On the nature of GPS draconitic year periodic pattern in multivariate position time series

A. R. Amiri-Simkooei1,2

Received 9 November 2012; revised 12 April 2013; accepted 21 April 2013; published 22 May 2013.

Plate tectonics studies using GPS require proper analysis of time series, in which all functional effects are understood and all stochastic effects are captured using an appropriate noise assessment technique. Both issues are addressed in this contribution. Estimates of spatial correlation, time correlated noise, and multivariate power spectrum for daily position time series of 350, 150, and 50 permanent GPS stations, respectively, collected between 2000–2007, 1998–2007, and 1996–2007 are obtained. The daily GPS global solutions were processed by the GPS Analysis Center at JPL. The detection power of the common-mode effects is improved by including the time- and space-correlated noise into the least squares power spectrum. Previous signals, such as those with periods of 13.63, 14.2, 14.6, and 14.8 days, are identified in the multivariate analysis. Significant signal with period of 351.6 ± 0.2 days and its higher harmonics are detected in the series, which closely follows the GPS draconitic year. The variation range of this periodic pattern for the north, east, and up components are about ±3, ±3.2, and ±6.5 mm, respectively. Three independent criteria confirm that this periodic pattern is of similar nature at adjacent stations, indicating its independence of the station-related effects such as multipath. It is likely due to the other causes of the GPS draconitic year period driven into GPS time series. The multivariate power spectrum shows a cluster of signals with periods ranging from 5 to 6 days (quasiperiodic signals). In their aliased forms, the effects are likely partly responsible for the time-correlated noise and partly for the periodic patterns at lower frequencies.


1. Introduction

[2] Horizontal and vertical velocities of permanent GPS stations are commonly estimated from the available position time series. Proper analysis of position time series is of particular interest for many geophysical applications that require unbiased estimates of velocities and their uncertainties. There are several issues known to affect the estimation of site velocity and its uncertainty. The ultimate goal of the GPS position time series studies is to discriminate between the functional and the stochastic effects in the series. Both effects are relevant in geophysical phenomena and hence the subject of the present contribution. Functional effects, such as a linear trend, offsets, and potential periodicities, can be well explained by a deterministic model, while the remaining unmodeled effects can be described by a proper stochastic model. Both models should optimally be selected and analyzed for proper analysis of the time series.

[3] Seasonal variations in site positions consist of signals from various geophysical sources and systematic modeling errors [Dong et al., 2002]. They showed that 40% of the seasonal power can be explained by redistributions of geophysical fluid mass loads. Penna and Stewart [2003] and Stewart et al. [2005] showed how unmodeled periodic errors at tidal frequencies can result in spurious longer periodic effects in the resultant time series. Yuan et al. [2009] reported that the K2 body tide and ocean tide loading periodic errors can significantly bias the site velocities. Amiri-Simkooei et al. [2007] showed that the seasonal variations, to a large extent, can be modeled by a set of harmonic functions. The colored noise of the series was shown to mimic periodic patterns, and therefore it could be compensated by using a set of harmonic functions. The results from Ray et al. [2008], Collilieux et al. [2007], Amiri-Simkooei et al. [2007], Tregoning and Watson [2009], King and Watson [2010], and Santamaría-Gómez et al. [2011] revealed harmonics of around 351 days and its higher harmonics in the series, which coincides with the “GPS draconitic year” period, i.e., the 351.4 days required for a GPS orbit to repeat its inertial orientation with respect to the Sun. This periodic pattern has been observed in the power spectra of nearly all products of the IGS [Griffiths
2. Harmonic Estimation Models

The temporal noise characteristics of GPS time series is well described as a combination of white noise and power law noise [Zhang et al., 1997; Williams et al., 2004]. Spatial correlation between time series is also considered to be significant [Williams et al., 2004]. Amiri-Simkooei [2009] presents a multivariate noise assessment of GPS time series in which both time and space correlated noise components are simultaneously estimated using the least squares variance component estimation [Teunissen and Amiri-Simkooei, 2008; Amiri-Simkooei, 2007]. In the studies of Ray et al. [2008] and Amiri-Simkooei et al. [2007], the power spectrum of the series has been estimated on the assumption that temporal and spatial correlation are absent. We now consider these issues on temporal and spatial noise components to optimally estimate the power spectrum density using a multivariate analysis. The presence of colored noise as well as the small amplitude of the signals makes the detection power of the signals inefficient. This is because color noise can also mimic periodic variations [Amiri-Simkooei et al., 2007; Williams, 2007], and hence it can be mixed with real periodic patterns. Multivariate analysis solves these issues.

