Inverting geodetic time series with a principal component analysis-based inversion method

A. P. Kositsky[1,2] and J.-P. Avouac[1]

[1] The Global Positioning System (GPS) system now makes it possible to monitor deformation of the Earth's surface along plate boundaries with unprecedented accuracy. In theory, the spatiotemporal evolution of slip on the plate boundary at depth, associated with either seismic or aseismic slip, can be inferred from these measurements through some inversion procedure based on the theory of dislocations in an elastic half-space. We describe and test a principal component analysis-based inversion method (PCAIM), an inversion strategy that relies on principal component analysis of the surface displacement time series. We prove that the fault slip history can be recovered from the inversion of each principal component. Because PCAIM does not require externally imposed temporal filtering, it can deal with any kind of time variation of fault slip. We test the approach by applying the technique to synthetic geodetic time series to show that a complicated slip history combining coseismic, postseismic, and nonstationary interseismic slip can be retrieved from this approach. PCAIM produces slip models comparable to those obtained from standard inversion techniques with less computational complexity. We also compare an afterslip model derived from the PCAIM inversion of postseismic displacements following the 2005 8.6 Nias earthquake with another solution obtained from the extended network inversion filter (ENIF). We introduce several extensions of the algorithm to allow statistically rigorous integration of multiple data sources (e.g., both GPS and interferometric synthetic aperture radar time series) over multiple timescales. PCAIM can be generalized to any linear inversion algorithm.

1. Introduction

[2] Faults slip in a variety of ways, such as during sudden seismic events or as a result of aseismic creep. Fault slip rates can therefore vary over a wide range of timescales, from the typical 10–100 s duration of large earthquakes, to the weeks or years duration of slow earthquakes and postseismic relaxation. Monitoring how fault slip varies with time is thus key to improving our understanding of fault behavior. Fault slip at depth results in surface deformation that can be observed with geodetic techniques [e.g., *Lisowski et al.*, 1991; *Segall and Davis*, 1997], paleogeodetic techniques [e.g., *Taylor et al.*, 1987; *Sieh et al.*, 1999], or remote sensing techniques [e.g., *Massonnet and Feigl*, 1998]. How faults slip at depth can thus be derived indirectly through modeling of surface deformation.

[3] Theoretical surface displacements expected from some fault slip at depth is generally computed based on the theory of linear elasticity [e.g., *Savage*, 1983; *Okada*, 1985; *Cohen*, 1999]. This formulation is linear and easily inverted using standard algorithms. The distribution of fault slip is generally parameterized based on some discretization of the fault geometry. The cumulative fault slip needed to explain displacements that have occurred between two epochs for which geodetic data are available can then be obtained from some least squares inversion. Because the number of parameters generally exceeds the number of observations, regularization constraints are generally added; for example, the roughness of the slip distribution can be penalized or a positivity constraint can be added. One way to invert geodetic time series for time-dependent slip distribution thus consists in inverting the displacements measured between each two successive epochs. This method is computationally very intensive when the number of epochs is large, especially when nonlinear regularization criteria are used. Furthermore, this method considers each epoch individually, so measurement errors at different time steps are not properly balanced. In addition, the method also requires geodetic time series to be sampled at each site at the same epochs, limiting the possibility of analyzing a mixed data set which could include campaign data or interferometric synthetic aperture radar (InSAR) data.

[4] P. Segall and colleagues proposed a variation of the epoch-by-epoch inversion called the extended network inversion filter (ENIF) [*Segall and Matthews*, 1997; *McGuire and Segall*, 2003] specifically for GPS measurements. ENIF takes

---

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, USA.
[2]Now at Ashima Research, Pasadena, California, USA.

into account a stochastic description of local benchmark motion, a nonparametric description of slip rate as a function of time, estimation and removal of reference frame errors, and furthermore makes use of an extended Kalman filter to smooth out noise-related temporal variations. This approach has been applied with great success in a number of studies [e.g., *Miyazaki et al.*, 2004]. However, the method has some limitations. One is that it involves a number of hyperparameters necessary for the model, making it a cumbersome tool requiring somewhat subjective choices which define the space of possible solutions. Another is that the technique is computationally costly. The inversion of a fairly modest data set can take hours or days to complete on a desktop machine. For example, the 400 epoch 10 continuous GPS (cGPS) station data set used to infer the afterslip distribution following the Nias earthquake [*Hsu et al.*, 2006] took approximately 2 h to run on a typical 2005 laptop. ENIF as formulated by *Segall and Matthews* [1997] is also restricted to the use of GPS time series, though it could theoretically be extended to allow analysis of any single type of spatiotemporal data (e.g., InSAR data). Another limitation is that the method is not easily applicable to the analysis of complex time series that would include interseismic, coseismic and postseismic deformation.

[5] In this study we describe and then test a principal component analysis-based inversion method (PCAIM) designed to overcome some of the aforementioned limitations. This technique is easy to use and computationally fast, allowing the user to quickly explore a suite of slip models with different geometries and other characteristics on a desktop computer. Note that PCAIM only deals with the inversion for fault slip problem, and is therefore not a substitute for ENIF since PCAIM does not incorporate any tool to estimate and correct benchmark wobble or reference frame errors. The corresponding MATLAB scripts are available from the Tectonic Observatory Web site (http://www.tectonics.caltech.edu/).

[6] Hereafter, we first describe the PCAIM approach. Next, we test the approach by applying it to synthetic geodetic time series and to cGPS records of postseismic deformation following the M_w 8.6 Nias, Indonesia, earthquake, already analyzed by *Hsu et al.* [2006]. We also show how PCAIM performs when applied to an spontaneous aseismic transient, taking the example of the 1999 Cascadia slow slip event, already analyzed by *McGuire and Segall* [2003].

2. Principle and Implementation of PCAIM
2.1. Principle of PCAIM

[7] Let us consider a set of geodetic stations (Figure 1) recording surface displacement over a series of time epochs. Each station is assumed to have successfully recorded data at all epochs. This assumption is necessary to implement the simplest version of the method described below. In the case of missing points or of data with different temporal sampling, the technique can be adapted, as discussed later. We place time series with identical sampling epochs in a $m \times n$ matrix X_0 where each row corresponds to the time series from one component (east, north or vertical) at one station and each column corresponds to all data measured at a given epoch. We suppose that displacements are due to an unknown, time-dependent slip distribution on a fault with known geometry α which is discretized as set of rectangular subpatches. Slip on each subfault is decomposed into a strike and a dip component. We assume that the medium surrounding the fault is elastic, and we represent fault slip by a matrix L where each row refers to slip (strike-slip or dip-slip) on a given subpatch and each column to an epoch. Surface displacements then obey

$$X_0 = G_\alpha L,\qquad(1)$$

where G_α denotes the Green's functions relating surface displacements with fault slip at depth (decomposed into a strike-slip component and a dip-slip component), given a fault geometry α. G_α can be computed from the semi-analytical solutions of *Okada* [1985] for a dislocation embedded in an elastic homogeneous half-space, as is the case in this study. This formulation can be substituted with any other which would relate linearly fault slip at depth and surface displacements. For example, one could use Green's functions computed for a medium with heterogeneous elastic properties and for any geometry of the free surface.

[8] Determination of the time-dependent slip model corresponding to the measurements requires inversion of that linear system. The principal component analysis based inversion method described in this study relies on the following principles:

[9] 1. The data sets can be decomposed as the sum of components, each component being associated with a pattern of surface displacement and a time function.

[10] 2. Only a small number of components is generally necessary to explain most of the data.

[11] 3. The pattern of surface displacements associated with each component can be inverted for some principal slip distribution.

[12] 4. The fault slip distribution corresponding to the original data set can be derived by linear combination of the principal slip distributions.

[13] In practice, PCAIM flows as follows:

[14] 1. Center X_0 along its rows and call the centered matrix X (see section 2.2.1).

[15] 2. We decompose X into the product of three matrices $X = USV^t$ such that the columns of U and V are of unit length and S is diagonal (although S is nonsquare, it is diagonal in the sense of all nondiagonal entries are nonzero). Such a decomposition can be obtained through a standard Singular Value Decomposition of X, as assumed hereafter to facilitate the presentation (see section 2.2.2). However, any other decomposition can be substituted for singular value decomposition (SVD). In practice, it is advantageous to use the weighted low-rank approximation of *Srebro and Jaakkola* [2003] to allow easy handling of missing data and uncertainties as discussed in section 6.

[16] 3. Determine the number r of principal components necessary to fit the data within uncertainties. Filter X by defining

$$X_r = U_r S_r V_r^t \qquad(2)$$
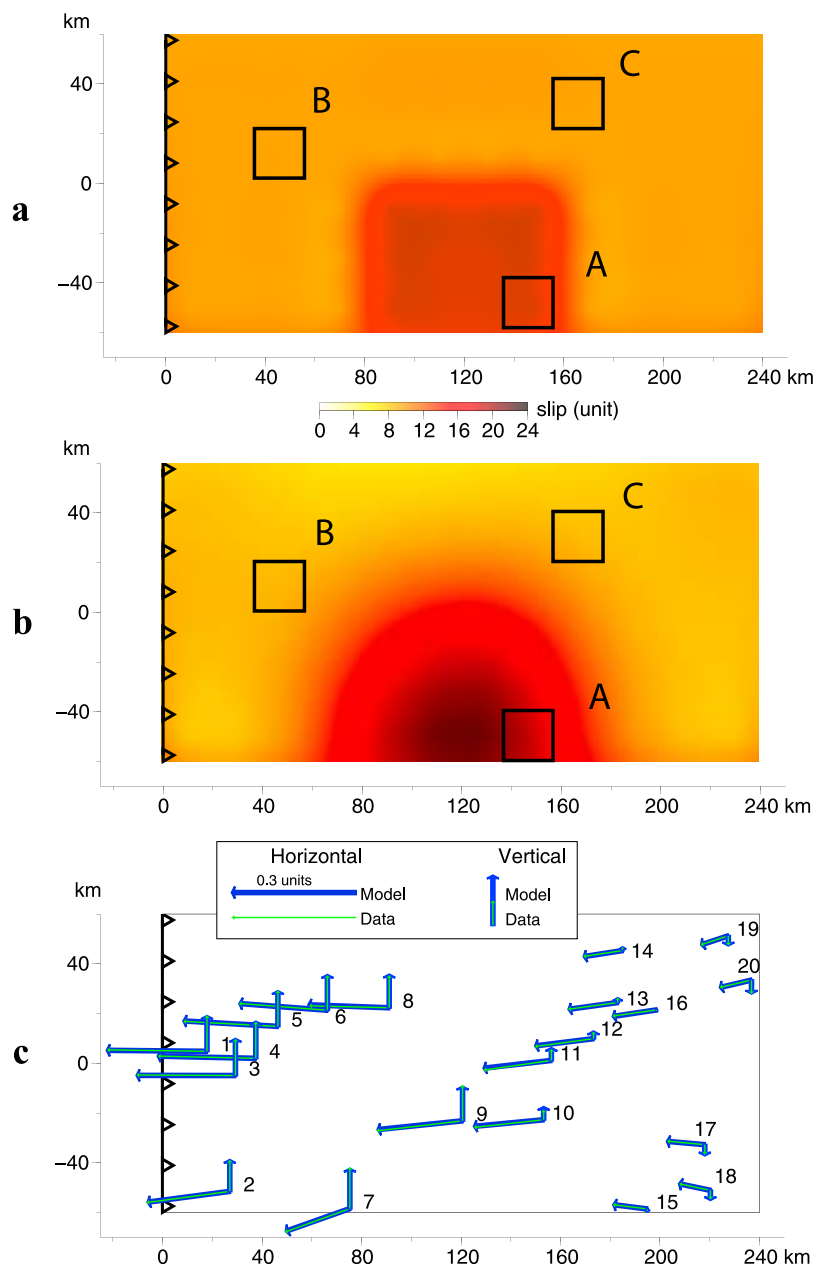
Figure 1. (a) Cumulative slip on fault corresponding to synthetic scenario used to test principal component analysis based inversion method (PCAIM) described in this study. Color scale shows slip amplitude. Scenario considers a 120 km × 240 km planar fault with a dip angle of 12°, mimicking a portion of a megathrust. (b) Cumulative slip on fault corresponding to slip model derived from PCAIM inversion of surface displacements at 20 geodetic stations. (c) Location of geodetic stations. Vectors show displacements corresponding to synthetic scenario (labeled "data") and computed from best fitting slip model derived from PCAIM inversion (labeled "model"). Fault slip and surface displacements are expressed in an arbitrary unit. Boxes show locations of patches A, B, C used to compare slip evolution in synthetic scenario and inverted slip model.

the truncated singular value decomposition of X (see section 2.2.3).

