Algorithms for testing of fractional dynamics: a practical guide to ARFIMA modelling

This content has been downloaded from IOPscience. Please scroll down to see the full text.

(http://iopscience.iop.org/1742-5468/2014/10/P10036)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 156.17.79.2
This content was downloaded on 30/10/2014 at 16:56

Please note that terms and conditions apply.
Algorithms for testing of fractional dynamics: a practical guide to ARFIMA modelling

Krzysztof Burnecki and Aleksander Weron

Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wroclaw University of Technology, Wyspianskiego 27, 50-370 Wroclaw, Poland
E-mail: krzysztof.burnecki@pwr.edu.pl and aleksander.weron@pwr.edu.pl

Received 30 June 2014
Accepted for publication 12 September 2014
Published 30 October 2014

Online at stacks.iop.org/JSTAT/2014/P10036
doi:10.1088/1742-5468/2014/10/P10036

Abstract. In this survey paper we present a systematic methodology which demonstrates how to identify the origins of fractional dynamics. We consider three mechanisms which lead to it, namely fractional Brownian motion, fractional Lévy stable motion and an autoregressive fractionally integrated moving average (ARFIMA) process but we concentrate on the ARFIMA modelling. The methodology is based on statistical tools for identification and validation of the fractional dynamics, in particular on an ARFIMA parameter estimator, an ergodicity test, a self-similarity index estimator based on sample p-variation and a memory parameter estimator based on sample mean-squared displacement. A complete list of algorithms needed for this is provided in appendices A–F. Finally, we illustrate the methodology on various empirical data and show that ARFIMA can be considered as a universal model for fractional dynamics. Thus, we provide a practical guide for experimentalists on how to efficiently use ARFIMA modelling for a large class of anomalous diffusion data.

Keywords: stochastic particle dynamics (theory), dynamical processes (theory), fluctuations (theory), stochastic processes (theory)
1. Introduction

A phenomenon observed in recent single-molecule experiments is anomalous diffusion, which largely departs from the classical Brownian diffusion theory since the mean-squared displacement (MSD) is nonlinear. The most popular theoretical models commonly employed are: continuous-time random walk (CTRW) [1], obstructed diffusion (OD) [2,3], fractional Brownian motion (FBM) [4,5], fractional Lévy $\alpha$-stable motion (FLSM) [6], fractional Langevin equation (FLE) [7] and autoregressive fractionally integrated moving average (ARFIMA), see [8] and references therein. For the latter process, we note that the acronyms ‘ARFIMA’ and ‘FARIMA’ are often used interchangeably in literature. Other popular anomalous diffusion models should be mentioned, for example those related to fluorescence recovery after photobleaching (FRAP): a time-dependent diffusion coefficient model [9,10] and scaled Brownian motion [11,12].

References

doi:10.1088/1742-5468/2014/10/P10036
Traditionally, fractional dynamics is related to the concept of fractional dynamic equations. This is an active field of study in physics, mechanics, mathematics and economics, and involves investigating the behaviour of objects and systems that are described by using differentiation of fractional orders. The celebrated fractional Fokker–Planck equation (FFPE) corresponding to the popular CTRW model, describing anomalous diffusion in the presence of an external potential $V(x)$, is given by the following formula

$$\frac{\partial w(x,t)}{\partial t} = \alpha D_t^{1-\alpha} \left[ \frac{\partial}{\partial x} \frac{V'(x)}{\eta} + K \frac{\partial^2}{\partial x^2} \right] w(x,t). \tag{1}$$

It was derived explicitly in [13,14], where methods of its solution were introduced and for some special cases exact solutions were calculated. Here, the operator $\alpha D_t^{1-\alpha}$, $\alpha \in (0,1)$, is the fractional derivative of the Riemann–Liouville type [15]. It is known that $\alpha D_t^{1-\alpha}$ introduces a convolution integral with a slowly decaying power-law kernel, which is typical for memory effects in complex systems [16]. In (1), $w(x,t)$ denotes the probability density function (PDF) and the prime (') stands for the derivative with respect to the space coordinates relating the force $F(x)$ and the potential through $F(x) = -V'(x)$. The constant $K$ denotes the anomalous diffusion coefficient, whereas $\eta$ is the generalized friction constant. For $\alpha \to 1$, (1) becomes the ordinary Fokker–Planck equation. The FFPE describes subdiffusion in accordance with the mean-squared displacement in the force-free limit and it obeys some generalized fluctuation-dissipation theorem. Moreover, a generalization of the Einstein–Stokes–Smoluchowski relation $K = k_B T/\eta$ connects the generalized friction and diffusion coefficients [1].

Derivatives and integrals of fractional orders are used to describe objects that can be characterized by long (power-like) memory or self-similarity. In recent years, there has been much interest in long-range dependent and self-similar processes, in particular FBM, FLSM and ARFIMA. As a candidate suitable for extensive statistical analysis of the fractional dynamics we choose the ARFIMA model [17–19]. It is a discrete time analogue of the overdamped fractional Langevin equation [20] that allows for the non-Gaussian law (Lévy $\alpha$-stable) and a long memory.

The objective of this work is to convince the readers that the ARFIMA process can serve as a universal and simple discrete time model for fractional dynamics of empirical data and the celebrated FBM and FLSM form the limiting case of ARFIMA. Moreover, since the ARFIMA models were successful in analysing data in other fields (econometrics, see 2003 Nobel Prize in Economic Sciences for C W J Granger and R Engel; finance and engineering [26–28]), many statistical tools (and computer packages, e.g. ITMS [18]) are available and are presented in appendix D. We also provide a basic checklist to advise the inexperienced user as to what statistical techniques must be applied in various stages of the analysis of the empirical data. In table 1 we present a list of the different physical environments along with relevant mathematical models.

The paper is structured as follows. In section 2 three models for fractional dynamical systems are considered, namely FBM, FLSM and an ARFIMA time series. The former two models are self-similar and their increments form long-range dependent processes. The latter discrete-time ARFIMA process is stationary, and when aggregated, in the limit, it converges to either FBM or FLSM. In this sense it generalises both models. In contrast to them, it allows for different light- and heavy-tailed distributions, and, as a stationary
Table 1. Physical environment versus mathematical models, see [10, 21–24] (RWRW stands for the random walk on random walk [25]).

<table>
<thead>
<tr>
<th>Physical environment</th>
<th>Mathematical models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trapping, crowded environment</td>
<td>CTRW, FFPE, subordinated BM</td>
</tr>
<tr>
<td>Labyrinthine environment</td>
<td>OD, percolation, RWRW</td>
</tr>
<tr>
<td>Viscoelastic system</td>
<td>FBM, FLSM, FLE, ARFIMA</td>
</tr>
<tr>
<td>System with time-dependent diffusion</td>
<td>scaled BM, scaled FBM, ARFIMA with (d(t))</td>
</tr>
<tr>
<td>System with transient diffusion</td>
<td>BM with transient subordinators</td>
</tr>
</tbody>
</table>

process, provides tools for calculating predictions. In section 3 we discuss three data sets related to the biological, telecommunication and astrophysical problems. They are successfully resolved by means of the universal ARFIMA time series model with various noises. Even if the finer details of the estimation procedures are case specific, we hope that the suggested checklist will still be of great use as a practical guide.

In appendices A–F we present useful fractional dynamics identification and validation methods. We present an estimation algorithm for the ARFIMA parameters in the case of noise belonging to the domain of attraction of the stable law for both positive and negative memory parameters. We also present methods for the testing of stationarity and ergodicity, distribution type, self-similarity and long memory. In particular we present two efficient methods for estimation of the self-similarity index and memory parameter with the help of \(p\)-variation and sample mean-squared displacement respectively.

2. Basic models for fractional dynamics

In this section we discuss three models for fractional dynamics: fractional Brownian motion (section 2.1), fractional Lévy stable motion (section 2.2), and ARFIMA (section 2.3). The former two models are self-similar and their increments form long-range dependent processes. The discrete-time ARFIMA process is stationary and generalizes both models, since aggregated in the limit, it converges to either fractional Brownian or Lévy stable motion. In contrast to the previous two models, it allows for different light- and heavy-tailed distributions, and both long (power-like) and short (exponential) dependencies [8]. Moreover, as a stationary process, it provides prediction tools.

