

## Testing the Markov hypothesis in fluid flows

Daniel W. Meyer\* and Frédéric Saggini

*Institute of Fluid Dynamics, ETH Zürich Sonneggstrasse 3, CH-8092 Zurich, Switzerland*

(Received 14 November 2015; revised manuscript received 21 March 2016; published 9 May 2016)

Stochastic Markov processes are used very frequently to model, for example, processes in turbulence and subsurface flow and transport. Based on the weak Chapman-Kolmogorov equation and the strong Markov condition, we present methods to test the Markov hypothesis that is at the heart of these models. We demonstrate the capabilities of our methodology by testing the Markov hypothesis for fluid and inertial particles in turbulence, and fluid particles in the heterogeneous subsurface. In the context of subsurface macrodispersion, we find that depending on the heterogeneity level, Markov models work well above a certain scale of interest for media with different log-conductivity correlation structures. Moreover, we find surprising similarities in the velocity dynamics of the different media considered.

DOI: [10.1103/PhysRevE.93.053103](https://doi.org/10.1103/PhysRevE.93.053103)

### I. INTRODUCTION

Markov processes are a special category of random processes  $v(t)$  with realization  $V(t)$ , where the future state  $v_1 \equiv v(t_1)$  at time  $t_1$  depends solely on the present state  $v_2$  at time  $t_2 \equiv t_1 - \Delta t$  and no further state in the past, i.e.,  $v_i$  at times  $t_i \equiv t_1 - (i - 1)\Delta t$  with  $i = 3, 4, \dots, n$ . Markov models are often applied as surrogates for complex physical processes, e.g., Brownian motion. Expressed in terms of transition probability density functions (PDFs), Markovianity means

$$p(v_1|v_2, v_3, v_4, \dots) = p(v_1|v_2). \quad (1)$$

Moreover, in a stationary Markov process, the transition PDFs are independent of the actual time, e.g.,  $p(v_1|v_2) = p(v_i|v_{i+1})$  for any  $i$ . Stationary Markov processes are used in a range of different physical applications. Pope [1] summarized corresponding modeling efforts in connection with turbulent dispersion and turbulent reactive flows, and Minier [2] reviewed Lagrangian stochastic methods for turbulent two-phase flows. Specific applications include the models proposed by Pozorski and Minier [3] and Meyer [4] for the fluid velocity seen by inertial particles. More recently a similar model was outlined to describe subgrid scale oscillations of the velocity seen by the particle in two-phase large eddy simulation (LES) [5]. In the context of atmospheric dispersion, stochastic fluid or tracer particles are successfully applied in the U.K. Meteorological Office's NAME model [6]. Moreover, for the simulation of turbulent reactive flows, Pope [7] has developed Markovian fluid particle velocity processes. Lastly, to model fluid particle acceleration fluctuations, Sawford [8] and Zamansky *et al.* [9] have proposed Markov models.

Based on the success of Markov models in turbulence, corresponding models were introduced more recently in the area of subsurface flow and transport. Le Borgne *et al.* [10] and Meyer and Tchelepi [11] introduced Markov velocity processes to model the transport of passive tracers in the heterogeneous subsurface. Both contributions focus on high Péclet numbers, where advection is more important than pore-scale dispersion, which is typical in most applications [12, section 10.5.2]. The presented velocity processes are

useful to gain a better understanding of the underlying physical effects and to arrive at macrodispersion predictions at low computational costs. For example, the refined Markov model outlined by Meyer *et al.* [13] provides predictions that run several orders of magnitude faster compared to conventional Monte Carlo sampling. However, since macrodispersion is governed by the subsurface characteristics that determine the flow field, e.g., the conductivity correlation structure, Markov models will not work equally well for all types of media. For example, Le Borgne *et al.* [10] found that tracer particle dynamics in two-dimensional multi-Gaussian fields with Gaussian correlation structure are non-Markovian in time due to a high degree of correlation at low velocities.

Despite the listed and numerous other examples [14,15], the applicability of the Markov hypothesis is rarely verified most likely due to a lack of a testing framework that is sufficiently general. In the few cases where it is verified though, the Chapman-Kolmogorov (CK) equality

$$p(v_1|v_3) = \int_{\Omega} p(v_1|v_2)p(v_2|v_3) dv_2 \quad (2)$$

is typically used. CK equation (2) is a weak test for the Markov condition

$$p(v_1|v_2, v_3) = p(v_1|v_2), \quad (3)$$

since multiplication of equation (3) with  $p(v_2|v_3)$  and integration over  $v_2$  leads to Eq. (2). CK equation (2) is a weak test of the Markov condition (3), since non-Markovian processes like the ones constructed, for example, by Feller [16], Parzen [17, p. 203], and Van Kampen [18, p. 79] satisfy the CK equation but do not satisfy the Markov condition and therefore are not Markovian. Accordingly, CK equation (2) is a necessary but not a sufficient condition for Markovianity.

In the case of a univariate random process with discrete states, the CK equation reduces to an equality between two transition matrices, i.e.,  $\mathbf{P}_{1,3} = \mathbf{P}_{1,2}\mathbf{P}_{2,3}$ , where the element with index  $(k,l)$  in matrix  $\mathbf{P}_{i,i+j}$  represents the transition probability from state  $k$  at time  $t_i$  to state  $l$  at  $t_{i+j}$  with  $j \geq 1$ . If the transition probabilities can be estimated reliably, the verification of the weak CK equation is straightforward and was, for example, conducted by Le Borgne *et al.* [10, Sec. IV C] and Kang *et al.* [19, Fig. 2] in the context of subsurface flow and transport as well as by Horstmeyer *et al.* [20, Figs. 6

\*meyerda@ethz.ch

and 7] in connection with speckle image analysis for optical encryption.

However, when dealing with a process whose state  $v$  is continuous, multiple issues complicate the verification of the Markov hypothesis by means of the CK equation and the Markov condition. Firstly, typically a limited number of samples  $n$  give rise to statistical errors when estimating transition PDFs  $p(v_i|v_{i+j})$  and  $p(v_i|v_{i+j}, v_{i+2j})$ . This aspect becomes particularly problematic in the Markov condition, where three-dimensional PDFs arise. Secondly, an accurate and computationally efficient method is needed to numerically evaluate the convolution in CK equation (2) at different points  $(v_3, v_1)$ . These issues become critical when verifying the CK equation and the Markov condition for small lags  $\Delta t$ , where the process states  $v_1$  and  $v_2$  or  $v_3$  are highly correlated and accordingly the transition PDFs are very peaked along the diagonals, e.g.,  $v_1 = v_3$  in  $v_3$ - $v_1$  space. To resolve these transition PDFs with histograms, very fine bins are required. In connection with a Cartesian equispaced bin arrangement, the integral evaluation in Eq. (2) translates then (as in the previous paragraph) to a matrix-matrix product that scales with the number of grid cells per  $v_i$  direction to the third power. Since the histogram bins have equal side lengths, reducing their size while maintaining the same sample size  $n$  will lead to increasing statistical errors. In summary, the typical method to inspect the Markov hypothesis based on the CK equation in connection with histograms is unable to adapt to transition PDFs with a diagonal structure, suffers from statistical errors, and represent only a weak test of Markovianity.

In this work, we propose a methodology that is based on a two-stage Markov test, where first the CK equation is verified and if successful the Markov condition is inspected. To reduce statistical and numerical errors in both stages, kernel density estimation and Gauss quadrature integration are applied. This methodology can account for the strong anisotropy in the transition PDFs and enables a locally adapted high-order integral evaluation. We demonstrate the capabilities of our approach for applications in both turbulence and subsurface dispersion. The latter area is of particular interest since predictive Markov models have been developed recently while the applicability of the Markov hypothesis is still being discussed [10,11].

