Università degli Studi di Padova
CENTRO INTERDIPARTIMENTALE *"Centro Ricerche Fusione"*

Universidade Técnica de Lisboa
INSTITUTO SUPERIOR TÉCNICO

JOINT RESEARCH DOCTORATE IN FUSION SCIENCE AND ENGINEERING
CYCLE XXVI

Techniques for prediction of disruptions on TOKAMAKS

Coordinator: Ch.mo Prof. Piero Martin
Supervisor: Ch.mo Prof. Paolo Bettini
Supervisor: Ch.ma Prof.ssa Alessandra Fanni

Doctoral student: Alessandro Pau

Introduction

The physicist Andreevich Artsimovich in the 1970 wrote that "thermonuclear [fusion] energy will be ready when mankind needs it". Considering the actual world energy situation and the effect on the environment due to the present harnessing of the different sources of energy, the hope is that time for fusion is finally arrived.

Background and Motivation

The activities carried out in the framework of this thesis regarded the development, implementation and application of algorithms for classification and prediction of disruptions in Tokamaks.

The balance of plasmas in a magnetic field can be described by the theory of magneto-hydro-dynamic (MHD). MHD instabilities are among the most serious factors that limit fusion devices operation in magnetic confinement configurations. When they occur on a large scale can degrade the performance of the plasma and lead to loss of confinement and control.

A disruption is a sudden loss of stability or confinement of tokamak plasma; it is a critical event in which the plasma energy is lost within a time span of few milliseconds exposing the plasma facing components to severe thermo-mechanical stresses and conductors surrounding the vessel to huge electromagnetic forces. Therefore, it becomes of primary importance to avoid or mitigate disruptions in order to preserve the integrity of the machine. This aspect and the understanding of disruptive phenomena play a key role in design and running of new experimental devices as ITER, currently under construction in Cadarache (France), which will have the task of demonstrating the feasibility of fusion energy production from a technical and engineering point of view.

These considerations motivate a strong interest in developing methods and techniques aimed to minimize both number and severity of disruptions. Furthermore when a disruption occurs it would be particularly important to be able to distinguish among its different types in order to improve avoidance

and mitigation strategies. Since physical models able to reliably recognize and predict the occurrence of disruptions are currently not available, the research carried out fits in the broad framework of machine learning techniques that have been exploited as an alternative approach to disruption prediction and automatic classification.

Promising approaches to prediction and classification are represented by the so-called "data-based" methods: to this purpose, existing systems have been applied and further developed and new approaches have been investigated.

The mentioned activity has been carried out in collaboration with the University of Cagliari and European Research Centers for nuclear fusion, taking as case study some of the most important experimental machines such as JET and ASDEX Upgrade (AUG), with several months of research spent at the Culham Science Centre.

List of publications and contributions of this thesis

[1] B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias, and JET-EFDA Contributors, "Manifold learning to interpret JET high-dimensional operational space", Plasma Phys. Control. Fusion 55 045006, 2013.

[2] B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias, and JET-EFDA Contributors, "*Automatic disruption classification based on manifold learning for real-time applications on JET*", Nuclear Fusion 53 093023, 2013.

[3] A. Pau, B. Cannas, A. Fanni, A. Murari, G. Sias, and JET-EFDA Contributors, "*Advances in disruption classification at JET*", 8th Workshop on Fusion Data Processing, Validation and Analysis, November 4-6, 2013, Ghent (Belgium).

[4] G. Sias, R. Aledda, B. Cannas, A. Fanni, A. Pau, G. Pautasso, and ASDEX Upgrade Team, "*Data visualization and dimensionality reduction methods for disruption prediction at ASDEX Upgrade*", 8th Workshop on Fusion Data Processing, Validation and Analysis, November 4-6, 2013, Ghent (Belgium).

[5] B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias, and JET-EFDA Contributors, "*Overview of manifold learning techniques for the investigation of*

disruptions on JET", JET Pin-board, to be submitted to Plasma Phys. Control. Fusion - "Physics-based optimization of plasma diagnostic information" Cluster.

[6] A. Murari, J. Vega, P. Boutot, B. Cannas, S. Dormido-Canto, A. Fanni, J. M. Lopez, R. Moreno, A. Pau, G. Sias, J. M.Ramirez, G. Verdoolaege, ASDEX Upgrade Team and JET EFDA contributors, "*Latest Developments in Data Analysis Tools for Disruption Prediction and for the Exploration of Multimachine Operational Spaces*", Proc. of 24th IAEA Fusion Energy Conference Abstracts, San Diego, USA, 8-13 October 2012.

One of the most demanding activities, especially in terms of required time, has been the building of representative and reliable databases which results to be fundamental for successfully apply data-driven methods. For AUG a database was already available and it is constantly updated by the researchers of University of Cagliari.

Regarding JET, in order to analyze and investigate its high-dimensional operational space, a reliable database has been built up on the base of 10 real time signals, which are representative of the disruptive behavior of the plasma. For the Carbon Wall (CW) data comes from plasma discharges selected from JET campaigns from 2005 to 2009, whereas ITER-like Wall (ILW) database is based on the same set of signals belonging to the campaigns from 2011 to 2013.

Several criteria and statistical analysis have been considered in order to properly select a reduced representative number of discharges. Different data reduction algorithm have been developed in order to obtain a reasonable amount of data, keeping at the same time the diversity and the representativeness of data in statistical terms. Only non-intentional disruptions have been considered with plasma current above 1MA. The resulting CW database is composed of 243 disruptions, whereas ILW database consists of 149 disruptions, where each signal has been sampled at a frequency of 1 kHz.

The high dimensional operational space of JET has been analyzed and visualized using different linear projection methods such as Grand Tour (GT) and Principal Component Analysis (PCA), and mapped through non-linear manifold learning techniques as Self-Organizing Map (SOM) and Generative Topographic Map (GTM). The use of the "Manifold Learning" finds its motivation in the fact that high-dimensional data can lie on an embedded, eventually non-linear, low-dimensional manifold, which can be easily visualized and understood if we consider a 2 or 3 dimensional space. Hence, in this PhD Thesis, Manifold Learning methods have been successfully applied both for classification and prediction of disruptions, showing their potentiality in

the analysis and the visualization of the operational space.

The SOM and/or GTM maps can be exploited to identify characteristic regions of the plasma scenario and for discriminating between regions with high risk of disruption and those with low risk of disruption. This part of the work has been supported with the implementation of tools for data analysis and data visualization with which it is also possible to quantify and evaluate the effectiveness of the mapping itself. The results show quite clearly that nonlinear manifold learning techniques are more suitable for mapping the JET high dimensional operational space; in particular GTM exhibits a higher capability of discriminating between safe and disruptive regions [1]. An important result of this analysis is represented by the fact that the two nonlinear methods seem to converge on the same manifold, which means that we are actually looking at the intrinsic properties hidden in the high dimensional data.

The tools developed for data analysis and visualization, in particular for GTMs, could be particularly useful in the study of the operational space where the relevant physics takes place, allowing the perception of eventual similarities among the different variables. The identification of dependencies and complex relations among the variables is made possible by analysis and comparison of similar patterns in the relative component distributions of the input variables onto the 2-D maps.

By applying such techniques, another relevant part of the Ph.D. activities has been spent in the analysis of the different types of disruption that can occur in JET, making reference to the manual classification that has been done in [*P.C. de Vries, et al., Nucl. Fusion 51 (2011) 053018*], where specific chains of events have been detected and used to classify disruptions, grouping those that follow specific paths. The classification is part of a particularly complex scenario whose analysis has required a considerable amount of time. The characterization of the operational space in terms of the different disruption classes may lead to better overall understanding and more focussed prevention and mitigation methods. A preliminary analysis carried out both with SOMs and GTMs has shown that the maps seem to self-organize in such a way that the disruptions which belong to the same class tend to aggregate, defining in this way regions where a certain class results to be predominant with respect to the others.

As described in [2], the potentiality of the GTM mapping of the JET operational space has been exploited to develop an automatic disruption classification of seven disruption classes occurred with the Carbon Wall. Each disruption is projected on the map and the probabilities of belonging to the different disruption classes are monitored during the time evolution, returning the class that the disrupted pulse more likely belongs to. Using the GTM

trained on the CW dataset with ILW discharges selected from the JET ILW campaigns C28-C30 significantly deteriorates the classification performance, particularly on certain disruption classes. Hence, a new GTM has been built to represent the new operational space of JET. Using this map with all the disruptions occurred in these last campaigns the very high classification performance is confirmed and therefore, the prospects for the deployment of this tool in real time are very promising [3].

GTMs potentiality has also been exploited for the prediction of disruptions at ASDEX Upgrade [4]: a 2-D-GTM has been built to represent the 7D AUG operational space on the base of discharges performed between May 2007 and April 2011. As it has been obtained in the case of JET, the GTM clearly proves to be able to separate non-disruptive and disruptive states of plasma. Therefore, likewise the SOM, the GTM can be used as a disruption predictor by tracking the temporal sequence of the samples on the map, depicting the movement of the operating point during a discharge. Various criteria have been studied to associate the risk of disruption of each map region with a disruption alarm threshold. The prediction performance of the proposed predictive system has been evaluated on a set of discharges coming from experimental campaigns carried out at AUG from May 2011 to November 2012.

Some measures have been used to evaluate the performance of the proposed methodologies. To calculate the precision of the clustering over the entire dataset, the average quantization error, a common index of the map resolution, has been applied. Moreover, to control the conservation of topology two different aspects have been analyzed, i.e., the trustworthiness of the projected neighborhood and the preservation of the resulting neighborhood. Moreover, an outlier analysis has been performed on the available data in order to quantify goodness and effectiveness of the projection [5].

In the last years, significant efforts have been devoted to the development of advanced data analysis tools to both predict the occurrence of disruptions and to investigate the operational spaces of devices, with the long term goal of advancing the understanding of the physics of these events and to prepare for ITER. Manifold learning tools are also producing very interesting results in the comparative analysis of JET and AUG operational spaces, on the route of developing predictors capable of extrapolating from one device to another, as foreseen in the framework of cross-machine approach [6].

Outline of the Thesis

In **chapter 1** the perspectives of fusion in the world energy context as an almost unlimited source of energy for the future are discussed, with particular reference to the role of magnetic confinement. Furthermore, the bases of fusion reactions have been introduced.

In **chapter 2** the main aspects of plasma stability in tokamaks configurations are described with the aim to provide an adequate reference for all the discussions of the following chapters. In particular, the main parameters related to plasma stability, which have been used for the construction of the databases, have been introduced.

The **chapter 3** is focused on the description of the operational limits with reference to the main quantities which should be maximized to improve plasma performance. Everything, also in the previous chapters, has been framed to introduce the key problems which this thesis has addressed: analysis, prediction and classification of disruptions. After the main considerations about the operational limits, the main phases, the causes and the consequences of disruptions have been discussed, trying to integrate the stability concepts introduced in the previous chapter.

The **chapter 4** is finalized to provide an insight of the Machine Learning methods which represent the starting point of all the analysis and algorithms implemented for disruption prediction and classification. Today the large amount of data available from fusion experiments and their character of high-dimensionality make particularly difficult handling, processing, understanding and extracting properly what is really important among all the available information. Machine Learning allows to deal with the problem in efficient way. Therefore, a framework of all the techniques exploited for the analysis has been provided, with particular reference to the Manifold Learning algorithms as Self Organizing Maps (SOMs) and Generative Topographic Mappings (GTMs). Also reference methods such as k-Nearest Neighbor (k-NN) or more recent methods such as Conformal Predictors, exploited for validation and reliability assessment purposes, have been described.

In **chapter 5** the state of the art of machine learning techniques applied to disruption prediction and classification is presented, describing in particular the main applications with the widely employed Neural Networks, such Multi Layer Perceptrons (MLPs), Support Vector Machines (SVMs) and Self Organizing Maps (SOMs), and statistical methods such as Discriminant Analysis or Multiple Threshold technique. Strengths and weaknesses have also been discussed with reference to a possible solution to overcome the drawbacks of these methods: the multi-machine approach.

Chapter 6 is dedicated to the description of the databases used for all

the analysis presented in the following chapters. In particular, the statistical analysis and the data-reduction algorithms that have been needed to build a reliable and statistically representative database have been discussed in detail.

The last three chapters contain all the analysis and all the algorithms implemented for the mapping of the operational space, disruption classification and prediction. In **chapter 7** the mapping of the JET operational space is presented. The first sections deal with projections and data-visualization with linear projection methods such as Grand Tour (GT) and Principal Component Analysis (PCA). In the central part, the same aspects have been taken into account by exploiting nonlinear Manifold Learning techniques, SOM and GTM, on the base of which a detailed analysis of the operational space has been performed. Such analysis, showing the potentiality of the methods, has been performed, regarding GTM model, through the implementation of a dedicated tool. Finally, an outliers' analysis and performance indexes appositely proposed have been considered for evaluating the overall performance of the mapping.

In the **chapter 8** the developed automatic disruption classification for JET has been described. The chapter is divided in two parts: the first one describes the classification of disruptions belonging to the Carbon Wall (CW) campaigns, whereas in the second part the classification of disruptions with the ITER-like Wall (ILW) is framed in the assessment of the suitability of the automatic classifier for real time applications, in conjunction with prediction systems working online at JET. The reliability of the results has been validated by comparison with a k-NN based reference classifier and through the recent conformal predictors, with which is possible to provide, in addition to the prediction/classification, the related level of confidence.

Chapter 9 is dedicated to the disruption prediction at ASDEX Upgrade. The first part is related to the description of the database and the data-reduction technique used to select a representative and balanced dataset. Self-Organizing Map and the Generative Topographic Mapping have been exploited to map ASDEX Upgrade operational space and to build a disruption predictor, introducing at the same time their potentiality for disruptions classification. Furthermore, the use of this two methods combined with a Logistic model has been proposed to realize a predictive system able to exploit the complementary behaviors of the two approaches, improving the overall performance in prediction.

Prefazione

Il fisico Andreevich Artsimovich nel 1970 scrisse che "l'energia da fusione nucleare sarà disponibile quando l'umanità ne avrà bisogno". Considerando l'attuale scenario energetico mondiale e l'impatto sull'ambiente dovuto allo sfruttamento delle diverse risorse energetiche, la speranza è che quel momento sia finalmente arrivato.

Background e Motivazione

Le attività svolte nell'ambito di questa tesi hanno riguardato lo sviluppo, l'implementazione e l'applicazione di algoritmi per la classificazione e la predizione di disruzioni nei Tokamak.

L'equilibrio dei plasmi nei campi magnetici può essere descritto dalla teoria magneto-idro-dinamica (MHD). Le instabilità MHD sono tra i fattori che limitano più seriamente le operazioni nelle macchine a fusione a confinamento magnetico.

Una disruzione è un'improvvisa perdita di stabilità e di confinamento nei tokamak; è un evento critico durante il quale l'energia immagazzinata nel plasma viene persa nell'arco di pochi millisecondi, esponendo i componenti della parete interna della camera da vuoto a severi stress termo-meccanici, e i conduttori circostanti a enormi forze elettromagnetiche. Quindi diventa di primaria importanza l'avoidance e la mitigazione delle disruzioni al fine di preservare l'integrità della macchina. Questo aspetto e la comprensione dei fenomeni disruttivi giocano un ruolo chiave nel progetto e nel funzionamento delle nuove macchine sperimentali come ITER, attualmente in costruzione a Cadarache (Francia), la quale avrà la finalità di dimostrare la fattibilità tecnica ed ingegneristica della produzione di energia da fusione.

Queste considerazioni motivano un forte interesse nello sviluppo di metodi e tecniche atti a minimizzare sia il numero che l'entità delle disruzioni. Inoltre, quando si verifica una disruzione, sarebbe veramente importante riuscire a distinguere tra i diversi tipi di disruzione, al fine di migliorare le strategie di avoidance e mitigazione. Dal momento che ad oggi non esistono mod-

elli fisici in grado di riconoscere e predire in maniera affidabile l'arrivo di una disruzione, la ricerca portata avanti in questi anni si integra nel più ampio contesto delle tecniche di Machine Learning, le quali sono state utilizzate come approccio alternativo alla predizione ed alla classificazione automatica delle disruzioni.

Approcci promettenti alla predizione ed alla classificazione sono rappresentati dai cosidetti approcci "data-based": a questo proposito sono state applicate e ulteriormente sviluppate diverse tecniche, e si è indagato su nuovi approcci.

Le attività citate sono state svolte in collaborazione con l'Università di Cagliari e importanti centri di ricerca europei sulla fusione, prendendo in esame alcune delle più importanti macchine sperimentali, quali il JET (Regno Unito) e ASDEX Upgrade (Germania), con diversi mesi trascorsi al Culham Science Centre (Abingdon, Regno Unito).

Elenco delle pubblicazioni e dei contributi legati alla tesi

[1] B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias, and JET-EFDA Contributors, "Manifold learning to interpret JET high-dimensional operational space", Plasma Phys. Control. Fusion 55 045006, 2013.

[2] B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias, and JET-EFDA Contributors, "*Automatic disruption classification based on manifold learning for real-time applications on JET*", Nuclear Fusion 53 093023, 2013.

[3] A. Pau, B. Cannas, A. Fanni, A. Murari, G. Sias, and JET-EFDA Contributors, "*Advances in disruption classification at JET*", 8th Workshop on Fusion Data Processing, Validation and Analysis, November 4-6, 2013, Ghent (Belgium).

[4] G. Sias, R. Aledda, B. Cannas, A. Fanni, A. Pau, G. Pautasso, and ASDEX Upgrade Team, "*Data visualization and dimensionality reduction methods for disruption prediction at ASDEX Upgrade*", 8th Workshop on Fusion Data Processing, Validation and Analysis, November 4-6, 2013, Ghent (Belgium).

[5] B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias, and JET-EFDA Contributors, "*Overview of manifold learning techniques for the investigation of*

disruptions on JET", JET Pin-board, to be submitted to Plasma Phys. Control. Fusion - "Physics-based optimization of plasma diagnostic information" Cluster.

[6] A. Murari, J. Vega, P. Boutot, B. Cannas, S. Dormido-Canto, A. Fanni, J. M. Lopez, R. Moreno, A. Pau, G. Sias, J. M.Ramirez, G. Verdoolaege, ASDEX Upgrade Team and JET EFDA contributors, "*Latest Developments in Data Analysis Tools for Disruption Prediction and for the Exploration of Multimachine Operational Spaces*", Proc. of 24th IAEA Fusion Energy Conference Abstracts, San Diego, USA, 8-13 October 2012.

Una delle attività più onerese, specie in termini di tempo, è stata la costruzione di un database affidabile e rappresentativo, che risulta essere fondamentale per un'applicazione coerente dei metodi "data-driven". Nel caso di ASDEX Upgrade un database era già disponibile e costantemente aggiornato da ricercatori dell'Università di Cagliari.

Per quanto riguarda il JET invece, al fine di analizzare il suo spazio operativo ad altà dimensionalità, è stato costruito un database affidabile sulla base di dieci segnali disponibili in tempo reale, che sono rappresentativi del comportamento disruttivo del plasma. Per la parete in carbone (CW), i dati provengono dalle campagne sperimentali che vanno dal 2005 al 2009, mentre per quanto riguarda la parete metallica (ILW), il database è basato sugli stessi segnali relativi agli esperimenti delle campagne dal 2011 al 2013.

Sono inoltre stati valutati diversi criteri e analisi statistiche al fine di selezionare in maniera appropriata un numero ridotto di scariche. Sono inoltre sono stati sviluppati diversi algoritmi di data-reduction al fine di ottenere una quantità di dati ragionevole, preservando al tempo stesso la diversità e la rappresentatività del database in termini statistici. Sono state considerate unicamente le disruzioni non intenzionali con una corrente di plasma non inferiore ad 1MA. Il risultante database per la parete in carbone è composto da 243 disruzioni, mentre quello relativo alla parete metallica é costituito da 149 impulsi disrotti, per i quali ciascun segnale è campionato alla frequenza di 1kHz.

L'analisi e la visualizzazione dello spazio ad alta dimensionalità di JET è stata ottenuta sia con metodi di proiezione lineari, quali il Grand Tour (GT) e la Principal Component Analysis (PCA), che con metodi di proiezione non lineari, detti di "Manifold Learning", quali la Self Organizing Maps (SOM) e le Generative Topographics Mappping (GTM). L'uso del Manifold Learning trova la sua ragion d'essere nel fatto che dati ad alta dimensionalità possono giacere in una struttura o spazio eventualmente non lineare a minore dimensionalità che può essere facilmente visualizzato e "compreso" se si considera

uno spazio 2-D o 3-D. Dunque, in questa tesi, le tecniche di Manifold Learning sono state applicate con successo per la predizione e la classificazione di disruzioni, evidenziando in particolare le loro potenzialità nell'analisi e nella visualizzazione degli spazi operativi.

Le mappe SOM e GTM possono essere utilizzate per identificare regioni caratteristiche e per discriminare tra quelle con alto e quelle con basso rischio di disruzione. Questa parte del lavoro è stato supportata con l'implementazione di strumenti per l'analisi e la visualizzazione dei dati, con cui è anche possibile quantificare e valutare l'efficacia del mapping stesso. I risultati mostrano chiaramente che le tecniche di Manifold Learning non lineari si dimostrano più adeguate nel mapping dello spazio operativo di JET ad alta dimensionalità; in particolare, la GTM presenta una maggiore capacità di discriminazione tra regioni "safe" e regioni disrotte [1]. Un importante risultato di questa analisi è rappresentato dal fatto che i due metodi non lineari sembrano convergere nell'identificazione dello stesso manifold, il che significa che stiamo realmente osservando le proprietà intrinseche nascoste nei dati ad alta dimensionalità.

Il tool sviluppato per la data-analysis e la visualizzazione, in particolare per le GTM, potrebbe essere particolarmente utile nello studio dello spazio operativo dove la fisica di rilievo, relativamente ai fenomeni considerati, ha luogo, consentendo l'individuazione di eventuali similarità tra le diverse variabili. L'identificazione di particolari dipendenze è resa possibile dall'analisi di pattern simili nella distribuzione relativa delle variabili in ingresso al sistema nelle mappe 2-D.

Una parte rilevante delle attività di dottorato è stata spesa applicando queste tecniche all'analisi dei diversi tipi di disruzioni che possono verificarsi al JET, facendo riferimento alla classificazione manuale che è descritta in [*P.C. de Vries, et al., Nucl. Fusion 51 (2011) 053018*], dove specifiche catene di eventi sono state individuate e utilizzate per classificare le disruzioni. La classificazione è parte di uno scenario particolarmente complesso per la cui analisi è stata molto onerosa in termini di tempo. La caratterizzazione dello spazio operativo in termini di diverse classi di disruzione può portare ad una migliore comprensione globale del fenomeno, nonchè a metodi di prevenzione e mitigazione più mirati. Un'analisi preliminare condotta sia con SOM che con GTM ha dimostrato che le mappe sembrano auto-organizzarsi in modo tale che le disruzioni che appartengono alla stessa classe tendono ad aggregarsi, definendo in questo modo delle regioni in cui una determinata classe risulta essere predominante rispetto alle altre.

Come descritto in [2], le potenzialità del mapping dello spazio operativo di JET con le GTM, possono essere sfruttate per sviluppare una classificazione automatica relativamente alle sette classi di disruzione considerate

per la parete in carbone. Ogni disruzione viene proiettata sulla mappa e le probabilità di appartenenza alle diverse classi vengono monitorate durante l'evoluzione nel tempo, restituendo la classe più probabile per l'impulso considerato. Utilizzando la stessa mappa GTM per scariche con la parete metallica (ILW), selezionate dalle campagne sperimentali C28-C30, si è trovato che le prestazioni in classificazione si deteriorano in modo significativo, specie per quanto riguarda alcune classi di disruzione. Quindi, è stata costruita una nuova mappa GTM per rappresentare il nuovo spazio operativo del JET. Utilizzando questa mappa sulla totalità delle disruzioni verificatesi nelle ultime campagne, si ritrovano prestazioni molto elevate in classificazione, e quindi le prospettive per l'utilizzo di questo strumento in tempo reale, sono molto promettenti [3].

Le potenzialità delle GTM sono state utilizzate anche per la predizione delle disruzioni ad ASDEX Upgrade [4]: è stata realizzata una mappa GTM 2-D per rappresentare lo spazio operativo 7D di ASDEX relativamente alle scariche effettuate tra maggio 2007 e aprile 2011. Come si è ottenuto nel caso del JET, la GTM dimostra chiaramente di essere in grado di discriminare tra gli stati disruttivi e quelli non disruttivi del plasma. Pertanto, analogamente alla SOM, la GTM può essere usata come predittore di disruzioni monitorando la sequenza temporale dei campioni sulla mappa, che raffigura l'evoluzione del punto di lavoro durante una scarica. Diversi criteri sono stati studiati per associare il rischio di disruzione di ogni regione della mappa con una specifica soglia di allarme. Le performance del sistema predittivo proposto sono state valutate su una serie di scariche provenienti dalle campagne sperimentali effettuate ad ASDEX da maggio 2011 al novembre 2012.

Per valutare le prestazioni delle metodologie di mapping proposte sono stati utilizzati alcuni indicatori. Per calcolare la precisione del clustering è stato valutato l'errore medio di quantizzazione sull'intero insieme di dati, un indice comunemente utilizzato per la valutazione della risoluzione delle mappe. Inoltre, per verificare la conservazione della topologia, sono stati analizzati due aspetti differenti, ossia l'affidabilità del vicinato mappato e la conservazione del vicinato originale. Inoltre, è stata eseguita sui dati disponibili l'analisi degli outlier, al fine di quantificare la bontà e l'efficacia della proiezione [5].

Negli ultimi anni sono stati dedicati notevoli sforzi allo sviluppo di strumenti di analisi avanzata dei dati, sia per predire il verificarsi di una disruzione, sia per studiare gli spazi operativi delle macchine, con l'obiettivo a lungo termine di far progredire la comprensione della fisica che sta dietro a questi eventi in vista di ITER. Gli strumenti di Manifold Learning stanno producendo risultati molto interessanti anche per quanto concerne l'analisi

comparativa degli spazi operativi di JET e ASDEX, in relazione alla sviluppo di predittori in grado di estrapolare da un macchina ad un altra, come previsto nell'ambito della strategia cross-machine [6].

Outline della Tesi

Nel **capitolo 1** vengono discusse le prospettive della fusione nel contesto energetico mondiale come fonte quasi illimitata di energia per il futuro, con particolare riferimento al ruolo del confinamento magnetico. Inoltre, sono state introdotte le basi sulle reazioni di fusione.

Nel **capitolo 2** vengono descritti gli aspetti principali della stabilità del plasma nelle configurazioni tokamak, con l'obiettivo di fornire un riferimento adeguato per tutte le discussioni dei capitoli successivi. In particolare vengono introdotti i principali parametri relativi alla stabilità del plasma, che sono stati utilizzati per la costruzione dei database.

Il **capitolo 3** è incentrato sulla descrizione dei limiti operativi con riferimento ai principali parametri che dovrebbero essere ottimizzati per migliorare le performance del plasma. Tutto, anche nei capitoli precedenti, è contestuale all'introduzione dei principali problemi che questa tesi si pone l'obiettivo di affrontare: analisi, predizione e classificazione delle disruzioni. Dopo le considerazioni sui limiti operativi, vengono discusse le fasi principali, le cause e le conseguenze dei processi disruttivi, cercando di integrarvi i concetti sulla stabilità introdotti nel capitolo precedente.

Il **capitolo 4** è invece finalizzato a fornire una panoramica sui metodi di Machine Learning che rappresentano il punto di partenza per tutte le analisi e gli algoritmi implementati per la predizione e la classificazione delle disruzioni. Oggi la grande quantità di dati disponibili dagli esperimenti sulla fusione e il loro carattere di alta dimensionalità, rendono particolarmente difficile la gestione, l'elaborazione, la comprensione e l'estrazione di quelle informazioni che sono veramente importanti tra tutte quelle disponibili.
Il Machine Learning consente di affrontare il problema in modo efficiente. Viene quindi fornito un quadro generale di tutte le tecniche utilizzate per l'analisi, con particolare riferimento agli algoritmi di Manifold Learning come la Self Organizing Map (SOM) e la Generative Topographic Mapping (GTM). Vengono inoltre descritti metodi di riferimento come il k-Nearest Neighbor (k-NN) o metodi più recenti come i predittori conformali, utilizzati per scopi di validazione e valutazione dell'affidabilità.

Nel **capitolo 5** viene presentato lo stato dell'arte relativamente alle tecniche di Machine Learning applicate alla predizione e alla classificazione di disruzioni, descrivendo in particolare le principali applicazioni con le ampia-

mente utilizzate Reti Neurali, quali Multi Layer Perceptrons (MLP), Support Vector Machines (SVM) e Self Organizing Maps (SOM), e i metodi statistici come la Discriminant Analysis o la tecnica Multiple Threshold. Vantaggi e svantaggi vengono discussi anche con riferimento ad una possibile soluzione per superare gli svantaggi di questi metodi: l'approccio multi-machine.

Il **capitolo 6** è dedicato alla descrizione dei database utilizzati per tutte le analisi che verranno presentate nei capitoli seguenti. In particolare vengono discussi in dettaglio l'analisi statistica e gli algoritmi di data-reduction che si sono resi necessari per costruire un database affidabile e statisticamente rappresentativo.

Gli ultimi tre capitoli contengono le analisi e gli algoritmi implementati per il mapping degli spazi operativi, la classificazione e la predizione delle disruzioni. Nel **capitolo 7** viene descritto il mapping dello spazio operativo di JET. Le prime sezioni si occupano di proiezione e visualizzazione dei dati con metodi di proiezione lineari come Grand Tour (GT) e Principal Component Analysis (PCA). Nella parte centrale sono stati trattati gli stessi aspetti sfruttando tecniche non lineari di Manifold Learning, SOM e GTM, sulla base delle quali è stata effettuata una dettagliata analisi dello spazio operativo. Tale analisi, mostrando la potenzialità dei metodi, è stata eseguita, per quanto riguarda il modello GTM, mediante la realizzazione di un tool dedicato. Infine, le performance nel mapping sono state valutate attraverso l'analisi degli outlier e di indici di performance appositamente proposti.

Nel **capitolo 8** viene descritta la classificazione automatica implementata per le disruzioni al JET. Il capitolo è diviso in due parti: la prima descrive la classificazione delle disruzioni appartenenti alle campagne con la parete in carbonio, mentre nella seconda parte è descritta la classificazione con la parete metallica (ILW) contestualmente alla valutazione della idoneità del classificatore automatico per applicazioni in tempo reale, unitamente ai sistemi di predizione on-line al JET. L'affidabilità dei risultati è stata validata attraverso il confronto con un classificatore di riferimento basato sulla tecnica k-NN, e attraverso i più recenti predittori conformali, con cui è possibile fornire in aggiunta alla predizione/classificazione il relativo livello di confidenza.

Il **capitolo 9** invece è dedicato alla predizione delle disruzioni ad ASDEX Upgrade. La prima parte è relativa alla descrizione del database e della tecnica di data-reduction utilizzata per selezionare un insieme di dati rappresentativo ed bilanciato. SOM e GTM sono stati utilizzate per mappare lo spazio operativo di ASDEX Upgrade e per costruire un predittore di disruzioni, introducendo al stesso tempo le loro potenzialità in termini di classificazione. Inoltre è stato proposto l'uso combinato di questi due metodi con un regressore logistico al fine di realizzare un sistema predittivo in grado

di sfruttare i comportamenti complementari dei due approcci, migliorando le prestazioni complessive in predizione.

Contents

Chapter 1

Fusion and magnetic confinement

1.1 Fusion energy

Fusion is a form of nuclear energy that powers the Sun and the stars and has the potential to provide an almost unlimited source of energy for the Earth.

The physicist Andreevich Artsimovich in the 1970 wrote that "thermonuclear fusion [energy] will be ready when mankind needs it". Considering the actual world energy situation and the effect on the environment due to the present harnessing of the different sources of energy, the hope is that time for fusion is finally arrived.

Fusion represents a source of energy really attractive first of all because the fuels which have to be used in a typical fusion power plant, water and lithium, are clean and environmentally sustainable not producing atmospheric pollution as the greenhouse gases. Another important point on the side of fusion is that such fuels are particularly abundant in the Earth, such that their supply will not represent a problem in the future. Unlike fission, low atomic number elements can react in such a way to convert mass to energy through fusion processes, as it happens for example in the Sun, where massive gravitational forces gives rise to the adequate conditions for fusion.

In the picture 1.1 we can see a graph representing the nuclear binding energy per nucleon plotted against the total number of protons and neutrons in the nucleus, i.e. the atomic mass. Nuclear binding energy is the energy required to separate a nucleus of an atom into its individual protons and neutrons. The mass defect is related to the energy released when the nucleus is formed according to the well known Einstein law $E = \Delta m \cdot c^2$. The most important feature of figure 1.1 is the maximum around mass number 56 corresponding to Fe element. This means that energy can be released if two

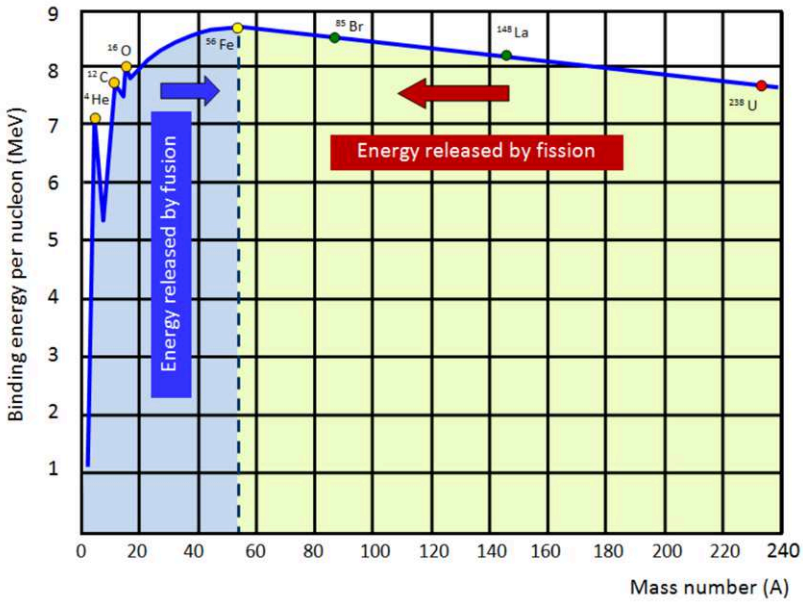


Figure 1.1: Nucleon binding energy. [from: *www.schoolphysics.co.uk*]

lighter atoms join to form a heavier one (moving from the left side toward the maximum of the curve) giving rise to a fusion reaction. On the other hand, according to the graph, energy can also be released if, moving from the right side toward the maximum of the curve, very heavy atom splits to form lighter fragments in a fission process.

On Earth conditions for fusion unfortunately are much harder to achieve. Low atomic number elements, as hydrogen and its isotopes, have to be heated to very high temperatures for reaching the right conditions for fusion. When these conditions are met gas mixture evolves into another state of the matter named plasma, where the negatively charged electrons are separated from the positively charged atomic nuclei (ions). One of the reasons that makes fusion not possible normally is that the strongly repulsive electrostatic forces which arises between the positively charged nuclei prevent them from getting close enough for fusion to occur. But when the temperature increases to a certain extent, the positively charged nuclei gain energy up to the point where attractive nuclear forces exceed electrostatic repulsive forces allowing fusion between the nuclei and the resulting release of energy.

If we consider on Earth a fuel of isotopes of hydrogen, we have not only to heat such a mixture of gas at temperatures of the order of 100 million degrees Celsius, but we have also to confine and keep it sufficiently dense in order to make fusion between nuclei possible. The fusion of hydrogen is the main reaction that powers the sun too, but in this case the strong gravita-

tional forces provides a fundamental contribution to achievement of fusion conditions.

Beyond the very important advantages of fusion energy connected with environmental impact, one has to consider also the aspect of fuel reserves. As it will be described in the following, the most convenient reaction through which fusion can be achieved is the one between an equal mix of Deuterium and Tritium. Deuterium can be found naturally in ocean water, can be extracted at relatively low cost and in particular, assuming the present rate of total world energy consumption, its supply can be guaranteed for something like 2 billion years!
The actual limit in terms of fuel reserve is represented by the Tritium: it is a radioactive isotope with a half-life of roughly 12 years, thus, in practical terms, it is not available naturally and has to be produced in situ in the power plant. But it is possible to obtain Tritium by breeding with the isotope Li^6 of lithium which the blanket of the future fusion devices will consist of. Always assuming the present rate of total world energy consumption, estimates indicate that Li^6 will be available on Earth for something like 20 millennia, before which, very likely, efficient technological solutions to employ D-D reactions will be ready, even if they produce less energy than D-T reactions.

Fusion energy has the potential to provide large amounts of base load electricity, changing deeply and in large scale the way in which the world consumes energy. The scientific feasibility of thermonuclear fusion via magnetic confinement has already been demonstrated, and presently also inertial confinement experiments are very promising. But in order to make fusion feasible also from the technological and engineering point of view, several critical issues have to be addressed, many of which will be dealt with in the framework of next generation of fusion reactors such as ITER and DEMO, which represent one of the most challenging scientific experiments of the upcoming future. Developing proper technologies and transposing all the scientific achievement to demonstrate not only the technical but also the economical feasibility of a fusion power plant which provides energy to electric grid, will require a considerable effort and further improvements of present technologies.

In the picture 1.2 a schematic representation of a future fusion power plant is reported. Deuterium and tritium fuel burns at a very high temperature in the central reaction chamber. The energy is released as charged particles, neutrons, and radiation and it is absorbed in a lithium blanket surrounding the reaction chamber. The neutrons convert the lithium into tritium fuel. A conventional steam-generating plant is used to convert the nuclear energy to electricity. The waste product from the nuclear reaction is helium.

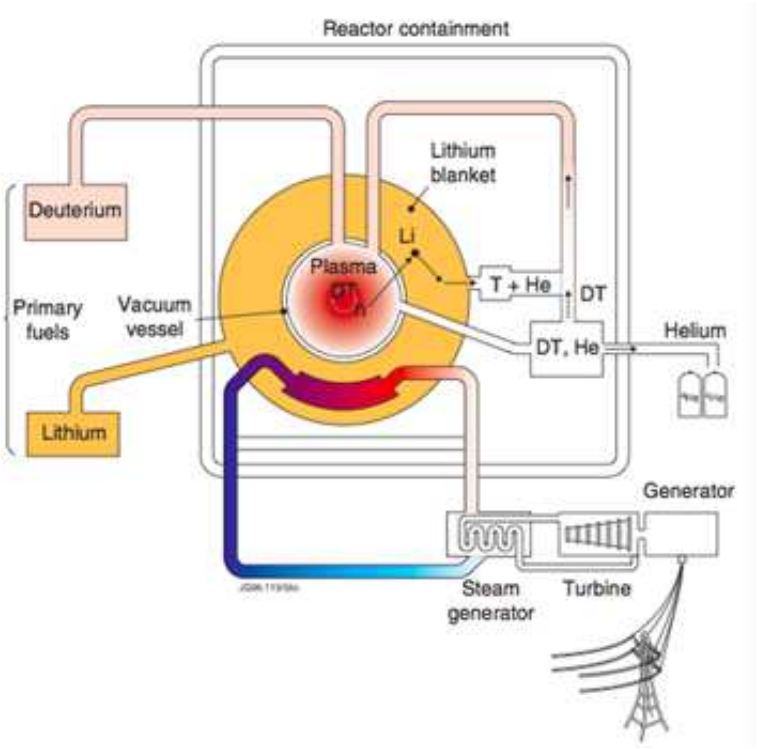


Figure 1.2: Schematic diagram of a proposed nuclear fusion power plant. [Fusion: The Energy of the Universe]

1.2 Basis of Fusion reaction

As it has been discussed in the previous section fusion can represent an almost unlimited source of energy for the future. The strong interest in fusion reactions has been motivated not only by considerations about fuel reserve and environmental impact, but also by the enormous potential in terms of produced energy with respect to other fuels and sources of energy. If we take into account the energy equivalence of different types of fuels, it is very easy to understand the reason of such a strong interest: the energy produced with 0.14 tons of Deuterium by fusion reactions is equivalent to the one produced by burning 10^6 tons of fossil oil or 0.8 tons of Uranium by nuclear fission. Among the relevant nuclear fusion reactions, as anticipated in the previous section, we have those ones among hydrogen isotopes such as D-D reactions (1.1, 1.2), which produce energy by the nuclear interaction between two deuterium nuclei according to the two equally likely reactions:

$$\mathrm{D} + \mathrm{D} \rightarrow \mathrm{He}^3 + \mathrm{n} + 3.27\ \mathrm{MeV} \tag{1.1}$$

$$\mathrm{D} + \mathrm{D} \rightarrow \mathrm{T} + \mathrm{p} + 4.03\ \mathrm{MeV} \tag{1.2}$$

D-D are the most desirable reactions, since theoretically their supply is economical and practically unlimited. Instead the D-T reaction (1.3) is based on the nuclear interaction between a deuterium nucleus and a tritium nucleus (figure 1.3). Among the possible fusion reactions it is the one with the highest likelihood of occurrence and it is usually written in the following way:

$$\mathrm{D} + \mathrm{T} \rightarrow \alpha + \mathrm{n} + 17.6\ \mathrm{MeV} \tag{1.3}$$

The 17.6 MeV of energy released through the D-T reaction is in the form of kinetic energy in part associated with the neutron (14.1 MeV) and in part with the alpha particle (3.5 MeV). Alpha particle should be confined within the plasma and transfer its energy by collisions to plasma ions and electrons. In this case the reaction releases 3.52 MeV per nucleon, whereas for the D-D (1.2) we have roughly 1.01 MeV released per nucleon. One of the problems in this case is associated to high energetic neutrons which pose serious problems of material activation and radiation damages, but the main drawbacks are related to tritium, Tritium is radioactive, it undergoes beta decay with a half-life (approximately 12.5 years), and is not naturally present on Earth. Nevertheless the high likelihood of occurrence with respect to the others, makes this reaction the main option of worldwide fusion research. In the figure 1.4 the probability that a fusion reaction will take place is represented in terms of cross sections for a wide range of energies. In particular at lower energies the probability for the D-T reaction is much higher than for the

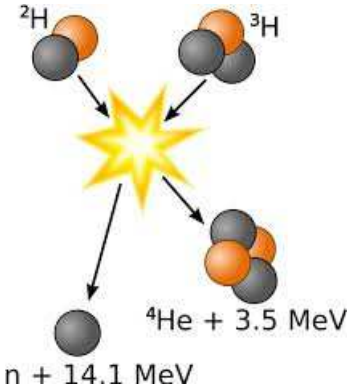


Figure 1.3: Deuterium-Tritium reaction. [from: *www.schoolphysics.co.uk*]

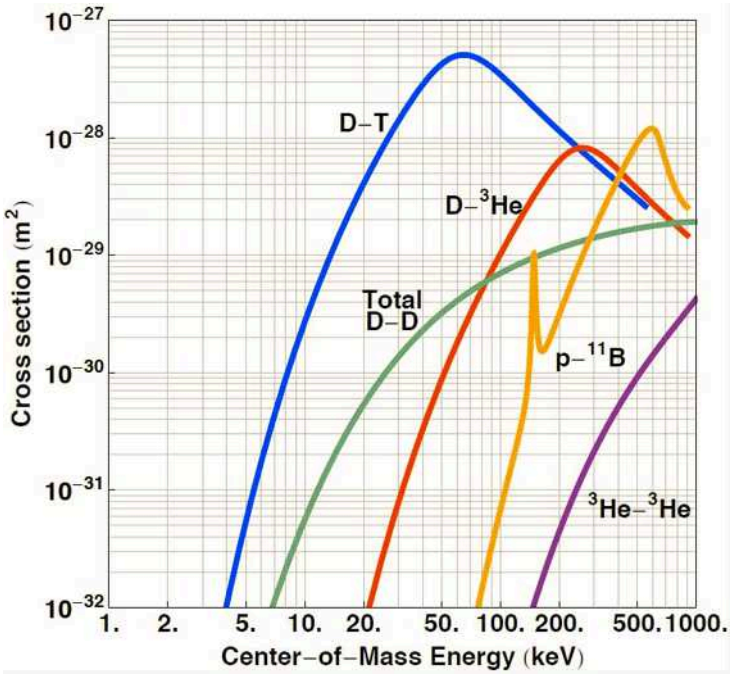


Figure 1.4: Cross sections versus center-of-mass energy for key fusion reactions. [from *http://iec.neep.wisc.edu/*]

other reactions.

As anticipated in the previous section tritium has to be supplied directly in situ, in the fusion power plant, by neutron capture in lithium, that is the most favorable chemical element for breeding tritium. In this context, the primary reactions through which tritium can be produced are the following (1.4, 1.5):

$$\mathrm{Li}^6 + \mathrm{n(slow)} \rightarrow \alpha + \mathrm{T} + 4.8\ \mathrm{MeV} \tag{1.4}$$

$$\mathrm{Li}^7 + \mathrm{n(fast)} \rightarrow \mathrm{T} + \alpha + \mathrm{n} - 2.5\ \mathrm{MeV} \tag{1.5}$$

Both reactions give rise to the production of tritium, even if the first one releases energy whereas the second one consumes it. On the other hand, the reaction with Li^7 is particularly important as well, because it doesn't consume a neutron allowing the possibility for self-sufficient tritium production in a fusion reactor, that is each neutron gives rise to the production of at least one new tritium nucleus. Naturally there is much more Li^7, but the reaction related to Li^6 has an higher likelihood of occurrence, therefore, it is the reaction which dominates in the breeding of tritium.

1.3 Magnetic confinement in fusion

Presently, two main experimental approaches are being studied: magnetic confinement and inertial confinement. The first approach in order to keep confined the hot plasma uses strong magnetic fields, whereas in the second approach small pellets containing fusion fuel are compressed to extremely high densities through strong lasers or particle beams.

Regarding magnetic confinement the widely investigated concepts are tokamaks (and spherical tokamaks), stellarators, reversed field pinches, spheromaks, field reversed configurations and levitated dipoles. All the machines are basically 2-D axisymmetric toroidal configurations, except the stellarator, that is an inherently 3-D configuration. Among all the configurations tokamaks have achieved the best overall performance, followed by stellarators. These configurations (figure 1.5) are all characterized by strong magnetic fields, reasonable transport losses and can operate in stable conditions with acceptable performance. Unlike tokamaks, stellarators do not require toroidal current drive in a reactor but the complexity and the costs related to the achievement of the 3-D magnetic configuration are a not negligible disadvantage.

In general, the presence of large toroidal magnetic fields implies reactors of certain size, and this means higher costs, whereas in the case of other configurations as the reverse field pinch, the toroidal magnetic field is much

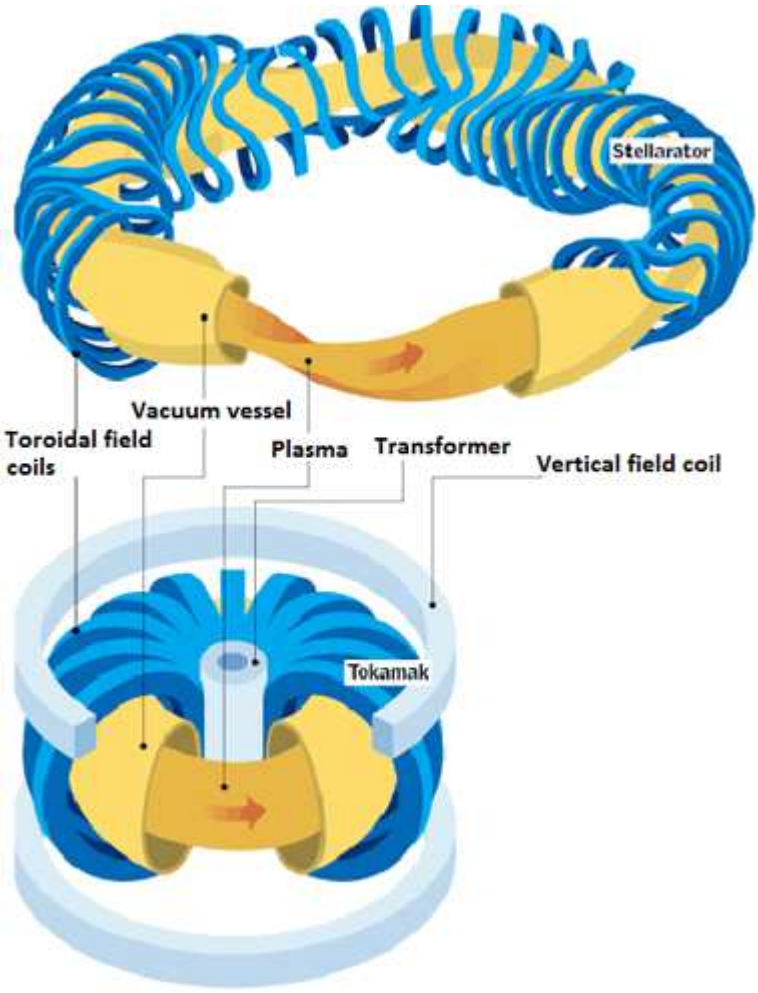


Figure 1.5: Tokamak and stellarator concepts.

smaller, as well as the costs, but unfortunately it corresponds to poor plasma performance with higher transport losses. Furthermore, tokamaks and stellarators configurations can be MHD stable even without the presence of a conducting wall near the plasma, whereas devices as reverse field pinches would require ideally a perfectly conducting wall with control feedback system for steady state operations. There are advantages and drawbacks for all the configurations, but so far tokamaks remain the most attractive configuration for a reactor, in fact ITER, which should demonstrate technological and engineering feasibility of a burning plasma experiment, will be a tokamak.

