Nonlinear time series analysis and clustering for jet axis identification in vertical turbulent heated jets

A. K. Charakopoulos,[1] T. E. Karakasidis,[1,*] P. N. Papanicolaou,[2] and A. Liakopoulos[1]

[1]*Department of Civil Engineering, Laboratory of Hydromechanics and Environmental Engineering, University of Thessaly, 38334 Volos, Greece*

[2]*School of Civil Engineering, Department of Water Resources and Environmental Engineering, National Technical University of Athens, Heroon Polytechniou St., 15780 Zografos, Greece*

(Received 10 June 2013; published 14 March 2014)

In the present work we approach the hydrodynamic problem of discriminating the state of the turbulent fluid region as a function of the distance from the axis of a turbulent jet axis. More specifically, we analyzed temperature fluctuations in vertical turbulent heated jets where temperature time series were recorded along a horizontal line through the jet axis. We employed data from different sets of experiments with various initial conditions out of circular and elliptical shaped nozzles in order to identify time series taken at the jet axis, and discriminate them from those taken near the boundary with ambient fluid using nonconventional hydrodynamics methods. For each temperature time series measured at a different distance from jet axis, we estimated mainly nonlinear measures such as mutual information combined with descriptive statistics measures, as well as some linear and nonlinear dynamic detectors such as Hurst exponent, detrended fluctuation analysis, and Hjorth parameters. The results obtained in all cases have shown that the proposed methodology allows us to distinguish the flow regime around the jet axis and identify the time series corresponding to the jet axis in agreement with the conventional statistical hydrodynamic method. Furthermore, in order to reject the null hypothesis that the time series originate from a stochastic process, we applied the surrogate data method.

DOI: 10.1103/PhysRevE.89.032913 PACS number(s): 05.45.Tp, 47.27.−i, 47.52.+j

I. INTRODUCTION

Turbulent jet flows constitute a class of complex phenomena in physics and are encountered in a variety of engineering applications in atmospheric, oceanic, and other cases of environmental fluid dynamics. Due to noteworthy implications the study of turbulent jets has been the subject of investigation by several researchers in the past. The basic flow configuration for a turbulent jet flow is shown in Fig. 1, where in the case of fully developed turbulence one can observe a three region behavior. The first region corresponds to large distances from the jet axis, actually at the boundary with ambient water named boundary region (BR), the second one, the inner region (IR), concerns the region between the boundary region and the core of the jet, and the third region, the jet axis region (JR), is the region near the core of the jet. The dynamics of these regions is characterized by the presence of small and large scale structures (vortices). Large scale structures occupy the full width of the jet and appear with a certain low frequency that is a function of the distance from the nozzle due to continuous pairing. They contain a major part of jet energy. Small scale eddies constitute the large, energy carrying structures, appear at higher frequencies, and receive energy from the immediately bigger ones mainly at the central part of the flow.

Determining the location of the axis in a turbulent jet is a classical hydrodynamics problem in the literature with the majority of the previous approaches requiring hydrodynamics knowledge and methods [1,2]. These studies were devoted to vertical and horizontal turbulent jet flows by investigating the changes over time in velocity, density, temperature, or tracer concentration in the jet flow field, based on experimental research and numerical simulation methods [3–6]. More recently, a few works attempt to study the turbulent flow from the complex network perspective. Particularly, Gao and Jin [7] have employed the complex network theory to investigate the dynamic characteristics of gas-liquid two-phase flow patterns and characterize and distinguish patterns from inclined water-oil flow. Also, Gao *et al.* [8] proposed a method to distinguish patterns from inclined water-oil flow experiments based on the concept of network motifs. Liu *et al.* [9] studied the statistical properties of complex networks constructed from time series of energy dissipation rates in three-dimensional fully developed turbulence and Small *et al.* [10] have employed nonlinear methods to characterize chaotic dynamics from simulations of large strain behavior of a granular material under biaxial compression.

During the preceding decades a different way to obtain insight in the properties of a dynamical system (in our case the jet) was to analyze the time series of measurable quantities, which depend on the underlying and usually unknown dynamics of the physical system, in order to identify its characteristics. Time series analysis may include the use of simple descriptive statistics, linear, and nonlinear methods. Linear analysis incorporates simple measures, such as the autocorrelation function and power spectrum, while nonlinear methods based on the reconstruction of phase spaces include the mutual information and correlation dimension. A concise review of these methods can be found in Refs. [11–13]. Moreover, calculation of some nonlinear dynamic detectors, such as Hurst exponent, detrended fluctuation analysis (DFA), and Hjorth parameters, has shown significant power in the analysis of time series [14–18].

The aim of the present work is to propose a method of identifying and discriminating different regions of the jet through the analysis of time series measured at various positions within

*thkarak@uth.gr

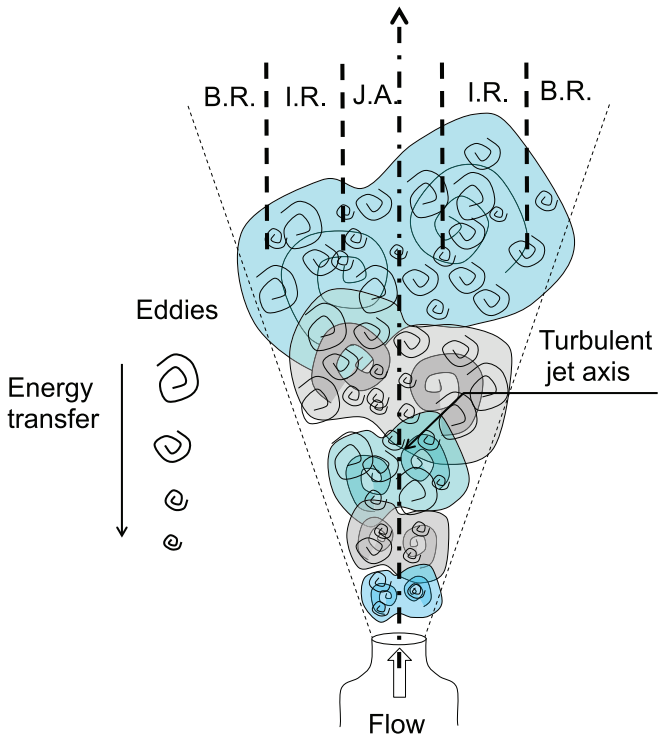

FIG. 1. (Color online) Schematic diagram of turbulent jet flow.

the jet, using mainly nonlinear time series analysis techniques and nonconventional hydrodynamics methods based mainly on statistics. More specifically, we analyzed the temperature fluctuations from vertical turbulent heated jets where temperature time series were recorded along a horizontal line through the jet axis. We employed data from five different sets of experiments with different initial conditions. In two cases the nozzle shape was circular and in the remaining three elliptical with 1:2 axis ratio. We concentrated our study in the investigation of the question if one can distinguish the time series corresponding to regions near the jet axis where conditions of fully developed turbulence are expected, from time series corresponding to regions that are more distant and from those near the boundary with the ambient water where intermittent flow of jet fluid occurs. In addition, the null hypothesis was tested for dynamic characteristics of the temperature time series performing a surrogate data test [19–22].

From the results of the proposed methodology we cannot only locate the jet axis region, but also differentiate from the time series analysis the dynamical behavior of various other regions.

The structure of the paper is as follows. In Sec. II we discuss briefly the theoretical background and the experimental setup for temperature measurements. In Sec. III we present the methodology employed for data analysis along with the linear and nonlinear measures. The results and discussion are presented in Sec. IV. Finally, the conclusions are presented in Sec. V.

II. THEORY AND EXPERIMENTAL SETUP

A. Theory

A heated round jet of diameter D and density ρ_o flows out of a nozzle with uniform velocity U parallel to jet axis in a calm ambient fluid of density ρ_α. The initial jet parameters are the volume and the specific (per unit mass) momentum and buoyancy fluxes are defined as

$$\begin{aligned} Q &= \frac{\pi D^2}{4} U, \\ M &= QU, \\ B &= \left(\frac{\rho_\alpha - \rho_o}{\rho_\alpha} g \right) Q = g'_o Q, \end{aligned} \tag{1}$$

respectively, where g is the gravitational acceleration and g'_o the effective gravity that will subsequently produce vertical momentum flux. Two characteristic length scales are defined as [23]

$$l_Q = \frac{Q}{M^{1/2}} \quad \text{and} \quad l_M = \frac{M^{3/4}}{B^{1/2}}. \tag{2}$$

l_M is a momentum length scale formed from the momentum and buoyancy fluxes and l_Q is the source length scale formed from the volume and momentum fluxes. Their ratio l_Q/l_M is the initial buoyant jet Richardson number R_o

$$\begin{aligned} R_o &= \frac{l_Q}{l_M} = \left(\frac{p}{4} \right)^{1/4} \frac{\sqrt{g'_o D}}{U} = \left(\frac{p}{4} \right)^{1/4} \frac{1}{F_o} \quad \text{and} \\ F_o &= \frac{U}{\sqrt{g'_o D}}, \end{aligned} \tag{3}$$

where F_o is the initial densimetric Froude number.

The temperature difference between the jet and ambient fluid produces the density deficiency that is responsible for the initial jet specific buoyancy flux. The mean dilution S at a point of the jet flow field is defined to be the ratio

$$S = \frac{T_o - T_a}{T - T_a}, \tag{4}$$

where T_o is the initial jet temperature, T_α the ambient temperature, and T the local time-averaged temperature. Jirka [24] has defined the jet axis to be the point of minimum dilution S_c

$$S_c = \frac{T_o - T_a}{T_c - T_a}, \tag{5}$$

where T_c is the maximum time-averaged (center line) temperature. We define x as the horizontal distance from jet axis and z the vertical distances from the nozzle.

B. Description of the experiments

Experiments have been carried out in a transparent orthogonal tank with dimensions 1.00 m × 0.80 m and 0.70 m deep, equipped with a peripheral overflow to remove excess water at the Applied Hydraulics Laboratory of the National Technical University of Athens. A perspective view of the experimental setup is shown in Fig. 2.

