EFDA–JET–CP(13)03/27

A. Shabbir, G. Verdoolaege, G.Van Oost, J.-M. Noterdaeme
and JET EFDA contributors

Visualization of Tokamak Operational Spaces Through the Projection of Data Probability Distributions

Visualization of Tokamak Operational Spaces Through the Projection of Data Probability Distributions

A. Shabbir[1], G. Verdoolaege[1], G.Van Oost[1], J.-M. Noterdaeme[1,2]
and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]Department of Applied Physics, Ghent University, B-9000 Ghent, Belgium
[2]Max-Planck-Institut für Plasmaphysik, EURATOM-Assoziation, D-85748 Garching, Germany
** See annex of F. Romanelli et al, "Overview of JET Results",*
(24th IAEA Fusion Energy Conference, San Diego, USA (2012)).

Preprint of Paper to be submitted for publication in Proceedings of the
40th EPS Conference on Plasma Physics, Espoo, Finland.
1st July 2013 – 5th July 2013

ABSTRACT.
Information visualization is becoming an increasingly important tool for making inferences from large and complex data sets describing tokamak operational spaces. Landmark MDS, a computationally efficient information visualization tool, well suited to the properties of fusion data, along with a comprehensive probabilistic data representation framework, is shown to provide a structured visual map of plasma confinement regimes, plasma disruption regions and plasma trajectories. This is aimed at contributing to the understanding of underlying physics of various plasma phenomena, while providing an intuitive tool for plasma monitoring.

1. INTRODUCTION

Exploration of the operational space of fusion devices is an essential activity in fusion research for establishing the conditions under which specific operational regimes and plasma instabilities develop. Visualization of machine operational spaces entails the representation of multidimensional diagnostic fusion data in a low-dimensional space, usually of two or three dimensions for facilitating human perception emerging from a visual description. This provides a natural and intuitive tool for plasma physicists and machine operators to perceive and interpret diagnostic information and plasma behaviour.

In this work, a comprehensive framework rooted in probability theory has been adopted for the representation of fusion data. Diagnostic data in nuclear fusion experiments is subject to considerable measurement uncertainty and a possible loss of valuable information may incur if the error bar associated with the measurement is regarded as a mere side effect. Our proposed framework uses the full probability distribution of plasma signals and hence includes all information that is contained in the signal.

Furthermore, a computationally efficient adaptation of multidimensional scaling (MDS), known as Landmark MDS (LMDS) [1], has been investigated and applied for a 2-D visualization of the operational space of fusion plasmas. MDS is a well-developed information visualization tool which yields a configuration of points in the Euclidean plane with minimal distortion of all pairwise distances between data points. The provided mapping is quantitative in terms of the proximity between the data points, which makes it advantageous over other visualization tools such as self-organising maps (SOMs) [2] and discriminant analysis. However, MDS suffers from a polynomial computational and memory complexity and hence when the data size increases, it becomes too computationally expensive, for all practical purposes. This problem is overcome by LMDS which reduces the complexity from polynomial to linear. It is therefore well suited for the problem at hand, i.e. visualization of massive and non-linearly correlated fusion data sets.

Finally, this paper presents the application of Landmark MDS based on a probabilistic framework for visualization of confinement regimes and plasma disruptions. Trajectories for disruptive plasmas can also be visualised for JET disruptive pulses.

2. GEOMETRIC-PROBABILISTIC FRAMEWORK

For effective data representation, we adopt the notion that the fundamental object ensuing from a measurement is, actually, a probability distribution [3]. The visualization tool presented in this work essentially requires a measure of similarity between data points, which in our framework are thus probability distributions. The field of information geometry offers a natural similarity measure between probability density functions (PDFs), which are interpreted as points on a Riemannian differentiable manifold [3]. The PDF parameters provide a coordinate system on the manifold and the geodesic distance (GD) between the probability distributions is an adequate and well suited similarity measure.

3. LANDMARK MDS

Landmark MDS [1], a marked improvement over MDS in terms of computational complexity, at the expense of a negligible reduction in accuracy, is presented as a suitable tool for the massive amounts of fusion data. LMDS has the following workflow:

- A set of n_L landmark points are randomly selected from the data. In our experiments, $n_L = 10$ is sufficient.
- Metric MDS is applied to the chosen nL landmark points and a k-dimensional embedding is obtained for the p-dimensional landmark points, where $k \leq p$.
- A distance-based triangulation (DBT) is then used for embedding the remaining data points in the k-dimensional subspace. DBT is a procedure through which a low-dimensional embedding is obtained by an affine linear transformation of the squared distances between the data and the landmark points.

