponds to a large distance in the input space while a lighter shading indicates a proximity. Light regions, therefore, can be considered as macro-clusters of input data while dark areas as separators between these macro-clusters. This representation allows you to locate macro clusters without having a priori information.
Figure 18 a) reports the U-matrix representation of the GTM of the C36 dataset where a clear dark boundary between two lighter macro-clusters can be identified (highlighted with a black dashed line). Using the automatically evaluated warning times, the GTM has been colored on the basis of the node composition and shown in Figure 18 b). From Figures 18 a) and 18 b), it can be noted that the boundary in the U-matrix is very similar to the boundary between the green (safe) and the red (disrupted) regions. Moreover, the map performs a clear separation of the safe and disrupted regions with very high discrimination capability as reported in Table 3.

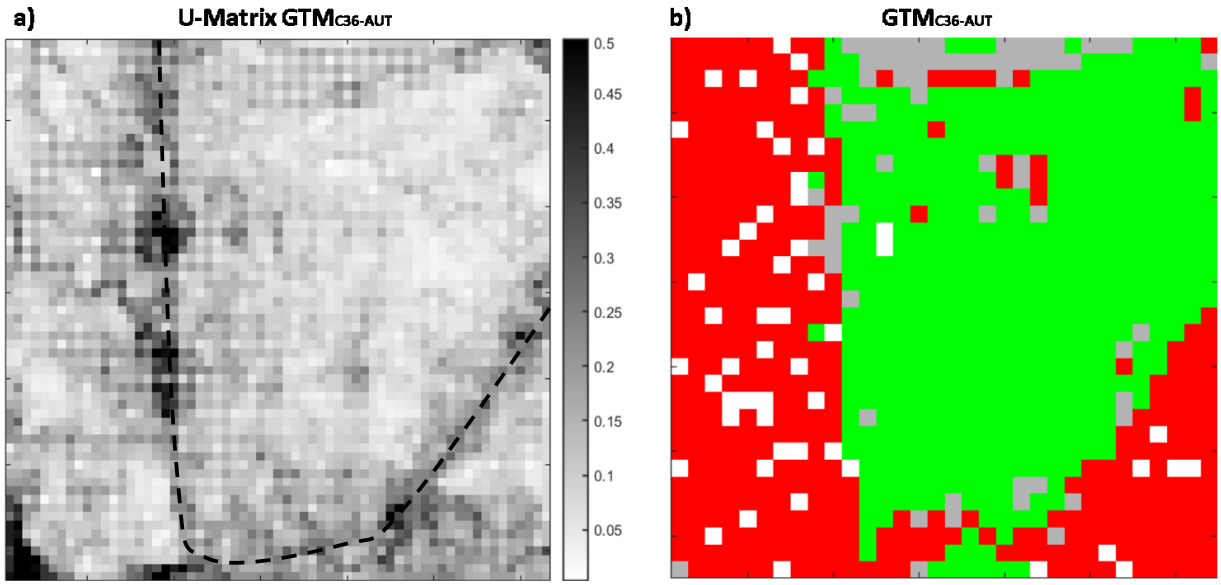


**Figure 18.** a) U-matrix of the $GTM_{C36\text{-}AUT}$. Lighter colors indicate smaller distance between clusters, while darker colors indicate higher distances. b) $GTM_{C36\text{-}AUT}$ obtained coloring the clusters using the automatically evaluated warning times $T_{i\text{-}AUT}$.

**Table 3**. $GTM_{C36\text{-}AUT}$ composition (using $T_{i\text{-}AUT}$)

| GTM | % safe samples belonging to safe (green) clusters | % disr. samples belonging to disr. (red) clusters | % samples in the grey clusters | % empty clusters |
|---|---|---|---|---|
| **$GTM_{C36\text{-}AUT}$** | 96.45 | 86.34 | 8.06 | 5.08 |

Due to the limited number of discharges in the C36 dataset, it was not possible to build wide independent test set for this map.

