

# Inference on fractal processes using multiresolution approximation

Kenneth Falconer and Carmen Fernández

*Mathematical Institute, University of St Andrews, North Haugh,  
St Andrews, Fife, KY16 9SS, United Kingdom*

and

*Instituto Español de Oceanografía  
Cabo Estai-Canido, 36200 Vigo, Spain*

## Abstract

We consider Bayesian inference via Markov chain Monte Carlo (MCMC) for a variety of fractal Gaussian processes on the real line. These models have unknown parameters in the covariance matrix, requiring inversion of a new covariance matrix at each MCMC iteration. The processes have no suitable independence properties so this becomes computationally prohibitive. We surmount these difficulties by developing a computational algorithm for likelihood evaluation based on a ‘multiresolution approximation’ to the original process. The method is computationally very efficient and widely applicable, making likelihood-based inference feasible for large datasets. A simulation study indicates that this approach leads to accurate estimates for underlying parameters in fractal models, including functional parameters in the recently introduced multifractional Brownian motion. We apply the method to a variety of real data sets and illustrate its application to prediction and to model selection.

*Keywords:* Fractal process; Fractional Brownian motion; Graphical model; Markov chain Monte Carlo; Multiresolution approximation; Temporal modelling.

## 1 Introduction

Many data sets, for example time series from meteorology, finance and internet traffic, display such a degree of irregularity that modelling by smooth processes is inappropriate. In this case, commonly used processes are GARCH-type and stochastic volatility models (see *e.g.* the collections edited by Engle, 1995, and by Shephard, 2005), or non-Gaussian models driven by Lévy motion (Barndorff-Nielsen and Shephard, 2001). An alternative, which we explore in this paper, is the use of fractal models, for example the many variants on Brownian motion such as fractional Brownian motion, fractional Gaussian noise, and multifractional Brownian motion. Such

models have (possibly functional) parameters in the covariance matrix that determine their fractal properties, and our aim is to perform Bayesian inference on these parameters. In an MCMC setting, this requires the inversion of a new covariance matrix at each iteration, and since these processes have no convenient independence properties this is computationally very expensive.

In this paper, we develop a computational algorithm based on a ‘multiresolution approximation’ (hereafter, MRA) to the original processes. Daniel and Willsky (1997) introduced the MRA to fractional Brownian motion for the purpose of (approximate) process simulation. Here, we consider the MRA in a more general setting, with possibly irregularly spaced time points and fractal processes other than fractional Brownian motion. More importantly, we consider an inference context and use the MRA as a basis for developing a computational algorithm to evaluate (approximately) likelihoods in a very efficient recursive manner. As a consequence, Bayesian inference and prediction via MCMC becomes feasible for large sets of observations. This likelihood-based approach constitutes a departure from earlier work on inference on parameters underlying fractal processes (see e.g. Benassi *et al.*, 1998, and references therein, and Kent and Wood, 1997), which often constructs estimators based on specific features of the processes and focuses on ‘infill’ (also called fixed-domain) asymptotic properties.

The MRA method is widely applicable, and we illustrate different possibilities using a few specific processes which we describe in Section 2. In particular, we consider fractional Brownian motion (Mandelbrot and Van Ness, 1968), multifractional Brownian motion (Benassi *et al.*, 1997, Peltier and Lévy Véhel, 1995), a generalisation that allows the regularity properties of the process to vary in time, and fractional Gaussian noise, a stationary process obtained as the increments of fractional Brownian motion. Section 3 describes in detail the MRA, which may be regarded as a Gaussian graphical model (Whittaker, 1990) of the original process, with a much simplified dependence structure, but which reflects the correlations of the original process at all scales well-enough to permit accurate inference. Section 4 introduces the sampling model, which has a smooth trend with an error modelled by a fractal process. It then describes the MCMC implementation, where our computational algorithms for posterior inference and prediction are developed. Section 5 examines the appropriacy of the approximation for the purpose of inference. We set up a simulation experiment where data are generated from a variety of fractal processes and inference is conducted on the basis of the MRAs to the original processes. The results clearly suggest that the approximation is effective for inference in such cases. Later in the section, we provide theoretical arguments that lend additional support to the appropriacy of the approximation. Section 6 illustrates the use of the method in the context of two real datasets.

We model wind speed data by multifractional Brownian motion, deducing a significantly varying regularity index over time, and we demonstrate the predictive possibilities of the method. We then examine exchange rates using fractional Brownian motion. We use Reversible Jump to compute posterior probabilities for alternative models that have been proposed for the trend, touching upon the issue of arbitrage in financial models based on fractional Brownian motion. Possibilities for further development are detailed in Section 7.

In recent years a substantial body of literature has developed in connection with multi-resolution ideas. Some such models are often called multiscale and are used as models on their own right rather than as approximations to some other model. Their structure allows accommodation of nonstationarities in the data and often permits fast computations, making them attractive for handling many large sets of data arising in practical situations, particularly in spatial and spatio-temporal contexts. See, for example, Huang *et al.* (2002), Johannesson *et al.* (2007), Patil and Taillie (2001) and references therein.

## 2 Fractal processes

This section gives an overview of the fractal processes to which we will apply the MRA method, although it is applicable to fractal processes other than those explicitly considered in the paper.

### 2.1 Fractional Brownian motion

Index- $h$  fractional Brownian motion (henceforth denoted by fBm), where  $0 < h < 1$ , is the Gaussian process  $\{Z(t) : t \in \mathbb{R}\}$  such that

- (i) with probability 1,  $Z(0) = 0$  and the sample paths are everywhere continuous,
- (ii) for any  $t$  and  $u$ ,  $Z(t) - Z(u) \sim N(0, |t - u|^{2h})$ .

Fractional Brownian motion was introduced by Mandelbrot and Van Ness (1968) and the books by Adler (1981) and Mandelbrot (2002) are general references. Many financial models are based on fBm (see e.g. Shiryaev, 1999), which has also been used to model network traffic, fluid turbulence and in hydrology.

When  $h = 1/2$ , fBm reduces to Brownian motion. From the definition, fBm is a Gaussian process with

$$E[Z(t)] = 0, \quad V[Z(t)] = |t|^{2h}, \quad Cov[Z(t), Z(u)] = \frac{|t|^{2h} + |u|^{2h} - |t - u|^{2h}}{2}. \quad (2.1)$$

It is immediate that fBm is not a stationary process. Most commonly, only positive times are considered, in which case the correlation between  $Z(t)$  and  $Z(u)$  is positive and increasing in

$\min\{t, u\}/\max\{t, u\}$ . Correlation also depends on  $h$ .

With probability 1, sample paths of fBm are nowhere differentiable and have everywhere Hölder exponent  $h \in (0, 1)$ . Thus, the smaller  $h$ , the more irregular the sample graphs will be. This can be appreciated from Figure 1, which displays realisations corresponding to 1,025 randomly spaced times for several values of  $h$  (here, as when data are simulated elsewhere in the paper, a standard algorithm for multivariate Normal random variate generation via Cholesky decomposition of the covariance matrix is used).

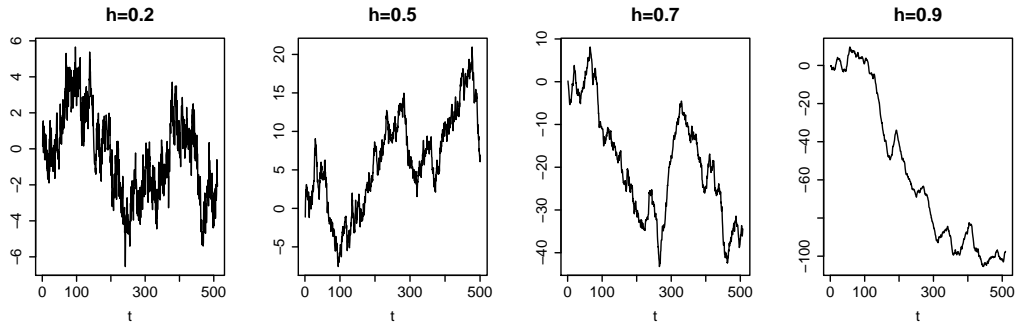


Figure 1: fBm realisations corresponding to  $h = 0.2, 0.5, 0.7, 0.9$ .

Given a set of observations corresponding to  $n$  time points of a realisation of fBm, we shall consider Bayesian inference on  $h$ .

## 2.2 Multifractional Brownian motion

Multifractional Brownian motion (denoted as mBm) is a recent generalisation of fBm that allows for a varying degree of regularity along sample graphs. This process was introduced by Benassi *et al.* (1997) and Peltier and Lévy Véhel (1995). Let  $h : \mathbb{R} \rightarrow [h_{\min}, h_{\max}] \subset (0, 1)$  be a differentiable function. Then multifractional Brownian motion with functional parameter  $h(t)$  is the process  $\{Z(t) : t \in \mathbb{R}\}$  defined in harmonizable form by

$$Z(t) = \text{Re} \int_{-\infty}^{\infty} \frac{e^{it\zeta} - 1}{|\zeta|^{h(t)+1/2}} dW(\zeta), \quad (2.2)$$

where  $W$  is white noise and  $\text{Re}$  denotes Real part. There is an equivalent definition of mBm as a moving average with respect to white noise, see Cohen (1999).

From the definition above, mBm is a zero-mean Gaussian process. The harmonizable form (2.2) guarantees a valid covariance metric, which was calculated by Ayache *et al.* (2000) as

$$\text{Cov}[Z(t), Z(u)] = D(h(t), h(u)) \left( |t|^{h(t)+h(u)} + |u|^{h(t)+h(u)} - |t-u|^{h(t)+h(u)} \right) \quad (2.3)$$

where

$$D(x, y) = \frac{\sqrt{\Gamma(2x+1)\Gamma(2y+1)\sin(\pi x)\sin(\pi y)}}{2\Gamma(x+y+1)\sin[\pi(x+y)/2]} \quad (0 < x, y < 1). \quad (2.4)$$

Note that  $V[Z(t)] = |t|^{2h(t)}$ , whereas the correlation between  $Z(t)$  and  $Z(u)$  now depends on  $t/u$ ,  $h(t)$  and  $h(u)$ . When  $h(t) = h$  mBm reduces to index- $h$  fBm. Benassi *et al.* (1997) showed that, with probability 1, the sample paths have Hölder exponent equal to  $h(t)$  at time  $t$ . Unlike fBm, the process mBm does not have stationary increments. The various figures in Section 5.2 provide examples of sample graphs corresponding to different choices of the function  $h(t)$ .

Potential applications of mBm include topography (such as mountain skylines) with varying geology, financial data with varying volatility, micro-scale properties of wind and other meteorological data with changing weather regimes.

