

11-1-2006

# Bayesian Wavelet-Based Methods for the Detection of Multiple Changes of the Long Memory Parameter

Kyungduk Ko  
*Boise State University*

# Bayesian Wavelet-Based Methods for the Detection of Multiple Changes of the Long Memory Parameter

Kyungduk Ko and Marina Vannucci

**Abstract**—Long memory processes are widely used in many scientific fields, such as economics, physics, and engineering. Change point detection problems have received considerable attention in the literature because of their wide range of possible applications. Here we describe a wavelet-based Bayesian procedure for the estimation and location of multiple change points in the long memory parameter of Gaussian autoregressive fractionally integrated moving average models (ARFIMA( $p, d, q$ )), with unknown autoregressive and moving average parameters. Our methodology allows the number of change points to be unknown. The reversible jump Markov chain Monte Carlo algorithm is used for posterior inference. The method also produces estimates of all model parameters. Performances are evaluated on simulated data and on the benchmark Nile river dataset.

**Index Terms**—ARFIMA models, Bayesian inference, change point, reversible jump, wavelets.

## I. INTRODUCTION

LONG memory processes have been widely used in many fields, such as economics, finance, and telecommunications, to cite a few, to model characteristic phenomena. Data from long memory processes have the distinctive feature that dependencies between distant observations are not negligible. Common models for long memory behavior are the fractional Brownian motion and fractional Gaussian noise. Discrete-time long memory processes are the autoregressive, fractionally integrated, moving average models with fractional differencing parameter  $d$ , or ARFIMA( $p, d, q$ ), first introduced by Granger and Joyeaux [9] and Hosking [11]. For these models the value of the spectral density function goes to infinity as the frequency goes to zero, and classical time series methods for estimation and testing cannot be applied. Also, the full structure of the variance-covariance matrix makes inferential methods computationally expensive. In early stages, approximate maximum likelihood methods were used by Li and McLeod [17] and Fox and Taqqu [6]. Sowell [26] calculated the exact covariance matrix to compute the likelihood function under the assumption that the roots of the autoregressive polynomial are simple. Beran [1] investigated asymptotic sampling theory properties of exact and approximate maximum likelihood methods. As for Bayesian approaches, Pai and Ravishanker [21], [22] adopted the Metropolis algorithm

to estimate the model parameters, Koop *et al.* [16] used importance sampling with the exact form of the variance-covariance matrix of Sowell [26], and Ko and Vannucci [15] proposed wavelet-based estimation procedures. Also, Jensen [14] investigated Bayesian estimators of the long memory parameter in the context of stochastic volatility models.

While estimation of model parameters has been widely investigated, less work has been done in designing methods for change point analysis of the long memory parameter in ARFIMA( $p, d, q$ ) models. Change points in a given series result from unexpected changes in the physical process that generates the data. Estimation techniques for model parameters may be inaccurate when change points are not properly located. Beran and Terrin [2] proposed a test for detecting a single change in ARFIMA( $p, d, q$ ) models. Ray and Tsay [25] investigated Bayesian estimation procedures for multiple change points analysis of the mean level and the long memory parameter. They considered ARFIMA( $0, d, 0$ ) models only and used a time-dependent Kalman filter approach with a truncated moving average (MA) approximation to evaluate the likelihood. Their method allows accurate estimation only if the change points occur at the ends of prespecified data blocks. Moreover, they used a griddy Gibbs sampler algorithm to estimate the long memory parameter, a procedure that can lead to inaccurate estimates. Liu and Kao [18] also adopted griddy Gibbs sampler estimation algorithm but allowed the number of change points to vary with the other model parameters and used the reversible jump Markov chain Monte Carlo (MCMC) method of Green [10] for inference. These authors considered the special case of ARFIMA( $1, d, 0$ ) models with conditionally heteroscedastic GARCH(1, 1) innovations.

Here we propose a wavelet-based Bayesian method for the detection of multiple change points of the long memory parameter for ARFIMA( $p, d, q$ ) models. Wavelets, being self-similar, have a strong connection with long memory processes and have been previously employed in the estimation of time-varying long memory parameters; see Whitcher and Jensen [31] and Wang *et al.* [30]. We use discrete wavelet transform (DWT) as a tool to get a sparse form of the covariance matrix, therefore simplifying the calculation of the likelihood. Following Liu and Kao [18], we allow an unknown number of change points and implement a reversible jump MCMC algorithm to carry out posterior inference in the wavelet domain. We deal with general types of ARFIMA( $p, d, q$ ) models and design a Metropolis algorithm to sample the long memory parameters and the model parameters. Our method combines existing tools, such as DWT and reversible jump MCMC, in a novel manner to produce estimates of the number and locations of multiple change points. It also provides posterior estimates

Manuscript received June 26, 2004; accepted February 6, 2006. The work of M. Vannucci was supported in part by the National Science Foundation under CAREER award DMS-0093208. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anamitra Makur.

K. Ko is with the Department of Mathematics, Boise State University, Boise, ID 83725-1555, USA.

M. Vannucci is with the Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA (e-mail: ko@math.boisestate.edu; mvannucci@stat.tamu.edu).

Digital Object Identifier 10.1109/TSP.2006.881202

of the long memory parameters as well as the autoregressive (AR) and MA parameters. Results from Vannucci and Corradi [29] allow us to use the exact form of the variance-covariance matrix of ARFIMA models instead of approximated formulas as frequently done in wavelet-based estimation of long-memory processes. The reversible jump MCMC method we implement for posterior inference is largely similar to the one adopted by Liu and Kao [18]. However, unlike ours, their method is not wavelet-based.

This paper is organized as follows. Section II provides a brief introduction to ARFIMA models and to discrete wavelet transformations. Section III describes the change point model and the prior distributions. We also report there the derivations of the posterior distributions necessary for our inferential strategy. Section IV introduces the reversible jump MCMC procedure. Mathematical derivations of posterior and proposal ratios are reported in the Appendix. Section V describes the simulation study and the results on the Nile River dataset. Section VI provides some concluding remarks.

## II. PRELIMINARIES

### A. ARFIMA Processes

A long memory process is characterized by the slow decay in the autocovariance function of the type  $\gamma(\tau) = C_x \tau^{-\alpha}$ , where  $C_x > 0$  is a constant depending on the process,  $0 < \alpha < 1$ , and  $\tau$  is large. Fractional ARIMA( $p, d, q$ ) or ARFIMA( $p, d, q$ ), first introduced by Hosking [11] and Granger and Joyeux [9], are well-known examples of long memory processes. Let us first define the fractional difference operator  $(1-B)^d$ , where  $d \in (-0.5, 0.5)$ , as the binomial series expansion

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

where  $B$  is the backshift operator and where the coefficients are square summable

$$\binom{d}{j} (-1)^j = \frac{\Gamma(d+1)(-1)^j}{\Gamma(d-j+1)\Gamma(j+1)} = \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)}. \quad (2)$$

Here  $\Gamma(\cdot)$  denotes the gamma function, which is defined as  $\Gamma(x) = \int_0^{\infty} t^{x-1} \exp(-t) dt$  for  $x > 0$ , as  $\Gamma(x) = \infty$  for  $x = 0$ , and as  $\Gamma(x) = x^{-1} \Gamma(1+x)$  for  $x < 0$ ; see Brockwell and Davis [3]. For nonnegative integers  $p$  and  $q$ , a fractional ARIMA( $p, d, q$ ) process is defined by the stationary solution of the equation

$$\Phi(B)(1-B)^d x_t = \Theta(B) \varepsilon_t \quad (3)$$

where

$$\begin{aligned} \Phi(B) &= 1 + \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p \\ \Theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \end{aligned}$$

and  $\varepsilon_t$  is a Gaussian white noise process with zero mean and variance  $\sigma^2$ . Here we assume a finite mean  $\mu$ , that is  $\mu = 0$

without loss of generality. Differencing  $d$  times the process produces an ARMA( $p, q$ ) model. Fractional ARIMA processes are stationary and invertible for  $d < 0.5$ . They exhibit positive dependency between distant observations for  $0 < d < 0.5$  (long memory) and negative dependency for  $-0.5 < d < 0$  (intermediate memory), and reduce to short memory ARMA( $p, q$ ) processes for  $d = 0$ . A special class is given by the fractionally integrated processes obtained for  $p = 0$  and  $q = 0$ , also called fractionally differenced white noise, or  $I(d)$ , in that differencing  $d$  times produces a white noise process.

