# An Evaluation of a Monte Carlo Markov Chain Method for the Statistical Analysis of GPS Time Series


**G. Olivares[(1)], F.N. Teferle[(1)], S.D.P. Williams[(2)]**

1) Geophysics Laboratory, University of Luxembourg, Luxembourg

2) National Oceanography Center

Contact: german.olivares@uni.lu


UNIVERSITÉ DU LUXEMBOURG

National Oceanography Centre

NATURAL ENVIRONMENT RESEARCH COUNCIL


## Abstract

**One of the objectives of TIGA is to compute precise station coordinates and velocities for GPS stations of interest. Consequently, a comprehensive knowledge of the stochastic features of the GPS time series noise is crucial, as it affects the velocity estimation for each GPS station. For that, we present a Monte Carlo Markov Chain (MCMC) method to simultaneously estimate the velocities and the stochastic parameters of the noise in GPS time series. This method allows to get a sample of the likelihood function and thereby, using Monte Carlo integration, all parameters and their uncertainties are estimated simultaneously.**

**We propose this method as an alternative to optimization methods, such as the Maximum Likelihood Estimator (MLE) method implemented in the widely used CATS software, whenever the likelihood and the parameters of the noise are to be estimated in order to obtain more realistic uncertainties for all parameters involved. Furthermore, we assess the MCMC method through comparison with the widely used CATS software using daily height time series from the British Isles continuous GNSS Facility (BIGF) level 2 products.**


## Introduction

We have analyzed 10 stations from the BIGF Level 2 products with CATS[1] and MCMC[2] software in order to obtain parameters such as the rate, the intercept and the amplitudes of annual and semi-annual signals. In both analysis it has been assumed that the noise signal is composed by time-correlated (coloured) and random (white) noise. We have estimated the spectral index and the power amplitudes of both noise components simultaneously with the aforementioned parameters.

Both methods estimate similar rates but as MCMC estimates the uncertainty associated to the spectral index (whereas CATS does not) the uncertainties for the parameters are different.

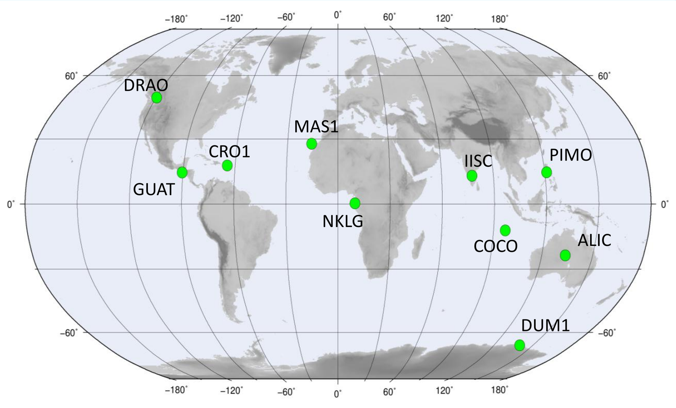


*Figure 1. Selected GPS stations from the BIGF.*

### Maximum Likelihood Estimator

The likelihood of getting the observational data given some parameters is defined as:

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

- $y$ : Observational data
- $\hat{y}$ : Model data
- $\theta$ : Parameter
- $C$ : Covariance matrix

The parameters that better fit the data are estimated by computing the maximum of the likelihood:

$$\hat{\theta} \equiv \arg\ max[L]$$

Very often there is no closed-form formula for the Likelihood function and numerical computation is needed.

For that purpose CATS has been developed (Williams 2008)..