[17] 4. Invert the left singular vectors of X_r (columns of U_r) for slip distributions as if they were ordinary displacement vectors. We call these the *principal slip distributions* of X_r and name the matrix of principal slip distributions L_r (see section 2.2.4).

[18] 5. Sum the weighted principal slip distribution multiplied by their associated *time functions* over all components $((L_r S_r) V_t^t)$. Assuming the problem is fully or overdetermined, this results in a unique cumulative slip state at each time up to setting an arbitrary and time-independent "zero" for the slip (see section 2.2.5).

[19] It is useful to think of PCAIM based on the diagram in Figure 2. The left-hand track in Figure 2 represents
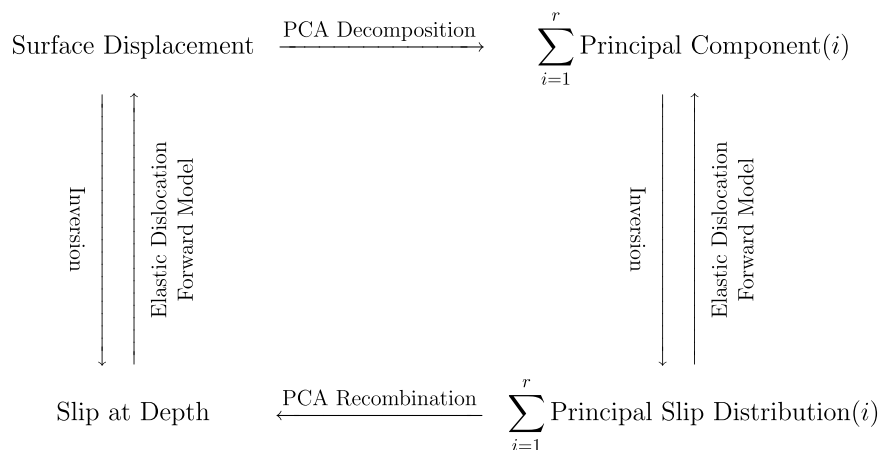
Figure 2. Diagram showing principle of PCAIM described in this study. Conventional approaches follow left-hand track, directly translating displacement data into a slip model through some inversion procedure based on theory of dislocations in an elastic half-space [e.g., *Okada*, 1985]. PCAIM decomposes displacement data into sum of so-called principal components. Each of components is individually modeled and translated into a corresponding principal slip distribution model. Note that the slip model associated with any one particular component does not have any particular physical significance. PCAIM takes advantage of linearity of formulation that converts fault slip into surface displacement (based on theory of linear elasticity). It makes it possible to retrieve the physical slip model by linear combination of principal slip models derived from inversion of each component as described in text.

directly translating displacement data into a slip model by inverting the difference in surface displacement between consecutive epochs for incremental fault slip. PCAIM instead decomposes the displacement data into the sum of principal components. Each of the components can be inverted individually into a corresponding slip distribution. It should be noticed that each individual component corresponds to a linear combination of the contributions from various sources and not to a particular, identifiable physical source. In general, each component has no obvious physical meaning considered alone, although the various components might be recombined to extract the contribution of particular sources [*Kawamura and Yamaoka*, 2006, 2009]. One common exception is when nearly all of the signal can be explained by the first component, meaning the slip distribution is spatially stationary and only the amplitude of the signal varies over time. In that case, the signal is essentially the result of one single process that is represented by the first component. If we use all components in PCAIM instead of only the most statistically significant components, PCAIM results in exactly the same model as that produced via consecutive epoch-by-epoch inversions (the proof is given in Appendix A).

2.2. Justification and Implementation of PCAIM
2.2.1. Preparation of X
[20] We center the original data matrix X_0 by defining

$$X(i,j) = X_0(i,j) - \frac{\sum_{k=1}^{m} X_0(i,k)}{m}. \tag{3}$$

Note that i corresponds to the spatial index and j corresponds to the temporal index.
2.2.2. Principal Component Analysis of X
[21] We proceed with a principal component analysis (PCA) of X following a standard linear algebra approach

(see any standard textbook on linear algebra for details). We find the eigendecomposition of the spatial and temporal covariance matrices, $XX^t = US^2U^t$ and $X^tX = VS^2V^t$, respectively. That is to say, the column vectors of U are the eigenvectors of the spatial covariance matrix XX^t, the column vectors of V are the eigenvectors of the temporal covariance matrix X^tX, and S is a diagonal matrix with diagonal elements equal to the positive square root of the eigenvalues of X^tX and XX^t.
[22] This decomposition leads us to write

$$X_{m \times n} = U_{m \times m} S_{m \times n} V_{n \times n}^t. \tag{4}$$

In this decomposition, the ith component is associated with a pattern of surface displacement (represented by the ith column of U) and a time function (represented by the ith column of V). Components are ordered according to the amount of variance explained in the initial geodetic data set with the first component being the one with the most variance explained, i.e., corresponding to the largest eigenvalue. We assume for the moment that the singular values of X are distinct so that the singular decomposition is unique, which is almost always the case in practice. In fact, PCAIM still works when there are nondistinct singular values, but due to the resulting complications we omit the discussion for the sake of brevity. PCA is one of the most intuitive ways to decompose the data when we are trying to maximize the amount of variance explained in the first i components in an unbiased manner, but the user may deem it useful or necessary to define U, S and V differently for different circumstances.

2.2.3. Determine the Number of Useful Components
[23] PCA leaves uncorrelated noise and deformation localized to small groups of stations in the higher-order components. Thus, PCA decomposition is a method to filter data by reconstructing the time series with only the lower-
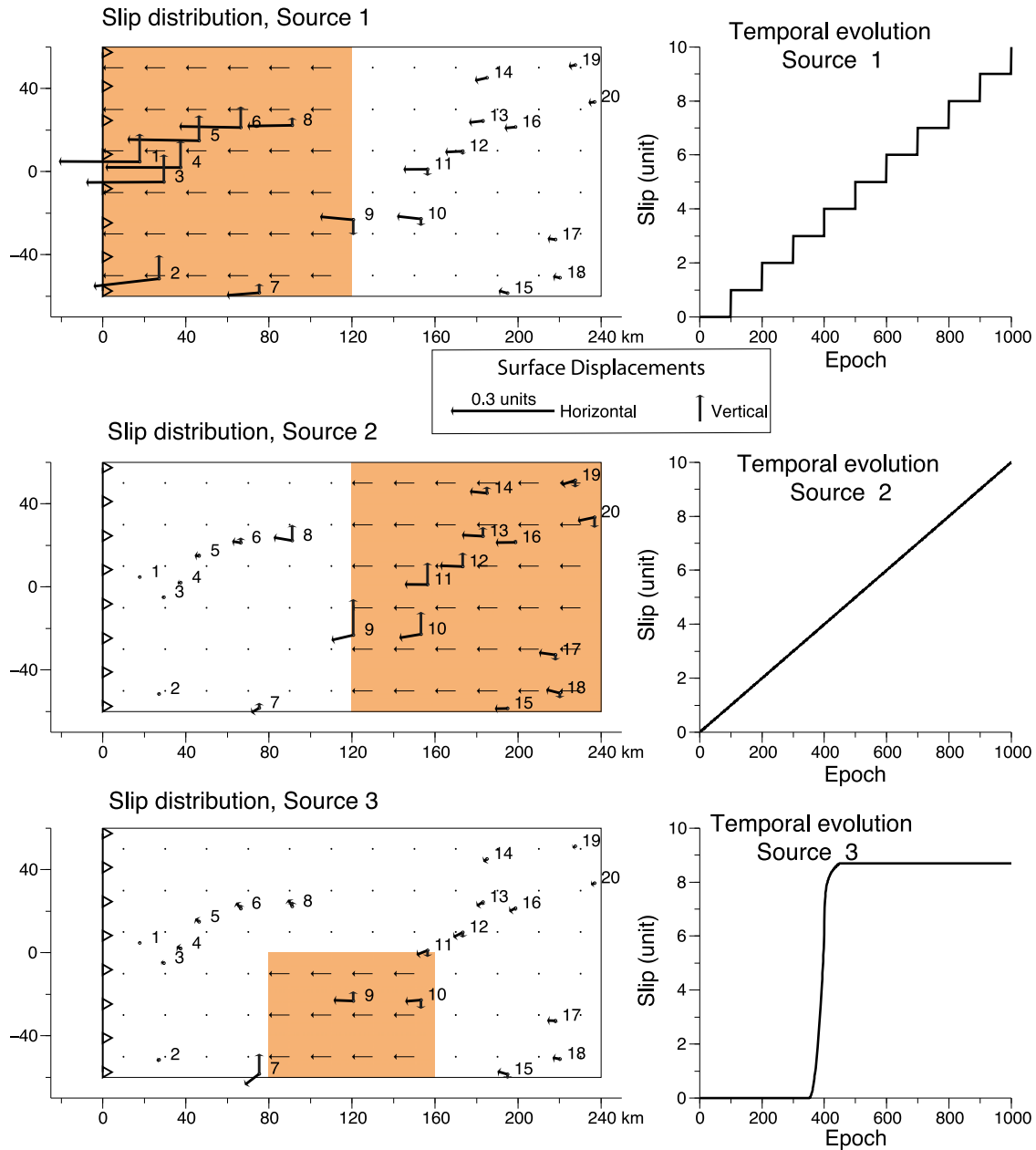
Figure 3. Graphical representation of three sources that are added to constitute synthetic fault slip model. Source 1 corresponds to updip half of fault and mimics a stick-slip behavior. It produces a 1 unit slip event every 100 epochs. Source 2 corresponds to creep at constant rate, arbitrarily set to 10 units over 1000 epochs, on downdip portion of fault. Source 3 is a transient event with a gradual accumulation of 9 units of slip over 50 epochs. Rake is assumed constant to 90°, corresponding to purely dip-slip motion.

order components [e.g., *Savage*, 1988; *Savage and Svarc*, 1997; *Aoki and Scholz*, 2003; *Kawamura and Yamaoka*, 2006, 2009; *Fukuda et al.*, 2008]. That is, we approximate the original data X, by selecting the first r components (also called modes by certain authors [e.g., *Kawamura and Yamaoka*, 2006, 2009]),

$$X \approx X_r = U_r S_r V_r^t, \qquad (5)$$

where U_r is the first r columns of U, S_r is the leading principal $r \times r$ submatrix of S, and V_r is the first r columns of V.

[24] The number of components can be chosen based on the reduced chi-square statistics. This allows us to determine the number of components such that the misfits between the filtered and the original data are, on average, of the order of magnitude of the measurement uncertainties. In practice, this implies selecting the number of components so that the reduced chi squared statistic, χ_{red}^2, of $X - X_r$ is approximately equal to one, i.e.,

