

**Improved estimation of Fokker-Planck equations through optimization**

A. P. Nawroth and J. Peinke

*Institut for Physics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany*

D. Kleinhans and R. Friedrich

*Institut for Theoretical Physics, University of Münster, D-48149 Münster, Germany*

(Received 11 May 2007; published 5 November 2007)

An improved method for the description of hierarchical complex systems by means of a Fokker-Planck equation is presented. In particular the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm for constraint problems is used to minimize the distance between the numerical solutions of the Fokker-Planck equation and the empirical probability density functions and thus to estimate properly the drift and diffusion term of the Fokker-Planck equation. The optimization routine is applied to a time series of velocity measurements obtained from a turbulent helium gas jet in order to demonstrate the benefits and to quantify the improvements of this optimization routine.

DOI: [10.1103/PhysRevE.76.056102](https://doi.org/10.1103/PhysRevE.76.056102)

PACS number(s): 02.50.Ey, 05.45.Tp, 05.10.Gg

**I. INTRODUCTION**

Most complex systems can be assigned to the two following classes: the time dependent complex systems and the scale dependent complex systems [1]. Examples for the first class are nonlinear chaotic dynamical systems, while systems with a scaling behavior over a wide range of different scales, like turbulence, financial markets, or earthquakes are examples for the second class. Besides the characterization of new features of these complex systems it is a challenge to derive effective underlying equations for their description. A successful approach to such systems is a description through stochastic equations (Langevin or corresponding Fokker-Planck equations), which may involve nonlinearity in the deterministic as well as in the stochastic part. This approach has become particularly interesting, as it has been shown that it is possible to estimate the underlying stochastic equations directly by data analysis.

The verification of the preconditions and the application of this approach to time dependent systems has been described in [2–6]. It was successfully applied to the description of noisy electrical circuits [4], systems with feedback delay [7], traffic flow data [8], physiological time series [9,10], and metrological processes [11], to mention just a few. Also the second class, the scale dependent complex systems, which in general are not stationary in scale, can be analyzed successfully by this approach. In this class, stochastic processes evolving in scale are reconstructed. A complete statistical description, i.e., general  $n$  scale joint statistics, for certain classes of systems, such as the roughness of surfaces [12,13], turbulence [14–17], cosmic microwave background radiation [18], and finance [19–21], can be obtained. Though in general a reconstruction of time series for the scale dependent complex systems is not possible in such a simple way as for the first class, certain promising attempts have been made [13,22].

The use of Langevin and Fokker-Planck equations is therefore a very promising method for time series analysis. The critical part in this method is the correct estimation of the coefficients of the Langevin or the corresponding Fokker-Planck equation, which are the so-called Kramers-Moyal co-

efficients. A correct estimation of these coefficients is crucial to a good description of the underlying processes. The estimation of the Kramers-Moyal coefficients is complicated by the fact that the approach itself is based on the assumption of Markov properties. This assumption is valid for many systems for big and small but finite time steps. The main difficulties arise from the fact that for the estimation of the underlying equations it is necessary to calculate the limit of infinitely small time steps, where the Markov properties are often no longer valid. For more details concerning this discussion see [23–25]. Further concerns about systematic estimation problems were discussed in [26]. Due to these problems in the estimation process it was necessary until now to apply manual corrections [20] to the determined Kramers-Moyal coefficients in some cases in order to get an optimal description. For time dependent systems these problems were addressed in [27] by proposing an improved estimation method for the necessary parameters. This improved estimation method utilizes the comparison of the probability density functions (PDFs) generated by the Langevin equation and those computed directly from the empirical data.

In this work we address the crucial problem of the correct estimation of the coefficients for the Fokker-Planck equation for the second class of systems with scale dependent complexity. It should be noted that there is no principle problem to transfer the results of this work to the class of time dependent complex systems. Furthermore, the methods proposed in this work can be regarded as a systematic way to include the manual corrections described, for example, in [20]. In general for the class of scale dependent complex systems the processes are not stationary in scale and so it is often necessary to use numerical instead of analytic solutions. Therefore the knowledge of the solution to the Fokker-Planck equation will not be global but pointwise in the space spanned by the coefficients of the Fokker-Planck equation. In order to utilize the comparison of the PDFs, as it has been suggested for the first class of complex systems [27], optimization routines are proposed to find the optimal set of parameters, which implies the best agreement between the numerical solutions of the Fokker-Planck equation and the PDFs computed directly from the data.

In detail, in Sec. II the basic features for stochastic processes evolving in scale are discussed. A description of the optimization routines is given in Sec. III. First results for turbulence data are shown in Sec. IV and insights are pointed out for this type of data. Further applications of the discussed methods are shown in Sec. V. We finish with some concluding remarks in Sec. VI.

## II. FOKKER-PLANCK EQUATION

We start with a situation where for a complex system some amount of data  $x(t)$  is given. Here  $x$  denotes the describing quantity, such as heights for surfaces or velocity for turbulent fields and  $t$  denotes a time or a space variable. For simplicity we assume that  $x$  is a one-dimensional quantity, noting that higher dimensional systems can be treated in a similar way [28]. The scale dependent features are described by  $y(t, \tau)$ , where  $\tau$  denotes the selected scale and  $y$  a quantity describing the disorder (complexity) of  $x(t)$  in a  $\tau$ -neighborhood.  $y$  may be a local roughness or any other local quantity (see, for example, [29]). Here we define  $y$  as a simple increment

