Timothy Graves, Christian L.E. Franzke, Nicholas W. Watkins, Robert B. Gramacy, Elizabeth Tindale

Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models

Article (Accepted version) (Refereed)

Original citation: Graves, Timothy, Franzke, Christian L.E., Watkins, Nicholas W., Gramacy, Robert B. and Tindale, Elizabeth (2017) *Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models.* Physica A: Statistical Mechanics and Its Applications . ISSN 0378-4371

DOI: 10.1016/j.physa.2017.01.028

Accepted Manuscript

Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models

Timothy Graves, Christian L.E. Franzke, Nicholas W. Watkins, Robert B. Gramacy, Elizabeth Tindale

PII: S0378-4371(17)30029-8
DOI: http://dx.doi.org/10.1016/j.physa.2017.01.028
Reference: PHYSA 17921

To appear in: *Physica A*

Received date : 16 July 2016
Revised date : 14 December 2016

Please cite this article as: T. Graves, et al., Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models, *Physica A* (2017), http://dx.doi.org/10.1016/j.physa.2017.01.028.

Highlights

- Self-Similarity has two contributions: Long-range dependence and heavy-tailed jumps
- Systematic simultaneous estimation of long-range dependence and heavy-tail distribution parameters
- Development of a novel Bayesian method for estimation of these two parameters
- Method flexible to allow choice of heavy-tailed distribution (e.g. t- or α-stable distributed)
- Successful demonstration of effectiveness and accuracy on synthetic data

Systematic Inference of the Long-Range Dependence and Heavy-Tail Distribution Parameters of ARFIMA models

Timothy Graves[a], Christian L. E. Franzke[b], Nicholas W. Watkins[c,d,e,f], Robert B. Gramacy[g], Elizabeth Tindale[d]

[a] *Arup, London, UK*
[b] *Meteorological Institute and Center for Earth System Research and Sustainability, University of Hamburg, Germany*
[c] *London School of Economics, London, UK*
[d] *University of Warwick, Coventry, UK*
[e] *Open University, Milton Keynes, UK*
[f] *Institute of Physics and Astronomy, University of Potsdam, Potsdam-Golm, Germany*
[g] *Virginia Tech, USA*

Abstract

Long-Range Dependence (LRD) and heavy-tailed distributions are ubiquitous in natural and socio-economic data. Such data can be self-similar whereby both LRD and heavy-tailed distributions contribute to the self-similarity as measured by the Hurst exponent. Some methods widely used in the physical sciences separately estimate these two parameters, which can lead to estimation bias. Those which do simultaneous estimation are based on frequentist methods such as Whittle's approximate maximum likelihood estimator. Here we present a new and systematic Bayesian framework for the simultaneous inference of the LRD and heavy-tailed distribution parameters of a parametric ARFIMA model with non-Gaussian innovations. As innovations we use the α-stable and t-distributions which have power law tails. Our algorithm also provides parameter uncertainty estimates. We test our algorithm using synthetic data, and also data from the Geostationary Operational Environmental Satellite system (GOES) solar X-ray time series. These tests show that our algorithm is able to accurately and robustly estimate the

*Corresponding author
Email addresses: christian.franzke@uni-hamburg.de (Christian L. E. Franzke)

LRD and heavy-tailed distribution parameters.

Key words: Long-Range Dependence, Heavy-Tails, Bayesian Estimation, ARFIMA

1. Introduction

Long-range dependence (LRD) is an ubiquitous property of many physical, biological and financial systems [1, 31]. Hurst's observation (the "Hurst effect") of the anomalous rate of growth of range in hydrological time series, such as the height of the river Nile, was one of the first natural phenomena for which the need for a non-Brownian statistical description was recognised. Mandelbrot & Van Ness [28] explained the Hurst effect as being due to long range dependence in time, which Mandelbrot & Wallis [29] then dubbed the "Joseph effect". Mandelbrot and his co-authors encapsulated the Joseph effect in their seminal model, fractional Brownian motion (fBm), using its stationary increments, fractional Gaussian noise, to model the Nile time series. Like the more familar Wiener Brownian motion, fBm has the property of self-similarity under a dilation in time where Δt is replaced by $\lambda\Delta t$:

$$x(\lambda\Delta t) \stackrel{d}{=} \lambda^H x(\Delta t) \tag{1}$$

Throughout our paper we will follow Embrechts & Maejima [11], by defining H as the self-similarity exponent. In fBm H takes values between 0 and 1, with $H = 1/2$ being the Brownian case. To describe the growth of rescaled range (R/S) ("the Joseph effect") due to the persistence seen in the increments of fBm, Mandelbrot & Van Ness [28] used a second exponent J, where

$$R/S \sim \tau^J. \tag{2}$$

The presence of LRD affects the predictability of systems and their long-term behaviour and has thus continued to be controversial.

For reasons that are as much historical as technical [16, 27], the parameters of Gaussian LRD models such as fractional Gaussian noise have typically been inferred either indirectly from the self-similarity exponent H, or directly using J (frequently called the Hurst exponent and also denoted by H), because in such models the self-similarity and Hurst exponents happen to coincide [34]. However, what is frequently still not appreciated is that the self-similarity exponent actually has two contributions: (i) one from

LRD (also called the Joseph effect) and (ii) one from non-Gaussian jumps which are power-law distributed (also called the 'Noah' effect); e.g. α-stable increments, which have probability density function (pdf) tails decaying as $f(x) \sim x^{-(1+\alpha)}$ where the index α runs from 0 to 2. As Mandelbrot emphasised [e.g 27, p. 157] H can be different from J, and R/S only measures the latter. We will thus avoid the potentially confusing term 'Hurst exponent' in this paper and label the contribution of memory to the self-similarity exponent by J, as Mandelbrot recommended after the ambiguity became clear to him.

A second type of non-Brownian phenomenon had also been recognised by Mandelbrot [25]. This was the non-Gaussian increments, with "heavy" power-law tails in the pdf,

$$f(x) \sim x^{-(1+\alpha)} \tag{3}$$

seen in financial time series [31] and also in many natural ones. In contrast to the LRD he called this the "Noah effect". He proposed a second paradigmatic model, ordinary Levy motion (oLm), for cases when the anomalous behaviour of the time series originates entirely from this effect, rather than long temporal memory.

Real time series do not necessarily exhibit just one or the other of these two limiting cases. Mandelbrot & Wallis [30] thus proposed that the effects modelled by fBm and oLm could be combined in a more general self-similar additive model, "fractional hyperbolic" [30] motion, a descendent of which is now referred to as linear fractional stable motion, LFSM, [e.g. 11]. LFSM has by now been applied to problems as diverse as communications traffic [24], geophysics [38], magnetospheric physics [47] and solar flares [44]. The "ambivalent" [4] dual behaviour of such models makes it important to develop methods which can simultaneously estimate both the Joseph and Noah effects and their corresponding exponents J and α.

