OPEN ACCESS
IOP Publishing
Plasma Physics and Controlled Fusion
Plasma Phys. Control. Fusion **64** (2022) 084005 (12pp)
https://doi.org/10.1088/1361-6587/ac7a5d

Fast characterization of plasma states in W7-X with permutation entropy

J F Guerrero Arnaiz[1,2,*], A Dinklage[1,2], J Geiger[1], M Hirsch[1], U Höfel[1], C Brandt[1], H Thomsen[1], J Schilling[1], K Rahbarnia[1], U Neuner[1], M Zanini[1] and the W7-X Team[1]

[1] Max-Planck-Institut für Plasmaphysik, Greifswald, Germany
[2] Universität Greifswald, Institut für Physik, Greifswald, Germany

E-mail: juan.fernando.guerrero.arnaiz@ipp.mpg.de

Received 28 February 2022, revised 2 June 2022
Accepted for publication 20 June 2022
Published 4 July 2022

Abstract

Detecting changes in plasmas is compulsory for control and the detection of novelties. Moreover, automated novelty detection allows one to investigate large data sets to substantially enhance the efficiency of data mining approaches. To this end we introduce permutation entropy (PE) for the detection of changes in plasmas. PE is an information-theoretic complexity measure based in fluctuation analysis that quantifies the degree of randomness (resp. disorder, unpredictability) of the ordering of time series data. This method is computationally fast and robust against noise, which allows the evaluation of large data sets in an automated procedure. PE is applied on electron cyclotron emission and soft x-ray measurements in different Wendelstein 7-X low-iota configuration plasmas. A spontaneous transition to high core-electron temperature (T_e) was detected, as well as a localized low-coherent intermittent oscillation which ceased when T_e increased in the transition. The results are validated with spectrogram analysis and provide evidence that a complexity measure such as PE is a method to support *in-situ* monitoring of plasma parameters and for novelty detection in plasma data. Moreover, the acceleration in processing time offers implementations of plasma-state-detection that provides results fast enough to induce control actions even during the experiment.

Keywords: permutation, entropy, data-mining, W7-X

(Some figures may appear in colour only in the online journal)

1. Introduction

Wendelstein 7-X (W7-X) is an *optimized*, superconducting stellarator [8, 17, 23] with the main goal of demonstrating that stellarators are capable of plasma steady-state operation at fusion-relevant plasma conditions. The optimization was performed according to different reactor relevant criteria, e.g. minimizing losses due to neoclassical transport and maximizing the confinement of fast particles in the plasma center, while maintaining good magneto-hydrodynamic (MHD) stability at finite β equilibria. Regardless of the complexity and challenges that are faced in stellarator research, these devices are promising for reactor purposes since the magnetic field topology is formed only by external coils, which means that the plasma can thus be sustained continuously.

To ensure steady-state operation, it is necessary to detect, understand and react accordingly to plasma state changes that could compromise the plasma confinement. For this, novelty/anomaly detection methods based on neural networks are currently implemented e.g. for the detection of magnetic instabilities caused by Alfvén eigenmodes [20] and for the detection of disruption precursors in tokamaks [7]. However,

* Author to whom any correspondence should be addressed.

these methods can be computationally expensive and may require large data sets to operate appropriately. Here, we propose an extremely fast and robust detector based on permutation entropy (PE) for *in-situ* detection of transient events, as well as for offline bulk data analysis. To study the spatio-temporal evolution of the plasma and demonstrate the feasibility and benefits of the PE analysis, we implemented it on highly sampled T_e and soft x-ray (SXR) data.

2. Measuring predictability in multivariate time series

Measuring predictability of time series allows a better understanding of the characteristics and properties of any underlying stochastic processes and the characteristics of dynamical systems. Time series can be regarded to be more or less ordered. A generic task of time series analysis is to predict a value from measurements, that is e.g. calculating the probability to measure a value x_t given measurements $\{\ldots, x_{t-2\tau}, x_{t-\tau}\}$ preceding x_t or to unravel characteristic invariants of the time series. Entropy has been shown to be one of the most powerful quantities to evaluate the degree of randomness and predictability of a signal. However, with continuous data, the entropy of a time series can only be calculated to provide an estimation. The accuracy of which is bound to the advantages and drawbacks of the method implemented [6]. Various estimators have been developed, one of these being the PE which is a statistical measure that takes all statistical dependencies into account to describe the degree of randomness in a data set. PE was firstly introduced in [1] and has been successfully implemented in different science branches e.g. medical sciences [12, 13, 18] and financial economics [10, 24, 25], including plasma physics, e.g. for unraveling the chaotic nature of L-mode edge plasma density fluctuation dynamics [14] and characterizing low-aspect-ratio reversed-field pinch plasma through the identification of stochastic and chaotic regions [19]. In this paper, we apply PE on non-stationary data to characterize plasma states and their evolution. In particular, PE is employed as a method for time series analysis and the method is computationally effective to allow one for fast even real-time characterization.

In a sense, PE estimates the degree of randomness of a signal. The salient feature of PE is to identify ordinal patterns of m-tuples separated by a delay time τ_D [1, 2]. For a PE of order m, m measurements are analyzed at a time forming a *time comb* to determine the ordering of consecutive measurements. To do so, distinct measurements are assigned a *rank* or a *symbol* depending on their relative ordinal arrangement such that the order pattern π of the m-tuple can be identified. As an example, for the order $m=3$ one could arrange three (different) values into 3! different patterns π. In practice, PE is determined by counting the number of occurrences of each π and the calculating the Shannon information H_m of the probability distribution p_π:

$$H_m = -\sum_{\pi} p_\pi \log p_\pi. \tag{1}$$

The Shannon information varies within $0 \leqslant H_m \leqslant \log_2 m!$. $H_m = 0$ is attained for either a monotonously increasing or decreasing time series. For completely random time series (independent and identically distributed processes, *i.i.d.*) each permutation has probability $(m!)^{-1}$ leading to $H_m = \log_2 m!$. However, if there is some structure in the data, the permutations are not uniformly distributed $H_m < \log_2 m!$. And the more irregular the time series is, the closer is $H_m/\log_2 m!$ to one. For experimental data, an ordinal technique such as PE is less affected by noise and data outliers.

For a numerical implementation, PE is based on sorting algorithms rather than distance calculations which are way faster to compute. The main determinant of the computing speed of the PE is the permutation order m given that the sorting algorithm has to assign an order pattern to each m-tuple from $m!$ distinct possible patterns. The choice of m is a trade-off between details in ordering pattern and computational costs: while low m lead to very small computation times, a low permutation order does not allow a detailed separation of ordering patterns and may not accurately describe the underlying dynamics. If one is interested in details of the ordering patterns, a higher m would be required and a larger computation time is expected. In principle, one wants to ensure that all possible patterns $m!$ become visible (appear at least once). From the literature [5, 21] it can be seen that appropriate values of m can be chosen depending on the length of the time series to be analyzed, which e.g. for the case of 10^4 samples, a permutation order $m=4$ would be adequate. The effect of the hyperparameter τ_D on the PE will be discussed along a practical example below.

To further illuminate the PE technique, an analysis of synthetic data is presented. In this use case, the PE of two uncoupled systems is calculated, namely of $X_1(t_n)$ and $X_2(t_n)$ defined as:

$$X_1(t_n) = \xi_1 \tag{2}$$

$$X_2(t_n) = \xi_2 + \alpha X_2(t_n - 1), \tag{3}$$

where ξ_i is Gaussian noise with expectation value μ_i and standard deviation σ_i ($\xi_i \sim N_i(\mu_i, \sigma_i)$) and α is an autoregressive (AR) coefficient limited to $|\alpha| < 1$. Both systems consist of random signals, however $X_2(t_n)$ has an additional first order AR term creating a dependency between the last and the upcoming realization. The time traces of the first 100 samples of both systems are depicted in figure 1. For this example, we will consider $\mu_{1,2} = 0$, $\sigma_{1,2} = 1$, $\alpha = 0.9$ and $t_n = 1, 2, \ldots, 10^4$. From the waveforms it can be seen that $X_2(t_n)$ (red) is more structured and seem less random than $X_1(t_n)$ (blue).