$$y(t, \tau) := x(t + \tau) - x(t). \quad (1)$$

In order to obtain a statistically complete description of the system with respect to  $y$ , the joint probability density function  $p(y_1, \tau_1, \dots, y_n, \tau_n)$  of  $y$  at different scales  $\tau_i$  has to be known. The joint PDF is constructed from the set of  $y_i$  obtained at the same  $t$  value. In the following the joint statistics of these increment processes are considered. Because of the involved scales the dimension of the joint probability density can be very high. Therefore it is in general very difficult to compute this joint PDF from empirical time series. However, the description and computation can be highly simplified if Markov properties hold. This is the case if

$$p(y_i, \tau_i | y_{i+1}, \tau_{i+1}, \dots, y_n, \tau_n) = p(y_i, \tau_i | y_{i+1}, \tau_{i+1}) \quad (2)$$

is true for all  $i$  and  $n > i$ . Without loss of generality we take  $\tau_i < \tau_{i+1}$ . It should be noted that the Markov property can be tested for a given data set [3,15,17,20,30]. For valid Markov properties the joint probability density can be substantially simplified:

$$p(y_1, \tau_1, \dots, y_n, \tau_n) = p(y_1, \tau_1 | y_2, \tau_2) \cdots p(y_{n-1}, \tau_{n-1} | y_n, \tau_n) \times p(y_n, \tau_n). \quad (3)$$

Because the conditional PDFs of first order [the right side of Eq. (2)] provide a complete description of a Markov process, they are the basic quantity to measure the correctness of the description of a Markov process. This issue and the importance of using conditional PDFs and not unconditional PDFs for the verification of the estimated process are discussed in [15,19,20,31].

For Markov processes the conditional probability density satisfies a master equation, which can be put into the form of a Kramers-Moyal expansion for which the Kramers-Moyal coefficients  $D^{(k)}(y, \tau)$  are defined as the limit  $\Delta\tau \rightarrow 0$  of the conditional moments:

$$D^{(i)}(x, \tau) = \lim_{\Delta\tau \rightarrow 0} \frac{\tau}{i! \Delta\tau} \int_{-\infty}^{+\infty} (x' - x)^i p(x', \tau - \Delta\tau | x, \tau) dx'. \quad (4)$$

For a general stochastic process, all Kramers-Moyal coefficients are different from zero. According to Pawula's theorem, however, the Kramers-Moyal expansion stops after the second term, provided that the fourth order coefficient  $D^{(4)}(y, \tau)$  vanishes [15,16,32]. In that case, the Kramers-Moyal expansion reduces to a Fokker-Planck equation [32,33]

$$-\tau \frac{\partial}{\partial \tau} p(y, \tau | y_0, \tau_0) = \left[ -\frac{\partial}{\partial y} D^{(1)}(y, \tau) + \frac{\partial^2}{\partial y^2} \times D^{(2)}(y, \tau) \right] p(y, \tau | y_0, \tau_0). \quad (5)$$

Note that, in contrast to the usual definition of the Fokker-Planck equation, here both sides are multiplied by  $\tau$ . This corresponds to a logarithmic length scale as used in [34]. This choice is convenient for analyzing fractal scaling features, but does not imply any loss of generality.  $D^{(1)}(x, \tau)$  and  $D^{(2)}(x, \tau)$  are the drift or diffusion coefficients, respectively. It should be noted that the conditional PDF  $p(x', \tau - \Delta\tau | x, \tau)$  can be estimated directly from the data, and therefore the Kramers-Moyal coefficients can be determined by using Eq. (4).

## III. OPTIMIZATION

For the optimized estimation of the Kramers-Moyal coefficients an implementation of the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm for constraint problems (L-BFGS-B algorithm) [35–38] is used, which is described in the Appendix in detail. The starting point is the approximation of the Kramers-Moyal coefficients determined by the evaluation of Eq. (4). The coefficients are approximated by functions with free parameters  $q_i^{(j)}$ ,

$$\tilde{D}^{(1)}(y, \tau) = f(y, \tau, q_0^{(1)}, \dots, q_m^{(1)}), \quad (6)$$

$$\tilde{D}^{(2)}(y, \tau) = g(y, \tau, q_0^{(2)}, \dots, q_n^{(2)}). \quad (7)$$

Solving Eq. (5) as proposed in [15] by using  $\tilde{D}^{(1)}(y, \tau)$  and  $\tilde{D}^{(2)}(y, \tau)$  as the drift and diffusion coefficient, respectively, leads to a conditional PDF of first order  $p_{num}(y_{i-1}, \tau_{i-1} | y_i, \tau_i, q_0^{(1)}, \dots, q_m^{(1)}, q_0^{(2)}, \dots, q_n^{(2)})$ . In order to compare the numerical solutions of the Fokker-Planck equation with empirical probabilities, it is better to use joint instead of conditional probabilities. Therefore the two following joint PDFs will be compared:

$$p_n = p_{num}(y_{i-1}, \tau_{i-1} | y_i, \tau_i, q_0^{(1)}, \dots, q_n^{(2)}) p(y_i, \tau_i), \quad (8)$$

$$p_{ref} = p(y_{i-1}, \tau_{i-1}; y_i, \tau_i). \quad (9)$$

As a measure, to calculate the agreement between these two, the weighted mean square error in logarithmic space is used, which is defined as

$$d_M(p_n, p_{ref}) := \frac{\int_R (p_n + p_{ref})(\ln p_n - \ln p_{ref})^2 dy_i dy_{i-1}}{\int_R (p_n + p_{ref})(\ln^2 p_n + \ln^2 p_{ref}) dy_i dy_{i-1}}. \quad (10)$$

Here  $R$  denotes the subspace, where an estimate of  $p_{ref}$  from empirical data is possible and  $p_{ref} > 0$ . The L-BFGS-B algorithm minimizes the nonlinear function  $d_M(p_n, p_{ref}, q)$  under the constraint that  $D^{(2)}(y, \tau)$  remains positive and the slope of  $D^{(1)}(y, \tau)$  with respect to  $\tau$  remains negative. This may be denoted for each component  $i$  of  $q$  as  $L_i \leq q_i \leq U_i$ , where  $q$  is the  $N_q = \tilde{m} + \tilde{n}$  dimensional vector of all the parameters  $q_0^{(1)}, \dots, q_{\tilde{m}}^{(1)}, q_0^{(2)}, \dots, q_{\tilde{n}}^{(2)}$  and  $L$  and  $U$  are chosen in accordance with the constraints.

