Extraction of quasi-coherent modes based on reflectometry data

Luigui Salazar, Stéphane Heuraux, Roland Sabot, Andreas Krämer-Flecken, Tore Supra Team

Extraction of quasi-coherent modes based on reflectometry data

Luigui Salazar[1,2,*], Stéphane Heuraux[1], Roland Sabot[2], Andreas Krämer-Flecken[3], Tore Supra Team and TEXTOR Team

[1] IJL UMR 7198 CNRS, Université de Lorraine, F-54000 Nancy, France
[2] CEA, IRFM, F-13108 Saint-Paul-Lez-Durance, France
[3] Institut für Energie- und Klimaforschung/Plasmaphysik, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany

E-mail: luigui.salazar@cea.fr

Received 16 March 2022, revised 20 June 2022
Accepted for publication 20 July 2022
Published 6 September 2022

Abstract

The identification of turbulence sources would drive to a deeper understanding of confinement dynamics in tokamak plasmas. Turbulence results from a mixture of instabilities corresponding to sources at different timescales and spatial scales. Using poloidal correlation reflectometry and multi-pin Langmuir probe, it was shown in the T-10 and the Tokamak Experiment for Technology Oriented Research (TEXTOR) tokamaks that the reflectometry frequency spectrum is the superposition of several components: broadband component, quasi-coherent (QC) modes and low-frequency components. The relevance of QC modes is associated with their link with the trapped electron mode instability. This link was exhibited in the transition from the linear ohmic confinement (LOC) to the saturated ohmic confinement (SOC) regime. A method is presented in this paper to extract the QC mode component from the reflectometry data, enabling its separation from the broadband component and the study of its time evolution. It is a first step toward the discrimination of turbulence sources. The central idea explores a way to combine the approach of signal processing and machine learning. The continuous wavelet transform on the basis of complex Morlet wavelet has proved to be efficient in providing a decomposition of a signal at different scales over time for fluctuation tackling; clustering techniques, such as the mini-batch *K*-means, are able to tackle clusters at different scales. The method was applied to Tore Supra and TEXTOR reflectometry data. In Tore Supra, the amplitude of the extracted QC mode component decreases during the LOC–SOC transition. In TEXTOR, the amplitude of the coherent spectra of the extracted QC mode component is similar to the experimental coherent spectra obtained through correlation reflectometry. The developed method permits the extraction of components, preserving their physical and statistical properties.

Keywords: plasma turbulence, reflectometry, diagnostic, tokamak

1. Introduction

Turbulence is the root of confinement degradation in tokamaks because it induces anomalous transport of energy and particles [1]. It is strongly nonlinear and the result of instabilities, such as ion temperature gradient (ITG), trapped electron mode (TEM) and electron temperature gradient (ETG). The energy sources generating these instabilities are believed to reign in turbulence transport at different regimes for time scale and spatial scale. Consequently, tackling these sources would lead to a better understanding of the dynamics among instabilities. According to harmonic analysis, they are the superposition of various spectra at several frequency bands [2]. It is generally

assumed that there is no or unknown coupling among these various instabilities [3], although geodesic acoustic modes and zonal flows are two examples of instabilities that exchange with turbulence. A prior step is the separation of the different components from experimental turbulence data.

Among the different diagnostics able to probe turbulence and fluctuations of the plasma parameters, reflectometry, a radar-like technique based on the propagation of electromagnetic waves in plasmas, has been found to be very versatile [4] for the analysis of density fluctuations, as it requires simple access to the plasma and provides good time and space resolution.

For the purpose of investigating turbulence properties, a study was carried out using poloidal correlation reflectometry (PCR), multi-pin Langmuir probe and heavy ion beam probe in the T-10 tokamak [5]. This work has succeeded in splitting the frequency spectra into several components: broadband, low- and high-frequency quasi-coherent (QC) modes, low-frequency components and some oscillations at 15–30 kHz. Similar observations were reported in the Tokamak Experiment for Technology Oriented Research (TEXTOR) [6] and Tore Supra [7] tokamaks. In both tokamaks, the amplitude of the high-frequency QC mode is much lower than the low-frequency one.

It is relevant to have a deeper look at the low-frequency QC modes. It was shown that they are a signature of the TEM instability [8]. Indeed, disappearance of these modes was reported in several tokamaks [9, 10] during transitions from the linear ohmic confinement (LOC) to the saturated ohmic confinement (SOC) regime. The SOC–LOC transition is attributed to the stabilization of the TEM [11]. However, the results from [12] were based on an average behavior, excluding the short timescale evolution of the frequency spectra.

A complementary study has been performed on TEXTOR. The QC modes exhibit short- and long-range correlations, as shown by the significant values of the magnitude squared coherence (MSC) between antennas at long poloidal and toroidal distances [6, 13]. Long time series permit to observe this long-range coherence and in the following offer the possibility to investigate the short timescale behavior.

These first investigations on spectrum decomposition were continued at the Tore Supra tokamak. The aim was to automatically decompose the reflectometry frequency spectrum into a set of four components [14]—direct current component, low-frequency fluctuations, broadband turbulence and noise level—to perform systematic analysis. A database of 6000 discharges and more than 350 000 reflectometry measurements was built to extract trends from statistical analysis instead of a dedicated shot-to-shot analysis [15]. Nevertheless, this decomposition includes some assumptions, such as stationary turbulence behavior during the reflectometry probing time ($\approx$10 ms) for each acquisition. Moreover, side-lobe components, such as QC modes and high-frequency magnetohydrodynamic modes, were integrated into the broadband component.

From this point of view, the long-term objective is to overcome the stationary assumption, and thereby offer a way to tackle the dynamics of the turbulence components. The continuous wavelet transform (CWT) was used to capture density fluctuation properties from the reflectometry raw complex signals at various scales [16].

The extraction of the low-frequency QC mode component is based on a screening technique applied to the CWT decomposition. The screening is composed of two threshold functions: cross-correlation (CC) threshold and density energy threshold. A union of the patterns of both screening domains is chosen to combine correlation and energy density (ED) properties. However, this choice can include irrelevant elements for QC mode characterization.