Bibliography

[1] World Energy Outlook-Special Report 2013 IEA, *Redrawing the Energy Climate Map*

[2] W.M. Stacey, 2010 WILEY-VCH Verlag GmbH & Co. KGaA, *Fusion: An Introduction to the Physics and Technology of Magnetic Confinement Fusion*

[3] J.P. Freidberg, 2007 Cambridge University Press, *Plasma Physics and Fusion Energy*

[4] G. McCracken and P. Stott, 2005 Elsevier Academic Press, *Fusion: The Energy of the Universe*

Chapter 2

Plasma stability in tokamak configurations

2.1 Introduction

The equilibrium of plasmas embedded in a magnetic field can be described by the magneto-hydro-dynamic (MHD) theory. MHD instabilities have the effect to strongly restrict fusion performance in magnetic confined plasmas, mainly because of the operational limits they impose. Such instabilities do not only limit the device operational domain reducing the achievable performance, but when they occur on a large scale often they can degrade the confinement leading to a sudden loss of plasma current and energy, that is a *disruption*. Plasma energy is lost within a time span of few milliseconds exposing the plasma facing components to severe thermo-mechanical stresses and conductors surrounding the vessel to huge electromagnetic forces. The deposition of the plasma energy on the vessel walls can cause deformations, structural damages, and eventually melting or evaporating of the in-vessel components. All these aspects not only drive structural and mechanical design of the machine, but also make necessary to avoid or mitigate disruptions in order to preserve the integrity of the machine.

2.2 Magnetic confinement with Tokamak configuration

A thermonuclear fusion plasma, due to its high temperature, is not allowed to come directly in contact with the wall, because the materials eroded by the plasma itself would quickly cool this latter. One solution to overcome this

problem is to confine and thermally insulate the fuels in a chamber by magnetic fields. A charged particle q moving in a magnetic field will undergo a Lorentz force $\mathbf{F} = q(\mathbf{E}+\mathbf{v}\times\mathbf{B})$ perpendicular to both the direction of particle motion (with velocity $\mathbf{v}$) and magnetic field $\mathbf{B}$, where $\mathbf{E}$ is the electric field. This force is responsible of producing a circular particle motion in the plane perpendicular to the magnetic field line. In other words a charge particle in a magnetic field will move along the field line with a spiral trajectory (figure 2.1), whose radius is called gyro-radius (or Larmor radius) and is inversely proportional to the strength of the magnetic field. As we have discussed in

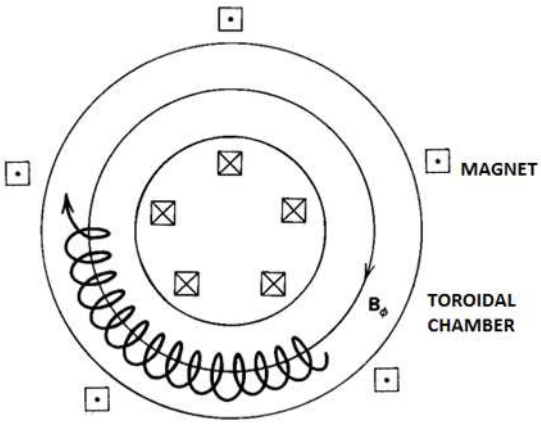


Figure 2.1: Charged particle motion along a magnetic field line in a toroidal configuration

the previous sections, tokamaks represent one of the possible approaches to magnetic confinement of plasmas. As it is well known, tokamak configuration allows to overcome the inherent end losses that we have in cylindrical geometry. In a pure toroidal system with only a toroidal field, intrinsic factors as magnetic field curvature and gradient gives rise to a vertical drift in opposite direction for ions and electrons (with velocities $\mathbf{v_{d,i}}$ and $\mathbf{v_{d,e}}$ respectively), as it is shown in the sketch in Figure 2.2. The electric field resulting by the charge separation, determines an outward $\mathbf{E}\times\mathbf{B}$ drift of plasma particles (with velocity $\mathbf{v_{E\times B}}$). In other words, a toroidal configuration with purely a toroidal magnetic field is intrinsically unstable: to avoid radially outward drift motions and thus that particles hit the wall, it is necessary to twist magnetic field lines through some additional component. A poloidal magnetic field must be superimposed upon the toroidal magnetic field in order to compensate these drifts. The result is to have helical magnetic field lines entirely contained within the toroidal chamber. Such a poloidal field, in the case of the tokamak is produced by a toroidal current flowing in the plasma, whereas in a stellarator is produced by external coils.

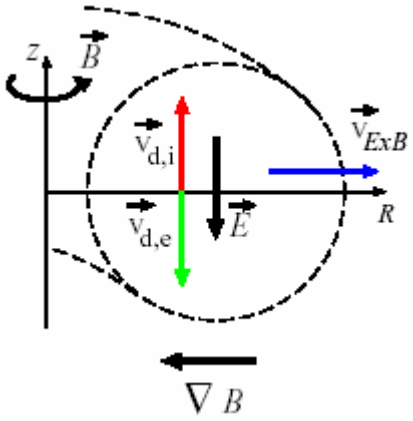


Figure 2.2: Particles drift in a toroidal configuration

2.2.1 Tokamak coordinate system

Given suitable operative conditions, it has been proved that tokamaks are stable. Nevertheless in order to increase plasma performance and confinement it is necessary to push relevant plasma parameters close to their limits, as for example pressure, current and density. This can determine the onset of different instabilities that affect significantly the confinement leading in certain circumstances to the abrupt termination of the discharge. Considering a torus, it is usual to work in a cylindrical coordinate system (R, ϕ, z), where R is the radial coordinate, ϕ is the toroidal angle and z is vertical axis of the torus (figure 2.3a). When all quantities results to be independent with respect to the toroidal angle ϕ we are in a condition of axisymmetric. In figure 2.3b we can see the quantities of interest in the poloidal cross section, the coordinate along the minor radius r and the poloidal angle θ.

2.2.2 Shafranov shift and equilibrium in a toroidal magnetic configuration

Unfortunately, when we connect the ends of a cylinder obtaining a torus, the condition of MHD equilibrium is no longer satisfied. In fact plasma has the tendency to expand outward in the direction of the major radius basically for two reasons. The first one is related to the fact that the pressure forces inside the cylinder are in first approximation equally distributed on the boundary of the poloidal cross section, but in a toroidal geometry, the outer surface has a larger area than the inner one, so that the net force is outwards. The second one is due to circuit theory considerations, in fact since we are consid-

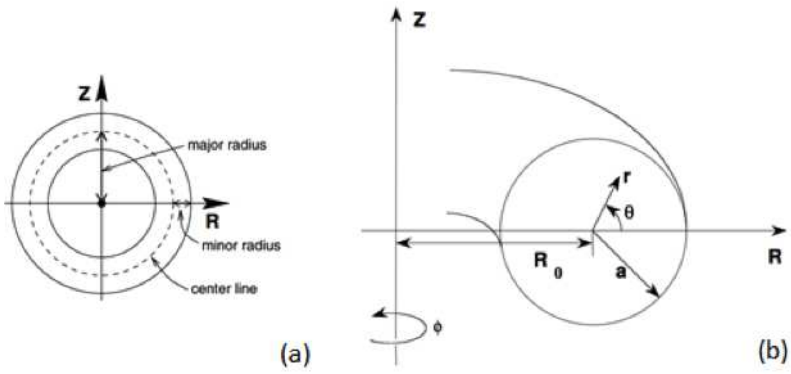


Figure 2.3: (a) axisymmetric coordinate system in a toroidal geometry; (b) poloidal cross section coordinates.

ering a ring where a current is flowing through along the toroidal direction, we have that current elements shifted by an angle $\phi = \pi$ repel each other because the current is in opposite direction. Also in this case the net force is outwards, or in other words, it tends to expand the plasma ring along the radial direction. Therefore, a toroidal plasma column is not in equilibrium because of the magnetic effect given by the current inside and because of the kinetic effect associated to the pressure of the plasma.

As we have seen in the introductory chapter dedicated to different devices in relation to the magnetic confinement, magnetic surfaces in tokamak toroidal geometry are essentially circular tubes around the main axis of the machine (z axis), and the current field lines lie on these magnetic surfaces that are isobaric surfaces too.

The magnetic field in a geometry as the toroidal one has three components: the radial one along the R axis (major radius), the vertical one along the z axis and the toroidal one along the coordinate corresponding to the toroidal direction, i.e. along the angle ϕ (along which all the physical parameters should be equal in every point since we are assuming an axisymmetric configuration). The basic condition for plasma equilibrium requires that forces in every point are zero [2], as reported by the following relation (eq 2.1):

$$\mathbf{j} \times \mathbf{B} - \boldsymbol{\nabla} p = 0 \tag{2.1}$$

where $\mathbf{j}$ is the current density, $\mathbf{B}$ the magnetic field and $\boldsymbol{\nabla} p$ is the pressure gradient.

Flux surfaces in a tokamak configuration look like nested toroidal flux tubes and are the solution of the Grad-Shafranov equation, which is a differential equation in terms of a poloidal flux function ψ. Grad Shafranov equation

can be numerically solved under simple geometrical assumptions (circular plasma and large aspect ratio, that is the ratio between major and minor radius), as most of the codes for the equilibrium reconstruction do, as EFIT for example. Being the plasma enclosed in an electrically conductive shell, the most important effect associated with the fact that plasma tends to expand outwards, is that field lines are compressed in the outboard side. This compression gives rise to an increase of the magnetic pressure that has the effect to counteract the tendency of the plasma to expand. The resulting equilibrium state is then characterized by a shift of the fluid outwards with respect to the geometric center of the circumference related to the poloidal cross section, which does not correspond anymore to the axis of the magnetic configuration in the new equilibrium state. This deviation, defined as Shafranov shift (Δ), is shown in Figure 2.4.

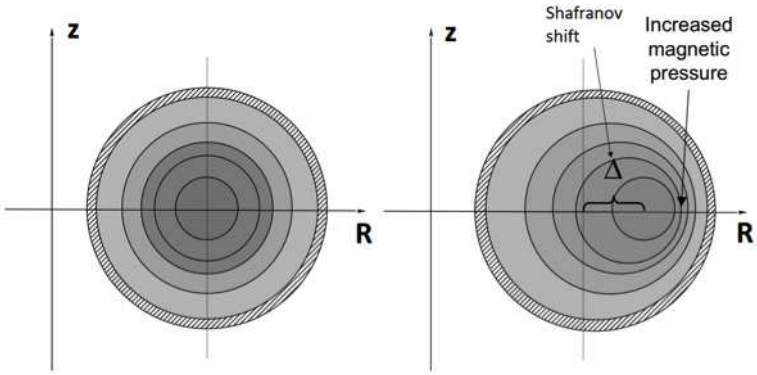


Figure 2.4: Shafranov shift.

2.2.3 Stabilization with external vertical field and beta parameter

As introduced in the previous paragraph, in the outboard side the poloidal field lines are closer each other than the inboard side: this means that the poloidal field is stronger in the outer region, and, being the magnetic pressure proportional to the square of the magnetic induction, the resulting force is inwards and opposes the expansion of the plasma. To counteract the forces which tends to expand the plasma, the practical solution is represented by the addition of a vertical magnetic field along the zeta axis, whose interaction with the toroidal plasma current gives rise to a $\mathbf{j} \times \mathbf{B}$ force in the opposite direction, i.e. inwards. The fact that a plasma ring is not in equilibrium alone but we need a magnetic field produced by external sources to keep the

equilibrium is not a particular property of the toroidal configuration, but it is a general property of the plasma independently on its configuration. This general principle is expressed by the Virial's theorem, which says that it is impossible to sustain any MHD equilibrium without currents external to the plasma. At the equilibrium, under certain hypothesis (single fluid under stationary conditions, circular cross section and large aspect ratio), the poloidal field on plasma surface at minor radius a and angle θ is given by the following formula (2.2):

$$B_\theta(a,\theta) = \frac{\mu_0 I}{2\pi a} \cdot (1 + \frac{a}{R_0}\Lambda \cdot cos\theta) \quad \text{where} \quad \Lambda = \beta_\theta + \frac{l_i}{2} - 1 \tag{2.2}$$

In the expression of Λ, they appear two quantities of fundamental importance in relation to plasma stability and equilibrium: the poloidal beta β_θ and the internal inductance l_i. The parameter β is defined as the ratio between kinetic plasma pressure, averaged over the plasma volume, and the corresponding magnetic pressure:

$$\beta = \frac{\langle p \rangle}{\frac{B^2}{2\mu_0}} \tag{2.3}$$

The poloidal beta simply refers to the poloidal magnetic field B_θ. This parameter represents a measure of the quality and economic efficiency of the confinement, and plays a key role in stability. If we consider a plasma ring with a current flowing inside, the current density inside is fixed, but usually is not uniform in the cross section. The temperature in the core region of the plasma is higher than the one in the edge region, and it's known that the plasma resistivity, differently by the conductor material like copper, decreases as the temperature increases. The current tends to flow where the resistivity is lower, so it tends to concentrate in the center of the plasma column. Regarding the inductance we can say that in general is defined as the ratio of the linked flux divided by the corresponding current. Anyway it turns out quite difficult to define the internal inductance of a plasma column because normally it's defined for current filaments. In these cases, when the current is not filamentary, we can define the internal inductance l_i in terms of magnetic energy, as the following ratio (2.4):

$$l_i = \frac{\langle B_\theta^2 \rangle}{B_\theta^2(a)} \tag{2.4}$$

where a is the minor radius. In other words the internal inductance is a normalized parameter that gives an indication about radial profile and peaking

of plasma current, since the poloidal field depends on the plasma current in the toroidal direction (I_p). Returning to the considerations about the need of external sources according to the Virial theorem, the vertical field necessary to maintain the plasma in equilibrium is expressed by the following equation (2.5):

$$B_z = \frac{\mu_0 I_p}{4\pi a} \cdot (ln\frac{8R_0}{a} + \Lambda - \frac{1}{2}) \tag{2.5}$$

Its effect is to provide an inward force able to counteract the outward hoop force that acts on the plasma because of the aforementioned reasons.

2.2.4 The safety factor q

Another very important parameter for the analysis of the equilibrium and the confinement properties of the plasma is the safety factor q, which is defined as:

$$q = \frac{\Delta\Phi}{2\pi} \tag{2.6}$$

Such a parameter is indicative of the helicity of the field lines, determining how many toroidal rotations (indicated by the variation of the toroidal angle $\Delta\Phi$) are necessary for a single rotation of a magnetic field line in the poloidal direction (2π). If $q = m/n$ and m and n are the integer values corresponding respectively to the toroidal and the poloidal turns after which a field line rejoins up on itself, we say that the field line lies on a rational surface, otherwise we speak about ergodic surfaces. As we will deal with in the next section, rational surfaces of q and its radial profile play a key role in the stability of the plasma. The figure 2.5 shows typical profiles of the main quantities in a large aspect-ratio tokamak. In general, making reference to the equation of the field lines we have that for tokamaks with large aspect ratio ($\frac{R}{a} \gg 1$) safety factor can be approximated as follow (2.7)

$$q(r) = \frac{r}{R}\frac{B_\phi}{B_\theta} \tag{2.7}$$

Taking into account the elongation k of the plasma shape, on the base of which $r = a\sqrt{k}$, and the Ampere Law, according to which the poloidal magnetic field B_θ is defined as

$$B_\theta = \frac{\mu_0 I_p}{2\pi r} \tag{2.8}$$

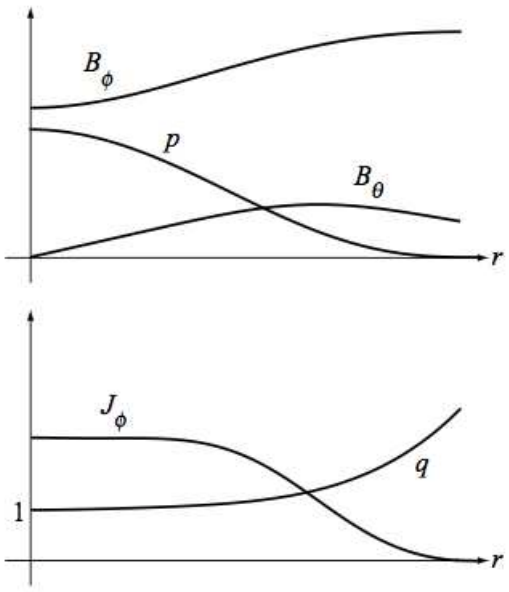


Figure 2.5: Typical profiles in a tokamak in the large-aspect-ratio limit $R/a \to \infty$, where B_Φ is the toroidal component of the magnetic field, B_θ is the poloidal component, p is the pressure, J_Φ is the toroidal current density and q is the safety factor [3].

the safety factor at the edge, can be written as (2.9):

$$q_{edge} = \frac{B_\phi \cdot a^2 k}{2RI_p \cdot 10^{-7}} \tag{2.9}$$

Rational values and radial profile of the safety factor are essential in MHD stability considerations, as well as the so called magnetic shear, defined as follow (2.10):

$$s(r) = \frac{r}{q(r)} \frac{dq(r)}{dr} \tag{2.10}$$

The magnetic shear is strictly related to the resonance concept and has important implications in MHD stability: it describes basically the variation of the magnetic field winding angle moving radially through subsequent magnetic surfaces. In this context therefore, a strong magnetic shear is generally good for stability; conversely it results really dangerous conversely when close surfaces has the same safety factor, because these surfaces can couple with each other giving rise to resonance phenomena and instabilities. Therefore,the q-radial profile plays a key role in governing several MHD instabili-

ties. Furthermore it has also several implications in the characterization of confinement modes: particularly important examples to this purpose are the reverse, the optimized and the negative central shear which are associated to enhanced confinement regimes.

2.3 MHD stability

2.3.1 Basic classifications of MHD instabilities

The macroscopic equilibrium of a fusion plasma can be described by MHD theory. As it has been described in the previous sections, the equilibrium in a toroidal configuration is characterized by a set of nested flux surfaces on which magnetic and current field lines lie. MHD considers the plasma as a single, globally quasi-neutral fluid, composed of charged particles which can conduct electrical currents and react to magnetic fields.

MHD equations can be seen as the union of fluid dynamics equations and Maxwell's equations of electromagnetism, and can be properly elaborated in order to describe in stationary conditions MHD equilibria. The equilibria configurations are linked to a specific device and are defined for a certain set of boundary conditions. In particular, for the toroidal pinch devices, the configurations characterizing the equilibria can be found by solving the Grad-Shafranov equation, which is expressed in terms of the poloidal flux function ψ (2.11):

$$R\frac{\partial}{\partial R}\frac{1}{R}\frac{\partial \psi}{\partial R} + \frac{\partial^2 \psi}{\partial z^2} = -\mu_0 \, R^2 \frac{dp(\psi)}{d\psi} - {\mu_0}^2 f(\psi)\frac{df(\psi)}{d\psi} \qquad (2.11)$$

Moreover,we have seen moreover that a vertical field produced by an external source is needed to balance the intrinsic tendency of a plasma in a toroidal configuration to expand outwards along the major radius R. In a confined plasma, an instability is driven by the free energy contained in the equilibrium configuration. In a tokamak, there are two main sources of free energy: the kinetic energy of the plasma and the energy of the magnetic field generated by the plasma. Instabilities can therefore, be driven by the radial gradient of either the pressure or the current profile. At low β, the magnetic energy is much higher than the kinetic energy and the instabilities will mainly be current-driven; at high β, we expect the pressure driven instabilities to become significant.

Current driven and pressure driven instabilities

MHD instabilities influence the achievable β of a configuration, therefore, they have to be avoided or kept under control ideally. An example is the external kink driven by current gradients near the edge imposing restrictions on the possible current profiles. Restrictions on the pressure profile can come from the so-called interchange instabilities or from the ballooning instabilities. Pressure driven instabilities basically depend on the entity of the pressure gradient and on the field line curvature.

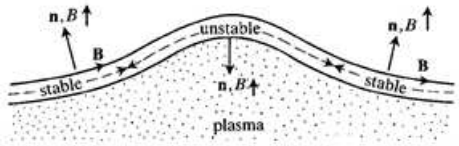


Figure 2.6: Bad and good curvature for pressure driven instabilities

In figure 2.6 we can see that in relation to the interchange of free energies between the field lines we may have bad curvatures to which is associated an unstable situation as the central one, or good curvatures (as the side ones), where the interchange of magnetic field and plasma works very well to release free energies providing in this way a stabilizing effect. In other words when the radius of curvature is parallel to the pressure gradient (so-called *bad curvature*)we have a destabilizing effect, while if the radius of curvature is anti-parallel to the pressure gradient, an interchange of plasma and magnetic field will increase the magnetic energy and thus be stabilizing (*good curvature*).

Pressure driven instabilities are often associated to internal modes, that is they occur within the plasma without affecting macroscopically the entire surface region of the plasma column. Ballooning modes are generally the most unstable pressure driven instabilities and in a tokamak usually are characterized by a larger amplitude on the low field side of a flux surface, whereas kink modes, on the contrary, have more or less the same amplitude along the flux surface. Their stability depends especially by the curvature of the magnetic field lines. In general pressure driven instabilities are particularly important because they set a limit to the maximum achievable β in a fusion plasma.

Current driven instabilities are connected to the parallel current, and are typically the so called kink modes, because of the shape associated to the deformation of the plasma column. In the figure 2.7 for example we can see a kink instability in presence of a conductive wall, which through the eddy

currents that flow in its surface gives rise to a restoring force which has the effect to stabilize the kink. In particular if there were no vacuum region between the plasma and the conducting wall we could't have any displacement of the plasma surface and only internal kink modes would be possible.

Obviously, to have a perfectly conducting wall surrounding plasma surface is not a viable option with fusion plasmas. We have to consider a wall with finite resistivity which has the effect to slow down the growth rate of instabilities as the external kink modes for example. Therefore, a perfectly *conducting wall* could greatly improve stability beyond the limit of the ideal no-wall case, whereas a resistive wall in practice does not change the limit with respect to the no-wall case but it changes the time scale slowing down the growth rate. Current driven instabilities in general can be associated to

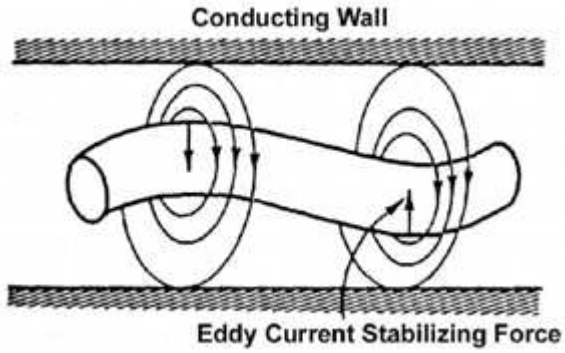


Figure 2.7: kink stability in presence of a conducting wall

internal or external modes, and, as it will deal with in the chapter dedicated to the operational limits, especially external kink modes are very important because they limit the maximum toroidal plasma current in stable conditions.

Ideal and resistive MHD

The previous distinction was based on the source of the instabilities. Another basic distinction is made on the basis of the time scale of the characteristics phenomena and is between *ideal MHD* and *resistive MHD* instabilities. In the ideal case, we consider the plasma perfectly conductive and, therefore, we refer to the Alfvén time scale, where the evolution of the instability is limited only by the inertia of the plasma, which is very small because the mass of the plasma itself is very small (order of microseconds or tens of microseconds). If instead we consider resistive MHD instabilities, the time scale is of order of milliseconds because even if the plasma is not perfectly conductive the resistivity is low.

This distinction is particularly relevant since even if the plasma in certain conditions should be stable in ideal MHD approximation, it could be unstable because of resistive effects. Furthermore, the flux conservation law, which is valid in ideal MHD, dictates that magnetic field lines move with the plasma flow, and therefore magnetic topology is conserved ("frozen"), which in other words means that magnetic field lines cannot tear or reconnect, as instead happens in resistive MHD.

2.4 General concepts of linear MHD stability

2.4.1 Mode numbers

Another basic classification is related to the mode numbers and the resonance position. In the simple case of a circular tokamak with large aspect-ratio, the modes, or in other words the helicity of the perturbations, are in the form $e^{(m\theta - n\phi)}$, where m and n are respectively the poloidal and the toroidal mode numbers. A mode m, n is resonant in the plasma if inside it or close to its surface there are magnetic surfaces satisfying the condition $m/n = q$, where q is the safety factor describing the helicity of such a surface. Avoiding resonant modes in the plasma is fundamental for stability in tokamaks.

In the figure 2.8(A) a sketch of the set of Mirnov coils installed on the STOR-M tokamak for the investigation of MHD instabilities is reported. In particular two sets of poloidal arrays of 12 Mirnov coils regularly spaced with a step of 30° at two opposite toroidal sections, allow the measure of poloidal mode numbers up to $m = 6$. Toroidal mode numbers can instead be analyzed by four sets of toroidal arrays, each one composed by 4 discrete Mirnov coils toroidally separated each one from the others by 90°. This distribution allows the determination of toroidal mode numbers up to $n = 2$. In the figure 2.8(B) some schematic pictures of toroidal ($n = 1$) and poloidal ($m = 1, 2, 4$) modes numbers have been reported.

2.4.2 Main formulations of linear stability

Also in the MHD framework the most reliable definition of stability is the one of exponential stability, on the base of which a system is unstable if any of the modes eigenfrequencies correspond to exponential growth, that is when the related imaginary part is greater than zero [1] [5]. Beyond eigenfunctions and corresponding eigenfrequencies calculation, another theoretical procedure for stability analysis is the energy principle, based on the potential energy variation for a certain plasma displacement $\xi(x)$. In particular,

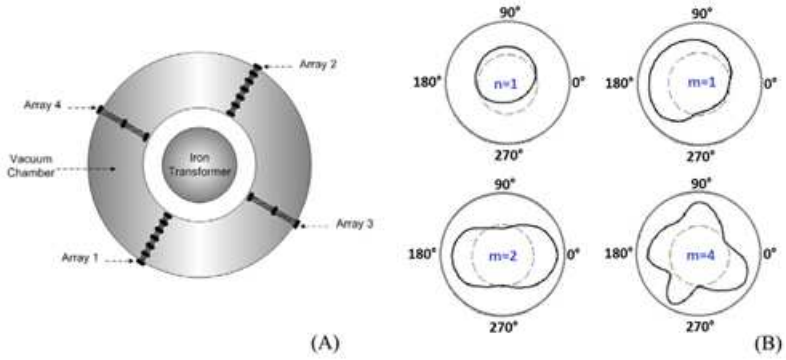


Figure 2.8: (A) Distribution of toroidal and poloidal Mirnov coils; (B) representation of toroidal ($n = 1$) and poloidal ($m = 1, 2, 4$) modes numbers from [4]

linear stability can be evaluated by linearization of the equations governing the system and by analyzing the response to sufficiently small perturbations around equilibrium conditions. The amplitude of such perturbations saturates above a certain level because of nonlinear processes. Possible sources of perturbations are typically the onset of an instability or the presence of a magnetic field error. Let's consider for example the simple case of a tokamak with helical field lines where on a poloidal section the magnetic topology is described by concentric circles and the magnetic field has the following form (2.12):

$$\mathbf{B} = B_\phi \hat{\boldsymbol{\phi}} + B_\theta \hat{\boldsymbol{\theta}} \tag{2.12}$$

If we add now to an equilibrium condition (subscript "0") a small radial perturbation of small amplitude b_r, such as $b_r/B_0 \ll 1$, the resulting magnetic field will be given by (2.13):

$$\mathbf{B} = \mathbf{B_0}(r) + b_r(r) sin(m\theta - n\phi)\hat{\mathbf{r}} \tag{2.13}$$

Being the wave vector of the perturbation (2.14)

$$\mathbf{k} = \frac{m}{r}\hat{\boldsymbol{\theta}} - \frac{n}{R}\hat{\boldsymbol{\phi}} \quad \text{where} \quad k_\parallel = 0 \tag{2.14}$$

if 2.15 is satisfied, which is analog to consider $q = m/n$ for the considered field line, then the mode is resonant in the plasma, and a small magnetic perturbation can give rise to a large field line excursion.

$$\mathbf{k} \cdot \mathbf{B} = \frac{m}{r} B_\theta - \frac{n}{R} B_\phi = 0 \tag{2.15}$$

Considering the resistive effect, we have a broader range of accessible states and magnetic field lines can tear giving rise to significant changes of topology such as reconnection phenomena and the formation of magnetic islands as shown for example in figure 2.9.

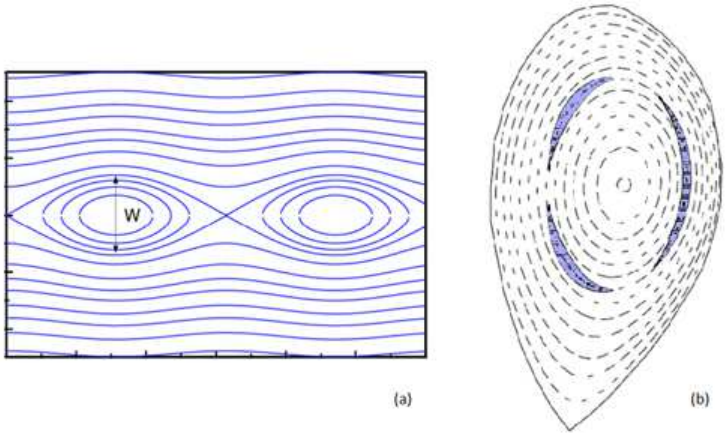


Figure 2.9: (a) Field lines tearing and reconnection; (b) $m = 3$ magnetic islands (from [6]).

The energy principle

Energy principle for ideal MHD is based on the fact that if a perturbation of a given equilibrium condition reduces the potential energy associated to the configuration, the considered equilibrium is unstable. Given a certain equilibrium condition, if F represents the force arising because of a displacement ξ, the potential energy will be defined by the following equation (2.16):

$$\delta W = -\frac{1}{2}\int_{Vol} \mathbf{F} \cdot \boldsymbol{\xi}\, d\tau \tag{2.16}$$

The linearized force is given by the relation 2.17

$$\mathbf{F} = \mathbf{j_1} \times \mathbf{B_0} + \mathbf{j_0} \times \mathbf{B_1} - \boldsymbol{\nabla} p_1 \tag{2.17}$$

where the equilibrium is indicated by the subscript 0, whereas the perturbation by the subscript 1. Regarding the perturbed terms, p_1 is obtained by integrating the linearized adiabatic equation, B_1 is obtained by integrating

Faraday's induction law, whereas the perturbed current density j_1 is calculated through Ampere's law. The final formulation for the variation of the potential energy is the following 2.18:

$$\delta W = \frac{1}{2}\int_{plasma}\left(\gamma p_0(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 + (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}p_0)\,\boldsymbol{\nabla}\cdot\boldsymbol{\xi} + \frac{\mathbf{B_1}^2}{\mu_0} - \mathbf{j_0}\cdot(\mathbf{B_1}\times\boldsymbol{\xi})\right)d\tau + \\ + \int_{vacuum}\left(\frac{\mathbf{B_v}^2}{2\mu_0}\right)d\tau \tag{2.18}$$

As it has been said above, plasma equilibrium is considered unstable if for any physically allowable displacement $\boldsymbol{\xi}$ the corresponding variation of potential energy is negative. In particular we can distinguish in the previous equation the pressure driven term $(\boldsymbol{\xi}\cdot\boldsymbol{\nabla}p_0)\,\boldsymbol{\nabla}\cdot\boldsymbol{\xi}$ and the current driven term $\mathbf{j_0}\cdot\mathbf{B_1}\times\boldsymbol{\xi}$: depending on the term which results to be prevalent between the two, the mode will be considered pressure driven or current driven. The integral in the second row of the equation (2.18) represents instead the transfer of energy to the vacuum region ($\mathbf{B_v}$ is the magnetic field in the vacuum).

Stability problem is usually addressed by considering the behavior and the time evolution in response to perturbation of small amplitudes. Consequently it is possible to linearize the considered systems of partial differential equations for which there exist several numerical techniques that make complex problems of stability analysis tractable. Obviously, the theory of linear stability cannot predict or extrapolate the behavior of the system interested by a non-linear evolution of stability, but experiments have proved that a plasma unstable according to linear MHD stability, often evolve unavoidably to a state of dramatic deterioration of confinement. In this sense therefore linear stability provides a strong base for such an analysis.

Bibliography

[1] J. Wesson J. Oxfor University Press 2004 *Tokamaks 4th edn*

[2] V. S. Mukhovatov, V. D. Shafranov, 2011 *Plasma Equilibrium in a Tokamak*, Nuclear Fusion 11

[3] J.P. Freidberg, 2007 Cambridge University Press, *Plasma Physics and Fusion Energy*

[4] C. Xiao *et al.* IOP-IAEA 2008 *Investigation of MHD instabilities on STOR-M tokamak*

[5] J.P. Friedberg 1987 Plenum Press, New York, "*Ideal MHD*

[6] `http://home.physics.ucla.edu/calendar/conferences/cmpd/talks/forest.pdf`

Chapter 3

Operational limits and disruptions on Tokamaks

The goal of fusion research is to achieve the conditions for a magnetically confined burning plasma. In particular when adequate conditions are provided, in a D-T plasma, α particles heating alone is sufficient to sustain plasma temperature against energy losses, without the further need of additional heating. This operational point is called *ignition*, and basically identifies the condition at which a nuclear fusion reaction becomes self-sustaining, as it can be expressed through the figure of merit represented by the triple product of density, temperature and confinement time (3.1). For a D-T plasma the value has to be of a certain order, as it is expressed by the following condition:

$$nT\tau_E > 10^{21} m^{-3}\ keV\ s \tag{3.1}$$

The value can change depending mainly on the considered profile for density and temperature. Such a condition is derived on the base of the well known *Lawson* criterion. The aim is to give rise to the conditions needed to self-sustain a plasma with a temperature of $10keV$, a confinement time of several seconds and a fuel density of the order of $10^{20}\ particles/m^3$. The fusion energy gain factor Q, is defined as the ratio of fusion power produced by nuclear fusion reactions to the power needed to sustain the plasma. The condition of Q = 1 is referred to as break-even (figure 3.1). Commercial power plants would require a Q value between 20 and 30, whereas the technical objective for ITER will be a minimum gain of $Q = 10$ for at least 300 seconds and to demonstrate steady-state operation with a $Q = 5$ for several thousands of seconds [1]. The fusion power in a tokamak device has different constraints, some of them are technical and economical constraints as the machine size or the applied magnetic field, others are related to physical limits connected

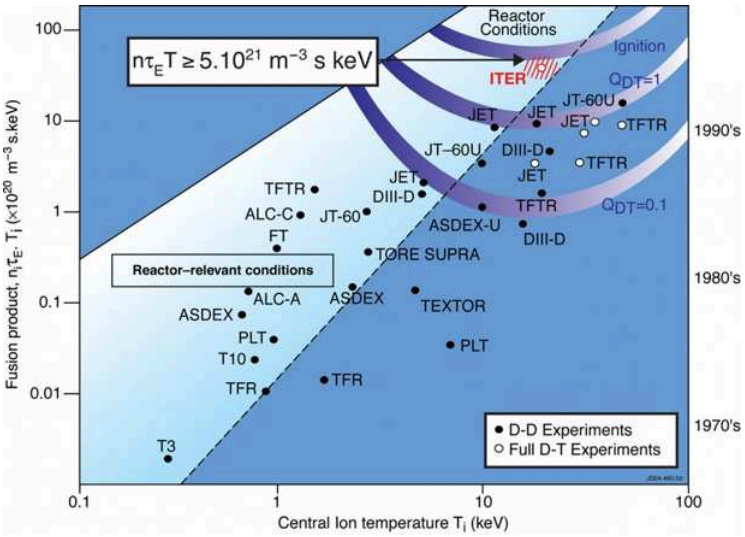


Figure 3.1: Figure of merit of fusion performance (Triple Product $nT\tau_E$) [www.efda.org].

with the stability concept, as for example β which should be maximized for the reasons already discussed in the previous chapter.

In order to achieve the target of nuclear fusion, the experiment have to maximize three basic quantities: fuel density n, energy confinement time τ_E, and the normalized pressure beta. The optimization of this parameters is often limited by the onset of MHD instabilities that can be driven basically by the gradients of the plasma current or the pressure profiles. MHD instabilities on macroscopic scales can lead to the degradation of the confinement (soft limit) or in the worst case to the abrupt termination of the discharge with a *disruption* (hard limit). Maximizing β requires to increase as more as possible plasma pressure and stored energy and to do this we have only the possibility to act on the available external control parameters. To increase the temperature for example we have to apply more auxiliary heating, if we want to set the density (feedback controlled) to a certain level, we have to act on the gas fuelling rate, whereas the control of the plasma current has to be done through the induced loop voltage and therefore through flux regulation.

Furthermore, depending on the regimes or foreseen operational scenarios, the optimization of plasma performance comes up against different constraints. There are scenarios where such a optimization requires a proper shaping and control of pressure and current density profiles, as well as we have for example in the case of optimized or reversed shear scenarios [2]. An extensive discussion about operational limits is beyond the scope of this

thesis, nevertheless they will briefly discussed and summarized all the main constraints with particular reference to their connection with disruptions.

3.1 Operational limits

Disruption-free operations in a tokamak are limited by well known operational limits not necessarily related directly to a violation of an MHD stability boundary: the current and the pressure limit, which are a direct consequence of development of an ideal MHD instability, and the density limit which is more directly a consequence of an excessive radiation from the plasma, accompanied by a progressive deterioration of plasma energy and confinement. This does not mean that during the chain of events that leads eventually to disruptions there is no an intermediate MHD instability which contributes to the final loss of confinement.

The violation of these operational boundaries in tokamaks leads to the onset of MHD instabilities, often characterized by helical perturbations, as those ones described in the previous chapter. These MHD instabilities grow non-linearly in the final phase until a major disruption occurs. In the present generation of medium-size tokamaks the loss of thermal energy has a typical time scale of $\sim 100\mu s$. The rapid cooling due to the *thermal quench* and the consequent increase of plasma resistivity gives rise to the fast decay of the current, known as *current quench*. Coming back to the description of the three basic operational boundaries, the current limit and the density limit can be described making reference to the well known *Hugill diagram*, where the inverse of the safety factor at the edge $1/q_a$ is plotted against the so-called Murakami parameter nR/B_ϕ, that is basically a normalized line averaged density (figure 3.2).
Since the temperature has an optimum value at $\sim 20keV$, n, that here indicates the line averaged density, should be as high as possible. But density is limited by disruptions due too excessive edge cooling: for a given plasma current there is a maximum achievable line averaged density.

The density limit, also known as *Greenwald limit* [3], is expressed by the condition $n_{GW}(10^{20}m^{-3}) \sim I_p(MA)/[\pi a^2(m^2)]$. This is an empirical boundary, and especially in the last years has increased due to application of additional heating and advanced wall conditioning methods that reduce the strong radiated power related to impurities. In fact, as it will be discussed in the section dedicated to the analysis of the causes and the chain of events of disruption, there is a strong connection with the radiation instabilities such as the radiative collapse and MARFE limit.

Regarding the current limit instead, as we can see in the Hugill diagram

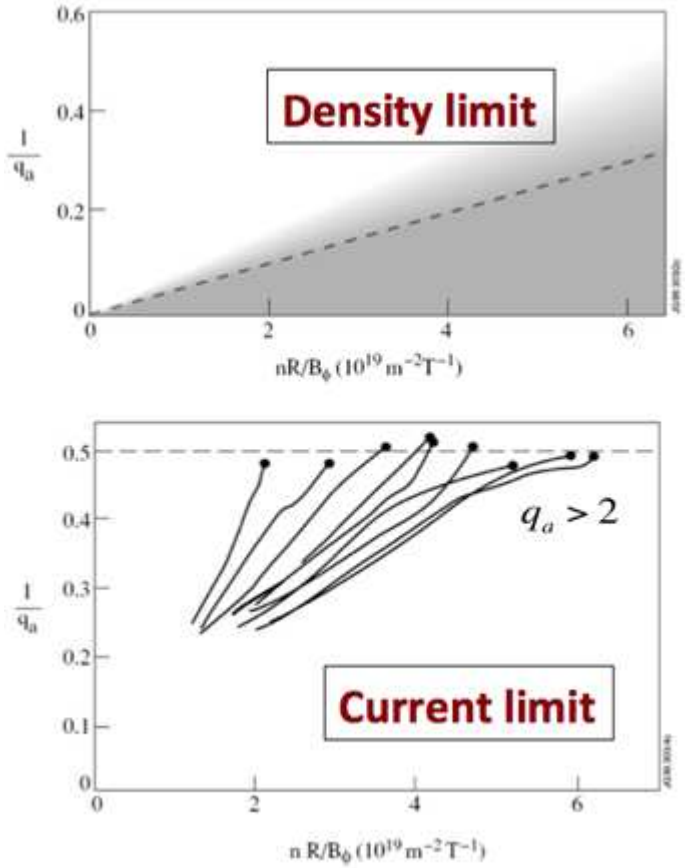


Figure 3.2: Hugill diagram: density limit (top); current limit (bottom) [4].

[5], the condition $1/q_a > 0.5$ or analogously $q_a > 2$ represents an hard limit because in the region where this condition is not satisfied, the external kink mode $m = 2, n = 1$ becomes unstable and the discharge will unavoidably disrupt. Considering the dependence between the safety factor at the edge and the plasma current, this is actually a limit on the maximum current for a given magnetic field. This mode could in theory be stabilized with a highly conductive wall surrounding closely the plasma, which is not possible because of the need to reduce the interaction of the plasma with the wall.

As the latter limit, also the pressure limit has an MHD origin. In particular, it is related to the *Troyon* ideal MHD limit [6] on the volume averaged toroidal beta β_t, which is, in other words, a limit on the maximum plasma pressure that can be confined by a given magnetic field. The calculation had been done taking into account ideal MHD instabilities as ballooning modes and Mercier criterion [7] for optimized plasma current and pressure profiles, and what had been found was that n=1 free boundary kink modes set a limit on the maximum achievable β.

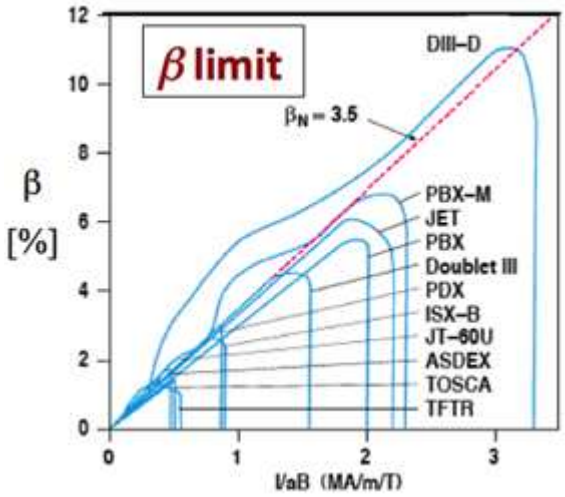


Figure 3.3: Beta limit in different tokamaks

As it is reported in Figure 3.3, the normalized volume average beta $\beta_N = \beta_{t(\%)}/[I_p(\mathrm{MA})/a(\mathrm{m})B(\mathrm{T})]$ should not exceed the value of approximately $3.5\%\mathrm{MA}/(\mathrm{m}\cdot\mathrm{T})$ [8]. It is important to highlight that these boundaries must not be considered rigidly, in fact there exist conditions in which the described limits can be exceeded, and, on the other hand, there are conditions far from these boundaries where the plasma however disrupts. This is due basically to the high complexity of the underlying physics and MHD stability on the base of the processes which drive disruption phenomenon, and

this is the reason whereby it is so important to develop different approaches as data-driven methods for disruption prediction.

3.2 Disruptions

3.2.1 Introduction

A disruption is a sudden loss of stability or confinement of a tokamak plasma: plasma energy is lost within a time span of few milliseconds exposing the plasma facing components to severe thermo-mechanical stresses and conductors surrounding the vessel to huge electromagnetic forces. As it has been introduced in the previous section, the operational space accessible to a tokamak is highly restricted by disruptive events. Moreover, disruptions, in addition to affecting the execution of the research program, can constitute a risk for the structural integrity of the machine, especially in large devices.

Therefore, it is particularly important, especially in view of ITER, to improve the understanding of the processes which lead to disruption. Deeply investigations have been carried out on precursors, causes and consequences of disruptive events. The main phases preceding a disruption are represented in figure 3.4.

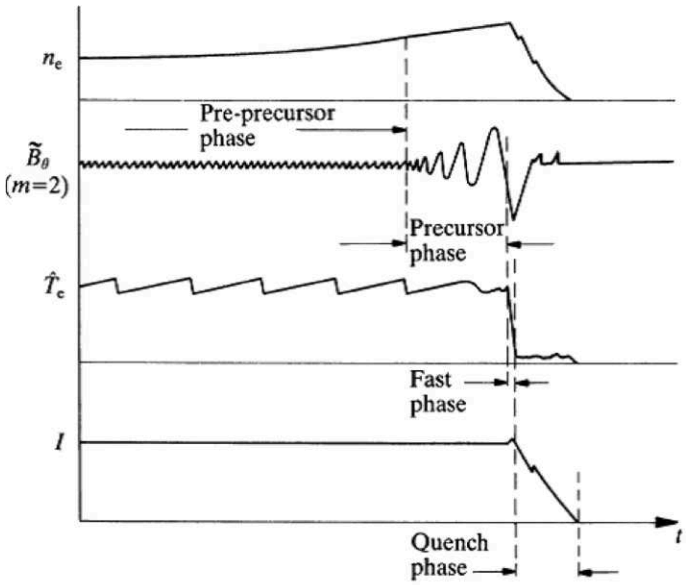


Figure 3.4: Main phases of a disruptions [5]

The pre-precursor and the precursor phase, which are often considered a

unique phase, refer to a change in the underlying conditions up to a critical point when there is the onset of an MHD instability.

The fast phase, also referred as thermal quench, is characterized by a central temperature collapse in few millisecond, followed by a rapid increase of plasma resistivity with a redistribution of the toroidal current and a flattening of its radial profile. The resulting flattened current profile is also associated to a consequent decrease of plasma internal inductance that, for flux conservation, gives usually rise to a characteristic spike of the plasma current and large transient negative loop voltage.

The final phase, referred as current quench, is characterized by the decay to zero of the plasma current: it is not uncommon to have current decays greater than 100[MA/s], whereas time scales are determined by the particular conditions in which the process sets up [5].

3.2.2 Main causes and mechanisms

The main physics instabilities which lead to disruption are directly related to the overcoming of the operational limits described in the initial section of this chapter. Moreover it is of primary importance to understand the underlying mechanisms at the base of the chain of events which characterize disruptions [9].

Depending on the conditions in which the discharge is evolving, the same modes which are often observed as precursors can lead to disruption or not. The complexity of the mechanisms which can get the plasma unstable makes the prediction very challenging.

The density limit, for example, is strongly connected to the mechanisms of radiation instability that builds up when the total radiated power exceeds the heating power. Plasma radiated power has different origins: Bremsstrahlung radiations, cyclotron radiations and the radiations due to line emissions. Radiated power from impurity ions represent the most important source of radiation in the plasma: besides enhancing Bremsstrahlung losses, the presence of impurities produces further losses due to line radiation and recombination with a power density equal to $P_R = R(T_e)n_e n_i$, where n_e is the plasma density, n_i id impurity ion density and $R(T_e)$ is the radiation efficiency.
Radiation instabilities can set up with different mechanisms [5] [10]. One of these is by radiation cooling of the plasma edge where impurity ions are not fully ionized: as the density increases at the edge, the temperature decreases and the line radiation from low-Z impurities is strongly enhanced. As we can see in the picture 3.5, radiation efficiencies have a peak at low temperatures. This produces a poloidally symmetric radiation at the plasma edge, where as more the temperature is reduced due to strong radiation losses, the more the

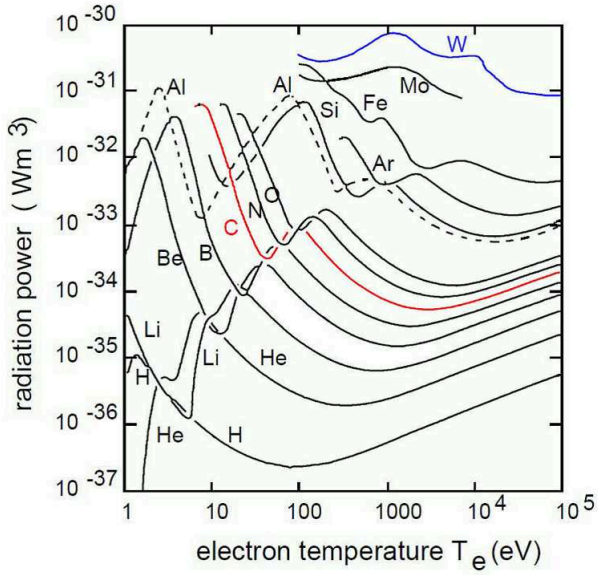


Figure 3.5: Radiation efficiency of impurities [3]

plasma radiation losses are enhanced, and this gives rise to further decrease of the temperature self-feeding the instability process.

When the density limit is reached, or, in other words, when radiation losses exceeds the heating power, the temperature collapse and the contraction of the plasma current profile by cooling edge makes the plasma unstable to MHD modes, leading eventually to disruption. This is the basic mechanism at the base of a radiative collapse.

Critical density scales with heating power and low effective charge state Z_{eff} [11]; therefore, increasing the heating power and reducing the impurity content in the plasma, it is possible to achieve higher values of density before to get into the density limit. In this conditions there can be the onset of another radiation limit, the MARFE (Multifaceted Asymmetric Radiation From Edge) [12], a poloidally asymmetric radiation instability which develops usually on the High Field Side (HFS) or near the X-point.

The conditions for the onset of a MARFE depend on plasma-wall interaction, flux of recycling neutrals of the working gas and heat flow from the plasma centre to the edge [13]. In this case the maximum achievable density does not depend on the input heating power as we have for a poloidally symmetric radiative collapse, but depends directly on the average current density, as well as it is clearly expressed by the Greenwald limit. The linear dependency between density and plasma current density is clearly shown in the Hugill diagram.

Another important cause of instability related to radiation is the impurity accumulation [14] [15]. High-Z impurity accumulation in the plasma centre gives rise to strong radiation due to the fact that atoms are not fully ionized. This in turn give rise to flattening or even a hollowing of the temperature profiles with a consequent decreasing of the current density in the centre due to raising of plasma resistivity. This picture is also characterized by hollow q profiles, with values of the safety factor on axis greater than one, and thus no sawthooth crashes. When this mechanism is amplified beyond a certain level the central temperature collapses causing internal disruptions due to the onset of MHD activity.