2.1. Fractional Brownian motion

Fractional Brownian motion introduced by Kolmogorov in 1940 [29,30] is a generalization of the classical Brownian motion (BM). Most of its statistical properties are characterized by the self-similarity (Hurst) exponent \(0 < H < 1\). For any \(0 < H < 1\), fractional Brownian motion (FBM) of index \(H\) (Hurst exponent) is the mean-zero Gaussian process \(B_H(t)\) with the following integral representation [30]:

\[
B_H(t) = \int_{-\infty}^{\infty} \left\{ (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}} \right\} dB(u),
\]

(2)

where \(B(t)\) is a standard Brownian motion and \((x)_+ = \max(x,0)\).
FBM is $H$-self-similar, namely for every $c > 0$ we have $B_H(ct) \overset{D}= c^H B_H(t)$ in the sense of all finite dimensional distributions, and has stationary increments. It is the only Gaussian process satisfying these properties. With probability 1, the graph of $B_H(t)$ has both a Hausdorff dimension and a box dimension of $2 - H$.

For $H > 1/2$, the increments of the process are positively correlated and exhibit positive long-range dependence, whereas for $H < 1/2$, the increments of the process are negatively correlated and exhibit negative long-range dependence (see appendix C). For the second moment of the FBM we have $\langle B_H^2(t) \rangle = \sigma^2 t^{2H}$, where $\sigma > 0$.

The precise simulation of such a process is of great interest. The most commonly used approaches can be split into two categories. The first one, related to theoretically exact methods, has so far been composed only of a matrix factorization technique based on the Cholesky decomposition of the covariance matrix. Unfortunately, this technique has a complexity of $O(N^2)$ and requires high computational resources even for moderate trajectory lengths. The other category is composed of non-exact techniques. All of the above methods have their own particular drawbacks and advantages and the choice between them boils down to a tradeoff between speed and accuracy. However, Davies and Harte [31] proposed both a fast and exact synthesis method for stationary Gaussian processes, called the circulant embedding method. It can be easily applied for a FBM [32]. Since it is based on the fast Fourier transform (FFT) algorithm, its complexity is only $O(N \log N)$.

2.2. Fractional Lévy stable motion

FBM can be generalized to a fractional Lévy stable motion (FLSM) [33,34]. FLSM is the $\alpha$-stable process $L_H^\alpha(t)$ with the following integral representation:

$$L_H^\alpha(t) = \int_{-\infty}^{\infty} \left\{(t-u)^d_+ - (-u)^d_+\right\} dL_\alpha(u),$$

where $L_\alpha(t)$ is an $\alpha$-stable motion (SM), $0 < \alpha \leq 2$, $0 < H < 1$, and $d = H - 1/\alpha$.

The process for $\alpha = 2$ becomes a FBM, is $H$-self-similar and has stationary increments. Analogous to the FBM case, we say the increments of the process exhibit positive long-range dependence if $d > 0$ ($H > 1/\alpha$), and negative dependence when $d < 0$ ($H < 1/\alpha$) [35]. This is due to the behaviour of the integrand (kernel function) in (3).

When $d < 0$ the kernel function has singularities at $u = 0$ and $u = t$. These singularities magnify the stable noise process $dL_\alpha(t)$, causing large spikes in the paths of the FLSM. Their dependence structure resembles that of a negatively correlated process and thus we shall refer to this case as the negative dependence scenario. In the case when $d > 0$ the kernel is bounded and positive, for all $t > 0$. Thus the jumps in the paths of $L_H^\alpha(t)$ due to the fluctuations of the noise process are not as magnified as in the case $d < 0$. In this case especially for large values of $H$, the kernel function decays slowly as $u \rightarrow -\infty$. This implies that the past fluctuations in the process $dL_\alpha(t)$ significantly influence the present values of the process $L_H^\alpha(t)$. This case is referred to as the positive long-range dependence scenario. Therefore, as in the Gaussian case, the parameter $d$ controls the sign of dependence.

Simulation of such a process is even more difficult than in the FBM case. This is due to the fact that stable processes cannot be characterized by 2D distributions, which is the case in the Gaussian case. Namely, any Gaussian random vector is fully defined
Algorithms for testing of fractional dynamics by the mean and covariance matrix, whereas stable vectors are defined in terms of so-called spectral measure [33], which is a much more complicated object. Hence, in the literature there are no exact methods for the simulation of FLSM. The most commonly used approaches apply integral representation of FLSM. Stoev and Taqqu [35] proposed a fast method based on the FFT algorithm with complexity only $O(N \log N)$. It can also be applied to ARFIMA processes and more general moving average stable processes.

2.3. ARFIMA process

Autoregressive fractionally integrated moving average (ARFIMA) time series were introduced by Granger and Joyeux [17] and Hosking [36]. The ARFIMA process $\{X_t\}$, denoted by ARFIMA$(p, d, q)$, is defined by:

$$\Phi_p(B)X_t = \Psi_q(B)(1 - B)^{-d}Z_t,$$ (4)

where innovations $Z_t$ are i.i.d. random variables with either finite or infinite variance. $\Phi_p(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$ is the autoregressive polynomial and $\Psi_q(z) = 1 + \psi_1 z + \psi_2 z^2 + \ldots + \psi_q z^q$ is the moving average polynomial. The operator $(1 - B)^{-d}$ is the integrating operator and has the infinite binomial expansion

$$(1 - B)^{-d}Z_t = \sum_{j=0}^{\infty} b_j(d)Z_{t-j},$$ (5)

where the $b_j(d)$’s are the coefficients in the expansion of the function $f(z) = (1 - z)^{-d}$, $|z| < 1$, i.e.

$$b_j(d) = \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}, \quad j = 0, 1, \ldots,$$ (6)

where $\Gamma$ is the gamma function.

The sequence $\{Z_t\}$ is often called the noise process (sequence) [18]. We assume that innovations $Z_t$ are i.i.d. and belong to the domain of attraction of an $\alpha$-stable law with $0 < \alpha \leq 2$. The series given by (5) is convergent almost surely (a.s.) and the ARFIMA definition (4) is correct if and only if

$$\alpha(d - 1) < -1 \iff d < 1 - \frac{1}{\alpha}.$$ (7)

In particular, in the Gaussian case ($\alpha = 2$) we have $d < 1/2$. Moreover, if all roots of the polynomial $\Phi_p$ lie outside the unit circle, the ARFIMA$(p, d, q)$ time series defined by (4) is stationary and has a causal moving average form

$$X_t = \sum_{j=0}^{\infty} c_j(d)Z_{t-j},$$ (8)

for details see [36,37]. Stationary ARFIMA processes are also ergodic [35]. In this paper we concentrate on such a case.

Such processes are asymptotically $H$-self-similar with the parameter $H = d + 1/\alpha$. Precisely, we have $N^{-H} \sum_{k=1}^{[N]} X_k \to L^H(t)$ for $t \geq 0$ and the convergence is with respect to finite-dimensional distributions [35]. Hence the ARFIMA process can be considered as a discrete-time generalization of FLSM (in particular of FBM).

doi:10.1088/1742-5468/2014/10/P10036
In the Gaussian case, the rate of decay of the autocovariance function of the ARFIMA model is
\[ r(k) = \langle X(0)X(k) \rangle \sim k^{2d-1}, \quad \text{as } k \to \infty. \] (9)
Therefore, for \( d > 0 \) we have \( \sum_{k=0}^{\infty} |r(k)| = \infty \). This serves as a classical definition of long memory and is equivalent to the case of the increments of FBM with \( H > 1/2 \). Similarly, for \( d < 0 \) we arrive at the negative short (but power-like) dependence, which corresponds to the increments of FBM with \( H < 1/2 \). The case \( d = 0 \) leads to the autoregressive moving average (ARMA) model, which has an exponentially decaying autocorrelation function \([18,38]\).

Finally, the linear predictor for the ARFIMA process based on the finite past \( X_n, \ldots, X_0 \) takes the form:
\[ \hat{X}_{n+h} = \sum_{j=0}^{n} a_j X_{n-j}, \] (10)
where the sequence \( \{a_0, a_1, \ldots, a_n\} \) is given by \( a_j = -\sum_{t=0}^{k-1} c_t h_{j+t} \), the \( c_j \)'s are defined by (8) and \( h_j \)'s are given by
\[ \frac{\Phi_p(z)(1-z)^d}{\Theta_q(z)} = \sum_{j=0}^{\infty} h_j z^j, \quad |z| < 1, \] (11)
see \([39]\) for the discussion of the prediction problem in the infinite variance ARFIMA case.

