Unbiased characterization of noise in geodetic data

Trever T. Hines · Eric A. Hetland

Abstract Geodetic time series contain temporally correlated noise that must be quantified before the data can be used to make geophysical inferences. If the noise is not accurately quantified then there is a risk of underestimating the uncertainties on inferred geophysical parameters. The maximum likelihood estimation (MLE) method is commonly used to characterize noise in geodetic time series; however, this method is known to be biased. Specifically, the MLE method has a tendency to underestimate the amplitude of random walk noise. This bias is most pronounced when estimating the noise in shorter time series. We discuss an unbiased alternative to the MLE method, which is known as the restricted maximum likelihood (REML) method. We use synthetic tests to demonstrate that the REML method does not suffer from the bias inherent in the MLE method. Since the computational costs of the REML and MLE method are nearly equivalent, the REML method should always be prefered over the MLE method when quantifying noise in geodetic data.

1 Introduction

Before geodetic data can be used to make geophysical inferences, it is necessary to have an accurate noise model. Here we consider noise to be any observed deformation that is not representative of the cohesive crustal block. The noise in geodetic

Trever T. Hines
Department of Earth and Environmental Sciences
University of Michigan, Ann Arbor
Ann Arbor, MI, USA
E-mail: hinest@umich.edu

Eric A. Hetland
Department of Earth and Environmental Sciences
University of Michigan, Ann Arbor
Ann Arbor, MI, USA
E-mail: ehetland@umich.edu

time series is temporally correlated and its power spectrum can often be described by a power law relationship (Agnew, 1992),

$$P(f) = P_o f^{-k}, \tag{1}$$

where f is frequency, k is the spectral index and P_o is the noise amplitude. For data recorded by strain and tilt meters (Wyatt, 1982, 1989) and electronic distance measuring (EDM) instruments (Langbein and Johnson, 1997), the temporally correlated noise can be modeled as a random walk ($k = 2$). In these studies, the random walk noise was attributed to unstable geodetic monuments. Global Positioning System (GPS) data, which is prone to additional non-physical sources of error, often has temporally correlated noise that is best modeled as flicker noise ($k = 1$) (Zhang et al, 1997; Mao et al, 1999; Williams et al, 2004). However, the most appropriate model can vary between stations. Generally, the noise in GPS data is best described as white noise plus some combination of random walk and flicker noise (Langbein, 2008).

If temporally correlated noise is mismodeled or ignored then the uncertainties in geophysical parameters inferred from geodetic data may be underestimated (Zhang et al, 1997). Since no single noise model is universally appropriate, it may be preferable to determine a noise model for each station before attempting to study any underlying signal. Langbein and Johnson (1997) introduced a maximum likelihood estimation (MLE) method to determine the hyperparameters (e.g., P_o and k) that best characterize the noise in geodetic time series. Furthermore, the MLE method can be used to discern which type of stochastic process (e.g., power law or Gauss-Markov) is most appropriate (Langbein, 2004). There are other methods for determining noise models, such as the least squares variance component estimation method (Amiri-Simkooei et al, 2007) and the network noise estimator (Dmitrieva et al, 2015). However, the MLE method from Langbein and Johnson (1997) is the most widely used (e.g., Langbein, 2004, 2008; Zhang et al, 1997; Mao et al, 1999; Williams et al, 2004; Hill et al, 2009; King and Williams, 2009; Murray and Svarc, 2017).

One deficiency with the MLE method, which was recognized by Langbein and Johnson (1997), is that it can be biased towards underestimating the amplitude of random walk noise. The MLE method is biased because it assumes that residual geodetic time series, which had geophysical signals estimated and removed, are representative samples of noise. This is not always a fair assumption because estimating and removing geophysical signals will inevitably also remove low frequency components of noise. Langbein (2012) further explored the bias in the MLE method and how it propagates into the uncertainties for estimated tectonic rates of deformation. They demonstrated with synthetic data, consisting of white noise and random walk noise, that the bias is stronger for shorter time series. Langbein (2012) emphasized the role of the crossover frequency, f_c, which is the frequency where the power of the white noise is equal to the power of the random walk noise. They suggested that a time series should be about 10 times longer than f_c^{-1} in order to accurately quantify its random walk noise.

