Unbiased Noise Characterization for Geodetic Data

Trever T. Hines and Eric A. Hetland

March 20, 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 described the frequency content of noise in geodetic data (Agnew, 1992). The power spectral density of power-law noise is described as

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

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. The referenced studies have found that localized motions of the monument can be modeled as a random walk process ($n = 2$). Global or regional 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 provide a correction to the ML method which removes the bias without adding any significant computational burden.

2 Maximum likelihood methods

Let $\mathbf{d}_*$ denote a column vector of n observations made at times $\{t_1, t_2, \ldots, t_n\}$. 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 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 $\lambda \to \infty$). 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 find appropriate values for. Ideally, we would selects 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 ML method from Langbein and Johnson (1997) chooses θ such that the probability of sampling the least squares residual vector,

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

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

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

with respect to θ, where

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

The ML method implicitly 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

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

which is distributed as

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

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. Thus 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. 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 θ. The REML method was introduced by Patterson and Thompson (1971) and is now accepted in the Kriging literature 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 has the properties $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ and $\mathbf{R}\mathbf{G} = \mathbf{0}$. 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 then conveniently choose $\mathbf{R}$ to satisfy $\mathbf{R}^T\mathbf{R} = \mathbf{I} - \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$. The probability density function for $\mathbf{x}$ can then be written as

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

Note the similarity between eq. (8) and eq. (4). If programmed efficiently (see Appendix 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 remediates 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 the hyperparameters from snythetic 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

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

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 cross-over frequency for the synthetic noise noise is then

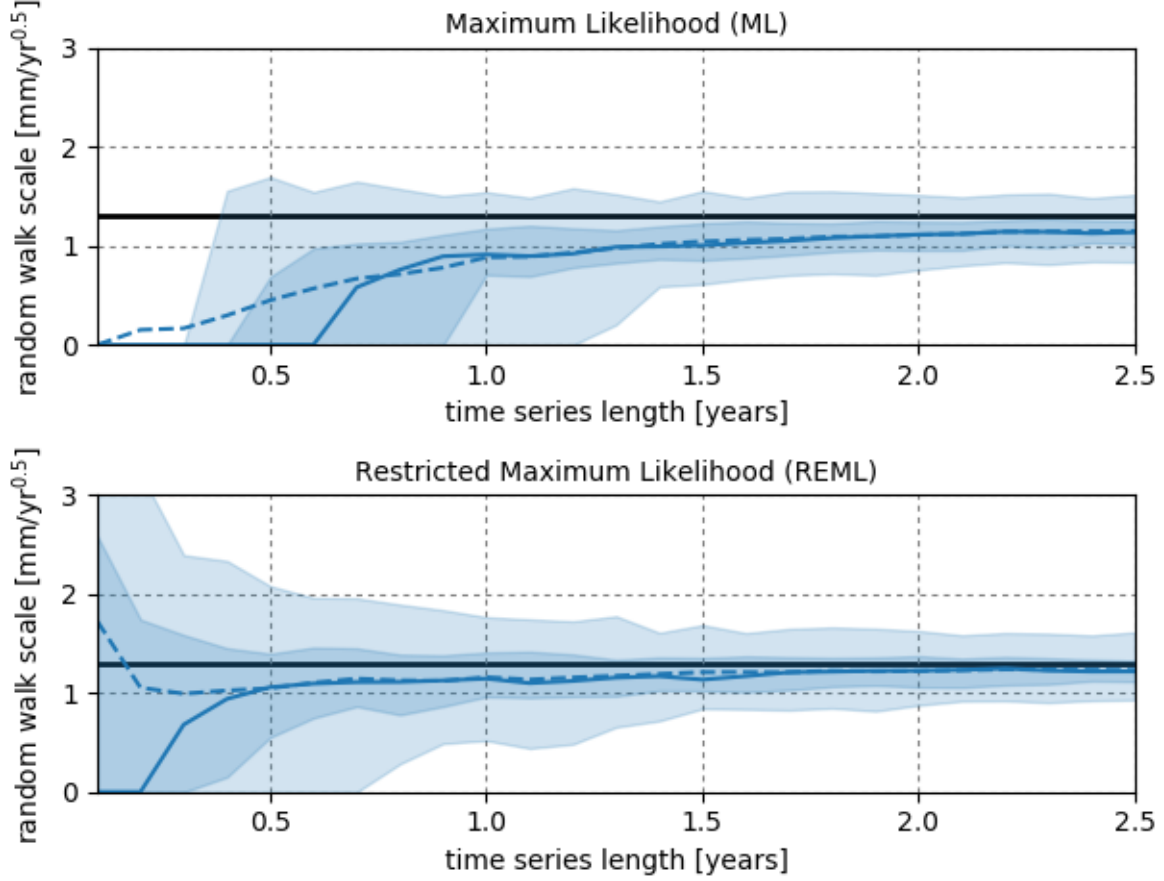$$f_o = \frac{1}{2\pi\sqrt{\Delta t}}\frac{\sigma_{rw}}{\sigma_w}. \tag{10}$$