#### IV. RESULTS FOR TURBULENCE

The procedure described above is now applied to experimental data. The data considered were obtained from a cryogenic axisymmetric helium gas jet at a Reynolds number of  $7.6 \times 10^5$ . The data set contain  $1.6 \times 10^7$  measurements of the velocity in the center of a free jet, where the distance between the anemometer and the nozzle was 40  $D$  and the diameter  $D$  of the nozzle was 2 mm. The Markov-Einstein coherence length  $\tau_{Markov}$  is 8 sample steps for this data set.  $\tau_{Markov}$  denotes the length, below which the data set does not exhibit Markov properties, as defined in Eq. (2). The integral length, which corresponds to the largest correlated structure, is calculated from the autocorrelation function and is for our data 715 sample steps. For further details we refer to [39]. This high Reynolds number data set has the benefit that the region between the Markov-Einstein coherence length [30] and the integral length spans a large interval of scales. This is important because below the Markov-Einstein coherence length Eq. (5) cannot be applied due to the missing Markov properties. Above the integral length the properties of the turbulent cascade are not present anymore.

A first approximation of the Kramers-Moyal coefficients, for the turbulence data used here, is determined by means of their definition in Eq. (4). Examining these first estimates of the Kramers-Moyal coefficients for a fixed scale  $\tau$ , it becomes evident that the first coefficient can be represented very well with a polynomial of first order and the second one with a polynomial of second order, cf. also [15]. Therefore the Kramers-Moyal coefficients are parametrized as

$$\tilde{D}^{(1)}(y) = q_0^{(1)} + q_1^{(1)}y, \quad (11)$$

$$\tilde{D}^{(2)}(y) = q_0^{(2)} + q_1^{(2)}y + q_2^{(2)}y^2, \quad (12)$$

for a fixed scale  $\tau$ . Examining the dependence of the Kramers-Moyal coefficients on the scale  $\tau$ , the picture is much more ambiguous. Therefore the range of scales is divided into small half-open intervals  $[\tau_i^{left}, \tau_i^{right}]$ , where  $\tau_i^{left}$  denotes the left and  $\tau_i^{right}$  the right border of the scale interval. The optimization is performed for each scale inter-

val independently. Thus as a good local approximation there is no dependency of the Kramers-Moyal coefficients on the scale in each of these intervals, but arbitrary changes are allowed if going from one scale interval to the next. In such a way we account for a more complex dependency on the scale.

Though working with constant coefficients  $q_i^{(j)}$  in each scale interval constitutes an approximation it has two advantages. The first is that the smoothness of the resulting functions  $q_i^{(j)}$  with respect to the scale  $\tau$  provides a first assessment of the robustness of the optimization process, if we assume the true coefficients to be smooth functions with respect to the scale. The second and more important advantage is that the number of variables  $N_q$ , or in other words, the dimension of the space where the optimization takes place, is smaller. This is important because the optimization in a lower dimensional space can be much faster than one in a high dimensional space. In addition the number of local minima may increase rapidly with the addition of more variables and therefore the localization of the global minimum becomes more difficult.

The first estimates are then used to reconstruct the conditional probability density in each scale interval. In order to assess the quality of the solution the distance  $d_M$  between the probability density of the data and the reconstructed probability density by the estimated process equation is calculated. Thereby, the distance  $d_M(p_n, p_{ref})$  given in Eq. (10) is applied. Now the iterative algorithm, described in the Appendix is used to minimize  $d_M(p_n, p_{ref})$  with respect to  $q$ .

The optimization is performed for scales ranging from  $\tau_{Markov}$  to 80 750 sample steps. The scale intervals are chosen here in such a way that

$$\tau_i^{right} = \max \left\{ \frac{\tau_i^{left}}{0.9}; \tau_i^{left} + \tau_{Markov} \right\}. \quad (13)$$

The initial estimate of the Kramers-Moyal coefficients by means of their definition in Eq. (4) can only be performed for a scale larger than  $\tau_{limes}$ , where  $\tau_{limes} \geq \tau_{Markov}$ . In our case  $\tau_{limes} = 5\tau_{Markov}$ . This is due to the procedure to perform the limit in Eq. (4) numerically, for more details see, e.g., [15]. Therefore as initial estimates the values of the Kramers-Moyal coefficients at the scale  $\tau_{limes}$  are used for scales  $\tau \leq \tau_{limes}$ . The limit in Eq. (4) is calculated without the use of possible refinements in order to test the robustness of the optimization routine. The average number of iterations before the optimization stopped was around 25. The values of the distance measure between the PDFs of the original data and the reconstructed ones using the initial estimates of the Kramers-Moyal coefficients are displayed in Fig. 1 as open symbols. It should be noted that each symbol represents one of the scale intervals, where an independent optimization took place.

