Nonlin. Processes Geophys., 22, 679–700, 2015
www.nonlin-processes-geophys.net/22/679/2015/
doi:10.5194/npg-22-679-2015

Nonlinear Processes in Geophysics
Open Access
EGU

Efficient Bayesian inference for natural time series using ARFIMA processes

T. Graves[1], R. B. Gramacy[2], C. L. E. Franzke[3], and N. W. Watkins[4,5,6,7]

[1]URS Corporation, London, UK
[2]The University of Chicago, Booth School of Business, Chicago, IL, USA
[3]Meteorological Institute and Center for Earth System Research and Sustainability (CEN), University of Hamburg, Hamburg, Germany
[4]Centre for the Analysis of Time Series, London School of Economics and Political Science, London, UK
[5]Centre for Fusion Space and Astrophysics, University of Warwick, Coventry, UK
[6]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
[7]Faculty of Mathematics, Computing and Technology, Open University, Milton Keynes, UK

Correspondence to: C. L. E. Franzke (christian.franzke@uni-hamburg.de)

Received: 9 February 2015 – Published in Nonlin. Processes Geophys. Discuss.: 27 March 2015
Revised: 23 September 2015 – Accepted: 5 November 2015 – Published: 18 November 2015

Abstract. Many geophysical quantities, such as atmospheric temperature, water levels in rivers, and wind speeds, have shown evidence of long memory (LM). LM implies that these quantities experience non-trivial temporal memory, which potentially not only enhances their predictability, but also hampers the detection of externally forced trends. Thus, it is important to reliably identify whether or not a system exhibits LM. In this paper we present a modern and systematic approach to the inference of LM. We use the flexible autoregressive fractional integrated moving average (ARFIMA) model, which is widely used in time series analysis, and of increasing interest in climate science. Unlike most previous work on the inference of LM, which is frequentist in nature, we provide a systematic treatment of Bayesian inference. In particular, we provide a new approximate likelihood for efficient parameter inference, and show how nuisance parameters (e.g., short-memory effects) can be integrated over in order to focus on long-memory parameters and hypothesis testing more directly. We illustrate our new methodology on the Nile water level data and the central England temperature (CET) time series, with favorable comparison to the standard estimators. For CET we also extend our method to seasonal long memory.

1 Introduction

Many natural processes are sufficiently complex that a stochastic model is essential, or at the very least an efficient description (Watkins, 2013). Such a process will be specified by several properties, of which a particularly important one is the degree of memory in a time series, often expressed through a characteristic autocorrelation time over which fluctuations will decay in magnitude. In this paper, however, we are concerned with specific types of stochastic processes that are capable of possessing long memory (LM) (Beran, 1994a; Palma, 2007; Beran et al., 2013). Long memory is the notion of there being correlation between the present and *all* points in the past. A standard definition (Beran, 1994a; Palma, 2007; Beran et al., 2013) is that a (finite variance, stationary) process has *long memory* if its autocorrelation function (ACF) has power-law decay: $\rho(\cdot)$ such that $\rho(k) \sim c_\rho k^{2d-1}$ as $k \to \infty$, for some non-zero constant c_ρ, and where $0 < d < \frac{1}{2}$. The parameter d is the memory parameter; if $d = 0$ the process does not exhibit long memory, while if $-\frac{1}{2} < d < 0$ the process is said to be anti-persistent.

The asymptotic power-law form of the ACF corresponds to an absence of a characteristic decay timescale, in striking contrast to many standard (stationary) stochastic processes where the effect of each data point decays so fast that it rapidly becomes indistinguishable from noise. An exam-

ple of the latter is the exponential ACF, where the *e*-folding timescale sets a characteristic correlation time. The study of processes that *do* possess long memory is important because they exhibit unusual properties, because many familiar mathematical results fail to hold, and because of the numerous examples of data sets where LM is seen.

The study of long memory originated in the 1950s in the field of hydrology, where studies of the levels of the Nile (Hurst, 1951) demonstrated anomalously fast growth of the rescaled range of the time series. After protracted debates[1] about whether this was a transient (finite time) effect, the mathematical pioneer Benoît B. Mandelbrot showed that if one retained the assumption of stationarity, novel mathematics would then be essential to sufficiently explain the Hurst effect. In doing so he rigorously defined the concept of long memory (Mandelbrot and Van Ness, 1968; Mandelbrot and Wallis, 1968).

Most research into long memory and its properties has been based on classical statistical methods, spanning parametric, semi-parametric, and non-parametric modeling (see Beran et al., 2013, for a review). Very few Bayesian methods have been studied, most probably due to computational difficulties. The earliest works are parametric and include Koop et al. (1997), Pai and Ravishanker (1998), and Hsu and Breidt (2003). If computational challenges could be mitigated, the Bayesian paradigm would offer advantages over classical methods including flexibility in specification of priors (i.e., physical expertise could be used to elicit an informative prior). It would offer the ability to marginalize out aspects of a model apparatus and data, such as short-memory or seasonal effects and missing observations, so that statements about long-memory effects can be made unconditionally.

Towards easing the computational burden, we focus on the autoregressive fractional integrated moving average (ARFIMA) class of processes (Granger and Joyeux, 1980; Hosking, 1981) as the basis of developing a systematic and unifying Bayesian framework for modeling a variety of common time series phenomena, with particular emphasis on (marginally) detecting potential long-memory effects (i.e., averaging over short-memory and seasonal effects). ARFIMA has become very popular in statistics and econometrics because it is generalizable and its connection to the autoregressive moving average (ARMA) family and to fractional Gaussian noise is relatively transparent. A key property of ARFIMA is its ability to simultaneously yet separately model long and short memory.

Here we present a Bayesian framework for the efficient and systematic estimation of the ARFIMA parameters. We provide a new approximate likelihood for ARFIMA processes that can be computed quickly for repeated evaluation on large time series, and which underpins an efficient Markov chain Monte Carlo (MCMC) scheme for Bayesian inference. Our sampling scheme can be best described as a modernization of a blocked MCMC scheme proposed by Pai and Ravishanker (1998) – adapting it to the approximate likelihood and extending it to handle a richer form of (known) short-memory effects. We then further extend the analysis to the case where the short-memory form is unknown, which requires trans-dimensional MCMC, in which the model order (the p and q parameters in the ARFIMA model) varies and, thus, so does the dimension of the problem. This aspect is similar to the work of Ehlers and Brooks (2008), who considered the simpler autoregressive-integrated moving average (ARIMA) model class, and to Holan et al. (2009), who worked with a non-parametric long-memory process. Our contribution has aspects in common with Eğri˙oğlu and Günay (2010), who presented a more limited method focused on model selection rather than averaging. The advantage of averaging is that the unknown form of short-memory effects can be integrated out, focusing on long memory without conditioning on nuisance parameters.

The aim of this paper is to introduce an efficient Bayesian algorithm for the inference of the parameters of the ARFIMA(p,d,q) model, with particular emphasis on the LM parameter d. 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, e.g., α stable or t distribution (to be published in a companion paper). The remainder of the paper is organized as follows. Section 2 discusses the important numerical calculation of likelihoods, representing a hybrid between earlier classical statistical methods and our new contributions towards a full-Bayesian approach. Section 3 describes our proposed Bayesian framework and methodology in detail, focusing on long memory only. Then, in Sect. 4, we consider extensions for additional short memory and the computational techniques required to integrate them out. Empirical illustration and comparison of all methods is provided in Sect. 5 via synthetic and real data including the Nile water level data and the central England temperature (CET) time series, with favorable comparison to the standard estimators. In the case of the Nile data, we find strong evidence for long memory. The CET analysis requires a slight extension to handle *seasonal* long memory, and we find that the situation here is more nuanced in terms of evidence for long memory. The paper concludes with a discussion in Sect. 7 focused on the potential for further extension.

2 Likelihood evaluation for Bayesian inference

2.1 ARFIMA model

We provide here a brief review of the ARFIMA model. More details are given in Appendix A.

[1] For a detailed exposition of this period of mathematical history, see Graves et al. (2014).

An ARFIMA model is given by

$$\Phi(\mathcal{B})(1-\mathcal{B})^d X_t = \Theta(\mathcal{B})\varepsilon_t. \quad (1)$$

We define the backshift operator $\mathcal{B}$, where $\mathcal{B}X_t = X_{t-1}$, and powers of $\mathcal{B}$ are defined iteratively: $\mathcal{B}^k X_t = \mathcal{B}^{k-1} X_{t-1} = \cdots = X_{t-k}$. Φ is the autoregressive component and Θ is the moving average component and constitute the short-memory components of the ARFIMA model. These are defined in more detail in Appendix A and in Graves (2013).

2.2 Likelihood function

For now, we restrict our attention to a Bayesian analysis of an ARFIMA(0,d,0) process, having no short-ranged ARMA components ($p = q = 0$), placing emphasis squarely on the memory parameter d. As we explain in our Appendix, the resulting process is identical to a *fractionally integrated* processes with memory parameter d.

Here we develop an efficient and new scheme for evaluating the (log) likelihood, via approximation. Throughout, the reader should suppose that we have observed the vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ as a realization of a stationary, causal and invertible ARFIMA(0,d,0) process $\{X_t\}$ with mean $\mu \in \mathbb{R}$. The innovations will be assumed to be independent, and taken from a zero-mean *location-scale probability density* $f(\cdot; 0, \sigma, \boldsymbol{\lambda})$, which means the density can be written as $f(x; \delta, \sigma, \boldsymbol{\lambda}) \equiv \frac{1}{\sigma} f(\frac{x-\delta}{\sigma}; 0, 1, \boldsymbol{\lambda})$. The parameters δ and σ are called the location and scale parameters, respectively. The m-dimensional $\boldsymbol{\lambda}$ is a shape parameter (if it exists, i.e., $m > 0$). A common example is the Gaussian $\mathcal{N}(\mu, \sigma^2)$, where $\delta \equiv \mu$ and there is no $\boldsymbol{\lambda}$. We classify the four parameters μ, σ, $\boldsymbol{\lambda}$, and d into three distinct classes: (1) the mean of process, μ; (2) innovation distribution parameters, $\boldsymbol{\upsilon} = (\sigma, \boldsymbol{\lambda})$; and (3) memory structure, d. Together, $\boldsymbol{\psi} = (\mu, \boldsymbol{\upsilon}, \boldsymbol{\omega})$, where $\boldsymbol{\omega}$ will later encompass the short-range parameters p and q.

Our proposed likelihood approximation uses a truncated autoregressive model (AR) (∞) approximation (cf. Haslett and Raftery, 1989). We first re-write the AR(∞) approximation of ARFIMA(0,d,0) to incorporate the unknown parameter μ, and drop the (d) superscript for convenience: $X_t - \mu = \varepsilon_t - \sum_{k=1}^{\infty} \pi_k (X_{t-k} - \mu)$. Then we truncate this AR(∞) representation to obtain an AR(P) one, with P large enough to retain low frequency effects, e.g., $P = n$. We denote $\Pi_P = \sum_{k=0}^{P} \pi_k$ and, with $\pi_0 = 1$, rearrange terms to obtain the following modified model:

$$X_t = \varepsilon_t + \Pi_P \mu - \sum_{k=1}^{P} \pi_k X_{t-k}. \quad (2)$$

It is now possible to write down a *conditional* likelihood. For convenience the notation $\boldsymbol{x}_k = (x_1, \ldots, x_k)^\top$ for $k = 1, \ldots, n$ will be used (and $\boldsymbol{x}_0$ is interpreted as appropriate where necessary). Denote the unobserved $\boldsymbol{P}$ vector of random variables $(x_{1-P}, \ldots, x_{-1}, x_0)^\top$ by $\boldsymbol{x}_\mathrm{A}$ (in the Bayesian context these will be auxiliary, hence "A"). Consider the likelihood $L(\boldsymbol{x}|\boldsymbol{\psi})$ as a joint density, which can be factorized as a product of conditionals. Writing $g_t(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{\psi})$ for the density of X_t conditional on $\boldsymbol{x}_{t-1}$, we obtain $L(\boldsymbol{x}|\boldsymbol{\psi}) = \prod_{t=1}^{n} g_t(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{\psi})$.

This is still of little use because the g_t may have a complicated form. However, by further conditioning on $\boldsymbol{x}_\mathrm{A}$, and writing $h_t(x_t|\boldsymbol{x}_\mathrm{A}, \boldsymbol{x}_{t-1}, \boldsymbol{\psi})$ for the density of X_t conditional on $\boldsymbol{x}_{t-1}$ *and* $\boldsymbol{x}_\mathrm{A}$, we obtain $L(\boldsymbol{x}|\boldsymbol{\psi}, \boldsymbol{x}_\mathrm{A}) = \prod_{t=1}^{n} h_t(x_t|\boldsymbol{x}_\mathrm{A}, \boldsymbol{x}_{t-1}, \boldsymbol{\psi})$. Returning to Eq. (2) observe that, conditional on both the observed and *un*observed past values, X_t is simply distributed according to the innovations' density f with a suitable change in location: $X_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_\mathrm{A} \sim f(\cdot; [\Pi_P \mu - \sum_{k=1}^{P} \pi_k x_{t-k}], \sigma, \boldsymbol{\lambda})$. Then using location-scale representation:

$$\begin{aligned} h_t(x_t|\boldsymbol{x}_\mathrm{A}, \boldsymbol{x}_{t-1}, \boldsymbol{\psi}) &\approx f\left(x_t; \left[\Pi_P \mu - \sum_{k=1}^{P} \pi_k x_{t-k}\right], \sigma, \boldsymbol{\lambda}\right) \\ &\equiv \frac{1}{\sigma} f\left(\frac{c_t - \Pi_P \mu}{\sigma}; 0, 1, \boldsymbol{\lambda}\right), \\ &\text{where } c_t = \sum_{k=0}^{P} \pi_k x_{t-k},\ t = 1, \ldots, n. \quad (3) \end{aligned}$$

Therefore, $L(\boldsymbol{x}|\boldsymbol{\psi}, \boldsymbol{x}_\mathrm{A}) \approx \sigma^{-n} \prod_{t=1}^{n} f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \boldsymbol{\lambda}\right)$, or equivalently

$$\ell(\boldsymbol{x}|\boldsymbol{\psi}, \boldsymbol{x}_\mathrm{A}) \approx -n \log \sigma + \sum_{t=1}^{n} \log\left\{ f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \boldsymbol{\lambda}\right)\right\}. \quad (4)$$

Evaluating this expression efficiently depends upon efficient calculation of $\mathbf{c} = (c_1, \ldots, c_n)^t$ and $\log f(\cdot)$. From Eq. (3), $\mathbf{c}$ is a convolution of the augmented data, $(\boldsymbol{x}, \boldsymbol{x}_\mathrm{A})$, and coefficients depending on d, which can be evaluated quickly in the R language for statistical computing via `convolve` via fast Fourier transform (FFT). Consequently, evaluation of the *conditional* on $\boldsymbol{x}_A$ likelihood in the Gaussian case costs only $\mathcal{O}(n \log n)$ – a clear improvement over the exact method. Obtaining the unconditional likelihood requires marginalization over $\boldsymbol{x}_\mathrm{A}$, which is analytically infeasible. However, this conditional form will suffice in the context of our Bayesian inferential scheme, presented below.