$$\chi_{red}^2 = \frac{1}{N - r(n + m + 1)} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(X(i,j) - X_r(i,j))^2}{\sigma(i,j)^2} \approx 1, \qquad (6)$$
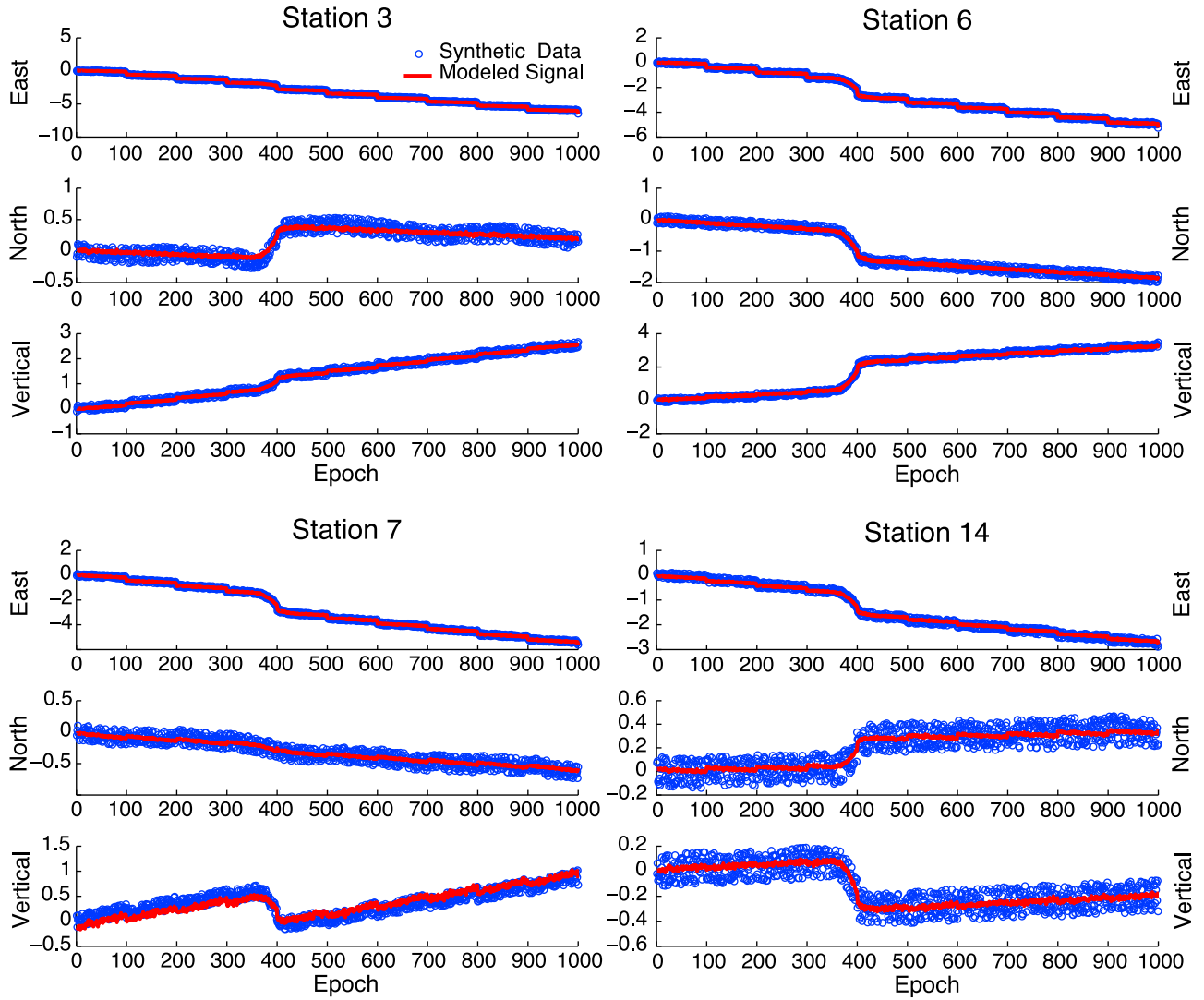
Figure 4. Geodetic time series computed from synthetic fault model of Figure 3 with addition of a randomly generated noise with uniform distribution in range $[-0:15; 0:15]$ (blue circles), at stations 3, 6, 7, and 14. See Figure 1 for station locations. Series are constituted of 1000 samples, 1 per epoch. Red lines show theoretical displacement curves computed from slip models derived from PCAIM inversion of synthetic data using three components.

where $\sigma(i, j)^2$ is the 1-sigma uncertainty on the i,jth data point.

[25] The appropriate number of components to be selected can also be chosen based on the F test [e.g., *Press et al.*, 2002] which quantifies whether the additional variables introduced with the $r + 1$th component is worth the variance it explains. The criterion is based on the ratio of the difference between the chi-square obtained before and after adding the new component, divided by the chi-square without the additional component and multiplied by a ratio of degrees of freedom, i.e.,

$$F = \frac{\chi_k^2 - \chi_{k+1}^2}{\chi_{k+1}^2} \frac{N - p_{k+1}}{p_{k+1}p_k}, \qquad (7)$$

where χ_k^2 is the χ^2 of the model with k components, N is the number of data, and p_k is the number of parameters in the model with k components. This criterion has the advantage that it is independent of any renormalization of the uncertainties.

2.2.4. Determination of Principal Slip Distributions

[26] Each left-singular vector represents a spatial pattern of displacements at the surface. We show in Appendix A that, provided that it is valid to model the epoch-by-epoch displacements as resulting from fault slip at depth, any principal component found using PCA can be modeled as surface displacements associated with some slip distribution at depth as well.

[27] At this point we compute the principal slip distribution (l_i) of the ith component of X_r, u_i (i.e., the ith column of U_r) as a solution to

$$G_\alpha l_i = u_i. \qquad (8)$$
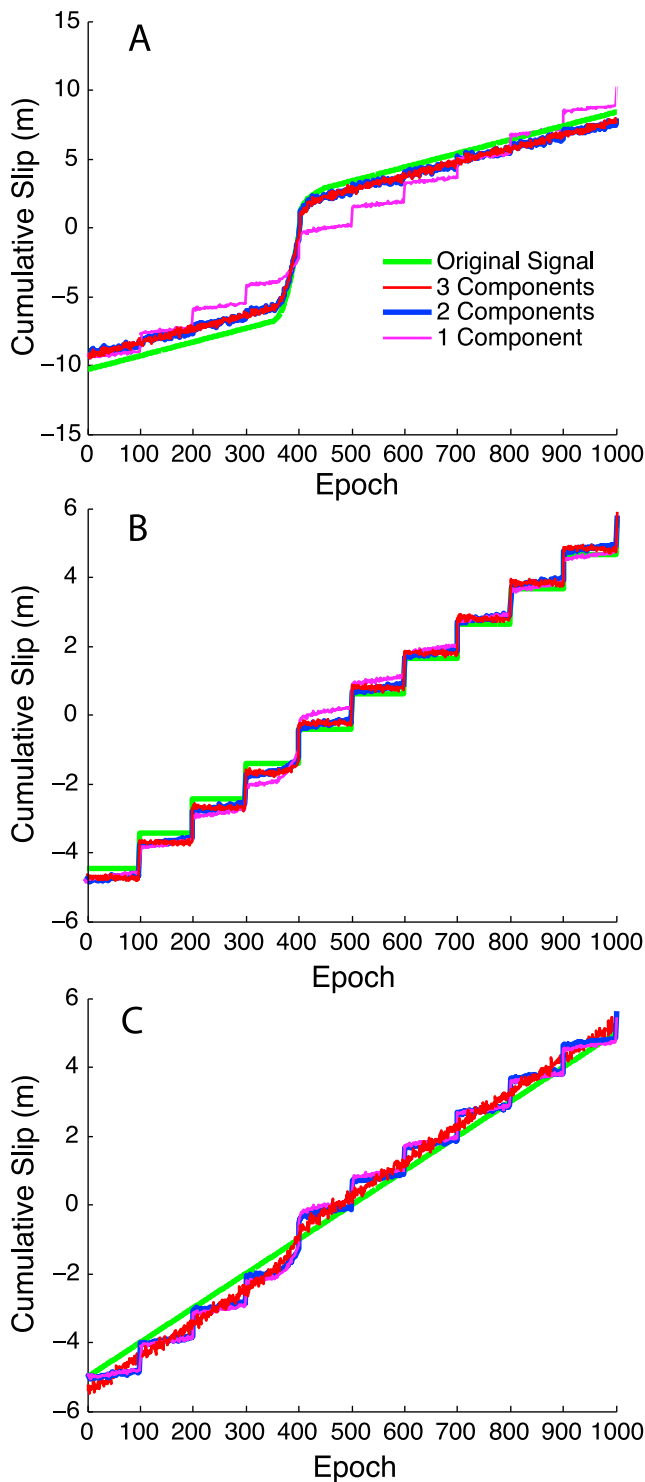
Figure 5. Imposed slip history on patches A, B, and C (green curves). See Figure 1 for location of patches. Patch A corresponds to a region affected by both constant creep behavior and transient event. Patch B is within domain with only stick-slip motion. Patch C is within domain with only steady state creep. Also shown is slip history retrieved from principal component analysis inversion method (PCAIM) applied to synthetic time series generated at 20 stations shown in Figure 4, using one, two, and three principal components.

These slip distributions by themselves mean that given the data in X, l_i represents the ith most important slip distribution at depth in terms of explaining the variance of X that have orthogonal surface displacements at the measurement stations.

[28] In practice, the inversion of equation (8) is often an ill-posed problem. It is indeed fairly usual to invert displacements at approximately a dozen stations (in which case $m = 36$ for GPS data) on a fault surface consisting of hundreds of subpatches, each of which can accommodate slip in any combination of two independent directions. Some constraints are then needed to regularize the inversion. We choose a common approach which consists in seeking "smooth" slip solutions, but any other linear criteria could be used at the user's discretion.

[29] For each inversion, we minimize a cost function that is a linear combination of the weighted squared misfits between the u_i and the model predictions $G_\alpha \cdot l_i$ and the Laplacian of the principal slip distribution,

$$Cost\ Function = \sum_{k=1}^{3m} ([G_\alpha \cdot l_i](k) - u_i(k))^2 + \frac{1}{\gamma^2} ||\Delta l_i||^2, \quad (9)$$

where $[G_\alpha \cdot l_i](k)$ is the model prediction corresponding to the kth data point of the ith principal component of data and $1/\gamma^2$ allows us to set the weight of the smoothness of the slip distribution, and $\Delta = (\partial^2/\partial^2 x) + (\partial^2/\partial^2 y)$. With this smoothness constraint, the solution to the least squares problem is always unique though the solution depends on the chosen value of γ. It is worth noting that this is only one of many possible regularization procedures to make the original ill-posed problem solvable.

2.2.5. Determination of Final Slip Model

[30] Because we assume linear elasticity, the relationship between fault slip at depth and surface displacements is linear. It is thus possible to combine the slip distributions into a model of the spatiotemporal slip evolution on the fault using time functions defined by the PCA of X. Now that we have inverted the left singular vectors we can proceed in the space of slip distributions. $l_i \cdot s_i$, $i = 1, \ldots, r$ define properly scaled slip distributions. In matrix form, we have inverted the columns of U_r from $U_r S_r V_r^t$ for L_r and multiplied by S_r.

[31] To find the time evolution of the slip distribution, we need to distribute the time functions that make up V_r. This effectively states that at epoch j, the cumulative contribution to the total slip from the kth component is $l_k \cdot s_k \times V^t(k, j)$.

2.3. Computational Complexity

[32] As demonstrated in Appendix A, in the limit of $r \rightarrow rank(X)$, PCAIM generates the same results as if we had performed an inversion of the displacement between epochs $i \leq j$. At most $rank(X) \leq \min(m, n - 1)$ inversion(s) must be performed to find the time-dependent slip pattern equal to that derived from epoch-by-epoch difference inversions. Furthermore, for any $r \leq rank(X) \leq n - 1$, performing exactly r inversions allows approximately $\sum_{i=1}^{r} \frac{\sigma_i^2}{var}$ of the variance to be explained, where var is the total variance of the data set ($var = \sum_{i=1}^{rank(X)} \sigma_i^2$) and σ_i is the ith singular value. We have only performed r inversions instead of $n - 1$ as is necessary for an epoch-by-epoch inversion algorithm. Since some kind of time filtering is often imposed
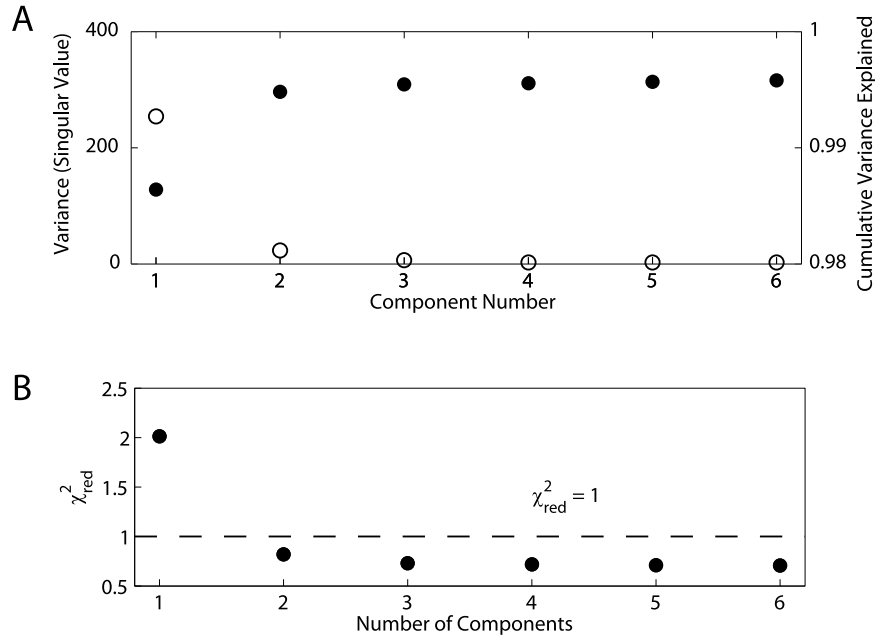
Figure 6. (a) Open circles show total variance (given by singular value) explained by each single principal component. Solid circles show cumulative fractional variance when components are successively added to reconstruct signal, starting from first component. (b) Misfit between measured and reconstructed displacements as number of components used to reconstruct signal increases. Misfit is quantified from reduced chi-square as defined in text (equation (6)). In that case it is necessary to take into account the first two components for filtered data to fit original data within uncertainties on average. On the basis of F test, the first three components are actually found significant.

