J Stat Phys (2014) 156:896–907
DOI 10.1007/s10955-014-1035-6

A Method for Identifying Diffusive Trajectories with Stochastic Models

D. O'Malley · V. V. Vesselinov · J. H. Cushman

Received: 12 March 2014 / Accepted: 29 May 2014 / Published online: 19 June 2014

Abstract Single particle tracking is a tool that is being increasingly used to study diffusive or dispersive processes in many branches of natural science. Often the ability to collect these trajectories experimentally or produce them numerically outpaces the ability to understand them theoretically. On the other hand many stochastic models have been developed and continue to be developed capable of capturing complex diffusive behavior such as heavy tails, long-range correlations, nonstationarity, and combinations of these things. We describe a computational method for connecting particle trajectory data with stochastic models of diffusion. Several tests are performed to demonstrate the efficacy of the method, and the method is applied to polymer diffusion, RNA diffusion in *E. coli*, and RAFOS dispersion in the Gulf of Mexico.

Keywords Anomalous diffusion · Model identification · Polymer · E. coli · Gulf of Mexico

1 Introduction

Lagrangian dynamics in mixing processes are frequently observed and simulated. Often particle trajectories are driven by unknown and/or innumerable physical processes and hence are often modeled by stochastic processes. Perhaps the most famous such example was Brown's observation of the highly irregular motion of pollen grains in water [1]. A mathematical characterization of this motion was subsequently provided [2], and the motion is now known as Brownian motion (despite earlier observations of the "confused, continual & violent" motion of crushed coal in alcohol [3]). Modern techniques make it possible not only to observe these motions, but to quantify these observations (*i.e.*, turn them into time series). To name a few

D. O'Malley (✉) · V. V. Vesselinov
Computational Earth Science, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: omalled@lanl.gov

J. H. Cushman
Departments of Earth, Atmospheric and Planetary Sciences and Mathematics, Purdue University, West Lafayette, IN, USA

examples, experimental "particle" tracking has been carried out in finance [4,5], microbiology [6–8], macrobiology [9,10], oceanography [11], and porous media [12,13]. Numerical simulation techniques are often used to represent particle trajectories in place of experiments [14–17].

In the example of Brown's pollen grains, theory was used to derive an appropriate stochastic model. However, for many problems there is an insufficient understanding of the underlying stochastic dynamics to theoretically identify a model. Nonetheless, it is frequently desirable to associate a stochastic model with the motion of the particles. For instance, some of the advantages of the stochastic models are that they may make it possible to apply new analytical tools [18] to the problem and that it is much less expensive to obtain samples from a stochastic process than it is to obtain them from an experiment or a simulation of the full physics. Here, we describe a methodology to identify an appropriate stochastic process for diffusive trajectories based on computational/statistical techniques that can be used when theory fails to accomplish this.

2 Methodology

To achieve the goal of identifying experimental or simulated trajectories with a mathematical description of a stochastic process, we combine maximum likelihood methods with analytical information criteria. One of the challenges in using this approach is to develop a rich set of stochastic processes that can model a diverse set of physical phenomena (see Sect. 2.2). Another is estimating the maximum-likelihood parameters for each model (see Sect. 2.3). Once the maximum-likelihood parameters are determined, an information criteria is computed (see Sect. 2.1). The information criteria can then be used to select an appropriate model or provide weights for each of the models. The weights can be used to combine several stochastic models to produce an ensemble representation of the trajectories.

General processes with a large number of parameters naturally have a higher likelihood to reproduce trajectory observations, but the analytical information criteria (*e.g.*, the Akaike information criteria) strike a balance between the number of process parameters and the goodness of fit. This guards against overfitting, and guides the selection process toward an appropriately simple model for the data.

2.1 Information Criteria

Information criteria provide a quantitative version of the parsimony principle or Occam's razor. That is, they attempt to find the balance between the goodness-of-fit in a model and the complexity of the model (mostly in terms of the number of parameters, but also possibly in terms of things like the Fisher information). The Akaike information criterion (AIC) for a model with n parameters is defined to be

$$AIC = 2n - 2\log(L) \tag{1}$$

where L gives the likelihood of the model given some data [19]. Models with lower AIC (fewer parameters/greater likelihood) are preferred. A "corrected" version of the AIC, denoted by AICc, is defined as

$$AICc = 2n - 2\log(L) + \frac{2n(n+1)}{m-n-1} \tag{2}$$

where m is the number of data points [20]. For a fixed number of model parameters, the AICc converges to the AIC as the number of data points increases to infinity.