The tank was equipped with a peripheral overflow to remove excess water. The hot water jet supply consists of a water heater made of stainless steel, which is well insulated and pressurized by air, to provide adequate constant head pressure to drive to jet. During the water heating, a recirculating pump was used to ensure that the hot water is well mixed and there are

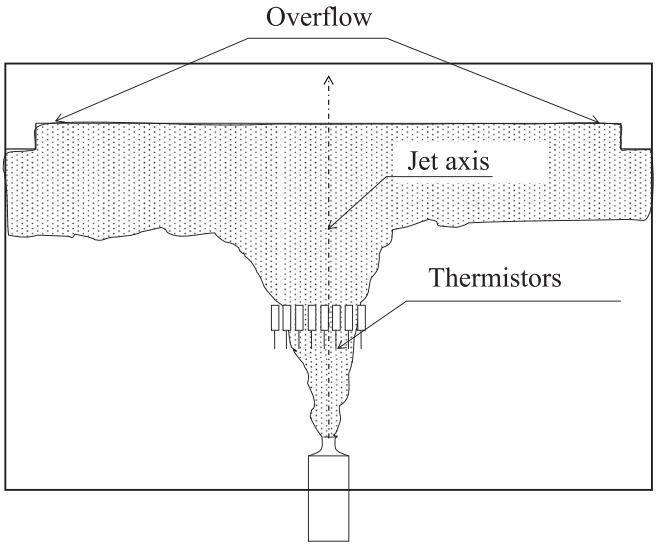

FIG. 2. Perspective view of the experimental setup.

no temperature gradients. An insulated pipe drives the water from the heater into the jet plenum, through a calibrated flow meter. The jet plenum was a cylinder of 4.5 cm i.d. equipped with 5 mm honeycomb and sponge to destroy the large scale turbulence structures. At the downstream end of it different types of nozzles could be mounted. Two different nozzles have been used—a circular of 1.5 cm diameter and an elliptical nozzle with 2:1 axis ratio 2 cm × 1 cm. They have been shaped in a way to provide smooth transition from the plenum diameter to the nozzle exit to avoid flow separation. The jet water temperature ranged between 58.60 °C and 61.44 °C, while the ambient water temperature ranged between 18.40 °C and 24.6 °C. Temperature measurements were obtained by an array of fast response thermistors positioned at constant elevation from the nozzle, on the plane of symmetry of the buoyant jet. The jet flow field was determined using a slide projector and a semitransparent paper sheet (shadowgraph) in order to place the rake of thermistors properly. In this paper, we use the data recorded at an elevation of 20 cm above the nozzle axis. The basic initial conditions of the flow are tabulated in Table I. The first, second, and fifth case studies correspond to buoyancy driven, plumelike type of flow ($z/l_M \approx 10$), where the momentum that has been produced by buoyancy is much greater than the initial jet momentum. The other two case studies (third and fourth) correspond to the buoyant jet ($1 < z/l_M < 10$), where initial jet momentum and buoyancy flux are in balance and both important for the jet movement.

III. TIME SERIES ANALYSIS

A. Methodology

To distinguish the time series that correspond to regions near the jet axis from those corresponding to regions near the boundary we used mainly nonlinear measures. More specifically, we focused on a set of measures such as the mutual information and the cumulative mutual information. We have also estimated simple descriptive statistics measures, calculated some linear and nonlinear dynamic detectors such as the Hurst exponent, detrended fluctuation analysis (DFA), and Hjorth parameters, and we have evaluated their discriminating power using clustering approach. In the proposed method, the separation of boundary regime from the inner region is obtained without the use of fluid mechanics properties. As indicated earlier, five experiments of turbulent jets with different initial conditions and nozzle shapes were used to illustrate the method. Below we give a brief description of the various measures employed in the present time-series analysis.

B. Linear and nonlinear measures

1. Descriptive statistics

We calculated some simple descriptive statistics and we try to summarize and describe our data set. More specifically, we estimated the mean, median, variance, standard deviation, interquartile range, skewness and kurtosis of each time series.

2. Mutual information and cumulative mutual information

Mutual information *I(t)* is the most popular nonlinear measure used in time series analysis in order to select the appropriate delay time τ for state space reconstruction [25] and is defined as

$$I(\tau) = \sum_{x(t_i),x(t_i+\tau)} P(x(t_i),x(t_i+\tau))\log\left[\frac{P(x(t_i),x(t_i+\tau))}{P(x(t_i)x(t_i+\tau))}\right], \tag{6}$$

where $x(t_i)$ is the ith data point of time series $t = kDt(K = 1,2,\ldots,k_{\max})$; $P(x(t_i))$ is the probability density at $x(t_i)$ and $P(x(t_i),x(t_i+\tau))$ is the joint probability density at $x(t_i)$, $x(t_i+\tau)$; τ is the delay time. The delay τ of the first minimum is chosen as a delay time for the reconstruction of phase space.

We also used a nonlinear measure, the cumulative mutual information $M(\tau_{\max})$, defined as the sum of mutual information $I(\tau)$ [26] for a time delay τ as

$$M(\tau_{\max}) = \sum_{\tau=1}^{\tau_{\max}} I(\tau). \tag{7}$$

TABLE I. Experimental conditions. (*) f = temperature data sampling frequency.

Case study	Shape of nozzle	D (cm)	U (cm/s)	T_o (°C)	T_a (°C)	Re	l_m (cm)	z/l_m	R_o	f (Hz)*
1st	Round	1.5	6.74	58.60	24.60	2040	2.167	9.231	0.614	80
2nd	Round	1.5	6.74	58.60	24.60	2040	2.167	9.231	0.614	80
3rd	Elliptical	2:1	18.25	60.60	18.40	4888	5.236	3.820	0.239	80
4th	Elliptical	2:1	11.85	61.44	19.10	3203	3.365	5.944	0.372	100
5th	Elliptical	2:1	7.58	61.40	19.20	2048	2.156	9.278	0.581	100

3. Estimation of Hurst exponent

The Hurst exponent (H) has been used first as a measure of long-range correlation in the analysis of time series, and second to classify them. Estimation of the Hurst exponent was initially introduced by Hurst [14] while investigating the discharge time series of the Nile River. The oldest method to calculate the Hurst exponent called rescaled range or R/S analysis was proposed by Mandelbrot and Wallis [15] and estimates H are based on the R/S statistics.

Suppose we have a time series of N measurements. We divide the time series into N_s shorter subseries of length $n = N, N/2, N/4$. For each subseries we define the range R_n

$$R_n = \max_{1\leqslant k\leqslant s}\left[\sum_{i}^{k}(x_{ns+i} - \overline{x_n})\right] - \min_{1\leqslant k\leqslant s}\left[\sum_{i}^{k}(x_{ns+i} - \overline{x_n})\right], \tag{8}$$

where

$$n = 0,1,\ldots,N_s - 1, \quad N_s = N/S,$$

and

$$\overline{x_n} = \frac{1}{s}\sum_{i=1}^{s} x_{ns+i}. \tag{9}$$

The sample standard deviation is defined as

$$S_n = \sqrt{\frac{1}{s}\sum_{i=1}^{s}(x_{ns+i} - \overline{x_n})^2}. \tag{10}$$

Then the rescaled range is R_n/S_n.

The Hurst exponent is estimated by calculating the average rescaled range for all subseries of length n. It can be shown that the R/S statistics follows the relation

$$(R/S)_n = c \cdot n^H \quad \text{as} \quad n \to \infty, \tag{11}$$

where c is a constant. In plotting $(R/S)_n$ statistics against n on a log-log scale, we expect to get a line the slope of which determines the Hurst exponent. With this definition the Hurst exponent of 0.5 indicates a random series, when a value of H lies within the interval $0.5 < H < 1$ indicates persistent behavior, while a Hurst exponent value between 0 and 0.5 exists for time series with antipersistent behavior.

4. Detrended fluctuation analysis

Detrended fluctuation analysis (DFA) method was proposed by Peng *et al.* [18] and it is useful for analyzing time series that originate from long-memory processes. The advantage of DFA over R/S analysis is that it permits the detection of long-range correlations embedded in nonstationary time series.

Suppose we have a time series length N. We divide the time series into N_s shorter subseries of length n. For each time subseries $m = 1,\ldots,N_s$ first we create a cumulative time series

$$Y_{i,m} = \sum_{j=1}^{i} X_{j,m}, \quad i = 1,2,\ldots,n. \tag{12}$$

Next we fit the series with a least squares line,

$$\tilde{Y}_m(x) = a_m x + b_m, \tag{13}$$

and then we calculate the root mean square fluctuation of the integrated and detrended time series,

$$F(m) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{i,m} - a_m i - b_m)^2}. \tag{14}$$

The root mean square fluctuation is calculated by

$$\bar{F}(n) = \frac{1}{N_s}\sum_{m=1}^{N_s} F(m). \tag{15}$$

As in the case of R/S analysis a linear relationship on a log-log plot indicates the presence of power-law scaling $F(n) \approx c \cdot n^{\alpha}$ where c is a constant [18]. The DFA function can be characterized by a scaling exponent α as the slope of a line on a plot log$F(n)$ to log(n). If the process is persistent the slope is greater than 0.5, while if it is antipersistent then the slope is less than 0.5. In the case of white noise time series the slope is roughly 0.5.

5. Hjorth parameters

The Hjorth descriptors, namely activity, mobility, and complexity, were mainly developed for quantification of an electroencephalogram (EEG) by Hjorth [16,17]. The Hjorth parameters, activity (H_A), mobility (H_M), and complexity (H_C) are called normalized slope descriptors because they can be defined as first and second derivatives and are respectively defined as follows:

$$H_A = m_0, \tag{16}$$

$$H_M = \sqrt{\frac{m_2}{m_0}}, \tag{17}$$

$$H_C = \sqrt{\frac{m_4/m_2}{m_2/m_0}}, \tag{18}$$

where m_0 is the variance (square of the standard deviation) of the variable, m_2 is the variance of the first derivative of the variable, and m_4 is the variance of the second derivative of the variable.

