Complex networks for the analysis of the synchronization of time series relevant for plasma fusion diagnostics

T. Craciunescu¹, A. Murari²-³, E. Peluso⁴, M. Gelfusa⁴, M. Lungaroni⁴, P. Gaudio⁴ and JET Contributors

EUROfusion Consortium, JET, Culham Science Centre, Abingdon, OX14 3DB, UK

¹National Institute for Laser, Plasma and Radiation Physics, Magurele-Bucharest, Romania
²Consorzio RFX (CNR, ENEA, INFN, Universita’ di Padova, Acciaierie Venete SpA), Padova, Italy
³EUROfusion Programme Management Unit
⁴University of Rome “Tor Vergata”, Rome, Italy

*See the author list of “Overview of the JET results in support to ITER” by X. Litaudon et al. to be published in Nuclear Fusion Special issue: overview and summary reports from the 26th Fusion Energy Conference (Kyoto, Japan, 17-22 October 2016)

Abstract—A pair of methods, based on joint recurrence plots and a modified version of cross visibility networks, is proposed for the determination of the efficiency of Edge Localized Modes (ELM) triggering by pellet injection. The main issue in the evaluation of the synchronization between pellet injection and ELM occurrence resides in the fact that ELMs are quasi periodic in nature and therefore, after any pulsed perturbation, if enough time is allowed to elapse, an ELM always occurs. The proposed methods allow the efficiency evaluation on a shot to shot basis and the identification of different phases in a discharge. The performance of the methods is illustrated using JET discharges with the ITER Like Wall (ILW).

Keywords—ELM control, pellet injection, complex networks.

I. INTRODUCTION

The time series analysis and the determination of causal-effect relationship represent an important topic in various fusion plasma studies like: i) the investigation of instabilities, including the assessment of disruption causes, ii) the study of plasma dynamics and of the impurity control, particularly important since the installation of the new ILW iii) the L-H transition, for which several models have been proposed but a dynamical theory is still unavailable.

ELMs play a crucial role in present-day tokamak operation and their control is a very important factor for the development of reactor scenarios. Moreover the Type-I ELMy H-mode has been chosen as the standard operation scenario for ITER. The control of ELMs is very important as they determine a significant reduction of the energy confinement. The confinement degradation, which occur during a time interval of the order of a millisecond, leads to significant particle expelling and to a large heat deposition on the vessel wall. At the scale of the next generation of devices these phenomena will determine an unacceptable vessel wall erosion level.

Various forms of ELM pacing based on external perturbations, have been proposed. One of the most promising is the pacing of ELMs by injecting small pellets of frozen fusion fuel into the plasma edge at high frequency [2]. An illustration of the time series describing the sequence of ELMs and the pellet injection is presented in Fig. 1. In order to determine the efficiency of the control method it is therefore important to assess the synchronization of the two time series. The main difficulty of the problem resides in the fact that ELMs are quasi periodic phenomena. After a sufficiently long time interval the ELMs will reoccur independently on pellet injection, as a natural effect of the plasma dynamic. Therefore the evaluation of the effectiveness of ELM pacing techniques should determine the time interval when their triggering effect has a real effect.

Recently several approaches, able to work in realistic experimental conditions have been proposed [3]. These methods are based on statistical approaches which implement a concept of causality as increased predictability. Granger has implemented this concept in the following manner: if the process \( X \) described by the time series values \( \{x_i\} \) is the ‘Granger cause’ of the process \( Y \), generating time series values \( \{y_i\} \), then the past evolution of \( X \) and \( Y \) provide more information for the prediction of \( Y \) future evolution than the past evolution of \( Y \) alone. A further refined approach, especially developed for the information transfer between time series and consequently for evaluating their causal interaction, in term of predictability, is the transfer entropy (TE) [5].
In this paper we present two approaches based on complex networks. The mathematical background of the methods is presented in section II. The methods are then applied to ELM pacing with pellet in Section III using data of JET discharges with the ITER Like Wall (ILW). Conclusions are drawn at the end of the paper.

