ORIGINAL ARTICLE

Fast error analysis of continuous GPS observations

**M. S. Bos · R. M. S. Fernandes ·
S. D. P. Williams · L. Bastos**

Abstract It has been generally accepted that the noise in continuous GPS observations can be well described by a power-law plus white noise model. Using maximum likelihood estimation (MLE) the numerical values of the noise model can be estimated. Current methods require calculating the data covariance matrix and inverting it, which is a significant computational burden. Analysing 10 years of daily GPS solutions of a single station can take around 2 h on a regular computer such as a PC with an AMD Athlon™ 64 X2 dual core processor. When one analyses large networks with hundreds of stations or when one analyses hourly instead of daily solutions, the long computation times becomes a problem. In case the signal only contains power-law noise, the MLE computations can be simplified to a $\mathcal{O}(N \log N)$ process where N is the number of observations. For the general case of power-law plus white noise, we present a modification of the MLE equations that allows us to reduce the number of computations within the algorithm from a cubic to a quadratic function of the number of observations when there are no data gaps. For time-series of three and eight years, this means in practise a reduction factor of around 35 and 84 in computation time without loss of accuracy. In addition, this modification removes the implicit assumption that there is no environment noise before the first observation. Finally, we present an analytical expression for the uncertainty of the estimated trend if the data only contains power-law noise.

M. S. Bos (✉) · L. Bastos
University of Porto, Monte da Virgem,
4430-146 Vila Nova de Gaia, Portugal
e-mail: msbos@fc.up.pt

L. Bastos
e-mail: lcbastos@fc.up.pt

R. M. S. Fernandes
University of Beira Interior, CGUL, IDL,
R. Marquês d'Ávila e Boloma, 6201-001 Covilhã, Portugal

R. M. S. Fernandes
Delft University of Technology, DEOS,
Kluyverweg 1, 2629 HS Delft, The Netherlands
e-mail: rmanuel@di.ubi.pt

S. D. P. Williams
Proudman Oceanographic Laboratory,
6 Brownlow Street, Liverpool L3 5DA, UK
e-mail: sdwil@pol.ac.uk

1 Introduction

Error analysis of the estimates of continuous GPS positions have received a lot of attention in the last few years (Zhang et al. 1997; Mao et al. 1999; Williams et al. 2004). It is now established that GPS residuals show temporal correlation and that this has to be taken into account to produce realistic error bars on the GPS-derived velocities. This temporal correlation in the noise is usually described by a power-law plus white noise model. The latter has equal power at all frequencies, while the power-law is defined by its power spectrum P (Agnew 1992; Kasdin 1995):

$$P(f) = \frac{P_0}{f^\alpha} \tag{1}$$

where f is the frequency, P_0 a constant and α the spectral index.

The problem is then reduced to estimating the spectral index of the power-law and the estimation of the amplitudes of the power-law and white noise random variables. Mao et al. (1999) showed that the most accurate results are obtained with maximum likelihood estimation (MLE). However, the current formulation of the MLE for GPS time-series data is computationally intensive and grows with the cube of the number of observations.

This paper presents a simple modification of the MLE equations, which enables us to reduce the required number of operations to a quadratic function when there are no data gaps. At the same time, the new algorithm avoids the standard implicit assumption that there is no noise before the first observation has been made.

2 Review of current GPS error analysis techniques

If each data point x_i is independent, normally distributed and if it has a standard deviation of σ_i, then the following equation follows a χ^2 probability distribution with $N - M$ degrees of freedom:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{x_i - \hat{x}_i}{\sigma_i} \right)^2 \tag{2}$$

where N is the number of observations and M is the number of parameters used to estimate $\hat{\mathbf{x}}$.

Since Eq. (2) involves taking the difference between the observations and the estimated model, it produces an indication of the goodness of fit. For a large number of data points, say 30, the value χ^2 is approximately $N - M$. The reduced chi-square is defined as $\chi^2/(N - M)$ and consequently has a value around one. When the reduced χ^2 value is much larger than one, it indicates that the estimated model is wrong or that the assumed errors are too small.

GPS analysis software provides standard deviations for each estimated station position, but in almost all cases these are too small to be realistic. To get better error values, they are usually scaled by $\sqrt{\chi^2/(N - M)}$. This empirical reduced chi-square method is widely used within the geodetic community, but is not the correct solution to improve the error bars. The problem is that one assumes that the GPS residuals are independent from each other, while in reality they show a strong correlation in time (Johnson and Agnew 1995). Mao et al. (1999) have shown that after applying the reduced χ^2 value, the real error can still be underestimated by a factor of 5–11.

Figure 1 shows the North component of the GPS time-series at Kootwijk (KOSG), The Netherlands. It also shows the residuals after subtracting the least-squares estimated linear motion. From these residuals, the power-spectrum was computed and is shown in Fig. 2. It illustrates the need for an
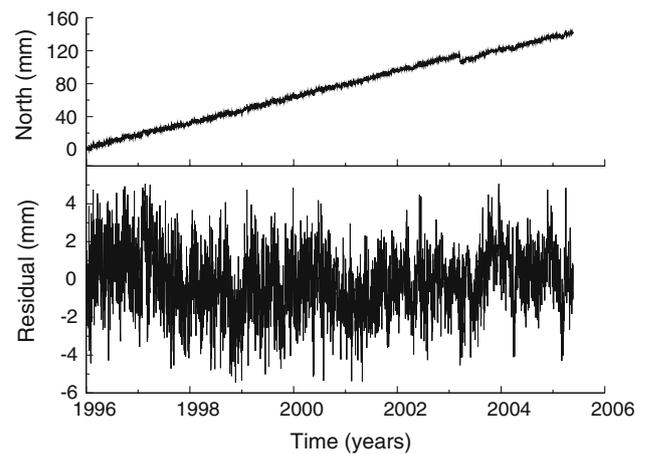
Fig. 1 The upper panel shows the North component of the GPS position time-series at KOSG. The *lower panel* shows the same data set after subtraction of a linear trend, a yearly signal and one offset
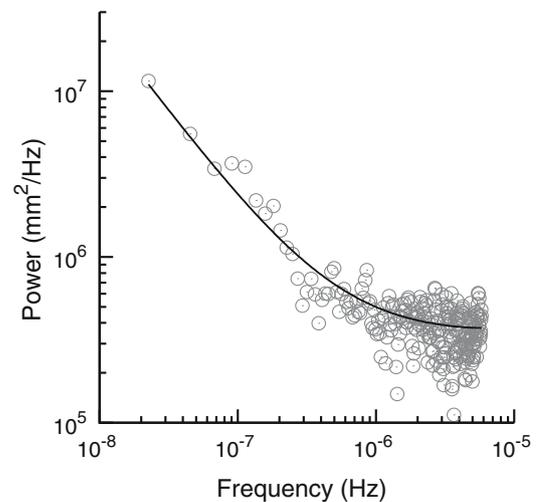
Fig. 2 Power spectrum of the GPS residuals at KOSG, North component. The *circles* denote the spectrum computed from the observations and the *solid line* is the fitted power-law plus white noise model

