

## Can high-order moments be meaningfully estimated from experimental turbulence measurements?

T. Dudok de Wit\*

Laboratoire de Physique et Chimie de l'Environnement, CNRS and University of Orléans, 3A, Avenue de la Recherche Scientifique, F-45071 Orléans Cedex 2, France

(Received 8 December 2003; revised manuscript received 7 September 2004; published 18 November 2004)

Although high-order moments are widely used in the study of fully developed turbulence, their statistical properties remain poorly known. It is well known that beyond a given order, moment estimates based on finite samples cannot be trusted. We provide an empirical criterion for determining that order and illustrate it using a long record of boundary layer turbulence. The results show that even with modest levels of intermittency, structure functions in the inertial range of turbulence cannot be meaningfully assessed for orders as low as 5 or 6.

DOI: 10.1103/PhysRevE.70.055302

PACS number(s): 47.27.-i, 05.40.-a

For several decades, the modeling and the experimental measurement of moments have been a key problem in turbulence research [1,2]. The attention has progressively shifted to high-order moments, as these can help discriminate concurrent models of turbulent cascades. Yet, these developments are often pursued without substantiating evidence that such moments can actually be measured experimentally.

More than three decades ago, Tennekes and Wyngaard [3] warned against the danger of inferring high-order moments from experimental data. Although many pitfalls can, nowadays, be overcome by adequate experimental set up and suitable data processing, two basic problems remain: the lack of ergodicity and the finite sample size. Both have been overlooked, if not neglected, in the literature.

The validation of moment estimates is indeed a difficult task; analytical results only exist for distributions that are close to Gaussian [4]. Tennekes and Lumley [5] introduced a method for determining the uncertainty of a  $q$ th order moment by using knowledge of the  $2q$ th order moment and the integral time scale. Their approach, however, requires good estimates of the low-order moments. Qualitative insight can be gained by investigating the probability density of the data [3,6], an approach that has been used by several others [7–9]. Other approaches involve the central limit theorem [10], assume an algebraic decay of the distribution [11], or consider the problem from a dynamical system point of view [12]. We shall focus here on finite sample size effects and derive an empirical criterion for evaluating their impact.

To illustrate our approach, we consider velocity measurements made by hot wire anemometry in a turbulent boundary layer. This data set has already been analyzed in Refs. [12–14]. The air velocity is recorded at a constant rate of 37.5 kHz, giving a string  $\{v_i\}$  of data. Of particular interest for turbulence studies are the wave-field velocity increments  $u_i = |v_{i+\tau} - v_i|$ , so we shall concentrate on the statistical properties of the string of increments  $\{u_i\}_{i=1}^N$ . Let us consider a time lag  $\tau$  of 12 sampling periods to start with, as this value falls right within the inertial range; the length of the string of increments is then  $N=442\,349$ . In what follows, all velocity

increments are normalized versus their standard deviation (before taking the absolute value), so that  $u$  is dimensionless.

The  $q$ th order moment of the velocity increments is better known as the structure function, whose formal definition and empirical estimate are, respectively,

$$S_q = \int_0^\infty p(u)u^q du, \quad (1)$$

$$\hat{S}_q = \frac{1}{N} \sum_{i=1}^N u_i^q. \quad (2)$$

Here,  $p(u)$  is the probability density of  $u$ . In Fig. 1, the estimated probability density is compared to a Gaussian distribution with the same mean and same variance. A weak departure from Gaussianity is apparent, which may be an indication for short-scale intermittency.

Tennekes and Wyngaard [3] suggested validating the moment estimates by plotting the integrand  $p(u)u^q$  for various orders  $q$ . The surface spanned by this integrand equals the value of the corresponding structure function. This is illustrated in Fig. 1 [discrete functions are plotted, since histograms were used to estimate  $p(u)$ ].

For low orders ( $q \leq 4$ ), the surface spanned by each integrand is regular and well bounded. As the order increases, however, so does the contribution of rare events until the boundary becomes too ragged for the surface to be well defined.

Among the reasons for this degradation are the large uncertainty associated with rare events and the unavoidable truncation of the integral in Eq. (1). We conclude from Fig. 1 that the highest accessible order should be around  $q=6$ , a value that is quite small as compared to what one would expect from such a long time series. Our objective is to estimate this quantity more objectively.