To finish the QC mode extraction, we rely on artificial intelligence (AI). AI is an area that has been evolving fast during the last years. The first step of the AI procedure is to find a vector space for merging properties of CWT similar to the density energy and wavelet scales. The second AI step is the mini-batch K-means [17] (unsupervised clustering technique) to tackle clusters corresponding to the QC modes.

More 'classical' methods, such as short-time Fourier transform (STFT), principal component analysis [18] and variational auto-encoder neural network [19], were first explored, but they did not succeed because of the prohibitive calculation time for an application to a large database and the low signal-to-noise ratio (SNR).

Our method has been successfully applied to Tore Supra and TEXTOR reflectometry data for extracting the QC mode components. The decrease of the QC mode amplitude during LOC–SOC transition in Tore Supra is recovered. In TEXTOR, the MSC obtained from two poloidaly separated antennas is very similar to the coherence obtained after applying the QC mode extraction.

The extraction method is thus able to eliminate other components from the spectrum without affecting the QC component. In this sense, a better decomposition of the reflectometry data spectrum is reachable by removing the QC modes from the broadband component, offering a simpler signal to analyze for further studies.

2. Exploratory part and final data analysis method

2.1. General approach

Before finalizing the method, different attempts and approaches were tested to extract the spectrum components. Since the AI algorithms are numerous and diverse, we think it is worthy to briefly present the ones we tried and their shortcomings. Indeed, application of AI to reflectometry and more generally fusion plasma turbulence is still ongoing, and thus the most suitable methods have not yet been discovered.

2.1.1. Metric density-based spatial clustering of applications with noise. First, density-based spatial clustering of applications with noise (DBSCAN) [20] was tested to find the QC modes directly from the raw complex signal ($f(t)$) . DBSCAN is a technique to identify clusters in a multidimensional space.

To find a space where the different components would be separated, we tried the general relativity metric concept, in which the tensor field varies among elements [21]. The central idea was to find invariant values by optimizing α and β coefficients (1). After this process, DBSCAN was applied to seek for density clusters over the metric representation:

$$\text{Signal} : f(t) = A_{(t)}.e^{\phi_{(t)}}, dS^2 = \alpha_{(\phi)}.dA^2 + \beta_{(t)}.dt^2. \quad (1)$$

Although some structures emerged, it was difficult to identify the modes, with the main issue associated with the computing time and the huge quantities of data. We did not pursue this path in this first step.

2.1.2. Statistical methods. This statistical approach belongs to the classical methods, in which the idea is to smooth the raw signal spectrum to improve its shape and reduce the noise. The moving average filter and the local average filter were able to achieve some smoothing, but the uncertainty was too high due to the lack of information on the shape of the different spectral components.

To overcome these issues, a time–frequency representation (TFR) has the potential to separate the different components by their dynamics. It also allows access to short-time evolution.

2.2. Time–frequency representation

2.2.1. Short-time Fourier transform. A natural approach for TFR is STFT. STFT gives a 2D representation (image) on which convolutional filters $g(t)$ can be applied:

$$Sf(u,\varepsilon) = \langle f, g_{u,\varepsilon} \rangle = \int_{-\infty}^{+\infty} f(t)g(t-u)e^{-i\varepsilon t}dt$$
$$u : \text{time step}$$
$$\varepsilon : \text{frequency}. \quad (2)$$

Nevertheless, the Heisenberg box associated with this transform is constant along the frequency and time axes, resulting in a mix of time scales and frequency scales due to the convolution between the time-window fast fourier transform (FFT) and the observed frequencies leading to a blurred TFR. Wavelet transform [16] provides a variable Heisenberg box in the TFR.

2.2.2. Continuous wavelet transform. CWT [22] was chosen rather than the discrete wavelet transform (DWT), as it permits analysis of time series that contain a nonstationary power at many different frequencies [23], in which the scales can be adjusted to those of the observed fluctuations. The scales of the DWT are limited to a discrete set [24].

The parameters essential in the definition of this transform are the type of wavelet and the wavelet dictionary ($\psi_{u,s}$) (i.e all wavelet versions that are going to be used) [16]. This transform is based on a non-orthogonal analysis:

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right)dt$$
$$u : \text{time step}$$
$$s : \text{scale}. \quad (3)$$

- Wavelet

Because reflectometry signals are complex, a complex wavelet must be used to tackle angle (position) and module (energy). Two wavelets were brought out for this operation: the Paul wavelet and the Morlet wavelet [25]. The Paul wavelet is efficient for phase detection, whereas the Morlet wavelet is prioritized for detection of fluctuations, which was our main goal. The Morlet wavelet is defined as

$$\psi_{\text{Morlet}}(t) = |s|^{-0.5} \cdot \pi^{-0.25} \cdot e^{6(\frac{t \cdot f_s}{s})j} \cdot e^{-0.5(\frac{t \cdot f_s}{s})^2}$$
$$s : \text{scale}$$
$$f_s : \text{frequency sampling}. \quad (4)$$

- Dictionary

The wavelet dictionary is created from scales and time steps. In this whole process the time step is set to be the sampling period ($u = \delta(t)$), then scales are defined in such a way that the calculation time is minimized and the frequency band of concern is covered (20 kHz$-f_s/2$), removing the low-frequency components (0$-$20 kHz) :

$$D = \left\{ \psi_{u,s(t)} = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \right\}. \quad (5)$$

It should be pointed out that negative and positive scales are presented because of complex raw signals. Negative scales correspond to negative frequency bandwidth and likewise for positive scales.

The first approach in CWT was to take every scale and to perform independent component analysis at every time step to separate the CWT spectrum into additive subcomponents by assuming that they are non-Gaussian signals. The main constraints were the lack of precision for reconstruction through the inverse continuous wavelet transform (ICWT) of the CWT to obtain QC modes.

2.2.3. Other approach. This last approach relies on a variational auto-encoder network for performing encoding of CWT into a multivariate latent distribution and decoding for reconstruction along the time axis, the idea of which was to divide CWT into several windows to obtain a pattern window evolution. The main constraints were calculation time and learning process.

