34th EPS Conference on Plasma Phys. Warsaw, 2 - 6 July 2007 ECA Vol.**31F**, P-5.144 (2007)

Mapping of the ASDEX Upgrade Operational Space using Clustering Techniques

M. Camplani[1], B. Cannas[1], A. Fanni[1], G. Pautasso[2], G. Sias[1], P. Sonato[3], M. Zedda[1] and the Asdex-Upgrade Team[2]

[1]*Electrical and Electronic Engineering Dept. - University of Cagliari, Italy*

[2] *Max-Planck-Institut fur Plasmaphysik, EURATOM Association, Garching - Germany*

[3] *Consorzio RFX, Associazione EURATOM-ENEA sulla Fusione, Padova - Italy*

Abstract -Two clustering techniques are used to map ASDEX Upgrade high-dimensional operational space. The produced mapping exploits the similarities of the data by grouping similar data items together, allowing to identify characteristic regions for plasma scenarios. In particular, the 2D SOM projects high-dimensional plasma states to a 2D map that can be easily visualised and understood. The database consists of 7 plasma parameters from 229 shots carried out during the operation of ASDEX between June 2002 and July 2004.

I. INTRODUCTION

The range of plasma states accessible to a tokamak is highly restricted by disruptive events. This is firstly because disruptions limit the range of operation in current and density, and secondly because their occurrence leads to a large mechanical stress and intense heat loads. In the past, empirical explorations of the operational boundaries were performed in order to determine the disruption-free operational space [1] [2]. In this paper, an alternative approach, based on clustering techniques, is proposed to map the ASDEX Upgrade operational space, which allows to use more than two plasma parameters at a time. In fact, when the correct association between input data and class membership is not available, the identification of distinct classes from features can be approached with a clustering procedure. The clustering process subdivides the input space in subsets, each assembling data that present common features. In this case, only the information embedded in the input data can be used to identify some aggregation or properties of the input patterns. In this paper, traditional K-means and Self-Organising Maps algorithms have been used [3]. The database used in this work is composed of 149 flat-top disruptive shots and 80 safe shots occurred during the experimental campaigns performed at ASDEX Upgrade between June 2002 and July 2004 (shot range 16000-19000). Seven plasma parameters have been selected to describe the plasma regime during the current flat-top: safety factor; total input power; the ratio between radiated power and the input power; internal inductivity; poloidal beta; locked mode signal; electron density.

II. DATA CLUSTERING

Let us consider an *n*-dimensional input space *X* and *N* samples $\bar{x}_i \in X$, $i = 1,2,\ldots,N$, which have to be partitioned in *K* clusters. *K-means* clustering is the most popular and simply hard clustering algorithm. It assigns each point $\bar{x}_i$ to the cluster whose center is nearest in terms of Euclidean distance. That center is the average of all the points in the cluster and it is associated to an *n*-dimensional prototype vector $\bar{m}_j$. At the first iteration the prototype vectors are randomly chosen. Then, each point is assigned to the nearest prototype vector (the winner $\bar{m}_w$) and the new prototype vector is recomputed. The previous steps are repeated until a prefixed convergence criterion is met (usually, the algorithm stops when there is no further change in the cluster composition).

Self Organizing Map (SOM) defines a mapping from the *n*-dimensional input space X onto a regular (usually 2D) array of neurons (corresponding to the clusters), preserving the topological properties of the input. This means that points close to each other in the input space are mapped to the same or neighbouring neurons in the output space. The SOM is iteratively trained using competitive learning. During the SOM learning, a weight protype vector with the same dimensionality as the input space is associated to each neuron. At each training step, a pattern is randomly chosen from the input data set and its Euclidean distance to all the prototype vectors is computed. The neuron with prototype vector closest to the selected input pattern is called the Best Matching Unit (BMU). The weights of the BMU and of the neurons close to it in the SOM are adjusted towards the chosen input.

In order to know whether a clustering technique has properly adapted itself to the training data, several quality mapping indexes can be evaluated. A common measure used to calculate the precision of the mapping is the average quantization error E_q over the entire data set: $E_q = \frac{1}{N}\sum_{i=1}^{N}\|\bar{x}_i - \bar{m}_w\|$. Moreover, it is possible to evaluate how well the SOM preserves the data set topology evaluating the topographic error E_t: $E_t = \frac{1}{N}\sum_{i=1}^{N}u(\bar{x}_i)$ where $u(\bar{x}_i) = 1$ if the first and second winner prototype vectors of $\bar{x}_i$ are not next to each other, otherwise $u(\bar{x}_i) = 0$. E_t represents the proportion of all data vectors for which first and second BMUs are not adjacent units.

III. THE OPERATIONAL SPACE MAPPING

In order to reduce the dimensionality of the ASDEX-Upgrade data set, a preliminary clustering of the samples is performed using SOM and K-means algorithms. Each cluster is supposed to contain samples belonging to similar plasma states. This procedure allows one to automatically select a limited and representative number of samples for the subsequent operational space mapping.