### Monte Carlo Markov Chain (MCMC)

A Metropolis-Hasting algorithm is used to get a sample of the a posteriori distribution that, according to Bayes Theorem, is related to the Likelihood:

$$P(\hat{\theta}|y) = \frac{L(y|\hat{\theta})P(\hat{\theta})}{P(y)}$$

where, $P(\hat{\theta})$ and $P(y)$ are the a priori distributions of the estimated parameters and the data. Concerning the parameters we have chosen an uniform distribution for them, whereas it is not necessary to know $P(y)$ for our algorithm. Thus the MCMC method provides histograms for all parameters (including the spectral index) and cross-correlation between these can be easily visualized with 2D histograms. An example is given in Fig. 2 where histograms for the spectral index (Fig. 2a) and the power amplitude (Fig. 2b) are shown for the GPS station FRAE in the UK. For the same station, Fig. 2c. shows how 2D histograms help to detect cross-correlations (for that case between coloured and white noise power amplitudes).

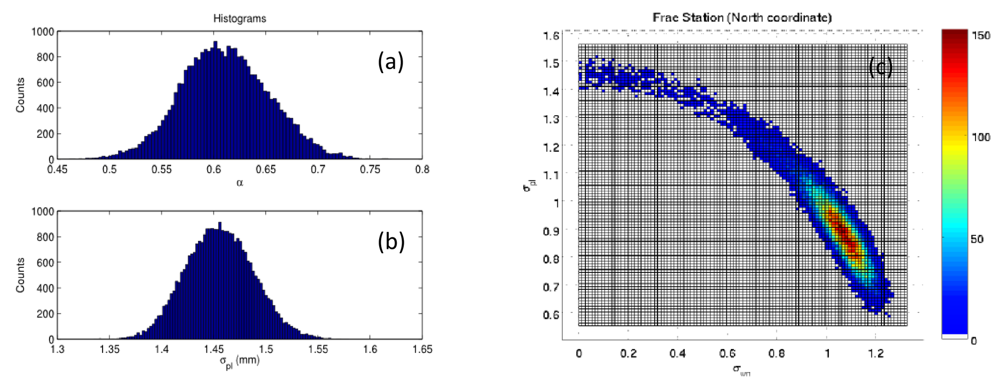


*Figure 2. Histograms for spectral index (a), the power amplitude (b) and 2D Histogram (c) for coloured noise ($\sigma_{pl}$) and white noise ($\sigma_{wn}$) amplitudes ) for GPS station FRAE.*

### The MCMC algorithm

1) Compute the likelihood at $\theta_i$, i.e. $L_i \equiv L(y|\theta_i)$.
**2)** Take a random step in parameter space to obtain a new value for the parameter $\theta_{i+1}$. The probability distribution of the step is taken to be Gaussian centered over $\theta_i$ with variance $\sigma_\theta$.
**3)** Compute $L_{i+1}$.
**4)** If $L_{i+1}/L_i > 1$, take the step, i.e. save the new set of parameters $\theta_{i+1}$ and $L_{i+1}$ as part of the chain, then go to step 2 after the substitution $\theta_i \rightarrow \theta_{i+1}$.
**5)** If $L_{i+1}/L_i < 1$, draw a random number $x \in [0,1]$ from a uniform distribution. If $L_{i+1}/L_i < x$, do not take the step, i.e. keep the previous parameter value as part of the chain and return to step 2. If $L_{i+1}/L_i > x$, take the step, i.e. do as in **4**.

Rule 4 leads the Markov chain towards the maximum of the Likeilihood (and the a posteriori distribution) . In order to get a sample of it we allow the chain to explore the region around the maximum by including rule 5. Thus, once a sample of $P(\hat{\theta}|y)$ is obtained, we proceed to estimate the parameters by means of Monte Carlo integration.

### Methodology:

For all stations we have considered the Up coordinate only.

We have estimated the following parameters:

- The spectral index $\alpha$.
- The power amplitude for the colored noise $\sigma_{pl}$ and the white noise $\sigma_{wn}$.
- The rate $v$.
- The intercept

As all time series are at least eight years in length, we do not model any periodic terms, as the effect on the rate estimates is negligible (Blewitt and Lavallee, 2002; Bos et al., 2010).