Regarding the MHD stability, as it has been discussed in the previous section, two basic restrictions on the accessible operational domain are imposed by the limit on the safety factor at the edge,which is a current limit, and by the β limit, which is a limit on the maximum plasma pressure which can be confined for a given magnetic field. The first one is related to the unstable external kink modes for m=2, n=1, whereas the ideal limit on β is imposed by free-boundary kink modes for n=1.

Assuming a non-zero plasma resistivity, the instabilities which may eventually deteriorate plasma confinement leading to a disruption are the tearing modes. These resistive instabilities are characterized by the development of magnetic islands due to magnetic flux reconnections, as shown in Figure 3.6. When such modes are destabilized and grow up to a level whereby the island saturates, the changes in the plasma current profile can determine a loss of confinement in an always larger region causing eventually a disruption. Magnetic island stability and evolution is governed by the Rutherford

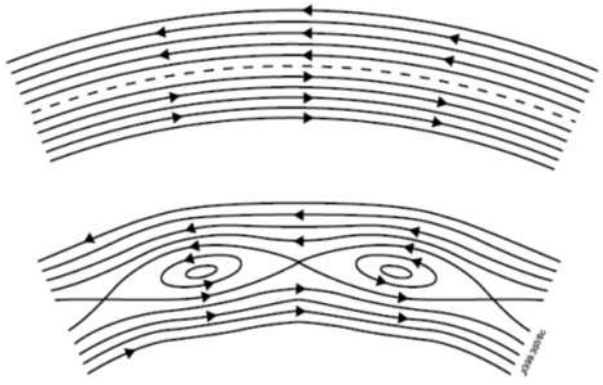


Figure 3.6: Reconnection and magnetic islands

equation (3.2):

$$\frac{\tau_R}{r_s^2}\frac{dw}{dt} = \Delta'(w) \tag{3.2}$$

where the growth or decay rate of an island of width w can be described in terms of local resistive time τ_R, minor radius r_s at the rational surface $q = m/n$ and a classical stability index Δ' (3.3).

$$\Delta'(w) = \frac{1}{B_r}\left(\frac{\partial B_r}{\partial r}\right)\Bigg|_{r_s+w/2}^{r_s-w/2} \tag{3.3}$$

There exist even situations where double tearing modes occur. Normally q profile is monotonous and increases from the axis, where it has its minimum, toward the edge, where it reaches its maximum value. This condition can also be described in terms of magnetic shear, always positive in these conditions. But in particular regimes or scenarios, as in the case of reversed shears, or strong impurity accumulation in the centre with hollow current density profiles, q profiles do not preserve the characteristic of monotonicity. We can have therefore coupling of the modes related to the same rational q-values and enhancing of the transport between the corresponding rational surfaces with the formation of magnetic islands, which, eventually, destroy the confinement and cause major disruptions.

Often tearing modes and magnetic islands are clear precursors of a disruption. When they start to stop, or do not rotate anymore together with the plasma fluid, they lock to the wall and grow with a time scale dependent on the resistive time constant of the surrounding vacuum vessel wall. The corresponding radial magnetic field perturbation induces eddy current in the wall whose magnetic field tends to oppose to magnetic island rotation exerting a force which has the effect to slow down and stop the island.

Besides eddy current forces, MHD instabilities such as locked modes, can also interact and be excited by *error fields (EFs)*, which are deviations of the magnetic fields from axi-symmetry. EFs are due mainly to non perfect alignment of the coils surrounding the plasma; they can excite modes making them grow until they lock to the wall and the plasma disrupts. Error fields can be compensated or reduced to a non-critical level through a dedicate system of external coils, the so-called Error Field Correction Coils (EFCCs).

Regarding the boundary on the maximum plasma pressure, ideal β limit is calculated for optimized current and pressure profiles; therefore, it depends on the particular conditions of operation. Some experiments showed that such a limit is only reached transiently. On the base of the boundary

conditions and the considered scenarios, different instabilities limit the maximum achievable β to a lower value, as Neoclassical Tearing Modes (NTMs) or Resistive Wall Modes (RWMs) [2].

NTMs are driven locally by the reduction of the bootstrap current which depends on the flattening of the pressure profile across a magnetic island with a consequent enhancing of the local radial transport. The most significant NTMs are characterized by mode numbers m=2, n=1, and m=3, n=2. They can be described by a modified Rutherford equation [16], where an additional term takes into account the reduction of bootstrap current. NTMs pose a serious problem for high performance scenarios, even if presently several techniques for their stabilizations have been studied and successfully applied as shown in the Figure (3.7). The effect on the energy confinement

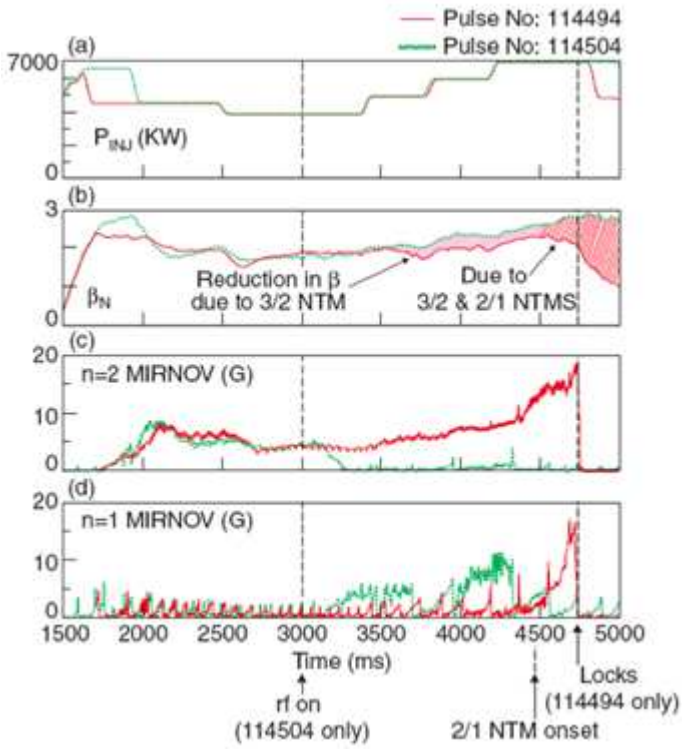


Figure 3.7: NTMs stabilization: two DIII-D discharges with (No. 114504, dotted lines) and without (No. 114494, solid lines) ECCD suppression of an m=3, n=2 NTM. (a) Neutral beam power, (b) β_N, (c) n = 2, (d) n = 1.(T.C. Hender et al., IPB2007, Chapter 3)

due to m=3, n=2 and m=2, n=1 NTMs can be seen in the evolution of β_N through the comparison of two discharges in DIII-D. The two pulses are more or less identical, with the presence of the same sequence of NTMs, but in the

discharge where stabilization by electron cyclotron current drive (ECCD) is performed the effect on β_N is well evident.

In conditions of high plasma pressure, RWMs can cause disruptions as well. High β plasmas are unstable to external kink modes, and this obviously represents a limit in the exploitation of high performance advanced scenarios with high bootstrap current fraction. External kink modes could be stabilized by a nearby conductive wall, allowing in principle to exceed the no-wall limit. In Figure 3.8 is reported the calculation of the time evolution of β_N for the no-wall limit with the corresponding measures of the MHD activity by Mirnov loops and photodiodes. But being a real wall characterized

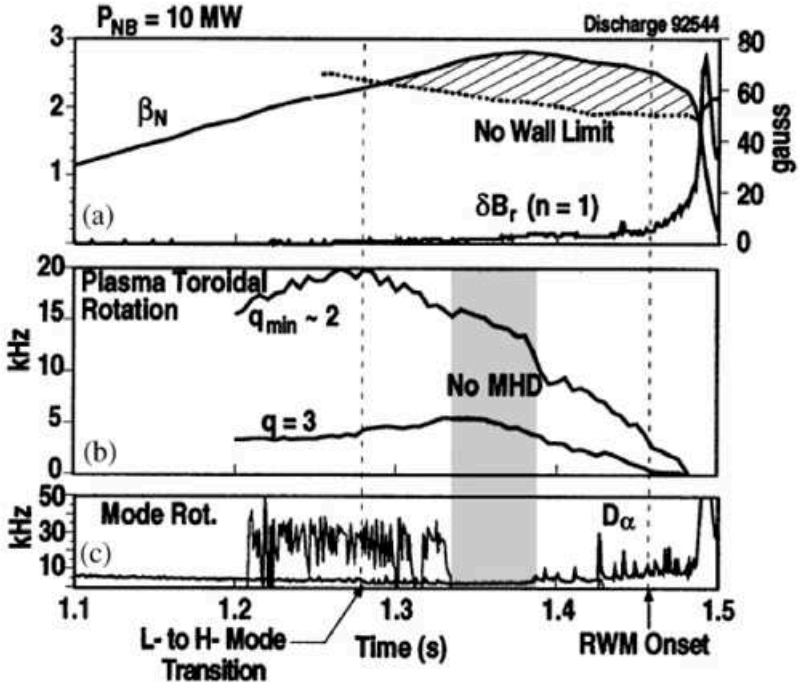


Figure 3.8: Time evolution of discharge No. 92544 showing (a) β_N relative to the computed no-wall limit and the saddle loop amplitude δ_{B_r} of the RWM, (b) measured plasma rotation from CER at $q = q_{min}$ and $q = 3$, and (c) MHD activity from Mirnov loops and photodiodes. (A.M. Garofalo et al., PRL. 82, 3811 (1999))

by a finite resistivity, the grow rate of the resulting resistive mode will now be governed by the resistive time constant of the wall. Also in the case of RWMs, different methods for stabilization have been demonstrated by several experiments, in particular by using active feedback coils systems and by plasma rotation.

In advanced scenarios with the presence of Internal Transport Barriers

(ITBs), Alfvenic growth rate instabilities can build up leading with very fast time scales to disruption. Being fast, they result to be particularly difficult to detect and typically give rise to the highest energies and heat loads. Plasmas characterized by ITBs exhibit radially localized regions of improved confinement with steep pressure gradients in the plasma core, which in turn could drive instabilities leading to disruption. In relation to the achievement of continuous operation it is well known that a large fraction of bootstrap current is necessary, and, that discharges exhibiting the formation of ITBs are favorable to this aim. Experimentally, the presence of such a current fraction is usually associated with high β discharges with a weakly positive or negative magnetic shear in the central region of the plasma column.

3.2.3 VDEs

Another cause of instability is the lost position control of plasma vertically elongated. It is well known that plasmas are elongated for reasons of stability and confinement. However, being unstable to vertical displacements, that is in the direction of elongation, it is necessary a feedback control stabilization system on plasma vertical position, based on poloidal field coils (see chapter 2). When the vertical control is lost, a Vertical Displacement Event (VDE) develops, inducing large forces on the surrounding structures. Therefore, these events are particularly dangerous for the integrity of the machine, even if the presence of conductive surrounding structures oppose to the displacement thanks to the induced currents, slowing down the vertical motion on the base of the resistive time. The loss of vertical control can be caused also by the rapid changes in plasma parameters during a disruption, but in some cases it occurs before the energy and the current quench, therefore it can be seen as a cause.

3.2.4 Consequences

As it has been discussed in the introductory section, disruptions represent a not negligible risk for the structural integrity of the machine. The thermal quench, that is the phase in which a large amount of thermal energy is lost from the plasma, can cause extremely high *thermal loads* on the plasma facing components (PFCs), on the diverter and in general on the first wall. Presently, no material could withstand all the thermal energy of a large device as ITER without being heavily damaged or directly melted. Obviously, there are mechanisms through which a consistent fraction of the thermal energy is dissipated before to be released on surface materials, as for example by radiation losses. Furthermore, we have to take into account that the total

heat flux has to be distributed on the largest possible area. Anyway, the foreseen heat loads are still too high for PFCs and divertor materials, therefore, further mitigation actions must be considered.

Besides heat loads, another serious issue is represented by the consequences of the plasma abrupt current quench: large *eddy currents* can be induced in the vacuum vessel and surrounding structures, creating forces potentially capable of damaging the device. Eddy currents are driven basically by the movement of the plasma column and by the variation of the plasma current values. Moreover, during disruptions the plasma can hit the first wall and a consistent fraction of plasma current can flows directly from the plasma to the vacuum vessel and the surrounding conductive structures through the wall-contacting region. The resulting currents flow mostly in the poloidal direction and are characterized typically by a toroidally symmetric component due to magnetic flux conservation, and eventually by a toroidally asymmetric component with mode number n=1, whose origin so far is not so clear (figure 3.9).

The forces induced by these so called *halo currents* can be very harmful. Eddy and halo currents give rise to vertical forces between the plasma column and the vacuum vessel and forces between the vacuum vessel and the coils. The problem of equilibrium and vertical stability has already been discussed in the chapter 3.
Finally, the production of *relativistic (runaway) electrons* during the current quench poses another threat to the integrity of the plasma facing components, especially in the case of high-current tokamaks as ITER. The conversion by Coloumb avalanche multiplication of plasma current to relativistic electron current can reach even 70% of the initial plasma current, leading to potential damages to PFCs.

Observations in present tokamaks have shown that runaway electron can cause damages due to the deposition of thermal energy on material surfaces. An additional fraction to this energy is originated from the conversion of the magnetic energy associated to the relativistic beam [17]. Thus, on the base of the always improving understanding of the nature of such a phenomenon, different avoidance or mitigation strategies are currently under study, with particular reference to their application in ITER.

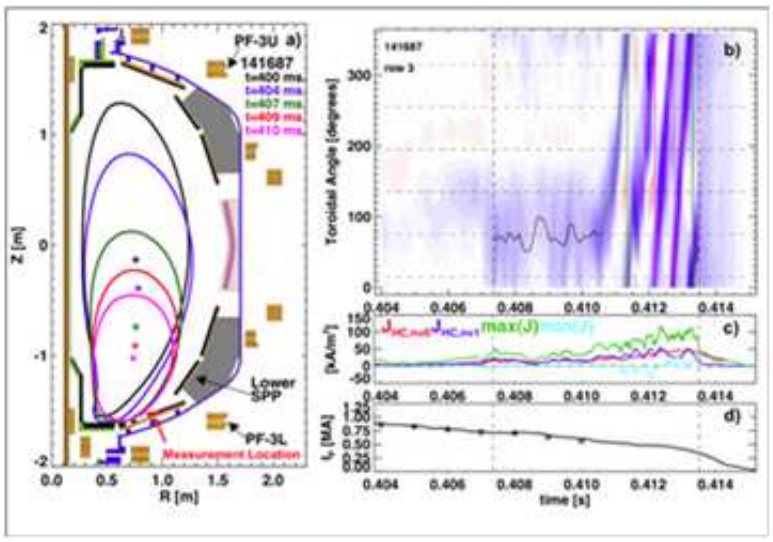


Figure 3.9: Example of halo current dynamics in NSTX: (a) vertical motion leading up to the disruption, (b) contours of halo current as a function of time and toroidal angle, (c) maximum and minimum current instantaneously measured on any tile, along with the amplitudes in a simple n = 1 decomposition, and (d) the plasma current.(from: [18])

Bibliography

[1] T.B. Kingery, John Wiley & Sons 2011, *Nuclear Energy Encyclopedia: Science, Technology, and Applications* a cura di Thomas B. Kingery

[2] H Zohm *et al.* 2003 Plasma Phys. Control. Fusion 45 A163, *MHD limits to tokamak operation and their control*

[3] M. Greenwald, 2002 Plasma Phys. Control. Fusion 44 R27, *Density Limits in Toroidal Plasmas*

[4] J. Wesson, 2000, *The Science of JET*, `http://www.jet.efda.org/documents/books/wesson.pdf`

[5] J. Wesson J. Oxford University Press 2004, *Tokamaks 4th edn*

[6] F. Troyon *et al.* 1984 Plasma Phys. Control. Fusion 26 209, *MHD-Limits to Plasma Confinement*

[7] J.P. Friedberg 1987 Plenum Press, New York, *Ideal MHD*

[8] T.C. Hender *et al.* 2007 Nucl. Fusion 47 S128, *Chapter 3: MHD stability, operational limits and disruptions*

[9] P.C. deVries *et al.* 2011 Nucl. Fusion 51 053018, *Survey of disruption causes at JET*

[10] H.R. Koslowsky 2008 Trans. Fusion Sci. Technol. 53 144, *Operational Limits and Limiting Instabilities in Tokamak Machines*

[11] F.C.Schuller, 1995 Plasma Phys. Control. Fusion 37 A135, *Disruptions in Tokamaks*

[12] B. Lipschultz, 1987 J. Nucl. Materials, 145-147 15, *Review of MARFE Phenomena in Tokamaks*

[13] J.Rapp 1999 Nucl. Fusion 39 765, *Density limit in Textor-94 auxiliary heated discharges*

[14] M.Z. Tokar *et al.* 1995 Plasma Phys. Control. Fusion 37 A241, *The Influence of Impurities on Limites Tokamak Plasmas and Relevant Mechanisms*

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

[16] O. Sauter *et al.* 1997 Phys. Plasmas 4 1654, *Beta limits in long-pulse tokamak discharges*

[17] Loarte A. *et al.* 2011 Nucl. Fusion 51 073004 *Magnetic energy flows during the current quench and termination of disruptions with runaway current plateau formation in JET and implications for ITER*

[18] S.P. Gerhardt 2013 Nucl. Fusion 53 023005 *Dynamics of the disruption halo current toroidal asymmetry in NSTX*

Chapter 4

Machine learning for mapping, prediction and classification

4.1 Introduction

Today the large amount of data available from fusion experiments and their character of high-dimensionality make particularly difficult handling, processing, understanding and extracting properly what is really important among all the available information. In fact very often data sets consists not only in a huge number of examples, but are also characterized by a consistent number of features necessary to exhaustively represent the behavior of a certain phenomenon for example. Obviously, not all the features have necessarily the same level of importance, or it can happen that some of them are redundant or completely useless in relation to a specific objective. This is a key point for several reasons: first of all, even if it is continuously increasing, there is a computational limit to the amount of data which can be handled because of the complexity of the algorithm, the required memory, etc. Furthermore, high-dimensionality makes data very difficult to interpret; scientists often have to deal with problems involving high-dimensional data.

The most obvious issue is visualization; when the data dimension is greater than three cannot be visualized and it becomes harder to perceive similarities and dissimilarities between different variables. Furthermore, the sampling of the space is harder due to the high number of possible data samples. Essentially, the amount of data to achieve a given spatial density of examples increases exponentially with the dimensionality of data space (empty space phenomenon). Generally speaking, algorithms that operate on high-dimensional data are faced with the "*curse of dimensionality*" and the associated issues, resulting in a very high complexity. For example, organiz-

ing and searching data relies on detecting areas where objects form groups with similar properties; in high-dimensional data however all objects appear to be sparse and dissimilar in many ways which prevents common data organization strategies from being efficient. One approach to simplification is to assume that the data of interest lies on a low-dimensional manifold, embedded in the high-dimensional space. Thus, data reduced to a small enough number of dimensions can be visualized in the low-dimensional embedding space. Attempting to uncover this manifold structure in a data set is referred to as manifold learning. It is worth mentioning that identifying the right manifold would also allow to better model the relevant physics. Therefore,manifold learning has the potential not only to improve the visualization and the intuitive estimation of problems but also to qualitatively increase the understanding of the relevant physics.

Moreover, beyond visualization, one has to take into account also the aspect of the computational burden required by pattern recognition, classification and prediction algorithms which usually are used immediately after the initial step of dimensionality reduction. In other words, reducing the quantity of relevant features in a data set is a fundamental step for the subsequent application of powerful data-analysis and machine learning techniques [1].

When we talk about data visualization and mapping, very often we are intrinsically making reference to the same concept, but sometimes some distinction are made among methods which provide just visualization and methods that provide a mapping. In the context of machine learning, mapping methods are considered mostly able to provide a preliminary feature extraction step, after which pattern recognition algorithms can be efficiently applied. Instead, data visualization methods can be considered as a subset of mapping methods based mostly on distance measurements and data proximity. Anyway, in many applications such a distinction becomes in practice inappreciable.

4.2 Manifold learning algorithms

In the last few years, many manifold learning techniques have been developed for dimensionality reduction. A number of supervised and unsupervised linear dimensionality reduction frameworks have been designed [2], which define specific procedures to choose interesting linear projections of the data such as PCA [3] and Grand Tour [4]. These linear methods can be powerful, especially in terms of data-visualization, but often miss important nonlinear structures in the data. Recently, several different algorithms have been de-

veloped to perform dimensionality reduction of nonlinear manifolds. Among them, there are powerful methods such as Self Organizing Map (SOM), Generative Topographic Mapping (GTM), Isomap and Locally Linear Embedding (LLE) [2].

Isomap is a simple method of nonlinear dimensionality reduction that extends metric multidimensional scaling (MDS) exploiting graph distance as an approximation of the geodesic distance, instead of the Euclidean distance. The main idea at the base of the method is to use the distance along a geodesic path onto the considered manifold as measure of dissimilarity. The mapping preserve the intrinsic metric of the data, therefore it can be defined as a distance preservation method.

LLE instead, similarly to SOM and GTM algorithms, is a topology preservation method. In mathematics, a topological variety or manifold is basically a topological space that resembles Euclidean space near each point, or in other words each point of an n-dimensional manifold has a neighborhood homeomorphic to the Euclidean space in $\mathbb{R}^n$. LLE defines a eigenvector based method, and its optimization don't involve an iterative algorithm, avoiding in this way the problem of eventual local minima.

The most important feature about manifolds is represented by their topology, or, in other words, the neighborhood relationships between subregions of the considered manifold. Nonlinear dimensionality reduction can be achieved also with distance preservation, but it turns out to be very constraining. In certain cases the embedding of a manifold requires that some regions has to be stretched or shrunk to be properly embedded in a lower dimensional space. This is the reason whereby generally topology preservation, even if more complex, seems to be more suitable in this framework.

Therefore, summarizing, dimensionality reduction is the process through which we can find a suitable representation of our original data, with the aim of discovering eventually particular structures or patterns which can lead to more targeted statistical analysis such as clustering, smoothing, probability density estimation and classification. In addition to these advantages, moreover, we have to consider the power of visualization if dimensionality is reduced to 2-D or 3-D.

LLE, unlike SOM and GTM, for preserving topology proposes a different approach based on the so-called *conformal mapping*, which, instead of preserving local distances, preserves local angles. In a certain way local distances and local angles are linked by scalar products, thus they may be interpreted as two different ways to preserve local scalar products [2]. Anyway, regarding the different methods, a not negligible point is the computational burden that has to be evaluated in relation to the specific application. Regarding Isomap and LLE for example, the spectral decomposition required by the

two algorithms can represent a heavy computational bottleneck. Going up with dimensionality and size of databases, their use becomes hard without a very powerful hardware configuration, and an eventual real-time application would be particularly challenging.

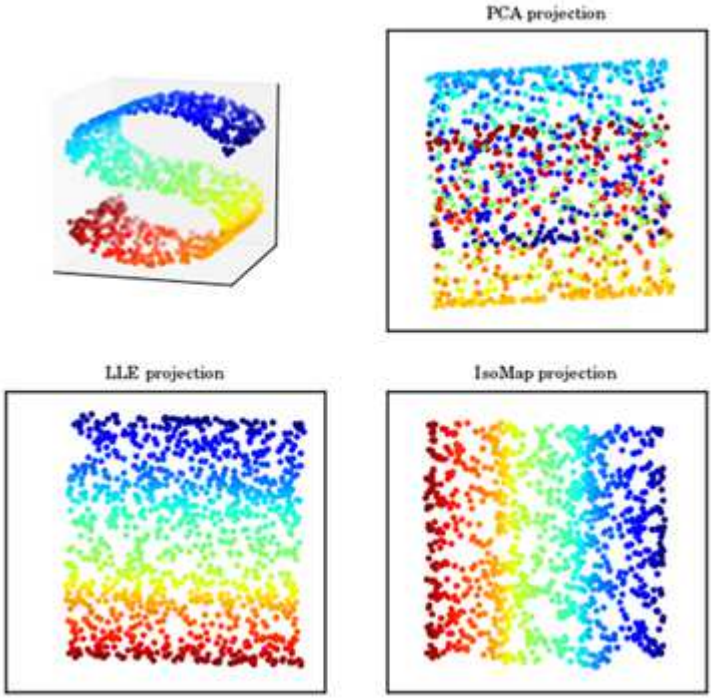


Figure 4.1: Comparison between PCA and Manifold Learning methods (LLE and Isomap). [from *www.astroml.org/book_figures*]

There exist several other algorithms for manifold learning, as well there exists also different variants of the cited algorithms, but an extensive discussion about all the methods is beyond the scope of this thesis. Therefore, only the methods applied for the analysis performed in the framework of this thesis will be described, in particular Grand Tour and Principal Component Analysis among the linear techniques, and Self Organizing Maps and Generative Topographic Mappings among the nonlinear ones. The linear techniques are simpler and easier to implement than more recent methods considering nonlinear transforms, but often miss important nonlinear structures in the data. In any case, they turn out to be very useful for an initial analysis about basic statistical properties and interesting linear structures hidden in data. Furthermore, some sections will be dedicated to the introduction of reference classification and prediction algorithms used in the framework of this thesis

in conjunction with manifold learning algorithms, such as k-Nearest Neighbor (kNN) technique and Conformal Predictors.

Let us consider the problem of reducing the dimensionality of a given data set consisting of N high-dimensional points in an Euclidean space. The high-dimensional input points will be referred to as $\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, ..., \mathbf{t_N}\}$ with $\mathbf{t_i} \in \mathbb{R}^D$. Let L be the dimensionality of the manifold that the input is assumed to lie on. The low-dimensional representations that the dimensionality reduction algorithms find will be referred to as $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}\}$ with $\mathbf{x_i} \in \mathbb{R}^L$.

4.2.1 Grand Tour (GT)

Usually, in order to discover some basic property of a dataset of interest, it is useful to start looking at data from different points of view, investigating the highest possible number of lower dimensional representations. This is a proper method of analysis, especially in those cases for which eventual structures hidden within data are totally unknown.

The Grand Tour method, introduced by Asimov [4] and Buja and Asimov [5], is a multivariate visualization method that generates a continuous sequence of low dimensional projections of a high dimensional data set. The animation obtained provides an overview of the high dimensional space in a sequence of 2-D plots. Data are looked from all possible viewpoints to get an idea of the overall distribution.
To create a two dimensional Grand Tour, a sequence of planes is generated. The set of planes has to be dense in the data space; the sequence of planes is also required to move continuously from one plane to the next so that the human visual system can smoothly interpolate the data and track individual points and structures in the data. Hence, the mathematics of the Asimov-Buja Grand Tour requires a continuous, space-filling path through the set of planes in the high-dimensional data space. Then, data has to be projected onto the planes and observed in a time-sequenced set of 2-2-DD images. Several algorithms have been proposed to achieve these two conditions, based on obtaining a general rotation in the high dimensional space. In this work, the MATLAB implementation in [6] of the Pseudo Grand Tour algorithm, firstly described in Wegman and Shen [7], has been used. The main advantages of the Pseudo Grand Tour, which is an approximate version of the Grand Tour, are speed, ease of calculation, uniformity of the tour, and ease of recovering the projection. However, the algorithm is not space filling, thus only a "pseudo" grand tour is obtained.

4.2.2 Principal Component Analysis (PCA)

The main purpose of Principal Component Analysis is to reduce dimensionality taking into account as much of the variance of our high-dimensional data as possible.

PCA finds the L directions (vectors) along which the data has maximum variance and the relative importance of these directions. If data lies perfectly along an embedding subspace of $\mathbb{R}^L$, PCA will reveal that subspace; otherwise, PCA will introduce some errors. Let the first L principal components of $\mathbf{T}$ be $\mathbf{P} = [\mathbf{p_1}, ..., \mathbf{p_L}]$ with $\mathbf{p_i} \in \mathbb{R}^D$.

The columns of $\mathbf{P}$ are the directions of maximum variation within the data, and they form an orthonormal basis that spans the principal subspace so there is no redundant information [3]. The data $\mathbf{x_i}$ can be approximated by linear combination of the principal components as $\mathbf{x_i} = \mathbf{P^T t_i}$, where $\mathbf{P^T t_i} = c_i$ are the linear coefficients obtained by projecting the training data onto the principal subspace; that is, $\mathbf{C} = [\mathbf{c_1}, ..., \mathbf{c_N}] = \mathbf{P^T T}$.

Despite PCA's popularity it presents a number of limitations. The main drawback is the requirement that the data lies on a linear subspace. Indeed, when data lies in a low-dimensional manifold, not in a low dimensional subspace, PCA does not correctly extract the low-dimensional structure. Manifold learning algorithms essentially attempt to duplicate the behavior of PCA, but on nonlinear manifolds instead of linear subspaces.

4.2.3 Self Organizing Map (SOM)

The SOM is a type of artificial neural network developed by Kohonen [8]. SOMs are widely applied as nonlinear dimensionality-reduction tools in order to convert complex nonlinear relationship between data items into a low-dimensional space. A SOM can be intuitively interpreted as some kind of nonlinear PCA. In a SOM the objective is more to preserve the topology, rather than the distance, in the distribution of the data.

One natural way to put this idea in practice consists of replacing the hyper-plane with a discrete (and bounded) grid or lattice defined by some points called prototypes. The prototypes have coordinates in both the embedding and the initial space. They are iteratively fitted inside the data cloud moving the prototypes together with their neighbors in the lattice toward the original data points as it is shown in Figure 4.2. Hence, the Self-Organizing Map is a nonlinear dimensionality reduction technique which performs two concurrent subtasks:

- Dimensionality reduction: high dimensional inputs are projected on a low-dimensional regular grid.

- Data clustering and topology preservation: points close to each other in the input space are mapped to the same or neighboring clusters in the output space.

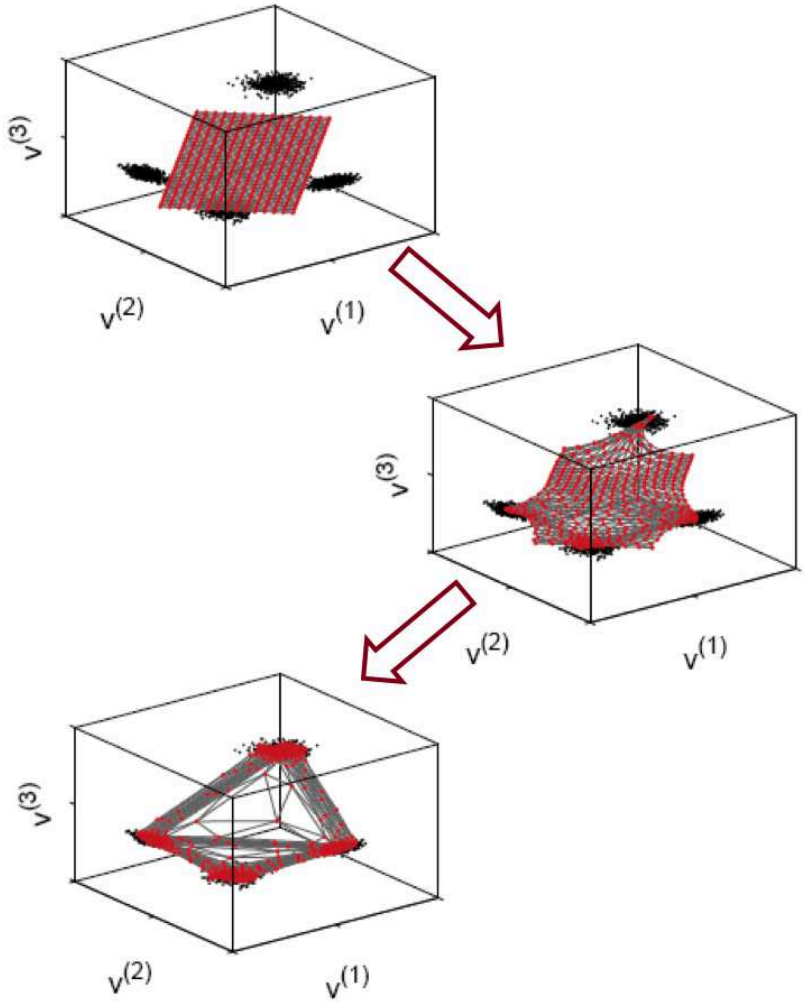


Figure 4.2: Self Organizing Map: prototypes iterative fitting inside the data cloud

Let us consider in more detail the problem of reducing the dimensionality of a given data set consisting of high-dimensional points in Euclidean space. The SOM replaces the set of points $\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, ..., \mathbf{t_N}\}$ in the D-dimensional input space $\mathbf{T}$ onto the smaller set of K prototypes points $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_K}\}$ with $\mathbf{x_i} \in \mathbb{R}^L$. Each prototype point in the low-dimensional regular lattice corresponds to a point in the original space. Moreover, SOM preserves the topological properties of the input. This means that points close to each other

in the input space are mapped on the same or neighboring prototypes in the embedding space. Preserving neighborhood's relations in the mapping makes possible to see more clearly the structure hidden in the high-dimensional data. The coordinates $\mathbf{x}$ are initialized and then updated iteratively during the SOM training procedure. The SOM runs through the data set $\mathbf{T}$ several times, called epochs. During each epoch, for each $\mathbf{t_i}$, the closest prototype vector $\mathbf{x_r}$ is determined. Then, the coordinates of all the prototypes are updated according to the learning rule

$$\mathbf{x_i} = \eta\Lambda(i,r)(\mathbf{t_i} - \mathbf{x_r}) \tag{4.1}$$

The neighborhood function $\Lambda(i,r)$ is equal to 1 for $i = r$, and falls off exponentially with the distance d_{ir} between prototypes i and r in the lattice. Thus, prototypes close to the winner r, as well as the winner itself, have their coordinates updated, whereas those further away, experience little effect. Learning generally proceeds in two broad stages: a shorter initial training phase, in which the map reflects the coarser and more general patterns in the data, followed by a much longer fine tuning stage, in which the local details of the organization are refined. We start with a wide range of $\Lambda(i,r)$ and η then both the range of $\Lambda(i,r)$ and the value of η are gradually reduced as the learning proceeds. A typical choice for$\Lambda(i,r)$ is:

$$\Lambda(i,r) = e^{-d_{ir}^2/2\sigma^2} \tag{4.2}$$

where σ is a width parameter that is gradually decreased. Thus, the SOM simultaneously performs the combination of three concurrent subtasks: vector quantization, dimensionality reduction and topology preservation.

4.2.4 Generative Topographic Mapping (GTM)

Generative Topographic Mapping belongs to the class of the so called "generative models", which try in a certain way to model the distribution of the data by defining a density model with low intrinsic dimensionality in the data space. Through a nonlinear mapping from the latent space to the data space it generates a mixture of Gaussians, whose centers are constrained to lie on, a low dimension space embedded in the high-dimensional one and has to be fitted to the data. This is usually achieved through a form of the Expectation Maximization algorithm (EM) by maximizing the likelihood or the log-likelihood function of the model [9].

In a certain way, GTM has been inspired by the SOM algorithm, attempting to overcome its limitations. In particular, SOM does not define a density model and the convergence of the prototype vectors are not based on the optimization of an objective function such as the likelihood function, in fact the preservation of the neighborhood structure is not guaranteed. Being a generative latent model, GTM basically tries to find a representation in terms of a small number of latent variables: in order to be able to visualize the lower dimensional representation of the data, the latent variable dimension must be 2 or 3. Since the mapping is defined from the latent space to the data space, for visualization purposes an inversion of the mapping itself is required and this is achieved computing the posterior probability in the latent space through the Bayes' theorem.
However, we have to take into account that a single data point correspond to a probability distribution in the latent space, not just to a single point, reason for which we usually make reference to condensed information such as the mean or the mode of the posterior distribution.

Let's describe now in more detail the basic mathematical formulation upon which GTM is based. GTM defines a mapping from the latent space (L-dimensional space) into the data space (D-dimensional space). So, given a dataset in the data space $\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, ..., \mathbf{t_N}\}$, the first step is to map the latent space, which consists of a regular grid of nodes $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_K}\}$, into the data space through a parameterized nonlinear function $y(\mathbf{x}; \mathbf{W})$, where W is the matrix of parameters representative of the mapping (see figure 4.3). The objective of the GTM is to define a probability distribution over the D- dimensional space in terms of latent variables:

$$p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{4.3}$$

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

$$p(\mathbf{t}|\mathbf{x}, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{-D/2} \cdot e^{\left\{-\frac{\beta}{2}||y(\mathbf{x};\mathbf{W})-\mathbf{t}||^2\right\}} \tag{4.4}$$

where the inverse of the β parameter is the noise variance. The distribution in $\mathbf{T}$-space, for a given value of $\mathbf{W}$, could then be obtained by integration over the x-distribution. Since the integral is generally not analytically tractable,

the latent variable distribution is replaced by a prior distribution $p(\mathbf{x})$ consisting of a superposition of delta functions, each one associated with one of the nodes of the regular grid in the latent space

$$p(\mathbf{x}) = \frac{1}{K} \cdot \sum_{k=1}^{K} \delta(\mathbf{x} - \mathbf{x_k}) \tag{4.5}$$

Substituting 4.4 and 4.5 in 4.3, the distribution function in the data space becomes:

$$p(\mathbf{t}|W, \beta) = \frac{1}{K} \cdot \sum_{k=1}^{K} p(\mathbf{t}|\mathbf{x_k}, \mathbf{W}, \beta) \tag{4.6}$$

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

$$y(\mathbf{x}, \mathbf{W}) = \mathbf{\Phi}(\mathbf{x}) \cdot \mathbf{W} \tag{4.7}$$

where $\mathbf{W}$ is a $D \times M$ matrix of weight parameters and M is the number of the basis functions.

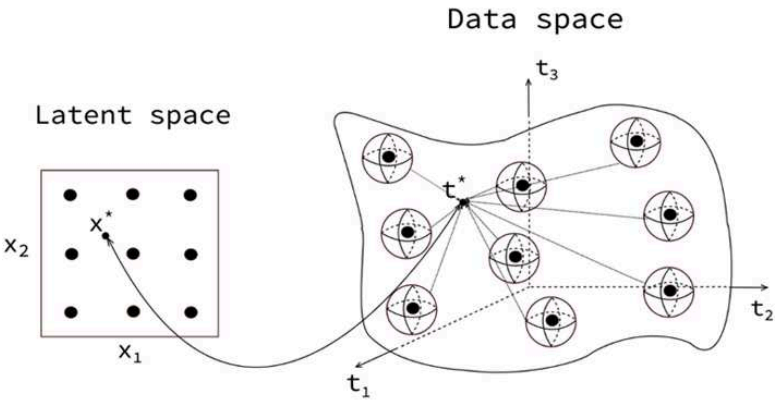


Figure 4.3: GTM mapping and manifold: each node located at a regular grid in the latent space is mapped to a corresponding point $y(\mathbf{x}; \mathbf{W})$ in the data space, and forms the centre of a corresponding Gaussian distribution. In the figure the correspondences between a data point in the manifold embedded in the data space and the mean of the posterior distribution in the latent space is also shown.

The adaptive parameters of the model are $\mathbf{W}$ and β. Since the GTM represents a parametric probability density model, it can be fitted to the data set by maximum likelihood, e.g. maximizing the log likelihood function. This can be performed, e.g., using the expectation-maximization algorithm.
The *likelihood function* for a set of i.i.d data points $\{\mathbf{t_1}, \mathbf{t_2}, ..., \mathbf{t_N}\}$ can be written as:

$$L = \prod_{n=1}^{N} (p(\mathbf{t}|\mathbf{W}, \beta)) = \prod_{n=1}^{N} \left(\frac{1}{K} \sum_{k=1}^{K} p(\mathbf{t_n}|\mathbf{x_k}, \mathbf{W}, \beta) \right) \tag{4.8}$$

therefore, the *log-likelihood function*, whose handling is usually more efficient, has the following form:

$$l = \sum_{n=1}^{N} ln \left(\frac{1}{K} \sum_{k=1}^{K} p(\mathbf{t_n}|\mathbf{x_k}, \mathbf{W}, \beta) \right) \tag{4.9}$$

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

Visualization

For visualization purposes, the resulting mapping in the high-dimensional space has to be transposed into the low-dimensional latent space, which is therefore chosen to be 2-D or three-dimensional (3-D). Extra dimensions would improve the quality of the results, but data with more than two or three dimensions can be difficult to interpret. The inversion of the mapping is performed by employing Bayes' theorem, which allows calculating the posterior probability in the latent space. Once we have found suitable values $\mathbf{W}^*$ and β^* for respectively the matrix of weight and biases for the nonlinear mapping and for the inverse of the noise variance, GTM defines a probability distribution in the data space conditioned on the latent variable, that is $p(\mathbf{t}|\mathbf{x_k})$ with $k = 1, 2, ..., K$. But what we are interesting in is the

corresponding posterior distribution in the latent space for any given data point $\mathbf{t}$, that is $p(\mathbf{x_k}|\mathbf{t})$; therefore, in order to compute this latter we can use the Bayes' theorem in conjunction with the prior distribution over the latent variable $p(\mathbf{x})$, as it is calculated in the following expression (4.10):

$$p(\mathbf{x_k}|\mathbf{t}) = \frac{p(\mathbf{t_n}|\mathbf{x_k}, \mathbf{W}^*, \beta^*) \cdot p(\mathbf{x_k})}{\sum_{k'=1}^{K} p(\mathbf{t_n}|\mathbf{x_k}, \mathbf{W}^*, \beta^*) \cdot p(\mathbf{x_{k'}})} \tag{4.10}$$

For visualizing all the data points in the latent space, it is then possible to plot the mean (4.11) or the mode (4.12) of the posterior probability distribution in the latent space.

$$\mathbf{x}_n^{mean} = \sum_{k=1}^{K} \mathbf{x}_k \cdot p(\mathbf{x_k}|\mathbf{t_n}) \tag{4.11}$$

$$\mathbf{x}_n^{mode} = argmax\left\{p(\mathbf{x_k}|\mathbf{t_n})\right\} \tag{4.12}$$

The mean position $\mathbf{x}_n^{mean}(t)$ in the latent space is calculated by averaging the coordinates of all nodes taking the posterior probabilities as weighting factors. In figure 4.3, the data point $\mathbf{t}^*$ is represented in the latent space as the mean weighted by the posterior probabilities.

Algorithm and implementation [10]

A scheme which summarizes the basic steps for the GTM construction model is given in the flowchart in figure 4.4.
The Matlab toolbox for the computation of the GTM which has been used as a base for the implementation of the data analysis and classification algorithms is part of Exploratory Data Analysis (EDA) toolbox described in [11]. The first step of the computation is the generation of the grids of the latent points and of the radial basis function centers. Regarding the radial basis functions in particular, the width σ is an important parameter, since in conjunction with their number and with the number of latent points, determines smoothness and flexibility of the mapping. Therefore,it is important to note that even if for computational reasons the algorithm works with a discrete number of latent points, the mapping is continuous over the latent space. In fact, it has to correspond to the manifold embedded in the data space where the centers of Gaussians (corresponding to the latent points) lie on. The choice of these parameters, as suggested by the main author of the tool [10], in general is not uniquely defined since it depends on the specific case, but the important point is that the choice will affect the final mapping.

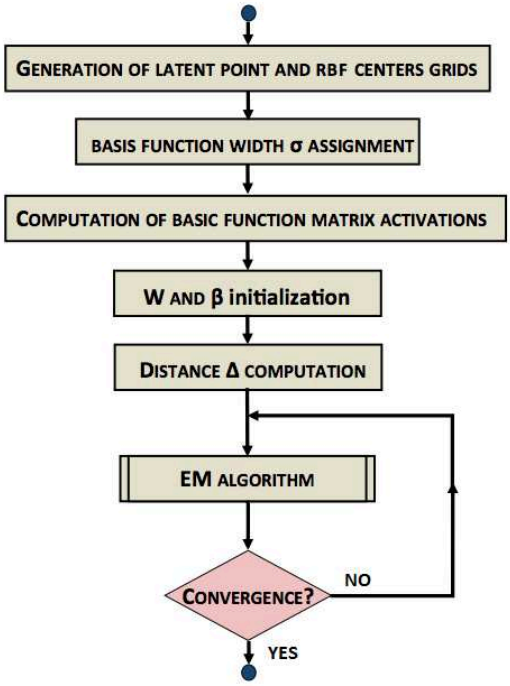


Figure 4.4: GTM algorithm flowchart

As general consideration, depending on the RBF width and their number, we can have larger or smaller overlapping among them, that means we can have more or less correlation. The more and broader the basis functions are, the higher the flexibility of the mapping will be, but we don't have to lose completely the correlation among RBFs, otherwise also the smoothness of the mapping will be lost. A measure of the overlapping among RBFs is given also by the number of points they have in common; therefore, in order to preserve the smoothness, we have to guarantee that the number of shared points is not too low. Regarding the number of latent points, the author suggests as a good rule to have $O(10^L)$ number of latent points as support of each basis function.

Regarding the nonlinear mapping (4.7), a generalized linear regression model is usually chosen as parametric nonlinear model, whereas regarding basis functions, several types could be used, but in the adopted implementation Gaussian basis functions are used. Once the matrix $\mathbf{\Phi}$ of basis functions has been computed, the initialization of W can be done randomly or PCA-based, and β has to be initialized coherently with respect to W initialization.

The next step is represented by the calculation of the distance $\mathbf{\Delta}$ between any given data point and the Gaussian centers to which latent points are mapped ($\Delta_{kn} = ||\mathbf{t}_n - \mathbf{\Phi}_k \mathbf{W}||^2$). At this point we enter in the iterative pro-

cedure for the mixture of Gaussians fitting through the EM algorithm. In the Expectation step the responsibilities r_{kn} that the n-th point $\mathbf{t}$ in the data space is generated from the k-th node of the grid are calculated according to the following expression:

$$r_{kn} = p(\mathbf{x_k}|\mathbf{t_n}, \mathbf{W}, \beta) = \frac{p(\mathbf{t_n}|\mathbf{x_k}, \mathbf{W}, \beta)}{\sum_{k'=1}^{K} p(\mathbf{t_n}|\mathbf{x_k}, \mathbf{W}, \beta) \cdot p(\mathbf{x_{k'}})} \tag{4.13}$$

Such responsibilities are the weights in function of which the parameters W and β are updated at each iteration until a convergence criterion will not be satisfied (usually the maximum number of iterations). In other words, in the Maximization step each component of the mixture of Gaussians is moved toward dataset points for which it results to have higher responsibility. A schematic representation of main steps of the Expectation Maximization for GTM building model is reported in the box of figure 4.5

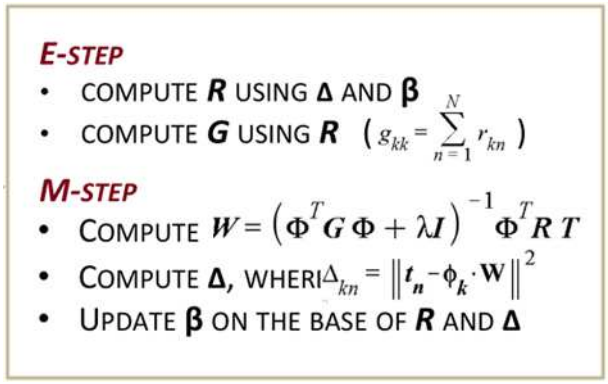


Figure 4.5: EM main steps.

4.2.5 Extension of the GTM tool for data analysis, prediction and classification

A not negligible part of the work carried out in the framework of this thesis has regarded the implementation of algorithms for data analysis, classification and prediction, which basically are an extension of the basic GTM tool (and can be applied also to SOMs). The developed tools, which will be described in conjunction with the results in the following chapters, provide additional functions related to the mapping of an high-dimensional space, in particular:

- evaluation and quantification of the effectiveness of the mapping (Quantization Error measure, Trustworthiness measure, and Topology Preservation measure).
- implementation of different types of representation (basic maps, Pie-planes and Component-planes).
- tracking of temporal evolution of a new object onto the map (online and real-time implementation)
- data-reduction algorithm based on the GTM model

GTM's tools could be particularly useful in the study of the operational space where the relevant physics takes place, allowing the perception of eventual similar patterns and the identification of dependencies or complex relations in the feature space. Furthermore, these tools have been used not only for analysis but also as "kernel" for the algorithms of prediction and classification, as it will described in the subsequent part of the thesis.

4.2.6 k-Nearest Neighbor (k-NN)

k-Nearest Neighbors algorithm (k-NN) is a reference non-parametric method used for classification and regression. In pattern recognition, it represents one of the simple but at the same time used learning algorithm. An object can be classified on the base of its neighbor by a majority vote: the class membership will indicate the class with the higher number of neighbors among the k nearest ones (figure 4.6).

k-NN is defined as an instance-based classifier, unlike GTM for example, which defines a generative latent model. There are several implementation of this algorithm, such us the weighted version for taking into account the different importance of the neighbors on the base of the distance to the test unlabeled point. k-NN technique requires the definition of a similarity measure, or in other words a distance measure. The most common used metrics is the Euclidean distance, but also other metrics such as Hamming distance can be used depending on structure and properties of the data of interest. It is a simple and flexible technique whose drawbacks are well known, as for example the application of the basic majority voting criterion for classification when the dataset is strongly unbalanced in terms of the different classes. In this case the class with higher frequency of occurrence can distort the majority vote among k nearest neighbors. One solution to overcome this problem is to take into account the distance of each of the k nearest neighbors with a weighted sum: a common rule is to multiply simply

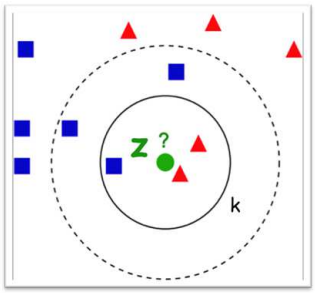


Figure 4.6: k-Nearest Neighbor technique with k=3: in thi case the test point z is classified as triangle.

for a factor proportional to the inverse of the distance from the considered point to the test unlabeled point.

Anyway k-NN has some strong consistency results. In particular the algorithm is guaranteed to yield an error rate no worse than twice the *Bayes error rate* if the amount of data tends to infinity [12]. Bayes error rate is referred to the optimal decision boundary that provides the lowest probability of error for a classifier, given distribution of data [13].

Mahalanobis distance [14]

A particular metrics which has been exploited as similarity measure with the k-NN technique, is the Mahalnobis distance, whose definition and intuitive picture are represented in Figure 4.7. If we are considering the problem of estimating the probability that a test point belongs to a certain set, intuitively, it is quite easy to deduce that the closer the point in question is to the center of mass of the distribution of points , the more likely it belongs to the set. When the considered distribution of points is not spherical then the probability of the test point to belong to the set, depends not only on the distance , but also on the direction.