Figure 2(a) shows the probability density functions (PDF) of the time series in figure 1. We observe that the PDF of $X_1(t_n)$ (blue) is narrower than the one from $X_2(t_n)$ (red); however, both PDFs have the same shape (normal distributions). It is noted here that, from this depiction of the one dimensional distribution, the degree of randomness of the signals cannot be determined.

For the assessment of the degree of randomness, the joint distributions of consecutive realizations from the time series

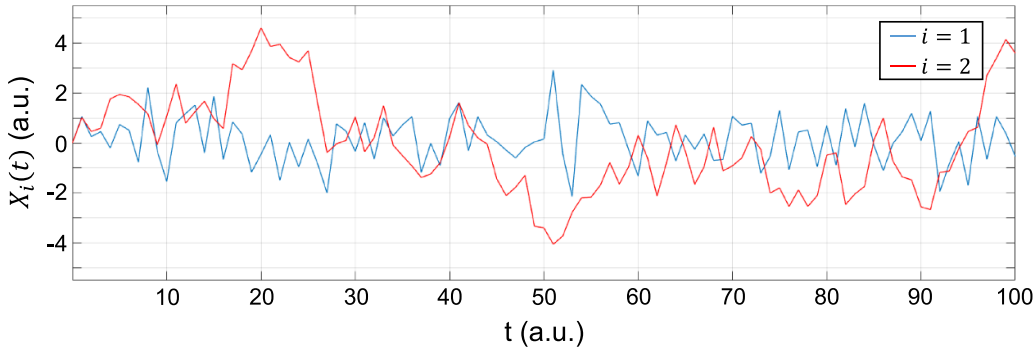

Figure 1. Time traces of synthetic data to be analyzed by means of PE. Data consists of the first 100 realizations of systems $X_1(t_n)$ and $X_2(t_n)$ as defined in equations (1) and (3).

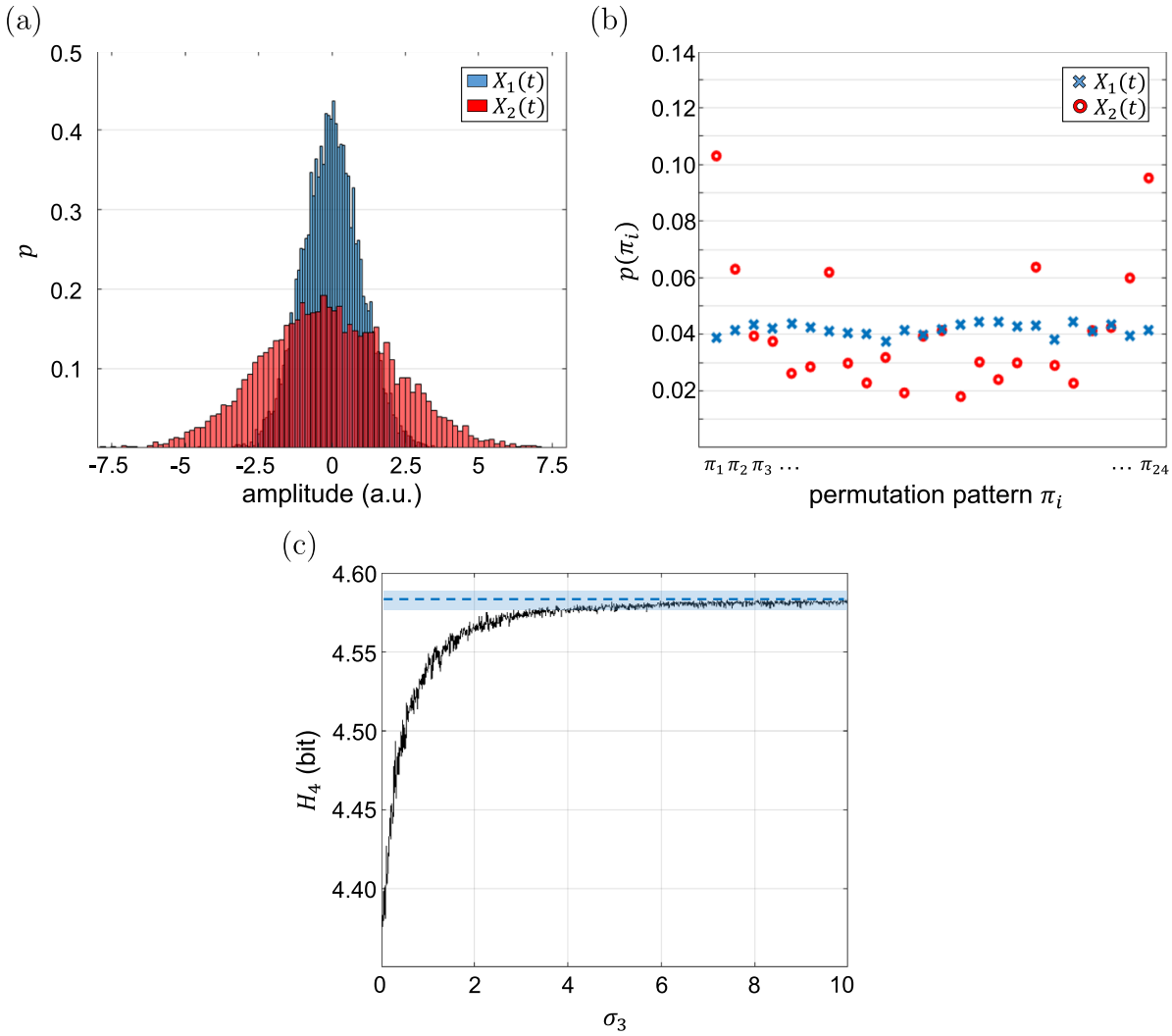

Figure 2. Analysis of uncoupled systems $X_1(t_n)$ and $X_2(t_n)$ for $t = 1, 2, \ldots, 10^4$. (a) Probability density function estimate of both systems. (b) Probability distribution of permutation patterns π of 4th order of consecutive measurements ($m = 4$, $\tau_D = 1$). The 4-tuples are arranged from monotonically decreasing to monotonically increasing order patterns such that $\pi_1 = (4,3,2,1), \pi_2 = (4,3,1,2), \ldots, \pi_{24} = (1,2,3,4)$. (c) Permutation entropy of system $X_2(t_n)$ with varying additional noise source $\xi_3 \sim N_3(0, \sigma_3)$. The blue dashed line indicates the entropy of system $X_1(t_n)$ ($H_4 = 4.583$ bit) and the shaded area illustrates the error with a confidence interval of 99.7%.