Three ranges can be identified. The first range spans from  $\tau_{Markov}$  to  $\tau_{limes}$ . Here the limit could not be calculated in a proper way and constant initial estimates of the Kramers-Moyal coefficients have been used, resulting in a nearly constant distance measure in this range. The second range spans from  $\tau_{limes}$  to  $\tau_{optimal}$ , where  $\tau_{optimal}$  is around 300 sample

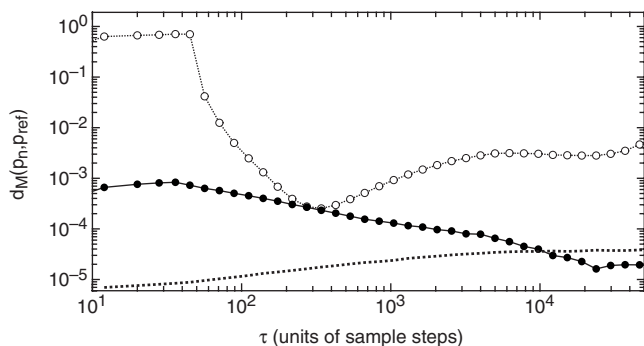


FIG. 1. Distance measure  $d_M$  between the PDFs of the original data and the reconstructed ones using the numerical solutions of the Fokker-Planck equation determined by the initial estimate of the Kramers-Moyal coefficients (open symbols) and the optimized estimate of the Kramers-Moyal coefficients (black dots). The dotted line provides the expected distance if both distributions have been produced by the same process.

steps. We will see later that for  $\tau_{optimal}$  the initial estimates seem already to be optimal. In this range the distance measure decreases monotonically with increasing scale  $\tau$ . This may be due to a better description of the data with increasing  $\tau$ , or due to a better performance of the initial estimate of the Kramers-Moyal coefficients, or due to both. In the third range the distance measure increases after the minimum at  $\tau_{optimal}$ , which marks the border between the second and the third range.

Performing the optimization routine described above, the distance measure between the PDFs of the original data and the reconstructed ones using the optimization routine is obtained. The distance measure for the optimized PDFs is displayed in Fig. 1 as black dots. For very small scales  $\tau \lesssim \tau_{times}$  the distance measure remains constant or increases slightly. For a very broad range of scales the distance measure then declines monotonically, until it saturates for very large scales. Interestingly a scale  $\tau \sim \tau_{optimal}$  exists, where the distance function has approximately the same value for the initially estimated coefficients and the optimized one. This indicates that for  $\tau_{optimal}$  the initial estimate of the Kramers-

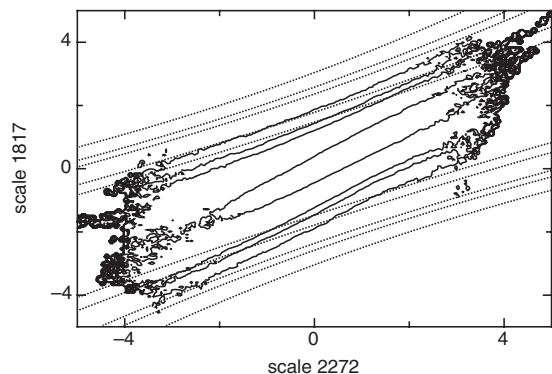


FIG. 2. Conditional probability density  $p[y(\tau=1817)|y(\tau=2272)]$  of given data (solid curves) and reconstructed by the numerical solution of the Fokker-Planck equation (dotted lines) using the initial estimates of the Kramers-Moyal coefficients.

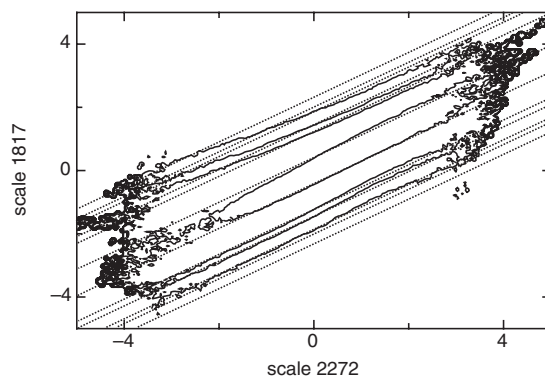


FIG. 3. Conditional probability density  $p[y(\tau=1817)|y(\tau=2272)]$  of given data (solid curves) and reconstructed by the numerical solution of the Fokker-Planck equation (dotted lines) using the optimized estimates of the Kramers-Moyal coefficients.

Moyal coefficients is already optimal, as mentioned above. We obtained similar results for other data sets.

In order to assess the significance of the results the intrinsic error is estimated. The data set is divided into subsets and the distance between the distributions belonging to the corresponding subsets is calculated. This is done for different sizes of subsets and then extrapolated to obtain the intrinsic error for the whole data set. As seen in Fig. 1 the intrinsic error is still smaller for scales up to  $10^4$  than the distance measure for the optimized coefficients, nevertheless to our interpretation the magnitude of the distance measure is with  $10^{-3}$  sufficiently small, see Figs. 2 and 3 for an example.

The graphs for the optimized coefficients  $q_1^{(1)}$ ,  $q_0^{(2)}$ , and  $q_2^{(2)}$  are shown in Figs. 4–6. For  $q_0^{(1)}$  and  $q_1^{(2)}$  the initial estimates as well as the optimized values are essentially equal to zero. This result has an interesting physical context. It has been shown that for a higher dimensional analysis a corresponding nonvanishing  $q_1^{(2)}$  term violates the second von Kármán equation [42] (see [40] and Eqs. (31) and (32) in [28]). In Fig. 6 two ranges can be identified. For smaller scales  $\tau < \tau_{additive}$ , where  $\tau_{additive}$  is around 2000 sample steps,  $q_2^{(2)}$  takes nonvanishing positive values, while it becomes zero for scales  $\tau > \tau_{additive}$ . The other nonvanishing term of the second Kramers-Moyal coefficient,  $q_0^{(2)}$ , in Fig. 5, exhibits in the same region a power-law behavior which

