Modelling and analysis of turbulent datasets using Auto Regressive Moving Average processes

Davide Faranda,[1,a)] Flavio Maria Emanuele Pons,[2] Bérengère Dubrulle,[1] François Daviaud,[1] Brice Saint-Michel,[3] Éric Herbert,[4] and Pierre-Philippe Cortet[5]

[1]*Laboratoire SPHYNX, Service de Physique de l'Etat Condensé, DSM, CEA Saclay, CNRS URA 2464, 91191 Gif-sur-Yvette, France*
[2]*Dipartimento di Scienze Statistiche, Universitá di Bologna, Via delle Belle Arti 41, 40126 Bologna, Italy*
[3]*Institut de Recherche sur les Phénomènes Hors Equilibre, Technopole de Chateau Gombert, 49 rue Frédéric Joliot Curie, B.P. 146, 13 384 Marseille, France*
[4]*Université Paris Diderot - LIED - UMR 8236, Laboratoire Interdisciplinaire des Énergies de Demain, Paris, France*
[5]*Laboratoire FAST, CNRS, Université Paris-Sud, France*

We introduce a novel way to extract information from turbulent datasets by applying an Auto Regressive Moving Average (ARMA) statistical analysis. Such analysis goes well beyond the analysis of the mean flow and of the fluctuations and links the behavior of the recorded time series to a discrete version of a stochastic differential equation which is able to describe the correlation structure in the dataset. We introduce a new index Υ that measures the difference between the resulting analysis and the Obukhov model of turbulence, the simplest stochastic model reproducing both Richardson law and the Kolmogorov spectrum. We test the method on datasets measured in a von Kármán swirling flow experiment. We found that the ARMA analysis is well correlated with spatial structures of the flow, and can discriminate between two different flows with comparable mean velocities, obtained by changing the forcing. Moreover, we show that the Υ is highest in regions where shear layer vortices are present, thereby establishing a link between deviations from the Kolmogorov model and coherent structures. These deviations are consistent with the ones observed by computing the Hurst exponents for the same time series. We show that some salient features of the analysis are preserved when considering global instead of local observables. Finally, we analyze flow configurations with multistability features where the ARMA technique is efficient in discriminating different stability branches of the system. © *2014 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4896637]

I. INTRODUCTION

For a long time, experimentally testable predictions of turbulence properties have been influenced by available measurements. For example, hotwire velocity measurements have motivated statistical analysis of turbulent spectra or velocity increments computation, allowing the evaluation of Kolomogorov direct or refined similarity hypothesis.[1] More recently, new sophisticated instruments and acquisition techniques, such as the Particle Image Velocimetry (PIV) and the Laser Doppler velocimetry (LDV), have made possible to measure instantaneous velocity fields with resolution equivalent to that of Large Eddy Simulations.[2–4] With these high quality datasets, it is now possible to reconstruct the large scale flow dynamics and compute global observables even in relatively complex geometries such as in the non-homogenous, non-isotropic von Kármán flow.[5,6] As more

a)Electronic mail: davide.faranda@cea.fr

26, 105101-1

spatial and temporal scales are becoming accessible to measurements, it is important to extract all the possible information from a statistical analysis of the data as it may eventually lead to test new theoretical predictions. In that respect, the integrated information obtained by measuring spectral features may not be sufficient to distinguish the contributions of different scales. Moreover, filter response functions used in spectral analysis may introduce spurious effects on the fast scales hiding an intricate structure.[7] Since present accessible measurements now give access to a large range of spatial scales, it therefore seems more promising to focus on turbulence properties in the physical space.

Alternative statistical description of turbulence in the physical space actually date back to Kolmogorov and Obukhov[1,8] and motivated formulation of stochastic models for the time evolution of turbulent observables. A now classical example is the Lagrangian stochastic model for the velocity of a passive tracer proposed by Thomson.[9] In that model, the inertial range of Lagrangian turbulent velocity is described through a Langevin equation, involving parameters that are determined via the so-called *Well Mixed Condition* (WMC). This model is in fact equivalent to an autoregressive process of order 1, usually denoted AR(1) or ARMA(1, 0) (see below) (ARMA = Auto Regressive Moving Average). Recent experiments however indicate that this simple model does not work for the velocity increments, which cannot be described by a simple standard Brownian motion as suggested by Obukhov.[10] Indeed, non normal corrections originate from long correlations due to the intermittent character of turbulent flows. There are several models that suggest a more refined description, based, e.g, on the Rapid Distorsion Theory,[11] on the account of the two-point two-time Eulerian acceleration-acceleration correlation,[12] on temporal memory kernel.[13] These approaches lead to excellent approximations of the experimentally determined velocity *pdf*'s, although an analytic solution for the model is still not available (for a review see Refs. 14 and 15). However, it is not clear whether these models directly correspond to the features really observed in turbulent experiments or, in other words, how far is a real experiment from the theoretical idealization.

To answer this question, as well as optimizing the information available from experimental measurements, it is mandatory to consider a more refined statistical analysis, able to account for temporal memory effects as well as velocity dependent diffusion coefficients. A good candidate is given by analysis in terms of ARMA(p,q) processes that have already been used to study problems ranging from geophysics to social science and finance.[16,17] This analysis aims to represent the statistical properties of a time series X_t using a model in which the value at time t is a combination of the p previous observations of the series—the so-called auto-regressive part AR(p) - and q noise terms - the moving average part MA(q)—with p and q chosen to be the lowest order to describe the series (see below). We observe that ARMA(p, q) processes are also good candidates to describe turbulent experimental data, since high p orders correspond to high temporal memory and high q orders correspond to a complicated structure of the diffusion coefficients. In the present paper, we will apply the ARMA modeling technique to large datasets obtained in the (inhomogeneous, anisotropic) von Kármán flow to illustrate the potential of this method.

The von Kármán experiment, in which the flow is generated in between two counter-rotating coaxial impellers, is a simple way to obtain experimentally a large Reynolds number ($Re \sim 10^6$) in a compact design.[18] In the equatorial shear layer, fluctuations are large and exhibit similar local properties as in large Reynolds number experimental facilities devoted to homogeneous turbulence. Away from the shear layer, one observes a decrease of the turbulence intensity. Overall, the flow is strongly turbulent, so that the instantaneous velocity fields, measured by means of a PIV system, strongly differ in a non-trivial manner from their time average.[6] Although significant advancements in understanding the physics of this system by statistical analysis[19] and from statistical mechanics approaches[20] have been made, several features of the flow remain unexplained and require further investigations. These include the nature of the phase transition recently discovered in the fully turbulent regime,[19] the forcing dependent stability of steady states[21] or the asymmetry of the torque probability distribution in different forcing conditions.[22,23] These features are based on both local measurements (such as velocity measurements using PIV or LDV techniques) or global measurements, such as total angular momentum, energy or torque applied to the rotating disks by the turbulence (drag friction). For any of these local and global measurements, we will define the ARMA(p, q) model which better represents the data, keeping in mind that the simplest model

explaining the Kolmogorov turbulent spectra is the ARMA(1,0) model (see below). We then try to answer the following questions:

- How far is a von Kármán flow velocity time series from an ARMA(1, 0) model?
- Can inhomogeneous anisotropic turbulence be better described by other ARMA(p, q) models and for which orders?
- Is there a spatial organization of ARMA(p, q) reflecting the spatial distribution of velocity inhomogeneities?
- Do different flow configurations correspond to different ARMA(p, q) models?
- At a pure statistical level, is there an amount of information that the ARMA modeling can extract with respect to other techniques?

The main achievement of the paper is to suggest that these questions can be positively answered with a rather simple analysis. Moreover, once the order p, q is identified, one has immediately a criterion to build continuous stochastic models similar to the ones introduced in Refs. 11 and 22 for the quantities analyzed. Our aim is thus to define a general technique which can be then used to analyze and critically extract information from any turbulence experiment. In the present paper we underline the general procedure, leaving specific applications to future publications. The paper is organized as follows: first we give an overview to present the relevance of the Obukhov model, then describe ARMA(p, q) models for turbulence by giving a survey of their statistical and mathematical properties. Then we present the experimental set up and the quantities analyzed with the algorithm. Finally, we present and discuss the results obtained, outlining perspectives for the analysis of general turbulent datasets.

II. ARMA MODELS OF TURBULENCE

A. From the Obukhov model to an ARMA(1, 0) process

The celebrated phenomenological theories of Kolmogorov and Obukhov[1,8] aimed to represent the complex phenomena of turbulence with a simple stochastic model. Thomson[9] was able to show that, in the inertial subrange, passive tracer Lagrangian velocities can be modeled by a Langevin equation (or Ornstein-Uhlenbeck process) with known coefficients; when discretizing this equation for simulation purposes, one can formally write it as an autoregressive process of order 1, usually denoted AR(1) or ARMA(1, 0). In particular, in the unidimensional case, the evolution of the velocity and of the position of a tracer particle (u, x) can be described by the stochastic differential equations:

$$\mathrm{d}u = a(x, u, t)\mathrm{d}t + b(x, u, t)\mathrm{d}W, \tag{1}$$

$$\mathrm{d}x = u\,\mathrm{d}t, \tag{2}$$

where $\mathrm{d}W$ are the increments of a Wiener process. In the same paper the determination of the coefficients a and b is discussed and it is found that, in Gaussian homogeneous turbulence, $a = -\frac{u}{T_L}$, where T_L is the Lagrangian decorrelation timescale, while $b = \sqrt{C_0 \epsilon}$, where C_0 is a universal constant and ϵ is the mean kinetic energy dissipation rate. This can be written, as suggested in Ref. 24, in terms of macroscopic quantities:

$$\epsilon = \frac{2\sigma_u^2}{C_0 T_L}, \tag{3}$$

where σ_u^2 is the fluid velocity variance (equal to the Eulerian variance) and can be seen as a measure of the turbulence intensity. Once the coefficients are known, one can write a discrete version of the Langevin equation (1):

$$\Delta u = -\frac{u}{T_L}\Delta t + \sqrt{C_0 \epsilon}\,\Delta W. \tag{4}$$

We are now considering a discrete-time stochastic difference equation, so we can use a discrete-time index $t \in \mathbb{Z}$ and, rearranging, Eq. (4) reads

$$u_t = \left(1 - \frac{\Delta t}{T_L}\right) u_{t-1} + \sqrt{C_0 \epsilon} \, \Delta W. \tag{5}$$

Denoting $\left(1 - \frac{\Delta t}{T_L}\right) = \phi$, $\sqrt{C_0 \epsilon \Delta t} = \sigma$ and recalling that $\{\Delta W\}$ are the increments of a Wiener process, the equation can be rewritten as follows:

$$u_t = \phi u_{t-1} + \varepsilon_t, \tag{6}$$

where $\{\varepsilon_t\}$ are independent variables, normally distributed. Equation (6) is the expression of an AR(1) process. To show that it is the simplest physical model which agrees both with Richardson law and the inertial range scaling proposed by Kolmogorov, it is sufficient to note that in an AR(1) process, the expected values of the velocity and the position scale in time, respectively, as

$$E[u^2(t)] \sim t, \quad E[x^2(t)] \sim t^3. \tag{7}$$

The second property is the Richardson law. Then, defining $\delta u = \sqrt{E[u^2(t)]}$ and $\ell = \sqrt{E[x^2(t)]}$, we get from Eq. (7) $\delta u \sim \ell^{1/3}$ which can be seen as an equivalent of the Kolmogorov scaling.