As an example, Figure 19 reports the temporal evolution of the disrupted discharge #90346, not used to train the $GTM_{C36\text{-}AUT}$: a) red (green) disrupted (non-disrupted) class membership function, which represents the percentage of samples of the disrupted and non-disrupted class respectively, in the cluster to which the sample is associated, with respect to the total number of samples in the cluster itself; b) trajectory of the discharge on the map. The circles depicting the evolution in time of the operating point are colored depending on the evolution time. The starting point is green, then the point becomes darker and darker as the discharge is approaching to the final point in red; c) Time evolution of the 6 plasma dimensionless parameters, together with the plasma current and the locked mode; the $GTM_{C36\text{-}AUT}$ alarm is marked with a vertical purple dashed line, the blue dashed line marks the mode lock time and the red dashed line marks the disruption time $t_D$. The disruptive discharge starts in a non-disruptive cluster, firstly evolving in the non-disrupted (green) region, enters the disruptive (red) region, returns in the green region and enters, at the very end, in a disruptive cluster, which corresponds to the disruption time. For the considered discharge, the GTM identifies, according to what observed during the experimental session, an impurity accumulation pattern well in advance to the disruption time and triggers the alarm. Moreover, the trajectory on the map highlights the observed subsequent stable phase followed by a very fast disruption due to a mode lock.

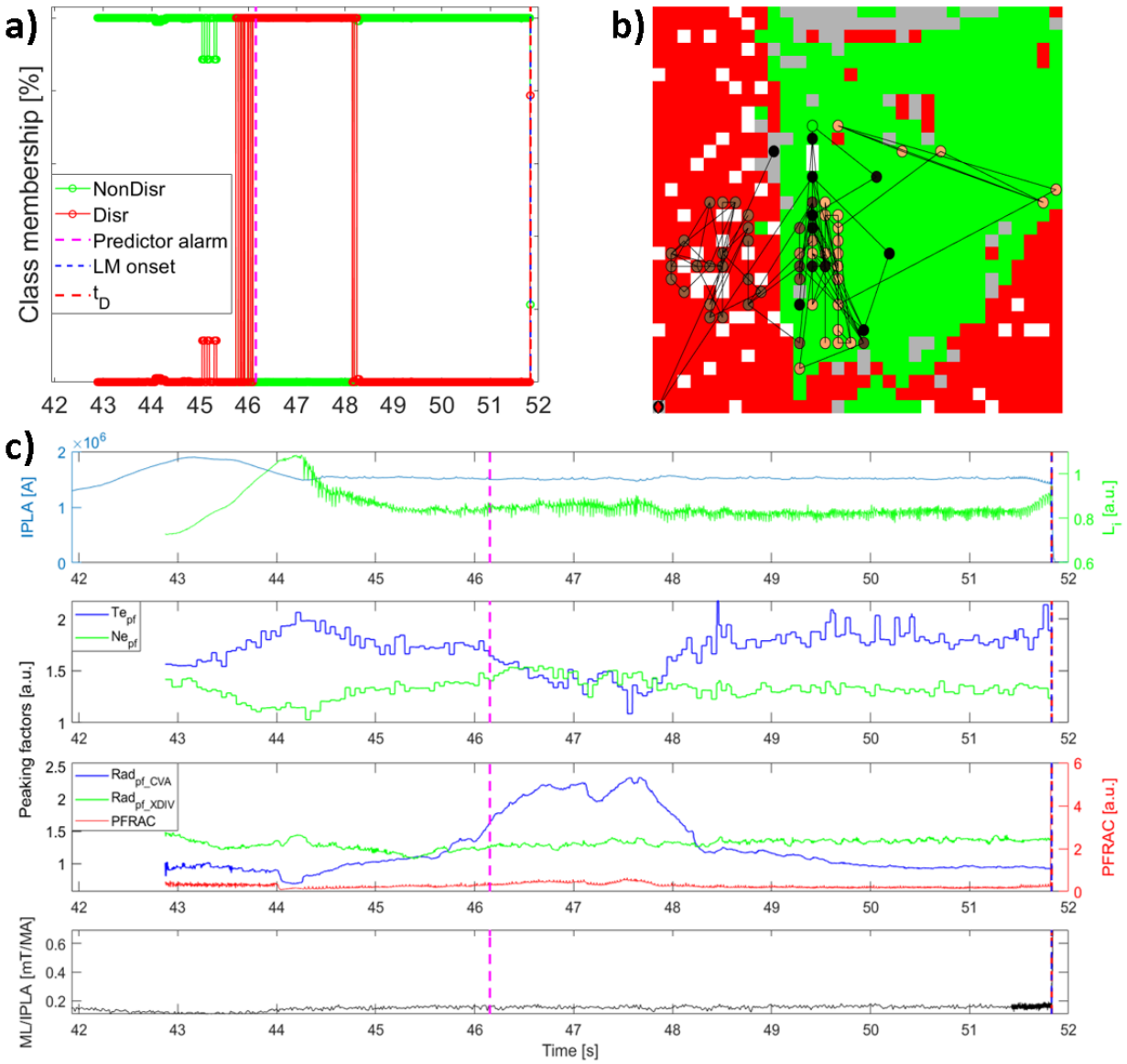


**Figure 19.** Disrupted discharge #90346: a) Membership function of the of non-disrupted (green) and disrupted (red) classes; b) Projection on the map; the lighter points correspond to the beginning of the discharge, whereas the darker one corresponds to the end, at the disruption time $t_D$; c) Time evolution of the 6 plasma dimensionless parameters, together

