Sørbye, S. H., Myrvoll-Nilsen, E., & Rue, H. (2019).
An approximate fractional Gaussian noise model with O(n) computational cost.
Statistics and Computing, 29, 821–833.

DOI: 10.1007/s11222-018-9843-1

Noname manuscript No.
(will be inserted by the editor)

An approximate fractional Gaussian noise model with $\mathcal{O}(n)$ computational cost

Sigrunn H. Sørbye[1] · **Eirik Myrvoll-Nilsen**[1] · **Håvard Rue**[2]

Abstract Fractional Gaussian noise (fGn) is a stationary time series model with long memory properties applied in various fields like econometrics, hydrology and climatology. The computational cost in fitting an fGn model of length n using a likelihood-based approach is $\mathcal{O}(n^2)$, exploiting the Toeplitz structure of the covariance matrix. In most realistic cases, we do not observe the fGn process directly but only through indirect Gaussian observations, so the Toeplitz structure is easily lost and the computational cost increases to $\mathcal{O}(n^3)$. This paper presents an approximate fGn model of $\mathcal{O}(n)$ computational cost, both with direct or indirect Gaussian observations, with or without conditioning. This is achieved by approximating fGn with a weighted sum of independent first order autoregressive (AR) processes, fitting the parameters of the approximation to match the autocorrelation function of the fGn model. The resulting approximation is stationary despite being Markov and gives a remarkably accurate fit using only four AR components. Specifically, the given approximate fGn model is incorporated within the class of latent Gaussian models in which Bayesian inference is obtained using the methodology of integrated nested Laplace approximation. The performance of the approximate fGn model is demonstrated in simulations and two real data examples.

✉ Sigrunn H. Sørbye

sigrunn.sorbye@uit.no

[1] Department of Mathematics and Statistics, UiT The Arctic University of Norway, 9037 Tromsø, Norway.

[2] CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

Keywords Autoregressive process · Gaussian Markov random field · Integrated nested Laplace approximation · Long-range dependence · Toeplitz matrix

1 Introduction

Many natural processes observed in either time or space exhibit a long-range dependency structure, here referred to as long memory. One way to characterise long memory is in terms of the autocorrelation function having a slower than exponential decay as a function of increasing temporal or geographical distance between observational points. In second-order stationary time series, long memory implies that the autocorrelations are not absolutely summable (McLeod and Hipel 1978). Long memory behaviour has been observed within a vast variety of time series applications, like hydrology (Hurst 1951; Hosking 1984), geophysical time series (Mandelbrot and Wallis 1969b), network traffic modeling (Willinger et al. 1996), econometrics (Baillie 1996; Cont 2005) and climate data analysis (Franzke 2012; Rypdal and Rypdal 2014). For comprehensive introductions to long memory processes and their applications, see for example Doukhan et al. (2003) and Beran et al. (2013).

When introduced in Mandelbrot and Van Ness (1968), fractional Brownian motion (fBm) represented a first paradigmatic example of a long memory model. The fBm is a time-continuous Gaussian process which generalizes ordinary Brownian motion to allow for dependent increments. A unique property of fBm is that it is exactly self-similar and its long memory properties are characterized by the self-similarity parameter H, also named the Hurst exponent. Of specific interest is the discrete-time increment process of fBm, referred to as fractional Gaussian noise (fGn). This was the first stationary Gaussian process which could explain the famous Hurst phenomenon (Hurst 1951). Since its introduction, the fGn process has been applied in a variety of applications ranging from hydrology (Molz et al. 1997), analysis of functional magnetic resonance images (Maxim et al. 2005) and climate analysis (Rypdal and Rypdal 2014).

Inheriting the parsimonious parameterization of fBm, the autocorrelation function of fGn is fully specified as a function of H and the model has long memory when $1/2 < H < 1$. Exploiting the Toeplitz structure of the covariance matrix, the computational cost of likelihood-based inference in fitting an fGn process of length n is $\mathcal{O}(n^2)$, making use of the Durbin-Levinson or Trench algorithms (Levinson 1947; Durbin 1960; Trench 1964; Golub and Loan 1996; McLeod et al. 2007). However, the required Toeplitz structure is fragile to modifications of the Gaussian observational model and computations of conditional distributions. For example, the Toeplitz structure is destroyed if fGn is observed indirectly with Gaussian inhomogeneous noise, or has missing data. In these situations, the computational cost of likelihood-based inference would increase to $\mathcal{O}(n^3)$.

This paper presents an accurate and computationally efficient approximate fGn model of cost $\mathcal{O}(n)$, both with direct and indirect Gaussian observations, with or without additional conditioning. This allows for routinely use of fGn

models with large n, with negligible loss of accuracy. The new approximate model uses a weighted sum of independent first order autoregressive processes (AR(1)). The motivation is that aggregation of short-memory processes plays an important role to explain long memory behaviour in time series (Beran et al. 2010) and an infinite weighted sum of AR(1) processes will give long memory (Granger 1980). In practice, the number of aggregated processes might need to be rather large to reflect the underlying long-memory parameter (Haldrup and Valdés 2017). However, the new approximate model only needs a weighted sum of four AR(1) processes to be accurate. This is obtained by fitting the weights and the coefficients of the AR(1) processes to mimic the autocorrelation function of the exact fGn model, as a continuous function of H.

A key feature of the approximate fGn model is a high degree of conditional independence within the model. Specifically, the approximate model will be represented as a Gaussian Markov random field (GMRF), the computational properties of which are not destroyed by indirect Gaussian observations nor conditioning (Rue and Held 2005). The approximate model is also stationary, a desired property which is not common among GMRFs as they typically have boundary effects. Since the approximate model is a local GMRF, it also fits well within the framework of latent Gaussian models for which approximate Bayesian analysis is obtained with integrated nested Laplace approximations (INLA) (Rue et al. 2009) using the R-package `R-INLA` (`www.r-inla.org`). This provides a flexible modelling framework in which the approximate fGn model can be combined with for example time trends, linear and nonlinear covariate effects and other random effects to build realistically complex models for observed time series. A different aspect is that an aggregated model of a few AR(1) components could actually represent a more plausible and interpretable model than the theoretical fGn process in real data applications. Specifically, the approximate model can serve as a tool for automatic source separation in situations where the data at hand represent combined signals. Among others, this can be useful in continuous-time climate data analysis in which a weighted sum of Ornstein-Uhlenbeck processes have been linked with multibox energy balance models (Fredriksen and Rypdal 2017).

A flexible alternative in modelling long memory processes is to use the framework of autoregressive fractionally integrated moving average (ARFIMA) models (Granger and Joyeux 1980; Hosking 1981). These models represent a natural extension of the classical ARIMA-models (Box and Jenkins 1980) and can be used to model both short and long range dependency structures simultaneously. If the order of the autoregressive (AR) and moving average (MA) parts are both 0, the resulting ARFIMA$(0,d,0)$ model has very similar properties to fGn when the long memory parameter $d = H - 0.5$. Both of these models exhibit the same hyperbolic decay of the autocorrelation function. A conceptual distinc-

tion between these two models is that fGn can be considered as a discrete approximation to the fractional derivative of the time-continuous Brownian motion (Hosking 1981). In contrast, AFRIMA$(0,d,0)$ is obtained by fractional differencing of the ARIMA$(0,1,0)$ model which is by definition discrete. This natural extension of ARIMA$(0,1,0)$ can be seen as an advantage of ARFIMA$(0,d,0)$ models compared with fGn (Graves et al. 2017). On the other hand, many asymptotic relations of fGn processes also hold for finite samples (Taqqu et al. 1995) and fGn-based models are very popular due to their analytic tractability (Purczyński and Wlodarski 2006).

This paper focuses on approximating fGn but as demonstrated in Section 3.4 the same method can be used to approximate ARFIMA$(0,d,0)$ models. We note that by using algorithms based on the fast Fourier transform, ARFIMA-models can be fitted with $\mathcal{O}(n\log(n))$ computational cost (Jensen and Nielsen 2014) when the process is observed directly. The fast Fourier transform can also be used to give a computationally efficient infinite sum approximation of fGn, reducing the computational cost of the Whittle estimator (Purczyński and Wlodarski 2006). Chan and Palma (2006) gives a review of different likelihood-based methods to fit ARFIMA-models. These include an approximate state-space algorithm using the Kalman filter, which can also be modified to analyse time series with missing data (Palma and Chan 1997). The approximation uses a truncated state-space approach, representing the ARFIMA-model by a moving average process with $M \approx 30$ terms. The general cost of this algorithm is $\mathcal{O}(nM^3)$, including n updates and inversion of $M \times M$ matrices. Knorr-Held and Rue (2002) describe a general GMRF framework which includes state-space models for time series. They describe the relation between the Kalman filter and a Cholesky factor approach, stating that the latter is both conceptually simpler and computationally more efficient as calculations are done only once for a band-matrix of dimension nM.