In fact, activity represents the width of the signal. Mobility is defined as the squared root of the ratio between the variances of the first derivative and the amplitude, and represents the mean frequency of the time series. Complexity is the ratio between the mobility of the first derivative and the mobility of the nonlinear time series. Complexity indicates the deviation of the slope and can be seen as a measure of change in frequency of the signal. Although these measures have been used in the EEG analysis and constitute useful clinical tools for the quantitative description of an EEG [27–29] relevant studies where these parameters have been employed in the analysis of time series from hydrodynamics do not appear in the literature.

C. Clustering analysis

Apart from the performance of the above measures as individual indicators, the process of cluster analysis has been used for classification of the time series. Clustering is a division of data into groups of similar objects. Each group consists of

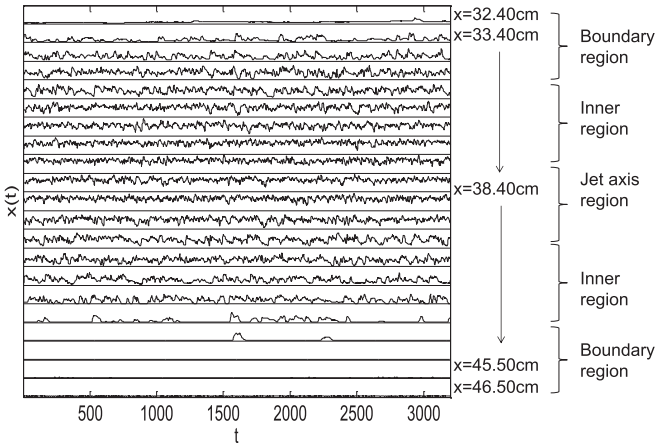

FIG. 3. Time series at various positions.

objects that are similar between themselves and dissimilar to objects of other groups. There have been several studies that employed different measures of similarity for clustering such as methods based on mutual information or correlation dimension [30–32]. In our study we have used a combination of all measures described in the previous sections as a measure of similarity. More specifically, we have employed all the simple statistical measures described above, the mutual and cumulative mutual information, Hurst exponent, detrended fluctuation analysis (DFA), and the Hjorth parameters.

We used the single linkage hierarchical clustering algorithm [33] in order to classify our data. One of the main advantages of hierarchical clustering is that a dendrogram can be drawn to find the appropriate number of clusters in a data set. The height at which two clusters are merged in the dendrogram reflects the distance of the two clusters.

Briefly using the single linkage hierarchical clustering algorithm we first calculate the distance from each object (point) to all other points, using Euclidean distance measure, and place the numbers in a distance matrix. Then we identify the two clusters with the shortest distance in the matrix, and

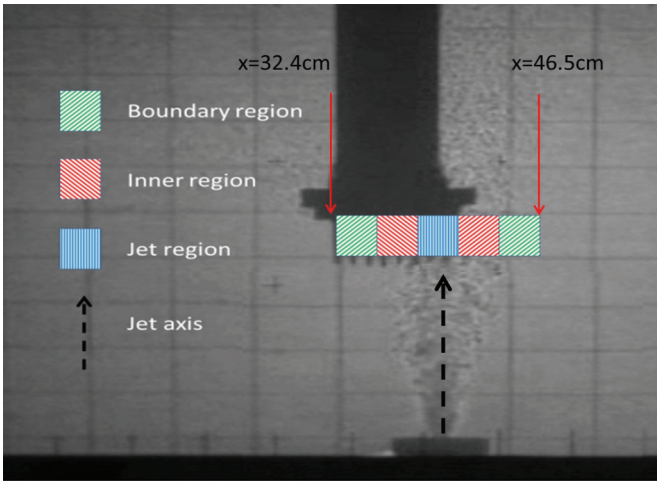

FIG. 4. (Color online) Shadowgraph view of the experimental setup. The green areas refer to the boundary region, while the blue and red areas refer to the jet axis region and inner regions, respectively. The red arrows indicate the limits of the measurement zone.

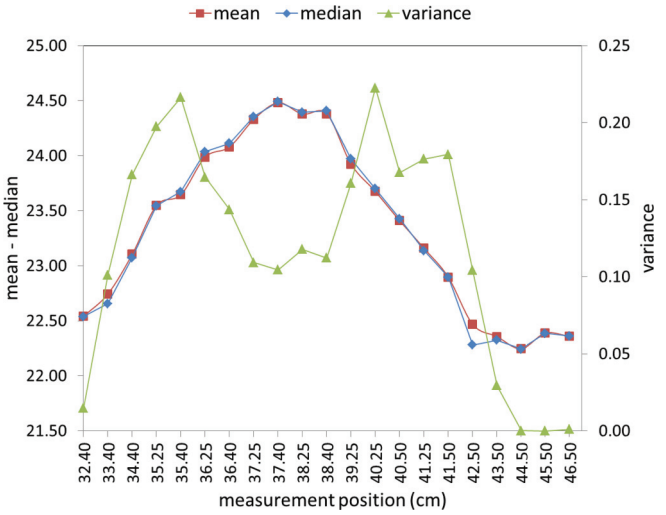

FIG. 5. (Color online) Mean, median (left vertical axis), and variance (right vertical axis) of the time series as a function of measurement positions for the various cases of jets.

merge them together. Subsequently, the distance matrix is recomputed, as those two clusters are now in a single cluster (and no longer exist by themselves). Finally, we compute the distances from each object (point) to all other points until we group all the data sets into clusters. The most dependent data are grouped together.

IV. RESULTS AND DISCUSSIONS

A. Case 1

We analyzed 21 recordings of temperature time series, one for each radial location of measurement, where the sampling time at each location was 40 s at a frequency of 80 Hz. The first time series recorded in a position $x = 32.40$ cm at horizontal axis and the last position in $x = 46.50$ cm. In Fig. 3 the time series is displayed. The horizontal axis refers to time (t) and the vertical axis refers to each time series located as we move from the left boundary ($x = 32.4$ cm) of the tank to the right ($x = 46.5$ cm).

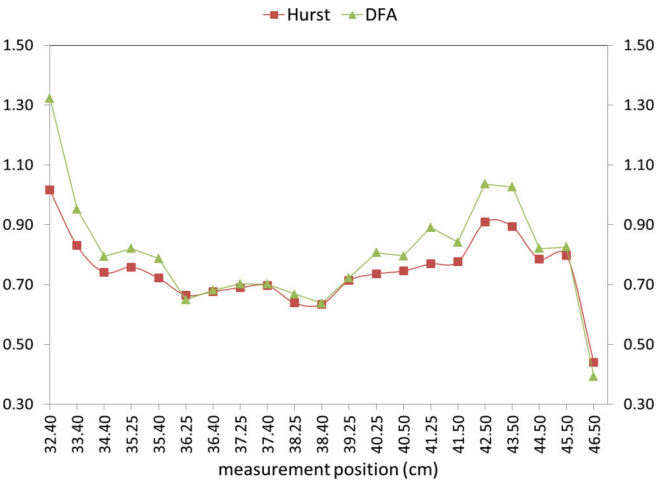

FIG. 6. (Color online) Hurst exponent of the time series as a function of measurement positions with R/S method and exponent with the detrended fluctuation analysis (DFA).

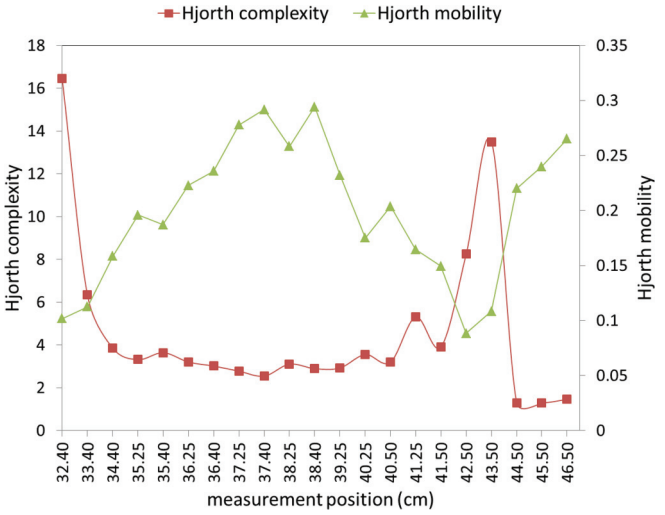

FIG. 7. (Color online) Hjorth parameters mobility and complexity of the time series as a function of measurement positions.

As we have mentioned before for each time series we first calculated the simple statistical measures such as the mean, median, variance, standard deviation, kurtosis, skewness, and interquartile range. Then we evaluated the nonlinear measures (mutual information and the cumulative mutual information). Finally, we estimated the linear and nonlinear dynamic detectors, i.e., the Hurst exponent, detrended fluctuation analysis (DFA), and Hjorth parameters. Subsequently, we constructed the normalized matrix of all the above measures, and evaluated the discrimination power of the measures using clustering analysis. The algorithm described in Sec. III C was used for clustering.

A representative shadowgraph view of the time series is presented in Fig. 4. Moreover, in this figure we present approximately the jet axis region and the rest regions and also the location of the centerline of the jet using different colors.

1. Estimation of the descriptive statistics measures

The statistical analysis of the time series is focused on evaluation of simple statistical measures such as the mean, median, variance, standard deviation, kurtosis, skewness, and interquartile range for each time series. The objective of this part of our work is to study the performance of simple statistics estimators and whether it is possible to discriminate one time series from another. The results of mean, median, and variance are shown in Fig. 5 as a function of the horizontal distance from the nozzle.

As we can clearly observe the mean and median values of the excess (above ambient) temperature profile is quite symmetrical. There is also a region consisting of the time series at $x = 37.25$, 37.40, 38.25, and 38.40 cm, where the maximum values correspond to the regime around the jet axis. Additionally, from this figure it is evident that the lowest values of the variance occurred in the same region.

