PAPER • OPEN ACCESS

A hybrid physics/data-driven logic to detect, classify, and predict anomalies and disruptions in tokamak plasmas

View the article online for updates and enhancements.

A hybrid physics/data-driven logic to detect, classify, and predict anomalies and disruptions in tokamak plasmas

R. Rossi[1,*] ⬤, M. Gelfusa[1] ⬤, T. Craciunescu[2] ⬤, I. Wyss[1] ⬤, J. Vega[3] ⬤ and A. Murari[4,5] ⬤ on behalf of JET Contributors[a]

EUROfusion Consortium, JET, Culham Science Centre, Abingdon OX14 3DB, United Kingdom of Great Britain and Northern Ireland

[1] Department of Industrial Engineering, University of Rome 'Tor Vergata', via del Politecnico 1, Roma, Italy
[2] National Institute for Laser, Plasma and Radiation Physics, Magurele-Bucharest, Romania
[3] Laboratorio Nacional de Fusión, CIEMAT. Av. Complutense 40, 28040 Madrid, Spain
[4] Consorzio RFX (CNR, ENEA, INFN, Università di Padova, Acciaierie Venete SpA), Corso Stati Uniti 4, 35127 Padova, Italy
[5] Istituto per la Scienza e la Tecnologia dei Plasmi, CNR, Padova, Italy

E-mail: r.rossi@ing.uniroma2.it

CrossMark

Abstract

Disruptions are abrupt collapses of the configuration that have afflicted all tokamaks ever operated. Reliable observers are a prerequisite to the definition and the deployment of any realistic strategy of countermeasures to avoid or mitigate disruptions. Lacking first principle models of the dynamics leading to disruptions, in the past decades empirical predictors have been extensively studied and some were even installed in JET real time network. Having been conceived as engineering tools, they were often very abstract. In this work, physics and data-driven methodologies are combined to identify the main macroscopic precursors of disruptions: magnetic instabilities, abnormal kinetic profiles and radiation patterns. Machine learning predictors utilising these observers can not only detect and classify these anomalies but also determine their probability of occurrence and estimate the time remaining before their onset. These tools have been applied to a database of about two thousand JET discharges with various isotopic compositions including DT, in conditions simulating in all respects real time deployment. Their performance would meet ITER requirements, and they are expected to be easily transferrable to larger devices, because they rely only on normalised quantities, form factors, and physical/empirical scaling laws.

[a] See Mailloux *et al* 2022 (https://doi.org/10.1088/1741-4326/ac47b4) for JET Contributors.
[*] Author to whom any correspondence should be addressed.

1. A data-driven, physics-based approach to prediction for proximity control

The sudden and violent collapses of the configuration called disruptions have proved to be unavoidable in all tokamaks ever operate in the world [1, 2]. Devices with metallic plasma facing components, including JET with a metallic wall, seem to be particularly vulnerable to this extreme form of global instability [3, 4]. Indeed, as will be shown in more detail later, the high current campaigns, leading to and including the last DT1, have been affected by completely unacceptable disruptivity levels, both in the baseline and hybrid scenario. Indeed the maximum tolerable percentage of disruptions in ITER at full current is certainly not higher than 5% [5]: in DEMO even a single full power non mitigated disruption can cause irreparable damage to the device [6]. Given the potential harm that these events can cause in the next generation of machines, devising a robust strategy to handle them is a top priority of the international community. In this perspective, enormous amounts of resources are being devoted to developing tools for the mitigation of disruptions' detrimental effects: electromagnetic forces, thermal loads and runaway electrons [7]. In addition to fast valves for massive gas injection, the technology of the shattered pellets has made very remarkable progress in the last decade [8–11].

One aspect not fully appreciated in the fight against disruptions is the fact that even the most effective remedial tools require a suitable trigger to be operated at the right moment and in the right circumstances. Moreover, an excessively cautionary use of mitigation techniques is not an option in future devices for at least two reasons. First, reliable operation should be guaranteed to maximise the scientific exploitation (and the continuity of the electricity production in the reactor). Secondly even mitigation actions entail a certain level of risk to the integrity of the machines and their use should be therefore minimised. Robust observers, for identifying the plasma state from the point of view of its proximity to the stability boundaries, are of course essential elements of any realistic strategy of control. Consequently, proximity detection and control are crucial aspects for the reactor perspective of the tokamak configuration.

Unfortunately, disruptions are a very complex phenomenon, characterised by the nonlinear interactions between multiple factors. The peaking of the electron density profile, the hollowing of the electron temperature profile, the transport of impurities, the pedestal properties, centrifugal forces, plasma rotation, local radiation emission are just some of the aspects, which can affect the magnetic fields and trigger the onset of macroscopic instabilities, finally leading to the collapse of the configuration [12–15]. The complexity of the plasma dynamics in the phase preceding the disruptions is such that no first principle models exist to reliably predicting their occurrence offline, let alone in real time (even if significant progress in modelling tokamak plasmas has been made in the last decades).

The lack of theory-based predictors has been compensated in the last years by the development of empirical classifiers. Those derived manually, by inspection of a limited amount of examples, have proved to have very poor predictive capabilities on JET with a metallic wall [16]. This fact has motivated the development of various predictors based on different machine learning techniques [17]. Disruption predictors for real time deployment are usually conceived as engineering binary classifiers in suitable multi-dimensional operational spaces, where each dimension corresponds to a different plasma quantity. The traditional training process is meant to divide the operational space into two regions (disruptive/non-disruptive), so that the classifier can determine the boundary between these two zones. This boundary is represented by an equation that model the relationship between the relevant quantities in the operational space. During the experiments, classifications are performed on a periodic basis with typical time resolution of a few ms.

In terms of artificial intelligence technologies, practically all the main machine learning threads have been tried to predict disruptions: artificial neural networks, support vector machines, fuzzy logic, generative topographic mapping, deep learning and even reinforcement learning [17]. With regard to real-time signal processing, the methods implemented have explored practically all known data analysis techniques for time series in the time domain, in the frequency domain, and in the combined time/frequency domains [18, 19]. Machine learning predictors with various technologies have been developed for all the following Tokamaks: TCV [20], ADITYA [21], AUG [19, 22–25], DIII-D [26–28], J-TEXT [29], NSTX [30], EAST [28, 31], ALCATOR C-MOD [27, 28], JT-60 U [32, 33] and JET [19, 34–44]. The advanced predictor of disruptions is the first disruption predictor that on JET obtained success rates >98%, false alarm rates <2% and average warning times of hundreds of ms [45]. It was also the first predictor, based on machine learning tools, implemented in the JET real time network. In the following years also SPAD and the predictor based on the centroid method were also installed in JET real time network [40, 46].

Notwithstanding their positive qualities, traditional supervised machine learning tools for disruption prediction are not completely satisfactory, when it comes to their deployment in future experimental reactors. Indeed, they require a lot of examples for the training, are very machine specific (not portable to other devices) and tend to age very quickly (their performances tend to degrade very rapidly with the evolution of the experimental programme). In the perspective of ITER,

these drawbacks are particularly severe for the following main reasons:

1. ITER will explore new operational regions (in both dimension and dimensionless parameters), and thus an algorithm trained with data of smaller devices (JET, AUG, etc) would have problems to maintain its performances in ITER experiments.
2. ITER plasmas will also evolve over time, starting at low current and energy, finally reaching the full performances ITER scenarios at 15 MA. This continuous evolution of the plasma operational regime implies that ITER disruption predictors would need to adapt to ITER changing circumstances.

In this work, a new approach to disruption prediction has been developed. It has been conceived with the specific aim of trying to address the aforementioned challenges. In this perspective, the predictors described in the rest of the present paper have the following characteristics:

1. *Hybrid physics and data-driven training.* A few works on how to scale anomaly indicators between machines have been published [47, 48]. Their scope is quite limited though. Moreover, at the beginning of operation, there is no specific data available and therefore data-driven predictors cannot be trained with experimental data of the new tokamak. In these situations, it is important to exploit to the maximum all the available knowledge, both theoretical and experimental. Consequently, in this work a hybrid physics/data-driven approach has been implemented. At the beginning of a new experimental campaign or device, the algorithm is trained with either physics models, obtained from theory, or with empirical equations, derived from previous tokamaks or campaigns. Then, during the evolution of the experimental campaigns and on a shot to shot basis, the training set is updated, and the algorithm is retrained. Note that with this approach, the predictor can work from the first pulse, without the need of any 'from scratch training' with data of the new device.
2. *Physics based indicators and anomaly classification.* In order to take the best measures to counter plasma anomalies potentially leading to disruptions, the identification of the plasma state it is clearly of crucial relevance. Contrary to most predictors in the literature, which are based on a black-box approaches, the observers developed in the context of the present work are based on specific and physically meaningful indicators. They allow a clear understanding of why an alarm has been raised, permitting the control system to take well informed decisions. These indicators, being based on clear physical quantities are also expected to be better transferable from one device to another. Some work in this direction was performed in the past but only for mitigation purposes [18, 19, 42, 46].
3. *Time-to-anomaly prediction.* The identification of an anomaly is not sufficient to take appropriate counteractions. Indeed, some countermeasures might require a minimum

time to have an effect on the plasma. Estimates of the interval remaining, before the onset of a major anomaly, would be extremely valuable to optimise the control strategies [49]. Therefore, the developed tools provide also time-to-anomaly estimates for some of the most relevant dangerous situations, which can occur in the discharge.

The developed predictors are meant to provide information useful for all the major classes of interventions envisaged to counteract disruptions: avoidance, prevention and mitigation. In the community avoidance indicates the actions of the control system to keep the plasma away from the stability boundaries, thus allowing the experiments to continue with their original programme. Preventions are the measures to be undertaken when there is not enough time for avoidance and the best strategy consists of terminating the discharge rapidly, to prevent the fatal instabilities from growing and causing a disruption. When a disruption has become unavoidable, the only remaining option is to activate the systems, such as massive gas injection or shattered pellet injectors, to mitigate its consequences.

With regard to the organisation of the paper, next section describes the main families of anomalies leading to disruptions, the indicators developed to detect them, and the solutions devised to estimate the time remaining before the onset of the various anomalies. The subject of section 3 are the techniques of adaptive learning required to keep pace with the changing experimental conditions. In section 4 the description of the database is provided, together with the criteria to assess the performance of the predictors. The results of the real time like application of the proposed methodology to a very large dataset of JET, including thousands of discharges, are provided in section 5. The conclusions and lines of future developments are discussed in the last section of the paper.

2. Physics-based indicators for real-time identification of anomalies and the estimation of the time to anomaly

As alluded to in the previous section, any successful control strategy must rely on the proper identification of the system to be controlled. The determination of the plasma state for the prediction of anomalies is a very delicate matter even for machines that have already produced a lot of data. However, even if the details of the plasma dynamics leading to disruptions are not fully understood, there is a quite general consensus about the main phases of the plasma evolution leading to a disruption. In metallic devices, the anomalies that typically occur earlier in the discharges are the ones involving the emitted radiation (due to impurity accumulation or excessively high density). As a consequence of specific radiation patterns or other causes, most disruptions are preceded by distortions of the kinetic profiles, mainly of the electron temperature (hollowness, edge cooling (EC)). In the last stages before the disruption, the previous problems cause the growth of magnetic modes, to a level at which the configuration cannot be sustained anymore.

Irrespective of the details and specificities of the plasma route toward a disruption, the experience of present-day devices indicates quite clearly that the most important measurements for prediction are diagnostics of the magnetic configuration quality, indicators of the temperature profile and quantifiers of the emitted radiation. The same quantities are also the most suited to determining the time remaining before the onset of the main macroscopic instabilities, potentially leading to disruptions. The solutions adopted for the observers, utilised in the rest of the work, are summarised in the following subsections.

2.1. Detecting anomalies in the magnetic configuration

The amplitude of the locked mode has traditionally been a workhorse for triggering mitigation actions in many devices and on JET in particular (see [16–18, 50] and references therein). In this work, the detection of anomalous locked mode signals is implemented with a new approach that combines physics understanding and experimental data. Indeed, at beginning of a new experimental campaign or of a new tokamak, there are not specific data and therefore the predictors must be trained with physical information provided by other tokamaks, other campaigns, theory and/or numerical information.

Large locked mode signals are correlated with the size of the magnetic islands due to small toroidal number (usually $n = 1$) macroscopic instabilities [1, 48]. It is now clear that when the velocity of these modes decreases in the laboratory frame of reference (finally locking to the wall), the amplitude of the radial perturbations increases exponentially and usually leads to a disruption. From a theoretical point of view, the threshold for the locked mode amplitude triggering a disruption is interpreted in terms of the Chirikov parameter [51]. However, in real time the correct computation of this quantity is quite challenging and in the last decades empirical models have been developed to understand what is the best mode locking threshold to be used in multimachine applications. A multimachine scaling in power law form for the mode locking, interpreting the various quantities in terms of the magnetic island maximum width, was formulated in [48]. Unfortunately, such a law does not perform very well for JET with the ITER like wall [47] and therefore alternatives have been proposed. In [47] a simple scaling law, derived with a machine learning based methodology, was proposed. It links the locked mode amplitude with the internal inductance. In this work, we have implemented this approach to find an empirical equation that correlates the normalised locked mode signal (the locked mode amplitude divided by the magnetic poloidal field at the geometrical minor radius, calculated through the plasma current, LM_N) with the internal inductance (l_i):

$$LM_{N,threshold} = a l_i^b. \tag{2.1}$$

The equation has been particularised using the same campaigns of the cited article [47] with exactly the same setup. To

this end, first the disruption probability is modelled with the sigmoid function reported in equation (2.2):

$$p_{LM} = e^{d_1(LM_N - LM_{N,\,threshold}) + d_2} \Big/ \left(1 + e^{d_1(LM_N - LM_{N,\,threshold}) + d_2}\right). \tag{2.2}$$

At the beginning of a new experimental campaign, no new data is available and therefore the locked mode predictor is trained to reproduce equation (2.2). The predictor consists of a simple feed forward neural network (3 hidden layers of 20 neurons each with tanh transfer functions plus one neuron output layer), which returns the probability (q_{LM}) of an anomalous locked mode amplitude. The cross-entropy loss function has been used to train the neural network:

$$loss_{physics} = -\sum p_{LM,1} \ln(q_{LM,1}) - \sum p_{LM,0} \ln(q_{LM,0}) \tag{2.3}$$

where the subscript 0 indicates a safe discharge, the subscript 1 a disruptive shot. Moreover, $p_{LM,0}$ is the expected probability that the mode locking is not anomalous (calculated as 1—p_{LM} and derived from equation (2.2); q_{LM} and $q_{LM,0}$ are the predicted probabilities (the neural network predicts q_{LM} and q_{LM0} is 1—q_{LM}).

Once the new device or campaign starts, pulse by pulse, the training set is meant to be updated (see section 3) and the neural network retrained adding a new data-driven term to the loss function:

$$loss = \frac{(loss_{physics} + \alpha \, loss_{data})}{1 + \alpha} \tag{2.4}$$

where:

$$loss_{data} = -\sum_{class=1} \ln(q_{LM,1}) - \sum_{class=0} \ln(q_{LM,0}). \tag{2.5}$$

And α is a weight parameter. In this work, α has been set to a constant value equal to 1. However, it is possible to adjust its value as a function of the amount of the training set data and their reliability.

In terms of interpretation, this predictor returns the probability that the time slice has an anomalous disruptive mode locked amplitude. Since the locking to the wall of the magnetic perturbation is the last event before a disruption, p_{LM} can also be conceived as the probability of the plasma disrupting. This is due to the fact that in general, once a macroscopic mode has locked to the wall, there is not enough time to undertake any other action except mitigation. As discussed in the next subsections, this correspondence, between probability of an anomaly and probability of disruption, does not apply to the outputs of the other predictors for avoidance and prevention.

