Alma Mater Studiorum · Università di Bologna

DOTTORATO DI RICERCA IN
SCIENZE STATISTICHE

Ciclo XXVIII

Settore Concorsuale di afferenza: 13/D1

Settore Scientifico disciplinare: SECS-S01

STATISTICAL ANALYSIS OF A CLOSE VON KÁRMÁN FLOW

Presentata da: Flavio Maria Emanuele Pons

Coordinatore dottorato:
Prof.ssa Alessandra Luati

Relatore:
Prof.ssa Alessandra Luati

Esame finale: anno 2016, sessione I

English abstract

This thesis addresses the statistical modeling of turbulence, focusing on three main aspects: the critical transition from laminarity to turbulence, the effects of the so-called intermittency and the energy dynamics of a turbulent flow.

The central part of the thesis consists of six papers. The first four papers (Papers A, B, C, D) have been already published. They have been written in collaboration with the research group in Physics of the Condensed Matter of the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) in Saclay, Paris and constitute the first part of the thesis. Papers E, F were developed during a six-months visiting period at the Department of Mathematics of the University of Aarhus (Denmark) under the supervision of Ole Barndorff-Nielsen and Jürgen Schmiegel, and constitute the second part. A detailed description of the contents of each paper is reported in Section 1.3; here in the following we shall provide an overview of the whole work.

In Part I we develop two new indices $(\mathcal{O}, \mathcal{R})$ to quantify the proximity to critical transitions in stochastic dynamical systems, with particular attention to the transition from laminarity to turbulence in fluids (Paper A). The two indices are tested on two toy models and then applied to the detection of the turbulent transition in a magnetised fluid. Motivated by the fact that similar stylised features are observed in turbulent and financial time series, we employ the indices $\mathcal{O}, \mathcal{R}$ to detect critical events in financial markets.

We define a third index Υ, which quantifies the effects of intermittency and does not require very long time series. This index turns out to be effective in recovering the structure of the turbulent flow (Papers B, C). In Paper D we show that Υ is also sensitive to the turbulent behavior of financial markets, providing a possible early warning indicator of the proximity to critical events.

In Part II we introduce a new local observable as the arrival times of tracer particles at a particular point in the fluid as a proxy of the turbulent velocity field (Paper E). We model the universal self-organising structure of this observable in an effective and parsimonious way, through the Normal Inverse Gaussian (NIG) distribution and the introduction of a Stochastic Equivalence Class (SEC).

In the second paper of Part II (Paper F), we address the continuous-time

dynamics of the energy budget of the turbulent field analised in Paper E. We consider a global observable, the turbulence intensity, in the framework of *ambit stochastics*, which is suitable for modeling spatio-temporal stochastic processes featuring intermittency. We show that this observable can be characterised as the exponential of a stochastic integral on a Lévy basis, under the assumption that the energy transmission across time scales is a multiplicative cascade process. The model is estimated based on the NIG fit to the marginal distribution of the time series and on second-order multiplicative correlations and then validated using third-order correlations.

For copyright reasons, Paper A and Paper D can be displayed in their final form, but not with the journal format. For the print version of the two papers, see [20] and [19].

Riassunto

La presente tesi è incentrata sulla modellazione statistica della turbolenza, con particolare attenzione a tre aspetti: la transizione critica da flusso laminare a turbolento, gli effetti dell'*intermittenza* e la dinamica del bilancio energetico di un flusso turbolento. La parte centrale della tesi è costituita da sei articoli. I primi quattro articoli (Paper A, B, C, D), tutti giÃ pubblicati, sono stati prodotti in collaborazione con il gruppo di ricerca in Fisica della Materia Condensata del Commissariat á l'Énergie Atomique et aux Énergies Alternatives (CEA) situato a Saclay (Parigi), e costituiscono la prima parte della tesi. I Paper E, F sono stati sviluppati durante un periodo, della durata di sei mesi, trascorso come visiting Ph.D. student presso il Dipartimento di Matematica dell'Università di Aarhus (Danimarca), sotto la supervisione dei Proff. Ole Bardnorff-Nielsen e Jürgen Schmiegel, e costituiscono la seconda parte. In Sez. 1.3 si trova una descrizione dettagliata dei contenuti di ciascun articolo; qui riportiamo invece una visione di insieme dell'intera tesi.

Nella Parte I vengono sviluppati due nuovi indici $(\mathcal{O}, \mathcal{R})$ per quantificare la prossimità di un sistema dinamico ad una transizione critica, con particolare attenzione alla transizione da laminarità a turbolenza nei fluidi (Paper A). I due indici sono testati su due modelli ideali e, in seguito, utilizzati per la rilevazione della transizione turbolenta in un fluido magnetizzato. Sfruttando il fatto che le serie storiche turbolente e finanziarie presentano un insieme di fatti stilizzati comuni ad entrambe, gli indici $\mathcal{O}, \mathcal{R}$ sono impiegati per rilevare eventi critici in serie storiche di indici finanziari.

Viene definito un terzo indice, Υ, avente come caratteristica principale quella di riuscire a cogliere gli effetti dell'intermittenza senza richiedere il campionamento di serie storiche molto lunghe, spesso difficili da ottenere, riuscendo così a ricostruire in maniera efficace la struttura del flusso turbolento (Paper B, C). Nel Paper D si mostra che questo indice è sensibile anche a comportamenti turbolenti dei mercati riflessi da indici finanziari, costituendo quindi un possibile indicatore di prossimità ad eventi critici.

Nella Parte II viene introdotto un nuovo osservabile locale come proxy del campo di velocità turbolenta: tale osservabile è costituito dalla sequenza dei tempi di misura di uno strumento sensibile alla distribuzione di particelle di tracciante passivo rilasciate nel fluido, legata a sua volta al campo di velocità (Paper E). Si mostra che questo osservabile presenta una struttura di auto-organizzazione che ammette una modellazione parsimoniosa

ed universale, mediante la distribuzione Normal Inverse Gaussian (NIG) e l'introduzione di una Classe di Equivalenza Stocastica (SEC).

Nel secondo articolo di questa parte (Paper F) viene caratterizzata la dinamica a tempo continuo del bilancio energetico del campo turbolento analizzato nel Paper E. In questo caso si considera un osservabile globale, ovvero l'intensità della turbolenza, nel contesto modellistico dell' *ambit stochastics*, particolarmente adeguato per descrivere processi stocastici spazio-temporali caratterizzati da intermittenza. Viene dimostrato che, assumendo che la propagazione di energia attraverso le diverse scale sia un processo a cascata moltiplicativa, questo osservabile può essere caratterizzato come l'esponenziale di un integrale stocastico di una base di Lévy. Il modello è stimato a partire dalla distribuzione marginale della serie storica e dalla struttura di correlazione a due tempi, e in seguito validato tramite la correlazione a tre tempi.

Per ragioni di copyright, Paper A e Paper D possono essere allegati nella loro versione definitiva, ma non nel formato della rivista. Per la versione di stampa, si vedano [20] e [19].

Contents

Chapter 1

Introduction

1.1 Overview on turbulence

The aim of this section is to provide a synthetic introduction to the phenomenology and to the basic theoretical concepts about turbulence. In particular, we describe the main physical facts about the phenomenon without highly technical elements, while we highlight which of these facts have a statistical counterpart. For more detail, see, for example, the interesting and pedagogic review contained in [21].

An apocriphal story attributes to the German physicist Werner Heisenberg this sentence:

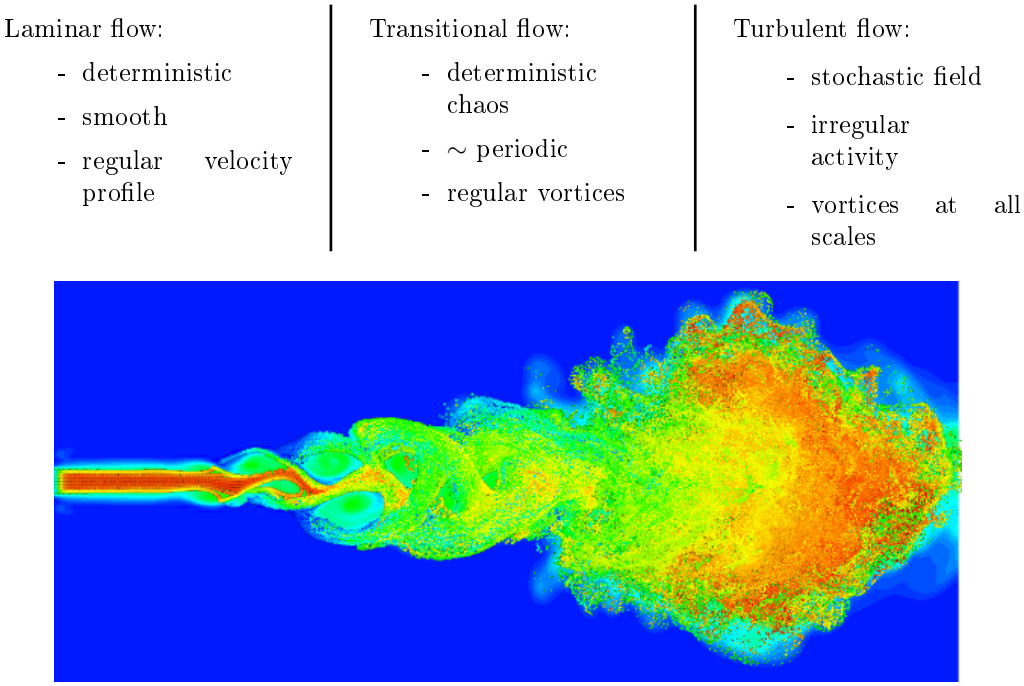
"*When I meet God, I am going to ask him two questions: Why relativity? And why turbulence? I really believe he will have an answer for the first.*".

A very similar quote is also attributed to the English matematician Horace Lamb, involving a comparison with Quantum Electrodynamics. Although we cannot be sure that these were the real words of the two scientists, these anecdotes have become famous among the researchers studying turbulence, as a demonstration of how complex and elusive this phenomenon still appears nowadays.

Turbulence is a physical phenomenon which can arise in the motion of a fluid, but also of a plasma or of a superfluid, even though with different features. Turbulence in fluids was the first to be observed and studied, and is the object of investigation of this thesis. On the other hand, the motion of a fluid is not necessarily turbulent and, all the more so, we can

identify three types of flows: laminar, transitional and turbulent. In order to provide a graphical example, the visualisation of a turbulent jet is displayed in Fig. 1.1. Starting from the left of the image, we can observe the three flow regimes. The transition from one regime to another is controlled by a combination of the flow parameters, known as the Reynolds number, which will be introduced later in this section.

Figure 1.1: Transition from laminar to turbulent flow.



The laminar flow takes its name from the fact that the moving fluid can be thought of as a pile of infinitesimal sheets (or laminae): the velocity profile is smooth and deterministic, while the mixing of matter or physical properties (e.g. the momentum) is negligible. A practical example of a strongly laminar flow is a layer of honey sliding on a slightly inclined plane: it is clear that its motion will be slow and regular, with no vortices inside the fluid. If the Reynolds number is too high to sustain a laminar motion, a transitional motion arises: this is characterised by deterministic chaos, so that quasi-periodic vortices or travelling waves can appear. If also the chaotic motion cannot be supported anymore, the flow becomes further dis-

organised: the velocity field becomes a stochastic field, in which no periodic or regular structures can be recognised. The activity in time of this field appears very irregular, even though a degree of organisation is preserved through the developement of continuously dissipating vortices. These introduce locally coherent structures at different time and spatial scales, thus generating a complex spatio-temporal correlation structure.

Figure 1.2: Transition from laminar to turbulent flow in the smoke of incense bars.

A simple, yet impressing real-world image of the transitions can be obtained observing the smoke from a cigarette or from an incense bar, as the one shown in Fig. 1.2. However, turbulence is an ubiquitous phenomenon which, given the right conditions, affects the motion of any fluid at very different spatial scales, sometimes producing noticeable effects, as shown in Fig. 1.3, 1.4: for example, the helices of an airplane produce a sequence of vortices known as 'wake turbulence', possibly affecting the stability of following aircrafts.

Another example is the wake turbulence developing in wind farms, where it can decrease the efficiency of the wind turbines located downstream to the wake-generating ones (see, for example, [22]). For these reasons, besides the pure research objective, a deeper understanding of the phenomenon or an effective way to model it in detail are required and continuously searched,

Figure 1.3: Wake turbulence from an airplane ([28]).

so that the effects of turbulence on the stability or efficiency of mechanical systems can be quantified and taken into account.

Figure 1.4: Wake turbulence in a wind farm in Denmark.

The Reynolds number

As already stated, the transition from laminar to turbulent motion is controlled by a non-dimensional parameter of the flow, introduced by George Stokes in 1851, but named after Osborne Reynolds, who made it popular in his paper [31] published in 1883. In his work, Reynolds describes a simple experiment in which the transition to turbulence is observed injecting ink into a container filled with water (as in Fig. 1.5), modulating the diameter of the pipe and the velocity of injection.

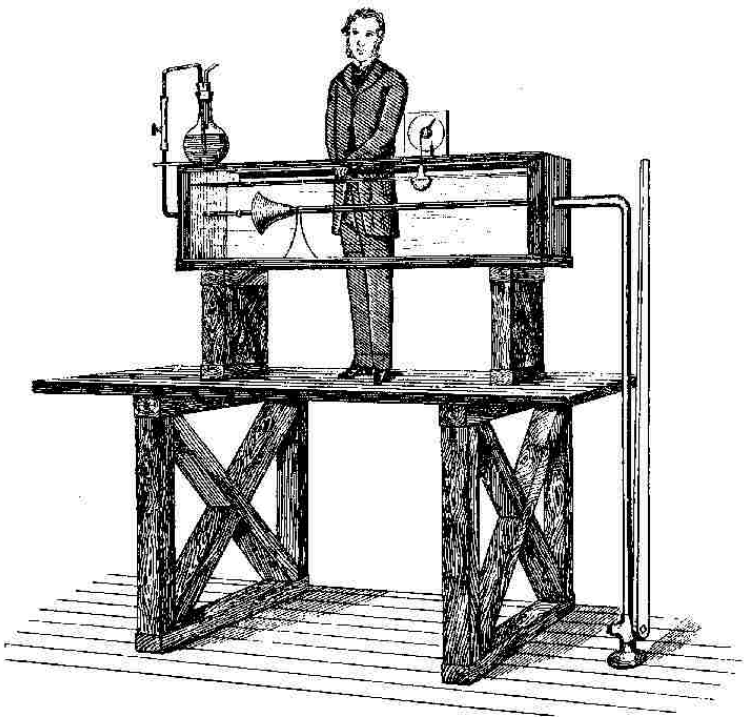


Figure 1.5: Reynolds' experimental apparatus.

By means of this experiment, Reynolds was able to compute the critical thresholds at which the transitions to chaos and to turbulence take place. In general, the Reynolds number is obtained as the ratio of the inertial forces to the viscous forces acting in the fluid and its expression depends on the geometry of the domain. For a pipe with diameter D and a mean flow

velocity V, for a fluid with kinematic viscosity ν, the Reynolds number is

$$Re = \frac{VD}{\nu}.$$

The observed typical values of the parameter for a pipe flow are:

$$\begin{aligned} & Re && < 2300 : \text{laminar flow} \\ 2300 < \; & Re && < 4000 : \text{transitional flow} \\ & Re && > 4000 : \text{turbulent flow.} \end{aligned}$$

The so-called *fully developed turbulence* is obtained when $Re \to \infty$ and can be observed in the atmosphere or in the ocean. Any artificial experimental setup has a necessarily finite Re, and one of the crucial problems when reproducing turbulence in a laboratory experiment is to reach values of the Reynolds number high enough to reproduce fully developed turbulence in good approximation.

The Navier-Stokes equation

The main observable of the motion of a fluid is the fluid velocity field $\mathbf{v}(\mathbf{x}, t)$. The evolution in time of the velocity vector is governed by the equation introduced by Navier in 1839 in [26]. The Navier-Stokes equation for an incompressible fluid is

$$\begin{aligned} \partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} &= -\nabla p + \nu \nabla^2 \mathbf{v} \\ \nabla \cdot \mathbf{v} &= 0. \end{aligned}$$

The properties and even the existence of solutions of this equation are still unknown and the latter can be solved only numerically, through the so-called Direct Numerical Simulation (DNS), which is extremely demanding from a computational point of view, so that it is difficult to obtain large samples of the observables.

The second term in the l.h.s. is the advective term, nonlinear in the velocity field, responsible for the energy transfer in the fluid. The second term on the r.h.s. is the dissipative term, proportional to the kinematic viscosity of the fluid, responsible for the energy dissipation. When the viscosity of the fluid is very large, or the other parameters defining the Reynolds number are small enough, the dissipative term dominates the

equation and the nonlinear effects are negligible. On the other hand, when the mean velocity of the fluid is large enough, the equation is dominated by the nonlinear term, which is responsible for the critical transition to a less organised motion. In view of this, turbulence could be seen as a mechanism to dissipate the energy injected in the system when the friction due to the kinematic viscosity of the fluid molecules is not sufficient to the purpose.

Measuring turbulence: Eulerian and Lagrangian approach

The description of the velocity field of a continuum, thus including fluids, can be given in two different frameworks, named Eulerian and Lagrangian approach. Following the Eulerian approach, the observer is located at fixed points in space and records the time evolution of the field in each point: if required, also the spatial structure of the field can be recovered. In this way, the total acceleration acting on a fluid particle is given by the local variation of the velocity field

$$\frac{d\mathbf{v}(\mathbf{x},t)}{dt} = \frac{\partial \mathbf{v}(\mathbf{x},t)}{\partial t}.$$

On the other hand, in the Lagrangian approach the observer follows a single fluid particle along its trajectory. In this way, he samples the shocks coming from the surrounding of the particle in different points at different times. Thus, the total acceleration requires the inclusion of ad advective term

$$\frac{d\mathbf{v}(\mathbf{x},t)}{dt} = \frac{\partial \mathbf{v}(\mathbf{x},t)}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}.$$

In practice, Lagrangian sampling is more complicated to obtain than the Eulerian sampling, but it can be obtained through DNS and in some cases it may be necessary or useful to go back and forth between the two frameworks. In this sense, a bridge between Eulerian and Lagrangian approach is given by the Taylor's 'frozen turbulence hypothesis', discussed in [36]:

"*If the velocity of the air stream which carries the eddies is very much greater than the turbulence velocity, one may assume that the sequence of changes in v at a fixed point are simply due to the passage of an unchanging pattern of turbulent motion over the point.*"

In few words, if the mean flow is fast enough with respect to the typical magnitude of the fluctuations, the following transformation is possible and often useful to pass from space to time domain and vice-versa

$$v(t) = v\left(\frac{x}{V}\right),$$

where v is the component of $\mathbf{v}$ along the mean flow direction, $V = E[v]$ and x is measured at $t = 0$ upstream to the point at which v is measured.

Empirical facts and basic theory

The nonlinear effects that appear when the advective term is dominant give raise to a great variety of empirical facts. Since their quantitative observation has been made easier in recent years thanks to more advanced experimental and computational devices, thus requiring appropriate statistical analysis. Moreover, once the flow experiences the transition to turbulence, the velocity field becomes stochastic, so that a statistical approach is the most natural way to model the phenomenon, as will be pointed out later in this section.

From a physical point of view, one of the main features of turbulence, as already anticipated, is its capability to dissipate the kinetic energy injected in the system. The anomalous dissipation is linked to the irregular activity of the turbulent velocity field, which results in vorticity fluctuations and in vortex stretching. This leads to the continuous formation and dissipation of small vortices (or *eddies*) shaped as filaments, characterised by extreme values of the intensity of the velocity vector and responsible for the brief and local peaks of kinetic energy dissipation.

Also the diffusivity of a turbulent fluid is anomalous. It is well known that the diffusion of a passive tracer in a still fluid follows the Fourier's heat law. The variance of the trajectories of an ensemble of passive particles released in the fluid at time t_0 is $V[\vec{x}] \sim t^2$ (ballistic regime) at small intervals from t_0, while it reaches the so-called diffusive limit $V[\vec{x}] \sim t$ for time intervals large enough. It is observed that the variance of such an ensemble released in a turbulent fluid shows anomalous scaling, resulting in a hyper-diffusivity.

Spectra

One of the most important empirically observed facts is the typical shape of the spectrum of the turbulent velocity field. The main feature displayed by turbulent spectra is an interval of wave numbers, named *inertial (sub)range*, such that the spectral density has a constant negative slope, approximatively equal to -5/3. This feature appears when the Reynolds number is large enough, leading to a neat separation between the large scales and the small, molecular scales.

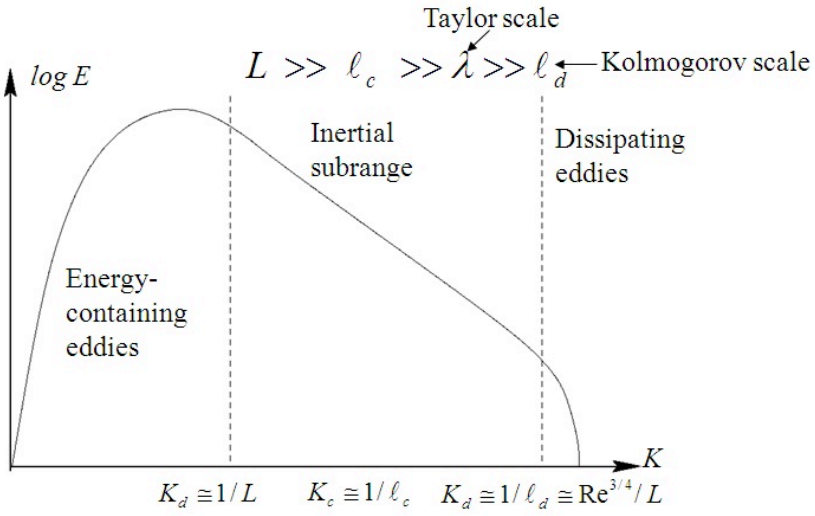


Figure 1.6: Energy transmission mechanism in turbulence.

An example of the typical aspect of a turbulent spectrum in space, as a function of the wave number, is shown if Fig. 1.6: this shape has been associated with a direct cascade of energy, which is a natural hypothesis in systems described by nonlinear equations and was first introduced in turbulence theory by Richardson in [32]. At small wave numbers we observe the large structures which arise due to the injection of energy in the fluid: these vortices progressively break into smaller eddies, thus losing some of their kinetic energy and transferring it to a smaller scale. This process continues across the whole inertial range (which, at relatively low Re can be very thin or completely missing) until the so-called Kolmogorov scale, or viscous scale

η is reached: at this point, the kinetic energy propagated from the large scales through the inertial range is finally dissipated by the intermittent eddies. Since turbulence is a continuous phenomenon, the smallest active scales are usually much larger than the smallest possible scale, that is the molecular one: in order to detect the flow features associated to anomalous dissipation, the sampling frequency should be larger than the one delimiting the Kolmogorov scale. It is worth to mention that, due to Taylor's hypothesis, we can observe an analogous behavior in the spectra computed from time series data.

Time series

Investigation in the time domain, rather than in the frequency domain, makes the stochastic nature of turbulence more evident and lets us notice some statistical features of the flow. Suppose we deal with a single realisation of an Eulerian measurement of the main component of the velocity vector, that is a univariate time series $v(t)$, sampled at a frequency falling in the dissipative range in a stationary turbulent flow. Suppose also that the time series is long enough to build sequences of increments $[v(t+s) - v(t)]$ such that the lag s ranges from the dissipative to the largest scale.

The inspection of the velocity time series will reveal the presence of clusters of high-frequency activity, which is a fingerprint of the peaks of anomalous dissipation, due to the passage of the small eddies. This phenomenon takes the name of **intermittency**, in reference to the intermittent activity of the eddies, even though it produces a great variety of effects also from the statistical point of view. First of all, these activity clusters introduce local changes in the variance of $v(t)$, while their irregular nature makes such changes stochastic: this means that the data generating process is modulated by stochastic volatility, which translates in conditional heteroskedasticity in the sample time series.

Moreover, both the velocity and the increment series at different lags display long-range dependence, which can be quantified through the Hurst exponent, often taking values different from $H = 1/2$, corresponding to the case of short memory. These two facts suggest that, in order to obtain a stochastic model for the Eulerian time series, we need to rule out both linear and Markovian or semi-Markovian processes. In addition, it is observed that, in the inertial range, the moments of the increments, also referred to as *structure functions* in turbulence, display a *multifractal and universal*

scaling:

$$S_n(s) = \mathrm{E}\{[v(t+s) - v(t)]^n\} \propto |t|^{\zeta(n)}.$$

Another evident property of turbulent velocity time series and of their increments is the non-Gaussianity. The increment process displays asymmetric distributions with heavy tails, linked to the excess of extreme events due to the intermittency. Both the skewness and the kurtosis change with the lag, but the densities tend to converge to Gaussian-like shapes as $s \to \infty$.

It is clear, observing this plurality of stylised features, that also the statistical modeling of the phenomenon, besides the deterministic physical solution of the problem, can be very challengy.

This empirical features makes clear not only that the deterministic knowledge about turbulence is strongly limited by the nonlinearity of the Navier-Stokes equation, but also that even the statistical modeling of the phenomenon can be extremely challengy.

Kolmogorov theory

The acknowledgment that a statistical approach is essential in the study of turbulence is largely due to Kolmogorov, who published in [24] his celebrated theory based on two similarity hypotheses, often denoted 'K41'. Kolmogorov starts from the assumption of local homogeneity and local isotropy of turbulence, considering a flow with high Re. Moreover, he obtains an explicit expression for the average kinetic energy dissipation $\overline{\varepsilon}$. The main contribution of that paper is the statement of the two hypotheses of similarity of turbulence:

H1 : for locally isotropic turbulence the distribution function of the velocity depends only on $\overline{\varepsilon}$ and the kinematic viscosity ν.

H2 : if the velocity field is sampled with spatial increments $l > \eta$, the distribution function depends only on $\overline{\varepsilon}$ and not on the viscosity.

Here η is the already mentioned Kolmogorov scale, that is the spatial scale of the smallest pulsations of the flow, below which the energy is directly dissipated by viscosity.

Kolmogorov also predicts a universal scaling relation for the variogram, i.e. the structure function of order 2, and thus for the spectrum:

$$\begin{aligned} S_2(l) &= C_2 \varepsilon^{\frac{2}{3}} l^{\frac{2}{3}} \\ \mathcal{E}(k) &= C \varepsilon^{\frac{2}{3}} k^{-\frac{5}{3}}. \end{aligned}$$

Under the assumptions and the hypotheses considered by Kolmogorov, the scaling relation for the variogram can be extended to structure functions of any order, i.e. $S_n(l) = C_n \varepsilon^{\frac{n}{3}} l^{\frac{n}{3}}$. All the constants C_n are universal. In 1962 Kolmogorov publishes a refined version of his original theory, denoted 'K62' and presented in [25], to include further results by Obukhov (see [27]) and to keep into account some observations received by Landau. We do not discuss the refinement of K41 here, but it is worth to mention the new assumption that $\ln \varepsilon \sim N(\mu, \sigma^2)$, where ε is the instantaneous energy dissipation.

In observed turbulence, the scaling relations found in K41 are approximatively true in the inertial range, as we can easily observe from data. In general, it is well known that the predictions of K41 and K62 on the flow statistics can never be observed in real data: Gaussian velocity increments, Markovian evolution of the Lagrangian velocity, universal scaling in spectra and structure functions, lognormal kinetic energy dissipation cannot cover the complex variety of features displayed by turbulent velocity time series. All these aspects show important deviations from Kolmogorov's predictions in real data and it is widely recognised that intermittency is responsible for such deviations, so that an actual challenge is to model the intermittent pulsation of the eddies and the consequent statistical features of the velocity field.

1.2 Data

All the papers contain and are often motivated by real data analysis. We sketch here the experimental setting and the structure of the resulting datasets.

The data, provided by the CEA, are realizations of the von Kármán Experiment (VKE), in which a turbulent flow is produced in a closed cylindric vessel. The turbulent velocity vector is sampled in time on a plane grid passing for the vertical axis of the cylindric vessel.

The VKE produces a so-called von Kármán flow, by means of two counter-rotating co-axial impellers. This configuration allows us to reach very high Reynolds numbers ($Re \sim 10^6$) using a compact experimental setup. The resulting flow is stationary in mean, strongly inhomogeneous and anisotropic. This implies that sampling the velocity field in different points in space results in mean-stationary time series virtually produced in

different turbulent experiments. The main feature of the von Kármán flow is that it develops an equatorial shear layer with local properties similar to the ones obtained in much larger experiments at high Re designed to study homogeneous isotropic turbulence.

A schematic view of the instrumentation used to produce the von Kármán flow is displayed in Fig. 1.7. The dimensions are expressed as a ratio with respect to the radius of the cylindric vessel containing the fluid, which has a real dimension R=100 mm. The flow is produced by two disks, each one carrying 16 blades; the height of the blades is 20 mm and their curvature radius is $R_c = 46.25$ mm.

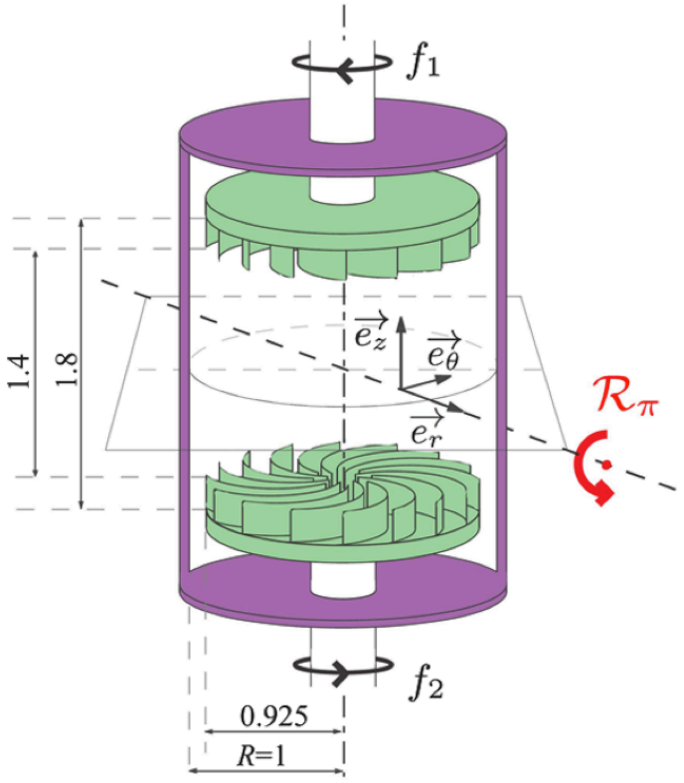


Figure 1.7: Schematic view of the VKE, Fig. from [16].

The disks are put in motion by two impellers, rotating at typical frequencies around 10 Hz. The frequencies of the two impellers, f_1 and f_2, can be

equal or different, producing slightly different flows. The two impellers have a power of 1.8 kW, thus leading to a maximum torque of 11.5 N·m. Given the structure of the vessel and the two frequencies, the Reynolds number of the experiment can be easily controlled, since it can be expressed as

$$Re = \frac{\pi(f_1 + f_2)R^2}{\nu}.$$

Since the rotation can be directed forward or backward and the blades are not straight, two different kinds of flow can be obtained and are marked by a $(-)$ or a $(+)$:

$(-)$ the advancing face of the blades is the concave one and the flow is called *anti-contrary*;

$(+)$ the advancing face is the convex one and the flow is called *contrary*.

These two configurations correspond to different forcing conditions; some modifications can also be induced in the flow geometrically, by insertion of an annulus in the equatorial region, thus increasing the magnitude of the fluctuations.

The fluid used in this experiment is a mixture of water and glycerol in different concentrations and at temperatures between 15°C and 35°C, so that the viscosity can vary, producing a range of possible Reynolds number $50 \leq Re \leq 1.2 \cdot 10^6$.

The problem is axysimmetric and symmetric towards rotations of π with respect to every radial axis passing for the centre of the vessel (R_π-symmetry).

Mean flow topology and flow regimes

The topology of the mean flow is governed by a parameter defined by the forcing condition and called *rotation number*, expressed as

$$\theta = \frac{f_1 - f_2}{f_1 + f_2}.$$

Two cases can be recognised, depending on whether the two frequencies are equal or not:

$\theta = 0$ corresponds to an exactly counter-rotating system, characterized by exact R_π-symmetry: two toric recirculation cells develop and are separated by an azimuthal shear layer. In this case the system has a $O(2)$ symmetry.

$\theta \neq 0$: the flow changes breaking the symmetric circulation. When $\theta = \theta_c$, where θ_c is a critical value of the parameter depending on the forcing and on the geometry of the experiment, a single cell forms in the vessel. When $\theta \neq 0$ the system has a $SO(2)$ symmetry.

A synthetic view of the two configurations is shown in fig. 1.8:

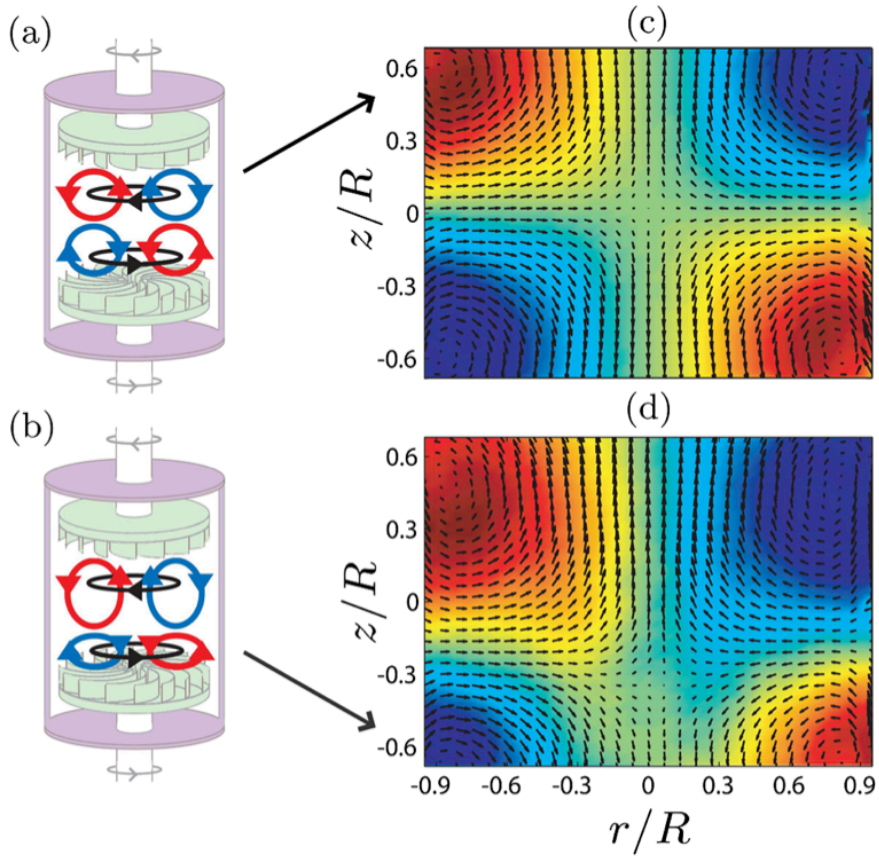


Figure 1.8: Mean flow topology for $\theta = 0$ and $\theta \neq 0$, Fig. from [16].

Depending on the chosen values for the experimental parameters, a wide range of Re can be investigated, as shown in Table 1.1.

C	μ at 15°C	μ at 30°C	ρ	Re range
99%	1700	580	1260	50 - 2 000
93%	590	210	1240	130 - 5 600
85%	140	60	1220	550 - 19 000
81%	90	41	1210	840 - 28 000
74%	43	20	1190	1 800 - 56 000
0%	1.1	0.8	1000	570 000 - 1 200 000

Table 1.1: Dynamic viscosity μ (10^{-3} Pa·s) at various temperatures, density ρ (kg·m^{-3}) at 20°C and achievable Reynolds number range for various mass concentrations C of glycerol in water.

Different ranges of Re correspond to different flow regimes, which we report here from [30]. Since some of the regimes lead to the development of waves, we need to introduce the wavenumber m:

Reynolds number	wave number and flow regime
$Re < 175$	$m = 0$, axisymmetric, R_π-symmetric steady basic flow
$175 < Re < 330$	$m = 2$, discretely R_π-symmetric steady flow
$330 < Re < 389$	$m = 2$, non R_π-symmetric, equatorial-parity-broken traveling waves
$389 < Re < 400$	modulated traveling waves
$400 < Re < 408$	chaotic modulated traveling waves
$408 < Re \lesssim 1000$	chaotic flow
$1000 < Re \lesssim 3300$	transition to turbulence
$3300 < Re \lesssim 10^4$	inertially-driven fully turbulent flow
$Re \gtrsim 10^4$	multivalued inertial turbulence regimes

Table 1.2: Main flow regimes and corresponding intervals of the Reynolds number.

In our analysis we shall consider only datasets produced by fully turbulent flows with $Re \gtrsim 10^4$.

Acquisition of the velocity data

Data about the evolution in time of the velocity field are obtained through two different techniques: Stereoscopic Particle Image Velocimetry (SPIV) and Laser Doppler Velocimetry (LDV). Both of them measure the velocity vector on a plane grid, so that for each point of the grid a time series is obtained. In the following, a summary of the two methodologies is proposed.

SPIV. The radial, axial and azimuthal components of the velocity are measured on a 95×66 points grid covering a whole meridian plane. The spatial resolution of this grid is $\Delta x = 2.08$ mm. The sampling frequency is $1 \leq f_s \leq 15$ Hz, thus falling in the inertial range, and the total length of the time series goes from 400 to 27000 records. The field is made non-dimensional by scaling it on a typical velocity:

$$v_0 = \frac{2\pi R(f_1 + f_2)}{2}$$

and then windowed in order to remove the spurious velocities on the blades and the boundaries; at this point the resulting grid is made of 58×58 points. Finally, two filters are applied: a global one, in order to remove all the velocities $v > 3v_0$ and a local nearest-neighbour one to remove spurious vectors. In this way about the 1% of the time series is lost.