The CWT approach proved to be suitable to highlight patterns in TFR, but additional algorithms tackling different properties of the QC modes were needed to extract them.

2.3. Final approach and data analysis procedure

Finally, in this section, a description of the algorithm process from the reflectometry database to the extraction of QC

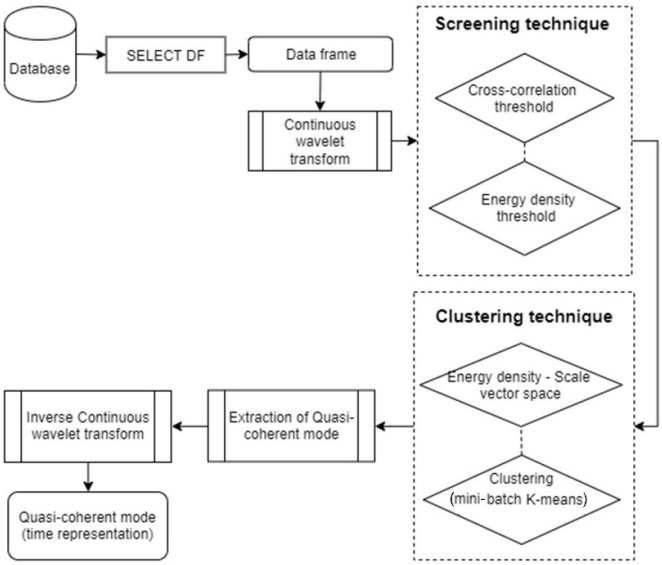

Figure 1. Data-flow diagram: rounded rectangle for data, rectangle double line for process and rhombus for proposed techniques.

modes is done. The high-frequency QC (HF-QC) modes, when detected, are observed to be around double the frequency peak of the low-frequency QC modes. For the rest of the paper, the QC mode term refers to the low-frequency QC modes (i.e. $f < 150$ kHz).

Complex raw signals ($f(t)$) are selected (data frame) based on predefined plasma conditions and undergo a CWT, a screening, a clustering and finally an ICWT. Figure 1 illustrates the algorithm procedure. As an example, this will be applied to a reflectometry acquisition (#32093 $t = [11.1151; 11.1397\text{s}]$; $F = 118.8$ GHz) that exhibits strong QC modes.

2.3.1. Screening technique. Once the CWT is obtained (figure 2), the chosen reference scale is the one for which the average ED over the time duration is a local maximum. This is called high-ED scale (s_{hed}), and it is located in the frequency range associated to the QC mode. This is done separately for negative (s^{-}_{hed}) and positive (s^{+}_{hed}) scales (see figure 2).

For screening, hard threshold techniques are used. The CC threshold that will be used for all signals is estimated from reflectometry measurements with strong QC amplitude because the QC modes are more likely to be discriminated. Those signals $g_n(t)$ belong to the Tore Supra shot range of #32263–#32265. The ED threshold is calculated for each signal separately.

- CC threshold

The minimum CC is used as a threshold to enhance the SNR, and the latter can be deduced from (6):

$$\text{SNR} = \frac{\mu(|Wg_n(u,s)|)}{\sigma(|Wg_n(u,s)|)} \tag{6}$$

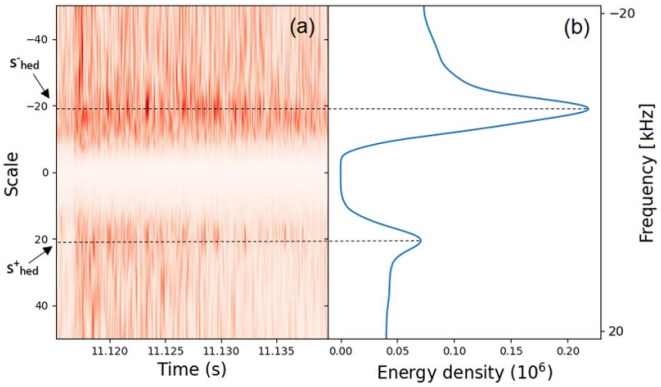

Figure 2. CWT modulus of (a) $f(t)$ and (b) the averaged ED as a function of scales.

where $Wg_n(u,s)$ is the CWT of the chosen signal $g_n(t)$; μ and σ correspond to the mean and standard deviation, respectively.

The threshold is then estimated:

$$\text{Threshold}_{\text{CC}} \approx \frac{\sum_{n=1}^{N} \text{SNR}_n}{N} \tag{7}$$

in which N corresponds to the number of signals $g_n(t)$. Then, a CC operation is performed between s_{hed} and other scales s (either positive or negative scales) and then normalized by their standard deviation, as shown by equation (8):

$$\text{CC}_{\text{normalized}} = \frac{|s_{\text{hed}}(t)|^2 \otimes |s(t)|^2}{\sigma_{\text{hed}} \cdot \sigma}. \tag{8}$$

Finally, hard thresholding (T_{h1}) is applied (figure 3):

$$T_{\text{h1}} = \begin{cases} Wf(u,s), & |\text{CC}_{\text{normalized}}| \geqslant \text{Threshold}_{\text{CC}} \\ 0, & |\text{CC}_{\text{normalized}}| < \text{Threshold}_{\text{CC}}. \end{cases} \tag{9}$$

- ED threshold:

This threshold is based on the minimum ED for existence of QC modes. In that regard, for a signal, we take every $|Wf_n(u,s)|^2$ from all u and s, and project it into a one-dimensional vector. In the beginning, we tried to estimate its probability density function by doing convolutions with predefined kernels, such as Gaussian and cosine. Nevertheless, it is not efficient because of the low precision in distinguishing components. This led us to a parametric approach, in which we could predefine the number of components or clusters.