This paper is organized as follows. In section 2, we derive the formulation of the harmonic estimation for a multivariate linear model. The goal is to detect common-mode signals that are assumed to be present in all GPS position time series. To detect significant signals in GPS time series, the existing spectral analysis methods (e.g., power spectra estimated using the Lomb-Scargle periodogram) are usually formulated based on the assumption that the time series have only white noise and that they are uncorrelated to each other. We give an extension of the least squares harmonic estimation for a multivariate and multiharmonic model, which includes both temporal and spatial correlation of the series. Section 3 applies this theory to daily position time series of 350, 150, and 50 permanent GPS stations. The multivariate power spectrum of the series is estimated using the multivariate harmonic estimation model. Such a model will also provide us with the time- and space-correlated noise of GPS time series. We provide some useful observations on the nature of the periodic pattern reported by Amiri-Simkooei [2007], Ray et al. [2008], Collilieux et al. [2007], King and Watson [2010], and Santamaría-Gómez et al. [2011]. A (more) precise estimate of the period of this signal along with its spatial variations is highlighted in the multivariate analysis. Further, in plate tectonics studies, for an unbiased nature of the periodic pattern reported by Santamaría-Gómez et al. [2007], due to [2011], the covariance matrix of observables \( y \), the \( n \) vector of unknown parameters is to be estimated, and \( E \) and \( D \) are the expectation and dispersion operators, respectively.

In hypotheses testing procedures, equation (1) is assumed to be an appropriate model under the null hypothesis, while it is to be improved under the alternative hypothesis. When there are still undetected periodic patterns in time series, the functional part of the model \( Ax \) can be extended to

\[
E(y) = Ax + A_2 x_k, D(y) = Q_y
\]

where the \( m \times 2 \) matrix \( A_2 \) contains two columns corresponding to the signal \( a_k \cos \omega_k t + b_k \sin \omega_k t \) at frequency \( \omega_k \) \( (x_k = [a_k, b_k]^T \) are two unknown coefficients). This problem is performed using the statistical testing on the basis of likelihood ratio test to see whether or not the introduced two unknowns \( a_k \) and \( b_k \) are indeed statistically significant [Teunissen, 2000].

The problem of finding the unknown frequencies \( \omega_k \) is the task of LS-HE, which can be obtained using the following maximization problem:

\[
\omega_k = \arg \max_{\omega_k} P(\omega_k)
\]

where

\[
P(\omega_k) = \hat{e}^T Q_y^{-1} A_2^T Q_y^{-1} P_\Delta^+ A_2 \hat{e}
\]

is called the spectral values, a terminology introduced in the Fourier spectral analysis [Priestley, 1981]. They are unitless quantities if the proper covariance matrix \( Q_y \) is used in equation (4). They are functions of \( \omega_k \), due to \( A_2 \), the covariance matrix \( Q_y \), and the least squares residuals \( \hat{e} = P_\Delta^+ y \), with \( P_\Delta^+ = I - A (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \) an orthogonal projector. A plot of these spectral values versus a set of discrete values for \( \omega_k \) is used as a tool to investigate the contribution of different frequencies in the construction of the original time series. Therefore, we compute the spectral values for different discrete frequencies using equation (4). The frequency at
which the spectrum achieves its maximum value is the frequency \( \omega_k \) of the periodic pattern.

[10] After detection of \( \omega_k \) using the numerical maximization of equation (3), one has to test the null hypothesis against the alternative one to see whether the detected signal at this frequency is indeed significant. The test statistic in equation (4) at the argument \( \omega_k \) of the maximum power is used:

\[
T_k = P(\omega_k) = e^T A_k^{T} \left( A_k Q_k^{-1} A_k^T \right)^{-1} A_k Q_k^{-1} e,
\]

which, under \( H_0 \), has a central chi-square distribution with 2 degrees of freedom provided that \( Q_k \) is known and that the original time series has multivariate normal distribution [Teunissen, 2000]. If the null hypothesis is rejected, it implies that the detected signal is in fact significant, and we may perform the same procedure for finding yet other frequencies. Matrix \([A \; A_k] \rightarrow A\) is now the new design matrix.

[11] **Special case**: As a special case we may consider the following assumptions to simplify the power spectrum

[12] 1. In the functional model, when dealing with a zero-mean time series, it is concluded that \( A=0 \), and therefore \( P^A = I \).

[13] 2. In the stochastic model, if the random process contains only stationery white noise, one may choose \( Q_k = I \).