It is of interest that the behavior of the above measures, which are based on the measurement of the time series at various locations, permits one to discriminate in a clear way the spatial behavior of turbulence across the line of measurement.

2. Evaluation of Hurst exponent

So far the parameters we have investigated are capable of extracting information from the time series. However, for a clear discrimination between the time series we needed extra information. The obtained results for the Hurst exponent for each time series with the R/S method and with detrended fluctuation analysis (DFA) are displayed in Fig. 6.

The Hurst exponents take the lowest values in the region 36.25–38.40 cm which corresponds to the jet axis region allowing one to thus distinguish these time series (and the corresponding measurement regions) from the others. The values of the Hurst exponent are in general greater than 0.5 (except for the location $x = 46.5$ which lies outside the jet region and in fact corresponds to the ambient water region) indicating a persistent behavior. We have to mention that the R/S method and DFA method show almost the same behavior. It is of interest to discuss the physical interpretation of this behavior. In fact, close to jet boundaries the presence of long-life large vortices creates a strongly persistent behavior in time, while as we move towards the jet axis region the

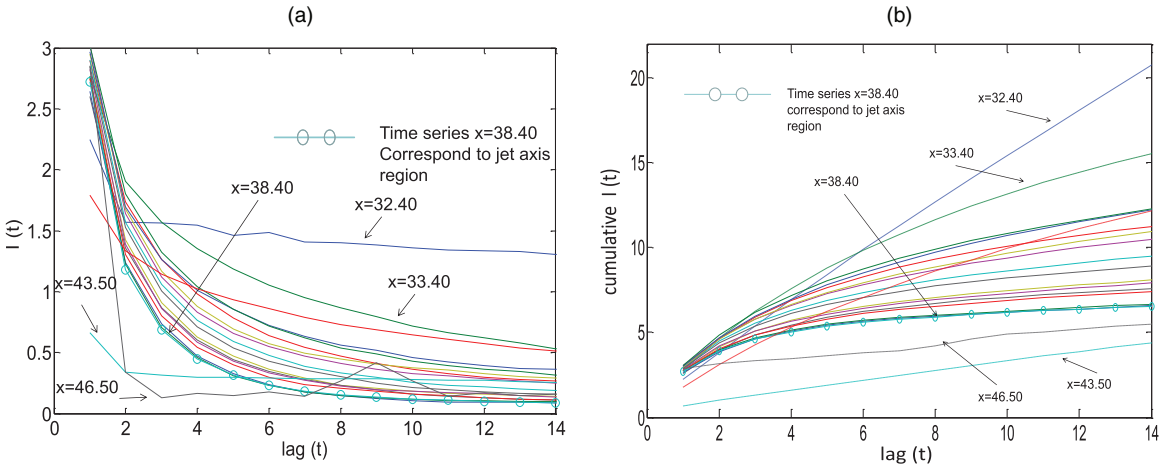

FIG. 8. (Color online) (a),(b) Mutual information and cumulative mutual information of the time series as a function of measurement positions.

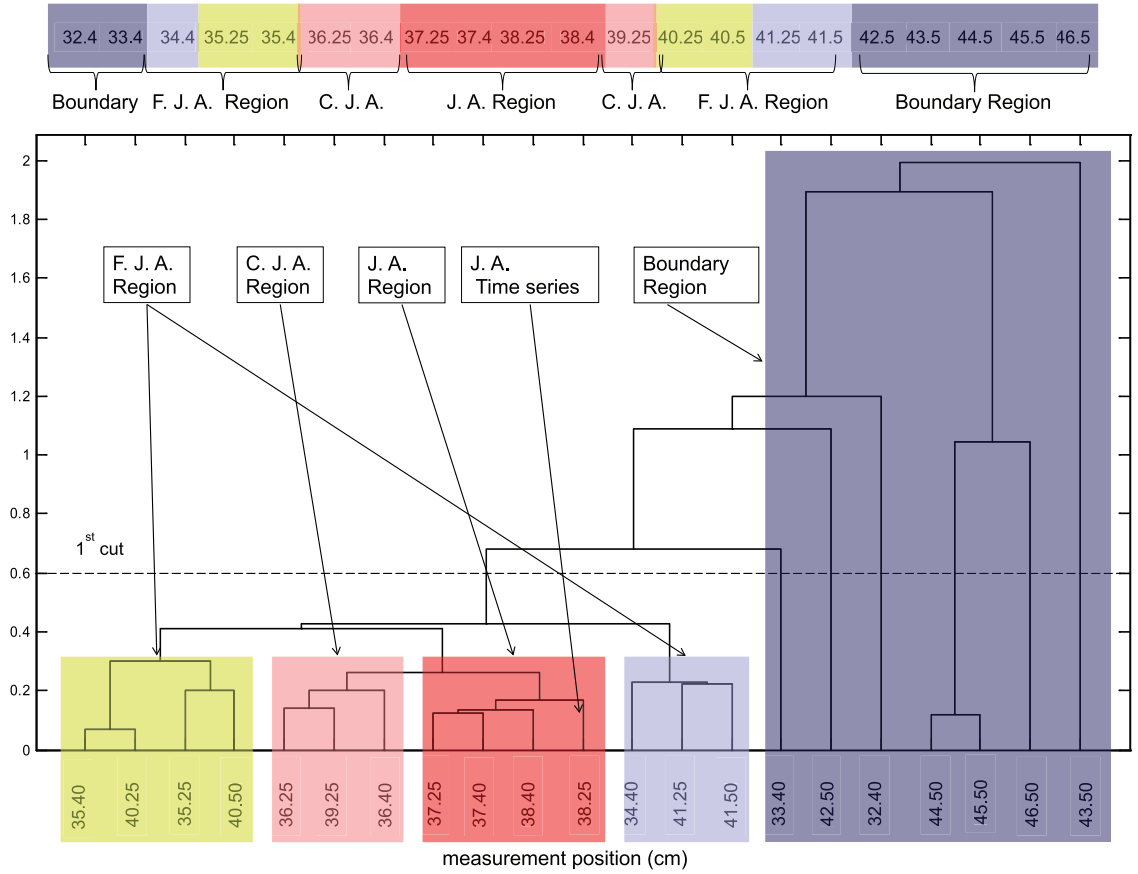

FIG. 9. (Color online) Dendrogram based on the measures obtained from the time series. The sketch above is a schematic representation of the location of the various regions in the measurement setup (FJA: far from jet axis region; CJA: close to jet axis region; JA: jet axis region).

presence of short-lived and small size vortices reduces the persistent behavior.

3. Hjorth parameters

The results for Hjorth parameters of the time series mobility and complexity are presented in Fig. 7. It is interesting to note that the mobility of the groups with the highest values appears in the time series from the core of the jet, where the lowest values of the complexity occurred.

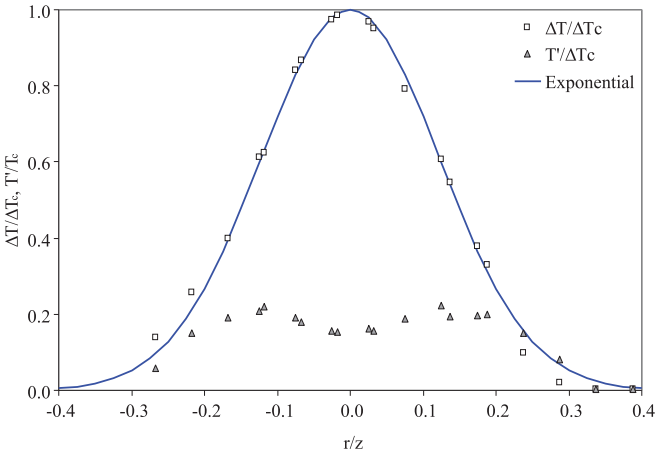

FIG. 10. (Color online) Normalized excess (above ambient) and rms temperature by the maximum center line time-averaged temperature is plotted versus r/z, along with exponential fit to the time-averaged normalized temperature.

4. Mutual and cumulative mutual information

Further, we have calculated the mutual information function and the cumulative mutual information for each time series and the results are presented in Figs. 8(a) and 8(b). In Fig. 8(a) one can observe that there are time series presenting the smallest local minimum but not the lowest values of average mutual information which correspond in fact to regions in or close to the ambient water, while time series close to the jet axis (near $x = 37.40$ cm) present the lowest values of average mutual information (the two far distant measurement positions $x = 46.50$ and 43.50 cm are exempt). As we get far from the jet axis, but always in the turbulent jet region, average mutual information increases and time lags of the minimum are shifted toward larger times. Such behavior is consistent with what is expected since near the jet center line turbulence is fully developed and there appear many short-lived small scale turbulent structures, while near the jet boundary the large scale flow structures live longer. At the jet axis region the memory of the flow structures is lost fast, while at locations close to the boundaries memory lasts longer.

A close look in Fig. 8(b), where the cumulative mutual information for each time series is presented, indicates that some time series are closer to some other forming groups corresponding to different measurement areas.

TABLE II. Surrogate.

Discriminating statistics	Time series from jet axis	Time series located in the inner region
First minimum of mutual information	$D = 10.04 > 1.96$	$D = 20.76 > 1.96$
Hjorth mobility	$D = 97.92 > 1.96$	$D = 107.77 > 1.96$
Hjorth complexity	$D = 154.99 > 1.96$	$D = 363.57 > 1.96$
Hurst exponent (R/S)	$D = 4.53 > 1.96$	$D = 7.85 > 1.96$
Hurst exponent (DFA)	$D = 5.36 > 1.96$	$D = 8.75 > 1.96$

TABLE III. Parameters of case studies.

Case study	Shape of nozzle	Number of time series recorded	Location of the first and last time series along the horizontal axis
2nd	Round	21	31.7–46.2
3rd	Elliptical	19	33.5–43.5
4th	Elliptical	18	33.0–43.0
5th	Elliptical	14	33.0–42.5

5. *Clustering*

From the results already presented above it is evident that we can separate various regions in the jet from the analysis of the corresponding time series. However, since each measure captures different characteristics and with different sensitivity, we combined all the above measures in an effort to better discriminate the various regions of the jet as well as locate the jet axis. As a measure of similarity we used a combination of all the simple statistical measures described above, the mutual and cumulative mutual information, and the Hurst exponent and the Hjorth parameters.