improved noise model. In the high frequencies, the noise is flat (property of white noise), while for the lower frequencies, the spectrum seems to obey a power-law. The slope of the power-law is around one, which is sometimes called flicker noise. Williams et al. (2004) have shown that the power spectra in Fig. 2 is representative for many GPS residuals.

Following Langbein and Johnson (1997) and Kasdin (1995), the spectral index is represented by α. A power-law noise time-series, say $\mathbf{r}$, with a spectral index between -1 and 1 can be generated by convolving it with independent and identically distributed (IID) noise w_k (Hosking 1981):

$$r_k = \sum_{i=0}^{\infty} h_i \, w_{k-i} \tag{3}$$

where, following Kasdin (1995)

$$h_0 = 1$$
$$h_i = \left(\frac{\alpha}{2} + i - 1\right)\frac{h_{i-1}}{i} \qquad (4)$$

These relations follow from Hosking's definition of the fractional difference operation:

$$
\begin{aligned}
(1 - B)^{\alpha/2} &= \sum_{i=0}^{\infty} \binom{\alpha/2}{i}(-B)^i \\
&= \sum_{i=0}^{\infty} \frac{\Gamma(\alpha/2 + 1)}{\Gamma(\alpha/2 - i + 1)\, i!}(-B)^i \\
&= 1 - \frac{\alpha}{2}B - \frac{1}{2}\frac{\alpha}{2}\left(1 - \frac{\alpha}{2}\right)B^2 - \cdots \qquad (5)
\end{aligned}
$$

where Γ is the Gamma function and B is the backward-shift operator ($Bx_k = x_{k-1}$). From this definition, one can deduce that:

$$(1 - B)^{\alpha/2} r_k = w_k \qquad (6a)$$

$$(1 - B)^{-\alpha/2} w_k = r_k \qquad (6b)$$

In order to apply the backward-shift operator, Eq. (6a) assumes the existence of an infinite amount of $\mathbf{r}$ before r_k, while Eq. (6b) assumes the same for $\mathbf{w}$.

In this research, since we observe a finite amount of signal with coloured noise $\mathbf{r}$, we will concern ourselves mostly with Eq. (6b). The covariance between r_k and r_l only depends on $\tau = l - k$. It thus forms a Toeplitz matrix, which can be written as $\mathrm{Cov}(r_k, r_l) = \gamma(\tau)$. A Toeplitz matrix is a matrix in which each descending diagonal from left to right is constant. If one assumes an infinite sequence of zero-mean white noise $\mathbf{w}$, with variance σ_{pl}^2, then the covariance is (Hosking 1981; Appendix A):

$$
\begin{aligned}
\gamma(0) &= \sigma_{pl}^2 \frac{\Gamma(1 - \alpha)}{(\Gamma(1 - \frac{\alpha}{2}))^2} \\
\gamma(\tau) &= \frac{\frac{\alpha}{2} + \tau - 1}{-\frac{\alpha}{2} + \tau}\gamma(\tau - 1) \quad \text{for } \tau > 0 \qquad (7)
\end{aligned}
$$

The covariance of Eq. (7) only exists for $\alpha < 1$. For $\alpha \geq 1$, the noise is not stationary and therefore produces an infinite variance. Since the power-law noise in GPS may have a spectral index larger than one, this causes a problem. For this reason, one usually assumes that $w_i = 0$ for $i < 0$, which means that the sum only runs to k to produce finite covariance values. With this assumption, the covariance between r_k and r_l with $l > k$ is:

$$\mathrm{Cov}(r_k, r_l) = \sigma_{pl}^2 \sum_{i=0}^{k} h_i h_{i+(l-k)} \qquad (8)$$

Equation (8) can also be written as a matrix multiplication $\mathbf{C} = \mathbf{U}^T \mathbf{U}$. Matrix $\mathbf{U}$ is an upper triangular matrix and looks like:

$$
\mathbf{U} = \begin{pmatrix}
h_0 & h_1 & \ldots & h_N \\
0 & h_0 & & h_{N-1} \\
\vdots & & \ddots & \vdots \\
0 & 0 & \ldots & h_0
\end{pmatrix} \qquad (9)
$$

Using Eq. (3) and again assuming that $w_i = r_i = 0$ for $i < 0$, one has the following relations:

$$\mathbf{r} = \mathbf{U}^T \mathbf{w} \qquad (10a)$$

$$\mathbf{w} = \mathbf{U}^{-T} \mathbf{r} \qquad (10b)$$

For clarity, we have omitted the scaling factor $\Delta T^{\alpha/2}$ used by Williams (2003), where ΔT is the sampling interval, since we will only discuss daily GPS solutions here.