### B. Discrete Wavelet Transforms

Suppose we observe a time series as a realization of a random process and let us indicate the data vector as  $\mathbf{X} = (x_1, \dots, x_n)^T$  with  $n = 2^J$  and  $J$  a positive integer<sup>1</sup> denoting the scale of the data. A discrete wavelet transform [19] can be used to reduce the data to a set of wavelet coefficients. Although it operates via recursive applications of filters, for practical purposes a DWT of order  $r$  is often represented in matrix form as  $\mathbf{Z} = \mathcal{W}\mathbf{X}$ , with  $\mathcal{W}$  an  $n \times n$  orthogonal matrix of the form

$$\mathcal{W} = [\mathcal{W}_1^T, \mathcal{W}_2^T, \dots, \mathcal{W}_r^T, \mathcal{V}_r^T]^T \quad (4)$$

that decomposes the data into sets of coefficients

$$\mathbf{Z} = [\mathbf{z}_1^T, \mathbf{z}_2^T, \dots, \mathbf{z}_r^T, \mathbf{x}_r^T]^T \quad (5)$$

where  $\mathbf{z}_j = \mathcal{W}_j \mathbf{X}$  is of dimension  $n_j = n/2^j$ ,  $j = 1, 2, \dots, r$  with  $r \leq J$ ,  $\mathbf{x}_r = \mathcal{V}_r \mathbf{X}$  is of dimension  $n/2^r$ , and  $\sum_{j=1}^r n_j + n/2^r = n$ . Coefficients  $\mathbf{x}_r$  are scaling coefficients representing a coarser approximation of the data, while coefficients  $\mathbf{z}_1, \dots, \mathbf{z}_r$  are wavelet coefficients representing local features of the data at different resolution scales. An inverse transformation exists to reconstruct a set of data from its wavelet decomposition.

## III. BAYESIAN CHANGE POINT ANALYSIS

### A. The Model

We want to allow the long memory parameter  $d$  in (3) to change over time. Ray and Tsay [25] investigated models that allow random shifts of this parameter as  $d_t = d_0 + \sum_{j=1}^t \delta_j \beta_j = d_{t-1} \beta_t$ , where the  $\delta_t$ s are independent and identically distributed Bernoulli random variables and  $\beta_j$ s are random observations with known distribution. In their implementation the authors allowed shifts to occur only at time points  $t = b+1, 2b+1, \dots, n-b+1$  with  $b$  prespecified. Here we adopt a more flexible model that uses a latent discrete state variable to indicate the different states (or “regimes”) induced by the change points. This model has a simple and easy interpretation and it has become very popular in the change point literature; see, for example, [4] and references therein. According to this model, we have

$$\Phi(B)(1-B)^{d_t} x_t = \Theta(B) \varepsilon_t$$

<sup>1</sup>Using  $2^J$  points, with  $J$  integer, is not a real restriction and methods exist to overcome the limitation allowing wavelet transforms to be applied to any length of data.

where

$$d_t = \begin{cases} d_1, & \text{if } 1 \leq t < c_1 \\ d_2, & \text{if } c_1 \leq t < c_2 \\ \vdots & \vdots \\ d_{k+1}, & \text{if } c_k \leq t \leq n \end{cases}$$

with  $0 \leq k \leq k_{\max}$  and  $k_{\max}$ , the maximum number of changes of the long memory parameter, to be prespecified. The elements of the vector  $\mathbf{C}_k = (c_1, \dots, c_k)$  indicate the positions at which changes in the long memory parameter occur and those of  $\mathbf{d}_k = (d_1, d_2, \dots, d_{k+1})$  the values of the long memory parameter in the corresponding sub-intervals.

### B. Likelihood

We model wavelet coefficients, rather than the original data. The DWT is a linear and orthogonal transformation, and wavelet coefficients therefore inherit the distribution of the data—specifically, they are zero-mean Gaussian. Let  $\Psi = (\phi, \mathbf{d}_k, \theta)$ . We write the likelihood function in the wavelet domain as

$$f(\mathbf{z}|\phi, \mathbf{d}_k, \theta, \sigma^2) = (\sqrt{2\pi})^{-n} \prod_{i=1}^n \left( \frac{1}{\sigma^2 \sigma_{z_i}^2(\Psi)} \right)^{1/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^n \frac{z_i^2}{\sigma^2 \sigma_{z_i}^2(\Psi)} \right] \quad (6)$$

where  $z_i$  is the  $i$ th component of the vector  $\mathbf{z}$  of the wavelet coefficients. We compute the exact variances  $\sigma_{z_i}^2$  of the wavelet coefficients via a recursive algorithm that uses the DWT filters [29, Prop. 1, p. 974]. The algorithm has an interesting link to the two-dimensional discrete wavelet transform (DWT2) that makes computations simple. In the context of this paper, the variances  $\sigma_{z_i}^2$  can be computed as the diagonal elements of the matrix that results from applying the DWT2 to the theoretical variance–covariance matrix whose element values can be calculated based on the specific model under consideration and the values of the model parameters.

We use the following argument in order to line the wavelet coefficients up in time with their location in the original data domain; see, for example, Percival and Walden [24, p. 13], and Wang *et al.* [30]. There are  $n_j$  wavelet coefficients at scale  $j = 1, 2, \dots, r$  of the DWT and they correspond to the times  $(2m+1)2^{j-1}$ ,  $m = 0, 1, \dots, n_j - 1$ . We therefore line up all wavelet coefficients accordingly and use the reordered vector in the implementation of the MCMC procedure later described. Hereafter we assume that  $\mathbf{z} = (z_1, z_2, \dots, z_n)$  in (6) is the vector of lined up wavelet coefficients.

The advantage of using wavelets in our context is that the DWT serves as a tool to simplify the likelihood. Long memory data have, in fact, a dense covariance structure that makes the exact likelihood of the data difficult to handle; see, for example, Beran [1]. Simpler models, instead, can be used for wavelet coefficients. Decorrelation properties of the wavelets for long memory processes are, in fact, well documented in the literature. Tewfik and Kim [27] proved that the correlation between wavelet coefficients decreases exponentially fast across scales and hyperbolically fast along time. Jensen [12], [13] provides evidence that these rates of decay allow the DWTs to do a credible job at decorrelating the highly autocorrelated long memory processes. In addition, wavelets have an adaptive behavior in

the time domain in that the time support of the wavelet coefficients at high scales is smaller than the time support of those at lower scales. This makes wavelets well suitable for estimating time-varying parameters.

We want to emphasize that wavelet transforms do not decorrelate long memory data exactly and that (6) is indeed an approximate model. For further investigation, we looked into empirical measures of this approximation. We essentially applied the inverse of the Vannucci and Corradi algorithm to the variances of the wavelet coefficients only, i.e., to a diagonal matrix, therefore reconstructing an approximate covariance matrix in the data domain, and then computed reconstruction errors as mean squares ( $D_1$ ) and mean absolute deviations ( $D_2$ ) between the original and the reconstructed covariance matrices. These errors were typically very small. For example, we got  $D_1 = 0.0038$  and  $D_2 = 5.3 \times 10^{-5}$  for the mean absolute deviations and mean squares, respectively, for ARFIMA(1,  $d$ , 1) with  $d = 0.2$ ,  $\phi = 0.5$ ,  $\theta = -0.8$ , and  $D_1 = 0.0231$  and  $D_2 = 6.6 \times 10^{-4}$  for  $d = 0.4$ ,  $\phi = 0.1$ ,  $\theta = -0.8$ . We obtained similar results for different values of the parameters ( $d, \theta, \phi$ ). Overall, reconstructions from diagonal structures in the wavelet domain are reasonably close to the original ones, the only exceptions being extreme cases where  $d$  is close to 0.5 and the MA and AR parameters are close to the unitary root. These arguments provide further evidence for the whitening properties of the wavelets, suggesting that the approximation to uncorrelated coefficients can be reasonable, especially when considering the computational advantage implied by such approximation. Other evidence can be found in the fairly good detection performance of our method (see the applications section). These arguments, of course, strictly depend on the long memory structure and do not apply to other processes. Percival *et al.* [23] and Gabbanini *et al.* [7] have looked at wavelet packets as a way to decorrelate processes for which the standard DWTs fail, such as for short memory processes.