Suppose we have observations at  $n$  times  $t_1 < t_2 < \dots < t_n$  corresponding to a realisation of mBm. We consider Bayesian inference on  $h(t)$ . Since the latter is a smoothly varying function, we will parameterise it via a linear combination of orthonormal polynomials, subsequently transformed so that the image lies within an interval  $[h_{\min}, h_{\max}] \subset (0, 1)$ , as required. Thus, let  $p_j(t)$ ,  $j = 1, \dots, J$ , be an orthonormal family of polynomials (for example, Chebychev or Legendre polynomials), defined on  $[t_1, t_n]$ , where  $p_j(t)$  is of degree  $j - 1$ , and define the function

$$h(t) = h_{\min} + (h_{\max} - h_{\min}) \frac{\exp[\sum_{j=1}^J \gamma_j p_j(t)]}{1 + \exp[\sum_{j=1}^J \gamma_j p_j(t)]}. \quad (2.5)$$

We will consider the prior distribution

$$\gamma_j \sim N(0, \sigma^2/j), \quad j = 1, \dots, J, \quad \text{independently}, \quad (2.6)$$

for a suitably chosen value of  $\sigma^2$ , where the term  $1/j$  in the variance reflects the decay in the coefficients when a differentiable function is expanded in terms of orthonormal polynomials. In all our applications, we fix the value of  $J$ , but  $J$  could also be subject to inference [computationally, this could be handled using reversible jump (Green, 1995)].

### 2.3 Fractional Gaussian noise

Fractional Gaussian noise (denoted as fGn) is defined as the process corresponding to the increments of fBm for a fixed lag  $\delta > 0$ , that is, the process  $\{Y(t) \equiv Z(t + \delta) - Z(t) : t \in \mathbb{R}\}$  where  $\{Z(t) : t \in \mathbb{R}\}$  is index- $h$  fBm.

From the definition, it follows that fGn is a stationary zero-mean Gaussian process with

$$\text{Cov}[Y(t), Y(u)] = \delta^{2h} K_h \left( \frac{|t - u|}{\delta} \right), \quad \text{where} \quad K_h(s) = \frac{|1 + s|^{2h} + |1 - s|^{2h} - 2|s|^{2h}}{2}. \quad (2.7)$$

Hence,  $V[Y(t)] = \delta^{2h}$ . Long-range dependence (non-summable correlations) occurs when  $h > 0.5$ , in which case correlations are positive and increasing in  $h$ . fGn with  $h > 0.5$  is often used as a statistical model with long memory. In finance it may represent increments of share prices or exchange rates (see Mandelbrot, 1997). Classical applications are to annual river levels (Mandelbrot and Wallis, in Mandelbrot, 2002; Beran, 1994) and internet traffic (Beran, 1994).

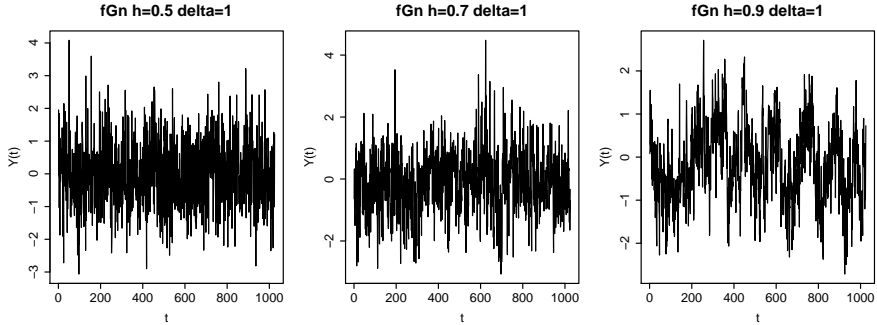


Figure 2: Realisations of fGn with  $\delta = 1$  for various values of  $h$ .

Figure 2 displays traces of fGn (1,025 time points at distance 1 apart) for several values of  $h$  with  $\delta = 1$ . The case  $h = 0.5$  corresponds to independence (white noise), whereas it is visually apparent from the figure that for  $h > 0.5$  temporal autocorrelation increases as  $h$  increases.

Most studies relate to ‘discrete-time fractional noise’, where the times  $t$  considered are consecutive integers and  $\delta = 1$ . However, particularly if the observations are not equally spaced, it is natural to regard them as a sample from ‘continuous-time fractional noise’ for an appropriate value of  $\delta$ , not necessarily equal to 1. This increases the amount of flexibility allowed in the correlation structure of the model. In this paper, we perform inference on  $\delta$  as well as on  $h$ .

### 3 Multiresolution approximation to the likelihood

Let  $\{\varepsilon_\theta(t) : t \in \mathbb{R}\}$  be a family of Gaussian fractal processes, where  $\theta$  parameterizes the covariance structure. We wish to conduct inference on  $\theta$  given observations at times  $t_1 < t_2 < \dots < t_n$ . The  $t_i$  could be irregularly spaced (either as a consequence of observations taken at irregular time intervals or due to missing data) and, typically, the number of observations  $n$  may be very large. As a result, calculating the inverse covariance matrix of the random vector  $(\varepsilon_\theta(t_1), \dots, \varepsilon_\theta(t_n))$  is computationally intensive, rendering implementation of MCMC, or any other numerical procedure that requires repeated evaluation of the likelihood function, prohibitively time-consuming. To overcome this difficulty, we approximate the joint density of  $(\varepsilon_\theta(t_1), \dots, \varepsilon_\theta(t_n))$  by another

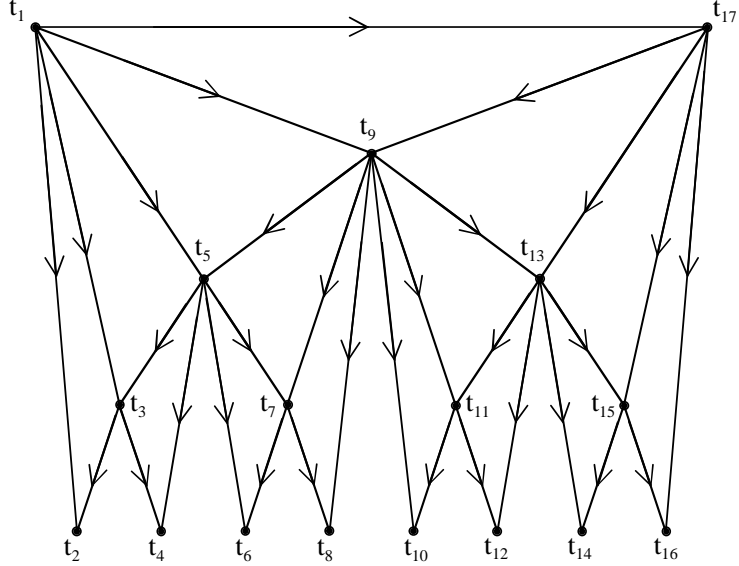


Figure 3: Multiresolution graph with  $n = 17$  observational times.

Gaussian density which permits rapid likelihood calculations. This approximation is very similar to that introduced by Daniel and Willsky (1997), who used it exclusively for the purpose of approximate simulation of fBm. We adopt their terminology ‘multiresolution’ approximation. The basic idea is to simplify the dependence structure of the random vector, as described below.

We construct inductively a directed graph  $(V, E^<)$  with vertex set  $V = \{t_1, t_2, \dots, t_n\}$  and directed edge set  $E^<$ , that will represent the dependence structure of the MRA. The set of vertices  $V$  is partitioned into ‘levels’  $V_1, V_2, \dots$ . We set  $V_1 = \{t_1, t_n\}$  and start with  $E^< = \{(t_1, t_n)\}$ . For  $l = 2, 3, \dots$  we define  $V_l$  as follows. For each consecutive pair of points  $t_L < t_R$  in  $V_1 \cup V_2 \cup \dots \cup V_{l-1}$  such that  $R - L > 1$ , we put into  $V_l$  the point  $t_C$ , where  $C = (L + R)/2$  if  $R - L$  is even and either  $C = (L + R - 1)/2$  or  $C = (L + R + 1)/2$  (choosing according to some convention) if  $R - L$  is odd. The directed edges  $(t_L, t_C)$  and  $(t_R, t_C)$  are then added to  $E^<$ . Thus  $t_L$  and  $t_R$  may be thought of as the ‘parents’ of ‘child’  $t_C$ . This process continues until all the  $t_i$  are in some  $V_l$ . This is represented in Figure 3 for  $n = 17$  (when  $n \neq 1 + 2^l$  for an integer  $l$ , the bottom row of the graph is not ‘full’, but this poses no problems).

Using the dependence structure corresponding to the directed graph, the MRA has density

$$p_{\text{MRA}}(\varepsilon_\theta(t_1), \dots, \varepsilon_\theta(t_n)) = p(\varepsilon_\theta(t_1), \varepsilon_\theta(t_n)) \prod_{l \geq 2} \prod_{t_C \in V_l} p(\varepsilon_\theta(t_C) | \varepsilon_\theta(t_L), \varepsilon_\theta(t_R)), \quad (3.1)$$

where  $t_L$  and  $t_R$  are the parents of  $t_C$ . Each of the factors on the right hand side of (3.1) is taken to be the corresponding density according to the original Gaussian fractal process.

Note that, by construction, the MRA preserves the distribution of  $(\varepsilon_\theta(t_L), \varepsilon_\theta(t_C), \varepsilon_\theta(t_R))$

whenever  $(t_L, t_C, t_R)$  is a parents-child triple. This implies that the marginal distributions at each time point  $t_i$  are not affected by the approximation and the covariances  $Cov[\varepsilon_\theta(t_i), \varepsilon_\theta(t_j)]$  are also preserved whenever  $(t_i, t_j) \in E^\leftarrow$ . The remaining covariances of the MRA may be calculated using the directed graph structure, relating the variables to ancestors until a common parents-child triple of ancestors is reached, as shown in Appendix A. For likelihood calculations, we do not need to evaluate these covariances, a fact that contributes to the computational efficiency of this MRA. However, they will be required for prediction at unobserved times.

Let  $\Sigma$  and  $\Omega$  denote the covariance matrices of the original random vector and its MRA, respectively. Daniel and Willsky (1997) observed that for fBm the entries of  $\Sigma$  and  $\Omega$  are remarkably close. For Brownian motion, which is a Markov process, the approximation is exact, but all entries of  $\Sigma$  and  $\Omega$  are very close even if the index  $h$  of fBm is far from  $1/2$ . For example, with  $h = 0.9$  and 513 equally spaced  $t_i$ , we obtain  $1 \leq \Omega_{i,j}/\Sigma_{i,j} \leq 1.108$  for all entries  $(i, j)$ .

Using the MRA in place of the original process achieves two things. Likelihood calculations are computationally fast, using the density factorisation on the right hand side of (3.1). For example, evaluating the likelihood can be achieved in  $O(n)$  operations compared to  $O(n^3)$  that would be required using Cholesky decomposition to invert the  $n \times n$  covariance matrix. This results in a computational gain that is crucial in order to handle large sets of data. Details will be provided in the following section. On the other hand, for many fractal processes, this gives an adequate likelihood approximation for inference, as will be demonstrated in Section 5.