4. INFORMATION VISUALIZATION

For visualization of confinement regimes, data from the International Tokamak Physics (ITPA) Global H-mode Confinement Database (ITPA database) is used, containing more than 10,000 validated measurements of various plasma and engineering parameters, during discharges in 19 tokamaks [4]. Our aim was to distinguish between low and high confinement regimes, or L– and H– mode. In our data representation framework, each measurement is regarded as a sample from a Gaussian probability distribution with the mean the measurement itself and standard deviation the error bar. MDS and LMDS based on the geodesic distance between the respective distributions are used for creating the visualizations shown in Figures 1a and 1b. It can be observed that the mapping obtained with LMDS (Fig. 1b), is only slightly less accurate when compared with the mapping obtained with MDS, which is considered as the reference map. Meanwhile the execution time for obtaining the given visualization was reduced by LMDS with a factor 300 compared to MDS. Furthermore, the mapping obtained without the proposed probabilistic representation of the data in conjunction with the GD, is much more informative, with little overlap between the classes, than using the measurement values only together with the Euclidean distance (Fig.1c). Clearly, the visualisation with the proposed methodology exhibits structure and provides a clear indication of the type of confinement regime.

In the second experiment, data acquired in disruptive shots at JET between campaign C21 and

C27 was used for obtaining visually informative plots for disruptions. Wavelet decomposition using 4 wavelet scales was carried out for 4 indicative signals, namely the plasma current, mode lock amplitude, electron density and total radiated power. The distribution of wavelet coefficients was described by a zero-mean Laplace distribution, again allowing a fast calculation of geodesic distances. Figure 2a shows the visualization obtained using LMDS with 10 landmarks points for disruptive and non-disruptive data points. Non-disruptive points were obtained from the indicator signals at 2 to 1 seconds before disruption, whereas the data points obtained from 210 to 30 milliseconds preceding disruption constitute the disruptive points. Figure 2b provides a deeper, quantitative insight into the distribution of disruptive and non-disruptive plasma states by marking the density contours for the points in each cluster. The maximum density of points for each cluster is taken as the reference level 1, and the contours are defined with respect to the respective maximum level. Finally, a trajectory of Pulse No: 78015, a JET pulse that disrupted at 16.32 seconds due to the onset of a neoclassical tearing mode, is mapped. Figure 2b follows the trajectory from about 1.5 seconds preceding the discharge.

CONCLUSIONS

This work provides a visualization mechanism, based on a probabilistic data representation framework and applies it to plasma confinement regimes and plasma disruptions. The developed tool allows for enhanced understanding of underlying physics alongside facilitating plasma control and monitoring.

ACKNOWLEDGEMENTS

This work was supported by EURATOM and carried out within the framework of the European Fusion Development Agreement. Additional support by the European Commission is acknowledged within the framework of the Erasmus Mundus International Doctoral College in Fusion Science and Engineering. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

REFERENCES

[1]. V.de Silva and J.B.Tenenbaum, Geometric Algorithms Seminar, Stanford University, Stanford, CA, 2004.

[2]. R. Aledda et al., IEEE Transactions on Plasma Sciences, **40**, pp. 570–576, 2012.

[3]. G. Verdoolaege et al., Plasma Physics and Controlled Fusion, **54**, art. no. 124006 (6 pp.), 2012.

[4]. The International Global H Mode Confinement Database, see http://efdasql.ipp.mpg.de/hmodepublic.