2.2. The characterisation of the kinetic profiles

The two electron temperature anomalies considered in this work are EC and core hollowness. Both anomalies leave a clear signature in the electron temperature profiles. In the case of hollowness, the core electron temperature is not peaked in the centre but off-axis, becoming bimodal. In the case of EC, a

strong decrease of the peripheral temperature is observed [12, 41, 42, 52]. The two indicators proposed in [52] to quantify these anomalies are the ones used in this work. They are briefly described in the following and they are calculated using the data of JET high resolution Thomson scattering.

Regarding the electron temperature hollowness (ETH), the 'Gaussian fitted hollowness' indicator has been implemented. The indicator is derived by fitting a bimodal symmetric Gaussian function to the electron temperature profile. If the profile is unimodal (no hollowness), the separation of the two Gaussians is in principle zero. On the contrary, for hollow profiles, the separation of the two Gaussians increases with and it is directly proportional to the degree of hollowness. The separation is quantified by the Bhattacharya indicator particularised for symmetric Gaussians. Then, by using a sigmoid function, the probability to have an ETH is computed (the coefficients d_{1ETH} and d_{2ETH} have been calculated using numerical cases reported in the reference paper [52]):

$$p_{ETH} = e^{d_{1ETH}(ETH - ETH_{threshold}) + d_{2ETH}}$$
$$/ \left(1 + e^{d_{1ETH}(ETH - ETH_{threshold}) + d_{2ETH}} \right). \quad (2.6)$$

EC is detected using the cumulative distribution function (CDF) of the normalised electron temperature profile and by calculating an 'edge radius' ρ_{98} as the radius containing the 98% of the CDF. Then, the lowest is the radius, the more severe is the EC. Again, by using a sigmoid function, the probability of the plasma being in an EC dangerous situation can be computed (the coefficients d_{1EC} and d_{2EC} have been calculated using numerical cases shown in [52]):

$$p_{EC} = e^{d_{1EC}(EC - EC_{threshold}) + d_{2EC}} / \left(1 + e^{d_{1EC}(EC - EC_{threshold}) + d_{2EC}} \right). \quad (2.7)$$

To interpret equations (2.6) and (2.7) two comments are in place. First, it should be noticed that both anomalies deform the temperature profile independently from the absolute value of this quantity. The relevant information resides in the shape of the profile. Indeed, for example, what really matters for the control system is how hollow a temperature profile is and this can be well quantified by a simple form factor. Consequently, the detection of electron temperature anomalies does not require any machine learning method (no training set, etc), because the critical values can be determined on the basis of previous experiments and do not need to be adjusted during the campaigns. Secondly, the probabilities expressed by equations (2.6) and (2.7) have to be interpreted differently from the one of the mode locking, described in the previous subsection. Indeed, p_{ETH} and p_{EC} quantify how anomalous the temperature profile is, not directly the plasma tendency to disrupt. Indeed, typically the mode locked also must become anomalous before the plasma configuration collapses. As will be discussed in section 2.4, therefore p_{ETH} and p_{EC} can be utilised to provide information about the propensity of the plasma to developing macroscopic magnetic instabilities, not directly about the plasma propensity to disrupt.

2.3. Quantifying anomalies in the emitted radiation

Contrary to mode locking and electron temperature anomalies, radiation anomalies do not lend themselves easily to a straightforward interpretation for various reasons. First, the phenomena influencing the radiation emission in a tokamak are extremely complex. Consequently neither physical nor empirical equations of radiative disruptive thresholds have been found, which can be applied in real time and easily transferred from one device to another. Moreover, contrary to temperature hollowness or EC, radiation cannot simply be eliminated. Therefore, an alternative physics-guided semi-supervised methodology has been developed to detect anomalous radiation patterns.

Another difficulty resides in the fact that the plasma radiation emission in tokamaks is usually estimated with the inversion of bolometric signals, utilising quite sophisticated tomographic algorithms. This approach allows obtaining high spatial resolution information and the advanced method based on the maximum likelihood can also associate confidence intervals to the results [53–56]. Unfortunately, the available tomographic algorithms are too time consuming and their deployment in real-time is, at the moment, unrealistic [57]. The workaround solution of calculating radiation peaking factors, as developed in [42], can be useful but, their interpretation can be delicate in some situations and they lack sufficient spatial resolution to identify all the potentially disruptive emission patterns detected on JET. A systematic analysis of the peaking factor limitations can be found in [58]. A possible alternative to estimating the local emissivity is the use of deep learning [59, 60]. However, being these methodologies based on supervised techniques, they may not return reliable results in anomalous situations. Moreover, transfer learning to new tokamaks should be developed to render this technology applicable from day 0 of a new device.

The alternative adopted in the present work consists of trading space resolution for time resolution. The emission is calculated for macro-pixels derived by macro-views, obtained by summing the corresponding lines of sight of the bolometers [58, 61]. Given the topology of JET bolometric cameras, it is possible to divide the projections into three vertical and three horizontal macro views, whose intersections identify the eight regions reported in figure 1. Obtaining the emission from these macro/pixels requires solving a system of six equations with eight degrees of freedom. By imposing that the emissivity must be positive, this system of equations can be solved by making recourse to non-negative least square minimisation methods [62]. Given the types of emissivities encountered on JET, the iterative routine implemented always converges in less than 50 μs, more than adequate since the cycle time of JET real time network is 2 ms. The absolute values of the radiation emission from the macro-pixels have been checked with the most accurate tomographic method available, the one based on the maximum likelihood [55]. Since it has been found that the average error is around 15%–20% in most cases, the proposed simplified inversion method has the spatial resolution, the time
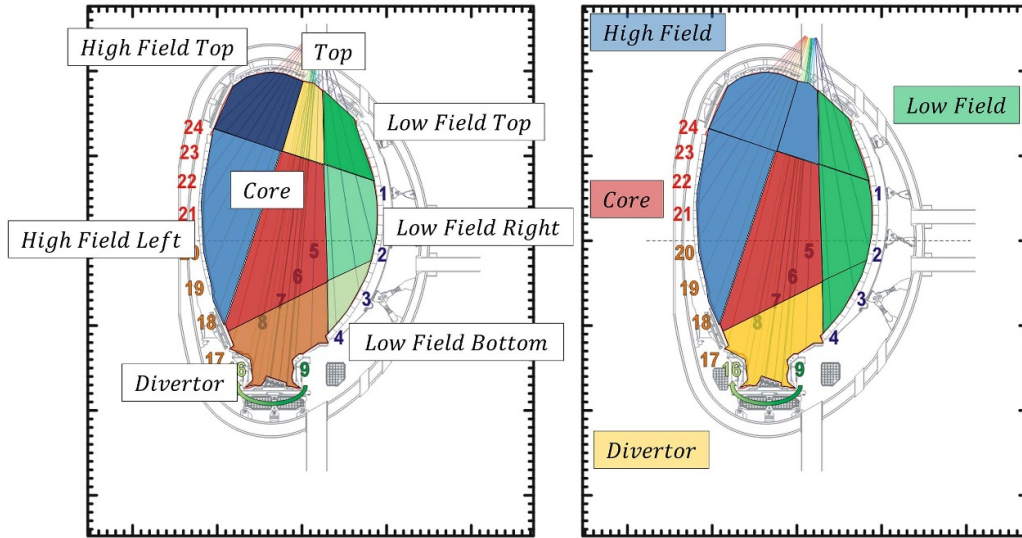
Figure 1. The macro-pixels resulting from intersecting the macro-views described in section 2.3 (left) and the four macro-regions to detect radiation anomalies (right).

resolution and the accuracy to detect all the most important radiation anomalies potentially leading to disruptions on JET.

Then, the eight macro-pixels are used to calculate the radiation in four macro-regions of interest:

- Core: this region is exactly the one shown in figure 1 left. This region allows detecting anomalous core radiation due to heavy impurity accumulation and is usually correlated with the onset of ETH.
- Divertor (Div): also this region is exactly the one shown in figure 1 left. This region allows evaluating excessive radiation in the region of the divertor.
- High-field: calculated as the sum of high-field left, high-field top and top. This region is used to detect MARFE and MARFE-like radiation patterns, phenomena that sometimes precede electron temperature edge reductions at the edge.
- Low-field: calculated as the sum of low-field bottom, low-field right and low-field top, this region is useful to monitor the radiation emitted by the frequently formed radiative blobs on the low-field side, which sometimes precede core radiation and hollowness or EC.

To predict the onset of radiation anomalies, the simple emission of each macro-pixel is not enough though. The really relevant quantity is the ratio of the radiated power divided by the local internal energy:

$$\Lambda_i = \frac{P_{\mathrm{rad}}\left(\rho,\theta,t\right)}{E_p\left(\rho,\theta,t\right)} \tag{2.8}$$

where ρ and θ are the coordinates of the corresponding macro-pixel barycentre. The Λ_i quantities have the units of s^{-1} and can therefore be interpreted as the inverse of cooling factors. How to calculate the denominator of equation (2.8), the average plasma energy of a macro-pixel, is explained in appendix A.

On a shot-to-shot basis, the training sets of 'safe' and 'anomalous' plasmas are updated and the cooling factors recalculated (see training set update logic section). However, a typical classification algorithm of classification cannot be used because:

1. The anomalous training set would be very unbalanced, containing a lot (most) of time slices when the plasma is not anomalous.
2. The safe training set may contain some time slices where the plasma is anomalous but the plasma does not disrupt for various reasons (for example because the ramp down has already started and the plasma dynamics is slow enough to allow the safe termination of the discharge).

This implies that the classification training set could be inaccurate and classification algorithms may fail. Therefore, the following simple but more robust approach has been developed. The probability density function of the various cooling factors is calculated in the two cases of safe and disruptive discharges. So, for each region, the pdfs of 'safe' and '(potentially) anomalous' Λ_i are derived from the available examples. What it is expected is that the pdfs would diverge at higher Λ_i values as shown pictorially in figure 2(*a*) (the pdf of anomalous data should contain more cases at larger Λ_i). An example for the core radiation anomaly is shown in figure (*b*, left). From the pdfs the corresponding CDFs are calculated and the potential true positive and false alarm (in percentage) trends are obtained (figures (*a*) and (*b*), right). The true positive percentage (or sensitivity) trend is calculated as one minus the anomalous CDF, while false alarm percentage (or 1—specificity) is one minus the safe CDF.

Once the false alarms and sensitivity trends are derived (figure 2(*b*) right), the $\Lambda_{\mathrm{threshold},i}$ is evaluated by using the following rules:
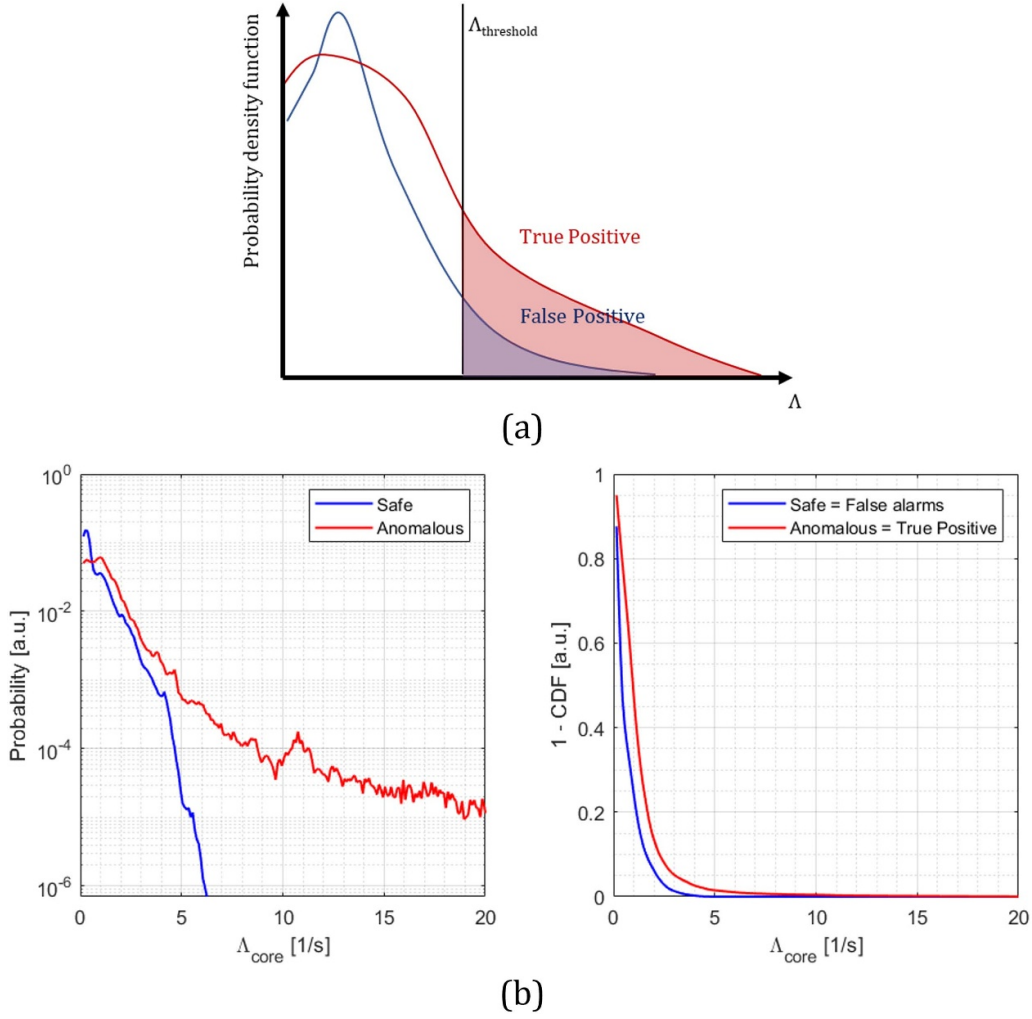
Figure 2. The method to calculate the $\Lambda_{\text{threshold}}$. Top: a cartoon describing the main principle for an unspecified region: the anomaly is calculated in terms of a threshold in the cumulative distribution function (see text for more details). Bottom: experimental pdf and CFD for the core radiation anomaly. The figure on the left clearly shows that there is a strong divergence between save and anomalous for Λ_{core} larger than 5, while the right figure illustrates the false alarm and true positive rates as a function of the threshold.

1. Search for the threshold that ensures a false alarm rate lower than 0.05% and a true positive larger than the false alarm rate.
2. If no threshold satisfies the previous rule, the false alarm limit is increased to 0.1%.
3. If no threshold satisfies the previous rules, the false alarm limit is increased to 1%.

Once the threshold has been determined on the basis of the CDFs, the probability to have a local radiative anomaly is calculated with a sigmoid function:

$$p_\Lambda = e^{d_{1\Lambda}(\Lambda_i - \Lambda_{\text{threshold},i}) + d_{2\Lambda}} / \left(1 + e^{d_{1\Lambda}(\Lambda_i - \Lambda_{\text{threshold},i}) + d_{2\Lambda}} \right). \tag{2.9}$$

Note that a $\Lambda_{\text{threshold},i}$ is found for each macro region (Core, HF, LF and Divertor) and an alarm is raised if one of the p_Λ exceeds 0.5.

As in the case of p_{ETH} and p_{EC}, also equation (2.8) is to be interpreted as the probability of the radiation being anomalous in a certain region, compared to the value assumed in

safe discharges. Therefore, as discussed in the next subsection, p_Λ provides information about the probability of anomalies developing in the temperature profile not directly about the propensity of the plasma to disrupt.