[14] These simplifications, with equation (4), yield

\[
P(\omega_k) = E^{T} A_k^{T} A_k^{-1} A_k^{T} E
\]

[15] The above simplified formula has been used in the study by Vanícek [1969, 1971], Lomb [1976], and Scargle, [1982, 1989, 1997] for the power spectrum estimation of unevenly spaced data. With LS-HE, we may in addition include the following terms into the analysis: (1) the linear trend \( Ax \), as a deterministic part of the model, and the covariance matrix \( Q_x \), as a stochastic part of the model. In the following subsections, we introduce the multivariate and multiharmonic LS-HE methods, which are the unique features of this method. The multivariate model is used to detect common-mode signals in multiple time series, while the multiharmonic model is aimed at detecting a periodic pattern using all its constitutive harmonics. Both models will be applied to the multiple series in section 3.

### 2.2. Multivariate Model

[16] When dealing with multiple time series, one may apply the LS-HE to obtain the multivariate power spectrum for detecting the common-mode periodic patterns of all series. If in the linear model, instead of one time series \( y \), there exist several \( r \) time series with identical \( A \) and \( Q_x \), and the corresponding parameter vectors have to be determined, the model is referred to as a multivariate linear model. For a multivariate model, equation (2) is generalized to

\[
E(\text{vec}(Y)) = (I_r \otimes A) \text{vec}(X) + (I_r \otimes A_k) \text{vec}(X_k)
\]

with the multivariate covariance matrix

\[
D(\text{vec}(Y)) = \Sigma \otimes Q
\]

where vec is the vector operator and \( \otimes \) is the Kronecker product, which is an operation that for two matrices \( R \) and \( S \) of arbitrary size results in a block matrix as \( R \otimes S = [r_{ij} S] \).

For the properties of the vector operator and the Kronecker product, we refer to Magnus [1988]. The \( m \times r \) matrix \( Y=[y_1 \; y_2 \; \ldots \; y_r] \) collects observations from \( r \) number of series, and so do the \( n \times r \) matrices \( X=[x_1 \; x_2 \; \ldots \; x_r] \) and \( X=[x_{1k} \; x_{2k} \; \ldots \; x_{rk}] \) for the unknown parameters.

[17] If \( \Sigma = I \), i.e., if the multiple time series are uncorrelated to each other, the multivariate LS-HE reduces to the univariate LS-HE. In general, the full unknown structure of the \( r \times r \) matrix \( \Sigma \) and a few unknowns of the \( m \times m \) matrix \( Q \) can be estimated using a multivariate noise assessment method [Amiri-Simkooei, 2009].

[18] The multivariate structure of \( I_r \otimes A_k \) indicates that there exists common periodic signals in all of the series, which need to be detected using the LS-HE method. The common signals have the same frequency but they are possibly of different amplitudes and phases. In case that the frequency, the amplitude, and the phase of the signals are identical, one may use the \( e_r \otimes A_k \) structure, where \( e_r = [1 \; \ldots \; 1]^T \). This is however not of interest in GPS time series analysis because usually the amplitudes and phases are different.

[19] For power spectrum of the multivariate model one just needs to substitute the terms in equation (4) from the multivariate model as follows: \( I_r \otimes A \rightarrow A, I_r \otimes A_k \rightarrow A_k, \Sigma \otimes Q \rightarrow \Sigma \otimes Q, \text{vec}(E) \rightarrow e \). One then obtains \( P^E = I_r \otimes P^A \), and hence equation (4) yields

\[
P(\omega) = \text{tr} \left( \hat{E}^T Q^{-1} A_k^{T} Q^{-1} P^A A_k^{-1} A_k^{T} Q^{-1} \Sigma^{-1} \right)
\]

with \( \hat{E} = P^A Y \) the least squares residuals and \( P^A = I - A(A^T Q^{-1} A)^{-1} A^T Q^{-1} \) the orthogonal projector. The power spectrum obtained from equation (8) is referred to as the multivariate power spectrum, which simultaneously uses all time series and takes into account the cross correlation (through \( \Sigma \)) and time correlation (through \( Q \)) in an optimal least squares sense. In section 3, we make use of equation (8) in which \( \Sigma \) and \( Q \) are to be estimated using the multivariate noise assessment.

[20] To test the significance of the detected signal, the following test statistic can be used:

\[
T = P(\omega) = \text{tr} \left( \hat{E}^T Q^{-1} A_k^{T} Q^{-1} P^A A_k^{-1} A_k^{T} Q^{-1} \Sigma^{-1} \right)
\]

under the null hypothesis is distributed as a central chi-square distribution with \( 2r \) degrees of freedom: \( T \sim \chi^2(2r, 0) \), provided that both \( \Sigma \) and \( Q \) are known and that the original observables are normally distributed. When \( \Sigma \) and \( Q \) are unknown the distribution has a complicated form. In case that GPS position time series are long enough (e.g., 10 years), one may ignore the randomness of the estimated \( \Sigma \) and \( Q \), and hence the chi-squares distribution can be used as an approximation. Also, multiple time series can further help getting better approximation.