LDV. Measurements of the velocity field are obtained through a He-Ne Flowlite Laser with $\lambda = 632.8$ nm. Either the axial or the azimuthal component of the velocity is directly measured; the radial component is recovered using $\nabla \cdot \mathbf{v} = 0$, which is legitimate by the fact that the time-averaged field is axisymmetric and solenoidal. The resulting data consist of the time evolution of the velocity field on a 11×17 points grid, with an average data rate of 1 kHz, providing information about the small, dissipative scales. On the other hand, the capability of the LDV device to perform a measure of the velocity depends on the presence of passive tracers (i.e. particles with negligible inertia) in the fluid. More in particular, the instrument can perform a measurement when a particle passes through one of the grid points. This fact prevents us to control at which times the velocity is recorded, thus producing irregularly spaced time series.

1.3 Conclusions

In the following we summarise the content and the main contributions of each paper that constitutes this thesis.

Part I

Paper A addresses the problem of defining effective early warning indicators for critical transitions. In the VKE framework, such indicators reveal the relation between the experimental parameters and the nature of the flow (laminar, chaotic, stochastic). Our approach assumes a prior baseline time series model for the considered observable far from the transition. In particular, we assume that, in the baseline or equilibrium state, the continuous-time dynamics of the system can be described by an Ornstein-Uhlenbeck (OU) process. The OU dynamics is Markovian in time and its sample version is an AR(1) process. We also assume that there exists a parameter of the experiment, λ, controlling the proximity of the system to the transition, which happens at a critical value $\lambda = \lambda_c$.

A set of time series of the observable, sampled in different experimental conditions (i.e. at different values of λ), is analysed by fitting the most appropriate linear ARMA(p,q) model to each time series. The model selection is performed using the Bayesian Information Criterion (BIC). Once the models are estimated, we quantify how "distant" each time series is from the baseline model. To this purpose we define two measures of divergence: the total order of the model $\mathcal{O} = (p + q)$ and the total persistence $\mathcal{R}$, given as the sum of the absolute values of all the autoregressive and moving-average parameters.

We validate our method considering two toy models and we apply it to two real-world cases. The first toy model is a classical system featuring bistability under the effect of random noise: its trajectory is obtained by simulating an OU process with a biquadratic drift term. As a second test case, we choose a classical 2D Ising model with a nearest-neighbor interaction on a square lattice. In the first case, the Markovian dynamics is preserved also when approaching the critical transition, while the total persistence steadily increases approaching $\mathcal{R} = 1$, which is the upper boundary of the admissible values for the parameter of a stationary AR(1) process. On the other hand, the analysis of the Ising dynamics shows a steep increase of both indicators near the transition.

After preliminary testing, we apply the indicators to two real systems. The first one is the von Kármń Sodium Experiment (VKS), i.e. the same configuration of the VKE, but featuring a magnetised fluid instead of water: it is possible to find a value (here known *a priori*) of the control parameter such that a *dynamo effect* is observed. This phenomenon affects the time series of the modulus of the magnetic field, which is chosen as the observable of interest. The transition clearly emerges from our analysis, around the expected value of the control parameter, with a steep divergence of both the indicators from the baseline value. On the contrary, the analysis of an analogous experiment in which the dynamo could not be produced, results in a complete absence of any indication of critical transition. In the second case, we consider the log returns of the EUR/USD exchange rate hourly datasets. This choice is motivated by the fact that financial and turbulent time series show very similar features: we investigate if the aforementioned indicators can produce a warning for future financial crises of endogenous origin. Both the indicators, but in particular $\mathcal{O}$, show three sudden increases which are followed, after a few months delay, by official warnings from the European Central Bank.

In Paper B we exploit the approach described above to obtain a detailed characterisation of the turbulence in the SPIV data of the VKE. We assume a linear OU equation as a reference model for the evolution of a component of the turbulent velocity vector. This hypothesis is consistent with the classical probabilistic theory of turbulence. We interpret the deviations from this model as an effect of the intermittency, a crucial and still not completely understood phenomenon, which arises when the turbulence is well developed. In order to measure the deviation from the baseline model, we introduce a new index Υ, constructed as a normalized distance between the Bayesian Information Criterion (BIC) of the best ARMA(p,q) describing each time series and the BIC for the AR(1) model.

After a brief review of the basics of ARMA time series modeling and an introduction about the experimental setup and the sampling techniques, we show that the index Υ catches many features of the turbulent field. We also show that these results are stable with respect to phase randomization, which means that the information we obtain about intermittency through this index comes from amplitude intermittency, rather than from phase intermittency. Since high values of Υ correspond to higher-order models for the time series (i.e. ARMA(p,q) with $p > 1, q > 0$), we argue that the deviations from the reference model may be explained by anomalous memory

in the stochastic process. In view of this, we compare the performance of Υ and of the Hurst exponent H in recovering the flow structure. The results indicate the presence of a long memory component in the most active turbulent regions, corresponding to high values of Υ, and no memory or anti-persistence in the areas associated with low values of the index, with strong gradients in highly inhomogeneous regions. We conclude that some salient features of the analysis are preserved when considering global besides local observables. Finally, we analyse flow configurations displaying multistability and we find that the technique is efficient in discriminating different stability branches of the system.

Paper C contains a brief but focused discussion about the index Υ, compared to a more traditional index of intermittency, computed from the structure functions. In particular, we show that the two indices are proportional, thus providing the same information about the turbulence. On the other hand, Υ ranges over a wider interval of values and can be easily computed also from relatively short time series, while the higher-order structure functions require very large samples to be estimated. Of particular interest is the capability of our index, unexpected *a priori*, to catch the wandering shear layer in the bistable realization of the VKE considered for the application.

In Paper D we jointly exploit the good performance of the index Υ, introduced in Paper B, and the capability of the early warning indicators, introduced in Paper A, to anticipate critical financial events. We consider the daily changes in three financial indices, i.e. the Dow Jones, the Standard and Poor's and the EURO STOXX 50.

The best model for the whole time series, selected with the BIC, is an AR(1) for all the three indices. We then divide the time series in windows of 60 days and we show that also the large majority of these subsamples can be modeled by a stationary AR(1) process. We compute Υ for each window and we choose $\Upsilon = 0.3$ as a critical value of the index: we find that such high values appear to anticipate financial crises of endogenous origin. This is possibly attributed to the fact that time series of turbulent observables and of financial indices sampled in "turbulent" markets share many stylised facts, such as long memory, distribution of the increments changing with the lag and stochastic volatility. In turbulence, all these features are deviations from the classical theory and are tipically linked to the effect of intermittency, which is caught by the index Υ, as discussed in Papers B,C.

Part II

In Paper E we consider the LDV dataset from the VKE: in this case we only have $j = 1, 2, \cdots, 18$ points in space, but the time series are much longer, due to a high sampling frequency, which enables us to describe the small-scale behavior of turbulence. While it can guarantee a high average sampling frequency, the device can measure the turbulent velocity only when one of the particles released in the fluid passes through one of the measuring points: this makes the time step irregular. In order to overcome this problem, we consider the sequence of sampling times, also recorded by the device, and investigate if this can be used as a turbulent observable. In particular, the natural logarithm of the increments of the process at lag $k = 1, 2, \cdots, K$, denoted τ_j^k, can be interpreted as log-waiting times for the next k particles to be detected by the device. We consider this variable, in analogy with the more classical logarithm of the turbulent velocity increments.

As a first step, we consider the Hurst exponent and the tail heaviness of the distribution of the increments at all k. The former highlights the long memory of the increments at all lags, with different values at different lags and a peculiar behavior in each time series. Its variability among the samples shows a minimum around $k = 800$. The latter is quantified fitting a stretched exponential relation both to the left and right tails and computing the nonlinear least squares estimate of the stretching exponent. In contrast with the classical findings for the velocity increments, the tail heaviness decays with the lag starting from a near-Gaussian value and evolving into light tails.

The main part of the paper provides a parsimonious and universal description of the waiting times. By *universal* we mean that we can observe a collapse of the density of the increments at different lags of series sampled at different points in space on the same law. The parsimonious description is given by the NIG distribution, a four-parameter probability density function admitting asymmetry, non-Gaussian kurtosis and non-Gaussian tails. We show that the NIG distribution fits well τ_j^k for every j, k. We then highlight the self-organising nature of this observable defining a change-of-lag through the second and fourth order cumulants, showing that it is possible to observe a collapse of the densities on the same law.

Paper F completes the statistical analysis of the turbulence in the VKE considering a global observable. In particular, we compute the turbulence

intensity over small, constant time intervals. This observable quantifies the instantaneous energy budget of the system and it is positive definite by construction. These properties make the turbulence intensity suitable for *ambit stochastics* modeling. In this framework, we assume a multiplicative cascade process underlying the injection of energy across time scales. Then, we describe the continuous-time dynamics as the exponential of the stochastic integral of a Lévy basis, computed on the so-called *ambit set*. This model is completely specified once the ambit set and the Lévy basis are known.

The probability density function of the Lévy basis is the same as the marginal distribution of the observable, with their moments linked through the Lebesgue measure of the ambit set. We find that the marginal distribution of the time series is well described by the NIG law, as already known for the kinetic energy dissipation rate. The shape of the boundary of the ambit set is recovered from the time scaling of two-point correlators, a correlation measure introduced for multiplicative processes. Its volume gives us the relationship between the moments of the Lévy basis and the moments of the observed process. We find that the boundary of the ambit set can be described by a simple polynomial equation. We also verify the existence of a clear self-scaling property of the two-point correlators, which is a fingerprint of an underlying multiplicative cascade process. Finally, we compare the empirical and the theoretical three-point correlators, as a more stringent indication of the cascade-like nature of the energy dynamics, observing a good matching.

Summary and main contributions

In the following we summarise the main contributions contained in each part.

Part I

I. We introduce and validate two new early warning indicators $(\mathcal{O}, \mathcal{R})$ of critical transition. These indicators are based on linear time series analysis and appear to be effective for systems experiencing transitions between two states, at least when the fundamental state can be assumed to be the realisation of a linear Markovian stochastic process (Paper A).

II. We introduce a new intermittency index Υ, based on the Bayesian Information Criterion, and we employ it to characterise the turbulence in the VKE. We show that this index captures amplitude intermittency and we also show that its value is linked to a slow decay of the ACF of the time series, measured through the Hurst exponent. Moreover, we prove its effectiveness with respect to a traditional, physics-based index, overcoming the problem of computing the structure functions from short time series (Paper B, C).

III. We apply Υ to the early detection of critical financial events, finding that high values of the index appear before or at most together with financial crises of endogenous origin (Paper D).

Part II

I. We solve the issue of the irregular sampling of the velocity field by introducing a new local observable, the log-waiting times for the arrival of tracer particles at the detection point of the LDV instrument. We prove that this novel observable shares many stylised features with turbulent velocity and that it admits a parsimonious and universal description through the Normal Inverse Gaussian distribution and the definition of a Stochastic Equivalence Class. The main discrepancy with the expected behavior lays in the light-tailed distribution of the increments, which may be justified through physical arguments (Paper E)

II. We characterise the energy budget dynamics of the VKE through the realised volatility of the velocity field, which can be physically interpreted as the instantaneous kinetic energy of the flow. We model this observable by means of a multiplicative cascade process. We show that fitting Normal Inverse Gaussian to the marginal distribution of the observable and modeling the time-scaling of the two-point correlators is sufficient to specify the model, so that the process can be written as the exponential of a stochastic integral. We validate our model through the self-scaling of two-point correlators and the prediction of three-point correlators, finding a very good agreement between model and observations (Paper F).

Chapter 2

Part I

2.1 Overview on Part I

This first part of the thesis is focused on the statistical detection of critical transitions and of intermittency in turbulence. The work has involved a collaboration with the research group in Physics of the Condensed Matter at the CEA in Paris. In particular, the motivation for this part arises from the issue of defining a statistical index, based on linear time series analysis, able to detect the critical transition from laminar to turbulent flow in the VKE. We then extend our investigation to other systems and exploit our findings to extract information from datasets concerning turbulent flows.

In the following, we briefly describe the underlying framework. We consider dynamical systems that can experience a critical transition (or phase transition) from one state to another. The state of the system is controlled by a - possibly unobservable - parameter λ and the transition happens when the parameter reaches a critical value λ_c. We argue that the stochastic process that approximates a time series sampled in the considered system should be different as long as the system is in different states.

In many cases of two-state systems, one can be thought of as the basic state, the other as the complex state: the stochastic behavior in turbulence, opposed to the deterministic one in laminar flows, is an example of this fact, where the control parameter is the Reynolds number. We start considering cases in which we may assume a linear and Markovian evolution in time for the system in the basic state. We can think of such systems as realisations of an Ornstein-Uhlenbeck process, which in its discrete time version is an

AR(1) process.

Concerning the sampling of the system, we discriminate two cases. In the first case, the system is sampled in different experiments, each of them corresponding to a fixed value of λ: we then have a time series for each experiment. Otherwise, the parameter changes in time during the sampling, thus generating a single time series. Notice that the non-constant value of the control parameter does not imply any kind of non-stationarity in the time series, at least a priori. In the first case we can analyse each time series separately, while in the second we may need to isolate portions of the time series and analyse them sequentially with a moving window method. In the following, we summarise the procedure and the results contained in this part.

Detection of critical transitions

We implicitly assume an AR(1) model as our reference for systems in their basic states. We expect that, when nonlinear effects become dominant and the system experiences a transition, its more complex dynamics will be better captured by more complicated processes, for example higher order ARMA(p,q) models, selected with the Bayesian Information Criterion (BIC). We construct two indices to summarise the features of the chosen ARMA(p,q) model with parameters $(\vec{\phi}, \vec{\theta})$:

$$\begin{aligned} \mathcal{O} &= p + q; \\ \mathcal{R} &= \sum_{i=1}^{p} |\phi_i| + \sum_{j=1}^{q} |\theta_j|. \end{aligned}$$

While doing this, we do not imply that such model is well specified: on the contrary, we expect that any simple linear model will be a misspecification for the real underlying process, thus requiring an overparameterisation and then higher order models.

We test our method on two toy models: a bistable system and an Ising model. The former is generated by an equation analogous to the Ornstein-Uhlenbeck process, but featuring a biquadratic drift term, representing a double-well potential. Here, the dynamics is forced to be stationary and Markovian even after the transition: in this case we observe that, while the selected process is always an AR(1), the autoregressive parameter steadily grows to its limit value $|\phi| = 1$, above which stationarity is not preserved.

In the Ising model, on the contrary, we observe a steep increase in the value of both indices corresponding to the transition as soon as $\lambda \geq \lambda_c$.

We then apply our method to two real systems. The first is a variation of the VKE, the von Kármán Sodium Experiment (VKS), in which the vessel contains liquid sodium: when the forcing condition is strong enough, a dynamo develops in the fluid, causing strong changes in the modulus of the magnetic field, which is chosen as observable of this system. In the second case, we consider the log returns of the EUR/USD exchange rate hourly datasets. In this analysis, our indices appear effective in recovering the dynamo effect and in anticipating critical events in the financial markets.

Assessing intermittency

In view of the results described above, we argue that, in case of turbulent systems, strong deviations from the baseline Markovian linear model may be due to intermittency, like many other well known discrepancies between observations and the Kolmogorov theory of turbulence, above all non-Gaussianity. In this sense, we expect that our method should at least highlight the fluid areas with highly intermittent activity, where steady large-scale vortices form and can break into smaller ones, feeding the energy cascade. We then focus on the SPIV datasets from the VKE at high *Re*, in which the velocity field is sampled on a refined grid, with time-steps falling in the inertial range. In this way, we have an instantaneous picture of the field, or at least one of its plane sections, and also its evolution in time.

Here, we decide to explicitly use the model misspecification to define a new index Υ. Since we assume an AR(1) as the baseline model, we construct the index as a normalised difference between the BIC corresponding to the AR(1) and the BIC corresponding to the selected ARMA(p,q) model

$$\Upsilon = 1 - \exp\{|\mathrm{BIC}(p+1, q) - \mathrm{BIC}(\overline{p}, \overline{q})|\}/n,$$

where ARMA$(\overline{p}, \overline{q})$ is the reference model and n is the sample size. The larger this discrepancy, the farther the system is expected to be from the baseline state. The new index proves to be effective in recovering the flow structure, highlighting fluid areas characterised by strong activity.

Moreover, since from our first results we deduced that deviations from the baseline model may be due to an increase of the persistence of the

process, i.e. a slow decay of the autocorrelation function, we investigate if long-range dependence is present in the time series and if so, what information can provide about the flow. Long-range dependence is associated to fractional integration in the time series, while in continuous time we may think, for example, of an Ornstein-Uhlenbeck-like process, but innovated by a fractional Gaussian noise instead of a standard Wiener process. We compute the Hurst exponent for every time series, finding that in the most intermittent areas it assumes values compatible with long memory, while in areas without eddies its value is around $H = 1/2$ (short memory) or smaller (anti-persistence). Its value appears well correlated to the value of Υ, so that also the Hurst exponent appears effective in recovering the flow structure. Finally, following the suggestion of a referee, we also investigate if these findings are stable with respect to phase randomisation: the phase-randomised time series produce very similar results, leading to the idea that the index Υ captures effects due to amplitude rather than phase intermittency.

The main asset of the aforementioned method is that the effects of intermittency often requires long, high-frequency time series for the computation of the structure functions of order $n > 2$. On the other hand, our index proved to be effective in recovering intermittent activity from the SPIV dataset, composed of time series of length $N \sim 5 \cdot 10^3$. In order to check whether the features captured by Υ are due to intermittency or to other phenomena, we investigate if the index can be used to obtain the intermittency correction to the Kolmogorov theory. In order to do this, we compare Υ with an intermittency index μ computed following the ESS (Extended Self Similarity) method introduced in [34]. Since the computation of such index requires the structure functions, it can only be used if the samples are large enough, while Υ proved to be effective also on the short SPIV time series: we now use the much longer LDV series and compare the results. The intermittency index is obtained as follows. First, we estimate the relative scaling exponents ζ_p^* from $G_p(\tau) \sim \langle |\delta u_\tau|^3 \rangle^{\zeta_p^*}$; then, we compute the index as $\mu = \zeta_2^* - \frac{2}{3}\zeta_3^*$. Here τ is a time increment, $G_p(\tau)$ is the structure function of order p and $\delta u_\tau = u_{t+\tau} - u_t$, for coherence with the notation in the paper. We find that the two indices are well correlated, while Υ ranges over a wider interval of values, making it easily interpretable. We also consider a SPIV dataset sampled in a VKE configuration producing a wandering shear layer, meaning that the most intermittent part of the flow has two possible stable positions and switches continuously between them.

We can see, from the graphical inspection of the spatial distribution of Υ (Fig. 4 in [19]), the two positions captured by the index, despite the short available time series, as a further proof of its capability to detect turbulent activity.

Beyond turbulence

Considering the results obtained so far, we go beyond turbulence and we analyse financial time series, still through our index Υ. This is motivated by two reasons. First, the initially introduced indices $\mathcal{O}, \mathcal{R}$ proved to be effective in discriminating portions of financial time series preceding official warnings in crisis events. Moreover, as pointed out, for example, in [7], turbulent and financial time series share some stylised features, possibly due to intermittency.

Here we sketch our method. We consider the daily returns of three stock indices over a period of at least 20 years. The idea is that, when the financial system is approaching a crisis or a critical event (of endogenous origin), the behavior of the market changes, together with the dynamics of the time series. Since for each index we only have one time series, and the control parameter λ is, in this case, unobservable, we need to apply a moving window method. We choose a window of 60 days, after verifying that our findings are stable for windows between 40 and 80 days.

Since each window is quite short, we expect that, even though financial time series usually display complex features, a linear Markovian model may still be appropriate, at least for sub-periods of non-critical activity. On the other hand, departures from the simple reference model on such short sub-samples may be indication of strong intermittency-like effects. Here we do not assume a markovian model a priori, but we perform a BIC-based model selection on each subsample, finding that an AR(1) is the most appropriate model at least for the 75% of the subsamples of each index. From this, we decide to assume such model as a baseline, and we compute the intermittency index for each sub-series. We choose $\Upsilon = 0.3$ as threshold value: this choice is made after observation of the empirical density of the values of the index. The high values appear to precede or at most to coincide with financial crises of economic origin.

We leave for future work the problem of defining an analogous quantity for which the asymptotic distribution can be known.

Paper A

Faranda, D., Dubrulle, B. and Pons, F. M. E. (2014). Statistical early-warning indicators based on autoregressive moving-average models. *Journal of Physics A: Mathematical and Theoretical* **47(25)**, 252001.

Statistical early-warning indicators based on Auto-Regressive Moving-Average processes

Davide Faranda, Bérengère Dubrulle*
Laboratoire SPHYNX, Service de Physique de l'Etat Condensé, DSM, CEA Saclay, CNRS URA 2464, 91191 Gif-sur-Yvette, France

Flavio Maria Emanuele Pons
Department of Statistics, University of Bologna, Via delle Belle Arti 41, 40126 Bologna, Italy .

We address the problem of defining early warning indicators of critical transition. To this purpose, we fit the relevant time series through a class of linear models, known as Auto-Regressive Moving-Average (ARMA(p,q)) models. We define two indicators representing the total order and the total persistence of the process, linked, respectively, to the shape and to the characteristic decay time of the autocorrelation function of the process. We successfully test the method to detect transitions in a Langevin model and a 2D Ising model with nearest-neighbour interaction. We then apply the method to complex systems, namely for dynamo thresholds and financial crisis detection.

Many experimental or natural systems undergo *critical transitions* - sudden shifts from one to another dynamical regime. In some instances, *e.g.* global changes in climate science, species extinction in ecology, spin glasses, it is of crucial importance to build early warning indicators, i.e. estimates of the transition threshold based on finite time-series corresponding to situations where the bifurcation did not happen yet. The statistical approach to this issue traditionally involves so-called indicators of criticality [1, 2]. They are based on specific properties of ideal statistical systems (such as the Langevin or Ising model) near the transition: critical slowing down, modifications of the auto-correlation function or of the fluctuations [3], increase of variance and skewness [4], diverging susceptibility [5–7], diverging correlation length. However, it is known that, in some complex systems, these indicators fail to detect the transition: in spin glasses, no diverging correlation length has been found so far, and one has to resort to finer statistical tools (such as four point dynamical susceptibility [8, 9]) to detect transitions. In addition, traditional early warning indicators may be inapplicable in datasets containing a small number of observations (see *e.g.* [10, 11]), which is usually the case in several applications where the experiment is unique (as in financial or climate time series), difficult to repeat or to sample for a long time (as for atmospheric measurements, laboratory turbulence, etc). This suggests that indicators based on single statistical properties may not be sufficient for detecting transitions in complex systems, so that one should rather consider all the information contained in the finite-time series.

The main idea of the present letter is therefore to introduce a class of indicators of critical transitions based on a statistical model for the observed data when approaching a tipping point. To be interesting for applications, these indicators must satisfy certain properties: i) they must generalize the well-know indicators based on single statistical properties and ii) they must be built using a statistical model that is simple to implement and works for limited data set. Here, we show that these properties are satisfied for indicators based on the auto-regressive moving-average processes of order p, q ARMA(p,q), modeling a time series $X_t(\lambda)$, experiencing a transition at $\lambda = \lambda_c$. In the first part of the paper, we first recall some basics on ARMA(p,q) modeling and define corresponding early-warning indicators. We then check that these indicators are able to detect the transition in simple theoretical models, such as Langevin double-well model or Ising model. We then apply our indicators to the analysis of complex systems for the detection of turbulent dynamo thresholds and financial crisis.

Theory. Let us consider a series of observations X_t of an observable with unknown underlying dynamics, controlled by a parameter λ. We further assume that for $\lambda < \lambda_c$ the time series $X_t(\lambda)$ represents a stationary phenomenon. The critical threshold λ_c is defined through the condition that for $\lambda \geq \lambda_c$, there is a bifurcation, in the sense that there exists no smooth transformation of the physical measure through the transition. Since $X_t(\lambda)$ is stationary, we may then model it by an ARMA$(p(\lambda), q(\lambda))$ process such that for all t:

$$X_t(\lambda) = \sum_{i=1}^{p} \phi_i(\lambda) X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j(\lambda) \varepsilon_{t-j} \qquad (1)$$

with $\varepsilon_t \sim WN(0, \sigma^2)$ - where WN stands for white noise - and the polynomials $\phi(z) = 1 - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p}$ and $\theta(z) = 1 - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q}$, with $z \in \mathbb{C}$, have no common factors. Notice that, hereinafter, the noise term ε_t will be assumed to be a white noise, which is a very general condition [12]. For a general stationary time series, this model is not unique. However there are several standard procedures for selecting the model which fits at best the data. The one we exploit in this paper is the Box-Jenkis procedure [13]. We chose the lowest

p and q such that the residuals of the series filtered by the process ARMA(p,q) are delta correlated. This fixes p and q, and thus our statistical model. There are other model selection procedures based on information criteria (Bayesan or Akaike information criteria). Unfortunately, in our case none of them gives clear indications for discriminating the model to be used as it is not clear which range p and q must be tested to get reliable results. Intuitively, p and q are related to memory lag of the process, while the coefficients $\phi_i(\lambda)$ and $\theta_i(\lambda)$ represent the persistence: the higher their sum (in absolute value), the slower the system is forgetting its past history. In the sequel, we now present early warning indicators based on these parameters.

Early-warning indicators. Far from the transition, the time series of a generic physical observable can be described by an ARMA(p,q) model with a reasonably low number of p,q parameters and coefficients. On the other hand, for $\lambda \to \lambda_c$, the critical value corresponding to a transition, the statistical properties (such as the shape and/or the persistence of the autocorrelation function) of the system change, leading to different characteristics of the ARMA(p,q) model which can describe the data series or to an inadequacy of the model itself. Specifically, several changes in the dynamics occur near the transition, as the system is allowed to explore a larger portion of the phase space with higher excursions in the direction of the new stable state. First, the distributions of the observables become skewed towards the maxima or the minima, depending on the direction of the shift. Second, the system may experience a *critical slowing down* with diverging memory effects. This phenomenon is traditionally quantified by the autocorrelation function (ACF) of the time series X_t defined (assuming a zero-mean process) as:

$$\mathrm{ACF}(h) = E[X_{t+h}X_t]/E[X_t^2]. \tag{2}$$

Far from the transition, the ACF tends to be 0 after a finite number of lags $\bar{h}$. As $\lambda \to \lambda_c$, critical increase of memory of the system makes $\bar{h} \to \infty$. The ARMA$(p(\lambda), q((\lambda))$ model of the corresponding time series will then be characterized by two properties:

- $\sum_{i=1}^{p} |\phi_i|$ and $\sum_{j=1}^{1} |\theta_j|$ increase for $\lambda \to \lambda_c$ as the direct consequence of $\bar{h} \to \infty$.
- $p+q$ increases for $\lambda \to \lambda_c$ because of additional new time scales associated to the trajectories moving towards the potential barrier between the two attracting states.

This rather simple observation allows us to define two indicators: $\mathcal{O}(\lambda) = p(\lambda) + q(\lambda)$, which diverges for $\lambda \to \lambda_c$, and the total persistence of the process:

$$\mathcal{R}(\lambda) = \sum_{i=1}^{p} |\phi_i(\lambda)| + \sum_{i=1}^{q} |\theta_i(\lambda)|$$

which also show a divergent behavior at the transition, unless $\mathcal{O}(\lambda) = 1$. In this latter case $\mathcal{R}(\lambda) = \phi_1 \to 1$ for $\lambda \to \lambda_c$. These indicators present several advantages with respect to the traditional ones reported, for example, in [1]. First, by computing the functional form for $p(\lambda)$ and $q(\lambda)$ and for the coefficients $\phi_i(\lambda)$ and $\theta_i(\lambda)$ one has also an effective statistical toy model for describing the phenomenon and to produce data with analogous statistical properties. This may be very useful for series or data which can hardly be reproduced (laboratory experiments) or integrated by new observations (climate datasets, stock market titles). Second, if several series at different λ's are available, one can extrapolate the characteristics of the process at not yet measured λ's. This property can be very useful for devising new experiments knowing the possible location of the transitions. Third, if the transition is marked by the fact that $\mathcal{R}(\lambda) \to \mathcal{O}(\lambda)$ rather than by a divergence of $\mathcal{O}(\lambda)$, one may argue that the potential landscape for the observable X does not change significantly when approaching the transition and therefore a Langevin reduction to a double well system is possible. If, on the contrary, the order changes significantly approaching the transition, such a low dimensional reduction is not appropriate and one should be very careful in pursuing such a model as shown, for a relevant climatic example, in [14].

A toy model for critical transitions. We start considering a classical system featuring bistability under the effect of random noise, i.e.

$$\mathrm{d}X = -V'(X)\mathrm{d}t + \epsilon \mathrm{d}W \tag{3}$$

with potential $V(X) = aX^4 - bX^2 + \lambda X$, where $\lambda \geq 0$ and W is a Wiener process with unit variance. We consider system (3) for values of λ such that, in the deterministic limit, it features two stable fixed points ($\bar{X}_1 < 0$ and $\bar{X}_2 > 0$) and an unstable fixed point $\tilde{X}$. The asymptotic behavior of the system can be assessed in terms of the solution of a Fokker-Plank equation [15]. Here we are rather interested in the finite-time behavior and we consider only the simulations such that the noise does not push the system across the bifurcation, *i.e.* the system is confined in one of the two wells. For this system we perform the following numerical experiments: for each value of λ we compute an ensemble of 500 trajectories $X(\lambda)$ finding, for each of them, the best ARMA$(p(\lambda), q(\lambda))$ in the sense specified by the Box-Jenkins procedure [13]. Then, $\langle \mathcal{O}(\lambda) \rangle$ and $\langle \mathcal{R}(\lambda) \rangle$ have been computed, here $\langle \cdot \rangle$ stands for the ensemble average.

In Fig. 1-a we report the results of this analysis, which clearly show that the average order is not affected in this case and $\langle \mathcal{O}(\lambda) \rangle \simeq 1$, whereas the transition is well highlighted by $\langle \mathcal{R}(\lambda) \rangle$ which approaches $\langle \mathcal{O}(\lambda) \rangle = 1$ for $\lambda \to \lambda_c$. There is a simple way to understand this behavior by linking the orders p, q and of the coefficients ϕ_i to the autocorrelation function ACF (see [12]- Chapter

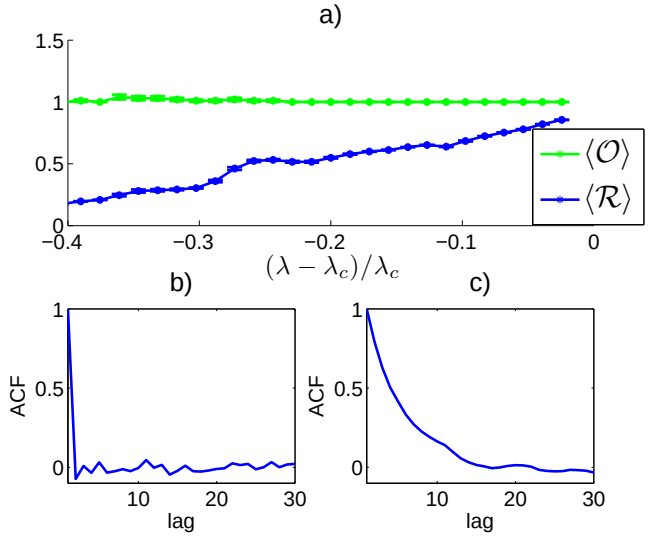


FIG. 1: (a):$\langle\mathcal{O}\rangle$ and $\langle\mathcal{R}\rangle$ for the system defined in Eq. 3. Each error-bar represents the average of 15 realizations and the standard deviation of the mean. b) ACF for $(\lambda-\lambda_c)/\lambda_c = -0.271$. c) ACF for $(\lambda-\lambda_c)/\lambda_c = -0.05$.

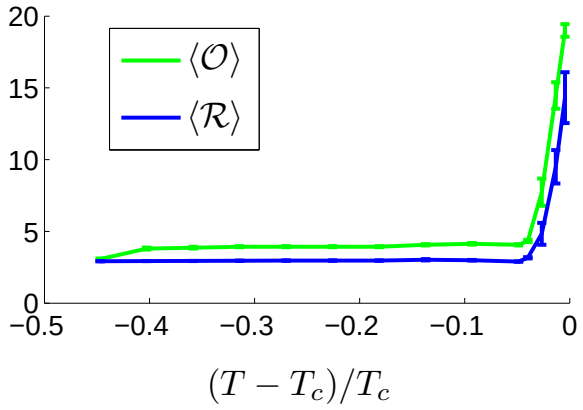


FIG. 2: $\langle\mathcal{O}\rangle$ and $\langle\mathcal{R}\rangle$ for the system defined in Eq. 4, L=256. Each error-bar represents the average of 15 realizations and the standard deviation of the mean.

3 for more details). The orders are linked to the functional form of the ACF whereas the values of ϕ_i depend on the decay rate. In the case of system given by Eq. 3, the shape of the ACF is exponential both far from the transition (Fig 1-b) and when approaching it (Fig 1-c). However, in the latter case, the decay is much slower, this causing the increase $\mathcal{R}(\lambda) \to 1$.

The Ising model. As a second test of the indicator, we consider a classical 2-D Ising dynamics with a nearest-neighbor interaction on a square lattice of size L. At each site j, a discrete spin is allowed to have two values $\sigma_j \in \{+1, -1\}$. The energy of the configuration is given by the Hamiltonian:

$$\mathcal{H} = -J \sum_{\text{neighbors}} \sigma_i \sigma_j \qquad (4)$$

under the interaction J. We consider only the case $J = 1$ and evolve the system by using Metropolis algorithm [16]. A second order phase transition is expected at the temperature $T = T_c = 2/\ln(1+\sqrt{2}) \simeq 2.269$. To apply our early warning indicators, we performed 100 simulations for a square lattice of size $L = 256$ at different $T > T_c$. We checked that our results do not depend sensitively on the size of the lattice, provided that $L > 128$. After discarding 100000 time iterations necessary to reach a clearly identifiable stationary state, for each temperature, an ensemble of 15 time series consisting of 200000 time units of $\mathcal{H}(t, T)$ is analyzed with the procedure described above. Stationarity has been tested performing a Dickey-Fuller test on each time series. The results for $\langle\mathcal{O}\rangle$ and $\langle\mathcal{R}\rangle$ are reported in Fig. 2. It is evident that $\langle\mathcal{O}\rangle$ and $\langle\mathcal{R}\rangle$ increase when moving towards the critical temperature T_c. In this case, not only the persistence of the correlations $\mathcal{R}$, but also the number of terms $\mathcal{O}$ necessary to describe the process increases. This means that the transition cannot be modeled by a simple Langevin equation as other time scales become important. In other words, this transition is associated to a non-trivial unknown potential landscape.

An example of complex system. Up to now we have analyzed toy systems, extensively studied both analytically and numerically and for which the threshold are analitically predictable. However, interesting systems, such as turbulence or finance, lie on another level of complexity and one naturally wonders whether the technique provides reliable results. We focus on the data of the Von Kármán Sodium (VKS) experiment, a successful attempt to get a transition to dynamo in a laboratory turbulent liquid-metal experiment. The control parameter for the transition is the magnetic Reynolds number Rm. The interesting characteristic of this experiment is that several dynamo and no dynamo configurations have been obtained by changing the material of the impellers and of the cylinder [7, 17]. Here we focus on two different configurations: (i) one producing a well-documented stationary dynamo at $Rm \approx 44$, thereby providing a fair test of our method and (ii) one that failed to produce the dynamo within the accessible values of Rm. The time series is constructed using the modulus of the magnetic field $|B|(t)$ as a function of time t, measured within six detectors in the core of the vessel. From this, we extract the quantities $\mathcal{O}$ and $\mathcal{R}$, averaged over the six detectors. The results are plotted in Fig. 3: the main figure for the configuration (i), the inset for the configuration (ii). They depend quantitatively on the sensors chosen, but not qualitatively as the transition is always detected at the same Rm. The transition is very net in terms of divergence of $\mathcal{O}$ and $\mathcal{R}$ and can be located at $Rm_f = 47$, when the dynamo is observed. In the non-dynamo case,

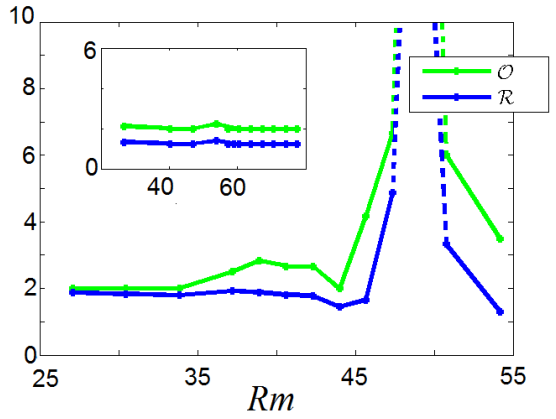


FIG. 3: Averaged $\mathcal{O}$ and $\mathcal{R}$ for the Von Karman - Sodium experiment. Solid lines refer to the experiments for increasing values of Rm. Inset: same as the main figure but for a configuration where the dynamo has not been observed.

no sign of transition is visible.