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 30-70 percentile of estimates, the solid blue line indicates the median, and the dashed blue line indicates the mean. Note that $f_o^{-1} \approx 0.3$.

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

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

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 10,000 synthetic noise timeseries, which each have a length of 2.5 yr.

We consider σ_w to be known, for the sake of simplicity, and we just want to estimate σ_{rw} from the synthetic data. While 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 estimates σ_{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 σ_{rw} is shown in Figure 1.

We use the ML and REML methods to estimate σ_{rw}, and we vary the portion of the . considered known, we assume that the white noise component We then use the ML method and the REML method to estimate , which corresponds to $f_o^{-1} \approx 0.3$ yr.

We compare the two remains to be determined is whether the REML metho

We disagree with the heuristic that the length of the timeseries sould be 5 times larger than the cutoff period.

virtually equivalent to for will both equations can be evaluated with essentially equivalent computation cost. valuating the two expressions use In order to make this problem tractable, we assume some functional form for Σ. For example, if we assume that ϵ is Brownian motion then the components of Σ can be written as

$$\Sigma_{ij} = \sigma^2 \min(t_i, t_j). \tag{12}$$

Not

3

With this assumption we have converted to which converts the intractible problem of optimizing $\mathbf{C_d}$ into a more tractible problem of optimizing the hyperparameter σ.

can be described components of the covariance matrix can be if the $\mathbf{d}_*$ was observed at times $\mathbf{t} = t_1, t2, ..., t_n$, then we can

The MLE method,

f the data which we do not consider to be noise (which and its covariance, Cd, is what we want to estimate.

In fact, the bias in the MLE method is well recognized in the Kriging literature (e.g Cressie, 1993) and the correction that we propose has been known of since the work of Patterson and Thompson (1971). Nonetheless, we find

In this paper we an alternate explanation for the bias in the MLE method and we also

be high variance in estimated colored noise should have a high variance but not a bias. and we reasoning can how find this to be a satisfactory explanation for the bais

This is not a satisfying explanation bsatisfactory explanation for the bias, since it only explains why we would expect a high variance in estimated colored noise components.

Rather, this rationale would suggest that estimated components of colored noise would have a high variance not not biased. in

rather we find it as an explanation

the power spectral density is dominated by white noise and there will be little evidence for a colored noise component. Langbein (2012) reasoned that this bias can be noise power spectral density for such a short timeseries is dominated by the white noise, making the colored noise component indistinguishable. We do not find this explanation to be satifactory. If the absense of evidence for colored noise does not , explains why there should be a high variance in the estimating a smla

would be dominate by the white component masked by the white component

The rational being that the frequency content of any colored noise is white noise power and colored noise

the time series istowards inferring baised the esimates to showed the MLE method tends to estimate an for used many studies on noise analysis

rigorous studies on geodetic it should be common practice to estimate and optimal noise model. no clear model which

In To determine which noise model is most appropriate,

in who used longer GPS time series. Langbein (2008) showed that the amplitude of the power law noise and its spectral index can vary depending on the station location and the how the station was installed. Thus a rigorous study involving geodetic data would require that a noise model be determined for each station. would the most rigorous

The most common approach to techniques for determinng

There is not a generally accepted noise model for geodetic data, and the most rigorous geodetic analysis would require estimating a noise model for each geodetic time series.