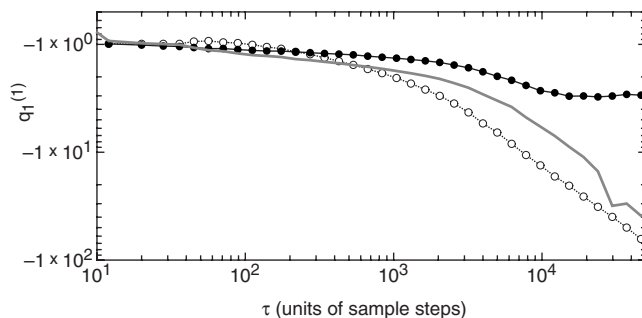


FIG. 4. The parameter  $q_1^{(1)}$ . The initial estimate is denoted with white circles while the optimized one is denoted with black circles. The gray line shows the results for centered increments using optimized coefficients.



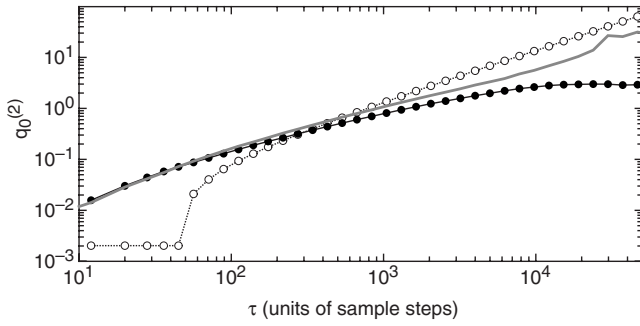


FIG. 5. The parameter  $q_0^{(2)}$ . The initial estimate is denoted with white circles while the optimized one is denoted with black circles. The gray line shows the results for centered increments using optimized coefficients.

saturation for larger scales. The same is true for  $q_1^{(1)}$  in Fig. 4, but with a much higher accuracy. The exponent for  $q_1^{(1)}$  in this region is 0.069.

For the interpretation of  $\tau_{additive}$  we note that  $q_2^{(2)}$  represents the multiplicative noise in the system. Because this term becomes zero for scales larger than  $\tau_{additive}$  and also  $q_1^{(2)}$  is zero, the system exhibits only additive noise above  $\tau_{additive}$ . By investigating the moments of the system, in Fig. 7, it can be noted that the moments start to saturate at a scale, which is comparable to  $\tau_{additive}$ . This indicates that  $\tau_{additive}$  is related to the integral length of the system. Further  $q_2^{(2)}=0$  for  $\tau > \tau_{additive}$  is in agreement with a Gaussian shape of the PDF of velocity increments for large scales [43].

## V. APPLICATIONS

The method above provides a much better answer to the central question of determining the correct Kramers-Moyal coefficients; but besides this it enables us to discuss further important questions arising from the description of scale dependent systems with a Fokker-Planck equation. The first of these questions is the optimal increment definition for the stochastic process as given by Eq. (1). We started our analysis using the left-justified increments, which are more common in the literature. Using left-justified increments means that the smaller increment is nested inside the larger incre-

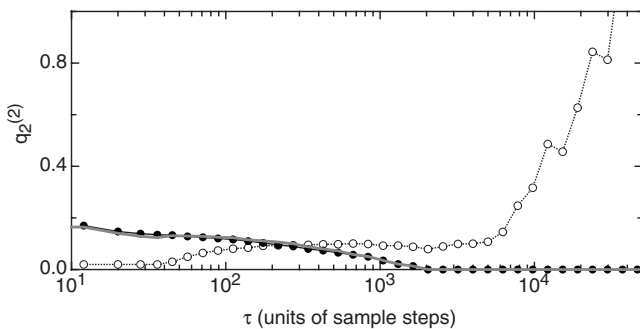


FIG. 6. The parameter  $q_2^{(2)}$ . The initial estimate is denoted with white circles while the optimized one is denoted with black circles. The gray line shows the results for centered increments using optimized coefficients.

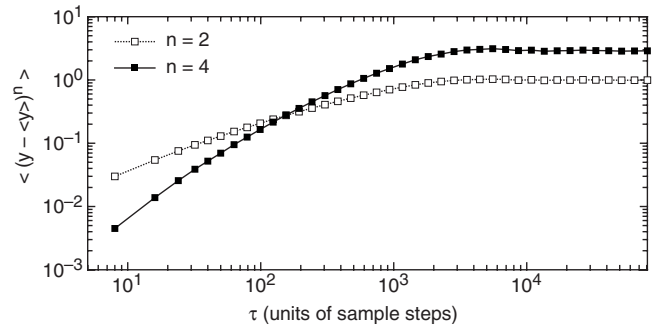


FIG. 7. The second and fourth moment of the data set.

ment and that both increments have the left end point in common. For certain classes of systems this may introduce additional correlations between the increments that are not desired [6]. Thus it has been proposed to use centered increments instead of left-justified ones,

$$y(t, \tau) := x\left(t + \frac{\tau}{2}\right) - x\left(t - \frac{\tau}{2}\right). \quad (14)$$

It is now possible to investigate the improvement of the description of the system by using centered increments. As a criterion we use the distance between the numerical solution of the Fokker-Planck equation, where the coefficients have been optimized, and the empirical PDF. As can be seen in Fig. 8, the distance measure exhibits smaller values for centered increments with the exceptions of very small and very large scales. This indicates that the description of this special system indeed can be improved using centered increments. Reanalyzing the coefficients that are shown in Figs. 4–6 no principal changes are found for the use of centered increments. In Figs. 4–6 additionally the results for the centered increment analysis are shown. The biggest change is found for  $q_1^{(1)}$  and  $q_0^{(2)}$  and for scales larger than the integral scale, which are less important.

