AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE

10.1002/2015JB012424

Data-adaptive detection of transient deformation in geodetic networks

Damian Walwer[1], Eric Calais[1], and Michael Ghil[1,2]

[1]Ecole normale supérieure, Department of Geosciences, PSL Research University, Paris, France, [2]ASOS Department and IGPP, University of California, Los Angeles, California, USA

Key Points:
- Data-adaptive analysis of GPS time series to detect transient deformation and seasonal oscillations
- Use of the M-SSA to analyze GPS time series without using any a priori hypotheses
- Detection of two cycles of uplift and subsidence at Akutan volcano likely to be of magmatic origin

Supporting Information:
- Supporting Information S1

Correspondence to:
D. Walwer,
walwer@geologie.ens.fr

Citation:
Walwer, D., E. Calais, and M. Ghil (2016), Data-adaptive detection of transient deformation in geodetic networks, *J. Geophys. Res. Solid Earth*, *121*, doi:10.1002/2015JB012424.

Received 5 AUG 2015
Accepted 22 FEB 2016
Accepted article online 24 FEB 2016

Abstract The recent development of dense and continuously operating Global Navigation Satellite System (GNSS) networks worldwide has led to a significant increase in geodetic data sets that sometimes capture transient-deformation signals. It is challenging, however, to extract such transients of geophysical origin from the background noise inherent to GNSS time series and, even more so, to separate them from other signals, such as seasonal redistributions of geophysical fluid mass loads. In addition, because of the very large number of continuously recording GNSS stations now available, it has become impossible to systematically inspect each time series and visually compare them at all neighboring sites. Here we show that Multichannel Singular Spectrum Analysis (M-SSA), a method derived from the analysis of dynamical systems, can be used to extract transient deformations, seasonal oscillations, and background noise present in GNSS time series. M-SSA is a multivariate, nonparametric, statistical method that simultaneously exploits the spatial and temporal correlations of geophysical fields. The method allows for the extraction of common modes of variability, such as trends with nonconstant slopes and oscillations shared across time series, without a priori hypotheses about their spatiotemporal structure or their noise characteristics. We illustrate this method using synthetic examples and show applications to actual GPS data from Alaska to detect seasonal signals and microdeformation at the Akutan active volcano. The geophysically coherent spatiotemporal patterns of uplift and subsidence thus detected are compared to the results of an idealized model of such processes in the presence of a magma chamber source.

1. Introduction and Motivation

The past decade has seen a rapid increase in the number and spatial density of continuously operating Global Navigation Satellite System (GNSS) stations to monitor crustal deformation in tectonically active regions or simply serve as reference stations for surveying applications. The leading technique as of today is the Global Positioning System (GPS), with two major deformation monitoring networks currently in operation, the Plate Boundary Observatory (PBO) in the Western U.S. with roughly 2000 stations and the GEONET (http://www.gsi.go.jp/) network in Japan with ~5000 stations.

These networks record linear site displacements due to plate motions and interseismic strain accumulation on active faults, as well as transient-deformation events of various origins—volcanic, tectonic, or hydrologic [e.g., *Feng and Newman*, 2009; *King et al.*, 2007; *Miller et al.*, 2002]—whose spatiotemporal scales span several orders of magnitude. Our ability to objectively and reliably detect these transients is of primary importance because they contain information on the rheological properties and stress state of near-surface faults, as well as of the Earth's crust and mantle. Detecting transients, however, is challenging for two reasons. First, the very large amount of data now available in large geodetic networks make the visual inspection of time series at each site very time consuming and well nigh impossible. Second, the amplitude of these transients can be close to—or lie even below—the background noise inherent to geodetic time series.

Several methods have recently been proposed to detect transient deformation in GPS time series. *Ohtani et al.* [2010] model GPS time series as the sum of a spatially coherent signal described as a sum of spatial wavelets weighted by temporally varying coefficients plus errors from different sources, such as reference frame, site-specific noise, and observation error. *Ji and Herring* [2013] first smooth the time series using a Kalman filter to increase signal-to-noise ratio, while explicitly estimating annual and semiannual harmonic terms to model the seasonal oscillations commonly observed in GPS time series. They analyze the data in

the temporal dimension first by estimating time-correlated signals, then, in a separate step, use a principal component analysis (PCA) to exploit the spatial correlations between the smoothed and filtered time series.

The approach of *Riel et al.* [2014] is to parameterize position time series using a dictionary of temporally varying nonorthogonal functions assumed to represent all possible signal types present in the data. The dictionary contains functions to model known signals, such as sinusoidal functions for seasonal oscillations or logarithmic functions for postseismic deformation. It also contains B-spline functions with various periods and starting times to represent transient displacements. The method accounts for spatial correlations using common B-spline functions to describe transient mode of variability shared across several time series.

These methods all aim at improving the signal-to-noise ratio while using the spatial or temporal correlation inherent to GPS position time series to search for transient events. All three use a priori hypotheses on the spatial or temporal structure of the transients to build basis functions—e.g., wavelets or B-splines—onto which the data can then be projected to detect and extract those transients. In *Ji and Herring* [2013] the a priori hypothesis is the stochastic process that gives rise to the noise and the transient signal.

All these methods also parameterize the seasonal signal as a sum of harmonic functions with annual and semi-annual periods; this assumption prevents one, however, from accounting for interannual variability. Moreover, the stochastic properties of GPS position time series and the temporal and spatial shapes of transient events are not really known at this time. Making educated guesses on the stochastic properties of GPS time series or on the basis functions that best describe their temporal correlations is a valid approach, but one would like in fact to extract this information from the GNSS data sets in a data-adaptive way, i.e., purely based on information contained in the data, without a priori assumptions.

The purpose of this paper is to show that some of these shortcomings can be overcome using Multichannel Singular Spectrum Analysis (M-SSA). M-SSA is an advanced time series analysis method that simultaneously takes advantage of the spatial and temporal correlations in geophysical fields to extract empirical basis functions that represent the common modes of spatiotemporal variability of the data set. The main benefit of this method is to allow unraveling oscillations and trends with nonconstant slopes embedded in time series without using any a priori knowledge about their period and amplitude or their spatiotemporal structures. M-SSA was first proposed, in the version used here, in the field of climate studies by *Keppenne and Ghil* [1993] and *Plaut and Vautard* [1994] to extract low-frequency oscillations of the atmospheric system from geopotential height field data. It was also used to study surface winds [*Jiang et al.*, 1995] and sea level data [*Unal and Ghil*, 1995]. *Ghil et al.* [2002] provide a review of the methodology and of many applications; see also http://web.atmos.ucla.edu/tcd/ssa/ for further details and references.

We show in the following that M-SSA is well suited to extract geophysically relevant information from GPS position time series. In the next section, we briefly review the methodology and illustrate its application to a set of synthetic time series whose properties resemble those of GPS records. This methodology is then applied in section 3 to a set of recorded displacements at eight GPS sites on the Akutan volcano in the Aleutian Islands. In section 4, we show that M-SSA can be successfully applied to larger data sets, by using 80 stations from the PBO network covering Alaska, as well as to the extraction of noise characteristics from the Akutan signals.

2. Multichannel Singular Spectrum Analysis (M-SSA)

2.1. Formulation

Chen et al. [2013] have already proposed using single-channel SSA for the extraction of the seasonal signal from a single GPS time series and reviewed the univariate methodology. The emphasis here is on the often more powerful multivariate case.

Let

$$\{X_l(t) : l = 1, \ldots, L; t = 1, \ldots, N\} \tag{1}$$

be an ensemble of GPS time series: L is the number of channels—each channel here being a time series—and N the number of data points in each channel. The sampling time Δt is constant; hence, $(N-1)\Delta t$ is the duration of the time series.

The key idea of M-SSA is to exploit the covariance information contained in a series of lagged copies of all $X_l(t)$ over a sliding M-point window [*Broomhead and King*, 1986a, 1986b; *Ghil et al.*, 2002]. In M-SSA one starts therefore by considering the matrix $\tilde{\mathbf{X}}_l$ that includes M time-delayed copies of the original time series $X_l(t)$, written as:

$$\tilde{\mathbf{X}}_l = \begin{pmatrix} X_l(1) & X_l(2) & \cdots & X_l(M) \\ X_l(2) & X_l(3) & \cdots & X_l(M+1) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ X_l(N') & X_l(N'+1) & \cdots & X_l(N) \end{pmatrix} \tag{2}$$

with $N' = N - M + 1$. Here M is the length of the window used to embed the original time series and it must be chosen to optimize the quantity of information extracted while maintaining satisfactory statistical confidence in that information.

The covariance matrix $\mathbf{T}_{l,l'}$ between two time series $X_l(t)$ and $X_{l'}(t)$ is given by

$$(T_{l,l'})_{j,j'} = \frac{1}{\tilde{N}} \sum_{n=\max(1,1+j-j')}^{\min(N,N+j-j')} X_l(t) X_{l'}\left(t + j'_j\right), \tag{3}$$

where

$$\tilde{N} = \min(N, N + j - j') - \max(1, 1 + j - j') + 1. \tag{4}$$

These matrices are the blocks of a grand covariance matrix $\tilde{\mathbf{T}}$, given by

$$\tilde{\mathbf{T}} = \begin{pmatrix} \mathbf{T}_{1,1} & \mathbf{T}_{1,2} & \cdot & \cdot & \cdot & \mathbf{T}_{1,L} \\ \mathbf{T}_{2,1} & \mathbf{T}_{2,2} & \cdot & \cdot & \cdot & \mathbf{T}_{2,L} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \mathbf{T}_{l,l'} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{T}_{L,1} & \mathbf{T}_{L,2} & \cdot & \cdot & \cdot & \mathbf{T}_{L,L} \end{pmatrix}. \tag{5}$$

The resolution of the eigenvalue problem

$$\tilde{\mathbf{T}}\mathbf{E}^k = \lambda_k \mathbf{E}^k \tag{6}$$

yields the $L \times M$ eigenvalues λ_k and eigenvectors $\mathbf{E}^k$ of $\tilde{\mathbf{T}}$. Each $\mathbf{E}^k$ can be seen as a succession of L segments $\mathbf{E}^k_l$ of length M. We can thus associate a segment $\mathbf{E}^k_l$ to each time series $X_l(t)$. The eigenvectors $\mathbf{E}^k$ are called the spatiotemporal empirical orthogonal functions (ST-EOFs). Each eigenvalue carries a given amount of variance from the overall data set.

2.2. Synthetic Example

To illustrate the method, we consider first, for the sake of simplicity, the 10 year long univariate time series plotted in Figure 1, with $N = 3652$. This synthetic time series is the sum of an arctangent function that simulates a transient displacement, harmonic functions with annual and semiannual periods that simulate seasonal oscillations, along with a white noise and a colored noise. The latter is associated with the presence of correlations in the time series and it is chosen to be a stochastic process characterized in the spectral domain by the power law [e.g., *Agnew*, 1992]

$$S(f) = S_0 \left(\frac{f}{f_0}\right)^{-\alpha}. \tag{7}$$

Here $S(f)$ is the spectral density, f is the frequency, P_0 and f_0 are constants and α is the spectral index: the larger α, the faster the roll-off, i.e., the faster the decay of the spectral density with increasing frequency f. For this first synthetic example we choose $\alpha = 0.5$. The variance associated with the colored noise account for 50% of the total noise variance.

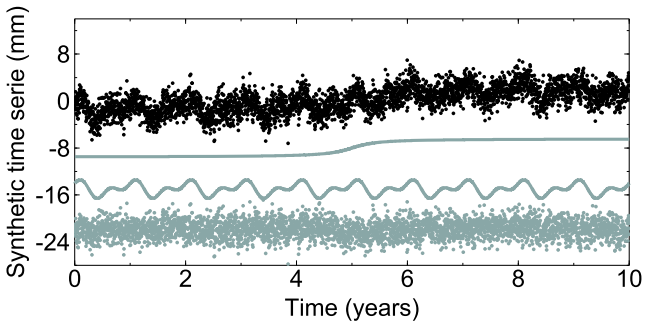

Figure 1. Univariate synthetic time series (black dots) composed of, from top to bottom in gray, a transient displacement, annual and semiannual harmonic functions, as well as colored and white noise.

We apply M-SSA with $M = 400$ days in order to capture the annual signal contained in the synthetic time series. The resulting singular spectrum in Figure 2 displays the eigenvalues calculated by solving equation (6), sorted in decreasing order and normalized so that they represent their corresponding portion of the total variance of the time series. Note that the singular values μ_k are actually the square roots of the eigenvalues λ_k, i.e., $\mu_k^2 = \lambda_k$. We prefer, though, plotting the λ_k's, since they sum to unity, but will still refer to the plots as singular spectra, which is the more customary terminology in the M-SSA context.

One observes that most of the variance in the time series is contained in the first six eigenvalues, with a drop and a plateau after the sixth eigenvalue (Figure 2). These six leading eigenvalues represent the signal contained in the time series, while the lower order ones correspond to noise. One also notices that eigenvalues 2–3 and 4–5 are forming pairs, which indicate the presence of oscillatory modes [*Vautard and Ghil*, 1989; *Ghil et al.*, 2002]. These results illustrate the capacity of M-SSA to extract periodic components, such as the seasonal signals often present in GPS time series.

In studying a single time series one encounters only temporal EOFs (T-EOFs), while using several time series would have given rise to ST-EOFs, cf. equations (5) and (6). The five leading T-EOFs are shown in Figure 3. Note that T-EOFs 2–3 and 4–5, associated with the two pairs of eigenvalues mentioned above, are in phase quadrature. Such pairs of T-EOFs can be seen as data-adaptive counterparts of the sine and cosine functions in the usual Fourier analysis of time series. The advantage of these pairs over the Fourier sine and cosine functions is to allow the representation of periodic signals with time-varying amplitudes with a few EOFs only [*Vautard and Ghil*, 1989; *Ghil et al.*, 2002].