### C. Prior Distributions

We specify priors for the unknowns, i.e., the vectors  $\mathbf{d}_k$  and  $\mathbf{C}_k$ , the autoregressive coefficients  $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ , the moving average coefficients  $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ , the parameter  $k$ , and the error variance  $\sigma^2$ . Let  $\zeta_k = (k, \omega_k, \sigma^2) = (k, \mathbf{C}_k, \mathbf{d}_k, \phi, \theta, \sigma^2)$ , where  $\omega_k = (\mathbf{C}_k, \mathbf{d}_k, \phi, \theta)$ . We assume

$$\begin{aligned} \pi(k, \omega_k, \sigma^2) &= \pi(k) \pi(\omega_k, \sigma^2 | k) \\ &= \pi(k) \pi(\mathbf{C}_k) \pi(\mathbf{d}_k | k, \mathbf{C}_k) \pi(\phi) \pi(\theta) \pi(\sigma^2). \end{aligned}$$

We use uniform distributions over the set  $(0, 0.5)$  as the priors for  $d_k$ ,  $k = 0, \dots, k_{\max}$ , and uniform priors for the  $\phi$ s and  $\theta$ s defined in the ranges that satisfy the causality and invertibility properties of the ARMA process. We also specify a uniform distribution over the set of integers  $\{0, 1, 2, \dots, k_{\max}\}$  for the prior on  $k$ . As for the prior distribution of  $\mathbf{C}_k$ , suppose that  $k^*$  is generated at a particular iteration from  $U\{0, 1, 2, \dots, k_{\max}\}$ . Then the number of possible ways we can pick  $k^*$  change points out of  $n-2$  candidate locations is  $\binom{n-2}{k^*}$ . A reasonable choice for a prior distribution is therefore

$$\mathbf{C}_k | k \sim \binom{n-2}{k}^{-1}$$

which is noninformative. All the above priors are proper. We set an improper prior on  $\sigma^2$  as  $\pi(\sigma^2) \propto \sigma^{-2}$ .

#### D. Posterior Distributions

Given the likelihood (6) and the priors as described in Section III-C, the joint posterior distribution of all parameters given the data can be expressed as

$$\begin{aligned} & f(k, \boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta}, \mathbf{C}_k, \sigma^2 | \mathbf{z}) \\ & \propto f(\mathbf{z} | \boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta}, \sigma^2) \cdot \pi(k, \omega_k, \sigma^2) \\ & = \left( \frac{1}{\prod_{i=1}^n \sigma^2 \sigma_{z_i}^2(\Psi)} \right)^{1/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^n \frac{z_i^2}{\sigma^2 \sigma_{z_i}^2(\Psi)} \right] \\ & \quad \cdot (1 + k_{\max})^{-1} \cdot \binom{n-2}{k}^{-1} \cdot \sigma^{-2}. \end{aligned}$$

We treat  $\sigma^2$  as a nuisance parameter and integrate it out to get a proper marginal posterior distribution of  $\boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta}$  and  $\mathbf{C}_k$  as

$$\begin{aligned} f(k, \boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta}, \mathbf{C}_k | \mathbf{z}) & \propto \left( \frac{1}{\prod_{i=1}^n \sigma_{z_i}^2(\Psi)} \right)^{1/2} \\ & \quad \times \left[ \sum_{i=1}^n \frac{z_i^2}{\sigma_{z_i}^2(\Psi)} \right]^{-n/2} \cdot (1 + k_{\max})^{-1} \cdot \binom{n-2}{k}^{-1}. \end{aligned}$$

In order to implement our MCMC estimation procedure, we need the following full conditional distributions. Suppose we set the maximum number of change points to  $k$ , so that  $\mathbf{C}_k = \{c_1, c_2, \dots, c_{m-1}, c_m, c_{m+1}, \dots, c_k\}$ . The conditional posterior distribution of a change point  $c_m$  is

$$\begin{aligned} & P(c_m | \mathbf{z}, k, c_1, c_2, \dots, c_{m-1}, c_{m+1}, \dots, c_k, \Psi) \\ & \propto \left( \frac{1}{\prod_{i=c_{m-1}}^{c_m-1} \sigma_{z_i}^2(\phi, d_m, \theta)} \right)^{1/2} \left( \frac{1}{\prod_{i=c_m}^{c_{m+1}-1} \sigma_{z_i}^2(\phi, d_{m+1}, \theta)} \right)^{1/2} \\ & \quad \times \left[ \sum_{i=c_{m-1}}^{c_m-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_m, \theta)} \right]^{-(c_m - c_{m-1})/2} \\ & \quad \times \left[ \sum_{i=c_m}^{c_{m+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_{m+1}, \theta)} \right]^{-(c_{m+1} - c_m)/2} \quad (7) \end{aligned}$$

where  $c_m = c_{m-1} + 1, c_{m-1} + 2, \dots, c_{m+1} - 1, c_0 = 1, c_{k+1} = n$ , and  $m = 1, 2, \dots, k$ . This conditional distribution is used to update the change points.

The conditional joint posterior distribution of  $\Psi = (\boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta})$  is

$$\begin{aligned} P(\boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta} | \mathbf{z}, k, \mathbf{C}_k) & \propto \prod_{j=1}^{k+1} \left( \frac{1}{\prod_{i=c_{j-1}}^{c_j-1} \sigma_{z_i}^2(\phi, d_j, \theta)} \right)^{1/2} \\ & \quad \times \left[ \sum_{i=c_{j-1}}^{c_j-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_j, \theta)} \right]^{-(c_j - c_{j-1})/2}. \quad (8) \end{aligned}$$

This conditional distribution is used to implement a reversible jump MCMC procedure with an embedded Metropolis step as described in the next section.

## IV. REVERSIBLE JUMP MCMC

For inference, we adopt a reversible jump Markov chain Monte Carlo method, Green [10]. Reversible jump MCMC is a generalization of the Metropolis–Hastings algorithm to cases where model parameters have variable dimensions. In our model, the number of change points is not fixed a priori, and therefore the dimension of the parameter space  $\xi_k = \zeta_k - \{\sigma^2\}$  changes at each MCMC iteration. At each step the update of the model dimension and model parameters is carried out by computing an acceptance probability designed to ensure a detailed balance equation, which in turn allow the Markov chain to have a specified invariant distribution.

