Recurrence quantification analysis of electrostatic fluctuations in fusion plasmas

Z.O. Guimarães-Filho a, I.L. Caldas a, R.L. Viana b,*, J. Kurths c, I.C. Nascimento a, Yu.K. Kuznetsov a

a Instituto de Física, Universidade de São Paulo, 05315-970 São Paulo, São Paulo, Brazil
b Departamento de Física, Universidade Federal do Paraná, 81531-990 Curitiba, Paraná, Brazil
c Institut für Physik, Universität Potsdam, PF 601553, D-14415 Potsdam, Germany

Received 17 April 2007; accepted 27 July 2007
Available online 11 September 2007
Communicated by C.R. Doering

Abstract

We have investigated plasma turbulence at the edge of a tokamak plasma using data from electrostatic potential fluctuations measured in the Brazilian tokamak TCABR. Recurrence quantification analysis has been used to provide diagnostics of the deterministic content of the series. We have focused our analysis on the radial dependence of potential fluctuations and their characterization by recurrence-based diagnostics. Our main result is that the deterministic content of the experimental signals is most pronounced at the external part of the plasma column just before the plasma radius. Since the chaoticity of the signals follows the same trend, we have concluded that the electrostatic plasma turbulence at the tokamak plasma edge can be partially explained by means of a deterministic nonlinear system.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Recurrence plots (RPs) are graphical representations of the matrix [1–3]

\[ R_{i,j} = \Theta(\epsilon - \|x_i - x_j\|), \quad i, j = 1, 2, \ldots, N, \]

where \( x \in \mathbb{R}^D \) represents a dynamical state in the \( D \)-dimensional phase space of the system under consideration at time \( i \), \( \epsilon \) is a predetermined threshold, \( \Theta(.) \) is the unit step function, \( \| \cdots \| \) stands for the Euclidean norm, and \( N \) is the total number of points. The RP is thus obtained by assigning a black (white) dot to the points for which \( R_{i,j} = 1 (0) \).

RPs have been originally introduced as a numerical tool to calculate the maximum Lyapunov exponent related to a time series [1]. Later it has been shown that RPs can also be used to study non-stationarity of a time series as well as to indicate its degree of aperiodicity [4]. In particular, there are several measures proposed by Webber and Zbilut to make recurrence quantification analysis from an RP, turning it into an extremely useful tool in nonlinear time series analysis [5]. Recurrence quantification analysis has been extensively used in a plethora of problems ranging from meteorology [6], finance [7], and geophysics [8] to cardiology [9]. In particular, RPs have become important to characterize data from space plasmas [10].

In such disciplines one usually considers a univariate time series from which one can make an embedding using the vectors [11]

\[ x_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(d-1)\tau}\}, \]

where \( d \) is the embedding dimension and \( \tau \) is the delay. There are standard methods devised to obtain reliable estimates from both \( d \) and \( \tau \) [12]. The basic idea of an RP is to start from such a phase space embedding and compare the embedding vectors with each other, drawing pixels when the Euclidean distance between vectors is below some threshold \( \epsilon \), defined as a small fraction of the standard deviation of the time series being considered [13,14].
The taxonomy of dynamical behaviors can be explored in a large extent by using RPs and the quantifiers associated with them (recurrence quantification analysis, or RQA for short) [5]. As an example, stationary time series yield RPs which are homogeneous along a diagonal line. Moreover, if the RP shows a cloud of points with a homogeneous yet irregular distribution, then the time series has a pronounced stochastic nature. On the other hand, the formation of patterns in RPs may indicate stationary chaotic behavior. Moreover, RPs enable the computation of dynamical invariants, such as the second order Rényi entropy and correlation dimension [15,16].