3 A Bayesian approach to long-memory inference

We are now ready to consider Bayesian inference for ARFIMA(0,d,0) processes. Our method can be succinctly described as a modernization of the blocked MCMC method of Pai and Ravishanker (1998). Isolating parameters by blocking provides significant scope for modularization,

which helps to accommodate our extensions for short memory. Pairing with efficient likelihood evaluations allows much longer time series to be entertained than ever before. Our description begins with the appropriate specification of priors, which are more general than previous choices, yet still encourages tractable inference. We then provide the relevant updating calculations for all parameters, including those for auxiliary parameters $\boldsymbol{x}_{\mathrm{A}}$.

We follow earlier work (Koop et al., 1997; Pai and Ravishanker, 1998) and assume a priori independence for components of ψ. Each component will leverage familiar prior forms with diffuse versions as limiting cases. Specifically, we use a diffuse Gaussian prior on μ: $\mu \sim \mathcal{N}(\mu_0,\ \sigma_0^2)$, with σ_0 large. The improper flat prior is obtained as the limiting distribution when $\sigma_0 \to \infty$: $p_\mu(\mu) \propto 1$. We place a gamma prior on the precision $\tau = \sigma^{-2}$ implying a *root-inverse gamma* distribution $\mathcal{R}(\alpha_0,\ \beta_0)$ for σ, with density $f(\sigma) = \frac{2}{\Gamma(\alpha)} \beta_0{}^{\alpha_0}\, \sigma^{-(2\alpha_0+1)} \exp\left(-\frac{\beta_0}{y^2}\right),\ \sigma > 0$. A diffuse/improper prior is obtained as the limiting distribution when $\alpha_0,\ \beta_0 \to 0$: $p_\sigma(\sigma) \propto \sigma^{-1}$, which, in the asymptotic limit, is equivalent to a log uniform prior. Finally, we specify $d \sim \mathcal{U}(-\frac{1}{2},\ \frac{1}{2})$.

Updating μ: following Pai and Ravishanker (1998), we use a symmetric random walk (RW) Metropolis–Hastings (MH) update with proposals $\xi_\mu \sim \mathcal{N}(\mu,\ \sigma_\mu^2)$, for some σ_μ^2. The acceptance ratio is

$$A_\mu\left(\mu, \xi_\mu\right) = \sum_{t=1}^{n} \log\left\{ f\left(\frac{c_t - \Pi_P \xi_\mu}{\sigma}; \boldsymbol{\lambda}\right)\right\} - \sum_{t=1}^{n} \log\left\{ f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \boldsymbol{\lambda}\right)\right\} + \log\left[\frac{p_\mu\left(\xi_\mu\right)}{p_\mu(\mu)}\right] \tag{5}$$

under the approximate likelihood.

Updating σ: we diverge from Pai and Ravishanker (1998) here, who suggest independent MH with moment-matched inverse gamma proposals, finding poor performance under poor moment estimates. We instead prefer a RW MH approach, which we conduct in log space since the domain is $\mathbb{R}^+$. Specifically, we set: $\log \xi_\sigma = \log \sigma + \upsilon$, where $\upsilon \sim \mathcal{N}(0,\ \sigma_\sigma^2)$ for some σ_σ^2. $\xi_\sigma | \sigma$ is log-normal and we obtain $\frac{q(\sigma;\xi_\sigma)}{q(\xi_\sigma;\sigma)} = \frac{\xi_\sigma}{\sigma}$. Recalling Eq. (5), the MH acceptance ratio under the approximate likelihood is

$$A_\sigma\left(\sigma, \xi_\sigma\right) = \sum_{t=1}^{n} \log\left\{ f\left(\frac{c_t - \Pi_P \mu}{\xi_\sigma}; \boldsymbol{\lambda}\right)\right\} - \sum_{t=1}^{n} \log\left\{ f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \boldsymbol{\lambda}\right)\right\} + \log\left[\frac{p_\sigma\left(\xi_\sigma\right)}{p_\sigma(\sigma)}\right] + (n-1)\log\left[\frac{\sigma}{\xi_\sigma}\right].$$

The MH algorithm, applied alternately in a Metropolis-within-Gibbs fashion to the parameters μ and σ, works well. However, *actual* Gibbs sampling is an efficient alternative in this two-parameter case (i.e., for known d; see Graves, 2013).

Update of d: updating the memory parameter d is far less straightforward than either μ or σ. Regardless of the innovations' distribution, the conditional posterior $\pi_{d|\psi_{-d}}(d|\boldsymbol{\psi}_{-d}, \boldsymbol{x})$ is not amenable to Gibbs sampling. We use RW proposals from truncated Gaussian $\xi_d \sim \mathcal{N}^{(a,b)}(d,\ \sigma_d^2)$, with density

$$f(x; \mu, \sigma, a, b) = \frac{1}{\sigma} \frac{\phi^{(\mathcal{N})}[(x-\mu)/\sigma]}{\Phi^{(\mathcal{N})}[(b-\mu)/\sigma] - \Phi^{(\mathcal{N})}[(a-\mu)/\sigma]}, \quad a < x < b, \tag{6}$$

where $\Phi^{(\mathcal{N})}$ and $\phi^{(\mathcal{N})}$ are the standard normal cumulative density function (CDF) and probability density function (PDF), respectively. In particular, we use $\xi_d | d \sim \mathcal{N}^{(-1/2,1/2)}(d,\ \sigma_d^2)$ via rejection sampling from $\mathcal{N}(d,\ \sigma_d^2)$ until $\xi_d, \in (-\frac{1}{2},\ \frac{1}{2})$. Although this may seem inefficient, it is perfectly acceptable; for example, if $\sigma_d = 0.5$ the expected number of required variates is still less than 2, regardless of d. More refined methods of directly sampling from truncated normal distributions exist – see for example Robert (1995) – but we find little added benefit in our context.

A useful cancellation in $q(d;\ \xi_d)/q(\xi_d;\ d)$ obtained from Eq. (6) yields

$$A_d = \ell\left(\boldsymbol{x} | \xi_d, \boldsymbol{\psi}_{-d}\right) - \ell\left(\boldsymbol{x} | d, \boldsymbol{\psi}_{-d}\right) + \log\left[\frac{p_d\left(\xi_d\right)}{p_d(d)}\right] + \log\left\{\frac{\Phi^{(\mathcal{N})}\left[\left(\frac{1}{2} - d\right)/\sigma_d\right] - \Phi^{(\mathcal{N})}\left[\left(-\frac{1}{2} - d\right)/\sigma_d\right]}{\Phi^{(\mathcal{N})}\left[\left(\frac{1}{2} - \xi_d\right)/\sigma_d\right] - \Phi^{(\mathcal{N})}\left[\left(-\frac{1}{2} - \xi_d\right)/\sigma_d\right]}\right\}.$$

Denote $\xi_{c_t} = \sum_{k=0}^{P} \xi_{\pi_k} x_{t-k}$ for $t = 1,\ \ldots,\ n$, where $\{\xi_{\pi_k}\}$ are the proposed coefficients $\{\pi_k^{(\xi_d)}\}$; $\pi_k^{(d)} = \frac{1}{\Gamma(k+1)} \frac{\Gamma(k-d)}{\Gamma(-d)}$; furthermore, $\xi_{\Pi_P} - \sum_{k=0}^{P} \xi_{\pi_k}$. Then in the approximate case

$$A_d = \sum_{t=1}^{n} \log\left\{ f\left(\frac{\xi_{c_t} - \xi_{\Pi_P \mu}}{\sigma}; \boldsymbol{\lambda}\right)\right\} - \sum_{t=1}^{n} \log\left\{ f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \boldsymbol{\lambda}\right)\right\} + \log\left[\frac{p_d\left(\xi_d\right)}{p_d(d)}\right] + \log\left\{\frac{\Phi^{(\mathcal{N})}\left[\left(\frac{1}{2} - d\right)/\sigma_d\right] - \Phi^{(\mathcal{N})}\left[\left(-\frac{1}{2} - d\right)/\sigma_d\right]}{\Phi^{(\mathcal{N})}\left[\left(\frac{1}{2} - \xi_d\right)/\sigma_d\right] - \Phi^{(\mathcal{N})}\left[\left(-\frac{1}{2} - \xi_d\right)/\sigma_d\right]}\right\}. \tag{7}$$

Optional update of $\boldsymbol{x}_{\mathrm{A}}$: when using the approximate likelihood method, one must account for the auxiliary variables $\boldsymbol{x}_{\mathrm{A}}$, a $\boldsymbol{P}$ vector (e.g., $P = n$). We find that, in practice, it is not necessary to update all the auxiliary parameters at each iteration. In fact the method can be shown to work perfectly well, empirically, if we *never* update them, provided they are given a sensible initial value (such as the sample mean of the observed data $\overline{x}$). This is not an uncommon tactic in the long-memory (big-n) context (e.g., Beran, 1994b); for further discussion refer to Graves (2013, Appendix C).

For a full-MH approach, we recommend an independence sampler to backward project the observed time series. Specifically, first relabel the observed data: $y_{-i} = x_{i+1}$, $i = 0, \ldots n-1$; furthermore, use the vector $(y_{-(n-1)}, \ldots, y_{-1}, y_0)^t$ to generate a new vector of length n, $(Y_1, \ldots, Y_n)^t$ where Y_t via Eq. (2): $Y_t = \varepsilon_t + \Pi_P\, \mu - \sum_{k=1}^{n} \pi_k\, Y_{t-k}$, where the coefficients $\{\pi\}$ are determined by the current value of the memory parameter(s). Then take the proposed $\boldsymbol{x}_\mathrm{A}$, denoted $\boldsymbol{\xi}_{x_\mathrm{A}}$, as the reverse sequence: $\xi_{x_{-i}} = y_{i+1}$, $i = 0, \ldots, n-1$. Since this is an independence sampler, calculation of the acceptance probability is straightforward. It is only necessary to evaluate the proposal density $q(\boldsymbol{\xi}_{x_\mathrm{A}} | \boldsymbol{x}, \boldsymbol{\psi})$. But this is easy using the results from Sect. 2. For simplicity, we prefer uniform prior for $\boldsymbol{x}_\mathrm{A}$.

Besides simplicity, justification for this approach lies primarily in is preservation of the autocorrelation structure – this is clear since the ACF is symmetric in time. The proposed vector has a low acceptance rate, and the potential remedies (e.g., multiple-try methods) seem unnecessarily complicated given the success of the simpler method.

4 Extensions to accommodate short memory

Simple ARFIMA(0,d,0) models are mathematically convenient but have limited practical applicability because the entire memory structure is determined by just one parameter, d. Although d is often of primary interest, it may be unrealistic to assume no short-memory effects. This issue is often implicitly acknowledged since semi-parametric estimation methods, such as those used as comparators in Sect. 5.1, are motivated by a desire to circumvent the problem of specifying precisely (and inferring) the form of short memory (i.e., the values of p and q in an ARIMA model). Full parametric Bayesian modeling of ARFIMA(p,d,q) processes represents an essentially untried alternative, primarily due to computational challenges. Related, more discrete, alternatives show potential. Pai and Ravishanker (1998) considered all four models with $p, q \leq 1$, whereas Koop et al. (1997) considered 16 with $p, q \leq 3$.

Such approaches, especially ones allowing larger p, q, can be computationally burdensome as much effort is spent modeling unsuitable processes towards a goal (inferring p, q), which is not of primary interest (d is). To develop an efficient, fully parametric, Bayesian method of inference that properly accounts for varying models, and to marginalize out these nuisance quantities, we use reversible-jump (RJ) MCMC (Green, 1995). We extend the parameter space to include the set of models (p and q), with chains moving *between* (i.e., changing p and/or q) and within (sampling $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ given particular fixed p and q) models, and focus on the marginal posterior distribution of d obtained by (Monte Carlo) integration over all models and parameters therein. RJ methods, which mixes so-called *trans-dimensional*, between-model moves with the conventional within-model ones, have previously been applied to both autoregressive models (Vermaak et al., 2004), and full-ARMA models (Ehlers and Brooks, 2006, 2008). In the long-memory context, Holan et al. (2009) applied RJ to fractional exponential processes (FEXP). However for ARFIMA, the only related work we are aware of is by Eğri˙oğlu and Günay (2010) who demonstrated a promising if limited alternative.

Below we show how the likelihood may be calculated with extra short-memory components when p and q are known, and subsequently how Bayesian inference can be applied in this case. Then, the more general case of unknown p and q via RJ is described. The result is a Monte Carlo inferential scheme that allows short-memory effects to be marginalized out when summarizing inferences for the main parameter of interest: d, for long memory.

4.1 Likelihood derivation and inference for known short memory

Recall that short-memory components of an ARFIMA process are defined by the AR and moving average (MA) polynomials, Φ and Θ, respectively (see Sect. 2.1). Here, we distinguish between the polynomial, Φ, and the vector of its coefficients, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$. When the polynomial degree is required explicitly, bracketed superscripts will be used: $\Phi^{(p)}$, $\boldsymbol{\phi}^{(p)}$, $\Theta^{(p)}$, $\boldsymbol{\theta}^{(p)}$.

We combine the short-memory parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ with d to create a single memory parameter, $\boldsymbol{\omega} = (\boldsymbol{\phi}, \boldsymbol{\theta}, d)$. For a given unit-variance ARFIMA(p,d,q) process, we denote its autocovariance (ACV) by $\gamma_\omega(\cdot)$, with $\gamma_d(\cdot)$ and γ_ϕ, $\boldsymbol{\theta}(\cdot)$ those of the relevant unit-variance ARFIMA(0,d,0) and ARMA(p,q) processes, respectively. The spectral density function (SDF) of the unit-variance ARFIMA(p,d,q) process is written as $f_\omega(\cdot)$, and its covariance matrix is $\boldsymbol{\Sigma}_\omega$.

An exact likelihood evaluation requires an explicit calculation of the ACV $\gamma_\omega(\cdot)$; however, there is no simple closed form for arbitrary ARFIMA processes. Fortunately, our proposed approximate likelihood method of section 2 can be ported over directly. Given the coefficients $\{\pi_k^{(d)}\}$ and polynomials Φ and Θ, it is straightforward to calculate the $\{\pi_k^{(\omega)}\}$ coefficients required by again applying the numerical methods of Brockwell and Davis (1991, Sect. 3.3).