In our paper we use a newer, more flexible time series model: the well-known Autoregressive Fractional Integrated Moving Average (ARFIMA) [e.g. [1]] with non-Gaussian increments [e.g. [8]], which also allows an adjustable high frequency component. In this model J is encapsulated by the standard LRD parameter d used in statistics, which ranges from $-1/2$ to $1/2$, with $d = 0$ being the uncorrelated, white noise case. Our algorithm allows for α-stable but also t-distributed increments and can also easily be extended to use any distribution characterized by only a shape and scale parameter.

Our method is based on the Bayesian ARFIMA inference algorithm of Graves et al. [17] for which we developed a new approximate likelihood for the efficient parameter inference. We will show how nuisance parameters (e.g. short memory effects) can be integrated out in order to focus systematically on the long memory parameter. Here we extend our method to simultaneously estimate the LRD and the heavy-tailed parameter. As heavy-tailed distributions we use the t-distribution and the α-stable distribution. For computational reasons we have to restrict our inference to the finite mean ($1 < \alpha \leq 2$) case.

It was realised as early as 1969 by Mandelbrot & Wallis [30] that non-parametric LRD estimators are not "fooled" by the presence of non-Gaussianity [14], not least because they measure J rather than H. However, it is still advantageous to perform simultaneous parameter estimation in order to minimize estimation bias, and to provide direct estimates of α rather than having to estimate the tail exponent by some other means such as a measurement of the pdf or cdf. This is especially important for ARFIMA type models which contain both Short-Range Dependence (SRD) and LRD characteristics, in contrast to the pure mono-fractal approach originally taken by Mandelbrot with fBm [28].

The standard approach [e.g. [1, 46, 7]] in statistics to estimating the parameters of finite variance ARMA models is ultimately derived from a variant of Whittle's method proposed by Hannan [20]. Successive developments have encompassed Gaussian ARFIMA [13], ARMA with α-stable noise [35], and ARFIMA with such noise [23]. These developments are accessibly summarised by [7] which constructs a new estimator for ARFIMA and importantly can access the negative d range which is not accessible to the original Whittle estimator proposed in [20].

Our new inference algorithm is based on the one put forward in [16, 17]. The algorithm consists of a systematic Bayesian framework, a new approximate likelihood for ARFIMA processes and an efficient blocked Monte Carlo Markov Chain (MCMC) sampler. Our Bayesian inference algorithm has been designed in a flexible fashion so that, for instance, the innovations can come from a wide class of different distributions such as the α-stable "Levy" class [31], or the t distribution that is also widely employed in finance [3]. Our algorithm can also estimate the SRD parameters, although these can be integrated out if one is only interested in the LRD and heavy-tail parameters. To our knowledge, a t-distribution has not been considered in LRD models so far.

The Bayesian approach allows us the direct computation of the probability of a theory or model parameter [19]. For our purposes, the important difference between frequentist and Bayesian approaches is that in the former the parameters ψ of a statistical model are taken as fixed, whereas in the latter they are uncertain variables. Assume we have data $\mathbf{x}$ which is a realisation from an unknown distribution. A given hypothesised model distribution will have a likelihood $L(\mathbf{x}|\psi)$, i.e. the likelihood of getting that data $\mathbf{x}$ given a set of values of the parameters ψ. The best estimate of these parameters, e.g. the LRD parameter d, is what we want.

Bayes' theorem states that:

$$\pi_{\psi,\mathbf{x}}(\psi|x) \propto p_{\psi}(\psi)L(\mathbf{x}|\psi) \tag{4}$$

where p_{ψ} is a *prior* probability density on the parameters ψ, and π_{ψ} is the desired *posterior* density. The generic 3 stage approach this allows us to use is i) postulate a p_{ψ}, then ii) multiply the prior by the calculated likelihood function for that model L and normalise, i.e. apply Bayes' theorem; and so iii) generate the posterior π_{ψ}. In principle this could be completely analytically calculable, but in practice one usually has to use computational methods like Markov Chain Monte Carlo (MCMC) algorithms because it becomes analytically intractable.

Our paper is structured as follows: In section 2 we describe our inference approach for the memory parameter d which is related to J as $J = d + 0.5$. Section 3 is the main part of the paper and describes the extensions of the ARFIMA inference algorithm of Graves [16], Graves et al. [17] to stable innovations. We summarize in section 4.

2. Bayesian inference on Gaussian ARFIMA model for *d*.

We first briefly describe the Bayesian inference of an ARFIMA model with Gaussian innovations [16, 17].

An ARFIMA model has three classes of parameters: those governing the location (here the mean μ); the innovation distribution (here just the scale σ); and memory structure (here LRD parameter d, and the AR and MA series, ϕ and θ respectively, which are of order p and q in an ARFIMA(p, d, q) model).

We choose flat priors for μ, $\log \sigma$ and d. Flat priors are non-informative; i.e. a flat prior in the coin-tossing heads or tails case is $1/2$, while in an N category case it is $1/N$. As we have no analytic form for the posterior distribution π, we use MCMC sampling. MCMC [9] is a way to simulate complex, nonstandard multivariate distributions. There are several types of MCMC, including Metropolis-Hastings [9] which we use extensively.

Our approach has several advantages. First, we don't need to assume the order of the ARFIMA(p, d, q) model, i.e. pre-specify p, q. Rather we use the reversible jump (RJ) MCMC approach [18]. In this the parameter space of ψ is extended to include the set of possible models. The Markov chains move *between* models as well as within them. Reversible-jump MCMC allows the sampling of the posterior distribution on spaces of varying dimensions using a transdimensional Markov Chain. Thus, the simulation is possible even if the number of parameters in the model is not known.

Second, our approach allows the reparameterisation of the model to enforce stationary constraints on ϕ and θ. This reparameterization improves the computational efficiency of our algorithm [16].

Third, our approach allows a fast approximate Gaussian likelihood calculation. The LRD correlation structure, which considerably enlarges the dimension of the covariance matrix, prevents use of standard likelihood methods. Previously we [17] proposed a method for the fast evaluation of conditional likelihoods, e.g. $(n \log n)$ in the Gaussian case. Our use of the Metropolis-Hastings algorithm requires the careful selection of proposal distributions. In order to propose new parameters we use a truncated Normal random walk because this sampler has a finite range of $-1/2 < d < 1/2$ for the LRD parameter.

3. Non-Gaussian innovations

In the literature on ARFIMA models, Gaussianity of the innovations is typically assumed. This assumption is made for at least three reasons. Firstly, Gaussian analysis often turns out to be mathematically convenient because the form of the multivariate normal likelihood allows many problems to be solved exactly. Secondly, the normal distribution is a reasonable model for many "real-life" applications. Thirdly, one often appeals to the role played by the Gaussian distribution in the central limit theorem (CLT). By considering the stochastic elements of a problem to be actually composed of a large number of non-Gaussian small disturbances, then, provided the variance is finite, the CLT enables us to assume their aggregation is approximately Gaussian.

