Geophys. J. Int. (2008) **174,** 1052–1064 doi: 10.1111/j.1365-246X.2008.03846.x

Instantaneous deformation from continuous GPS: contributions from quasi-periodic loads

Richard A. Bennett

Department of Geosciences, University of Arizona, Tucson, AZ 85721-0077, *USA. E-mail: rab@geo.arizona.edu*

Accepted 2008 May 7. Received 2008 April 17; in original form 2008 January 23

SUMMARY

Continuous GPS (CGPS) coordinate time-series are known to experience repeating deformation signals with seasonal and other periods. It is unlikely that these signals represent perfect sinusoids with temporally constant amplitude. We develop an analysis method that accommodates temporal variations in the amplitudes of sinusoidal signals. We apply the method to simulated coordinate time-series to numerically explore the potential consequences of neglecting decadal variation in amplitude of annual motions on the residual-error spectra of CGPS measurements, as well as potential bias in estimates for secular site velocity. We find that secular velocity bias can be appreciable for shorter time-series, and that residual-error time-series of longer duration may contain significant power in a broad band centred on semi-annual frequency if temporal variation in the amplitude of annual motions is not accounted for in the model used to reduce the observations to residuals. It may be difficult to differentiate the bandpass filtered signature of mismodelled loading signals from power-law noise, using residual-error spectra for shorter time-series. We provide an example application to a ∼9-yr coordinate time-series for a CGPS station located in southern California at Carbon Creek Control Structure (CCCS), which is known to experience large amplitude seasonal motions associated with the Santa Ana aquifer system.

Key words: Inverse theory; Numerical approximations and analysis; Satellite geodesy; Transient deformation.

1 INTRODUCTION

Precise coordinate time-series from permanent networks of continuously-recording GPS (CGPS) stations, such as the EarthScope Plate Boundary Observatory Facility, can be used to study numerous geophysical processes, including the motions of tectonic plates, strain accumulation around active faults, coseismic or aseismic fault slip, viscous relaxation of stresses in the deeper parts of the crust or upper mantle, magmatic and hydrologic processes, the flow fields of ice sheets and other sources of motion. Many of these processes cause Earth's surface to move with complex spatial and temporal patterns, not well described by constant rates of change of coordinates. An important limitation to our ability to use geodetic measurements to understand these processes is that, in many cases, we do not know *a priori* the source of the observed deformation. Thus, it can be difficult to specify an appropriate set of kinematic model parameters or to model observed motions directly in terms of density or stress distributions on or beneath Earth's surface. However, direct studies of observed crustal motions in the absence of a physical model for crustal motion can sometimes reveal distinctive clues critical to identification of the source processes and ultimately to the specification of a physical model.

Many sources of deformation of the solid Earth's surface (including surface area covered by ice) are associated with periodic loading. Loading effects have been studied for more than a century (e.g. Darwin 1882) but have been studied most intensely over the past decade, coincident with the emergence of CGPS networks specifically designed for geodynamics (e.g. van Dam *et al.* 1994; van Dam & Wahr 1998; Blewitt *et al.* 2001; van Dam *et al.* 2001; Blewitt & Lavallée 2002; Dong *et al.* 2002; Elósegui *et al.* 2003; Penna & Stewart 2003; Davis *et al.* 2004 Stewart *et al.* 2005; Nicolas *et al.* 2006; Watson *et al.* 2006; Clarke *et al.* 2007; Penna *et al.* 2007; van Dam *et al.* 2007; and others). Positioning errors associated with unaccounted for or mismodelled multipath, atmospheric delays, GPS ephemerides, tidal deformations and other potential sources of repeating error could also contribute to periodicity in CGPS time-series.

Periodic signals can severely hamper estimation of secular site velocity if unaccounted for. Blewitt & Lavallée (2002) showed that annual signals of constant amplitude in GPS coordinate time-series will result in non-negligible biases in estimates for secular crustal motion if periodic terms are not included in the kinematic model. They also showed how neglecting to account for constant amplitude periodic motions could lead to correlations in residual-error time-series, with spectral characteristics similar to those of power-law noise models. Periodic surface motions derive from a number of load sources, such as atmospheric pressure changes, ocean and solid Earth tides, ice mass fluctuations and variations in water table levels

and are significant at semi-monthly, monthly and semi-annual periods (e.g. van Dam *et al.* 2001). Although they may wax and wane with some regularity, it is unlikely that any of these variations have temporally constant amplitudes, given that they are associated with the highly dynamic fluid parts of Earth and could also arise from temporally-variable positioning errors. If unaccounted for, these departures from purely periodic site motion could also lead to biased estimates for secular velocity. Moreover, such 'quasi-periodic' site motion could potentially dominate the time-variable component of instantaneous coordinate velocity, causing instantaneous velocity to fluctuate between positive and negative values, obscuring any underlying signals of a non-secular non-periodic nature. Thus, a complete understanding of site motions caused by quasi-periodic loads may be critical to our ability to use CGPS time-series to investigate a wide class of potential non-periodic time-dependent crustal deformation processes. Finally, if the responses to quasi-periodic loads are unaccounted for, inferences regarding the observation–error spectrum based on residual-error time-series could be inaccurate.

In this paper, we develop a flexible semi-parametric model suitable for investigating quasi-periodic signals in CGPS coordinate time-series. We chose a method that is flexible in that it allows for small (bounded) temporal variations in the amplitudes of the signals. A limited number of previous investigations have accounted for the possibility of time-variable amplitudes in seasonal signals, with the goal of mitigating their effect on estimates for crustal deformation (e.g. Murray & Segall 2005). Our investigation differs in that it concentrates on an assessment of (1) our ability to estimate periodic amplitude variations; (2) their effects on secular velocity and (3) their effects on estimates for residual-error statistics.

Although we concentrate on annual signals in this paper, the analysis method that we develop should be applicable to any sinusoidal signal of period relevant to CGPS time-series, whether real or apparent. The method should be useful for mitigating biases associated with complex quasi-periodic signals when estimating secular velocities or other non-periodic signals using coordinate time-series. The theory, upon which our analysis is based, is fairly general and could be modified for investigations of nominally periodic signals with waveforms other than sinusoids. Moreover, although our discussion focuses on the analysis of CGPS coordinate time-series, the method may have application to other types of data streams containing periodic signals subject to small magnitude variations in waveform amplitude.

Section 2 introduces the analysis procedure. We set up the problem, provide a solution algorithm and briefly discuss the issues of model selection and error analysis. Section 3 provides three examples using simulated data, with emphasis on the accuracy of secular velocity estimates and error analyses for the case of time-variable quasi-periodic signals. Section 4 shows an application of the method to an observed CGPS coordinate time-series from southern California. Finally, we discuss the limitations of the analysis approach and the general problem of modelling quasi-periodic site motion in Section 5.

2 THEORY

2.1 Statement of the problem

Consider a kinematic model for the time-evolution of a geodetic coordinate *x(t)* given by

$$x(t) = x_0 + vt + s(t), \tag{1}$$

where x_0 is a constant initial offset, v is a constant velocity and *s(t)* is a model for quasi-periodic motion of the form

$$s(t) = a \sin\left(\frac{2\pi}{T}t\right) + b \cos\left(\frac{2\pi}{T}t\right) + c(t) \sin\left(\frac{2\pi}{T}t + \theta\right). \tag{2}$$

a and b are unknown constants representing time-averaged wave amplitudes and *c(t)* is an unknown amplitude deviation function with time-averaged value of zero. T is the period of the signal. For simplicity, we assume that the time-variable part of the signal is in phase with the time-averaged part, such that $\theta = \tan^{-1}(b/a)$. It is possible to relax this restriction by simultaneously estimating amplitude deviations for both sine and cosine terms (with a common period), though at the expense of increased computation time and model variance. However, quasi-periodic site motion driven by environmental loads may, in many instances, be adequately accounted for by assuming constant phase. An example of this is provided below using real data from CGPS site Carbon Creek Control Structure (CCCS) in southern California.

Given a set of measurements d_i of the coordinate x at a series of M different times t_i:

$$d_i = x_0 + vt_i + s(t_i) + e_i, \quad i = 1, \ldots, M, \tag{3}$$

we seek to determine the constants a and b and the deviation function *c(t)*, as well as the unknown constants x_0 and v. We assume that the data are sufficiently abundant and distributed through time that x_0, v, a and b are estimable without regularization in the absence of amplitude deviation c. We also assume that the data are ordered according to increasing time and that the t_i for which we have measurements define the closed interval $[0, T_M]$. For the purposes of developing the problem, we assume that the errors represented by e_i are zero-mean random variables with variance–covariance $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. It is straightforward to consider a more general error variance–covariance matrix $\sigma^2\mathbf{V}$ if required by a specific data set, as we describe in more detail below. We also discuss the use of residual-errors to make inferences about the statistics of the true observational errors below.

Estimation of the constants x_0, v, a and b is straightforward and constitutes a core component of routine GPS time-series analysis. A more challenging problem is the determination of the continuous time-varying deviation function *c(t)*. We here address this latter problem.

To begin, it is convenient to define the linear functionals ϕ_i such that

$$\phi_i(c) = \int_0^{T_M} \delta(t - t_i) \sin\left(\frac{2\pi}{T}t + \theta\right) c(t)\, \mathrm{d}t, \tag{4}$$

where $\delta(t)$ is the Dirac delta function. We temporarily ignore the dependence of the ϕ_i on the initial phase θ and thus the unknowns a and b as if the phase were known *a priori*. This dependence is straightforward to address as we demonstrate below.

Our problem is severely underdetermined, given that we seek to estimate a continuous function from a finite number of data. One approach to addressing this 'ill-posedness' is to impose restrictions on the class of admissible amplitude deviations. Toward this end, we define a Hilbert space of amplitude deviations $\mathcal{A}$ such that

$$\mathcal{A} = \left\{ c : [0, T_M] \to \mathbb{R}, \int_0^{T_M} c(t)\mathrm{d}t = 0, \int_0^{T_M} \dot{c}(t)^2\, \mathrm{d}t < \infty \right\}, \tag{5}$$

where $\mathbb{R}$ is the set of real numbers and $\dot{c}$ is the time-derivative of c. We choose the (squared) norm on $\mathcal{A}$ to be

$$\|c\|_{\mathcal{A}}^2 = \int_0^{T_M} [\dot{c}(t)]^2 \, \mathrm{d}t. \tag{6}$$

The inner product associated with this norm is

$$\langle f, g \rangle_{\mathcal{A}} = \int_0^{T_M} \dot{f}(t)\dot{g}(t) \, \mathrm{d}t, \tag{7}$$

for all $f, g \in \mathcal{A}$. Under this norm, the amplitude functionals ϕ_i are bounded on $\mathcal{A}$, meaning that

$$|\phi_i(f) - \phi_i(g)| \leq K \|f - g\|_{\mathcal{A}}, \tag{8}$$

for some positive real number K known as the 'operator norm'.