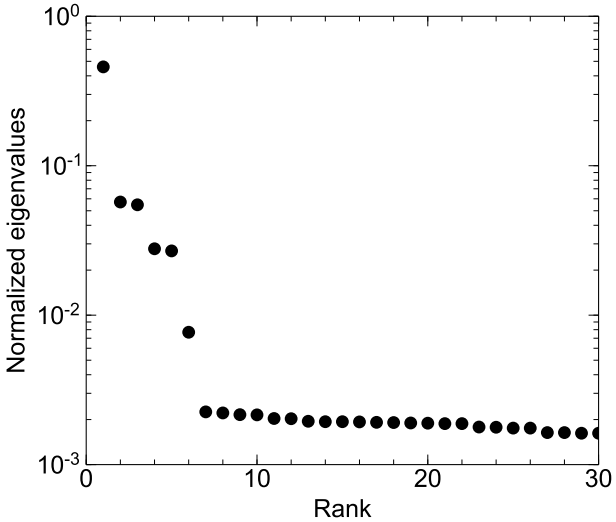

Figure 2. Singular spectrum of the synthetic time series shown in Figure 1. Eigenvalues are normalized to represent the appropriate fraction of the total variance of the time series, and they sum to 1.

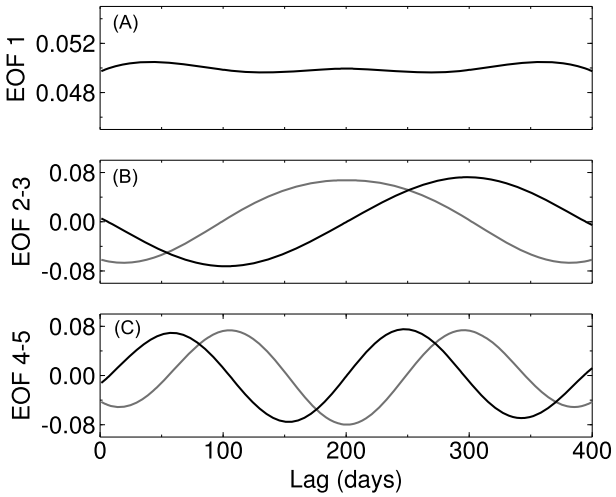

Figure 3. The five leading temporal empirical orthogonal functions (T-EOFs) extracted from the synthetic time series; they are associated with the first five eigenvalues in Figure 2.

The projection of the rows of matrix $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \ldots, \tilde{\mathbf{X}}_l)$ onto the eigenvectors $\mathbf{E}_k$ gives the coefficients $A^k(t)$:

$$A^k(t) = \sum_{j=1}^{M} \sum_{l=1}^{L} X_l(t+j-1)E_l^k(j). \tag{8}$$

The $L \times M$ vectors $A^k(t)$ are called the spatiotemporal principal components (ST-PCs) in M-SSA or temporal PCs (T-PCs) in the case of univariate data. They are time series of weighting coefficients associated with their corresponding T-EOFs.

Figure 4 shows the five leading T-PCs recovered from the synthetic time series. As for the T-EOFs, T-PCs 2–3 and 4–5 are in phase quadrature, with annual and semiannual periods, respectively. They capture the seasonal oscillations present in the synthetic time series, while the first T-PC captures the transient displacement.

2.3. Signal Reconstruction

The ST-PCs and the ST-EOFs allow us to partially reconstruct the time series $X_l(t)$ [*Ghil and Vautard*, 1991; *Vautard et al.*, 1992; *Plaut and Vautard*, 1994]. The part of the signal $R_l^k(t)$ associated with the kth ST-EOF and

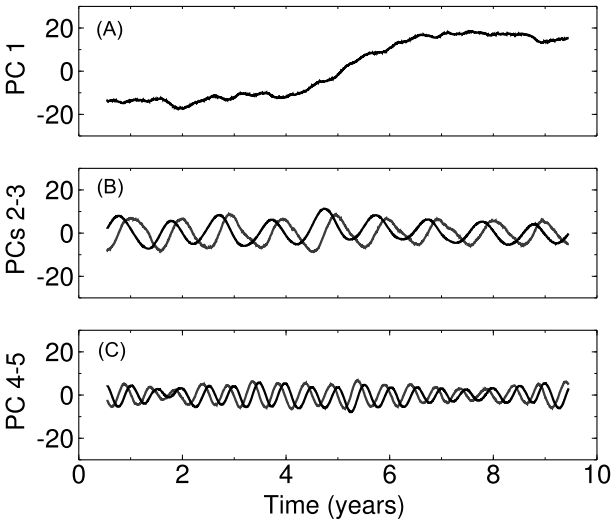

Figure 4. The five leading temporal principal components (T-PCs) associated with the five T-EOFs of Figure 3. These T-PCs are time coefficients that weigh the corresponding T-EOFs in reconstructing the original time series; they represent the different modes of variability of the time series.

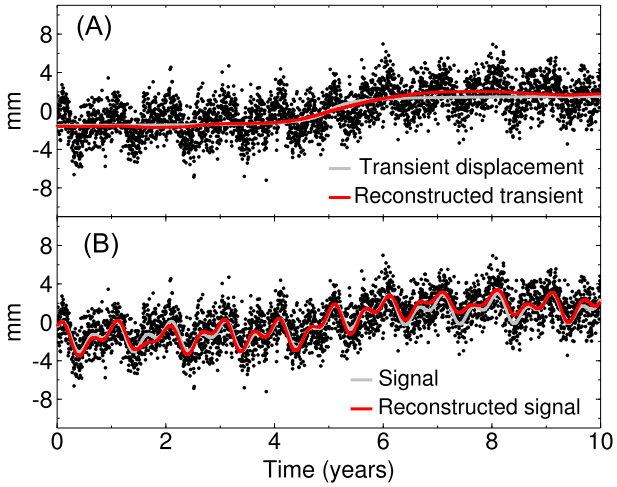

Figure 5. Reconstruction of the (a) transient displacement and the (b) seasonal oscillations in our synthetic example. Black dots = synthetic time series, gray lines = synthetic transient and oscillatory signals embedded in the synthetic time series, red lines = reconstructed signals using M-SSA. (Figure 5a) The reconstruction using the first T-PC and corresponding T-EOF captures the transient displacement signal. (Figure 5b) The reconstruction using the first five T-PCs and corresponding T-EOFs additionally captures the seasonal signal.

ST-PC is given by

$$\begin{cases} R_l^k(t) = \frac{1}{M}\sum_{j=1}^{M} A^k(t-j+1)E_l^k(j), & M \le t \le N-M+1 \\ R_l^k(t) = \frac{1}{i}\sum_{j=1}^{i} A^k(t-j+1)E_l^k(j), & 1 \le t \le M-1 \\ R_l^k(t) = \frac{1}{N-i+1}\sum_{j=1-N+M}^{M} A^k(t-j+1)E_l^k(j), & N-M+2 \le t \le N. \end{cases} \quad (9)$$

These partial reconstructions are filtered versions of the original time series and capture a given portion of their total variance. No information is lost in the process since one can reconstruct the full time series by summing the $L \times M$ $R_l^k(t)$.

Figure 5 shows the reconstructions of the transient displacement and the seasonal oscillations of our example. They faithfully reproduce the transient and oscillatory signals embedded in the synthetic data.

In summary, M-SSA allows one to decompose time series into statistically uncorrelated components that can be classified into trends, oscillatory patterns, and noise. As shown above with the synthetic example of a transient event, trends need not be linear, as emphasized already by *Ghil and Vautard* [1991]. Moreover, oscillations can be amplitude and phase modulated, as will be shown below.

2.4. M-SSA Versus SSA and PCA

At this stage, we have illustrated the method using a univariate synthetic time series, which only gives rise to temporal EOFs and PCs. This particular case of the more generic M-SSA corresponds to $L = 1$ in equation (1) and is simply called SSA, cf. *Broomhead and King* [1986a] and *Vautard and Ghil* [1989]. One can also note that the generic case reduces to the classical principal component analysis (PCA) [*Preisendorfer et al.*, 1981], commonly used in statistical analysis, when $M = 1$ in equation (2).

M-SSA extracts, however, relevant modes of variability in time series more efficiently when using jointly several time series with correlated information. In this case, M-SSA simultaneously exploits the spatial and the temporal correlations in the time series, which allows for a more efficient separation of noise from trends and oscillatory modes.

We illustrate this point by simulating 15 time series in a manner similar to the one described above, all sharing a common transient signal. We apply M-SSA to ensembles of 1 to 15 of those time series and compute the misfit between the synthetic and reconstructed transient signal. As shown in Figure 6a, the misfit systematically decreases as the number of time series that are jointly analyzed increases.

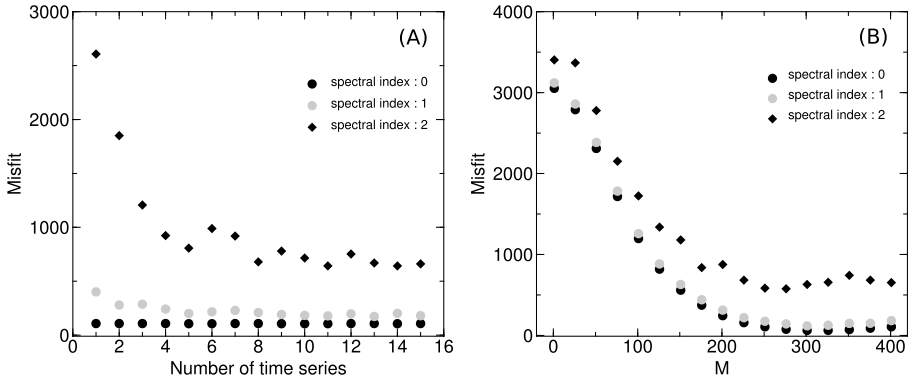

Figure 6. Evaluation of the improvement in M-SSA reconstruction as a function of (a) the number L of time series and (b) of window width M. (Figure 6a) Misfit between the synthetic transient signal and its reconstruction versus the number L of synthetic time series used in M-SSA. (Figure 6b) Misfit between the synthetic transient signal and its reconstruction versus M, the maximum lag used to compute the covariance matrix. All the synthetic time series are similar to the one in Figure 1 and share the same transient displacement. Black dots correspond to the results of an experiment that uses a white noise ($\alpha = 0$), while gray dots correspond to colored noise with a spectral index of $\alpha = 1$ and black diamonds correspond to a noise with a spectral index of $\alpha = 2$. The separating power of M-SSA improves as the number of the series sharing a transient signal increases as well as M increases.

We also illustrate the benefit of using both spatial and temporal correlations jointly over using spatial correlations only, as in the case of PCA, by varying M. Figure 6b shows that the misfit is maximum for $M = 1$ (PCA case) and systematically decreases as M increases. These two simple synthetic tests show that simultaneously exploiting spatial and temporal correlations in time series, instead of using only one of them, significantly improves our ability to extract the transient signals embedded in the data.

We varied the spectral index α of the noise introduced in the synthetic time series by using the values of $\alpha = 0, 1$, and 2. These values correspond to white noise, flicker noise, and random walk noise (sometimes called brownian or red noise), respectively, all of which are commonly found in GPS time series [e.g., *Williams*, 2003]. We observe in Figure 6 that the transient is captured more easily when α is lower and that separating a transient requires more information, i.e., a larger number L of time series, when α is higher. This type of dependence on α is expected because a higher spectral index means that—for the same total spectral power, given by $\int_0^\infty f\, S(f)\mathrm{d}f$—the spectral density, cf. equation (7), will be higher at low frequencies f. This reflect the occurrence of more time-correlated patterns with long correlation times, i.e., more patterns that can be similar to the transient signal.

We also conducted a more realistic multivariate synthetic test that is reported in the supporting information. This test was formulated by the Southern California Center community and was also performed by *Riel et al.* [2014]. It consists in a set of synthetic data simulating time series recorded at cGPS sites in California. These time series contains simulated colored noise, seasonal signals, and transient signals related to a slip event on the Santa Monica fault.

3. Application to Transient Deformation of the Akutan Volcano, Alaska

We applied M-SSA to a set of GPS data recorded at the Akutan volcano, Alaska, in order to test the method in the case of GPS position time series. Akutan is an active volcano of the Aleutian arc equipped with eight continuously operating GPS receivers since the late 2006; see maps in Figure 7. The data are taken from the Plate Boundary Observatory, http://pbo.unavco.org/.

We chose this case because *Ji and Herring* [2011] were able to identify a transient inflation episode in early 2008 given only a few millimeters of horizontal and vertical displacements, i.e., close to the noise level in the GPS time series. To do so, they first smoothed each time series independently by modeling it as the sum of a constant rate, annual and semiannual harmonic functions, and a first-order Gauss-Markov (FOGM) process that accounts for temporally correlated noise and transient signal. In a second step, they used PCA to extract spatially correlated information from the FOGM state estimates. They also removed some of the noise common to all Alaska continuous GPS (cGPS) stations—hence not specific to Akutan's deformation—by estimating a seven-parameter reference frame transformation to the original time series in a first, preanalysis, stage.

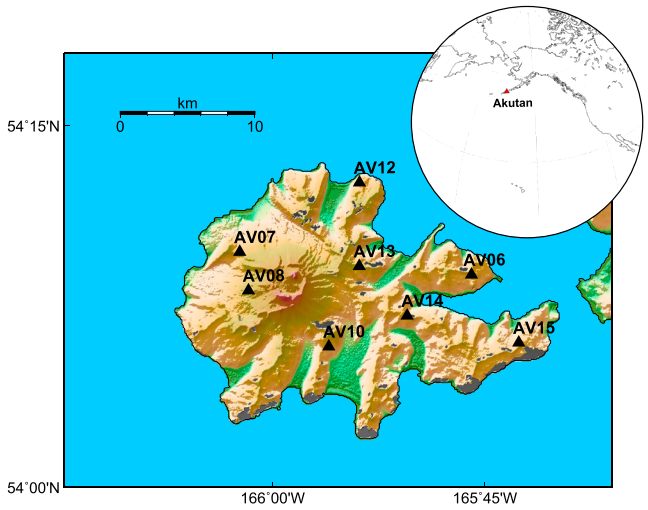

Figure 7. Maps showing the location of the Akutan volcano within the Aleutian arc (inset) and of the eight continuously recording GPS receivers (black triangles) used in this study on the island (main map).

