Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors. FIMA approach

Krzysztof Burnecki

Hugo Steinhaus Center
Wroclaw University of Technology
Poland

KIAS Workshop on Anomalous Dynamics in Biological Systems, 2015

Outline

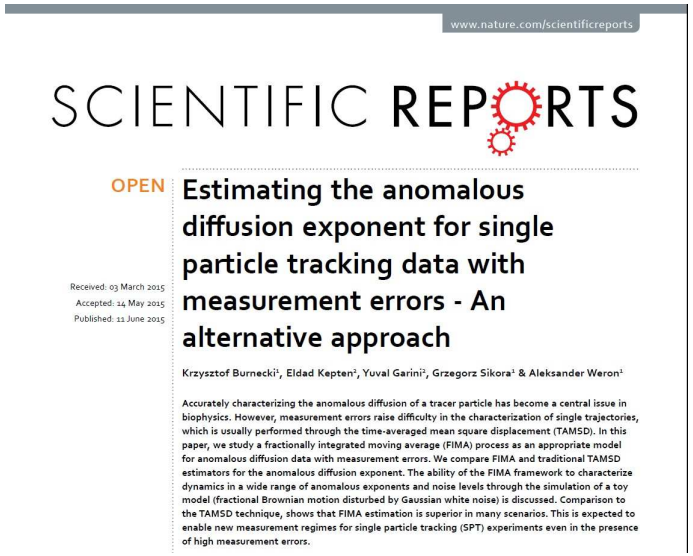
Outline

- Autoregressive fractionally integrated moving average (ARFIMA) model

Outline

- Autoregressive fractionally integrated moving average (ARFIMA) model

- Single particle tracking data with measurement errors. MSD vs. FIMA approach (Burnecki, Kepten, Garini, Sikora and Weron, Scientific Reports, 2015)



www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Received: 03 March 2015
Accepted: 14 May 2015
Published: 11 June 2015

Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors - An alternative approach

Krzysztof Burnecki[1], Eldad Kepten[2], Yuval Garini[2], Grzegorz Sikora[1] & Aleksander Weron[1]

Accurately characterizing the anomalous diffusion of a tracer particle has become a central issue in biophysics. However, measurement errors raise difficulty in the characterization of single trajectories, which is usually performed through the time-averaged mean square displacement (TAMSD). In this paper, we study a fractionally integrated moving average (FIMA) process as an appropriate model for anomalous diffusion data with measurement errors. We compare FIMA and traditional TAMSD estimators for the anomalous diffusion exponent. The ability of the FIMA framework to characterize dynamics in a wide range of anomalous exponents and noise levels through the simulation of a toy model (fractional Brownian motion disturbed by Gaussian white noise) is discussed. Comparison to the TAMSD technique, shows that FIMA estimation is superior in many scenarios. This is expected to enable new measurement regimes for single particle tracking (SPT) experiments even in the presence of high measurement errors.

ARFIMA (another acronym FARIMA) process

Autoregressive fractionally integrated moving average (ARFIMA) process $\{X_t\}$, denoted by ARFIMA(p, d, q), introduced by Granger and Joyeux (J. Time Ser. Anal., 1980) and Hosking (Biometrika, 1981), is defined by equation

$$\begin{aligned}(1-B)^d \quad & [X(t) - \phi_1 X(t-1) - \ldots - \phi_p X(t-p)] \\ = \quad & Z(t) - \psi_1 Z(t-1) - \ldots - \psi_q Z(t-q),\end{aligned}$$

where

- B is the shift operator, i.e. $B^j X_t = X_{t-j}$,
- $Z(t)$ is the noise sequence with either finite or infinite variance,
- $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is the autoregressive (AR) polynomial and
- $\Psi_q(B) = 1 - \psi_1 B - \psi_2 B^2 - \ldots - b_q B^q$ is the moving average (MA) polynomial.

ARFIMA, fractional differencing

What is new in the definition is the fractional difference operator $(1-B)^d$ with the exponent d, called the memory parameter.

The operator has the infinite binomial expansion

$$(1-B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-B)^j = \sum_{j=0}^{\infty} \pi_j B^j,$$

where

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)},$$

and Γ is the gamma function.

Noise sequence

If the noise $Z(t)$ consists of i.i.d. random variables belonging to the domain of attraction of the Lévy α-stable law, then $d < 1 - 1/\alpha$ and $0 < \alpha \leq 2$. The noise may be either Gaussian, non-Gaussian with finite variance or it may have infinite variance.

- For finite variance case, the typical choices are exponential, log-normal, normal-inverse Gaussian and truncated stable distributions.
- For infinite variance variables one may consider symmetric and skewed Lévy stable, and Pareto distributions.

ARFIMA and long memory

- It's one of the most popular examples of long memory (long-range dependent) processes.
- In the Gaussian case ($\alpha = 2$) the autocovariance function $\tau(n) = \langle X(n)X(0)\rangle - \langle X(n)\rangle\langle X(0)\rangle \sim n^{2d-1}$. Moreover, for $d > 0$ we have $\sum_{n=0}^{\infty} |\tau(n)| = \infty$. This serves as a classical definition of long memory.
- For $\alpha < 2$ the covariance does not exist and one has to replace it, e.g., with the codifference. It was proved that for $d > 1 - 2/\alpha$ ARFIMA possesses long-term dependence in the classical sense.

ARFIMA model. Applications

ARFIMA times series have been widely studied in the literature in

- telecommunication,
- industry,
- economics,
- astronomy,
- hydrology,
- image processing,
- geoscience,
- medicine.

More ARFIMA properties

- Aggregated ARFIMA$(0,0,0) \Longrightarrow$ Lévy stable motion, in particular Brownian motion.
- Aggregated ARFIMA$(0,d,0) \Longrightarrow$ fractional Lévy stable motion (FLSM) with $H = d + 1/\alpha$, in particular fractional Brownian motion (FBM) with $H = d + 1/2$.
- Contrary to FBM and FLSM it can also describe different light and heavy-tailed distributions and a short (exponential-like) memory.
- It can also model a measurement error (MA part includes the information on the error).