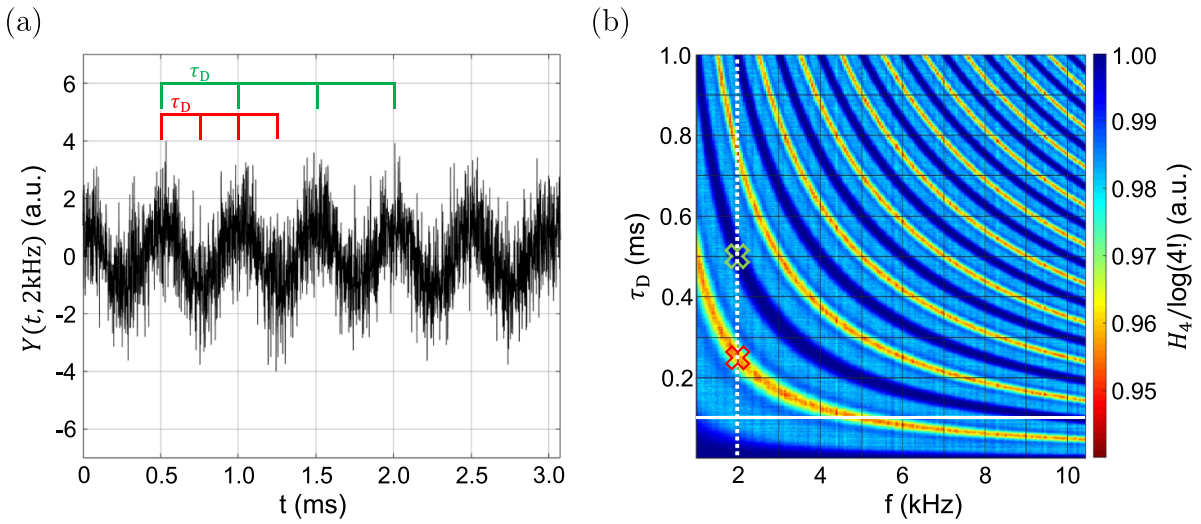

Figure 3. Calculation of permutation entropies of order $m = 4$ of signals $Y(t,f) = \sin(2\pi ft) + \xi$, where $\xi \sim N(0,1)$, for the assessment of best suited embedding delay τ_D to detect structures with different frequencies. (a) Time trace of the signal with $f = 2$ kHz. Minimum and maximum PE values were achieved with $\tau_D = 0.25$ ms (red) and $\tau_D = 0.50$ ms (blue); normalized to the maximum achievable entropy, $H_4/\log 4! = 0.950$ and $H_4/\log 4! = 0.999$ respectively. (b) $H_4/\log 4!$ of signals with varying f and τ_D. White solid line indicates values of PE for time series with different f and $\tau_D = 0.1$ ms; white dotted line, for the specific case $f = 2$ kHz with color coded markers pinpointing embedding delays shown in (a).

illustrated in figure 1 are shown in figure 2(b). For this, a PE with $m = 4$ and $\tau_D = 1$ is implemented. It can be seen for the case of $X_1(t_n)$ (blue) that all $m! = 24$ possible permutation patterns π have roughly equal probabilities ($p(\pi) \approx \frac{1}{24}$). On the other hand, for $X_2(t_n)$ (red) some probabilities deviate significantly from an equal distribution indicating that monotonically decreasing and increasing sequences are more frequent. This reflects the effect of the AR term (α in equation (3)) leading to more ordered patterns in $X_2(t_n)$ (red) compared to $X_1(t_n)$ (blue). Calculating the PE of the systems $X_1(t_n)$ and $X_2(t_n)$ results in 4.583 bit and 4.359 bit, respectively. This indicates a higher entropy in the time series $X_1(t_n)$ compared to $X_2(t_n)$. In other words, $X_1(t_n)$ is more random (and thus less predictable) than $X_2(t_n)$.

Now, we assess the robustness of PE. To this end, $X_2(t_n)$ is exposed to different noise levels, figure 2(c). For this, an additional i.i.d. signal $\xi_3 \sim N_3(0,\sigma_3)$ with varying σ_3 is incorporated into the time series of $X_2(t_n)$ depicted in figure 1. After a steep for low values of σ_3, it gradually approaches the maximum PE value of $X_2(t_n)$ (blue dashed line, $H_4 = 4.583$ bit). At about $\sigma_3 > 4$ the entropies of $X_1(t_n)$ and $X_2(t_n)$ are virtually the same. For this example with normal distributions at $\sigma_{1,2} = 1$, PE is able to detect autodependencies for noise levels up to $\sigma_3 \approx 4\sigma_1$.

Now we discuss the role of the hyperparameters in the PE calculation, namely the embedding delay τ_D and the order m. The assessment of a well-suited embedding delay τ_D leading to the detection of underlying structures at different frequencies is presented in figure 3. For this, synthetic data are generated consisting of noisy sine waves $Y(t,f) = \sin(2\pi ft) + \xi$, with frequencies $f \in [1, 11]$ kHz and $\xi \sim N(0,1)$. The waveform of $Y(t, 2\text{ kHz})$ is shown in figure 3(a). For a time window of 10 ms, the PE of order $m = 4$ is calculated for varying $\tau_D \in [0.01, 1.00]$ ms. This results in a minimum PE value with $\tau_D = 0.25$ ms (depicted in red) and maximum with $\tau_D = 0.50$ ms (green). These values, normalized to the maximum achievable entropy, are $H_4/\log 4! = 0.950$ and $H_4/\log 4! = 0.999$ respectively. This indicates an embedding delay of $\tau_D = 0.25$ ms (1/2 of the period T of the sine wave) to be best suited for the detection of the underlying structure with $f = 2$ kHz, whereas an analysis with $\tau_D = 0.50$ ms $= T$ would not be able to differentiate between this structure and a signal fully comprised of white noise. The reason becomes obvious from figure 3(a): if the time comb samples periodic data at a fixed phase, a ranking and thus an ordinal pattern cannot be derived.

The full impact of sampling at a fixed phase is seen in figure 3(b). First, we follow the dotted line to discuss the variation of a specific delay time τ_D. This is equivalent to changing the extent of the time comb in figure 3(a). For the two indicated time combs, the PE is maximum and minimum respectively. The difference in PE (as depicted by red/green symbols in figure 3(b)) affects cases with the previously discussed sampling at the same phase of the noisy sine wave (min. PE) and at the largest variation of phases (max. PE). Expanding the discussion to an extended range of frequencies, provides a similar impact of sampling at a fixed phase but at different frequencies. This result is seen as a consecutive structure of 'fringes' in figure 3(b).

In practical application, combinations of frequencies f and delay times τ_D leading to normalized PE ≈ 1 are not sensitive to detect irregularities. The white line shows the choice

made for the analysis in this paper, indicating a 'blind spot' for $f \gtrsim 9$ kHz (but without impact on the results). A strategy to overcome this deficiency is obvious: adding an additional PE analysis at a somewhat higher τ_D does ensure a coverage at any blind spot for any frequency of interest.

3. Multivariate measurements on Wendelstein 7-X for the characterization of plasma states

Now, we analyze data taken from plasma discharges of Wendelstein 7-X. The specific discharges were conducted with magnetic field settings that lead to a comparably low value of the rotational transform (low ι configuration). Figure 4 shows an overview of the temporal evolution of some plasma quantities from the plasma discharge 20180829.24. The salient feature apparent from the waveforms of the electron temperature is a step-wise increase of T_e at the discharge time $t \approx 1.4$ s. In this time range, the heating power P and diamagnetic energy W_{dia} are roughly constant. The sudden increase is observed with different diagnostics namely the electron cyclotron emission (ECE) and the SXR diagnostic both measurements depending on T_e. This has been observed to occur in various plasma discharges with different plasma conditions.

To examine the spatio-temporal characteristics of the temperature change, highly sampled data from an ECE radiometer and from SXR diagnostic have been analyzed by means of PE. The measurements are described along with the analysis in the subsequent sections.

3.1. ECE radiometer at W7-X

ECE radiometry is employed to measure the electron temperature and to derive T_e profiles in stellarators [9]. The ECE diagnostic in W7-X measures the second harmonic X-mode emission at 2.5 T employing a heterodyne radiometer. The radiometer has a line of sight (LoS) that crosses the plasma center and consists of 32 signal channels that measure a frequency range from 126 to 162 GHz. With a sampling frequency up to 2 MHz for temperature localized measurements, the LoS of the radiometer is selected such that the magnetic field increases monotonically towards the inner side of the torus. The correspondence of electron gyration frequency and $|B|$, allows one a spatial 1D allocation of temperature measurements. Here, the specific assignment of frequencies to $|B|$ is made by calculating the field with VMEC [11], a 3D MHD equilibrium code that solves the force balance of ideal MHD equilibrium.