[e.g., *Fukuda et al.*, 2008], the computation time is reduced from

$$O(\mathrm{SVD}(X)) + (n-1)O\big(K\big(x_i - x_j\big)\big) \qquad (10)$$

to

$$O(\mathrm{SVD}(X)) + rO\big(K\big(x_i - x_j\big)\big), \qquad (11)$$

where K is the single inversion function. In practice, we have found $r \leq 3$ often holds in examples when $n \approx 300$. PCAIM then allows a 1×10^{-2} reduction in the number of inversions without significantly changing the result of the inversion.

3. Testing PCAIM With Synthetic Data

[33] To evaluate how well this methodology performs, we put together a synthetic scenario with a relatively complex arbitrary fault slip history. We prescribe the fault geometry, impose a spatiotemporally varying slip distribution, and then generate a synthetic displacement field. We add noise and use these data as input to the PCAIM and compare the inferred spatiotemporally varying slip model to the original imposed slip model.

[34] Our synthetic example is inspired by the setting of the Sunda megathrust which is monitored by the SUGAR cGPS experiment (http://www.tectonics.caltech.edu/sugarkml.html). Paleogeodetic records have also been retrieved in this area. These are records of vertical displacements retrieved from the study of coral growth [e.g., *Sieh et al.*, 1999; *Natawidjaja et al.*, 2004] which typically cover from several decades to a few centuries.

[35] For our synthetic test we consider a network of 20 sites overlaying a ~15° dipping fault plane (Figure 1). In practice, the sites might correspond to sites where paleo-geodetic data would be available or to the location of GPS SuGaR stations. In the first case, the typical sampling rate would be annual and records could cover several centuries. In the case of actual GPS data, the sampling rate would be typically between 1 s and 1 day, and the record could extend up to a few decades.

3.1. Data Generation

[36] We impose three sources of slip on the fault (Figure 3). One source, that extends from the trench to 31 km depth, mimics simple stick-slip behavior (Source 1) with a slip event every 100 epochs. The deeper portion of the fault creeps at a constant rate (Source 2). In addition, we add one slip transient that evolves over time (Source 3). The imposed slip history is obtained by summing all three terms. We calculate the resulting surface displacement field using *Okada*'s [1985] formulation and sample this field at 1000 epochs at the 20 locations (Figures 3 and 4). Because of the linearity of the equations, fault slip and displacements at the surface are expressed in the same arbitrary length unit. We add randomly generated, evenly distributed noise from the interval $[-0.15, 0.15]$ unit to each data point. The standard
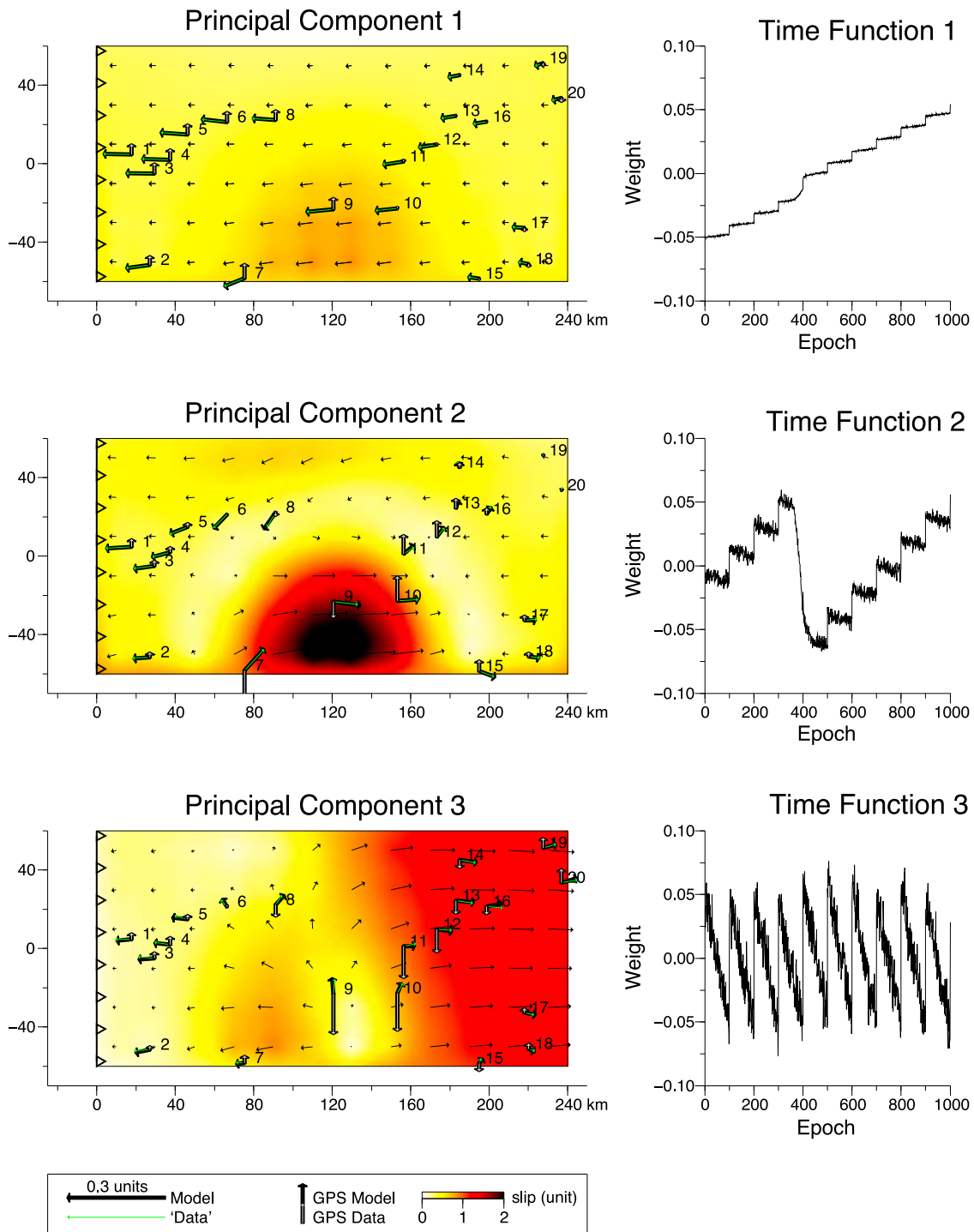
Figure 7. (left) Slip model and (right) source time function associated with each of first three principal components. Green arrows show surface displacements at various stations corresponding to each principal component (i.e., corresponding column of U obtained from SVD decomposition of data matrix). Thin black vectors show slip vector on fault obtained from inversion of these surface displacements, and thick vectors show reconstructed displacements at each stations.

deviation of the added noise is 0.087 unit and is considered to represent the 1-sigma uncertainty on the data.

[37] This results in 60 time series of 1000 samples each (Figure 4). Figure 5 shows the slip history imposed at the location of the three patches, A, B and C, the locations of which are reported in Figure 1.

3.2. Results of the Synthetic Test

[38] Figure 6 shows the contribution to the data variance of the various components derived from the PCA analysis of the 60 time series. The first component dominates clearly and explains about 99% of the data variance, and the reduced
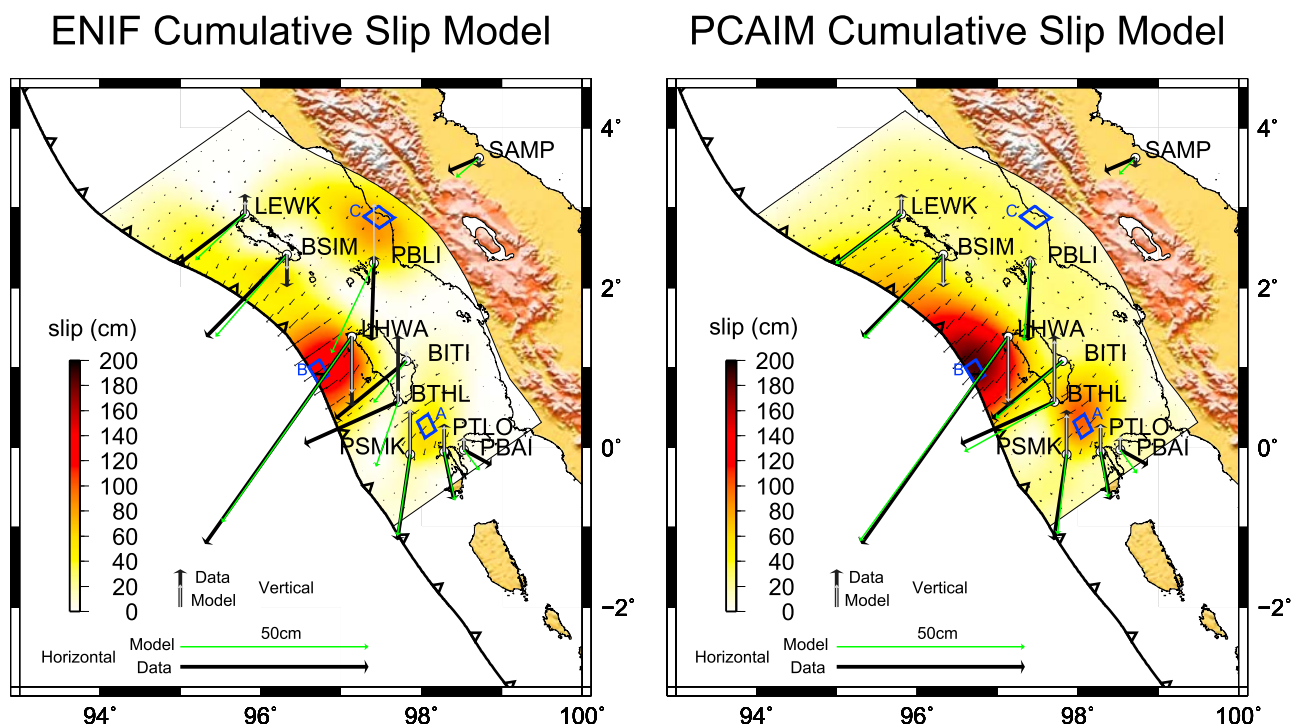
Figure 8. Location of cGPS stations that recorded postseismic deformation following 2005 M_w 8.6 Nias earthquake and comparison of cumulative slip distributions 334 days after 2005 Nias earthquake obtained from (left) ENIF [*Hsu et al.*, 2006] and (right) PCAIM inversions. Grid shows fault subdivision in 20 km × 16 km patches. Fault dips to northeast. Dip angle increases from 10° near trench to 30° near bottom edge of fault [*Hsu et al.*, 2006]. Blue boxes show locations of patches A, B, and C used to compare slip evolution derived from two models (Figure 12).

χ^2 value falls below 1 with two components, suggesting that two components are sufficient to fit the data within uncertainties. However, a quick look at the synthetic time series and postfit residuals reveal that the misfits are patterned and do not represent white noise if only the first two components are considered. This is consistent with the temporal function associated with the third component which shows a temporal pattern which suggests that this component does not only account for white noise. The appropriate number of components to be selected is more suitably chosen based on the *F* test. In this case, the *F* test determines we should use three components. More precisely, adding the third component improves the fit better than if it were random noise at the probability level 1-$p_{1,2}$ with $p_{1,2}$ = 0 (to MATLAB's computational accuracy). Similarly, we obtain $p_{2,3} = 2.6 \times 10^{-23}$, and $p_{3,4} = 1$ (because of the smoothness constraint on the slip distribution, the addition of a fourth component actually increases the chi-square statistics. The first three components are sufficient to explain about 99.5% of the data variance. Higher-order components account mostly for the noise added to the synthetic data. Figure 7 shows the slip pattern and the time functions derived from the inversion of components 1, 2, and 3. We compare the true (prescribed) cumulative slip to the inverted cumulative slip over the 1000 unit time period in Figure 5. We see that the inverted slip models associated with each of these three components indeed sum to fit the imposed slip history quite well. This confirms that very little signal is left in the higher-order components. So this test demonstrates that PCAIM can faithfully reproduce complicated spatiotemporal signals, without the a priori assumptions of a stationary

displacement field or characteristics of the time evolution of slip at depth.