To focus the exposition, consider the simple, yet useful, ARFIMA(1,d,0) model where the full memory parameter is $\boldsymbol{\omega} = (d, \phi_1)$. Because the parameter spaces of d and ϕ_1 are independent, it is simplest to update each of these parameters separately; d with the methods of Sect. 3 and ϕ_1 similarly: $\xi_{\phi_1} | \phi_1 \sim \mathcal{N}^{(-1,1)}(\phi_1, \sigma^2_{\phi_1})$, for some $\sigma^2_{\phi_1}$. In practice however, the posteriors of d and ϕ_1 typically exhibit significant correlation so independent proposals are inefficient. One solution would be to parametrize to some d^* and orthogonal ϕ_2^*, but the interpretation of d^* would not be clear. An alternative to explicit reparametrization is to update the parameters jointly, but in such a way that proposals are aligned

with the correlation structure. This will ensure a reasonable acceptance rate and mixing.

To propose parameters in the manner described above, a two-dimensional, suitably truncated Gaussian random walk, with covariance matrix aligned with the posterior covariance, is required. To make proposals of this sort, and indeed for arbitrary $\boldsymbol{\omega}$ in larger p and q cases, requires sampling from a *hypercuboid*-truncated multivariate normal (MVN) $\mathcal{N}_r^{(a,b)}(\boldsymbol{\omega}, \Sigma_\omega)$, where $(\mathbf{a}, \mathbf{b})$ describe the coordinates of the hypercube. We find that rejection sampling-based unconstrained similarly parametrized MVN samples (e.g., using `mvtnorm`, Genz et al., 2012) works well, because in the RW setup the mode of the distribution always lies inside the hypercuboid. Returning to the specific ARFIMA(1,d,0) case, $r = 2$, $\mathbf{b} = (0.5, 1)$, and $\mathbf{a} = -\mathbf{b}$ are appropriate choices. Calculation of the MH acceptance ratio $A_\omega(\boldsymbol{\omega}, \boldsymbol{\xi}_\omega)$ is trivial; it simply requires numerical evaluation of $\Phi_r^{(\mathcal{N})}(\cdot; \cdot, \boldsymbol{\Sigma}_\omega)$, e.g., via `mvtnorm`, since the ratios of hypercuboid normalization terms would cancel. We find that initial values $\phi^{[0]}$ chosen uniformly in $\mathcal{C}_1$; i.e., the interval $(-1, 1)$, and $d^{[0]}$ from $\{-0.4, -0.2, 0, 0.2, 0.4\}$ work well. Any choice of prior for $\boldsymbol{\omega}$ can be made, although we prefer flat (proper) priors.

The only technical difficulty is the choice of proposal covariance matrix $\boldsymbol{\Sigma}_\omega$. Ideally, it would be aligned with the posterior covariance; however, this is not a priori known. We find that running a pilot chain with independent proposals via $\mathcal{N}_r^{(a,b)}(\boldsymbol{\omega}, \sigma_\omega^2 \mathbf{I}_r)$ can help choose a $\boldsymbol{\Sigma}_\omega$. A rescaled version of the sample covariance matrix from the pilot posterior chain, following Roberts and Rosenthal (2001), works well (see Sect. 5.2).

4.2 Unknown short-memory form

We now expand the parameter space to include models $M \in \mathcal{M}$, the set of ARFIMA models with p and q short-memory parameters, indexing the size of the parameter space $\Psi^{(M)}$. For our trans-dimensional moves, we only consider adjacent models, on which we will be more specific later. For now, note that the choice of bijective function mapping between model spaces (whose Jacobian term appears in the acceptance ratio) is crucial to the success of the sampler. To illustrate, consider transforming from $\Phi^{(p+1)} \in \mathcal{C}_{p+1}$ down to $\Phi^{(p)} \in \mathcal{C}_p$. This turns out to be a non-trivial problem, however, because (for $p > 1$) $\mathcal{C}_p$ has a very complicated shape. The most natural map would be $(\phi_1, \ldots, \phi_p, \phi_{p+1}) \longmapsto (\phi_1, \ldots, \phi_p)$. However, there is no guarantee that the image will lie in $\mathcal{C}_p$. Even if the model dimension is fixed, difficulties are still encountered; a natural proposal method would be to update each component of $\boldsymbol{\phi}$ separately but, because of the awkward shape of $\mathcal{C}_p$, the allowable values for each component are a complicated function of the others. Non-trivial proposals are required.

A potential approach is to parametrize in terms of the inverse roots (poles) of Φ, as advocated by Ehlers and Brooks (2006, 2008): by writing $\Phi(z) = \prod_{i=1}^{p}(1 - \alpha_i z)$, we have $\boldsymbol{\phi}^{(p)} \in \mathcal{C}_p \Longleftrightarrow |\alpha_i| < 1$ for all i. This looks attractive because it transforms $\mathcal{C}_p$ into $D^p = D \times \cdots \times D$ (p times) where D is the open unit disc, which is easy to sample from. But this method has serious drawbacks when we consider the RJ step. To decrease dimension, the natural map would be to remove one of the roots from the polynomial. But because it is assumed that Φ has real coefficients (otherwise the model has no realistic interpretation), any complex α_i must appear as conjugate pairs. There is then no obvious way to remove a root; a contrived method might be to remove the conjugate pair and replace it with a real root with the same modulus; however, it is unclear how this new polynomial is related to the original, and to other aspects of the process, like ACV.

4.2.1 Reparametrization of Φ and Θ

We therefore propose reparametrization Φ (and Θ) using the bijection between $\mathcal{C}_p$ and $(-1, 1)^p$ advocated by various authors, e.g., Marriott et al. (1995) and Vermaak et al. (2004). To our knowledge, these methods have not previously been deployed towards integrating out short-memory components in Bayesian analysis of ARFIMA processes.

Monahan (1984) defined a mapping $\boldsymbol{\phi}^{(p)} \longleftrightarrow \boldsymbol{\varphi}^{(p)}$ recursively as follows:

$$\phi_i^{(k-1)} = \frac{\phi_i^{(k)} - \phi_k^{(k)} \phi_{k-i}^{(k)}}{1 - \left(\phi_k^{(k)}\right)^2}, \; k = p, \ldots, 2, \; i = 1, \ldots, k-1. \quad (8)$$

Then set $\varphi_k^{(p)} = \phi_k^{(k)}$ for $k = 1, \ldots, p$. The reverse recursion is given by

$$\phi_i^{(k)} = \begin{cases} \varphi_k^{(p)} & \text{for} \quad i = k \qquad\quad k = 1, \ldots, p \\ \phi_i^{(k-1)} + \varphi_k^{(p)} \phi_{k-i}^{(k-1)} & \text{for} \quad i = 1, \ldots, k-1 \quad k = 2, \ldots, p \end{cases}.$$

Note that $\phi_p^{(p)} = \varphi_p^{(p)}$. Moreover, if $p = 1$, the two parametrizations are the same, i.e., $\phi_1 = \varphi_1$ (consequently the brief study of ARFIMA(1,d,0) in Sect. 4.1 fits in this framework). The equivalent parametrized form for $\boldsymbol{\theta}$ is $\boldsymbol{\vartheta}$. The full memory parameter $\boldsymbol{\omega}$ is parametrized as $\overline{\Omega} = (-1/2, \; 1/2) \times$ (the image of $\mathcal{C}_{p,q}$). However, recall that in practice, $\mathcal{C}_{p,q}$ will be assumed equivalent to $\mathcal{C}_p \times \mathcal{C}_q$, so the parameter space is effectively $\overline{\Omega} = (-1/2, 1/2) \times (-1, 1)^{p+q}$.

Besides mathematical convenience, this bijection has a very useful property (Kay and Marple, 1981, cf.), which helps motivate its use in defining RJ maps. If $d = q = 0$, using this parametrization for $\boldsymbol{\varphi}$ when moving between different values of p allows one to automatically choose processes that have very closely matching ACFs at low lags. In the MCMC context this is useful because it allows the chain to propose models that have a similar correlation structure to the current one. Although this property is nice, it may be of limited value for full-ARFIMA models, since the proof of the main result does not easily lend itself to the inclusion of either a MA or long-memory component. Nevertheless, our empirical results

similarly indicate a near match for a full-ARFIMA(p,d,q) model.

4.2.2 Application of RJ MCMC to ARFIMA(p,d,q) processes

We now use this reparametrization to efficiently propose new parameter values. Firstly, it is necessary to propose a new memory parameter $\boldsymbol{\varpi}$ while keeping the model fixed. Attempts at updating each component individually suffer from the same problems of excessive posterior correlation that were encountered in Sect. 4.1. Therefore, the simultaneous update of the entire $r=(p+q+1)$-dimensional parameter $\boldsymbol{\varpi}$ is performed using the hypercuboid-truncated Gaussian distribution from definition $\boldsymbol{\xi}_{\varpi}|\boldsymbol{\varpi} \sim \mathcal{N}_r^{\mathcal{H}_r}(\boldsymbol{\varpi}, \boldsymbol{\Sigma}_{\varpi})$, where $\mathcal{H}_r$ defines the r-dimensional rectangle. The covariance matrix $\boldsymbol{\Sigma}_{\varpi}$ is discussed in some detail below. The choice of prior $p_{\varpi}(\cdot)$ is arbitrary. Pai and Ravishanker (1998) used a uniform prior for $\boldsymbol{\omega}$, which has an explicit expression in the $\boldsymbol{\varpi}$ parametrization (Monahan, 1984). However, their expression is unnecessarily complicated since a uniform prior over Ω holds no special interpretation. We therefore prefer uniform prior over $\overline{\Omega}$: $p_{\varpi}(\boldsymbol{\varpi}) \propto 1$, $\boldsymbol{\varpi} \in \overline{\Omega}$.

Now consider the between-model transition. We must first choose a model prior $p_{\mathcal{M}}(\cdot)$. A variety of priors are possible; the simplest option would be to have a uniform prior over $\mathcal{M}$, but this would of course be improper. We may in practice want to restrict the possible values of p, q to $0 \le p \le P$ and $0 \le q \le Q$ for some P, Q (say 5), which would render the uniform prior proper. However, even in this formulation, a lot of prior weight is being put onto (larger) more complicated models that, in the interests of parsimony, might be undesired. As a simple representative of potential priors that give greater weight to smaller models, we prefer a truncated joint Poisson distribution with parameter λ: $p_{\mathcal{M}}(p, q) \propto \frac{\lambda^{p+q}}{p!q!}\mathbb{I}(p \le P, q \le Q)$.

Now, denote the probability of jumping from model $M_{p,q}$ to model $M_{p',q'}$ by $U_{(p,q),(p',q')}$. U could allocate non-zero probability for every model pair, but for convenience we severely restrict the possible jumps (while retaining irreducibility) using a two-dimensional bounded birth and death process. Consider the subgraph of $\mathbb{Z}^2$: $G=\{(p,q): 0 \le p \le P, 0 \le q \le Q\}$, and allocate uniform non-zero probability only to neighboring values, i.e., if and only if $|p-p'|+|q-q'|=1$. Each point in the body of G has four neighbors, each point on the line boundaries has three, and each of the four corner points has only two neighbors. Therefore, the model transition probabilities $U_{(p,q),(p',q')}$ are either 1/4, 1/3, 1/2, or 0.

Now suppose the current $(p+q+3)$-dimensional parameter is $\boldsymbol{\psi}^{(p,q)}$, given by $\boldsymbol{\psi}^{(p,q)}=(\mu, \sigma, d, \boldsymbol{\varphi}^{(p)}, \boldsymbol{\vartheta}^{(q)})$, using a slight abuse of notation. Because the mathematical detail of the AR and MA components are almost identical, we consider only the case of decreasing/increasing p by 1 here; all of the following remains valid if p is replaced by q, and $\boldsymbol{\varphi}$ replaced by $\boldsymbol{\vartheta}$. We therefore seek to propose a parameter $\boldsymbol{\xi}^{(p+1,q)}=(\xi_\mu, \xi_\sigma, \xi_d, \boldsymbol{\xi}_{\varphi}^{(p+1)}, \boldsymbol{\xi}_{\vartheta}^{(q)})$, that is somehow based on $\boldsymbol{\psi}^{(p,q)}$. We further simplify by regarding the other three parameters (μ, σ, and d) as having the same interpretation in every model, choosing $\xi_\mu=\mu$, $\xi_\sigma=\sigma$ and $\xi_d=d$. For simplicity we also set $\boldsymbol{\xi}_{\vartheta}^{(q)}=\boldsymbol{\vartheta}^{(q)}$. Now consider the map $\boldsymbol{\varphi}^{(p)} \to \boldsymbol{\xi}_{\varphi}^{(p+1)}$. To specify a bijection, we match dimensions by adding in a random scalar u. The most obvious map is to specify u, so that its support is the interval $(-1, 1)$ and then set: $\boldsymbol{\xi}_{\varphi}^{(p+1)}=(\boldsymbol{\varphi}^{(p)}, u)$. The corresponding map for decreasing the dimension $\boldsymbol{\varphi}^{(p+1)} \to \boldsymbol{\xi}_{\varphi}^{(p)}$ is $\boldsymbol{\xi}_{\varphi}^{(p)}=(\varphi_1^{(p+1)}, \ldots, \varphi_p^{(p+1)})$. We either add, or remove the final parameter, while keeping all others fixed with the identity map, so the Jacobian is unity. The proposal $q(u|\boldsymbol{\psi}^{(p,q)})$ can be made in many ways – we prefer the simple $\mathcal{U}(-1,1)$. With these choices the RJ acceptance ratio is

$$A = \ell_{(p',q')}\left(\boldsymbol{x}|\boldsymbol{\xi}^{(p',q')}\right) - \ell_{(p,q)}\left(\boldsymbol{x}|\boldsymbol{\psi}^{(p,q)}\right) + \log\left\{\frac{p_{\mathcal{M}}(p',q')}{p_{\mathcal{M}}(p,q)}\frac{U_{(p',q'),(p,q)}}{U_{(p,q),(p',q')}}\right\},$$

which applies to both increasing and decreasing dimensional moves.