, other tmodeling assumptions made in deriving the displacement time ich can have non-physical sources of error related to modeling assumptions in deriving the d, be contanimated sources of errors related to the GPS system, Studies of geodetic mon

can be derived from the geodetic data can be overly confident (e.g Mao et al., 1999). can be unduly cbiased is necessary to accurately quantify this noise using the data to make geophysical inferences

This can be attributed to localized motion of the geodetic monument (e.g. Wyatt, 1982, 1989; Agnew, 1992; King and Williams, 2009). For GPS data, there are several assumptions and models involved ?Langbein (2012) has suggested that non-physical, temporally correlated noise could be introduced by errors in deriving the displacement time series.

It is generally accepted that the frequency content of this noise can be accurately described with a power law relationship. There is no generally accepted spectral index, since that can vary depending on the location and the type of instrument being used. For example, short baseline instruments such as BSM or LSM, tend to have a random walk noise, while GPS data, which measures displacements over regional scales, tends to have a flicker noise. The amplitude of the noise can vary considerably between stations.

For the most rigorous of analysis, the noise model should be estimated for each station. Langbein and Johnson (1997) introduce a maximum likelihood method. THis method works by doing this

Here is the problem with it. It assumes residuals have a noise model.

THis paper merely reiterates a method which has been around since 1971. Given the widespead use of Langbeins method, we find it necessary to bring forth this method and demonstrate how it does not bias shit.

s tend borehole strain meters tend to have

The power law that best describes the geodetic noise can vary depending on the location of the monument and the geodetic technique. on by not The spectral index that best describes the noise varies depending on geodetic technique. For example, borehole strain meters has a XXX noise, while GPS data tends to be better described by flicker noise (Zhang et al., 1997; Mao et al., 1999). There noise also has regional While there seems

to be some consensus on the the spectral index, the amplitude of the noise can vary by station. For example in nevada there is nothing going on.

and the the ctends to be somewhere between 1 and 2 depending on the location (Langbein, 2008) , where (Mao et al., 1999)

By inspecting the frequency content of the noise, researchers have generally the power spectrum Models

could also be the result of errors in the GPS has suggested that non-physical te GPS

to the origin, Geodetic data contains temporally correlated noise. and it is necessary to accurately quantify this noise before can be used to draw

Before the data can be used to make inferences about geophysical phenomina, it is necessary to

If this noise is not accurately quantified, then it can then is and it is necessary to accurately quantify this noise before the data can be used for geophysical. It is necessary to accurately quantify this noise in order to make

(Langbein and Johnson, 1997; Zhang et al., 1997; Mao et al., 1999). This is often attributed to localized motion of the geodetic monument (e.g Wyatt, 1982, 1989; Agnew, 1992; King and Williams, 2009). GPS data may suffer from additional, non-physical sources of temporally correlated noise which are introduced in deriving the displacement time series.

It is important for us to first define what we consider to be noise. We consider noise to be any measured displacements that are localized or seasonal. of displacement that are not representative of the cohesive crust. That is, measurements that are localized.

As suggested by (Langbein, 2008, 2012), some of the temporally correlated noise could be a non-physical result of errors in deriving the GPS time series. and a the result errors introduced in deriving the GPS processing

some of the temporal correlation could be th and, for GPS data, . and nonphysical errors introduced in deriving the GPS displacement time series. intrumentation error correlated noise in geodetic data has been It is now widely recognized that geodetic data contains

Many people have used the MLE method by Langbein 1997,

fuck (Hines and Hetland, 2016)

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.

Hines, T. T. and Hetland, E. A. (2016). Rapid and simultaneous estimation of fault slip and heterogeneous lithospheric viscosity from post-seismic deformation. *Geophysical Journal International*, 204(1):569–582.

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.

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).