B. Generalization: ARMA(*p*, *q*) model for turbulence

The ARMA(1, 0) leads to a Markovian evolution for the Lagrangian turbulent velocity, and is unable to describe the intermittency or memory that have been shown to exist in real flows. In most laboratory turbulent flows, available datasets are time series of values of a physical observable at a fixed point or obtained by tracking Lagrangian particles. In our case, time series are obtained at fixed points in space; in this work, no spatial velocity profiles are studied. This historically motivated the shift of paradigm from *space* velocity increments to *time* velocity increments defined as $\delta u_\tau = u(t + \tau) - u(t)$ and motivated computations of the time structure function. Of course, in situations where measurements are made on the background of a strong mean velocity U, scale velocity increments and time velocity increments can be directly related through the Taylor hypothesis $\ell = U\tau$. In situations such that the fluctuations are of the same order than the mean flow, however, the Taylor hypothesis fails. A suggestion has been made by Pinton and Labbé[25] to then resort to a *local Taylor Hypothesis*, in which $\ell = \int dt u(t)$ where u is the local rms velocity. This is equivalent to consider a scale such that $\ell \sim \tau \delta u_\tau$ and may be seen as equivalent to modifying the *space* Kolmogorov refined hypothesis into a *time* hypothesis.

A natural generalization to take into account these features is thus to consider higher order ARMA(*p*, *q*) models, exhaustively treated, in example, in Ref. 26. A summary of useful notions about ARMA(*p*, *q*) modeling is provided in the Appendix. An ARMA(*p*, *q*) model corresponds to discrete time, stationary stochastic processes $\{X_t\}$ such that, for all t:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \vartheta_j \varepsilon_{t-j} + \varepsilon_t. \tag{8}$$

Here $\{\varepsilon_t\}$ is assumed to be a white noise of variance σ^2 and the polynomials $\phi(z) = 1 - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p}$ and $\vartheta(z) = 1 + \vartheta_1 z_{t-1} + \cdots + \vartheta_q z_{t-q}$, with $z \in \mathbb{C}$, have no common factors. Notice the white noise assumption is a very general condition and $X(t)$ will be normally distributed, resulting by a linear combination of independent and identically distributed random variables.

From a physical point of view Eq. (8) is the natural extension of the ARMA(1,0) model corresponding to the Obukhov model by introducing a temporal memory structure: Intuitively, the autoregressive part of the process expresses a dependence of the value of the process at time t on a linear combination of its own p previous values, while the moving average component introduces, at time t, a linear dependence of the q previous values of the noise term. The quantification of memory effect in real turbulent flows will then be made through fits of the data by an ARMA(*p*, *q*) model, and measurements of how *far* this model is from the ARMA(1,0) model. For this, we first need to

define the notion of *best* ARMA(p, q) fit, and then the notion of *distance* between the ARMA(1,0) and a given ARMA(p, q).

C. Model selection and characterization via correlation analysis

The main idea of time series analysis through ARMA models is to select the linear model that fits the data in the most parsimonious way, so that diagnosis of the nature of the generating process, forecasting or Monte Carlo simulations can be performed. Model selection is a non-trivial step of the procedure that can be addressed essentially by two means: through correlation analysis or through information based criteria, such as the Bayesian Information Criterion (*BIC*) described below.

For very simple processes, one can get access to the time dependence structure through computation of the auto-correlation function (ACF) and of the partial autocorrelation function (PACF) formally defined in the Appendix. In particular, for a MA(q) process, the theoretical ACF is characterized by q non-zero peaks, while the PACF decays exponentially or as a damped trigonometric function; for AR(p) processes the PACF is characterized by p non-zero peaks while the ACF decays exponentially. Hence, this fact allows to rule out or confirm the validity of an AR(p) or MA(q) hypothesis by a simple inspection of the ACF and PACF.

In the general ARMA(p, q) case, the simple correlation analysis described just above is not insightful. The model choice and the parameters estimation can be assessed by using the procedure introduced in Ref. 27, which also takes into account more complicated (such as integrated and seasonal) models:

1. preliminary analysis: the series is plotted in order to identify possible trends in mean and variance or periodic behaviors. Since here we deal with physically stationary processes, no trends are expected;
2. identification on the basis of the estimated ACF and PACF (or applying information criteria, such as the *BIC*);
3. estimation through maximum likelihood techniques;
4. diagnostic checking, that is, testing the estimated sequence for residual correlations (and normality or other distributive hypotheses, if required).

In the following analysis we perform the second step of the procedure fitting an ensemble of models with different (p, q) couples; we then choose the ARMA(p, q) model with the lowest total order $p + q$ producing not correlated residuals. The serial independence of the residuals series is tested as described in the Appendix. As already mentioned, this phase could be based on the value of the *BIC*. In this case, the information criterion is computed for each model: the best fit is the minimum *BIC* after the *steepest descent*. The two methods provide the same results. First of all, we tested them on a synthetic time series of 10^5 values simulated from an ARMA(3,1), obtaining a correct estimation of the model with both methods. To ensure that this technique is stable also for shorter time series, in Fig. 1 we show a *BIC* profile as a function of p and q for one of the analyzed velocity samples, consisting of 600 observations: both the methods lead to the choice of
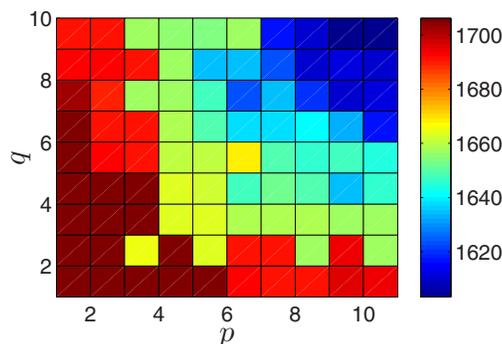
FIG. 1. *BIC* values resulting from different fits of ARMA(p,q) model for a velocity time series consisting of 600 observations.

an ARMA(7,7). All the PIV time series have length $n = 600$, while the typical LDV sample size is $n \sim 5 \times 10^5$.

D. A measure of distance from Kolmogorov theory based on the Bayesian information criterion

It is useful to concentrate the information obtained by analysis in a single index. We want to obtain a measure of the distance of the selected ARMA(p, q) model from the ARMA(1, 0), namely the Thomson-Obukhov model.

For a given dataset, the relative quality of a statistical model can be measured by the *BIC*, defined as

$$BIC = -2\ln \hat{L} + k[\ln(n) + \ln(2\pi)], \tag{9}$$

where $\hat{L}$ is the likelihood function for the investigated model. For an exhaustive definition of this quantity, see Ref. 28. Since the likelihood function is maximized when the correct model is found, while goes to zero in case of misspecification, its logarithm grows for well-specified models, while diverges to $-\infty$ otherwise. Thus, the first term globally tends to become negative or to assume small values once the best model form is identified. On the other hand, the second term grows with the number of parameters times the sample size: so it serves as a penalization for the number of parameters, in order to avoid overfitting. In brief, when testing an ensemble of models for a certain dataset, the best one is identified by the minimum value of the BIC.

For a Gaussian ARMA(p, q) model, it is expressed as follows:

$$BIC(n, \hat{\sigma}^2, p, q) = (n - p - q)\ln\left[\frac{n\hat{\sigma}^2}{n - p - q}\right] + n(1 + \ln\sqrt{2\pi}) + \tag{10}$$

$$+ (p + q)\ln\left[\frac{(\sum_{t=1}^{n} X_t^2 - n\hat{\sigma}^2)}{p + q}\right].$$

Notice that n is fixed by the experiment. The sample variance $\hat{\sigma}^2$ is computed from the sample and is a series-specific quantity. Thus, in order to obtain a meaningful definition of the distance from Kolmogorov model, the $BIC(n, \hat{\sigma}^2, p, q)$ must be normalized with respect to the Obukhov case $BIC(n, \hat{\sigma}^2, 1, 0)$:

$$\Upsilon = 1 - \exp\left\{|BIC(n, \hat{\sigma}^2, p + 1, q) - BIC(n, \hat{\sigma}^2, 1, 0)|\right\}/n, \quad 0 \le \Upsilon \le 1. \tag{11}$$

This quantity tends to zero if the dataset is well described by the Obukhov model and tends to one in the opposite case. We introduce the $p + 1$ correction to magnify small Υ values.