Therefore,for a multivariate vector $\mathbf{x} = (x_1, x_2, ..., x_N)$, assuming a generic distribution of points with center of mass $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_N)$ and whose probability distribution is represented by the covariance matrix $\mathbf{S}$, Mahalanobis distance is defined as $D_M = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{S}^{-1} ((\mathbf{x} - \boldsymbol{\mu})}$.

In Figure 4.7 for example we are assuming that the distribution of points is ellipsoidal. In those directions where the ellipsoid has a short axis the test point must be closer, while in those ones where the axis is long the test point can be further away from the center, always maintaining the same probability to belong to the considered set. The ellipsoid that best represents the set's probability distribution can be estimated by building the covariance matrix

of the samples. The Mahalanobis distance is simply the distance of the test point from the center of mass divided by the width of the ellipsoid in the direction of the test point.

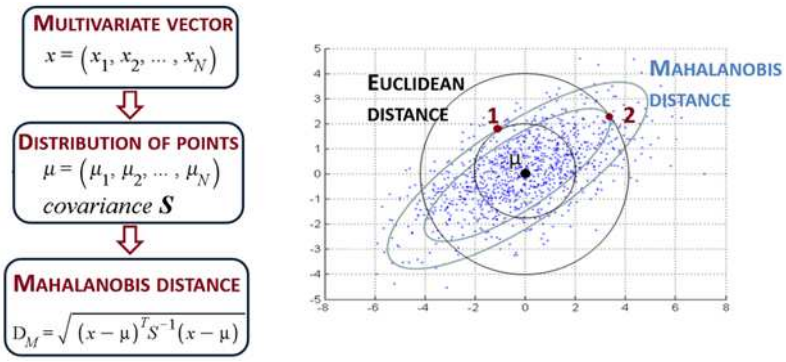


Figure 4.7: Comparison between Mahalanobis distance and Euclidean distance.

4.2.7 Conformal Predictors

Conformal predictors are placed in the wide framework of the machine learning techniques that have been developed for prediction and classification purposes. Unlike others methods, they have the peculiarity to provide together with prediction or classification also the corresponding level of confidence. The theory of Conformal Predictions is based on the principles of algorithmic randomness, and on the Kolmogorov complexity of an i.i.d. (identically independently distributed) sequence of data instances [15].

Conformal predictors can be used in principle with any method of prediction, such as support vector machines, neural networks, decision trees, nearest neighbor classifiers, etc. To determine the confidence level for the classification of a new object, it is necessary to estimate how different a new object is from the old examples: to this purpose, usually a nonconformity score is calculated on the base of a defined nonconformity measure. In particular we are interested to predictions using features of the new object; let's consider successive n ordered pairs $(\mathbf{t}_1,y_1)$, $(\mathbf{t}_2,y_2)$, ..., $(\mathbf{t}_n,y_n)$, where $\mathbf{z}_i = (\mathbf{t}_i,y_i)$ represents the generic example, which consists of an object $\mathbf{t}_i$ and the corresponding label y_i. Both the object and the labels belong to measurable spaces, respectively the object and the label space.

Conformal prediction requires firstly the definition of a nonconformity measure, which quantifies how different a new example is from old examples [16]. A bag of size $n \in \aleph$ is a collection of n elements that may be identical and can be given in any order. In the following we will refer to a bag of size

n with the notation $\langle \mathbf{z}_1, ..., \mathbf{z}_n \rangle$. The first step of the conformal prediction algorithm is the computation of the nonconformity scores α_i for any object of the given bag on the base of a defined nonconformity measure A:

$$\alpha_i := A(\langle \mathbf{z}_1, ..., \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, ..., \mathbf{z}_n \rangle, \mathbf{z}_i) \tag{4.14}$$

Nevertheless, nonconformity scores have not an absolute value, being relative to the particular case considered for the given bag of objects $\langle \mathbf{z}_1, ..., \mathbf{z}_n \rangle$. Therefore, in order to generalize and give a measure of how unusual an element z_i is with respect to the other elements of the bag, its score must be compared with the one of all the other objects. This can be done, for example, by computing the so-called p-value, which is defined by the fraction:

$$p\text{-}value = \frac{\#|\{j = 1, ..., n : \alpha_j \geq \alpha_i\}|}{n} \tag{4.15}$$

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

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

$$\begin{aligned} \textbf{Confidence} &= 1 - 2^{nd}\ largest\ p\text{-}value \\ \textbf{Credibility} &= largest\ p\text{-}value\ (max(p_j),\ j = 1, ..., N) \end{aligned} \tag{4.16}$$

The values of credibility and confidence are indicative of the reliability with which the classification is provided. In particular, assuming that each class is statistically well represented in the training set, a low value of credibility means that the new object (test) is not representative of any class of objects in the bag (training set). Another important point is represented by the fact that the maximum *p-value* is not necessarily defined in unique way, in the

sense that the maximum *p-value* could be attributed to more than one class. This is a case of ambiguity, that means the conformal predictor for the given training set, on the base of the defined nonconformity measure, is not able to discriminate among the classes which the maximum *p-value* is associated with.

As it has been anticipated at the beginning of this section, the nonconformity score can be computed in different ways. For the classification purpose of this work the conformal predictor will be based on the nearest neighbor technique. When a new example $\mathbf{z}_n = (\mathbf{t}_n, y_n)$ is given to the conformal predictor for classification, the nearest neighbor technique finds the object $\mathbf{t}_i$ of the training set closest to the new one ($\mathbf{t}_n$) and assign its label y_i to the label y_n to be predicted. At this point, in order to quantify the goodness of the prediction, we have to compare the distance of the nearest object $\mathbf{t}_i$ with the distance of the nearest neighbor with a different label with respect to the one previously attributed to the test object. According to this considerations, the nonconformity scores can be computed as:

$$\alpha_i = \frac{min\{|\mathbf{t}_j - \mathbf{t}_i| : 1 \leq j \leq n \;\&\; j \neq i \;\&\; y_i = y_j\}}{min\{|\mathbf{t}_j - \mathbf{t}_i| : 1 \leq j \leq n \;\&\; j \neq i \;\&\; y_i \neq y_j\}} \tag{4.17}$$

$$= \frac{distance\ to\ z's\ nearest\ neighbour\ with\ the\ same\ label}{distance\ to\ z's\ nearest\ neighbour\ with\ a\ different\ label}$$

4.2.8 Logistic regression

Classification is one of the most important topics in statistic and machine learning, and a simple approach to it is to come up with a rule which provide a discrete output (binary if the discrimination is between two classes) depending on the input variables. But in many cases, for example if we want to take into account the eventual presence of noise in our data, a discrete output is not the best rule, but probably we would like to provide an answer with a probability or a level of confidence.
If we define $\mathbf{t}$ as the input variable and Y as the output variable, this could be done simply by considering the conditional distribution of Y given the input variable $\mathbf{t}$, that is $P(Y|\mathbf{t})$.
Let's consider Y as a binary or dichotomous output variable which is coded as 0 or 1. The logistic regression models the probability that a generic sample belongs to a class 0 or 1 using $\mathbf{t}$ as independent variable or predictor. This probability is formally defined as:

$$log\frac{p(\mathbf{t})}{1 - p(\mathbf{t})} = \alpha + \beta\mathbf{t} \tag{4.18}$$

Therefore,$p(\mathbf{t})$ will be given by:

$$p(\mathbf{t}) = \frac{e^{\alpha+\beta\mathbf{t}}}{1+e^{\alpha+\beta\mathbf{t}}} \tag{4.19}$$

To minimize the misclassification rate, we should predict $Y = 1$ when $p \geq 0.5$ and $Y = 0$ when $p < 0.5$. Therefore, logistic regression gives us a linear classifier, whose decision boundary separating the two predicted classes is nothing else that the solution of $\alpha + \beta\mathbf{t} = 0$. In Figure 4.8 the logistic curve is represented.

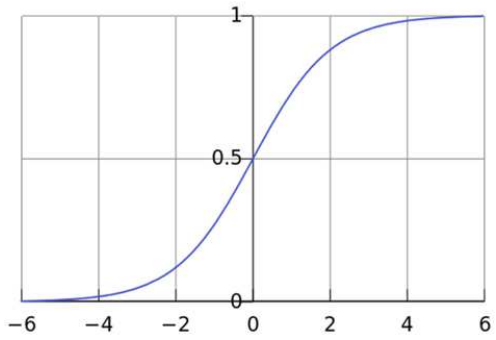


Figure 4.8: Logistic curve (from *http://en.wikipedia.org/wiki/Logistic_ regression*).

If $p(\mathbf{t})$ is the probability of the event, the *odds* of the event is defined as:

$$odds = \frac{p(\mathbf{t})}{1-p(\mathbf{t})} \tag{4.20}$$

The logistic model (*logit*) is based on a linear relationship between the natural logarithm of the odds of an event and a numerical independent variable; therefore, we can express the logistic regression as:

$$logit = log(odds) = \alpha + \beta\mathbf{t} \tag{4.21}$$

Bibliography

[1] L. Cayton 2005 *Algorithms for manifold learning*, available from `psu.edu`

[2] Lee J.A. and Verleysen M. Springer 2007 *Nonlinear Dimensionality Reduction*

[3] Jolliffe I.T. 1986 Berlin: Springer, *Principal Component Analysis*

[4] Asimov D. 1985 SIAM Journal on Scientific and Statistical Computing, 6(1):128- 143, *The Grand Tour: a Tool for Viewing Multidimensional Data*

[5] A. Buja and D. Asimov 1986 In D.M. Allen, editor, Computer Science and Statistics: Proc. of the 17th Symposium on the Interface, pages 63-67, Amsterdam: North Holland, Elsevier Science Publisher B.V., *Grand Tour Methods: An Outline.*

[6] Martinez W.L. and Martinez A.R. 2005 London: CRC Press, *Exploratory Data Analysis with Matlab*

[7] Wegman E.J. 2003 Stat Med. May 15;22(9):1383-97,*Visual data mining*

[8] Kohonen M.T. 1989 New York: Springer, *Self-Organization and Associative Memory*

[9] Bishop, C., Svensen, M., & Williams, C. 1998 Neural Computation, 10(1), 215-234., *GTM: The Generative Topographic Mapping*

[10] M. Svensen 1998 University of Aston in Birmingham, *GTM: The Generative Topographic Mapping*

[11] W.L. Martinez, A.R. Martinez, J.L. Solka, 2010 CRC Press, *Exploratory Data Analysis with MATLAB*

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

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

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

[15] V. Vovk, A. Gammerman, G. Shafer Springer 2010 *Algorithmic Learning in a Random World*

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

Chapter 5

State of the art: techniques applied to disruption classification and prediction

5.1 Introduction

In tokamaks the disruption of a discharge can induce large forces on the surrounding structure and large heat loads on in-vessel components, especially in large devices as ITER. In this framework, being able to predict and classify disruptions would be of primary importance for improving avoidance and mitigation strategies. Physical models able to reliably recognize and predict the occurrence of disruptions are currently not available, therefore in the last decade, various machine learning techniques have been exploited as an alternative approach to disruption prediction and automatic classification.

Presently, the systems for detection of disruptions are based on more or less complex combinations of signals that, on the base of a predefined rules or thresholds, allow to take proper actions for terminating the discharge with the lowest possible risk for eventual damages on the machine. In ASDEX and in JET, for example, there is a control system in closed loop based on a threshold on the locked mode amplitude, which triggers a mitigation system (a massive gas injection valve in ASDEX). But, very often, what can be detected by these systems is unfortunately the final part of the chain of events which leads to disruption, and this is not sufficient in many case to avoid potential damages to machine structures.
In the following section, the most important techniques for disruptions prediction and classification will be reviewed.

5.2 Machine Learning

5.2.1 Main applications

Neural Networks (NNs) are one of the first techniques used in the framework of disruption prediction and classification. Neural Networks are basically an information processing system which try to resemble the way biological nervous systems, such as the brain, process information. Their structure is typically based on a large number of highly interconnected processing elements (neurons), arranged in different layers. Typically they have an input layer, an output layer and one or more hidden layers, whose number depends basically by the complexity of the specific task or application. The interconnections among neurons of different layers are called synapses and are characterized by "weights" which are updated during the learning process. The output of each neuron is computed in function of its weighted inputs through an activation function.

One of the first predictors of disruptions based on neural networks has been built for TEXT tokamak [1]. In this work the authors propose a Multi Layer Perceptron (MLP) to predict the fluctuations of the poloidal magnetic field measured through Mirnov coils, in order to identify MHD modes m = 2, which are widely recognized as important precursors of the disruptions. The proposed neural network, trained with one disruptive and a one non-disruptive pulse was able to predict a disruption in another shot 1 ms in advance. This approach has been extended with better results by adding to Mirnov coils measurements the soft X-ray signals [2]: in this case the system was able to predict some disruptions 3 ms in advance.

Always the same approach has been adopted in another tokamak, ADITYA, where in addition to Mirnov coils and soft X-ray signals, Balmer α (H_α) signals were used to increase prediction performance, extending to 8 ms the time in advance with which precursors of density limit disruptions were predicted [3].

In DIII-D tokamak instead, a three layer MLP was trained on the base of 33 input magnetic measurements, using a training set of 56 and a test set of 28 β-limit disruptions. The prediction were performed on the base of a parameter function of the normalized β_N, and the system was optimized maximizing true positive detection and minimizing false detection. About 90% of the disruptions were correctly predicted.

A NN-based disruptions predictor has been implemented also in the tokamak JT-60. Its objective was to predict disruptions caused by density limit, ramp down of the plasma current, locked modes due to low density, and β-limit. The neural network was trained with 9 input parameters, by adopting

a successive retraining procedure on the base of a stability level produced by the first procedure of training. This led to a success rate of about 97% of correct prediction 10 ms in advance, not considering disruptions due to β-limit, that didn't show clear precursors before the actual disruptions. The false detections were about the 2% [4]. A separated NN was trained appositely to predict β limit disruptions but with lower performance.

In ASDEX Upgrade NN-based methods have been widely employed for disruption prediction. In [5] a NN had the function to trigger a pellet injection system for the mitigation of the disruptions. It was trained on 99 disruptive discharges and 386 non-disruptive discharges, taking in input signals representative of the stable behavior of plasma, such as the locked mode or the q_{95}. The online system was able to correctly recognize 79% of disruptions.

In [6] a neural network predictor has been built using plasma discharges selected from two years of ASDEX Upgrade experiments. In order to test the real-time prediction capability of the system, its performance has been evaluated using discharges coming from different subsequent experimental campaigns. The large majority of selected disruptions are of the cooling edge type and typically preceded by the growth of tearing modes, degradation of the thermal confinement and enhanced plasma radiation. A very small percentage of them happen at large beta after a short precursor phase. For each discharge, seven plasma diagnostic signals have been selected from numerous signals available in real-time [7]. During the training procedure, a self-organizing map has been used to reduce the database size in order to improve the training of the neural network. Moreover, an optimization procedure has been performed to discriminate between safe and pre-disruptive phases. Such a system was able to achieve about 82% of success rate on the pulses of the same campaigns, but it deteriorated significantly when applied to subsequent campaigns.

The degrade of performance was almost entirely overcome through a retraining procedure [8]. The adaptive system contains a Self Organizing Map, which determines the 'novelty' of the input of the MLP predictor module. The answer of the MLP predictor will be inhibited whenever a novel sample is detected. Furthermore, it is possible that the predictor produces a wrong answer although it is fed with known samples. In this case, a retraining procedure will be performed to update the MLP predictor in an incremental fashion using data coming from both the novelty detection, and from wrong predictions. In particular, a new update is performed whenever a missed alarm is triggered by the predictor with which the non-recognize disruptive discharges were integrated to the training set in the adaptive procedure. The performance has been calculated on a test set of 536 safe discharges and 128

disruptive ones, giving a total prediction success rate greater than 93% with a missed alarm rate of about 13%.

Also for JET there are several experiences where NN-based predictors have been used to predict disruptions. In [9] a MLP was trained on 86 disruptive discharges and 400 discharges successfully terminated. A balanced training set was computed selecting randomly 400 samples from each safe discharge and the samples of the last 400 ms for each disruptive discharge. The most important input parameters were found to be the plasma current, the total input power, poloidal β and the internal inductance of the plasma. 84% of the disruptions belonging to the test set were correctly predicted at least 100 ms in advance.

Anyway, by testing the proposed approach with the whole pulses, the performance of the system deteriorates probably because of the fact that the reduced (for computational reasons) dataset used in the training was not representative enough of all the possible features for discriminating a non-disruptive behavior from a disruptive one. In order to overcome this inconvenient, a clustering method based on a Self Organizing Map was used to reduce more coherently the size of the training set, allowing the predictor to reach 77% of correct predictions with only 1% of false detections on a test set [10].

One of the major drawbacks of the NN approaches is that the network performance normally deteriorates when new plasma configurations are presented to the network. The ageing of a neural prediction system is unavoidable for the machines, such as JET, where new the plasma configurations are explored. Improvements might be possible using Novelty Detection (ND) techniques. In [11], both the prediction and the novelty detection tasks are performed by the same system using a Support Vector Machine (SVM). The SVM predictor shows a null percentage of false alarms, while the percentage of missed alarms is not negligible. However, using the knowledge acquired during the training phase of the predictor, the system is able to detect the novelty of new pulses increasing the performance of the entire system. In particular, the novelty detector is able to justify many of the missed alarms of the predictor as they are recognized as belonging to new regions of the operational space.

In [12] the mapping of the 7-dimensional plasma parameter space of ASDEX Upgrade (AUG) has been performed using a 2-D self-organizing map, which reveals the map potentiality in data visualization. The proposed approach allowed the definition of simple displays capable of presenting meaningful information on the actual state of the plasma, but it also suggested to use the SOM as a disruption predictor by analyzing the trajectories described over the map by the discharges under test. Various criteria have been studied

to associate the risk of disruption of each region of the map to a disruption alarm threshold. The data for this study came from AUG experiments executed between July 2002 and November 2009. The prediction performance of the proposed system has been evaluated on a test set of discharges (199 disrupted and 1070 non disrupted) different from those used for the map training, obtaining a very good prediction success rate close to 90%.

A successful experience in JET is represented by the real-time Advanced Predictor Of DISruptions (APODIS) [13]. In its most recent configuration it consists of a combination of supervised classification systems, based on SVM (Support Vector Machines) organized in two layers. The first layer contains a series of three different SVM predictors, analyzing three consecutive time windows (each 32 ms long) of data to take into account the history of the discharge. The outputs of these three evaluations are used as inputs to the second layer classifier, which takes the final decision whether or not to launch an alarm. APODIS was trained/tested with 8169 discharges (7648 safe discharges and 521 unintentional disruptions), working in open loop during the ITER-like wall campaigns of JET (2011-2012). This predictor achieved a success rate of about 98% with a false alarm rate of 0.92%. with an average warning time of 426 ms. Regarding the minimum time to perform mitigation actions in JET, which is 30 ms [14], the fraction of disruptions correctly detected 30 ms in advance has been 87.50%.

Regarding the framework of disruption classification, a first attempt of automatic classifier based on NN has been proposed in [15]. Such a classifier, based on pattern recognition techniques, was trained to discriminate among 4 classes of disruptions: mode lock, density limit/high radiated power, H-L transition and ITB plasma disruptions. The considered methods referred to clustering techniques as Self-Organizing Maps and K-means, and classification techniques such as Multi-Layer Perceptrons, Support Vector Machines, and k-Nearest Neighbours. In particular, to improve the robustness and the reliability, a Multiple Classifiers system consisting of five MLPs was implemented.

Recently, a new clustering method, based on the geodesic distance on a probabilistic manifold, has been applied to the JET disruption database for classification purposes [16]. The proposed approach allows to take into account also the error bars of the measurements and, through the nearest neighbor approach, was able to achieve a success rate of about 85% in the identification of the different types of disruptions, with no type of disruption classified with a success rate lower than 70%.

5.3 Statistical methods

Besides Neural Networks, other methods have been applied with remarkable results in the prediction and the classification of disruptions. In ASDEX, a very interesting application related to the prediction of cooling edge disruptions is described in [17]. The proposed method is based on discriminant analysis, a model-based clustering that can be used to estimate probability density functions within a supervised learning framework. In this specific application a threshold has been set to discriminate between disruptive and non disruptive pulses. The parameters which appear in the equation, being related to the causes of cooling edge disruptions, allow through their relations to find also some characteristic behaviors of the phenomenon, as for example the increasing of the internal inductance associated to a contraction of the current profile. Such a method allowed to detect 80% of cooling edge disruptions 20 ms in advance.

The data driven techniques described in [12] require a number of safe and disrupted pulses to build the predictive model. However, for ITER only a limited number of disruptions are acceptable to avoid irreversible damage to structures surrounding the plasma. A new view on disruption prediction has been proposed in [18] using Fault Detection and Isolation technique, which is a well-tested industrial technique. The prediction is based on the analysis of the residuals of an auto regressive exogenous input model of the system in Normal Operating Conditions . Hence, the disruption prediction is formalized as a fault detection problem, where the non disrupted pulses are assumed as the normal operation conditions and the disruptions are assumed as status of fault. The main advantage with respect to the literature is the fact that the model does not need disruptions to train the system but only a limited number of safe pulses. The input for the model are the time series of the radiated fraction of the total input power, the internal inductance and the poloidal beta coming from ASDEX Upgrade data between 2002 and 2009. Results are promising but lower false alarm rates are needed.

Recently another very promising application has been developed for the prediction of disruptions based on diagnostic data in the high-ß spherical torus NSTX [19], where an approach of combining multiple threshold tests has been developed on the base of the values of many signals. The starting point has been that no single signal or calculation and associated threshold value give rise to the basis for disruption prediction in NSTX. The main difficulty was related to the fact that the combination of thresholds that produce an acceptable false-positive rate have too large a missed or late-warning rate and viceversa. Therefore, an algorithm for optimizing the tuning of the multiple threshold tests has been developed allowing to achieve a false-positive

rate of 2.8%, with a late + missed warning rate of 3.7%, and thus a total failure rate of 6.5%. Such a methods has been tested on a database of about 2000 disruptions, during the plasma current flat top, collected from three run campaigns.

In JET, besides several NN applications, also other approaches have been beaten, as for example the fuzzy logic approach. The prediction of the probability of disruption was based on 12 input signals and 36 logic based rules, where both input and output signals were categorized according to a certain ranking among 3 or 5 available ones [20]. This method has the additional value to provide the possibility of transposing on the rules some basic physics related to operational limits for example, even if the optimization of the categorization of the input variables is achieved by training on a set of disruptive and non-disruptive discharges, with all the drawbacks previously discussed about the representativeness of the training set.

5.4 General comments and multi-machine approach

One of the main critical aspects of the application of these methods, NN-based and not, is represented by the need itself to require a representative training set in order to perform efficiently. Having available a representative training set means basically to have had a certain number of disruptions, but in larger devices, especially in the case of ITER, they are anything but desired events.

Another important point is represented by the tendency to deteriorate as more as we move away from the operative conditions in which the training has been performed. A possible solution could be to develop a "cross-machine" predictor which can be trained with data of certain machines allowing to extrapolate to other machines, independently on their size. In order to be able to do this, first of all the input plasma parameters must be not only well representative of the disruptive behaviour of the plasma, but in addition they must be made dimensionless.

There are already parameters which intrinsically satisfy these requirements, as q_{95} and β_N for example, and others that can be made dimensionless by dividing for a quantity with the same dimension. For example the radiated power can be divided by the input power to define a radiated power fraction parameter. For this approach to be really applicable, there should be a representative set of dimensionless plasma parameters defined in the same way in all the machine, and eventual scaling factors have to be defined to be able

to apply the systems in different machines. Furthermore, a first attempt to realize a "cross-tokamak" predictor has been described in [21], where a NN trained on a tokamak was used to predict the time to disruption of another tokamak (JET and ASDEX Upgrade). The best performance was achieved with seven dimensionless parameters in input. The results of this study are quite encouraging, even if, as it would be expected, the system performed significantly better when tested on the same machine used for the training too.

Again in this direction, the work presented in [22] described the latest developments in data-analysis tools for disruption prediction and exploration of multi-machine operational spaces. In this framework, manifold learning tools already showed in several applications their potentiality, allowing a very efficient investigation of the operational space where the relevant physics takes place, unlike most of the other approaches described in this chapter. Therefore, even if the aforementioned drawbacks keep to be valid also for manifold learning techniques, they provide the possibility to strongly improve the understanding about the underlying physics and mechanisms at the base of disruptions, and they can represent a fundamental resource for extrapolation studies in the framework of multi-machine approach.

Bibliography

[1] Hernandez J.V. *et al.* 1996 Neural network prediction of some classes of tokamak disruption Nucl. Fusion 36 1009 – 17

[2] Vannucci A. *et al* 1999 *Forecast of TEXT plasma disruptions using soft X rays as input signal in neural network* Nucl. Fusion 39 255 – 62

[3] Sengupta A. and Ranjan P. 2000 *Prediction of density limit dis- ruption boundaries from diagnostic signals using neural net- works* Nucl. Fusion 40 1993 – 2008

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

[5] Pautasso G. and Gruber O. 2003 *Study of disruptions in ASDEX Upgrade* Fusion Sci and Tech 44 716

[6] Cannas B. *et al* 2010 *An adaptive real-time disruption predictor for ASDEX Upgrade* Nucl. Fusion 50 075004

[7] Cannas, A. Fanni, G. Pautasso, G. Sias, P. Sonato and ASDEX-Upgrade Team 2009 *Criteria and algorithms for constructing reliable data bases for statistical analysis of disruptions at ASDEX-Upgrade*, Fusion Engineering and Design, vol. 84, no. 2 – 6, pp.534 – 539.

[8] Cannas B., A. Fanni, G. Pautasso, G. Sias and the ASDEX Upgrade Team 2011 *Disruption prediction with adaptive neural networks for ASDEX UPGRADE*, Fusion Engineering and Design 86 1039 – 1044.

[9] Cannas B. *et al.* 2004 *Disruptions forecasting at JET using neu- ral networks* Nucl. Fusion 44 68 – 76

[10] Cannas B. *et al.* 2007 *A prediction tool for real-time application in the disruption prediction system at JET* Nucl. Fusion 47 1559-69

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

[12] R. Aledda, B. Cannas, A. Fanni, G. Pautasso ,G. Sias and the ASDEX Upgrade Team 2012 *Mapping of the Asdex Upgrade Operational Space for Disruption Prediction*, IEEE Trans. on Plasma Science, Vol. 40, no.3, pp. $570-576$.

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

[14] P. C. de Vries, M. F. Johnson, I. Segui and JET-EFDA Contributors 2009 Nuclear Fusion 49 055011 (12 pp).

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

[16] Murari A., Boutot P., Vega J., Gelfusa M., Moreno R., Verdoolaege G., de Vries P.C. and JET-EFDA Contributors 2013 *Clustering based on the Geodesic Distance on Gaussian Manifolds for the Automatic Classification of Disruptions*, Nucl. Fusion 53 033006

[17] Zhang Y. *et al.* 2011 *Prediction of disruptions on ASDEX Up- grade using discriminant analysis* Nucl. Fusion 51 063039

[18] R. Aledda, B. Cannas, A. Fanni, G. Sias, G. Pautasso and the ASDEX Upgrade team, 2013 *Multivariate statistical models for disruption prediction at ASDEX Upgrade*, Fusion Engineering and Design, 88 $1297-1301$.

[19] Gerhardt S.P *et al.* 2013 *Detection of disruptions in the high-β spherical torus NSTX* Nucl. Fusion 53 063021

[20] Vaglisindi G. *et al.* 2008 *A disruption predictor based on fuzzy logic applied to JET database* IEEE Trans Plas Sci 36 253-62

[21] Windsor C.G. *et al.* 2005 *A cross-tokamak neural network dis- ruption predictor for the JET an ASDEX Upgrade tokamaks* Nucl. Fusion 45 337-50

[22] Murari A. *et al.* 2013 *Latest developments in data analysis tools for disruption prediction and for the exploration of multimachine operational spaces in Fusion Energy* 2012 (Proc. 24th Int.Conf. San Diego, 2012) (Vienna: IAEA) CD-ROM EX/P8-04

Chapter 6

The database for JET

6.1 Introduction

A crucial issue for analysis, exploration and mapping of high operational spaces is represented by the quality of the database in terms of reliability, and representativeness. One of the main problem in the construction of a database characterized by high dimensionality and a large amount of observations, is how to "reduce" coherently available data preserving statistical significance. Two separated databases have been built with discharges belonging to the *Carbon Wall* (CW) configuration and to the new *ITER-like Wall* (ILW) configuration. The distinction is motivated basically by the need to analyze what is changed moving from a configuration to the other one in terms of the underlying physics and operational space. This point will be addressed in the following discussing also from a statistical point of view the observed differences.

For the Carbon Wall, data comes from plasma discharges selected from JET campaigns C15 (2005) - C27 (2009), whereas ITER-like Wall (ILW) database is based on the same set of signals belonging to the campaigns C28 (2011) - C30 (2013). The aim, as it has been discussed in detail in the chapter dedicated to Manifold Learning, is to learn the possible manifold structure embedded in the data, to create some representations of the plasma parameters on low-dimensional maps, which are understandable and which preserve the essential properties owned by the original data. Therefore, proper criteria have been used to select suitable signals downloaded from JET databases in order to obtain a data set of reliable observations. Moreover, a statistical analysis has been performed to recognize the presence of outliers. Finally, data reduction, based on clustering methods, has been performed to select a limited and representative number of samples for the operational space

mapping.

6.2 JET CW database

The database is built taking into account a set of signals recorded by several diagnostics and available from JET experimental campaigns. For the selection of the signals to be considered, an analysis based on physical considerations and the availability in real time is carried out, also with reference to the plasma parameters used from various authors for disruption prediction both on JET [1] and [2]. The selected signals are representative of the behavior of both the plasma "safe" configurations, i.e. when the pulses are correctly terminated, and when a disruption occurs. Thus, the database contains both safe and disruptive pulses selected during the current flat-top, which are classified making reference mainly to the JET disruption database. Discharges for which the plasma current remained below 1MA were discarded as for disruptive events these are usually insignificant at JET.

The parameters considered to build the database are available in real time in the JET pulse file (JPF) system or can be directly calculated by other signals available in real time except the q_{95} signal. The set of considered signals is shown in table 6.1. Among all the pulses available from JET campaigns, only those belonging to the campaigns from C15 to C27 are taken into account, because, during the shutdown following the campaign C14, changes were made to in-vessel components such as divertor tiles. In the aforementioned interval, 10366 pulses are selected, including safe and disruptive shots. Only the non-intentional disruptions are taken into account. In the campaigns C15-C27, 428 non-intentional disruptions are retained, for which all the 10 signals reported in table 6.1 are available. Note that the

JPF Signal	**Acronym**	**Unit**
Plasma current	I_p	A
Poloidal beta	β_p	$a.u.$
Mode lock amplitude	LM	T
Safety factor at 95% of poloidal flux	q_{95}	$a.u.$
Total input power	P_{tot}	W
Plasma internal inductance	l_i	$a.u.$
Plasma centroid vertical position	Z_{cc}	m
Line-integrated plasma density	ne_{lid}	m^{-2}
Stored diamagnetic energy time derivative	$\mathrm{d}W_{dia}/\mathrm{d}t$	W
Total radiated power	P_{rad}	W

Table 6.1: Set of considered signals

plasma current in JET reference frame is negative. In this case we make reference to its absolute values.

According to the literature [1], in order to synchronize the signals on the same time base vector, a sampling frequency of 1kHz is chosen.
A statistical analysis is carried out in order to identify eventual anomalous signals and a not negligible number is found to be unusable because of the excessive presence of outliers or a time evolution with no physical meaning, probably due to a fault of the corresponding diagnostic during the acquisition. Such a selection has given rise to a final dataset of 243 non-intentional disruptions among all those ones available in the considered campaigns. In fact the resulting database consists of a subset of all the non intentional disruptions corresponding to specific types, whose composition will be discussed in the chapter dedicated to the automatic classification. A distinction in different types for JET has been described in the survey in [3]: it is based on a manual classification, where specific chain of events have been detected and used to classify disruptions, grouping those that follow specific paths. A more detailed picture of disruption classification will be provided again in the chapter related to the automatic classification.

By analyzing the distributions of the signal values, a proper range of variation for each signal is assumed to clean the data. These ranges are validated with the help of JET physicists.

A time instant $t_{pre-disr}$ has been defined for the disrupted discharges, which discriminates between the non-disruptive and the disruptive phase. In this discussion, $t_{pre-disr}$ is assumed equal for all the discharges, and it is set equal to 210 ms following some suggestions reported in the literature [4]. The choice of using a unique $t_{pre-disr}$ for all disruptive pulses is widely shared in the literature and in different machines [1], [2], [5]. The assessment of a specific $t_{pre-disr}$ for each disruptive discharge represents one of the most relevant issues in understanding the disruptive events. However,the relevance of the topic and the problem complexity led us to consider it a main topic of future work.

The dataset for each disruptive pulse consists of 210 points for each of the 10 signals(one sample every 1ms), in the time interval $[t_D - 210, t_D]$ms,where t_D is the time in which the disruption takes place. The main statistical parameters of the cleaned data in the time interval $[t_D - 210, t_D]$ms are reported in table 6.2.

Then, confidence limits at 1% and 99% are used for each signal through the *quantile* function.

The introduction of a confidence level is widely employed as reported in the literature [6]. Regarding the utilization of confidence level, it is very important to point out that practically all the thresholds for cleaning the

Signal	Min	Max	Mean	Median	Std
I_p	7.00E+05	3.85E+06	1.86E+06	1.87E+06	4.08E+05
β_p	3.03E-06	3.16E+00	3.17E-01	2.07E-01	3.41E-01
LM	1.00E-04	4.65E-03	5.58E-04	4.55E-04	3.97E-04
q_{95}	2.02E+00	9.66E+00	4.01E+00	3.80E+00	1.06E+00
P_{tot}	1.46E+05	4.04E+07	5.76E+06	3.48E+06	5.77E+06
l_i	3.62E-01	2.60E+00	1.09E+00	1.12E+00	1.78E-01
Z_{cc}	1.26E-03	1.17E+00	2.71E-01	2.76E-01	6.75E-02
ne_{lid}	4.02E+18	2.68E+21	1.02E+20	7.47E+19	1.73E+20
$\mathrm{d}W_{dia}/\mathrm{d}t$	-2.39E+07	1.19E+07	-1.09E+06	-7.45E+05	1.91E+06
P_{rad}	1.01E+05	1.99E+08	4.83E+06	2.76E+06	8.75E+06

Table 6.2: CW non-intentional disruptions statistics ($[t_D$ - 210, $t_D]$ms).

data are chosen with consistent margin with respect to the real limit values of the signals. The final number of disruptive samples is 38900.

6.2.1 Safe discharges data-reduction

In the considered interval of campaigns ($C15 - C27$), all the 10 signals included in table 6.1 are available for the flat-top of 10366 safe discharges. The pulses for which the plasma current is less than 1MA are discarded obtaining 9000 safe discharges. Moreover, all the pulses for which the signals to be used are not consistent, from a physical point of view or in relation to a suitable range of values, are discarded. Being each signal sampled at 1kHz, a huge amount of data are available for describing the safe operational space.

A first shot selection is performed taking into account that several shots are repeated with similar settings of the parameters. This analysis is based on various statistical parameters (mean, median, minimum, maximum and standard deviation of each signal for all the pulses), and the resulting selection is widely validated by visual inspection. Finally, 1467 safe discharges are retained, which results in more than 20M samples. Note that this number is too large to be handled by the data visualization algorithms. Furthermore, it is much larger than the number of disrupted samples, for which only the last 210 ms for each discharge are considered. For this reason,data reduction has to be performed on the safe samples in order to obtain a balanced dataset. First of all, as for disruptive shots, a data cleaning is performed discarding the outliers. Then, the k-means clustering technique [7] is employed as a base for the development of the data reduction algorithm.

Data reduction algorithm

The application of the k-means algorithm requires the normalization of data in order to maximize the effectiveness of the clustering. Here, the variables are normalized between 0 and 1. For each pulse, the samples are grouped in a fixed number of clusters. Such a number is chosen by optimizing the value of a clustering validation index (the Dunn Index [8]) for a limited number of pulses. Here, 10 clusters are used; note that by increasing such a number, no performance improvement is reached while a greater computational burden ensues. Then, in each cluster,the samples are selected in such a way to undersample the space ensuring to cover the 10-D parameter space.

Let h be the data reduction rate, N_C is the number of the safe samples in the cluster, and R is the cluster radius. The cluster is partitioned through N_C/h 10-D hyper spheres, with radius r_i centered in the cluster centroid, where

$$r_i = i\frac{R}{N_c/h} \quad \text{for} \quad i = 1, ..., N_c/h \tag{6.1}$$

For each hyper sphere, the sample closest to its surface is retained. In this case, a reduction rate $h = 70$ is set. In figure 6.1, the result obtained by the data reduction algorithm for the shot No. 66389 is visualized through a PCA projection. Figures 1(a) and (b) show the two principal components of the ten-dimensional samples before and after data reduction, respectively.

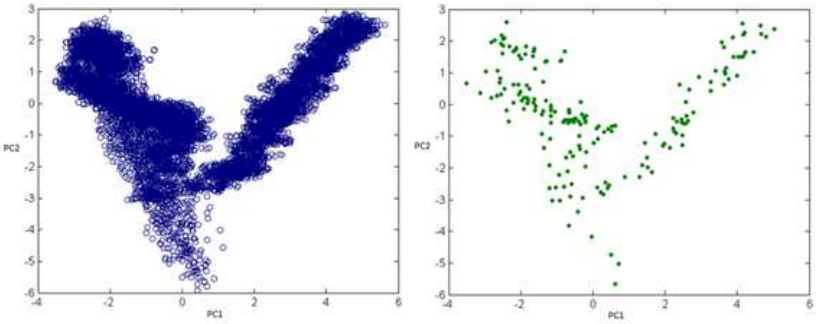


Figure 6.1: Result of the data reduction algorithm visualized through PCA (shot No. 66389): scatter plot of the first two PCs (a) of the dataset; (b) of the dataset after data reduction by k-means.

In addition to the visual investigation, for evaluating the goodness of the algorithm, the distributions of the original and reduced data sets are evaluated showing that reduced data statistically reflect the information of original data, even if with a lower density with reference to the same range of the considered variables. Figure 6.2 shows the distribution of starting data ad

reduced data for the q_{95}.

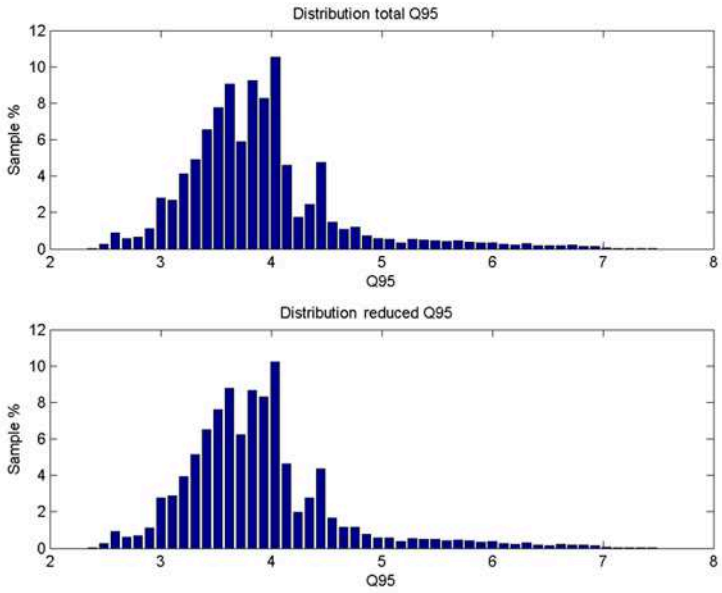


Figure 6.2: Distribution of q_{95} before and after the data reduction: the statistical distribution is preserved.

The data reduction algorithm allows one to reduce the original database from 20M to about 0.3M samples. The main statistical parameters for the cleaned data of the selected safe pulses are reported in table 6.3. Also here, confidence limits at 1% and 99% are used for each signal, leading to about 240000 samples.

6.3 JET ILW database

ITER-like Wall database is based on the same set of signals of table 6.1 belonging to the campaigns C28 (2011) - C30 (2013). Presently, regarding the safe discharges, the database is still under construction: in the considered interval of campaigns ($C28-C30$), after a initial selection, mainly on the base of the availability of all the 10 signals included in table 6.1, the resulting set of safe discharges consists of approximately 1200 discharges. All the procedures for shot selection and data reduction have to be applied yet.

Regarding the disruptive discharges, the database consists of 149 non intentional disruptions from ILW campaign ($C28-C30$), whose composition in terms of different classes is summarized in the table 6.5.

A more detailed discussion will be addressed in the chapter dedicated to

Signal	Min	Max	Mean	Median	Std
I_p	8.16E+05	3.86E+06	2.03E+06	1.96E+06	3.12E+05
β_p	6.47E-06	2.95E+00	3.10E-01	1.92E-01	2.70E-01
LM	1.00E-04	5.00E-03	2.16E-04	2.06E-04	9.18E-05
q_{95}	2.33E+00	7.49E+00	3.89E+00	3.81E+00	6.86E-01
P_{tot}	1.00E+05	3.26E+07	4.33E+06	1.42E+06	5.30E+06
l_i	5.87E-01	2.40E+00	1.06E+00	1.11E+00	1.16E-01
Z_{cc}	1.39E-01	4.71E-01	2.91E-01	3.00E-01	3.66E-02
ne_{lid}	4.01E+18	4.80E+20	6.71E+19	5.10E+19	4.35E+19
$\mathrm{d}W_{dia}/\mathrm{d}t$	-2.40E+07	2.52E+07	2.37E+04	2.36E+04	9.27E+05
P_{rad}	1.00E+05	9.96E+07	2.47E+06	6.10E+05	6.39E+06

Table 6.3: CW safe discharge statistics.

Signal	Min	Max	Mean	Median	Std
I_p	9.61E+05	3.42E+06	1.95E+06	1.92E+06	3.67E+05
β_p	3.07E-08	1.22E+00	2.99E-01	2.31E-01	1.74E-01
LM	1.00E-04	4.28E-03	4.88E-04	4.21E-04	3.38E-04
q_{95}	2.17E+00	5.89E+00	3.65E+00	3.58E+00	6.28E-01
P_{tot}	1.15E+05	2.80E+07	5.27E+06	4.26E+06	4.14E+06
l_i	3.79E-01	2.01E+00	1.08E+00	1.07E+00	1.98E-01
Z_{cc}	1.21E-03	4.23E-01	2.52E-01	2.48E-01	4.60E-02
ne_{lid}	4.02E+18	1.13E+21	1.35E+20	1.23E+19	1.06E+20
$\mathrm{d}W_{dia}/\mathrm{d}t$	-1.76E+07	3.95E+06	-1.49E+06	-9.50E+05	1.88E+06
P_{rad}	1.16E+05	1.16E+08	5.02E+06	3.67E+06	5.50E+06

Table 6.4: ILW non-intentional disruptions statistics ($[t_D$ - 210, $t_D]$ms).

ILW Disruptions			
Labels	**Classes**	**Tot**	**%**
ASD	Auxiliary Power Shut-Down	2	1.34
GWL	Greenwald Limit	0	0.00
IMC	Impurity Control Problem	109	73.15
ITB	Too Strong Internal Transport Barrier	0	0.00
LON	Low Density and Low q	7	4.70
NC	Density Contol Problem	22	14.77
NTM	Neo-classical Tearing Model	9	6.04

Table 6.5: Composition of the ILW Database in terms of different classes.

the automatic classification, nevertheless, as a general consideration which is important to highlight, in the first campaigns with the full metallic wall there were no disruptions due to too strong ITB and disruptions due to Greenwald limit.

Bibliography

[1] Rattá G A, Vega J, Murari A, Vagliasindi G, Johnson M F, de Vries P C and EFDA-JET Contributors 2010 Nucl. Fusion 50 025005-15

[2] Zhang Y, Pautasso G, Kardaun O, Tardini G, Zhang X D and the ASDEX Upgrade Team 2011 Nucl. Fusion 51 063039-41

[3] de Vries P C, Johnson M F, Alper B, Buratti P, Hender T C, Koslowski H R, Riccardo V and JET-EFDA Contributors 2011 Nucl. Fusion 51 053018-30

[4] Murari A, Vega J, Rattà G A, Vagliasindi G, Johnson M F, Hong S H and JET-EFDA Contributors 2009 Nucl. Fusion 49 055028-39

[5] Cannas B, Fanni A, Pautasso G, Sias G and Sonato P 2010 Nucl. Fusion 50 075004

[6] Cox D R, Hinkley D V 1974 Theoretical Statistics (Chapman & Hall)

[7] Lloyd S P 1982 IEEE Trans. on Information Theory 2 12937

[8] Dunn J 1974 J Cybern 495-104

Chapter 7

Mapping of JET operational space

7.1 Introduction

In this chapter the high dimensional operational space of JET with the Carbon Wall (CW) will be described and visualized using different linear projection methods such as Grand Tour (GT) and Principal Component Analysis (PCA), and mapped through non-linear manifold learning techniques as Self-Organizing Map (SOM) and Generative Topographic Map (GTM). The potentiality of manifold learning methods will be discussed showing several types of representations, also with reference to the data analysis and visualization tools developed for GTM presented in chapter 5. As integration of the analysis of the operational spaces there will be also a comparison with classical scatter plots identifying operational limits and boundaries for the considered database. All the algorithms have been described in the chapter 5.

Both SOM and GTM maps can be exploited to identify characteristic regions of the plasma scenario and for discriminating between regions with high risk of disruption and those with low risk of disruption, quantify and evaluate the effectiveness of the mapping itself. Some measures have been implemented to evaluate the performance of the proposed methodologies. In particular, the precision of the clustering over the entire dataset has been calculated through the average quantization error for both the nonlinear mappings, as well as the trustworthiness of the projected neighborhood and the preservation of the resulting neighborhood.
Moreover, an outlier analysis has been performed on the available data in order to compare how the two mapping techniques relate in terms of mapping

of observations not representative of the considered datasets.

The results show quite clearly that nonlinear manifold learning techniques are more suitable for mapping the JET high dimensional operational space, and what is really interesting is represented by the fact that the two nonlinear methods seem to converge on a manifold with similar characteristics, which means that such characteristics are strongly related to the intrinsic properties hidden in the high dimensional data.

The final dataset (all the details about construction and statistical analysis are reported in the chapter 6) which we will refer to consists of:

- 222 flat-top disruptions (38900 samples)
- 1467 safe discharges (239965 samples)

7.2 Data visualization with linear projection methods

7.2.1 Projection with Grand Tour

Grand Tour algorithm provides a multivariate visualization method generating a continuous sequence of 2-D or 3-D projections of a high dimensional data set. The animation is particularly useful for discovering eventual structures hidden in high dimensional data, allowing to look at data from practically all possible points of view.

To investigate the 10-D JET CW data, a sequence of 2-D images has been generated using Grand Tour algorithm. Figure 7.1 shows four 2-D scatter plots corresponding to different iterations of the algorithm, i.e. to different viewpoints, where blue points correspond to safe samples whereas red points correspond to disruptive samples. As can be noted, safe regions (blue) and disrupted regions (red) can be identified, even if overlaps are present.

7.2.2 Projection and mapping with PCA

Principal Component Analysis is one of the most popular and mostly used dimensionality reduction methods. The technique performs an orthogonal linear transformation of the components of the original input data in such a way that they are uncorrelated one with each other. The resulting principal components are ordered and those ones which explain most of the variance of dataset are retained. For obvious constraints of visualization, only the first two (2-D visualization) or the first three (3-D visualization) principal components can be used as new coordinate axes for providing a graphical

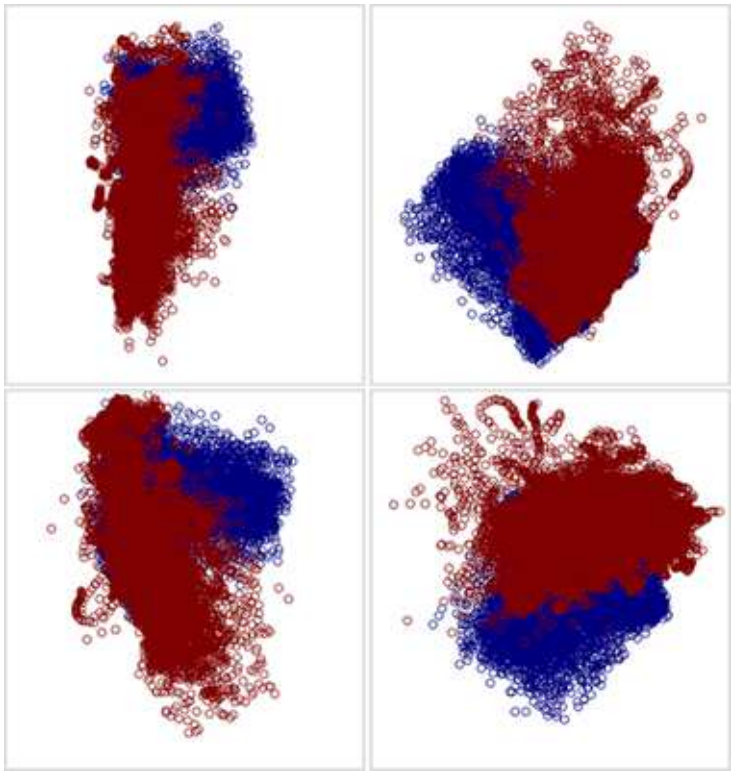


Figure 7.1: Grand Tour projections of 10-D training disruptive (red) and safe (blue) samples at different iterations.

representation of the dataset.

The Dimensionality Reduction Toolbox of Matlab, already mentioned in chapter 5, has been used. The analysis by PCA of the 10-dimensional dataset allowed among the other things to get an indication of the actual dimensionality of the considered operational space. Therefore,the variance retained by each principal component and the cumulative variance retained by a progressive number of components have been reported in Table 7.1.