To get the total noise, pure white noise with variance σ_w^2 has to be added to the power-law noise. Writing Eq. (8) as a matrix $\mathbf{E}(\alpha)$ and using $\mathbf{I}$ for the unit matrix, the new covariance matrix $\mathbf{C}$ for the GPS data can be written as (Langbein and Johnson, 1997; Langbein, 2004; Williams, 2003):

$$\mathbf{C} = \sigma_w^2 \mathbf{I} + \sigma_{pl}^2 \mathbf{E}(\alpha) \qquad (11)$$

The east, north and up components all have a separate covariance matrix $\mathbf{C}$. The GPS solutions provide a spatial correlation between the components but this is normally neglected when Eq. (11) is considered because they are small compared to the temporal correlations. Furthermore, it is assumed that, for each component, the variances are equal at each data point. This condition is mostly met in practise.

2.1 Maximum likelihood estimation

To estimate the parameters that describe the trend in the GPS observations and the parameters of the noise model, we must maximise the probability p that their values have occurred for given observations $\mathbf{x}$. If a Gaussian distribution is assumed, then this probability is:

$$
\begin{aligned}
p(\mathbf{x}) = &\frac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{C})} \\
&\times \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}_{ot})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}_{ot})\right] \qquad (12)
\end{aligned}
$$

where the vector $\hat{\boldsymbol{\theta}}_{ot}$ contains the parameters for the nominal value, which is the offset of the whole time-series and the trend. The observation matrix $\mathbf{H}$ has N rows and 2 columns. The first contains solely ones, while the second column contains a linear trend. Langbein (2004) notes that it is straightforward to include also a yearly signal and offsets in the estimation process. Note that the observed minus the estimated trend will produce the residuals $\mathbf{r}$ discussed in the Sect. 2: $\mathbf{r} = \mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}_{ot}$.

In practise, the maximum of the logarithm of this probability p is computed, which is an equivalent problem because the logarithm is a monotonically increasing function. The covariance matrix $\mathbf{C}$ depends on the noise and not on $\hat{\boldsymbol{\theta}}_{ot}$. Therefore, the maximum can be found by setting the derivative of the logarithm of Eq. (12) to zero.

After some rearrangement, one finds:

$$\hat{\boldsymbol{\theta}}_{ot} = \left(\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} \tag{13}$$

This is the formula for weighted least squares. Using Eq. (13), the values for the noise parameters α, σ_{pl} and σ_w must be found by numerically maximising:

$$\ln(p(\mathbf{x})) = -\frac{1}{2}\left[N \ln(2\pi) + \ln \det(\mathbf{C}) \right.$$
$$\left. + (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}_{ot})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}_{ot})\right] \tag{14}$$

Details about this computation are given by Langbein and Johnson (1997) and Langbein (2004).

3 Remaining problems

MLE provides a reliable error estimate for the trend (Mao et al. 1999). However, its main disadvantage is the computation time. Analysing the North, East and Up component of 10 years of daily GPS solutions of one station can take as long as two hours on a regular computer such as a PC with an AMD Athlon™ 64 X2 Dual Core Processor. The error analysis of large networks or the analysis of hourly solutions can therefore take several days.

The causes of the relatively long computation time are the construction of the covariance matrix and the computation of Eqs. (13) and (14), all requiring $\mathcal{O}(N^3)$ operations. For this reason, the spectral index is often a priori fixed to a value of one to accelerate the process.

Another approach is to estimate the noise properties directly from the power-spectrum (Mao et al. 1999), which is extremely fast because the power-spectrum can be computed with only $\mathcal{O}(N \log N)$ operations for evenly spaced data. However, we prefer the MLE method because it allows consistent estimates of the components of the covariance function and the parameters of the time-dependent model of the data (Langbein 2004).

4 Fast power-law noise analysis

If the noise consists only of pure power-law noise, then the logarithm of the determinant of the covariance matrix is, for all α, equal to $2N \ln \sigma_{pl}$. To compute the likelihood function (Eq. 14), it would be desirable to have a fast way to compute $\mathbf{C}^{-1}$. Hosking (1981) describes how the moving average equation for the noise can be rewritten as an autoregressive representation:

$$w_k = \sum_{i=0}^{\infty} p_i r_{k-i} \tag{15}$$

where

$$p_0 = 1$$
$$p_i = \left(-\frac{\alpha}{2} + i - 1\right)\frac{p_{i-1}}{i} \tag{16}$$

The advantage of this formulation is that it allows us to compute the inverse of the covariance matrix $\mathbf{C}$ directly. Remember that we had assumed that $w_i = r_i = 0$ for $i < 0$, which has the consequence that the summation of Eq. (15) must be truncated to k. Furthermore, using the relation $\mathbf{C}^{-1} = \mathbf{U}^{-1}\mathbf{U}^{-T}$ and Eq. (10b), we get with $p_i = 0$ for $i < 0$:

$$(\text{Cov}(r_k, r_l))^{-1} = \sum_{i=0}^{N-k} p_i \, p_{i+(k-l)} \tag{17}$$

In the likelihood function (Eq. 14), we need to compute $\mathbf{r}^T \mathbf{C}^{-1} \mathbf{r}$. If one writes the elements p_i in a lower triangular matrix $\mathbf{L} = \mathbf{U}^{-T}$, this becomes: $(\mathbf{L}\mathbf{r})^T(\mathbf{L}\mathbf{r})$. A major reduction in computation time is achieved by taking advantage of the fact that $\mathbf{L}\mathbf{r}$ is a convolution operation with filter p_i, which 'whitens' the residuals $\mathbf{r}$. In Eq. (13) the same filter operation is applied to each column of $\mathbf{H}$. Press et al. (1988) describe how this can be efficiently computed with fast Fourier transform in $\mathcal{O}(N \log N)$ operations. Thus the MLE for pure power-law noise can be reduced from $\mathcal{O}(N^3)$ to $\mathcal{O}(N \log N)$ operations for this special case.