III. EXPERIMENTAL SET-UP AND DATA PROCESSING

In order to illustrate and apply these concepts, we have worked with a specific axisymmetric turbulent flow: the von Kármán flow generated by two counter-rotating impellers in a cylindrical vessel. The experimental set-up is described in Refs. 21 and 29. Here, we consider a configuration where the disks are exactly counter-rotating at frequency $f_1 = f_2 = F$, in the two forcing conditions associated with the concave (resp., convex) face of the blades going forward, denoted in the sequel by sense ($-$) (resp., ($+$)). The resulting mean velocity fields are quite similar, with two toric recirculations separated by a mean shear layer (see Fig. 2). The forcing conditions however strongly influence the level of fluctuations, which are much higher in the ($-$) case. The working fluid is water, with viscosity $\nu = 1.0 \times 10^{-6}$ m^2 s^{-1}. The Reynolds number is defined as

$$Re = 2\pi F R^2 \nu^{-1},$$

where R is the cylinder radius. We introduce a cylindrical system of coordinates $\vec{x} = (R, \varphi, Z)$ with its origin at the center of the cylinder and the z-axis aligned with the impeller's rotation axis (see Fig. 1 of Ref. 30).
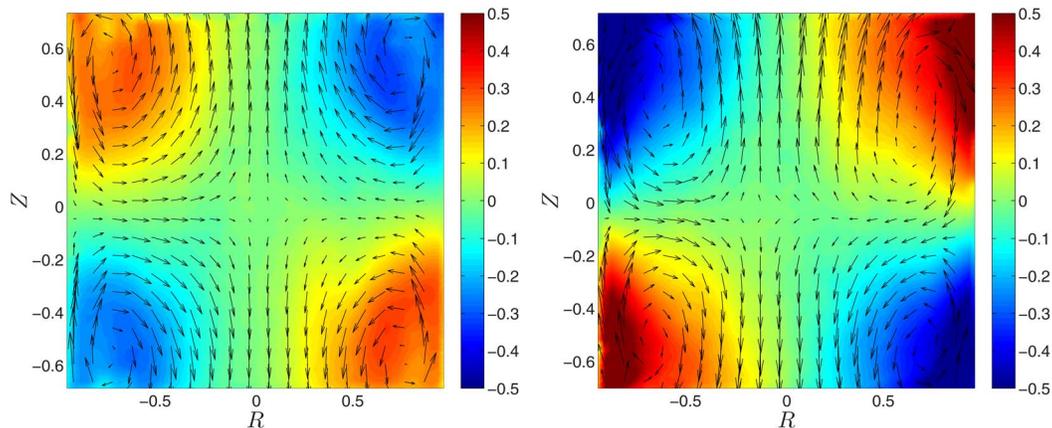
FIG. 2. Structure of the mean velocity field for $\theta = 0$. The arrows represents directions and intensities of the velocity components in the PIV plan averaged over time $(\bar{u}, \bar{v})$. The orthogonal component $\bar{w}$ is represented by the color scale. Left: (+) sense of rotation. Right: (−) sense of rotation.

In the sequel, we analyze both local and global observables. As local observable, we will consider the time series of the modulus of the velocity fields

$$\vec{V}(\vec{x}, t) = [u(\vec{x}, t), v(\vec{x}, t), w(\vec{x}, t)]$$

obtained by PIV measurements. Here u is the component in the PIV plane described in terms of R, the radial distance from the center of the cylinder; v in terms of Z, the vertical distance from the center, and w is the normal component to the PIV plane (the azimuthal velocity in cylindrical coordinates). For the comparisons between PIV and LDV measurements we will consider the normal component only w. We will also address two important aspects of statistical modeling of turbulence: the appearance of multifractal cascades, usually studied via the computation of the Hurst Exponents, and the role of phase randomization, which permit to isolate the effects of intermittency related only to the phase of the signals.

As global observables, we consider first the normalized kinetic energy introduced by Cortet *et al.*:[6]

$$\delta(t) = \frac{\langle V^2(t) \rangle}{\langle \bar{V}^2 \rangle}. \tag{12}$$

Here the brackets indicate the spatial average, and the bar a time average. $\delta(t)$ represents the ratio of the total kinetic energy of the instantaneous flow to the total kinetic energy of the mean flow. As a second global observable, we also consider the torque time series $C_1(t)$ and $C_2(t)$ experienced by the two motors. The goal is to compare which part of information about the flow is carried by observables built using the velocity fields (such as $\delta(t)$) and which is carried by dissipation measurements such as the torques.

The impeller speed F and the applied torques C_1 and C_2 are related to the average dissipation rate in the experiment, ϵ, through the injected power $\mathcal{P}$. The typical kinetic energy in the experiment, $\langle \bar{V}^2 \rangle$, can be directly computed from the PIV data. Knowing these global quantities, it is possible to obtain rough estimates of two typical quantities of the turbulent flow, the Kolmogorov typical length and time scales η and t_η and the Taylor typical scale λ, using the dimensional analysis inspired by Kolmogorov[1] and the identities of Taylor.[31] These relations are only valid for homogeneous and isotropic turbulence, which is not the case here: we will use them anyway to get rough estimates, presented in Table I for experiments conducted in water at $F = 5$ Hz, with curved blades and under various forcing conditions.

TABLE I. Typical length scales of the flow, using only global average quantities and usual turbulent identities.[31] Data are obtained with curved blades in both rotation senses, for $f_1 = f_2 = 5$ Hz, using water. Two turbulent states coexist in the $(-)$ sense due to hysteresis,[32] one with one recirculation cell and another one with two cells. The one-cell state is unstable in the $(+)$ direction for $f_1 = f_2$.

Sense	Cells	ϵ (m^2 s^{-3})	$\langle \bar{V}^2 \rangle$ (m^2 s^{-2})	η (m)	t_η (s)	λ (m)
$(-)$	1	28	7.8	1.4×10^{-5}	1.9×10^{-4}	2.0×10^{-3}
$(-)$	2	9.2	2.6	1.8×10^{-5}	3.3×10^{-4}	2.0×10^{-3}
$(+)$	2	2.5	1.3	2.5×10^{-5}	6.3×10^{-4}	2.8×10^{-3}

IV. RESULTS

We begin the analysis of the datasets by showing how the ARMA procedure, described in Sec. II, works on two velocity series extracted at two different locations from the same PIV experiment at $Re = 10^5$ and for the $(+)$ sense of rotation. The series $|\vec{V}(t)| = \sqrt{\vec{V}(t)^2}$ and their ACF and PACF are represented in Fig. 3. They have been obtained by sampling the data at 15 Hz. By analyzing the structure of the ACFs and the PACFs, one observes immediately that they do not consist of a small number of discrete peaks out of the confidence bands. This excludes the possibility that the series can be represented by pure AR(p) or MA(q) processes. Moreover, it is clear that a by-eye determination of the order (p, q) is not possible.

This result is consistent with the non-Markovian behavior used to describe the torque measurements via stochastic models in Refs. 11 and 22. By implementing our best fit procedure, we find that the best ARMA model to fit the data depends on the measurement points: the series on the left-hand side of Fig. 3 is fitted by an ARMA(1,1) model whereas the other one by an ARMA(4,2) model. This is of course not surprising, because the von Kármán flow is highly inhomogeneous. In the remaining of this section, we analyze the relationships between the flow inhomogeneous spatial structure and the ARMA fit structure by mapping the ARMA parameters.

A. Velocity fields

Let us now analyze the spatial structures obtained by applying the procedure described in Secs. II and III for a PIV field taken at $Re = 2 \times 10^5$, with mean velocity field provided in Fig. 2
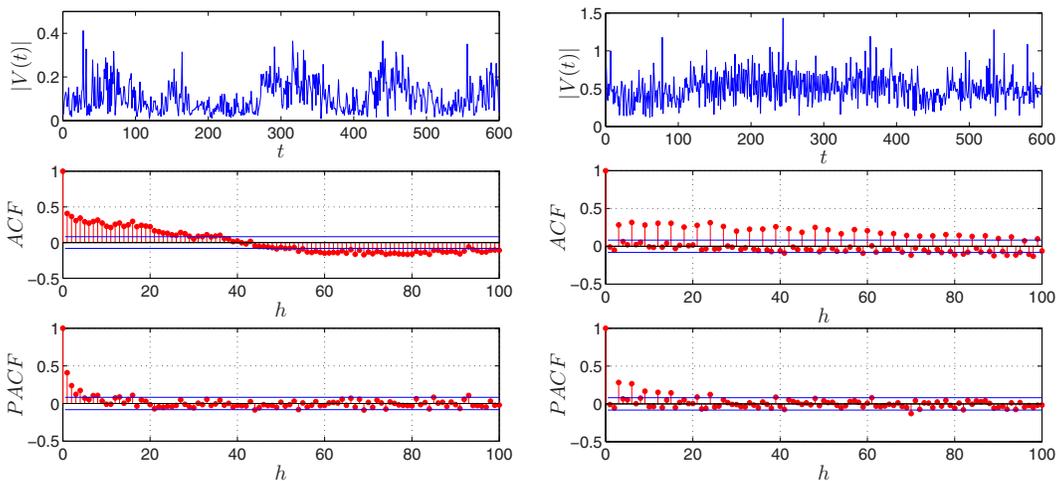
FIG. 3. Two time series of $|V(t)|$ (upper panels) with their respective ACF functions (middle panels) and PACF (lower panels). $Re \simeq 10^5$, $(+)$ sense of rotation. Blue lines in the ACFs and PACFs represent the confidence bands at the 95% confidence level. Sample frequency: 15 Hz. X-axis is in sample index.
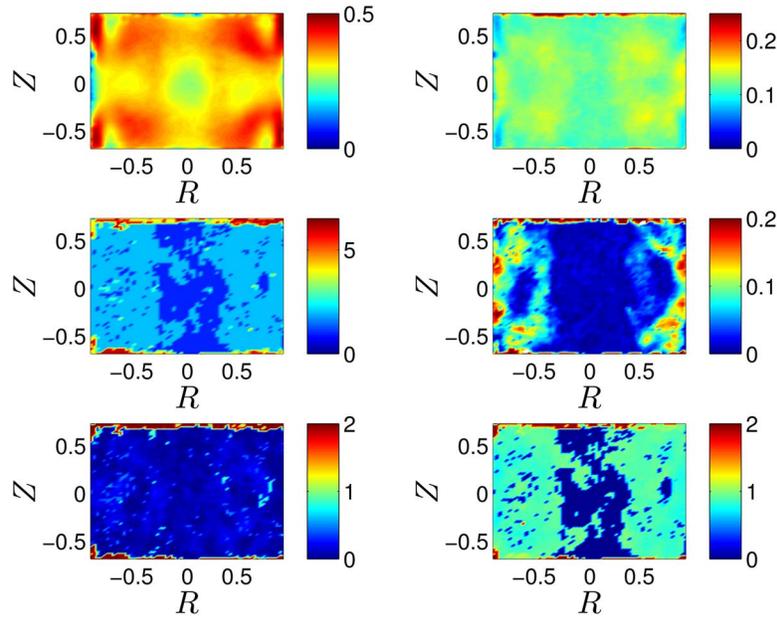
FIG. 4. ARMA analysis for the (+) sense of rotation. Top left: $|\bar{V}|$. Top right: $|\vec{V}(t)|$ standard deviation. Center left: Total order $\mathcal{O}$ found by fitting an ARMA(p,q) to the $|\vec{V}(t)|$ data. Center right: Distance from the Kolmogorov model Υ for the $|\vec{V}(t)|$ data. Bottom left: Sum of the autoregressive coefficient Φ. Bottom right: Sum of the moving average coefficients Θ.

for the (+) sense of rotation (left) and (−) (right). The two pictures look extremely similar: one can immediately recognize the cells structure of the flows described in Sec. III.