Financial crisis. We conclude by discussing the performances of the ARMA early warning indicators applied to the EUR/USD exchange rate hourly datasets (Figure 4-b). The chosen observable is the log-return of the time series, a quantity commonly examined in finance as the series obtained this way do not contain non-stationarities:

$$R(t) = log(X_t) - log(X_{t-1}).$$

Here X_t is the EUR/USD exchange hourly rate. Monthly values for $\mathcal{O}$ and $\mathcal{R}$ have been extracted from the time series of $R(t)$ and results are shown in figure 4-a. Our analysis can be safely performed on these series since they are stationary, as it results from the Dickey-Fuller test [18]. The technique clearly points to three distinct warnings (marked by the red dotted lines). Interestingly, they are followed after a few months delay, by official warnings of the European Central Bank (ECB). The first warning corresponds to the Sub-prime American crisis, the second to the Greek crisis, and the third one to the Irish crisis. The crisis for the real market falls immediately after the ECB announcements. If we compare these results with the ones arising from physical systems, the warning seems to appear *too early*. We may argue that indicators which provide similar warnings are available also at the ECB. The time between the early warning discover and the ECB announcements may serve to the ECB for trying corrections and avoid an immediate financial crisis which is announced only when the crisis itself is unavoidable. Similar behaviors have been discovered for early warning indicators applied to financial datasets, as reported in [19, 20] .

Discussion. In this work we have proposed a new method to detect early warnings of critical transitions via a statistical approach which allows also for incorporating the information of several statistical indicators analysed in [2]. Here, we exploit the properties of linear (*i.e.* stationary and invertible) autoregressive moving-average processes, denoted ARMA$(p(\lambda), q(\lambda))$, being λ the system control parameter. More specifically, we have defined two indicators representing the total order and the total persistence of the stochastic process. An increase of the former is indicative of the impossibility to represent the data series in a parsimonious way, thus leading to the idea that the linearity hypothesis fails and the decay of the autocorrelation function of the process is no longer exponential; an increase of the latter is linked to a longer decay time of the correlation, and can be due to the increase of the total order or just of the persistence itself. The two phenomena are very different and, up to our knowledge, the traditional indicators exploit only the increase of the memory of the system (critical slowing down) to identify the threshold λ_c. Here, the possibility of detecting substantial modifications to the shape of the autocorrelation function via the change in the orders p, q, allows for understanding whether reductions to simple low dimensional models are relevant

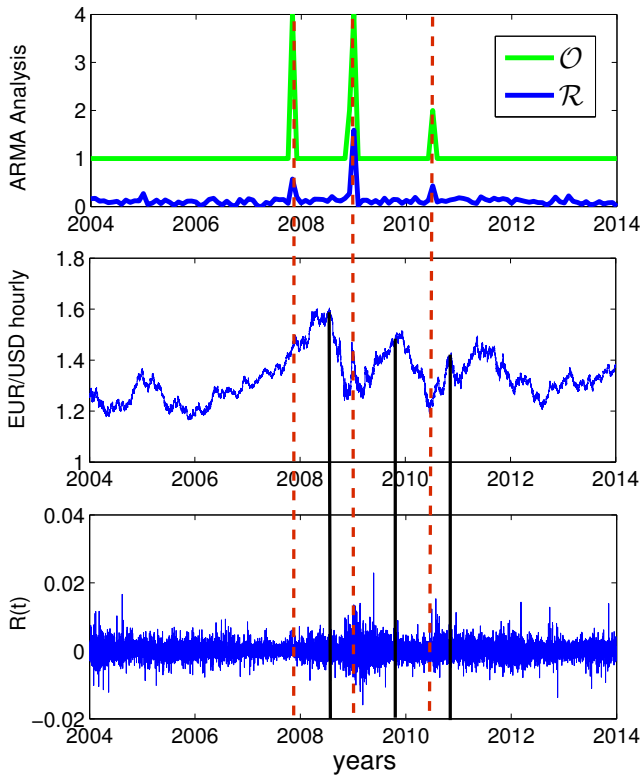


FIG. 4: Upper panel: Average $\mathcal{O}$ and $\mathcal{R}$ for the log-return series $R(t)$ of the EUR/USD hourly exchange rate. Central panel: EUR/USD hourly exchange rate. Lower panel: $R(t)$. Red dotted lines correspond to early warning of the crisis. Black continuous lines correspond to actions taken by the ECB.

or not for describing the dynamics. We have combined these two indicators to detect critical transitions both in models and in real systems. In all the cases considered, the behavior of the indicators has shown to be an effective way to investigate the proximity of the system to a critical transition; thus, they seem to be a useful tool to study critical transitions, since their estimation involves well-known, standard statistical techniques characterized by a low computational cost and applicable to relatively short time series.

The application to finance seems promising. It would be interesting to extend this approach to other financial indicators as well as to climate data. On a theoretical level, one could use the technique to understand how transitions are modified when systems originally in equilibrium are driven out of equilibrium by forcing-dissipation mechanisms, starting from conceptual toy model of out-of-equilibrium Ising dynamics [21, 22].

* Electronic address: davide.faranda@cea.fr

[1] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, Nature **461**, 53 (2009).

[2] V. Dakos, S. R. Carpenter, W. A. Brock, A. M. Ellison, V. Guttal, A. R. Ives, S. Kefi, V. Livina, D. A. Seekell, E. H. Van Nes, et al., PLoS One **7**, e41010 (2012).

[3] D. Faranda, V. Lucarini, P. Manneville, and J. Wouters, Arxiv Preprint 1211.0510 (2013).

[4] C. Kuehn, Physica D **240**, 1020 (2011).

[5] R. Monchaux, M. Berhanu, S. Aumaître, A. Chiffaudel, F. Daviaud, B. Dubrulle, F. Ravelet, S. Fauve, N. Mordant, F. Pétrélis, et al., Physics of fluids **21**, 035108 (2009).

[6] M. Berhanu, B. Gallet, R. Monchaux, M. Bourgoin, P. Odier, J. Pinton, N. Plihon, R. Volk, S. Fauve, N. Mordant, et al., Journal of Fluid Mechanics **641**, 217 (2009).

[7] S. Miralles, N. Bonnefoy, M. Bourgoin, J.-F. P. Odier, P. Nicolas, G. Verhille, J. Boisson, F. Daviaud, and B. Dubrulle, Preprint (2013).

[8] S. Franz, C. Donati, G. Parisi, and S. C. Glotzer, Philosophical Magazine B **79**, 1827 (1999).

[9] G. Parisi, EPL (Europhysics Letters) **40**, 357 (1997).

[10] D. Faranda, V. Lucarini, G. Turchetti, and S. Vaienti, Journal of Statistical Physics **145**, 1156 (2011).

[11] D. Faranda and S. Vaienti, Geophysical Research Letters **40**, 1 (2013).

[12] P. J. Brockwell and R. A. Davis, *Time series: theory and methods* (Springer, 2009).

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

[14] V. Lucarini, D. Faranda, and M. Willeit, Nonlinear Processes in Geophysics **19**, 9 (2012).

[15] H. Risken, *Fokker-Planck Equation* (Springer, 1989).

[16] H. Haario, E. Saksman, and J. Tamminen, Bernoulli pp. 223–242 (2001).

[17] J. Boisson, S. Aumaitre, N. Bonnefoy, M. Bourgoin, F. Daviaud, B. Dubrulle, P. Odier, J. Pinton, N. Plihon, and G. Verhille, New Journal of Physics **14**, 013044 (2012).

[18] D. A. Dickey and W. A. Fuller, Econometrica: Journal of the Econometric Society pp. 1057–1072 (1981).

[19] S. Fischer, Review of world economics **139**, 1 (2003).

[20] E. P. Davis and D. Karim, Journal of Financial stability **4**, 89 (2008).

[21] M. Pleimling, B. Schmittmann, and R. Zia, EPL (Europhysics Letters) **89**, 50001 (2010).

[22] D. Faranda, B. Dubrulle, B. Saint-Michel, and S. Thalabard, arXiv preprint arXiv:1309.3437 (2013).

Paper B

Faranda, D., Pons, F. M. E., Dubrulle, B., Daviaud, F., Saint-Michel, B., Herbert, É. and Cortet, P. P. (2014). Modelling and analysis of turbulent datasets using Auto Regressive Moving Average processes. *Physics of Fluids* **26(10)** 105101.

PHYSICS OF FLUIDS **26**, 105101 (2014)

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*

(Received 25 March 2014; accepted 16 September 2014; published online 6 October 2014)

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. [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

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

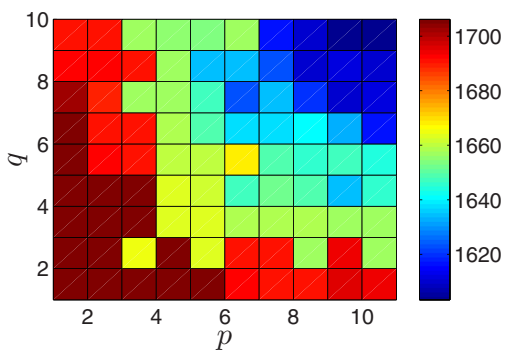


FIG. 1. *BIC* values resulting from different fits of ARMA(p,q) model for a velocity time series consisting of 600 observations.

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:

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

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

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).

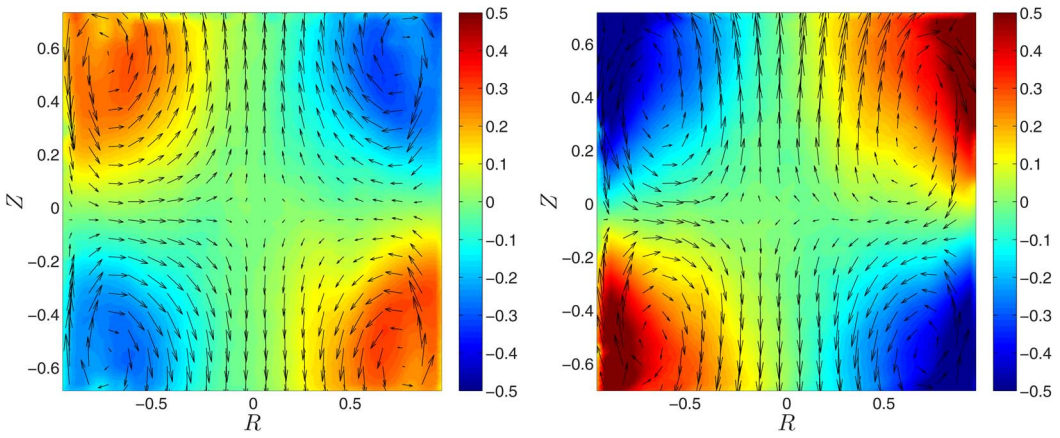


FIG. 2. Structure of the mean velocity field for $\theta = 0$. The arrows represents directions and intensities of the velocity components in the PIV plan averaged over time $(\bar{u}, \bar{v})$. The orthogonal component $\bar{w}$ is represented by the color scale. Left: (+) sense of rotation. Right: (−) 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

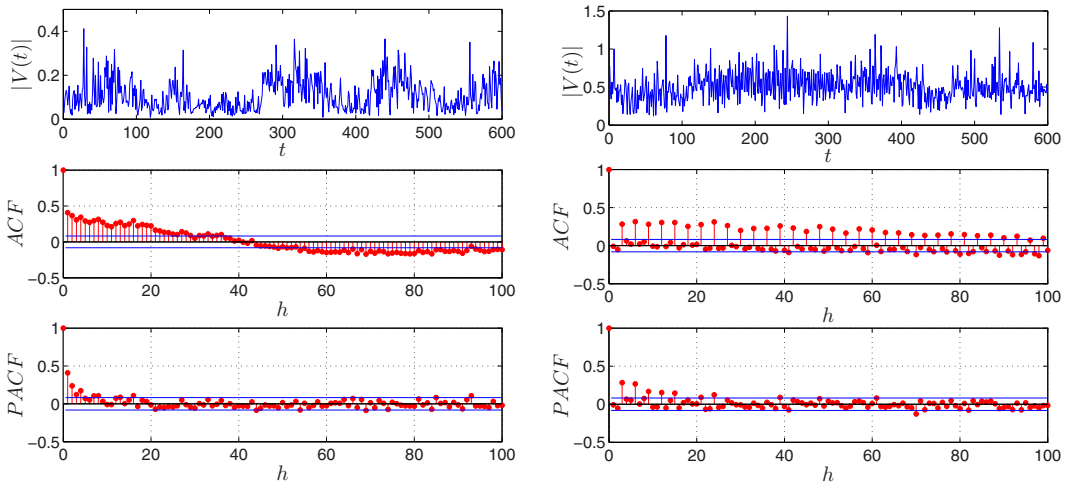


FIG. 3. Two time series of $|V(t)|$ (upper panels) with their respective ACF functions (middle panels) and PACF (lower panels). $Re \simeq 10^5$, (+) sense of rotation. Blue lines in the ACFs and PACFs represent the confidence bands at the 95% confidence level. Sample frequency: 15 Hz. X-axis is in sample index.

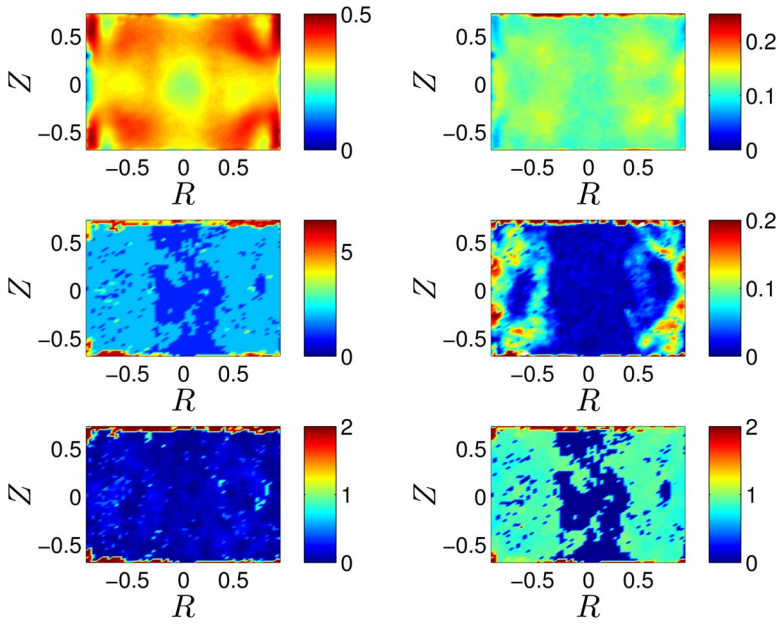


FIG. 4. ARMA analysis for the (+) sense of rotation. Top left: $|\bar{V}|$. Top right: $|\vec{V}(t)|$ standard deviation. Center left: Total order $\mathcal{O}$ found by fitting an ARMA(p,q) to the $|\vec{V}(t)|$ data. Center right: Distance from the Kolmogorov model Υ for the $|\vec{V}(t)|$ data. Bottom 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 $|\bar{V}|$

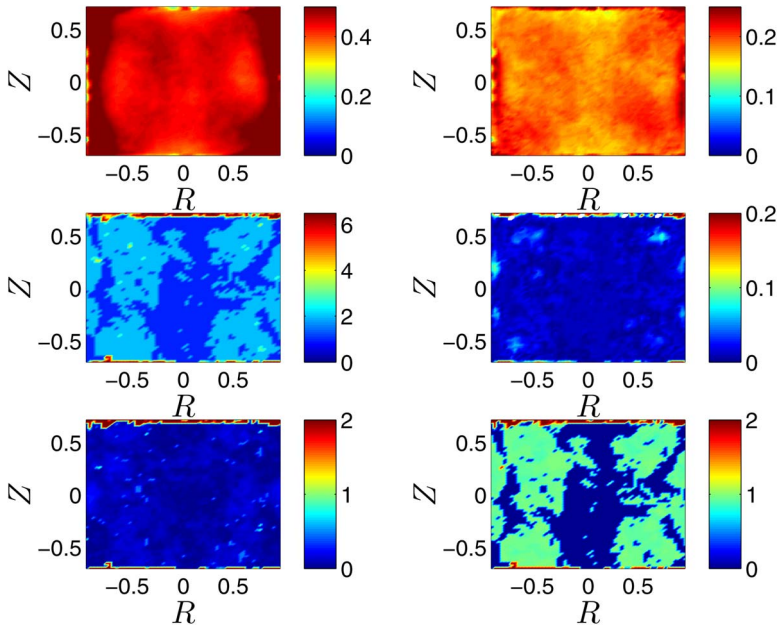


FIG. 5. ARMA analysis for the (−) sense of rotation. Top left: $|\bar{V}|$. Top right: $|\vec{V}(t)|$ standard deviation. Center left: Total order $\mathcal{O}$ found by fitting an ARMA(p,q) to the $|\vec{V}(t)|$ data. Center right: Υ for the $|\vec{V}(t)|$ data. Bottom left: Sum of the autoregressive 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.

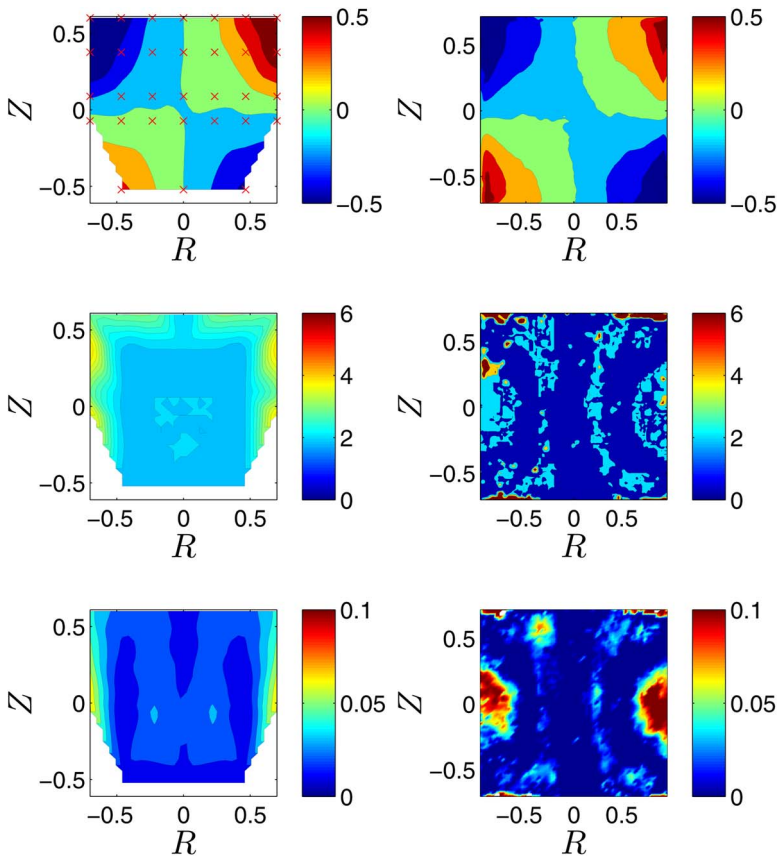


FIG. 6. Comparison between the ARMA analysis for LDV data (left) and PIV data (right) for the azimuthal component of the velocity w. $\bar{w}$ (upper panels), total order $\mathcal{O}$ (central panels) and distance from Kolmogorov model Υ (lower panels); $(-)$ sense of rotation; $Re \simeq 10^5$. The 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

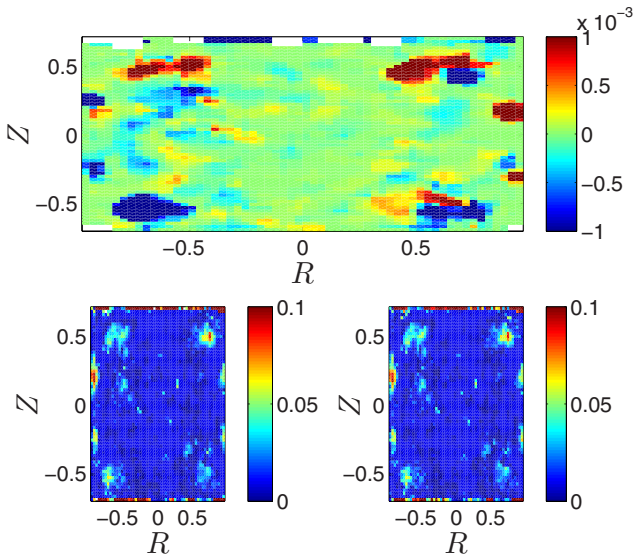


FIG. 7. Upper panel: Difference between Υ computed for the $|\vec{V}(t)|$ data for the $(-)$ sense of rotation and Υ on surrogate date of the same experiment obtained with a phase randomization procedure. Lower panel: Υ for the original data (left) and for the surrogates (right).

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

$$\mathrm{d}u = a(x,u,t)\mathrm{d}t + b(x,u,t)\mathrm{d}W^{2H}, \tag{13}$$

where $\mathrm{d}W^H$ 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

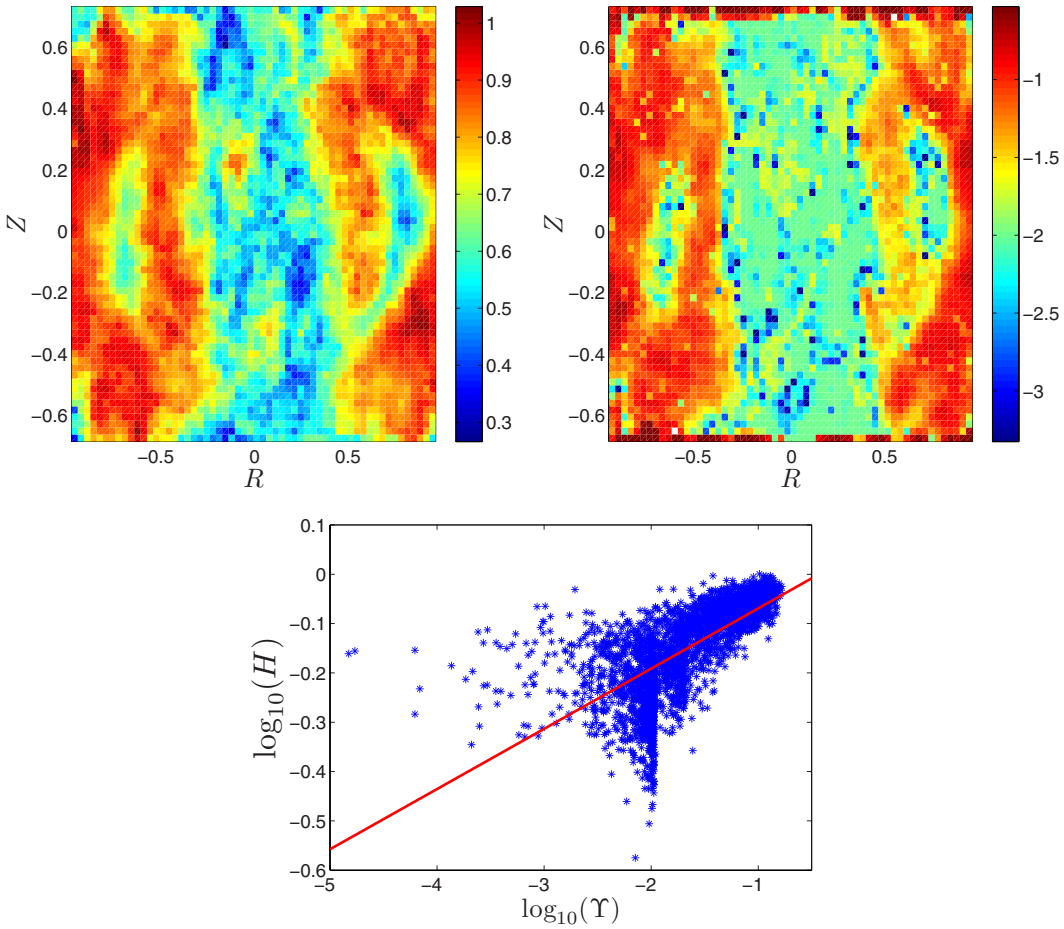


FIG. 8. Upper panels: Comparison between the Hurst exponents H (left) and the deviation from the Kolmogorov model in log-scale $\log_{10}(\Upsilon)$ (right) found by fitting an ARMA(p,q) to the $|\vec{V}(t)|$ data for the (+) sense of rotation. Lower panel: Scatter plot of the Hurst Exponent H and $\log_{10}(\Upsilon)$. 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

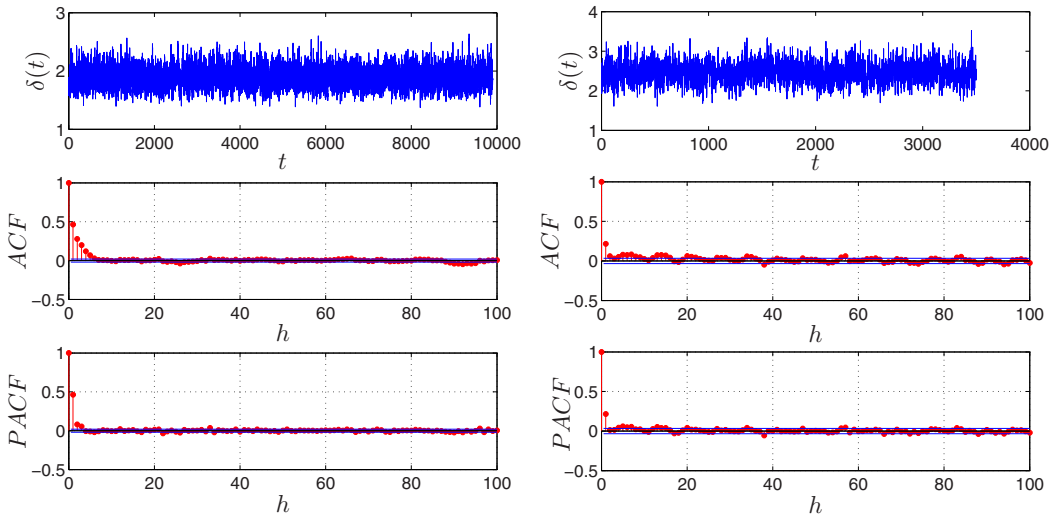


FIG. 9. Time series of $\delta(t)$ (upper panels) with their respective ACF functions (middle panels) and PACF (lower panels). Left: (+) sense of rotation. Right: (−) sense of rotation. $Re \simeq 10^5$. Blue lines in the ACFs and PACFs represent the confidence bands at the 95% confidence level. Sample 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

$$\mathcal{R} = \sum_{i=1}^{p} |\phi_i| + \sum_{i=1}^{q} |\theta_i|.$$

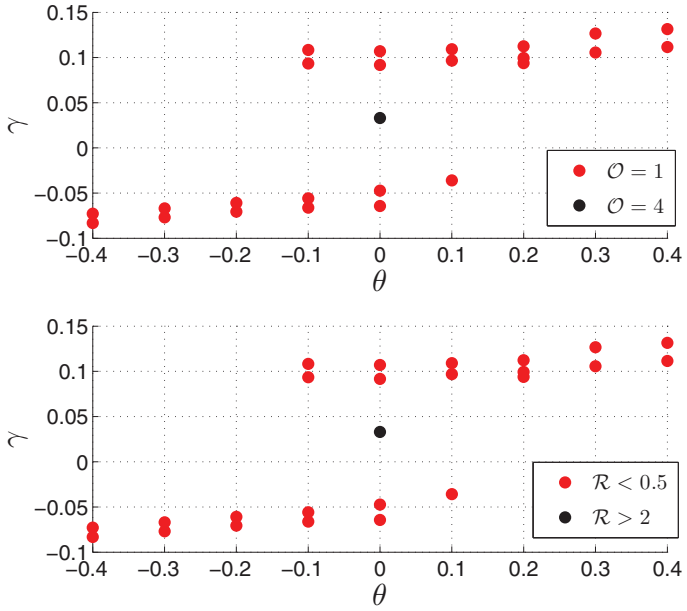


FIG. 10. Total order $\mathcal{O}$ and total persistence $\mathcal{R}$ for the von Kármán experiment under the speed control. The points represent the averaged γ and θ obtained for each experiment.

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

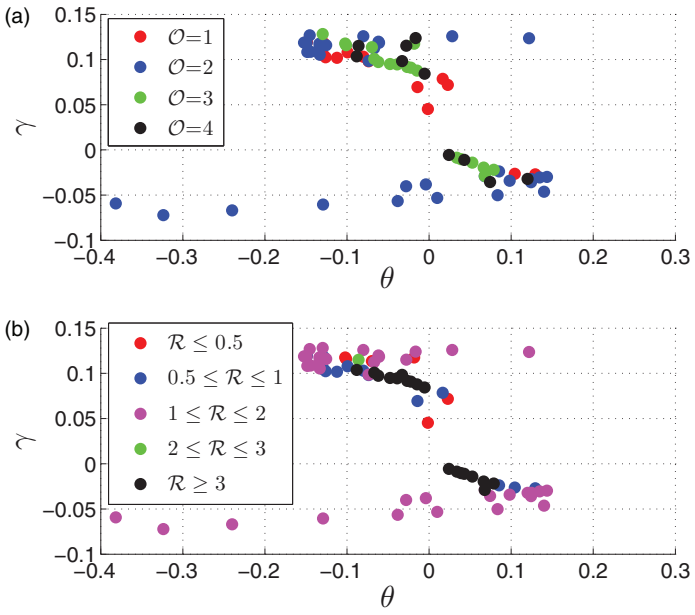


FIG. 11. (a) Total order $\mathcal{O}$ and (b) total persistence $\mathcal{R}$ for the von Kármán experiment under the torque control. The points represent the averaged γ and θ obtained for each experiment.

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

We acknowledge the other members of the VKE collaboration who performed the experiments: C. Wiertel and V. Padilla. We acknowledge two anonymous referees whose comments greatly improved the results and the consistency of this paper. D.F. acknowledges the support of a CNRS postdoctoral grant.

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

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

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).

Paper C

Faranda, D., Dubrulle, B., Daviaud, F. and Pons, F. M. E. (2014). Probing turbulence intermittency via autoregressive moving-average models. *Physical Review E* **90(6)** 061001.

PHYSICAL REVIEW E **90**, 061001(R) (2014)

Probing turbulence intermittency via autoregressive moving-average models

Davide Faranda,* Bérengère Dubrulle, and François Daviaud
Laboratoire SPHYNX, Service de Physique de l'Etat Condensé, DSM, CEA Saclay, CNRS URA No. 2464, 91191 Gif-sur-Yvette, France

Flavio Maria Emanuele Pons
Department of Statistics, University of Bologna, Via delle Belle Arti 41, 40126 Bologna, Italy
(Received 4 July 2014; published 3 December 2014)

We suggest an approach to probing intermittency corrections to the Kolmogorov law in turbulent flows based on the autoregressive moving-average modeling of turbulent time series. We introduce an index Υ that measures the distance from a Kolmogorov-Obukhov model in the autoregressive moving-average model space. Applying our analysis to particle image velocimetry and laser Doppler velocimetry measurements in a von Kármán swirling flow, we show that Υ is proportional to traditional intermittency corrections computed from structure functions. Therefore, it provides the same information, using much shorter time series. We conclude that Υ is a suitable index to reconstruct intermittency in experimental turbulent fields.

DOI: 10.1103/PhysRevE.90.061001 PACS number(s): 47.27.E–

Introduction. One of the few exact results known for isotropic, homogeneous, and mirror-symmetric turbulence is the 4/5 law derived by Kolmogorov. It links the longitudinal velocity increments $\delta u_\ell = u(x+\ell) - u(x)$ to the mean rate of energy dissipation $\langle \epsilon \rangle$ via

$$\left\langle \delta u_\ell^3 \right\rangle = -\tfrac{4}{5}\langle \epsilon \rangle \ell, \tag{1}$$

where angular brackets denote averaging. This exact relation was generalized by Kolmogorov [1] as a scaling law $\delta u_\ell \equiv (\epsilon \ell)^{1/3}$, where $\equiv$ means *has the same statistical properties*. Should ϵ be a nonstochastic constant, the scaling law would imply self-similar behavior for the structure functions of order p, $S_p(\ell) = \langle \delta u_\ell^p \rangle$, which would scale like

$$F_p(\ell) \sim \epsilon^{p/3} \ell^{p/3}. \tag{2}$$

For $p = 3$, we recover the 4/5 law. For $p = 2$, this equation predicts a second-order structure function that varies like $\ell^{2/3}$. By a Fourier transform, this is equivalent to a one-dimensional energy spectrum scaling with wave number k as $E(k) \sim k^{-5/3}$, also known as the Kolmogorov spectrum [2,3]. Both the 4/5 law and the Kolmogorov spectrum have been measured and checked in many natural and laboratory isotropic turbulent flows [4]. More generally, Eq. (2) predicts a linear law for the exponent of the structure functions $\zeta(p) = d \ln F_p(\ell)/d \ln \ell = p/3$. However, as pointed out by Landau and recognized by Kolmogorov [1], there is no reason to assume that ϵ is a constant over space and/or time, so it should rather be viewed as a stochastic process, which depends upon the scale ℓ at which it is measured $\epsilon \equiv \epsilon(\ell)$. In such a case, the correct scaling of the structure function is rather

$$F_p(\ell) \sim \langle \epsilon(\ell)^{p/3} \rangle \ell^{p/3}. \tag{3}$$

This modified law predicts correction to the linear law $\zeta(p) = p/3$, which is connected to the intermittent nature of the dissipation. For example, a log-normal model for the dissipation (a suggestion by Landau and Obukhov) implies quadratic corrections for the $\zeta(p)$. Other models have been suggested and lead to different corrections [5–7]. Intermittency corrections up to $p = 4$ have been measured in a variety of experimental and numerical flows and appear to be robustly consistent from one experiment to another (see, e.g., the review of [8]). Corrections for larger values of p are subject to resolution and statistical convergence issues: The larger the scaling exponent, the larger the statistical sampling must be in order to capture the rare events. There is presently no general consensus about the behavior of intermittency corrections at large order. This hinders progress in the understanding of the statistical properties of the energy dissipation. In this Rapid Communication we suggest an approach to probing intermittency corrections based on the autoregressive moving-average (ARMA) modeling of turbulent time series. We introduce an index Υ that measures the distance from a Kolmogorov-Obukhov model in the ARMA space. Applying our analysis to velocity measurements in a von Kármán swirling flow, we show that this index is proportional to the traditional intermittency correction computed from the structure function and provides the same information, using shorter time series.

Intermittency parameters. In most laboratory turbulent flows, data sets are time series of values of a physical observable at a fixed point or obtained by tracking Lagrangian particles. This 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 measurements of time structure functions $G_p(\tau) = \langle (\delta u_\tau)^p \rangle$ and its local exponent $\chi_p = d \ln G_p(\tau)/d \ln \tau$. 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 $\ell = U\tau$. When fluctuations are of the same order as the mean flow, however, the Taylor hypothesis fails. A suggestion has been made in [9] 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 considering a scale such that $\ell \sim \tau \delta u_\tau$ and may be seen as equivalent to modifying the space-refined Kolmogorov hypothesis into a time hypothesis $\delta u_\tau \equiv (\epsilon \tau)^{1/2}$,

*davide.faranda@cea.fr

which leads to

$$G_p(\tau) \sim \langle \epsilon^{p/2} \rangle \tau^{p/2}. \quad (4)$$

Such scaling is equivalent to the scaling obtained using the Lagrangian structure function. In any case, we may define the intermittency as the deviation of the local exponents $\zeta_p^* = \zeta_p$ (space increments) or $\zeta_p^* = \chi_p$ (time increments) with respect to a linear behavior and may be quantified to first order by the parameter

$$\mu = \zeta_2^* - \tfrac{2}{3}\zeta_3^*. \quad (5)$$

This factor is proportional to the logarithm of the β parameter of the log-Poisson model [5,6], or to the μ parameter of the log-normal model [1]. It is also valid when the scaling exponents have been computed using extended self-similarity (ESS) [10], which is especially interesting in situations where turbulence is inhomogeneous and when the Taylor hypothesis does not hold. In the following, we compare this intermittency index with another one, built in a purely statistical framework.

Indeed, Thomson [11] showed that, in the Lagrangian framework, the time-refined Kolmogorov hypothesis is in fact equivalent to a stochastic description in terms of an Ornstein-Uhlenbeck process with suitable drift and noise term

$$du = -\frac{u}{T}dt + \sqrt{C_0 \epsilon} dW, \quad (6)$$

where T is a decorrelation time scale, C_0 is a universal constant, and ϵ is the mean dissipation. Indeed, taking into account the definition of the particle position x, $dx = udt$, we get a scaling of the time averages of velocity and position as

$$\overline{u^2(t)} \sim t, \quad \overline{x^2(t)} \sim t^3. \quad (7)$$

The second property is the Richardson law. Then, defining $\delta u = [\overline{u^2(t)}]^{1/2}$ and $\ell = [\overline{x^2(t)}]^{1/2}$, we get from Eq. (7) $\delta u \sim \ell^{1/3}$, which leads to the space-refined Kolmogorov hypothesis.

The discrete-time version of Eq. (6) can be written as

$$u_t = \phi u_{t-1} + \psi_t, \quad (8)$$

where t is a discrete-time label, dW are the increments of a Brownian motion, $\phi = (1 - \frac{\Delta t}{T})$, and ψ_t are independent variables, normally distributed. Equation (8) is the expression of an autoregressive process of order one, denoted by AR(1). Such a model is described by a single decorrelation time. So it cannot describe real flows intermittency, which involves a whole range of time scales corresponding to the turnover times of the turbulent eddies with memory effects. To capture these effects, it is mandatory to consider a projection of the velocity data on higher-order ARMA(p,q) models. This enables a quantification of intermittency effects as a distance with respect to the insufficient AR(1) model in this space.