One fertile field of study where RPs are potentially advantageous is the analysis of fusion plasmas such those generated by tokamaks [17]. A major goal in the study of such systems has been to understand the causes and associated rates of anomalously large cross-field transport, which is thought to be caused by plasma turbulence [18,19]. One experimental signature of plasma turbulence in the plasma edge of a tokamak is the fluctuating behavior of the electrostatic floating potential. Such signals often display a broad fluctuation spectra, as observed in the Brazilian tokamak TCABR [20]. A number of probes have been built in TCABR to measure the particle density and temperature fluctuations in the edge region. The experimental results suggest that the turbulent transport is mainly electrostatic in nature [21], as similarly observed in other tokamaks [22]. For example, electrostatic turbulence has been studied in the plasma edge of TCABR under the influence of radio frequency excited waves [23] and bias electrode polarization [24].

The deterministic content of the electrostatic fluctuations in the plasma edge has been assigned to physical processes governed by nonlinear mechanisms, like the interaction between drift waves which appear due to the steep density gradients in the plasma edge region of a tokamak [25]. These drift wave interactions are known to depend critically on the radial position such that the electrostatic turbulent fluctuations should also exhibit some radial dependence. However, the radial variation of the deterministic content of the plasma turbulence may not be immediately apparent from the experimental data obtained in tokamaks. Hence in this work we used RPs as a tool to quantify the recurrence properties of such series to provide a quantification of the degree of determinism and other related measures of RQA and their dependence on the radial position at the plasma edge. It turned out that this radial dependence, formerly elusive in conventional analysis using, e.g. spectral methods, is best observed using RQA.

The Letter is organized as follows: in Section 2 we briefly describe the RQA numerical diagnostics used to analyze the experimental data. Section 3 outlines the experimental setting, and the kind of data we obtain from the tokamak TCABR. Section 4 presents the results of RQA, as applied to data, and the last section contains our conclusions.

2. Recurrence quantification analysis

RQA comprises many quantitative diagnostics of the distribution of points (actually pixels) in a recurrence plot, focusing on three basic kinds of structures [26]. The first kind are single, or isolated points, which occur if the dynamical states are rare, do not persist for any time, or fluctuate heavily. The recurrence rate (RR) is the probability of finding a black recurrence point (for which $R_{i,j} = 1$), or

$$RR = \frac{1}{N(N-1)} \sum_{i,j=1, i \neq j}^{N} R_{i,j}$$

where $N^2$ is the total number of pixels (black or white) in an RP [27]. We remark that the main diagonal points are excluded from the double sum, since each point is recurrent with itself.

Most of the RQA diagnostics deal with diagonal lines, which are structures parallel to the line of identity $R_{i,j} = 1$, $i = 1, 2, \ldots, N$, and formally defined as

$$R_{i+k,j+k} = 1 \quad (k = 1, 2, \ldots, \ell),$$

$$R_{i,j} = R_{i+\ell+1,j+\ell+1} = 0.$$   

where $\ell$ is the length of the diagonal line, occurs when a segment of a given trajectory (in phase space) runs parallel to another segment. In other words, when an RP presents a diagonal line, two pieces of a trajectory undergo for a certain time (the length of the diagonal) a similar evolution and visit the same region of phase space at different times. This is the key idea of recurrence and thus a clearcut signature of determinism. Accordingly, we compute $P(\ell) = |\ell_i; i = 1, 2, \ldots, N_\ell|$, which is the frequency distribution of the lengths $\ell_i$ of diagonal lines, and $N_\ell$ is the absolute number of diagonal lines, with the exception of the main diagonal line which always exist by construction.

The determinism (DET) is defined as

$$DET = \frac{\sum_{\ell=\ell_{\text{max}}}^{\ell_{\text{min}}} \ell P(\ell)}{\sum_{\ell=1}^{\ell_{\text{max}}} \ell P(\ell)}$$

where $\ell_{\text{min}}$ is the minimum length allowed for a diagonal line, whereas the maximum diagonal length is $\ell_{\text{max}} = \max(\ell_i; i = 1, 2, \ldots, N_\ell))$. Thus DET measures the percentage of points in an RP belonging to diagonal lines. Other related quantities are the ratio (RATIO) between DET and RR and the average diagonal length (Ldm),

$$Ldm = \frac{\sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} \ell P(\ell)}{\sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} P(\ell)}.$$