The simulation of ARFIMA processes is by no means straightforward. For such processes, except for the special case: ARMA\((p, q)\) model, there are no exact methods. The existing algorithms mostly concentrate on efficient ways to calculate the infinite sum given in (8). Stoëv and Taqqu \([35]\) proposed a fast method based on the FFT algorithm with complexity only \( O(N \log N) \). It can also be applied to FLSM and more general moving average stable processes.

The ARFIMA\((0, d, q)\) is called FIMA\((d, q)\) and is represented by the following equation
\[ (1 - B)^d X(t) = Z(t) - \psi_1 Z(t-1) - \ldots - \psi_q Z(t-q), \] (12)
where \( t = 0, \pm 1, \ldots, B \) is the shift operator: \( BX(t) = X(t-1) \), \(-1/2 < d < 1/2\), which takes fractional values, either positive or negative, and \( \{Z(t)\} \) is a white noise sequence \([18]\).

In many applications the ARFIMA\((1, d, 1)\) model is sufficient to describe the data well, see, e.g. \([8,40]\). The basic building blocks of the ARFIMA\((1, d, 1)\) model are the AR\((1)\) process: \( X(t) = \phi_1 X(t-1) + Z(t) \), and the MA\((1)\) process: \( X(t) = Z(t) - \psi_1 Z(t-1) \). AR\((1)\) is a causal or future-independent function of noise and stands for the regression. The explanatory variable is the observation immediately prior to our current observation. In order to get an idea about the role of the MA part let us concentrate on the case when \( Z(t) \) is a white noise. It appears that if \( X(t) \) is a stationary 1-correlated time series, i.e. \( X(s) \) and \( X(t) \) are independent whenever \( |t - s| > 1 \) (in contrast to an i.i.d. sequence which is zero-dependent), then it can be represented as the MA\((1)\) process \([18]\). The dependence (correlation in the white noise case) is only one lag long and its intensity is fully controlled by the parameter \( \psi_1 \). Hence, ARMA models introduce the short memory of the process. In general, the MA\((q)\) process may reconstruct an arbitrary short time (finite lag) correlation structure from the experimental data. The fractional integration introduces the long (power-law) memory, which is defined by the memory parameter \( d \).
Its special case, FIMA\((d, 1)\) is given by the following formula:

\[
\frac{1}{(1 - B)^d}X(t) = Z(t) - \psi Z(t - 1),
\]

where \(t = 0, \pm 1, \ldots\).

Brownian motion (BM) corresponds, in the limit sense [35], to FIMA\((0, 0)\). Similarly, FBM corresponds to FIMA\((d, 0)\) with \(d = H - 1/2\), where \(H\) is the self-similarity parameter. Hence, it is possible to model and characterize more complex processes than using FBM alone. The FIMA processes offer flexibility in modelling long power-law and one-lag dependencies by choosing the memory parameter \(d\) and the appropriate moving average coefficient \(\psi\) in (13).

3. Modelling of fractional dynamics

In this section we illustrate the applicability of ARFIMA processes in three different kinds of real data. They are related to viscoelastic systems (molecular biology), telecommunication (electromagnetic field measurements) and astrophysical phenomena. We will study the data by applying the following checklist for testing for fractional dynamics.

**Checklist. Testing for fractional dynamics**

- **L1** Stationarity and ergodicity (prerequisites for ARFIMA modelling) (see appendices A and B)
- **L2** Sample MSD (to check for the presence of the power-law dependence and anomalous diffusion) (see appendix C)
- **L3** Fitting the ARFIMA parameters (usually it is enough to fit at most ARFIMA\((1, d, 1)\), so three parameters are to be estimated) (see appendix D).
- **L4** Fitting the distribution to the noise (Gaussian, non-Gaussian stable, other) (see appendix E)
- **L5** (optional) \(H\)-self-similarity for the partial sum process (along with the concept of surrogate data, it provides a double check for both the memory parameter \(d\) and stability index \(\alpha\)) (see appendix F)

3.1. Modelling of single particle tracking dynamics in molecular biology

Anomalous diffusion in crowded fluids, e.g. in the cytoplasm of living cells, is a frequently occurring phenomenon [21]. Crowded fluids form a wide class of soft condensed matter systems. Prominent examples provide dense polymer solutions and the cytoplasm of living cells [41].

The motion of the tracer particles may switch from normal to anomalous diffusion, i.e. the tracer’s mean square displacement (MSD) exhibits a sublinear growth with time MSD\((t) \sim t^a\), where \(a < 1\) for pure anomalous subdiffusion. In the transient anomalous subdiffusion, one mechanism operates at short times and another at long times, depending
Algorithms for testing of fractional dynamics on a crossover time. Despite the frequent observation of such a behaviour in crowded fluids, the origin of the anomaly has remained controversial [22].

Why does anomalous subdiffusion matter? According to [22], first, it affects reaction kinetics through the time dependent diffusion coefficient. Secondly, it is a probe of submicroscopic organization, though unfortunately far from uniquely invertible. Moreover, ubiquitous observation in cell biology shows that the diffusive motion of macromolecules and organelles is anomalous, see [21].

The progress made in understanding the molecular nature of intracellular processes in recent years is astonishing, and is due to tight collaboration between physicists, chemists, biologists and mathematicians, all working in the same direction and using specific methods of their sciences [10, 42].

As a universal candidate suitable for extensive statistical analysis of the subdiffusive dynamics in biological cells, we propose the ARFIMA model which was described in section 2.3. We apply the introduced methodology to the single mRNA molecule time series from the Golding and Cox microscopy video [40, 43]. The analyzed video has 1801 frames of size 59 × 76 pixels, which present the location of the mRNA molecule at the time $t_k = k \times s$, for $k = 1, \ldots, 1801$. The centre of the pixel in the upper-left and bottom-right corner of each frame has the coordinates (1,1) and (59,76), respectively. We identified the position of a molecule at the time $t_k$ with the position of its mass centre $(\bar{x}(t_k), \bar{y}(t_k))$, which is calculated based on image segmentation results obtained by applying the approach introduced in [44].

Consequently, we obtained a 2D data set of $(\bar{x}(t_k), \bar{y}(t_k))$, for $k = 1, \ldots, 1801$. These data form the time series of a single mRNA molecule in the analyzed microscopy video [40]. We analyze a 2D trajectory $\{(X_n, Y_n) = (\bar{x}(t_n), \bar{y}(t_n)) : n = 1, \ldots, 1801\}$ and concentrate on the increments of $x$ and $y$ coordinates. All analysis is based on two sets of increments $\{X_{n+1} - X_n : n = 1, \ldots, 1800\}$ and $\{Y_{n+1} - Y_n : n = 1, \ldots, 1800\}$ of coordinates $x$ and $y$, respectively.

**Step L1.** The trajectories and their increments of both coordinates of the data are presented in figure 1. A simple examination of the presented increments tells us that the data are not stationary in general. They demonstrate two different regimes. Hence, by using the variance change point test [45] we split the data for two stationary subsets. The changing points are the 702nd and 715th observation for the $x$ and $y$ coordinates, respectively. In total, we obtain four subsets which we denote by $X_1$, $X_2$, $Y_1$ and $Y_2$. We checked the stationarity of the subsets by means of the autocovariance function (see appendix A). They reveal a stationary pattern, which is illustrated in figure 2 for the $X_1$ part.

Next, we applied the ergodicity test (see appendix B) to all four subsets of the data to check the necessary condition for ergodicity (B.10), which requires only one appropriately long trajectory. In figure 3 we see the result of the test for the $X_1$ and $X_2$ subsets. The necessary condition for mixing (B.9) is clearly satisfied. This implies that ergodicity cannot be rejected. We also checked that all subsets satisfy the necessary conditions for mixing (B.9).

**Step L2.** In order to check the power-law dependence and subdiffusive behaviour of the data we calculated the sample MSD, see appendix C. First, we calculated sample MSD for all four subsets of the data: $X_1$, $X_2$, $Y_1$ and $Y_2$ and obtained the following $d$ estimates:

\[\text{doi:10.1088/1742-5468/2014/10/P10036}\]
Figure 1. A plot of the trajectories (top panels) and their increments (bottom panels) with a change point in variance (red dashed line), for $x$ (left panels) and $y$ coordinates (right panels). The variance changing points are the 702nd and 715th observations for the $x$ and $y$ coordinates, respectively. Source: [40].

