A Bayesian Monte Carlo Markov Chain Method for Parameter Estimation of Fractional Differenced Gaussian Processes

G. Olivares, F.N. Teferle

May 8, 2013

Abstract

We present a Bayesian Monte Carlo Markov Chain method to simultaneously estimate the spectral index and power amplitude of a fractional differenced Gaussian process at low frequency, in presence of white noise, and a linear trend and periodic signals. This method provides a sample of the likelihood function and thereby, using Monte Carlo integration, all parameters and their uncertainties are estimated simultaneously. We test this method with simulated and real Global Positioning System height time series and propose it as an alternative to optimization methods currently in use. Furthermore, without any mathematical proof, the results from the simulations suggest that this method is unaffected by the stationary regime and hence, can be used to check whether or not a time series is stationary.

1 Introduction

Long-range dependence (LRD) processes are present in many different research fields as, for instance, biology [1], geophysics [2], hydrology and economics [3]. This ubiquity enhances the interest in parameter estimation methods for time series with temporal correlation at long scale.

There are different theoretical approaches to characterize LRD processes. If continuous time evolution is considered, fractional Brownian motion is used [4]. Its discrete counterparts are fractional Gaussian noise (fGn) [4] and fractionally differenced Gaussian noise (fdGn) [5].

All those processes can be characterized, at low frequencies, i.e. $\omega \to 0$, by the power law spectrum

$$S(\omega) = S_0 \left(\frac{\omega_0}{\omega} \right)^{\alpha} , \tag{1}$$

*The authors are with the Geophysics Laboratory, Université du Luxembourg, Luxembourg.

2

where ω is the angular frequency, ω_0 is any given reference angular frequency, S_0 the power amplitude at $\omega = \omega_0$, and $\alpha \in (\mathbf{R}^+ + \{0\})$ the spectral index. Equation (1) includes some very well known cases [2], e.g.: white noise ($\alpha = 0$), Flicker noise ($\alpha = 1$) and Random Walk noise ($\alpha = 2$).

Previously published parameter estimation methods are based either on empirical tools such as, for instance, the rescaled adjusted range or R/S-statistic [6], aggregated variance method [3], periodogram method [7], Allan variance [8, 9], or optimization methods like Maximum Likelihood Estimation (MLE) [10, 11] and Least Squares (LS) [12].

So far, MLE has been widely used, although, without a closed-form expression for the likelihood function and a requirement for assuming an a priori power amplitude [10], or an a priori spectral index [11, 12]. Moreover, without a closed-form expression, some of these methods do not provide an uncertainty associated with the spectral index estimate [13, 14, 15]. [16]-[22] shown that MLE yields asymptotically normally distributed estimates for the parameters and, based on that, confidence intervals can be derived. Nevertheless, our interest is to find a suitable method to analyze fdGn processes whose spectral index spans from stationary to nonstationary regime values, i.e. for $\alpha > 0$. As there are no theoretical proofs concerning the asymptotic behaviour of MLE for nonstationary time series, the confidence intervals are computed numerically, by means of a grid-like exploration of the parameter space. This increases the computational time of MLE geometrically with the number of parameters, whereas for the Bayesian Monte Carlo Markov Chain (MCMC) method it does linearly, suggesting it to be a more suitable method for dealing with higher numbers of parameters. [23] and [24] noticed that the properties of a particular estimator may change according to the value of the spectral index α. Without any mathematical proof, the results from the simulations below suggest that the MCMC method performs well, regardless of the value of the spectral index, i.e. it can also estimate whether the time series is stationary or not. Another difference is that in case of non-standard distribution functions of the parameters, the optimization method can get stuck in a local maximum, whereas the MCMC method can deal with multimodallity as it explores the neighbourhood of the maxima of the likelihood.

Other problems with the above published methods arise from pre-analysis detrending (if need be) as this removes significant amounts of low-frequency energy from the original signal, thereby yielding underestimated spectral indices [25, 26].

Considering the impact the stochastic nature of the noise has on parameter estimation [27, 28], we used the MCMC method for LRD processes that, unlike current optimization methods, simultaneously estimates all parameters (including the spectral index) and their uncertainties. Moreover, MCMC can detect non-Gaussianity on parameter distributions and any correlation between them. We also noticed that the simulations systematically give underestimated uncertainties for all parameters if an a priori value for the spectral index is assumed.

This paper is organized as follows: section II includes the fdGn process, section III presents a brief introduction to the MCMC method with a description of our algorithm. Then in section IV some results are given and discussed from simulated data and real data from height time series obtained from Global Positioning System (GPS) measurements. Finally, section V contains the concluding remarks.

2 Fractionally Differenced Gaussian Process

According to [5], fdGn is defined as the fractional difference of discrete-time Gaussian noise, namely

$$r(k) = (1 - B)^{-\alpha/2} w(k) \,. \tag{2}$$

where $w(k) \in \mathcal{N}(0, \sigma)$, α is the spectral index and B is the difference operator $B\, r(k) = r(k-1)$. By expressing the difference operator power as a power series on B, (2) can also be conceived as linear convolution of white noise or as a moving average (MA) process of infinite order through the equation

$$r(k) = \sum_{i=0}^{\infty} h(i)\, w(k - i) \,, \tag{3}$$

with

$$h(i) = \frac{(i + \alpha/2 - 1)!}{i!(\alpha/2 - 1)!} \,. \tag{4}$$

The process is completely characterized by just two parameters, the spectral index α, and the power amplitude σ of the convoluted white noise process. Indeed, the spectral density $S(\omega)$ of $r(k)$ can be easily obtained from (2) as

$$S(\omega) = \frac{2^{-\alpha}\sigma^2}{\left(\sin\left(\frac{\omega}{2}\right)\right)^{\alpha}} \,. \tag{5}$$

At low frequencies, when $\omega \to 0$, $\sin(\omega) \sim \omega$, hence (5) yields the power law spectrum, i.e. (1)

$$S(\omega) \sim \frac{\sigma^2}{\omega^{\alpha}} \,. \tag{6}$$

For calculus purposes (3) is re-written in matrix form

$$\mathbf{r} = L\mathbf{w} \,, \tag{7}$$

where $\mathbf{r} \in \mathbf{R}^N$ is a fdGn process, $\mathbf{w} \in \mathbf{R}^N$ a white noise process, N is the number of data (or time series length) and L, the convolution matrix filter, is a lower triangular Toeplitz matrix (a matrix made of constant diagonals) with coefficients

$$L_{ij} = \begin{cases} h_{i-j} \equiv h(i - j) & \forall (i - j) \geq 0 \\ 0 & \forall (i - j) < 0 \end{cases} \tag{8}$$

Thus the covariance matrix of fdGn is $\mathbf{C} = <\mathbf{r}\,\mathbf{r}^T> = \sigma^2 L\, L^T$, with matrix coefficients

4

$$\mathbf{C}_{lk} = \sigma^2 \sum_{i=0}^{l} h_i h_{i+|k-l|} \,, \tag{9}$$

where $l, k \in [0, N-1] \subset \mathbf{N}$. As the signal length N increases, (9) becomes a Toeplitz matrix, i.e. it only would depend on the lag between data, but for any sample of the process it is not. Some MLE methods exploit this Toeplitz-like symmetry and thereby increasing the computational speed [10, 29]. Nevertheless, by doing so it may yield biased estimates of the parameters.