In this paper, the algorithm randomly evolves by choosing at each iteration one of three move steps, according to pre-specified probabilities:  $P(\text{birth of a change point}) = \alpha_k$ ,  $P(\text{death of a change point}) = \beta_k$ , and  $P(\text{transition of } \boldsymbol{\omega}_k) = \gamma_k$ , where  $\alpha_k + \beta_k + \gamma_k = 1$ . The three types of moves we use are as follows.

i) *Birth of a change point*: The birth step adds a new change point to the current parameter state  $\xi_k = (k, \omega_k)$ , where  $\omega_k = (\mathbf{C}_k, \mathbf{d}_k, \boldsymbol{\phi}, \boldsymbol{\theta})$ , that is, the number of change points increases from  $k$  to  $k+1$ . Let the locations of the  $(k+1)$  change points be  $\mathbf{C}'_{k+1} = (c'_1, c'_2, \dots, c'_{k+1})$  and the corresponding long memory parameters be  $\mathbf{d}'_{k+1} = (d'_1, d'_2, \dots, d'_{k+2})$ . Then the proposed parameter space becomes  $\xi'_{k+1} = (k+1, \boldsymbol{\omega}'_{k+1})$ , where  $\boldsymbol{\omega}'_{k+1} = (\mathbf{C}'_{k+1}, \mathbf{d}'_{k+1}, \boldsymbol{\phi}, \boldsymbol{\theta})$ . We accept a proposed transition from  $\xi_k$  to  $\xi'_{k+1}$  with probability shown in (9) at the bottom of the next page. See the Appendix for the derivation of this acceptance probability.

ii) *Death of a change point*: The death step drops one point among the existing change points, that is, the number of change points decreases from  $k$  to  $k-1$ . Let the proposed locations with  $(k-1)$  change points be  $\mathbf{C}^*_{k-1} = (c^*_1, c^*_2, \dots, c^*_{k-1})$  and the corresponding long memory parameters be  $\mathbf{d}^*_{k-1} = (d^*_1, d^*_2, \dots, d^*_k)$ . The resulting parameter space is  $\xi^*_{k-1} = (k-1, \boldsymbol{\omega}^*_{k-1})$ , where  $\boldsymbol{\omega}^*_{k-1} = (\mathbf{C}^*_{k-1}, \mathbf{d}^*_{k-1}, \boldsymbol{\phi}, \boldsymbol{\theta})$ . We accept a transition from  $\xi_k$  to  $\xi^*_{k-1}$  with probability shown in (10) at the bottom of the next page. See the Appendix for more details.

iii) *Transition of the model parameters*  $\omega_k = (\mathbf{C}_k, \mathbf{d}_k, \boldsymbol{\phi}, \boldsymbol{\theta})$ : For the transition step, we update the model parameters by sampling  $\mathbf{C}_k$  from the conditional distribution (7) and  $(\boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta})$  from the joint (8). For the latter we use a Metropolis algorithm with a multivariate normal distribution as proposal distribution, centered at the maximum likelihood estimates of the parameters and with covariance matrix given by the observed Fisher information. We use rejection sampling to take into account the restrictions on the values of the parameters.

In the design and implementation of the birth and death moves, we largely followed the work of Liu and Kao [18]. We report in the Appendix the necessary calculations to derive the acceptance probabilities of the two moves for our particular case. Also, following Green [10], we want to ensure  $\alpha_k \pi(k) = \beta_{k+1} \pi(k+1)$  so that

$$\alpha_k = c \cdot \min \left\{ 1, \frac{\pi(k+1)}{\pi(k)} \right\}, \beta_{k+1} = c \cdot \min \left\{ 1, \frac{\pi(k)}{\pi(k+1)} \right\}$$

where the constant  $c$  is as large as possible subject to  $\alpha_k + \beta_k \leq 0.9$  for all  $k = 0, 1, \dots, k_{\max}$ . Since in our prior setting  $\pi(k) = \pi(k+1) = 1/(k_{\max} + 1)$  for  $k$ ,  $\alpha_k = \beta_{k+1} = c \leq 0.9$ ,  $\alpha_k$ , and  $\beta_{k+1}$  can be set to 0.45. Therefore, we set up each probability as shown in the equation at the bottom of the page and use these probabilities to choose one of the three moves at each MCMC iteration.

## V. APPLICATIONS

### A. Simulation Study

There are a number of ways to generate a time series that exhibits long memory properties. A computationally simple one was proposed by McLeod and Hipel [20] and involves the Cholesky decomposition of the correlation matrix  $R_X(i, j) = [\rho(|i-j|)]$ . Given  $R_X = MM'$  with  $M = [m_{i,j}]$  a lower triangular matrix, if  $\epsilon_t$ ,  $t = 1, \dots, n$  is a Gaussian white noise series with zero mean and unit variance, then the series

$$X_t = \gamma_0^{1/2} \sum_{i=1}^t m_{t,i} \epsilon_i \quad (11)$$

will have autocorrelation  $\rho(\tau)$ . We used the McLeod and Hipel method to simulate data with  $\rho(\tau)$  computed as in Sowell [26]. We set  $\sigma^2 = 1$  and sample size  $n = 512$ .

$$\begin{aligned} & \alpha_{\text{birth}}(\xi_k, \xi_{k+1}^l) \\ &= \min(1, \text{posterior ratio} \times \text{proposal ratio} \times |J|) \\ &= \min \left\{ 1, \prod_{j=m^*-1}^{m^*} \frac{\left( \frac{\left[ \sum_{i=c_j^l}^{c_{j+1}^l-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_{j+1}^l, \theta)} \right]^{- (c_{j+1}^l - c_j^l)}}{\prod_{i=c_j^l}^{c_{j+1}^l-1} \sigma_{z_i}^2(\phi, d_{j+1}^l, \theta)} \right)^{1/2}}{\left( \frac{\left[ \sum_{i=c_j^l}^{c_{j+1}^l-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_{m^*}, \theta)} \right]^{- (c_{j+1}^l - c_j^l)}}{\prod_{i=c_j^l}^{c_{j+1}^l-1} \sigma_{z_i}^2(\phi, d_{m^*}, \theta)} \right)^{1/2}} \times \frac{4R_{u_1} \beta_{k+1}}{\alpha_k} \right\} \quad (9) \end{aligned}$$

$$\begin{aligned} & \alpha_{\text{death}}(\xi_k, \xi_{k-1}^*) \\ &= \min(1, \text{posterior ratio} \times \text{proposal ratio} \times |J|) \\ &= \min \left\{ 1, \prod_{j=u-1}^u \frac{\left( \frac{\left[ \sum_{i=c_j}^{c_{j+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_u^*, \theta)} \right]^{- (c_{j+1} - c_j)}}{\prod_{i=c_j}^{c_{j+1}-1} \sigma_{z_i}^2(\phi, d_u^*, \theta)} \right)^{1/2}}{\left( \frac{\left[ \sum_{i=c_j}^{c_{j+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_{j+1}, \theta)} \right]^{- (c_{j+1} - c_j)}}{\prod_{i=c_j}^{c_{j+1}-1} \sigma_{z_i}^2(\phi, d_{j+1}, \theta)} \right)^{1/2}} \times \frac{\alpha_{k-1}}{4R_u \beta_k} \right\} \quad (10) \end{aligned}$$

$$\begin{cases} \alpha_k = 0.0, \beta_k = 0.9, \text{ and } \gamma_k = 0.1, & \text{if } k = k_{\max} \\ \alpha_k = 0.9, \beta_k = 0.0, \text{ and } \gamma_k = 0.1, & \text{if } k = 0 \\ \alpha_k = 0.45, \beta_k = 0.45 \text{ and } \gamma_k = 0.1, & \text{if } k = 1, 2, 3, \dots, k_{\max} - 1 \end{cases}$$

We looked at both single and multiple change points. For the single change point case, we simulated datasets with a change at  $t = 257$ . For multiple changes, we induced changes at  $t = 129$  and  $t = 257$ . We simulated data from ARFIMA(1,  $d$ , 1), ARFIMA(1,  $d$ , 0), and ARFIMA(0,  $d$ , 1) models with one change point and from ARFIMA(1,  $d$ , 1) with two change points. We used  $\phi = 0.1, 0.3$ ,  $\theta = 0.3, 0.4, 0.5$ , and  $d = 0.05, 0.2, 0.3, 0.35, 0.4, 0.45$ . In all analyses we used discrete wavelet transforms with Daubechies's minimum phase wavelets with seven vanishing moments [5].