Figure 2. Autocorrelation (left panel) and partial autocorrelation (right panel) functions corresponding to the subset $X_1$ of the data. Both functions tend to zero indicating stationarity of the sample. Moreover, all values of the functions lie between 95% confidence bounds calculated for the white noise case which suggests the ARFIMA$(0, d, 0)$ model.

\[ -0.1, -0.19, -0.28, -0.22, \text{ respectively.} \] Next, we did it for the 2D data: $(X_1, Y_1)$, $(X_2, Y_2)$ and $(X = X_1 \cup X_2, Y = Y_1 \cup Y_2)$, where the distance was calculated according to the Euclidean norm. The results were: $-0.15, -0.20, -0.16$, respectively. We can see that in all cases $d$ is essentially negative (the negative power-law dependence case), hence the process clearly follows the subdiffusive dynamics.

Step L3. We fitted the ARFIMA$(1, d, 1)$ model (see appendix D) to the four time series [40]. The best-fitted ARFIMA models were: ARFIMA$(0, -0.13, 0)$, ARFIMA$(0, -0.14, 1)$ ($\psi = 0.13$), ARFIMA$(0, -0.38, 0)$, and ARFIMA$(1, -0.07, 1)$ ($\phi = 0.48$, $\psi = 0.67$). The
Figure 3. Real (left panels) and imaginary (right panels) parts of the function \( \hat{E}(n) \) corresponding to the subsets X1 (top panels) and X2 (bottom panels) of the data. The necessary condition for mixing (B.9) is clearly satisfied. This implies that ergodicity cannot be rejected.

orders of the model were chosen looking at the autocorrelation and partial autocorrelation functions (see appendix A and checking the independence of the noise. The noise was calculated by means of an ITSM package [18]. The independence for the above models could not be rejected for the Ljung–Box, turning points, difference-sign and rank tests (for information about the tests, see, e.g. [18,46,47]).

Step L4. Next, we fitted different distributions to the noise, namely Gaussian, Lévy stable and normal-inverse Gaussian (NIG). A random variable \( X \) is said to have a NIG distribution if it has density

\[
f(x) = \frac{\alpha \delta}{\pi} \frac{\delta \sqrt{\alpha^2 - \beta^2 + \beta(x-\mu)}}{\delta^2 + (x-\mu)^2} \text{K}_1(\alpha \sqrt{\delta^2 + (x-\mu)^2}), \quad -\infty < x < \infty. \tag{14}
\]

The NIG distribution, introduced in [48], is described by four parameters \((\alpha, \beta, \delta, \mu)\), where \( \alpha \) stands for tail heaviness, \( \beta \) for asymmetry, \( \delta \) is the scale parameter, and \( \mu \) is the location. The normalizing constant \( \text{K}_\lambda(t) \) in (14) is the modified Bessel function of the third kind with index \( \lambda \), also known as the MacDonald function.

The NIG law is the only one that cannot be rejected for any part of the data, see [40]. To confirm the excellent fit of the model, we also calculated sample MSD for 1000 simulated trajectories of ARFIMA models with the fitted parameters given and compared the results with the MSD values obtained for the original data. We can see in figure 4 that the fitted ARFIMA processes reproduce the sample MSD well.
Figure 4. Boxplots of sample MSD calculated for 10,000 simulated trajectories of the ARFIMA processes fitted to the four subsets of the analyzed data. We can observe that the ARFIMA model leads to MSD values which are close to those from the empirical data.

The proposed ARFIMA model is universal. Nevertheless, it turns out that the four stationary subsets of the 2D trajectory have different parameters ($\phi_1, d, \psi_1$). This is due to the fact that our single particle tracking analysis is not a mean-value approach as was done in [43]. Observe that the shape of the cells and crowded fluid characteristic of the cytoplasm influence the dynamics of the labelled mRNA molecules. In particular, we believe that the parameters of the fitted ARFIMA models can provide some insight into the physical reasons for the subdiffusive motion of the molecule. Namely, the parameters $d$ in both directions $x$ and $y$ are influenced by the shape of the cell. Simulations show that as the width of the biological cell gets smaller, the memory parameter becomes ‘more negative’. Two additional parameters $\phi_1$ and $\psi_1$ which are responsible for short-time effects are influenced by short-distance interactions in a crowded fluid environment in the cytoplasm [40].

Step L5. Finally, we studied the surrogate data of the four subsets. From the results for the surrogate data (see appendix E), the corresponding estimates for the parameter $1/\alpha$ were: $1/\alpha_{p\text{Var}} = H_{p\text{Var}} = 0.53, 0.5, 0.47, 0.5$, respectively, where $H_{p\text{Var}}$ is the self-similarity estimator obtained via the $p$-variation method (see appendix F). We observe that the values confirm that the distributions of the noises belong to the domain of attraction of the Gaussian ($\alpha = 2$) law.

3.2. Modelling of electromagnetic field

Here we analyze a set of Universal Mobile Telecommunications System (UMTS) data which is plotted in figure 5. The electromagnetic field intensity was measured in Wroclaw in an urban area every minute from 12.01.2011 22:40 to 19.01.2011 21:18 (9,999 observations) [49].

Step L1. We can clearly see that the data possess a seasonal trend with a daily period which is related to a daily pattern of the telephone calls. After removing the daily seasonality (by fitting and subtracting a sine function) the autocorrelation function still displayed a non-stationary behaviour (it did not converge to zero). Next, we studied the...
volatility of the data \( \{ \hat{X}_n : n = 1, 2, \ldots, 9999 \} \). A fluctuating variance was observed in a squared data plot, especially when the data had a mean close to 0. In order to remove this effect and make the volatility constant we, first, calculated the sample moving variance \( S^2_k := \frac{1}{60} \sum_{i=k}^{k+60} (\hat{X}_k - \bar{X})^2 \), for \( k = 1, 2, \ldots, 9939 \), and, next, computed the series \( \tilde{X}_n = \hat{X}_{n+30}/\sqrt{S^2_n} : n = 1, 2, \ldots, 9939 \). The autocorrelation function for the modified data revealed a stationary pattern. We also applied the ergodicity test (see appendix B) to the data to check the necessary condition for mixing (B.9). The necessary condition was clearly satisfied, hence ergodicity cannot be rejected.

**Step L2.** Next, we studied the memory structure of the series. To this end we applied the sample MSD method described in appendix C of estimating the memory parameter \( d \). We obtained that \( d_{\text{MSD}} = 0.31 \). The estimator is clearly positive, which indicates the long memory property. This could also be observed from the autocorrelation function, which decayed slowly to zero.

**Step L3.** We fitted the ARFIMA model to the data \( \{ \tilde{X}_n : n = 1, 2, \ldots, 9939 \} \). For the model orders \( p \leq 1 \) and \( q \leq 1 \), we applied the ARFIMA estimator defined in appendix D. Next, we calculated the corresponding noise. It appeared that only in the case of \( p = 1 \) and \( q = 0 \) the noise could be treated as white noise (the sequence was tested for randomness with the Ljung-Box test [46,50]). Hence this suggests that the ARFIMA(1, 0.34, 0) model describes the data well.

**Step L4.** The distribution underlying the noise could not be identified. Such an ARFIMA process can be simulated by generating the noise from its empirical distribution function.

### 3.3. Modelling of high solar flare activity

Observations of solar flare phenomena in x-rays became possible in the 1960s with the availability of space-borne instrumentation. Since 1974 broad-band soft x-ray emissions from the Sun have been measured almost continuously by the meteorology satellites operated by the National Oceanic and Atmospheric Administration (NOAA) as well as the Synchronous Meteorological Satellite (SMS) and the Geostationary Operational Environment Satellite (GOES).
From 1974 to 1986 the soft x-ray records were obtained by at least one from the series of GOES-type satellites. Since 1983, data from two and even three co-operating GOES are generally available. The x-ray sensor, part of the space environment monitor system aboard the GOES, consists of two ion chamber detectors which provide whole-sun x-ray fluxes in the 0.05–0.3 nm and 0.1–0.8 nm wavelength bands. The solar soft x-ray flares data is widely available from the NOAA Space Environment Center website (http://goes.ngdc.noaa.gov/data/avg/).