Whilst the first and third of these reasons are both reasonably sound and objective, the second has sometimes been the product of modelers' wishes rather than observational evidence. In practice many real-life processes simply cannot be modelled as Gaussian [3, 31, 48, 38]. In particular, overdispersion is a common problem, i.e. a Gaussian model cannot account for all the observed variability. Consequently, if the data suggest leptokurtosis, the Gaussianity assumption may be inappropriate [12].

We define the tail behaviour as follows:

$$\mathbb{P}(X > x) \sim Cx^{-a}, \tag{5}$$

for some positive constants a and C. Such a distribution will be referred to as 'heavy-tailed' if α is between 0 and 2. Clearly for such distributions, moments only exist up to the a-th one. If X is heavy-tailed with parameter exponent a then:

$$\begin{array}{lll} \mathbb{E}|X|^p < \infty & \text{for any} & 0 < p < a, \\ \mathbb{E}|X|^p = \infty & \text{for any} & p \geq a. \end{array} \tag{6}$$

One example of a heavy-tailed distribution is the family of stable distributions.

3.1. Stable distributions

A random variable X is said to have a stable distribution, denoted $X \sim \mathcal{S}_{\alpha,\beta}(\gamma,\delta)$, if there are parameters $0 < \alpha \leq 2$, $-1 < \beta < 1$, γ positive and δ real, such that its characteristic function has the following form:

$$\log\left[\varphi_{\mathcal{S}}(\theta)\right] = \begin{cases} -\gamma^\alpha |t|^\alpha \left(1 - i\beta(\operatorname{sign} t)\tan\frac{\pi\alpha}{2}\right) + i\delta t & \text{if } \alpha \neq 1 \\ -\gamma|t|\left(1 + i\beta\frac{2}{\pi}(\operatorname{sign} t)\log|t|\right) + i\delta t & \text{if } \alpha = 1 \end{cases}. \tag{7}$$

The support for the stable distribution is the whole real line, except in the case where $\alpha < 1$ and $\beta = \pm 1$, in which case it is limited to a semi-infinite interval of the real line. Note also that if $\alpha = 2$ the parameter β is irrelevant, in which case it is convention to set $\beta = 0$. There are many different parametrisations of the stable distribution; the article by [37] provides an excellent summary. Unfortunately neither the probability density nor distribution functions have generally applicable analytic forms, with a few known exceptions [48, 49]: $\mathcal{S}_{2,0}(\gamma,\delta)$ is the $\mathcal{N}(\delta, 2\gamma^2)$, $\mathcal{S}_{1,0}(\gamma,\delta)$ is the Cauchy distribution, and $\mathcal{S}_{1/2,1}(\gamma,\delta)$ is the Lévy distribution.

Because for $\alpha < 1$ the process has no mean, which causes many difficulties in LRD parameter inference, we will assume henceforth that $1 < \alpha \leq 2$. We denote by parameters δ and γ the 'location' and 'scale' parameters respectively. Although the density is nearly always non-analytical, the stable distribution does satisfy a location-scale density [16] of the form:

$$f(x; \delta, \sigma, \boldsymbol{\lambda}) \equiv \frac{1}{\sigma} f\left(\frac{x-\delta}{\sigma}; 0, 1, \boldsymbol{\lambda}\right). \tag{8}$$

Consequently one need only be concerned with the 'standardised' stable distributions with $\delta = 0$ and $\gamma = 1$, which will be denoted using the shorthand $\mathcal{S}_{\alpha,\beta}$ with corresponding density $f_{\mathcal{S}}(\cdot; \alpha, \beta\ )$.

Typically the parameter controlling the tail decay, α, is referred to as the 'index of stability'. The parameter β is conventionally called the 'skew' parameter since non-zero values induce skewness in the distribution.

Throughout the remainder of this section, it will be assumed that a process $\{X_t\}$ is both causal and invertible and has Wold expansion:

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}, \tag{9}$$

where the coefficients $\{\psi_k\}$ are real and ℓ^2-convergent, and the innovations $\{\varepsilon_t\}$ are independent and identically distributed $\mathcal{S}_{\alpha,\beta}(\gamma, 0)$ for some $1 < \alpha < 2$, $-1 \leq \beta \leq 1$ and positive γ. Recall that for such processes, a stably distributed process has long memory if its Wold expansion decays as a power-law [16]. Throughout most of this paper, only stably distributed ARFIMA$(0, d, 0)$ processes will be considered. We note that in the Gaussian case the condition:

$$-\frac{1}{2} < d < \frac{1}{2}, \tag{10}$$

was required to ensure causality and invertibility. In the stable case, the following stronger condition exists:

$$-\left(1 - \frac{1}{\alpha}\right) < d < 1 - \frac{1}{\alpha}. \tag{11}$$

Note that this is consistent with the $\alpha = 2$, Gaussian, case (10). The region of allowable values of the pair (α, d) is shown in figure 1.

3.1.1. Statistical inference from stable processes

We first discuss how to draw inference in the simplest possible scenario of independent and identically distributed random variables. Because of the lack of an analytical density, and consequently likelihood, stable distributions are notoriously difficult to work with. In many cases, it may only be the parameter α that is of interest and therefore it may seem reasonable to estimate this directly from the tail behaviour (5). The naive approach of calculating the empirical distribution function and using log-regression was developed by [21] and [10] amongst others. But [32] later showed that this method is seriously flawed because the tail behaviour is truly *asymptotic* and for some parametrisations and combinations of parameter values, the power-law behaviour does not occur until far out into the tail.

Bayesian analysis of stable distributions has been limited; [5] considered the problem of finding the joint posterior distribution of the parameters from a collection of n independent and identically distributed stable random variables. Unsurprisingly the posterior is intractable so [5] developed an MCMC method requiring n auxiliary variables and complicated sampling regimes.

3.1.2. Bayesian inference for stably-distributed ARFIMA processes

The most significant challenge in the stably-distributed ARFIMA processes scenario is the efficient computation of the log-likelihood. We need to be able to compute the logarithm of the density $f_{\mathcal{S}}(x;\alpha,\beta)$ for any $\alpha > 1$, $-1 < \beta < 1$ and real x. To compute the log-likelihood efficiently, note that we actually seek to evaluate the same density at n points simultaneously. For this purpose we use the approach developed by Mittnik et al [36], in which the stable density is calculated on a regular grid, from which the density at all the x_k can be interpolated. Their method takes advantage of the fact that the characteristic function φ of stable processes (7) is the 'Fourier-dual' of the probability density function:

$$f_{\mathcal{S}}(x;\alpha,\beta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\varphi_{\mathcal{S}}(t;\alpha,\beta)\,dt. \tag{12}$$

In particular, [36] showed that this integral can be approximated by a sum which, using an N-sized FFT, can calculate (to an arbitrary precision) the values of $f_{\mathcal{S}}(y_k;\alpha,\beta)$ on an N-grid of equally spaced values $\{y_k\}$ where:

$$y_k = h\left(k - 1 - \frac{N}{2}\right), \qquad k = 1,\ldots,N$$

for some N and h. Once this grid has been calculated, the densities $f_{\mathcal{S}}(x_1;\alpha,\beta),\ldots,f_{\mathcal{S}}(x_n;\alpha,\beta)$ can be evaluated by linear interpolation.