The first row of the covariance matrix is the correlation function for every lag of the process. Fig. 1 shows the dependency of that correlation with the spectral index for three different signals normalized at the origin. It can be seen that the higher the spectral index α is, the slower the correlation decays.

3 The MCMC method

The MCMC method is the use of simulations made with Markov chains to get a good sample of the posterior distribution of the parameter given the data. Then by means of Monte Carlo integrals on that sample, estimates for the parameters and their uncertainties are obtained.

The theory of Markov chains is well developed and further information can be found in e.g. [30]. Much of the elaboration of the theory can be avoided for our purposes, since we are not interested in discovering the properties of some arbitrary Markov chain, but rather in constructing a Markov chain with the properties described in subsection 3.2.

3.1 MCMC background

With the distribution function for each parameter, given the data, we could estimate their expected values. Nevertheless, to do so, repeated observational data would be needed at every epoch. This is usually not the case with natural phenomena in geophysics or biology where the events represented by time series can not be repeated at will. Even in engineering, the conditions under which the signal is produced may change in time. Consequently, sometimes the aforementioned distribution (a.k.a. the posterior distribution) function can not be obtained right away from the data.

On the other hand, a sample of the conditional distribution function for the data can be obtained with a priori distributions of the parameters.

According to the Bayes theorem [31] the above distributions are related as

$$\mathcal{P}(\theta|\mathbf{y}) = \frac{\mathcal{L}(\mathbf{y}|\theta)\mathcal{P}(\theta)}{\mathcal{P}(\mathbf{y})} \,, \tag{10}$$

where $\mathcal{P}(\theta|\mathbf{y})$ is the posterior distribution, $\mathbf{y}$ is the observational data, i.e. the process, θ is the vector that contains the parameters, $\mathcal{L}(\mathbf{y}|\theta)$ is the conditional probability function for the data or likelihood, $\mathcal{P}(\theta)$ is the prior probability distribution for the parameters and $\mathcal{P}(\mathbf{y})$ is the probability function for the data.

Once a sample of the posterior distribution is obtained, the expected values of the parameters of the model can be estimated by means of Monte Carlo integration. More precisely,

$$< \theta_i >= \int \mathcal{P}(\theta|\mathbf{y})\theta_i d\theta = \frac{\sum_{t=1}^{M} \theta_{t,i}}{M} \ , \tag{11}$$

where M is the number of points in the Markov chain and $\theta_{t,i}$ denotes the value of the ith component of the parameter vector θ at the tth step of the chain. The $100(1 - 2p)\%$ confidence interval $[c_p, c_{1-p}]$ for a parameter is estimated by setting c_p to the pth quantile of $\theta_{t,i}$, $t = 1, .., N$ and c_{1-p} to the $(1-p)^{th}$ quantile. Moreover, (11) yields an unbiased estimator for θ_i and its accuracy does not depend on the number of parameters [32].

3.2   The MCMC algorithm

In this section we introduce the MCMC method used to estimate the time series parameters. Firstly, we consider a homogeneous distribution $\mathcal{P}(\theta)$ for all the parameters within an a priori interval $(\theta_{min}, \theta_{max})$. The algorithm, the so-called Metropolis-Hasting method [30], that leads the Markov chain through the posterior distribution surface is summarized below:

1. Compute the likelihood at θ_i, i.e. $\mathcal{L}_i \equiv \mathcal{L}(\mathbf{y}|\theta_i)$.

2. Take a random step in parameter space to obtain a new value for the parameter θ_{i+1}. The probability distribution of the step is taken to be Gaussian in each direction i with variance σ_i, i.e. the step size.

3. Compute $\mathcal{L}_{i+1}$.

4. If $\mathcal{L}_{i+1}/\mathcal{L}_i > 1$, take the step, i.e. save the new set of parameters θ_{i+1} and $\mathcal{L}_{i+1}$ as part of the chain, then go to step 2 after the substitution $\theta_i \to \theta_{i+1}$.

5. If $\mathcal{L}_{i+1}/\mathcal{L}_i < 1$, draw a random number x from a uniform distribution from 0 to 1. If $x > \mathcal{L}_{i+1}/\mathcal{L}_i$ do not take the step, i.e. keep the previous parameter value θ_i as part of the chain and return to step 2. If $x < \mathcal{L}_{i+1}/\mathcal{L}_i$ take the step, i.e. do as in 4.

Should the algorithm follow all steps from 1 to 4, it would find the maximum of the likelihood. Nevertheless, in order to get confidence levels for the estimates, the algorithm is set up to explore the surroundings of the maximum, hence the fifth step.
As $\mathcal{P}(\theta)$ is the same for all the values of θ, and $\mathcal{P}(\mathbf{y})$ is constant along the MCMC process, the ratio $\mathcal{L}_{i+1}/\mathcal{L}_i$, by the Bayes theorem (10), is equivalent to $\mathcal{P}(\theta_{i+1}|\mathbf{y})/\mathcal{P}(\theta_i|\mathbf{y})$, therefore the sample obtained for the likelihood with the algorithm above is the same for the posterior distribution. Also notice that with $\mathcal{P}(\mathbf{y})$ being constant during the whole process it is irrelevant for (11) and it can be ignored.