Before ending this section, we will present the inverse of the covariance matrix $\mathbf{C}$ when the number of observations goes to infinity. This inverse is well behaved for $\alpha > -1$ and converges to the following Toeplitz matrix, $\text{Cov}^{-1}(r_k, r_l) = \gamma^{-1}(\tau)$, with $\tau = l - k$:

$$\gamma^{-1}(0) = \frac{1}{\sigma_{pl}^2} \frac{\Gamma(1+\alpha)}{\left(\Gamma(1+\frac{\alpha}{2})\right)^2}$$
$$\gamma^{-1}(\tau) = \frac{\tau - 1 - \frac{\alpha}{2}}{\tau + \frac{\alpha}{2}} \gamma^{-1}(\tau - 1) \quad \text{for } \tau > 0 \tag{18}$$

Using Eqs. (6a) and (6b), one can see that this is the result of changing $\alpha/2$ with $-\alpha/2$ in Eq. (7). Equation (18) only represents the approximation of the inverse of the covariance matrix when the number of observations grows to infinity. Thus, in practise one still has to use Eq. (17) where the summations are taken up to $N - k$ since w_k, and therefore r_k, are zero for $k < 0$.

If there really is no noise before the first observation is made, it would be better to stop and restart the observations as often as possible to avoid any power-law noise to develop

in the time-series. Intuitively this is incorrect since the start of the observations has no physical effect on the GPS monumentation, which Langbein and Johnson (1997) consider to be the cause of some of the power-law noise. This problem is solved in the next section.

5 A fast alternative

In Sect. 2, it was stated that power-law noise can be generated by convolving white noise with a transfer function h. Since we do not know anything about the signal before the first observation, the convolution in Eq. (3) was taken only over all the available data points. However, now we make the assumption that the noise is present prior to the start of the observations and that the convolution summation needs be taken to infinity.

Unfortunately, for $\alpha \geq 1$, the noise is non-stationary and the summation would grow to infinity. For this reason, Agnew (1992) used a structure function that is defined as the difference between two observations at time t and time $t + T$. Here, only the discrete case will be discussed and the difference of the GPS residuals will be restricted to one time step, $\Delta r_k = r_{k+1} - r_k$. Using Eq. (3), this can be written as (Kasdin 1995)

$$\Delta r_k = w_{k+1} + \left(\frac{\alpha}{2} - 1\right) \sum_{i=0}^{k} \frac{h_i}{i+1} w_{k-i} \tag{19}$$

The same difference can be taken between the points r_{l+1} and r_l. The power-law covariance between these two first differences is, with $\tau = k - l$, $k \geq l$:

$$\text{Cov } (\Delta r_k, \Delta r_l)$$
$$= \sigma_{pl}^2 \left[1 + \left(\frac{\alpha}{2} - 1\right)^2 \sum_{i=0}^{l} \frac{h_i^2}{(i+1)^2}\right] \quad \text{for } k = l$$
$$= \sigma_{pl}^2 \left[\left(\frac{\alpha}{2} - 1\right) \frac{h_{\tau-1}}{\tau} + \left(\frac{\alpha}{2} - 1\right)^2\right.$$
$$\left. \times \sum_{i=0}^{l} \frac{h_{i+\tau} h_i}{(i+\tau+1)(i+1)}\right] \quad \text{for } k \neq l \tag{20}$$

Equation (20) is the exact equivalent to the covariance matrix of Eq. (8), but for first-differenced data.

Let us now apply our assumption that the first observation is already affected by noise and that this noise, because we took the first difference, is stationary. In this case, we can extend our summations in Eq. (20) to infinity. Using Eq. (4) and some hypergeometric function relations (Abramowitz and Stegun 1965) the covariance $\text{Cov}(\Delta r_k, \Delta r_l) = \text{Cov}(\Delta r_l,$

$\Delta r_k) = \gamma_{pl}(\tau)$ can be written as, see Appendix A:

$$\gamma_{pl}(0) = \sigma_{pl}^2 \frac{\Gamma(3-\alpha)}{(\Gamma(2-\frac{\alpha}{2}))^2}$$
$$\gamma_{pl}(\tau) = \frac{\frac{\alpha}{2} + \tau - 2}{1 - \frac{\alpha}{2} + \tau} \gamma_{pl}(\tau - 1) \quad \text{for } \tau > 0 \tag{21}$$

Equation (21) is the same as Eq. (7) but with $\alpha/2$ replaced by $\alpha/2 - 1$, which causes it to be valid for $\alpha < 3$. The advantage of the covariance matrix given by Eq. (21) is that its construction takes $\mathcal{O}(N)$ instead of $\mathcal{O}(N^3)$ operations because only the vector γ_{pl} needs to be known to describe the whole matrix. Looking back at Eq. (6a), one can see that Eq. (21) is the result of rewriting the definition of the fractional difference:

$$(1 - B)^{\alpha/2} \mathbf{r} = (1 - B)^{\alpha/2-1} (1 - B)\mathbf{r}$$
$$= (1 - B)^{\alpha/2-1} \Delta \mathbf{r} \tag{22}$$

The covariance between two first differenced white noise observations is:

$$\gamma_w(0) = 2\sigma_w^2$$
$$\gamma_w(1) = -\sigma_w^2$$
$$\gamma_w(\tau) = 0 \quad \text{for } \tau > 1. \tag{23}$$

The sum of power-law and white noise still produces a Toeplitz matrix. Using the algorithm of Bojanczyk et al. (1995), the Cholesky decomposition of the Toeplitz covariance matrix can be performed in $\mathcal{O}(N^2)$ operations. Let us write this decomposition as $\mathbf{C} = \mathbf{U}^T \mathbf{U}$, where $\mathbf{U}$ is an upper triangular matrix. A further reduction in computation time can be obtained by rewriting Eq. (13) as

$$\hat{\boldsymbol{\theta}}_{ot} = \left(\mathbf{H}^T (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^T (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{x}$$
$$= \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{y} \tag{24}$$

where

$$\mathbf{U}^T \mathbf{A} = \mathbf{H} \tag{25a}$$
$$\mathbf{U}^T \mathbf{y} = \mathbf{x} \tag{25b}$$

Equations (25a) and (25b) can be directly solved by back-substitution because $\mathbf{U}$ is an upper triangular matrix. Also note that $\mathbf{H}$ and $\mathbf{x}$ now represent the first-differenced design matrix and the first-differenced observations.