[39] Also, this example illustrates clearly that the three principal slip distributions do not coincide with the elementary sources used to construct the synthetic signal. In cases with real data, we similarly do not expect a single principal slip distribution to coincide with a single physical phenomenon.

4. Testing PCAIM on a Postseismic Scenario

[40] In this section we use real data and compare how PCAIM results compare to ENIF results on a postseismic relaxation scenario. The data set consists of GPS records of postseismic relaxation following the March 2005 M_w 8.6 Nias earthquake (Figure 8). A previous analysis of this data set has shown that the recorded surface deformation reflects afterslip on the megathrust in the area surrounding the rupture area of the main shock [*Hsu et al.*, 2006]. We use the same 10 cGPS time series as in the ENIF model (Figure 9), and the same fault geometry. Six of the stations have fairly complete time series. Three stations (BITI, BTHL, and PBLI) were installed 160 to 240 days after the Nias earthquake and thus did not record the early period of postseismic relaxation. One of the stations (PLTO) did not function in the period between 10 and 50 days after the main shock. Some of the time series are shown in Figure 10 (the other time series are shown in the auxiliary material).[1]

---

[1]Auxiliary materials are available in the HTML. doi:10.1029/2009JB006535.
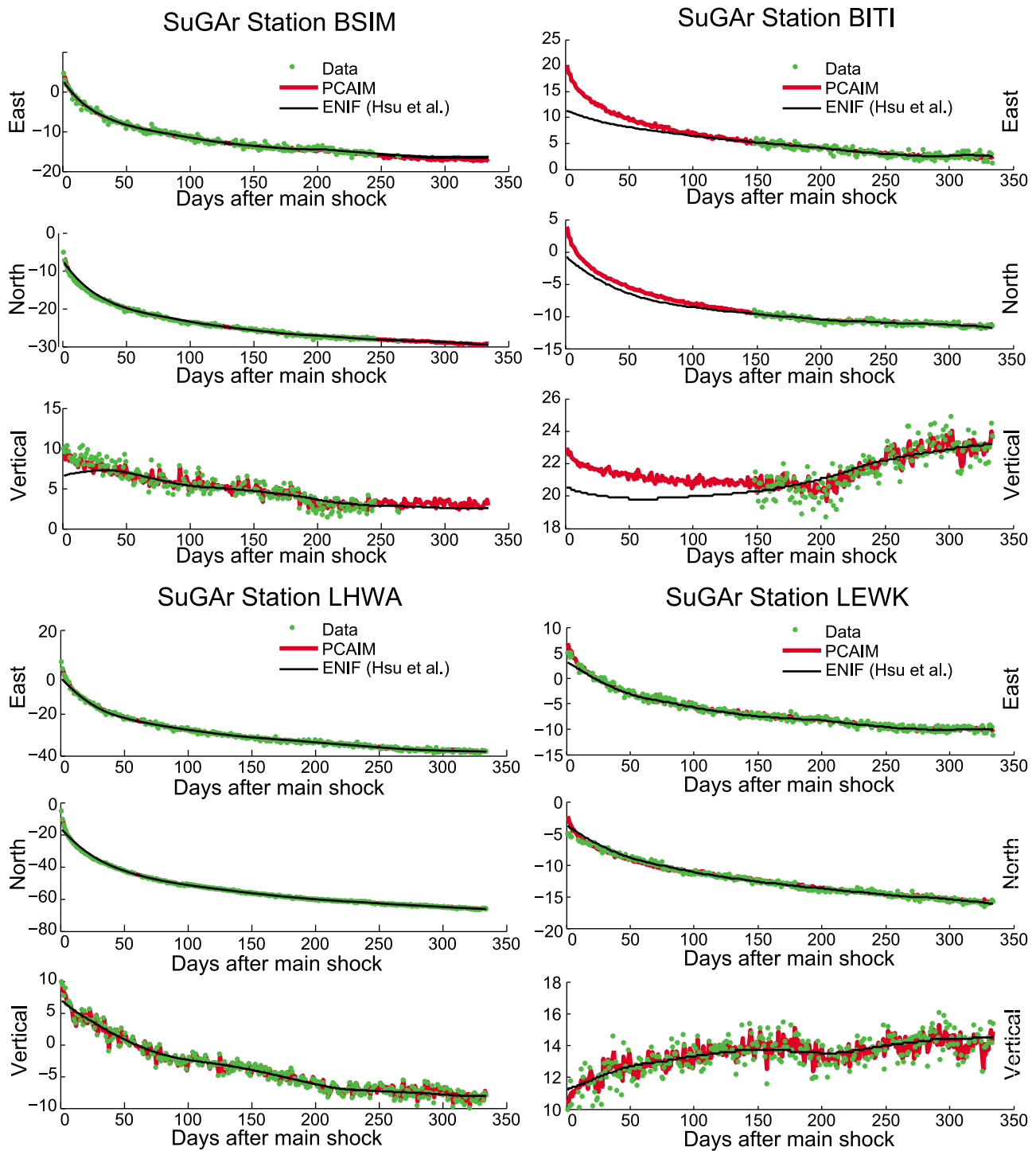
Figure 9. Observed and modeled geodetic times series of postseismic displacement following 2005 M_w 8.6 Nias earthquake at stations BSIM, BITI, LHWA, and LEWK. Original data points are shown in green. Red lines show displacements predicted from PCAIM slip model obtained in this study, and black lines show displacements predicted from ENIF slip model of *Hsu et al.* [2006]. Two models differ significantly only where there are missing points or in early postseismic period where ENIF slip model poorly tracks rapidly varying velocities due to temporal filtering. Times series and modeled predictions at all stations located in Figure 8 are shown in the auxiliary material.

4.1. Data Preprocessing

[41] As implemented for this study, PCAIM requires that the data matrix has no missing entries. Some procedure is thus needed to estimate the missing data point. Here we interpolate the data using a relatively flexible law the "log-exponential" function [*Perfettini and Avouac*, 2004] which describes the theoretical displacement that should be observed
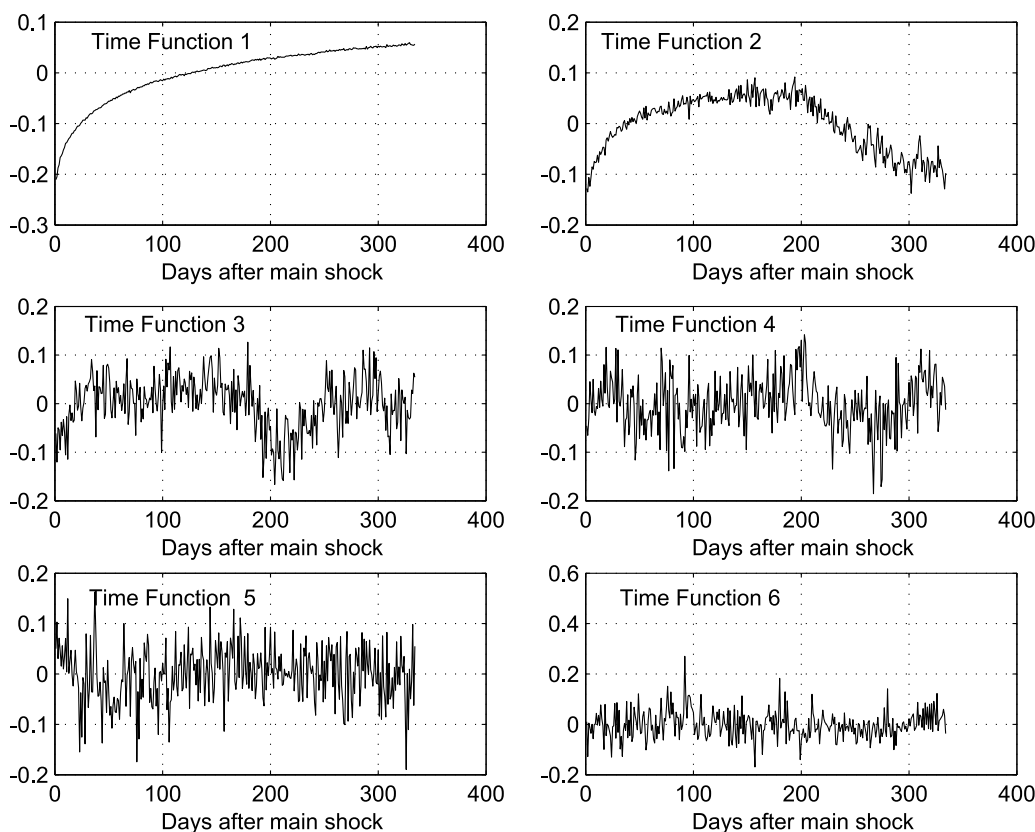
Figure 10. Time functions of first six principal components. Note that higher-order components (>3) are more erratic than lower-order components.

in case of frictional afterslip obeying rate-strengthening friction. Another approach, computationally more costly, would be to first analyze only the epochs with measurements recorded at all stations and then use this model to estimate the missing data points. Several iterative procedures can then be envisioned to estimate the missing values. In either case, we end up with a full data matrix. A more general approach is outlined in section 6. It relies on a decomposition in which the data are weighted according to their uncertainties. In this case, missing data are replaced with arbitrary values with very large uncertainties (several orders of magnitude larger than the typical uncertainties on the measurements).

4.2. Results of Postseismic Test

[42] PCA is then run on the filled-in data matrix. The time functions associated with the first six components are shown in Figure 10. Qualitatively, it shows that the coherent signal within the network is probably accounted for by only the first three or four components.

[43] Next, we choose the number of components to be used in the inversion. Figure 11 shows the contribution to the data variance of the various components derived from the PCA analysis of the 30 time series. The first component is clearly dominant and explains about 97% of the data variance. The reduced chi-square criteria suggest that many more components are needed for the filtered data to fit the original data within uncertainties (Figure 11). It is highly probable that in fact the formal uncertainties assigned to the data are underestimated. This is common since formal uncertainties do not account for a number of sources of colored

noise [e.g., *Zhang et al.*, 1997]. In this case, the *F* test is particularly appropriate because it quantifies the improvement in the fit obtained by adding a new component based on the ratio of the chi-square obtained before and after adding the new component. This criterion is thus independent of any renormalization of the uncertainties. The *F* test yields that four components are significant in that case. More precisely, adding the second component improves the fit better than if it were random noise at the probability level 1-$p_{1,2}$ with $p_{1,2}$ = 1.06×10^{-8}. Similarly we obtain $p_{2,3} = 9.31 \times 10^{-4}$, $p_{3,4} = 1.33 \times 10^{-6}$ and $p_{4,5} = 0.693$. To facilitate comparison with the ENIF solution of *Hsu et al.* [2006], we used three components since this is the number of components that approaches the same level of explanation of variance of the data as the ENIF model of *Hsu et al.* [2006].

[44] Figure 12 shows the slip patterns derived from the inversion of the first three components. These principal slip models were obtained with no constraints on the rake aside from that implicit in penalizing the Laplacian of the slip distribution. When we add the inverted slip history associated with each of these three components, PCAIM yields a cumulative slip model following the earthquake similar to the ENIF cumulative slip model (Figure 8). The slip histories inferred from the PCAIM and ENIF models at the location of patches A, B and C (see Figure 8 for location) are compared in Figure 13. The PCAIM and ENIF models both predict theoretical displacements in good agreement with the observations (Figure 9). Their respective predictions might, however, differ very significantly where data are missing (for example, at station BITI in Figure 9). These
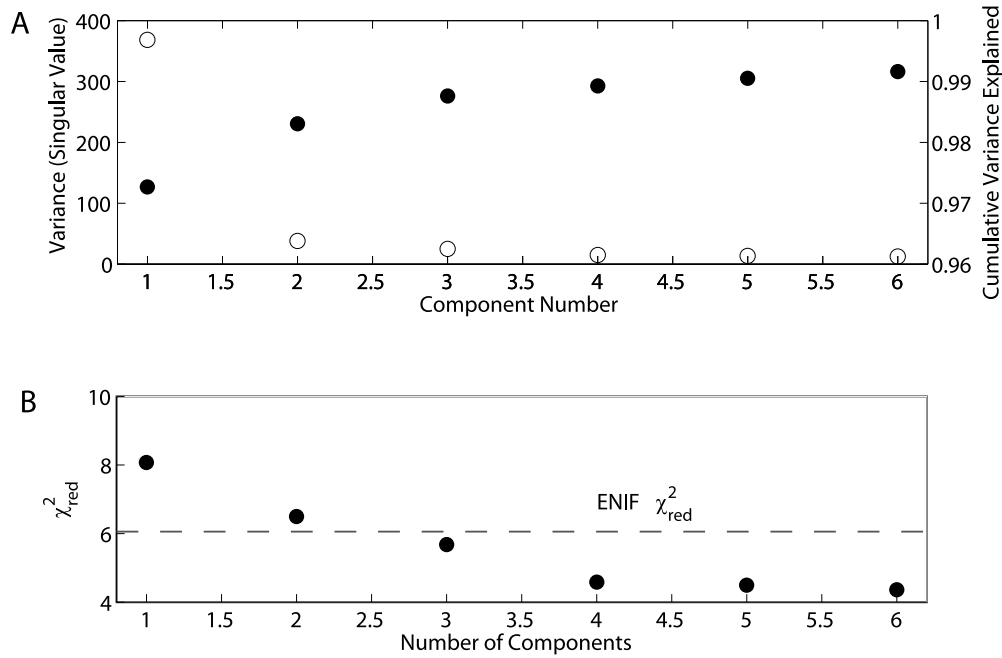
Figure 11. (a) Open circles show total variance (given by singular value) explained by each single principal component. Solid circles show cumulative fractional variance when components are successively added to reconstruct signal, starting from first component. (b) Misfit between measured and reconstructed displacements as number of components used to reconstruct signal increases. Misfit between measured and filtered displacements is quantified from reduced chi-square as defined in text (equation (6)). In that case, it is necessary, to take into account first three components for filtered data to fit original data as well as reference model of *Hsu et al.* [2006] which yielded a reduced chi-square of 6. Originally, assigned uncertainties are probably underestimated. On the basis of *F* test, the first four components are actually found significant as detailed in text.

differences do not impede the quality of the fit to the original data. Thus, the two models might be considered as equally possible slip histories.