First, let us reorder the array of velocity increments and rank them in decreasing order:  $u_1 \geq u_2 \geq \dots \geq u_N$ . The empirical structure function of order  $q$  is still given by Eq. (2) but with reordered indices, noted  $k$ . As before, the area spanned by the series  $\{u_k^q\}_{k=1}^N$  converges for large  $N$  toward the value of the  $q$ th order structure function. This area is

\*Electronic address: ddwit@cnrs-orleans.fr

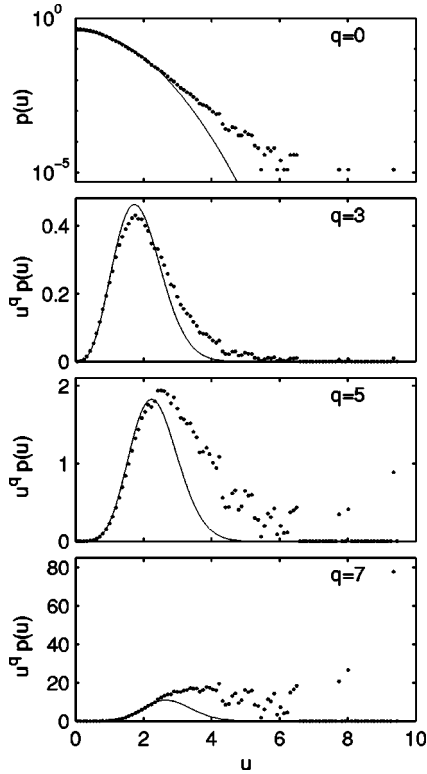


FIG. 1. The integrand of Eq. (1) for various orders  $q$  as measured (dots) and as calculated from a Gaussian distribution (line) whose mean and variance equals that of the measurements. The top panel shows the probability density (vertical axis is logarithmic). In the next panels, the value of the empirical structure function equals the area spanned by the dots. The probability density was estimated using a histogram with 100 equispaced bins; the time lag is  $\tau=12$  sampling periods. Here, and in all following plots, velocities are in dimensionless units.

displayed in Fig. 2. It can be divided into two parts: a long linear tail that is dominated by velocity increments belonging to the bulk of the distribution, and for small  $k$  a sharp peak that is made of rare events. The accuracy of our structure function estimate largely depends on our ability to properly capture the surface of that peak.

It turns out that the peak in Fig. 2 is remarkably well described by a power law,

$$u_k = \alpha \left( \frac{k}{N} \right)^{-\gamma}, \quad (3)$$

as evidenced by a representation with logarithmic axes. Such a scaling invariance would *a priori* be expected only from distributions that exhibit algebraic asymptotic decay, such as Lévy-type distributions. What we observe recalls the empirical Zipf law [15,16], which appears with astonishing ubiquity in rank-ordered statistics [17–20]. The origin of the scaling invariance we observe, however, is essentially rooted in the statistical properties of the extremes in a ranked distribution [21], and so does not necessarily reflect some property of the underlying physics. Indeed, this scaling holds almost regardless of the true probability density of the record. Among the physically meaningful distributions, only those

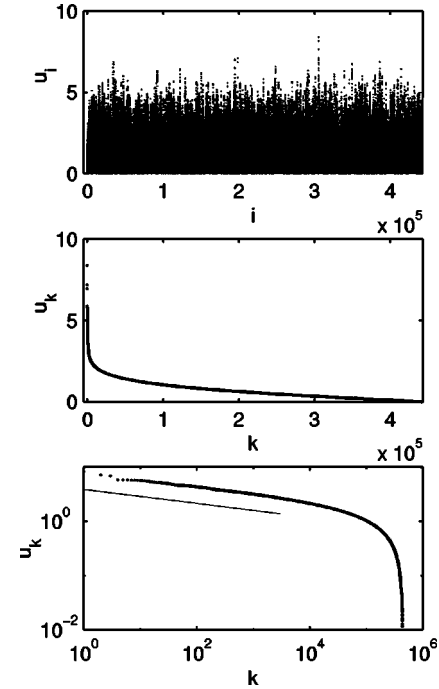


FIG. 2. From top to bottom: the string of velocity increments  $\{u_i\}$  vs their index  $i$ , the string of ranked velocity increments  $\{u_k\}$  with linear and logarithmic axes. A least-squares fit over the range  $10 \leq k \leq 1000$  gives the scaling exponent  $\gamma=0.128 \pm 0.004$  (slope shown by thin line).

which are strictly Gaussian were found to exhibit a clear departure from this scaling. Such distributions, however, are of marginal interest here. We compared a variety of data sets obtained from experimental and simulated neutral fluid and plasma turbulence, using various lags  $\tau$ ; all of them confirmed the robustness of this power-law scaling. We shall therefore take the existence of this scaling as our main working hypothesis.