2.4. Prediction of the time to anomaly

By using the anomaly indicators described in the previous subsections, the algorithm is able to detect an anomaly and to classify it in one of the classes: mode locking, EC, hollowness, core radiation, HF radiation, LF radiation, divertor radiation. As mentioned, in the case of mode locking, the associated probability can be interpreted as the propensity of the plasma to disrupt. For the other classes, the probabilities quantify the how severe the corresponding anomaly is with respect to safe discharges. This is the basic information that the control system requires to take the proper countermeasures (mitigation, prevention, and avoidance). However, in order to accurately control the plasma and optimise feedback schemes, it would be very useful to have an idea of how much time is available

to undertake a certain action. For this reason, three specific predictors have been developed that predict the time to hollowness, time to EC and time to mode locking ($t_{r,\text{LM}}$, $t_{r,\text{ETH}}$ and $t_{r,\text{EC}}$).

The task is a typical supervised regression problem and for each predictor a feed-forward neural network is used. The architecture used is a three hidden layers neural network with 20, 10 and 5 neurons respectively. The output, i.e. the time to the anomaly, are predicted in logarithm scale in the range from 1 ms to 2 s (2 s is considered the time horizon). The loss function is a weighted MSE aiming at weighting more shorter times, to compensate for the fact that in shorter intervals there are less data (for example, from 1 ms to 10 ms there are 10 points per pulse, while from 100 ms to 1 s there are 900 points):

$$\text{loss}_{W,\text{MSE},j} = \frac{\sum W_i (\log_{10} t_{j,i} - \log_{10} y_{j,i})^2}{\sum W_i} \tag{2.10}$$

where:

$$W_i = \frac{1}{y_{j,i}}. \tag{2.11}$$

While $t_{j,i}$ and $y_{j,i}$ are the predicted and the effective time to anomaly j of the ith observation. These times are calculated continuously for the entire discharge providing each network with the following signals as inputs: the inputs are the plasma current (I_p), the magnetic toroidal field (B_t), the dimensionless locked mode (LM_N), the internal inductance (l_i), the electron temperature and density in the core, middle and edge regions, the electron temperature profile anomaly indicators (ETH and EC), the Λ_i in the four regions (Core, HF, LF and Div) and the Λ_{heating}.

The electron temperature and density in the core, middle and edge regions are the average values calculated considering the radii proposed in [12] (core for $R_0 < R < 3.4$, middle for $3.4 < R < 3.6$, and edge for $3.6 < R < 3.8$), while Λ_{heating} is the total input power normalised to the plasma energy (note that analogously to radiative Λ, this value is the inverse of a time that represents, in a certain sense, a characteristic heating time).

3. Adaptive implementation of proximity control

All the previous predictors are meant to be run in parallel during the execution of the experiments. Except the electron temperature anomaly indicators that are simple form factors, they are based on data driven or hybrid physics-data driven approaches. These predictors need to adapt to the changing circumstances, ensuing from the evolution of the experimental campaigns. To this end, the implemented approach consists of updating the training set on a shot-to-shot basis, not only to maintain but also to improve the performances of the predictors. Specifically, an automatic simple training set update logic, requiring no human intervention, has been developed. Of course, this implies that the results obtained in this work are conservative. Indeed, some post pulse analysis, performed by experts to better classify data before retraining, would lead

to various improvements of the complete algorithm performances. Such corrective manual interventions could be easily implemented in ITER between subsequent discharges. Such human revisions could also remedy problems with the diagnostic signals, even if the data quality and reliability will have to be much better on ITER and DEMO than on JET. In any case, the automated adaptive approach is the one implemented to obtain the results reported in the rest of the work. The logic of this automatic updating of the training set is described in detail in the following.

3.1. Mode locking training set update

The mode locking predictor is based on the locked mode signal and it typically needs to intervene at the end of the pre-disruptive phase, to trigger the measures of last resort, the mitigation actions. Consequently, the training set must be update only in clear cases:

1. If the previous pulse is disruptive and no remedial actions have been undertaken (no mitigation, prevention or avoidance), the last 100 ms are inserted in the training set and classified as anomalous (class = 1).
2. If the previous pulse is safe and no alarms have been raised, the pulse is stable and 100 random points are sampled sequentially, added to the training set sequentially (respecting the time ordering) and classified as safe (class = 0).

3.2. Radiation anomaly training set update

In the case of radiation anomaly, the training set update is less restrictive, and it follows the rules:

1. If a mode locking or an electron temperature anomaly is detected in the last pulse, the previous 2000 points are added to the training set and labelled as anomalous (class = 1). It has indeed proved more effective to train the predictor with the information about the anomalies caused by abnormal radiation than with the disruptive time slices.
2. If the last pulse triggered no alarms and it did not disrupt (safe pulse), all the points of the pulse are added in the training set and labelled as not anomalous (class = 0).

3.3. Time to anomaly training set update

The training set, to estimate the time remaining before an anomaly, is updated following the criteria:

For time to mode locking:

1. If the mode locking detection algorithm predicts a locked mode anomaly, the 2000 points before the alarm are taken and the targets are given as a function of the actual difference between the alarm time and the actual beginning of the current quench.
2. If no mode locking alarms are triggered and the pulse is not disruptive, 200 points are added to the training set with the target equal to the time horizon (2 s).

For Time to EC

1. If the EC detection algorithm predicts an EC anomaly, the 2000 points before the alarm are taken and the targets are given as a function of the actual difference between the alarm time and the actual beginning of the current quench.
2. If no EC alarms are triggered and the pulse is not disruptive, 200 points are added to the training set with the target equal to the time horizon (2 s).

For time to hollowness:

1. If the hollowness detection algorithm predicts a hollowness anomaly, the 2000 points before the alarm are taken and the targets are given as a function of the actual difference between the alarm time and the actual beginning of the current quench.
2. If no hollowness alarms are triggered and the pulse is not disruptive, 200 points are added to the training set with the target equal to the time horizon (2 s).

The dataset is randomly divided into training set (70%) and test set (30%).

As mentioned all these updates are implemented after each discharge. The computational times, being of the order of minutes, are fully compatible with the time between discharges.

Note that disruptive pulses are fewer than safe pulses and this may lead to unbalanced dataset. To alleviate this problem, different methodologies have been used. For mode locking classification, the cross-entropy loss function has been used. For time to anomaly prediction, on a safe pulse we sample fewer points than disruptive points and we also weighted the one closer to the disruption (equation (2.10)).

4. The database and the criteria to evaluate predictors' performances

This section is devoted first to the description of the large database built to test the proposed approach (section 4.1). Then the criteria to evaluate the performances of the developed tools are overviewed (section 4.2).

4.1. Description of the database

The database consists of 1683 JET pulses, spanning from the high-power DD campaign to the full tritium and DT campaign (C38, C39, C40, C41). The DB includes 1141 not disruptive pulses and 542 disruptive. The flat-top plasma current ranges from 1 MA to 3.5 MA, while the toroidal magnetic field goes from 1.7 T to 3.9 T. The maximum input power is 37.8 MW, where 31.6 MW were provided by the NBI and 5.1 MW by the ICRH. See table 1 for more details.

All disruptive shots have been included, except intentional disruptions and a few cases, in which essential data are missing. Note that, even if some of these pulses are in an operational regime where disruption would not involve danger for the tokamak, they have been included to test the capability of the predictor to perform well in a large range of plasma parameters. A graphical overview of the entire database is provided in figure 3. The bins cover 0.5 MA centred on the values indicated on the *x* axis (for example the bin centred at 2.5 MA includes discharges in the interval 2.25–2.75 MA). The absolute frequencies (figure 3(*a*)) are calculated dividing the number of safe (disruptive) cases in the bin e by the total number of safe (disruptive) pulses in the entire database. The relative frequencies (figure 3(*b*)) are calculated as the number of safe (disruptive) discharges in each bin divided by the total number of discharges in the same bin.

4.2. The criteria to evaluate the performances of the predictors in simulated real-time conditions

In order to evaluate the performances of the proposed predictors, they have been run in conditions exactly the same as the ones they would have faced in real life. To simulate an online situation, it is important to define the performance metrics and the alarm priorities.

The alarm priorities are in order of decreasing priority: mitigation, followed by prevention and then avoidance. It means that, for example, if both mitigation and prevention alarms are raised at the same time, it is assumed that the pulse must be mitigated because there is no time left for anything else. Therefore, the performance of the predictors will be evaluated according to the following classification, in which *mitigation time threshold* indicates the minimum time needed for implementing mitigation actions, *prevention time threshold* indicates the minimum time needed for implementing prevention actions and *avoidance time threshold* the minimum time needed for undertaking avoidance strategies.

In the case of *not disruptive pulses*:

1. If no alarms are triggered the pulse is classified as *safe no false alarms*.
2. If the predictor launches *only avoidance alarms*, the pulse is classified as *safe avoidance*.
3. If the predictor triggers only a *prevention alarm*, the pulse is classified as *safe false prevention*.
4. If the predictor triggers only *a mitigation alarm*, the pulse is classified as *safe false mitigation*.
5. If the predictor triggers *both prevention and mitigation alarms*, there are two possibilities. If the *prevention alarm time* is larger than the *prevention time threshold*, the pulse is classified as *safe false prevention*, otherwise it is classified as *safe false mitigation*.
6. If the predictor triggers *avoidance alarms* and *prevention and/or mitigation alarms*, there are three possibilities. If

Table 1. Database disruption/Safe pulse statistics. C38, C39, C40, C41 are the names of JET campaigns spanning from the high-power DD campaign (C38) to the full tritium and DT campaign (C41).

	C38	C39	C40	C41	Total
Total	907	168	310	298	1683
Safe	653	133	184	171	1141
Disruptive	254	35	126	127	542
Ramp up disruptions	0	0	0	0	0
Flat top disruptions	62	23	33	33	151
Ramp down disruptions	192	12	93	94	391
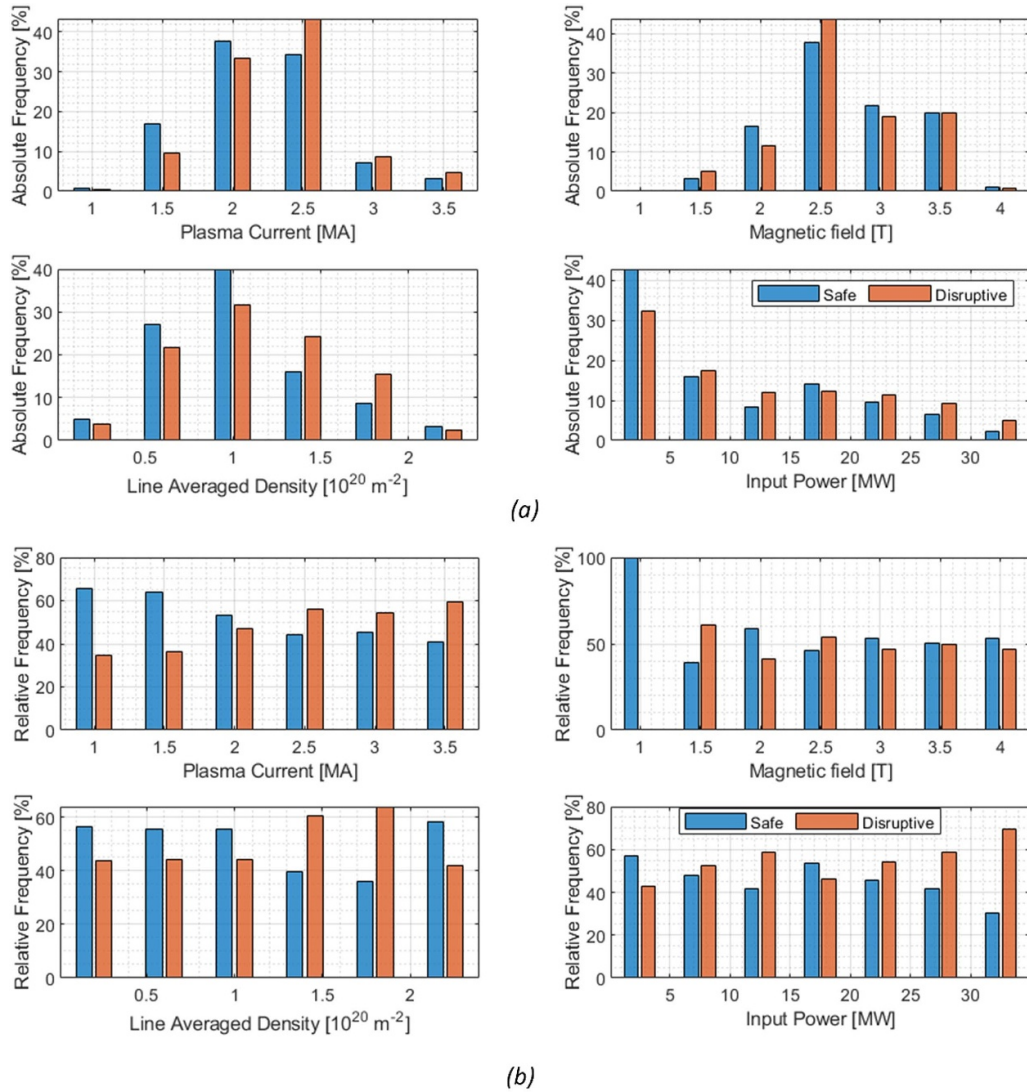
Figure 3. Absolute (*a*) and relative (*b*) frequencies of safe and disruptive discharges as a function of the main plasma parameters (see text for the details about the calculation of the percentages).

the *avoidance alarm time* is larger than *avoidance time threshold*, the pulse is classified as *safe avoidance*, otherwise it is classified as *safe false mitigation* or *safe false prevention* according to the previous rules.

In the case of *disruptive pulses*:
1. If no alarms are triggered the pulse is classified as *missed*.

2. If only a mitigation alarm has been triggered and the mitigation warning time is smaller than the minimum time for mitigation, the pulse is classified as *tardy mitigation*.

3. If only a mitigation alarm is triggered and the mitigation warning time is larger than the mitigation time threshold, the pulse is classified as *good mitigation*.
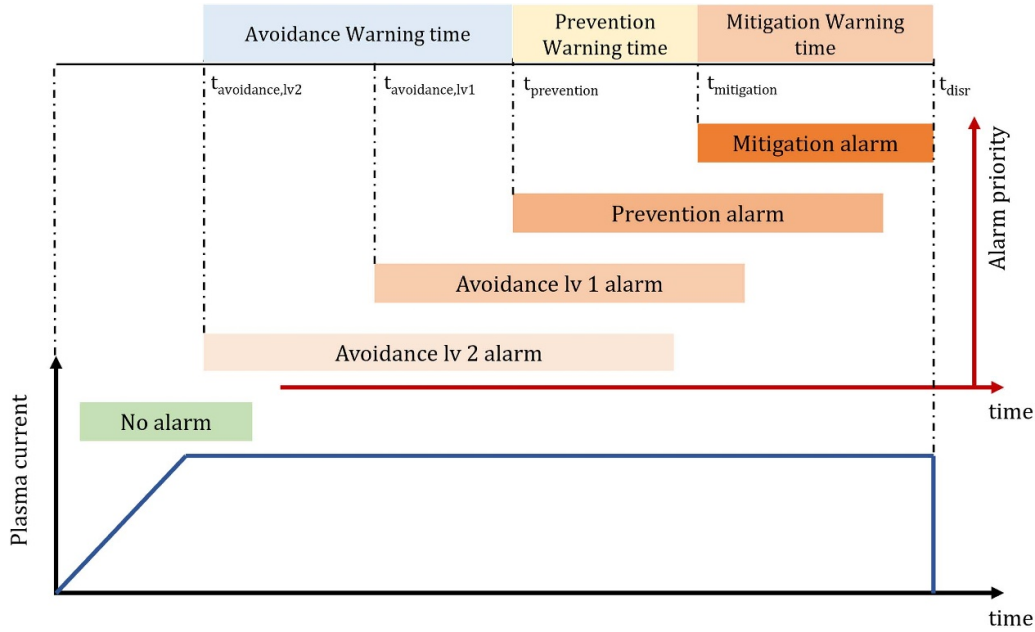
Figure 4. The logic of the typical sequence and priority of the alarms.