with the plasma current and the locked mode: the $GTM_{C36\text{-}AUT}$ alarm, corresponding to an impurity influx, is marked with a vertical magenta dashed line, the blue dashed line marks the mode lock time and the red dashed line marks the disruption time $t_D$.

All the presented results confirm the validity of the algorithm proposed for the evaluation of the warning times, mandatory for the updating of the model. The strategy to continuous learning the model is all but trivial and is out of the scope of the present paper.

## 6. Conclusions

In this paper an algorithm for the automatic identification of the pre-disruptive phase of tokamak discharges has been proposed. This work is framed in the complex and broad field of disruption prediction and classification; the field addresses the issues related to the integrity preservation of the tokamaks and to the better understanding of the physical mechanisms which destabilize the plasma. Presently, a general physical model for clearly recognizing disruptive behavior does not exist, and this sometimes produces ambiguity on the manual classification task as well. Hence, the interest is not only towards the classification task (as a plethora of different models exist, and many of them provide satisfying performance) but also in the properties of the parameter space where the relevant disruption physics takes place, its visualization and interpretative analysis. The challenge of the understanding of very complex high dimensional spaces led researchers to the use of manifold learning techniques such as Self-Organizing Map and Generative Topographic Mapping. Especially with the latter, the encouraging results led to the application of the method in a real-time framework. On the other hand, these models are trained using manually labelled data, which is necessary for the training step. The label identifies the reference warning time, the moment when the final chain of events destabilizes the plasma. The use of these times allows the machine learning methods to compare the regular terminated discharges with the pre-disruptive phase of the disruptive ones: the use of information inherent to the non-disrupted evolution would introduce uncertainty in the model. The manual identification of the warning times is very time consuming and complicated; it can also be uncertain due to the possible interplay of many different mechanisms. In this context, using a set of features, synthetized to detect some of the main known disruption precursors in fusion experiments, an algorithm for the automatic identification of the warning times has been developed and tested. The algorithm is based on the use of similarity measures between distributions, and it weights the contribution of each input feature to construct a *Warning Time Indicator*. The study of the *WTI* distribution in the regular discharges allows to optimize a coherent threshold value for the identification of the warning times.

The encouraging results led to the use of the automatic warning times as the new inputs of the GTM algorithm, in place of the manually detected ones. The shape and the composition of the GTMs trained with the manual and the automatic ones were comparable, as well as the data distribution obtained with the mapping and univariate analysis of the signals.

The results obtained with the GTM confirm the efficacy of the method and validate the proposed algorithm. The general principle of the algorithm seemed to work quite well, leading to a coherent discrimination of the non-disrupted and pre-disruptive phases of discharges, also referring to more recent experimental campaigns. The Machine Learning models generally suffer from ageing whether the input parameter space of the machine changes, and this is also valid for different experimental campaigns, where the operational scenarios can be different. The presented results, together with the map composition, confirms the possibility to complement effectively the cumbersome and time-consuming identification of off-normal states in the evolution of disruption discharges with the objective of implementing continuous learning in a binary classification scheme. Control systems need to be informed about the occurrence of specific events, in order to map an off-normal state into a corresponding reaction to avoid a disruption. In this respect, a more detailed analysis is still required to characterize how different events chains develop during the pre-disruptive phase. Nevertheless, being able to identify an off-normal condition represents decidedly a step forward in this direction.

Hence, in future works, this tool, together with a set of data analysis and clustering algorithms, could help in finding fundamental differences in the input parameters spaces, retraining the models and synthetizing more general features or indicators, to limit the performance degradation of the models.

## Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium and received funding from the EURATOM research and training programme 2014–2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

## References

[1] B. Cannas, A. Fanni, P. Sonato, K. Zedda, (2007), " (Cannas, Fanni, Sonato, & Zedda, 2007)", Nuclear Fusion, 47, 11, 1559-1569.

[2] B. Cannas, A. Fanni, G. Pautasso, G. Sias, P. Sonato, (2010), "An adaptive real-time disruption predictor for ASDEX upgrade", Nuclear Fusion, 50, 7, 075004.