From the equations, given two equally likely models, the one with fewer parameters would be preferred based on this criterion. Similarly, given two models with the same number of parameters, the one with a greater likelihood would be preferred.

Other information criteria such as the Bayesian information criterion (BIC) and Kashyap information criterion (KIC) are also available. The BIC and the KIC are more complicated than AIC and AICc because they require a prior distribution. The KIC additionally employs the Fisher information matrix. Different arguments can be made for and against the various information criteria (see, *e.g.*, [21] and [22] for two examples). The evaluation of the information criterion is essentially a post-processing step, and, if desired, multiple information criteria can be explored. We echo the sentiments in [22] that the choice to use some information criteria is more important than the choice of which criteria, and the AIC will be employed here. While the KIC is potentially the most robust of the information criteria mentioned here, it has the disadvantage that prior probabilities must be supplied. Meaningful prior probabilities are frequently unavailable. The AIC's relative simplicity makes it more practical, and it appears to be capable of reliably choosing the appropriate model in the cases considered and for a related problem [23]. The AICc and BIC depend on the same variables (maximum likelihood, number of parameters, and number of data points) as the AIC, so they can be used as drop-in replacements, if desired.

2.2 Stochastic Models

Let $X(t)$ denote a sample trajectory. Without loss of generality, we assume zero mean drift, *i.e.*, $\langle X(t)\rangle = 0$. The classical Lagrangian Brownian model of diffusion is defined via three properties [24]

- Every increment $X(t+s) - X(s)$ is normally distributed with mean zero and variance $\sigma^2 t$, *i.e.*, $X(t+s) - X(s) \sim N(0, \sigma^2 t)$.
- For every pair of disjoint time intervals (t_1, t_2) and (t_3, t_4), the increments $X(t_4) - X(t_3)$ and $X(t_2) - X(t_1)$ are independent random variables with the distribution given above.
- With probability 1, $X(0) = 0$ and $X(t)$ is continuous.

When one of the first two conditions does not hold for a diffusive process, the process is considered anomalous. That is, anomalous diffusion occurs when one or more of the following holds

1. The increments are not normally distributed.
2. The increment distribution is not stationary.
3. The increments over disjoint time intervals are not independent.

Anomalous diffusive processes are often characterized by power-law mean square displacements. The conditions for Brownian motion are stringent while the conditions for anomalous diffusion are broad. It is therefore not surprising that anomalous diffusion is often found, *e.g.*, [7,8,25–29].

Stochastic models of anomalous diffusion can often be described by perturbing the properties that define Brownian motion. Three models of anomalous diffusion can be defined that are similar to Brownian motion, except that the models have one of the properties of anomalous diffusion listed above. These models are α-stable Lévy motion, Brownian motion with a nonlinear clock, and fractional Brownian motion. Of course, there are many more models of anomalous diffusion—see [30] for a review of some properties of a number of commonly employed stochastic models of anomalous diffusion.

α-stable Lévy motion is defined similarly to Brownian motion, except the normally distributed increments are replaced by α-stable increments [31]. This has the effect of transforming the tail of the particle distribution from one with exponential decay in the Brownian case to one with power-law decay in the α-stable case. One of the implications of this change is that α-stable Lévy motion has infinite mean square displacement, rather than the mean square displacement of Brownian motion, which is linear in time. It will also have an infinite first moment if $0 < \alpha \leq 1$. The increments of α-stable Lévy motion are not normally distributed, but they are stationary and independent. That is, α-stable Lévy motion has the first property listed above, but not the second or third.

Fractional Brownian motion [32] can be also defined in a way similar to Brownian motion, except that the assumption of independent increments is not used [24]. Instead, fractional Brownian motion has a heavy-tailed correlation structure.

$$E[B_H(t)B_H(s)] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right) \tag{3}$$

where $B_H(t)$ is a fractional Brownian motion with Hurst exponent H. The mean square displacement of a fractional Brownian motion is a power-law in time, t^{2H}. Note that the increments of fractional Brownian motion are not independent, but they are normally distributed and stationary. That is, fractional Brownian motion has the third property listed above, but it does not have the first or the second. Fractional Brownian motion was popularized by Mandelbrot and Van Ness [32], but was studied previously by Kolmogorov [33]. A historical account of fractional Brownian motion is found in [34].

Time-rescaled Brownian motion [35], which we call Brownian motion with a nonlinear clock [36] can be defined similarly, except that the assumption of stationary increments is not used [37]. This process can be understand most easily via function composition with a Brownian motion,

$$X(t) = B(C(t)) \tag{4}$$

where $B(t)$ is a Brownian motion, and $C(t)$ is a deterministic subordinator called the clock. Brownian motion with a nonlinear clock does not have stationary increments, but it does have uncorrelated, normally distributed increments. That is, it has the first and third properties listed above, but not the second.