Let us then assume that the ranked velocity increments obey a power-law scaling for the  $M < N$  first elements. The exact value of  $M$  is not essential for what follows; we merely introduce it to separate the velocity increments into two classes. The empirical structure function can be rewritten as

$$\hat{S}_q = \frac{1}{N} \sum_{k=1}^M u_k^q + \frac{1}{N} \sum_{k=M+1}^N u_k^q, \quad (4)$$

$$= \underbrace{\frac{1}{N} \sum_{k=1}^M \alpha^q \left( \frac{k}{N} \right)^{-q\gamma}}_{\hat{S}_q^{(1)}} + \underbrace{\frac{1}{N} \sum_{k=M+1}^N u_k^q}_{\hat{S}_q^{(2)}}. \quad (5)$$

Rare events contribute to  $\hat{S}_q^{(1)}$ , and the bulk of the distribution to  $\hat{S}_q^{(2)}$ . One can readily show that the second term in Eq. (4) does not significantly depend on  $N$ , as  $u_k \approx (1-k/N)/p(u=0)$ . The first term can be rewritten as

$$\hat{S}_q^{(1)} = \alpha^q N^{q\gamma-1} \sum_{k=1}^M k^{-q\gamma}. \quad (6)$$

For  $0 \leq q\gamma < 2$ , this is well approximated by [22]

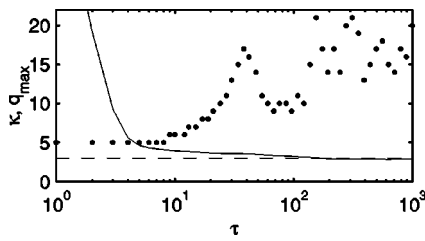


FIG. 3. The maximum order  $q_{max}$  as computed for various time lags  $\tau$  from the same data set (dots). The normalized fourth-order structure function  $\kappa = S_4(\tau)/S_2^2(\tau)$  or kurtosis is shown as a measure of deviation from Gaussianity (full line). For a Gaussian process, the expectation of the kurtosis is 3 (dashed line). Time lags are expressed in units of sampling periods. The inertial range approximately spans from  $\tau=8$  to  $\tau=70$ .

$$\hat{S}_q^{(1)} = \alpha^q N^{q\gamma-1} \left( \frac{7}{12} - \frac{1}{1-q\gamma} [1 - (rN)^{1-q\gamma}] \right), \quad (7)$$

where we introduced  $r=M/N < 1$ . This estimate, to be consistent, should converge toward its expectation as  $N$  increases. The parameters  $\alpha$ ,  $r$ , and  $\gamma$  only weakly depend on  $N$ . A simple consistency criterion thus consists of determining whether  $\hat{S}_q^{(1)}$  converges or diverges for increasing  $N$ . Equation (7) shows that divergence occurs for  $q\gamma > 1$ , with a limiting case for  $q\gamma=1$ . We conclude that the maximum order of the structure function estimate is essentially a function of the scaling index  $\gamma$  of the ranked distribution.

In practice,  $q$  must be an integer, so the maximum order for which a structure function can be meaningfully estimated from a finite data set should be

$$q_{max} = \left\lfloor \frac{1}{\gamma} \right\rfloor - 1, \quad (8)$$

where  $\lfloor \cdot \rfloor$  denotes the integer part.

Applying this criterion to our boundary layer data gives a maximum order of  $q=6$ , in full agreement with the previous qualitative analysis. Its application to a variety of data sets always gave an excellent agreement with the qualitative analysis, even when the power scaling could be applied to a small fraction only (typically a few tens of samples) of the data set.

The main asset of this criterion is its self-consistency, in the sense that no assumptions need to be made either on the validity of the central limit theorem, or on the functional dependence of the tails of the distribution. The main hypothesis is the power-law scaling of the ranked increments  $\{u_k\}$  for small  $k$ .

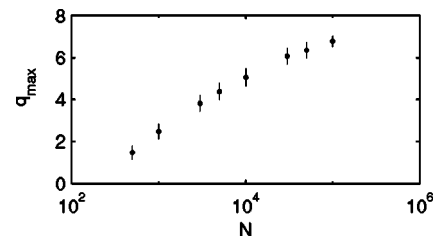


FIG. 4. Scaling of the maximum order  $q_{max}$  with the sample size  $N$ , for  $\tau=12$ . For each value of  $N$ , nonoverlapping sequences were taken from the same time series. Error bars represent  $\pm 1$  standard deviation of  $q_{max}$ .