With this composite method, *Ji and Herring* [2011] were able to successfully identify an inflation event with 1–10 mm of radial horizontal displacement and 0–11 mm of uplift centered on the active crater; this event started in February 2008 and ended 4 months later. Their approach, however, makes a number of a priori assumptions, in particular, on the shape of the seasonal oscillations with constant amplitude and on the a priori probability density function of the noise and the transient signal. The fact that they are able to identify a transient likely to be of magmatic origin justifies their hypotheses in this specific case.

We would like to test whether one can also identify this transient inflation event by jointly exploiting the spatial and temporal correlations contained in the data, using the M-SSA on the raw, unfiltered time series, with no preprocessing whatsoever. Moreover, we will show that M-SSA can detect additional inflation events in more recent data, as well as being applicable to much larger data sets.

3.1. The GPS Data

We use position time series from cGPS stations of the PBO network, sampled on a daily basis and expressed in the Stable North America Reference Frame [*Herring et al.*, 2008]. Since the cGPS measure displacement in the northward, eastward, and upward directions, and since there are eight such stations on the island (see Figure 7), the number of channels in our M-SSA analysis is $L = 8 \times 3 = 24$.

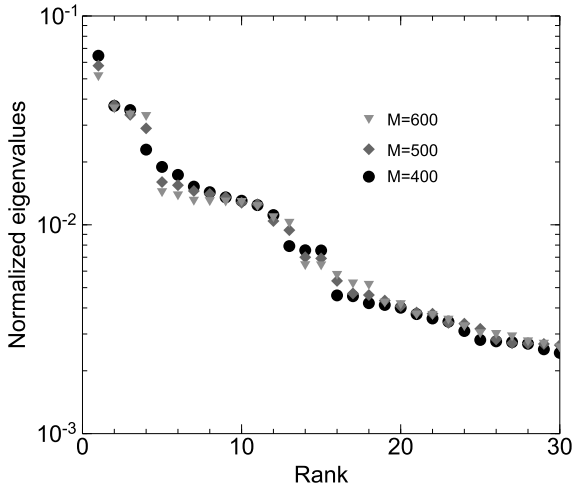

Figure 8. Singular spectrum of the GPS position time series at 8 cGPS sites on Akutan volcano from early 2006 to late 2014. The first 30 M-SSA eigenvalues are plotted for the window widths $M = 400$, 500, and 600 days. Eigenvalues are normalized as in Figure 2, so that their sum is unity.

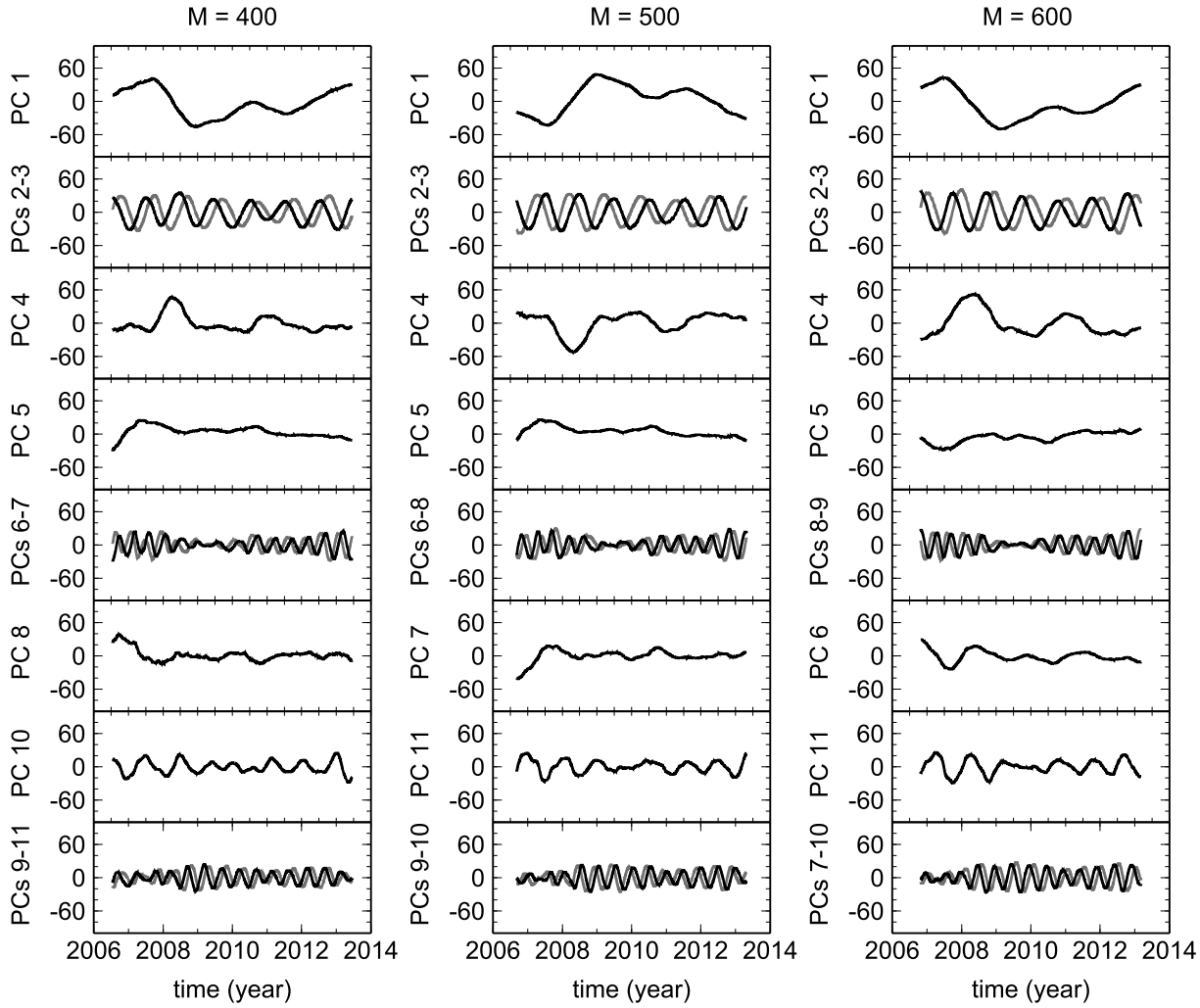

Figure 9. The 11 leading ST-PCs associated either with the trend with nonconstant slope or the seasonal oscillations. They are calculated using (left column) $M = 400$, (middle column) $M = 500$, and (right column) $M = 600$ days. The persistence of the shape of the ST-PCs for the different values of M—ignoring the occasional change of sign, which is a numerical artifact of PCA in general—is an indicator of the significance of the extracted modes. Regardless of the value of M, ST-PCs 1, 4, 5, and 8 jointly capture the data-adaptive trend, while the ST-PC pair 2–3 represents the annual oscillations. The four ST-PCs 6–7 and 9–11 capture the semiannual oscillations for $M = 400$, whereas this mode is captured by ST-PCs 6–7 and 9–10 for $M = 500$ and ST-PCs 8–10 and ST-PCs 6–9 for $M = 600$.

M-SSA requires that all time series analyzed have the same length. Hence, we filled missing data in any of the channels over an interval of length at epoch t_m by using the following model:

$$x(t_m) = at_m + b + \sigma r, \tag{10}$$

where $x(t_m)$ is an estimation of the missing data, a and b are the slope and intercept of the regression line, respectively, r is a Gaussian random variable with zero mean and unit standard deviation, while σ is the standard deviation of the available data over an interval of the same length.

Filling gaps in this way is justified by the fact that Gaussian noise is not correlated in time or space and is thus going to be seen by M-SSA as part of the noise, rather than of the trends or oscillatory signals [*Unal and Ghil*, 1995].

We also removed the linear trend in each time series before applying the M-SSA to the data set. The linear trend, associated with stationary deformation, captures in general the largest proportion of variance in the data set. Because of this, it is useful to remove it to focus on transient deformation that account often for much less variance.

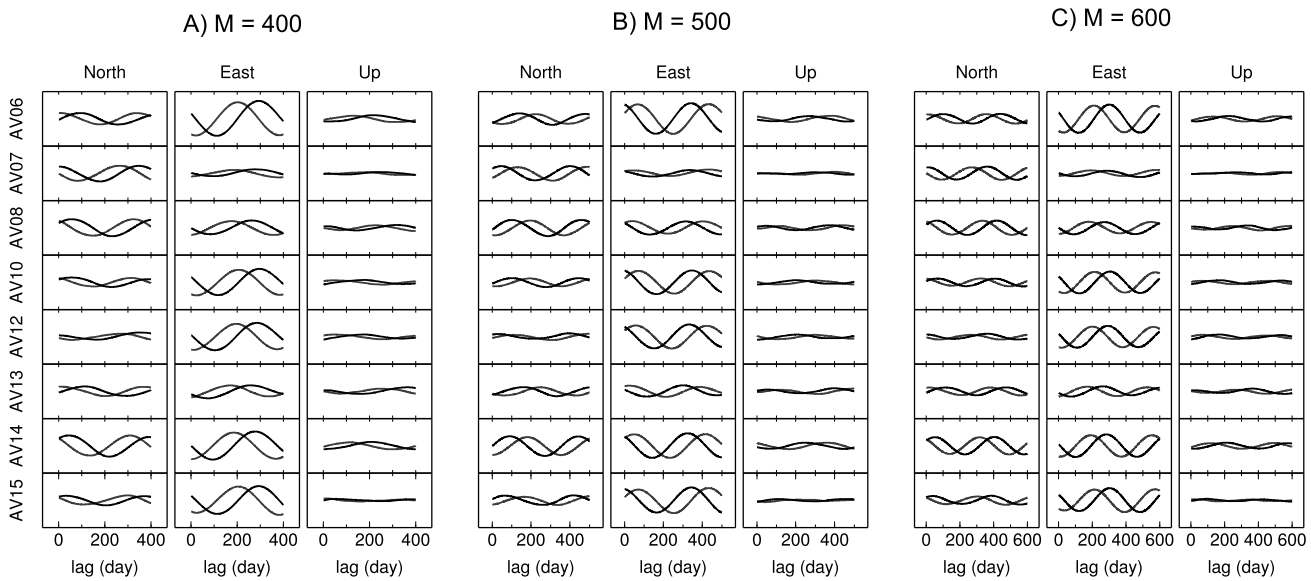

Figure 10. Spatiotemporal empirical orthogonal functions (ST-EOFs) associated with eigenvalues 2 (black) and 3 (gray) (see Figure 8) calculated using (a) $M = 400$, (b) $M = 500$, and (c) $M = 600$ days. The subpanels show the components of these EOFs for each cGPS site and each direction of motion being measured. These two ST-EOFs capture the annual oscillatory mode of annual period present in the data. The differences between estimates using different values of M are small.

3.2. Choice of Window Width *M*

Before we describe the modes of variability, based on the ST-EOFs and ST-PCs, we discuss the choice of the window width M; it equals the number of lagged time series used to construct the grand covariance matrix $\tilde{\mathbf{T}}$ and it is a key parameter of M-SSA. M determines, in particular, the length of the ST-EOFs and, therefore, it limits the maximum period of the oscillations that can be extracted from the data, which equals approximately $M\Delta t$.

The optimal value of M is a trade-off between the statistical significance of the extracted modes and the maximum period of the oscillations one aims to extract. The larger the N/M ratio—i.e., the smaller the value of M—the larger the statistical significance of the extracted modes [*Ghil et al.*, 2002]. This is so because the uncertainty associated with the estimation of the covariances decreases as the number of points used in the estimation, i.e., $N - M + 1$—increases. On the other hand, extracting the annual signal present in the GPS time series obviously requires that M be greater than 365 days.

We applied M-SSA to the eight GPS time series recorded at Akutan volcano from late 2006 to early 2014, for a total of 3000 data points in the three displacements—N-S, E-W, and up-down—and obtained the singular spectrum shown in Figure 8. The singular spectrum in the figure was obtained for three different values of M: 400, 500, and 600 days. The three spectra show a similar shape, with a first eigenvalue well above the others and accounting for a large portion of the data variance. It is followed by a first plateau formed by two eigenvalues for $M = 400$ or three eigenvalues for $M = 500$ and $M = 600$. *Vautard and Ghil* [1989] already noticed that, as M increases, the same feature in the spectrum will require a larger number of modes to be described fully.

All three spectra show a second plateau, sloping down to a third one that ends at eigenvalue 15. The ST-EOFs and ST-PCs for these value of M are plotted in Figures 10 and 9, respectively. They also show similar shapes, for instance, the letter M-like trend visible in the first ST-PC for all values of the window width M, cf. Figure 9.

The similarity of the spectra, together with the robustness of the shape of the ST-EOFs and ST-PCs, suggests that they capture the same dynamic processes. This parametric robustness—in addition to standard statistical significance tests—gives us further confidence in the extracted modes being physically meaningful [*Ghil et al.*, 2002, section 5.2]. In the following, we will discuss results obtained with $M = 400$.