The features of these processes can be stacked, producing the following variations

1. Brownian motion
2. α-stable Lévy motion
3. Fractional Brownian motion
4. Brownian motion run with a nonlinear clock
5. Fractional α-stable Lévy motion
6. α-stable Lévy motion run with a nonlinear clock
7. Fractional Brownian motion run with a nonlinear clock
8. Fractional α-stable Lévy motion run with a nonlinear clock

If it is not assumed that there is mean zero drift, this list can be doubled in length with the second half being identical to the above except that "with drift" is appended to each line.

2.3 Maximum Likelihood

Let $X(t)$ be the trajectory of a particle that takes positions $x_0, x_1, \ldots, x_m$ at times $t_0, t_1, \ldots, t_m$, respectively. The likelihood function for a Brownian motion is given by a product of probability density functions for normal distributions,

$$f_X(\sigma) = \prod_{i=1}^{m} \frac{\exp[-(x_i - x_{i-1})^2/2\sigma^2(t_i - t_{i-1})]}{\sqrt{2\pi\sigma^2(t_i - t_{i-1})}} \tag{5}$$

Similarly, the likelihood function for an α-stable Lévy motion is given by a product of probability density functions for α-stable distributions,

$$f_X(\alpha, \beta, \gamma) = \prod_{i=1}^{m} f_{\alpha,\beta,\gamma(t_i - t_{i-1})^{1/\alpha},0}(x_i - x_{i-1}) \tag{6}$$

where $f_{\alpha,\beta,\gamma,0}$ is the probability density function of an α-stable distribution with skewness β, scale parameter σ, and location parameter 0. $f_{\alpha,\beta,\gamma,0}$ has no closed analytical form ([31]), but can be computed numerically. Fractional Brownian motion has correlated increments, and thus, the likelihood function cannot be decomposed into the product of the likelihoods of the increments. Since fractional Brownian motion is a Gaussian process [32], the likelihood function is given by the probability density function of a multivariate normal distribution

$$f_X(\sigma, H) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp(-\mathbf{x}' \Sigma^{-1} \mathbf{x}/2) \tag{7}$$

where Σ is the covariance matrix whose elements are given by [31]

$$\Sigma_{i,j} = \frac{\sigma^2}{2}(|t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H}) \tag{8}$$

The likelihood function for processes with a nonlinear clock can be computed using the clock, $\mathfrak{C}(t; \mathbf{p})$ (where $\mathbf{p}$ are the parameters for the clock), and the likelihood function for the underlying stochastic process by simply transforming the time coordinate. The likelihood function for Brownian motion with a nonlinear clock is given by

$$f_X(\sigma, \mathbf{p}) = \prod_{i=1}^{m} \frac{\exp[-(x_i - x_{i-1})^2/2\sigma^2(\mathfrak{C}(t_i; \mathbf{p}) - \mathfrak{C}(t_{i-1}; \mathbf{p}))]}{\sqrt{2\pi\sigma^2(\mathfrak{C}(t_i; \mathbf{p}) - \mathfrak{C}(t_{i-1}; \mathbf{p})}} \tag{9}$$

A technique for computing the maximum likelihood parameters associated with the stochastic models is necessary in order to apply the information criteria. Since the goal is to develop a framework that stochastic processes can be added to with minimal effort, analytical techniques for determining maximum likelihood parameters are passed over in favor of a numerical techniques. The likelihood functions are implemented in a Python/C code called stochastic process identification for diffusion (SPROID). Numerical optimization procedures are used to determine the maximum-likelihood parameters for a given trajectory. A number of different constrained optimization routines including truncated Newton [38], L-BFGS-B [39], and a brute force approach are available within the code. The default method is the truncated Newton method which has been used to obtain the results presented here. The truncated Newton method proceeds iteratively. In brief, at each iteration, a search direction is obtained by solving

$$H(f)\mathbf{p} = -\nabla f \tag{10}$$

where f is the likelihood function, $H(f)$ is the Hessian of f. A line search along $\mathbf{p}$ is then performed to obtain a better estimate of the maximum likelihood parameters. The iteration proceeds until an approximately optimal value is obtained. This method has proven sufficient for finding maximum likelihood parameters in the problems we have explored. It is prudent to test the different optimization techniques that are available within