## II. TESTING MARKOVIANITY

The CK equation (2) and the Markov condition (3) focus only on the immediate next state in the past, i.e.,  $v_3$ . The Markov hypothesis (1), however, accounts for dependencies on all past states. To test the independence of the transition PDF  $p(v_1|v_2, v_3, v_4, \dots)$  for states  $v_i$  with  $i > 3$ , we can write, for example,

$$\begin{aligned} p(v_1|v_4) &= \iint_{\Omega} p(v_1, v_2, v_3|v_4) dv_2 dv_3 \\ &= \iint_{\Omega} p(v_1, v_2|v_3, v_4) p(v_3|v_4) dv_2 dv_3 \\ &= \iint_{\Omega} p(v_1|v_2, v_3, v_4) p(v_2|v_3, v_4) p(v_3|v_4) dv_2 dv_3 \end{aligned}$$

or in analogy to Eq. (2)

$$p(v_1|v_4) = \iint_{\Omega} p(v_1|v_2) p(v_2|v_3) p(v_3|v_4) dv_2 dv_3 \quad (4)$$

to verify that  $p(v_1|v_2, v_3, v_4) = p(v_1|v_2)$  and  $p(v_2|v_3, v_4) = p(v_2|v_3)$ , where the latter corresponds for a stationary process to Eq. (3). In expression (4), repeated integration has to be conducted, which is a disadvantage in view of accumulating integration errors. Instead, we inspect

$$\begin{aligned} p(v_1|v_5) &= \iiint_{\Omega} p(v_1|v_2, v_3, v_4, v_5) p(v_2|v_3, v_4, v_5) \\ &\quad \times p(v_3|v_4, v_5) p(v_4|v_5) dv_2 dv_3 dv_4, \end{aligned}$$

which can be written for a stationary process and with  $p(v_1|v_2, v_3, v_4, v_5) = p(v_1|v_2)$  as

$$\begin{aligned} p(v_1|v_5) &= \int_{\Omega} \left[ \int_{\Omega} \overbrace{p(v_1|v_2) p(v_2|v_3)}^{=p(v_1|v_3)} dv_2 \right. \\ &\quad \left. \times \int_{\Omega} \underbrace{p(v_3|v_4) p(v_4|v_5)}_{=p(v_3|v_5)} dv_4 \right] dv_3. \quad (5) \end{aligned}$$

This is similar to the CK equation (2), but with a lag  $2\Delta t$  instead of  $\Delta t$ . Accordingly, to inspect the validity of the Markov hypothesis for a certain lag  $\Delta t$  and an increasing number of past states  $v_i$  with  $i > 2$ , the CK equation can be tested for a sequence of lags with increasing sizes, i.e.,  $\Delta t, 2\Delta t, 4\Delta t, \dots$ . Similarly, to keep the dimensionality of the probability space low, the Markov condition can be verified sequentially like  $p(v_1|v_2, v_3) = p(v_1|v_2)$ ,  $p(v_1|v_2, v_4) = p(v_1|v_2)$ ,  $\dots$ .

However, with the autocorrelation function of many physical processes being a decaying function for increasing lags, it is likely that if  $v_1$  is independent of  $v_3$  [Eqs. (2) and (3)], it will be independent of  $v_4, v_5, \dots$  as well [Eq. (5)]. Nevertheless, for a stationary process  $v(t)$ , we can replace  $v_3$  by  $v_i$  with  $i > 2$  in Eq. (2). In the limit of  $i \rightarrow \infty$ , the transition PDFs  $p(v_1|v_i)$  and  $p(v_2|v_i)$  become equal to the stationary PDF  $p(v)$  of process  $v(t)$  and Eq. (2) reduces to the eigenproblem

$$p(v_1) = \int_{\Omega} p(v_1|v_2) p(v_2) dv_2, \quad (6)$$

which could be used to inspect the independence for lags going to infinity.

## III. ESTIMATING TRANSITION PDFS

For the estimation of transition PDFs, e.g.,

$$p(v_i|v_{i+j}) \equiv \frac{p(v_i, v_{i+j})}{p(v_{i+j})} \quad (7)$$

with  $j$  being, for example, equal to 1 or 2 as in Eqs. (2) and (5), we propose the application of kernel density estimators (KDEs)

$$p(v) \approx \frac{1}{nh} \sum_{i=1}^n K\left(\frac{v - V^{(i)}}{h}\right)$$

[21–23]. Here  $n$  is the number of samples,  $h$  the kernel bandwidth or smoothing parameter,  $K$  the kernel, and  $V^{(i)}$

the sample with index  $i$ . Scott [23, Sec. 6.2.3] points out that the particular choice of the kernel shape is secondary, but the selection of the kernel bandwidth is of primary importance. Here we work with Gaussian kernels and in the univariate case, that is, for  $p(v_{i+j})$ , choose the smoothing parameter based on Silverman's rule [22, Eq. (3.28)], i.e.,  $h \equiv 1.06\alpha\sigma n^{-1/5}$ . Here  $\alpha = 1$  unless stated otherwise and  $\sigma$  is the sample standard deviation. In cases of higher dimensionality, e.g., for  $p(v_i, v_{i+j})$ , it is of particular importance, as will become clear later, to choose a kernel that is aligned with the joint PDF to be estimated. The corresponding Gaussian KDE is given by

$$p(\mathbf{v}) = \frac{\sum_{i=1}^n k[h^{-2}(\mathbf{v} - \mathbf{V}^{(i)})^T \mathbf{C}^{-1}(\mathbf{v} - \mathbf{V}^{(i)})]}{\sqrt{\det(\mathbf{C})n^d}}$$

with  $k(y) \equiv \frac{\exp(-y/2)}{(2\pi)^{d/2}}$ ,

$$h \equiv \alpha \left( \frac{4}{d+2} \right)^{1/(d+4)} n^{-1/(d+4)}, \quad (8)$$

dimensionality  $d$ , and sample covariance matrix  $\mathbf{C}$  [22, Eqs. (4.7) and (4.14)]. Equivalent methods were outlined by Wand and Jones [24] and Haerdle *et al.* [21, Eq. (3.71)]. For example, in the CK equation,  $d = 2$ ,  $\mathbf{v} \equiv (v_1, v_j)^T$  with  $j = 2$  or  $3$ . By assuming ergodicity in  $t$ , the sample  $\mathbf{V}^{(i)}$  represents all possible pairs  $(V_i, V_{i+j})^T$  that can be extracted from the process realization  $V_i \equiv V(t_i)$  with lag  $j\Delta t$ . When estimating the PDFs involved on the right-hand side of Eq. (7) with KDEs, it is important to make sure that the sample size  $n$  is sufficiently high such that the KDEs converge to unique densities independent of  $n$ . Corresponding convergence checks were conducted in the present work.

To evaluate the right-hand side of CK equations (2) and (5), for example, at point  $(v_1, v_3)^T$ , Gauss-Jacobi quadrature is applied on a confined interval bounded by  $(v_1 + v_3)/2 \pm L_v$ . In the present work,  $L_v$  was determined based on the standard deviation of  $v_2$  conditional on  $v_3$ . In order to check if the integration region  $2L_v$  is sufficiently large, it can be verified whether the resulting PDF  $p(v_1|v_3)$  integrates over  $v_1$  to one. Moreover, convergence checks involving the number of quadrature points and  $L_v$  should be performed. All these safeguards were implemented in the present work.

A first set of application examples, where the outlined test method was deployed in the context of three-dimensional turbulence is contained in the Appendix.

#### IV. ARE TRACER PARTICLES IN THE SUBSURFACE MARKOVIAN?

In this part, we use the methodology outlined in the previous sections to inspect the Markovianity of the Lagrangian velocity process of tracer particles in heterogeneous subsurface flows. In the context of two-dimensional isotropic multivariate Gaussian log-conductivity (or log-transmissivity) fields, we inspect the temporal log-velocity-magnitude process  $v(t)$  defined below. This process is the main driver of longitudinal macrodispersion in a spatially constant regional flow [13]. We demonstrate that depending on the size of the time lag  $\Delta t$ , Markovianity of  $v(t)$  can be found for both exponentially decaying and Gaussian correlation structures at mild and high heterogeneity levels. It is pointed out that we chose a two-dimensional setup because of its high relevance in hydrological applications. For advection-dominated transport results, that is, predictions of plume shape and spreading, we refer the interested reader to Refs. [10,11,13], where Markovian velocity process models are presented and their transport predictions are evaluated in space-stationary and non-space-stationary cases. The limiting dispersion behavior in space-stationary setups was investigated in Refs. [11,13], and Fickian behavior was found in most cases.