Following Eckmann et al. [1] the lengths of the diagonal lines are related to the inverse of the largest positive Lyapunov exponent of the system. Likewise, the divergence $DIV = 1/\ell_{\text{max}}$ is related with the Kolmogorov–Sinai (KS) entropy of a dynamical system, or the sum of its positive Lyapunov exponents. Moreover, we can use RPs to compute reliable estimates for the Shannon entropy (ENTR),

$$\text{ENTR} = - \sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} p(\ell) \ln p(\ell),$$

where

$$p(\ell) = \frac{P(\ell)}{\sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} P(\ell)}.$$
is the probability distribution of the diagonal line lengths. \( \text{ENTR} \) reflects the complexity of the deterministic structure in the system. However, since it depends sensitively on the bin number, the value of \( \text{ENTR} \) may differ for different realizations of the same process, like distinct data preparations.

The third kind of interesting structures in an RPs is vertical lines, representing time intervals for which a dynamical state does not change or changes very slowly, and they turn to be a typical behavior of laminar states in intermittency scenarios. Vertical lines are defined by:

\[
\begin{align*}
R_{i,j+k} &= 1 \quad (k = 1, 2, \ldots, \nu), \\
R_{i,j} &= R_{i,j+k+1} = 0,
\end{align*}
\]

where \( \nu \) is the length of a vertical line. It suffices to consider vertical lines, since the RP is symmetric under interchange of indexes \( i \) and \( j \). Analogously to diagonal lines, we can obtain the frequency distribution of the lengths \( \nu_i \) of vertical lines,

\[
P(\nu) = \{\nu_i; \ i = 1, 2, \ldots, N_\nu\},
\]

where \( N_\nu \) is the absolute number of vertical lines. The laminarity (LAM) is the percentage of RP points forming vertical lines, or

\[
\text{LAM} = \frac{\sum_{\nu=\nu_{\text{min}}}^{\nu_{\text{max}}} \nu P(\nu)}{\sum_{\nu=1}^{N_\nu} \nu P(\nu)}
\]

where \( \nu_{\text{min}} \) is the minimum lengths of a vertical line, whereas the maximum vertical length is \( \nu_{\text{max}} = \max\{\nu_i; \ i = 1, 2, \ldots, N_\nu\} \). The trapping time (TT) is the average length of a vertical line

\[
\text{TT} = \frac{\sum_{\nu=\nu_{\text{min}}}^{\nu_{\text{max}}} \nu P(\nu)}{\sum_{\nu=\nu_{\text{min}}}^{\nu_{\text{max}}} P(\nu)}
\]

3. Electrostatic turbulence in tokamak plasmas

The knowledge of the transport properties of the plasma in the tokamak edge (i.e., the region comprising the outer portion of the plasma column and the vacuum region that separates it from the vessel wall) is essential to the stable operation of the tokamak [25,28]. Field line chaos plays here a major role in the interpretation of the experimental results, since it was long recognized that turbulent transport is particularly important in the plasma edge [29]. In particular, drift waves can be destabilized due to the confining magnetic field so as to yield a turbulent spectrum [30].

Quantitative investigations of the electrostatic turbulent spectrum have been made using both standard approaches, like spectral analysis and wavelets, as well as dynamical diagnostics like the return-time statistics [31]. The use of the latter approach is encouraged by the role of chaotic behavior in temporal scales on the onset and development of large-scale turbulence. One of such toolboxes is just the recurrence quantification analysis, which we apply to the electrostatic turbulence measurements available for the tokamak edge region.

The experiments were performed in a hydrogen circular plasma in the Brazilian tokamak TCABR [20] (major radius \( R = 61 \) cm and minor radius \( a = 18 \) cm). The plasma current reaches a maximum value of 100 kA, with duration 100 ms, the hydrogen filling pressure is \( 3 \times 10^{-3} \) Pa, and toroidal magnetic field \( B_T = 1.1 \) T. Three Langmuir probes measure the density and the plasma electrostatic potential: two probes measure the floating potential fluctuations and the third probe measures ion saturation current fluctuations. The probes are mounted on a movable shaft that can be displaced radially from \( r = 15 \) cm to 23 cm, with respect to the center of the plasma column. In this work we shall focus on the range from 16 to 21 cm so as to cover both the plasma edge and the so-called scrape-off layer, the latter comprising part of the vacuum layer existent between the plasma column and the vessel wall. The probe displacement, however, occurs only for separate discharges, in order not to disturb the plasma due to the movement of the probe. The measurements were performed at a sampling frequency of 1 MHz, and the measuring circuit has a 300 kHz bandwidth to avoid aliasing, such that in every discharge out of 10^5 points can be recorded [21,23].