In this paper we discuss an alternative to the MLE method, which is known as the restricted maximum likelihood (REML) method. We use synthetic tests to demonstrate that the REML method does not suffer from the bias that is inherent in the MLE method. With the REML method, we can accurately quantify

the random walk noise in time series that are as short as f_c^{-1}. Furthermore, the REML and MLE methods have practically equivalent computational costs. For these reasons, we argue that there is no reason to prefer using the MLE method over the unbiased REML method.

2 Maximum likelihood methods

In this section we briefly describe the MLE method and explain why it is biased. We then provide a description of the REML method. Let $\mathbf{d}_*$ denote a column vector of n observations. We treat $\mathbf{d}_*$ as a realization of the random vector

$$\mathbf{d} = \mathbf{Gm} + \epsilon, \tag{2}$$

where ϵ is the data noise vector, $\mathbf{G}$ is an $n \times m$ matrix with linearly independent columns that are used to describe geophysical signal in $\mathbf{d}$ (e.g., secular rates, coseismic offsets, postseismic transience), and $\mathbf{m}$ is a column vector of m model parameters which have uninformative priors (i.e. $\mathbf{m} \sim \mathcal{N}(\mathbf{0}, \lambda \mathbf{I})$ in the limit as $\lambda \to \infty$). We assume that the data noise can be described as $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\theta))$, where θ are the hyperparameters which we want to estimate appropriate values for. If we had chosen an informed prior for $\mathbf{m}$, we would select θ such that the probability of drawing $\mathbf{d}_*$ from $\mathbf{d}$, $p_\mathbf{d}(\mathbf{d}_*|\theta)$, is maximized. However, the uninformed prior on $\mathbf{m}$ makes $\mathbf{d}$ improper and $p_\mathbf{d}$ is zero for all choices of θ. We must instead seek an alternative likelihood function to maximize.

The MLE method chooses θ such that the probability of sampling the least squares residual vector,

$$\mathbf{r}_* = \left(\mathbf{I} - \mathbf{G} \left(\mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G} \right)^{-1} \mathbf{G}^T \mathbf{\Sigma}^{-1} \right) \mathbf{d}_*, \tag{3}$$

from ϵ is maximized. To put it explicitly, The MLE method maximizes the probability density function

$$p_\epsilon(\mathbf{r}_*|\theta) = \left(\frac{1}{(2\pi)^n |\mathbf{\Sigma}(\theta)|} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \mathbf{d}_*^T \mathbf{K}(\theta) \mathbf{d}_*} \tag{4}$$

with respect to θ, where

$$\mathbf{K} = \mathbf{\Sigma}(\theta)^{-1} - \mathbf{\Sigma}(\theta)^{-1} \mathbf{G} \left(\mathbf{G}^T \mathbf{\Sigma}(\theta)^{-1} \mathbf{G} \right)^{-1} \mathbf{G}^T \mathbf{\Sigma}(\theta)^{-1}. \tag{5}$$

Implementations of the MLE method typically maximize the logarithm of eq. (4) with the downhill simplex method (Press et al, 2007). It is important to recognize that the MLE method assumes that $\mathbf{r}_*$ is a representative sample of ϵ. This assumption is only valid when n is sufficiently large. To elaborate, we note that $\mathbf{r}_*$ is a sample of the random variable

$$\mathbf{r} = \left(\mathbf{I} - \mathbf{G} \left(\mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G} \right)^{-1} \mathbf{G}^T \mathbf{\Sigma}^{-1} \right) \mathbf{d}, \tag{6}$$

which is distributed as

$$\mathbf{r} \sim \mathcal{N} \left(\mathbf{0}, \mathbf{\Sigma} - \mathbf{G} \left(\mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G} \right)^{-1} \mathbf{G}^T \right). \tag{7}$$

The term being subtracted in eq. (7) is the covariance of the least squares prediction vector, which will typically get smaller as n increases. The distribution of $\mathbf{r}$ will then tend towards that of ϵ as n increases. Hence, we can only assume that $\mathbf{r}_*$ is a representative sample of ϵ when n is sufficiently large. We can also observe from eq. (7) that the variance of $\mathbf{r}$ will always be less than the variance of ϵ. This is the reason why the MLE method is biased towards underestimating the noise in short time series.

