A novel computational method (MASD: Moving Average Squared Displacement) to distinguish Brownian and non-Brownian trajectories

J. Palhalmi*

*East-West Technologies Ltd.,
Budapest, Hungary

Abstract. In bioinformatics and computational biology, the mean square displacement (MSD) analysis is one of the most commonly applied methods to describe systems dynamics and to determine the diffusion coefficient of one, two, or three-dimensional trajectories. Under biological circumstances the time domain is rarely stationary, so the length of trajectories is extremely short. To our best knowledge, in the case of extremely short (N=100 and N=15) trajectory lengths there is no systematic comparative statistical study which demonstrates the positive performance of any analytical method compared with mean square displacement analysis. A new approach has been developed to distinguish slightly different trajectories at the population level.

1. Introduction

Under ideal conditions the solution of Einstein's diffusion equation is the normalized Gaussian probability density function (PDF), and the variance of this function is equal to the mean square displacement (MSD) [1]. In this case, if the particle trajectories are long enough, the ensemble averages (EAMSD) are equal to the time averages (TAMSD). In the case of biological trajectories, the conditions are fundamentally different, therefore the MSD analysis has its natural limitations.

Although TAMSD and EAMSD methods provide various useful information, these have certain limitations. They are unable to discriminate between trajectory patterns with slight differences (Hurst parameter 0.01). In addition, if the system is non-Brownian and the trajectories are short, further uncertainties are introduced in the calculations [2]. Accordingly, there is an urgent need to develop a conceptually novel approach for the analysis of trajectories sampled from different fields of applications.

A systematic computational statistical approach was worked out to develop a new computational method (MASD: moving average squared displacement) to characterize trajectories. The sensitivity of MASD method to quantify slight differences in displacement patterns of trajectories was compared with the sensitivity of the classical MSD method.

2. Methods

Fractional Brownian motion equation was used to simulate classical Brownian and non-Brownian trajectories by adjusting Hurst parameter (H) between 0.9 and 0.1 levels. Based on the FBM equation, generation of classical Brownian and non-Brownian trajectories was achieved by the adjustment of the Hurst parameter [3].

$$B_H(t) := \frac{1}{\Gamma(H+1/2)} \left(\int_{-\infty}^{0} [(t-s)^{H-1/2} - (-s)^{H-1/2}] \, dB(s) + \right.$$
$$\left. + \int_{0}^{t} (t-s)^{H-1/2} \, dB(s) \right), \tag{1}$$

where Γ represents the gamma function, H is the Hurst parameter, t is time, s is the increment and B is the displacement of the particles.

The classical Brownian motion was simulated with a Hurst parameter of 0.5 and it shows linear MSD relation. In case of non-Brownian data sees ($0 < H < 0.5$) the limit value is dependent on the initial value, which determines the nonlinearity of the MSD curves. Trajectory simulations were coded under Octave / Matlab environment by applying the wavelet based FBM function [4].

For the comparative statistical experiments, one dimensional trajectories with different lengths ($N = 15$ and $N = 100$, data points) were simulated in the above mentioned Hurst parameter ranges.

Determination of diffusion coefficient based on mean square displacement (MSD) calculation (D MSD):

$$MSD_{(n\Delta t)} = \frac{1}{N-n-1} \sum_{i=1}^{N-n-1} (x(i+n\Delta t) - x(i))^2, \tag{2}$$
$$MSD_{(n\Delta t)} = 2dD_{MSD(n\Delta t)}\Delta t = (x(i+n\Delta t) - x(i))^2$$

where d is degree of freedom (it is 1 in one dimensional simulated trajectories), $x(i)$ is the coordinate of the trajectory time series, Δt is time interval, n is number of a given data point within the trajectory, $D_{MSD(n\Delta t)}$ is diffusion coefficient estimated from the slope of the MSD curve.

Determination of diffusion coefficient based on moving average squared displacement (MASD) calculation (D MASD):

Since the limit value of the ensembled and time averaged MSD is equal if the system is ergodic and equal with $2D_{MSD(n\Delta t)}\Delta t$ for each coordinates if the trajectory is long enough, performing the moving window averaging on (2) fixing the time lag at $n = 1$, establishes the possibility of observing the fluctuations on the microscale displacement events. In other words, averaging this window means over the trajectories in the population unifies the advantages of the time averaged and ensemble MSD methods.

Two sample Kolmogorov-Smirnov test was used to estimate the difference between the distributions of trajectory populations represented by diffusion coefficients (D) calculated by MASD and MSD methods ($n = 1000$).