The random log-conductivity fields  $Y(\mathbf{x})$  with spatial position  $\mathbf{x}$  used in this work were generated with a spectral random field generator [25]. The resulting isotropic fields (see Fig. 1) are characterized in terms of the correlation structure with log-conductivity variance  $\sigma_Y^2$  and correlation length  $l_Y$ . Since we report normalized flow statistics in the absence of pore-scale dispersion, the log-conductivity mean is irrelevant. The hydraulic head  $h(\mathbf{x})$  was calculated based on the continuity equation  $\nabla \cdot (K \nabla h) = 0$  with  $K(\mathbf{x}) = \exp[Y(\mathbf{x})]$  and a driving mean head gradient in  $x_1$  direction [13]. Note that in general the hydraulic conductivity  $K$  is a tensorial quantity that is assumed here to be isotropic and thus reduces to a scalar [26, Sec. 5]. Eventually, the position of tracer particles was tracked with

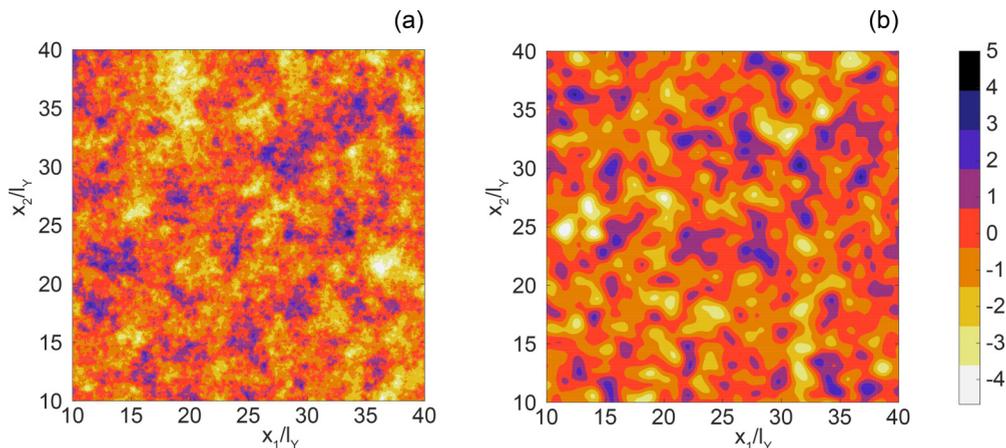


FIG. 1. Sections of log-conductivity fields  $Y(\mathbf{x})$  resulting from (a) the exponentially decaying and (b) the Gaussian correlation structure. In both cases  $\sigma_Y^2 = 1$  and the spatial directions are normalized with the correlation length  $l_Y$ .

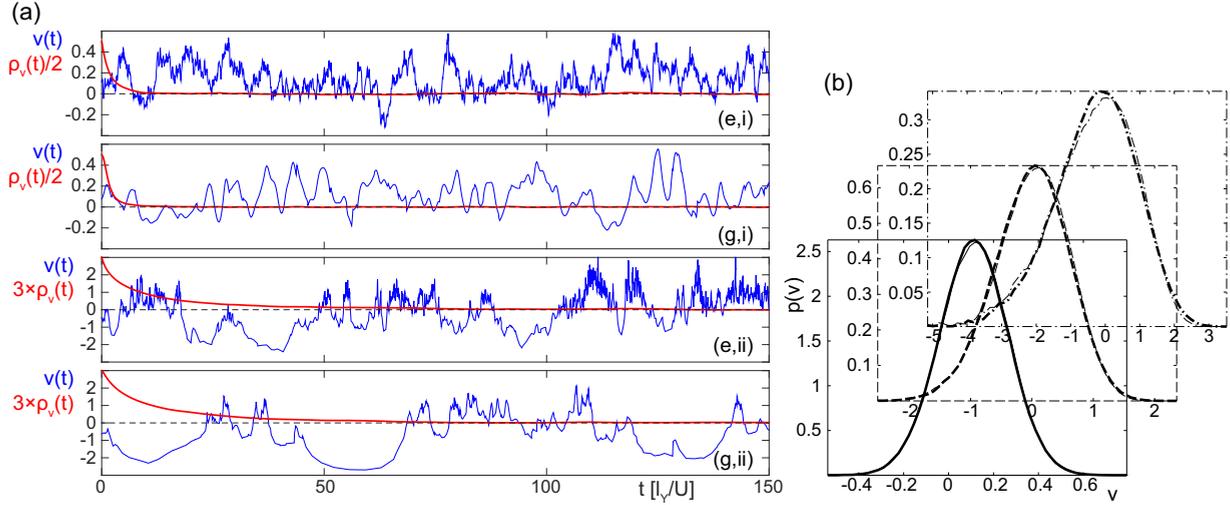


FIG. 2. (a) Log-velocity-magnitude trajectories  $v(t)$  (thin blue) and autocorrelation functions  $\rho_v(t)$  (thick red) for the exponentially decaying (e) and Gaussian (g) correlation structures and heterogeneity levels  $\sigma_Y^2 = 1/16$  (i) and 4 (ii). (b) Histograms of the normalized log-velocity-magnitude for the exponentially decaying (thick) and Gaussian (thin) correlation structures and three different  $\sigma_Y^2 = 1/16$  (solid), 1 (dashed), and 4 (dash-dot).

$\mathbf{X}(t) = \int_0^t \mathbf{u}[\mathbf{X}(t'), t'] dt'$ , where the flow field  $\mathbf{u}$  is resulting from Darcy's law, i.e.,  $\mathbf{u} = -(K/n)\nabla h$  with constant porosity  $n$ . Accordingly, in the absence of Brinkman and Forchheimer corrections in Darcy's law, we consider laminar low-velocity flow in porous media whose characterization is based on representative elementary volumes [27, Sec. 2]. Moreover, we focus on the high Péclet number limit, where the Péclet number is defined as  $Pe \equiv Ul_Y/D$  based on the mean flow velocity  $U$ , the correlation length  $l_Y$ , and the pore-scale dispersion coefficient  $D$ . However, at small or moderate Péclet numbers, we expect that the inclusion of pore-scale dispersion in the form of a random walk component in  $\mathbf{X}(t)$  [28] will rather enhance Markovianity at short time lags due to an increased temporal decorrelation of the Lagrangian velocity  $\mathbf{u}[\mathbf{X}(t), t]$ . In the following, we focus on the dynamics of the Lagrangian velocity magnitude process  $u(t) \equiv |\mathbf{u}(t)|$  or more precisely  $v(t) \equiv \ln[u(t)/U]$ , where  $U$  is the mean downstream velocity in  $x_1$  direction.

Unlike  $u(t)$ , whose stationary PDF is very skewed in heterogeneous cases with high  $\sigma_Y^2$  [29],  $v(t)$  is approximately Gaussian as illustrated in Figs. 2(b) and 3 with log scaling, which will facilitate the subsequent formulation of a Markov model. Here histograms of  $v(t)$  for different heterogeneity levels  $\sigma_Y^2$  are plotted for both correlation structures. Interestingly, both correlation structures lead to almost identical log-velocity-magnitude histograms despite the quite different log-conductivity field structures (see Fig. 1) and flow dynamics shown in Fig. 2(b).

In a first step, we test the validity of the Markov hypothesis for the mildest heterogeneity level considered, i.e.,  $\sigma_Y^2 = 1/16$ . For closer inspection of the validity of CK equation (2), most of the data in the following figures are plotted along the diagonal  $v_1 = v_3$  in  $v_3$ - $v_1$  space and a logarithmic scaling is used in certain transition PDF contour plots. For the exponentially decaying correlation structure, it is shown in Fig. 4 that the Markov hypothesis is not supported by the Lagrangian particle

statistics for a lag  $\Delta t = 0.11l_Y/U$ , i.e., the left-hand side of CK equation (2) represented by the filled contours in Fig. 4 does not match the right-hand side (dashed contour lines). However, for lags  $\Delta t = 0.22l_Y/U$  and  $0.44l_Y/U$  the contours from both equation sides coincide, and thus the CK equation is satisfied to a very good approximation.