According to Sun *et al*'s spectrum decomposition [14], once the two low-frequency components (direct current and low-frequency fluctuations) are filtered out, the following two components remain: broadband turbulence and noise. Since we consider QC modes as a third component, we choose three clusters to apply the K-means algorithm [26], an unsupervised machine learning algorithm that minimizes the within-cluster

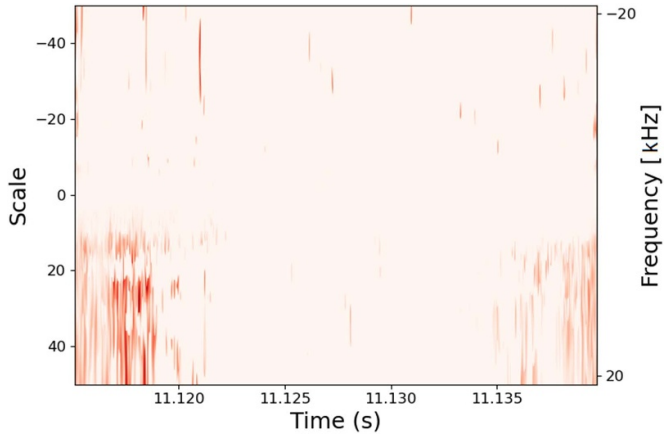

Figure 3. $|Wf(u,s)|$ after CC threshold (T_{h1}).

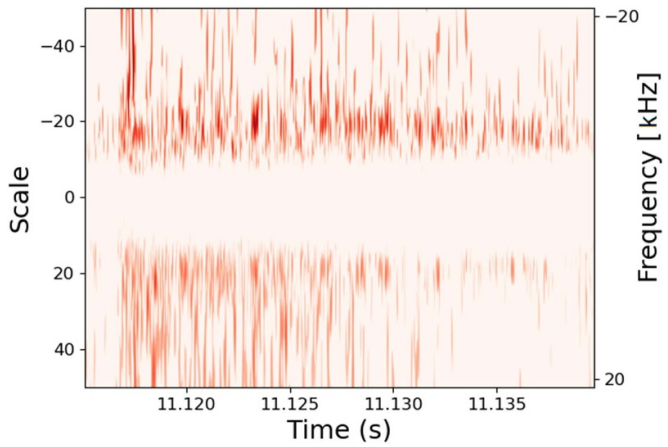

Figure 4. $|Wf(u,s)|$ after ED threshold (T_{h2}).

variances. Consequently, as there are three clusters, two limits are defined among them:

$$\begin{aligned} &\mathrm{ED}_{(\mathrm{cluster1})} \in [\min 1, \max 1\rangle; \\ &\mathrm{ED}_{(\mathrm{cluster2})} \in [\max 1, \max 2\rangle; \\ &\mathrm{ED}_{(\mathrm{cluster3})} \in [\max 2, \max 3]. \end{aligned} \tag{10}$$

Then,

$$\mathrm{Threshold}_{\mathrm{ED}} \approx \max 1. \tag{11}$$

Finally, hard thresholding (T_{h2}) is applied (figure 4):

$$T_{h2} = \begin{cases} Wf(u,s), & |Wf(u,s)|^2 \geqslant \mathrm{Threshold}_{\mathrm{ED}} \\ 0, & |Wf(u,s)|^2 < \mathrm{Threshold}_{\mathrm{ED}}. \end{cases} \tag{12}$$

The objective is to perform a mesh able to catch at least the QC modes and thereby eliminate most of the noise. To achieve this objective, the ED and CC 'patterns' resulting from the hard thresholding undergo a union operation to combine the properties of the ED and CC thresholds:

$$Wf(u,s)_{\mathrm{screened}} = T_{h1} \cup T_{h2}. \tag{13}$$

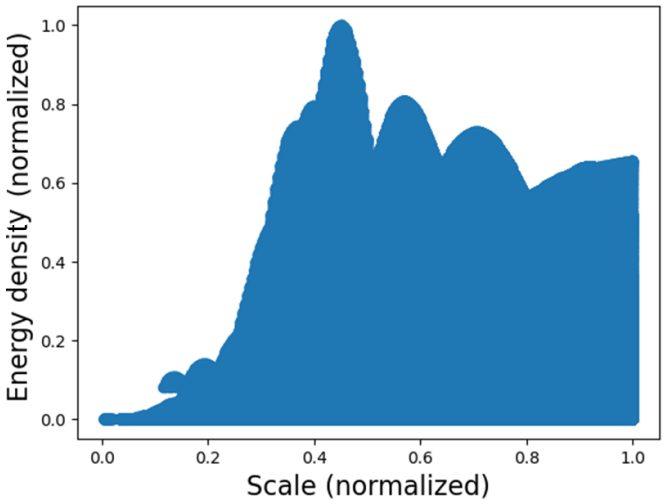

Figure 5. Projection of $Wf(u,s)_{\mathrm{screened}}$ into ED-scale vector space.

2.3.2. Clustering technique. A second process is needed to extract the QC modes from the screened CWT representation.

- ED-scale vector space

The choice of this vector space relies on combining the properties of the ED and CWT scales (named s) identified by the screening technique. This vector space follows the characteristics of a Euclidean space $\mathbb{R}^2$. First, all $|Wf(u,s)|^2$ and s belonging to $Wf(u,s)_{\mathrm{screened}}$ are normalized as the following:

$$\begin{aligned} s_{\mathrm{normalized}} &= \frac{s}{s_{\max}} \\ |Wf(u,s)|^2_{\mathrm{normalized}} &= \frac{|Wf(u,s)|^2}{|Wf(u,s)|^2_{\max}}. \end{aligned} \tag{14}$$

Then, the vector space is set as the following:

$$\boldsymbol{V} = (s_{\mathrm{normalized}}, |Wf(u,s)|^2_{\mathrm{normalized}}). \tag{15}$$

It is represented as (figure 5):

- Clustering

This part seeks to identify the QC mode cluster in the vector space $\boldsymbol{V}$. It is based on the mini-batch K-means; compared to the K-means algorithm, the main idea is to use small random batches of examples of a fixed size, so that they can be stored in memory. In each iteration, a new random sample from the data set is obtained and used to update the clusters, and this is repeated until convergence [27].