The presented approach takes advantage of both the GMRF-framework (Rue and Held 2005) and the Cholesky factorisation (Rue 2001; Knorr-Held and Rue 2002) to provide the approximate fGn model using a truncation with only four terms. The plan of this paper is as follows. Section 2 reviews the computational cost in fitting the exact fGn model to direct and indirect Gaussian observations. Section 3 presents the new approximate fGn model and derives the computational cost and memory requirement for evaluating the log-likelihood. The performance of the approximate model is demonstrated by simulations in Section 4, also including a study of its predictive properties. In Section 5, we use the implicit source separation ability in decomposing the historical dataset of annual water level minima for the Nile river (Toussoun 1925; Beran 1994). Implementation within the class of latent Gaussian models is demonstrated by analysing a monthly mean surface air temperature

series for Central England (Manley 1953, 1974; Parker et al. 1992). For comparison, we also include results using the approximate ARFIMA$(0,d,0)$ model for these two datasets. Concluding remarks are given in Section 6.

2 The computational cost of fGn

Let $\boldsymbol{x} = (x_1,\ldots,x_n)^T$ be a zero-mean fGn process, $\boldsymbol{x} \sim \mathcal{N}_n(0,\boldsymbol{\Sigma})$. The elements of the covariance matrix, $\Sigma_{ij} = \sigma^2\gamma_{\boldsymbol{x}}(k)$ where $k = |i-j|$, are defined by the autocorrelation function

$$\gamma_{\boldsymbol{x}}(k) = \frac{1}{2}\left(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\right), \quad k = 0,1,\ldots,n-1.$$

This function has a hyperbolic decay, $\gamma_{\boldsymbol{x}}(k) \sim H(2H-1)k^{2(H-1)}$ as $k \to \infty$. The fGn process is persistent when $1/2 < H < 1$, while it reduces to white noise when $H = 1/2$. When $0 < H < 1/2$, the fGn model has anti-persistent properties, but we do not discuss this case here.

When fGn is observed directly, we estimate H by maximizing the log-likelihood function

$$\log(\pi(\boldsymbol{x})) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{Q}| - \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x},$$

where $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ is the precision matrix of $\boldsymbol{x}$. Making use of the Toeplitz structure of $\boldsymbol{\Sigma}$, the likelihood is evaluated in $\mathcal{O}(n^2)$ flops using the Durbin-Levinson algorithm (Golub and Loan 1996, alg.4.7.2). Also, the precision matrix $\boldsymbol{Q}$ can be calculated in $\mathcal{O}(n^2)$ flops by the Trench algorithm (Golub and Loan 1996, Algorithm 4.7.3). In general, the Trench algorithm can be combined with the Durbin-Levinson recursions (Golub and Loan 1996, alg. 5.7.1), to calculate the exact likelihood of general linear Gaussian time series models (McLeod et al. 2007).

A major drawback of relying on these algorithms for Toeplitz matrices is that the Toeplitz structure is easily destroyed if the time series is observed indirectly. Consider a simple regression setting in which an fGn process is observed with independent Gaussian random noise,

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\boldsymbol{0},\boldsymbol{D}^{-1})$ and $\boldsymbol{D}$ is diagonal. The log-density of $\boldsymbol{x} \mid \boldsymbol{y}$ is

$$\log\pi(\boldsymbol{x} \mid \boldsymbol{y}) = \frac{1}{2}\log|\boldsymbol{Q}+\boldsymbol{D}| - \frac{1}{2}\boldsymbol{x}^T(\boldsymbol{Q}+\boldsymbol{D})\boldsymbol{x} + \boldsymbol{y}^T\boldsymbol{D}\boldsymbol{x} + \text{constant}. \tag{2}$$

The conditional mean of $\boldsymbol{x}$ is found by solving $(\boldsymbol{Q}+\boldsymbol{D})\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{D}\boldsymbol{y}$ with respect to $\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}}$, while the marginal variances equal $\text{diag}\{(\boldsymbol{Q}+\boldsymbol{D})^{-1}\}$. The Toeplitz structure of $\text{Cov}(\boldsymbol{y}) = \boldsymbol{Q}^{-1} + \boldsymbol{D}^{-1}$ is only retained when the noise term has homogeneous variance, i.e. $\boldsymbol{D}^{-1} \propto \boldsymbol{I}$. With non-homogenous observation variance or missing data, the computational cost in fitting (1) would require general algorithms of cost $\mathcal{O}(n^3)$. This makes analysis of many real data sets infeasible, or at best challenging.

The motivation for expressing the log-likelihood function in terms of the precision matrix $\boldsymbol{Q}$, is to prepare for an approximate GMRF representation of the fGn model. We have already noted that aggregation of an infinite number

of short-memory processes can explain long memory behaviour in time series. This implies that $\boldsymbol{Q}$ is (or can be approximated with) a sparse band matrix, but with a larger dimension (still denoted by n) for a finite sum. Assume for a moment that such an approximation exists. We can then apply general numerical algorithms for sparse matrices which only depend on the non-zero structure of the matrix. This implies that the numerical cost in dealing with $\boldsymbol{Q}$ or $\boldsymbol{Q}+\boldsymbol{D}$, is the same. Conditioning on subsets of $\boldsymbol{x}$ implies nothing else than working with a submatrix of $\boldsymbol{Q}$ or $\boldsymbol{Q}+\boldsymbol{D}$, and does not add to the computational costs; see (Rue and Held 2005, chap. 2) for details. Specifically, we can make use of the Cholesky decomposition, in which the relevant precision matrix $\boldsymbol{Q}+\boldsymbol{D}$ is factorised as $\boldsymbol{Q}+\boldsymbol{D}=\boldsymbol{L}\boldsymbol{L}^T$, where $\boldsymbol{L}$ is a lower triangular matrix. The log-likelihood is then evaluated with negligible cost (Rue 2001), as the log-determinant is $\log|\boldsymbol{Q}+\boldsymbol{D}| = 2\sum_{i=1}^{n} L_{ii}$. The conditional mean is found by solving $\boldsymbol{L}\boldsymbol{u}=\boldsymbol{D}\boldsymbol{y}$ and $\boldsymbol{L}^T\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}}=\boldsymbol{u}$. The numerical cost in finding the Cholesky decomposition depends on the non-zero structure of the matrix. For time-series models (or long skinny graphs), the cost is $\mathcal{O}(n)$ (Rue and Held 2005). The explicit construction of such an approximation is discussed next.

3 An approximate fGn model

This section presents an approximate fGn model which is a weighted sum of a few independent AR(1) processes. We will fit the parameters of the approximation to match the autocorrelation structure of fGn up to a given finite lag. The resulting approximate model is a GMRF with a banded precision matrix of fixed bandwidth, which gives a computational cost of $\mathcal{O}(n)$.

3.1 Fitting the autocorrelation function

Define m independent AR(1) processes by

$$z_{j,t} = \phi_j z_{j,t-1} + \nu_{j,t}, \qquad j = 1,\ldots,m, \quad t = 1,\ldots,n, \tag{3}$$

where $0<\phi_j<1$ denotes the first-lag autocorrelation coefficient of the jth AR(1) process. Further, let $\{\nu_{j,t}\}_{j=1}^m$ be independent zero-mean Gaussians, with variance $\sigma^2_{\nu,j} = 1-\phi_j^2$. Define the cross-sectional aggregation of the AR(1) processes,

$$\tilde{\boldsymbol{x}}_m = \sigma \sum_{j=1}^{m} \sqrt{w_j}\boldsymbol{z}^{(j)}, \tag{4}$$

where $\boldsymbol{z}^{(j)} = (z_{j,1},\ldots,z_{j,n})^T$ and $\sum_{j=1}^m w_j = 1$. The finite-sample properties of a similar aggregation of AR(1) processes are studied in Haldrup and Valdés (2017), where $w_j = 1/m$, $\sigma_\nu^2 = 1$ and where the coefficients ϕ_j are Beta distributed. They conclude that "*First of all, one should be aware that cross-sectional aggregation leading to long memory is an asymptotic feature that applies for the cross-sectional dimension tending to infinity. In finite samples and for moderate cross-sectional dimensions the observed memory of the series can be rather different from the theoretical memory*" (Haldrup and Valdés 2017, pp. 7-8).