The second important question concerns the parametrization of the Kramers-Moyal coefficients. It is now possible to determine whether a more complex parametrization of the Kramers-Moyal coefficients, for example, by using higher

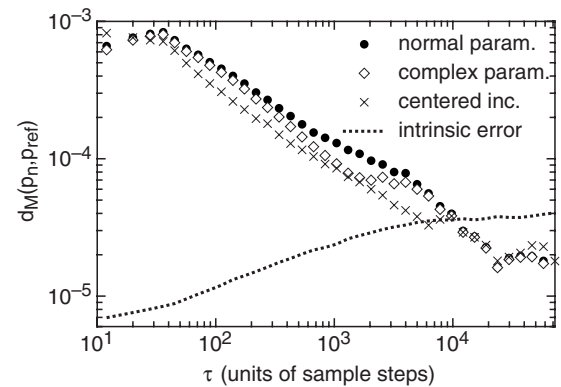


FIG. 8. Comparison of the distance function  $d_M$  for different settings. The dotted line provides the expected distance if both distributions have been produced by the same process.

order polynomials, yields a better description of the system. As a simple example to illustrate this, the question of asymmetric Kramers-Moyal coefficients is considered. If the initial estimates of the Kramers-Moyal coefficients are examined, the functional form appears to be asymmetric in some cases. This seems to be especially true for very large scales, where the number of independent events becomes smaller. In order to verify if the underlying stochastic process can be better described by a separate parametrization for negative and positive increments, the following parametrization is chosen for the optimization:

$$\tilde{D}^{(1)}(y, \tau) = \begin{cases} q_0^{(1)} + q_1^{(1)}y & \text{if } y < 0 \\ q_2^{(1)} + q_3^{(1)}y & \text{if } y \geq 0, \end{cases} \quad (15)$$

$$\tilde{D}^{(2)}(y, \tau) = \begin{cases} q_0^{(2)} + q_1^{(2)}y + q_2^{(2)}y^2 & \text{if } y < 0 \\ q_3^{(2)} + q_4^{(2)}y + q_5^{(2)}y^2 & \text{if } y \geq 0. \end{cases} \quad (16)$$

As depicted in Fig. 8 the distance function takes smaller values than for the original optimization, although the improvement is not as large as when using centered increments. It should further be noted that an improvement is in this case not surprising since the optimization is now performed in a higher dimensional space and the space used for the original optimization is a subspace of this second optimization. Nevertheless this finding is in accordance with the proposed importance of higher odd order terms in the diffusion coefficient [17].

Therefore it may be inferred that a further improvement of the description of this system may be provided by using an appropriate increment definition rather than adopting the assumption of asymmetric Kramers-Moyal coefficients. The first provides a better description of the system by using fewer free variables compared to the second. Another questions that may be answered in such a fashion is the use of higher order Kramers-Moyal coefficients, especially the fourth order coefficient because of its importance for the application of the theorem of Pawula [32].

## VI. CONCLUSIONS

In this work we have shown a practical way to implement an optimization routine to improve the description of hierarchical systems by means of a Fokker-Planck equation. In order to do so, first an estimate of the Kramers-Moyal coefficients using their definition in Eq. (4) is calculated. This initial estimate is then used to solve the Fokker-Planck equation numerically and to obtain as a solution the conditional probability density functions (PDFs) of first order. As a next step the distance between this reconstructed conditional probability and the one obtained directly from the time series is determined using Eq. (10). This procedure forms the basis of our optimization routine. A parametrization of the initial estimate of the Kramers-Moyal coefficients is chosen with a specified number of variables  $N_q$ . The L-BFGS-B algorithm is employed to minimize the distance between the numerical solutions of the Fokker-Planck equation and the empirical PDFs by adjusting the free variables. The L-BFGS-B algorithm is an algorithm which is very effective in the case of an

optimization of many variables, which may be constrained. Therefore the method proposed here will also be effective for very complex parametrization, as long as these parametrizations are not misspecified.

We applied the optimization routine to a time series of velocity measurements obtained from a cryogenic axisymmetric helium gas jet. We demonstrated the benefits of this optimization routine. First, it is possible to obtain values of the Kramers-Moyal coefficients for much smaller scales, due to the fact that it is no longer necessary to calculate a limit in scale which is the bottleneck of the original Kramers-Moyal method. Second, the optimized coefficients produce numerical solutions of the Fokker-Planck equation that are much closer to the empirical PDFs than those produced by the initial estimates. Third, possible systematic errors in the classical estimation routine of the Kramers-Moyal coefficients that have been pointed out in the literature can be avoided using this optimization routine. Fourth, the optimized coefficients show remarkable simple functional forms in a large scaling region, while the behavior of the initial estimates is much more ambiguous. At last the results produced by this optimization routine are remarkably stable. Independent optimizations have been performed for small intervals in scale bordering on each other, producing estimates which are very smooth with respect to the scale. Therefore this method provides the means to determine the Kramers-Moyal coefficients with much more accuracy or to determine correct Kramers-Moyal coefficients for small data sets.