Table 1 Results from the tests for Brownian motion (BM), fractional Brownian motion (FBM), symmetric α-stable Lévy motion (SLM) and Brownian motion with a power-law clock (BMPLC)

	BM	FBM	SLM	BMPLC
# Correct	68	96	97	97
True parameters	$\overline{\sigma} = 1$	$\overline{\sigma} = 1$	$\overline{\gamma} = 1$	$\overline{\sigma} = 1$
		$\overline{H} = 0.75$	$\overline{\alpha} = 1.5$	$\overline{p} = 0.5$
Average ML parameters	$\overline{\sigma} = 1.01$	$\overline{\sigma} = 1.01$	$\overline{\gamma} = 0.99$	$\overline{\sigma} = 0.96$
		$\overline{H} = 0.75$	$\overline{\alpha} = 1.53$	$\overline{p} = 0.53$
SD	$\sigma : 0.08$	$\sigma : 0.24$	$\gamma : 0.16$	$\sigma : 0.20$
		$H : 0.07$	$\alpha : 0.16$	$p : 0.13$

The # Correct gives the number of times the model was chosen out of 100 given that it was the true model

the code to see which one works best for a given problem. If all else fails, the brute force technique can be used which evaluates the likelihood function on a relatively dense grid.

It is necessary to specify bounds for each of the parameters within which the maximum-likelihood parameters are sought. The code also allows for a log-transformation of any or all of the parameters in the case that there is uncertainty in the order of magnitude of the parameter. At present, all the items listed in Sect. 2.2 except number 5 and 8 are implemented in the SPROID code. SPROID is under active development, and more processes are being added. SPROID is in the process of being released under an open source license and will be available at http://sproid.lanl.gov.

3 Results

3.1 Tests

To verify the reliability of the proposed methodology several tests were carried out for a variety of stochastic models (Brownian motion, symmetric α-stable Lévy motion, fractional Brownian motion, and Brownian motion with a power-law clock [40]). For each process, 100 sample trajectories over a dimensionless time interval from 0 to 9.9 with a step size of 0.1 were computed and analyzed. After that we applied our code to perform blind identification. Each of the 400 sample trajectories are analyzed with the four stochastic models, the maximum-likelihood parameters are identified and the best model is selected using AIC. The results are summarized in Table 1. The combined running time of the tests was approximately 28 minutes on a 3 GHz dual-core laptop.

In the cases of the more complex models (FBM, SLM, BMPLC), the AIC does a good job of selecting the correct process—96 or 97 times out of 100. For Brownian motion—68 times out of 100. However, in this case, the maximum likelihood parameters for the more complex models essentially reduce them (all of which contain Brownian motion as a special case) to something very close to Brownian motion. For example, the average value of H was 0.49 ($H = 1/2$ corresponds to Brownian motion), the average value of α was 1.98 ($\alpha = 2$ corresponds to Brownian motion), and the average value of p was 0.94 ($p = 1$ corresponds to Brownian motion). The overall results indicate that the AIC is capable of correctly selecting the correct model and that the maximum likelihood parameters are close to the true values with a relatively small standard deviation.

3.2 Applications

3.2.1 Polymer Diffusion

The first application we consider is based on a Rouse model [41] in one dimension of the diffusion of the central monomer in a long polymer [14]. The Hamiltonian for a polymer chain with N monomers is given by

$$H(\mathbf{x}) = \frac{K}{2}\Sigma_{i=1}^{N-1}(x_{i+1} - x_i)^2 \tag{11}$$

and the Boltzmann factor for the separation between two neighboring monomers is

$$\exp\left[\frac{-\beta K(x_{i+1} - x_i)^2}{2}\right] \tag{12}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the absolute temperature, K is a spring constant, and x_i are the positions of the monomers. The model is simulated using the same Monte Carlo (MC) approach described in [14]. One MC time unit/step is said to elapse after N MC moves. Each MC move consists of randomly selecting a monomer, proposing a displacement, then accepting or rejecting it based on the ratio of the Boltzmann factors.

This is a good test case because it is a physics-based model that exhibits anomalous diffusion on a short time scale and classical diffusion on a long time scale. On the short time scale, there are indicators that fractional Brownian motion with a Hurst exponent of $H = 1/4$ is an appropriate stochastic model based on the limit of an infinitely long polymer [42], and on the behavior near an absorbing boundary for a finite polymer [40]. A study for long polymers was carried out using the same approach as in Sect. 3.1. To study the short time scale behavior, the trajectories of the central monomer of 100 polymers containing 129 monomers were tracked for 1,000 Monte Carlo steps with every 10th step being recorded. The relaxation time from anomalous to classical behavior for this monomer is on the order of 1.4×10^4 Monte Carlo steps [14]. To study the long time scale behavior, the trajectories of the central monomer of 100 polymers containing 33 monomers were tracked for 10^6 Monte Carlo steps with every 10,000th step being recorded. A shorter polymer was used for the long time scale because fewer steps are required to approach the long time limit.

