PHYSICAL REVIEW E **70**, 055103(R) (2004)

Increment definitions for scale-dependent analysis of stochastic data

Matthias Waechter,[1,*] Alexei Kouzmitchev,[2,†] and Joachim Peinke[1,‡]
[1]*Institute of Physics, Carl-von-Ossietzky University, D-26111 Oldenburg, Germany*
[2]*Institute of Theoretical Physics, University of Münster, D-48149 Münster, Germany*

(Received 29 April 2004; published 16 November 2004)

It is common for scale-dependent analysis of stochastic data to use the increment $\Delta(t,r)=\xi(t+r)-\xi(t)$ of a data set $\xi(t)$ as a stochastic measure, where r denotes the scale. For joint statistics of $\Delta(t,r)$ and $\Delta(t,r')$ the question of how to nest the increments on different scales r,r' is investigated. Here we show that in some cases spurious correlations between scales can be introduced by the common left-justified definition. The consequences for a Markov process are discussed. These spurious correlations can be avoided by an appropriate nesting of increments. We demonstrate this effect for different data sets and show how it can be detected and quantified. The problem allows to propose a unique method to distinguish between experimental data generated by a noiselike or a Langevin-like random-walk process, respectively.

DOI: 10.1103/PhysRevE.70.055103 PACS number(s): 02.50.−r, 05.10.−a, 95.75.Wx

I. INTRODUCTION

The complexity of most disordered systems depends on the scale at which they are observed. Therefore, stochastic analysis of those systems uses scale-dependent quantities for their characterization. The term "scale" here means for a data set $\xi(t)$ the distance r between two arbitrary points t,t' with $t'-t=r$ (t may denote time as well as space in this context). The *increment* $\Delta(t,r)=\xi(t+r)-\xi(t)$ is a common scale-dependent measure of complexity and disorder. Well-known examples for other scale-dependent measures of complexity are the autocorrelation function $R(r)=\langle\xi(t)\xi(t+r)\rangle$, the rms width $w_r(t)=\langle[\xi(t)-\xi]^2\rangle_r^{1/2}$, or wavelet functions.

Traditionally, the investigation of statistical properties is performed on distinct scales, e.g., by means of the structure functions $\langle\Delta(t,r)^n\rangle$ given by the probability density functions (PDF) $p(\Delta(t,r))$. An advanced approach is to try to describe the joint statistics of the chosen measure on many different scales. This is achieved by the knowledge of the joint PDF $p(\Delta(t,r_1);\ldots;\Delta(t,r_n))$. By these joint PDF also the correlations between scales are worked out, showing how the complexity is linked between scales.

If the statistics of the scale-dependent measure can be regarded as a Markov process evolving in r, the knowledge of two-scale conditional PDF is sufficient for a complete description of multiscale joint PDF [1]. The conditional PDF $p(\Delta_1(t,r_1)|\Delta_0(t,r_0))$ denotes the probability of finding an increment $\Delta(t,r_1)=\Delta_1$ on the scale r_1 under the condition that at the same time t on a different scale r_0 another increment $\Delta(t,r_0)=\Delta_0$ has been found. The validity of the Markov property can be tested by the investigation of conditional PDF [2], of the Chapman-Kolmogorov equation [3], or of reconstructed noise [4]. If, furthermore, the noise involved in the process is Gaussian distributed, the whole joint statistics can be grasped by a Fokker-Planck or Langevin equation [1–5]. This approach has been used by different researchers in a number of applications [2–9]. In some cases also the question of increment definitions has been discussed, as in [6].

In this paper we want to address the question of whether the relative location of the increments may introduce spurious correlations between different scales. In particular, we investigate the two cases of the left-justified increment

$$\Delta_l(t,r)=\xi(t+r)-\xi(t) \tag{1}$$

and the centered increment

$$\Delta_c(t,r)=\xi(t+r/2)-\xi(t-r/2). \tag{2}$$

We will discuss the implications of these increment definitions on two different types of stochastic processes and compare the results for experimental data.

