

## A statistical analysis approach of plasma dynamics in gyrokinetic simulations of stellarator turbulence

A. D. Papadopoulos<sup>1</sup>, J. Anderson<sup>2</sup>, E-J. Kim<sup>3</sup>, M. Mavridis<sup>4</sup> and H. Isliker<sup>4</sup>

<sup>1</sup> *School of Electrical and Computer Engineering, National Technical University of Athens, 157 80, Greece*

<sup>2</sup> *Department of Space, Earth and Environment, Chalmers University of Technology, SE-412 96 Göteborg, Sweden*

<sup>3</sup> *Centre for Fluid & Complex Systems, Coventry University, Priory St, Coventry CV1 5FB, UK*

<sup>4</sup> *Dept. of Physics, Aristotle University of Thessaloniki, 541 24, Thessaloniki, Greece*

### Abstract

A geometrical method is used for the analysis of stochastic processes in plasma turbulence. Distances between thermodynamic states can be computed according the thermodynamic length methodology which allows the use of a Riemannian metric on the phase space. A geometric methodology is suitable in order to understand stochastic processes involved in e.g. order-disorder transition, where a sudden increase in distance is expected. Gyrokinetic simulations of Ion-Temperature-Gradient (ITG) mode driven turbulence in the core-region of the stellarator W7-X, with realistic quasi-isodynamic topologies are considered. In gyrokinetic plasma turbulence simulations avalanches, e.g. of heat and particles, are often found and in this work a novel method for detection is investigated. This new method combines the Singular Spectrum Analysis algorithm and Hierarchical Clustering such that the gyrokinetic simulation time series is decomposed into a part of useful physical information and noise. The informative component of the time series is used for the calculation of the Hurst exponent, the Information Length and the Dynamic Time. Based on these measures the physical properties of the time series is revealed.

### The SSA & HC methods

The SSA analysis for 1D series is a special case of the 2D SSA method. Therefore only the latter is presented. The Gene-data  $Q(t, \rho)$  forms a matrix of size  $N_x \times N_y$ . A window  $W$  of size  $L_x \times L_y$  is considered, running inside  $Q(t, \rho)$ , that is  $L_x \leq N_x$ ,  $L_y \leq N_y$ . In the two dimensional SSA, the trajectory matrix is defined as  $\mathbf{X} = [X_1, X_2, \dots, X_{K_x \cdot K_y}]$ . It is a matrix of size  $L_x \cdot L_y \times K_x \cdot K_y$  where  $K_x = N_w - L_w + 1$ ,  $w = x, y$  and its columns  $X_i$  are vectorizations (vec) of the  $L_x \times L_y$  submatrices  $W$ ,  $X_{k+(l-1)K_x} = \text{vec}(W)$ ,  $W = X_{k,l}^{(L_x \times L_y)} \equiv (Q_{i,j})_{i=k, j=l}^{L_x+k-1, L_y+l-1}$ . More compactly  $\mathbf{X} = \mathbf{T}_{2D}(Q)$ , where  $\mathbf{T}$  is the embedding linear operator. It is assumed that the  $x$ -axis is oriented to the bottom and the  $y$ -axis to the right of the  $Q$  matrix and the origin is the

left upper corner.

The next step of the SSA analysis consists of the decomposition of the trajectory matrix  $\mathbf{X}$ . In particular the Singular Value Decomposition is applied on the trajectory matrix leading to the decomposition  $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2, \dots, \mathbf{X}_d$ , where  $\mathbf{X}_j = U_j V_j^T \sqrt{\lambda_j}$ ,  $j = 1, 2, \dots, d$ , the  $j^{\text{th}}$  singular vectors and value ( $\lambda_j$ ) respectively and  $d$  the rank of  $\mathbf{X}$ . The  $d$  values and corresponding matrices  $\mathbf{X}_j$  are grouped into  $m$  sets:  $\{1, 2, \dots, d\} = \cup_{i=1}^m I_i$ . The grouping of the indices,  $j = 1, 2, \dots, d$ , is achieved via the agglomerative HC method. In the agglomerative HC method a hierarchy of clusters is produced following a bottom-up approach. In particular each observation (index) forms its own cluster and then observations are merged in an additive manner, moving up in the hierarchy. Each observation corresponds to a particular reconstructed matrix  $\tilde{X}_j$  related to the  $j^{\text{th}}$  component  $\mathbf{X}_j$ , of the trajectory matrix  $\mathbf{X}$ . The observation merging is based on a weighted distance matrix  $d_w \equiv 1 - \rho_w$ , where  $\rho_w(i, j) \equiv (\tilde{X}_i, \tilde{X}_j)_w / (\|\tilde{X}_i\| \|\tilde{X}_j\|)$ ,  $\|\tilde{X}_i\| \equiv \sqrt{(\tilde{X}_i, \tilde{X}_i)_w}$ ,  $(X, Y)_w \equiv \sum_l w_l X_{l,l}^T Y_{l,l}$ . The distance between observations  $i, j$  is given by  $(d_w)_{i,j}$ . In the first HC step, pairs of observations  $i, j$  form a cluster if  $i = \text{agr}_{i*} \min (d_w)_{i*,j}$ . In the second HC step, since clusters have been formed, there must be a definition of cluster distance. This is provided by the linkage clustering, for which there are many choices, single-linkage, complete-linkage, average-linkage, Ward-linkage, etc. The single-linkage clustering is used here, where clusters  $A, B$ , are at a distance  $\min \left( (d_w)_{i,j}, i \in A, j \in B \right)$ . The HC steps continue up-wards clustering until the desirable number of clusters,  $m$ , are formed (the HC stopping criterion in this work) or until the clusters are too far apart to merge. In this study  $m = 3$  and we consider the classes (clusters), oscillations, trend and noise. Eventually the noise part is removed.