There are several issues to note regarding implementation of this scheme. Firstly, the choice of parameters N and h is important. The number of points in the grid is N, so a larger N generates a larger grid (at mild computational expense since the FFT has complexity $\mathcal{O}(N\log N)$). The spacing between points is h, so a smaller h produces a more detailed grid. If the data are fat-tailed, the maximum of $|y_k|$ would be expected to be large which means that Nh must also be large. This means making N large, at the expense of slowing the FFT, or making h large, at the expense of losing detail in the interpolation. We will therefore use fixed values for N and h ($N=2^{13}$, $h=2^{-10}$) and use a series expansion to calculate the remaining outliers by noting that for $1<\alpha\le 2$ the density $f_{\mathcal{S}}(x;\alpha,\beta)$ has the following asymptotic expansion for $x\to\infty$ [2, 33]:

$$
\begin{aligned}
f_{\mathcal{S}}(x) =& \frac{1}{\pi}\sum_{k=1}^{\infty} c a_k (cx)^{-k\alpha-1}\\
\text{where } c =& \cos(\beta^*)^{1/\alpha},\\
a_k =& (-1)^{k-1}\frac{\Gamma(1+k\alpha)}{k!}\sin\left(\frac{k}{2}(\pi\alpha+2\beta^*)\right),\\
\text{and } \tan(\beta^*) =& -\beta\tan(\pi\alpha/2).
\end{aligned}
$$

Further refinements to this FFT-based approximation of stable densities were described by [33], which included numerically calculating the integral in (12) using Simpson's rule, and replacing the linear interpolation with cubic splines. In practice, we found that there was no noticeable advantage to using these more costly techniques in the long memory context.

To simplify matters, we assume a two-dimensional uniform joint prior over the allowable region (Fig. 1):

$$
p_{d,\alpha}(d,\alpha)\propto \mathbb{1}\left\{|d|<1-\frac{1}{\alpha}, 1<\alpha<2\right\}. \tag{13}
$$

Note that this prior places zero probability on $\alpha=2$, i.e. the Gaussian. Because of their qualitatively different behaviours, we will not try and include the cases of $\alpha=2$ and $\alpha<2$ in the same analysis.

The prior (13) results in the marginal prior for d being no longer uniform:

$$\begin{aligned} p_d(d) &= \int_1^2 p_{d,\alpha}(d,\alpha)\,d\alpha \\ &= \frac{1}{2-2\log 2}\left(2-\frac{1}{1-|d|}\right), \qquad |d| < \frac{1}{2}. \end{aligned}$$

Similarly, the marginal prior for α:

$$p_\alpha(\alpha) = \frac{1}{1-\log 2}\left(1-\frac{1}{\alpha}\right), \qquad 1 < \alpha < 2.$$

d and α are updated as follows:

$$\begin{aligned} \xi_d|(d,\alpha) &\sim \mathcal{N}^{\left(-1+\frac{1}{\alpha},1-\frac{1}{\alpha}\right)}(d,\sigma_d^2) \\ \xi_\alpha|(d,\alpha) &\sim \mathcal{N}^{\left(\frac{1}{1-|d|},2\right)}(\alpha,\sigma_\alpha^2), \end{aligned}$$

for some σ_α^2. These proposals are accepted/rejected according to:

$$\begin{aligned} A_d(d,\xi_d) &= \Delta\ell + \log\left\{\frac{\Phi[(1-\frac{1}{\alpha}-d)/\sigma_d]-\Phi[(\frac{1}{\alpha}-1-d)/\sigma_d]}{\Phi[(1-\frac{1}{\alpha}-\xi_d)/\sigma_d]-\Phi[(\frac{1}{\alpha}-1-\xi_d)/\sigma_d]}\right\} \\ A_\alpha(\alpha,\xi_\alpha) &= \Delta\ell + \log\left\{\frac{\Phi[(2-\alpha)/\sigma_\alpha]-\Phi[(\frac{1}{1-|d|}-\alpha)/\sigma_\alpha]}{\Phi[(2-\xi_\alpha)/\sigma_\alpha]-\Phi[(\frac{1}{1-|d|}-\xi_\alpha)/\sigma_\alpha]}\right\}. \end{aligned}$$

with $\Delta\ell = \ell(\mathbf{x}|\xi_d,\boldsymbol{\psi}_{-d}) - \ell(\mathbf{x}|\boldsymbol{\psi})$

An alternative approach would be to propose the pair (d,α) jointly, but this is unnecessarily complicated in practice. It should be noted that, due to numerical issues for α near to 1, the lower bound of $\alpha = 1$ used throughout this procedure is actually replaced by $\alpha = 1.02$ in the computer code. The code we have written also allows α to be fixed (the Bayesian interpretation would be a unit mass prior). In this case, any prior can be used for d. Tuning of the proposal variance σ_α^2 is achieved using the same automatic tuning procedure outlined in Graves et al. [17]. Initial values of (d,α) are chosen uniformly randomly over the allowable region. Comparison of our Figure 1 with Figure 2 of [7] shows that our set of allowable parameter values is smaller, and in particular they can access part of the infinite mean range of α for some negative d values, however for both cases they coincide for positive d.

3.1.3. Application to synthetic data

In this section we evaluate our algorithm. Initially, 60 time series of length $n = 2^{10}$ were simulated with each value of $\alpha_I \in \{1.75, 1.50, 1.25\}$ being used 20 times. Values of d_I were chosen randomly, conditional on the pair (d_I, α_I) being in the allowable range.

The residuals of d and α are presented in figure 2. From plot (a) we see that the Bayesian estimate of d appears to be approximately unbiased and the residual appears to be independent of d_I. It is immediately clear that accuracy of the Bayesian estimator appears to increase as α decreases. To confirm this, the average posterior standard deviations for d were 0.015, 0.009, 0.004 for $\alpha_I = 1.75$, 1.50, 1.25 respectively. In summary therefore, our knowledge about d increases when α is small, i.e. detection of long memory is easier in the presence of fat tails. Although this may seem initially surprising, there is a strong intuitive explanation. Long memory is characterised by the slow decay of the influence of each innovation, or 'shock'. In the Gaussian framework, no shock is very much larger than any other so the traceability of each shock is hard because it gets lost in the 'noise'. Yet in the fat-tailed case, extreme shocks are to be expected, and their effects will be easier to observe through time. As α decreases, such shocks appear more frequently and are more dramatic, and consequently their decay profile is easier to determine.