Construction of $\boldsymbol{\Sigma}_{\varpi}$: much of the efficiency of the above scheme, including within- and between-model moves, depends on the choice of $\boldsymbol{\Sigma}_{\varpi} \equiv \Sigma^{(p,q)}$, the within-model move RW proposal covariance matrix. We first seek an appropriate $\Sigma^{(1,1)}$, as in Sect. 4.1, with a pilot tuning scheme. That matrix is shown on the left below, where we have blocked it out

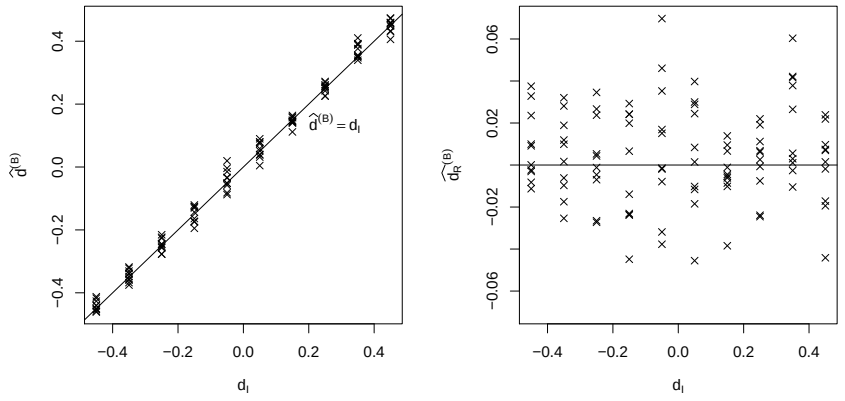
$$\Sigma^{(1,1)} = \left(\begin{array}{c|c|c} \sigma_d^2 & \sigma_{d,\varphi_1} & \sigma_{d,\vartheta_1} \\ & \sigma_{\varphi_1}^2 & \sigma_{\varphi_1,\vartheta_1} \\ & & \sigma_{\vartheta_1}^2 \end{array}\right),$$

$$\Sigma^{(p,q)} = \left(\begin{array}{c|c|c} \sigma_d^2 & \Sigma_{d,\varphi}^{(p)} & \Sigma_{d,\vartheta}^{(q)} \\ & \boldsymbol{\Sigma}_{\varphi}^{(p)}, \boldsymbol{\varphi}^{(p)} & \boldsymbol{\Sigma}_{\varphi}^{(p)}, \boldsymbol{\vartheta}^{(q)} \\ & & \Sigma_{\vartheta}^{(q)}, \boldsymbol{\vartheta}^{(q)} \end{array}\right) \quad (9)$$

(where each block is a scalar), so that we can extend this idea to the (p, q) case in the obvious way – on the right above – where $\boldsymbol{\Sigma}_{\varphi}^{(p)}, \boldsymbol{\varphi}^{(p)}$ is a $p \times p$ matrix, $\Sigma_{\vartheta}^{(q)}, \boldsymbol{\vartheta}^{(q)}$ is a $q \times q$ matrix, etc. If either (or both) $p, q=0$ then the relevant blocks are simply omitted. To specify the various sub-matrices, we propose $\varphi_2, \ldots, \varphi_p$ with equal variances, and *independently* of d, φ_1, ϑ_1, (and similarly for $\vartheta_2, \ldots, \vartheta_q$). In the context of Eq. (9), the following holds true:

$$\Sigma_{d,\varphi}^{(p)} = \left(\begin{array}{c|c} \sigma_{d,\varphi_1} & \mathbf{0} \end{array}\right),$$

$$\Sigma_{d,\vartheta}^{(q)} = \left(\begin{array}{c|c} \sigma_{d,\vartheta_1} & \mathbf{0} \end{array}\right),$$



(a) (b)

Figure 1. Posterior outputs; **(a)** Bayesian estimate $\widehat{d}^{(\mathrm{B})}$ values on the y axis against the true d_I on the x axis, **(b)** residuals $\widehat{d_{\mathrm{R}}}^{(\mathrm{B})}$ from the Bayesian estimate from the truth against that truth, d_I. Each "x" plotted represts one estimate or residual.

$$\boldsymbol{\Sigma}_{\varphi}^{(p)}, \boldsymbol{\varphi}^{(p)} = \left(\begin{array}{c|c} \sigma_{\varphi_1}^2 & \mathbf{0} \\ \mathbf{0} & \sigma_{\varphi}^2 I_{p-1} \end{array} \right),$$

$$\Sigma_{\vartheta}^{(q)}, \boldsymbol{\vartheta}^{(q)} = \left(\begin{array}{c|c} \sigma_{\vartheta_1}^2 & \mathbf{0} \\ \mathbf{0} & \sigma_{\vartheta}^2 I_{q-1} \end{array} \right),$$

$$\boldsymbol{\Sigma}_{\varphi}^{(p)}, \boldsymbol{\vartheta}^{(q)} = \left(\begin{array}{c|c} \sigma_{\varphi_1,\vartheta_1} & \mathbf{0} \\ .\mathbf{0} & \mathbf{O} \end{array} \right),$$

where the dotted lines indicate further blocking, $\mathbf{0}$ is a row vector of zeros, and $\mathbf{O}$ is a zero matrix. This choice of $\boldsymbol{\Sigma}_{\varphi}$ is conceptually simple, computationally easy and preserves the positive definiteness as required (see Graves, 2013).

5 Empirical illustration and comparison

Here we provide empirical illustrations for the methods above: for classical and Bayesian analysis of long-memory models, and extensions for short memory. To ensure consistency throughout, the location and scale parameters will always be chosen as $\mu_I = 0$ and $\sigma_I = 1$. Furthermore, unless stated otherwise, the simulated series will be of length $n = 2^{10} = 1024$. This is a reasonable size for many applications; it is equivalent to 85 years of monthly observations. When using the approximate likelihood method we set $P = n$.

5.1 Long memory

Standard MCMC diagnostics were used throughout to ensure, and tune for, good mixing. Because d is the parameter of primary interest, the initial values $d^{[0]}$ will be chosen to systematically cover its parameter space, usually starting five chains at the regularly spaced points $\{-0.4, -0.2, 0, 0.2, 0.4\}$. Initial values for other parameters are not varied: μ will start at the sample mean $\overline{x}$; σ at the sample standard deviation of the observed series $\boldsymbol{x}$.

5.1.1 Efficacy of approximate likelihood method

We start with the null case; i.e., how does the algorithm perform when the data are not from a long-memory process? One hundred independent ARFIMA(0,0,0), or Gaussian white noise, processes are simulated, from which marginal posterior means, standard deviations, and credibility interval end points are extracted. Table 1 shows averages over the runs.

The average estimate for each of the three parameters is less than a quarter of a standard deviation away from the truth. Credibility intervals are nearly symmetric about the estimate and the marginal posteriors are, to a good approximation, locally Gaussian (not shown). Upon, applying a proxy credible-interval-based hypothesis test, one would conclude in 98 of the cases that $d = 0$ could not be ruled out. A similar analysis for μ and σ shows that hypotheses $\mu = 0$ and $\sigma = 1$ would each have been accepted 96 times. These results indicate that the 95 % credibility intervals are approximately correctly sized.

Next, consider the more interesting case of $d_I \neq 0$. We repeat the above experiment except that 10 processes are generated with d_I set to each of $\{-0.45, -0.35, \ldots, 0.45\}$, giving 100 series total. Figure 1 shows a graphical analog of results from this experiment. The plot axes involve a Bayesian residual estimate of d, $\widehat{d_{\mathrm{R}}}^{(\mathrm{B})}$, defined as $\widehat{d_{\mathrm{R}}}^{(\mathrm{B})} = \widehat{d}^{(\mathrm{B})} - d_I$, where $\widehat{d}^{(\mathrm{B})}$ is the Bayesian estimate of d.

From the figure is clear that the estimator for d is performing well. Figure 1a shows how tight the estimates of d are around the input value – recall that the parameter space for d is the whole interval $(-\frac{1}{2}, \frac{1}{2})$. Moreover, Fig. 1b indicates that there is no significant change of posterior bias or variance as d_I is varied.

Next, the corresponding plots for the parameters σ and μ are shown in Fig. 2. We see from Fig. 2a that the estimate of σ also appears to be unaffected by the input value d_I. The sit-

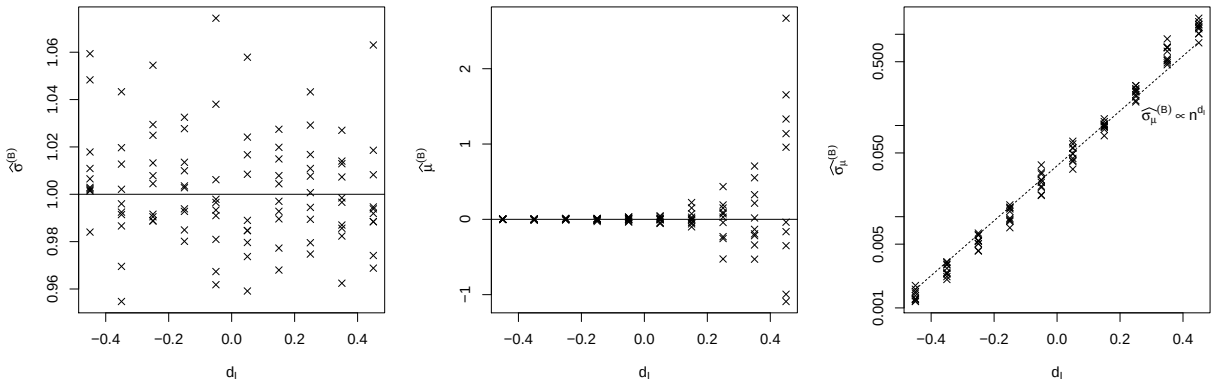


Figure 2. Posterior outputs: **(a)** Bayesian estimated standard deviation $\widehat{\sigma}^{(\mathrm{B})}$ against true d_I values; **(b)** Bayesian estimated mean $\widehat{\mu}^{(\mathrm{B})}$ against d_I; and **(c)** uncertainty in the posterior for μ, the standard deviation $\widehat{\sigma_{\mu}}^{(\mathrm{B})}$ against d_I (semi-log scale). Each "x" plotted corresponds to an estimate.

Table 1. Posterior summary statistics for an ARFIMA(0,0,0) process. Results are based on averaging over 100 independent ARFIMA(0,0,0) simulations for the long-memory parameter d, mean μ and noise variance σ.

	Mean	SD	95 % CI	
d	0.006	0.025	−0.042	0.055
μ	−0.004	0.035	−0.073	0.063
σ	1.002	0.022	0.956	1.041

uation is different however in Fig. 2b for the location parameter μ. Although the bias appears to be roughly zero for all d_I, the posterior variance clearly *is* affected by d_I. To ascertain the precise functional dependence, consider Fig. 2c, which shows, on a semi-log scale, the marginal posterior standard deviation of μ, $\widehat{\sigma_{\mu}}^{(\mathrm{B})}$ against d_I.

It appears that the marginal posterior standard deviation $\widehat{\sigma_{\mu}}^{(\mathrm{B})}$ is a function of d_I; specifically, $\widehat{\sigma_{\mu}}^{(\mathrm{B})} \propto A^{d_I}$, for some A. The constant A could be estimated via least-squares regression. Instead however, inspired by asymptotic results in literature concerning classical estimation of long-memory processes (Beran, 1994a), we set $A = n$ and plotted the best-fitting such line (shown in Fig. 2c). Observe that, although not fitting exactly, the relation $\widehat{\sigma_{\mu}}^{(\mathrm{B})} \propto n^{d_I}$ holds reasonably well for $d_I \in (-\frac{1}{2}, \frac{1}{2})$. Indeed, Beran (1994a) motivated long memory in this way, and derived asymptotic consistency results for optimum (likelihood-based) estimators and found indeed that the standard error for μ is proportional to $n^{d-1/2}$ but the standard errors of all other parameters are proportional to $n^{-1/2}$.

5.1.2 Effect of varying time series length

We now analyze the effect of changing the time series length. For this we conduct a similar experiment but fix $d_I = 0$ and vary n. The posterior statistics of interest are the posterior standard deviations $\widehat{\sigma_d}^{(\mathrm{B})}$, $\widehat{\sigma_{\mu}}^{(\mathrm{B})}$, and $\widehat{\sigma_{\sigma}}^{(\mathrm{B})}$. For each $n \in \{128 = 2^7,\ 2^8,\ \ldots,\ 2^{14} = 16\,384\}$, 10 independent ARFIMA(0,0,0) time series are generated. The resulting posterior standard deviations are plotted against n (on log–log scale) in Fig. 3.

Observe that all three marginal posterior standard deviations are proportional to $\frac{1}{\sqrt{n}}$, although the posterior of μ is less reliable. Combining these observations with our earlier deduction that $\sigma_{\mu}^{(\mathrm{B})} \propto n^{d_I}$, we conclude that for an ARFIMA(0,d_I,0) process of length n, the marginal posterior standard deviations follow those of Beran (1994a).

5.1.3 Comparison with common estimators

In many practical applications, the long-memory parameter is estimated using non-/semi-parametric methods. These may be appropriate in many situations, where the exact form of the underlying process is unknown. However, when a specific model form is known (or at least assumed) they tend to perform poorly compared with fully parametric alternatives (Franzke et al., 2012). Our aim here is to demonstrate, via a short Monte Carlo study involving ARFIMA(0,d,0) data, that our Bayesian likelihood-based method significantly outperforms other common methods in that case. We consider the following comparators: (i) rescaled adjusted range, or R/S (Hurst, 1951; Graves, 2013) – we use the R implementation in the `FGN` (McLeod et al., 2007) package; (ii) semi-parametric Geweke–Porter–Hudak (GPH) method (Geweke and Porter-Hudak, 1983) – implemented in R package `fracdiff` (Fraley et al., 2012); (iii) detrended fluctuation analysis (DFA), originally devised by Peng et al. (1994) – in the R package `PowerSpectrum` (Vyushin et al., 2009); and (iv) wavelet-based semi-parametric estimators (Abry et al., 2003) available in R package `fARMA` (Wuertz, 2012).