ARFIMA trajectories

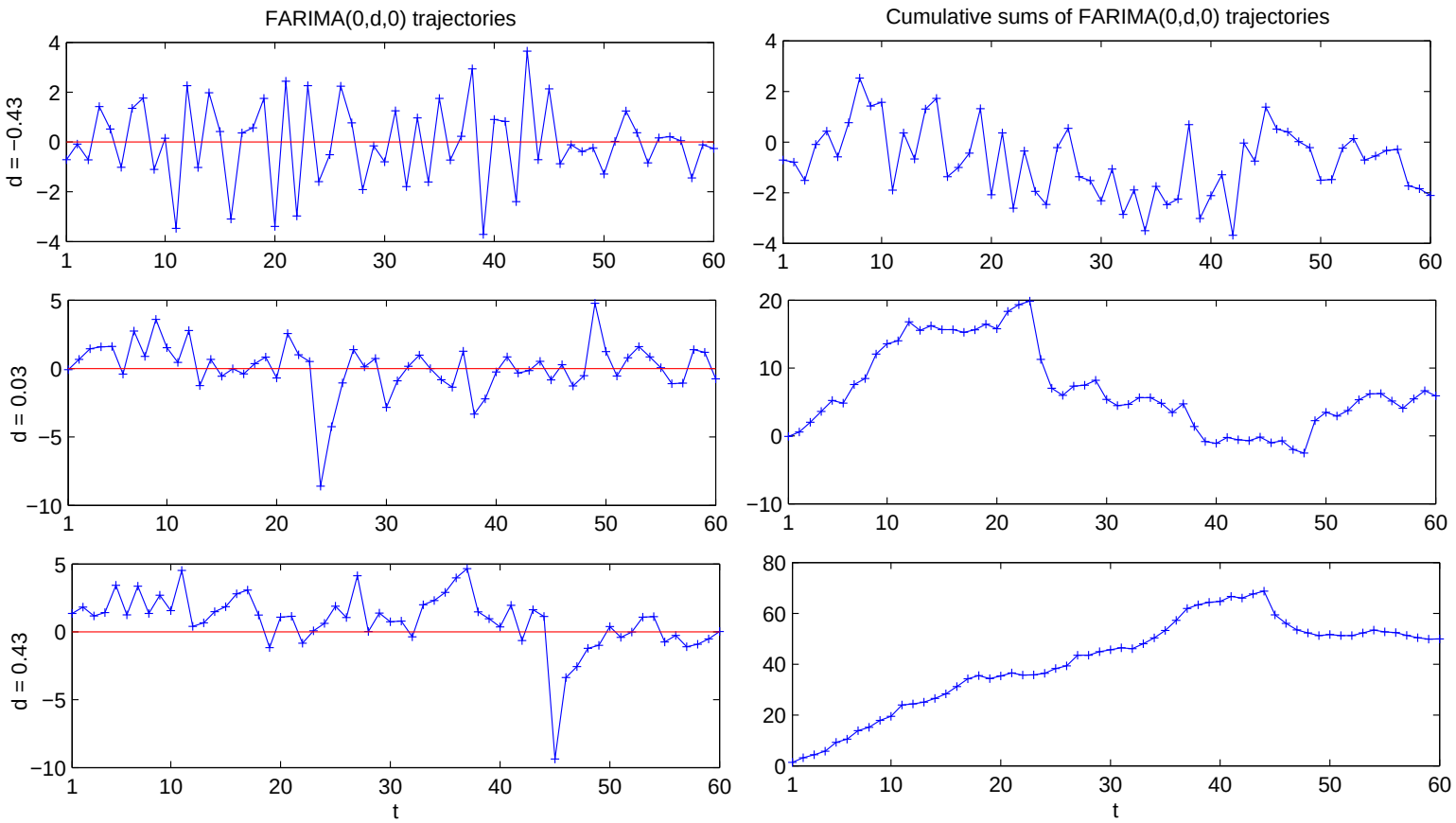


Figure 1 : Simulated trajectories of symmetric 1.8-stable ARFIMA$(0, d, 0)$ series with the memory parameter $d \in \{-0.43, 0.03, 0.43\}$ (left panel), and their cumulative sums (right panel).

ARFIMA process for $p = 0$. FIMA(d, q)

- FIMA(d, q):=ARFIMA$(0, d, q)$ is represented by the fractional difference equation:

$$(1 - B)^d X(t) = Z(t) - \psi_1 Z(t-1) - \ldots - \psi_q Z(t-q).$$

- In many applications FIMA$(d, 1)$ model is sufficient to describe the data well. FIMA$(d, 1)$ can be considered a first-order approximation of the arbitrary short memory structure (q lags) taking into account only the first lag. In this case the model reduces to:

$$(1 - B)^d X(t) = Z(t) - \psi Z(t-1), \quad t = 0, \pm 1, \ldots$$

FIMA$(d, 1)$ parameter estimation

The estimator of the parameter vector β is defined as the vector argument $\tilde{\beta}$, for which the following function attains its minimum value:

$$g(\beta) = \int_0^{\pi} \frac{\tilde{I}_N(\lambda)}{W(\lambda, \beta)} d\lambda,$$

where

$$W(\lambda, \beta) = \frac{(1 + 2\psi \cos \lambda + \psi^2)}{(2 - 2\cos\lambda)^d}$$

is the spectral density of the FIMA process.