At present, accurate solar data is available for the four most recent cycles only. Individual solar cycles are different in form, amplitude and length. Hence understanding the long-term solar variability and predicting the solar activity is a real problem for solar physics. It is associated with a variety of space weather effects [51]. Solar activity is known to correlate with flare activity, and a variety of flare properties (time and strength) could be incorporated into predictions. Moreover, these disturbances pose serious threats to man-made spacecrafts, disrupt electronic communication channels and can even create huge electrical currents in power grids [52].

Our aim is to present a proper statistical model for analyzing and predicting soft x-ray solar emission activity in the period of solar maximum. Here, we study a high activity period, when energy was transmitted by x-rays emitted during blasts on a solar surface from January 1st 2000 to December 31st 2002. For details, see [53].

Step L1. The plot of the data showed no non-stationary effects. We also plotted autocorrelation and partial autocorrelation functions. They decayed slowly to zero starting from positive values suggesting the power-law positive correlation structure. The ergodicity of the data was checked by applying the ergodicity test (see appendix B). The necessary condition for mixing (B.9) was satisfied, hence the ergodicity cannot be rejected.

Step L2. Next, we studied the memory structure of the series by applying the sample MSD method described in appendix C. The estimated memory parameter is $d_{MSD} = 0.29$ indicating the long memory property, which coincided with the output of the autocorrelation function.

Step L3. The parameters of the ARFIMA process were fitted to minimize the prediction error, see appendix D. We concluded that the proper model is ARFIMA(2, $d$, 0) with the parameters $\phi_1 = 0.03$, $\phi_2 = 0.03$, and $d = 0.19$.

Step L4. We fitted different distributions to the noise. The one not rejected by any of the tests was the Pareto law with $\alpha = 1.25$. The Pareto distribution is a proper choice for the present analysis because it gives a description of positive random variables whereas the stable one for $1 < \alpha < 2$ is related to both positive and negative random variables, but a time series of x-ray flare energy is quite positive by definition.

As the power-law distributions belong to the domain of attraction of stable law (see, e.g. [33]), the resulting distribution of the ARFIMA process should be close to the stable one. The value of $\alpha$ for the data was estimated to be 1.21. One may check that the estimated value of $\alpha$ for simulated ARFIMA times series with Pareto variables is usually underestimated. This agrees with our situation.

Step L5. We studied the surrogate data of the empirical time series. From the results for the surrogate data (see appendices E and F), the corresponding estimate for the parameter
Algorithms for testing of fractional dynamics

Figure 6. Solar flare data (blue line with asterisks) and 1 d-ahead prediction (red thick solid line) calculated under the ARFIMA(2, 0.19, 0) model. Source: [53].

1/α is: 1/α_{pVar} = H_{pVar} = 0.85, where $H_{pVar}$ is the self-similarity estimator obtained via the $p$-variation method (see appendix F). We observe that the estimators are close to the one assumed in our model: $1/\alpha = 0.8$. Moreover, $d = 0.19$ is the highest admissible value of $d < 1/\alpha$ for the ARFIMA model.

Finally, we calculated the 1 day-ahead prediction (see section 2.3) for the ARFIMA time series and setting $h = 1$. The results are depicted in figure 6. The error of the prediction is not low but we have to take into account there is no strong deterministic component in the data, only a stochastic component with strongly persistent noise.

Acknowledgments

The authors would like to acknowledge the support of NCN Maestro Grant No. 2012/06/A/ST1/00258. They would also like to thank the Editor and two anonymous referees for their constructive criticism leading to an improved presentation of this article.

Appendix A. Stationarity

All tools presented here assume that the data is either stationary or increment stationary [42]. Stationary and non-stationary processes are very different in their properties, and they require different inference procedures. First, note that a simple and useful method to determine if a process is stationary in empirical studies is to plot the data. Loosely speaking, if a series does seem to have a linear trend, seasonal trend or a varying volatility, then it is very likely it is not stationary. In this case one has to transform the data to make them stationary. To this end, it is usually enough to remove the trends by means of fitting and subtracting a polynomial and a trigonometric function.

For a more refined diagnostic of dependence and stationarity, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are examined. The sample
autocorrelation function at lag $k$ is defined as
\begin{equation}
    r_k = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{N} (X_t - \bar{X})}, \quad \text{for } k = 0, 1, \ldots, K
\end{equation}
and stands for a measure of the linear dependence between observations with a time lag $k$ [18]. The partial autocorrelation function $\eta_k$, which, at lag $k$, is defined as a correlation between the predictor errors of values $X_n$ and $X_{n+k}$ represented in terms of $X_{n+1}, \ldots, X_{n+k-1}$ [18]. For any ergodic process, both functions, ACF and PACF, should approach zero as the lag tends to infinity. They also provide information about the order of the underlying ARMA model. This leads to a simple visual test for stationarity.

**Algorithm 1.** Stationarity test for one sample

(i) Plot the data. If the data possesses a clear linear or seasonal trend, try to remove it.

(ii) Calculate the ACF and PACF. If they converge to zero, this suggests stationarity of the underlying process. In particular:

(a) For i.i.d. data the autocorrelation function should be zero for all $k \neq 0$. The partial autocorrelation function should be equal to zero except for the lag $k = 0$.

(b) For moving average processes of order $q$ (MA($q$)), the autocorrelation function should be zero for lags beyond $q$. Hence, it can be used for estimating the $q$ parameter.

(c) For autoregressive processes of order $p$ (AR($p$)), the partial autocorrelation function should be zero for lags beyond $p$. Hence, it can be used for estimating the $p$ parameter.

We also note that if the model possesses both the AR and MA parts then the reasoning about the orders of the models (estimation of $p$ and $q$) can be quite misleading [18].

If a number of realizations of a phenomena are recorded, the stationarity property of the analyzed datasets can be checked by means of so-called quantile lines. Let us assume that we observe $M$ samples of length $N$ and denote their values by $\{Z_k^n\}$, $n = 1, 2, \ldots N$, $k = 1, \ldots, M$, and $0 < p_j < 1$, $j = 1, \ldots, J$ are given probabilities.

**Algorithm 2.** Stationarity test for many samples [54]

(i) Derive estimators of the corresponding quantiles $q_j(n) = F_n^{-1}(p_j)$, where $F_n = F_n(x)$ denotes the CDF of the random variable $Z_n$ represented by the statistical sample $Z_n^k$, $k = 1, \ldots, M$.

(ii) Obtain the approximation of the so-called quantile lines, i.e. the curves $q_j = q_j(n)$ defined by the condition $P\{Z_n \leq q_j(n)\} = p_j$. In layman terms, the quantile lines represent the value $q_j$ for which $p_j \ast 100\%$ of the data are below at a certain time point $n$.

doi:10.1088/1742-5468/2014/10/P10036
(iii) For a stationary process the quantile lines $q_j(n) = \text{const}$, whereas for a self-similar process they behave like $n^H$.

**Appendix B. Mixing and ergodicity**

We will describe now two further fundamental properties of the data: ergodicity and mixing. Ergodicity of the stationary process $Y(n)$ means that its phase space cannot be divided into two nontrivial sets such that a point starting in one set will never get to the second set. Let us emphasize that for every stationary and ergodic process the Boltzmann ergodic hypothesis, enabling better analysis of the data characteristics is satisfied, i.e. the temporal and ensemble averages coincide.

Another fundamental property is mixing, i.e. the asymptotic independence of the random variables $Y(n)$ and $Y(0)$ as $n$ goes to infinity. It is well-known that mixing is a stronger property than ergodicity [55]. Thus to show ergodicity it is enough to prove mixing, which is easier in many cases.

To this end, we use the dynamical functional (DF) test recently developed in [56]. It is based on a concept of the dynamical functional [33]. The dynamical functional $D(n)$ corresponding to the process $Y(n)$ is defined as

$$D(n) = \langle \exp\{i[Y(n) - Y(0)]\} \rangle.$$

Thus, $D(n)$ is actually a Fourier transform of $Y(n) - Y(0)$ evaluated for the Fourier-space variable $k = 1$. The following result illustrates the strength of the dynamical functional [57]. The stationary ID process $Y(n)$ is mixing if and only if

$$\lim_{n \to \infty} D(n) = |\langle \exp\{iY(0)\} \rangle|^2.$$

The above condition should be viewed as the asymptotic independence of $Y(n)$ and $Y(0)$ as $n \to \infty$. Moreover, if $Y(n)$ is Gaussian, then the dynamical functional is equal to

$$D(n) = \exp\{\sigma^2[r(n) - 1]\},$$

where $r(n)$ is the autocorrelation function of $Y$ and $\sigma^2$ is the variance of $Y(0)$. Thus, in the Gaussian case, the condition (B.2) is equivalent to the fact that $r(n) \to 0$ as $n \to \infty$.