Having demonstrated that the MLE method is biased, we move on to discuss the REML method for selecting θ. The REML method was introduced by Patterson and Thompson (1971), and is now established in the Kriging community as an unbiased method for estimating covariance functions (e.g., Cressie, 1993). The REML method can be understood by first considering an $(n - m) \times n$ matrix $\mathbf{R}$ which satisfies $\mathbf{R}\mathbf{G} = \mathbf{0}$. We then consider the random variable $\mathbf{x} = \mathbf{R}\mathbf{d}$, which can be interpreted as the projection of $\mathbf{d}$ onto a subspace that is orthogonal to the columns of $\mathbf{G}$. As opposed to $\mathbf{d}$, $\mathbf{x}$ is a proper random variable since it is independent of the prior on $\mathbf{m}$. The REML method chooses θ such that the probability of drawing $\mathbf{x}_* = \mathbf{R}\mathbf{d}_*$ from $\mathbf{x}$, $p_{\mathbf{x}}(\mathbf{x}_*|\theta)$, is maximized. As noted by Harville (1974), the particular choice for $\mathbf{R}$ does not matter because it will only change the likelihood function which we are maximizing by a scale factor. Following Harville (1974), we let $\mathbf{R}$ have the properties $\mathbf{R}^T\mathbf{R} = \mathbf{I} - \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$ and $\mathbf{R}\mathbf{R}^T = \mathbf{I}$. The probability density function for $\mathbf{x}$ can then be written as

$$p_{\mathbf{x}}(\mathbf{x}_*|\theta) = \left(\frac{\left|\mathbf{G}^T\mathbf{G}\right|}{(2\pi)^{n-m}\left|\mathbf{\Sigma}(\theta)\right|\left|\mathbf{G}^T\mathbf{\Sigma}(\theta)^{-1}\mathbf{G}\right|} \right)^{\frac{1}{2}} e^{-\frac{1}{2}\mathbf{d}_*^T\mathbf{K}(\theta)\mathbf{d}_*}. \tag{8}$$

Note the similarity between eq. (8) and eq. (4). If programmed efficiently (see Appendix 1) and if $m \ll n$, the computational cost of the REML method is practically equivalent to that of the MLE method. What remains to be determined is whether the REML method remediates the bias in the MLE method. We demonstrate that this is indeed the case with a numerical test.

3 Synthetic demonstration

We compare the REML and MLE methods by using them to estimate hyperparameters from synthetic data. This demonstration is modeled after the demonstration from Langbein (2012) which highlights bias in the MLE method. Our synthetic noise is a combination of white and random walk noise, which has a power spectral density described by

$$P(f) = \frac{\sigma_{rw}^2}{2\pi^2 f^2} + 2\sigma_w^2 \Delta t, \tag{9}$$

where σ_{rw} and σ_w are hyperparameters for the random walk and white noise components, respectively. Δt is the sampling period, which is set at one day. The crossover frequency for the synthetic noise is then

$$f_c = \frac{1}{2\pi\sqrt{\Delta t}} \frac{\sigma_{rw}}{\sigma_w}. \tag{10}$$

In order to use the MLE or REML method, we must express the power law relationship in the frequency domain as a covariance matrix in the time domain. A

general procedure for doing so can be found in Langbein (2004). The components of the covariance matrix corresponding to eq. (9) can be concisely written as

$$\Sigma_{ij} = \sigma_{rw}^2 \min(i\Delta t, j\Delta t) + \sigma_w^2 \delta_{ij}, \tag{11}$$

where δ_{ij} is the Kronecker delta function. Similar to Langbein (2012), we set $\sigma_{rw} = 1.3$ mm/yr$^{0.5}$ and $\sigma_w = 1.1$ mm. We generate 5,000 synthetic noise time series, which each have a length of 2.5 yr.

We consider σ_w to be known, and we want to estimate σ_{rw} from the synthetic data. Although our synthetic data just consists of noise, we assume that the unknown underlying geophysical signal (i.e. **Gm**) consists of an offset plus a linear trend. We estimate σ_{rw} with the MLE and REML methods using varying lengths of the synthetic time series. The time series lengths range from 0.1 yr to 2.5 yr at 0.1 yr increments. The distribution of estimated σ_{rw} is shown in Figure 1.