Of the infinite number of constants x_0, v, a, $b \in \mathbb{R}$ and deviations $c(t) \in \mathcal{A}$, we concentrate on that subset minimizing the functionals Ψ_ρ given by

$$\Psi_\rho(\mathbf{m}, c) = \frac{1}{\sigma^2} \sum_{i=1}^{M} [d_i - \mathbf{g}_i \mathbf{m} - \phi_i(c)]^2 + \rho \|c\|_{\mathcal{A}}^2, \tag{9}$$

where $\mathbf{g}_i \in \mathbb{R}^4$ is the row vector with elements 1, t_i, $\sin(2\pi t_i/T)$ and $\cos(2\pi t_i/T)$ and $\mathbf{m} \in \mathbb{R}^4$ is the column vector with elements x_0, v, a, b, respectively. ρ is a regularization parameter, controlling the relative importance of matching the data points d_i and minimizing the variability of the deviation function as measured by the Hilbert space norm $\| \cdot \|_{\mathcal{A}}$.

Our choice of Hilbert space $\mathcal{A}$ and performance functional Ψ_ρ allow us to construct a finite dimensional basis set for our solutions, using the Reisz Representation Theorem (Naylor & Sell 1982). According to this theorem, there exist unique functions $\Phi_i \in \mathcal{A}$, such that

$$\phi_i(c) = \langle \Phi_i, c \rangle_{\mathcal{A}}, \tag{10}$$

for each bounded linear functional ϕ_i on $\mathcal{A}$.

Functions $\Phi_i(t)$ satisfying (10) are said to be the 'representers' of the corresponding functionals ϕ_i. To identify expressions for the representers Φ_i, it is useful to first consider the representers associated with point evaluation on $\mathcal{A}$, that is, with the functionals $p_i(c)$ given by

$$p_i(c) \equiv c(t_i) \tag{11}$$

$$= \int_0^{T_M} \delta(t - t_i) c(t) \, \mathrm{d}t \tag{12}$$

$$= \langle \mathcal{P}_i, c \rangle_{\mathcal{A}}. \tag{13}$$

The representers of point evaluation associated with the Hilbert space $\mathcal{A}$ are

$$\mathcal{P}_i(t) = k_1(t_i)k_1(t) + k_2(|t_i - t|), \quad t, t_i \in [0, T_M], \tag{14}$$

where $k_1(t) = t/T_M - 1/2$ and $k_2(t) = 1/2[k_1^2(t) - 1/12]$ for $t \in [0, T_M]$. We derived our $\mathcal{P}_i(t)$ from expressions provided by Craven & Wahba (1979), which are valid for the unit interval [0,1], by simple scaling of the time-variable. An important property of the $\mathcal{P}_i(t)$ is that they repeat periodically outside of the interval $[0, T_M]$ with period T_M [The periodic property of the point evaluation representers is completely independent of the (quasi-)periodic nature of the signals under investigation. The fact that we are investigating periodic signals is entirely coincidental]. The functions $\mathcal{P}(t_i, t) \equiv \mathcal{P}_i(t)$ are analogous to Green's functions for differential operators in that the representers for more general bounded linear functionals, such as ϕ_i, may be derived from these representers of point evaluation. For our case,

$$\Phi_i(t) = \int_0^{T_M} \delta(u - t_i) \sin(2\pi u/T + \theta) \mathcal{P}(u, t) \, \mathrm{d}u \tag{15}$$

$$= \sin(2\pi t_i/T + \theta) \, \mathcal{P}_i(t), \tag{16}$$

as is readily verified:

$$\phi_i(c) = \langle \Phi_i, c \rangle_{\mathcal{A}} \tag{17}$$

$$= \sin(2\pi t_i/T + \theta) \langle \mathcal{P}_i, c \rangle_{\mathcal{A}} \tag{18}$$

$$= \sin(2\pi t_i/T + \theta) p_i(c) \tag{19}$$

$$= \sin(2\pi t_i/T + \theta) c(t_i) \tag{20}$$

$$= \int_0^{T_M} \delta(t - t_i) \sin\left(\frac{2\pi}{T} t + \theta\right) c(t) \, \mathrm{d}t. \tag{21}$$

Fig. 1 shows 10 examples of the amplitude deviation representers Φ_i evenly spaced in the interval 0–10 yr. Each Φ_i integrates to zero over the interval 0–10 yr.

Given these representers, all estimators of the deviation $c(t)$ minimizing (9) subject to $c \in \mathcal{A}$ may be expressed as

$$c(t) = \sum_{i=1}^{M} \xi_i \Phi_i(t), \tag{22}$$

where the M $\xi_i \in \mathbb{R}$ are unknown coefficients. To show that this is the case, consider an amplitude variation of the form $c(t) = c_\perp(t) + \sum_{i=1}^{M} \xi_i \Phi_i(t)$, such that $c_\perp(t)$ is orthogonal to each of the representers $\Phi_i(t)$. This means that $\langle \Phi_i, c_\perp \rangle_{\mathcal{A}} = \phi_i(c_\perp) = 0$. We see immediately that non-zero $c_\perp$ cannot contribute in any way to fitting the data. Furthermore, from the Projection Theorem (Naylor & Sell 1982)

$$\|c\|_{\mathcal{A}}^2 = \|c_\perp\|_{\mathcal{A}}^2 + \left\| \sum_{i=1}^{M} \xi_i \Phi_i \right\|_{\mathcal{A}}^2 \geq \left\| \sum_{i=1}^{M} \xi_i \Phi_i \right\|_{\mathcal{A}}^2. \tag{23}$$

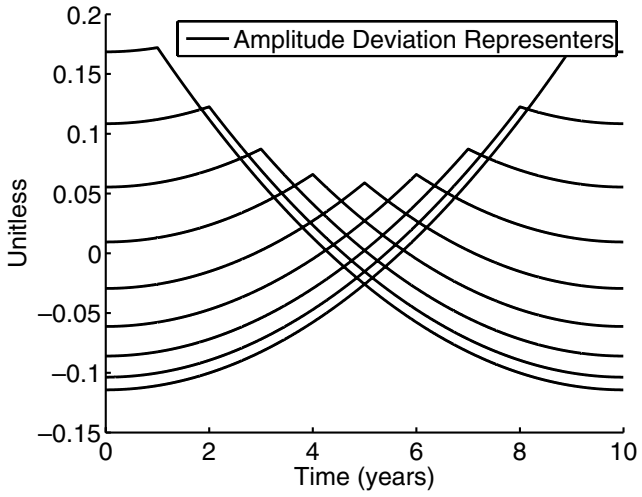

Figure 1. Amplitude deviation representers (solid curves) for select epochs evenly spaced in the interval 0–10 yr. Each coordinate estimate of the time-series is associated with a representer, which depends on the epoch of measurement. Deviation function estimates are constructed from linear combinations of these representers. The area under each curve is zero.

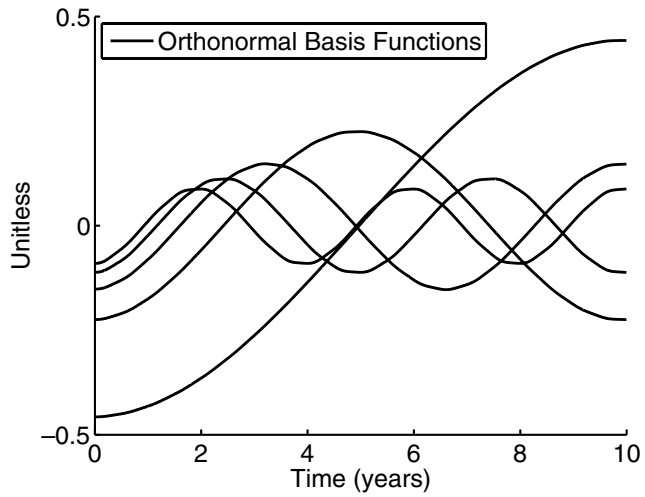

Figure 2. Orthonormal basis functions (solid curves) derived from the representers of amplitude deviation shown in Fig. 1. Only the first five basis functions are shown for clarity.

That is, $c_\perp$ only serves to increase the norm of the amplitude variation. For any given ρ and $\sum_{i=1}^{M} \xi_i \Phi_i$, we can always make $\Psi_\rho(c)$ smaller by reducing $\|c_\perp\|_{\mathcal{A}}$. Therefore, any amplitude variation minimizing $\Psi_\rho(c)$ must be such that $\hat{c}_\perp = 0$.

For computation and analysis, it is advantageous to orthonormalize the amplitude representers, such that

$$c(t) = \sum_{i=1}^{N} c_i \eta_i(t), \tag{24}$$

where $N \leq M$ and the functions $\eta_i(t)$ form an orthonormal basis on $\mathcal{A}$. Following Matthews & Segall (1993), we determine the orthonormal basis functions from the eigenvectors of the Gram matrix $\mathbf{\Gamma}$, which has elements $\gamma_{ij} = \langle \Phi_i, \Phi_j \rangle_{\mathcal{A}}$. Explicitly,

$$\eta_i(t) = \lambda_i^{-1/2} \sum_{j=1}^{M} v_{ji} \Phi_j(t), \tag{25}$$

where v_{ji} represents the jth component of the ith eigenvector $\mathbf{v}_i$, and $\mathbf{v}_i$ and the corresponding eigenvalues λ_i satisfy $\mathbf{\Gamma v}_i = \lambda_i \mathbf{v}_i$. Fig. 2 shows the first five orthobasis functions over the interval 0–10 yr, derived from the complete set of amplitude deviation representers for daily measurements in this interval. The orthobasis functions are ordered such that the least complex basis functions attain the largest amplitudes and are associated with the largest eigenvalues.

2.2 Solution algorithm

Let $\mathbf{n} \in \mathbb{R}^N$ be the column vector with elements $c_1, \ldots, c_N$. Substitution of (24) into the performance functional (9) leads to the minimization problem

$$\Psi_\rho(\hat{\mathbf{m}}, \hat{\mathbf{n}}) = \min_{\mathbf{m},\mathbf{n}} \Psi_\rho(\mathbf{m}, \mathbf{n}). \tag{26}$$

The solution to this problem requires that we address two fundamental issues. First, the functionals ϕ_i depend on the unknown initial phase θ, as noted above. Thus, $\Psi_\rho(\mathbf{m}, \mathbf{n})$ is not a linear function of the unknown parameters and linear inverse theory is strictly not applicable. Second, determination of a preferred solution requires the specification of a preferred $\rho = \rho^*$.