II. COMPLEX NETWORKS

The transformation of the observational time series into complex network representations aims to characterize the underlying dynamics of time series by means of topological features. A plethora of different methods have been proposed to accomplish this mapping. For the particular case of the synchronization between pellet injection and ELM occurrence the Recurrence Plots (RP) and Complex Networks (CN) seem to be quite effective.

A. Recurrence Plots

Recurrence plots construct a sparse binary matrix from a time series in order to allow the easy visualization of the information that characterizes the behavior of trajectories in the phase space [6]. Each time when the trajectory is visiting roughly the same point in the phase space is marked in the recurrence plot matrix by a unit value:

\[ R_{ij}(\varepsilon) = \Theta(\varepsilon - \| x_i - x_j \|) \]

where: \( x_i \) and \( x_j \) are the point in phase space, \( i \) and \( j \) indexes the corresponding time points, \( N \) is the total number of points, \( \varepsilon \) is a threshold defining the size of the neighborhood and \( \Theta(x) \) is the Heaviside function. In an attempt to provide a unified framework for studying time series as complex networks, Donner et al. proposed to interpret the recurrence matrix (1) as the adjacency matrix of an unweighted and undirected complex network. In this way links are created between neighboring points in the phase space [7].

A multivariate extension of RP is the joint recurrence plots (JRP). For two sub-systems characterized by independent RPs, the joint recurrence plot is defined as the Hadamard product of the recurrence matrices:

\[ JR(i,j) = \Theta(\varepsilon_x - \| x_i - x_j \|) \cdot \Theta(\varepsilon_y - \| y_i - y_j \|) \] (2)

JRP can be used to detect the simultaneous occurrence of recurrences in two sub-systems. An example for a discharge devoted to ELM pacing is presented in Fig. 2.

Besides allowing the visualization of the dynamical system periodicities, RPs allow the definition of several quantities indicators [9]. The indicators are constructed starting from the distribution of diagonal, vertical and horizontal lines appearing in the RP. In this paper this recurrence quantification analysis (RQA) is performed by means of the entropy of the diagonal length which is defined by the following relation:

\[ ENTR = \frac{\sum_{l=1}^{N} p(l) \ln p(l)}{\sum_{l=1}^{N} p(l)} \] (3)

Where \( p(l) = \frac{\sum_{l_{min}}^{l} p(l)}{\sum_{l_{min}}^{N} p(l)} \)

Fig. 2. Example of recurrence plot for discharge 82439 devoted to ELM pacing with pellets.

The definition of ENTR is based on the frequency distribution of the diagonal lengths in the RP. Therefore it accounts for the complexity of the RP with respect to the diagonal lines. For uncorrelated noise (low complexity) ENTR has small values while for RPs developing an increased number of longer diagonals ENTR has higher values.

B. Weighted cross-visibility networks (WC-VN)

A widely used complex network approach relies on the visibility graphs (VG) algorithm which creates the complex network by linking visible elements in a series [10]. Every component in the time series \( \{ y_i \} \), \( i=1, \ldots , N \) is mapped to a node in the graph and node \( i \) is connected with node \( j \) if the following relation is true:

\[ y_k < y_i + \frac{y_j - y_i}{j-i} (k-i), \forall i < k < j \] (4)

The construction of VG is illustrated in Fig. 3.

A recent extension of VGs are the so called ‘Cross-Visibility’ Networks (CVN) recently introduced in [11] which transform a set of two time series \( \{ x_i \} \) and \( \{ y_i \} \) in a graph in order to reveal their coupling. For each time series points \( \{ x_i \} \) a node is created in the graph. A connection between two
nodes $i$ and $j$ is created if one of the following relations are valid:

$$y_k = y_i + x_j - x_i (k - i), \quad i < \forall k < j \quad (5)$$

$$y_k = y_i + x_j - x_i (k - i), \quad i < \forall k < j \quad (6)$$

According to Eqs. (5) and (6) the node is looking at the components of $\{x_i\}$ time series, through the obstacles of the shifted time series $\{y_k\} = \{y_k - y_i + x_i\}$. The visibility is ensured if, $tg(\alpha) > tg(\beta)$ for any $i < k < j$ (visibility from the top) or $tg(\alpha) < tg(\beta)$ for any $i < k < j$ (visibility from beneath). The construction of CVN is illustrated in Fig. 4. CVN is an measure of the mutual information relating the two time series. The adjacency matrix has values equal to 1 if Eq. (5) or Eq. (6) is satisfied and 0 otherwise. In our approach we have modified the network matrix by weighting the connections with the metric distance between two connected values in the time series.

$$A_{ij} = \{\text{dist}(y_i - y_j), \text{if Eq. (5) or Eq. (6) is satisfied}\}$$

This allows the use of a metric distance which is able to take into account the noise inherently superimposing on the experimental data (as described below, in subsection C).