Naturally we are also interested in the posterior of α. We see from figure 2(b) that the Bayesian estimate of α is essentially unbiased, and the posterior standard deviation of α is roughly independent of d_I and α_I (although there is some suggestion that the posterior standard deviation is smallest when $\alpha_I = 1.25$, this is not practically significant). Interestingly, the joint posterior of (d, α) shows no correlation between the two parameters, indicating that the posteriors are independent. This fact helps to justify not proceeding with a joint proposal of (d, α) in the Metropolis–Hastings step in the previous subsection. Furthermore, it inspires the question: "can we improve our knowledge about d if we *know* the value of α?" A simple Monte Carlo comparison shows that this is not the case (see figure 3) leading to the interesting result that, if we assume no knowledge of α, we sacrifice no information about d.

There is some mild correlation between the posteriors of α and γ (not shown). This is not surprising since both parameters affect the 'variability' of the underlying stable distribution. However the algorithm is able to successfully disentangle the scaling effects of γ and the shaping effects of

α. Finally, an analysis varying the length of times series n, was performed. From Fig. 4 we see that a $n^{-1/2}$ rule applies for stably distributed processes. Note also the relative decreases in posterior standard deviation obtained by decreasing α_I and increasing n. For example, to obtain the same level of confidence about the parameter d when $\alpha_I = 1.5$ and $n = 2^{10} = 1024$, one would have to use a time series of length about $2^{14} = 16384$ in the Gaussian case.

3.2. Asymmetric stable distributions

We will now briefly consider the asymmetric case, where $\beta \neq 0$. Again, because of the careful modularisation of the method, adding in this parameter to the Metropolis–Hastings algorithm presents no difficulty. Any prior on the support $[-1, 1]$ can be used but for convenience we will use the simplest form:

$$p_\beta(\cdot) \sim \mathcal{U}(-1, 1).$$

The proposal distribution is:

$$\xi_\beta|\beta \sim \mathcal{N}^{(-1,1)}(\beta, \sigma_\beta^2),$$

for some σ_β^2. To test the efficacy of the method in this framework, we simulated twenty processes with $\alpha_I = 1.5$, $\beta_I = 0.5$ and d_I randomly in the allowable range. The summary statistics are presented in table 1.

It is clear that the method can accurately determine the 'skewness' parameter β. Further investigation reveals that the posterior of β is uncorrelated with any other parameter (not shown). Also, as $\alpha \to 2$ the marginal posterior standard deviation of β becomes increasingly large, and the distribution actually approaches the uniform on $(-1, 1)$ (also not shown). This is because, as remarked upon earlier, the parameter β becomes increasingly unidentifiable as α increases, and at the Gaussian limit it is irrelevant.

3.3. t-distribution

To demonstrate the flexibility of our Bayesian MCMC algorithm, we will now briefly consider using t-distributed innovations. To our knowledge, there is no literature concerning long memory models with t-distributed innovations, most likely because of the reasons given at the end of the introduction. The t-distribution acts as a useful intermediate between the Gaussian and the power law-tailed α stable distributions. To see this, consider its probability density function:

$$f(x;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\pi\nu}} \left\{1 + \frac{x^2}{\nu}\right\}^{-\frac{\nu+1}{2}}. \tag{14}$$

As $x \to \infty$ the probability density function behaves as $\sim Ax^{-\nu-1}$ for some A and consequently the tail function behaves as $\mathbb{P}(X > x) \sim Bx^{-\nu}$ for some B. By comparison with (5) we see that such distributions are power law-tailed. However, unlike the stable distribution which allowed the tail exponent to be only $0 < a < 2$, here ν may take any positive value, leading to power law tail distributions that can have an arbitrary number of finite moments. In particular, for $\nu > 2$ the t-distribution has finite variance yet is still power law tailed, in direct contrast to stable distributions (and unlike them being attracted to the Gaussian under convolution). It is worth remarking that the limiting distribution of $\nu \to \infty$ is the standard Gaussian. Furthermore, the case $\nu = 1$ is the standardised Cauchy distribution. Recall that these two distributions also correspond to particular values of α-stable distributions ($\alpha = 1$ and 2 respectively).

Turning attention to t-distributed long memory processes, it will be useful to generalise (14) to obtain a scale-location distribution satisfying (8):

$$f(x;\delta,\gamma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\gamma\sqrt{\pi\nu}} \left\{1 + \frac{1}{\nu}\left(\frac{x-\delta}{\gamma}\right)^2\right\}^{-\frac{\nu+1}{2}}. \tag{15}$$

Such a t-distribution has variance $\gamma^2\frac{\nu}{\nu-2}$ for $\nu > 2$ (and infinite for $\nu \leq 2$). Throughout the remainder of this section we will restrict attention to the 'intermediate' t-distributions that have finite variance, $\nu > 2$. As with the stable distribution, the scale parameter will be notated as γ rather than σ to avoid implying that it is also a standard deviation.

Due to the modularisation of the method outlined [17], it is relatively trivial to incorporate the t distribution into the Bayesian framework. Calculation of the log-likelihood is straightforward given the density. The prior for ν can be chosen to be anything supported on the positive half-line. There is no standard non-informative prior for ν, so we will use an exponential prior truncated to the right of $\nu = 2$:

$$p_\nu(\nu) = \lambda e^{-\lambda(\nu-2)}, \qquad \nu > 2, \tag{16}$$

for some λ to be chosen. Using a prior that is independent of the other model parameters again allows simplification in the Metropolis–Hastings step. To propose new values of ν, we will use the same exponentiated random-walk as for the scale parameter described, although restricted to $\nu > 2$:

$$\log(\xi_\nu - 2) = \log(\nu - 2) + \upsilon,$$

where $\upsilon \sim \mathcal{N}(0, \sigma_\nu^2)$ for some σ_ν^2. Calculation of the relevant acceptance probability is trivial. The parameter σ_ν^2 can be automatically 'tuned' to obtain a desired acceptance rate. A suitable initial value for the pair (γ, ν) are the approximate MLEs which can be found crudely by treating the data as independent and identically t-distributed and maximising the log-likelihood numerically.

A small Monte Carlo study was conducted, for which we generated 50 t-distributed ARFIMA$(0, d, 0)$ time series with $\nu_I = 5$ and $d_I \in \{-0.45, -0.35, \ldots, 0.45\}$. Two different priors were used, setting λ in (16) to be either 0.1 or 0.2. The interesting summary statistics are presented in table 2.