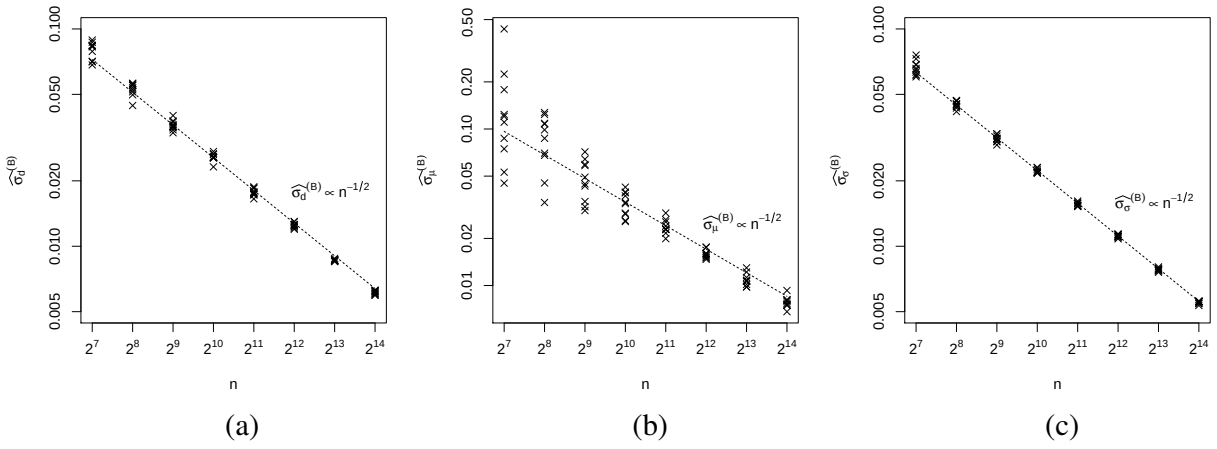


Figure 3. Posterior outputs from an ARFIMA(0,0,0) series: **(a)** the posterior standard deviation in d, $\widehat{\sigma_{\mathrm{d}}}^{(\mathrm{B})}$ against the sample size n; **(b)** posterior standard deviation in μ, $\widehat{\sigma_{\mu}}^{(\mathrm{B})}$ against n; and **(c)** $\widehat{\sigma_{\sigma}}^{(\mathrm{B})}$ against n (log–log scale).

n	mean difference
128	0.057
256	0.029
512	0.015
1024	0.007

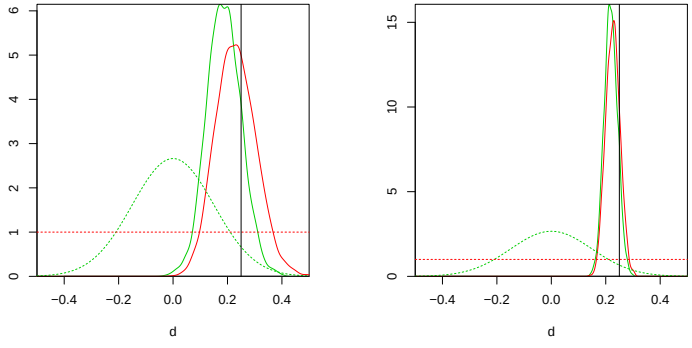


Figure 4. Table: mean difference of estimates $\hat{d}^{(\mathrm{B})}$ under alternative prior assumption. Plots: comparison of posteriors (solid lines) obtained under different priors (dotted lines). Time series used: ARFIMA(0,0.25,0) – **(a)** $n = 2^7 = 128$, **(b)** $n = 2^{10} = 1024$.

Each of these four methods will be applied to the same 100 time series with varying d_I as were used earlier experiments above. We extend the idea of a residual, $\widehat{d_{\mathrm{R}}}^{(\mathrm{R})}$, $\widehat{d_{\mathrm{R}}}^{(G)}$, $\widehat{d_{\mathrm{R}}}^{(D)}$, and $\widehat{d_{\mathrm{R}}}^{(W)}$, to accommodate the new comparators, respectively, and plot them against $\widehat{d_{\mathrm{R}}}^{(\mathrm{B})}$ in Fig. 5.

Observe that all four methods have a much larger variance than our Bayesian method, and moreover the R/S is positively biased. Actually, the bias in some cases would seem to depend on d_I: R/S is significantly (i.e., > 0.25) biased for $d_I < -0.3$ but slightly negatively biased for $d > 0.3$ (not shown); DFA is only unbiased for $d_I > 0$; both the GPH and wavelet methods are unbiased for all $d \in (-\frac{1}{2}, \frac{1}{2})$.

5.2 Extensions for short memory

Known form: we first consider the MCMC algorithm from Sect. 4.1 for sampling under an ARFIMA(1,d,0) model where the full memory parameter is $\boldsymbol{\omega} = (d, \phi_1)$. Recall that method involved proposals from a hypercuboid MVN using a pilot-tuned covariance matrix. Also recall that it is a special case of the reparametrized method from Sect. 4.2.

In general, this method works very well; two example outputs are presented in Fig. 6, under two similar data-generating mechanisms.

Figure 6a shows relatively mild correlation ($\rho = 0.21$) compared with Fig. 6b, which shows strong correlation ($\rho = 0.91$). This differential behavior can be explained heuristically by considering the differing data-generating values. For the process in Fig. 6a the short-memory and long-memory components exhibit their effects at opposite ends of the spectrum; see Fig. 7a. The resulting ARFIMA spectrum, with peaks at either end, makes it easy to distinguish between short and long-memory effects, and consequently the posteriors of d and ϕ are largely uncorrelated. In contrast, the parameters of the process in Fig. 7b express their behavior at the same end of the spectrum. With negative d these effects partially cancel each other out, except very near the origin where the negative memory effect dominates; see Fig. 7b. Distinguishing between the effects of ϕ and d is much more

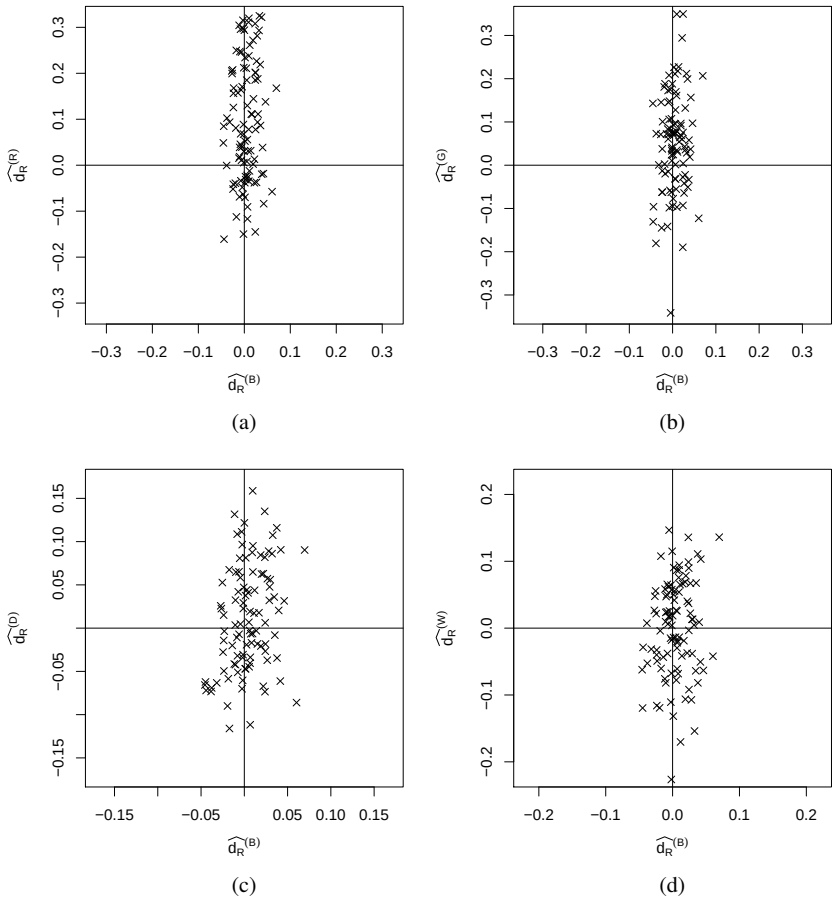


Figure 5. Comparison of Bayesian estimator with common classical estimators: **(a)** R/S, **(b)** GPH, **(c)** DFA, and **(d)** wavelet.

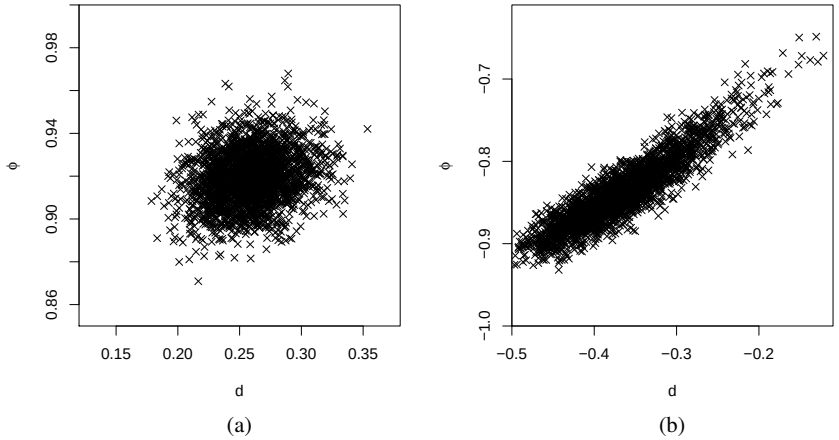


Figure 6. Posterior samples of (d, ϕ): input time series **(a)** $(1+0.92\mathcal{B})(1-\mathcal{B})^{0.25}\,X_t = \varepsilon_t$, **(b)** $(1-0.83\mathcal{B})(1-\mathcal{B})^{-0.35}\,X_t = \varepsilon_t$.

difficult in this case; consequently the posteriors are much more dependent.

In cases where there is significant correlation between d and ϕ, it arguably makes little sense to consider only the marginal posterior distribution of d. For example the 95 % credibility interval for d from Fig. 7b is $(-0.473,\ -0.247)$, and the corresponding interval for ϕ is $(-0.910,\ -0.753)$, yet these clearly give a rather pessimistic view of our joint knowledge about d and ϕ; see Fig. 7c. In theory an ellipsoidal credibility set could be constructed although this is clearly less practical when $\sim \boldsymbol{\omega} > 2$.

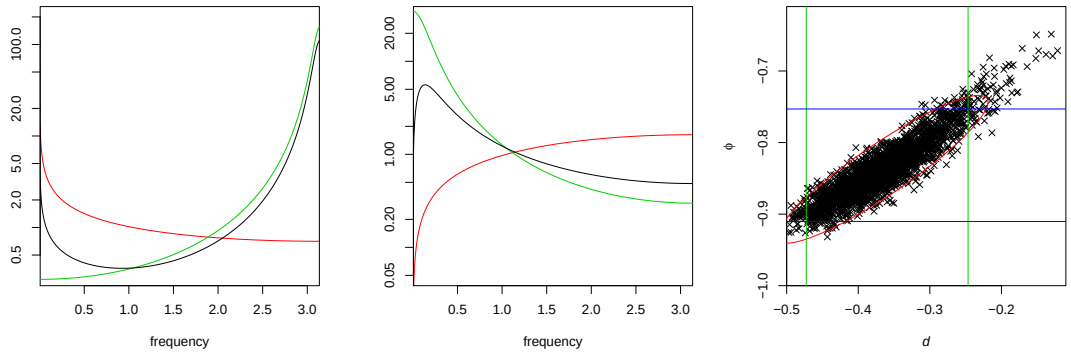


Figure 7. Spectra for processes in Fig. 6. Green line is relevant ARMA(1,0) process, red line is relevant ARFIMA(0,d,0) process, black line is ARFIMA(1,d,0) process: **(a)** $(1+0.92\mathcal{B})(1-\mathcal{B})^{0.25}\,X_t=\varepsilon_t$; **(b)** $(1-0.83\mathcal{B})(1-\mathcal{B})^{-0.35}\,X_t=\varepsilon_t$. **(c)** Shows posterior samples of (d,ϕ) from series considered in **(b)** with credibility sets: red is 95 % credibility set for (d,ϕ), green is 95 % credibility interval for d, blue is 95 % credibility interval for ϕ.

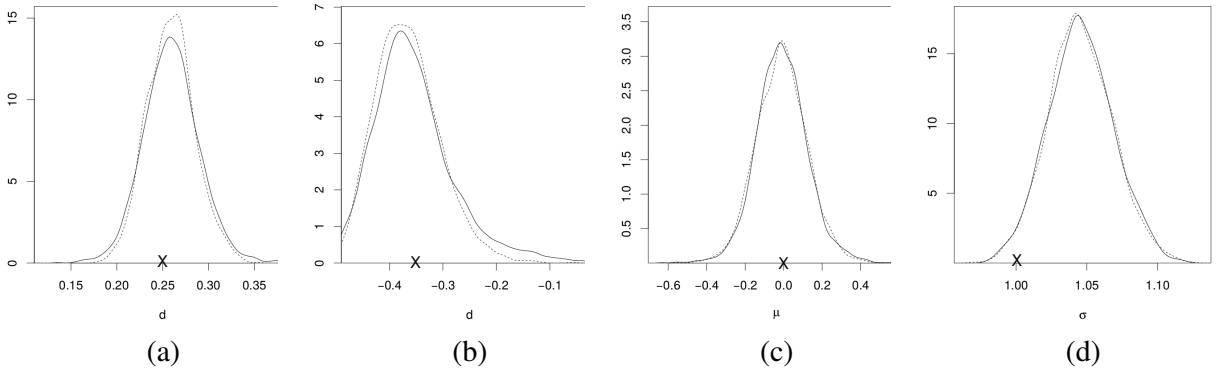


Figure 8. Marginal posterior density of d from series in Fig. 6, **(a, b)**. Solid line is density obtained using reversible-jump algorithm. Dotted line is density obtained using fixed $p=1$ and $q=0$. The true values are $d_l=0.25$ and -0.35, respectively. **(c, d)** Shows the posterior densities for μ and σ, respectively, corresponding to the series in Fig. 6a; those for Fig. 6b look similar. The true values are $\mu=0$ and $\sigma=1$. True values are marked by an X.

Unknown form: the RJ scheme outlined in Sect. 4.2 works well for data simulated with p and q up to 3. The marginal posteriors for d are generally roughly centered around d_l (the data-generating value) and the modal posterior model probability is usually the correct one. To illustrate, consider again the two example data-generating contexts used above.

For both series, kernel density for the marginal posterior for d are plotted in Fig. 8a and b, together with the equivalent density estimated assuming unknown model orders.

Notice how the densities obtained via the RJ method are very close to those obtained assuming $p=1$ and $q=0$. The former are slightly more heavy tailed, reflecting a greater level of uncertainty about d. Interestingly, the corresponding plots for the posteriors of μ and σ do not appear to exhibit this effect; see Fig. 8c and d. The posterior model probabilities are presented in Table 2, showing that the correct modes are being picked up consistently.

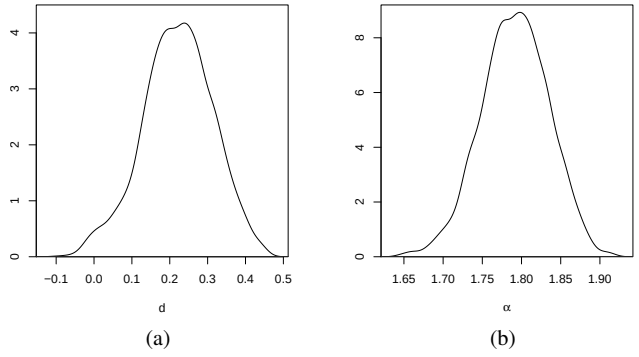


Figure 9. Marginal posterior densities **(a)** d, **(b)** α from the model Eq. (10).

Table 2. Posterior model probabilities for time series from Figs. 6a, b and 8a, b for the autoregressive parameter p and moving average parameter q. Bold numbers denote the true model.