Sample MSD

- Let $\{Z_i,\ i = 1, \ldots, N\}$ be a sample of length N. The sample (time average) MSD is defined as

$$M_N(\tau) = \frac{1}{N-\tau} \sum_{k=1}^{N-\tau} (Z_{k+\tau} - Z_k)^2.$$

- If the sample comes from an ARFIMA(p, d, q) process with noise belonging to the domain of attraction of the Lévy α-stable law, then for its partial sum process for large N

$$M_N(\tau) \sim \tau^{2d+1} S_{\alpha/2},$$

where $\sim$ means similarity in distribution, and $S_{\alpha/2}$ is a totally skewed $\alpha/2$-stable random variable (Burnecki and Weron, PRE 2010).

Sample MSD, cont.

- $d > 0 \Rightarrow$ superdiffusion
- $d = 0 \Rightarrow$ diffusion
- $d < 0 \Rightarrow$ subdiffusion

- **In general the memory parameter d controls the type of anomalous diffusion regardless of the underlying distribution**

Estimating the anomalous diffusion exponent with measurement errors (Burnecki, Kepten, Garini, Sikora and Weron, Scientific Reports, 2015)

We consider the toy model:

$$M(t) = B_H(t) + \epsilon(t), \quad t = 1, 2, \ldots, T,$$

where

- $\{B_H(t) : t = 1, 2, \ldots, T\}$ is the standard FBM, $E\left(B_H(t)^2\right) = t^{2H}$,
- $\{\epsilon(t) : t = 1, 2, \ldots, T\}$ is the measurement noise = white Gaussian noise, i.i.d. $\epsilon(t) \sim N(0, \sigma^2)$,
- $B_H(t)$ and $\epsilon(t)$ are independent.

Influence of the noise on the toy model

(Loading Video...)

Toy model

- For the toy model

$$M(t) = B_H(t) + \epsilon(t), \quad t = 1, 2, \ldots, T$$

$$Cov\left(M(t+h), M(t)\right) = Cov\left(B_H(t+h), B_H(t)\right), \quad h \neq 0$$

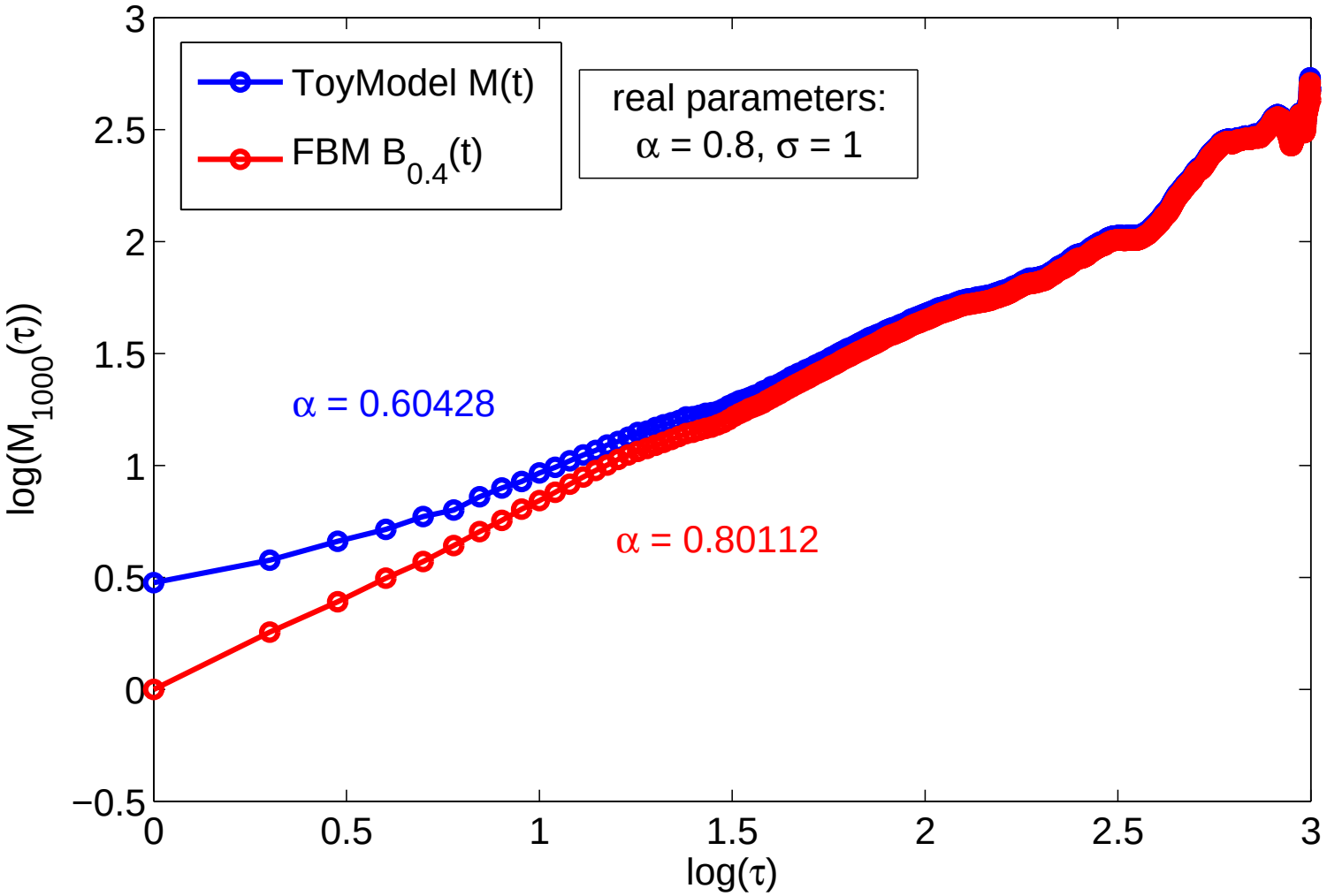
$$E\left(M(t)^2\right) = E\left(B_H(t)^2\right) + E\left(\epsilon(t)^2\right) = t^{2H} + \sigma^2$$