4. If only a prevention alarm is triggered and the prevention warning time is smaller than the prevention time threshold, the pulse is classified as *Tardy prevention*.

5. If only a prevention alarm is triggered and the prevention warning time is larger than the prevention time threshold, the pulse is classified as *good prevention*.

6. If both mitigation and prevention alarms are triggered, if the prevention warning time is larger than the prevention time threshold, the pulse is classified as *good prevention*. Otherwise, if the mitigation warning time is larger than the mitigation time threshold, the pulse is classified as *good mitigation*, else it is a *tardy mitigation*.

7. If an avoidance alarm is raised and the avoidance warning time is larger than the avoidance time threshold, the pulse is classified as *good avoidance*, otherwise the previous rules apply.

Figure 4 summarises the metrics and the main parameters implemented to define the metrics. Note that two avoidance levels have been considered in the scheme. They are just triggered as a function of the alarm. If a radiation anomaly is detected without an ETH alarm, the plasma should be in an operational region more stable than the operational region where hollowness is detected. Therefore, different avoidance schemes may be used. Consequently, with avoidance lv2 it is intended a radiation anomaly with no electron temperature anomalies, while anomaly lv1 is triggered when hollowness is detected (irrespective of the presence of radiation anomalies). All the details and a table about the evaluation criteria are provided in appendix B.

5. Results

This section provides a detailed overview of the results achieved with the developed techniques. section 5.1 describes how the predictors operate in practice and their overall performances. Some representative shots are specifically discussed in section 5.2. The relation between the estimates of the presented tools and JET control systems is the subject of section 5.3. The sensitivity of the predictors' results on the required warning times is investigated in section 5.4.

5.1. Predictor operation and performances

In this section, an example of how the predictor evolves during the campaign, for a fixed value of the hyper-parameters, is provided. In section 5.4, how performances change as a function of the required warning times for prevention and avoidance is analysed. The results in this section have been obtained assuming that the minimum warning time to undertake avoidance and prevention actions are 100 ms. The minimum time required to undertake successful mitigation actions is assumed to be 10 ms, to reflect the capabilities of JET tools.

At the beginning of the campaign, it is assumed that no experimental data is available. Thus, the mode locking predictor is trained with the physics loss function (the operational space after the first training is shown in figure 5—top left). The hollowness and EC indicators are determined by the thresholds that are self-determined (no training is needed). The radiation anomalies cannot be detected because it is considered too difficult to extrapolate the impurity and radiation behaviour to new, possibly much larger devices (however, if future studies will find some reliable scaling laws for radiation anomalies,
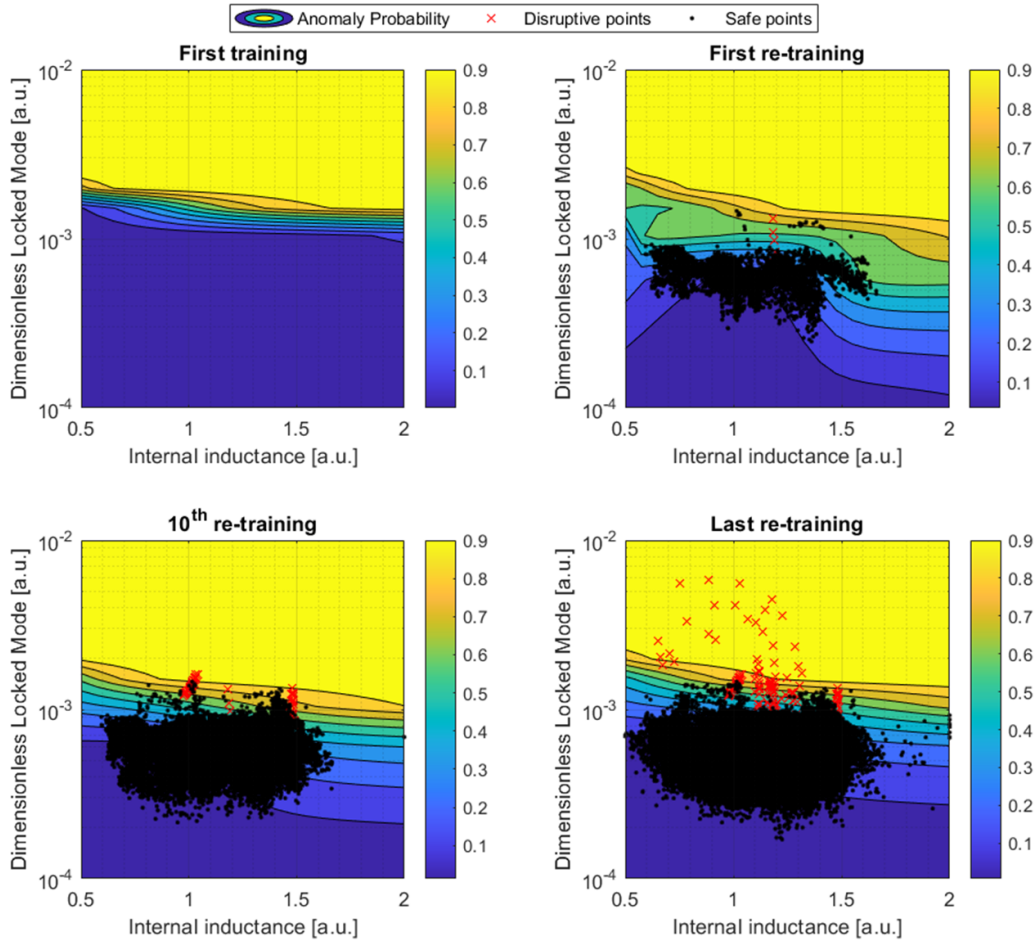
Figure 5. Mode locking—internal inductance predictor operational space as a function of retraining. The colour code indicates the probability of disruption. Black dots: safe time slices. Red crosses: disruptive time slices.

they may be implemented following a logic similar to the one used in this work for mode locking). It is assumed that also for the time-to-anomaly predictors there is not enough prior knowledge to train them without specific data of the new device.

In the analysed database, the evolution of the experiments is such that only after 30 discharges it is meaningful to perform the first training with the new data. Up to that point (before the first training), 14 disruptive pulses have been observed and 6 were mitigated, 3 prevented, 5 avoided (ETH) without neither tardy alarms nor missed disruptions. Regarding the 16 safe pulses, 3 were mitigated and 3 were avoided. However, by analysing the 6 pulses where 'false alarms' have been triggered, it can be observed that the anomalies are actually present, even if they do not lead to a major disruption. A typical example is pulse 94 162, the 6th pulse in the database. In this case, a 'false avoidance' due to ETH and a following false mitigation alarms are raised. However, both anomalies actually occur during the discharge. Figure 6 shows the main signals and anomaly indicators for this pulse. Both core radiation anomaly and hollowness indicators increase from $t = 13$ s, with a quite deep hollow profile from 13.22 s to 14.042 s, when the $n = 1$ mode locking amplitude increases (triggering a mitigation alarm). However, in this case unusually the mode locking does not lead to a disruption. Moreover in this shot mitigation actions would not

have had any major detrimental effect because the plasma current was already very low at the moment of the alarm.

After 30 pulses, all the predictors (mode locking, temperature profile anomalies, radiation anomaly and the time-to-anomaly) have been retrained by combining the physics loss function and the data driven loss function, derived from the training set automatically extracted from the first 30 pulses of the campaign. From this discharge and on the shot to shot basis, the entire algorithm detects and classifies the various anomalies and is automatically updated, by taking decisions according to the rules described in the previous section. The evolution of the operational space of the LM predictor and the radiation thresholds for the core region are shown in figures 5 and 7 respectively. The trends of this macropixel are representative of those of the other 3 regions. Figure 7 must be interpreted as described in figure 2 and the relative text.

The overall performances as a function of the progressive pulse number are shown in figure 8. At first, it can be observed that the sensitivity of the predictor is very high. At the end of the campaign, only three disruptive pulses were not correctly avoided, prevented, or mitigated (they were all tardy mitigations). Of the disruptive pulses, 190 have been mitigated, 81 prevented and 268 avoided. Of the three pulses for which the mitigation alarm was launched less than 10 ms before the
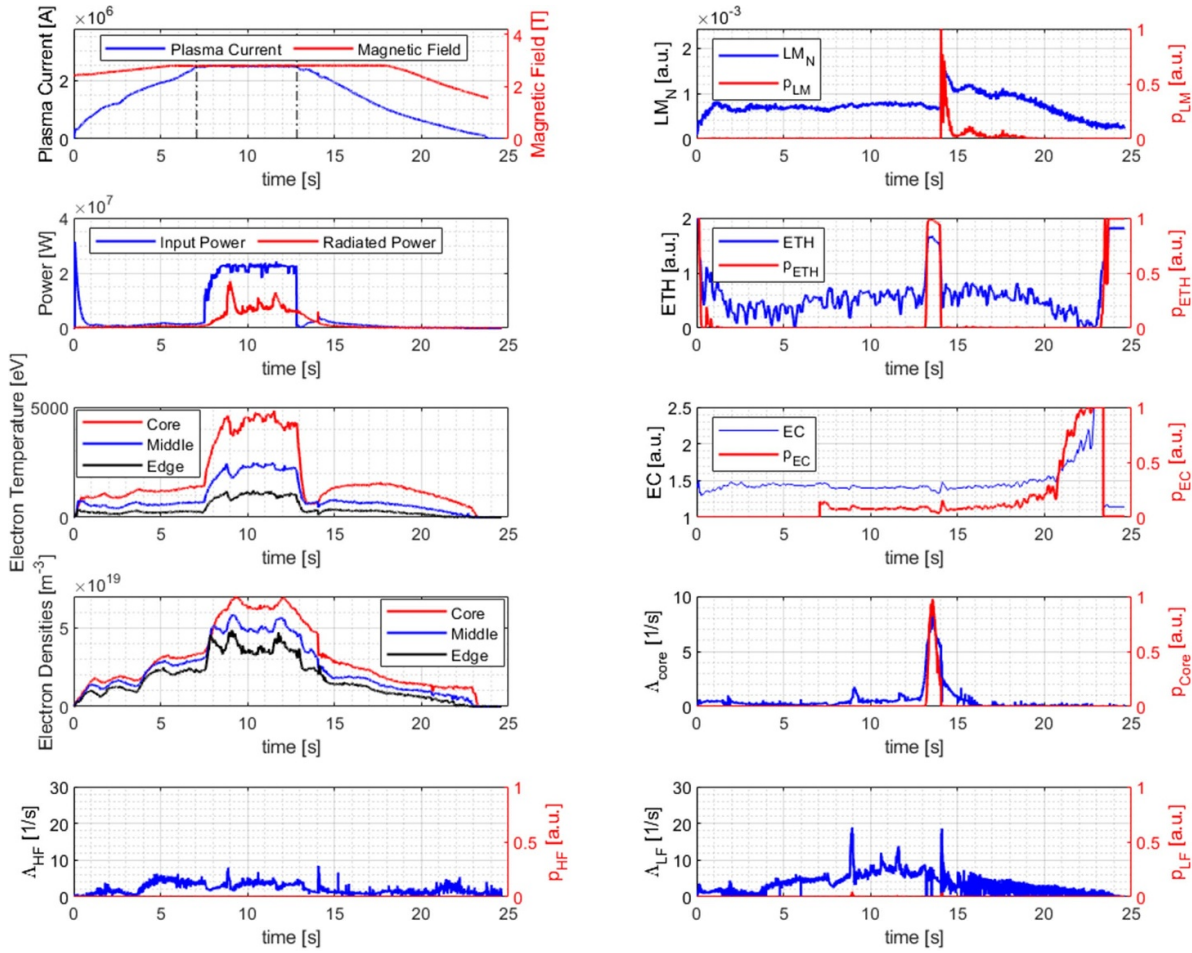
Figure 6. Pulse 94 162: main plasma parameters and implemented anomaly indicators. In the 4th row 2nd column it is possible to observe the core radiation peak detected as anomalous together with the electron temperature hollowness (3rd row, 2nd column). The mode lock signal (1st row, 2nd column) increases at the end of the hollowness/core radiation detection, suggesting that the mode has been triggered by the anomalous kinetic profile.
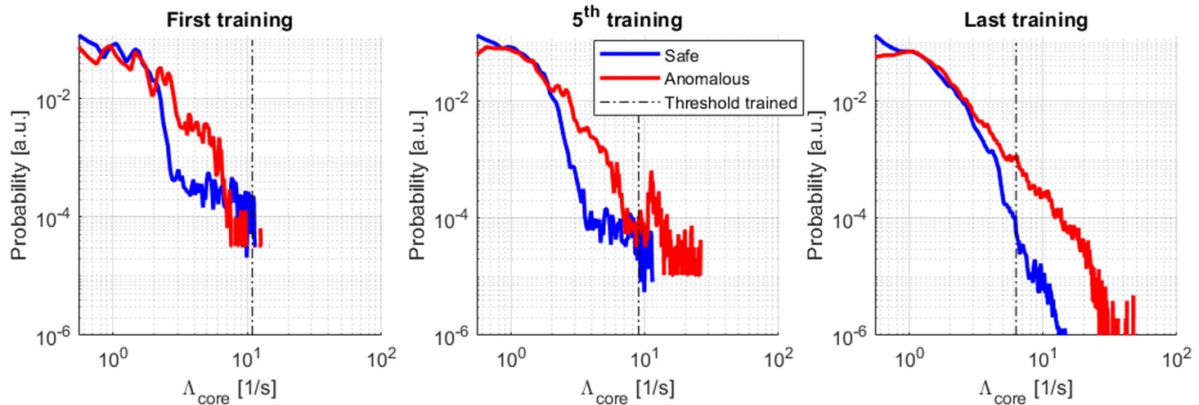

Figure 7. Probability distributions of anomalous and safe Λ_{core} after the first, the 5th and the last retraining.

beginning of the current quench, in two cases the warning time was 9 ms and in one case 2 ms. Given the length of the current quench (at JET, most of current quenches last for at least 30 ms, and usually they are between 50 and 150 ms), mitigation would have been potentially effective also in these three discharges [63, 64]. Moreover, current quenches should

be longer for large size future tokamaks (in ITER it is expected that they will be at least one order of magnitude longer) [63, 64].