The results are recorded in Tables 2 and 3 for the short and long time scales, respectively. The table indicates that the method used here identifies fractional Brownian motion with $H = 0.25$ as the preferred model for the short time scale. The table indicates that the method employed here chooses Brownian motion as the preferred model on the long time scale 64 out of 100 times which is similar to the result for when true Brownian motion is used. As before, the estimated parameter values for the more complex models also indicate that Brownian motion or a model very close to it is the preferred model for the long time scale (the preferred complex models are with $\bar{H} \approx 0.5$, $\bar{\alpha} \approx 2$ and $\bar{p} \approx 1$, for the FBM, SLM and BMPLC, respectively, i.e. these are very close to Brownian motion).

This analysis further supports the case that fractional Brownian motion with $H = 1/4$ is an appropriate stochastic model for the trajectory of the central monomer on the short time scale, and Brownian motion is an appropriate model on the long time scale. It also demonstrates the applicability of the developed method to real data sets.

Table 2 Results from the analysis of the trajectories of the central monomer on short time scales

	BM	FBM	SLM	BMPLC
# Selected	0	100	0	0
Average ML parameters	$\overline{\sigma} = 0.32$	$\overline{\sigma} = 0.57$	$\overline{\gamma} = 0.22$	$\overline{\sigma} = 0.94$
		$\overline{H} = 0.25$	$\overline{\alpha} = 1.98$	$\overline{p} = 0.41$
SD	$\sigma : 0.02$	$\sigma : 0.10$	$\gamma : 0.02$	$\sigma : 0.17$
		$H : 0.05$	$\alpha : 0.05$	$p : 0.09$

Table 3 Results from the analysis of the trajectories of the central monomer on long time scales

	BM	FBM	SLM	BMPLC
# Selected	64	17	2	17
Average ML parameters	$\overline{\sigma} = 0.08$	$\overline{\sigma} = 0.13$	$\overline{\gamma} = 0.06$	$\overline{\sigma} = 0.24$
		$\overline{H} = 0.47$	$\overline{\alpha} = 1.99$	$\overline{p} = 0.96$
SD	$\sigma : 0.005$	$\sigma : 0.04$	$\gamma : 0.006$	$\sigma : 0.39$
		$H : 0.07$	$\alpha : 0.04$	$p : 0.19$

Table 4 Results from the diffusion of RNA molecules in living *E. coli* cells in the x direction

	BM	FBM	SLM	BMPLC
# Selected	0	15	11	1
Average ML parameters	$\overline{\sigma} = 0.064$	$\overline{\sigma} = 0.065$	$\overline{\gamma} = 0.035$	$\overline{\sigma} = 0.079$
		$\overline{H} = 0.34$	$\overline{\alpha} = 1.75$	$\overline{p} = 1.00$
SD	$\sigma : 0.043$	$\sigma : 0.045$	$\gamma : 0.014$	$\sigma : 0.058$
		$H : 0.044$	$\alpha : 0.15$	$p : 0.32$

3.2.2 Diffusion of RNA Molecules in Living E. coli Cells

Observations of the diffusion of RNA molecules in living *E. coli* cells [7] indicate that the motion is subdiffusive with mean square displacement growing proportional to $t^{0.35}$. There is some evidence to suggest that the motion of the RNA molecules is self-similar with self-similarity index equal to 0.35 [43], and a test has been performed to determine whether the motion arises from fractional Brownian motion or a subordinated Brownian motion both with self-similarity index equal to 0.35 [44]. The conclusion of this test was that the motion is likely a fractional Brownian motion.

We applied the technique developed here to examine how appropriate Brownian motion, fractional Brownian motion, symmetric Lévy motion and Brownian motion with nonlinear clocks are as models for the motion of the RNA molecules. The results are summarized in Tables 4 and 5 for the x and y coordinates respectively. As expected, fractional Brownian motion is the model that is most often selected (32 out of 54). These results as well as the results in [44] indicate that fractional Brownian motion with $H = 0.35$ does a good job of capturing some of the non-Markovian effects of the trajectories. It is surprising, however, that symmetric Lévy motion is selected in 19 out of 54 trajectories. This indicates that fractional Brownian motion is missing the tendency for the molecules to move relatively large distances in short times (see, *e.g.*, Fig. 1b in [7]). A model such as fractional Lévy motion may be able to capture both of these effects, but we will explore this possibility in future research.