The above condition (B.2) can be written in the equivalent form

$$\lim_{n \to \infty} E(n) = 0,$$

where

$$E(n) = D(n) - |\langle \exp\{iY(0)\} \rangle|^2.$$

The DF test holds for all infinitely divisible stationary processes [56].

Consequently, using the formula (B.2) we find that the stationary ID process $Y(n)$ is ergodic if and only if

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} D(k) = |\langle \exp\{iY(0)\} \rangle|^2.$$

Equivalently, $Y(n)$ is ergodic if and only if

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} E(k) = 0.$$
It should be mentioned that in order to check the necessary and sufficient conditions for mixing (B.3) and ergodicity (B.6) in the language of the dynamical functional (B.1), the reasonable length of each analyzed trajectory should not be shorter than 500 points. The number of trajectories needed to calculate ensemble averages depends strongly on the underlying distribution. For the Gaussian case it is sufficient to have about 100 trajectories, but it is not sufficient for a heavy-tailed $\alpha$-stable law. Every distribution needs to be analyzed separately.

It gets much more complicated, when there are not enough trajectories to calculate ensemble averages. Suppose that we have only one realization of the process $Y(n)$, $n = 0, 1, ..., N$, where $N$ is an appropriately large integer. If we assume that $Y(n)$ is mixing, then the Boltzmann ergodic hypothesis is satisfied—the temporal and ensemble averages coincide.

**Algorithm 3. Testing of mixing and ergodicity properties on one trajectory [56]**

(i) Approximate the dynamical functional $D(n)$ by

\[
\hat{D}(n) = \frac{1}{N-n+1} \sum_{k=0}^{N-n} \exp\{i[Y(n+k) - Y(k)]\}, \quad (B.7)
\]

(ii) Approximate the ensemble average on the right side of (B.2) by

\[
\hat{a} = \left| \frac{1}{N+1} \sum_{k=0}^{N} \exp\{iY(k)\} \right|^2 \quad (B.8)
\]

(iii) If

\[
\hat{E}(n) = \hat{D}(n) - \hat{a} \approx 0 \quad (B.9)
\]

for a large $n$, then the process can be mixing. Violation of (B.9) implies that $Y(n)$ does not have the mixing property.

(iv) Check the following condition

\[
\frac{1}{n} \sum_{k=0}^{n-1} \hat{E}(k) \approx 0. \quad (B.10)
\]

for large $n$. If the condition is satisfied, then the process can be ergodic. Its violation implies ergodicity breaking.

It should be underlined that (B.9) and (B.10) are by no means sufficient for mixing and ergodicity, respectively. This means that having only one trajectory of a random process, we can prove a lack of mixing or ergodicity. To show that the process is mixing or ergodic, ensemble averages need to be calculated [5].

We note that the non-stationary (and non-ergodic) models like CTRW can be distinguished from the ergodic and stationary models (as exemplified by FBM) by applying tests aimed at checking the stationarity of increments or ergodicity. In [58] a test, based on a fractal structure, was introduced to discriminate between FBM and OD.
Appendix C. Memory parameter estimator based on sample mean-squared displacement

In the literature, different methods of assessing long-range dependence and estimating the memory parameter $d$ have been developed [19]. It is important to realize what the assumptions and limitations of various tools are and to realize what the exact output is of the different estimators. This also refers to self-similarity index estimators. For example, a very well-known RS method, in the general stable case, does not return $H$, which is true only in the Gaussian case, but the value $d + 1/2$, where $d = H - 1/\alpha$ and $\alpha$ is the index of stability.

The sample mean-squared displacement (MSD) for the general Lévy stable case was calculated in [6]. Let $\{X_i, i = 0, \ldots, N\}$ be a sample of length $N + 1$. The sample MSD $M_N(\tau)$ for lag $\tau$ is defined as

$$M_N(\tau) = \frac{1}{N - \tau + 1} \sum_{k=0}^{N-\tau} (X_{k+\tau} - X_k)^2. \quad (C.1)$$

The sample MSD is a time average MSD on a finite sample regarded as a function of difference $\tau$ between observations. It is a random variable in contrast to the ensemble average which is deterministic. The following fact describes the behaviour of the sample MSD for the partial sum process of the ARFIMA time series. Let $\{Y_i, i = 0, \ldots, N\}$ be a ARFIMA($p, d, q$) time series with $\alpha$-stable noise and $1 < \alpha \leq 2$. We define its partial sum process $\{X_k = \sum_{i=1}^{k} Y_i, k = 0, \ldots, N\}$. Then for large $N/\tau$:

$$M_N(\tau) \overset{D}{\sim} \tau^{2d+1}. \quad (C.2)$$

Since the normalized partial sum of the ARFIMA process converges to a FLSM with $H = d + 1/\alpha$ [35], we obtain the following fact [6]. Let $\{X_i, i = 0, \ldots, N\}$ be a FLSM with $0 < H < 1$ and $1 < \alpha \leq 2$. Then, for large $N$,

$$M_N(\tau) \overset{D}{\sim} \tau^{2d+1}, \quad (C.3)$$

where $d = H - 1/\alpha$.

In particular, for a FBM we obtain the well-known result that $M_N(\tau) \sim \tau^{2H}$, and for both BM and LSM we arrive at the diffusion case, namely $M_N(\tau) \sim \tau$ since $d = 0$.

As a consequence, we see that the memory parameter $d$ controls the type of anomalous diffusion [6]. If $d < 0$ ($H < 1/\alpha$), in the negative dependence case, the process follows the subdiffusive dynamics, if $d > 0$ ($H > 1/\alpha$), the character of the process changes to superdiffusive. Moreover, it appears that $\alpha$-stable processes for $\alpha < 2$ can serve both as examples of subdiffusion and superdiffusion. The subdiffusion pattern arises when the dependence is negative, so possible large positive jumps are quickly compensated for by large negative jumps, and on average the process travels shorter distances than the light-tailed Brownian motion.

Now, we introduce a method of estimation for the memory parameter $d$ based on the notion of the sample MSD. It can be applied to the ARFIMA time series and fractional stable noise for both Gaussian ($\alpha = 2$) and non-Gaussian ($\alpha < 2$) cases.

Algorithm 4. Estimation of the memory parameter based on sample MSD [46]

(i) Calculate $M_N(\tau)$ for $\tau = 1, 2, \ldots, 10$. 

doi:10.1088/1742-5468/2014/10/P10036
(ii) Fit the linear regression line according to \( \ln(M_N(\tau)) = \ln(C) + a \ln(\tau) \), \( \tau = 1, 2, \ldots, 10 \), where \( C \) and \( a \) are constants.

(iii) The estimated value \( \hat{d} = (a - 1)/2 \).

The introduced sample MSD estimator is consistent and works remarkably well especially for the FBM case. Other analyzed processes need more attention due to approximation errors caused by simulation methods and not exact but limiting results. In general, one may check that the variance of the estimator is very low compared with other well-known methods of estimation of \( d \), like, e.g. R/S [49].

Appendix D. ARFIMA parameter estimation

Following [59, 60], we define the \((p + q + 1)\)-dimensional vector \( \beta_0 = (\phi_1, \phi_2, \ldots, \phi_p, \psi_1, \psi_2, \ldots, \psi_q, d) \), where \( \phi_1, \phi_2, \ldots, \phi_p \) and \( \psi_1, \psi_2, \ldots, \psi_q \) are the coefficients of the polynomials \( \Phi_p \) and \( \Psi_q \) respectively. The vector \( \beta_0 \) is from the parameter space \( E = \{ \beta : \phi_p, \psi_q \neq 0, \Phi_p(z)\Psi_q(z) \neq 0 \text{ for } |z| \leq 1, \Phi_p, \Psi_q \text{ have no common roots}, d \in (-1/2, 1-1/\alpha) \} \).

The estimation procedure can be summarized in the following steps.