II. INCREMENT DEFINITIONS AND CORRELATIONS BETWEEN SCALES

First, two idealized types of stochastic processes for discrete times are defined. The first one is called white-noise-like random walk (WNR) with the random variable $\xi(t_i)$ given by

$$P(\xi,t_i)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\xi^2/(2\sigma^2)\},$$

$$P(\xi,t_i;\xi',t_k)=P(\xi,t_i)P(\xi',t_k)\quad \text{for } i\neq k, \tag{3}$$

which implies that

$$\langle\xi(t_i)\xi(t_k)\rangle=\sigma^2\delta_{ik} \tag{4}$$

$$\text{and } \langle\xi(t_k)|\xi(t_i)=x_0\rangle=0\quad \text{for } k>i. \tag{5}$$

The second model, called Langevin-like random walk (LRW), is the cumulative sum of the first one, i.e.,

---

*Electronic address: matthias.waechter@uni-oldenburg.de
†Electronic address: kuz@uni-muenster.de
‡Electronic address: peinke@uni-oldenburg.de

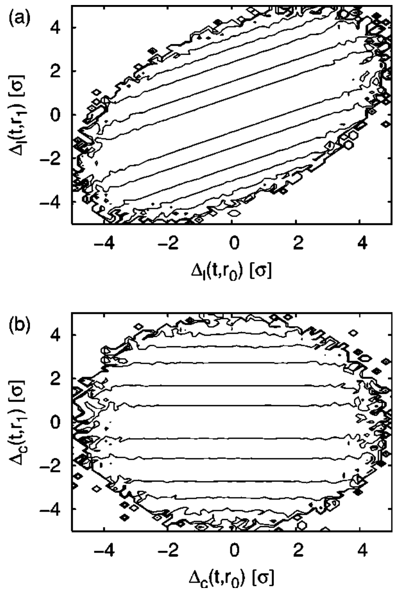

FIG. 1. Conditional PDF of (a) left-justified and (b) centered increments of a white-noise-like random walk. Scales r_0 and r_1 are 116 and 100 sample steps, respectively. PDF are displayed as contour lines, the levels differ by a factor of 10, with an additional level at 0.3.

$$\xi(t_i) = \sum_{k=1}^{i} w(t_k), \tag{6}$$

with $w(t_k)$ distributed as in (3). This implies that here,

$$\langle \xi(t_i)\xi(t_k)\rangle = i\,\sigma^2 \quad \text{for } k \geqslant i \tag{7}$$

$$\text{and } \langle \xi(t_k) | \xi(t_i) = x_0 \rangle = x_0 \quad \text{for } k > i. \tag{8}$$

The correlations between different scales for both increment functions $\Delta_l(t,r)$ and $\Delta_c(t,r)$ can be calculated easily. Taking increments of the WNR (3) and $r_1 > r_0 > 0$ we get

$$\langle \Delta_l(t,r_0)\Delta_l(t,r_1)\rangle = \sigma^2,$$

$$\langle \Delta_c(t,r_0)\Delta_c(t,r_1)\rangle = 0. \tag{9}$$

Note that for arbitrary different scales of Δ_l correlations are present even though there are no correlations at all in the data $\xi(t)$. These spurious correlations are introduced by the left-justified increment because for a fixed value t the increments $\Delta_l(t,r)$ of all scales r have the term $\xi(t)$ in common, see Eq. (1). For the LRW we obtain

$$\langle \Delta_l(t,r_0)\Delta_l(t,r_1)\rangle = \langle \Delta_c(t,r_0)\Delta_c(t,r_1)\rangle = r_0\sigma^2 \tag{10}$$

with $r_1 > r_0 > 0$. In contrast to the WNR, identical correlations of both increment definitions result here.

Correlations between two scales r_0, r_1 can directly be observed as dependence of $p(\Delta_1 | \Delta_0)$ on Δ_0. Increment statistics of the WNR are presented in Fig. 1 as conditional PDF of left-justified and centered increments. Data have been generated using the program GASDEV from [10] and normalized by σ. As expected from Eq. (9), for the *left-justified* increments (a) a correlation between both scales r_1, r_0 is evident because $p(\Delta_l(t,r_1) | \Delta_l(t,r_0))$ strongly depends on the value of $\Delta_l(t,r_0)$. In contrast, the conditional PDF of the *centered* increments (b) is independent of $\Delta_c(t,r_0)$ and thus both scales are uncorrelated for centered increments. For corresponding diagrams calculated for the LRW process (not shown here) we find identical PDF for left-justified and centered increments, similar to Fig. 1(a), and in accordance with Eq. (10).

III. CONSEQUENCES FOR MARKOV PROPERTIES

For the analysis and description of stochastic data by means of a Fokker-Planck equation, as mentioned in the Introduction, the underlying process has to be Markovian. In the previous section we have shown that the left-justified increment definition can introduce additional correlations. This effect also influences the Markov properties of these increments.