We address the first of these issues by pre-solving the well-posed problem,

$$\Psi_\infty(\tilde{\mathbf{m}}) = \min_{\mathbf{m}} \frac{1}{\sigma^2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|^2, \tag{27}$$

where $\mathbf{d} \in \mathbb{R}^M$ is the column vector with elements $d_1, \ldots, d_M$, the rows of the $M \times 4$ matrix $\mathbf{G}$ are given by the vectors $\mathbf{g}_i$ defined above and $\|\cdot\|$ represents the Euclidian norm. The solution vector $\tilde{\mathbf{m}}$ contains the parameter estimates $\tilde{x}_0, \tilde{v}, \tilde{a}, \tilde{b}$. For the case of more general error variance–covariance matrix $\sigma^2\mathbf{V}$, replace $\mathbf{d}$ and $\mathbf{G}$ with $\mathbf{L}^{-1}\,\mathbf{d}$ and $\mathbf{L}^{-1}\,\mathbf{G}$, respectively, where $\mathbf{L}$ is the lower diagonal Cholesky square root of the variance–covariance matrix, that is, $\mathbf{V} = \mathbf{L}\mathbf{L}^T$. Below, we refer to the problem (27) and the solution $\tilde{\mathbf{m}}$ as the traditional problem and solution, respectively.

We then estimate the full solution represented by $\hat{\mathbf{m}}$ and $\hat{\mathbf{n}}$ as follows. First, we set $\theta = \tilde{\theta} = \tan^{-1}(\tilde{b}/\tilde{c})$. Second, we substitute θ and eq. (24) into the performance functional (9). Finally, we solve the resulting minimization problem for $\hat{\mathbf{m}}$ and $\hat{\mathbf{n}}$ for a given ρ. With these substitutions, we may rewrite eq. (9) as

$$\Psi_\rho(\mathbf{z}) = \|\mathbf{d} - \mathbf{A}\mathbf{z}\|^2 + \rho\sigma^2\|\mathbf{n}\|^2, \tag{28}$$

where $\mathbf{A} = [\mathbf{G}\ \ \mathbf{\Gamma}]$ and $\mathbf{z}$ is the column vector formed by concatenation of the column vectors $\mathbf{m}$ and $\mathbf{n}$. $\mathbf{\Gamma}$ is the $M \times N$ matrix with ijth entry given by $\gamma_{ij} = \phi_i(\eta_j)$. For each value of ρ, there is a unique solution $\hat{\mathbf{z}}(\rho)$ minimizing the performance functional (28):

$$\Psi_\rho(\hat{\mathbf{z}}) = \min_{\mathbf{z}} \Psi_\rho(\mathbf{z}). \tag{29}$$

The result of this minimization is

$$\hat{\mathbf{z}}(\rho) = \left(\mathbf{A}^T\mathbf{A} + \rho\sigma^2\mathbf{B}\right)^{-1} \mathbf{A}^T\mathbf{d}, \tag{30}$$

where $\mathbf{B}$ is the $(N+4)$-square diagonal matrix with first four entries equal to zero and the remaining N entries equal to 1. For the case of a more general error variance–covariance matrix $\sigma^2\mathbf{V}$, replace $\mathbf{d}$ and $\mathbf{A}$ with $\mathbf{L}^{-1}\mathbf{d}$ and $\mathbf{L}^{-1}\mathbf{A}$, respectively. Below, we refer to the problem (28) and the solution $\hat{\mathbf{z}}$ alternately as the new, semi-parametric, or full problem and solution, respectively.

$\hat{\mathbf{z}} = \hat{\mathbf{z}}(\rho)$ depends critically on the regularization parameter ρ. For larger values of ρ, fluctuations in *c(t)* will be small in magnitude and smoothly varying through time, whereas for smaller values of ρ fluctuations will be larger in amplitude and more irregular through time. From eq. (28) it is clear that σ^2 plays the same role as ρ. For fixed ρ, estimators deriving from less precise data will exhibit less variation. Estimators based on more precise data may be more variable. In practice, however, one simply estimates a scaled regularization parameter $\rho' \equiv \rho\sigma^2$ for the case $\sigma^2 \neq 1$.

Because we re-estimate a and b in the full solution, one may wonder if it is necessary to iterate until the *a priori* value of θ used to calculate the representers is equal to the *a posteriori* estimate $\hat{\theta} = \tan\hat{b}/\hat{a}$. In our experience, however, the phase estimates $\tilde{\theta}$ and $\hat{\theta}$ are equal within their uncertainties as of the first full solution. Therefore, we have found no need for iteration. Nevertheless, we make no claim regarding the generality of this result and advocate that it be verified with each application.

2.3 Choosing the solution

There are many algorithms in the literature for determining the 'best' value of ρ based on principles of statistics. We here consider two popular prescriptions. One method, which is commonly applied in the geophysics literature, is known as cross validation. The method is based on the notion that the best model will be the one that best

predicts independent subsets of the data. One recipe for assessing the predictive power of the model is to purposely withhold each data point one by one during a series of parameter estimate determinations. The best model is then considered to be the model that minimizes the overall sum of squared prediction errors. Direct application of this method would be costly in that it would require estimation of one model for each data point for each candidate value of ρ. Fortunately, there is a significantly less costly solution that involves evaluation of one model using the full data set for each value of ρ (e.g. Wahba 1990). The best choice ρ^* according to 'leave-one-out' cross validation is the one that minimizes the cross-validated sum of squared errors CVSS given by

$$\mathrm{CVSS}(\rho) = \sum_{i=1}^{M} \left[\frac{d_i - \hat{d}_i(\rho)}{1 - h_{ii}(\rho)} \right]^2, \tag{31}$$

where $\hat{d}_i(\rho)$ is the ith component of the vector $\hat{\mathbf{d}}(\rho) = \mathbf{A}\hat{\mathbf{z}}(\rho)$, containing the predicted values of the data points, given the model parameters $\hat{\mathbf{z}}(\rho)$ associated with a given value of ρ and $h_{ii}(\rho)$ are the diagonal elements of the matrix $\mathbf{H}(\rho) = \mathbf{A}(\mathbf{A}^T\mathbf{A} + \rho\mathbf{B})^{-1}\,\mathbf{A}^T$. Some advantages of the CVSS statistic are that it is fairly straightforward to calculate and has an intuitively appealing interpretation. Another satisfying property of CVSS is that, for large M, it produces results that are equivalent to those determined by minimizing the Akaike Information Criterion (AIC; Akaike 1974) with philosophical foundations in information theory (e.g. Boltzmann 1877; Fisher 1922; Shannon 1948; Kullback & Leibler 1951). For our problem, the AIC score is

$$\mathrm{AIC} = M \log\left(\frac{\|\mathbf{d} - \mathbf{A}\mathbf{z}\|^2}{M} \right) + 2\,\mathrm{tr}\mathbf{H}(\rho), \tag{32}$$

where the notation tr$(\cdot)$ indicates the matrix trace operation. Models minimizing AIC are said to suffer the least Kullback–Leibler information loss relative to the true model.

One important drawback of the AIC/CVSS prescription is that it performs poorly when the model is not known exactly or when the errors in the data are correlated (e.g. Altman 1990; Hart 1991). Thus, the AIC/CVSS criterion works well for simulations where both model and errors are known but can be unstable when applied to real data. Instability is obvious in that the 'optimal' model overpredicts the data. This is apparent in (1) the value of ρ^*, which is so small that the effective number of model parameters approaches M; (2) the power spectral density of residual-errors, which attains a positive slope, and (3) the resulting fit to the data, which shows that the data points are interpolated with little smoothing (averaging) of errors.

The Bayesian Information Criterion (BIC; Schwarz 1978) overcomes this limitation of the AIC and CVSS statistics by more stringently penalizing models with smaller degrees of freedom. Minimization of the BIC statistic selects the model with maximum *a posteriori* probability. For our problem, BIC may be expressed as

$$\mathrm{BIC} = M \log\left(\frac{\|\mathbf{d} - \mathbf{A}\mathbf{z}\|^2}{M} \right) + \log(M)\,\mathrm{tr}\mathbf{H}(\rho). \tag{33}$$

As the number of data approaches infinity, the probability of selecting the true model using BIC approaches certainty.

In Section 3, we apply both the AIC/CVSS and BIC criteria to simulated data sets to demonstrate their utility and assess their differences. In Section 4, we apply only the BIC criterion to a real GPS time-series. Fig. 3 shows an example of the dependence of AIC, CVSS and BIC on ρ, for estimators based on simulated data representing a 10-yr time-series. The specifics of the simulation are described in detail in the next section. The statistics in Fig. 3 have been adjusted by subtracting their minimum values such that the adjusted minimum is zero for both to facilitate comparison. The close agreement between the values of ρ associated with these minima for AIC and CVSS informally demonstrates their general interchangeability. BIC results in a slightly larger value of ρ^* which produces smoother models.

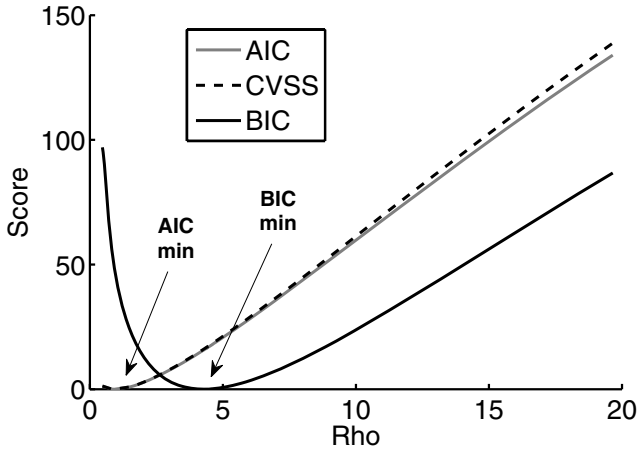

Figure 3. AIC, CVSS and BIC regularization statistics defined in the text. Minimization of any of these statistics determines an 'optimal' estimate for the smoothing parameter ρ. The BIC score consistently penalizes models with fewer degrees of freedom more heavily, resulting in larger values for ρ^* and smoother estimators of amplitude deviation. BIC appears to be more stable when applied to real data.