The approximation presented here only needs a small value of the cross-sectional dimension m to be accurate. The key idea to our approach is to fit the weights $\boldsymbol{w} = \{w_j\}_{j=1}^m$ and the autocorrelation coefficients $\boldsymbol{\phi} = \{\phi_j\}_{j=1}^m$ in (4) to match the autocorrelation function of fGn, as a function of H. The autocorrelation function of (4) follows directly as

$$\gamma_{\tilde{\boldsymbol{x}}_m}(k) = \sum_{j=1}^m w_j \phi_j^{|k|}, \quad k = 0, 1, \ldots, n-1.$$

Now, fix a value of $1/2 < H < 1$. We fit the weights and coefficients $(\boldsymbol{w}, \boldsymbol{\phi})_H$ by minimizing the weighted squared error

$$(\boldsymbol{w}, \boldsymbol{\phi})_H = \underset{(\boldsymbol{w},\boldsymbol{\phi})}{\operatorname{argmin}} \sum_{k=1}^{k_{\max}} \frac{1}{k} \big(\gamma_{\tilde{\boldsymbol{x}}_m}(k) - \gamma_{\boldsymbol{x}}(k)\big)^2, \tag{5}$$

where $k_{\max}$ represents a user-specified upper limit (we use $k_{\max} = 1000$). The squared error is weighted by $1/k$ to ensure a good fit for the autocorrelation function close to lag 0, while less weight is given to tail behaviour as the autocorrelation function is decaying slowly.

By a quite huge calculation done only once, we find $(\boldsymbol{w}, \boldsymbol{\phi})_H$ for a fine grid of H-values. Spline interpolation is used for values of H in between, to represent the weights and coefficients as continuous functions of H. The interpolation and fitting are performed using reparameterised weights and coefficients to ensure uniqueness and improved numerical behaviour. These reparameterisations are defined as

$$w_j = \frac{e^{v_j}}{\sum_{i=1}^m e^{v_i}} \quad \text{and} \quad \phi_j = \frac{1}{1 + \sum_{i=1}^j e^{-u_i}},$$

where $j = 1, \ldots, m$ and where $v_1 = 0$. The Hurst exponent is transformed as $H = 1/2 + 1/2 \exp(h)/(1+\exp(h))$. This ensures a stable and unconstrained parameter space on $\mathbb{R}^{2m-1}$ for fixed h, where $\phi_1 > \cdots > \phi_m$. Note that the error of the fit tends to zero, when H goes to 1 or $1/2$. The resulting coefficients and weights for $m = 3$ and $m = 4$ are displayed in Figure 1. The fitted weights and coefficients are also available in R using the function `INLA::inla.fgn`.

3.2 A Gaussian Markov random field representation

We will now discuss the precision matrix for the approximate fGn model. We start with one AR(1) process (3) of length n, with unit variance and a tridiagonal precision matrix

$$\boldsymbol{R}(\phi_j) = \frac{1}{1-\phi_j^2} \begin{pmatrix} 1 & -\phi_j & & & \\ -\phi_j & 1+\phi_j^2 & -\phi_j & & \\ & \ddots & \ddots & \ddots & \\ & & -\phi_j & 1+\phi_j^2 & -\phi_j \\ & & & -\phi_j & 1 \end{pmatrix}.$$

For the approximate fGn model, we have m such processes and their sum. Hence we need the $(m+1)n \times (m+1)n$ precision matrix of the vector

$$(\tilde{\boldsymbol{x}}_m^T, \boldsymbol{z}^{(1)T}, \ldots, \boldsymbol{z}^{(m)T}). \tag{6}$$

To ensure a non-singular distribution, we will add a small Gaussian noise term to the sum, i.e. we let

$$\tilde{\boldsymbol{x}}_m = \boldsymbol{\sigma} \left(\sum_{j=1}^m \sqrt{w_j} \boldsymbol{z}^{(j)} + \boldsymbol{\varepsilon} \right), \tag{7}$$

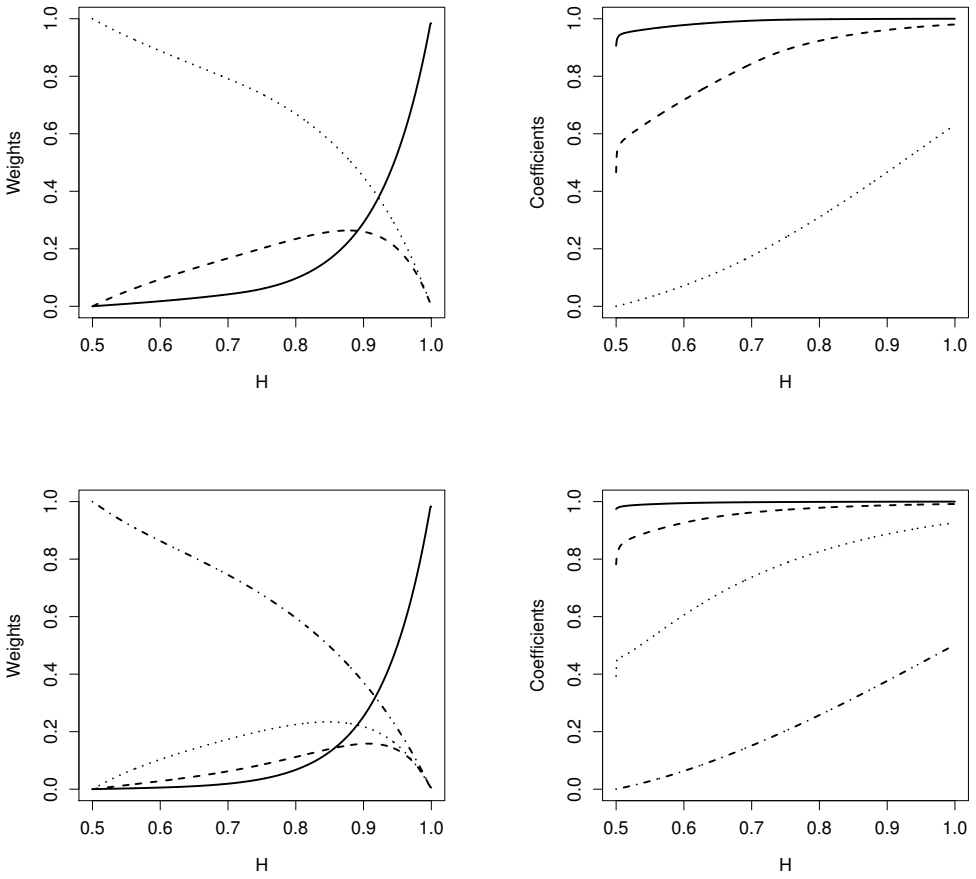

Fig. 1 The optimized weights and coefficients of the approximation as a function of H. These are found by solving (5) using $m = 3$ (upper panels) and $m = 4$ (lower panels) components in the approximation.

where the precision of $\boldsymbol{\varepsilon}$ is high, like $\kappa = \exp(15)$. The (upper part of the) precision matrix is found as

$$\begin{pmatrix} \kappa\boldsymbol{I}/\sigma^2 & -\sqrt{w_1}\kappa\boldsymbol{I}/\sigma & -\sqrt{w_2}\kappa\boldsymbol{I}/\sigma & \dots & -\sqrt{w_m}\kappa\boldsymbol{I}/\sigma \\ & \boldsymbol{R}(\phi_1)+w_1\kappa\boldsymbol{I} & \sqrt{w_1w_2}\kappa\boldsymbol{I} & \dots & \sqrt{w_1w_m}\kappa\boldsymbol{I} \\ & & \boldsymbol{R}(\phi_2)+w_2\kappa\boldsymbol{I} & \ddots & \vdots \\ & & & \ddots & \sqrt{w_{m-1}w_m}\kappa\boldsymbol{I} \\ & & & & \boldsymbol{R}(\phi_m)+w_m\kappa\boldsymbol{I} \end{pmatrix}.$$

The non-zero structure is displayed in Figure 2 (left panel) for $m = 3$ and $n = 10$. Even though the matrix is sparse, a more optimal structure can be achieved by grouping the $m+1$ variables associated with each of the n time points,

$$\boldsymbol{v}^T = \left(\tilde{x}_{m1}, z_1^{(1)}, \dots, z_1^{(m)}, \tilde{x}_{m2}, z_2^{(1)}, \dots, z_2^{(m)}, \dots, \tilde{x}_{mn}, z_n^{(1)}, z_n^{(2)}, \dots, z_n^{(m)}\right).$$

The benefit of this reordering is that the corresponding precision matrix $\boldsymbol{Q}_v$ is a band matrix, see Figure 2 (middle panel). Doing the Cholesky decomposition, $\boldsymbol{Q}_v = \boldsymbol{L}_v\boldsymbol{L}_v^T$, the lower triangular matrix $\boldsymbol{L}_v$ will inherit the lower bandwidth of $\boldsymbol{Q}_v$ (Rue 2001; Golub and Loan 1996, thm. 4.3.1), see Figure 2 (right panel). This leads to the following key re-