Table 5 Results from the diffusion of RNA molecules in living *E. coli* cells in the y direction

	BM	FBM	SLM	BMPLC
# Selected	0	17	8	2
Average ML parameters	$\overline{\sigma} = 0.058$	$\overline{\sigma} = 0.059$	$\overline{\gamma} = 0.036$	$\overline{\sigma} = 0.075$
		$\overline{H} = 0.36$	$\overline{\alpha} = 1.81$	$\overline{p} = 1.00$
SD	$\sigma : 0.024$	$\sigma : 0.025$	$\gamma : 0.014$	$\sigma : 0.062$
		$H : 0.050$	$\alpha : 0.13$	$p : 0.23$

Table 6 Results from the longitudinal dispersion of RAFOS floats in the Gulf of Mexico

	BM	FBM	SLM	BMPLC
# Selected	0	58	4	1
Average ML parameters	$\overline{\sigma} = 0.0075$	$\overline{\sigma} = 0.036$	$\overline{\gamma} = 0.046$	$\overline{\sigma} = 0.21$
		$\overline{H} = 0.81$	$\overline{\alpha} = 1.86$	$\overline{p} = 0.59$
SD	$\sigma : 0.0036$	$\sigma : 0.0033$	$\gamma : 0.0026$	$\sigma : 0.18$
		$H : 0.14$	$\alpha : 0.17$	$p : 0.56$

Table 7 Results from the latitudinal dispersion of RAFOS floats in the Gulf of Mexico

	BM	FBM	SLM	BMPLC
# Selected	0	56	3	4
Average ML parameters	$\overline{\sigma} = 0.0058$	$\overline{\sigma} = 0.0030$	$\overline{\gamma} = 0.0036$	$\overline{\sigma} = 0.17$
		$\overline{H} = 0.78$	$\overline{\alpha} = 1.89$	$\overline{p} = 0.61$
SD	$\sigma : 0.0029$	$\sigma : 0.0014$	$\gamma : 0.0022$	$\sigma : 0.15$
		$H : 0.13$	$\alpha : 0.15$	$p : 0.55$

3.2.3 Dispersion of RAFOS Floats in the Gulf of Mexico

The data used to study the dispersion of RAFOS floats [11] in the Gulf of Mexico has previously been found to exhibit superdispersive scaling behavior [45]. A description of the data collection and an extensive analysis can be found in [46]. The floats travel with the current at various depths ranging from 1,000 to 3,000 m for the different floats. The initial positions and the times at which the floats were released vary as well. The method was applied to 63 trajectories with float-position data available every 8 hours. Some floats provided trajectory data for about a week while the longest trajectory lasts for approximately 7 months.

Our method indicates (Tables 6 and 7 that fractional Brownian motion is the most viable model for the dispersion of these drifters among the candidate models. Fractional Brownian motion was selected 58 and 56 times out of 63 for the longitudinal and latitudinal increments, respectively. While observing animations of these trajectories (see the supporting information in [45]), the non-Markovian nature of the dispersion is clear—there are strong correlations in the velocity that appear to persist for some time. The selection of fractional Brownian motion does not appear to depend strongly upon the time scale. For trajectories that lasted more than 100 days, FBM was chosen 59 out of 62 times. In the shorter trajectories, it was chosen 55 out of 64 times.

4 Conclusion

A method for selecting stochastic models for diffusive and dispersive trajectories has been described. Synthetic tests indicate that the method is capable of selecting the appropriate model with fairly good reliability. The examination of the polymer diffusion served as a bridge between these synthetic test cases and physical processes. The reliability of the method for the polymer diffusion example was very similar to the reliability in the synthetic examples. In the polymer example, we were able to correctly identify fractional Brownian motion on the short time scale and Brownian motion on the long time scale. An analysis was undertaken of diffusion of RNA molecules in *E. coli*. This analysis indicated that fractional Brownian motion with $H \approx 0.35$ is the best among these two-parameter models for this type of diffusion and this supports previous results [44]. However, there was also evidence that fractional Brownian motion was not capturing the potential for the RNA molecules to move large distances in short times. We suggest that fractional Lévy motion is a more complicated model that may be able to capture this diffusive behavior more accurately, but leave this avenue to be explored in future work. A final application to dispersion in the Gulf of Mexico was considered, and the method indicated that fractional Brownian motion with $H \approx 0.8$ is the most appropriate model for this dispersion among the candidates. This corroborates the scaling behavior found in [45].