We set the initial value for the number of change points  $k$  to zero and the maximum number of change points allowed  $k_{\max}$  to three. For MCMC sampling, we used maximum likelihood estimates as initial values for  $\phi$ ,  $\theta$ , and  $d$  and the inverse of Fisher's information matrix as the variance-covariance matrix of the multivariate normal proposal distribution for the Metropolis step. We used 20 000 iterations with a burn-in of 10 000. At each iteration, a move type is randomly chosen among moves i)–iii), as described in Section IV, and the values of new parameters are obtained according to the chosen move type. When sampling  $c_m$  from (7), only the  $c$ s that changed during the reversible jump step are recalculated and standardized so to sum up to one. Then only one point among the candidate points is sampled from a multinomial distribution with the standardized values as multinomial probabilities and passed to the next step.

We assessed convergence by looking at plots of the autocorrelation functions of the sampled parameters. Figs. 1 and 2 show these plots for two of the models we simulated. We also computed Geweke's  $z_G$  as the standardized difference between the initial and final portions of the MCMC chains. According to Geweke [8], the  $z_G$  statistic is distributed as a standard Gaussian  $N(0, 1)$  if the chain has reached convergence. Therefore, values that are less than approximately  $3\sigma$  for all parameters indicate convergence. Values for our chains are reported in Tables II–V, together with the estimates of the parameters. Correlation plots and  $z_G$  values for all simulated cases were consistent with convergence to the stationary distribution. In particular,  $z_G$  values in our results are all smaller than 2.42 and therefore are consistent with convergence of the chains ( $\alpha = 0.01$ ).

The MCMC algorithm produces a sample on all parameters, from which the number of change points  $k$  can be estimated first, as the posterior mode, denoted  $\hat{k}$ . Inference on the model parameters is then obtained conditionally on  $\hat{k}$ . Table I, second column, shows the estimates of the number of change points  $k$  for an ARFIMA(1,  $d$ , 1) with  $\phi = 0.1$  and  $\theta = 0.5$  and one change point in the long memory parameter at  $c = 257$  with  $d_1 = 0.2$  and  $d_2 = 0.4$ . Table II reports the posterior estimates and posterior confidence intervals for the model parameters and change points corresponding to  $\hat{k} = 1$ , the most probable value from Table I. Table I, third column, and Table III show results for ARFIMA(1,  $d$ , 0) with  $\phi = 0.3$ ,  $d_1 = 0.2$ ,  $d_2 = 0.3$ , and one change point at  $c = 257$ . Table I, fourth column, and Table IV are for an ARFIMA(0,  $d$ , 1) with  $\theta = 0.3$ ,  $d_1 = 0.2$ ,  $d_2 = 0.3$ , and one change point at  $c = 257$ . Finally, Table I, fifth column, and Table V refer to an ARFIMA(1,  $d$ , 1) with  $\phi = 0.1$ ,  $\theta = 0.4$

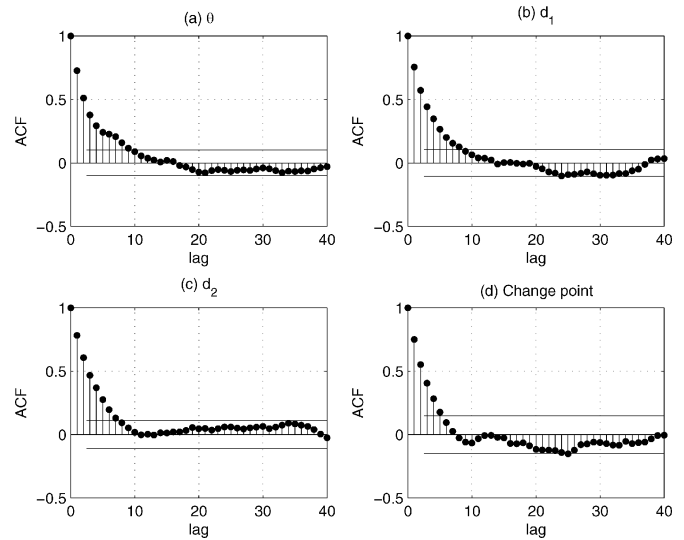


Fig. 1. ARFIMA(0,  $d$ , 1): Autocorrelation functions of sampled parameters.

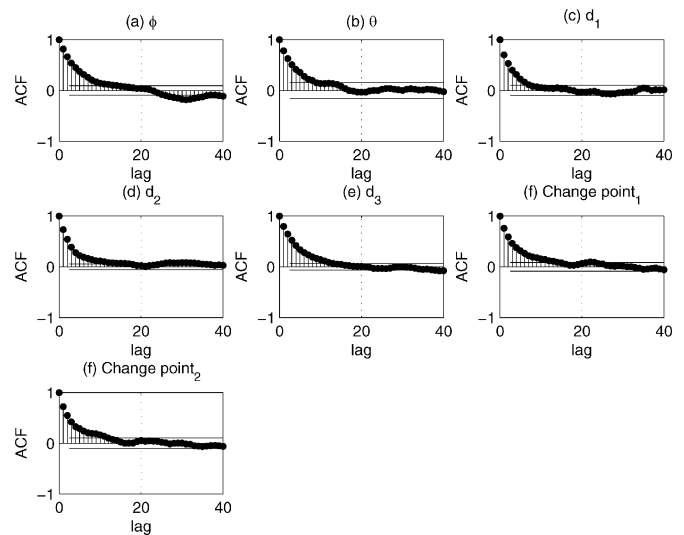


Fig. 2. ARFIMA(1,  $d$ , 1) with two change points: Autocorrelation functions of sampled parameters.

and two change points, at  $c_1 = 129$  and  $c_2 = 257$  with  $d_1 = 0.45$ ,  $d_2 = 0.35$ , and  $d_3 = 0.05$ .

Results are overall encouraging. The number of change points is correctly estimated in all cases considered. The method also shows good performance in the estimation of the model parameters and in the location of the change points, although there appears to be a slight overestimation of the location for ARFIMA(0,  $d$ , 1) with one change point and for the multiple change point case.

Finally, we tested the performance of the method in the case of a long memory model with a constant  $d$  by simulating data from an ARFIMA(0,  $d$ , 0) with  $d = 0.25$  and no change point. Our method correctly estimated  $\hat{k} = 0$ .

## B. Nile River Dataset

Probably the most well known example of a time series that exhibits long memory behavior is the Nile River minimum water levels dataset [28]. This time series consists of 663 yearly values, from A.D. 622 to A.D. 1284; see Fig. 3. Beran [1, p.

TABLE I  
ARFIMA(1, d, 1), ARFIMA(1, d, 0), ARFIMA(0, d, 1) AND ARFIMA(1, d, 1) (SEE TEXT FOR PARAMETER VALUES): POSTERIOR ESTIMATES OF THE NUMBER OF CHANGE POINTS  $k$

$\hat{k}$	ARFIMA(1, d, 1)	ARFIMA(1, d, 0)	ARFIMA(0, d, 1)	ARFIMA(1, d, 1)
	$\pi(\hat{k} \cdot)$	$\pi(\hat{k} \cdot)$	$\pi(\hat{k} \cdot)$	$\pi(\hat{k} \cdot)$
0	0.0251	0.3378	0.2295	0.0269
1	0.4937	0.4258	0.4223	0.4196
2	0.3375	0.1541	0.2565	0.5108
3	0.1437	0.0823	0.0917	0.0427

TABLE II  
ARFIMA(1, d, 1) WITH  $\phi = 0.1, d_1 = 0.2, d_2 = 0.4, \theta = 0.5$  AND ONE CHANGE POINT AT  $c = 257$ : PARAMETER ESTIMATES FOR  $\hat{k} = 1$ , TOGETHER WITH GEWEKE'S MEASURE FOR CONVERGENCE DIAGNOSTIC