Intermittency as a distance in ARMA space. A stationary time series X_t is said to follow an ARMA(p,q) process if it satisfies the discrete equation

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad (9)$$

with $\varepsilon_t \sim N_W(0,\sigma^2)$, where N_W stands for white noise and the polynomials $\phi(z) = 1 - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p}$ and $\theta(z) = 1 - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q}$, with $z \in \mathbb{C}$, have no common factors. Notice that the noise term ε_t will be assumed to be a white noise, which is a general condition [12]. We ensure unicity by applying the Box-Jenkis procedure [13]: We choose the lowest p and q such that the residuals of the series filtered by the process ARMA(p,q) are not correlated. To define a suitable distance in the space of ARMA(p,q) models, we introduce the Bayesian information criterion C_{BI}, measuring the relative quality of a statistical model, as

$$C_{BI} = -2 \ln \hat{L}(n,\hat{\sigma}^2,p,q) + k[\ln(n) + \ln(2\pi)], \quad (10)$$

where $\hat{L}(n,\hat{\sigma}^2,p,q)$ is the likelihood function for the investigated model, $k = p + q$, and n is the length of the sample. The variance $\hat{\sigma}^2$ is computed from the sample and is a series-specific quantity. The normalized distance between the fit ARMA($p+1$,q) and the Kolmogorov AR(1) model is then defined as the normalized difference between the $C_{BI}(n,\hat{\sigma}^2,p+1,q)$ and the AR(1) $C_{BI}(n,\hat{\sigma}^2,1,0)$:

$$\Upsilon = 1 - \exp\{|C_{BI}(p+1,q) - C_{BI}(1,0)|\}/n. \quad (11)$$

The $p+1$ serves to magnify Υ near zero. Note that $0 \leqslant \Upsilon \leqslant 1$: It goes to zero if the data set is well described by an AR(1) model and tends to one in the opposite case. In the case of velocity increments time series, it measures deviations from the Kolmogorov model.

Application to turbulent data. We apply the index defined in Eq. (11) to velocity time series obtained in a von Kármán turbulent swirling flow. The experimental setup consists of two sets of blades mounted on two counterrotating coaxial impellers at the top and bottom of a cylindric vessel of diameter $R = 0.1$ m. The operating fluid is water and the rotation frequency of the impellers can reach $F = 15$ Hz, resulting in large Reynolds numbers (Re $= 2\pi F R^2 \nu^{-1} \sim 10^6$). A detailed description of the experiment can be found in [14–16]. Two techniques are used to measure the fluid velocity on a grid: the particle interferometry velocimetry (PIV) and the laser Doppler velocimetry (LDV), mapped on a regular sampling time applying a sample-and-hold algorithm. The stereoscopic PIV measures the three components of the velocity field in a plane, while the LDV measurements provide the out-of-plane velocity component V_ϕ in a plane. The PIV produces regularly sampled time series at intervals of 0.1 s over a sample size at most of order 10^4 and a spatial resolution of the order of 1 mm, i.e., 10–100 times larger than the dissipation scale. The LDV time series are sampled over a time scale of the order of 0.001 s, producing a sample size up to 10^6 data on a grid of spatial resolution of the order of 1 cm. Given these resolution constraints, we compute spatial (temporal) velocity increments for the PIV (LDV) data. The idea is to compute at each spatial grid location the classical intermittency index μ, compare it to Υ, and see how they vary. All the analyses presented in this Rapid Communication are done using three components for the PIV and V_ϕ for the LDV. Since the von Kármán flow is inhomogeneous and anisotropic with large fluctuations [16], we expect that the time and space velocity structure functions depend on the measurement points. This is illustrated in Fig. 1 for the second- and fourth-order spatial and time structure functions. For the spatial case, deviations from the Kolmogorov scaling (solid lines) are small for the spatial structure functions, near the symmetry plane $Z = 0$. This plane

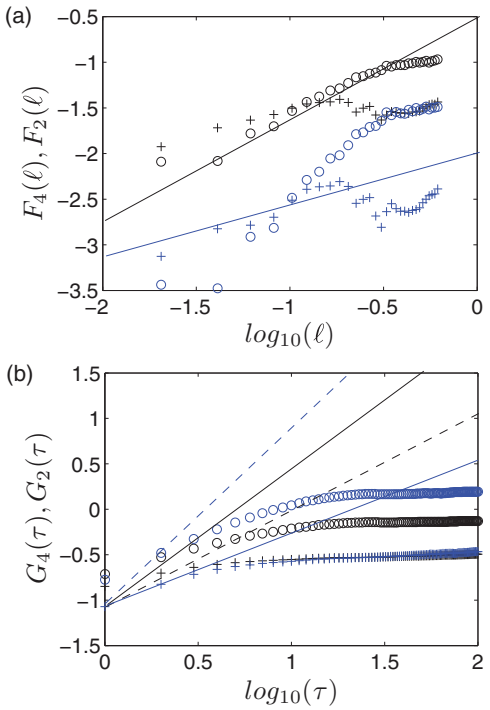


FIG. 1. (Color online) (a) Space structure function $F_p(\ell)$ of order 2 in black and of order 4 in blue (gray) at two PIV grid points of coordinates $R = -0.10$, $Z = -0.14$ (circles) and $R = 0.98$, $Z = 0.70$ (crosses). (b) Time structure function $G_p(\tau)$ of order 2 in black and of order 4 in blue (gray) at two LDV grid points of coordinates $R = 0$, $Z = 0.35$ (circles) and $R = 0.52$, $Z = 0.61$ (crosses). Lines represent the Kolmogorov predictions: solid is Eulerian with $\ell^{2/3}$ and $\tau^{2/3}$ in black and $\ell^{4/3}$ or $\tau^{4/3}$ in blue (gray) and dotted is Lagrangian with τ in black and τ^2in blue (gray).

is the location of an intense shear layer and has traditionally been used to perform isotropic homogeneouslike measurements. Outside this plane, deviations from the Kolmogorov scaling are large. For the time case, one observes two distinct behaviors: Outside the shear layer, where a mean velocity is well defined, one observes close to Eulerian-Kolmogorov scaling at the smallest time increments $\tau^{p/3}$; in the shear layer, where no Taylor hypothesis holds, the scaling is closer to Lagrangian scaling $\tau^{p/2}$. However, as already noted by [9] and shown in Fig. 2, the relative scaling exponents ζ_p^* computed as $G_p(\tau) \sim \langle|\delta u_\tau|^3\rangle^{\zeta_p^*}$ (ESS method) are in most of the flow close to the universal scaling exponents found by [8], in a variety of homogeneous turbulent flows, even those with no obvious inertial range. Using these ESS scaling exponents to compute the μ index, we may then draw a map of the intermittency and compare it with Υ. This is done in Fig. 3 for an LDV experiment at Re $\sim 10^5$. The spatial patterns look indeed similar. Moreover, the plot of Υ as a function of μ [Fig. 3(c)] evidences a linear relation between them; the linear regression represented by the red line leads to a linear correlation coefficient $r \simeq 0.69$. This means that Υ traces the same intermittency characteristics as the time structure functions. The comparison of Υ with the intermittency index μ computed for spatial structure functions is also informative: Because of convergency issues, we have to use a data set

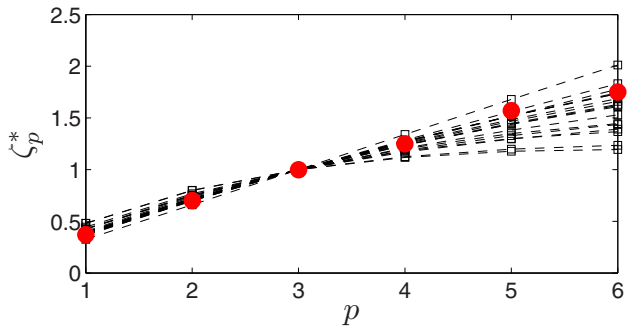


FIG. 2. (Color online) Plot of ζ_p^* computed for the LDV experiments. Different lines correspond to different measure points. Red (gray) spots mark the scaling exponents reported in [8].

of about 10^5–10^6 data points to converge the estimate of μ, while only 10^3 are needed to converge Υ. To illustrate this, we use the longest data set available: 9000 velocity fields of a PIV experiment performed at Re $\simeq 5 \times 10^4$. At this value, the von Kármán flow experiences the equivalent of a phase transition [17], with time wandering of the shear layer in between $Z = 0.3$ and -0.3. This corresponds to a very large time intermittency and is detected by the Υ index as shown in Fig. 4, under the shape of two patches at $R \simeq 0$, $Z = 0.3$ and $R \simeq 0$, $Z = -0.3$. This pattern is unique to the phase transition and is not present in other PIV experiments [18]. Besides, one observes a fairly symmetric structure, with maxima corresponding to the flow's four-cell structure. The time intermittency prevents the convergence of the spatial

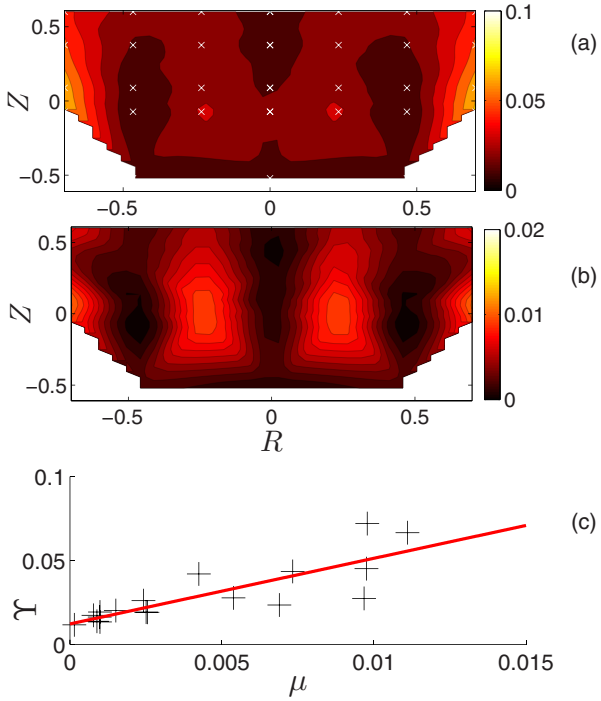


FIG. 3. (Color online) (a) Index Υ vs (b) the intermittency index $\mu = \zeta_2^* - \frac{2}{3}\zeta_3^*$. White crosses show measurement points. (c) Scatter plot of Υ vs μ. The red (gray) line shows a linear regression of the data.

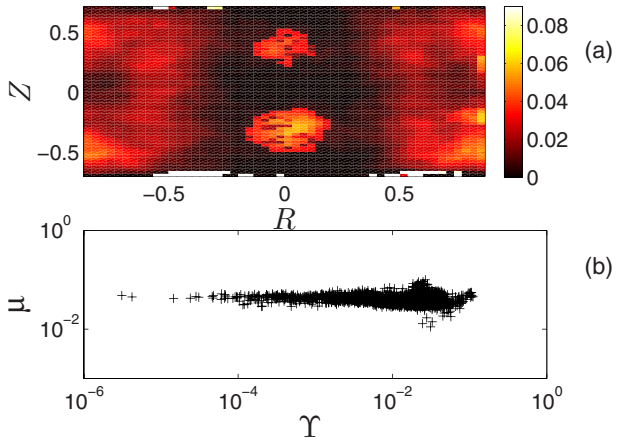


FIG. 4. (Color online) (a) Index Υ computed for a PIV experiment. The intermittency Υ maxima are localized near the walls and trace the position of the time wandering shear layer at $R = 0$. (b) Scatter plot of Υ vs μ, the former displaying a much greater sensitivity than the traditional intermittency index, despite the short length of the time series.

structure functions, resulting in a lack of symmetry of the μ field (not shown). As a result, μ fluctuates over a decade around a value of about 0.05, while Υ spans several orders of magnitude, as can be seen in Fig. 4(b). This shows that Υ is a more sensitive tool to detect intermittency than μ.

Discussion. We have introduced an intermittency index Υ that can be interpreted as a statistical distance between the best-fit linear ARMA model for a turbulent time series and the simplest possible process, i.e., AR(1). We have compared such an index with a classical intermittency index $\mu = \zeta_2^* - \frac{2}{3}\zeta_3^*$. In statistically converged LDV series, the two parameters are linearly related, with a regression coefficient $R \simeq 0.69$. In shorter PIV time series, Υ catches important characteristics of the mean flow in situations where μ cannot be computed due to a lack of convergence. Therefore, the main advantage of this index is the applicability to cases in which no big data sets are available. Moreover, ARMA models contain the information on nonlocal interactions between large and small scales in the parameters ϕ_i and θ_i. In general, the AR(p) part of an ARMA(p,q) process is the contribution of the large scales and represents the persistence of the process. If the process is more persistent than 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. When $\Phi \ll \Theta$ there is a clear separation between large and small scales. On the other hand, when their magnitude is similar such separation is not ensured anymore [18]. Our results reinforce the hypothesis of [19] that intermittency *propagates* in direct interactions between large and small scales, rather than in cascades. Finally, our method can be used to validate models based on the stochastic differential equation [20] with respect to experimental data.

Acknowledgments. Experimental data were provided with the help of E. Herbert, P. Ph. Cortet, C. Wiertel, and V. Padilla. D.F. acknowledges the support of a Centre National de la Recherche Scientifique postdoctoral grant.

[1] A. N. Kolmogorov, J. Fluid Mech. **13**, 82 (1962).
[2] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **30**, 299 (1941).
[3] A. Obukhov, Dokl. Akad. Nauk SSSR **32**, 22 (1941).
[4] U. Frisch, *Turbulence* (Cambridge University Press, Cambridge, 1996), Vol. 1, p. 310.
[5] Z.-S. She and E. Leveque, Phys. Rev. Lett. **72**, 336 (1994).
[6] B. Dubrulle, Phys. Rev. Lett. **73**, 959 (1994).
[7] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani, Phys. Rev. Lett. **17**, 3521 (1984).
[8] A. Arneodo, C. Baudet, F. Belin, R. Benzi, B. Castaing, B. Chabaud, R. Chavarria, S. Ciliberto, R. Camussi, F. Chilla *et al.*, Europhys. Lett. **34**, 411 (1996).
[9] J.-F. Pinton and R. Labbé, J. Phys. II **4**, 1461 (1994).
[10] R. Benzi, S. Ciliberto, C. Baudet, G. R. Chavarria, and R. Tripiccione, Europhys. Lett. **24**, 275 (1993).
[11] D. J. Thomson, J. Fluid Mech. **180**, 529 (1987).
[12] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. (Springer, Berlin, 1990).
[13] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 1st ed. (Wiley, New York, 1970).
[14] F. Ravelet, A. Chiffaudel, F. Daviaud *et al.*, J. Fluid Mech. **601**, 339 (2008).
[15] P.-P. Cortet, A. Chiffaudel, F. Daviaud, and B. Dubrulle, Phys. Rev. Lett. **105**, 214501 (2010).
[16] P.-P. Cortet, P. Diribarne, R. Monchaux, A. Chiffaudel, F. Daviaud, and B. Dubrulle, Phys. Fluids **21**, 025104 (2009).
[17] P. Cortet, E. Herbert, A. Chiffaudel, F. Daviaud, B. Dubrulle, and V. Padilla, J. Stat. Mech. (2011) P07012.
[18] D. Faranda, F. M. E. Pons, B. Dubrulle, F. Daviaud, B. Saint-Michel, É. Herbert, and P.-P. Cortet, Phys. Fluids **26**, 105101 (2014).
[19] J. Laval, B. Dubrulle, and S. Nazarenko, Phys. Fluids **13**, 1995 (2001).
[20] D. Faranda, B. Dubrulle, and F. M. E. Pons, J. Phys. A: Math. Theor. **47**, 252001 (2014).

Paper D

Faranda, D., Pons, F. M. E., Giachino, E., Vaienti, S. and Dubrulle, B. (2015). Early Warnings Indicators of financial crises via Auto Regressive Moving Average models. *Communications in Nonlinear Science and Numerical Simulation.*

Early Warnings Indicators of financial crises via Auto Regressive Moving Average models

Davide Faranda*
Laboratoire des Sciences du Climat et de l'Environnement,
CEA Saclay, 91191 Gif-sur-Yvette , France
and
Flavio Maria Emanuele Pons
Department of Statistics, University of Bologna,
Via delle Belle Arti 41, 40126 Bologna, Italy
and
Eugenio Giachino
and
Sandro Vaienti
Aix Marseille Université, CNRS, CPT,
UMR 7332, 13288 Marseille, France
and
Université de Toulon, CNRS, CPT,
UMR 7332, 83957 La Garde, France.
and
Bérengère Dubrulle
Laboratoire SPHYNX, SPEC, DSM, CEA Saclay,
CNRS URA 2464, 91191 Gif-sur-Yvette, France

January 7, 2015

Abstract

We address the problem of defining early warning indicators of financial crises. To this purpose, we fit the relevant time series through a class of linear models, known as Auto-Regressive Moving-Average (ARMA(p, q)) models. By running such a fit on intervals of the time series that can be considered stationary, we first determine

*The authors gratefully acknowledge F. Daviaud for useful discussions and comments.

the *typical* ARMA($\overline{p}, \overline{q}$). Such a model exists over windows of about 60 days and turns out to be an AR(1). Then, we define a distance Υ from such typical model in the space of the likelihood functions and compute it on short intervals of stocks indexes. Such a distance is expected to increase when the stock market deviates from its *normal* state for the modifications of the volatility which happen commonly before a crisis. We observe that Υ computed for the Dow Jones, Standard and Poor's and EURO STOXX 50 indexes provides an effective early warning indicator which allows for detection of the crisis events that showed precursors.

Keywords: Stochastic modeling, Tipping points detection, Extreme events

1 Introduction

In the late '70s, a succession of currency crises generated interest in Early warning indicators [22, 23]. Over the year, the indicators spread more generally to financial and economic crisis, generating methodological debates [17, 15]. Traditional statistical approach to this issue are based on specific properties of ideal statistical systems near critical transition: critical slowing down, modifications of the auto-correlation function or of the fluctuations [13], increase of variance and skewness [18], diverging susceptibility [21, 4, 20], diverging correlation length (see the book [25] for a comprehensive review). However, in many cases, these approaches fail to detect the financial crisis. First of all, such methods are *attractor* based, i.e. they assume that the system can be well described by relating the observation at the time t with the observation at the time $t + \tau$ by an empirical deterministic law describing a stationary state of the system (the so called attractor). This approach fails in describing financial data because, such processes involve a family of time scales rather than a single scale τ, [9, 3]. A second origin for the failure of traditional early warning indicators is due to the presence of human feed-backs on the system i.e. the constant attempt to keep economy in a state fit to make profits. Such feed-backs create some delays between the first early warning signals and the time at which the crisis is observed. In addition, traditional early warning indicators may be inapplicable in datasets containing a small number of observations (see *e.g.* [14]), which is usually the case in financial time series. This suggests that indicators based on single statistical properties are not well suited for financial analysis and that crisis detection must involve indicators based on global properties of the whole stochastic process. Here, we build a class of indicators based on the auto-regressive moving-average processes of order p, q ARMA(p, q), widely used to model and forecast the behavior of financial time series. We remark that the goal of this paper will not be to find the best model to describe stock indexes and make predictions: this would require at least the estimation of fractionally integrated (ARFIMA(p, d, q)) or conditionally heteroskedastic (GARCH(p)) models, among others. We will rather assess a *typical* ARMA(p, q) model able to capture the general features of the analyzed stock market and define the early warning indicators as deviations from such a model in a suitable likelihood space. In the first part of the paper, we recall some basics on ARMA(p, q) modeling and define corre-

sponding early-warning indicators. We then check that these indicators are able to detect the transition in theoretical financial models. We conclude the paper by presenting and discussing the results of the analysis for some stock indexes.

2 The method

Let us consider a series X_t of an observable with unknown underlying dynamics. We further assume that for a time scale τ of interest, the time series $X_{t_1}, X_{t_2}, ..., X_{t_\tau}$ represents a stationary phenomenon. Since X_t is stationary, we may then model it by an ARMA(p, q) process such that for all t:

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

with $\varepsilon_t \sim WN(0, \sigma^2)$ - where WN stands for white noise - and the polynomials $\phi(z) = 1 - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p}$ and $\theta(z) = 1 - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q}$, with $z \in \mathbb{C}$, have no common factors. Notice that, hereinafter, the noise term ε_t will be assumed to be a white noise, which is a very general condition [7]. For a general stationary financial time series, this model is not unique. However there are several standard procedures for selecting the model which fits at best the data. The one we exploit in this paper is the Box-Jenkis procedure [5]. We choose the lowest p and q such that the residuals of ARMA(p, q) fit are uncorrelated: to this purpose, we perform a Ljung-Box test for the absence of serial correlation (see, for example, [7]). This fixes p and q, and thus our statistical model. There are other model selection procedures based on information criteria (Bayesan or Akaike information criteria). We tested, that they all give clear indications for discriminating the model and that they provide qualitatively the same results of the Box-Jenkis procedure. Intuitively, p and q are related to memory lag of the process, while the coefficients ϕ_i and θ_i represent the persistence: the higher their sum (in absolute value), the slower the system is forgetting its past history.

Our definition of early warning indicators requires first the identification of the basic ARMA$(\overline{p},\overline{q})$ process (with $\overline{p}$ and $\overline{q}$ fixed) which is best suited to describe a Stock index, for

a time interval such that it can be considered stationary. This basic process plays the role of an *attractor* in the sense that it contains the information related to the dynamical properties of the system. With respect to the common attractors used in dynamical systems theory, dynamical indicators as the Lyapunov exponents are replaced by the coefficients ϕ_i and θ_j and the analogous of the attractor dimension is the number of terms p and q to be considered. We will comment on these analogies and on the possibility of choosing a reliable ARMA($\overline{p}$,$\overline{q}$) for stock market indexes in the next section. Now we turn to the ARMA based definition of the early warning indicators.

ARMA based Early warning indicators We consider a given Stock index, that is assumed to faithfully reflect the financial or economic conjuncture. In the absence of crisis, such index can be considered as stationary over a given time interval τ and can be fitted by a reference ARMA model. When crisis approach, the volatility of the index increases, and the best ARMA model describing the market will deviate from the basic one. The strongest the crisis, the larger the deviations will be. This suggest to introduce an early warning indicator as a suitable distance in the ARMA space from the reference model. For this, we introduce the Bayesian information criterion (BIC), measuring the relative quality of a statistical model, as:

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

where $\hat{L}(n, \hat{\sigma}^2, p, q)$ is the likelihood function for the investigated model and in our case $k = p + q$ and n the length of the sample. The variance $\hat{\sigma}^2$ is computed from the sample and is a series-specific quantity.

We can define a normalized distance between the referece ARMA($\overline{p}$, $\overline{q}$) and any other ARMA(p, q) model as the normalized difference between the $BIC(n, \hat{\sigma}^2, p+1, q)$ and the ARMA($\overline{p}$, $\overline{q}$) $BIC(n, \hat{\sigma}^2, \overline{p}, \overline{q})$:

$$\Upsilon = 1 - \exp\left\{|BIC(p+1, q) - BIC(\overline{p}, \overline{q})|\right\}/n. \tag{3}$$

with $0 \leq \Upsilon \leq 1$: it goes to zero if the dataset is well described by an ARMA($\overline{p}$, $\overline{q}$) model and tends to one in the opposite case.

It has been already observed that such indicators perform well in different physical systems, providing more information than the usual ones, based on the critical slow down due to the increase of correlations in the systems at the transition. These analyses have been reported in [12] where indicators similar to Υ have been used to model different physical systems: Ising and Langevin models, climate and turbulence.

3 Assessment of the reference model

We now perform the analysis on different stock indexes: the Dow Jones, the Standard & Poor's and the EURO STOXX 50. We consider databases of such indexes from January 1st 1990 to June 30th 2014 containing daily observations. The considered EURO STOXX 50 series is slightly longer (January 1st 1986 to June 30th 2014).

We first determine the time interval on which the series can be considered stationary. We run the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test for a unit root on the time series with increasing τ for $\tau > 10$ days. The test is successful for $10 < \tau < 90$ days but, as expected, results depend on how close to a crisis the window is taken. In order to preserve enough statistical information we chose a time window $\tau = 60$ days. We tested that results presented are in fact stable for $40 < \tau < 80$ days. We then compute for each index the reference ARMA($\overline{p}$, $\overline{q}$) model using a statistical strategy : for each window $\{X_t, ...X_{t+\tau}\}$ for $t = 1, 2, ..., T - \tau$, being T the total length of the series, we fit the best ARMA(p, q) describing the series in this time lag. We then count the frequency of all ARMA(p,q) that have the same total order $p + q$ and compute histograms of $p + q$. This is shown in Fig. 1 for each of the three indexes. From the figure, we see that there is a peak of probability around $\overline{p} + \overline{q} = 1$. Since p and q are integers, and since we exclude pure moving average (MA(1)) fits, this means that the most probable model has $\overline{p} = 1$, $\overline{q} = 0$. So, for each index, we choose the reference model as an AR(1).

Once the reference model is established, we can now compute the Υ indicator over all the time series, and see how it perform over a database of financial crises and by comparison with other financial indicators linked to the volatility of the markets. We thus consider the database [1] which reports the crises between 1940 and 1999, and the database [19] for the

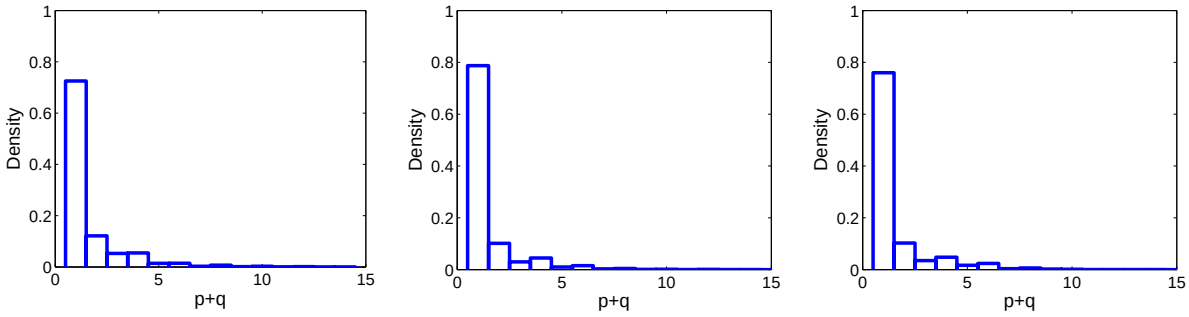


Figure 1: Assessment of the best ARMA($\overline{p}$,$\overline{q}$) model. The histograms represent the number of times (normalized to 1) that a certain $p+q$ ARMA order appear as best fit for sub-series of the originals in running windows of $\tau = 60$ days. Left panel: analysis for the Dow Jones index. Central panel: analysis for the Standard & Poor's index. Right panel: analysis for the EURO STOXX 50 index.

most recent ones. Naturally, we do not expect to get early warning indicators for crises connected to exogenous shocks, such as natural disasters (hurricanes, floods, earthquakes) or terrorism (Oklahoma bombs or September 11th 2001 events). The early warning, if any, should appear before the corresponding crisis.

4 Analysis

For the daily series of the Dow Jones, the Standard & Poor's and the EURO STOXX 50 indexes, we compute the quantity Υ defined in Eq. (3) using a running time windows of $\tau = 60$ days. At day $t + \tau$ we associate the Υ index computed using the observations $\{X_t, ..X_{t+\tau}\}$. We then smooth Υ data by using the moving average method with span of 5 days. We consider only values of $\Upsilon > 0.3$.

For the Dow Jones index results are shown in Figure 2. The upper panel shows the series of Dow Jones from January 1st 1990 to June 30th 2014, the central panel the daily changes of the Dow Jones index and the lower panel Υ early warning indicator (blue), thresholded as described before. The red stars indicate crisis of economical origins as derived from the databases [1], whereas yellow stars refer to crisis of non-economical origin. Early warnings provided by Υ are empirically associated to the respective crisis by red lines. Question

marks represent early warnings not associated to effective economical crises. The first thing to remark is that more than one warning is associated to the same crisis and that the interval between the early warning and the crisis is not constant. If we compare these results with the ones arising from physical systems and discussed in [12], the warnings for financial crises seem to appear *too early* with respect to what observed in controllable physical systems. The main difference of markets with respect to natural systems is that the former feel and react to the effect of the crisis by delaying its emergence until the time an official institution points to an economical problem. This usually happens when the crisis itself is unavoidable. For the Dow Jones time series, the matching between crisis and early warnings seem satisfactory although some warnings are of difficult interpretation.

A slightly different scenario appears when the Standard & Poor's stock index analysis is considered, as reported in Fig. 3. For this index some crises are well anticipated by an increase on Υ, some others instead are not captured. In order to explain the difference between the Dow Jones and the Standard & Poor's behavior we have to recall the way they are constructed: The major difference between them is that the Dow Jones includes a price-weighted average of 30 stocks whereas the Standard & Poor's is a market value-weighted index of 500 stocks. We can speculate that for indexes computed on a larger number of companies, crisis early warnings may be averaged out. In fact, if a relevant number of them will not suffer the effects of the crisis, no warning will be provided. Another possibility is that not all these companies have access to latest speculations on the market and therefore most of them cannot react in advance producing no early warnings at all.

The latter analysis concerns the EURO STOXX 50 index, extensevily analysed in [6], and it is reported in Fig. 4. As for the Dow Jones, the number of stocks considered is quite limited and crisis are well highlighted. We can remark that the average delay between early warning and crisis is shorter for the EURO STOXX than for the American indexes. This might be due to the fact that the European market usually follows the warnings and the speculations happening on the American side.

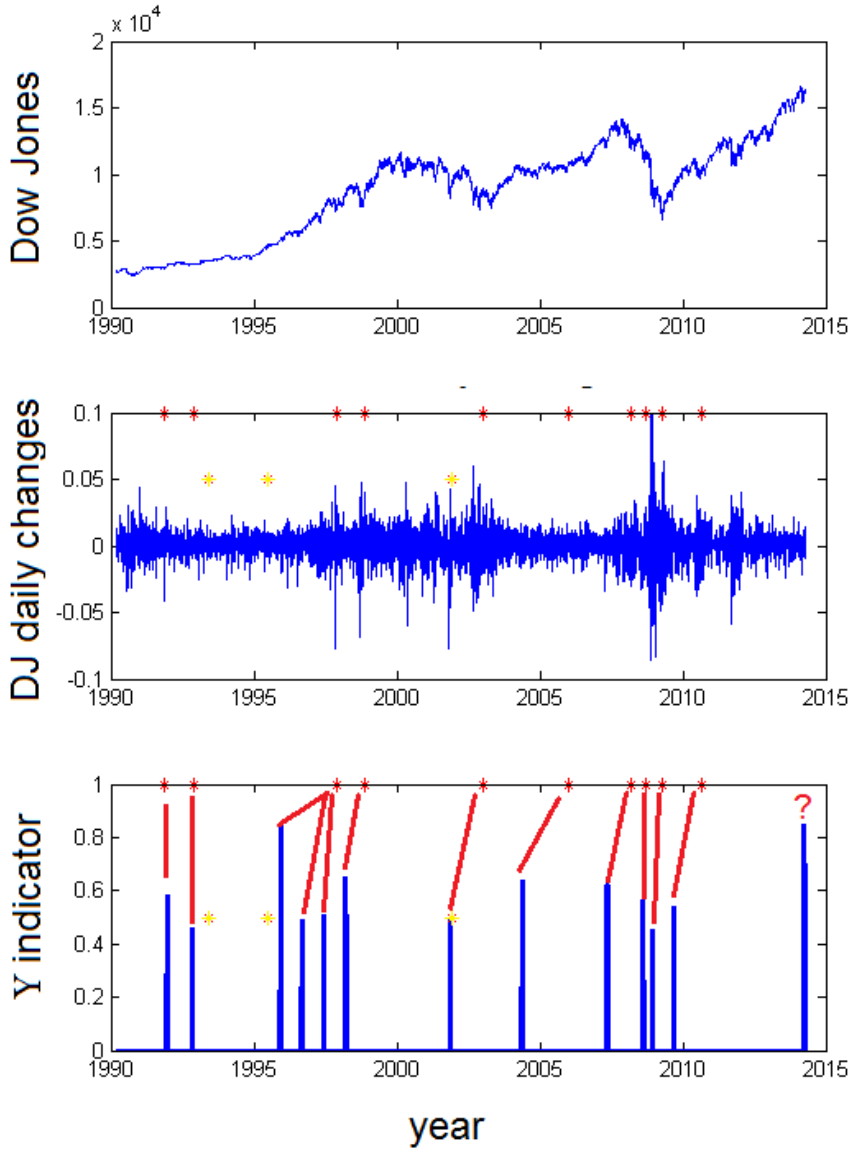


Figure 2: Early detection of crisis based on the Dow Jones stock index analysis. Upper panel: the series of Dow Jones from January 1st 1990 to June 30th 2014. Central panel: daily changes of the Dow Jones index, red stars indicate crisis of economical origin found in the databases [1, 19], yellow stars refer to crisis of non-economical origin. Lower panel: Υ early warning indicator (blue) associated to the respective crisis by red lines. Question marks represent early warnings not associated to effective economical crisis. See text for more descriptions.

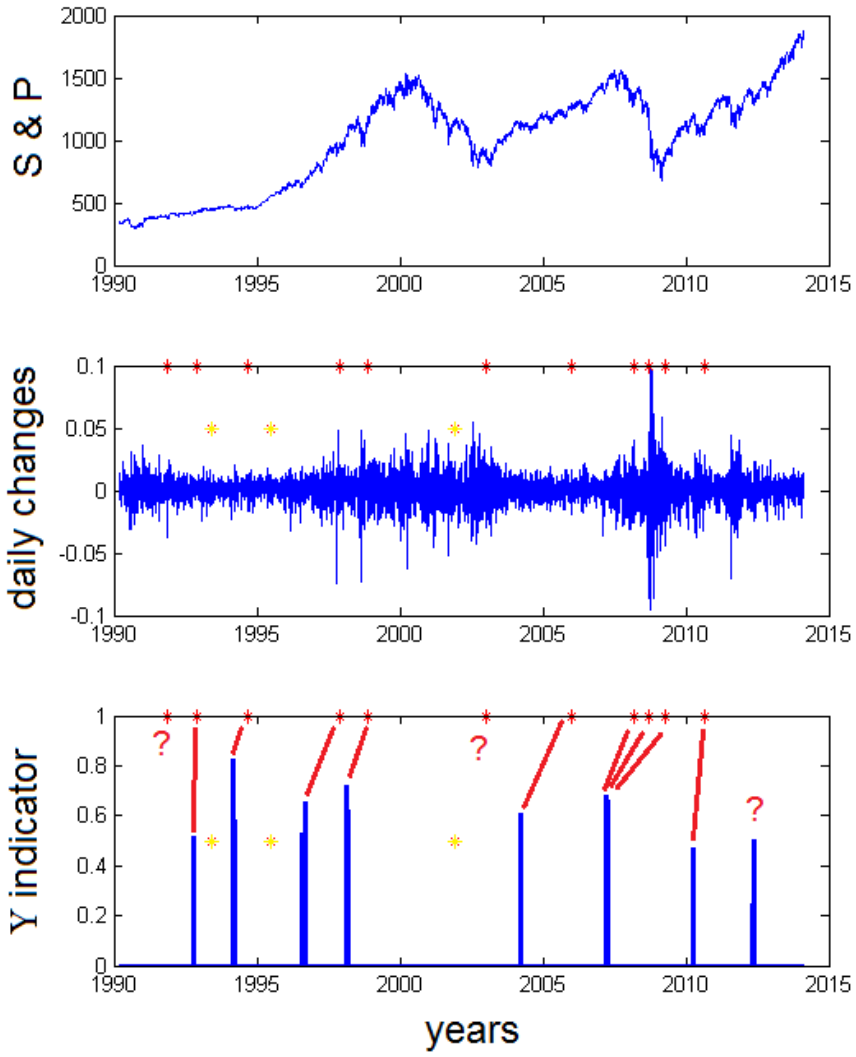


Figure 3: Early detection of crisis based on the Standard & Poor's stock index analysis. Upper panel: the series of Standard & Poor's from January 1st 1990 to June 30th 2014. Central panel: daily changes of the Standard & Poor's index, red stars indicate crisis of economical origin found in the databases [1, 19], yellow stars refer to crisis of non-economical origin. Lower panel: Υ early warning indicator (blue) associated to the respective crisis by red lines. Question marks represent early warnings not associated to effective economical crisis. See text for more descriptions.

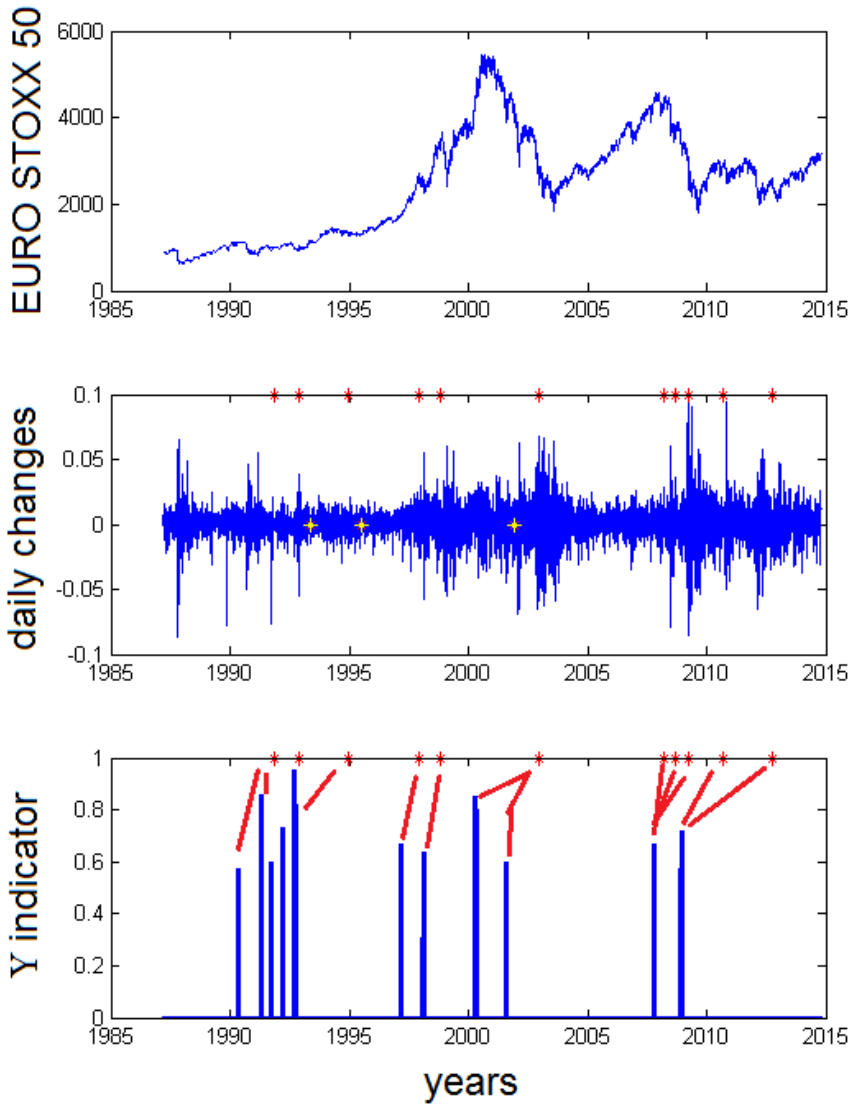


Figure 4: Early detection of crisis based on the EURO STOXX 50 index analysis. Upper panel: the series of EURO STOXX 50 from January 1st 1986 to June 30th 2014. Central panel: daily changes of the EURO STOXX 50 index, red stars indicate crisis of economical origin found in the databases [1, 19], yellow stars refer to crisis of non-economical origin. Lower panel: Υ early warning indicator (blue) associated to the respective crisis by red lines. Question marks represent early warnings not associated to effective economical crisis. See text for more descriptions.