3.3 The Sample of the Likelihood

We have estimated the parameters for three different processes:

1. **Case I**. Pure fdGn:

$$\mathbf{y} = L\mathbf{w} \ , \mathbf{w} \in \mathcal{N}(0, \sigma_{pl}) \tag{12}$$

2. **Case II**. Pure fdGn with white noise:

$$\mathbf{y} = L\mathbf{w} + \mathbf{u} \ , \mathbf{u} \in \mathcal{N}(0, \sigma_{wn}) \tag{13}$$

3. **Case III**. Pure fdGn with white noise in combination with linear and periodic signals:

$$\mathbf{y} = y_0 + v_0\,\mathbf{t} + A_c\cos(\omega_0\mathbf{t}) + A_s\sin(\omega_0\mathbf{t}) + L\mathbf{w} + \mathbf{u}\,, \tag{14}$$

where y_0 is the intercept, v_0 the slope, $\mathbf{t}$ the epoch, A_c and A_s the cosine and sine amplitudes, respectively, and ω_0 is the angular frequency of the periodic signal. Though the above algorithm is not necessarily restricted to one specific distribution (according to [33] it is one case among many M-estimators), the Gaussian joint likelihood was chosen for all three processes:

$$\mathcal{L}(\mathbf{y}|\hat{\theta}) = \frac{1}{(2\pi)^{N/2}\,|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}(\mathbf{y}-\hat{\mathbf{y}})^T\,\mathbf{C}^{-1}(\mathbf{y}-\hat{\mathbf{y}})}\,, \tag{15}$$

where N is the number of data, $\mathbf{y}$ the data, $\hat{\theta} \equiv (\hat{\alpha}, \hat{\sigma}_{pl})$, $(\hat{\alpha}, \hat{\sigma}_{pl}, \hat{\sigma}_{wn})$, $(\hat{\alpha}, \hat{\sigma}_{pl}, \hat{\sigma}_{wn}, \hat{v}_0, \hat{y}_0, \hat{A}_c, \hat{A}_s)$ the estimated parameters for cases I, II and III, respectively; $\hat{\mathbf{y}} \equiv \mathbf{y}(\hat{\theta})$ the estimated data, $|\mathbf{C}|$ the determinant of the covariance matrix, and the covariance matrix

$$\mathbf{C} = \sigma_{pl}^2 L\,L^T + \sigma_{wn}^2\mathbf{I}\,, \tag{16}$$

with $\mathbf{I}$ being the identity matrix.

The drawback of this method is that for every MCMC point, i.e every computation of $\mathcal{L}_i$, the covariance matrix has to be computed. Clearly, this increases the computational time, but in doing so, unlike MLE methods, it provides the covariance for the estimate of the spectral index, can detect local maxima and does not underestimate the uncertainties of the parameters [27, 11].

It is noted here that case III follows the general strategy in the modelling of geodetic position time series using both empirical and optimization methods, therefore we also show the performance of the MCMC method for one representative GPS height time series.

3.4 Convergence criteria

The algorithm used for this MCMC method does not provide any stop-condition by itself, consequently we must monitor the Markov chains in order to know whether they might have already reached a stage at which any statistical information obtained from them can not be significantly improved anymore. According to [34], two aspects characterize Markov chains in order to get a good sample of the posterior distribution function: "convergence" and "good mixing".

Convergence is achieved once the algorithm spots the maximum (or maxima) and gets samples around it. Good mixing refers to the idea of getting a homogeneous sample around the maximum in order to not have artificially biased estimators.

In this study we have used the spectral analysis method for the Markov chain presented in [35]. At small scale, the MCMC is mainly a random-walk process, therefore the MCMC points are correlated and any estimate for the parameters at that scale will be biased. In order to get a non-biased estimate we must allow the MCMC to reach the convergence state and a good mixing. By the ergodic theorem [30], the MCMC at long-scale provides a homogeneous sample of the distribution function, i.e. when the MCMC length is long enough it yields a white-noise-like spectrum. This spectrum can be written as

$$P(k) = P_0 \frac{(k^*/k)^\beta}{(k^*/k)^\beta + 1} \, , \tag{17}$$

where $\beta > 0$ is the spectral index of the spectrum of the Markov chain (it has nothing to do with the spectral index α of the time series itself), $k = j(2\pi/M)$ is the scale (with $j \in \mathbf{N}$), M the MCMC length, k^* is the cross-over scale, i.e. the inverse of the length for which two points of the Markov chain that distance apart (at least) are uncorrelated, and $P_0 = P(k \to 0)$.

According to [35], the Markov chain has converged and has a good mixing when the next two requirements are satisfied:

1. $k_{min} \equiv 1/M$ must be in the white noise regime, i.e. $k_{min} < k^*$, as it guarantees that the chain is long enough.

2. The convergence criteria is fulfilled when $r = P_0/M < 0.01$.

In section 4 the above parameters are given in Table 9 for each case.

Finally, as it takes some time for the chain to achieve the stationary state which guarantees the ergodic theorem to hold [30] (and under which the "convergence" and "good mixing" criteria can be used), we have discarded $\sim 33\%$ of the first points in all Markov chains.

4 Results

In order to show its performance, the MCMC method is used to estimate the parameters of, firstly, synthetic pure fdGn processes, secondly, synthetic fdGn with white noise and, thirdly, additionally including a linear trend and periodic signals. Finally, we analyze a sample geodetic time series consisting of daily height estimates obtained from GPS measurements.

4.1 Synthetic Data

For the synthetic data **r** we used (3) and (4) with length $N = 500$. Table 1 shows the priors on the parameters (in arbitrary units). The a priori intervals are chosen wide enough to show the robustness in the performance of the algorithm for starting points of the Markov chain very far away from the maximum of the likelihood. The true values of the parameters of the synthetic data are shown in the last column of tables 2 and 3 (in arbitrary units) and $\omega = 20\pi$.

Table 2 shows the estimates and their uncertainties for pure fdGn (cases Ia and Ib) and fdGn with white noise (case II). The second column (case Ib) shows the results for a Markov chain that started to explore at $\alpha = 6.5$ and $\sigma_{pl} = 6.5$ ($\alpha = 0.5$ and $\sigma_{pl} = 0.5$ for case Ia), to show the robustness of the process even when the starting points are far away from the true values (fourth column). Both cases yield estimates that, at their respective 1σ confidence levels, include the true values.

For the second case the uncertainties are bigger than for cases Ia and Ib because by adding a new parameter, the parameter space to be explored by the MCMC gets wider.

Cases $IIIa$ and $IIIb$ additionally introduce a linear trend and periodic signals thereby increasing the number of parameters from two (case I) up to seven. An alternative to MCMC would be to compute the likelihood over a grid, but as the number of parameters increases the computational time becomes prohibitive. Therefore, the MCMC method is more suitable for dealing with a high number of parameters.

Table 3 presents the results for these cases with an a priori spectral index equal to the real value $\alpha = 1.1$ (case $IIIa$), and without any a priori value on any parameter (case $IIIb$) to investigate the underestimation, when some a priori values are assumed on some parameters [36, 37, 11].

A number of effects are noticeable after adding trends and periodic signals. Firstly, the periodic signal introduces extra temporal-correlation, hence the higher value for $\hat{\alpha}$ for case $IIIb$. Another common effect is that the 1σ confidence level gets wider for all three estimates (α, σ_{pl}, σ_{wn}), due to the fact that with more parameters, the MCMC algorithm has to explore a bigger parameter space.