As discussed in section 2.3.1, the number of components or clusters existing is equal to 3. Some other parameters have to be set also for the mini-batch K-means algorithm: k-means++ method [28] for initialization of cluster centers, batch size equal to 1024 and the random state for random number generator equal to 0. This algorithm is thus applied into the ED-scale vector space $\boldsymbol{V}$ (figure 6).

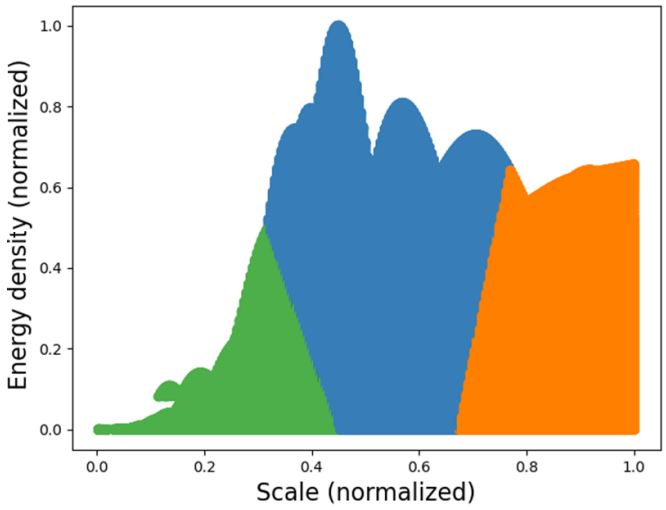

Figure 6. Clustering technique on V by mini-batch K-means, noise cluster (green), QC cluster (blue) and BB cluster (orange).

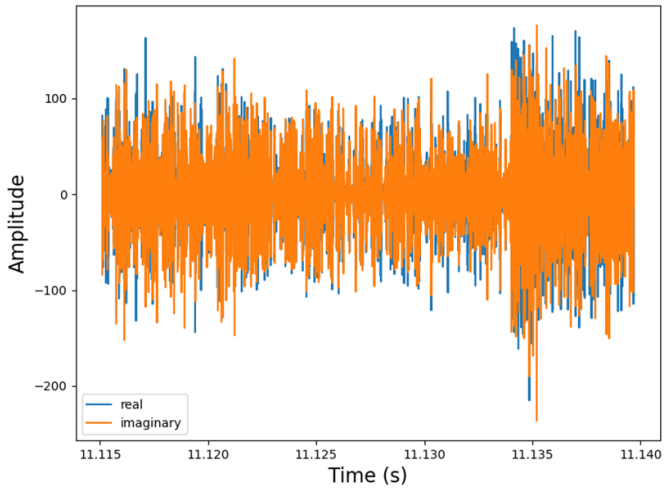

Figure 7. Time representation of the QC modes.

The QC mode cluster is identified, thanks to the s_{hed} chosen previously. This scale has to be present in the QC cluster.

As mentioned previously, this procedure is performed separately for negative and positive scales. Once the QC mode clusters are identified for both of them, they are back projected into the CWT space for reconstruction, and ICWT is applied to obtain the QC mode cluster in time representation (figure 7).

Thus, we can obtain the frequency spectrum of the initial signal, the QC modes and the initial signal without QC modes (figure 8).

3. Application of the algorithm

This part describes the application of our method to the experimental data from Tore Supra and TEXTOR reflectometers.

In Tore Supra, the reflectometry system worked in the frequency range of 105–155 GHz in X-mode [29]. The acquisition was performed at 1 MHz sample frequency. Typically, a series of 20–30 frequency steps were performed several times

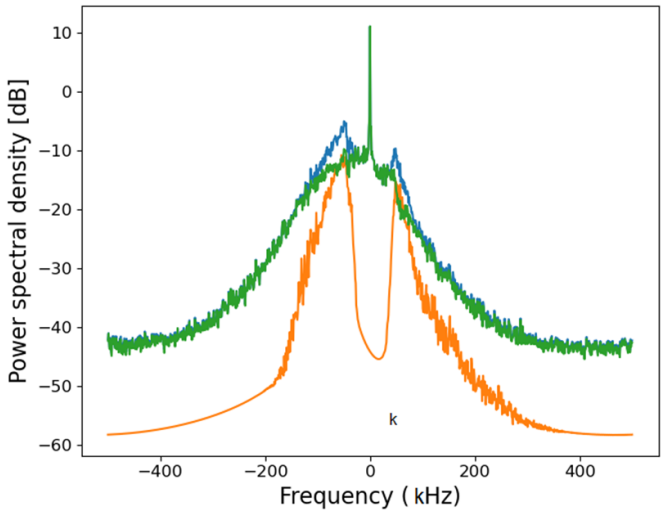

Figure 8. Frequency spectrum of initial signal (blue curve), QC modes (orange curve) and signal after extraction of QC modes (green curve).

per shot. The frequency step acquisition time windows were short about 10–20 ms.

In TEXTOR, the reflectometer worked in O-mode in the frequency range of 26–40 GHz. An antenna system built with one emitter and two receiver antennas was installed symmetrically to the emitter in the same cross section, enabling the performance of short-range poloidal correlation. The sampling frequency was 500 kHz, and we collected the data for a time interval of 2 s within a TEXTOR discharge [6].

3.1. LOC–SOC transition

The LOC–SOC transition was first identified by the saturation of the energy confinement time with increasing plasma density; see Rice *et al*'s review [11]. The transition is associated with the change of the dominant turbulent instability. Linear stability analysis performed with the GYRO code shows that the TEM instability is dominant for long-wavelength turbulence in the LOC regim and that the ITG modes are dominant in the SOC regime [30]. Recent results from gyro-kinetic codes have reached similar conclusions [31]. Different kinds of signatures are also associated with this behavior change: flip of the turbulent phase velocity [32, 33], change of the plasma toroidal velocity direction [34], rollover in density peaking occurs [35] and change in the electron fluctuation frequency spectra [30].