Fig. 1 shows the time evolution of a typical tokamak plasma discharge in TCABR. The plasma current [Fig. 1(a)] grows rapidly in the first 20 ms and reaches a plateau where the current stays at a 100 kA level, decaying slowly during the second half until the eventual disruption. The electron density evolution, indicated by Fig. 1(b), exhibits a similar evolution, with a plateau level of \( n_e \sim 10^{19} \) m\(^{-3}\). The signals we are particularly interested to study are for the floating electrostatic potential \( V_f \), a representative example being depicted by Fig. 1(c), which shows highly irregular fluctuations in the \(-200\) V to \(+200\) V range. The window chosen for measurements of the floating potential was indicated by vertical lines. We choose this window, indicated by vertical bars in Fig. 1, so as to avoid the discharge phase where external perturbations are applied for other kinds of investigations. Fig. 1, however, is rather exceptional since it represents a tokamak discharge where no such perturbations have been applied to the plasma.

The nature of the electrostatic potential fluctuation is dependent on the radial location where the probe is placed, as suggested by Fig. 2, where 2 ms of the time window in the
The radial dependence of the electrostatic turbulence level at the vicinity of the plasma radius is a signature of the role played by radial density gradients in the generation of drift waves which is the essential cause of turbulence in the plasma edge. In fact, the presence of steep density gradients in the plasma edge can give rise to fully developed drift-wave turbulence, which is considered a likely candidate for explaining anomalous transport observed in experiments [25]. However, characterizing turbulence in order to quantify its level and radial dependence is a difficult task, what can be illustrated by Figs. 2(d)–(f), where we show the power spectra of the potential fluctuations in the three radial positions just analyzed. All of them are broadband, which is already expected from the chaotic behavior related to turbulence but, apart from some unessential rippling, those spectra do not show a distinguished feature which could be used to quantify the turbulence level and specific different dynamical regimes.

The usefulness of linear approaches like power spectra is naturally limited in view of the strong nonlinear character of plasma turbulence, and encourages the use of chaos-based diagnostics, like the correlation dimension, Lyapunov exponents, and entropies. However, the existing methods for evaluating those quantities assume that the time series being investigated is stationary, long enough, and with low noise. These requirements are difficult to fulfill in plasma experiments [12]. To overcome these difficulties, other approaches have been proposed, like using the statistical properties of the return times to explore the recurrent behavior typical of turbulent phenomena [31].

The success of using these recurrence methods in characterizing the turbulence level observed in tokamak experiments suggests that RQA can be also of interest, especially because it does not impose stationarity nor long series as necessary conditions, and can also work satisfactorily with moderate noise levels. Moreover, RQA yields reliable estimates of the Shannon entropy and can also indicate the amount of determinism in a given time series, what gives us an idea of the noise level added to the chaotic signal. The recurrence plots for the first 1000 points (corresponding to a 1 ms interval) of the series depicted in Figs. 2(a)–(c) are shown in Figs. 3(a)–(c), respectively. We choose the embedding dimension as $d = 4$ and the time delay $\tau$ was selected by considering the first local minimum of the autocorrelation function [12]. We can recognize the changes in the turbulent behavior at the plasma radius by comparing the diagonal and vertical structures of Fig. 3(b) with the scattered nature of the recurrence plots depicted by Figs. 3(a) and (c), which suggest a pronounced stochastic effect, probably related to noise and/or other mechanisms not accounted for in a determinisitic theory. Figs. 3(d), (e), and (f) contain a magnification of the lower left corner of Figs. 3(a), (b), and (c), respectively, and illustrate qualitatively the different recurrence patterns as we consider the turbulent fluctuations in the vicinity of plasma edge.