However, the main difference between the results for cases $IIIa$ and $IIIb$ is the bigger uncertainty for all estimated parameters in case $IIIb$. These observed differences are consistent with the fact that having less parameters to estimate leads to smaller uncertainties, as the parameter space to be explored is narrower. According to the ergodic theorem, all the Markov chains converge towards the same final state (for stationary final states) [30]. Therefore, to have several different starting points (or several Markov chains) is equivalent to a longer Markov Chain. Nevertheless, more examples are provided in order to give more statistical significance to our results. Namely, four different spectral index values were analyzed $\alpha = (0.50, 1.17, 1.83, 2.5)$ for all three cases, and for each of these values we run ten different Markov chains, with their initial points randomly chosen in the interval $\alpha \in [0.5, 2.5]$. Table 4 summarizes the results for the estimates of the spectral index only, though we have estimated all parameters. All real values (header row) are within the 1σ confidence level of their respective estimates, except for $\alpha = 1.17$ in case I, and $\alpha = 0.5$ in case III, for these two cases the real values are within their respective 2σ confidence level. Sys-

9

tematically the uncertainties get bigger as the number of parameters increases (two, three and seven parameters for cases I, II and III, respectively). Another important result infered from Table 4 is that the MCMC method for stationary ($\alpha = 0.5$ in the first column) and nonstationary time series ($\alpha = 1.17, 1.83, 2.5$) performs alike. These results suggest that, unlike other estimators (Gaussian semiparametric and log-periodogram estimators in [23] and [24], respectively), the MCMC method seems unaffected by non-stationarity, hence being a potential tool to detect nonstationarity in time series.

With further detail, Tables 5, 6 and 7 show the estimates of the spectral index for each Markov chain, i.e. for ten different starting points. Table 5 (case I) shows how the estimates of each chain systematically contain the real values within their 1σ confidence level, except for the second column, whose real value $\alpha = 1.17$ is slightly beyond this confidence level (the smallest value at 1σ is $\hat{\alpha} = 1.18$). In Table 6 (case II) all chains yield results that are either systematically above the real values (first, second and fourth columns) or below (third column), and the real values are within the 1σ confidence level. All the variances are bigger than those from case I, which is consistent with a wider parameter space as a consequence of introducing a third parameter (σ_{wn}). Finally, Table 7 summarizes the results for each Markov chain for case III. The results are quite homogeneous and all estimates are systematically below (second column) or above the real values (first, third and fourth columns). All real values are within the 1σ confidence level, except for the first column ($\alpha = 0.5$). Our interpretation is that the temporal-correlation introduced by the periodic signals might increase the value of the estimated spectral index. However, this is not confirmed by the other columns. Another source of this bias might be that for the algorithm it is more difficult to tell the fdGn process from the white noise as the spectral index decreases and the parameter space gets wider due to the new parameters in case III, (v_0, y_0, A_c, A_s). Each row of Table 4 is in good agreement with the ten different Markov chains shown in Tables 5, 6 and 7, respectively, as it was expected, according to the ergodic theorem, thus allowing us to take either some different starting points or just one long Markov chain.

Although the sample for the posterior distribution is obtained using the Bayes Theorem, the mean of the estimate and their uncertainties are computed in a Frequentist-school fashion with the Monte Carlo integration, i.e. with (11), and quantiles, respectively. Furthermore, we can get information about the estimates by looking at their histograms. Those histograms represent the marginalized likelihoods, i.e. the likelihood integrated over all parameters except for the parameter we want to analyze. For example, in Fig. 2, the histograms for all parameters in case $IIIb$ are shown. These figures show that unlike the model parameters (v_0, y_0, A_c, A_s), the parameters corresponding to the stochastic process ($\alpha, \sigma_{pl}, \sigma_{wn}$) do not follow a Gaussian distribution. Though this might not be important for detrending in processes like case III, for example, geodetic time series, the non-Gaussianity behaviour of those parameters might still affect the estimates in other applications. [28] shown that classical localization methods are less efficient under multipath effects, than methods that estimate the error distribution.

Histograms for marginalized likelihoods also provide information about cross-correlation among parameters. For example, Fig. 3a confirms that α and σ_{pl} are uncorrelated without white noise within the signal (case I). On the other hand, Figs. 3b and 3c show the correlation between α, σ_{pl} and σ_{wn} when white

noise is included (case II).

Figs. 4a and 4b suggest that adding deterministic parameters for modelling such as v_0, y_0, A_c, A_s does not affect the correlations among the stochastic parameters (the above mentioned). Fig. 4c shows how the histogram for $\mathcal{L}(\mathbf{y}|v_0, y_0)$ projected onto the $v_0 - y_0$ plane gives an ellipse that indicates the correlation between them; the higher the estimated slope $\hat{v}_0$ is, the smaller the estimated intercept $\hat{y}_0$ of the straight line with the time series.

According to Fig. 4d, the periodic signal parameters are completely uncorrelated with respect each other.

4.2 Real Data: Geodetic Time Series

Since 1997, it has been shown that noise in geodetic time series can be characterized as long-term time series with spectral index values that span from stationary ($\alpha < 1$) to nonstationary regimes ($\alpha > 1$) [25, 40, 41, 11]. Therefore, it is necessary to develop a parameter estimation method that performs well regardless of the stationary regime. According to [23, 24], log-periodogram and Gaussian semiparametric estimators change its properties at going through $\alpha = 1.5$. Nevertheless, the results from the simulations suggest that the MCMC method could be used to check whether a time series is stationary or not, hence the interest at using it for the analysis of geodetic time series analysis.

In order to check the performance of the MCMC method with real data we analyzed the daily height time series obtained from GPS measurements. Current state-of-the-art global GPS analyses result in a day-to-day scatter of $1 - 2$ mm and 6 mm for the horizontal and vertical position components, respectively. We tested the MCMC method on the daily height estimates as this component is generally noisier than the horizontal components and is of primary interest to many geophysical studies of changes in global sea level and mass balance [38, 39]. The height time series used here can be regarded as a typical case from a high-precision continuous GPS station and were provided by the British Isles continuous GPS Facility[1]. For time efficiency, we selected a station with four years of GPS measurements and as with most GPS time series some data gaps are present ($\sim 18\%$ for that case). Still, gaps are not of great concern for this MCMC method, provided we do not use any Toeplitz-like symmetry for the covariance matrix that could otherwise be broken [37]. Moreover, according to [36], a time series spanning four years is long enough to consider a bias from the periodic signals to be negligible on the linear trend estimates.