[21] **Special case**: As a special case we may consider the following assumptions to simplify the multivariate power spectrum

[22] 1. When dealing with a zero-mean time series, one has \( A = 0 \), and therefore \( P^A = I \).

[23] 2. If the random process contains only stationery white noise, one may have \( Q_r = I \) (time series are temporarily uncorrelated),
3. Results and Discussions

The multivariate analysis is applied to the GPS position time series of daily global solutions. The solutions are processed using the precise point positioning method in the GPS Inferred Positioning System software (GIPSY) [Zumberge et al., 1997] by the GPS analysis center at Jet Propulsion Laboratory (JPL) [Beutler et al., 1999]. The reader is referred to the JPL Web site [http://sideshow.jpl.nasa.gov/post/series.html]. Prior to the analysis, a methodology based on the Detection, Identification, and Adaptation technique was employed to detect offsets in the time series [Tennissen, 2000]. The series was then corrected for the presence of offsets. Three data sets were employed. The data sets used include 350, 150, and 50 permanent GPS stations of which the time spans are 8, 10, and 12 years, respectively. Equation (8) is used to obtain the multivariate power spectrum of all time series. For this analysis, the matrices $\Sigma$ and $Q$ are estimated using the multivariate variance and covariance estimation developed by Amiri-Simkooei [2009]. Different results are presented for these three data sets.

3.1. Noise Assessment

The matrices $\Sigma$ and $Q$ are estimated using the multivariate noise assessment technique in an iterative procedure [see Amiri-Simkooei, 2009, algorithm in Figure 1]. The matrix $\Sigma$ expresses the cross correlation (e.g., spatial correlation) among the series, while $Q = \sigma^2_v I + \sigma^2_f Q_f + \sigma^2_w Q_w$ expresses time correlation among the observables within individual series. This includes white noise, flicker noise, and random walk noise amplitudes. All cross correlations along the three noise components can be estimated in the multivariate noise assessment.

The Hosking flicker noise covariance matrix introduced and used in Williams [2003], Langbein [2004], Williams et al. [2004], Beavan [2005], and Bos et al. [2008] will be employed in this study. In the multivariate harmonic estimation using equation (8), one needs to use $\Sigma$ and $Q$. One then has to use $\Sigma$ and $Q$ estimated from the same data as the other parameters are estimated. The estimated matrices are however of high precision because the time series used in this study are very long. For example, the precision of an estimated correlation coefficient is $\sigma_{r_p} = (1 - r^2)/(m - n)^{0.5}$ [Amiri-Simkooei, 2009]. For 10 years of daily solutions, this gives $\sigma_{r_p} < 1/60$. Therefore, the randomness of the estimated components is neglected in the multivariate analysis.

For the 350 GPS stations, matrix $\Sigma$ will be of size $\times = 1050 \times 1050$, where $r$ is the total number of series, and matrix $Q$ will be of size $m \times m$, where $m$ is the length of the series. Each of the three 350×350 block diagonal matrices of $\Sigma$ are the spatial correlation of the individual components, while the other three 350×350 off-diagonal blocks show the spatial cross covariance of the components, i.e., among north, east, and up components. Due to the Kronecker structure used in $\Sigma \otimes Q = (\sigma^2_v \Sigma) \otimes I + (\sigma^2_f \Sigma) \otimes Q_f + (\sigma^2_w \Sigma) \otimes Q_w$, the cross correlation in general and the spatial correlations in particular induced by white, flicker, and random walk noise are expressed as $\sigma^2_w \Sigma$, $\sigma^2_f \Sigma$, and $\sigma^2_w \Sigma$, respectively.

3.1.1. Spatial Correlation

GPS position time series are shown to be significantly spatially correlated [Williams et al., 2004; Amiri-Simkooei, 2009]. The results for (cross) correlation among series are
shown in Figure 1. It consists of spatial correlation between components north-north (NN), EE, UU, NE, NU, and EU obtained from $\Sigma$ estimated using the multivariate analysis. [33] The spatial correlations of individual components are significant over an angular range of $30^\circ$ (corresponding to about 3000 km). The maximum correlations have been obtained between the nearest sites, confirming that the noise has a common physical basis. Among the components, the spatial correlation of NN is higher than those of EE and UU components. Significant correlation of individual components has already been reported by Williams et al. [2004]; Amiri-Simkooei [2009]. The spatial (cross) correlations between components are however not significant; the mean correlation curve is around 0.1 for all components. We would have intuitively suspected that coordinate components of a station (or of adjacent stations) would be correlated. This correlation is expected because the components have been simultaneously estimated from the same data set through one functional model. The statement is correct for one epoch of observations or for a couple of adjacent epochs. When considering all observations together (24 h), we have a good GPS geometry with which the estimated coordinates are approximately uncorrelated.