5 Discussion

We have introduced an early warning indicator Υ for financial crises of economical origin based on the ARMA models. The indicator provides, for each day, a [0-1] distance with respect to a reference model ARMA($\overline{p}$, $\overline{q}$) which is able to fit the data on a certain window τ far from the crisis events.

The first result of the paper is that it is possible to statistically deduct such ARMA($\overline{p}$,$\overline{q}$) model which turns out to be the simple AR(1) often emerging in the description of natural phenomena driven by a Langevin equation, such as heavy particles in gas [2], polymers [24] and even turbulence [10]. We strongly remark that such model is not the best model to describe all the data of a specific index, but is the best one on a certain window τ chosen as parameter of the method. In some sense, the introduction of this model plays the role of the *attractor* of a physical system: attractor dimension is replaced by the total terms p and q needed to describe the series and Lyapunov exponents are linked to the magnitude of coefficients ϕ_i and θ_i. In physical systems crisis happen when the system departs from its attractor and explore new portions of the phase space. In terms of ARMA processes, crises happen when the model departs from the reference one and the series is fitted by other orders than $\overline{p}, \overline{q}$. It has been reported in [12] that this picture is true for physical systems ranging from toy models (Ising dynamics, Langevin problem) up to complex systems (turbulent flows).

For stock indexes, the indicator is useful for most of the crises reported in the databases, at least for the years we tested (1990 - 2014). The Υ indicator has different response functions according to the indexes to which it is applied. The larger the number of stocks used to construct the index, the lower is the early warning power of Υ. We have conjectured that such phenomenon is due to the fact that in indexes including more companies early effects of the crisis get averaged out either because some companies is not interested by the crisis, either because they do not have at their disposal any instruments to figure it out.

The instruments currently used to keep track of the markets volatility have been introduced by the CBOE (Chicago Board Options Exhange). Among such indexes, one of the most used is the VIX [16], which measures the expected volatility at 30 days for the

Standard & Poor's. This index has been created for the investors to have an information on the pure volatility of the market in the next month. In some sense, the VIX index should *feel* the modification of the markets and anticipate their change. However, as reported by CBOE [11, 8] VIX was not devised to predict stock prices, the direction of the market and highs or lows. Moreover, our Υ indicator is entirely based on the available data, rather on a forecast as the CBOE indexes. For a such a reason Υ will not be used as a substitute for the VIX index but rather as a completely different financial tool. As explained in [16], the VIX index is used to foresee on a month window the change of volatility in the market, not systemic changes in the behavior of the market. When a statistical comparison between the two indexes is done, the two indexes appear to be not correlated and they do not have the same statistical behavior: by construction the VIX index possess the same long memory and correlation structures of the S& P data series, whereas Υ consists of a series of peaks whose significance appears only when the original data series deviates from the AR(1) model. Υ therefore does not replace the VIX indexes: it is intended to forecast market shocks on the long term and not to foresee short terms volatility fluctuations.

Acknowledgments. DF acknowledges the support of a CNRS postdoctoral grant. DF and SV were supported by the ANR- Project Perturbations and by the project MOD TER COM of the french Region PACA

References

[1] Crisis events, djia declines and subsequent performance. `http://www.xpertadvice.com/Advisors/xpertadvice/market\%20reactions\%20to\%20crisis.pdf`. Accessed: 2014-09-30.

[2] S. Adelman and J. Doll. Generalized langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *The Journal of Chemical Physics*, 64(6):2375–2388, 1976.

[3] I. Andreadis and A. Serletis. Evidence of a random multifractal turbulent structure in the dow jones industrial average. *Chaos, Solitons & Fractals*, 13(6):1309–1315, 2002.

[4] M. Berhanu, B. Gallet, R. Monchaux, M. Bourgoin, P. Odier, J. Pinton, N. Plihon, R. Volk, S. Fauve, N. Mordant, et al. Bistability between a stationary and an oscillatory dynamo in a turbulent flow of liquid sodium. *Journal of Fluid Mechanics*, 641:217, 2009.

[5] G. E. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-D. iv, 1970.

[6] E. C. Brechmann and C. Czado. Risk management with high-dimensional vine copulas: An analysis of the euro stoxx 50. *Statistics & Risk Modeling*, 30(4):307–342, 2013.

[7] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer, 2009.

[8] J.-C. Duan and C.-Y. Yeh. Jump and volatility risk premiums implied by vix. *Journal of Economic Dynamics and Control*, 34(11):2232–2244, 2010.

[9] F. B. Duarte, J. T. MacHado, and G. M. Duarte. Dynamics of the dow jones and the nasdaq stock indexes. *Nonlinear Dynamics*, 61(4):691–705, 2010.

[10] B. Dubrulle, J.-P. Laval, S. Nazarenko, and O. Zaboronski. A model for rapid stochastic distortions of small-scale turbulence. *Journal of Fluid Mechanics*, 520:1–21, 2004.

[11] C. B. O. Exchange. The cboe volatility index–vix. *White Paper*, 2009.

[12] D. Faranda, B. Dubrulle, and F. M. E. Pons. Statistical early-warning indicators based on autoregressive moving-average models. *Journal of Physics A: Mathematical and Theoretical*, 47(25):252001, 2014.

[13] D. Faranda, V. Lucarini, P. Manneville, and J. Wouters. On using extreme values to detect global stability thresholds in multi-stable systems: The case of transitional plane couette flow. *Arxiv Preprint 1211.0510*, 2013.

[14] D. Faranda and S. Vaienti. A recurrence-based technique for detecting genuine extremes in instrumental temperature records. *Geophysical Research Letters*, 40:1–5, 2013.

[15] J. A. Frankel and A. K. Rose. Currency crashes in emerging markets: An empirical treatment. *Journal of international Economics*, 41(3):351–366, 1996.

[16] G. J. Jiang and Y. S. Tian. Extracting model-free volatility from option prices: An examination of the vix index. *The Journal of Derivatives*, 14(3):35–60, 2007.

[17] G. Kaminsky, S. Lizondo, and C. M. Reinhart. Leading indicators of currency crises. *Staff Papers-International Monetary Fund*, pages 1–48, 1998.

[18] C. Kuehn. A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics. *Physica D*, 240(12):1020–1035, 2011.

[19] L. Laeven and F. Valencia. Systemic banking crises database. *IMF Economic Review*, 61(2):225–270, 2013.

[20] S. Miralles, N. Bonnefoy, M. Bourgoin, J.-F. P. Odier, P. Nicolas, G. Verhille, J. Boisson, F. Daviaud, and B. Dubrulle. Dynamo threshold detection in the vks experiment. *Preprint*, 2013.

[21] R. Monchaux, M. Berhanu, S. Aumaître, A. Chiffaudel, F. Daviaud, B. Dubrulle, F. Ravelet, S. Fauve, N. Mordant, F. Pétrélis, et al. The von kármán sodium experiment: turbulent dynamical dynamos. *Physics of fluids*, 21:035108, 2009.

[22] R. H. Pettway and J. F. Sinkey. Establishing on-site bank examination priorities: An early-warning system using accounting and market information. *The Journal of Finance*, 35(1):137–150, 1980.

[23] S. Sharma and V. Mahajan. Early warning indicators of business failure. *The Journal of Marketing*, pages 80–89, 1980.

[24] I. Sokolov. Cyclization of a polymer: first-passage problem for a non-markovian process. *Physical review letters*, 90(8):080601, 2003.

[25] D. Sornette. *Why stock markets crash: critical events in complex financial systems*. Princeton University Press, 2009.

Chapter 3

Part II

3.1 Overview on Part II

This second part of the thesis consists of two papers, developed during a visiting period at the Department of Mathematics of the University of Aarhus (Denmark), under the supervision of Proff. Ole Barndorff-Nielsen and Jürgen Schmiegel, who is also co-author of both papers.

The object of this part is the statistical analysis of the fine structure of turbulence, thus assessing features displayed at small scales, exploiting the LDV high-frequency data. In particular, we construct a novel turbulent observable and we show that it can be used as a proxy of the turbulent velocity field. In order to do this, we prove that the time series of this new variable shear various stylised features with turbulent velocity time series. We provide a statistical description of the effects of intermittency and of the universal features of the observable over a large range of time scales. The main advantage of this variable is that it overcomes the problem of LDV measurements mentioned in Sec. 1.2, providing information at the maximum available sampling frequency. Moreover, we characterise the continuous-time dynamics of the energy budget of the system. This is done starting from the computation of the realised volatility of the velocity field, interpreted as the kinetic energy of the flow. In the framework of *ambit stochastics*, we show that this quantity evolves in time as a multiplicative cascade process, thus providing a powerful link between the statistical modeling of turbulence and the underlying physical process.

Particle arrival time and waiting time

Here, we still rely on datasets sampled in the VKE, in particular a panel of time series sampled in a flow with $Re \sim 10^5$. More in detail, we have 18 time series of the vertical component of the velocity field, sampled through the LDV technique for a total time $T = 3600$ s at an average rate of order ~ 1 kHz, so that the sample size of each time series ranges from around 1.5 to 2.5 million high-frequency data. This allows us to investigate the structure of turbulence at very small time scales.

As already mentioned in Sec. 1.2, the drawback of this kind of data acquisition technique is the fact that the LDV tool samples the local velocity when one of the passive tracer particles released in the fluid passes through the grid point. This prevents us from controlling the time step, which is also why the time series have different sample sizes. In order to extract the maximum possible information from this dataset, we have to overcome this problem, without filtering the time series, which could change the properties of the data. In order to do this, we decide to exploit the sequence of recording times of the LDV instrument instead of the velocity: the intuition is that, since the tracers follow the streamlines exactly, when a small, intermittent eddy passes through the recording point, the velocity spikes correspond to a high number of particles carried in a small time, thus resulting in frequent sampling and small time steps between subsequent measurements. This variable can be also seen as the sequence of arrival times of the tracer particles: in our procedure, we consider the first differences of this sequence at lags $k = 1, \cdots, K$, so that for each lag we can physically interpret them as the waiting times for the next k particles to arrive at the grid point.

We start from an exploratory analysis in order to assess the stationarity of such increments and then to check if they share some of the stylised features displayed by turbulent velocity time series. The increment sequences are stationary and all of them present long-memory traits, quantified by the Hurst exponent: its value is always $H > 1/2$ and changes in a smooth way with the lag. We also observe clearly non-Gaussian pdf's, with asymmetry and kurtosis changing with the lag. As a first step to characterise the shape of these distributions, we limit our analysis to the tail heaviness, quantified by fitting a stretched exponential $p(\tau) \propto \exp\{a\tau^b\}$ and estimating the stretching exponent b. In velocity increments, small values, corresponding to heavy tails, are usually observed, together with a trend to increase at

increasing lags and approaching the Gaussian limit $b = 2$. In this case, instead, we observe Gaussian tales at $k = 1$ in all the samples, and increasing values, corresponding to light tails, at increasing lags.

About this fact, we may argue that velocity spikes due to eddies are linked to particle clusters and so to small waiting times. On the other hand, the smallest possible time step that can be recorded by the instrument is limited and such small values may be not much smaller than the ones characterising the centre of the density, so that extreme values in the velocity may be correlated to almost average values in the waiting times. This problem still requires further investigation both from a statistical and physical point of view in order to be fully understood.

A more complete characterisation of the density of the increments is obtained through the NIG distribution. This is a four-parameter probability density function which can be thought of as a Normal distribution whose variance is allowed to vary randomly. This density admits both negative and positive asymmetry and non-Gaussian kurtosis, including the Normal distribution (and others) as a special case. For more details, see Appendix B in Paper F and references therein. We fit the NIG to all the increment sequences and we estimate the parameters through maximum likelihood, obtaining good fitting in all cases. However, while asymmetry and kurtosis usually vary in a smooth way with the lag for the velocity increments, the same behavior is not observed here.

The NIG thus provides a parsimonious model for this process, at least from a marginal point of view, at all lags. However, as already stated, the variation of the estimated NIG parameters is such that no common behavior is recognisable, making this description experiment-specific. On the other hand, the universal nature of turbulence has been empirically recognised at least since spectral (or variogram) estimations are available and theoretically predicted in K41, even if only approximatively and in the inertial range. Further attempts to extend universal scaling relations out of the inertial range while keeping into account intermittency effects include the concepts of Extended Self-Similarity (ESS) and its generalised version (GESS), see [11]-[14] and [34] for more detail and rigor.

Here, rather than finding universal scaling relations in the increment statistics, we assess universality in terms of a *time change*, or change of lag. Notice that each of our original series, and thus their increments, are sampled in different points of a strongly inhomogeneous and anisotropic flow, so that we can consider them as sampled in completely different ex-

periments. Loosely speaking, we look for some relationship that maps the lag sequence of one increment process to the lags of another, so that the distribution of the appropriately normalised samples at lags linked through the time change are identical. We show that such a universal structure can be successfully recovered, as already known for turbulent velocity and financial time series, over a very large range of lags.

Other than providing a detailed description of important features of the turbulence in the VKE, the main contribution of this analysis is that we show how the sequence of measurement times of the instrument can be exploited to obtain knowledge about the turbulence structure, overcoming the problem of the irregular sampling of the velocity without filtering the velocity time series. Future extensions of this work may include an effort to better understand the relationship between this observable and the velocity field. Moreover, a global observable can be built from these series, as the total number of measurements counted in a certain time step through all the points of the grid: such observable, which takes integer values, displays long memory and overdispersion, leading to the idea that might be itself a global turbulent observable, which we leave for future investigation, also given its analogy with integer-valued spread measures in finance.

Energy dynamics and ambit stochastics

In the last part of the thesis, we assess the continuous-time dynamics of the energy budget of the same flow considered above. In particular, we build a global positive definite observable from the original time series, as the normalised global realised volatility of the velocity field. Physically, it is interpreted as the normalised instantaneous kinetic energy of the whole flow (actually, the contribution from the sampled grid points), also known as turbulence intensity. As already mentioned, in turbulent flows the energy is propagated from large to small scales through a direct cascade: we look for fingerprints of such energy cascade by assuming a multiplicative cascade model in the framework of the ambit stochastics. This is a recently developed branch of stochastic processes (see, for example, [9] for recent advancements) that allows us, in general, to write a spatio-temporal stochastic process $\{Y_t(\mathbf{x})\}_{t\in\mathbb{R}}$ at point $\mathbf{x}$ and time t in the form

$$Y_t(\mathbf{x}) = \mu + \int_{A_t(\mathbf{x})} h(\mathbf{x},\xi,t,s)\sigma_s(\xi)L(\mathrm{d}\xi,\mathrm{d}s) + \int_{D_t(\mathbf{x})} q(\mathbf{x},\xi,t,s)a_s(\xi)\mathrm{d}\xi,\mathrm{d}s.$$

Here $A_t(\mathbf{x})$, $D_t(\mathbf{x})$ are *ambit sets*, containing information on the dependence structure and usually estimated through sample correlations; h, q are deterministic weight functions. If $A_t(x) = A + (t, x)$ they are convolutions; $\sigma_t \geq 0$, a_t are time-stationary stochastic fields, σ_t being the volatility field; L is a Levy basis.

It is clear that such framework is much more versatile than assuming an Ornstein-Uhlenbeck process for the continuous-time evolution of the observable. The present case admits some simplifications. Since our observable is positive definite, we can model its logarithm in this framework, so that the observable will be characterised as the exponential of a stochastic integral. In particular, we assume a multiplicative cascade model as the one proposed in [33] for the energy dissipation ε, which can be expressed as

$$\delta(t) = \exp\left\{\int_{A'(t)} Z(\mathrm{d}a)\right\}.$$

Such model is specified once the shape of the ambit set $A'(t)$ and the Lévy basis Z are determined. The first is estimated by the scaling in time of two-point multiplicative correlations and contains the information about the dependence structure of the process; once this is known, only the marginal distribution of the Lévy basis is required. We have no physical prescription for the distribution of $\ln \delta(t)$, while we know from K62 that the expected distribution for the natural logarithm of the dissipation is Gaussian. Notice that, if we could assume Gaussian components for the velocity field, we would have a prescription on the distribution of the intensity; since the Gaussian hypothesis is not satisfied by turbulence because of intermittency, we exploit once again the versatility of the NIG distribution.

We obtain a very good NIG fit for the marginal distribution of $\ln \delta(t)$, with an exponential left tail. On the other hand, the two point multiplicative correlations (also known as correlators) display a clear scaling in time, which can be approximated by a simple polynomial function of the lag: this function defines the shape of the ambit set. At this point, the model is estimated and the process can be simulated, if needed. However, a model validation is required, since our model is based on second order properties, while cascade processes leave precise fingerprints, also on higher-order properties. The first fingerprint we look for is the self-scaling of the two-point correlators of different orders, while the second one is the capability to predict the three-point correlator from the two-point ones. As shown in

the data analysis section of Paper F, both fingerprints are clearly displayed by the considered dataset, so that our model is well specified in such a way to provide an effective characterisation of the continuous-time dynamics of the energy of the flow.

Paper E

Statistical Analysis of a Close von Kármán Flow - Waiting Times.
Preprint

Statistical Analysis of a Close von Kármán Flow - Waiting Times

Flavio Pons
Jürgen Schmiegel

December 12, 2015

Abstract

In this paper we discuss the statistical analysis of an inhomogeneous and anisotropic turbulent flow. The analysis considers eighteen time series from the von Kármán Experiment, sampled using the Laser Doppler Velocimetry technique. First, we identify the recording times of the instrument as a proxy of the turbulent velocity field and we consider the log-increments of this process in analogy with the classical velocity increments. We show that this variable shares some stylized features with the velocity increments, which in turn often display many analogies with log-returns from financial markets. In particular, the increments at all lags are well fitted by densities encompassed by the Normal Inverse Gaussian distribution. Moreover, both the process and its increments at all scales display long memory-like behavior. On the other hand, while the tail heaviness also changes with the lag as in the case of turbulent velocity and financial markets, its value implies Gaussian to lighter tails, rather than heavy tails with a convergence towards Gaussianity at large scales. We discuss the universal nature of the evolution of the densities across time scales. We show that the densities of the log-increments of time series from different points of the inhomogeneous experiment, which are virtually different turbulent experiments, behave in a universal way in terms of a deterministic time change.

Contents

1 Introduction

Since the fundamental work by [18], the need for a statistical approach to the study of turbulence has become evident. The realization of a turbulent field, such as the fluid velocity or a passive tracer (e.g. the temperature), appears to be generated by a stochastic process: thus, the characterization of any variable sampled in a turbulent experiment requires a statistical approach.

In the following years, growing attention has been devoted to reach a better understanding of the intermittency of the turbulent velocity field, i.e. the presence of clusters of strong, high-frequency fluctuations around the mean velocity. This is one of the most peculiar features of turbulence, even though it is also known to appear in financial markets data, as stochastic volatility characterizing the time series of the log returns, as discussed, for example, in [6]. As described in [17], intermittency results in an approximate multifractal and universal scaling of the moments of the velocity increments, also known as structure function:

$$S_n(s) = \mathrm{E}[(v(t+s) - v(t))^n] \propto |t|^{\zeta(n)}, \tag{1}$$

where $v(t)$ is the main component of the three-dimensional velocity vector at time t, $\zeta(n)$ is the multifractal exponent, depending only on the moment order, s is a time increment falling in the inertial range. The latter is defined as the interval of time scales corresponding to frequencies such that the spectrum of the velocity field satisfies the scaling relation

$$E(k) \propto k^{-5/3}.$$

These properties hold in the limit of very large Reynolds number, defined as the non-dimensional quantity $Re = \frac{UL}{\nu}$, where U is the mean velocity, L is the linear length scale of the flow domain and ν is the kinematic viscosity of the fluid. Notice that a direct link exists between these two concepts: the higher Re, the larger the interval of time scales (or frequencies) characterized by the inertial range scaling law.

Besides the aforementioned multifractal scaling of the structure functions, we briefly recall some other stylized facts defining intermittency in turbulence.

- High-pass filtering can highlight intermittency in a time series of the turbulent velocity: if the filtering frequency corresponds to a time scale smaller than the upper limit of the dissipation range,

the filtered signal displays evident clusters of activity characterized by spikes.

- The flatness of the high-pass filtered turbulent velocity grows unboundedly with the filtering frequency.
- The local slopes of the high-order structure functions show deviations from the prediction of the celebrated 'K41 theory' presented in [18].
- The energy dissipation shows a very irregular behavior, with clusters of activity featured by high spikes.
- The probability density function (pdf) of the velocity increments $[v(t+s) - v(t)]$ changes with the lag s. In particular, the heaviness of the tails decays with the lag and, at the large scales, the distribution converges towards a Gaussian-like distribution.

The last phenomenon has been addressed in detail in [4], where the authors propose a more complete description of the distributions of the velocity increments at all available time scales through the Normal Inverse Gaussian law (NIG).

The scaling relation (1) is said to be *universal* because, once the Reynolds number is high enough to let the inertial range appear, the exponents $\zeta(n)$ are universal, with no dependence on the experimental setup. In the past decades, many efforts have been devoted to extend tise concept of universality to smaller Reynolds numbers and to a set of time scales wider than the inertial range. These include the hierarchical models introduced by [20] (She-Leveque Hierarchical Structure, SLHS), the concept of Extended Self-Similarity (ESS) proposed by [7] and [8] and its generalization (GESS) by [9] and [10].

More recently, in [5] ([2],[3],[6]) it is shown that SLHS and GESS can be considered as the consequence of a more general concept termed Stochastic Equivalence Class (SEC), relating the density of velocity increments at different scales and from different experiments by applying a change of scale. The proposed change of scale results in a collapse of the density of velocity increments onto a universal evolution across scales. This result complements the fact that the pdf of velocity increments can be well approximated within the class of NIG distributions at all time scales and all amplitudes, and for a wide range of experimental set-ups: NIG distributions provide a parametric description of the density of velocity increments and the time change describes the relationship among different experiments. This framework can

be applied to detect universality in time series sampled across different experiments. The purpose of the present paper is to investigate the potential of SEC and the corresponding NIG analysis to describe the statistics of the flow sampled at different points of a single non-homogeneous and anisotropic experiment. Such a situation is given in the close von Kármán flow produced in the von Kármán Experiment (VKE), as described by [19]. In this paper we consider a realization of the VKE in which the velocity is sampled using Laser Doppler Velocimetry (LDV), a technique that allows to obtain high frequency data on a spatial grid, but not to control the time step. In [16] many features of the VKE turbulent flow have been investigated by applying linear time series analysis and the estimation of the Hurst exponent. This procedure was feasible because the data were collected using Stereoscopic Particle Image Velocimetry (SPIV), which produces regularly spaced time series for the three components of the velocity vector. On the other hand, LDV allows for much larger sampling frequencies. This provides information about the small scales of turbulence and makes robust estimation possible, but the resulting time series are not regularly spaced.

To account for the problem of the uneven spacing of the time series, we consider the LDV sampling time instead of the velocity. The increments of this variable at lag k can be interpreted as the waiting times for the next k measures to be performed. This variable is not traditionally exploited in turbulence. We argue that it can be linked to the flow structure, since each measurement corresponds to a disturbance produced by a passive tracer when it interacts with the LDV system. Since passive tracers follow the streamlines with no deviations due to inertia, the series of times between subsequent measurements are directly linked to the distribution of the particles in the flow.

In this paper we show that, besides providing a statistical description of the flow in the VKE, waiting times have many features in common with the velocity sampled in a turbulent flow, letting us exploit the LDV information in a more complete manner. In particular, we show that waiting times display long-range dependence and heavy tails that vary with the lag. We also address, in detail, the evolution of the densities of waiting times across lags from the perspective of NIG distributions. Based on that we provide evidence of the appropriateness of SEC to directly relate the statistics of the LDV measurements at different spatial locations.

The paper is structured as follows. In Section 2 we describe the

experimental setup, the sampling technique and the resulting dataset. We then properly define the waiting times and perform an explorative analysis. In Section 4 two main stylized facts about intermittency are recalled, with focus on the behavior of the increments: the long memory and the tail heaviness. As a measure of the long memory we compute the Hurst exponent, while the tail heaviness is quantified by stretching exponents. Differences and analogies between waiting times and turbulent velocity measurements are discussed. In Section 5 we briefly introduce NIG distributions and analyse the densities of waiting times with this flexible class of laws. Finally, we apply the SEC framework and compare the results to the corresponding finding for turbulent velocity. Section 6 contains a summary of the results and the conclusions.

2 Description of the data

We analyze eighteen datasets, consisting of recording times and one-point measurements of the vertical component of the fluid velocity field produced in the von Kármán Experiment (VKE), designed and performed at the Commissariat Ã l'Énergie Atomique (CEA) in Paris. A detailed review of the experiment can be found, for example, in [19], [15] and [14]. The main feature of the VKE framework is its capability to produce very high Reynolds numbers ($Re \sim 10^6$) with a compact experimental setup.

The fluid consists of a solution of water and glycerol, making it possible to modulate the viscosity, seeded with passive particles, so that the measurements can be performed through Laser Doppler Velocimetry (LDV). Measurements of the velocity field are obtained through a He-Ne Flowlite Laser with $\lambda = 632.8$ nm. Either the axial or the azimuthal component of the velocity is directly measured; the radial component is recovered using $\nabla \cdot \mathbf{v} = 0$, which is legitimate by the fact that the time-averaged field is axisymmetric and solenoidal. The data consist of measurements of the field on a 11×17 points grid, with an average data rate of about 0.5 kHz; however, only 18 time series are retained discarding spurious data and boundary effects.

The Reynolds number of the considered experiment is $Re = 10^5$; the fluid is water kept at a temperature $\theta = 20$ Â°C, with corresponding kinematic viscosity $\nu = 1.0 \cdot 10^{-6}\,\mathrm{m^2 s^{-1}}$.

The total measuring time is fixed at $T = 3600$ s, but the number of

measurements in each time series depends on the number of particles passing athrough the corresponding measuring point during the experiment. This implies that the length of the time series is not fixed, as shown in table 1.

Series	Sample size	Average time-step (ms)
1	2462311	1.462039
2	2181160	1.650496
3	2143198	1.679727
4	2443088	1.473544
5	2006618	1.794062
6	2292121	1.570594
7	1994104	1.805317
8	2148151	1.675859
9	2119348	1.698633
10	2060018	1.747554
11	1644393	2.189256
12	1885173	1.909638
13	1978781	1.819298
14	2116351	1.701036
15	1490146	2.415867
16	1435270	2.508236
17	1736179	2.073516
18	1633155	2.204321

Table 1: Sample size at each measuring point (labelled by time series number) and corresponding average time-step in milliseconds.

Also, and most important, the device can perform a measurement only when a particle passes through a grid point, so that the time step cannot be controlled.

Let N be the length of the time series, expressed in number of measurements. At each point we obtain two series, t_j and v_{t_j}: v_{t_j} is the vertical component of the velocity vector, indexed by the measuring times t_j with $j = 1, \ldots, N$ and $t_N = T$. It is clear that v_{t_j} are irregularly spaced time series; on the other hand, t_j is a regular sequence, since j runs on the positive integers. We aim to study the behavior of this variable. The sequence $\{t_j\}_{j \in \mathbb{N}^+}$ is increasing and can be written

as

$$t_j = \mu_j + t'_j$$

where $\mu_j = j\Delta t$, $\Delta t = T/N$ is the average time-step and t' is the fluctuating part of the process. The term μ_j can be easily removed to obtain a series of realizations of the stochastic disturbances $\{t'_j\}$. The resulting series are presented in Fig. (1) and Fig. (2), where in each panel the realization of $\{t'_j\}$ (in seconds) is plotted against the normalized index j/N for each of the measuring points. For a first characterization, we apply basic time series analysis methods, as described, for example, in [12] and [13].

An augmented Dickey-Fuller test has been performed to test the unit root null hypothesis, resulting in a non-rejection with p-values ranging from 0.56 to 0.61 for all the time series. After first-differencing, the augmented Dickey-Fuller test is always significant with all the p-values smaller than 0.01: we may assume stationarity of the differenced series.

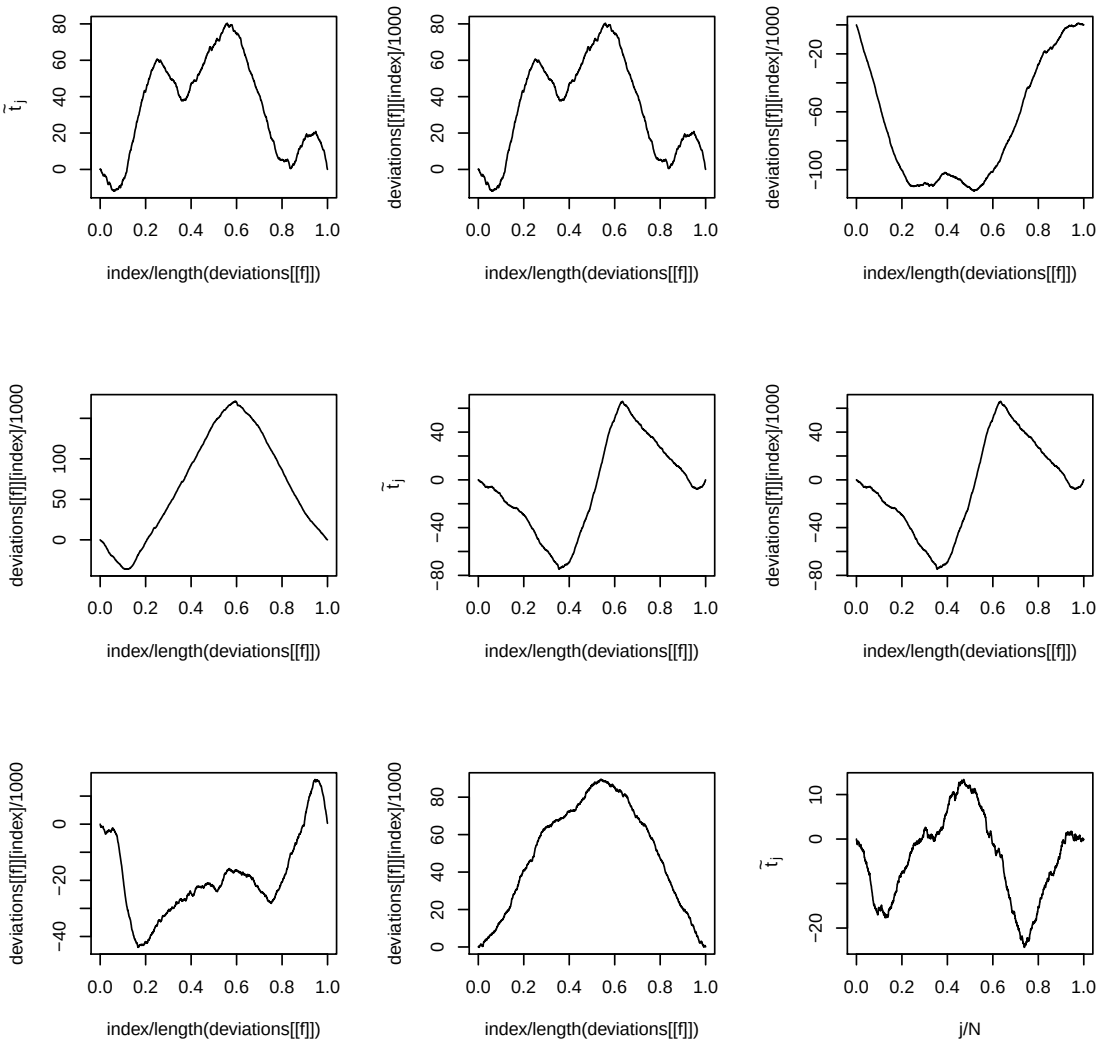


Figure 1: Disturbances $\{t'_j\}$, series 1-9.

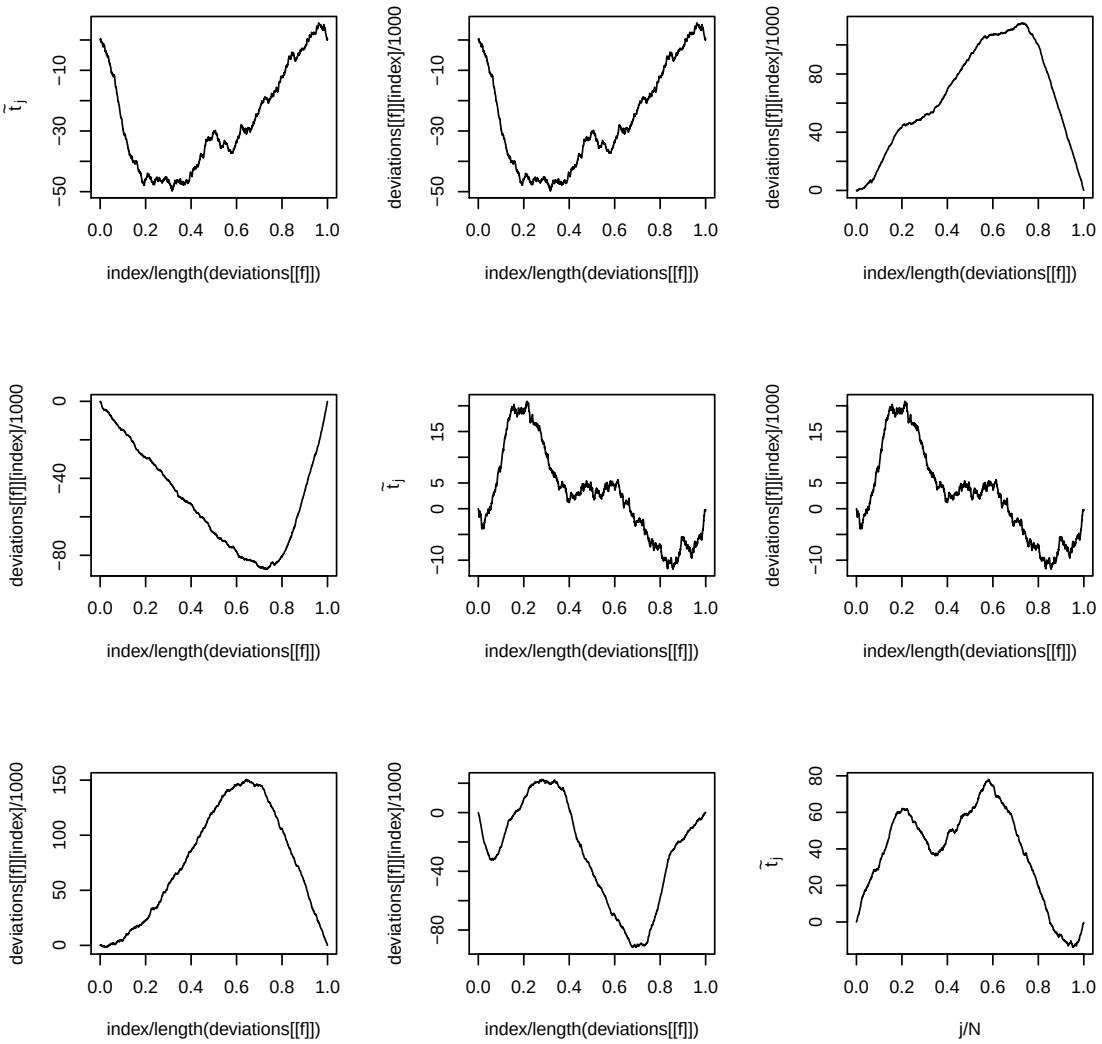


Figure 2: Disturbances $\{t'_j\}$, series 10-18.

3 The increment process: waiting times

We consider the sequence of measuring times $\{t_j\}$, from which we define the k-lag first order differences, $k \in \mathbb{N}$,

$$\Delta t_j^k = t_{j+k} - t_j \tag{2}$$

To cover a wide range of time scales, we chose the set of lags $k \in \mathcal{K} = \{1, 10, 20, \ldots, 1200\}$. For $k = 1$, $\{\Delta t^1\}_j$ is the process describing the waiting time for the next particle to pass through the measuring point; so, for each k, $\{\Delta t^k\}_j$ is the sequence of times in which the next k particles will pass through the given grid point.