JET operational space PCA		
Component	**Variance**	**Cumulative variance (%)**
1°	0.3625	36.25
2°	0.1699	53.24
3°	0.1350	66.74
4°	0.1007	76.81
5°	0.0727	84.08
6°	0.0473	88.81
7°	0.0424	93.05
8°	0.0348	96.53
9°	0.0196	98.49
10°	0.0150	100.00

Table 7.1: Variance retained by each component for JET respectively and corresponding cumulative variance.

Figure 7.2 shows the projection of the JET data onto the first two principal components. Here too, blue points correspond to safe samples whereas red points correspond to disruptive samples. On the left hand side of the figure, the safe points have been plotted before the disruptive ones, conversely, on the right hand side, the disruptive points have been plotted before the safe ones. As it can be noted, with this representation, two principal components are not enough to clearly separate the disruptive operational space from the safe one. The 10-D training samples have been also projected on the first three principal components, giving a 3-D visualization of the operational space of JET. Figure 7.3 reports the 3-D PCA projection. The visualization power of this map is higher than the previous one, even if the overlapping is still present.
In order to compare the discrimination capability of this projection method with the mapping obtained with SOM and GTM, which will be discussed in the following sections, a 2-D mapping has been realized on the base of the PCA projection with respect to the two first principal components.

The mapping is built on a regular grid of 4900 cells (comparable with respect to the number used for SOM and GTM units) in the 2-D plane, where

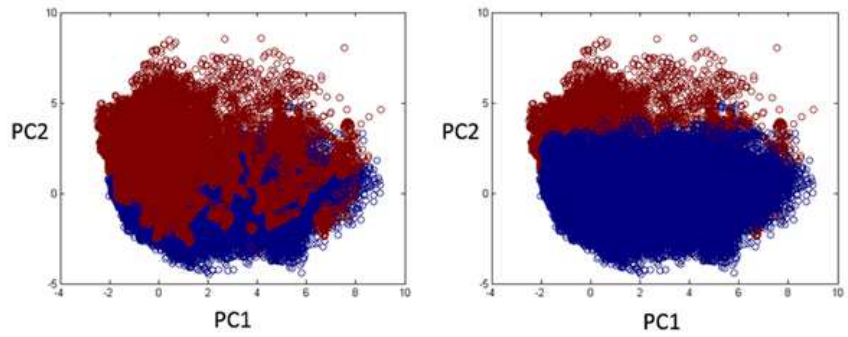


Figure 7.2: PCA projection of the 10-D training samples on the 2-D PCA; safe samples (blue), disruptive samples (red).

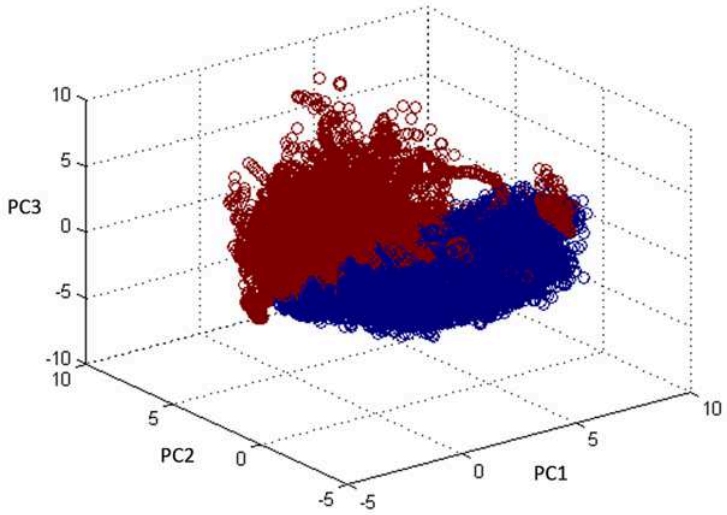


Figure 7.3: PCA projection of the 10-D training samples on the 3-D PCA; safe samples (blue), disruptive samples (red).

each cell is colored depending on its composition: blue cells contain only safe samples; red cells contain only disruptive samples; gray cells contain both safe and disruptive samples; white cells are empty (see Figure **??**(a)). As it can be noted, with this representation, two principal components are not enough to clearly separate the disruptive operational space from the safe one. In figure 7.4(b), the composition of the PCA representation in terms of samples into the cells is reported. The color code is the same as used in the previous 2-D and 3-D representations.

As can be seen, the blue (safe) cells contain 22.14% of the total samples and the red (disruptive) cells contain 3.04% of the total samples. Hence, 74.81% of the samples belong to mixed cells, which are the large majority of the total samples. Note that PCA performs a linear transformation of the input variables; in order to handle and discover nonlinear relationships between variables, nonlinear algorithms for dimensionality reduction are more effective.

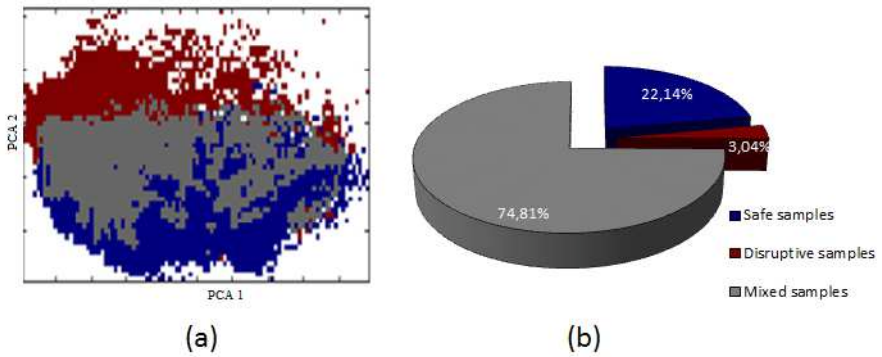


Figure 7.4: (a) PCA projection of the 10-D training samples on the 2-D PCA. Safe cells (blue), disruptive cells (red), mixed cells (gray); (b) composition of the 2-D PCA projection in terms of samples into the cells: safe cells/samples (blue), disruptive cells/samples (red), mixed cells/samples (gray).

7.3 Mapping with nonlinear methods

7.3.1 Mapping with SOM

In this section, SOM algorithm has been applied to visualize and analyze the structure of the 10-dimensional JET operational space. As the range of variation of the signals is very different, even several orders of magnitude, and since the manifold learning algorithms make use of space metrics, scaling of variables is mandatory. Hence, before projecting data, each signal in

the data base has been normalized between 0 and 1 by using the min-max normalization.
Further knowledge can be added to the intrinsic knowledge contained by the 10-D data associating a label to each sample in the data set: a safe state is associated to each non disruptive sample, whereas a disruptive state is associated to each disruptive sample.

As previously cited, the SOM is a non-linear dimensionality reduction method that produces a low-dimensional map of data by preserving their topology. The map consists of components called node or clusters. First of all, the map dimension, i.e., the number of clusters in the SOM, has to be properly selected. This has been done optimizing some performance indexes commonly used in literature to evaluate how appropriate the clustering performed by the SOM is [1]. Moreover, limiting the number of clusters preserves the generalization capability of the map. It is mandatory to choose the map dimension to maximize its capacity to discriminate among patterns with different features, keeping in the meanwhile a high generalization capability. A good tradeoff between these requirements is achieved with 4998 clusters.

The resulting map has 10 input neurons and 4998 neurons in the 2-D Kohonen layer. In this work, the SOM Toolbox 2.0 for Matlab [1] has been used to train the SOM. The safe or disruptive label associated to each sample can be used to identify four main categories of clusters in the SOM, depending on their composition: empty clusters, which contain no samples; disruptive clusters, which contain disruptive samples; safe clusters, which contain safe samples; mixed clusters, which contain both safe and disruptive samples. The same color code used in the previous section has been associated to each cluster of the map: depending on its class membership (see Figure 7.5): safe clusters are blue; disruptive clusters are red, mixed clusters are grey, and empty clusters are white. Each color, which is representative of a particular cluster composition, corresponds to a different disruption risk.

The 2-D SOM in figure 7.5(a) clearly highlights the presence of a large safe region (blue) with an associated low risk of disruption, some disruptive regions (red), with a high risk of disruption well separated from the safe region by transition and empty regions. Therefore, safe and disruptive states of plasma seem quite well separated in the SOM.
The SOM composition is reported in Figure 7.5(b) in terms of samples into the clusters. As it can be seen, safe clusters contain the 60.03% of the total samples, the disruptive region contains the 7.61% of the total samples and the transition region contains the remaining 32.36% of the samples. Note that, the 69.76% of the safe samples falls in the safe region and the 54.55% of the disruptive samples falls in the disruptive region. The remaining sam-

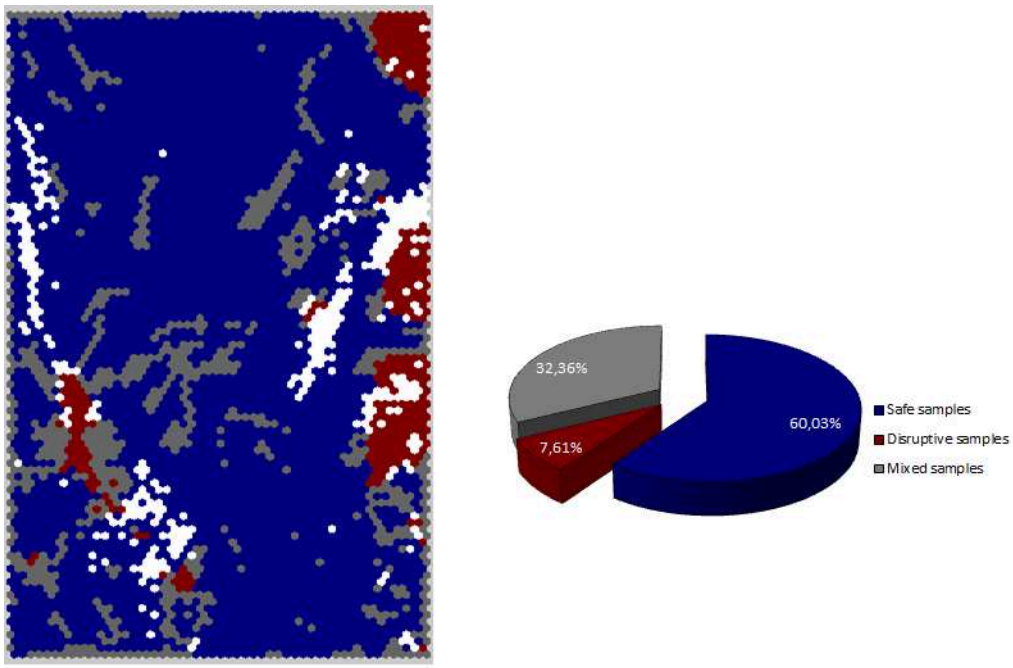


Figure 7.5: (a) 2-D SOM of the 10-D JET operational space: safe clusters (blue), disruptive clusters (red), mixed clusters (grey), empty clusters (white); (b) Composition of the SOM in terms of samples into the clusters: safe clusters/samples (blue), disruptive clusters/samples (red), mixed clusters/samples (grey), empty clusters (white).

ples identify the transition region that is mainly composed by safe samples: the 80.41% of samples in the mixed clusters are safe and the others are disruptive.

One of the causes of the presence of transition clusters is the choice of a unique value of $t_{pre-disr}$ for all the discharges. This choice is due to the lack of information on the length of the pre-disruptive phase for each shot, and can lead to incorrectly label some samples of disruptive discharges or to miss some information. Further effort can be devoted in order to reduce the transition region and better define the boundary between safe and disruptive regions. Note that the coordinates of the prototypes are known in the original multidimensional space, allowing identifying the values of plasma parameters along the boundaries between safe and disruption regions.

7.3.2 Mapping with GTM

The Generative Topographic Mapping algorithm is a probabilistic reformulation of the SOM introduced in [2]. Unlike the SOM, GTM has not been developed in the context of neural networks but in a statistic framework. As discussed in the dedicated section in chapter 5, GTM model addresses some limitations of the SOM such as the lack of a cost function, the lack of a theoretical basis for parameters, and the lack of a proof of convergence. Furthermore, in SOM hard assignments are used instead of soft ones (probabilities).
The parameters of the low-dimensional probability distribution and the smooth map are learned from the training data using the expectation-maximization (EM) algorithm [3].

The projection of the JET data onto the 2-D GTM map has been obtained using the Exploratory Data Analysis toolbox for MATLAB [4]. Also here, in order to compare GTM with SOM mapping, a regular grid of 4900 cells (comparable with respect to the number of SOM units) has been considered in the GTM plane, and the same color code has been adopted: blue cells contain only safe samples; red cells contain only disruptive samples; grey cells contain both safe and disruptive samples; white cells are empty (see figure 7.6(a)).

As in the SOM, the GTM presents a large safe region (blue), some disruptive regions (red), well separated from the safe region by transition and empty regions. In figure 7.6)(b) the composition of the GTM in terms of samples into the cells is reported.
The safe cells contain 79.95% of the total samples and the disruptive cells contain 10.27% of the total samples. Only 9.77% of samples belong to mixed cells. Note that, the 92.93% of the safe samples falls in the safe cells and the

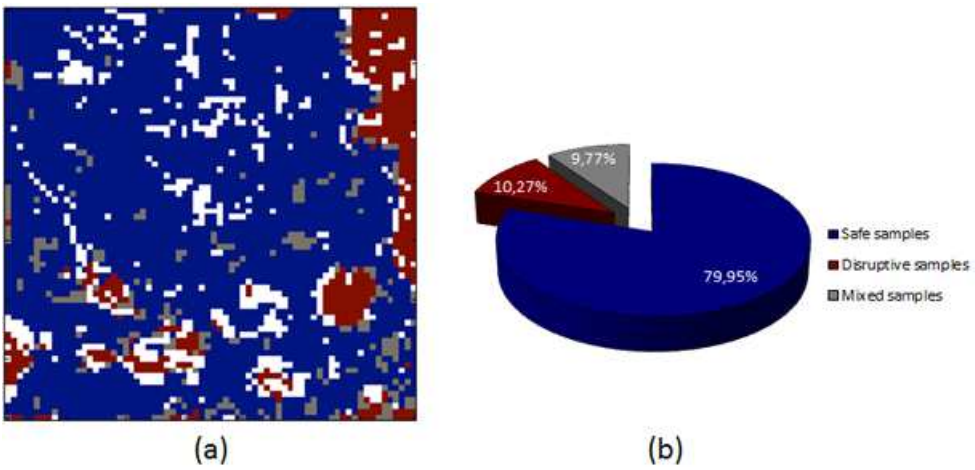


Figure 7.6: (a) GTM map of the 10 D JET operational space: safe cells (blue), disruptive cells (red), mixed cells (grey), empty cells (white); (b) Composition of the GTM map in terms of samples into the cells: safe cells/samples (blue), disruptive cells/samples (red), mixed cells/samples (grey).

73.62% of the disruptive samples falls in the disruptive cells. The remaining samples identify the transition region that is mainly composed by safe samples: the 62.35% of samples are safe and the others are disruptive. Then, the capability of the GTM to discriminate between safe and disrupted samples seems to be quite better than the SOM.

SOMs are widely used for data visualization and analysis, a lot of tools are available to explore the maps properties, and the computational complexity is limited also when managing huge amount of data, as in the problem at hand. To train the SOM only few minutes are needed by a double 6-core computer. To obtain the GTM map using the same data used for the SOM mapping, more than 1 hour of computation time was used by the same computer, and the algorithm turned out to be particularly demanding in terms of required memory. Therefore, from a computational point of view there are without doubt stronger constraints for GTM's model construction. Furthermore, there were no comparable tools available for data analysis, but part of the work carried out in the framework of this thesis has been dedicated exactly to this purpose.

Regarding the comparison of the two considered topographic maps of the JET 10-D operational space, it is very interesting to observe how clearly, by shrinking the SOM along the vertical axis, the manifold identified with the two mapping techniques look very similar (figure 7.7). Of course, the different approach in the non-linear mapping gives rise to differences, but to be

able to recover so likewise the underlying structure of the data, represents a good starting point that allows us to deal with the obtained mappings with a certain level of confidence.

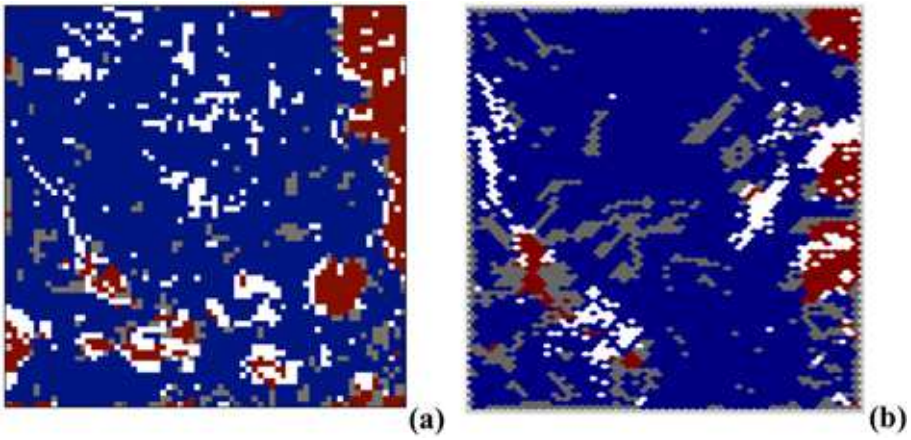


Figure 7.7: a) GTM of the 10-D JET operational space: safe units (blue), disruptive units (red), mixed units (grey), empty units (white); b) Shrunk version of the SOM in Figure 7.5(a)

7.4 Comparison with classical scatter plots

In literature, several efforts have been done to define a relationship between disruption risk and operational ranges. The most common diagrams concerning the tokamak operational ranges are related to the low-q and density limit (Hugill diagram), and to the β-limit, whose theory has already been discussed in the chapter 4.

The Hugill diagram shows the operational ranges with respect to the low-q limit and the density limit. The boundary of operation as limited by disruptions is plotted against the inverse edge safety factor $1/q_a$ and the Murakami parameter n_eR/B_t , where n_e is the line averaged plasma density (in m^{-3}). Disruptions generally restricts operation to a region $q_a > 2$ and to electron density such that $(n_eR/B_t)q_a$ is below a critical value in the range $10 \div 20 \cdot 10^{19}\mathrm{m}^{-2}\mathrm{T}^{-1}$ or higher when additional heating is applied. At JET, a critical value of $40 \cdot 10^{19}\mathrm{m}^{-2}\mathrm{T}^{-1}$, independent of the power, has been empirically found as shown in [5].

Figure 7.8(a) shows the Hugill diagram for the safe samples, whereas figure 7.8(b) shows the same Hugill diagram for the disruptive samples. Note that, the points in the diagram correspond to the safe and disruptive samples in the original space. Darker colors correspond to regions with high data

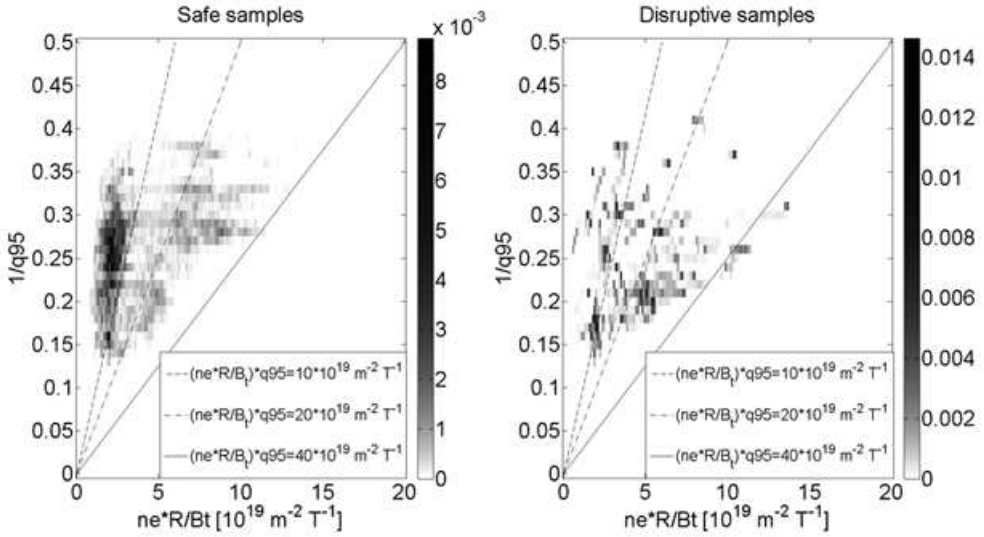


Figure 7.8: Hugill Diagram showing the operating regime for: (a) safe discharges; (b) last 210 ms of disruptive discharges.

concentration, as quantified in logarithmic scale by the color bar. An off-line signal for the line averaged density is available, even if for a limited number of the discharges considered in our data base. All the safe data lies in the region where $q_{95} > 2$. Few disruptive samples exceed the Greenwald limit, as shown in literature [6]. As it can be noted from the Hugill diagrams, several plasma configurations, leading to disruptions in less than 210 ms, are positioned in the same region of the safe discharges, confirming that a scatter plot of two parameters at a time is not suitable to distinguish between regions with high risk of disruption, and those with low risk of disruption.

Another operational boundary is the β-limit. Usually, tokamaks operate under the levels of $\beta_N = \beta_t(aB_t/I_p) = 4l_i$ where β_N is the normalized β_N and β_t is the toroidal β. Figure 7.9 (a) and (b) report data for the safe samples and for the disrupted samples respectively, in the plane $\beta_t\%$ versus $l_i I_p/aB_t$; here too, darker colors correspond to regions with high data concentration, as quantified in logarithmic scale by the color bar.

The β limit is given by the black line. As it can be noted, few samples are over the limit on these graphs. This is mainly because no real high β disruptions seem to have happened during the considered period, as shown in [5]. Moreover, the operational space is more complex, hence, it is not possible to distinguish safe and disruptive configurations looking at their position in the diagram, as highlighted also in [7].
The Hugill and the β-limit diagrams for safe and disruptive samples clearly show that it is not possible to distinguish safe and disruptive regions look-

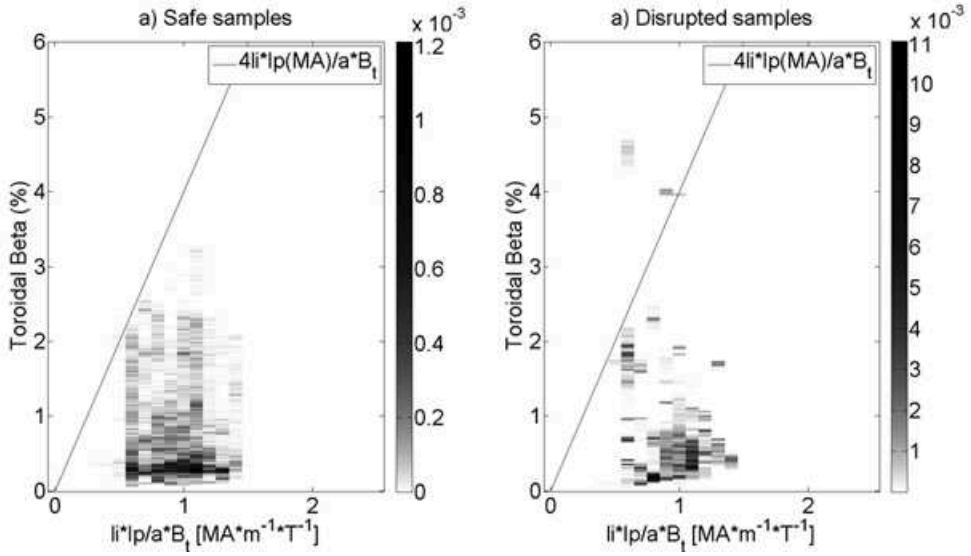


Figure 7.9: Scatter plot of the toroidal $\beta_t\%$ versus $l_i I_p/aB_t$: (a) safe discharges; (b) last 210 ms of disruptive discharges.

ing at their position in the diagrams, hence using only two parameters at a time. On the contrary, SOM and GTM maps, whose mapping project the information of a 10-D parameter space, are able to perform a better separation. These results point out the effective visualization capabilities of nonlinear data reduction methods for extracting valuable information from a large amount of high-dimensional data.

7.5 Analysis of JET operational space

7.5.1 Self Organizing Map analysis

Component plane

The Component Plane is one of the tools available to analyze the SOM results [8]. It allows a global view of the database and supports the user in detecting if there is any relation among variables through the analysis of similar patterns.
The Component Plane representation expresses the relative component distributions of the input data on the 2-D map. The dependencies among different variables can be identified by comparing the corresponding component planes: similar patterns (the colors corresponding to the values of the variables) in identical locations on the component planes are consistent with

correlated components.

In figure 7.10 the component planes for I_p, q_{95}, l_i, LM and dW_{dia}/dt are shown, together with the SOM. Note that, by picking the same cluster in each plane (in the same location), we could assemble the relative values of the plasma parameters of the cluster prototypes.

In figure 7.10 the disruptive regions have been marked with boxes A, B, and C. Figure 7.11 reports the probability density functions of the values of the prototypes of the clusters in the disrupted regions (region A: dotted line; region B: dashed line; region C: dash-dot line) and in the safe region (solid line) for the five variables considered in figure 7.10. From these functions an exact quantification of the range of the plasma parameters in the different regions of the map can be done. The analysis of figures 7.10 and 7.11 confirms well known operational limits. For example, a parameter which is very often linked with the upcoming disruption is the locked mode. Such aspect is underlined very well by the SOM. In fact, as we should expect, the disruptive regions marked with boxes A and B in figure 7.10 are characterized by high values of the locked mode signal.

Besides the considerations about the operational boundaries, what we can observe in the components plane is the presence of common patterns or regions where we have a correspondence among the distributions of different signals. For example, the disrupted region marked with the box A in figure 7.10 is characterized by high values of locked mode and internal inductance, low values of plasma current and negative values of the time derivative of the diamagnetic energy. The disruptive region marked with the box B has similar correspondences.

Moreover, it is well known that plasma current and safety factor are strongly correlated. This is confirmed by the probability density functions of the high disruption risk regions A, B, and C. Moreover, the disruption risk region C corresponds to a different operational configuration with respect to A and B. The first has high values of I_p and, as expected, low values of q_{95}, the last ones the opposite. However, although q_{95} and I_p are inversely proportional, the correlation between the two variables is not straightforward and they supply complementary information. For example, q_{95} allows one to discriminate regions A and B, whereas I_p does not (see figure 7.11). Thus we have a very complex behavior which cannot be reduced to simple correspondences or dependencies between two variables. The same information contained in figure 7.11 is reported also in table 7.2.

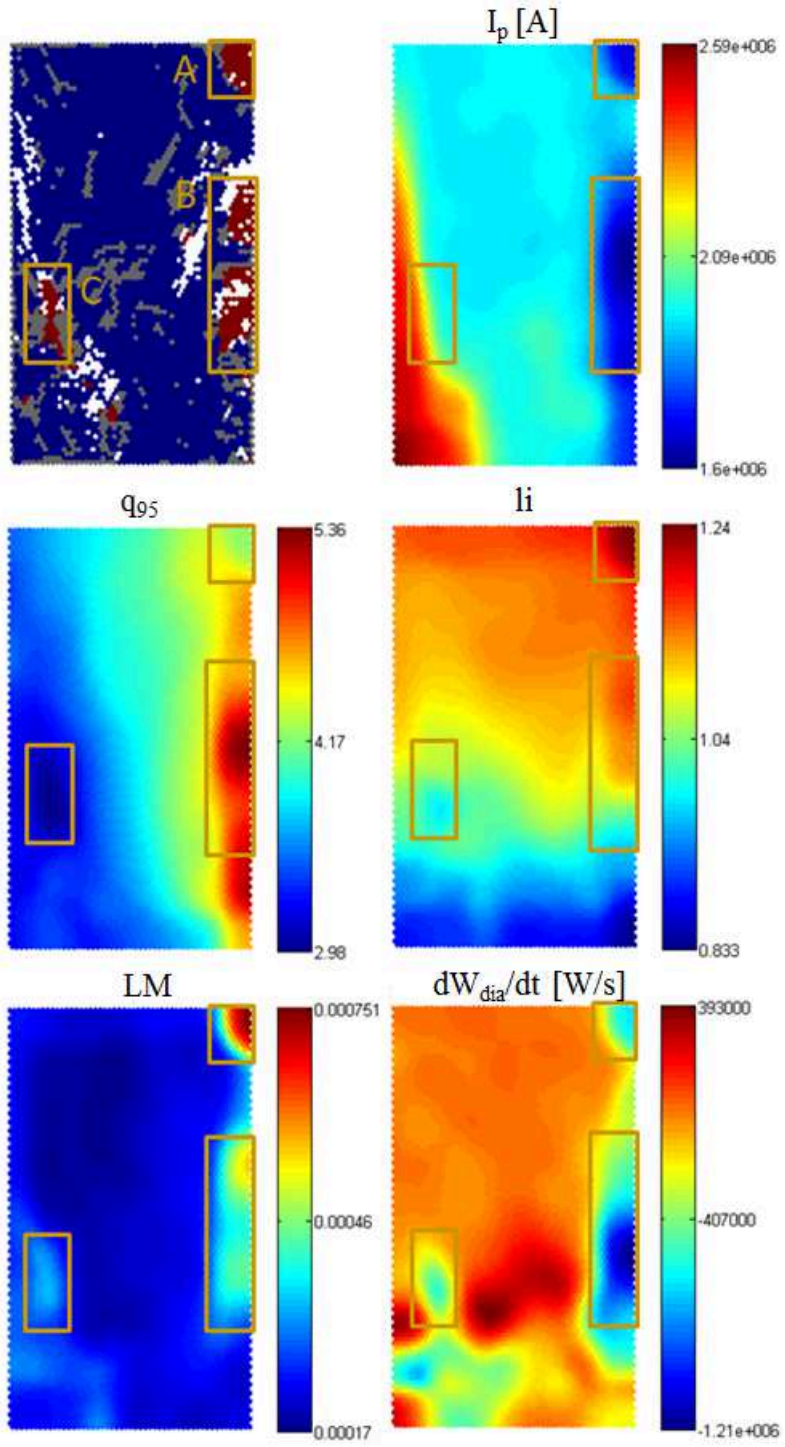


Figure 7.10: SOM and Component plane for I_p, q_{95}, l_i, LM, and dW_{dia}/dt.

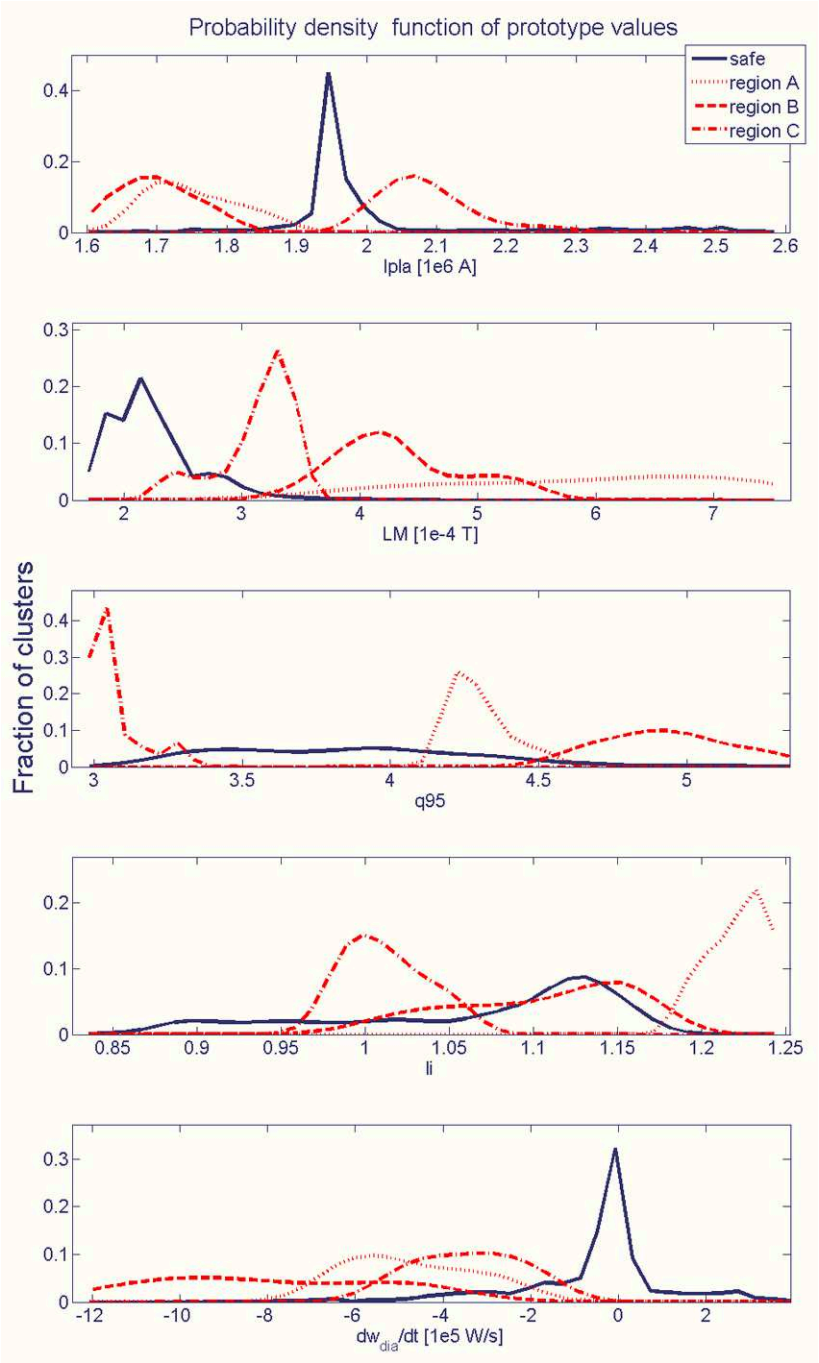


Figure 7.11: Probability density functions of the values of the prototypes of the clusters in the disrupted regions (region A: dotted line; region B: dashed line; region C: dash-dot line) and in the safe region (solid line) for I_p, q_{95}, l_i, LM, and dW_{dia}/dt.

	Region			
Signal	**Safe**	**A**	**B**	**C**
I_p(MA)	(1.9 to 2)	<2	<1.85	(1.9 to 2.3)
LM(10^{-4}T)	<3	>3	(3 to 6)	(2.1 to 3.8)
q_{95}	-	(4.1 to 4.6)	(4.5 to 6)	<3.3
l_i	-	(1.17 to 1.25)	(0.95 to 1.2)	(0.95 to 1.1)
dW_{dia}/dt	<2	(-8 to -1)	< -2	(6.5 to 0)

Table 7.2: Range of plasma parameters in safe and disruptive regions.

D-matrix

The component planes of the remaining variables P_{rad}, P_{tot}, ne_{lid}, β_p, and Z_{cc} are reported in figure 7.12, together with the D-matrix, is another type of representation available for the SOM toolbox, which visualizes the median distance between a cluster and adjacent ones.

Thus, the D-matrix allows one to display the similarity of data elements into one cluster with respect to the data into nearest ones. With this representation, it is possible to detect if there are macro-clusters of data and to judge if eventually they are well separated or not.
In figure 7.12, the D-Matrix corresponding to the SOM in figure 7.5 is shown. In the same figure 7.12, the component planes of P_{rad}, P_{tot}, ne_{lid}, β_p, and Z_{cc} are reported. Light areas in the D-matrix, where the distances between clusters are minimal, can be thought as macro-clusters and dark areas as separators. The high disruption risk regions in the top-right corner (box A) and in the right side (box B) of the SOM in figure 7.5 are well identified in the same location in figure 7.12. Other separated regions (marked with the boxes in figure 7.12) can be identified in the bottom of the D-Matrix display, which do not correspond to further high disruption risk regions. Nevertheless, the component planes of P_{rad}, P_{tot}, ne_{lid}, β_p, clearly show that these regions correspond to modifications in the operational parameters of the machine.

Moreover, the analysis of the SOM carried out through Component Planes and D-matrix highlights that only the variable Z_{cc} does not give any visually evident information in the perspective of defining the boundaries between disruptive and safe regions or distinguishing among different disruptive regions. Note that, Z_{cc} is crucial in predicting Vertical Displacement disruptions (VDDs), as demonstrated in [9]. Anyway, VDDs have not been considered because there is no a particular interest for them in the framework of disruption prediction and classification since their prediction can be done

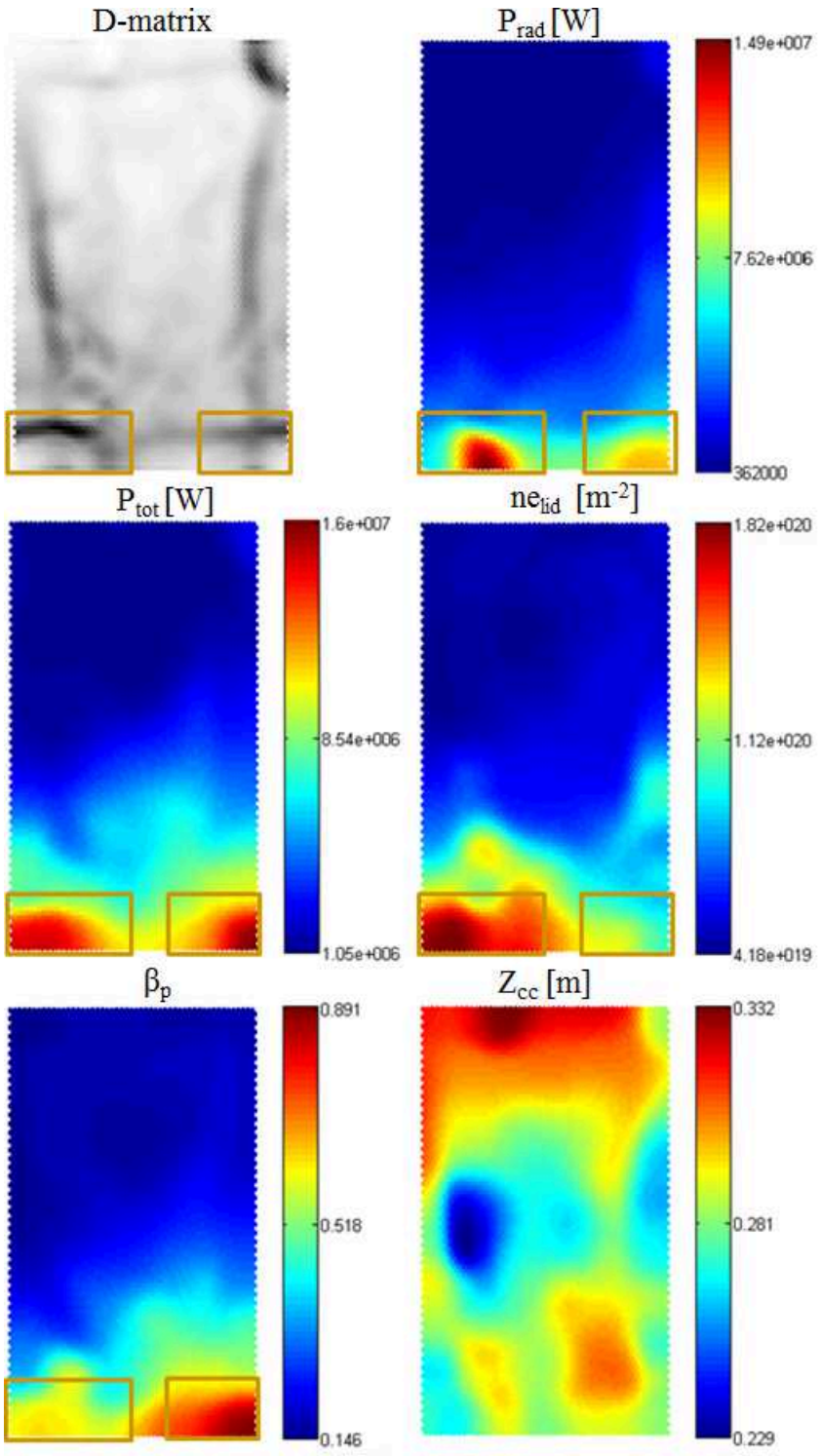


Figure 7.12: D-Matrix and Component Plane for P_{rad}, P_{tot}, ne_{lid}, β_p, and Z_{cc}.

quite easily on the base of a threshold on Z_{CC}, as shown in [9].

Discharge tracking

The potentiality of the available toolbox for the SOM suggests the possibility to track the temporal sequence of the samples on the map, depicting the movement of the operating point during a discharge. Following the trajectory in the SOM, it will be possible to eventually recognize the proximity to an operational region where the risk of an imminent disruption is high. In figure 7.13, the trajectories of a safe discharge (No. 78000) and of a disruptive discharge (No. 73851) are reported. As can be noted, the disruptive discharge (magenta trajectory) starts in a safe (blue) cluster, crosses mixed clusters, and arrives in a disruptive (red) cluster. The safe discharge (cyan trajectory) starts in a safe cluster, and evolves with the time moving into the safe region.

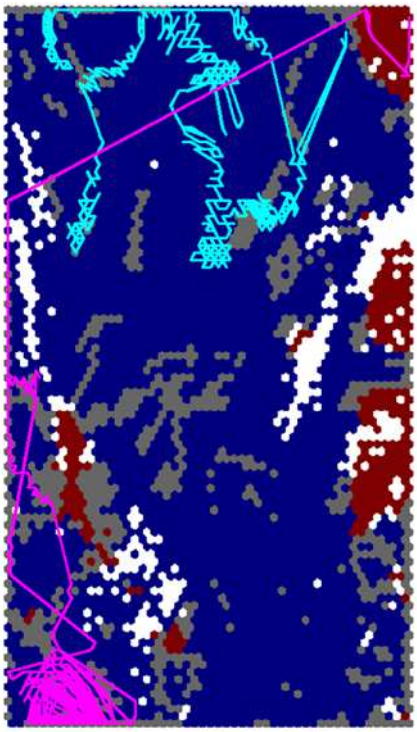


Figure 7.13: Tracks of the disruptive pulse No. 73851 (magenta) and of the safe pulse No. 78000 (cyan) on the 2-D SOM.

7.5.2 Generative Topographic Mapping analysis

Component plane

Besides the map representation, other tools analogous to those ones available for SOMs have been developed for GTM model. Among these tools, similarly to the SOMs' case, it turned out to be very interesting the analysis with the component planes. In figure 7.14, the component planes for I_p, q_95, l_i, LM, and dW_{dia}/dt are shown, together with the GTM's map. Some disrupted regions have been marked with boxes labeled from A to C.
Note that, even if the numbers of points in the latent spaces are about the same, GTM and SOM are based on a different non-linear relationship between the latent space and the data space. Although in both cases, points close to each other in the input space are mapped on the same or neighboring points in the latent space, the algorithms applied to define the mappings are different. Therefore, no direct correspondence was expected among the disruptive areas detectable on the GTM and the ones detectable on the SOM.

The analysis of the Component Planes for the GTM leads to considerations similar to those done for the SOM, even if the situation in this case is more complex. All the three disruptive regions highlighted in figure 7.14 correspond to high values of LM and negative values of dW_{dia}/dt; moreover, the region B is clearly characterized by low values of I_p, but high values of q_{95} and l_i.

Discharge tracking

Similarly to SOMs, also for GTMs a function to track the temporal sequence of the samples on the map has been developed. In figures 7.15 and 7.16, the trajectories of the same discharges projected on the SOM in figure 7.13 (No. 73851 and No. 78000) are reported.

Also in this case, the disruptive discharge starts in a safe (blue) cluster, crossing mixed clusters, and ends up in a disruptive (red) cluster. The safe discharge instead starts in a safe cluster, and evolves with the time moving within the safe region. Therefore, the considerations are basically the same of the tracking performed onto the SOM map. Furthermore, it is interesting to observe that corresponding discharges evolve approximately on the same regions in the operational space, and this reinforces the considerations about the similarity of the manifold identified by the two methods.

Finally, it is worth emphasizing that, compared to other disruption prediction approaches such as those in [10] and [11], the SOM and the GTM maps provide significant additional value. Whereas the tools in the reference paper are black boxes, which provide a prediction but are very difficult to

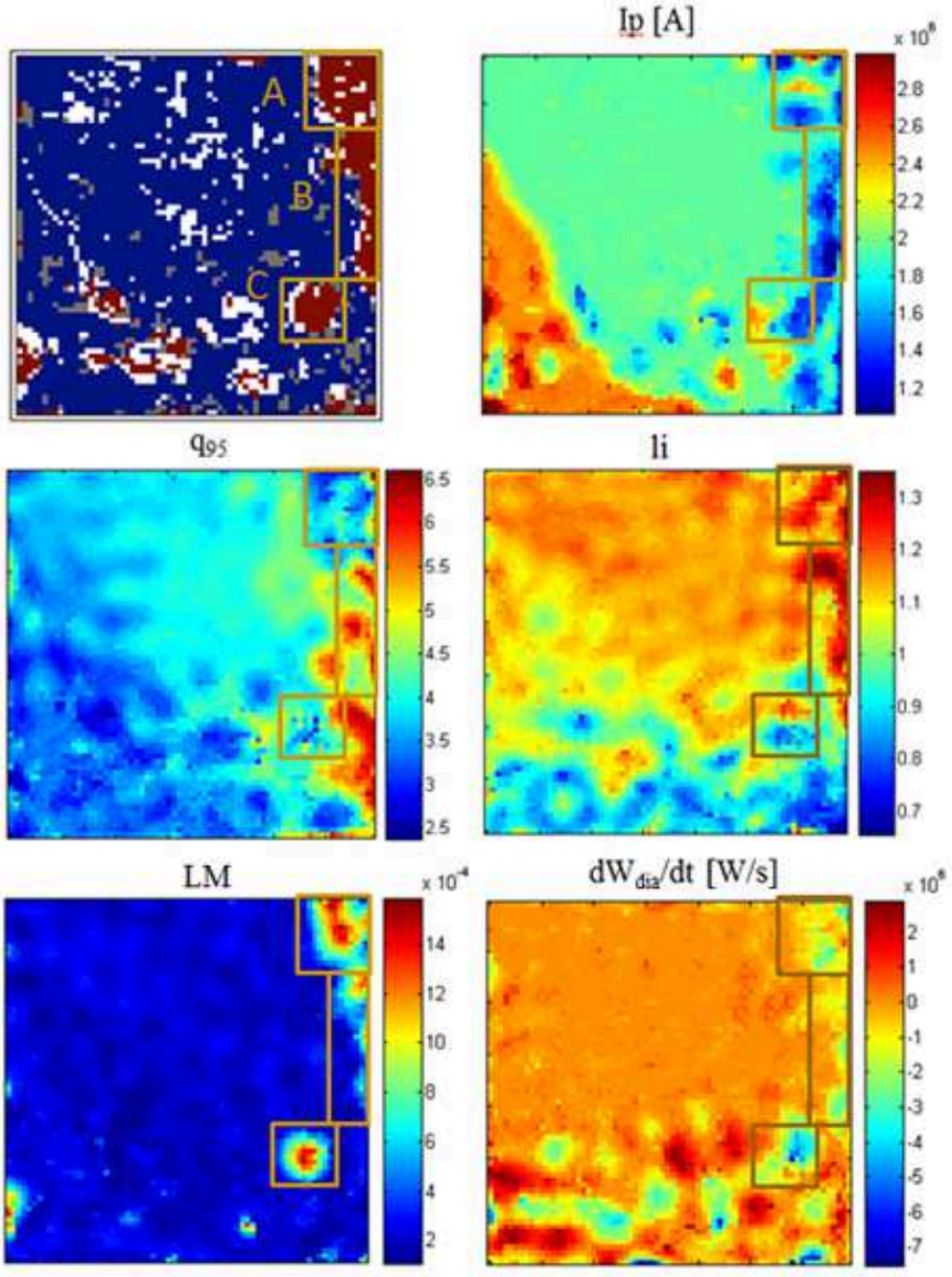


Figure 7.14: GTM and Component plane for I_p, q_{95}, l_i, LM, and dW_{dia}/dt.

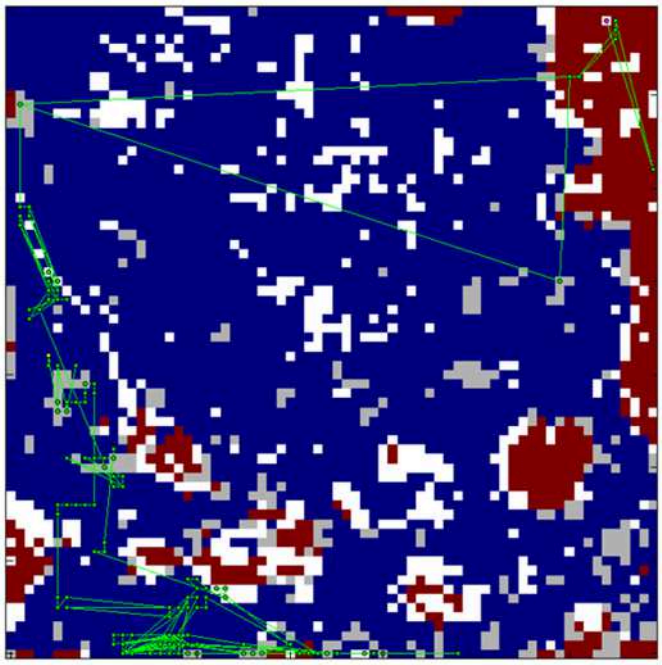


Figure 7.15: Track of the disruptive pulse No. 73851 (green) from the start of the flat-top phase (yellow dot) to the time of disruption (pink dot).