The estimate for the parameter $\theta_i$ is obtained by Monte Carlo integration, i.e.

$$\hat{\theta}_i = \int P(\theta|y)\, \theta_i\, d\theta = \frac{\sum_{t=1}^{M} \theta_{t,i}}{M}$$

where the sum is along the Markov chain and $M$ stands for its length.

### The noise analysis

Following Zhang et al. (1997) and Williams (2004, 2008) , we have considered the covariance of the signal to be a combination of white noise and coloured noise (a.k.a. power-law noise)

$$Cov = \sigma_{wn} I + \sigma_{pl} LL'$$

Where $\sigma_{wn}$ and $\sigma_{pl}$ are the power amplitudes for the white noise and the power law noise, respectively; $I$ is the identity matrix and $L$ is a lower triangular matrix in the form of:

$$L_{ij} = \begin{cases} \frac{(i-j+\alpha/2+1)!}{(i-j)!\,(\alpha/2-1)!}, \forall (i-j) \geq 0 \\ 0, \forall (i-j) < 0 \end{cases}$$

### Preliminary results:

Fig. 3 shows the differences between the estimates obtained from CATS and MCMC for the spectral index (Fig. 3a), the relative differences for the power amplitude (Fig. 3b), the rate of the GPS station (Fig. 3c) and its uncertainty (Fig. 3d).

Fig. 3a shows that the MCMC estimate for the spectral index estimate is higher than that provided by CATS for all GPS stations but MAS1. Note that for DRAO the large difference stems from CATS absorbing the white noise component into the power-law noise, thereby yielding a lower spectral index for the power-law noise.

Fig. 3b shows that the power amplitude for the coloured noise is also higher than the counterpart provided by CATS. Namely, the relative difference of the power amplitude for DRAO is around 40%. Figs. 3a and 3b suggest that the larger the difference in the spectral index estimates, the larger the difference in the power amplitude estimates.

Although not shown, the rate estimates agree on the sub-millimetre per year level with the largest rate difference of 0.9 mm/yr for IISC. Fig. 3c shows that the range for the relative difference in the rate estimates is quite large, going from 0.3% (MAS1) to 227% (IISC) and 230% (GUAT). Note that the absolute rate difference for GUAT is 0.3 mm/yr. For DRAO, the station for which CATS absorbs the white noise into the coloured noise component, the estimated rate according to MCMC is 60% larger than the one obtained with CATS.

*Figure 3. Differences of spectral index estimates (a), relative difference of power amplitude (b), rate (c) and rate uncertainty.*

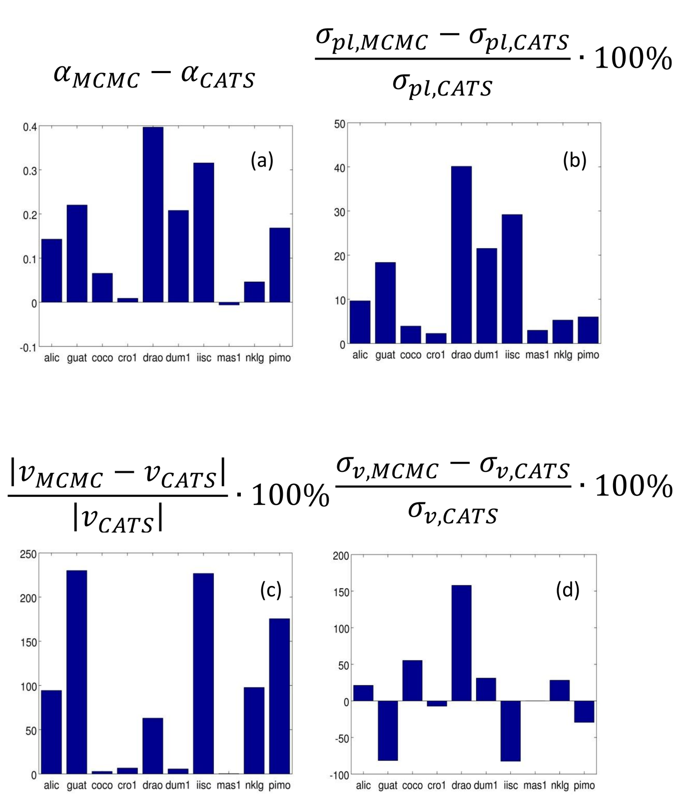


*Figure 4. Time series of the Up coordinate for GPS station IIGS (a) and GUAT (b) with the linear fit models from CATS (green) and MCMC (red). The vertical black line marks at epoch 2008.6 the offset.*