For what concerns the specificity, it is important to discern between **'false anomaly detection'** and **'wrong action'**. 'False anomaly detection' occurs when an indicator detects an
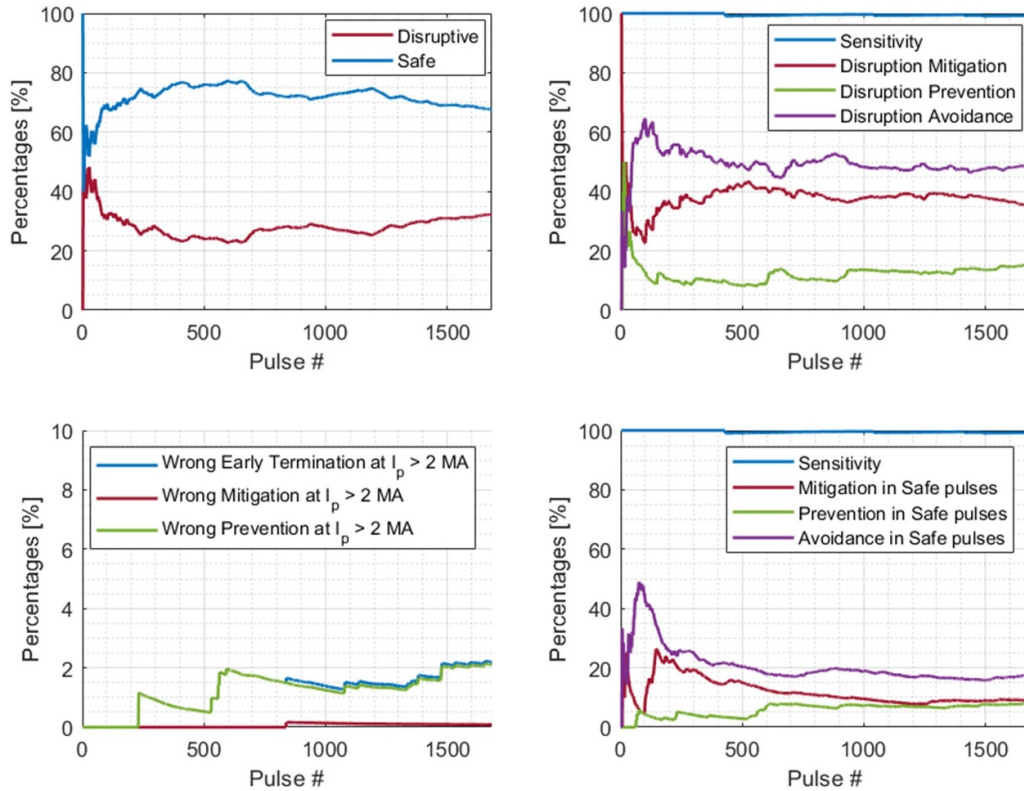
Figure 8. Top-left: percentage of disruptive and safe pulses in the database. Top-right: total sensitivity (avoided + prevented + mitigated disruptions) and percentage of correct actions triggered on disruptive pulses. Bottom-left: percentages of triggered 'hard actions' (mitigation + prevention) on safe pulses with high-plasma current. Bottom-right: sensitivity and total number of alarms triggered on safe pulses. The x axis reports the progressive discharge number, because the predictors are adaptive and run to simulate deployment in real time during the evolution of the experimental programme.

anomaly that is not actually present in the discharge. On the contrary, 'wrong action' is when a control/mitigation action is triggered but the plasma does not actually need it. For example, if the LM indicator detects an alarm and a locked mode is actually present, the outcome is a 'true anomaly detection'. However, if a mitigation action is triggered but the plasma would have survived without this action, this is considered a 'wrong action'.

With regard to the anomaly detection specificity, this value is extremely high. For example, every time the hollowness indicator detects a hollow profile the hollowness is actually present (this can be easily confirmed by visual inspection, and a detailed analysis is performed in [52]). Therefore, the specificity of the indicators is of the order of 100% for mode locking, EC and hollowness. Note that these statistics cannot be provided for radiation anomalies, since today there is not a physics-based definition of radiation anomaly. However, it is possible to estimate the reliability of radiation indicators through a comparison with the other anomaly indicators. For example, in 86% of safe pulses where core radiation anomaly has been detected also electron temperature anomaly and/or mode locking is observed. In the case of high-field and low-field radiation anomaly, this percentage drop to 62% and 71% respectively. However, this estimation is conservative since it is expected that radiation is only one of the drivers (other are input power and transport) of electron temperature anomalies.

Different is the case of 'wrong actions'. In this work, it is assumed that all actions by the control system would have been undertaken at a probability threshold of 50% for each anomaly. For example, if the mode locking amplitude probability exceeds 0.5, mitigation actions would have been triggered. However, it is clear that in a real-case, when different control strategies can be taken, performances can be improved. For example, there are several studies regarding the possibility of de-locking magnetic islands, recovering from this instability, without the need to mitigate the plasma [20, 65, 66]. In such a case, the pulse could have been recovered by the appropriate control action, without the need of mitigating. This would reduce significantly the amount of false mitigation and prevention actions. Moreover, it has to be considered that triggers of avoidance do not lead to 'hard' actions in the plasma but only 'soft' ones, i.e. control actions that should allow to run the pulse in a safer situation. Most of them would improve the plasma performances even in discharges not disrupting. In other words, a 'false avoidance' trigger is much less harmful than a 'false mitigation' trigger and must be interpreted differently.

Another important consideration is about the feedback actions actually implemented by JET real time control system. After years of work, on JET several controllers are operational to mitigate or early terminate the discharge [67–70]. This implies that the predictors developed in the present work may
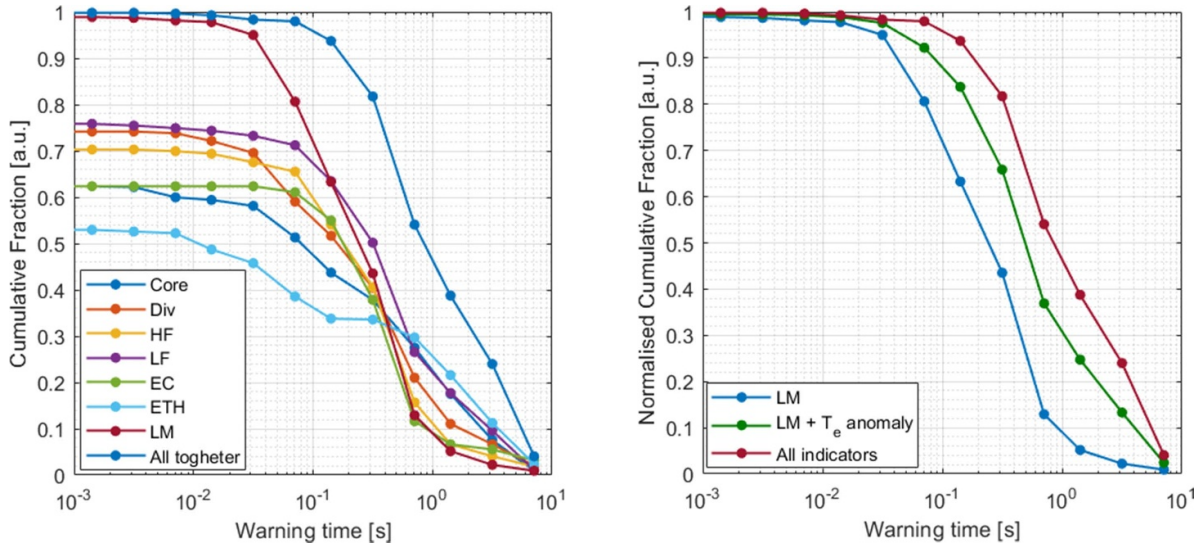
14

Figure 9. Left: cumulative fraction of true positive alarms (normalised with respect to the total number of disruptive pulses) for the various anomaly indicators as a function of the warning time. Right: cumulative fraction of true positive alarms for the predictor using only mode locking indicator (blue line), combining mode locking and electron temperature anomaly (green line), and using all indicators (LM, Te and radiation) shown with the red line.

detect anomalies that did not lead to a disruption because of JET control system intervening with the appropriate remedial actions. In this part of the work, the statistics have been calculated assuming that safe pulses are stable pulses and that disruptive ones are not. Therefore, the statistical results previously reported are to be considered conservative, since many safe pulses, in which anomalies are detected, are classified as 'wrong actions' even if the detected anomaly is absolutely real (but the plasma does not disrupt for various reasons including JET real time system interventions). To investigate this aspect in more details, section 5.3 reports the statistics for a subset of pulses, which have been analysed manually, to assess in detail the performances of the developed predictors and how they fare compared with JET control/prevention/mitigation actions.

Figure 9 (left) shows the cumulative fraction of warning times for each indicator, while figure 9 (right) show the cumulative fraction of three different predictors: the first one uses only the mode locking indicators, the second implements also the electron temperature anomalies, while the third combines all the indicators (magnetic, kinetic profiles, radiation patterns). The mode locking-based predictor LM returns the typical warning time curve observed in other works, with a median warning time around 200 ms. The use of hollowness and EC indicators allow to increase the median warning time from $\sim$200 ms to $\sim$500 ms, where the strongest contribution in the large warning time is given by the hollowness (EC are quite fast phenomena, as also discussed by [12], with a warning time between EC and mode locking of the order of 50–100 ms). The implementation of radiative indicators allows to increase the median warning time to 900 ms, with at least 90% of pulse predicted at least 200 ms before the beginning of the current quench.

Concerning the time to anomaly predictions, figure 10 (left) shows the median error committed by the predictors as a

function to the actual time to anomaly. The errors decrease linearly in logarithmic scale (as expected, being the predictor trained in log-scale), showing the capability of the neural networks to estimate correctly the order of magnitude of the time remaining before any of the three anomalies is observed. It is worth noting that the time-to-mode-locking is the most accurate prediction since mode locking is normally preceded by the other anomalies (radiation and temperature) with a quite large margin, allowing a reliable evaluation of the plasma stability. On the contrary, prediction of time-to-hollowness is the most difficult because this modification of the temperature profile can appear very close to core radiation anomaly and it is not preceded by other (real-time) clear features on the signals. For this reason, the 'time horizon' of the time-to-hollowness prediction is very small and therefore the error of the prediction quite high. The EC error is in between, since it is usually preceded by a clear anomaly radiation on the high-field side (MARFE-like radiation) and/or low field side radiation.

Figure 10 (right) shows the coefficient of determination R^2 calculated for each predicted pulse as a function of the pulse numbers [71]. Here, it can be seen that, on average, the prediction performances increase with the pulse number, and this is to be expected since more pulses are used to train the predictors and therefore they gain in accuracy. However, it can be observed that performances are quite good even just after the few pulses used for the training. It has to be highlighted that the time to anomaly prediction is based on a fully data-driven methodology, since theoretical/empirical equations deployable in real-time have not been developed yet. However, future works on this subject may achieve a better prediction of the time to anomaly from day-0, by exploiting physical or empirical models of these quantities.
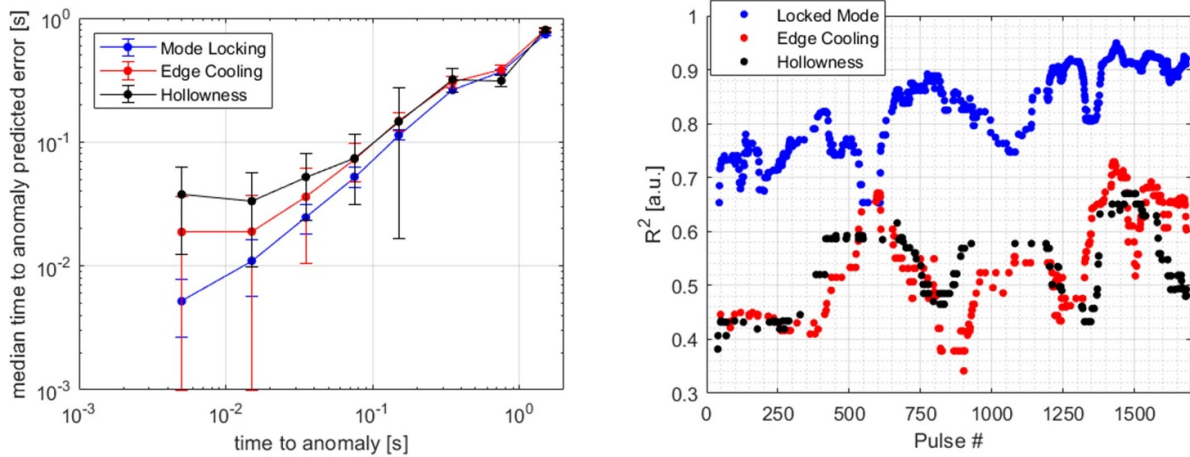
Figure 10. Median time to anomaly prediction error as a function of the actual time to anomaly (left) and the actual vs predicted time to anomaly R^2 as a function of the pulse # of the campaign (right). R^2 is the traditional coefficient of determination. In the plot on the right, the x axis reports the progressive discharge number, because the predictors are adaptive and run to simulate deployment in real time during the evolution of the experimental programme.

5.2. Pulse examples

In this section, some examples are reported to illustrate the functioning of the predictors, the various cases leading to recovering actions, and the amount of information that they are able to provide for the control of the plasma.

The first example is the pulse 96 486, a baseline scenario with a 3.5 MA flat top plasma current and ∼35 MW input power (30 MW from NBI, 4.2 MW from ICRH). The main signals are shown in figure 11. From $t = 13$ s, a second before the ramp-down phase, core radiation starts to slightly increase. At the beginning of the ramp down, due to the input power decrease, the ELMy mode terminates (figure 11, top-left) with an increase of impurity accumulation in the core. At $t = 14.179$ s, anomalous core radiation is detected, and at $t = 14.246$ s also the ETH alarm is triggered. The core radiation anomaly grows exponentially and at $t = 15.47$ the mode locking alarm is triggered. In this pulse, the only disruption prevention/mitigation action undertaken by JET control system was the activation of the disruption mitigation valve at $t = 15.492$. In figure 11 (bottom-left), the various alarms raised (following the logic described in the previous sections) are shown. In brief, the predictor raised an avoidance alarm 1.292 s in advance with respect to the mode locking. Then, the mitigation alarm is triggered 31 ms before the beginning of the current quench. It has to be mentioned that this is a pulse mitigated by JET control system; consequently, the actual warning time provide by the developed predictor may be larger than the one just quoted.

The core-electron temperature anomaly behaviour can be observed also in figure 12 (right). It can be observed that at first the core radiation anomaly slightly increases. This leads to an increase of the ETH that amplifies the core radiation for two reasons. First the impurities have a radiation function which increases when the temperature goes down. Secondly to maintain pressure equilibrium, during the onset of the ETH

the core density peaking is observed to increase, favouring the accumulation of impurities in the core. Both effects contribute to establishing a clear positive feedback loop that makes the thermal-radiative instability grow exponentially.

Figure 13 shows the time to anomaly predictions. When the plasma in a stable operational regime, far away from the anomaly onset, the time to anomaly prediction return a very high value, comparable with the time horizon (between 1 s and 2 s). When core radiation starts increasing ($t > 13$ s), a drop of the time to hollowness ETH signals is observed and when an anomalous core radiation is detected ($t = 14.179$), the time to ETH is around 35 ms. After 67 ms, the ETH anomaly is detected. The good agreement between predicted and actual time to mode locking is shown in figure 13 (right). As anticipated in the previous section, the prediction of the time to mode locking anomaly is much more reliable than the time to electron temperature anomaly since the mode locking is preceded by clear and measurable effects (radiation, temperature).

The second example is pulse 94 650. This is another baseline scenario at 3 MA plasma current and 25 MW input power (22.3 MW from NBI and 2.4 MW from ICRH).