Finally, one can compute the determinant and matrices $\mathbf{A}$ and $\mathbf{y}$ while performing the Cholesky factorisation. The factorisation computes each column i of matrix $\mathbf{U}$ with only the information of the column $i - 1$, except for the first column which is computed directly using γ [Eq. (21) plus Eq. (23)]. Thus only two columns of $\mathbf{U}$ need to be stored during the Cholesky decomposition, which is again beneficial for reducing the computation time. The full algorithm is given in Appendix B.

It must be noted that the first-differenced design matrix **H** no longer contains the nominal value of the whole time-series because we took the difference of the observations. Since this was a nuisance parameter, this is only to our advantage. Furthermore, the observation matrix can easily be extended to include other signals such as, for example, a yearly signal.

6 Analytical expressions for the slope uncertainty

The variance of the estimated slope can be computed with:

$$\sigma_{\hat{r}}^2 = \left(\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H} \right)^{-1} \tag{26}$$

Let us assume that the data only contains power-law noise. The design matrix **H** consists for first-differenced data only out of one column with ones. Using Eq. (21), one can now derive the following recurrence relations for the slope uncertainty $\sigma_{\hat{r}}$, for a time-series with N observations:

$$\sigma_{\hat{r}}^2(2) = \frac{\sigma_{pl}^2}{\Delta T^{2-\frac{\alpha}{2}}} \frac{\Gamma(3-\alpha)}{\Gamma(2-\frac{\alpha}{2})^2}$$

$$\sigma_{\hat{r}}^2(N) = \frac{N-2}{N+1-\alpha} \sigma_{\hat{r}}^2(N-1) \quad \text{for } N > 2 \tag{27}$$

Note that we have introduced here the scaling factor ΔT, the sampling period, to facilitate the comparison with earlier results. The explicit form of Eq. (27) is:

$$\sigma_{\hat{r}}^2 = \frac{\sigma_{pl}^2}{\Delta T^{2-\frac{\alpha}{2}}} \frac{\Gamma(3-\alpha)\Gamma(N-1)\Gamma(4-\alpha)}{(\Gamma(2-\frac{\alpha}{2}))^2 \Gamma(N+2-\alpha)} \tag{28}$$

For large values of N, the Gamma function is difficult to compute. A good approximation, however, can be obtained with the use of Stirling's formula (Abramowitz and Stegun 1965). Using this approximation, we get the following equation for the variance of the estimated slope:

$$\sigma_{\hat{r}}^2 \approx \frac{\sigma_{pl}^2}{\Delta T^{2-\frac{\alpha}{2}}} \frac{\Gamma(3-\alpha)\Gamma(4-\alpha)(N-1)^{\alpha-3}}{(\Gamma(2-\frac{\alpha}{2}))^2} \tag{29}$$

By setting $\alpha = 0$, one can verify that Eq. (28) produces the same expression for white noise as calculated by Zhang et al. (1997) among others:

$$\sigma_{\hat{r}}^2 = \frac{12\sigma_{pl}^2}{\Delta T^2(N^3-N)} \tag{30}$$

and the same expression for random walk noise ($\alpha = 2$):

$$\sigma_{\hat{r}}^2 = \frac{\sigma_{pl}^2}{\Delta T(N-1)} \tag{31}$$

Thus, for these two special cases, there is no effect when taking into account the noise before the first observation. This can be explained by the fact that white noise does not depend on past values, while the random walk noise before $t = 0$ will only cause a change in the nominal value. However, for all
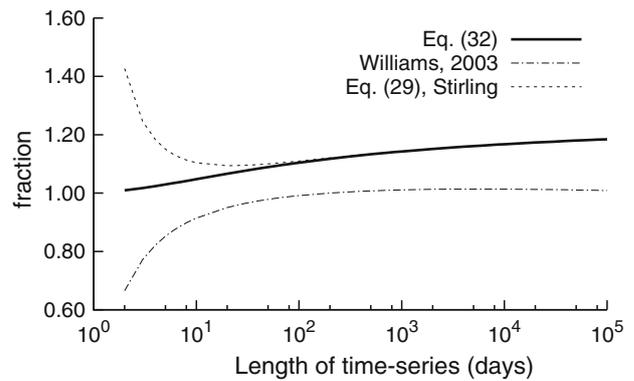
Fig. 3 Trend uncertainties computed using different formulas, divided by our reference trend uncertainty as defined in the text. All for flicker noise ($\alpha = 1$)

other values of the spectral index, there is a slight difference. A new result is that we now have an analytical expression for the case of flicker noise ($\alpha = 1$):

$$\sigma_{\hat{r}}^2 = \frac{8\sigma_{pl}^2}{\pi \Delta T^{\frac{3}{2}}(N^2-N)} \tag{32}$$

Equations (29) and (32) are new results, and it is interesting to compare them with the standard trend uncertainty that is computed, without taking first-differences, by constructing the covariance matrix using Eq. (8) and inserting it into Eq. (26). As was mentioned before, flicker noise is the most common type of power-law noise observed in GPS data, thus α will be set to one. This standard trend uncertainty will be our reference to which the new equations will be compared.

Williams (2003) provided an approximation for the reference trend uncertainty, which will also be included in the comparison. Figure 3 shows the standard deviation of the estimated slope using Eq. (32), the Stirling approximation Eq. (29) and the approximation of Williams (2003) as function of the length of the time-series. These three trend uncertainties have been divided by our reference trend uncertainty to make it easier to see their differences.