Acknowledgments The authors wish to thank two anonymous reviewers for feedback that substantially improved the manuscript. DO and VVV wish to acknowledge the Environmental Programs Directorate of the Los Alamos National Laboratory; the Advanced Simulation Capability for Environmental Management (ASCEM) project, Department of Energy, Environmental Management; and the Integrated Multifaceted Approach to Mathematics at the Interfaces of Data, Models, and Decisions (DiaMonD) project, Department of Energy, Office of Science. JHC wishes to acknowledge NSF Grant #EAR 1314828. The authors wish to thank Ido Golding and Edward C. Cox for the RNA data as well as BOEMRE for the Gulf of Mexico float data.

References

1. Brown, R.: A brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Philos. Mag. Ann. Chem. Math. Astron. Nat. Hist. Gen. Sci. **4**(21), 161–173 (1828)
2. Einstein, A.: On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. Annalen der Physik **17**(549–560), 16 (1905)
3. Ingenhousz, J.: Nouvelles Expériences et Observations sur Divers Oobjets de Physique. chez P. Théophile Barrois le jeune, Paris (1785)
4. Tsay, R.S.: Analysis of Financial Time Series. Wiley, Hoboken (2005)
5. Taylor, S.J.: Modelling Financial Time Series. World Scientific Publishing, Singapore (2008)
6. Saxton, M.J., Jacobson, K.: Single-particle tracking: applications to membrane dynamics. Ann. Rev. Biophys. Biomol. Str. **26**(1), 373–399 (1997)
7. Golding, I., Cox, E.C.: Physical nature of bacterial cytoplasm. Phys. Rev. Lett. **96**(9), 098102 (2006)
8. Regner, B.M., Vučinić, D., Domnisoru, C., Bartol, T.M., Hetzer, M.W., Tartakovsky, D.M., Sejnowski, J.: Anomalous diffusion of single particles in cytoplasm. Biophys. J. **104**(8), 1652–1660 (2013)
9. Viswanathan, G.M., Afanasyev, V., Buldyrev, S.V., Prince, P.A., Stanley, H.E.: Levy flight search patterns in animal behavior. Nature **381**, 413–415 (1996)
10. Sims, D.W., Southall, E.J., Humphries, N.E., Hays, G.C., Bradshaw, C.J.A., Pitchford, J.W., James, A., Ahmed, M.Z., Brierley, A.S., Hindell, M.A., et al.: Scaling laws of marine predator search behaviour. Nature **451**(7182), 1098–1102 (2008)
11. Rossby, T., Dorson, D., Fontaine, J.: The rafos system. J. At. Ocean. Technol. **3**(4), 672–679 (1986)
12. Moroni, M., Cushman, J.H.: Three-dimensional particle tracking velocimetry studies of the transition from pore dispersion to fickian dispersion for homogeneous porous media. Water Resour. Res. **37**(4), 873–884 (2001)