Let us consider how the new variable Δt^1 brings information about the turbulent field. The intuition is that, since the particles are passive tracers, not only their velocity will be equal to the local fluid velocity, but also their spatial distribution will follow the streamlines, marking the local structure of the flow. Consider, as an example, the case of a laminar flow seeded with well-mixed passive particles: we would expect a process such that the deviations from the mean of Δt^1 (and thus from the average time-step) would be white noise. For the available dataset, the null hypothesis of white noise for Δt^1 is rejected in a Ljung-Box test with p-value $< 2.2 \cdot 10^{-16}$ for all the series.

Accordingly, the inspection of the global and partial autocorrelation functions (hereinafter ACF and PACF, respectively) suggests the presence of a serial dependence in the datasets. In Fig. 3 both autocorrelation functions are shown for series 17; a similar behavior is found in all the series. The values of the correlations are low, but still significantly different from zero, since the confidence bands are very thin, due to the large sample size. This is true for a wide range of lags, suggesting the presence of long range dependence. This idea is confirmed by the estimated values of the Hurst exponent of all the eighteen series, ranging from 0.56 to 0.62 as shown in Table 2. The exponents have been computed according to the Whittle MLE estimator (see, for instance, Section 6.1 of [11].

Series	Hurst exponent
1	0.5978067
2	0.6238351
3	0.6188352
4	0.6060019
5	0.5914473
6	0.6059191
7	0.6012453
8	0.5897616
9	0.5823103
10	0.5856560
11	0.6106381
12	0.6116846
13	0.5906467
14	0.5771938
15	0.5963672
16	0.5931036
17	0.5733799
18	0.5649695

Table 2: Hurst exponent.

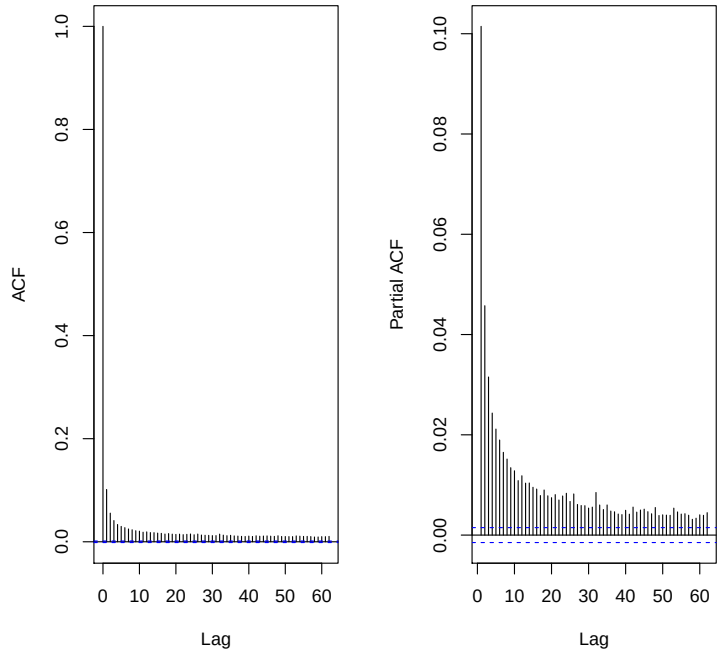


Figure 3: ACF and PACF of Δt^1, series 17.

For a fixed lag k the increments can be defined for overlapping or non-overlapping observation windows. In the overlapping case, the simple first-order difference of the series is computed, as in Equation 2. For non-overlapping increments, the differences are obtained from

$$\widetilde{\Delta t}_j^k = t_{jk+1} - t_{(j-1)k+1}, \quad j \in 1, \ldots, L,\ L = \lfloor N/k \rfloor - 1. \tag{3}$$

Whatever the definition adopted, the variable of interest in our analysis will be the natural logarithm of this quantity

$$\tau_j^k = \log(\Delta t_j^k) \tag{4}$$

and centered by subtracting the sample mean from each series.

Note that non-overlapping increments series are much shorter.

On the other hand, overlapping increments introduce spurious serial correlations in the datasets. This can be easily observed comparing

the ACF and the PACF in the two cases. In Fig. 4 we show this for the increments at lag 10 for series 17. It is clear that overlapping increments define a process with a much stronger dependence, evident from the smoothness of the series, from the linear decay of the ACF and the high-valued spikes in the PACF with period equal to 90 lags and starting from 1 at lag 1. On the contrary, the plot for the non-overlapping increments suggests at least weak stationarity for the series, the ACF displays long memory but no random walk-like persistence and the PACF rapidly decays.

The main part of our analysis, presented in Section 5, aims to fit a proper distribution to the pdf of the increments and to compare the fits obtained at different lags at different points. To prepare for that, we run a two-sided Kolmogorov-Smirnov test under the null hypothesis of identical distribution of overlapping and non-overlapping increments for all the series and all $k \in \mathcal{K}$. For all datasets and lags, the null hypothesis is not rejected with p-values always larger than 0.2. Thus, we may choose overlapping or non-overlapping increments for the density estimation.

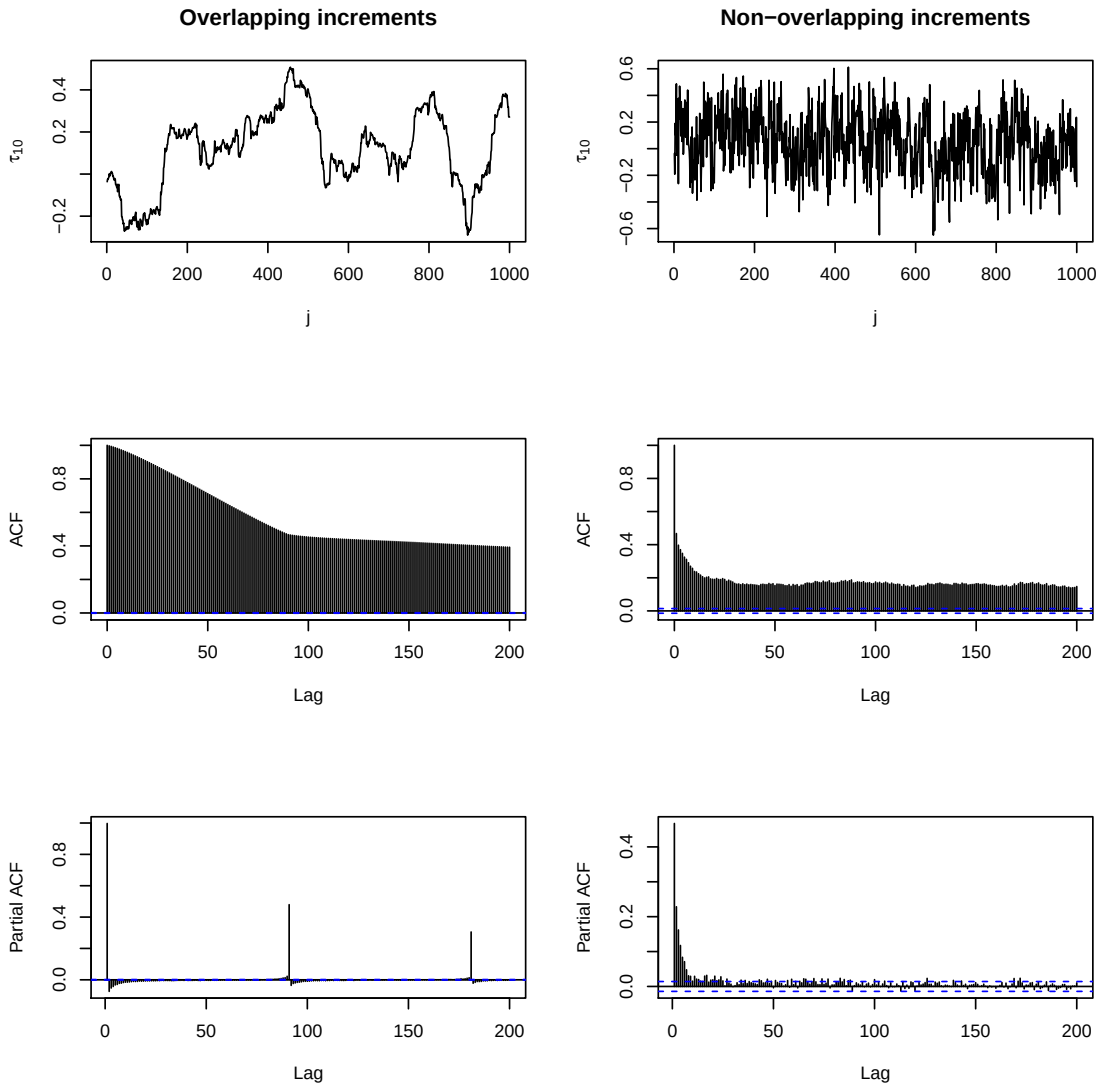


Figure 4: Comparison between overlapping and non-overlapping increments at lag 10 for series 17. Plot of the first 1000 data (top panels), ACF up to lag 200 (central panels) and PACF up to lag 200 (lower panels).

4 Preliminary analysis: long memory and tail heaviness

For a preliminary characterization of the evolution of the densities of increments across lags we now consider, in more detail, long memory and heaviness of the tails. We expect an analogy to what is observed for velocity increments (see [6] and references therein), where the densities evolve from heavy tails at small scales/lags towards an approximate Gaussian shape at larger scales/lags. This behavior is sometimes called *aggregational Gaussianity* in the literature.

4.1 Long memory of the increments

We have already pointed out that Δt^1 is a long memory process, with Hurst exponent larger than 0.57 for all time series. We investigate the behavior of the Hurst exponent of the non-overlapping increments in more detail for $k \in \mathcal{K}$. As in the previous section, the Hurst exponent is computed via Whittle MLE estimation, as in [11]. Results are shown in Fig. (5).

While the values of the Hurst exponent are quite close to 0.6 for the increments of all time series at lag 1, as already shown in Table (2), the values at increasing time scales show specific features depending on the sampling point in space. Three types of specific behavior can be observed. For some of the datasets (1,2,3,4,5,7,9,11,12) the value of H steadily increases with the lag. Most of the other datasets show a plateau or a slight decrease after a steep ascent at small lags. Dataset 18 is an exception, reaching a high peak around lag 300 and then rapidly decreasing. As a synthetic measure of these discrepancies, we may compute the variance of H among the datasets, σ_H^2. As shown in Fig. (6), this quantity varies smoothly as a function of the lag, starting from a value one order of magnitude smaller than the ones resulting for larger lags.

The different behavior of the Hurst exponent with the lag for different time series might be linked to the inhomogeneous and anisotropic nature of the flow: [16] show that the estimated Hurst exponent for the SPIV velocity time series traces efficiently the local features of the flow.

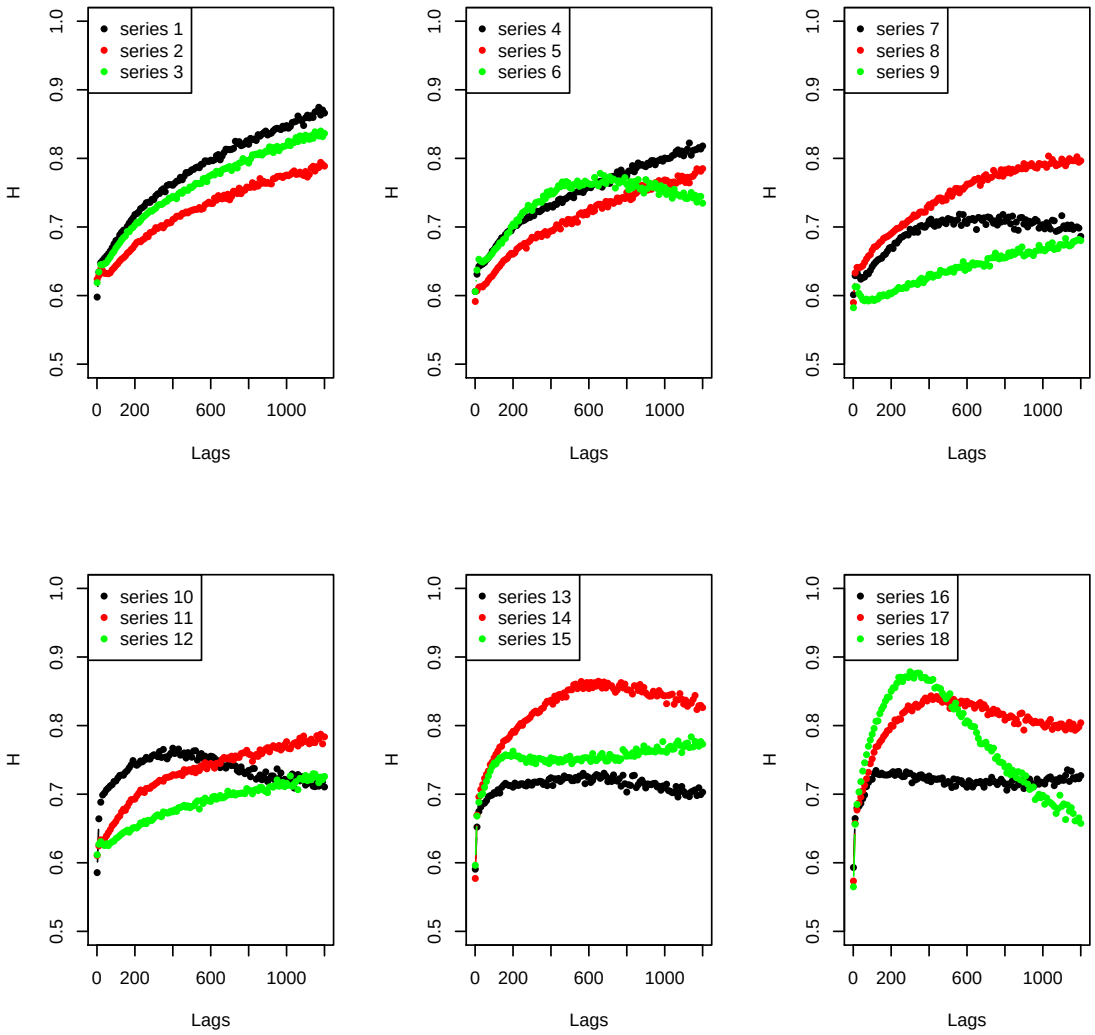


Figure 5: Estimated Hurst exponent for the increments of each time series as a function of lag.

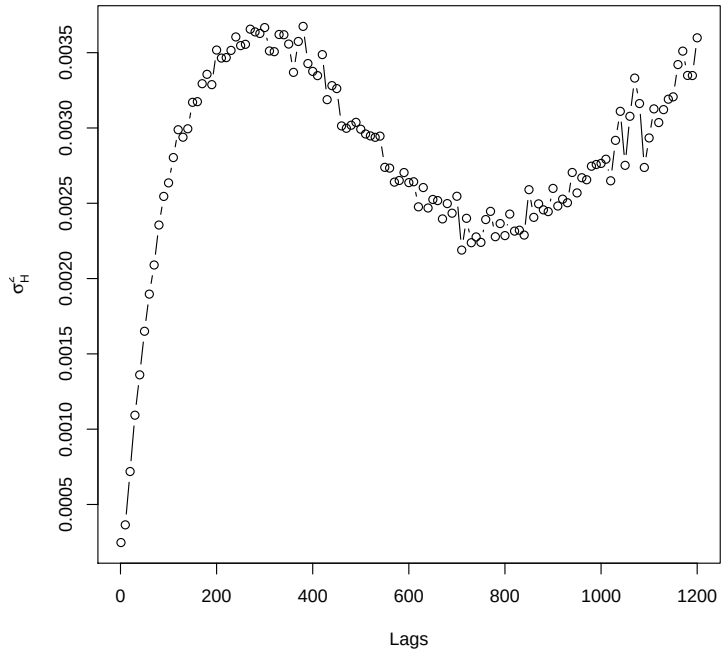


Figure 6: Variance of H among the dataset, σ_H^2, as a function of lag.

4.2 Stretched exponential tails

Besides the long memory of the process, also the shape of the distribution of the log-waiting times is expected to change with the lag, if this variable has traits in common with the turbulent velocity. We quantify this change by fitting a stretched exponential

$$p(\tau) \propto \exp\{a\tau^b\} \tag{5}$$

to the tails of the distributions and consider the value of the stretching exponent b. Here we consider large amplitudes that are more extreme than 10% of the minimum/maximum value of the corresponding tail. Classical findings for velocity increments (see [6] and references therein) show values of the exponent growing as a function of the lag, approaching an approximate Gaussian limit $b \approx 2$ at large time scales.

We perform the estimation of the stretching exponent b both on the left and right tail for large amplitudes, considering values more extreme than, respectively, the 0.1 and 0.9 percentile. Results are summarized in Fig. (7) and (8) for the left and right tail, respectively.

As for velocity increments, the stretching exponent is an increasing function of the lag. On the other hand, the values are systematically larger; in particular, at lag 1 the value is always very close to $b = 2$, so that the transition is from Gaussian to lighter tails, rather than from heavy to Gaussian ones.

It is worth to mention that the fits for the right tail result in relatively small values of the residual sum of squares (RSS) for all the series at all the lags, while for the left tail we have values increasing with the time scale, in particular for series 9 and 18. More in detail, the overall range of RSS is $[2.37 \cdot 10^{-5}; 3.13]$ for the right tail and $[2.1 \cdot 10^{-4}; 41]$ for the left tail; the maximum values for the left tail of datasets 9 and 18 are, respectively, 36.5 and 41, even though values larger than 10 result also in other series at moderate and large lags.

Thus, an analysis based on the tail heaviness results in a smooth transition from Gaussian to very fastly decaying tails. This fact is in contrast with the results obtained in the past for the velocity increments: while the heaviness of the tails still decreases with the time scale, for the turbulent velocity the Gaussian constitute the upper, rather than the lower limit of the tail evolution.

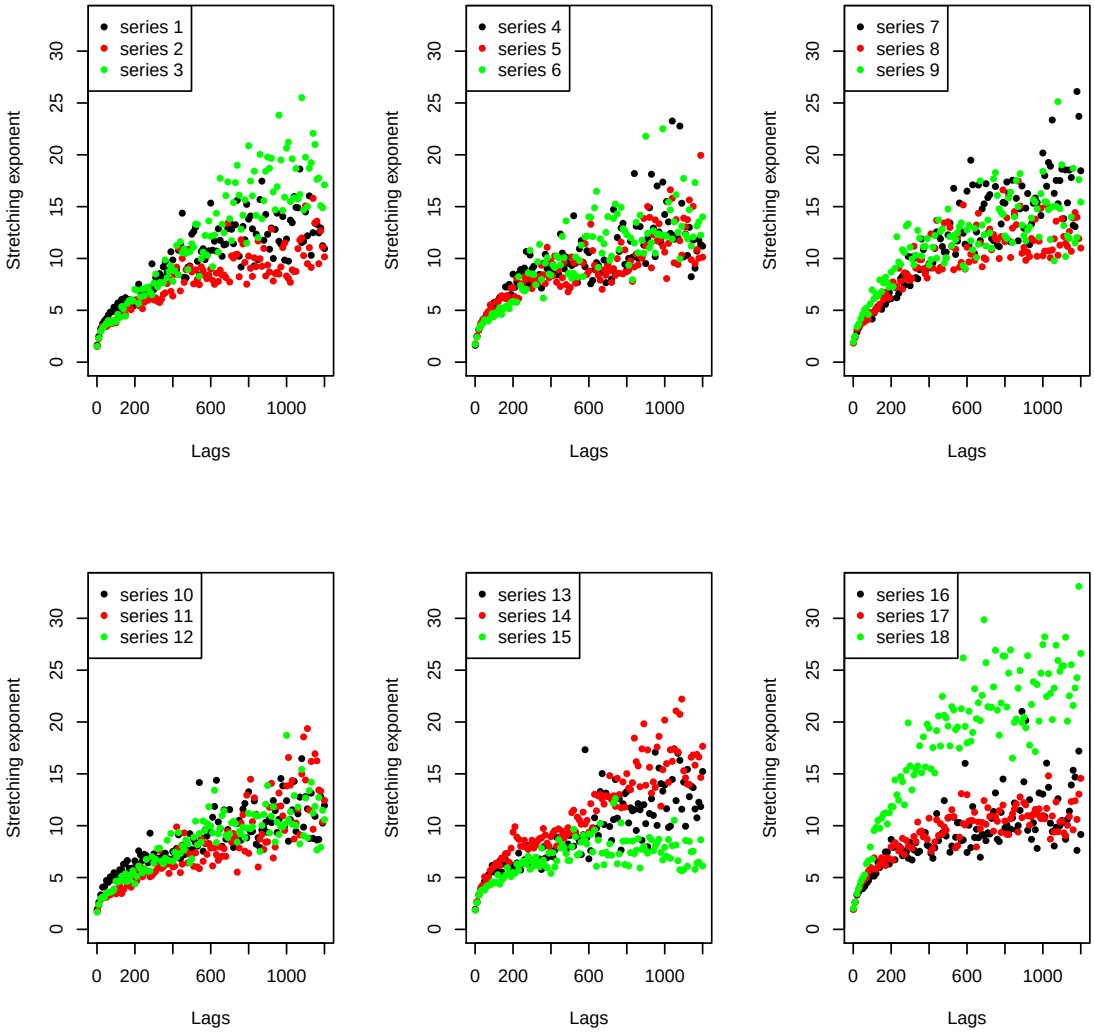


Figure 7: Estimated stretching exponent for the left tail of the distribution of the increments of each time series as a function of lag.

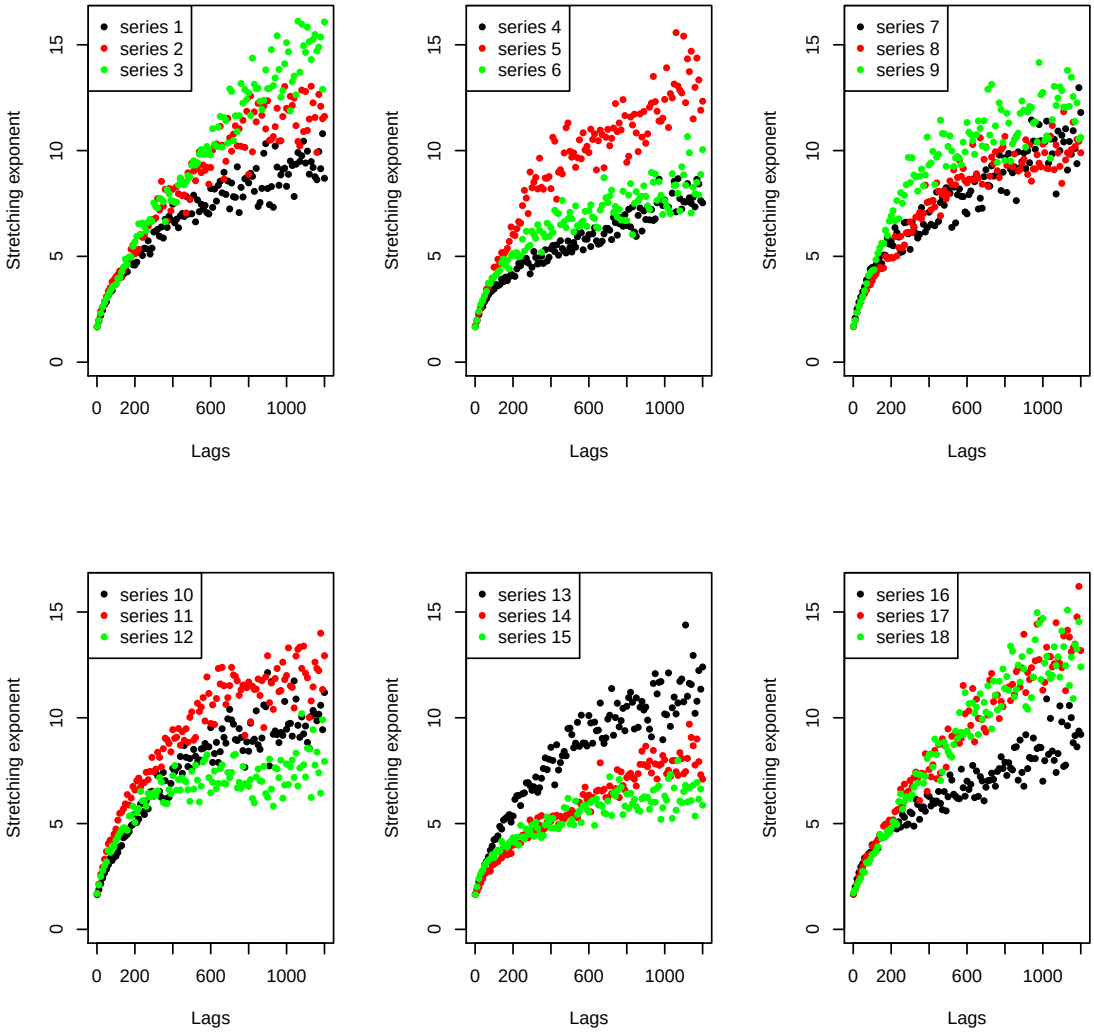


Figure 8: Estimated stretching exponent for the right tail of the distribution of the increments of each time series as a function of lag.

5 Parsimonious and universal description of the waiting times

In section 4 we gave a preliminary description of the behavior of the waiting times addressing the long memory and the tail heaviness.

In this section, after a brief introduction to the NIG distribution, we further investigate the distributions of waiting times. In particular, we address the universal features connecting the different spatial locations.

5.1 Normal Inverse Gaussian Distribution

The NIG distribution was introduced in [1]. This flexible four-parameter family of distributions in the real line is given as

$$p(x;\alpha,\beta,\mu,\sigma) = a(\alpha,\beta,\mu,\sigma)q\left(\frac{x-\mu}{\sigma}\right)^{-1} K_1\left\{\sigma\alpha q\left(\frac{x-\mu}{\sigma}\right)\right\} e^{\beta x}, \tag{6}$$

where α is a shape parameter, β a skewness parameter, μ is a location parameter and σ is a scale parameter. Here,

$$q(x) = \sqrt{1+x^2}$$
$$a(\alpha,\beta,\mu,\sigma) = \pi^{-1}\alpha \exp\left\{\sigma\sqrt{\alpha^2-\beta^2} - \beta\mu\right\}$$

and K_1 is the modified Bessel function of the third kind and index 1. We have that $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $0 \leq |\beta| < \alpha$.

The shape of the NIG distribution can be described by the symmetry χ and the steepness ξ, defined as

$$\chi = \rho\xi$$
$$\xi = (1-\overline{\gamma})^{-1/2}$$

where $\rho = \beta/\alpha$ and $\overline{\gamma} = \sigma\gamma = \sigma\sqrt{\alpha^2-\beta^2}$. By construction, $\chi \in (-1,1)$ and $\xi \in (0,1)$. Moreover, the equation $\xi = |\chi|$ defines a triangle in the rectangular interval $(-1,1)\times(0,1)$ which contains all the possible NIG laws.The lower edge of the triangle corresponds to the Normal distribution.

The flexibility of the pdf (6) admits asymmetric distributions (parameter χ)and heavy, semi-heavy and Gaussian tails (parameter ξ). This feature makes it an effective tool to describe turbulent velocity increments, which usually show a transition from heavy to Gaussian tails, when moving from the smallest to the largest time scale.

5.2 NIG fits

The typical behavior of the log-histogram of waiting times across lags is shown in Fig. 9, for dataset number 7 at lags $k = 1, 10, 20, 100, 200, 500, 800, 1000, 1200$. Most of the other spatial locations show a similar behavior. We summarize the main features as follows.

- All datasets display a negative asymmetry at lag $k = 1$, which tends to decrease for increasing lags.
- Three of the datasets, corresponding to measuring points number 4, 14 and 18, manifest a range of lags in which the distribution is positively skewed (e.g., see Fig. 10 for the shape triangle of series 4).
- Very large lags ($k = 5000, 10000$) have been additionally tested, still not leading to normality of the distributions. The null hypothesis of normality has been tested through a Pearson's chi-squared test, leading to rejection at all levels, with p-values virtually equal to 0 for all the datasets.
- The fits appear to be good at all lags; for lags of order $k \simeq 10$ some outliers appear in the left tail of the distribution. It is however worth to notice that the value of the log-histogram, in these cases, is never higher than -7.

These results confirm that a representation of the log-waiting times distribution by the NIG law is adequate.The corresponding representation within the NIG shape triangle is exemplified in Fig. (10) - (11). It appears that, for increasing lags, the distribution of the log-waiting times tends to depart from normality. This feature, opposed to what might be expected from the law of large numbers, reflects the long memory of the process. Further considerations on the distribution of the log-waiting times, in particular on the nature of the right tail of the distribution of large amplitude data, will be presented in the next Subsection.

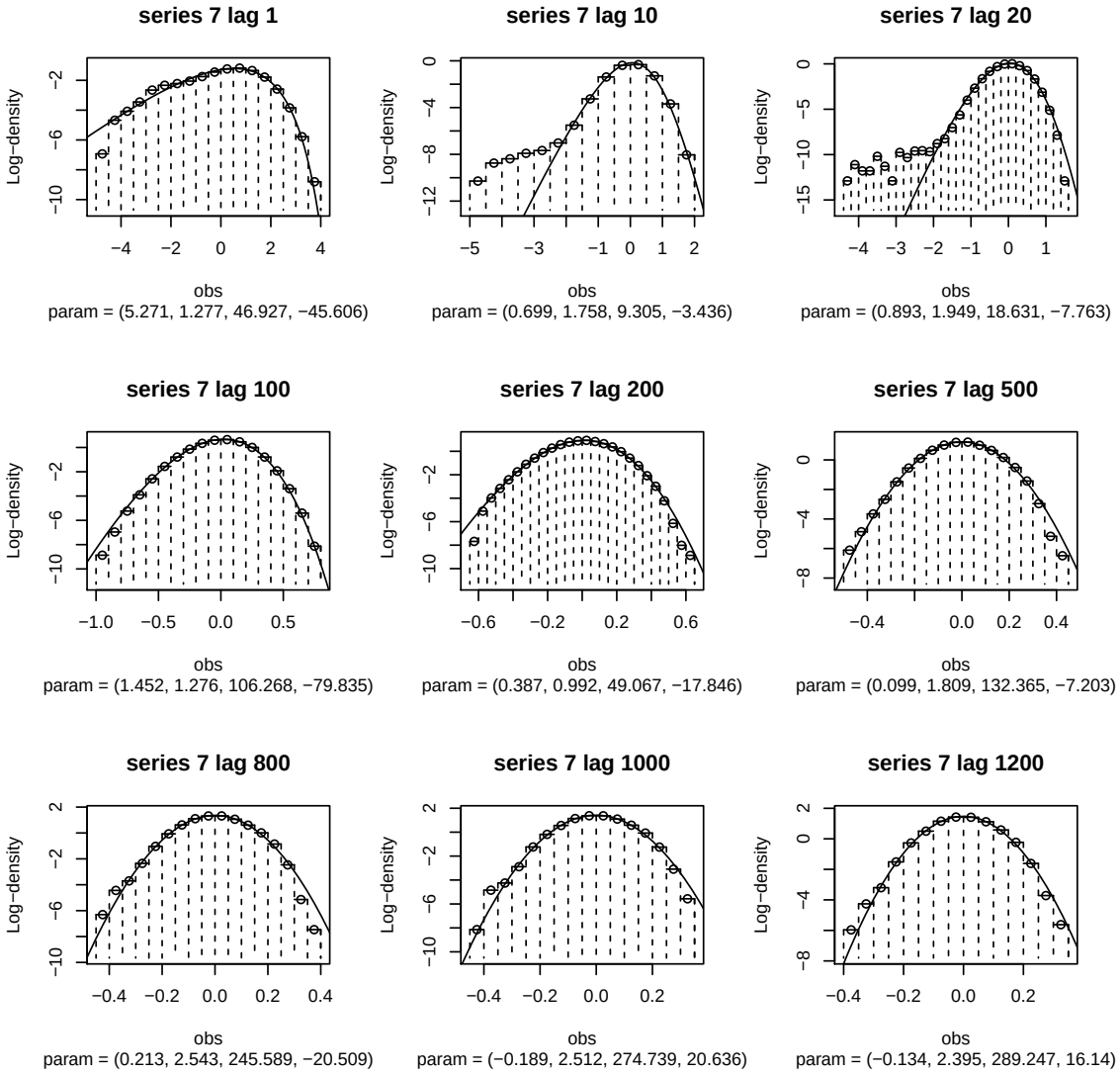


Figure 9: Log-histograms and superimposed NIG fits for series 7 at lags $k =$ 1, 10, 20, 100, 200, 500, 800, 1000, 1200.

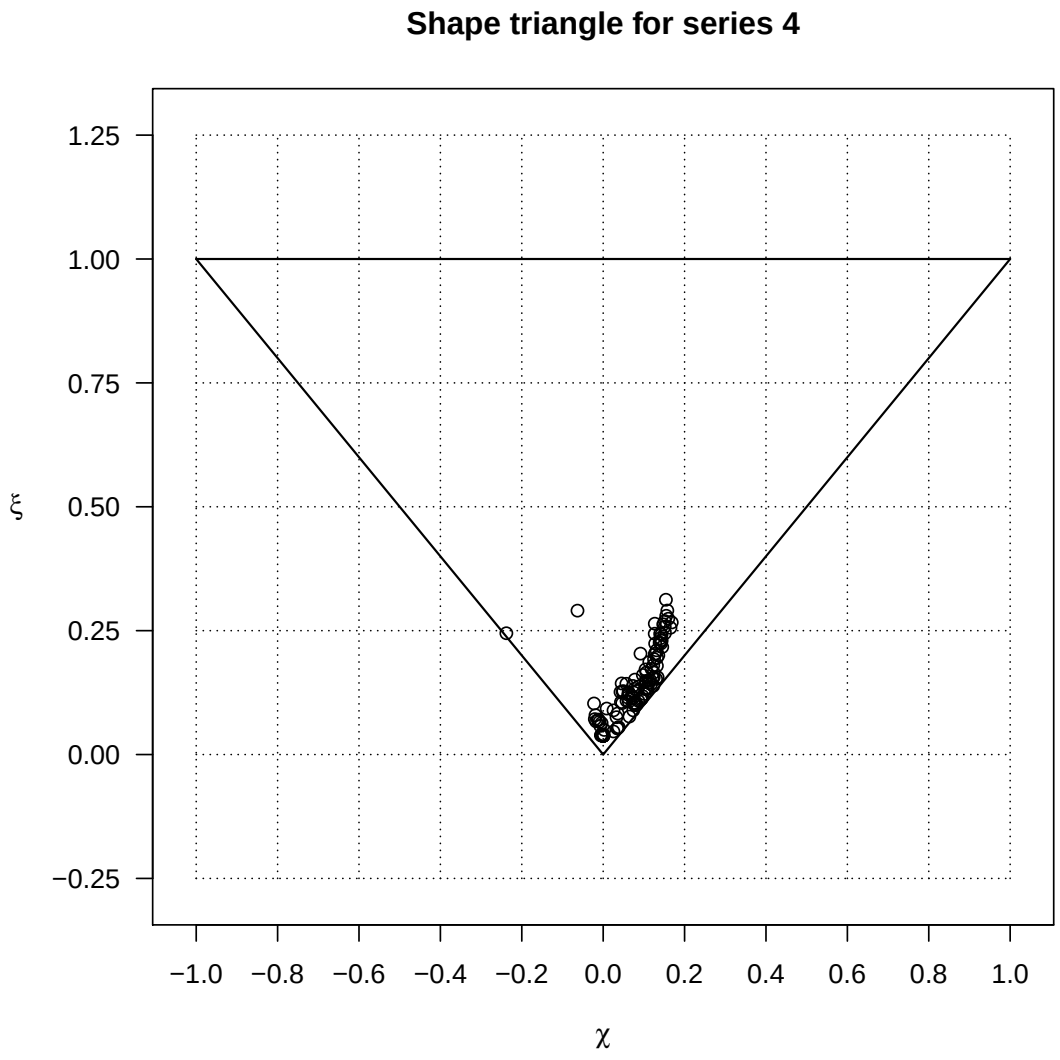


Figure 10: NIG shape triangle for series 4, clearly showing the transition across the half of the triangle characterized by positive asymmetry.

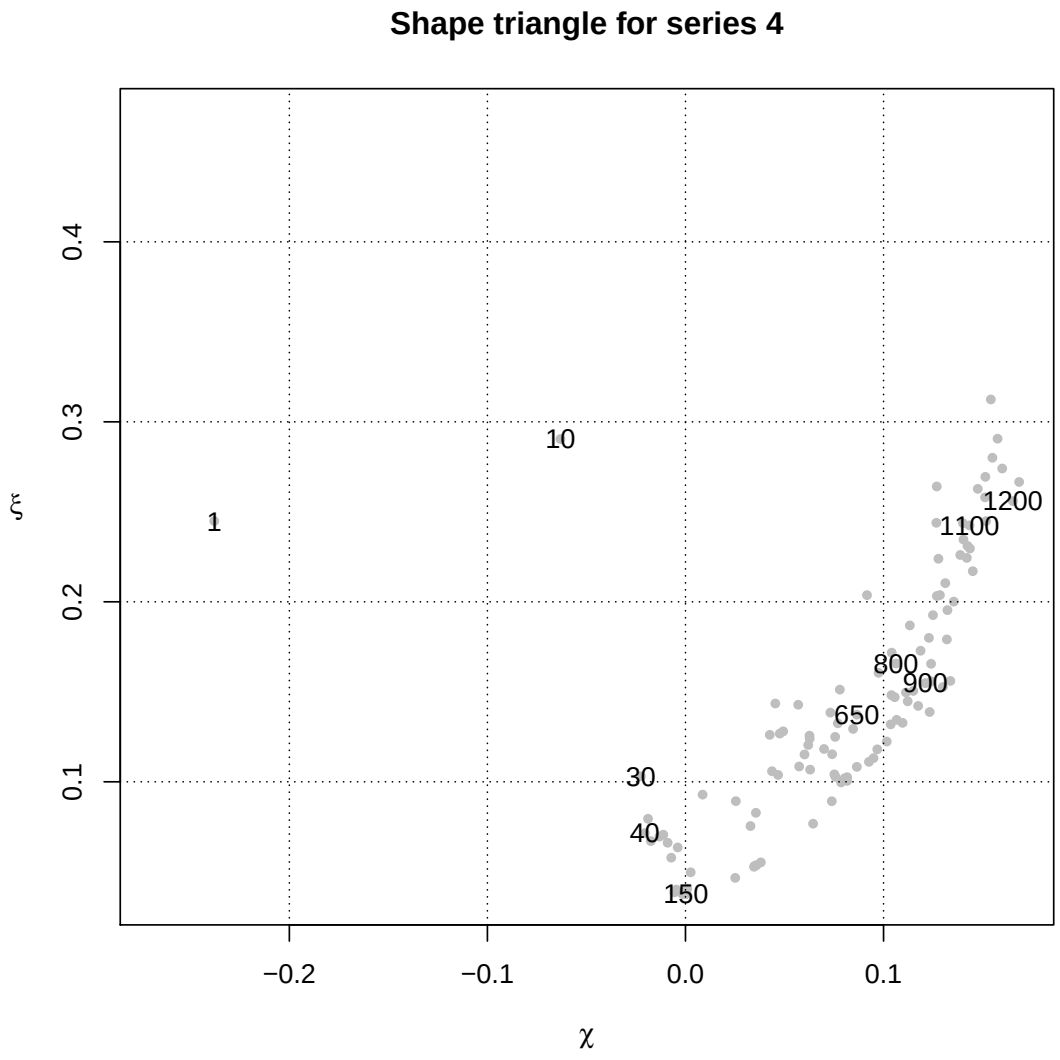


Figure 11: NIG shape triangle for series 4, clearly showing the transition across the half of the triangle characterized by positive asymmetry.