$p\backslash q$	0	1	2	3	4	5	Marginal
				(a)			
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	**0.805**	0.101	0.003	0.000	0.000	0.000	0.908
2	0.038	0.043	0.001	0.000	0.000	0.000	0.082
3	0.005	0.004	0.000	0.000	0.000	0.000	0.009
4	0.000	0.001	0.000	0.000	0.000	0.000	0.001
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Marginal	0.848	0.148	0.004	0.000	0.000	0.000	
				(b)			
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	**0.829**	0.125	0.002	0.000	0.000	0.000	0.956
2	0.031	0.013	0.000	0.000	0.000	0.000	0.044
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Marginal	0.860	0.138	0.002	0.000	0.000	0.000	

As a test of the robustness of the method, consider a complicated short-memory input combined with a heavy-tailed α-stable innovations distribution. Specifically, the time series that will be used is the following ARFIMA(2,d,1) process

$$\left(1-\frac{9}{16}\mathcal{B}^2\right)(1-\mathcal{B})^{0.25}X_t = \left(1+\frac{1}{3}\mathcal{B}\right)\varepsilon_t, \quad \text{where } \varepsilon_t \sim \mathcal{S}_{1.75,0}. \tag{10}$$

For more details, see Graves (2013, Sect. 7.1). The marginal posterior densities of d and α are presented in Fig. 9.

Performance looks good despite the complicated structure. The posterior estimate for d is $\widehat{d}^{(\mathrm{B})} = 0.22$, with 95 % CI (0.04, 0.41). Although this interval is admittedly rather wide, it is reasonably clear that long memory is present in the signal. The corresponding interval for α is (1.71, 1.88) with estimate $\widehat{\alpha}^{(\mathrm{B})} = 1.79$. Finally, we see from Table 3 that the algorithm is very rarely in the wrong model.

6 Observational data analysis

We conclude with the application of our method to two long data sets: the Nile water level minima data and the CET. The Nile data are part of the R package "longmemo" and the CET time series can be downloaded from http://www.metoffice.gov.uk/hadobs/hadcet/.

6.1 The Nile data

Because of the fundamental importance of the Nile river to the civilizations it has supported, local rulers kept measurements of the annual maximal and minimal heights obtained by the river at certain points (called gauges). The longest uninterrupted sequence of recordings is from the Roda gauge (near Cairo), between AD 622 and 1284 ($n = 663$).[2] These data are plotted in Fig. 10.

Table 3. Posterior model probabilities based on simulations of model Eq. (10) for the autoregressive parameter p and moving average parameter q. Bold numbers denote the true model.

$p\backslash q$	0	1	2	3	4	5	Marginal
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	**0.822**	0.098	0.001	0.000	0.000	0.921
3	0.014	0.056	0.004	0.000	0.000	0.000	0.075
4	0.003	0.001	0.000	0.000	0.000	0.000	0.004
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Marginal	0.017	0.880	0.102	0.002	0.000	0.000	

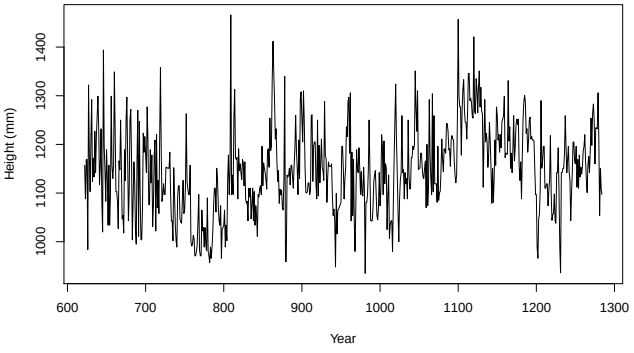


Figure 10. Annual Nile minima time series.

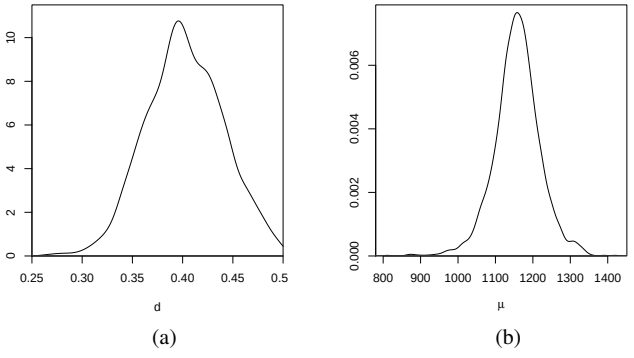
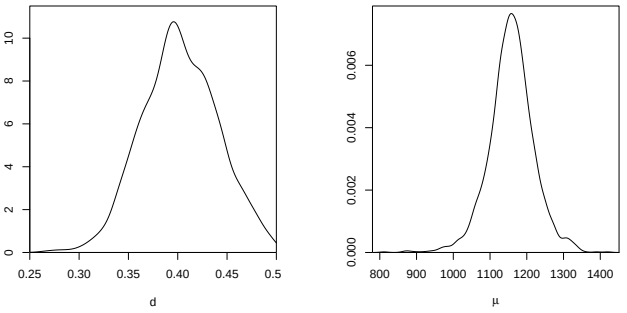


Figure 11. Marginal posterior densities for Nile minima; **(a)** d, **(b)** μ.

We immediately observe the apparent low frequency component of the data. The data appear to be on the "verge" of being stationary; however, the general consensus amongst the statistical community is that the series *is* stationary. The posterior summary statistics are presented in Table 5, density estimates of the marginal posteriors of d and μ are presented in Fig. 12, and the posterior model probabilities are presented in Table 4.

[2] There is evidence (e.g., Ko and Vannucci, 2006b) that the sequence is not actually homogeneous.

	mean	95% CI	
d	0.402	0.336	0.482
μ	1158	1037	1284
σ	70.15	66.46	73.97



(a) (b)

Figure 12. Table: summary posterior statistics for Nile minima. Plots: marginal posterior densities for Nile minima – **(a)** d, **(b)** μ.

Table 4. Posterior model probabilities for Nile minima time series for the autoregressive parameter p and moving average parameter q. Bold numbers denote the best fit model.

$p\backslash q$	0	1	2	3	4	5	Marginal
0	**0.638**	0.101	0.010	0.000	0.000	0.000	0.750
1	0.097	0.124	0.011	0.000	0.000	0.000	0.232
2	0.007	0.010	0.000	0.000	0.000	0.000	0.018
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Marginal	0.742	0.236	0.022	0.000	0.000	0.000	

Table 5. Summary posterior statistics for Nile minima time series for the long-memory parameter d, mean μ, and noise variance σ.

	Mean	SD	95 % CI end points	
d	0.402	0.039	0.336	0.482
μ	1158	62	1037	1284
σ	70.15	1.91	66.46	73.97

The posterior summary statistics and marginal densities of d and μ for the Nile data are presented in Fig. 12. Posterior model probabilities are presented in Table 6. We see that the model with the highest posterior probability is the ARFIMA(0,d,0) model with $d \approx 0.4$. This suggests a strong, pure, long-memory feature. Our results compare favorably with other studies (Liseo et al., 2001; Hsu and Breidt, 2003; Ko and Vannucci, 2006a).

It is interesting to compare these findings with other literature. Liseo et al. (2001) used a semi-parametric Bayesian method on the first 512 observations of the sequence and obtained an estimate for d of 0.278. Hsu and Breidt (2003) used a similar method to Pai and Ravishanker (1998) to estimate d (within an ARFIMA(0,d,0) model) at 0.416 with an approximate credibility interval of (0.315, 0.463). Ko and Vannucci (2006a) similarly found using wavelets $\widehat{d^{\mathrm{B}}} = 0.379$ with a credibility interval of (0.327, 0.427). Palma (2007) obtained $\widehat{d^{\mathrm{B}}} = 0.420$. Holan et al. (2009) obtained $\widehat{d^{\mathrm{B}}} = 0.387$ with

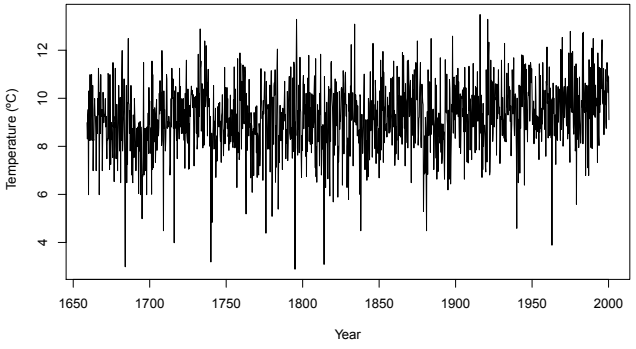


Figure 13. CET time series (deseasonalized).

a credibility interval of (0.316, 0.475) using their Bayesian FEXP method.

We note that the interpretation as persistence of the $d \approx 0.4$ ($H \approx 0.9$) value that we and others have obtained has been challenged by Kärner (2001). In his view the analysis should be applied to the increments of the level heights rather than the level heights themselves, giving an anti-persistent time series with a negative d value. The need for a short-range-dependent component that he argues for is, however, automatically included in the use of an ARFIMA model. Although ARFIMA was originally introduced in econometrics as a phenomenological model of LM, very recent progress is being made in statistics and physics on building a bridge between it and continuous time linear dynamical systems (see e.g., Slezak and Weron, 2015).

In conclusion, our findings agree with all published Bayesian long-memory results (except for the anomalous finding of Liseo et al. (2001)). Moreover, these findings agree with numerous classical methods of analysis (e.g., Beran, 1994a) that have found the best model fit is an ARFIMA(0,d,0) model with $d \approx 0.4$. We note that it is a result of our data analysis method that short-memory can be neglected, rather than being an a priori assumption.

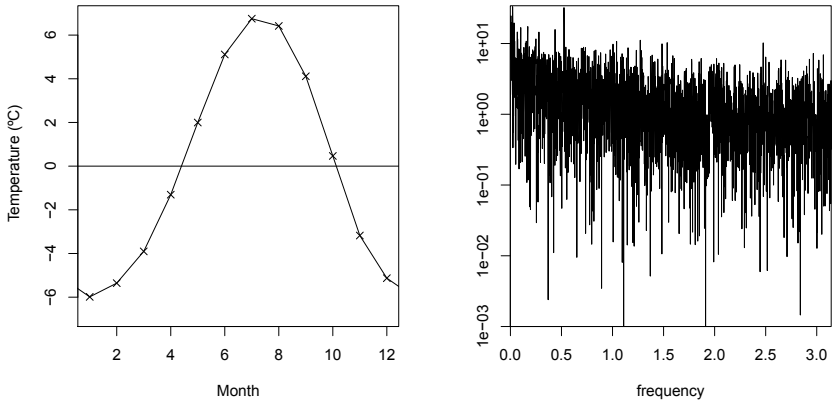


Figure 14. CET time series; **(a)** assumed deterministic seasonal component $S(t)$, **(b)** spectrum of deseasonalized index.

Table 6. Posterior model probabilities for Nile minima time series for the autoregressive parameter p and moving average parameter q.

$p\backslash q$	0	1	2	3	4	5	Marginal
0	0.638	0.101	0.010	0.000	0.000	0.000	0.750
1	0.097	0.124	0.011	0.000	0.000	0.000	0.232
2	0.007	0.010	0.000	0.000	0.000	0.000	0.018
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Marginal	0.742	0.236	0.022	0.000	0.000	0.000	

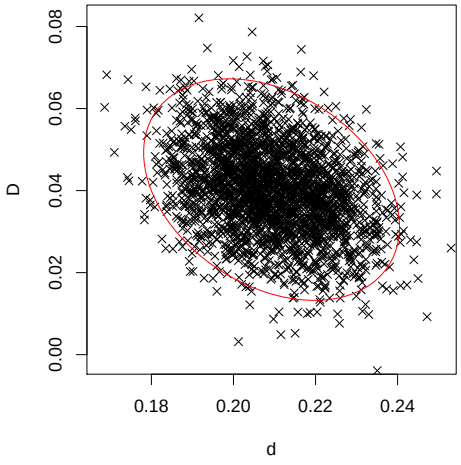


Figure 15. Joint posterior samples of (d, D) with 95 % credibility set in red for CET time series.

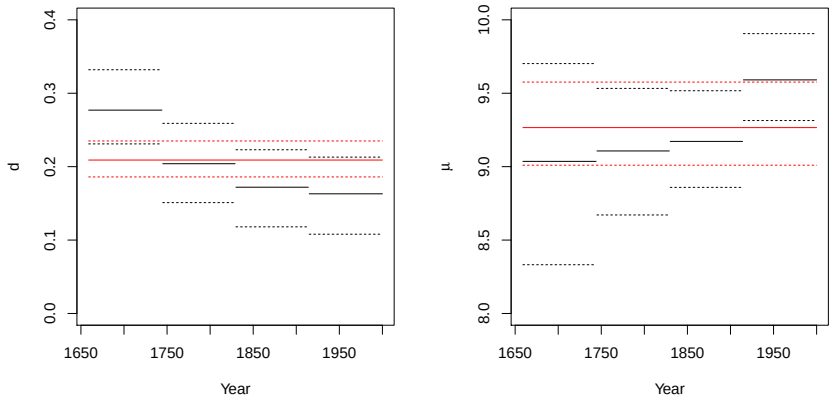
6.2 Central England temperature

There is increasing evidence that surface air temperatures posses long memory (Gil-Alana, 2003, 2008; Bunde et al., 2014; Franzke, 2010, 2012) but long time series are needed to get robust results. The CET index is a famous measure of the *monthly* mean temperature in an area of southern-central England dating back to 1659 (Manley, 1974). Given to a precision of 0.5 °C prior to 1699 and 0.1 °C thereafter, the index is considered to be the longest reliable known temperature record from station data. As expected, the CET exhibits a significant seasonal signal, at least some of which must be considered as deterministic. Following the approach of Montanari et al. (2000), the index is first deseasonalized using the additive "STL" method (Cleveland et al., 1990). This deseasonalized CET index is shown in Fig. 13.

The estimated seasonal function $S(t)$ that was removed is shown in Fig. 14a. The spectrum of the deseasonalized process is shown in Fig. 14b. D denotes the seasonal long-memory parameter. Notice that, in addition to the obvious spectral peak at the origin, there still remains a noticeable peak at the monthly frequency $\omega = \frac{\pi}{6}$. However, there are no further peaks in the spectrum, which would appear to rule out a seasonal ARFIMA (SARFIMA) model. These observations therefore suggest that a simple two-frequency Gegenbauer(d,D;$\frac{\pi}{6}$)$_2$ process might be an appropriate model. See Appendix B for more details about seasonal long memory.

Applying this model, the marginal posterior statistics are presented in Table 7 and the joint posterior samples of (d, D) from this model are plotted in Fig. 15. These clearly indicate that both d and D are non-zero (albeit small in the case of D) suggesting the presence of long memory in both the conventional and seasonal sense.

In order to compare these results with other publications', it is important to note that to remove annual seasonality from the CET, the series of annual means is often used instead of the monthly series. This of course reduces the fidelity of the analysis. Hosking (1984) found (using rather crude estimation procedures) that the best-fitting model for the annual means of the CET was the ARFIMA(1,0.33,0) model with $\phi = 0.16$. Pai and Ravishanker (1998) used the same series as test data for their Bayesian method; they fitted each of



(a) (b)

Figure 16. CET time series; posterior estimate (solid line) and 95 % credibility interval (dotted line) for four blocks (black) and whole index (red) for **(a)** d, **(b)** μ.