## 4 Bayesian computations via multiresolution approximation

We consider Bayesian inference given observations  $y(t_i)$ ,  $t_1 < t_2 < \dots < t_n$ , from

$$y(t_i) = \mu(t_i) + \omega^{-1/2} \varepsilon_\theta(t_i), \quad (4.1)$$

where  $(\varepsilon_\theta(t_1), \dots, \varepsilon_\theta(t_n))$  is a partial realisation of a zero-mean Gaussian fractal process. The parameter  $\omega > 0$  represents precision and  $\mu(t_i)$  is the sampling mean at time  $t_i$ . In most of our applications we shall choose

$$\mu(t) = \sum_{k=1}^K p_k(t) \beta_k, \quad (4.2)$$

where  $p_k(t)$  are orthonormal polynomials and  $\beta_k \in \mathbb{R}$ , so that  $\mu(t)$  is a smooth function of time.

We consider the prior density

$$p(\beta, \omega, \theta) \propto \omega^{-1} p(\theta), \quad (4.3)$$

which corresponds to the standard non-informative prior for  $(\beta, \omega)$ , whereas  $p(\theta)$  is left unspecified for the moment. The joint posterior distribution admits the following factorisation:

$$\beta|\omega, \theta, y \sim N_K \left( (X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}y, (X'\Sigma_\theta^{-1}X)^{-1}/\omega \right), \quad (4.4)$$

$$\omega|\theta, y \sim \text{Gamma} \left( \frac{n-K}{2}, \frac{y'\Sigma_\theta^{-1}y - y'\Sigma_\theta^{-1}X(X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}y}{2} \right), \quad (4.5)$$

$$p(\theta|y) \propto p(\theta) \left( |X'\Sigma_\theta^{-1}X| |\Sigma_\theta| \right)^{-1/2} \left\{ y'\Sigma_\theta^{-1}y - y'\Sigma_\theta^{-1}X(X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}y \right\}^{-(n-K)/2}, \quad (4.6)$$

where  $'$  denotes transpose,  $|\cdot|$  stands for determinant,  $X$  is the  $n \times K$  design matrix (assumed to be of full column-rank with  $K < n$ ) associated with  $\beta$  [thus,  $X$  is determined by (4.2)] and  $\Sigma_\theta$  is the sampling covariance matrix. We shall take  $p(\theta)$  to be a proper prior defined on a compact set in order to guarantee integrability of  $p(\theta|y)$  and, thus, propriety of the posterior distribution.

#### 4.1 Posterior inference

The posterior distribution can be computed using MCMC, drawing  $\beta$  and  $\omega$  directly from the distributions in (4.4) and (4.5), respectively, and using a Metropolis-Hastings algorithm for updating  $\theta$  (we note that  $\theta$  can be multi-dimensional and non-trivial to sample effectively). This requires evaluating  $\Sigma_\theta^{-1}$  at each iteration of the MCMC algorithm, which is prohibitively time-consuming if  $n$  is moderately large. To circumvent this problem, we replace the matrix  $\Sigma_\theta$  corresponding to the original process by the covariance matrix  $\Omega_\theta$  of the MRA in (3.1).

From (4.4)-(4.6) the quantities, now involving  $\Omega_\theta$ , that we need to compute are

$$|\Omega_\theta|, \quad y'\Omega_\theta^{-1}y, \quad X'\Omega_\theta^{-1}X, \quad X'\Omega_\theta^{-1}y. \quad (4.7)$$

This reduces to finding the determinant of  $\Omega_\theta$  and evaluating bilinear forms involving  $\Omega_\theta^{-1}$ .

The key idea is to combine the directed graph structure of the MRA with the following standard linear algebra results for partitioned matrices. For a matrix  $\Omega$  and vectors  $u, v$ , partitioned as

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$

where  $\Omega_{22}$ ,  $u_2$  and  $v_2$  are scalars, we obtain that

$$|\Omega| = |\Omega_{11}|z \quad \text{and} \quad u'\Omega^{-1}v = u_1'\Omega_{11}^{-1}v_1 + \frac{(u_2 - Bu_1)(v_2 - Bv_1)}{z} \quad (4.8)$$

where  $B = \Omega_{21}\Omega_{11}^{-1}$  and  $z = \Omega_{22} - B\Omega_{12}$ . We can use the results in (4.8) to compute the quantities in (4.7) in an inductive way, starting with the time points in  $V_1$  and going down the levels  $V_l$  of the vertex set  $V$ . For each time  $t_C \in V_l$  ( $l \geq 2$ ),  $z$  and  $B$  in (4.8) will depend only on the times  $t_L$  and  $t_R$ , the parents of  $t_C$ . Hence, recalling (3.1), this procedure leads to

$$|\Omega_\theta| = |\Sigma_\theta(t_1, t_n)| \prod_{l \geq 2} \prod_{t_C \in V_l} z_\theta(t_C),$$

and

$$\begin{aligned}
u' \Omega_\theta^{-1} v &= \left( u(t_1), u(t_n) \right) \left( \Sigma_\theta(t_1, t_n) \right)^{-1} \begin{pmatrix} v(t_1) \\ v(t_n) \end{pmatrix} \\
&+ \sum_{l \geq 2} \sum_{t_C \in V_l} \frac{1}{z_\theta(t_C)} \left( u(t_C) - B_\theta(t_C) \begin{pmatrix} u(t_L) \\ u(t_R) \end{pmatrix} \right) \left( v(t_C) - B_\theta(t_C) \begin{pmatrix} v(t_L) \\ v(t_R) \end{pmatrix} \right),
\end{aligned}$$

where

$$B_\theta(t_C) = \left( \Sigma_\theta(t_C t_L), \Sigma_\theta(t_C t_R) \right) \left( \Sigma_\theta(t_L, t_R) \right)^{-1}, \quad z_\theta(t_C) = \Sigma_\theta(t_C t_C) - B_\theta(t_C) \begin{pmatrix} \Sigma_\theta(t_C t_L) \\ \Sigma_\theta(t_C t_R) \end{pmatrix}, \quad (4.9)$$

and we have used the notation  $u(t_i)$  for the entry of the vector  $u$  (similarly for  $v$ ) corresponding to time  $t_i$ ,  $\Sigma_\theta(t_i, t_j)$  for the  $2 \times 2$  submatrix of  $\Sigma_\theta$  corresponding to times  $t_i$  and  $t_j$  and  $\Sigma_\theta(t_i t_j)$  for the entry of the matrix  $\Sigma_\theta$  corresponding to the covariance between times  $t_i$  and  $t_j$ .

Note that as we consider each time point  $t_C$ , we only need to invert the  $2 \times 2$  covariance matrix corresponding to its parents  $(t_L, t_R)$ . Thus, the number of operations required to evaluate the quantities in (4.7) is of order  $n$  (the number of observations).

## 4.2 Prediction

Now we consider prediction of future unobserved values  $y(t)$ . Computationally this poses new challenges, as the multiresolution graph should be more naturally set up over the entire range of times, including those at which we wish to predict. In other words, the MRA is set up over times  $t_1 < \dots < t_n < \dots < t_{\tilde{n}}$ , but we have only observed data for the first  $n < \tilde{n}$  of them. This complicates the dependence structure of the observed data vector as explained next.

We will call *borderline observed times* those times  $t_C \leq t_n$  whose parent to the right does not correspond to an observed datum (i.e.  $t_R > t_n$ ), together with  $t_1$  (who has no parents). The dependence graph corresponding to the times with observed data (the first  $n$ ), will be the same graph as before but will also contain edges between each pair of borderline observed times. As an example, suppose that only the first 12 times in Figure 3 correspond to observed data. Drawing a vertical line between  $t_{12}$  and  $t_{13}$  in the figure, the borderline observed times are  $t_1$ ,  $t_9$ ,  $t_{11}$  and  $t_{12}$ . The directed graph for the first 12 times only, contains the additional edges  $(t_1, t_{11})$ ,  $(t_1, t_{12})$  and  $(t_9, t_{12})$ .

Let us first concentrate on posterior inference. When computing the quantities in (4.7) for the  $n$  observed times,  $\Omega_\theta$  corresponds to the more complicated directed graph mentioned in the previous paragraph. We will still use the basic results in (4.8) to compute these quantities inductively, working down the levels  $V_l$  of the vertex set  $V$ , but using only the observed times.

For the observed times  $t_C \in V_l$  that are not borderline, the algorithm is exactly as before. However, for each borderline observed time,  $t_C \in V_l$ , it is considerably more complicated, due to its dependence on all previous (i.e. higher up in the multiresolution graph) borderline observed times. In order to deal with this, we have modified the algorithm as described below.

First, we define *borderline unobserved times* to be those times  $t_C > t_n$  whose parent to the left corresponds to an observed datum (i.e.  $t_L \leq t_n$ ), together with  $t_{\bar{n}}$  (who has no parents). For example, going back to Figure 3 with only the first 12 times observed, the borderline unobserved times are  $t_{17}$  and  $t_{13}$ . The borderline unobserved times will play a critical role in the new algorithm, for the following reason. For a borderline observed time  $t_C \in V_l$  ( $l \geq 2$ ), its parent to the left,  $t_L$ , is always the previous borderline observed time, whereas its parent to the right,  $t_R$ , is the latest borderline unobserved time (this may be better visualized by referring back to Figure 3). Hence, if we knew the conditional distribution (or mean and variance) at time  $t_R$  given the borderline observed times above  $t_C$  in the multiresolution graph, we could easily relate  $t_C$  to all previous borderline observed times, as required. Thus, when working on level  $V_l$ , our algorithm will keep track of the conditional distribution of the latest borderline unobserved time given all the observed borderline times from  $V_1$  (top of the graph) down to the level  $V_l$ . For each level  $V_l$  ( $l \geq 2$ ), there is exactly one borderline observation, that can either be observed or unobserved. Thus, in the algorithm below, only one of steps (2b) or (2c) is required at each level. The algorithm is as follows (where the symbol “=” will denote assignment):

**1.** Start by setting  $|\Omega_\theta| = \Sigma_\theta(t_1 t_1)$ ,  $u' \Omega_\theta^{-1} v = u(t_1) (\Sigma_\theta(t_1 t_1))^{-1} v(t_1)$  and number of observed borderline times  $(nb) = 1$ . The latest borderline unobserved time is  $t_{\bar{n}}$ ; hence, compute the coefficient of conditional mean and the conditional variance at time  $t_{\bar{n}}$  given previous borderline observed times  $(t_1)$ :  $(cm)_1 = \Sigma_\theta(t_1 t_{\bar{n}}) (\Sigma_\theta(t_1 t_1))^{-1}$  and  $(cv) = \Sigma_\theta(t_{\bar{n}} t_{\bar{n}}) - (cm)_1 \Sigma_\theta(t_1 t_{\bar{n}})$ .