- For its increments

$$\rho_M(h) := Cov\left(M(t+1) - M(t), M(t+1+h) - M(t+h)\right)$$

$$\rho_M(h) = \begin{cases} \rho_H(h) + 2\sigma^2 & \text{if } h = 0 \\ \rho_H(h) - \sigma^2 & \text{if } h = 1 \\ \rho_H(h) & \text{if } h > 1 \end{cases}, \quad \rho_H(h) \sim H(2H-1)h^{2H-2}$$

Toy model MSD



Toy model covariance structure

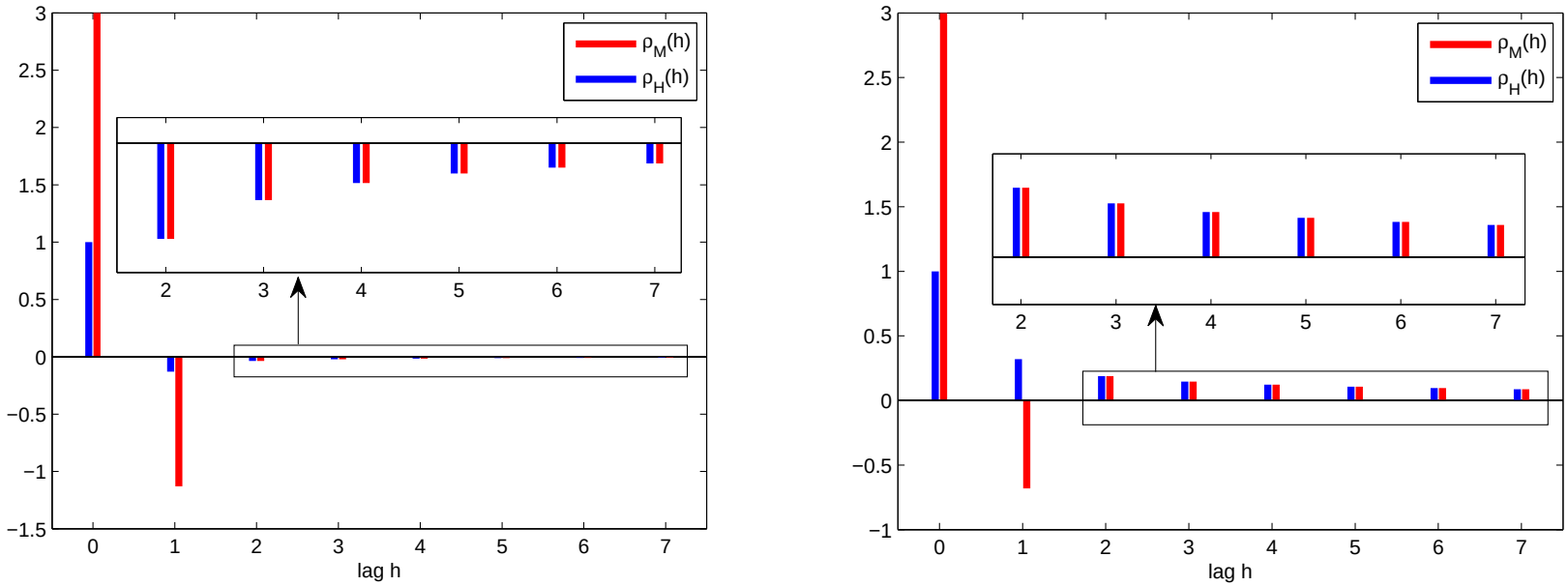


Figure 2 : Covariance function of the increments of the toy model $M(t) = B_H(t) + \epsilon(t)$ and increments of the corresponding $B_H(t)$: a) with Hurst index $H = 0.4$ and measurement error $\sigma = 1$ (left), b) with Hurst index $H = 0.7$ and measurement error $\sigma = 1$ (right).