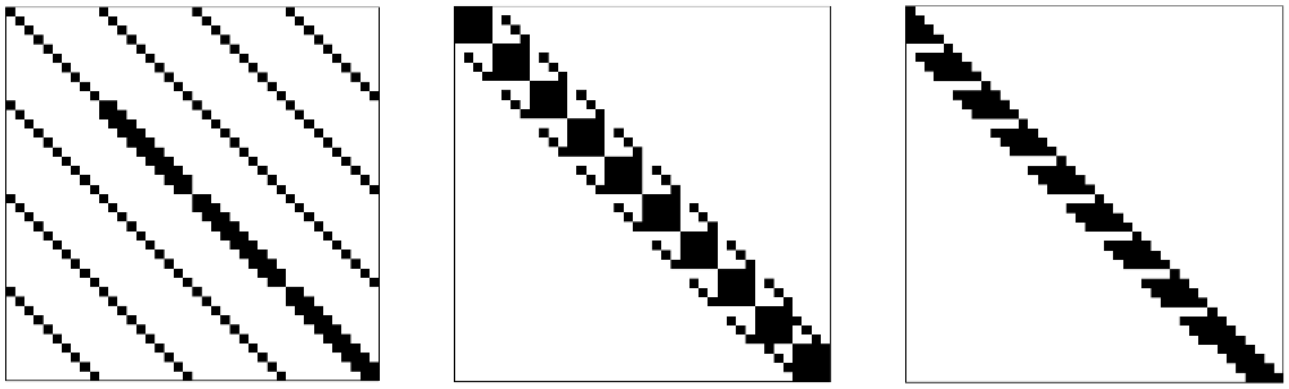

Fig. 2 The structure of the precision matrix of the vector in (6) (right panel), the structure of the precision matrix of the reordered vector in (8) (middle panel) and the resulting structure of the lower triangular matrix in the Cholesky decomposition (right panel). The matrices are illustrated for the case $m = 3$ and $n = 10$.

sult concerning the computational cost of the approximate model, with a trivial proof.

Theorem 1 *The number of flops needed for Cholesky decomposition of* $\boldsymbol{Q}_v$ *is* $n(m+1)^3$. *The memory requirement for the Cholesky triangle is* $n(m+1)(m+2)$ *reals.*

Proof $\boldsymbol{Q}_v$ is a band matrix with dimension $d = n(m+1)$ and bandwidth $b = m+1$. The computational cost of the Cholesky factorisation, $\boldsymbol{Q}_v = \boldsymbol{L}_v\boldsymbol{L}_v^T$ is $db^2 = n(m+1)^3$ and the memory requirement needed is $d(b+1) = n(m+1)(m+2)$ (Golub and Loan 1996, sec. 4.3.5).

The computational cost and memory requirement of the Cholesky decomposition do not change if the approximate fGn model is observed indirectly, like in the regression model (1). Also, the computational cost is much smaller than using the Kalman recursions for a truncated ARFIMA process. Notice that it is possible to construct a GMRF approximation which has an even lower computational cost by using the cumulative sums of $\sigma\sum_{j=1}^{m}\sqrt{w_j}\boldsymbol{z}^{(j)}$ to form a sparse $mn \times mn$ precision matrix, with the same bandwidth. However, this approach does not allow for automatic source separation in situations where the fGn can be seen to represent combined signals. This feature of the approximate model is demonstrated in Section 5.1.

3.3 Choosing the number of AR(1) components in the approximation

The choice of m in (7) reflects a trade-off between computational efficiency and approximation error. This implies that m should be as small as possible but still large enough to give a reasonably accurate approximation of the autocorrelation function of fGn. Figure 3 illustrates the autocorrelation function of fGn compared with the approximate model when $m = 3$ and $m = 4$, using $k_{\max} = 1000$ in (5). We only show

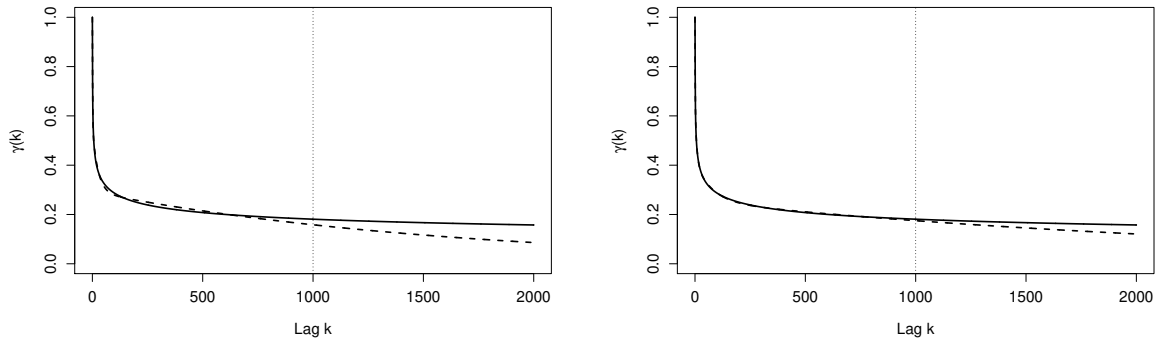

Fig. 3 The exact autocorrelation function (solid) when $H = 0.9$, versus the autocorrelation function of the approximate model (dashed), using $m = 3$ (left panel) and $m = 4$ (right panel). The autocorrelations of the approximate and exact models are matched up to lag $k_{max} = 1000$.

results for $H = 0.9$ as the differences between the curves will be less visible using smaller values of H.

We do notice that $m = 4$ gives an almost perfect match of the autocorrelation function up to $k_{\max}$. For larger lags, the autocorrelation function of the approximate fGn model will have an exponential decay, hence we cannot match the hyperbolic decay of the exact fGn. As a consequence, k_{max} can be seen as a trade-off between having a good fit for the first part of the autocorrelation function versus tail behaviour.

A different way to illustrate the difference between the approximate and exact fGn models is in terms of the Kullback-Leibler divergence. This is a measure of complexity between probability distributions, which here measures the information lost when the approximate fGn model is used instead of the exact fGn model. Figure 4 displays the square-root of the Kullback-Leibler divergence for $n = 500$ and $n = 2000$, as a function of H. We notice that $m = 4$ clearly gives an improvement over $m = 3$, in particular for larger values of H. The loss in information when $n = 2000$ compared to $n = 500$ is small, despite the fact that the autocorrelation function is fitted only up to lag $k_{max} = 1000$.

3.4 A note on ARFIMA-models

The presented approximation is especially suitable in fitting a parsimoniously parameterised model like fGn. The fact that the autocorrelation function of fGn is specified by only one parameter can be seen both as a modeling benefit but also as a limitation. An ARFIMA(p,d,q) model can be used to model both short and long memory dependency structures, both simultaneously and separately. The parameters p and q give the orders of the short memory autoregressive and moving average parts of the model. The long memory property is prescribed by the parameter d, giving the order of the fractional differencing of the underlying autoregressive moving average model.

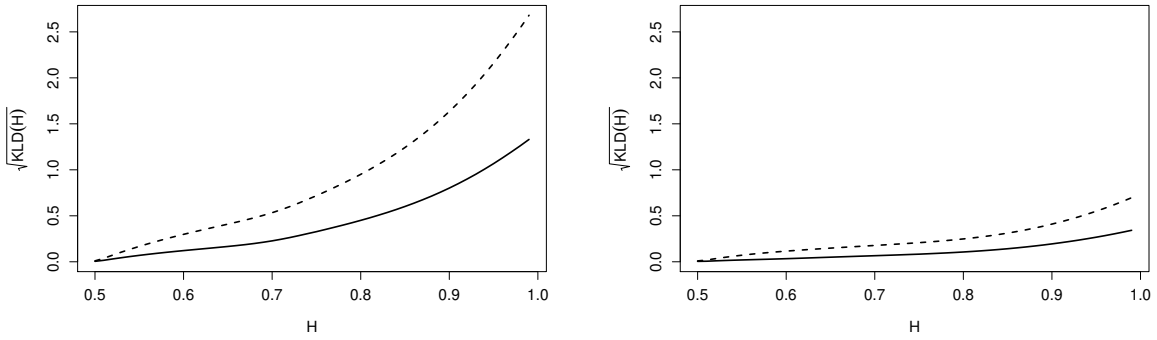

Fig. 4 The square-root of the Kullback-Leibler divergence as a function of H for time series of length $n = 500$ (solid) and $n = 2000$ (dashed), using the approximate fGn model with $m = 3$ (left panel) and $m = 4$ (right panel).