Possible applications for this refined approach have been shown. First, the question of the appropriate increment definition has been considered. It has been shown that by using centered increments instead of left-justified ones, the description of the underlying stochastic process for our example system can be improved. Second, the question of the optimal parametrization of Kramers-Moyal coefficients in the Fokker-Planck equation has been considered. It was shown that in our case an asymmetric parametrization provides only a slight improvement. This aspect interests us because it is directly related to closure of the higher order moments, see Eq. (4.13) in [15]. With our findings here we see that a perturbing linear term for the diffusion coefficient may have no significance; thus the reported contradiction of the reconstructed Fokker-Planck equation with the second Kármán equation seems to have no significance, or saying it in other words, this discrepancy is just a result of an inaccurate estimation. Further applications may include the analysis of more complex parametrizations of the Kramers-Moyal coefficients and the influence of higher order Kramers-Moyal coefficients, thereby offering new insights into the complexity of turbulence; but it should also be mentioned that this method of optimization can be applied to other complex systems, like these mentioned in the Introduction.

## ACKNOWLEDGMENT

We thank B. Chabaud and O. Chanal for providing us with excellent data.

## APPENDIX: L-BFGS-B ALGORITHM

For the optimized estimation of the Kramers-Moyal coefficients we apply an iterative procedure, called L-BFGS-B

algorithm [35–37]. Solving Eq. (5) as proposed in [15] leads to a conditional PDF of first order  $p_{num}(y_{i-1}, \tau_{i-1} | y_i, \tau_i, q)$ . In order to maximize the agreement between the numerical solutions  $p_n$  and the PDFs from the empirical data  $p_{ref}$ , a measure is needed. Here  $d_M(p_n, p_{ref}, q)$  is used, which is defined in Eq. (10). The L-BFGS-B algorithm minimizes the nonlinear function  $d_M(p_n, p_{ref}, q)$ , here  $d(q)$  is used as a shorthand notation, under the constraint  $L \leq q \leq U$ .  $L$  and  $U$  represent the lower and upper bound on  $q$ , respectively.

Using the details provided above, an iterative procedure is started to find the vector  $q$  which minimizes the distance function  $d(q)$ . As a first step of each iteration ( $k$ ) a quadratic model  $m_{(k)}(q)$  of  $d(q)$  at the iterate  $q_{(k)}$  is computed,

$$m_{(k)}(q) := d(q_{(k)}) + g_{(k)}^T(q - q_{(k)}) + \frac{1}{2}(q - q_{(k)})^T B_{(k)}(q - q_{(k)}), \quad (\text{A1})$$

where  $g_{(k)}$  denotes the gradient of  $d(q)$  with respect to  $q$  and  $B_{(k)}$  is a limited-memory BFGS approximation to the Hessian.

Because the following steps have to be repeated for each ( $k$ ), the index ( $k$ ) is omitted for these steps if it does not change the meaning. As a second step a set of active bounds has to be found using the gradient projection method. The projection of an arbitrary point  $q$  onto the feasible region is defined by

$$P(q, L, U)_i = \begin{cases} L_i, & q_i < L_i, \\ q_i & \text{if } q_i \in [L_i, U_i], \\ U_i, & q_i > U_i. \end{cases} \quad (\text{A2})$$

Therefore a piecewise linear path  $q(s)$ , which is the projection of the steepest descent direction at the starting point  $q^0$  onto the feasible region, defined by Eq. (A2), is denoted by

$$q(s) = P(q^0 - sg, L, U). \quad (\text{A3})$$

As a third step the generalized Cauchy point  $q^c$ , which is defined as the first local minimizer of the function  $m_{(k)}(q(s))$  on the piecewise linear path  $q(s)$ , is computed. The components of  $q^c$  which are at their upper or lower bound,  $U$  or  $L$ , comprise the active set  $\mathcal{A}(q^c)$  of variables.

As a fourth step the following quadratic problem over the subspace of free variables is considered.

$$\min\{m_{(k)}(q) | q_i = q_i^c \quad \forall i \in \mathcal{A}(q^c)\}, \quad (\text{A4})$$

$$\text{subject to } L \leq q \leq U \quad \forall i \notin \mathcal{A}(q^c). \quad (\text{A5})$$

Equation (A4) is solved approximately without the condition of Eq. (A5) using a direct primal method [36]. The solution of this unconstrained problem is denoted by  $\delta^u$ . Therefore the solution of the constrained problem can be written as

$$q_{(k+1),i}^s = \begin{cases} q_{(k),i}^c & \text{if } i \notin \mathcal{F}, \\ q_{(k),i}^c + (Z\delta^s)_{(k),i} & \text{if } i \in \mathcal{F}, \end{cases} \quad (\text{A6})$$

where

$$\delta^s = \alpha^* \delta^u \quad (\text{A7})$$

and

$$\alpha^* := \max\{\alpha | \alpha \leq 1, L_i - q_i^c \leq \alpha \delta_i^u \leq U_i - q_i^c, i \in \mathcal{F}\}. \quad (\text{A8})$$

$Z$  denotes the  $N_q \times N_q^c$  matrix of unit vectors (i.e., each column is a column of the identity matrix) that span the subspace of free variables at  $q^c$ , where  $N_q^c$  denotes the number of free variables at  $q^c$ .  $\mathcal{F}$  denotes the set of indices corresponding to the set of free variables.

As a last step a line search between the current  $q_{(k)}$  and the approximate minimizer  $q_{(k+1)}^s$  is performed, which satisfies the strong Wolfe conditions

$$d(q_{k+1}) \leq d(q_k) + c_1 \lambda_k g_k^T (q_{k+1}^s - q_k), \quad (\text{A9})$$

$$|g_{k+1}^T (q_{k+1}^s - q_k)| \leq c_2 |g_k^T (q_{k+1}^s - q_k)|, \quad (\text{A10})$$

where  $\lambda_{(k)}$  is the step length and  $c_1 = 10^{-4}$  and  $c_2 = 0.9$ . Here the algorithm of More and Thuente [41] is used. The solution of this line search is used as the next iterate  $q_{(k+1)}$ .

