

# Unbiased Noise Characterization for Geodetic Data

Trever T. Hines and Eric A. Hetland

March 27, 2017

## 1 Introduction

The noise in geodetic data, which we consider to be any observed deformation that is not representative of the cohesive crustal block, is temporally correlated. It is necessary to accurately quantify this noise before the data is used to make geophysical inferences. A power-law relationship is often used to describe the frequency content of noise in geodetic data (Agnew, 1992). The power spectral density of power-law noise is described as

$$P(f) = P_o f^{-n} \quad (1)$$

where  $f$  is frequency,  $n$  is the spectral index and  $P_o$  is the amplitude. Analysis of time series from strain and tilt meters (Wyatt, 1982, 1989), electronic distance measurements (Langbein and Johnson, 1997), and short-baseline GPS (King and Williams, 2009) indicates that temporally correlated noise can be attributed, at least in part, to an unstable geodetic monument. With the exception of King and Williams (2009), the referenced studies have found that localized motions of the monument can be modeled as a random walk process ( $n = 2$ ). GPS data, which is prone to additional non-physical sources of error (e.g. King and Watson, 2010), has been described as a combination of white noise ( $n = 0$ ) and flicker noise ( $n = 1$ ) (Zhang et al., 1997; Mao et al., 1999; Williams et al., 2004). Langbein (2008), who used longer GPS timeseries, found that GPS data in Southern California and Nevada contains white noise and some combination of flicker noise and random walk noise that varies between stations. It was suggested by Langbein (2008) that random walk noise in GPS data can be attributed to monument instability while flicker noise can be attributed to non-physical errors introduced in deriving the displacement timeseries.

No single noise model is universally appropriate for geodetic data, and the most rigorous of studies involving geodetic data should estimate a noise model for each station. Langbein and Johnson (1997) introduce a maximum likelihood (ML) method for determining the optimal values for  $P_o$  and  $n$  (or the hyperparameters for any other assumed model). This method has become the standard technique for characterizing noise in geodetic data (Langbein, 2004, 2008; Zhang et al., 1997; Mao et al., 1999; Williams et al., 2004; King and Williams, 2009; Murray and Svarc, 2017).

Recently, Langbein (2012) demonstrated with synthetic data consisting of white and colored noise that the ML method is biased towards inferring a small component of colored noise in short timeseries. The bias is appreciable when the length of the timeseries is less than or comparable to the cross-over period, which is the period at which the power of the colored noise exceeds the power of the white noise. Langbein (2012) claims that the colored noise tends to be underestimated because it is indistinguishable in the power spectra of the data. We do not find this explanation to be satisfying. If the length of the timeseries is comparable to or less than the cross-over period then we would expect the estimated colored noise to have a high variance but this does not explain why there is a bias. In this paper we explain why the ML method from Langbein and Johnson (1997) is biased, and we demonstrate that an alternative approach, known as the restricted maximum likelihood (REML) method, does not suffer from this bias and has practically no additional computational burden.

## 2 Maximum likelihood methods

Let  $\mathbf{d}_*$  denote a column vector of  $n$  observations. We treat  $\mathbf{d}_*$  as a realization of the random vector

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \epsilon, \quad (2)$$

where  $\epsilon$  is the data noise vector,  $\mathbf{G}$  is a  $n \times m$  matrix with linearly independent columns that are used to describe geophysical signal in  $\mathbf{d}$  (e.g. secular rates, coseismic offsets, postseismic transience, etc.), 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  $\lambda \rightarrow \infty$ ). We assume that the data noise can be described as  $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma(\theta))$ , where  $\theta$  are the hyperparameters which we want to find appropriate values for. Ideally, we would select  $\theta$  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  $\theta$ . We must instead seek an alternative likelihood function to maximize.

The ML method from Langbein and Johnson (1997) chooses  $\theta$  such that the probability of sampling the least squares residual vector,

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

from  $\epsilon$  is maximized. To put it explicitly, The ML method maximizes the probability density function

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

with respect to  $\theta$ , where

$$\mathbf{K} = \boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{\Sigma}(\theta)^{-1} \mathbf{G} (\mathbf{G}^T \boldsymbol{\Sigma}(\theta)^{-1} \mathbf{G})^{-1} \mathbf{G}^T \boldsymbol{\Sigma}(\theta)^{-1}. \quad (5)$$

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

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

which is distributed as

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

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  $\epsilon$  as  $n$  increases. Thus we can only assume that  $\mathbf{r}_*$  is a representative sample of  $\epsilon$  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  $\epsilon$ . We believe this is the reason why the ML method from Langbein and Johnson (1997) is biased towards underestimating the noise in short timeseries.