An important fact associated with the LOC–SOC transition is the disappearance of the QC modes in the reflectometry frequency spectra [9, 10]. Furthermore, the QC frequency bandwidth is intermediate between narrow coherent modes and the broadband turbulence component. It was recently shown that this bandwidth can evolve from a coherent peak to a broadband turbulence with increasing collisionality in electron cyclotron resonance heating plasmas [36].

In the LOC regime (figures 9(a)–(c)), QC modes can be distinguished from the BB component, and the ratio of the QC

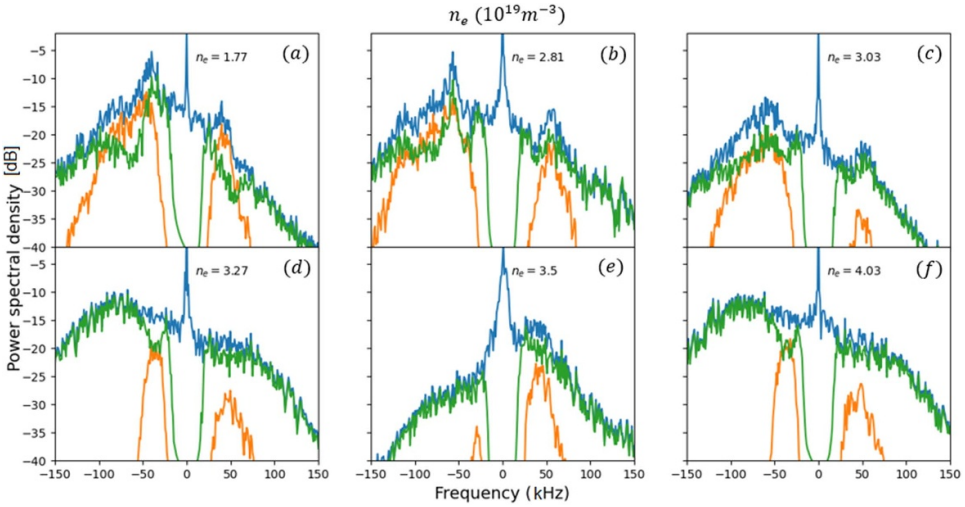

Figure 9. Tore Supra reflectometry spectra during LOC–SOC transition at $I_p = 1.2MA$ for $r/a = 0.3 \pm 0.025$ and $B_t = 3.8T$ from the shot range of #41003–#41013.

peak to the BB amplitude is meaningful (blue curve). As we mentioned previously, experimental and gyro-kinetic codes support this peak. Nevertheless, experimentally, it has been observed that the QC modes disappear in the SOC regime. In the SOC regime (figures 9(d)–(f)), there is a very low ratio between the QC peak and the BB amplitude (blue curve).

In previous studies, no quantification was performed. We apply our method to the same acquisitions as those shown in figure 1(c) from [7] to quantify the energy contribution of the QC mode to the frequency spectrum. Figure 9 shows the extracted QC modes (orange curve). In general, the main part of the QC mode contribution is extracted. However, the method could not be applied successfully to the last spectrum $n_e = 4.6 \times 10^{19}\,\mathrm{m}^{-3}$ because the spectrum is affected by Doppler effects [37]. For example, this can occur when the reflectometer does not probe the plasma perfectly perpendicular to the cutoff layer (i.e. a misalignment of the reflectometer compared to the equatorial plane or a vertical displacement of the plasma). As a consequence, the method detects a QC mode with the wrong amplitude (also in figures 9(d) and (f)). A harmonic or an HF-QC mode is present (figures 9(a) and (b)), which disturbs the extraction.

Finally, to obtain an estimation of the LOC–SOC transition (figure 10), a density step-by-step scan is recommended instead of analyzing a current ramp-up [11]. Therefore, the same shot range as in [7] is taken over an extended set of reflectometry measurements. Because the QC mode extraction process is quite fast (15 s for 10 000 time steps), it was applied to all the reflectometry measurements in the LOC–SOC shot range for which the cutoff layer was at $r/a = 0.3 \pm 0.025$. After excluding data with low SNR and those having a Doppler shift associated with the central frequency disappearance, 20 signals from a range of densities going from 1.8×10^{19} to $4.6 \times 10^{19}\,\mathrm{m}^{-3}$ are chosen.

For the first time, the QC mode energy is quantified. The QC mode energy normalized to the broadband energy shows a clear decrease in the QC mode contribution with increasing density (figure 10(b)). The transition is around the line average density $< n_l >= 3 \times 10^{19}\,\mathrm{m}^{-3}$. This threshold is similar to the density threshold for the confinement time observed in figure 10(a). The extraction of the QC modes in the LOC–SOC transition highlights four points around $< n_l >= 3 \times 10^{19}\,\mathrm{m}^{-3}$. These points exhibit large fluctuations of this normalized energy with respect to the electron density; this could be related to the plasma parameters reigning the LOC–SOC phenomenology. This would not depend only on parameters, such as the plasma current, the device major radius and the edge safety factor, but some other parameters, such as Z_{eff} and background ion species, might be considered as well [34]. The confinement saturation density remains ill-defined.

To verify that the signal properties are conserved after applying our method, we take a look at the coherence conservation between two antennas of PCRs developed at TEXTOR [13].

3.2. Short radial correlation

The short radial correlation (SRC) lengths for turbulence measured with reflectometry were analyzed in TEXTOR, and it revealed some spectrum properties linked to the QC modes by means of the MSC [6]. It has been observed that MSC allows to separate spectrum components and particularly shows a high correlation for frequencies from 50 to 150 kHz and -150 to -50 kHz corresponding to the QC modes. The MSC is defined as

$$\Gamma(f) = \frac{|G_{xy}(f)|^2}{G_{xx(f)}.G_{yy(f)}} \tag{16}$$

where $G_{xy}(f)$ denotes cross-spectral density and $G_{xx}(f)$, and $G_{yy}(f)$ auto-spectral density.