The ray-tracing code TRAVIS [15] is used to take into account the radiation transport effects when the temperature profiles are derived from emissivity signals of the ECE channels. The complex refraction index along the line of sight of the ECE radiometer is determined from electron density and -temperatures, mapped on magnetic coordinates from a MHD equilibrium provided by the VMEC code. Technically, the radiation transport equation is solved such that the emissivity (and thus the respective temperatures profiles) contributing to each frequency channel is predicted. TRAVIS provides the location and width of the emissive layer at any frequency. This results in a typical spatial mapping to an effective measurement location (see figure 5(a)) with a resolution of the ECE measurements of $\Delta r \approx 1$ cm in the plasma core.

3.2. Analysis of ECE time series

The T_e profile for the discharge presented in figure 4 is shown in figure 5(a). Negative values of the normalized effective radius ρ represent regions in the high-field side (inboard), whereas positive values, regions in the low-field side (outboard). The ECE channel numbers are depicted above each data point in the profile[3]. We obtain a peaked T_e profile and also observe that measurements retrieved through ECE channel-#10 originate near the plasma core, in the high T_e gradient region on the low-field side.

The time series of measurements of channel-#10 ($\rho = 0.20$) recorded at a sampling frequency of 1 MHz is displayed in figure 5(b). Figure 5(b) shows a significant fluctuation level, partly due to measurement noise and partly due to plasma fluctuations. The signal is seen to fluctuate during the plasma start up phase, before reaching a stationary state from $t \approx 0.4$ s to $t \approx 1.4$ s. A spontaneous T_e increase is observed at $t \approx 1.4$ s (red dashed line) and the temperature is maintained at a second plateau in a stationary phase until $t = 2.5$ s when the heating power is modulated.

As a first step analysis, PE was applied to the waveform shown in figure 5(b). This analysis, as well as all further PE analysis presented in this paper, is conducted with a permutation order $m = 4$, an embedding delay $\tau_D = 0.1$ ms and a sliding time window of 10 ms of duration for the computation of each PE value. The calculations are performed in *MATLAB* with a code based on the algorithm presented in [22]. To claim that structures within a signal have been detected by means of PE, a detection threshold must be defined. The choice of the threshold is oriented at a significance level derived from noise data analysis. More specifically this is done by calculating the PE of 100 time series consisting of Gaussian noise ($\xi \sim N(0,1)$) with the same hyperparameters ($m = 4$, $\tau_D = 0.1$ ms) as for the analysis of experimental data. Obviously, the most appropriate threshold will depend on the data to be analyzed, but the method employed here is well-suited to establish a basis for an analysis that requires minimal previous knowledge about the nature of the data.

The result of the PE analysis is shown in figure 6(a). For an intuitive illustration of the results, the complement of the PE

[3] It can be seen that the data of channel-#16 ($\rho = -0.08$) is absent. This is due to the heating system operating at about the same frequency (≈140 GHz) as channel-#16 and thus, the measurements obtained through this channel are not accurate and are therefore neglected.

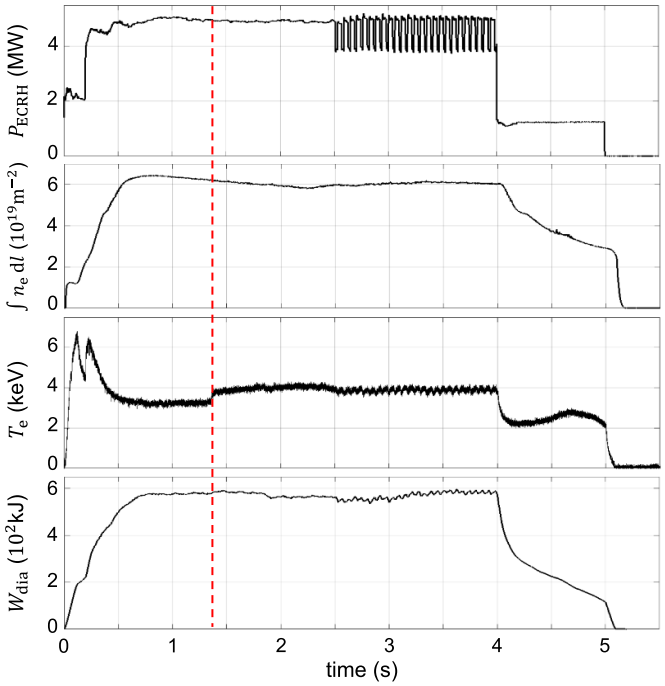

Figure 4. Time traces of low iota configuration plasma discharge 20180829.24 in W7-X (configuration B: DBM, $\langle\beta\rangle = 1.02\,\%$). Top-to-bottom: electron cyclotron heating power P_{ECRH}, line integrated electron density n_e, electron temperature T_e and diamagnetic energy W_{dia}. Red dashed line pinpoints the instant when a spontaneous T_e increase is observed.

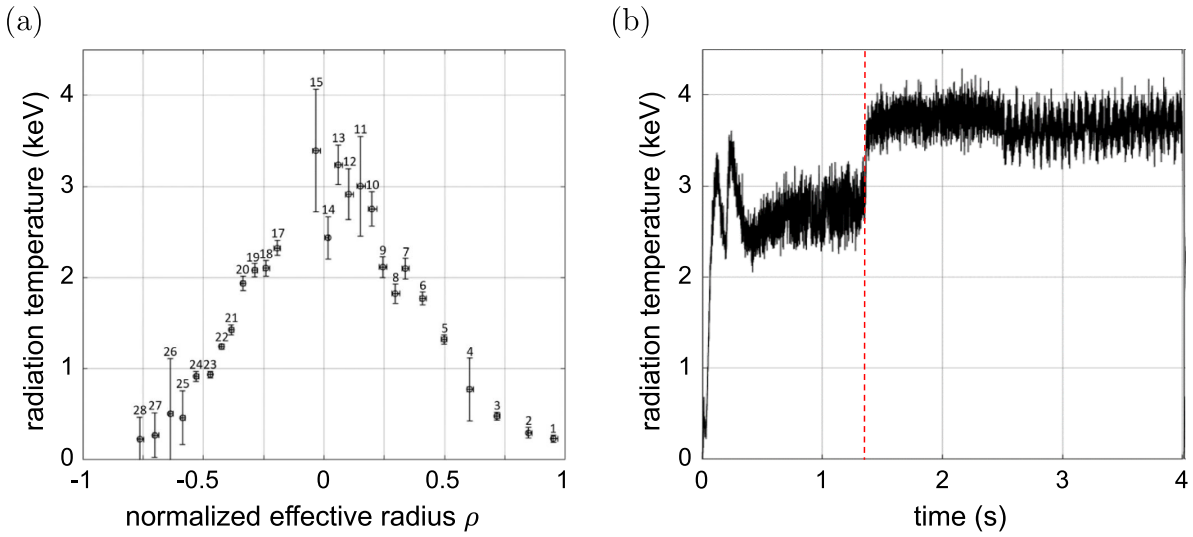

Figure 5. (a) Measured radiation temperature mapped to the effective measurement location calculated by TRAVIS in W7-X plasma discharge 20180829.24 from $t = 1.00$ s to $t = 1.05$ s. The number above each data point indicates the respective ECE channel number. (b) Radiation temperature measurements obtained through channel-#10 for the whole plasma discharge. Red dashed line indicate instant when spontaneous T_e increase is observed.

values normalized through the maximum achievable entropy is plotted in logarithmic scale. The significance threshold is $\log(1-s) = 6.7\times10^{-4}$ (magenta line) and this value, multiplied by three to prevent random spikes in PE that would yield false positive detections, results in the detection threshold for the T_e transition (blue line). We observe that the PE (black line) for $t \in [0.4, 1.4]$ s is significantly greater than for $t > 1.4$ s and is above both thresholds. This is also observed for a moving average (MA) over a sliding window of 10 ms of the PE signal (green line). Hence, PE differentiates between the time