A full overview of the quantities computed by using the ARMA analysis is presented in Fig. 4 for the (+) sense of rotation and in Fig. 5 for the (−) sense. Obviously, even if the four cells structure presented in Fig. 2 is recovered in both the situations, the average over time of $|\vec{V}(t)|$ denoted as $|\bar{V}|$
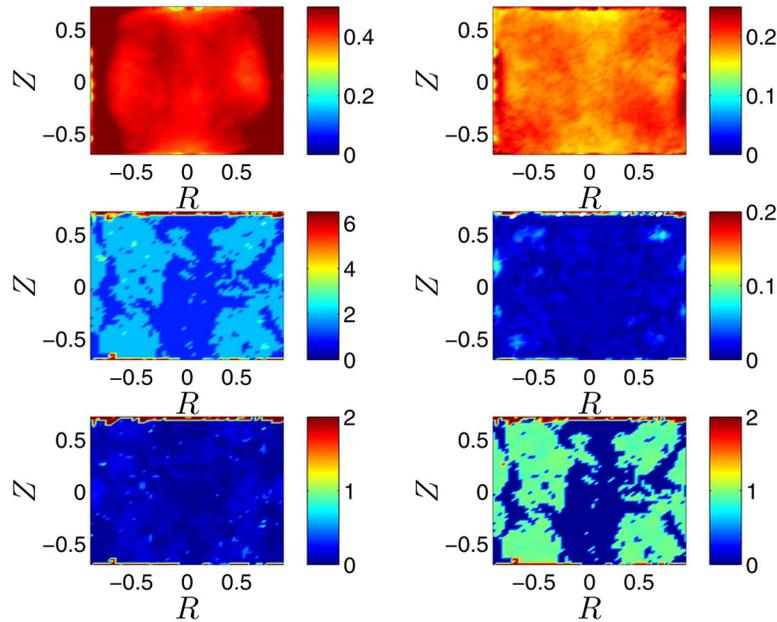
FIG. 5. ARMA analysis for the (−) sense of rotation. Top left: $|\bar{V}|$. Top right: $|\vec{V}(t)|$ standard deviation. Center left: Total order $\mathcal{O}$ found by fitting an ARMA(p,q) to the $|\vec{V}(t)|$ data. Center right: Υ for the $|\vec{V}(t)|$ data. Bottom left: Sum of the autoregressive coefficient Φ. Bottom right: Sum of the moving average coefficients *Theta*.

(top left panels) and the standard deviation of $|\vec{V}(t)|$(top right panels) show remarkable differences between the two configurations. In the $(+)$ sense of rotation the four cells structure is appreciable whereas higher mean values and fluctuations of $|\vec{V}(t)|$ are recorded in the proximity of the wall of the cylinder in the $(-)$ rotation. In this latter configuration, the fluid is pushed to the side of the cylinder and higher turbulent fluctuations are registered, as described in Ref. 5. Let us now analyze what happens to the quantities introduced by the ARMA analysis.

1. Order of ARMA

We start by describing the behavior of the total order $\mathcal{O} = p + q$ of the processes fitted for each time series (middle left panels). In both senses of rotation, the highest orders are concentrated near the impellers. However, differences appear in the other regions of the domain. In the $(+)$ sense of rotation, the highest orders are found in correspondence to the highest fluctuations. Near the center of the cylinder the orders are low and, for some of the series, the signal is indistinguishable from noise (order 0). In the $(-)$ set up, the highest level of turbulent fluctuations contribute to homogenize the behavior in larger areas such that a weak four cells structure is recognizable. This effect is probably linked to the presence of more homogeneous fluctuations in the flow (top right panel of Fig. 5). Even if the $\mathcal{O}$ color scale has been limited at $p + q = 6$ for comparison with the $(+)$ situation, we underline that much higher orders appear in the $(-)$ setups near the impellers and, locally, at the walls of the cylinder. High p, q orders are directly connected to the vortices introduced by the rotations of the impellers and whose appearance is explainable in terms of Goertler instabilities.[5] We will see in Sec. IV D that these effects are recovered also for global observables.

2. Distance from Kolmogorov model

The difference between the $(-)$ and $(+)$ configuration is also highlighted by the results of the Υ computations reported in the middle right panels of Figs. 4 and 5. The distance from Kolmogorov model is lowest and almost zero near the boundaries, where the fluctuations are modest, and increases towards the center.

Evident differences appear if one compares Υ values for $(+)$ and $(-)$ senses of rotation. As expected, the highest values are found in the $(+)$ case, which is the one preserving a spatial four cells pattern in the fluctuations. This suggests that the coherent structures visualized using bubble air seedings are responsible for deviations from the Obukhov model.

In the $(-)$ set-up, the region of values of $\Upsilon \geq 0.1$ clearly traces the area with maximal azimuthal velocities. We have therefore a clear connection between coherent structures. In the present data set, we do not observe obvious signature of the influence of the shear layer dynamics. However, by using a much larger data set, we have been able to evidence the signature of the wandering of the shear layer in between to metastable position. This is reported in Ref. 33.

3. Physical interpretation of ARMA(p, q) coefficients

The bottom panels refer to the sum of the coefficients $\Phi = \sum_{i=1}^{p} |\phi_i|$ (bottom left panels) and $\Theta = \sum_{i=1}^{q} |\vartheta_i|$ (bottom right panels). Φ and Θ may be regarded as a representation of the total persistence of the phenomena, i.e., how much the system remembers of its past history. In order to get a better understanding of this idea and thus obtain a physical interpretation of the AR and MA parts of the process, we exploit, once again, Thomson's model. Equation (1) implies a Markovian evolution of the Lagrangian velocity in the inertial subrange, which is linked to an exponential behavior of the ACF:

$$\rho(t) \sim e^{-\frac{t}{T_L}}.$$

We can observe that the first term of the rhs of the equation contains some information about the global correlation structure of the process, which is even more evident in the discrete time (see Eq. (5)), since the autoregressive coefficient $\phi = (1 - \frac{\Delta t}{T_L})$ is the Taylor expansion of the exponential ACF. We have already mentioned that $|\phi|$ is a measure of the persistence of the process; here, this

persistence is driven by the large eddies, since T_L is the Lagrangian decorrelation time scale, which is $T_L \gg \Delta t$, with Δt lying in the inertial subrange.

On the other hand, since the noise term in Eq. (1) is a standard Brownian motion, the innovations of the time series in Eq. (5) ΔW, are normally identically distributed and δ-correlated. Thus, the second term of the equation is simply a noise driving the process, with no linear dependence between any couple of values $(\Delta W_i, \Delta W_j)$ with $i \neq j$; this assumption is due to the fact that the stochastic kicks come from the viscous eddies, which live in the viscous time scale $\tau_\eta \ll \Delta t$: since the particle samples the turbulent field with a frequency linked to a characteristic time $\sim \Delta t$, the viscous eddies are completely uncorrelated between two steps. This means that here $q = 0$ and the innovations of the process are a pure (Gaussian) noise, and all the information is contained in the mean and the variance.

In general, we may say that the AR(p) part of an ARMA(p, q) process is linked to the contribution of the large scales and represents the persistence of the process. Notice that, if $p = 1$, it must be $|\phi| < 1$ in order to satisfy the stationarity condition; if the process is more persistent than an AR(1) with $|\phi| < 1$, higher values of p are required to explain all the correlation coming from the large scales. Analogous considerations hold for the MA(q) part: if $q > 0$, a linear combination of previous values of the noise appears in the equation, introducing a correlation structure in the innovation term, i.e., a higher persistence of the noisy contributions. This means that the small eddies do not decorrelate completely between two *sampling times*, so we should assume to have eddies at all scales.

4. Comparison with high-resolution datasets (LDV)

The results obtained with the PIV technique must be validated and checked against higher temporal resolution datasets. In fact, although the possibility of defining a distance from the Kolmogorov model in the physical space rather than in the Fourier space seems appealing, we must be sure that the results obtained with the ARMA analysis are stable with respect to an increase of resolution.

In the previous discussion, we have pointed out that the Obukhov model, representing homogeneous and isotropic turbulence, can be written as an ARMA(1,0) model. This corresponds to have a purely power-law spectrum which does not contain any other features than the decay predicted by Kolmogorov. Since Eq. (5) contains an explicit dependence on Δt only for the coefficient ϕ, we do not expect to see a change in the order of the process when increasing the resolution, but rather changes in ϕ and ϑ coefficients. This consideration holds unless the spectrum changes slope or has peaks for some of the frequencies we add to the spectrum by increasing the resolution. In this case we expect to see also a change of the autoregressive and moving average polynomials.