Table 8 summarizes the estimates obtained for the daily height time series. The estimated spectral index $\alpha = 0.67 \pm 0.07$ indicates that the time series is stationary at the 1σ confidence level, and the power of the fdGn process $\sigma_{pl} = 12.4 \pm 0.8 \, mm/yr^{0.17}$ [2] is bigger than that of the white noise $\sigma_{wn} = 1.3 \pm 0.8 \, mm$. Concerning the deterministic parameters (v_0, y_0, A_c, A_s), the linear trend $\hat{v}_0 = 2.6 \pm 0.9 \, mm/yr$ is the same order of magnitude as the estimates for the periodic signals $\hat{A}_c = 0.7 \pm 0.5 \, mm$ and $\hat{A}_s = 1.3 \pm 0.3 \, mm$. Finally, with the values of the uncertainties, we can conclude that, at least at 1σ confidence

---

[1]http://www.bigf.ac.uk/

[2]Here the power law amplitude is scaled by $\Delta T^{\alpha/4}$ in order to be sure that any noise source with spectral index α will cross at the same frequency given the same sampling interval ΔT and equal noise amplitude.

level, all estimates have Gaussian distribution.

Histograms for the marginalized likelihoods of every parameter are given in Fig. 5. Figs. 5a and 5b show the non-symmetric distribution for α and σ_{pl} respectively. This is in good agreement with the results for synthetic data (see Figs. 2a and 2b), but it is not possible to assess how much comes from the GPS signal itself and how much from the post-processed height time series. Fig. 5c shows the distribution for the white noise power amplitude σ_{wn}. In this case, the maximum is too close to the axis to either disclaim or claim any resemblance with respect to its synthetic counterpart in Fig. 2c. Figs. 5d, 5e, 5f and 5g respectively show the deterministic parameters v_0, y_0, A_c and A_s have Gaussian-like histograms as it was expected according to the synthetic results previously shown. Concerning correlations between parameters, Figs. 6a and 6b show that α, σ_{pl}, and σ_{wn} are correlated in a similar way as their synthetic counterparts in Figs. 4b and 4c, respectively. Fig. 6c shows how the slope and the intercept are anti-correlated and, finally, Fig. 6d gives the same Gaussian solution as Fig. 4d with the condition that A_c, $A_s > 0$ in order to avoid artificial multimodality.

To summarize, the MCMC method applied to real data from a sample GPS height time series provides similar results to the synthetic data (case III). The individual distributions of the parameters and their correlations are Gaussian for the deterministic parameter, but not for the stochastic ones, though, at 1σ confidence level they can be considered Gaussian as Table 8 shows.

Concerning the convergence and good mixing criteria aforementioned in Section 3.4, Table 9 summarizes the results for cases I, II, III and with real data. There it is shown that for all cases the two conditions are fulfilled, i.e. $r < 0.01$ and $k^* > k_{min}$ in columns two, three and four respectively.

5   Conclusions

We have presented a Metropolis-Hasting-based Bayesian MCMC method that simultaneously estimates all parameters alongside their uncertainties for fdGn processes. The method performs well for estimating the characteristic parameters of this process, namely α and σ_{pl}, even when additioned white noise (see Table 2), linear trend and periodic signals are present (see Table 3).

This method also provides a tool for measuring non-Gaussianity on the parameter distributions (see Figs. 2 and 5), and correlation between the parameters (see Figs. 3, 4 and 6). We have also shown that setting a priori parameter values, as some optimization methods allow [10, 13], yields uncertainties that are too optimistic (see Table 3), whereas MCMC takes into account the effect of estimating the noise as well, thus giving rise to bigger uncertainties.

Previous works have proved that geodetic time series can be characterized as nonstationary LRD processes of unknown spectral index [25, 40, 41, 11]. For that, analytical or numerical methods capable of detecting nonstationarity are needed. From the results shown in Tables 4, 5, 6 and 7, and without any mathematical proof, we propose the Bayesian MCMC method as a tool to detect nonstationarity in time series.

Finally, it must be emphasized that the spectral index α determines the estimation of parameters from geodetic time series and their uncertainties as well [27, 25, 11]. A recent study [28] shown that it also affects the estimation of

real-time vehicles positions from Global Navigation Satellite System (GNSS) signals [28] (a different application of GNSS as presented here). These examples highlight the importance of estimating the spectral index and its uncertainty. For that, and without any closed-form function for the likelihood, should the stochastic noise be estimated, MCMC provides more realistic estimations than optimization methods do.

Acknowledgement

References

[1] B. B. Mandelbrot, "The Fractal Geometry of Nature", W. H. Freeman & Co Ltd, 1982, 978-0716711865.

[2] D. C. Agnew, "The time-domain behaviour of power-law noises", *Geophys. Res. Lett.*, vol. 19, no. 4, pp. 333–336, February, 1992.

[3] J. Beran, "Statistics for Long-Memory Processes", *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, 1994.

[4] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications", *SIAM Review*, vol. 10, no. 4, pp. 422–437, October, 1968.

[5] J. R. M. Hosking, "Fractional Differencing", *Biometrika*, vol. 68, no. 1, pp. 165–176, April, 1981.

[6] B. B. Mandelbrot and J. R. Wallis, "Computer experiments with fractional Gaussian noises", *Water Resour. Res.*, vol. 5, no. 1, pp. 228–267, 1968.

[7] J. Geweke and S. Porter-Hudak, "The estimation and application of long memory times series models", *Journal of Time Series Analysis*, vol. 4, no. 4, pp. 221–238, 1983.

[8] A. Caporali, "Average strain rate in the Italian crust inferred from a permanent GPS network - I. Statistical analysis of the time-series of permanent GPS stations, *Geophysical Journal International*, no. 155, pp. 241–253, 2003.

[9] M. Hackl, R. Malservisi, U. Hugentobler and R. Wonnacott, "Estimation of velocity uncertainties from GPS time series: Examples from the analysis of the South African TrigNet network ", *Journal of Geophysical Research*, vol. 116, no. B11404, doi:10.1029/2010JB008142, November, 2011.

[10] M. Deriche and A. H. Tewfik, "Maximum Likelihood Estimation of the Parameters of Discrete Fractionally Differenced Gaussian Noise Process", *Signal Processing, IEEE Transactions on*, vol. 41, no. 10, pp. 2977–2989, October, 1993.

[11] S. P. D. Williams and Y. Bock and P. Fang and P. Jamason and R. M. Nikolaidis and M. Miller and D. J. Johnson, "Error analysis of continuos GPS position time series", *J. Geophys. Res.*, vol. 109, no. B03, doi: 10.1029/2003JB002741 2004.

[12] A.R. Amiri-Simkoeei, "Noise in multivariate GPS position time-series", *J Geod*, vol. 83, pp. 175–187, 2009.

[13] S. P. D. Williams, "CATS: GPS coordinate time series anaylsis software", *GPS Solutions*, no. 12, pp. 147–153, 2008.

[14] F. P. Craigmile and P. Guttorp, "Space-time modelling of trends in temperature series", *Journal of Time Series Analysis*, vol. 32, no. 4, pp. 378–395, April, 2011.