Choosing between the two priors suggested, or indeed any other prior, is of course up to the modeller. However one fact that may influence this choice is that, when erroneously applied to Gaussian data, the prior with $\lambda = 0.2$ tends to produce point estimates for ν that are less than 30, whilst $\lambda = 0.1$ leads to most being larger than 30. Since a basic rule-of-thumb is that, for $\nu > 30$, the t-distribution is practically indistinguishable from the Gaussian, this might suggest that the prior with $\lambda = 0.1$ might be more useful. It should be noted however that this analysis would be sensitive to the length of time series n.

3.4. Comparison with other estimators

In [40, 45, 41, 14] various estimator methods such as the Variable Bandwidth method [40], wavelets [14], Rescaled range (R/S) [22], Detrended Fluctuation Analysis (DFA) [39], the Whittle estimator [43, 42] and a semi-parametric power spectral method [15] have been used for estimating the LRD parameter d in an ARFIMA(1,d,1) model with α-stable innovations. Comparison with our results shows that the d parameter is well estimated with relative small uncertainty bounds compared with the classical estimators [14].

4. Application to Solar X ray data

The GOES geostationary meteorological satellites have been used to observe X rays from solar flares over several decades starting in the mid 1970s. Burnecki et al. [6], Stanislavsky et al. [44] fitted an ARFIMA model with α-stable innovations to a time series comprising daily aggregates of solar flare events derived from GOES data, and inferred the H, d and α values using the finite impulse response (FIRT), variance of residuals (VaR, or DFA), and McCulloch quantile methods. For this interval [6] found d to be 0.21 and α to be 1.2674

Unfortunately the flare series studied by the previous authors was not available at the time of writing from the original public National Oceanic and Atmospheric Administration (NOAA) archive sites. Instead we have obtained the full, 1 minute resolution GOES solar X-ray irradiance (in W/m^2) in the 0.1-0.8 nanometre long wavelength channel for the period that [6] identified, and we show its daily mean values in figure 5.

For Solar Cycle 23 the primary and secondary science satellites were GOES 8, 10 and 12, and a correction factor of 0.7 was applied as per the recommendations at the NOAA archive. The data is archived at `http://satdat.ngdc.noaa.gov/sem/goes/data/new_avg/`

We note that direct comparison with [6] is thus not possible, as our full X ray series includes the sharp rises due to the onset of the flares, the decay of flares, and the background X ray flux, and so our inferred ARFIMA model should be seen as describing the daily mean of this rather than the daily aggregated flares. For the daily mean data we find posterior mean estimates of α, d to be 1.65 and 0.205.

For comparison we have used the maximum likelihood method implemented in MATLAB to find a α value of 1.16, and from a power-spectral estimator [15] we find $d = 0.37$. The differences to the Bayesian estimator could be due to our estimator doing a joint estimate and/or they could be due to finite time series length. However, our earlier tests using simulated surrogate data gave very good results. Further work will also be needed to compare in more detail the high resolution X ray time series with the derived flare time series studied by [6], but our results suggest that an α-stable ARFIMA model is indeed appropriate and useful for this higher resolution dataset.

5. Conclusions

Self-similarity is by now well known and well studied, and has found many applications in physics and elsewhere in the sciences of complexity. However, as we discussed, although expressed by a single exponent, H, self-similarity can arise both from long-range dependence and heavy-tailed jumps respectively, thus giving two potential contributions to the exponent. In consequence there is a need to simultaneously estimate both the long-range dependence and heavy-tail distribution parameters, d and α. Although best statistical practice allows joint estimation by some frequentist approaches, the estimation is still sometimes done in the science literature by measuring H and one of d or α. In this paper we presented a novel Bayesian method to directly infer d and α on the hypothesis of an ARFIMA model with heavy tailed innovations. Our method is flexible enough to allow the choice of heavy-tailed distribution (e.g. α- or t-distributed), and we gave a demonstration of its effectiveness and accuracy on synthetic data and solar X-ray data.

6. Acknowledgements

CF was supported by the German Research Foundation (DFG) through the cluster of excellence CliSAP (EXC177), the SFB/TRR181 and grant FR3515/3-1. CF and NWW acknowledge travel funding from the Norwegian Research Council KLIMAFORSK project no 229754. NWW acknowledges travel support from the London Mathematical Laboratory, Office of Naval Research grant NICOP-N62909-15-1-N143 at Warwick and Potsdam, and useful discussions with Sandra Chapman. ET is supported by an STFC student ship.

References

[1] Beran, J. (1994). *Statistics for Long Memory Processes.* Chapman & Hall, New York.

[2] Bergström, H. (1952). On some expansions of stable distribution functions. *Arkiv för Matematik*, 2(4), 375–378.

[3] Bouchaud, J.-P. & Potters, M. (2003). *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management.* Cambridge University Press.

[4] Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439, 462–465.

[5] Buckle, D. J. (1995). Bayesian inference for stable distributions. *Journal of the American Statistical Association*, 90(430), 605–613.

[6] Burnecki, K., Klafter, J., Magdziarz, M., & Weron, A. (2008). From solar flare time series to fractional dynamics. *Physica A: Statistical Mechanics and its Applications*, 387(5), 1077–1087.

[7] Burnecki, K. & Sikora, G. (2013). Estimation of farima parameters in the case of negative memory and stable noise. *IEEE Transactions on Signal Processing*, 61(11), 2825–2835.

[8] Burnecki, K. & Weron, A. (2014). Algorithms for testing of fractional dynamics: a practical guide to arfima modelling. *Journal of Statistical Mechanics: Theory and Experiment*, (pp. P10036).

[9] Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), 327–335.

[10] DuMouchel, W. H. (1983). Estimating the stable index α in order to measure tail thickness: A critique. *The Annals of Statistics*, 11(4), 1019–1031.

[11] Embrechts, P. & Maejima, M. (2002). *Selfsimilar Processes.* Princeton University Press.

[12] Feller, W. (1957). *An Introduction to Probability Theory and Its Applications, 3rd ed.*, volume 1. Wiley, New York.

[13] Fox, R. & Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, (pp. 517–532).

[14] Franzke, C. L. E., Graves, T., Watkins, N. W., Gramacy, R. B., & Hughes, C. (2012). Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity. *Philosophical Transactions of the Royal Society A*, 370(1962), 1250–1267.

[15] Geweke, J. & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4), 221–238.

[16] Graves, T. (2013). *A systematic approach to Bayesian inference for long memory processes*. PhD thesis, University of Cambridge, UK.

[17] Graves, T., Franzke, C. L. E., Watkins, N. W., & Gramacy, R. B. (2015). Efficient Bayesian inference for long memory processes. *Nonlin. Proc. Geophys.*, 22, 679–700.

[18] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732.

[19] Gregory, P. (2005). *Bayesian Data Analysis for the Physical Sciences*. Cambridge University Press.

[20] Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probab.*, (pp. 130–145).