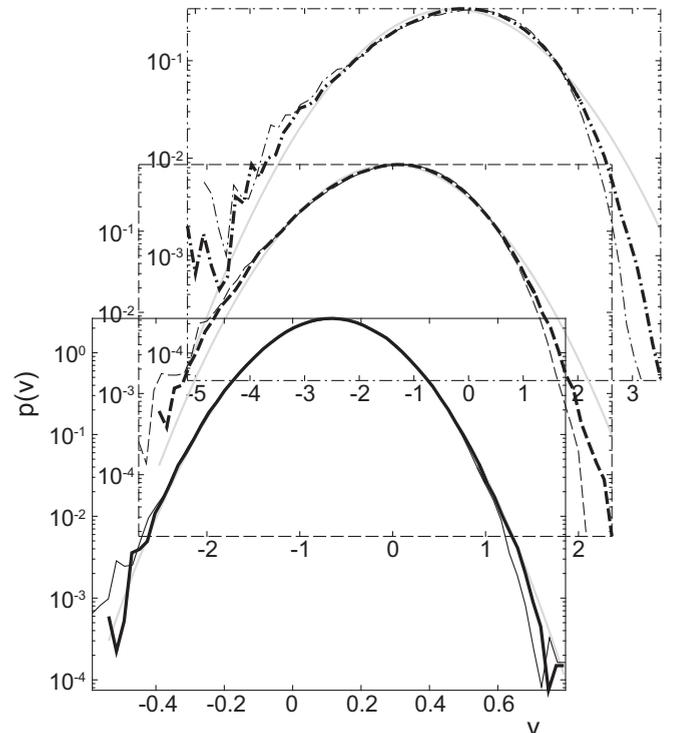


FIG. 3. See Fig. 2(b). For comparison, Gaussian probability densities (gray solid) with means and variances equal to the data sets from the exponentially decaying correlation structure are plotted as well.

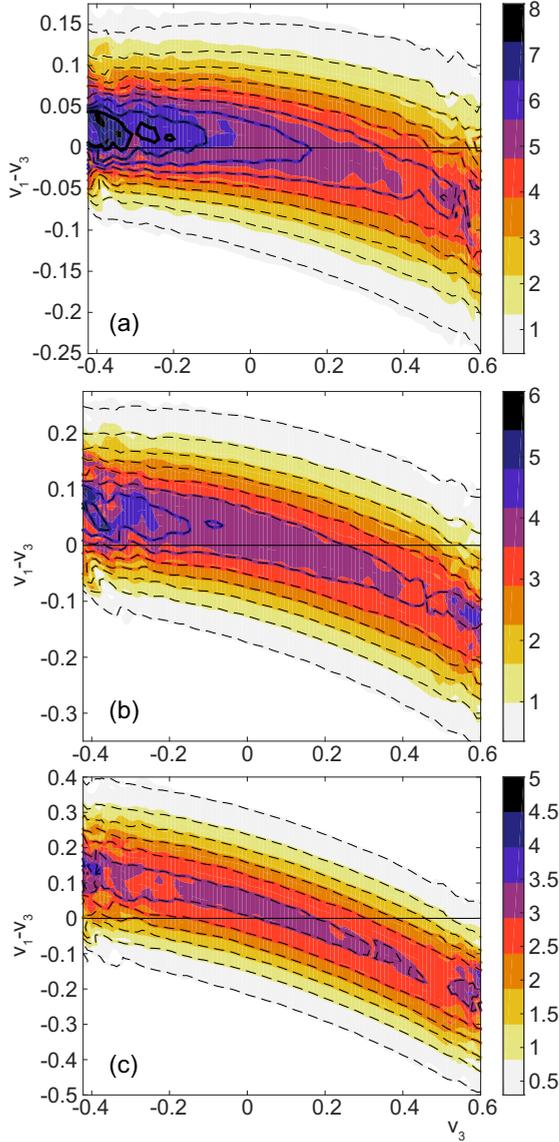


FIG. 4. Transition PDFs  $p(v_1|v_3)$  resulting from (filled contours) the KDE, i.e., the left-hand side of CK equation (2), and (dashed contour lines) the right-hand side of CK equation (2). Results for the exponentially decaying correlation structure with  $\sigma_Y^2 = 1/16$  and three different time lags (a) to (c) corresponding to  $\Delta t = 0.11l_Y/U$ ,  $0.22l_Y/U$ , and  $0.44l_Y/U$ , respectively, are depicted.

Subsequent tests of the Markov condition (3) confirm the validity of the Markov hypothesis for lag  $\Delta t = 0.44l_Y/U$ . The corresponding data are provided in Fig. 5, where the cubical domain containing the grid points used for the estimation of  $p(v_1|v_2, v_3)$  is depicted. The contour surfaces of equal probability in  $p(v_1|v_2, v_3)$  shown in panel (b) are to a good approximation parallel to the  $v_3$  direction, which illustrates that the Markov condition (3) is satisfied. Given the fact that the vertical grid lines in the domain are parallel to the  $v_3$  direction, further evidence is provided in panels (c) and (d), where the distribution of  $p(v_1|v_2, v_3)$  on the top and bottom planes of the domain, respectively, is shown. Both PDF distributions are approximately equal, which provides

further evidence that the dependence of  $p(v_1|v_2, v_3)$  on  $v_3$  is insignificant.

For the Gaussian correlation structure, as can be seen from Fig. 6 and based on similar results as depicted in Fig. 5, the Markov hypothesis becomes applicable for roughly 10 times larger lags at  $\Delta t = 2l_Y/U$ . This is not surprising given the smooth log-conductivity fields and accordingly  $v(t)$  realizations that are resulting from the Gaussian correlation structure [see Figs. 1 and 2(a)]. Since  $p(v_1|v_3)$  at  $\Delta t = 4l_Y/U$  is essentially a Gaussian PDF with constant standard deviation and conditional mean  $\langle v_1|v_3 \rangle$  that is linear in  $v_3$ , the corresponding  $v(t)$  process can be accurately modeled by a simple linear diffusion or Langevin process. For the exponentially decaying correlation structure, however, a nonlinear process or model is necessary (see Fig. 4).

When increasing the level of heterogeneity, the transition PDFs become more complex. This is shown in Fig. 7, where results for two lags  $\Delta t$  and the exponentially decaying correlation structure with  $\sigma_Y^2 = 4$  are depicted. Especially for the larger lag, two regimes can be identified. For high velocities or  $v_3 > -1$ , the process is similar to the ones inspected for  $\sigma_Y^2 = 1/16$ , i.e., the standard deviation of  $v_1$  conditional on  $v_3$  is approximately constant, and the mean has an approximate linear dependence on  $v_3$ . For small velocities with  $v_3 < -1$ , the conditional standard deviation becomes small, leading to a more correlated behavior [compare Fig. 2(e,ii)], which is in agreement with earlier observations [10,11]. While the Markov hypothesis is supported by the data for high velocities (compare Fig. 7), deviations between the sides of the CK equation appear at small velocities (inset in top panel of Fig. 7). This is the case as well for the Markov condition verified in Fig. 8, where a slight dependence on  $v_3$  becomes apparent at small  $v_1$ , whereas the dependence is negligible for large  $v_1$  values.

Very similar observations can be made for the Gaussian correlation structure. Like in the mildly heterogeneous case, the Markov hypothesis becomes a good approximation for larger time lags, i.e.,  $\Delta t \geq 1.6l_Y/U$ , compared with the exponentially decaying structure, where good agreement in the CK equation and the Markov condition was found already for  $\Delta t \geq 0.64l_Y/U$ . Low- and high-velocity regimes are present, as is shown in Fig. 9, where the transition PDF  $p(v_1|v_3)$  and the joint PDF  $p(v_1, v_3)$  are plotted, respectively [see Fig. 2(g,ii) as well].