Since we have selected a 1 ms portion of each signal for constructing a recurrence plot, an immediate question arises as how we could assure that this piece of the original signal constitutes a stationary series. There are standard diagnostics of stationarity for nonlinear time series [12], but we have performed such a test using the consistency of recurrence-based diagnostics, like the determinism (but practically any other of the quantities described in Section 2 could be equally used). Fig. 4(a) exhibits the results of these tests, where the determinism (DET) of a sequence of recurrence plots have been determined, using Eq. (5), for consecutive pieces of an original series, each of them with 1 ms of length (equivalent to 1000 points, as before). For data acquired at $r = 16.5$ cm [depicted as empty circles in Fig. 4(a)] the values of DET for each series segment are distributed around a mean value of $\approx 0.17$ and with a small dispersion, indicating a consistency of such diagnostic against the total extension of the series and thus suggesting that the original series is stationary enough for our purposes. This stationarity test was repeated for other two radii, just like those considered in Figs. 2 and 3, the results pointing to values of determinism around 0.23 and 0.36, respectively.

Another issue to be taken into account when analyzing data from tokamak plasma is that each data set is related to a specific plasma discharge (tokamaks operate in pulsate regime). Since there is a plethora of physical factors affecting the nature of a discharge, we expect that such factors can be adequately managed out, such that the results are reproducible enough to give trustworthy values for the diagnostics we perform. This can be obtained, in practice, by making measurements in successive discharges, for which the tokamak and plasma parameters are...
Fig. 3. Recurrence plots for the first 1000 points (corresponding to a 1 ms time window) of the series for floating potential at radial positions (a) \( r = 16.5 \) cm, (b) \( r = 18.0 \) cm, and (c) \( r = 21.0 \) cm. Magnifications of the lower left corners of these plots are shown in (d), (e), and (f), respectively.

held constant. Possible differences, however, may arise from instabilities and other effects related to the highly nonlinear (and actually turbulent) plasma behavior. In Fig. 4(b) we consider potential fluctuation data from out of 57 discharges involving 9 radial positions of the probe (in each discharge the probe position is held fixed at a given position). The mean values of DET obtained in different discharges are plotted versus the radial positions. The line joining those points is a polynomial fit just to guide the eye, i.e. it is not intended to give a radial profile DET \( r \) but rather a trend: the degree of determinism increases significantly as we approach the plasma radius, and decreases afterwards. This suggests a diminishing contribution of stochastic effects at the plasma edge, in conformity with the visual inspection of Fig. 3.

One key factor when constructing a recurrence plot is the choice of the embedding dimension \( d \). We have compared different values of \( d \) in Figs. 5(a) to (d) to plot radial profiles of DET for different discharge, in much the same way as in Fig. 4(b). The consistency of the radial profile of DET with a clear maximum near the plasma radius is demonstrated qual-
Fig. 4. (a) Variation of determinism with the time window (each of them containing 1000 points) selected from the original signal for \( r = 16.5 \) cm (circles), \( r = 18 \) cm (squares) and \( r = 21 \) cm (stars). The mean values are indicated as full lines. (b) Radial profile for the mean value of determinism for various discharges (see text for details). The full line is a polynomial fit and the horizontal bar placed in the bottom represents the scrape-off layer region, after the plasma radius (determined by a material limiter).

Fig. 5. Radial profiles of determinism for embedding dimensions (a) \( d = 1 \), (b) \( d = 2 \), (c) \( d = 3 \), and (d) \( d = 5 \). Circles stand for original data, whereas squares refer to surrogate data. The horizontal bar indicates the scrape-off layer region.