$\hat{k}$	$\pi(\hat{k} \cdot)$	estimate (90% posterior interval)	$z_G$
1	0.4937	$\hat{\phi}=0.1649$ (0.0321, 0.3547)	1.880
		$\hat{\theta}=0.5226$ (0.3880, 0.6474)	1.684
		$\hat{d}_1=0.1083$ (0.0175, 0.2427)	1.986
		$\hat{d}_2=0.4187$ (0.2735, 0.4823)	1.870
		$\hat{c}_1=251$	2.051

TABLE III  
ARFIMA(1, d, 0) WITH  $\phi = 0.3, d_1 = 0.2, d_2 = 0.3$ , AND ONE CHANGE POINT AT  $c = 257$ : PARAMETER ESTIMATES FOR  $\hat{k} = 1$ , TOGETHER WITH GEWEKE'S MEASURE

$\hat{k}$	$\pi(\hat{k} \cdot)$	estimate (90% posterior interval)	$z_G$
1	0.4258	$\hat{\phi}=0.2462$ (0.1385, 0.3692)	1.216
		$\hat{d}_1=0.1362$ (0.0201, 0.3233)	-1.543
		$\hat{d}_2=0.3282$ (0.1008, 0.4627)	0.381
		$\hat{c}_1=252$	1.847

TABLE IV  
ARFIMA(0, d, 1) WITH  $\theta = 0.3, d_1 = 0.2, d_2 = 0.3$ , AND ONE CHANGE POINT AT  $c = 257$ : PARAMETER ESTIMATES FOR  $\hat{k} = 1$ , TOGETHER WITH GEWEKE'S MEASURE

$\hat{k}$	$\pi(\hat{k} \cdot)$	estimate (90% posterior interval)	$z_G$
1	0.4223	$\hat{\theta}=0.2900$ (0.1725, 0.3499)	-0.890
		$\hat{d}_1=0.1082$ (0.0035, 0.2873)	1.431
		$\hat{d}_2=0.3853$ (0.1824, 0.4932)	2.424
		$\hat{c}_1=266$	2.012

TABLE V  
ARFIMA(1, d, 1) WITH  $\phi = 0.1, \theta = 0.4, d_1 = 0.45, d_2 = 0.35, d_3 = 0.05$ , AND TWO CHANGE POINTS AT  $c_1 = 129$  AND  $c_2 = 257$ : PARAMETER ESTIMATES FOR  $\hat{k} = 2$ , TOGETHER WITH GEWEKE'S MEASURE

$\hat{k}$	$\pi(\hat{k} \cdot)$	estimate (90% posterior interval)	$z_G$
2	0.5108	$\hat{\phi}=0.0836$ (0.0451, 0.1939)	1.276
		$\hat{\theta}=0.3852$ (0.3294, 0.4963)	1.483
		$\hat{d}_1=0.4212$ (0.2922, 0.4768)	1.863
		$\hat{d}_2=0.3128$ (0.1278, 0.4584)	1.725
		$\hat{d}_3=0.0923$ (0.0226, 0.2125)	2.109
		$\hat{c}_1=138$	2.088
		$\hat{c}_2=269$	1.954

117–118] showed that the data can be modeled as fractionally differenced process  $I(d)$  and obtained an estimate of  $d = 0.4$  using Whittle's approximate maximum likelihood approach. However, he suspected a change of the long memory parameter in the time series. Looking at the plot of the data, he noticed that the first part of the series seemed to fluctuate much more independently than the subsequent measurements. Beran and Terrin [2] tested this hypothesis by segmenting the time series into six consecutive disjoint sets of observations of length 100 and estimating  $d$  on each set. Their estimated values of  $d$  were 0.0433 for the first 100 years and 0.3531, 0.3652, 0.3281, 0.3435, and 0.4354 for the subsequent sets. Ray and Tsay [25] reported an estimate of  $\hat{d} = 0.05$  for the first 100 years and of  $\hat{d} = 0.45$  afterward. These results were all obtained by testing for change points at prespecified locations.

We used the Nile River yearly minima from A.D. 622 to A.D. 1134 ( $n = 512$ ). Data are displayed in plot Fig. 3(a). We fitted an ARFIMA(0,  $d_t$ , 0) model, allowing the number of change points and locations to be unknown. Posterior estimates for  $k$  were (0.3218, 0.3917, 0.1923, 0.0942) for  $k = 0, 1, 2, 3$ , respectively. The posterior mode was therefore at  $\hat{k} = 1$ , suggesting the existence of one change point in the data. Table VI reports the estimates of the model parameters and of the location of the change point for  $\hat{k} = 1$ . Our method puts the change at the year A.D. 726, slightly later than suggested by Beran and Terrin [2]. Our estimates of the long memory parameter before and after

the estimated change point are  $\hat{d}_1 = 0.0547$  and  $\hat{d}_2 = 0.4235$ , respectively. Our estimate of  $d$  for  $\hat{k} = 0$  was  $\hat{d} = 0.3827$ , i.e., a value close to Beran and Terrin's estimate of 0.40 under the assumption that there is no change point. This value is also very close to the average estimate, 0.3651, for the five consecutive disjoint segments after the first 100 years, as found by Beran and Terrin [2].

Bayesian analysis allows inference on model parameters via full posterior distributions. Our inference is summarized in Fig. 3(b) and (c), where kernel estimates of the posterior densities of  $d$  before and after A.D. 726, the estimated change point, are shown, respectively. Fig. 3(d) shows the trace of the location of the change point. Fig. 4 shows plots of the autocorrelation functions for the sampled parameters. Values of Geweke's  $z_G$  statistic are reported in Table VI and are consistent with convergence to the stationary distribution.

When setting the prior on the long memory parameter  $d$ , we have restricted ourselves to the long memory range given by the interval (0, 0.5). This may result in an overestimate of the parameter when  $d$  is close to zero. Upon suggestion from one of the referees, we reran this application with a uniform prior on the interval (-0.5, 0.5). Our method still favored one change point, although the estimate of  $d_1$  was  $\hat{d}_1 = -0.0173$  (-0.1545, 0.1037). Estimates of the other model parameters were very close to those reported in Table VI.



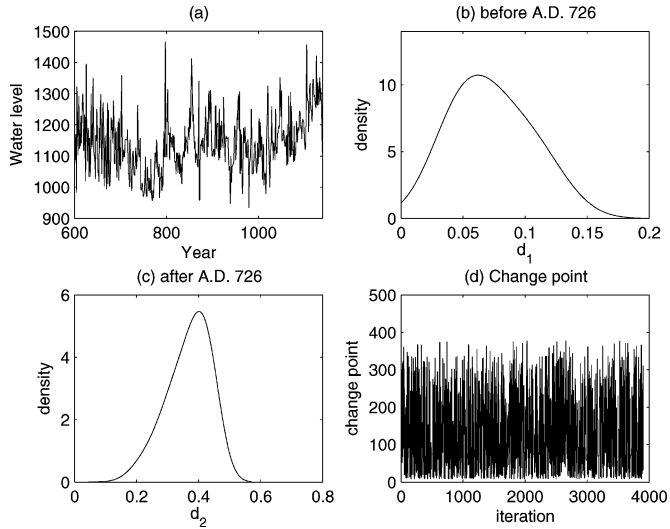


Fig. 3. Nile River data: (a) Original data and kernel estimates of the posterior density of  $d$  (b) before A.D. 726 and (c) after A.D. 726. (d) shows the MCMC trace of the location of the change point.

TABLE VI  
NILE RIVER DATASET: ESTIMATES OF THE ARFIMA(0,  $d$ , 0) MODEL  
PARAMETERS AND CHANGE POINT  $c$  FOR  $k = 1$ , TOGETHER WITH GEWEKE'S  
MEASURE

$k$	$\pi(\hat{k} \cdot)$	estimate (90% posterior interval)	$z_G$
		$\hat{d}_1=0.0547$ (0.0072, 0.1347)	0.442
1	0.3917	$\hat{d}_2=0.4235$ (0.2667, 0.4901)	1.382
		$\hat{c}_1=104$ (A.D. 726)	-1.928

## VI. CONCLUDING REMARKS

We have proposed a wavelet-based Bayesian method for the detection of multiple change points of the long memory parameter for ARFIMA( $p, d, q$ ) models. We have allowed an unknown number of change points and have implemented a reversible jump MCMC algorithm to carry out posterior inference in the wavelet domain. Our method combines existing tools, such as DWT and reversible jump MCMC, in a novel manner to produce estimates of the number and locations of multiple change points in the long memory parameter. It also provides posterior estimates of the long memory parameters as well as the AR and MA parameters.