2.4 Assessing error variance

Throughout this paper, we use the term 'residual-errors' in reference to the predicted errors $\tilde{e} = \mathbf{d} - \mathbf{G}\tilde{\mathbf{m}}$ and $\hat{e} = \mathbf{d} - \mathbf{A}\hat{\mathbf{z}}$ to emphasize their dependence on the model used to reduce the data. Residual-errors are commonly used to investigate the nature of the true observation-errors e, as if the model for site motion were known exactly. Residual-errors will be correlated if the model for kinematics is incomplete, or if the observation-errors are correlated. For each of the estimations that we consider below, we assess the power spectral density of the residual-errors to illustrate their temporal correlations.

It is also useful to consider the *a posteriori* estimates for the observational-error variance based on the assumption of an exact model. The traditional estimate of the *a posteriori* variance $\tilde{\sigma}^2$ is

$$\tilde{\sigma}^2 = \frac{1}{\sigma^2} \frac{\|\mathbf{d} - \mathbf{G}\mathbf{m}\|^2}{M - 4}. \tag{34}$$

An estimate for the *a posteriori* variance $\hat{\sigma}^2$ associated with the cross-validated time-variable analysis is provided by Wahba (1990):

$$\hat{\sigma}^2 = \frac{1}{\sigma^2} \frac{\|\mathbf{d} - \mathbf{A}\hat{\mathbf{z}}(\rho^*)\|^2}{\mathrm{tr}\,[\mathbf{I} - \mathbf{H}(\rho^*)]}. \tag{35}$$

3 NUMERICAL EXAMPLES

In this section, we apply the theory developed in Section 2 to three synthetic data sets. The goals of this section are to (1) demonstrate the utility of the theory presented above to probe the effects of quasi-periodic loads; (2) assess the consequences of neglecting annual-amplitude variation on the power spectral density of residual time-series (i.e. the residual-error spectra) and (3) assess secular

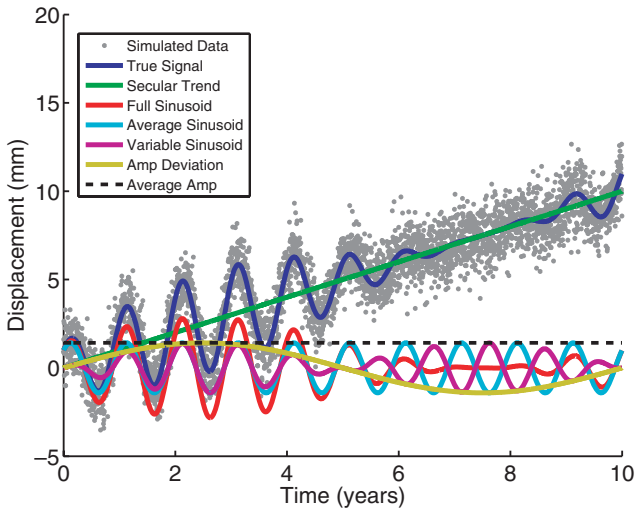

Figure 4. Simulated data set representing the sum of several contributing factors including a constant velocity term, annual motion with constant amplitude, annual motion with time-variable amplitude and errors sampled from a zero mean Gaussian distribution of unit variance. The time variation of the annual amplitude is modelled as sinusoid with period of 10 yr and amplitude equal to that of the constant part of the annual signal.

velocity biases when using short time-series in the presence of long wavelength variations in the amplitude of annual signals. A more elegant analytical assessment of these questions along the lines of the study of Blewitt & Lavallée (2002) is complicated for our problem due to the larger number of model parameters. Instead, our investigation relies on numerical experiments.

The simulated data are composed of several parts, including (1) a secular trend determined from a constant velocity; (2) constant amplitude purely periodic motion representing the time-averaged part of the quasi-periodic signal and (3) another sinusoidal variation representing the time-variable part of the quasi-periodic signal. This latter, time-variable signal, has an amplitude that varies with decadal period. Each of these components of the signal is illustrated in Fig. 4.

Fig. 4 shows the longest of the synthetic data sets that we consider. The simulated data are intended to represent CGPS coordinate time-series over a 10 yr span of time. The data were generated using eqs (1) and (2) plus errors sampled from a zero mean Gaussian distribution with unit variance (in units of mm^2). Different values for the *a priori* variance σ^2 would have an effect on the values of ρ^*, which we estimate below, and the *a posteriori* estimate of the variance but not on the parameter estimates themselves.

We assume one measurement per day, which is typical of CGPS time-series. The parameters of the kinematic model are listed in Table 1. For simplicity, all parameter values used in the simulation were either zero (y-intercept in units of mm), or one (secular velocity in units of mm yr^{-1}, amplitudes of individual periodic terms in units of mm). The total magnitude of the time-averaged part of the periodic signal is thus $\sqrt{1+1} \approx 1.4$ mm. We chose to model the time-variable deviation function as a sine function with 10 yr period and amplitude equal to the amplitude of the time-averaged component. We make no claim that this choice of amplitude deviation is more realistic than any other potential deviation. The initial phase is $\theta = \tan^{-1}(1) \simeq 0.785$.

We also consider two shorter time-series of duration 2.5 and 1.5 yr. We use the same model as for the longer 10-yr time-series described above, but independent error sequences were used for all three series. These shorter time-series serve to illustrate the effects of decadal amplitude variation on the estimates for secular velocity and show how the semi-parametric estimator performs on shorter series for the ideal case of uncorrelated measurement errors. We use the longer 10-yr time-series to explore the effects of neglecting the decadal variation on the inferred power spectrum. We also explore the ability of the estimator to recover decadal variations in the amplitude of periodic signals, again under the ideal case of uncorrelated measurement errors.

We begin our discussion with the 10-yr time-series shown in Fig. 4. The period of ten years was chosen for two reasons. First, 10 yr represents one complete cycle for our assumed decadal amplitude deviation function. Second, and more importantly, a large number of continuous GPS stations installed in the mid-1990s have recently achieved their tenth year of operation, with relatively few stations having been in operation for much longer than ten years.

Table 1. Simulated and estimated model constants and variances.

Model param	True value	Traditional method estimates 1.5 yr	2.5 yr	10 yr	New method estimates[a] 1.5 yr	2.5 yr	10 yr
x_0	0.00	−0.19 (0.09)	−0.25 (0.07)	0.02 (0.03)	−0.03 (0.13)	−0.07 (0.09)	0.04 (0.04)
					−0.04 (0.11)	−0.11 (0.08)	0.02 (0.04)
v	1.00	1.24 (0.10)	1.19 (0.05)	1.00 (0.01)	0.96 (0.14)	1.05 (0.06)	0.99 (0.01)
					0.99 (0.12)	1.08 (0.05)	0.99 (0.01)
a	1.00	1.02 (0.06)	0.96 (0.05)	1.00 (0.02)	1.08 (0.07)	0.99 (0.05)	0.99 (0.02)
					1.07 (0.06)	0.98 (0.05)	0.99 (0.02)
b	1.00	0.99 (0.06)	1.09 (0.05)	0.98 (0.02)	1.02 (0.06)	1.11 (0.05)	0.98 (0.02)
					1.01 (0.06)	1.11 (0.05)	0.98 (0.02)
θ	0.79	0.77 (0.06)	0.85 (0.05)	0.78 (0.02)	0.75 (0.06)	0.84 (0.05)	0.78 (0.02)
					0.76 (0.06)	0.84 (0.05)	0.78 (0.02)
σ^2	1.00	1.17	1.08	1.49	1.10	0.99	1.00
					1.12	0.99	1.02
SSR[b]		639	977	5444	599	891	3661
					605	899	3697
P[c]		4	4	4	7	9	23
					6	6	13

Note: The numbers in parentheses are the standard deviations.
[a] There are two values listed for each parameter associated with the AIC/CVSS and BIC solutions, respectively. [b] SSR is the sum of squared residual-errors. [c] P is the effective number of model parameters.

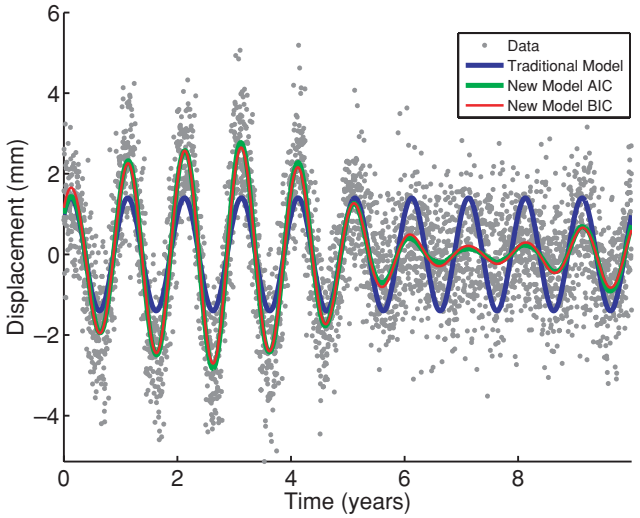

Figure 5. Model fits to the synthetic data set of Fig. 4 for both the traditional model and the new semi-parametric models developed in Section 2. Semi-parametric models are shown for both the AIC/CVSS and BIC regularization criteria. The synthetic data have been detrended.

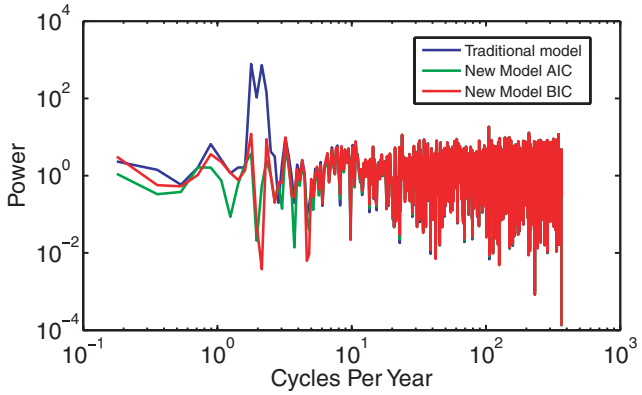

Figure 6. Power spectral density function estimates for the misfit to the synthetic data of Fig. 4 for both the traditional model and the new semi-parametric models based on AIC/CVSS and BIC. All three of the spectral density estimates are nearly indistinguishable for frequencies higher than about 4 cycles yr^{-1}. There is a large amount of power in the frequency band 0.2 to 2 cycles yr^{-1} in the residuals associated with the traditional model that is not present in the residuals associated with the semi-parametric models. This result suggests that assessments of the error spectra of GPS time-series based on residual fits to time-series may be biased if sites experience periodic motions with time-variable amplitudes unless such signals are accounted for.