3.3. Singular Spectrum, ST-EOFs, and ST-PCs

The ST-PCs associated with the first 11 eigenvalues are plotted in Figure 9. We observe that the first ST-PC describes 6% of the total variance of the data set and corresponds to an M-shaped trend and a relatively long periodic temporal variability. ST-EOFs 2 and 3 carry 7% of the variance and, as shown in Figure 10, they are in

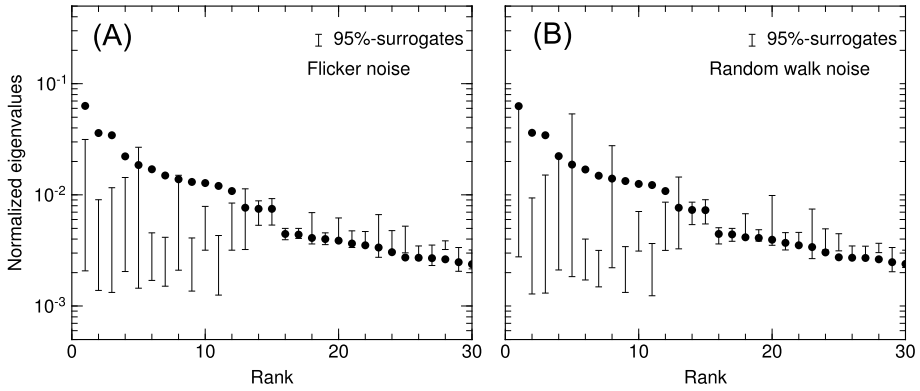

Figure 11. MC-MSSA test of statistical significance using surrogate data containing (a) flicker noise, with spectral index $\alpha = 1$; and (b) random walk noise, with $\alpha = 2$. Lower and upper ticks on the error bars correspond to the 2.5% and 97.5% quantiles of an ensemble of 500 surrogate data that are different realizations of the null hypothesis under consideration. These tests show that modes associated with eigenvalues 5 and 8 that lie within the corresponding confidence intervals are possibly generated by the presence of colored noise in the GPS time series recorded at Akutan volcano.

phase quadrature. They form a pair that captures an oscillatory component of annual period and represents 7% of the total variance (Figure 9).

ST-PCs 4, 5, and 8, together with ST-PC 1, capture the trend with nonconstant slope; this triplet contains 4% of the variance, for a total of 12% in the data-adaptive trend. ST-PCs 6-7 and ST-PCs 9–11 form two pairs that together capture an oscillatory component of semiannual period; together, these four ST-PCs contain 6% of the data variance (Figure 9). ST-PC 10 captures also an oscillatory component with mostly annual period and contains 1% of the data variance.

3.4. Testing the Statistical Significance of the Extracted Modes

As first pointed out by *Allen and Smith* [1994], colored noise in geophysical time series can lead to the false detection of modes that resemble those of geophysical origin when applying the M-SSA. To make sure that the modes that we consider when reconstructing our time series are not associated with colored noise, appropriate tests of statistical significance have to be applied.

Monte Carlo SSA (MC-SSA) was first proposed by *Allen and Smith* [1994] as such a test for the univariate SSA and then extended by *Allen and Robertson* [1996] to the multivariate case. This test helps discriminate between the modes associated with signals and those associated with noise. MC-SSA relies on constructing sets of "surrogate data" [*Theiler et al.*, 1991] and it allows one to formulate several distinct null hypotheses.

The main idea of MC-SSA is to compare the partial variances in the data set being analyzed, in the directions spanned by its ST-EOFs, with the corresponding variances resulting from a particular null hypothesis. The null hypothesis here takes the form of an ensemble of surrogate data $\mathbf{X}_R$ of the same length N and dimension L as the data set we wish to analyze.

The partial variances of the null hypothesis in the directions defined by the ST-EOFs $\mathbf{E}^k$ are estimated by the diagonal elements of

$$\mathbf{\Lambda}_R = \mathbf{E}'\tilde{\mathbf{T}}_R\mathbf{E}. \tag{11}$$

Here $\mathbf{E}'$ denotes the transpose of the matrix $\mathbf{E}$ whose $L \times M$ columns are the ST-EOFs $\mathbf{E}^k$ deduced from the covariance matrix of the data via equation (6), and $\tilde{\mathbf{T}}_R$ is the covariance matrix of the null hypothesis, calculated using the ensemble of surrogate data $\mathbf{X}_R$ and equation (3).

By calculating $\mathbf{\Lambda}_R$ for a sufficiently large ensemble of surrogate data, one can estimate confidence intervals for the diagonal elements of $\mathbf{\Lambda}_R$. If the eigenvalues of the data covariance matrix lie inside or below these confidence intervals, we conclude that the associated modes are not significant with respect to the null hypothesis under consideration.

This test, however, turns out to be too lenient in the multivariate case; in particular when the covariance matrix $\tilde{\mathbf{T}}$ deduced from the data is rank deficient, i.e., when $L \times M \gg N$ [*Allen and Robertson*, 1996]. This limitation,

however, can be overcome by an improvement of the test proposed by *Groth and Ghil* [2015]. The MC-SSA results reported in the following use the improved MC-SSA algorithm of *Groth and Ghil* [2015].

The difficulty in this kind of test is the choice of the null hypothesis against which the data are tested. Here we choose to test only the statistical significance of the first 11 modes described above and represented in Figure 9. It is indeed among these modes that we identified the different signals that are discussed in the next sections of the paper.

The surrogate data are therefore constructed in two steps. First, we reconstruct the GPS time series using all ST-EOFs and ST-PCs except the first 11 ones. In other words, we reconstruct the time series using equation (9), with k ranging from 12 to $L \times M$. In the second step, we add to these reconstructed time series a noise model against which we wish to test the significance of the first 11 modes. Here two canonical types of noise models are considered : flicker noise with $\alpha = 1$ and random walk noise with $\alpha = 2$. These two types of colored noise are often considered to be present in GPS time series [e.g., *Williams et al.*, 2004].

Figure 11 shows the MC-SSA test results. Error bars represent the two-sided 95% confidence intervals derived from 500 realizations of the surrogate data set, and we concentrate on the first 11 eigenvalues. For the null hypothesis that contains flicker noise (Figure 11a), only eigenvalues 5 and 8 lie within the error bars. They are, however, relatively close to the upper ticks of their associated error bars, i.e., to the 97.5% quantile, and especially so in the case of the eighth eigenvalue, which lies almost on the upper tick. In this case, for eigenvalues 5 and 8, our test does not reject the null hypothesis. However, the proximity of the eigenvalues to the upper tick may be indicative of a type II error or false negative.

For the MC-SSA test against random walk noise (Figure 11b) eigenvalues 5 and 8 clearly lie within their associated error bars. Hence, ST-EOFs 5 and 8 and the corresponding ST-PCs (Figure 9) may have been generated by the presence of random walk noise in the data set and are consequently not used to reconstruct the signals considered in the next sections.

3.5. Seasonal Signals

Seasonal signals with annual and semiannual periods are ubiquitous in GPS position time series and are known to be caused in a large part by mass transport and redistribution in the atmosphere, oceans, and continental water reservoirs, whose load triggers deformation of the Earth surface that is measurable with space-geodetic techniques [e.g., *van Dam and Wahr*, 1998; *Dickey*, 2002; *Bevis et al.*, 2005; *Fu et al.*, 2012].

Seasonal signals contain important geophysical information on the rheological structure of the Earth's crust and mantle [*Chanard et al.*, 2014] and on the spatiotemporal characteristics of the load [e.g., *Blewitt*, 2001; *Heki*, 2004]. They must, therefore, be extracted as objectively as possible from the data.

However, our ability to extract transient deformations from GPS time series depends on our capacity to identify correctly the seasonal signal, since the transients we are interested in may share some of the characteristics of seasonal signals. In particular, the spectral content of the seasonal and transient signals may overlap, which renders classic time series analysis based on the Fourier transform unsuitable. Thus, for instance, *Heki* [2004] showed that Fourier analysis of interannually modulated seasonal oscillations may yield transient-like residuals.

Seasonal signals in GPS time series are commonly modeled using the sum of sinusoidal functions with annual and semiannual periods fitted to the data, which provides an estimate of their phase and amplitude. This procedure is efficient in the case of purely harmonic signals but cannot account, for instance, however, for interannual amplitude modulations or asymmetric oscillations, two characteristics that often are present in actual GPS time series [*Heki*, 2004; *Bennett*, 2008].

M-SSA has long been used in meteorology and climatology to extract such anharmonic and amplitude-modulated oscillatory modes of variability for the atmosphere and oceans [e.g., *Keppenne and Ghil*, 1993; *Plaut and Vautard*, 1994; *Unal and Ghil*, 1995; *Jiang et al.*, 1995]. Here we take advantage of this capacity of M-SSA to extract and reconstruct the seasonal signal present in GPS time series without constraining a priori their period, shape, or time-dependent amplitude.

Figure 10 shows the pair of ST-EOFs 2–3 that captures the annual oscillations in the Akutan GPS time series. The phase quadrature between the two members of the pair emphasizes their being data-adaptive counterparts of the fixed sine and cosine basis functions of Fourier transforms. Their main advantage is that these two ST-EOFs alone describe the annual oscillations in the $L = 24$ time series and their interannual amplitude

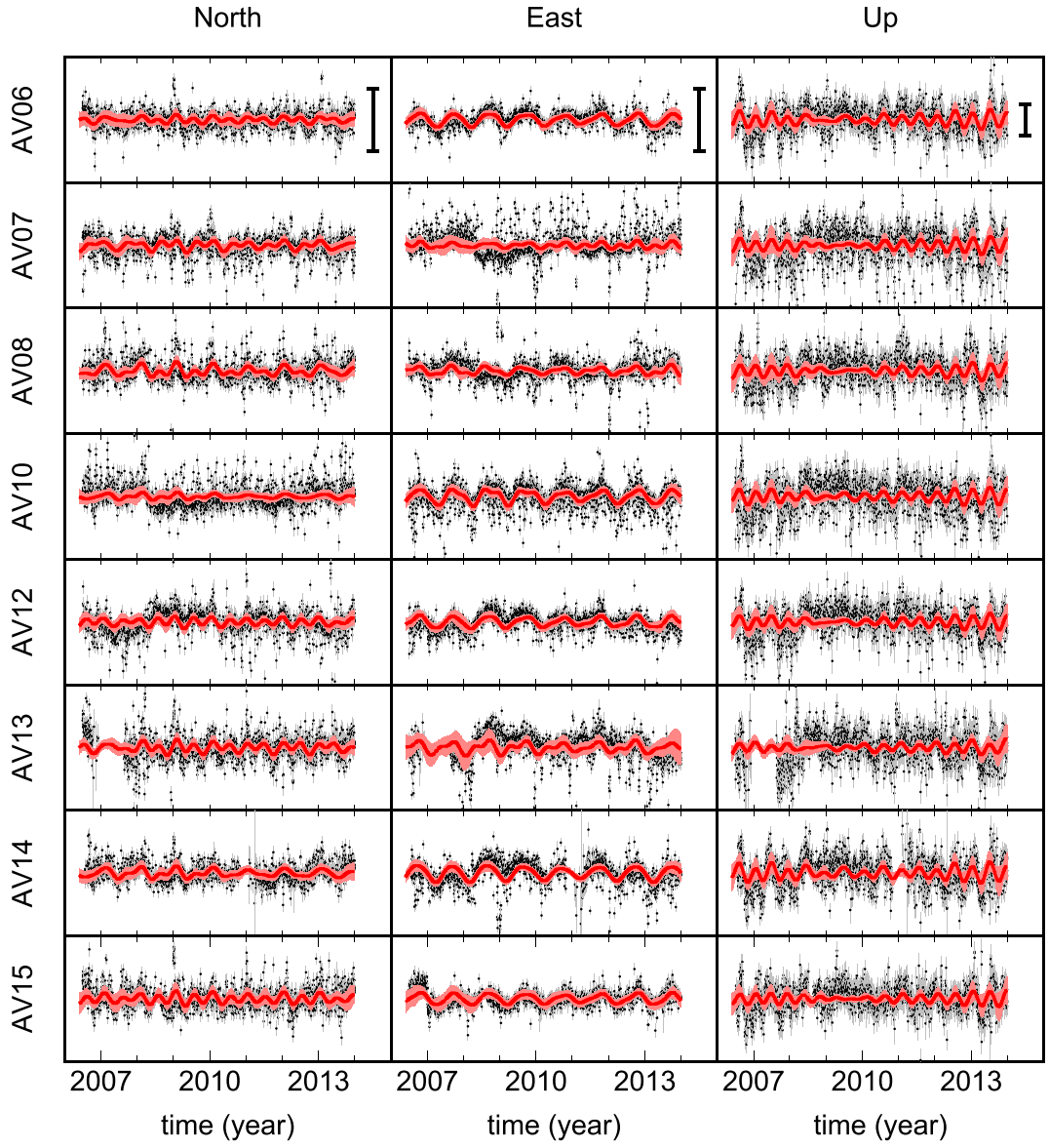

Figure 12. Seasonal variability of the GPS data set for Akutan. Black dots: detrended daily GPS position time series at eight sites on Akutan volcano, Alaska, with their 1 σ error bars. Red lines: reconstructed series after M-SSA using the combination of ST-PCs 2–3 (annual) and 6–7 plus 9–11 (semiannual). The black vertical lines in the topmost panels represent 10 mm. The subpanels correspond to the eight cGPS sites and three directions of motion as in Figure 10. Note the good visual fit to the data. The pluriannual modulation of the seasonal signal is clearly visible, for instance, in the lower amplitude vertical signal for 2009, which appears across all the stations. The light red areas represent the variability of the reconstructions of the seasonal oscillations when the raw time series are perturbated by colored noise with spectral index of 1.

variation [*Ghil et al.*, 2002, and references therein]. The three panels of the figure show very small differences in results for the window widths M = 400, 500, and 600 days, a finding that supports the robustness of the M-SSA results. ST-EOFs 6–7 and 9–11 (not shown) jointly capture the semiannual variability.