In the simulation study, we have used Daubechies wavelets with seven vanishing moments. In our previous work we found that wavelets with high degrees of regularity produce slightly better estimates of the long memory parameter for large sample sizes [15]. Wavelets with higher numbers of vanishing moments ensure wavelet coefficients approximately uncorrelated. On the other hand, the support of the wavelets increases with the regularity and boundary effects may arise in the DWT, so that a tradeoff is often necessary.

For inference we have mainly focused on the long memory parameter and the autoregressive and moving average parameters. Our inferential procedure, however, could be generalized to include inference on other model parameters, such as the noise variance and the process mean, either by implementing a full Gibbs or by using a Rao-Blackwellization procedure. This

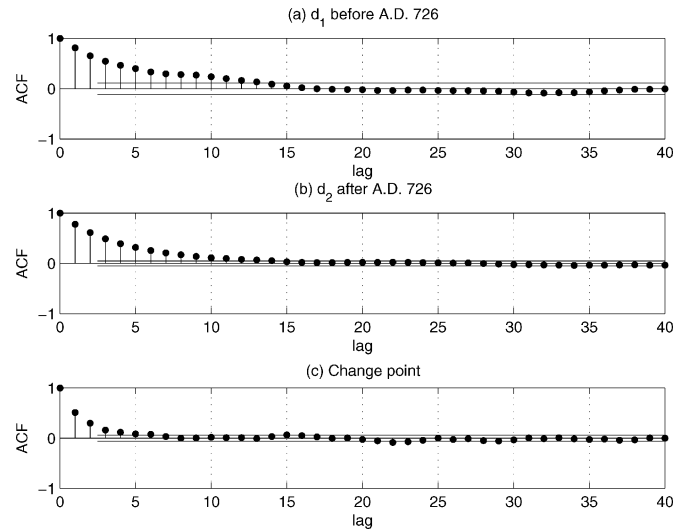


Fig. 4. Nile River data: Autocorrelation functions of sampled parameters.

would also make the investigation of prediction performances of the estimated models possible, for example, by approximating the predictive distribution via Monte Carlo integration using the output from the MCMC sampler; see, for example, Pai and Ravishanker [21], [22]. This is currently under investigation. Finally, while our method has focused on changes in the long memory parameter only, extensions may be possible that include changes in the model parameters.

## APPENDIX

### BIRTH AND DEATH MOVES FOR REVERSIBLE JUMP MCMC

#### A. Birth of a Change Point

The birth step adds a new change point to the current parameter state  $\xi_k = (k, \omega_k)$ , that is, the number of change points increases from  $k$  to  $k+1$ . Two uniform random variables  $u_1$  and  $u_2$  are generated to match the dimension of the parameters such that  $u_1 \sim U\{2, 3, \dots, n-1\} \cap \{c_1, c_2, \dots, c_k\}^c$ , where  $\{c_1, \dots, c_k\}$  is the set of existing change points and  $u_2 \sim U(-R_{u_1}, R_{u_1})$ , with  $R_{u_1} = \min\{d_{m^*-1}, 0.5 - d_{m^*-1}\}$  and  $m^* = \min\{i | u_1 < c_i\}$ .

Let the locations of the  $(k+1)$  change points be  $\mathbf{C}'_{k+1} = (c'_1, c'_2, \dots, c'_{k+1})$  and the corresponding long memory parameters be  $\mathbf{d}'_{k+1} = (d'_1, d'_2, \dots, d'_{k+2})$ , where

$$d'_i = \begin{cases} c_i, & \text{if } 1 \leq i \leq m^* - 1 \\ u_1, & \text{if } i = m^* \\ c_{i-1}, & \text{if } m^* + 1 \leq i \leq k + 1 \end{cases}$$

$$d'_i = \begin{cases} d_i, & \text{if } 1 \leq i \leq m^* - 1 \\ d_{m^*} - u_2, & \text{if } i = m^* \\ d_{m^*} + u_2, & \text{if } i = m^* + 1 \\ d_{i-1}, & \text{if } m^* + 2 \leq i \leq k + 2. \end{cases}$$

The proposed parameter space becomes  $\xi'_{k+1} = (k+1, \omega'_{k+1})$ , where  $\omega'_{k+1} = (\mathbf{C}'_{k+1}, \mathbf{d}'_{k+1}, \phi, \theta)$ . Let  $T(d_{m^*}, u_2) = (d_{m^*} -$

$u_2, d_{m^*} + u_2$ ). The Jacobian for the acceptance probability for birth step is

$$|J| = \left| \frac{\partial T(d_{m^*}, u_2)}{\partial (d_{m^*}, u_2)} \right| = \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} = 2.$$

The posterior ratio is shown in the first equation at the bottom of the page. The proposal ratio can be written as

$$\text{proposal ratio}_{\{k \rightarrow (k+1)\}} = \frac{\beta_{k+1} \cdot (k+1)^{-1}}{\alpha_k \{2R_{u_1}(n-k-2)\}^{-1}} = \frac{2R_{u_1} \beta_{k+1} (n-k-2)}{\alpha_k (k+1)}.$$

The denominator in the above formula comes from the fact that when a transition is made from  $\xi_k$  to  $\xi'_{k+1}$ , there are three components that need to be computed: the probability of a birth  $\alpha_k$ , the probability that we choose one candidate point for a birth among the  $(n-k-2)$  points, and the probability to split the long memory parameter into two. The numerator consists of the probability of a death  $\beta_{k+1}$  and the probability of dropping one point among the existing  $(k+1)$  points.

The acceptance rate for the birth of a change point is the product of the above three components: the posterior ratio, the proposal ratio, and the Jacobian. This completes our derivation of (9).

**B. Death of a Change Point**

The death step drops one point among the existing change points, that is, the number of change points decreases from  $k$  to  $k-1$ . A uniform random variable  $u$  is generated to match the dimension of the parameters in the range  $\{c_1, c_2, \dots, c_k\}$ . Let the proposed locations with  $(k-1)$  change points be  $\mathbf{C}_{k-1}^* =$

$(c_1^*, c_2^*, \dots, c_{k-1}^*)$  and the corresponding long memory parameters be  $\mathbf{d}_{k-1}^* = (d_1^*, d_2^*, \dots, d_k^*)$ , where

$$c_i^* = \begin{cases} c_i, & \text{if } 1 \leq i \leq u-1 \\ c_{i+1}, & \text{if } u \leq i \leq k-1 \end{cases}$$

and

$$d_i^* = \begin{cases} d_i, & \text{if } 1 \leq i \leq u-1 \\ \frac{(d_u + d_{u+1})}{2}, & \text{if } i = u \\ d_{i+1}, & \text{if } u+1 \leq i \leq k. \end{cases}$$

The resulting parameter space is  $\xi_{k-1}^* = (k-1, \boldsymbol{\omega}_{k-1}^*)$ , where  $\boldsymbol{\omega}_{k-1}^* = (\mathbf{C}_{k-1}^*, \mathbf{d}_{k-1}^*, \phi, \theta)$ . Let  $T^*(d_u, d_{u+1}) = ((d_u + d_{u+1})/2, d_u)$ . Then the Jacobian for the acceptance probability is

$$|J| = \left| \frac{\partial T^*(d_u, d_{u+1})}{\partial (d_u, d_{u+1})} \right| = \begin{vmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{vmatrix} = \frac{1}{2}.$$

The posterior ratio is shown in the last equation at the bottom of the page. The proposal ratio can be written as

$$\text{proposal ratio}_{\{k \rightarrow (k-1)\}} = \frac{\alpha_{k-1} \{2R_u(n-k-1)\}^{-1}}{\beta_k k^{-1}} = \frac{\alpha_{k-1} k}{2R_u \beta_k (n-k-1)}$$

where  $R_u = \min\{d_u^*, 0.5 - d_u^*\}$ .