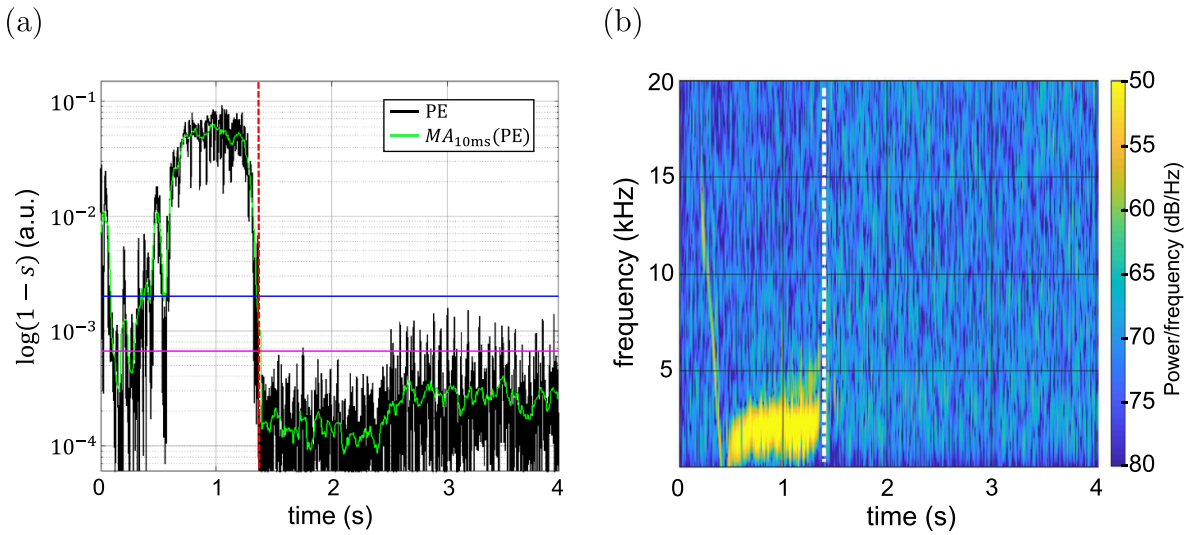

Figure 6. Analysis of ECE measurements obtained through signal channel-#10 from plasma discharge 20180829.24 in W7-X. Red and white dashed lines indicate instant when spontaneous T_e increase is observed. (a) Permutation entropy analysis with permutation order $m = 4$ and embedding delay $\tau_D = 0.1$ ms for a sliding time window of 10 ms of duration: complement of PE values s normalized through the maximum achievable entropy. Green line depicts moving average over a sliding window of 10 ms, magenta line displays the PE significance threshold derived from noise data analysis and blue line indicates the T_e transition detection threshold. (b) Spectrogram computed through the short-time Fourier transform.

windows before and after the T_e transition (red dashed line). Other than this distinction, a small fluctuation in the PE analysis can be seen between the significance threshold and the T_e transition detection threshold, which goes along with the heating power modulation starting from $t \approx 2.5$ s. This MA signal is better suited than the original one to serve as *trigger signal* in data mining by avoiding large excursions (compare black and green lines in figure 6(a)). Figure 6(b) shows spectral analysis of the same data computed using the *MATLAB* built-in function *spectrogram* [16]. All spectral analysis throughout this paper are computed through short-time Fourier transforms with hamming windows of 1 ms of duration and 10% overlap between windows. Through spectral analysis, a quasi coherent mode structure with a fluctuation frequency between $f \approx 2$ kHz and $f \approx 3$ kHz is observed between $t \approx 0.4$ s and $t \approx 1.4$ s, which vanishes once the T_e transition occurs (white dashed line). Comparing figures 6(a) and (b) indicates a smaller PE to result from a more ordered time series as reflected by the spectral maximum in figure 6(b).

Now, PE and spectral analysis are applied on all signal channels such that the localization of the structures can be examined. Both methods identified the same ECE channels for which the MHD activity was observed (see figure 7). More specifically, the T_e profile is separated in three types of regions: MHD activity was observed (red shaded area), no activity was observed (white shaded area) and no fluctuation analysis was possible due to low signal-to-noise ratio (gray shaded area). Even though, the analysis could not be performed in all regions, MHD activity is localized in the core. Thus, through both methods, spectral analysis and PE the mode activity could be allocated along the radial coordinate and lies in the high gradient region near the plasma core.

3.3. XMCTS at W7-X

The SXR multi-camera tomography system (XMCTS) measures emissions through pinhole cameras arranged in an up-down symmetry in a poloidal plane [3], where the flux surfaces have a triangular shape as seen in figure 8. This non-invasive diagnostic consists of 20 SXR cameras arranged poloidally to cover the plasma from multiple views and each camera is equipped with 18 photodiodes to measure radiation in the energy range of approximately 1–12 keV [4]. Figure 8 depicts the flux surfaces for the low-iota configuration B (DBM, $\beta = 1.02\%$) and the line of sight geometry of this diagnostic, specifically, the region covered by SXR camera 4A (gray shaded region) and LoS-#8 (white dashed line). The SXR emissions measured emerge mainly from contributions from bremsstrahlung ($\propto n_e^2 \sqrt{T_e}$), recombination and line radiation. Through high sampling rate and spatial resolution, this diagnostic is able to resolve MHD instabilities in a poloidal cross section of the plasma.

3.4. Analysis of SXR time series

The radiation recorded through SXR camera 4A for plasma discharge 20180829.24 in W7-X is presented in figure 9. The waveforms in red (figure 9(a)) correspond to measurements closest to the vessel's inner wall (high-field side), whereas

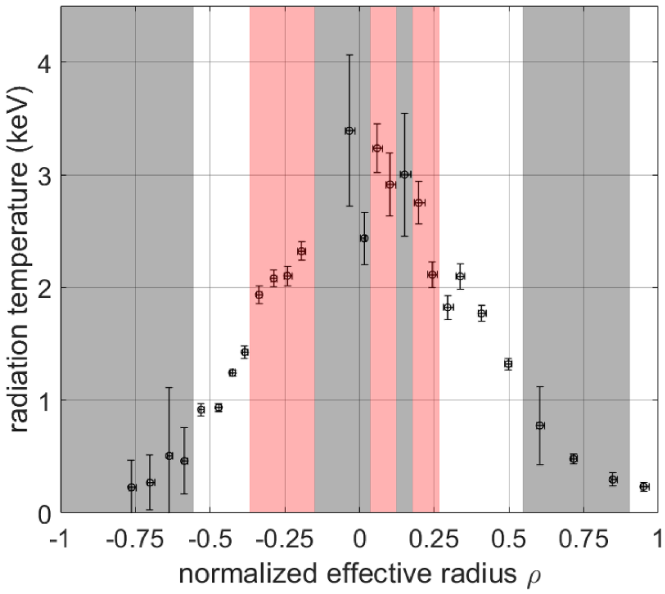

Figure 7. Measured radiation temperature mapped to the effective measurement location calculated by TRAVIS in W7-X plasma discharge 20180829.24 from $t = 1.00$ s to $t = 1.05$ s. Red shaded areas indicate regions where MHD activity was observed. In white are regions where this activity was not observed and, in gray, regions where no fluctuation analysis was possible due to low signal-to-noise ratio.