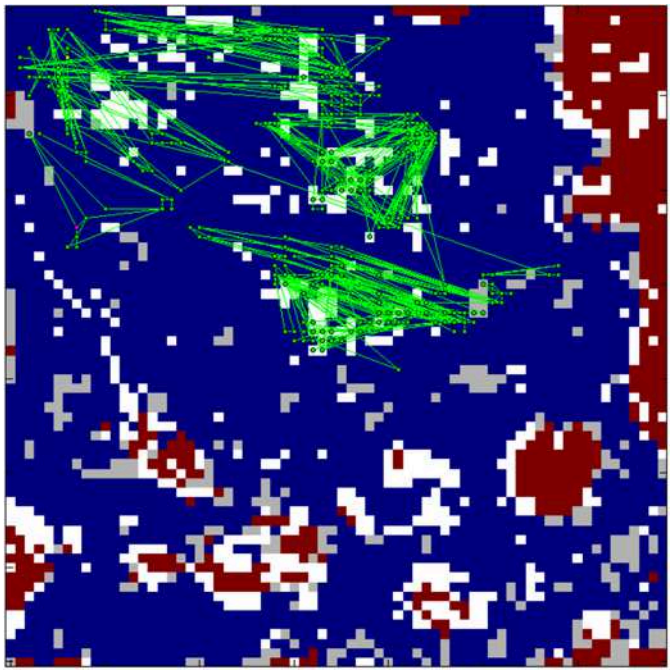


Figure 7.16: Track of the disruptive pulse No. 73851 (green) from the start (yellow dot) to the end (pink dot) of the flat-top phase.

interpret, on the contrary, the maps allow to follow the trajectory of the plasma and to study its behavior leading to a disruption. So the developed maps have the potential to provide much more than a simple prediction in the understanding of the operational space and the causes of the disruptions.

7.6 Mapping performance analysis

7.6.1 Introduction

In order to compare the mappings obtained with SOM and GTM, some measures can be used to evaluate the performance of each methodology. Special emphasis is put on the position of outliers and extreme points in the maps, and on quantization and topological errors. In particular, some novel measures such as Quantization Error measure, Trustworthiness measure, and Topology Preservation measure will be defined to provide an objective means by which the mappings can be compared. Until now, it does not appear that these methodologies have been compared in a setting in which the underlying structure of the data may not be known a-priori. Moreover, an outlier analysis has been performed on the available data in order to quantify the goodness of the projection. In fact, in order to preserve the shape of the data cloud, the outliers in the original space should be projected on the margin of the map.

7.6.2 Outliers' analysis

An outlier is an observation that numerically deviates abnormally from other values of the rest of the population it belongs to. For characterizing abnormal observations there exist different techniques, and, among the graphical ones, scatter plots and box plots are widely employed, revealing outliers' location and distance with respect to the other points of the population. In the following, the outlier analysis has been used to evaluate the goodness of the mapping. In fact, the topological shape of the data cloud in the original space is preserved during the mapping if extreme points of the data cloud are mapped to extreme units, located at the borders of the map.

The Mahalanobis distance is a measure of statistical distance in a multidimensional space. The points with the greatest Mahalanobis distance are considered outliers. Figure 7.17 reports the Mahalanobis distance for each point of the dataset with respect to the mean value of the same JET dataset. As it can be seen, the Mahalanobis distances of the safe and disruptive samples are quite different for the two macro-sets.

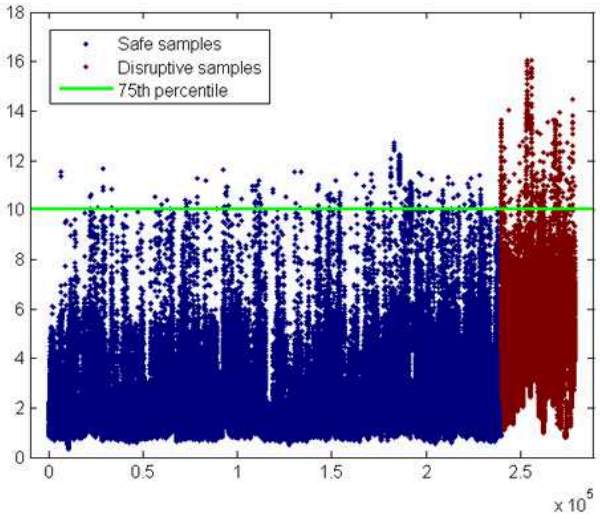


Figure 7.17: Mahalanobis distance of the safe (blue) and disruptive samples (red) with respect to the mean value of the entire JET dataset.

This is confirmed by using the box plot representation (see Figure 7.18) of the Mahalanobis distances [12]. On each box, the central mark is the median value, the edges of the box are the 25^{th} (lower quartile) and 75^{th} percentiles (upper quartile), the whiskers extend to the most extreme data points not considered outliers, whereas outliers are plotted individually. If the lower quartile is Q_1 and the upper quartile is Q_3, then the difference $(Q_3 - Q_1)$ is called the interquartile range or IQR. A data point is usually marked as outlier if it is beyond the following quantity called inner fence:

- Upper/lower inner fence: $Q_3 \pm 1.5 \cdot IQR$

Another more severe condition for identifying outliers takes into consideration a larger threshold on the previous definition that is the outer fence:

- Upper/lower outer fence: $Q_3 \pm 3 \cdot IQR$

A point beyond an outer fence is considered an extreme outlier.
In Figure 7.18, both for safe and disruptive samples, outliers (marked in red) with respect to inner fences are identified, and, as it can be seen, they are all above the upper one.

In the maps in Figure 7.19, the green map units contain samples with Mahalanobis distance greater than the upper outer fence. It can be noted that both in the GTM and the SOM, part of the identified outliers are mapped in the borders of the map, whereas the other part is mostly associated with

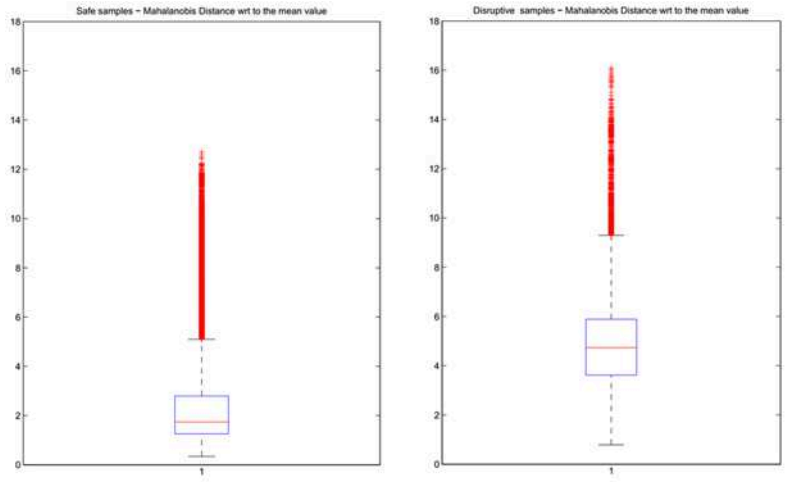


Figure 7.18: Box plot of the Mahalanobis distance for safe samples (on the left) and disruptive samples (on the right) of JET CW dataset with outliers marked with respect to the upper outer fence.

disruptive map units, well separated by transition and empty regions from the safe ones. Moreover, outliers' location in the learned manifolds is similar for the GTM and the SOM. The difference is emphasized because of the different geometric shape factor of the two maps.

7.6.3 Performance indexes

The Average quantization error Eq is a common measure used to calculate the precision of the SOM clustering over the entire dataset [13]:

$$E_q = \frac{1}{N} \sum_{j=1}^{K} \sum_{p=1}^{N_j} ||\mathbf{t_p} - \mathbf{b_j}|| \tag{7.1}$$

This error evaluates the fitting of the map to the data and it is determined by averaging the distance of each data vector $\mathbf{t_p}$ from the barycenter $\mathbf{b_j}$ of the N_j data associated to the map unit j to whom $\mathbf{t_p}$ is associated. Thus, the optimal map is expected to yield the smallest average quantization error. Partitions with a good resolution are characterized by low values of Eq.

Literature reports several error indexes to control the conservation of topology, (see [14] and the references therein). Topology preservation has, however, turned out to be quite difficult to be defined for a discrete grid. Here, the "Trustworthiness" of the projected neighborhood and the "Preservation" of the resulting neighborhood have been taken into account. Trust-

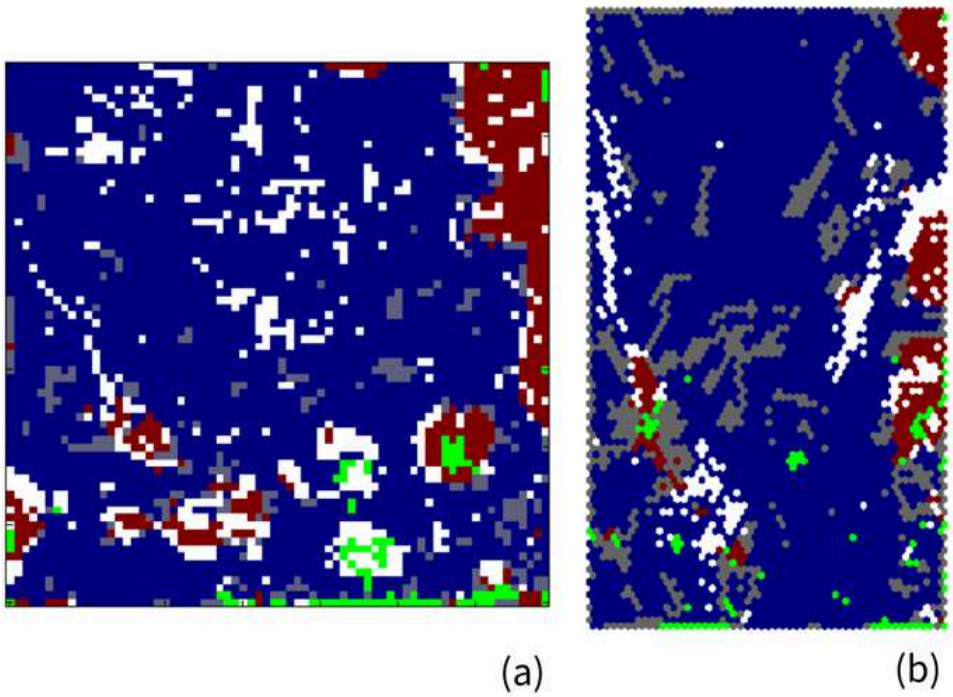


Figure 7.19: Data points with Mahalanobis distance greater than the upper outer fence (green map units) in the GTM (a) and SOM (b)

worthiness measures if data points mapped closed by on the maps are close by in the input space as well [14]. For each data point in the latent space, the set of N_j points belonging to the same map unit constitutes the considered neighborhood. The Trustworthiness of the neighborhood is quantified by measuring how far the data points belonging to the neighborhood in the latent space are from their barycenter in the original space. The distances are measured as rank orders. A measure of the error on the trustworthiness can be expressed as

$$E_{t1} = \frac{1}{K^*} \sum_{j=1}^{K^*} \frac{1}{N_j(N - N_j)} \sum_{\mathbf{t_i} \in U_{N_j}} [rank(\mathbf{t_i}, \mathbf{b_j}) - N_j] \tag{7.2}$$

where K^* is the number of no-empty map units, N_j is the neighborhood size, i.e., the number of samples associated with the j^{th} map unit, $\mathbf{b_j}$ is the barycenter of the N_j vectors in map unit j, U_{Nj} is the set of the $\mathbf{t_i}$ vectors associated with the map unit j which are not in the N_j closest to $\mathbf{b_j}$ in the original space, $rank(\mathbf{t_i}, \mathbf{b_j})$ is the position of vector $\mathbf{t_i}$ within the sorted list of increasing Euclidean distances from $\mathbf{b_j}$.

A second type of measure analyzes if the original neighborhood is preserved when data are projected. In particular, in the latent space, for the GTM all the points belonging to a certain map unit j will be characterized by the corresponding mode of the posterior probability $mode_j$, whereas in the case of the SOM they will be characterized by the corresponding prototype vector $\mathbf{x_j}$. For the GTM the error on the preservation of the original neighborhood can be expressed as

$$E_{t2(GTM)} = \frac{1}{K^*} \sum_{j=1}^{K^*} \frac{1}{N_j(N - N_j)} \sum_{\mathbf{t_i} \in V_{N_j}} [rank(mode(\mathbf{t_i}), mode_j) - N_j] \tag{7.3}$$

where V_{Nj} is the set of the $\mathbf{x_i}$ vectors among the N_j closest to $mode_j$ in the original space which are not associated with the map unit j, whereas $rank(mode(\mathbf{t_i}, mode_j))$ is the position of $mode(\mathbf{t_i})$ within the sorted list of increasing Euclidean distances from $mode_j$. For Self Organizing Maps the corresponding error is

$$E_{t2(SOM)} = \frac{1}{K^*} \sum_{j=1}^{K^*} \frac{1}{N_j(N - N_j)} \sum_{\mathbf{t_i} \in V_{N_j}} [rank(\mathbf{x}(\mathbf{t_i}), \mathbf{x_j}) - N_j] \tag{7.4}$$

where $\mathbf{x}(\mathbf{t_i})$ is the prototype vector associated with $\mathbf{t_i}$. Note that all the

points associated with the same map unit have the same rank.

The quantization errors and the errors on the trustworthiness of the projected data and on the preservation of the original neighborhood are reported in Table 7.3. GTM presents a lower quantization error, i.e., a better fitting of the map to data with respect to SOM. In this case, the map units better represent the data set. GTM is characterized by a more reliable visualization of the proximities, being the Trustworthiness error one-order lower than SOM. On the contrary, SOM has better performance on the preservation of the original neighborhood. This is not surprising because each dimensionality reduction method necessarily achieves a compromise between the last two kinds of errors.

Quality Index	**GTM**	**SOM**
E_q	0.063	0.155
E_{t1}	0.0011	0.0121
E_{t2}	0.0082	0.0016

Table 7.3: Quantization and topological errors for GTM and SOM.

Bibliography

[1] SOMtoolbox 2005 Adaptive Informatics Research Centre, Helsinky Univ. of Technology, Finland. http://www.cis.hut.fi/projects/somtoolbox

[2] Bishop C, Svensén M, Williams C 1998 Neural Computation 10 215-34

[3] Dempster AP; Laird NM, Rubin DB 1977 Journal of the Royal Statistical Society. Series B (Methodological) 39 1-38

[4] Exploratory Data Analysis Toolbox for MATLAB, `http://www.icml.uach.cl/jgarces/BIMA285/Libros/Exploratory%20Data%20Analysis%20with%20MATLAB/C3669APPb.pdf`

[5] De Vries P, Johnson M F, Segui I and JET EFDA Contributors 2009 Nucl. Fusion 49 055011-13

[6] Wars D J and Wesson L A 1992 Nucl. Fusion 32 1117-23

[7] DeVries P C, Johnson M F, Alper B, Buratti P, Hender T C, Koslowski H R, Riccardo V and JET-EFDA Contributors 2011 Nucl. Fusion 51 053018-30

[8] Vesanto J, Himberg J, Alhoniemi E and Parhankangas J 2000 SOM toolbox for Matlab5, Helsinki Univ. of Tech., `http://www.cis.hut.fi/somtoolbox/package/papers/techrep.pdf`

[9] Zhang Y, Pautasso G, Kardaun O, Tardini G, Zhang X D and the ASDEX Upgrade Team 2011 Nucl. Fusion 51 063039-41

[10] Rattá G A, Vega J, Murari A, Vagliasindi G, Johnson M F, de Vries P C and EFDA-JET Contributors 2010 Nucl. Fusion 50 025005-15

[11] Cannas B, Fanni A, Pautasso G, Sias G and Sonato P 2010 Nucl. Fusion 50 075004

[12] Bartkowiak 2004 A. *Visualizing large data by the SOM and GTM methods - what are we obtaining?* Intelligent Information Processing and Web Mining, Advances in Soft Computing 25, 399-403

[13] Kohonen MT 1989 *Self-Organization and Associative Memory*, Springer-Verlag, New York

[14] J. Venna and S. Kaski. 2007 *Nonlinear dimensionality reduction as information retrieval.* In Marina Meila and Xiaotong Shen, editors, Proc. of the 11th International Conference on Artificial Intelligence and Statistics (AISTATS 2007), San Juan, Puerto Rico, March 21-24, pp. 568-575

Chapter 8

Automatic Disruption Classification for real-time applications on JET

8.1 Introduction

In the previous chapters, it has been highlighted from several points of view how much complex the understanding of disruption physics is. The development of a physical model to reliably recognize and predict the occurrence of this hazardous plasma behaviour is presently beyond reach, due to several reasons, as in particular the amount of available signals in experiments and the nonlinear relationship between various instabilities. Therefore, in the last decade, various machine learning techniques, mainly artificial neural networks and support vector machines (SVMs), have been used as an alternative approach to disruption prediction [1]- [4]. The progress has been quite remarkable and recently a new predictor, called APODIS [4], has been very successfully deployed in JET real time network. Notwithstanding the considerable success rate, predictors such as APODIS can foresee the occurrence of a disruption but are not designed to identify its type.

On the other hand, to optimize the effectiveness of mitigation systems, it is important to predict the type of disruptive event about to occur. Indeed the best strategy to handle a disruptive plasma evolution triggered by an ITB (Internal Transport Barrier), for example, is not necessarily the same as the one to mitigate a radiative collapse. Reliable prediction of the disruption type would allow the control and the mitigation systems to optimize the strategy to land the plasma safely and reduce to a minimum the probability of damage to the device.

In [5] a survey of the disruption causes has been carried out over the last decade of JET operations. Each disruption has been manually analyzed and associated to a particular disruption class. In particular, specific chains of events have been detected and used to classify disruptions, grouping those that follow specific paths. For JET unintentional disruptions, various characteristic sequences of events have been identified. Among them, a number of clear paths could be identified that can be associated with a specific disruption class, e.g., those due to a too strong internal transport barrier and a too fast current rise, as it can be seen in Figure 8.1.

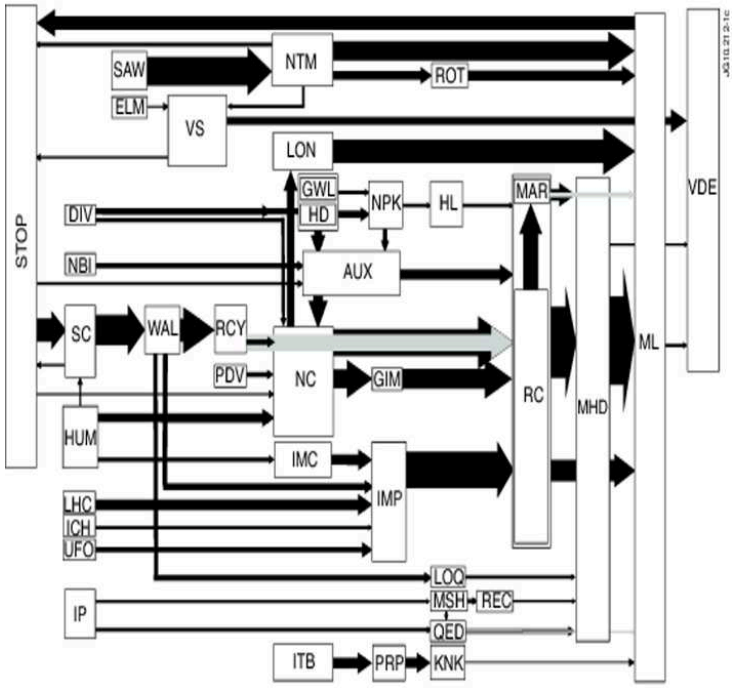


Figure 8.1: Schematic overview showing the statistics of the chain of events for non-intentional disruptions with the CW from 2000 to 2010 [5].

It should be noted, however, that the complexity of the disruption process makes this manual classification very difficult, time consuming and sometimes ambiguous. A few disruptions were not able to be classified at all [5]. Furthermore there are cases where multiple destabilizing factors acted at the same time, therefore the determination of the sequence of events between the root cause and the final disruption process turned out to be not so straightforward. Nevertheless, this basic work is essential to develop an automated classification able to help identifying a strategy for disruption avoidance or mitigation.

A first attempt to automatically classify disruptions at JET was proposed in [9] using pattern recognition techniques. Disruptions for training were manually classified by the authors, in collaboration with physicists at JET, in four classes: mode lock, density limit/high radiated power, H-mode/L-mode transition, and internal transport barrier plasma disruptions. In this chapter, an approach to automatic disruptions classification based on the nonlinear manifold learning methods described in the previous chapters, will be described, with a detailed discussion about the results.

The proposed approach for the discrimination of disruption types consists of identifying characteristic regions in the operational space where the plasma undergoes a disruption. To this purpose, given the results in the mapping of JET 10-dimensional space, SOM and GTM potentialities have been extensively investigated and an algorithm for automatic classification has been developed for both the methods.

Also in this case, the already described database of 243 non-intentional disruptions occurred in the JET CW campaigns from C15 to C27 (pulse range 63718-79853) have been considered. Each disruption is projected on the maps described in the previous chapter, and the probabilities of belonging to the different disruption classes are monitored during the time evolution, returning the class which the disrupted pulse more likely belongs to.

8.2 Automatic classification with the Carbon Wall

8.2.1 The database of disruption types

For many of the disruptive shots available on JET database, in addition to the time of the disruption, also disruption classes, which are associated to typical chain-of-events, were identified [5]. In particular, 243 disruptive discharges belonging to campaigns performed at JET from C15 (year 2005) and up to C27 (year 2009), in the range between shot number 63718 and 79853, have been considered. Table 8.1 reports the seven disruption types identified in the database, and their acronyms, reported in [5]. Moreover, in the same table, the number of shots in each class, and the percentage of occurrence in the database, are reported. This established classification is based on the macroscopic symptoms exhibited by the discharges prior to the disruption and allows comparing the results of the proposed automated clustering with the expert classification.

The plasma quantities used to automatically classify these discharges are the same ones already described in chapter 7 and also used for the mapping

CW Disruptions			
Labels	**Classes**	**Tot**	**Tot %**
ASD	Auxiliary Power Shut-Down	50	20.58
GWL	Greenwald Limit	9	3.70
IMC	Impurity Control Problem	83	34.16
ITB	Too Strong Internal Transport Barrier	10	4.12
LON	Low Density and Low q	12	4.94
NC	Density Contol Problem	58	23.87
NTM	Neo-classical Tearing Model	21	8.64
TOT		**243**	**100**

Table 8.1: Composition of the CW Database in terms of different classes.

of JET operational space. The choice of these quantities is basically due to their relation to plasma stability and their availability in real-time. The set of quantities has already been used in literature for disruption prediction purposes [1].

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

8.2.2 Analysis of the disrupted regions

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

Effective real time strategies have been developed to use the JET mapping for classification purposes. An analysis has been made to find whether the different disruption classes lie in confined regions of the map, i.e., whether the different disrupted regions of the map are associated to particular disruption classes. To this purpose, making reference to the manual disruption classification as reported in [5], a label (corresponding to the disruption types reported in Table 8.1) has been associated to each sample of a disruption.

Monitoring the evolution of each disruptive discharge on the maps, it has been found that many of them evolve within the same region. However, some regions can contain samples belonging to different disruption classes, as can be seen in figures 6 a-b, where the Auxiliary power shut-down (ASD) and Density control problem (NC) classes are represented. In particular, the

clusters in the SOM and in the GTM maps, which contain samples of ASDs 8.2(a) and NCs 8.2(b), are marked with different shades of bright red and green, in such a way to identify at the same time the class of disruption and the percentage of samples of the considered class with respect to the total number of disruptive samples. ASDs and NCs are two among the most numerous classes of disruption in the considered database. Qualitatively, it can be seen that the two classes mainly occupy different areas in the maps.
The other class with a high frequency of occurrence, both in the considered database and in the totality of non-intentional disruptions on JET, is the Impurity control problem (IMC). These last three classes are quite widespread all over the disruptive regions in the operational space, even if we can find regions where a specific class results to be predominant with respect to the others (see figure 8.2). This can be seen also making reference to figures 8.3-8.4 where the SOM and GTM maps are visualized using the so-called "pie planes". In such visualization, each node is represented by a pie chart describing the percentage composition in terms of number of samples belonging to safe and disruptive shots. The samples belonging to safe discharges are represented in blue, while the ones belonging to disruptive discharges are diversified according to the color code reported on the legend in the same figures, with reference to the different classes of disruptions. From this figure, it can be seen for example that the nodes in the regions marked with boxes relate to samples mainly coming from IMC and NC disrupted discharges.
This very heterogeneous picture could be partially due to the uncertainty of the manual classification or, more likely, to the complexity of the chain-of-events that the disruptions follow during their temporal evolution.

For example, the well-known mechanism leading to an edge cooling disruption could take place because of different reasons, such as a too high edge density or a high impurity density at the edge. In the case of density control problem (NC) and impurity control problem (IMC) disruptions, the two processes could be quite distinct even if both characterized by a high level of radiation. In particular, for a density limit disruption, radiation can be poloidally asymmetric and the instability is often linked to the stability of the divertor detachment and to the formation of MARFEs. Instead, in the case of radiative collapse by impurities, the radiation collapse is poloidally symmetric, shrinking the plasma column and increasing the plasma inductance [13], [14].

The previous considerations are confirmed by looking again at figure 8.4, which reports also the Component Planes of the internal inductance and the locked mode. The Component Plane representation, as described in the previous chapter, expresses the relative component distribution of the input data on the 2-D map. For each signal, the Component Plane associates each node

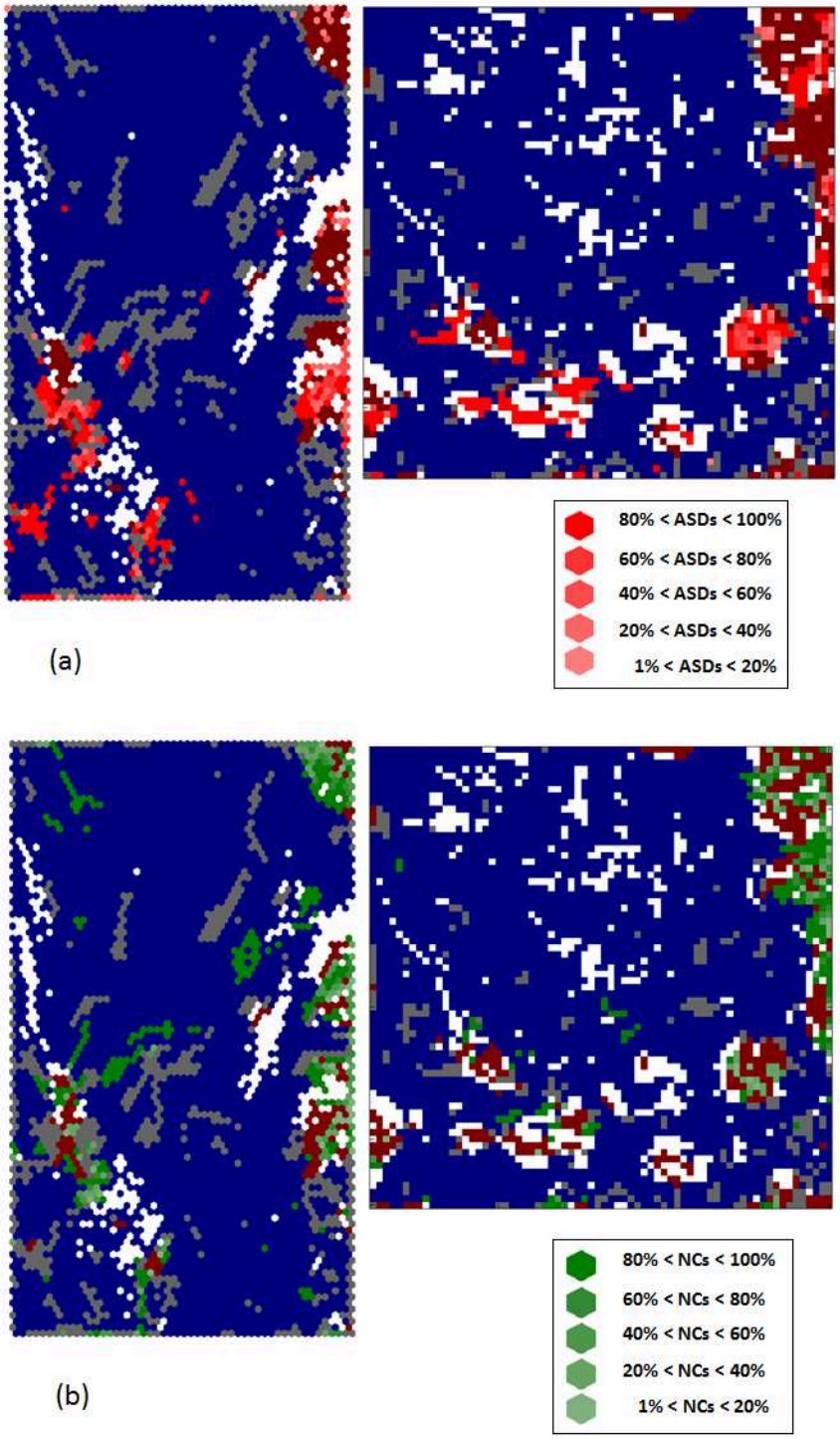


Figure 8.2: SOM (left) and GTM (right) maps coloured depending on disruption class: (a) clusters marked by shades of red contain ASD samples; (b) clusters marked by shades of green contain NC samples.

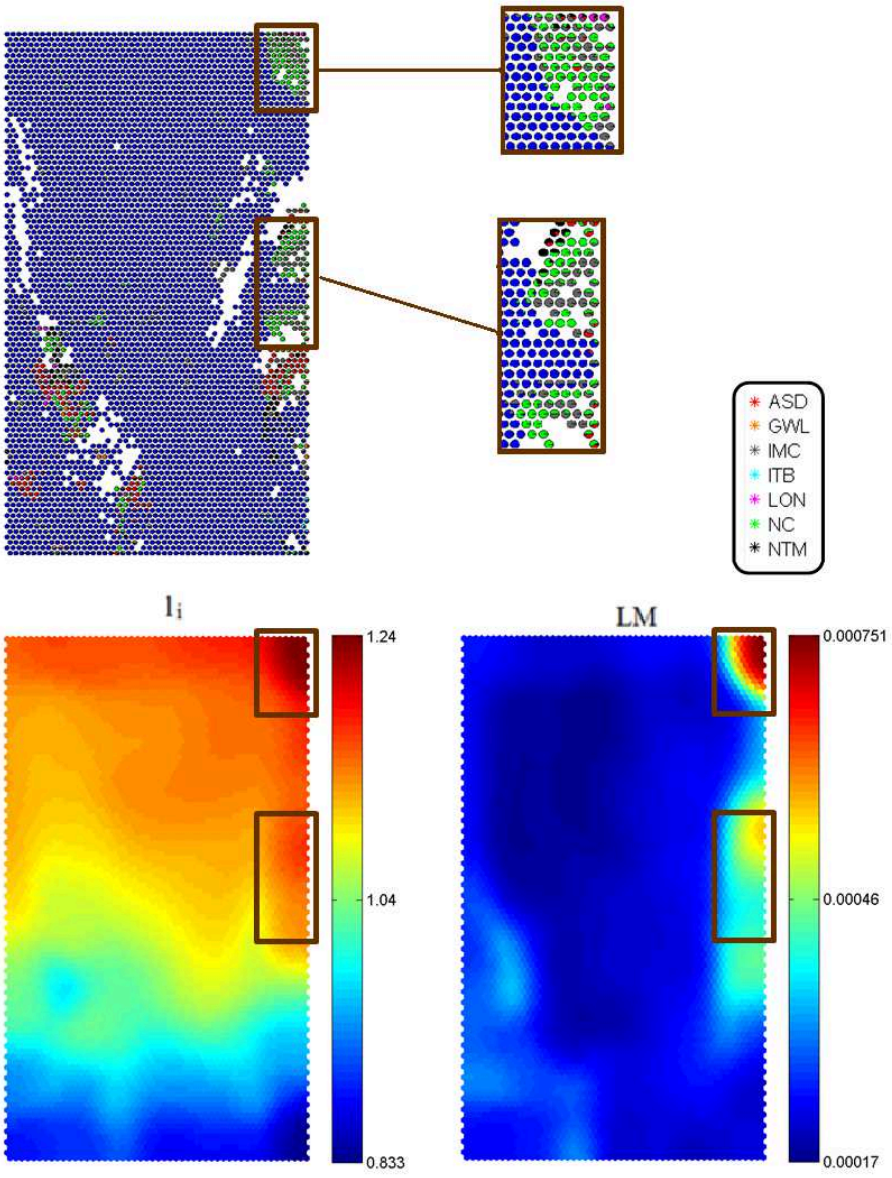


Figure 8.3: On the top: SOM map (left side) using a pie chart representation. Zoom of the regions in the boxes (right side). - On the bottom: GTM component planes of the internal inductance (left side) and the locked mode (right side).

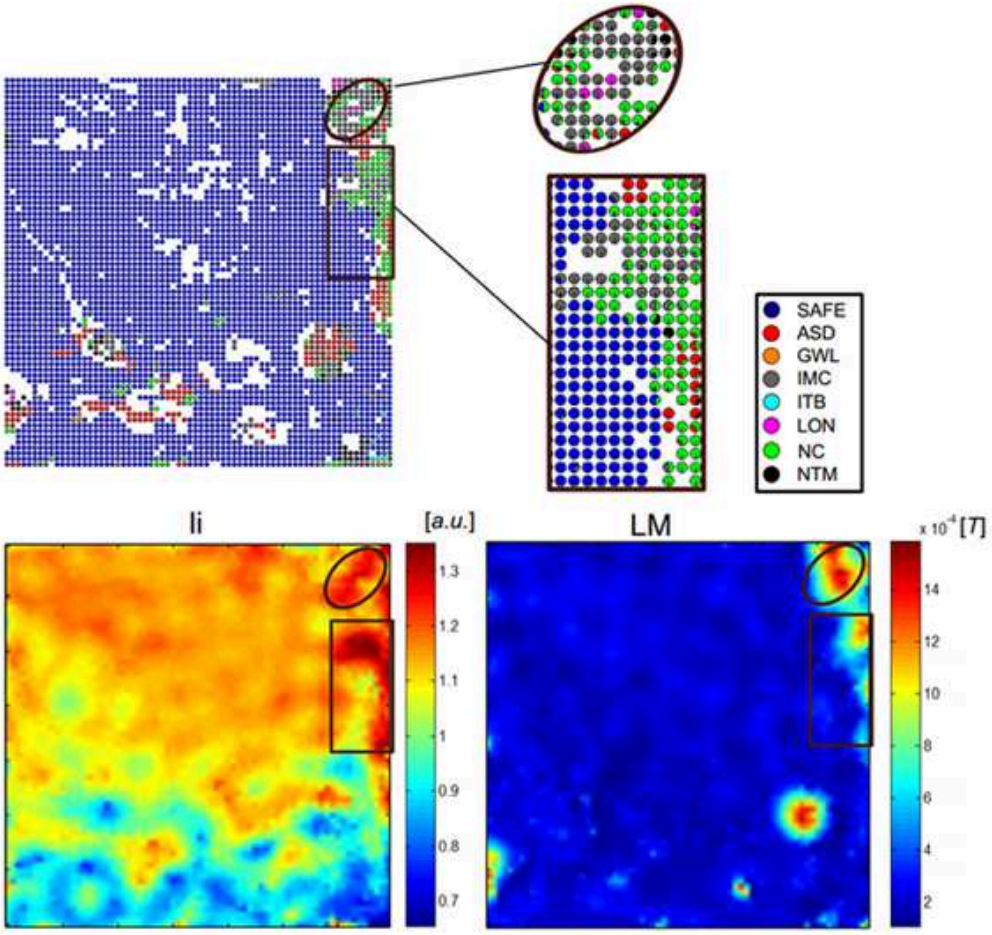


Figure 8.4: On the top: GTM map (left side) using a pie chart representation. Zoom of the regions in the boxes (right side). - On the bottom: GTM component planes of the internal inductance (left side) and the locked mode (right side).

with the mean value of the corresponding signal for all the samples belonging to such node. From these figures, it can be seen that the regions marked with boxes are mainly characterized by density control problem (NC) and impurity control problem (IMC) disruptions, and show high values of internal inductance and locked mode. The presence of both classes in the described region could be due to the connection with high radiation, even if the processes that lead to disruption are different. In this case, the signals contained in the database do not seem to allow a further distinction of the two classes.

Unlike the previously analyzed classes, disruptions due to a too strong internal transport barriers (ITB), occupy a confined region in the right bottom corner in GTM map (see Figures 8.6 and 8.5). Disruptions due to too strong internal transport barriers (ITB) constitute an important class of disruptions to which one of the shortest duration of the chain-of-events is associated. Being fast, they result to be particularly difficult to detect and typically exhibit the highest energies and heat loads. As it has been shown in [5], disruptions due to too strong internal transport barriers are characterized by the highest ratio between the plasma energy at the time of the disruption and the maximum energy during the last one second of plasma.

Plasmas characterized by internal transport barriers exhibit radially localized regions of improved confinement with steep pressure gradients in the plasma core, which in turn could drive instabilities that lead to a disruption. In relation to the achievement of continuous operation, it is well known that a large fraction of bootstrap current is necessary, and that discharges exhibiting the formation of ITBs are favourable to this aim. Experimentally, the presence of such a current fraction is usually associated with high β_p discharges with a weakly positive or negative magnetic shear in the central region of the plasma column. High values of q are probably due to the fact that advanced scenarios are typically run at q=5 and 6.

In figures 8.5 and 8.6, the nodes associated with the samples of the disruptions due to too strong internal transport barriers (ITB) are visualized respectively on the SOM and the GTM map. The different cyan shades represent different percentages of samples of the considered class with respect to the total number of disrupted samples associated with the same node. It can be seen that disruptions due to too strong internal transport barriers mainly occupy the region marked with the boxes in figures 8.5 and 8.6. These regions have also been represented by means of the Component Planes of the poloidal beta and the safety factor. As expected, disruptions due to too strong ITBs are characterized by high values of these two parameters.

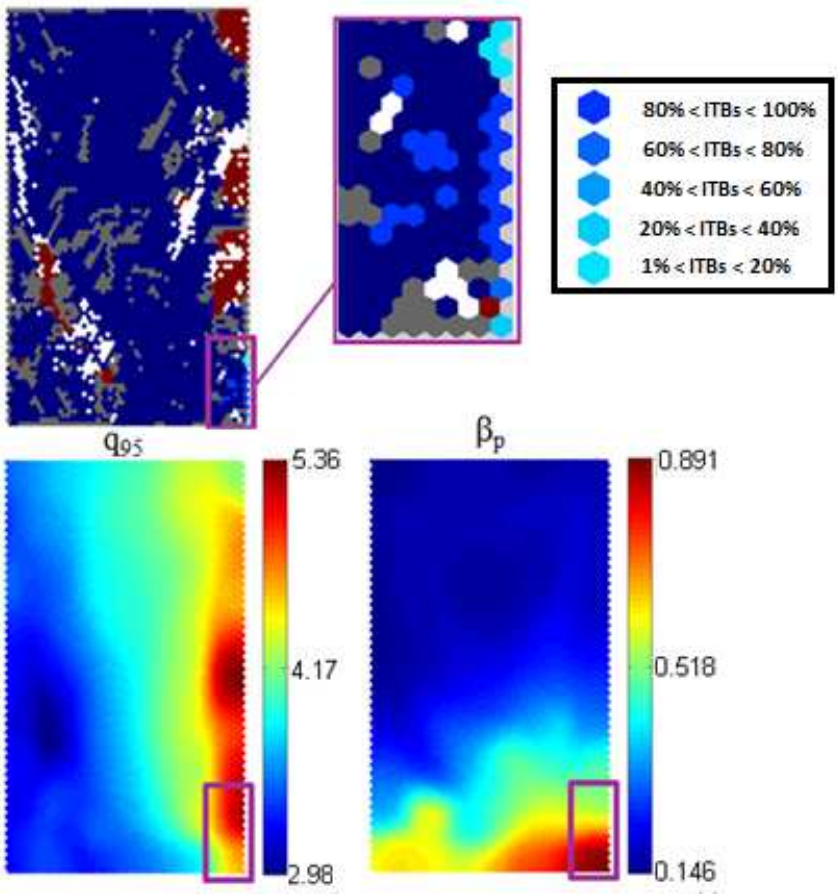


Figure 8.5: Analysis of the node composition for ITB disruptions and Component Planes of poloidal beta and safety factor in the SOM.

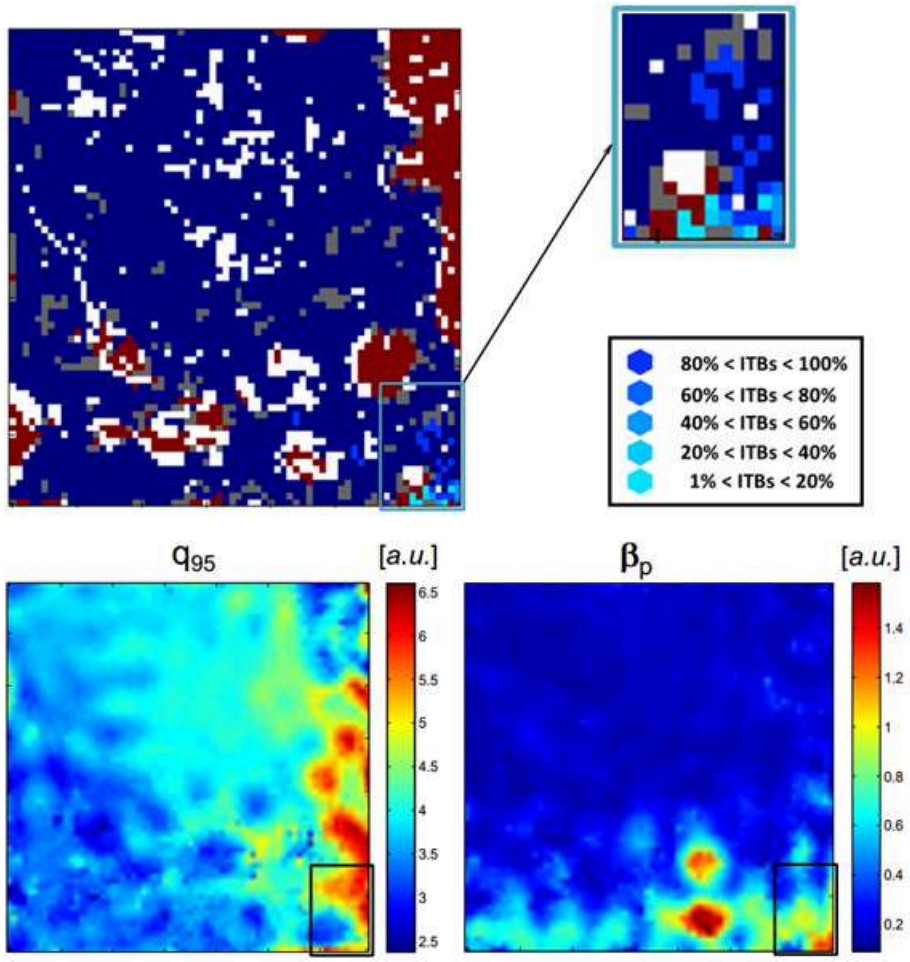


Figure 8.6: Analysis of the node composition for ITB disruptions and Component Planes of poloidal beta and safety factor in the GTM.

8.2.3 Results of the automatic disruption classification

Each node in the map (SOM or GTM) is related to samples coming from different disruption types and/or safe samples. By following on the map the temporal sequence of the samples of a disruption (the last 210 samples of the disrupted shots), each sample will be associated with a node. For each sample and each class, a class membership can be defined, being the percentage of samples of the considered class in the node to which the sample is associated, with respect to the total number of disruptive samples in the node itself. In figure 8.7 a-b the temporal evolution of the class membership of the seven classes (Class membership function) during the JET discharge No.66313 is reported.

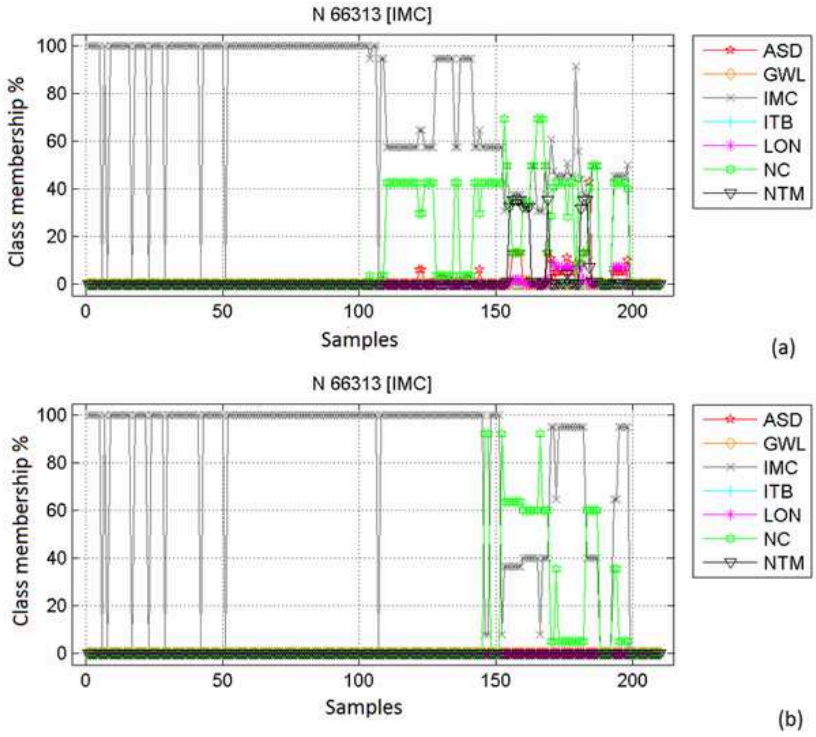


Figure 8.7: Class membership functions for disruption No.66313; (a) SOM (b) GTM

As it can be noted, for the majority of samples of this shot, the greatest class membership value corresponds to impurity control problem disruption (IMC), which is the same class assigned to this shot in [5]. This is true for a relatively long interval before the disruption time, especially in the case of the GTM. Note that, during this pulse, excessive Neon is introduced into

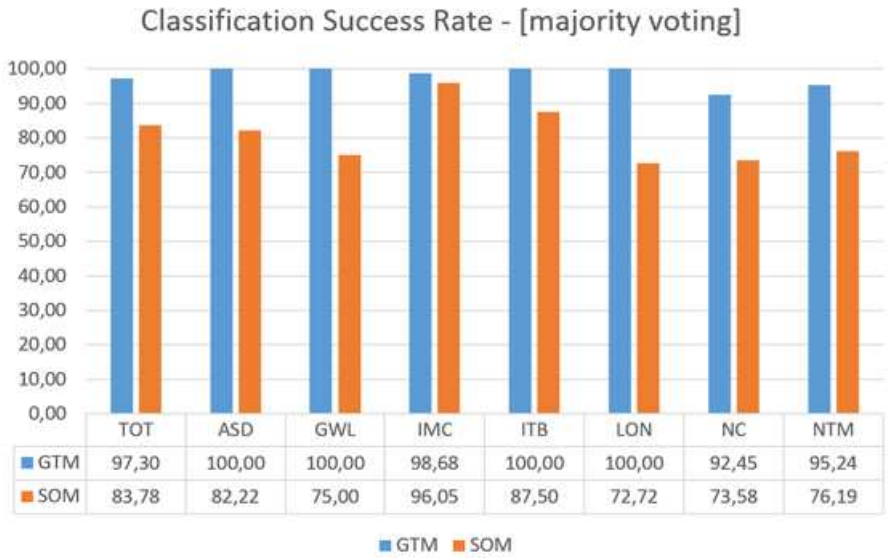
the plasma during a phase with no auxiliary heating, resulting in a radiative collapse. As Neon also increases the density significantly, this could justify the presence of relatively high NC class membership values close to the disruption in figure 8.7.

The previous analysis shows the power and the versatility of the proposed technique; the different classes of disruption tend to aggregate according to the self-organization of the map in such a way that each class results to be predominant with respect to the others in particular regions of the operational space. In order to perform an automatic disruption classification using topographic maps output and to quantify its effectiveness, a proper classification, criterion has been introduced. In particular, to classify a disruptive shot a majority voting algorithm has been adopted based on the class membership of each class in the whole time interval before the disruption (210 ms).

In the histograms in figure 8.8, the results obtained by applying the majority voting to all the 222 disrupted pulses, are reported in terms of percentage success rate. A pulse has been considered correctly classified if the automatic system produces the same classification given in [5]. As it can be noted, the success rate of GTM is higher than that of SOM for all the considered classes, reaching in some cases even the percentage of 100%. These results show the higher discrimination capability of the GTM model with respect to the considered classification, that is exactly what has been found in terms of separation between safe and disruptive regions in the mapping of JET operational space.

8.2.4 Discussion of the results

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



	TOT	ASD	GWL	IMC	ITB	LON	NC	NTM
GTM	97,30	100,00	100,00	98,68	100,00	100,00	92,45	95,24
SOM	83,78	82,22	75,00	96,05	87,50	72,72	73,58	76,19

Figure 8.8: Comparison between the percentage success rates of the off-line automatic classification performed by GTM and SOM.

Hence, this limits the discrimination capability of the classification system when approaching to the disruption time.

For example, in our classification, a not uncommon phenomenon is observed for density control problem (NC) and impurity control problem (IMC) disruptions, which initially evolve in a region where they could be correctly classified with a high level of confidence, and then evolve in the auxiliary shut down (ASD) class when approaching the disruption time. In figure 8.9 the class membership functions for the shot No. 67322, manually classified in [5] as NC, are shown. As it can be seen, the discharge initially evolves in nodes where all the samples belong to the NC class, while, as it approaches to the disruption time, it moves towards nodes where the majority of the samples belong to ASD class. Note that, an ASD disruption is basically a density control problem during/after the switch of the heating system. Hence, at first glance, the traces of an ASD and of the considered disruption would follow very similar paths.