[34] This high spatial correlation of the GPS time series should be taken into account when estimating the multivariate power spectrum using equation (8). We however note that the stacked normalized spectra given by Amiri-Simkooei et al. [2007] and Ray et al. [2008] are based on the assumption of uncorrelated time series ($\Sigma$ diagonal). Though neglecting these correlations cannot likely introduce biases to the detected signals, they can give rise to imprecise estimates of the detected frequency of the signal.

3.1.2. Temporal Correlation

[35] The white, flicker, and random walk noise amplitudes are estimated through matrices $Q$ and $\Sigma$. Previous work has shown that the noise in GPS position time series can be taken as a combination of white noise and flicker noise. Random walk noise, which is mainly related to monument instability, was shown not to be present in the global GPS solutions [Williams et al., 2004]. This could be due to the shortness of the time series or to the dominance of the other noise types such as flicker noise, which masks the (small) amplitude of random walk noise. Using the multivariate model, we have now the possibility to include the three noise components white, flicker, and random walk and estimate their amplitudes simultaneously.

[36] The results for time correlation are shown in Figure 2. Due to the Kronecker structure used in $\Sigma \otimes Q$, the amplitudes of flicker and random walk noise over different stations are multiples of the white noise amplitudes. The variances of the series along with their cross correlations induced by white, flicker, and random walk noise can be obtained from the $r \times r$ matrices $\sigma^2_w \Sigma$, $\sigma^2_f \Sigma$, and $\sigma^2_{rw} \Sigma$, respectively. We highlight, in real situation, that this should not indicate all stations contain random walk noise, because the estimated values are an average value (over all stations) due to the special structure used.

[37] The average (over 350 series) amplitudes of white noise components along with their estimated standard deviations (or formal errors denoted by $\pm$) are $2 \pm 0.03$, $2.64 \pm 0.04$, and $5.85 \pm 0.1$ mm for north, east, and up components, respectively. The corresponding values for flicker and random walk noise are $0.9 \pm 0.02$, $1.18 \pm 0.02$, and $2.63 \pm 0.04$ mm/yr$^{1/4}$, and $0.08 \pm 0.001$, $0.1 \pm 0.002$, and $0.22 \pm 0.004$ mm/yr$^{1/2}$, respectively. These results indicate that white and flicker noise have the largest contributions of the noise structure and random walk noise has the smallest contribution. The small standard deviations and the high convergence speed—when estimating the noise amplitudes—indicate that the estimated noise amplitudes are in fact reliable. We however note that the absolute random walk noise amplitudes are small, indicating no significant effects on rate uncertainties. The noise amplitudes of the east component are, on average, larger than those of the north components, which may be due to the
ambiguity fixing effects. Also, the noise amplitudes of the up component are larger than those of the north and east components at least by a factor of 2. Similar results can also be obtained using the other two data sets.

3.2. Power Spectrum

3.2.1. Multivariate Spectrum

Equation (8) is now used to obtain the multivariate power spectrum (Figure 3). Ideally, for a white noise structure, the power spectrum becomes flat, which indicates that the spectrum has a constant power at different frequencies. This situation also holds when the (correct) covariance matrix of the time series is used in estimating the power spectrum of equation (8). The flatness of the power spectrum is thus due to the use of the correct covariance matrices $Q$ and $\Sigma$. This flat spectrum indicates that the signals at higher frequencies can statistically be significant through the statistical hypothesis testing, which is due to a precise estimate of the detected signals at higher frequencies. With an immature stochastic model (i.e., uncorrelated series with white noise structure), the peaks at higher frequencies cannot likely be statistically significant due to their lower spectral values at these frequencies [see Amiri-Simkooei et al., 2007, Figure 7].

The multivariate spectrum shows signals with periods of 13.63, 14.2, 14.6, and 14.8 days reported also in previous work. These signals are mainly due to the aliasing effects of which subdaily unmodeled periodic signals lead to spurious long-period signals in GPS position time series Penna and Stewart [2003]. The mechanism of the aliasing effects was investigated in detail by Stewart et al. [2005]. The errors sources mainly come from the deficiencies of body tide models and ocean tide loading models [Yuan and Chao, 2012] and unmodeled diurnal and semidiurnal atmospheric tidal loading deformation model [Tregoning and Watson, 2009, 2011]. For example, the 13.63 day peak indicates a propagated signal caused by mismodeling of the M2 or O1 constituents within ocean tide loading or solid Earth tide models.