The hierarchy built by the clustering algorithm from each time series is represented by the dendrograms given in Fig. 9. The horizontal axis refers to each time series and the vertical axis refers to the distance. On top of the dendrogram the location of each time series measurement along the horizontal axis is plotted in a schematic way.

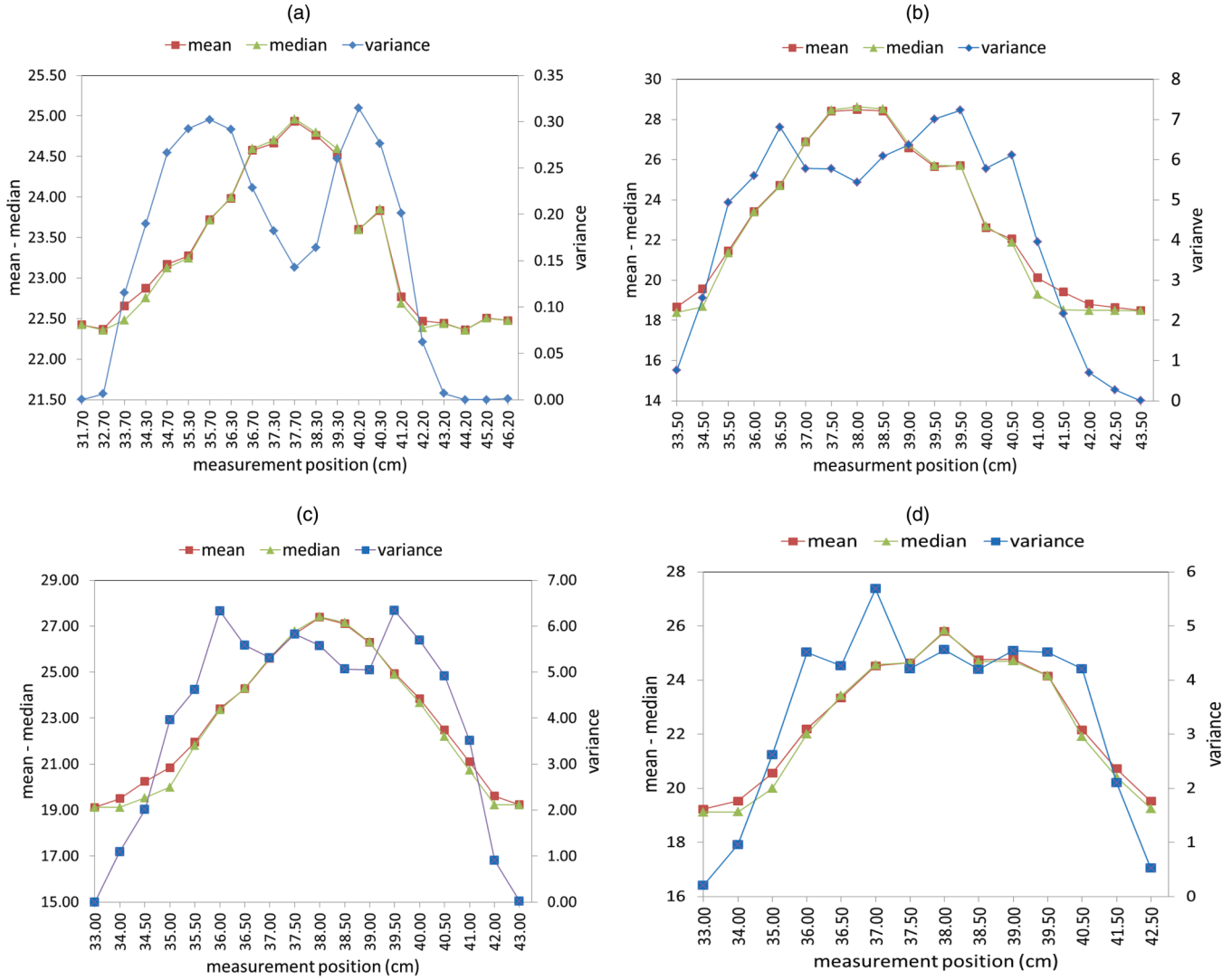

FIG. 11. (Color online) Mean, median (left vertical axis), and variance (right vertical axis) of the time series as a function of measurement positions for the various cases of jets (see Table III). (a) Second case, (b) third case, (c) fourth case, and (d) fifth case.

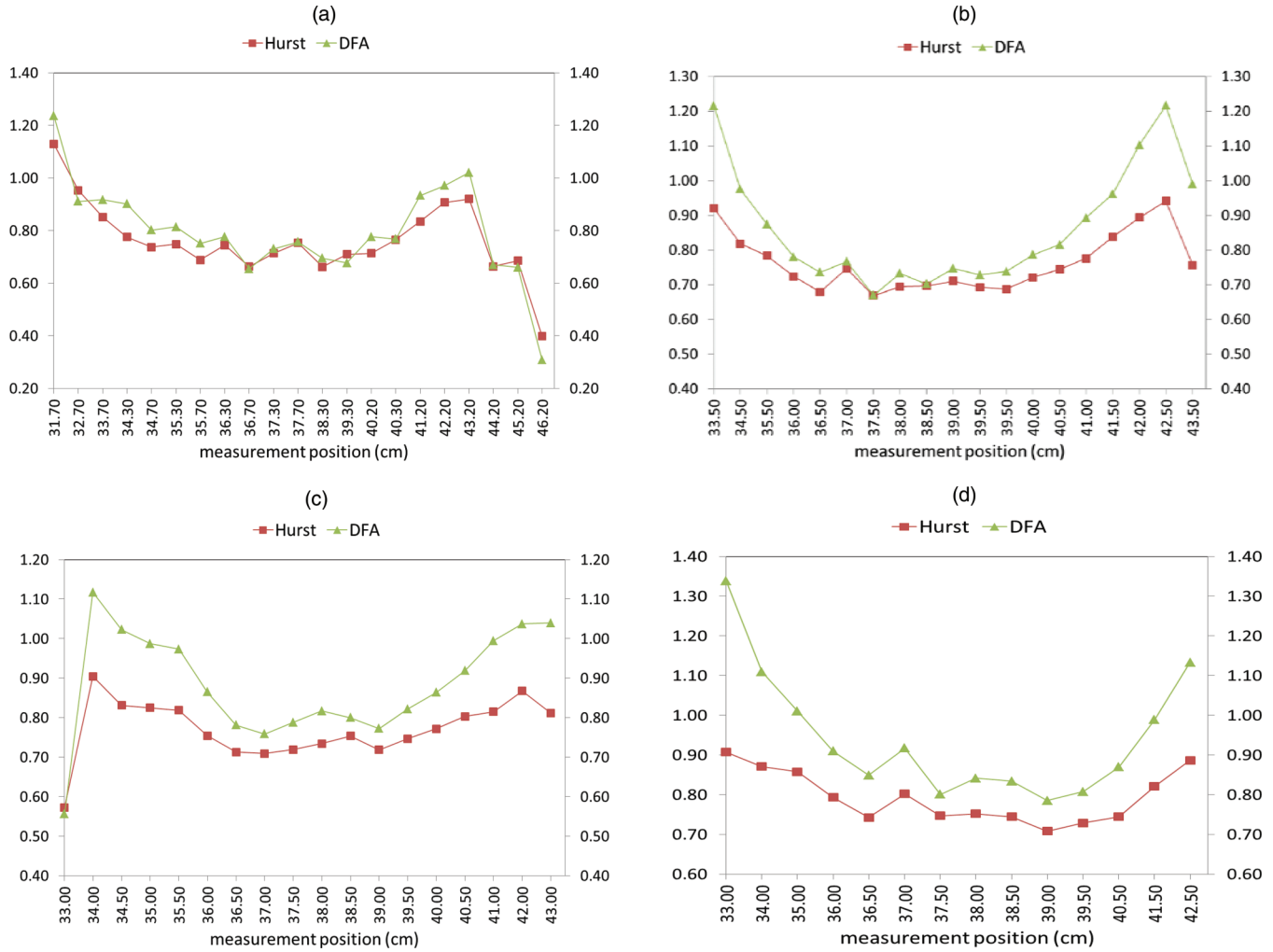

FIG. 12. (Color online) Hurst exponent of the time series as a function of measurement positions with R/S method and with the detrended fluctuation analysis (DFA) for the various cases of jets (see Table III). (a) Second case, (b) third case, (c) fourth case, and (d) fifth case.

According to the clustering method applied, separation between the jet inner and boundary region is based on the distance between the clusters formed based on the properties of each measurement position. In order to avoid any misunderstanding with the measurement distance from the nozzle we refer to this distance as cluster distance. We decided to make successive "*cuts*" at the dendrogram at different levels of cluster distance (vertical axis). The first "*cut*" is made at distance ∼0.6, where one can clearly see two main partitions. One main group consist of the time series at distances $x = 32.40$ cm, $x = 33.40$ cm, $x = 42.50$ cm, $x = 43.50$ cm, $x = 44.50$ cm, $x = 45.50$ cm, and $x = 46.50$. These time series correspond to long measurement distances from the axis and we denote the corresponding region as the boundary region (BR). The second main cluster includes the time series that are members of the same group at smaller distances from the jet axis and correspond to the inner region of the jet. This first step is important because we can exclude the time series corresponding to the edges of the measuring area.