In our analysis, we compared the PIV data for the ($-$) sense of rotation with the LDV experiment performed in the same conditions. Since for the LDV series only the w component is measured, we will compare this quantity to the same recorded for the PIV experiments. The LDV data allow for exploring frequencies of order of the kHz, whereas the PIV is limited to a frequency of 15 Hz, so that we extend significantly the range of frequencies analyzed. Whereas the time resolution of the LDV data is very high, the spatial resolution is indeed low: we have w measurements only at the 18 points represented by the red crosses in the top-left panel of Fig. 6. For this reason, the quantities obtained from the LDV analysis (left panels of Fig. 6) have been interpolated on a finer spatial grid. Anyway, the level of details remains lower if compared with the PIV results (right panels of Fig. 6). The top panels of Fig. 6 show a comparison between the averaged velocity field as obtained from the LDV and the PIV analysis. They both show not only the familiar cells structure, but also that the order of magnitude of the velocity fields is extremely close for the two different techniques. The analysis of the orders $\mathcal{O}$ (reported in the central panels of Fig. 6) shows consistency between the two techniques: the highest orders are located at the walls of the cylinder and near the impellers. Moreover, if we average the total order on all the available points, we get $\mathcal{O} = 2.4 \pm 0.7$ for the LDV and $\mathcal{O} = 1.8 \pm 0.8$, values compatible within a standard deviation. Finally, the analysis of Υ (lower panels of Fig. 6) reveals that the maxima are located, both for the PIV and the LDV, near the walls of the cylinder around $Z = 0$. By computing the average of Υ over all the points we find 0.03 ± 0.02 for the LDV data and 0.02 ± 0.02 for the PIV, again consistent within a standard deviation.
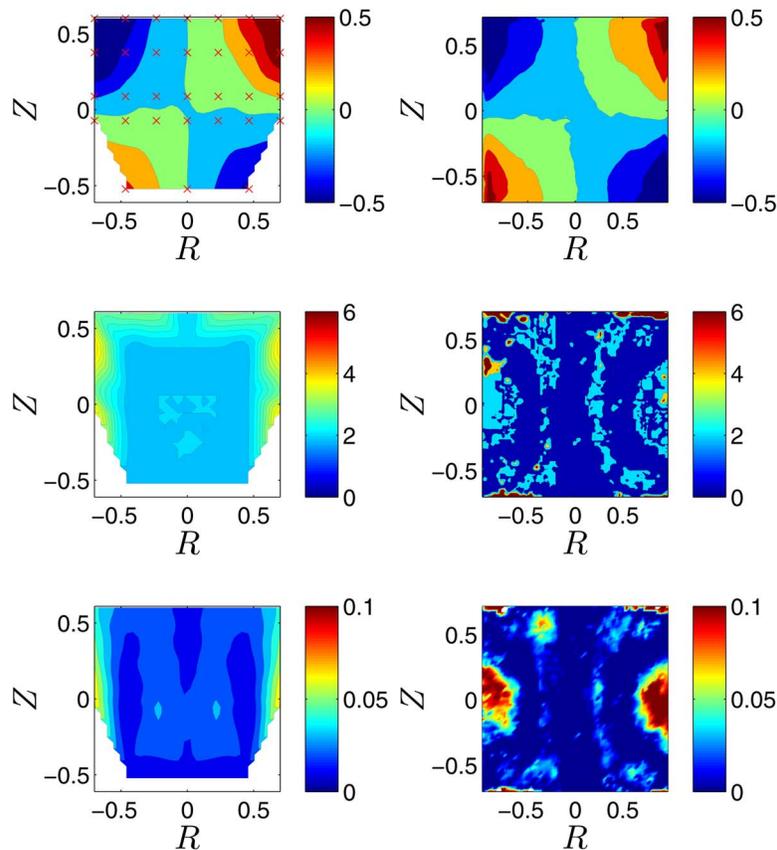
FIG. 6. Comparison between the ARMA analysis for LDV data (left) and PIV data (right) for the azimuthal component of the velocity w. $\bar{w}$ (upper panels), total order $\mathcal{O}$ (central panels) and distance from Kolmogorov model Υ (lower panels); $(-)$ sense of rotation; $Re \simeq 10^5$. The red crosses in the top-left panel show the locations of the measurement points for the LDV experiment.

B. ARMA analysis of phase randomized data: Phase intermittency

Phase randomization is often used in turbulence to destroy the intermittency effects related to Fourier phases while preserving the intermittency effect related to Fourier amplitude:[34] by applying such procedure one preserves up to the second order statistics (covariance and spectrum). It is therefore interesting to apply the ARMA analysis to phase randomized data sets to quantify the relative influence of phase and amplitude intermittency in turbulence.

A simple and efficient way to perform this phase randomization is to compute the Fourier transform of the time series, then randomize the phase (while preserving the anti-symmetry of the phase with respect to the frequency variable resulting from the real nature of the data) and going back to the physical space by means of an inverse Fourier transform. In order to perform this task, we have used the MATLAB code provided by Carlos Gias, based on the procedure described in Ref. 35.

The results we present correspond to the $(-)$ sense of rotation for the PIV data already analyzed in Sec. IV. After generating surrogate velocity data, we compute Υ for the phase-randomized data and compare it with the original one. This is done in the lower panel of Fig. 7, with the original Υ (left) and the Υ for phase-randomized data (right). Both panels show the same structure meaning that most of the contribution to the intermittency parameter is associated to intermittency amplitude through first and second order statistics (presumably through the advection and shearing effect of the large scale flow). To get information about phase intermittency, we subtract the intermittency index from the phase-randomized data to the original ones, obtaining $\Delta\Upsilon$. Results of such a difference are reported in the upper panel of Fig. 7. It is about two orders of magnitude smaller than the amplitude
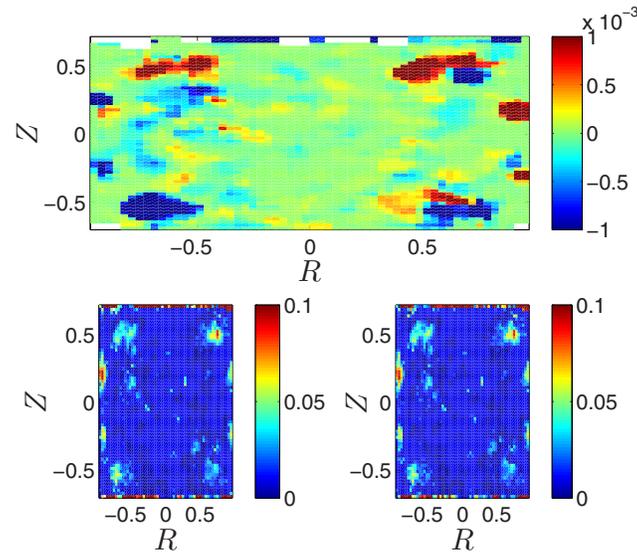
FIG. 7. Upper panel: Difference between Υ computed for the $|\vec{V}(t)|$ data for the $(-)$ sense of rotation and Υ on surrogate date of the same experiment obtained with a phase randomization procedure. Lower panel: Υ for the original data (left) and for the surrogates (right).

intermittency. Its resulting spatial structure is highly organized, showing some association with vortices. This feature may be connected with the observation that anomalous scaling, in linearly advected hydrodynamical models, is connected to the existence of statistically preserved structures with highly complex geometrical properties.[36,37] We leave this for future investigation.

C. Hurst exponents

A generalized version of the first equation in the system (1) can be written as

$$ du = a(x, u, t)dt + b(x, u, t)dW^{2H}, \tag{13} $$

where dW^H is the increment of a fractional Brownian motion (fBm) and H is the so-called Hurst exponent.[38] The fBm, first introduced by Mandelbrot and Van Ness,[39] is a generalization of Brownian motion where the increments are not independent. It has zero mean and the following covariance function:

$$ E[W_H(t)W_H(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}). $$

The exponent H is a real number in $(0, 1)$ and its value determines the memory of the stochastic process. For $H = 1/2$, the standard Brownian motion is recovered. For $0 < H < 1/2$ the process is anti-persistent, i.e., an increase will most likely be followed by a decrease or vice-versa. Finally, for $1/2 < H < 1$, the series is persistent, i.e., increases generally follow increases.

We want to investigate if the behavior displayed by the total order and total persistence of the ARMA(p, q) and by the distance index Υ can be better explained by the fractional nature of the underlying stochastic process. In order to do this, we compute H for the $|\vec{V}(t)|$ data in the $(+)$ sense of rotation and we compare it to Υ in Fig. 8. The computation of the Hurst exponents follow the methods presented in Ref. 40 which we found to be all consistent with each other. Since Υ values span 5 orders of magnitude while H is always of order 1, we consider the $\log_{10}(\Upsilon)$. Not only the spatial structure of H and $\log_{10}(\Upsilon)$ are very similar (upper panels of Fig. 8), but also a linear relation can be found between these two quantities (lower panel of the same figure). The linear correlation coefficient is $r = 0.70$ and these results hold also for the LDV experiments with almost identical fit coefficients and $r = 0.81$. From this analysis we argue that a fBm description of the phenomenon might be used to explain the nature of the correlations in the series and it could be useful to further
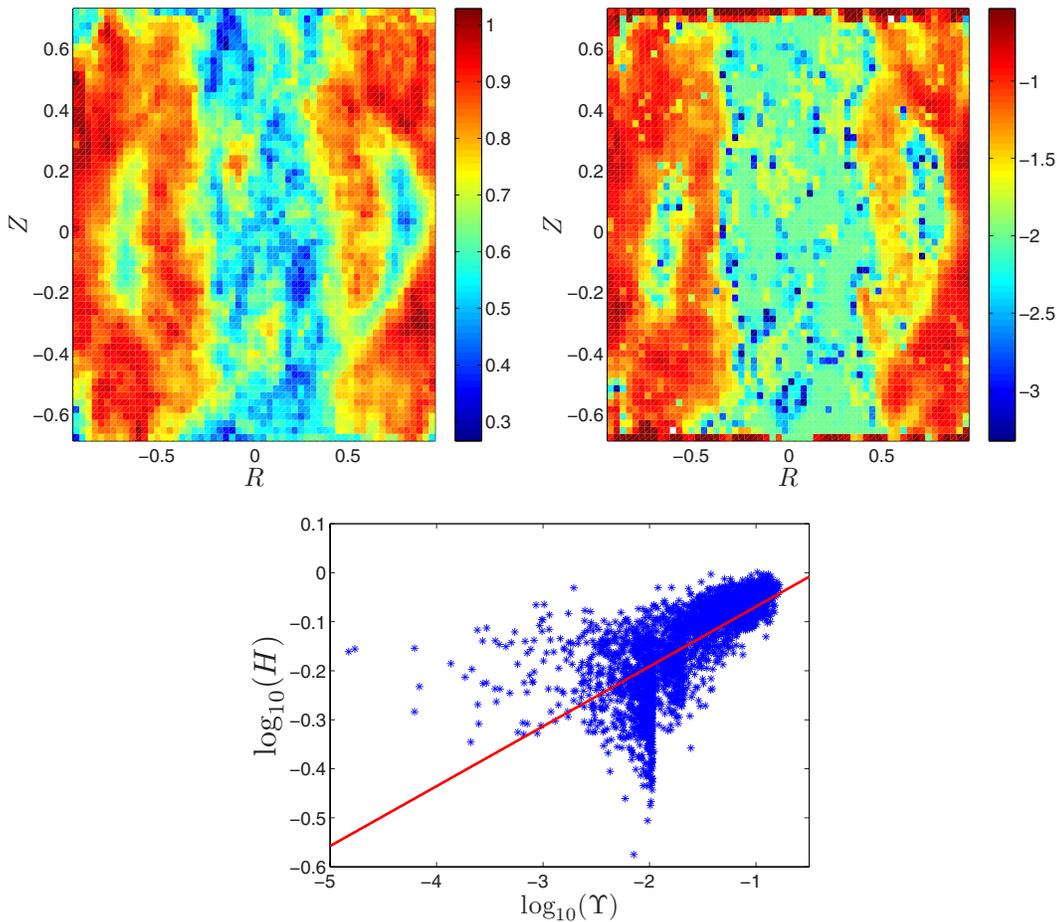
FIG. 8. Upper panels: Comparison between the Hurst exponents H (left) and the deviation from the Kolmogorov model in log-scale $\log_{10}(\Upsilon)$ (right) found by fitting an ARMA(p,q) to the $|\vec{V}(t)|$ data for the ($+$) sense of rotation. Lower panel: Scatter plot of the Hurst Exponent H and $\log_{10}(\Upsilon)$. The red solid line shows a linear fit to the data.

improve the modeling of inhomogeneous and anisotropic turbulence. However, the results show that Υ, based only on an ARMA(p, q) estimation, is equally effective in quantifying deviations from the Kolmogorov model which could not be due to fBm effects.