[15] H. Sheng, Y. Q. Chen and T. Qiu, "On the robustness of Hurst estimators", *IET Signal Processing*, vol. 5, no. 2, pp. 209–225, 2011.

[16] R. Dahlhaus, "Efficient parameter estimation for self-similar processes", *Ann. Statist.*, vol. 17, pp. 1749–1766, 1989.

[17] F.B. Sowell, "Maximum likelihood estimation of stationary univariate fractionally integrated time series models", *J. Econometrics*, vol. 53, pp. 165–188, 1992.

[18] J. Beran, "On a class of M-estimators for long-memory Gaussian models" *Biometrika*, vol. 81, pp. 755–766, 1994.

[19] G.A. Koop, E. Ley, J. Osiewalski, M.F.J. Steel, "Bayesian analysis of long memory and persistence using ARFIMA models", *J. Economet.*, vol. 76, no. 1-2, pp. 149–169, 1997.

[20] B. Liseo, D. Marinucci, L. Petrella, "Bayesian semiparametric inference on long-range dependence", *Biometrika*, vol. 88, no. 4, pp. 1089–1104, 2001.

[21] O. Lieberman, R. Rosemarin, J. Rousseau, "Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes", *Economet. Theo.*, vol. 28, pp. 457–470, 2012.

[22] J. Rousseau, N. Chopin, B. Liseo, "Bayesian nonparametric estimation of the spectral density of a long or intermediate memory Gaussian process", *Ann. Statist.*, vol. 40, no. 2, pp. 964–995, 2012.

[23] C. Velasco, "Gaussian semiparametric estimation of nonstationary time series", *Journal of Time Series Analysis*, vol. 20, pp. 87–127, 1999.

[24] C. Velasco, "Non-stationary log-periodogram regression", *J. Econometrics*, vol. 91, pp. 325–371, 1999.

[25] J. Langbein and H. O. Johnson, "Correlated errors in geodetic time series: Implications for time-dependent deformation", *J. Geophys. Res.*, vol. 102, no. B1, pp. 591–603, January 10, 1997.

[26] R. W. Hamming, Digital Filters, Dover Publications, Inc., 3rd Edition, 1998.

[27] M.S. Bos, L. Bastos and R. M. S. Fernandes, "The influence of seasonal signals on the estimation of the tectonic motion in short continuous GPS time-series", *Journal of Geodynamics*, no. 49, pp. 205–209, 2010.

[28] A. Rabaoui and N. Viandier and E. Duflos and J. Marais and P. Vanheeghe, "Dirichlet Process Mixtures for Density Estimation in Dynamic Nonlinear Modeling: Application to GPS Positioning in Urban Canyons", *Signal Processing, IEEE Transactions on*, vol 60, no. 4, pp. 1638–1655, April, 2012.

[29] D. B. Rigling, "Maximum Likelihood Estimation of Band-Limited Power Law Spectrums", *Signal Processing, IEEE Letters,*, vol. 19, no. 5, pp. 307–310, May, 2012.

[30] W. R. Gilks and S. Richardson and D. Spiegelhalter, "Markov Chain Monte Carlo in Practice", Chapman & Hall, 1996.

[31] D.S. Sivia, J. Skilling, "Data Analysis: A Bayesian Tutorial", Oxford University Press, 2006.

[32] M. Chen and Q. Shao and J. G. Ibrahim, "Monte Carlo Methods for Bayesian Computation", Springer Ed., 2001.

[33] R. A. Maronna, R. D. Martin and V. J. Yohai, "Robust statistics: theory and methods ", Wiley series in probability and statistics, 2006.

[34] M.K. Cowles, B.P. Carlin, "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review", *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 883–904, June, 1996.

[35] J. Dunkley, M. Bucher, P. G. Ferreira and K. Moodley and C. Skordis, "Fast and reliable MCMC for cosmological parameter estimation", *Mon. Not. R. Astron. Soc.*, vol. 356, pp. 925–936, 2005.

[36] G. Blewitt and D. Lavallée, "Effect of annual signals on geodetic velocity", *J. Geophys. Res.*, vol. 107, no. B7, pp. ETG 9-1–ETG 9-11, July, 2002.

[37] M. S. Bos, R. M. S. Fernandes, S. P. D. Williams and L. Bastos, "Fast error analysis of continuous GPS observations", *J Geod*, vol. 82, pp. 157–166, 2008.

[38] T. van Dam and J. Wahr and P. C. D. Milly and A. B. Shmakin and G. Blewitt and D. Lavallee and K. M. Larson, "Crustal displacements due to continental water loading", *Geophys. Res. Lett.*, vol. 28, no. 4, pp. 651–654, February, 2001.

[39] F.N. Teferle, R.M. Bingley, E.J. Orliac, S.D.P. Williams, P.L. Woodworth, D. McLaughlin, T.F. Baker, I. Shennan, G.A. Milne, S.L. Bradley, D.N. Hansen, "Crustal motions in Great Britain: evidence from continuous GPS, absolute gravity and Holocene sea level data", *Geophys J Int* Vol. 178, pp. 23-46, doi:10.1111/j.1365-246X.2009.04185.x, 2009.

[40] J. Zhang, Y. Bock, P. Fang, S.P.D. Williams, J. Genrich, S. Wdowinski and J. Behr, "Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities", *J. Geophys. Res.*, vol. 102, no. B2, pp. 18,035–18,055, 1997.

[41] A. Mao, C.G.A. Harrison and H.D. Timothy, "Noise in GPS coordinate time series", *J. Geophys. Res.*, vol. 104 no. B2, pp. 2797–2816, 1999.

Table 1: Priors on the parameters in terms of the real values for the synthetic data

Parameter	Interval
$\hat{\alpha}$	$\alpha > 0$
$\hat{\sigma}_{pl}$	$[-10^2, 10^2] \times \sigma_{pl}$
$\hat{\sigma}_{wn}$	$[-10^2, 10^2] \times \sigma_{wn}$
$\hat{v}_0$	$[-10^2, 10^2] \times v_0$
$\hat{y}_0$	$[-10^2, 10^2] \times y_0$
$\hat{A}_c$	$[-10^2, 10^2] \times A_c$
$\hat{A}_s$	$[-10^2, 10^2] \times A_s$

Table 2: Estimated Parameters in arbitrary units for cases I and II ($N = 500$)

Parameter	Ia	Ib	II	Real Value
$\hat{\alpha}$	1.07 ± 0.07	1.11 ± 0.06	1.11 ± 0.12	1.10
$\hat{\sigma}_{pl}$	1.01 ± 0.03	1.00 ± 0.03	$0.93^{+0.09}_{-0.10}$	1.00
$\hat{\sigma}_{wn}$	$--$	$--$	0.36 ± 0.18	0.20