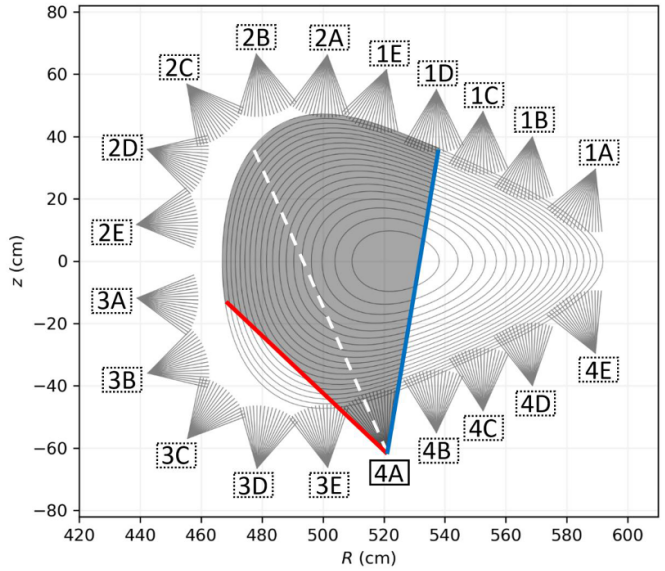

Figure 8. Setup of the soft x-ray multi-camera tomography system depicting the lines of sight geometry for all cameras installed ($1A - 4E$) and the flux surfaces for the low iota configuration B (DBM, $\beta = 1.02\%$). The shaded area indicates the region from which the radiation measured through SXR camera '4A' comes from (delimited by the red and blue lines). The white dashed line depicts one of the 18 lines of sight (specifically, LoS-#8) available for the measurements.

waveforms in blue, correspond to measurements near the plasma center (low-field side, see figure 8).

The channels showing the innermost lines of sight are arranged in figure 9(b), where SXR signals measuring similar intensities but from opposing locations w.r.t. the plasma center can be observed. When T_e spontaneously rises (black dashed line), a significant branching between the off-axis measurements and their respective counterparts is seen. Radiation measurements taken more laterally on the high-field side (red) drop slightly in intensity, whereas measurements in the low-field side (blue) increase. The difference in the temporal evolution of the ECE and SXR signals provides clear evidence that the emissivity change is localized to the innermost channels. Moreover, while ECE data showed a

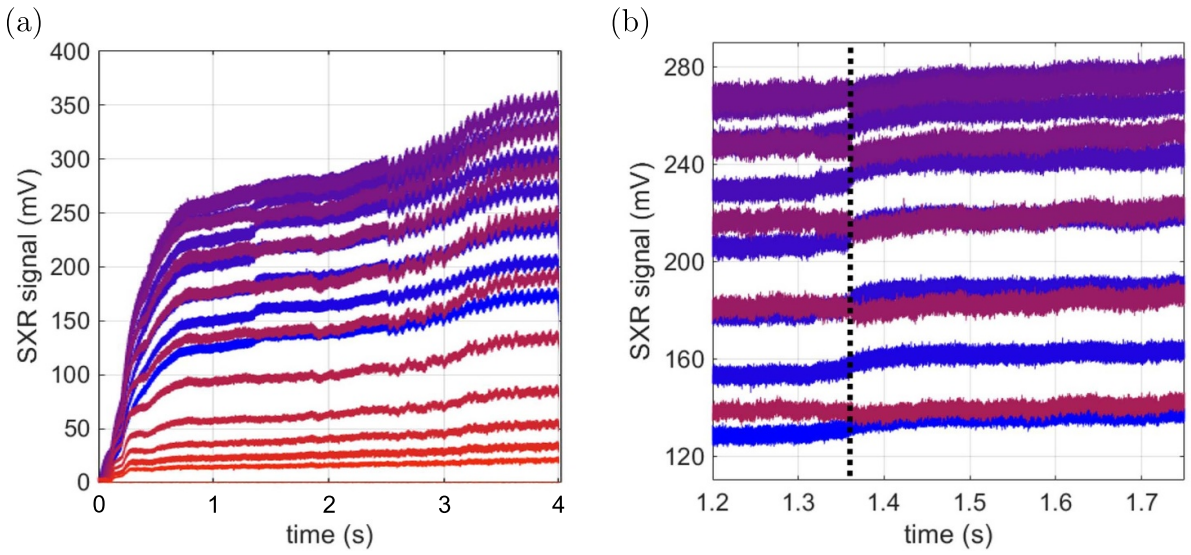

Figure 9. SXR measurements obtained through camera '4A' for low-iota configuration plasma discharge 20180829.24 in W7-X. Black dashed line indicates instant when spontaneous T_e increase is observed. (a) Time series recorded through all 18 photodiodes: red plots depicting lines of sight closer to the periphery and blue plots, regions near the plasma center (see figure 8). (b) Magnification of (a) exhibiting bifurcation of signals after T_e transition.

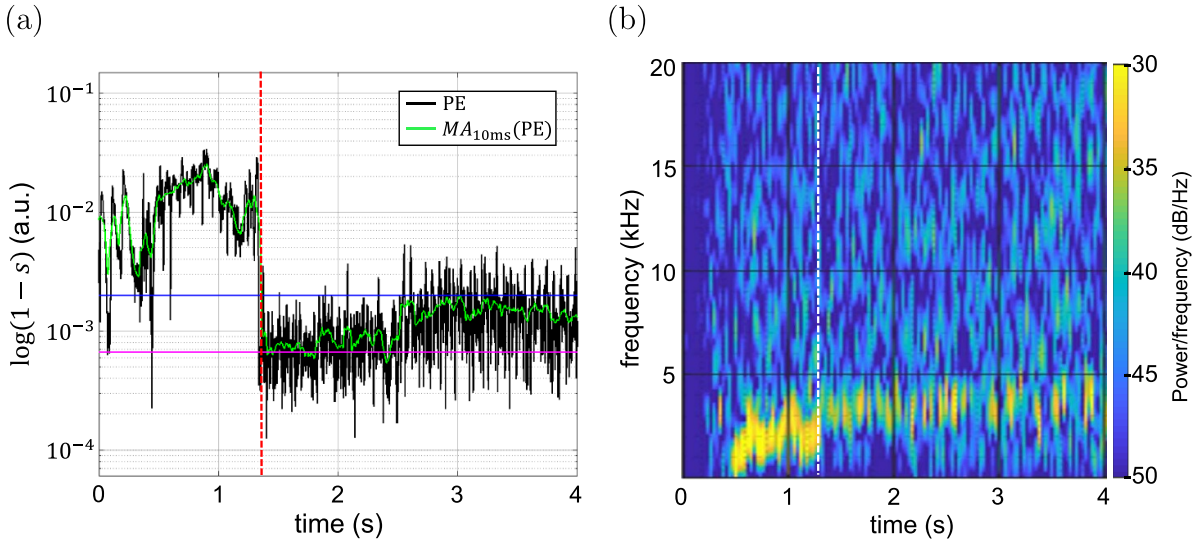

Figure 10. Analysis of SXR measurements obtained through camera '4A' LoS-#8, from plasma discharge 20180829.24 in W7-X. Red and white dashed lines indicate instant when spontaneous T_e increase is observed. (a) Permutation entropy analysis with permutation order $m = 4$ and embedding delay $\tau_D = 0.1$ ms for a sliding time window of 10 ms of duration: complement of PE values s normalized through the maximum achievable entropy. Green line depicts moving average over a sliding window of 10 ms, magenta line displays the PE significance threshold derived from noise data analysis and blue line indicates the T_e transition detection threshold. (b) Spectrogram computed through short-time Fourier transform revealing two structures around 1–4 kHz.

localized increase of T_e, the simultaneously observed increase and decrease of SXR signal ($I \propto n_e^2 \sqrt{T_e}$) allows one to conclude that the plasma density is also affected in the transition.