Figure 14 shows the most relevant signals for pulse 94 650. At first, two spikes are observed in the edge regions (HF and LF) at $t = 6.8$ s and $t = 13.45$ s, probably caused by some impurity influxes. From $t = 15.5$, an increase of edge radiation is observed and at $t = 15.8$ s an intense radiation activity is observed on the high field side (MARFE-like radiation, also confirmed by a visible camera). After 175 ms, the edge radiation leads to an EC of the electron temperature profile (detected a $t = 16.142$ s) and a consequent mode locking at $t = 16.515$ s. The pulse disrupts at $t = 16.611$ s, with a plasma current equal to 1.73 MA. Therefore, in this case, the avoidance warning time (with respect to the prevention alarm) lasts for 175 ms, followed by a prevention warning time (with respect to the mitigation alarm) of 373 ms, and a mitigation warning time of 96 ms.
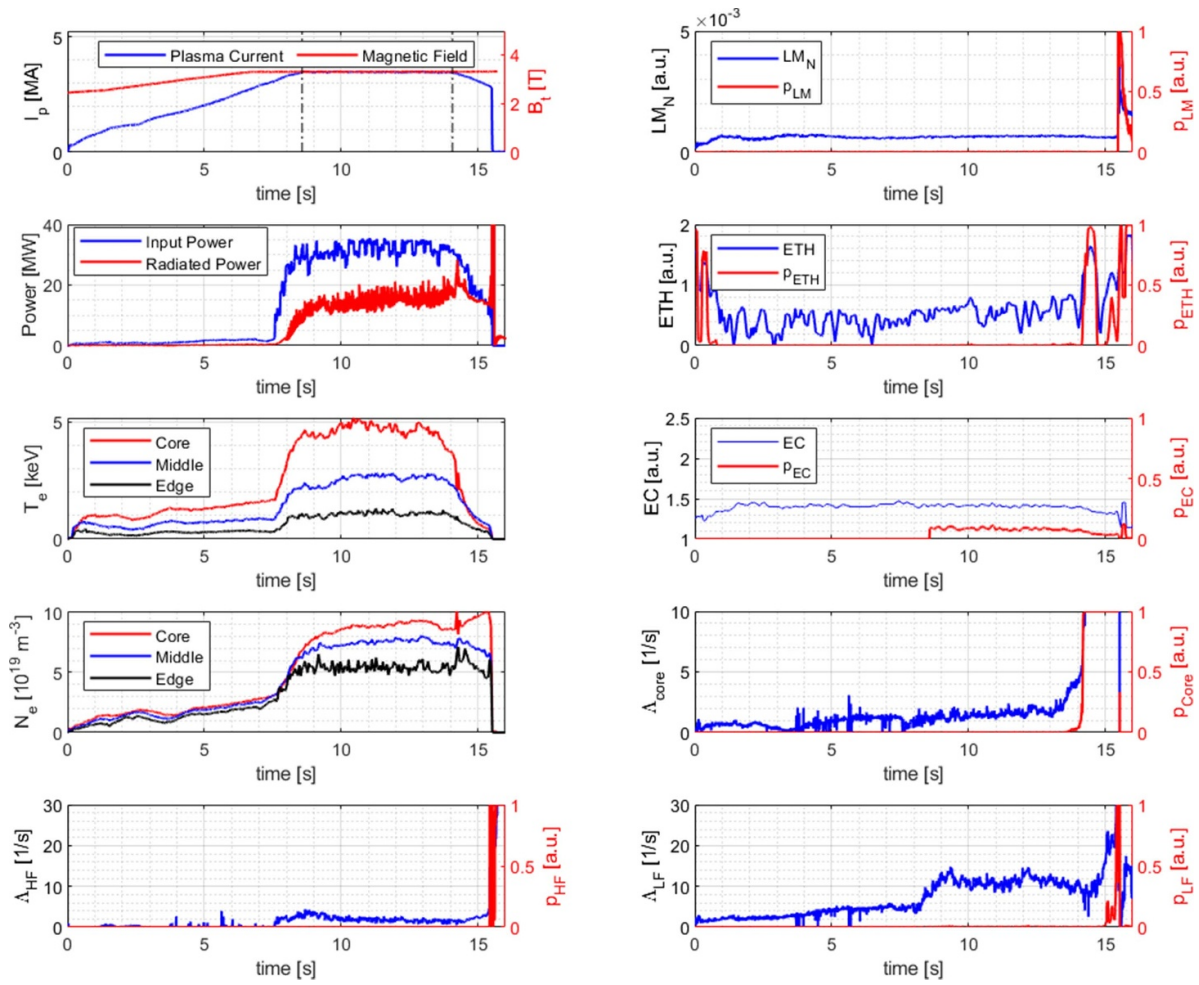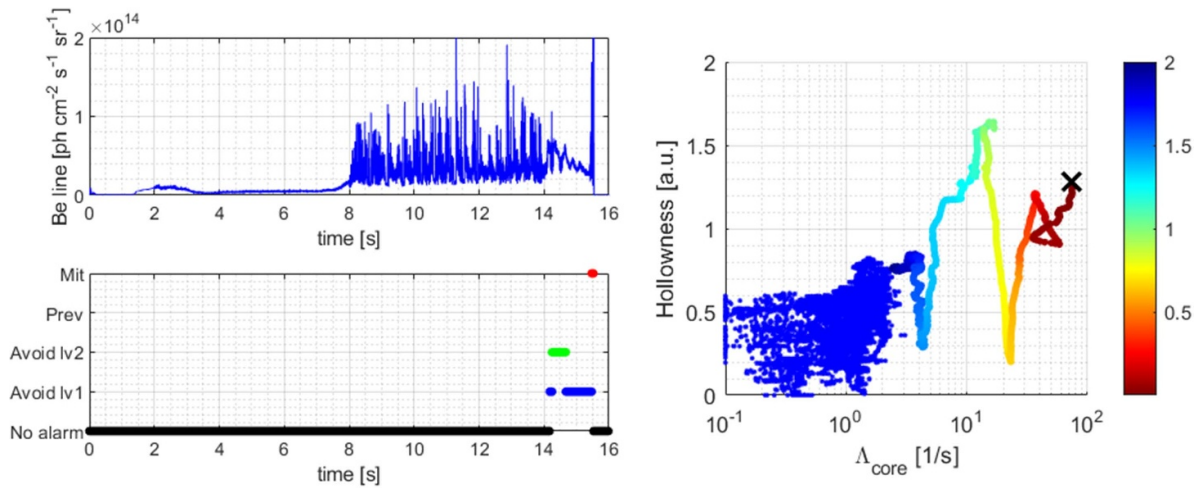
Figure 11. Pulse 96 486—main signals.

Figure 12. Pulse 96 486—Top-left: beryllium spectroscopic line; bottom left: alarms triggered by the predictor. Right: pulse time traces in the phase-space Λ_{core}—Hollowness indicator.

Figure 15 reports the signal in the space HF radiation indicator—EC indicator. Also in this case, it is possible to observe that the radiation and temperature anomalies increase together, with an exponential growth of the EC signals, since the local decrease of the electron temperature causes a local increase of the electron density (pressure balance), with a consequent amplification of the local radiation (confirming the positive feedback loop).
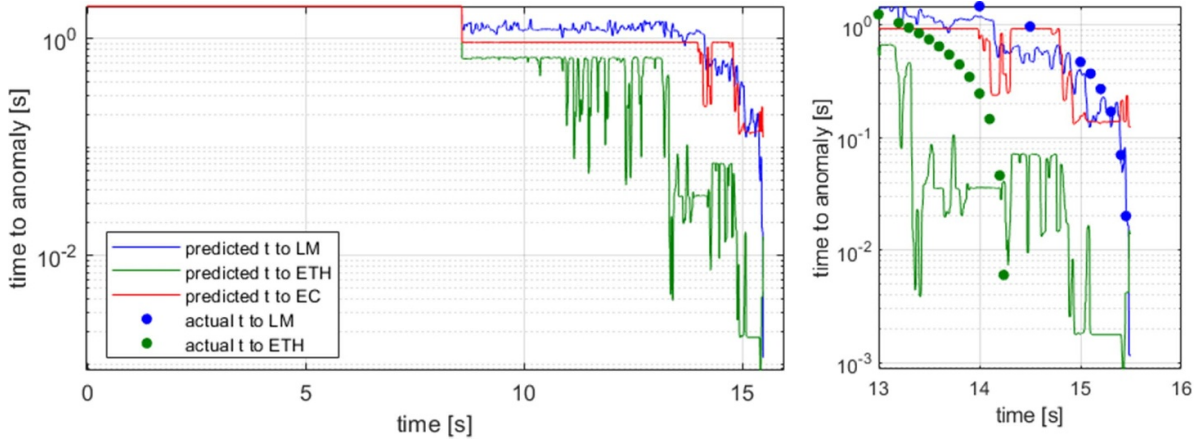
Figure 13. Pulse 96 486—Time to anomaly predictions and targets (left: entire pulse: right: zoom near disruption).
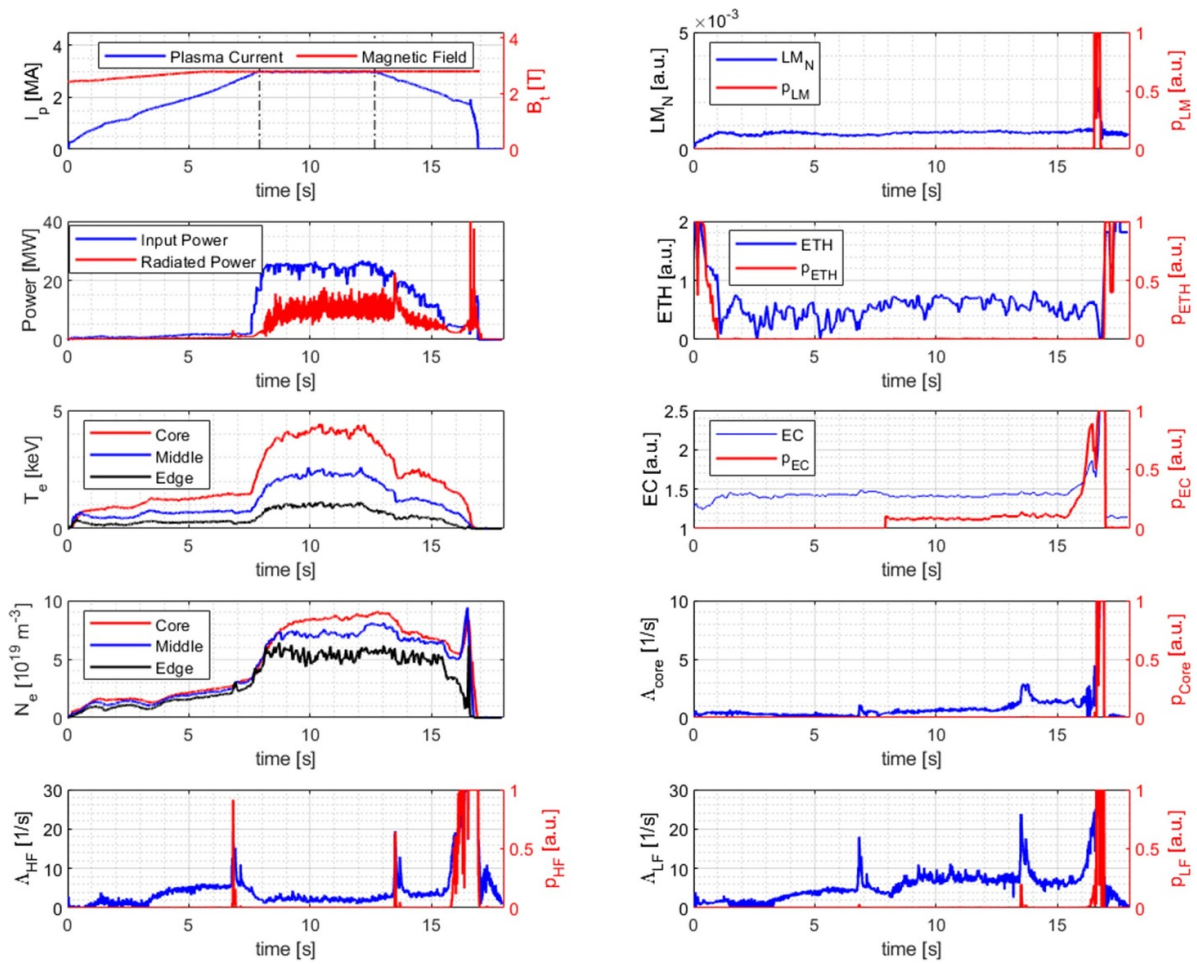

Figure 14. Pulse 94 650—main signals.

In figure 16, the time to anomaly predictions for pulse 94 650 are shown. In this case, both the time to mode locking and time to EC predictions follow quite well the actual evolution of the respective anomalies, when the pulse is in the unstable phase.

5.3. Comparison with JET control system

In this subsection, a subset of the entire database is considered, to compare the results of the predictors with the already implemented real-time JET controllers and
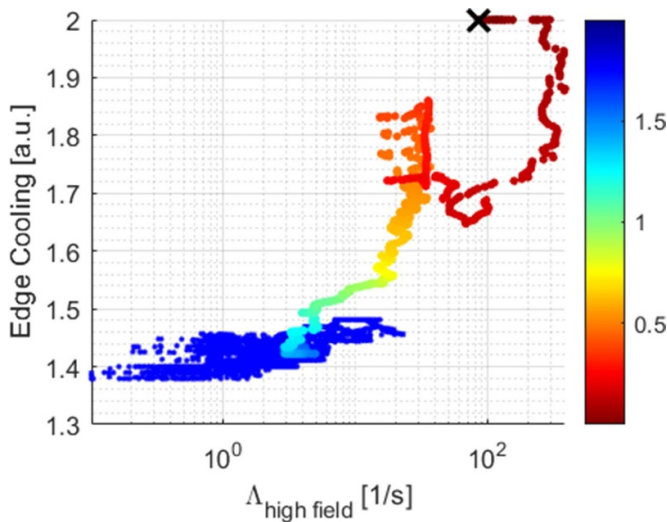
Figure 15. Pulse 94 650—Pulse time traces in the phase-space $\Lambda_{\text{high-field}}$—Edge cooling indicator.

prevention/mitigation systems. The pulses included in this analysis belong to the DD campaign for DT preparation (the so-called C38) and they were performed in the framework of to the two main JET tasks of developing the 'ITER baseline scenario' and the 'hybrid scenario'. In total, this data set comprises 75 disruptive and 211 safe discharges. These are discharges in which the interventions of JET control system was very clear and properly documented.

On JET, different indicators are already implemented to detect plasma anomalies during a discharge and trigger some actions to mitigate or prevent disruptions.

Regarding the mitigation actions, on JET there are disruption mitigation valves (DMV) that inject gas inside the plasma. This gas causes the plasma to irradiate, loosing energy, and so reducing the loads on the plasma facing components, since most of the plasma energy is lost isotropically by the radiation channel [72, 73].

A second relevant action, that can be considered a prevention action, is the so called 'jump to termination' or JTT; it is used to terminate the plasma discharge early in a controlled way when some anomalous radiation events occur (e.g. tungsten impurity accumulation).

Detailed descriptions of the JET real-time implemented algorithms and control actions can be found in the literature [67–70].

Table 2 shows the comparison between the present predictor and the JET termination actions in the dataset considered and the main findings are reported below.

All the pulses where the JET control fired the disruption mitigation valve are detected significantly earlier by the predictors. In 28 pulses, the avoidance alarms are triggered with sufficient warning time to make avoidance strategies feasible, while 5 of them would require prevention and 3 mitigation. The median anticipation between the avoidance alarm and the DMV is 689 ms, the average is 1.115 s, and the minimum value is 135 ms. For prevention, the median is 118 ms, the average

135 ms, and the minimum warning time is 80 ms. Finally, for mitigation there is a median value of 165 ms, an average of 481 ms and a minimum value of 25 ms.

In 19 disruptive pulses a JTT procedure was undertaken and in 18 of them successful avoidance alarms would have been triggered by our predictors, whereas in one only the mitigation alarm would have been triggered. The median avoidance warning time (with respect to JTT) is 120 ms, with an average of 350 ms. The minimum value is negative (-0.45 ms), since two pulses triggered an avoidance alarm slightly delayed with respect to JTT.

In 20 disruptive discharges, the JET system did not trigger neither a JTT nor DMV action. However, all of these disruptions occur below the JET plasma current threshold (2 MA) and therefore actions might not have been triggered for this reason. In any case, of these 5 pulses would have satisfied the criterion for mitigated, 4 for prevention, and 11 for avoidance.

Out of all 33 safe pulses where JET triggered a JTT action, the mitigation predictor triggers seven times: it has been checked that mode locking was actually present in the ramp down, below the plasma current threshold for DMV triggering. In ten discharges the avoidance predictor properly launches and alarm. In all safe pulses where JET did not take any actions, also the predictors trigger no alarms. This confirms that the algorithm is very specific, and that the false alarms are mostly triggered in situations where the plasma is actually unstable even if it does not disrupt.