The structural complexity of the CVN may be characterized by the Shannon entropy of the normalized adjacency matrix:

$$S = -\sum_{i,j} A_{ij} ln A_{ij} \quad (8)$$

For uncorrelated times series, when the visibility of two points in the time series $\{x_i\}$ are more frequently interrupted by obstacles in the time series $\{y_k\}$, the complex network is characterized by a broad distribution of the link’s lengths. This corresponds to lower values of the entropy. When the time series are synchronized the number of longer paths in the network increases and this is reflected by higher values of the entropy.

C. Geodesic Distance on Gaussian manifolds

RP and WC-VN require the calculation of a distance. The most used approach is based on the Euclidean distance, due to its simple geometric meaning and its straightforward implementation. However, the experimental data is always affected by noise and can include outliers. In order to cope with this problem we have recently proposed [12] to use the Geodesic Distance on Gaussian Manifolds (GDGM), which takes into account intrinsically the additive Gaussian noise accompanying the experimental data. In the view of the GDGM concept the individual measurements are interpreted as Gaussian distributions. Therefore a Gaussian probability density function (pdf), characterized by a specific mean $\mu$ and standard deviation $\sigma$, can be associated to each point in the time series data. In this view the distance between two time series points is the distance between the corresponding Gaussian distributions. An appropriate definition of this distance is the geodesic distance on the probabilistic manifold containing the data [13]. The Fischer-Rao metric allows the derivation of the geodesic distance on this Gaussian manifold [14]. For the case of two univariate Gaussian distributions $p_1(x|\mu_1, \sigma_1)$ and $p_2(x|\mu_2, \sigma_2)$, the geodesic distance is given by the relation:

$$DGM(p_1||p_2) = \sqrt{2\ln \frac{1+\delta}{1-\delta}} = 2\sqrt{\tanh^{-1}\delta} \quad (9)$$

where $\delta = \left[\frac{(\mu_1-\mu_2)^2+2(\sigma_1-\sigma_2)^2}{(\mu_1-\mu_2)^2+2(\sigma_1+\sigma_2)^2}\right]^2$.

III. RESULTS

The methods based on RP and WC-VN, have been tested on JET discharges with sufficient statistics and for cases when plasmas are sufficiently stationary. The time series used in the experiments rely on the measurement of the $D_0$ emission, which has been used to determine both the occurrence of the ELMs and the arrival time of the pellets in the plasma [15, 16], as shown in Fig. 1.

An illustrative example is presented in Fig. 5. The methods gives coherent and similar outputs. RP identifies two different lag times at 2.9 ms and 3.7 ms respectively, while WC-VN finds the synchronizing lags 2.8 ms and 3.5 ms. The discrepancy, typical for all the analyzed discharges, is the limit of the accuracy which can be achieved in this type of experiments. The results for the discharge #82854 (Fig. 5) show that the time resolution allow the detection of two phases during a single shot. The importance of using GDGM metric, which takes into account the presence of noise in the experimental data is illustrated in Fig. 6.
IV. CONCLUSIONS

The determination of the time interval for which a causal–effect relationship between pellet injection and ELMs occurrence for each particular discharge represents a more flexible approach in comparison with the traditional heuristic criteria reported in the literature which use a fixed lag time of 2 ms (see Ref. 3 for details). The time resolution is appropriate for the identification of the occurrence of different phases in the same discharge. The results show also that the cross visibility networks, including weights based on metric distances and amplitude values, may be, in general, an efficient tool for the investigation of time series synchronization experiments.

ACKNOWLEDGMENT

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

REFERENCES