[45] One major difference between the two slip models is observed in the very early days of the postseismic relaxation process. Close inspection of Figure 9 shows that the ENIF slip model does not manage to track well the rapid decrease in geodetic velocities during the first 10–20 days (see stations LEWK and LHWA). In that regard, PCAIM does a better job modeling the data as the fit to these early epochs is improved. It thus makes sense the cumulative ENIF and PCAIM slip models over the whole analyzed period show
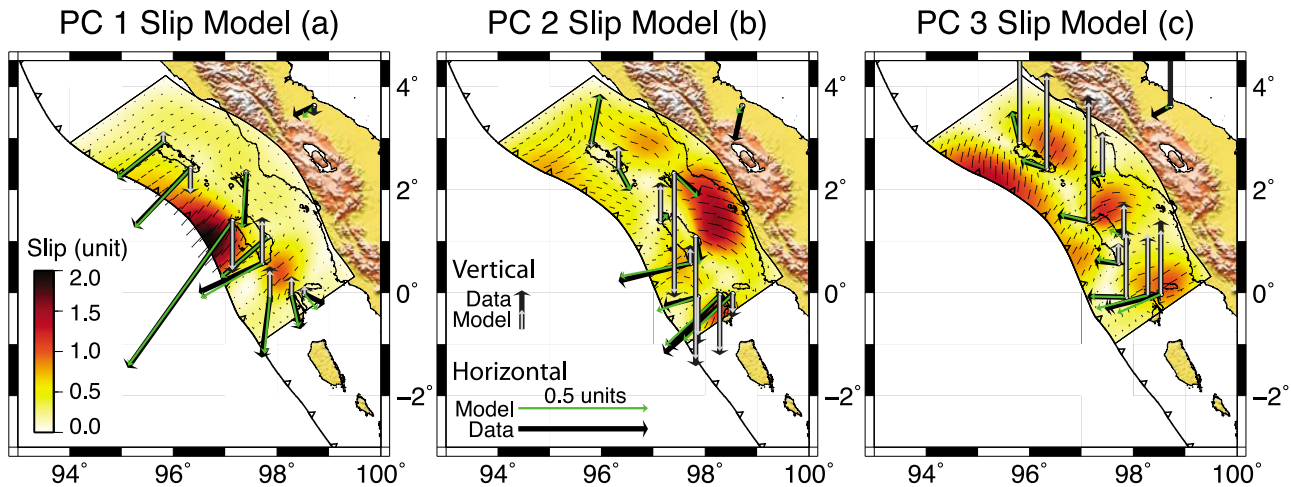
Figure 12. Slip distribution and associated pattern of surface displacements associated with each of first three components. Notice that complexity of slip distribution increases as we move to higher components. This results from the attempt to fit more isolated and disparate events or noise as we move into higher components.
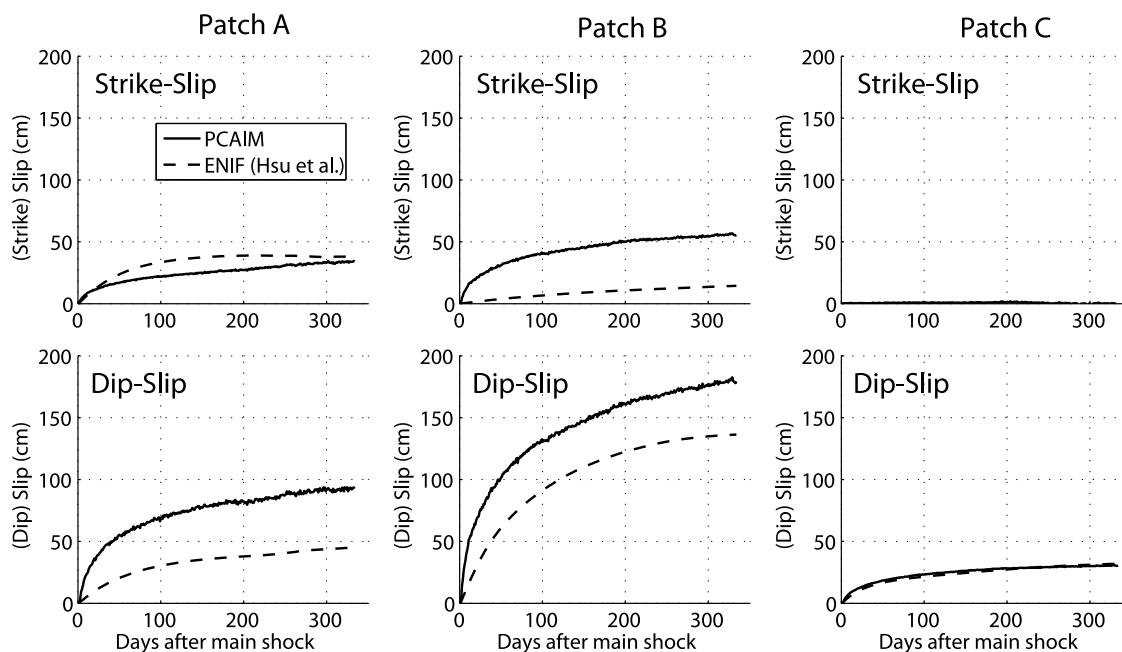
Figure 13. Comparison of slip evolution as determined from ENIF inversion (dashed line) and PCAIM inversion (continuous line) at three different fault patches A, B, and C. See Figure 8 for the locations of the patches.

similar patterns but have quite different amplitudes (Figure 8). The slip potency (integral of cumulative slip over area) is 5.7×10^{10} m^3 for the PCAIM slip model and 3.9×10^{10} m^3 for the ENIF slip model. The PCAIM slip model over 400 days following the main shock amounts to as much as 28% of the 2.0×10^{11} m^3 coseismic slip potency [*Briggs et al.*, 2006], versus 20% for the ENIF slip model. Another reason for some of the difference between the two slip models is that the ENIF model was obtained with some constraints on rake, positivity of slip and moment minimization, whereas no such constraints were imposed on the PCAIM model. This "positivity" constraint on slip is practically unnecessary given the constraints coming from the data themselves and the penalty on spatial roughness. Finally the PCAIM slip model does not involve any externally imposed smoothing or predetermined functional form of the time evolution of slip while the ENIF model has imposed temporal smoothness due to Kalman filtering.

5. Testing PCAIM on a Slow Slip Event

[46] In this section we evaluate how PCAIM performs compared to ENIF in the case of a transient aseismic event (also called a "slow earthquake"). We focused on the 1999 slow event on the Cascadia subduction zone [*Dragert et al.*, 2001] for which a relatively detailed source model has been obtained with ENIF [*McGuire and Segall*, 2003] (Figures 14, 15, and 16). This scenario is quite different from the post-seismic scenario analyzed in section 4 in that the signal-to-noise ratio is much smaller and the source is more complicated, characterized by a nonstationary spatial distribution of slip. The data set analyzed by *McGuire and Segall* [2003] consists of continuous GPS records at 12 stations of the Pacific Northwest Geodetic Array [*Miller et al.*, 2001] over 115 days (Figure 14).

[47] We used the same 3-D fault geometry as *McGuire and Segall* [2003], originally from *Fluck et al.* [1997], which consists of 72 approximately 25 km × 15 km subfaults. We have run inversions on the original data set and obtained models which were showing differences with the ENIF model of *McGuire and Segall* [2003]. Those were difficult to interpret because we were not able to reproduce the preprocessing of *McGuire and Segall* [2003], who removed interseismic secular velocities from the original time series, and also because PCAIM, unlike ENIF, does not account for framework errors and benchmark wobble. To facilitate the comparison we show here a model obtained by using, as an input in the inversion, the theoretical displacements predicted from the ENIF model. We found that the PCAIM source model matches the input data to better than 0.15 mm, when four components are used (the standard deviation of the residuals is 0.352 mm for one component, 0.151 mm for two components, and 0.138 mm for three components). This means that, if three or more components are taken into account, the PCAIM model and the ENIF models predict displacements which differ insignificantly in view of the 1–2 mm spread in the filtered data, as the graphical comparison of Figure 14 shows. One corollary of that finding is that for PCAIM to work on real data, the noise should be corrected for within components of order 4 and lower. This can only be achieved with efficient filtering of any spatially correlated noise in the data (framework errors and tropospheric effects in particular).

[48] Figure 15 shows the slip rates derived for both the PCAIM and ENIF models for the same time periods as those chosen by *McGuire and Segall* [2003]. Additionally, Figure 16 shows the cumulative slip models over the full 115 day period corresponding to the two models. These models have similar cumulative slip potencies of 7.0×10^8 m^3 for the PCAIM model and 8.9×10^8 m^3 for the
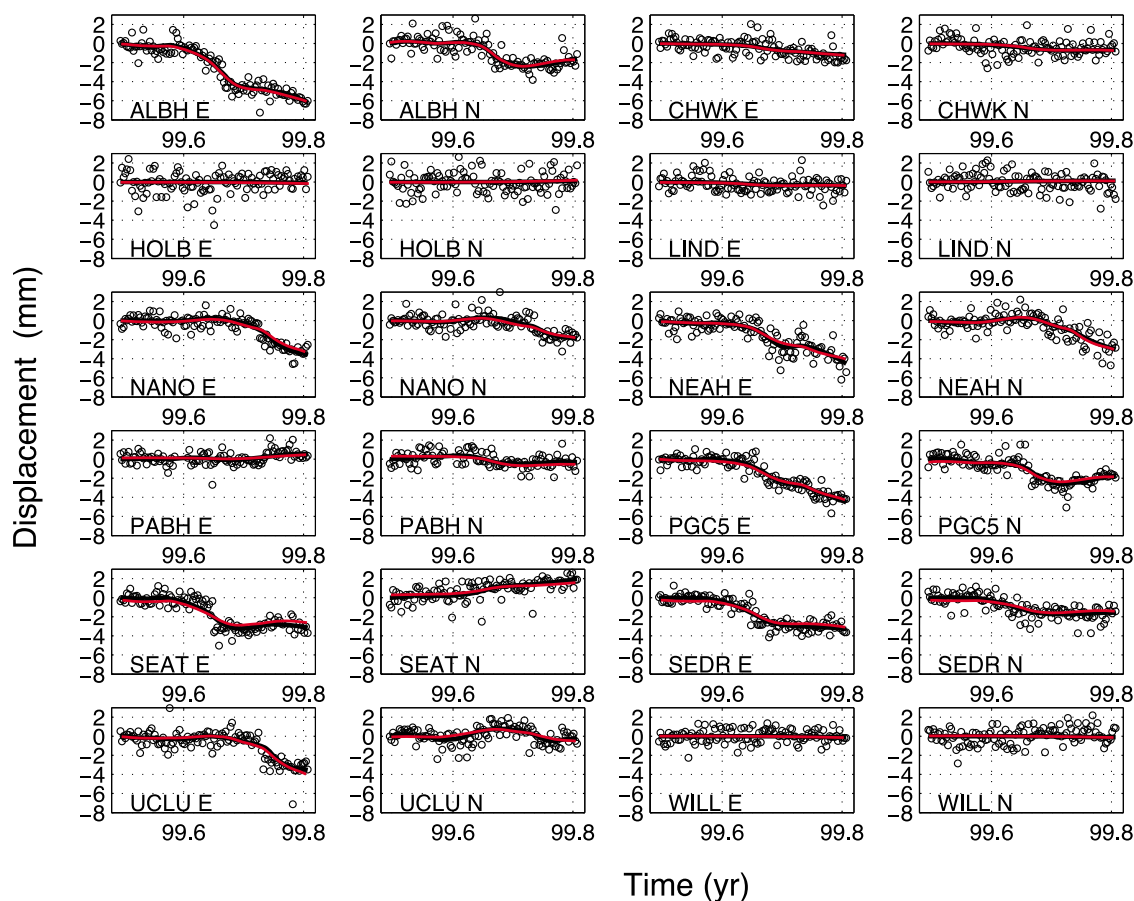
Figure 14. Continuous GPS records of 1999 Cascadia slow earthquake. Plots show east and north displacements recorded at 12 stations (located in Figure 15) over a 115 day period spanning the event. Dots are time series of *McGuire and Segall* [2003] in which contributions from secular velocity, reference frame errors, and benchmark wobble have been removed. Black solid lines show predictions of best fitting slip model determined from ENIF inversion of these data. Red solid lines show theoretical displacements predicted from PCAIM slip model derived from inversion of times series predicted by ENIF model. First three principal components were retained.