For the rest of the paper, the word *coherence* is used to denote the MSC. As coherence spectrum is well shaped for

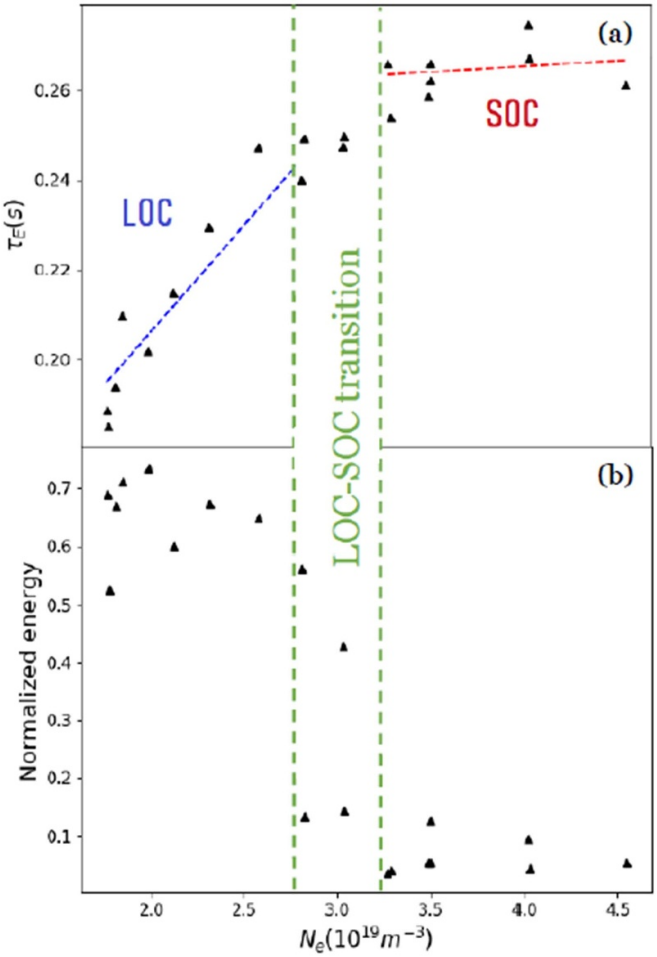

Figure 10. Tore Supra confinement time versus density during the LOC–SOC transition at $I_p = 1.2MA$ for $r/a = 0.3 \pm 0.025$ and $B_t = 3.8T$ from the (a) shot range of #41003–#41013 and the (b) QC mode energy normalized by the BB energy as a function of the density.

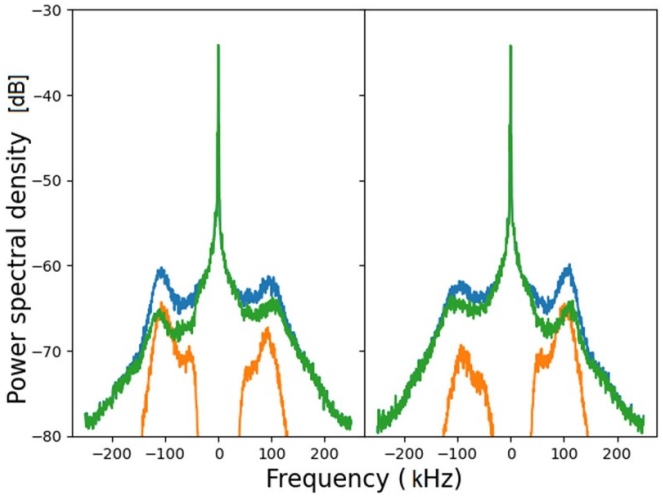

Figure 11. Frequency spectrum of initial signal from both antennas (blue curve), QC modes (orange curve) and signal after extraction of QC modes (green curve).

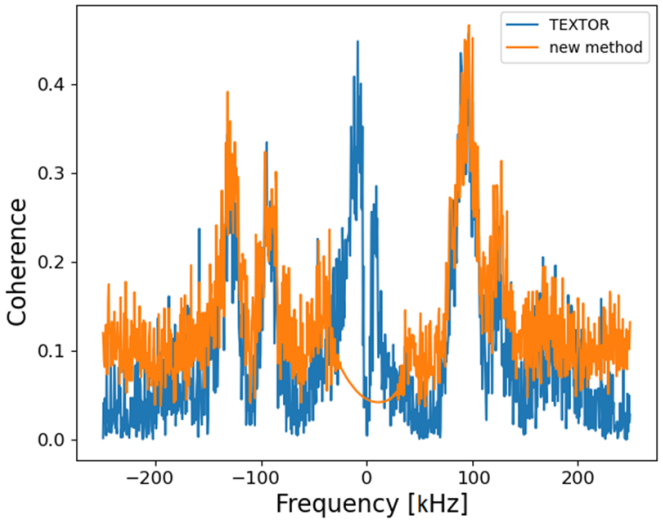

Figure 12. Coherence spectrum: TEXTOR method (blue curve) and proposed method (orange curve).

QC modes because of its SRC properties, a comparison can be made between the coherence from correlation reflectometry data at TEXTOR and the coherence using this work's method.

This method allows the extraction of QC modes coming from the signal received by one of the correlation reflectometry antennas. Figure 11 shows the initial and extracted spectra for both antennas. Due to the symmetry and the wave emission pattern, a mirror effect can be seen on the spectrum.

After the QC mode extraction for both antenna signals have been processed separately, we performed MSC following the procedure described as follows.

For the orange curve, the signals from the two reflectometry antennas (approx. 80 000 data points) are split into eight sets of 10 000 points, and then our method is applied to each of them. Thereafter, coherence is calculated pairwise between both antenna sets, and finally averaging of all coherence spectra is done. For the blue curve, equation (16) is performed between the whole signals of both antennas.

As shown in figure 12, both coherence spectra follow each other in shape and amplitude, showing a good agreement. This comparison confirms that our method preserves the signal properties in the extracted QC mode signal.

4. Conclusions and forthcoming studies

A first extraction of QC modes from reflectometry raw signals has been performed. The properties of QC modes and other components are preserved after application of our algorithm, as shown in section 3.1 for the physical perspective and in section 3.2 for the statistical one.