3. Results

Simulated trajectory populations deviating from Brownian state (Hurst parameter $H < 0.5$) were analyzed to compare the sensitivity of the MASD and the classical MSD methods. The population statistical values of diffusion coefficients represented by boxplots are plotted in the function of Hurst parameter ($n = 1000$, $N = 15$) (Fig. 1). Note the remarkable difference between D MASD and D MSD values in the range of $0.1 < H < 0.5$.
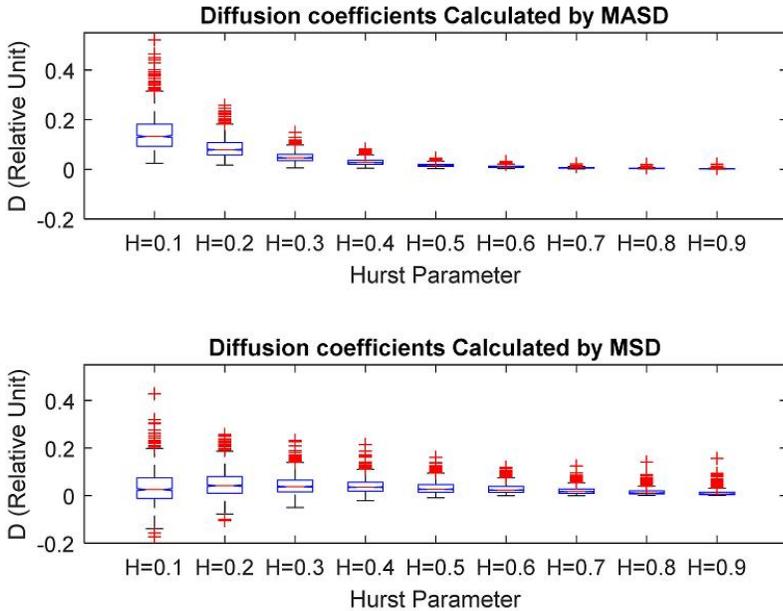

Figure 1

To show further evidence for the capability of the MASD method to distinguish trajectory patterns, the distributions of the D values calculated from non-Brownifn ($H = 0.49$) and Brownian trajectories ($H = 0.5$) were compared with two sample Kolmogorov–Smirnov test ($n = 1000$). Even in

case of extremely short ($N = 100$ and $N = 15$) trajectory populations, the MASD method proved to be mere sensitive to find statistically significant difference ($N = 15$: D MASD: $p = 0.0019$, D MSD: $p = 0.2350$, $N = 100$: D MASD: $p = 2.383 * 10^{-35}$, D MSD: $p = 0.0106$, see Fig. 2.)
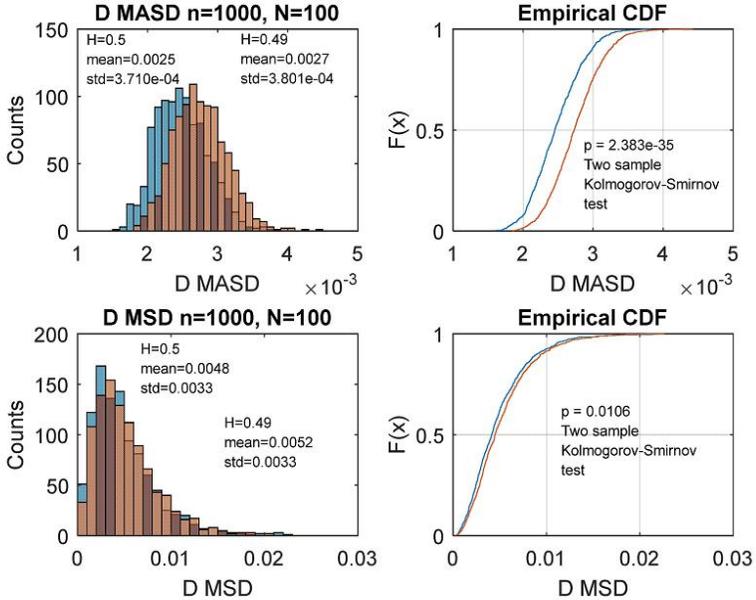

Figure 2

4. Conclusions

The comparative analytical results clearly demonstrate the robustness of the MASD calculation method to quantify differences at Hurst parameter $= 0.01$ level in trajectory patterns even in case of extremely short ($N = 100$ and $N = 15$) trajectories. The performance of MASD calculation method will be discussed compared with the classical MSD based calculation methods under different simulation circumstances.

References

1. *Einstein A.* On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat // Annalen der Physik. — 1906. — Vol. 17. — P. 549–560.
2. *Grebenkov D. S. et al.* Time-averaged quadratic functionals of a Gaussaan process // Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys. — 2011. — Vol. 83, no. 6, part 1. — Article ID 061117.
3. *Mandelbrot B. B., Van Ness J. W.* Fractional Brownian Motions, Fractional Noises and Applications // SIAM Review. — 1968. — Vol. 10. — P. 422–437.
4. *Abry P., Sellan F.* The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation // Appl. and Comp. Harmonic Anal. — 1996. — Vol. 3, no. 4. — P. 377–383.