For a disruptive shot, let us define as *precursor time* (t_{PREC}) the time instant that discriminates between safe and pre-disruptive phases. On the basis of previous experiences on disruption prediction at ASDEX Upgrade, t_{PREC} has been set equal to 45ms [4]. The samples belonging to the interval [$0 \div t_{PREC}$] are considered *safe samples* whereas those corresponding to the subsequent time instants until the *disruption time* t_D are considered *disruptive samples*. Following this classification, four kinds of clusters have been identified: disruptive clusters (DCs) containing disruptive samples, safe clusters (SCs) containing safe samples, transition clusters (TCs) containing both disruptive and safe samples, and empty clusters (ECs) containing no samples.

IV. RESULTS AND CONCLUSIONS

The distribution of the 1196 clusters resulting from the clustering procedure, is shown in Table 1. Note that all the 149 disruptive shots and 80 safe shots are mapped in the same space.

	DCs(%)	SCs(%)	TCs(%)	ECs(%)
SOM	2.25 (27/1196)	73.24 (876/1196)	19.90 (238/1196)	4.61 (55/1196)
K-means	3.43(41/1196)	82.78 (990/1196)	13.71 (164/1196)	0.08 (1/1196)

Table 1: Percentage of disruptive (DCs), safe (SCs), transition (TCs) and empty (ECs) clusters

As expected, the majority of clusters is constituted by safe samples. K-means mapping results in a smaller number of transition clusters. In particular, in the SOM, 31.09% (#74) of TCs contains samples belonging to safe shots. Conversely, only 13.88% (#23) of K-means TCs contains samples belonging to safe shots. Thus, the K-means clustering better discriminates safe regions from disruptive regions of the shots.

Table 2 reports the mapping precision E_q for both clustering techniques and the SOM topographic error E_t.

	E_q	E_t
SOM	0.10	0.05
K-means	0.03	-

Table 2: Mapping quality measures

As can be noted, K-means mapping is characterized by a better precision, i.e., by a smaller average distance between each data vector and its prototype. The SOM topographic error highlights that the input topology is preserved in the 2D-map for 95% of samples.

Figure 1(a) shows the 2D SOM. The blue clusters contain the samples belonging only to the safe pulses, the red and green clusters contain safe and disruptive samples of disruptive shots respectively and the white clusters are empty. For each cluster the color density is proportional to the number of samples contained within the cluster. The map clearly evidences the presence of a disruptive region (green) and of a safe region composed by samples from safe shots (blue) and *safe* samples from the disruptive shots (red) that are at the boundary of the two regions. Figure 1(b) shows the map of the D-Matrix. The Median Distance-Matrix (D-Matrix) visualization method plots the median distance of the prototype vectors of each cluster to its neighbours. Thus, the D-Matrix allows one to display the similarity of data elements into clusters with respect to the data into nearest clusters. The growing gray level indicates the growing of the median distance from each map unit to its neighbours. Macro regions of similar plasma states are constituted by the lighter clusters and they can be demarcated by darker clusters. As can be noted by the SOM, the region on the right top corner holds the majority of disruptive samples. The D-matrix map shows that this disruptive region is characterized by plasma configurations different from those of the samples on the nearest clusters.

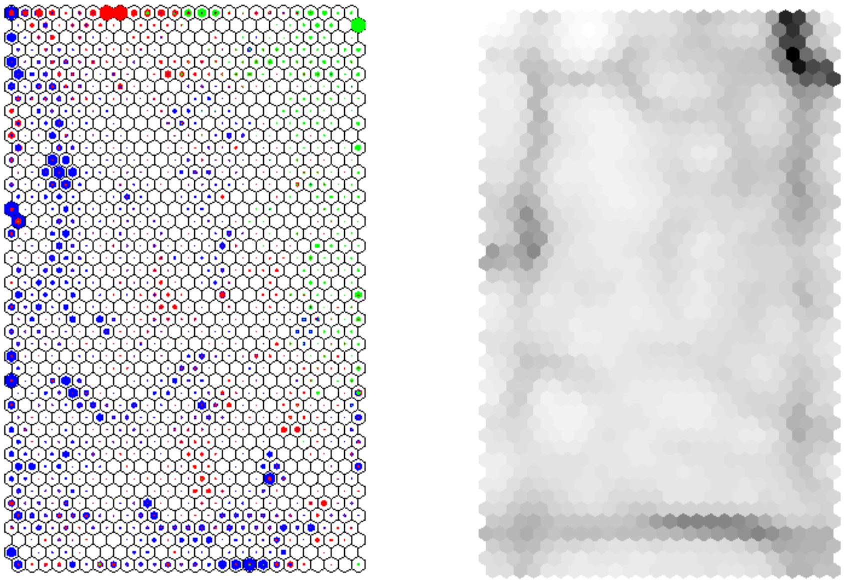

Figure1: (a) SOM; (b) D-matrix map

VI. REFERENCES

[1] M. Murakami,et al, "Some observation on maximum Density in Tokamaks Experiments" Nucl. Fusion 16 347-348, 1976.
[2] S.J.Fielding et al, "High-Density discharges with gettered Torus walls in DITE" Nucl. Fusion 17 1382-1385, 1977.
[3] Duda R.O., Hart P.E. Stork D.G., "Pattern classification", Prentice-Hall, 1999.
[4] G.Sias, "A Disruption Prediction System for ASDEX-Upgrade Based On Neural Networks", Ph.D. Dissertation, University of Padova, 2007.

"Acknowledgement This work was supported by the Euratom Communities under the contract of Association between EURATOM/ENEA. The views and opinions expressed herein do not necessarily reflect those of the European Commission "