5.4. On the influence of required warning times on the performances

The results reported in the previous subsection have been obtained assuming that the warning times required for successful avoidance and prevention are both 100 ms (and 10 ms for mitigation). However, since there is still some research on control schemes for avoidance and prevention, it is clear that the minimum warning times may be different in future experiments. It has also to be considered that different warnings time may be required as a function of the anomaly typology and intensity. For these reasons, a parametric study of how the performances vary as a function of the required warning times has been performed.

Figure 17 shows how the performances in disruptive pulses change as a function of the required avoidance warning times (left) and prevention warning times (right). As expected, avoided disruptive pulse percentages decrease with increasing warning times. However, this decrease is limited and does not affect the missed/tardy pulses, but simply implies that the numbers of mitigated/prevented pulses increase. Similar results are observed for the increase of required prevention warning times. If the required warning increases, less pulses will be prevented, but mitigation actions will suffice to avoid unmitigated disruptions. It has to be note that in this case, since prevention priority is higher than avoidance (section 4.2— figure 4), the pulses, in which avoidance alarms are triggered, remain the same.
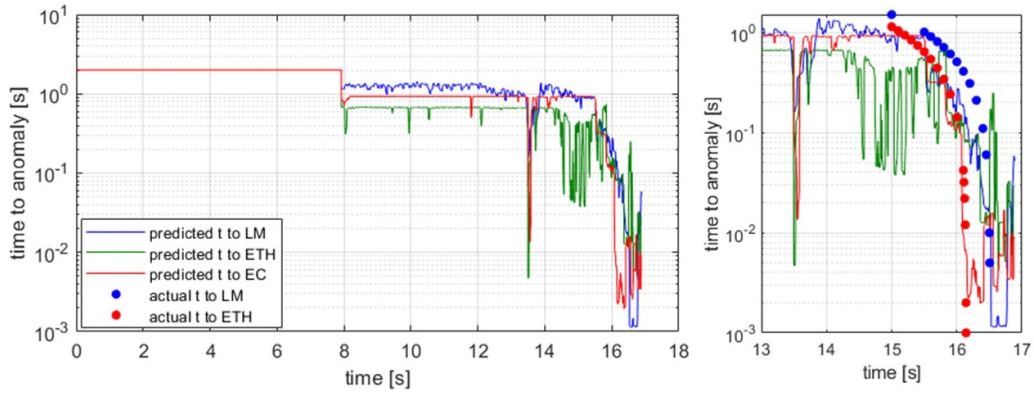
Figure 16. Pulse 94 650—Time to anomaly predictions and targets (left: entire pulse: right: zoom near disruption).

Table 2. Comparison between JET termination actions and the predictor response. DMV is the disruption mitigation valve. JTT is the acronym indicating the jump to termination procedure.

	Total	Mitigation	Prevention	Avoidance
Disruptive JET DMV	36	3	5	28
Disruptive JET JTT	19	1	0	18
Disruptive JET no actions	20	5	4	11
Safe JET JTT	33	7	0	10
Safe JET no actions	178	0	0	0

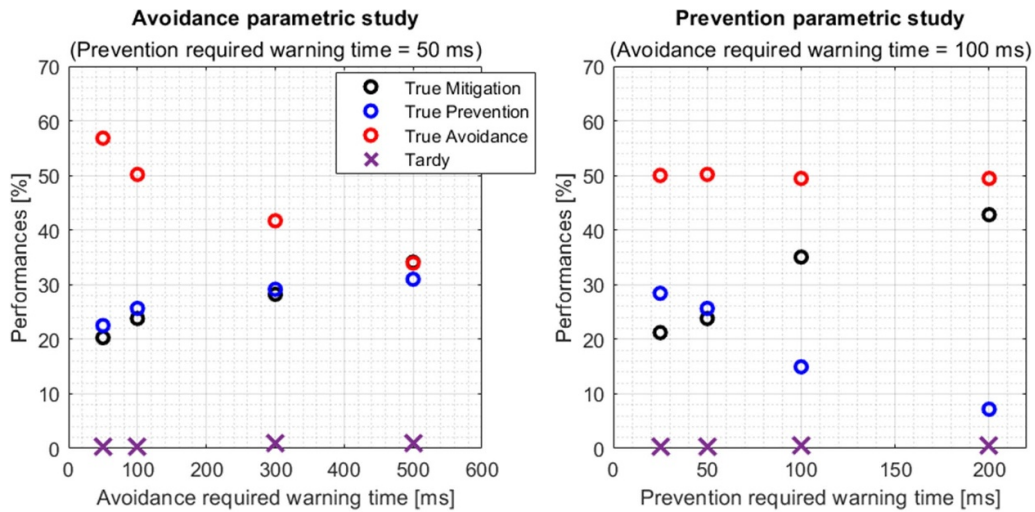| Disruptive Time vs DMV	Mitigation	Prevention	Avoidance
Mean (ms)	481	135	1115
Median (ms)	165	118	689
Min (ms)	25	80	135

Figure 17. Performances on disruptive pulses as a function of required avoidance warning time (left) and required prevention warning time (right). The y axes report the number of the successful actions indicated in the legend as percentage of the total number of disruptive discharges.

The results shown in figure 17 clearly demonstrate that the adaptive training logic has a good reliability and that all the predictors work in such a way as to ensure that disruptive pulses are at least mitigated, favouring actions that aim at triggering prevention/mitigation actions.

6. Conclusions and next steps

A set of predictors have been developed, which combine physics and data driven methodologies and utilise only signals and technologies compatible with real time deployment. In

addition to providing alarms to trigger specific types of actions, they also estimate the probability of occurrence and the time to the various anomalies. The ones devoted to mitigation and prevention can be operated from the first discharge of a campaign or the operation of new devices.

The tests performed in real time like situations provide very good results in terms of all the most relevant metrics: sensitivity, specificity and warning times. Their performances do not show any significant dependence on the isotopic composition, proving the generality of the technical solutions adopted. Comparison with the actions taken by JET control system indicates two important facts. First the predictors return very reliable and detailed information, in line with previously developed and implemented methodologies, but are typically more informative. Moreover, the vast majority of the alarms triggered by the developed predictors precede JET remedial control actions by a very significant margin.

The combination of physics and data-driven techniques makes it possible to develop predictors that maximise the transfer of information from studies in different machines, also allowing the integration of knowledge from theoretical studies and numerical simulations. In addition, the use of scaling laws (experimental or empirical) in the models is expected to facilitate transfer learning and adaptive learning. Being better grounded in the physics, these models should be less prone to errors, particularly at the start of operations of a new machine or when there is a large variation in the machine's operational regime.

The results obtained in this work are conservative, because the physics equation has been kept constant and not updated with new data. However, on a shot-to-shot basis, the physical/empirical scaling laws could be updated by the scientists and therefore the extrapolation error could be significantly reduced. In terms of future developments, it is indeed planned to test the improvements that can be achieved by manual intervention, by correcting some errors of the predictors. This is an activity that could take place between discharges in real life, and it is expected to have a quite significant impact on the already quite good performances.

The most natural next step in terms of application is certainly the transfer of the proposed predictors to other devices, a task that should not be prohibitively difficult, given the nature of the indicators implemented. Another route to explore is the potential of state aware predictors, particularised for the main phases of the discharge, ramp up, flat top and ramp down. This would require implementing trajectory learning, a series of techniques which have proven to be quite effective in the past [22, 34].

Aknowledgements

Appendix A. The estimate of the macro-pixels internal energy

In the simple indicator expressed by equation (2.1), the most delicate aspect is the determination of the internal energy of each macro-pixel with sufficient accuracy. The basic alternatives, to estimate this quantity in real-time, are described in this appendix.

The equation to calculate the plasma internal energy per unit volume can be written as:

$$u = \frac{p_i + p_e}{\gamma - 1} = \frac{n_i T_i + n_e T_e}{\gamma - 1} \tag{A.1}$$

where the subscript e indicates the electrons and i the ions. In the traditional assumptions of a deuterium fully ionised plasmas ($n_e = n_i = n$ e $\gamma = 5/3$) with $T_e = T_i = T$, one obtains:

$$u = \frac{3}{2} nT. \tag{A.2}$$

The energy inside a given volume V can then be calculated as:

$$E_V = \int_V \frac{3}{2} nT \mathrm{d}V \tag{A.3}$$

An alternative, to determine the plasma energy in a given region, consists therefore of using the magnetic topology provided by a real time equilibrium code and the temperature and density profiles from kinetic measurements (typically the Thomson scattering). Under the usual assumption of the magnetic surfaces being isobars, the temperature, density and pressure fields can be calculated as shown in the example reported in figure A1. These bidimensional maps allow determining the energy in any region of the poloidal cross section with discrete integrals of the form:

$$E_k = \int_{V_k} \frac{3}{2} nT \mathrm{d}V = \frac{3}{2} 2\pi \sum \sum 2 n_{e,i,j} T_{e,i,j} R_{i,j} \mathrm{d}R \mathrm{d}Z \tag{A.4}$$

where $n_{e,i,j}$ and $T_{e,i,j}$ are the density and temperature of the pixel, whose barycentre has the coordinates (i, j). $\mathrm{d}R$ e $\mathrm{d}Z$ are the dimensions along R and Z and $R_{i,j}$ is the major radius of the corresponding pixel.

Unfortunately, on JET no real-time equilibrium reconstruction code is routinely available. A simplified but quite effective alternative solution has proved to be fitting the High-Resolution Thomson Scattering (HRTS) profiles in the region from $R = 3$ to $R = 3.9$ with a second order polynomial. Such a fit allows determining the typical percentage of the plasma
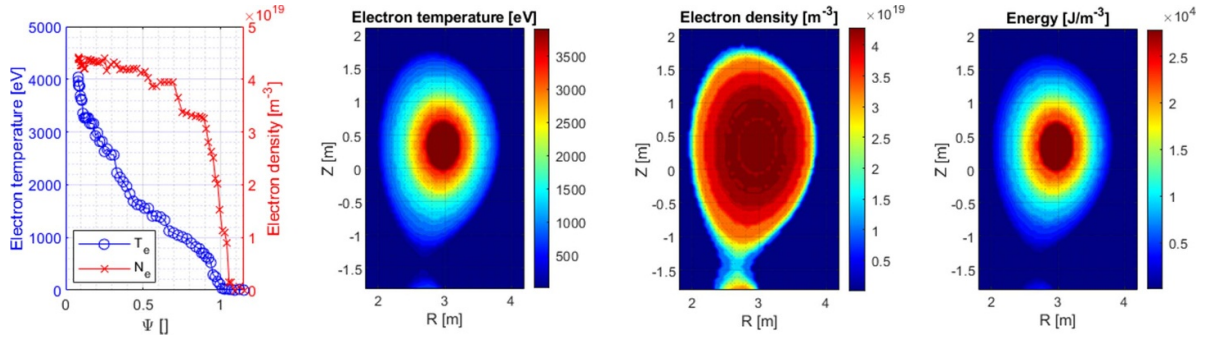
Figure A1. Reconstruction of the temperature, density and pressure fields for discharge number 95 998 at t = 18 s.

Table AI. Relative error of the internal energy fraction estimate with the fit of the HRTS profiles compared to the more sophisticated calculations using the equilibrium reconstruction.

	LFB	LFR	LFT	Core	Div	Top	HFL	HFT
Relative error using HRTS	17.03%	8.47%	23.00%	4.08%	13.85%	17.41%	11.80%	23.38%

internal energy contained in each macro-pixel. For a new discharges, the energy in each macro-pixel can be determined as a fraction of the total plasma energy, which is available in real time (for example from the measurements of the diamagnetic effect). Table AI reports the relative error on the energy fraction estimated using the HRTS fitting against the equilibrium-based method for each macro-pixel. It is worth mentioning that this alternative determination of the macro-pixels energy keeps the number of diagnostics required by the entire methodology to a minimum, since the HRTS is also the system used to calculate the indicators of hollowness and EC (see section 2).

Appendix B: predictor performances evaluation details

The following table provides all the decision rules for evaluating the performances of the developed predictors. The times indicate the instants of the corresponding detections. $\Delta t_{\text{min,mitigation}}$ is the minimum time interval required by the mitigation tools to intervene, $\Delta t_{\text{min,prev}}$ is the minimum time interval required to implement prevention actions, $\Delta t_{\text{min,avoid}}$ is the minimum time interval required to implement avoidance strategies.

Table B1. Table describing the rules for the evaluation of the predictor performances.

Pulse type	Alarm(s)	Condition	Classification	Rule #
Safe	No alarms		Safe no false alarms	S1
Safe	Avoidance		Safe avoidance	S2
Safe	Prevention		Safe false prevention	S3
Safe	Mitigation		Safe false mitigation	S4
Safe	Prevention, mitigation	If $t_{\text{mitigation}} - t_{\text{prevention}} \geqslant \Delta t_{\text{min,prev}}$	Safe false prevention	S5
Safe	Prevention, Mitigation	If $t_{\text{mitigation}} - t_{\text{prevention}} < \Delta t_{\text{min,prev}}$	Safe false mitigation	S6
Safe	Avoidance, prevention, mitigation	If $t_{\text{prevention}} - t_{\text{avoidance}} \geqslant \Delta t_{\text{min,avoid}}$ & $t_{\text{mitigation}} - t_{\text{avoidance}} \geqslant \Delta t_{\text{min,avoid}}$	Safe avoidance	S7
Safe	Avoidance, prevention, mitigation	If $t_{\text{avoidance}} - t_{\text{prevention}} < \Delta t_{\text{min,avoid}}$ or $t_{\text{avoidance}} - t_{\text{mitigation}} < \Delta t_{\text{min,avoid}}$	Safe false prevention or safe false Mitigation (see rules S5, S6)	S8
Disruptive	No alarms		Missed	D1
Disruptive	Avoidance	If $t_{\text{disr}} - t_{\text{avoidance}} \geqslant \Delta t_{\text{min,avoid}}$	Good avoidance	D2
Disruptive	Avoidance	If $t_{\text{disr}} - t_{\text{avoidance}} < \Delta t_{\text{min,avoid}}$	Tardy avoidance	D3
Disruptive	Prevention	If $t_{\text{disr}} - t_{\text{prevention}} \geqslant \Delta t_{\text{min,prevention}}$	Good prevention	D4
Disruptive	Prevention	If $t_{\text{disr}} - t_{\text{prevention}} < \Delta t_{\text{min,prevention}}$	Tardy prevention	D5
Disruptive	Mitigation	If $t_{\text{disr}} - t_{\text{mitigation}} \geqslant \Delta t_{\text{min,mitigation}}$	Good mitigation	D6
Disruptive	Mitigation	If $t_{\text{disr}} - t_{\text{mitigation}} < \Delta t_{\text{min,mitigation}}$	Tardy mitigation	D7
Disruptive	Prevention, mitigation	If $t_{\text{mitigation}} - t_{\text{prevention}} \geqslant \Delta t_{\text{min,prev}}$	Good prevention	D8
Disruptive	Prevention, mitigation	If $t_{\text{mitigation}} - t_{\text{prevention}} < \Delta t_{\text{min,prev}}$ & $t_{\text{disr}} - t_{\text{mitigation}} \geqslant \Delta t_{\text{min,mitigation}}$	Good mitigation	D9
Disruptive	Prevention, mitigation	If $t_{\text{mitigation}} - t_{\text{prevention}} < \Delta t_{\text{min,prev}}$ & $t_{\text{disr}} - t_{\text{mitigation}} < \Delta t_{\text{min,mitigation}}$	Tardy mitigation	D10
Disruptive	Avoidance, prevention, mitigation	If $t_{\text{prevention}} - t_{\text{avoidance}} \geqslant \Delta t_{\text{min,avoid}}$ & $t_{\text{mitigation}} - t_{\text{avoidance}} \geqslant \Delta t_{\text{min,avoid}}$	Good avoidance	D11
Disruptive	Avoidance, prevention, mitigation	If $t_{\text{avoidance}} - t_{\text{prevention}} < \Delta t_{\text{min,avoid}}$ or $t_{\text{avoidance}} - t_{\text{mitigation}} < \Delta t_{\text{min,avoid}}$	Good prevention, good mitigation, tardy prevention or tardy mitigation according with previous D rules.	D12