Based on this existing method, we can study the dynamics and propagation of QC modes using poloidal correlation reflectometers.

Moreover, an analysis of each extracted spectrum component can be carried out after the application of our method. The physical and statistical properties of each component can then be determined separately, and their interplay can be studied as well.

This method opens up the doors to automatize the extraction of modes for statistical analysis (big data) and to study the coupling with other spectrum components [2]. The method can first be used to detect whether the QC mode contribution is important or negligible, and consequently a machine learning algorithm can be built to discriminate spectrum with and without QC modes. It is also necessary to develop a self-determination of thresholds using AI methods. In this regard, the screening technique is expected to improve with the concept of threshold learning, in which the system would learn from the whole database and thereby offer a better discrimination of QC modes for each signal. Likewise, the clustering technique is expected to integrate another parameter in the vector space representation: the metric concept. It would lead to a separation of coupling modes.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

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). The views and opinions expressed are, however, those of the authors 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

Luigui Salazar https://orcid.org/0000-0002-1830-5991
Stéphane Heuraux https://orcid.org/0000-0001-7035-4574
Roland Sabot https://orcid.org/0000-0002-7419-9871
Andreas Krämer-Flecken https://orcid.org/0000-0003-4146-5085

References

[1] Truc A *et al* 1986 *Nucl. Fusion* **26** 1303
[2] Garbet X 2006 *C. R. Physique* **7** 573–83
[3] Diamond P H, Itoh S I and Itoh K 2010 *Modern Plasma Physics: Volume 1, Physical Kinetics of Turbulent Plasmas* (Cambridge: Cambridge University Press)
[4] Heuraux S and Clairet F 2009 *Instrum. Mes. Métrologie* **9** 59–86
[5] Vershkov V *et al* 2005 *Nucl. Fusion* **45** S203
[6] Krämer-Flecken A *et al* 2004 *Nucl. Fusion* **44** 1143
[7] Arnichand H *et al* 2014 *Nucl. Fusion* **54** 123017
[8] Arnichand H, Kraemer-Flecken A, Hacquin S and Sabot R 2013 Quasi coherent modes in textor and tore supra tokamaks *Proc. 11th Int. Reflectometry Workshop-IRW11 (Palaiseau)*
[9] Arnichand H *et al* 2015 *Plasma Phys. Control. Fusion* **58** 014037
[10] Zhong W *et al* 2016 *Phys. Plasmas* **23** 060702
[11] Rice J, Citrin J, Cao N, Diamond P, Greenwald M and Grierson B 2020 *Nucl. Fusion* **60** 105001
[12] Arnichand H 2015 Identication of trapped electron modes in frequency fluctuation spectra of fusion plasmas *PhD Thesis* Aix Marseille Université
[13] Krämer-Flecken A *et al* 2015 *New J. Phys.* **17** 073007
[14] Sun Y, Sabot R, Hornung G, Heuraux S, Hacquin S and Verdoolaege G 2018 *Rev. Sci. Instrum.* **89** 073504
[15] Sun Y, Sabot R, Heuraux S, Garbet X, Hacquin S, Hornung G and Verdoolaege G 2019 *Phys. Plasmas* **26** 032307
[16] Mallat S 1999 *A Wavelet Tour of Signal Processing* (Amsterdam: Elsevier)
[17] Feizollah A, Anuar N B, Salleh R and Amalina F 2014 Comparative study of k-means and mini batch k-means clustering algorithms in android malware detection using network traffic analysis *2014 Int. Symp. on Biometrics and Security Technologies (ISBAST)* (IEEE) pp 193–7
[18] Abdi H and Williams L J 2010 *Wiley Interdiscip. Rev.-Comput. Stat.* **2** 433–59
[19] An J and Cho S 2015 *Spec. Lecture IE* **2** 1–18 (available at: www.semanticscholar.org/paper/Variational-Autoencoder-based-Anomaly-Detection-An-Cho/061146b1d7938d7a8dae70e3531a00fceb3c78e8)
[20] Schubert E, Sander J, Ester M, Kriegel H P and Xu X 2017 *ACM Trans. Database Syst.* **42** 1–21
[21] Vasiliev V V and Fedorov L V 2018 *J. Mod.Phys.* **9** 2482–94
[22] Kaiser G 2011 Continuous wavelet transforms *A Friendly Guide to Wavelets* (Berlin: Springer) pp 60–77
[23] Daubechies I 1990 *IEEE Trans. Inf. Theory* **36** 961–1005
[24] Farge M 1992 *Annu. Rev. Fluid Mech.* **24** 395–458
[25] Torrence C and Compo G P 1998 *Bull. Am. Meteorol. Soc.* **79** 61–78
[26] Li Y and Wu H 2012 *Phys. Proc.* **25** 1104–9
[27] Sculley D 2010 Web-scale k-means clustering *Proc. 19th Int. Conf. on World Wide Web* pp 1177–8
[28] Arthur D and Vassilvitskii S 2006 k-means++: the advantages of careful seeding *Technical Report* (Stanford)
[29] Sabot R, Sirinelli A, Chareau J M and Giaccalone J C 2006 *Nucl. Fusion* **46** S685
[30] Sung C *et al* 2013 *Nucl. Fusion* **53** 083010
[31] Grierson B *et al* 2019 *Phys. Plasmas* **26** 042304
[32] Angioni C, Peeters A, Ryter F, Jenko F, Conway G, Dannert T, Fahrbach H, Reich M, Suttrop W, Team A U 2005 *Phys. Plasmas* **12** 040701
[33] Conway G *et al* 2006 *Nucl. Fusion* **46** S799
[34] Rice J *et al* 2013 *Nucl. Fusion* **53** 033004
[35] Erofeev I *et al* 2017 *Nucl. Fusion* **57** 126067
[36] Lee W *et al* 2020 *Nucl. Fusion* **61** 016008
[37] Rhodes T, Peebles W, Nguyen X, VanZeeland M, Degrassie J, Doyle E, Wang G and Zeng L 2006 *Rev. Sci. Instrum.* **77** 10E922