Independent of the correlation structure, particularly the transition PDFs for  $\sigma_Y^2 = 4$  involve a degree of complexity that greatly complicates a verification of the CK equation and the Markov condition with conventional methods based on equidistant histograms and matrix products. On one hand, the low-velocity regions necessitate the use of fine grid cells and accordingly many samples to keep statistical errors low. On the other hand, coarse structures in the high-velocity regions require sufficient spatial extension in probability space and therefore induce a high number of grid cells. With the methodology outlined in this work, these difficulties can be overcome. Since KDEs provide semianalytical PDF estimates, the transition PDFs can be quantified at relevant grid points only. For example, to capture the narrow low-velocity regions [see, for example, Fig. 7(a)], a small spacing of points

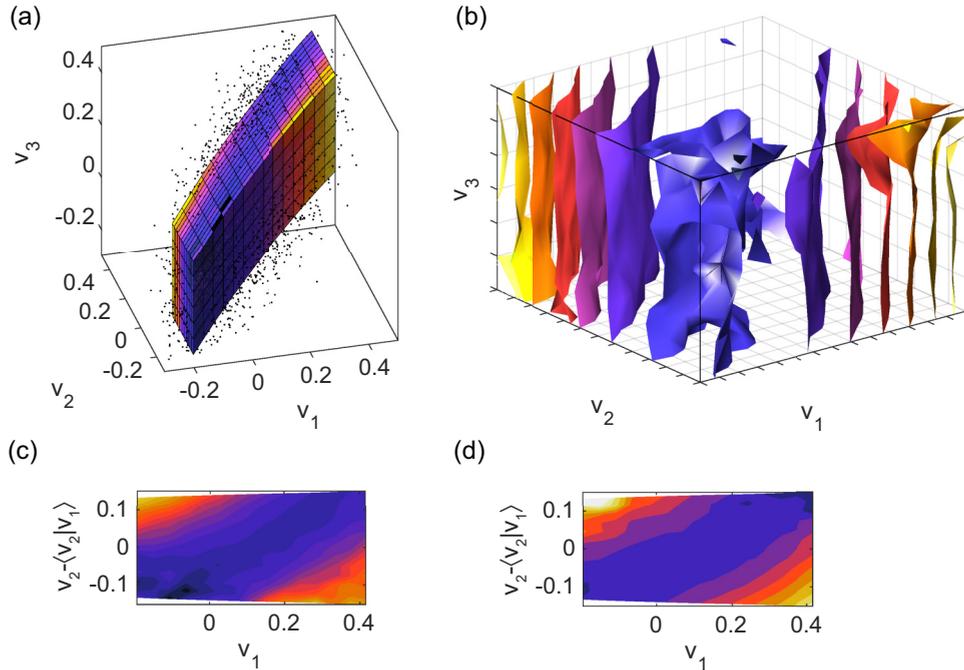


FIG. 5. Transition PDF  $p(v_1|v_2, v_3)$  of the exponentially decaying correlation structure with  $\sigma_Y^2 = 1/16$  and a time lag  $\Delta t = 0.44l_Y/U$ . The color shading corresponds to  $\log_{10}$  values of  $p(v_1|v_2, v_3)$ . (a) The dots are 5000 exemplary samples, and the cube represents the domain with the grid points where the PDF was estimated. The grid points are arranged on planes parallel to the  $v_2$ - $v_3$  plane and the vertical grid lines are parallel to the  $v_3$  direction. Contour surfaces of equal probability inside the cubical domain are depicted in (b). The probability density distributions on the top and bottom faces of the domain are given in panels (c) and (d), respectively.

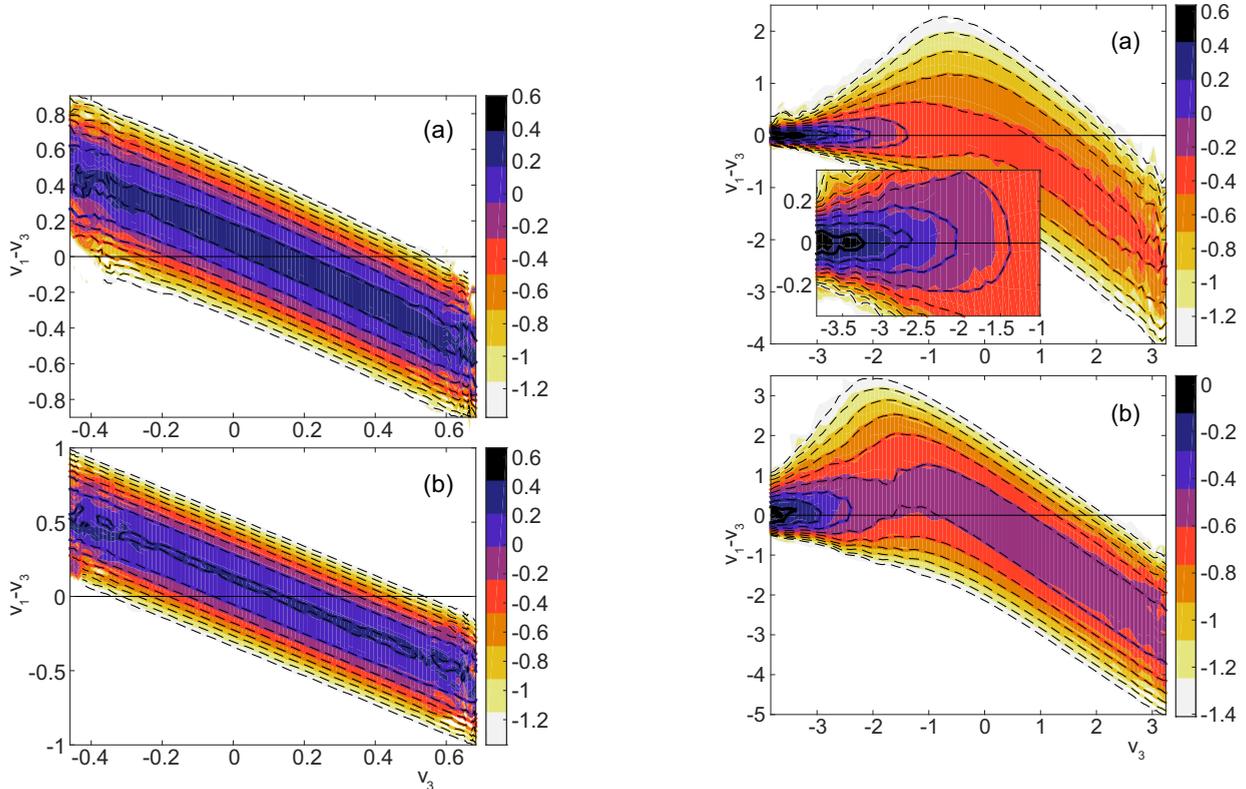


FIG. 6. See Fig. 4. The contour levels represent  $\log_{10}$  values of  $p(v_1|v_3)$ . Results for the Gaussian correlation structure with  $\sigma_Y^2 = 1/16$  and two different time lags (a) and (b) corresponding to  $\Delta t = 2l_Y/U$  and  $4l_Y/U$ , respectively, are depicted.

FIG. 7. See Fig. 4. The contour levels represent  $\log_{10}$  values of  $p(v_1|v_3)$ . Results for the exponentially decaying correlation structure with  $\sigma_Y^2 = 4$  and two different time lags (a) and (b) corresponding to  $\Delta t = 0.64l_Y/U$  and  $2.56l_Y/U$ , respectively, are depicted. The inset shows a detail view of the PDF depicted in the top panel.