FIMA model covariance structure

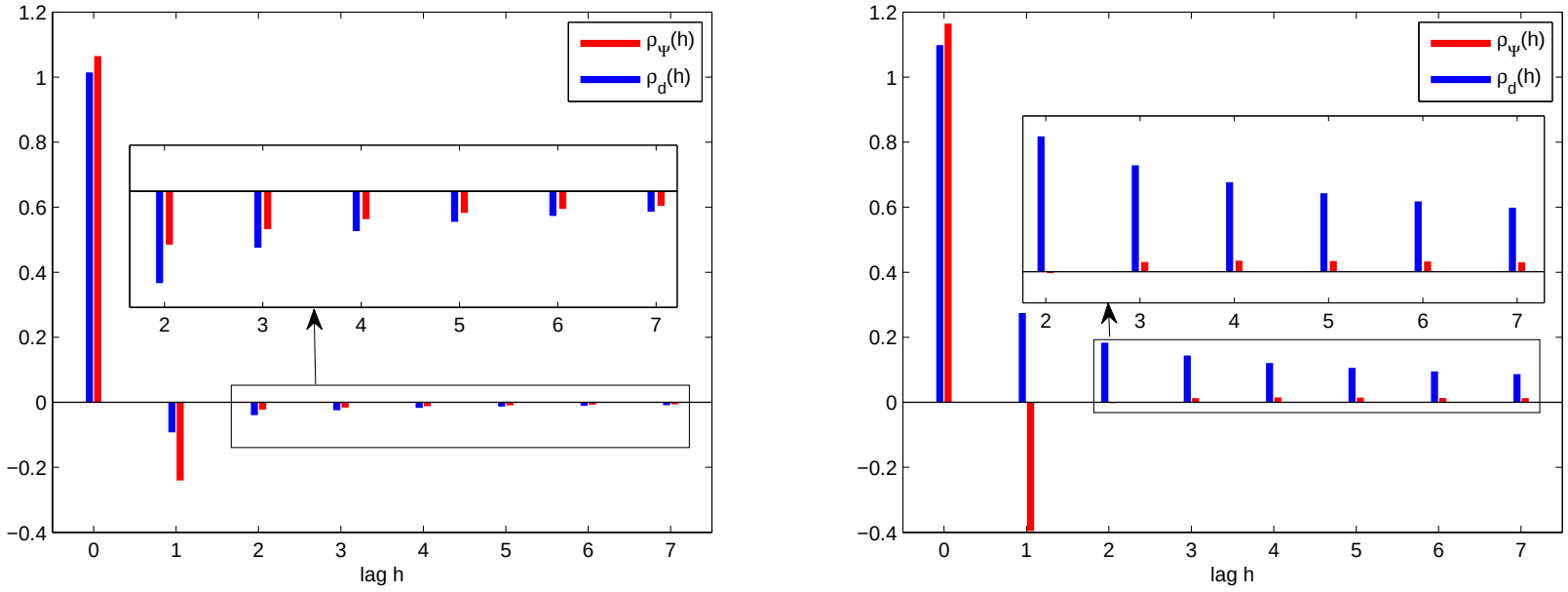


Figure 3 : Covariance function of the FIMA$(d, 1)$ model and the corresponding FI(d) a) with memory parameter $d = -0.1$ and $\Psi = 0.15$ (left), b) with memory parameter $d = 0.2$ and $\Psi = 0.6$ (right).

Estimation

Our goal is to estimate from simulated trajectories of the toy model

$$M(t) = B_H(t) + \epsilon(t), \quad t = 1, 2, \ldots, T,$$

the anomalous diffusion exponent $\alpha = 2H$

$$E\left(M(t)^2\right) = E\left(B_H(t)^2\right) + E\left(\epsilon(t)^2\right) = t^{\overbrace{2H}^{\alpha}} + \sigma^2$$

We apply two different estimators of α :

- time average mean square displacement (MSD)
- estimator of the memory parameter d of the FIMA process

MSD vs. FIMA

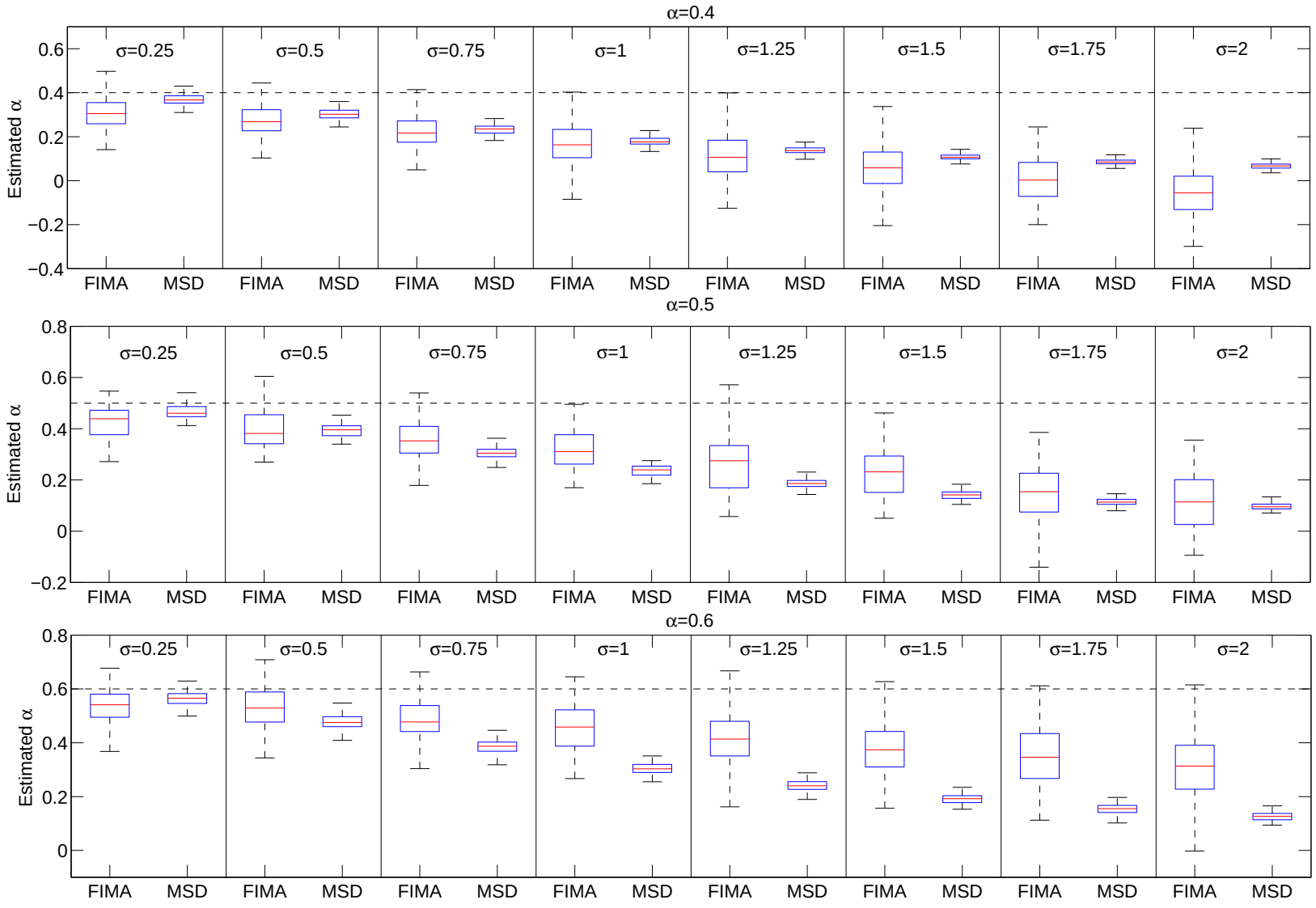
Regimes:

- strong subdiffusion: $\alpha \in \{0.4, 0.5, 0.6\}$
- weak subdiffusion: $\alpha \in \{0.7, 0.8, 0.9\}$
- classical diffusion: $\alpha = 1$
- weak superdiffusion: $\alpha \in \{1.1, 1.2, 1.3\}$
- strong superdiffusion: $\alpha \in \{1.4, 1.5, 1.6\}$