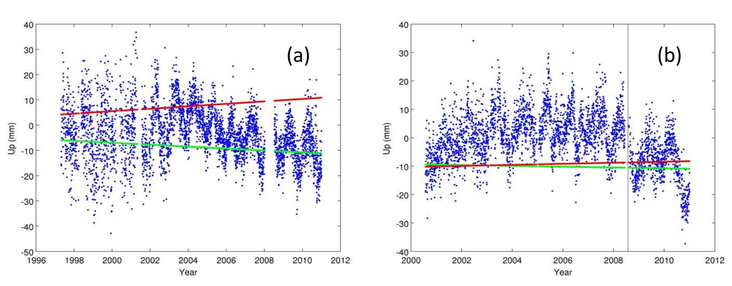


Fig. 3d clearly shows that absorbing the white noise into the coloured noise component within the time series leads to an underestimated rate uncertainty. Indeed, for DRAO, the uncertainty computed by MCMC is almost 160% higher. In Fig. 3d the rate uncertainties for GUAT and IISC estimated by CATS are clearly larger (>200%) than those from MCMC. These differences between the MCMC and CATS estimates in GUAT and IISC stem from, firstly, the fact that their time series show non-linear behaviour (Fig. 4a), and secondly, from inadequate modelling of the time series (i.e. not accounting for offsets) (Fig. 4b). For IISC, according to CATS, the rate is -0.38 mm/yr, whereas, from MCMC, it is 0.42 mm/yr (Fig. 4a). Fig. 4b shows how an offset (vertical black line at epoch 2008.6), that was not taken into account, leads to a discrepancy between the rate estimates. Based on these preliminary results, the MCMC and CATS methods agree very well for time series with linear behaviour and when correctly modelled.


## References

1. S.D.P. Williams, "CATS: GPS coordinate time series anaylsis software", GPS Solutions, no. 12, pp. 147–153 (2008).
2. G. Olivares , N.F. Teferle, "A Monte Carlo Markov Chain Method for Parameter Estimation of Fractional Differenced Gaussian Processes", submitted to Transactions on Signal Processing.
3. J. Zhang, Y. Bock, H. Johnson, 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" Journal of Geophysical Research, Vol. 102, No. B8, pp. 18,035-18,055 (1997).
4. S.D.P. Williams, Y. Bock, P. Fang, P. Jamason, R.M. Nikolaidis, L. Prawirodirdjo, M. Miller and D.J. Johnson, Journal of Geophysical Research, Vol. 109, B03412 (2004).



## Acknowledgements

The authors would like to express their gratitude to the British Isles continuous GNSS Facility (BIGF) for providing the station coordinate time series (Level 2 product).


## Conclusions

The preliminary results of this study suggest:

1. MCMC and CATS generally give rate estimates that are in good agreement at the sub-millimetre per year level.
2. Compared to MCMC, CATS underestimates the spectral indices and the coloured power amplitudes for all the stations but MAS1.
3. For DRAO, for which the white noise component was absorbed into the coloured noise component by CATS, MCMC provides larger rate uncertainties.
4. The biggest disagreements between both methods stem from cases where time series showed non-linear behaviour (IISC) or were mis-modelled (GUAT).


***Presented at:*** *The IGS Workshop 2012, Olsztyn, Poland, July 23-27, 2012*


UNIVERSITÉ DU LUXEMBOURG