Tectonophysics 665 (2015) 118–126

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Monte Carlo SSA to detect time-variable seasonal oscillations from GPS-derived site position time series

Chang Xu [a], Dongjie Yue [b]

[a] School of Surveying & Municipal Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China
[b] School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China

ARTICLE INFO

Article history:
Received 14 January 2015
Received in revised form 20 September 2015
Accepted 22 September 2015
Available online 8 October 2015

Keywords:
GPS
MCSSA
Seasonal *oscillations*
MLE
Mass loading

ABSTRACT

We explore the capability of singular spectrum analysis (SSA) to extract time-variable seasonal *oscillations* from continual GPS observations and demonstrate the statistical assessment on the colored noise (in particular the first-order autoregressive AR(1) noise) using Monte Carlo SSA (MCSSA) methodology. We provide example applications to ~15-year vertical coordinate time series for 36 globally distributed International GNSS Service (IGS) sites. We find the SSA-filtered seasonal signals can easily pass the confidence interval and hypothesis tests of MCSSA. However, maximum likelihood estimate (MLE) results show that 72% of sites have their flicker noise amplitudes reduced after removing SSA-filtered annual signal, implying that the SSA-filtered seasonal signals may contain an artificial signal driven by colored noise. Therefore, the AR(1) null hypothesis noise model may be misleading in surrogate data tests for GPS seasonal signals. Moreover, comparison between SSA-filtered GPS annual signals and joint geophysical model predictions (non-tidal atmospheric loading + non-tidal ocean loading + hydrological loading) confirms that seasonal signals are resulting from a combination of mass loading and systematic error.

1. Introduction

Site-position time series generated from continuous Global Positioning System (CGPS) arrays reveal significant seasonal variations with annual and semi-annual periods (especially in the vertical component). The study of seasonal signals in GPS coordinate time series has been pursued for a number of important reasons. First, seasonal signals can bias the site velocity if unaccounted for (Blewitt and Lavallée, 2002). Second, seasonal signals in the frame sites can bias the frame realization and further aliased into the site coordinates (Freymueller, 2009). Finally, seasonal signals are vital for understanding the underlying geophysical processes associated with the large-scale transport of terrestrial fluids, such as atmospheric pressure (Tregoning and van Dam, 2005; Tregoning and Watson, 2009; Van Dam et al., 1994) and hydrology loading (Dill and Dobslaw, 2013; Dong et al., 1997, 2002; van Dam et al., 2001). In this case, these seasonal effects must be taken into account for reliable estimated seasonal signals, in both GPS coordinate time series and environmental loading. However, conventional least-squares fitting (LSF) method solves seasonal terms with constant amplitude, while real seasonal signals usually display amplitude modulation due to non-uniform variation of the seasonal excitation sources. Additionally, the noise in GPS coordinate time series and environmental variables are time-correlated. If unaccounted for, the uncertainties of the harmonic terms are estimated to be too optimistic (Langbein, 2012; Mao et al., 1999; Zhang et al., 1997). Therefore, to obtain accurate harmonic estimators and their uncertainties require a prior appropriate background noise assumption. Mao et al. (1999) and Williams et al. (2004) found that the stochastic properties of GPS coordinate time series can be best described as a combination of white plus flicker noise. However, other noise (e.g., power-law noise, random walk noise, and first-order Gauss–Markov noise) is also detected in GPS coordinate time series (King and Williams, 2009; Langbein, 2008). Though the noise content can be estimated using the maximum likelihood estimation (MLE) method, unfortunately, current methods (e.g., CATS software (Williams, 2008) have a significant computational burden. For 10 years of daily GPS solutions of a single station with missing data, a stochastic model such as power-law noise plus white where the spectral index is estimated takes over a day to process.

Extraction of time-variable seasonal signals in GPS coordinate time series is challenging. The problem has been explored in some details in several recent studies. Bennett (2008) developed a flexible semi-parametric model to investigate quasi-periodic signals in CGPS coordinate time series, while the model depends on three main assumptions: smoothly deviation functions, constant phase, and "best" regularization parameter, respectively. Davis et al. (2012) proposed a Kalman filter to capture the stochastic seasonal behavior of geodetic time series, while the random walk process assumed in Kalman filtering should be estimated in advance, which leads to a heavy computational burden. Chen et al. (2013) applied singular spectrum analysis (SSA) to extract time-variable seasonal signals from GPS coordinate time series. However, simply using "pair selection criteria" (Vautard et al., 1992) to identify oscillatory empirical

http://dx.doi.org/10.1016/j.tecto.2015.09.029
0040-1951/

orthogonal functions (EOFs) may be problematic, because the annual signal can be contaminated by the draconitic spectrum with period of 351.2 days (Ray et al., 2008). More recently, Klos et al. (2015) used wavelet decomposition to determine seasonal GPS curves that change in their amplitude. However, wavelet decomposition may absorb colored noise. Therefore, it is necessary to further extend our research to apply Monte Carlo SSA (MCSSA) to discriminate between the seasonal signals and colored noise exist in GPS coordinate time series, and to investigate whether SSA-filtered GPS seasonal signals reflect geophysical signals.