**2.** For  $l \geq 2$ :

**(2a)** For each observed time  $t_C \in V_l$  that is not borderline, the algorithm works as before. Thus, compute  $B_\theta(t_C)$  and  $z_\theta(t_C)$  as in (4.9) and set

$$\begin{aligned} |\Omega_\theta| &= |\Omega_\theta| z_\theta(t_C) \\ u' \Omega_\theta^{-1} v &= u' \Omega_\theta^{-1} v + \frac{1}{z_\theta(t_C)} \left( u(t_C) - B_\theta(t_C) \begin{pmatrix} u(t_L) \\ u(t_R) \end{pmatrix} \right) \left( v(t_C) - B_\theta(t_C) \begin{pmatrix} v(t_L) \\ v(t_R) \end{pmatrix} \right). \end{aligned}$$

**(2b)** If  $V_l$  contains an unobserved borderline time,  $t_C$ , compute the coefficients of the conditional mean and the conditional variance at time  $t_C$  given previous borderline observed times. First, compute  $B_\theta(t_C)$  and  $z_\theta(t_C)$  as in (4.9), noting that  $t_L$  is the latest borderline observed time,

whereas  $t_R$  is the previous borderline unobserved time. The required coefficients are

$$\begin{aligned}(cm)_j &= (cm)_j (B_\theta(t_C))_2, \quad \text{for } j = 1, \dots, (nb) - 1 \\ (cm)_{(nb)} &= (B_\theta(t_C))_1 + (cm)_{(nb)} (B_\theta(t_C))_2, \\ (cv) &= z_\theta(t_C) + (cv) \{(B_\theta(t_C))_2\}^2,\end{aligned}$$

where  $(B_\theta(t_C))_1$  and  $(B_\theta(t_C))_2$  denote the two components of  $B_\theta(t_C)$ .

**(2c)** If  $V_i$  contains an observed borderline time,  $t_C$ , compute the coefficients of the conditional mean and the conditional variance at time  $t_C$  given previous borderline observed times. For this, first compute  $B_\theta(t_C)$  and  $z_\theta(t_C)$  as in (4.9), and note that  $t_L$  is the previous borderline observed time, whereas  $t_R$  is the latest borderline unobserved time. Then set

$$\begin{aligned}|\Omega_\theta| &= |\Omega_\theta| \tilde{z}_\theta(t_C), \quad \text{where } \tilde{z}_\theta(t_C) = z_\theta(t_C) + (cv) \{(B_\theta(t_C))_2\}^2, \\ u' \Omega_\theta^{-1} v &= u' \Omega_\theta^{-1} v \\ &+ \frac{1}{\tilde{z}_\theta(t_C)} \left( u(t_C) - B_\theta(t_C) \begin{pmatrix} u(T_{(nb)}) \\ \sum_{j=1}^{(nb)} (cm)_j u(T_j) \end{pmatrix} \right) \left( v(t_C) - B_\theta(t_C) \begin{pmatrix} v(T_{(nb)}) \\ \sum_{j=1}^{(nb)} (cm)_j v(T_j) \end{pmatrix} \right),\end{aligned}$$

where  $T_1, \dots, T_{(nb)}$  are the previous borderline observed times starting from the top of the graph.

In addition, update the number of borderline observed times and the conditional distribution of  $t_R$  (the latest borderline unobserved time) to condition also on  $t_C$  (the new borderline observed time). This means updating  $(nb)$ ,  $(cv)$  and  $(cm)_j$  by setting:

$$\begin{aligned}(nb) &= (nb) + 1, \\ (\tilde{cv}) &= \left( \frac{\{(B_\theta(t_C))_2\}^2}{z_\theta(t_C)} + \frac{1}{(cv)} \right)^{-1}, \\ (cm)_j &= (\tilde{cv}) \frac{(cm)_j}{(cv)}, \quad \text{for } j = 1, \dots, (nb) - 2 \\ (cm)_{(nb)-1} &= (\tilde{cv}) \left( \frac{(cm)_{(nb)-1}}{(cv)} - \frac{(B_\theta(t_C))_1 (B_\theta(t_C))_2}{z_\theta(t_C)} \right) \\ (cm)_{(nb)} &= (\tilde{cv}) \frac{(B_\theta(t_C))_2}{z_\theta(t_C)} \\ (cv) &= (\tilde{cv}).\end{aligned}$$

Let us now focus on predicting the value of the observable  $y(t_f)$  at some future time  $t_f$ , where  $n < f \leq \tilde{n}$ . Thus, the aim is to compute the conditional distribution of  $y(t_f)$  given the vector of observations  $(y(t_1), \dots, y(t_n))$ . Conditioning on the parameters  $(\beta, \omega, \theta)$ , the dependence structure for all  $\tilde{n}$  times is given by the directed graph of the MRA. Hence, conditioning on the parameters, the predictive distribution of  $y(t_f)$  depends only on the observations corresponding to borderline observed times  $(y(T_1), \dots, y(T_{(nb)}))$ . Using standard results for multivariate

Gaussian distributions, this leads to a Normal distribution for  $y(t_f)$  with mean and variance

$$\begin{aligned}\mu_{t_f} &= \mu(t_f) + \left(\Omega_\theta(t_f T_1), \dots, \Omega_\theta(t_f T_{(nb)})\right) \left(\Omega_\theta(T_1, \dots, T_{(nb)})\right)^{-1} \begin{pmatrix} y(T_1) - \mu(T_1) \\ \vdots \\ y(T_{(nb)}) - \mu(T_{(nb)}) \end{pmatrix} \\ v_{t_f} &= \omega^{-1} \left\{ \Omega_\theta(t_f t_f) \right. \\ &\quad \left. - \left(\Omega_\theta(t_f T_1), \dots, \Omega_\theta(t_f T_{(nb)})\right) \left(\Omega_\theta(T_1, \dots, T_{(nb)})\right)^{-1} \left(\Omega_\theta(t_f T_1), \dots, \Omega_\theta(t_f T_{(nb)})\right)' \right\},\end{aligned}$$

where  $\mu(t) = \sum_{k=1}^K p_k(t) \beta_k$  in accordance with (4.2),  $\Omega_\theta(T_1, \dots, T_{(nb)})$  is the covariance matrix corresponding to the borderline observed times, and  $\Omega_\theta(t_f T_j)$  is the covariance between times  $t_f$  and  $T_j$ , all based on the MRA over the times  $\{t_1, \dots, t_{\tilde{n}}\}$ . Thus, in order to evaluate  $\mu_{t_f}$  and  $v_{t_f}$  we first need to compute the covariance between the prediction time  $t_f$  and each of the borderline observed times; Appendix A describes how covariances between any two time points can be computed. Next, we need to compute bilinear forms involving the inverse covariance matrix corresponding to only the borderline observed times. This can be done using the algorithm described above, ignoring step (2a). Finally, with a set of draws from the posterior distribution of  $(\beta, \omega, \theta)$  given the  $n$  observations  $y(t_1), \dots, y(t_n)$  (these draws are obtained using the algorithm described above), the predictive distribution of  $y(t_f)$  is a mixture of Normals, where the mean and variance of each component are given by  $\mu_{t_f}$  and  $v_{t_f}$  corresponding to each of the posterior draws of  $(\beta, \omega, \theta)$  and all components are equally weighted. If preferred,  $\beta$  and  $\omega$  could be integrated out analytically, and the predictive distribution of  $y(t_f)$  would then be mixture of  $t$  distributions where the mixing is done using the posterior draws of  $\theta$ .

## 5 Appropriacy of the multiresolution approximation

In this section, we examine the appropriacy of the MRA for the purpose of posterior inference. We start with simulation experiments, where realisations from fractal processes are simulated and inference is conducted using the corresponding MRA. Posterior results are compared with the true values used to generate the data. Although for completeness we also present posterior results for the trend, our main focus is in estimating the fractal parameters ( $h$  for fBm,  $h(t)$  for mBm,  $h$  and  $\delta$  for fGn). Unless when stated otherwise, all posterior results in the paper are based on MCMC runs with 120,000 iterations, of which the first 20,000 were discarded. The programs were coded in Fortran 77 and ran on a 1,666 MHz personal computer.

## 5.1 Results for simulated fractional Brownian motion

We simulated realisations of (exact) fBm with  $h = 0.7$  for a range of sample sizes. The corresponding datasets are plotted in Figure 4. To each of these datasets, we added the (cubic polynomial) trends displayed via a broken curve in the relevant panels of Figure 5.

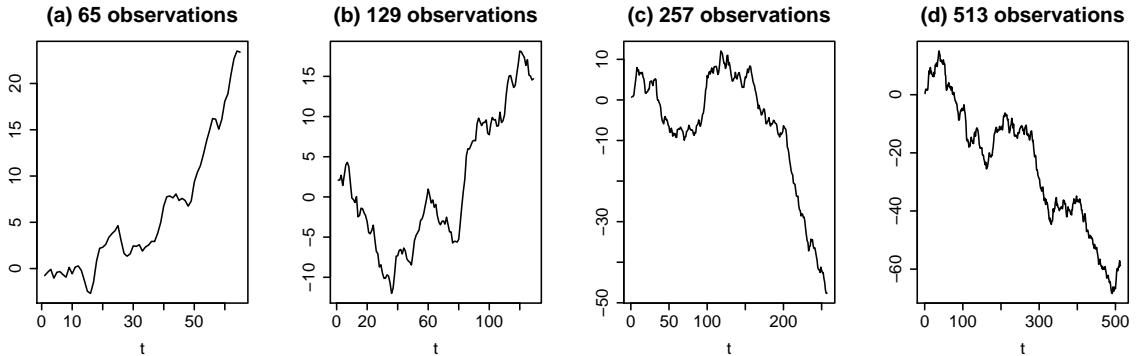


Figure 4: Realisations of fBm with  $h = 0.7$

The resulting datasets (consisting of fBm plus trend) are analysed according to the sampling model in (4.1)-(4.2), with  $K = 7$  Chebychev polynomials used for  $\mu(t)$ , and the standard non-informative prior for  $(\beta, \omega)$  described in (4.3). For the ‘error term’  $\varepsilon_\theta$  of the sampling model (4.1) we consider mBm. The function  $h(t)$  is parameterized according to (2.5), again using Chebychev polynomials, with the prior on the coefficients given by (2.6). We make two different choices for  $h(t)$ , both with  $h_{\min} = 1 - h_{\max} = 0.05$ : the two left columns of Figure 5 correspond to  $J = 1$  (that is, fBm) and we take  $\sigma^2 = 3$ , which implies a prior distribution for  $h$  very close to uniform; the two right panels of Figure 5 correspond to  $J = 7$  and  $\sigma^2 = 2$ , which implies that the prior distribution of  $h(t)$  is also very near uniform for all values of  $t$ .