The distribution of σ_{rw} estimated by the MLE method indicates that there is a bias towards underestimating σ_{rw} when the length of the time series is comparable to f_c^{-1}, which is 0.3 yr in this demonstration. The bias is appreciable when the time series is shorter than $\sim$1 yr, and estimates of σ_{rw} cluster around 0.0 when the time series is shorter than f_c^{-1}. The distribution tightens up around the true value and remains relatively constant for time series with length greater than $\sim$1 yr. This is consistent Langbein and Johnson (1997) who said that the time series should be at least 5 times greater than f_c^{-1} to get a good estimate of the random walk component. However, the mean and median of the distribution tends to be slightly less than the true value even when the full length of the time series is used.

In contrast, the REML method does significantly better at estimating σ_{rw}. For every time series length considered, the true value for σ_{rw} is within the 25-75 percentile of estimated σ_{rw}. For time series longer than f_c^{-1}, the mean and median of estimated σ_{rw} closely resembles the true value, indicating that the REML method is indeed unbiased. When the length of the time series is less than f_c^{-1}, the mean and median deviate from the true value and the variance of estimated σ_{rw} sharply increases. For such short time series, the random walk component cannot be resolved because it is being masked by the white noise.

4 Discussion and conclusion

The MLE method is the most commonly used method for quantifying noise in geodetic time series, despite the fact that it is known to be biased. The bias in the MLE method can result in underestimated uncertainties in geophysical parameters derived from geodetic time series (Langbein, 2012). The intention of this paper is to bring the REML method to light in the geodetic community as an unbiased alternative to the MLE method. Given the widespread usage of the MLE method, it may not be realistic to suggest that researchers abandon it in favor of the REML method. However, the REML method is nearly identical to the MLE method in terms of its computational cost and in terms of its implementation. Indeed, the only difference between the log likelihood functions being maximized by the REML and MLE methods is that the REML method includes two additional, easily computed, terms (See Appendix 1). We can therefore view the REML method as merely an unbiased correction to the MLE method.
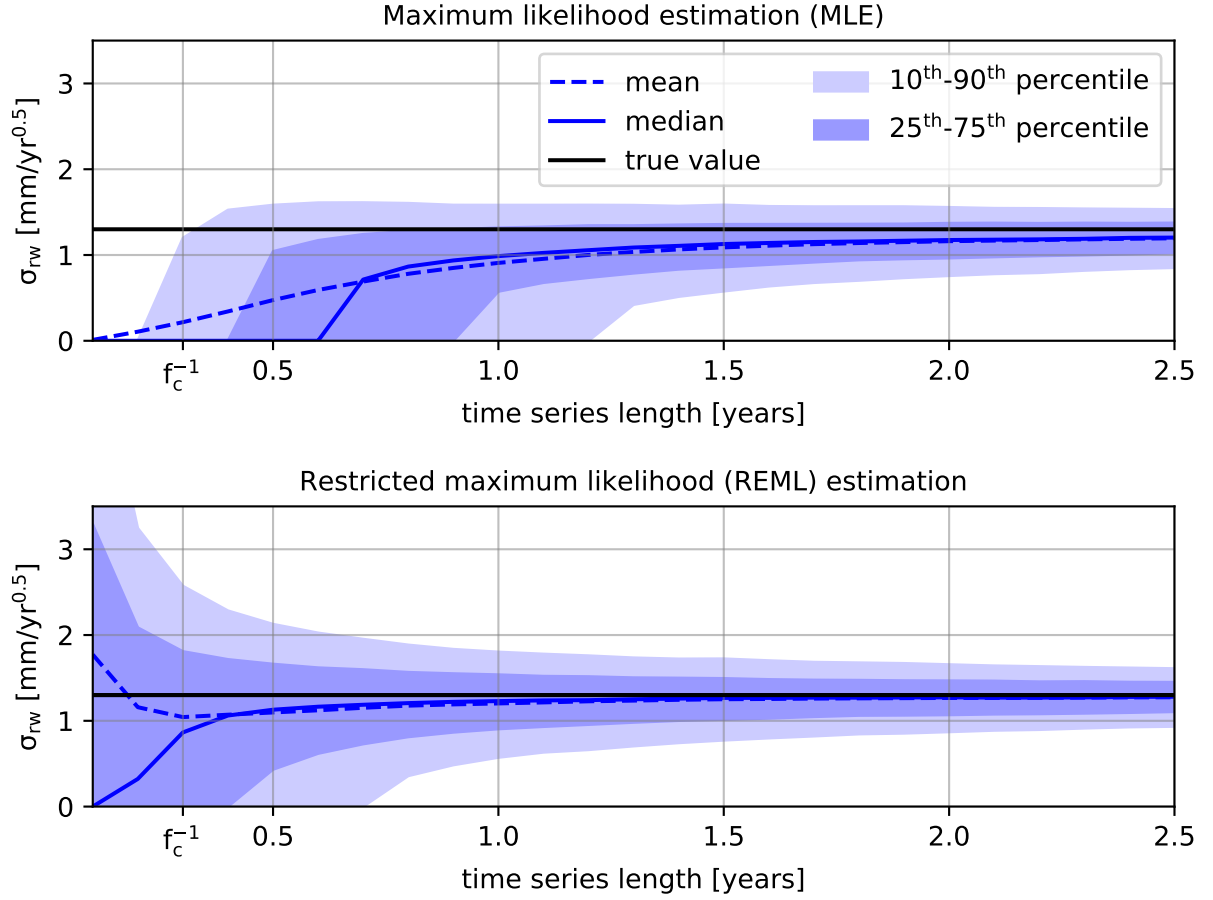
Fig. 1 Random walk amplitudes, σ_{rw}, estimated by the MLE and REML methods from synthetic data. The length of the synthetic time series used to estimate σ_{rw} is varied from 0.1 yr to 2.5 yr. The black line indicates the true random walk scale ($\sigma_{rw} = 1.3$), the light blue region shows the 10-90 percentile of estimates, the dark blue region shows the 25-75 percentile of estimates, the solid blue line indicates the median, and the dashed blue line indicates the mean.