D. Global observables

An interesting question to address when dealing with spatial-temporal extended systems, is how the information on the single trajectories is transmitted to integrated quantities. In particular, one may ask whether the differences found in the ARMA analysis for the local observables of PIV fields are preserved for scalar quantities, i.e., if high ARMA(p, q) orders found locally in the proximity of the impellers and the cylinder walls give a contribution to global observables or whether they average out. In this section we present results obtained for the quantity $\delta(t)$ introduced in Eq. (12). We have further tested that our results are independent of the choice of the global observable, whether derived from PIV measurements — Angular momentum — or measured independently like torques measurements.

We have carried out the analysis on global observables at several Reynolds numbers around $Re = 10^5$, that is in a fully turbulent regime. The typical behavior of the time series of $\delta(t)$ is represented in Fig. 9 for the ($+$) sense of rotation (left), and the ($-$) one (right). The top panel refers to the time series obtained by averaging the spatial velocity fields and shows no particular
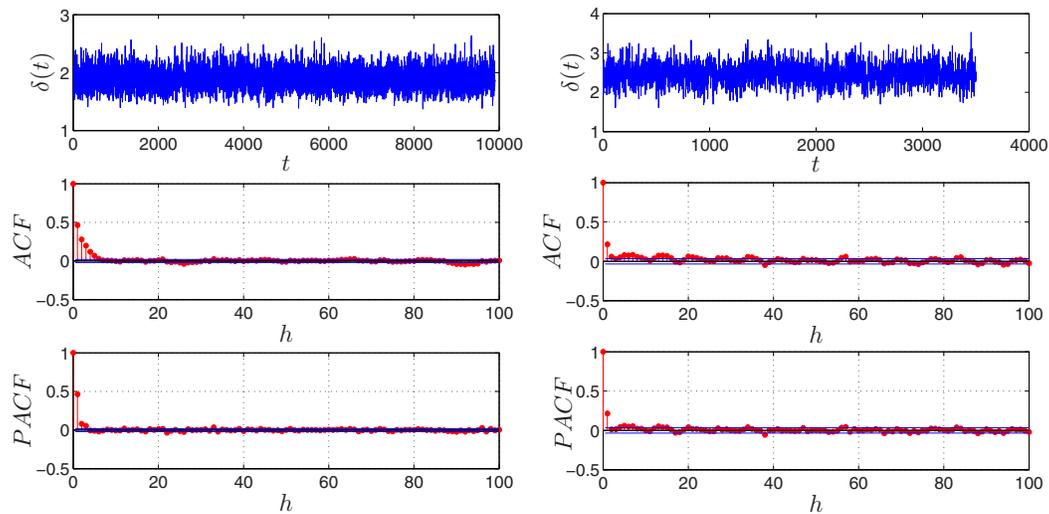
FIG. 9. Time series of $\delta(t)$ (upper panels) with their respective ACF functions (middle panels) and PACF (lower panels). Left: (+) sense of rotation. Right: (−) sense of rotation. $Re \simeq 10^5$. Blue lines in the ACFs and PACFs represent the confidence bands at the 95% confidence level. Sample frequency: 15 Hz. X-axis is in sample index.

differences at first sight, as we have seen for the examples of the velocity series shown in Fig. 3. However, the ACF (middle panels) and PACF (lower panels) are remarkably different. The ACF of the (+) sense of rotation decays quickly and the PACF shows only one peak significantly different from zero: an ARMA(1,0) is enough to explain the correlation structure. On the other hand, an oscillatory behavior of both the ACF and PACF is clearly recognizable for the (−) rotation. The orders p, q needed to decorrelate the latter time series are higher in the (−) rotation, namely $p = 2$, $q = 1$. These results hold generally by varying Re and changing observables and point to the intrinsic differences between the two senses of rotation.

One can notice that some characteristic features appear in the ACF of the (−) sense of rotation. We can speculate that, for scalar quantities, the highest orders get averaged out if their contribution is substantially different at different (r, z) as it happens for the (+) rotation. However, when the same kind of features are present in the ACF and PACF for series at different (r, z) the contribution sums up and is well visible in the behavior of global observables.

V. MULTISTABILITY

Another interesting question is whether the application of ARMA techniques to turbulence is helpful to discriminate between different stability regimes. The simple guess is that by increasing the instability of a configuration, higher orders $\mathcal{O}$ arise as we introduce in the system new time scales linked to the presence of nearby attracting states. In order to check this idea, let us consider again a von Kármán swirling flow with the same geometry described before, with the Reynolds number fixed at $Re \sim 10^5$. In this system, one can impose either the speeds f_1 and f_2 of the motors or the torques C_1, C_2 and define two natural dimensionless quantities:

$$\theta = (f_1 - f_2)/(f_1 + f_2), \quad \gamma = (C_1 - C_2)/(C_1 + C_2)$$

which are, respectively, the reduced impeller speed difference and the reduced shaft torque difference. In Ref. 21, the authors found that different forcing conditions change the nature of the stability of the steady states. Here we complete the results represented in Fig. 1 of Ref. 21, with the ARMA analysis in terms of the quantities $\mathcal{O}$ and the total persistence of the process defined as

$$\mathcal{R} = \sum_{i=1}^{p} |\phi_i| + \sum_{i=1}^{q} |\theta_i|.$$
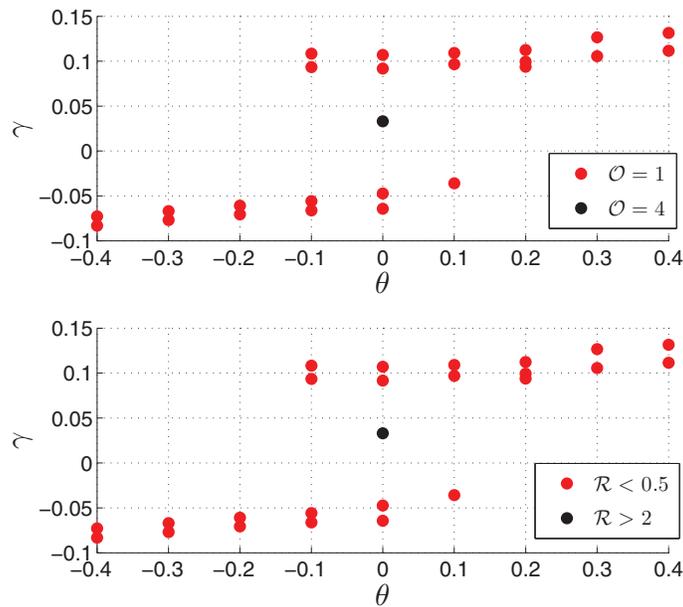
FIG. 10. Total order $\mathcal{O}$ and total persistence $\mathcal{R}$ for the von Kármán experiment under the speed control. The points represent the averaged γ and θ obtained for each experiment.

Speed control: In this case all the turbulent flows are steady. By plotting the averaged θ's and γ's measured in several experiments, one obtains the diagram shown in Fig. 10. The colors refer to different $\mathcal{O}$ (Fig. 10, upper panel), and $\mathcal{R}$ (Fig. 10, lower panel). Starting both impellers at $\theta \simeq 0$ leads to a marginally stable state, which consists of two symmetric recirculation cells separated by a shear layer. If one waits enough time, a fluctuation may force a jump of the system to one of the two bifurcated states represented by the red points. The instability of the symmetric state is reflected by the order of the ARMA processes fitted for the series of $\gamma(t)$ at $\gamma \simeq \theta \simeq 0$. For this experiment we found $\mathcal{O} = 4$ and $\mathcal{R} \simeq 3$, values definitely larger than the ones found in the bifurcated states where always $\mathcal{O} = 1$ $\mathcal{R} < 0.5$. From the available data one can argue that the potential barrier—the repellor in dynamical system—at $\theta = \pm 0.1$ is somehow impenetrable as we do not get any increase in $\mathcal{O}$ and $\mathcal{R}$ for the series $\gamma(t)$ recorded at such values of θ.

Torque control: By imposing the torque control one gets access to new attracting states, located in correspondence to the repellor found in the speed control. The results for the torque control have been obtained by analyzing time series of $\theta(t)$ and the results in terms of $\mathcal{O}$ and $\mathcal{R}$ are reported in Fig. 11. Before commenting on the new states, we begin by analyzing the states which are attracting in both the configurations. In Ref. 21, the authors assert that the properties of the attracting states in the speed control set up are analogous to the ones found for the torque control. However, by applying the ARMA analysis, we found, as one would expect, remarkable differences. In particular, the symmetric state (which was marginally stable in the speed control) is now stable as one can go to the bifurcated states in a continuous way. This is confirmed by the low values of $\mathcal{O}$ and $\mathcal{R}$ found for the torque control where $\mathcal{O} = 1$ (Fig. 11(a), found in correspondence of $\gamma \simeq 0.005$ and $\theta \simeq 0$) and $\mathcal{R} < 1$ (Fig. 11(b)). The bifurcated states have a different characteristic order (typically $\mathcal{O} = 2$) and persistence (typically $1 < \mathcal{R} < 2$). These are not linked to the presence of transitions, as they persist further away from the unstable range of parameters. They are linked to the modifications in the dynamics induced by the change of control which affects also the stable regions in a fine way, which has not been discussed in Ref. 21 but is evident by applying the ARMA technique capabilities. Let us now comment on the new states which appear in the torque control in correspondence to the repellor for the speed control. These states feature multistability as detailed in Ref. 21. In terms of ARMA analysis they are characterized by higher orders (green branches in the upper panel of Fig. 11 with $\mathcal{O} = 3$), and persistence (black branches of the lower panel in Fig. 11). Even if an
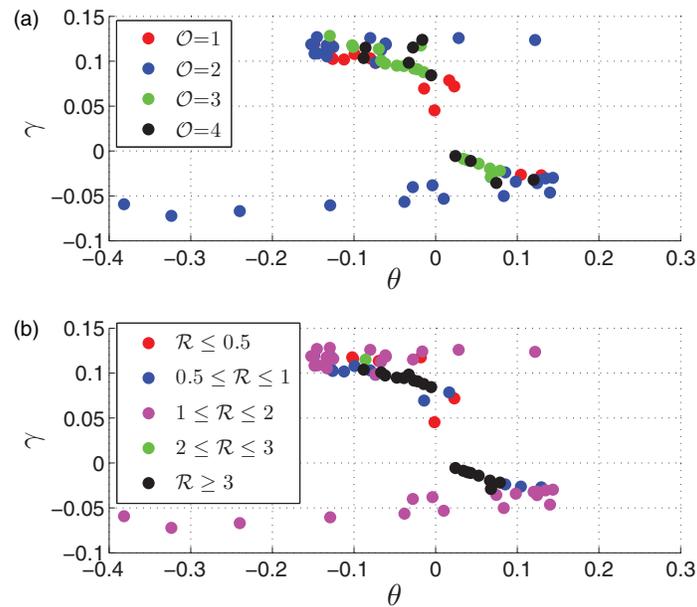
FIG. 11. (a) Total order $\mathcal{O}$ and (b) total persistence $\mathcal{R}$ for the von Kármán experiment under the torque control. The points represent the averaged γ and θ obtained for each experiment.

order is found for multistable time series, an ARMA model cannot just be stuck to the data in this case, as it will not reproduce a multistable behavior, but rather a process with one stable state whose correlation properties are similar to the ones found for the multistable time series.