5.3 Universality and Change of Lag

An important and interesting issue is the question whether the evolution across lags at different spatial locations can be modeled by selecting different lags at different spatial locations. Such a problem has been addressed in [2,3,6] for distributions of velocity increments. Here, we follow this procedure.

5.3.1 Universality and the Stochastic Equivalence Class

Let $u^{(i)}(k) = \tau^k_{j,(i)}$ denote the waiting times at lag k and at spatial location $i = 1, 2, \cdots, 18$. The key idea exploited in [2,3,6] is the question of the existence of deterministic functions $F^{(i)}$ and $g^{(i)}$ for each spatial location such that the following Stochastic Equivalence Class (SEC) is fulfilled

$$\frac{u^{(i)}(k^{(i)})}{g^{(i)}(k^{(i)})} \stackrel{d}{=} \frac{u^{(j)}(k^{(j)})}{g^{(j)}(k^{(j)})} \iff F^{(i)}(k^{(i)}) = F^{(j)}(k^{(j)}). \tag{7}$$

Here we assume that $F^{(i)}$ is monotonic. The existence of such a SEC then implies a change of lags relating different spatial locations $k^{(i)} = (F^{(i)})^{-1}(F^{(j)}(k^{(j)}))$. It is this change of scales that results in a collapse of the corresponding densities. Provided that such a SEC exists, one can estimate the corresponding change of lags by using the procedure suggested in [2,3,6]:

1. Choose a reference dataset (j), in this case series 2.
2. Compare $\overline{c}_4^{(i)}(k)$ and $\overline{c}_4^{(j)}(k)$ with $\overline{c}_4 = c_4/(c_2)^2$ and $k \in \mathcal{K}$.
3. Choose, for each series, the couples (k_1, k_2) such that $\overline{c}_4^{(i)}(k^{(i)}) = \overline{c}_4^{(j)}(k^{(j)})$.
4. In addition, we choose: $g^{(j)}(k) = 1$ and $g^{(i)}(k^{(i)}) = \left\{ \frac{c_2^{(i)}(k^{(i)})}{c_2^{(j)}(k^{(j)})} \right\}^{\frac{1}{2}}$.
5. Kolmogorov-Smirnov test with $\alpha = 0.01$ for equality in distribution of the scaled log-waiting times, as in the l.h.s. of Eq. 7.

According to these rules, we use the standardized fourth order cumulants as the functions F and the second order cumulants as the deterministic functions g to rescale the variable. Two remarks are in order. A strict equality for the time change functions, as required

at point (3), can hardly be found. It can be replaced with a looser argument, based, for example, on the relative error:

$$\Delta_{ij} = \frac{|\overline{c}_4^{(i)}(k^{(i)}) - \overline{c}_4^{(j)}(k^{(j)})|}{\frac{|\overline{c}_4^{(i)}(k^{(i)})| + |\overline{c}_4^{(j)}(k^{(j)})|}{2}}. \tag{8}$$

After running some different test analyses, a conservative condition $\Delta_{ij} < 0.2$ seemed adequate to our case. In fact, the condition could be more strict, but it appeared that this does not affect the conclusions presented below.

5.3.2 Results

The behavior of $\overline{c}_4^{(i)}(k)$ is monotonically decreasing for all the series. The ranges of values can be very different from one dataset to another, as an effect of the inhomogeneity of the VKE. Fig. 12 shows $\overline{c}_4^{(i)}(s)$ for some datasets, compared to $\overline{c}_4^{(2)}(s)$ for the reference dataset.

Analysing all spatial locations we formed 37837 possible combinations that satisfie the condition (8).

All these combinations have been tested for equality in distribution through a two-sided Kolmogorov-Smirnov test at a level $\alpha = 0.01$. Let us recall that, given two samples of size m and n, with empirical cumulative distribution functions F_m and G_n respectively, the Kolmogorov-Smirnov test statistic is defined as:

$$D_{mn} = \left(\frac{mn}{m+n}\right)^{\frac{1}{2}} \sup_x |F_m(x) - G_n(x)|.$$

It is clear from this equation that, when $mn \gg m+n$, i.e. when both samples are very large, the test statistic assumes high values, so that the test highly tends to over-reject the null hypothesis of identical distributions. Here we exploit the smaller sample size of the non-overlapping increments to perform the test. The null hypothesis of identical distribution at $\alpha = 0.01$ is not rejected for 26629 on 37837 couples. This means that, despite of the looseness of condition (8) and of the still very large sample size at small lags, we observe a collapse of the densities in the 70% of the tested couples of sequences.

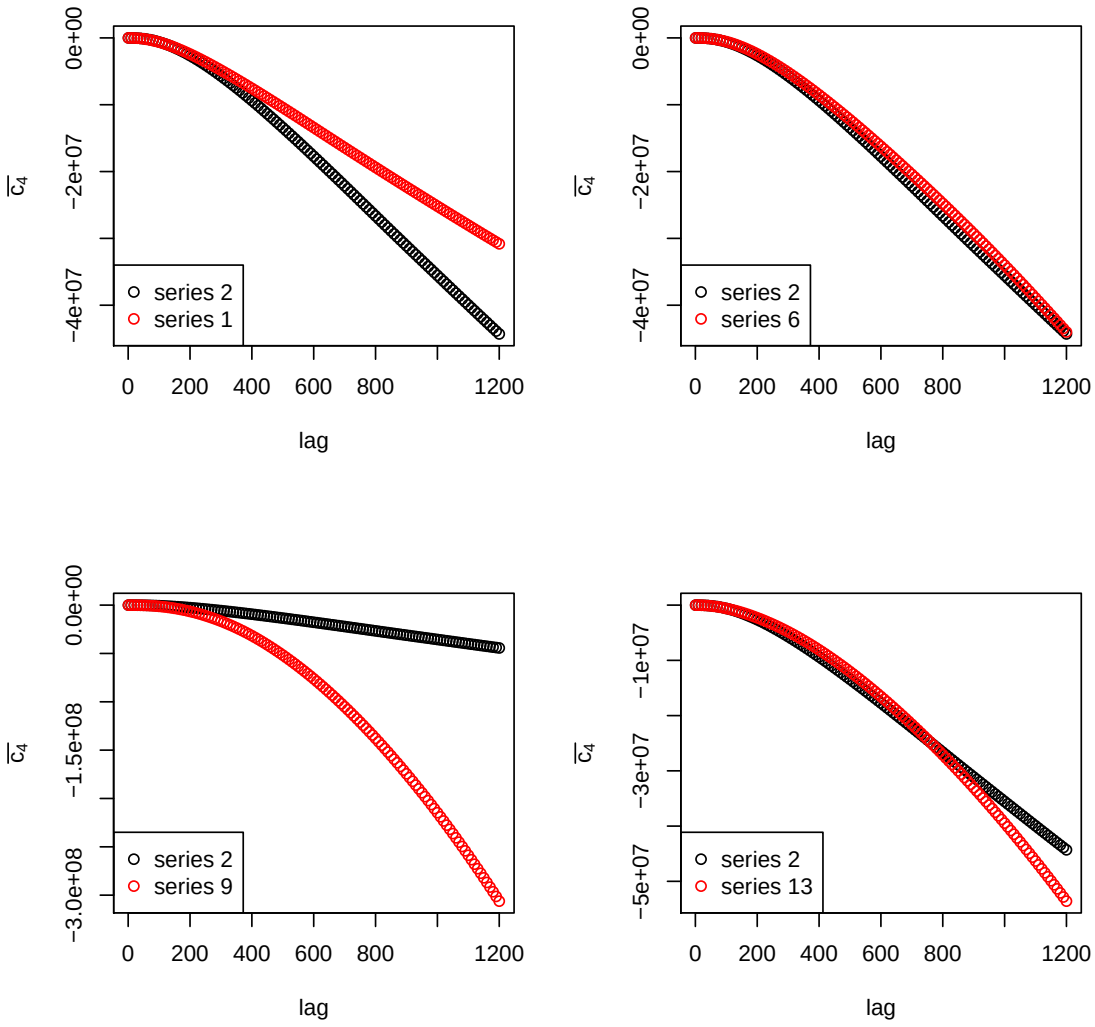


Figure 12: Comparison of $\overline{c}_4^{(2)}(s)$ and $\overline{c}_4^{(i)}(s)$ for $i = 1, 6, 9, 13$.

6 Conclusions

We perform a statistical analysis of the turbulent field in a close von Kármán flow at a very high Reynolds number, sampled in eighteen points in space. Thanks to the inhomogeneity of the von Kármán flow, the resulting time series can be considered as realizations of different turbulent experiments.

We exploit the recording times of a Laser Doppler Velocimetry device as a proxy of the turbulent field, using the log-waiting times in analogy with the velocity increments in classical turbulence analysis and with the log-returns in finance. We show that the considered proxy variable shares some stilyzed features both with turbulent velocity and financial markets: long memory-like behavior at all lags, shape of the pdf and tail heaviness changing with the time scale. The tail heaviness displays a clear difference with respect to the classical findings, according to which the tails can be described by a stretched exponential with stretching exponent approaching the Gaussian value $b = 2$ at large lags: in the case of the waiting times, the Gaussian limit is recovered at lag 0 for all the available time series, while growing values characterize the larger time scales, leading to light tails.

We then perform the analysis of the pdf of the log-waiting times. We find that a pdf of the class of the Normal Inverse Gaussian distribution describes well the log-waiting times at all lags. On the other hand, the representation of the evolution with the lag of the shape of the NIG density for each dataset through the shape triangle, results in a complex and non linear (in terms of straightness of the path across the triangle) behavior. However, if the log-waiting times are a good proxy for the turbulent field, we expect to be able to recognize a universal evolution of the pdf across the time scales. To this purpose, we introduce a deterministic time change in terms of the fourth order cumulants of the process and we show that such time change results in a collapse of the densities of the log-waiting times, at different lags and from different samples, to tha same law over a large range of time scales, exceeding the inertial range.

Possible further developments of the work should include a better understanding of the physical meaning of the log-waiting times and their relation with the turbulent velocity field. Moreover, an investigation of the link between the universal description given in this paper and the more classical frameworks encompassing universality but valid only in the inertial range, such as GESS, may be appropriate, as al-

ready been proposed in [5].

References

[1] Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. Scandinavian Journal of statistics, 151-157.

[2] Barndorff-Nielsen, O. E. (2001). Superposition of Ornstein–Uhlenbeck type processes. *Theory of Probability and Its Applications* **45(2)**, 175–194.

[3] Barndorff-Nielsen, O. E. and Schmiegel, J. (2004). LÃ©vy-based spatial-temporal modelling, with applications to turbulence. *Russian Mathematical Surveys* **59(1)**, 65.

[4] Barndorff-Nielsen, O. E., Blæsild, P., and Schmiegel, J. (2004). A parsimonious and universal description of turbulent velocity increments. *The European Physical Journal B - Condensed Matter and Complex Systems* **41(3)**, 345-363.

[5] Barndorff-Nielsen, O., Schmiegel, J., and Shephard, N. (2006). Time change and universality in turbulence and finance. TN Thiele Centre, University of Aarhus.

[6] Barndorff-Nielsen, O. E., and Schmiegel, J. (2008). Time change, volatility, and turbulence (pp. 29-53). Springer Berlin Heidelberg.

[7] Benzi, R., Ciliberto, S., Baudet, C., Chavarria, G. R., and Tripiccione, R. (1993). Extended self-similarity in the dissipation range of fully developed turbulence. *EPL (Europhysics Letters)* **24(4)**, 275.

[8] Benzi, R., Ciliberto, S., Tripiccione, R., Baudet, C., Massaioli, F., and Succi, S. (1993). Extended self-similarity in turbulent flows. *Physical review E* **48(1)**, R29.

[9] Benzi, R., Biferale, L., Ciliberto, S., Struglia, M. V., and Tripiccione, R. (1995). On the intermittent energy transfer at viscous scales in turbulent flows. *EPL (Europhysics Letters)* **32(9)**, 709.

[10] Benzi, R., Biferale, L., Ciliberto, S., Struglia, M. V., and Tripiccione, R. (1996). Generalized scaling in fully developed turbulence. *Physica D: Nonlinear Phenomena* **96(1)**, 162-181.

[11] Beran, J. (1994). Statistics for long-memory processes (Vol. 61). CRC Press.

[12] Box, G. E., Jenkins, G. M., and Reinsel, G. C. (2011). Time series analysis: forecasting and control (Vol. 734). John Wiley and Sons.

[13] Brockwell, P. J., and Davis, R. A. (2013). Time series: theory and methods. Springer Science and Business Media.

[14] Cortet, P. P., Diribarne, P., Monchaux, R., Chiffaudel, A., Daviaud, F. and Dubrulle, B. (2009): Normalized kinetic energy as a hydrodynamical global quantity for inhomogeneous anisotropic turbulence. *Physics of Fluids* **21(2)**, 25104.

[15] Cortet, P. P., Chiffaudel, A., Daviaud, F., and Dubrulle, B. (2010). Experimental evidence of a phase transition in a closed turbulent flow. *Physical review letters* **105(21)**, 214501.

[16] Faranda, D., Pons, F. M. E., Dubrulle, B., Daviaud, F., Saint-Michel, B., Herbert, É., and Cortet, P. P. (2014). Modelling and analysis of turbulent datasets using Auto Regressive Moving Average processes. *Physics of Fluids*, 26(10), 105101.

[17] Frisch, U. (1995). Turbulence: the legacy of AN Kolmogorov. Cambridge university press.

[18] Kolmogorov, A. N. (1941, January). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. In Dokl. Akad. Nauk SSSR (Vol. 30, No. 4, pp. 299-303).

[19] Ravelet, F., Chiffaudel, A. and Daviaud, F. (2008): Supercritical transition to turbulence in an inertially-driven von Kármán closed flow. *Journal of Fluid Mechanics* **601**, 339–364.

[20] She, Z. S., and Leveque, E. (1994). Universal scaling laws in fully developed turbulence. *Physical review letters*, **72(3)**, 336.

Paper F

Cascade-like fingerprints in an anisotropic closed turbulent flows. *Preprint*

Cascade-like fingerprints in an anisotropic closed turbulent flow

Flavio Pons and Jürgen Schmiegel

December 12, 2015

Abstract

We model the evolution in time of the energy budget of a turbulent flow, quantified by the turbulence intensity. We assume a multiplicative cascade as the data generating process, in analogy with the turbulent energy dissipation. In few words, the natural logarithm of the process can be written as the integral of a Lévy basis over an ambit set, a class of models proposed for the energy dissipation in [12]. The model is specified by the distribution of the Lévy basis and by the scaling with the lag of two-point correlators. The self-scaling of two-point correlators and the prediction of three-point correlators are used as fingerprints to evaluate the appropriateness of the model.

1 Introduction

In this paper we characterize the energy budget of an inhomogeneous anisotropic flow in terms of a continuous cascade process. In particular, we analyze a dataset produced in the von Kármán Experiment (VKE), performed at the Commissariat á l'Énergie Atomique (CEA) in Paris. The experiment consists of a realization of the von Kármán flow, obtained with a compact experimental setup, described in detail by [11]. This is designed to control and increase the Reynolds number up to $Re \sim 10^6$, in order to obtain a fully turbulent stationary flow; the fluid is a solution of water and glycerol, so that the viscosity can be modulated. The Reynolds number of the considered experiment is $Re = 10^5$; the fluid is pure water at a temperature $\theta = 20$ C and the corresponding kinematic viscosity is $\nu = 1.0 \cdot 10^{-6}\,\mathrm{m^2 s^{-1}}$. The fluid is

seeded with passive particles, exploited to sample either the axial or the azimuthal component of the fluid velocity field on a 11×17 point grid by means of Laser Doppler Velocimetry (LDV). The measurement is performed by a He-Ne Flowlite Laser with $\lambda = 632.8$ nm, with an average data rate of ~ 0.5 kHz when a particles crosses a grid point. The radial component of the velocity vector is recovered using the fact that $\nabla \cdot \mathbf{v} = 0$, since the mean velocity field is axisymmetric and solenoidal. The time series affected by spurious data and boundary effects are discarder, so that of the original 187, only 18 are retained; in this paper we consider measurements of the azimuthal component of the velocity vector.

The total duration of the experiment is fixed at $T = 3600$ s, but the number of measurements in each time series depends on the number of passive tracer particles passing at the corresponding measuring point during the experiment, due to the LDV technique. This implies that the sample size N of the time series is not fixed, ranging from $N = 1435270$ to $N = 2462311$, and that the time step between two subsequent measurements of the same time series is not constant.

1.1 The turbulence intensity

A suitable measure of the global energy budget of the VKE is the turbulence intensity, introduced in [8] as

$$i = \sqrt{\frac{\overline{v^2} - \overline{v}^2}{\overline{v}^2}}, \tag{1}$$

were v is a component of the velocity vector and $\overline{(\cdot)}$ denotes time averaging. Phisically, this measures the (normalized) difference between the average kinetic energy due to the turbulent fluctuations and the kinetic energy of the mean field and it is scalar and positive definite by construction.

In [8], the authors consider some different realizations of the VKE, all sampled through Stereoscopic Particle Image Velocimetry (SPIV). This technique produces time series of the velocity field on a 58×58 point grid; the time series are shorter and sampled with a lower data rate than in the LDV case, while the time step is fixed. The authors then introduce a modified version of the turbulence intensity, taking into account the anisotropic and inhomogeneous nature of the VKE:

$$\tilde{\delta}(t) = \frac{\langle v^2(t) \rangle}{\langle \overline{v}^2 \rangle} \tag{2}$$

where $\langle\cdot\rangle$ denotes ensamble averaging, which in this case is equivalent to spatial averaging. This quantity is still scalar and positive definite, but its value is instantaneous instead of global, since it measures the contribution of the instantaneous kinetic energy of the field to the kinetic energy of the mean field. Let us notice that the spatial averaging in eq. (2) makes it a global observable for the VKE.

In order to overcome the problem of the irregular sampling in time in the available dataset, we modify Eq. (2) to obtain a regular partition of the total time interval $[0, T]$ in each series. We choose a constant time interval Δt as the global average waiting time for 5 measurements in the whole dataset. We verify that in this way we avoid to consider time intervals without any measurement, with no drastic decrease in the length of the resulting time series and of the corresponding sampling frequency. The resulting value is $\Delta t = 9.27$ ms. We then obtain evenly spaced time series, each of length $M = \lfloor \frac{T}{\Delta t} \rfloor = 388267$, indexed by a discrete time $t_m = m \cdot \Delta t$, $m = 0, 1, \cdots, M$.

Considering this partition, we replace the instantaneous value of the kinetic energy $v^2(t)$ in the numerator of eq. (2) with the kinetic energy resulting from the mean velocity inside each time step

$$\overline{v}^2(\Delta t_m) = \frac{1}{n} \sum_{h=1}^{n} v_h,$$

where $\Delta t_m = t_m - t_{m-1}$ and n is the number of measurements in Δt_m. The new expression for the turbulence intensity reads

$$\delta(t) = \frac{\langle \overline{v}^2(\Delta t_m) \rangle}{\langle \overline{v}^2 \rangle}. \tag{3}$$

2 Continuous cascade processes

Continuous cascade processes can be considered as special cases of a much wider class of stochastic processes that have been termed *stochastic intermittency fields* in [14]. These stochastic intermittency fields are characterized by a nested hierarchy of correlations expressed in terms of an *ambit set* together with an underlying weightfunction. Depending on the choice of the shape of the ambit set and the weight of the independent noise contributions, a wide range of correlations can be modelled, including continuous cascade structures defined in terms of scaling and self-scaling of correlators. It is these correlators

we use as fingerprints of the underlying models in the analysis of the data.

2.1 Model construction

The basic notion for the construction of continuous cascade processes, widely used for the modelling of the statistics of the energy dissipation in turbulent flows, is that of an *independently scattered random measure* (i.s.r.m). They provide a natural basis for describing uncorrelated noise processes in space and time (but are not restricted to that geometry). Loosely speaking, the measure associates a random number with any subset of the underlying space S. Whenever two subsets are disjoint, the associated measures are independent, and the measure of a disjoint union of sets almost certainly equals the sum of the measures of the individual sets (see Appendix A for more detail and rigor).

A special class of i.s.r.m.'s is that of a homogeneous Lévy basis, where the distribution of the measure of each set is infinitely divisible and does not depend on the localisation of the subset. In this case, it is easy to handle integrals with respect to the Lévy basis using the well-known Lévy-Khintchine and Lévy-Itô representation for Lévy processes.

Let Z be a homogeneous Lévy basis on S. Then, for $S' \subset S$ we have the fundamental relation

$$\mathrm{E}\left\{\exp\left[\int_{S'} h(a)Z(\mathrm{d}a)\right]\right\} = \exp\left\{\int_{S'} \mathrm{K}[h(a)]\mathrm{d}a\right\}, \tag{4}$$

where h is a suitable deterministic function (subject to some minor condition to ensure the existence of the above integral), and K denotes the cumulant function of $Z(\mathrm{d}a)$, defined by

$$\ln \mathrm{E}\left\{\exp(\xi Z(\mathrm{d}a))\right\} = \mathrm{K}[\xi]\mathrm{d}a. \tag{5}$$

This relation allows to explicitely calculate and model the correlation structure of the integrated and h-weighted noise field Z.

A particular simple example is given by the model

$$\varepsilon(x) = \exp\left\{\int_{S'(x)} Z(\mathrm{d}a)\right\}, \tag{6}$$

where $S'(x) \subset S$ is a finite and attached to each point $x \in S$. This type of model has been used in [12] to model the statistics of the energy dissipation in a turbulent flow.

2.2 The self-scaling property

The multiplicative structure inferent to (6) can be characterized using k-point correlators of order $\vec{n} = (n_1, \ldots, n_k)$ defined as

$$c_{\vec{n}}(x_1, \ldots, x_k) = \frac{\mathrm{E}\left\{\varepsilon(x_1)^{n_1} \cdots \varepsilon(x_k)^{n_k}\right\}}{\mathrm{E}\left\{\varepsilon(x_1)^{n_1}\right\} \cdots \mathrm{E}\left\{\varepsilon(x_k)^{n_k}\right\}}. \tag{7}$$

These correlators can all be expressed in terms of volumes of overlaps $V(S'(x_i) \cap V(S'(x_j)$, $i, j = 1, \ldots, k$ and the corresponding constants

$$\overline{\mathrm{K}}[n_i, n_j] = \mathrm{K}[n_i + n_j] - \mathrm{K}[n_i] - \mathrm{K}[n_j]. \tag{8}$$

For $k = 2$ and $\vec{n} = (n_1, n_2)$ one obtains

$$c_{\vec{n}}(x_1, x_2) = \exp\left\{\overline{\mathrm{K}}[n_1, n_2] V(S'(x_1) \cap S'(x_2))\right\}. \tag{9}$$

Here the properties of the underlying Lévy basis are separated from the properties of the associated sets S' which allows to represent correlators of order $\vec{m} = (m_1, m_2)$ as a scaling relation to correlators of order $\vec{m} = (m_1, m_2)$

$$c_{\vec{m}}(x_1, x_2) = c_{\vec{n}}(x_1, x_2)^{r(\vec{m}, \vec{n})} \tag{10}$$

where

$$r(\vec{m}, \vec{n}) = \frac{\overline{\mathrm{K}}[m_1, m_2]}{\overline{\mathrm{K}}[n_1, n_2]}. \tag{11}$$

This self-scaling property is independent of the shape and size of the associated sets S' and only depends on the properties of the underlying Lévy basis. In the sequel we will test and verify this property for our data-set as a first fingerprint of an underlying cascade-like structure.

2.3 3-point correlations

The second fingerprint of an underlying cascade-like structure concerns the behaviour of three point correlators. For simplicity we set $\mathrm{E}\{\varepsilon(x)\} = 1$, i.e. $\mathrm{K}[1] = 0$ without loss of generality. We also specify the underlying space to be $S = \mathbb{R}$ (for the sake of simplicity) and the associated sets $S(x) = [x - L, x]$ where $L > 0$ denotes a kind of decorrelation distance. The three point correlators of order $\vec{n} = (1, 1, 1)$ for

$x_1 \leq x_2 \leq x_3$ and $x_3 - x_1 \leq T$ can then be expressed as

$$\begin{aligned} c_{\vec{n}}(x_1, x_2, x_3) =& \exp\left\{\mathrm{K}[2] V(S'(x_1) \cap S'(x_2)) \setminus S'(x_3))\right\} \\ & \exp\left\{\mathrm{K}[2] V(S'(x_2) \cap S'(x_3)) \setminus S'(x_1))\right\} \\ & \exp\left\{\mathrm{K}[3] V(S'(x_1) \cap S'(x_3))\right\} \\ =& \frac{c_{1,1}(x_1, x_2)}{c_{1,1}(x_1, x_3)} \frac{c_{1,1}(x_2, x_3)}{c_{1,1}(x_1, x_3)} c_{2,1}(x_1, x_3) c_{1,1}(x_1, x_1) \\ =& c_{1,1}(x_1, x_2) c_{1,1}(x_2, x_3) c_{1,1}(x_1, x_1) \\ & c_{1,1}(x_1, x_3)^{r(2,1)} c_{1,1}(x_1, x_3)^{-2}. \end{aligned} \tag{12}$$

Similarly, for $0 \leq x_2 - x_1 \leq T$, $x_3 - x_1 > T$ and $x_3 - x_2 \leq T$ we obtain

$$c_{\vec{n}}(x_1, x_2, x_3) = c_{1,1}(x_1, x_2) c_{1,1}(x_2, x_3) \tag{13}$$

and for $0 \leq x_2 - x_1 \leq T$, $x_3 - x_2 > T$

$$c_{\vec{n}}(x_1, x_2, x_3) = c_{1,1}(x_1, x_2). \tag{14}$$

In all cases, the three-point correlators of order $(1, 1, 1)$ are completely determined by the two-point correlators of order $(1, 1)$ and the self-scaling exponent $r((2, 1)$. Equations (12), (13) and (14) are the second fingerprint of an underlying cascade process we will confront our data with.

3 Data analysis

In the following, we show that the turbulence intensity can be modelled by a process as described in Eq. 6, so that its evolution in time can be written as

$$\delta(t) = \exp\left\{\int_{S'(t)} Z(\mathrm{d}a)\right\} = \exp\left\{Z(S'(t))\right\}. \tag{15}$$

In this framework, we find convenient to model the natural logarithm of the observable. An additional remark about notation is required. The model setting illustrated in Section 2 is very general. In the following, we focus on the purely temporal dynamics of the observable, so we replace x with t. Moreover, since we will consider time lags $\Delta = t_2 - t_1$, we will write $c_{n_1,n_2}(\Delta)$ rather than $c_{n_1,n_2}(t_1, t_2)$.

In this section, we describe how the shape of the ambit set S' and the marginal distribution of the Lévy basis can be estimated from data.

We also evaluate the appropriateness of the continuous multiplicative cascade model through the analysis of the self-scaling of the two-point correlators and the predictability of the three-point correlators from the two-point correlators.

3.1 Marginal distributions

Eq. 5 implies a relation between the cumulants (or the moments) of the Lévy basis and of the observable through the volume of the ambit set. In particular, we have $\mathrm{K}[Z(S'(t))] = \mathrm{K}[Z]\mathrm{V}(S')$, assuming a stationary ambit set, so that $S'(t) = S'(0) + t$.

There is no physical prescription on the distribution of the turbulence intensity. A visual inspection of the histogram of $\ln \delta(t)$ shows a negative asymmetry, while the log-histogram suggests a heavy, exponentially decaying left tail (see Fig. 1).

In [9], the authors discuss a similar problem concerning the kinetic energy dissipation $\varepsilon(t)$. On the one hand, Kolmogorov-Obukhov theory prescribes a log-normal model for the energy dissipation, so that we expect $\ln \varepsilon(t) \sim N(\mu, \sigma^2)$. On the other hand, the histograms show a clear non-Gaussian nature of the data. The authors consider three possible candidate distributions: Gaussian, Stable and Normal Inverse Gaussian (NIG), a four-parameter distribution featuring asymmetry, anomalous kurtosis and semi-heavy tails. The NIG can be regarded as a generalisation of the Normal distribution, in which the variance is allowed to be random. They show that the NIG provides the best fit among the considered models.

Following these results, we propose a NIG model for $\ln \delta(t)$. The fit is shown in Fig. 1, superimposed to the histogram and log-histogram. We also display the maximum likelihood estimates of the parameters, $\mu = 0.916$, $\sigma = 0.726$, $\alpha = 5.381$, $\beta = 4.47$. The fit shows a good adaptation of the NIG also for very small values of the log-density. Moreover, the exponential decay of the left tail is clear from the log-histogram and the relative log-density fit. Notice that two-point correlators c_{n_1,n_2} exist only if $n_1 + n_2 < \alpha - \beta$, thus possibly requiring a constrained estimation of the NIG parameters. In our case we compute the correlators up to a maximum order $n_1 + n_2 = 4$, so that no constrain is needed, since $\alpha - \beta = 9.851 > 4$.

Let us remark that if $Z \sim NIG(\alpha, \beta, \mu, \sigma)$, Eq. 5 implies that $\ln \delta(t) \sim NIG(\alpha, \beta, \mathrm{V}(S)\mu, \mathrm{V}(S)\sigma)$: thus, once the shape of the ambit set S is known, the distribution of the Lévy basis is specified.

3.2 Two-point correlators

As already mentioned in Section 2, two-point correlators can be used to specify the model. In particular, Eq. 9 links the two-point correlator at lag $t_2 - t_1 = \Delta$ to the volume of the overlap of the ambit sets $S'(t_1)$ and $S'(t_2)$. On the other hand, the ambit set is bounded by a decreasing continuous function $g(t) > 0$. Then, if the correlators display a scaling with the lag, we can relate the slope of the two-point correlators to the boundary of the ambit set as follows

$$\frac{\partial}{\partial \Delta} \ln c_{n_1,n_2}(\Delta) = \overline{\mathrm{K}}[n_1, n_2] g^{(-1)} \left(\frac{\Delta}{2} \right). \tag{16}$$

In Fig. 2 we display the scaling of $c_{1,1}$, $c_{1,2}$, $c_{2,2}$ in double logarithmic scale. The plot suggests a self-scaling property in the two-point correlators, which we analyse further in this section.

Concerning the scaling, two behaviors can be recognised: the decay with the lag is quadratic up to $\Delta \sim 25$, while it is linear for $\Delta > 25$. The maximum lag of interest for the ambit set construction, at which decorrelation happens so that $\ln c_{n_1,n_2}(\Delta) = 0$, is $\Delta \sim 400$. A proper model for the scaling in time of the two-point correlators leads to the specification of the ambit set. We verified that a combination of a quadratic fit for $0 < \Delta \leq 25$ and a linear fit for $25 < \Delta < 400$ may be a suitable choice. Nevertheless, a discontinuity appears at lag 25, requiring some manipulation to be removed. Since the combination of the two models, including the intercepts, requires the estimation of five parameters, we choose instead a polynomial fit of order 4 for the whole interval $0 < \Delta < 400$. The superposition of the resulting model $c_{1,1}(\Delta) = \sum_{j=0}^{4} a_j \Delta^j$ to the empirical correlator is displayed in Fig. 3. The value of the parameters, estimated through nonlinear least squares, are shown in Table 1.

a_0	a_1	a_2	a_3	a_4
0.3290	-0.2328	0.7755	-0.0122	0.0007

Table 1: Estimated coefficients for the polynomial fit of $c_{1,1}$.

As already mentioned in Section 2, two-point correlators may display self-scaling as a fingerprint of an underlying multiplicative cascade process. This self-scaling relation is clear in our data, as shown in the plot of $c_{1,1}$ vs. $c_{1,2}$ (Fig. 4), $c_{1,1}$ vs. $c_{2,2}$ (Fig. 5), $c_{1,2}$ vs. $c_{2,2}$

(Fig. 6). The values of the coefficient of determination R^2 for the linear fit in the three cases are, respectively $R_1^2 = 0.993$, $R_2^2 = 0.9985$, $R_3^2 = 0.999$.

3.3 Three-point correlators

A further fingerprint, involving higher order moments, of a multiplicative cascade process comes from the matching between the empirical and theoretical three-point correlators. In fact, under the model ansatz in Eq. 6, Eq 12 - 14 give us a theoretical prediction of the empirical value of $c_{1,1,1}$ computed as in Eq. 7 with $\vec{n} = (1, 1, 1)$.

The comparison between the estimated three-point correlator and its theoretical prediction under model 7, displayed in Fig. 7, shows a good matching, resulting in a strong indication of the appropriateness of the modeling framework proposed in Section 2.

4 Summary and conclusions

The energy budget of a turbulent flow can be measured through the turbulence intensity. In this paper we characterise the continuous-time dynamics of this quantity for a dataset from the von Kármán Experiment. After a sketch of the experimental setting, we construct the instantaneous global (in space) turbulence intensity. We then introduce a suitable class of models for continuous multiplicative cascade processes, specified through the marginal distribution of the available time series and the slope of two-point correlators. We then check the appropriateness of the model through the self-scaling property of two-point correlators and the predictability of the three-point correlator from the two-point correlators, since both are fingerprint of an underlying multiplicative cascade model.

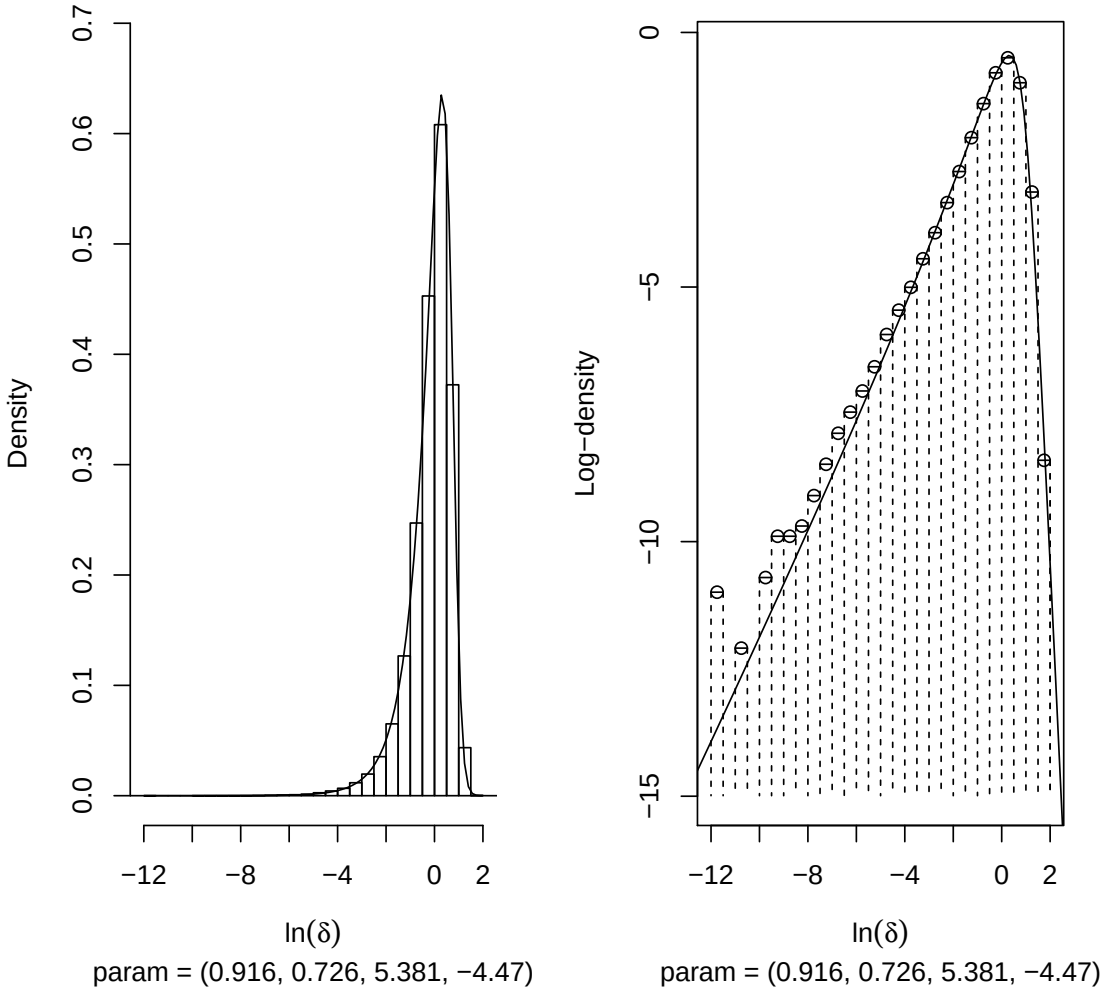


Figure 1: NIG fit for the marginal distribution of $\ln \delta(t)$.