Posterior inference is conducted replacing the assumed fBm or mBm by its MRA. The results are displayed in Figure 5. It is clear that, even for relatively small sample sizes, posterior inference via the MRA is quite reasonable with posterior moments quite near the true values. Taking  $J = 1$  does better than  $J = 7$ , which is unsurprising since the datasets were generated according to fBm. For fBm, we find that the posterior standard deviation of the trend is generally much smaller for lower values of  $t$ . This is explained by the fact that the variance and correlation of fBm increase with  $t$ , so the data provide more information about the trend for smaller values of  $t$ . For mBm the variance is  $t^{2h(t)}$ , and we find that the larger the estimated value of  $h(t)$  the more uncertainty in the estimated trend at time  $t$  (for  $t > 1$ ). The exaggerated curvature at the ends of the intervals in the third column of the figure are likely to be due to the polynomial

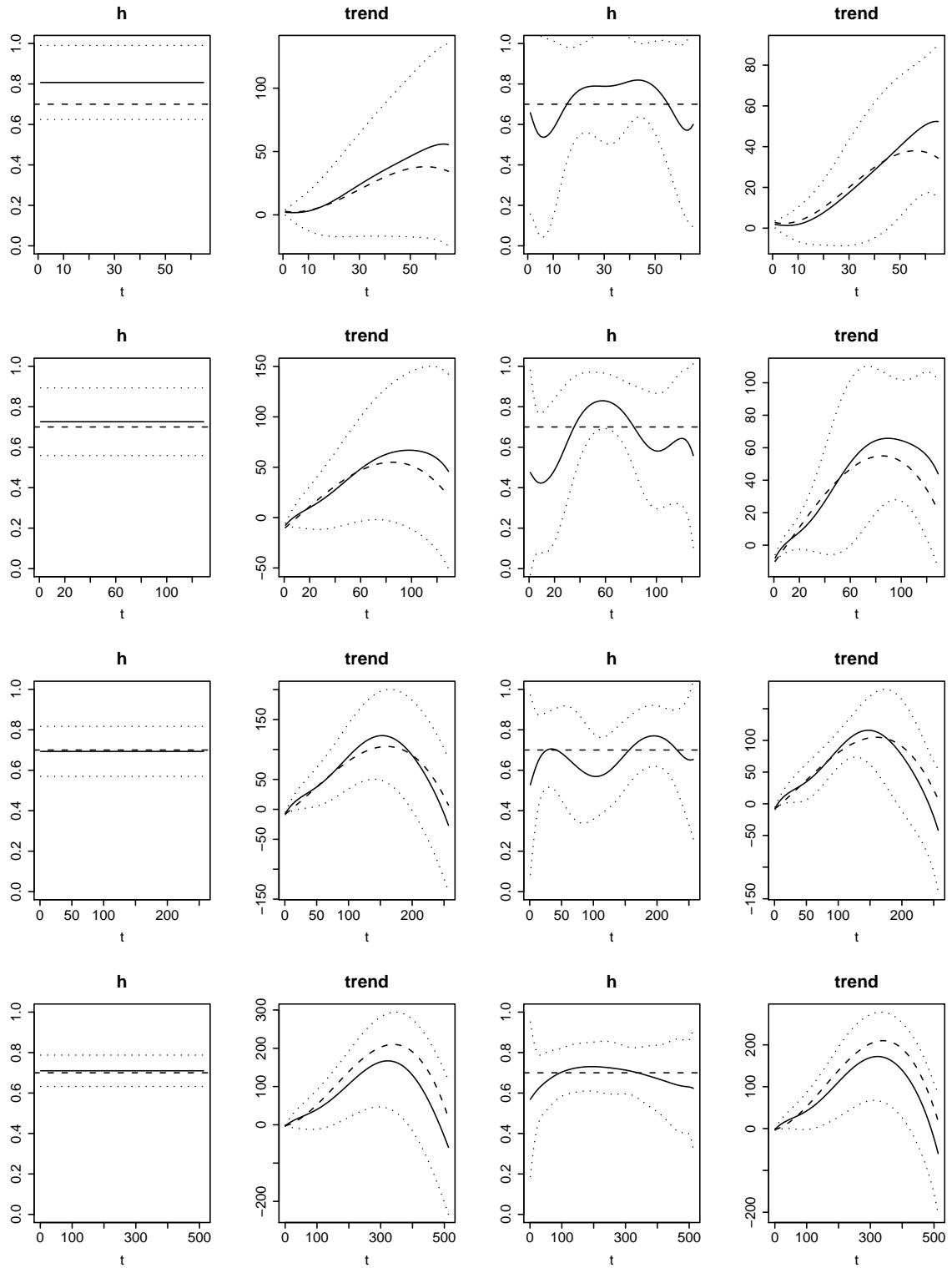


Figure 5: Posterior inference for data in Figure 4. Rows from top to bottom correspond to panels (a), (b), (c) and (d) in Figure 4, respectively. Continuous lines are posterior means; dotted lines are posterior means plus or minus 2 standard deviations; broken lines are true values.

regression used for  $h(t)$ . The use of more complex functions, with local support, to model  $h(t)$  is an avenue to be explored (although we shall not do so in this paper).

In order to assess more clearly the impact of the sample size on posterior inference, we simulated 250 independent realisations of index  $h = 0.7$  (exact) fBm for each of the sample sizes  $n = 65, 129, 257$  and  $513$ . For each of the 250 realisations, we ran the sampler, obtaining the posterior results displayed in Figure 6 in the form of their distribution over the 250 realisations. The left and middle panels correspond to a model with  $J = 1$ , and respectively report the posterior mean and root mean square error for  $h$ , illustrating that inference on  $h$  becomes more accurate as sample size increases. The right panel corresponds to a model with  $J = 7$ . In this case, mean square error for  $h(t)$  is computed by averaging the mean square errors across the observational times and then taking square root. Again we observe that increasing sample size leads to improved inference.

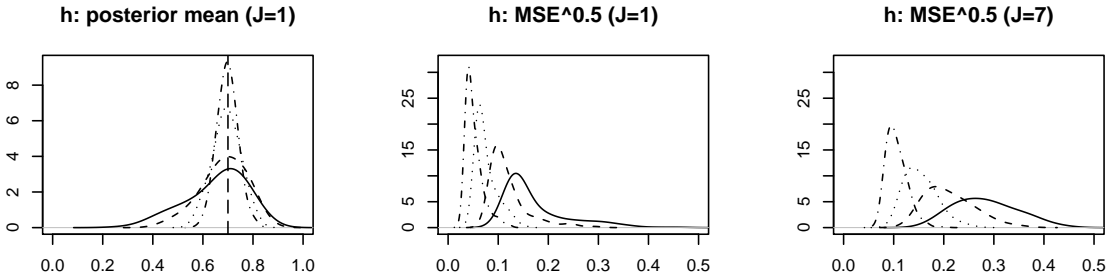


Figure 6: Distribution of posterior results over 250 independent realisations of fBm with  $h = 0.7$  of sizes  $n = 65, 129, 257$  and  $513$ . Left panel: increasing  $n$  corresponds to a tighter distribution. Middle and right panels: increasing  $n$  corresponds to a distribution that is shifted to the left.

## 5.2 Results for simulated multifractional Brownian motion

The left panels of Figure 7 display two datasets, each of 1,025 randomly spaced observations from exact mBm, to which we added the trends plotted with broken lines in the right panels. The data are analysed assuming the sampling model in (4.1)-(4.2), taking  $K = 7$  Chebychev polynomials for  $\mu(t)$ , and assuming the non-informative prior in (4.3) for  $(\beta, \omega)$ . The ‘error term’ in (4.1) is taken to be mBm. The function  $h(t)$  is parameterised as in (2.5), using  $J = 7$  Chebychev polynomials, and taking the prior in (2.6) with  $\sigma^2 = 2$  and  $h_{\min} = 1 - h_{\max} = 0.05$ . Posterior inference is conducted replacing the mBm model by its MRA. The results, displayed

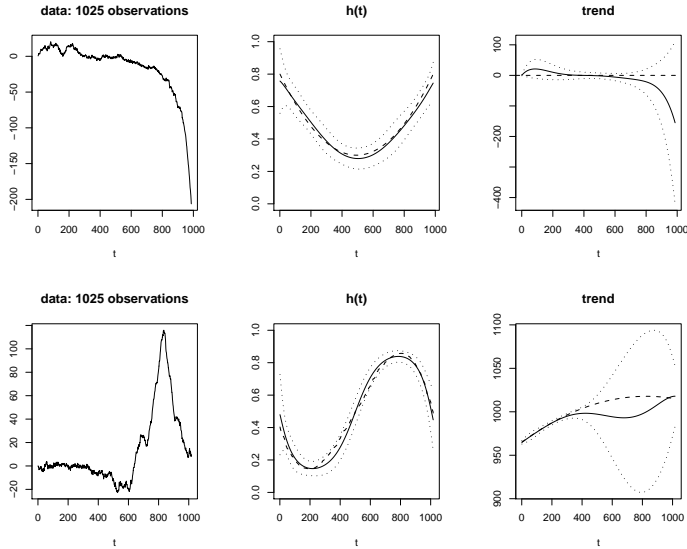


Figure 7: Posterior inference for data in left panel. Continuous lines are posterior means; dotted lines are posterior means plus or minus 2 standard deviations; broken lines are true values.

in the middle and right panels of Figure 7, testify to the excellent performance of posterior inference on  $h(t)$  via MRA. Results with a sample of half the size (not shown) were also good. We see that the uncertainty in the posterior estimates of the trend is larger when  $t^{2h(t)}$  is large, as explained in the previous subsection.

We compared our estimates with several other methods and the results in Figure 8 are typical of what we have found. The data in the left panel were obtained as a realisation of mBm with  $h(t)$  given by the broken line in the middle and right panels. The Bayesian posterior mean using MRA is the smooth solid curve in the central panel, which is virtually identical to the true function. To obtain a good fit, we required the use of  $J = 16$  Chebychev polynomials in (2.5) and we took  $\sigma^2 = 1$  in (2.6). For comparability with the other methods we have assumed  $\mu(t) = 0$  and  $\omega = 1$  in (4.1). Software and documentation for the other methods can be found within the web package *FracLab* (n.d.). Here, we provide a succinct description of these methods.

For  $n$  equally spaced time points in  $[0, 1]$ , *i.e.*  $t_i = i/n$  for  $i = 1, \dots, n$ , Benassi *et al.* (1998) propose the following generalised quadratic variation (GQV) estimator of  $h(t)$ :

$$\hat{h}_{n,\epsilon}(t) = \frac{1}{2} \left( 1 + \log_2 \frac{V_{n/2,\epsilon}(t)}{V_{n,\epsilon}(t)} \right) \quad \text{where}$$

$$V_{n,\epsilon}(t) = \sum_{i: |(i/n)-t| < \epsilon} \left\{ y\left(\frac{i+1}{n}\right) - 2y\left(\frac{i}{n}\right) + y\left(\frac{i-1}{n}\right) \right\}^2$$

and show that  $\hat{h}_{n,\epsilon}(t)$  converges almost surely to  $h(t)$  as  $n \rightarrow \infty$ , with  $\epsilon = n^{-1/3}$  the optimal

value to balance bias and variance. The results (not displayed) are not good, leading to erratic estimates, often even outside the range  $[0, 1]$ . Barrière and Lévy Véhel (in preparation) refine the GQV approach, the essence of their idea being to consider  $V_{n,\epsilon}(t)$  for several values of  $n$  and to fit a least squares regression line of  $\log(V_{n,\epsilon}(t))$  on  $\log(n)$ . The estimate using refined GQV (RGQV) is the somewhat rugged line in the right panel of Figure 8. The oscillation-based (OB) method uses  $\sup\{y(s) : s \in [t - \epsilon, t + \epsilon]\} - \inf\{y(s) : s \in [t - \epsilon, t + \epsilon]\}$  instead of GQV, and leads to the somewhat rugged line in the middle panel of the figure. The discrete wavelet transform (DWT) method is based on the fact that the asymptotic behaviour of local wavelet coefficients reflects the Hölder exponents (see Jaffard and Mayer 1996); this estimate is shown as the highly irregular trace in the right panel.