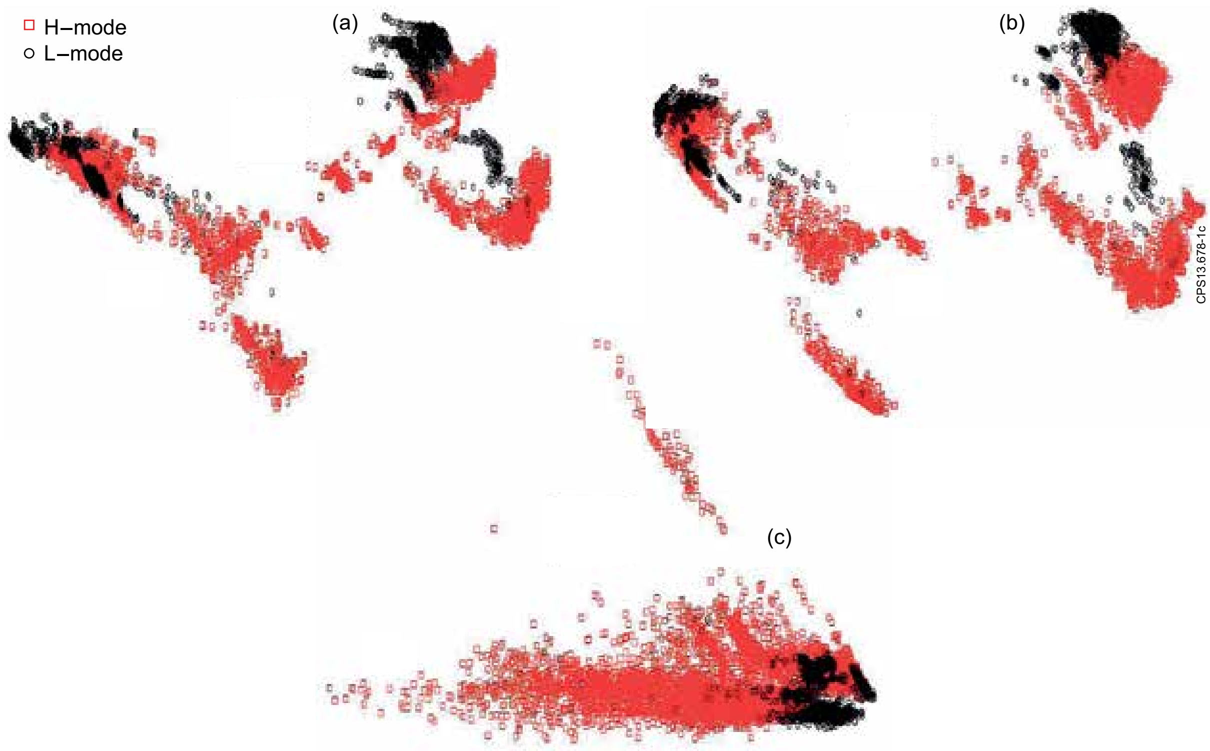

Figure 1: Two-dimensional projections of the ITPA database, indicating L- and H-mode clusters. (a) Using MDS and geodesic distance. (b) Using LMDS with 10 landmark points and geodesic distance. (c) Using LMDS with 10 landmark points and Euclidean distance without measurement uncertainty.

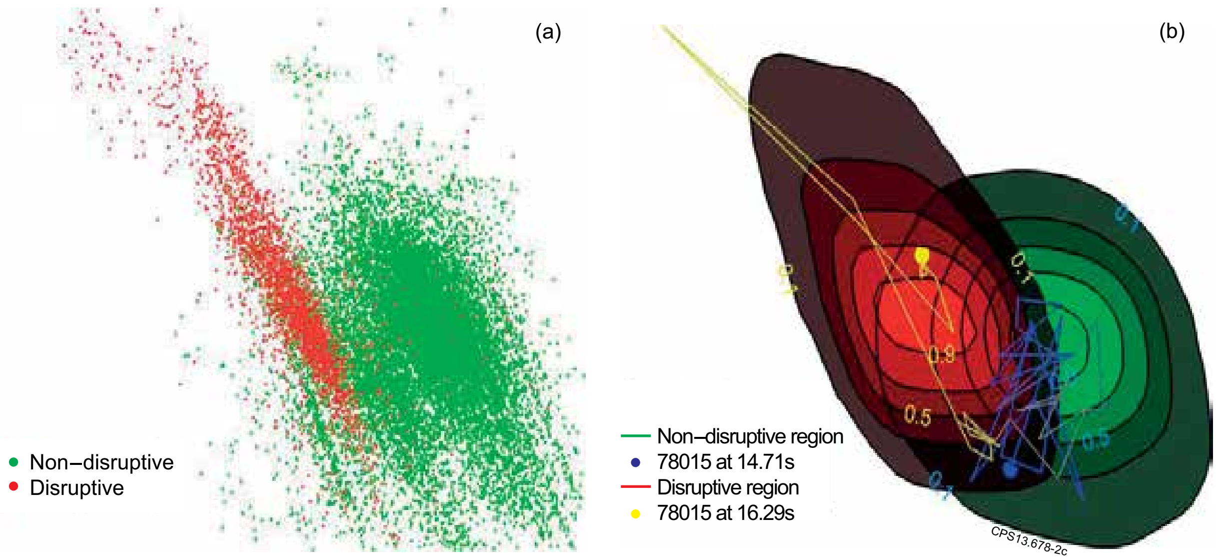

Figure 2: Two-dimensional projections from JET disruption data (a). Disruptive and non-disruptive clusters of data points are mapped using LMDS with 10 landmark points. (b). Density contours for the data points belonging to disruptive and non-disruptive clusters are drawn. The numbers give a qualitative measure of the degree of the disruptiveness of the region. A trajectory for Pulse No: 78015 is also traced.