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

In JET several experiments have been carried out for the beta limit assessment, varying the pressure and the q profiles, ranging from Hybrid to

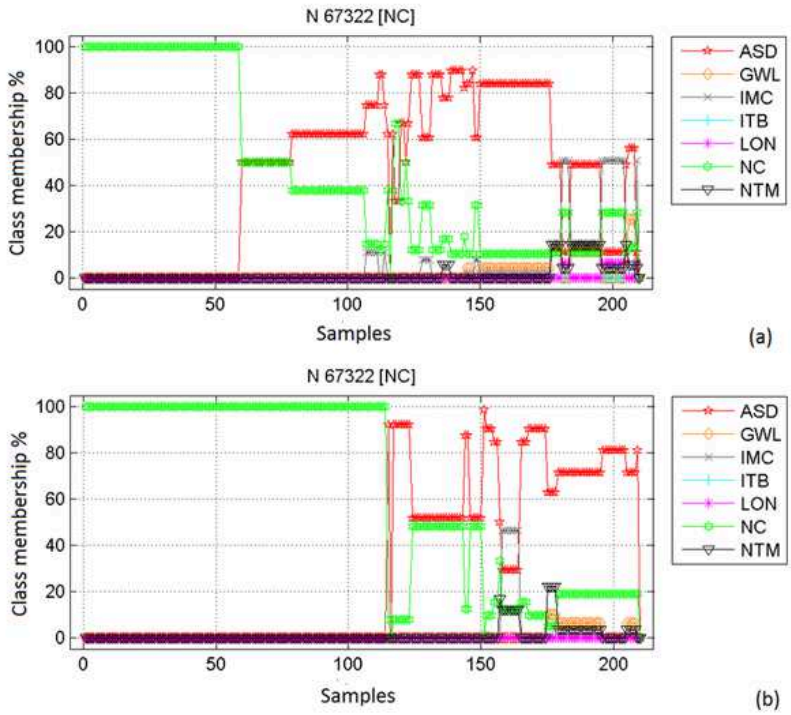


Figure 8.9: Class membership functions for disruption No.67322; (a) SOM (b) GTM.

ITB configuration, in order to investigate advanced scenarios. During these experiments, among the other instabilities, NTMs with m/n=2/1, 3/2, 4/3 have been observed [15]. In these conditions, the value of q_{95} is about 4-5, while the currents assume values in a range slightly above 1MA.

Figure 8.10a shows the nodes interested by the discharge No. 72670, manually classified as NTM. Such a discharge evolves in the right bottom corner of the GTM. As shown in figure 8.4 (see pie planes representations), the majority of disruptions in this area is due to NTMs and too strong ITBs. Figure 8.10b shows the class membership during the time evolution of the shot. As predicted by the previous considerations, NTM and ITB classes are characterized by a comparable value of the class membership functions. In this case, the adopted criterion allows one to correctly classify the disruption as NTM, even if with a reduced level of confidence. As it can be noted by the component planes for β_p, P_{tot}, and I_p, the right bottom covering both the maps is characterized by high values of poloidal beta (figure 8.6) and total input power (figure 8.10c), as well as low values of plasma current (figure 8.10d). Discrimination capability could be improved by considering further information, e.g., for example the measurement of pressure profile peaking, and this is true for all the classes in general. But many information, in particular those one related to several profiles, are not always available in real time, or need to be processed a posteriori.
Two other very interesting cases, are represented by the shot No.79772, manually classified as NC, and the shot No.79770 manually classified as LON [5]. In figure 8.11a and 8.11b the class membership functions returned by the GTM are reported. Such discharges were performed for investigating the physics of ELM control with magnetic perturbation fields (EFCC). Our system classified the pulse No. 79770 as LON, as in [5], whereas, regarding the pulse No. 79772, classified as NC in [5], the GTM recognizes the presence of the NC class for the whole considered time interval, but the highest class membership is associated with the Low density and low q problem disruption (LON) class. The pulse in the final phase is indeed characterized by low values of the edge safety factor and disrupted at $q_{95} \simeq 2.5$. Actually, during this pulse, when the NBI was switched off, erroneous density control gave rise to too fast a density decrease, leading to too low a density and an error field locked mode.

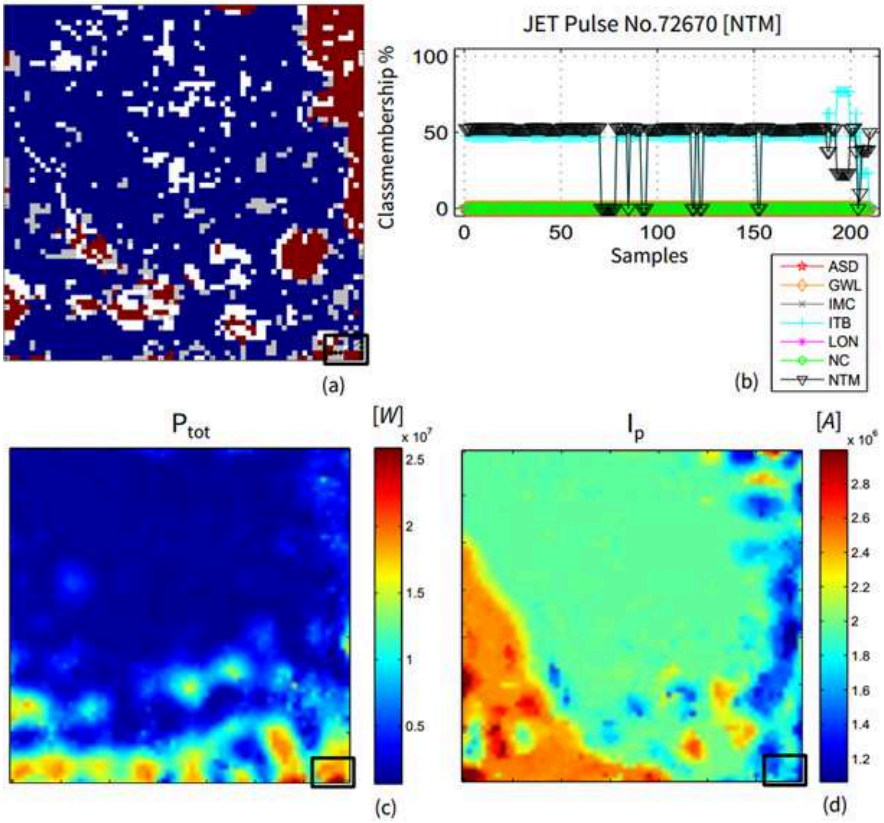


Figure 8.10: (a) clusters (black box) occupied in the GTM by the disruption No. 72670; (b) Class membership functions.(c-d): component planes of the total input power (left) and the plasma current (right).

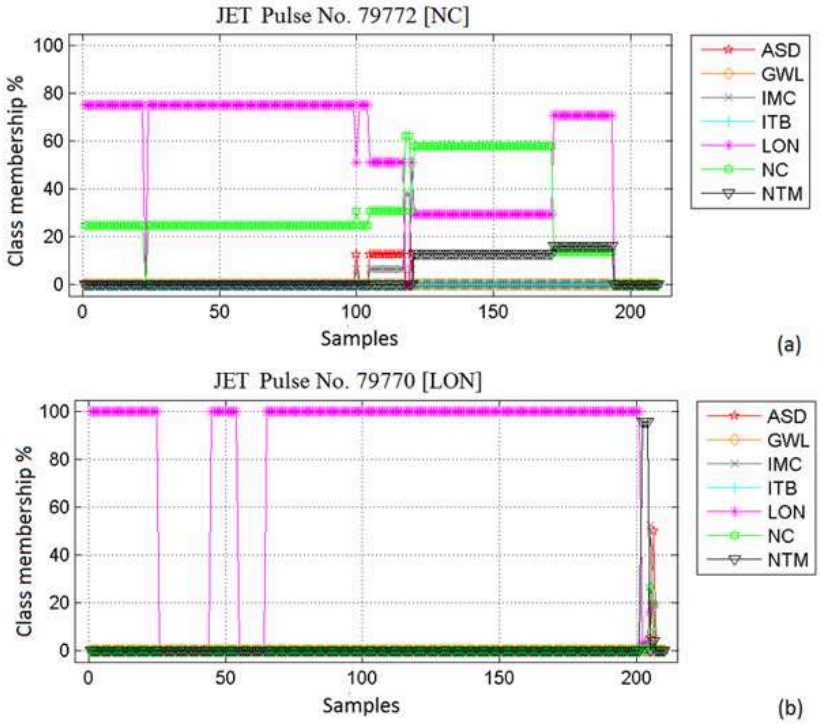


Figure 8.11: Class membership functions for disruptions No. 79772(a) and No. 79770 (b).

8.3 Automatic classification with the ITER-like Wall

8.3.1 Introduction

Also the 10-dimensional operational space of JET with ITER Like wall has been explored using the Generative Topographic Mapping method. A new 2-dimensional map has been exploited to develop automatic disruption classifications of 7 classes manually identified in [5] and described in the previous section. In particular, among all the non-intentional disruptions, the subset indicated in the Table 8.2 has been considered, that occurred in the JET campaigns from C28 (2011) to C30 (2013) after the installation of the new ITER Like Wall (ILW).

A statistical analysis has been performed on the plasma parameters describing the operational spaces of JET with CW and ILW and some physical considerations have been made on the difference of these two operational spaces and the disruption classes which can be identified.

The performance of the ILW GTM classifier is tested in a real time fashion in conjunction with a disruption predictor presently operating at JET. Moreover, to validate and analyze the results, another reference classifier has been developed, based on the k-Nearest Neighbor technique.

Finally, in order to verify the reliability of the performed classification, a conformal predictor has been developed which is based on non-conformity measures.

8.3.2 Impact of the ITER-like Wall on disruptions

In the $2010 - 2011$ all the plasma-facing components (PFCs) have been replaced with a metallic wall, the so-called ITER-like Wall (ILW). The new wall is composed of beryllium tiles in the main chamber and tungsten in the divertor. The change of materials has significantly modified the physics of disruptions. ILW have caused first of all an enhancing of heat loads and vessel forces, and this is due basically to the lower fraction of energy that is radiated during the disruption process. Consequently, in fact, a larger fraction of thermal and magnetic energy is "available" to be conducted to the even more fragile PFCs. The lower fraction of radiated energy gives rise to higher temperatures during the post-thermal quench phase, which means longer current quench times, since this latter is inversely proportional to the plasma resistivity [6].

The disruption rate is increased, especially in the first period of operations after the installation of the new metallic wall, since the properties of the

new materials affected not only the physics of disruptions, but also control schemes and operational scenarios turned out to be different with respect to what we had for the CW.

Also operational limits, such as the density limit, have been affected. In fact, as described in chapter 4, this limit is determined mainly by the radiation instability which build up at the plasma edge. In the case of the ILW it develops at significantly lower edge temperatures, since the radiation efficiency of Be is lower than for C and shows its highest values at lower temperatures. In conjunction with the different properties of recycling, this produced an increase of the densities at which is possible to operate JET [7].

Another important aspect related to the installation of the ILW, is the presence of new causes and new chains of events which lead to disruption [8]. An analysis identical to the one described for the Carbon Wall has been carried out, as shown in the scheme in Figure 8.12. Therefore, a statistical analysis on the root causes has been performed and characteristic chains of events have been identified to group those disruptions which follow the same path [8]. Besides the changes in the composition in terms of different classes, which will be described in the following section, the main peculiarity is related to the onset of new causes, among which the dominant is the one indicated in the scheme in Figure 8.12 as "RPK", that is strong radiation peaking. This phenomenon has occurred in 4.6% of all the discharges operated in $2011 - 2012$. Several discharges disrupted due to this high core radiation, which in part happened during the main heating phase and in part after the switch-off of the auxiliary heating.

Although the causes for the strong radiation peaking are not completely clear, it is thought to be associated mainly to the transport properties of high Z-impurities, which give rise to a strong accumulation in the core, or in other cases, the radiation increase occurs for a sudden influx of impurities due to an enhancing of the divertor source. This two processes have a different time scale, in particular, the first one develops on a much slower resistive time scale and is characterized by the hollowing of the temperature profile, with at the same time, a strong peaking of the density profile. On this time scale, becoming hollow, the temperature profile starts to modify the q and the current density profiles, which could be driven unstable by the broadening of the q-profile itself [8]. This broadening can also be observed in terms of reduction of the plasma internal inductance.

Another important point that has to be taken into account, especially from the point of view of the prediction, is related to the fact that with the ILW the current quench is significantly slower than what we had with the CW. In particular, if we make reference to the threshold of $dI_p/dt > 5MA/s$ for defining disruption, there are even cases in which a thermal quench is not

followed by any current quench. Or in other cases it can happen that, after a first thermal quench, the temperature recovers and another chain of events not directly connected with the previous one, can eventually takes place and lead to disruption.

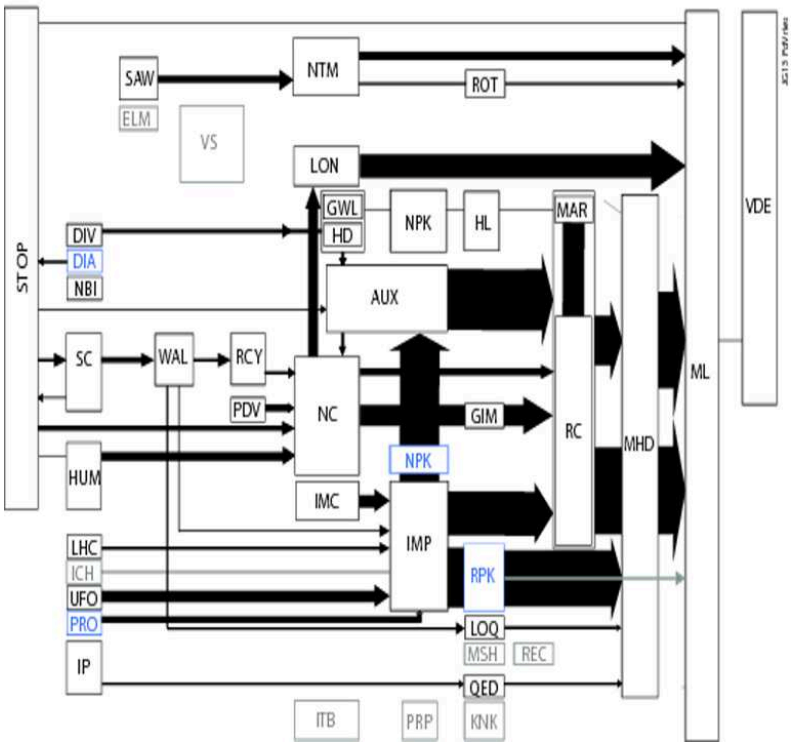


Figure 8.12: Schematic overview showing the statistics of the chain of events for non-intentional disruptions with the ILW (2011 – 2012) [8].

8.3.3 ILW versus CW operational space of JET

After campaign C27, JET installed the new ITER Like wall (ILW). The first attempt was to project the disruptions of the ILW campaigns onto the GTM map trained with the CW discharges, but the performance of the map in classifying the new disruptions significantly deteriorated for certain classes, probably because of the fact that the operational space, or at least, the considered feature space is changed.

Therefore, a more detailed analysis has been performed to investigate how the operational space of JET with the new ITER Like wall eventually changed and if the disruption physics modified with respect to the CW experiments. The whole database consists of 243 non intentional disruptions

belonging to the CW campaigns from C15 to C27, and of 149 non intentional disruptions of the ILW campaigns from C28 to C30. In Table 8.2, the composition and the occurrence in terms of different classes is reported for both the CW and the ILW databases.

Disruptions		CW		ILW	
Labels	**Classes**	**Tot**	**Tot %**	**Tot**	**Tot %**
ASD	Auxiliary Power Shut-Down	50	20.58	2	1.34
GWL	Greenwald Limit	9	3.70	0	0.00
IMC	Impurity Control Problem	83	34.16	109	73.15
ITB	Too Strong Internal Transport Barrier	10	4.12	0	0.00
LON	Low Density and Low q	12	4.94	7	4.70
NC	Density Contol Problem	58	23.87	22	14.77
NTM	Neo-classical Tearing Model	21	8.64	9	6.04

Table 8.2: CW vs. ILW Database.

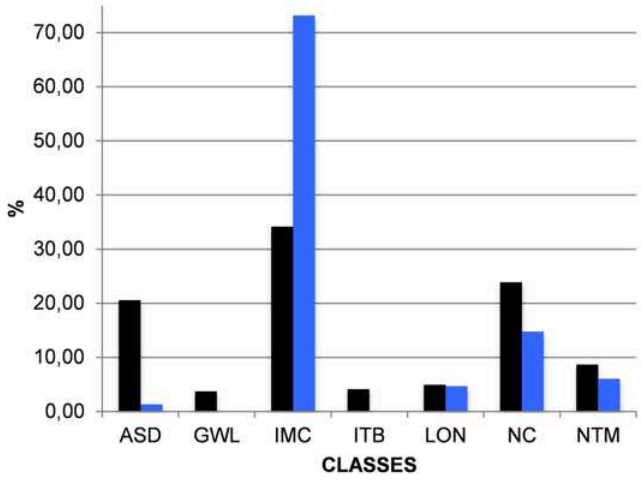


Figure 8.13: Distribution of disruptions in the CW (black) and ILW (blue) campaigns.

As it can be seen from Table 8.2 and Figure 8.13, the composition of the two data bases is quite different: in particular, disruptions due to Greenwald limit or due to too strong ITB are no longer present in the new campaign, whereas the number of disruptions due to IMC consistently increased.

Moreover, a new class has been identified, characterized by strong radiation peaking due to impurity tungsten accumulation in the core of the plasma

(new Impurity Control Problems disruptions: IMC_new). The distinction in different classes is based on the manual classification described in [8], where specific chains of events have been detected and used to classify disruptions, grouping those that follow specific paths.

A statistical analysis has been then performed on the plasma parameters describing the operational spaces of JET with CW and with ILW. In Figure 8.14 the probability density distributions of the last 210 ms of Plasma current I_p (a), Safety Factor at 95% of Poloidal Flux q_{95} (b), Plasma Internal Inductance l_i (c) and Line Integrated Plasma Density ne_{lid} (d) have been reported for the IMC disruptions with the CW (red lines) and with the ILW (grey dashed lines), and for the new impurity type (IMC_new) with the ILW (blue dashed lines).

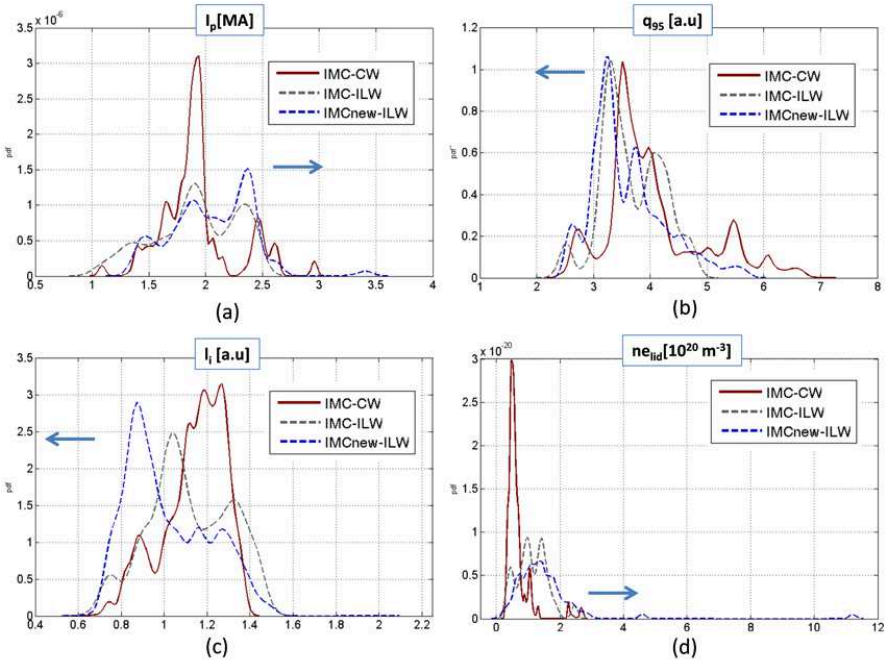


Figure 8.14: Probability density distributions of: (a) Plasma current I_p; (b) Safety Factor at 95% of Poloidal Flux q_{95}; (c) Plasma Internal Inductance l_i; (d) Line Integrated Plasma Density ne_{lid}.

The analysis highlights interesting features in particular for the new impurity type class, confirming that a new GTM is needed to represent the ILW operational space of JET. From Figure 8.14, it can be seen that it is quite difficult to discriminate among classes just from the distribution of the signals. In fact it is well known that what is really important is the combination of the signals.

Moreover, looking at the previous figures, some interesting features can be found: for the new impurity type class the distribution function of internal inductance is shifted towards lower values, whereas the electron density function is shifted toward higher values. This can be probably due to the fact that the impurity accumulation of the tungsten in the core, in certain cases, when a certain concentration is reached, starts to modify the temperature and the current profiles giving rise to instabilities followed, in some cases, by disruptions. Further analysis can be made to compare different disruption classes behavior passing from CW to ILW. Regarding density control problem and impurity control problem classes, Figure 8.15 reports the probability density functions of I_p and l_i for the IMC and NC disruptions with CW, whereas Figure 8.16 reports the distributions of the same signals for the IMC, IMC_new and NC disruptions with ILW.

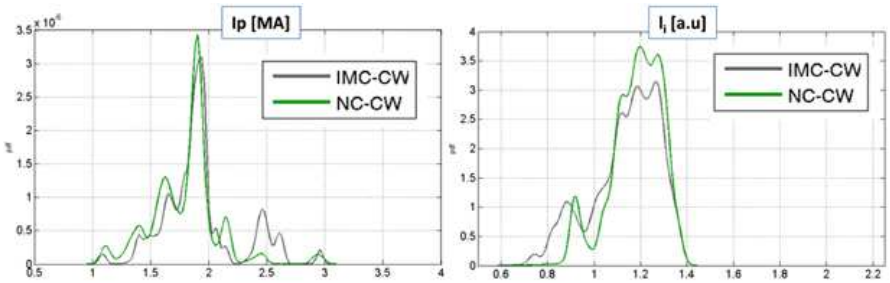


Figure 8.15: Probability density functions of Ip (left side) and li (right side) for the IMC (grey) and NC (green) disruptions with CW.

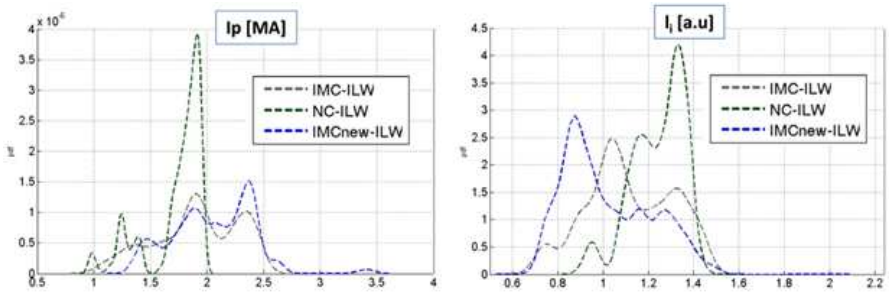


Figure 8.16: Probability density functions of Ip (left side) and li (right side) for the IMC (dashed grey), IMC_new (dashed blue) and NC (dashed green) disruptions with ILW.

From figure 8.15, it can be seen that, with the CW, both I_p and l_i signal distributions are more or less overlapped and this is coherent with the analysis

of the disrupted regions presented in the section 8.2.2. Conversely, for the ILW, I_p and l_i distributions result to be quite different, especially if we compare NC and IMC_new classes. In particular, for the plasma current, it can be seen that, at least in the last 210 ms, there are no NC disruptions above 2 MA. Moreover, the new impurity type basically occurs for lower values of the internal inductance.

8.3.4 Mapping of the JET ITER-like Wall operational space

Starting from the previous statistical analysis and the physical considerations on the new disruption types, a new GTM has been trained to represent the ILW operational space of JET. The training set consists of the last 210 ms of the 149 non intentional ILW disruptions (29137 samples), which have been mapped through 81 radial basis functions (Gaussian shape) with a 1.5 width, over a latent space of 36x36 grid.

In Figure 8.17(a) the Mode representation of the GTM is reported. Figure 8.17(b) shows the GTM Pie Plane representation. In such visualization, each node is represented by a pie chart describing the percentage composition in terms of number of samples belonging to the different disruption classes. The samples are diversified according to the color code reported on the legend in the same figure, with reference to the different classes of disruptions.
Both representations highlight a high level of separation among the different classes with respect to what has been found for the Carbon Wall. In Table 8.3, the level of separation of the different classes is reported in terms of percentage of samples of each class which is projected in nodes entirely composed by samples of the considered class.

Classes	**Class Samples (%)**
ASD	15.86
IMC	93.51
LON	68.16
NC	77.57
NTM	60.38

Table 8.3: Discrimination capability of GTM model for the considered classes.

Figure 8.18 shows the same map (Mode (a) and Pie Plane (b) representations), trained with the same training parameters, where the IMC_new class has been introduced.
It can be seen that the new class is even better separated with respect to

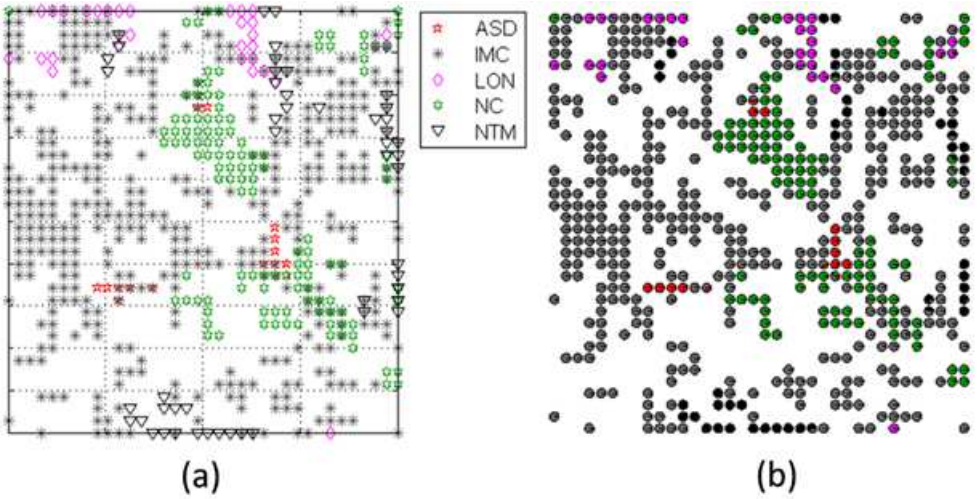


Figure 8.17: 2-D GTM of the 10-D JET ILW operational space: (a) Mode Representation. The nodes are represented with different color and symbols as indicated in the legend, empty nodes are white; (b) Pie Plane Representation. The nodes composition in terms of the five different classes of disruptions is represented according to the color code reported on the legend.

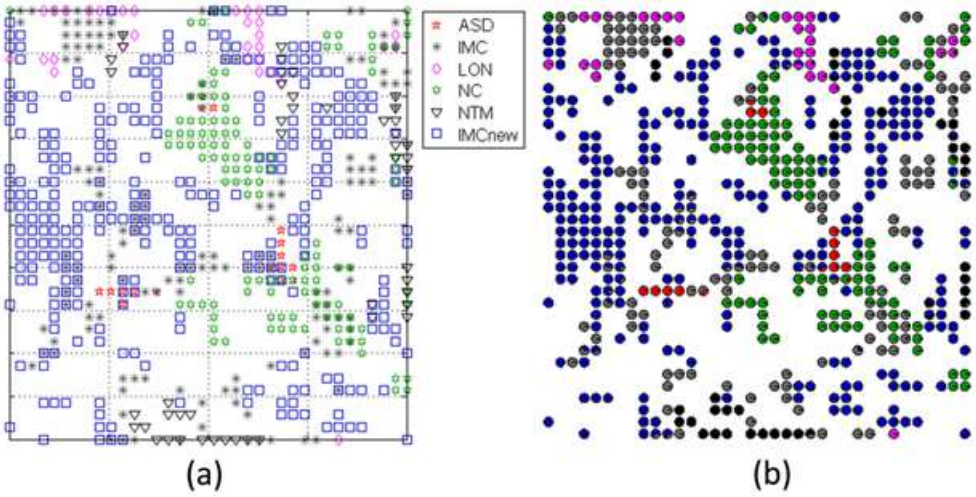


Figure 8.18: 2-D GTM of the 10-D JET ILW operational space with the IMCnew disruption class: (a) Mode Representation. The nodes are represented with different color and symbols as indicated in the legend, empty nodes are white; (b) Pie Plane Representation. The nodes composition in terms of the six different classes of disruptions is represented according to the color code reported on the legend.

other classes. Table 8.4 reports the same information of Table 8.3, but with the addition of the new impurity type class.

Classes	Class Samples (%)
ASD	15.86
IMC	72.90
LON	68.16
NC	77.57
NTM	55.36
IMCnew	91.18

Table 8.4: Discrimination capability of GTM model for the considered classes.

It is interesting to observe in fact, that, coherently with what has been found for the CW operational space, the main contribution to the nodes shared by samples of density control problem and impurity control problem disruptions is given by the old "IMC" class, whereas the overlapping on the map presented by the new impurity type is mainly with the IMC class itself.

8.3.5 Real time application in conjunction with APODIS

On the base of only the mapping of the operational space, having zoomed on the disruptive space, one can guess that by applying the majority voting algorithm to the last 210 ms, practically all of the disruptions can be correctly classified according to the manual classification; in fact apart one isolated case it is what it happens.

But one of the main objective of this study is to develop a system that can be used in real time and can be integrated eventually with the other systems already working in real time at JET. Therefore, in order to test the performance in classification of the new maps, a real time application has been simulated in conjunction to APODIS [4], improving at the same time the efficiency of the codes and assessing finally the suitability for real time applications.

APODIS (Advanced Predictor Of DISruptions) is a disruption predictor mainly constituted of three different Support Vector Machine (SVM) predictors, which analyze three consecutive time windows (each one 32 ms long) of data to take into account the history of the discharge. It has been deployed in JET's real-time system during the last campaigns with the ILW with very good results (well above 90% of Success Rate) and it is presently working in the ATM network in open loop.

During the simulation, the majority voting algorithm has been applied

to the class membership function of a time window of respectively 32 or 64 ms right before the time in which APODIS triggers the alarm. Note that, in several cases APODIS gives the alarm significantly in advance with respect to the thermal quench time, even hundreds of ms in advance. Table 8.5 reports the results of the real time automatic classification performed by the GTM trained considering the classes previously defined for the CW campaigns, therefore without any distinction in the different impurity types disruptions. As can be seen, the Success Rate is quite high reaching more than 90%, thus in very good agreement with the manual classification.

GTM	GLOBAL	ASD	IMC	LON	NC	NTM
32ms	93.23	100.00	94.00	66.67	100.00	85.71
64ms	94.07	100.00	95.10	66.67	100.00	85.71

Table 8.5: Percentage success rates of the real time automatic classification performed by GTM on the classes identified for the CW campaigns.

The classification performances slightly deteriorates when the new class is considered, as shown in Table 8.6. This is due to the difficulty to discriminate in certain cases the new class from the previous impurity control problem one.

GTM	GLOBAL	ASD	IMC	LON	NC	NTM	IMCnew
32ms	87.22	100.00	67.86	66.67	100.00	83.33	93.15
64ms	85.93	100.00	71.43	66.67	100.00	83.33	89.33

Table 8.6: Percentage success rates of the real time automatic classification performed by GTM considering the IMC_new disruption class.

8.3.6 Validation and comparison

In order to validate and analyze the results obtained with GTM, another reference classifier has been developed based on k-NN which uses as kernel the Mahalonobis distance (see chapter 5). The simulations have been performed using as kernel different metrics, such as the Euclidean or the Hamming distances, but, at least for this specific problem, Mahalanobis distance performs quite better with respect to the other tested metrics. k-NN is a reference instant based classifier, unlike GTM that builds a generative latent model. In this case the majority voting is applied to the k closest points in the high

dimensional space. Table 8.7 reports the performance of the k-NN classifier for the classes identified for the CW campaign. Table 8.8 shows the k-NN performance when the new impurity type is considered. Also in this case, the global performance is above 90% when the new impurity control problem class is not considered, whereas the performance slightly deteriorates when the new class is considered.

k-NN	GLOBAL	ASD	IMC	LON	NC	NTM
32ms	92.91	100.00	95.19	71.43	90.48	85.71
64ms	92.20	100.00	95.19	71.43	90.48	85.71

Table 8.7: Percentage success rates of the real time automatic classification performed by k-NN classifier considering the classes identified for the CW campaign.

k-NN	GLOBAL	ASD	IMC	LON	NC	NTM	IMCnew
32ms	90.78	100.00	82.14	71.43	95.24	83.33	94.81
64ms	87.94	100.00	82.14	71.43	90.48	83.33	90.91

Table 8.8: Percentage success rates of the real time automatic classification performed by k-NN classifier considering the IMC_new disruption class.

The class membership function of the disruptions gives us useful information. As an example, in Figure 8.19 the class-memberships of the pulse No. 82867 is reported for both GTM and k-NN, which is a IMC disruption according to the manual classification.
It is possible to note a transition among different classes, and in particular the one between NCs and IMCs or vice versa, which is not uncommon both for CW and ILW. Note that APODIS alarm is triggered almost two seconds before the thermal quench. It is also very important to point out that both the classifiers converge onto the same results, even if in this specific case we can observe that for GTM based classifiers the phase where we can associate the highest probability to the correct class is about 400 ms before t_D, whereas in the k-NN is more than 700 ms in advance.

In Figure 8.20 the time evolution of some of the signals which are part of the database is reported for the discharge No. 82867, whereas in Figure 8.21 a zoom of the previous plots is reported with respect to the time window analyzed in Figure 8.19.

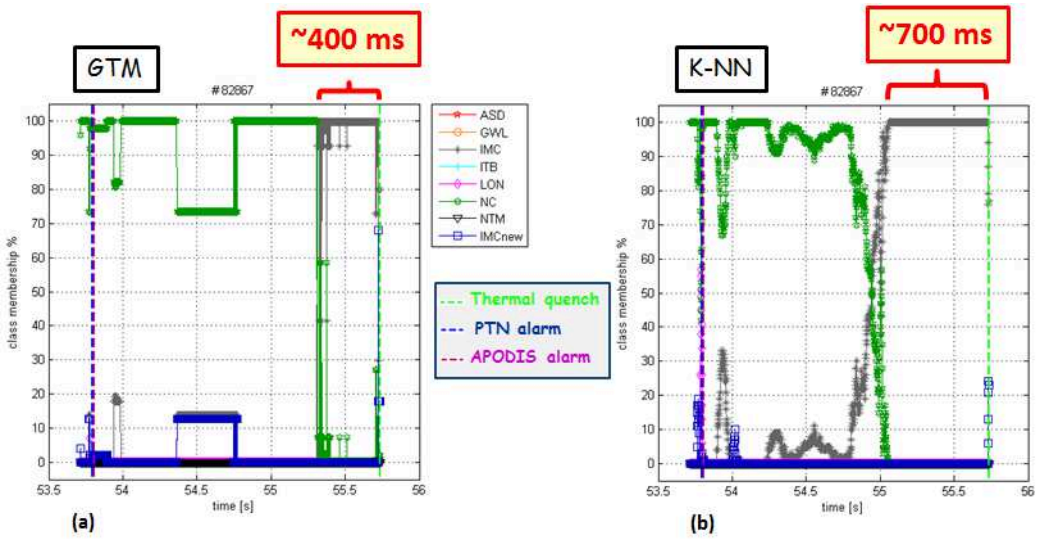


Figure 8.19: Class-membership functions of the shot No. 82867 (IMC) for GTM (left side) and k-NN (right side). According to the legend, the vertical green line identifies the thermal quench, the blue line the PTN alarm, and the pink line the APODIS alarm.

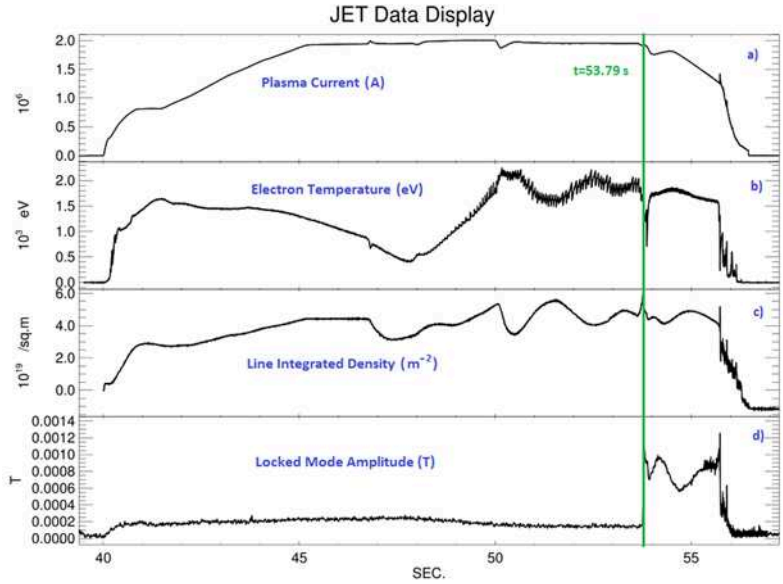


Figure 8.20: Time evolution of a) plasma current, b) central electron temperature from Electron Cyclotron Emission (ECE) measurements, c) line integrated density and d) locked mode amplitude for the current flat-top phase of the shot No. 82867; the vertical line in green represents the time of the locked mode.

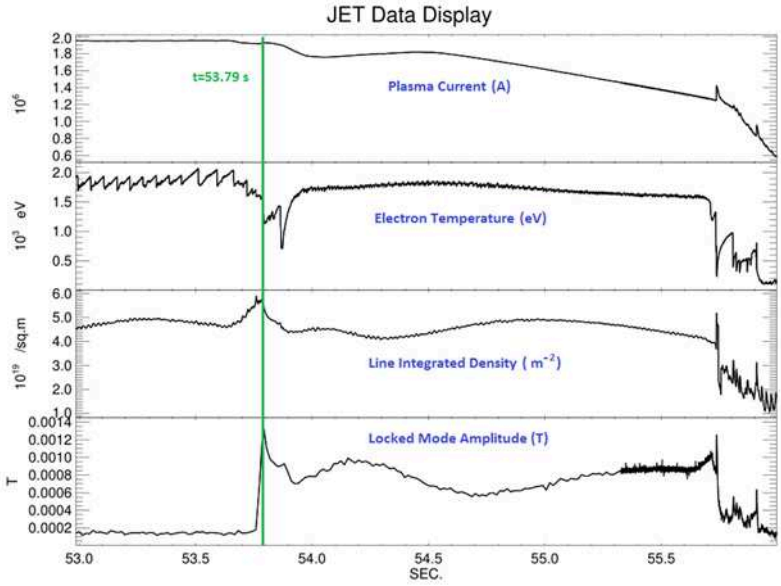


Figure 8.21: Zoom of Figure 8.20 (time interval $[53-56]$s - shot No. 82867).

As it can be seen from these pictures, a Reciprocating Langmuir Probes (RCP) caused a locked mode at $t \simeq 53.79s$, time around which a rapid change of the density occurs, followed by a quench of the temperature that in the subsequent phases recovers up to the final thermal quench at $t \simeq 55.73s$. Both PTN and APODIS trigger the alarm when the mode locks (see Figure 8.19)and for both the classifiers the discharge is evolving as a NC disruption up to the final phase where is correctly classified as IMC, according to the manual classification.

Given the complex behaviors which often characterize the evolution of a discharge, it is important to know the reliability and the confidence of the classification. Literature provides recent methods, such as the conformal predictors, already described in chapter 5, which allow us to take into account also this aspect. To this purpose, a conformal predictor has been developed which is based on non-conformity measures.

Regarding classification, the conformal predictors can provide the level of reliability of classification itself with two parameters: the *credibility* and the *confidence* which are defined on the base of the *p-values* (see chapter 5). In figure 8.22 the label provided by the classification, the credibility and the confidence levels are reported for pulse No. 82867. As it can be seen the credibility, which is the parameter with more variability, is quite low for all the initial phase, then it rises constantly during the last $\sim 400ms$, according to the results obtained with the GTM based classifier.

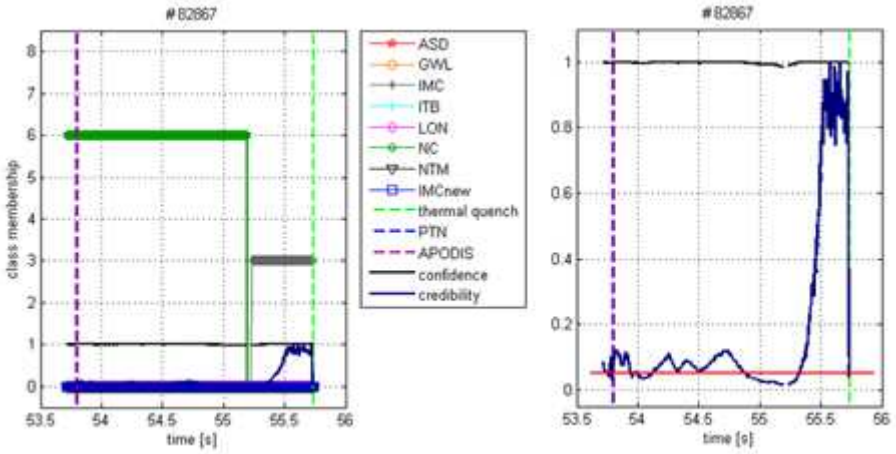


Figure 8.22: Left side: class-membership provided by the conformal predictor for the shot No. 82867, confidence level (blue) and credibility (black). The vertical green line identifies the thermal quench, the blue line the PTN alarm, and the pink line the APODIS alarm (left side). Right side: zoom representing the confidence level (blue), the credibility (black) and the threshold of 0.05 (red) (right side).

The credibility, even if low in the phase where the conformal predictor is assigning the label corresponding to the NC class, is mostly above 0.05, which in literature is often used as threshold for trusting or not a prediction (Figure 8.22 (right side)). In general, if the credibility is less than 5%, the considered samples are not representative of the training set, or in other words we cannot consider that they are generated independently from the same distribution. In particular, the credibility falls under the considered threshold in correspondence of the transition between NC and IMC classes, behavior that could depend on a rapid reconfiguration or a change in the considered parameters' space. Further analysis are needed to clarify this point.

In Figure 8.23, the class memberships function obtained with the GTM (a) and with the k-NN (b) based classifiers are reported for the pulse No. 82569, which has been manually classified as IMC disruption.
What is particularly interesting to observe in this case is the fact that, apart the agreement in the classification provided by the two methods, if we look at the confidence level plotted in Figure 8.24, we find that remains very high for a long phase. In fact looking at the projection on the map (Figure 8.25), we can see that the discharge is evolving in a not extended region of the operational space, and this mean that the parameters are not changing too much in the considered time interval, at least up to the last phases just before the disruption. This is confirmed by the time evolution of some of the considered signals, as we can see in Figure 8.26.

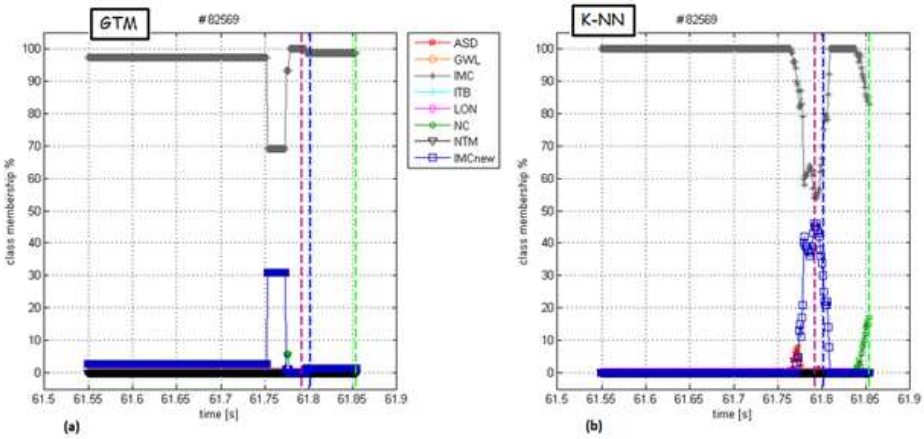


Figure 8.23: Class-membership functions of the shot No. 82569 (IMC) for GTM (left side) and k-NN (right side). According to the legend, the vertical green line identifies the thermal quench, the blue line the PTN alarm, and the pink line the APODIS alarm.

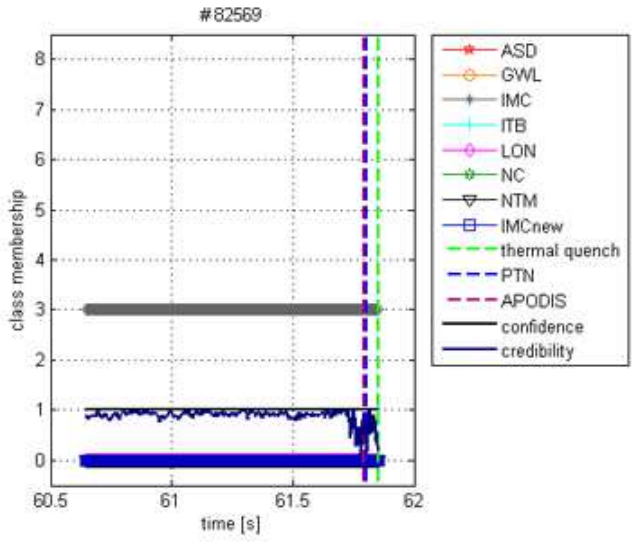


Figure 8.24: Class-membership provided by the conformal predictor for the shot No. 82569, confidence level (blue) and credibility (black). The vertical green line identifies the thermal quench, the blue line the PTN alarm, and the pink line the APODIS alarm.

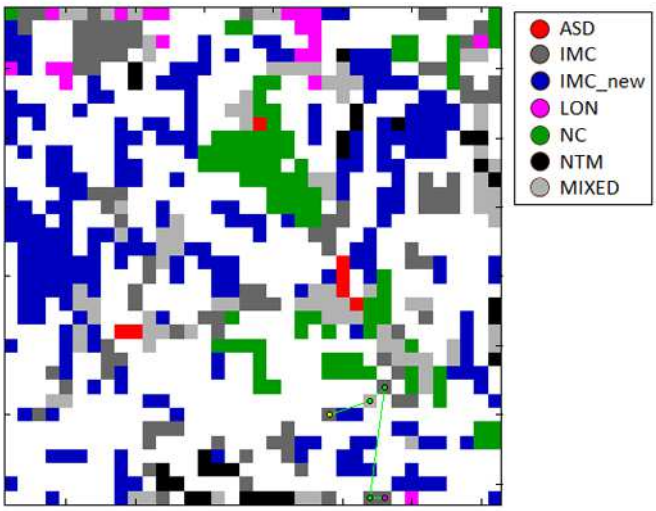


Figure 8.25: Projection of the discharge No. 82569 on the GTM map

Regarding the disruptions belonging to the new impurity type, in Figure 8.27 is reported an example of disruption due to impurity accumulation. In this case the accumulation of W occurs after a step-down of the Neutral Beam Injection power, and the hollowing of the temperature profile can be observed in correspondence of the increasing of radiation by tungsten. This, on the slow time scale of the transport process, affects the current density and the q profiles, driving MHD modes unstable until we have a locked mode.
By computing the automatic classification it is possible to verify that the pulse is correctly classified as belonging to the new impurity type by all the implemented systems. Furthermore it is interesting to see that when the mode locks, there are "jumps" in the class-membership calculated by the conformal predictor, and the credibility in this interval drops practically to zero. In the interval prior to the locked mode, again all the classifiers are clearly recognizing the new impurity type, as it is shown in Figures 8.28 and 8.29.

8.3.7 Discussion of the results

The challenge to automatically discriminate the type of disruptions at JET both in the Carbon wall (CW) campaigns and in the ITER Like wall (ILW) ones has been tackled using a GTM manifold learning method. The disruption classes in the ILW have been deeply analyzed and compared with those

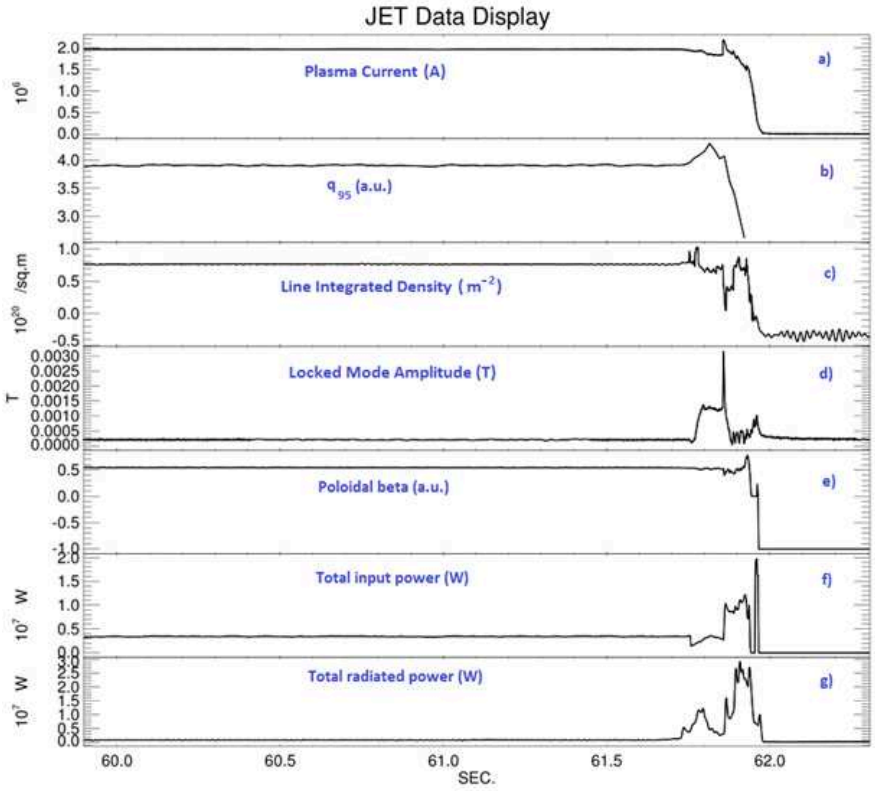


Figure 8.26: Time evolution of a)plasma current, b) q_{95}, c) line integrated density, d) locked mode amplitude, e) poloidal beta, f)total input power and g) total radiated power measure by bolometer for the shot No. 82569.