The reconstructed matrices  $\tilde{X}_j$  are the result of antidiagonal averaging [2] followed by inverse embedding,  $\mathbf{T}_{2D}^{-1}$ , applied to the  $\mathbf{X}_i$  matrices.

### Evaluation of Information Length, Dynamic Time and Hurst Exponent

From SSA we obtain time subseries  $\tilde{X}_i(\rho)$ ,  $i = 1, \dots, d$  of the various SSA components and it is possible to calculate their dynamic time and information length. The dynamic time  $\tau(t, \rho)$  is a time scale over which the probability of the  $\tilde{X}_i(\rho)$  changes on average at time  $t$  (denoted as  $p(\tilde{X}_i(\rho), t)$ ). The probability  $p(\tilde{X}_i(\rho), t)$  is estimated from a subset (window) of  $\tilde{X}_i(\rho)$  samples of length  $W_L$  produced around the time index  $t$ . In particular, first the samples of  $\tilde{X}_i(\rho)$  are interpolated to equally spaced time instances (separated by  $dt$ ), then from the samples of each window  $w$  of size  $W_L$  the  $p(\tilde{X}_i(\rho), t)$  is calculated at the time instant  $t$  at the middle of the window's time interval. Then the window  $w$  is moved (running window) by one time sample and the  $p(\tilde{X}_i(\rho), t + dt)$  is calculated at the time instant  $t + dt$ , and so on. The calculation

of  $p(\tilde{X}_i(\rho), t)$  is done based on the histogram of  $w$ -samples. Usually the histogram-produced  $p(\tilde{X}_i(\rho), t)$  is non-smooth and a smoothing Gaussian kernel is applied. Then  $\tau_i(t, \rho)$  of the  $\tilde{X}_i(\rho)$  subsequence, is given by [4] :

$$\tau_i(t, \rho)^2 = 1 / \int d\tilde{X}_i(\rho) \frac{1}{p(\tilde{X}_i(\rho), t)} \left( \frac{\partial p(\tilde{X}_i(\rho), t)}{\partial t} \right)^2. \quad (1)$$

From the dynamic time the information length,  $L_i(t, \rho)$ , can be directly calculated [4]:

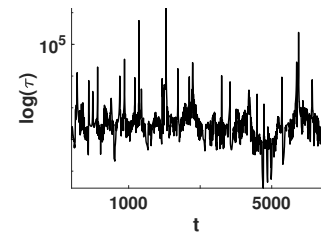
$$L_i(t, \rho) = \int_0^t ds \frac{1}{\tau_i(s, \rho)}. \quad (2)$$

Another quantity of interest is the Hurst exponent,  $H$ , of  $Q(\rho)$ . In general  $0 < H < 1$  but if  $H = 0.5$  the series is considered random (uncorrelated), if  $H > 0.5$  the series has long term positive autocorrelation meaning that high (low) values in series  $Q(\rho)$  will have a higher probability of being followed by another high (low) value whereas if  $H < 0.5$  in the long run with high probability high (low) values in  $Q(\rho)$  will have a higher probability of being followed by another low (high) value. The Hurst exponent is calculated by the rescaled range (RS) method as the exponent  $H$  such that:  $\mathbb{E}[R(n)/S(n)] = Cn^H$  for  $n \rightarrow \infty$ , where  $C$  is a constant,  $\mathbb{E}[\cdot]$  is the expected mean,  $S(n)$  is the standard deviation of the series  $Q_1(\rho), Q_2(\rho), \dots, Q_n(\rho)$  and  $R(n)$  is the range of the  $n$  cumulative deviations from the mean. That is  $R(n) = \max(Z_1, Z_s, \dots, Z_n) - \min(Z_1, Z_s, \dots, Z_n)$ ,  $Z_j = \sum_{i=1}^j (\tilde{X}_i - m)$ ,  $m = (\sum_{i=0}^n \tilde{X}_i) / n$ . Then  $H$  is calculated as the slope of the line that fits the  $\log(R(n)/S(n))$  data as a function of  $\log(n)$ .