Algorithm 5. ARFIMA parameter estimation [60]

(i) Calculate the normalized periodogram by

\[
I_n(\lambda) = \left| \sum_{t=1}^{n} X_t e^{-i\lambda t} \right|^2, \quad -\pi \leq \lambda \leq \pi. \tag{D.1}
\]

(ii) The estimator \( \beta_n \) of the true parameter vector \( \beta_0 \) is defined as

\[
\beta_n = \arg \min_{\beta \in E} \int_{\frac{-\pi}{n}}^{\frac{\pi}{n}} \frac{I_n(\lambda)}{g(\lambda, \beta)} d\lambda, \tag{D.2}
\]

where

\[
g(\lambda, \beta) = \left| \frac{\Psi_q(e^{-i\lambda}, \beta)}{\Phi_p(e^{-i\lambda}, \beta)(1 - e^{-i\lambda})^d} \right|^2. \tag{D.3}
\]

If the orders of the model are small, which is often the case, the estimation procedure simplifies. Let us now concentrate on ARFIMA(1, d, 1). In such a case the parameter space \( E \) has 3D, i.e. \( \beta = (\phi_1, \psi_1, d) \in E \), polynomials \( \Phi_p(z) = 1 - \phi_1 z, \Psi_q(z) = 1 + \psi_1 z \) and the power transfer function has the form

\[
g(\lambda, \beta) = \left| \frac{1 + \psi_1 e^{-i\lambda}}{1 - \phi_1 e^{-i\lambda}} \right|^2 \frac{1}{(2 - 2 \cos \lambda)^d}. \tag{D.4}
\]
Therefore, the estimator $\hat{\beta}_n$ can be rewritten as

$$\arg\min_{\beta \in E} \int_{\frac{\pi}{n}}^{\pi} I_n(\lambda) \left[ \frac{1 - 2\phi_1 \cos \lambda + \phi_1^2}{1 + 2\psi_1 \cos \lambda + \psi_1^2} \right] (2 - 2 \cos \lambda)^d d\lambda. \quad (D.5)$$

In particular, for the FIMA$(d,1)$ model, which is feasible in applications, the estimator is

$$\arg\min_{\beta \in E} \int_{\frac{\pi}{n}}^{\pi} I_n(\lambda) \frac{(2 - 2 \cos \lambda)^d}{1 + 2\psi_1 \cos \lambda + \psi_1^2} d\lambda. \quad (D.6)$$

In order to calculate $\hat{\beta}_n$, we used the `fminsearch` function implemented in Matlab, which applies the simplex search method of [61].

The above estimation procedure requires information about the order of the model, namely the $p$ and $q$ parameters. In practice, in order to estimate the parameters, one can investigate the autocorrelation and partial autocorrelation functions (see appendix A) or for the increasing orders of the model, stop the search when the noise becomes close to the white noise sequence. Calculation of the noise can be performed using standard mathematical packages, like, e.g. ITSM [18]. One can check its independence via standard tests, like the Ljung-Box [47].

Another way to estimate the parameters (including the order) of the ARFIMA model is to find such a set of parameters which minimizes the prediction error (the problem is related to so-called backtesting in statistics) [53]. The prediction formula for the ARFIMA processes was discussed in section 2.3. This is one of the most time-consuming fitting procedures but it has an advantage that it is very meaningful to practitioners.

**Appendix E. Fitting the distribution to the noise**

We now consider in detail two possible probability laws underlying the noise (and, consequently, the data): Gaussian and $\alpha$-stable. To determine if the noise comes from a population with a different distribution, one can perform various statistical tests based on the empirical distribution function, e.g. the Kolmogorov–Smirnov and Cramer–von Mises [47].

Stable distributions, also called $\alpha$-stable, are ubiquitous in nature due to the generalized central limit theorem. It says that the stable distributions, like the Gaussian one, attract distributions of the sums of independent identically distributed random variables [33]. Due to this reason, stable distributions naturally appear when the evolution of a system or the result of an experiment are determined by a sum of many random factors.

A random variable $X$ is stable if for some $\alpha \in (0; 2]$, $\sigma \in (0; \infty)$, $\beta \in [-1; 1]$, $\mu \in \mathbb{R}$ its characteristic function has the following form [33]:

$$\phi_X(t) = \begin{cases} 
\exp(-\sigma^\alpha |t|^\alpha (1 - i\beta \text{sgn}(t) \tan \frac{\pi \alpha}{2}) + it\mu) & \text{if } \alpha \neq 1, \\
\exp(-\sigma |t|(1 + i\beta^2 \text{sgn}(t) \ln |t|) + it\mu) & \text{if } \alpha = 1. 
\end{cases} \quad (E.1)$$

Therefore, each stable distribution is characterized by four parameters. The parameter $\alpha$ is called the index of stability, $\sigma$ is the scale parameter, $\beta$ is the skewness parameter and $\mu$ is the location parameter. Stable variable $X$ is denoted by $X \sim S_\alpha(\sigma, \beta, \mu)$, whereas the symmetric $\alpha$-stable random variable is denoted by $X \sim S_\alpha S$. In the latter case $\beta = \mu = 0$. 

doi:10.1088/1742-5468/2014/10/P10036
When $\alpha = 2$, the distribution reduces to a Gaussian. For the Gaussian distribution $\beta$ is irrelevant, and it is characterized by standard deviation and the mean.

We know that for general stable distributions, although they are continuous, there is no elementary form of probability density function. Computational formulae for stable densities and distribution functions are given in [62]. The Fox function representation for the stable distributions is exhibited in [63] and stable densities in terms of the incomplete hypergeometric function are expressed in [64]. The only stable distributions with elementary probability density functions are: the Gaussian distribution, namely $X \sim S_2(\sigma, 0, \mu) = N(\mu, 2\sigma^2)$, the Cauchy distribution, namely $X \sim S_1(\sigma, 0, \mu)$, and the Lévy distribution, namely $X \sim S_{1/2}(\sigma, 1, \mu)$.

The tails of non-Gaussian stable distributions decrease like a power function: $|x|^{-\alpha}$. Due to this reason, they appear naturally in the description of random processes with large outliers. The rate of decay mainly depends on the parameter $\alpha$. The smaller the $\alpha$, the slower the decay and the heavier the tails. Consequently, for a stable random variable $X$ with index $\alpha < 2$ one has $\langle |X|^{\delta} \rangle = \infty$ for any $\delta \geq \alpha$ and $\langle |X|^{\delta} \rangle < \infty$ for $0 < \delta < \alpha$.

The comprehensive theory of $\alpha$-stable distributions is presented in [33].

There are at least three standard procedures for estimating stable law parameters: (i) the maximum likelihood method based on numerical approximation of the stable likelihood function; (ii) the quantile method using tabulated quantiles of stable laws; and (iii) the method using regression on the sample characteristic function. The regression method is considered to be both fast and accurate [65].

If the analyzed data comes from FLSM or ARFIMA with Lévy stable noise, one can apply the computer test from [34] to estimate the stability index $\alpha$ of the data. The test applies the concept of surrogate data [66], which refers to data that preserves certain linear statistic properties of the experimental time series, without the deterministic component. If the data comes from FLSM or ARFIMA, then we should observe a change to $1/\alpha$ in the self-similarity estimator values for FLSM and aggregate ARFIMA.

**Algorithm 6. Computer test to estimate the stability index [34]**

(i) Obtain the surrogate data, e.g. by random shuffling of the original data positions.

(ii) Estimate the self-similarity index $H$, e.g. via the $p$-variation algorithm (see appendix F). The resulting $\alpha = 1/H$.

The test can also be also to distinguish between diffusion (BM and LSM) and fractional (anomalous) diffusion (FBM and FLSM) models.

To test whether the data follows a Gaussian distribution it is sufficient to apply one of the standard tests implemented in many mathematical packages, like, e.g. Jarque–Bera or Kolmogorov–Smirnov [68]. For the non-Gaussian stable distribution, one can apply a strict testing procedure for an arbitrary distribution or employ specific properties of the stable law. We concentrate here on the latter case. One way to test for stability is to check the distinctive property of stable random variables, namely summability, see, e.g. [67]. The second array of tests checks the power-law behaviour of the underlying data. In this context an issue arises of recognizing stable distributions with the stability index close to 2. In this case, if the sample is not long enough, the shape of empirical PDF is close to a Gaussian [68], and both log–log scale analysis and standard estimators of the power-law exponent estimators like Hill give an overestimated value of $\alpha$ for the
number of observations less than $10^6$ [69]. In applications, the number of observations is often less.