The vertical component of the annual signal in Figure 9 is weaker than its horizontal components. The weakness of the vertical component is consistent with Akutan being an island of small size, where vertical deformation caused by atmospheric or hydrological loading—the latter in the form of continental water or snow—is expected to be small [*van Dam and Herring*, 1994]. In such a setting, nontidal ocean mass loading, unaccounted for in the cGPS analysis used here, should represent the largest contribution to seasonal deformation [*Munekane and Matsuzaba*, 2004].

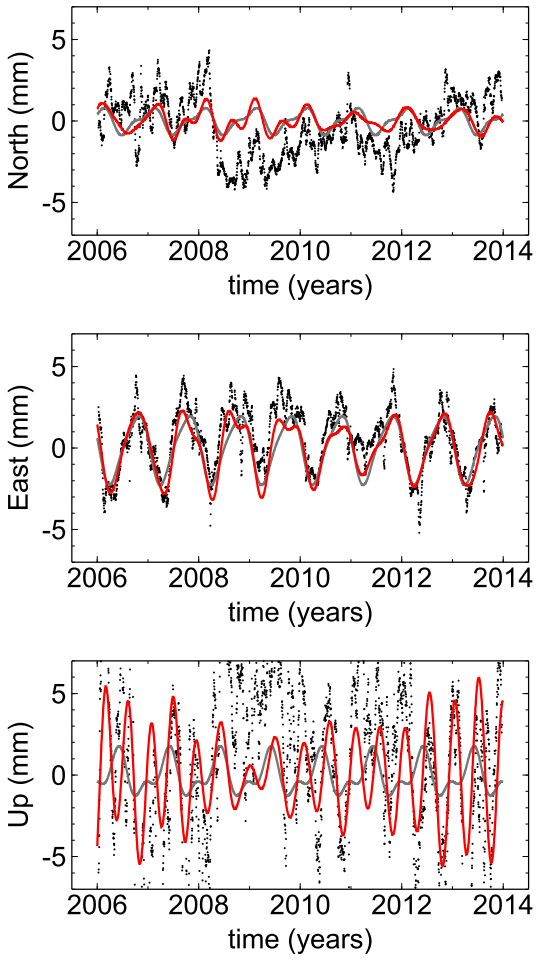

Figure 13. Comparison between the reconstruction of the seasonal signal obtained with M-SSA (red solid curve) and a commonly used fit of the sum (black solid curve) of an annual and a semiannual harmonic function for each component recorded at site AV10: (top) northward, (middle) eastward, and (bottom) upward. We applied a 20 day moving average to the raw data (black dots) to reduce the data scatter and be able to better distinguish the difference between the M-SSA reconstructions and the four-term Fourier series. The two models are, to first order, very close, but M-SSA reconstructions show a pluriannual modulation in amplitude that is not present in the constant-amplitude harmonic functions. Note also that these harmonic functions significantly underestimate the observed range of values in the vertical component. The weighted root mean square (WRMS) are 3.29 mm, 2.99 mm, and 8.44 mm for the M-SSA reconstructions of the seasonal signal for the North, East, and Up components, respectively. The harmonic functions fit have a WRMS of 3.32 mm, 3.05 mm, and 9.14 mm which is systematically higher.

The reconstructed series using the combination of ST-EOFs 2–3, 6–7, and 9–11 and the corresponding ST-PCs (Figure 9) together capture well the seasonal oscillations that are readily apparent in the data, as shown in Figure 12. In spite of the small number of ST-EOFs used, the reconstructed time series exhibits fairly complex waveforms; in particular, a lower amplitude episode in the vertical component of the seasonal signal in the rightmost column in Figure 12 is shared in 2009 by all the stations. This lower amplitude episode is also visible in Figure 13 for the specific site AV10.

We find, in fact, that this modulation is present in the semiannual oscillations associated with ST-PCs 6–7 but not in those associated with ST-PCs 2–3 and 9–11, an observation that may indicate that the ST-PC pair 6–7 captures a specific dynamic process that is different from the one captured by ST-PCs 2–3 and 9–11. Further investigation of these seasonal signals and their multiannual modulation is beyond the scope of the present paper, but our observations illustrate that M-SSA can extract information relevant to the dynamics and the physics driving these signals.

We compare in Figure 13 the seasonal signals extracted using M-SSA from the three components recorded at station AV10, on the one hand, with a curve fit to the data that uses the classical four-term Fourier series

$$x(t) = A_1 \cos(2\pi t) + A_2 \sin(2\pi t) + B_1 \cos(\pi t) + B_2 \sin(\pi t), \tag{12}$$

on the other. Here A_1, A_2, B_1, and B_2 are the coefficients for the annual and semiannual terms estimated from the daily positions $x(t)$, and the fits were done for each one of the three components — north, east, and up — separately.

The difference between M-SSA reconstructions of the seasonal oscillations and the four-term Fourier expansion is most pronounced for the vertical component, which shows interannual modulations of larger amplitude. Furthermore, the M-SSA reconstruction and four-term Fourier estimates for the seasonal oscillations in the vertical are not always exactly in phase, since the coefficients $A_{1,2}$ and $B_{1,2}$ are constant, and thus the phase, as well as the amplitude, is constant, while the M-SSA results have both phase and amplitude modulations.

In addition, the reconstruction of the seasonal modes and their interannual modulations owes to ST-PCs 2–3 and 6–7 for the vertical but ST-PCs 2–3 and 9–11 for the horizontal components. This shows that M-SSA is able to extract annual, semiannual, and interannual oscillations with actually fewer parameters than a least squares fit to a Fourier for which at least six sinusoidal functions would have been necessary to account for such variability.

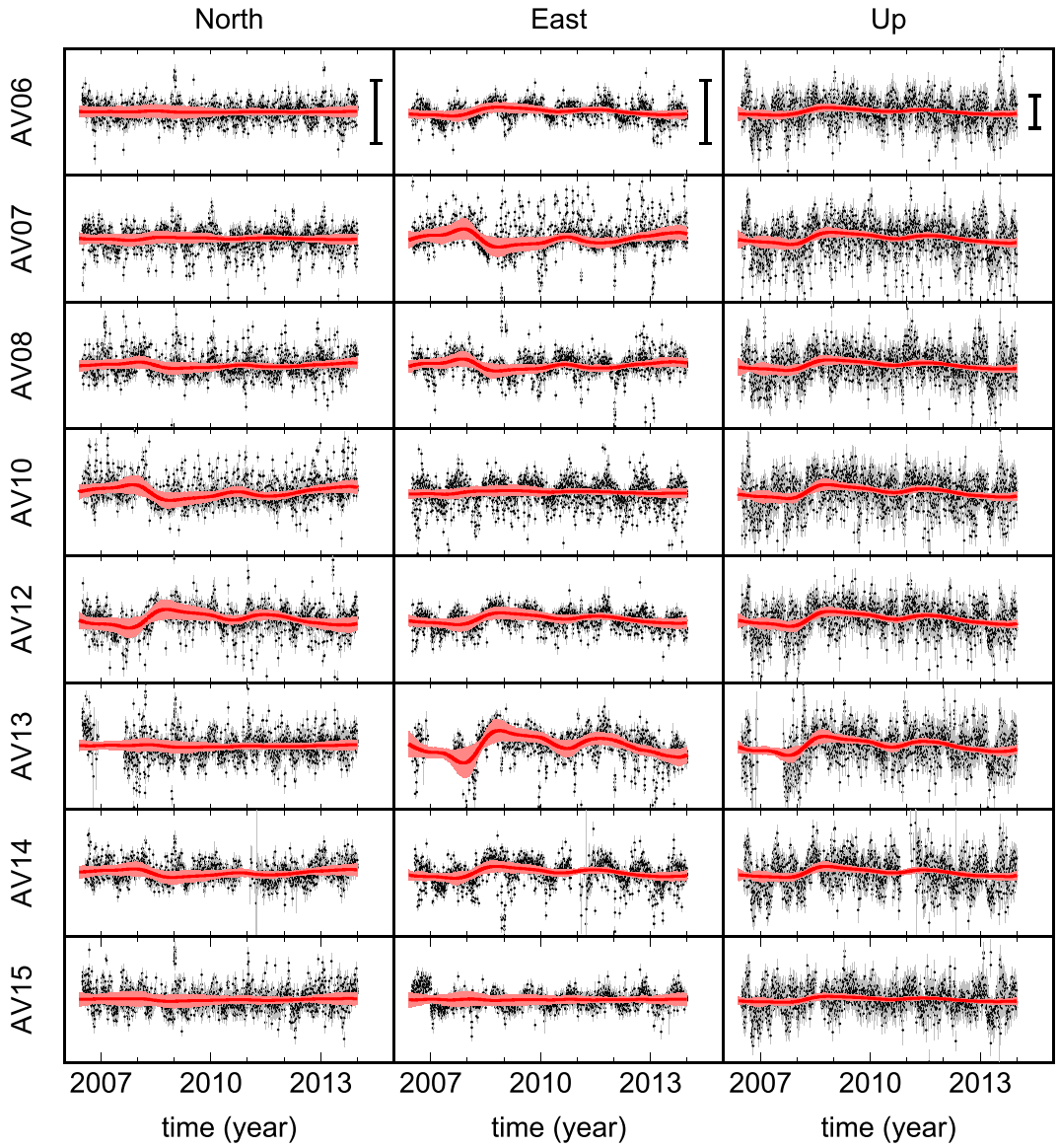

Figure 14. Transient motion at Akutan volcano, Alaska. The subpanels correspond to the eight cGPS sites and three directions of motion as in Figure 10. Note the good visual fit to the data. Black dots: detrended daily GPS position time series at the eight cGPS sites; with their 1 σ error bars. Red lines: M-SSA reconstructions, using the trend ST-PCs 1 and 4, which together carry 11% of the total data variance. The black vertical lines in the topmost panels represent 10 mm. The light red areas represent the variability of the reconstruction of the data-adaptive trends when the raw time series are perturbated by colored noise with spectral index of 1.

These observations show that the M-SSA reconstructions faithfully capture the interannual variability of the seasonal oscillations, with respect to both amplitude and phase modulation. This pluriannual modulation of the seasonal signal identified by M-SSA may be indicative of multiannual oscillations in the regional hydroclimatic regime; see *Ghil and Robertson* [2002, and references therein] including, in particular, *Wallace and Gutzler* [1981].

3.6. Transient Signals

We now focus on ST-PCs 1 and 4 (Figure 9), which carry the data-adaptive trend shared by all 24 GPS time series at Akutan volcano. These ST-PCs together capture 11% of the total variance in the data set (Figure 8) and are associated with transient behavior, mostly described by the shape of ST-PC 1. This transient is shown by the reconstructions in Figure 14.

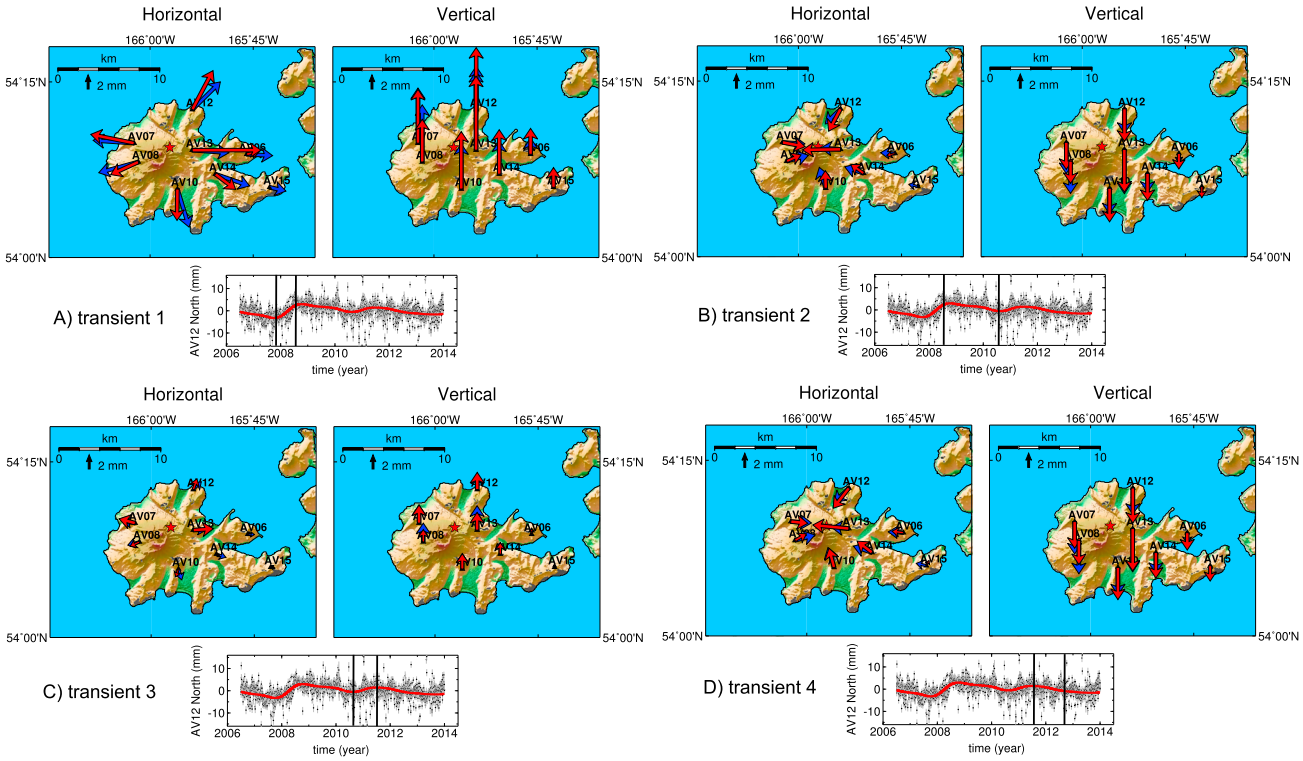

Figure 15. Maps of the horizontal and vertical displacements at Akutan volcano that are associated with the (a–d) four dominantly monotonic parts of the reconstructed trends derived from ST-PCs 1 and 4. For each event, the two top panels plot the vector displacements at the eight cGPS sites, while the bottom panel plots the time series of northward displacements at site AV12. In this latter panel, black dots show the raw data and the red solid curves show the M-SSA-reconstructed trend signals, while the blue solid curves show the best fit prediction from a simple, idealized model of magma chamber expansion [*Mogi*, 1958]. In the top panels, the red arrows are based on the M-SSA reconstructions and the blue ones are based on the *Mogi* [1958] model. Note the spatial consistency of the displacements for each of the four successive episodes with a radial pattern centered on the volcano that is consistent with a magmatic origin.