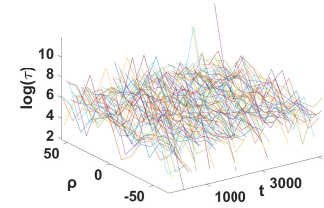
## Numerical Results

The analysis is performed using ion-heat flux data  $Q(\rho)$  time series from the gyrokinetic code GENE. The collected data are for normalized magnetic flux radii  $s = 0.5$ , normalized temperature gradient  $R/L_T = 1$  and density gradient  $R/L_n = 0$ . In addition the electrons are considered adiabatic. The  $Q(\rho)$  time series are analyzed with the 1D-SSA ( $Q$  data averaged over the radius  $\rho$ ) or with the 2D-SSA (non averaged  $Q$  data).

The dynamic time stores the instantaneous changes in the information length, this is indicative of sudden changes depending on the present resolution, see Figure 1a and 1b and [4]. In Figure 1a the dynamic time in the 1D case is presented, this is average behaviour of the strongly



(a) Dynamic time function of time samples  $t$ .



(b) Dynamic time as function of time samples  $t$  and varying radius  $\rho$ .

fluctuating signal in Figure 1b. However, the dynamic time exhibit rapid fluctuations and to have a consistent change in dynamics a significant change in the information over a short interval is needed. It is therefore suggested that only by analyzing the information and the dynamic time in tandem, an indication of change in dynamics can be detected. The information length is the additive effect of a series of consistent fluctuations, see Fig 2a and 2b.

A sudden increase in information length is found at around  $t = 5000$  in the 1D case. The information is almost linearly increasing until the change in the dynamics is suggested by a sudden increase in information. In 2D a more varied picture is found. At each radii the information length is calculated, which then gives a almost linear increase in the information length for each radii. It is observed that at certain radii there are sharp increases in information length e.g.  $\rho = 0$  after a certain time similar to the 1D case.

## Discussion

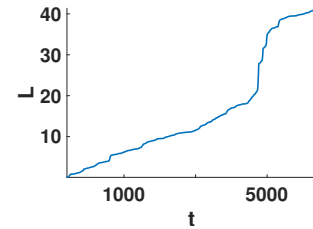
The information length measures the distance between two consecutive states in plasmas. It is thus believed that this measure could be used as indication that some large structure is formed e.g. and is mediating significant transport. In a previous effort the 1D model was developed however it was deemed difficult to directly compare with larger events found in the simulations thus a 2D version is now presented. It is seen that large fluctuations in dynamic time, the instantaneous change in length Fig 1, can be observed if they are consistent c.f. Hurst exponent Fig 3. It is found that at certain radii and times there is a significant increase in information length.

## Acknowledgement

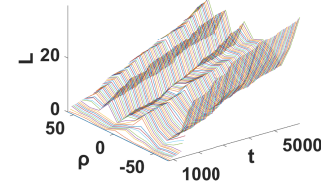
AP was supported by the European Union via the Euratom Research and Training Programme under Grant 101052200 EUROfusion through EUROfusion Consortium.

## References

- [1] The GENE Development Team, <http://genecode.org/>
- [2] N. Golyandina, V. Nekrutkin, A. Zhigljavsky, CHAPMAN & HALL/CRC, (2001).
- [3] J. Anderson, et al, Phys. Plasmas **21**, 122306 (2014). 122306 (2014).
- [4] J. Anderson, et al, Physics of Plasmas **27**, 022307 (2020).



(a) Information length as function of time samples.



(b) Information length, as function of time samples, and varying radius  $\rho$ .

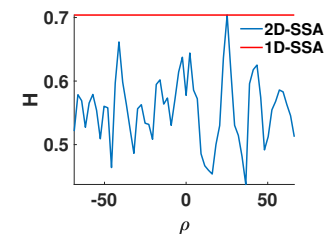


Figure 3: Hurst exponent,  $H(\rho)$  calculated by the 1D and 2D SSA .