Table 3: Estimated Parameters for case $IIIa$ (with an a priori $\alpha = 1.1$) and case $IIIb$ (any a priori parameter value)

Parameter	IIIa	IIIb	Real Value
$\hat{\alpha}$	1.10	1.26 ± 0.10	1.10
$\hat{\sigma}_{pl}$	0.99 ± 0.04	0.95 ± 0.07	1.00
$\hat{\sigma}_{wn}$	0.19 ± 0.13	0.27 ± 0.17	0.20
$\hat{v}_0$	$20.21^{+1.21}_{-1.28}$	$20.30^{+1.97}_{-2.06}$	20.00
$\hat{y}_0$	-0.74 ± 0.70	0.32 ± 0.80	0.00
$\hat{A}_c$	10.11 ± 0.16	10.02 ± 0.18	10.00
$\hat{A}_s$	5.12 ± 0.16	5.09 ± 0.18	5.00

Table 4: Estimated spectral index for cases I, II and III ($N = 500$)

Case	$\alpha = 0.5$	$\alpha = 1.17$	$\alpha = 1.83$	$\alpha = 2.5$
I	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.07
II	0.57 ± 0.13	1.23 ± 0.11	1.82 ± 0.10	2.58 ± 0.10
III	0.67 ± 0.14	1.11 ± 0.11	1.84 ± 0.10	2.42 ± 0.11

Table 5: Estimated spectral index for case I ($N = 500$)

Chain	$\alpha = 0.5$	$\alpha = 1.17$	$\alpha = 1.83$	$\alpha = 2.5$
1st	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.08
2nd	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.08
3rd	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.08
4th	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.07
5th	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.08
6th	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.08
7th	0.45 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.07
8th	0.45 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	$2.54^{+0.08}_{-0.07}$
9th	0.45 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.08
10th	0.46 ± 0.07	1.25 ± 0.07	1.80 ± 0.08	2.54 ± 0.07

Table 6: Estimated spectral index for case II ($N = 500$)

Chain	$\alpha = 0.5$	$\alpha = 1.17$	$\alpha = 1.83$	$\alpha = 2.5$
1st	$0.57^{+0.10}_{-0.12}$	1.24 ± 0.11	1.81 ± 0.10	2.58 ± 0.10
2nd	$0.57^{+0.12}_{-0.11}$	1.23 ± 0.10	1.82 ± 0.10	2.58 ± 0.10
3rd	0.57 ± 0.11	1.22 ± 0.10	1.81 ± 0.10	2.58 ± 0.10
4th	$0.56^{+0.10}_{-0.11}$	1.24 ± 0.11	1.82 ± 0.10	2.58 ± 0.10
5th	0.56 ± 0.10	1.22 ± 0.10	1.82 ± 0.10	2.58 ± 0.10
6th	0.56 ± 0.12	1.24 ± 0.11	1.82 ± 0.10	2.58 ± 0.10
7th	$0.57^{+0.12}_{-0.11}$	1.23 ± 0.11	1.82 ± 0.10	2.58 ± 0.10
8th	0.56 ± 0.10	1.23 ± 0.11	1.82 ± 0.10	2.58 ± 0.10
9th	$0.57^{+0.12}_{-0.11}$	1.23 ± 0.11	1.82 ± 0.10	2.58 ± 0.10
10th	0.59 ± 0.13	1.23 ± 0.10	1.82 ± 0.10	2.58 ± 0.10

Table 7: Estimated spectral index for case III ($N = 500$)

Chain	$\alpha = 0.5$	$\alpha = 1.17$	$\alpha = 1.83$	$\alpha = 2.5$
1st	$0.69^{+0.16}_{-0.15}$	$1.10^{+0.10}_{-0.11}$	1.84 ± 0.09	2.43 ± 0.12
2nd	0.66 ± 0.11	1.1 ± 0.10	1.86 ± 0.10	2.41 ± 0.11
3rd	$0.71^{+0.17}_{-0.15}$	1.11 ± 0.10	1.85 ± 0.09	2.42 ± 0.11
4th	0.66 ± 0.11	1.12 ± 0.11	1.85 ± 0.09	$2.41^{+0.12}_{-0.11}$
5th	0.67 ± 0.12	1.11 ± 0.11	1.84 ± 0.09	2.41 ± 0.11
6th	0.65 ± 0.11	1.10 ± 0.10	1.84 ± 0.09	2.41 ± 0.11
7th	0.65 ± 0.11	1.11 ± 0.11	1.84 ± 0.09	2.42 ± 0.11
8th	$0.67^{+0.12}_{-0.11}$	1.11 ± 0.10	1.83 ± 0.09	$2.45^{+0.13}_{-0.12}$
9th	$0.65^{+0.11}_{-0.12}$	1.10 ± 0.10	1.84 ± 0.09	2.41 ± 0.11
10th	$0.71^{+0.17}_{-0.15}$	1.11 ± 0.11	1.84 ± 0.09	2.41 ± 0.11

Table 8: Estimated Parameters for GPS data case

Parameter	Estimate
$\hat{\alpha}$	0.67 ± 0.07
$\hat{\sigma}_{pl}$	$12.4 \pm 0.8 \ mm/yr^{0.17}$
$\hat{\sigma}_{wn}$	$1.3 \pm 0.8 \ mm$
$\hat{v}_0$	$2.6 \pm 0.9 \ mm/yr$
$\hat{y}_0$	$-4.1 \pm 1.6 \ mm$
$\hat{A}_c$	$0.7 \pm 0.5 \ mm$
$\hat{A}_s$	$1.3 \pm 0.7 \ mm$