The M-SSA-filtered time series (red curves in the figure) display a transient deformation event starting in early 2008 and lasting for about half a year. This transient event is similar in shape and duration to the one identified by *Ji and Herring* [2011], and it is visible in Figure 14 at all eight sites and in all three directions of motion; the amplitude of the motion and its direction vary, though, from site to site.

The event is already visible in some of the raw time series, such as the northward component at AV12, or the eastward component at AV07. But its shape and duration in the raw data, where all the signals and the noise are mixed, sometimes resemble part of a seasonal cycle—e.g., in the eastward component at AV12—or a sharp step function that could be mistaken for an equipment problem—e.g., in the northward component at AV10.

In Figure 15, we consider four transient episodes dominated by vertical uplift. We compare the associated ground motions (red arrows in the figure) with those given by a calculation that uses the simple model of a point source volume change, simulating a magma chamber, embedded in an elastic half-space (blue arrows in the figure). Such a model—also used by *Ji and Herring* [2011]—was first proposed by *Mogi* [1958] as a simple way of calculating surface deformation in active volcanic systems.

The spatial representation of the site displacements from 2008.0 to 2008.7 in Figure 15a shows that this event is associated with a radial extension and uplift centered on the volcano, with amplitudes decreasing outward. This is consistent with the observations reported by *Ji and Herring* [2011] and validates the extraction of this transient by M-SSA.

Figures 15a–15d also show other features shared among the filtered time series, across the four episodes, although these features are more subtle than the one described above. We observe that the first transient, from 2008.0 to 2008.7 (Figure 15a), is followed at all sites by another monotonic trend with an opposite slope, from 2010.5 to 2010.8 (Figure 15b), then by a shorter monotonic trend, from 2010.8 to 2011.5 (Figure 15c),

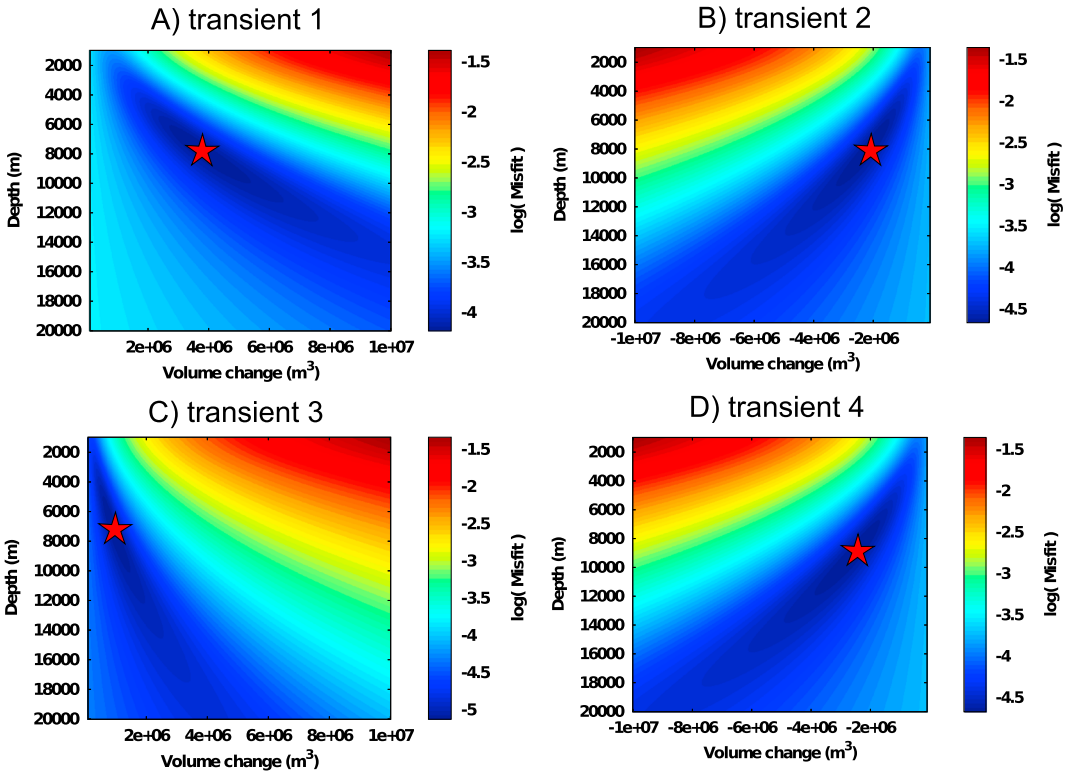

Figure 16. Misfit plots between the deformation predicted by a Mogi source and the observed displacements for each of the four transient events shown in Figure 15. The four events are well explained by inflation (uplift and radial extension at the surface) or deflation (subsidence and radial contraction at the surface) of the same point source at 8 ± 1 km. These results strongly support the hypothesis that the small but coherent deformations extracted by M-SSA are of magmatic origin.

with the same slope as in Figure 15a. Finally, we observe fourth monotonic trend, from 2011.5 to the end of the data used here (Figure 15d), whose slope is again opposite to the previous one.

The amplitudes and slopes of these trends vary from site to site: they are subtle and very difficult to see in the raw time series, in particular, in the noisier vertical component, where they are buried in the noise but still extracted by our M-SSA analysis. The match between the horizontal component of the raw series and the filtered reconstruction, though, is readily visible at those sites and for those components for which the amplitude of these trends is relatively large, such as the eastward component at AV07, the northward component at AV12, or the eastward component at AV13.

The site displacements associated with the three following transients are plotted in Figures 15b–15d. They also show a well-organized pattern of radial contraction (episode 2), extension (episode 3), and compression (episode 4), associated with subsidence, uplift, and subsidence, respectively. All three events have amplitudes of less than 4 mm, i.e., they are much smaller than the first. Event 3 is the least pronounced one, with displacements that do not exceed 2 mm. In spite of these very small displacements, their common features—to wit, the radial pattern centered on the volcano, the good correlation between horizontal and vertical components, and the decreasing amplitudes with distance from the volcano summit reinforce our confidence that they represent a signal of geophysical origin rather than noise.

We are, therefore, observing—over the roughly 8 years of data analyzed—two cycles of uplift and radial extension, followed by subsidence and radial contraction, and both cycles consistently display features that strongly suggest a magmatic origin. We test this hypothesis by running a series of "Mogi models," as described above, for each of the four transient deformation episodes. For each episode we vary the depth and volume change of the source, while its horizontal position is kept constant at the latitude and longitude used by *Ji and Herring* [2011]. Figure 16 shows that the minimum misfit between model and data analysis results is obtained

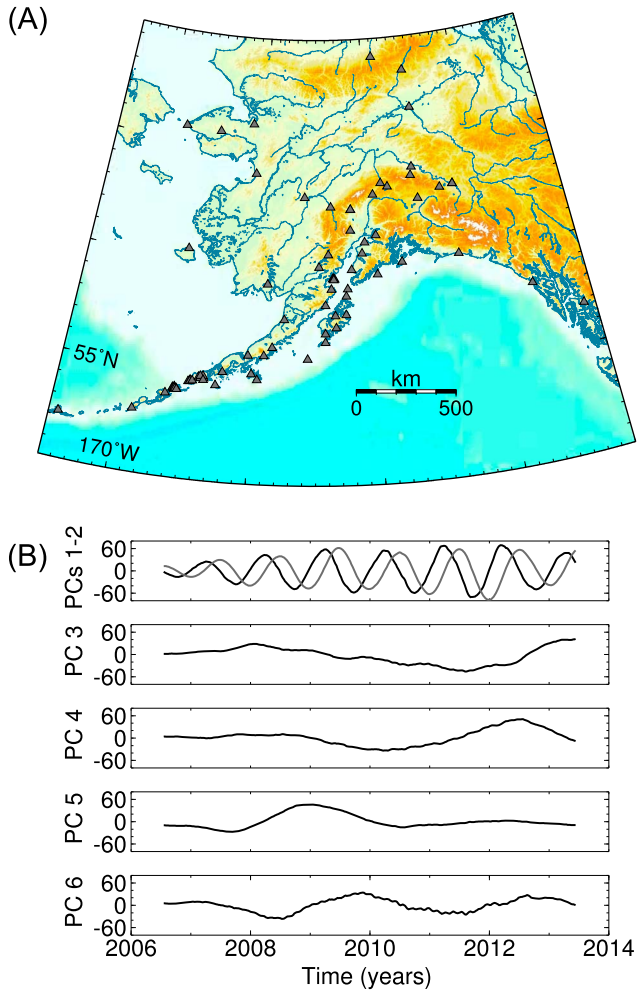

Figure 17. M-SSA analysis of the PBO network of 80 GPS stations from Alaska: (a) Geographic distribution of the stations; and (b) the six leading ST-PCs calculated by applying M-SSA to the position time series from this network. ST-PCs 1 and 2 are in phase quadrature and represent a seasonal oscillation with an annual period. The next four ST-PCs represent data-adaptive trends.

for a source located at a depth of close to 8 km for all four sources. The fact that a single source location can describe all four transient events is a strong indication that they are, indeed, of magmatic origin and, therefore, that the very small transients extracted by M-SSA are indeed representative of a geophysical signal.

4. Discussion

4.1. Implications for Transient Detection in Large GPS Networks

We have so far been using GPS stations located within roughly 25 km of each other and thus strong spatial and temporal correlations are expected, making M-SSA particularly efficient at extracting common modes of variability from the data. We now address the question of whether the transient signal described above would have been detected across a larger station footprint and with no a priori information on its location.

The latter situation is particularly relevant given the large number of GPS stations now commonly deployed across wide areas. Visually combing through these large data sets for transient events is becoming more and more difficult, and it is therefore necessary to develop automated approaches to identifying geophysical signals in a systematic and objective way.

We applied, therefore, M-SSA to GPS position time series from 80 sites distributed across Alaska, including those on the Akutan volcano; Their geographical distribution is plotted in Figure 17a. The computational burden increases rapidly because of the dimension of the spatiotemporal lag covariance matrix $\tilde{\mathbf{T}}$, which is

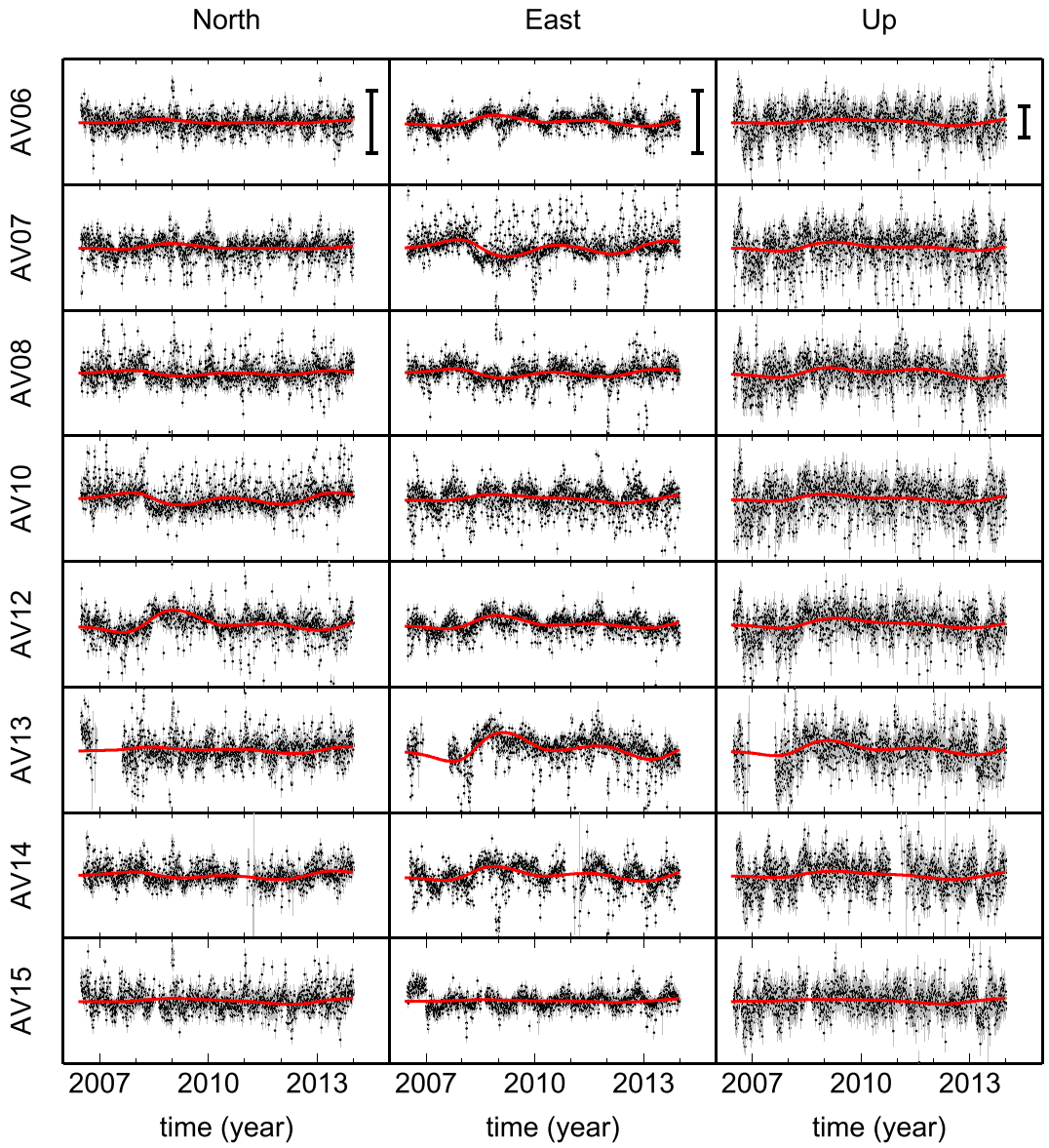

Figure 18. Reconstructed time series (solid red lines) using ST-PCs 3–6 of Figure 17b, which represent data-adaptive trends. The black vertical lines in the topmost panels represent 10 mm. The reconstructions here are for the same eight cGPS sites on Akutan volcano as plotted in Figure 14 and are presented in the same format. Here, however, the reconstructions are calculated from the 80 × 3 time series from the PBO network covering Alaska and shown in Figure 17a. Note the similarity with the reconstructed series derived solely from the 8 × 3 Akutan GPS time series, plotted in Figure 14. For 15 (∼60%) of them the correlation coefficients between the reconstructed time series from eight stations and 80 stations are larger than 0.7. For seven of them (∼30%) the correlation coefficients are less than 0.5. However, these seven time series are those for which the signal is very weak such as in the three components recorded at AV15.