13. Moroni, M., Cushman, J.H.: Statistical mechanics in 3d-ptv experiments in the study of anomalous dispersion: part ii experiment. Phys. Fluids **13**(1), 81–91 (2001)
14. Kantor, Y., Kardar, M.: Anomalous diffusion with absorbing boundary. Phys. Rev. E **76**, 061121 (Dec 2007)
15. Crandall, C.A., Kauffman, L.J., Katz, B.G., Metz, P.A., McBride, W.S., Berndt, M.P.: Simulations of ground-water flow and particle tracking analysis in the area contributing recharge to a public supply well near Tampa, FL. US Geol Surv Sci Invest Rep, 5231, (2008)
16. Bijeljic, Branko, Mostaghimi, Peyman, Blunt, Martin J.: Signature of non-fickian solute transport in complex heterogeneous porous media. Phys. Rev. Lett. **107**, 204502 (Nov 2011)
17. Ono, J., Ohshima, K.I., Uchimoto, K., Ebuchi, N., Mitsudera, H., Yamaguchi, H.: Particle-tracking simulation for the drift/diffusion of spilled oils in the sea of okhotsk with a three-dimensional, high-resolution model. J. Oceanogr. **69**(4), 413–428 (2013)
18. O'Malley, D., Vesselinov, V.V.: Analytical solutions for three-dimensional reaction-advection-dispersion with anomalous dispersion. Adv. Water Resour. In Submission.
19. Akaike, Hirotugu: A new look at the statistical model identification. Autom. Control IEEE Trans. **19**(6), 716–723 (1974)
20. Hurvich, C.M., Tsai, C.L.: Regression and time series model selection in small samples. Biometrika **76**(2), 297–307 (1989)
21. Burnham, K.P., Anderson, D.R.: Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. Springer, Berlin (2002)
22. Ye, M., Meyer, P.D.: On model selection criteria in multimodel analysis. Water Resour. Res. **44**(3), 07 (2008)
23. O'Malley, D., Cushman, J.H.: The ubiquity of, and geostatistics for, nonstationary increment random fields. Water Resour. Res. **49**(7), 4525–4529 (2013)
24. Falconer, K.: Fractal Geometry: Mathematical Foundations and Applications. Wiley, Hoboken (2007)
25. Gefen, Yuval, Aharony, Amnon, Alexander, Shlomo: Anomalous diffusion on percolating clusters. Phys. Rev. Lett. **50**(1), 77 (1983)
26. Bouchaud, Jean-Philippe, Georges, Antoine: Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. Phys. Rep. **195**(4), 127–293 (1990)
27. Solomon, T.H., Weeks, E.R., Swinney, H.L.: Observation of anomalous diffusion and lévy flights in a two-dimensional rotating flow. Phys. Rev. Lett. **71**(24), 3975 (1993)
28. Tsallis, C., Bukman, D.J.: Anomalous diffusion in the presence of external forces: exact time-dependent solutions and their thermostatistical basis. Phys. Rev. E **54**(3), R2197 (1996)
29. Metzler, Ralf, Klafter, Joseph: The random walk's guide to anomalous diffusion: a fractional dynamics approach. Physics Rep. **339**(1), 1–77 (2000)
30. Sokolov, I.M.: Models of anomalous diffusion in crowded environments. Soft Matter **8**(35), 9043–9052 (2012)
31. Samorodnitsky, G., Taqqu, M.S.: Stable Non-Gaussian Processes. Chapman & Hall, New York (1994)
32. Benoit, B.: Fractional brownian motions, fractional noises and applications. SIAM Rev. **10**(4), 422–437 (1986)
33. Kolmogorov, A.N.: Wienersche spiralen und einige andere interessante kurven im hilbertschen raum. C. R. (Doklady) Acad. Sci. URSS (N.S.) **26**, 115–118 (1940)
34. Taqqu, M.S.: Benoît mandelbrot and fractional brownian motion. Statist. Sci. **28**(1), 1–5 (2013)
35. Lim, S.C., Muniandy, S.V.: Self-similar gaussian processes for modeling anomalous diffusion. Phys. Rev. E **66**(2), 021114 (2002)
36. Cushman, J.H., O'Malley, D.: Anomalous diffusion as modeled by a nonstationary extension of brownian motion. Phys. Rev. E **79**(3), 032101 (2009)
37. O'Malley, D., Cushman, J.H.: Fractional brownian motion run with a nonlinear clock. Phys. Rev. E **82**(3), 032102 (2010)
38. Nash, S.G.: Preconditioning of truncated-newton methods. SIAM J. Sci. Statist. Comput. **6**(3), 599–616 (1985)
39. Byrd, R., Lu, P., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. **16**(5), 1190–1208 (1995)
40. O'Malley, D., Cushman, J.H.: Scaling laws for fractional brownian motion with power-law clock. J. Statist. Mech.: Theory Exper. **01**, L01001 (2011)
41. Rouse, P.E.: A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. J. Chem. Phys. **21**(7), 1272–1280 (1953)
42. Panja, D.: Probabilistic phase space trajectory description for anomalous polymer dynamics. J. Phys. **23**(10), 105103 (2011)

43. O'Malley, D., Cushman, J.H.: Random renormalization group operators applied to stochastic dynamics. J. Statist. Phys. **149**(5), 943–950 (2012)
44. Magdziarz, Marcin, Weron, Aleksander, Burnecki, Krzysztof, Klafter, Joseph: Fractional brownian motion versus the continuous-time random walk: a simple test for subdiffusive dynamics. Phys. Rev. lett. **103**(18), 180602 (2009)
45. O'Malley, D., Cushman, J.H., Johnson, G.: Random renormalization groups and bayesian scaling of dispersion/diffusion in lake michigan and the gulf of mexico. Geophys. Research Lett. **40**(17), 4638–4642 (2013)
46. Donohue, K., Hamilton, P., Leaman, K., Leben, R., Prater, M., Waddell, E., Watts, R.: Exploratory study of deepwater currents in the Gulf of Mexico. Technical Report MMS 2006–073, U.S. Department of the Interior, Minerals Management Service, Gulf of Mexico OCS, Region, 2006.