The superior performance of the Bayesian approach seems clear, a pattern that has been repeated in all other cases that we examined.

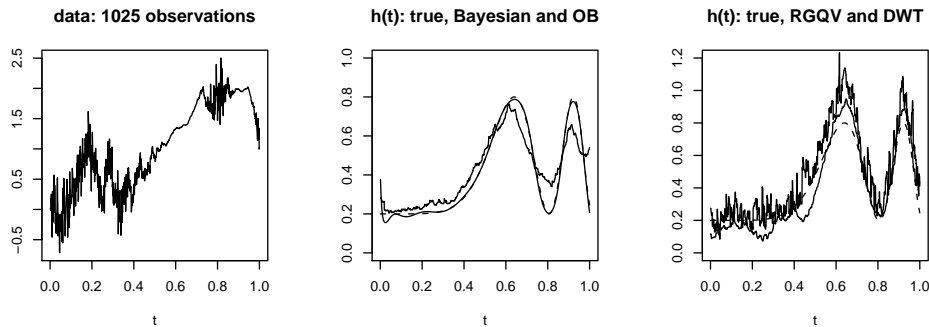


Figure 8: Inference for data in left panel. Middle panel: True  $h(t)$  (discontinuous), Bayesian posterior mean (smooth solid), OB method (rugged). Right panel: True  $h(t)$  (discontinuous), RGQV (rugged) and DWT (very rugged).

### 5.3 Results for fractional Gaussian noise

We generated data from fGn with  $h = 0.7$  and  $\delta = 1$ , without adding any trend, and considered posterior inference under the model (4.1)-(4.2). In the model, we used  $K = 7$  Chebychev polynomials for the trend, and the standard non-informative prior distribution for  $(\beta, \omega)$ . The ‘error’ process model was replaced by a MRA to fGn normalised to have unit variance (in this way, the sampling variance is given by  $1/\omega$ , and  $\delta$  and  $h$  only affect the correlation structure).

We simulated samples of size  $n = 129$  and  $n = 513$ . For each of these sample sizes, we considered both equally spaced and randomly spaced observations (always with fixed  $\delta = 1$ ) as

follows: we generated a sample of length  $2n$  corresponding to times  $\bigcup_{i=1}^n \{i-1+u_i, i\}$ , where  $u_i$  are random draws from a Uniform(0,1) distribution. Then we split this sample into two sub-samples of equal length, corresponding to the integer and non-integer times, respectively.

We consider posterior inference on  $\delta$  as well as on  $h$ , and in order to guarantee propriety of the posterior distribution we choose a proper prior for  $\theta = (h, \delta)$  on a compact set, as indicated in Section 4. In particular,  $h$  is assigned a Uniform prior distribution on the interval  $[0.5, 0.95]$  and  $\delta$  a Uniform distribution on the interval  $[0.1, 100]$ .

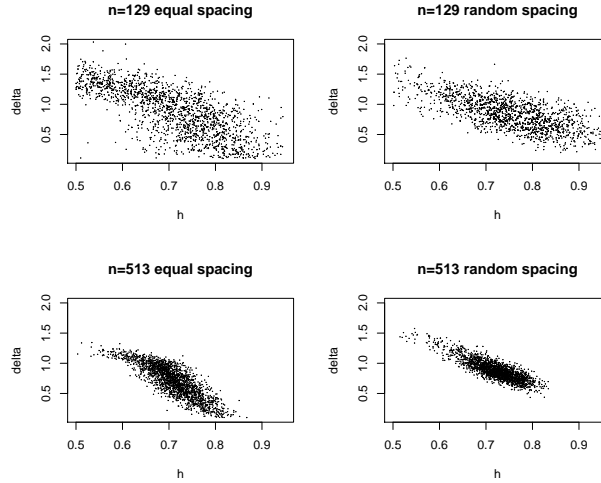


Figure 9: Scatterplot of posterior distribution of  $(h, \delta)$  for simulated data. True values are  $h = 0.7$  and  $\delta = 1$ .

Figure 9 displays scatterplots of the posterior distribution of  $(h, \delta)$  for each of the simulated datasets. It is clear from the figure that increasing sample size leads to better inference and that, for a given sample size, randomly spaced data is more informative than equally spaced data. This is not surprising as having randomly spaced data means that there is a bigger range of distances to help discriminate among the various possible correlation functions. It is also obvious from the pictures that there is strong posterior negative correlation between  $h$  and  $\delta$ . This is also unsurprising: increasing  $\delta$  is equivalent to scaling down the distances between observations [see equation (2.7)] and this can to some extent be offset by decreasing the value of  $h$ , as this implies lower correlations. Note that in all the plots in Figure 9 fixing the value of  $\delta$  at 1 (the true value) centres the posterior distribution of  $h$  at 0.7 (the true value of  $h$ ).

## 5.4 Theoretical arguments

In this section we indicate some theoretical notions that support the MRA approach. Let  $\Sigma$  and  $\Omega$  respectively denote the covariance matrices corresponding to the original fractal process and its MRA. A common way to measure discrepancy between probability distributions is through the Kullback-Leibler (KL in the sequel) divergence criterion, given by

$$\text{KL}(f, g) = \int \log \left[ \frac{f(x)}{g(x)} \right] f(x) dx, \quad (5.1)$$

if the distributions have densities  $f(\cdot)$  and  $g(\cdot)$ . For Gaussian distributions with the same mean and covariance matrices  $\Sigma$  and  $\Omega$ , we shall denote this by  $\text{KL}(\Sigma, \Omega)$ .

Of relevance to us is the following result on the best approximation to a multivariate Gaussian distribution by a Gaussian graphical model. The proof is a simple adaptation of that in Theorem 6.6.1 of Whittaker (1990), replacing the empirical covariance matrix  $S$  therein by  $\Sigma$ .

**Theorem 5.1** *Consider an  $n$ -variate Gaussian distribution with mean zero and covariance matrix  $\Sigma$ . Further, consider an undirected graph  $(V, E)$  with vertices  $\{1, \dots, n\}$  and a Gaussian graphical model with mean zero and independence graph  $(V, E)$ . Then the covariance matrix  $\Omega$  of the graphical Gaussian model for which the Kullback-Leibler divergence  $\text{KL}(\Sigma, \Omega)$  is minimum, satisfies the equations  $(\Omega^{-1})_{i,j} = 0$  if  $(i, j) \notin E$  and  $\Omega_{i,j} = \Sigma_{i,j}$  for all  $(i, j) \in A$ , for all cliques  $A$  in  $(V, E)$ . These conditions determine  $\Omega$  uniquely.*

Consider the ‘moral’ graph  $(V, E)$  obtained from the multiresolution graph  $(V, E^{\leftarrow})$  (introduced in Section 3) by dropping edge directions and creating additional undirected edges between ‘unmarried’ parents. In our case, the moral graph is the same as the directed one with the edge directions removed. It is known that the moral graph always contains the conditional independence graph. Since for Gaussian distributions, conditional independence is characterised by a zero in the corresponding entry of the inverse of the covariance matrix, we have  $(\Omega^{-1})_{i,j} = 0$  if  $(t_i, t_j) \notin E$ . Now recalling that a clique is a subgraph where there is an edge between each pair of vertices and such that this property would be lost if any other vertex was added to the subgraph, it is clear that the cliques in the graph  $(V, E)$  are the parents-child triples. As noted in Section 3, by construction,  $\Omega_{i,j} = \Sigma_{i,j}$  whenever  $(t_i, t_j)$  belongs to a parents-child triple. Thus, by application of Theorem 5.1, the multiresolution process is the best approximation (in the KL divergence sense) to the original process within the class of Gaussian models with the given independence graph  $(V, E)$ .

Suppose that we have a family of Gaussian processes with covariance matrices  $\Sigma_\theta$ , parameterized by  $\theta$ , and let  $\Omega_\theta$  be the covariance matrices of the corresponding multiresolution

approximations. For fixed  $\theta_0$ , it follows from Theorem 5.1 that  $\text{KL}(\Sigma_{\theta_0}, \Omega_{\theta})$  attains its minimum when  $\theta = \theta_0$ . From the definition of KL divergence in (5.1), the result implies that if the data were generated by the original process with  $\theta = \theta_0$  but we considered the likelihood function of  $\theta$  based on the MRA, the expected value of the log-likelihood function would be maximised when  $\theta = \theta_0$ . Thus, at least in the sense of KL divergence, we might expect that using the MRA in place of the original process would allow comparable likelihood-based estimates. This is supported by our empirical findings with simulated data, shown in previous subsections.

## 6 Analysis of real data

Now that we have established the good inferential properties for data simulated from fractal processes, we illustrate the performance of the MRA method in the context of problems arising from two real data sets. In particular, for the wind speed data, we implement the prediction algorithm described in Subsection 4.2, whereas for the financial data, we compute posterior model probabilities for two competing models.

### 6.1 Wind speed: prediction

The dataset plotted in the top left panel of Figure 10 corresponds to mean wind speed measurements taken at half hour intervals at the top of Cairngorm (Scottish Highlands). The dataset contains  $n = 2,712$  observations starting on day 206 (last week of July) of 1992, and was extracted from the Cairngorm weather webpage (n.d.). We have subtracted the first observation so that the data pass through the origin at time zero and taken as observational times all integers from 1 to 2,712. In addition to conducting posterior inference, we will predict wind speed up to 12 hours ahead. Thus, we will require the use of the more complex computational algorithms described in Subsection 4.2.

We modelled these data in exactly the same way as was done for the simulated data at the beginning of Subsection 5.2, considering again a MRA to the original mBm process. The MRA will be set up over  $\tilde{n} = n + 24$  times, require to encompass the 12 additional hours for forecasting purposes. When running the MCMC sampler, we found that mixing was a bit slow for the parameters  $\gamma_j$  that define  $h(t)$ , so we have used 1,200,000 iterations, discarding the first 200,000. Posterior results for  $h(t)$  and the trend  $\mu(t)$  are presented in the two top right panels of Figure 10. The estimated function  $h(t)$  fluctuates substantially over the time period considered.

The bottom panels of Figure 10 display the predictive densities for several time periods ahead, together with the actual values recorded (indicated by a bullet point on the horizontal

axis). As we try to predict further into the future, the predictive distribution becomes wider and it is generally always quite wide. Predicting wind speed at the top of a Scottish mountain is a rather difficult exercise and these results are presented here as an illustration of the capabilities of our method rather than as a serious attempt at forecasting wind speed. The latter would require taking relevant factors into account (e.g. incorporated as explanatory variables in the mean and, if available, taking other spatially correlated data into consideration).