It is indeed straightforward to see that the left-justified increment $\Delta_l(t,r)$ of a WNR is consequently not a Markov process in the scale variable r. A necessary condition for a stochastic process to be Markovian is the Chapman-Kolmogorov equation [1] (here we use the notation $\Delta_l(t,r_i) = \Delta_{l,i}$)

$$p(\Delta_{l,3}|\Delta_{l,1}) = \int_{-\infty}^{\infty} p(\Delta_{l,3}|\Delta_{l,2})p(\Delta_{l,2}|\Delta_{l,1})d\Delta_{l,2} \tag{11}$$

for any triplet $r_1 < r_2 < r_3$.

For the WNR process we first derive from (1) and (3) the correlation matrix $\boldsymbol{S}$ for $\Delta_{l,1}, \Delta_{l,2}$ with elements $s_{ij} = \langle \Delta_{l,i}\Delta_{l,j}\rangle$, $i,j = 1, 2$:

$$\boldsymbol{S} = \begin{pmatrix} 2\sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 \end{pmatrix}. \tag{12}$$

Because the difference of two Gaussian distributed random variables is also Gaussian, we can derive two-dimensional PDF of $\Delta_{l,1}, \Delta_{l,2}$ using the general two-dimensional form of the Gaussian distribution (after [1])

$$p(\Delta_{l,i};\Delta_{l,j}) = \frac{1}{2\pi\sqrt{\det \boldsymbol{S}}}\exp\left\{-\frac{1}{2}\sum_{i,j=1}^{2}(\boldsymbol{S}^{-1})_{ij}\Delta_{l,i}\Delta_{l,j}\right\}. \tag{13}$$

Using (12) and (13), we can now explicitly calculate both sides of Eq. (11), namely

$$p(\Delta_{l,3}|\Delta_{l,1}) = \frac{1}{\sqrt{2\pi}\sqrt{3/2}\sigma}\exp - \frac{(\Delta_{l,3} - \Delta_{l,1}/2)^2}{3\sigma^2} \tag{14}$$

and

$$\int_{-\infty}^{\infty} p(\Delta_{l,3}|\Delta_{l,2})p(\Delta_{l,2}|\Delta_{l,1})d\Delta_{l,2} = \frac{1}{\sqrt{2\pi}\sqrt{15/8}\sigma}\exp - \frac{(\Delta_{l,3} - \Delta_{l,1}/4)^2}{(15/4)\sigma^2}. \tag{15}$$

Obviously the Chapman-Kolmogorov equation (11) is vio-

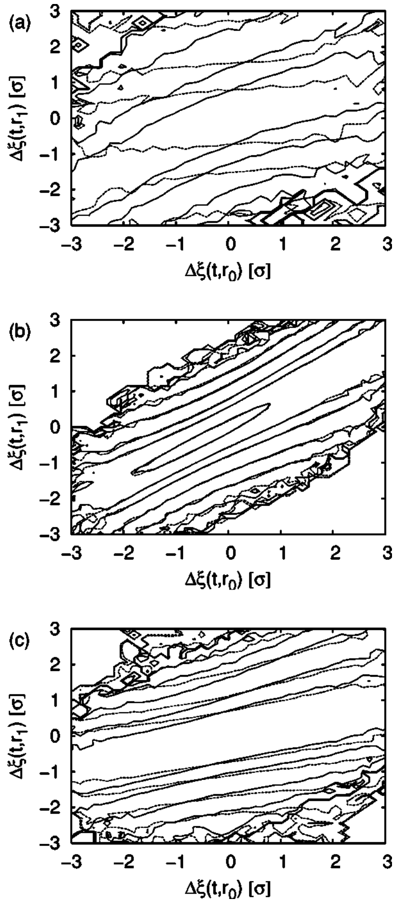

FIG. 2. Conditional PDF of left-justified (solid lines) and centered (broken lines) increments of experimental data. (a) Height profiles from a smooth asphalt road using scales $r_0=137$, $r_1=104$ mm, (b), (c) velocity time series from a turbulent free jet, with small (b) and large (c) scale differences $\delta r=r_0-r_1$ (see text). PDF are plotted as in Fig. 1.

lated for left-justified increments of the WNR on any scales $r_1<r_2<r_3$. The same procedure can be used to see that for centered increments the Markov property holds.