The rest of the article is structured as follows. Section 2 briefly outlines the methodology of SSA, MCSSA, and LSF, respectively. Section 3 demonstrates the ability of MCSSA in detecting seasonal signals from noised GPS coordinate time series, and then follows the comparison between SSA and LSF in extracting seasonal signals. Noise contents estimated from the stochastic model using MLE are used to confirm whether MCSSA really works. Example applications to ~15 years vertical coordinate time series for 36 globally distributed International GNSS Service (IGS) sites are presented. The potential consequence of neglecting seasonal signals on the estimation of site velocity is explored. Displacements predicted from atmospheric pressure loading (ATML), non-tidal ocean mass loading (NTOL), and hydrology loading (HYDL) are used to investigate whether SSA-filtered GPS seasonal signals reflect geophysical significance. Finally, Section 4 presents concluding remarks.

2. Methodology

2.1. SSA and MCSSA

Vautard and Ghil (1989) originally introduce the idea of using pairs of sinusoidal EOFs in quadrature to indicate a physical oscillation. Since then, SSA has received considerable attention in climate data analysis. For a standardized time series X, where sample index t varies from 1 to N, we obtain an $N \times M$ trajectory matrix D by sliding an M-point window, and then define an $M \times M$ lagged correlation matrix:

$$C_X = \eta D^T D \tag{1}$$

where the superscript T indicates transposition of vectors and matrices, η is a normalization constant.

Due to less variance and convenient associating frequencies with EOFs, we adopt the VG algorithm (Ghil et al., 2002) to compute the eigenelements of C_X:

$$C_{ij} = \frac{1}{N-|i-j|} \sum_{t=1}^{N-|i-j|} d_t d_{t+|i-j|} \tag{2}$$

We diagonalize the lag-covariance matrix C_X and rank the eigenvalues in decreasing order:

$$\Lambda_X = (E_X)^T C_X E_X \tag{3}$$

where Λ_X is diagonal, the kth diagonal element being the kth largest eigenvalue and the kth column of E_X being the corresponding eigenvector or EOF.

Projecting the time series onto each EOF yields the corresponding principal components (PCs) A_k:

$$A_k(t) = \sum_{j=1}^{M} X(t+j-1) E_k(j) \tag{4}$$

The entire time series or parts of it that correspond to trends, periodic terms, or noise can be reconstructed by linear combining the reconstructed components (RCs) R_k:

$$R_K = \begin{cases} \dfrac{1}{i} \displaystyle\sum_{j=1}^{i} A_k(t+j-1) E_k(j) & 1 \le i \le M-1 \\ \dfrac{1}{M} \displaystyle\sum_{j=1}^{M} A_k(t+j-1) E_k(j) & M \le i \le N-M+1 \\ \dfrac{1}{N-i+1} \displaystyle\sum_{j=i-N+M}^{M} A_k(t+j-1) E_k(j) & N-M+2 \le i \le N \end{cases}, \tag{5}$$

where k is the set of EOFs on which the reconstruction is based on.

The following question is how to select k when a data series is contaminated with colored noise, and this issue is not well solved in Chen et al. (2013). For the purposes of developing the problem, we assume the signal and noise are linearly independent, and define the expected lag-covariance matrix of the data series as:

$$\xi(C_X) = \xi(C_S) + \xi(C_R) \tag{6}$$

We find that signal-to-noise (S/N) separation using slope break in a "scree diagram" of eigenvalues versus k does not work well when a data series turns out not to be white, because the high-ranked EOFs of C_X can no longer be expected to approximate to the EOFs of the lag-covariance matrix of signal C_S. Given this, conventional eigenvalue rank order is no longer a reliable indicator of statistical or physical significance. Alternatively, "Pair selection criteria" (Vautard et al., 1992) may be a solution for indicating potential oscillatory EOFs. However, Allen (1992) notes sinusoidal EOF pairs may also be attributable to noise, and a linear plus a first-order autoregressive (AR(1)) noise can easily pass the "pair selection criteria." These problems that arise with red noise motivate Allen and Smith (1996) to develop MCSSA, a method of discriminating signals from arbitrary noise process via SSA, based on the idea of surrogate data testing.

The first step in MCSSA is testing series against a pure AR(1) noise null hypothesis. Before simulating red noise data (i.e., surrogate data), noise parameters (e.g., square on the mean u^2, lag-1 autocorrelation γ, and noise covariance α, respectively) must be unbiasedly estimated in advance. Following Allen and Smith (1996), we define the expected lag-covariance matrix of this noise realization by

$$C_N = c_0 W \tag{7}$$

where $W_{ij} = \gamma^{|i-j|} - \mu^2(\gamma)$ and $\mu^2(\gamma) = -\frac{1}{N} + \frac{2}{N^2}\left[\frac{N-\gamma^N}{1-\gamma} - \frac{\gamma(1-\gamma^{N-1})}{(1-\gamma)^2}\right]$.

We estimate the explicitly γ, namely $\tilde{\gamma}$, by solving

$$\frac{tr_1(W)}{tr_0(W)} = \frac{tr_1(C_X)}{tr_0(C_X)} \tag{8}$$

using Newton–Raphson iteration with a start from $\hat{\gamma} = tr_1(C_X)/tr_0(C_X)$, where tr_j is a generalized trace operator, which applied to a $M \times M$ symmetric matrix, is defined as:

$$tr_j(C) = \frac{1}{M-j} \sum_{k=1}^{M-j} C_{k,k+j} \tag{9}$$

Then the unbiased lag-1 autocorrelation $\tilde{c}_0$ is given by $\tilde{c}_0 = tr_0(C_X)/tr_0(W)$, and unbiased $\tilde{\alpha}$ is obtained from $\tilde{\alpha} = \sqrt{\tilde{c}_0(1-\gamma)}$ simultaneously.