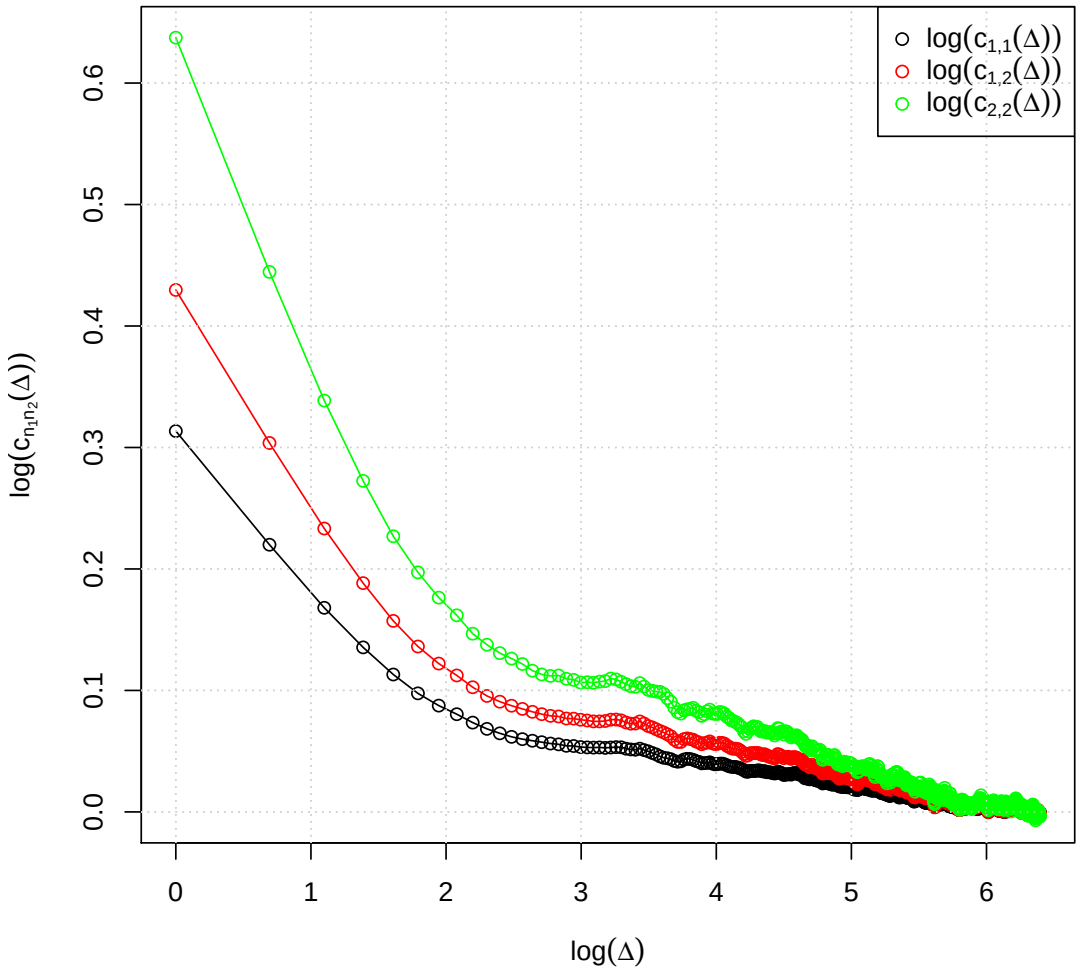


Figure 2: Scaling property of $\ln c_{1,1}$, $\ln c_{1,2}$, $\ln c_{2,2}$.

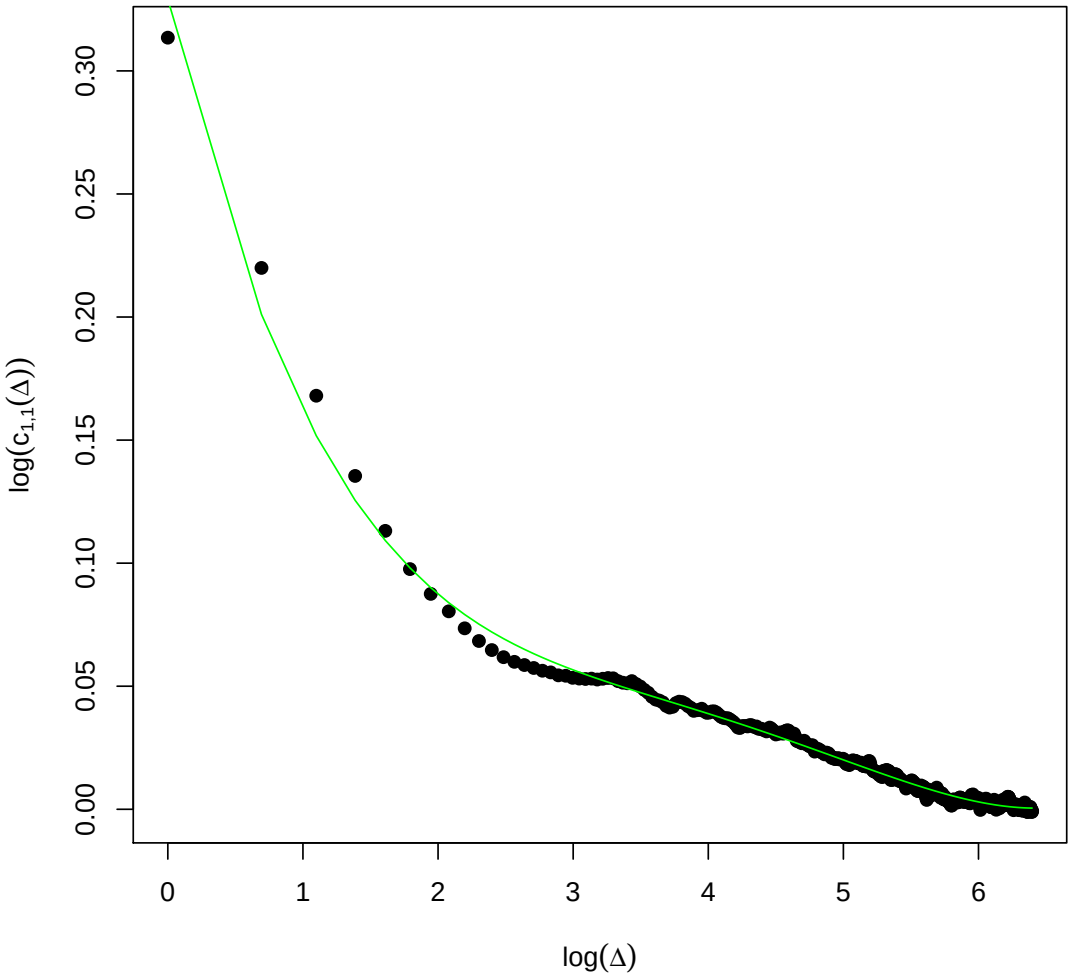


Figure 3: Polynomial fit of order 4 for $\ln c_{1,1}$.

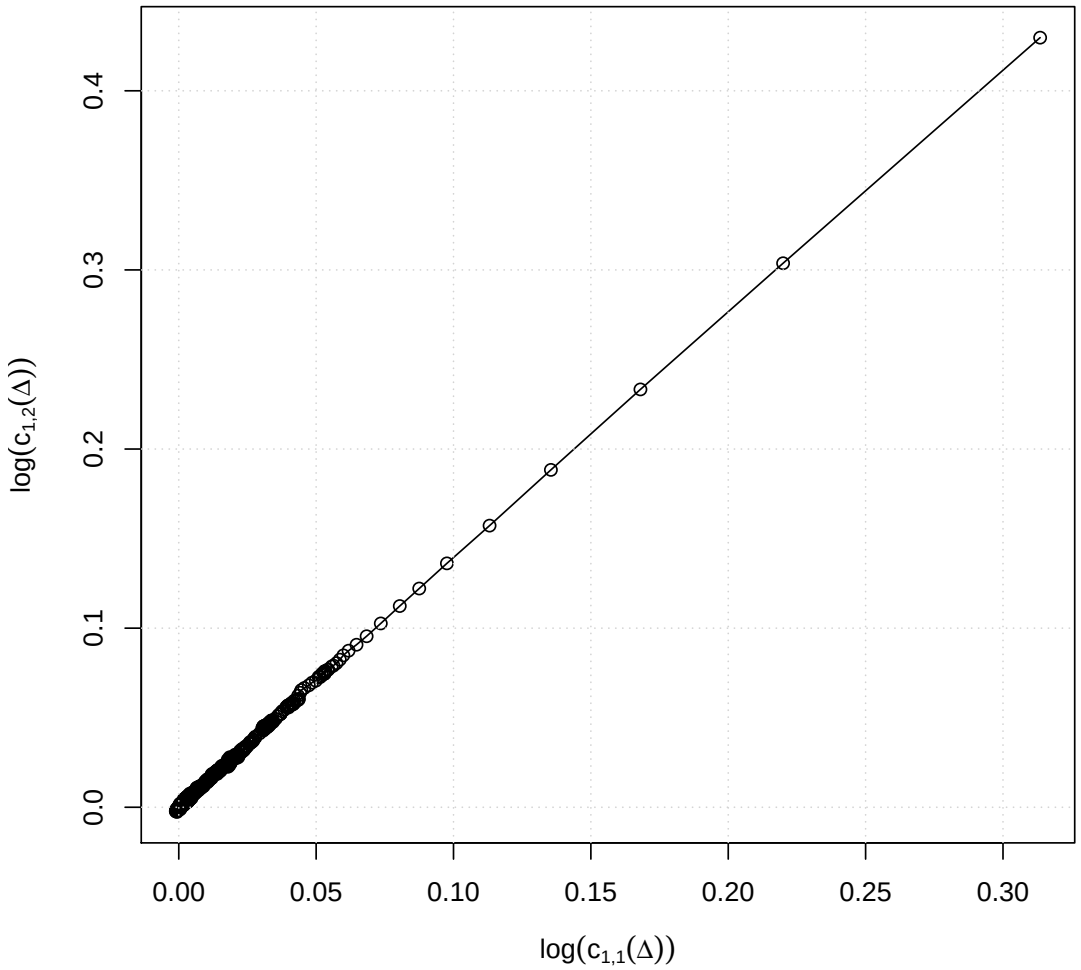


Figure 4: Self-scaling property: $c_{1,1}$ vs. $c_{1,2}$ in double logarithmic scale.

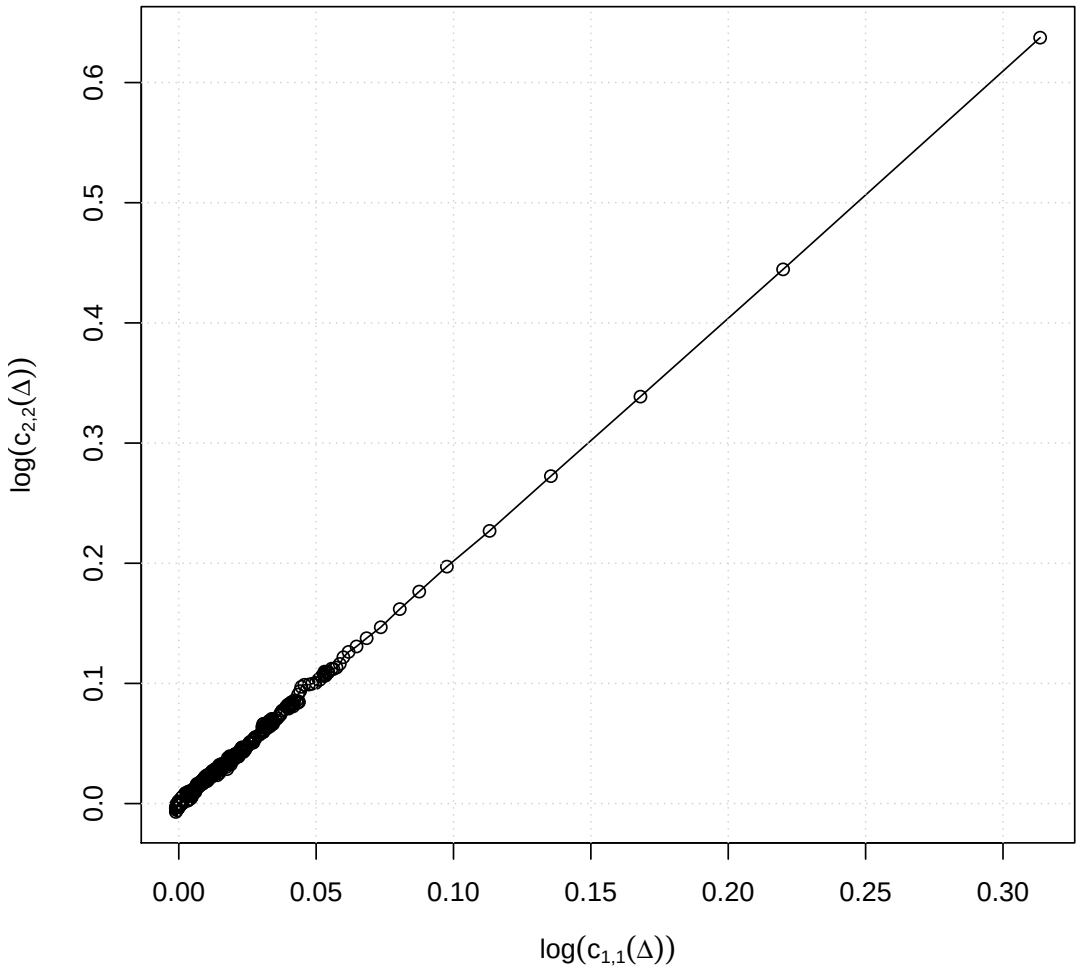


Figure 5: Self-scaling property: $c_{1,1}$ vs. $c_{2,2}$ in double logarithmic scale.

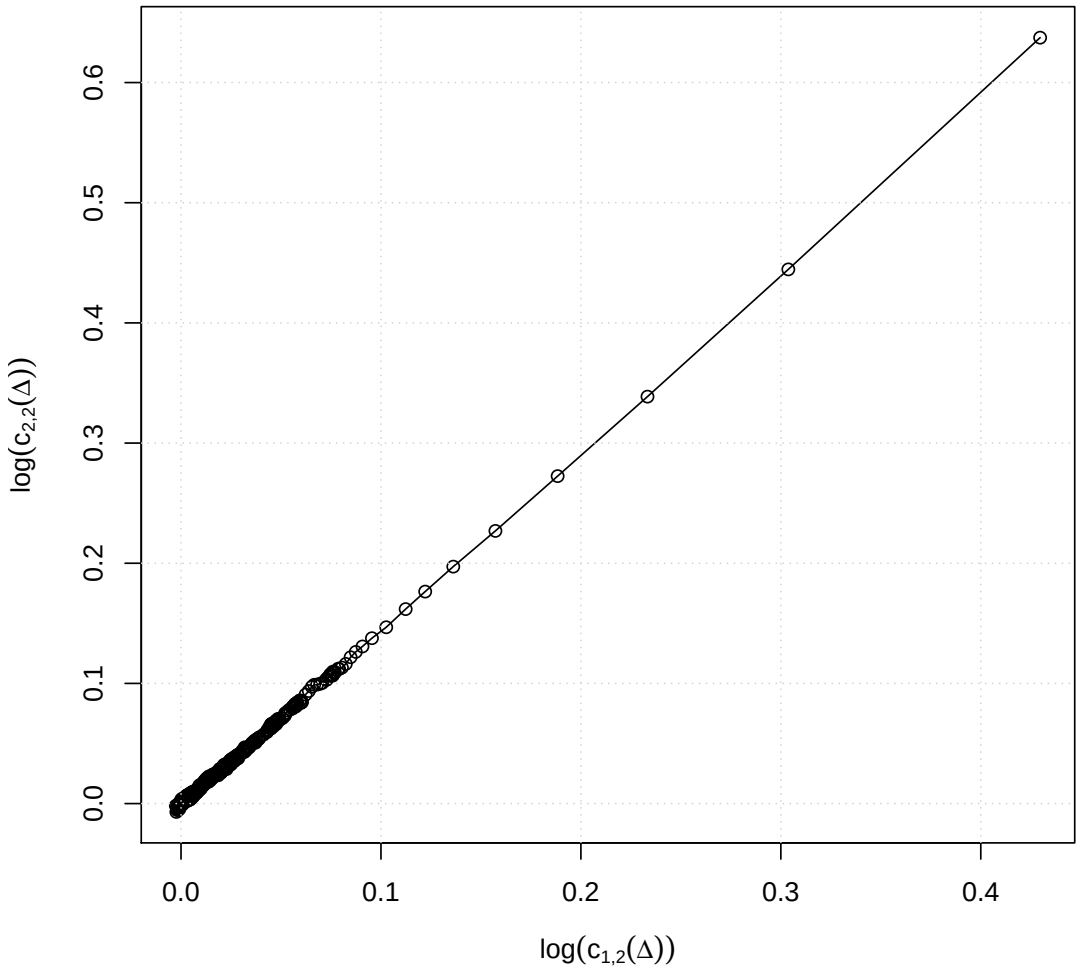


Figure 6: Self-scaling property: $c_{1,2}$ vs. $c_{2,2}$ in double logarithmic scale.

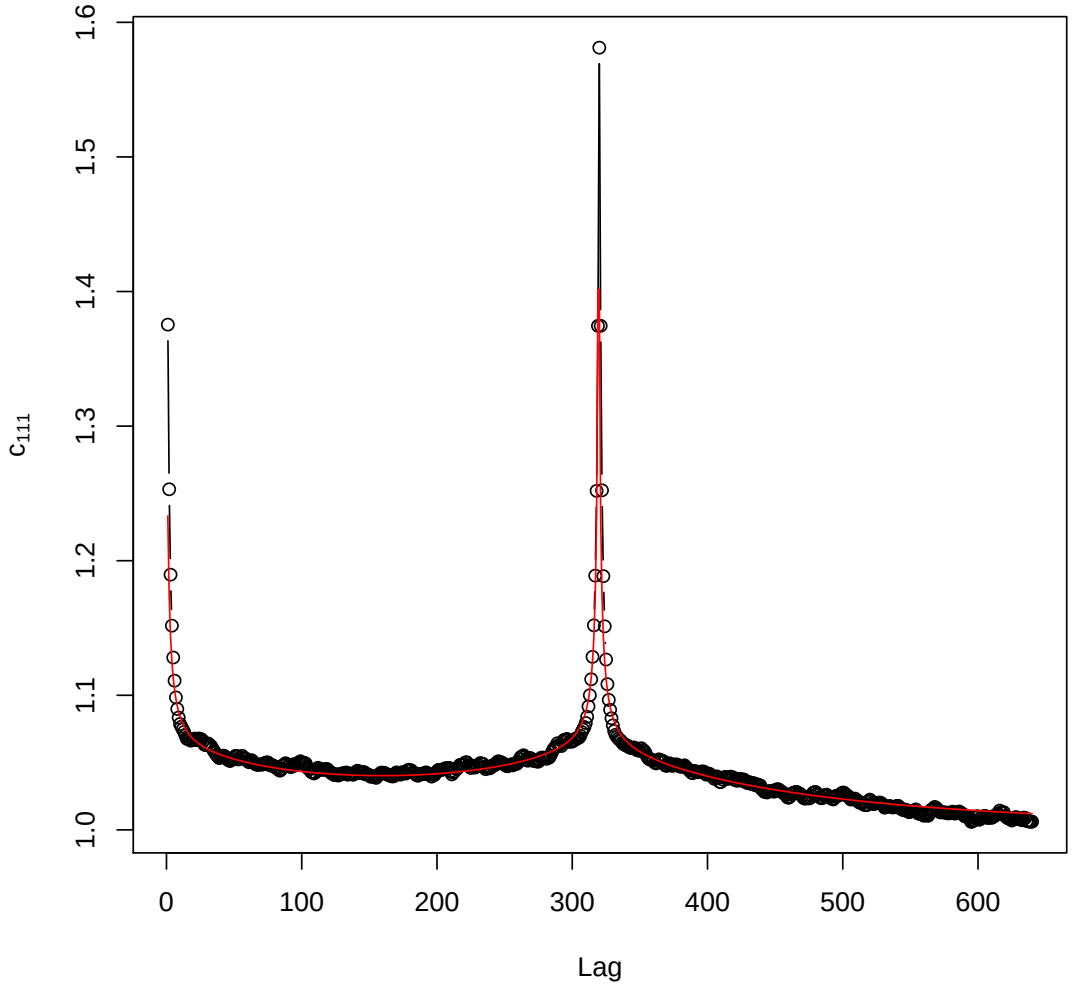


Figure 7: Comparison between empirical (black circles) and theoretical (red line) three-point correlator $c_{1,1,1}$.

A Integration with respect to a Lévy basis

Ambit processes were introduced in [3] as a framework for tempo-spatial modelling. These processes are defined in terms of integrals with respect to a Lévy basis. Here, we restrict our attention to those ambit processes defined as the stochastic integral of a deterministic function with respect to a homogeneous Lévy basis defined on $\mathbb{R}^2$.

Denote by $\mathcal{B}_b\left(\mathbb{R}^2\right)$ the set of bounded Borel subsets of $\mathbb{R}^2$. A Lévy basis Λ on $\mathbb{R}^2$ is an infinitely divisible, independently scattered random measure on $\mathbb{R}^2$, i.e. $\left(\Lambda\left(A\right)\right)_{A\in\mathcal{B}_b(\mathbb{R}^2)}$ is a stochastic process such that: (i) $\Lambda\left(A\right)$ is infinitely divisible; (ii) $\Lambda\left(A\right)$ and $\Lambda\left(B\right)$ are independent if $A\cap B=\varnothing$; and, (iii) If $A_1,\ldots,A_n\in\mathcal{B}_b\left(\mathbb{R}^2\right)$ are disjoint and such that $\cup_{i=1}^n A_i\in\mathcal{B}_b\left(\mathbb{R}^2\right)$, then

$$\Lambda\left(\bigcup_{i=1}^n A_i\right)\overset{a.s.}{=}\sum_{i=1}^n\Lambda\left(A_i\right).$$

A Lévy basis Λ on $\mathbb{R}^2$ is called homogeneous if $\Lambda\left(A\right)\overset{d}{=}\Lambda\left(A+x_0\right)$, for $x_0\in\mathbb{R}^2$.

The stochastic integral $\int f d\Lambda$ of a deterministic measurable function $f:\mathbb{R}^2\to\mathbb{R}$ with respect to a homogeneous Lévy basis Λ is defined in two steps: (a) If $f=\sum_{i=1}^n a_i 1_{A_i}$ is a real simple function on $\mathbb{R}^2$ with $A_1,\ldots,A_n$ disjoint, for $A\in\mathcal{B}\left(\mathbb{R}^2\right)$, we define

$$\int_A f d\Lambda=\sum_{i=1}^n a_i\Lambda\left(A_i\cap A\right).$$

(b) If $f:\mathbb{R}^2\to\mathbb{R}$ can be approximated almost everywhere (with respect to the Lebesgue measure) by a sequence of simple functions $\{f_n\}$ as in (a), provided that the limit exist, we define

$$\int_A f d\Lambda=P-\lim\int_A f_n d\Lambda, \tag{17}$$

for $A\in\mathcal{B}\left(\mathbb{R}^2\right)$. We say that a measurable function $f:\mathbb{R}^2\to\mathbb{R}$ is Λ-integrable if the integral (17) exists.

Let $K\left\{z\ddagger X\right\}:=\log E\left\{\exp\left(zX\right)\right\}$ and $C\left\{z\ddagger X\right\}:=\log E\left\{\exp\left(izX\right)\right\}$ denote the log-moment generating function and the log-characteristic function, respectively, of the random variable X. The functions K and

C will be called the kumulant and cumulant function, respectively. For each homogeneous Lévy basis Λ, we can associate a random variable Λ' to Λ such that

$$K\left\{z \ddagger \Lambda\left(da\right)\right\} = K\left\{z \ddagger \Lambda'\right\} da,$$

and

$$C\left\{z \ddagger \Lambda\left(da\right)\right\} = C\left\{z \ddagger \Lambda'\right\} da.$$

The random variable Λ' is called the Lévy seed of Λ.

The stochastic integral $\int f d\Lambda$ and the Lévy seed Λ' satisfy the next relation (see [10] for a proof).

Proposition 1 *Let Λ be a Lévy basis on $\mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}$ a Λ-integrable function. Then*

$$K\left\{z \ddagger \int_A f\left(a\right) d\Lambda\left(a\right)\right\} = \int_A K\left\{z f\left(a\right) \ddagger \Lambda'\right\} da$$

and

$$C\left\{z \ddagger \int_A f\left(a\right) d\Lambda\left(a\right)\right\} = \int_A C\left\{z f\left(a\right) \ddagger \Lambda'\right\} da.$$

For the purposes of this paper, an ambit process is a stochastic process $\left(Y_t\right)_{t\geq 0}$ of the form

$$Y_t = \int_{A_t} f\left(\left(0,t\right) - a\right) d\Lambda\left(a\right),$$

where $A \in \mathcal{B}_b\left(\mathbb{R}^2\right)$ and $A_t = A + (0,t)$. For a more general definition of ambit processes and a discussion of their mathematical properties, we refer to [6].

B Normal inverse Gaussian distribution

The normal inverse Gaussian (NIG) distribution is a four-parameter family of continuous probability distributions whose probability density function is given by

$$f_{\mathrm{NIG}(\alpha,\beta,\mu,\delta)}\left(x\right) = \frac{\alpha e^{\delta\gamma}}{\pi} e^{\beta(x-\mu)} \frac{K_1\left(\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right)}{q\left(\frac{x-\mu}{\delta}\right)}, \tag{18}$$

where $\gamma = \alpha^2 - \beta^2$, $q(x) = \sqrt{1+x^2}$ and K_1 denotes the modified Bessel function of the second kind with index 1. The domain of variation of the parameters is given by $\mu \in \mathbb{R}$, $\delta \in \mathbb{R}_+$, and $0 \leq |\beta| < \alpha$. The parameters α and β are shape parameters, μ determines the location, and δ determines the scale. The distribution is denoted by $\mathrm{NIG}(\alpha, \beta, \mu, \delta)$.

The NIG distribution is a particular case of the GH law that arises when $\lambda = 1/2$ in $f_{\mathrm{GH}(\lambda,\chi,\theta,\mu,\Sigma,\gamma)}$. However, the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization of the GH law produces a parametrization for the NIG distribution which differs from the above parametrization $f_{\mathrm{NIG}(\alpha,\beta,\mu,\delta)}$. The relation between these two parameterizations can be found in [7].

The cumulant function $K(z; \alpha, \beta, \mu, \delta) = \log E[\exp\{zV\}]$ of a random variable V with distribution $\mathrm{NIG}(\alpha, \beta, \mu, \delta)$ is given by

$$K(z; \alpha, \beta, \mu, \delta) = z\mu + \delta\left(\gamma - \sqrt{\alpha^2 - (\beta + z)^2}\right). \tag{19}$$

It follows immediately from this that the normal inverse Gaussian distribution is infinitely divisible. Namely, if $X_i \sim \mathrm{NIG}(\alpha, \beta, \mu_i, \delta_i)$, $i = 1, 2$, are independent random variables, then we have $X_1 + X_2 \sim \mathrm{NIG}(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$.

It is often of interest to consider alternative parameterizations of the normal inverse Gaussian laws. In particular, letting $\overline{\alpha} = \delta\alpha$ and $\overline{\beta} = \delta\beta$, we have that $\overline{\alpha}$ and $\overline{\beta}$ are invariant under location-scale changes.

Sometimes it is useful to represent NIG distributions in the so-called shape triangle. Consider the alternative asymmetry and steepness parameters χ and ξ defined by

$$\xi = (1 + \overline{\gamma})^{-1/2}, \quad \chi = \rho\xi,$$

where $\rho = \beta/\alpha$ and $\overline{\gamma} = \delta\gamma = \delta\sqrt{\alpha^2 - \beta^2}$. These parameters are invariant under location-scale changes. Their range defines the NIG shape triangle

$$\{(\chi, \xi) : 0 < \xi < 1, -\xi < \chi < \xi\}.$$

When $\chi = 0$ the NIG distribution is symmetric. Values $\chi > 0$ indicate a positively skewed distribution and $\chi < 0$ a negatively skewed law. The steepness parameter ξ measures the heaviness of the tails of the NIG distribution. The limiting case $\xi = 0$ corresponds to a normal distribution.

The NIG law has a wide range of applications. For more details about this distribution and their applications, we refer to [1, 2, 4, 5].

References

[1] Barndorff-Nielsen, O. E. (1977): Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A* **353**(1674), 401–419.

[2] Barndorff-Nielsen O.E. (1997): Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.* **24**, 1–13.

[3] Barndorff-Nielsen, O. E. and Schmiegel, J. (2003): Lévy-based Tempo-Spatial Modelling; with Applications to Turbulence. *Uspekhi Mat. Nauk*. **159**, 65-91.

[4] Barndorff-Nielsen O.E., Blæsild, P. Schmiegel J. (2004): A parsimonious and universal description of turbulent velocity increments. *Eur. Phys. J. B* **41**, 345–363.

[5] Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S.I. (Eds.) (2012): Lévy processes: theory and applications. Springer Science & Business Media, New York.

[6] Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2015): Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. =Banach Center Publ. **104**, 25–60.

[7] Breymann, W. and Lüthi, D. (2010): ghyp: A package on generalized hyperbolic distributions. Manual for the R package ghyp. Web-site: https://cran.r-project.org/web/packages/ghyp/ghyp.pdf.

[8] Cortet, P. P., Diribarne, P., Monchaux, R., Chiffaudel, A., Daviaud, F. and Dubrulle, B. (2009): Normalized kinetic energy as a hydrodynamical global quantity for inhomogeneous anisotropic turbulence. *Physics of Fluids* **21(2)**, 25104.

[9] Hedevang, E. and Schmiegel, J. (2013). A causal continuous-time stochastic model for the turbulent energy cascade in a helium jet flow. Journal of Turbulence **14(11)**, 1–26.

[10] Rajput, B.S. and Rosinski, J. (1989): Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* **82** (3), 451–487.

[11] Ravelet, F., Chiffaudel, A. and Daviaud, F. (2008): Supercritical transition to turbulence in an inertially-driven von Kármán closed flow. =Journal of Fluid Mechanics **601**, 339–364.

[12] Schmiegel, J., Cleve, J., Eggers, H. C., Pearson, B. R. and Greiner, M. (2004). Stochastic energy-cascade model for (1+ 1)-dimensional fully developed turbulence. Physics Letters A **320(4)**, 247–253.

[13] Barndorff-Nielsen, O. E. (2001). Superposition of Ornstein–Uhlenbeck type processes. Theory of Probability & Its Applications **45(2)**, 175–194.

[14] Barndorff-Nielsen, O. E. and Schmiegel, J. (2004). Lvy-based spatial-temporal modelling, with applications to turbulence. Russian Mathematical Surveys **59(1)**, 65.

Bibliography

[1] Barndorff-Nielsen, O. E. (1977): Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A* **353**(1674), 401–419.

[2] Barndorff-Nielsen O.E. (1997): Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.* **24**, 1–13.

[3] Barndorff-Nielsen, O. E. (2001). Superposition of Ornstein–Uhlenbeck type processes. Theory of Probability & Its Applications **45(2)**, 175–194.

[4] Barndorff-Nielsen, O. E. and Schmiegel, J. (2003): Lévy-based Tempo-Spatial Modelling; with Applications to Turbulence. *Uspekhi Mat. Nauk*. **159**, 65-91.

[5] Barndorff-Nielsen, O. E. and Schmiegel, J. (2004). LÃ©vy-based spatial-temporal modelling, with applications to turbulence. Russian Mathematical Surveys **59(1)**, 65.

[6] Barndorff-Nielsen O.E., Blæsild, P. Schmiegel J. (2004): A parsimonious and universal description of turbulent velocity increments. *Eur. Phys. J. B* **41**, 345–363.

[7] Barndorff-Nielsen, O., Schmiegel, J., and Shephard, N. (2006). Time change and universality in turbulence and finance. TN Thiele Centre, University of Aarhus.

[8] Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S.I. (Eds.) (2012): Lévy processes: theory and applications. Springer Science & Business Media, New York.

[9] Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2015): Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. =Banach Center Publ. **104**, 25–60.

[10] Breymann, W. and Lüthi, D. (2010): ghyp: A package on generalized hyperbolic distributions. Manual for the R package ghyp. Web-site: https://cran.r-project.org/web/packages/ghyp/ghyp.pdf.

[11] Benzi, R., Ciliberto, S., Baudet, C., Chavarria, G. R., and Tripiccione, R. (1993). Extended self-similarity in the dissipation range of fully developed turbulence. *EPL (Europhysics Letters)* **24(4)**, 275.

[12] Benzi, R., Ciliberto, S., Tripiccione, R., Baudet, C., Massaioli, F., and Succi, S. (1993). Extended self-similarity in turbulent flows. *Physical review E* **48(1)**, R29.

[13] Benzi, R., Biferale, L., Ciliberto, S., Struglia, M. V., and Tripiccione, R. (1995). On the intermittent energy transfer at viscous scales in turbulent flows. *EPL (Europhysics Letters)* **32(9)**, 709.

[14] Benzi, R., Biferale, L., Ciliberto, S., Struglia, M. V., and Tripiccione, R. (1996). Generalized scaling in fully developed turbulence. *Physica D: Nonlinear Phenomena* **96(1)**, 162-181.

[15] Cortet, P. P., Diribarne, P., Monchaux, R., Chiffaudel, A., Daviaud, F. and Dubrulle, B. (2009): Normalized kinetic energy as a hydrodynamical global quantity for inhomogeneous anisotropic turbulence. *Physics of Fluids* **21(2)**, 25104.

[16] Cortet, P. P., Chiffaudel, A., Daviaud, F. and Dubrulle, B. (2010). Experimental evidence of a phase transition in a closed turbulent flow. *Physical review letters* **105(21)** 214501.

[17] Faranda, D., Dubrulle, B., Daviaud, F. and Pons, F. M. E. (2014). Probing turbulence intermittency via autoregressive moving-average models. *Physical Review E* **90(6)** 061001.

[18] Faranda, D., Pons, F. M. E., Dubrulle, B., Daviaud, F., Saint-Michel, B., Herbert, É. and Cortet, P. P. (2014). Modelling and analysis of turbulent datasets using Auto Regressive Moving Average processes. *Physics of Fluids* **26(10)** 105101.

[19] Faranda, D., Pons, F. M. E., Giachino, E., Vaienti, S. and Dubrulle, B. (2015). Early Warnings Indicators of financial crises via Auto Regressive Moving Average models. *Communications in Nonlinear Science and Numerical Simulation.*

[20] Faranda, D., Dubrulle, B. and Pons, F. M. E. (2014). Statistical early-warning indicators based on autoregressive moving-average models. *Journal of Physics A: Mathematical and Theoretical* **47(25)**, 252001.

[21] Frisch, U. (1995). Turbulence: the legacy of AN Kolmogorov. Cambridge university press.

[22] Hedevang, E. (2012). Stochastic modelling of turbulence-with applications to wind energy. PhD Dissertations, Department of Mathematical Sciences, University of Aarhus, 2012.

[23] Hedevang, E. and Schmiegel, J. (2013). A causal continuous-time stochastic model for the turbulent energy cascade in a helium jet flow. Journal of Turbulence **14(11)**, 1–26.

[24] Kolmogorov, A. N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk SSSR* **30(4)**, 299–303.

[25] Kolmogorov, A. N. (1962). A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *Journal of Fluid Mechanics*, **13(01)**, 82–85.

[26] Navier, L. M. H. (1839). Resume des Lecons donnees a I'Ecole des Ponts etChaussees sur I'Application de la Mecanique a I'Etablissement des Constructions et des Machines (new ed.).

[27] Obukhov, A. M. (1941). Spectral energy distribution in a turbulent flow. *Dokl. Akad. Nauk SSSR* **32(1)**, 22–24.

[28] Perry, R. B., Hinton, D. A. and Stuever, R. A. (1997). NASA wake vortex research for aircraft spacing. AIAA paper, 57.

[29] Rajput, B.S. and Rosinski, J. (1989): Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* **82** (3), 451–487.

[30] Ravelet, F., Chiffaudel, A. and Daviaud, F. (2008): Supercritical transition to turbulence in an inertially-driven von Kármán closed flow. *Journal of Fluid Mechanics* **601**, 339–364.

[31] Reynolds, Osborne (1883). "An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels". *Philosophical Transactions of the Royal Society* **174 (0)**: 935–982

[32] Richardson, L. F. (1922). Weather prediction by numerical process. *Cambridge University Press.*

[33] Schmiegel, J., Cleve, J., Eggers, H. C., Pearson, B. R. and Greiner, M. (2004). Stochastic energy-cascade model for (1+ 1)-dimensional fully developed turbulence. Physics Letters A **320(4)**, 247–253.

[34] She, Z. S., and Leveque, E. (1994). Universal scaling laws in fully developed turbulence. *Physical review letters*, **72(3)**, 336.

[35] Stokes, George (1851). "On the Effect of the Internal Friction of Fluids on the Motion of Pendulums". *Transactions of the Cambridge Philosophical Society* **9**, 8–106.

[36] Taylor, G. I. (1938). The spectrum of turbulence. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **164**, 476–490