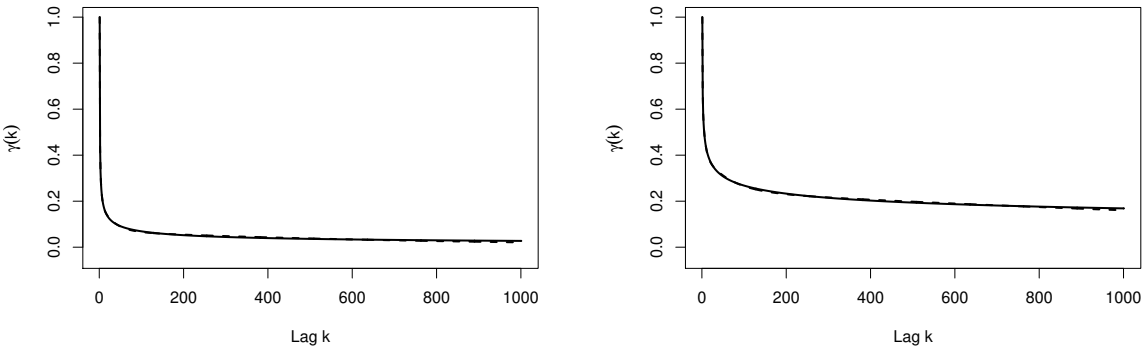

Fig. 5 The exact (solid) and approximated (dashed) autocorrelation function for ARFIMA(0,d,0), where $d = 0.3$ (left) and $d = 0.4$ (right).

The method of matching autocorrelation functions can easily be used to provide an approximation of ARFIMA$(0,d,0)$. Figure 5 displays the exact and approximated autocorrelation functions when $d = 0.3$ and $d = 0.4$. As illustrated, the approximation method is very accurate also in this case. This is not surprising as the ARFIMA$(0,d,0)$ model has very similar properties to fGn when $d = H - 0.5$. Extensions of the given approximation idea to the case of estimating all parameters of the ARFIMA model represents an interesting future project but we have considered this to be beyond the scope of this paper. We also note that ARFIMA-models have been criticised as having an atypical long-range dependency property, offering "*no meaningful diversity beyond fGn*" (Veitch et al. 2013, pp. 2). Even though the use of fGn processes or ARFIMA$(0,d,0)$ might seem limited, we gain flexibility by incorporating the given GMRF approximation within the general class of latent Gaussian models (Rue et al. 2009). This gives a user-friendly framework in which an fGn component can be combined with other explanatory effects in modeling real time series.

4 Simulation results

To evaluate the properties of the approximate fGn model, we now study the loss of accuracy when using the approximate versus the exact fGn model, for estimation and prediction. The results will demonstrate an impressive performance for both the estimation and prediction exercises, using the approximate fGn model with $m = 4$.

4.1 Maximum likelihood estimation of H

We first study the loss of accuracy using the approximate versus the exact fGn model in maximum likelihood estimation of H. We fit the approximate model using $m = 3$ and $m = 4$, to simulated fGn series of length $n = 500$, with $N = 1000$ replications. The error is evaluated in terms of the root mean squared error (RMSE) and the mean absolute error (MAE) of $\tilde{H}_i - \hat{H}_i$, where $\tilde{H}_i$ and $\hat{H}_i$ denote the estimates using the approximate versus the exact fGn, for the ith replication.

The results are summarised in Table 1 in which the true Hurst exponent ranges from 0.60 to 0.95. Using $m = 3$, the Hurst exponent is underestimated and the error is seen to increase with H, at least up to 0.90. The situation really improves for $m = 4$, in which the error is small for all values of H. The standard deviation estimates found from the empirical Fisher information, are more similar than the estimates themselves (results not shown).

Figure 6 displays scatterplots of the maximum likelihood estimates for the approximate model with $m = 3$ and 4, versus the estimates using the exact model, when $H = 0.7$, 0.8 and 0.9. The inaccuracy for $m = 3$ is clearly visible and increases with increasing values of H, while $m = 4$ shows very good performance. We have noticed that the same general remarks also hold when we increase the length of the series to $n = 2000$. The series then contain more information about H, and the error due to using $k_{max} = 1000$ is negligible. In conclusion, we do get a very low loss of accuracy using the approximate model with $m = 4$. This is impressive, especially as it applies for all reasonable values of H in the long memory range.

4.2 Predictive properties

This section investigates the effect of the approximation error when we observe an fGn process of length n with fixed H, and then want to predict future time points. The approximate model is implemented with $m = 4$. To evaluate the properties of the predictions, we consider the empirical mean of the standardised absolute prediction error,

$$\text{err}_\mu(p) = \frac{1}{N}\sum_{i=1}^{N} \frac{|\tilde{\mu}_{p,i} - \mu_{p,i}|}{\sigma_p}, \tag{8}$$

where N is the number of replications. $\tilde{\mu}_{p,i}$ is the conditional expectation for p time points ahead from the ith replication using the approximate fGn model. Correspondingly, $\mu_{p,i}$ is the conditional expectation using the exact fGn model while

Table 1 Maximum likelihood estimation of H in simulations. The columns show the average of the maximum likelihood estimates of H using the exact versus the approximate models with $m = 3$ and $m = 4$, the corresponding root mean squared error and the absolute mean error. The generated fGn processes are of length $n = 500$ with $N = 1000$ replications.

	Average MLE of H			RMSE($\tilde{H}$)		MAE($\tilde{H}$)	
H	Exact	$m = 3$	$m = 4$	$m = 3$	$m = 4$	$m = 3$	$m = 4$
0.60	0.5998	0.5998	0.5998	0.0019	0.0007	0.0015	0.0006
0.65	0.6481	0.6478	0.6480	0.0026	0.0008	0.0021	0.0006
0.70	0.7004	0.6997	0.7003	0.0033	0.0008	0.0026	0.0006
0.75	0.7488	0.7472	0.7487	0.0032	0.0007	0.0025	0.0006
0.80	0.7998	0.7974	0.7996	0.0031	0.0006	0.0026	0.0005
0.85	0.8503	0.8471	0.8500	0.0035	0.0004	0.0032	0.0004
0.90	0.8999	0.8965	0.8997	0.0035	0.0003	0.0034	0.0003
0.95	0.9500	0.9475	0.9499	0.0025	0.0002	0.0025	0.0001

σ_p is the conditional standard deviation. To measure the error in the conditional standard deviation, we use

$$\text{err}_\sigma(p) = \frac{\tilde{\sigma}_p}{\sigma_p} - 1, \tag{9}$$

which does not depend on the replication.

The left panel of Figure 7 illustrates the empirical prediction error in (8) for $p = 1, \ldots, 250$ time-points ahead, following either $n = 500$ or $n = 2000$ observations. The right panel shows the corresponding error in the prediction standard deviation (9). We only report results for $H = 0.8$ as other values of the Hurst exponent give similar results. We notice that the mean prediction error increases slightly when $n = 2000$ compared to $n = 500$, which is explained by the increased error for lags larger than $k_{\max} = 1000$. Otherwise, both errors are relatively small and also quite stable with p.

5 Real data applications: Source separation and Bayesian inference

This section demonstrates two different aspects of the approximate fGn model in real data applications. First, the approximate model can be used as a tool for source separation of a combined signal, for example representing underlying cycles or variations for different time scales. This will be illustrated in analysing the Nile river dataset (available in `R` as `longmemo::NileMin`). These data give annual water level minimas for the period 622 - 1284, measured at the Roda Nilometer near Cairo. Second, the approximate model can easily be combined with other model components within the general framework of latent Gaussian models and fitted efficiently using `R-INLA`. This is demonstrated in analysing the Hadley Centre Central England Temperature series (Had-