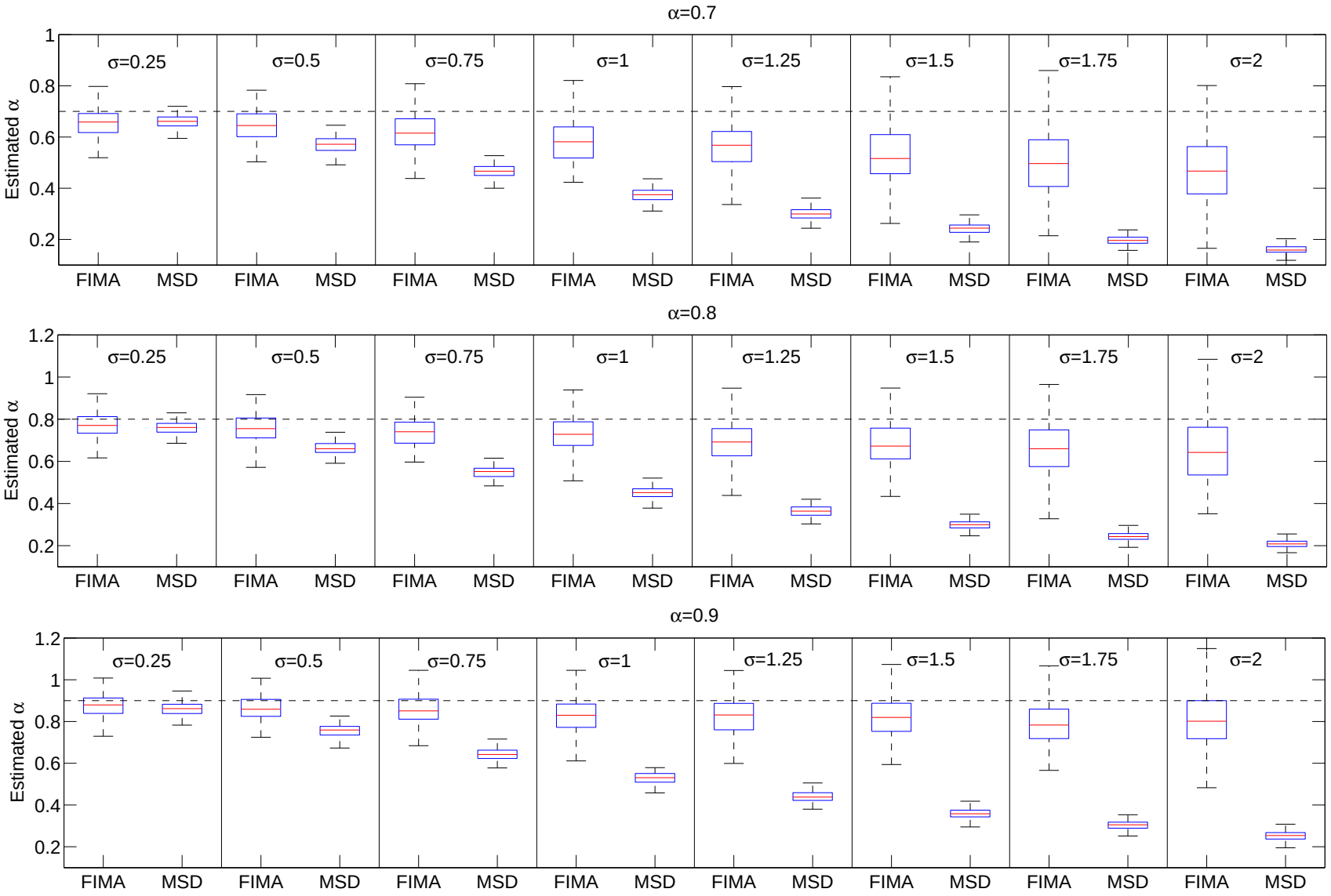
- Measurement error $\sigma \in \{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$ and the length of the trajectory $T = 1024$.

We simulate 1000 trajectories of the toy model for each case, estimate α from MSD and FIMA and compare the two methods.