The optimization procedure is stopped if the value for  $d(q)$  cannot be reduced by a certain percentage, which will be in our case around  $10^{-8}$  or a certain number of iterations will be reached, which is 50.

- 
- [1] J. Peinke, C. Renner, and R. Friedrich, in *Complexity from Microscopic to Macroscopic Scales: Coherence and Large Deviations*, edited by A. T. Skjeltorp and T. Vicsek, NATO Science Series II (Plenum, New York, 2002), Vol. 63, pp. 151–169.
- [2] S. Siegert, R. Friedrich, and J. Peinke, *Phys. Lett. A* **243**, 275 (1998).
- [3] R. Friedrich, J. Zeller, and J. Peinke, *Europhys. Lett.* **41**, 153 (1998).
- [4] R. Friedrich, S. Siegert, J. Peinke, S. Lueck, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister, *Phys. Lett. A* **271**, 217 (2000).
- [5] M. Siefert, A. Kittel, R. Friedrich, and J. Peinke, *Europhys. Lett.* **61**, 466 (2003).
- [6] M. Waechter, A. Kouzmitchev, and J. Peinke, *Phys. Rev. E* **70**, 055103(R) (2004).
- [7] T. D. Frank, P. J. Beek, and R. Friedrich, *Phys. Lett. A* **328**, 219 (2004).
- [8] S. Kriso, J. Peinke, R. Friedrich, and P. Wagner, *Phys. Lett. A* **299**, 287 (2002).
- [9] T. Kuusela, *Phys. Rev. E* **69**, 031916 (2004).
- [10] F. Ghasemi, M. Sahimi, J. Peinke, and M. R. R. Tabar, *J. Biol. Phys.* **32**, 117 (2006).
- [11] P. Sura and S. T. Gille, *J. Mar. Res.* **61**, 313 (2003).
- [12] M. Waechter, F. Riess, H. Kantz, and J. Peinke, *Europhys. Lett.* **64**, 579 (2003).

- [13] G. R. Jafari, S. M. Fazeli, F. Ghasemi, S. M. Vaez. Allaei, M. Reza Rahimi Tabar, A. Iraj Zad, and G. Kavei, *Phys. Rev. Lett.* **91**, 226101 (2003).
- [14] R. Friedrich and J. Peinke, *Phys. Rev. Lett.* **78**, 863 (1997).
- [15] C. Renner, J. Peinke, and R. Friedrich, *J. Fluid Mech.* **433**, 383 (2001).
- [16] M. Tutkun and L. Mydlarski, *New J. Phys.* **6**, 49 (2004).
- [17] P. Marcq and A. Naert, *Phys. Fluids* **13**, 2590 (2001).
- [18] F. Ghasemi, A. Bahraminasab, M. S. Movahed, S. Rahvar, K. R. Sreenivasan, and M. R. R. Tabar, *J. Stat. Mech.: Theory Exp.* P11008 (2006).
- [19] R. Friedrich, J. Peinke, and C. Renner, *Phys. Rev. Lett.* **84**, 5224 (2000).
- [20] C. Renner, J. Peinke, and R. Friedrich, *Physica A* **298**, 499 (2001).
- [21] M. Ausloos and K. Ivanova, *Phys. Rev. E* **68**, 046122 (2003).
- [22] A. P. Nawroth and J. Peinke, *Phys. Lett. A* **360**, 234 (2006).
- [23] M. Ragwitz and H. Kantz, *Phys. Rev. Lett.* **87**, 254501 (2001).
- [24] R. Friedrich, C. Renner, M. Siefert, and J. Peinke, *Phys. Rev. Lett.* **89**, 149401 (2002).
- [25] M. Ragwitz and H. Kantz, *Phys. Rev. Lett.* **89**, 149402 (2002).
- [26] P. Sura and J. Barsugli, *Phys. Lett. A* **305**, 304 (2002).
- [27] D. Kleinhans, R. Friedrich, A. Nawroth, and J. Peinke, *Phys. Lett. A* **346**, 42 (2005).
- [28] M. Siefert and J. Peinke, *J. Turbul.* **7**, 1 (2006).
- [29] M. Karth and J. Peinke, *Complexity* **8**, 34 (2003).
- [30] S. Lück, C. Renner, J. Peinke, and R. Friedrich, *Phys. Lett. A* **359**, 335 (2006).
- [31] C. Renner, J. Peinke, R. Friedrich, O. Chanal, and B. Chabaud, *Phys. Rev. Lett.* **89**, 124502 (2002).
- [32] H. Risken, *The Fokker-Planck Equation* (Springer-Verlag, Berlin, 1996).
- [33] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985).
- [34] R. Friedrich and J. Peinke, *Physica D* **102**, 147 (1997).
- [35] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 1999).
- [36] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *SIAM J. Sci. Comput. (USA)* **16**, 1190 (1995).
- [37] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, *ACM Trans. Math. Softw.* **23**, 550 (1997).
- [38] R-2.2.1 (2006), [www.r-project.org](http://www.r-project.org)
- [39] O. Chanal, B. Chabaud, B. Castaing, and B. Hébral, *Eur. Phys. J. B* **17**, 309 (2000).
- [40] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics: Mechanics of Turbulence*, (MIT Press, Cambridge, MA, 1975), Vol. 2.
- [41] J. J. Moré and D. J. Thuente, *ACM Trans. Math. Softw.* **20**, 286 (1994).
- [42] The Kármán equation is one of the few rigorous results in idealized turbulence, relating longitudinal and transversal increments.
- [43] For turbulence it is a well-known feature that there is a large length scale, the integral length scale, above which the disorder behaves essentially like uncorrelated Gaussian noise.