In this paper, we have used synthetic tests to demonstrate that the REML method does not suffer from the bias inherent in the MLE method from Langbein and Johnson (1997). We also show that the REML method is able to characterize random walk noise in geodetic time series that are as short at f_c^{-1}. In contrast, the MLE method can only accurately quantify random walk noise for time series that are several times longer than f_c^{-1}. Furthermore, the MLE method and REML method have nearly identical computational costs. We believe that the REML method should always be preferred over the MLE method for quantifying noise in geodetic time series.

Appendix 1: REML algorithm

Algorithm 1 demonstrates how to efficiently compute the log of the REML likelihood function. For comparison, we include Algorithm 2, which evaluates the log of the MLE likelihood function. If we assume that $m \ll n$, then the main computational burden in both algorithms is computing the Cholesky decomposition of $\mathbf{\Sigma}$.

Algorithm 1 Function that takes $\mathbf{d}$, $\mathbf{\Sigma}$, and $\mathbf{G}$ as input and returns the logarithm of eq. (8). We use the notation $\mathbf{X}\backslash\mathbf{Z}$ to denote solving the system of equations $\mathbf{XY} = \mathbf{Z}$ for $\mathbf{Y}$.

---

function $reml\_log\_likelihood(\mathbf{d},\mathbf{\Sigma},\mathbf{G})$
$\quad \mathbf{A} \leftarrow cholesky(\mathbf{\Sigma})$
$\quad \mathbf{B} \leftarrow \mathbf{A}\backslash\mathbf{G}$
$\quad \mathbf{C} \leftarrow cholesky(\mathbf{B}^T\mathbf{B})$
$\quad \mathbf{D} \leftarrow cholesky(\mathbf{G}^T\mathbf{G})$
$\quad \mathbf{a} \leftarrow \mathbf{A}\backslash\mathbf{d}$
$\quad \mathbf{b} \leftarrow \mathbf{C}\backslash(\mathbf{B}^T\mathbf{a})$
$\quad$ **return** $\sum_i^m \log(D_{ii}) - \sum_i^n \log(A_{ii}) - \sum_i^m \log(C_{ii}) - \frac{1}{2}\mathbf{a}^T\mathbf{a} + \frac{1}{2}\mathbf{b}^T\mathbf{b} - \frac{n-m}{2}\log(2\pi)$
end function

Algorithm 2 Function that takes $\mathbf{d}$, $\mathbf{\Sigma}$, and $\mathbf{G}$ as input and returns the logarithm of eq. (4).