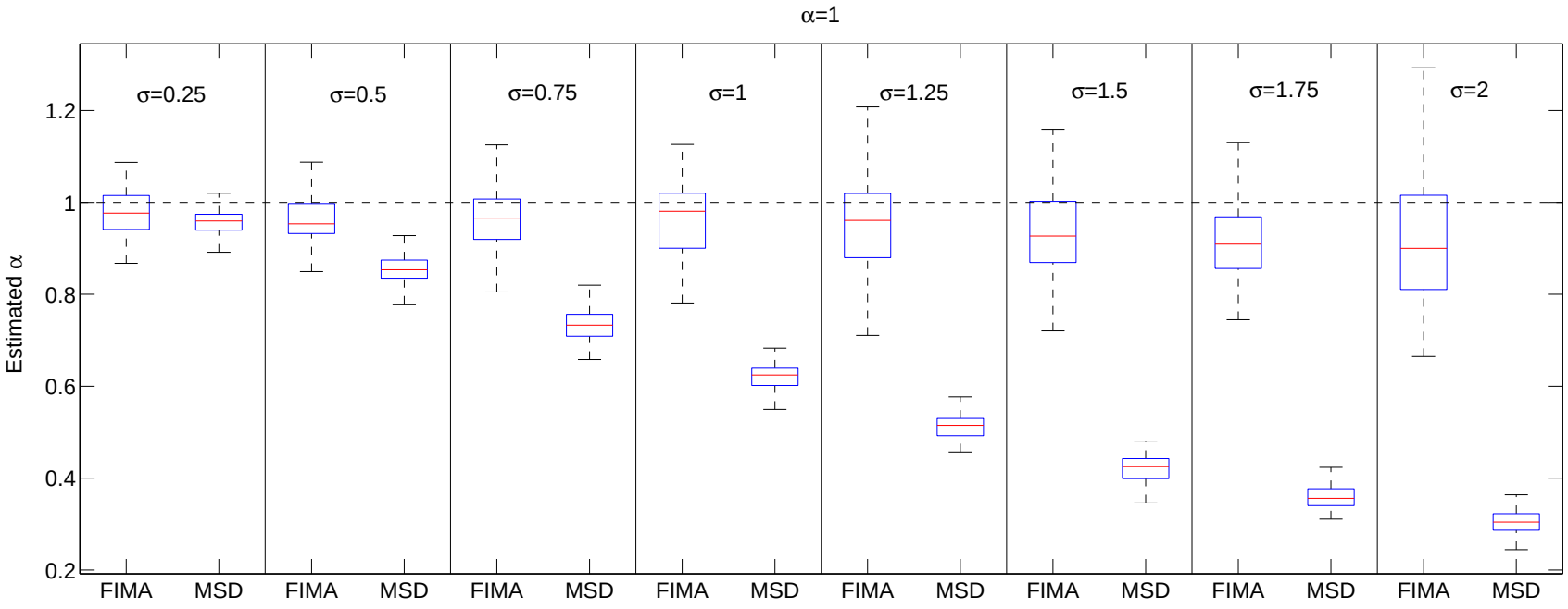
Estimation results: strong subdiffusion



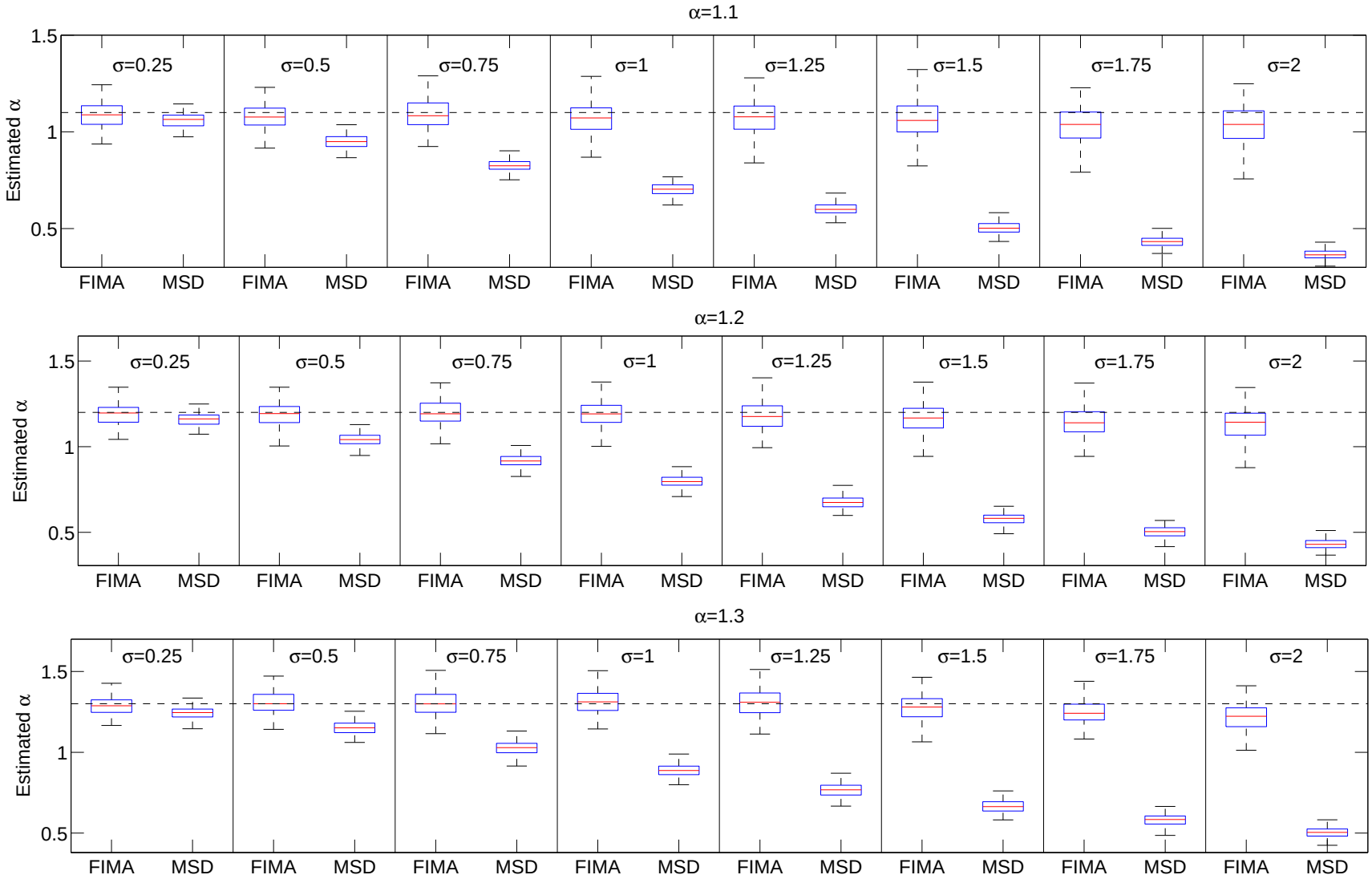
Estimation results: weak subdiffusion



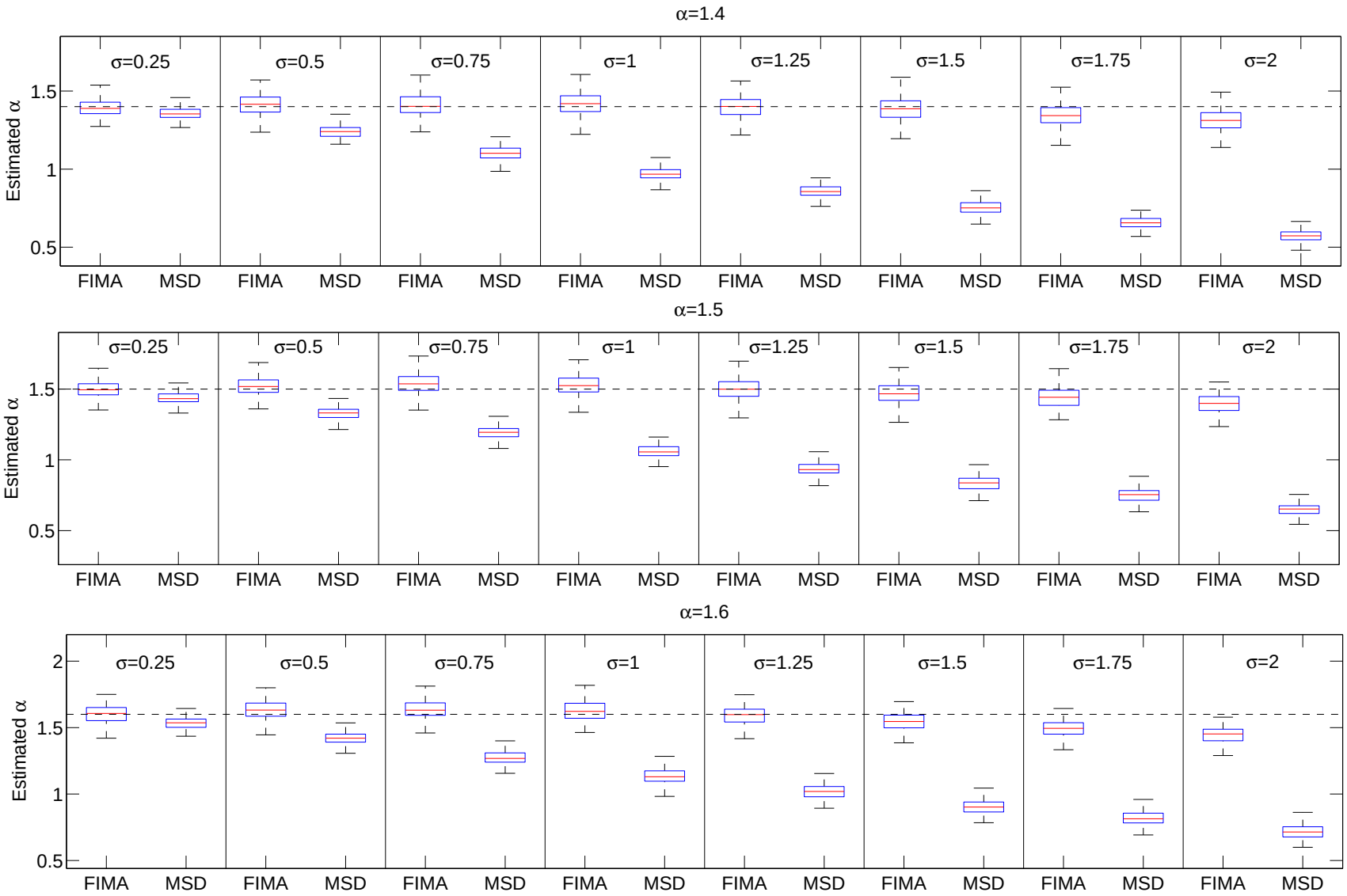
Estimation results: classical diffusion



Estimation results: weak superdiffusion



Estimation results: strong superdiffusion



Conclusions

- FIMA estimator is better than MSD for $\alpha \geq 0.7$.
- For strong and weak subdiffusion both FIMA and MSD estimators are accurate only for small σ values.
- For weak and strong superdiffusion FIMA estimator is very accurate with low variance.

- K. Burnecki, E. Kepten, Y. Garini, G. Sikora, A. Weron, *Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors - An alternative approach*, Scientific Reports, 2015.
- E. Kepten, A. Weron, K. Burnecki, G. Sikora, Y. Garini, *Guidelines for the Fitting of Anomalous Diffusion Mean Square Displacement Graphs from Single Particle Tracking Experiments*, PLOS ONE, 2015.
- K. Burnecki, A. Weron, Algorithms for testing of fractional dynamics: a practical guide to ARFIMA modelling, J. Stat. Mech., 2014.
- K. Burnecki, G. Sikora, *Estimation of FARIMA parameters in the case of negative memory and stable noise*, IEEE Trans. Signal Process. **61**, 2013, 2825–2835.
- K. Burnecki, G. Sikora, A. Weron, *Fractional process as a unified model for subdiffusive dynamics in experimental data*, Phys. Rev. E **86**, 2012, 041912.

Thank you!

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Received: 03 March 2015
Accepted: 14 May 2015
Published: 11 June 2015

Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors - An alternative approach

Krzysztof Burnecki[1], Eldad Kepten[2], Yuval Garini[2], Grzegorz Sikora[1] & Aleksander Weron[1]

Accurately characterizing the anomalous diffusion of a tracer particle has become a central issue in biophysics. However, measurement errors raise difficulty in the characterization of single trajectories, which is usually performed through the time-averaged mean square displacement (TAMSD). In this paper, we study a fractionally integrated moving average (FIMA) process as an appropriate model for anomalous diffusion data with measurement errors. We compare FIMA and traditional TAMSD estimators for the anomalous diffusion exponent. The ability of the FIMA framework to characterize dynamics in a wide range of anomalous exponents and noise levels through the simulation of a toy model (fractional Brownian motion disturbed by Gaussian white noise) is discussed. Comparison to the TAMSD technique, shows that FIMA estimation is superior in many scenarios. This is expected to enable new measurement regimes for single particle tracking (SPT) experiments even in the presence of high measurement errors.