Having demonstrated that the ML method from Langbein and Johnson (1997) is biased, we move on to discuss the restricted maximum likelihood (REML) method for selecting  $\theta$ . 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 a  $(n - m) \times n$  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 mapping 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  $\theta$  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

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

Note the similarity between eq. (8) and eq. (4). If programmed efficiently (see Appendix A) and if  $m \ll n$ , the computational cost of the REML method is practically equivalent to that of the ML method from Langbein and Johnson (1997). What remains to be determined is whether the REML method remedies the bias in the ML method. We demonstrate that this is indeed the case with a numerical test.

### 3 Synthetic demonstration

We compare the REML method and the ML method from Langbein and Johnson (1997) by using the two methods to estimate hyperparameters from synthetic data. This demonstration is modeled after the demonstration from Langbein (2012) which highlights bias in the ML method. Our synthetic noise is a combination of white and random walk noise, which has a power spectral density described by

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

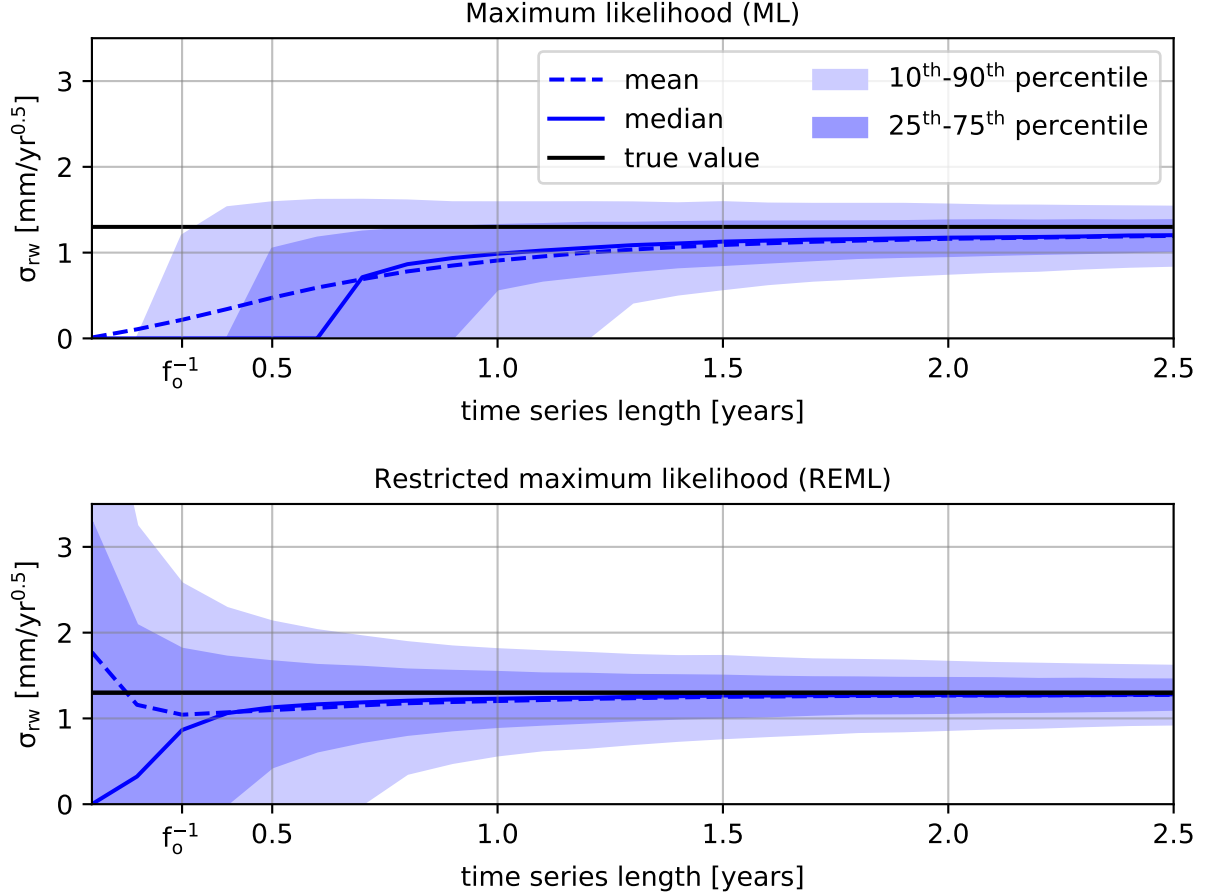


Figure 1: Comparison of ML and REML method over varying lengths of the synthetic timeseries. The black line indicates the true random walk scale ( $\sigma_{rw} = 1.3$ ), the light blue region shows the 10-90 percentile of estimates, the dark blue region shows the 25-75 percentile of estimates, the solid blue line indicates the median, and the dashed blue line indicates the mean. Note that  $f_o^{-1} \approx 0.3$ .

where  $\sigma_{rw}$  and  $\sigma_w$  are hyperparameters for the random walk and white noise components, respectively.  $\Delta t$  is the sampling period, which is set at one day. The cross-over frequency for the synthetic noise is then

$$f_o = \frac{1}{2\pi\sqrt{\Delta t}} \frac{\sigma_{rw}}{\sigma_w}. \quad (10)$$

In order to use the ML 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

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

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

We consider  $\sigma_w$  to be known, for the sake of simplicity, and we want to estimate  $\sigma_{rw}$  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 a constant and linear trend. We estimate  $\sigma_{rw}$  with the ML and REML methods using varying lengths of the synthetic timeseries. The timeseries lengths range from 0.1 yr to 2.5 yr at 0.1 yr increments. The distribution of estimated  $\sigma_{rw}$  is shown in Figure 1.