This example clearly shows that one can find, far from the bifurcation, a typical ARMA(p, q) process ($\mathcal{O} = 1$ for the speed control and $\mathcal{O} = 2$ for the torque control) which describe the data-sets whereas p and q are evidently modified by the stability properties. We remark that one must pay extraordinary care when the goal is to find a model for the data-set.

VI. FINAL REMARKS

In this paper we have shown how to extract information from turbulent data-sets by applying an ARMA statistical analysis. Such analysis goes well beyond the analysis of the mean flow and of the fluctuations; in fact, it is possible to link the behavior of the recorded time series to a discrete version of a stochastic differential equation which is able to describe the correlation structure in the data-set. We have tested the method on data-sets produced by the experiments on the von Kármán swirling flow, already analyzed in several publications.

We have shown how the anisotropies and inhomogeneities present in real experiments as well as the finite resolution of the data-sets influence the order p, q of the process which better describes the data. We find that data are suitably described by ARMA(p, q) processes whose orders are generally different from the Obukhov model although with a very limited number of auto-regressive and moving average terms (generally p, $q = 1$ or 2). We have introduced a new index $0 \leq \Upsilon \leq 1$ to measure and quantify this difference. The value of Υ increases in areas where large scale coherent structures are present. It would be interesting to rely the statistics of Υ to the computation of refined statistics of velocity increments, possible only for time high-resolved experiments.[41] In particular, we aim to compare Υ with classical intermittency parameters based on structure functions. This idea follows from the hypothesis first proposed by Laval *et al.*[42] that intermittency *propagates* in direct interactions between large and small scales, rather than in cascades. Preliminary analysis carried on the LDV data-sets show that there is a linear proportionality between Υ and the classical intermittency parameters defined on the velocity increments.

The inhomogeneous structure of the PIV experiments is reflected by the range of different orders p, q found in our analysis: a great part of the flow can be described in terms of noise, whereas higher orders concentrate around cells ((+) sense of rotation) or near the walls ((−) rotation). The analysis of global observables shows that most of the information about the local structure of the flow is preserved, including the differences found between the two senses of rotation. This correspondence between local and global quantities is very important and it will be further exploited for challenging systems for which only global observables are available as in the SHREK data-sets experiment with super-fluid Helium.[43]

We have also checked our results against the change in time resolution by comparing them against the LDV experiments, whose average sampling frequency is order of the kHz. Not only the values of $\mathcal{O}$ found with this technique are consistent, but also the values of Υ and the spatial structures observed. Finally, we have commented on the effects of multistability on the ARMA analysis by considering two different kind of forcings for the von Kármán experiment.

In the Obukhov model, the coefficients of the stochastic model are given from the turbulence theory, resulting in a simple Langevin equation which describes the process in the continuous time. Here, we have applied estimation methods to obtain a parametric description in the discrete time, but the passage to continuous time stochastic differential equations is not trivial for a general ARMA(p, q) process. Obtaining an expression of the model's parameters in terms of physical quantities of turbulence theory is presently not possible. In fact, it is likely that the MA(q) part of our processes represents the contribution of the shortest time scales detectable with the available techniques. One way to test this idea is to verify that only the structure of the MA(q) part of the process changes by changing only small scales feature of the flow. In order to do that, we are currently testing impellers with a fractal structure. Preliminary analysis show that MA(q) orders are different between fractal and non-fractal impellers whereas the AR(p) do not change. Details will be reported in a future publication.

Several generalizations of ARMA models exist and they allow for taking into account the possible multi-fractal behavior of turbulence. The comparison of the Hurst exponent and the Υ index suggests that it will be interesting to extend the analysis to fractional integrated ARMA or ARFIMA(p, H, q) models. The first ones can be appropriated for studying problems of non-stationary turbulence[44] whereas a SARMA models analysis could be suitable for studying problems of wave turbulence.[45]

ACKNOWLEDGMENTS

APPENDIX: THEORETICAL SURVEY ON ARMA MODELING

We have already mentioned, in the definition of the ARMA(p, q) process, that $\{X_t\}$ must be stationary. Usually, two definitions of stationarity are given when treating discrete time stochastic processes: *strong stationarity* implies stationarity of the whole joint probability distribution of the stochastic process, while *weak stationarity* requires the first two moments of the process to be finite and constant in time. The results about ARMA(p, q) processes are usually proved requiring weak stationarity, that is of course implied by strong stationarity; in our data analysis we will assume stationarity on a physical basis, by studying the system when the dynamics has reached well identifiable stationary states.

First of all, we observe that Eq. (8) can be written in a very compact form, introducing the backward operator B such that $B^j X_t = X_{t-j}$, $j \in \mathbb{Z}$:

$$\phi(B)X_t = \vartheta(B)\varepsilon_t. \tag{A1}$$

If $\vartheta(z) \equiv 1$ the process reduces to an AR(p), if $\phi(z) \equiv 1$ the process is said to be a MA(q) process.

Formally, existence and uniqueness of a stationary solution $\{X_t\}$ of Eq. (A1) are satisfied if and only if

$$\phi(z) \neq 0 \quad \forall \, |z| = 1. \tag{A2}$$

Two other important features of a discrete time stochastic process are causality and invertibility. Causality refers to the possibility of recovering the value of the process at time t as a function of the innovations ε_s, with $s \leq t$. Formally, $\{X_t\}$ is causal if there exists a succession of absolutely summable coefficients $\{\psi_j\}$ so that the process at time t can be written as

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{A3}$$

which implies, in terms of the auto-regressive polynomial:

$$\phi(z) \neq 0 \quad \forall \, |z| \leq 1. \tag{A4}$$

Invertibility could be regarded as the property specular to causality, so $\{X_t\}$ is invertible if the series of the innovations $\{\varepsilon_t\}$ can be recovered from the process. This requires the existence of a succession of summable coefficients $\{\pi_j\}$, which allows us to write

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \tag{A5}$$

and in this case, condition (A4) is required on the moving-average polynomial:

$$\vartheta(z) \neq 0 \quad \forall \, |z| \leq 1. \tag{A6}$$

In case of the presence of d unit roots in the auto-regressive polynomial, the process becomes non-stationary; however, the d-differenced process $(1 - B)^d X_t$ can be stationary. In particular, if $(1 - B)^d X_t$ is an ARMA(p, q) process, X_t is said to be an ARIMA(p, d, q) process (where the "I" stands for *integrated*); the particular case of an AR(1) with $\phi = 1$ reduces to the well-known random walk. Taking the differences of a time series is a drastic operation and a careful testing for the presence of unit roots must be performed if this kind of non-stationarity (also called *stochastic trend*) is supposed to exist. The most used test to this purpose is the Augmented Dickey-Fuller test; notice that a unit root in the moving average can also be taken into account through the hypothesis that the innovations are an integrated autoregressive process. Extensions to more complicated models can be found in literature, but these basics ARIMA processes are sufficient for the data analysis proposed in the present work.

The main idea of time series analysis through ARMA models is to select the linear model that fits the data in the most parsimonious way, so that diagnosis of the nature of the generating process, forecasting or Monte Carlo simulations can be performed. Model selection is a non-trivial step of the procedure and should be discussed after the introduction of some fundamental tools for the investigation of the time-dependence structure of the stochastic process. This issue can be addressed through spectral analysis, decomposition of the time series in trend, cyclical, periodical, and irregular components and, most of all, correlation analysis. In this approach the dependence structure is studied analyzing the global and the partial autocorrelation functions.

The (global) auto-covariance function (ACVF) at lag h of a zero-mean stochastic process is defined as

$$\gamma(h) = E[X_{t+h} X_t] \tag{A7}$$

and, when normalized over the variance, gives the (global) ACF:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}. \tag{A8}$$

The concept of PACF is less intuitive; formally, it can be written as

$$
\begin{aligned}
\alpha(0) &= 1, \\
\alpha(h) &= \phi_{hh} \quad h = 1, 2, \ldots,
\end{aligned}
\tag{A9}
$$

where ϕ_{hh} is the last component of $\phi_h = \Gamma_h^{-1} \gamma_h$ with $\Gamma_h = [\gamma(i - j)]_{i,j=1}^h$ and $\gamma_h = [\gamma(1), \ldots, \gamma(h)]'$. In practice, the PACF quantifies the correlation between the prediction errors at lag 0 and h, given that it can be shown that the conditional expected value is the best linear predictor:

$$
\phi_{hh} = CORR[X_h - P(X_h | X_1, \ldots, X_{h-1}), X_0 - P(X_0 | X_1, \ldots, X_{h-1})].
$$

For very simple processes, the ACF and the PACF give strong hints about the time dependence structure. In particular, for a MA(q) process, the theoretical ACF is characterized by q non-zero peaks, while the PACF decays exponentially or as a damped trigonometric function; for AR(p) processes specular considerations are valid. On the (+), if the process is characterized by both autoregressive and moving-average polynomials, few information can be obtained by a simple by-eye evaluation of the ACF and PACF. In this case, statistical information criteria must be used; one of the most widely known is the *BIC*, which will be presented after the introduction of the concept of estimation for ARMA(p, q) models.