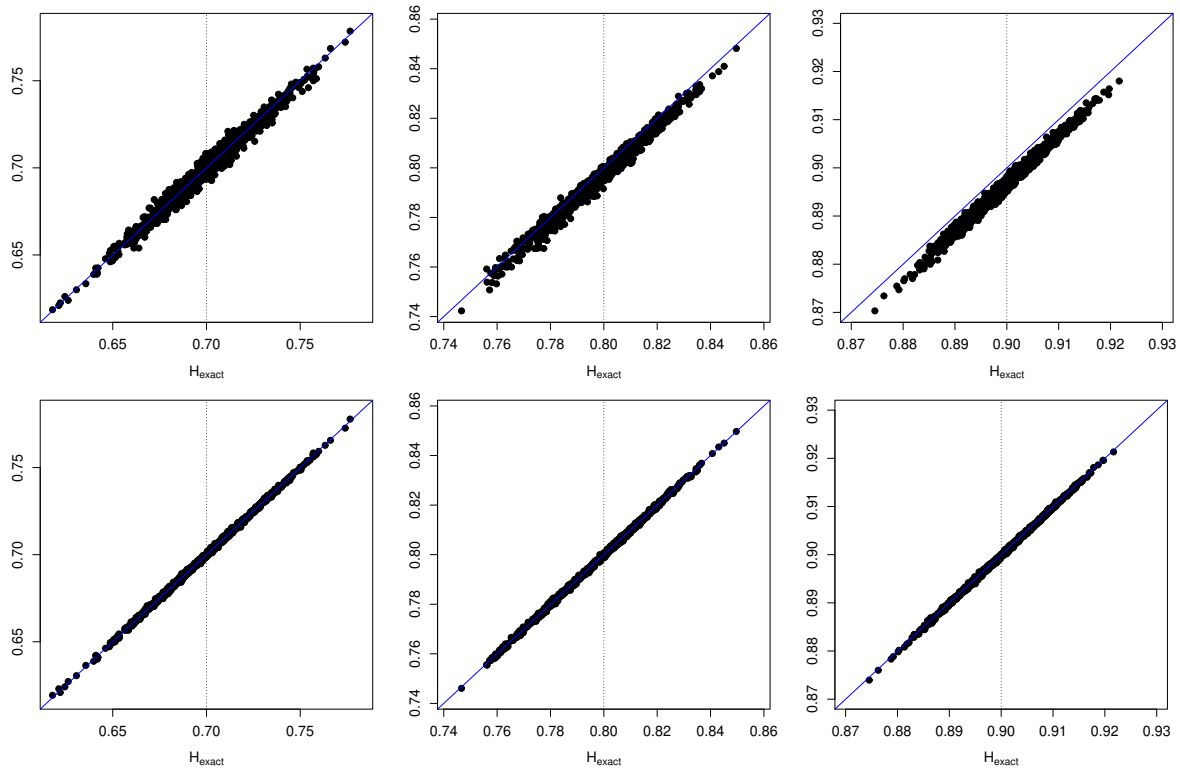

Fig. 6 The maximum likelihood estimates of H for $N = 1000$ replications using the approximate fGn model with $m = 3$ (upper panels) and $m = 4$ (lower panels), versus the exact fGn. The true H-values are $H = 0.7$ (left), $H = 0.8$ (middle) and $H = 0.9$ (right) and the generated series have length $n = 500$.

CET), available at `www.metoffice.gov.uk/hadobs`. These data give mean monthly measurements of surface air temperatures for Central England in the period 1659 - 2016. The two datasets are illustrated in Figure 8.

5.1 Signal separation for the Nile river annual minima

The Nile river dataset is a widely studied time series (Mandelbrot and Wallis 1969b; Beran 1994; Eltahir 1996) often used as an example of a real fGn process (Koutsoyiannis 2002; Benmehdi et al. 2011). Analysis of this dataset led to the discovery of the Hurst phenomenon (Hurst 1951). For hydrological time series, this phenomenon has been explained as the tendency of having irregular clusters of wet and dry periods and can be related to characteristics of the fluctuations of the series at different temporal scales (Koutsoyiannis 2002).

We can easily fit the exact fGn model to this dataset as the process is observed directly and the length of the series is only $n = 663$. In this case, the maximum likelihood estimate for the Hurst exponent is $\hat{H} = 0.831$. Using the approximate fGn model with $m = 4$, we get $\hat{H} = 0.829$. This illustrates that the approximate and exact models give very similar estimates. A value of $H \approx 0.83$ also corresponds well with results reported in literature (Man-

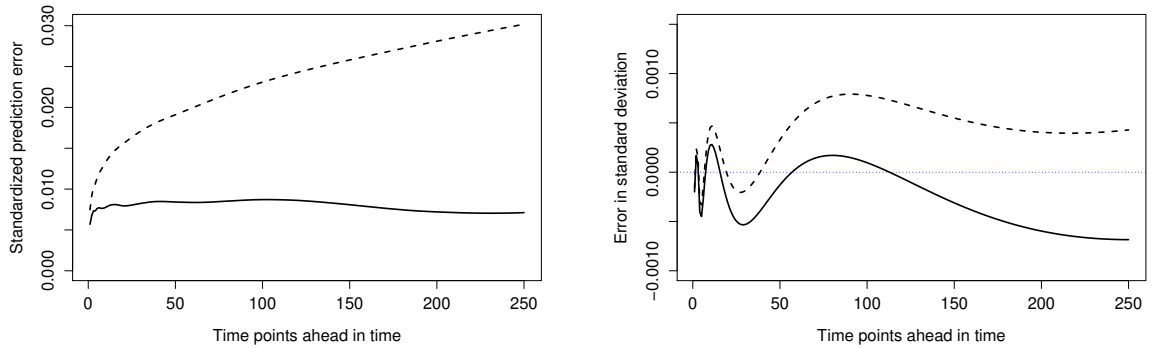

Fig. 7 The prediction error for the mean (8), predicting up to 250 points ahead, when $H = 0.8$ (left panel). The similar error in the predictive standard deviation (9) (right panel). The observed time series is either of length $n = 500$ (solid) or $n = 2000$ (dashed).

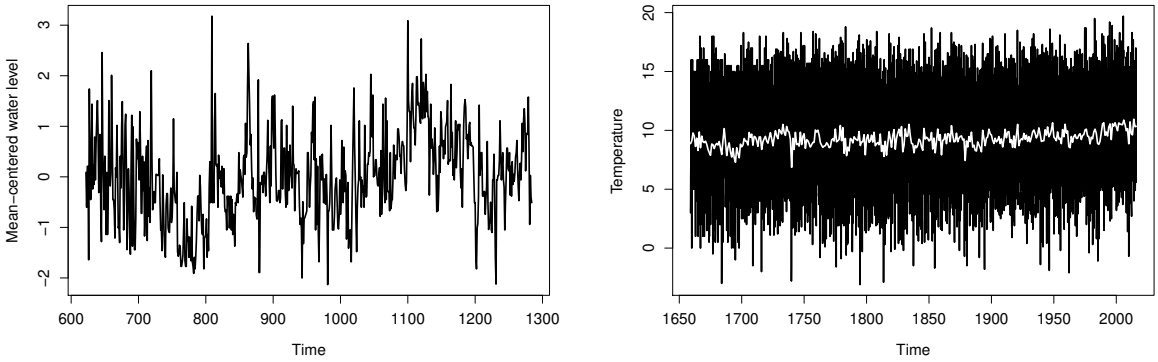

Fig. 8 Mean-centered annual minimum water level of the Nile river (left panel). Monthly mean surface air temperatures for Central England, also including the annual mean temperatures in white (right panel).

delbrot and Wallis 1969a; Benmehdi et al. 2011). Beran and Terrin (1996) states that both a fractional Gaussian noise with $H \approx 0.83$ and an ARFIMA$(0,d,0)$ model with $d = 0.4$ fits well for this dataset. Graves et al. (2015) analyse the given dataset fitting an ARFIMA$(0,d,0)$ model using an MCMC-approach. They get a posterior estimate of $d = 0.402$ with a 95% credible interval equal to $(0.336, 0.482)$. Our results using the approximated ARFIMA$(0,d,0)$ model is very similar giving $d = 0.399$ with a 95% credible interval equal to $(0.350, 0.438)$.

An advantage of the approximate versus the exact model is that the decomposition given by the mixture of AR(1) can be used as a source separation tool. Figure 9 illustrates the four estimated weighted AR(1) components in fitting model (5). The estimated autocorrelation coefficients for these components equal $\boldsymbol{\phi} = (0.999, 0.982, 0.847, 0.291)$, while the weights

are $\boldsymbol{w} = (0.099, 0.129, 0.232, 0.540)$. The estimated standard deviation is $\hat{\sigma} = 0.888$.

As illustrated in Figure 1, the first autocorrelation coefficient will always be quite close to 1. This gives a slowly varying trend, which in this case basically represents the mean. The second component also reflects a slowly varying signal, which can be interpreted to represent cycles of the water level fluctuations of about 200 - 250 years. The third component seems to reflect shorter cycles of length 30 - 100 years. These cycles are seen to appear more irregularly and we also notice the tendency of having clusters of years with high and low water levels, respectively. The fourth component, which has the smallest autocorrelation coefficient and the largest weight, can be interpreted as weakly correlated annual noise. In comparison, Kondrashov et al. (2005) study oscillatory modes of Nile River records using empirical orthogonal functions. Fitting a non-linear data-adaptive trend to the data for the period 622 - 1470, they detect a 256-year cycle. They also find quasi-quadrennial and quasi-biennial cycles in addition to periodicities of 64, 19, 12 and 7 years. A disadvantage of using the approximate fGn model as a source separation tool is that the number of modes is fixed to be equal to the number of AR(1) components used in the approximation. An advantage is that the resulting components do have a quite clear interpretation as these can be linked directly to the weights and autocorrelation coefficients of the approximation.