IV. INDICATORS FOR SPURIOUS CORRELATIONS CAUSED BY INCREMENT DEFINITION

The question if, or if not, the above-mentioned spurious correlations between different scales are introduced by the increment definition is of practical importance for the analysis of measured data. For data which behave like the LRW, i.e., $\langle \xi(t+r) | \xi(t)=x_0\rangle = x_0$ rather than $\langle \xi(t+r) | \xi(t)=x_0\rangle = 0$ as for the WNR, the increment definition should be unimportant. No spurious correlations would be created in either case. In contrary, for data which behave more like the WNR, the increment definition should be more important.

As shown above, the conditional PDF can serve as a means to discriminate between true and spurious correlations between scales if we compare conditional PDF of left-justified and centered increments. In Fig. 2 conditional PDF are shown for these increments of two experimental data sets. Figure 2(a) displays PDF of both increment types obtained from surface height profiles of a smooth asphalt road. The distance between consecutive data points is 1.04 mm; further details of the measurement are found in [5,11]. The difference between both types of increments is evident and similar to that for the WNR in Fig. 1. In Figs. 2(b) and 2(c) conditional PDF are shown in the same manner for velocity increments measured in a turbulent free jet at $\mathrm{Re}=2.7\times 10^4$ (for details see [2]). In both cases scale r_1 is $L/2$. The scale difference $\delta r=r_0-r_1$ is small (1.5 λ) for (b) and large (L) for (c) [12]. It can be seen in Fig. 2(b) that here conditional PDF of left-justified and centered increments are identical and the increment definition does not influence the statistics. For $\delta r=L$ at the end of the inertial range, Fig. 2(c), a slight difference of both conditional PDF has already occurred. The definition of L [12] provides that for large scale differences $\delta r > L$ a transition to a noiselike behavior as in Figs. 1 and 2(a) can be expected. Nevertheless, only a small fraction of the correlation between the scales r_0, r_1 is detected as spurious here.

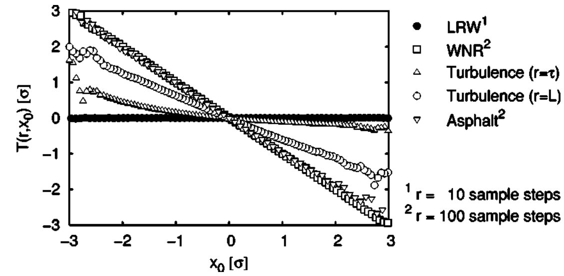

FIG. 3. The conditional expectation value $T(r,x_0)=\langle \Delta_l(t,r) | \xi(t)=x_0\rangle$ for different data sets. T is shown as function of x_0 for fixed scales r.

As a second indicator the conditional expectation value

$$T(r,x_0) = \langle \Delta_l(t,r) | \xi(t) = x_0 \rangle \qquad (16)$$

can be estimated from the measured data. $T(r,x_0)$ quantifies the influence of the value $\xi(t)=x_0$ on the left-justified increment $\Delta_l(t,r)$. It follows immediately that $T(r,x_0)=\langle \xi(t+r) | \xi(t)=x_0\rangle - x_0$. With Eqs. (5) and (8) we obtain the ideal cases $T(r,x_0)=-x_0$ for the WNR and $T(r,x_0)=0$ for the LRW. If for experimental data there is a strong dependence of T on x_0 the data must be regarded as noiselike in the sense of the WNR in the respective length scale, and for scale-dependent analysis the use of left-justified increments is not appropriate. If otherwise T is independent of x_0 the data behave like a LRW, and thus the increment definition is not important.

In Fig. 3 we present the dependence of T on $\xi(t)=x_0$ for different data sets. Data of both ideal cases were generated as in Fig. 1. Turbulence and asphalt road data have already been shown in Fig. 2. As expected, we see that for the LRW there is no dependence of $T(r,x_0)$ on x_0. In contrast, for the WNR as well as for the surface data the dependence is clear with $T(r,x_0)=-x_0$. For the turbulent velocity increments it can be seen that on the small scale $\Delta r=\lambda$ the influence of x_0 on T is only small, while on the large scale $\Delta r=L$ the dependence is more pronounced. This finding corresponds to Fig. 2, where for small scales [Fig. 2(b)] the conditional PDF of left-justified and centered increments were identical, while for large scales [Fig. 2(c)] a difference occurred.