ENIF model and they both suggest two subevents with similar latitudinal location and extent. Slow slip first started in the south and then moved to the north. They do, however, differ with regard to the detail of the slip distribution, in particular the depth range. Slip is shallower in the PCAIM model than in the ENIF model. We interpret this difference as due different methods use to regularize the inversions and which therefore tend to favor different solutions, but equivalent in terms of the fit to the data. This test shows that a source model as complicated as the one proposed for the 1999 Cascadia slip event can be reproduced from PCAIM. However, in such a low signal-to-noise case, it is necessary to appropriately filter the correlated noise in the GPS times series, as can be done with filtering tools built-in to ENIF [*McGuire and Segall*, 2003].

6. Discussion: Performance, Limitations and Extensions of PCAIM

[49] Our study demonstrates the validity of the PCAIM approach and that it yields results comparable to the ENIF method. In addition, PCAIM is more flexible since it can be used to invert any kind of temporally varying fault slip including coseismic, postseismic, and transient aseismic slip events, and it is computationally very efficient. The total number of independent variables for the r component PCAIM algorithm on a data set of m GPS stations on a fault with N subpatches recording for n epochs is $r(3m + n)$ (r components × [3 per station +1 per change in epoch +1 singular value]), whereas for ENIF it is $3m(n - 1)$ (three per station per change in epoch). This enumeration ignores the hyperparameters as the number of these is small and does not scale with the size of the data set. Because r is typically less than 5, for even moderately small values of n and m PCAIM represents a great reduction over an epoch-to-epoch type of inversion in the number of parameters necessary to describe slip on the fault.

[50] The method we describe above is based on a standard SVD and can therefore only deal with data that are sampled at each station at each epoch and assumed to have equal uncertainties. While the epochs do not have to be evenly spaced, inverting jointly the daily GPS and single epoch measurements from campaign GPS measurements, is not possible this way. However, the technique can be adapted to allow this kind of joint inversion. A partial solution is employing weighted SVD if each component
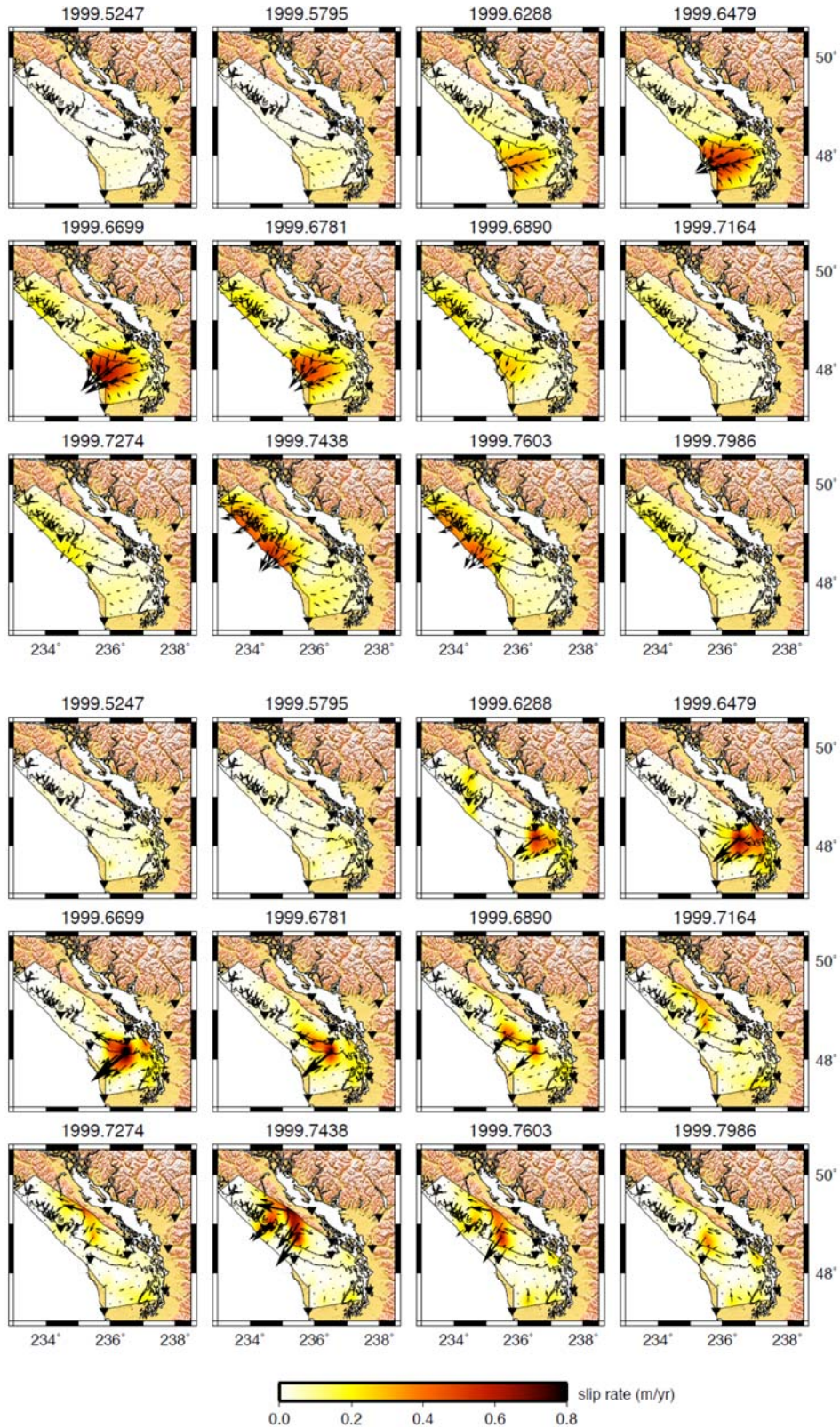
Figure 15. Time evolution of slip rate distribution corresponding ENIF slip model of (bottom) *McGuire and Segall* [2003] and (top) to PCAIM model obtained in this study. Black triangles show locations of GPS stations.

PCAIM (this study)

ENIF (McGuire and Segall, 2003)

slip rate (m/yr)

0   2   4   6   8
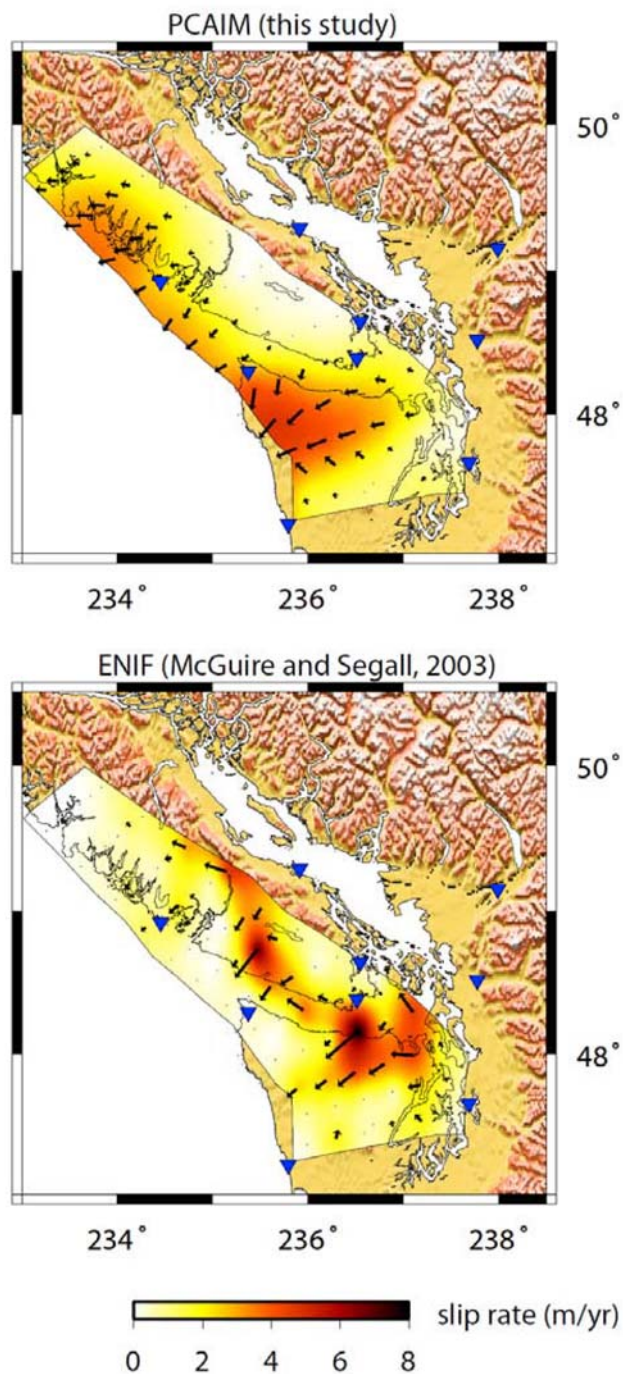
Figure 16. Cumulative slip distribution of ENIF slip model (top) of *McGuire and Segall* [2003] and (bottom) of PCAIM model obtained in this study, corresponding to 115 days covered by cGPS data of Figure 14. Blue triangles show locations of GPS stations.

at a given station has relatively constant error with time. A better option is to replace the standard SVD with a more sophisticated decomposition. In that regard, the decomposition of *Srebro and Jaakkola* [2003] is particularly adapted to take into account individual measurement errors and deal with missing data points. In this approach the data matrix X is decomposed into components U, S, V, with each data point weighted according to the square of its standard error.

This decomposition allows us to appropriately take into account formal uncertainties on the data as well as deal with missing data by assigning them infinite standard errors. For joint GPS and InSAR time series, we can assign errors of infinity to periods at which we have no InSAR scenes. This alternative decomposition, although not used in the present study, is implemented in the PCAIM code available from the Tectonics Observatory web site.

[51] Instead of inverting each component separately, we can construct a large block-diagonal system of equations to simultaneously invert any number of components we wish. As long as no additional linear constraints are added, the block-diagonal construction does not increase the computational complexity. However, if we are willing to increase the computational time of the algorithm, we can add linear equality or inequality constraints to the model. For example, these could include adding additional data points at particular times and locations not included in X or impose a positivity constraint on the model.

[52] The analysis of the Nias earthquake postseismic data set shows a significant amount of highly correlated noise is incorporated into the first several components of the PCAIM Nias model (for example, time function 2 in Figure 10). This noise can be smoothed out through temporal smoothing of the time function associated with each principal component. This procedure should, however, be handed with caution since it can lead to filtering out meaningful signal as well as noise. Another option would be to preprocess the time series to remove the correlated noise within the network, as done in the ENIF procedure.

7. Conclusion

[53] Using synthetic tests and comparison to current methods, we have demonstrated the effectiveness of the PCAIM algorithm for inverting geodetic time series for temporal variations of fault slip at depth. Several main advantages of PCAIM over other existing methods are: it is computationally efficient; it can deal with complex temporal variations and irregularly spaced sampling epochs; and it does not require adjustable parameters other than the number of components and those necessary to regularize the inversion (here the weight on Laplacian-based smoothing of the slip distribution). We have roughly outlined several extensions and variations of the PCA method described above that can make PCAIM an efficient and versatile inversion algorithm capable of handling missing data, and multiple types of data on multiple timescales.