5.2 Full Bayesian analysis of a temperature series

The HadCET series is the longest existing instrumental record of monthly temperatures in the world. The observations started in January 1659 and have been updated monthly. The observed temperatures do have uncertainties (Parker and Horton 2005) and has been revised several times (Manley 1953, 1974; Parker et al. 1992). Especially, the measurements up to 1699 have a precision of 0.5 °C, while the precision is 0.1 °C thereafter (Graves et al. 2015). We analyse temperatures up to December 2016, which gives a total of $n = 4296$ observations.

We assume that the mean of the given temperatures can be modelled by

$$\mathrm{E}(y_t) = \beta_0 + \beta_1 t + s_t + x_t, \quad t = 1, \ldots, 4296,$$

where y_t is the temperature in month t. The given linear predictor includes an intercept β_0, a linear trend β_1 and a seasonal effect s_t of periodicity $q = 12$ which captures monthly variations. This seasonal effect is modelled as an intrinsic GMRF of rank $n-q+1$, having precision parameter τ_s (Rue and Held 2005, p. 122) and scaled to have a generalized variance equal to 1 (Sørbye and Rue 2014). The term x_t denotes the approximate fGn model with $m = 4$, having precision parameter τ_x.

The parameters β_0 and β_1 are assigned vague Gaussian priors, $\beta_i \sim N(0, 10^3)$, while we use penalised complexity priors (PC priors) (Simpson et al. 2017) for all hyperpa-

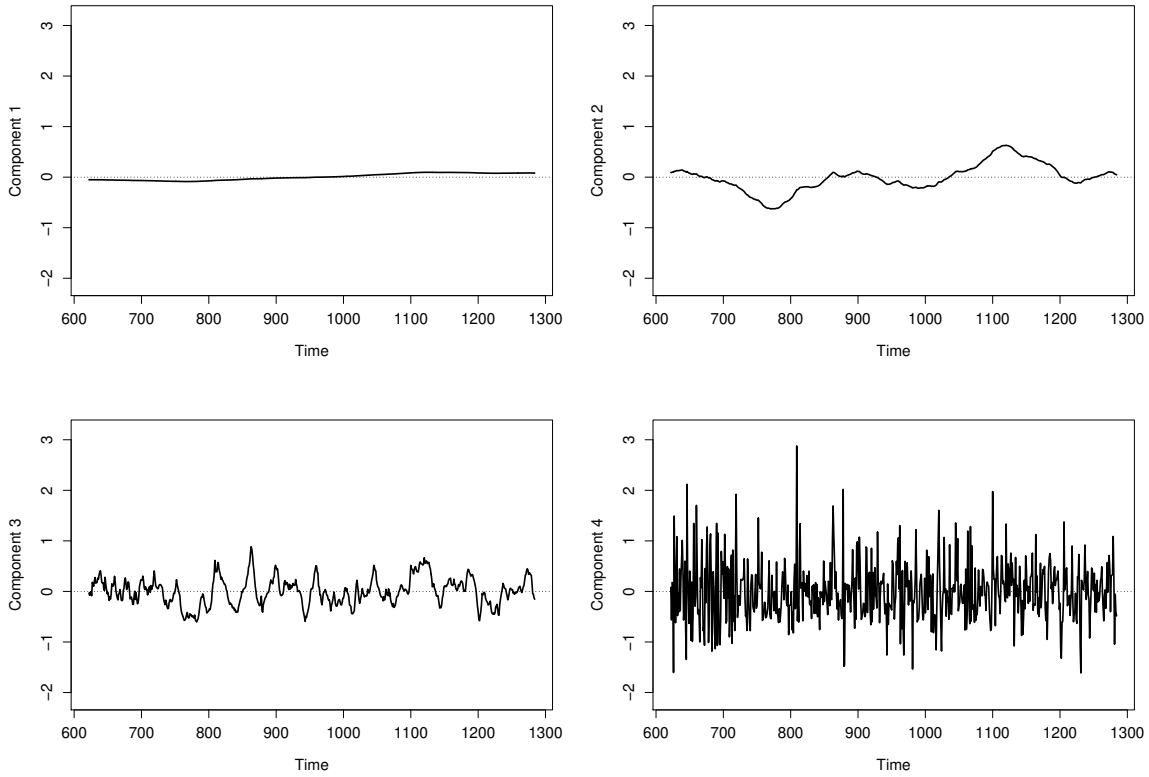

Fig. 9 The estimated weighted AR(1) components for mean-centered annual minimum water levels of the Nile river, using the approximate fGn model with $m = 4$.

rameters. This implies a type II Gumbel distribution for the precision parameters τ_s and τ_x, scaled using the probability statement $P(\tau^{-1/2} > 1) = 0.01$. The PC prior for H (Sørbye and Rue 2017) is scaled by assuming the tail probability $P(H > 0.9) = 0.1$. The resulting analysis has proven to be robust to prior choices. Among others, this has been investigated by using Gamma$(1, 5 \cdot 10^{-5})$ priors for the precision parameters (default in INLA) and the common choice of using a uniform prior for H (Benmehdi et al. 2011). Still, we do prefer to use PC priors as these represent a principle-based choice of priors (Simpson et al. 2017) which also facilitates interpretation of hyperprior parameters (Sørbye et al. 2018).

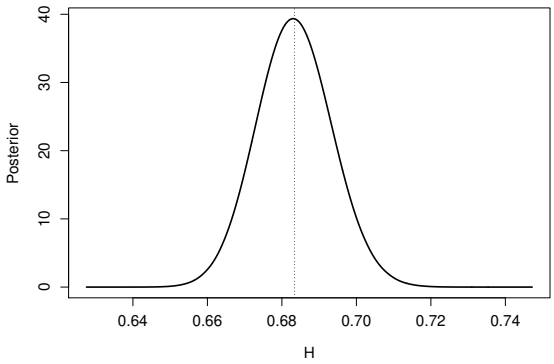

Fig. 10 Analysing the HadCET series: The marginal posterior $\pi(H \mid \boldsymbol{y})$

Analysis of the given model using an exact fGn term is infeasible in terms of computational cost and memory usage. A MacBook Pro with 16GB of RAM crashes due to memory shortage when analysing exact fGn processes of length $n > 2500$. Using the approximate fGn term with $m = 4$, the full Bayesian analysis takes about 14 seconds. The inference gives a significantly positive trend with posterior mean $\hat{\beta}_1 = 2.4 \cdot 10^{-4}$ with 95% credible interval $(1.2 \cdot 10^{-4}, 3.7 \cdot 10^{-4})$. This corresponds to an increased temperature of approximately 0.29 ± 0.15 °C per century. This is in correspondence with the result in Gil-Alana (2003) who estimated an increase of about 0.23 °C per century using the same dataset up the year of 2001.

The marginal posterior for H is illustrated in Figure 10. The posterior mean is $\hat{H} = 0.683$ with a 95% equi-tailed credible interval equal to $(0.664, 0.704)$. We have also fitted the model where x_t is the approximated ARFIMA$(0,d,0)$ model giving a posterior mean estimate of d equal to 0.229 with 95% credible interval equal to $(0.205, 0.253)$. These results are quite similar to Graves et al. (2015) who fitted the ARFIMA$(0,d,0)$ model to a deseasonalized version of the temperatures, giving a posterior mean of $d = 0.209$ with a 95% credible interval equal to $(0.186, 0.235)$. However, in fitting the model to the whole series they assume mean-stationarity. In analysing the series for four different time periods, Graves et al. (2015) report a significantly higher value of the long memory parameter ($d = 0.277$) for the first time period (1659 – 1744) compared to the other time periods. We also observed this, both in fitting the fGn process ($H = 0.740$) and the ARFIMA(0,d,0) model ($d = 0.288$). The higher-value of the long-memory parameter in this case might be explained by a lack of resolution for the first time period (Graves et al. 2015).

6 Concluding remarks

In this paper, we obtain a remarkably accurate approximation of fGn using a weighted sum of only four AR(1) components. The resulting approximate fGn model has a small loss of accuracy for the whole long memory range of H. The key idea to obtain this is to ensure that the approximate model captures the most essential part of the autocorrelation structure of the exact fGn model. This is achieved by appropriate weighting, matching the autocorrelation structure up to a specified maximum lag. The same idea can be used to approximate other models, like ARFIMA$(0,d,0)$.

By construction, the autocorrelation function of the approximate model has an exponential decay for lags larger than the specified maximum lag. This implies that the approximate model does not satisfy formal definitions of long memory processes. However, this trade-off is needed to make analysis of realistically complex models computationally feasible. The great benefit of the resulting approximation is that it has a GMRF structure. This is crucial, especially as

computations can be performed equally efficient in unconditional and conditional scenarios.