The vertical dashed lines show the 10 harmonics of 1.04 cycle per year. The peaks clearly match with all of the 10 frequencies. The fits are more apparent at higher frequencies. To explain these signals, we recall the theory of Fourier series expansion of periodic functions. Let the function $f(x)$ be an arbitrary periodic function with period of $T$ (i.e., $f(x + T) = f(x)$). Because such a periodic function is not purely sinusoidal, it can theoretically be written as an infinite sum of sine and cosine functions on the interval $[-T/2, T/2]$. The results presented are in fact an example of a Fourier series decomposition of a periodic function of $T = 351.4$ days into a truncated sum (up to 10) of simple oscillating functions sines and cosines.

The resulting significant signals are close to the harmonics of the GPS draconitic year period (351.4 days), which is the revolution period of the GPS constellation.
in inertial space with respect to the Sun. The aliasing signals may contribute to parts of the dracontic signal. In fact, there are a few possible causes of the mapping/aliasing/propagation of the draconitic year period into GPS position time series. These mappings may be driven by orbit mismodeling, atmospheric loading, or station-dependent multipath errors [Ray et al., 2008; Amiri-Simkooei et al., 2007; Tregoning and Watson, 2009; King and Watson, 2010; Santamaría-Gómez et al., 2011]. For example, a possible cause is multipath error. The GPS constellation is considered to have orbital periods of 12 sidereal hours, in which case, the repeat period would have a frequency of one cycle per year. The daily GPS solutions are averaged over one solar day (24 h). However, the repeat time of the GPS constellation has an average value over the day corresponding to a daily advance of 246.8 s [Agnew and Larson, 2007]. Therefore, for daily sampling, this repeat period will alias to a frequency of 246.8/86400 = 0.0028565 cycles/d, or 1.04 cycles/yr.

The lowest frequency is intuitively expected to have the largest and sharpest peak. Figure 3 shows that the estimated spectrum for the first harmonic is flattened to some extent, in agreement with the findings of Santamaría-Gómez et al. [2011] where the detected signals were more scattered at lower frequencies. The reason is due to the leakage effect in spectral analysis, which expresses that in order to detect two signals with periods of $T_1$ and $T_2$, the time span of the series should be at least equal to $T_1 T_2/(T_2 - T_1)$. The length of the GPS time series is not yet long enough to distinguish between the annual signal and the draconitic year period; the time series should be at least 25 years long. Parts of the power spectrum at this frequency have been absorbed by the annual signal, which has already been removed from the time series. A sharper peak at the principal frequency for the bottom frame is due to the longer time series (12 years) used for this spectrum. We however note that at higher harmonics of this periodic pattern, shorter durations of data allow the separation of the frequencies as can be seen in Figure 3 (e.g., for the first harmonic the data should be only 12.5 years long).

The spectrum shows a cluster of periods between 5 and 6 days along with their first harmonics (Figure 3, top frame). There exist many small peaks indicating no clear, sharp, and unique peak for these signals. Therefore, they are likely of quasiperiodic nature and perhaps station-dependent which explain local phenomena.

### 3.2.2. Multiharmonic Spectrum

To find a more precise estimate of the period of this periodic pattern, we use the multivariate multiharmonic model. It is “multivariate” because use is made of multiples time series. Also, it is “multiharmonic” because the effect of all constitutive harmonics is used to detect the principal frequency of the periodic pattern. In the case that one assumes $Q = I$ (while it is not true due to the presence of power law noise), the higher-frequency signals will be downweighted. This indicates that higher-frequency signals cannot significantly contribute to detect the principal frequency using the multiharmonic model introduced in equation (10). This is however not the case because of the use of the estimated matrices $Q$ (time correlation) and $\Sigma$ (spatial correlation). The multivariate power spectrum is flat (Figure 3), and hence the higher frequencies can significantly contribute to detect the periodic pattern.

![Figure 3](image-url)

**Figure 3.** Multivariate least squares power spectrum expressed in equation (8). Vertical axes are normalized, such that the maximum power is equal to one; (top) 350 stations, (middle) 150 stations, (bottom) and 50 stations.

![Figure 4](image-url)

**Figure 4.** Multivariate multiharmonic spectrum expressed in equation (8), where $A_j$ is obtained from equation (10) with $p = 10$. Vertical axes are normalized such that the maximum power is equal to one; (top) 350 stations, (middle) 150 stations, and (bottom) 50 stations.
The results are presented in Figure 4, where the three spectra correspond to the three data sets considered. As expected, the largest peak of this periodic pattern should locate at the principal frequency. The peaks at the top, middle, and bottom frames are at 351.46, 351.56, and 351.80 days, respectively. The average and standard deviation of these three numbers are 351.60 ± 0.2 days, which nicely follow the GPS draconitic year period, i.e., 351.4 days.