Clearly, the more significant the tails of the probability density are, the smaller the threshold order  $q_{max}$  will be. Because of that, the maximum order will depend on the time lag  $\tau$ , usually increasing with it until the distribution of the velocity increments becomes Gaussian (see Fig. 3). As a consequence, high-order moments are easier to estimate at large lags, when the distribution is close to Gaussian. The nonmonotonic increase we observe in Fig. 3 is not generic and is most likely due to the lack of exact self-similarity in the inertial range.

These results can be extended to the case of signed velocity increments, the only difference being that both wings of the probability distribution must be treated separately, possibly giving rise to different values of  $q_{max}$ .

Finally, let us investigate how the maximum order depends on the sample size  $N$ . To do so, we estimate  $q_{max}$  from nonoverlapping subsets of various lengths, taken from the same record, for a given lag  $\tau=12$ . Figure 4 shows the average value of  $q_{max}$  and its standard deviation versus the length  $N$  of the subsets. The increase of  $q_{max}$  with  $N$  is rather slow and we conclude that in the inertial range of this particular data set, moments with orders larger than 10 are practically beyond reach.

To summarize, we found a simple and empirical criterion for determining the maximum order for which one can reasonably estimate moments of a given data set. It is based on the observational evidence that the ranked distribution of rare events tends to follow a power law. Even for weakly turbulent fields and long records, the lack of sound statistics on rare events shows that the inference of moments as low as 5 or 6 can be a meaningless task.

*Note added in proof.* A nonparametric estimator of heavy-tailed distributions, based on wavelet transforms, has recently been proposed by Gonçalves and Riedli [23].

I gratefully acknowledge Fabien Anselmet (IRPHE, Marseille) for providing the data and the dynamical systems team (CPT, Marseille) for many stimulating discussions.

- [1] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT Press, Cambridge, MA, 1975).  
 [2] U. Frisch, *Turbulence, the Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, 1995).

- [3] H. Tennekes and J. Wyngaard, *J. Fluid Mech.* **55**, 93 (1972).  
 [4] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1 (Griffin, London, 1977).  
 [5] H. Tennekes and J. L. Lumley, *A First Course in Turbulence*

- (MIT Press, Cambridge, MA, 1972).
- [6] F. N. Frenkiel and P. S. Klebanoff, *Boundary-Layer Meteorol.* **8**, 173 (1975).
  - [7] F. Anselmet, Y. Gagne, E. J. Hopfinger, and R. A. Antonia, *J. Fluid Mech.* **140**, 63 (1984).
  - [8] T. Dudok de Wit and V. V. Krasnosel'skikh, *Nonlinear Processes Geophys.* **3**, 262 (1996).
  - [9] T. S. Horbury and A. Balogh, *Nonlinear Processes Geophys.* **4**, 185 (1997).
  - [10] R. Camussi, C. Baudet, R. Benzi, and S. Ciliberto, *Phys. Rev. E* **54**, R3098 (1996).
  - [11] A. V. Chechkin and V. Y. Gonchar, *Chaos, Solitons Fractals* **11**, 2379 (2000).
  - [12] E. Ugalde, *J. Phys. A* **29**, 4425 (1996).
  - [13] M. Ould-Rouis, F. Anselmet, P. L. Gal, and S. Vaienti, *Physica D* **85**, 405 (1995).
  - [14] S. Vaienti, M. Ould-Rouis, F. Anselmet, and P. L. Gal, *Physica D* **73**, 99 (1994).
  - [15] G. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, MA, 1949).
  - [16] D. Sornette, *Critical Phenomena in Natural Sciences* (Springer, Berlin, 2000).
  - [17] B. Mandelbrot, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk* (Springer, Berlin, 1997).
  - [18] W. Li, *IEEE Trans. Inf. Theory* **38**, 1842 (1992).
  - [19] R. Guenther, L. Levitin, B. Schapiro, and P. Wagner, *Int. J. Theor. Phys.* **35**, 395 (1996).
  - [20] G. Troll and P. beim Graben, *Phys. Rev. E* **57**, 1347 (1998).
  - [21] E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
  - [22] J. Spanier and K. B. Oldham, *An Atlas of Functions* (Springer, Berlin, 1987).
  - [23] P. Gonçalves and R. Riedli (unpublished).