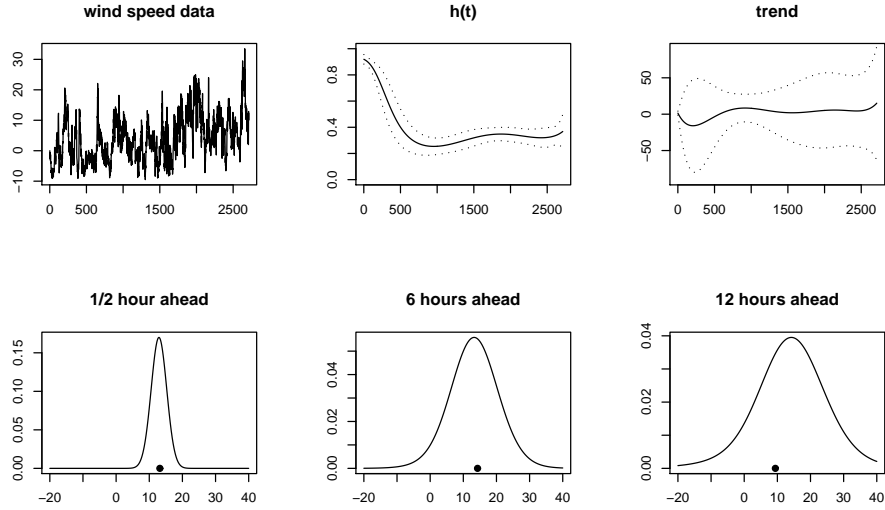


Figure 10: Top row: Posterior inference for wind speed data. Continuous lines are posterior means; dotted lines are posterior means plus or minus 2 standard deviations. Bottom row: Predictive densities and actual values.

A concern about the MRA, which arises in particular in the context of prediction, is that the dependence graph obtained for the joint distribution of times  $t_1 < \dots < t_n$  is generally different when the MRA is set up over the  $n$  times only from when it is set up over a larger set of times  $t_1 < \dots < t_n < \dots < t_{\tilde{n}}$ . As a consequence, posterior inference on parameters and prediction of unobserved values are not invariant to the range of times over which the MRA is set up. Nevertheless, our empirical experience indicates that, at least for moderately large values of  $n$ , the differences are not large. As an illustration, Figure 11 provides some comparisons for the wind speed data when  $n = 2,712$  times are observed but the MRA is set up over  $\tilde{n} \geq n$  times. The left panel corresponds to posterior inference on  $h(t)$ . The continuous lines are the posterior mean and 2 standard deviations when  $\tilde{n} = n$ , whereas the various discontinuous lines indicate posterior means when  $\tilde{n} = n + 24$ ,  $\tilde{n} = n + 288$  and  $\tilde{n} = n + 576$ . The middle and right panels display predictive distributions when  $\tilde{n} = n + 24$  (continuous lines),  $\tilde{n} = n + 288$  and  $\tilde{n} = n + 576$

(with various discontinuous lines). We recommend that the MRA be set up taking the smallest value of  $\tilde{n}$  that is ‘natural’ for the problem at hand. For example, take  $\tilde{n} = n$ , the number of observations, when interest centres on posterior inference, whereas, if prediction is an objective, take  $\tilde{n}$  only just large enough to encompass the range of times at which one wishes to predict.

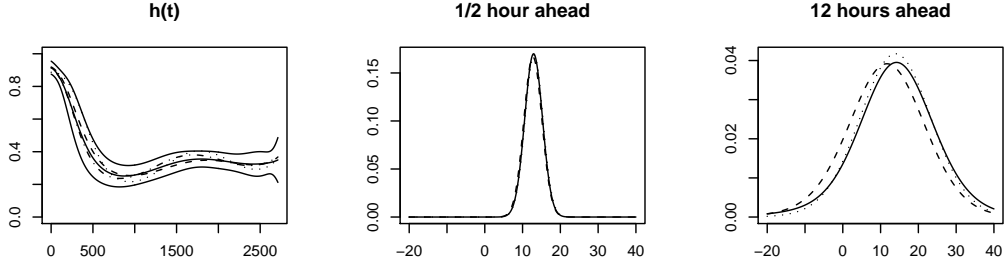


Figure 11: Wind speed data: comparison of results under different multiresolution graphs. Left panel: continuous lines are posterior mean and 2 standard deviations when  $\tilde{n} = n$ ; discontinuous lines are posterior means under larger values of  $\tilde{n}$ . Middle and right: continuous lines correspond to  $\tilde{n} = n + 24$ ; discontinuous lines correspond to larger values of  $\tilde{n}$ .

## 6.2 Exchange rates: model comparison

The data displayed on the left panel of Figure 12 are logarithms of the exchange rate between the Japanese Yen and the U.S.A. Dollar, daily for the years 1982-1986. It was downloaded from the Time Series Data Library (Hyndman, n.d.). We have 1,253 observations and we have subtracted the first one so that the data go through zero at the time origin. Traditionally, the dynamics of exchange rates are assumed to be well described by the stochastic differential equation

$$ds(t) = \beta s(t)dt + \omega^{-1/2}s(t)d\varepsilon(t), \quad (6.1)$$

where  $s(t)$  denotes the exchange rate at time  $t$ ,  $\varepsilon(t)$  is Brownian motion and  $\beta \in \mathbb{R}$  and  $\omega > 0$  are parameters. The usual solution of this equation uses the Itô formula, which depends on Brownian motion being a martingale, and leads to the model  $\log[s(t)] = \beta t + \omega^{-1/2}\varepsilon(t)$ . Because Brownian motion has independent increments, this model implies that the log-returns,  $\log[s(t)/s(t-1)]$ , are independent. It has been argued that this is often unrealistic (see e.g. Mandelbrot, 1997, and Willinger *et al.*, 1999), and fBm has been proposed as an alternative to Bm for the term  $\varepsilon(t)$  in (6.1). This poses difficulties: for  $h \neq 1/2$ , fBm is not a semimartingale and, thus, the previous techniques for solving equation (6.1) do not apply. Two different solutions have been proposed,

one based on path-wise integration ideas and another one based on Malliavin calculus:

$$\log[s(t)] = \beta t + \omega^{-1/2} \varepsilon(t), \quad \text{and} \quad (6.2)$$

$$\log[s(t)] = \beta t - \frac{1}{2\omega} t^{2h} + \omega^{-1/2} \varepsilon(t), \quad (6.3)$$

where  $\varepsilon(t)$  now denotes index- $h$  fBm. From the point of view of finance, there is a crucial difference between these solutions: if a Black-Scholes pricing model is constructed based on these solutions, the second solution leads to a model without arbitrage, but the first solution allows for arbitrage possibilities (see Sottinen and Valkeila, 2003, and references therein). It then becomes relevant to ascertain which of the two models seems more plausible given a set of data. This was investigated by Kukush *et al.* (2005) who, assuming  $h$  known, considered likelihood ratio tests though, as far as we are aware, they did not apply them to any dataset.

Here, we consider this problem from a Bayesian perspective, and estimate  $\beta$ ,  $\omega$  and  $h$  simultaneously to computing posterior model probabilities. Note that model (6.2) fits directly in our framework in Section 4, whereas the model in (6.3) has an additional term in the trend that involves both  $\omega$  and  $h$ . Computationally, we will handle the latter model by sampling  $\omega$  jointly with  $h$  instead of using equations (4.5) and (4.6). As usual, we replace the original fBm ‘error term’ process by its MRA. We consider the same prior distributions for  $\beta$ ,  $\omega$  and  $h$  as throughout the paper, and give prior probability 1/2 to each of the two models. Computing the posterior distribution now requires the use of Reversible Jump (see Green, 1995) as it is necessary to consider both models and to jump between them. The reversible jump step of the MCMC algorithm can be implemented very easily: given that we are at a certain model, we always proposed to jump to the other one, while keeping the same values of  $\beta$ ,  $\omega$  and  $h$ . This led to extremely high acceptance probabilities and very good mixing.

Posterior inference on  $h$  is displayed in the middle panel of Figure 12, with the posterior mean of  $h$  being equal to 0.576. The posterior probabilities of the models are very close, 0.485 for the first model and 0.515 for the second one, indicating that the data cannot distinguish between them. The reason appears to be a combination of  $\omega$  being very large (posterior median of  $\omega$  is 26543) and  $h$  being not too far from 0.5, which makes the models almost identical. This pattern was repeated for virtually all the financial datasets we tried.

However, the results are very different if the fBm ‘error term’ in the models (6.2) and (6.3) is replaced by mBm, and  $t^{2h}$  is substituted by  $t^{2h(t)}$  in the trend in (6.3). As before, we model  $h(t)$  using (2.5) with  $J = 7$  Chebychev polynomials, and take all priors as specified previously. Posterior inference on  $h(t)$  is summarised in the right panel of Figure 12, and we obtain a non-constant estimate, suggesting that mBm is a better model than fBm. The data are now much

more informative with respect to the two postulated models: the posterior probabilities are 0.750 for the first model and 0.250 for the second model. We refrain from drawing any conclusions regarding the potential implications of our findings for finance, since this is a complex issue outside the scope of this paper. One could, of course, investigate other models that exhibit long range dependence both with and without arbitrage, see for example, Rogers (1997), but our main aim here was to show how our method can be used to compare different models.

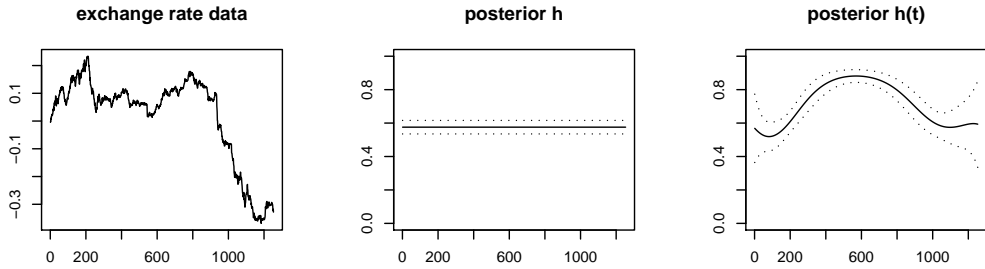


Figure 12: Posterior inference for data in left panel. Continuous lines are posterior means; dotted lines are posterior means plus or minus 2 standard deviations. Middle panel corresponds to inference with fBm and right panel to mBm.

## 7 Conclusions and further directions

Modelling by fractal processes has the potential for development and application in many directions. A major difficulty of simulation of such stochastic models is the complexity of the covariance structure, which makes the evaluation of the likelihood slow. The approach described in this paper, via MRA, reduces computation time enormously, while leading to very good accuracy for inference of fractal parameters.

Although we have focused on Bayesian inference, the approximation should be useful in any likelihood-based context where repeated evaluation of the likelihood is required. In addition, the same ideas could be applied if the fractal processes were used as building blocks in complex hierarchical models. As an example, the method could be straightforwardly applied to the case where the sampling model is as in (4.1) with an additional additive Normal error, which is independent for different times  $t_i$ . This additional noise could be interpreted as measurement error (or, in traditional geostatistics terminology, a nugget effect). Since the additional term does not change any correlations between the observations, this model can be accommodated

with only minor modifications of our current programs.

We conclude by mentioning some directions for future research.

1. The accuracy of approximation could be improved markedly by a modest enhancement of the independence graph. A graphical structure that retains the correlations between the process at a point and a sequence of other points going away at a geometric rate might be expected to incorporate dependencies at all scales very accurately indeed.