Finally, a comment on the multivariate spectra in Figures 3 and 4 is in order. These plots have been made when considering the north, east, and up components simultaneously. We may try to make these plots per component. Because the results are quite similar to the case when we treat them together, they are not presented in this contribution. For example if we make the multiharmonic spectrum individually for each component, the peaks for the north, east, and up components occur at 351.7, 351.7, and 351.4 days, respectively (350 stations). These numbers change to 351.1, 352.1, and 351.9 days (150 stations), and 351.6, 352.2, and 351.1 (50 stations). The mean and standard deviation of these nine numbers closely follow those specified above.

3.3. On the Nature of Periodic Pattern

This subsection focuses on the nature of this periodic pattern. In the linear model $y = Ax$, we may partition the design matrix and the unknown parameters as $A = [A_1 : A_2]$ and $x = [x_1^T x_2^T]^T$, respectively, where $x_1$ contains unknowns of the linear trend terms plus annual and semiannual signals, and $x_2$ contains those for the draconitic year signals and all its harmonics up to 10. We now investigate the signal estimated for this periodic signal using $\hat{y}_2 = A_2 \hat{x}_2$. Such results can be obtained for all time series, of which the results are collected in the $m \times r$ matrix $\hat{Y}_2 = A_2 \hat{X}_2$. An investigation on this
matrix shows that the mean range (over all stations) of variations of this periodic pattern for the north, east, and up components are $-2.9$ to 3.0 mm, $-3.3$ to 3.0 mm, and $-6.3$ to 6.7 mm, respectively. These ranges show the amplitudes (the minimum and maximum values) of this periodic pattern. The mean amplitudes (mean absolute values over one draconitic year) are 1.4, 1.3, and 2.8 mm for the north, east, and up components, respectively.

Three ways for further investigation of this phenomenon are presented as follows.

3.3.1. Visual Inspection

The simplest way to compare the effect of this periodic pattern on different GPS stations is the visual inspection. Figures 5 and 6 show two typical examples of the behavior of this periodic pattern for adjacent (e.g., < 10 km) and for very far (e.g., > 3000 km) GPS stations, respectively. The plots show that this periodic pattern is of similar behavior for the adjacent stations, but they are different for distant stations. This indicates that the dominant effect is independent of the station-related effects such as multipath.

3.3.2. Correlation Analysis

Furthermore, we investigate the behavior of this periodic pattern using the correlation analysis. We obtain the spatial correlation induced by the “zero mean” time series collected in the matrix $Y_2$ using $Y_2^T Y_2/m$. We made it zero mean by using the sinusoidal functions over one full cycle (it is thus not zero mean from statistical viewpoint). The results are presented in Figure 7 for the data set with 350 stations. Significant correlation for the adjacent stations can be observed for the three components over the angular range of 0° to 20° (2000 km). This is second way of confirming that the periodic pattern behaves alike for the adjacent stations but can be different for very distant stations.

3.3.3. Principal Component Analysis and k-means Clustering

Principal component analysis (PCA) is a standard mathematical tool that transforms a number of different but possibly correlated variables into a smaller number of uncorrelated variables called principal components (PCs). The method extracts relevant information from complex data sets consisting of different attributes (features). High dimensionality of the features makes their interpretation difficult. PCA involves the calculation of the eigenvalue decomposition of the data covariance matrix. The order of PCs is based on the amount of variations they represent. The first PC has the largest variations. Each component can then be interpreted as the direction, uncorrelated to previous components, which maximizes the variance of the samples when projected onto that component. For a short but comprehensive discussion on PCA, we refer to Amiri-Simkooei et al. [2011]. The first PCs are fed to the $k$-means clustering algorithm to find data samples (here GPS stations) of similar spatial pattern [see Seber, 1984].

Data over one GPS draconitic year (351 days) of the $Y_2$ are used for the PCA process. Collecting all data of the 350 GPS sites make a $350 \times 351 \times 3$ data matrix as $\hat{Y} = \begin{bmatrix} \hat{Y}_N^T & \hat{Y}_E^T & \hat{Y}_U^T \end{bmatrix}$. The rows of $\hat{Y}$ correspond to the 350 GPS stations, and the columns of $\hat{Y}$ correspond to the total effect of the 10 harmonics of the periodic pattern over the GPS draconitic year period for the north, east, and up components. Prior to applying a PCA process, each column of the matrix $\hat{Y}$ is standardized to have a zero mean and unit variance. The correlation matrix is then extracted from $R = \hat{Y}^T \hat{Y}$. Using the eigenvalue decomposition, i.e., $R = U \Lambda U^T$, the first 36 PCs of the data matrix, which account for 98% of the variability in the data, is fed to the $k$-means clustering algorithm.