Given a parametric hypothesis ARMA(p, q) for a time series, the corresponding discrete-time equation is fitted to the data and all the parameters are estimated with maximum likelihood techniques, well described also in the more practical volume.[46] At this point, two sets $\{\hat{\phi}_j\}_{j=1}^p$ and $\{\hat{\vartheta}_j\}_{j=1}^q$ of estimated parameters are available, as well as a time series of estimated residuals $\{\hat{\varepsilon}\}$ of the same length of the original time series; if the tested ARMA(p, q) fits the data, $\{\hat{\varepsilon}\}$ must be a sequence of independent random variables. Notice that, if the orders p and q are too high, the time series is over-fitted, so the analyst must be careful in choosing the most parsimonious model in terms of number of parameters. Thus, once the residual sequence has been obtained, inference must be made on the null hypothesis H_0 of uncorrelated residuals. At this point, the sample ACF $\hat{\rho}(j)$ is computed; then, one of the most used test statistics is the Ljung-Box Test:

$$
Q_{LB} = n(n+2) \sum_{n-j}^{h} \frac{\hat{\rho}(j)^2}{n-j} \sim_{H_0} \chi^2(h),
\tag{A10}
$$

where n is the length of the time series and h is a fixed number of lags at which the sample ACF is computed. If H_0 is not rejected at a given level (usually $\alpha = 0.01$ or $\alpha = 0.05$), the tested ARMA(p, q) fits the time series.

As already mentioned, in case of complex or high-order processes, ACF and PACF are not sufficient to obtain a hint on the possible order (p, q) at glance; in this case, some different hypothetical ARMA(p, q) models can be fitted and for each one the *BIC* is computed:

$$
\begin{aligned}
BIC = {} & (n - p - q) \ln\left[\frac{n\hat{\sigma}^2}{n - p - q}\right] + n(1 + \ln\sqrt{2\pi}) + \\
& + (p + q) \ln\left[\frac{\left(\sum_{t=1}^{n} X_t^2 - n\hat{\sigma}^2\right)}{p + q}\right].
\end{aligned}
\tag{A11}
$$

This quantity is minimized by the most parsimonious model providing a good fit to the time series. However, an other possibility is to fit many different ARMA(p, q) models and choose the one for which the null hypothesis of uncorrelated residuals is not rejected and the total number $(p + q)$ of parameters is minimum.

[1] A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," Dokl. Akad. Nauk SSSR **30**, 299–303 (1941).

[2] S. Kato, H. Ito, T. Ichikawa, and M. Kagami, "Laser doppler velocimeter," U.S. patent 5,587,785 (24 December 1996).

[3] R. J. Adrian and J. Westerweel, *Particle Image Velocimetry* (Cambridge University Press, 2010), Vol. 30.

[4] R. J. Goldstein, *Fluid Mechanics Measurements* (Taylor & Francis, 1996).

[5] F. Ravelet, A. Chiffaudel, and F. Daviaud, "Supercritical transition to turbulence in an inertially driven von Kármán closed flow," J. Fluid Mech. **601**, 339–364 (2008).

[6] P.-P. Cortet, P. Diribarne, R. Monchaux, A. Chiffaudel, F. Daviaud, and B. Dubrulle, "Normalized kinetic energy as a hydrodynamical global quantity for inhomogeneous anisotropic turbulence," Phys. Fluids **21**(2), 025104 (2009).

[7] Y. Zhou, "Interacting scales and energy transfer in isotropic turbulence," Phys. Fluids A: Fluid Dyn. **5**(10), 2511–2524 (1993).

[8] A. Obukhov, "On the distribution of energy in the spectrum of turbulent flow," Dokl. Akad. Nauk SSSR **32**, 22–24 (1941).

[9] D. J. Thomson, "Criteria for the selection of stochastic models of particle trajectories in turbulent flows," J. Fluid Mech. **180**, 529–556 (1987).

[10] U. Frisch, *Turbulence* (Cambridge University Press, Cambridge, UK, 1996), Vol. 1, p. 310.

[11] J.-P. Laval, B. Dubrulle, and J. McWilliams, "Langevin models of turbulence: Renormalization group, distant interaction algorithms or rapid distortion theory?," Phys. Fluids **15**, 1327 (2003).

[12] R. Friedrich, "Statistics of lagrangian velocities in turbulent flows," Phys. Rev. Lett. **90**(8), 084501 (2003).

[13] A. Baule and R. Friedrich, "Joint probability distributions for a class of non-Markovian processes," Phys. Rev. E **71**(2), 026101 (2005).

[14] A. Aringazin and M. Mazhitov, "Stochastic models of lagrangian acceleration of fluid particle in developed turbulence," Int. J. Mod. Phys. B **18**(23n24), 3095–3168 (2004).

[15] A. Aringazin and M. Mazhitov, "One-dimensional Langevin models of fluid particle acceleration in developed turbulence," Phys. Rev. E **69**(2), 026305 (2004).

[16] P. H. Franses and D. Van Dijk, *Non-linear Time Series Models in Empirical Finance* (Cambridge University Press, 2000).

[17] J. Guiot, "ARMA techniques for modelling tree-ring response to climate and for reconstructing variations of paleoclimates," Ecol. Modell. **33**(2), 149–171 (1986).

[18] L. Marié and F. Daviaud, "Experimental measurement of the scale-by-scale momentum transport budget in a turbulent shear flow," Phys. Fluids **16**, 457 (2004).

[19] P.-P. Cortet, A. Chiffaudel, F. Daviaud, and B. Dubrulle, "Experimental evidence of a phase transition in a closed turbulent flow," Phys. Rev. Lett. **105**(21), 214501 (2010).

[20] A. Naso, R. Monchaux, P.-H. Chavanis, and B. Dubrulle, "Statistical mechanics of Beltrami flows in axisymmetric geometry: Theory reexamined," Phys. Rev. E **81**(6), 066318 (2010).

[21] B. Saint-Michel, B. Dubrulle, L. Marié, F. Ravelet, F. Daviaud *et al.*, "Evidence for forcing-dependent steady states in a turbulent swirling flow," Phys. Rev. Lett. **111**, 234502 (2013).

[22] N. Leprovost, L. Marié, and B. Dubrulle, "A stochastic model of torques in von kármán swirling flow," Eur. Phys. J. B **39**(1), 121–129 (2004).

[23] J. H. Titon and O. Cadot, "The statistics of power injected in a closed turbulent flow: Constant torque forcing versus constant velocity forcing," Phys. Fluids **15**(3), 625–640 (2003).

[24] H. Tennekes, "Similarity relations, scaling laws and spectral dynamics," in *Atmospheric Turbulence and Air Pollution Modeling*, edited by F. T. M. Nieuwstadt and H. van Dop (Reidel, 1982), pp. 37–68.

[25] J.-F. Pinton and R. Labbé, "Correction to the taylor hypothesis in swirling flows," J. Phys. II **4**(9), 1461–1468 (1994).

[26] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. (Springer, 1990).

[27] G. E. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-Day, 1970).

[28] G. Casella and R. L. Berger, *Statistical Inference* (Duxbury Press, Belmont, CA, 1990), Vol. 70.

[29] B. Saint-Michel, F. Daviaud, and B. Dubrulle, "A zero-mode mechanism for spontaneous symmetry breaking in a turbulent von Kármán flow," New J. Phys. **16**(1), 013055 (2014).

[30] P.-P. Cortet, E. Herbert, A. Chiffaudel, F. Daviaud, B. Dubrulle, and V. Padilla, "Susceptibility divergence, phase transition and multistability of a highly turbulent closed flow," J. Stat. Mech.: Theory Exp., P07012 (2011).

[31] G. I. Taylor, "Statistical theory of turbulence," Proc. R. Soc. London, Ser. A **151**(873), 421–444 (1935).

[32] F. Ravelet, L. Marié, A. Chiffaudel, and F. Daviaud, "Multistability and memory effect in a highly turbulent flow: Experimental evidence for a global bifurcation," Phys. Rev. Lett. **93**(16), 164501 (2004).

[33] D. Faranda, D. B. Pons, F. M. Emanuele, and F. Daviaud, "Probing turbulence intermittency via auto-regressive moving-average model," preprint arXiv:1407.5478 (2014).

[34] V. Nikora, D. Goring, and R. Camussi, "Intermittency and interrelationships between turbulence scaling exponents: Phase-randomization tests," Phys. Fluids **13**(5), 1404–1414 (2001).

[35] D. Prichard and J. Theiler, "Generating surrogate data for time series with several simultaneously measured variables," Phys. Rev. Lett. **73**(7), 951 (1994).

[36] L. Biferale, I. Daumont, A. Lanotte, and F. Toschi, "Anomalous and dimensional scaling in anisotropic turbulence," Phys. Rev. E **66**(5), 056306 (2002).

[37] A. Celani and M. Vergassola, "Statistical geometry in scalar turbulence," Phys. Rev. Lett. **86**(3), 424 (2001).

[38] J. Beran, *Statistics for Long-Memory Processes* (CRC Press, 1994), Vol. 61.

[39] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," SIAM Rev. **10**(4), 422–437 (1968).

[40] R. Weron, "Estimating long-range dependence: finite sample properties and confidence intervals," Phys. A (Amsterdam, Neth.) **312**(1), 285–299 (2002).

[41] A. Arneodo, C. Baudet, F. Belin, R. Benzi, B. Castaing, B. Chabaud, R. Chavarria, S. Ciliberto, R. Camussi, F. Chilla *et al.*, "Structure functions in turbulence, in various flow configurations, at Reynolds number between 30 and 5000, using extended self-similarity," EPL (Europhys. Lett.) **34**(6), 411 (1996).

[42] J. Laval, B. Dubrulle, and S. Nazarenko, "Nonlocality and intermittency in three-dimensional turbulence," Phys. Fluids **13**(7), 1995–2012 (2001).

[43] B. Saint-Michel, E. Herbert, J. Salort, C. Baudet, M. B. Mardion, P. Bonnay, M. Bourgoin, B. Castaing, L. Chevillard *et al.*, "Probing quantum and classical turbulence analogy through global bifurcations in a von Kármán liquid Helium experiment," preprint arXiv:1401.7117 (2014).

[44] C. Connaughton, A. C. Newell, and Y. Pomeau, "Non-stationary spectra of local wave turbulence," Phys. D **184**(1), 64–85 (2003).

[45] S. Nazarenko, *Wave Turbulence* (Springer, 2011), Vol. 825.

[46] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed. (Springer, 2002).