[21] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5), 1163–1174.

[22] Hurst, H. E., Black, R. P., & Simaika, Y. M. (1965). *Long-Term Storage: An Experimental Study*. Constable, London.

[23] Kokoszka, P. S. & Taqqu, M. S. (1996). Parameter estimation for infinite variance fractional arima. *The Annals of Statistics*, 24(5), 1880–1913.

[24] Laskin, N., Lambadatis, I., Harmantzis, F. C., & Devetsikiotis, M. (2002). Fractional levy motion and its application to network traffic modeling. *Computer Networks*, 40, 363–275.

[25] Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), 394–419. Correction: Mandelbrot [26].

[26] Mandelbrot, B. B. (1972). Correction of an error in Mandelbrot [25]. *The Journal of Business*, 45(4), 542–543.

[27] Mandelbrot, B. B. (2002). *Gaussian Self-Affinity and Fractals: Globality, The Earth, $1/f$ Noise, and R/S*, volume H of *Selecta*. Springer; New York.

[28] Mandelbrot, B. B. & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4), 422–437.

[29] Mandelbrot, B. B. & Wallis, J. R. (1968). Noah, Joseph and operational hydrology. *Water Resources Research*, 4(5), 909–918.

[30] Mandelbrot, B. B. & Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5(5), 967–988.

[31] Mantegna, R. & Stanley, H. E. (2000). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press.

[32] McCulloch, J. H. (1997). Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business & Economic Statistics*, 15(1), 74–81.

[33] Menn, C. & Rachev, S. T. (2006). Calibrated FFT-based density approximations for α-stable distributions. *Computational Statistics and Data Analysis*, 50(8), 1891–1904.

[34] Mercik, S., Weron, K., Burnecki, K., & Weron, A. (2003). Enigma of self-similarity of fractional Lévy stable motions. *Acta Physica Polonica B*, 34(7), 3773–3791.

[35] Mikosch, T., Gadrich, T., Kluppelberg, C., & Adler, R. J. (1995). Parameter estimation for arma models with infinite variance innovations. *The Annals of Statistics*, (pp. 305–326).

[36] Mittnik, S., Doganoglu, T., & Chenyao, D. (1999). Computing the probability density function of the stable Paretian distribution. *Mathematical and Computer Modelling*, 29(10–12), 235–240.

[37] Nolan, J. P. (1998). Parameterizations and modes of stable distributions. *Statistics and Probability Letters*, 38(2), 187–195.

[38] Painter, S. & Paterson, L. (1994). Fractional Lévy motion as a model for spatial variability in sedimentary rock. *Geophysical Research Letters*, 21(25), 2857–2860.

[39] Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2), 1685–1689.

[40] Schmittbuhl, J., Vilotte, J.-P., & Roux, S. (1995). Reliability of self-affine measurements. *Physical Review E*, 51(1), 131.

[41] Sheng, H., Chen, Y., & Qiu, T. (2011). On the robustness of hurst estimators. *IET Signal Processing*, 5(2), 209–225.

[42] Shimotsu, K. & Phillips, P. C. (2006). Local whittle estimation of fractional integration and some of its variants. *Journal of Econometrics*, 130(2), 209–233.

[43] Shimotsu, K., Phillips, P. C., et al. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics*, 33(4), 1890–1933.

[44] Stanislavsky, A. A., Burnecki, K., Magdziarz, M., Weron, A., & Weron, K. (2095). Farima modeling of solar flare activity from empirical time series of soft x-ray solar emission. *Astrophysical Journal*, 693(2), 1877–1882.

[45] Stoev, S. & Taqqu, M. S. (2005). Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations. *Journal of Time Series Analysis*, 26(2), 211–249.

[46] Taqqu, M. S. & Teverovsky, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance time series. *A practical guide to heavy tails: statistical techniques and applications*, 177, 218.

[47] Watkins, N. W., Credgington, D., Hnat, B., Chapman, S. C., Freeman, M. P., & Greenhough, J. (2005). Towards synthesis of solar wind and geomagnetic scaling exponents: A fractional Lévy motion model. *Space Science Reviews*, 121(1-4), 271–284.

[48] Waxman, D. & Feng, J. (2005). Implications of long tails in the distribution of mutant effects. *Physica D*, 206(3), 265–274.

[49] Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*, volume 65 of *Translations of Mathematical Monographs*. American Mathematical Society.

Table 1: Posterior summary statistics for asymmetric-stable ARFIMA$(0, d, 0)$ process. Average of 20 runs.

	mean	std	95% CI endpoints	
d_R	0.002	0.008	-0.014	0.018
α	1.481	0.044	1.398	1.570
β	0.483	0.084	0.316	0.643

Table 2: Comparison of posterior summary statistics for t_5-distributed ARFIMA$(0,d,0)$ process using two different priors. Average of 50 runs.

		mean	std	95% CI endpoints	
$\lambda = 0.1$	d_R	0.003	0.022	-0.039	0.046
$\lambda = 0.2$	d_R	0.003	0.022	-0.039	0.046
$\lambda = 0.1$	γ	1.005	0.039	0.931	1.080
$\lambda = 0.2$	γ	1.002	0.038	0.929	1.077
$\lambda = 0.1$	ν	5.665	1.061	3.857	7.760
$\lambda = 0.2$	ν	5.543	0.987	3.850	7.511