$(L \times M)^2$. We then resampled the data weekly. It allows us to reduce the window width M without making the ST-EOFs length smaller than 365 days, which is necessary to extract the seasonal oscillations as explained in section 3.2. The M-SSA analysis proceeds the same way as described in section 3 above and it produces six leading ST-PCs, shown in Figure 17b; this set of ST-PCs describes 31% of the data variance.

We observe that over such a large spatial scale the first two ST-PCs correspond to an oscillatory component with an annual period, while the remaining four ST-PCs represent trends. The time series reconstructed using these trends, plotted in Figure 18, display the same transients as described above using the eight-site M-SSA of Figure 9. The corresponding site displacements, plotted in Figure 19, stand out as a local anomaly at the

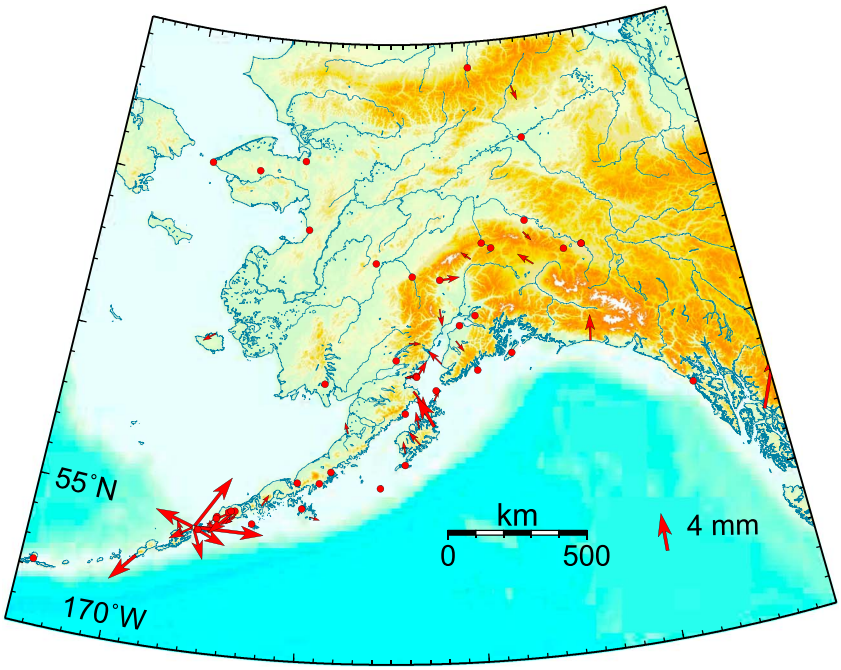

Figure 19. Map of the horizontal displacements associated with the first transient deformation event identified above for the Alaskan PBO network; this event corresponds to the time interval 2008.0–2008.7, cf. Figure 18. The red arrow over the Gulf of Alaska indicates a horizontal displacement of more than 1 mm. To make them visible we indicate GPS sites with displacements less than 1 mm with red dots. Note that displacements at Akutan stand out as anomalously large, with a radial pattern. Figure 20 below shows a close-up on Akutan.

scale of Alaska. The displacements over Akutan, as obtained by the Alaska-wide M-SSA analysis, are plotted in Figure 20. They show a radial pattern of extension as seen by the M-SSA with the eight sites on Figure 15a.

Having applied M-SSA to the 80 GPS stations covering a large footprint, we thus found that the method is still able to identify common modes of variability at the specific location of the Akutan volcano. As a consequence, applying M-SSA to the uncorrected raw data of a regional network of stations—i.e., without removing a seasonal signal, modifying the frame of reference, or assuming a given set of noise characteristics—should allow one to detect statistically significant local anomalies. Such anomalies can then be further investigated by zooming in on them and verifying that the corresponding spatial patterns are consistent with a geophysical process [e.g., *Ji and Herring*, 2011]. This result opens the way for the automated detection of transient events in large GNSS networks.

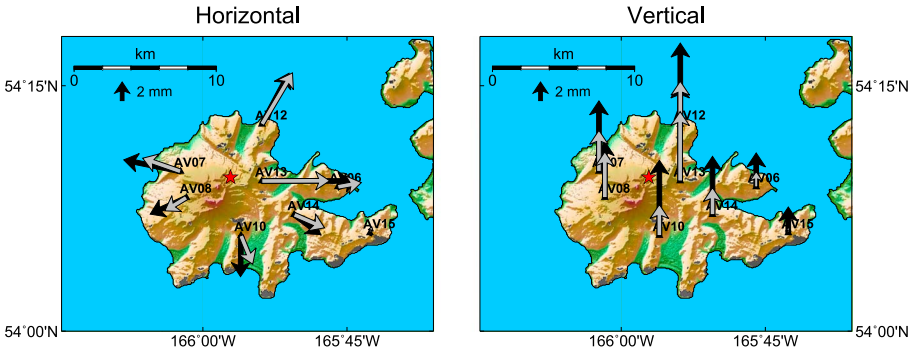

Figure 20. Comparison of the horizontal and vertical displacements associated with the first transient-deformation event identified above (2008.0 to 2008.7), as extracted from the eight-site M-SSA (black arrows versus the 80-site, Alaska-wide M-SSA (white arrows). Note the agreement between the local and the regional-scale M-SSA estimates: both show a radial pattern of extension and uplift centered on the volcano. This pattern is consistent with a magmatic origin, as seen in Figure 15 above.

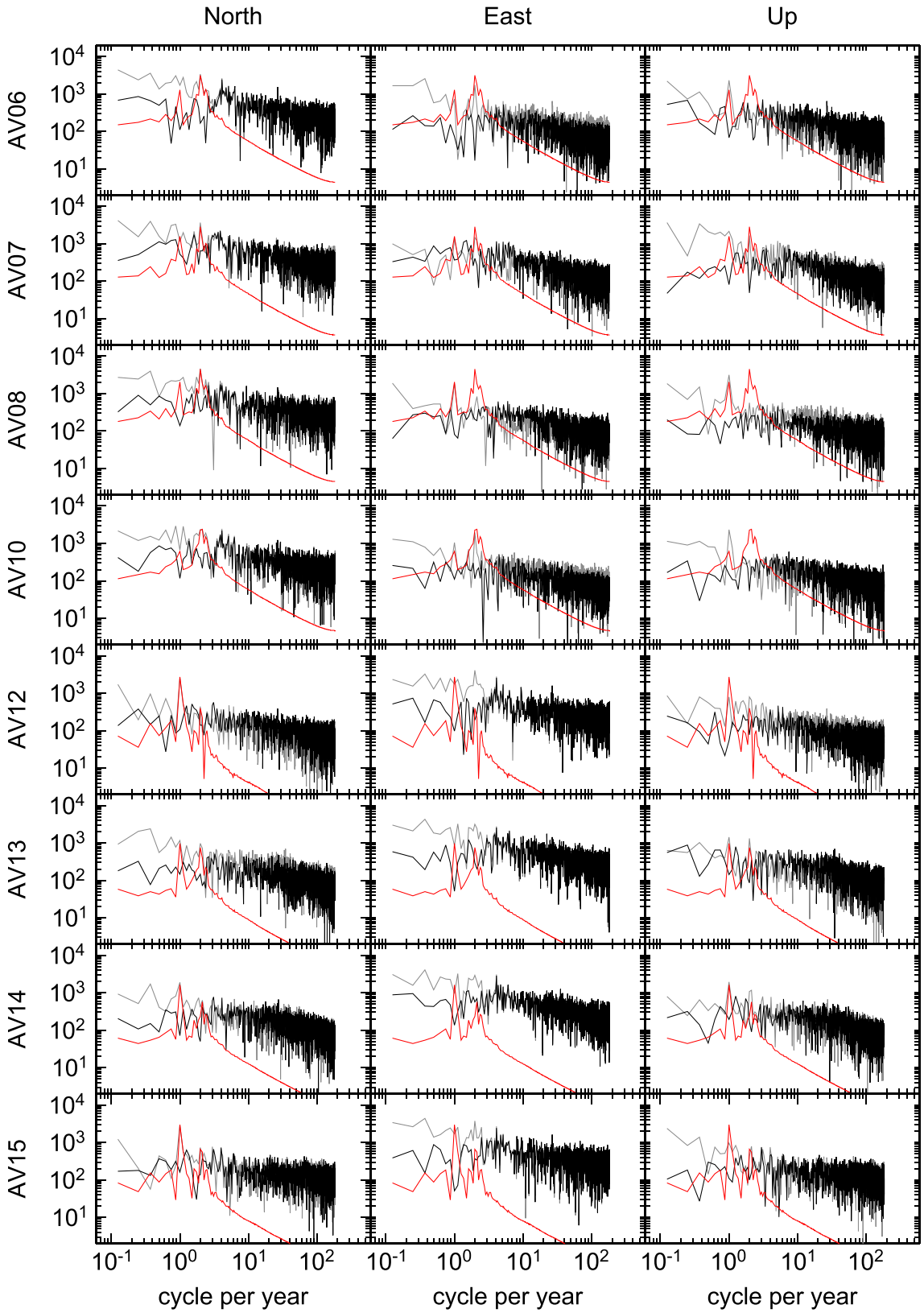

Figure 21. Power spectral density plots of the raw time series (gray lines) and M-SSA-based seasonal signal (red); the residual component, i.e., raw minus the M-SSA reconstruction of the transient displacements plus the seasonal signals are plotted as black lines. Note the flattening of the spectrum of the residuals at low frequency, indicating that the trend and the oscillatory components are extracted by M-SSA while preserving the high-frequency noise content.

4.2. Noise Extraction From GPS Time Series

Noise in GPS position time series is temporally correlated with more power at low frequencies than at high frequencies [e.g., *Zhang et al.*, 1997; *Mao et al.*, 1999; *Williams*, 2003]. The power spectrum of GPS time series on the Akutan volcano (gray lines in Figure 21) follows this characteristic behavior with a slope of roughly −1 of the spectral density at high frequencies indicative of temporally correlated noise.

Such temporal correlations can lead to time-dependent patterns similar to geophysical signals, in particular, when the spectral index is greater than 1 [*Agnew*, 1992]; see also the discussion of the spectral index α in section 2.2, cf. equation (7) and Figure 6. It is clear from the figure, furthermore, that the power spectrum of the raw time series at low frequencies is not flat and includes the contributions from trends and oscillatory signals described above.

We take advantage of the ability of M-SSA to extract common modes of variability among time series jointly in space and time. In particular, we do expect it to discriminate between the spectral content associated with time-dependent noise, on the one hand, and signals such as transient deformation and seasonal motions, on the other hand.

This expectation is tested in Figure 21, where black lines show the power spectrum of the noise, i.e., the raw time series minus the reconstruction of the transient plus the seasonal signals. We observe that the noise carries much less low frequency variance than the raw time series and that its spectrum is quite flat at low frequencies. This finding confirms our expectation that the low-frequency part of the raw-data spectrum (gray lines in Figure 21) is mostly due to the sum of the seasonal and transient signals extracted by M-SSA. The high-frequency content of the residual time series is, however, unaffected, with a slope identical to that of the raw data. This shows that M-SSA act here as a low-pass filter.

The power spectra of the reconstructed seasonal oscillations in Figure 12 are plotted in red in Figure 21. Two well-defined peaks are visible at annual and semiannual periods. The transient signal also contributes to the variance present at these periods, but M-SSA is able to unravel the spectral content associated with these transients from the one associated with the seasonal oscillations.

This efficient extraction of trends, oscillatory modes, and noise by M-SSA is possible because the method captures the different temporal and spatial scales of the physical processes responsible for the generation of the time series.

4.3. M-SSA, GPS Time Series, and Nonlinear Dynamical Systems

The idea that time series analysis could be applied to extract information about underlying nonlinear dynamical systems was proposed by *Mañé* [1981] and by *Takens* [1981]. *Broomhead and King* [1986a, 1986b] and subsequent authors [*Vautard and Ghil*, 1989; *Ghil and Vautard*, 1991; *Vautard et al.*, 1992] reformulated the early "method of delays" in the setting of the classical Karhunen-Loéve representation of time series, as reviewed by *Ghil et al.* [2002], which is at the basis of the M-SSA methodology.