V. CONCLUSIONS

We found that for scale-dependent analysis of stochastic data, where the connections between different scales are investigated using increment statistics, the definition of the increment can be important, depending on the nature of the data. Apparent correlations between scales may be introduced by the left-justified increment. The importance of the increment definition varies between the ideal cases of the LRW (6), where it is nonrelevant, and the WNR (3), where it is crucial. In this case the use of left-justified increments leads to biased results for correlations between different scales. Especially, the surface measurement data we have studied require the centered definition on all accessible scales [5,13]. For turbulent velocities this influence depends on the regarded length (or time) scale r. In previous publications [2,14] no significant difference between the drift and diffusion coefficients of the Fokker-Planck equation of Δ_l and Δ_c was found. This is in accordance with our findings in Fig. 2(b), where the PDF of Δ_l and Δ_c are shown to be identical for small scale differences $r_0 - r_1$, and only at the integral length scale a difference occurs [see Fig. 2(c)]. Detailed consequences are currently being investigated [15].

The conditional expectation value $T(r,x_0)$ allows one to quantify the influence of a left-justified increment. Nevertheless, the specification of a threshold in a statistically meaningful way is still an open question.

While in this paper we used the increments (1) and (2) to demonstrate the introduction of spurious correlations, we expect that these considerations can be applied to general scale-dependent measures of complexity, such as the rms width $w_r(t) = \langle (\xi(t) - \overline{\xi})^2 \rangle_r^{1/2}$ or wavelet functions. One could generally distinguish between measures which are orthogonal on different scales and those which are not [16]. We expect similar results for correlations between scales as demonstrated here for left-justified and centered increments.

ACKNOWLEDGMENTS

We experienced helpful discussions with R. Friedrich, M. Siefert, M. Haase, and A. Mora. Financial support by the German Volkswagen Foundation is kindly acknowledged.

[1] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).

[2] C. Renner, J. Peinke, and R. Friedrich, J. Fluid Mech. **433**, 383 (2001).

[3] R. Friedrich and J. Peinke, Physica D **102**, 147 (1997).

[4] P. Marcq and A. Naert, Phys. Fluids **13**, 2590 (2001).

[5] M. Waechter, F. Riess, H. Kantz, and J. Peinke, Europhys. Lett. **64**, 579 (2003).

[6] A. Naert, R. Friedrich, and J. Peinke, Phys. Rev. E **56**, 6719 (1997).

[7] G. R. Jafari, S. M. Fazeli, F. Ghasemi, S. M. Vaez Allaei, M. R. R. Tabar, A. Iraji Zad, and G. Kavei, Phys. Rev. Lett. **91**, 226101 (2003).

[8] F. Ghasemi, A. Bahraminasab, S. Rahvar, and M. Reza Rahimi Tabar, e-print astro-ph/0312227.

[9] M. Ausloos and K. Ivanova, Phys. Rev. E **68**, 046122 (2003).

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, England, 1992).

[11] M. Waechter, F. Riess, and N. Zacharias, Veh. Syst. Dyn. **37**, 3 (2002).

[12] The Taylor length λ and the integral length L denote the lower and upper bound of the so-called inertial range of scales, where a universal behavior of turbulent flows is found. L is defined by the autocorrelation function and estimates a correlation length such that the turbulent fluctuations on scales $r > L$ can be considered as uncorrelated.

[13] M. Waechter, F. Riess, T. Schimmel, U. Wendt, and J. Peinke, Eur. Phys. J. B (to be published), e-print physics/0404015.

[14] C. Renner, Ph.D. thesis, Carl-von-Ossietzky University, Oldenburg, Germany, 2001, http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2002/renmar02/renmar02.html.

[15] M. Siefert and J. Peinke, e-print physics/0409035.

[16] Orthogonality can be defined if we construct for $\Delta_l(t,r)$ a generating function $F_l(t,r,t') = \delta(t+r,t') - \delta(t,t')$ such that $\Delta_l(t,r) = \int F_l(t,r,t')\xi(t')\mathrm{d}t'$. Analogously $F_c(t,r,t') = \delta(t+r/2,t') - \delta(t-r/2,t')$ is constructed for the centered increment $\Delta_c(t,r)$. Now, for any $r_0 \neq r_1, r_0, r_1 \neq 0$ the scalar product $F_l(t,r_0,t') \cdot F_l(t,r_1,t') = \int F_l(t,r_0,t') F_l(t,r_1,t')\mathrm{d}t'$ for the left-justified increment is obviously different from zero, while $F_c(t,r_0,t') \cdot F_c(t,r_1,t') = 0$.