2. A fractal model on a higher dimensional domain may be appropriate for spatio-temporal situations. A common instance is where there are a small number of spatially dispersed recording sites at which frequent or continuous-time observations are made. Examples of this include meteorological data recorded at a number of stations; see Haslett and Raftery (1989) for an application to wind speed and Lovejoy and Schertzer (1995) for a discussion of rainfall. Fractality may be apparent in the time series at the individual stations, and incorporating appropriate spatial correlations would allow inference across the entire region.

3. The method could be applied to other classes of fractal processes, for example, processes with parameters that explicitly separate short and long range properties, such as those defined in Gneiting and Schlather (2003).

4. Another potential topic of future research is the asymptotic properties of likelihood estimators based on MRA. In this context, one could consider fixed-domain asymptotics (that is, increasing sample size within a fixed time domain) or increasing-domain asymptotics (where the range of times increases with the sample size). As pointed out by a referee, Theorem 5.1 suggests that consistency of the maximum likelihood estimator based on MRA holds under certain regularity assumptions. However, this result would need to be established rigorously and a study of the asymptotic covariance matrix of this estimator should be undertaken. The latter might be done via some variant of the information sandwich approach (Liang and Zeger, 1996), as used recently by Stein *et al.* (1994) in a spatial context. From a Bayesian viewpoint, one might study the behaviour of the posterior distribution as sample size increases, although the Bayesian approach generally places more emphasis on exact finite sample results.

## 8 Acknowledgements

We are grateful to two anonymous Referees for comments that helped us improve the paper. We also thank Professor Jacques Lévy Véhel for his help with comparisons with other estimation methods. Carmen Fernández was supported by a Nuffield Foundation grant (NAL/00419/G), under the scheme ‘Awards to Newly Appointed Science Lecturers’.

## 9 Appendix A: covariance matrix of multiresolution approximation

From the conditional properties of the multivariate Gaussian distribution, the MRA is equivalent to defining

$$\varepsilon_\theta(t_C) = (B_\theta(t_C))_1 \varepsilon_\theta(t_L) + (B_\theta(t_C))_2 \varepsilon_\theta(t_R) + \eta(t_C), \quad (9.1)$$

where  $(B_\theta(t_C))_1$  and  $(B_\theta(t_C))_2$  are the entries of  $B_\theta(t_C)$  in (4.9) in Section 4.1, and where  $\eta(t_C)$  has a zero-mean Normal distribution [with variance given by  $z_\theta(t_C)$  in (4.9)]. We can find the covariance between any two  $\varepsilon_\theta(t_i)$  and  $\varepsilon_\theta(t_j)$  in the MRA by using this construction iteratively, starting at times  $t_i$  and  $t_j$  and going upwards in the graph until the first common parents-child triple  $(t_L, t_C, t_R)$  is reached. Assuming  $t_i < t_j$ , we can then express  $\varepsilon_\theta(t_i)$  in the MRA as a linear combination of  $\varepsilon_\theta(t_L)$  and  $\varepsilon_\theta(t_C)$  plus a zero-mean Normal random error, and the same holds for  $\varepsilon_\theta(t_j)$  with  $\varepsilon_\theta(t_C)$  and  $\varepsilon_\theta(t_R)$  in the linear combination. The two Normal random errors are independent and using that the distribution of the MRA equals the original process for each parents-child triple, the covariance between  $\varepsilon_\theta(t_i)$  and  $\varepsilon_\theta(t_j)$  in the MRA follows immediately.

This is better understood with an example, so referring to Figure 3, we show how to compute the covariance between  $\varepsilon_\theta(t_{10})$  and  $\varepsilon_\theta(t_{16})$  in the MRA. Starting from the bottom of the graph,  $(t_9, t_{13}, t_{17})$  is the first parents-child triple that is a common ancestor. Applying (9.1) subsequently to  $\varepsilon_\theta(t_{10})$  and  $\varepsilon_\theta(t_{11})$ , we get

$$\begin{aligned} \varepsilon_\theta(t_{10}) &= (B_\theta(t_{10}))_1 \varepsilon_\theta(t_9) + (B_\theta(t_{10}))_2 \varepsilon_\theta(t_{11}) + \eta(t_{10}) \\ &= (B_\theta(t_{10}))_1 \varepsilon_\theta(t_9) + (B_\theta(t_{10}))_2 \left[ (B_\theta(t_{11}))_1 \varepsilon_\theta(t_9) + (B_\theta(t_{11}))_2 \varepsilon_\theta(t_{13}) + \eta(t_{11}) \right] + \eta(t_{10}) \\ &= b_\theta(t_{10}, t_9) \varepsilon_\theta(t_9) + b_\theta(t_{10}, t_{13}) \varepsilon_\theta(t_{13}) + \psi(t_{10}), \end{aligned}$$

with the appropriate definitions of the coefficients  $b_\theta(\cdot)$  and where  $\psi(t_{10})$  is a zero-mean Normal random variable. Similarly, we obtain  $\varepsilon_\theta(t_{16}) = b_\theta(t_{16}, t_{13}) \varepsilon_\theta(t_{13}) + b_\theta(t_{16}, t_{17}) \varepsilon_\theta(t_{17}) + \psi(t_{16})$ . Using the independence between  $\psi(t_{10})$  and  $\psi(t_{16})$ , together with the fact that the approximation is exact on each parents-child triple, we obtain the covariance in the MRA:

$$\begin{aligned} \text{Cov}[\varepsilon_\theta(t_{10}), \varepsilon_\theta(t_{16})] &= b_\theta(t_{10}, t_{13}) b_\theta(t_{16}, t_{13}) \Sigma_\theta(t_{13}t_{13}) + b_\theta(t_{10}, t_9) b_\theta(t_{16}, t_{13}) \Sigma_\theta(t_9 t_{13}) \\ &\quad + b_\theta(t_{10}, t_9) b_\theta(t_{16}, t_{17}) \Sigma_\theta(t_9 t_{17}) + b_\theta(t_{10}, t_{13}) b_\theta(t_{16}, t_{17}) \Sigma_\theta(t_{13} t_{17}). \end{aligned}$$

## 10 References

Adler, R.J. (1981) *Random Fields*. Chichester: Wiley.

- Ayache, A., Cohen, S. and Lévy Véhel, J. (2000) The covariance structure of multifractional Brownian motion, with application to long range dependence. ICASSP'00, June 2000, Istanbul.
- Barndorff-Nielsen, O.E., and Shephard, N. (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. Roy. Statist. Soc., Ser. B*, **63**, 167-241.
- Barrière, O. and Lévy Véhel, J. (in preparation) Improving the bias-variance trade-off for regression-based estimators.
- Benassi, A., Cohen, S., and Istas, J. (1998) Identifying the multifractional function of a Gaussian process. *Statist. Probab. Lett.*, **39**, 337-345.
- Benassi, A., Jaffard, S., and Roux, D. (1997) Elliptic Gaussian random processes and pseudodifferential elliptic operators. *Rev. Mat. Iberoamericana*, **13**, 19-89.
- Beran, J. (1994) *Statistics for Long-Memory Processes*. New York: Chapman and Hall.
- Cairngorm weather webpage (n.d.). Department of Physics, Heriot Watt University.  
<http://www.phy.hw.ac.uk/resrev/aws/awsarc.htm> Accessed on 9 June, 2003.
- Cohen, S. (1999) From self-similarity to local self-similarity: the estimation problem. In *Fractals - Theory and Applications in Engineering* (eds M. Dekking, J. Lévy Véhel, E. Lutton and C. Tricot), pp 3-16. London: Springer.
- Daniel, M.M. and Willsky, A.S. (1997) Modelling and estimation of fractional Brownian motion using multiresolution stochastic processes. In *Fractals in Engineering* (eds J. Lévy Véhel, E. Lutton and C. Tricot), pp 124-137. London: Springer.
- Engle, R.F. (1995) *ARCH: Selected Readings*. Oxford University Press.
- FracLab (n.d.) A fractal analysis toolbox for signal and image processing.  
<http://www.irccyn.ec-nantes.fr/hebergement/FracLab/> Accessed on 2 February 2006.
- Gneiting, T. and Schlather, M. (2003) Stochastic models that separate fractal dimension and Hurst effect. *SIAM Review*, **46**, 269-282.
- Green, P.J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711-732.
- Haslett, J. and Raftery, A.E. (1989) Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. *Appl. Statist.*, **38**, 1-50.
- Hyndman, R.J. (n.d.) *Time Series Data Library*. (FVD2.DAT)  
<http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/> Accessed on 24 April, 2003.
- Huang, H.-C., Cressie, N. and Gabrosek, J. (2002) Fast, resolution-consistent spatial prediction

- of global processes from satellite data. *J. Comput. Graph. Statist.*, **11**, 63-88.
- Jaffard, S. and Meyer, Y. (1996) Wavelet methods for pointwise regularity and local oscillations of functions. *Mem. Amer. Math. Soc.*, **123**, no. 587.
- Johannesson, G., Cressie, N. and Huang, H.-C. (2007) Dynamic multi-resolution spatial models. *Environm. Ecolog. Statist.*, forthcoming.
- Kent, J.T. and Wood, A.T.A. (1997) Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. Roy. Statist. Soc., Ser. B*, **55**, 679-699.
- Kukush, A., Mishura, Y. and Valkeila, E. (2005) Statistical inference with fractional Brownian motion. *Statist. Decisions*, **21**, 219-260.
- Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.
- Lovejoy, S. and Schertzer, D. (1995) Multifractals and rain. In *New Uncertainty Concepts in Hydrology and Hydrological Modelling* (ed Z.W. Kundzewicz) pp 109-113. Cambridge: Cambridge University Press.
- Mandelbrot, B.B. (1997) *Fractals and Scaling in Finance*. New York: Springer.
- Mandelbrot, B.B. (2002) *Gaussian Self-Affinity and Fractals*. New York: Springer.
- Mandelbrot, B.B. and Van Ness, J. (1968) Fractional Brownian motion, fractional noises and applications. *SIAM Rev.*, **10**, 422-437.
- Patil, G.P. and Taillie, C. (2001) A multiscale hierarchical Markov transition matrix model for generating and analyzing thematic raster maps. *Environm. Ecolog. Statist.*, **8**, 71-84.
- Peltier, R.F. and Lévy Véhel, J. (1995) Multifractional Brownian motion: definition and preliminary results. *Rapport de Recherche de l'INRIA*, No. 2645.
- Rogers, L.C.G. (1997) Arbitrage with fractional Brownian motion. *Math. Finance*, **7**, 95-105.
- Shephard, N. (2005) *Stochastic Volatility: Selected Readings*. Oxford University Press.
- Shiryaev, A.N. (1999) *Essentials of Stochastic Finance. Facts, Models, Theory*. Singapore: World Scientific.
- Sottinen, T. and Valkeila, E. (2003) On arbitrage and replication in the fractional Black-Scholes pricing model. *Statist. Decis.*, **21**, 93-107.
- Stein, M.L., Chi, Z. and Welty, L.J. (2004) Approximating likelihoods for large spatial data sets. *J. Roy. Statist. Soc., Ser. B*, **66**, 275-296.
- Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.
- Willinger, W., Taqqu, M.S. and Teverovsky, V. (1999) Stock market prices and long-range dependence. *Finance Stoch.*, **3**, 1-13.