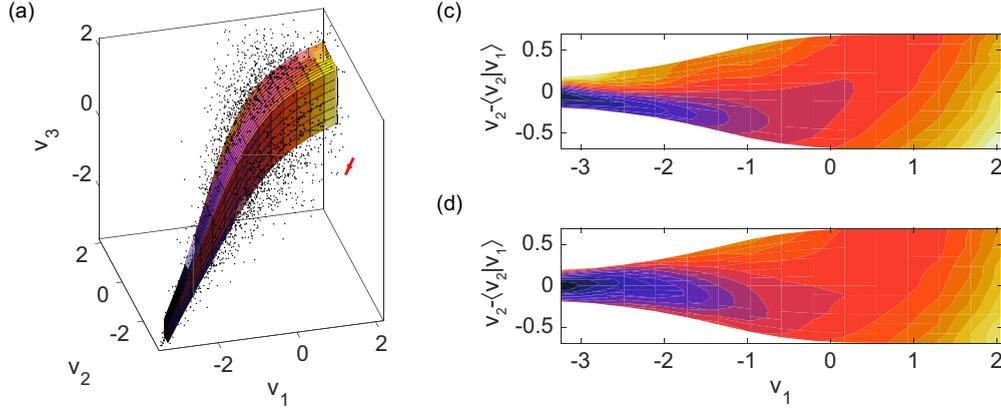


FIG. 8. Transition PDF  $p(v_1|v_2, v_3)$  of the exponentially decaying correlation structure with  $\sigma_Y^2 = 4$  and a time lag  $\Delta t = 0.64l_Y/U$ . See Fig. 5.

is chosen, whereas in the high-velocity regions points are coarsely distributed. For a graphical illustration see Figs. 8(a) and 10. By using a suitable kernel shape, the KDE can account for the narrow diagonal PDF structure and the PDF can be accurately estimated with a relatively small sample size. However, since we use the same kernel shape and size in the entire probability space, its size has to be selected such that the narrow PDF section in the low-velocity region is resolved. For this reason  $\alpha$  in Eq. (8) was reduced to 0.5 for the case with the exponentially decaying correlation structure with  $\sigma_Y^2 = 4$  and  $\Delta t = 0.64l_Y/U$ .

V. CONCLUSIONS

In this work, we have systematically tested the Markov hypothesis for applications in the areas of turbulence modeling (see the Appendix) and subsurface dispersion. Our results confirm the suitability of Markov models for the description of tracer particle velocities and the fluid velocities seen by inertial particles. In the context of subsurface dispersion, unlike previous contributions, we show that Markov models become applicable above a lowest limiting time lag. For the smooth Gaussian correlation structure, this lag is significantly larger compared to the discontinuous exponentially decaying

structure. For low heterogeneity levels, the transition PDFs have relatively simple shapes for both structures. However, at high heterogeneity levels, these PDFs are more complex and low- and high-velocity regimes coexist. At high velocities, the transition PDF is wide and the Markov hypothesis is accurate for small lags. However, at low velocities, the transition PDF is narrow and the dynamics become Markovian at elevated lags only. It is remarkable that despite the different log-conductivity correlation structures, the transition PDFs are very similar for both correlation structures at high heterogeneity levels. This similarity opens the prospect of a unified Markovian macrodispersion model for both structures and blends thereof like the Matern covariance. This is because each transition PDF essentially represent a blueprint of a model process.

For the estimation of transition PDFs, that are crucial when verifying the Markov hypothesis, a numerical framework that is based on suitably selected kernel density estimators and a Gauss quadrature formula was outlined and successfully applied.<sup>1</sup> Our framework is particularly useful when testing Markovianity at small lags, where conventional histogram-

<sup>1</sup>The according computer programs are available by contacting the corresponding author.

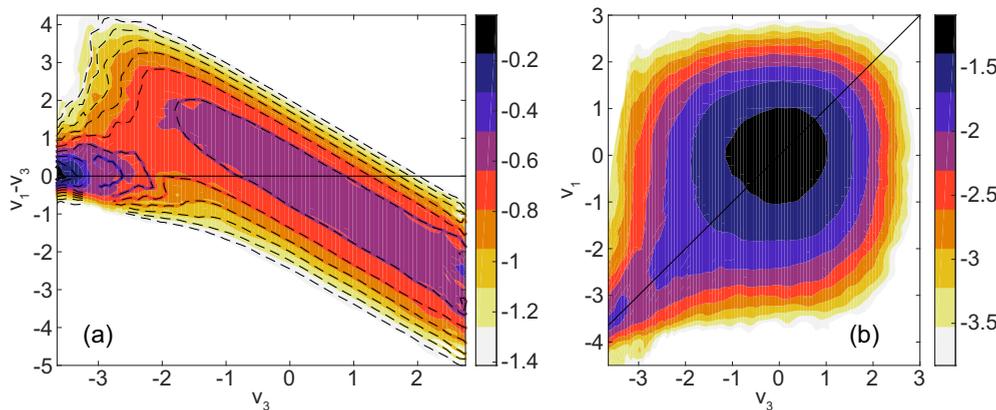


FIG. 9. Results for the Gaussian correlation structure with  $\sigma_Y^2 = 4$  and a lag  $\Delta t = 6.4l_Y/U$  are depicted. (a) See Fig. 4. The contour levels represent  $\log_{10}$  values of  $p(v_1|v_3)$ . (b) The contour levels represent  $\log_{10}$  values of the joint PDF  $p(v_1, v_3)$ .

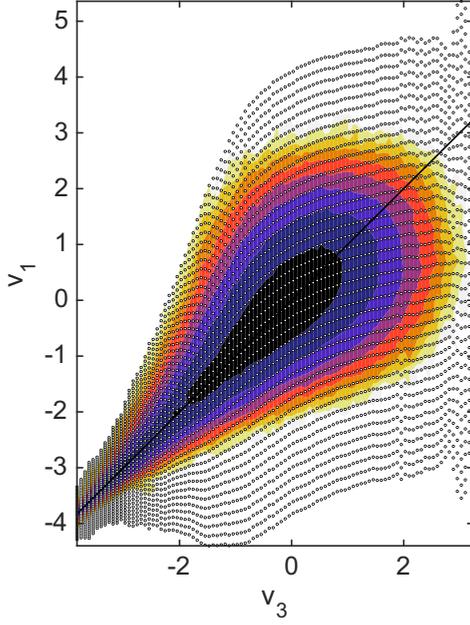


FIG. 10. Joint PDF  $p(v_1, v_3)$  resulting from kernel density estimation. Contour levels represent  $\log_{10}$  values of the joint PDF. The dots represent points  $(v_1, v_3)^T$ , where the right-hand side of the Chapman-Kolmogorov equation was evaluated with Gauss quadrature leading to the dashed contours in Fig. 7(a). Results for the exponentially decaying correlation structure with  $\sigma_v^2 = 4$  and time lag  $\Delta t = 0.64l_Y/U$  are depicted.

based methods become inaccurate. Moreover, the presented methodology includes testing of the Markov condition. This is unlike most existing contributions that focus on the weak Chapman-Kolmogorov equation, which may falsely identify processes as Markov.

#### ACKNOWLEDGMENTS

We are very grateful to Philipp Weiss from ETH Zürich, who provided us with the two-phase turbulence data.

Moreover, we acknowledge funding by ETH Zürich and very helpful comments of the anonymous reviewers.

#### APPENDIX: APPLICATIONS IN TURBULENT FLOWS

To demonstrate the suitability of our methodology in the context of turbulent flows, we use data stemming from direct numerical simulation of forced homogeneous isotropic turbulence. This is a prototype flow representative of turbulence in many applications. The computational domain is a three-dimensional box that is periodic in each spatial direction and has a side length of  $2\pi$  (all values given in SI units). To arrive at a statistically stationary turbulent flow in time, the linear forcing scheme of Rosales and Meneveau [30] was applied. The turbulent flow field used in this study has a Taylor-scale Reynolds number of  $Re_\lambda = 48.5$ , a Kolmogorov length scale  $\eta = 0.0273$  and time scale  $\tau_\eta = 0.126$ , a fluid density  $\rho = 1$ , and a viscosity  $\nu = 4.491 \times 10^{-3}$ . The inertial particles were represented as point particles with diameter  $d_p = 3.19 \times 10^{-3}$  and densities  $\rho_p = 1 \times 10^3$  and  $80 \times 10^3$  corresponding to Stokes numbers  $St \equiv \tau_p/\tau_\eta = 1$  and  $80$ , respectively. Here

$$\tau_p \equiv \frac{2}{9} \frac{\rho_p}{\rho} \frac{1}{v} \left( \frac{d_p}{2} \right)^2 \quad (\text{A1})$$

is the Stokes relaxation time of the particle. We point out that the point particle approximation is applicable in the limit of  $d_p \ll \eta$ . For the numerical solution of the Navier-Stokes equation, a spectral element solver [31] was applied and supplemented by a one-way-coupling particle scheme. One-way coupling is accurate if inertial particles are suspended at low mass loadings. The Stokes drag force was determined based on the Schiller-Naumann drag correction as a function of the particle Reynolds number. The resulting particle evolution equations are given by

$$\frac{d\mathbf{x}_p}{dt} = \mathbf{u}_p \text{ and } \frac{d\mathbf{u}_p}{dt} = -\frac{c_p}{\tau_p} (\mathbf{u}_p - \mathbf{u}_f), \quad (\text{A2})$$

where  $\mathbf{x}_p(t)$  and  $\mathbf{u}_p(t)$  are the particle position and velocity, respectively,  $c_p \equiv 1 + 0.15Re_p^{0.687}$  is the Schiller-Naumann