The clustering results are presented in Figure 8 for having two (or three) clusters. The top frame gives the results for the two clusters in which the first cluster (blue circles) contains a series of adjacent sites (64%) and the second cluster (red asterisks) consists of other sites (the remaining 36%). When using three clusters, the first cluster (blue circles), second cluster (green pluses), and third cluster (red asterisks) have the percentages of 52%, 22%, and 26% of the total 350 stations, respectively. Most of the clustered stations are again close to each other, indicating a similar spatial pattern at adjacent stations.

For many geophysical applications of GPS position time series, separation between the functional and stochastic effects is of high importance. The significant detected signal with the GPS draconitic year period is an example of the functional effect. Its large amplitude along with its similar spatial pattern at adjacent stations indicate that this effect should be compensated for geophysical phenomena in the
functional part of the GPS time series model. Similar to the annual and semiannual signals, which can significantly bias the site velocity and its uncertainty, the draconitic pattern can also bias these two parameters. The amount of induced bias on site velocity is inversely proportional to the length of the time series. For the current study (because the series are long enough), this bias is negligible on site velocities. However, it can significantly affect the site velocities for shorter time series (e.g., 1 year series). Our results indicate that the impact of the periodic pattern on noise assessment is about 8%, which is overestimated if it is neglected.

4. Concluding Remarks

[56] Daily position time series of 350, 150, and 50 permanent GPS stations, respectively, with the time span of 8, 10, and 12 years were analyzed. Spatial correlation, time correlated noise, and power spectrum of the series were estimated using a multivariate noise assessment and harmonic estimation technique. The results indicated that the GPS position time series are significantly temporarily and spatially correlated. The spatial correlation is higher in the north components than the east and up components. It was also shown that the random walk amplitude can be estimated in the multivariate noise assessment, though its amplitude was shown to be significantly smaller than the white and flicker noise amplitudes.

[57] For signal detection, in the multivariate analysis, we aimed to improve the detection power of the common-mode signals in a least squares sense by including the time- and space-correlated noise into the least squares power spectrum. The results showed that all previous signals detected previously in GPS time series were identified using the multivariate analysis. Using the multivariate multiharmonic model, it was shown that the period of the periodic pattern closely follow 351.4 days, the period of GPS draconitic year. The range of variations for this periodic pattern was shown to be $-2.9$ to $3.0$ mm, $-3.3$ to $3.0$ mm, and $-6.3$ to $6.7$ mm for the north, east, and up components, respectively. Three independent measures were used to
evaluate the behavior of the periodic pattern. They include a visual inspection, a spatial correlation analysis, and a principal component analysis. We showed that this periodic pattern is of similar nature at adjacent stations (Figure 5) but differs significantly for very distant stations (Figure 6). Therefore, this effect does not likely depend on the station-related effects such as multipath. We hypothesize that this pattern is in fact due to other causes of the GPS draconitic year period driven into GPS time series. Other possible causes of the mapping of the draconitic year period into GPS position time series include orbit mismodeling and atmospheric loading effects.

The power spectrum showed a cluster of signals with periods ranging from 5 to 6 days obtained using the multivariate analysis. Such quasiperiodic signals are likely dependent on site positions, and they are likely partly responsible for the time-correlated noise and partly for the periodic patterns of the time series in their aliased form.

It is highlighted that proper plate tectonics studies using GPS position time series require an appropriate functional model, in which all deterministic effects are modeled, and a realistic stochastic model, in which all noise components are estimated. The ultimate goal of the GPS position time series studies is to discriminate between the functional and the stochastic effects. Both effects are relevant and were addressed in this contribution. Functional effects such as a linear trend, offsets, and potential periodicities can be well explained by a deterministic model. Signals with periods close to GPS draconitic year period is an example of deterministic signal, to be included in the functional part of the model. The remaining unmodeled effects can best be described by a stochastic model. The significant spatial correlation among the series (Figure 1) will seriously affect the interpretation of many geophysical phenomena–strain parameters for instance. The spatial correlation induced from the estimated periodic pattern (Figure 7) was shown to be of a similar nature to the spatial correlation of the series (Figure 1). Parts of such significant spatial correlation are expected to be due to this periodic pattern.

Acknowledgments. I am very grateful to Peter Teunissen at Curtin University for his valued comments on earlier versions of this manuscript. His remarks improved the quality and presentation of this paper, which is kindly appreciated. I would also like to acknowledge the associate editor and two anonymous reviewers for their valuable comments on this paper.

References


Williams, S. D. P. (2007), Personal communication, Proudman Oceanographic Laboratory, Bidston, UK.