An approximate model can never reflect the properties of the exact model perfectly, but neither does a theoretical model in explaining an observed data set. In theory, the fGn model corresponds to an aggregation of an infinite number of AR(1) components which indicates that the model is difficult to interpret in practice. The given decomposition of just a few AR(1) terms might provide a more realistic model. As an example, we have provided a decomposition of the Nile river data, which reflects fluctuations and cycles for different temporal scales. Such a decomposition could also be valuable in analysing other time series. For example, long memory in temperature series has been related to an aggregation of a few simple underlying geophysical processes (Fredriksen and Rypdal 2017).

Implementation of the approximate fGn model in `R-INLA` provides an easy-to-use tool to analyse models with fGn structure. As demonstrated in the temperature example, we can easily combine the fGn model component with other terms in an additive linear predictor, for example nonlinear effects of covariates, random error terms and and random effects including both temporally and/or spatially structured components. In Myrvoll-Nilsen et al. (2018), the mean function of the approximate fGn model is modified to include climate forcing and additional hyperparameters to provide realistic models for temperature series. Among others, this makes it possible to estimate equilibrium climate sensitivity in a computationally efficient way (Rypdal et al. 2018). Further modifications include extensions to spatio-temporal analysis. Especially, we do see a potential in incorporating fGn model components in the analysis of spatial time series combining the given approximate model with the methodology in Lindgren et al. (2011).

References

Baillie RT (1996) Long memory processes and fractional integration in econometrics. J Econom 73:5–59

Benmehdi S, Makarava N, Menhamidouche N, Holschneider M (2011) Bayesian estimation of the self-similarity exponent of the Nile River fluctuation. Nonlinear Process Geophys 18:441–446

Beran J (1994) Statistics for Long-Memory Processes, 1st edn. Chapman & Hall/CRC, New York

Beran J, Terrin N (1996) Testing for a change of the long-memory parameter. Biometrika 83:627–638

Beran J, Schützner M, Ghosh S (2010) From short to long memory: Aggregation and estimation. Comput Stat Data Anal 54:2432–2442

Beran J, Feng Y, Ghosh S, Kulik R (2013) Long-Memory Processes - Probabilistic Properties and Statistical Methods. Springer, Heidelberg

Box GEP, Jenkins GM (1980) Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco

Chan NH, Palma W (2006) Estimation of long-memory time series models: A survey of different likelihood-based methods. Adv Econom 20:89–121

Cont R (2005) Long range dependence in financial markets. In: Fractals in Engineering, Springer, London, pp 159–180

Doukhan P, Oppenheim G, Taqqu M (2003) Theory and Applications of Long-Range Dependence. Birkhauser Boston, c/o Springer-Verlag, New York Inc.

Durbin J (1960) The fitting of time-series models. Rev Inst Int Stat 28:233–244

Eltahir EAB (1996) El Niño and the natural variability in the flow of the Nile River. Water Resour Res 32:131–137

Franzke C (2012) Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. J Clim 25:4172–4182

Fredriksen HB, Rypdal M (2017) Long-range persistence in global surface temperatures explained by linear multibox energy balance models. J Clim 30:7157–7168

Gil-Alana LA (2003) An application of fractional integration to a long temperature series. International Journal of Climatology 23:1699–1710

Golub G, Loan CV (1996) Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore

Granger CWJ (1980) Long memory relationships and the aggregation of dynamic models. J Econom 14:227–238

Granger CWJ, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. J Time Ser Anal 1:15–29

Graves T, Gramacy RB, Franzke CLE, Watkins NW (2015) Efficient Bayesian inference for natural time series using ARFIMA processes. Nonlin Processes Geophys 22:679–700

Graves T, Gramacy R, Watkins N, Franzke C (2017) A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951 – 1980. Entropy 19:1–21

Haldrup N, Valdés JEV (2017) Long memory, fractional integration and cross-sectional aggregation. J Econom 199:1–11

Hosking JRM (1981) Fractional differencing. Biometrika 68:165–176

Hosking JRM (1984) Modeling persistence in hydrological time series using fractional differencing. Water Resour Res 20:1898–1908

Hurst HE (1951) Long-term storage capacities of reservoirs. T Am Soc Civ Eng 116:770–799

Jensen AN, Nielsen MØ (2014) A fast fractional difference algorithm. J Time Ser Anal 35:428–436

Knorr-Held L, Rue H (2002) On block updating in Markov random field models for disease mapping. Scand J Stat 29:597–614

Kondrashov D, Feliks Y, Ghil M (2005) Oscillatory modes of the extended Nile River records (a.d. 622 - 1922). Geophy Res Lett 32:1–4

Koutsoyiannis D (2002) The Hurst phenomenon and fractional Gaussian noise made easy. Hydrolog Sci J 47:573–595

Levinson N (1947) The Wiener (root mean square) error criterion in filter design and prediction. J Math Phys 25:261–278

Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion). J Royal Stat Soc B 73:423–498

Mandelbrot BB, Van Ness JW (1968) Fractional Brownian motions, fractional noises and applications. SIAM Rev 10:422–437

Mandelbrot BB, Wallis JR (1969a) Computer experiments with fractional gaussian noises. Water Resour Res 5:228–267

Mandelbrot BB, Wallis JR (1969b) Global dependence in geophysical records. Water Resour Res 5:321–340

Manley G (1953) The mean temperature of central England, 1698 to 1952. Q J Royal Meteorol Soc 79:242–261

Manley G (1974) Central England temperatures: Monthly means 1659 to 1973. Q J Royal Meteorol Soc 100:389–405

Maxim V, Sendur L, Fadili J, Suckling J, Gould R, Howard R, Bullmore E (2005) Fractional Gaussian noise, functional MRI and Alzheimer's disease. NeuroImage 25:141–158

McLeod AI, Hipel KW (1978) Preservation of the rescaled adjusted range: 1. A reassessment of the Hurst phenomenon. Water Resour Res 14:491–508

McLeod AI, Yu H, Krougly ZL (2007) Algorithms for linear time series analysis: With R Package. J Stat Softw 23:1–26

Molz FJ, Liu HH, Szulga J (1997) Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions. Water Resour Res 33:2273–2286

Myrvoll-Nilsen E, Sørbye SH, Rypdal M, Fredriksen HB, Rue H (2018) Method for estimating climate response under scaling assumption. Preprint

Palma W, Chan NH (1997) Estimation and forecasting of long-memory processes with missing values. J Forecast 16:395–410

Parker DE, Horton EB (2005) Uncertainties in the central England temperature series since 1878 and some changes to the maximum and minimum series. Int J Climatol 25:1173–1188

Parker DE, Legg TP, Folland CK (1992) A new daily central England temperature series, 1772-1991. Int J Climatol 12:317–342

Purczyński J, Wlodarski P (2006) On fast generation of fractional Gaussian noise. Comput Stat Data Anal 50:2537–2551

Rue H (2001) Fast sampling of Gaussian Markov random fields. J Royal Stat Soc B 63:325–338

Rue H, Held L (2005) Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC, London

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J Royal Stat Soc B 71:319–392

Rypdal M, Rypdal K (2014) Long-memory effects in linear response models of Earth's temperature and implications for future global warming. J Clim 27:5240–5258

Rypdal M, Fredriksen HB, Myrvoll-Nilsen E, Rypdal K, Sørbye SH (2018) Emergent scale invariance and climate sensitivity. Under revision in Climate DOI 10.20944/preprints201810.0542.v1

Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH (2017) Penalising model component complexity: A principled, practical approach to constructing priors. Stat Sci 232:1–28

Sørbye SH, Rue H (2014) Scaling intrinsic Gaussian Markov random field priors in spatial modelling. Spat Stat 8:39–51

Sørbye SH, Rue H (2017) Fractional Gaussian noise: Prior specification and model comparison. Environmetrics DOI 10.1002/env.2457

Sørbye SH, Illian JB, Simpson DP, Burslem D, Rue H (2018) Careful prior specification avoids incautious inference for log-gaussian cox point processes. J Royal Stat Soc C DOI 10.1111/rssc.12321

Taqqu MS, Teverovsky V, Willinger W (1995) Estimators for long-range dependence: An empirical study. Fractals 3:785–798

Toussoun O (1925) M'emoire sur l'histoire du Nil. In: M'emoires a l'Institut d'Egypte, vol 18, pp 366–404

Trench WF (1964) An algorithm for the inversion of finite Toeplitz matrices. J Soc Ind Appl Math 12:515–522

Veitch D, Gorst-Rasmussen A, Gefferth A (2013) Why FARIMA models are brittle. Fractals 21, DOI 10.1142/S0218348X13500126

Willinger W, Paxson V, Taqqu MS (1996) Self-similarity and heavy tails: Structural modeling of network traffic. In: A Practical Guide to Heavy Tails: Statistical Techniques and Applications, Birkhauser, Boston, pp 27–53