Now we conduct the PE and spectral analysis of density and temperature dependent SXR data. The PE analysis ($m = 4$, $\tau_D = 0.1$ ms) as shown in figure 10(a) indicates a clear detection of structures from plasma start up until the instance when the T_e suddenly rises (red dashed line). From this point in time onwards, PE (black line) detects structures with values above the significance threshold (magenta line) and partially above the T_e detection threshold (blue line). The MA signal also detects structures above the significance threshold after $t \approx 1.4$ s, however not above the T_e detection threshold. The

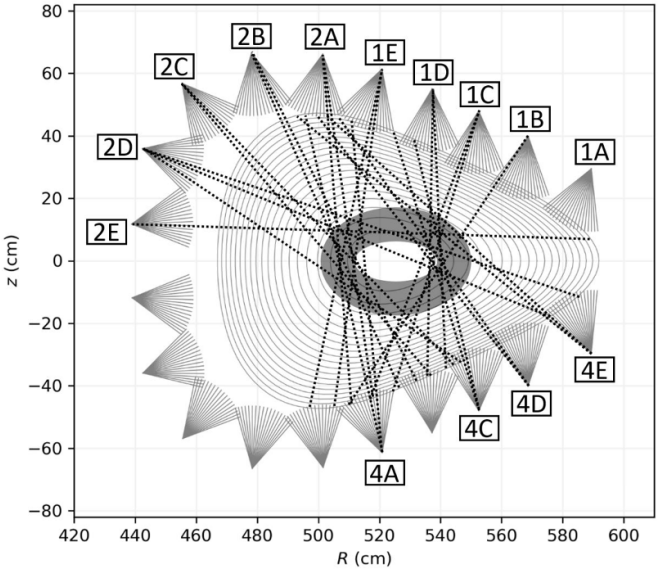

Figure 11. Localization of MHD activity in the poloidal plane through the analysis of SXR measurements in W7-X from plasma discharge 20180829.24. Dotted lines indicate lines of sight that detected MHD activity. Gray area depicts locations with high density of crossings between dotted lines suggesting a 2D localization of the activity.

next step is to identify the structures detected by the PE analysis through spectral analysis as shown in figure 10(b). From this analysis, two different structures can be observed: a predominant structure in a frequency range between $f = 2$ kHz and $f = 4$ kHz taking place between plasma start up and the T_e transition (white dashed line), and a weaker in power structure in the slightly higher frequency range (up to $f \approx 5$ kHz) that extends until the plasma is terminated. The second structure is observed in the PE analysis, however it lies between the significance threshold and the T_e transition detection threshold. Therefore, the distinction between the time windows before and after the T_e transition is in the PE analysis most evident. The computing times for both of the analysis lie around $t_{PE} = 0.49$ s and $t_{FFT} = 13.83$ s, PE being clearly faster for the analysis of a ≈5.5 s long time series with a sampling rate of 2 MHz.

Assuming that SXR data is proportional to $n_e^2\sqrt{T_e}$, the ECE signals increase only locally and the total energy did not show any response, the previous observations are an indication of the redistribution of the electron pressure. Expanding the time series analysis to all available photodiodes from the SXR cameras, the MHD activity can be localized in a 2D tomographic reconstruction of a plasma poloidal cross section. The cameras available for the analysis are shown in figure 11 as well as the lines of sight (black dotted lines) where the activity preceding the transition was identifiable by means of PE. From the density of the crossings between these lines of sight, a rough idea of the spatial location of the activity can be estimated (gray shaded area). From this simplistic tomographic approach, we observe that the activity lies near the plasma center as also observed from the ECE data analysis in the previous section. Furthermore, the tomographic approach allows to exclude the occurrence of the MHD activity on axis.

3.5. Data mining with PE

Now, we investigate the capability of PE for an automated detection of the temperature transition in a large amount of data. Purpose of the survey is to identify the temperature bifurcation using PE at settings used in the previous section ($m = 4$, $\tau_D = 0.1$ ms). The success rate of this data mining procedure is derived from the number of successful detections of transitions and the number of false positives. To this end, ECE data from a full shot-day on W7-X are analyzed to detect spontaneous T_e transitions. Since the ECE signal reflects the electron temperature rather than a combination of n_e and T_e, only the ECE data were chosen for the data mining. The PE (at fixed settings) was applied on ECE data. Table 1 summarizes for which discharges PE detected a temperature transition at a time t_B, and for which discharges no transition could be revealed. The result from the findings in table 1 is that PE detects all transitions and no false positives were found, i.e. PE detected all T_e transitions when using a MA of the PE over a sliding time window of 10 ms as trigger signal as shown in figures 6 and 10. It is worth mentioning, that the first 100 ms of every discharge were excluded from the analysis to avoid the detection of structures due to plasma start-up.

To validate the PE findings, spectrogram were calculated and the frequency of a accompanying MHD activity is added to the transition in table 1. The spectrogram confirms the transition detected by PE but also reveals additional power spectral densities also at varying frequencies. E.g. in figure 6, a peak steadily decreasing in the plasma build-ip phase adds signal contributions that are hard to discriminate. Even more ambiguous, the spectrogram of soft-x ray data (figure 10) indicate spurious power spectral densities even after the T_e transition. Contrasting the more ambiguous findings from spectral analysis, the PE analysis offered (for the data set in table 1)

Table 1. Detection of spontaneous T_e transitions in bulk data analysis by means of permutation entropy in W7-X plasma discharges. PE analysis performed on ECE data with permutation order $m = 4$ and embedding delay $\tau_D = 0.1$ ms for a sliding time window of 10 ms. Plasma parameters before the transition are shown, including increment in electron temperature measured through ECE signal channel-#10 $\Delta T_{e,ch\#10}$ and frequency f of the MHD activity preceding the transition.

Discharge ID	P_{ECRH} (MW)	$\int$ ndl (10^{19} m^{-2})	t_B (s)	T_e (keV)	ΔT_e (keV)	f_{ECE} (kHz)
20180829.5	5.0 ± 0.2	4.0 ± 0.1	—	—	—	—
20180829.6	5.02 ± 0.02	3.43 ± 0.03	3.7 ± 0.1	3.35 ± 0.02	0.25 ± 0.02	7.5–10
20180829.7	3.71 ± 0.02	3.01 ± 0.03	2.1 ± 0.1	3.47 ± 0.03	0.33 ± 0.03	7–10
20180829.8	3.62 ± 0.02	2.60 ± 0.03	0.9 ± 0.1	4.04 ± 0.02	0.17 ± 0.02	7–8
20180829.9	3.72 ± 0.03	2.21 ± 0.02	0.9 ± 0.1	4.22 ± 0.04	1.00 ± 0.04	7–8
20180829.12	3.5 ± 0.2	2.0 ± 0.2	—	—	—	—
20180829.15	2.0 ± 0.1	2.0 ± 0.2	—	—	—	—
20180829.16	5.01 ± 0.02	5.01 ± 0.02	1.6 ± 0.1	2.79 ± 0.02	0.64 ± 0.02	2.5–3.5
20180829.17	4.83 ± 0.03	3.61 ± 0.03	1.2 ± 0.1	2.83 ± 0.01	0.18 ± 0.01	8–9
20180829.18	3.50 ± 0.02	4.10 ± 0.02	1.7 ± 0.1	2.81 ± 0.02	0.38 ± 0.02	8–9
20180829.19	3.5 ± 0.2	4.0 ± 0.2	—	—	—	—
20180829.20	2.42 ± 0.03	4.02 ± 0.02	2.6 ± 0.1	1.94 ± 0.03	0.37 ± 0.03	2–2.5
20180829.21	3.55 ± 0.02	4.34 ± 0.01	1.2 ± 0.1	3.02 ± 0.02	0.37 ± 0.02	2–3
20180829.22	1.5 ± 0.1	4.0 ± 0.2	—	—	—	—
20180829.23	2.0 ± 0.1	4.0 ± 0.1	—	—	—	—
20180829.24	4.97 ± 0.04	6.23 ± 0.04	1.4 ± 0.1	2.81 ± 0.03	0.85 ± 0.03	2–3
20180829.25	5.04 ± 0.03	6.33 ± 0.03	1.1 ± 0.1	1.81 ± 0.03	0.31 ± 0.03	2–3
20180829.26	5.00 ± 0.04	6.30 ± 0.04	1.4 ± 0.1	2.94 ± 0.03	0.88 ± 0.03	2–3
20180829.27	3.61 ± 0.03	6.04 ± 0.03	2.5 ± 0.1	2.57 ± 0.03	0.41 ± 0.03	4–5
20180829.28	3.6 ± 0.2	6.0 ± 0.5	—	—	—	—
20180829.31	3.5 ± 0.2	6.5 ± 0.4	—	—	—	—
20180829.32	2.04 ± 0.02	3.98 ± 0.02	2.1 ± 0.1	1.75 ± 0.02	0.25 ± 0.02	3–4
20180829.33	2.05 ± 0.02	3.96 ± 0.02	2.1 ± 0.1	1.82 ± 0.03	0.25 ± 0.03	3–4
20180829.34	2.05 ± 0.01	4.05 ± 0.02	2.0 ± 0.1	1.74 ± 0.03	0.23 ± 0.03	2–4
20180829.35	2.02 ± 0.01	3.73 ± 0.01	2.3 ± 0.1	1.80 ± 0.03	0.25 ± 0.03	1–4
20180829.37	5.0 ± 0.2	4.3 ± 0.1	—	—	—	—
20180829.38	5.4 ± 0.2	3.5 ± 0.1	—	—	—	—
20180829.39	5.36 ± 0.04	2.08 ± 0.01	1.6 ± 0.1	4.81 ± 0.04	1.21 ± 0.04	6–7
20180829.40	5.51 ± 0.03	2.02 ± 0.03	3.1 ± 0.1	5.02 ± 0.03	1.06 ± 0.03	5–9
20180829.41	5.3 ± 0.2	2.1 ± 0.1	—	—	—	—
20180829.42	5.3 ± 0.1	2.3 ± 0.2	—	—	—	—
20180829.43	5.3 ± 0.2	2.6 ± 0.2	—	—	—	—
20180829.48	3.0 ± 0.3	3.5 ± 0.1	—	—	—	—
20180829.49	3.1 ± 0.2	3.0 ± 0.1	—	—	—	—
20180829.50	3.2 ± 0.2	3.1 ± 0.2	—	—	—	—