ORCID iDs

R. Rossi ⬤ https://orcid.org/0000-0003-4414-6119

M. Gelfusa ⬤ https://orcid.org/0000-0001-5158-7292

T. Craciunescu ⬤ https://orcid.org/0000-0002-0012-4260

I. Wyss ⬤ https://orcid.org/0000-0001-7625-8401

J. Vega ⬤ https://orcid.org/0000-0002-1622-3984

A. Murari ⬤ https://orcid.org/0000-0002-3932-3865

References

[1] Hender T.C. *et al* 2007 MHD stability, operational limits and disruptions *Nucl. Fusion* **47** S128–202

[2] Lehnen M. *et al* 2015 Disruptions in ITER and strategies for their control and mitigation *J. Nucl. Mater.* **463** 39–48

[3] De Vries P.C. *et al* 2012 The impact of the ITER-like wall at JET on disruptions *Plasma Phys. Control. Fusion* **54** 124032

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

[5] Lehnen M. 2017 Challenges of disruption mitigation in ITER *44th EPS Conf. on Plasma Physics*

[6] Lehnen M. 2019 Loads due to disruptions and prospects for mitigation *6th IAEA DEMO Programme Workshop*

[7] Schuller F.C. 1995 Disruptions in tokamaks *Plasma Phys. Control. Fusion* **37** A135–62

[8] Strait E.J. *et al* 2019 Progress in disruption prevention for ITER *Nucl. Fusion* **59** 112012

[9] Jachmich S. *et al* 2022 Shattered pellet injection experiments at JET in support of the ITER disruption mitigation system design *Nucl. Fusion* **62** 026012

[10] Li Y. *et al* 2021 Comparison of disruption mitigation from shattered pellet injection with massive gas injection on J-TEXT *Nucl. Fusion* **61** 126025

[11] Mailloux J. *et al* 2022 Overview of JET results for optimising ITER operation *Nucl. Fusion* **62** 042026

[12] Pucella G. *et al* 2021 Onset of tearing modes in plasma termination on JET: the role of temperature hollowing and edge cooling *Nucl. Fusion* **61** 046020

[13] Angioni C. *et al* 2014 Tungsten transport in JET H-mode plasmas in hybrid scenario, experimental observations and modelling *Nucl. Fusion* **54** 083028

[14] Puiatti M.E. *et al* 2002 Radiation pattern and impurity transport in argon seeded ELMy H-mode discharges in JET *Plasma Phys. Control. Fusion* **44** 1863–78

[15] Lascas Neto E., Graves J.P., Raghunathan M., Sommariva C. and Pfefferlé D. 2022 Heavy impurity transport in tokamaks subject to plasma rotation, NTV and the influence of

saturated ideal MHD perturbations *Plasma Phys. Control. Fusion* **64** 014002

[16] Vega J. *et al* 2016 Review of disruption predictors in nuclear fusion: classical, from scratch and anomaly detection approaches *IECON 2016–42nd Annual Conf. IEEE Industrial Electronics Society* (IEEE) (https://doi.org/10.1109/IECON.2016.7794130)

[17] Vega J. *et al* 2022 Disruption prediction with artificial intelligence techniques in tokamak plasmas *Nat. Phys.* **18** 741–50

[18] Peluso E., Rossi R., Murari A., Gaudio P. and Gelfusa M. 2020 Alternative detection of n = 1 modes slowing down on ASDEX upgrade *Applied Sciences* **10** 7891

[19] Sias G., Cannas B., Fanni A., Murari A. and Pau A. 2019 A locked mode indicator for disruption prediction on JET and ASDEX upgrade *Fusion Eng. Des.* **138** 254–66

[20] Kong M., Felici F., Sauter O., Galperti C., Vu T., Ham C.J., Hender T.C., Maraschek M. and Reich M. 2022 Physics-based control of neoclassical tearing modes on TCV *Plasma Phys. Control. Fusion* **64** 044008

[21] Agarwal A., Mishra A., Sharma P., Jain S., Daniel R., Ranjan S., Manchanda R., Ghosh J. and Tanna R. 2021 Deep sequence to sequence learning-based prediction of major disruptions in ADITYA tokamak *Plasma Phys. Control. Fusion* **63** 115004

[22] Murari A., Rossi R., Peluso E., Lungaroni M., Gaudio P., Gelfusa M., Ratta G. and Vega J. 2020 On the transfer of adaptive predictors between different devices for both mitigation and prevention of disruptions *Nucl. Fusion* **60** 056003

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

[24] Pautasso G. *et al* 2001 Prediction and mitigation of disruptions in ASDEX Upgrade *J. Nucl. Mater.* **290–293** 1045–51

[25] Cannas B., Fanni A., Pautasso G. and Sias G. 2011 Disruption prediction with adaptive neural networks for ASDEX Upgrade *Fusion Eng. Des.* **86** 1039–44

[26] Rea C., Montes K.J., Erickson K.G., Granetz R.S. and Tinguely R.A. 2019 A real-time machine learning-based disruption predictor in DIII-D *Nucl. Fusion* **59** 096016

[27] Rea C., Granetz R.S., Montes K., Tinguely R.A., Eidietis N., Hanson J.M. and Sammuli B. 2018 Disruption prediction investigations using machine learning tools on DIII-D and Alcator C-Mod *Plasma Phys. Control. Fusion* **60** 084004

[28] Montes K.J. *et al* 2019 Machine learning for disruption warnings on Alcator C-Mod, DIII-D, and EAST *Nucl. Fusion* **59** 096015

[29] Zheng W. *et al* 2018 Hybrid neural network for density limit disruption prediction and avoidance on J-TEXT tokamak *Nucl. Fusion* **58** 056016

[30] Gerhardt S.P., Darrow D.S., Bell R.E., LeBlanc B.P., Menard J.E., Mueller D., Roquemore A.L., Sabbagh S.A. and Yuh H. 2013 Detection of disruptions in the high- β spherical torus NSTX *Nucl. Fusion* **53** 063021

[31] Bao N.N., Huang Y., Xiao B.J., Yuan Q.P., Zhuang H.D., Luo Z.P., Wang Y.H. and Zhang R.R. 2020 A real-time disruption prediction tool for VDE on EAST *IEEE Trans. Plasma Sci.* **48** 715–20

[32] Vega J., Hernández F., Dormido-Canto S., Isayama A., Joffrin E., Matsunaga G. and Suzuki T. 2019 Assessment of linear disruption predictors using JT-60U data *Fusion Eng. Des.* **146** 1291–4

[33] Yoshino R. 2003 Neural-net disruption predictor in JT-60U *Nucl. Fusion* **43** 1771–86

[34] Murari A., Lungaroni M., Peluso E., Gaudio P., Vega J., Dormido-Canto S., Baruzzo M. and Gelfusa M. 2018 Adaptive predictors based on probabilistic SVM for real time disruption mitigation on JET *Nucl. Fusion* **58** 056002

[35] Murari A., Vagliasindi G., Arena P., Fortuna L., Barana O. and Johnson M. 2008 Prototype of an adaptive disruption predictor for JET based on fuzzy logic and regression trees *Nucl. Fusion* **48** 035010

[36] Rattá G.A., Vega J., Murari A. and Gadariya D. (JET Contributors) 2021 PHAD: a phase-oriented disruption prediction strategy for avoidance, prevention, and mitigation in JET *Nucl. Fusion* **61** 116055

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

[38] Rattá G., Vega J., Murari A. and Gadariya D. (JET Contributors) 2021 PHAD: a phase-oriented disruption prediction strategy for avoidance, prevention, and mitigation in JET *Nucl. Fusion* **61** 116055

[39] Murari A., Rossi R., Lungaroni M., Baruzzo M. and Gelfusa M. 2021 Stacking of predictors for the automatic classification of disruption types to optimize the control logic *Nucl. Fusion* **61** 036027

[40] Esquembri S., Vega J., Murari A., Ruiz M., Barrera E., Dormido-Canto S., Felton R., Tsalas M. and Valcarcel D. 2018 Real-time implementation in JET of the SPAD disruption predictor using MARTe *IEEE Trans. Nucl. Sci.* **65** 836–42

[41] Fontana M., Challis C.D., Conway N.J., Felton R., Goodyear A., Hogben C., Peacock A. and Schmuck S. 2020 Real-time applications of electron cyclotron emission interferometry for disruption avoidance during the plasma current ramp-up phase at JET *Fusion Eng. Des.* **161** 111934

[42] Pau A. *et al* 2018 A first analysis of JET plasma profile-based indicators for disruption prediction and avoidance *IEEE Trans. Plasma Sci.* **46** 2691–8

[43] Aymerich E., Sias G., Pisano F., Cannas B., Carcangiu S., Sozzi C., Stuart C., Carvalho P.J. and Fanni A. (JET Contributors) 2022 Disruption prediction at JET through deep convolutional neural networks using spatiotemporal information from plasma profiles *Nucl. Fusion* **62** 066005

[44] Pau A., Fanni A., Carcangiu S., Cannas B., Sias G., Murari A. and Rimini F. 2019 A machine learning approach based on generative topographic mapping for disruption prevention and avoidance at JET *Nucl. Fusion* **59** 106017

[45] Vega J., Dormido-Canto S., López J.M., Murari A., Ramírez J.M., Moreno R., Ruiz M., Alves D. and Felton R. 2013 Results of the JET real-time disruption predictor in the ITER-like wall campaigns *Fusion Eng. Des.* **88** 1228–31

[46] Vega J., Murari A., Dormido-Canto S., Hernández F., Cruz T., Gadariya D. and Rattá G.A. 2020 A linear equation based on signal increments to predict disruptive behaviours and the time to disruption on JET *Nucl. Fusion* **60** 026001

[47] Murari A., Peluso E., Lungaroni M., Rossi R. and Gelfusa M. 2020 Investigating the physics of tokamak global stability with interpretable machine learning tools *Appl. Sci.* **10** 6683

[48] de Vries P.C. *et al* 2016 Scaling of the MHD perturbation amplitude required to trigger a disruption and predictions for ITER *Nucl. Fusion* **56** 026007

[49] Sabbagh S.A. *et al* 2023 Disruption event characterization and forecasting in tokamaks *Phys. Plasmas* **30** 032506

[50] Rattá G.A., Vega J. and Murari A. 2019 A multidimensional linear model for disruption prediction in JET *Fusion Eng. Des.* **146** 2393–6

[51] Chirikov B.V. 1979 A universal instability of many-dimensional oscillator systems *Phys. Rep.* **52** 263–379

[52] Rossi R., Gelfusa M., Flanagan J. and Murari A. 2022 Development of robust indicators for the identification of electron temperature profile anomalies and application to JET *Plasma Phys. Control. Fusion* **64** 045002

[53] Gelfusa M., Craciunescu T., Peluso E., Giacomelli L., Kiptily V., Reux C., Szepesi G., Murari A. and Contributors J. 2021 A maximum likelihood tomographic method applied to JET gamma ray emission during the current quench *Fusion Eng. Des.* **168** 112637

[54] Craciunescu T., Bonheure G., Kiptily V., Murari A., Soare S., Tiseanu I. and Zoita V. 2008 The maximum likelihood reconstruction method for JET neutron tomography *Nucl. Instrum. Methods Phys. Res.* A **595** 623–30

[55] Craciunescu T., Peluso E., Murari A. and Gelfusa M. 2018 Maximum likelihood bolometric tomography for the determination of the uncertainties in the radiation emission on JET TOKAMAK *Rev. Sci. Instrum.* **89** 053504

[56] Peluso E., Craciunescu T., Murari A., Carvalho P. and Gelfusa M. 2019 A comprehensive study of the uncertainties in bolometric tomography on JET using the maximum likelihood method *Rev. Sci. Instrum.* **90** 123502

[57] Ruiz M., Nieto J., Costa V., Craciunescu T., Peluso E., Vega J. and Murari A. 2022 Acceleration of an algorithm based on the maximum likelihood bolometric tomography for the determination of uncertainties in the radiation emission on JET using heterogeneous platforms *Appl. Sci.* **12** 6798

[58] Wyss I., Murari A., Spolladore L., Peluso E., Gelfusa M., Gaudio P. and Rossi R. (JET-contributors) 2023 Comparison of a fast low spatial resolution inversion method and peaking factors for the detection of anomalous radiation patterns and disruption prediction *Fusion Eng. Des.* **193** 113625

[59] Matos F.A., Ferreira D.R. and Carvalho P.J. 2017 Deep learning for plasma tomography using the bolometer system at JET *Fusion Eng. Des.* **114** 18–25

[60] Ferreira D.R., Carvalho P.J., Carvalho I.S., Stuart C.I. and Lomas P.J. 2022 Investigating the physics of disruptions with real-time tomography at JET *Plasma Sci. Technol.* **24** 035103

[61] Rossi R. *et al* 2023 A systematic investigation of radiation collapse for disruption avoidance and prevention on JET tokamak *Matter Radiat. Extremes* **8** 046903

[62] Lawson C.L. and Hanson R.J. 1995 Solving least squares problems *Society for Industrial and Applied Mathematics* (https://doi.org/10.1137/1.9781611971217)

[63] Riccardo V. *et al* 2010 JET disruption studies in support of ITER *Plasma Phys. Control. Fusion* **52** 124018

[64] Loarte A. *et al* 2007 Transient heat loads in current fusion experiments, extrapolation to ITER and consequences for its operation *Phys. Scr.* **T128** 222–8

[65] La Haye R.J., Prater R., Buttery R.J., Hayashi N., Isayama A., Maraschek M.E., Urso L. and Zohm H. 2006 Cross–machine benchmarking for ITER of neoclassical tearing mode stabilization by electron cyclotron current drive *Nucl. Fusion* **46** 451–61

[66] Poli F.M., Fredrickson E.D., Henderson M.A., Kim S.-H., Bertelli N., Poli E., Farina D. and Figini L. 2018 Electron cyclotron power management for control of neoclassical tearing modes in the ITER baseline scenario *Nucl. Fusion* **58** 016007

[67] Piron L. *et al* 2021 Progress in preparing real-time control schemes for deuterium-tritium operation in JET *Fusion Eng. Des.* **166** 112305

[68] Sozzi C. *et al* Termination of discharges in high performance scenarios in JET *28th IAEA Fusion Energy Conf. (FEC2020)*

[69] Piron L. *et al* 2023 Radiation control in deuterium, tritium and deuterium-tritium JET baseline plasmas—part I *Fusion Eng. Des.* **193** 113634

[70] Piron L. *et al* 2023 Radiation control in tritium and deuterium-tritium JET baseline plasmas—part II *Fusion Eng. Des.* **192** 113695

[71] Devore J L 2011 *Probability and Statistics for Engineering and the Sciences* 8th edn (Cengage Learning) pp 508–10

[72] Jachmich S., Kruezi U., Card P., Deakin K., Kinna D., Koslowski H.R., Lambertz H.T. and Lehnen M. 2015 Implementation of a new disruption mitigation system into the control system of JET *Fusion Eng. Des.* **96–97** 633–6

[73] Lehnen M. *et al* 2011 Disruption mitigation by massive gas injection in JET *Nucl. Fusion* **51** 123010