Fig. 5 shows best fit models to the 10-yr time-series based on the traditional constant amplitude approach represented by the minimization problem (27) and our time-variable semi-paramteric approach represented by problem (29). Semi-parametric models are shown for both the AIC/CVSS and BIC choices of ρ^*. Table 1 lists the parameter estimates and their uncertainties. It is clear from this example that the best fit model resulting from the new approach is significantly more variable than that of the traditional approach for both AIC/CVSS and BIC. The amplitude of the seasonal variation in the semi-parametric model is as much as a factor of two larger than the traditional model during the early period between ∼2 and 2.5 yr and is near-zero later in the time-series. These variations coincide with the decadal variation in the amplitude deviation function described in the previous paragraph and shown in Fig. 4. Whereas the traditional solution includes estimates for four independently determined parameters, the semi-parametric models contain estimates for approximately 23 and 13 independently determined model parameters for AIC/CVSS and BIC, respectively. It is interesting to note that had we used our knowledge of the mathematical form of the simulated amplitude deviation, we could have estimated parameters describing the quasi-periodic motion using a total of 8 model parameters (x_0, v, a, b and four additional coefficients describing the amplitude variation). In general, however, the form of the amplitude deviation function is unknown.

Various aspects of the misfit of the model may be considered. The pointwise match of the model to the data is significantly better using the new semi-parametric method relative to the traditional method; the weighted sum of squared residuals (SSR) for the traditional method is 5444, whereas the SSR for the new time-variable method is appreciably smaller at 3661 and 3697 for AIC/CVSS and BIC solutions, respectively. Errors in the estimates for secular velocity v and trend of site motion $x_0 + vt$ are comparable for this solution, reflecting the fact that sinusoidal signals tend to average to zero with longer time-series (e.g. Blewitt & Lavallée 2002), even for the case of non-constant amplitudes. However, this averaging effect does not apply to statistical measures of misfit, for example the *a posteriori* estimate for the variance of the data errors. The traditional estimator $\tilde{\sigma}^2$ overestimates the true error variance by 50 per cent (1.5 mm^2), whereas the new estimator $\hat{\sigma}^2$ provides an accurate estimate of 1.0 mm^2 for both AIC/CVSS and BIC models (Table 1). Fig. 6 shows the power spectral density estimates for the residuals associated with the traditional and semi-parametric models. Power spectra are similar for all three models, with the exception of a broad band of frequencies centred on the semi-annual frequency for which the traditional errors have appreciably more power. Residual (positioning error) time-series are commonly used to infer the noise characteristics of GPS measurements. Such estimates depend on the model that was used to reduce the observed time-series to residual time-series. As Fig. 6 shows, omission of the time-dependent part of periodic signals leaves appreciable power in the lower frequencies.

An important feature of the semi-parametric estimator is that it decomposes rather complex signals, such as that shown in Fig. 4, into their constituent parts. Of particular interest is the possibility of learning something about the variability in the amplitude of environmental (or other) loads, as characterized by the deviation function c. The estimate for the deviation function resulting from the analysis of our 10 yr simulation is shown in Fig. 7. As this figure illustrates, the estimators do a remarkable job of tracking the variation in the amplitude of the quasi-periodic signal, with typical point uncertainty between 0.1 and 0.2 mm.

Fig. 8 shows examples of the dependence of the regularization statistics on ρ, for estimators based on the simulated data over 2.5 (Fig. 8a) and 1.5 yr (Fig. 8b) intervals. The specifics of the simulation are described in detail above. As in Fig. (3), the statistics in Fig. 8 have been adjusted by subtracting their minimum values such that the adjusted minimum is zero for both to facilitate comparison. The close agreement between the values of ρ corresponding to these minima for AIC and CVSS informally demonstrates the general interchangeability of these statistics even for relatively short time-series. Optimal values of ρ associated with the BIC minima are larger than the values obtained by AIC/CVSS for both the 2.5 and 1.5 year intervals.

Figs 9(a) and (b) show the 2.5- and 1.5-yr time-series associated with the statistics presented in Figs 8(a) and (b), respectively. The parameter estimates associated with the inversions using these

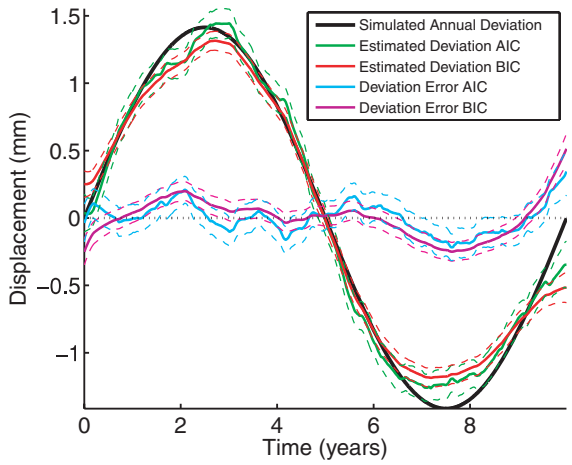

Figure 7. Deviation function inferred from the synthetic data set of Fig. 4 estimated using the semi-parametric analysis method described in the text. The true deviation function and the misfit of the deviation function estimates (labelled Deviation Error) are shown for comparison. The dashed lines represent one standard deviation error bounds. The estimator of amplitude deviation does a reasonable, though not perfect, job of recovering the true function to within error.

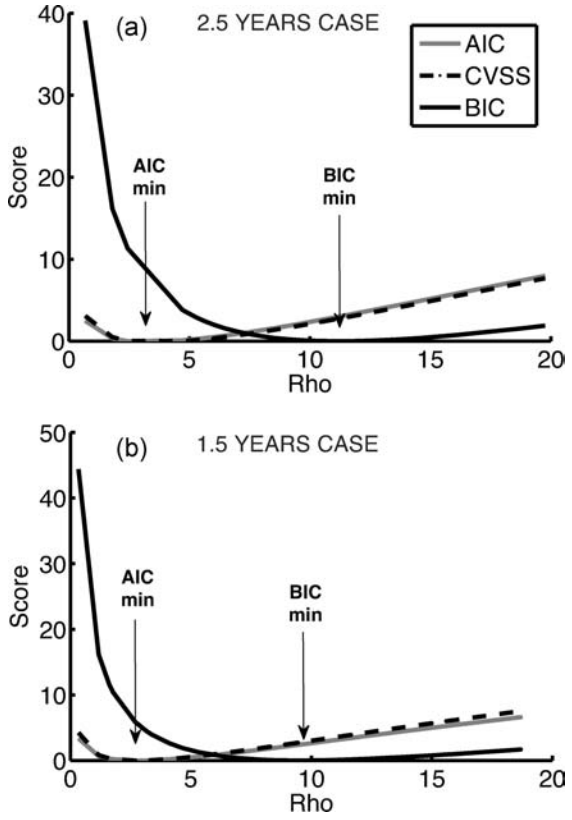

Figure 8. Regularization parameter estimates for synthetic (a) 2.5- and (b) 1.5-yr duration time-series. The close agreement between the AIC and CVSS statistics is apparent even when the time-series under investigation have short duration.

time-series are listed in Table 1. As for the case of the full 10-yr time-series, we also plot the constituent parts of the synthetic signal, including a secular trend, a constant amplitude sinusoidal term with annual period, a time-variable sinusoid and the corresponding

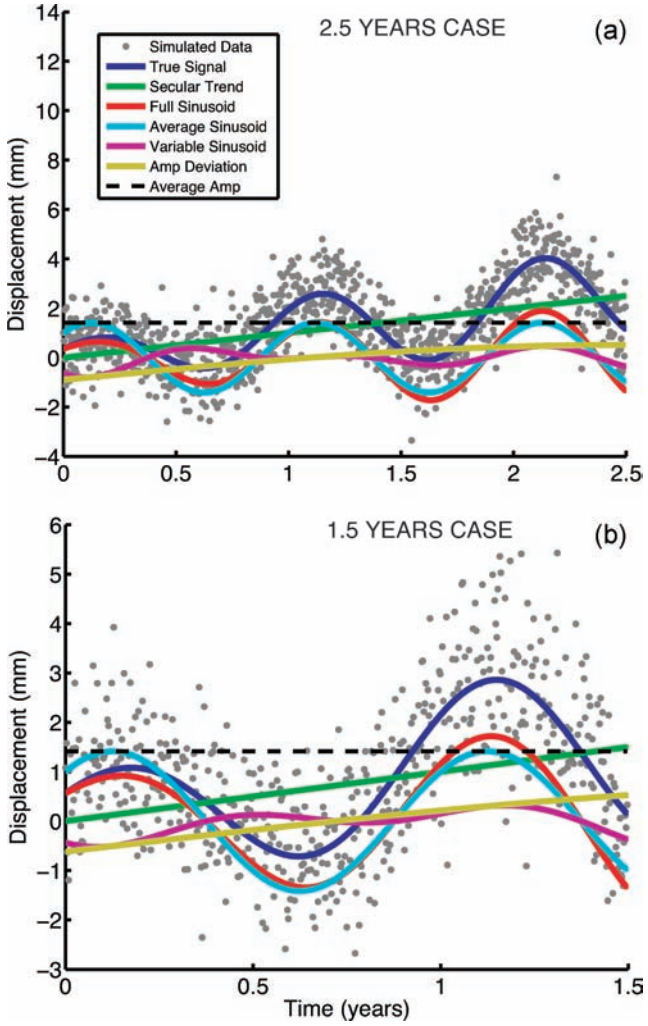

Figure 9. Synthetic time-series for (a) 2.5 and (b) 1.5 yr intervals using the same kinematic model as for the the series in Fig. 4 but with differing random error sequences.

amplitude deviation function. Figs 10(a) and (b) show the results of the traditional and semi-parametric inversions for these subsets of data. As was observed using the 10 yr series, the semi-parametric models exhibit significantly more structure than the models obtained by traditional means. The effective number of independently determined parameters P for the time-variable models are listed in Table 1.