The acceptance rate for the death of one of the existing change points is the product of the above three components: the posterior ratio, the proposal ratio, and the Jacobian. This completes our derivation of (10).

ACKNOWLEDGMENT

The authors wish to thank the Associate Editor and referees for the thorough review process and in particular one of the ref-

$$\text{posterior ratio}_{\{k \rightarrow (k+1)\}} = \frac{f(k+1, \boldsymbol{\phi}, \mathbf{d}'_{k+1}, \boldsymbol{\theta}, \mathbf{C}'_{k+1} | \mathbf{z})}{f(k, \boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta}, \mathbf{C}_k | \mathbf{z})} \propto \prod_{j=m^*-1}^{m^*} \frac{\left( \frac{1}{\prod_{i=c'_j}^{c'_{j+1}-1} \sigma_{z_i}^2(\phi, d'_{j+1}, \theta)} \right)^{1/2} \left[ \sum_{i=c'_j}^{c'_{j+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d'_{j+1}, \theta)} \right]^{-(c'_{j+1}-c'_j)/2}}{\left( \frac{1}{\prod_{i=c_j}^{c_{j+1}-1} \sigma_{z_i}^2(\phi, d_i, \theta)} \right)^{1/2} \left[ \sum_{i=c_j}^{c_{j+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_i, \theta)} \right]^{-(c_{j+1}-c_j)/2}} \cdot (k+1)$$

$$\text{posterior ratio}_{\{k \rightarrow (k-1)\}} = \frac{f(k-1, \boldsymbol{\phi}, \mathbf{d}_{k-1}^*, \boldsymbol{\theta}, \mathbf{C}_{k-1}^* | \mathbf{z})}{f(k, \boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta}, \mathbf{C}_k | \mathbf{z})} \propto \prod_{j=u-1}^u \frac{\left( \frac{1}{\prod_{i=c_j}^{c_{j+1}-1} \sigma_{z_i}^2(\phi, d_u^*, \theta)} \right)^{1/2} \left[ \sum_{i=c_j}^{c_{j+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_u^*, \theta)} \right]^{-(c_{j+1}-c_j)/2}}{\left( \frac{1}{\prod_{i=c_j}^{c_{j+1}-1} \sigma_{z_i}^2(\phi, d_{j+1}, \theta)} \right)^{1/2} \left[ \sum_{i=c_j}^{c_{j+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\phi, d_{j+1}, \theta)} \right]^{-(c_{j+1}-c_j)/2}} \cdot k$$

erees for pointing out a technical mistake in the MCMC procedure.

#### REFERENCES

- [1] J. Beran, *Statistics for Long-Memory Processes*. New York: Chapman and Hall, 1994.
- [2] J. Beran and N. Terrin, "Testing for a change of the long-memory parameter," *Biometrika*, vol. 83, pp. 627–638, 1996.
- [3] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1991.
- [4] S. Chib, "Estimation and comparison of multiple change-point models," *J. Econometrics*, vol. 86, pp. 221–241, 1998.
- [5] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992, vol. 61, CBMS-NSF Conference Series.
- [6] R. Fox and M. S. Taqqu, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Ann. Statist.*, vol. 14, pp. 517–532, 1986.
- [7] F. Gabbanini, M. Vannucci, G. Bartoli, and A. Moro, "Wavelet packet methods for the analysis of variance of time series with application to crack widths on the Brunelleschi dome," *J. Comp. Graph. Statist.*, vol. 13, no. 3, pp. 639–658, 2004.
- [8] J. Geweke, "Evaluating the accuracy of sampling-based approaches to the calculation of margin densities," in *Bayesian Statist.*, J. M. Bernardo, J. O. Berger, J. O. Dawd, and A. F. M. Smith, Eds. Oxford, U.K.: Oxford Univ. Press, 1992, vol. 4, pp. 169–193.
- [9] C. W. Granger and R. Joyeux, "An introduction to long memory time series models and fractional differencing," *J. Time Series Anal.*, vol. 1, pp. 15–29, 1980.
- [10] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.
- [11] J. R. M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, pp. 165–176, 1981.
- [12] M. Jensen, "Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter," *J. Forecast.*, vol. 18, pp. 17–32, 1999.
- [13] —, "An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets," *J. Econ. Dyn. Contr.*, vol. 24, no. 3, pp. 361–386, 2000.
- [14] —, "Semiparametric Bayesian inference of long-memory stochastic volatility models," *J. Time Series Anal.*, vol. 25, no. 6, pp. 895–922, 2004.
- [15] K. Ko and M. Vannucci, "Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes," *J. Statist. Plan. Infer.*, vol. 136, pp. 3415–3434.
- [16] G. Koop, E. Ley, J. Osiewalski, and M. F. J. Steel, "Bayesian analysis of long memory and persistence using ARFIMA models," *J. Econometrics*, vol. 76, pp. 149–169, 1997.
- [17] W. K. Li and A. I. McLeod, "Fractional time series modelling," *Biometrika*, vol. 73, pp. 217–221, 1986.
- [18] S. I. Liu and M. H. Kao, "Bayesian Analysis for Multiple Changes of the Long Memory Parameter", Tech Rep. National Central Univ., Graduate Inst. of Statistics, Chung-Li, Taiwan, R.O.C., 1999.
- [19] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.
- [20] A. I. McLeod and K. W. Hipel, "Preservation of the rescaled adjusted range, parts 1, 2 and 3," *Water Resources Res.*, vol. 14, pp. 491–512, 1978.
- [21] J. S. Pai and N. Ravishanker, "Bayesian modelling of ARFIMA processes by Markov chain Monte Carlo methods," *J. Forecast.*, vol. 15, pp. 63–82, 1996.
- [22] —, "Bayesian analysis of autoregressive fractionally integrated moving-average processes," *J. Time Series Anal.*, vol. 19, no. 1, pp. 99–112, 1998.
- [23] D. B. Percival, S. Sardy, and A. C. Davison, "Wavestrapping time series: Adaptive wavelet-based bootstrapping," in *Nonlinear and Nonstationary Signal Processing*, W. J. Fitzgerald, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [24] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. New York: Cambridge Univ. Press, 2000.
- [25] B. K. Ray and R. S. Tsay, "Bayesian methods for change-point detection in long-range dependent processes," *J. Time Series Anal.*, vol. 23, pp. 687–705, 2002.
- [26] F. Sowell, "Maximum likelihood estimation of stationary univariate fractionally integrated time series models," *J. Econometrics*, vol. 53, pp. 165–188, 1992.
- [27] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inf. Theory*, vol. 38, pp. 904–909, 1992.
- [28] O. Toussoun, "Mémoire sur l'histoire du Nil," *Mémoires a l'Institut d'Egypte*, vol. 18, pp. 366–404, 1925.
- [29] M. Vannucci and F. Corradi, "Covariance structure of wavelet coefficients: Theory and models in a Bayesian perspective," *J. Roy. Statist. Soc.*, ser. B, vol. 61, no. 4, pp. 971–986, 1999.
- [30] Y. Wang, J. Cavanaugh, and C. Song, "Self-similarity index estimation via wavelets for locally self-similar processes," *J. Statist. Plan. Infer.*, vol. 99, pp. 91–110, 2001.
- [31] B. Whitcher and M. Jensen, "Wavelet estimation of a local long memory parameter," *Explor. Geophys.*, vol. 31, pp. 94–103, 2000.

**Kyungduk Ko**, photograph and biography not available at the time of publication.

**Marina Vannucci**, photograph and biography not available at the time of publication.