a straight forward setting to define a statistically based PE threshold for systematically detecting previously unrevealed temperature bifurcation.

4. Discussion and conclusions

The applicability of PE as a first step analysis on plasma data over a wide range of plasma conditions is broad and serves as a tool to rapidly and robustly detect underlying dynamics in time series. This method allowed an automated analysis of bulk data, which resulted in the detection of localized T_e fluctuations in low-iota configuration discharges in W7-X. PE as a data mining tool allowed one to discriminate shots with and without T_e transitions. The results of the PE analysis were supported by spectral analysis, however, the former was shown to be at least 25 times faster to compute for this case. We conclude that PE can be used to report changes of the plasma state as fast as $\sim$0.1 ms which is much faster than the energy confinement time ($\sim$100 ms). It is concluded that PE has a large potential for real-time detection of changes to, e.g. induce interlocks or other control interventions.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No. 101052200—EUROfusion). Views and opinions expressed are however those of the author(s) only and

do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

ORCID iDs

J F Guerrero Arnaiz https://orcid.org/0000-0002-7650-2918
A Dinklage https://orcid.org/0000-0002-5815-8463
J Geiger https://orcid.org/0000-0003-4268-7480
C Brandt https://orcid.org/0000-0002-5455-4629
J Schilling https://orcid.org/0000-0002-6363-6554
K Rahbarnia https://orcid.org/0000-0002-5550-1801
M Zanini https://orcid.org/0000-0002-8717-1106

References

[1] Bandt C and Pompe B 2002 Permutation entropy: a natural complexity measure for time series *Phys. Rev. Lett.* **88** 174102
[2] Bandt C and Shiha F 2007 Order patterns in time series *J. Time Ser. Anal.* **28** 646–65
[3] Brandt C, Broszat T, Thomsen H, Laube R, Marquardt M, Franz P, Schülke M, Sieber T and Weißflog S 2017 Installation of the soft x-ray multi-camera tomography system (XMCTS) in the Wendelstein 7-X stellarator *Fusion Eng. Des.* **123** 887
[4] Brandt C *et al* 2020 Soft x-ray tomography measurements in the Wendelstein 7-X stellarator *Plasma Phys. Control. Fusion* **62** 035010
[5] Cuesta-Frau D, Murillo-Escobar J P, Orrego D and Delgado-Trejos E 2019 Embedded dimension and time series length. Practical influence on permutation entropy and its applications *Entropy* **21** 385
[6] Dudewicz E, Gyorfi L and van der Meulen E 1997 Nonparametric entropy estimation: an overview *Int. J. Sci. Res. Math. Stat. Sci.* **6** 17
[7] Ferreira D R, Carvalho P J, Sozzi C, Lomas P J and Contributors J 2020 Deep learning for the analysis of disruption precursors based on plasma tomography *Fusion Sci. Technol.* **76** 901–11
[8] Grieger G *et al* 1992 Physics optimization of stellarators *Phys. Fluids* B **4** 2081
[9] Hartfuss H J, Geist T and Hirsch M 1997 Heterodyne methods in millimetre wave plasma diagnostics with applications to ECE, interferometry and reflectometry *Plasma Phys. Control. Fusion* **39** 1693
[10] Henry M and Judge G 2019 Permutation entropy and information recovery in nonlinear dynamic economic time series *Econometrics* **7** 10
[11] Hirshman S, van RIJ W and Merkel P 1986 Three-dimensional free boundary calculations using a spectral green's function method *Comput. Phys. Commun.* **43** 143
[12] Jordan D, Stockmanns G and Kochs E 2007 Permutation entropy of the EEG indicates increase and decrease of the anesthetic level *Anesthesiology* **107** A800
[13] Li X, Cui S and Voss L 2008 Using permutation entropy to measure the electroencephalographic effects of sevoflurane *Anesthesiology* **109** 448
[14] Maggs J E, Rhodes T L and Morales G J 2015 Chaotic density fluctuations in L-mode plasmas of the DIII-D tokamak *Plasma Phys. Control. Fusion* **57** 045004
[15] Marushchenko N, Turkin Y and Maassberg H 2014 Ray-tracing code TRAVIS for ECR heating, EC current drive and ECE diagnostic *Comput. Phys. Commun.* **185** 165
[16] *MATLAB and Statistics Toolbox Release 2018a* (Natick, MA: The Mathworks, Inc.) (available at: www.mathworks.com/help/signal/)
[17] Nührenberg J and Zille R 1986 Stable stellarators with medium β and aspect ratio *Phys. Lett.* A **114** 129
[18] Olofsen E, Sleigh J and Dahan A 2008 Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect *Br. J. Anaesth.* **101** 810
[19] Onchi T, Fujisawa A, Sanpei A, Himura H and Masamune S 2017 Permutation entropy and statistical complexity in characterising low-aspect-ratio reversed-field pinch plasma *Phys. Scr.* **92** 055601
[20] Škvára V, Šmídl V, Pevný T, Seidl J, Havránek A and Tskhakaya D 2020 Detection of alfvén eigenmodes on COMPASS with generative neural networks *Fusion Sci. Technol.* **76** 962–71
[21] Staniek M and Lehnertz K 2007 Parameter selection for permutation entropy measurements *Int. J. Bifurc. Chaos* **17** 3729–33
[22] Unakafova V A and Keller K 2013 Efficiently measuring complexity on the basis of real-world data *Entropy* **15** 4392–415
[23] Wolf R C *et al* 2017 Major results from the first plasma campaign of the Wendelstein 7-X stellarator *Nucl. Fusion* **57** 102020
[24] Zhang N, Sun Y, Zhang Y, Yang P, Lin A and Shang P 2019 Distinguishing stock indices and detecting economic crises based on weighted symbolic permutation entropy *Fluct. Noise Lett.* **18** 1950026
[25] Zunino L, Fernández Bariviera A, Guercio M B, Martinez L B and Rosso O A 2012 On the efficiency of sovereign bond markets *Physica* A **391** 4342