In the inner region a more refined structure exists with several groups appearing for smaller cluster distance cuts. More specifically, we observed four groups of time series. The first consists of the time series at $x = 34.40$ cm, $x = 41.25$ cm, and $x = 41.50$ cm and the second of the time series at $x = 35.25$ cm, $x = 35.40$ cm, $x = 40.25$ cm, and $x = 40.50$ cm. The first and second groups include the time series from locations far from jet axis and we denote it as far from jet axis (FJA) region. The third group consists of the time series at $x = 36.25$ cm, $x = 36.40$ cm, and $x = 39.25$ cm. The third group includes the time series closer to the jet region than the first and the second cluster and denotes the corresponding regions as close to jet axis (CJA) region. Next we can distinguish one cluster which includes the time series at $x = 37.25$ cm, $x = 37.40$ cm, $x = 38.25$ cm, and $x = 38.40$ cm. This group of time series is at the smallest cluster distance (vertical axis of the graph). For this cluster we can notice that first the time series at $x = 37.25$ cm joined to $x = 37.40$ cm, then joined to $x = 38.40$ cm, and finally all these time series joined with the time series at $x = 38.25$ cm. This result suggests that the jet axis location is near to the time series at $x = 38.25$, which is the first subgroup as the cluster distance reduces. During the experiment the jet axis was located by optical measurements close to the midpoint $x = 37.70$ cm. This was also supported by the behavior of the average

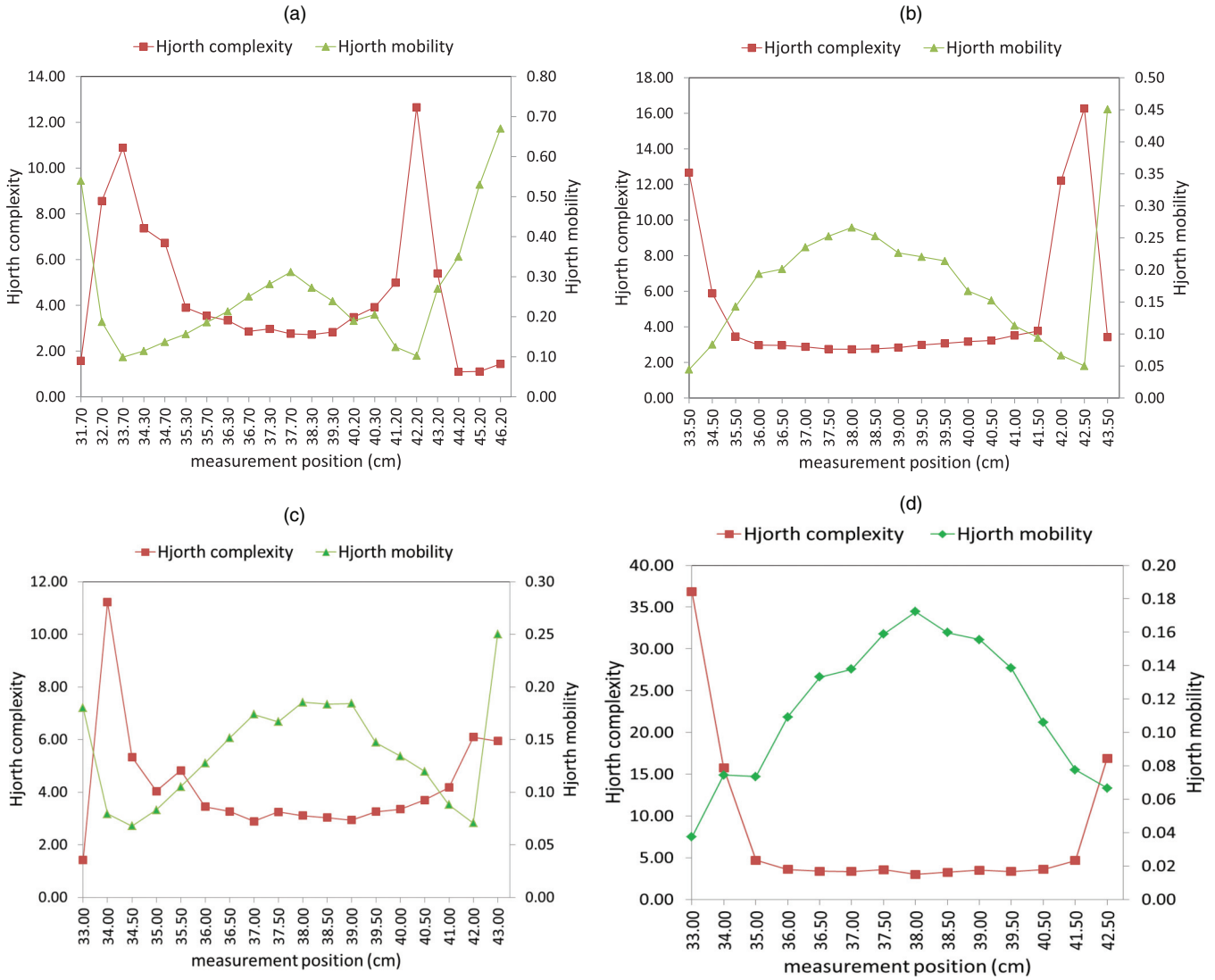

FIG. 13. (Color online) Hjorth parameters of the time series mobility and complexity as a function of measurement positions for the various cases of jets (see Table III). (a) Second case, (b) third case, (c) fourth case, and (d) fifth case.

temperatures observed in these time series. In Fig. 10 the normalized excess (above ambient) and rms temperature by the maximum center line time-averaged temperature is plotted versus r/z, along with an exponential fit to the time-averaged normalized temperature.

It is of interest to note here that the position of the axis from the hydrodynamics aspect can be considered where the maximum normalized temperature or the minimum intensity of turbulence between the two peak values occurred, i.e., at $x = 37.75$ cm. We observe that following our analysis the jet axis is located close to $x = 38.25$ cm, but we can notice that the analysis of the time series allowed us to discriminate with great detail the time series corresponding to different turbulence states along the horizontal line of measurement where the thermistors are located. Also we could detect the time series corresponding to locations near jet axis.

Summarizing the above discussion, from the dendrogram in Fig. 9 one can distinguish the flow region in the boundary region from that in inner region, based on cluster distance. We consider as BR the time series corresponding to longer distances. The inner region contains all the time series that are members of the same group for smaller distances (usually around 0.6). In this inner region a more refined structure exists with several groups appearing for smaller distance cuts. It is of interest that the jet axis region corresponds to a group whose members stay together up to very small distances.

6. Surrogate data method

In order to reject the hypothesis that the time series originate from a random process, we applied the surrogate data test. We generated two sets of 30 surrogate data, one from time series recorded in the inner region and another from time series measured at the jet axis, using one of the most common techniques, the Fourier transformed surrogates. This method generates time series with the same power spectrum as the original data set. We assume that the time series originates from a stochastic process (null hypothesis) and then we attempt to

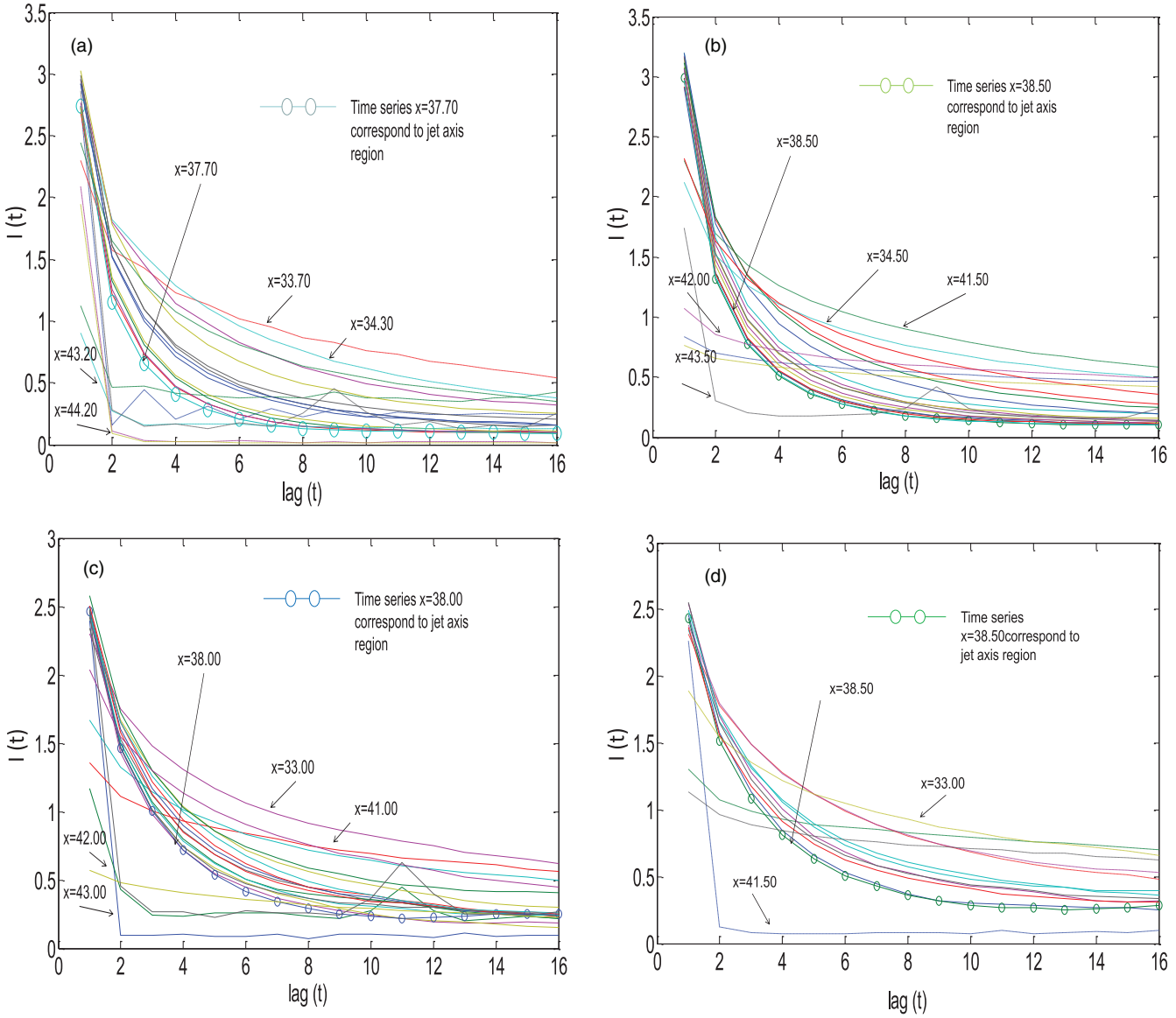

FIG. 14. (Color online) Mutual information of the time series as a function of measurement positions for the various cases of jets (see Table III). (a) Second case, (b) third case, (c) fourth case, and (d) fifth case.

reject this hypothesis. According to this method, if the estimate s_o of a quantity resulting from analysis of the original data does not lie within the empirical distribution of s under H_o formed by the estimates $s_1, s_2, \ldots, s_N$ on the N surrogates, then H_o is rejected. The quantity $D = \frac{|s_o - \bar{s}|}{d_s}$, where $\bar{s}$ and d_s are the average and standard deviation of $s_1, s_2, \ldots, s_N$, allows us to reject H_o at a confidence level higher than 95%, if $D > 1.96$.