The pointwise match of the model to the data is significantly better, using the new method relative to the traditional method. The SSR are listed in Table 1. In the case of the 1.5 yr series, the velocity error is within a few standard deviations, but for the 2.5-yr time-series the velocity error associated with the traditional estimate is greater than 3 standard deviations (>99.7 per cent confidence). Thus, the random variable $\hat{v}$ associated with the traditional method has a mean that almost surely differs from the true velocity, that is, $\hat{v}$ is a biased estimator. Errors in the estimates for secular velocity are significantly reduced using the semi-parametric model. Errors in positioning and trend are also reduced significantly. These results demonstrate that decadal variations in annual amplitude, if unaccounted for, could bias velocity estimates based on short time-series, even if the time-averaged part of the repeating signal's amplitude is accounted for. The amplitude deviation function that

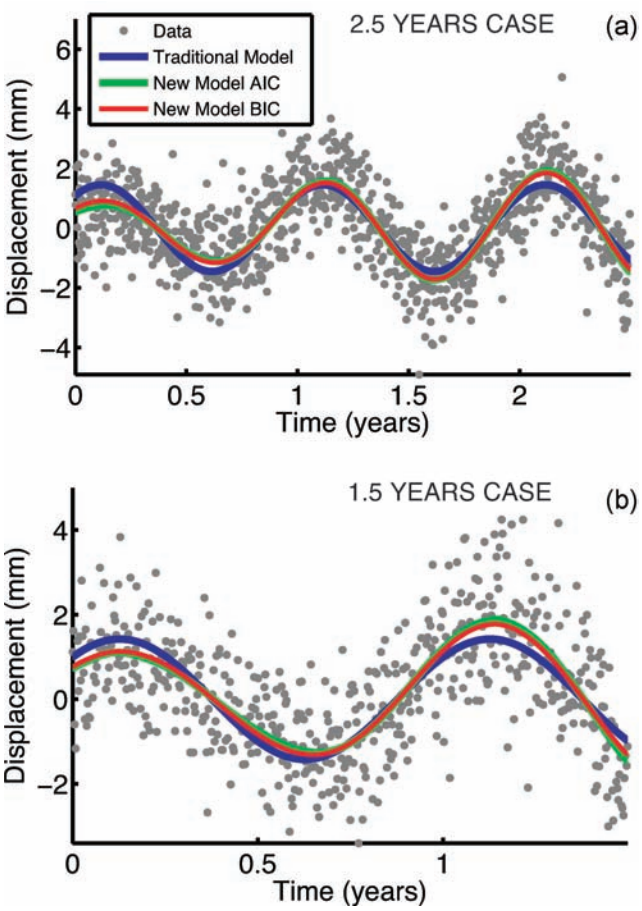

Figure 10. Model fits to the synthetic data sets of Figs 9(a) and (b), respectively, for both the traditional model and the semi-parametric models based on AIC/CVSS and BIC regularization criteria. The difference in amplitude of annual motions between the models is subtle and could be difficult to identify by visual inspection of the observed time-series prior to modelling.

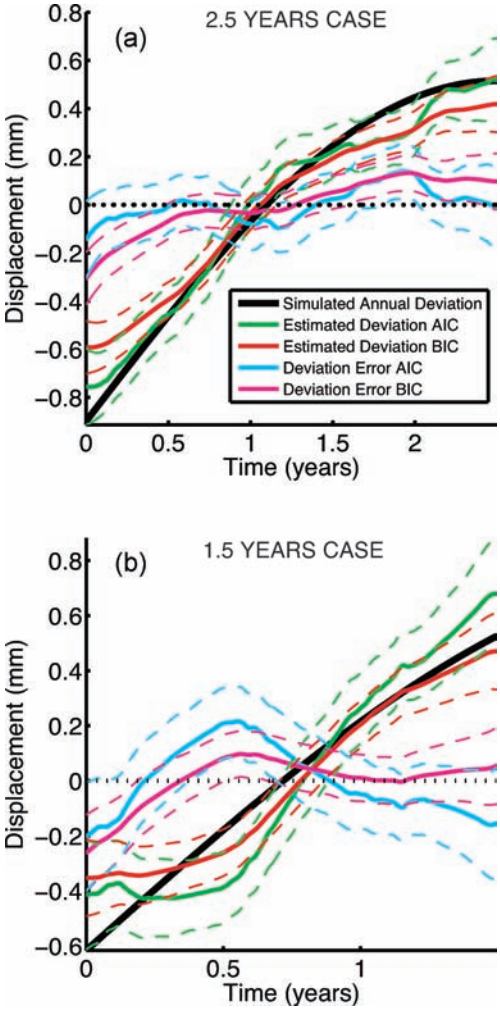

Figure 11. Deviation function estimators for the synthetic data set of Fig. 9, estimated using the semi-parametric analysis method described in the text based on AIC/CVSS and BIC regularization criteria. The true deviation function and the misfit of the deviation function estimates (labelled Deviation Error) are show for comparison. The dashed lines represent one standard deviation error bounds. The estimators of amplitude deviation do a reasonable job of recovering the true function to within error even for short time-series for both AIC/CVSS and BIC.

we used for simulation was arbitrary, but we surmise that a broad class of plausible amplitude deviations could bias velocity estimates. Further study is required to assess the general importance of such effects.

Fig. 11 shows the resulting deviation function estimates for the 2.5- and 1.5-yr time-series together with the actual deviation functions used in the simulations and the associated model misfits. Overall the deviation function estimators provide close matches to the true functions even for these rather short time-series; the errors in the estimates are most of the time smaller than one standard deviation. The errors are larger near the beginning of the interval where the simulated deviation function has the steepest slope. The largest errors occur for the smoother BIC model. Fig. 12 shows the power spectral density estimates for the residuals associated with both the traditional model and the semi-parametric model for the 2.5 and 1.5 yr cases. Power spectra are similar for both inversion methods with the exception of the lowest frequencies. Due to the short duration of the time-series, however, the limited bandwidth of the error arising from unmodelled time-variable annual amplitude is not as strikingly apparent as for the case of the 10-yr time-series shown in Fig. 6. For the shorter time-series it might therefore be difficult to differentiate power-law residual-error spectra from bandpass filtered spectra.

4 APPLICATION TO REAL DATA

In this section, we apply the new analysis method to a real CGPS time-series. For demonstration, we have selected a long-running CGPS station CCCS located in southern California. The station was constructed as part of the Southern California Integrated GPS Network. The monument consists of a set of five deep-anchored rods set in an alluvial fan. The station is located along the northeastern edge of the Santa Ana aquifer system and experiences very large (4–5 mm) amplitude quasi-annual variation in site motion (Fig. 13) (Argus *et al.* 2005). We here consider time-series for the north component of position for site CCCS provided by Scripps Orbit and Permanent Array Center (SOPAC). We corrected time-series for three jumps associated with earthquakes or other disturbances identified in SOPAC's 'Refined Model' for site kinematics. These were 3.0 ± 0.3 mm at epoch ∼1999.8, 0.3 ± 0.3 mm at epoch ∼2006.0 and 14.1 ± 0.3 mm at epoch ∼2006.8. We have selected this site and time-series solution for two reasons. First, this site exhibits clear quasi-periodic motion. Second, Langbein

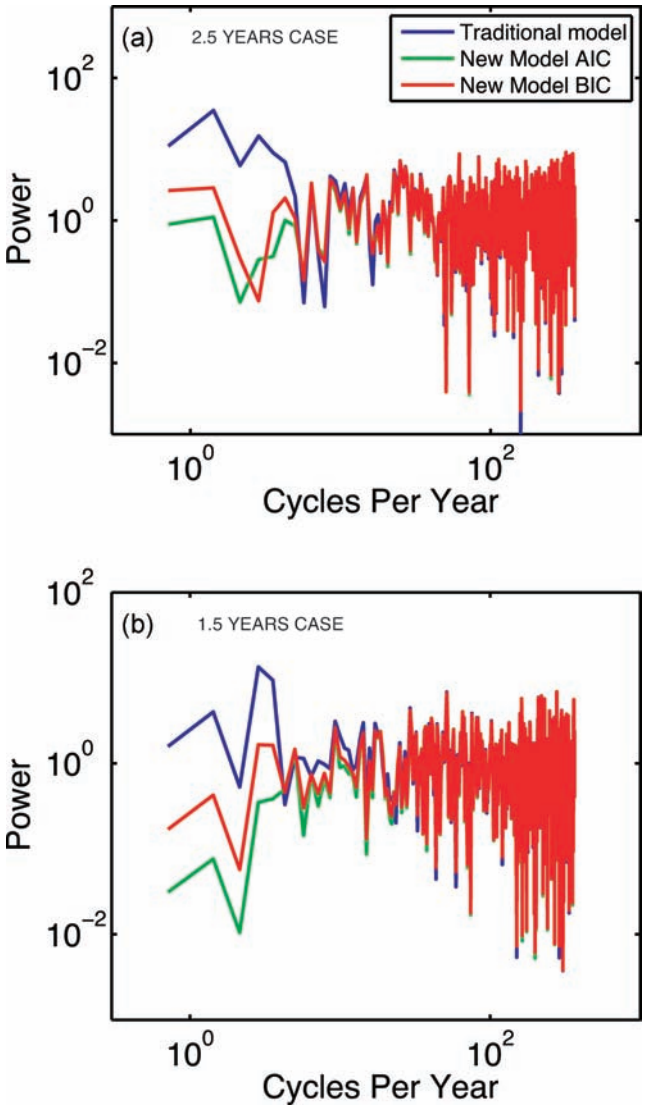

Figure 12. Power spectral density function estimates for the misfit to the synthetic data of Fig. 9 for both the traditional model and the semi-parametric models. All three of the spectral density estimates are nearly indistinguishable for frequencies higher than about 4 cycles yr^{-1}. There is very little difference between the spectra among the models due to the short duration of the time-series.

(2008) analysed the residual time-series for this site as part of his analysis of SOPAC time-series, finding that the residual-error spectrum for CCCS was best described as a combination of bandpass filtered and power-law noise.

For this analysis of real data, we included terms for repeating sinusoids at two periods, representing annual and semi-annual variation. Thus, for both the traditional and new semi-parametric methods, we estimate six constant parameters (x_0, v, a_a, b_a, a_s, b_s), where a_a and b_a represent the constant amplitudes of the annual variation and a_s and b_s represent the amplitudes of the constant part of the semi-annual terms. For the new semi-parametric method, we also estimate two deviation functions c_a and c_s, respectively, in addition to the constant parameters. We used one regularization parameter determined by BIC to damp both deviation functions. The fit to the data achieved using a single regularization parameter is sufficient to illustrate our main points.