The problem of recognizing a $\alpha$-stable distribution with $\alpha$ close to 2 from experimental data, when the sample size of available data is not large, was addressed in [68]. Following [68] we introduce a testing procedure combining a simple visual test based on the fourth empirical moment, and the Anderson–Darling (AD) and Jarque–Bera (JB) statistical tests. Namely, we calculate the empirical cumulative fourth moment (ECFM) of the simulated data sets, which for a sample of observations $\{x_1, \ldots, x_n\}$ is defined as follows:

$$C(k) = \frac{1}{k} \sum_{i=1}^{k} (x_i - \bar{x})^4, \quad k = 1, 2, \ldots, n, \quad (E.2)$$

where $\bar{x}$ is the mean of the random sample.

The formula (E.2) can be calculated for any sample obtained from an arbitrary probability distribution. For a fixed $k$ it forms a random variable. For distributions with finite fourth moment (e.g. Gaussian), the ECFM, as a function of $k$, converges to a constant, whereas for distributions with infinite fourth moment (e.g. stable with $\alpha < 2$) it diverges to infinity. The latter, for a finite sample, can be observed as an irregular chaotic behaviour. The resulting procedure is presented in the following algorithm.

**Algorithm 7. Recognizing $\alpha$-stable distribution with $\alpha$ close to 2 [68]**

(i) Calculate the ECFM for the sample.

(ii) If the ECFM tends to a constant, we check for the Gaussian distribution by using the JB test.

   (a) If its $p$-value exceeds the confidence level (usually 5%), then we can assume the underlying distribution of the time series is Gaussian. In this case we estimate its parameters by using the standard maximum likelihood estimation method.

   (b) If the JB test shows the data cannot be modelled by a Gaussian distribution, then we test them for the stable distribution.

   1. If the AD test gives a $p$-value that exceeds the confidence level, then we can assume the time series can be described by the $\alpha$-stable distribution. In this case we estimate its parameters via the regression approach.

   2. If the stability is rejected, other distributions have to be taken into considerations.

(iii) If the ECFM exhibits chaotic-like behaviour, then we test for the stable distribution by means of an AD test.

   (a) If the $p$-value is greater than the confidence level, then we can assume the data follows the stable law. In this case we estimate its parameters via the regression approach.

   (b) If the stability is rejected, other distributions have to be taken into considerations.
Appendix F. Self-similarity index estimator based on sample $p$-variation

Let us now discuss the idea of $p$-variation, $p > 0$. The concept of $p$-variation generalizes the well-known notions of total or quadratic variations, which have found applications in various areas of physics, mathematics and engineering [23, 70]. Let $X(t)$ be a stochastic process analyzed on the time interval $[0, T]$. Then, the $p$-variation of $X(t)$ is defined as the limit of sum of increments of $X(t)$ taken to the $p$-th power over all partitions $P$ of the interval $[0, T]$, when the mesh of the partitions goes to zero. When $p = 1$ it reduces to the total variation, whereas $p = 2$ leads to the notion of quadratic variation.

In practice, one calculates sample $p$-variation [6] taking differences between every $m$-th element of the data.

Let $\{X_i, \ i = 0, \ldots, N\}$ be a sample of length $N + 1$. Sample $p$-variation $V_m^{(p)}$ for lag $m$ is defined as

$$V_m^{(p)} = \sum_{k=0}^{N/m-1} |X_{(k+1)m} - X_{km}|^p. \quad (F.1)$$

Note the essential difference between increments of $X_i$ in (F.1) and (C.1). The latter is a moving window, whereas the former is defined on disjoint subintervals (blocks).

The following fact describes the behaviour of the $p$-variation for the partial sum process of the ARFIMA time series. Let $\{Y_i, \ i = 0, \ldots, N\}$ be a ARFIMA($p, d, q$) time series with $\alpha$-stable noise and $1 < \alpha \leq 2$. We define its partial sum process $\{X_k = \sum_{i=0}^k Y_i, \ k = 0, \ldots, N\}$. Then for large $N/m$: if $\alpha = 2$, then $V_m^{(p)} \sim m^{H_p-1}$; if $1 < \alpha < 2$, then $V_m^{(p)} \sim m^{H_p-1}$ for $d > 0$ and $V_m^{(p)} \sim m^{p(H-1/\alpha)}$ for $d < 0$. The symbol $\sim$ denotes similarity in distribution and $H = d + 1/\alpha$.

Since the normalized partial sum of the ARFIMA process converges to a FLSM with $H = d + 1/\alpha$ [35], we obtain the following fact. Let $\{X_i, \ i = 0, \ldots, N\}$ be a FLSM with $0 < H < 1$ and $1 < \alpha \leq 2$. Then for large $N/m$: if $\alpha = 2$, then $V_m^{(p)} \sim m^{H_p-1}$; if $1 < \alpha < 2$, then $V_m^{(p)} \sim m^{H_p-1}$ for $H > 1/\alpha$ and $V_m^{(p)} \sim m^{p(H-1/\alpha)}$ for $H < 1/\alpha$. The symbol $\sim$ denotes similarity in distribution.

This implies that in the case of Gaussian ARFIMA or FBM for $p > 1/H$ sample $p$-variation is a strictly increasing function of $m$ (it tends to zero as $m$ gets smaller), whereas for $p < 1/H$ it is a strictly decreasing function of $m$ (it diverges to infinity when $m$ gets smaller). For a stable non-Gaussian ARFIMA or FLSM the situation differs from the Gaussian case and depends on whether $d$ is positive or negative. It appears that the sample $p$-variation is always a decreasing function with respect to $m$ when $d < 0$. If $d > 0$, the situation is the same as in the Gaussian case: if $p > 1/H$, then sample $p$-variation is an increasing function of $m$, if $p < 1/H$ it is a decreasing function of $m$.

Now, we introduce a method of estimation of the self-similarity index $H$ based on sample $p$-variation for the ARFIMA process and fractional stable motion. For $d > 0$ the method applies to both Gaussian ($\alpha = 2$) and non-Gaussian ($\alpha < 2$) cases. For $d < 0$ it is defined only for the Gaussian case. The idea of the method is to find $p = 1/H$ for which $V_m^{(p)}$ as a function of $m$ changes its monotonic behaviour, i.e. becomes constant. To this end, as a simple tool, we propose to calculate the difference between $V_m^{(p)}$ for the smallest $m$, namely $m = 1$, and $V_m^{(p)}$ for a larger $m$ (but not too large to ensure $N/m$ is sufficiently high), e.g. $m = 8$. This is done for different $p$’s. The smallest distance defines
the value of the estimator $\hat{p}$ and consequently $\hat{H}$. The procedure can be summarized as follows.

**Algorithm 8.** Estimation of the self-similarity index based on sample $p$-variation [49]

(i) Calculate $V_m^{(p)}$ for $p = 1/(0.01 : 0.01 : 0.99)$ and $m = 1$ and 8.

(ii) Find $\hat{p}$ that minimizes $(V_8^{(p)} - V_1^{(p)})^2$.

(iii) The estimated value $\hat{H} = 1/\hat{p}$.

The introduced $p$-variation estimator seems to be consistent and works remarkably well for the FBM case. It is worth-mentioning that the variance of the estimator is very low in comparison to other methods of estimation of the self-similarity index [49].

Finally, we note that the idea of $p$-variation can be used to distinguish between two mechanisms leading to anomalous diffusion, namely FBM and CTRW. In [70] a simple test, based on one realization of the unknown process, was introduced and applied to the data of Golding and Cox [43].

**References**

[22] Saxton M J 2012 Biophys. J. 103 2411
[31] Davies R and Harte D 1987 Biometrika 74 95

doi:10.1088/1742-5468/2014/10/P10036
Algorithms for testing of fractional dynamics

[35] Stoev S and Taqqu M S 2004 Fractals 12 95
[40] Burnecki K, Muszkieta M, Sikora G and Weron A 2012 Europhys. Lett. 98 10004
[50] Ljung G M and Box G E P 1978 Biometrika 65 297
[52] Clark S 2006 Nature 441 402
[66] Chang T, Sauer T and Schiff S 1995 Chaos 111 111

doi:10.1088/1742-5468/2014/10/P10036