The distribution of  $\sigma_{rw}$  estimated by the ML method indicates that there is a bias towards underestimating  $\sigma_{rw}$  when the length of the timeseries is comparable to  $f_o^{-1}$ , which is 0.3 yr in this demonstration. The bias is appreciable when the timeseries is shorter than  $\sim 1$  yr, and estimates of  $\sigma_{rw}$  cluster around 0.0 when the timeseries is shorter than  $f_o^{-1}$ . The distribution tightens up around the true value and remains relatively constant for timeseries with length greater than  $\sim 1$  yr. This is consistent Langbein and Johnson (1997) who said that the timeseries should be at least 5 times greater than  $f_o^{-1}$  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 timeseries is used.

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

## 4 Discussion and conclusion

We have demonstrated that the restricted maximum likelihood (REML) method for characterizing noise does not suffer from the bias inherent in the maximum likelihood (ML) method from Langbein and Johnson (1997). Furthermore, the ML and REML methods have practically equivalent computational costs. We therefore argue that there is no reason to prefer the ML method over the REML method. Despite the fact that the REML method has been well established in the Kriging community and elsewhere (e.g. Wahba, 1985), it does not seem to be appreciated in the geodetic community. It is our intention to bring the REML method to light. We believe this is particularly necessary because the ML method has been frequently used in research and has been used in deriving public data products (e.g. Murray and Svarc, 2017).

We also emphasize that the REML method is able to accurately characterize colored noise in timeseries that are shorter than what is required for the ML method. Langbein and Johnson (1997) and Langbein (2012) have suggested that the length of a timeseries must be at least 5 to 10 times  $f_o^{-1}$  in order to accurately characterize colored noise. Based on our synthetic demonstration, the REML method remains accurate for time series as short as  $f_o^{-1}$ .

## 5 Appendix

(pseudo-code for REML)

## References

- Agnew, D. C. (1992). The time-domain behavior of power-law noises. *Geophysical Research Letters*, 19.
- Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, rev. ed. edition.
- Harville, D. A. (1974). Bayesian Inference for Variance Components Using Only Error Contrasts. *Biometrika*, 61(2):383–385.
- King, M. A. and Watson, C. S. (2010). Long GPS coordinate time series: Multipath and geometry effects. *Journal of Geophysical Research: Solid Earth*, 115(4):1–23.
- King, M. A. and Williams, S. D. P. (2009). Apparent stability of GPS monumentation from short-baseline time series. *Journal of Geophysical Research: Solid Earth*, 114(10):1–21.
- Langbein, J. (2004). Noise in two-color electronic distance meter measurements revisited. *Journal of Geophysical Research B: Solid Earth*, 109(4):1–16.
- Langbein, J. (2008). Noise in GPS displacement measurements from Southern California and Southern Nevada. *Journal of Geophysical Research: Solid Earth*, 113(5):1–12.
- 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.
- Langbein, J. and Johnson, H. (1997). Correlated errors in geodetic time series: Implications for time-dependent deformation. *Journal of Geophysical Research*, 102(B1):591–603.
- Mao, A., Harrison, G. A., and Dixon, H. (1999). Noise in GPS coordinate time series. *Journal of Geophysical Research*, 104(B2):2797–2816.
- Murray, J. R. and 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).

- Patterson, H. D. and Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58(3):545–554.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (2007). *Numerical Recipes: The Art of Scientific Computing*, volume 31. Cambridge University Press, Cambridge, 3 edition.
- Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Annals of Statistics*, 13(4):1378–1402.
- Williams, S. D. P., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R. M., Prawirodirdjo, L., Miller, M., and Johnson, D. J. (2004). Error analysis of continuous GPS position time series. *Journal of Geophysical Research-Solid Earth*, 109(B3):B03412.
- Wyatt, F. (1982). Displacement of Surface Monuments: Horizontal Motion. *Journal of Geophysical Research*, 87:979–989.
- Wyatt, F. K. (1989). Displacement of surface monuments: Vertical motion. *Journal of Geophysical Research*, 94(B2):1655–1664.
- Zhang, J., Bock, Y., Johnson, H., Fang, P., Williams, S., Genrich, J., Wdowinski, S., and Behr, J. (1997). Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities. *Journal of Geophysical Research*, 102(97).