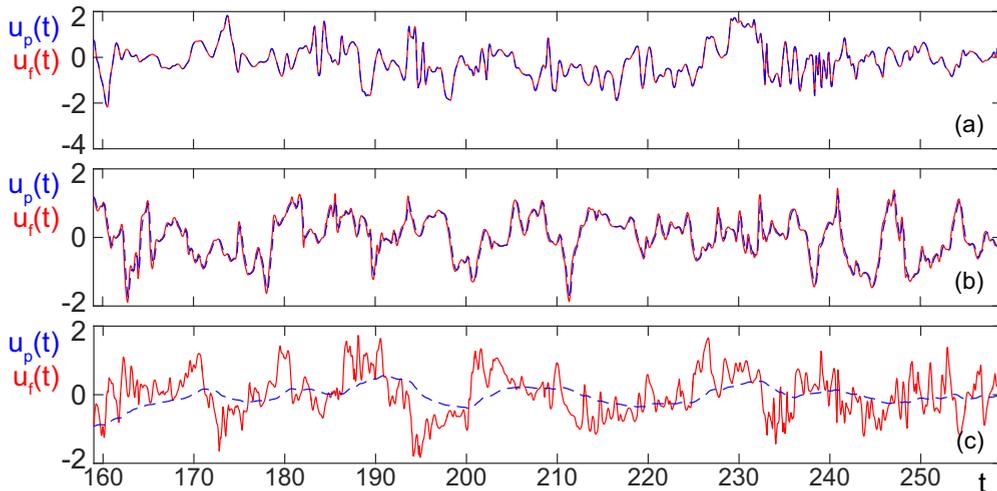


FIG. 11. Time series of the particle velocity (blue dashed line) and the local fluid velocity seen by the particle (red solid line). Depicted are three types of particles, (a) fluid particle and inertial particles with (b)  $St = 1$  and (c)  $St = 80$ .

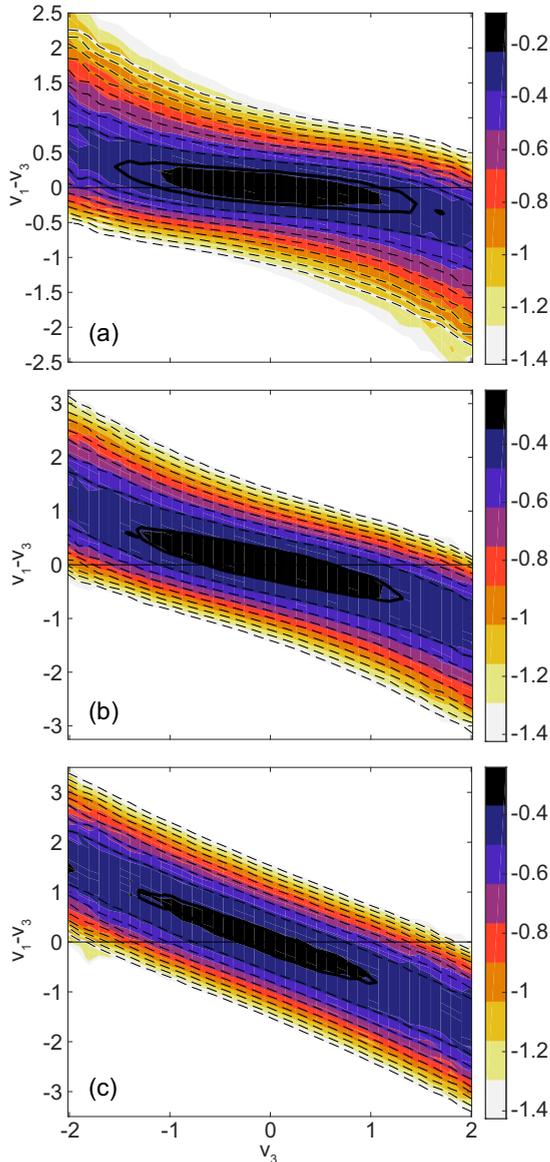


FIG. 12. Transition PDFs  $p(v_1|v_3)$  resulting from (filled contours) the KDE, i.e., the left-hand side of the CK equation, and (dashed contour lines) the right-hand side of the CK equation. The contour levels represent  $\log_{10}$  values of  $p(v_1|v_3)$ . Results for fluid or tracer particles at three different time lags (a) to (c) corresponding to  $\Delta t = 0.24, 0.48$ , and  $0.96$ , respectively, are depicted.

drag correction applicable for  $\text{Re}_p \equiv |\mathbf{u}_p - \mathbf{u}_f|d_p/\nu$  up to 1000 [32], and  $\mathbf{u}_f(t) \equiv \mathbf{u}[\mathbf{x}_p(t), t]$  is the local fluid velocity seen by the particle. Exemplary time series of components of the particle velocity and the fluid velocity seen by the particle are depicted in Fig. 11 for  $\text{St} = 1$  and 80. It is visible that for the larger Stokes number, which corresponds to heavier particles, the particle velocity fluctuates less (blue dashed curve). However, as the heavy particle moves through the fluid, the velocity of the surrounding fluid changes more rapidly (red solid curve) compared to the lighter particle with  $\text{St} = 1$  and thus becomes subject to an increasingly random uncorrelated motion (RUM) [33,34]. In the case of tracer or

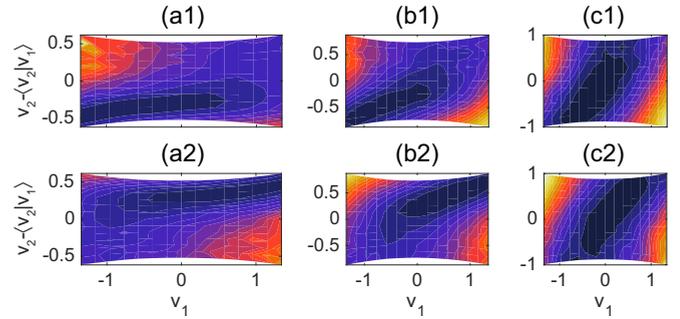


FIG. 13. Transition PDFs  $p(v_1|v_2, v_3)$  at two different values of  $v_3$ , (1) and (2), analogous to the ones depicted in Fig. 5, panels (c) and (d). Results for fluid or tracer particles at three different time lags (a) to (c) corresponding to  $\Delta t = 0.24, 0.48$ , and  $0.96$ , respectively, are depicted.

fluid particles, the fluid particle position was evolved based on the fluid velocity seen by the particle, i.e.,  $\mathbf{u}_p(t) = \mathbf{u}_f(t)$ . Corresponding time series are included in Fig. 11 as well.