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

[4] B. Cannas, A. Fanni, G. Pautasso, G. Sias, (2011) "Disruption prediction with adaptive neural networks for ASDEX Upgrade", Fusion Engineering and Design, 86, 6-8, 1039-1044.

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

[6] R. Aledda, B. Cannas, A. Fanni, A. Pau, G. Sias, (2015) "Improvements in disruption prediction at ASDEX Upgrade", Fusion Engineering and Design, 96-97, 1 698-702.

[7] W. Zheng et al (2018), "Hybrid neural network for density limit disruption prediction and avoidance on J-TEXT tokamak", Nuclear Fusion, 58, 056016.

[8] C. Rea, et al. (2018) "Disruption prediction investigations using Machine Learning tools on DIII-D and Alcator C-Mod", Plasma Physics and Controlled Fusion, 60, 084004 (13pp).

[9] K. J. Montes, et al. (2019) "Machine learning for disruption warning on Alcator C-Mod, DIII-D, and EAST", Nucl. Fusion, in press, https://doi.org/10.1088/1741-4326/ab1df4.

[10] B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias, (2013) "Automatic disruption classification based on manifold learning for real-time applications on JET", Nuclear Fusion, 53, 9, 093023.

[11] Murari A. et al (2013) "Clustering based on the geodesic distance on Gaussian manifolds for the automatic classification of disruptions", Nuclear Fusion, 53, 033006.

[12] B. Cannas, P. De Vries, A. Fanni, A. Murari, A. Pau, G. Sias, (2015), "Automatic disruption classification in JET with the ITER-like wall", Plasma Physics and Controlled Fusion, 57, 12, 125003.

[13] A. Pau, A. Fanni, S. Carcangiu, B. Cannas, G. Sias, A. Murari, F. Rimini and JET Contributors, "A Machine Learning approach based on Generative topographic mapping for disruption prevention and avoidance at JET," Nucl. Fusion 59 (2019) 106017 (22pp).

[14] R. Coelho et al., "Synthetic Diagnostics in the European Union Integrated Tokamak Modelling," Fusion Science and Technology, vol. 63, no. 1, pp. 1-8, 2013.

[15] J. Kates-Harbeck, A. Svyatkovskiy, and W. Tang (2019) “Predicting disruptive instabilities in controlled fusion plasmas through deep learning”, Nature – Letter Research, 568, (18pp).

[16] F. Matos, D. Ferreira, P. Carvalho and JET Contributors, “Deep learning for plasma tomography using the bolometer system at JET,” Fusion Engineering and Design, vol. 114, pp. 18-25, Volume , January 2017, pp. 18-25.

[17] A. Pau, A. Fanni, B. Cannas, S. Carcangiu, G. Pisano, G. Sias, P. Sparapani, M. Baruzzo, A. Murari, F. Rimini, M. Tsalas, P.C. de Vries, (2018), “A first analysis of JET plasma profile-based indicators for disruption prediction and avoidance”, IEEE Transactions on Plasma Science, DOI:10.1109/TPS.2018.2841394.

[18] C. Sozzi, et al., “Early identification of disruption paths for prevention and avoidance”, 27th IAEA Fusion Energy Conference (FEC 2018), 22-27 October 2018, Ahmedabad, India, pp. 1-8, https://conferences.iaea.org/indico/event/151/papers/6273/files/4867-sozzi_paper_IAEA2018_v6.pdf .

[19] A. Pau, B. Cannas, A. Fanni, G. Sias, M. Baruzzo, A. Murari, G. Pautasso, M. Tsalas, (2017) “A tool to support the construction of reliable disruption databases”, Fusion Engineering and Design, 125, 139-153.

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

[21] L. Barrera et al. (2010), “Inboard and outboard electron temperature profile measurements in JET using ECE diagnostics”, Plasma Physics and Controlled Fusion, 52 085010.

[22] A. Huber, et al. (2007), “Upgraded bolometer system on JET for improved radiation measurements”, Fusion Engineering and Design, 82, 1327–1334.

[23] P. C. de Vries, et al. (2012) “The impact of the ITER-like wall at JET on disruptions”, Plasma Physics and Controlled Fusion, 54, 124032 (9pp).

[24] S.-H. Cha, “Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions,” International Journal of Mathematical Models and Methods in Applied Sciences, pp. 300-307, 2007.

[25] A. Ultsch and H.P. Siemon. Kohonen's self organizing feature maps for exploratory data analysis. In Proc. INNC'90, Int. Neural Network Conf., pages 305-308, Dordrecht, Netherlands, 1990. Kluwer.