Figure 3 shows that after 100 days, the Williams approximation is in very good agreement with our reference trend uncertainty because the fraction goes to unity. After 100 days, the Stirling approximation, Eq. (29), is in good agreement with Eq. (32). Figure 3 also shows that the new method gives, for time-series longer than 1,000 days, a value for the trend uncertainty that is around 15–20% larger than the reference trend uncertainty. One could argue that the reference trend uncertainty is in fact 15–20% too small and that the new method, with its associated uncertainty, is in fact more realistic because the original method does not account for noise before the start of the series.

Figure 4 shows the error in the estimated trend as function of the data span using the noise properties that are observed
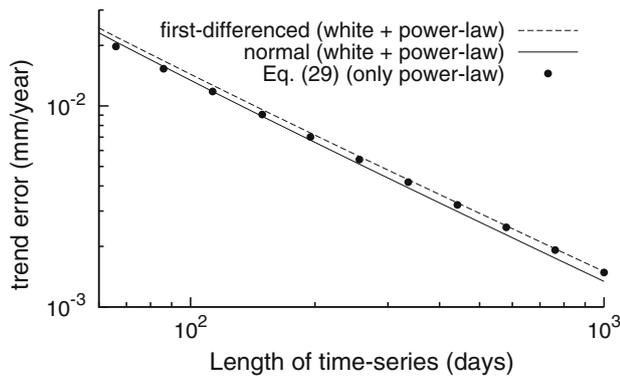
Fig. 4 The trend uncertainty at KOSG for the North component as function of the observation span, using the white plus power-law noise properties observed in these data. The *solid line* is computed using undifferenced data, the *dashed line* for first-differenced data. The points have been computed under the assumption that only power-law noise is present

in the North component of the GPS time-series at KOSG. In Fig. 4, the uncertainty was computed using Eq. (26) with the normal covariance matrix of Eq. (11) and the first-differenced covariance matrix, which is the sum of Eqs. (21) and (23). The trend uncertainty was also computed using only power-law noise (Eq. 29).

As was observed in Fig. 3, Fig. 4 shows that the new covariance matrix for first-differenced data produces a larger trend uncertainty than the reference trend uncertainty. Figure 4 also shows that, after 2 years, the trend uncertainties for first-differenced data of the power-law noise only and the power-law plus white noise are equal. Thus after 2 years, the power-law noise completely dominates the uncertainty, and the contribution of the white noise can be neglected.

Mao et al. (1999) concluded that the velocity error in time-series could be underestimated by factors of 5–11 if a pure white noise model is assumed. Taking into account that Eq. (32) produces larger values than the reference trend uncertainty used by Mao et al. (1999), these factors can grow up to 6–13.

7 Numerical results

Section 6 presented analytical expressions for the trend uncertainty for given noise parameters of the power-law noise. In reality, we are confronted with power-law plus white noise in the GPS data of which the parameter values are unknown. To test the performance of the normal MLE and the new first-differenced MLE approach, one hundred synthetic time-series with 1,000 and 3,000 days were generated, which have the same trend values and noise properties as observed in the North component of KOSG data.

The mean values of the estimated parameters and their observed standard deviation using the two MLE approaches

are given in Table 1. These values were obtained by implementing the algorithms described in Sects. 2 and 5 and using the BLAS and LAPACK libraries (Anderson et al. 1999), which are optimised for matrix computations.

From Table 1, one can see that for time-series of 1,000 days, both MLE methods underestimate the spectral index. This was already observed by Williams et al. (2004) and is attributed to the fact that the trend will absorb some of the very long periods of the power-law noise. For time-series of 3,000 days, this bias is less.

For a very large number of synthetic time-series, the observed standard deviation of the trend value should be equal to the predicted one for given values of the spectral index and the variances of the power-law and white noise. Table 1 shows that, using 100 time-series of 1,000 days, both methods have an ensemble mean trend value of 15.56 mm/year with a standard deviation of 0.473 and 0.489 mm/year for the normal and first-differenced MLE respectively. Using the estimated noise properties, Eqs. (11), (21), (23) and (26), the mean predicted errors are 0.398 and 0.456 mm/year, again for the normal and first-differenced MLE, respectively.

The predicted errors from the first-differenced MLE are much closer to the standard deviation of the trends for both methods, again reinforcing the notion that the new method is more realistic. Taking the uncertainty in the predicted errors into account, the observed and predicted trend uncertainties are in good agreement. For time-series of 3,000 days, the agreement is even better.

Table 1 also lists the computation time, and shows that the new method is indeed significantly faster. The computer used for these computations contained an AMD Athlon™ 64 X2 Dual Core Processor 4200+, with 2 Gbyte of memory. The computation times for the normal MLE and the first-differenced MLE methods for different lengths of the time-series are plotted in Fig. 5.
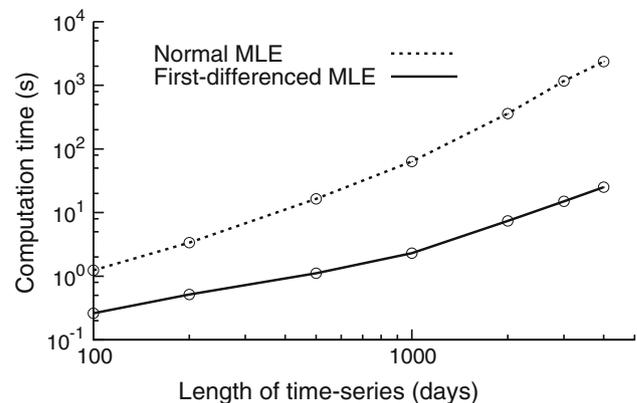
Fig. 5 The computation time of the normal and first-differenced MLE methods for different lengths of the time-series