Single-channel SSA and M-SSA have thus been linked to the framework of nonlinear dynamical system theory. This connection helps explain the fact that M-SSA allows one to extract physically relevant modes of variability common to several GPS time series because these modes reflect the structure of the underlying attractor.

This make the M-SSA particularly suitable to analyze GPS time series generated by geophysical nonlinear dynamical systems and a fortiori by volcanoes. Volcanoes are intrinsically chaotic nonlinear dynamical systems as highlighted by *Sparks* [2003]. The processes often invoked to explain inflations and subsidences observe with geodetic data are crystallization, melting, and gas exsolutions along with the dynamic of the hydrothermal systems [e.g., *Caricchi et al.*, 2014]. All these phenomena can be coupled with each other and account (among others) for the nonlinearity of volcanic systems [*Sparks*, 2003].

The deformation of the Akutan volcano that has been extracted with the M-SSA reflect well this nonlinear dynamic. They are cyclic — we observe successions of inflations and subsidences — but not periodic. This kind of chaotic behavior arise typically from nonlinear dynamical systems that have at least 3 degrees of freedom as first shown by *Lorenz* [1963] in his famous paper.

GPS time series can obviously provide information not only on magmatic processes — as shown here — but also on tectonic, hydrological, and seismological processes that affect systems constituted of faults, volcanoes, crust, and mantle. Such systems are complex and can be seen as well as nonlinear dynamical systems for which we often have limited knowledge of the physics that precisely controls their evolution in time.

Regardless of this limitation, M-SSA allows us to efficiently and objectively separate processes and to rank them as a function of their importance, i.e., of the fraction of variance in the observed time series that they generate. Some of these processes correspond to what is classically considered as the "signal," i.e., as the deterministic part, while others correspond to what is classically considered as "noise," i.e., as the stochastic part.

5. Conclusion

We have shown that M-SSA allows us to extract signals of geophysical significance from GPS position time series and to efficiently separate them from noise. The data-adaptive nature of the method results in a simple and parsimonious representation, using only a few spatiotemporal EOFs (ST-EOFs) to describe complex spatiotemporal fluctuations, such as anharmonic oscillations and trends with nonconstant slopes. This representation is achieved without any a priori assumptions on the stochastic characteristics of the noise or on the cause of the underlying physical processes.

Like other methods, M-SSA is sensitive to the quality of the data. In particular, the presence of colored noise can lead to the occurrence of modes that resemble those of geophysical origins. We have shown, however, that hypothesis tests, such as the MC-SSA, help us to discriminate between modes associated with correlated noise and modes associated with signals. The presence of discrete offsets, due, for example, to equipment changes at the GPS sites can also lead to bad reconstructions of the time series. Visual inspection of the data may, therefore, be necessary, although new techniques are being proposed to automatically detect and correct such shifts in time series [*Montillet et al.*, 2015; *Tran*, 2013].

Another limitation of the M-SSA is that the size of the $(M \times L)^2$ lag covariance matrix $\tilde{\mathbf{T}}$ grows rapidly with the size of the GPS network considered. This can lead to computational issues in solving the eigenvalue problem described in equation (6).

Despite these limitations M-SSA presents an attractive alternative to existing methods to objectively extract information embedded in geodetic time series. Its ability to filter and compress the information contained in GPS time series through the calculation of robust and parsimonious representations opens the perspective to use it as a tool to detect signals hidden in the large amount of geodetic data now available worldwide.

Acknowledgments
This research is supported by a doctoral grant from PSL Research University (D.W.), by research grant 9AD01274 from the INSU/CNRS program "PNTS" (Programme National de Télédétection Spatiale) (E.C.) and by National Science Foundation grant OCE-1243175 and Office of Naval Research grant N00014-12-1-0911 (M.G.). This work uses data products provided by the Plate Boundary Observatory operated by UNAVCO for EarthScope (www.earthscope.org) and supported by the National Science Foundation grants EAR-0350028 and EAR-0732947.

References

Agnew, D. C. (1992), The time-domain behavior of power-law noises, *Geophys. Res. Lett.*, *19*(4), 333–336, doi:10.1029/91GL02832.
Allen, M. R., and A. W. Robertson (1996), Distinguishing modulated oscillations from coloured noise in multivariate datasets, *Clim. Dyn.*, *12*(11), 775–784.
Allen, M. R., and L. A. Smith (1994), Investigating the origins and significance of low-frequency modes of climate variability, *Geophys. Res. Lett.*, *21*(10), 883–886, doi:10.1029/94GL00978.
Bennett, R. A. (2008), Instantaneous deformation from continuous GPS: Contributions from quasi-periodic loads, *Geophys. J. Int.*, *174*(3), 1052–1064.
Bevis, M., D. Alsdorf, E. Kendrick, L. P. Fortes, B. Forsberg, R. Smalley, and J. Becker (2005), Seasonal fluctuations in the mass of the Amazon River system and Earth's elastic response, *Geophys. Res. Lett.*, *32*, L16308, doi:10.1029/2005GL023491.
Blewitt, G. (2001), A new global mode of Earth deformation: Seasonal cycle detected, *Science*, *294*(5550), 2342–2345, doi:10.1126/science.1065328.
Broomhead, D. S., and G. P. King (1986a), Extracting qualitative dynamics from experimental data, *Physica D*, *20*(23), 217–236, doi:10.1016/0167-2789(86)90031-X.
Broomhead, D. S., and G. P. King (1986b), On the qualitative analysis of experimental dynamical systems, in *Nonlinear Phenomena and Chaos*, edited by S. Sarkar, pp. 113–144, Adam Hilger, Bristol, England.
Caricchi, L., J. Biggs, C. Annen, and S. Ebmeier (2014), The influence of cooling, crystallisation and re-melting on the interpretation of geodetic signals in volcanic systems, *Earth Planet. Sci. Lett.*, *388*, 166–174.
Chanard, K., J. P. Avouac, G. Ramillien, and J. Genrich (2014), Modeling deformation induced by seasonal variations of continental water in the Himalaya region: Sensitivity to Earth elastic structure, *J. Geophys. Res. Solid Earth*, *119*, 5097–5113, doi:10.1002/2013JB010451.
Chen, Q., T. van Dam, N. Sneeuw, X. X. Collilieux, M. M. Weigelt, and P. Rebischung (2013), Singular spectrum analysis for modeling seasonal signals from GPS time series, *J. Geodyn.*, *72*, 25–35.
Dickey, J. O. (2002), Interdisciplinary space geodesy: Challenges in the new millennium, in *Vistas for Geodesy in the New Millennium, Int. Assoc. of Geod.*, edited by J. O Dickey, pp. 590–594, Springer, Berlin.
Feng, L., and A. V. Newman (2009), Constraints on continued episodic inflation at Long Valley Caldera, based on seismic and geodetic observations, *J. Geophys. Res.*, *114*, B06403, doi:10.1029/2008JB006240.
Fu, Y., J. T. Freymueller, and T. Jensen (2012), Seasonal hydrological loading in southern Alaska observed by GPS and GRACE: Seasonal loading in Southern Alaska, *Geophys. Res. Lett.*, *39*, L15310, doi:10.1029/2012GL052453.
Ghil, M., and A. W. Robertson (2002), "Waves" vs "particles" in the atmosphere's phase space: A pathway to long-range forecasting?, *Proc. Natl. Acad. Sci. U.S.A.*, *99*(Suppl. 1), 2493–2500.
Ghil, M., and R. Vautard (1991), Interdecadal oscillations and the warming trend in global temperature time series, *Nature*, *350*, 324–327.
Ghil, M., M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou (2002), Advanced spectral methods for climatic time series, *Rev. Geophys.*, *40*(1), 1003, doi:10.1029/2000RG000092.

Groth, A., and M. Ghil (2015), Monte Carlo Singular Spectrum Analysis (SSA) revisited: Detecting oscillator clusters in multivariate datasets, *J. Clim.*, *28*(19), 7873–7893.
Heki, K. (2004), Dense GPS array as a new sensor of seasonal changes of surface loads, in *The State of the Planet: Frontiers and Challenges in Geophysics*, pp. 177–196, AGU, Washington, D. C.
Herring, T., M. Craymer, G. Sella, R. Snay, G. Blewitt, D. Argus, Y. Bock, E. Calais, J. Davis, and M. Tamisiea (2008), SNARF 2.0: A regional reference frame for North America, *Eos Trans. AGU*, *89*(Joint Assem. Suppl.), G21–01.
Ji, K. H., and T. A. Herring (2011), Transient signal detection using GPS measurements: Transient inflation at Akutan volcano, Alaska, during early 2008, *Geophys. Res. Lett.*, *38*, L06307, doi:10.1029/2011GL046904.
Ji, K. H., and T. A. Herring (2013), A method for detecting transient signals in GPS position time-series: Smoothing and principal component analysis, *Geophys. J. Int.*, *193*(1), 171–186, doi:10.1093/gji/ggt003.
Jiang, N., J. D. Neelin, and M. Ghil (1995), Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific, *Clim. Dyn.*, *12*(2), 101–112.
Keppenne, C. L., and M. Ghil (1993), Adaptive filtering and prediction of noisy multivariate signals: An application to subannual variability in atmospheric angular momentum, *Intl. J. Bifurcation Chaos*, *3*, 625–634.
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*, B03409, doi:10.1029/2006JB004448.
Lorenz, E. N. (1963), Deterministic nonperiodic flow, *J. Atmos. Sci.*, *20*(2), 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
Mañé, R. (1981), On the dimension of the compact invariant sets of certain non-linear maps, in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898, pp. 230–242, Springer, Berlin.
Mao, A., C. G. A. Harrison, and T. H. Dixon (1999), Noise in GPS coordinate time series, *J. Geophys. Res.*, *104*(B2), 2797–2816, doi:10.1029/1998JB900033.
Miller, M. M., T. Melbourne, D. J. Johnson, and W. Q. Sumner (2002), Periodic slow earthquakes from the Cascadia subduction zone, *Science*, *295*(5564), 2423–2423.
Mogi, K. (1958), Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them, *Bull. Earthquake Res. Inst.*, *36*, 99–134.
Montillet, J.-P., S. D. P. Williams, A. Koulali, and S. C. McClusky (2015), Estimation of offsets in GPS time-series and application to the detection of earthquake deformation in the far-field, *Geophys. J. Int.*, *200*(2), 1205–1219, doi:10.1093/gji/ggu473.
Munekane, H., and S. Matsuzaba (2004), Nontidal ocean mass loading detected by GPS observations in the tropical Pacific region, *Geophys. Res. Lett.*, *31*, L08602, doi:10.1029/2004GL019773.
Ohtani, R., J. J. McGuire, and P. Segall (2010), Network strain filter: A new tool for monitoring and detecting transient deformation signals in GPS arrays, *J. Geophys. Res.*, *115*, B12418, doi:10.1029/2010JB007442.
Plaut, G., and R. Vautard (1994), Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere, *J. Atmos. Sci.*, *51*(2), 210–236, doi:10.1175/1520-0469(1994).
Preisendorfer, R. W., F. Zwiers, and T. Barnett (1981), *Foundations of Principal Component Selection Rules, SIO Reference Ser.*, Scripps Inst. of Oceanogr., 192 pp., Univ. of Calif.
Riel, B., M. Simons, P. Agram, and Z. Zhan (2014), Detecting transient signals in geodetic time series using sparse estimation techniques, *J. Geophys. Res. Solid Earth*, *119*, 5140–5160, doi:10.1002/2014JB011077.
Sparks, R. (2003), Forecasting volcanic eruptions, *Earth Planet. Sci. Lett.*, *210*(1), 1–15.
Takens, F. (1981), Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898, pp. 366–381, Springer, Berlin.
Theiler, J., B. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer (1991), Testing for nonlinearity in time series: The method of surrogate data, *Tech. Rep.*, Los Alamos National Laboratory, Los Alamos, N. M.
Tran, D. T. (2013), Analyse rapide et robuste des solutions GPS pour la tectonique, PhD thesis, Univ. of Nice Sophia Antipolis, Nice, France.
Unal, Y. S., and M. Ghil (1995), Interannual and interdecadal oscillation patterns in sea level, *Clim. Dyn.*, *11*(5), 255–278.
van Dam, T. M., and T. A. Herring (1994), Detection of atmospheric pressure loading using very long baseline interferometry measurements, *J. Geophys. Res.*, *99*(B3), 4505–4517.
van Dam, T. M., and J. Wahr (1998), Modeling environment loading effects: A review, *Phys. Chem. Earth*, *23*(910), 1077–1087, doi:10.1016/S0079-1946(98)00147-5.
Vautard, R., and M. Ghil (1989), Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, *35*(3), 395–424.
Vautard, R., P. Yiou, and M. Ghil (1992), Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica D*, *58*, 95–126.
Wallace, J. M., and D. S. Gutzler (1981), Teleconnections in the geopotential height field during the Northern Hemisphere winter, *Mon. Weather Rev.*, *109*(4), 784–812, doi:10.1175/1520-0493(1981).
Williams, S. D. P. (2003), The effect of coloured noise on the uncertainties of rates estimated from geodetic time series, *J. Geod.*, *76*(9–10), 483–494, doi:10.1007/s00190-002-0283-4.
Williams, S. D. P., Y. Bock, P. Fang, P. Jamason, R. M. Nikolaidis, L. Prawirodirdjo, M. Miller, and D. J. Johnson (2004), Error analysis of continuous GPS position time series, *J. Geophys. Res.*, *109*, B03412, doi:10.1029/2003JB002741.
Zhang, J., Y. Bock, H. Johnson, P. Fang, S. Williams, J. Genrich, S. Wdowinski, and J. Behr (1997), Southern California permanent GPS geodetic array: Error analysis of daily position estimates and site velocities, *J. Geophys. Res.*, *102*(B8), 18,035–18,055, doi:10.1029/97JB01380.