Table 7. Posterior summary statistics for CET index for the long-memory parameter d, seasonal long-memory parameter D, mean μ, and noise variance σ.

	Mean	SD	95 % CI end points	
d	0.209	0.013	0.186	0.235
D	0.040	0.011	0.018	0.062
μ	9.266	0.144	9.010	9.576
σ	1.322	0.015	1.294	1.353

the ARFIMA models with $p, q \leq 1$ and found that all models were suitable. Their estimates of d ranged from 0.24 for $p = q = 0$ to 0.34 for $p = 0$, $q = 1$.

Of course all these studies assume the time series is stationary and, in particular, has a constant mean. The validity of this assumption was considered by Gil-Alana (2003) who used formal hypothesis testing to consider models:

$$Y_t = \beta_0 + \beta_1 t + X_t, \tag{11}$$

where $\{X_t\}$ is an ARFIMA(0,d,0) process. For values of $d = 0$, 0.05, 0.10, 0.15, β_1 was found to be significantly non-zero (at about 0.23 °C per century) but for $d \geq 0.20$, statistical significance was not found. Gil-Alana (2008) later extended this work by replacing the ARFIMA(0,d,0) process in Eq. (11) with a Gegenbauer(d;ω) process to obtain similar results. However, choice of ω was rather ad hoc, likely influencing the results.

In order to consider the stationarity of the time series, we divided the series up into four blocks of length 1024 months (chosen to maximize efficiency of the fast Fourier transform) and analyzed each block independently. The posterior statistics for each block are presented in Table 8 with some results presented graphically in Fig. 16.

Table 8. Posterior summary statistics for four blocks of CET index for the long-memory parameter d, seasonal long-memory parameter D, mean μ, and noise variance σ.

		Mean	SD	95 % CI end points	
1659–1744	d	0.277	0.026	0.231	0.332
	D	0.054	0.022	0.013	0.097
	μ	9.036	0.347	8.332	9.702
	σ	1.217	0.027	1.167	1.271
1744–1829	d	0.204	0.028	0.151	0.259
	D	0.017	0.023	−0.028	0.063
	μ	9.107	0.216	8.671	9.533
	σ	1.348	0.031	1.290	1.409
1829–1914	d	0.172	0.027	0.118	0.223
	D	0.036	0.022	−0.010	0.076
	μ	9.172	0.168	8.859	9.517
	σ	1.364	0.030	1.312	1.429
1914–2000	d	0.163	0.027	0.108	0.213
	D	0.063	0.022	0.023	0.109
	μ	9.591	0.152	9.314	9.906
	σ	1.348	0.030	1.291	1.406

It is interesting to note that the degree of (conventional) long memory is roughly constant over the last three blocks but appears to be larger in the first block. Of particular concern is that there is no value of d that is included in all four 95 % credibility intervals; this would suggest non-stationarity. Although this phenomenon may indeed have a physical explanation, it is more likely caused by the inhomogeneity of the time series. Recall that the first 50 years of the index are only given to an accuracy of 0.5 °C compared to 0.1 °C afterwards; this lack of resolution clearly has the potential to bias in favor of strong autocorrelation when compared with later periods.

Interestingly, the seasonal long-memory parameter D has 95 % credibility intervals that include zero for the both the second and third blocks. Finally, note that the 95 % credibility intervals for μ all include the range (9.314, 9.517), in other words it is entirely credible that the mean is non-varying over the time period.

7 Conclusions

We have provided a systematic treatment of efficient Bayesian inference for ARFIMA models, the most popular parametric model combining long- and short-memory effects. Through a mixture of theoretical and empirical work we have demonstrated that our method can handle the sorts of time series data with possible long memory that we are typically confronted with.

Many of the choices made throughout, but in particular those leading to our likelihood approximation, stem from a need to accommodate further extension. For example, in future work we intend to extend them to cope with heavy-tailed innovation distributions. For more evidence of potential in this context, see Graves (2013, Sect. 7).

Finally, an advantage of the Bayesian approach is that it provides a natural mechanism for dealing with missing data, via data augmentation. This is particularly relevant for long historical time series, which may, for a myriad of reasons, have recording gaps. For example, some of the data recorded at other gauges along the Nile have missing observations although otherwise span a similarly long time frame. For a demonstration of how this might fit within our framework, see Sect. 5.6 of Graves (2013).

Appendix A: ARFIMA model

We define an autocovariance ACV $\gamma(\cdot)$ of a weakly stationary process as $\gamma(k) = \mathrm{Cov}(X_t\,(E)[X])(X-(E)[X])]$ is the lag-covariance matrix. The (normalized) ACF $\rho(\cdot)$ is defined as $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$. A stationary process $\{X_t\}$ is said to be *causal* if there exists a sequence of coefficients $\{\psi_k\}$, with finite total mean square $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ such that for all t, a given member of the process can be expanded as a power series in the backshift operator acting on the innovations, $\{\varepsilon_t\}$:

$$X_t = \Psi(\mathcal{B})\varepsilon_t, \text{ where } \Psi(z) = \sum_{k=0}^{\infty} \psi_k z^k. \tag{A1}$$

The innovations are a white (i.e., stationary, zero mean, iid) noise process with variance σ^2. Causality specifies that for every t, X_t can only depend on the past and present values of the innovations $\{\varepsilon_t\}$.

A process $\{X_t\}$ is said to be an autoregressive process of order p, AR(p), if for all t:

$$\begin{aligned}\Phi(\mathcal{B})X_t =& \varepsilon_t, \text{ where } \Phi(z) = 1 + \sum_{k=1}^{p} \phi_k z^k, \\ &\text{and } \left(\phi_1, \ldots, \phi_p\right) \in \mathbb{R}^p.\end{aligned} \tag{A2}$$

AR(p) processes are invertible, stationary and causal if and only if $\Phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. $\{X_t\}$ is said to be a moving average process on the order of q, MA(q), if

$$\begin{aligned}X_t =& \Theta(\mathcal{B})\varepsilon_t, \text{ where } \Theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k, \\ &\text{and } \left(\theta_1, \ldots, \theta_p\right) \in \mathbb{R}^q,\end{aligned} \tag{A3}$$

for all t.[3] MA(q) processes are stationary and causal, and are invertible if and only if $\Theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. A natural extension of the AR and MA classes arises by combining them (Box and Jenkins, 1970).

The process $\{X_t\}$ is said to be an ARMA process of orders p and q, ARMA(p,q), if for all t:

$$\Phi(\mathcal{B})X_t = \Theta(\mathcal{B})\varepsilon_t. \tag{A4}$$

Although there is no simple closed form for the ACV of an ARMA process with arbitrary p and q, so long as the process is causal and invertible, then $|\rho(k)| \leq C\, r^k$, for $k > 0$; i.e., it decays exponentially fast. In other words, although correlation between nearby points may be high, dependence between distant points is negligible.

Before turning to long memory, we require one further result. Under some extra conditions, stationary processes with ACV $\gamma(\cdot)$ possess a SDF $f_{\mathrm{sd}}(\cdot)$ defined such that $\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f_{\mathrm{sd}}(\lambda)\, d\lambda$, $\forall k \in \mathbb{Z}$. This can be inverted to obtain an explicit expression for the SDF (e.g., Brockwell and Davis, 1991, §4.3): $f_{\mathrm{sd}}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k)\, e^{-ik\lambda}$, where $-\pi \leq \lambda \leq \pi$.[4] Finally, the SDF of an ARMA process is

$$f_{\mathrm{sd}}(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta\left(e^{-i\lambda}\right)|^2}{|\Phi\left(e^{-i\lambda}\right)|^2},\ 0 \leq \lambda \leq \pi. \tag{A5}$$

For an ARFIMA process (Eq. 1) the restriction $|d| < \frac{1}{2}$ is necessary to ensure stationarity; clearly if $|d| \geq \frac{1}{2}$ the ACF would not decay. The continuity between stationary and non-stationary processes around $|d| = \frac{1}{2}$ is similar to those that occur for the AR(1) process with $|\phi_1| \to 1$ (such processes are stationary for $|\phi_1| < 1$, but the case $|\phi_1| = 1$ is the non-stationary random walk).

There are a number of alternative definitions of LM, one of which is particularly useful, as it considers the frequency domain: a stationary process has long memory when its SDF follows $f_{\mathrm{sd}}(\lambda) \sim c_f\, \lambda^{-2d}$, as $\lambda \to 0^+$ for some positive constant c_f, and where $0 < d < \frac{1}{2}$.

The simplest way of *creating* a process that exhibits long memory is through the SDF. Consider $f_{\mathrm{sd}}(\lambda) = |1 - e^{i\lambda}|^{-2d}$, where $0 < |d| < \frac{1}{2}$. By simple algebraic manipulation, this is equivalently $f_{\mathrm{sd}}(\lambda) = \left(2 \sin \frac{\lambda}{2}\right)^{-2d}$, from which we deduce that $f(\lambda) \sim \lambda^{-2d}$ as $\lambda \to 0^+$. Therefore, assuming stationarity, the process that has this SDF (or any scalar multiple of it) is a long-memory process. More generally, a process having spectral density

$$f_{\mathrm{sd}}(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d},\ 0 < \lambda \leq \pi, \tag{A6}$$

is called *fractionally integrated* with memory parameter d, Fractionally Integrated FI(d) with memory parameter d (Barnes and Allan, 1966; Adenstedt, 1974). The full trichotomy of negative, short, and long memory is determined solely by d.

In practice this model is of limited appeal to time series analysts because the entire memory structure is determined by just one parameter, d. One often therefore generalizes it by taking any short-memory SDF $f^*_{\mathrm{sd}}(\cdot)$, and defining a new SDF: $f_{\mathrm{sd}}(\lambda) = f^*_{\mathrm{sd}}(\lambda) \left|1 - e^{i\lambda}\right|^{-2d}$, $0 \leq \lambda \leq \pi$. An obvious class of short-memory processes to use this way is ARMA. Taking f^* from Eq. (A5) yields so-called autoregressive fractionally integrated moving average process with parameter d, and orders p and q (ARFIMA(p,d,q)), having SDF:

[3] Many authors define $\Phi(z) = 1 - \sum \phi_k z^k$. Our version emphasizes connections between Φ and Eqs. (A2)–(A3).

[4] Since ACV of a stationary process is an even function of lag, the above equation implies that the associated SDF is an even function. One therefore only needs to be interested positive arguments: $0 \leq \lambda \leq \pi$.

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta\left(e^{-i\lambda}\right)|^2}{|\Phi(e^{-i\lambda})|^2} |1 - e^{i\lambda}|^{-2d},\ 0 \le \lambda \le \pi. \tag{A7}$$

Choosing $p = q = 0$ recovers FI(d) $\equiv$ ARFIMA(0,d,0).

Practical utility from the perspective of (Bayesian) inference demands finding a representation in the temporal domain. To obtain this, consider the operator $(1-\mathcal{B})^d$ for real $d > -1$, which is formally defined using the generalized form of the binomial expansion (Brockwell and Davis, 1991, Eq. 13.2.2):

$$(1-\mathcal{B})^d =: \sum_{k=0}^{\infty} \pi_k^{(d)} \mathcal{B}^k, \text{ where } \pi_k^{(d)} = (-1)^k \frac{1}{\Gamma(k+1)} \frac{\Gamma(d+1)}{\Gamma(d-k+1)}. \tag{A8}$$

From this observation, one can show that $X_t = (1-\mathcal{B})^{-d} Z_t$, where $\{Z_t\}$ is an ARMA process, has SDF (Eq.). The operator $(1-\mathcal{B})^d$ is called the fractional differencing operator since it allows a degree of differencing between the zeroth and first order. The process $\{X_t\}$ is fractionally inverse differenced; i.e., it is an integrated process. The operator is used to re-define both the ARFIMA(0,d,0) and more general ARFIMA(p,d,q) processes in the time domain. A process $\{X_t\}$ is an ARFIMA(0,d,0) process if for all t: $(1-\mathcal{B})^d X_t = \varepsilon_t$. Likewise, a process $\{X_t\}$ is an ARFIMA(p,d,q) process if for all t: $\Phi(\mathcal{B})(1-\mathcal{B})^d X_t = \Theta(\mathcal{B})\varepsilon_t$, where Φ and Θ are given in Eqs. (A2) and (A3), respectively.

Finally, to connect back to our first definition of long memory, consider the ACV of the ARFIMA(0,d,0) process. By using the definition of spectral density to directly integrate Eq. (), and an alternative expression for $\pi_k^{(d)}$ in Eq. (A8)

$$\pi_k^{(d)} = \frac{1}{\Gamma(k+1)} \frac{\Gamma(k-d)}{\Gamma(-d)}, \tag{A9}$$

one can obtain the following representation of the ACV of the ARFIMA(0,d,0) process:

$$\gamma_d(k;\sigma) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(1+k-d)}. \tag{A10}$$

Because the parameter σ^2 is just a scalar multiplier, we may simplify notation by defining $\gamma_d(k) = \gamma_d(k;\ \sigma)/\sigma^2$, whereby $\gamma_d(\cdot) \equiv \gamma_d(\cdot;\ 1)$. Then the ACF is

$$\rho_d(k) = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(1+k-d)}, \tag{A11}$$

from which Stirling's approximation gives $\rho_d(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}$, confirming a power-law relationship for the ACF. Finally, note that Eq. (A9) can be used to represent ARFIMA(0,d,0) as an AR(∞) process, as $X_t + \sum_{k=1}^{\infty} \pi_k^{(d)} X_{t-k} = \varepsilon_t$. Furthermore, noting that in this case $\psi_k^{(d)} = \pi_k^{(-d)}$ leads to the following MA(∞) analog: $X_t = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)} \frac{\Gamma(k+d)}{\Gamma(d)} \varepsilon_{t-k}$.

Appendix B: Seasonal long-memory models

We define a seasonal differencing operator $(1-\mathcal{B}^s)$, as a natural extension to a SARFIMA processes by combining seasonal and non-seasonal fractional differencing operators (Porter-Hudak, 1990):

$$(1-\mathcal{B})^d \left(1-\mathcal{B}^s\right)^D X_t = \varepsilon_t.$$

The generalization to include both seasonal and non-seasonal short-memory components is obvious (Porter-Hudak, 1990):

$$\Phi^{(p)}(\mathcal{B})\Phi_s^{(P)}\left(\mathcal{B}^s\right)(1-\mathcal{B})^d\left(1-\mathcal{B}^s\right)^D X_t = \Theta^{(q)}(\mathcal{B})\Theta_s^{(Q)}\left(\mathcal{B}^s\right)\varepsilon_t.$$

Focusing on the first of these issues, Hosking (1981) considered generalising the ARFIMA(0,d,0) process in a different manner by retaining only one pole but at any given frequency in [0, π]. The model he suggested was later studied and popularized by Anděl (1986) and Gray et al. (1989, 1994), and became known as the "Gegenbauer process".