Table 1 The results of estimating the trend and noise parameters in 100 time-series of 1,000 and 3,000 days where the spectral index $\alpha = 1.105$, the standard deviation of the power-law noise $\sigma_{pl} = 0.691$ mm, the standard deviation of the white noise $\sigma_w = 1.393$ mm and the trend is 15.621 mm/year

| | 1,000 days | | | | 3,000 days | | | |
| | Normal MLE | | Differenced MLE | | Normal MLE | | Differenced MLE | |
Parameter	Mean	σ	Mean	σ	Mean	σ	Mean	σ
α	0.917	± 0.233	0.950	± 0.224	1.058	± 0.095	1.067	± 0.093
σ_{pl} (mm)	0.858	± 0.232	0.827	± 0.223	0.724	± 0.096	0.717	± 0.093
σ_w (mm)	1.283	± 0.171	1.304	± 0.152	1.372	± 0.051	1.375	± 0.050
Trend (mm/year)	15.560	± 0.473	15.564	± 0.489	15.631	± 0.171	15.632	± 0.180
Predicted error (mm/year)	0.398	± 0.155	0.456	± 0.172	0.154	± 0.032	0.178	± 0.037
Mean time (s)	76		2		1348		16	

The mean time indicates the average computation time of each run

8 Data gaps

So far, we have not mentioned data gaps, which are unfortunately present in most GPS time-series. To correct for these gaps, one can construct, both for the normal and the first-differenced MLE method, first the full covariance matrix for the complete time-series and afterwards delete the rows and columns for which no data are available (Williams 2003).

Unfortunately, the deletion of rows and columns destroys the Toeplitz structure of the covariance matrix for the first-differenced MLE method and with that one loses the gain in computation speed. The exact solution of this problem is outside the scope of this research, but we have experimented with two possible remedies.

The first one is simply filling the data gaps by linear interpolation. We analysed 164 GPS stations with a wide range in the length of the observation span and interpolated the data gaps. For 90% of the stations, the trends estimated by the normal and the first-differenced method differed by less than one standard deviation from each other for all three components. Another approach is to ignore the gaps in the stochastic model and keep the Toeplitz covariance matrix. In this solution approximately 94% of the trends differed from the normal method by less than one standard deviation.

These results show that already with some simple methods, the data gap problem can be solved in a satisfactory manner.

9 Conclusions

It has been demonstrated that if there is only power-law noise in a continuous GPS time-series, then the normal MLE equations can be computed with $\mathcal{O}(N \log N)$ operations. Furthermore, by taking the first difference of the observations, power-law noise with a spectral index value around one,

which is commonly observed in GPS observations (Williams et al. 2004), is made stationary. The new covariance matrix can be written as a convenient recursive expression, the construction of which only takes $\mathcal{O}(N)$ operations.

Another advantage of the new covariance matrix is that it is a Toeplitz matrix. The Cholesky factorisation of this type of matrices can be performed in $\mathcal{O}(N^2)$ operations, instead of $\mathcal{O}(N^3)$ for positive definite matrices. Finally, the likelihood function can be computed without the explicit construction of the covariance matrix or its Cholesky factorisation, which again reduces the computation time. It has been shown that for time-series of 1,000 and 3,000 days, a reduction factor of around 35 and 84 respectively in computation time can be achieved.

The modification also removed the implicit assumption of no noise before the first observation. This generalisation made it possible to derive an analytical expression for the uncertainty of the estimated trend in the case of pure power-law noise. Using these expressions, we conclude that the trend uncertainty for flicker noise presented by Williams (2003) is 15–20% too small for time-series longer than three years and that the new method, with its associated uncertainty, is more realistic.

Appendix A Derivation of the covariance

The coefficients h_i of Eq. (4) can be written as (Kasdin 1995):

$$h_i = \frac{(\frac{\alpha}{2})_i}{(1)_i} = \frac{\Gamma(\frac{\alpha}{2} + i)\Gamma(1)}{\Gamma(\frac{\alpha}{2})\Gamma(1 + i)} = \frac{\Gamma(\frac{\alpha}{2} + i)}{\Gamma(\frac{\alpha}{2}) \, i!} \tag{33}$$

where $(a)_i = 1 \times a \times (a+1) \times \cdots \times (a+i-1)$ denotes the Pochammer symbol. Furthermore, the hypergeometric function is defined as (Abramowitz and Stegun 1965):

$$F(a, b; c; z) = \sum_{i=0}^{\infty} \frac{(a)_i (b)_i}{(c)_i} \frac{z^i}{i!}$$

$$= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{i=0}^{\infty} \frac{\Gamma(a+i)\Gamma(b+i)}{\Gamma(c+i)} \frac{z^i}{i!} \quad (34)$$

Thus, Eq. (8) can be written as:

$$\sum_{i=0}^{\infty} h_i h_{i+\tau} = \frac{1}{\left(\Gamma(\frac{\alpha}{2})\right)^2} \sum_{i=0}^{\infty} \frac{\Gamma(\frac{\alpha}{2}+i)\Gamma(\frac{\alpha}{2}+\tau+i)}{\Gamma(1+\tau+i)} \frac{1^i}{i!} \quad (35)$$

Using the following relation (Abramowitz and Stegun 1965):

$$F(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \quad (36)$$

Eq. (35) can be written as:

$$\sum_{i=0}^{\infty} h_i h_{i+\tau} = \frac{\Gamma(\frac{\alpha}{2}+\tau)\Gamma(1-\alpha)}{\Gamma(\frac{\alpha}{2})\Gamma(1+\tau-\frac{\alpha}{2})\Gamma(1-\frac{\alpha}{2})} \quad (37)$$

Finally, with the relation $\Gamma(\tau+1) = \tau\Gamma(\tau)$, one obtains Eq. (7). A similar derivation produces Eq. (21) from Eq. (20).

Appendix B New algorithm

Bojanczyk et al. (1995) describe a algorithm for performing a Cholesky factorisation of Toeplitz matrices. However, as shown in the text, we only need to know the determinant of the covariance matrix and the matrices **A** and **y** to compute our likelihood function; see Eqs. (14) and (24).