Table 9: Parameters for convergence and good-mixing test

Case	r	k^*	k_{min}	β
I	10^{-4}	81 ± 4	0.03	1.73 ± 0.08
II	10^{-3}	2.7 ± 0.2	0.03	1.02 ± 0.02
III	0.009	2.8 ± 0.3	0.06	1.3 ± 0.1
GPS	0.002	8.9 ± 0.7	0.06	1.47 ± 0.07
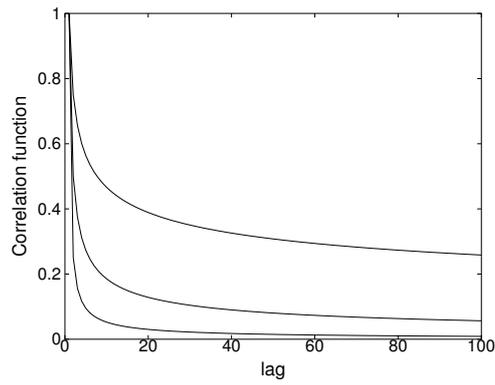
Figure 1: Correlation function vs. lag, from bottom to top, for $\alpha = 0.5$, 1.0, 1.5, respectively.
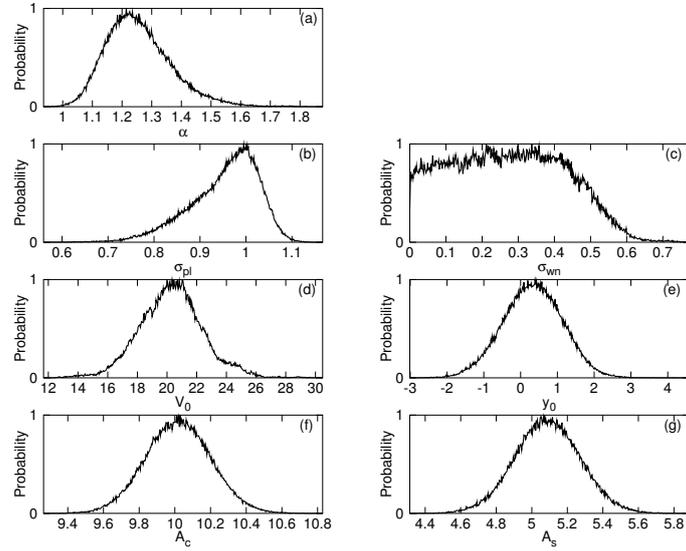
Figure 2: Histograms of parameters for synthetic data for case *IIIb*.
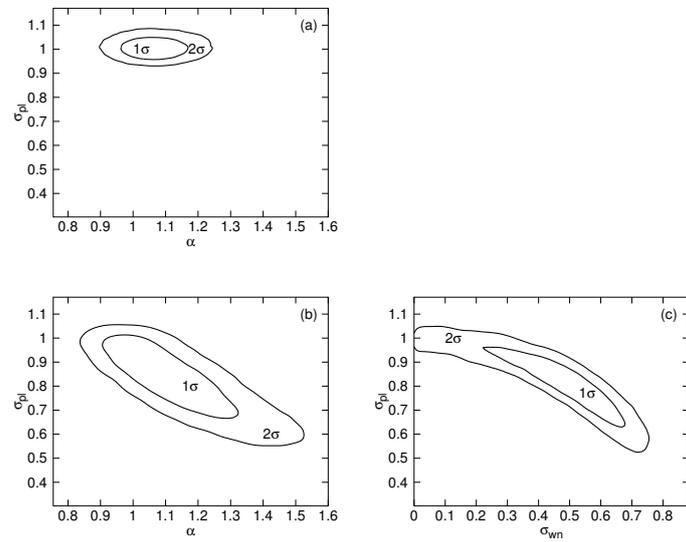

Figure 3: Confidence levels for 2D marginalized likelihoods for case *I* (a) and *II* (b, c).
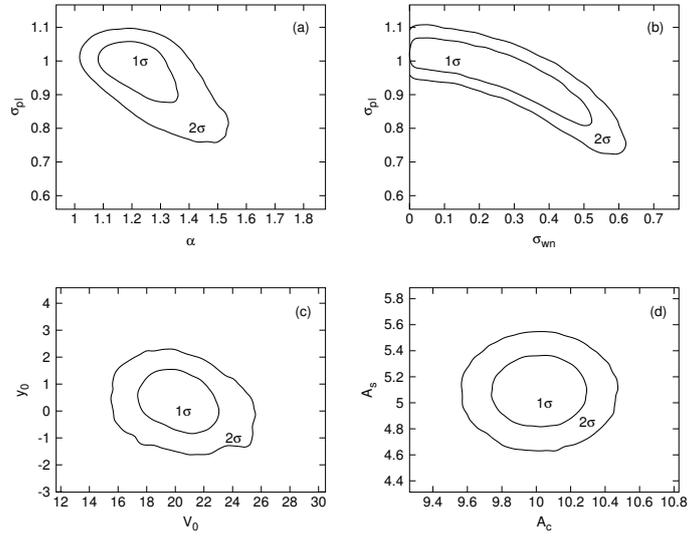
Figure 4: Confidence levels for 2D marginalized likelihoods for case IIIb.
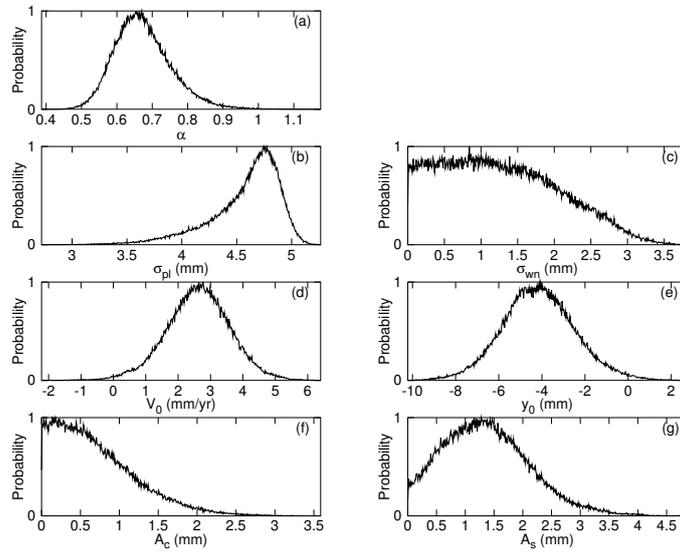

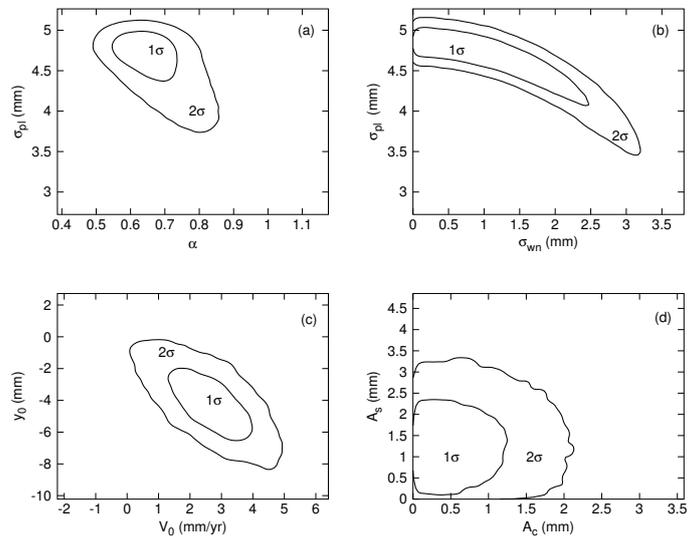Figure 5: Histograms of parameters for GPS data (Up coordinate).

Figure 6: Confidence levels for 2D marginalized likelihoods for the GPS data case.