A process $\{X_t\}$ is a Gegenbauer (d; ω) process if for all t:

$$\left(1 - 2u\mathcal{B} + \mathcal{B}^2\right)^d X_t = \varepsilon_t, \tag{B1}$$

where $\omega = \cos^{-1} u$ is called the Gegenbauer frequency. The obvious extension to include short-memory components $\Phi^{(p)}$ and $\Theta^{(p)}$ is denoted GARMA(p,d,q;ω).

The term "Gegenbauer" derives from the close relationship to the Gegenbauer polynomials, a set of orthogonal polynomials useful in applied mathematics. The Gegenbauer polynomials are most usefully defined in terms of their generating function. The Gegenbauer polynomial on the order of k with parameter d, $G_k^{(d)}(\cdot)$ satisfies

$$\left(1 - 2uz + z^2\right)^{-d} \equiv \sum_{k=0}^{\infty} G_k^{(d)}(u) z^k. \tag{B2}$$

The spectral density function of the Gegenbauer(d;ω) process is (Gray et al., 1989)

$$f(\lambda) = \frac{\sigma^2}{2\pi} |2(\cos\lambda - \cos\omega)|^{-2d},\ 0 \le \lambda \le \pi.$$

Note that Gegenbauer(d;ω) processes possess a pole at the Gegenbauer frequency ω. Gegenbauer processes may be considered to be somewhat ambiguous in terms of long memory. Non-trivial (i.e., $\omega \neq 0$) Gegenbauer processes have bounded spectral density functions at the origin, and therefore do not have long memory according to our strict definition. Consequently a more general Gegenbauer process was developed: let $\mathbf{d} = (d_1, \ldots, d_k)$ and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)$, and for all j, $u_j = \cos\omega_j$ (assumed distinct). Then a process $\{X_t\}$ is a k-factor Gegenbauer($\mathbf{d}$;$\boldsymbol{\omega}$) process if for all t (Woodward et al., 1998):

$$\prod_{j=1}^{k}\left(1-2u_j\mathcal{B}+\mathcal{B}^2\right)^{d_j}X_t=\varepsilon_t. \tag{B3}$$

The spectral density function of the k-factor Gegenbauer($\mathbf{d};\boldsymbol{\omega}$) process is (Woodward et al., 1998)

$$f(\lambda)=\frac{\sigma^2}{2\pi}\prod_{j=1}^{k}\left|2\left(\cos\lambda-\cos\omega_j\right)\right|^{-2d_j},\ 0\le\lambda\le\pi.$$

Indeed, k-factor Gegenbauer models are very flexible, and include nearly all other seasonal variants of ARFIMA processes such as the flexible-seasonal ARFIMA (Hassler, 1994) and fractional ARUMA (Robinson, 1994; Giraitis and Leipus, 1995) processes. Importantly, they also includes SARFIMA processes (Reisen et al., 2006): a SARFIMA$(0,d,0)\times(0,D,0)_s$ process is equivalent to a $\left\lfloor\frac{s+2}{2}\right\rfloor$ factor Gegenbauer($\mathbf{d};\boldsymbol{\omega}$) process where:

$$\omega_j=\frac{2\pi(j-1)}{s},\ j=1,\ldots,k,$$

and $d_1=\frac{d+D}{2}$, $d_j=D$ for $j=2,\ldots,k$, unless s is even in which case $d_k=\frac{D}{2}$.

Although k-factor Gegenbauer models are very general, one particular sub-model is potentially very appealing. This is the two-factor model, with one pole at the origin and one at a non-zero frequency. In order to conform with notation for ARFIMA$(0,d,0)$ processes, we will slightly redefine this model: a process $\{X_t\}$ is a *simple two-frequency* Gegenbauer process with parameters d, D, and ω, denoted Gegenbauer$(d,D;\omega)_2$ if for all t:

$$\left(1-(2\cos\omega)\mathcal{B}+\mathcal{B}^2\right)^{D}(1-\mathcal{B})^{d}X_t=\varepsilon_t.$$

The Bayesian MCMC methodology developed here is easily extended to incorporate these seasonal fractional models. It is assumed that the frequency ω, or seasonal period s, is a priori known.

Acknowledgements. We thank one anonymous reviewer and M. Crucifix for their comments, which helped to improve this manuscript. C. L. E. Franzke is supported by the German Research Foundation (DFG) through the cluster of excellence CliSAP (EXC177), N. W. Watkins is supported by ONR NICOP grant N62909-15-1-N143, and both are supported by the Norwegian Research Council KLIMAFORSK project 229754. N. W. Watkins thanks the University of Potsdam for hospitality.

Edited by: Z. Toth
Reviewed by: M. Crucifix and another anonymous referee

References

Abry, P., Flandrin, P., Taqqu, M. S., and Veitch, D.: Self-similarity and long-range dependence through the wavelet lens, in: Theory and Applications of Long-Range Dependence, Birkhaeuser, Boston, USA, 527–556, 2003.

Adenstedt, R. K.: On Large-sample Estimation for the Mean of a Stationary Random Sequence, Ann. Stat., 2, 1095–1107, 1974.

Anděl, J.: Long Memory Time Series Models, Kybernetika, 22, 105–123, 1986.

Barnes, J. A. and Allan, D. W.: A statistical model of flicker noise, Proc. IEEE, 54, 176–178, 1966.

Beran, J.: Statistics for Long Memory Processes, Chapman & Hall, New York, 1994a.

Beran, J.: On a Class of M-Estimators for Gaussian Long-Memory Models, Biometrika, 81, 755–766, 1994b.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R.: Long Memory Processes, Springer, Heidelberg, 2013.

Box, G. E. P. and Jenkins, G. M.: Time Series Analysis, Forecasting and Control, Holden-Day, Wiley, Hoboken, USA, 1970.

Brockwell, P. J. and Davis, R. A.: Time Series: Theory and Methods, 2nd Edn., Springer, New York, 1991.

Bunde, A., Ludescher, J., Franzke, C., and Schellnhuber, H.-J.: How significant is West Antarctic warming?, Nat. Geosci., 7, 246–247, 2014.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: STL: A Seasonal-Trend Decomposition Procedure Based on Loess, J. Offic. Stat., 6, 3–73, 1990.

Eğri˙oğlu, E. and Günay, S.: Bayesian model selection in ARFIMA models, Expert Sys. Appl., 37, 8359–8364, 2010.

Ehlers, R. S. and Brooks, S. P.: Bayesian analysis of order uncertainty in ARIMA models, Tech. rep., Department of Statistics, Federal University of Paraná, http://www.est.ufpr.br/rt/ehlb06.pdf (last access: November 2015), 2006.

Ehlers, R. S. and Brooks, S. P.: Adaptive proposal construction for reversible jump MCMC, Scand. J. Stat., 35, 677–690, 2008.

Fraley, C., Leisch, F., Maechler, M., Reisen, V., and Lemonte, A.: `fracdiff`: Fractionally differenced ARIMA aka ARFIMA(p,d,q) models, R package version 1.4-1, http://CRAN.R-project.org/package=fracdiff (last access: November 2015), 2012.

Franzke, C. L. E.: Long-range Dependence and Climate Noise Characteristics of Antarctic Temperature Data, J. Climate, 23, 6074–6081, 2010.

Franzke, C. L. E.: Nonlinear trends, long-range dependence and climate noise properties of surface air temperature, J. Climate, 25, 4172–4183, 2012.

Franzke, C. L. E., Graves, T., Watkins, N. W., Gramacy, R. B., and Hughes, C.: Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity, Philos. T. Roy. Soc. A, 370, 1250–1267, 2012.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T.: `mvtnorm`: Multivariate Normal and t Distributions, R package version 0.9-9992, http://CRAN.R-project.org/package=mvtnorm (last access: November 2015), 2012.

Geweke, J. and Porter-Hudak, S.: The estimation and application of long-memory time series models, J. Time Ser. Anal., 4, 221–238, 1983.

Gil-Alana, L. A.: An Application of Fractional Integration to a Long Temperature Series, Int. J. Climatol., 23, 1699–1710, 2003.

Gil-Alana, L. A.: Cyclical long-range dependence and the warming effect in a long temperature time series, Int. J. Climatol., 28, 1435–1443, 2008.

Giraitis, L. and Leipus, R.: A generalized fractionally differencing approach in long-memory modeling, Lithuan. Math. J., 35, 53–65, 1995.

Granger, C. W. J. and Joyeux, R.: An Introduction to Long-memory Time Series Models and Fractional Differencing, J. Time Ser. Anal., 1, 15–29, 1980.

Graves, T.: A systematic approach to Bayesian inference for long memory processes, PhD thesis, University of Cambridge, Cambridge, UK, 2013.

Graves, T., Gramacy, R. B., Watkins, N. W., and Franzke, C.: A brief history of long memory, Tech. rep., The University of Chicago, http://arxiv.org/abs/1406.6018, last access: 4 November 2014.

Gray, H. L., Zhang, N.-F., and Woodward, W. A.: On generalized fractional processes, J. Time Ser. Anal., 10, 233–257, 1989.

Gray, H. L., Zhang, N.-F., and Woodward, W. A.: Correction to: Gray et al. (1989), J. Time Ser. Anal., 15, 561–562, 1994.

Green, P. J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika, 82, 711–732, 1995.

Haslett, J. and Raftery, A. E.: Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource, J. Roy. Stat. Soc. Ser. C, 38, 1–50, 1989.

Hassler, U.: (Mis)specification of long memory in seasonal time, J. Time Ser. Anal., 15, 19–30, 1994.

Holan, S., McElroy, T., and Chakraborty, S.: A Bayesian Approach to Estimating the Long Memory Parameter, Bayesian Analysis, 4, 159–190, 2009.

Hosking, J. R. M.: Fractional differencing, Biometrika, 68, 165–176, 1981.

Hosking, J. R. M.: Modeling persistence in hydrological time series using fractional differencing, Water Resour. Res., 20, 1898–1908, 1984.

Hsu, N.-J. and Breidt, F. J.: Bayesian Analysis of Fractionally Integrated ARMA with additive noise, J. Forecast., 22, 491–514, 2003.

Hurst, H. E.: Long-term storage capacity of reservoirs, T. Am. Soc. Civ. Eng., 116, 770–808, 1951.

Kärner, O.: Comment on Hurst Exponent, Geophys. Res. Lett., 28, 3825–3826, doi:10.1029/2001GL013104, 2001.

Kay, S. M. and Marple, S. L.: Spectrum analysis – A modern perspective, Proc. IEEE, 69, 1380–1419, 1981.

Ko, K. and Vannucci, M.: Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes, J. Stat. Plan. Inf., 136, 3415–3434, 2006a.

Ko, K. and Vannucci, M.: Bayesian Wavelet-Based Methods for the Detection of Multiple Changes of the Long Memory Parameter, IEEE T. Sig. Proc., 54, 4461–4470, 2006b.

Koop, G., Ley, E., Osiewalski, J., and Steel, M. F.: Bayesian analysis of long memory and persistence using ARFIMA models, J. Econ., 76, 149–169, 1997.

Liseo, B., Marinucci, D., and Petrella, L.: Bayesian semiparametric inference on long-range dependence, Biometrika, 88, 1089–1104, 2001.

Mandelbrot, B. B. and Van Ness, J. W.: Fractional Brownian Motions, Fractional Noises and Applications, SIAM Review, 10, 422–437, 1968.

Mandelbrot, B. B. and Wallis, J. R.: Noah, Joseph and operational hydrology, Water Resour. Res., 4, 909–918, 1968.

Manley, G.: Central England temperatures: monthly means 1659 to 1973, Q. J. Roy. Meteorol. Soc., 100, 389–405, 1974.

Marriott, J. M., Ravishanker, N., Gelfand, A. E., and Pai, J. S.: Bayesian analysis for ARMA processes: complete sampling based inference under exact likelihoods, in: Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner, edited by: Berry, D., Chaloner, K., and Geweke, J., Wiley, New York, 243–256, 1995.

McLeod, A. I., Yu, H., and Krougly, Z.: Algorithms for Linear Time Series Analysis: With R Package, J. of Statistical Software, http://www.jstatsoft.org/v23/i05/ (last access: November 2015), 2007.

Monahan, J. F.: A note on enforcing stationarity in autoregressive-moving average models, Biometrika, 71, 403–404, 1984.

Montanari, A., Rosso, R., and Taqqu, M. S.: A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan, Water Resour. Res., 36, 1249–1259, 2000.

Pai, J. S. and Ravishanker, N.: Bayesian analysis of autoregressive fractionally integrated moving-average processes, J. Time Ser. Anal., 19, 99–112, 1998.

Palma, W.: Long Memory Time Series, Wiley, Hoboken, USA, 2007.

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L.: Mosaic organization of DNA nucleotides, Phys. Rev. E, 49, 1685–1689, 1994.

Porter-Hudak, S.: An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates, J. Am. Stat. Assoc., 85, 338–344, 1990.

Reisen, V. A., Rodrigues, A. L., and Palma, W.: Estimation of seasonal fractionally integrated processes, Comput. Stat. Data Anal., 50, 568–582, 2006.

Robert, C. P.: Simulation of truncated normal variables, Stat. Comput., 5, 121–125, 1995.

Roberts, G. O. and Rosenthal, J. S.: Optimal Scaling for Various Metropolis–Hastings Algorithms, Stat. Sci., 16, 351–367, 2001.

Robinson, P. M.: Efficient Tests of Nonstationary Hypotheses, J. Am. Stat. Assoc., 89, 1420–1437, 1994.

Slezak, J. and Weron, A.: From physical linear systems to discrete-time series. A guide for analysis of the sampled experimental data, Phys. Rev. E, 91, 053302, doi:10.1103/PhysRevE.91.053302, 2015.

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S.: Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes, J. Time Ser. Anal., 25, 785–809, 2004.

Vyushin, D., Mayer, J., and Kushner, P.: `PowerSpectrum`: Spectral Analysis of Time Series, R package version 0.3, http://www.atmosp.physics.utoronto.ca/people/vyushin/mysoftware.html (last access: November 2015), 2009.

Watkins, N. W.: Bunched black (and grouped grey) swans: Dissipative and non-dissipative models of correlated extreme fluctuations in complex geosystems., Geophys. Res. Lett., 40, 1–9, doi:10.1002/GRL.50103, 2013.

Woodward, W. A., Cheng, Q. C., and Gray, H. L.: A k-factor GARMA long-memory model, J. Time Ser. Anal., 19, 485–504, 1998.

Wuertz, D.: `fArma`: ARMA Time Series Modelling, R package version 2160.77, http://CRAN.R-project.org/package=fArma (last access: November 2015), 2012.