In a first step, the validity of the Markov hypothesis is tested for fluid or tracer particles. Corresponding models are applied for example for the simulation of atmospheric dispersion [6] or turbulent reactive flows with PDF methods [1,7]. Here variable  $v$  represents one component of the particle or Lagrangian fluid velocity vector  $\mathbf{u}_p(t) = \mathbf{u}_f(t)$ . The resulting transition PDFs  $p(v_1|v_3)$  are depicted in Fig. 12 for different time lags  $\Delta t$ . For closer inspection, these PDFs are plotted along the diagonal  $v_1 = v_3$  in  $v_3$ - $v_1$  space and a logarithmic contour level distribution is applied. For lags  $\Delta t \geq 0.48$ , it is visible that a Markov model can be considered since the left-hand side of CK equation (2), represented by the filled contours in Fig. 12, matches with the right-hand side (dashed contour lines). Moreover, by going from small to large lags, we can observe that  $p(v_1|v_3)$  becomes, except for a linear dependence of the conditional mean  $\langle v_1|v_3 \rangle$ , essentially independent of  $v_3$ . Additionally, since the shape of  $p(v_1|v_3)$  approaches a Gaussian for increasing  $\Delta t$ , our results are consistent with the linear diffusion processes or Langevin models that are typically applied in the previously listed applications (e.g., Ref. [1, Eq. (28)]). In Fig. 13 we can observe that for a growing time lag  $\Delta t$ , the transition PDF  $p(v_1|v_2, v_3)$  becomes independent of  $v_3$  and eventually satisfies to a good approximation the Markov condition (3) at  $\Delta t = 0.96$ . This outcome is more conservative compared to the previous result from the CK equation, i.e.,  $\Delta t \geq 0.48$ .

The transition PDFs for the fluid velocity seen by the particle for inertial particles with  $\text{St} = 1$  are very similar compared to the fluid particle case shown in Fig. 12. For the PDFs of the heavy particle with  $\text{St} = 80$  given in Fig. 14, however, the Markov and diffusion process regimes are reached roughly at lags half the size compared with  $\text{St} = 1$ . This behavior is consistent with our discussion of Fig. 11. There, we observed that for  $\text{St} = 80$ ,  $\mathbf{u}_f(t)$  evolves more rapidly and thus loses memory more quickly than for  $\text{St} = 1$ . In summary, our results support the use of stochastic diffusion processes to model  $\mathbf{u}_f(t)$  over time periods  $\gg \tau_\eta$  [3,4].

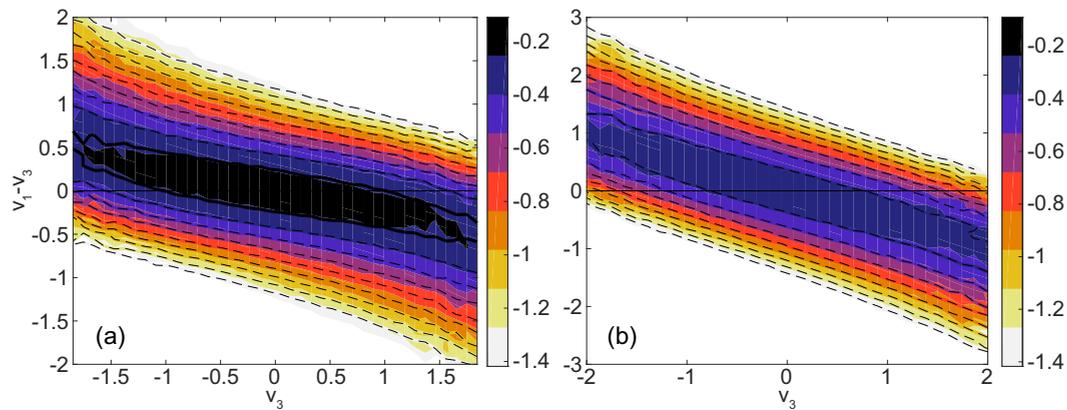


FIG. 14. See Fig. 12. Results for inertial particles with  $St = 80$  and two different time lags (a) and (b) corresponding to  $\Delta t = 0.24$  and  $0.48$ , respectively, are depicted.

- 
- [1] S. B. Pope, *Phys. Fluids* **23**, 011301 (2011).
- [2] J.-P. Minier, *Prog. Energy Combust. Sci.* **50**, 1 (2015).
- [3] J. Pozorski and J. P. Minier, *Phys. Rev. E* **59**, 855 (1999).
- [4] D. W. Meyer, *J. Fluid Mechanics* **706**, 251 (2013).
- [5] J. Pozorski and S. V. Apte, *Int. J. Multiphase Flow* **35**, 118 (2009).
- [6] A. Jones, D. Thomson, M. Hort, and B. Devenish, The U.K. Met Office's next-generation atmospheric dispersion model, NAME III, in *Air Pollution Modeling and Its Application XVII*, edited by C. Borrego and A.-L. Norman (Springer, New York, 2007), Sec. 62, pp. 580–589.
- [7] S. B. Pope, *Prog. Energy Combust. Sci.* **11**, 119 (1985).
- [8] B. L. Sawford, *Phys. Fluids A* **3**, 1577 (1991).
- [9] R. Zamansky, I. Vinkovic, and M. Gorokhovski, *J. Fluid Mech.* **721**, 627 (2013).
- [10] T. Le Borgne, M. Dentz, and J. Carrera, *Phys. Rev. E* **78**, 026308 (2008).
- [11] D. W. Meyer and H. A. Tchelepi, *Water Resour. Res.* **46**, W11552 (2010).
- [12] Y. Rubin, *Applied Stochastic Hydrogeology* (Oxford University Press, Oxford, 2003), p. 391.
- [13] D. W. Meyer, H. A. Tchelepi, and P. Jenny, *Water Resour. Res.* **49**, 2359 (2013).
- [14] J. W.-B. Lin and J. D. Neelin, *Geophys. Res. Lett.* **30**, 1162 (2003).
- [15] Y. Xue and A. Leetmaa, *Geophys. Res. Lett.* **27**, 2701 (2000).
- [16] W. Feller, *Ann. Math. Stat.* **30**, 1252 (1959).
- [17] E. Parzen, *Stochastic Processes*, Classics in Applied Mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 1999), p. 324.
- [18] N. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam, 2007).
- [19] P. K. Kang, M. Dentz, T. Le Borgne, and R. Juanes, *Phys. Rev. Lett.* **107**, 180602 (2011).
- [20] R. Horstmeyer, R. Y. Chen, B. Judkewitz, and C. Yang, *Opt. Express* **20**, 26394 (2012).
- [21] W. Haerdle, A. Werwatz, M. Mueller, and S. Sperlich, *Nonparametric and Semiparametric Models*, Springer Series in Statistics (Springer, Berlin, 2004).
- [22] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability (Chapman and Hall/CRC, Boca Raton, FL, 1998), p. 174.
- [23] D. W. Scott, *Multivariate Density Estimation* (John Wiley and Sons, New York, 1992).
- [24] M. P. Wand and M. C. Jones, *J. Am. Stat. Assoc.* **88**, 520 (1993).
- [25] F. Mueller, P. Jenny, and D. W. Meyer, *J. Comput. Phys.* **250**, 685 (2013).
- [26] J. Bear, *Dynamics of Fluids in Porous Media* (American Elsevier, New York, 1972), p. 764.
- [27] G. Dagan, *Flow and Transport in Porous Formations* (Springer, Berlin, 1989), p. 465.
- [28] P. Salamon, D. Fernandez-Garcia, and J. J. Gomez-Hernandez, *J. Contam. Hydrol.* **87**, 277 (2006).
- [29] P. Salandin and V. Fiorotto, *Water Resour. Res.* **34**, 949 (1998).
- [30] C. Rosales and C. Meneveau, *Phys. Fluids* **17**, 095106 (2005).
- [31] P. F. Fischer, J. W. Lottes, and S. G. Kerkemeier, Nek5000 a spectral element code for CFD, <http://nek5000.mcs.anl.gov> (2015).
- [32] L. Schiller and A. Naumann, *Verein Deutscher Ingenieure* **77**, 318 (1933).
- [33] R. H. A. Ijzermans, E. Meneguz, and M. W. Reeks, *J. Fluid Mech.* **653**, 99 (2010).
- [34] K. Gustavsson, E. Meneguz, M. Reeks, and B. Mehlig, *New J. Phys.* **14**, 115017 (2012).