In order to validate our estimation of rejection of the presence of a linear process, we used as discriminating statistics the first minimum of mutual information, the Hjorth mobility, the Hjorth complexity, and the Hurst exponent calculated by R/S analysis and DFA analysis. The resulting values are summarized in Table II. As we can see, the values obtained from the original time series (one from the inner region and one from the jet axis) are distinctly different from the ones resulting from the surrogates. This is confirmed from the value of the parameter $D > 1.96$, which allows us to reject the null hypothesis with a confidence level higher than 95%.

B. Other cases

In order to establish the methodology described previously in detail, we have also employed data from different sets of experiments with different initial conditions as shown in Table I. However, some major parameters for each case study can be found in Table III.

We briefly report the main part of the methodology as follows.

(i) Estimation of simple statistical measures.

(ii) Evaluation of mutual information and cumulative mutual information function.

(iii) Calculation of Hurst exponent, detrended fluctuation analysis (DFA), and Hjorth parameters.

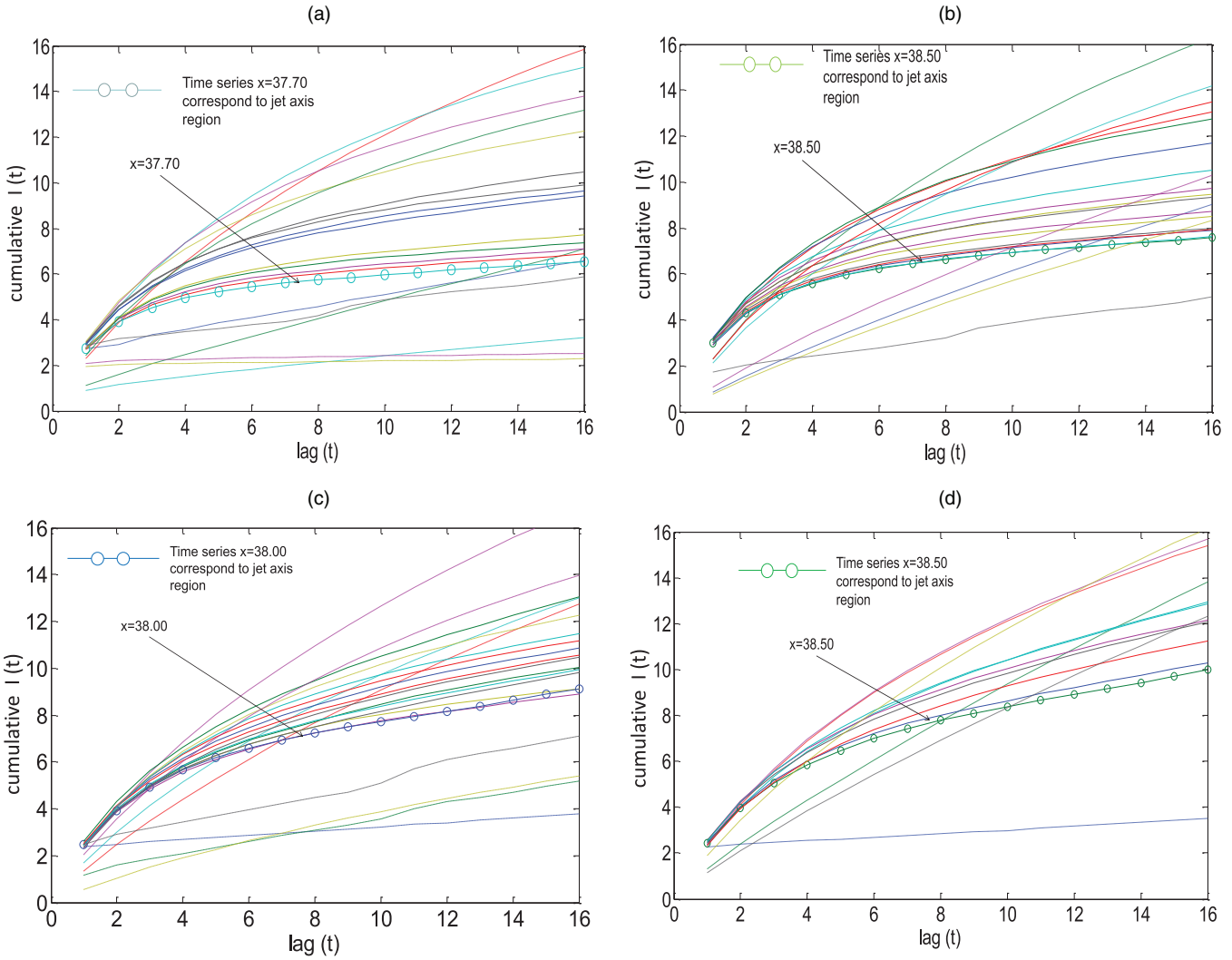

FIG. 15. (Color online) Cumulative mutual information of the time series as a function of measurement positions for the various cases of jets (see Table III). (a) Second case, (b) third case, (c) fourth case, and (d) fifth case.

(iv) Combination of above measures using the clustering techniques.

In Figs. 11(a)–11(d) the mean, median, and variance of the time series as a function of measurement positions for the various cases of jets, (a) second case, (b) third case, (c) fourth case, and (d) fifth case, are presented. The profiles of the measures in all cases are very similar. The mean and median increase while moving from the boundary to the center of the jet where they take a peak value, while moving to the other boundary of the jet they decrease. This is expected since near jet axis the maximum mean temperature occurs. Regarding the variance we have a central region where the fluctuations are smaller than that of the surrounding regions. This is constant with the physical state of the fluid, while close to the jet axis the variations of the temperature are smaller.

Figures 12(a)–12(d) show the Hurst exponent of the time series with R/S method and the detrended fluctuation analysis (DFA) of the second, third, fourth, and fifth case studies. One may observe that all these experimental sets have almost the same behavior.

The profile of the Hjorth parameters of the time series mobility and complexity for the second, third, fourth, and fifth case studies is plotted in Figs. 13(a)–13(d). Close to the boundary the complexity is greater from that in the inner region. On the contrary, mobility takes the highest value in the near the center region of the jet.

Figures 14 and 15(a)–15(d) display the mutual information function and the cumulative information function for each set of the second, third, fourth, and fifth case studies. As a common characteristic in all cases we can observe that the time series take the lowest values of average mutual information near jet axis (the two far distant measurement positions at $x > 40.00$ cm are exempt).

The results for the cumulative mutual information show the existence of groups of point measurements presenting similar behavior and which could form specific groups and subgroups.

The dendrograms obtained from a clustering procedure for the second, third, fourth, and fifth case studies are presented in Figs. 16(a)–16(d), respectively. Again, as we have already discussed in the first case study (Fig. 9) described in detail in Sec. IV A, it turns out that we can separate the flow region in