Fig. 13 shows time-series for the north component of CCCS site position together with the fitted models resulting from the traditional and the new semi-parametric analysis approaches. The traditional model clearly does a poor job of tracking the observed variation in site motion. The SSR for the traditional model is 11 623 and the *a posteriori* variance estimate for this model is $\tilde{\sigma}^2 = 1.9$ mm. The semi-parametric model, on the other hand, is much more capable of tracking the variations; the SSR is 792 and the *a posteriori* estimate for the standard deviation associated with the new model is $\hat{\sigma}^2 = 0.5$ mm, nearly a factor of 4 smaller than for the fit of the traditional model. The effective number of model parameters for the semi-parametric model is ∼202. The significantly improved fit to the data provided by the semi-parametric model qualitatively suggests that our restriction to constant phase θ was an acceptable approximation, at least for this site.

The difference between secular velocity estimates for the traditional and new models is -0.08 mm yr^{-1} (traditional minus new). The uncertainty in both is 0.03 mm yr^{-1}. This rate difference, although rather small, is significant, given that both estimates used the same data and are thus highly correlated. Argus *et al.* (2005) estimated a correction for the velocity of site CCCS due to non-tectonic deformation associated with hydrology of the Santa Ana aquifer. Their correction is 0.8 mm yr^{-1} oriented 113° counter-clockwise from north. The north component of this anthropogenic motion is -0.3 mm yr^{-1}. Our estimated difference is in the same direction (southwards) though ∼4 times smaller than Argus *et al.*'s correction, resolved onto the north coordinate direction. Southward motion is consistent with long-term subsidence of the basin, which pulls stations on the periphery of the basin in towards the area of maximum

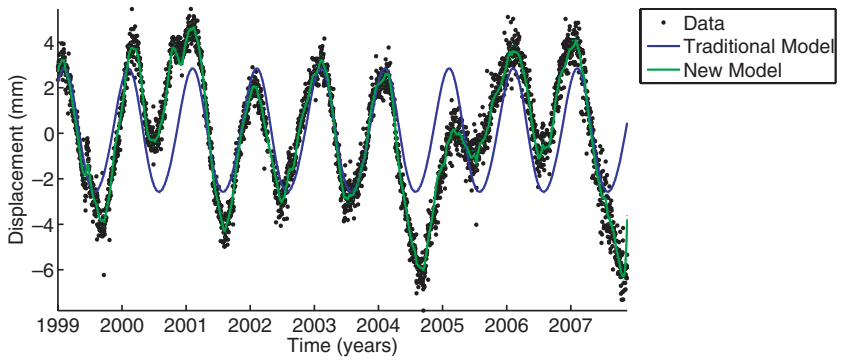

Figure 13. Time-series for the north component of position for CGPS site CCCS located in southern California. Also shown are the best fit models resulting from the traditional and semi-parametric (BIC) analysis methods. Both models include terms for both annual and semi-annual site motion.

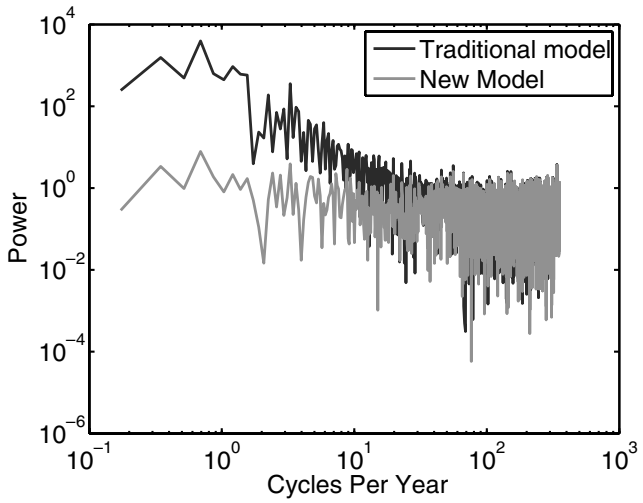

Figure 14. Power spectral density function estimates for the residuals to the model fits to the north component time-series for site CCCS. There is an appreciable difference between the traditional and semi-parametric model residual power spectra at low frequencies.

subsidence. Secular subsidence is expected as fine grained silt and clay layers are irreversibly compacted. There are two main reasons for the difference between Argus *et al.*'s correction and the velocity estimate difference that we compute between traditional and new models. First, the Argus *et al.* correction was based on the cumulative non-tectonic deformation at site CCCS, whereas the deviation functions of our analysis are modulated to annual and semi-annual periods. Therefore, our estimates may not entirely characterize the cumulative deformation of the basin which could have long-period and/or secular components. Second, Argus *et al.*'s analysis was based on a 3 yr subset of the 8-yr time span considered in our analysis. We expect any potential differences between the SOPAC GPS solution considered here and the Jet Propulsion Laboratory solution considered by Argus to be insignificant relative to these other two effects.

The power spectral density of residual-errors associated with the traditional and semi-parametric models for CCCS are shown in Fig. 14. The mistfit of the new model contains significantly less low frequency power than for the traditional constant parameter model. The most striking difference occurs in the frequency band of ∼0.2 to ∼2 cycles yr^{-1}, though there is also an appreciable difference in power in the higher frequency range of ∼2 to ∼20 cycles yr^{-1}. The misfit spectra at the highest frequencies (>20 cycles yr^{-1}) are comparable between the two solutions, reflecting the underlying (mostly) uncorrelated error.

Fig. 15 shows the estimates for the quasi-periodic part of the model. Unlike the simulation study presented in the previous section, wherein we considered deviations with 10 yr period, the interannual variability for both annual and semi-annual deviation function estimates are quite high. Although there are periods of time where both deviation functions have small magnitude relative to the constant amplitude part of the annual variation, the deviations occasionally attain large values. The pattern is complex, suggesting that the underlying processes responsible for the deviations are themselves complex. The largest anomaly in the annual variation occurs during the late-Fall and Winter of 2004–2005. The largest anomaly in semi-annual variation occurs during the late 2007 period, coincident with the largest departures in the observed time-series from their time-averaged values. It is notable that the southern California region experienced 947 mm of rain during the 2004–2005 season, the second highest annual total since 1878, and water levels in wells near the bases of local mountain ranges rose by 10s of metres (King *et al.* 2007). However, this wet year was followed by one of the driest years on record for 2005–2006 and more normal amounts of rainfall during the 2006–2007 water year. Moreover, the Santa Ana aquifer is managed by the Orange County Water District, who recharge the basin using water from the Santa Ana river (Argus *et al.* 2005). Thus, high correlation between levels of rainfall and estimated amplitude deviations are not necessarily expected for site CCCS. A more meaningful physical interpretation of the deviation functions will require a detailed investigation of site motion in three dimensions for a network of stations and is beyond the scope of this paper.

5 DISCUSSION AND CONCLUSIONS

Explanation of observed periodic and quasi-periodic motions in terms of loading sources is challenging. The problem has been explored in some detail in several recent studies. For example, van Dam *et al.* (2007) compared annual signals in CGPS height time-series with annual displacements predicted using gravity data from Gravity Recovery and Climate Experiment (GRACE). van Dam *et al.*

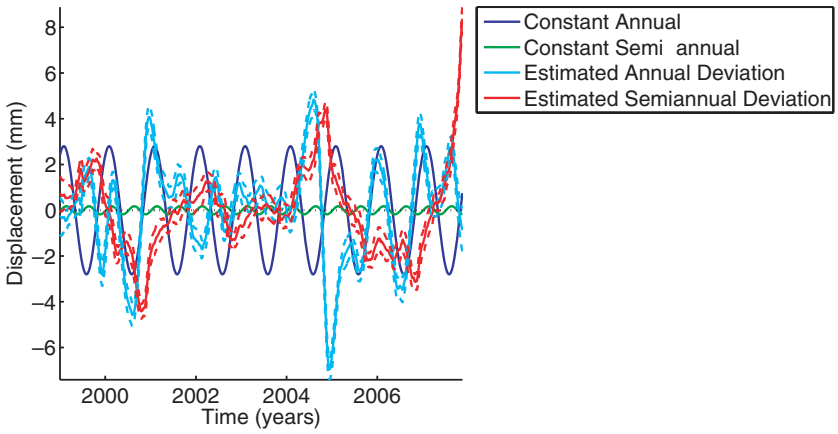

Figure 15. Deviation function estimates for annual and semi-annual variation obtained from the analysis of site CCCS time-series based on the BIC regularization criterion. The dashed lines represent one standard deviation error bounds. Also shown for comparison are the predicted annual and semi-annual motions associated with the constant time-averaged parts of the periodic amplitudes.

found poor agreement in both amplitude and phase between observed and calculated displacements. Dong *et al.* (2002) compared observed GPS coordinate time-series with model predictions for a number of load sources, finding that less than half of the power of the observed annual vertical site motion could be attributed to known loading sources. These studies may indicate that periodic and quasi-periodic motion observed using CGPS is not entirely a result of crustal loading, or that loads inferred by GPS are complicated by aliasing effects. For example, Penna & Stewart (2003), Stewart *et al.* (2005) and Penna *et al.* (2007) have demonstrated that short-period displacements can alias into spurious fortnightly, semi-annual and annual signals in daily GPS time-series. An important objective in developing the method presented above is that it accommodates fairly complex repeating signals, without having to identify and parametrize the source of the signal(s).

Our solution method depends on three main assumptions. First, we assume that the deviation functions vary smoothly through time. Second, we made the approximation that the phase θ is constant, but this restriction could be relaxed by simultaneously estimating amplitude deviations for both sine and cosine terms (with a common period), though at the expense of increased computation time and model variance. Third, we identify one 'best' model by minimizing the BIC statistic, which appears to perform well using both simulated and real data.

The conclusions of this study are that, if periodic motion has time-variable amplitude, then it may result in biased estimates for secular velocity and/or observation-error spectra even if the time-averaged amplitudes of the periodic motions are estimated. Biases in secular rate estimates appear to decrease with time, based on the limited number of solutions presented here, but generalization of this result is not warranted as the persistence of such biases is certain to depend on the nature of the deviation function. However, biases in estimates for error statistics based on residual time-series are not systematically reduced as time-series increase in duration. Error estimates are affected even when using long time-series.