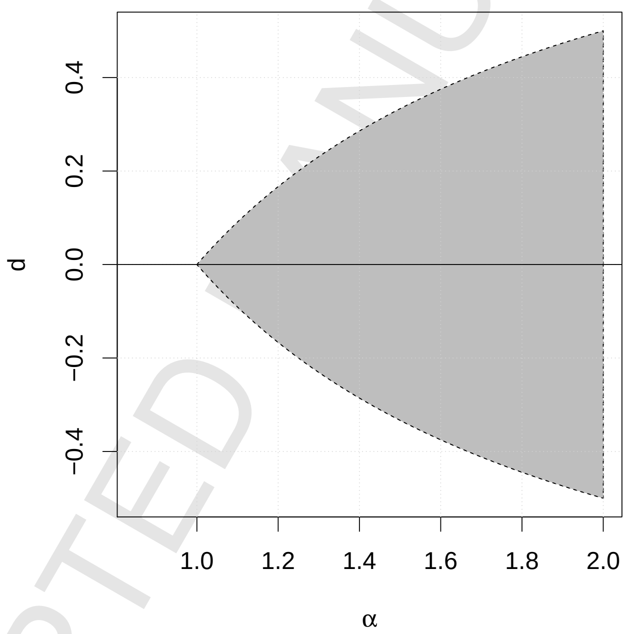
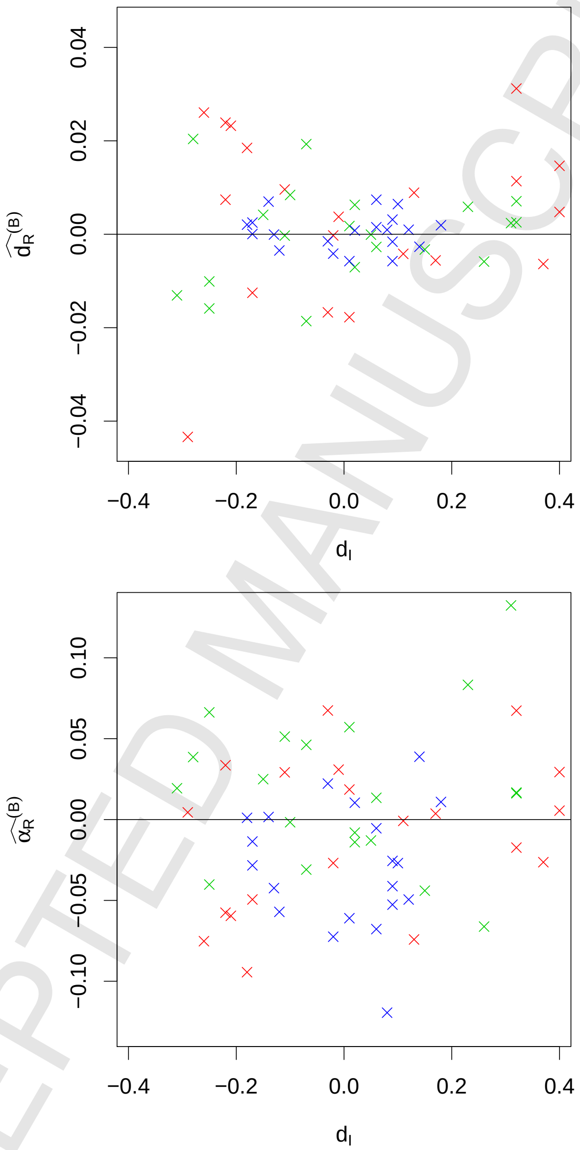


Figure 1: Set of allowable values for (α, d); does not include the dotted boundary.



(a)
(b)

Figure 2: Posterior outputs from α-stable ARFIMA$(0, d, 0)$ series; (a) $\widehat{d_R}^{(B)}$ against d_I, (b) $\widehat{\alpha_R}^{(B)}$ against d_I. Red is used for $\alpha_I = 1.75$, green for $\alpha_I = 1.50$ and blue for $\alpha_I = 1.25$.

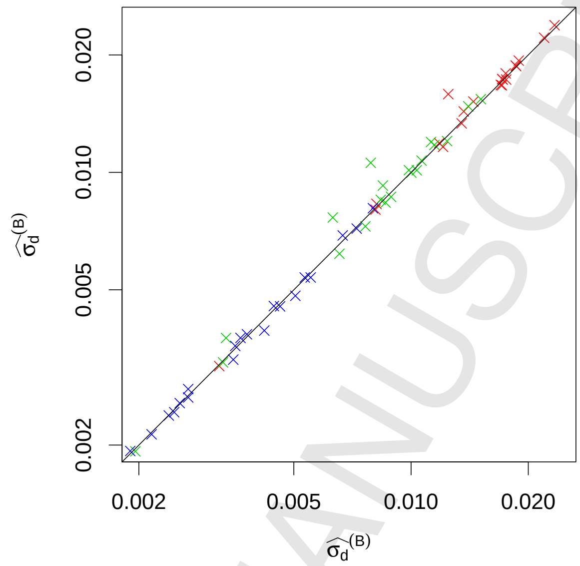


Figure 3: Posterior outputs; log-scale comparison of $\widehat{\sigma_d}^{(B)}$ when α is assumed unknown (y-axis) against unknown (x-axis). Red is used for $\alpha_I = 1.75$, green for $\alpha_I = 1.50$ and blue for $\alpha_I = 1.25$.

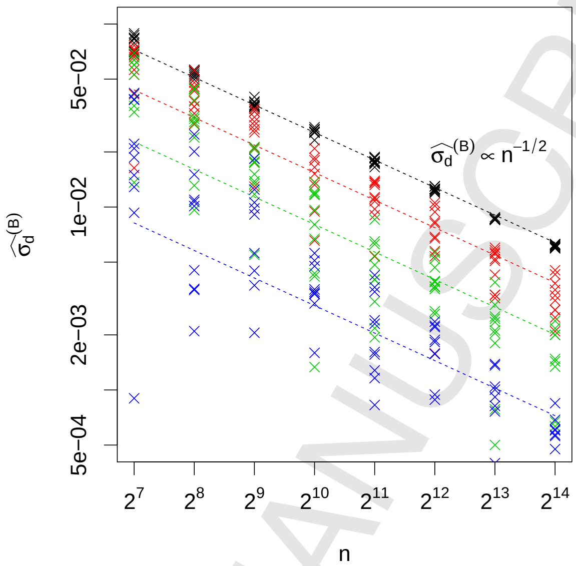


Figure 4: Posterior outputs from ARFIMA$(0,0,0)$ series; (a) $\widehat{\sigma_d}^{(B)}$ against n. Black is used for $\alpha_I = 2$, red for $\alpha_I = 1.75$, green for $\alpha_I = 1.5$ and blue for $\alpha_I = 1.25$. (log-log scale).

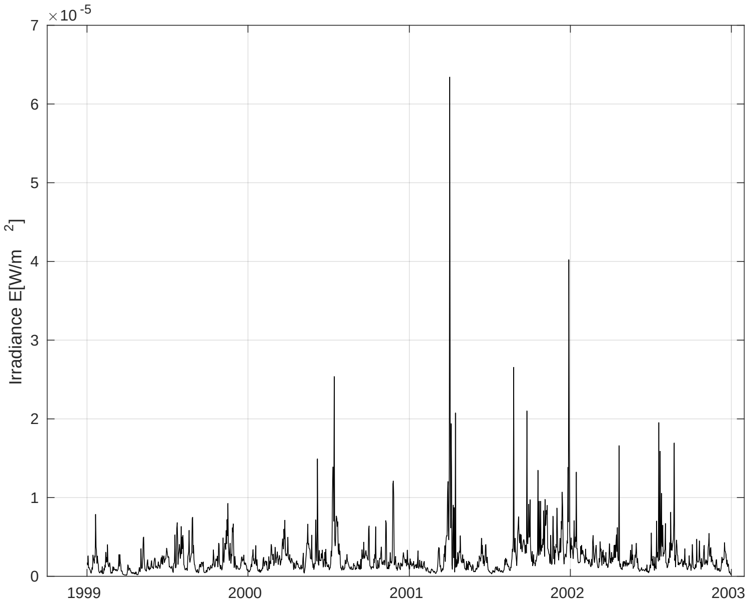


Figure 5: GOES time series.