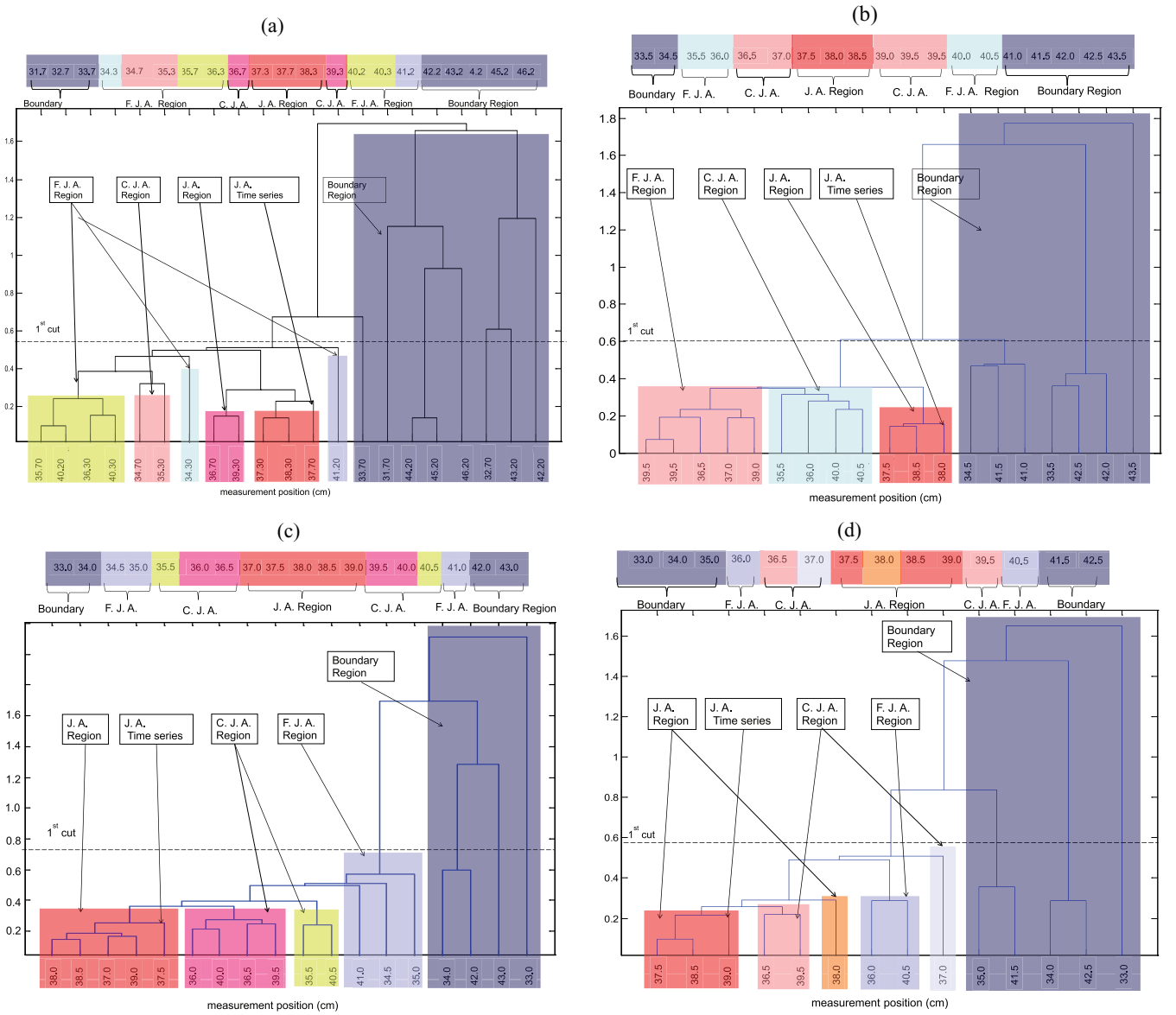

FIG. 16. (Color online) Dendrogram based on the measures obtained from the time series of (a) the second case, (b) third case, (c) fourth case, and (d) fifth case studies. The sketch above is a schematic representation of the location of the various regions in the measurement setup (FJA: far from jet axis region; CJA: close to jet axis region; JA: jet axis region).

a boundary and an inner region based on the cluster distance (vertical axes of dendrograms). We consider as BR the time series corresponding to greater distances. The inner region contains all the time series that are members of the same group for smaller distances (usually around 0.6). In this inner region a more refined structure exists with several groups appearing for smaller distance cuts. It is of interest noting that the jet axis region corresponds to a group whose members stay together up to very small distance "cuts." Also, it is worth mentioning that the jet axis is related to separations at small dendrogram distances in the jet axis region (JA), where one can distinguish the jet axis time series as the time series corresponding at the highest distance within this group of time series. Thus we could attribute the jet axis at the measurement position 37.7 cm in Fig. 15(a), which is close to the value 38.0 cm estimated through a Gaussian fit to the temperatures.

In Table IV we summarize the results of all cases; to identify the position of the axis jet along the horizontal line, we attributed the measurement station corresponding to a jet axis behavior. For comparison we present the estimated axis location obtained through a Gaussian fit on the measured average temperature profile. We can see in all cases that our estimation lays within the estimated values. Summarizing in all cases the clustering procedure results suggests that it is capable to discriminate the existence of several fluid regions which refer to different physical states. Comparing Figs. 8 and 15(a)–15(d) we observe that the results present a similar behavior. Furthermore, in every case we can distinguish the measurement station close to the jet axis in agreement with the conventional hydrodynamics methods.

The interest of the proposed methodology is that positioning of jet axis is mainly based on the analysis of the time series

TABLE IV. Results of jet axis location using the clustering procedure and comparison with estimations from hydromechanics methods.

Case study	Shape of nozzle	Measurement station attributed to the jet axis using the clustering procedure (present study)	Estimated location of jet axis using a Gaussian fit
1st	Round	38.25	37.75
2nd	Round	37.70	38.00
3rd	Elliptical	38.00	38.00
4th	Elliptical	37.75	38.00
5th	Elliptical	39.00	38.20

of a scalar measured, without the need to make a hypothesis regarding the statistical distribution of the scalar, such as a Gaussian fit to the temperature profile. The various indicators (measures) can also be employed separately in order to locate characteristic regions, but the clustering procedure seems to provide a broader approach. However, since each indicator may represent different aspects of the underlying system dynamics at each location, the combined approach through the clustering procedure seems to be more effective.

V. CONCLUSIONS

In the present study we have proposed a methodology in the study turbulent jet flows, to distinguish the region near jet axis from those near the boundary (ambient water) and the one in between, based on temperature time series analysis through nonlinear analysis and nonconventional statistical methods. Five data sets of temperature time series measured in experiments with different initial conditions were used to illustrate the method. In contrast to several previous studies the main advantage of the method presented is that it does not require prior specialist hydrodynamic knowledge.

For each time series, first we have performed conventional descriptive statistical measures and then we have calculated nonlinear measures such as mutual information and cumulative mutual information. Also, we have estimated some linear and nonlinear dynamic detectors such as the Hurst exponent, detrended fluctuation analysis (DFA), and Hjorth parameters. We have evaluated their discriminating power using clustering analysis. Furthermore, we applied the surrogate data method in the temperature time series. The results allowed us to reject the null hypothesis that the time series originate from a linear process, with a confidence level higher than 95%.

The methodology presented here provides results that are in agreement with those of applied hydrodynamics methods. It turns out that the various measures can provide information about various regions of the jet, as well as about the location of jet axis. The combined use of all the measures along with a clustering procedure can discriminate far better various regions of the flow based on a different behavior, and lead to a methodology for obtaining the location of jet axis. In fact the results suggest that the above analysis is capable of extracting information and can be useful for a more clear discrimination of the time series measured near the jet axis from those that correspond to regions near the boundaries with ambient fluid. More specifically, the time series near the jet axis have considerably different range of values from the time series near the boundary with ambient fluid. It seems that the jet axis is related to separations at small distances in the jet axis region, and generally it only corresponds to the first one member group that appears first in the JA region.

We conclude that, using a combination of simple descriptive statistics measure along with a linear and a nonlinear indicator, one could discriminate time series that correspond to the region near the jet axis with a high confidence level. This methodology seems quite promising for application in complex flows, as well as in applications where several different state zones exist in a physical system where one can have access to spatiotemporal data. Since the various indicators reflect an aspect of the underlying system dynamics at various regions, the combined use of them through the clustering procedure can provide the means to better discriminate these regions, and to identify specific localized regions such as the jet axis in the present study. Generally speaking, such methodologies could also be employed in the case of other spatiotemporal phenomena in order to localize different regions based on their dynamical behavior. A characteristic example is the rainfall structure, the dynamics of which is one of the most interesting environmental phenomena.

[1] P. Bradshaw, *Turbulence*, Topics in Applied Physics Vol. 12 (Springer, Berlin, 1978).

[2] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT Press, Cambridge, MA, 1975), Vol. 2.

[3] G. Broze and F. Hussain, J. Fluid Mech. **311**, 37 (1996).

[4] S. C. Crow and F. H. Champagne, J. Fluid Mech. **48**, 547 (1971).

[5] A. Mataoui and R. Schiestel, J. Fluids Struct. **25**, 60 (2009).

[6] X. Wang and Tan S. Keat, J. Hydrodynam. **22**, 1009 (2010).

[7] Z. Gao and N. Jin, Phys. Rev. E **79**, 066303 (2009).

[8] Z. K. Gao, N. D. Jin, W. X. Wang, and Y. C. Lai, Phys. Rev. E **82**, 016210 (2010).

[9] C. Liu, W.-X. Zhou, and W.-K. Yuan, Physica A **389**, 2675 (2010).

[10] M. Small, D. M. Walker, A. Tordesillas, and C. K. Tse, Chaos **23**, 013113 (2013).

[11] H. D. I. Abarabnel, *Analysis of Observed Chaotic Data* (Springer, Berlin, 1996).

[12] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 1995).

[13] J. C. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press, New York, 2003).

[14] H. E. Hurst, Trans. Am. Soc. Civ. Eng. **116**, 770 (1951).

[15] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 967 (1969).
[16] B. Hjorth, Electroencephalogr. Clin. Neurophysiol. **29**, 306 (1970).
[17] B. Hjorth, Electroencephalogr. Clin. Neurophysiol. **34**, 321 (1973).
[18] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).
[19] J. Theiler, S. Eubanka, A. Longtina, B. Galdrikiana, and D. Farmera, Physica D **58**, 77 (1992).
[20] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **79**, 1475 (1997).
[21] D. Kugiumtzis and C. Aifantis, J. Mech. Behav. Mater. **15**, 135 (2004).
[22] M. Small, K. Judd, and T. Stemler, arXiv:1306.4064.
[23] H. B. Fischer, E. J. List, R. C. Y. Koh, J. Imberger, and N. H. Brooks, *Mixing in Inland and Coastal Waters* (Academic Press, New York, 1979).
[24] G. H. Jirka and G. H. Env, Fluid Mech. **4**, 1 (2004).
[25] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[26] D. Kugiumtzis, I. Vlachos, A. Papana, and P. G. Larsson, Int. J. Bioelectromagn. **9**, 134 (2007).
[27] P. P. Balestrassi, A. P. Paiva, A. C. Zambroni de Souza, J. B. Turrioni, and E. Popova, J. Appl. Stat. **38**, 327 (2011).
[28] M. A. Navascues and M. V. Sebastian, Int. J. Comput. Math. **86**, 10 (2009).
[29] F. S. Bao, D. Y.-C. Lie, and Y. Zhang, in *Proceedings of the 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI-2008)*, Vol. 2 (IEEE, Piscataway, NJ, 2008), p. 482.
[30] A. Kraskov, H. Stogbauer, R. G. Andrzejak, and P. Grassberger, Europhys. Lett. **70**, 278 (2005).
[31] I. Priness, O. Maimon, and I. Ben-Gal, BMC Bioinformatics **8**, 111 (2007).
[32] G. Tourassi, E. Frederick, M. Markey, and C. Floyd, Med. Phys. **28**, 2394 (2001).
[33] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 4th ed. (Prentice Hall, Upper Saddle River, NJ, 1998).