The determinant can be computed by multiplying the diagonal elements of the Cholesky decomposition. Matrix **A** and **y** are formed with back substitution using the Cholesky decomposition. The following algorithm, in GNU Octave notation (also provided as Electronic Supplementary Material with this article), computes the logarithm of the determinant and the required matrices for given covariance matrix `gamma_x`: sum of Eqs. (21) and (23) and first-differenced matrices **H** and **x**. n is the number of observations and m the number of columns of **H**.

```
function [lndeterminant,A,y] =
  mactrick(gamma_x,H,x)

U       = zeros(n,2); % C = U'*U
V       = zeros(n,2);
dummyH  = zeros(n,m);
dummyx  = zeros(n,1);
A       = zeros(n,m);
y       = zeros(n,1);
```

```
%--- define the generators u and v
U(:,1)   = gamma_x/sqrt(gamma_x(1));
V(2:n,1) = U(2:n,1);

%--- First element on diagonal
lndeterminant = log(U(1,1));

%--- First solution vector values
A(1,:) = H(1,:)/U(1,1);
y(1,1) = x(1,1)/U(1,1);
dummyH(2:n,:) = U(2:n,1)*A(1,:);
dummyx(2:n,1) = U(2:n,1)*y(1,1);

k_old =1;
k_new =2;
for k=1:n-1
  sin_theta = V(k+1,k_old)/U(k,k_old);
  cos_theta = sqrt(1.0-sin_theta^2);
  U(k+1:n,k_new) = ( U(k:n-1,k_old) -   \
      sin_theta*V(k+1:n,k_old))/cos_theta;
  V(k+1:n,k_new) = ( V(k+1:n,k_old) -   \
      sin_theta*U(k:n-1,k_old))/cos_theta;

  %--- Update determinant
  lndeterminant += log(U(k+1,k_new));

  %--- Extend back-substitution
  A(k+1,:)=1/U(k+1,k_new)* \
          (H(k+1,:)-dummyH(k+1,:));
  y(k+1,1)=1/U(k+1,k_new)* \
          (x(k+1,1)-dummyx(k+1,1));
  dummyH(k+2:n,:)+=U(k+2:n,k_new)*A(k+1,:);
  dummyx(k+2:n,1)+=U(k+2:n,k_new)*y(k+1,1);

  k_old = 3-k_old;
  k_new = 3-k_new;
end

lndeterminant *= 2.0;  % remember C=U'*U !!
```

For example, assume one has the following observations **x** and design matrix **H**:

$$\mathbf{x} = \begin{pmatrix} 3.5 \\ 1.1 \\ 0.4 \end{pmatrix} \qquad \mathbf{H} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \quad (38)$$

The first-differenced observations and design matrix are:

$$\Delta\mathbf{x} = \begin{pmatrix} -2.4 \\ -0.7 \end{pmatrix} \qquad \Delta\mathbf{H} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (39)$$

Assuming, for example, that $\alpha = 1$, $\sigma_{pl} = 0.7$ and $\sigma_w = 1.4$, one gets using Eqs. (21) and (23) for the covariance matrix:

$$\gamma = \begin{pmatrix} 4.543 \\ 2.168 \end{pmatrix} \qquad \mathbf{C} = \begin{pmatrix} 4.543 & -2.168 \\ -2.168 & 4.543 \end{pmatrix} \quad (40)$$

Inserting $\Delta\mathbf{x}$, $\Delta\mathbf{H}$ and γ into the above algorithm, one obtains $\ln(\det(\mathbf{C})) = 2.7693$ and:

$$\mathbf{y} = \begin{pmatrix} -1.126 \\ -0.985 \end{pmatrix} \qquad \mathbf{A} = \begin{pmatrix} 0.469 \\ 0.788 \end{pmatrix} \quad (41)$$

One can verify that Eq. (24) holds:

$$\left(\Delta\mathbf{H}^T\mathbf{C}^{-1}\Delta\mathbf{H}\right)^{-1}\Delta\mathbf{H}^T\mathbf{C}^{-1}\Delta\mathbf{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y} = -1.55$$

$$(42)$$

References

Abramowitz M, Stegun IA (1965) Handbook of mathematical functions with formulas, graphs and mathematical tables. National Bureau of standards applied mathematics series 55. US Government Printing Office, Washington DC

Agnew DC (1992) The time-domain behaviour of power-law noises. Geophys Res Lett 19(4):333–336

Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A, Sorensen D (1999) LAPACK users' guide, 3rd edn. Society for Industrial and Applied Mathematics, Philadelphia

Bojanczyk AW, Brent RP, de Hoog FR, Sweet DR (1995) On the stability of the Bareiss and related Toeplitz factorization algorithms. SIAM J Matrix Anal Appl 16(1):40–57

Hosking JRM (1981) Fractional differencing. Biometrika 68(1):165–176

Johnson HO, Agnew DC (1995) Monument motion and measurements of crustal velocities. Geophys Res Lett 22(21):2905–2908

Kasdin NJ (1995) Discrete simulation of colored noise and stochastic processes and $1/f^\alpha$ power law noise generation. Proc IEEE 83(5):802–827

Langbein J (2004) Noise in two-color electronic distance meter measurements revisited. J Geophys Res 109, B04406, doi:10.1029/2003JB002819

Langbein J, Johnson H (1997) Correlated errors in geodetic time series: Implications for time-dependent deformation. J Geophys Res 102(B1):591–603

Mao A, Harrison CGA, Dixon TH (1999) Noise in GPS coordinate time series. J Geophys Res 104(B2):2797–2816

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical recipes in C. 2nd edn. University Press, Cambridge

Williams SDP (2003) The effect of coloured noise on the uncertainties of rates from geodetic time series. J Geod 76(9–10):483–494, doi:10.1007/s00190-002-0283-4

Williams SDP, Bock Y, Fang P, Jamason P, Nikolaidis RM, Prawirodirdjo L, Miller M, Johnson DJ (2004) Error analysis of continuous GPS position time series. J Geophys Res 109, B03412, doi:10.1029/2003JB002741

Zhang J, Bock Y, Johnson H, Fang P, Williams S, Genrich J, Wdowinski S, Behr J (1997) Southern California permanent GPS geodetic array: error analysis of daily position estimates and site velocities. J Geophys Res 102(B8):18035–18055