Appendix A: Proofs

[54] Consider a slip distribution a assumed to be a solution of $G_\alpha a = d$ where d is known data. Assume that epoch-by-epoch inversion is valid and the solution is unique. The results still holds when there is a nontrivial subspace of solutions, but we stick to the unique solution case for clarity of presentation. We prove here the two following propositions:

[55] 1. As $r \rightarrow rank(X)$ PCAIM gives the same result as day-by-day inversion.

[56] 2. For each $1 \leq k \leq rank(X)$, the kth principal slip distribution exists and is a linear combination of the slip distributions derived from epoch-by-epoch inversions.

[57] We will demonstrate the latter result first.

[58] We show that each left singular vector is the sum of the columns of X, then based on the assumption that epoch-by-epoch inversion has a solution we show that there exists a solution to $G_\alpha a = x_i$. We conclude that each equation of the form $G_\alpha a = u_i$ has a solution and thus there is at least one slip distribution corresponding to each left singular vector.

[59] Consider the full principal component decomposition of

$$X_{m \times n} = U_{m \times m} S_{m \times n} V^t_{n \times n}. \tag{A1}$$

From linear algebra we know that U and V are of full rank and thus invertible, so

$$X_{m \times n} V_{n \times n} = U_{m \times m} S_{m \times n} V^t_{n \times n} V_{n \times n} = U_{m \times m} S_{m \times n}. \tag{A2}$$

Since S is a diagonal matrix, equation A2 tells us that each column of U multiplied by a constant is the sum of weighted averages of the columns of X. In particular, for $S(i, i) \neq 0$,

$$u_i S(i, i) = \sum_{k=1}^{n} x_k v(k, i) \tag{A3}$$

$$u_i = \sum_{k=1}^{n} x_k \frac{v(k, i)}{S(i, i)}. \tag{A4}$$

Because we assume that the epoch-by-epoch inversions are valid, we know for each $2 \leq k \leq rank(X)$, the equation $G_\alpha a_k = x_k - x_{k-1}$ has a solution a_k. By the definition of linearity of $G\alpha$, it's clear that $G_\alpha(b \cdot a_k + c \cdot a_{k'}) = bG_\alpha(a_k) + cG_\alpha(a_{k'})$. By induction on the number of terms in parentheses, we can find a slip distribution resulting in surface displacement equal to any linear combination of $\{x_k - x_{k-1}\}$. In particular for any $k' \leq k$, $(x_k - x_{k-1}) + (x_{k-1} - x_{k-2}) + \ldots + (x_{k'}+1 - x_{k'}) = x_k - x_{k'}$ and $-(x_k - x_{k'}) = x_{k'} - x_k$ are both in the span of these epoch-by-epoch surface displacements. Thus for any k', k, $x_{k'} - x_k$ has a corresponding slip distribution.

[60] Recall X is centered and in particular this implies

$$\sum_{j=1}^{n} x_j = 0. \tag{A5}$$

We divide by n and subtract x_k to each side and see the follow statements are equivalent:

$$\sum_{j=1}^{n} x_j = 0, \tag{A6}$$

$$\frac{1}{n} \sum_{j=1}^{n} x_j = 0, \tag{A7}$$

$$-x_k + \frac{1}{n} \sum_{j=1}^{n} x_j = -x_k, \tag{A8}$$

$$\frac{1}{n} \sum_{j=1}^{n} x_j - x_k = -x_k, \tag{A9}$$

$$\frac{1}{n} \sum_{j=1}^{n} G_\alpha(a_j - a_k) = -x_k. \tag{A10}$$

We combine equations (A10) and (A4) to find a weighted sum of slip distributions from epoch-by-epoch inversions which results the displacement u_i at the surface:

$$u_i = \sum_{k=1}^{n} x_k \frac{v(k, i)}{S(i, i)}, \tag{A11}$$

$$u_i = \sum_{k=1}^{n} \left(-\frac{1}{n} \sum_{j=1}^{n} G_\alpha(a_j - a_k) \right) \frac{v(k, i)}{S(i, i)}, \tag{A12}$$

$$u_i = G_\alpha \left[\sum_{k=1}^{n} \left(-\frac{1}{n} \sum_{j=1}^{n} (a_j - a_k) \right) \frac{v(k, i)}{S(i, i)} \right]. \tag{A13}$$

[61] Since $r = \text{rank}(X)$ implies $X = U_r S_r V'_r$, it is clear that the columns of X and the columns of U span the same space. The previous equations demonstrate each column of U has an associated slip distribution. Thus we can write a slip solution to any linear combination of the column of X (in particular the epoch-by-epoch input data $x_k - x_{k-1}$) in terms of a linear combination of the slip solutions of U. Thus since the slip solution to $x_k - x_{k-1}$ is unique, we arrive at the same result regardless of whether we apply PCAIM or epoch-by-epoch inversion to X. This ends our proof of (2).

[62] Let us assume that it is valid to model the epoch-by-epoch displacements as resulting from fault slip at depth. Here show that, if $r = rank(X)$, PCAIM generates the same results as if we had performed an inversion of the displacement between epochs $i \leq j$.

[63] As the rank of X is r, we know that $X = U_r S_r V_r^t$ holds exactly. By the previous proof, we know that each column vector in U_r has an inverse. That is, there exists a matrix L_r such that $G_\alpha \cdot L_r = U_r$. This implies that $X = G_\alpha \cdot L_r S_r V_r^t$ holds exactly. Thus we have a slip distribution $L_r S_r V_r^t$ that describes the incremental slip along the fault at each epoch. But by the assumption of uniqueness of the epoch-by-epoch inversion solutions $L_r S_r V_r^t$ must be the same as the epoch-by-epoch slip distribution. This concludes our proof of (1.).

References

Aoki, Y., and C. H. Scholz (2003), Interseismic deformation at the Nankai subduction zone and the Median Tectonic Line, southwest Japan, *J. Geophys. Res.*, *108*(B10), 2470, doi:10.1029/2003JB002441.

Briggs, R. W., et al. (2006), Deformation and slip along the Sunda Megathrust in the great 2005 Nias-Simeulue earthquake, *Science*, *311*, 1897–1901, doi:10.1126/science.1122602.

Cohen, S. C. (1999), Numerical models of crustal deformation in seismic zones, *Adv. Geophys.*, *41*, 134–231.

Dragert, H., K. Wang, and T. S. James (2001), A silent slip event on the deeper Cascadia subduction interface, *Science*, *292*, 1525–1528, doi:10.1126/science.1060152.

Fluck, P., R. D. Hyndman, and K. Wang (1997), Three-dimensional dislocation model for great earthquakes of the Cascadia subduction zone, *J. Geophys. Res.*, *102*(B9), 20,539–20,550, doi:10.1029/97JB01642.

Fukuda, J., et al. (2008), Geodetic inversion for space-time distribution of fault slip with time-varying smoothing regularization, *Geophys. J. Int.*, *173*(1), 25–48, doi:10.1111/j.1365-246X.2007.03722.x.

Hsu, Y. J., M. Simons, J. P. Avouac, J. Galetzka, K. Sieh, M. Chlieh, D. Natawidjaja, L. Prawirodirdjo, and Y. Bock (2006), Frictional afterslip following the 2005 Nias-Simeulue earthquake, Sumatra, *Science*, *312*, 1921–1926, doi:10.1126/science.1126960.

Kawamura, M., and K. Yamaoka (2006), Spatiotemporal characteristics of the displacement field revealed with principal component analysis and the mode-rotation technique, *Tectonophysics*, *419*(1–4), 55–73, doi:10.1016/j.tecto.2006.03.019.

Kawamura, M., and K. Yamaoka (2009), Temporal relationship between the 2000 Miyake-Kozu seismovolcanic activity and the 2000 Tokai slow-slip event, *Tectonophysics*, *465*(1–4), 45–59, doi:10.1016/j.tecto.2008.10.009.

Lisowski, M., J. C. Savage, and W. H. Prescott (1991), The velocity field along the San Andreas Fault in central and southern California, *J. Geophys. Res.*, *96*, 8369–8389, doi:10.1029/91JB00199.

Massonnet, D., and K. L. Feigl (1998), Radar interferometry and its application to changes in the Earth's surface, *Rev. Geophys.*, *36*, 441–500, doi:10.1029/97RG03139.

McGuire, J. J., and P. Segall (2003), Imaging of aseismic fault slip transients recorded by dense geodetic networks, *Geophys. J. Int.*, *155*, 778–788, doi:10.1111/j.1365-246X.2003.02022.x.

Miller, M. M., D. J. Johnson, C. M. Rubin, H. Dragert, K. L. Wang, A. Qamar, and C. Goldfinger (2001), GPS-determination of along-strike variation in Cascadia margin kinematics: Implications for relative plate motion, subduction zone coupling, and permanent deformation, *Tectonics*, *20*(2), 161–176, doi:10.1029/2000TC001224.

Miyazaki, S., P. Segall, J. Fukuda, and T. Kato (2004), Space time distribution of afterslip following the 2003 Tokachi-oki earthquake: Implications for variations in fault zone frictional properties, *Geophys. Res. Lett.*, *31*, L06623, doi:10.1029/2003GL019410.

Natawidjaja, D. H., K. Sieh, S. N. Ward, H. Cheng, R. L. Edwards, J. Galetzka, and B. W. Suwargadi (2004), Paleogeodetic records of seismic and aseismic subduction from central Sumatran microatolls, Indonesia, *J. Geophys. Res.*, *109*, B04306, doi:10.1029/2003JB002398.

Okada, Y. (1985), Surface deformation to shear and tensile faults in a half-space, *Bull. Seismol. Soc. Am.*, *75*, 1135–1154.

Perfettini, H., and J.-P. Avouac (2004), Postseismic relaxation driven by brittle creep: A possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the Chi-Chi earthquake, Taiwan, *J. Geophys. Res.*, *109*, B02304, doi:10.1029/2003JB002488.

Press, W., A. Teukolsky, W. Vetterling, and B. Flannery (2002), *Numerical Recipes in C: the Art of Scientific Computing*, Cambridge Univ Press, Cambridge, U. K.

Savage, J. C. (1983), A dislocation model of strain accumulation and release at a subduction zone, *J. Geophys. Res.*, *88*, 4984–4996, doi:10.1029/JB088iB06p04984.

Savage, J. C. (1988), Principal Component Analysis of geodetically measured deformation in Long Valley caldera, Eastern California, 1983–1987, *J. Geophys. Res.*, *93*(B11), 13,297–13,305, doi:10.1029/JB093iB11p13297.

Savage, J. C., and J. L. Svarc (1997), Postseismic deformation associated with the 1992 M_w = 7.3 Landers earthquake, southern California, *J. Geophys. Res.*, *102*, 7565–7578, doi:10.1029/97JB00210.

Segall, P., and J. L. Davis (1997), GPS applications for geodynamics and earthquake studies, *Annu. Rev. Earth Planet. Sci.*, *25*, 301–306, doi:10.1146/annurev.earth.25.1.301.

Segall, P., and M. Matthews (1997), Time dependent inversion of geodetic data, *J. Geophys. Res.*, *102*, 22,391–22,409, doi:10.1029/97JB01795.

Sieh, K., S. N. Ward, D. Natawidjaja, and B. W. Suwargadi (1999), Crustal deformation at the Sumatran subduction zone revealed by coral rings, *Geophys. Res. Lett.*, *26*, 3141–3144, doi:10.1029/1999GL005409.

Srebro, N. N., and T. Jaakkola (2003), Weighted low-rank approximations, paper presented at 20th International Conference on Machine Learning, Assoc. for the Adv. of Art. Intell., Washington, D. C.

Taylor, F. W., C. Frohlich, J. Lecolle, and M. Strecker (1987), Analysis of partially emerged corals and reef terraces in the central Vanuatu arc: Comparison of contemporary coseismic and nonseismic with Quaternary vertical movements, *J. Geophys. Res.*, *92*, 4905–4933, doi:10.1029/JB092iB06p04905.

Zhang, J., et al. (1997), Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities, *J. Geophys. Res.*, *102*, 18,035–18,055, doi:10.1029/97JB01380.

_____

J.-P. Avouac (corresponding author), Tectonics Observatory, Division of Geological and Planetary Sciences, California Institute of Technology, Mail Code 100-23, Pasadena, CA 91125, USA. (avouac@gps.caltech.edu)

A. P. Kositsky, Ashima Research, 600 S. Lake Ave., Pasadena, CA 91106, USA. (apk@ashimaresearch.com)