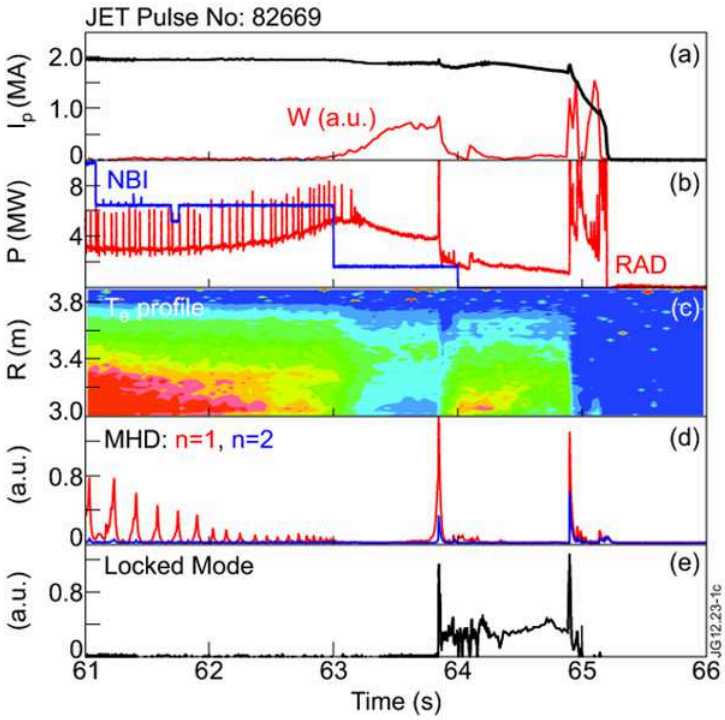


Figure 8.27: Example of disruption caused by impurity accumulation (discharge No. 82669) [6]

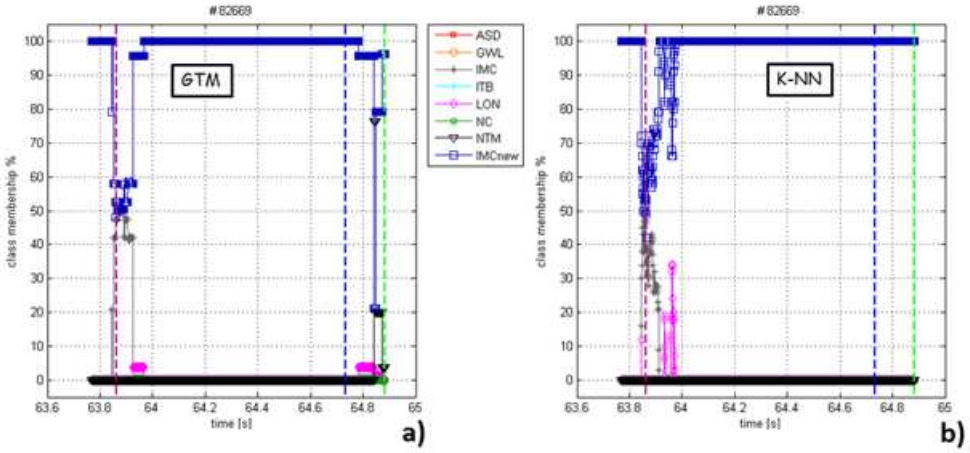


Figure 8.28: Class-membership functions calculated through a) GTM and b) k-NN for discharge No. 82669.

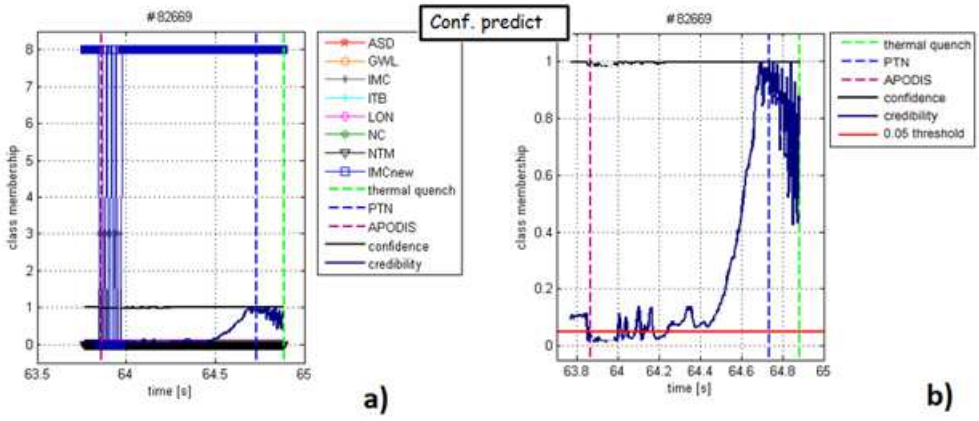


Figure 8.29: Class-membership functions calculated through a) conformal predictor for discharge No. 82669; in b) a zoom of a) is reported regarding the confidence level (black) and the credibility (blue).

in the CW JET campaigns. In particular, the probability density functions of the different plasma parameters highlight the different behaviors of the new impurity control problem disruptions, due to high-Z impurity accumulation in the core of the plasma column, with respect to the old IMC ones. Moreover, the statistical analysis showed the variation of the operational space of JET with ILW with respect to that with CW.

For this reason, two different GTM maps have been trained for CW and ILW campaigns. The latter has been used to simulate a real time behaviour of the GTM classifier in conjunction with the prediction system APODIS, which is successfully working on line at JET. The obtained results assess the suitability of the GTM based classifier for real time application with very good results: the prediction success rate is quite high (above 90%) according to the manual classification. However, the performance worsened when the new IMC class is introduced, because it is quite difficult to distinguish this new class from the previously defined IMC class. Furthermore, in order to validate and analyze the obtained results, another reference classifier has been developed, based on k-NN, which uses as kernel the Mahalanobis distance. The performance of the reference classifier is still above 90%, but, also for it, the success rate deteriorates when the new IMC class is introduced.

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

perform a statistical analysis of the transitions among different classes.

Finally, in order to verify the reliability of the performed classification, a conformal predictor has been developed which is based on non-conformity measures. The obtained results indicate the suitability of the conformal predictors to assess the reliability of the GTM classification even if the calculation time allowed their use only in an off line fashion.

The mapping of the JET operational space has been built on the base of a set of signals which are available in real time with an high reliability, and every device should provide for each discharge in standard operations.
This does not mean that the performance of the system in mapping and classification could not improve if additional information are taken into account. In [8] the disruptivity, which gives the likelihood of a disruption within a specific parameter space, has been calculated in terms of different parameters, as for example the temperature peaking and the radiation peaking.
Some parameters representative of the profiles of certain quantities which would help to improve the discrimination capability of the proposed systems, but often they require a post-pulse validation. Therefore a trade-off between performance and reliability has to be carefully considered, without forgetting, on the other hand, real-time and computational constraints.

Bibliography

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

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

[3] 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.

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

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

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

[7] Huber A *et al.* 2012 *Impact of the ITER-like wall on divertor detachment and on the density limit in the JET tokamak* 20th Conf. on Plasma Surface Interaction (Aachen, Germany, 2012) J. Nucl. Mater. submitted

[8] de Vries *et al.* 2013 5th Annual Meeting of the APS Division of Plasma Physics, Denver, Colorado, USA, *The influence of an ITER-like wall on disruptions at JET*

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

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

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

[12] Zhang Y., Pautasso G., Kardaun O., Tardini G., Zhang X.D. and the ASDEX Upgrade Team 2011 *Prediction of ASDEX Upgrade disruptions using discriminant analysis* Nucl. Fusion 51 063039-41

[13] Rapp J. *et al.* 1999 *Density limits in TEXTOR-94 auxiliary heated discharges*, Nucl. Fusion 39 765

[14] Borrass K. et al. 2004 *Recent H-mode density limit studies at JET*, Nucl. Fusion 44 752.

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

[16] Gammerman A. and Vovk V. 2007 *Hedging Predictions in Machine Learning*

Chapter 9

Disruption prediction at ASDEX Upgrade

9.1 Introduction

In this chapter, two different approaches are proposed as disruption predictors at ASDEX Upgrade. The first method consists in extracting information from the complex multidimensional operational space of the machine by means of data visualization and dimensionality reduction methods. The second method allows to build a black-box predictor which provides a statistic predictive model.

In this study, among the visualization and dimensionality reduction methods, the Self-Organizing Map and the Generative Topographic Mapping are investigated. The 2-D mappings provided by SOM and GTM are used with good results as disruption predictor by associating the risk of disruption of each cluster in the map to a disruption alarm threshold. Furthermore, following the trajectory of the plasma on the maps it is possible to study its behavior leading to a disruption; thus, it can be taken advantage of this additional value to realize a single system for disruption prediction and classification.

Among the multivariate statistical models the Logistic regression approach is proposed. The Logistic model works as disruption predictor by monitoring the probability of a disruptive event during the experiments. Despite its simplicity, good results have been achieved, but being a probabilistic classification model the logistic predictor does not make available any other additional information on the plasma state evolution.

In addition, the two methods have been combined to realize a predictive system able to exploit the complementary behaviors of the two approaches.

The data base for this study represents the 7-D ASDEX Upgrade operational space described by means of disrupted and safe discharges selected in the shot range $21654 - 26891$, and performed in ASDEX Upgrade between May 2007 and April 2011.

9.2 Database

Data for this study were extracted from the AUG experimental campaigns performed between 2007 and 2012, in the shots range $21654 - 28832$. The database has been divided in three subsets, following a temporal progression as reported in Table 9.1. The Training set, containing only discharges performed between May 2007 and April 2011, has been used to build the maps and to optimize the coefficients of the logit model; the data set Test_1, containing shots performed in the same time period of Training set, has been used to test the generalization capability of the maps and the logit model; the set Test_2, containing shots performed after those of Training set, has been used to evaluate the ageing of the models when used during more recent campaigns.

Data Set	**Time Period**	**Safe Discharges**	**Disrupted Discharges**
Training	May 2007 - April 2011	310	121
Test_1	May 2007 - April 2011	155	60
Test_2	April 2011 - November 2012	271	106

Table 9.1: Data base composition.

Only disruptions which occurred in the flat-top phase or within the first 100 ms of the plasma ramp-down phase, and characterized by a plasma current greater than 0.8 MA, are considered. Disruptions mitigated by massive gas injection, triggered by the locked mode detector, and those caused by vertical instabilities, so called Vertical Displacement Disruptions (VDDs), were excluded. Each of the three data sets is composed by time series related to the seven plasma parameters reported in Table 9.2. All the parameters are sampled making reference to the time base of the plasma current signal. The sampling rate is equal to 1 kHz.

Signal	Acronym
Safety factor at 95% of poloidal flux	q_{95}
Total input power	P_{TOT}
Locked Mode signal	LM signal
Radiated fraction of the input power	$P_{frac} = P_{rad}/P_{TOT}$
Plasma density divided by the Greenwald limit	$ne_{Greenwald}$
Internal inductance	l_i
Poloidal β	β_p

Table 9.2: Plasma parameters considered in the data base.

9.3 2-D AUG operational space mapping

One of the viable way to realize a disruption predictor consists in extracting information from the multidimensional operational space of the reactor by means of data visualization and dimensionality reduction methods as SOM and GTM. Looking at the good results on disruption prediction accomplished by the authors with SOM on a foregoing AUG databases [1], and on operational space mapping with GTM on the JET database [2], in this work, both SOM and GTM have been employed to realize a 2-D mapping of the 7-D AUG operational space on the considered database.

In order to project the 7-D AUG operational space onto the 2-D SOM and GTM, further knowledge is added to the intrinsic knowledge owned by the 7-D data space, which consists in associating a label to each sample in the training set. Samples coming from safe discharges have been labeled as safe samples *(ss)*. For disrupted discharges, a time value, called $t_{PRE-DISR}$, has to be assumed to discriminate between the safe and disruptive phases. On the basis of previous experiences [3], an empirical value equal to 45 ms from the disruption time (t_D) has been taken for each discharge. Therefore, samples preceding $t_{pre-disr}$ are considered as safe samples (ss), whereas samples in the interval $[t_{pre-disr} \div t_D]$ are labeled as disruptive samples (ds). Only the disruptive samples and safe samples from safe discharges have been included in the Training set; the safe phase of disruptive discharges is assumed to be well represented by the safe samples of safe discharges.

The training set results in a large amount of data, 310 safe discharges make 1094697 (ss) available, whereas 121 disruptive discharges make 5267 (ds) available. As it can be noted the group of safe sample is 210 time larger than the disruptive one. Thus, in order to avoid the predominant influence of safe samples with respect the disruptive ones during the training phase, and with the aim to aggregate the expected redundant information contained in a so large database into a more manageable and efficient one, a data re-

duction on safe samples has been performed. The goal of the data reduction procedure has been to achieve a ratio $ss/sd < 10$; that value comes from the authors previous experiences on AUG and JET database, in [1] a data reduction with $ss/sd = 7.1$ and in [5] a data reduction with $ss/sd = 6.4$ were performed respectively.

A reduced number of representative safe samples have been selected by a procedure where firstly each signal has been quantized in a range of values depending on its own distribution. Then, the 7-D space has been partitioned with a 7-D grid where each node is the combination of the quantized signal values. Finally, a fraction of the samples corresponding to each node has been selected. The nodes to which correspond a number of samples lower than the mean value of samples per node on the entire grid have been excluded from the selection. The adopted criterion allows us to exclude from the database the samples related to unusual signal combinations. The data reduction procedure results in 39115 safe samples, with $ss/sd = 7.43$.

Figure 9.1 shows the pdf of the considered plasma parameters for safe samples before (blue line) and after the data reduction (green dashed line). As it can be noted the trend of the seven signals remain unchanged after the data reduction. Figure 9.2 shows the projection of the 7-D space of safe sample before the data reduction (blue points) and after the data reduction (green points) onto the first three principal components. As can be noted, only regions with low density are uncovered after the data reduction. Figure 9.3 displays the two mappings obtained with the reduced safe samples group and the disruptive samples belonging to the training set; figure (a) reports the GTM with 1600 map units and the figure (b) reports the SOM with 1674 map units. The dimension of SOM, i.e. the number of map units, has been selected with an heuristic rule proposed in [4]; for comparison purposes also the GTM size has been chosen applying the same rule. On the two maps four types of map unit can be identified depending on their composition: safe map units containing safe samples, disruptive map units containing disruptive samples, mixed map units containing both safe and disruptive samples and empty map units containing no samples. A color code has been adopted to identify the four map unit categories on the map. The safe map units are green, those disruptive are red, the mixed map units are grey and finally the empty ones are white.

For both maps, a large safe region (the green one) where the risk of disruption is low, two mainly disruptive regions (in red) where the risk of disruption is high, and transition regions as boundary between the previous ones, can be identified. Tracking the temporal evolution of plasma discharges both on the GTM and the SOM, it has been observed that the great majority of the safe discharges evolves within the safe region, as the yellow trajectories reported

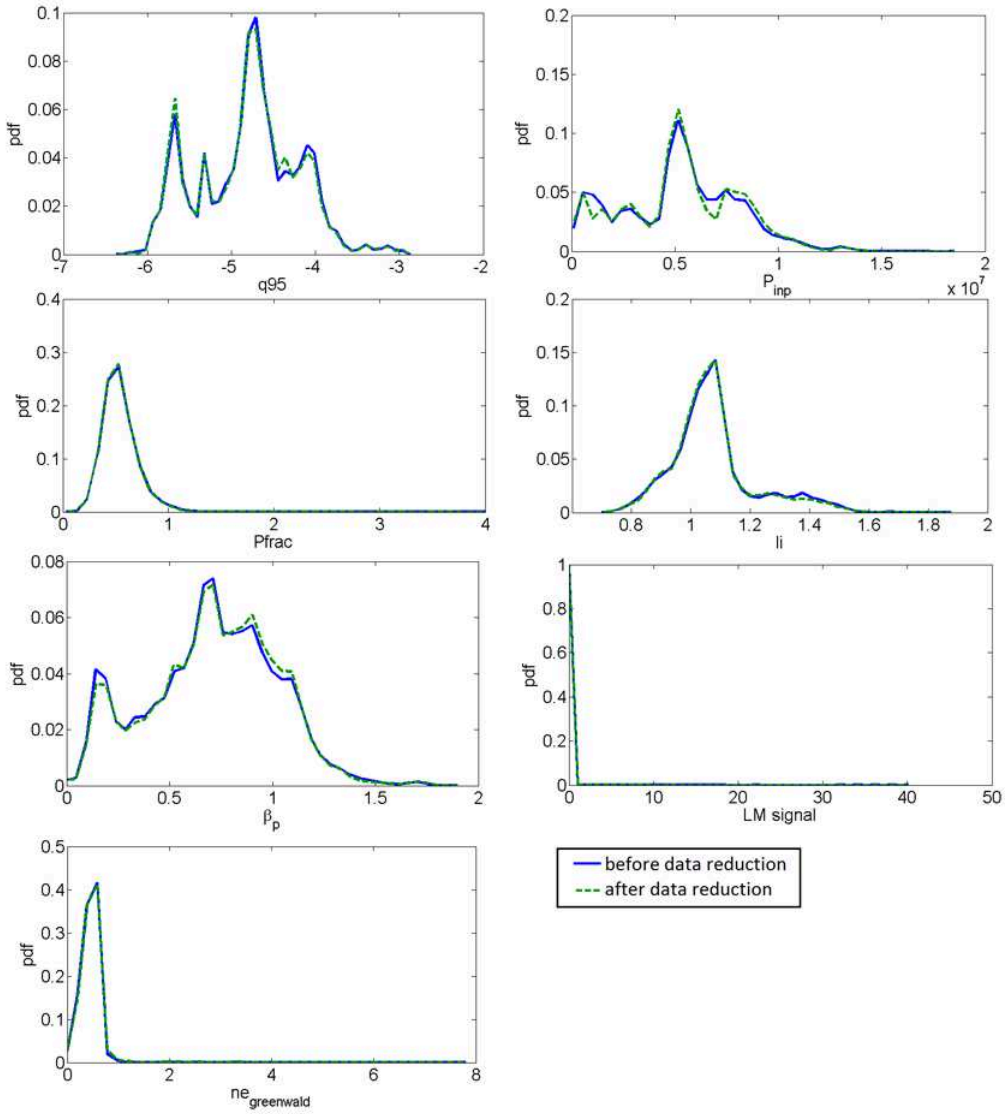


Figure 9.1: Probability density functions of the plasma parameters included in the data base for safe samples, before the data reduction (blue line) and after the data reduction (green dashed line).

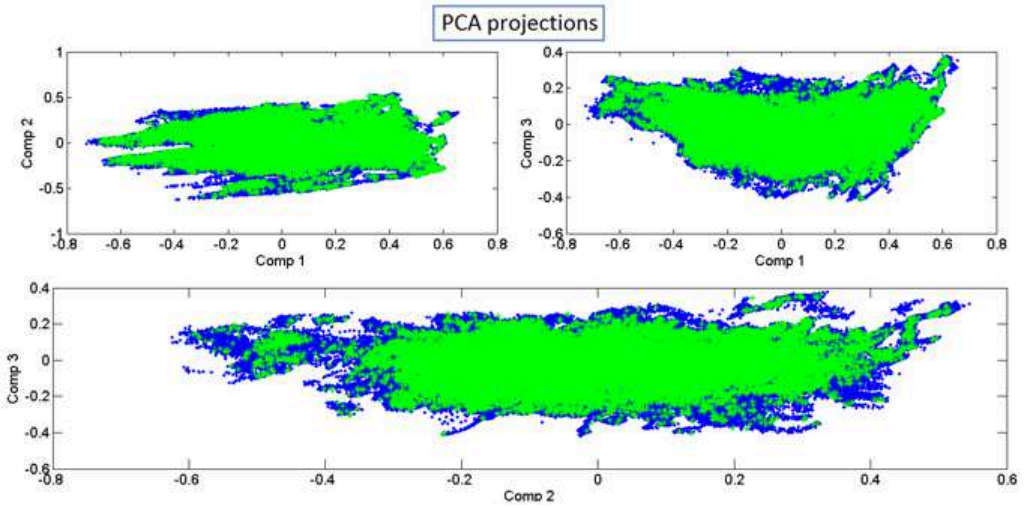


Figure 9.2: PCA projection of the 7-D AUG safe space of training set before the data reduction (blue points) and after the data reduction (green points) onto the first three principal components.

on the two maps in Figure 9.3. On the contrary, for the majority of disruptive discharges, the trajectory starts in the safe region and, passing through the transition region, ends in a disruptive region, as the blue trajectories on figure 9.3. This suggested us to use both maps as disruption predictors by linking the disruption alarm to the disruption risk of the different regions.

9.4 Disruption Predictors

In order to evaluate the prediction performance of the two maps as predictors some performance indices have been introduced:

- Successful Predictions (SPs): the fraction of safe or disruptive discharges correctly predicted.
- Tardy Detections (TDs): the fraction of disruptive discharges which triggers the alarm too late.
- Missed Alarms (MAs): the fraction of disruptive discharges predicted as non-disruptive.
- False Alarms (FAs): the fractions of safe discharges predicted as disruptive.

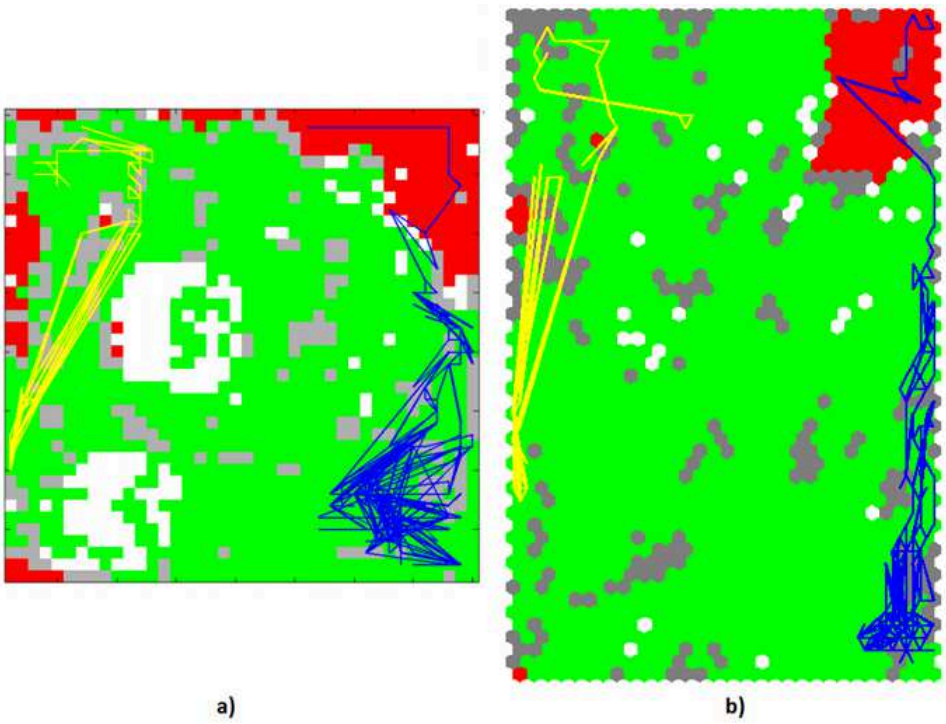


Figure 9.3: 2-D mappings of AUG operational space, a) GTM whit 1600 map units; b) SOM with 1674 map units; safe map units (green), disruptive map units (red), mixed map units (grey), empty map units (white). On both the maps the projection of a safe discharge (yellow line) and a disruptive discharge (blue line) on the GTM (a) and the SOM (b) is reported.

- Successful Rate (SR): the fraction of discharges (safe and disruptive) correctly predicted.

At AUG a disruption is considered correctly predicted if the prediction system is able to trigger the alarm 2ms before t_D. Two ms is the time needed to the mitigation systems to intervene [6]. Conversely, a safe discharges is correctly predicted when the alarm is not triggered at all. One of the mail goals of experimental reactors, as AUG, is to exploit its own potentialities. A conservative disruption predictor could limit the exploration capability of the machine; in order to avoid this drawback, the percentage of disruption triggered too much in advance has to be limited as well as the false alarms. To this purpose, the Early Detection (ED) index has been defined as the fraction of disruptive discharges which triggers the alarm too much in advance. In this study, a disruption is considered predicted too much in advance if the alarm is activated within the time window $[t_{FLAT-TOP}, t_D - 500]$ ms [7], where $t_{FLAT-TOP}$ is the flat top beginning time of plasma current.

9.5 SOM and GTM

In order to employ the two mappings as disruption predictors, a suitable alarm criteria which links the disruption risk of the different regions to the percentage of disruptive samples (DS%) into the map unit, has been optimized. In particular, the alarm is triggered when the trajectory stays in a disruptive or a mixed map unit containing at least 95% of disruptive samples for at least h consecutive samples. For each cluster the parameter h is evaluated by means of 9.1 for the GTM and 9.2 for the SOM.

$$h_{GTM} = round(-5.6 \cdot \mathrm{DS}\% + 562) \tag{9.1}$$

$$h_{SOM} = round(-3.2 \cdot \mathrm{DS}\% + 322) \tag{9.2}$$

where *round()* is the nearest integer function.
The coefficients of these linear functions have been optimized maximizing the Successful Rate (SR) on the training set. Table 9.3 reports the prediction performance for SOM and GTM on the three data sets simulating the on-line operation.
Table 9.3 shows that the SR obtained with GTM results to be better than the ones with the SOM. Moreover, the GTM has always better performance on early detections and false alarms than the SOM, instead the SOM achieves always lower MAs than the GTM.

Method	Data set	Disruptive discharges			Safe discharges			
		TD[%]	MA[%]	SP[%]	SP[%]	FA[%]	SR[%]	ED[%]
SOM	Training	1.61	5.65	92.74	94.52	5.48	94.01	16.93
	Test_1	4.76	6.35	88.89	89.60	10.32	89.50	17.46
	Test_2	0	1.83	98.17	84.13	15.87	88.16	16.51
GTM	Training	0	8.26	91.74	97.42	2.58	95.82	9.1
	Test_1	0	11.67	88.33	91.61	8.39	90.70	10
	Test_2	0	3.77	96.23	89.67	10.33	91.51	12.26

Table 9.3: Prediction performance for SOM and GTM on the three data sets.

9.6 Logit model

Besides SOM and GTM, a Logit model has been trained to predict, starting from the seven variables listed in table 9.2, the probability of a generic sample to belong to a safe or a disruptive phase. During the training of the model, the dichotomous output has been set equal to 0 for safe samples and 1 for disruptive samples. Thus, the Logit model realizes a mapping from the input variables to a continuous output, which should assume large negative values for samples belonging to safe phase and large positive values for those belonging to disruptive phase. From a preliminary analysis, it has been observed that, for the majority of safe discharges belonging to the training phase, the Logit model output (LMO) is always smaller than 0 throughout the discharge. Conversely, for the great majority of the of disrupted discharges the time evolution of the LMO remains at low values for the majority of the discharge and begins to grow when the pulse approaches the disruption time. As an example, Fig. 9.4 reports the LMO for a disruptive (No. 21886) and a safe (No. 21718) discharge.
This behavior suggests us to use the logit model as disruption predictor by introducing a suitable threshold value that discriminates between the safe and the disruptive phase. Figure 9.5 reports the probability density function of LMO for samples belonging to the training set. It can be seen that for the great majority of safe samples belonging both to safe and disruptive discharges (blue and green line respectively), LMO remains smaller than 0. Conversely, for the majority of disrupted samples, LMO is greater than zero. Figure 9.5 shows that an LMO value can be set as alarm threshold to discriminate between safe and disruptive phase. Thus, the adopted alarm criteria consists in triggering a disruption alarm when the LMO exceeds the threshold value. To avoid false alarms sometimes caused by spikes in the diagnostic signals, a time delay has been introduced that inhibits the alarm

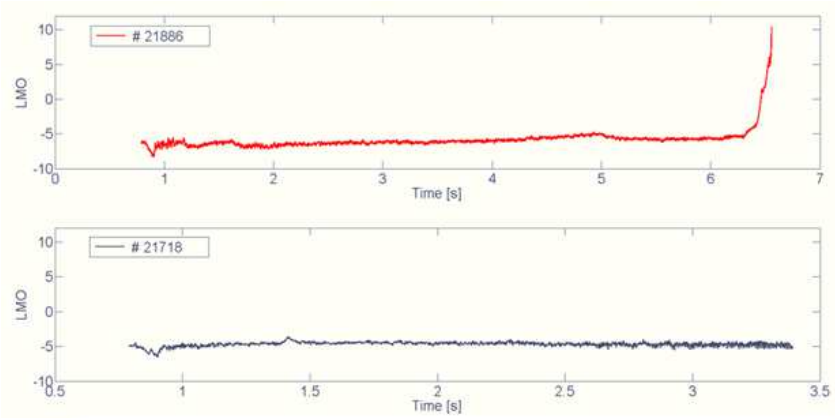


Figure 9.4: Logit model output (LMO) for a disruptive discharge (upper plot) and a safe discharge (lower plot).

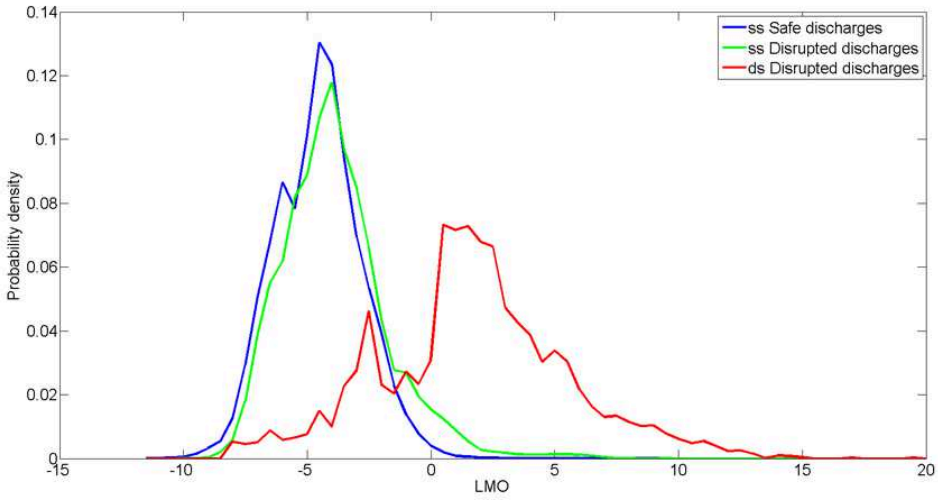


Figure 9.5: Probability density of LMO for samples belonging to the training set. Three sample groups are represented: *ss* of safe discharges (blue), *ss* of disruptive discharges (green) and *ds* of disruptive discharges (red)

for h samples after the alarm is activated. The optimum threshold value has been optimized in the range $[1 \div 3]$ maximizing the Successful Rate on the training set. The parameter h has been optimized in the range $[1 \div 10]$. The best performance of the Logit model as disruption predictor has been achieved with an alarm threshold equal to 2.3 and $h = 2$ (see Table 9.4). It can be noted that the SRs and FAs obtained with the Logit model are always slightly worse than those achieved with the SOM and the GTM, but among the three methods, the Logit model reaches the best performance on early detections.

Method	**Data set**	**Disruptive discharges**			**Safe discharges**			
		TD[%]	MA[%]	SP[%]	SP[%]	FA[%]	SR[%]	ED[%]
Logit	Training	0.80	8.87	90.32	94.51	5.48	93.32	4.84
	Test_1	1.61	11.29	87.09	89.68	10.32	88.94	4.84
	Test_2	0.90	0.00	99.10	82.28	17.71	86.95	10.8

Table 9.4: Prediction performance for the Logit model on the three data sets.

9.7 Combined predictors

Comparing tables 9.3 and 9.4, the three methods achieve comparable values of SRs, but no one method can be stated as the most suitable. In fact, the Logit regressor maintains limited the early detections, always lower than 10.8%, but false alarms overcome 17% (on Test_3); the GTM presents the best performance on safe discharges, with FA always lower than 10% , but EDs are higher than 9% on the three data set; the SOM reaches the worst results on EDs, always higher than 16%, and FAs are higher than 10% on Test_2 and Test_3. Thus, SOM and GTM have poor results on early detections where the Logit model achieves good results. On the contrary the Logit model has poor results on false alarms where the GTM obtains good results.
The complementary behavior previously pointed out, suggest to combine both the SOM and the GTM with the Logit regressor in order to realize two combined predictors able to maintain limited early detections and false alarms as well as maximizing the SRs. In the combined predictors the alarm is triggered only when the trajectory evolving on the map (either SOM or GTM) stays in disrupted or mixed map units containing at least 90% of disruptive samples for at least h consecutive samples and the LMO is higher than a suitable alarm threshold. The parameters h and the alarm thresh-

old have been optimized maximizing the SR on the training set. The alarm threshold has been optimized in the range $[1 \div 3]$. The predictor consisting in the combination of the SOM and the Logit model achieves the best performance (see Table 9.5) with an alarm threshold equal to 1.8 and h is evaluated in each cluster by means of the function $h_{SOM} = round(-2.6 \cdot \mathrm{DS}\% + 262)$, where DS% is the percentage of disruptive sample in the cluster.

The predictor consisting in the combination of the GTM and the Logit model achieves the best performance, reported in Table 9.5 too, with an LMO threshold equal to 1.7 and h is evaluated in each cluster by means the function $h_{GTM} = round(1.3 \cdot \mathrm{DS}\% + 132)$.

Method	**Data set**	**Disruptive discharges**			**Safe discharges**			
		TD[%]	MA[%]	SP[%]	SP[%]	FA[%]	SR[%]	ED[%]
SOM	Training	1.65	8.26	90.08	98.39	1.61	96.06	5.79
&	Test_1	6.56	8.20	85.25	93.55	6.45	91.20	3.28
LOGIT	Test_2	0.00	1.85	98.15	91.88	8.12	93.67	6.48
GTM	Training	0.00	8.33	91.67	97.74	2.26	96.05	5.00
&	Test_1	3.23	11.29	85.48	92.26	7.74	90.32	4.84
LOGIT	Test_2	0.00	2.80	97.20	89.67	10.33	91.80	7.48

Table 9.5: Prediction performance for the combined predictors on the three data set.

Table 9.5 shows that, with respect to SOM and GTM, the corresponding combined predictors accomplish slightly better SRs, but it has to be highlighted that EDs and FAs significantly improve. In particular, combining the Logit model with the SOM allows both early detections and false alarms to fall down 7%, instead MAs slightly deteriorate on Trainig set and Test_2. Combining the Logit model with the GTM allows the early detections to remain below 8%, false alarms have been reduced only for the training set and Test_1, no enhancement are attained for Test_2. Regarding MAs, no considerable improvements are achieved on the three data set.

9.8 Disruption classification

A preliminary analysis shows that during the disruptive phase different types of disruptions evolve in different disruptive map regions. As an example, figure 9.6 reports the time evolution of a density limit disruption (black trajectory) and a beta limit disruption (blue trajectory) on both GTM and SOM. The black point represents the trajectory starting point, the yellow point is

the trajectory ending point. As it is shown by the yellow squares, on both maps, the density limit disruption ends in the disruptive region on the right upper corner, instead the beta limit ends in the small disruptive region on the left map side. This means that on both SOM and GTM, disruptions happening at low thermal energy (density limit) end in a disruptive region different from those happening at high thermal energy (beta limit).

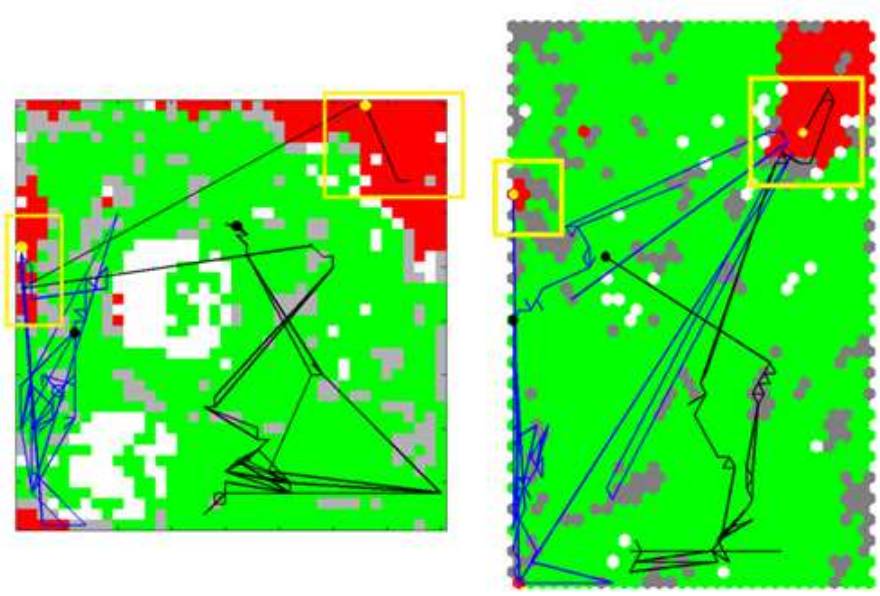


Figure 9.6: Time evolution of two disruption types on GTM and SOM; black trajectory: density limit disruption (No. 28727); blue trajectory: beta limit disruption (No. 25172).

This preliminary study shows that among the disruptive regions, areas depicting the behavior of a certain disruption class could be identified. Therefore, following the trajectory on the map, it is possible to eventually recognize non only the proximity of disruption but also the class it belongs to. These results confirm the potentiality of SOM and GTM as disruption classifier too, as well as it is shown in JET database in [5].

9.9 Conclusions

This study shows that it is possible to describe the 7-D AUG operational space on a 2-D map (SOM and GTM), where regions with different risk of disruption can be identified. A criterion has been optimized to associate the risk of disruption of each map region with a disruption alarm threshold. The

prediction performance of the proposed predictive systems has been evaluated on a test set of discharges coming from experimental campaigns carried out at ASDEX Upgrade from May 2011 to November 2012.
The GTM predictor results to have better Success Rate than the SOM predictor on both the Test sets, reaching on average $\sim 91\%$. Furthermore, GTM has always better performance on Early Detections and False Alarms than SOM, although on the test sets both indexes remain above 8%. Conversely, the SOM achieves better performance on Missed Alarms than the GTM.
Additionally a statistic predictor has been trained and tested on the same data set. This predictor, based on a Logistic Regressor model, achieves slightly worse performance than SOM and GTM, except on early detections, which are much lower than those achieved by SOM and GTM.
Finally, the good performance of SOM and GTM and the tendency of Logit model to limit the early detections have been exploited combining each map with the statistical model. A new alarm criterion has been optimized, in particular, the alarm triggered on the map (either SOM or GTM) is activated only if Logit Model Output is greater than a certain threshold.
An overall improvement of the performance has been obtained both for SOM and GTM. The new predictor involving the SOM gets the best performances, on test sets, the SR exceeds 91%, FAs remain lower than 8.2% and EDs reach at most 6.5%.

Bibliography

[1] Aledda R. *et al.* 2012 IEEE Transactions On Plasma Science, Vol. 40, No. 3.

[2] Cannas B., Fanni A., Murari A., Pau A., Sias G., and JET EFDA Contributors 2013 *Manifold learning to interpret JET high-dimensional operational space*, Plasma Phys. Control. Fusion 55 art. no 045006

[3] Cannas B. *et al.* 2010 Nuclear Fusion 50 075004

[4] SOMtoolbox 2005 Adaptive Informatics Research Centre, Helsinky Univ. of Technology, Finland. http://www.cis.hut.fi/projects/somtoolbox

[5] Cannas B., Fanni A., Murari A., Pau A., Sias G. and JET EFDA Contributors 2013 *Automatic disruption classification based on manifold learning for real-time applications on JET* Nucl. Fusion 53 093023

[6] Pautasso G., *et al.* 2009 Disruptions studies in ASDEX Upgrade in view of ITER Plasma Phys. Control Fusion, vol. 51, no. 12, p. 124 056

[7] Zhang Y., Pautasso G., Kardaun O., Tardini G., Zhang X.D. and the ASDEX Upgrade Team 2011 Prediction of ASDEX Upgrade disruptions using discriminant analysis, Nucl. Fusion 51 063039-41

Conclusions and future work

The activities carried out in the framework of this thesis regarded the development, the implementation and the application of algorithms for classification and prediction of disruptions in Tokamaks.

Disruptions can expose the plasma facing components to severe thermo-mechanical stresses and conductors surrounding the vessel to huge electromagnetic forces; therefore, it becomes of primary importance to avoid or mitigate disruptions in order to preserve the integrity of the machine. This aspect turns out to be particularly relevant in design and running of new experimental devices as ITER.

These considerations motivate and still motivate a strong interest in developing methods and techniques aimed to minimize both number and severity of disruptions. But, besides the prediction, it is particularly important to be able to distinguish among their different types in order to improve avoidance and mitigation strategies. Since physical models able to reliably recognize and predict the occurrence of disruptions are currently not available, machine learning techniques have been exploited as an alternative approach to disruption prediction and automatic classification, both with the application and further development of existing systems and with the investigation of new approaches.

One of the first problems which have to be addressed when working with data-based methods is the construction of a reliable and representative database. This is true especially in fusion, where the character of high dimensionality and the huge amount of available observations, poses a serious problem about how to "reduce" coherently available data. Therefore, proper criteria have been used to select suitable signals downloaded from JET databases in order to obtain a data set of reliable observations. Finally, data-reduction, based on clustering methods, has been performed to select a limited and representative number of samples for the operational space mapping. Two separated databases have been built with discharges belonging to the *Carbon Wall* (CW) configuration (campaigns from 2005 to 2009) and to the new *ITER-like Wall* (ILW) configuration (campaigns from 2011

to 2013). The distinction allowed to analyze what is changed moving from a configuration to the other one in terms of the underlying physics and operational space, and this is reflected in the different behaviour of disruptions, coherently to what has been found with all the analysis carried out.

One of the most important part of the work regarded the mapping of the high dimensional operational space of JET, which has been described and visualized using different linear projection methods such as Grand Tour (GT) and Principal Component Analysis (PCA), and mapped through non-linear manifold learning techniques as Self-Organizing Map (SOM) and Generative Topographic Map (GTM). The potentiality of manifold learning methods has been discussed showing several types of representations, also with reference to the data analysis and visualization tools developed for GTM. The power of the proposed techniques has also been highlighted through a comparison with classical scatter plots identifying operational limits and boundaries for the considered database.

In particular, both SOM and GTM maps can be exploited to identify characteristic regions of the plasma scenario and for discriminating between regions with high risk of disruption and those with low risk of disruption, to quantify and evaluate the effectiveness of the mapping itself. In addition, some measures have been used to evaluate the performance of the proposed methodologies. To calculate the precision of the clustering over the entire dataset the average quantization error, a common index of the map resolution, has been applied. Furthermore, to control the conservation of topology two different aspects have been analyzed, i.e., the trustworthiness of the projected neighborhood and the preservation of the resulting neighborhood. Moreover, an outlier analysis has been performed on the available data in order to quantify goodness and effectiveness of the projection.

Regarding the results achieved with the mapping, both the SOM and the GTM presents a large safe region well separated from some disruptive regions by transition regions, which consists of map units that contains both safe and disruptive samples, and empty regions. In particular GTM model turned out to have both higher capability of discriminate between safe and disruptive samples (less than 10% of the samples are projected in transition regions) and better performance in the mapping.

Given the results in the mapping of JET 10-dimensional space, SOM and GTM potentialities have been extensively investigated and an algorithm for automatic classification has been developed for both the methods. The proposed approach for the discrimination of disruption types consists of identifying characteristic regions in the operational space where the plasma undergoes a disruption.

A relevant part of the activities carried out in the framework of this thesis

has been spent in the analysis of the different types of disruptions that can occur in JET, making reference to the manual classification made available by physicists, where specific chains of events have been detected and used to classify disruptions, grouping those that follow specific paths.
The characterization of the operational space in terms of the different disruption classes may lead to better overall understanding and more focussed prevention and mitigation methods. The maps obtained with SOM and GTM algorithms have shown to self-organize in such a way that the disruptions which belong to the same class tend to aggregate, defining in this way regions where a certain class results to be predominant with respect to the others. Each disruption has been projected on the maps, and the probabilities of belonging to the different disruption classes have been monitored during the time evolution, returning, among the seven considered classes, the one which the disrupted pulse more likely belongs to. In order to perform the classification, a majority voting algorithm has been applied the the class-membership, computed for each shot. The algorithm associates to each sample a probability to belong to the seven classes. The success rate of GTM is high for all the considered classes (above 97%), reaching in some cases even the percentage of 100%.

After the campaign in 2009, JET installed the new ITER Like wall (ILW). The first attempt has been to project the disruptions of the ILW campaigns onto the GTM map trained with the CW discharges; the classification performance for the new disruptions significantly deteriorated for certain classes. Thus, the disruption classes with the ILW have been deeply analyzed and compared with those in the CW JET campaigns. In particular, the probability density functions of the different plasma parameters clearly highlighted the different behavior of a new class, an impurity type due to high-Z impurity accumulation in the core of the plasma column. Instead, in the considered period, some disruption classes present in the CW data bases are no longer present in the ILW data base, as those due to too strong internal transport barrier and the ones due to Greenwald limit.
The performance of the new ILW GTM classifier has been tested in conjunction with APODIS, a prediction system working on-line at JET, simulating the application in real time, that is, by synchronizing a time windows of 32 and 64 ms in advance with respect to the time in which APODIS triggers the alarm. By applying the majority voting algorithm to the class membership in the considered time window, the prediction success rate is still quite high (above 90%) according to the manual classification. The performance slightly worsened when the new impurity type is introduced, because in certain cases it turns out to be quite difficult to distinguish this new class from the previously defined impurity control problem class.

Furthermore, in order to validate and analyze the obtained results, another reference classifier has been developed based on k-NN which uses as kernel the Mahalanobis distance. The performance of the reference classifier is still above 90%, but, also for it, the success rate deteriorates when the new IMC class is introduced.

Finally, in order to verify the reliability of the performed classification, a conformal predictor has been developed, which is based on non-conformity measures. The preliminary results indicate the suitability of the conformal predictors to assess the reliability of the GTM classification even if the calculation time allows their use only in an off-line fashion.

GTMs potentiality has also been exploited for the prediction of disruptions at ASDEX Upgrade: a 2-D GTM has been built to represent the 7D AUG operational space on the base of discharges performed between May 2007 and April 2011. As it has been obtained in the case of JET, the GTM clearly proves to be able to separate non-disruptive and disruptive states of plasma. Therefore, likewise the SOM, the GTM can be used as a disruption predictor by tracking the temporal sequence of the samples on the map, depicting the movement of the operating point during a discharge. Various criteria have been studied to associate the risk of disruption of each map region with a disruption alarm threshold. The prediction performance of the proposed predictive system has been evaluated on a set of discharges coming from experimental campaigns carried out at AUG from May 2011 to November 2012.

The GTM predictor achieves the best overall performance, above the 91% on the considered Test sets. Furthermore, GTM has always better performance on Early Detections and False Alarms than SOM, although, on the test sets both indexes remain above 8%. Conversely, the SOM achieves better performance on Missed Alarms than the GTM.

Additionally, a statistic predictor has been trained and tested on the same data set. This predictor, based on a Logistic Regressor model, achieves slightly worse performance than SOM and GTM, except on early detections, which are much lower than those achieved by SOM and GTM.

Therefore, the good performance of SOM and GTM and the tendency of Logit model to limit the early detections have been exploited combining each map with the statistical model. A new alarm criterion has been optimized, in particular, the alarm triggered on the map (either SOM or GTM) is activated only if the Logit model output is greater than a certain threshold, achieving an overall improvement of the performance. The combined predictor involving the SOM gets the best performances, on test sets, the Success Rate exceeds 91%, False Alarms remain lower than 8.2% and Early Detections reach at most 6.5%.

The high performance of the proposed methods gives rise to the perspective of a deployment of these tools in real time: regarding this point, even if a porting of the Matlab codes should be needed for the integration in the real time frameworks, the suitability for real time applications has been already assessed. Furthermore, the algorithms described in this work have been developed in a tool for Matlab which allows, given a database, to perform all the analysis presented in this thesis almost automatically, from the data-reduction, going through the mapping of operational spaces up to the projection of new data and the assessment of mapping performance.
This techniques represent a powerful tool for data-analysis and could be very useful not only in the framework of disruption prediction and classification, but also in other fields, such as, for example, one would like to distinguish or discriminate a particular behavior or plasma state. To conclude, still regarding the future perspectives, machine learning tools are also producing very interesting results in the comparative analysis of different fusion devices operational spaces, as in the case of JET and ASDEX Upgrade, on the route of developing predictors capable of extrapolating from one device to another, as foreseen in the framework of a cross-machine approach.

List of Tables

List of Figures

Acknowledgements

During these three years of research work a lot of things have happened, not always particularly positive, but I'm really glad to have persevered in what I was doing, without being discouraged in the moments of difficulty. But what I'd really like to do at this stage, is to express my gratitude to all the people who have contributed and made this Thesis possible.

First of all, I would like to thank Prof. **Alessandra Fanni**, who has been my supervisor here in Cagliari and has given me the opportunity to perform this challenging work under her supervision. This thesis without her outstading support and her personal and professional advice would not have been possible. I have to say the same things for Dr. **Barbara Cannas** and **Giuliana Sias**, who have made me really appreciate my research work, giving me their immeasurable support during the whole time. The University of Cagliari has represented a really stimulating environment where I have grown as person and as researcher.

I would also thank all the people in Consorzio RFX, starting from my supervisor Prof. **Paolo Bettini**, who have been always available and, even from afar, have supported me during this three years.

A great part of this Thesis have been made at JET laboratories, in the Culham Science Center, where I have found an extraordinary scientific and human environment: I would like to thank in particular way Dr. **Andrea Murari** for his constant support and supervision, and for having invested a lot of time in introducing me into the fascinating world of fusion. I would also like to express my gratitude to all the people who helped me improving knowledges and better understand the basics of the object of my research, in particular Dr. **Gabriella Pautasso** in IPP and **Peter de Vries** in JET. All the discussions with them have contributed in fundamental way to my research work.

I would like to thank also my office mates here in Cagliari **Fabio**, **Raffaele**, **Sara** and **Giuditta** for having bearing me especially during the periods of hard work, and all my collegues in Padova, **Greg**, **Giovanni**, **Tullio** and **Winder**, for having helped me every time I needed and I was here in

Cagliari.

I would like to express all my gratitude to my brother **Marco** for having always supported me and, not negligible thing, for his precious support with LaTeX. I don't have the words to thank my parents **Antonino** and **Maria Chiara** for everything, but in particular for their encouragement about the path I have chosen to follow. Lastly, I would like to thank and to dedicate everything of these three years to my partner **Stefania**, for having always made me feel her closeness and love, especially all the time I spent abroad without her. Even if I wanted I could not find the right words to express how much you mean for me.

Alessandro Pau