Blewitt & Lavallée (2002) showed how annual signals appear in power spectral density as peaks above the background spectrum at annual frequencies and their harmonics. The power of these peaks decays roughly as an inverse power-law function of frequency with spectral index near 1, similar to flicker noise. Langbein (2008) evaluated temporal correlations in residual CGPS time-series for long-running stations in southern California. He used residual time-series obtained from SOPAC, from which we also obtained the CCCS time-series discussed in Section 4. The durations of the time-series used by Langbein ranged from 2.5 to 10 yr. He found that, even after correcting for annual and semi-annual motions of constant amplitude, power spectral densities of the residual-error for many sites could be interpreted as bandpass filtered noise, centred on seasonal frequencies plus (inverse) power-law noise. Williams *et al.* (2004) analysed residual time-series, corrected for annual and semi-annual signals of constant amplitude, for a super set of the sites analysed by Langbein but for a shorter duration of time on average (<5 yr). Williams *et al.* estimated spectral indices of 0.9 ± 0.4, showing that, as in the Langbein study, the power-law character of residual errors persists even if constant amplitude seasonal motions are accounted for. Williams *et al.* did not estimate bandpass filtered noise.

Langbein also found that, although there is some tendency for residual time-series associated with deep-braced monuments to exhibit lower levels of correlation (and thus smaller standard error in rate), there is also an apparent dependence on site location; the highest levels of correlation (highest standard rate errors) in residual error time-series were found for sites located in regions of active pumping. Finally, Langbein found a marginally significant correlation between annual average rainfall and the level of correlations of residual time-series. Pumping and seasonal groundwater effects on crustal deformation in southern California and Nevada are large and well known and have been studied in some detail using GPS and InSAR (Amelung *et al.* 1999; Bawden *et al.* 2001; Lu & Danskin 2001; Watson *et al.* 2002; Lanari *et al.* 2004; Argus *et al.* 2005; Gourmelen *et al.* 2007; King *et al.* 2007). Spurious signals arising from aliasing associated with GPS processing methods have also been demonstrated theoretically, as well as through simulations and studies of real data (Penna & Stewart 2003; Stewart *et al.* 2005; Penna *et al.* 2007). It is unlikely that all of these periodic signals can be accurately represented by sinusoids of constant amplitude. Based on the body of previous work and the analysis results presented above, we suspect that some fraction of the observed temporal correlations among residual-errors for CGPS time-series for stations in southern California and elsewhere may reflect the unmodelled time-variable part of quasi-periodic site motions driven by meteorological, hydrological or a variety of other potential physical processes distinct from observational error. If so, inferences about the nature of GPS observation-errors based on the power spectral densities of residual time-series may need to be re-evaluated using more general models for site kinematics.

ACKNOWLEDGMENTS

This research was funded by NSF grant EAR-0545519 from the EarthScope Program. Sigrún Hreinsdóttir and Austin Holland made constructive comments on an early version of the manuscript. Mark Brandon suggested the use of the BIC statistic for model regularization. Constructive comments by Simon Williams, an anonymous reviewers and editor John Beavan significantly improved this manuscript.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification, *IEEE, Trans. Auto. Control,* **19,** 716–723.

Altman, N., 1990. Kernel smoothing of data with correlated errors, *J. Am. Stat. Assoc.,* **85,** 749–759.

Amelung, F., Devin, L., Galloway, Bell, J.W., Zebker, H.A. & Randell, J.L., 1999. Sensing the ups and downs of Las Vegas: InSAR reveals structural control of land subsidence and aquifer-system deformation, *Geology,* **27,** 483–486.

Argus, D.F., Heflin, M.B., Peltzer, G., Crampe, F. & Webb, F.H., 2005. Interseismic strain accumulation and anthropogenic motion in metropolitan Los Angeles, *J. geophys. Res.-Solid Earth,* **110,** doi:10.1029/2003JB002934.

Bawden, G.W., Thatcher, W., Stein, R.S. & Hudnut, K.W., 2001. Tectonic contraction across Los Angeles after removal of groundwater pumping effects, *Nature,* **412,** 812–815.

Blewitt, G. & Lavallée, D., 2002. Effect of annual signals on geodetic velocity, *J. geophys. Res.,* **107,** doi:10.1029/2001JB000570.

Blewitt, G., Lavallée, D., Clarke, P. & Nurutdinov, K., 2001. A new global mode of Earth deformation: seasonal cycle detected, *Science,* **294,** 2342–2345.

Boltzmann, L., 1877. Über die Beziehung eines allgemeine mechanischen Satzes zum zweiten Haupsatze der Wärmetheorie, in *Sitzungsberichte der Akademie der Wissenschaften*, Vol. 75, pp. 67–73, Pergamon Press, New York.

Clarke, P.J., Lavallee, D.A., Blewitt, G. & van Dam, T., 2007. Basis functions for the consistent and accurate representation of surface mass loading, *Geophys. J. Int.,* **171,** 1–10, doi:10.1111/j.1365–246X.2007.03493.x.

Craven, P. & Wahba, G., 1979. Smoothing noisy data with spline functions, *Numerische Mathematik,* **31,** 377–403.

Darwin, G.H., 1882. On variations in the vertical due to elasticity of the Earth's surface, *London Edinburgh Dublin Phil. Mag. J. Sci.,* **14,** 409–427.

Davis, J.L., Elosequi, P., Mitrovica, J.X. & Tamisiea, M.E., 2004. Climate-driven deformation of the solid Earth from GRACE and GPS, *Geophys. Res. Lett.,* **31,** doi:10.1029/2004GL021435.

Dong, D., Fang, P., Bock, Y., Cheng, M.K. & Miyazaki, S., 2002. Anatomy of apparent seasonal variations from GPS-based site position time series, *J. geophys. Res.,* **107,** 10.1029/2001JB000573.

Elósegui, P., Davis, J.L., Mitrovica, J.X., Bennett, R.A. & Wernicke, B.P., 2003. Crustal loading near Great Salt Lake, Utah, *Geophys. Res. Lett.,* **20,** doi:10.1029/2003GL016579.

Fisher, R.A., 1922. On the mathematical foundations of theoretical statistics, *Phil. Trans. R. Soc. Lond. A,* **222,** 309–368.

Gourmelen, N., Amelung, F., Casu, F. & Manzo, M., 2007. Mining-related ground deformation in Crescent Valley, Nevada: implications for sparse GPS networks, *Geophys. Res. Lett.,* **34,** doi:10.1029/2007GL029427.

Hart, J.D., 1991. Kernel regression estimation with time series errors, *J. R. Stat. Soc. B,* **53,** 173–187.

King, N.E. *et al.*, 2007. Space geodetic observation of expansion of the San Gabriel Valley, California, aquifer system, during heavy rainfall in winter 2004–2005, *J. geophys. Res.,* **112,** doi:10.1029/2006JB004448.

Kullback, S. & Leibler, R.A., 1951. On information and sufficiency, *Ann. Math. Stat.,* **22,** 79–86.

Lanari, R., Lundgren, P., Manzo, M. & Casu, G., 2004. Satellite radar interferometry time series analysis of surface deformation for Los Angeles, California, *Geophys. Res. Lett.,* **31,** doi:10.1029/2004GL021294.

Langbein, J., 2008. Noise in GPS displacement measurements from southern California and southern Nevada, *J. geophys. Res.,* **113,** doi: 10.1029/2007JB005247.

Lu, Z. & Danskin, W.R., 2001. InSAR analysis of natural recharge to define structure of a ground-water basin, San Bernardino, California, *Geophys. Res. Lett.,* **28,** 2661–2664.

Matthews, M.V. & Segall, P., 1993. Estimation of depth-dependent fault slip from measured surface deformation with application to the 1906 San Francisco earthquake, *J. geophys. Res.,* **98,** 12 153–12 163.

Murray, J.R. & Segall, P., 2005. Spatiotemporal evolution of a transient slip event on the San Andreas fault near Parkfield, California, *J. geophys. Res.-Solid Earth,* **110,** doi:10.1029/2005JB003651.

Naylor, A. & Sell, G., 1982. *Linear Operator Theory in Engineering and Science*, Vol. 40 of Applied Mathematical Sciences, Springer-Verlag.

Nicolas, J. *et al.*, 2006. Seasonal effect on vertical positioning by Satellite Laser Ranging and GLobal Positioning System and on absolute gravity at the OCA geodetic station, Grasse, France, *Geophys. J. Int.,* **167,** 1127–1137.

Penna, N.T. & Stewart, M.P., 2003. Aliased tidal signatures in continuous GPS height time series, *Geophys. Res. Lett.,* **30,** doi:10.1029/2003GL018828.

Penna, N.T., King, M.A. & Stewart, M.P., 2007. GPS height time series: short-period origins of spurious long-period signals, *J. geophys. Res.,* **112,** doi:10.1029/2005JB004047.

Schwarz, G., 1978. Estimating the dimension of a model, *Ann. Stat.,* **6,** 461–464.

Shannon, C.E., 1948. A mathematical theory of communication, *Bell Syst. Tech. J.,* **27,** 379–423.

Stewart, M.P., Penna, N.T. & Lichti, D.D., 2005. Investigating the propagation mechanism of unmodelled systematic errors on coordinate time series estimated using least squares, *J. Geod.,* **79,** 479–489.

van Dam, T. & Wahr, J., 1998. Modeling environmental loading: a review, *Phys. Chem. Earth,* **23,** 1077–1087.

van Dam, T.M., Blewitt, G. & Heflin, M.B., 1994. Atmospheric pressure loading effects of GLobal Positioning System coordinate determinations, *J. geophys. Res.,* **99,** 23 939–23 950.

van Dam, T., Wahr, J., Milly, P.C.D., Shmakin, A.B., Blewitt, G., Lavallee, D. & Larson, K., 2001. Crustal displacements due to continental water loading, *Geophys. Res. Lett.,* **28,** 651–654.

van Dam, T., Wahr, J. & Lavallee, D., 2007. A comparison of annual vertical crustal displacements from GPS and Gravity Recovery and Climate Experiment (GRACE) over Europe, *J. geophys. Res.,* **112,** doi:10.1029/2006JB004335.

Wahba, G., 1990. *Spline Models for Observational Data*, Vol. 59 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics.

Watson, K.M., Bock, Y. & Sandwell, D.T., 2002. Satellite interferometric observations of displacements associated with seasonal groundwater in the Los Angeles basin, *J. geophys. Res.,* **107,** 10.1029/2001JB000470.

Watson, C., Tregoning, P. & Coleman, R., 2006. Impact of solid Earth tide models on GPS coordinate and tropospheric time series, *Geophys. Res. Lett.,* **33,** doi:10.1029/2005GL025538.

Williams, S., Bock, Y., Feng, P., Jamason, P., Nikolaidis, R.M. & Prawirodirdjo, L., 2004. Error analysis of continuous GPS position time series, *J. geophys. Res.,* **109,** doi:10.1029/2003JB002741.