---

function $mle\_log\_likelihood(\mathbf{d},\mathbf{\Sigma},\mathbf{G})$
$\quad \mathbf{A} \leftarrow cholesky(\mathbf{\Sigma})$
$\quad \mathbf{B} \leftarrow \mathbf{A}\backslash\mathbf{G}$
$\quad \mathbf{C} \leftarrow cholesky(\mathbf{B}^T\mathbf{B})$
$\quad \mathbf{a} \leftarrow \mathbf{A}\backslash\mathbf{d}$
$\quad \mathbf{b} \leftarrow \mathbf{C}\backslash(\mathbf{B}^T\mathbf{a})$
$\quad$ **return** $-\sum_i^n \log(A_{ii}) - \frac{1}{2}\mathbf{a}^T\mathbf{a} + \frac{1}{2}\mathbf{b}^T\mathbf{b} - \frac{n}{2}\log(2\pi)$
end function

References

Agnew DC (1992) The time-domain behavior of power-law noises. Geophysical Research Letters 19

Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2007) Assessment of noise in GPS coordinate time series: Methodology and results. Journal of Geophysical Research: Solid Earth 112(7):1–19, DOI 10.1029/2006JB004913

Cressie N (1993) Statistics for Spatial Data, rev. ed. edn. John Wiley & Sons, New York, DOI 10.1111/j.1365-3121.1992.tb00605.x

Dmitrieva K, Segall P, DeMets C (2015) Network-based estimation of time-dependent noise in GPS position time series. Journal of Geodesy 89(6):591–606, DOI 10.1007/s00190-015-0801-9

Harville DA (1974) Bayesian Inference for Variance Components Using Only Error Contrasts. Biometrika 61(2):383–385

Hill EM, Davis JL, Elósegui P, Wernicke BP, Malikowski E, Niemi NA (2009) Characterization of site-specific GPS errors using a short-baseline network of braced monuments at Yucca Mountain, southern Nevada. Journal of Geophysical Research: Solid Earth 114(11):1–13, DOI 10.1029/2008JB006027

King MA, Williams SDP (2009) Apparent stability of GPS monumentation from short-baseline time series. Journal of Geophysical Research: Solid Earth 114(10):1–21, DOI 10.1029/2009JB006319

Langbein J (2004) Noise in two-color electronic distance meter measurements revisited. Journal of Geophysical Research B: Solid Earth 109(4):1–16, DOI 10.1029/2003JB002819

Langbein J (2008) Noise in GPS displacement measurements from Southern California and Southern Nevada. Journal of Geophysical Research: Solid Earth 113(5):1–12, DOI 10.1029/2007JB005247

Langbein J (2012) Estimating rate uncertainty with maximum likelihood: Differences between power-law and flicker-random-walk models. Journal of Geodesy 86(9):775–783, DOI 10.1007/s00190-012-0556-5

Langbein J, Johnson H (1997) Correlated errors in geodetic time series: Implications for time-dependent deformation. Journal of Geophysical Research 102(B1):591–603, DOI 10.1029/96JB02945

Mao A, Harrison GA, Dixon H (1999) Noise in GPS coordinate time series. Journal of Geophysical Research 104(B2):2797–2816

Murray JR, Svarc J (2017) Global Positioning System Data Collection , Processing , and Analysis Conducted by the U . S . Geological Survey Earthquake Hazards Program. Seismological Research Letters 88(3), DOI 10.1785/0220160204

Patterson HD, Thompson R (1971) Recovery of Inter-Block Information when Block Sizes are Unequal. Biometrika 58(3):545–554

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (2007) Numerical Recipes: The Art of Scientific Computing, 3rd edn. Cambridge University Press, Cambridge, DOI 10.1137/1031025

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. Journal of Geophysical Research-Solid Earth 109(B3):B03,412, DOI 10.1029/2003JB002741

Wyatt F (1982) Displacement of Surface Monuments: Horizontal Motion. Journal of Geophysical Research 87:979–989

Wyatt FK (1989) Displacement of surface monuments: Vertical motion. Journal of Geophysical Research 94(B2):1655–1664, DOI 10.1029/JB094iB02p01655

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. Journal of Geophysical Research 102(97)