Iteratively for different embedding dimensions (a quantitative agreement is clearly impossible since the values of \( d \) affect the topology of the reconstructed attractor). Fig. 5 also shows another numerical experiment we performed, by computing DET for the same time series as before, but with shuffled points. This procedure creates a surrogate series for which the statistical distributions are preserved, but destroying any time correlation due to determinism. Hence, a difference between the values of DET between the original series and its surrogate indicates a deterministic content which yields the (chaotic) dynamics we observe. Comparing those differences for a fixed radius, say \( r = 18 \) cm, shows that this deterministic content increases when we pass from \( d = 1 \) to 3, but saturates afterwards. We concluded that \( d = 3 \) is enough for our purposes. This obviously does not imply that \( d = 3 \) is an optimal embedding dimension, in the sense we require, for example, for obtaining the correlation dimension through standard procedures [32], but recurrence plots with \( d = 3 \) already furnish results as good as with other higher-dimension (and time-consuming) embeddings. We have varied other parameters as the cutoff radius (\( \epsilon \)) and minimum diagonal and vertical lengths (\( \ell_{\text{min}} \) and \( \nu_{\text{min}} \)), without noticeable changes in the profiles.
Up to here, we restricted our analysis to the determinism DET, but the general trend exhibited by the latter is also clearly seen in other diagnostics, as illustrated by Fig. 6. We considered also the laminarity [Fig. 6(a)], entropy [Fig. 6(b)], average diagonal length [Fig. 6(c)], and trapping time [Fig. 6(d)]. We have, in fact, obtained also the radial profiles for the remaining diagnostics defined in Section 2, but we have chosen to show only four since the results do not differ in a significative way from each other.

4. Conclusions

Recurrence quantification analysis (RQA) is a powerful approach for the investigation of nonlinear time series appearing in a variety of physically interesting applications, since RQA quantifies the number and duration of recurrences of a dynamical system presented by its phase space trajectory. One of the advantages from using RQA is to get measures for quantifying the information content of turbulent signals. Those recurrence-based diagnostics, like determinism, Shannon entropy, laminarity, trapping time, etc., have been shown to better distinguish the origins of complex behavior in experimental time series by evaluating the deterministic and stochastic contents of them, if compared with other techniques like power spectra, conditional analysis, etc.

The experimental time series considered in this Letter are electrostatic potential fluctuations in the edge plasma of a tokamak. The stability and reproducibility of RQA results obtained in this work indicate that the recurrence-based diagnostics are appropriate to analyze data from plasma edge turbulence obtained from Langmuir probes. The power of RQA diagnostics was used to clearly show that the deterministic content of these fluctuations is not spatially uniform, but it is more pronounced just before the plasma border. Our results suggest that theoretical models for describing such fluctuations, using nonlinear mode coupling of drift waves, can be used in the vicinity of the plasma border, but may not give reliable results when applied in the internal edge plasma or the far scrape-off layer separating the plasma column from the tokamak wall. The radial location of the maximum level of determinism is within the region of steep density gradients observed in tokamaks, the latter being considered as the main cause of drift wave instabilities leading to electrostatic plasma turbulence.

Thus, our findings are in favor of theoretical models describing plasma turbulence in terms of nonlinear interaction of drift waves and their predictions as the coherent structure propagation and the onset of zonal flows. On the other hand, usual techniques of fluctuation analysis may not give the same information we got from RQA about the radial dependence of the deterministic content of the signal. As an example, if we consider that the standard deviation of the fluctuations would quantify the stochastic content of the signal we would be led to a rather different conclusion. In fact, the results show a higher standard deviation in the scrape-off layer, slightly after the plasma border. Hence this conventional statistical approach would not recommend the use of deterministic models in the vicinity of the plasma border, which is just the region for which RQA warrants the usefulness of such models. Moreover, while our experimental data do not show stationary behavior from the stochastic point of view, their statistical moments are not constant, we found evidences of stationarity from the deterministic point of view.
Acknowledgements

This work was made possible with partial financial help from FAPESP, CNPq, Fundação Araucária, and CAPES (Brazilian Government Agencies).

References

    V. Bellintani, M. Machida, M.V.A.P. Heller, R.M.O. Galvão, E.K.
    Bengtson, W.L. Hodge, R.D. Durst, S.C. McCool, B. Richards, K.W. Gentile,
    Instum. 59 (1988) 1739;